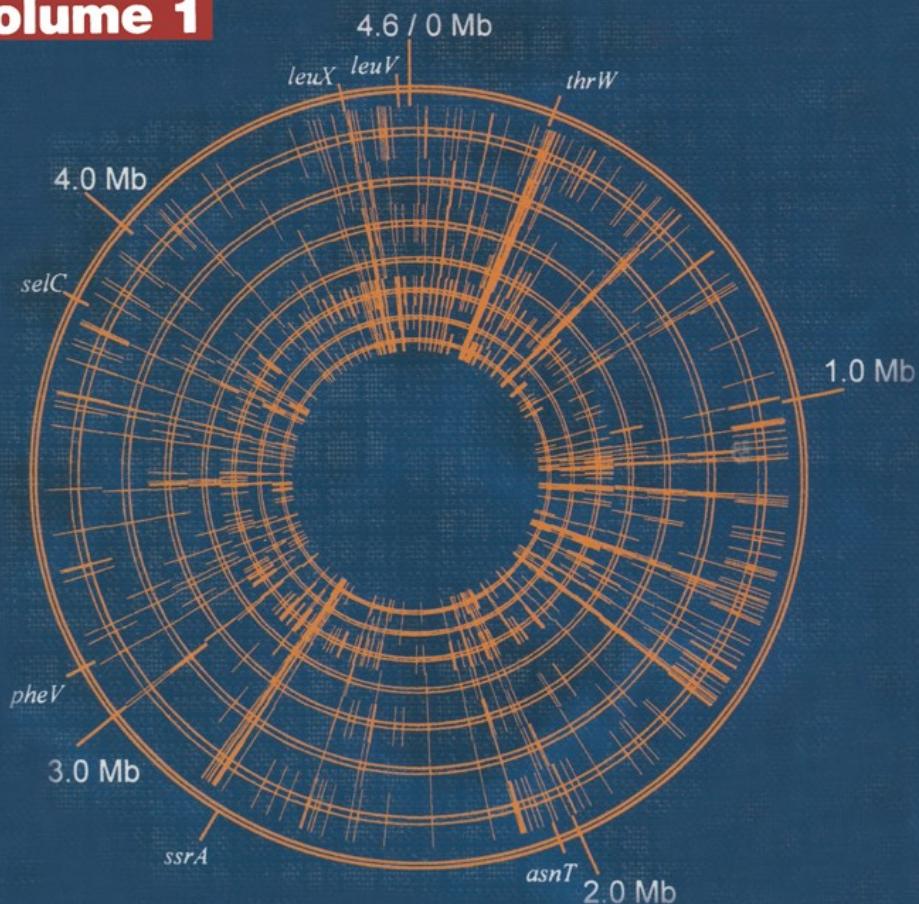


J. Hacker J. B. Kaper (Eds.)

# Pathogenicity Islands and the Evolution of Pathogenic Microbes

**Volume 1**



Springer

# Current Topics in Microbiology 264/I and Immunology

## Editors

R.W. Compans, Atlanta/Georgia

M. Cooper, Birmingham/Alabama · Y. Ito, Kyoto

H. Koprowski, Philadelphia/Pennsylvania · F. Melchers, Basel

M. Oldstone, La Jolla/California · S. Olsnes, Oslo

M. Potter, Bethesda/Maryland

P.K. Vogt, La Jolla/California · H. Wagner, Munich

Springer-Verlag Berlin Heidelberg GmbH

# **Pathogenicity Islands and the Evolution of Pathogenic Microbes**

**Volume I**

Edited by J. Hacker and J.B. Kaper

With 34 Figures and 7 Tables



**Springer**

Professor Dr. JÖRG HACKER  
Universität Würzburg  
Institut für Molekulare Infektionsbiologie  
Röntgenring 11  
97070 Würzburg  
Germany  
*e-mail:* j.hacker@mail.uni-wuerzburg.de

Professor Dr. JAMES B. KAPER  
University of Maryland School of Medicine  
Center for Vaccine Development  
685 W. Baltimore Street  
21201 Baltimore, MD  
USA  
*e-mail:* jkaper@umaryland.edu

*Cover Illustration:* Relative position on the *S. enterica* serovar *Typhimurium* LT-2 genetic map of virulence determinants which are absent from *E. coli* K-12. The *large circle* represents the chromosome, the *small circle* the virulence plasmid, pSLT. The positions of 0 and 100min are indicated. The *sspH1* and *sopE1* genes are absent from strain LT-2 but are present in other serovar *Typhimurium* isolates. (MIAO et al. 1999; MIROLD et al. 1999).

ISSN 0070-217X  
ISBN 978-3-642-07656-5 ISBN 978-3-662-09217-0 (eBook)  
DOI 10.1007/978-3-662-09217-0

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag Berlin Heidelberg GmbH.

Violations are liable for prosecution under the German Copyright Law.

<http://www.springer.de>  
© Springer-Verlag Berlin Heidelberg 2002  
Originally published by Springer-Verlag Berlin Heidelberg New York in 2002  
Softcover reprint of the hardcover 1st edition 2002  
Library of Congress Catalog Card Number 15-12910

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Product liability: The publishers cannot guarantee the accuracy of any information about dosage and application contained in this book. In every individual case the user must check such information by consulting other relevant literature.

Cover Design: *design & production GmbH*, Heidelberg  
Typesetting: Scientific Publishing Services (P) Ltd, Madras  
Production Editor: Angélique Gcota  
Printed on acid-free paper SPIN: 10765335 27/3020 - 5 4 3 2 1 0

## Preface

In the year 1972 the famous evolutionary biologist Theodor Dobzhansky wrote that “nothing makes sense in biology except in the light of evolution”. This sentence holds true even for the new developments in molecular biology: recombinant DNA technology, cloning of prokaryotic and eukaryotic genes, establishment of whole genome sequences, and reprogramming of cellular processes following transfer of chromosomes.

The evolutionary aspects of modern biology are of particular importance when interpreting the results from the different genome sequencing projects. As of the year 2001 more than 45 bacterial genomes have been sequenced, including more than 30 genomes of pathogenic bacteria such as *Helicobacter pylori*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus*, or pathogenic variants of the species *Escherichia coli*. The interpretation of these sequences reveal that the Darwinian laws of evolution – establishment of genetic variability, formation of new phenotypes, and natural selection of the newly formed variants – are also true for both pathogenic and nonpathogenic microbes. The ongoing genome sequencing projects as well as genetic studies performed over the last 10–15 years, however, also provided new insights into the mechanisms which led to the generation of new genetic variants. It became clear that, in addition to the generation of point mutations and the processes of genetic rearrangements, horizontal gene transfer plays a tremendous role in the evolution of prokaryotes and, presumably, also in the development of eukaryotic organisms. This volume of *Current Topics in Microbiology and Immunology* entitled “Pathogenicity Islands and the Evolution of Pathogenic Microbes” also focuses on these general mechanisms of the evolution of prokaryotes. The genetic processes which drive the evolution of bacteria are summarized in the first chapter of the book “Evolution of Prokaryotic Genomes”, written by W. Arber.

A few years ago one of the great pioneers of molecular pathogenesis research, S. Falkow, postulated that lateral gene transfer processes in bacteria lead to “evolution in quantum

leaps". Our publication aims to summarize the current knowledge on the impact of lateral gene transfer in the evolution of pathogenic microbes, with a particular focus on new genetic elements, termed pathogenicity islands (PAI). PAIs were first described more than 10 years ago for pathogenic variants of the species *Escherichia coli*. PAIs have now been described for more than 30 species of pathogenic microbes and the ongoing discussion of the definition of PAIs and their significance in microbial evolution is also reflected in our book. From our point of view, PAIs represent formerly transferred or still mobile genetic elements which encode virulence factors and are present in pathogenic bacteria, but not present in related nonpathogenic specimens. PAIs are often flanked by particular boundary regions (direct repeat sequences, IS elements); they are often located near tRNA genes and often unstable. PAIs encode so-called virulence-associated genes, which are parts of PAIs together with "mobility genes" encoding transposases, integrases, or other enzymes active in genetic recombination. As the first PAIs have been described in pathogenic strains of *Escherichia coli*, the two chapters by Redford and Welch and by Torres and Kaper, entitled "Extraintestinal *Escherichia coli* As a Model System for the Study of Pathogenicity Islands" and "Pathogenicity Islands of Intestinal *E. coli*" summarize our knowledge on the occurrence and distribution of PAI in extraintestinal and intestinal *E. coli*. The newly established DNA sequence of the enterohemorrhagic *E. coli* O157:H7 isolate also contributes to the discussion on PAIs in pathogenic *E. coli*.

It has been known for a number of years that not only PAIs but also plasmids and bacteriophages are able to carry genes whose products are involved in pathogenic processes. Accordingly, such elements and their products play an important role in pathogenesis due to intestinal *E. coli* as well due to *Shigella*. The role of PAIs in shigellosis is summarized in the article "Pathogenicity Islands of *Shigella*" by Ingersoll et al. Besides *E. coli* and the closely related species of the *Shigella* group, additional members of the enterobacteria such as *Yersinia*, *Salmonella*, and *Erwinia* are described in various contributions. Thus, in the article by Kingley and Bäumler "Pathogenicity Islands and Host Adaptation of *Salmonella* Serovars" the different genetic elements leading to the various variants of *Salmonella* are discussed. In this respect it seems useful to distinguish between larger "islands" which exhibit a number of specific features (see above) and smaller genetic regions termed "islets". PAI and islets as well as bacteriophages and even plasmids all play a role in the formation of lipopolysaccharides (LPS) of gram-negative bacteria and

capsules, which are produced by a variety of bacteria, both gram-negative and gram-positive. While Reeves and Wang describe the genomic organization of LPS-specific loci in their chapter, the one by Barrett et al. deals with the genomic structure of capsular determinants. Both describe common structures of LPS and capsules in different pathogens but also focus on the differences among various species and the genetic basis of these cell surface components. In both of these chapters, LPS and capsular-specific gene clusters are considered as “genetic modules” which have the capacity to recombine new variants of the cell surface. Vogel and Frosch in their contribution “The Genus *Neisseria*: Population, Structure, Genome Plasticity and Evolution of Pathogenicity” consider the respective loci as part of plasticity regions in the genomes of pathogenic *Neisseria* which may be introduced into new recipients by transformation rather than by transduction with phages or by plasmid-driven conjugation.

Another interesting aspect which is reflected in various chapters is that genomes evolve by acquisition of new pieces of DNA following gene transfer, and also by genome reduction. Different mechanisms include the deletion of sequences or the elimination of functions by the accumulation of point mutations or rearrangements. In the contribution by Carniel “Plasmids and Pathogenicity Islands of *Yersinia*” the importance of mobile genetic elements for the pathogenic phenotypes of *Yersinia* is described. Here, a reduction of functions is considered to be a prerequisite for the evolution of the very dangerous bacterial species *Yersinia pestis* from *Y. pseudotuberculosis* over the last 2000–10,000 years. In addition, the genomes of pathogenic shigellae exhibit deleted regions compared to those of *E. coli*, termed “black holes”, whose absence is required for a fully virulent phenotype. In their chapter “Phylogenetic Relationships and Virulence Evolution in the Genus *Bordetella*” von Wintzingerode and coworkers provide evidence that the genome of the causative agent of whooping cough, *Bordetella pertussis*, is 100kb smaller than the genome of the less pathogenic species *Bordetella bronchiseptica*, another example of genome evolution by reduction.

It has been known for a long time that the occurrence of bacteriophages and plasmids is not restricted to pathogenic bacteria. Rather such genetic elements are quite widespread in the prokaryotic world. The same holds true for elements which exhibit a general composition which is similar to that of PAI. Thus, J. Rood in the contribution “Genomic Islands of *Dichelobacter nodosus*” describes the genome structure of an animal pathogen, in which he was able to identify DNA elements very similar to PAIs, but with unknown pathogenic properties. In addition,

Labig et al. in their chapter “Pathogenicity Islands and PAI-like Structures in *Pseudomonas* Species” describe genomic islands in various species of the *Pseudomonas* group, where their products may contribute to the pathogenic potency of *Pseudomonas aeruginosa*, but also to the metabolic diversity of nonpathogenic *Pseudomonas* species such as *Pseudomonas putida*. Furthermore, the chapter by Dobrindt et al.. “Genome Plasticity of Pathogenic and Nonpathogenic Enterobacteria” demonstrates that symbiotic bacteria may also carry elements which correspond to genomic islands. Kim and Alfano demonstrate in their work “Pathogenicity Islands and Virulence Plasmids of Bacterial Plant Pathogens” that plant pathogens also carry such genetic elements.

The general mechanism of the type III secretion system of delivering molecules from bacteria to host cells is not restricted to pathogens either. Thus, type III-specific secretion systems are produced by pathogens such as intestinal *E. coli*, *Yersinia*, *Shigella*, and *Salmonella* but also by different plant pathogens and by symbiotic *Pseudomonas* bacteria. Furthermore, type IV pathways are expressed by *Helicobacter pylori*, *Legionella pneumiae*, and *Agrobacterium tumefaciens*. The chapters by Odenbreit and Hass “*Helicobacter pylori*: Impact of Gene Transfer and the Role of the cag Pathogenicity Island for Host Adaptation and Virulence” and Heuner et al. “Genome Structure and Evolution of *Legionella* Species” present examples of the occurrence, structure, and function of type IV secretion systems in various human pathogens. As demonstrated by Kim and Alfano, similar systems exist in the plant pathogen *Agrobacterium tumefaciens* which has the capacity to transfer DNA from bacteria to plant cells. This mechanism reflects another essential mechanism in the evolution of host-pathogen relationship.

Another function that is mediated by horizontal gene transfer is antibiotic resistance. As shown in the chapters “Phages and Other Mobile Virulence Elements in Gram-positive pathogens” by Gentry-Weeks et al. and “Impact of Integrons and Transposons on the Evolution of Resistance and Virulence” by Rowe- Magnus et al. various genetic elements, such as plasmids, “islands”, transposons, integrons and “super integrons” play a role in the distribution of resistance determinants. In particular cases, resistance- and virulence-specific genes can be part of the same genetic entity, which speaks for a co-evolution of the two properties.

The fact that PAIs and PAI-like elements such as plasmids and phages are constitutive parts of the genomes of gram-positive pathogens is supported by the data given in the chapter by Gentry-Weeks and also by the contributions “Pathogenicity

Islands and Other Virulence Elements in *Listeria*" by Kreft et al. and "Genome Structure and Evolution of the *Bacillus cereus* Group" by Kolsto et al.. Both groups of organisms, *Listeria* and *Bacillus*, comprise pathogenic as well as nonpathogenic members and the respective genetic elements – phages, plasmids, and islands – truly behave as pathogens or as commensals of the normal microbial flora. Furthermore, we propose that genetic elements such as genomic islands may also be present in eukaryotes. Köhler et al. in "Genome Structure of Pathogenic Fungi" discuss new data which support this view.

In this volume of *Current Topics in Microbiology and Immunology* a number of genome structures of pathogenic microbes are presented. It will be our future task to understand the functions of the known genes and their products and to identify new ones as more sequence information becomes available. This will enable us to speculate on the processes of natural selection, which, in consequence, lead to the development of new species and pathotypes. The Darwinian laws will need to be considered when new data on these processes are obtained in order to demonstrate that PAIs and other genetic elements really "make sense in the light of evolution".

J. HACKER and J.B. KAPER

## List of Contents

<b>W. ARBER</b>	
Evolution of Prokaryotic Genomes . . . . .	1
<b>P. REDFORD and R.A. WELCH</b>	
Extraintestinal <i>Escherichia coli</i> as a Model System for the Study of Pathogenicity Islands . . . . .	15
<b>A.G. TORRES and J.B. KAPER</b>	
Pathogenicity Islands of Intestinal <i>E. coli</i> . . . . .	31
<b>M. INGERSOLL, E.A. GROISMAN, and A. ZYCHLINSKY</b>	
Pathogenicity Islands of <i>Shigella</i> . . . . .	49
<b>R.A. KINGSLY and A.J. BÄUMLER</b>	
Pathogenicity Islands and Host Adaptation of <i>Salmonella</i> Serovars . . . . .	67
<b>E. CARNIEL</b>	
Plasmids and Pathogenicity Islands of <i>Yersinia</i> . . . . .	89
<b>P.P. REEVES and L. WANG</b>	
Genomic Organization of LPS-Specific Loci . . . . .	109
<b>B. BARRETT, L. EBAH, and I.S. ROBERTS</b>	
Genomic Structure of Capsular Determinants . . . . .	137
<b>U. DOBRINDT, U. HENTSCHEL, J.B. KAPER, and J. HACKER</b>	
Genome Plasticity in Pathogenic and Nonpathogenic Enterobacteria . . . . .	157
<b>F. VON WINTZINGERODE, G. GERLACH, B. SCHNEIDER, and R. GROSS</b>	
Phylogenetic Relationships and Virulence Evolution in the Genus <i>Bordetella</i> . . . . .	177
<b>K. LARBIG, C. KIEWITZ, and B. TÜMMLER</b>	
Pathogenicity Islands and PAI-Like Structures in <i>Pseudomonas</i> Species . . . . .	201
<b>Subject Index</b> . . . . .	213

# List of Contents of Companion Volume 264/II

S. ODENBREIT and R. HAAS <i>Helicobacter pylori</i> : Impact of Gene Transfer and the Role of the <i>cag</i> Pathogenicity Island for Host Adaptation and Virulence . . . . .	1
U. VOGEL and M. FROSCH The Genus <i>Neisseria</i> : Population Structure, Genome Plasticity, and Evolution of Pathogenicity . . . . .	23
J.I. ROOD Genomic Islands of <i>Dichelobacter nodosus</i> . . . . .	47
K. HEUNER, M. STEINERT, R. MARRE and J. HACKER Genomic Structure and Evolution of <i>Legionella</i> Species . . . . .	61
C. GENTRY-WEEKS, P.S. COBURN, and M.S. GILMORE Phages and Other Mobile Virulence Elements in Gram-Positive Pathogens . . . . .	79
A.-B. KOLSTØ, D. LERECLUS, and M. MOCK Genome Structure and Evolution of the <i>Bacillus cereus</i> Group . . . . .	95
J. KREFT, J.-A. VÁZQUEZ-BOLAND, S. ALTROCK, G. DOMINGUEZ-BERNAL, and W. GOEBEL Pathogenicity Islands and Other Virulence Elements in <i>Listeria</i> . . . . .	109
J.F. KIM and J.R. ALFANO Pathogenicity Islands and Virulence Plasmids of Bacterial Plant Pathogens . . . . .	127
G. KÖHLER, J. MORSCHHÄUSER, and J. HACKER Genome Structure of Pathogenic Fungi . . . . .	149
D.A. ROWE-MAGNUS, J. DAVIES, and D. MAZEL Impact of Integrons and Transposons on the Evolution of Resistance and Virulence . . . . .	167
Subject Index . . . . .	189

## List of Contributors

(Their addresses can be found at the beginning of their respective chapters.)

ARBER, W.	1	KINGSLEY, R.A.	67
BARRETT, B.	137	LARBIG, K.	201
BÄUMLER, A.J.	67	REDFORD, P.	15
CARNIEL, E.	89	REEVES, P.P.	109
DOBRINDT, U.	157	ROBERTS, I.S.	137
EBAH, L.	137	SCHNEIDER, B.	177
GERLACH, G.	177	TORRES, A.G.	31
GROISMAN, E.A.	49	TÜMMLER, B.	201
GROSS, R.	177	VON WINTZINGERODE, F.	
HACKER, J.	157		177
HENTSCHEL, U.	157	WANG, L.	109
INGERSOLL, M.	49	WELCH, R.A.	15
KAPER, J.B.	31, 157	ZYCHLINSKY, A.	49
KIEWITZ, C.	201		

## **List of Contributors of Companion Volume 264/II**

ALFANO, J.R.	127	KOLSTØ, A.-B.	95
ALTROCK, S.	109	KREFT, J.	109
COBURN, P.S.	79	LERECLUS, D.	95
DAVIES, J.	167	MARRE, R.	61
DOMINGUEZ-BERNAL, G.		MAZEL, D.	167
	109	MOCK, M.	95
FROSCH, M.	23	MORSCHHÄUSER, J.	149
GENTRY-WEEKS, C.	79	ODENBREIT, S.	1
GILMORE, M.S.	79	ROOD, J.I.	47
GOEBEL, W.	109	ROWE-MAGNUS, D.A.	
HAAS, R.	1		167
HACKER, J.	149, 61	STEINERT, M.	61
HEUNER, K.	61	VÁZQUEZ-BOLAND, J.-A.	
KIM, J.F.	127		109
KÖHLER, G.	149	VOGEL, U.	23

# Evolution of Prokaryotic Genomes

W. ARBER

1	Introduction . . . . .	1
2	The Thesis of the Existence of Evolution Genes . . . . .	2
3	A Look at the Definitions and at the Effects of Mutations . . . . .	3
4	Natural Strategies for the Generation of Genetic Variation . . . . .	4
4.1	Local Sequence Change . . . . .	4
4.2	DNA Rearrangement . . . . .	4
4.2.1	Transposition of Mobile Genetic Elements . . . . .	5
4.2.2	Site-Specific Recombination . . . . .	6
4.2.3	General Recombination . . . . .	7
4.3	DNA Acquisition . . . . .	7
4.3.1	Horizontal Gene Transfer . . . . .	8
4.3.1.1	DNA Transformation . . . . .	8
4.3.1.2	Bacterial Conjugation . . . . .	8
4.3.1.3	Virus-Mediated Transduction . . . . .	8
4.3.2	Barriers Against DNA Acquisition . . . . .	9
4.4	Qualitative Differences Between the Three Strategies for the Generation of Genetic Variants . . . . .	9
5	Updated View on Molecular Evolution . . . . .	10
5.1	Evolutionary Tree . . . . .	10
5.2	Coexistence in Bacterial Genomes of Genes Serving for the Benefit of Individuals and of Genes Serving Evolutionary Development . . . . .	11
5.3	Evolution Genes Are Proposed to be Submitted to Second-Order Selection . . . . .	11
5.4	Conclusions on the Elements of the Theory of Molecular Evolution . . . . .	12
	References . . . . .	13

## 1 Introduction

The oldest traces of life on earth go back to almost 4 billion years ago. These archaic forms of life were precursors of the prokaryotic unicellular organisms which today populate an extremely wide range of habitats. The diversity of prokaryotic organisms is accordingly enormous. This diversification finds an explanation in the evolutionary theory (WOESE 1987; BROCK et al. 1994).

The theory of biological evolution was developed largely on the basis of comparative anatomy and physiology of higher eukaryotic organisms. Paleontological studies on fossils thereby played an important role. Until recently, molecular evolution was hardly a topic of investigation by evolutionary biologists, and relatively little attention was given to microbial evolution. This situation has started to change with the introduction of molecular genetics, which has its roots in microbial genetics in the middle of the twentieth century. However, it was only in the 1990s that an increasing amount of DNA sequence data, including sequences of entire genomes, became available. Almost instantly, computerized sequence comparisons began to offer new insights into the evolutionary relatedness of single genes and of groups of genes with functional relations, and this has become one pillar of molecular evolution.

The second pillar of molecular evolution resides in the investigation of molecular processes leading to genetic variation. Genetic variation is the driving force of biological evolution, since populations formed by individuals with various types of genetic alterations together with their genetically unchanged ancestors form the substrate for natural selection, which is steadily exerted on these populations. Modern genetic methodologies render prokaryotic genomes accessible to studies of individual steps of spontaneous genetic variation. Bacterial genomes are particularly attractive for such studies not only because of their relatively small size, but also because of their haploid nature, procuring a rapid phenotypic manifestation of genetic alterations affecting phenotypic traits. In addition, the short generation times of bacteria, and thus the production of large populations in short periods of time, allow one to carry out competition experiments with mixed microbial populations under different life conditions. It is the aim of this article to review available data on the generation of genetic variation in bacteria and to draw relevant conclusions concerning the theory of molecular evolution.

## 2 The Thesis of the Existence of Evolution Genes

The availability of DNA sequences of entire genomes represents the basis for experimental investigations on the functional relevance of all the open reading frames contained on a genome. Such investigations, known as functional genomics and as proteomics, are targeted at finding out the functions exerted by the products of each of the genes carried in the genome. Many of these gene products will be found to carry out a more or less essential task relevant for the life of each cell in the population, if we define here life as bringing the unicellular organism from one generation to the next. We may wonder, however, if other bacterial genes rather fulfil evolutionary tasks, which could imply a role in the occasional generation of genetic variants. Such a function may not routinely be exerted in each cell going from one generation to the next, but rather in a small proportion of individuals of a cell population.

Biological evolution is an important process for the long-term maintenance, proliferation and diversification of life in an environment offering living conditions which are neither temporally nor spatially constant. Since nature has developed life, it might also take care of its evolutionary development. The thesis of the existence of specific genes serving the purposes of evolution of life is based on this reflection. Here we will first look for circumstantial evidence in favour of this hypothesis and then incorporate the thesis into the theory of molecular evolution. We will thereby encounter genetically encoded systems that actively generate genetic variants and other systems that reduce the efficiency of natural production of genetic variants, while some systems can be seen to do both.

### 3 A Look at the Definitions and at the Effects of Mutations

Classical genetics defines mutation as a sudden and stable alteration in a phenotypic trait of the organism. This change at the functional level, i.e. at the level of the gene product, becomes inherited by the progeny. Its cause can be located by genetic crosses on the genetic map of the organism, without a necessity to know the chemical nature of the carrier of genetic information. In contrast, molecular genetics defines a mutation as any sudden and stable alteration in the nucleotide sequence of the genome, whether this alteration does or does not affect a phenotypic trait of the concerned organism. Hence, this definition includes the mutations seen in classical genetics by their impact on the phenotype. It also includes all those alterations in genome sequences that have no direct influence on fitness, i.e. on the effect of natural selection under the relevant living conditions.

For our understanding of the molecular nature of processes contributing to natural mutagenesis, the molecular genetics definition of mutation is more relevant than the classical definition based on functional impacts. As a matter of fact, there is no good evidence available to suggest that mutagenesis events would, in general, either be limited to functional alterations or even more specifically be targeted to specific genes in need of being functionally adapted. Rather, mutagenesis can result from a number of quite different molecular processes which we are going to discuss in more detail below. Although the criteria for the use of a target for the mutation may vary for different mechanisms, all processes lead, as far as we know, only relatively rarely to a selective advantage, i.e. to an evolutionarily useful beneficial mutation. More frequent are mutations providing selective disadvantage and, in the extreme case causing lethality. Also relatively frequent are DNA sequence alterations that have no immediate influence on life processes, i.e. no effect on natural selection.

The generally observed prevalence of mutations with selective disadvantage over those providing a selective advantage represents an argument against a directedness of spontaneous mutations, and it requires a stringent limitation to mutation frequencies per genome. The latter proposal implies, as we will see later,

the need for efficient repair and restriction systems able to prevent certain mutagenesis processes from exerting their full potential activities.

## 4 Natural Strategies for the Generation of Genetic Variations

Data accumulated from microbial genetics allow us to discern three major strategies serving in nature for the generation of genetic variations in prokaryotes (ARBER 2000). These strategies are: (a) small local changes in the nucleotide sequences of a genome; (b) intragenomic reshuffling of DNA segments, sometimes accompanied by small local sequence changes; and (c) the acquisition of a segment of DNA originating in another more or less closely related organism. The latter process implies horizontal gene transfer and the uptake of transferred genetic information into the genome of the recipient bacterium.

### 4.1 Local Sequence Change

Local alterations of nucleotide sequences generally affect only one or a few adjacent positions in the inherited sequences of the genome. Alterations can be a nucleotide substitution, the insertion or the deletion of one or a few nucleotides, or a local scrambling of the inherited sequences. A number of causes for such changes have been identified and at least in part experimentally confirmed. The implied processes include replication infidelities, such as those attributed to short-lived tautomeric forms of the nucleotides and those due to replication slippage. Other causes are the intrinsic chemical instability of nucleotides, particularly of cytosine, or the effect of chemical mutagens and of some radiation mutagens.

This kind of sequence change can affect principally any position of a genome. However, some DNA regions have been shown to be more sensitive (hot spots) than others to this mutagenesis (DRAKE 1991). Since many of these mutations go back to intrinsic structural flexibility and stability properties of matter and to reactivity of matter to mutagenic stimuli, the frequency of basic events per genome is more or less given and is proportional to the size of the genome. However, efficient bacterial enzymatic repair systems keep the overall mutagenesis at low rates that are tolerable to the organism but evolutionarily useful. One can thus interpret the role of these repair systems as serving biological evolution, since they guarantee a certain degree of genetic stability in the context of natural mutagenesis, although they allow local mutagenesis to occur at low rates. This helps evolution to proceed steadily.

### 4.2 DNA Rearrangement

Early on, evolutionary biology identified two major sources of genetic diversity in eukaryotic organisms. One of these sources was seen in the already described

generation of local sequence changes. The other source is fully relevant only for sexually reproducible higher organisms. In these diploids the assortment of the parental chromosomes into the gametes and the general recombination between the homologous paternal and maternal chromosomes results at each generation in a novel distribution of pre-existing alleles. In haploid bacteria, where sexuality is not compulsory for reproduction, general recombination also occurs, but it does not play the same efficient role as in higher organisms to foster genetic diversity within a species. However, several other recombination processes have been thoroughly explored with bacteria, and, as we will outline below, these can also be interpreted as evolutionary functions.

#### 4.2.1 Transposition of Mobile Genetic Elements

The phenomenon of transposition, defined as the occasional enzyme-mediated change of the chromosomal location of a particular segment of the genome, was originally discovered in a plant: maize (MCCLINTOCK 1950). In the late 1960s this phenomenon was identified as also being the cause of bacterial mutations displaying somewhat unconventional properties, including polar effects (JORDAN et al. 1968; SHAPIRO 1969). In these and subsequent investigations it was rapidly seen that bacteria such as *Escherichia coli* strains contain in their genomes several different mobile genetic elements, called IS (inserted sequence) elements, each in one to several copies. Altogether, roughly 1% of the *E. coli* chromosome represents IS elements (IIDA et al. 1983).

Transposition of an IS element is mediated by the enzyme transposase, which is the product of a gene carried on the IS element itself. Each specific IS element has its own criteria for the selection of an insertion target, where either the maternal element itself or a copy thereof becomes integrated into DNA. In doing so, usually a short nucleotide sequence of the target becomes duplicated. Transposition is thus mutagenic in a double sense: it disrupts the inherited genetic information at the target, and it inserts a few additional nucleotides at the site of novel integration into the genome. IS and other mobile genetic elements also make important contributions to the DNA acquisition strategy of genetic variation (see Sect. 4.3).

Interestingly, IS transposition was shown to occur also in resting bacteria which were kept for long periods of time in stab cultures at room temperature (NAAS et al. 1994). Under these conditions, the number of transposition variations appearing in the surviving bacteria increased linearly with time over 30 years of storage (NAAS et al. 1995).

IS transposition can often lead to lethality, as shown by a study of P1 prophage mutants that had lost their ability to give rise to bacteriophage production upon induction (SENGSTAG and ARBER 1983). In this experiment 95% of the investigated lethal phage mutants were caused by the transposition of an IS element from a chromosomal location into the P1 prophage.

Several different genetically controlled strategies ensure that the event of transposition is rare (ARINI et al. 1997; MAHILLON and CHANDLER 1998). This

provides a reasonable genetic stability to the host bacterium, while the occasional transposition events meaningfully contribute to spontaneous mutagenesis serving the purposes of evolution. It can happen that subpopulations displaying much greater rates of transposition than usual are formed for limited times. This can give rise to bursts of transposition which can affect different sites on the genome (NAAS et al. 1994). This boosting effect in bacterial subpopulations found a mechanistic explanation by the temporal formation of particular sequence configurations containing structurally unstable IS dimers. This resulted in transposition frequencies several orders of magnitude higher than are normally seen for a single IS element (OLASZ et al. 1993; ARBER et al. 1994).

#### 4.2.2 Site-Specific Recombination

Many bacterial strains, as well as virus genomes and plasmids, are known to encode site-specific recombination systems. Typically, consensus DNA sequences of a length in the order of 20 nucleotides serve as sites of crossing over in the enzymatically mediated, relatively efficient recombination reaction. This process serves, for example, to integrate some temperate viruses as prophages into the host chromosome (WEISBERG and LANDY 1983). Integration can also occur, although with decreasing efficiency, into sequences deviating from the optimal consensus. Upon the site-specific recombinational interaction between two consensus sequences carried within the same DNA molecule the segment of DNA carried between the consensus sequences can become either deleted or inverted. This depends on the orientation of the crossing-over sites and often also on the intervention of a particular host protein which can drive the recombination into its DNA inversion mode. Examples of this kind of recombination are the flip-flop systems (GLASGOW et al. 1989). Besides rapidly and reproducibly bringing about mixed populations displaying two (or sometimes more) different genetic abilities, the role of DNA inversion can also be looked at with regard to its contribution to evolutionary development. This is based on the occasional use of one of a large number of so-called secondary crossing-over sites. These can deviate considerably from the consensus sequence (IIDA and HIESTAND-NAUER 1987). As a matter of fact, the degree of deviation influences the frequency of its use (ARBER 1991). Consequently, DNA inversion can occasionally bring about novel gene fusions and operon fusions. By chance, such fusions might be at the origin of either a novel, functionally relevant gene fusion or an improved gene expression control. Such events are quite rare and only statistically reproducible. This is precisely what gene products serving as generators of genetic variations are expected to do, namely, to occasionally produce genetic variants. This must occur inefficiently, nonreproducibly and without a fixed finality. The usefulness of novel sequences will, as in any event of spontaneous mutagenesis, be tested by natural selection. The involved recombination enzyme is a prototype of a genetic variation generator, and its evolutionary relevance should be obvious.

#### 4.2.3 General Recombination

Systems of general or homologous recombination are widespread in bacteria. They play a role in the repair of some damage caused to DNA, for example, by radiation. Typically, undamaged genomes can result from general recombination between homologous sequences of damaged sister chromosomes in a propagating cell. While bacteria normally have low basic levels of enzymes for general recombination, the efficiency of recombination can be boosted by SOS induction, which can, for example, result in response to UV irradiation (LITTLE and MOUNT 1982).

General recombination requires sequence homologies extending usually over 50 or more adjacent nucleotides. The shorter the homology is, the less efficient recombination will be. However, inefficiency is not an argument against the evolutionary relevance of recombinational events. Think, for example, of rare novel DNA sequence fusions that can occasionally be formed by intragenomic DNA inversion or deletion involving short segments of sequence homology. More frequent are events of recombination between extended DNA segments of homology, such as copies of a given IS element carried at different locations in the chromosome. Upon intramolecular recombination, this reaction can result in deletion as well as inversion. This process can also give rise to unequal crossing over between two sister chromosomes, resulting in either deletion or evolutionarily relevant gene duplication (IIDA et al. 1983).

### 4.3 DNA Acquisition

While the two already outlined natural strategies for the generation of genetic variations represent internal, intragenomic alterations in the inherited genome sequences, the third strategy resides in the uptake of a segment of DNA originating from the genome of another organism. This so-called donor organism can be more or less closely related genetically to the recipient bacterium. Processes of mobilisation of donor DNA from the donor genome often involving a viral or plasmid gene vector have been quite intensively studied, as have the actual processes of horizontal (or lateral) gene transfer and the incorporation of transferred DNA segments into the recipient genome (PORTER 1988). The three major mechanisms involved in horizontal gene transfer, DNA transformation (AVERY et al. 1944), bacterial conjugation (LEDERBERG 1947) and phage-mediated transduction (ZINDER and LEDERBERG 1952), were discovered about 50 years ago and form the fundaments of bacterial genetics. In the DNA acquisition strategy mobile genetic elements play a prominent role in the mobilisation of genes from the donor genome as well as in the subsequent fixation of transferred genes in the recipient genome, but other recombination processes are also involved for this purpose. DNA acquisition encounters a number of natural limitations, which can be seen as ensuring a reasonable degree of genetic stability to the recipient bacteria while allowing their population to profit from the possibility of trying out biological functions which have been developed by other organisms and

might be novel to the recipient. This is obviously an interesting evolutionary strategy.

### 4.3.1 Horizontal Gene Transfer

#### 4.3.1.1 DNA Transformation

Upon DNA transformation, fragments of free DNA originating from the donor bacterium are taken up into the recipient cell and later become incorporated into the recipient genome (HOTCHKISS and GABOR 1970). This sequence of events is usually required if selection is made for the phenotype of the acquired gene functions. Much is known about the individual steps, including the liberation of DNA fragments by the donor, about the penetration of a DNA fragment into the recipient cell and about the integration of acquired genetic information into the recipient genome, as well as about various obstacles to be overcome before a successful result is reached. Here, it should be remembered only that the state of competence for DNA uptake can vary considerably depending on the particular bacterial strain. For example, DNA uptake in the state of competence can be an active process helped by gene products of the recipient, or it can be more passive and depend rather on the physiological condition and on the impact of environmental factors. Remember, however, that for the purpose of evolutionary progress, rare events are of relevance.

#### 4.3.1.2 Bacterial Conjugation

Upon conjugation, two or more bacterial cells undergo intimate contact, and DNA becomes transferred in a polar manner from a donor into a recipient cell (JACOB and WOLLMAN 1961). A conjugative plasmid serves thereby as a mediator of the process and as a vector for the transferred bacterial genes. Subsequent recombinational processes can integrate part of the transferred genetic information into the donor chromosome, while in other cases the conjugative plasmid, together with a donor DNA segment, becomes established as an autonomous plasmid. Conjugative plasmids can span relatively wide host ranges. Again in line with the common evolutionary strategy of inefficiency in the generation of genetic variants, many conjugative plasmids are found in nature in a repressed state. Another observation relevant for our discussion is the capacity of conjugative plasmids to shift from the state of autonomy to passive propagation by integration into the host chromosome and vice versa upon excision. This is often mediated by IS elements carried in the conjugative plasmid and in the chromosome. This slow dynamics is a source of diversity in the genes becoming most readily transferred upon conjugation.

#### 4.3.1.3 Virus-Mediated Transduction

Many bacterial viruses (bacteriophages) are known to occasionally transfer segments of host DNA from a donor cell to a recipient becoming infected by the virus

particle (ARBER 1994). Uptake of bacterial DNA into a virus can follow different strategies. In general transduction, a viral particle can become packaged with a segment of the host chromosome instead of with the viral genome. Alternatively, in specialized transduction, hybrid DNA molecules composed of DNA of viral and of bacterial host origin are packaged into viral particles. Again, various types of recombinational integration into the recipient genome or the establishment of the hybrid vector as a plasmid must ensure inheritance in order to render the transferred genes functionally relevant to the new host.

#### **4.3.2 Barriers Against DNA Acquisition**

DNA acquisition encounters several natural barriers. Let us focus our attention here only on limitations affecting the DNA uptake by recipient cells and the inheritance and functional expression of the acquired genes. In all three DNA acquisition processes (transformation, conjugation and transduction) the cell surface of the recipient plays a key role in the penetration of the foreign DNA. Once donor DNA has entered the recipient cell, restriction/modification systems screen the DNA for its appropriate modification (ARBER 1965; ARBER and LINN 1969). Modification is usually a sequence-specific methylation of either adenine or cytosine nucleotides. If the modification is not according to the specificity of the recipient cell, the incoming DNA is recognized as foreign and subsequently endonucleolytically cut into fragments. Intracellular DNA fragments are rapidly degraded into nucleotides by exonucleases. However, open-ended DNA fragments are recombinogenic, and once in a while a fragment of donor DNA can become incorporated into the recipient genome. Thus, bacterial restriction enzymes drastically reduce the frequency of DNA acquisition, but the fragmentation fosters the occasional fixation of usually small segments of foreign DNA. DNA acquisition in small steps is precisely what can ensure good tolerance upon natural selection exerted on the resulting hybrid genome. By the addition to the genome of only one or a very few novel functions per step, the functional harmony of the recipient has a fair chance of not being disturbed. Under these conditions, the acquisition of a novel functional capacity can prove favourable and represent a selective advantage.

#### **4.4 Qualitative Differences Between the Three Strategies for the Generation of Genetic Variants**

The strategies of local sequence change, of DNA rearrangement and of DNA acquisition have quite different qualities. Local sequence changes are a source of stepwise improvement of available biological functions, which are steadily submitted to the forces of natural selection. If one allows for a number of silent and neutral steps of mutation to occur, the strategy might, in exceptional cases, eventually bring about a novel biological function. The measurement of evolutionary distances by the molecular clock is based on the strategy of local changes. It compares degrees of homology between sequences of usually one gene. Distance

measurements made by the molecular clock may be misled in cases of genes that have become changed by conversion in a process of DNA acquisition. Therefore, the measurement should be applied to several independent genes rather than to a single gene.

The strategy of DNA rearrangement can also improve available capacities, but it often brings about novel combinations such as operon and gene fusions. In an operon fusion a given reading frame is brought under the control of a different expression-control signal. Gene fusion can bring together two different functional domains or motifs, and the resulting sequences can occasionally result in a novel function, which may subsequently undergo further improvement by the strategy of local sequence change.

As already explained, DNA acquisition is a very powerful evolutionary strategy. In a single step a recipient cell may acquire a functional gene from a donor cell. This is a sharing in successful developments made by others.

One may postulate that it is good for bacteria to be equipped with one or a few systems for each of the three strategies for the generation of genetic variants. This is indeed what genetically well studied bacterial strains reveal. Being equipped with a system here means to possess in the genome or on accessory plasmids or viruses the appropriate genes to carry out the described reactions. Independent systems belonging to the same strategy for the generation of genetic variations may be able to substitute for each other, at least to some extent. But this is not the case for systems belonging to different strategies because of the different qualities in their contribution to the production of genetic variants.

## 5 Updated View on Molecular Evolution

We possess increasing knowledge on each of the specific mechanisms contributing to the generation of genetic variations. In addition, specific studies can reveal the rates at which each of these mechanisms – or at least the general strategies of local sequence change, of DNA rearrangement and of DNA acquisition – contribute to the overall mutagenesis. Molecular evolution thereby obtains a solid quantitative as well as mechanistic basis. In its establishment it is important to appropriately weigh the qualitatively different contributions made by each of the three strategies.

### 5.1 Evolutionary Tree

As to the tree of evolution, which should be imagined to have  $n$  dimensions rather than two as in the classical drawings, it has to carry randomly placed horizontal connectors between pairs of two branches (ARBER 1991). This can symbolize the occasional flux of one or a few genes by horizontal transfer, representing a contribution to the DNA acquisition strategy. In contrast, entire genomes become

vertically transmitted from generation to generation along the branches of the tree. Any of the genomic DNA sequences may occasionally become a target for local sequence change or undergo DNA reshuffling. Hence, in the vertical flux of genes entire genomes are subjected to occasional mutagenesis.

## 5.2 Coexistence in Bacterial Genomes of Genes Serving for the Benefit of Individuals and of Genes Serving Evolutionary Development

As was outlined in Sect. 4, a considerable number of genetically encoded systems contribute in diverse ways to generate genetic variations and to keep the frequency of such variations low enough to ensure a necessary genetic stability, but high enough to provide a tolerable degree of genetic diversity in populations so that evolution can steadily proceed. We call these genes, the products of which exert obvious evolutionary tasks, evolution genes. Clearly, they must be located, together with the housekeeping and accessory genes, on the bacterial chromosome or on accessory elements such as plasmids and viral DNA.

Evolutionary functions can be roughly classified into two groups: One of these concerns enzyme systems acting as generators of occasional genetic variations. Members of this group are involved mainly in the strategies of DNA rearrangement and of DNA acquisition. The other group of evolutionary functions has the task of keeping the frequencies of genetic variation low. Typical members of this second group are DNA repair systems and restriction/modification systems which mainly modulate the frequencies of genetic variation by the strategies of local sequence change and DNA acquisition.

A clear distinction between evolution genes on the one hand and housekeeping and accessory genes on the other is not always easy to make. One can note that housekeeping functions are in general required in all cells of a bacterial population. This may, at least to some degree, also apply to modulators of the frequency of genetic variation, while generators of genetic variation should be active only in rare cells of a bacterial population. A clear classification is also hampered by the fact that some gene products work both to the benefit of life requirements of all individuals and to the benefit of evolutionary developments, as exemplified by DNA topoisomerases and DNA ligases. But the products of a number of evolution genes are in no way required for normal cellular life, as exemplified by transposases.

## 5.3 Evolution Genes Are Proposed to be Submitted to Second-Order Selection

Housekeeping genes, the products of which are required in all cells, are submitted to direct selection for their evolutionary adaptation. Selection is thereby exerted on the functions of the gene products in the context of the activities of all expressed genes. This is obviously not possible for evolution genes that generate genetic variations. Rather, such genes are postulated to undergo evolutionary improve-

ment by second-order selection, i.e. based on the consequences of their occasional mutagenic action on directly selected genes in the context of the bacterial population (WEBER 1996). In other words, evolution genes are likely to be fine-tuned for their activity by their mutagenic action exerted on housekeeping and other essential genes. Bacteria which were already well equipped with properly functional variation generators had, in the past, the best chances of adapting the functions essential for each cell, so that it is precisely the descendants of these bacteria carrying properly functional variation generators that we encounter today. A similar argument can be made for evolution genes keeping the frequencies of mutagenesis low, but at levels favourable for a steady evolution. It is thought that fine-tuning for the proper functioning of evolution genes is brought about by the steady pressure made by second-order selection. One can imagine that for those genes which encode for evolution functions as well as for a function serving each individual to fulfil essential needs, both types of selection (i.e. direct selection and second-order selection) are simultaneously exerted so that both functions can become optimised.

#### 5.4 Conclusions on the Elements of the Theory of Molecular Evolution

According to the theory of evolution, biological evolution depends on the availability of genetic variations and on the natural selection exerted on mixed populations of variants and their ancestors. Evolution also depends on reproductive and geographic isolations which can modulate the evolutionary process, while the availability of genetic variations drives the process, and natural selection determines its directions. Evolution genes have their impact mainly on the generation of genetic variants, i.e. on mutagenesis. Evolution genes can also affect reproductive isolation. A good example are the restriction/modification systems. Interestingly, reproductive isolation as exerted by restriction endonucleases is not strict, but allows for the acquisition of foreign DNA in small steps.

The products of evolution genes are not alone in mediating biological evolution. They do it in cooperation with a number of nongenetic elements. Prominent contributions to mutagenesis are brought about by the chemical instability and the structural flexibility of biologically active molecules such as nucleotides and the enzymes interacting with DNA. Other contributions to genetic variation depend on random encounter, as, for example, of donor and recipient bacteria in conjugation or of a vector virus with a recipient bacterium in transduction. Other obvious nongenetic elements involved in the generation of genetic variants are environmental mutagens.

It is a close and harmonious cooperation of all of these nongenetic factors with the products of functionally fine-tuned evolution genes acting either as generators of genetic variations or as modulators of the frequencies of genetic variations, or both, which must have allowed the bacterial world to reach the astonishing diversity encountered today. Genetic evolution functions must have been developed in the course of time and they must still be submitted to improvement. They provide the bacteria with the capacity to steadily adapt to changing living condi-

tions. Among these conditions is the relatively recent appearance of higher plants and animals and eventually human beings as potential hosts for micro-organisms. In many cases this provided benefit to the novel host, but in other cases the association was detrimental to it. Pathogenicity of bacteria for their host is likely to result from adaptive evolution. A solid knowledge of the process of molecular evolution of bacteria represents a good basis for understanding pathogenic interactions and the origin of emerging infectious diseases.

## References

- Arber W (1965) Host-controlled modification of bacteriophage. *Annu Rev Microbiol* 19:365–378
- Arber W (1991) Elements in microbial evolution. *J Mol Evol* 33:4–12
- Arber W (1994) Bacteriophage transduction. *Encyclopedia of Virology*, Academic Press, New York, pp 107–113
- Arber W (2000) Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev* 24:1–7
- Arber W, Linn S (1969) DNA modification and restriction. *Annu Rev Biochem* 38:467–500
- Arber W, Naas T, Blot M (1994) Generation of genetic diversity by DNA rearrangements in resting bacteria. *FEMS Microbiol Ecol* 15:5–14
- Arini A, Keller MP, Arber W (1997) An antisense RNA in IS30 regulates the translational expression of the transposase. *Biol Chem* 378:1421–1431
- Avery OT, MacLeod CM, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med* 79:137–158
- Brock TD, Madigan MT, Martinko JM, Parker J (1994) *Biology of microorganisms*, 7th edn. Prentice-Hall International, pp 692–717
- Drake JW (1991) Spontaneous mutation. *Annu Rev Genet* 25:125–146
- Glasgow AC, Hughes KT, Simon MI (1989) Bacterial DNA inversion systems. In: Berg DE, Howe MM (eds) *Mobile DNA*. Amer Soc Microbiol, Washington DC, pp 637–659
- Hotchkiss RD, Gabor M (1970) Bacterial transformation, with special reference to recombination process. *Annu Rev Genet* 4:193–224
- Iida S, Hiestand-Nauer R (1987) Role of the central dinucleotide at the crossover sites for the selection of quasi sites in DNA inversion mediated by the site-specific Cin recombinase of phage Pl. *Mol Gen Genet* 208:464–468
- Iida S, Meyer J, Arber W (1983) Prokaryotic IS elements. In: Shapiro JA (ed) *Mobile genetic elements*. Academic Press, New York, pp 159–221
- Jacob F, Wollman EL (1961) Sexuality and the genetics of bacteria. Academic Press, New York
- Jordan E, Saedler H, Starlinger P (1968) 0° and strong polar mutations in the *gal* operon are insertions. *Mol Gen Genet* 102:353–363
- Lederberg J (1947) Gene recombination and linked segregations in *Escherichia coli*. *Genetics* 32:505–525
- Little JW, Mount DW (1982) The SOS regulatory system of *Escherichia coli*. *Cell* 29:11–22
- Mahillon J, Chandler M (1998) Insertion sequences. *Microbiol Mol Biol Rev* 62:725–774
- McClintock B (1950) The origin and behavior of unstable loci in maize. *Proc Natl Acad Sci USA* 36:344–355
- Naas T, Blot M, Fitch WM, Arber W (1994) Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. *Genetics* 136:721–730
- Naas T, Blot M, Fitch WM, Arber W (1995) Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12. *Mol Biol Evol* 12:198–207
- Olasz F, Stalder R, Arber W (1993) Formation of the tandem repeat (IS30)<sub>2</sub> and its role in IS30-mediated transpositional DNA rearrangements. *Mol Gen Genet* 239:177–187
- Porter RD (1988) Modes of gene transfer in bacteria. In: Kuperlapati R, Smith GR (eds) *Genetic recombination*. Amer Soc Microbiol, Washington DC, pp 1–41

- Sengstag C, Arber W (1983) IS2 insertion is a major cause of spontaneous mutagenesis of the bacteriophage P1: non-random distribution of target sites. *EMBO J* 2:67–71
- Shapiro JA (1969) Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. *J Mol Biol* 40:93–105
- Weber M (1996) Evolutionary plasticity in prokaryotes: a panglossian view. *Biol Phil* 11:67–88
- Weisberg R, Landy A (1983) Site-specific recombination in phage lambda. In: Hendrix RE, et al. (eds) Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 211–250
- Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51:221–271
- Zinder ND, Lederberg J (1952) Genetic exchange in *Salmonella*. *J Bacteriol* 64:679–699

# **Extraintestinal *Escherichia coli* as a Model System for the Study of Pathogenicity Islands**

P. REDFORD and R.A. WELCH

1	Introduction . . . . .	16
2	Discovery of Pathogenicity Islands . . . . .	16
3	PAI Characteristics . . . . .	17
4	PAIs in Uropathogenic <i>E. coli</i> . . . . .	18
4.1	UPEC Strain 536 . . . . .	18
4.2	UPEC Strain J96 . . . . .	18
4.3	UPEC Strain CFT073 . . . . .	19
5	PAIs in Other Extraintestinal <i>E. coli</i> . . . . .	19
5.1	Sepsis Strain AL862 . . . . .	19
5.2	Meningitis Strain EV36 . . . . .	20
5.3	Meningitis Strain C5 . . . . .	20
6	Advantages of Extraintestinal <i>E. coli</i> for the Study of PAIs . . . . .	20
6.1	Diversity . . . . .	20
6.2	Laboratory Workhorse . . . . .	20
6.3	In Vitro and In Vivo Models . . . . .	21
6.4	Human Studies . . . . .	21
6.5	Epidemiology . . . . .	22
6.6	Sequence Data . . . . .	22
7	Current Issues in the Study of Extraintestinal <i>E. coli</i> and PAIs . . . . .	23
7.1	Virulence . . . . .	23
7.2	Association of tRNAs and PAIs . . . . .	23
7.3	Phage and PAI Evolution . . . . .	24
7.4	Deletions and PAI Evolution . . . . .	25
7.5	PAIs and Hypervariable Loci . . . . .	26
7.6	Why tRNAs? . . . . .	26
7.7	The Case of pheV. . . . .	27
8	The Problem of Nomenclature . . . . .	27
9	Conclusion . . . . .	28
	References . . . . .	28

## 1 Introduction

The extraintestinal *Escherichia coli* comprise one of at least six pathogenic subtypes of *E. coli*. The group includes strains causing kidney and bladder infections, as well as sepsis and meningitis. The morbidity associated with these *E. coli* is vast. The uropathogenic strains (UPEC) are responsible for 80% of urinary tract infections in outpatients, resulting in more than 8 million visits each year to physicians in the United States alone (WENZEL 1991). *E. coli* remain a leading cause of abscess and septic shock. The meningitis *E. coli* strains account for 30% of these infections in neonates, resulting in death in one of five cases despite treatment (DE LOUVOIS 1994).

Pathogenicity islands (PAIs) were first described in extraintestinal *E. coli*. Since the first two islands were described as chromosomal elements in UPEC strain 536, two more have been identified in that strain, and at least two others are present in UPEC strain J96 and two in UPEC CFT073. Two islands have recently been described in the blood isolate AL862, and a single island has been found in a K1-encapsulated *E. coli* associated with meningitis (CIESLEWICZ and VIMR 1997; LALIOUI and BOUGUÉNEC 2001). Some of these PAIs superficially resemble each other in the genes that are present. For example, genes for hemolysin and P pili are found in many of the islands in the UPEC strains. Otherwise, the islands are quite diverse, with marked differences in size, chromosomal location, genetic content, and recombination potential.

Numerous other disease-causing bacteria have PAI-like features, but at present our understanding of the variety and number of these genetic elements in extra-intestinal *E. coli* is rivaled by few other bacteria. Extraintestinal *E. coli* therefore present an excellent model for determining the role of these gene clusters in disease. This review will describe what the study of extraintestinal strains has shown so far, and what it may yet reveal about the phenomenon of pathogenicity islands.

## 2 Discovery of Pathogenicity Islands

The first hint that the strains causing urinary tract infections differ from other *E. coli* came early in the twentieth century, when it was observed that such strains are often hemolytic, an uncommon phenotype in fecal *E. coli* strains (LYON 1917). Subsequent epidemiological studies have shown that, compared with fecal isolates, pyelonephritic strains are approximately three times more likely to display hemolysis and six times more likely to show mannose-resistant hemagglutination (MRHA) (DONNENBERG and WELCH 1996). Pyelonephritis-associated pili (P pili) are one cause of the MRHA phenotype. The linkage of genes encoding P pili and hemolysin was recognized by Low et al., who found clones containing both the P pilus and hemolysin genes in cosmid libraries of genomic DNA from UPEC strains. Welch

et al. and Low et al. looked outside the operons encoding these two factors and saw sequences in virulent strains not present in fecal *E. coli* isolates (Low et al. 1984; WELCH and FALKLOW 1984). Hacker et al. demonstrated that the P pili and hemolysin determinants could be lost after spontaneous deletions of up to 200kb in length from the chromosomes of different uropathogenic isolates (HACKER et al. 1983). It was also shown that the genes for P pili are linked to those for hemolysin in a chromosomal region of UPEC strain 536 not present in the laboratory *E. coli* strain K-12. This large, apparently distinct region was called a pathogenicity island, and two were described in strain 536, one with a copy of the hemolysin operon (*hly*) and the other with *hly* as well as genes for the P pili (*prf*) (HACKER et al. 1990). With *hly* as a guidepost, two more PAIs were subsequently described in UPEC strain J96 and one in UPEC strain CFT073 (SWENSON et al. 1996; KAO et al. 1997). Meanwhile, PAIs have been found in other subtypes of disease-causing *E. coli*, as well as in *Salmonella* species and bacterial genera outside the Enterobacteriaceae. Detailed discussions of these organisms and their islands are presented elsewhere in this volume.

### 3 PAI Characteristics

The many examples of PAIs found since their discovery in strain 536 has lead Hacker and colleagues to make some generalizations about them (HACKER et al. 1997). The UPEC islands are usually quite large, ranging from 10 to 200kb. As the name suggests, they contain genes associated with virulence. They are located in the chromosome at specific loci. Genes for transfer RNAs are often at these loci. A subset of tRNA genes (*pheV*, *pheU*, *leuX*, *thrW*, *serT*, *serW*, *serX*, *valV*, *asnT*) is found repeatedly associated with islands in *E. coli* as well as in other gram-negative bacteria. Further, the loci often feature *att* sites for different temperate phages (e.g., P22, P2 and P4). The UPEC islands usually have a guanosine plus cytosine (G + C) content different from that of the rest of the chromosome. Many islands feature direct repeats at each end, suggesting that a recombination event is associated with their creation. Indeed, the first two islands described in strain 536 are unstable and can delete spontaneously from the chromosome at a frequency of  $\sim 10^{-4}$ . These observations, taken together, suggest that horizontal gene transfer played a big part in the history of these PAIs and that significant evolutionary change at these sites is still possible. Ten years after the first use of the term, pathogenicity islands appear to be widespread among all clonal types of *E. coli*. Extraintestinal *E. coli* continue to be a rich hunting ground for new or variant examples of these genomic phenomena. An overview follows of specific islands in extraintestinal *E. coli*, including UPEC strains, a blood isolate, and a K1 meningitis strain.

## 4 PAIs in Uropathogenic *E. coli*

### 4.1 UPEC Strain 536

*E. coli* strain 536 was cultured from a patient with pyelonephritis. Initially, two pathogenicity islands were found in its chromosome, as described above. Recently, two additional islands have been described (HACKER et al. 1999). These discoveries have led to a proposal to name the islands with Roman numerals and apply this numbering system to other strains with multiple PAIs.

PAI I<sub>536</sub> is at minute 82.6 on the K-12 map and is 70kb in length. At this position in K-12 is the tRNA for selenocysteine, *selC*. The island is flanked by direct repeats, 16bp in length, and sequence analysis of the right junction showed that the direct repeat here runs into the 3' end of the *selC* sequence. Spontaneous deletions of this island have been observed, and these deletions disrupt the *selC* sequence. Near the right junction on PAI I<sub>536</sub> is a sequence homologous to the integrase of the retronophage ΦR73. PAI I<sub>536</sub> contains a copy of the *hly* operon, encoding hemolysin (HACKER et al. 1997).

PAI II<sub>536</sub>, at 190kb, is one of the largest known islands. This second island is at 96.9', immediately upstream of the leucine tRNA, *leuX*. Like PAI I<sub>536</sub>, PAI II<sub>536</sub> is flanked by direct repeats, in this case 18bp long. Spontaneous deletion mutants of PAI II<sub>536</sub> can readily be isolated. Deletions in this case disrupt the *leuX* tRNA at its 3' terminus. PAI II<sub>536</sub> has a copy of the *hly* operon and has as well *prf*, the operon for the P pilus. It also contains a P4 integrase homolog, though it may not be functional due to the presence of several premature stop codons (HACKER et al. 1999).

PAIs III<sub>536</sub> and IV<sub>536</sub> are inserted at 5.6' and 44', respectively, and are approximately 25 and 45kb. These inserts are at loci for the tRNAs *thrW* and *asnT*. PAI III<sub>536</sub> features genes for the S fimbriae and a sequence with homology to that for bacteriophage DLP12 integrase. PAI IV<sub>536</sub> shares features with the high-pathogenicity island (HPI) of *Yersinia*. The main difference is a deletion of sequence in 536 at the right junction, including the direct repeat there. Spontaneous deletions of PAI III<sub>536</sub> or PAI IV<sub>536</sub> have not been observed, although PAI IV<sub>536</sub> carries a P4 integrase which, in contrast to the one in PAI II<sub>536</sub>, is not defective. As in the HPI, PAI IV<sub>536</sub> is composed almost entirely of genes for the yersiniabactin siderophore and its receptor (HACKER et al. 1999).

### 4.2 UPEC Strain J96

UPEC strain J96 is also from a patient with acute pyelonephritis. Two large islands are known in this isolate, PAI I<sub>J96</sub> and II<sub>J96</sub>. The first of these is the bigger one, at >170kb, and contains a *hly* operon and a *pap* operon for the P pilus. PAI II<sub>J96</sub> is 110kb and also contains a hemolysin operon, although the expression of this *hly* is much lower than that of PAI I<sub>J96</sub>. A P pilus operon is in PAI II<sub>J96</sub> as well, though it is *prs*, a different class. PAI II<sub>J96</sub> also carries a sequence coding for the toxin

cytotoxic necrotizing factor type 1 (CNF1). PAI I<sub>J96</sub> is located adjacent to the tRNA *pheV* (67'), and PAI II<sub>J96</sub> is at *pheU* [94': this was designated *pheR* before publication of the genomic K-12 map (BLATTNER et al. 1997)]. No deletion mutants of either J96 PAI have been observed in our laboratory, although PAI II<sub>J96</sub> deletions have been observed with the isolate of J96 in the Hacker laboratory. The deletion here possibly involves the long (135bp) indirect repeats at each end of this island (SWENSON et al. 1996).

### 4.3 UPEC Strain CFT073

UPEC strain CFT073 was isolated from the blood and urine of a woman with pyelonephritis. It was selected from a collection of 61 such strains because it was the most cytotoxic to renal tubular epithelial cells in culture (MOBLEY et al. 1990). Initially, one island was described in this strain. PAI I<sub>CFT073</sub> is 58kb in size, with 44 open reading frames and segments of K-12 sequence interspersed within the island. At each junction there is a short (9-bp) direct repeat. No spontaneous deletions of the island have been observed. PAI I<sub>CFT073</sub> contains an *hly* operon, the only one known in this strain. It also has *pap*, one of two copies in the chromosome (KAO et al. 1997). When it was first described, mapping its location relative to the K-12 genome gave a confusing picture, since the right junction appeared to be at 63', while the left was at 26.5'. Preliminary data from the genomic DNA sequence analysis of this strain, however, place *hly* at 67', suggesting that this may represent another island at the *pheV* locus (R.A. Welch, unpublished). The other copy of a *pap*-like operon has recently been described as present on a second island. The preliminary report indicates PAI II<sub>CFT073</sub> is at least 70kb and contains 41 ORFs, many related to iron acquisition (PHILLIPS et al. 2000).

## 5 PAIs in Other Extraintestinal *E. coli*

### 5.1 Sepsis Strain AL862

Recently, *E. coli* strain AL862, isolated from the blood of a cancer patient, was found to contain two islands at the *pheV* and *pheU* loci, just as in strain J96 (LALIOUI and BOUGUÉNEC 2001). However, these PAIs bear little other resemblance to those in J96. They were found during the mapping of *afa-8*, which codes for AfaE-VIII, a recently described member of the afimrial adhesin family found in extraintestinal *E. coli* at a prevalence three times greater than that of fecal isolates. PAI I<sub>AL862</sub> is 61kb, and *afa-8* is the only known virulence factor encoded on it. The island features a P4 integrase and has 14-bp direct repeats at each end. PAI II<sub>AL862</sub> is also 61kb and contains *afa-8*. In fact, it appears to be a duplicate of PAI I, though this will not be certain until both islands are entirely sequenced. The duplication of an entire island

had not been previously observed in *E. coli*, although the recently published genome sequence of *E. coli* O157:H7 strain EDL933 reveals another instance of island duplication, where a large, 106-gene cluster encoding tellurite resistance and urease production resides at *serW* and also at *serX* (PERNA et al. 2001).

## 5.2 Meningitis Strain EV36

The tRNA *pheV* is also associated with what appears to be a PAI in *E. coli* EV36 (CIESLEWICZ and VIMR 1997). This is a strain causing neonatal meningitis, and most of the genes on this 20-kb element, which is inserted adjacent to *pheV*, appear to direct the synthesis and export of the polysialic acid capsule (K1), a key virulence factor in meningitis isolates. The typical features of PAIs such as terminal repeats, integrases, and unusual G + C content have yet to be reported for this putative island.

## 5.3 Meningitis Strain C5

Another meningitis strain, *E. coli* C5, has been studied by representational difference analysis, comparing its genome with that of two nonpathogenic strains (ECOR 4 and 15). Six unique regions in C5 were found. Their full nature awaits further characterization, but one of the regions resembles PAI II<sub>J96</sub> in that it contains *hly*, *prs* and *cnf1* (BONACORSI et al. 2000).

# 6 Advantages of Extraintestinal *E. coli* for the Study of PAIs

## 6.1 Diversity

This overview shows that PAIs in extraintestinal *E. coli* are diverse in content and genome location, even among strains causing the same infection, e.g., pyelonephritis. This is the foremost advantage of using these strains as a model. From this variety patterns can be perceived, and as more examples of islands accumulate, greater insight into their formation, architecture, and role in disease can be expected.

## 6.2 Laboratory Workhorse

Extraintestinal *E. coli* offer other advantages, however. The fact that they are *E. coli* makes them genetically tractable, since this genus has a long history in the molecular genetics laboratory. Wild-type *E. coli* are often not as easy to manipulate as laboratory strains, but many of the same techniques apply. In addition to

genetics, *E. coli* physiology has a large literature. Knowledge of *E. coli* metabolism may aid in determining how islands might add to bacterial fitness, while gene expression technologies may reveal how the genes contained within these highly variable loci interact with genes elsewhere on the chromosome.

### 6.3 In Vitro and In Vivo Models

Another major advantage of studying the extraintestinal *E. coli* is the availability of an array of relevant tissue culture and animal models. This is in addition to in vitro phenotypes which, for example, led to the discovery of hemolysin and adhesins. Cytotoxicity and invasion assays have been developed using cell cultures ranging from primary bladder epithelium to human brain microvascular endothelial cells (BADGER et al. 2000a; MULVEY et al. 1998). Most in vivo studies use rodents. For example, with UPEC, the ascending urinary tract infection in mice has become the animal model most frequently used, though some laboratories have relied upon rat models (HAGBERG et al. 1983; MARRE et al. 1986). Induction of peritonitis in mice and direct injection into the mouse bloodstream are other infection systems where valuable information has been gained concerning pathogenesis (ALKAN et al. 1986; RITTER et al. 1995).

### 6.4 Human Studies

Extraintestinal *E. coli*, or at least the uropathogenic subtypes, possess the additional advantage of being amenable to study in human beings. Experiments done so far in volunteers have focused on the role of adhesins and establishment of urinary tract infection (ANDERSSON et al. 1991; WULLT et al. 2000). These human subjects underwent intravesicular inoculation with strain 83972, an *E. coli* cultured from a child with long-standing, asymptomatic bacteriuria. They received the wild-type or a transformant expressing P pili from a recombinant plasmid. In this setting, P pili appeared to decrease both the time and the number of bacteria needed to establish infection. The number of subjects was not large and the studies have the drawback one expects from a small human trial where the subjects have highly diverse genetic backgrounds and immune histories. Nonetheless, these experiments show that significant questions about UPEC pathogenesis can be examined in the definitive host population under carefully chosen conditions.

Another example of UPEC pathogenesis studied with human subjects is the examination of the type 1 pilus phase switch in genomic DNA isolated directly from bacteria in urine (LIM et al. 1998). The study revealed that UPEC in the urine of infected individuals are predominantly in the type 1 off-phase, although in vitro these same isolates are capable of existing in either the on- or the off-phase. Such studies raise the prospect of broader analysis of UPEC virulence gene expression in human beings. For example, bacterial RNA isolated from patient samples could give an *in toto* view of the genes induced during infection.

## 6.5 Epidemiology

Historically, however, the study of UTI in human beings has used them not as experimental subjects, but as epidemiological data points. Decades of such clinical prevalence studies have created a profile of UPEC, with a wealth of information on serotype, capsule, serum resistance, and the prevalence of factors such as the different adhesins, toxins, and siderophores (DONNENBERG and WELCH 1996). The abundance of human epidemiology stems from the fact that UTIs are common and the bacteria that cause them are easily obtained. In comparison to the past, epidemiological studies now and in the future may be more informative, because as the complete UPEC and other *E. coli* pathogen genome sequences become available, patterns in the prevalence of genetic traits unique to UPEC can be determined.

## 6.6 Sequence Data

We are approaching a critical and exciting juncture in *E. coli* genomics research. The *E. coli* K-12 genome sequence was the crucial first step (BLATTNER et al. 1997). Now, with the complete sequence of an enterohemorrhagic *E. coli* (EHEC) strain (PERNA et al. 2001), and the sequence of UPEC CFT073 nearing completion (<http://www.genetics.wisc.edu>), there is anticipation that new insights into *E. coli* pathogenesis are on the way. This new sequence information, on top of the available database, will be fodder for molecular evolutionists who are attempting to discern how different pathogenic *E. coli* arose. Phylogenetic trees of *E. coli* will continue to be refined, with special emphasis on the evolution of pathogenic strains via horizontal genetic transfer mechanisms. These studies will be invaluable in helping us shape our view on the significance of the islands to pathogenesis (OCHMAN and BERGTHORSSON 1998).

As noted above, new techniques and methodologies are providing unprecedented opportunities for analysis of PAIs in extraintestinal *E. coli*. Commercially prepared microarrays containing all of the genes present in the *E. coli* K-12 genome are available, with arrays covering the pathogenic *E. coli* likely to follow. These will enable discovery of the regulatory interconnections for bacteria growing *in vivo* as well as *in vitro*. To complement these experimental methods, whole-genome-based genetic selection and screening methods for virulence gene discovery, such as signature-tagged mutagenesis, representational difference analysis, and differential fluorescence induction are successfully being carried out with the different pathogenic types of *E. coli* (BADGER et al. 2000a,b; BONACORSI et al. 2000).

## 7 Current Issues in the Study of Extraintestinal *E. coli* and PAIs

### 7.1 Virulence

The central issue at present in the study of extraintestinal *E. coli* and pathogenicity islands is defining the role of PAIs in virulence. This has been hindered in part by their size and complexity. UPEC strain 536, with its spontaneous island deletion mutants (PAI I<sub>536</sub><sup>-</sup>, PAI II<sub>536</sub><sup>-</sup>, or both islands deleted) has offered the best opportunity for studies of the role of whole islands in virulence. Some of the results of such studies have been unexpected. The surprise stems from the major part that at least two of the tRNAs associated with PAIs appear to play. Strain 536-21, a deletion mutant lacking both PAI I<sub>536</sub> and PAI II<sub>536</sub>, is greatly diminished in its ability to infect the kidneys of rats. Much of the ability to infect the kidney can be restored, however, by plasmid complementation of 536-21 with the tRNA genes that are disrupted during PAI deletions. In vitro studies have suggested that the reason for this might be that type 1 fimbriae expression, enterochelin iron uptake, and serum resistance in 536-21 can be restored by *leuX* complementation, and anaerobic fermentation restored by *selC*. Notably, these complemented strains do not restore hemolysin or P pili expression, or presumably any of the other >200 genes that are lost with the island deletions. Yet, in the rat pyelonephritis model, the *leuX*<sup>+</sup> 536-21 strain yielded an infection intensity (cfu/g kidney) nearly equal to that of 536 (RITTER et al. 1995). Thus, surprisingly, the large sets of genes on PAIs, including *pap* and *hly* operons which are closely linked to pyelonephritis in epidemiological studies, are not virulence factors here. It may be that this style of animal model, where different strains are used separately to infect different groups of animals, lacks sensitivity. An alternative model is one done competitively; i.e., the strains are mixed together, and the combined preparation is inoculated into the bladder. Our laboratory has been carrying out competitive studies, and preliminary experiments support the idea that loss of genes within an island such as PAI I<sub>536</sub> result in a significant loss of virulence in a mouse model of ascending UTI.

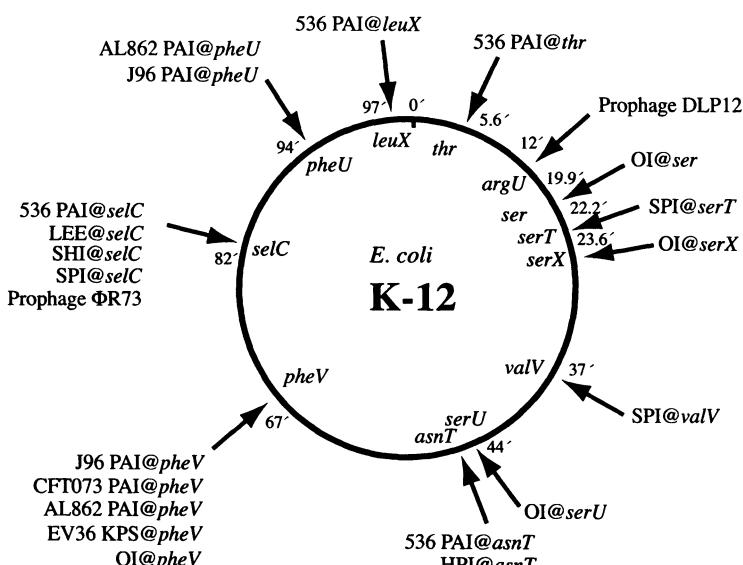
### 7.2 Association of tRNAs and PAIs

Specific tRNAs occupy a pivotal place not only in virulence gene expression, but clearly also in the evolution of pathogenicity islands. Initial attention was focused on *selC* because it was associated with PAI I<sub>536</sub> and then found to be at the site for the attaching and effacing locus (LEE) in EPEC and EHEC (McDANIEL et al. 1995). It was also found to be associated with the *Salmonella typhimurium* island, SPI-3 and with the SHI-2 in *Shigella flexneri* (Moss et al. 1999). As more PAIs are described, other tRNA genes show similar patterns of associations with multiple PAIs. In strain J96, *pheV* is associated with PAI I<sub>J96</sub>, and it also appears to be adjacent to a potential island in *E. coli* strain AL862 and another possible

island in meningitis strain EV36 (CIESLEWICZ and VIMR 1997; LALIOUI and BOUGUÉNEC 2001; SWENSON et al. 1996). *pheU* is associated with PAI II<sub>J96</sub> and with the other potential island in AL862, as well as with LEE in some EPEC and EHEC strains (SPERANDIO et al. 1998). In UPEC strain 536, *asnT* is associated with PAI IV<sub>536</sub>, which is the island that closely matches the HPI in *Yersinia* species. In *Yersinia*, the HPI is present at the *asnT* locus (or associated with two other asparagine tRNAs, *asnU* or *asnW*) (RAKIN et al. 1999). The polygamous association of the same tRNA with different islands is common enough that it has been used as a way to find new islands. SPI-3 in *S. typhimurium*, for example, was found by examining sequence around the *selC* locus (BLANC-POTARD and GROISMAN 1997).

### 7.3 Phage and PAI Evolution

PAIs are not the only genome-altering phenomenon linked to specific tRNAs. *leuX*, *thrW* and *selC*, which have been observed to be PAI loci (Fig. 1), all feature as well *att* sites, where phages attach and integrate in K-12 (BERLYN 1998). For example, the bacteriophage ΦR73 uses an *att* in the 3' sequence of *selC* as an integration site. Furthermore, a cryptic ΦR73 integrase gene has been found on PAI I<sub>536</sub> (HACKER et al. 1997). As noted above, many other islands contain prophage integrases.



**Fig. 1.** Ectochromosomal DNA sequences (“ectons”). Map shows sequences in pathogenic strains not present in MG1655 (K-12) chromosome. Preliminary data from sequencing of UPEC CFT073 show ectons at all of these tRNA loci. *PAI*, pathogenicity island; *LEE*, locus of enterocyte effacement; *KPS*, capsule polysaccharide synthesis cluster; *HPI*, high-pathogenicity island; *SHI*, *Shigella* island; *SPI*, *Salmonella* pathogenicity island; *OI*, EHEC O island

These findings have often been cited as evidence that temperate phages play a major role in genome evolution. Insertion and deletion at attachment sites creates hot spots for changes in the genetic potential of bacteria. The recent description of an *S. typhimurium* temperate phage that carries the virulence gene *sopE* provides an excellent example of how incremental changes at one of the P2 temperate phage insertion sites can come about (MIROLD et al. 1999).

The integration of phages or islands at tRNA gene loci may not be the whole story of island evolution, however. The *pheV* locus, for example, is where multiple examples of PAIs reside, yet it is not known to represent any prophage *att* site. How the different islands at the *pheV* locus came independently to be there is unknown, but a mechanism different from classic prophage transduction with integration at an *att* site may be operating. Many PAIs appear to be genetically immobilized, since they lack integrases or have disrupted or partial transposase gene sequences. Such observations have led investigators to hypothesize that there are successive stages to PAI evolution: from genetically active elements early on to mutated states, where the gene blocks are locked or “homed” to a site (HACKER et al. 1999). This may not always be the case, however. Perhaps the best evidence to date that genomic islands behave as live genetic elements (and are not necessarily the remnants of piecemeal prophage recombination processes) is a recent report by SULLIVAN and RONSON describing a very large 500-kb symbiosis island that transfers and inserts into a possible *pheU* locus in *Mesorhizobium loti* (SULLIVAN and RONSON 1998).

## 7.4 Deletions and PAI Evolution

Deletions, as well as integration, appear to play a significant role in island evolution. Maurelli and colleagues found that when the genomes of *Shigella* species are compared with *E. coli* K-12, there are large deletions (with varying endpoints in different species) of nearly 90kb that encompass the gene for lysine decarboxylase, *cadA* (MAURELLI et al. 1998). The loss of the lysine decarboxylase activity appears to be critical for enterotoxicity, because when *S. flexneri* is complemented with the *cadA* gene, there is attenuation of virulence function. This is attributable to extracellular production of cadaverine, which inhibits the enterotoxin ShET. The *cadC* gene in *E. coli* strains is closely linked to the *pheU* site at 94', and several deletions described by Maurelli and co-workers remove genes in this region, including *pheU* and genes past it, such as *cutA*. In UPEC strains J96 and CFT073, the *cadA-pheU* region is intact, but what lies between the *pheR-cutA* region differs between the two strains. The *pheU-cutA* region encompasses what we have previously described in J96 as PAI II<sub>J96</sub>. This region contains a defective P4 integrase, hemolysin, and P-related pili operons. Hacker's laboratory has observed deletion of PAI II<sub>J96</sub>, and such a deletion would mimic the creation of the “black hole” described by Maurelli and colleagues in the evolution of *Shigella* species from an *E. coli* progenitor.

## 7.5 PAIs and Hypervariable Loci

Data from the EHEC and UPEC genome sequencing projects at the University of Wisconsin (PERNA et al. 2001) as well as other genome sequences available show that the island integration hotspots at *pheV*, *pheU*, *leuX*, *thrW*, *serT*, *valV*, and *asnT* are hypervariable loci. For example, in addition to the known PAI I<sub>CFT073</sub> at *pheV*, there are sequences present in CFT073 at all the above-mentioned loci that are not present in *E. coli* K-12. Whether the reference point for genome comparison is K-12 or any wild strain of *E. coli*, *Salmonella*, *Shigella*, or *Yersinia*, we predict that each of the loci among this subset of tRNA genes will show tremendous variability in the sequences that are present. Examination of the sequence differences suggests a great diversity of genetic events, from transposition, duplication, homologous intrachromosomal recombination and, importantly, loss of function by deletions and inactivation by insertional disruption.

## 7.6 Why tRNAs?

Why is a specific subset of tRNA genes involved in island evolution? An oft-mentioned hypothesis is that the tRNA genes are linked to PAIs because the conserved nature of tRNAs across different species creates reliable targets for temperate phage integration. Why, then, are only some tRNAs involved? Are there only a limited number of phages integrating at only a limited number of *att* sites? Are there selective advantages for gene acquisition as well as for deletions at these sites? Is it simply the case that cross-species sequence homology and redundancy in tRNA functions permits genetic tinkering without a fitness cost, as some have suggested? For example, the asparagine tRNAs *asnT*, *asnU*, and *asnW* all have the same anticodon, so insertion or deletion at *asnT* might render it nonfunctional, but its loss would not greatly affect translation (HACKER et al. 1999). The redundancy argument can also be made with the *pheV-pheU* and *serX-serW* gene pairs, since each pair possesses the same anticodon, and each member of the pair is some distance apart from its homolog on the chromosomal map. Thus, disruption of one of these should not affect translation. But with *selC*, *leuX*, and *thrW* there are no functional homologs for use as a fallback. Specific tRNAs may have functions affecting virulence (RITTER et al. 1995), as appears to be the case with *leuX* and *selC* (see above), so disruption at these loci may be driven by factors more complex in terms of bacterial fitness than simple tolerance of their loss. Hou offers another explanation, based on a molecular hypothesis (Hou 1999). He posits that these hypervariable sites offer enhanced opportunities for DNA invasion and recombination through the participation of the 3' end of the tRNA in DNA-RNA hybrids. These may provide better substrates for integrases than DNA-DNA hybrids.

## 7.7 The Case of *pheV*

As mentioned above, *pheV* lacks a known *att* site. It is adjacent, however, to a locus named *rdgB*. In 1987, Clyman and Cunningham noted that mutations at or very near this locus result in increased intrachromosomal homologous recombination, as observed by recombination between duplicated but different *lacZ* mutant genes (CLYMAN and CUNNINGHAM 1987). The mechanism is unclear: further characterization of *rdgB* has not been reported. However, it is tempting to hypothesize that recombination within the chromosome may account for the duplications of genes (and possibly entire islands) seen in UPEC and other strains of *E. coli* such as EHEC. An intrachromosomal phenomenon like this *rdgB*-associated recombination could also generate the diversity seen with some virulence genes, such as the different classes of P pili and the hemolysin, or the several examples of AIDA-1-like autotransporters observed by genomic sequencing.

## 8 The Problem of Nomenclature

As DNA sequence information on *E. coli* pathogenicity islands rapidly accumulates, the present ad hoc system of naming them and their homologs is becoming strained. The terminology has evolved from clusters to islands to hemolysin islands to pathogenicity islands, and now there are genomic islands, symbiosis islands, islets, and archipelagoes. And these are only the generic names: LEE and HPI, for example, are more specific terms for islands in EPEC and *Yersinia* spp., respectively, while SHI and SPI describe islands in *Shigella* and *Salmonella*. On top of this is the problem of multiple islands in one chromosome, which has led to numbering systems using Roman numerals for the UPEC strains, or Arabic numerals to designate *Salmonella* islands.

The terminology is becoming more bewildering because it is being overwhelmed by the diversity of these DNA segments in size, number, gene composition, chromosomal location, and evolution. Results from the genome sequences suggest a wealth of new examples even more varied than those described so far. In the analysis of the *E. coli* O157:H7 strain EDL933 genome there are 177 unique DNA segments not found in the *E. coli* K-12 genome and, conversely, 234 K-12-specific segments not found in the genome of EDL933 (PERNA et al. 2001). The former were labeled “O islands”, while the latter are “K islands”. Overall there is a total of greater than 1.34Mb of unique, O-island DNA. This report shows not only the staggering number of these DNA segments, but also that K-12 itself, with its unique sequences, is not an essential or minimalist archetype of an *E. coli*.

We believe that specific terms such as PAI, LEE, and SPI will remain descriptive and useful. However, the generic terms need more definition. For example, how big is an island? The word suggests an element of a certain size, for example 10–20kb. What criteria are needed to add “pathogenicity” or “symbiosis” to it? Do

segments without apparent structure, for example, no *IS* elements or junctional direct repeats, qualify? We suggest that a broader, more inclusive term is needed to describe the phenomenon of DNA sequences present in a strain but not present in the reference strain. One such term is “ectochromosomal DNA”, with the prefix “ecto-” giving the connotation of coming from the outside as well as being outside the norm (i.e., the reference). We also suggest that an alternative system be developed for naming multiple ectochromosomal sequences within the same strain. Instead of numbers, it might be more descriptive to include the locus at which the “ecton” is found. Thus PAI I<sub>J96</sub> could be denoted J96 PAI@*pheV*, while PAI IV<sub>536</sub> could be 536 PAI@*asnT*. Figure 1 shows such a locus-based nomenclature. It also highlights the central role specific loci play, with a diverse array of ectons occupying the same locus in different strains.

## 9 Conclusion

In summary, during the next several years we can expect an unprecedented increase in genomic knowledge. We will have multiple, fully sequenced genomes and will be able to compare them with each other. The differences will be key in determining why some Enterobacteriaceae are benign and others pathogenic. The precedent for such comparative genomics in the field of medical microbiology was established by the discovery and description of pathogenicity islands in UPEC. Extraintestinal *E. coli* will continue to prove invaluable as analysis moves beyond sequence differences and into functional significance. The discovery of pathogenicity islands coincided with what appears to have been the start of a golden age of research in microbial pathogenesis. At this point, continual genetic and genomic revelations make it clear that this era is nowhere near its end.

## References

- Alkan ML, Wong L, Silverblatt FJ (1986) Change in degree of type 1 pilation of *Escherichia coli* during experimental peritonitis in the mouse. *Infect Immun* 52:549–554
- Andersson P, Engberg I, Lidin-Jansson G, Lincoln K, Hull R, Hull S, Svartbom C (1991) Persistence of *Escherichia coli* bacteriuria is not determined by bacterial adherence. *Infect Immun* 59:2915–2921
- Badger JL, Wass CA, Kim KS (2000a) Identification of *Escherichia coli* K1 genes contributing to human brain microvascular endothelial cell invasion by differential fluorescence induction. *Mol Micro* 36:174–182
- Badger JL, Wass CA, Weissman SJ, Kim KS (2000b) Application of signature-tagged mutagenesis for identification of *Escherichia coli* K1 genes that contribute to invasion of human brain microvascular endothelial cells. *Infect Immun* 68:5056–5061
- Berlyn MK (1998) Linkage map of *Escherichia coli* K-12, Edition 10: the traditional map. *Micro Mol Biol Rev* 62:814–984
- Blanc-Potard AB, Groisman EA (1997) The *Salmonella* *selC* locus contains a pathogenicity island mediating intramacrophage survival. *EMBO J* 16:5376–5385

- Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collada-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–1474
- Bonacorsi SPP, Clermont O, Tinsley C, Gall IL, Beaudoin JC, Elion J, Nassif X, Bingen E (2000) Identification of regions of the *Escherichia coli* chromosome specific for neonatal meningitis-associated strains. *Infect Immun* 68:2096–2101
- Cieslewicz M, Vimr E (1997) Reduced polysialic acid capsule expression in *Escherichia coli* K1 mutants with chromosomal defects in *kpsF*. *Mol Microbiol* 26:237–249
- Clyman J, Cunningham RP (1987) *Escherichia coli* K-12 Mutants in which viability is dependent on *recA* function. *J Bacteriol* 169:4203–4210
- De Louvois J (1994) Acute bacterial meningitis in the newborn. *J Antimicrob Chemother* 34 [Suppl A]:61–73
- Donnenberg M, Welch RA (1996) Virulence determinants of uropathogenic *E. coli*. In: Mobley HLT, Warren JW (eds) Urinary tract infections: molecular pathogenesis and clinical management. ASM Press, Washington, DC
- Hacker J, Knapp S, Goebel W (1983) Spontaneous deletions and flanking regions of the chromosomally inherited hemolysin determinant of an *Escherichia coli* O6 strain. *J Bacteriol* 154:1145–1152
- Hacker J, Bender L, Ott M, Wingender J, Lund B, Marre R, Goebel W (1990) Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal *Escherichia coli* isolates. *Microbial Pathog* 8:213–225
- Hacker J, Blum-Oehler G, Mühlendorfer I, Tschäpe H (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* 23:1089–1097
- Hacker J, Blum-Oehler G, Janke B, Nagy G, Goebel W (1999) Pathogenicity islands of extraintestinal *Escherichia coli*. In: Kaper JB, Hacker J (eds) Pathogenicity islands and other mobile virulence elements. ASM Press, Washington, DC
- Hagberg L, Engberg I, Freter R, Lam J, Olling S, Svanborg-Eden C (1983) Ascending, unobstructed urinary tract infection in mice caused by pyelonephritogenic *Escherichia coli* of human origin. *Infect Immun* 40:273–283
- Hou YM (1999) Transfer RNAs and pathogenicity islands. *Trends Biochem Sci* 24:295–298
- Kao JS, Stucker DM, Warren JW, Mobley HL (1997) Pathogenicity island sequences of pyelonephritogenic *Escherichia coli* CFT073 are associated with virulent uropathogenic strains. *Infect Immun* 65:2812–2820
- Laloui L, Bouguénec CL (2001) *afa-8* gene cluster is carried by a pathogenicity island inserted into the tRNA-Phe of human and bovine pathogenic *Escherichia coli* isolates. *Infect Immun* 69:937–948
- Lim JK, Gunther NW, Zhao H, Johnson DE, Keay SK, Mobley HLT (1998) In vivo phase variation of *Escherichia coli* type 1 fimbrial genes in women with urinary tract infection. *Infect Immun* 66:3303–3310
- Low D, David V, Lark D, Schoolnik G, Falkow S (1984) Gene clusters governing the production of hemolysin and mannose-resistant hemagglutination are closely linked in *Escherichia coli* serotype O4 and O6 isolates from urinary tract infections. *Infect Immun* 43:353–358
- Lyon MW (1917) A case of cystitis caused by *Bacillus coli*-hemolyticus. *JAMA* 69:353–358
- Marre R, Hacker J, Henkel W, Goebel W (1986) Contribution of cloned virulence factors from uropathogenic *Escherichia coli* strains to nephropathogenicity in an experimental rat pyelonephritis model. *Infect Immun* 54:761–767
- Maurelli AT, Fernandez RE, Bloch CA, Rode CK, Fasano A (1998) “Black holes” and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc Natl Acad Sci USA* 95:3943–3948
- McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB (1995) A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci USA* 92:1664–1668
- Mirold S, Rabsch W, Rohde M, Stender S, Tschäpe H, Russmann H, Igwe E, Hardt WD (1999) Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. *Proc Natl Acad Sci USA* 96:9845–9850
- Mobley HLT, Green DM, Trifillis AL, Johnson DE, Chippendale GR, Lockatell CV, Jones BD, Warren JW (1990) Pyelonephritogenic *Escherichia coli* and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. *Infect Immun* 58:1281–1289
- Moss JE, Cardozo TJ, Zychlinsky A, Groisman EA (1999) The *selC*-associated SHI-2 pathogenicity island of *Shigella flexneri*. *Mol Microbiol* 33:74–83
- Mulvey MA, Lopez-Boado, Wilson CL, Roth R, Parks WC, Heuser J, Hultgren SJ (1998) Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. *Science* 282:1494–1497

- Ochman H, Bergthorsson U (1998) Rates and patterns of chromosome evolution in enteric bacteria. *Curr Opin Microbiol* 1:580–583
- Perna NT, Plunkett G, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Posfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Limk A, Dimalanta ET, Potamitis KD, Apodaca J, Anantharaman TS, Lin JY, Yen G, et al (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529–533
- Phillips JA, Li X, Guyer DM, Mobley HLT (2000) Identification of a second pathogenicity island in uropathogenic strain CFT073. *ASM 100th General Meeting*, Abstract B-172, p 80
- Rakin A, Schubert S, Pelludat C, Brem D, Heesemann J (1999) The high-pathogenicity island of *Yersinia*. In: Kaper JB, Hacker J (eds) *Pathogenicity islands and other mobile virulence elements*. ASM Press, Washington, DC
- Ritter A, Blum G, Emody L, Kerenyi M, Bock A, Heuhierl B, Rabsch W, Scheutz F, Hacker J (1995) tRNA genes and pathogenicity islands: influence on virulence and metabolic properties of uropathogenic *Escherichia coli*. *Mol Microbiol* 17:109–121
- Sperandio V, Kaper JB, Bortolini MR, Neves C, Keller R, Trabulsi LR (1998) Characterization of the locus of enterocyte effacement (LEE) in different enteropathogenic *Escherichia coli* (EPEC) and Shiga-toxin producing *Escherichia coli* (STEC) serotypes. *FEMS Microbiol Lett* 164:133–139
- Sullivan JT, Ronson CW (1998) Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates at the phe-tRNA gene. *Proc Natl Acad Sci USA* 95:5145–5149
- Swenson D, Bukanov N, Berg D, Welch R (1996) Two pathogenicity islands in uropathogenic *Escherichia coli*: cosmid cloning and sample sequencing. *Infect Immun* 64:3736–3743
- Welch, RA, Falkow S (1984) Characterization of *Escherichia coli* hemolysins conferring quantitative differences in virulence. *Infect Immun* 43:156–160
- Wenzel R (1991) Epidemiology of hospital-acquired infection. In: Balows A, Hausler W, Herrmann K, Isenberg H, Shadomy H (eds) *Manual of clinical microbiology*. ASM Press, Washington DC
- Wullt B, Bergsten G, Connell H, Röllano P, Gebretsadik N, Hull R, Svanborg C (2000) P fimbriae enhance the early establishment of *Escherichia coli* human urinary tract. *Mol Microbiol* 38:456–464

# Pathogenicity Islands of Intestinal *E. coli*

A.G. TORRES and J.B. KAPER

1	Introduction . . . . .	31
2	Enteropathogenic <i>E. coli</i> . . . . .	32
2.1	LEE . . . . .	33
2.1.1	Structure . . . . .	33
2.1.2	Regulation . . . . .	36
2.2	EspC PAI . . . . .	37
3	Enterohemorrhagic <i>E. coli</i> . . . . .	38
3.1	LEE . . . . .	39
3.2	Other O-islands . . . . .	39
4	Enterotoxigenic <i>E. coli</i> . . . . .	40
4.1	Tia-PAI . . . . .	41
5	Enteroaggregative <i>E. coli</i> . . . . .	41
5.1	High-Pathogenicity Island . . . . .	42
5.2	she PAI . . . . .	43
6	Enteroinvasive <i>E. coli</i> . . . . .	43
6.1	SHI-2 . . . . .	43
6.2	Black Holes . . . . .	44
7	Concluding Remarks . . . . .	44
	References . . . . .	45

## 1 Introduction

*Escherichia coli* is a consistent inhabitant of the human colonic flora, and it is the predominant facultative organism in the human gastrointestinal tract; however, it makes up a very small proportion of the total intestinal bacterial content, which is dominated by anaerobic species. The regular presence of *E. coli* in the human intestine and feces has led to tracking the bacterium in nature as an indicator of fecal pollution and water contamination. This is especially important in developing countries, where *E. coli* infections are one of the leading causes of infant and

childhood morbidity and mortality due to ingestion of contaminated food and water. As a pathogen, *E. coli*, is best known for its ability to cause diarrhea. Five main categories of *E. coli* that cause diarrheal diseases are now recognized: enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), and enteroinvasive *E. coli* (EIEC). The ability of each category to cause disease depends on a specific array of genes encoding virulence factors. These virulence factors are usually encoded on bacteriophages, plasmids, transposons, and pathogenicity islands (PAIs) (reviewed in KAPER et al. 1999). In this chapter, we will focus on describing the main categories of diarrheagenic *E. coli* strains and their various PAIs that differentiate these organisms from nonpathogenic *E. coli* strains (Table 1). A sixth category not discussed in this chapter, diffusely adherent *E. coli* (DAEC), is a heterogeneous group of organisms with conflicting results about their significance in pathogenesis and with no PAIs yet reported. Additional categories have been suggested, but conclusive epidemiological and clinical information is so far lacking for the consideration of these diarrheagenic *E. coli* strains as new major categories. A complete overview of the pathogenesis, epidemiology, clinical significance, detection, and diagnosis of these categories of diarrheagenic *E. coli* strains has recently been published (NATARO and KAPER 1998).

## 2 Enteropathogenic *E. coli*

EPEC is a major etiological agent of infant diarrhea in developing countries and the prototype for a family of pathogens exhibiting the unique virulence mechanism known as “attaching and effacing” (A/E) (reviewed in NATARO and KAPER 1998). Adherence of EPEC strains and the formation of the A/E lesions in the intestinal mucosa are processes associated with localized degeneration of the brush border microvilli and assembly of highly organized pedestal-like actin structures in the epithelial cells beneath intimately attached bacteria (reviewed in NATARO and KAPER 1998). The chromosomal region found to be necessary and sufficient for the

**Table 1.** Characterized and putative pathogenicity islands in intestinal *E. coli* strains

Category of <i>E. coli</i>	Pathogenicity island	Phenotype
EPEC	LEE	Attaching and effacing (A/E) histopathology
EPEC	EspC-PAI	EspC enterotoxin
EHEC	LEE and eight large O-islands	A/E histopathology and several putative virulence factors
ETEC	Tia-PAI	Invasion and enterotoxin secretion
EAEC/EHEC	HPI ( <i>Yersinia</i> ) <sup>a</sup>	Iron utilization
EAEC	<i>she</i> PAI ( <i>S. flexneri</i> ) <sup>a</sup>	Pic protease and ShET1 enterotoxin
EIEC	SHI-2 ( <i>S. flexneri</i> ) <sup>a</sup>	Iron utilization

<sup>a</sup> These PAIs were initially characterized in the pathogens listed in parentheses.

induction of the A/E lesions on epithelial cells is the locus of enterocyte effacement (LEE) (McDANIEL et al. 1995; McDANIEL and KAPER 1997). In addition to the virulence factors encoded in the LEE, EPEC produces a bundle-forming pilus (BFP) (GIRÓN et al. 1991) encoded on a large 60-MDa plasmid. BFP was originally proposed to mediate initial adherence to the host cell, but subsequent studies in a human intestinal organ culture model have indicated that BFP mediates bacterium-to-bacterium interaction rather than epithelial cell adherence (HICKS et al. 1998). We will discuss in this section only the chromosomal elements associated with the A/E lesion and enterotoxin production.

## 2.1 LEE

### 2.1.1 Structure

The first description of a gene in the LEE PAI associated with intimate adherence of EPEC to epithelial cells was reported by JERSE et al. (1990), who screened *TnphoA* mutants of the prototype EPEC E2348/69 for loss of the A/E phenotype. The chromosomal gene identified in this study, *eae*, encodes intimin, a 94-kDa outer membrane protein required for intimate adherence that was recognized by sera from volunteers convalescing from experimental EPEC infections (JERSE et al. 1990, 1991). Further analysis of additional mutants in this region that were also negative for the A/E phenotype led McDANIEL et al. (1995) to identify a 35-kb locus containing numerous genes involved in the formation of these lesions. The LEE is present in all EPEC strains as well as in most A/E-positive strains, including *E. coli* O157:H7, and contains all pathogen-specific genes necessary for inducing A/E lesions (McDANIEL et al. 1995; McDANIEL and KAPER 1997). The LEE is not present in normal flora *E. coli*, uropathogenic *E. coli*, ETEC, EIEC, or EAEC. In EPEC E2348/69 and EHEC O157:H7 the LEE is inserted at the selenocysteine (*selC*) tRNA locus, the same location as other PAIs in ETEC, uropathogenic *E. coli*, *Shigella flexneri*, and *Salmonella enterica* serovar Typhimurium (see descriptions below and in Chaps. 2, 4, and 5 of this volume).

The complete sequence of the LEE from EPEC strain E2348/69 has been determined (ELLIOTT et al. 1998) and contains 35,624-bp with an average G+C content of 38.36% (far below the *E. coli* chromosome which averages 50.8%). Transfer of this region to a nonpathogenic *E. coli* K-12 strain is sufficient to confer the A/E phenotype (McDANIEL and KAPER 1997). The LEE contains 41 ORFs (of >50 amino acids) arranged in at least five polycistronic operons (Fig. 1). For functional purposes, the LEE is divided into three domains.

The middle domain contains the *eae* and *tir* genes. The *eae* encodes the intimin protein that mediates intimate attachment to the host cell (JERSE et al. 1991). Its receptor in the host cell is a 90-kDa tyrosine-phosphorylated protein named Tir (translocated intimin receptor), which was initially identified as a host protein (ROSENSHINE et al. 1992, 1996) but subsequently shown to be a bacterial protein that is translocated into the host cell via a type III secretion system (KENNY et al.

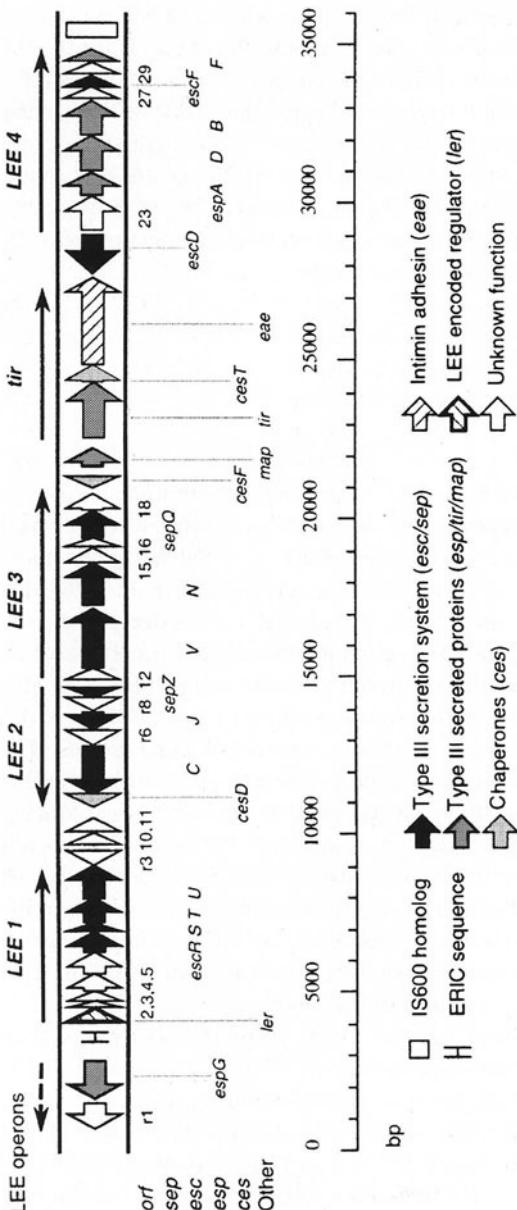


Fig. 1. Map of the LEE PAI of EPEC strain E2348/69. The different patterns of the large arrows indicate the functions of the proteins encoded by these genes

1997b). Tir is implicated in several processes, including serving as a receptor for intimin, thereby linking the bacterium to the host cell cytoskeleton and initiating other signaling events within the host cell (KENNY et al. 1997). Between *tir* and *eae* genes is the *cesT* (previously called *orfU*) gene, which encodes a chaperone for the Tir protein (ELLIOTT et al. 1999a).

The second domain is located upstream of the *eae/tir* domain and encodes a type III secretion system. At least 12 genes have been implicated as encoding putative components of the type III apparatus, based on homology to components of type III secretion systems of *Yersinia*, *Salmonella*, *Shigella*, and other pathogens (ELLIOTT et al. 1998). These genes were originally called *sep* (*E. coli* secretion) to follow the nomenclature proposed for the *Yersinia* type III secretion genes and two genes, *sepZ* and *sepQ*, retained their original name (ELLIOTT et al. 1998). In addition to the type III secretion apparatus, this domain contains the *cesD* and *cesF* (previously called *orf10*) genes. CesD is a chaperone protein essential for the secretion of EspD and EspB (WAINWRIGHT and KAPER 1998) and CesF is the type III secretion chaperone for EspF (ELLIOTT et al. 2001b). Recently, Map and EspG, the protein products of *map* (*orf19*) and *espG* (*orf2*) genes, were identified as type III-secreted proteins encoded in this domain. Map is an effector protein targeted to the host mitochondria (KENNY and JEPSON 2000), and EspG is functionally equivalent to VirA of *Shigella* and plays an unknown role in EPEC pathogenesis (ELLIOTT et al. 2001a). Finally, the *ler* gene encodes the LEE-encoded regulator (MELLIES et al. 1999), which is discussed below in the section on regulation.

The third domain located downstream of the *eae/tir* domain encodes the EPEC secreted proteins (Esps), which are secreted through the type III secretion apparatus and encoded by the *espA*, *espB*, *espD*, and *espF* genes. EspA is a 25-kDa protein that forms a filamentous structure on the surface of EPEC which is involved in protein translocation into epithelial cells (KNUTTON et al. 1998). *espB* (originally named *eaeB*) was initially identified by a *TnphoA* mutation that abolished the A/E phenotype (DONNENBERG et al. 1993). The 37-kDa protein encoded by the *espB* gene is produced and secreted upon contact with tissue culture cells and has been localized both in the cytoplasm and in the cell membrane of the host cell (TAYLOR et al. 1998; WOLFF et al. 1998). EspB has been suggested to be involved in triggering signal transduction events in the host cells (TAYLOR et al. 1998). EspD is a 40-kDa protein that is inserted into epithelial cell membranes (LAI et al. 1997; WACHTER et al. 1999). Recently, it has been proposed that EspB and EspD form a pore in the host membrane at the distal end of the EspA filament (WACHTER et al. 1999; WARAWA et al. 1999; HARTLAND et al. 2000). Finally, EspF is an effector protein associated with the disruption of the host intestinal barrier functions (MCNAMARA and DONNENBERG 1998; MCNAMARA et al. 2001).

In contrast to PAIs found in uropathogenic *E. coli*, the LEE is a stable element in the chromosome of EPEC E2348/69 that lacks obvious IS elements or phage sequences. Regardless of this apparent stability, WIELER et al. (1997) found that the insertion site of the LEE varies according to the evolutionary lineage. In EPEC O127:H6 and EHEC O157:H7, the LEE is inserted downstream of *selC* at 82 min in

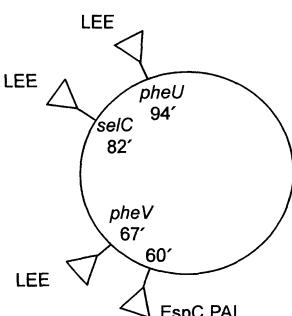
the K-12 chromosome, but in EPEC O111:H8/H<sup>-</sup> and Shiga toxin-producing *E. coli* (STEC) O26:H11/H<sup>-</sup>, the LEE is inserted at the *pheU* tRNA locus at 94 min (BENKEL et al. 1997; SPERANDIO et al. 1998). A third insertion site of LEE has recently been described by JORES et al. (2001) in a bovine STEC O103:H2 strain. The LEE in this strain is inserted in the *pheV* tRNA at 67 minutes. In addition to the 35-kb “core” LEE sequences seen in EPEC E2348/69 and EHEC O157:H7, this LEE contains additional DNA so that the total size is at least 80kb. This element is also flanked by apparently intact IS elements, suggesting potential mobility of this PAI (JORES et al. 2001). The relative positions of these different LEE insertion sites on the *E. coli* chromosome are given in Fig. 2.

### 2.1.2 Regulation

The LEE PAI from EPEC and EHEC strains is organized in five major predicted operons (MELLIES et al. 1999). The components of the type III secretion system are transcribed from three polycistronic operons designated *LEE1*, *LEE2*, and *LEE3* (Fig. 1). The secreted Esp molecules are part of a polycistronic operon designated *LEE4* and the fifth operon, *tir*, includes *tir*, *cesT* and *eae*. The structure of these operons has been confirmed by reverse transcriptase-polymerase chain reaction (RT-PCR) in EPEC E2348/69 (MELLIES et al. 1999).

In EPEC, there is a plasmid-encoded regulator of virulence genes called Per, which is encoded by three ORFs, *perA*, *perB* and *perC* (GÓMEZ-DUARTE and KAPER 1995). Per up-regulates the expression of BFP and intimin in EPEC (GÓMEZ-DUARTE and KAPER 1995; TOBE et al. 1996) and regulates Esp's secretion in response to different environmental cues (KENNY et al. 1997a). Per also activates the expression of *ler*, which then activates the expression of *LEE2*, *LEE3*, *tir* and *LEE4* in a cascade fashion (MELLIES et al. 1999).

*ler*, the first gene in the *LEE1* operon (originally termed *orfI*), encodes a protein able to regulate virulence gene expression and is part of a regulatory cascade involving Per (MELLIES et al. 1999; ELLIOTT et al. 2000). Ler activates the transcription of several LEE operons, with the exception of *LEE1*, and is required for expression of LEE-encoded proteins (FRIEDBERG et al. 1999; MELLIES et al. 1999; ELLIOTT et al. 2000). Activation of the divergently transcribed *LEE2* and



**Fig. 2.** Relative position of the LEE and EspC PAIs in the *E. coli* chromosome. The map coordinates in minutes are based on the *E. coli* K-12 chromosome. The LEE insertion site at *selC* is that of EPEC O127:H6 (E2348/69) and EHEC O157:H7 (McDANIEL et al. 1995), the LEE insertion site at *pheU* is that of EPEC O111:H8/H<sup>-</sup> and STEC O26:H11/H<sup>-</sup> (BENKEL et al. 1997, SPERANDIO et al. 1998), and the LEE insertion site at *pheV* is that of STEC O103:H2 (JORES et al. 2001). The insertion site of the EspC PAI is that of EPEC O127:H6 (MELLIES et al. 2001)

*LEE3* operons by Ler has recently been characterized (SPERANDIO et al. 2000; BUSTAMANTE et al. 2001). Ler shows similarity with a family of histone-like proteins, such as H-NS (ELLIOTT et al. 2000). H-NS is a 15.6-kDa DNA-binding protein that modulates, usually negatively, the expression of genes involved in many different cellular processes. It has been shown that in the absence of H-NS, expression of the *LEE2* and *LEE3* operons became Ler independent, suggesting that Ler acts as an antirepressor protein that overcomes the H-NS-mediated silencing on the *LEE2/LEE3* promoter region (BUSTAMANTE et al. 2001).

Another level of regulation of the LEE genes occurs with the host integration factor protein (IHF). This protein acts as a global regulator and is essential for the activation of *ler* (FRIEDBERG et al. 1999). In addition, quorum sensing is also involved in the regulation of virulence gene expression in EPEC and EHEC, as recently shown by SPERANDIO et al. (1999). Quorum sensing controls the *LEE1* and *LEE2* operons, and regulation of *LEE1* in turn increases the expression of the *LEE3* and *LEE4* operons via Ler (SPERANDIO et al. 1999). The regulation of the LEE genes in EPEC is summarized in Fig. 3.

## 2.2 EspC PAI

EspC is a 110-kDa protein secreted by EPEC E2348/69 that is not involved in the A/E phenotype (STEIN et al. 1996) and does not require the type III secretion system for delivery into the extracellular milieu (JARVIS et al. 1995; STEIN et al. 1996). This protein is a member of the immunoglobulin A (IgA) protease family of autotransporters, and it has recently been determined that EspC possesses enterotoxic activity (MELLIES et al. 2001). Culture supernatants from a laboratory strain of *E. coli* expressing EspC increased tissue transepithelial electrical potential

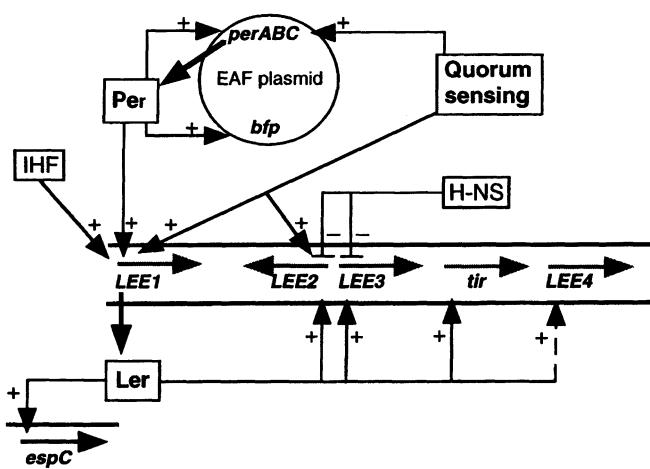


Fig. 3. Summary of regulation of the LEE in EPEC E2348/69. See text for details

difference (PD) and total tissue conductance (Isc) in rat jejunal tissue mounted in Ussing chambers (MELLIES et al. 2001).

*espC* is located in a 15,195-bp region with a G+C content of 40.5%. This region, designated EspC PAI, contains two loci potentially associated with virulence, *espC* and *orf3*, a gene encoding a protein with homology to VirA in *S. flexneri* (MELLIES et al. 2001). In addition, several ORFs with predicted protein similarity to a variety of mobile elements were identified in this PAI. The EspC PAI was mapped to a region between 59.35 and 60.06 min of the sequence of the *E. coli* K-12 chromosome. This observation suggests that approximately 33-kb of DNA between this location in *E. coli* K-12 are missing in EPEC E2348/69.

Southern blot analysis using *espC* probes detected homologous sequences in EPEC serotypes O127:H6 and O55:H6, but not in non-DEC1 EPEC strains and not in other A/E pathogens, including EHEC, rabbit diarrheagenic *E. coli* (RDEC-1), or *Citrobacter rodentium* strains (MELLIES et al. 2001), suggesting that the EspC PAI is only found in a subset of these pathogens. Further analysis is under way to determine the role of this PAI as an accessory virulence factor in this subset of EPEC strains.

### 3 Enterohemorrhagic *E. coli*

EHEC strains have recently been recognized as a cause of serious disease in the United States and in several industrialized countries due to the ingestion of contaminated food or water. EHEC, represented by the prototypic serotype O157:H7, colonizes the intestine and can cause nonbloody diarrhea or a diarrheal syndrome characterized by a copious bloody discharge. A subset of patients, particularly the very young or very old, can go on to develop hemolytic uremic syndrome (HUS), a potentially fatal kidney disease (reviewed in NATARO and KAPER 1998).

EHEC strains produce potent cytotoxins (Stx) that are virtually identical to the Shiga toxin produced by *S. dysenteriae*. Stx toxins are responsible for the severe damage to epithelial and endothelial cells by EHEC and are associated with the ability to cause HUS. These toxins are phage or chromosomally encoded and possess the basic A-B toxin subunit structure (reviewed in NATARO and KAPER 1998). Little is known about the colonization factors of EHEC other than the LEE-encoded intimin and Tir, but recent data indicate that fimbriae participate in the adherence of *E. coli* O157:H7 to eukaryotic cells and play a role in microcolony formation (TORRES et al. 2001). In addition, EHEC strains express the same LEE-encoded virulence factors found in EPEC (see below) and additional potential virulence factors encoded in a large virulence plasmid called pO157. The complete DNA sequence of the plasmid (BURLAND et al. 1998; MAKINO et al. 1998) and the chromosome of EHEC O157:H7 (PERNA et al. 2001) have been completely elucidated. The DNA sequence analysis led to the identification of new candidate genes potentially involved in the pathogenesis of disease due to EHEC. We will discuss in this section those genes located in the LEE and other eight putative islands containing virulence genes in the chromosome of EHEC.

### 3.1 LEE

PERNA et al. (1998) determined the complete nucleotide sequence of the LEE from the prototype EHEC O157:H7 strain EDL933. Similar to the EPEC LEE, this PAI is involved in the production of proteins associated with the intimate adherence to intestinal epithelial cells, initiation of host signal transduction pathways, and the formation of A/E lesions (reviewed in NATARO and KAPER 1998). The EHEC LEE comprises 43,359-bp in comparison to 35,624-bp for the EPEC LEE with a G + C content of 40.91%. The EHEC O157:H7 LEE is inserted in the same chromosomal location as the LEE in EPEC E2348/69. The EHEC LEE contains 54 predicted ORFs, of which 41 genes are common to both the EHEC and the EPEC LEEs. Of the 13 ORFs that are not present in the EPEC LEE, all fall within a putative prophage, designated 933L, that is located next to the *selC* locus. The 933L prophage is a member of the P4 family of cryptic prophages, serving as the site of integration for the related retronphage FR73 and another PAIs in pathogenic *E. coli* and *Shigella* strains (PERNA et al. 1998, 2001).

Other than the putative prophage, the EPEC and EHEC LEEs share 93.9% overall nucleotide identity with a few genes showing markedly higher variability (reviewed in KAPER et al. 1999). They include the *espB* (25.99% difference), *espD* (19.64%), and *espA* (15.37%) genes encoding secreted proteins. Other genes that showed marked differences were the intimin gene *eae* (12.77% difference) and the intimin receptor gene, *tir* (33.52%). (The variation in the intimin protein is heavily concentrated in the C-terminal region, which shares only 49% identity in the last 25% of the intimin protein compared with 94% identity in the first 75% of the protein.) In EPEC and EHEC O26:H11/H<sup>-</sup> strains, the Tir protein is translocated into the host cell and becomes phosphorylated but in EHEC O157:H7, Tir is not tyrosine phosphorylated upon translocation, due to the substitution of a serine for a tyrosine residue in the EHEC O157:H7 Tir (KENNY 1999).

Another difference found between the EPEC and the EHEC LEEs was observed when the two PAIs were cloned and transformed into an *E. coli* K-12 strain (McDANIEL and KAPER 1997; ELLIOTT et al. 1999b). While the cloned EPEC LEE confers the A/E phenotype when tested on Hep-2 cells, the cloned EHEC LEE was unable to confer the attaching and effacing phenotype (ELLIOTT et al. 1999b). Together, these differences suggest that the phenotypes produced by the EPEC and EHEC LEEs have been subjected to natural selection for adaptation to the host microenvironments or to evade the host immune system (PERNA et al. 1998).

### 3.2 Other O-Islands

The recent completion and analysis of the chromosomal sequence of EHEC strain EDL933 allowed Perna and colleagues (2001) to identify DNA segments, designated O-islands, that are present in EHEC but not in *E. coli* K-12 strain MG1655 as well as K-islands that are unique segments present in *E. coli* K-12 but not in EDL933. The “backbone” of sequences shared by EHEC and K-12 total 4.1-Mb. In addition to this

shared backbone, there 1.34-Mb of DNA that are considered O-islands and 0.53-Mb of DNA that are K-islands. The EDL933 chromosome contains 177 O-islands and 234 K-islands greater than 50-bp in length (PERNA et al. 2001). Of the O-islands, nine are larger than 15-kb and were found to encode putative virulence factors. One of this islands, designated OI#148, encodes the previously described LEE PAI (PERNA et al. 1998). The other eight putative islands encode a macrophage toxin and ClpB-like chaperone (OI#7), a RTx-toxin-like exoprotein and transport system (OI#28), two urease gene clusters (OI#43 and #48), an adhesin and polyketide or fatty-acid biosynthesis system (OI#47), a type III secretion system and secreted proteins similar to the *Salmonella-Shigella inv-spa* host cell invasion genes (OI#115), two toxins and a PagC-like virulence factor (OI#122), and a fatty acid biosynthesis system (OI#138). Among these nine large islands, four include genes encoding a member of the P4-family of integrases and are inserted next to tRNA genes (OI#43-*serW*, #48-*serX*, #122-*pheV* and #148-*selC*) (PERNA et al. 2001). Besides the LEE PAI, none of these O-islands have been associated with virulence to date, but we can speculate that these O-islands provide EHEC O157:H7 with a range of putative proteins for the survival in the environment and colonization within the human host.

In addition to these islands in EHEC O157:H7, a high-pathogenicity island (HPI), first characterized in pathogenic *Yersinia* strains, has been identified in EHEC O26:H11 (KARCH et al. 1999). The *E. coli* HPI, which was first described in EAEC, will be described in the EAEC section of this chapter.

## 4 Enterotoxigenic *E. coli*

ETEC strains are a major cause of childhood diarrheal disease in developing countries and a frequent cause of enteric illness among travelers to those countries (reviewed in SACK 1975 and NATARO and KAPER 1998). ETEC are acquired by ingestion of contaminated food and water and produce a range of diseases that vary from minor discomfort to severe cholera-like symptoms. ETEC colonize the proximal small intestine by means of the plasmid-encoded fimbrial colonization factor antigens, e.g., CFA I and CFA II (or CS antigens) in ETEC infections of humans or K-88 and K-99 in the case of animal ETEC infection (reviewed in NATARO and KAPER 1998). After colonization, ETEC strains elaborate two classes of plasmid-encoded enterotoxins that are responsible for net secretion of fluid into the intestine and diarrhea. The first class comprises the high-molecular-weight, heat-labile enterotoxins (LTI and LTII), that are structurally and functionally related to cholera toxin and the second class comprises low-molecular-weight, heat-stable enterotoxins called ST<sub>A</sub> and ST<sub>B</sub> (reviewed in SEARS and KAPER 1996 and NATARO and KAPER 1998). The LTII and ST<sub>B</sub> toxins are found nearly exclusively in animal ETEC strains and human strains may produce either LTI, ST<sub>A</sub> or both. In addition, ETEC strains have been reported to invade cultured human epithelial cells, although there are no clinical data to suggest that ETEC invades in vivo (ELSINGHORST and KOPECKO 1992;

ELSINGHORST and WEITZ 1994). The virulence factors associated with this invasion phenotype were identified with the characterization of the Tia-PAI.

#### 4.1 Tia-PAI

The initial observation by ELSINGHORST and KOPECKO (1992), suggesting that the prototypic ETEC strain H10407 invades human intestinal cell lines, led to the identification and characterization of two chromosomally encoded invasion loci, designated *tia* and *tib* (toxigenic invasion loci A and B) (ELsinghorst and Kopecko 1992; ELSINGHORST and WEITZ 1994). These two distinct loci were cloned and showed to confer to a noninvasive *E. coli* strain the ability to adhere to and invade intestinal epithelial cell lines (ELsinghorst and Kopecko 1992). The *tia* locus directs the synthesis of Tia, a 25-kDa outer membrane protein that is sufficient to confer the adherence and invasive phenotypes (FLECKENSTEIN et al. 1996). Recent studies with the purified protein indicate that Tia acts as an adhesin and invasin that binds to a specific receptor on the human intestinal cell line HCT8 (MAMMARAPPALLIL and ELSINGHORST 2000). Further analysis of the region flanking the *tia* locus revealed that this locus is located on a large chromosomal element of approximately 46-kb that is inserted at *selC*, the same insertion site of several PAIs in pathogenic *E. coli* and *Shigella* strains (FLECKENSTEIN et al. 2000). Sequence data of a 37-kb portion of this region revealed a G+C content of 43.7% compared with 50.8% overall composition for *E. coli* K-12. The presence in this region of genes encoding proteins with homology to the prophage P4 integrase, CP4-like integrase, and IS2 and the low G+C content, suggests horizontal transfer of the island (FLECKENSTEIN et al. 2000). Sequence analysis downstream of the *tia* gene identified four candidate open reading frames (ORFs) encoding putative proteins associated with bacterial secretion apparatuses. One candidate gene identified as *leoA* (labile enterotoxin output) was found to be required for maximal secretion of the LT toxin (FLECKENSTEIN et al. 2000). *tib*, the second locus in the chromosome of ETEC H10407 and specific ETEC strains, directs the synthesis of TibA, a 104-kDa outer membrane glycoprotein involved in epithelial cell adhesion and invasion (ELsinghorst and WEITZ 1994; LINDENTHAL and ELSINGHORST 2001). The *tib* locus also appears to be encoded on a large chromosomal region with a low G+C composition that is distinct from the *tia* locus (FLECKENSTEIN et al. 2000), but more information is required to determine if this locus is indeed a PAI.

### 5 Enteropathogenic *E. coli*

The distinctive feature of EAEC strains is their ability to attach to cultured HEp-2 cells in an aggregative adherence (AA) manner. EAEC strains have been

associated with persistent diarrhea in young children in developing countries, although recent reports have also implicated these organisms in sporadic endemic diarrhea worldwide (reviewed in NATARO and KAPER 1998). EAEC resemble ETEC strains in their ability to cause nonbloody diarrhea without invading or causing inflammation, but EAEC do not secrete the enterotoxins LT or ST.

Most of the virulence factors associated with the pathogenesis of EAEC have been localized to a 65-MDa plasmid (reviewed in KAPER et al. 1999). These factors included two fimbrial antigens (AAF/I and AAF/II) implicated in the adherence to human epithelial cells and two putative enterotoxins, Pet and EAST. Pet is a member of the autotransporter family of secreted proteins and is a serine protease involved in damage to human colonic tissue. EAST1 is a ST-like toxin that is present in both pathogenic and nonpathogenic *E. coli* strains with no role in enteric pathogenesis yet described. In addition, three chromosomal loci associated with virulence have been identified and linked to putative PAIs.

## 5.1 High-Pathogenicity Island

The HPI was first identified in pathogenic *Yersinia* strains as a chromosomal determinant absolutely essential for virulence in mice (FETHERSTON et al. 1992; CARNIEL et al. 1996; BUCHRIESER et al. 1998). The *Yersinia* HPI is extensively described in the chapter by Carniel in this volume. This 35- to 45-kb chromosomal region contains a cluster of 11 genes (*irp1-9*, *ybtA*, *fyuA*) involved in regulation, biosynthesis, and uptake of the siderophore yersiniabactin. HPI is inserted adjacent to an asparagine-specific tRNA loci and contains a P4-like integrase gene, *int*, often found at the integration site of several PAIs (CARNIEL et al. 1996; BUCHRIESER et al. 1998).

The HPI is frequently found in EAEC as well as in extraintestinal *E. coli* strains associated with urinary tract infections and sepsis, but it is also found in some EIEC, ETEC, and EPEC strains and a subset of STEC isolates (SCHUBERT et al. 1998; KARCH et al. 1999). In EIEC strains, the HPI is always inserted in the *asnT* tRNA gene, in contrast to *Y. pseudotuberculosis* where the HPI can be found in any of three *asn* tRNA genes (BUCHRIESER et al. 1998; SCHUBERT et al. 1999). The stability of the HPI in the *asnT* tRNA gene in EIEC and other *E. coli* strains has been associated with the inability of this island to transpose into a new *asn* tRNA gene. The DNA sequence analysis of the island in *E. coli* revealed some deletions in the region downstream of the *asnT* tRNA gene and the absence of two 17-bp direct repeats. This 17-bp duplication is found flanking the HPI in *Y. pseudotuberculosis* and has been implicated with the mobility of the island (SCHUBERT et al. 1999; BACH et al. 2000).

The role of HPI in *E. coli* virulence has not been demonstrated. Instead, HPI has been proposed to contribute to the fitness of *E. coli* strains in certain ecological niches (KARCH et al. 1999).

## 5.2 *she* PAI

The *she* PAI is a 46-kb unstable chromosomal locus inserted next to the *pheV* tRNA gene in *S. flexneri* serotype 2a. This PAI carries genes encoding the ShET1 enterotoxin, three autotransporter proteins called SigA, Pic, and Sap, a P4-like integrase, and other ORFs of unknown function (AL-HASANI et al. 2001). CZECZULIN et al. (1999) showed that 57% of the EAEC strains they examined carried genes homologous to the *she* PAI. Although the complete sequence of the *she* PAI has not been reported for any EAEC strain, both hybridization analysis and functional studies have suggested that these genes are located in a similar but not identical PAI in EAEC. For example, Pic is a secreted protease implicated in mucinase activity, serum resistance, and hemagglutination in both *S. flexneri* and EAEC strains (HENDERSON et al. 1999). The *pic* gene, an allele that overlaps the *set1A* and *set1B* genes encoding the two subunits of ShET1, is located in the *she* PAI in *S. flexneri*, but the flanking sequence in EAEC differs from that of the identical *pic* gene in *Shigella*. A number of IS-like elements were found flanking the *pic* gene in EAEC that are not present in *S. flexneri* (HENDERSON et al. 1999; AL-HASANI et al. 2001). Additional evidence for these differences arises from the fact that EAEC lacks the *sigA* gene, which encodes another secreted autotransporter protein involved in intestinal fluid accumulation (AL-HASANI et al. 2000). The exact structure of the *she* PAI in EAEC and its precise contribution to pathogenesis and epidemiology is under further investigation.

A third potential virulence locus has been identified in the EAEC O42 strain. The *tia* locus, present in the Tia-PAI in ETEC strains and discussed above, was found by FLECKENSTEIN et al. (1996) in this prototypic EAEC strain but not in other EAEC strains. The location of this PAI in EAEC or the presence of additional genes within the island have not yet been reported.

## 6 Enteroinvasive *E. coli*

EIEC closely resemble *Shigella* species in their pathogenic mechanisms and the kind of clinical illness they produce. EIEC penetrate and multiply within epithelial cells of the colon, causing widespread cell destruction. The clinical manifestations of the disease and the virulence factors involved in the pathogenesis are reviewed in the chapter by Ingersoll et al., in this volume.

### 6.1 SHI-2

EIEC, as all *E. coli* strains, synthesize and transport the siderophore enterobactin, a low-molecular-weight iron chelator with a high affinity for iron (reviewed in EARHART 1996). A second siderophore, aerobactin, is frequently found among

invasive *E. coli* strains and has been associated with increased virulence (EARHART 1996). The genes involved in synthesis and transport of aerobactin, which are often plasmid encoded, were recently found within a pathogenicity island in two *S. flexneri* strains from serotypes 2a and 5 (Moss et al. 1999; VOKES et al. 1999). This island, designated SHI-2 (*Shigella* pathogenicity island 2), is located downstream of *selC* and occupies 23.8-kb in *S. flexneri* serotype 5 and 30-kb in *S. flexneri* serotype 2a. The G + C content of the islands is slightly lower than that of the rest of the *Shigella* chromosome (51%) and varies from 48.5% in serotype 5 to 46% in serotype 2. In addition to the aerobactin transport and synthesis genes, SHI-2 contains an *int* gene nearly identical to the integrase gene found in EHEC O157:H7, several IS elements, and genes encoding proteins which confer immunity to colicins I and V (Moss et al. 1999; VOKES et al. 1999). Southern hybridization and PCR analysis indicated that the aerobactin genes are not located within SHI-2 in EIEC. Instead, these genes were found in a different chromosomal location in EIEC (VOKES et al. 1999). The presence of the aerobactin genes in multiple locations and the fact that they are flanked by insertion sequences suggest the possibility that these genes are associated with a separate PAI in EIEC.

## 6.2 Black Holes

One common biochemical property of EIEC and *Shigella* strains is their lack of lysine decarboxylase (LDC) activity. The LDC<sup>+</sup> phenotype has been proposed to be associated with virulence, and MAURELLI et al. (1998) found that introduction of *cadA* (the gene for LDC) in *S. flexneri* 2a causes attenuation in its virulence and inhibition of enterotoxin production. Comparison of the *cadA* region of *E. coli* K-12 with the corresponding regions in EIEC and *S. flexneri* revealed a large chromosomal deletion in these two pathogens of up to 90-kb in the vicinity of the *cadA* gene. This novel concept of "black holes" is suggested to provide an evolutionary advantage that enables *S. flexneri* and EIEC to enhance virulence, by deleting genes that are detrimental to the pathogenesis of the organisms.

## 7 Concluding Remarks

Recent completion of the *E. coli* O157:H7 genome sequence and the rapid progress in the elucidation of the genomes of other intestinal *E. coli* strains will allow the analysis and identification of putative virulence genes within unidentified PAIs. This is particularly important to determine the true contribution of the PAIs in the pathogenesis of diarrheagenic *E. coli*. Although PAIs have been identified in all the main diarrheagenic *E. coli* categories, a comprehensive analysis at the molecular level has been done only in the LEE of EPEC E2348/69 and EHEC O157:H7. More studies are needed to determine, for example, the regulatory networks controlling

the expression of the genes in the other PAIs, or to define those selective factors that contribute to the stable maintenance of these PAIs. Comparative analysis of the genomes of these pathogenic *E. coli* strains with other pathogens and non-pathogenic *E. coli* strains will help to predict what traits of virulence may emerge in the future.

*Acknowledgements.* Work in the laboratory was supported by Public Health Service grants AI-21657 and AI-41325 to J.B.K. A.G.T. is the recipient of a research supplement for underrepresented minorities from the NIAID, NIH. The authors thank Carl Brinkley and Simon Elliott for their assistance with the figures.

## References

- Al-Hasani K, Henderson IR, Sakellaris H, Rajakumar K, Grant T, Nataro JP, Robins-Browne R, Adler B (2000) The *sigA* gene which is borne on the *she* pathogenicity island of *Shigella flexneri* 2a encodes an exported cytopathic protease involved in intestinal fluid accumulation. *Infect Immun* 68:2457–2463
- Al-Hasani K, Rajakumar K, Bulach D, Robins-Browne R, Adler B, Sakellaris H (2001) Genetic organization of the *she* pathogenicity island in *Shigella flexneri* 2a. *Microb Pathog* 30:1–8
- Bach S, de Almeida A, Carniel E (2000) The *Yersinia* high-pathogenicity island is present in different members of the family Enterobacteriaceae. *FEMS Microbiol Lett* 183:289–294
- Benkel P, Hudel M, Chakraborty T (1997) The location of the LEE-region in the O26:H-Shiga toxin-producing *Escherichia coli* strain 413/89-1. Congress of the DGHM, Jena, Germany
- Buchrieser C, Brosch R, Bach S, Guiyoule A, Carniel E (1998) The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal *asn tRNA* genes. *Mol Microbiol* 30:965–978
- Burland V, Shao Y, Perna NT, Plunkett G, Sofia HJ, Blattner FR (1998) The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Res* 26:4196–4204
- Bustamante VH, Santana FJ, Calva E, Puente JL (2001) Transcriptional regulation of type III secretion genes in enteropathogenic *Escherichia coli*: Ler antagonizes H-NS-dependent repression. *Mol Microbiol* 39:664–678
- Carniel E, Guilvout I, Prentice M (1996) Characterization of a large chromosomal “high-pathogenicity island” in biotype 1B *Yersinia enterocolitica*. *J Bacteriol* 178:6743–6751
- Czeczulin JR, Whittam TS, Henderson IR, Nataro JP (1999) Phylogenetic analysis of enteroaggregative and diffusely adherent *Escherichia coli*. *Infect Immun* 67:2692–2699
- Donnenberg MS, Yu J, Kaper JB (1993) A second chromosomal gene necessary for intimate attachment of enteropathogenic *Escherichia coli* to epithelial cells. *J Bacteriol* 175:4670–4680
- Earhart CF (1996) Uptake and metabolism of iron and molybdenum. In: *Escherichia coli* and *Salmonella*. Neidhardt, FC. Washington, DC, American Society for Microbiology Press, pp 1075–1090
- Elliott SJ, Wainwright LA, McDaniel TK, Jarvis KG, Deng YK, Lai LC, McNamara BP, Donnenberg MS, Kaper JB (1998) The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol Microbiol* 28:1–4
- Elliott SJ, Hutcheson SW, Dubois MS, Mellies JL, Wainwright LA, Batchelor M, Frankel G, Knutton S, Kaper JB (1999a) Identification of CesT, a chaperone for the type III secretion of Tir in enteropathogenic *Escherichia coli*. *Mol Microbiol* 33:1176–1189
- Elliott SJ, Yu J, Kaper JB (1999b) The cloned locus of enterocyte effacement from enterohemorrhagic *Escherichia coli* O157:H7 is unable to confer the attaching and effacing phenotype upon *E. coli* K-12. *Infect Immun* 67:4260–4263
- Elliott SJ, Sperandio V, Giron JA, Shin S, Mellies JL, Wainwright L, Hutcheson SW, McDaniel TK, Kaper JB (2000) The locus of enterocyte effacement (LEE)-encoded regulator controls expression of both LEE- and non-LEE-encoded virulence factors in enteropathogenic and enterohemorrhagic *Escherichia coli*. *Infect Immun* 68:6115–6126

- Elliott SJ, Krejany EO, Mellies JL, Robins-Browne RM, Sasakawa C, Kaper JB (2001a) EspG, a novel type III secreted protein from enteropathogenic *E. coli* with similarities to VirA of *Shigella*. *Infect Immun* 69:4027–4033
- Elliott SJ, O'Connell CB, Savkovic SD, Koutsouris A, Donnenberg MS, Hecht G, Kaper JB (2001b) A novel gene required for EPEC to increase tight-junction permeability encodes a chaperone for EspF. (in preparation)
- Elsinghorst EA, Kopecko DJ (1992) Molecular cloning of epithelial cell invasion determinants from enterotoxigenic *Escherichia coli*. *Infect Immun* 60:2409–2417
- Elsinghorst EA, Weitz JA (1994) Epithelial cell invasion and adherence directed by the enterotoxigenic *Escherichia coli* tib locus is associated with a 104-kilodalton outer membrane protein. *Infect Immun* 62:3463–3471
- Fetherston JD, Schuetze P, Perry RD (1992) Loss of the pigmentation phenotype in *Yersinia pestis* is due to the spontaneous deletion of 102kb of chromosomal DNA which is flanked by a repetitive element. *Mol Microbiol* 6:2693–2704
- Fleckenstein JM, Kopecko DJ, Warren RL, Elsinghorst EA (1996) Molecular characterization of the tia invasion locus from enterotoxigenic *Escherichia coli*. *Infect Immun* 64:2256–2265
- Fleckenstein JM, Lindler LE, Elsinghorst EA, Dale JB (2000) Identification of a gene within a pathogenicity island of enterotoxigenic *Escherichia coli* H10407 required for maximal secretion of the heat-labile enterotoxin. *Infect Immun* 68:2766–2774
- Friedberg D, Umanski T, Fang Y, Rosenshine I (1999) Hierarchy in the expression of the locus of enterocyte effacement genes of enteropathogenic *Escherichia coli*. *Mol Microbiol* 34:941–952
- Girón JA, Ho AS, Schoolnik GK (1991) An inducible bundle-forming pilus of enteropathogenic *Escherichia coli*. *Science* 254:710–713
- Gómez-Duarte OG, Kaper JB (1995) A plasmid-encoded regulatory region activates chromosomal *eaeA* expression in enteropathogenic *Escherichia coli*. *Infect Immun* 63:1767–1776
- Hartland EL, Daniell SJ, Delahay RM, Neves BC, Wallis T, Shaw RK, Hale C, Knutton S, Frankel G (2000) The type III protein translocation system of enteropathogenic *Escherichia coli* involves EspA-EspB protein interactions. *Mol Microbiol* 35:1483–1492
- Henderson IR, Czeczulin JR, Eslava C, Noriega F, Nataro JP (1999) Characterization of Pic, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*. *Infect Immun* 67:5587–5596
- Hicks S, Frankel G, Kaper JB, Dougan G, Phillips AD (1998) Role of intimin and bundle-forming pili in enteropathogenic *Escherichia coli* adhesion to pediatric intestinal tissue in vitro. *Infect Immun* 66:1570–1578
- Jarvis KG, Girón JA, Jerse AE, McDaniel TK, Donnenberg MS, Kaper JB (1995) Enteropathogenic *Escherichia coli* contains a specialized secretion system necessary for the export of proteins involved in attaching and effacing lesion formation. *Proc Natl Acad Sci U S A* 92:7996–8000
- Jerse AE, Yu J, Tall BD, Kaper JB (1990) A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. *Proc Natl Acad Sci USA* 87:7839–7843
- Jerse AE, Gicquelais KG, Kaper JB (1991) Plasmid and chromosomal elements involved in the pathogenesis of attaching and effacing *Escherichia coli*. *Infect Immun* 59:3869–3875
- Jores J, Roumer L, Kaper JB, Wieler LH (2001) Identification of a new pathogenicity island inserted in the *pheV* tRNA gene of the bovine Shiga toxin-producing *E. coli* strain RW1374 (O103:H2) harboring a locus of enterocyte effacement that is flanked by intact insertion elements. (Submitted)
- Kaper J, Mellies JL, Nataro JP (1999) Pathogenicity islands and other mobile genetic elements of diarrheagenic *Escherichia coli*. In: Kaper JB, Hacker J (eds) Pathogenicity islands and other mobile genetic elements. ASM Press, Washington DC
- Karch H, Schubert S, Zhang D, Zhang W, Schmidt H, Olschlager T, Hacker J (1999) A genomic island, termed high-pathogenicity island, is present in certain non-O157 Shiga toxin-producing *Escherichia coli* clonal lineages. *Infect Immun* 67:5994–6001
- Kenny B, Abe A, Stein M, Finlay BB (1997a) Enteropathogenic *Escherichia coli* protein secretion is induced in response to conditions similar to those in the gastrointestinal tract. *Infect Immun* 65:2606–2612
- Kenny B, DeVinney R, Stein M, Reinscheid DJ, Frey EA, Finlay BB (1997b) Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell* 91:511–520
- Kenny B (1999) Phosphorylation of tyrosine 474 of the enteropathogenic *Escherichia coli* (EPEC) Tir receptor molecule is essential for actin nucleating activity and is preceded by additional host modifications. *Molec Microbiol* 31:1229–1241
- Kenny B, Jepson M (2000) Targeting of an enteropathogenic *Escherichia coli* (EPEC) effector protein to host mitochondria. *Cell Microbiol* 2:579–590

- Knutton S, Rosenshine I, Pallen MJ, Nisan I, Neves BC, Bain C, Wolff C, Dougan G, Frankel G (1998) A novel EspA-associated surface organelle of enteropathogenic *Escherichia coli* involved in protein translocation into epithelial cells. *EMBO J* 17:2166–2176
- Lai L-C, Wainwright LA, Stone KD, Donnenberg MS (1997) A third secreted protein that is encoded by the enteropathogenic *Escherichia coli* pathogenicity island is required for transduction of signals and for attaching and effacing activities in host cells. *Infect Immun* 65:2211–2217
- Lindenthal C, Elsinghorst EA (2001) Enterotoxigenic *Escherichia coli* TibA glycoprotein adheres to human intestine epithelial cells. *Infect Immun* 69:52–57
- Makino K, Ishii K, Yasunaga T, Hattori M, Yokoyama K, Yutsudo CH, Kubota Y, Yamaichi Y, Iida T, Yamamoto K, Honda T, Han C-G, Ohtsubo E, Kasamatsu M, Hayashi T, Kuhara S, Shinagawa H (1998) Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic *Escherichia coli* O157:H7 derived from Sakai outbreak. *DNA Res* 5:1–9
- Mammarrappallil JG, Elsinghorst EA (2000) Epithelial cell adherence mediated by the enterotoxigenic *Escherichia coli* *Tia* protein. *Infect Immun* 68:6595–6601
- Maurelli AT, Fernandez RE, Bloch CA, Rode CK, Fasano A, Beltrametti F, Kresse AU, Guzman CA (1998) “Black holes” and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc Natl Acad Sci U S A* 95:3943–3948
- McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB (1995) A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci U S A* 92:1664–1668
- McDaniel TK, Kaper JB (1997) A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. *Mol Microbiol* 23:399–407
- McNamara BP, Donnenberg MS (1998) A novel proline-rich protein, EspF, is secreted from enteropathogenic *Escherichia coli* via the type III export pathway. *FEMS Microbiol Lett* 166:71–78
- McNamara BP, Koutsouris A, O’Connell CB, Nougaréde JP, Donnenberg MS, Hecht G (2001) Translocated EspF protein from enteropathogenic *Escherichia coli* disrupts host intestinal barrier function. *J Clin Invest* 107:1–10
- Mellies JL, Elliott SJ, Sperandio V, Donnenberg MS, Kaper JB (1999) The Per regulon of enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). *Mol Microbiol* 33:296–306
- Mellies JL, Navarro-Garcia F, Okeke I, Frederickson J, Nataro JP, Kaper JB (2001) *espC* pathogenicity island of enteropathogenic *Escherichia coli* encodes an enterotoxin. *Infect Immun* 69:315–324
- Moss JE, Cardozo TJ, Zychlinsky A, Groisman EA (1999) The *selC*-associated SHI-2 pathogenicity island of *Shigella flexneri*. *Mol Microbiol* 33:74–83
- Nataro JP, Kaper JB (1998) Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 11:142–201
- Perna NT, Mayhew GF, Posfai G, Elliott S, Donnenberg MS, Kaper JB, Blattner FR (1998) Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* 66:3810–3817
- Perna NT, Plunkett G, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Posfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch RA, Blattner FR (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529–532
- Rosenshine I, Donnenberg MS, Kaper JB, Finlay BB (1992) Signal transduction between enteropathogenic *Escherichia coli* (EPEC) and epithelial cells: EPEC induces tyrosine phosphorylation of host cell proteins to initiate cytoskeletal rearrangement and bacterial uptake. *EMBO J* 11:3551–3560
- Rosenshine I, Ruschkowski S, Stein M, Reincheid DJ, Mills SD, Finlay BB (1996) A pathogenic bacterium triggers epithelial signals to form a functional bacterial receptor that mediates actin pseudopod formation. *EMBO J* 15:2613–2624
- Sack RB (1975) Human diarrheal disease caused by enterotoxigenic *Escherichia coli*. *Annu Rev Microbiol* 29:333–353
- Schubert S, Rakin A, Karch H, Carniel E, Heesemann J (1998) Prevalence of the “high-pathogenicity island” of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infect Immun* 66:480–485
- Schubert S, Rakin A, Fischer D, Sorsa J, Heesemann J (1999) Characterization of the integration site of *Yersinia* high-pathogenicity island in *Escherichia coli*. *FEMS Microbiol Lett* 179:409–414
- Sears CL, Kaper JB (1996) Enteric bacterial toxins: mechanisms of action and linkage to intestinal secretion. *Microbiol Rev* 60:167–215

- Sperandio V, Kaper JB, Bortolini MR, Neves BC, Keller R, Trabulsi LR (1998) Characterization of the locus of enterocyte effacement (LEE) in different enteropathogenic *Escherichia coli* (EPEC) and Shiga-toxin producing *Escherichia coli* (STEC) serotypes. FEMS Microbiol Lett 164:133–139
- Sperandio V, Mellies JL, Nguyen W, Shin S, Kaper JB (1999) Quorum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic *Escherichia coli*. Proc Natl Acad Sci USA 96:15196–15201
- Sperandio V, Mellies JL, Delahay RM, Frankel G, Crawford JA, Nguyen W, Kaper JB (2000) Activation of enteropathogenic *Escherichia coli* (EPEC) *LEE2* and *LEE3* operons by Ler. Mol Microbiol 38:781–793
- Stein M, Kenny B, Stein MA, Finlay BB (1996) Characterization od EspC, a 110-kilodalton protein secreted by enteropathogenic *Escherichia coli* which is homologous to members of the immunoglobulin A protease-like family of secreted proteins. J Bacteriol 178:241–250
- Taylor KA, O'Connell CB, Luther PW, Donnenberg MS (1998) The EspB protein of enteropathogenic *Escherichia coli* is targeted to the cytoplasm of infected HeLa cells. Infect Immun 66:5501–5507
- Tobe T, Schoolnik GK, Sohel I, Bustamante VH, Puente JL (1996) Cloning and characterization of *bfpTVW*, genes required for the transcriptional activation of *bfpA* in enteropathogenic *Escherichia coli*. Mol Microbiol 21:963–975
- Torres AG, Giron JA, Perna NT, Burland V, Blattner FR, Kaper JB (2001) Identification and characterization of *lpfABCC'DE*, a fimbrial operon of enterohemorrhagic *Escherichia coli* O157:H7. (in preparation)
- Vokes SA, Reeves SA, Torres AG, Payne SM (1999) The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island. Mol Microbiol 33:63–73
- Wachter C, Beinke C, Mattes M, Schmidt MA (1999) Insertion of EspD into epithelial target cell membranes by infecting enteropathogenic *Escherichia coli*. Mol Microbiol 31:1695–1707
- Wainwright LA, Kaper JB (1998) EspB and EspD require a specific chaperone for proper secretion from enteropathogenic *Escherichia coli*. Mol Microbiol 27:1247–1260
- Warawa J, Finlay BB, Kenny B (1999) Type III secretion-dependent hemolytic activity of enteropathogenic *Escherichia coli*. Infect Immun 67:5538–5540
- Wieler LH, McDaniel TK, Whittam TS, Kaper JB (1997) Insertion site of the locus of enterocyte effacement in enteropathogenic and enterohemorrhagic *Escherichia coli* differs in relation to the clonal phylogeny of the strains. FEMS Microbiol Lett 156:49–53
- Wolff C, Nisan I, Hanski E, Frankel G, Rosenshine I (1998) Protein translocation into host epithelial cells by infecting enteropathogenic *Escherichia coli*. Mol Microbiol 28:143–155

# Pathogenicity Islands of *Shigella*

M. INGERSOLL<sup>1</sup>, E.A. GROISMAN<sup>2</sup>, and A. ZYCHLINSKY<sup>3</sup>

1	Introduction . . . . .	50
1.1	<i>Shigella</i> and Dysentery . . . . .	50
1.2	Genetic Determinants of Virulence . . . . .	50
1.3	<i>Shigella</i> vs. <i>Escherichia coli</i> . . . . .	51
1.4	Pathogenicity Islands of <i>Shigella</i> . . . . .	51
2	Serotype Converting Locus; Shi-O . . . . .	52
2.1	Background . . . . .	52
2.2	Organization of Shi-O . . . . .	53
2.3	Gene Products of Shi-O . . . . .	53
2.4	Summary of Shi-O . . . . .	54
3	Shi-1 . . . . .	54
3.1	Background . . . . .	54
3.2	Identification of Shi-1 . . . . .	55
3.3	Organization and Characterization of Shi-1 . . . . .	55
3.4	Gene Products of Shi-1 . . . . .	55
3.4.1	The Autotransporters SigA and Pic . . . . .	56
3.4.2	The Enterotoxin ShET1 . . . . .	57
3.5	Summary of Shi-1 . . . . .	57
4	Shi-2 . . . . .	57
4.1	Background . . . . .	57
4.2	Identification of Shi-2 . . . . .	58
4.3	Organization and Characterization of Shi-2 . . . . .	58
4.4	Gene Products of Shi-2 . . . . .	59
4.4.1	The Aerobactin Operon . . . . .	59
4.4.2	Colicin Immunity . . . . .	60
4.4.3	Novel ORFs . . . . .	60
4.5	Summary of Shi-2 . . . . .	60
5	“Black Holes” . . . . .	61
6	Concluding Remarks . . . . .	61
	References . . . . .	62

<sup>1</sup> The Skirball Institute and Department of Microbiology, New York University Medical Center, 540 First Avenue, New York, NY 10016, USA

<sup>2</sup> Howard Hughes Medical Institute, Department of Molecular Microbiology, Washington University School of Medicine, 660 Euclid Ave, St. Louis, MO 63110, USA

<sup>3</sup> Max Planck Institute for Infection Biology, Schumannstraße 21/22, Campus Charité Mitte, 10117 Berlin, Germany

## 1 Introduction

### 1.1 *Shigella* and Dysentery

*Shigella* species are the causative agents of bacillary dysentery. Signs of disease range from mild diarrhea to a severe form of disease including fever, abdominal cramps, and stools containing blood, pus and mucus. *Shigella* are primarily human pathogens but can produce disease symptoms in other primates (SANSONETTI 1992).

After ingestion, shigellae traverse the intestinal epithelial barrier through specialized cells, called M-cells, at the level of the colon (WASSEF et al. 1989). These cells transport antigens, including enteric pathogens, across the epithelium. Following transcytosis, micro-organisms gain access to lymphoid follicles containing resident tissue macrophages (JARRY et al. 1989; SOESTAYO et al. 1990). After phagocytosis, shigellae rapidly destroy the membrane of the phagosome and are liberated into the host cell cytoplasm (FINLAY and FALKOW 1988; MAURELLI and SANSONETTI 1988).

In the cytoplasm of the host cell, the bacteria secrete IpaB, a protein that binds to caspase-1 (interleukin-1 $\beta$  converting enzyme). The IpaB-dependent activation of caspase-1 induces macrophage apoptosis and the release of large amounts of active IL-1 $\beta$  and IL-18 (HILBI et al. 1997, 1998). Release of these cytokines from the dying macrophages is the primary cause of the acute colonic inflammation characteristic of shigellosis (SANSONETTI et al. 2000). The inflammatory response damages the colonic mucosa and exacerbates the infection (PERDOMO et al. 1994).

There are four groups of *Shigella*, *S. flexneri*, *S. sonnei*, *S. dysenteriae*, and *S. boydii*, which differ in their epidemiology. *S. dysenteriae* is associated primarily with epidemics. *S. flexneri* is the predominant group found in areas of endemic infection, while *S. sonnei* is the group implicated in source outbreaks in developed countries (HALE 1991). *S. boydii* is not frequently isolated but may also be associated with epidemics. *S. flexneri* has traditionally been, as is reflected in this review, the focus of molecular-genetic research.

### 1.2 Genetic Determinants of Virulence

*Shigella* invasiveness and cytotoxicity genes are encoded on a large virulence plasmid. Plasmid-cured strains are no longer pathogenic (SANSONETTI et al. 1982). The plasmid encodes the *invasion plasmid antigens (ipa)* operon which includes IpaB, C, and D, three proteins essential for invasion, vacuolar escape, induction of macrophage apoptosis and pathogenicity in animal models (HIGH et al. 1992; MÉNARD et al. 1993; SASAKAWA et al. 1988). IpaB, C, and D are secreted by a type III secretion apparatus encoded by the *mxi* and *spa* operons that are also present on the virulence plasmid (ALLAOUI et al. 1992, 1993; ANDREWS et al. 1991). IpaB and IpaC form a complex (MÉNARD et al. 1994) that is sufficient for the bacteria to invade epithelial cells (MÉNARD et al. 1996).

In addition to the genes encoded on the large virulence plasmid, other chromosomally located genes are required for the full array of virulence phenotypes caused by *Shigella*. The multi-operon plasmid locus, encoding the type III system and its effectors (for review in BUCHRIESER et al. 2000; HUECK 1998), has been considered a pathogenicity island (PAI) by some (MECSAS and STRAUSS 1996). However, part of the defining tenet of a PAI is its localization to the bacterial chromosome (HACKER et al. 1997). For this reason, the *Shigella* large virulence plasmid will not be discussed here.

### 1.3 *Shigella* vs. *Escherichia coli*

*Shigella* and *Escherichia coli* are genetically very closely related and are identified as individual species mainly because of clinical considerations. Genetically, *E. coli* and *Shigella* differ because of (a) the presence of the virulence plasmid in *Shigella* and (b) the distribution and variety of distinct PAIs between the two species. Pathogenic *E. coli* strains contain PAIs that confer discrete phenotypes from *Shigella*, such as the locus of enterocyte effacement PAI, LEE, of enteropathogenic *E. coli* (EPEC) and enterohemorrhagic *E. coli* (EHEC) (NATARO and KAPER 1998).

Classic experiments demonstrated that chromosomal differences were responsible in part for the distinct virulence properties of *Shigella* and *E. coli*. Conjugative transfer of three separate loci from a nonpathogenic strain of *E. coli* attenuated *Shigella* virulence (FALKOW et al. 1963; FORMAL et al. 1965, 1971). Moreover, transfer of both the equivalent *Shigella* chromosomal loci and the large virulence plasmid was sufficient to confer virulence upon nonpathogenic *E. coli* (SANSONETTI et al. 1983). These experiments indicate that *Shigella* harbors species-specific sequences or alleles of genes which are required for virulence, and/or that it is missing certain *E. coli* genes which have the capacity to attenuate virulence (MAURELLI et al. 1998; SAKELLARIS et al. 2000).

### 1.4 Pathogenicity Islands of *Shigella*

To date, three PAIs have been identified in *Shigella flexneri*. Genes within these islands encode LPS modification enzymes (ALLISON and VERMA 2000), enterotoxins, proteases (FASANO et al. 1995; RAJAKUMAR et al. 1997), genes required for iron acquisition, and colicin resistance genes (MOSS et al. 1999; VOKES et al. 1999).

*Shigella* PAIs were named by different investigators without the development of a common nomenclature. Upon identification of a PAI at the *selC* locus (MOSS et al. 1999; VOKES et al. 1999), together with Shelly Payne's group at UT Austin, we suggested the nomenclature Shi for *Shigella* pathogenicity island following the *Salmonella* pathogenicity island nomenclature, SPI. Here, for the sake of clarity, we will use a common nomenclature for *Shigella* PAIs, Shi-1 for the *she* PAI (RAJAKUMAR et al. 1997) and Shi-2 for the island at the *selC* locus (MOSS et al. 1999; VOKES et al. 1999). The locus harboring the genes involved in serotype

conversion will be referred to as Shi-O (ADHIKARI et al. 1999), to reflect the role of this island in converting O antigen and the fact that it was described before both Shi-1 and Shi-2.

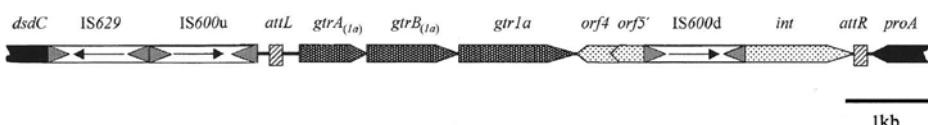
## 2 Serotype Converting Locus; Shi-O

### 2.1 Background

LPS is an important virulence factor in *Shigella* (LINDBERG et al. 1991). The humoral response to *Shigella* is specific to the O antigen of LPS (PHALIPON et al. 1995); therefore, immunity to a specific O antigen does not protect against other serotypes. Consequently, the capacity to change O antigens may be important for *Shigella* to avoid the humoral immune response.

There are 15 *Shigella flexneri* serotypes (NORIEGA et al. 1999), defined by differences in the basic O antigen structure. With the exception of serotype 6, all serotypes share the same common polysaccharide backbone (ALLISON and VERMA 2000). The differences in serotype arise through glucosylation and/or O acetylation of the basic O antigen (SIMMONS and ROMANOWSKA 1987). The basic O antigen, without any additions, is referred to as serotype Y. The enzymes that modify serotype Y, creating serotypes 2a, 3b, 5a, or X, are encoded in three sequential genes in lysogenic bacteriophage (GUAN et al. 1999; HUAN et al. 1997a; MAVRIS et al. 1997; VERMA et al. 1991, 1993; for review see ALLISON and VERMA 2000). Serotype 1a is a remarkable exception to the phage precedent. Its O antigen modification proteins are encoded by three genes similar to the phage genes found in serotypes 2a, 3b, 5a, and X but are localized to a PAI (Fig. 1), Shi-O.

In their review of serotype-converting bacteriophages in *S. flexneri*, ALLISON and VERMA suggest a common nomenclature for these three genes (Fig. 1): *gtrA<sub>(type)</sub>*, *gtrB<sub>(type)</sub>*, and *gtr<sub>(type)</sub>* (ALLISON and VERMA 2000). This nomenclature will be used in this review. The function of the first gene product, *gtrA<sub>(type)</sub>*, is unknown (ALLISON and VERMA 2000); the second gene, *gtrB<sub>(type)</sub>*, encodes a bactoprenol glucosyltransferase (GUAN et al. 1999) and the third gene, *gtr<sub>(type)</sub>*, encodes a glucosyltransferase (ALLISON and VERMA 2000).



**Fig. 1.** The serotype converting locus of *Shigella flexneri* 1a, Shi-O. The □ denotes insertion sequences and the arrows indicate their orientation, the ■ denotes genes involved in serotype conversion, the ▨ denotes other phage genes, the hatched rectangles denote the attachment sequences (*att*) and the solid black denotes bacterial genes with *E. coli* homologs. (Modified from ADHIKARI et al. 1999)

## 2.2 Organization of Shi-O

Shi-O appears to be an integrated bacteriophage that lost the ability to excise from the chromosome (ADHIKARI et al. 1999). Most of the phage's genome is missing from the insertion site. The two attachment sites (*attL* and *attR*) flanking the genes are identical to bacteriophage *attP* sites but are unusually spaced about 6.5kb apart and define the boundaries of the genes in this region. Normally, upon reversible phage integration into the chromosome, the separation between the two attachment sites by the phage genome is about 40kb.

Shi-O (Fig. 1) contains three ORFs, *gtrA<sub>(1a)</sub>*, *gtrB<sub>(1a)</sub>*, and *gtr1a*, transcribed in the same direction with a putative promoter sequence upstream of *gtrA<sub>(1a)</sub>*, ribosomal binding sites upstream of each ORF, and a rho-independent transcriptional terminator downstream of *gtr1a* (ADHIKARI et al. 1999). This conformation suggests that the genes are organized in an operon. Additionally, there are two ORFs encoded on the complementary strand. The first, *orf4*, is nearly identical to *orf3* of SfV, the bacteriophage of serotype 5a, which is believed to be involved in phage tail fiber assembly. The second, *orf5'*, is disrupted by the insertion sequence, IS600, but resembles *orf2* of SfV. The three *gtr* genes, *orf4*, and *orf5'* are flanked by IS600 both upstream and downstream and by IS629 upstream only. The IS600 elements are highly homologous to other IS600 elements from *S. flexneri*, *S. sonnei*, and *S. dysenteriae*. The IS629 bears significant homology to the *S. sonnei* IS629. A phage integrase gene (*int*), located downstream of *gtr1a*, is highly similar to other integrase genes present in SfV and in SfII and is interrupted by the downstream IS600 (ADHIKARI et al. 1999).

Upstream, 3' to the tandem insertion sequences, is a homolog of the *E. coli* gene *dsdC*, the D-serine dehydratase transcriptional activator. Downstream, 5' to the *attR*, the region displays homology to the *thrWproA* tRNA locus of the *E. coli* chromosome. The *dsdC* gene and the *thrWproA* locus do not map closely to each other in *E. coli*. ADHIKARI et al. speculate that the insertion sequences in this PAI may have facilitated a chromosomal rearrangement, or that this *Shigella* chromosome locus is organized differently than the comparable locus in *E. coli* (ADHIKARI et al. 1999).

The overall G+C content of the Shi-O ORFs is approximately 40%, lower than the typical 49%–53% described for the *Shigella* chromosome. The insertion sequences are about 40% G+C and the three serotype converting ORFs are 42.7%, 42%, and 33.4% for *gtrA<sub>(1a)</sub>*, *gtrB<sub>(1a)</sub>*, and *gtr1a*, respectively (ADHIKARI et al. 1999), suggesting that the locus was acquired through horizontal transfer.

## 2.3 Gene Products of Shi-O

Generally, the *gtrA<sub>(type)</sub>* and *gtrB<sub>(type)</sub>* genes of bacteriophages are highly conserved and interchangeable among serotype converting phages. The function of *gtrA<sub>(type)</sub>* is currently unknown, but it may be the flipase for the UndP-glucose

precursor (GUAN et al. 1999). It is predicted to encode an integral membrane protein with four transmembrane domains (ALLISON and VERMA 2000). The *gtrB*<sub>(type)</sub> gene encodes a bactoprenol glucosyltransferase, which catalyzes the transfer of glucose to bactoprenol phosphate from UDP-glucose (ALLISON and VERMA 2000; GUAN et al. 1999). The third gene in bacteriophages is unique in each case, and there are no significant homologies to other *gtr*(*type*) genes or to other sequences in the database. Each *gtr*(*type*) however, encodes a glucosyltransferase gene. The activity of the three gene products is required for the modification of the O antigen (ALLISON and VERMA 2000).

Adhikari et al. found that the GtrA<sub>(1a)</sub> and GtrB<sub>(1a)</sub> in the *S. flexneri* 1a Shi-O PAI were 88%–99% identical to serotype converting proteins found in *S. flexneri* bacteriophages (ADHIKARI et al. 1999), specifically SfII (MAVRIS et al. 1997), SfV (HUAN et al. 1997b), and SfX (GUAN et al. 1999). As expected, Gtr1a had no significant homology to any database sequences. Functional analysis of the region revealed that transformation of the three genes into *S. flexneri* serotype Y converts the bacteria to serotype 1a (ADHIKARI et al. 1999). However, the virulence properties of the converted strain remain to be tested.

## 2.4 Summary of Shi-O

This unique region of the chromosome is approximately 10.6kb and forms a distinct genetic unit, defined by the insertion sequences and the *attL* and *attR* sequences. The organization of the serotype converting genes as well as the location of *orf4*, *orf5'*, the *attP* sites, and the integrase indicate that this PAI is derived from an ancestral bacteriophage. The serotype converting locus contains multiple cryptic mobility elements including insertion sequences and an integrase gene, as well as having a different G + C content from *Shigella*'s overall 49%–53%. The deletion of the majority of the phage was probably accomplished through recombination events between the insertion sequences which contain repeated sequences (ADHIKARI et al. 1999). This locus is sufficient for serotype conversion (ADHIKARI et al. 1999) and therefore an important virulence determinant.

## 3 Shi-1

### 3.1 Background

A second pathogenicity island, characterized in *Shigella* serotype 2a, encodes *sigA*, *pic* (formerly known as *she*), *set1A*, and *set1B*, as well as two novel ORFs. The PAI was named *she* because it contains the *she* gene (RAJAKUMAR et al. 1997). As mentioned previously, the Shi-1 nomenclature will be used exclusively to avoid confusion.

### 3.2 Identification of Shi-1

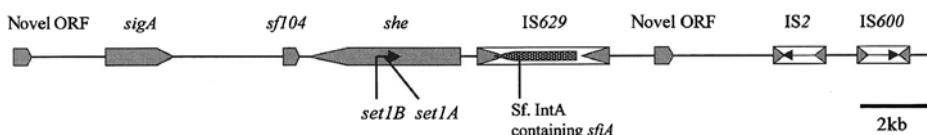
Identification of the Shi-1 PAI was accomplished through an approach termed island probing, a method that capitalizes on the genetic instability of some PAIs. A gene suspected of being part of a PAI is disrupted using an antibiotic resistance gene, and the incidence of loss of antibiotic resistance can be measured (RAJAKUMAR et al. 1997). The region was cloned and partially sequenced to reveal a unique (not homologous to *E. coli*) chromosomal region of *Shigella* (RAJAKUMAR et al. 1997) that has the characteristics of a pathogenicity island. The PAI spontaneously deletes from the chromosome at a frequency of  $10^{-5}$  to  $10^{-6}$  per cell generation under laboratory conditions (RAJAKUMAR et al. 1997). The spontaneous deletion is 51kb. The size of Shi-1 is not certain, however, because the whole locus has not been sequenced and the PAI flanking regions are not known (RAJAKUMAR et al. 1997).

### 3.3 Organization and Characterization of Shi-1

The elements of this locus, insertion sequences, multiple ORFs, and several cryptic mobility elements (RAJAKUMAR et al. 1997) define it as a PAI. As illustrated in Fig. 2, for the 25-kb region published thus far, Shi-1 encodes a novel ORF, *sigA*, *sf104*, *pic*, *set1A* and *set1B*, *sfiA* encoded within bacterial intron sequence Sf. IntA, a partial IS629, another novel ORF, and two complete insertion sequences, IS2 and IS600. The G + C contents of the genes in Shi-1 are: *sigA* 44.9%, *pic*, *set1A*, and *set1B* 49.1% and the two novel ORFs 55.7% and 53.7%. *sf104* (*Shigella flexneri* homolog of *Agrobacterium tumefaciens* *orf104*) is 68% similar to hypothetical proteins from *Rhizobium* and *Agrobacterium*. All of the insertion sequences bear similarity to published insertion sequences. IS629 is incomplete but has, within its 3' end, a sequence that is 99% homologous to group II intron-like sequences found in bacteria. Within the intron is an ORF, *sfiA*, potentially encoding a reverse transcriptase-like protein (RAJAKUMAR et al. 1997).

### 3.4 Gene Products of Shi-1

Shi-1 contains several genes that encode virulence factors. These include two putative autotransporters, SigA and Pic (formerly known as ShMu, the protein



**Fig. 2.** The Shi-1 PAI of *Shigella flexneri*. The □ denotes insertion sequences and the arrows indicate their orientation. The ■ denotes genes that have been sequenced. The ▨ inside of the IS629 is the bacterial intron containing the putative reverse transcriptase. The black line denotes regions that have not been sequenced. (Modified from RAJAKUMAR et al. 1997)

product of *she*), as well as ShET1, an enterotoxin of the CT-/LT-like toxin family.

### 3.4.1 The Autotransporters SigA and Pic

SigA (*Shigella IgA*-like protease homolog) and Pic (protein involved in intestinal colonization) are secreted through an autotransporter secretion system (AL-HASANI et al. 2000; HENDERSON et al. 1999). Autotransporter proteins encode an unusually long signal sequence, followed by a signal peptidase cleavage site, which enables them to utilize the Sec-dependent secretion pathway to traverse the inner membrane. From the periplasm, it is thought that the carboxyl-terminal of the protein inserts into the outer membrane forming a β-barrel pore with 14 transmembrane passes through which the rest of the protein is secreted. At this point, the protein may remain attached to the outer membrane or it may be cleaved, at a second cleavage site, into the environment (HENDERSON et al. 1998). SigA and Pic both contain a serine protease active site motif and are therefore classified as members of the subfamily of autotransporters called SPATEs, or serine protease autotransporters of Enterobacteriaceae.

SigA is a 139.6-kD protein with sequence similarity, especially in the carboxyl-terminal, to a number of bacterial IgA1 protease-like autotransporters (AL-HASANI et al. 2000). SigA contains a second cleavage signal indicating that it is a secreted protein. SigA, but not a mutant form missing the carboxyl-terminal domain, is secreted into the extracellular medium as a protein of approximately 103kD when expressed in *E. coli* (AL-HASANI et al. 2000). Secretion of SigA from *Shigella* is temperature regulated, like other SPATEs. The protease activity of SigA was demonstrated with a casein-based fluorogenic substrate and, in fact, is the sole protein in bacterial supernatants able to degrade casein. In the rabbit ileal loop model of infection, a *sigA* mutant produces 30% less fluid accumulation than the parent strain, suggesting that SigA has a potential role in pathogenesis.

Pic is another autotransporter encoded within Shi-1. The *Shigella pic* gene product is predicted to be 146kD. The homologous protein in *E. coli*, with 99.7% identity, has been characterized and is also predicted to be 146kD. The secreted form is 109.8kD. Like SigA, Pic contains an amino terminal signal sequence with a signal peptidase cleavage site, a putative serine protease active site, a carboxyl-terminal predicted to span the membrane 10–14 times, and a second signal cleavage site, allowing the protein to be released from the surface of the bacteria. The recombinant protein has protease activity on gelatin and bovine and murine mucin. It also confers serum resistance in laboratory strains of *E. coli*. The serum resistance is sensitive to phenylmethylsulfonyl fluoride (PMSF), demonstrating that the serine protease activity is necessary for the phenotype. Pic exhibits some hemagglutinase activity, especially with rat red blood cells (HENDERSON et al. 1999).

It is interesting to speculate that the protease activities of Pic could be involved in enabling *Shigella* to burrow through the mucin layer associated with the gut, thereby helping to establish initial colonization (HENDERSON et al. 1999).

### 3.4.2 The Enterotoxin ShET1

Experimental infections in monkeys led investigators to speculate that *S. flexneri* encoded an enterotoxin (KINSEY et al. 1976). Two genes, *set1B* and *set1A*, encode this enterotoxin (FASANO et al. 1995). These two genes are tandemly arranged, separated by only 6bp. Interestingly, the two genes are oppositely oriented and embodied entirely within *pic*. *set1B* is 186bp and encodes a protein of 7kD and *set1A* is 534bp, encoding a 20-kD protein. Together, they make a holoenzyme of 55kD organized in an A<sub>1</sub>-B<sub>5</sub> conformation (FASANO et al. 1997) typical of many bacterial enterotoxins of the CT-/LT-like toxin family (for review of CT-/LT-like toxins, see SCHMITT et al. 1999). Bacterial supernatants containing the toxin cause an increase in fluid accumulation in the rabbit ileal loop model. The same supernatants induce total tissue conductance and short-circuit current changes to a greater extent than control supernatants that do not contain the enterotoxin in Ussing chamber experiments indicating toxin activity (FASANO et al. 1997). ShET1 is prevalent in *S. flexneri* serotype 2a and rarely seen in other *S. flexneri* serotypes (NORIEGA et al. 1995). The watery diarrhea observed in animal experiments infected with this serotype may be explained by the presence of this particular enterotoxin (FASANO et al. 1997).

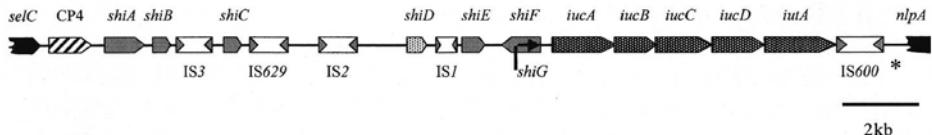
## 3.5 Summary of Shi-1

The identification of this chromosomal region as a PAI is still inconclusive since the putative borders and flanking regions have not been identified. The locus does fulfill several criteria for a PAI, however, including its potentially large size and the presence of cryptic mobility genes and ORFs with varying G + C contents. Furthermore, SigA, Pic, and ShET1 appear to be virulence factors (AL-HASANI et al. 2000; FASANO et al. 1997; HENDERSON et al. 1999). The activities of the ShET1 enterotoxin and SigA might help to explain the watery diarrhea that can precede the dysentery associated with *Shigella* infections (AL-HASANI et al. 2000; FASANO et al. 1995). Altogether, these experimental results and sequence data help establish the case for Shi-1 as a PAI, but only further sequencing can completely confirm the designation.

## 4 Shi-2

### 4.1 Background

The third PAI of *Shigella*, Shi-2, was discovered independently by two different labs in two different *Shigella* strains, SA100 (serotype 2a) (VOKES et al. 1999) and M90T (serotype 5a) (Moss et al. 1999). As illustrated in Fig. 3, Shi-2 encodes multiple cryptic mobility genes, an aerobactin operon, a colicin V immunity gene, and



**Fig. 3.** Shi-2 PAI of *Shigella flexneri* 5a. The ▨ denote the insertion sequences. The ▨ denote the aerobactin genes. The ▨ denotes the colicin V immunity gene. The ▨ denotes the integrase homolog. The ▨ denotes the unique genes and the black boxes denote the interrupted genes with *E. coli* homologs. The arrow indicates *shiG* as it is oriented opposite to *shiF*. The asterisk indicates where the serotype 2a Shi-2 PAI (VOKES et al. 1999) diverges in sequence from the serotype 5a. (Modified from Moss et al. 1999)

several novel ORFs. Shi-2 is inserted into the *selC* locus and, in both cases, is associated with a 3.8-kb deletion when compared with the *E. coli* genome (Moss et al. 1999; VOKES et al. 1999). Shi-2 from the two *Shigella* strains are identical in sequence except at the 3' end of the aerobactin operon. Although Shi-2 is present in all *S. flexneri* strains it is polymorphic; the aerobactin operon is always present in Shi-2, while other genes are present with different frequency (Moss et al. 1999; VOKES et al. 1999).

Interestingly, the experiments mentioned previously, using conjugal transfer of *Shigella* loci to *E. coli*, demonstrated that a large region in the *Shigella* chromosome that includes Shi-2 is required for the full inflammatory phenotype of *Shigella* (FALKOW et al. 1963; FORMAL et al. 1965, 1971; SANSONETTI et al. 1983).

#### 4.2 Identification of Shi-2

The identification of the Shi-2 locus was interesting because it demonstrated that an island can be independently discovered by two different methods. Shi-2 was located in *S. flexneri* serotype 2a in an effort to map the chromosomal locus of the aerobactin operon (VOKES et al. 1999). In *S. flexneri* serotype 5a, Shi-2 was discovered by probing the *selC* locus for large insertions, since *selC* is the site of multiple PAI insertions in other bacteria (BLANC-POTARD and GROISMAN 1997; Moss et al. 1999).

#### 4.3 Organization and Characterization of Shi-2

The integration of Shi-2 and other PAIs into the *selC* locus is very similar. A short region of Shi-2, 3' to *selC*, is virtually identical and, further downstream, well conserved among several PAIs integrated in the *selC* locus in other species. Other PAIs integrated at the *selC* locus include LEE, from both EHEC and EPEC, PAI-1 from UPEC, and SPI-3 from *Salmonella*. LEE, PAI-1, and Shi-2 integrate 16bp downstream from *selC*, while SPI-3 integrates 12bp downstream (Moss et al. 1999; VOKES et al. 1999). In *E. coli*, integration at the *selC* locus is associated with loss of chromosomal material (BLUM et al. 1994; McDANIEL et al. 1995; PERNA et al.

1998). The *Shigella selC* PAI integration seems to have caused the largest loss of material relative to *E. coli* (3.8kb) identified thus far (Moss et al. 1999). The sequence similarity between the beginning of Shi-2 and the *E. coli selC* associated PAIs ends shortly after the start of the island. Shi-2 and LEE possess an integrase with similarity to the CP4 family of phage integrases (PERNA et al. 1998). These similarities suggest a common mechanism of integration at the *selC* locus.

Additionally, Shi-2 contains multiple insertion sequences, including IS3, IS629, IS2, IS1 and IS600 in the 2a and 5a serotype, and a second IS2 at the 3' end of the 2a serotype PAI (Moss et al. 1999; VOKES et al. 1999). All of these insertion sequences exhibit extensive similarities to insertion sequences reported in the databases, but it appears that none of the insertion sequence transposase ORFs are intact. The insertion sequences of Shi-2 are interspersed between some of the novel ORFs and flank the aerobactin operon. The frequency and organization of these elements is unusual as compared with other PAIs characterized to date (Moss et al. 1999). This may indicate that Shi-2 was built in several pieces but is no longer able to mobilize itself (Moss et al. 1999).

As mentioned previously, the *S. flexneri* serotype 2a PAI contains an additional ORF (*orf35*) and insertion sequence (VOKES et al. 1999). *orf35* is present in strain M90T, as shown by PCR amplification (VOKES et al. 1999), but is not located at the same position in the chromosome (J. Moss, unpublished observation). Another difference between Shi-2 in serotypes 2a and 5a is that the first ORF in *Shigella* 5a encodes *shiA*, a 1043-bp gene that produces a 40-kD protein (Moss et al. 1999; J. Moss, unpublished observation), while in the 2a strain, the gene is disrupted by a stop codon. This results in a shorter sequence for this ORF and another ORF in the opposite orientation (VOKES et al. 1999). The significance of this difference is unclear.

## 4.4 Gene Products of Shi-2

The association of Shi-2 with virulence is clear for the aerobactin operon and genes coding for colicin V immunity. Whether any of the novel ORFs are important for *Shigella* virulence remains to be determined.

### 4.4.1 The Aerobactin Operon

The aerobactin operon is considered an important virulence locus because it encodes a siderophore that allows *Shigella* to acquire iron in the iron-limiting host environment (for review of siderophores see NEILANDS 1995). The aerobactin operon found in Shi-2 bears a strong resemblance to the *E. coli* pColV-K30 plasmid. The first four genes, *iucA-D* are 92%–97% homologous and the *iutA* gene is slightly less similar (85%) (Moss et al. 1999; VOKES et al. 1999). The proteins, IucA–D, form the siderophore which complexes with iron in the host environment. IutA is the bacterial receptor for the iron-siderophore complex. *Shigella* aerobactin mutants are dose-dependently attenuated in the rabbit ileal loop model (NASSIF

et al. 1987). Aerobactin deletion mutants were also found to be avirulent in the chicken embryo, but the deletions encompassed more than just the aerobactin genes and the phenotype was attributed to loss of adjacent genes (LAWLOR et al. 1987).

#### 4.4.2 Colicin Immunity

Colicins are bacterial toxins that bind and kill susceptible bacteria by forming pores in the cytoplasmic membrane, degrading DNA or inhibiting protein synthesis. Colicin-producing bacteria also make immunity proteins that confer resistance to the toxins for their own protection. Generally, the immunity genes and the colicin genes are linked (for review of colicins see RILEY and GORDON 1999). Shi-2 encodes a protein that confers protection to colicin V but no colicin-producing gene was identified in Shi-2 (Moss et al. 1999; VOKES et al. 1999). Serotype 2a does produce a colicin, called *S. flexneri* colicin, but it is not linked to the immunity gene (VOKES et al. 1999). The gene that confers colicin immunity to *colV*, *colIb*, and *S. flexneri* colicin (*orf21* or *imm* in *S. flexneri* 2A and *shiD* in *S. flexneri* 5) bears no resemblance to previously identified immunity genes (Moss et al. 1999; VOKES et al. 1999). Colicin immunity may play a role in survival of the bacteria in the intestinal lumen.

#### 4.4.3 Novel ORFs

The novel ORFs in Shi-2 bear little homology to any known genes and their functions are unknown. The 5a serotype PAI unique ORFs were designated *shiA-G* and this designation will be used here. Fitting the definition of pathogenicity islands, these novel ORFs exhibit a G + C content well below the rest of the *Shigella* chromosome (Moss et al. 1999). The structure prediction and the homologies to ShiA, albeit weak, suggest that it is similar to proteins in the large family of quinone reductases. ShiF resembles tetracycline-resistance transporters, while the additional ORF, *orf35*, found in the serotype 2a Shi-2 is homologous to an SFMD precursor protein (outer membrane usher protein associated with type 1 fimbriae) (VOKES et al. 1999).

### 4.5 Summary of Shi-2

The Shi-2 PAI is an interesting locus because of the unusual number and arrangement of cryptic elements, suggesting an elaborate history of horizontal acquisitions of genetic material. Its integration at the *selC* locus is also remarkable because of the number of previously characterized *selC* PAIs, implying that there may be some order to PAI creation and integration. The presence of several unknown ORFs is intriguing because this region of the *Shigella* chromosome is associated with the induction of inflammation. It is interesting to speculate that these unknown genes might encode proteins involved in host inflammation reactions. The functional characterization of this locus has yet to be completed.

## 5 “Black Holes”

Interestingly, a deletion of *cadA* was found in four representative *Shigella* strains and one enteroinvasive *E. coli* (EIEC) strain when compared with the published *E. coli* genome (MAURELLI et al. 1998). *cadA* encodes lysine decarboxylase, or LDC. The gene is present in over 80% of *E. coli* strains, with the exception of EIEC. Introduction of *cadA* into *Shigella flexneri* 2a had no effect on the ability of the strain to invade and colonize tissue, but it did decrease the conductance in Ussing chamber experiments, indicating that the gene may affect the activity of the two enterotoxins in *S. flexneri* 2a (MAURELLI et al. 1998). Cadaverine, the end-product of the reaction catalyzed by LDC, inhibited the effects of the enterotoxins in Ussing chambers as well, suggesting that LDC products protected the host cells rather than inhibiting the bacteria (MAURELLI et al. 1998). It is interesting to speculate whether these deletions serve to fine-tune bacterial virulence and are complementary to the addition of large genetic units, such as PAIs (MAURELLI et al. 1998).

Another example of the selective alteration of a genetic material is the curli locus. The curli, or *csg*, locus encodes a specific thin, aggregative type of fimbriae or curli involved in formation of biofilms in *E. coli* (VIDAL et al. 1998) or host cell attachment in *Salmonella* (SUKUPOLVI et al. 1997). In 60 strains that included all four groups of *Shigella* and in 11 EIEC strains tested, the *csg* locus showed either insertions, deletions, or mutations. This suggests that this interruption is necessary for virulence, although this hypothesis has not been tested experimentally.

## 6 Concluding Remarks

Currently, there are three PAIs described in *Shigella flexneri* strains, the serotype-converting locus, Shi-O, and the multigenic Shi-1 and Shi-2. These PAIs await a more thorough characterization. The Shi-O serotype-converting genes are known, but their mechanism of action is still not understood. The Shi-1 locus remains to be fully sequenced and mapped. Shi-2 contains seven unique genes, whose products and functions have yet to be elucidated.

*Shigella* strains might harbor more PAIs. Identification of a 99-kb region that spontaneously deletes in serotype 2a may be the next *Shigella* PAI described (RAJAKUMAR et al. 1996). Sequencing of the genome of *Shigella* strain serotype 2a is under way and will determine whether there are other PAIs in this strain. Because of the polymorphism of the *Shigella* chromosome, only comparison of the genomes of several strains will give the complete picture (LAN and REEVES 2000). An important challenge in the understanding of *Shigella* pathogenesis will be to assign virulence functions to genes in the PAIs. Currently, most of the *Shigella* in vitro tests assay for plasmid associated virulence phenotypes, such as invasion and

cytotoxicity. These tests will not be useful in testing the function of proteins encoded by *Shigella*'s islands which are associated with serotype conversion, protease activity, enterotoxicity, inflammation, and iron acquisition. The investigation of *Shigella* PAI functions is complicated by the lack of a small-animal diarrhea model where more of the PAI functions could be tested. The Sereny test in mice and the mouse lung-infection model can test only for acute inflammation. The rabbit ileal-loop model is also useful for investigating inflammation and fluid accumulation. In order to study chromosomally associated virulence, new tests need to be developed or existing tests need to be modified.

*Shigella* is currently a large threat, especially in developing countries with limited access to health care. *Shigella* PAIs are an important part of the virulence phenotype. Understanding the role of PAIs in *Shigella* virulence will ultimately result in both a better understanding of the disease process and the development of treatments and vaccines.

## References

- Adhikari P, Allison G, Whittle B, Verma NK (1999) Serotype 1a O-antigen modification: molecular characterization of the genes involved and their novel organization in the *Shigella flexneri* chromosome. *J Bacteriol* 181:4711–4718
- Al-Hasani K, Henderson IR, Sakellaris H, Rajakumar K, Grant T, Nataro JP, Robins-Browne R, Adler B (2000) The sigA gene which is borne on the she pathogenicity island of *Shigella flexneri* 2a encodes an exported cytopathic protease involved in intestinal fluid accumulation. *Infect Immun* 68:2457–2463
- Allaoui A, Sansonetti PJ, Parsot C (1992) MxiJ, a lipoprotein involved in secretion of *Shigella* Ipa invasins, is homologous to YscJ, a secretion factor of the *Yersinia* Yop proteins. *J Bacteriol* 174:7661–7669
- Allaoui A, Sansonetti PJ, Parsot C (1993) MxiD: an outer membrane protein necessary for the secretion of the *Shigella flexneri* Ipa invasins. *Mol Microbiol* 7:59–68
- Allison GE, Verma NK (2000) Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends Microbiol* 8:17–23
- Andrews GP, Hromockyj AE, Coker C, Maurelli AT (1991) Two novel virulence loci, mxiA and mxiB, in *Shigella flexneri* 2A facilitate excretion of invasion plasmid antigens. *Infect Immun* 59:1997–2005
- Blanc-Potard A, Groisman E (1997) The *Salmonella* selC locus contains a pathogenicity island mediating intramacrophage survival. *EMBO J* 16:5376–5385
- Blum G, Ott M, Lischewski A, Ritter A, Imrich H, Tschaep H, Hacker J (1994) Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect Immun* 62:606–614
- Buchrieser C, Glaser P, Rusniok C, Nedjari H, d'Hauteville H, Kunst F, Sansonetti PJ, Parsot C (2000) The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol Microbiol* 38:1–14
- Falkow S, Schneider H, Baron LS, Formal SB (1963) Virulence of *Escherichia-Shigella* genetic hybrids for the guinea pig. *J Bacteriol* 86:1251–1258
- Fasano A, Noriega FR, Maneval DR Jr, Chanasongram S, Russell R, Guandalini S, Levine MM (1995) *Shigella* enterotoxin 1: an enterotoxin of *Shigella flexneri* 2a active in rabbit small intestine in vivo and in vitro. *J Clin Invest* 95:2853–2861
- Fasano A, Noriega FR, Liao FM, Wang W, Levine MM (1997) Effect of *shigella* enterotoxin 1 (ShET1) on rabbit intestine in vitro and in vivo. *Gut* 40:505–511
- Finlay BB, Falkow S (1988) Comparison of the invasion strategies used by *Salmonella cholera-suis*, *Shigella flexneri* and *Yersinia enterocolitica* to enter cultured animal cells: endosome acidification is not required for bacterial invasion or intracellular replication. *Biochemie* 70:1089–1099

- Formal SB, LaBrec EH, Kent TH, Falkow S (1965) Abortive intestinal infection with an *Escherichia coli*-*Shigella flexneri* hybrid strain. *J Bacteriol* 89:1374–1382
- Formal SB, Gemski P Jr, Baron LS, LaBrec EH (1971) A chromosomal locus which controls the ability of *Shigella flexneri* to evoke keratoconjunctivitis. *Infect Immun* 3:73–79
- Guan S, Bastin DA, Verma NK (1999) Functional analysis of the O antigen glucosylation gene cluster of *Shigella flexneri* bacteriophage SfX. *Microbiology* 145:1263–1273
- Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H (1997) Pathogenicity islands of virulent bacteria: structure, function, and impact on microbial evolution. *Mol Microbiol* 23:1089–1097
- Hale TL (1991) Genetic basis of virulence in *Shigella* species. *Microbiol Rev* 55:206–224
- Henderson IR, Navarro-Garcia F, Nataro JP (1998) The great escape: structure and function of the autotransporter proteins. *Trends Microbiol* 6:370–378
- Henderson IR, Czeczułka J, Eslava C, Noriega F, Nataro JP (1999) Characterization of *pic*, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*. *Infect Immun* 67:5587–5596
- High N, Mounier J, Prevost MC, Sansonetti PJ (1992) IpaB of *Shigella flexneri* causes entry into epithelial cells and escape from the phagocytic vacuole. *EMBO J* 11:1991–1999
- Hilbi H, Chen Y, Thirumalai K, Zychlinsky A (1997) The interleukin 1 $\beta$ -converting enzyme, caspase 1, is activated during *Shigella flexneri*-induced apoptosis in human monocyte-derived macrophages. *Infect Immun* 65:5165–5170
- Hilbi H, Moss JE, Hersh D, Chen Y, Arondel J, Banerjee S, Flavell RA, Yuan J, Sansonetti PJ, Zychlinsky A (1998) *Shigella*-induced apoptosis is dependent on caspase-1 which binds to IpaB. *J Biol Chem* 273:32895–32900
- Huan PT, Bastin DA, Whittle BL, Lindberg AA, Verma NK (1997a) Molecular characterization of the genes involved in O-antigen modification, attachment, integration and excision in *Shigella flexneri* bacteriophage SfV. *Gene* 195:217–227
- Huan PT, Whittle BL, Bastin DA, Lindberg AA, Verma NK (1997b) *Shigella flexneri* type-specific antigen V: cloning, sequencing and characterization of the glucosyl transferase gene of temperate bacteriophage SfV. *Gene* 195:207–216
- Hueck CJ (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* 62:379–433
- Jarry A, Robaszkiewicz M, Brousse N, Potet F (1989) Immune cells associated with M cells in the follicle-associated epithelium of Peyer's patches in the rat. *Cell Tissue Res* 255:293–298
- Kinsey MD, Formal SB, Dammin GJ, Giannella RA (1976) Fluid and electrolyte transport in rhesus monkeys challenged intracutaneously with *Shigella flexneri* 2a. *Infect Immun* 14:368–371
- Lan R, Reeves PR (2000) Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends Microbiol* 8:396–401
- Lawlor K, Daskaleros P, Robinson R, Payne S (1987) Virulence of iron transport mutants of *Shigella flexneri* and utilization of host iron compounds. *Infect Immun* 55:594–599
- Lindberg AA, Karnell A, Weintraub A (1991) The lipopolysaccharide of *Shigella* bacteria as a virulence factor. *Rev Infect Dis* 13 [Suppl 4]:S279–284
- Maurelli AT, Sansonetti PJ (1988) Genetic determinants of *Shigella* pathogenicity. *Ann Rev Microbiol* 42:127–150
- Maurelli AT, Fernandez RE, Bloch CA, Rode CK, Fasano A (1998) “Black holes” and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc Natl Acad Sci USA* 95:3943–3948
- Mavris M, Manning PA, Morona R (1997) Mechanism of bacteriophage SfII-mediated serotype conversion in *Shigella flexneri*. *Mol Microbiol* 26:939–950
- McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB (1995) A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci USA* 92:1664–1668
- Mecsas JJ, Strauss EJ (1996) Molecular mechanisms of bacterial virulence: type III secretion and pathogenicity islands. *Emerg Infect Dis* 2:270–288
- Ménard R, Sansonetti PJ, Parsot C (1993) Nonpolar mutagenesis of the *ipa* genes defines IpaB, IpaC, and IpaD as effectors of *Shigella flexneri* entry into epithelial cells. *J Bacteriol* 175:5899–5906
- Ménard R, Sansonetti PJ, Parsot C, Vasselon T (1994) Extracellular association and cytoplasmic partitioning of the IpaB and IpaC invasins of *S. flexneri*. *Cell* 79:515–525
- Ménard R, Prevost MC, Gounon P, Sansonetti PJ, Dehio C (1996) The secreted Ipa complex of *Shigella flexneri* promotes entry into mammalian cells. *Proc Natl Acad Sci USA* 93:1254–1258
- Moss JE, Cardozo TJ, Zychlinsky A, Groisman EA (1999) The *selC*-associated SHI-2 pathogenicity island of *Shigella flexneri*. *Mol Microbiol* 33:74–83

- Nassif X, Mazert MC, Mounier J, Sansonetti PJ (1987) Evaluation with an *iuc::Tn10* mutant of the role of aerobactin production in the virulence of *Shigella flexneri*. *Infect Immun* 55:1963–1969
- Nataro JP, Kaper JB (1998) Diarrheagenic *Escherichia coli* [published erratum appears in Clin Microbiol Rev 1998, 2:403]. *Clin Microbiol Rev* 11:42–201
- Neilands JB (1995) Siderophores: structure and function of microbial iron transport compounds. *J Biol Chem* 270:26723–26726
- Noriega FR, Liao FM, Formal SB, Fasano A, Levine MM (1995) Prevalence of *Shigella* enterotoxin 1 among *Shigella* clinical isolates of diverse serotypes. *J Infect Dis* 172:1408–1410
- Noriega FR, Liao FM, Maneval DR, Ren S, Formal SB, Levine MM (1999) Strategy for cross-protection among *Shigella flexneri* serotypes. *Infect Immun* 67:782–788
- Perdomo OJ, Cavaillon JM, Huerre M, Ohayon H, Gounon P, Sansonetti PJ (1994) Acute inflammation causes epithelial invasion and mucosal destruction in experimental shigellosis. *J Exp Med* 180:1307–1319
- Perna NT, Mayhew GF, Posfai G, Elliott S, Donnenberg MS, Kaper JB, Blattner FR (1998) Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* 66:3810–3817
- Phalipon A, Kaufmann M, Michetti P, Cavaillon J, Huerre M, Sansonetti P, Krahenbuhl J (1995) Monoclonal immunoglobulin A antibody directed against serotype-specific epitope of *Shigella flexneri* lipopolysaccharide protects against murine experimental shigellosis. *J Exp Med* 182:769–778
- Rajakumar K, Sasakawa C, Adler B (1996) A spontaneous 99-kb chromosomal deletion results in multi-antibiotic susceptibility and an attenuation of contact haemolysis in *Shigella flexneri* 2a. *J Med Microbiol* 45:64–75
- Rajakumar K, Sasakawa C, Adler B (1997) Use of a novel approach, termed island probing, identifies the *Shigella flexneri she* pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. *Infect and Immun* 65:4606–4614
- Riley MA, Gordon DM (1999) The ecological role of bacteriocins in bacterial competition. *Trends Microbiol* 7:129–133
- Sakellaris H, Hannink NK, Rajakumar K, Bulach D, Hunt M, Sasakawa C, Adler B (2000) Curli loci of *Shigella* spp. *Infect Immun* 68:3780–3783
- Sansonetti PJ (1992) Molecular and cellular biology of *Shigella flexneri* invasiveness: from cell assay systems to shigellosis. *Curr Top Microbiol Immunol* 180:1–19
- Sansonetti PJ, Hale TL, Dammin GJ, Kapfer C, Collins HH, Formal SB (1983) Alterations in the pathogenicity of *Escherichia coli* K-12 after transfer of plasmid and chromosomal genes from *Shigella flexneri*. *Infect Immun* 39:1392–1402
- Sansonetti PJ, Kopecko DJ, Formal SB (1982) Involvement of a plasmid in the invasive ability of *Shigella flexneri*. *Infect Immun* 35:852–860
- Sansonetti PJ, Phalipon A, Arondel J, Thirumalai K, Banerjee S, Akira S, Takeda K, Zychlinsky A (2000) Caspase-1 activation of IL-1 $\beta$  and IL-18 are essential for *Shigella flexneri*-induced inflammation. *Immunity* 12:581–590
- Sasakawa C, Kamata K, Sakai T, Makino S, Yamada M, Okada N, Yoshikawa M (1988) Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase plasmid in *Shigella flexneri* 2A. *J Bacteriol* 170:2480–2484
- Schmitt CK, Meysick KC, O'Brien AD (1999) Bacterial toxins: friends or foes? *Emerg Infect Dis* 5:224–234
- Simmons DA, Romanowska E (1987) Structure and biology of *Shigella flexneri* O antigens. *J Med Microbiol* 23:289–302
- Soestayo M, Biewenga J, Kraal G, Sminia T (1990) The localization of macrophages subsets and dendritic cells in the gastrointestinal tract of the mouse with special reference to the presence of high endothelial venules. An immuno- and enzyme-histochemical study. *Cell Tissue Res* 259:587–593
- Sukupolvi S, Lorenz RG, Gordon JI, Bian Z, Pfeifer JD, Normark SJ, Rhen M (1997) Expression of thin aggregative fimbriae promotes interaction of *Salmonella typhimurium* SR-11 with mouse small intestinal epithelial cells. *Infect Immun* 65:5320–5325
- Verma NK, Brandt JM, Verma DJ, Lindberg AA (1991) Molecular characterization of the O-acetyl transferase gene of converting bacteriophage SF6 that adds group antigen 6 to *Shigella flexneri*. *Mol Microbiol* 5:71–75
- Verma NK, Verma DJ, Huan PT, Lindberg AA (1993) Cloning and sequencing of the glucosyl transferase-encoding gene from converting bacteriophage X (SFX) of *Shigella flexneri*. *Gene* 129:99–101
- Vidal O, Longin R, Prigent-Combaret C, Dorel C, Hooremans M, Lejeune P (1998) Isolation of an *Escherichia coli* K-12 mutant strain able to form biofilms on inert surfaces: involvement of a new *ompR* allele that increases curli expression. *J Bacteriol* 180:2442–2449

- Vokes SA, Reeves SA, Torres AG, Payne SM (1999) The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island. Mol Microbiol 33:63–73
- Wassef JS, Keren DF, Mailloux JL (1989) Role of M cells in initial antigen uptake and in ulcer formation in rabbit intestinal loop model of shigellosis. Infect Immun 57:858–863

# Pathogenicity Islands and Host Adaptation of *Salmonella* Serovars

R.A. KINGSLEY and A.J. BAUMLER

1	Introduction . . . . .	67
2	Horizontally Acquired Virulence Determinants of Serovar <i>Typhimurium</i> . . . . .	68
2.1	Pathogenicity Islands . . . . .	68
2.1.1	<i>Salmonella</i> Pathogenicity Islands 1 Through 5 . . . . .	68
2.1.2	Potential Candidates . . . . .	71
2.2	Pathogenicity Islet, Phage, and Plasmid-Encoded Virulence Determinants . . . . .	73
2.2.1	Effector Proteins of SPI-1 and SPI-2 . . . . .	73
2.2.2	Adhesins . . . . .	74
2.2.3	Other Horizontally Acquired Virulence Determinants . . . . .	74
2.3	Does Size Really Matter? . . . . .	75
3	Horizontal Gene Transfer and Host Adaptation . . . . .	76
3.1	Phylogenetic Studies . . . . .	76
3.2	Experimental Approaches . . . . .	78
3.2.1	Host Adaptation and Host Range . . . . .	78
3.2.2	Comparison of Host-Restricted and Broad-Host-Range Serovars . . . . .	79
3.2.3	The Quest for Genes Expanding the Host Range of Host-Restricted Serovars . . . . .	80
3.2.4	Comparing the Repertoires of Genes One Serovar Uses for Infection of Two Different Host Species . . . . .	80
3.3	Epidemiological Studies . . . . .	82
4	Conclusion . . . . .	82
	References . . . . .	83

## 1 Introduction

The term ‘pathogenicity island’ was coined in reference to large (70–190kb), unstable genomic regions encoding hemolysin and fimbrial adhesins present in uropathogenic *Escherichia coli* strains but absent from nonpathogenic isolates, such as the K-12 strain (BLUM et al. 1994; HACKER et al. 1983, 1990; HUGHES et al. 1987; KNAPP et al. 1986; RITTER et al. 1995). The concept helped to explain, in genetic and evolutionary terms, why closely related strains of *E. coli* may differ substantially in their pathogenic potential. *E. coli* contains commensal organisms that are part of our normal

intestinal flora as well as a number of intestinal and extraintestinal pathogens (CAUGANT et al. 1983; NATARO and KAPER 1998; OCHMAN and SELANDER 1984). In general terms, pathogenicity islands are quintessentially large DNA regions conferring a virulence trait which is absent from a closely related, nonpathogenic, reference species or strain. Analysis of pathogenicity islands in a variety of animal and plant pathogens has revealed a number of common features (HACKER et al. 1997). Pathogenicity islands are often (a) large ( $>30\text{kb}$ ), (b) inserted in tRNA genes, (c) associated with inverted repeats, transposases, integrases, or plasmid origin of replication, and (d) have a G + C content that is atypical for the pathogen's genome. In practice, many or none of these may be features of a pathogenicity island.

The above considerations illustrate that an important aspect of defining pathogenicity islands is the selection of a nonpathogenic reference strain. Due to their relatively close relatedness, comparison of pathogenic and nonpathogenic *E. coli* strains has proven useful in identifying genes involved in the pathogenic lifestyle in this group of organisms. Definition of pathogenicity islands encoded by *Salmonella* serovars, on the other hand, is complicated by the fact that they all are considered pathogenic for man (ALEKSIC et al. 1996). For this reason horizontally acquired virulence determinants in the genome of *Salmonella* serovars are commonly identified by comparison with *E. coli* K-12, the closest nonpathogenic relative of the genus *Salmonella*. However, *E. coli* isolates are considerably closer phylogenetically to each other than they are to *Salmonella* serovars (OCHMAN and SELANDER 1984; SELANDER and LEVIN 1980). Consequently, a comparison between *E. coli* isolates is in some aspects different from comparing the *E. coli* K-12 genome with that of a *Salmonella* serovar. It has been estimated that since its divergence from the *Salmonella* lineage, 755 genes have been introduced into the *E. coli* K-12 genome in at least 234 lateral transfer events (LAWRENCE and OCHMAN 1998). The large amount of genetic material obtained horizontally since the divergence of *E. coli* from *Salmonella* serovars increases the probability that DNA regions identified in only one of the two organisms may indeed have been generated by more than one lateral transfer event. Consistent with this idea, some pathogenicity islands identified in the genus *Salmonella* by comparison with *E. coli* K-12 were later shown to be composites of DNA regions with different phylogenetic histories (BLANC-POTARD et al. 1999, HENSEL et al. 1999).

## 2 Horizontally Acquired Virulence Determinants of Serovar *Typhimurium*

### 2.1 Pathogenicity Islands

#### 2.1.1 *Salmonella* Pathogenicity Islands 1 Through 5

To date five horizontally acquired regions of the serovar *Typhimurium* genome have been designated SPIs and are distinguished by numeration based on chronological

order of their discovery (Fig. 1). SPI-1 and SPI-2 are located at centisomes 63 and 30, respectively, on the serovar *Typhimurium* chromosome (Fig. 2). Both of these pathogenicity islands have many of the classic features of pathogenicity islands (Fig. 1) (HENSEL et al. 1997; MILLS et al. 1995; OCHMAN et al. 1996). SPI-1 and SPI-2 exhibit considerably lower G+C content compared with the average for the serovar *Typhimurium* genome. SPI-2 is inserted at a tRNA gene (*valV*), and SPI-1 encodes an IS3 element in its vicinity. Both SPI-1 and SPI-2 encode a type III export apparatus which mediates injection of secreted effector proteins into the cytosol of host cells (for a review on type III secretion see HUECK 1998).

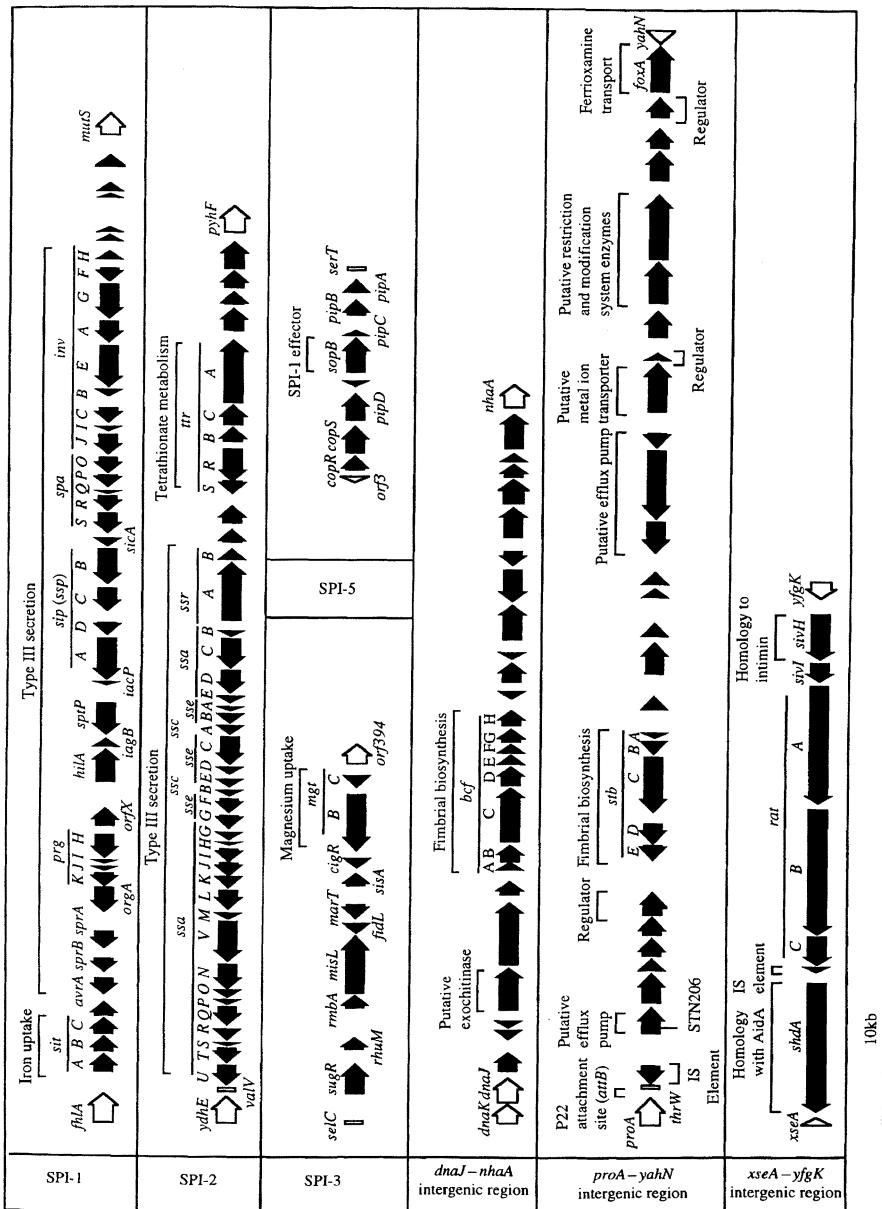
SPI-1 and SPI-2 are involved in disparate functions during pathogenesis. The SPI-1 encoded type III secretion system of serovar *Typhimurium* is required for invasion (GALÁN and CURTISS III 1989) and induction of cytokine production in human epithelial cell lines in vitro (HOBBIE et al. 1997). Furthermore, it is required for cytotoxicity observed at early time points after infection of murine macrophages in vitro (CHEN et al. 1996; MONACK et al. 1996). A functional SPI-1-encoded type III secretion system is essential for mortality, diarrhea, and histopathological lesions observed during oral serovar *Typhimurium* infections in calves (TSOLIS et al. 1999a). In contrast, mutations in SPI-1 cause only mild attenuation after oral inoculation of mice with serovar *Typhimurium* (GALÁN and CURTISS III 1989), a disease model in which diarrhea does not develop.

The SPI-2-encoded type III secretion system of serovar *Typhimurium* is required for survival in macrophages in vitro (OCHMAN et al. 1996), which is mediated by subverting vesicular trafficking (BEUZON et al. 2000; UCHIYA et al. 1999; VAZQUEZ-TORRES et al. 2000). A functional SPI-2-encoded type III secretion system is required for cytotoxicity observed at late time points in murine macrophages infected with serovar *Typhimurium* (VAN DER VELDEN et al. 2000). Furthermore, *Salmonella*-induced aggregation of host endosomal compartments in epithelial cells into tubules (lgp-tubules) requires an intact SPI-2-encoded type III secretion system (GUY et al. 2000). Mutational inactivation of SPI-2 results in avirulence of serovar *Typhimurium* during infection of mice (HENSEL et al. 1995). In contrast, inactivation of SPI-2 results in only modest reduction of serovar *Typhimurium* virulence for calves (TSOLIS et al. 1999a).

SPI-3 is an approximately 17-kb region with an atypically low G+C content which is inserted in the serovar *Typhimurium* chromosome at the *selC* tRNA locus (Fig. 1) (BLANC-POTARD and GROISMAN 1997). Two genes located on SPI-3, *mgtB* and *mgtC*, are required for high-affinity transport of magnesium, macrophage survival, and full virulence in mice. Mutational analysis of the remaining eight open reading frames on SPI-3 indicates that these are not required for full virulence in mice (BLANC-POTARD et al. 1999).

A 25-kb DNA region located at 92min on the serotype *Typhimurium* chromosome has been proposed to constitute a fourth pathogenicity island, SPI-4 (WONG et al. 1998). Currently, little is known about the role genes located on SPI-4 may play in *Salmonella* pathogenesis.

SPI-5 exhibits an atypical G+C content and is inserted adjacent to the tRNA gene *serT* at 20min on the serovar *Typhimurium* genetic map (Fig. 2). SPI-5 is an



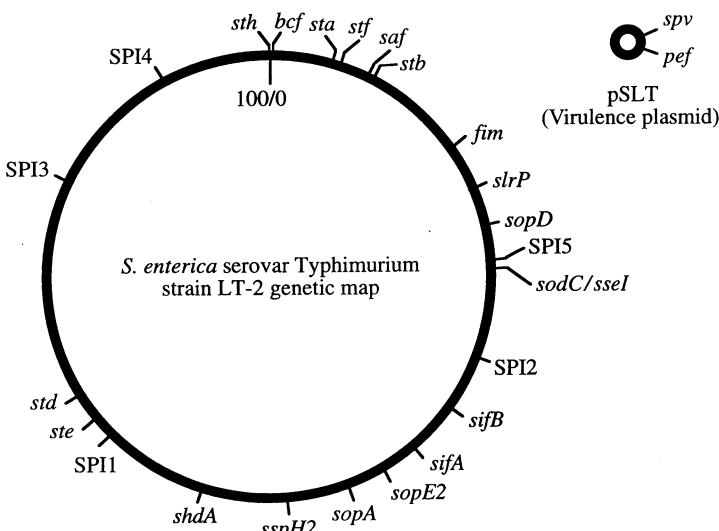
approximately 7-kb island encoding the effector protein SopB (SigD) which is secreted by the SPI-1-encoded type III secretion system (Fig. 1) (HONG and MILLER 1998; WOOD et al. 1998). SopB is an inositol phosphate phosphatase (NORRIS et al. 1998) which activates Akt, a serine-threonine kinase involved in the regulation of cell proliferation and survival of epithelial cells (STEELE-MORTIMER et al. 2000).

**Fig. 1.** Large genetic islands of *S. enterica* serovar *Typhimurium* carrying one or more virulence determinants. *Closed arrows* indicate open reading frames absent from the *E. coli* K-12 genome. *Open arrows* indicate open reading frames present in *E. coli* K-12. All islands are drawn to scale indicated by black bar. *Open boxes* indicate t-RNA genes. Known functions or putative functions indicated by sequence homology of open reading frames are indicated. Sizes and position of open reading frames are based on sequences determined by the Genome Sequencing Center, Washington University, St. Louis. (Genome Sequencing Center, personal communication) (McCLELLAND et al. 2000)

Mutations in *sopB* markedly reduce fluid accumulation and neutrophil influx elicited by *S. enterica* serovar *Dublin* in bovine ligated ileal loops (GALYOV et al. 1997).

### 2.1.2 Potential Candidates

The serovar *Typhimurium* genome contains a number of other DNA regions with one or several features of a pathogenicity island. For instance, the *fim* operon encoding type I fimbriae of serovar *Typhimurium* is encoded on a 31-kb DNA region absent from the *E. coli* genome. Furthermore, the insertion relative to the *E. coli* K-12 sequence is flanked on one side by a tRNA gene (Table 1) (Genome Sequencing Center, personal communication). The *saf* fimbrial operon is encoded on a large DNA region absent from the *E. coli* K-12 genome (FOLKESSON et al. 1999). This DNA region is flanked on one side by *aspV*, a tRNA gene (Table 1)



**Fig. 2.** Relative position on the *S. enterica* serovar *Typhimurium* LT-2 genetic map of virulence determinants which are absent from *E. coli* K-12. The large circle represents the chromosome, the small circle the virulence plasmid, pSLT. The positions of 0 and 100min are indicated. The *sspH1* and *sopE1* genes are absent from strain LT-2 but are present in other serovar *Typhimurium* isolates. (MIAO et al. 1999; MIROLD et al. 1999)

**Table 1.** Virulence determinants present in serovar *Typhimurium* but absent from *E. coli* K-12

Locus	Map location (min)	Insertion size <sup>b</sup> (kb)	Associated with	Function
SPI-1	63	45		SPI-1 type III secretion apparatus/ effector proteins/regulatory genes
SPI-5	25	7	<i>serT</i> (tRNA gene)	SPI-1 effector protein ( <i>sopB</i> )
<i>sopA</i>	44	7		SPI-1 effector protein
<i>sopD</i>	21	9		SPI-1 effector protein
<i>sopE1</i>	nd <sup>a</sup>	nd	Prophage ( <i>sopEΦ</i> )	SPI-1 effector protein
<i>sopE2</i>	42	18	Prophage	SPI-1 effector protein
<i>slrP</i>	18.5	3		SPI-1 effector protein
<i>sspH1</i>	nd	nd	Prophage (Gifsy-3)	SPI-1 effector protein
SPI-2	31	40	<i>valV</i> (tRNA gene)	SPI-2 type III secretion apparatus/ effector proteins/regulatory genes
<i>sspH2</i>	49	50	Prophage	SPI-2 effector protein
<i>sifA</i>	39	1.5		SPI-2 effector protein
<i>sifB</i>	35	3.5		SPI-2 effector protein
<i>sseI sodC</i>	24	50	Prophage (Gifsy-2)	SPI-2 effector protein, superoxide dismutase
<i>sseJ</i>	42	10		SPI-2 effector protein
SPI-3	82	18	<i>selC</i> (tRNA gene)	Mg <sup>2+</sup> transport
SPI-4	92	25		Putative type 1 secretion system
<i>shdA</i>	55	25	IS element	Intestinal colonization/shedding
<i>spv pef</i>	pSLT	90	Plasmid	ADP-ribosylating enzyme, fimbrial biosynthesis
<i>lpf</i>	80	7		Fimbrial biosynthesis
<i>bcf</i>	0.5	33		Fimbrial biosynthesis
<i>std</i>	66	13	<i>glyU</i> (tRNA gene)	Putative fimbrial biosynthesis
<i>stf</i>	5	7		Putative fimbrial biosynthesis
<i>sth</i>	100	6		Putative fimbrial biosynthesis
<i>fim</i>	15	31	<i>argF</i> (tRNA gene)	Fimbrial biosynthesis
<i>saf</i>	7	50	<i>aspV</i> (tRNA gene)	Putative fimbrial biosynthesis
<i>stb</i>	8	50	<i>thrW</i> (tRNA gene)	Putative fimbrial biosynthesis

<sup>a</sup> Not detected in serovar *Typhimurium* LT-2 genomic sequence.<sup>b</sup> Size based on comparison with the *E. coli* K-12 genomic sequence (McCLELLAND et al. 2000).

(Genome Sequencing Center, personal communication). Preliminary sequence data of the serovar *Typhimurium* LT-2 genome reveal that a third cluster of fimbrial biosynthesis genes (designated *std* operon) is located on a large DNA region absent from *E. coli* K-12 which is flanked by the tRNA gene *glyU* (Table 1) (Genome Sequencing Center, personal communication). Another large genetic island encoded at about 55min on the serovar *Typhimurium* genetic map encodes the *shdA* gene, which is required for efficient shedding of serovar Typhimurium from mice (KINGSLEY et al. 2000b). The island is approximately 25kb, contains the remnant of an IS1 element, and is flanked by the *xseA* and *yfgK* genes which are also present in *E. coli* K-12 (Fig. 1). While the role of *saf* and *std* in the pathogenesis of serovar *Typhimurium* infections is currently unknown, type 1 fimbriae and *shdA* are not required for virulence in the mouse model (KINGSLEY et al. 2000b; LOCKMAN and CURTISS III 1992). However, not all *Salmonella* pathogenicity islands contain genes that are required for full mouse virulence. For instance, mutations in SPI-5 of serovar *Dublin* do not increase the 50% lethal dose of serovar *Dublin* during

intraperitoneal infection of mice (WOOD et al. 1998). Similarly, a mutation in *sopB* (*sigD*) does not reduce lethality of serovar *Typhimurium* during oral infection of calves (TSOLIS et al. 1999a). Nonetheless, SPI-5 is considered a pathogenicity island because it is required for virulence associated phenotypes, such as fluid accumulation and neutrophil infiltration in bovine ligated ileal loops (WOOD et al. 1998). By analogy, it could be argued that the involvement of *shdA* in fecal shedding and intestinal colonization suffices for classifying the *xseA* and *yfgK* intergenic region as a pathogenicity island.

The *bcf* fimbrial operon was identified as required for colonization of bovine Peyer's patches in a signature-tagged mutagenesis screen (TSOLIS et al. 1999b). It is located between *dnaJ* and *nhoA* on an approximately 33-kb island that is not present in the *E. coli* K-12 genome (Fig. 1) (Genome Sequencing Center, personal communication). The *dnaJ-nhoA* intergenic region thus displays the basic features of a pathogenicity island. A second transposon insertion, reducing the ability of serovar *Typhimurium* to colonize murine spleen as shown by signature-tagged mutagenesis, is located in a gene encoding a putative efflux pump encoded on a 50-kb island in the *proA* and *yahN* intergenic region (STN206 in Fig. 1) (TSOLIS et al. 1999b). This serovar *Typhimurium* DNA region also contains fimbrial biosynthesis genes (*stb* operon) and the gene encoding the outer membrane ferrioxamine receptor protein FoxA (Genome Sequencing Center, personal communication). Characterization of a mutation in the *foxA* gene suggests a role of this DNA region in mouse virulence of serovar *Typhimurium* (KINGSLEY et al. 1999). The *proA-yahN* intergenic region is flanked on one side by a tRNA gene (*thrW*), an attachment site for bacteriophage P22 (*attP*) and an IS element. Therefore, the *proA-yahN* intergenic region also displays all the typical features of a pathogenicity island.

## 2.2 Pathogenicity Islet, Phage, and Plasmid-Encoded Virulence Determinants

### 2.2.1 Effector Proteins of SPI-1 and SPI-2

Although some type III-secreted effector proteins are encoded on SPI-1 and SPI-2, a considerable number are encoded by DNA regions located outside of these islands (Table 1). As described above, one of the effector proteins, SopB (SigD), is encoded on SPI-5 (Fig. 1). Several other effector genes, including *sopE1*, *sopE2*, *sseI*, *sspH1*, and *sspH2*, are flanked by genes with considerable sequence homology to genes of phages (HARDT et al. 1998; MIAO and MILLER 1999, 2000; WORLEY et al. 2000) (Genome Sequencing Center, personal communication). Indeed, SopE1 is carried by an intact bacteriophage capable of horizontal gene transfer between serovar *Typhimurium* strains through lysogenic conversion (MIROLD et al. 1999).

Finally, a number of effector proteins, including SopA, SopD, SlrP, SifA, SifB, and SseJ are encoded on DNA regions of less than 10kb which are not present in *E. coli* (Table 1) (MIAO and MILLER 2000, STEIN et al. 1996, TSOLIS et al. 1999b)

(Genome Sequencing Center, personal communication). The term pathogenicity islet has been proposed to describe these small DNA regions carrying virulence determinants (GROISMAN and OCHMAN 1997). Some pathogenicity islets consist of a single gene inserted between two housekeeping genes that are adjacent to each other in the *E. coli* K-12 genome. For instance, *sifA* is located on a 1.5-kb pathogenicity islet that is inserted within the *potABCD* operon, a housekeeping locus involved in transport of polyamines (STEIN et al. 1996). Support for the idea that this pathogenicity islet was acquired horizontally is the fact that it is flanked by 14-bp direct repeats and has a significantly lower G+C content (41%) than genes in the *potABCD* operon (51%). Similarly, *slrP* is encoded by a 3-kb pathogenicity islet inserted between *uvrB* and *yphK*, two genes that are adjacent in the *E. coli* K-12 genome (TSOLIS et al. 1999b). The remnant of a transposase gene located upstream of *slrP* and the G+C content of this pathogenicity islet (45%), which is considerably below that of the serovar *Typhimurium* genome (52%), are indicative of its acquisition by horizontal transfer.

### 2.2.2 Adhesins

As outlined above, several known or putative adherence determinants, including *fim*, *bcf*, *saf*, *stb*, *std*, and *shdA* are encoded on large DNA regions with one or several features of pathogenicity islands (Fig. 1, Table 1). Additional adherence determinants are encoded on pathogenicity islets or the serovar *Typhimurium* virulence plasmid. The *pef* fimbrial operon is located on the 90-kb virulence plasmid of serovar *Typhimurium* (FRIEDRICH et al. 1993). The finding that the plasmid is self-transmissible between serovar *Typhimurium* isolates underscores the importance of horizontal gene transfer in acquiring virulence genes (AHMER et al. 1999).

Additional serovar *Typhimurium* operons that are absent from *E. coli* K-12 and display homology to fimbrial biosynthesis genes include *stf*, *sth*, and *lpf* (BÄUMLER and HEFFRON 1995; EMMERTH et al. 1999; MORROW et al. 1999; Genome Sequencing Center, personal communication). These serovar *Typhimurium* operons form 6–7-kb insertions relative to the *E. coli* K-12 genome (Table 1). The role of *stf* and *sth* in the pathogenesis of serotype *Typhimurium* infections is currently unknown. However, the adhesin encoded by the *lpf* operon is required for full mouse virulence of serovar *Typhimurium*, and this DNA region is thus considered a pathogenicity islet (BÄUMLER et al. 1996, 1997b; VAN DER VELDEN et al. 1998).

### 2.2.3 Other Horizontally Acquired Virulence Determinants

In addition to the *pef* operon, the virulence plasmid encodes a cluster of five genes, the *spv* operon, which is required for full mouse virulence of serovar *Typhimurium* (Table 1) (GULIG and CURTISS 1987). The *spv* genes increase the intracellular growth rate of *S. typhimurium* within macrophages by an unknown mechanism (GULIG et al. 1998). The first gene, *spvR*, encodes a positive activator of *spvABCD* and negatively regulates an unknown repressor of the lpg-tubule formation

observable in epithelial cells (GUY et al. 2000; STEIN et al. 1996). The *spvB* gene encodes an ADP-ribosyl transferase (OTTO et al. 2000). The exact function of the proteins encoded by *spvA*, *spvC*, and *spvD* is not yet known.

The *sodC* gene encodes a Cu, Zn-superoxide dismutase which is required for survival in macrophages and full virulence of serovar *Typhimurium* in mice (DEGROOTE et al. 1997). This virulence gene is located on the same prophage (Gifsy-2) as *sseI* (Table 1).

## 2.3 Does Size Really Matter?

As discussed in the previous paragraphs, the serovar *Typhimurium* LT-2 genome contains numerous DNA regions (with a combined size of almost 600kb) which are absent from the *E. coli* K-12 genome and carry one or more virulence determinants (Table 1). Whether these DNA regions are classified as pathogenicity islands depends to a large extent on their respective sizes. An important rationale for using size as a criterion to define pathogenicity islands is the concept that the introduction of complex virulence factors which are encoded on large DNA regions may mark a quantum leap in the evolution of a bacterial pathogen (GROISMAN and OCHMAN 1996). As pointed out above, SPI-1 and SPI-2 appear to encode the prime virulence determinants for diarrheal disease (in calves) and typhoid fever-like disease (in mice), respectively, and acquisition of these determinants may thus have triggered such quantum leaps. The type III secretion systems encoded on SPI-1 and SPI-2 are highly complex, consisting of regulatory proteins, proteins associated with secretion of effector proteins across the bacterial inner and outer membrane, and proteins involved in the subsequent translocation of effector proteins across the cytoplasmic membrane of a host cell. Due to the complexity of these virulence factors, their introduction into the genome inevitably involves horizontal transfer of large DNA regions (> 35kb) (Fig. 1).

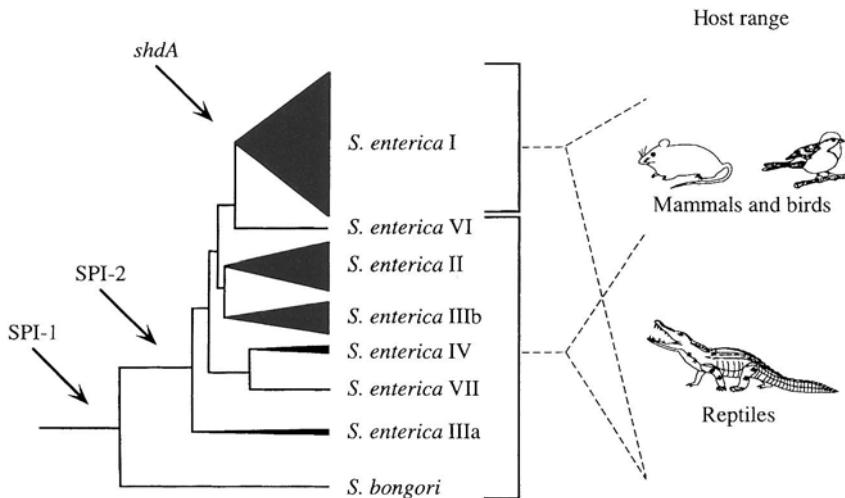
However, the concept that introduction of virulence determinants requires the transfer of large DNA regions does not apply to other pathogenicity islands of serovar *Typhimurium*. For instance, SPI-3 encodes only two virulence genes, namely *mgtB* and *mgtC* (BLANC-POTARD and GROISMAN 1997). Hybridization analysis with representative serovars of the genus *Salmonella* suggests that SPI-3 consists of several DNA regions with separate phylogenetic histories (BLANC-POTARD et al. 1999). This notion is further supported by the finding that genes encoded on the island exhibit a wide variability of G + C contents, ranging from 38% for *rmbA* to 53% in *misL*. It is therefore possible that various parts of SPI-3 were acquired through separate horizontal transfer events. If SPI-3 is viewed as being composed of several unrelated pieces of horizontally acquired DNA, it becomes clear that acquisition of each small piece (e.g., *mgtBC*) may represent the introduction of an islet rather than an island. The distinction between islets and islands based on size is further blurred by the fact that SPI-5 covers a mere 7kb (Fig. 1) (WOOD et al. 1998), a size similar to that of a number of pathogenicity islets present in the serovar *Typhimurium* genome (Table 1).

The distinction between pathogenicity islands and other horizontally acquired DNA regions encoding virulence determinants becomes even less meaningful when the functions of the respective genes in pathogenesis are considered. For instance, fimbrial operons of serovar *Typhimurium* may be located on a pathogenicity island (e.g., *bcf*), a plasmid (e.g., *pef*), or a pathogenicity islet (e.g., *lpf*) (Table 1). However, the location of these genes on different elements is not expected to alter their role during infection. Similarly, secreted targets of the type III secretion system encoded by SPI-1 may be encoded by a pathogenicity island (e.g., *sopB*), a pathogenicity islet (e.g., *slrP*), or a prophage (e.g., *sopE1*) (Table 1). Nonetheless, these genes are considered to be part of a single functional unit, namely the invasion-associated type III secretion system. Thus, it is not meaningful to exclude virulence determinants located on pathogenicity islets, phage, or plasmid when discussing the potential role of pathogenicity islands in host adaptation of *Salmonella* serovars.

### 3 Horizontal Gene Transfer and Host Adaptation

#### 3.1 Phylogenetic Studies

Acquisition of pathogenicity islands and islets by horizontal gene transfer was likely a key factor during evolution of virulence in bacterial pathogens (GROISMAN and OCHMAN 1996). Some lessons about the possible impact that acquisition of pathogenicity islands may have had can be inferred by determining the distribution of these DNA regions within the major evolutionary lineages present in the genus *Salmonella* (BÄUMLER 1997) (Fig. 3). For instance, based on the phylogenetic distribution and functional analysis of SPI-1 it has been speculated that a consequence of acquiring this pathogenicity island was the emergence of an ancestral diarrheal pathogen which gave rise to all extant *Salmonella* serovars (KINGSLEY et al. 2000a). Furthermore, horizontal gene transfer may have influenced speciation, since the formation of the two species assigned to the genus, *Salmonella bongori* and *S. enterica*, coincided with acquisition of SPI-2, as indicated by the presence of this pathogenicity island in only the latter species (HENSEL et al. 1997; OCHMAN and GROISMAN 1996). Investigation of the host range for extant *Salmonella* serovars suggests that the subsequent formation of one major evolutionary lineage within the genus *Salmonella*, *S. enterica* subspecies I, was accompanied by an expansion in host range to include warm-blooded animals (OCHMAN and WILSON 1987) (Fig. 3). That is, serovars belonging to *S. bongori* or *S. enterica* subspecies II–VII are mainly isolated from reptiles. In contrast, members of *S. enterica* subspecies I are in addition frequently isolated from mammals and birds and represent more than 99% of clinical isolates from human beings (ALEKSIC et al. 1996; BÄUMLER et al. 1998). The presence of a genetic island containing the *shdA* gene (Fig. 1) in 96% of *S. enterica* subspecies I isolates and its absence from *S. bongori* and *S. enterica* subspecies II–VII isolates indicates that formation of subspecies I was accompanied



**Fig. 3.** Phylogenetic relatedness and host range of extant *Salmonella* serovars. The branching structure of the phylogenetic tree shown on the left is based on comparative sequence analysis of housekeeping genes from different serovars (BOYD et al. 1996). Acquisition of SPI-1, SPI-2, and *shdA* by horizontal gene transfer likely occurred in branches indicated by arrows, as previously inferred from phylogenetic studies (HENSEL et al. 1997; KINGSLEY et al. 2000b; LI et al. 1995; OCHMAN and GROISMAN 1996). The lengths of vertical lines in triangles attached to the terminal branches of the tree are proportional to the number of serovars present in *S. bongori* and in the different subspecies of *S. enterica* (POPOFF and LE MINOR 1992). Roman numerals in the column adjacent to the tree indicate *S. enterica* subspecies. The host range of extant *Salmonella* serovars has been reported previously (ALEKSIC et al. 1996; POPOFF and LE MINOR 1992) and is shown on the right

by acquisition of this DNA region (KINGSLEY et al. 2000b) (Fig. 3). ShdA is a colonization factor which is required for prolonged shedding of *S. enterica* serovar *Typhimurium* from mice and displays sequence homology to AIDA, a nonfimbrial adhesin of diffuse adhering *E. coli* (DAEC). Thus, expansion in host range, characteristic of *S. enterica* subspecies I serovars, was accompanied by acquisition of at least one factor involved in host-pathogen interaction, suggesting that horizontal gene transfer may have played a role in adaptation to warm-blooded animals.

Analysis of *S. enterica* subspecies I serovars reveals that fimbrial adhesins and type III secreted effector proteins are frequently carried on phages, plasmids, pathogenicity islands, or pathogenicity islets (Table 1). Some of these virulence genes have a scattered phylogenetic distribution within *S. enterica* subspecies I, indicative of repeated acquisition or loss (BÄUMLER et al. 1997a; TSOLIS et al. 1999b). Shuffling of virulence determinants by horizontal transfer is expected to give rise to organisms expressing a particular gene in a new genetic background, thereby altering virulence traits. However, the question of whether acquisition of unique repertoires of fimbrial biosynthesis genes or genes encoding type III effector proteins by horizontal transfer is a mechanism to modulate the host range of a *Salmonella* serovar awaits further investigation.

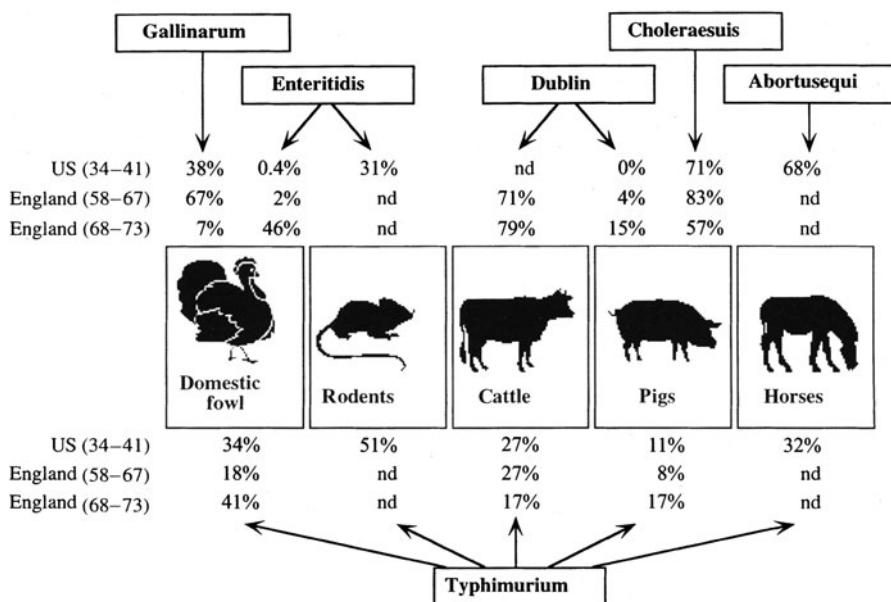
In summary, evolutionary studies are consistent with the idea that horizontal transfer of virulence genes was among the genetic mechanisms driving host adap-

tation, and they are therefore helpful in forming a framework in which to understand *Salmonella* pathogenesis. However, they provide little direct evidence to support this notion. The following sections will review several experimental strategies using various *S. enterica* subspecies I serovars which have been explored to improve our understanding of the molecular basis of host adaptation.

### 3.2 Experimental Approaches

#### 3.2.1 Host Adaptation and Host Range

Epidemiological surveys reveal that *S. enterica* subspecies I serovars display a spectrum of host specificity ranging from a narrow to a broad host range (Fig. 4). The prototypical host-restricted serovar is *Typhi*, which causes disease only in human beings and higher primates (GEOFFREY et al. 1960). Host restriction is less pronounced for other serovars displaying host specificity, such as the bovine-



**Fig. 4.** *S. enterica* subspecies I serovars most frequently associated with disease in various animal reservoirs. Data on the frequency at which serovars were associated with cases of disease in the United States (US) between 1934 and 1941 have been reported previously (EDWARDS and BRUNER 1943). Surveys on *Salmonella* serovars commonly associated with disease in England and Wales during 1957 and 1973 were performed by Sojka and co-workers (SOJKA and FIELD 1970; SOJKA and WRAY 1975). The frequency at which serovars were associated with illness is given as percent of the total number of *Salmonella* isolates. Arrows indicate the host range of *Salmonella* serovars suggested by these surveys. Data on serovar *Gallinarum* summarize information on both biotypes (*Pullorum* and *Gallinarum*). The reduced incidence of serovar *Gallinarum* between 1968–1973 compared with 1957–1967 was the result of a successful eradication program in England (Poultry Stock Improvement Plan) (SOJKA and FIELD 1970). nd, no data

adapted serovar *Dublin*, which not only causes disease in cattle, but is also frequently associated with illness in sheep and pigs (SOJKA and WRAY 1975). The other end of the spectrum is occupied by broad-host-range serovars, such as serovar *Typhimurium*, which is frequently associated with disease in a number of host species, including human beings, cattle, pigs, horses, poultry, rodents, and sheep (EDWARDS and BRUNER 1943; SOJKA and WRAY 1975; WRAY et al. 1981).

It should be pointed out that epidemiological surveys reveal the frequency at which a serovar is associated with cases of disease, but this does not necessarily indicate host adaptation. The reported frequency reflects a number of properties, including the ability of a pathogen to cause illness and its ability to either circulate or to be constantly introduced into a host population (KINGSLEY and BÄUMLER 2000). A serovar is considered host adapted if it is able to cause disease and circulate within a host population (Fig. 4). In contrast, if a serovar depends upon continuous reintroduction into a host population (i.e., through food) in order to persist then an association with disease may not be indicative of host adaptation. For example, serovar *Choleraesuis* causes a severe systemic infection in man with death rates exceeding even those among typhoid fever patients (SAPHRA and WASSERMANN 1954). However, despite its high virulence for man, serovar *Choleraesuis* is not considered human adapted because the organism is introduced into the human population from its porcine animal reservoir through food (person-to-person transmission is rare). On the other hand, serovar *Choleraesuis* is able not only to cause severe disease in pigs but also to persist in this animal reservoir through direct transmission. Therefore, serovar *Choleraesuis* is considered to be porcine adapted but not human adapted (Fig. 4). It becomes clear from these considerations that the severity of disease caused by a serovar in a host species is not by itself indicative of host adaptation. Although this example illustrates that features in addition to virulence are critically important for host adaptation, experimental approaches to the study of host adaptation commonly rely solely on virulence-associated phenotypes as readout.

### 3.2.2 Comparison of Host-Restricted and Broad-Host-Range Serovars

One approach used to study host adaptation is to compare *S. enterica* subspecies I serovars that differ with regard to host range in various animal or tissue-culture models. Comparison of various *S. enterica* subspecies I serovars in an animal model commonly reveals virulence-associated phenotypes that distinguish host-restricted serovars from those with a broad host range (BARROW et al. 1994; PASCOPELLA et al. 1995). For instance, this approach revealed that serovar *Typhi* is capable of invading M cells in murine Peyer's patches but, unlike serovar *Typhimurium*, it does not destroy these cells. Differences observed in animal models correlate in some cases with the behavior of those serovars in tissue-culture or organ-culture models (ALPUCHE-ARANDA et al. 1995; BOLTON et al. 1999a,b; ISHIBASHI and ARAI 1996; WATSON et al. 1995, 2000). However, a drawback of this type of analysis is that it is not suited for identification of virulence genes responsible for the characteristics that distinguish serovars during interaction with a particular host species. This

approach, therefore, does not reveal specific details as to whether pathogenicity islands or islets are important for host adaptation.

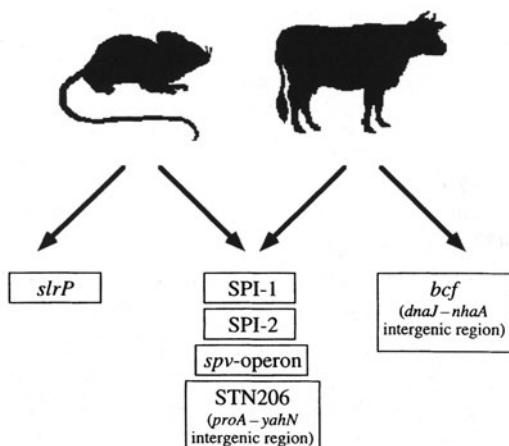
### **3.2.3 The Quest for Genes Expanding the Host Range of Host-Restricted Serovars**

A second strategy for studying the basis of host adaptation explored the possibility that the introduction of genes from a broad-host-range serovar may expand the host range of a host-restricted *Salmonella* serovar. Unlike host-restricted serovars, such as the human-adapted serovar *Typhi* or the avian-adapted serovar *Gallinarum*, the broad-host-range serovar *Typhimurium* is capable of causing disease in mice. In an attempt to identify genes that would confer mouse virulence DNA regions (i.e., cosmid clones or the virulence plasmid, respectively) of serovar *Typhimurium* were introduced into the host-restricted serovars *Gallinarum* and *Typhi* (PASCOPELLA et al. 1996; ROUDIER et al. 1990). However, no DNA regions important for mouse virulence were identified using this approach, suggesting that multiple unlinked genes may be responsible for the host-restricted phenotypes of serovars *Gallinarum* and *Typhi*.

Nonetheless, this approach may become feasible once genomic sequences of host-restricted and broad-host-range pathogens become available and can be compared. However, a clear understanding of what exactly can be learned from restoring mouse virulence to a host-restricted pathogen, such as serovar *Gallinarum*, will be important for interpreting the results of such a study. Based on phylogenetic analysis, it has been suggested that serovar *Gallinarum* evolved from a serovar *Enteritidis*-like ancestor which had a broader host range including rodents (LI et al. 1993). During its evolution from the hypothetical serovar *Enteritidis*-like ancestor, alterations in different genes may have accounted for the adaptation of serovar *Gallinarum* to one environment (i.e., poultry) and the loss of adaptation to another (i.e., mice). Consequently, by restoring mouse virulence of serovar *Gallinarum* we may learn little about pathogenicity islands or islets which represent adaptations to poultry. Instead, these studies will likely identify virulence functions that are dispensable during infection of poultry and were thus lost in the *Gallinarum* lineage.

### **3.2.4 Comparing the Repertoires of Genes One Serovar Uses for Infection of Two Different Host Species**

An alternate approach to studying host adaptation of a serovar is to compare the effect mutational inactivation of a gene has on virulence for two alternate host species. This strategy has been explored by comparing the virulence of randomly generated signature-tagged serovar *Typhimurium* mutants during oral infections of mice and calves (TSOLIS et al. 1999b). Although the bank analyzed was not comprehensive (260 mutants), this screen provided valuable insight into differences in the repertoires of virulence genes used by serovar *Typhimurium* during infection of two natural host species. Several transposon insertions that reduced the ability of



**Fig. 5.** Pathogenicity islands and islets of serovar *Typhimurium* identified by comparative screening of a random bank of signature-tagged transposon mutants during oral infection of mice and calves (TSOLIS et al. 1999b). Horizontally acquired DNA regions that are required only for infection of mice are listed on the left while those required for infection only of calves are shown on the right. DNA regions required for colonization of both bovine and murine tissues are listed in the center

mutants to colonize both bovine and murine tissues were located in well-characterized horizontally acquired virulence factors, including the *spv* operon, SPI-1 and SPI-2 (Fig. 5). Furthermore, one insertion (designated STN206 in Figs. 1 and 5) was identified in a genetic island located in the *proA*-*yahN* intergenic region (Genome Sequencing Center, personal communication).

In addition to virulence factors required for colonization of both mice and calves, the screen identified two *Salmonella*-specific DNA regions that were required for infection of only one of the two host species and may therefore represent host-range factors (Fig. 5) (TSOLIS et al. 1999b). One mutant defective for colonization of bovine but not murine intestinal tissues carried a transposon insertion in *bcfC*, a putative fimbrial biosynthesis gene located on a genetic island (Fig. 1, Table 1) (Genome Sequencing Center, personal communication). The transposon insertion in a mutant which displayed a colonization defect only during infection of mice was inserted in a 3-kb pathogenicity islet encoding SlrP, a type III effector protein (Table 1) (TSOLIS et al. 1999b). Thus, this study suggests that adhesins and type III effector proteins encoded by pathogenicity islands or islets may function as host-range factors of serovar *Typhimurium*.

Further support for the concept that type III secreted effector proteins contribute to host adaptation comes from comparative analysis of the effect mutations in *sspH1* and *sspH2* have on the virulence of serovar *Typhimurium* for mice and calves. Simultaneous inactivation of *sspH1* and *sspH2* reduces virulence during oral infection of calves but does not decrease the virulence of serovar *Typhimurium* during oral infection of mice (MIAO et al. 1999). Similarly, inactivation of the type III effector gene *sipA* (*sspA*) renders serovar *Typhimurium* unable to cause lethal oral infections in calves but does not reduce virulence of this pathogen during oral infections of mice (TSOLIS et al. 2000). These examples illustrate that different repertoires of type III effector proteins of serovar *Typhimurium* may be important for causing disease in mice and calves.

### 3.3 Epidemiological Studies

Epidemiological surveys of serovar *Typhimurium* reveal that persistence of this pathogen in animal reservoirs is characterized by a rise and fall of clonal populations that can be distinguished by phage typing. After dominating for a period of time, each epidemic strain is replaced by a new clone as indicated by the preponderance of a new phage type. For instance, serovar *Typhimurium* definitive phage type (DT) 204 was first isolated in 1972 from cattle and beef products in Germany (KUHN et al. 1982). During the 1970s, DT204 became the phage type most frequently isolated from pigs and cattle in Germany and subsequently spread among cattle populations in England and Wales (THRELFALL et al. 1978). The preponderance of this phage type ended in the 1980s, when it was replaced by phage types DT204c and DT49, which dominated in cattle populations in England and Wales during this decade (THRELFALL et al. 1990, 1993). Finally, in the 1990s these strains were replaced by DT104, a clone which first emerged in cattle populations in England and Wales in 1984 and currently represents the predominant serovar *Typhimurium* phage type in Europe (THRELFALL et al. 1994). Although the rise and fall of epidemic serovar *Typhimurium* clones is well documented, factors responsible for the temporal preponderance of individual phage types have remained an enigma.

Recent evidence suggests that horizontal gene transfer is a possible factor contributing to the temporal dominance of individual phage types. Hardt and co-workers screened a representative set of serovar *Typhimurium* phage types isolated in Germany for the presence of strains that are lysogenized by a bacteriophage carrying *sopE1*, a type III effector gene (Table 1) (HARDT et al. 1998; MIROLD et al. 1999). The *sopE1* gene was present in only a small number of phage types, including DT204, DT204c, and DT49. The finding that the *sopE1* gene is present in a small number of serovar *Typhimurium* strains representing epidemic cattle-associated phage types further supports the concept that type III effector genes are host-range factors important for circulation in the bovine animal reservoir. The presence of *sopE1* on an intact bacteriophage raises the possibility that acquisition of host-range factors by horizontal gene transfer may contribute to the emergence of new epidemic strains (MIROLD et al. 1999).

## 4 Conclusion

Horizontally acquired virulence genes may be carried by plasmids, prophages, pathogenicity islands, or pathogenicity islets. Since components of complex virulence determinants, such as type III effector genes, may be carried by any of these elements, a clear distinction between these mobile genetic elements is a matter of semantics. While the importance of virulence factors associated with mobile genetic elements for pathogenicity of *Salmonella* serovars is well established, little is known

about the genetic basis of host adaptation. Recent studies have furthered our understanding of how horizontal acquisition of virulence determinants may contribute to host adaptation, and two common themes are beginning to emerge. First, adhesins have been implicated in host adaptation by evolutionary studies (*shdA*) or through animal experiments (*bcf*). Second, the concept that type III effector genes contribute to host adaptation is supported by epidemiological arguments (*sopE1*) and comparative analysis of serovar *Typhimurium* mutants during infection of mice and cows (*slrP*, *sspH1*, *sspH2*). Horizontal transfer of adhesins and type III effector genes may thus have been an important mechanism to modulate host adaptation of *Salmonella* serovars.

*Acknowledgements.* The authors wish to thank the Genome Sequencing Center, Washington University, St. Louis, for communication of DNA sequence data prior to publication.

This material is based in part upon work supported by the Texas Advanced Research (Technology) Program under grant number 000089-0051-1999. Work in A.B.'s laboratory is supported by Public Health Service grants #AI40124 and #AI44170.

## References

- Ahmer BM, Tran M, Heffron F (1999) The virulence plasmid of *Salmonella typhimurium* is self-transmissible. *J Bacteriol* 181:1364–1368
- Aleksic S, Heinzerling F, Bockemühl J (1996) Human infection caused by salmonellae of subspecies II to VI in Germany, 1977–1992. *Zbl Bakter* 283:391–398
- Alpuche-Aranda CM, Berthiaume EP, Mock B, Swanson JA, Miller SI (1995) Spacious phagosome formation within mouse macrophages correlates with *Salmonella* serotype pathogenicity and host susceptibility. *Infect Immun* 63:4456–4462
- Barrow PA, Huggins MB, Lovell MA (1994) Host specificity of *Salmonella* infection in chickens and mice is expressed *in vivo* primarily at the level of the reticuloendothelial system. *Infect Immun* 62:4602–4610
- Bäumler AJ (1997) The record of horizontal gene transfer in *Salmonella*. *Trends Microbiol* 5:318–322
- Bäumler AJ, Heffron F (1995) Identification and sequence analysis of *lpfABCDE*, a putative fimbrial operon of *Salmonella typhimurium*. *J Bacteriol* 177:2087–2097
- Bäumler AJ, Tsolis RM, Heffron F (1996) The *lpf* fimbrial operon mediates adhesion to murine Peyer's patches. *Proc Natl Acad Sci USA* 93:279–283
- Bäumler AJ, Gilde AJ, Tsolis RM, van der Velden AWM, Ahmer BMM, Heffron F (1997a) Contribution of horizontal gene transfer and deletion events to the development of distinctive patterns of fimbrial operons during evolution of *Salmonella* serotypes. *J Bacteriol* 179:317–322
- Bäumler AJ, Tsolis RM, Valentine PJ, Ficht TA, Heffron F (1997b) Synergistic Effect of Mutations in *invA*, *lpfC* on the Ability of *Salmonella typhimurium* to Cause Murine Typhoid. *Infect Immun* 65:2254–2259
- Bäumler AJ, Tsolis RM, Ficht TA, Adams LG (1998) Evolution of host adaptation in *Salmonella enterica*. *Infect Immun* 66:4579–4587
- Beuzon CR, Meresse S, Unsworth KE, Ruiz-Albert J, Garvis S, Waterman SR, Ryder TA, Boucrot E, Holden DW (2000) *Salmonella* maintains the integrity of its intracellular vacuole through the action of SifA. [published erratum appears in EMBO J 2000 Aug 1, 19:4191] *Embo J* 19:3235–3249
- Blanc-Potard AB, Groisman EA (1997) The *Salmonella selC* locus contains a pathogenicity island mediating intramacrophage survival. *EMBO J* 16:5376–5385
- Blanc-Potard AB, Solomon F, Kayser J, Groisman EA (1999) The SPI-3 pathogenicity island of *salmonella enterica* [In Process Citation] *J Bacteriol* 181:998–1004
- Blum G, Ott M, Lichewski A, Ritter A, Imrich H, Tschäpe H, Hacker J (1994) Excision of large DNA regions termed pathogenicity islands from t-RNA-specific loci in the chromosome of an *Escherichia coli* wild type pathogen. *Infect Immun* 62:606–614

- Bolton AJ, Martin GD, Osborne MP, Wallis TS, Stephen J (1999a) Invasiveness of *Salmonella* serotypes Typhimurium, Choleraesuis and Dublin for rabbit terminal ileum in vitro. *J Med Microbiol* 48:801–810
- Bolton AJ, Osborne MP, Wallis TS, Stephen J (1999b) Interaction of *Salmonella choleraesuis*, *Salmonella dublin* and *Salmonella typhimurium* with porcine and bovine terminal ileum in vivo. *Microbiology* 145:2431–2441
- Boyd EF, Wang FS, Whittam TS, Selander RK (1996) Molecular genetic relationship of the *Salmonellae*. *Appl Environ Microbiol* 62:804–808
- Caugant DA, Levin BR, Lidin-Janson G, Whittam TS, Svanborg Eden C, Selander RK (1983) Genetic diversity and relationships among strains of *Escherichia coli* in the intestine and those causing urinary tract infections. *Prog Allergy* 33:203–227
- Chen LM, Kaniga K, Galan JE (1996) *Salmonella* spp are cytotoxic for cultured macrophages. *Mol Microbiol* 21:1101–1115
- DeGroote MA, Ochsner UA, Shiloh MU, Nathan C, McCord JM, Dinauer MC, Libby SJ, Vazquez TA, Xu Y, Fang FC (1997) Periplasmic superoxide dismutase protects *Salmonella* from products of phagocyte NADPH-oxidase and nitric oxide synthase. *Proc Natl Acad Sci USA* 94:13997–4001
- Edwards PR, Bruner, DW (1943) The occurrence and distribution of *Salmonella* types in the United States. *J Infect Dis* 72:58–67
- Emmerth M, Goebel W, Miller SI, Hueck CJ (1999) Genomic subtraction identifies *Salmonella typhimurium* prophages, F-related plasmid sequences, and a novel fimbrial operon, stf, which are absent in *Salmonella typhi*. *J Bacteriol* 181:5652–5661
- Folkesson A, Advani A, Sukupolvi S, Pfeifer JD, Normark S, Lofdahl S (1999) Multiple insertions of fimbrial operons correlate with the evolution of *Salmonella* serovars responsible for human disease. *Mol Microbiol* 33:612–622
- Friedrich MJ, Kinsey NE, Vila J, Kadner RJ (1993) Nucleotide sequence of a 139-kb segment of the 90-kb virulence plasmid of *Salmonella typhimurium*: the presence of fimbrial biosynthetic genes. *Mol Microbiol* 8:543–558
- Galán JE, Curtiss R III (1989) Cloning and molecular characterization of genes whose products allow *Salmonella typhimurium* to penetrate tissue culture cells. *Proc Natl Acad Sci USA* 86:6383–6387
- Galyov EE, Wood MW, Rosqvist R, Mullan PB, Watson PR, Hedges S, Wallis TS (1997) A secreted effector protein of *Salmonella dublin* is translocated into eukaryotic cells and mediates inflammation and fluid secretion in infected ileal mucosa. *Mol Microbiol* 25:903–912
- Geoffrey E, Gaines S, Landy M, Tigert WD, Sprintz H, Trapani RJ, Mandel AD, Benenson A S (1960) Studies on infection and immunity in experimental typhoid fever: typhoid fever in chimpanzees orally infected with *Salmonella typhosa*. *J Exp Med* 112:143–166
- Groisman EA, Ochman H (1996) Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* 87:791–794
- Groisman EA, Ochman H (1997) How *Salmonella* became a pathogen. *Trends Microbiol* 5:343–349
- Gulig PA, Curtiss R (1987) Plasmid-associated virulence of *Salmonella typhimurium*. *Infect Immun* 1987:2891–2901
- Gulig PA, Doyle TJ, Hughes JA, Matsui H (1998) Analysis of host cells associated with the Spv-mediated increased intracellular growth rate of *Salmonella typhimurium* in mice. *Infect Immun* 66:2471–2485
- Guy RL, Gonias LA, Stein MA (2000) Aggregation of host endosomes by salmonella requires SPI2 translocation of SseFG and involves SpvR and the fms-aroE intragenic. *Mol Microbiol* 37:1417–1435
- Hacker J, Knapp S, Goebel W (1983) Spontaneous deletions and flanking regions of the chromosomally inherited hemolysin determinant of an *Escherichia coli* O6 strain. *J Bacteriol* 154:1145–1152
- Hacker J, Bender L, Ott M, Wingender J, Lund B, Marre R, Goebel W (1990) Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal *Escherichia coli* isolates. *Microb Pathog* 8:213–225
- Hacker J, Blum-Oehler G, Muhrdorfer I, Tscharte H (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* 23:1089–1097
- Hardt WD, Urlaub H, Galan JE (1998) A substrate of the centosome 63 type III protein secretion system of *Salmonella typhimurium* is encoded by a cryptic bacteriophage. *Proc Natl Acad Sci USA* 95: 2574–2579
- Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW (1995) Simultaneous identification of bacterial virulence genes by negative selection. *Science* 269:400–403
- Hensel M, Shea JE, Bäumler AJ, Gleeson C, Blattner F, Holden DW (1997) Analysis of the boundaries of *Salmonella* pathogenicity island 2 and the corresponding chromosomal region of *Escherichia coli* K-12. *J Bacteriol* 179:1105–1111

- Hensel M, Nikolaus T, Egelseer C (1999) Molecular and functional analysis indicates a mosaic structure of *Salmonella* pathogenicity island 2. *Mol Microbiol* 31:489–498
- Hobbie S, Chen LM, Davis RJ, Galan JE (1997) Involvement of mitogen-activated protein kinase pathways in the nuclear responses and cytokine production induced by *Salmonella typhimurium* in cultured intestinal epithelial cells. *J Immunol* 159:5550–5559
- Hong KH, Miller VL (1998) Identification of a novel *Salmonella* invasion locus homologous to *Shigella* ipgDE. *J Bacteriol* 180:1793–1802
- Hueck CJ (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* 62:379–433
- Hughes C, Hacker J, Duvel H, Goebel W (1987) Chromosomal deletions and rearrangements cause coordinate loss of haemolysis, fimbriation and serum resistance in a uropathogenic strain of *Escherichia coli*. *Microb Pathog* 2:227–230
- Ishibashi Y, Arai T (1996) A possible mechanism for host-specific pathogenesis of *Salmonella* serovars. *Microb Pathogen* 21:435–446
- Kingsley RA, Bäumler AJ (2000) Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm. *Mol Microbiol* 36:1006–1014
- Kingsley RA, Reissbrodt R, Rabsch W, Ketley JM, Tsolis RM, Everest P, Dougan G, Baumler AJ, Roberts M, Williams PH (1999) Ferrioxamine-mediated iron(III) utilization by *Salmonella enterica*. *Appl Environ Microbiol* 65:1610–1618
- Kingsley RA, Tsolis RM, Townsend SM, Norris TL, TAF, Adams LG, Bäumler AJ (2000a) Impact of horizontal gene transfer on the evolution of *Salmonella* pathogenesis. In: Brogden KA, Roth JA, Stanton TB, Bolin CA, Minion FC, Wannemuehler MJ (eds) *Virulence mechanisms of bacterial pathogens*. ASM Press, Washington, DC, pp 227–240
- Kingsley RA, van Amsterdam K, Kramer N, Baumler AJ (2000b) The shdA gene is restricted to serotypes of *Salmonella enterica* subspecies I and contributes to efficient and prolonged fecal shedding. *Infect Immun* 68:2720–2727
- Knapp S, Hacker J, Jarchau T, Goebel W (1986) Large unstable inserts in the chromosome affect virulence properties of uropathogenic *Escherichia coli* 6 strain 536. *J Bacteriol* 168:22–30
- Kuhn H, Rabsch W, Tschape H, Tietze E (1982) Characterization and epidemiology of a *Salmonella typhimurium* epidemic strain (in German). *Z Arztl Fortbild (Jena)* 76:607–610
- Lawrence JG, Ochman H (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci USA* 95:9413–9417
- Li J, Smith NH, Nelson K, Crichton PB, Old DC, Whittam TS, Selander RK (1993) Evolutionary origin and radiation of the avian-adapted non-motile *salmonellae*. *J Med Microbiol* 38:129–139
- Li J, Ochman H, Groisman EA, Boyd EF, Solomon F, Nelson K, Selander RK (1995) Relationship between evolutionary rate and cellular location among the Inv/Spa invasion proteins of *Salmonella enterica*. *Proc Natl Acad Sci USA* 92:7252–7256
- Lockman HA, Curtiss III R (1992) Virulence of non-type 1-fimbriated and nonfimbriated nonflagellated *Salmonella typhimurium* mutants in murine typhoid fever. *Infect Immun* 60:491–496
- McClelland M, Florea L, Sanderson K, Clifton SW, Parkhill J, Churcher C, Dougan G, Wilson RK, Miller W (2000) Comparison of the *Escherichia coli* K-R genome with sampled genomes of a *Klebsiella pneumoniae* and three *Salmonella enterica* serovars, Typhimurium, Typhi and Paratyphi A. *Nucleic Acids Res* 28:4974–4986
- Miao EA, Miller SI (1999) Bacteriophages in the evolution of pathogen-host interactions. *Proc Natl Acad Sci USA* 96:9452–9454
- Miao EA, Miller SI (2000) A conserved amino acid sequence directing intracellular type III secretion by *Salmonella typhimurium*. *Proc Natl Acad Sci USA* 97:7539–7944
- Miao EA, Scherer CA, Tsolis RM, Kingsley RA, Adams LG, Baumler AJ, Miller SI (1999) *Salmonella typhimurium* leucine-rich repeat proteins are targeted to the SPI1 and SPI2 type III secretion systems. *Mol Microbiol* 34:850–864
- Mills DM, Bajaj V, Lee CA (1995) A 40-kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. *Mol Microbiol* 15:749–759
- Mirold S, Rabsch W, Rohde M, Stender S, Tschape H, Russmann H, Igwe E, Hardt WD (1999) Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. *Proc Natl Acad Sci USA* 96:9845–9850
- Monack DM, Raupach B, Hromockyj AE, Falkow S (1996) *Salmonella typhimurium* invasion induces apoptosis in infected macrophages. *Proc Natl Acad Sci USA* 93:9833–9838

- Morrow BJ, Graham JE, Curtiss R 3rd (1999) Genomic subtractive hybridization and selective capture of transcribed sequences identify a novel *Salmonella typhimurium* fimbrial operon and putative transcriptional regulator that are absent from the *Salmonella typhi* genome. *Infect Immun* 67:5106–5116
- Nataro JP, Kaper JB (1998) Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 11:142–201
- Norris FA, Wilson MP, Wallis TS, Galyov EE, Majerus PW (1998) SopB, a protein required for virulence of *Salmonella dublin*, is an inositol phosphate phosphatase [see comments]. *Proc Natl Acad Sci USA* 95:14057–14059
- Ochman H, Groisman EA (1996) Distribution of pathogenicity islands in *Salmonella* spp. *Infect Immun* 64:5410–5412
- Ochman H, Selander RK (1984) Evidence for clonal population structure in *Escherichia coli*. *Proc Natl Acad Sci USA* 81:198–201
- Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 26:74–86
- Ochman H, Soncini FC, Solomon F, Groisman EA (1996) Identification of a pathogenicity island for *Salmonella* survival in host cells. *Proc Natl Acad Sci USA* 93:7800–7804
- Otto H, Tezcan-Merdol D, Girisch R, Haag F, Rhen M, Koch-Nolte F (2000) The spvB gene-product of the *salmonella enterica* virulence plasmid is a mono(ADP-ribosyl)transferase. *Mol Microbiol* 37: 1106–1115
- Pascopella L, Raupach B, Ghori N, Monack D, Falkow S, Small PL (1995) Host restriction phenotypes of *Salmonella typhi* and *Salmonella gallinarum*. *Infect Immun* 63:4329–4335
- Pascopella L, Falkow S, Small PLC (1996) Identification of a genetic determinant from *Salmonella typhimurium* that confers upon *Salmonella gallinarum* enhanced survival in the mouse. 96th Annual ASM Meeting, New Orleans, Louisiana, American Society for Microbiology
- Popoff MY, Le Minor L (1992) Antigenic formulas of the *Salmonella* serovars. WHO Collaborating Center for Reference and Research on *Salmonella*, Institute Pasteur, Paris
- Ritter A, Blum G, Emody L, Kerenyi M, Bock A, Neuhierl B, Rabsch W, Scheutz F, Hacker J (1995) tRNA genes and pathogenicity islands: influence on virulence and metabolic properties of uropathogenic *Escherichia coli*. *Mol Microbiol* 17:109–121
- Roudier C, Krause M, Fierer J, Guiney DG (1990) Correlation between the presence of sequences homologous to the vir region of *Salmonella dublin* plasmid pSDL2 and the virulence of twenty-two *Salmonella* serotypes in mice. *Infect Immun* 58:1180–1185
- Saphra I, Wassermann M (1954) *Salmonella Cholerae suis*. A clinical and epidemiological evaluation of 329 infections identified between 1940 and 1954 in the New York *Salmonella* Center. *Am J Med Sci* 228:525–533
- Selander RK, Levin BR (1980) Genetic diversity and structure in *Escherichia coli* populations. *Science* 210:545–547
- Sojka WJ, Field HI (1970) Salmonellosis in England and Wales 1958–1967. *Vet Bull* 40:515–531
- Sojka WJ, Wray C (1975) Incidence of *Salmonella* infection in animals in England and Wales 1968–1973. *Vet Rec* 96:280–284
- Steele-Mortimer O, Knodler LA, Marcus SL, Scheid MP, Goh B, Pfeifer CG, Duronio V, Finlay BB (2000) Activation of Akt/protein kinase B in epithelial cells by the *Salmonella typhimurium* effector SigD. *J Biol Chem* 275:37718–37724
- Stein MA, Leung KY, Zwick M, Garcia-del Portillo F, Finlay BB (1996) Identification of a *Salmonella* virulence gene required for formation of filamentous structures containing lysosomal membrane glycoproteins within epithelial cells. *Mol Microbiol* 20:151–164
- Threlfall EJ, Ward LR, Rowe B (1978) Spread of multiresistant strains of *Salmonella typhimurium* phage types 204 and 193 in Britain. *Br Med J* 2:997
- Threlfall EJ, Frost JA, Ward LR, Rowe B (1990) Plasmid profile typing can be used to subdivide phage-type 49 of *Salmonella typhimurium* in outbreak investigations. *Epidemiol Infect* 104:243–251
- Threlfall EJ, Rowe B, Ward LR (1993) A comparison of multiple drug resistance in *salmonellas* from humans and food animals in England and Wales, 1981 and 1990. *Epidemiol Infect* 101:189–197
- Threlfall EJ, Frost JA, Ward LR, Rowe B (1994) Epidemic in cattle and humans of *Salmonella typhimurium* DT104 with chromosomally integrated multiple drug resistance. *Vet Rec* 134:577
- Tsolis RM, Adams LG, Ficht TA, Baumler AJ (1999a) Contribution of *Salmonella typhimurium* virulence factors to diarrheal disease in calves. *Infect Immun* 67:4879–4885
- Tsolis RM, Townsend SM, Miao EA, Miller SI, Ficht TA, Adams LG, Baumler AJ (1999b) Identification of a putative *Salmonella enterica* serotype *typhimurium* host range factor with homology to IpaH and YopM by signature-tagged mutagenesis. *Infect Immun* 67:6385–6393

- Tsolis RM, Adams LG, Hantman MJ, Scherer CA, Kimborough T, Kingsley RA, Ficht TA, Miller SI, Bäumler AJ (2000) SspA is required for lethal *Salmonella typhimurium* infections in calves but is not essential for diarrhea. *Infect Immun* 68:3158–3163
- Uchiya K, Barbieri MA, Funato K, Shah AH, Stahl PD, Groisman EA (1999) A *Salmonella* virulence protein that inhibits cellular trafficking. *Embo J* 18:3924–3933
- van der Velden AWM, Baumler AJ, Tsolis RM, Heffron F (1998) Multiple fimbrial adhesins are required for full virulence of *Salmonella typhimurium* in mice. *Infect Immun* 66:2803–2808
- van Der Velden AW, Lindgren SW, Worley MJ, Heffron F (2000) *Salmonella* pathogenicity island 1-independent induction of apoptosis in infected macrophages by *Salmonella enterica* serotype *typhimurium*. *Infect Immun* 68:5702–5709
- Vazquez-Torres A, Xu Y, Jones-Carson J, Holden DW, Lucia SM, Dinauer MC, Mastroeni P, Fang FC (2000) *Salmonella* pathogenicity island 2-dependent evasion of the phagocyte NADPH oxidase. *Science* 287:1655–1658
- Watson PR, Paulin SM, Bland AP, Jones PW, Wallis TS (1995) Characterization of intestinal invasion by *Salmonella typhimurium* and *Salmonella dublin* and effect of a mutation in the invH gene. *Infect Immun* 63:2743–2754
- Watson PR, Paulin SM, Jones PW, Wallis TS (2000) Interaction of *Salmonella* serotypes with porcine macrophages in vitro does not correlate with virulence. *Microbiology* 146:1639–1649
- Wong KK, McClelland M, Stillwell LC, Sisk EC, Thurston SJ, Saffer JD (1998) Identification and sequence analysis of a 27-kilobase chromosomal fragment containing a *Salmonella* pathogenicity island located at 92 minutes on the chromosome map of *Salmonella enterica* serovar *typhimurium* LT2. *Infect Immun* 66:3365–3371
- Wood MW, Jones MA, Watson PR, Hedges S, Wallis TS, Galyov EE (1998) Identification of a pathogenicity island required for *Salmonella* enteropathogenicity. *Mol Microbiol* 29:883–891
- Worley MJ, Ching KH, Heffron F (2000) *Salmonella* SsrB activates a global regulon of horizontally acquired genes. *Mol Microbiol* 36:749–761
- Wray C, Sojka WJ, Bell JC (1981) *Salmonella* infection in horses in England and Wales, 1973 to 1979. *Vet Rec* 109:398–401

# Plasmids and Pathogenicity Islands of *Yersinia*

E. CARNIEL

1	Introduction . . . . .	89
2	Plasmids of <i>Yersinia</i> . . . . .	90
2.1	The Virulence Plasmid of Pathogenic <i>Yersinia</i> (pYV) . . . . .	91
2.1.1	Adhesion to the Cell Surface . . . . .	91
2.1.2	Assembly of the Ysc Secretion Apparatus . . . . .	92
2.1.3	Injection of the Effector Proteins . . . . .	92
2.1.4	Attack of the Target Cells . . . . .	92
2.2	The Fraction 1 Plasmid of <i>Y. pestis</i> (pFra) . . . . .	93
2.3	The Plasminogen Activator Plasmid of <i>Y. pestis</i> (pPla) . . . . .	94
2.4	The pVM82 Plasmid of <i>Y. pseudotuberculosis</i> . . . . .	95
2.5	The Antibiotic Resistance Plasmids of <i>Y. pestis</i> . . . . .	96
3	The High-Pathogenicity Island . . . . .	96
3.1	Genetic Organization of the HPI . . . . .	97
3.1.1	The Yersiniabactin Locus . . . . .	97
3.1.1.1	The Yersiniabactin Biosynthesis Genes . . . . .	99
3.1.1.2	The Yersiniabactin Transport Genes . . . . .	99
3.1.1.3	The Yersiniabactin Regulation Genes . . . . .	99
3.1.2	The Right-Hand Extremity of the HPI . . . . .	100
3.1.3	The Left-Hand Region of the HPI . . . . .	100
3.2	Mobility of the HPI . . . . .	101
3.3	Identification of the HPI Among Other Enterobacteria . . . . .	102
4	Conclusion . . . . .	103
	References . . . . .	104

## 1 Introduction

The genus *Yersinia* is composed of 11 species, three of which are pathogenic for human beings: *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*. The former two are widely spread in temperate countries; they are transmitted by the fecal-oral route and cause digestive symptoms, usually of moderate intensity. In contrast, *Y. pestis* predominates in tropical and subtropical areas; it is transmitted mainly by flea bites and causes an extremely severe disease called plague.

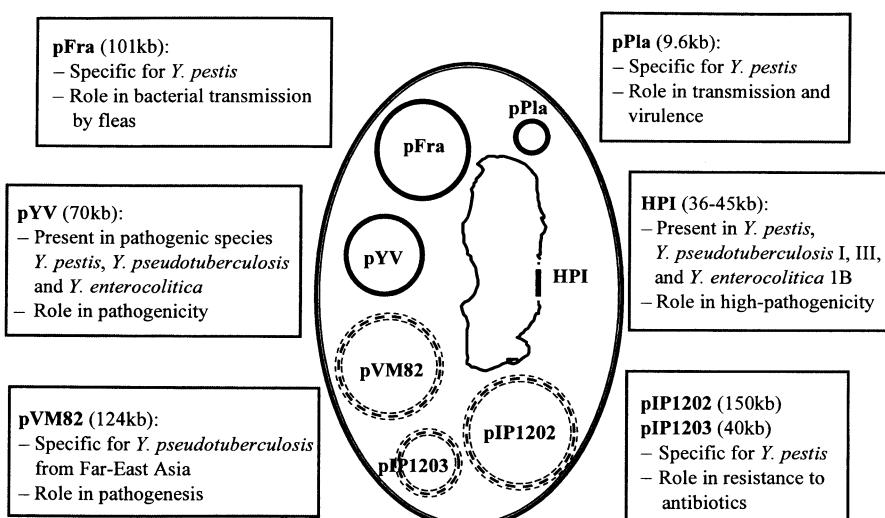
---

Laboratoire des *Yersinia*, Unité des Bactériologie Moléculaire et Médicale, Institut Pasteur, 28, rue du Dr. Roux, 75724 Paris Cedex 15, France

It is now clear that acquisition of genetic elements by horizontal transfer is crucial for bacterial evolution, virulence, and transmission. The genus *Yersinia* represents a good example of a bacterial group whose pathogenicity and transmission depend on horizontally acquired foreign DNA (Fig. 1). Hence, the virulence of the three pathogenic *Yersinia* species is strictly dependent on the presence of a virulence plasmid (pYV) which is absent from the nonpathogenic species. A subgroup of pathogenic *Yersinia* have the ability to cause severe and sometimes lethal infections. This high-pathogenicity phenotype is directly linked to the presence of a foreign piece of chromosomal DNA termed high-pathogenicity island (HPI). The ability of *Y. pestis* to be transmitted by flea bites is also dependent on the presence of two additional resident plasmids (pPla and pFra), specific to this species. Finally, other plasmids restricted to some strains confer on *Y. pestis* the ability to resist antibiotic treatment (pIP1202 and pIP1203) or on *Y. pseudotuberculosis* the capacity to cause a new type of disease (pVM82). This chapter will give a brief overview of these different acquired genetic elements.

## 2 Plasmids of *Yersinia*

A wide range of plasmids may be harbored by virulent as well as avirulent strains of *Yersinia*. Only those that are relevant in terms of public health (virulence, resistance, transmission) will be described here.



**Fig. 1.** Horizontally acquired genetic elements in *Yersinia*. These various elements are dispensable for bacterial survival but they are essential for the pathogenicity and/or the transmission of these microorganisms. Continuous lines indicate resident and ubiquitous genetic elements. Broken lines show plasmids that are found in localized areas and in specific isolates

## 2.1 The Virulence Plasmid of Pathogenic *Yersinia* (pYV)

This 70-kb plasmid, termed pYV (for plasmid *Yersinia* Virulence) or pCD1 (for plasmid encoding calcium dependence) is unique to pathogenic *Yersinia* species, i.e., *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica* (except strains of biotype 1A) and is crucial for their virulence. The pYV has a G+C content of 44.8%, slightly lower than that of the core genome, and contains several partial or complete insertion sequences scattered throughout the plasmid (Hu et al. 1998; PERRY et al. 1998). It has been estimated that seven copies of this replicon are present per cell. Its origin of replication is homologous to that of IncFIIA plasmids, although pYV is compatible with plasmid R100. The genes harbored by this replicon are highly conserved among the three pathogenic species, but the overall organization of the plasmid differs. This, in addition to a heterogeneous G+C content throughout the plasmid and the presence of numerous mobile elements, suggests a complex history of DNA acquisition, deletions, insertions, translocations, and internal rearrangements (Hu et al. 1998; PERRY et al. 1998).

The roles and targets of the pYV have been extensively studied over the past years and are now much better understood. The literature is so abundant on this subject that it will not be possible to cite each relevant reference in the context of this chapter. Several comprehensive reviews with voluminous references may provide more detailed information about the organization and function of pYV (BLEVES and CORNELIS 2000; BLOCKER et al. 2000; CORNELIS 1998, 2000; CORNELIS and VAN GIJSEGEM 2000; CORNELIS and WOLF WATZ 1997; FÄLLMAN et al. 1998; GALÁN and BLISKA 1996; STRALEY 1988; STRALEY and PERRY 1995).

The major functions of this replicon are to prevent phagocytosis of the adhering bacteria and to overcome the innate and specific immune response of the mammalian host, thus allowing better *in vivo* survival and multiplication of pathogenic *Yersinia*. The genes located on the pYV plasmid may be roughly divided into three major functional groups corresponding to the three main steps of interaction with the eukaryotic cells: bacterial adhesion to the target cells, injection of effector proteins into the cell cytosol by a type III secretion mechanism, and inhibition of several eukaryotic cell functions. During these different steps, the bacteria are in close contact with the target cell but they remain extracellular.

### 2.1.1 Adhesion to the Cell Surface

The first step in the process of eukaryotic cell shutdown and killing is initiated by the binding of bacteria to the cell membrane. This binding is mediated by a pYV-encoded outer membrane protein termed YadA (for *Yersinia* adhesin A) and by a chromosomally encoded invasin called Inv. Contact with the eukaryotic cell is a triggering event which activates *yop* gene transcription. Three pYV encoded proteins (YopN, LcrG, and TyeA) might be involved in the contact-induced control of Yop release. It is noteworthy that neither YadA nor Inv is produced in *Y. pestis* because of a point mutation or an IS insertion in their structural genes, respectively.

### 2.1.2 Assembly of the Ysc Secretion Apparatus

The *Yersinia* Ysc secretion apparatus forms a channel through which the effector Yop proteins cross the inner membrane, the peptidoglycan, and the outer membrane barriers in one step, most probably in a partially unfolded conformation. The secretion apparatus is composed of 29 proteins, ten of which (YscC, J, N, O, and Q-V) have homologues in other bacterial type III secretion systems. YscD and YscR-V seem to span the inner membrane, while YscC is a secretin which forms a ring-shaped pore with an internal diameter of  $\approx 50\text{\AA}$  in the outer membrane. Proper insertion of YscC in the outer membrane requires the lipoprotein YscW. YscN is a probable ATPase essential for the secretion process. The roles of the other Ysc proteins are not yet completely elucidated. Transcription of *ysc* genes is strongly thermoinduced at 37°C, probably by thermal modification of the structure of the chromatin, while the effect of the activator VirF seems to be dispensable.

### 2.1.3 Injection of the Effector Proteins

Translocation of the *Yersinia* effector proteins into the target cell cytosol requires their passage through the eukaryotic cell membrane. This passage appears to occur via a pore formed by a translocation apparatus composed of two hydrophobic translocators, YopB and YopD. LcrV, a Yop protein encoded on the same operon as YopB and YopD, is also required for efficient translocation of the effectors.

The genes encoding YopD and some other Yops (YopE, YopH, YopT, and YopN) are in the vicinity of loci that code for small acidic cytosolic chaperones designated Syc (for specific Yop chaperone). The roles of the Syc proteins are not completely understood and may not be the same for all Sycs. These small proteins may act as bodyguards to protect their cognate Yop from degradation, as secretion pilots to direct the Yops to the secretion apparatus, or as translocator-helpers to promote an efficient translocation of their partner Yops into the eukaryotic cytosol.

Yop transcription is activated by temperature (37°C) and VirF (*LcrF* in *Y. pestis*), a member of the AraC family of regulators. *virF*, itself is strongly thermoregulated; its product is active only at 37°C. The temperature-induced structural modifications of the chromatin characterized by an alteration of DNA supercoiling may render *yop* promoters more accessible to VirF and may dislodge a repressor such as the chromosomally encoded histone-like modulator YmoA. On the other hand, a feedback inhibition of *yop* genes transcription occurs when the secretion apparatus is closed and seems to be mediated by pYV-encoded repressors called YscM1 and YscM2 (*LcrQ* in *Y. pestis*).

### 2.1.4 Attack of the Target Cells

Six Yops (YopE, YopH, YopM, YopP, YpkA, and YopT) are injected by the type III secretion apparatus into the eukaryotic cytosol, where they have different effects. The target proteins for two effector Yops (YopM and YpkA) have not yet been identified. YopM is a leucine-rich protein located in the cell nucleus, while

YpkA (YopO in *Y. pestis*) is a serine-threonine kinase which is targeted to the inner surface of the plasma membrane and induces morphological alterations of the cell.

Three Yops (YopH, YopE, and YopT) exert their action on the cell cytoskeleton. YopH, a phosphotyrosine phosphatase having homologues in eukaryotes, dephosphorylates several eukaryotic proteins (p130<sup>cas</sup>, paxillin, and FAK). This Yop is specifically targeted to focal complexes, where it disrupts focal adhesion. YopE destroys actin filaments, probably by acting as a GTPase-activating protein for Rac-1 and Cdc42. YopT depolymerizes actin by modifying RhoA, a GTPase that acts on the formation of stress fibers. The result of the action of these three effector Yops is the inhibition of phagocytosis by macrophages and a reduced invasion of epithelial cells. In addition, YopH appears to act on the host-specific immune response by inhibiting the production of cytokines by T lymphocytes, and by decreasing surface expression of the co-stimulatory molecule B7.2 on activated B lymphocytes.

The last known effector, YopP (YopJ in *Y. pestis*), plays an essential role in down-regulating the inflammatory response of several cell types. This protein inhibits the activation of both the transcription activator NFκB and MAPK kinases. This results in decreased secretion of TNFα by macrophages and of IL8 by epithelial and endothelial cells, and in reduced presentation of ICAM1 and E-selectin on endothelial cells, which probably alter the recruitment of neutrophils at the site of infection. YopP has another important action, which is to induce apoptosis of macrophages through the caspase cascade.

## 2.2 The Fraction 1 Plasmid of *Y. pestis* (pFra)

This 101-kb plasmid is alternatively termed pFra (for plasmid-encoding fraction 1) and pMT1 (for plasmid-encoding murine toxin). The nucleotide sequence of pFra has recently been determined (Hu et al. 1998; LINDLER et al. 1998). This replicon has a G + C content (50.2%) close to that of the core genome but harbors some regions (essentially the murine toxin and the fraction 1 genes) with a much lower G + C content (38%–39%), suggesting a mosaic structure composed of different genetic elements acquired sequentially. Its origin of replication and partitioning functions resemble those of bacteriophages P1 and P7. pFra carries 115 potential coding regions, 38% of which have no significant homology to any protein present in the databases. The other genes harbored by this plasmid share various degrees of identity to known genes involved in protein and DNA metabolism, or regulation. Numerous mobility genes (integrase, transposases, resolvases, insertion sequences, and phage sequences) are carried by the plasmid. The best-characterized loci present on the pFra plasmids are those encoding the murine toxin and the fraction 1 antigen (F1 Ag).

F1 Ag forms a capsule-like structure surrounding the bacteria grown in vivo or in vitro at 37°C. The F1 Ag locus (*caf*) is composed of four genes which encode Caf1 (capsular subunits), Caf1M (a 28.7-kDa chaperone homologous to PapD), Caf1A (a 93.2-kDa protein homologous to PapC and involved in capsule

anchoring), and Caf1R (a 30-kDa activator of the AraC family). The capsular subunits assemble to form large polymers that can readily dissociate in vitro and most probably in vivo. This antigen has been widely used to develop diagnostic tests for plague and is included in several vaccine preparations currently under test. Although the presence of F1 Ag has been associated with resistance to phagocytosis by monocytes, mutational inactivation of its structural gene does not affect the virulence of the mutant in different animal models (DAVIS et al. 1996; DROZDOV et al. 1995), suggesting that F1 Ag is not a true virulence factor.

The second best characterized locus, *ynt*, codes for a so-called murine toxin which forms polymers of 240 and 120 kDa and may act as a  $\beta$ -adrenergic antagonist in rats and mice (MONTIE 1981). However, this protein is not toxic for other mammals such as guinea pigs, rabbits, and dogs, and its role in human infection has never been clearly established. The toxin is localized mainly in the bacterial cytoplasm and its synthesis is higher at 26°C than at 37°C (HINNEBUSCH et al. 2000). Recently, it was shown that this factor has a phospholipase D activity (HINNEBUSCH et al. 2000; RUDOLPH et al. 1999) and that its mutational inactivation does not affect the virulence of the mutant strain for mice. The main role of this protein appears to concern the flea, where it promotes colonization of the proventriculus of the insect (HINNEBUSCH et al. 2000), thus playing a role in the plague cycle by enhancing flea transmission of the bacterial agent.

pFra may spontaneously integrate into and excise from the chromosome (PROTSENKO et al. 1991). Some reports showing that a *Y. pestis* deprived of the entire pFra keeps an intact virulence potential for mice (FRIEDLANDER et al. 1995) or for African green monkeys (DAVIS et al. 1996), indicate that this plasmid may actually be more important for vector-borne plague transmission than for virulence.

### **2.3 The Plasminogen Activator Plasmid of *Y. pestis* (pPla)**

The plasmid encoding the plasminogen activator is alternatively termed pPla, pPCP1, or pPst. This 9.6-kb replicon, which is specific for *Y. pestis*, has a G+C content of 45.3% and a ColE1-like origin of replication (HU et al. 1998). It carries three loci: an IS100 insertion sequence present in multiple copies on the chromosome, a plasminogen activator gene (*pla*), and a region coding for a bacteriocin called pesticin (*psn*), and for its immunity protein (*pim*).

Pesticin is a 39.9-kDa bacteriocin with *N*-acetyl glucosaminidase activity which hydrolyzes the  $\beta$ -1,4 bond between *N*-acetylglucosamine and *N*-acetylmuramic acid in the glycan backbone of murein lipoprotein (FERBER and BRUBAKER 1979). It thus exerts its action on the peptidoglycan layer in the periplasmic space. This bacteriocin is active only on micro-organisms that carry the HPI, because the gene encoding the pesticin outer membrane receptor (designated *psn* in *Y. pestis*, and *fyuA* in *Y. enterocolitica*) is located on this chromosomal element (FETHERSTON et al. 1995; RAKIN et al. 1994). This bacteriocin may help the bacterium survive and compete with other organisms present in the same ecological niches, but it does not play a role in *Y. pestis* virulence (SODEINDE et al. 1992). Adjacent to the *pst* gene,

the *pim* gene encodes the pesticin immunity protein which protects the bacteria from the action of its own bacteriocin. The 14-kDa Pim protein is localized predominantly in the periplasmic space and may be associated with the outer membrane (RAKIN et al. 1996).

The plasminogen activator Pla protein is homologous to other proteins that activate plasminogen and has been involved in several activities. This outer membrane serine protease exhibits a proteolytic activity against complement component C3 and against several Yops during growth in culture medium, but the relevance of this activity *in vivo* remains unclear, as Yops are directly targeted into host cells and are not accessible to Pla degradation. The fibrinolytic activity of Pla predominates at 37°C, while the protein has a coagulase activity against rabbit plasma (but not human or mouse plasma) at 28°C. Pla also mediates weak binding to type IV collagen and promotes active and efficient invasion of HeLa cells by *Y. pestis* (COWAN et al. 2000). Although the *pla* gene was originally hypothesized to play a role in the blockage of the flea proventriculus, more recent studies have demonstrated that strains lacking pPla block fleas as efficiently as their wild-type counterparts (HINNEBUSCH et al. 1998). *pla* mutants of *Y. pestis* injected intravenously are fully virulent, while they are severely attenuated after subcutaneous infection (SODEINDE et al. 1992). It has been hypothesized that Pla may facilitate *Y. pestis* dissemination from its site of inoculation by cleaving fibrin deposits that trap the organisms and by reducing the chemoattraction of inflammatory cells, possibly by inhibiting IL-8 production (SODEINDE et al. 1992). Another study reported no difference in bacterial proliferation and local inflammatory response at the site of injection, but lower numbers of *pla* mutants at more distant sites, such as the popliteal lymph node and spleen (WELKOS et al. 1997). However, although some strains lacking pPla are severely attenuated (one million-fold) after subcutaneous inoculation, other isolates conserve their full virulence (KUTYREV et al. 1989; SAMOILOVA et al. 1996; WELKOS et al. 1997), suggesting that some factors may serve the function of Pla in these strains.

Thus, acquisition of pPla by a *Y. pestis* ancestor may have been important for the life cycle of the newly emerged pathogen, by facilitating its dissemination from the site of inoculation after the flea bite.

## 2.4 The pVM82 Plasmid of *Y. pseudotuberculosis*

A specific clinical form of *Y. pseudotuberculosis* infection is the so-called Far East scarlet-like fever (SOMOV 1976). This disease is encountered mainly in Siberia but has its counterpart, designated Izumi fever, in Japan (SATO et al. 1983). The onset of the infection, characterized by high fever, intestinal symptoms, and a hyperemia of the face, elbows and knees, is followed by a scarlet fever-like rash. A relapse occurs in half of the patients (MOLLARET et al. 1990). The *Y. pseudotuberculosis* strains that cause outbreaks of Far East scarlet-like fever possess the classical pYV plasmid and an additional plasmid termed pVM82 with a molecular mass of 82Mdal (SHUBIN et al. 1985). This plasmid appears to be a mosaic structure com-

posed of two parts: a genetic element of 57Mdal which may exist as an independent replicon (p57) in some strains, and a 25Mdal region which does not form a replicon alone but whose sequence may be at least partly found integrated into the chromosome of some isolates (SHOVADAEVA et al. 1991; SMIRNOV 1990). The specific clinical features of these strains seem to be associated with the 25-Mdal part of pVM82 rather than with the 57-Mdal portion. In contrast to p57, pVM82 is transferable to other strains of *Y. pseudotuberculosis* and to *E. coli* (SHOVADAEVA et al. 1990; SMIRNOV 1990). Acquisition of this conjugative additional plasmid by a *Y. pseudotuberculosis* clone from Far East Asia has thus conferred new pathogenic properties to the recipient bacteria.

## 2.5 The Antibiotic Resistance Plasmids of *Y. pestis*

A multiresistant strain of *Y. pestis* was isolated in 1995 in Madagascar, from a patient who presented with symptoms of bubonic plague (GALIMAND et al. 1997). Resistance involved not only all the antibiotics recommended for therapy (chloramphenicol, streptomycin, and tetracycline) and prophylaxis (sulfonamide and tetracycline) of plague, but also drugs that may have represented alternatives to classical therapy such as ampicillin, kanamycin, spectinomycin, and minocycline. The resistance determinants were carried by a 150-kb conjugative plasmid (pIP1202) which was closely related in structure to plasmid-borne determinants commonly found in enterobacteria and contained an Inc6-C origin of replication typical of broad-host-range plasmids of this group of bacteria.

More recently, a second *Y. pestis* strain which displayed high-level resistance to streptomycin alone was identified from another human case of bubonic plague in Madagascar (GUIYOULE et al. 2001). The resistance determinants were also carried by a self-transferable plasmid (pIP1203) but the plasmid (40kb), the streptomycin resistance determinants (adenylyl transferases), and the host bacterium (ribotype Q) were different from those associated with multiple antibiotic resistance. The streptomycin resistance genes in pIP1203 were inserted in R751, a broad-host-range plasmid belonging to the IncP $\beta$  group, and were part of the *tnpR-strA-strB-IR* cluster characteristic of the Tn5393 group of transposons.

In contrast to the other *Yersinia* plasmids described above, the two resistance plasmids were isolated from a single strain each. It is too early to know whether these plasmids will spread among the *Y. pestis* population or disappear. At least they demonstrate the capacity of *Y. pestis* to acquire resistance determinants in its natural habitat.

## 3 The High-Pathogenicity Island

Pathogenic *Yersinia* can be subdivided into low-pathogenicity strains, i.e., strains that induce a mild intestinal infection in human beings and are nonlethal for mice at

low doses, and high-pathogenicity strains which cause severe systemic infections in human beings and are mouse-lethal at low doses. One of the major differences between these two *Yersinia* pathotypes lies in the presence of a pathogenicity island termed “high-pathogenicity island” or HPI, because its presence correlates with the expression of a high-pathogenicity phenotype (CARNIEL et al. 1996). The island confers the ability to capture the iron molecules necessary for systemic dissemination of the bacteria in the host via a high-affinity iron-chelating system (e.g., siderophore) called yersiniabactin (HEESEMANN 1987). Mutations of several genes involved in yersiniabactin synthesis or transport (*irp2*, *fyuA/psh* or *ybtP*) demonstrate that these genes are important for expression of a high-pathogenicity phenotype (BEARDEN et al. 1997; CARNIEL et al. 1992; FETHERSTON et al. 1999; RAKIN et al. 1994).

The *Yersinia* HPI has all the criteria of a pathogenicity island (HACKER et al. 1997): (a) it is a large chromosomal DNA fragment (36kb in *Y. pseudotuberculosis* and *Y. pestis* and 43kb in *Y. enterocolitica*), (b) it carries genes essential for the expression of a high-pathogenicity phenotype, (c) it incorporates several repeated sequences (IS1328, IS1329, IS1222 and IS1400, or IS100) and a mobility gene (bacteriophage P4-like integrase gene), (d) it is bordered on one side by a *tRNA* locus (*asn tRNA*), and (e) the G + C content of the open reading frames composing the yersiniabactin locus (60%) is much higher than that of the core genome (46%–50%).

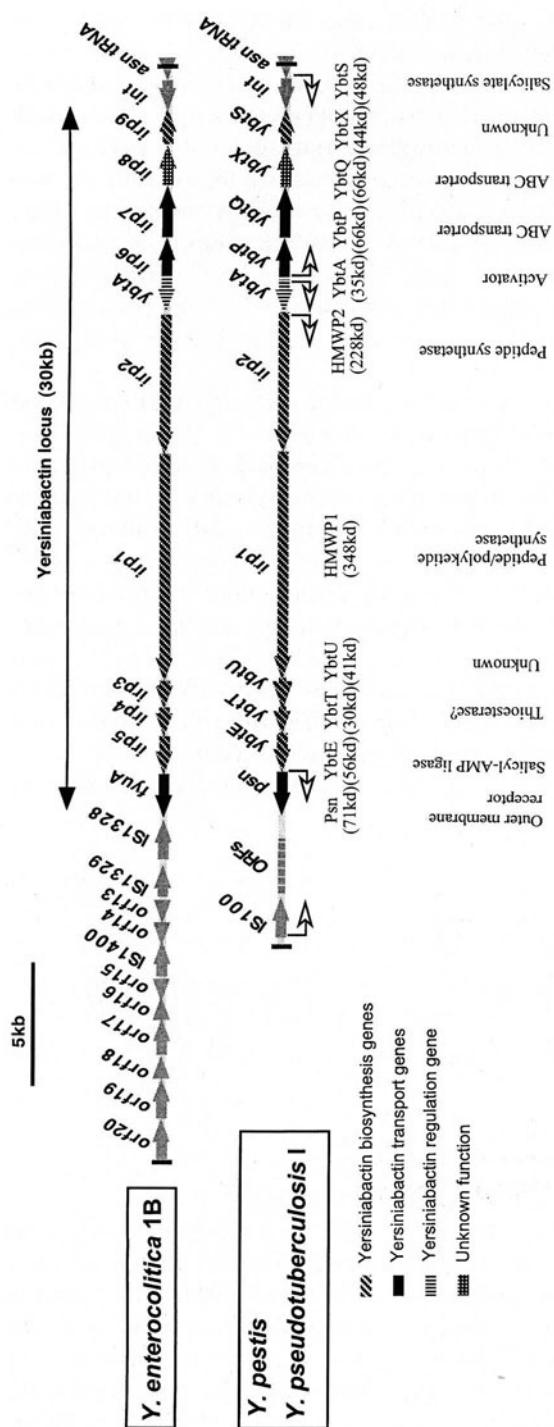
The HPI is present in the three pathogenic species of *Yersinia* but is restricted to certain subgroups of each species: serotypes I and III (with a 9-kb truncation at the left-hand part) of *Y. pseudotuberculosis* (BUCHRIESER et al. 1998a; RAKIN et al. 1995), biotype 1B of *Y. enterocolitica*, and all three biotypes of *Y. pestis* (DE ALMEIDA et al. 1993).

### 3.1 Genetic Organization of the HPI

The ≈30-kb right-hand part of the HPI, termed the yersiniabactin locus (Fig. 2), is highly conserved in the three pathogenic species *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica* while the left-hand part of the island is less conserved (CARNIEL 1999, 2001).

#### 3.1.1 The Yersiniabactin Locus

The siderophore yersiniabactin is a 482-dalton molecule which belongs to a small subgroup of phenolate siderophores and has an extremely high affinity for ferric iron ( $K_D = 4 \times 10^{-36}$ ) (GEHRING et al. 1998a; PERRY et al. 1999). The yersiniabactin molecule contains a phenolic group as well as one thiazolidine and two thiazoline rings (CHAMBERS et al. 1996; DRECHSEL et al. 1995; GEHRING et al. 1998a; HAAG et al. 1993). Its structure closely resembles those of pyochelin and anguibactin, produced by *Pseudomonas aeruginosa* and *Vibrio anguillarum*,



**Fig. 2.** Genetic organization of the HPI of the three highly pathogenic *Yersinia* species. The proven or putative function of the gene products and their sizes (in kilodaltons) are indicated below the maps. *White arrows* below the genetic map of *Y. pestis*/*Y. pseudotuberculosis* indicate the different operons. ORFs correspond to a cluster of short open reading frames that may be cryptic genes of bacteriophage origin

respectively. The structure of the yersiniabactin molecule is identical in *Y. enterocolitica* and *Y. pestis* (PERRY et al. 1999).

The locus involved in yersiniabactin-mediated iron-uptake is composed of 11 genes organized in four operons (Fig. 2). These genes can be roughly divided into three functional groups: yersiniabactin biosynthesis, transport into the bacterial cell (outer membrane receptor and transporters), and regulation. The correspondence between gene designations in *Y. pestis* and *Y. pseudotuberculosis* on the one hand and *Y. enterocolitica* 1B on the other, and the hypothetical or confirmed function for each gene are shown in Fig. 2.

### 3.1.1.1 The Yersiniabactin Biosynthesis Genes

The yersiniabactin machinery is extremely conserved in highly pathogenic *Yersinia* species, with more than 98% identity among the genes (BUCHRIESER et al. 1998a; GEHRING et al. 1998a; RAKIN et al. 1999). This genetic conservation is reflected by a functional conservation demonstrated by the fact that components of the yersiniabactin systems of the three species are interchangeable (CARNIEL et al. 1992; PERRY et al. 1999).

Formation of yersiniabactin occurs via a mixed polyketide synthase nonribosomal peptide synthetase strategy that assembles the siderophore in modular fashion from salicylate (GEHRING et al. 1998a,b; GUILVOUT et al. 1993; KEATING et al. 2000a,b; SUO et al. 1999). At least seven gene products are involved in yersiniabactin biosynthesis (Fig. 2): the high-molecular-weight proteins HMWP2 and HMWP1, YbtU, YbtT, YbtE, and YbtS (GEOFFROY et al. 2000).

### 3.1.1.2 The Yersiniabactin Transport Genes

Three genes are involved in yersiniabactin-mediated internalization of iron: *psn*/*fyuA*, *ybtP*, and *ybtQ* (Fig. 2). *Psn*/*FyuA* is a TonB-dependent outer membrane receptor for yersiniabactin, as well as for the bacteriocin pescicin (FETHERSTON et al. 1995; FETHERSTON and PERRY 1994; HEESEMANN et al. 1993; RAKIN et al. 1994). *YbtP* and *YbtQ* are inner-membrane permeases required for translocation of iron into the bacterial cytosol (FETHERSTON et al. 1999; GEHRING et al. 1998a). These two proteins are unique among the subfamily of ABC transporters of iron, since they both contain an amino-terminal membrane-spanning domain and a carboxy-terminal ATPase (FETHERSTON et al. 1999).

### 3.1.1.3 The Yersiniabactin Regulation Genes

Several systems regulate expression of the yersiniabactin locus. *YbtA*, which belongs to the AraC family of transcriptional regulators, activates expression from the *psn*, *irp2*, and *ybtP* promoters but represses expression of its own promoter (FETHERSTON et al. 1996). All four promoter regions (*psn*, *irp2*, *ybtA*, and *ybtP*) of the yersiniabactin region are negatively regulated by the Fur repressor in the presence of iron (GEHRING et al. 1998a; GUILVOUT et al. 1993; RAKIN et al. 1994).

Finally, there is evidence that yersiniabactin itself may up-regulate its own expression and that of *psn/fyuA* and *ybtPQXS* at the transcription level (BEARDEN et al. 1997; FETHERSTON et al. 1995, 1996, 1999; PELLUDAT et al. 1998; PERRY et al. 1999). It has been proposed that yersiniabactin functions as a signal molecule by binding to the YbtA activator, thus leading to transcriptional activation of other genes of the system and to transcriptional repression of *ybtA* (GEOFFROY et al. 2000).

### 3.1.2 The Right-Hand Extremity of the HPI

The region flanking the right-hand part of the yersiniabactin locus is also well conserved among the three pathogenic *Yersinia* species. This region contains a gene homologous to that of the integrase of bacteriophage P4 and an asparagine tRNA (*asn tRNA*) locus (BACH et al. 1999; BUCHRIESER et al. 1998a,b, 1999; CARNIEL et al. 1996; HARE et al. 1999; RAKIN et al. 1999). The P4-like *int* gene encodes a potentially functional protein in *Y. pestis* and *Y. pseudotuberculosis*, but the homologue in *Y. enterocolitica* contains a premature stop codon that results in a truncated peptide (BACH et al. 1999; RAKIN et al. 1999). The border of the HPI is formed by a 17-bp sequence homologous to the attachment site (*att*) of bacteriophage P4 and located at the 3' extremity of the *asn tRNA* locus. The *Y. pseudotuberculosis* HPI can insert into any of the three copies of the *asn tRNA* locus (BUCHRIESER et al. 1998a), while the HPI of *Y. enterocolitica* is inserted specifically into the *asnT* locus (CARNIEL et al. 1996; RAKIN et al. 1999) and that of *Y. pestis* into another copy of this locus (HARE et al. 1999).

### 3.1.3 The Left-Hand Region of the HPI

In contrast to the highly conserved 30-kb right-hand part of the HPI, the left-hand portion of the island is polymorphic among various strains and species of *Yersinia*. This left-hand region can be subdivided into two main groups: *Y. enterocolitica* on one hand, and *Y. pestis* and *Y. pseudotuberculosis* on the other hand (RAKIN et al. 1995).

In *Y. pestis* and *Y. pseudotuberculosis* I, the left-hand portion of the HPI is approximately 5kb long (Fig. 2). It carries one insertion sequence, designated IS100 (FETHERSTON and PERRY 1994; PODLADCHIKOVA et al. 1994; PRENTICE and CARNIEL 1995), and several short ORFs homologous to phage genes (BUCHRIESER et al. 1999; RAKIN et al. 1999). The left-hand part of the HPI terminates 248bp and 516bp downstream of the IS100 sequence in *Y. pseudotuberculosis* and *Y. pestis*, respectively (BUCHRIESER et al. 1998a, 1999; RAKIN et al. 1999). The larger size found in *Y. pestis* could be a scar of an IS630-like element insertion in this region (HARE et al. 1999). The border of the island is determined by a 17-bp sequence, repeated at the other extremity of the HPI (BUCHRIESER et al. 1998a, 1999; HARE et al. 1999).

The left-hand part of the *Y. enterocolitica* HPI extends 12.8kb downstream of the *fyuA* gene (Fig. 2) and is not well conserved among various *Y. enterocolitica* 1B strains (CARNIEL et al. 1996). It contains a cluster of four IS elements (IS1328, IS1329, a portion of IS1222 interrupted by IS1329, and IS1400), and seven other ORFs that have no significant homologies to any known genes (CARNIEL et al. 1996; RAKIN and HEESEMANN 1995; RAKIN et al. 1999, 2000). The left-hand extremity of the *Y. enterocolitica* HPI has a degenerate 17-bp repeat (BACH et al. 1999).

### 3.2 Mobility of the HPI

Precise excision of the HPI of *Y. pseudotuberculosis* I occurs spontaneously at a frequency of  $\approx 10^{-4}$  (BUCHRIESER et al. 1998a), generating a unique *att*-like sequence at the junction site, and restoring an intact *asn tRNA* locus. Excision of the *Y. pseudotuberculosis* island is most likely mediated by the HPI-encoded P4-like integrase and occurs at the *att*-like sites, probably by a mechanism of site-specific recombination. HPI insertion into different *asn tRNA* loci in individual colonies of the same isolate suggest that this element has retained the ability to excise and reintegrate into the chromosome (BUCHRIESER et al. 1998a).

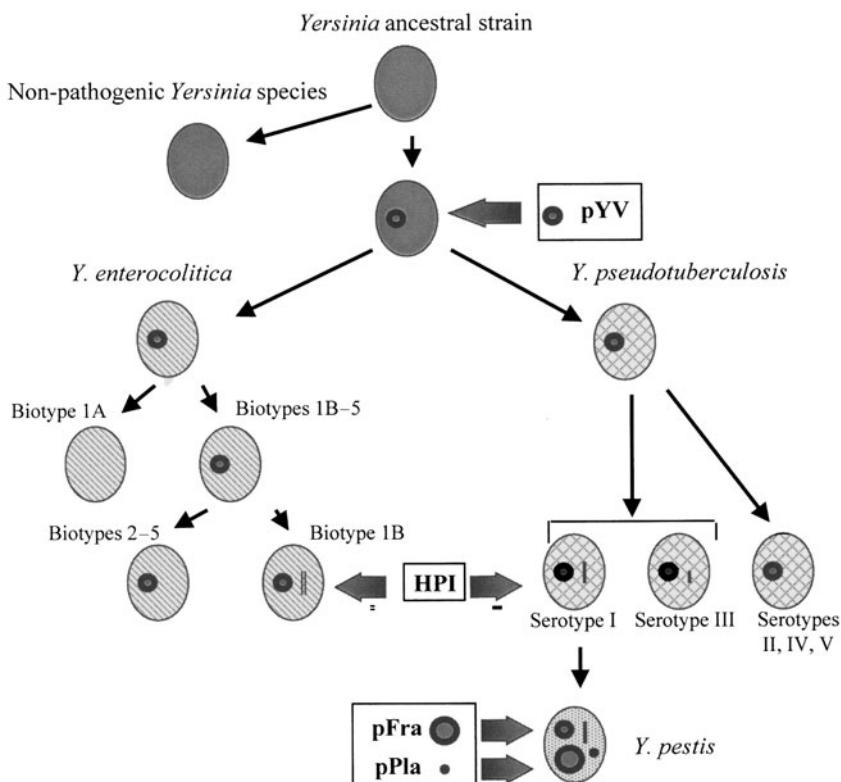
The excision of the *Y. pestis* HPI is not precise but occurs as part of a much larger chromosomal deletion of 102kb. This unstable region, designated “*pgm* locus” (FETHERSTON et al. 1992), encompasses most of the HPI extending from the IS100 copy to the right-hand border and extends further rightward over a  $\approx 68$ kb region termed pigmentation segment (BUCHRIESER et al. 1998b). Spontaneous excision of the 102-kb segment occurs at very high ( $2 \times 10^{-3}$ ) frequencies (HARE and McDONOUGH 1999), probably by homologous recombination between the two IS100 flanking copies (FETHERSTON and PERRY 1994; FETHERSTON et al. 1992). However, the HPI of *Y. pestis* has virtually the potential to excise precisely from the chromosome since it possesses a copy of the 17-bp repeat at each extremity, and an intact P4-like integrase gene (BUCHRIESER et al. 1999; HARE et al. 1999). The high deletion frequency of the 102-kb *pgm* locus may possibly mask the lower deletion frequency of the HPI alone.

The HPI of *Y. enterocolitica* 1B is the most stable (DE ALMEIDA et al. 1993). Although various deletions covering different portions of the HPI have been reported (BACH et al. 1999; RAKIN and HEESEMANN 1995), no precise excision of the island has ever been found. The stability of the HPI of *Y. enterocolitica* 1B may be explained by the absence of a functional integrase and of a conserved 17-bp direct repeat at the right-hand extremity of the island. Stabilization of the *Y. enterocolitica* HPI may represent an early step in a process of genetic fixation in the bacterial chromosome.

### 3.3 Identification of the HPI Among Other Enterobacteria

The HPI is the only pathogenicity island identified until now in different bacterial genera. After being discovered in *Yersinia* species, the island was subsequently found in other bacterial genera such as *E. coli*, *Enterobacter cloacae*, various species of *Klebsiella* and *Citrobacter* (BACH et al. 2000; KARCH et al. 1999; SCHUBERT et al. 1998, 2000; J. Hacker, personal communication), and non-I serotypes of *Salmonella enterica* (T. Öschläger et al., in preparation). In *E. coli*, the HPI is found in a wide variety of pathotypes: EAEC, EPEC, EIEC, ETEC, EHEC/STEC (where it is specific for certain serotypes such as O26:H11/H- and O128:H2/H-), UPEC, and septicemic and meningitis *E. coli* (KARCH et al. 1999; SCHUBERT et al. 1998).

The HPIs of *E. coli* and *Yersinia* are highly conserved, as demonstrated by the sequences of portions of *irp2* and *fyuA* that are more than 98% identical in the two genera (RAKIN et al. 1995; SCHUBERT et al. 1998, 1999). The *E. coli* HPI is more closely related to the island of *Y. pestis* and *Y. pseudotuberculosis* than to that of *Y. enterocolitica* (RAKIN et al. 1995). The overall organization of the HPI is also



**Fig. 3.** Hypothetical chronology of horizontal acquisition of resident foreign DNA by pathogenic *Yersinia*. The lengths of the arrows do not represent the time of evolution

well conserved among the different HPI-positive enterobacteria, although minor differences at the extremities of the island may be noted (BACH et al. 2000). An intact P4-like *int* gene is commonly present and shares 94% (STEC)–99% (UPEC) identity with the *Y. pestis* homologue (KARCH et al. 1999; SCHUBERT et al. 1999), although partial or total deletion of *int* has been reported in a few enterobacterial isolates (BACH et al. 2000; KARCH et al. 1999). With very few exceptions (BACH et al. 2000), the enterobacterial HPI is integrated into the *asnT* locus, which contains a perfectly conserved 17-bp *att*-like site. The left-hand boundary of the HPI is more variable among enterobacteria. Some *E. coli* strains that contain an *irp2* locus lack the *fyuA* gene, the IS100 element and/or the 17-bp direct repeat that forms the boundary of the HPI (BACH et al. 2000; KARCH et al. 1999; SCHUBERT et al. 1998, 1999), a situation reminiscent of the HPI of *Y. pseudotuberculosis* III and suggestive of a process of stabilization of the island on the chromosome of these enterobacteria. Furthermore, some STEC isolates that do harbor the *fyuA* locus still do not synthesize the protein (KARCH et al. 1999), indicating that FuyA may be dispensable under certain circumstances.

## 4 Conclusion

The various horizontally acquired genetic elements described in this chapter are not essential for bacterial survival. Nonpathogenic *Yersinia* species do not possess any of these elements but can still survive and multiply in the environment or in a mammalian host as saprophytes. Acquisition of these alien DNAs, although not required for survival, was probably a crucial step in the emergence of pathogenic variants of *Yersinia* and of novel species (*Y. pestis*) with new modes of transmission.

Strains harboring the pYV became able to circumvent the innate and specific immune responses of the host, and therefore to multiply in vivo and cause clinical symptoms. Since the pYV is present and very well conserved in three *Yersinia* species, it may be assumed that it was acquired by an ancestral strain of *Yersinia* (Fig. 3), prior to its divergence into *Y. enterocolitica* and *Y. pseudotuberculosis* 0.4–1.9 million years ago (ACHTMAN et al. 1999). Biotype 1A of *Y. enterocolitica* would have subsequently lost this replicon and returned to a nonpathogenic status. The global distribution of pYV-harboring strains also argues for an ancestral acquisition of this replicon. In contrast, the pVM82 plasmid is found in only some strains of *Y. pseudotuberculosis* from a specific and localized geographical area, suggesting its recent acquisition by a specific clone in Far East Asia. Presence of this plasmid conferred particular properties which rendered the recipient bacterium able to cause a new type of disease in human hosts. Even more localized is the acquisition of antibiotic resistance plasmids by *Y. pestis*. The isolation of these replicons from a single strain each is suggestive of their extremely recent acquisition.

The foreign chromosomal HPI is present in the three pathogenic species of *Yersinia* but, as opposed to pYV, its presence is restricted to certain subgroups of

*Y. enterocolitica* and *Y. pseudotuberculosis*, and this element is not entirely conserved in these two species. This is suggestive of an independent acquisition of two slightly different HPIs by subgroups of the two species after their divergence (Fig. 3), a larger one by *Y. enterocolitica* 1B and a smaller one by *Y. pseudotuberculosis* I and other Enterobacteria. The truncated HPI found in serotype III strains of *Y. pseudotuberculosis* either may have been acquired independently from the complete island present in serotype I strains, or may correspond to a partial deletion of the HPI in an ancestral strain of serotype III. The fact that the percentage of identity of HPI-borne genes among these various bacterial species is much higher than that of their core genome argues for a recent acquisition of this chromosomal element. *Y. pestis* has recently been shown to be a clone of *Y. pseudotuberculosis* that emerged 1,500–20,000 years ago (ACHTMAN et al. 1999). The presence of an HPI identical to that found in *Y. pseudotuberculosis* strains of serotype I suggests that *Y. pestis* arose from a strain of this serotype (Fig. 3). Since fleas ingurgitate *Y. pestis* during their blood meal, the HPI, by promoting systemic dissemination of the bacteria, was certainly determinant in the process of *Y. pestis* transmission by an insect. Sequential acquisitions of pFra and pPla were other essential steps in the vector-borne transmission of *Y. pestis*, the former by promoting colonization of the flea gut and the latter by facilitating dissemination of the micro-organisms from the site of inoculation after the flea bite.

During the evolution of *Yersinia* and their differentiation into various species, different events most likely took place in their genome, including point mutations, insertions, deletions, and translocation. However, in contrast to these gradual changes in the bacterial genome, acquisition of large pieces of foreign DNA (plasmids, pathogenicity islands) led to an evolution in “quantum leaps” (GROISMAN and OCHMAN 1996), with rapid and dramatic modifications in bacterial metabolism, resistance to environmental stresses, pathogenicity, or transmission properties.

## References

- Achtman M, Zurth K, Morelli G, Torrea G, Guiyoute A, Carniel E (1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc Natl Acad Sci USA 96:14043–14048
- Bach S, Buchrieser C, Prentice M, Guiyoute A, Msadek T, Carniel E (1999) The high-pathogenicity island of *Yersinia enterocolitica* Ye8081 undergoes low-frequency deletion but not precise excision, suggesting recent stabilization in the genome. Infect Immun 67:5091–5099
- Bach S, de Almeida A, Carniel E (2000) The *Yersinia* high-pathogenicity island is present in different members of the family *Enterobacteriaceae*. FEMS Microbiol Lett 183:289–294
- Bearden SW, Fetherston JD, Perry RD (1997) Genetic organization of the yersiniabactin biosynthetic region and construction of avirulent mutants in *Yersinia pestis*. Infect Immun 65:1659–1668
- Bleves S, Cornelis GR (2000) How to survive in the host: the *Yersinia* lesson. Microb Infect 2:1451–1460
- Blocker A, Holden D, Cornelis G (2000) Type III secretion systems: what is the translocator and what is translocated? Cell Microbiol 2:387–390

- Buchrieser C, Brosch R, Bach S, Guiyoule A, Carniel E (1998a) The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal *asn tRNA* genes. *Mol Microbiol* 30:965–978
- Buchrieser C, Prentice M, Carniel E (1998b) The 102-kb unstable region of *Yersinia pestis* comprises a high-pathogenicity island linked to a pigmentation segment which undergoes internal rearrangement. *J Bacteriol* 180:2321–2329
- Buchrieser C, Rusniok C, Frangeul L, Couve E, Billault A, Kunst F, Carniel E, Glaser P (1999) The 102-kb *pgm* locus of *Yersinia pestis*: sequence analysis and comparison of selected regions among different *Yersinia pestis* and *Yersinia pseudotuberculosis* strains. *Infect Immun* 67:4851–4861
- Carniel E (1999) The *Yersinia* high-pathogenicity island. *Int Microbiol* 2:161–167
- Carniel E (2001) The *Yersinia* high-pathogenicity island: an iron-uptake island. *Microb Infect* (in press)
- Carniel E, Guiyoule A, Guilvout I, Mercereau-Puijalon O (1992) Molecular cloning, iron-regulation and mutagenesis of the *irp2* gene encoding HMWP2, a protein specific for the highly pathogenic *Yersinia*. *Mol Microbiol* 6:379–388
- Carniel E, Guilvout I, Prentice M (1996) Characterization of a large chromosomal “high-pathogenicity island” in biotype 1B *Yersinia enterocolitica*. *J Bacteriol* 178:6743–6751
- Chambers CE, McIntyre DD, Mouck M, Sokol PA (1996) Physical and structural characterization of yersiniophore, a siderophore produced by clinical isolates of *Yersinia enterocolitica*. *Biometals* 9:157–167
- Cornelis GR (1998) The *Yersinia* deadly kiss. *J Bacteriol* 180:5495–5504
- Cornelis GR (2000) Molecular and cell biology aspects of plague. *Proc Natl Acad Sci USA* 97:8778–8783
- Cornelis GR, Wolf-Watz H (1997) The *Yersinia* Yop virulon: a bacterial system for subverting eukaryotic cells. *Mol Microbiol* 23:861–867
- Cornelis GR, Van Gijsegem F (2000) Assembly and function of type III secretory systems. *Ann Rev Microbiol* 54:735–774
- Cowan C, Jones HA, Kaya YH, Perry RD, Straley SC (2000) Invasion of epithelial cells by *Yersinia pestis*: evidence for a *Y. pestis*-specific invasin. *Infect Immun* 68:4523–4530
- Davis KJ, Fritz DL, Pitt ML, Welkos SL, Worsham PL, Friedlander AM (1996) Pathology of experimental pneumonic plague produced by fraction 1-positive and fraction 1-negative *Yersinia pestis* in African green monkeys (*Cercopithecus aethiops*). *Arch Pathol Lab Med* 120:156–163
- De Almeida AMP, Guiyoule A, Guilvout I, Iteman I, Baranton G, Carniel E (1993) Chromosomal *irp2* gene in *Yersinia*: distribution, expression, deletion and impact on virulence. *Microb Pathog* 14:9–21
- Drechsel H, Stephan H, Lotz R, Haag H, Zahner H, Hantke K, Jung G (1995) Structure elucidation of yersiniabactin, a siderophore from highly virulent *Yersinia* strains. In: Liebigs Ann. VCH Publishers Inc., 303 NW 12TH Ave, Deerfield Beach, FL 33442–1788, pp 1727–1733
- Drozdov IG, Anisimov AP, Samoilova SV, Yezhov IN, Yeremin SA, Karlyshev AV, Krasilnikova VM, Kravchenko VI (1995) Virulent non-capsulate *Yersinia pestis* variants constructed by insertion mutagenesis. *J Med Microbiol* 42:264–268
- Fällman M, Persson C, Schesser K, Wolf-Watz H (1998) Bidirectional signaling between *Yersinia* and its target cell. *Folia Microbiol (Praha)* 43:263–273
- Ferber DM, Brubaker RR (1979) Mode of action of pecticin: *N*-acetylglucosaminidase activity. *J Bacteriol* 139:495–501
- Fetherston JD, Perry RD (1994) The pigmentation locus of *Yersinia pestis* KIM6<sup>+</sup> is flanked by an insertion sequence and includes the structural genes for pecticin sensitivity and HMWP2. *Mol Microbiol* 13:697–708
- Fetherston JD, Schuetze P, Perry RD (1992) Loss of the pigmentation phenotype in *Yersinia pestis* is due to the spontaneous deletion of 102kb of chromosomal DNA which is flanked by a repetitive element. *Mol Microbiol* 6:2693–2704
- Fetherston JD, Lillard J, Perry RD (1995) Analysis of the pecticin receptor from *Yersinia pestis*: role in iron-deficient growth and possible regulation by its siderophore. *J Bacteriol* 177:1824–1833
- Fetherston JD, Bearden SW, Perry RD (1996) YbtA, an AraC-type regulator of the *Yersinia pestis* pecticin/yersiniabactin receptor. *Mol Microbiol* 22:315–325
- Fetherston JD, Bertolini VJ, Perry RD (1999) YbtP and YbtQ: two ABC transporters required for iron uptake in *Yersinia pestis*. *Mol Microbiol* 32:289–299
- Friedlander AM, Welkos SL, Worsham PL, Andrews GP, Heath DG, Anderson GW, Pitt MLM, Estep J, Davis K (1995) Relationship between virulence and immunity as revealed in recent studies of the F1 capsule of *Yersinia pestis*. *Clin Infect Dis* 21:S178–S181
- Galán JE, Bliska JB (1996) Cross-talk between bacterial pathogens and their host cells. *Annu Rev Cell Dev Biol* 12:221–255

- Galimand M, Guiyoule A, Gerbaud G, Rasoamanana B, Chanteau S, Carniel E, Courvalin P (1997) Multidrug resistance in *Yersinia pestis* mediated by a transferable plasmid. *N Engl J Med* 337:677–680
- Gehring AM, Demoll E, Fetherston JD, Mori I, Mayhew GF, Blattner FR, Walsh CT, Perry RD (1998a) Iron acquisition in plague – modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem Biol* 5:573–586
- Gehring AM, Mori I, Perry RD, Walsh CT (1998b) The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry* 37:11637–11650
- Geoffroy VA, Fetherston JD, Perry RD (2000) *Yersinia pestis* YbtU and YbtT are involved in synthesis of the siderophore yersiniabactin but have different effects on regulation. *Infect Immun* 68:4452–4461
- Groisman EA, Ochman H (1996) Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* 87:791–794
- Guilvout I, Mercereau-Puijalon O, Bonnefoy S, Pugsley AP, Carniel E (1993) High-molecular-weight protein 2 of *Yersinia enterocolitica* is homologous to AngR of *Vibrio anguillarum* and belongs to a family of proteins involved in nonribosomal peptide synthesis. *J Bacteriol* 175:5488–5504
- Guiyoule A, Gerbaud G, Buchrieser C, Galimand M, Rahalison L, Chanteau S, Courvalin P, Carniel E (2001) Transferable plasmid-mediated resistance to streptomycin in a clinical isolate of *Yersinia pestis*. *Emerg Infect Dis* 7:43–48
- Haag H, Hantke K, Drechsel H, Stojiljkovic I, Jung G, Zahner H (1993) Purification of yersiniabactin – a siderophore and possible virulence factor of *Yersinia enterocolitica*. *J Gen Microbiol* 139:2159–2165
- Hacker J, Blum-Oehler G, Mühldorfer I, Tschaep H (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* 23:1089–1097
- Hare JM, McDonough KA (1999) High-frequency RecA-dependent and -independent mechanisms of Congo red binding mutations in *Yersinia pestis*. *J Bacteriol* 181:4896–4904
- Hare JM, Wagner AK, McDonough KA (1999) Independent acquisition and insertion into different chromosomal locations of the same pathogenicity island in *Yersinia pestis* and *Yersinia pseudotuberculosis*. *Mol Microbiol* 31:291–303
- Heesemann J (1987) Chromosomal-encoded siderophores are required for mouse virulence of enteropathogenic *Yersinia* species. *FEMS Microbiol Lett* 48:229–233
- Heesemann J, Hantke K, Vocke T, Saken E, Rakine A, Stojiljkovic I, Berner R (1993) Virulence of *Yersinia enterocolitica* is closely associated with siderophore production, expression of an iron-repressible outer membrane polypeptide of 65000 Da and pesticin sensitivity. *Mol Microbiol* 8:397–408
- Hinnebusch BJ, Fischer ER, Schwan TG (1998) Evaluation of the role of the *Yersinia pestis* plasminogen activator and other plasmid-encoded factors in temperature-dependent blockage of the flea. *J Infect Dis* 178:1406–1415
- Hinnebusch J, Cherepanov P, Du Y, Rudolph A, Dixon JD, Schwan T, Forsberg A (2000) Murine toxin of *Yersinia pestis* shows phospholipase D activity but is not required for virulence in mice. *Int J Med Microbiol* 290:483–487
- Hu P, Elliott J, McCready P, Skowronski E, Garnes J, Kobayashi A, Brubaker RR, Garcia E (1998) Structural organization of virulence-associated plasmids of *Yersinia pestis*. *J Bacteriol* 180:5192–5202
- Karch H, Schubert S, Zhang D, Zhang W, Schmidt H, Ölschläger T, Hacker J (1999) A genomic island, termed high-pathogenicity island, is present in certain non-O157 shiga toxin-producing *Escherichia coli* clonal lineages. *Infect Immun* 67:5994–6001
- Keating TA, Miller DA, Walsh CT (2000a) Expression, purification, and characterization of HMWP2, a 229-kDa, six domain protein subunit of yersiniabactin synthetase. *Biochemistry* 39:4729–4739
- Keating TA, Suo ZC, Ehmann DE, Walsh CT (2000b) Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a hole-carrier protein domain. *Biochemistry* 39:2297–2306
- Kutyrev VV, Filippov AA, Shavina N, Protsenko OA (1989) Genetic analysis and simulation of the virulence of *Yersinia pestis*. *Mol Gen Mikrobiol Virusol*, pp 42–47
- Lindler LE, Plano GV, Burland V, Mayhew GF, Blattner FR (1998) Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIM5 plasmid encoding murine toxin and capsular antigen. *Infect Immun* 66:5731–5742
- Mollaret HH, Carniel E, Guilvout I (1990) La fièvre scarlatiniforme d'Extrême-Orient. *Med Mal Infect* 20:519–529
- Montie TC (1981) Properties and pharmacological action of plague murine toxin. *Pharmacol Ther* 12:491–499

- Pelludat C, Rakin A, Jacobi CA, Schubert S, Heesemann J (1998) The yersiniabactin biosynthetic gene cluster of *Yersinia enterocolitica*: organization and siderophore-dependent regulation. *J Bacteriol* 180:538–546
- Perry RD, Straley SC, Fetherston JD, Rose DJ, Gregor J, Blattner FR (1998) DNA sequencing and analysis of the low-Ca<sup>2+</sup> response plasmid pCD1 of *Yersinia pestis* KIM5. *Infect Immun* 66:4611–4623
- Perry RD, Balbo PB, Jones HA, Fetherston JD, DeMoll E (1999) Yersiniabactin from *Yersinia pestis*: biochemical characterization of the siderophore and its role in iron transport and regulation. *Microbiology* 145:1181–1190
- Podladchikova ON, Dikhanov GG, Rakin AV, Heesemann J (1994) Nucleotide sequence and structural organization of *Yersinia pestis* insertion sequence IS100. *FEMS Microbiol Lett* 121:269–274
- Prentice MB, Carniel E (1995) Sequence analysis of a *Yersinia pestis* insertion sequence associated with an unstable region of the chromosome. *Contr Microbiol Immunol* 13:294–298
- Protsenko OA, Filippov AA, Kutyrev VV (1991) Integration of the plasmid encoding the synthesis of capsular antigen and murine toxin into *Yersinia pestis* chromosome. *Microb Pathog* 11:123–128
- Rakin A, Heesemann J (1995) Virulence-associated *fyuA/irp2* gene cluster of *Yersinia enterocolitica* biotype 1B carries a novel insertion sequence IS1328. *FEMS Microbiol Lett* 129:287–292
- Rakin A, Saken E, Harmsen D, Heesemann J (1994) The pesticin receptor of *Yersinia enterocolitica*: a novel virulence factor with dual function. *Mol Microbiol* 13:253–263
- Rakin A, Urbitsch P, Heesemann J (1995) Evidence for two evolutionary lineages of highly pathogenic *Yersinia* species. *J Bacteriol* 177:2292–2298
- Rakin A, Boolgakowa E, Heesemann J (1996) Structural and functional organization of the *Yersinia pestis* bacteriocin pesticin gene cluster. *Microbiology* 142:3415–3424
- Rakin A, Noelting C, Schubert S, Heesemann J (1999) Common and specific characteristics of the high-pathogenicity island of *Yersinia enterocolitica*. *Infect Immun* 67:5265–5274
- Rakin A, Schubert S, Guiolvout I, Carniel E, Heesemann J (2000) Local hopping of IS3 elements into the A + T-rich part of the high-pathogenicity island in *Yersinia enterocolitica* 1B, O: 8. *FEMS Microbiol Lett* 182:225–229
- Rudolph AE, Stuckey JA, Zhao Y, Matthews HR, Patton WA, Moss J, Dixon JE (1999) Expression, characterization, and mutagenesis of the *Yersinia pestis* murine toxin, a phospholipase D superfamily member. *J Biol Chem* 274:11824–11831
- Samoilova SV, Samoilova LV, Yezhov IN, Drozdov IG, Anisimov AP (1996) Virulence of pPst(+) and pPst(−) strains of *Yersinia pestis* for guinea-pigs. *J Med Microbiol* 45:440–444
- Sato K, Ouchi K, Taki M (1983) *Yersinia pseudotuberculosis* infection in children, resembling Izumi fever and Kawazaki syndrome. *Pediatr Infect Dis J* 2:123–126
- Schubert S, Rakin A, Karch H, Carniel E, Heesemann J (1998) Prevalence of the “high-pathogenicity island” of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infect Immun* 66:480–485
- Schubert S, Rakin A, Fischer D, Sorsa J, Heesemann J (1999) Characterization of the integration site of *Yersinia* high-pathogenicity island in *Escherichia coli*. *FEMS Microbiol Lett* 179:409–414
- Schubert S, Cuenca S, Fischer D, Heesemann J (2000) High-pathogenicity island of *Yersinia pestis* in enterobacteriaceae isolated from blood cultures and urine samples: prevalence and functional expression. *J Infect Dis* 182:1268–1271
- Shovadaeva GA, Markov AP, Pokrovskaya MS, Chetina EV, Yanishevsky NV, Shubin FN, Gunzburg AL, Smirnov GB (1990) Structural and functional characteristics of *Yersinia pseudotuberculosis* pVM82. *Genetika* 26:621–629
- Shovadaeva GA, Shubin FN, Shaginian IA, Markov AP, Gintzburg AL, Smirnov GB (1991) Alternative localization of the determinant of pathogenicity coded by the *Yersinia pseudotuberculosis* pVM82 plasmid. *Molek Genet Mikrobiol Virusol*, pp 23–27
- Shubin FN, Sibirtsev YT, Rasskazov VA (1985) *Yersinia pseudotuberculosis* plasmids and their role in the realization of the epidemic process in pseudotuberculosis. *Zh Mikrob Epidemiol Immunol* 12:53–56
- Smirnov GB (1990) Molecular biology of the factors responsible for *Yersinia* virulence. *Biomed Sci* 1: 223–232
- Sodeinde OA, Subrahmanyam YVK, Stark K, Quan T, Bao YD, Goguen JD (1992) A surface protease and the invasive character of plague. *Science* 258:1004–1007
- Somov GP (1976) Far Eastern scarlatina-like fever (human pseudotuberculosis) (in Russian). *Zh Mikrob Epidemiol Immunol* 4:103–108
- Straley SC (1988) The plasmid-encoded outer-membrane proteins of *Yersinia pestis*. *Rev Infect Dis* 10:S323–S326

- Straley SC, Perry RD (1995) Environmental modulation of gene expression and pathogenesis in *Yersinia*. Trends Microbiol 3:310–317
- Suo ZC, Walsh CT, Miller DA (1999) Tandem heterocyclization activity of the multidomain 230-kDa HMWP2 subunit of *Yersinia pestis* yersiniabactin synthetase: Interaction of the 1–1382 and 1383–2035 fragments. Biochemistry 38:14023–14035
- Welkos SL, Friedlander AM, Davis KJ (1997) Studies on the role of plasminogen activator in systemic infection by virulent *Yersinia pestis* strain C092. Microb Pathog 23:211–223

# Genomic Organization of LPS-Specific Loci

P.P. REEVES and L. WANG

1	Introduction . . . . .	109
1.1	Nomenclature for <i>Escherichia</i> and <i>Salmonella</i> . . . . .	110
1.2	Nomenclature for LPS Biosynthetic Genes . . . . .	111
2	Overview of LPS Structure, Biosynthesis and Genetics . . . . .	111
2.1	LPS Structure . . . . .	111
2.2	Biosynthesis and Genetics of LPS . . . . .	114
3	O Antigen Genes in <i>E. coli</i> and <i>S. enterica</i> . . . . .	114
3.1	O Antigen Variation . . . . .	114
3.2	O-Antigen Gene Cluster . . . . .	115
3.3	Intraspecies Movement of O Antigen Genes . . . . .	118
3.4	Interspecies Movement of O Antigen Genes . . . . .	119
3.5	Origins of the Polymorphism in the O Antigen Locus . . . . .	120
3.6	The O-Antigen Gene Cluster as a PAI . . . . .	121
4	Lipid A/Core Genes of <i>E. coli</i> and <i>S. enterica</i> . . . . .	122
4.1	Genes for the Synthesis of Lipid A/Core Precursors . . . . .	123
4.2	(Kdo) <sub>2</sub> -Lipid A Synthesis Genes in K-12 and LT2 . . . . .	123
4.3	LPS Core Synthesis Genes in K-12 . . . . .	124
4.4	Variation in LPS Core Synthesis Genes . . . . .	125
4.5	Summary of Function, Location and G+C Content of Lipid A/Core Genes . . . . .	125
5	LPS Genes of Species Other than <i>E. coli</i> and <i>S. enterica</i> . . . . .	126
5.1	O Antigen Genes in <i>Yersinia</i> spp. . . . .	127
5.2	IS1358 and the Origins of the Ogawa Form of the <i>Vibrio cholerae</i> O1 Antigen . . . . .	129
5.3	LPS Core Genes of <i>Yersinia pestis</i> and <i>Yersinia pseudotuberculosis</i> . . . . .	129
6	Conclusions . . . . .	130
	References . . . . .	130

## 1 Introduction

Lipopolysaccharide (LPS) genes have many of the characteristics of PAIs but also differ in significant ways. Lipopolysaccharide differs from the products of most PAI genes in that it is an essential component of the cell, and mutants totally lacking LPS are not found. However parts of the molecule such as the O antigen are not

essential for growth in laboratory media, but in general are required for survival in the natural environment. For example, in many pathogens complete LPS is required for survival in the host, and in this these discretionary components do resemble the products of PAI genes. Many of the genes for these discretionary components occur in discrete clusters and these gene clusters share many properties with PAIs.

LPS is a key component of the outer membrane (OM) of gram-negative bacteria. It comprises three distinct regions: Lipid A, an oligosaccharide core and, commonly, a polysaccharide O antigen. Lipid A is embedded in the OM, in which LPS is the major amphipathic molecule in the outer leaflet. The role of LPS in the structure of the OM probably accounts for its being essential for the survival of bacteria. However, some clones and even species of bacteria lack O antigen, and many isolates lose the ability to make O antigen on cultivation. In other species a short oligosaccharide replaces the O polysaccharide and for these the complete molecule is commonly referred to as lipo-oligosaccharide, although we include them in this discussion of LPS.

Lipid A is the most conserved part of LPS. The core oligosaccharide links the Lipid A to the O antigen. Diversity in the core structure is limited, with five different structures found in *Escherichia coli*. The O antigen is highly polymorphic, with 190 serologically distinguished forms in *E. coli*. It has been observed that some O antigen forms are disproportionately represented in pathogenic clones and concluded that specificity of O antigen is important in determining pathogenicity. The gene clusters for these forms resemble PAIs in their distribution. Furthermore, there is evidence of association between O polysaccharide chain length and pathogenicity in *E. coli*.

In this chapter we will look at the genetics and evolution of O antigen, Lipid A and core genes, and at similarities to conventional PAIs. Some of these genes occur in PAI-like clusters and others are scattered in small operons. They are known from a range of species, and this, plus quite detailed knowledge of function, allows comparison and generalization on pattern of gene distribution. First, we must look briefly at nomenclature for both species and genes.

## 1.1 Nomenclature for *Escherichia* and *Salmonella*

We will be looking at variation within and between species and at the role of lateral transfer. In this context it is most important to be consistent in our concept of the species. This requires particular attention with *Escherichia* and *Salmonella*, which we will use for most of our examples. Sequence variation in *Salmonella* is that expected for a species, and the use of *Salmonella enterica* as the name for all serovars, with the possible exception of subspecies V, is now widely accepted (LE MINOR and POPOFF 1987; POPOFF and LE MINOR 1997), although 2312 serovars have been given full species status and that terminology is also in use; we use the single species name *S. enterica* and use the old species names as serovar names, which reflects current views on bacterial species and is consistent with usage in

*E. coli*. Shigella strains have sequence similarity to *E. coli* at a level which puts them clearly within the same species (PUPO et al. 1997, 2000). Further, they do not form a single discrete set of strains within *E. coli* (OCHMAN et al. 1983; PUPO et al. 1997, 2000). Indeed, sequence data for eight housekeeping genes shows that most Shigella strains fall into three groups, although outliers include such significant Shigella forms as Dysenteriae 1 and Sonnei. This raises the same problem faced above with *Salmonella* and we refer to them all as Shigella, Dysenteriae, or Sonnei etc. strains of *E. coli*. Boydii 13 is the exception and is probably the sole representative of another species. We will often use *E. coli* strain K-12 and *S. enterica* strain LT2 as examples and refer to them simply as K-12 and LT2.

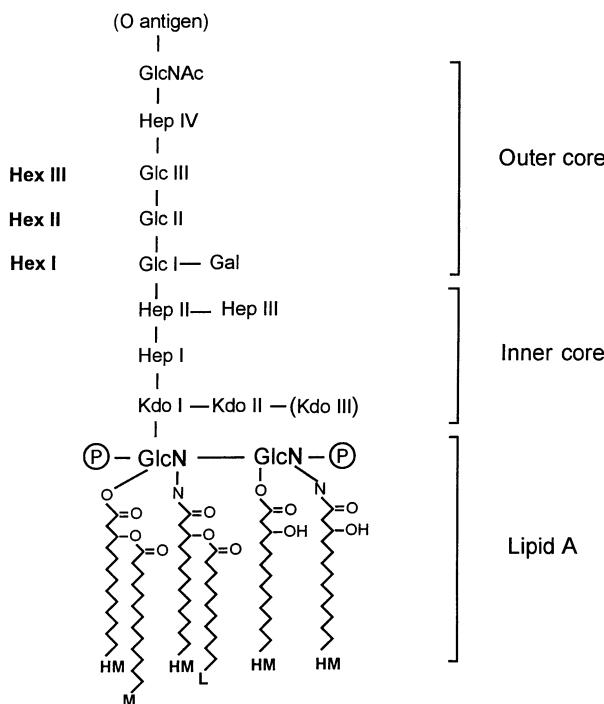
## 1.2 Nomenclature for LPS Biosynthetic Genes

LPS biosynthesis genes were originally given *rf\*\** names, but this system cannot cope with the large number of LPS genes now identified. We are using instead the Bacterial Polysaccharide Gene Nomenclature (BPGN) system, designed by researchers in the area (REEVES et al. 1996a,b). This system allows each functionally distinctive gene to have a unique name, genes in different clusters but with the same function having the same name, generally starting with "w". However, pathway genes have names that relate to the specific pathway, examples being *manABC* for the three genes of the GDP-mannose pathway and *ddhABCD* for the dTDP-dideoxyhexose pathway genes, both in function order. The older-style names were used in earlier publications for many of the examples we quote, and we refer readers to our web site ([www.angis.usyd.edu.au/BacPolGenes/welcome.html](http://www.angis.usyd.edu.au/BacPolGenes/welcome.html)), which gives synonyms and other information for such genes.

# 2 Overview of LPS Structure, Biosynthesis and Genetics

## 2.1 LPS Structure

Figure 1 shows the structure of Lipid A and core of K-12. The K-12 and the very similar LT2 Lipid A/core are the type specimens for the classical model of this component. The Lipid A comprises a phosphorylated  $\beta$ -1, 6-linked glucosamine disaccharide with one N- and one O-linked fatty acyl chain on each glucosamine (RAETZ 1996). These fatty acids are usually hydroxy fatty acids and may carry additional fatty acids as acyloxyacyl substituents on these hydroxyl groups. The fatty acids are embedded in the outer membrane and in *E. coli* and *S. enterica*, at least, LPS is thought to be the only major lipid in the outer leaflet of the OM. Lipid A is the most conserved part of LPS. In many species within the Enterobacteriaceae it has a common structure with a  $\beta$ (1-6)-linked disaccharide backbone of amino sugars, commonly glucosamine (TAKAYAMA and QURESHI 1992). Variation is

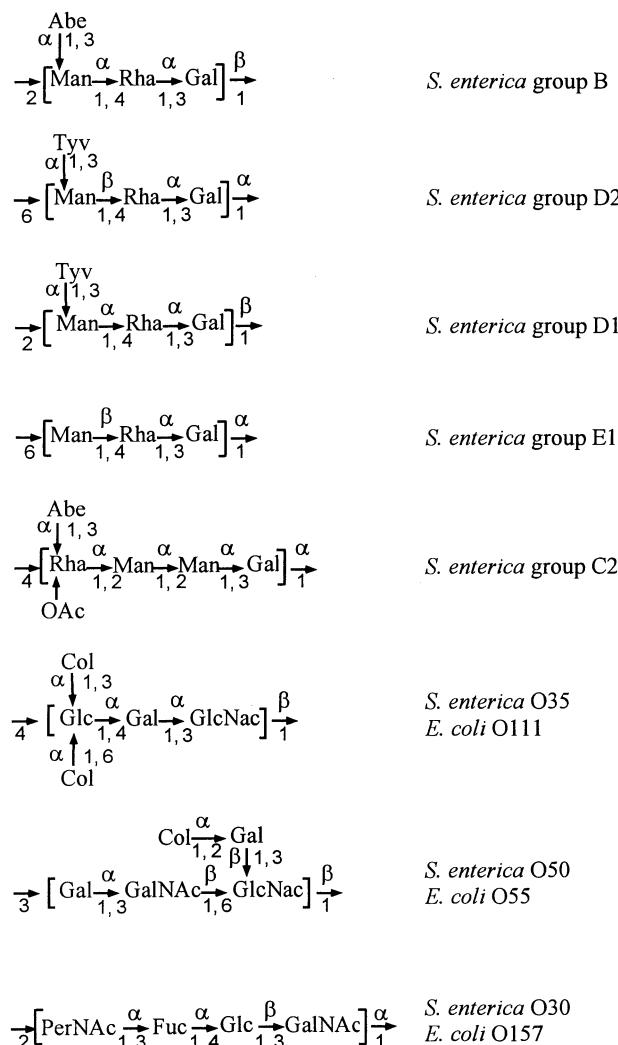


**Fig. 1.** Lipid A/core of *E. coli* K-12. *Glc*, Glucose; *GlcN*, glucosamine; *GlcNAc*, *N*-acetyl glucosamine; *Kdo*, 3-deoxy-D-manno-octulosonic acid; *Hep*, L-glycero-D-manno-heptose; *HM*, 3-hydroxymyristic acid; *M*, myristic acid; *L*, lauric acid. Note that the *KdoIII* substitution is not stoichiometric. Note also that almost all K-12 strains carry a mutation in the O-antigen gene cluster and so do not express O antigen (LIU and REEVES 1994a). See text, HOLST and BRADE (1992) and RICK and RAETZ (1999) for details

between, rather than within, species and lies mostly in the number and length of the fatty acyl chains, the degree to which the molecule is phosphorylated, the presence or absence of acyloxyacyl substituents, the occurrence of amino sugars other than glucosamine as a component of the backbone, and the presence of additional polar constituents (RIETSCHEL et al. 1992; TAKAYAMA and QURESHI 1992). However, species in other families may have more radical differences (PLÖTZ et al. 2000; RIETSCHEL et al. 1992).

The core is an oligosaccharide and can be divided into two structurally distinct regions: the inner (Lipid A proximal) and outer core (RIETSCHEL et al. 1992). The inner core of *E. coli* is comprised primarily of L-glycero-D-manno-heptose (heptose) and 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) residues. One of the Kdo residues links the core to Lipid A. This inner core structure is generally conserved among the Enterobacteriaceae (RIETSCHEL et al. 1992). The outer core displays more variation, and in *E. coli* there are five outer core types (AMOR et al. 2000) which share a basic structure of a three-hexose backbone and two side-chain residues. The order of backbone hexoses, and the nature, position and linkage of side-chain residues can all vary.

The O antigen, a repeat-unit polysaccharide with units known as O units, is one of the most variable cell constituents. The variations are in the types of sugars present, their arrangement within the O unit and the linkage between repeat units. The highly variable nature of the O antigen provides the basis for serotyping. We show in Fig. 2 a representative sample of different O unit structures. O antigen chain lengths are usually clustered around a modal value, with the Wzz protein



**Fig. 2.** Structure of selected *E. coli* and *S. enterica* O antigen repeat units. *Abe*, Abequose; *Col*, colitose; *Gal*, galactose; *Fuc*, fucose; *Man*, mannose; *Per*, perosamine; *Rha*, rhamnose; *Tyv*, tyvelose; others as in Fig. 1. References given in REEVES 1994, except for *S. enterica* O35/*E. coli* O111 (KENNE et al. 1983), *S. enterica* O50/*E. coli* O55 (KENNE et al. 1983; LINDBERG et al. 1981), and *S. enterica* O30/*E. coli* O157 (PERRY et al. 1986)

being responsible for the modal distribution (BASTIN et al. 1993; BATCHELOR et al. 1991).

## 2.2 Biosynthesis and Genetics of LPS

The three generally recognised components of LPS are synthesised as two rather than three components, Lipid A/core and O polysaccharide, which are then ligated to give the complete molecule, which itself may then be further modified.

Lipid A/core is made in one continuous process. The distinction is made between Lipid A and core largely because of the ease with which the linkage between Kdo and Lipid A can be cleaved by acid hydrolysis to separate the hydrophobic Lipid A and hydrophilic core/O antigen components, the starting points for much chemical analysis. Synthesis is initiated by transfer of R-3-hydroxymyristic acid to UDP-GlcNAc by LpxA to make UDP-3-O-monoacyl-GlcNAc, which is embedded in the cytoplasmic membrane, as are all subsequent intermediates in synthesis of Lipid A/core (see below).

O antigen synthesis proceeds independently with each O unit assembled on the C55 isoprenoid carrier lipid, Undecaprenol (Und) phosphate, the intermediates again being embedded in the cytoplasmic membrane with the hydrophilic O unit on the cytoplasmic face (OSBORN 1979; OSBORN et al. 1972).

Lipid A/core and O unit are separately translocated (flipped) to the periplasmic face of the membrane, where the O unit is polymerised to O polysaccharide on UndPP before transfer to Lipid A/core by WaaL (MÄKELÄ and STOCKER 1984; McGRATH and OSBORN 1991a,b; MULFORD and OSBORN 1983; OSBORN 1979). Little is known of the further steps in export of LPS to the outer face of the OM, but it has recently been shown that TolA protein plays a critical but as yet uncharacterised role in the surface expression of LPS (GASPAR et al. 2000).

## 3 O Antigen Genes in *E. coli* and *S. enterica*

### 3.1 O Antigen Variation

The O antigen is almost universally present in fresh isolates of *E. coli* and *S. enterica* and many other (but not all) species. However, it is often lost after prolonged cultivation in laboratory media, with the mutations characterised in the case of *E. coli* K-12. Most extant strains of K-12 have an IS5 insertion in the rhamnose transferase gene, but at least one strain has instead a deletion of the TDP-rhamnose pathway genes (LIU and REEVES 1994a), indicating two independent events in the period before K-12 became the favoured strain for bacterial genetics.

It is well established that loss of O antigen makes many pathogens serum sensitive or otherwise seriously impaired in virulence (FRANK et al. 1987;

MORRISON and KLINE 1977; PLUSCHKE et al. 1983), but there is very little direct evidence for the role of O-antigen specificity or the benefit of maintaining the large number of forms. There is indirect evidence in the distribution of O-antigen forms. Strains with each of the *E. coli* modes of pathogenicity have only a limited number of O antigens (JOHNSON et al. 1996; NATARO and KAPER 1998). This is not because the strains are closely related, as it has been shown in several studies that they may be quite divergent on multilocus enzyme electrophoresis (MLEE) (CAMPOS et al. 1994; RODRIGUES et al. 1996; WHITTAM 1996). It seems, then, that there is some advantage in particular O antigens for specific forms of pathogenicity. In the case of *E. coli* O1, O7 and O18, there is direct evidence that the O-antigen differences account for differences in the nature of pathogenicity (ACHTMAN and PLUSCHKE 1986; PLUSCHKE et al. 1983). It has also been shown that the virulence of *E. coli* Flexneri is reduced if the O antigen is changed (GEMSKI et al. 1972), and isogenic *S. enterica* serovar *Typhimurium* strains with antigen O4 are more virulent than those in which the O4 antigen has been experimentally substituted with antigen O9 (MÄKELÄ et al. 1973).

The O antigen is on the cell surface and appears to be a major target of both the immune system and bacteriophages, which must apply intense selection. This is probably a major factor in the origin and maintenance of the high level of variation. Each strain expresses only one O antigen form, and the variation is thought to allow the various clones of a species to each present a surface that offers a selective advantage in the niche occupied by that clone. We have estimated that a selective advantage of only 0.1% for one O antigen over another in a given niche is more than sufficient to maintain different alleles in different clones (REEVES 1992).

*S. enterica* and *E. coli* diverged about 140 million years ago (OCHMAN and WILSON 1987a,b) and have, respectively, 54 and 190 (including *Shigella*) known O-antigen forms recognised in their typing schemes (CDC 1999; LIOR 1994; POPOFF and LE MINOR 1997). Although *E. coli* and *S. enterica* are treated as closely related species, there are only three cases in which the O-antigen structure is identical in the two: *E. coli* O111 and *S. enterica* O35 (KENNE et al. 1983); *E. coli* O55 and *S. enterica* O50 (KENNE et al. 1983; LINDBERG et al. 1981); and *E. coli* O157 and *S. enterica* O30 (PERRY et al. 1986). This suggests that there has been extensive turnover of O antigens, presumably by lateral gene transfer, since species divergence. It is noteworthy that the three O antigens common to both species are associated with EPEC (and sometimes EHEC) strains in *E. coli*.

### 3.2 O-Antigen Gene Cluster

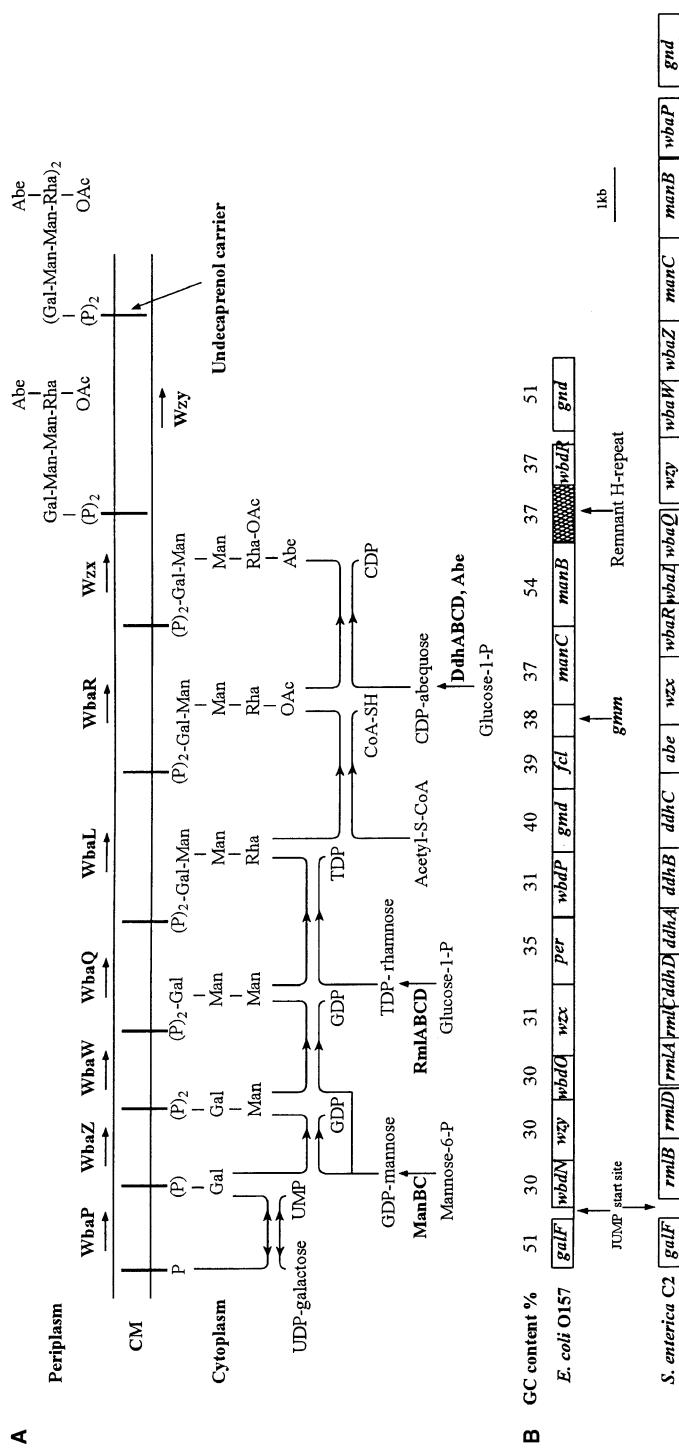
The differences between the many forms of O antigen are due almost entirely to genetic variation in the O-antigen gene cluster, although variation within a recognised form may be due to modification by phage-encoded or other genes outside of the cluster. We use *E. coli* and *S. enterica* for examples of the O antigen gene cluster, but the same principles seem to apply widely. Species which generally have an O antigen have a locus for all genes specific for that structure. That locus is

between *galF* and *gnd* in *E. coli* and *S. enterica* (REEVES 1994), but between *hemH* and *gsk* in *Yersinia* (KESSLER et al. 1991, 1993; ZHANG et al. 1997) and at other loci in other species. Sequence information for about 45 O-antigen gene clusters from 16 species is currently available ([www.angis.usyd.edu.au/BacPolGenes/welcome.html](http://www.angis.usyd.edu.au/BacPolGenes/welcome.html)).

The O-antigen gene cluster includes nucleotide sugar pathway genes, sugar transferase genes and O unit processing genes. Some O antigens include O acetyl groups or other residues, and the transferase genes for these may be in the cluster. Genes for early steps in pathways that are also involved in housekeeping functions are generally not duplicated in the O-antigen cluster (see REEVES 1994 for review).

Figure 3 shows the gene clusters for *S. enterica* C2 and *E. coli* O157 O antigens. The *S. enterica* C2 O unit has four different sugar residues (Fig. 2) and the genes for each step in synthesis of this structure have been identified as shown in Fig. 3. The 11 enzymes for synthesis of three nucleotide sugars, the six transferases to synthesize the O unit on UndPP, the O antigen flippase, and the O antigen polymerase are encoded by genes located within the O-antigen gene cluster (BROWN et al. 1992; LIU et al. 1993, 1995). The biosynthetic pathway genes for the three O antigen-specific nucleotide sugar precursors are themselves clustered within the O-antigen gene cluster (Fig. 3). *manB* and *manC* code for enzymes converting mannose-6-P to GDP-mannose; *rmlABCD* encode enzymes for making TDP-rhamnose from glucose-1-P; *ddhABCD* and *abe* encode enzymes for making CDP-abequose from glucose-1-P. Note that UDP-Gal is used in other pathways and is synthesised by housekeeping enzymes. Galactose transferase, encoded by *wbaP*, initiates the synthesis of O unit by transferring galactose phosphate from UDP-Gal to UndP. Transferases encoded by *wbaZ*, *wbaW*, and *wbaQ* then sequentially put on two mannose and one rhamnose residue before the rhamnose residue is O acetylated by WbaL (LIU et al. 1993, 1995). The abequose residue is then linked to O-acetyl rhamnose by WbaR to make a complete O unit. The O unit flippase, encoded by *wzx* (LIU et al. 1996), flips the UndP-linked O unit across the inner membrane before it is polymerised into long-chain O antigen by the O polymerase Wzy (BROWN et al. 1992).

Our second example is the *E. coli* O157 gene cluster, which is less well characterised, as the specificity of the three putative transferase genes has not been determined (WANG and REEVES 1998). The O unit of O157 antigen has four sugars (Fig. 2). Precursors UDP-GalNAc and UDP-Glc are synthesised by housekeeping enzymes. Fcl makes GDP-fucose from GDP-4-keto-6-D-deoxy-mannose, which is synthesised from mannose-6-P via a three-step pathway by ManB, ManC and Gmd (ANDRIANOPOULOS et al. 1998; STEVENSON et al. 1996). GDP-perosamine is made from GDP-4-keto-6-D-deoxymannose by Per (WANG and REEVES 1998). It is thought that *wbdR* encodes the enzyme for making GDP-PerNAc from GDP-perosamine (WANG and REEVES 1998). The *wecA* gene, located in the ECA gene cluster (MEIER and MAYER 1985), encodes the enzyme for transferring GalNAc-P onto UndP as the first sugar, and enzymes encoded by *wbdN*, *wbdO* and *wbdR* are assumed to put on the other three sugars to make the complete O unit (WANG and REEVES 1998). *wzx* and *wzy* genes are also found in the O157 gene cluster (Fig. 3).



**Fig. 3.** **A** Biosynthesis of the *S. enterica* C2 O antigen. Enzymes encoded by the C2 O antigen genes are in **boldface**. *C.M.*, cytoplasmic membrane; other abbreviations as in Figs. 1 and 2. See text for details and references. **B** O-antigen gene clusters of *E. coli* O157 and *S. enterica* C2. See text for details and references

As is commonly the case, the genes in the *S. enterica* C2 and *E. coli* O157 gene clusters have a low G+C content, mostly ranging from 30% to 46% for individual genes. This is in sharp contrast to the *galF* and *gnd* genes adjacent to the two gene clusters (Fig. 3), which have G+C contents typical of housekeeping genes of the host chromosome. *manB* of O157 is a notable exception, with a higher G+C content (54%), and *rmlD* of C2 has a typical G+C content of 50%. Both of these genes are subject to special circumstances because they are commonly found in polysaccharide gene clusters, and it appears that the original gene of an incoming gene cluster can be replaced by a resident copy of the gene with different G+C content. The O antigen *manB* genes of *E. coli* and *S. enterica* are often derived from the *manB* gene of the adjacent colanic acid gene cluster, which is of generally high G+C content (JENSEN and REEVES 2001; STEVENSON et al. 1996, 2000): the *manB* gene of *E. coli* O157 is clearly in this category. In the case of *rmlD* gene the replacement is by genes from other O-antigen gene clusters that appear to have been in the host species for long enough to have gained a typical G+C content (LI and REEVES 2000).

A 39-bp JUMPStart sequence is present upstream of many polysaccharide gene clusters (HOBBS and REEVES 1994) and also upstream of *E. coli* haemolysin and F-factor gene clusters. This sequence, or part of it, is required for the regulation of downstream genes by RfaH (LEEDS and WELCH 1997; NIETO et al. 1996), which works as an antiterminator (BAILEY et al. 1996; BEUTIN et al. 1981; LEEDS and WELCH 1996; STEVENS et al. 1997). In the case of the O-antigen gene cluster it has been shown that expression of all genes in the LT2 O-antigen gene cluster is enhanced by RfaH acting on JUMPStart (WANG et al. 1998).

### 3.3 Intraspecies Movement of O Antigen Genes

The many forms of O antigen in a species constitute a polymorphism in which the morphs vary in the set of genes present in the O-antigen gene cluster. The O antigen present in a strain can change by recombination involving the DNA flanking the locus: In the clonal population structure found in bacteria this has the effect of a strain gaining a new O antigen by lateral transfer from another clone within the species. This does not involve any transposition process. There is good evidence for such transfer in the common occurrence of strains with the same or very similar MLEE pattern but different O antigens. For example, two strains almost identical to K-12 in MLEE have very different O-antigen gene clusters (LIU and REEVES 1994b). Many other examples have been noted, for example by ACHTMAN et al. (1986), but perhaps best known is the case of the O157:H7 and O55:H7 clones. MLEE on a large number of *E. coli* strains has shown that *E. coli* O157:H7 does not resemble any other O157 strains (FENG et al. 1998; WHITTAM et al. 1993) but belongs to a distinct group of mostly O55 strains including *E. coli* O55:H7, to which it is most closely related (FENG 1993; PUPO et al. 1997; WHITTAM and WILSON 1988; WHITTAM et al. 1993). The H7 *fliC* genes of O55:H7 and O157:H7 strains are almost identical and different from those of strains with other O antigens (REID

et al. 1999; WANG et al. 2000), giving sequence data support to the close relationship of the O55:H7 and O157:H7 clones. This all indicates relatively recent acquisition of the O157 O antigen gene cluster by an O55:H7 strain to generate the O157:H7 clone. (Note that “recent” is in an evolutionary context: the data are mute on whether the acquisition was shortly before the O157:H7 clone came to prominence or decades, hundreds, or thousands of years before that.) This is the only case to date in which an O antigen donor has been identified to clone level. The transfer involved homologous recombination in the flanking *galF* gene and a more distant downstream site (Wang, Huskic, Rothemund and Reeves, unpublished data).

The movement by homologous recombination of O-antigen gene clusters within a species has been well documented in *S. enterica*, which has a well-defined subspecies structure, with housekeeping genes generally having a subspecies specific sequence (SELANDER et al. 1991). The *gnd* gene, which is adjacent to the O-antigen gene cluster, generally fits this pattern, but in some strains the gene is that of a different subspecies or partly of the subspecies of the strain and partly of a different subspecies. In the latter situation it is always the 5' end of the gene, adjacent to the O-antigen cluster, that is of the unexpected subspecies, indicating that it was co-transferred with the O-antigen cluster from a strain of that subspecies, with recombination occurring in the *gnd* gene (THAMPAPILLAI et al. 1994). We have extended the study to the *galF* gene at the other end of the O-antigen gene cluster, with comparable results (Lan, Ryan, Bouantoun and Reeves, unpublished data).

Intraspecies O-antigen gene cluster transfer by homologous recombination in the flanking genes, as discussed, generally substitutes one O antigen for another because most forms of the O antigen in the species are encoded by a gene cluster at the same locus.

### 3.4 Interspecies Movement of O Antigen Genes

O-antigen gene clusters also transfer between species. In closely related species this could involve homologous recombination in flanking DNA as for intraspecies transfer. The O8 and O9a O antigens of *E. coli* are examples. They are identical to O5 and O3 of *Klebsiella pneumoniae* and differ from typical *E. coli* O antigens in several respects, including in mapping between *gnd* and *his* instead of between *galF* and *gnd* (SUGIYAMA et al. 1997, 1998). It seems clear that they transferred from *K. pneumoniae*, where the homologues map at the same locus. However, for most interspecies transfer there will not be sufficient homology for this to occur. The only such case for which we have any data is *E. coli* Sonnei. The O antigen of Sonnei is unique in *E. coli*. It has a gene cluster very similar to those for most other *E. coli* O antigens, except that it is on a plasmid instead of on the chromosome between *gnd* and *galF*. It has *wzx* and *wzy* genes for O unit flippase and polymerase, showing that it is in the same class as most other O antigens, and is the only such *E. coli* gene cluster not thought to be between *galF* and *gnd*. (The O8, O9a and O9 gene clusters are in a different class.) It also has as the first gene of the cluster a *wzz* gene for the O antigen chain-length determinant. This is unusual in *E. coli*, in which

*wzz* is usually located outside of the gene cluster on the other side of the *gnd* and *ugd* genes (BASTIN et al. 1993; FRANCO et al. 1998). The *wzz* gene, however, can be in the main cluster in other species such as *Pseudomonas aeruginosa*, *Vibrio cholerae* and *Yersinia pseudotuberculosis*. The location of the gene cluster on a plasmid and the atypical location of *wzz* in the gene cluster suggest a relatively recent transfer from another species. The *Plesiomonas shigelloides* O17 antigen has the same structure as the Sonnei O antigen (RAUSS et al. 1970; TAYLOR et al. 1993), making this species a potential donor of the gene cluster. We have been able to confirm this by showing that most genes in the cluster are identical or nearly identical in the two species, suggesting a recent transfer from *P. shigelloides* (SHEPHERD et al. 2000). The genes at the ends of the locus do show some sequence differences, but these are within the usual range for intraspecies variation, which is more commonly present at the ends of O-antigen gene clusters. This divergence probably represents differences between the *P. shigelloides* strain sequenced and that which was donor for the transfer to *E. coli*. It is the identity or near identity for most of the genes which tells us that there has not been time for any detectable divergence since the interspecies lateral transfer event. There has also been a large deletion in Sonnei in the region where the O-antigen cluster is usually located, apparently caused by homologous recombination between a *manB* gene in the original O-antigen gene cluster and the *manB* gene in the upstream colanic-acid gene cluster (LAI et al. 1998). Thus, for Sonnei we have clear evidence of interspecies transfer of an O-antigen gene cluster on a plasmid and loss of function of the preexisting chromosomal gene cluster.

O54 of *S. enterica* may be in a similar stage to Sonnei as the single gene required for this O antigen is on a plasmid, but no potential donor has been identified. In this case both chromosomal and plasmid-encoded forms are expressed and the O54 gene cluster is of a different class to all others known in *S. enterica* (KEENLEYSIDE and WHITEFIELD 1996).

### 3.5 Origins of the Polymorphism in the O Antigen Locus

The intra- and interspecific lateral transfer events discussed above are in effect redistributing the polymorphic forms. We will now look at what we can infer of the generation of new forms. The lateral transfer events discussed above do not seem to have involved transposition as so often invoked for transfer of typical PAIs. However, transposable elements do seem to have been involved in generation of new O-antigen forms by reassortment of genes within gene clusters. The D2 O antigen of *S. enterica* is a good example, as it appears to have arisen by recombination between two other O-antigen gene clusters and both putative parents are known. The D2 O antigen can be treated as a molecular chimera of the D1 and E1 structures (see Fig. 2). It has the same polymerisation linkage between O units and the same mannosyl glycosidic linkage as E1, but has the side branch abequose of D1. The sequences show that the D2 O-antigen gene cluster is also a chimera (CURD et al. 1998; XIANG et al. 1994). The 5' end of the gene cluster is the same as that of

D1, and the 3' end the same as that of E1, except that a gene of unknown function at the 3' end of the E1 cluster is absent. A single recombination event suffices to provide the genes appropriate for the D2 structure. The genes which specify the properties common to all three gene clusters come from D1 or E1, depending on where they are in the gene cluster. At the junction of the two components there is an H-repeat type transposable element which was presumably involved in its assembly, and it is easy to envisage a mechanism whereby this could have occurred (XIANG et al. 1994).

There are several other cases of transposable elements at sites within an O-antigen gene cluster which suggest involvement in gene cluster formation. For example, in the *E. coli* O157 gene cluster, a remnant H-repeat is located upstream of *wbdR*, the putative *N* acetyl transferase gene (Fig. 3). The presence of a deletion and other mutations in the H-repeat unit sequence indicate that it has been associated with this gene cluster for a long period of time since last undergoing transposition, perhaps after mediating incorporation of *wbdR* into the gene cluster (WANG and REEVES 1998). *wbdR* is the only gene which one could envisage not being essential for synthesis of a functional O antigen, as if the perosamine were not acetylated it is quite likely that polymerisation would still occur to give an O antigen with perosamine in place of *N*-acetyl-perosamine. In itself, this does not make a strong case, but there are several such cases, and we speculated that the ancestral gene cluster had perosamine, with the *N*-acetyl transferase gene being added by lateral transfer mediated by the H-repeat, similar to the H-repeat-mediated gene transfer proposed for the *S. enterica* D2 O-antigen cluster (WANG and REEVES 1998). Another example from *V. cholerae* is discussed later.

### 3.6 The O-Antigen Gene Cluster as a PAI

Having looked at the O-antigen gene clusters of *E. coli* and *S. enterica*, we can now consider their similarities and differences in relation to typical PAIs. The O-antigen gene cluster is present in both pathogenic and nonpathogenic strains of *E. coli*, but some O antigens, such as O111 and O157, are commonly found in pathogenic clones and their gene clusters have characteristics of PAIs, in that they comprise a set of genes correlated with pathogenicity. However, gaining a new O-antigen gene cluster itself does not transform a normally benign organism into a pathogen, as there are other requirements and, in particular, the relevant PAI(s).

In *E. coli* PAIs are present specifically in pathogenic clones and the PAIs found correlate well with mode of pathogenicity (GROISMAN and OCHMAN 1996; HACKER and KAPER 2000). The situation in *S. enterica* is rather different, as all have SPI1 and all but subspecies V strains have SPI2. In one sense all *S. enterica* are pathogenic, having properties related to eucaryote cell invasion. However, they vary enormously in the manner in which this "pathogenicity" is expressed. Some, such as serovar *Sofia* resident in many flocks of chickens in Australia, are not known to cause symptoms of disease in any hosts, while serovars *Typhi* and *Typhimurium* cause systemic and potentially lethal infection in appropriate hosts. The serovars

causing disease in humans and most of those causing food poisoning have one of only a few O antigens, mostly B, C1, C2, D or E. Those not causing symptoms of disease are in reality commensal, and in *S. enterica* there is a better correlation of true pathogenicity with O antigen specificity than with presence of the named PAIs, although many of the more pathogenic forms carry a virulence plasmid (GULIG 1990; GULIG et al. 1993; ROTGER and CASADESUS 1999).

O-antigen gene clusters also resemble PAIs in commonly being of lower G + C content than typical, an indication of recent acquisition of these genes. As discussed above, we have recently identified a plasmid intermediate for transfer of an O-antigen gene cluster into *E. coli*. It is not known if this is the usual means of interspecies transfer, and nothing is known of how a gene cluster, once transferred, is later inserted into the chromosome at the locus typical for that species. It is clear that intraspecies transfer of O antigen genes is largely by homologous recombination in flanking DNA, and this provides a clue as to why the O-antigen gene clusters seem to end up the same site in any given species. Selection for a new O antigen is probably often in part selection against the old by the immune system. We suspect that integration at the appropriate site is driven by the combined need to have only one O-antigen gene cluster present, and by the advantage from time to time of changing that O antigen as the niche of the strain changes. Gene clusters at the locus common in that species will, on transfer, automatically displace the resident gene cluster, as in the O157 replacement of O55 in *E. coli* discussed above.

In contrast to PAIs, O-antigen clusters are stable in chromosomal location in that they do not move by transposition to other loci. They generally lack direct or inverted repeat elements at the ends of the cluster, and the gene clusters as a whole do not have the appearance of a transposable element. It is not common to find an IS in an O-antigen gene cluster but, as discussed above, those that have been found seem to be related to reassortment of genes within the gene cluster rather than to transposition of the whole gene cluster. Nevertheless, the process of loss and gain of O antigen forms is continuous and there has been an almost complete turnover of O antigen forms in *E. coli* and *S. enterica* in the past 140 million years (there are only three forms common to both species).

#### **4 Lipid A/Core Genes of *E. coli* and *S. enterica***

Lipid A/core is made in a continuous process, quite separate from the synthesis of O antigen. Lipid A/core plays an important role in the OM and a complete Lipid A/inner core molecule is crucial for OM stability in *E. coli* and *S. enterica*, although (Kdo)<sub>2</sub>-Lipid A is the minimal LPS structure required for viability in the laboratory. Study of gene function in Lipid A synthesis in particular has been much more difficult than study of O antigen genetics, due to the requirement of the basic structure for cell viability. However, the major pathways are now well understood. The genes show some interesting characteristics in the context of

PAIs as some are in a major cluster while others are present as single genes or small clusters.

#### 4.1 Genes for the Synthesis of Lipid A/Core Precursors

The structure of K-12 Lipid A and core is shown in Fig. 1. UDP-GlcNAc, UDP-Glc and UDP-Gal are present for other purposes, and their biosynthetic pathways are encoded by housekeeping genes. L-glycero-D-manno-heptose (heptose) is added from its ADP derivative, which is synthesised from sedoheptulose 7-phosphate via four steps (EIDELS and OSBORN 1971; EIDELS and OSBORNE 1974). The *gmhA* and *gmhD* genes encode enzymes for the first and last steps (BROOKE and VALVANO 1996; COLEMAN 1983), and VALVANO et al. (2000) have recently shown that WaaE is a bifunctional enzyme involved in synthesis of ADP-heptose. They postulate that it carries out the second and third steps (VALVANO et al. 2000); however, this is still not proven and the name *waaE* is being retained until details of the pathway are clear, when a *gmh\** name will be given (VALVANO et al. 2000). The three genes are conserved in K-12 and LT2, with nucleotide sequence similarity levels between 81% and 87%. *gmhA*, *gmhD* and *waaE* are located at 81.7, 5.5 and 68.8min, respectively, on the K-12 chromosome (BERLYN 1998), with G+C contents of 51%, 51% and 55.7%.

Kdo is transferred from CMP-Kdo, synthesised from arabinose-5-P and PEP via a three step pathway (for review see RAETZ 1990). Two of the genes, *kdsA* and *kdsB*, are well characterised (GOLDMAN et al. 1986; WOISETSCHLAGER et al. 1988; WOISETSCHLAGER and HOGENAUER 1987) and located at 27.3 and 20.9min on the K-12 chromosome, respectively (BERLYN 1998), with G+C contents of 51.6% and 52.7%. These two genes are conserved in LT2, with nucleotide sequence identity levels at 84% and 79% between each pair. The gene for the third step has not been identified.

#### 4.2 (Kdo)<sub>2</sub>-Lipid A Synthesis Genes in K-12 and LT2

(Kdo)<sub>2</sub>-Lipid A, also known as Re LPS, is an important intermediate, as it has the full complement of fatty acids, the last being added only after the Kdo (RAETZ 1996; RICK and RAETZ 1999). Its synthesis in *E. coli* involves ten reactions (nine enzymes), detailed in a recent review (RICK and RAETZ 1999). All nine genes have been identified in K-12. LpxA initiates synthesis by transfer of R-3-hydroxymyristic acid to UDP-GlcNAc to make UDP-3-O-monoacyl-GlcNAc. *lpxA*, *lpxD* and *lpxB* for steps 1, 3, and 5, respectively, map at 4min with *fabZ* and several genes involved in macromolecular synthesis. *lpxC*, *lpxK*, *waaM* and *waaN* (for steps 2, 6, 8 and 9, respectively) are located at 2.3, 20.9, 24.1 and 41.8min on the K-12 chromosome. The *waaA* gene, encoding a bifunctional Kdo transferase for addition of 2 Kdo residues (step 7), is in the *waa* gene cluster (see below). All of the above eight genes are of normal G+C content (between 51% and 54%) in K-12 and are conserved in

LT2 with nucleotide sequence identity levels between 75% and 92%. Recently, the *lpxH* gene, encoding the enzyme for the fourth step, has been identified at 12min on the K-12 chromosome (Babinski and Raetz, personal communication).

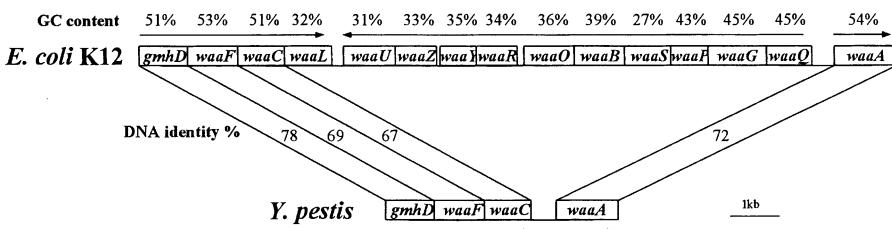
### 4.3 LPS Core Synthesis Genes in K-12

We refer the reader to a recent review (HEINRICHS et al. 1998b) and the references therein for details on this topic. Hep I and Hep II are added to (Kdo)<sub>2</sub>-Lipid A from ADP-Hep to form the Lipid A/inner core. The outer core is then synthesised by further sequential addition of sugars to Lipid A/inner core.

The transferase genes for the synthesis of core oligosaccharide on Lipid A are all located in the *waa* gene cluster, which maps at 81–82min on both the K-12 chromosome (BERLYN 1998) and the LT2 chromosome (SANDERSON et al. 1995). The *waa* gene cluster includes genes for all transferases for assembly of core, together with the gene for the last step of ADP-L-glycero-D-manno-heptose synthesis and the O antigen ligase gene *waaL* (RAETZ 1996). The gene cluster contains three operons, defined by their first genes as the *gmhD*, *waaQ* and *waaA* operons (Fig. 4).

The *gmhD* operon, located at the 5' of the *waa* gene cluster, contains *gmhD*, *waaF*, *waaC* and *waaL*. GmhD catalyzes the last reaction in the synthesis of ADP-L-glycero-D-manno-heptose, as mentioned above. WaaC is the transferase for HepI (KADRMAS and RAETZ 1998; see Fig. 1), and WaaF is a putative transferase for HepII (SCHNAITMAN and KLENA 1993). WaaL is the ligase for attachment of O polysaccharide onto Lipid A/core (KLENA et al. 1992).

The central *waaQ* operon contains ten genes (HEINRICHS et al. 1998b). WaaQ is the transferase for HepIII (YETHON et al. 1998). WaaG, WaaO and WaaR are transferases for GlcI, GlcII and GlcIII (see Fig. 1), respectively (HEINRICHS et al. 1998a; SCHNAITMAN and KLENA 1993), and WaaB is the transferase for the Gal residue (see Fig. 1) (SCHNAITMAN and KLENA 1993). WaaP and WaaY are involved in phosphorylation of the heptose residues (YETHON et al. 1998), while the functions of WaaU, WaaS and WaaZ are not clear. A JUMPStart sequence is also found in the upstream region of the *waaQ* operon, and genes in the *waaQ* operon are regulated by RfaH (SCHNAITMAN and KLENA 1993).



**Fig. 4.** Genetic organization of the *E. coli* K-12 *waa* gene cluster for oligosaccharide core biosynthesis and the corresponding region in *Y. pestis*. The *E. coli* K-12 sequence was extracted from GenBank entry U00096 and the *Y. pestis* map was drawn based on the genomic sequence data of *Y. pestis* strain CO-92 ([http://www.sanger.ac.uk/Projects/Y\\_pestis](http://www.sanger.ac.uk/Projects/Y_pestis))

The *waaA* operon contains the *waaA* gene for the bifunctional Kdo transferase and a gene of unknown function.

#### 4.4 Variation in LPS Core Synthesis Genes

Lipid A/inner core is the most conserved part of LPS, probably due partly to its importance in maintaining the stability of OM and partly to its being buried below the outer core and O antigen, and thus not as much involved in surface interactions. The nine genes involved in (Kdo)<sub>2</sub>-Lipid A synthesis, the three genes for synthesis of the heptose precursor, and the genes for the first two heptose transferases (*waaF* and *waaC*) are conserved between K-12 and LT2 and are of normal G+C content. The level of identity ranges from 75% to 92%. However, the K-12 *waaQ*, *waaP* and *waaY* genes, for the third heptose transferase and modification of heptose residues, also conserved in LT2, have DNA identity levels between 67% and 74% and lower G+C contents of 45%, 43% and 35%, respectively.

The outer core oligosaccharide is more variable, and five types have been described for *E. coli*, known as the K-12 and R1–R4 core types (AMOR et al. 2000). The sequences of the K-12 and LT2 *waa* gene clusters have been known for some time (see RAETZ 1996 for review) and recently, the region from *waaC* to *waaA* was sequenced for *E. coli* types R1–R4 (HEINRICHS et al. 1998b). The *waaC* and *waaA* genes, involved in the synthesis of LipdA/inner core, are conserved in all of them with amino acid identities greater than 84% and 96% for each gene.

There are seven to ten genes in the *waaQ* operons of the five *E. coli* core types and that of LT2. The LipdA/inner core plus the first Hex residue (glucose) is conserved in all six types and, consistent with this, the five inner core genes (*waaQ*, *waaF*, *waaC*, *waaP* and *waaY*) are conserved (HEINRICHS et al. 1998b), as is *waaG* encoding the HexI glycosyltransferase (HEINRICHS et al. 1998b). As expected, *waaG* is the only outer core gene conserved in all six types.

#### 4.5 Summary of Function, Location and G+C Content of Lipid A/Core Genes

The distribution of the genes for synthesis of Lipid A/core is very interesting. The *waa* gene cluster includes all the glycosyl transferase genes plus *gmhD*, which encodes the last step in ADP-L-glycero-D-manno-heptose synthesis. Their products carry out series of reactions to make Lipid A/core from Lipid A and the nucleotide sugars (including CMP-Kdo, but only from ADP-D-D-Hep which has to be converted by GmhD to ADP-L-D-Hep first). The G+C contents for the K-12 *waa* gene cluster genes are shown in Fig. 4. The G+C contents and location on the chromosome and within the cluster reflect the distribution of the genes among other species. All known structures of core from species of Enterobacteriaceae have the Kdo and first two Hep residues conserved (HOLST and BRADE 1992). The relevant

genes for these residues, which are part of the common Enterobacterial genome, are scattered around the chromosome or located at the ends of the *waa* gene cluster (*waaA*, *waaC* and *waaF* together with *gmhD*). The other genes, in the centre of the *waa* gene cluster, code for transferases that add Hep III and outer core, and the ligase that adds O antigen after the completion of core synthesis. This is the more variable part of Lipid A/core.

We thus have a hierarchy of genes. The genes for the most widespread functions, synthesis of ADP-L-glycero-D-manno-heptose and CMP-Kdo, are mostly away from the PAI-like *waa* gene cluster. The next most widespread genes, mostly for transferases conserved in the Enterobacteriaceae, are on the periphery of the gene cluster, and *waaG*, for the only hexose linkage common to all *E. coli* and *S. enterica*, is adjacent to these genes. The genes for the transferases which differentiate the six core forms found in *E. coli* and *S. enterica* are in the centre of the clusters, and it is only these genes that have a low G+C content. It appears that genes for what are effectively housekeeping genes in Enterobacteriaceae are not clustered, whereas those less widespread are in the *waa* gene cluster. *gmhD* is interesting in this context, as the first three steps in ADP-L-glycero-D-manno-heptose synthesis complete the synthesis of ADP-D-glycero-D-manno-heptose, which is present in the inner core of some nonenterobacterial species. GmhD converts ADP-D-D-Hep to ADP-L-D-Hep and hence may be less widespread than genes for the first three steps of the pathway, and the presence of *gmhD* in the *waa* gene cluster is also consistent with the pattern.

The pattern is reminiscent of the situation found for O antigen genes, with pathway genes for more common sugar precursors scattered around the chromosome, while genes specific to O antigen are in the O-antigen cluster. Within O-antigen gene clusters, pathway genes are grouped typically at the ends of the cluster and those unique to an O antigen (such as transferase genes) are located in the centre (see, for example, the *S. enterica* C2 O-antigen gene cluster in Fig. 3).

In both cases the genes that are present in a wide range of species and all or most cells of the species are not clustered, and within the clusters the genes found in all or many of the gene clusters are at the ends of the gene cluster. This arrangement facilitates lateral transfer of the more variable part of the cluster, as has been observed for the O-antigen clusters in *S. enterica* and *E. coli*.

## 5 LPS Genes of Species Other than *E. coli* and *S. enterica*

We have focussed on the LPS of *E. coli* and *S. enterica* because there are many more experimental data available on the gene clusters and function of individual genes for these species. However, we should look at some other examples as – while they have much in common – there are significant differences between species, and also there are several interesting examples of PAI-like situations in other species, of which we have singled out three examples.

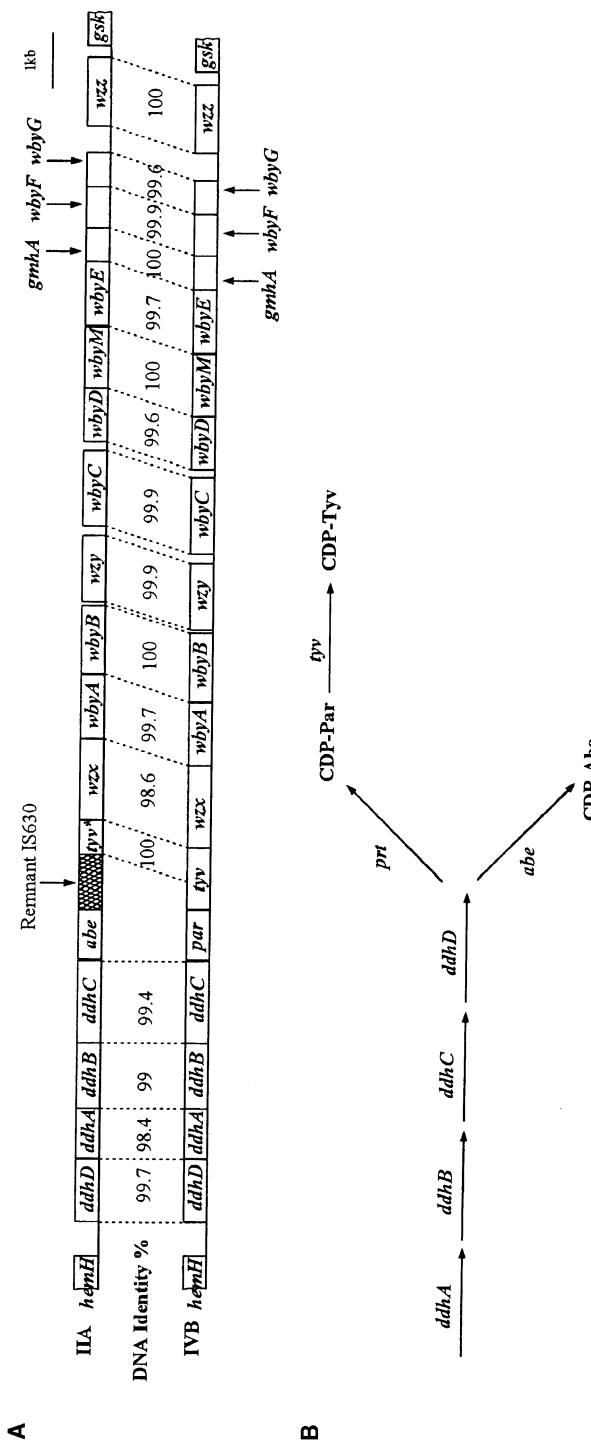
## 5.1 O Antigen Genes in *Yersinia* spp.

*Yersinia pseudotuberculosis* is unusual in that most of the O antigens reported include a dideoxyhexose (DDH) sugar in their structure. Indeed, all five naturally occurring 3,6 DDH sugars (abequose, colitose, paratose, tyvelose and ascarylose) are found in *Y. pseudotuberculosis*, whereas these highly immunogenic sugars are relatively rare in other species including *E. coli* and *S. enterica*. Note that in *S. enterica*, although only related O antigens A, B, D1, D2, D3 and C2 have a DDH residue, strains with these O antigens are very prevalent and DDH containing O antigens are common in that sense.

The O antigen gene cluster of both *Y. pseudotuberculosis* and *Y. enterocolitica* is located between *hemH* and *gsk* genes. (KESSLER et al. 1991, 1993; ZHANG et al. 1993, 1996). These two genes in *E. coli* are separated by one gene of unknown function. Also the gene immediately upstream of *hemH* is *adk* in both genomes. *Yersinia* and *Escherichia* are both in the Enterobacteriaceae and it seems that the gene order in this region has been conserved but the O-antigen gene cluster is inserted here in *Yersinia*. The converse situation is found for the *gnd* and *galF* genes. In *E. coli* and *S. enterica* these genes flank the locus for the O-antigen gene cluster, but perusal of the *Yersinia pestis* genome ([http://www.sanger.ac.uk/Projects/Y\\_pestis](http://www.sanger.ac.uk/Projects/Y_pestis)) shows that they are adjacent in the *Y. pestis* and presumably *Y. pseudotuberculosis*, as *Y. pestis* is, in effect, a clone of *Y. pseudotuberculosis* (ACHTMAN et al. 1999).

We have sequenced the gene clusters for the *Y. pseudotuberculosis* IIA and IVB O antigens (Pacinelli, Wang and Reeves, unpublished data). These O antigens have similar structures but contain tyvelose and abequose, respectively. The biosynthetic pathways for the activated nucleotide forms of these two DDH sugars are related (RAETZ 1996), with four common steps encoded by genes *ddhABCD* for conversion of glucose-1-phosphate to CDP-4-keto-3,6-dideoxy-D-glucose. This can then be converted either to CDP-abequose by CDP-abequose synthase (encoded by *abe*) or to CDP-paratose by paratose synthase (encoded by *prt*). CDP-paratose can then be converted into CDP-tyvelose by CDP-tyvelose epimerase (encoded by *tyv*) (Fig. 5). The IIA and IVB gene clusters are shown in Fig. 5. We know from older data that the gene order in *Y. pseudotuberculosis* is, as in *S. enterica*, *ddh-DABC*, followed by the *abe*, *prt* or *prt* plus *tyv* genes, depending on the DDH present.

The IIA gene cluster contains a truncated *tyv* gene which is almost identical to part of the IVB *tyv* gene. Upstream of this truncated *tyv* gene is a remnant IS sequence. A *tyv* gene is relevant only in the presence of a *prt* gene, as the function of Tyv is epimerisation of CDP-paratose. This indicates that the IIA gene cluster is derived from a form with tyvelose in the O antigen, and we postulate a recombination event mediated by an IS element in which the *prt* and *tyv* genes of the ancestral gene cluster were replaced by a functional *abe* gene to generate the IIA cluster (HOBBS and REEVES 1995). The other genes in the two clusters are almost identical (Fig. 5), and thus the IVB gene cluster has all the characteristics expected of the ancestral form for the IIA gene cluster. The source of the *abe* gene has not yet



**Fig. 5.** **A** Comparison of the O-antigen genes from *Y. pseudotuberculosis* groups IIA and IVB. Part of the sequences has been previously reported (GenBank entry L01777), and the remainder is unpublished data (Pucinella, Wang, and Reeves). **B** Biosynthetic pathway of CDP-Tyv and CDP-Abe derived from CDP-Glc, with genes for each step indicated. *Par*, paratose; other abbreviations as in Fig. 2

been established but was most probably a *Y. pseudotuberculosis* strain with an O antigen containing abequose.

### **5.2 IS1358 and the Origins of the Ogawa Form of the *Vibrio cholerae* O1 Antigen**

In the *Y. pseudotuberculosis* IIA and *S. enterica* D2 examples discussed above, we had both an apparent recombinant gene cluster and one or both presumptive parent gene clusters on which to base the argument that transposable elements had been involved in reassortment of genes to generate a new O antigen form.

There are other cases of transposable elements being adjacent to genes which are not essential for formation of an O unit and could have been inserted after the basic gene cluster was assembled, as they add to the basic repeating structure. In these cases also the postulated addition of a gene could have involved transposition. The only evidence is the location of a transposable element next to the gene, as we do not at present have presumptive parents. We have discussed the case of the *wbdR* gene of the *E. coli* O157 gene cluster. Another example is the *wbeT* gene of the *Vibrio cholerae* O1 O antigen. WbeT is the methyl transferase which converts the Inaba to the Ogawa form of the O1 O antigen (MANNING et al. 1993). The *wbeT* gene is adjacent to a defective IS1358 element which could have been involved in transfer of the *wbeT* gene as proposed above for the *E. coli* O157. In this case it is clear that if the gene were not present a fully functional O antigen would be made, as *wbeT* mutants make the Inaba form of O antigen (MANNING et al. 1993). Conversion of Ogawa strains to the Inaba form involves mutation of *wbeT* and occurs quite frequently, but the reverse change can occur only if the change in *wbeT* is reversible and thus will occur only in some Inaba strains and is relatively rare. Thus it is quite plausible that the original form lacked *wbeT*, which was gained later. Again the IS element is adjacent to the only gene for which absence is likely to leave a functional gene cluster, providing support for the hypothesis that it was involved in addition of the gene to the O1 gene cluster.

### **5.3 LPS Core Genes of *Yersinia pestis* and *Yersinia pseudotuberculosis***

The *waa* gene cluster of *Y. pestis* is easily located in the nearly complete genome sequence currently available ([http://www.sanger.ac.uk/Projects/Y\\_pestis](http://www.sanger.ac.uk/Projects/Y_pestis)) as the genes are very similar to those of K-12. What is remarkable is that only four genes are present, the whole of the low G + C region of K-12 being absent (Fig. 4). We have shown by PCR, using primers based on these four genes, that the same situation applies in *Y. pseudotuberculosis* (unpublished data). The genes present in the *waa* gene cluster are sufficient to give LPS with Kdo, HepI and HepII but no HepIII or outer core sugars. This is the component of LPS core common to all Enterobacteriaceae. *Y. pseudotuberculosis* has an outer core that differs substan-

tially from those of *E. coli* and *S. enterica*, with only the basic four residue enterobacterial components in common (HOLST and BRADE 1992). The genes for synthesis of the remainder of the *Y. pseudotuberculosis* core must map elsewhere. This raises interesting questions on the evolution of the enterobacterial LPS core. Is the situation in *Y. pseudotuberculosis* the primitive one, with the *E. coli/S. enterica* lineage having incorporated the genes for the outer core within the cluster for the inner core? Or were the central, largely outer core genes of the common ancestor replaced functionally in *Y. pseudotuberculosis* by genes elsewhere on the chromosome before being lost from the cluster by deletion? This glimpse of the changing pattern of LPS genes seen in the *Yersinia*, *E. coli/S. enterica* comparison heralds an exciting period ahead as we see more genome sequences, and the situation certainly has parallels with typical PAIs.

## 6 Conclusions

O-antigen gene clusters have many properties of PAIs but nonetheless differ in significant ways. Like traditional PAIs, the O antigen can be important for pathogenicity of a clone. Unlike traditional PAIs, however, the site of incorporation appears to be important. Almost all O-antigen gene clusters of any species map to the same locus. Clearly, this locus can change, as it differs in *Yersinia* and *Escherichia*, so in one or both the locus must have changed since these members of the Enterobacteriaceae diverged. However, that divergence occurred over a very long time frame, and in the time frame over which PAIs move between species, the variation between species in O-antigen locus must act as a major deterrent to interspecies transfer of O-antigen genes by homologous recombination. But despite conservatism in genetic location, there is substantial diversity in the O antigens of most species, and clones can gain new forms by homologous recombination with other strains of the same species. A similar situation probably applies to genes of the LPS core, as discussed for *E. coli*, but the frequency of change is probably much less.

We have the interesting situation that LPS gene clusters look more like alleles of a chromosomal locus than a PAI on the time frame we usually consider for PAI movement, but have more similarity to PAIs if considered on a longer time frame.

## References

- Achtman M, Pluschke G (1986) Clonal analysis of descent and virulence among selected *Escherichia coli*. *Annu Rev Microbiol* 40:185–210
- Achtman M, Heuzenroeder M, Kusecek B, Ochman H, Caugant D, Selander RK, Väisanen-Rhen V, Korhonen TK, Stuart S, Ørskov F, Ørskov I (1986) Clonal analysis of *Escherichia coli* O2:K1 isolated from diseased humans and animals. *Infect Immun* 51:268–276

- Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E (1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc Natl Acad Sci USA 96:14043–14048
- Amor K, Heinrichs DE, Fridrich E, Ziebell K, Johnson RP, Whitfield C (2000) Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. Infect Immun 68:1116–1124
- Andrianopoulos K, Wang L, Reeves PR (1998) Identification of the fucose synthetase gene in the colanic acid gene cluster of *Escherichia coli* K-12. J Bacteriol 180:998–1001
- Bailey MJ, Hughes C, Koronakis V (1996) Increased distal gene transcription by the elongation factor RfaH, a specialized homologue of NusG. Mol Microbiol 22:729–737
- Bastin DA, Brown PK, Haase A, Stevenson G, Reeves PR (1993) Repeat unit polysaccharides of bacteria: a model for polymerisation resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. Mol Microbiol 7:725–734
- Batchelor RA, Haraguchi GE, Hull RA, Hull SI (1991) Regulation by a novel protein of the bimodal distribution of lipopolysaccharide in the outer membrane of *Escherichia coli*. J Bacteriol 173:5699–5704
- Berlyn MKB (1998) Linkage map of *Escherichia coli* K-12, edition 10: the traditional map. Micro Mol Biol Rev 62:814–984
- Beutin L, Manning PA, Achtman M, Willetts N (1981) *sfrA* and *sfrB* products of *Escherichia coli* K-12 are transcriptional control factors. J Bacteriol 145:840–844
- Brooke JS, Valvano MA (1996) Biosynthesis of inner core lipopolysaccharide in enteric bacteria: identification and characterization of a conserved phosphoheptose isomerase. J Biol Chem 271:3608–3614
- Brown PK, Romana LK, Reeves PR (1992) Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar Muenchen (strain M67): genetic basis of the polymorphism between groups C2 and B. Mol Microbiol 6:1385–1394
- Campos LC, Whittam TS, Gomes TAT, Andrade JRC, Trabulsi RL (1994) *Escherichia coli* serogroup O111 includes several clones of diarrheagenic strains with different virulence properties. Infect Immun 62:3282–3288
- CDC (1999) Laboratory methods for the diagnosis of epidemic dysentery and cholera. Atlanta, Georgia
- Coleman WG (1983) The *rfaD* gene codes for ADP-L-glycero-D-manno-heptose-6-epimerase. J Biol Chem 258:1985–1990
- Curd H, Liu D, Reeves PR (1998) Relationships among the O-antigen gene clusters of *Salmonella enterica* groups B, D1, D2, and D3. J Bacteriol 180:1002–1007
- Eidels L, Osborn MJ (1971) Lipopolysaccharide and aldoheptose biosynthesis in transketolase mutants of *Salmonella typhimurium*. Proc Natl Acad Sci USA 68:1673–1677
- Eidels L, Osborne MJ (1974) Phosphoheptose isomerase, first enzyme in the biosynthesis of aldoheptose. Microbiol Rev 249:5642–5648
- Feng P (1993) Identification of *Escherichia coli* O157:H7 by DNA probe specific for an allele of *uidA* gene. Mol Cell Probes 7:151–154
- Feng P, Lampel KA, Karch H, Whittam TS (1998) Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. J Infect Dis 177:1750–1753
- Franco VA, Liu D, Reeves PR (1998) The Wzz (Cld) protein in *Escherichia coli*: amino acid sequence variation determines O antigen chain length specificity. J Bacteriol 180:2670–2675
- Frank MM, Joiner K, Hammer C (1987) The function of antibody and complement in the lysis of bacteria. Rev Infect Dis 9 [Suppl 5]:S537–545
- Gaspar JA, Thomas JA, Marolda CL, Valvano MA (2000) Surface expression of O-specific lipopolysaccharide in *Escherichia coli* requires the function of the TolA protein. Mol Microbiol 38:262–275
- Gemski PJ, Sheahan DG, Washington O, Formal SB (1972) Virulence of *Shigella flexneri* hybrids expressing *Escherichia coli* somatic antigens. Infect Immun 6:104–111
- Goldman RC, Bolling TJ, Kohlbrenner WE, Kim Y, Fox JL (1986) Primary structure of CTP: CMP-3-deoxy-D-manno-octulosonate cytidyl-transferase (CMP-KDO synthetase) from *Escherichia coli*. J Biol Chem 261:15831–15835
- Groisman EA, Ochman H (1996) Pathogenicity islands: bacterial evolution in quantum leaps. Cell Vol 87:791–794
- Gulig PA (1990) Virulence plasmids of *Salmonella typhimurium* and other salmonellae. Microb Pathog 8:3–11
- Gulig PA, Danbara H, Guiney DG, Lax AJ, Norel F, Rhen M (1993) Molecular analysis of spv virulence genes of the salmonella virulence plasmids. Mol Microbiol 7:825–830
- Hacker J, Kaper JB (2000) Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol 54:641–679

- Heinrichs DE, Monteiro MAM, Perry MB, Whitfield C (1998a) The assembly system for the lipopolysaccharide R2 core-type of *Escherichia coli* is a hybrid of those found in *Escherichia coli* K-12 and *Salmonella enterica*. *J Biol Chem* 273:8849–8859
- Heinrichs DE, Yethon JA, Whitfield C (1998b) Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. *Mol Microbiol* 30: 221–232
- Hobbs M, Reeves PR (1994) The JUMPstart sequence:a 39-bp element common to several polysaccharide gene clusters. *Mol Microbiol* 12:855–856
- Hobbs M, Reeves PR (1995) Genetic organisation and evolution of *Yersinia pseudotuberculosis* 3,6-dideoxyhexose biosynthetic genes. *Biochim Biophys Acta* 1245:273–277
- Holst O, Brade H (1992) Chemical structure of the core region of lipopolysaccharides. In: Morrison DC, Ryan JL (eds) *Bacterial endotoxic lipopolysaccharides*. CRC Press, Boca Raton, Florida
- Jensen SO, Reeves PR (2001) Molecular evolution of the GDP-mannose pathway genes (*manB* and *manC*) in *S. enterica*. *Microbiology* 147:599–610
- Johnson RP, Clarke RC, Wilson JB, Read SC, Rahn K, Renwick SA, Sandhu KA, Alves D, Karmali MA, Lior H, McEwen SA, Spika JS, Gyles CL (1996) Growing concerns and recent outbreaks involving non-O157:H7 serotypes of verotoxigenic *Escherichia coli*. *J Food Pro* 59:1112–1122
- Kadrymas JL, Raetz CRH (1998) Enzymatic synthesis of lipopolysaccharide in *Escherichia coli*. Purification and properties of heptosyltransferase I. *J Biol Chem* 273:2799–2807
- Keenleyside WJ, Whitefield C (1996) A novel pathway for O-polysaccharide biosynthesis in *Salmonella enterica* serovar Borreze. *J Biol Chem* 271:28581–28592
- Kenne L, Lindberg B, Soderholm E, Bundle DR, Griffith DW (1983) Structural studies of the O-antigens from *Salmonella greenside* and *Salmonella adelaide*. *Carbohydr Res* 111:289–296
- Kessler A, Brown PK, Romana LK, Reeves PR (1991) Molecular cloning and genetic characterization of the *rfb* region from *Yersinia pseudotuberculosis* serovar IIA, which determines the formation of the 3,6 dideoxyhexose abequose. *J Gen Microbiol* 137:2689–2695
- Kessler A, Haase A, Reeves PR (1993) Molecular analysis of the 3,6-dideoxyhexose pathway genes of *Yersinia pseudotuberculosis* serogroup IIA. *J Bacteriol* 175:1412–1422
- Klena JD, Asford II RS, Schnaitman CA (1992) Role of *Escherichia coli* K12 *rfa* genes and the *rfp* gene of *Shigella dysenteriae* 1 in generation of lipopolysaccharide core heterogeneity and attachment of O antigen. *J Bacteriol* 174:7297–7307
- Lai V, Wang L, Reeves PR (1998) *Escherichia coli* clone Sonnei (*Shigella sonnei*) had a chromosomal O-antigen gene cluster prior to gaining its current plasmid-borne O-antigen genes. *J Bacteriol* 180:2983–2986
- Le Minor L, Popoff MY (1987) Designation of *Salmonella enterica* sp. nov., nom. rev., as the type and only species of the genus *Salmonella*. *Int J Syst Bacteriol* 37:465–468
- Leeds JA, Welch RA (1996) RfaH enhances elongation of *Escherichia coli* *hlyCABD* mRNA. *J Bacteriol* 178:1850–1857
- Leeds JA, Welch RA (1997) Enhancing transcription through the *Escherichia coli* hemolysin operon, *hlyCABD*: RfaH and upstream JUMPStart DNA sequences function together via a postinitiation mechanism. *J Bacteriol* 179:3519–3527
- Li Q, Reeves PR (2000) Genetic variation of dTDP-L-rhamnose pathway genes in *Salmonella enterica*. *Microbiology* 146:2291–2307
- Lindberg B, Lindh F, Longren J, Lindberg AA, Svenson SB (1981) Structural studies of the O-specific side-chain of the lipopolysaccharide from *Escherichia coli* O55. *Carbohydr Res* 97:105–112
- Lior H (1994) Classification of *Escherichia coli*. In: Gyles CL (ed) *Escherichia coli* in domestic animals and humans. CAB International, Wallingford, UK
- Liu D, Reeves PR (1994a) *Escherichia coli* regains its O antigen. *Microbiology* 140:49–57
- Liu D, Reeves PR (1994b) Presence of different O antigen forms in three isolates of one clone of *E. coli*. *Genetics* 138:7–10
- Liu D, Haase AM, Lindqvist L, Lindberg AA, Reeves PR (1993) Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica*:identification and characterization of transferase genes of groups B, C2, and E1. *J Bacteriol* 175:3408–3413
- Liu D, Lindquist L, Reeves PR (1995) Transferases of O-antigen biosynthesis in *Salmonella enterica*: dideoxyhexosyl transferases of groups B and C2 and acetyltransferase of group C2. *J Bacteriol* 177:4084–4088
- Liu D, Cole R, Reeves PR (1996) An O-antigen processing function for Wzx(RfbX): a promising candidate for O-unit flippase. *J Bacteriol* 178:2102–2107

- Mäkelä PH, Stocker BAD (1984) Genetics of lipopolysaccharide. In: Rietschel ET (ed) *Handbook of endotoxin*. Elsevier Science, Amsterdam
- Mäkelä PH, Valtonen VV, Valtonen M (1973) Role of O-antigen (lipopolysaccharide) factors in the virulence of *Salmonella*. *J Infect Dis* 128 [Suppl]:S84-S85
- Manning PA, Stroher UH, Morona R (1993) Molecular basis for O-antigen biosynthesis in *Vibrio cholerae* O1: Ogawa-Inaba switching. In: Wachsmuth IK, Blake P, Olsvik O (eds) *Vibrio cholerae* and cholera. American Society for Microbiology, Washington DC
- McGrath BC, Osborn MJ (1991a) Evidence for energy-dependent transposition of core lipopolysaccharide across the inner membrane of *Salmonella typhimurium*. *J Bacteriol* 173:3134–3137
- McGrath BC, Osborn MJ (1991b) Localisation of the terminal steps of O-antigen synthesis in *Salmonella typhimurium*. *J Bacteriol* 173:649–654
- Meier U, Mayer H (1985) Genetic location of genes encoding enterobacterial common antigen. *J Bacteriol* 163:756–762
- Morrison DC, Kline LF (1977) Activation of the classical and properdin pathways of complement by bacterial lipopolysaccharides. *J Immunol* 118:362–368
- Mulford CA, Osborn MJ (1983) A intermediate step in translocation of lipopolysaccharide to outer membrane of *Salmonella typhimurium*. *Proc Natl Acad Sci USA* 80:1159–1163
- Nataro JP, Kaper JB (1998) Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 11:142–201
- Nieto JM, Bailey MJ, Hughes C, Koronakis V (1996) Suppression of transcription polarity in the *Escherichia coli* haemolysin operon by a short upstream element shared by polysaccharide and DNA transfer determinants. *Mol Microbiol* 19:705–713
- Ochman H, Wilson AC (1987a) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 26:74–86
- Ochman H, Wilson AC (1987b) Evolutionary history of enteric bacteria. In: Neidhardt FC (ed) *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*. American Society for Microbiology, Washington, DC
- Ochman H, Whittam TS, Caugant DA, Selander RK (1983) Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. *J Gen Microbiol* 129:2715–2726
- Osborn MJ (1979) Biosynthesis and assembly of the lipopolysaccharide of the outer membrane. In: Inouye M (ed) *Bacterial outer membranes*. John Wiley and Sons, New York
- Osborn MJ, Cynkin MA, Gilbert JM, Muller L, Singh M (1972) Synthesis of bacterial O-antigen. *Methods Enzymol* 28:583–601
- Perry MB, MacLean L, Griffith DW (1986) Structure of the O-chain polysaccharide of the phenol-phase soluble lipopolysaccharide of *Escherichia coli* 0:157:H7. *Biochem Cell Biol* 64:21–28
- Plötz BM, Lindner B, Stetter KO, Holst O (2000) Characterization of a novel Lipid A containing D-galacturonic acid that replaces phosphate residues. *J Biol Chem* 275:11222–11228
- Pluschke G, Mayden J, Achtman M, Levine RP (1983) Role of the capsule and the O-antigen in resistance of O18:K1 *Escherichia coli* to complement-mediated killing. *J Bacteriol* 42:907–913
- Popoff MY, Le Minor L (1997) Antigenic formulas of the *Salmonella* serovars, 7th revision. WHO Collaborating Centre for Reference and Research on *Salmonella*. Institut Pasteur Paris, France
- Pupo GM, Karaolis DKR, Lan R, Reeves PR (1997) Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. *Infect Immun* 65:2685–2692
- Pupo GM, Lan R, Reeves PR (2000) Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci USA* 97:10567–10572
- Raetz CRH (1990) Biochemistry of endotoxins. *Annu Rev Biochem* 59:129–170
- Raetz CRH (1996) Bacterial lipopolysaccharides: a remarkable family of bioactive macroamphiphiles. In: Neidhardt FD (ed) *Escherichia and Salmonella typhimurium: Cellular and molecular biology*, 2nd edn. American Society for Microbiology, Washington, DC
- Rauss K, Kontrohr T, Vertenyi A, Szendrei L (1970) Serological and chemical studies of *Sh. sonnei*, *Pseudomonas shigelloides* and C27 strains. *Acta Microbiol Acad Sci Hung* 17:157–166
- Reeves PR (1992) Variation in O antigens, niche specific selection and bacterial populations. *FEMS Microbiol Lett* 100:509–516
- Reeves PR (1994) Biosynthesis and assembly of lipopolysaccharide. In: Neuberger A, van Deenen LLM (eds) *Bacterial cell wall*. Elsevier, Amsterdam
- Reeves PR, Hobbs M, Valvano M, Skurnik M, Whitfield C, Coplin D, Kido N, Klena J, Maskell D, Raetz C, Rick P (1996a) Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol* 4:495–503

- Reeves PR, Hobbs M, Valvano M, Skurnik M, Whitfield C, Coplin D, Kido N, Klena J, Maskell D, Raetz C, Rick P (1996b) A new nomenclature for bacterial surface polysaccharide genes. <http://wwwangissuoza/BacPolGenes/welcome/html>
- Reid SD, Selander RK, Whittam TS (1999) Sequence diversity of flagellin (*fliC*) alleles in pathogenic *Escherichia coli*. *J Bacteriol* 181:153–160
- Rick PD, Raetz CRH (1999) Microbial pathways of lipid A biosynthesis. In: Brade H, Opal SM, Vogel SN, Morrison DC (eds) *Endotoxin in health and disease*. Marcel Dekker, New York
- Rietschel ET, Brade L, Lindner B, Zähringer U (1992) Biochemistry of lipopolysaccharides. In: Morrison DC, Ryan JL (eds) *Bacterial endotoxic lipopolysaccharides*. CRC Press, Boca Raton, Florida
- Rodrigues J, Scaletsky ICA, Campos LC, Gomes TAT, Whittam TS, Trabulsi LR (1996) Clonal structure and virulence factors in strains of *Escherichia coli* of the classic serogroup O55. *Infect Immun* 64:2680–2686
- Rotger R, Casadesus J (1999) The virulence plasmids of *Salmonella*. *Int Microbiol* 2:177–184
- Sanderson KE, Hessel A, Rudd KE (1995) Genetic map of *Salmonella typhimurium*, edition VIII. *Microbiol Rev* 59:241–303
- Schnaitman CA, Klena JD (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol Rev* 57:655–682
- Selander RK, Beltran P, Smith NH (1991) Evolutionary genetics of *Salmonella*. In: Selander RK, Clark AG, Whittam TS (eds) *Evolution at the molecular level*. Sinauer Associates, Sunderland, Mass.
- Shepherd JG, Wang L, Reeves PR (2000) Comparison of O-antigen gene clusters of *Escherichia coli* (*Shigella*) Sonnei and *Plesiomonas shigelloides* O17: Sonnei gained its current plasmid-borne O-antigen genes from *P. shigelloides* in a recent event. *Infect Immun* 68:6056–6061
- Stevens MP, Clarke BR, Roberts IS (1997) Regulation of the *Escherichia coli* K5 capsule gene cluster by transcription antitermination. *Mol Microbiol* 24:1001–1012
- Stevenson G, Andrianopoulos K, Hobbs H, Reeves PR (1996) Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *J Bacteriol* 178:4885–4893
- Stevenson G, Lan R, Reeves PR (2000) The Colanic Acid gene cluster of *Salmonella enterica* has a complex history. *FEMS Microbiol Lett* 191:11–16
- Sugiyama T, Kido N, Kato Y, Koide N, Yoshida T, Yokochi T (1997) Evolutionary relationship among *rfb* gene clusters synthesizing mannose homopolymer as O-specific polysaccharides in *Escherichia coli* and *Klebsiella*. *Gene* 198: 111–113
- Sugiyama T, Kido N, Kato Y, Koide N, Yoshida T, Yokochi T (1998) Generation of *Escherichia coli* O9a serotype, a subtype of *E. coli* O9, by transfer of the wb\* gene cluster of *Klebsiella* O3 into *E. coli* via recombination. *J Bacteriol* 180:2775–2778
- Takayama K, Qureshi N (1992) Chemical structure of lipid A. In: Morrison DC, Ryan JL (eds) *Bacterial endotoxic lipopolysaccharides*. CRC Press, Boca Raton, Florida
- Taylor DN, Trofa AC, Sadoff J, Chu C, Bryla D, Shiloach J, Cohen D, Ashkenazi S, Lerman Y, Egan W, Schneerson R, Robbins JB (1993) Synthesis, characterization, and clinical evaluation of conjugate vaccines composed of the O-specific polysaccharides of *Shigella dysenteriae* type 1, *Shigella flexneri* type 2a, and *Shigella sonnei* (*Plesiomonas shigelloides*) bound to bacterial toxoids. *Infect Immun* 61:3678–3687
- Thampapillai G, Lan R, Reeves PR (1994) Molecular evolution in the *gnd* locus of *Salmonella enterica*. *Mol Biol Evol* 11:813–828
- Valvano MA, Marolda CL, Bittner M, Glaskin-Clay M, Simon TL, Klena JD (2000) The *rfaE* gene from *Escherichia coli* encodes a bifunctional protein involved in biosynthesis of the lipopolysaccharide core precursor ADP-L-glycero-D-manno-heptose. *J Bacteriol* 182:488–497
- Wang L, Reeves PR (1998) Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect Immun* 66:3545–3551
- Wang L, Jensen S, Hallman R, Reeves PR (1998) Expression of the O antigen gene cluster is regulated by RfaH through the JUMPstart sequence. *FEMS Microbiol Lett* 165:201–206
- Wang L, Rothemund D, Curd H, Reeves PR (2000) Sequence diversity of the *Escherichia coli* H7 *fliC* genes: Implication for a DNA based typing scheme for *E. coli* O157:H7. *J Clin Microbiol* 38:1786–1790
- Whittam TS (1996) Genetic variation and evolutionary processes in natural populations of *Escherichia coli*. In: Neidhardt FC, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE (eds) *Escherichia and Salmonella. Cellular and molecular biology*. ASM Press, Washington, DC

- Whittam TS, Wilson RA (1988) Genetic relationships among pathogenic *Escherichia coli* of serogroup O157. *Infect Immun* 56:2467–2473
- Whittam TS, Wolfe ML, Wachsmuth IK, Ørskov F, Ørskov I, Wilson RA (1993) Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect Immun* 61:1619–1629
- Woisetschläger M, Hogenauer G (1987) The *kdsA* gene coding for 3-deoxy-D-manno-octulosonic acid 8-phosphate synthetase is part of an operon in *Escherichia coli*. *Mol Gen Genet* 207:369–373
- Woisetschläger M, Hodel-Neuhöfer A, Hogenauer G (1988) Localization of the *kdsA* gene with the aid of the physical map of the *Escherichia coli* chromosome. *J Bacteriol* 170: 5382–5384
- Xiang SH, Hobbs M, Reeves PR (1994) Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for its origin from an insertion sequence-mediated recombination event between group E and D1 strains. *J Bacteriol* 176:4357–4365
- Yethon JA, Heinrichs DE, Monteiro MA, Perry MB, Whitfield C (1998) Involvement of *waaY*, *waaQ*, and *waaP* in the modification of *Escherichia coli* lipopolysaccharide and their role in the formation of a stable outer membrane. *J Biol Chem* 273:26310–26316
- Zhang L, Al-Hendy A, Toivanen P, Skurnik M (1993) Genetic organization and sequence of the *rfb* gene cluster of *Yersinia enterocolitica* serotype O:3: similarities to the dTDP-L-rhamnose biosynthesis pathway of *Salmonella* and to the bacterial polysaccharide transport systems. *Mol Microbiol* 9: 309–321
- Zhang L, Toivanen P, Skurnik M (1996) The gene cluster directing O-antigen biosynthesis in *Yersinia enterocolitica* serotype O8: identification of the genes for mannose and galactose biosynthesis and the gene for the O-antigen polymerase. *Microbiology* 142:277–288
- Zhang L, Radziejewska-Lebrecht J, Krajewska-Pietrasik D, Toivanen P, Skurnik M (1997) Molecular and chemical characterization of the lipopolysaccharide O-antigen and its role in the virulence of *Yersinia enterocolitica* serotype O8. *Mol Microbiol* 23:63–76

# Genomic Structure of Capsular Determinants

B. BARRETT, L. EBAH, and I.S. ROBERTS

1	Introduction . . . . .	137
2	The Diversity of <i>E. coli</i> Capsules . . . . .	138
2.1	The Genetics of <i>E. coli</i> Group 1 and 4 Capsules . . . . .	140
2.2	The Genetics of <i>E. coli</i> Group 2 Capsules . . . . .	142
2.3	The Genetics of <i>E. coli</i> Group 3 Capsules . . . . .	144
2.4	Localisation of the Group 2 and 3 Capsule Gene Clusters on the <i>E. coli</i> Chromosome . . . . .	145
3	Capsule Gene Clusters in <i>Neisseria meningitidis</i> . . . . .	146
4	Capsule Diversity in <i>Streptococcus pneumoniae</i> . . . . .	148
4.1	Capsule Gene Clusters in <i>Streptococcus pneumoniae</i> . . . . .	149
4.2	The Role of Transposition and Recombination in the Emergence of <i>S. pneumoniae</i> Capsule Gene Clusters . . . . .	150
5	Capsule Gene Clusters in Other Gram-Positive Pathogens . . . . .	151
6	Concluding Remarks . . . . .	152
	References . . . . .	153

## 1 Introduction

The production of an extracellular polysaccharide capsule is a common feature of many bacteria (WHITFIELD and VALVANO 1993). The capsule, which often constitutes the outermost layer of the cell, mediates the interaction between the bacterium and its immediate environment and plays a crucial role in the survival of bacteria in hostile environments. One such environment is the human host, where interactions between the capsule and the host's immune system may be vital in deciding the outcome of an infection (MOXON and KROLL 1990). In the absence of specific antibody, a capsule offers protection against the nonspecific arm of the host's immune system by conferring increased resistance to complement-mediated killing and complement-mediated opsonophagocytosis (MICHALEK et al. 1988; MOXON and KROLL 1990). A small set of capsular polysaccharides which resemble poly-

saccharide moieties present in host tissue are poorly immunogenic (FINNE 1982; LINDAHL et al. 1994). The *Escherichia coli* K1 and *Neisseria meningitidis* serogroup B capsules, both of which contain *N*-acetyl-neuraminic acid and the heparin-like *E. coli* K5 capsule, all elicit a poor antibody response in infected individuals (JENNINGS 1990) and confer some measure of resistance to the host's adaptive humoral response. Aside from direct interactions with the host's immune system, capsules may promote the formation of biofilms and the colonisation of a variety of ecological niches, including indwelling catheters, prostheses and the formation of alginate-rich biofilms in the lungs of cystic fibrosis patients (ROBERTS 1995). In such instances the polysaccharide may present a permeability barrier to antibiotics and hinder the effective eradication of the bacteria (COSTERTON et al. 1999).

A striking feature of bacterial capsular polysaccharides is their diversity. There is a huge repertoire of structurally different polysaccharide molecules which can be synthesised by any one bacterial species; this is best typified by capsule diversity in *E. coli* and *Streptococcus pneumoniae*. Paradoxically, amongst this diversity there are capsular polysaccharide molecules which are conserved across different bacterial species, such as the *E. coli* K1 and *N. meningitidis* serogroup B capsular polysaccharide. Both the diversity of capsular polysaccharides and the conservation of certain polysaccharide structures across species barriers raise questions about the evolution of capsule gene clusters and the transmission of these genes between different bacterial pathogens. The increasing number of capsule gene clusters that have been cloned and analysed from a range of both gram-negative and -positive bacterial pathogens is beginning to shed light on processes by which capsule diversity has been achieved. This review will focus principally on the generation of capsule gene diversity in *E. coli*, *N. meningitidis* and *S. pneumoniae*.

## 2 The Diversity of *E. coli* Capsules

*Escherichia coli* produces more than 80 chemically and serologically distinct capsules, called K antigens (JANN and JANN 1992). These capsules have been separated into groups 1, 2, 3 and 4 (Table 1) on the basis of capsule gene organisation, regulation of expression and the biosynthetic mechanism (WHITFIELD and ROBERTS 1999). The primary structures of certain of these capsular polysaccharides are identical to capsules expressed by other gram-negative pathogens, suggesting possible lateral transfer of genes encoding polysaccharide biosynthesis between these bacteria. Typically, group 1 capsules resemble those expressed by *Klebsiella* and *Erwinia* species, whilst group 2 and group 3 capsules have structural homologues in *N. meningitidis* and *Haemophilus influenzae* (Table 1). Group 4 capsules comprise a subset of amino-sugar-containing capsules that were previously characterised as either group 1B or O-antigen capsules. A more detailed description of capsule diversity can be found in WHITFIELD and ROBERTS (1999).

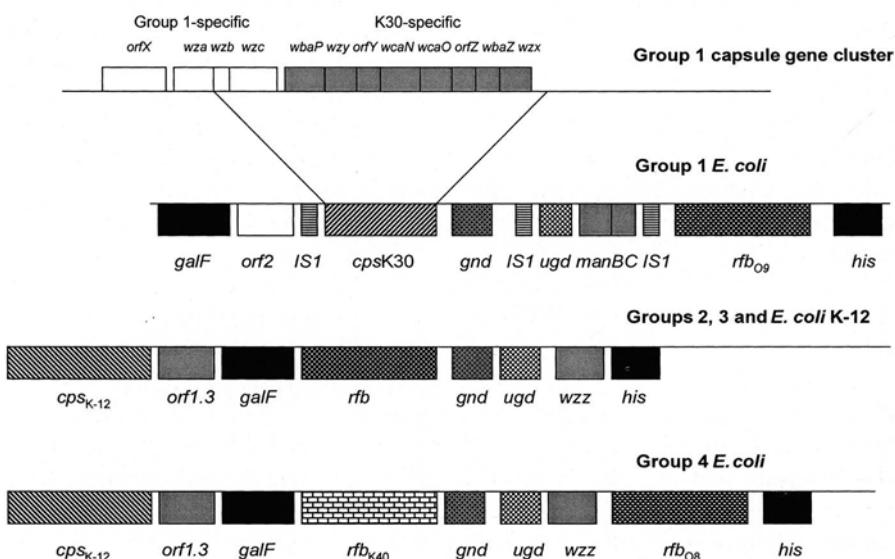
**Table 1.** Classification of *Escherichia coli* capsules

Characteristic	Group	1	2	3	4
Former K-antigen group	IA	II	III or III	IB (O-antigen capsules)	
Co-expressed with O serogroups	Limited range (O8, O9, O20, O101)	Many	Many	Often O8, O9 but sometimes none	
Co-expressed with colanic acid	No	Yes	Yes	Yes	Yes
Thermostability	Yes	No	No	Yes	Yes
Terminal lipid moiety	Lipid A-core in K <sub>LPS</sub> ; unknown for capsular K-antigen	α-glycerophosphate	α-glycerophosphate?	Lipid A-core in K <sub>LPS</sub> ; unknown for capsular K-antigen	
Reducing terminus	Reducing terminus	Non-reducing terminus	Non-reducing terminus?	Reducing terminus?	
Polymerization system	Wzy-dependent	Processive	Processive?	Wzy-dependent	
Trans-plasma membrane export	Wzx (PST2)	ABC-2 exporter?	ABC-2 exporter?	Wzx (PST2)	
Translocation proteins	Wza, Wzc	KpsD, KpsE (KpsF?)	KpsD, KpsE?	Unknown	
Elevated levels of CMP-Kdo synthetase	No	Yes	No	No	
Genetic locus	<i>cps</i> near <i>his</i> and <i>rfb</i>	<i>kps</i> near <i>serA</i>	<i>kps</i> near <i>serA</i>	<i>rfb</i> near <i>his</i>	
Thermoregulated – (i.e. not expressed below 20°C)	No	Yes	No	No	
Positively regulated by the Rcs system	Yes	Serotype K30	Serotypes K1, K5	Serotypes K10, K54	No
Model system		<i>Klebsiella, Erwinia</i>	<i>Neisseria, Haemophilus</i>	<i>Neisseria, Haemophilus</i>	Serotypes K40, O111
Similar to					Many genera

## 2.1 The Genetics of *E. coli* of Group 1 and 4 Capsules

The gene clusters for group 1 and group 4 capsules map near the *his* operon at 45 min on the *E. coli* chromosome. This region of the *E. coli* chromosome is highly polymorphic (Fig. 1) with gene clusters for the biosynthesis of a number of cell surface polysaccharide antigens including colanic acid (*cps*<sub>K-12</sub>), group 1 capsules (*cps*) and group 4 capsules and O-antigens (*rfb*). The particular genome organisation is dependent on the specific repertoire of expressed cell surface polysaccharides. In strains expressing group 1 capsules, the *cps* gene cluster is inserted between the *galF* and *gnd* genes at a site occupied by the *rfb* gene cluster in *E. coli* K-12 strains or those expressing group 2, 3 and 4 capsules (Fig. 1). In group 1-expressing strains the *rfb* gene cluster encoding for either the O8 or O9 antigen is atypically located 3' to the *gnd* gene proximal to the *his* operon (Fig. 1) (ARMOR and WHITFIELD 1997; DRUMMELSMITH et al. 1997). Additional genes encoding enzymes for the biosynthesis of specific nucleotide sugar precursors necessary for the biosynthesis of the particular group 1 polysaccharide may be located between the *gnd* gene and the O8/O9 *rfb* gene cluster (Fig. 1).

Detailed molecular analysis of *E. coli* group 1 capsule gene clusters has identified a conserved genetic organisation and revealed information on the acquisition of these capsule determinants in *E. coli* (DRUMMELSMITH and WHITFIELD 1999; RAHN et al. 1999). The K30 capsule gene cluster is the archetypal and most-studied group 1 capsule gene cluster and consists of 16 kb with 12 open reading frames (Fig. 1). The organisation of K30 capsule gene cluster is in keeping with a single



**Fig. 1.** The *his*-region of the *Escherichia coli* chromosome. The location of the group 1 (K30) and group 4 (K40) capsule gene clusters relative to *galF* *his* and the colanic acid biosynthesis gene cluster (*cps*<sub>K-12</sub>) is depicted. The organisation of the K30 gene cluster is shown in detail

transcriptional unit with the promoter located 5' to the *orfX* (Fig. 1). The identification of likely *cis*-acting regulatory sequences in this region is supportive of this notion. The first four genes *orfX-wza-wzb-wzc* are conserved between different group 1 capsule gene clusters (RAHN et al. 1999). The Wza, Wzb and WzC proteins are involved in the translocation of group 1 polysaccharides, whilst no specific role has yet been assigned to the protein encoded by *orfX* (DRUMMELSMITH and WHITFIELD 1999). The K30-specific genes *wbaP-wzx* (Fig. 1) are typical of a Wzy-independent O-antigen (*rfb*) gene cluster and explain the appearance of K30<sub>LPS</sub> as well as K30 capsules in such strains (DRUMMELSMITH and WHITFIELD 1999). Therefore, it is the group 1 translocation genes *wza*, *wzb* and *wzc* which provide the discriminatory feature that distinguishes group 1 capsule gene clusters from those for Wzy-independent O antigens. The mechanism by which group 1 capsule diversity is achieved and specific biosynthetic genes are inserted 3' to the *wzc* gene is as yet unclear.

Comparison of the group 1 capsule genes of *E. coli* with those of *K. pneumoniae* shows a high degree of conservation. The *wza*, *wzb* and *wzc* genes are 99.5% identical between different *E. coli* and *K. pneumoniae* isolates (RAHN et al. 1999) and in the case of the *E. coli* K30 and *K. pneumoniae* K20 capsules, which are structurally identical, the homology extends into the biosynthetic genes (RAHN et al. 1999). The high similarity between group 1 capsule gene clusters in *E. coli* and *K. pneumoniae* together with the identification of *orfX* in *E. coli* strains expressing group 1 capsules and *K. pneumoniae* but absent in laboratory K-12 *E. coli* strains (Fig. 1) suggests lateral gene transfer of these genes between these two micro-organisms. Analysis of the sequence 5' to group 1 capsule gene clusters in a number of *E. coli* serotypes identified a number of insertion sequences including IS1, IS2, IS30 and IS600 (DRUMMELSMITH et al. 1997; RAHN et al. 1999). The identification of IS elements adjacent to most of the *E. coli* group 1 capsule gene clusters and the rarity of IS elements 5' to capsule gene clusters in *Klebsiella* strains suggests that IS elements may have mediated the initial transfer of the capsule genes to *E. coli* and could be important for the exchange of capsule genes between strains. The flanking of the *ugd* gene and duplicated *manBC* genes in *E. coli* K30 strains by IS1 elements (Fig. 1) adds support to this notion of a role for IS elements in group 1 capsule gene evolution.

In this scenario the region between *galF* and *his* (*orf2-cps-ugd-rfb-gnd*) in *Klebsiella* has been introduced into and replaced the corresponding region (*cps<sub>K-12</sub>-rfb-gnd-ugd-wzz*) in a progenitor *E. coli* strain. This would result in the loss of the *cps<sub>K-12</sub>* gene cluster and the *wzz* gene in *E. coli* strains expressing group 1 capsules (Fig. 1). This hypothesis is supported by a number of observations. Firstly, the expression of group 1 capsules and colanic acid slime are mutually exclusive (DRUMMELSMITH et al. 1997). Secondly, strains expressing group 1 capsules lack a *wzz* gene (DODGSON et al. 1996; DRUMMELSMITH et al. 1997). Thirdly, group 1 capsules are expressed with a limited repertoire of O antigens (O8, O9, O9a, O20 and O101) (JANN and JANN 1997), which is in keeping with the limited range of O-antigens expressed by *Klebsiella* species (KELLY et al. 1995). Fourthly, there is high sequence diversity of the *gnd* gene within *E. coli* strains (NELSON and

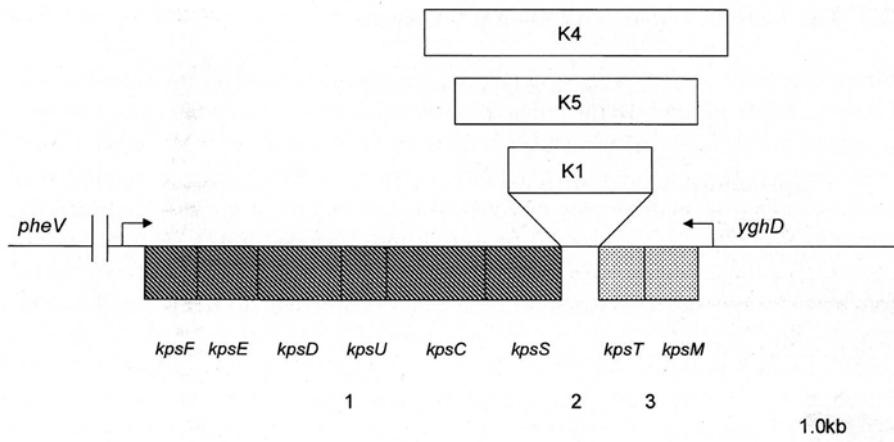
SELANDER 1994). This diversity is probably a consequence of numerous recombination events in this region of the chromosome involved in the co-transfer of *rfb* and *cps* gene clusters and it has been suggested that the *gnd* gene was transferred from *Klebsiella* into *E. coli* (RAHN et al. 1999).

Whilst it is likely that lateral gene transfer of *cps* genes between *Klebsiella* and *E. coli* strains has taken place, that alone cannot explain the evolution and diversity of group 1 capsules in *E. coli*. For instance, how does one explain the expression of group 1 capsules in *E. coli* for which there are no corresponding homologues in *Klebsiella*? It is likely that the situation is more complex with bi-directional exchange of *cps* genes between organisms other than merely *E. coli* and *Klebsiella* species.

The best-studied of the group 4 capsules is the *E. coli* K40 capsule. Analysis of the *galF-his* region of the chromosome shows that the K40 capsule genes (*rfb<sub>K40</sub>*) are located adjacent to the *cps<sub>K-12</sub>* gene cluster for colanic acid biosynthesis in the location normally occupied by the Wzy-dependent O antigen (*rfb*) gene cluster (Fig. 1) (ARMOR and WHITFIELD 1997). The accompanying O8/O9 *rfb* gene cluster is located between *gnd* and *his*, such that the genome organisation is *cps<sub>K-12</sub>-galF-rfb<sub>K40</sub>-gnd-ugd-wzz-rfb<sub>O8</sub>-his* (Fig. 1). This chromosome organisation is conserved amongst group 4 strains and there are no IS elements in the *gnd-rfb* region in strains expressing group 4 capsules (ARMOR and WHITFIELD 1997). The divergence of the *gnd-rfb* region between strains expressing group 1 and group 4 capsules indicates that the evolution of these two capsule lineages in *E. coli* cannot be accounted for by the single acquisition event of a particular capsule gene cluster. Rather, the polymorphic nature of this region of the chromosome suggests that a number of accretion and deletion events likely to be mediated in part by IS elements have taken place. The conservation of the *ugd* gene in both group 1 and group 4 capsule-expressing strains suggests that this gene was present in the progenitor organism from which lineages have been derived, that the starting gene order may have been *gnd-ugd-wzz*, and that the loss of *wzz* gene by IS-mediated rearrangements was the first step in the differentiation of group 1 and group 4 capsule-expressing strains.

## 2.2 The Genetics of *E. coli* Group 2 Capsules

The cloning and analysis of a large number of *E. coli* group 2 capsule gene clusters established that group 2 capsule gene clusters have a conserved modular genetic organisation consisting of three regions, 1, 2 and 3 (Fig. 2) (ROBERTS et al. 1986, 1988; ROBERTS 1996). This modular organisation, first demonstrated with *E. coli* group 2 capsule gene clusters, would now appear to applicable to capsule gene clusters from other bacteria (ROBERTS 1996). Regions 1 and 3 are conserved in all of the group 2 capsule gene clusters analysed and encode the eight Kps proteins involved in the transport of group 2 polysaccharides from their site of synthesis on the inner face of the cytoplasmic membrane onto the cell surface (Fig. 2). Homologues to Kps proteins are encoded by the capsule gene clusters of other gram-negative bacteria, indicating that there are conserved steps in the transport of group



**Fig. 2.** Group 2 capsule gene clusters in *Escherichia coli*. The conserved regions 1 and 3 encoding the Kps proteins are shown with a representative selection of different serotype-specific region 2 cassettes. The arrows denote the two major promoters that control group 2 capsule expression. The two vertical lines indicate that the linkage between group 2 capsule gene clusters and *pheV* has not been shown at the level of nucleotide sequence, only by alignment of restriction maps of the flanking DNA

2 capsular polysaccharides in these different gram-negative bacteria. For a detailed analysis of the roles of the Kps proteins in the transport of group 2, the reader is referred to the recent review by WHITFIELD and ROBERTS (1999). Region 2 is serotype specific and encodes enzymes for the polymerisation of the polysaccharide molecule and, where necessary, for the biosynthesis of the specific monosaccharide components that make up the polysaccharide. The size of the specific region 2 is variable and reflects in part the complexity of the polysaccharide to be synthesised (Fig. 2) (ROBERTS 1996). The region 2 DNA of the K5 and K1 capsule gene clusters have a high (66%) A + T content compared with that of regions 1 (50%) and 3 (57%) (ROBERTS 1996). This is typical of genes that encode enzymes for polysaccharide biosynthesis (ROBERTS 1995) and would suggest that group 2 capsule diversity has been achieved in part through the acquisition of different region 2 sequences. Amplification by PCR of sequences between regions 1 and 2 and between regions 2 and 3 from a number of group 2 capsule gene clusters failed to find any evidence for insertion sequences or site-specific recombination events playing a role in this process (ROBERTS 1996). Rather, the acquisition of new region 2 sequences may occur through homologous recombination between the flanking regions 1 and 3 of an incoming and resident capsule gene cluster. The observation that the 3' ends of the *kpsS* and *kpsT* genes which flank either side of region 2 (Fig. 2) show the greatest divergence amongst the conserved region 1 and 3 *kps* genes (ROBERTS 1996), would support this hypothesis for acquiring and losing region 2 sequences. The mechanism by which region 2 diversity, and therefore the diversity of *E. coli* group 2 capsular polysaccharides, has been achieved is still unknown.

### 2.3 The Genetics of *E. coli* Group 3 Capsules

Group 3 capsule gene clusters map near *serA* and are allelic to group 2 capsule gene clusters, being located at the same site on the chromosome. The only group 3 capsule gene clusters that have been studied in detail are the K10 and K54 capsule gene clusters (PEARCE and ROBERTS 1995; RUSSO et al. 1998). Group 3 capsule gene clusters have a segmental gene organisation reminiscent of group 2 capsule gene clusters. There are two conserved regions 1 and 3 which flank a serotype-specific central region 2 (PEARCE and ROBERTS 1995; CLARKE et al. 1999). However, within this arrangement the organisation of the genes is different. In the group 3 capsule gene clusters region 1 contains four genes encoding homologues of the group 2 region 1 and 3 proteins, *KpsD*, *E*, *M* and *T*, whilst the group 3 region 3 is comprised of two genes which encode homologues of the group 2 region 1 proteins *KpsC* and *S* (RUSSO et al. 1998; CLARKE et al. 1999). The A + T ratio of the group 3 region 1 is 61%, which is significantly higher than the average 51% for the *E. coli* genome (BLATTNER et al. 1997) and that of region 1 (50%) and region 3 (57%) of the group 2 capsule gene clusters (CLARKE et al. 1999). This suggests that region 1 of the group 3 capsule gene clusters may have been acquired by a route different to that of regions 1 and 3 of the group 2 capsule gene clusters. The A + T ratio of the group 3 region 3 is 57%, which is lower than that for region 1, suggesting that regions 1 and 3 of group 3 capsule gene clusters may have been acquired separately in two independent lateral gene transfers.

Analysis of the nucleotide sequence 5' to the first gene of region 1, *kpsD<sub>K10</sub>*, revealed the presence of sequences highly homologous to the 5' end of the *kpsM* gene and the region 3 promoter of group 2 capsule gene clusters (RUSSO et al. 1998; CLARKE et al. 1999). The extent of nucleotide homology between the group 2 and group 3 region 1 promoter regions extends approximately 190bp into the 5' coding sequence of the *kpsM* gene. This DNA sequence does not show significant homology to the functional *kpsM<sub>K10</sub>* and *kpsM<sub>K54</sub>* genes located 3' to *kpsD<sub>K10</sub>*. The identification of a cryptic *kpsM* gene suggests that the groups 2 and 3 promoter regions were acquired from a common ancestor expressing a group 2 capsule. It is possible that the group 3 capsule cluster was derived by the insertion of foreign capsule genes 3' of the start of the *kpsM* gene in an existing group 2 capsule gene cluster.

A remnant of IS110 from *Streptomyces coelicolor* is present 53bp 3' to the cryptic *kpsM* gene in K10 and K54 capsule gene clusters, and it has been postulated that this IS element was involved in the mobilisation of the group 3 determinants into the progenitor group 2 strain (RUSSO et al. 1998). In addition to the IS110 sequence, a region of 99% homology to IS3 (TIMMERMAN and TU 1985) was identified 5' to the K10 region 1 promoter. This insertion element was not identified in the K54 capsule gene cluster and would appear to be specific to the K10 capsule gene cluster. IS elements have been implicated in the duplication of genes in the group I capsule locus of *E. coli* K30 (DRUMMELSMITH et al. 1997) and they have been found near the capsule genes of *Klebsiella pneumoniae* (WACHAROTAYANKUN et al. 1993). In addition, remnants of IS600 and IS630 elements may have been

involved in lateral transfer of a pathogenicity island into enteropathogenic and enterohaemorrhagic *E. coli* (PERNA et al. 1998). Conceivably, a block of capsule genes could be mobilised through transposition if they were flanked by IS elements. Although numerous IS3 elements are present on the *E. coli* chromosome (BLATTNER et al. 1997), a second IS3 within or flanking the K10 capsule gene cluster was not identified (CLARKE et al. 1999). However, a second flanking IS3 element could have been lost through subsequent recombination events. Alternatively, group 3 capsule genes could be transferred by homologous recombination between IS elements located in *E. coli* and in DNA from another organism.

Analysis of DNA sequence 3' to region 3 of the K10 capsule gene cluster identified the presence of a prophage related to retronophage φR73 and other CP4-like cryptic prophages found in *E. coli* K-12 (BLATTNER et al. 1997). Numerous virulence determinants have been associated with lysogenic bacteriophages (CHEETHAM and KATZ 1995), and CP4-like cryptic prophages have been implicated in the acquisition of the LEE pathogenicity island of enterohaemorrhagic *E. coli* strain, EDL933 (PERNA et al. 1998). Therefore, it is possible that bacteriophage transduction may have played a role in the acquisition of the K10 capsule gene cluster and other group 3 capsule gene clusters.

In summary, the acquisition of group 3 capsule genes is likely to have occurred by the insertion of new capsule genes into an existing group 2 capsule gene cluster. The remnants of *kpsM* and the region 3 promoter support this hypothesis. The origin of the primordial group 3 capsule gene cluster is unknown, although the low G+C content of region 1 (39%) is more in keeping with capsule genes from *Haemophilus influenzae* (A+T 37%). The identification of IS elements and prophage sequences flanking the group 3 capsule gene cluster suggest a role for these elements in the acquisition of group 3 capsule genes at this site on the *E. coli* chromosome.

## 2.4 Localisation of the Group 2 and 3 Capsule Gene Clusters on the *E. coli* Chromosome

Analysis of the sequence 5' to region 3 of the group 2 capsule gene clusters identified sequences common to both group 2 and K-12 strains. The common *E. coli* sequence began 1058bp 5' to the ATG of *kpsM* and corresponded to the 3' end of the *yghD* gene (Fig. 2). This is in keeping with the previous mapping of group 2 capsule gene clusters to 64 min on the *E. coli* chromosome proximal to *serA* (ØRSKOV and NYMAN 1974). The YghD protein is presumed to be an M-type component of a general protein secretion pathway. A cryptic tRNA gene with homology to the L-seryl-tRNA selenium transferase was identified 3' to the *yghD* gene (data not shown), and it is possible that this may represent the remnants of a tRNA gene that was important in the insertion of the *kps* gene cluster at this site. It has been shown that tRNA genes are associated with pathogenicity islands and are recognised as sites for the accretion of DNA via horizontal transfer (OCHMAN and BERGTHORSSON 1998). No other signatures indicative of IS elements or REP

sequences were present at this junction site to offer clues on the likely mechanism by which specific insertion of the group 2 capsule genes 3' to *yghD* might have occurred. The same junction site 3' to the *yghD* is present in group 3 strains, confirming the notion that group 3 capsule genes inserted into an existing group 2 capsule gene cluster.

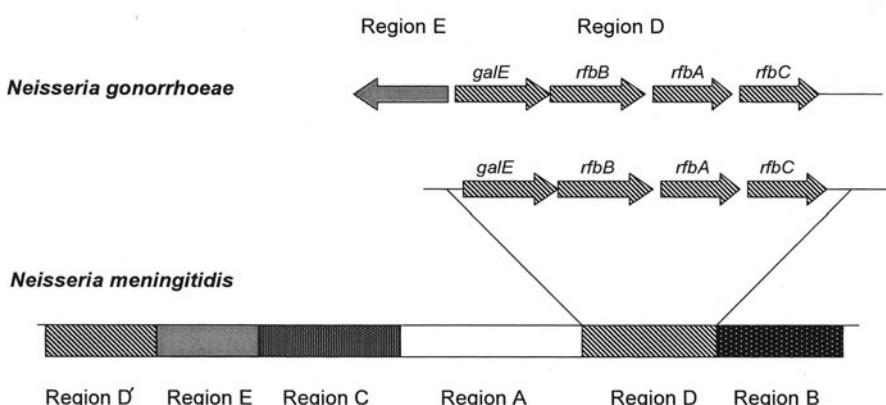
The *pheV* gene, coding for phenylalanine tRNA, is positioned 150bp 3' to *yghD* in *E. coli* K-12, suggesting that this gene should flank region 1 of group 2 and region 3 of group 3 capsule gene clusters (Fig. 2). To date, DNA homologous to the *pheV* gene has been detected in the sequenced chromosomal DNA flanking either group 2 or 3 capsule gene clusters. However, alignment of physical maps of the cosmid clones containing sequences flanking region 1 of the K1 capsule gene cluster show conserved restriction sites corresponding to the *pheV* region of the chromosome (data not shown). This confirms that group 2 capsule gene clusters (and by extrapolation group 3 capsule gene clusters) are inserted between *yghD* and *pheV*. The *pheV* tRNA gene has been reported as the site for the insertion of pathogenicity island (PAI) IV in *E. coli* uropathogenic strains (SVENSON et al. 1996). In addition, group 2 capsule determinants have been detected at the terminus of PAI IV (SVENSON et al. 1996). This suggests that group 2 capsule genes may have been introduced into uropathogenic *E. coli* on PAI IV. Analysis of the *E. coli* reference collection demonstrated that group 2 capsule genes are present in subgroups B2 and D together with other virulence genes *hly*, *sfa* and *pap*, with a high frequency for the co-existence of group 2 capsule gene clusters and the *pap* gene cluster (BOYD and HARTL 1998). Subgroups B2 and D have the largest genome size and it is likely that the acquisition of these multiple virulence factors is a consequence of PAIs. The group 2 capsule gene clusters were found in only four strains in subgroup A, which probably reflects the recent lateral transfer of these genes from a B2 or D strain (BOYD and HARTL 1998). Group 2 capsule genes were absent from 15 strains that comprise DEC collection of *E. coli* strains isolated from patients suffering from enteric *E. coli* infections (BOYD and HARTL 1998). This indicates that DEC strains require a set of virulence determinants different to those of uropathogenic *E. coli* and that *E. coli* strains do not carry PAIs for multiple disease categories. In other words, there are distinct lineages of *E. coli* which contain group 2 capsule gene clusters usually associated with other virulence factors necessary for urinary tract and extraintestinal infections (BOYD and HARTL 1998; PICARD et al. 1999).

### 3 Capsule Gene Clusters in *Neisseria meningitidis*

*N. meningitidis* is responsible for a number of life-threatening infections of man (KLEIN et al. 1996). Twelve meningococcal capsule serogroups have been described (ROBERTS 1996); with most infections being assigned to serogroups A, B, C, W135 and Y (JONES 1995). The capsule genes are located to a single *cps* locus and show a modular organisation consisting of five regions A–E with a partial duplication of

region D 5' to region C (Fig. 3) (FROSCH et al. 1991; HAMMERSCHMIDT et al. 1994; PETERING et al. 1996). Regions B and C encode proteins for the lipid modification and export of the capsular polysaccharide (FROSCH et al. 1991; FROSCH and MULLER 1993; HAMMERSCHMIDT et al. 1994), while region A is serogroup specific, encoding enzymes for the biosynthesis of the particular meningococcal polysaccharide (EDWARDS et al. 1994; CLAUS et al. 1997; SWARTLEY et al. 1998). Capsule switching from one serogroup to another can occur following DNA uptake by homologous recombination between conserved flanking regions (SWARTLEY et al. 1997). Analysis of region A from serogroups B, C, W135 and Y, all of which contain sialic acid, has shed light on the molecular divergence of these capsule gene clusters (CLAUS et al. 1997). The *sia A-C* genes encoding enzymes for the biosynthesis and activation of *N*-acetylneuraminic acid are conserved with divergence 3' to the *siaC* gene in the *siaD* gene encoding for the specific sialyltransferase (CLAUS et al. 1997). These four capsule serogroups were classified into two groups on the basis of the *siaD* gene and the nature of the polysaccharide molecule. The serogroup B and C sialyltransferase enzymes, which synthesise homopolymeric sialic acid-containing capsules, were homologous, whereas the corresponding enzymes from serogroups W135 and Y which synthesise heteropolymers containing sialic acid showed high homology (CLAUS et al. 1997). As such, variation by gene transfer in the *siaD* gene would convey specificity to the sialic acid-containing polysaccharide.

Mutations in the single large ORF in region E (Fig. 3) have no detectable phenotype (PETERING et al. 1996). Located within the capsule gene cluster is region D, containing the *galE* gene, together with homologues to the *rfbB*, *rfbA* and *rfbC* genes (Fig. 3) (HAMMERSCHMIDT et al. 1994). Mutations in the *galE* gene resulted in a truncated LOS, indicating that the GalE protein is functional whereas mutations in the *rfb* genes had no detectable phenotype and were not transcribed



**Fig. 3.** Capsule gene cluster in *Neisseria meningitidis* serogroup B and related sequences in *Neisseria gonorrhoeae*. The five regions of the group B capsule gene clusters are shown by hatched boxes. The genes within region D are shown in detail, but other genes apart from the single *orf* in region E are omitted for clarity. The conserved sequences and corresponding genes are shown in *N. gonorrhoeae*.

(HAMMERSCHMIDT et al. 1994). This mosaic-like arrangement with the insertion of the functional *gale* gene within the capsule gene cluster together with silent copies of *rfb* genes is curious and possibly reflects events involved in the acquisition of capsule genes by *N. meningitidis*. Analysis of the *Neisseria gonorrhoeae* chromosome identified a region E 5' to region D but without any intervening capsule genes (Fig. 3) (PETERING et al. 1996). There are two possible explanations for these observations: firstly, that the capsule genes have been lost from *N. gonorrhoeae*; secondly, that *N. meningitidis* acquired the capsule gene cluster at the D-E locus and that *N. gonorrhoeae* never expressed a capsule. The latter hypothesis is more tempting for a number of reasons. The G+C ratio of the *cps* gene cluster in *N. meningitidis* is 30%, which is considerably lower than that for neisserial species (EDWARDS et al. 1994), suggesting the acquisition of the capsule gene by horizontal transfer from another organism. This event could be a critical factor in the emergence of *N. meningitidis* from a progenitor unencapsulated gonococcal ancestor. Analysis of the chromosomes of *N. meningitidis* and *N. gonorrhoeae* identified three regions at which DNA specific to *N. meningitidis* had been inserted, with the capsule genes being located to the first of these regions (TINSLEY and NASSIF 1996). Comparison of the *N. meningitidis* group B capsule gene cluster with that of *E. coli* K1 identified differences in gene organisation and nucleotide sequence, which suggests that these capsule gene clusters do not have a common origin (STEENBERGEN and VIMR 1990; FROSCH et al. 1991; EDWARDS et al. 1994). The observations that there are functional differences in the sialyltransferase enzymes from these two organisms and that there are differences in the regulation of sialic acid metabolism support the notion that these capsule gene clusters, whilst encoding for the same capsular polysaccharide, may have different origins.

Regulation of capsule expression is important in the pathology of invasive meningococcal disease. For maximising attachment and invasion of epithelial cells it is preferable to reduce encapsulation, while during the bacteraemic stages of infection encapsulation will be important to confer resistance to host defences. Capsule expression in *N. meningitidis* is modulated in at least three ways: Firstly, phase variation through a poly(dC) repeat, which introduces a +1/-1 frameshift in the *siaD* gene causing premature translation (HAMMERSCHMIDT et al. 1996b); secondly, Rho-dependent intracistronic transcription termination in the *siaD* gene (LAVITOLA et al. 1999); thirdly, by the site-specific insertion of IS1301 in the *siaA* gene (HAMMERSCHMIDT et al. 1996a). The consequence of the insertion of IS1301 in the *siaA* gene is to abolish capsule expression and sialylation of the LOS molecules, both of which will promote interaction with epithelial cells.

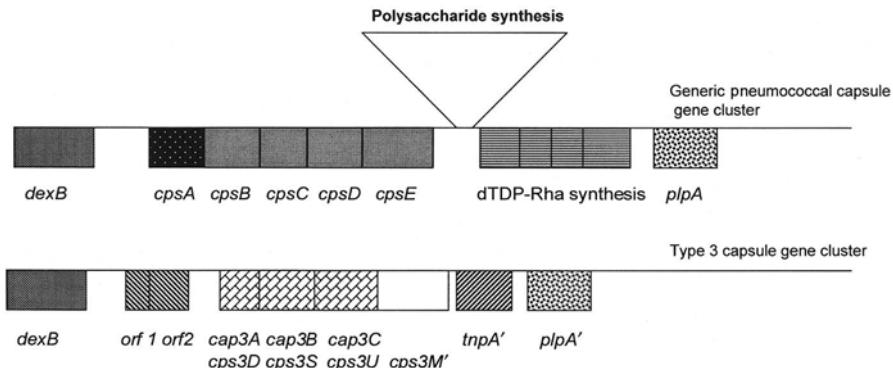
#### 4 Capsule Diversity in *Streptococcus pneumoniae*

For nearly 70 years since the pioneering experiments of Griffith on the transformation of avirulent unencapsulated *S. pneumoniae* to encapsulation and virulence,

it has been known that the expression of a capsule is an essential virulence factor in the pathogenesis of pneumococcal disease (GRIFFITH 1928). To date, 90 capsular serotypes have been described for *S. pneumoniae*, a number of which are included in pneumococcal vaccines. The striking diversity in pneumococcal capsular polysaccharide structure raises questions about the mechanisms by which this has been achieved and the selective pressures which drive the process of serotype diversity. The natural genetic transformation of *S. pneumoniae* means that there is the opportunity for recombination-mediated genetic plasticity and the intra- and interspecies movement of DNA (CLAVERTYS et al. 2000). It is likely that this process will contribute to the diversity of capsular serotypes and evolution of new capsular structures.

#### 4.1 Capsule Gene Clusters in *Streptococcus pneumoniae*

To date, 13 pneumococcal capsule gene clusters have been cloned and analysed in detail (GARCIA et al. 2000). With the exception of the type 37 capsule gene cluster, pneumococcal capsule gene clusters are inserted at the same site on the chromosome between the *dexB* and *plpA* (*aliA*) genes (CAIMANO et al. 1997; GARCIA et al. 2000). In keeping with group 2 capsule gene clusters from *E. coli*, there appears to be a modular-like organisation consisting of conserved genes flanking serotype-specific genes (GARCIA et al. 2000). The first gene *capA/cpsA* is conserved in all capsule gene clusters analysed to date, apart from the type 3 capsule gene cluster (Fig. 4). Based on an analysis of the predicted amino acid sequence of the CapA/CspA protein, it has been suggested that this protein is involved in the regulation of capsule gene expression, although this function has not been demonstrated biochemically (CAIMANO et al. 1997; GARCIA et al. 1997, 2000). Downstream from the *capA/cpsA* gene there are four additional genes (*cap/cpsBCDE*) that are conserved (Fig. 4). The CpsB, C and D proteins regulate capsule expression via auto-tyrosine phosphorylation of the CpsD protein, with CpsC being required for this reaction with CpsB effecting the dephosphorylation of CpsD (MORONA et al. 2000). The CpsE protein is likely involved in the initiation of polysaccharide biosynthesis. On the basis of differences in the *cap/cpsBCDE* genes it is possible to subdivide pneumococcal capsule gene clusters into two groups (MORONA et al. 1999). In the case of the type 3 capsule gene cluster, which in many ways is atypical, sequences 5' to the capsule-specific genes are also homologous to those flanking the type 1, 14 and 19F capsule gene clusters. However, in the case of type 3 the *orfs* contain multiple mutations, indicating that these sequences are not essential for type 3 capsule expression. Capsule gene clusters are transcribed from a single promoter 5' to the *cap/cpsA* gene, apart from the type 3 capsule gene cluster, in which the promoter is 5' to the *cap3A* gene (CAIMANO et al. 1997; GARCIA et al. 1997, 2000). The identification of genes located outside the capsule locus which encode proteins for the synthesis of precursors for capsule biosynthesis (HARDY et al. 2000) raises questions about the coordinate regulation of these genes and the interplay between basic metabolic pathways and the elaboration of a polysaccharide capsule.



**Fig. 4.** The generic pneumococcal capsule gene cluster and type 3 capsule gene cluster. The *top line* denotes the generic pneumococcal capsule gene cluster with the conserved *csp* genes (see text for details). The site of insertion of the serotype-specific genes is shown. The dTDP-rhamnose (dTDP-Rha) synthesis genes that are found 3' of the type 1, 2, 19F, 19A, 19B, 19C and 23F capsule gene clusters are shown. *Below* is depicted the type 3 capsule gene cluster with the two different nomenclatures *cps/cap*. For a comprehensive depiction of all of the pneumococcal capsule gene clusters the reader is referred to GARCIA et al. (2000)

#### 4.2 The Role of Transposition and Recombination in the Emergence of *S. pneumoniae* Capsule Gene Clusters

The presence of conserved genes flanking the serotype-specific biosynthetic genes readily permits recombination following DNA uptake and the wholesale replacement of the resident capsule gene cluster. This process will explain the observed switching of capsule serotype by pathogenic strains of *S. pneumoniae* (CLAVERYS et al. 2000). However, careful analysis of a number of *cps/cap* gene clusters indicates a possible role for IS elements in the generation of capsule gene diversity in *S. pneumoniae*. Flanking either side of the type 1 capsule gene cluster there is a directly repeated copy of IS1167, while IS1202 is located 5' to the type 19F, 19A and 23F capsule gene clusters. The type 3 capsule gene cluster provides additional evidence for transposition in the acquisition of capsule gene clusters in *S. pneumoniae* (CAIMANO et al. 1997). Located 3' to the type 3 capsule gene clusters is an *orf* termed *tnpA* which has homology with a number of putative transposases from a range of gram-positive IS elements including IS1167 from *S. pneumoniae*. The *tnpA* *orf* lacks translational signals although it is transcribed as part of the type 3 capsule gene cluster (CAIMANO et al. 1997). In type 3 strains the *plpA* gene 3' to the capsule gene cluster has a 5' deletion removing the first 281 amino acids of the encoded PlpA protein (CAIMANO et al. 1997). This deletion of the *plpA* gene, together with the presence of *tnpA*, is strongly suggestive of some form of IS-mediated recombination event, probably reflecting an earlier transposition event in the generation of the type 3 capsule gene cluster. The analysis of the *cap3A* (also referred to as *cps3D*) and *cap1G* genes has shed further light on the transmission of capsule genes between *S. pneumoniae* and other encapsulated pathogens. Both

genes encode for a UDP-glucose dehydrogenase (UDP-GlcDH) which catalyses the production of UDP-GlcA a component sugar of the capsular polysaccharide. The *cap3A* and *cap1G* genes are only 65% identical, indicating that there must be two sources for this gene in the evolution of capsule gene clusters in *S. pneumoniae* (GARCIA et al. 2000). This low homology is reflected in the lack of recombination between these two genes during transformation experiments (BERNHIMER and WERMUNDSEN 1969). Comparison of the predicted amino acid sequence of Cap1G with other UDP-GlcDH enzymes identified a high level of homology to the KfiD protein of *E. coli* K5. In addition the *kfiD* gene has a low G + C content atypical for *E. coli* and has a codon usage in keeping with that of the *cap1G* gene (GARCIA et al. 2000). Overall, this suggests that the *kfiD* gene has been acquired in *E. coli* by lateral transfer from gram-positive bacteria, probably a *Streptococcus* species (MUNOZ et al. 1998). The mechanism by which this transfer may have occurred is still unknown. The observation that many capsule biosynthesis genes in *E. coli* have a disproportionate G + C ratio (ROBERTS 1996) might indicate that more of these genes may have been acquired from a progenitor with low G + C gram-positive bacteria.

In summary, it is clear that capsule switching in *S. pneumoniae* is mediated by homologous recombination following natural transformation. The grouping of the capsule genes at one site on the chromosome flanked by homologous DNA facilitates this process. This recombination-mediated genetic plasticity offers a mechanism by which capsule switching could occur in *S. pneumoniae*, particularly in response to changes in environmental stimuli. It would permit the movement of new capsule genes into a *S. pneumoniae* with a particular genetic background which conveys increased virulence or antibiotic resistance such that virulent isolates expressing different capsule serotypes emerge.

## 5 Capsule Gene Clusters in Other Gram-Positive Pathogens

There is homology between the serotype 14 capsule gene cluster of *S. pneumoniae* and the capsule gene cluster of serotype Ia and III in *Streptococcus agalactiae* (CHAFFIN et al. 2000). At this stage it is impossible to identify the derivation of these genes and establish their origin. Interestingly, the ability to switch between serotype Ia and serotype III in *S. agalactiae* is mediated by a single polymerase enzyme, indicating that a single gene can confer serotype specificity (CHAFFIN et al. 2000). In *Staphylococcus aureus* the type 1 capsule gene cluster has been assigned to a discrete genetic element of approximately 35kb, termed the *cap1* element (LEE 1995). This element is found only in type 1-expressing strains but is not allelic to the *cap* locus at which the *cap5* and *cap8* gene clusters map, although it is contained on the same 175-kb *Sma*I fragment (LEE 1995; SAU et al. 1997). The nature of the *cap1* element is unknown,

although it would appear not to be a temperate bacteriophage (LEE 1995). The *cap5* and *cap8* gene clusters have a conserved genetic organisation with conserved flanking regions either side of a serotype-specific central region (SAU et al. 1997). The origin and evolution of the *cap5/cap8* locus is currently unknown. In *Staphylococcus epidermidis* the expression of the cell surface polysaccharide adhesin PIA is mediated by the incision and precise excision of the insertion sequence IS256 into the PIA encoding *ica* genes (ZIEBUHR et al. 1999). This is a further example of the role of IS elements in modulating the expression of cell surface polysaccharide structures.

## 6 Concluding Remarks

As more capsule gene clusters from a range of microbial pathogens are analysed, insights into the evolution and spread of these genes is becoming clearer. A unifying theme is the conservation within a given species of the genes encoding capsular polysaccharide export functions, with capsule specificity being imparted by serotype/group-specific cassettes. In many cases these cassettes are inserted between flanking regions to permit capsule switching by homologous recombination. Conservation in the functional steps involved in the assembly and export of capsular polysaccharides is reflected by homology between proteins involved in these processes across broad taxonomic boundaries. Analysis of A + T ratios of capsule gene clusters offers clues to likely progenitor organisms from which the genes may have been acquired, but definitive answers await to be found. One possibility is that *S. pneumoniae* (or a similar high A + T progenitor), with its ability to take up DNA across species barriers, may be the engine room which has driven capsule diversity. Subsequently, capsule biosynthesis genes have become disseminated across bacterial groupings. Whilst this is an appealing notion, it cannot in itself account for all of the observed capsule diversity.

The identification of IS elements within and flanking capsule gene clusters suggests a role for these elements in the evolution and spread of the complex gene clusters. In addition, it is evident that IS elements may moderate the control of capsule expression, permitting invasion of bacteria into host tissue. There is some evidence for bacteriophage-mediated transduction of capsule genes in *E. coli* and for their association with PAIs. The increasing availability of genome sequences will enhance the comparative genomics and allow a clearer picture of the evolution of capsule gene clusters to emerge.

*Acknowledgements.* The authors wish to thank Dr. Chris Whitfield for helpful discussions. The work in the laboratory of Ian Roberts is supported by the BBSRC, Wellcome Trust and the Lister Institute for Preventive Medicine. Lillian Ebah gratefully acknowledges a Commonwealth PhD Scholarship.

## References

- Armor PA, Whitfield C (1997) Molecular and functional analysis of genes required for the expression of group 1B K antigens in *Escherichia coli*: characterisation of the *his*-region containing gene clusters for multiple cell surface polysaccharides. *Mol Microbiol* 26:145–161
- Bernheimer HP, Wermundsen IE (1969) Unstable binary capsulated transformants in pneumococcus. *J Bacteriol* 98:1073–1079
- Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–1462
- Boyd EF, Hartl DL (1998) Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *J Bacteriol* 180:1159–1165
- Caimano MJ, Hardy GG, Yother J (1997) Capsule genetics in *Streptococcus pneumoniae* and a possible role for transposition in the generation of the type 3 locus. *Microb Drug Resist* 4:11–23
- Chaffin DO, Beres SB, Yim HY, Rubens CE (2000) The serotype of type Ia and III group B streptococci is determined by the polymerase gene within the polycistronic operon. *J Bacteriol* 182:4466–4477
- Cheetham BF, Katz ME (1995) A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Mol Microbiol* 18:201–208
- Clarke BR, Pearce R, Roberts IS (1999) Genetic organisation of the *Escherichia coli* K10 capsule gene cluster: identification and characterisation of two conserved regions in group III capsule gene clusters encoding polysaccharide transport functions. *J Bacteriol* 181:2279–2285
- Claus H, Vogel U, Muhlenhoff M, Gerardy-Schahn R, Frosch M (1997) Molecular divergence of the *sia* locus in different serogroups of *Neisseria meningitidis* expressing polysialic acid capsules. *Mol Gen Genet* 257:28–34
- Claverys J-P, Prudhomme M, Mortier-Barriere I, Martin B (2000) Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity. *Mol Microbiol* 35:251–259
- Costerton JW, Stewart, PS Greenberg EP (1999) Bacterial biofilms: a common cause of persistent infections. *Science* 284:1318–1322
- Dodgson C, Amor P, Whitfield, C (1996) Distribution of the *rol* gene encoding the regulator of lipo-polysaccharide O-chain length in *Escherichia coli* and its influence on the expression of group I capsular antigens. *J Bacteriol* 178:1895–1902
- Drummelsmith J, Whitfield C (1999) Gene products required for cell surface expression of the capsular form of the group 1 K antigen in *Escherichia coli* (O9a:K30). *Mol Microbiol* 31:1321–1332
- Drummelsmith J, Amor PA, Whitfield C (1997) Polymorphism, duplication and IS-1-mediated rearrangement in the *his-rfb-gnd* region of *Escherichia coli* strains with group IA capsular-K-antigens. *J Bacteriol* 179:3232–3238
- Edwards U, Muller A, Hammerschmidt S, Gerady-Schahn R, Frosch M (1994) Molecular analysis of the biosynthesis pathway of the alpha-2,8 polysialic acid capsule by *Neisseria meningitidis* serogroup B. *Mol Microbiol* 14:141–149
- Finne J (1982) Occurrence of unique polysialosyl carbohydrate units in glycoproteins of developing brain. *J Biol Chem* 257:11966–11970
- Frosch M, Müller A (1993) Phospholipid substitution of capsular polysaccharides and mechanisms of capsule formation in *Neisseria meningitidis*. *Mol Microbiol* 8:483–493
- Frosch M, Edwards U, Bousset K, Kauße B, Weisgerber C (1991) Evidence for a common molecular origin of the capsule gene loci in gram-negative bacteria expressing group II capsular polysaccharides. *Mol Microbiol* 5:1251–1263
- Garcia E, Arrecubieta C, Munoz R, Mollerach M, Lopez R (1997) A functional analysis of the *Streptococcus pneumoniae* genes involved in the synthesis of the type 1 and type 3 capsular polysaccharides. *Microb Drug Resist* 3:73–88
- Garcia E, Llull D, Munoz R, Mollerach M, Lopez R (2000) Current trends in capsular polysaccharide biosynthesis of *Streptococcus pneumoniae*. *Res Microbiol* 151:429–435
- Griffith F (1928) The significance of pneumococcal types. *J Hyg* 27:113–159
- Hammerschmidt S, Birkholz C, Zahringer U, Robertson BD, van Putten J, Ebeling O, Frosch M (1994) Contribution of genes from the capsule gene cluster complex (*cps*) to lipo-oligosaccharide biosynthesis and serum resistance in *Neisseria meningitidis*. *Mol Microbiol* 11:885–896

- Hammerschmidt S, Hilse R, van Putten JP, Gerardy-Schahn R, Unkmeir A, Frosch M (1996a) Modulation of cell surface sialic acid expression in *Neisseria meningitidis* via a transposable genetic element. *EMBO J* 15:192–198
- Hammerschmidt S, Muller A, Sillmann H, Muhlenhoff M, Borrow R, Fox A, van Putten JP, Zollinger WD, Gerardy-Schahn R, Frosch M (1996b) Capsule phase variation in *Neisseria meningitidis* serogroup B by slipped strand mispairing in the polysialyl transferase gene (*siaD*): correlation with bacterial invasion and the outbreak of meningococcal disease. *Mol Microbiol* 20:1211–1220
- Hardy GG, Caimano MJ, Yother J (2000) Capsule biosynthesis and basic metabolism in *Streptococcus pneumoniae* are linked through the cellular phosphoglucomutase. *J Bacteriol* 182:1854–1863
- Jann B, Jann K (1997) Capsules of *Escherichia coli*. In: Sussman M (ed) *Escherichia coli* mechanisms of virulence. Cambridge University Press, Cambridge, pp 113–143
- Jann K, Jann B (1992) Capsules of *Escherichia coli*, expression and biological significance. *Can J Microbiol* 38:705–710
- Jennings HJ (1990) Capsular polysaccharides as vaccine candidates. *Curr Top Microbiol Immunol* 150:97–127
- Jones DM (1995) Epidemiology of meningococcal disease in Europe and the USA. In: Cartwright K (ed) *Meningococcal disease*. John Wiley and Sons, Chichester, pp 147–157
- Kelly RF, MacLean LL, Perry MB, Whitfield C (1995) Structure of the O-antigens of *Klebsiella* serotypes O2(2a, 2e), O2(2a, 2e, 2h) O2(2a, 2f, 2g), members of a family of related D-galactan O-antigens in *Klebsiella* spp. *J Endotoxin Res* 2:131–140
- Klein NJ, Ison CA, Peakman M, Hammerschmidt S, Frosch M, Heyderman RS (1996) The influence of capsulation and lipo-oligosaccharide structure on neutrophil adhesion molecule expression and endothelial injury by *Neisseria meningitidis*. *J Infect Dis* 173:172–179
- Lavitola A, Bucci C, Salvatore P, Maresca G, Bruni CB, Alifano P (1999) Intracistronic transcription termination in polysialyltransferase gene (*siaD*) affects phase variation in *Neisseria meningitidis*. *Mol Microbiol* 33:119–127
- Lee CY (1995) Association of the staphylococcal type-1 capsule-encoding genes with a discrete genetic element. *Gene* 29:115–119
- Lindahl U, Lidholt K, Spillmann D, Kjellen L (1994) More to heparin than anticoagulation. *Thrombin Res* 75:1–32
- Michalek MT, Bremer EG, Mold C (1988) Effect of gangliosides on activation of the alternative pathway of human complement. *J Immunol* 140:1581–1587
- Morona JK, Morona R, Paton JC (1999) Analysis of the 5' portion of the type 19A capsule locus identifies two classes of *cpsC*, *cpsD* and *cpsE* genes in *Streptococcus pneumoniae*. *J Bacteriol* 181:3599–3605
- Morona JK, Paton JC, Miller DC, Morona R. (2000) Tyrosine phosphorylation of CpsD negatively regulates capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. *Mol Microbiol* 35:1431–1442
- Moxon ER, Kroll JS (1990) The role of bacterial polysaccharide capsules as virulence factors. *Curr Top Microbiol Immunol* 150:65–86
- Munoz R, Garcia E, Lopez R (1998) Evidence for horizontal transfer from *Streptococcus* to *Escherichia coli* of the *kfiD* gene encoding the K5-specific UDP-glucose dehydrogenase. *J Mol Evol* 46:324–326
- Nelson K, Selander RK (1994) Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. *Proc Natl Acad Sci USA* 91:10227–10231
- Ochman H, Bergthorsson U (1998) Rates and patterns of chromosome evolution in enteric bacteria. *Curr Opin Microbiol* 1:580–583
- Ørskov I, Nyman K (1974) Genetic mapping of the antigenic determinants of two polysaccharide K antigens, K10 and K54, in *Escherichia coli*. *J Bacteriol* 120:43–51
- Pearce R, Roberts IS (1995) Cloning and analysis of gene clusters for production of the *Escherichia coli* K10 and K54 antigens: identification of a new group of *serA*-linked capsule gene clusters. *J Bacteriol* 177:3992–3997
- Perna NT, Mayhew GF, Pósfal G, Elliott S, Donnenberg MS, Kaper JB, Blattner FR (1998) Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* 66:3810–3817
- Petering H, Hammerschmidt S, Frosch M, van Putten J, Ison CA, Robertson BD (1996) Genes associated with meningococcal capsule complex are also found in *Neisseria gonorrhoeae*. *J Bacteriol* 178:3342–3345
- Picard B, Garcia JS, Gouriou S, Duriez P, Brahimi N, Bingen E, Elion J, Denamur E (1999) The link between phylogeny and virulence in *Escherichia coli* extra-intestinal infection. *Infect Immun* 67:546–553

- Rahn A, Drummelsmith J, Whitfield C (1999) Conserved organization in the gene clusters for expression of *Escherichia coli* group 1 K antigens: relationship to the colanic acid biosynthesis locus and the genes from *Klebsiella pneumoniae*. *J Bacteriol* 181:2307–2313
- Roberts IS (1995) Bacterial polysaccharides in sickness and in health. *Microbiol* 141:2023–2031
- Roberts IS (1996) The biochemistry and genetics of capsular polysaccharide production in bacteria. *Ann Rev Microbiol* 50:285–315
- Roberts IS, Mountford R, High N, Bitter-Suerman D, Jann K, Timmis KN, Boulnois GJ (1986) Molecular cloning and analysis of the genes for production of the K5, K7, K12, and K92 capsular polysaccharides of *Escherichia coli*. *J Bacteriol* 168:1228–1233
- Roberts IS, Mountford R, Hodge R, Jann KB, Boulnois GJ (1988) Common organization of gene clusters for the production of different capsular polysaccharides (K antigens) in *Escherichia coli*. *J Bacteriol* 170:1305–1310
- Russo TA, Wenderoth S, Carlino UB, Merrick JM, Lessie AJ (1998) Isolation, genomic organization, and analysis of the group III capsular polysaccharide genes *kpsD*, *kpsM*, *kpsT*, and *kpsE* from an extraintestinal isolate of *Escherichia coli* (CP9, O4/K54/H5). *J Bacteriol* 180:338–349
- Sau S, Bhasin N, Wann ER, Lee JC, Foster TJ, Lee CY (1997) The *Staphylococcus aureus* allelic genetic loci for serotype 5 and 8 capsule expression contain type-specific genes flanked by common genes. *Microbiol* 143:2395–2405
- Steenbergen S, Vimr E (1990) Mechanism of polysialic acid chain elongation in *Escherichia coli* K1. *Mol Microbiol* 4:603–611
- Svenson D, Bukanov NO, Berg DE, Welch RA (1996) Two pathogenicity islands in uropathogenic *Escherichia coli* J96:cosmid cloning and sample sequencing. *Infect Immun* 64:3736–3743
- Swartley JS, Martin AA, Edupuganti S, Liu L-J, Cieslak P, Perkins B, Wenger JD, Stephens DS (1997) Capsule switching of *Neisseria meningitidis*. *Proc Natl Acad Sci USA* 94:271–276
- Swartley JS, Liu L-J, Miller YK, Martin LE, Edupuganti S, Stephens DS (1998) Characterisation of the gene cassette required for biosynthesis of the (2’-6)-linked N-acetyl-D-Mannosamine-1-phosphate capsule of serogroup A *Neisseria meningitidis*. *J Bacteriol* 180:1533–1539
- Timmerman KP, Tu CPD (1985) Complete sequence of IS3. *Nucleic Acids Res* 13:2127–2139
- Tinsley CR, Nassif X (1996) Analysis of the genetic differences between *Neisseria meningitidis* and *Neisseria gonorrhoeae*: two closely related bacteria expressing two different pathogenicities. *Proc Natl Acad Sci USA* 93:11109–11114
- Wacharotayankun R, Arakawa Y, Ohta M, Tanaka K, Akashi T, Mori M, Katō N (1993) Enhancement of extracapsular polysaccharide synthesis in *Klebsiella pneumoniae* by RmpA2, which shows homology to NtrC and FixJ. *Infect Immun* 61:3164–3174
- Whitfield C, Roberts IS (1999) Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol Microbiol* 31:1307–1319
- Whitfield C, Valvano MA (1993) Biosynthesis and expression of cell surface polysaccharides in gram-negative bacteria. *Adv Micro Physiol* 35:135–246
- Ziebuhr W, Krimmer V, Rachid S, Lossner I, Gotz F, Hacker J (1999) A novel mechanism of phase variation of virulence of *Staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS256. *Mol Microbiol* 32:345–356

# Genome Plasticity in Pathogenic and Nonpathogenic Enterobacteria

U. DOBRINDT<sup>1</sup>, U. HENTSCHEL<sup>1</sup>, J.B. KAPER<sup>2</sup>, and J. HACKER<sup>1</sup>

1	Introduction . . . . .	157
2	Evolution of Enterobacteria . . . . .	158
3	<i>Escherichia coli</i> as a Model Organism . . . . .	160
3.1	The Flexible and the Core Gene Pool of <i>E. coli</i> . . . . .	160
3.2	<i>E. coli</i> Pathotypes . . . . .	161
3.3	Gene Pools of Pathogenic <i>E. coli</i> . . . . .	162
3.4	Pathogenicity Islands of Extraintestinal <i>E. coli</i> . . . . .	165
3.5	Mobile Genetic Elements of Intestinal <i>E. coli</i> . . . . .	166
4	Gene Pools in Other Enterobacteria . . . . .	168
5	Conclusion . . . . .	170
	References . . . . .	172

## 1 Introduction

The Enterobacteriaceae comprise a distinct phylogenetic cluster that share a common ancestor with other  $\gamma$ -Proteobacteria. This prokaryotic family comprises 40 genera with 200 species (GARRITY 2001). Within this division many representatives live in intimate association with hosts either as pathogens, as commensals or as symbionts (STEINERT et al. 2000). The best-studied examples are the enterobacteria, which comprise the clinically relevant human and animal pathogenic species *Escherichia coli*, *Salmonella enterica*, and *Shigella* spp., as well as *Yersinia pestis*, *Y. pseudotuberculosis* and *Y. enterocolitica*. The entomopathogenic bacterium *Photorhabdus luminescens* also belongs to the Enterobacteriaceae. This bacterium is unusual in that it combines a symbiotic life style within the guts of nematodes with a pathogenic life style that results in the killing of insects. Among the  $\gamma$ -Proteobacteria there are many species establishing symbiotic interactions

<sup>1</sup> Institut für Molekulare Infektionsbiologie der Universität Würzburg, Röntgenring 11, 97070 Würzburg, Germany

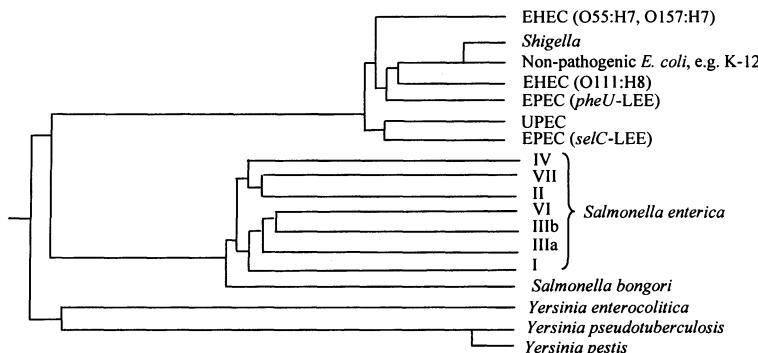
<sup>2</sup> Center for Vaccine Development and Department of Microbiology and Immunology, University of Maryland School of Medicine, 685 W. Baltimore Street, Baltimore, Maryland 21201, USA

mostly with invertebrate hosts, for example with insects, with bioluminescent squid and other marine invertebrates, and with nematodes. The genomes of several pathogens and symbionts have been sequenced recently and work is still in progress. In spite of the diverse manifestations of bacteria-host interactions, there are similar fundamental mechanisms that mediate the interaction and communication between the bacterial and eukaryotic partners (HENTSCHEL et al. 2000; STEINERT et al. 2000).

The enterobacteria are gram-negative, oxidase-negative, nonsporulating rods. They are facultative aerobes with relatively simple nutrient requirements that ferment sugars to a variety of end products. Also, enterobacteria produce many adhesive molecules such as pili and fimbriae which permit the attachment to surfaces. On an evolutionary scale, the emergence of enterobacteria is quite recent and coincides with the emergence of mammalian organisms. The original *Escherichia coli* were probably inhabitants of the mammalian intestinal flora. The selective pressures due to high bacterial densities ( $10^8$  bacteria/g biomass) and species competition (more than 500 bacterial species) combined with an ecological niche that is deprived in nutrients, were probably the driving forces to promote the expansion of strains into other ecological niches of a given host. Accordingly, while the commensal *E. coli* populate the large bowel, the pathogenic variants are found in additional locations such as the urogenital tract (UPEC), the mucosal surfaces of the gut (e.g. EHEC) and the blood (SEPEC). In this context it is interesting to note that all pathogenic strains of *E. coli*, *Shigella* spp., *Salmonella enterica* and *Yersinia* spp. have at least one, if not the majority of their essential virulence determinants on mobile genetic elements. From an evolutionary point of view, the enterobacteria can be considered a distinct phylogenetic group that contains certain physiological 'core properties' allowing their existence in or on mammalian hosts. With the acquisition of 'flexible properties', sometimes by single genetic events, they succeeded in conquering several other niches. This event led to the emergence of diseases over a relatively short evolutionary time affecting human beings and animals alike.

## 2 Evolution of Enterobacteria

The enterobacterial species *E. coli*, *Salmonella enterica* and *Shigella* spp. can be grouped into a phylogenetic cluster which diverged from other members of the  $\gamma$ -subdivision of gram-negative purple bacteria in coincidence with the emergence of mammalian organisms (WOESE 1987). The genus *Yersinia* is more distantly related but shares many genomic and phenotypic features with *E. coli* (see Fig. 1). While it has been estimated that *Y. enterocolitica* and *Y. pseudotuberculosis* diverged about 100 million years ago, recent data suggest that *Y. pestis* emerged relatively recently, about 1500–20,000 years ago, from *Y. pseudotuberculosis* (ACHTMAN et al. 1999). *Shigella* species are closely related to *Escherichia coli* and they could all be con-



**Fig. 1.** Evolution of enterobacteria from a common ancestor. The figure shows a rough schematic presentation of the evolutionary lineages according to recent publications on the phylogeny of pathogenic enterobacteria (ACHTMAN et al. 1999; BOYD et al. 1996; PUPO et al. 1997, 2000; REID et al. 2000). *EHEC*, Enterohaemorrhagic *E. coli*; *EPEC*, enteropathogenic *E. coli*; *UPEC*, uropathogenic *E. coli*; *LEE*, locus of enterocyte effacement. *Roman numbers* indicate different serovars of *Salmonella enterica*

sidered members of one species, as the distinction between the *E. coli* and *Shigella* spp. is based only on the pathogenic character of the bacteria. It has recently been described that *Shigella* evolved from *E. coli* within the past 35,000–270,000 years. Their chromosomal organization shares more than 90% homology, according to DNA-DNA reassociation experiments (PUPO et al. 2000). *E. coli* and *Salmonella* spp. are also closely related bacteria which are thought to have separated from a common ancestor about 140 million years ago (see also Fig. 1; OCHMAN and WILSON 1987; BOYD et al. 1996). The comparison of the chromosomes of *E. coli* K-12 and different *Salmonella* species showed that they share a stable core genome with similar gene orders and gene contents. Genetic maps of *E. coli* K-12 and *S. enterica* vs. *Typhimurium* are nearly 90% identical. However, the core genome is interspersed by many variable regions due to insertion/deletion events, and genome size in *S. enterica* and *E. coli* isolates has been shown to vary between isolates by up to 20% (RILEY and SANDERSON 1990; MCCLELLAND and WILSON 1998; WONG et al. 1999).

Enterobacterial populations are clonal. *E. coli* isolates can be grouped into particular clones that started to evolve under competition as distinct genetic types about 9 million years ago. A set of *E. coli* reference strains (ECOR strains) isolated from a variety of hosts and geographical locations has been compiled and analysed with respect to their genetic diversity and phylogeny. These strains are considered to represent the range of genotypic variation within the species *E. coli* (OCHMAN and SELANDER 1984; HERZER et al. 1990). Different clones of *E. coli* can be classified according to their pathotype and their host, as well as by using combinations of genotyping and phenotyping methods including multilocus enzyme electrophoresis (MLEE), multilocus sequence typing (MLST) and ribotyping (ACHTMAN and PLUSCHKE 1986; MASLOW et al. 1995; PUPO et al. 1997; WANG et al. 1997; BINGEN et al. 1998; BOYD and HARTL 1998; REID et al. 2000). These clones arose in parallel by the loss as well as by ordered gain of genetic information. The clones are maintained during adaptation to their respective niches and, because of horizontal

gene transfer, their further evolution is constantly in progress. Horizontal gene transfer plays an important role in the evolution of these enterobacterial species and results in very dynamic and diverse enterobacterial genome structures. In the case of the *E. coli* strain MG1655, it has been estimated that about 18% of the genome represents horizontally acquired sequences (LAWRENCE and OCHMAN 1998). The stepwise acquisition of 'foreign DNA' from distantly related organisms as well as the loss of DNA regions resulted in different metabolic and pathogenic features that distinguish the different genera, strains and pathotypes. Arising from a nonpathogenic *E. coli* ancestor, the loss of the *ompT* and *cadA* genes in combination with the acquisition of at least two pathogenicity islands and one virulence plasmid led to the evolution of pathogenic *Shigella* (NAKATA et al. 1993; MAURELLI et al. 1998; OCHMAN et al. 2000). Another detailed study of the evolutionary relationship between Shiga toxin-expressing *E. coli* strains showed that the nonpathogenic laboratory strain MG1655 and the EHEC strain EDL933 shared a common ancestor about 4.5 million years ago. The parallel gain and loss of mobile genetic elements, such as bacteriophages, plasmids and the LEE pathogenicity island, in different lineages of pathogenic *E. coli* enabled the evolution of separate clones which belong to different *E. coli* pathotypes (REID et al. 2000).

### 3 *Escherichia coli* as a Model Organism

#### 3.1 The Flexible and the Core Gene Pool of *E. coli*

The chromosome of *E. coli* K-12 is the best-studied microbial genome. Accordingly, that of the *E. coli* K-12 strain MG1655 was the first genome of an *E. coli* strain to be completely sequenced. Earlier results revealed an unexpected level of structural and genetic diversity among genomes of different *E. coli* strains. It has been demonstrated that the genome size within the species *E. coli* varies considerably between 4.6 and 5.5 Mb (BERGTHORSSON and OCHMAN 1998). The availability of three complete genome sequences of the nonpathogenic *E. coli* strain MG1655 and of two enterohaemorrhagic *E. coli* (EHEC) strains EDL933 and Sakai of the serogroup O157:H7 makes it now possible to compare in detail the genetic and structural variability of genomes of different *E. coli* strains (BLATTNER et al. 1997; HAYASHI et al. 2001; PERNA et al. 2001).

There is growing evidence that the genome of every *E. coli* strain can be considered to be composed of a universally present core of genes providing the backbone of genetic information which is generally present in all *E. coli* strains. This set also includes the 'minimal set' of genes needed for independent growth and replication of the bacterial cell, as previously described (HUTCHISON et al. 1999; HACKER and CARNIEL 2001). The core set of genes should be conserved in the majority of enterobacteria with respect to composition and position on the chromosome. In addition, the *E. coli* genome contains a flexible gene pool which is not common for

all strains and consists of an individual ‘assortment’ of strain-specific genetic information which may provide additional properties enabling these strains to adapt to special environmental conditions (see Table 1). Therefore, differences in genome size reflect the size variation of the flexible gene pool and are due mainly to the acquisition and loss of genomic DNA. Insertions and deletions of chromosomal regions ranging from a few base pairs to more than 100kb have been observed in *E. coli* (BLATTNER et al. 1997; HACKER and KAPER 2000; RODE et al. 1999; PERNIA et al. 2001). A surprisingly great proportion of the flexible gene pool consists of uncharacterized, unknown ORFs without any obvious function (in the case of the *E. coli* strain MG1655, about 38% of all ORFs have an unknown function). Another major constituent of the flexible gene pool is the group of accessory genetic elements, e.g. plasmids, transposons, insertion sequence elements, prophages and nonfunctional fragments thereof, as well as DNA elements termed genomic islands and islets. The latter DNA regions represent specific ‘foreign’ DNA entities and differ by their size. Genomic islands are at least 10kb. DNA regions smaller than 10kb are considered genomic islets (KAPER and HACKER 1999). These accessory genetic elements can be selfish (e.g. insertion sequence elements), they can provide a certain benefit or advantage for the host cell under specific conditions (e.g. genomic islands), or they can be both (e.g. resistance determinants on integrons or transposons, virulence-associated genes on bacteriophages or virulence plasmids). They can either be integrated into the chromosome at different locations or replicate independently as extrachromosomal elements. Several types of these elements can be laterally transferred and are present in probably all of the major bacterial phylogenetic groups, thus contributing to the inter- and intraspecies variability in genome content (KAPER and HACKER 1999; HACKER and KAPER 2000; OCHMAN et al. 2000).

### 3.2 *E. coli* Pathotypes

The flexible gene pool of *E. coli* encodes factors that play a role in host specificity and pathogenesis of the respective isolates. *E. coli* is the only bacterial species which is adapted to human beings as well as to many animals and which grows as a commensal organism but is also able to cause species-specific diseases (e.g. chicken

**Table 1.** Composition of bacterial genomes

Core gene pool		Flexible gene pool	
DNA elements	Encoded features	DNA elements	Encoded features
Chromosomes (plasmids)	Ribosomes Cell envelope Key metabolic pathways DNA replication Nucleotide turnover	Genomic islands Genomic islets Bacteriophages Plasmids Integrons Transposons	Pathogenicity Antibiotic resistance Secretion Symbiosis Degradation Secondary metabolism Restriction/modification Transposases/integrases

sepsis caused by O78-*E. coli* strains and the oedema disease of cattle caused by Shiga toxin-producing isolates of serotype O139) (ØRSKOV and ØRSKOV 1992; AARESTRUP et al. 1997; DHO-MOULIN and FAIRBROTHER 1999).

In human beings, different *E. coli* pathotypes are able to cause various infections. Five different *E. coli* pathotypes are involved in intestinal infections (for a detailed review see NATARO and KAPER 1998). Enterotoxigenic *E. coli* (ETEC) produce toxins similar to the cholera toxin while enteropathogenic *E. coli* (EPEC) cause specific types of diarrhoeal diseases, especially in developing countries. The enteroinvasive *E. coli* (EIEC) induce a disease very similar to shigellosis. An identical plasmid encoding an invasion system is present in genomes of EIEC and *Shigella* species. Different variants of enterohaemorrhagic *E. coli* (EHEC) are able to cause haemorrhagic colitis and the haemolytic uremic syndrome (HUS) in human beings. Interestingly, the source of EHEC bacteria is cows, which do not express the receptor for Shiga toxins (PRUIMBOOM-BREES et al. 2000). This is probably the reason why these bacteria do not cause disease in these animals. Enteroaggregative *E. coli* (EAEC) strains form another human-specific *E. coli* pathogroup. Besides intestinal infections, particular *E. coli* strains are able to cause extraintestinal infections in man. Uropathogenic *E. coli* (UPEC) are involved in infections of the bladder and the kidney. Similar *E. coli* strains are able to cause sepsis, and especially the K1 capsule type-producing strains are involved in newborn meningitis (JANN and JANN 1992; NATARO and LEVINE 1994). There are some indications that chronic diseases of the gut, such as Crohn's disease, are also associated with the occurrence of specific types of *E. coli* (BOUDEAU et al. 2001). It is our suggestion that the occurrence of the different pathotypes can be correlated with the presence of plasmids, phages and pathogenicity islands, which are part of the flexible gene pool and which direct the pathogenic potential of these strains.

### 3.3 Gene Pools of Pathogenic *E. coli*

Genomes of pathogenic *E. coli* strains are characterized by the presence of virulence-associated genes which are usually absent in nonpathogenic isolates. Therefore, this group of genes belongs to the flexible gene pool. Many virulence-associated genes are encoded by mobile genetic elements such as bacteriophages, plasmids or pathogenicity islands (MÜHLDORFER and HACKER 1994) and consequently can be frequently transferred to other bacteria. Specific combinations of virulence-associated genes are typical for specific *E. coli* pathotypes (for detailed descriptions of virulence-associated genes of extraintestinal and intestinal *E. coli* strains see also Sects. 3.4 and 3.5 of this chapter, as well as Chaps. 2 and 3 of this issue). These genomic regions can be easily identified due to their transferability and their sequence homology to known mobile DNA elements or virulence genes.

The analysis of the complete genome sequences of the *E. coli* strain MG1655 (4.639.221bp) and that of the EHEC strains Sakai (5.498.450bp) and EDL933 (5.529.376bp) reveals differences with respect to the genome size as well as to the number of translatable genes and structural RNA-encoding loci. Whereas for the

genome of the *E. coli* strain MG1655 4289 protein-encoding genes and 115 structural RNA loci have been determined, 5361 or 5349 translatable ORFs and 141 or 128 structural RNA loci can be found within the chromosomes of the EHEC strains Sakai and EDL933, respectively (BLATTNER et al. 1997; HAYASHI et al. 2001; PERNA et al. 2001).

Direct comparison of the complete chromosomal sequences of the K-12 strain MG1655 and of the O157:H7 EHEC strains Sakai and EDL933 shows that they share about 4.1Mb of DNA, which is similarly arranged in these strains with the exception of one inverted DNA region including the replication terminus that is present in the O157:H7 strain EDL933. This common chromosomal backbone is interrupted over the entire length by DNA regions which are specific for the O157:H7 strain Sakai or for the O157:H7 strain EDL933. For the latter, the presence of 0.53Mb of *E. coli* K-12 MG1655-specific sequences (528 genes) has been described, which is absent from the EDL933 genome. The chromosomes of the two EHEC strains harbour 1.39Mb (1.632 ORFs) and 1.34Mb (1.387 ORFs) of DNA, respectively, which are absent from the nonpathogenic *E. coli* strain MG1655 and include many known or putative virulence genes. In the strain EDL933, 35% of these ORFs have no known function. The chromosome of the nonpathogenic *E. coli* K-12 strain MG1655 contains one intact prophage ( $\lambda$ ), at least eight defective prophages and 42 insertion sequence elements or fragments thereof (BLATTNER et al. 1997). In the O157:H7 strain Sakai, 18 prophages or phage remnants have been identified, as well as six chromosomal regions which carry an integrase-encoding gene. Nine tRNA loci and the *ssrA* gene coding for the tmRNA are used as an integration site for bacteriophages or phage-like elements (HAYASHI et al. 2001). 18 chromosomal regions related to bacteriophages have also been identified in the *E. coli* strain EDL933; one of these represents a functional prophage whereas the others seem to be cryptic. With the exception of the prophage *rac* in *E. coli* MG1655 and that of CP-933R in *E. coli* EDL933, all other prophages of these two strains are strain specific. Four of the EDL933-specific DNA regions share some features of pathogenicity islands, as they contain a gene coding for a P4-like integrase and are associated with a tRNA gene (PERNA et al. 2001).

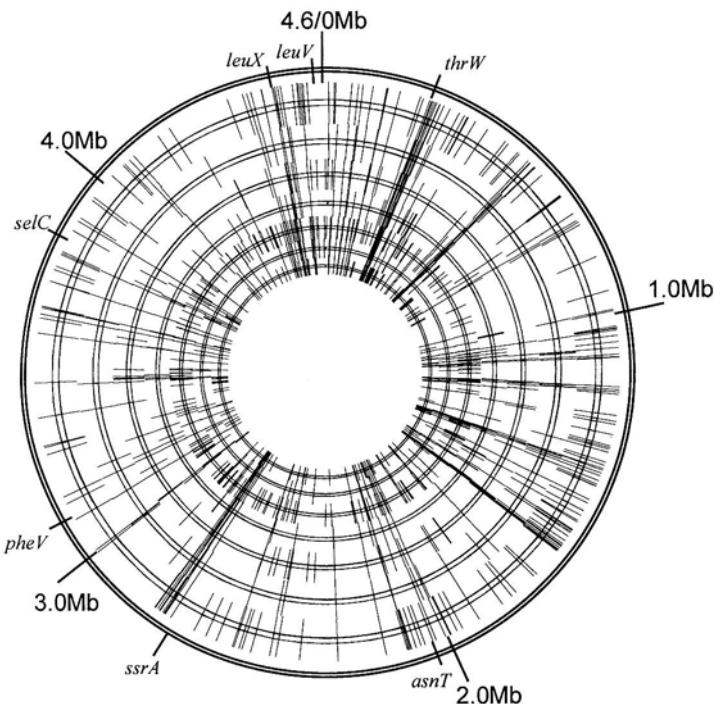
These data demonstrate the frequent presence of accessory genetic elements within *E. coli* genomes and the high variability with respect to number and type of these elements. The fact that several strain-specific DNA regions are associated with tRNA genes also underlines the fact that horizontal gene transfer is of great importance for the evolution of *E. coli* strains. Comparison of the genetic diversity of five nonpathogenic *E. coli* strains using *E. coli* MG1655-specific DNA macro-arrays also confirmed that the flexible gene pool consists of a high number of putative ORFs and mobile genetic elements and that their numbers are highly variable among different *E. coli* isolates (BLATTNER et al. 1997; OCHMAN and JONES 2000; PERNA et al. 2001). Nevertheless, many individual isolates of the species *E. coli* will have to be analysed in order to come to a reliable conclusion with respect to the composition of the flexible and the core gene pool of *E. coli*.

Other constituents of the flexible gene pool without homology to virulence genes or association with transferable genetic elements can be determined by the

comparison of whole genome contents of different pathogenic and nonpathogenic *E. coli* strains. Comparison of the genetic diversity of different pathogenic and nonpathogenic *E. coli* isolates not only describes groups of genes which are specific for individual pathotypes and may be essential for their virulence but also identifies genes which are present only in nonpathogenic *E. coli* K-12 strains. These genes may act as 'virulence repressors' and the analysis of their function may be crucial for the understanding of *E. coli* virulence. In order to assess the fraction of genes which represent the flexible or the common core gene pool among different *E. coli* pathotypes as well as genes which are specific for nonpathogenic *E. coli* strains, the genomes of seven different pathogenic *E. coli* isolates have been compared by DNA-DNA hybridisation with that of the nonpathogenic *E. coli* K-12 strain MG1655 using DNA macroarrays (DOBRINDT et al. 2001b). Between 5% and 10% of the 4290 translatable ORFs present in the *E. coli* K-12 strain MG1655 were shown to be absent in the different pathogenic *E. coli* isolates (see Table 2 and Fig. 2). The majority of these ORFs (about 50%–60%) can be functionally grouped as hypothetical and unclassified. ORFs which are categorized as components of mobile genetic elements such as bacteriophages, transposons or plasmids also show a very high variability with respect to their presence in different *E. coli* strains. In addition to these predominant groups of ORFs, several other determinants have to be considered as a part of the flexible gene pool of *E. coli* due to their variable distribution among different *E. coli* isolates. Among the different pathogenic isolates, a considerable variability has also been demonstrated with respect to the presence of genes required for LPS biosynthesis. This can be explained by different O serotypes of these strains. The *fec* operon encoding a ferric citrate uptake system and the *fim* gene cluster which codes for type 1 fimbriae, as well as the *hsd*-encoded type I restriction modification system and the *relEB*-encoded cytotoxin-antitoxin system were not detectable in some pathogenic *E. coli* isolates investigated in this study. This was also the case for *hip* genes conferring persistence to inhibition of peptidoglycan or DNA synthesis, the *mcr* operon involved in methylation restriction and several *csp* genes involved in cold shock adaptation. It has been reported for some of these determinants and also for several other *E. coli* genes that they are not generally present in *E. coli* (BARCUS et al. 1995; CHARBIT

**Table 2.** Number of translatable ORFs which are absent in different pathogenic *E. coli* strains but present in the nonpathogenic *E. coli* K-12 strain MG1655 according to DNA-DNA hybridisation experiments using an *E. coli* MG1655-specific DNA macroarray. (Data from DOBRINDT et al. 2001b)

Functional category	536 (UPEC)	IHE3034 (MENEC)	C9221a (ETEC)	E2348/69 (EPEC)	EDL1284 (EIEC)	4797/97 (EHEC)	DPA065 (EAEC)	All strains
Hypothetical, unclassified, unknown	206	114	196	119	184	125	145	40
Phage, transposon, plasmid	40	9	14	10	11	9	9	3
Others	137	79	199	80	151	78	118	11
Total	383	202	409	209	346	212	272	54



**Fig. 2.** Distribution of chromosomal alterations among different pathogenic *E. coli* isolates. The individual chromosomes are displayed in equal length in the following order, from outwards to inwards: 536 (UPEC), IHE3034 (MENEC), 4797/97 (EHEC), E2348/69 (EPEC), C9221a (ETEC), EDL1284 (EIEC), and DPA065 (EAEC). Missing ORFs are marked by vertical lines in the individual chromosomes. The position of the deleted ORFs refers to the *E. coli* MG1655 chromosome (outer circle). The positions of tRNA genes frequently used as chromosomal insertion sites of horizontally acquired DNA elements are marked within the map of the *E. coli* strain MG1655. (Data from DOBRINDT et al. 2001b)

and AUTRET 1998). They do not belong to a basic gene set common to all *E. coli* isolates, but their expression results in specific traits which are advantageous under special conditions.

According to the DNA-DNA hybridisation, about 1.3% of the translatable ORFs of the nonpathogenic *E. coli* strain MG1655 were shown to be absent in all pathogenic isolates tested (DOBRINDT et al. 2001a,b). This interesting group of *E. coli* K-12-specific ORFs may contribute to the virulence of pathogenic *E. coli* strains, as their encoded gene products may inhibit or repress cellular functions which are required for virulence in different *E. coli* pathotypes.

### 3.4 Pathogenicity Islands of Extraintestinal *E. coli*

Extraintestinal *E. coli* have the capacity to cause urinary tract infections, sepsis and newborn meningitis. As shown a number of years ago, the virulence factors of

extraintestinal *E. coli* are encoded by pathogenicity islands. In uropathogenic *E. coli*, pathogenicity islands have been described for various isolates including the strains 536 (O6:K15), J96 (O4:K6) and the strain CFT073 (BLUM et al. 1994, 1995; SWENSON et al. 1996; KAO et al. 1997; see also Chap. 2 of this issue). All these pathogenicity islands which harbour the genes coding for several virulence factors, e.g.  $\alpha$ -hemolysin, cytotoxic necrotizing factor 1, P-fimbriae, S-fimbriae or iron uptake systems, are associated with tRNA-specific loci (GROISMAN and OCHMAN 1996; HACKER et al. 1997). Interestingly, the same tRNA loci (e.g. *selC*) are also used by intestinal *E. coli* as well as by other enterobacteria (*S. enterica*, *Shigella flexneri*) as insertion sites for the incorporation of pathogenicity islands (BLANC-POTARD and GROISMAN 1997; MOSS et al. 1999). Similar pathogenicity islands have been detected in septicemic and uropathogenic *E. coli* (KAPER and HACKER 1999). In addition, a PAI which encodes afimbrial adherence factors has been described recently (LALIOUI and LE BOUGUENEC 2001). Meningitis-causing *E. coli* (MENEC) are characterized by the expression of the capsular antigen K1. The respective loci seem to be associated with the tRNA gene *pheV*, which is also used as a target for the incorporation of haemolysin and P-fimbriae specific pathogenicity islands (SWENSON et al. 1996; CIESLEWICZ and VIMR 1997). There are indications that the aerobactin gene cluster is also part of a pathogenicity island in extraintestinal *E. coli*. An aerobactin-specific pathogenicity island was first described in *S. flexneri* (MOSS et al. 1999; VOKES et al. 1999). The presence and distribution of pathogenicity islands of extraintestinal *E. coli* is a good example for gene transfer among different enterobacteria, because parts of these islands were also found in other enterobacteria, such as *Shigella* spp. or *S. enterica* (HACKER and KAPER 2000).

### 3.5 Mobile Genetic Elements of Intestinal *E. coli*

Several mobile genetic elements have been described for the different intestinal *E. coli* pathotypes to carry important genetic information which is required for their pathogenic features (see Table 3; NATARO and KAPER 1998; DOBRINDT and HACKER 1999). The genes coding for different adhesins, resistance genes and the heat-labile and heat-stable enterotoxins of ETEC strains are localized on plasmids (MAZAITS et al. 1981; BERTIN 1998; MAINIL et al. 1998; GÓMEZ-DUARTE et al. 1999). F17 fimbrial adhesin determinants have been discovered together with genes coding for the cytotoxic necrotizing factor 2 on virulence plasmids of necrotoxigenic *E. coli* strains (MAINIL et al. 2000). Many EPEC strains contain large EAF plasmids which carry the *bfp* locus coding for bundle-forming pilus and the *per* genes whose gene products are important for the regulation of LEE-encoded gene expression (DONNENBERG et al. 1992; MELLIES et al. 1999). The *astA* gene which encodes the EAST1 toxin can also be frequently found on EAF plasmids (NATARO and KAPER 1998). The presence of a so-called pO157 plasmid is characteristic for the majority of EHEC strains. This plasmid carries the *ehx* genes coding for an enterohemolysin and a catalase-peroxidase (*katP*), as well as for a type II secretion system (*etp*), a serine protease (*espP*) and a large toxin with homology to the large

**Table 3.** Properties encoded by genetic elements of the flexible gene pool of intestinal *E. coli* strains

Mobile genetic element	Encoded feature
Plasmid	Adhesins Toxins Siderophore systems Resistance to antibiotics, heavy metals Colicins Type II secretion system Type III secretion system and secreted effectors Enzymes (katalase, protease)
Bacteriophage	Toxins
IS element, transposon, integron	Toxins Antibiotic resistance
Pathogenicity island	Toxins Type III secretion system and secreted effectors Adhesins Siderophores Enzymes (proteases)

clostridial toxin (LCT) (BRUNNER et al. 1996; SCHMIDT et al. 1997; BURLAND et al. 1998). Several other plasmids ranging in size from 2 to 87 kb have been described in O157:H7 strains, too. Their importance for the virulence of these strains is unclear (WILLSHAW et al. 1992). Most of the EAEC strains carry large plasmids that share a high degree of homology (CZECZULIN et al. 1999). The AAF1 (aggregative adherence fimbriae 1)-encoding determinant has been detected on a 60-MDa plasmid which may also harbour the EAST1 toxin gene. (SAVARINO et al. 1994; NATARO and KAPER 1998). EIEC strains are closely related to *Shigella* with respect to their pathogenicity. Therefore, they share a large pInv plasmid (140-MDa) (SASAKAWA et al. 1992). This plasmid carries the invasion-related genes which encode a type III secretion apparatus (*mxi*, *spa*), as well as secreted proteins (*ipa*) involved in the invasion phenotype and a toxin designated *Shigella* enterotoxin 2 (NATARO et al. 1995). Other genes providing intestinal *E. coli* strains with advantageous properties, e.g. resistance to antibiotics, expression of colicins and siderophores, are plasmid encoded as well. Colicin-encoding plasmids can also harbour siderophore- or other virulence-associated determinants (WATERS and CROSA 1991; MARTINEZ et al. 1994; GOMEZ-LUS 1998; OTTO et al. 1998).

The different types of Shiga toxin (Stx), the major virulence factor of enterohaemorrhagic *E. coli* (EHEC) strains, are usually encoded on temperate bacteriophages. The Shiga toxin types 1- and 2-encoding genes are part of the bacteriophages H-19B and 933W, respectively (SCOTLAND et al. 1983; PLUNKETT III et al. 1999). In addition, the nucleotide sequences of the complete genomes of two O157:H7 EHEC strains show that their genomes also contain many defective prophages as well as different types of IS elements (see also Sect. 3.1 of this article; HAYASHI et al. 2001; Perna et al. 2001). Different members of the latter class of mobile genetic elements are widespread also among *E. coli* and can be associated with virulence-associated genes (MCVEIGH et al. 2000). Integrons carrying

antibiotic resistance cassettes have been described for intestinal *E. coli* strains as well (SUNDE and SORUM 1999).

Several pathogenicity islands have been described for the different intestinal *E. coli* pathotypes. The locus of enterocyte effacement (LEE) is a PAI widespread among attaching and effacing *E. coli* strains causing diarrhoea in human beings and animals. The comparison of the complete nucleotide sequences of the LEE-PAI of EPEC, EHEC and rabbit diarrhoeagenic *E. coli* (RDEC) revealed that these genetic elements share a common core of 40 ORFs. However, they differ with respect to their chromosomal insertion site and their flanking sequence context (ZHU et al. 2001). Another pathogenicity island (EspC-PAI) has been described for EPEC strains (MELLIES et al. 2001). The determination of the complete genome sequences of two O157:H7 EHEC strains demonstrated that in addition to the LEE-PAI several other large genetic elements are present in these genomes which share some features of a pathogenicity island such as the association with a tRNA gene as well as the presence of bacteriophage integrase and virulence-associated genes (HAYASHI et al. 2001; PERNA et al. 2001). The high-pathogenicity island (HPI) coding for a siderophore system has also been detected in several intestinal *E. coli* strains (EPEC, EHEC EAEC, EIEC, ETEC) (SCHUBERT et al. 1998, 1999; KARCH et al. 1999). The Tia-PAI which is required for adhesion and invasion of intestinal epithelial cells has been described for ETEC strains (FLECKENSTEIN et al. 2000). DNA elements with homology to two PAIs which were initially described for *S. flexneri* can be found in intestinal *E. coli* pathotypes. A homologue of the she-PAI encoding a protease and an enterotoxin is present in EAEC strains (HENDERSON et al. 1999). The aerobactin genes which are located on the SHI-2 PAI in *S. flexneri* are also present in EIEC strains. Although the SHI-2 PAI is not present in EIEC, the occurrence of the aerobactin determinant in a different chromosomal location in EIEC in association with IS elements may be indicative of a different type of PAI (VOKES et al. 1999). PAIs of intestinal *E. coli* strains are reviewed in detail in Chap. 3 of this volume, by Torres and Kaper.

## 4 Gene Pools in Other Enterobacteria

The minimal set of genes which is required and sufficient for the life of an independently replicating cell should be shared by the vast majority of bacteria and forms the basic composition of the core gene pool. The minimal gene set was determined from the available sequence data (MUSHEGIAN and KOONIN 1996), and, similarly, it has been estimated that between 265 and 350 sequences constitute the universal family of proteins shared by all bacterial species that had been completely sequenced at that time (KYRPIDES et al. 1999; HUTCHISON et al. 1999). However, the minimal set of genes is difficult to determine as it depends on the individual growth conditions, and there is no clear definition of the minimal requirements and processes of life. The investigation of the smallest bacterial genomes should help to

define certain aspects of the minimal gene requirements of bacterial cells. The genome of the aphid bacterial endosymbiont *Buchnera* sp. APS, whose closest free-living relatives belong to the Enterobacteriaceae, has been completely sequenced (SHIGENOBU et al. 2000). The genome of *Buchnera* sp. APS is considered to be a subset of the *E. coli* genome (SHIGENOBU et al. 2000). At about 640kb it is extremely small in comparison to free-living enterobacteria (4–5Mb) and reflects the obligate symbiotic lifestyle of this intracellular organism. The lack of considerable genome size variation among different *Buchnera* isolates (with genome sizes between 630 and 643kb) suggests that the minimal genome size, at least with respect to the specific conditions of this symbiosis, has been reached and that the isolated niche and the strictly vertical transmission of *Buchnera* prevent the acquisition of foreign DNA (WERNEGREN et al. 2000). The genome size reduction reflects the stable and nutrient-rich habitat of *Buchnera* in their obligate symbiotic relationship with aphids. The *Buchnera* genome lacks genes involved in the biosynthesis of cell-surface components, such as lipopolysaccharides and phospholipids, genes involved in fermentation, chemotaxis/motility, and osmotic adaptation, as well as those for the entry and exit of eukaryotic cells. *Buchnera* contains less genetic information required for the synthesis of amino acids than free-living enterobacteria, as it lacks the complete biosynthetic pathways for all amino acids. Only the pathways for the biosynthesis of essential amino acids are present in the genome. These amino acids are secreted and provided for the host. *Buchnera* does not produce nonessential amino acids. However, the proportion of genes involved in the production of amino acids is markedly higher than in intracellular parasites with reduced genome sizes such as *Rickettsia* and *Chlamydia*. With the exception of two small plasmids, there are no accessory genetic elements such as bacteriophages, IS sequence elements or other repetitive sequences present in the genome. The fraction of pseudogenes (8), unique genes (4) and regulatory genes is much smaller in *Buchnera* sp. SPA than in other bacteria (ZOMORODIPOUR and ANDERSSON 1999; SHIGENOBU et al. 2000). Another example of the contribution of genome content reduction to the evolution of bacterial species is the emergence of *Y. pestis* from *Y. pseudotuberculosis*. Several genes which are present in *Y. pseudotuberculosis* have been shown to be absent or nonfunctional in *Y. pestis* (ACHTMAN et al. 1999; E. Carniel, unpublished), underlining the importance of gene loss for the evolution of bacterial species.

In addition to the minimal common gene pool present in enterobacteria, similar common gene sets which are frequently organized as laterally transferable blocks on plasmids or pathogenicity islands can be found not only among isolates of clonal groups but also among different bacterial pathogens which cause similar types of infection. One example of a commonly expressed molecular virulence mechanism and an at least functional common gene pool among different enterobacteria are the determinants coding for type III secretion systems, which export toxins into host cells. They are present in different pathotypes of *E. coli*, as well as in *S. flexneri* and *S. enterica* (BARINAGA 1996). In the case of the LEE-PAI of attaching and effacing *E. coli* strains the type III secretion apparatus is encoded by the *sep* and *esc* genes (ELLIOTT et al. 1998). The *inv/spa* gene cluster of *S. enterica* is

located on a pathogenicity island and is required for the bacterial entry into mammalian intestinal cells. These genes correspond to the *mxi/spa* genes of the *S. flexneri* virulence plasmid, which can also be found in enteroinvasive *E. coli*. Its genomic location and base composition indicate that this gene cluster must have been independently acquired by *Salmonella* and *Shigella* and was kept within their genomes due to the similar lifestyles of these pathogens (OCHMAN and GROISMAN 1995).

Bacteria of the genus *Photobacterium* (Enterobacteriaceae) are closely related to *E. coli*, yet they have very different life strategies. *Photobacterium luminescens* bacteria live in symbiosis with soil nematodes. They undergo a complex life cycle that involves a symbiotic stage, during which the bacteria are carried in the guts of nematodes, and a pathogenic stage in which susceptible insects are killed by the combined action of nematode and bacteria. Partial sequencing of the *P. luminescens* genome has revealed many virulence factors that are similar to those of other enterobacteria (FRENCH-CONSTANT et al. 2000). These include insecticidal toxins, Rtx-like toxins, proteases, lipases and various antibiotics. Moreover, type III secretion mechanisms have been identified. The fact that virulence factors such as the Yops of *Yersinia* have been identified in *P. luminescens* suggests that horizontal gene transfer may also occur between vertebrate and invertebrate pathogens and even symbionts. Accordingly, many mobile genetic elements, such as insertion elements, transposons, and several large plasmids, have been identified on the genome of *P. luminescens*. As more sequence information becomes available, the gap between seemingly different life strategies of bacteria that live in intimate associations with animal hosts will likely become smaller.

Other widespread genetic elements which belong to the flexible gene pool among enterobacteria are accessory elements such as the insertion sequence element IS100, which can frequently be found within the genomes of enterobacteria. The so-called HPI which encodes a siderophore system was initially discovered in *Yersinia* spp. The core element of the HPI has also been detected in a broad range of other enterobacteria, such as *E. coli*, *Citrobacter* sp. and *Klebsiella pneumoniae*, representing a “broad host-range genomic island” which contributes to the adaptation to different niches. The presence of this island confers a growth advantage due to an increased capability of iron uptake, which is advantageous in the environment as well as in different hosts (SCHUBERT et al. 1998; KARCH et al. 1999; BACH et al. 2000).

## 5 Conclusion

With the advent of sequencing, a plethora of DNA sequence information has become available. It is now possible to compare entire microbial genomes and to search for unifying themes as well as compare their differences. First and foremost, a correlation between metabolic versatility and genome size has become evident.

The degenerated lifestyles of obligate intracellular pathogens and symbionts goes hand in hand with a phenomena that has been termed “evolution by reduction”. The more integrated lifestyles are correlated with significantly reduced genome sizes by as much as one order of magnitude. From these studies it is now possible to identify the minimal set of genes that is required for life. Other general features resulting from genome degeneration that have been determined in obligate intracellular symbionts of insects are (a) an extremely low G + C content, (b) accelerated sequence evolution, (c) shortened proteins, and (d) elimination of intergenic spaces (CLARK et al. 2001).

A second surprising finding that has come out of comparative genome analysis is that microbial genomes can host variable and frequently significant amounts of foreign DNA. Values range from virtually none in the genomes of permanently intracellular parasites such as *Mycoplasma* and the aphid symbiont *Buchnera* to as much as 18% in the enteric bacteria *E. coli* and *S. enterica*. The mechanisms which allow the horizontal transfer of DNA sequence can permanently alter bacterial genotypes. For example, the acquisition of pathogenicity islands many millions of years ago has affected the evolution of today's enterobacteria. Because this process has been instrumental in the creation in new species, it has also been termed “evolution in quantum leaps” (GROISMAN and OCHMAN 1996). Clearly, horizontal gene transfer in its diverse manifestations has been an important and frequent event throughout history; its magnitude and impact upon evolution are just beginning to be appreciated.

Thirdly, many new insights have been gained using the comparative approach. The enterobacteria have proven particularly useful in this regard because their very similar core genomes make it possible to pinpoint relevant differences. Accordingly, it was shown that single DNA segments can convert a commensal micro-organism into a pathogen. In contrast, virulence may also be caused by the absence of genes in the pathogenic variant, as may be anticipated for pathogenic *E. coli*. This would be conceivable if pathogens were lacking a repressor element that allowed virulence to proceed. Virulence by way of missing repressor elements will be revealed only via the whole genome approach, and it is expected that many more discoveries will be made.

According to Darwin, the evolution of organisms is driven by an overall need to increase fitness. Every living organism is the product of its genetic and phenotypic makeup, which is a reflection of the selective pressures of the environment. The evolutionary success of a genome should be governed by the need to increase its fitness with respect to the functions it encodes and also with respect to the structure itself. Structural parameters such as variations in size, G + C content, and amount of coding vs. noncoding sequence reflect evolutionary adaptations. With the tools that have become available it is now possible to apply Darwin's principles to a new dimension, namely the evolution of entire genomes.

*Acknowledgements.* Our own work related to this topic was supported by the *Deutsche Forschungsgemeinschaft* (SFB479) and the *Fonds der Chemischen Industrie*.

## References

- Aarestrup FM, Jorsal SE, Ahrens P, Jensen NE, Meyling A (1997) Molecular characterization of *Escherichia coli* strains isolated from pigs with edema disease. *J Clin Microbiol* 35:20–24
- Achtman M, Pluschke G (1986) Clonal analysis of descent and virulence among selected *Escherichia coli*. *Annu Rev Microbiol* 40:185–210
- Achtman M, Zurth K, Morelli G, Torrea G, Guiyole A, Carniel E (1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* 96:14043–14048
- Bach S, de Almeida A, Carniel E (2000) The *Yersinia* high-pathogenicity island is present in different members of the family *Enterobacteriaceae*. *FEMS Microbiol Lett* 183:289–294
- Barcus VA, Titheradge AJ, Murray NE (1995) The diversity of alleles at the *hsd* locus in natural populations of *Escherichia coli*. *Genetics* 140:1187–1197
- Barinaga M (1996) A shared strategy for virulence. *Science* 272:1261–1263
- Bergthorsson U, Ochman H (1998) Distribution of chromosome length variation in natural isolates of *Escherichia coli*. *Mol Biol Evol* 15:6–16
- Bertin A (1998) Phenotypic expression of K88 adhesion alone or simultaneously with K99 and/or F41 adhesins in the bovine enterotoxigenic *Escherichia coli* strain B41. *Vet Microbiol* 59:283–294
- Bingen E, Picard B, Brahim N, Matthy S, Desjardins P, Elion J, Denamur E (1998) Phylogenetic analysis of *Escherichia coli* strains causing neonatal meningitis suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strains. *J Infect Dis* 177:642–650
- Blanc-Potard AB, Groisman EA (1997) The *Salmonella* *selC* locus contains a pathogenicity island mediating intramacrophage survival. *EMBO J* 16:5376–5385
- Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Villes J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–1462
- Blum G, Ott M, Lischewski A, Ritter A, Imrich H, Tschaep H, Hacker J (1994) Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect Immun* 62:606–614
- Blum G, Falbo V, Caprioli A, Hacker J (1995) Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and alpha-hemolysin form the pathogenicity island II of the uropathogenic *Escherichia coli* strain J96. *FEMS Microbiol Lett* 126:189–195
- Boyd EF, Hartl DL (1998) Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *J Bacteriol* 180:1159–1165
- Boyd EF, Wang FS, Whittam TS, Selander RK (1996) Molecular genetic relationships of the *Salmonellae*. *Appl Environ Microbiol* 62:804–808
- Boudeau J, Barnich N, Darfeuille-Michaud A (2001) Type 1 pilus-mediated adherence of *Escherichia coli* strain LF82 isolated from Crohn's disease is involved in bacterial invasion of intestinal epithelial cells. *Mol Microbiol* 39:1272–1284
- Bruder W, Schmidt H, Karch H (1996) KatP, a novel catalase-peroxidase encoded by the large plasmid of enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology* 142:3305–3315
- Burland V, Shao Y, Perna NT, Plunkett G, Sofia HJ, Blattner FR (1998) The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Res* 26:4196–4204
- Charbit A, Autret N (1998) Horizontal transfer of chromosomal DNA between the marine bacterium *Vibrio furnissii* and *Escherichia coli* revealed by sequence analysis. *Microb Comp Genomics* 3:119–132
- Cieslewicz M, Vimr E (1997) Reduced polysialic acid capsule expression in *Escherichia coli* K1 mutants with chromosomal defects in *kpsF*. *Mol Microbiol* 26:237–249
- Clark MA, Baumann L, Thao ML, Moran NA, Baumann P (2001) Degenerative minimalism in the genome of a psyllid endosymbiont. *J Bacteriol* 183:1853–1861
- Czeculin JR, Whittam TS, Henderson IR, Navarro-Garcia F, Nataro JP (1999) Phylogenetic analysis of enteroaggregative and diffusely adherent *Escherichia coli*. *Infect Immun* 67:2692–2699
- Dho-Moulin M, Fairbrother JM (1999) Avian pathogenic *Escherichia coli* (APEC). *Vet Res* 30:299–316
- Dobrindt U, Hacker J (1999) Plasmids, phages and pathogenicity islands: lessons on the evolution of bacterial toxins. In: Alouf JE, Freer J (eds) *The comprehensive sourcebook of bacterial protein toxins*, 2nd edn. Academic, London, pp 3–23

- Dobrindt U, Blum-Oehler G, Hartsch T, Gottschalk G, Ron EZ, Fünfstück R, Hacker J (2001a) The S-fimbriae encoding determinant *sfaI* is located on the pathogenicity island III<sub>536</sub> of the uropathogenic *Escherichia coli* strain 536. *Infect Immun* 69:4248–4256
- Dobrindt U, Agerer F, Hacker J (2001b) Whole genome approach reveals genome optimization in pathogenic *Escherichia coli*. *Infect Immun* (submitted)
- Donnenberg MS, Girón JA, Nataro JP, Kaper JB (1992) A plasmid-encoded type IV fimbrial gene of enteropathogenic *Escherichia coli* associated with localized adherence. *Mol Microbiol* 6:3427–3437
- Elliott SJ, Wainwright LA, McDaniel TK, Jarvis KG, Deng YK, Lai LC, McNamara BP, Donnenberg MS, Kaper JB (1998) The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol Microbiol* 28:1–4
- Ffrench-Constant RH, Waterfield N, Burland V, Perna NT, Daborn PJ, Bowen D, Blattner FR (2000) A genomic sample sequence of the entomopathogenic bacterium *Photorhabdus luminescens* W14: potential implications for virulence. *Appl Environ Microbiol* 66:3310–3329
- Fleckenstein JM, Lindler LE, Elsinghorst EA, Dale JB (2000) Identification of a gene within a pathogenicity island of enterotoxigenic *Escherichia coli* H10407 required for maximal secretion of the heat-labile enterotoxin. *Infect Immun* 68:2766–2774
- Garrity GM (2001) Bergey's manual of systematic bacteriology, 2nd edn. Springer, Berlin Heidelberg New York
- Gómez-Duarte OG, Ruiz-Tagle A, Gomez DC, Viboud GI, Jarvis KG, Kaper JB, Girón JA (1999) Identification of *lrgA*, the structural gene of longus type IV pilus of enterotoxigenic *Escherichia coli*. *Microbiology* 145:1809–1816
- Gomez-Lus R (1998) Evolution of bacterial resistance to antibiotics during the last three decades. *Int Microbiol* 1:279–284
- Groisman EA, Ochman H (1996) Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* 87:791–794
- Hacker J, Carniel E (2001) Ecological fitness, genomic islands and bacterial pathogenicity – a Darwinian view on the evolution of microbes. *EMBO Rep* 2: (in press)
- Hacker J, Kaper JB (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* 51:641–679
- Hacker J, Blum-Oehler G, Mühlendorfer I, Tschaep H (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* 23:1089–1097
- Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han CG, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H (2001) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res* 8:11–22
- Henderson IR, Czeczulin J, Eslava C, Noriega F, Nataro JP (1999) Characterization of *pic*, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*. *Infect Immun* 67:5587–5596
- Hentschel U, Steinert M, Hacker J (2000) Common molecular mechanisms of symbiosis and pathogenesis. *Trends Microbiol* 8:226–231
- Herzer PJ, Inouye S, Inouye M, Whittam TS (1990) Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J Bacteriol* 172:6175–6181
- Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC (1999) Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* 286:2165–2169
- Jann K, Jann B (1992) Capsules of *Escherichia coli*, expression and biological significance. *Can J Microbiol* 38:705–710
- Kao JS, Stucker DM, Warren JW, Mobley HL (1997) Pathogenicity island sequences of pyelonephritogenic *Escherichia coli* CFT073 are associated with virulent uropathogenic strains. *Infect Immun* 65:2812–2820
- Kaper JB, Hacker J (1999) Pathogenicity islands, plasmids and other mobile elements, ASM Press, Washington, DC
- Karch H, Schubert S, Zhang D, Zhang W, Schmidt H, Ölschläger T, Hacker J (1999) A genomic island, termed high-pathogenicity island, is present in certain non-O157 Shiga toxin-producing *Escherichia coli* clonal lineages. *Infect Immun* 67:5994–6001
- Kyrpides N, Overbeek R, Ouzounis C (1999) Universal protein families and the functional content of the last universal common ancestor. *J Mol Evol* 49:413–423
- Lalioui L, Le Bouguenec C (2001) *afa-8* Gene cluster is carried by a pathogenicity island inserted into the tRNA(Phe) of human and bovine pathogenic *Escherichia coli* isolates. *Infect Immun* 69:937–948

- Lawrence JG, Ochman H (1998) Molecular archaeology of the *Escherichia coli* genome. Proc Natl Acad Sci USA 95:9413–9417
- Mainil JG, Daube G, Jacquemin E, Pohl P, Kaeckenbeeck A (1998) Virulence plasmids of enterotoxigenic *Escherichia coli* isolates from piglets. Vet Microbiol 62:291–301
- Mainil JG, Gerardin J, Jacquemin E (2000) Identification of the F17 fimbrial subunit- and adhesin-encoding (f17A and f17G) gene variants in necrotoxigenic *Escherichia coli* from cattle, pigs and humans. Vet Microbiol 73:327–335
- Martinez JL, Herrero M, de Lorenzo V (1994) The organization of intercistronic regions of the aerobactin operon of pColV-K30 may account for the differential expression of the *iucABCD iutA* genes. J Mol Biol 238:288–293
- Maslow JN, Whittam TS, Gilks CF, Wilson RA, Mulligan ME, Adams KS, Arbeit RD (1995) Clonal relationships among bloodstream isolates of *Escherichia coli*. Infect Immun 63:2409–2417
- Maurelli AT, Fernández RE, Bloch CA, Rode CK, Fasano A (1998) “Black holes” and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. Proc Natl Acad Sci USA 95:3943–3948
- Mazaitis AJ, Maas R, Maas WK (1981) Structure of a naturally occurring plasmid with genes for enterotoxin production and drug resistance. J Bacteriol 145:97–105
- McClelland M, Wilson RK (1998) Comparison of sample sequences of the *Salmonella typhi* genome to the sequence of the complete *Escherichia coli* K-12 genome. Infect Immun 66:4305–4312
- McVeigh A, Fasano A, Scott DA, Jelacic S, Moseley SL, Robertson DC, Savarino SJ (2000) IS1414, an *Escherichia coli* insertion sequence with a heat-stable enterotoxin gene embedded in a transposase-like gene. Infect Immun 68:5710–5715
- Mellies JL, Elliott SJ, Sperandio V, Donnenberg MS, Kaper JB (1999) The Per regulon of enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). Mol Microbiol 33:296–306
- Mellies JL, Navarro-Garcia F, Okeke I, Frederickson J, Nataro JP, Kaper JB (2001) *espC* pathogenicity island of enteropathogenic *Escherichia coli* encodes an enterotoxin. Infect Immun 69:315–324
- Moss JE, Cardozo TJ, Zychlinsky A, Groisman EA (1999) The *seIC*-associated SHI-2 pathogenicity island of *Shigella flexneri*. Mol Microbiol 33:74–83
- Mühldorfer I, Hacker J (1994) Genetic aspects of *Escherichia coli* virulence. Microb Pathog 16:171–178
- Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci USA 93:10268–10273
- Nakata N, Tobe T, Fukuda I, Suzuki T, Komatsu K, Yoshikawa M, Sasakawa C (1993) The absence of a surface protease, OmpT, determines the intercellular spreading ability of *Shigella*: the relationship between the *ompT* and *kcpA* loci. Mol Microbiol 9:459–468
- Nataro JP, Kaper JB (1998) Diarrheagenic *Escherichia coli*. Clin Microbiol Rev 11:142–201
- Nataro JP, Levine MM (1994) *Escherichia coli* diseases in humans. In: Gyles CL (ed) *Escherichia coli* in domestic animals and humans. CAB International, Wallingford, pp 285–333
- Nataro JP, Seriwatana J, Fasano A, Maneval DR, Guers LD, Noriega F, Dubovsky F, Levine MM, Morris JG Jr (1995) Identification and cloning of a novel plasmid-encoded enterotoxin of enteroinvasive *Escherichia coli* and *Shigella* strains. Infect Immun 63:4721–4728
- Ochman H, Groisman EA (1995) The evolution of invasion in enteric bacteria. Can J Microbiol 41:555–561
- Ochman H, Jones IB (2000) Evolutionary dynamics of full genome content in *Escherichia coli*. EMBO J 19:6637–6643
- Ochman H, Selander RK (1984) Standard reference strains of *Escherichia coli* from natural populations. J Bacteriol 157:690–693
- Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J Mol Evol 26:74–86
- Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304
- Ørskov F, Ørskov I (1992) *Escherichia coli* serotyping and disease in man and animals. Can J Microbiol 38:699–704
- Otto BR, van Dooren SJ, Nuijens JH, Luirink J, Oudega B (1998) Characterization of a hemoglobin protease secreted by the pathogenic *Escherichia coli* strain EB1. J Exp Med 188:1091–1103
- Perna NT, Plunkett III G, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Pósfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC,

- Welch RA, Blattner FR (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529–533
- Plunkett III G, Rose DJ, Durfee TJ, Blattner FR (1999) Sequence of Shiga toxin 2 phage 933 W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J Bacteriol* 181:1767–1778
- Pruimboom-Brees IM, Morgan TW, Ackermann MR, Nystrom ED, Samuel JE, Cornick NA, Moon HW (2000) Cattle lack vascular receptors for *Escherichia coli* O157:H7 Shiga toxins. *Proc Natl Acad Sci USA* 97:10325–10329
- Pupo GM, Karaolis DK, Lan R, Reeves PR (1997) Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. *Infect Immun* 65:2685–2692
- Pupo GM, Lan R, Reeves PR (2000) Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci USA* 97:10567–10572
- Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS (2000) Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* 406:64–67
- Riley M, Sanderson KE (1990) Comparative genetics of *Escherichia coli* and *Salmonella typhimurium*. In: Drlica K, Riley M (eds) The bacterial chromosome. ASM Press, Washington DC, pp 85–95
- Rode CK, Melkerson-Watson LJ, Johnson AT, Bloch CA (1999) Type-specific contributions to chromosome size differences in *Escherichia coli*. *Infect Immun* 19:230–236
- Sasakawa C, Buysse JM, Watanabe H (1992) The large virulence plasmid of *Shigella*. *Curr Top Microbiol Immunol* 180:21–44
- Savarino SJ, Fox P, Deng Y, Nataro JP (1994) Identification and characterization of a gene cluster mediating enteroaggregative *Escherichia coli* aggregative adherence fimbria I biogenesis. *J Bacteriol* 176:4949–4957
- Schmidt H, Henkel B, Karch H (1997) A gene cluster closely related to type II secretion pathway operons of gram-negative bacteria is located on the large plasmid of enterohemorrhagic *Escherichia coli* O157 strains. *FEMS Microbiol Lett* 148:265–272
- Schubert S, Rakim A, Karch H, Carniel E, Heesemann J (1998) Prevalence of the “high-pathogenicity island” of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infect Immun* 66:480–485
- Schubert S, Rakim A, Fischer D, Sorsa J, Heesemann J (1999) Characterization of the integration site of *Yersinia* high-pathogenicity island in *Escherichia coli*. *FEMS Microbiol Lett* 179:409–414
- Scotland SM, Smith HR, Willshaw GA, Rowe B (1983) Vero cytotoxin production in strain of *Escherichia coli* is determined by genes carried on bacteriophage. *Lancet* 2:216
- Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. *APS*. *Nature* 407:81–86
- Steinert M, Hentschel U, Hacker J (2000) Symbiosis and pathogenesis: evolution of the microbe-host interaction. *Naturwissenschaften* 87:1–11
- Sunde M, Sorum H (1999) Characterization of integrons in *Escherichia coli* of the normal intestinal flora of swine. *Microb Drug Resist* 5:279–287
- Swenson DL, Bukanov NO, Berg DE, Welch RA (1996) Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. *Infect Immun* 64:3736–3743
- Vokes SA, Reeves SA, Torres AG, Payne SM (1999) The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island. *Mol Microbiol* 33:63–73
- Wang F-S, Whittam TS, Selander RK (1997) Evolutionary genetics of the isocitrate dehydrogenase gene (*icd*) in *Escherichia coli* and *Salmonella enterica*. *J Bacteriol* 179:6551–6559
- Waters VL, Crosa JH (1991) Colicin V virulence plasmids. *Microbiol Rev* 55:437–450
- Wernegreen JJ, Ochman H, Jones IB, Moran N (2000) Decoupling of genome size and sequence divergence in a symbiotic bacterium. *J Bacteriol* 182:3867–3869
- Willshaw GA, Scotland SM, Smith HR, Rowe B (1992) Properties of Vero cytotoxin-producing *Escherichia coli* of human origin of O serogroups other than O157. *J Infect Dis* 166:797–802
- Woese CR (1987) The phylogeny of bacteria. *Microbiol Rev* 51: 221
- Wong RM, Wong KK, Benson NR, McClelland M (1999) Sample sequencing of a *Salmonella typhimurium* LT2 lambda library: comparison to the *Escherichia coli* K12 genome. *FEMS Microbiol Lett* 173:411–423
- Zhu C, Agin TS, Elliott SJ, Johnson LA, Thatte TE, Kaper JB, Boedeker EC (2001) Complete nucleotide sequence and analysis of the locus of enterocyte effacement from rabbit diarrheagenic *Escherichia coli* RDEC-1. *Infect Immun* 69:2107–2115
- Zomorodipour A, Andersson SG (1999) Obligate intracellular parasites: *Rickettsia prowazekii* and *Chlamydia trachomatis*. *FEBS Lett* 452:11–15

# **Phylogenetic Relationships and Virulence Evolution in the Genus *Bordetella***

F. VON WINTZINGERODE<sup>1</sup>, G. GERLACH<sup>2</sup>, B. SCHNEIDER<sup>2</sup>, and R. GROSS<sup>2</sup>

1 Taxonomy and Characteristics of Members of the Genus <i>Bordetella</i> . . . . .	177
2 Phylogenetic Relationship of Host-Associated <i>Bordetella</i> Species with <i>B. petrii</i> and Other Environmental Organisms . . . . .	180
3 Genomic Aspects of <i>Bordetella</i> Evolution . . . . .	184
4 Evolutionary Impact of Virulence Gene Regulation on Host Adaptation . . . . .	187
5 Coordinate and Differential Control of Virulence Gene Expression . . . . .	189
6 Evolution of the Virulence Control System . . . . .	192
7 Conclusions . . . . .	194
References . . . . .	194

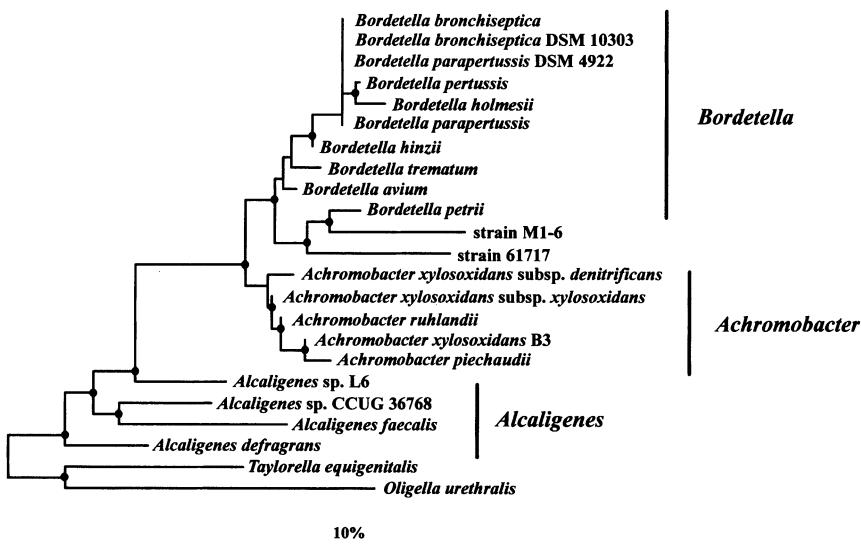
## **1 Taxonomy and Characteristics of Members of the Genus *Bordetella***

Within the beta-subclass of Proteobacteria, members of the genera *Bordetella*, *Achromobacter* and *Alcaligenes* form a group of closely related organisms (Fig. 1). To date, seven *Bordetella* species are described, all of which are gram-negative, strictly aerobic coccobacilli with a nonfermentative, asaccharolytic metabolism. They occur exclusively in close association with man or warm-blooded animals.

*B. pertussis* is the etiologic agent of whooping cough (pertussis), a highly contagious respiratory disease marked by severe, spasmodic coughing episodes (HEWLETT 1995). It is an obligate pathogen for man and there is no evidence for animal or environmental reservoirs. *B. parapertussis* causes a milder form of whooping cough in human beings and chronic, nonprogressive pneumonia in sheep (YUK et al. 1998). Unlike *B. pertussis* and *B. parapertussis*, *B. bronchiseptica* displays a broad host range and causes respiratory disease in various four-legged

<sup>1</sup> Institut für Mikrobiologie und Hygiene, Universitätsklinikum Charité, Dorotheenstrasse 96, 10117 Berlin, Germany

<sup>2</sup> Lehrstuhl für Mikrobiologie, Theodor-Boveri-Institut, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany



**Fig. 1.** Phylogenetic tree based on 1418 positions of the 16S rRNA gene of members of the AAB complex within the beta-subclass of Proteobacteria. The dendrogram was generated by the maximum likelihood approach using the FastDnaML software implemented in the ARB software package (LUDWIG and STRUNK 1998). Hypervariable positions were excluded from the analysis by use of a specific filter (ARB software). Branching points marked by solid circles are supported by a second tree generated by the maximum parsimony method (PHYLP software, implemented in the ARB package). For each species the respective type strain was analysed. Unclassified strains 61717 and M1-6 are phylogenetically affiliated to *B. petrii* despite having relatively low overall sequence similarities (96.5% and 96.9%, respectively). The 16 rRNA gene sequences of *Taylorella equigenitalis* and *Oligella urethralis* were used as an outgroup. Scale bar represents 10% estimated sequence divergence

mammalian species including dogs, rodents, horses and pigs but only rarely in human beings (GOODNOW 1980; GUEIRARD et al. 1995; WOOLFREY and MOODY 1991). Despite the differences in their host range and the severity of resulting disease, *B. pertussis*, *B. parapertussis* and *B. bronchiseptica* should not be considered true species as defined genetically. In fact, different methods including DNA-DNA hybridisation, comparative sequence analysis of 16S and 23S rDNA genes, of the beta subunits of RNA polymerase (RpoB) and of gyrase (GyrB), and of virulence genes, and multi-locus enzyme gel electrophoresis (MLEE) demonstrated that these bacteria most likely represent subspecies of a single species with different host adaptations (ARICO et al. 1987; KLOOS et al. 1981; MÜLLER and HILDEBRANDT 1993; MUSSER et al. 1986; YABUCHI et al. 1998; von Wintzingerode, unpublished results). To take account of this close relationship *B. pertussis*, *B. parapertussis* and *B. bronchiseptica* will therefore be referred to as members of the *B. bronchiseptica* cluster throughout this review.

A detailed analysis of the evolutionary relationships within the *B. bronchiseptica* cluster by MLEE and IS typing revealed that *B. parapertussis* comprises two distinct populations. Human *B. parapertussis* isolates (*B. parapertussis<sub>hu</sub>*) are highly

clonal and are more closely related to IS1001-containing *B. bronchiseptica* strains than to *B. parapertussis* isolates from sheep (*B. parapertussis<sub>ov</sub>*), which are genotypically more heterogeneous. Moreover, this analysis provided evidence that *B. pertussis*, *B. parapertussis<sub>hu</sub>*, and *B. parapertussis<sub>ov</sub>* have evolved independently from different *B. bronchiseptica* lineages (VAN DER ZEE et al. 1996, 1997).

In accordance with their close genetic relationship, members of the *B. bronchiseptica* cluster express a series of related virulence factors. Among them are adhesion or colonisation factors including the filamentous hemagglutinin (FHA), several serotypes of fimbriae (FIM) and several autotransporter proteins including pertactin (PRN). Moreover, they produce closely related toxins such as adenylate cyclase toxin (CYA), dermonecrotic toxin (DNT) and tracheal cytotoxin (TCT). As exemplified by pertussis toxin (PTX), despite the presence of the respective virulence genes there are significant differences in the pattern of expression of virulence factors in various strains of the *B. bronchiseptica* cluster (see below). Relevant characteristics of the individual virulence factors and their role during infection and disease are discussed in several recent reviews (HEWLETT 1995; PARTON 1999; WEISS 1992; COTTER and DiRITA 2000).

*B. avium* causes bordetellosis, an upper respiratory tract disease of young poultry. Bordetellosis causes severe problems in commercially raised poultry, especially turkeys (ARP and CHEVILLE 1984). A recent survey indicated that *B. avium* infection may also be common among different wild bird species (RAFFEL et al. 2000). As with members of the *B. bronchiseptica* cluster, *B. avium* exhibits a strong tropism for ciliated tracheal epithelial cells (ARP and CHEVILLE 1984). Little is known about relevant virulence factors of *B. avium*, although it has been shown to produce TCT and DNT (GENTRY-WEEKS et al. 1988). Phenotypic and genotypic characteristics demonstrated that *B. avium* is truly a distinct species related only distantly to members of the *B. bronchiseptica* cluster (KERSTERS et al. 1984; DELEY et al. 1986; MATTHEWS and PRESTON 1997).

Within the past 5 years, three new species have been included in the genus *Bordetella*, namely *B. hinzii*, *B. holmesii* and *B. trematum* (VANDAMME et al. 1995, 1996; WEYANT et al. 1995). *B. hinzii* is found mainly as a commensal of the respiratory tract of fowl. Only recently, *B. hinzii* was reported to be the causative agent of fatal septicemia in man (KATTAR et al. 2000). It was also isolated from immunocompromised patients, where it might have pathogenic potential (COOKSON et al. 1994; FUNKE et al. 1996; GADEA et al. 2000). A group of isolates previously collected in the CDC nonoxidizer group 2 (NO-2) were recently classified as *B. holmesii* on the basis of genotypic and chemotaxonomic analyses (WEYANT et al. 1995). *B. holmesii* strains have been isolated repeatedly from the blood of young adults and occasionally from sputum and nasopharyngeal specimens; more recently, they were shown to be associated with pertussis-like disease (WEYANT et al. 1995; TANG et al. 1998; YIH et al. 1999; MAZENGIA et al. 2000). Since the first isolation of *B. holmesii* strains in 1983, their frequency of isolation has increased substantially, raising the question of whether *B. holmesii* is an emerging pathogen (WEYANT et al. 1995; MAZENGIA et al. 2000). The most recently described member of the genus is *B. trematum*, which has been isolated from human ear infections and

wounds, but never from the respiratory tract. Little is known about the pathogenic significance of this organism (VANDAMME et al. 1996).

## 2 Phylogenetic Relationship of Host-Associated *Bordetella* Species with *B. pertii* and Other Environmental Organisms

The family Alcaligenaceae comprises the genera *Bordetella* and *Alcaligenes* (DELEY et al. 1986). In the past, different types of bacteria were assigned to the genus *Alcaligenes* due to an incomplete definition based solely on phenotypic traits. As a result, *Alcaligenes* spp. were found to be genotypically quite heterogeneous, exhibiting differences of more than 10% in their G + C content of DNA and different 16S rRNA-based phylogenetic affiliations (BUSSE and AULING 1992; DELEY et al. 1986). Based on comparative 16S rDNA sequence analysis and DNA G + C content data, YABUCHI et al. (1998) recently revived the genus *Achromobacter*, which accommodates those species formerly assigned to the genus *Alcaligenes* that are genotypically more closely related to *Bordetella* spp. Within the 16S rDNA-based phylogenetic tree the genera *Achromobacter* and *Bordetella* form two closely related clusters which are clearly separated from the quite heterogeneous *Alcaligenes* spp. (Fig. 1). Until now the genus *Achromobacter* has not been formally assigned to the family Alcaligenaceae; consequently, we use the term “*Alcaligenes-Achromobacter-Bordetella*” complex (AAB complex) to define this group of closely related bacteria.

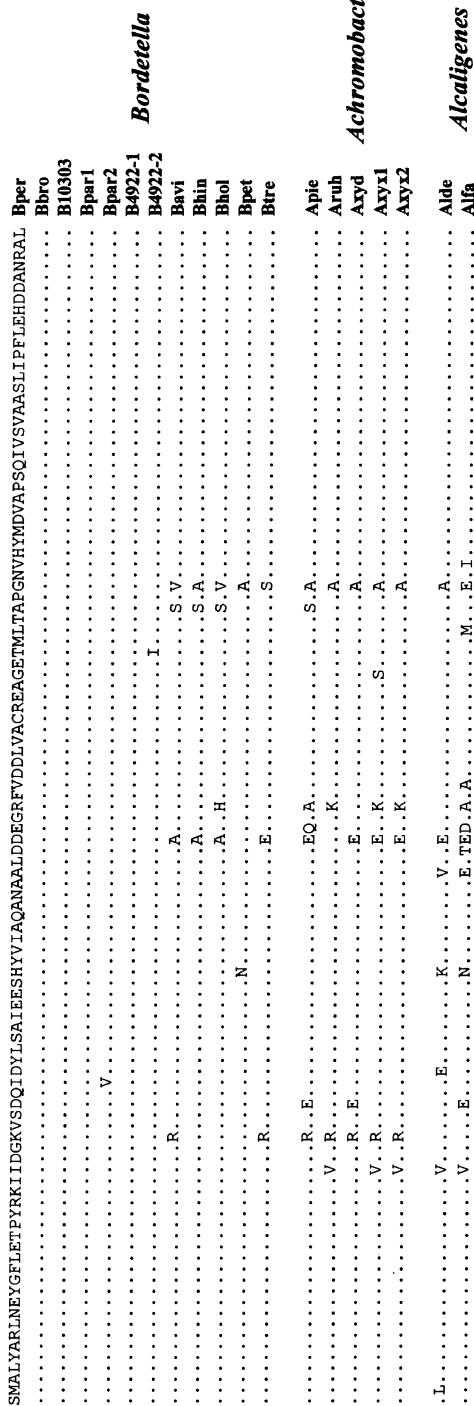
It is generally accepted that *Bordetella* spp. occur exclusively in close association with man and warm-blooded animals, whereas *Alcaligenes* and *Achromobacter* spp. are environmental bacteria with some facultatively pathogenic representatives (*Al. faecalis*, *Alcaligenes* sp. strain CCUG 36768, *Achromobacter piechaudii* and *A. xylosoxidans*) (KIREDJIAN et al. 1986; MANDELL et al. 1987; BUSSE and AULING 1992; WEISS 1992; KROONEMAN et al. 1996; GRANOWITZ and KEENHOLTZ 1998; Foss et al. 1998; KRONVALL et al. 2000; REINECKE et al. 2000). This is surprising, since *Bordetella* spp. have relatively simple nutritional requirements and share many physiological characteristics with other AAB species. With the exception of *B. pertussis*, they are easily cultivated in vitro on various media (WEISS 1992). Furthermore, Porter et al. (PORTER et al. 1991; PORTER and WARD-LAW 1993) observed that *B. bronchiseptica* is able to survive for several weeks in buffered saline and lake water without the addition of nutrients, which suggests freshwater habitats as potential reservoirs of this pathogen. However, despite several reports on *Bordetella*-like strains isolated from contaminated soil, paper machines or milk, until recently *Bordetella* species had not been isolated from the environment (SHEN et al. 1998; VÄISÄNEN et al. 1998; JAYARAO and WANG 1999). Since detailed phenotypic descriptions and 16S rDNA sequence data for these isolates are not available, the taxonomic position of these isolates remains unclear. This is best illustrated by the trichloroethylene-degrading soil isolate KP22, formerly classified as a *Bordetella* sp. (HANADA et al. 1998), which is actually an

*Achromobacter* strain exhibiting 99.7% 16S rDNA sequence similarity to the type strain of *A. piechaudii* (von Wintzingerode, unpublished results).

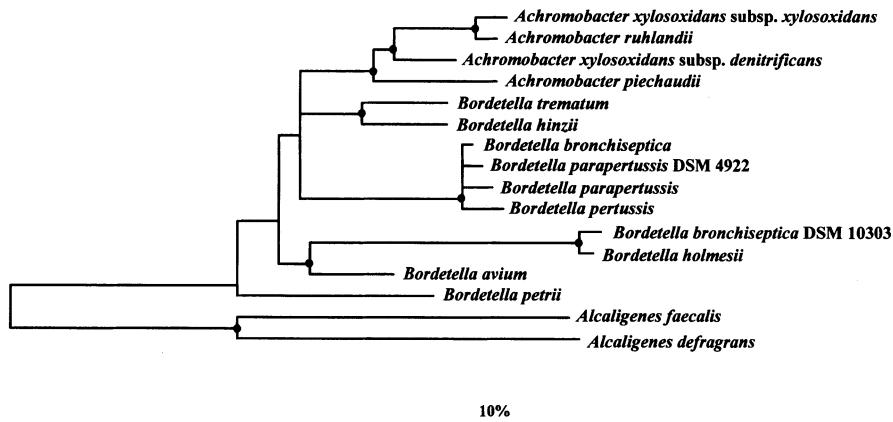
To our knowledge, the strain Se-1111R, which was recently isolated from an anaerobic, dechlorinating bioreactor, is the first true member of the genus *Bordetella* of environmental origin. Its biochemical and chemotaxonomic characteristics and its DNA G+C content affiliated strain Se-1111R to the *Achromobacter-Bordetella* group within the AAB complex. Comparative 16S rDNA sequence analysis and DNA-DNA hybridisation experiments revealed that strain Se-1111R in fact should be considered to represent a novel species within the genus *Bordetella*, for which the name *B. petrii* was proposed (VON WINTZINGERODE et al. 2001) (Fig. 1). Interestingly, a *B. petrii*-associated bacteriophage was isolated which morphologically resembles the temperate transducing bacteriophage of *B. avium* (SHELTON et al. 2000; VON WINTZINGERODE et al. 2001). The ability of *B. petrii* to grow anaerobically is noteworthy, since this feature links the strictly aerobic "classical" Bordetellae with facultatively anaerobic *Achromobacter* species (see Sect. 6).

Similar to ribosomal RNA, genes encoding the beta subunit of RNA polymerase (RpoB) and of gyrase (GyrB) are ubiquitous among bacteria. As the amino acid sequences of these proteins are highly conserved, their comparison has been used for bacterial classification (YAMAMOTO and HARAYAMA 1995; YAMAMOTO et al. 2000; MOLLET et al. 1997; KIM et al. 1999). Therefore, to investigate the genetic relationship of *B. petrii* and the other members of the AAB complex in more detail, we analysed their RpoB and GyrB sequences (von Wintzingerode, unpublished). *rpoB* gene fragments were PCR amplified from type strains of all AAB species and two additional *Bordetella* strains (*B. parapertussis* DSM 4922 and *B. bronchiseptica* DSM 10303) using primers described previously (DAHLLÖF et al. 2000). Comparative analysis of the deduced RpoB amino acid sequences supports the 16S rDNA-based classification presented previously: (a) *Bordetella* spp. are more closely related to *Achromobacter* spp. than to *Alcaligenes* spp. (Fig. 2). (b) With two exceptions (*B. trematum* and *A. xylosoxidans* subsp. *denitrificans*), members of the genera *Bordetella* and *Achromobacter* were most closely related to their respective type species (*B. pertussis* and *A. xylosoxidans*) thus confirming their taxonomic affiliations (data not shown). (c) Members of the *B. bronchiseptica* cluster were most closely related with each other, showing nearly identical or identical RpoB sequences (Fig. 2). It is noteworthy that preliminary results of comparative sequence analysis of genes encoding the RisA response regulator also show the genotypic coherence of the Bordetellae and support the affiliation of *B. petrii* to the genus *Bordetella* (Gerlach and Gross, unpublished).

For comparative GyrB sequence analysis *gyrB* gene fragments were PCR amplified from the above-mentioned strains using degenerate primers (YAMAMOTO and HARAYAMA 1995). A phylogenetic tree derived from deduced GyrB sequences is shown in Fig. 3. Several characteristics of this tree are consistent with the 16S rDNA-based phylogeny (compare Fig. 1 and Fig. 3): (a) *Achromobacter* spp. form a monophyletic group. (b) *Bordetella* species are more closely related to the *Achromobacter* cluster than to *Alcaligenes* spp. (c) GyrB sequences of the type strains



**Fig. 2.** Sequence alignment of 107 deduced amino acid positions of the beta subunit protein of RNA polymerase (RpoB) of members of the AAB complex. Differences in amino acid residues compared with the RpoB protein of *B. pertussis* are shown. For each species the respective type strain was analysed. *Bper*, *Borderella pertussis*; *Bbro*, *B. bronchiseptica*; *B10303*, *B. bronchiseptica* strain DSM 10303; *Bpar1* and *Bpar2*, *B. parapertussis* PCR clones 1 and 2; *B4922-1* and *B4922-2*, *B. parapertussis* strain DSM 4922 PCR clones 1 and 2; *Bav1*, *B. avium*; *Bhin*, *B. himazii*; *Bhol*, *B. holmesii*; *Bpet*, *B. petrii*; *Btre*, *B. trematum*; *Apie*, *Achromobacter piechaudi*; *Arah*, *A. rrahlandii*; *Axyd*, *A. xylosoxidans* subsp. *denitrificans*; *Axyx1* and *Axyx2*, *A. xylosoxidans* subsp. *xylosoxidans* PCR clones 1 and 2; *Alde*, *Alcaligenes defragnans*; *Alfa*, *A. faecalis*.



**Fig. 3.** Phylogenetic tree based on 379 deduced amino acid positions of the beta-subunit protein of gyrase (GyrB) of members of the AAB complex. The dendrogram was generated by the maximum parsimony method using the ARB-parsimony program of the ARB software package (LUDWIG and STRUNK 1998). Hypervariable positions were excluded from the analysis by use of a specific filter (ARB software). Branching points marked by solid circles are supported by a second tree generated by the maximum likelihood method (PHYLIP software, implemented in the ARB package). For each species the respective type strain was analysed. Scale bar represents 10% estimated sequence divergence between two species of a cluster. Branches connecting different clusters are not in scale. Detailed GyrB similarity values are available on request

of *B. pertussis*, *B. parapertussis*, and *B. bronchiseptica* differ in only a few amino acids positions, underlining the high overall similarity of the members of the *B. bronchiseptica* cluster. Furthermore, the GyrB sequence of the ovine *B. parapertussis* strain DSM 4922 is nearly identical to that of the *B. parapertussis* type strain, which is in line with the fact that their 16S rDNA sequences are identical (data not shown).

However, the analysis based on the GyrB sequences also revealed interesting differences as compared with the 16S rDNA-based phylogenetic tree: (a) In this analysis the genus *Bordetella* does not appear to be monophyletic. (b) As judged from GyrB sequence similarity with the type species of the genus *Bordetella*, only the type strains of *B. parapertussis* and *B. bronchiseptica* and *B. parapertussis* strain DSM 4922 are true *Bordetella* strains, whereas the GyrB sequences of the other *Bordetellae* are more similar to that of *A. xylosoxidans* subsp. *xylosoxidans*, the type species of the genus *Achromobacter*. Interestingly, *A. xylosoxidans* subsp. *xylosoxidans* is also placed within the *Bordetellae* by comparative sequence analysis of their genes encoding the outer-membrane protein OmpA (Gerlach and Gross, unpublished). (c) GyrB sequence analysis does not support the 16S rDNA-based close relationship of *B. holmesii* to the *B. bronchiseptica* cluster. (d) For *B. bronchiseptica* strain DSM 10303 comparative GyrB- and 16S rDNA-sequence analyses produced different results. Whereas strain DSM 10303 and the *B. bronchiseptica* type strain share identical 16S rDNA sequences, the GyrB sequence of strain DSM 10303 is nearly identical to that of *B. holmesii* and is only distantly related to the sequence of the *B. bronchiseptica* type strain.

Such differences in phylogenetic trees based on different genetic markers are not unusual. For instance, discrepancies in GyrB- and 16S rDNA-based phylogenetic trees were also observed among closely related *Pseudomonas* strains (YAMAMOTO and HARAYAMA 1998). In these cases the different tree structures were obtained only when hypervariable 16S rDNA positions were included in the analysis, and their appearance may therefore be explained by plesiomorphy of hypervariable residues (LUDWIG et al. 1998). However, in the present study differences between GyrB- and 16S rDNA-based trees of the AAB species were observed despite the fact that hypervariable positions were omitted from the sequence alignments. Therefore, the GyrB- and 16S rRNA-encoding genes present in the different AAB species must have been subject to different evolutionary constraints.

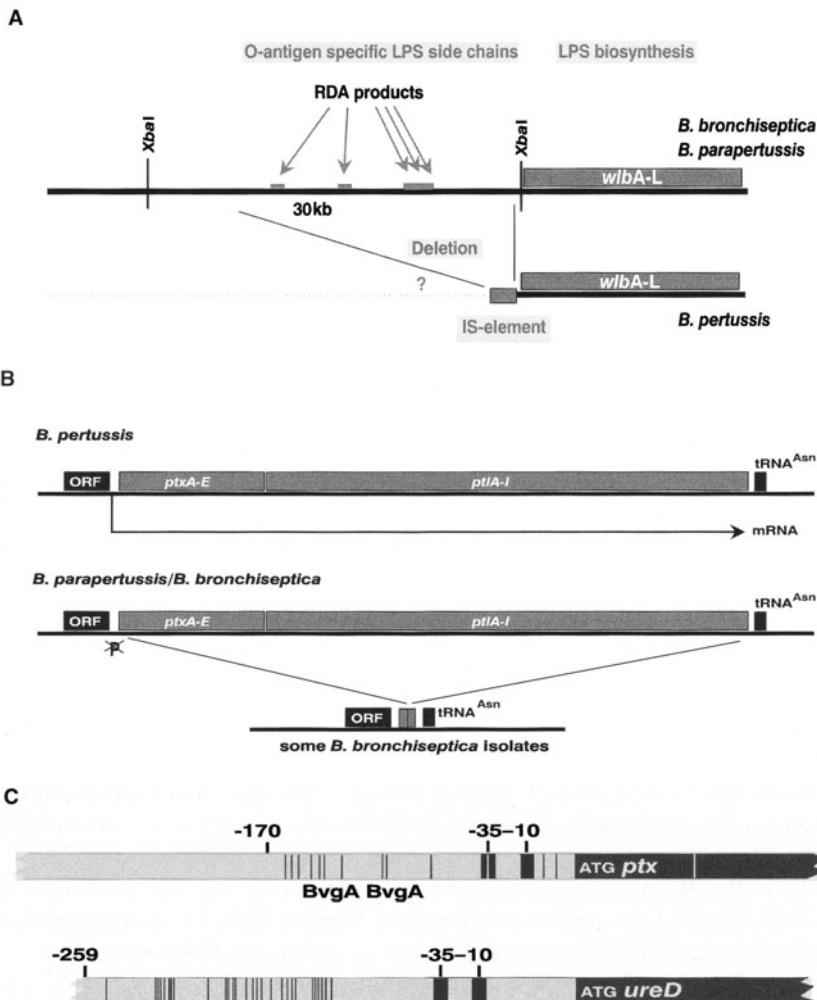
In summary, phylogenetic relationships among AAB species appear to be quite complex. Whereas the RpoB analysis generally supports the 16S-rDNA- and DNA-G + C-based definition of the genera *Bordetella* and *Achromobacter* as distinct groups, the GyrB-based analysis does not. Moreover, the genus *Alcaligenes* is not yet well defined and there are currently only two validated species (*Al. faecalis* and *Al. defragrans*). As judged from 16S rDNA analysis, strains classified as *Alcaligenes* sp. are only distantly related to the type species of the genus (*Al. faecalis*) and do not form a monophyletic group. Consequently, their current classification neglects the phylogenetic relationship of these organisms. The phylogenetic relationships of the novel members of the AAB complex such as *B. petrii* and the recently isolated strains 61717 and M1-6 (DRANCOURT et al. 2000; HUANG et al. 2000) (see Fig. 1) need to be further characterized, because they are placed between *Bordetella* spp. and *Achromobacter* spp. suggesting a bush-like topology of the evolutionary tree of the AAB complex. Such a genetic structure poses problems for the genus definition, since genus level differentiation based on phenotypic characteristics is impossible with these bacteria (VANDAMME et al. 1996; YABUCHI et al. 1998).

### 3 Genomic Aspects of *Bordetella* Evolution

Genetically, the bacteria belonging to the *B. bronchiseptica* cluster are closely related (Fig. 1). Nevertheless, the genome structures of these bacteria show a remarkable variability (STIBITZ and YANG 1997). For example, the characterisation of the genomic organisation of 14 clinical isolates of *B. pertussis* collected during an outbreak of whooping cough in Canada, as well as the comparison of several laboratory strains, demonstrated unusually frequent events of large chromosomal inversions (STIBITZ and YANG 1997, 1999). These genomes contain high copy numbers of mobile genetic elements which may be involved in the generation of such rearrangements [[http://www.sanger.uk/projects/B\\_pertussis](http://www.sanger.uk/projects/B_pertussis)]. Accordingly, the comparison of the genomic organisation of *B. pertussis* strains with a *B. parapertussis* strain did not reveal much similarity in gene order (STIBITZ and YANG 1997).

As shown by PFGE and whole-genome sequencing, the *B. pertussis* genome is several hundred kilobases smaller than that of other isolates of the *B. bronchiseptica* cluster [[http://www.sanger.uk/projects/B\\_pertussis](http://www.sanger.uk/projects/B_pertussis)]. Apparently, the long-lasting adaptation of *B. pertussis* to a single host species is correlated with a substantial reduction in its genome size. Genome reduction and loss of metabolic versatility were noted previously also in other obligate pathogens such as *Rickettsia* spp., *Mycoplasma* spp., and *Chlamydia* spp., as well as in intracellular endosymbionts such as *Buchnera aphidicola* (ANDERSSON and KURLAND 1998; SHIGENOBU et al. 2000). The lack of genetic information is also in agreement with the more fastidious growth requirements of *B. pertussis* isolates and their increased sensitivity against detrimental environmental conditions as compared with *B. bronchiseptica* strains (COTTER and MILLER 1997; PORTER and WARDLAW 1993; SCHNEIDER et al. 2000). Moreover, the recent attempt to identify strain-specific DNA fragments by representational difference analysis (RDA) resulted in the isolation of several gene loci encoding putative metabolic functions present in *B. bronchiseptica* and *B. parapertussis* isolates which are absent in *B. pertussis* strains (MIDDENDORF and GROSS 1999). Most notably, the RDA approach identified a genetic locus present in the genomes of *B. bronchiseptica* and *B. parapertussis* isolates which encodes functions required for the biosynthesis of the O-antigen-specific lipopolysaccharide (LPS) side chains that has been deleted from the *B. pertussis* genome (Fig. 4A) (MIDDENDORF and GROSS 1999; PRESTON et al. 1999). In fact, O-antigen-specific side chains are naturally produced by *B. bronchiseptica* and *B. parapertussis*, but they are not expressed by *B. pertussis* isolates. As recently described, in synergy with the tracheal cytotoxin, the LPS has a major impact on tissue destruction in the tracheal epithelium and plays a major role in pathology (FLAK et al. 2000). Moreover, as discussed below, changes in the LPS structure appear to be correlated with host adaptation in human *B. bronchiseptica* isolates (GUEIRARD et al. 1998).

One of the RDA fragments present in *B. pertussis* but absent from the *B. bronchiseptica* and *B. parapertussis* genomes encodes a restriction/modification system and is part of a *B. pertussis* specific lysogenic bacteriophage (MIDDENDORF and GROSS 1999; [http://www.sanger.uk/projects/B\\_pertussis](http://www.sanger.uk/projects/B_pertussis)). A second *B. pertussis* specific RDA fragment marks a small genomic region of about 5kb encoding metabolic and regulatory functions which are absent from all other *Bordetella* species and may therefore represent a small metabolic islet which has been acquired horizontally by *B. pertussis* (GERLACH et al. 2000; MIDDENDORF and GROSS 1999). Although these examples indicate the presence of horizontally acquired genetic material in the genome of *B. pertussis*, currently no major impact of horizontal gene transfer is apparent for the evolution and host adaptation of bacterial lineages within the *B. bronchiseptica* cluster. However, horizontal acquisition of genetic information probably has contributed to the evolution of the *B. bronchiseptica* cluster itself, because closely related organisms infecting birds such as *B. hinzi* or *B. avium* or the environmental isolate *B. petrii* seemingly do not contain DNA sequences homologous to the most prominent virulence genes present in the mammalian pathogens (GERLACH et al. 2000). Accordingly, the gene locus encoding the pertussis toxin genes (*ptxA-E*) and its type IV secretion machinery (*ptlA-I*) has



**Fig. 4A–C.** Variable genomic regions of *Bordetella* species. **A** Representational difference analysis (RDA) identified a gene locus involved in the biosynthesis of O-antigen-specific LPS side chains in *B. bronchiseptica* and *B. parapertussis*. This locus is replaced by an insertion element in *B. pertussis*, where only the adjacent *wlb* locus has been retained. **B** The *ptx* locus has some characteristics of a pathogenicity island. *B. parapertussis* and *B. bronchiseptica* do not express PTX due to inactivating mutations in their promoter regions (cf. C). In some *B. bronchiseptica* strains most of the *ptx/ptl* locus is deleted, conserving only the flanking regions on its left and a tRNA gene on its right side. **C** Accumulation of point mutations in the promoter regions of transcriptionally silent pertussis toxin and urease pseudogenes. The position of the point mutations is indicated by vertical bars

some features of a pathogenicity island (PAI) (ANTOINE et al. 2000b; BURNS 1999). This gene cluster is flanked on one side by a tRNA-encoding gene, and there are *B. bronchiseptica* strains that lack the entire *ptx/ptl* gene cluster but retain the upstream and downstream gene sequences (Fig. 4B) (ANTOINE et al. 2000b). The comparison of the completed genome sequences of *B. pertussis* and *B. bronchiseptica*

will provide sufficient information in the future to estimate whether and to which extent the adaptation process to a single host species – in the case of *B. pertussis*, human beings – has required the acquisition of relevant novel genetic information by horizontal transfer, or whether *B. pertussis* is merely a deletion variant of a *B. bronchiseptica* ancestor ([http://www.sanger.uk/projects/B\\_pertussis](http://www.sanger.uk/projects/B_pertussis)).

## 4 Evolutionary Impact of Virulence Gene Regulation on Host Adaptation

Little is known about virulence-related factors in *Bordetella* species outside of the *B. bronchiseptica* cluster. However, for the investigation of molecular mechanisms of host adaptation the *B. bronchiseptica* cluster itself represents an extremely interesting group of organisms. It provides an extraordinary collection of bacterial lineages which have already adapted or apparently are in the state of adaptation to specific mammalian host organisms or human beings (VAN DER ZEE et al. 1997). In fact, as shown by a combination of IS typing and multilocus enzyme electrophoreses (MLEE) there are bacterial lineages with a broad or a quite narrow host range, and there are lineages which are characterised by the adaptation to a single host species, such as the strictly human pathogens *B. pertussis* and *B. parapertussis* (VAN DER ZEE et al. 1997). In addition, the molecular analysis of host adaptation in the *B. bronchiseptica* cluster is particularly interesting because the genetic repertoire of these organisms is very similar. Contrary to the situation observed in several pathogenic Enterobacteriaceae, host adaptation seemingly did not involve horizontal acquisition of large amounts of genetic material (GERLACH et al. 2000).

What else is the molecular basis of host adaptation in the *B. bronchiseptica* cluster? Most importantly, major differences between bacterial lineages adapted to different host species were noted on the level of gene expression. In fact, several virulence genes and housekeeping factors were described which are expressed in one strain adapted to a particular host, but, despite the presence of the respective coding sequences, are transcriptionally silent or regulated differently in other strains. In Table 1 several examples of differentially expressed genes in *B. pertussis* and *B. bronchiseptica* strains are shown. Moreover, striking examples of differential gene expression during infection were recently reported for human *B. bronchiseptica* strains isolated from a patient with recurrent bronchopneumonia. With time these strains have lost their ability to express their major virulence factor CYA and the capacity to express the O-antigen-specific side chains characteristic for the LPS of *B. bronchiseptica*, indicating that successful adaptation to human beings as the host and the establishment of a chronic state required significant adaptations in the pattern of virulence gene expression by these strains (GUEIRARD et al. 1995, 1998; LEBLAY et al. 1997). Differential control of gene expression therefore appears to be a key mechanism involved in adaptational processes within the *B. bronchiseptica* cluster.

**Table 1.** Examples of differentially expressed factors in *B. pertussis* and *B. bronchiseptica*

Factors	<i>Bordetella pertussis</i>	<i>Bordetella bronchiseptica</i>
PTX		
Genes	+	+
Expression	+	-
Type IV secretion		
Genes	+	+
Expression	+	+
Tcf		
Gene	+	+
Expression	+	-
BrkA		
Gene	+	+
Expression	+	Some strains
Type III secretion		
Genes	+	+
Expression	-	+
Alcaligin biosynthesis		
Genes	+	+
Bvg-regulated expression	-	Some isolates
Flagellae		
Genes	+	+
Expression	-	+
Urease		
Genes	+	+
Expression	-	+
Vrg6		
Genes	+	+
Expression	+	-

Although little is known about the role of differential gene expression for the virulence properties of these bacteria and about the underlying molecular mechanisms, several interesting observations have been reported recently: (a) As mentioned above, the *Bordetella* genome contains a large number of mobile genetic elements generating significant genetic variability between different strains. In *B. pertussis* lack of gene expression in several cases is due either to integration of insertion elements in the respective genes/operons or to subsequent genetic rearrangements leading to scrambled operon structures and silent genes (Schneider and Gross, unpublished results). As exemplified by the LPS biosynthesis locus described above (Fig. 4A), such scrambled operons or pseudogenes may be subject to further degeneration and subsequent deletion from the chromosome. (b) There are genes which in different strains are regulated in a different manner on the transcriptional level. A striking example is the alcaligin siderophore biosynthesis operon which is under the control of the iron responsive FUR protein (BEALL and SANDEN 1995) but in certain *B. bronchiseptica* lineages is also regulated by the virulence regulatory BvgAS two-component system (GIARDINA et al. 1995). (c) In the *ptx* and *ure* operons silencing of gene expression appears to be caused by a striking

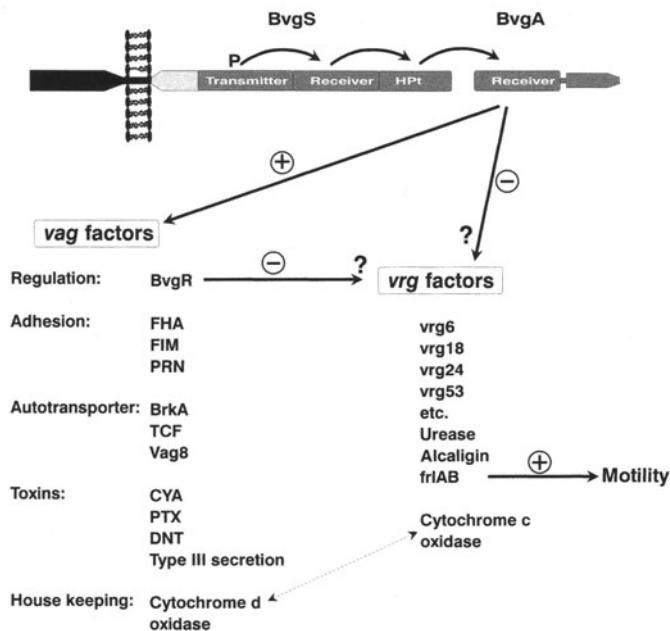
accumulation of point mutations in their promoter regions, possibly interfering with binding of transcription factors (Fig. 4C) (GROSS and RAPPOLI 1988; MACMILLAN et al. 1998). Similar cases of gene silencing were recently described for amino acid biosynthetic operons in some strains of *Buchnera*, the endosymbiont of aphids (LAI et al. 1996). Currently, however, no mechanistic explanation can be provided for the apparent selectivity of such point mutations which are confined to a small region in the upstream portions of these operons. In fact, there are only few additional point mutations present within the structural genes of the *ptx* operon in *B. bronchiseptica* and *B. parapertussis*. Accordingly, it was recently shown that, if expressed, PTX derived from *B. bronchiseptica* can be a functional toxin (HAUSMAN and BURNS 2000). In agreement with this finding, recent case descriptions of human *B. bronchiseptica* infections reported an elevated serum antibody titre against PTX indicating that, although not expressed in vitro, expression of their *ptx* operons must have occurred during infection in the human host (STEFANELLI et al. 1997).

(d) Different strains of the *B. bronchiseptica* cluster can express different fimbrial serotypes (Fim2, Fim3, FimA, FimN and FimX) (BOSCHWITZ et al. 1997; KANIA et al. 2000). The production of certain fimbriae may contribute to host adaptation because preliminary results indicated a correlation of their expression level and the host species colonised by different *B. bronchiseptica* isolates (BURNS et al. 1993). Expression of the fimbriae can be switched off reversibly by promoter-inactivating mutations consisting in nucleotide insertions or deletions within long C stretches in their transcriptional control regions (WILLEMS et al. 1990).

These examples indicate that differences in gene expression play a major role during adaptation to different host species within the members of the *B. bronchiseptica* cluster. Many virulence-related genes are regulated coordinately in their expression by the BvgAS two-component system (COTTER and DiRITA 2000). Interestingly, although there are many differences in the regulation of virulence-related functions in the different strains (Table 1), the BvgAS system itself appears to be functionally equivalent in the different lineages of the *B. bronchiseptica* cluster and was even exchanged between *B. pertussis* and *B. bronchiseptica* without obvious effects on the virulence properties of the respective mutant strain in a rat infection model (MARTINEZ DE TEJADA et al. 1996). Despite extensive surveys for other regulatory factors involved in virulence regulation, only very few additional regulators have been identified so far, including the *fur* gene encoding the global iron-responsive Fur protein and the RisAS two-component system, which apparently is involved in survival under harsh conditions and may contribute to persistent infection in the case of *B. bronchiseptica* (BEALL and SANDEN 1995; JUNGNITZ et al. 1998).

## 5 Coordinate and Differential Control of Virulence Gene Expression

The BvgAS system consists of the unorthodox BvgS histidine kinase and the transcriptional activator BvgA (Fig. 5) (PERRAUD et al. 1999). It controls several



**Fig. 5.** Schematic representation of the BvgAS two-component system and of some of its positively and negatively regulated factors

subsets of genes either in a positive (*vag*; virulence-activated genes) or in a negative manner (*vrg*; virulence-repressed genes) in response to changes in the growth conditions (ARICO et al. 1989; COTTER and DiRITA 2000; KNAPP and MEKALANOS 1988; WEISS and FALKOW 1984). Temperature is one of the signals perceived by the BvgAS system, and body temperature leads to maximal expression of the *vag* genes, whereas the *vrg* genes are expressed only at room temperature. In *B. pertussis*, among the positively controlled *vag* genes there are the major virulence factors including PTX, CYA, DNT and the adhesion and colonisation factors FHA and FIM, as well as several autotransporter proteins including PRN, TCF and Vag8 (Fig. 5) (FINN and AMBSBAUGH 1998; HEWLETT 1995; PARTON 1999; RAPPOLI 1994). Moreover, a recently performed "in silico scan" of the genome sequence of *B. pertussis* revealed some additional *bvg*-regulated genes (ANTOINE et al. 2000a). The BvgAS system also activates transcription of a putative repressor protein, BvgR, which probably controls expression of a series of *vrg* genes (MERKEL et al. 1998). Neither the genes belonging to the *vrg* regulon nor their functions and their transcriptional control system(s) are well characterized yet, and *vrg* regulation may also involve either a direct action of BvgA as a repressor and/or additional regulatory components including a newly discovered quorum-sensing system (MERKEL and KEITH 2000). Interestingly, regarding the expression of the *vrg* genes, many strain-specific differences were noted. For instance, in contrast to most other strains

of the *B. bronchiseptica* cluster, *B. pertussis* is not able to express the *bvg* negatively regulated urease and is not motile despite the presence of motility genes (AKERLEY and MILLER 1996; MACMILLAN et al. 1998), whereas only *B. pertussis* but not *B. bronchiseptica* expresses the *vrg6* gene (BEATTIE et al. 1993). In *B. bronchiseptica* isolates, several of the *vrg* genes may contribute to survival and growth under harsh environmental conditions (BANEMANN and GROSS 1997; COTTER and MILLER 1997).

Due to subtle differences in their regulation, the *vag* genes can be further subdivided (SCARLATO et al. 1993). These differences become apparent when the kinetics of transcription of these factors are investigated after shifting the bacteria from nonpermissive to permissive conditions. According to these differences in their transcription kinetics, several of the *vag* genes were classified as so-called early and late genes (SCARLATO et al. 1991). Whereas the major adhesin FHA is an early factor, the toxins PTX and CYA will be transcribed only several hours after the switch in the environmental conditions. This differential expression depends on the concentration of phosphorylated BvgA (BvgA-P), as the promoters have different affinities for the transcriptional regulator and their maximal expression either may be achieved with a low amount of BvgA-P (FHA) or may require high concentrations of the activator (PTX and CYA) (STEFFEN et al. 1996; ZU et al. 1996). Moreover, additional factors have been identified which, in an unknown fashion, may be implicated in the fine tuning of toxin transcription (DESHAZER et al. 1995; FUCHS et al. 1996). A role in virulence for these regulatory phenomena has not yet been formally demonstrated. However, it was noted that this differential pattern of virulence gene expression reflects the natural events occurring during an infection, when a pathogen residing in the environment enters a new host and first has to guarantee its adherence to host tissue; only in a second phase does it have to organise its defence against host attack, possibly requiring toxin expression (RAPPUOLI 1994; SCARLATO et al. 1993).

On the other hand, the strictly human pathogen *B. pertussis* is directly transmitted from one host to another via aerosol and is generally expected to live in a quite uniform environment. In fact, *Bordetella* mutants with a BvgS constitutive phenotype permanently expressing their virulence regulon were not distinguishable from wild-type bacteria in several animal models, at least during early phases of infection. Based on these results, it was suggested that the main task of the BvgAS system may be simply to guarantee expression of the virulence genes, but the fine tuning of these genes by the two-component system may be of less importance in pathogens which have lost the capacity to survive in the environment and during their life cycle apparently do not encounter dramatic differences in environmental conditions (MARTINEZ DE TEJADA et al. 1998). Does this mean that the quite complex regulatory network guided by the BvgAS system is merely an evolutionary relic of an ancestor which was a facultative coloniser of host organisms?

Recently, a third class of *vag* genes was identified which exhibits a different pattern of expression. Members of this class of genes are apparently transcribed only at intermediate BvgA-P concentrations, whereas their transcription is turned off again under high BvgA-P concentrations, when production of the other virulence factors including the toxins is maximal (COTTER and DiRITA 2000). This

so-called Bvg<sup>i</sup> phase of virulence gene expression appears to be highly conserved within the different members of the *B. bronchiseptica* cluster. The discovery of the Bvg<sup>i</sup> phase and the conservation of such subtle regulatory mechanisms between *B. pertussis* and *B. bronchiseptica* were taken as an indication that BvgAS-mediated fine tuning of virulence gene expression is still important, also for the strictly human pathogen *B. pertussis*. It is possible that slight differences in temperature within the lower respiratory tract and in the nasopharynx may be sufficient for the generation of different subsets of virulence genes in these host compartments. Different subsets of colonisation factors may in fact be required for colonisation of either the nasopharynx or deeper tissues of the respiratory tract, as the bacteria in the nasopharynx but not in the trachea should be readily transmitted to a new host.

## 6 Evolution of the Virulence Control System

Although the evolutionary origin of the virulence regulatory BvgAS system is unknown, this system may have enabled facultatively pathogenic ancestors of the *B. bronchiseptica* complex to distinguish between host and environmental locations due to its ability to sense differences in temperature. However, if the Bordetellae have evolved from environmental or facultative pathogenic microorganisms, it is likely that the BvgAS system originally may have been involved in the regulation of functions other than virulence. In fact, it is known that the BvgAS system also affects expression of housekeeping functions such as the cytochrome composition of the terminal oxidase of the respiratory chain (EZZELL et al. 1981) (Fig. 5). This finding is particularly interesting because of the recent identification of a PAS domain in the BvgS sensor protein by sequence similarity (TAYLOR and ZHULIN 1999). PAS domains are involved in sensing of signals as different as light, oxygen, redox potential and overall energy level of the cell and are found in all kingdoms of life. The domain is located in the cytoplasmic portion of BvgS between the membrane-spanning region and the transmitter domain. These data suggest that, at least in an evolutionary ancestor, the BvgS protein may have been involved either in sensing of oxygen or in the energy state of the cell and may be interconnected with the respiratory chain.

The BvgS protein therefore resembles other sensory proteins including the ArcB histidine kinase of *E. coli* which has a similar domain structure and is also a sensor of the redox potential (Fig. 6) (IUCHI and LIN 1993; PERRAUD et al. 1999). Together with the ArcA response regulator, the ArcB protein is engaged in the regulation of the aerobic metabolism modulon controlling expression of dehydrogenases, cytochrome complexes and enzymes involved in fatty acid metabolism and the tricarboxylic acid cycle (UNDEN and BONGAERTS 1997). Moreover, a detailed mutation analysis of the PAS domain provided some evidence for its involvement in ArcB-mediated signal transduction (MATSUSHIKA and MIZUNO 2000). Interestingly, the recent search for *virulence genes* in *B. bronchiseptica* identified a series

of genes and operons encoding metabolic functions very similar to those controlled by the ArcAB system in *E. coli* (Schneider and Gross, unpublished results). In addition, a role of the PAS domain for BvgS activity is indicated by the fact that the introduction of mutations in the so-called core region of the PAS domain caused an inactivation of the BvgS protein and mutations in the scaffold region of the PAS domain uncoupled the BvgS protein from its dependence on the signal input domain, causing a constitutive activity of the histidine kinase (BEIER et al. 1996; MANETTI et al. 1994; MILLER et al. 1992). Therefore, it is tempting to speculate that in a time when the bacteria acquired the ability to colonise host organisms, a two-component system involved in the adaptation to environments with different oxygen concentrations (and/or temperature) was linked to factors required for colonisation of a eucaryotic host. In fact, a change in the oxygen concentration or in intracellular redox conditions may have been an indicator of whether the bacteria were inside or outside a host organism as well as of differences in temperature. In this respect, it is important to note that *B. bronchiseptica* strains still have significant potential to survive harsh conditions outside of their host organisms for a prolonged period of time (PORTER and WARDLAW 1993). In contrast, *B. pertussis* lacks this capacity and, interestingly, several of the *vrg* genes encoding basic housekeeping functions related to the ArcAB regulon which were identified in *B. bronchiseptica* are in the process of degenerating in *B. pertussis* (Schneider and Gross, unpublished results). In the future, the characterisation of two-component systems present in other *Bordetella* species, and particularly in the environmental facultatively anaerobic species *B. petrii*, will show whether BvgAS-related systems are present in these organisms and, if so, what functions they have. The identification of *B. petrii* demonstrates the close relationship of the pathogenic *Bordetellae*

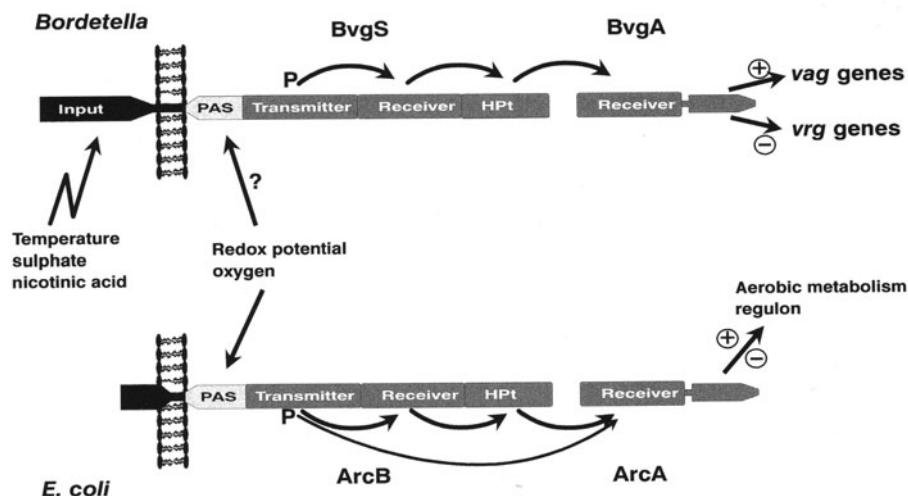


Fig. 6. Structural similarity of the BvgAS and ArcAB two-component phosphorelay systems

to very versatile environmental organisms capable even of anaerobic growth (VON WINTZINGERODE et al. 2001).

## 7 Conclusions

The genus *Bordetella* represents an extraordinary group of bacterial pathogens particularly suitable for the study of evolution of their pathogenic potential. Among the Bordetellae there are several closely related organisms with different host adaptations which also express different pathogenicities ranging from asymptomatic over acute to chronic forms of infections of the respiratory tract, but also closely related bacteria of environmental origin. Questions regarding the evolutionary origin of virulence traits, as well as their subsequent integration into regulatory networks of the bacteria required for their proper expression, and mechanisms of host adaptation are under investigation. Apart from general insights in evolutionary mechanisms of bacterial pathogenicity, these studies will also reveal whether the *B. bronchiseptica* cluster represents a reservoir for emerging pathogens such as *B. holmesii*, currently adapting to man as a host organism.

*Acknowledgements.* The authors would like to thank Dagmar Beier, Andreas Bock and Johannes Gross for their critical reading of the manuscript. F.v.W would like to thank Jacqueline Rudolf and Mareike Kurz for help with *rpoB* and *gyrB* sequencing. Financial support was provided by grants from the Deutsche Forschungsgemeinschaft (SFB479-A2) and the Fonds der Chemischen Industrie to R.G. and by the Deutsche Forschungsgemeinschaft (Schn 317/6-5) and the Studienstiftung des Deutschen Volkes to F.v.W.

## References

- Akerley BJ, Miller JF (1996) Understanding signal transduction during bacterial infection. *Trends Microbiol* 4:141–146
- Andersson SGE, Kurland CG (1998) Reductive evolution of resident genomes. *Trends Microbiol* 6: 263–268
- Antoine R, Alonso S, Raze D, Coutte L, Lesjean S, Willery E, Locht C, Jacob-Dubuisson F (2000a) New virulence-activated and virulence-repressed genes identified by systematic gene inactivation and generation of transcriptional fusions in *Bordetella pertussis*. *J Bacteriol* 182:5902–5905
- Antoine R, Raze D, Locht C (2000b) Genomics of *Bordetella pertussis* toxins. *Int J Med Microbiol* 290:301–305
- Arico B, Gross R, Smida J, Rappuoli R (1987) Evolutionary relationships in the genus *Bordetella*. *Mol Microbiol* 1:301–308
- Arico B, Miller J, Roy C, Stibitz S, Monack D, Falkow S, Gross R, Rappuoli R (1989) Sequences required for the expression of *Bordetella pertussis* virulence factors share homology with prokaryotic signal transduction proteins. *Proc Natl Acad Sci USA* 86:6671–6675
- Arp LH, Cheville NF (1984) Tracheal lesions in young turkeys infected with *Bordetella avium*. *Am J Vet Res* 45:2196–2201

- Banemann A, Gross R (1997) Phase variation affects long-term survival of *Bordetella bronchiseptica* in professional phagocytes. *Infect Immun* 65:3469–3473
- Beall BW, Sanden GN (1995) Cloning and initial characterization of the *Bordetella pertussis fur* gene. *Curr Microbiol* 30:223–226
- Beattie DT, Mahan MJ, Mekalanos JJ (1993) Repressor binding to a regulatory site in the DNA coding sequence is sufficient to confer transcriptional regulation of the *vir*-repressed genes (*vrg* genes) in *Bordetella pertussis*. *J Bacteriol* 175:519–527
- Beier D, Deppisch H, Gross R (1996) Conserved sequence motifs in the unorthodox BvgS two-component sensor protein of *Bordetella pertussis*. *Mol Gen Genet* 252:169–176
- Boschwitz JS, van der Heide HG, Mooi FR, Relman DA (1997) *Bordetella bronchiseptica* expresses the fimbrial structural subunit gene *fimA*. *J Bacteriol* 179:7882–7885
- Burns DL (1999) Biochemistry of type IV secretion. *Curr Opin Microbiol* 2:25–29
- Burns EH Jr, Norman JM, Hatcher MD, Bemis DA (1993) Fimbriae and determination of host species specificity of *Bordetella bronchiseptica*. *J Clin Microbiol* 31:1838–1844
- Busse HJ, Auling G (1992) The genera *Alcaligenes* and “*Achromobacter*”. In: Balows A, Trüper HG, Dworkin M, Harder W, Schleifer K-H (eds) *The prokaryotes*. Springer, Berlin Heidelberg New York, pp 2544–2555
- Cookson BT, Vandamme P, Carlson LC, Larson AM, Sheffield JVL, Kersters K, Spach DH (1994) Bacteremia caused by a novel *Bordetella* species, “*B. hinzii*”. *J Clin Microbiol* 32:2569–2571
- Cotter PA, DiRita VJ (2000) Bacterial virulence gene regulation: an evolutionary perspective. *Annu Rev Microbiol* 54:519–565
- Cotter PA, Miller JF (1997) A mutation in the *Bordetella bronchiseptica bvgS* gene results in reduced virulence and increased resistance to starvation, and identifies a new class of Bvg-regulated antigens. *Mol Microbiol* 24:671–685
- Dahlöf I, Bailie H, Kjelleberg S (2000) *rpoB*-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. *Appl Environ Microbiol* 66:3376–3380
- DeLey JP, Seger K, Kersters K, Mannheim W, Lievens A (1986) Intra- and intergeneric similarities of the *Bordetella* ribosomal ribonucleic acid cistrons: proposal for a new family, *Alcaligenaceae*. *Int J Syst Bacteriol* 36:405–414
- DeShazer D, Wood GE, Friedman RL (1995) Identification of a *Bordetella pertussis* regulatory factor required for transcription of the pertussis toxin operon in *Escherichia coli*. *J Bacteriol* 177:3801–3807
- Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP, Raoult D (2000) 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *J Clin Microbiol* 38:3623–3630
- Ezzell JW, Dobrogosz WJ, Kloos WE, Manclark CR (1981) Phase-shift markers in the genus *Bordetella*: loss of cytochrome d-629 in phase IV variants. *Microbiols* 31:171–181
- Ferry NS (1910) A preliminary report of the bacteria findings in canine distemper. *Am Vet Rev* 37:499–504
- Finn TM, Amsbaugh DF (1998) Vag8, a *Bordetella pertussis* bvg-regulated protein. *Infect Immun* 66:3985–3989
- Flak TA, Heiss LN, Engle JT, Goldman WE (2000) Synergistic epithelial responses to endotoxin and a naturally occurring muramyl peptide. *Infect Immun* 68:1235–1242
- Foss S, Heyen U, Harder J (1998) *Alcaligenes defragerans* sp. nov., description of four strains isolated on alkenoic monoterpenes ((+)-menthene, alpha-pinene, 2-carene, and alpha-phellandrene) and nitrate. *System Appl Microbiol* 21:237–244
- Fuchs TM, Deppisch H, Scarlato V, Gross R (1996) A new gene locus of *Bordetella pertussis* defines a novel family of prokaryotic transcriptional accessory proteins. *J Bacteriol* 178:4445–4452
- Funke G, Hess T, von Graevenitz A, Vandamme P (1996) Characteristics of *Bordetella hinzii* strains isolated from a cystic fibrosis patient over a 3-year period. *J Clin Microbiol* 34:966–969
- Gadea I, Cuena-Estralla M, Benito N, Blanco A, Fernandez-Guerrero ML, Valero-Guillem PL, Soriano F (2000) *Bordetella hinzii*, a “new” opportunistic pathogen to think about. *J Infect* 40:298–299
- Gentry-Weeks CR, Cookson BT, Goldman WE, Rimler RB, Porter SB, Curtiss R III (1988) Dermonecrotic toxin and tracheal cytotoxin, putative virulence factors of *Bordetella avium*. *Infect Immun* 56:1698–1707
- Gerlach G., von Wintzingerode F, Middendorf B, Gross R (2000) Evolutionary trends in the genus *Bordetella*. *Microbes Infect* 3:61–72
- Giardina PC, Foster LA, Musser JM, Akerley BJ, Miller JF, Dyer DW (1995) bvg repression of alcaligin synthesis in *Bordetella bronchiseptica* is associated with phylogenetic lineage. *J Bacteriol* 177:6058–6062
- Goodnow RA (1980) Biology of *Bordetella bronchiseptica*. *Microbiol Rev* 44:722–738

- Granowitz EV, Keenholz SL (1998) A pseudoepidemic of *Alcaligenes xylosoxidans* attributable to contaminated saline. Am J Infect Control 26:146–148
- Gross R, Rappuoli R (1988) Positive regulation of pertussis toxin expression. Proc Natl Acad Sci USA 85:3913–3917
- Gueirard P, Weber C, LeCoustumer A, Guiso N (1995) Human *Bordetella bronchiseptica* infection related to contact with infected animals: persistence of bacteria in host. J Clin Microbiol 33:2002–2006
- Gueirard P, Le Blay K, LeCoustumer A, Chaby R, Guiso N (1998) Variation in *Bordetella bronchiseptica* lipopolysaccharide during human infection. FEMS Microbiol Lett 162:331–337
- Hanada S, Shigematsu T, Shibuya K, Eguchi M, Hasegawa T, Suda F, Kamagata Y, Kanagawa T, Kurane R (1998) Phylogenetic analysis of trichloroethylene-degrading bacteria newly isolated from soil polluted with this contaminant. J Ferm Bioeng 86:539–544
- Hausman SZ, Burns DL (2000) Use of pertussis toxin encoded by ptx genes from *Bordetella bronchiseptica* to model the effects of antigenic drift of pertussis toxin on antibody neutralization. Infect Immun 68:3763–3767
- Hewlett EL (1995) The genus *Bordetella*. In: Mandell GL, Douglas RG, Bennett JE (eds) Principles and practice of infectious diseases. Churchill Livingstone, New York, pp 2078–2084
- Huang Z, Cui Z, Ishii M, Igarashi Y (2000) Microflora for efficient degradation of cellulolytic substrate Genbank accession no. AB039335
- Iuchi S, Lin EC (1993) Adaptation of *Escherichia coli* to redox environments by gene expression. Mol Microbiol 9:9–15
- Jayarao BM, Wang L (1999) A study on the prevalence of gram-negative bacteria in bulk tank milk. J Dairy Sci 82:2620–2624
- Jungnitz H, West NP, Walker MJ, Chhatwal GS, Guzman CA (1998) A second two-component regulatory system of *Bordetella bronchiseptica* required for bacterial resistance to oxidative stress, production of acid phosphatase, and in vivo persistence. Infect Immun 66:4640–4650
- Kania SA, Rajeev S, Burns EH, Odom TF, Holloway SM, Bemis DA (2000) Characterization of *fimN*, a new *Bordetella bronchiseptica* major fimbrial subunit gene. Gene 256:149–155
- Kattar MM, Chavez AP, Limaye AP, Rassoulian-Barret SL, Yarfitz SL, Carlson LC, Houze Y, Swanson S, Wood BL, Cookson BT (2000) Application of 16S rRNA gene sequencing to identify *Bordetella hinzzii* as the causative agent of fatal septicemia. J Clin Microbiol 38:789–794
- Kersters K, Hinz K-H, Hertle A, Segers P, Lievens A, Siegmann O, De Ley J (1984) *Bordetella avium* sp. nov., isolated from the respiratory tracts of turkeys and other birds. Int J Syst Bacteriol 34:56–70
- Kim BJ, Lee SH, Lyu MA, Kim SJ, Bai GH, Chae GT, Kim EC, Cha CY, Kook YH (1999) Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). J Clin Microbiol 37:1714–1720
- Kiredjian M, Holmes B, Kersters K, Guilvout I, De Ley J (1986) *Alcaligenes piechaudii*, a new species of the human clinical specimens and the environment. Int J Syst Bacteriol 36:282–287
- Kloos WE, Mohapatra N, Dobrogosz WJ, Ezzell JW, Manclark CR (1981) Deoxyribonucleotide sequence relationships among *Bordetella* species. Int J Syst Bacteriol 31:173–176
- Knapp S, Mekalanos JJ (1988) Two trans-acting regulatory genes (*vir* and *mod*) control antigenic modulation in *Bordetella pertussis*. J Bacteriol 170:5059–5066
- Kronvall G, Hansson H-S, von Stedingk LV, Törnqvist E, Falsen E (2000) Septic arthritis caused by a gram-negative bacterium representing a new species related to the *Bordetella-Alcaligenes* complex. APMIS 108:187–194
- Krooneman J, Wieringa EB, Moore ER, Gerritse J, Prins RA, Gottschal JC (1996) Isolation of *Alcaligenes* sp. strain L6 at low oxygen concentrations and degradation of 3-chlorobenzoate via a pathway not involving (chloro)catechols. Appl Environ Microbiol 62:2427–2434
- Lai CY, Baumann P, Moran N (1996) The endosymbiont (*Buchnera* sp.) of the aphid *Diuraphis noxia* contains plasmids consisting of *trpEG* and tandem repeats of *trpEG* pseudogenes. Appl Environ Microbiol 62:332–339
- LeBlay K, Gueirard P, Guiso N, Chaby R (1997) Antigenic polymorphism of the lipopolysaccharide from human and animal isolates of *Bordetella bronchiseptica*. Microbiology 143:1433–1441
- Ludwig W, Strunk O, Klugbauer S, Klugbauer N, Weizenegger M, Neumaier J, Bachleitner M, Schleifer KH (1998) Bacterial phylogeny based on comparative sequence analysis. Electrophoresis 19:554–568
- MacMillan DJ, Mau M, Walker MJ (1998) Characterisation of the urease gene cluster in *Bordetella bronchiseptica*. Gene 208:243–251
- Mandell WF, Garvey GJ, Neu HC (1987) *Achromobacter xylosoxidans* bacteraemia. Rev Infect Dis 9:1001–1005

- Manetti R, Arico B, Rappuoli R, Scarlato V (1994) Mutations in the linker region of BvgS abolish response to environmental signals for the regulation of the virulence factors in *Bordetella pertussis*. Gene 150:123–127
- Martinez de Tejada G, Miller JF, Cotter PA (1996) Comparative analysis of the virulence control systems of *Bordetella pertussis* and *Bordetella bronchiseptica*. Mol Microbiol 22:895–908
- Martinez de Tejada G, Cotter PA, Heininger U, Camilli A, Akerley BJ, Mekalanos JJ, Miller JF (1998) Neither the Bvg phase nor the vrg6 locus of *Bordetella pertussis* is required for respiratory infection in mice. Infect Immun 66:2762–2768
- Matsushika A, Mizuno T (2000) Characterization of three putative sub-domains in the signal-input domain of the ArcB hybrid sensor in *Escherichia coli*. J Biochem (Tokyo) 127:855–860
- Matthews RC, Preston NW (1997) *Bordetella*. In: Emmerson AM, Hawkey PM, Gillespie SH (eds) Principles and Practice of Clinical Bacteriology. John Wiley & Sons, Chichester, pp 323–336
- Mazengia E, Silva EA, Peppe JA, Timperi R, George H (2000) Recovery of *Bordetella holmesii* from patients with pertussis-like symptoms: use of pulsed-field gel electrophoresis to characterize circulating strains. J Clin Microbiol 38:2330–2333
- Merkel TJ, Keith JM (2000) The regulation of gene expression in *Bordetella pertussis* by quorum sensing. Int J Med Microbiol [Suppl] 30:A3
- Merkel TJ, Barros C, Stibitz S (1998) Characterization of the *bvgR* locus of *Bordetella pertussis*. J Bacteriol 180:1682–1690
- Middendorf B, Gross R (1999) Representational difference analysis identifies a strain-specific LPS biosynthesis locus in *Bordetella* spp. Mol Gen Genet 262:189–198
- Miller JF, Johnson SA, Black WJ, Beattie DT, Mekalanos JJ, Falkow S (1992) Constitutive sensory transduction mutations in the *Bordetella pertussis bvgS* gene. J Bacteriol 174:970–979
- Mollet C, Drancourt M, Raoult D (1997) *rpoB* sequence analysis as a novel basis for bacterial identification. Mol Microbiol 26:1005–1011
- Müller M, Hildebrandt A (1993) Nucleotide sequences of the 23S rRNA genes from *Bordetella pertussis*, *B. parapertussis*, *B. bronchiseptica* and *B. avium*, and their implications for phylogenetic analysis. Nucleic Acids Res 21:3320
- Musser JM, Hewlett EL, Peppler MS, Selander RK (1986) Genetic diversity and relationships in populations of *Bordetella* spp. J Bacteriol 166:230–237
- Parton R (1999) Review of the biology of *Bordetella pertussis*. Biologicals 27:71–76
- Perraud AL, Weiss V, Gross R (1999) Signalling pathways in two-component phosphorelay systems. Trends Microbiol 7:115–120
- Preston A, Allen AG, Cadisch J, Thomas R, Stevens K, Churcher CM, Badcock KL, Parkhill J, Barrell B, Maskell DJ (1999) Genetic basis for lipopolysaccharide O-antigen biosynthesis in bordetellae. Infect Immun 67:3763–3767
- Porter JF, Wardlaw AC (1993) Long-term survival of *Bordetella bronchiseptica* in lake water and in buffered saline without added nutrients. FEMS Microbiol Lett 110:33–36
- Porter JF, Parton R, Wardlaw AC (1991) Growth and survival of *Bordetella bronchiseptica* in natural waters and in buffered saline without added nutrients. Appl Environ Microbiol 57:1202–1206
- Raffel TR, Register KB, Temple LM (2000) Incidence of *Bordetella avium* in wild birds as measured by serology and tracheal cultures. abstr. C-57. Abstracts of the 100th General Meeting of the American Society for Microbiology. American Society for Microbiology, Los Angeles, Calif.
- Rappuoli R (1994) Pathogenicity mechanisms of *Bordetella*. Curr Top Microbiol Immunol 192:319–336
- Reinecke F, Groth T, Heise K-P, Joentgen W, Müller N, Steinbüchel A (2000) Isolation and characterization of an *Achromobacter xylosoxidans* strain B3 and other bacteria capable to degrade the synthetic chelating agent iminodisuccinate. FEMS Microbiol Lett 188:41–46
- Scarlato V, Arico B, Prugnola A, Rappuoli R (1991) Sequential activation and environmental regulation of virulence genes in *Bordetella pertussis*. EMBO J 10:3971–3975
- Scarlato V, Arico B, Domenighini M, Rappuoli R (1993) Environmental regulation of virulence factors in *Bordetella* species. Bioessays 15:99–104
- Schneider B, Gross R, Haas A (2000) Phagosome acidification has opposite effects on intracellular survival of *Bordetella pertussis* and *B. bronchiseptica*. Infect Immun 68:7039–7048
- Shelton CB, Crosslin DR, Casey JL, Ng S, Temple LM, Orndorff PE (2000) Discovery, purification, and characterization of a temperate transducing bacteriophage for *Bordetella avium*. J Bacteriol 182:6130–6136
- Shen Y, Stehmeier LG, Voordouw G (1998) Identification of hydrocarbon-degrading bacteria in soil by reverse sample genome probing. Appl Environ Microbiol 64:637–645

- Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. *Nature* 407:81–86
- Stefanelli P, Mastrantonio P, Hausman SZ, Giuliano M, Burns DL (1997) Molecular characterization of two *Bordetella bronchiseptica* strains isolated from children with cough. *J Clin Microbiol* 35: 1550–1555
- Steffen P, Goyard S, Ullmann A (1996) Phosphorylated BvgA is sufficient for transcriptional activation of virulence-regulated genes in *Bordetella pertussis*. *EMBO J* 15:102–109
- Stibitz S, Yang MS (1997) Genomic fluidity of *Bordetella pertussis* assessed by a new method for chromosomal mapping. *J Bacteriol* 179:5820–5826
- Stibitz S, Yang MS (1999) Genomic plasticity in natural populations of *Bordetella pertussis*. *J Bacteriol* 181:5512–5515
- Tang YW, Hopkins MK, Kolbert CP, Hartley PA, Severance PJ, Persing DH (1998) *Bordetella holmesii*-like organisms associated with septicemia, endocarditis, and respiratory failure. *Clin Infect Dis* 26:389–392
- Taylor BL, Zhulin IB (1999) PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev* 63:479–506
- Unden G, Bongaerts J (1997) Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. *Biochim Biophys Acta* 1320:217–234
- Väistönen OM, Weber A, Bennasar A, Rainey FA, Busse H-J, Salkinoja-Salonen MS (1998) Microbial communities of printing paper machines. *J Appl Microbiol* 84:1069–1084
- Vandamme P, Hommez J, Vancanneyt M, Monsieurs M, Hoste B, Cookson B, Wirsing von König CH, Kersters K, Blackall PJ (1995) *Bordetella hinpii* sp. nov., isolated from poultry and humans. *Int J Syst Bacteriol* 45:37–45
- Vandamme P, Heyndrickx M, Vancanneyt M, Hoste B, De Vos P, Falsen E, Kersters K, Hinz KH (1996) *Bordetella trematum* sp. nov., isolated from wounds and ear infections in humans, and reassessment of *Alcaligenes denitrificans* Rüger and Tan 1983. *Int J Syst Bacteriol* 46:849–858
- van der Zee A, Groenendijk H, Peeters M, Mooi FR (1996) The differentiation of *Bordetella parapertussis* and *Bordetella bronchiseptica* from humans and animals as determined by DNA polymorphism mediated by two different insertion sequence elements suggests their phylogenetic relationship. *Int J Syst Bacteriol* 46:640–647
- van der Zee A, Mooi FR, van Embden J, Musser JM (1997) Molecular evolution and host adaptation of *Bordetella* spp.: Phylogenetic analysis using multilocus enzyme electrophoresis and typing with three insertion sequences. *J Bacteriol* 179:6609–6617
- von Wintzingerode F, Schattke A, Siddiqui RA, Rösick U, Göbel UB, Gross R (2001) *Bordetella petrii* sp. nov., isolated from an anaerobic bioreactor and emended description of the genus *Bordetella*. *Int J Syst Evol Microbiol* 51:1257–1265
- Weiss AA (1992) The genus *Bordetella*. In: Balows A, Trüper HG, Dworkin M, Harder W, Schleifer K-H (eds) *The prokaryotes*. Springer, Berlin Heidelberg New York, pp 2530–2543
- Weiss AA, Falkow S (1984) Genetic analysis of phase change in *Bordetella pertussis*. *Infect Immun* 43:263–269.
- Weyant RS, Hollis DG, Weaver RE, Amin MFM, Steigerwalt AG, O'Connor SP, Whitney AM, Daneshvar MI, Moss CW, Brenner DJ (1995) *Bordetella holmesii* sp. nov., a new gram-negative species associated with septicemia. *J Clin Microbiol* 33:1–7
- Willems R, Paul A, van der Heide HG, ter Avest AR, Mooi FR (1990) Fimbrial phase variation in *Bordetella pertussis*: a novel mechanism for transcriptional regulation. *EMBO J* 9:2803–2809
- Woolfrey BF, Moody JA (1991) Human infections associated with *Bordetella bronchiseptica*. *Clin Microbiol Rev* 4:243–255
- Yabuuchi E, Kawamura Y, Kosako Y, Ezaki T (1998) Emendation of genus *Achromobacter* and *Achromobacter xylosoxidans* (Yabuuchi and Yano) and proposal of *Achromobacter ruhlandii* (Packer and Vishniac) comb. nov., *Achromobacter piechaudii* (Kiredjian et al.) comb. nov., and *Achromobacter xylosoxidans* subsp. *denitrificans* (Rüger and Tan) comb. nov. *Microbiol Immunol* 42:429–438
- Yamamoto S, Harayama S (1995) PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl Environ Microbiol* 61:1104–1109
- Yamamoto S, Harayama S (1998) Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *rpoD* and 16S rRNA genes. *Int J Syst Bacteriol* 48:813–819
- Yamamoto S, Kasai H, Arnold DL, Jackson RW, Vivian A, Harayama S (2000) Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. *Microbiology* 146:2385–2394

- Yih WK, Silva EA, Ida J, Harrington N, Lett SM, George H (1999) *Bordetella holmesii*-like organisms isolated from Massachusetts patients with pertussis-like symptoms. *Emerg Infect Dis* 5:441–443
- Yuk MH, Heininger U, Martinez de Tejada G, Miller JF (1998) Human but not ovine isolates of *B. parapertussis* are highly clonal as determined by PCR-based RAPD fingerprinting. *Infection* 26:270–273
- Zu T, Manetti R, Rappuoli R, Scarlato V (1996) Differential binding of BvgA to two classes of virulence genes of *Bordetella pertussis* directs promoter selectivity by RNA polymerase. *Mol Microbiol* 21: 557–565

# **Pathogenicity Islands and PAI-Like Structures in *Pseudomonas* Species**

K. LARBIG, C. KIEWITZ, and B. TÜMMLER

1	Introduction . . . . .	201
2	PAI-Like Structures in <i>P. aeruginosa</i> . . . . .	202
2.1	The <i>P. aeruginosa</i> PAO1 Genome . . . . .	202
2.2	The Exoenzyme S Regulon, a Cluster of Virulence Genes . . . . .	202
2.3	Islands of Clone- or Strain-Specific DNA Cause Inter- and Intraclonal Genomic Diversity in <i>P. aeruginosa</i> . . . . .	204
2.4	Origin of PAI-Like Structures . . . . .	205
3	PAIs or PAI-Like Structures in Other <i>Pseudomonads</i> . . . . .	207
3.1	The Hrp PAI of <i>P. syringae</i> . . . . .	207
3.2	Islands of Catabolic Versatility in <i>Pseudomonas</i> . . . . .	208
4	Conclusions . . . . .	209
	References . . . . .	210

## **1 Introduction**

The pseudomonads are a class of organisms which are ubiquitously distributed at low frequency in soil and aquatic habitats (OECD 1997). Some species (*Pseudomonas putida*, *Pseudomonas fluorescens*) are in addition profound colonizers of the rhizo- and phyllosphere and promote plant growth, whereas the phylogenetically related species *Pseudomonas syringae* belongs to the major bacterial phytopathogens. The type species *Pseudomonas aeruginosa* is an opportunistic pathogen for plants, animals and man (CAMPA et al. 1993).

Due to their metabolic versatility, pseudomonads are prime candidates for bioremediation of environmental pollutants such as halogenated hydrocarbons. Numerous *Pseudomonas* strains are employed for biotechnological purposes because of uncommon metabolic routes of synthesis and/or degradation of chemicals or secondary metabolites. It is tempting to assume that these peculiar metabolic features are encoded by some strain-specific extrachromosomal or chromosomal DNA, but, with a few exceptions, the genetic origin of versatility has not been

worked out for the pseudomonads. It is only the type species *P. aeruginosa* for which the genomic basis of its diversity has been analyzed in sufficient depth to allow some general conclusions.

## 2 PAI-Like Structures in *P. aeruginosa*

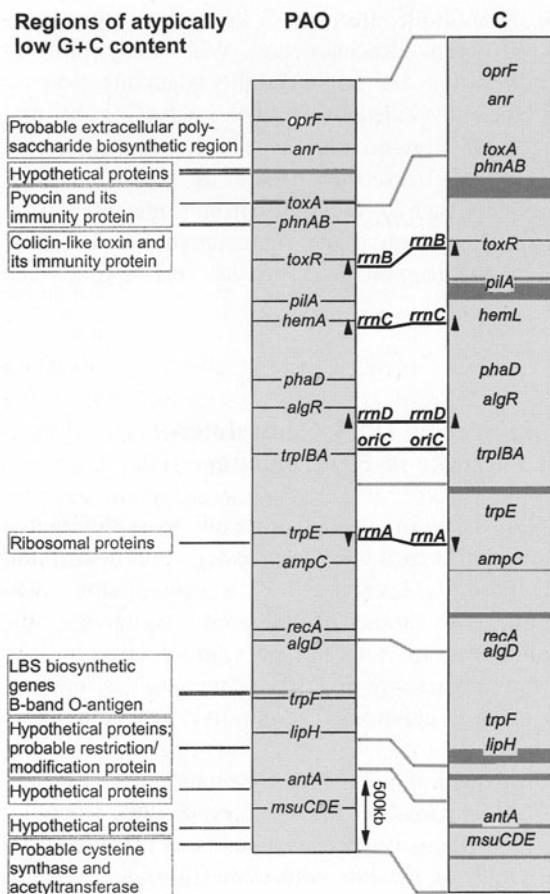
### 2.1 The *P. aeruginosa* PAO1 Genome

The whole genome sequence of the genetic reference strain *P. aeruginosa* PAO1 has been resolved (STOVER et al. 2000). The 6.26-Mb PAO genome differs from the other sequenced microbial genomes of lower size in its larger repertoire of regulatory genes and genes involved in the catabolism, transport and efflux of organic compounds. Numerous paralogs exist that encode the same enzymatic activity, albeit with probably a different substrate profile. These features endow *P. aeruginosa* with its remarkable metabolic versatility, particularly for degradation of uncommon nutrients such as toxic organic chemicals. These genes involved in uncommon catabolic pathways and the members of paralogous families share the same high G + C content with the genes encoding the essential pathways of intermediary metabolism, indicating that the paralogs and the catabolic versatility are a phylogenetically ancient signature of *P. aeruginosa*.

A uniformly high G + C content is characteristic for the genome. However, ten islands of 3.0kb or greater exhibit significantly lower G + C content and an unusual codon usage (Fig. 1), possibly indicative of recent horizontal transfer (STOVER et al. 2000). Some islands carry apparently dispensable genes that are not found in all *P. aeruginosa* strains and, for example, encode intraspecies toxins such as pyocins and/or proteins of yet unknown function (orphans). Other islands encode cellular appendages and elements of the outer membrane such as exopolysaccharide and LPS biosynthesis enzymes. The identified PAO1 islands contain the genetic elements whose phenotypic diversity is exploited by the classical methods of strain typing: LPS O antigen serotyping (HANCOCK et al. 1983) and pyocin typing (FYFE et al. 1984). Although systematic sequence and map information for strains other than PAO are scarce in *P. aeruginosa*, in all likelihood analogous islands will exist in other *P. aeruginosa* clones that confer the respective sero- and pyocin type.

### 2.2 The Exoenzyme S Regulon, a Cluster of Virulence Genes

*P. aeruginosa* PAO1 has a type III secretion system encoded by the 25,670-bp exoenzyme S (ExoS) regulon (FRANK 1997). Proteins encoded by the ExoS regulon and the *Yersinia* Yop regulon show a high level of amino acid homology. Like *Yersinia*, *P. aeruginosa* uses a contact-mediated translocation mechanism to transfer anti-host factors directly into eukaryotic cells (FRANK 1997). Intracellular



**Fig. 1.** Comparison of the chromosome maps of *P. aeruginosa* strains PAO and C. The circular maps were linearized and superimposed at the origin of replication (*oriC*) as the reference point to visualize differences in DNA contents on a linear scale. The connecting lines show the corresponding map coordinates of *oriC*, the four *rrn* operons (bold lines), and conserved *Spel* recognition sites (gray lines; SCHMIDT et al. 1996) on the PAO and C chromosomes. The localization of the ten regions with significantly lower G + C content in the PAO genome is indicated by the black lines on the left. The annotation was taken from the primary publication by STOVER et al. (2000) and from the Web site of the *P. aeruginosa* PAO1 genome project (<http://www.pathogenesis.com>). The shaded regions indicate size and map position of PAO- and clone C-specific islands of 5kb or greater. The sequence of common genes is conserved in strains C and PAO, but their physical distance in the two strains differs by the size of intervening PAO- or C-specific sequence

delivery of the anti-host factors ExoS and ExoT leads to disruption of eukaryotic signal transduction through ADP-ribosylation. For example, the cytotoxin ExoS uncouples the Ras-mediated signal transduction pathway. ExoS possesses two functional domains. The amino-terminal domain of ExoS elicits the disruption of actin, while the carboxyl-terminal domain possesses ADP-ribosyltransferase activity (PEDERSON et al. 1999). Another anti-host factor is ExoU. Production of ExoU correlates with acute cytotoxicity and lung injury; the presence of the *exoU* gene is an indicator for a more virulent strain (FINCK-BARBARCON et al. 1997). The type III secretion system transports at least two factors that kill macrophages: ExoU, which causes necrosis, and a second, yet unidentified protein, which induces apoptosis (HAUSER and ENGEL 1999). Moreover, the type III secretion system of some *P. aeruginosa* strains is able to induce rapid ExoU-independent oncrosis of macrophages and polymorphonuclear neutrophils: cellular and nuclear swelling, disintegration of the plasma membrane, and absence of DNA fragmentation (DACHEUX et al. 2000). These studies demonstrated that variations in cytotoxin

expression between strains cause the multiple effects of *P. aeruginosa* type III intoxication. However, the target cell type is also important. While lung epithelial cells show significant changes in morphology but not in viability when infected with *P. aeruginosa*, macrophages are efficiently killed by *P. aeruginosa* (COBURN and FRANK 1999). In summary, the ExoS regulon may be considered an ancient pathogenicity island (PAI). It has become irreversibly fixed in the genome because it lost all elements of genomic instability such as reversible chromosome integration signals. It acquired the species-specific codon usage characteristics and is now present in all *P. aeruginosa*, albeit with variations in cytotoxin expression between strains.

### **2.3 Islands of Clone- or Strain-Specific DNA Cause Inter- and Intraclonal Genomic Diversity in *P. aeruginosa***

Genomic diversity in *P. aeruginosa* is accomplished not only by a differential repertoire of phages or plasmids, but also by a considerable variation in chromosomal contents. The chromosome size of 11 unrelated *P. aeruginosa* clones was determined to vary between 5.2 and 7Mb (SCHMIDT et al. 1996). Apparently, the *P. aeruginosa* genome consists of a core of essential genes and a large bulk of dispensable genes which sustain the remarkable diversity of metabolism and lifestyle. Clone-specific DNA, which is not found in reference strain PAO1, contributes by about 20% to genome size.

Besides strain PAO, so far only the dominant *P. aeruginosa* clone C has been investigated in detail by physical genome analysis (SCHMIDT et al. 1996; RÖMLING et al. 1997). Members of clone C have frequently been isolated from inanimate and disease habitats, including the airways of patients with cystic fibrosis (RÖMLING et al. 1994). Figure 1 shows a comparison of the map of strain PAO with that of the most common clone C genotype, represented by strain C. The gene order and the gross backbone structure are conserved between PAO and C. However, the about 700kb larger strain C chromosome acquired 11 regions of 5kb or greater which are not present in strain PAO. Their size varies between 20kb and 160kb. Another three blocks of 5kb or greater are present in PAO but absent in strain C. They match exactly with the regions of significantly lower G + C-content (Fig. 1), substantiating their classification as a DNA island.

Intraclonal genome diversity was studied for 21 clone C isolates (RÖMLING et al. 1997). Clone C consists of closely related genotypes (also called clonal variants), each of which is characterized by a unique macrorestriction fragment pattern. Within clone C the total genome size varies at maximum by 300kb. The exchange of DNA blocks, which is a measurement for the actual gross-scale genome variability, took place to the same extent (> 10%) among clone C strains as among strains PAO and C. The acquisition and loss of DNA occurred preferentially around the terminus of replication but was not observed around the origin of replication (*oriC*), from about *rrnC* to *rrnA* (Fig. 1). Three regions close to the *phnAB*, *pilA* and

*lipH* loci were subject to extensive variation processes, characterizing them as hypervariable regions of the clone C chromosome.

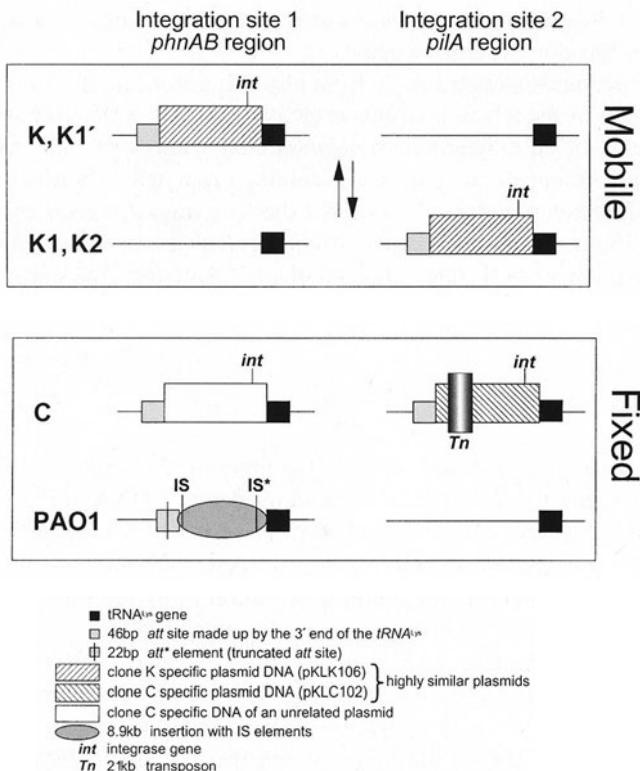
The conclusions on genome diversity drawn from physical genome mapping of clone C isolates also apply to the whole taxospecies (RÖMLING et al. 1995; HEUER et al. 1998). The screening of 60 different *P. aeruginosa* clones uncovered a conserved *oriC*-reactive *SpeI* fragment in almost all isolates examined, whereas a random distribution of fragment size was observed for the *SpeI* fragment encoding *phnAB*, *pilA* and *lipH*. Hence, these three hypervariable regions are probably the key players in the generation of both inter- and intraclonal genomic diversity in *P. aeruginosa*.

## 2.4 Origin of PAI-Like Structures

The evolution of PAIs is usually deduced from the comparative mapping and sequencing of lineages, because the ancestral donor and recipient of DNA are not available for analysis. *P. aeruginosa* is one of the few species for which the acquisition of a DNA island has been documented by analysis of sequential isolates from the same source. These primary data about genomic evolution in natural populations are not biased by any heuristic assumptions or models (KIEWITZ et al. 2000).

The *P. aeruginosa* genome contains two identical *tRNA<sup>Lys</sup>* genes in two hypervariable regions close to the *phnAB* and *pilA* loci (Fig. 1). *P. aeruginosa* clone K isolates from the airways of a patient with cystic fibrosis were found to repetitively exchange the incorporation of a 106-kb plasmid between the 3' ends of the two *tRNA<sup>Lys</sup>* genes (Fig. 2, KIEWITZ et al. 2000), whereby also the free plasmid was detectable to a variable extent in the sequential isolates. Temperate bacteriophages typically integrate into the bacterial chromosome via site-specific recombination at an *attB* site that is within, or overlaps with, the 3' end of a tRNA gene. The episomal clone K plasmid encodes an integrase *int* gene in the vicinity of a 46-bp *attP* sequence that is complementary to the 3' end of the *tRNA<sup>Lys</sup>* gene (*attB* site), which was utilized for the integration into the chromosome. The dominant clone C incorporated a highly similar plasmid at the *tRNA<sup>Lys</sup>* gene of the *pilA* region and an unrelated plasmid at the *tRNA<sup>Lys</sup>* gene of the *phnAB* region (Fig. 2). The plasmid sequence was stably fixed in the chromosome, if the plasmid contig was disrupted by a further sequence such as a transposon (Fig. 2). An *att*-spacer-*int* contig is characteristic for phage attachment sites but has so far not been demonstrated for plasmids from any other gram-negative taxospecies. In *P. aeruginosa* the plasmid *int*-*attB* sequences are utilized to reversibly incorporate the plasmid into the chromosome at the *attB* sites of tRNA genes. The plasmid sequence becomes irreversibly fixed in the chromosome by secondary processes such as transposon mutagenesis. These data demonstrate how a mobile extrachromosomal plasmid is first reversibly and later irreversibly captured by the host chromosome to become a PAI-like island.

Interestingly, a 8.9-kb island with atypically low G+C content was integrated into the strain PAO1 chromosome at the same *tRNA<sup>Lys</sup>* recognition site in the



**Fig. 2.** Evolution from plasmids into PAI-like structures in *P. aeruginosa*. Islands in *phnAB* and *pilA* regions of *P. aeruginosa* PAO, clone K and clone C chromosomes. *Upper panel*: Intragenomic rearrangements caused by reversible and sequential integration of plasmid pKLK106 at the *attB* sites of one of the two *tRNA<sup>Lys</sup>* genes. *Lower panel*: Stable islands. Clone C incorporated two different plasmids at the *attB* sites of the two *tRNA<sup>Lys</sup>* genes. Plasmid pKLC102 became irreversibly fixed by the insertion of a 21-kb transposon. Strain PAO1 harbors a small PAI-like structure in the *phnAB* region flanked by the *tRNA<sup>Lys</sup>* gene and a truncated *att* site

*phnAB* region (Fig. 1; STOVER et al. 2000). The discernible *att*\* element at the other border of the island is truncated (Fig. 2, bottom row; KIEWITZ et al. 2000) which could be why the island is stably maintained in the PAO1 strain. Not only the significantly different G+C content and codon usage characteristics, but also the annotation of the 8.9-kb insertion provides evidence that the DNA block was acquired by horizontal gene transfer (KIEWITZ et al. 2000). Two different IS elements are located close to the borders and encode transposases, and two other ORFs encode a colicin-like toxin and its immunity protein.

In summary, the DNA blocks in the *phnAB* and the *pilA* regions fulfill many characteristics of a PAI, as defined by HACKER et al. (1997): PAIs (a) include many virulence genes, (b) are selectively present in pathogenic strains, (c) have different G+C content than the global host chromosome, (d) occupy large chromosomal regions, (e) are often flanked by direct repeats, (f) are bordered by tRNA genes

and/or cryptic mobile genetic elements, and (g) are unstable. The DNA blocks found in *P. aeruginosa* clone K meet the criteria (c)–(g), but not (a) and (b). According to the analysis of clone K and clone C isolates the genomic organization of clone- or strain-specific DNA in *P. aeruginosa* is very similar to that of PAIs in pathogens, but the presence of the PAI-like blocks is associated neither with virulence nor with a disease phenotype or a peculiar habitat (RÖMLING et al. 1994, 1997; KIEWITZ and TÜMMLER 2000). The same spectrum of *P. aeruginosa* clones is found in the inanimate environment and in infected plants and animals, including man, which is probably favorable for the ubiquitous lifestyle of a bacterium that can colonize virtually all aquatic habitats with some minimal supply of minerals and carbon sources.

### 3 PAIs or PAI-Like Structures in Other *Pseudomonads*

PAIs and PAI-like structures should exist not only in the type species, but in the whole *Pseudomonas* genus. However, our knowledge is so far restricted to examples from *P. syringae* and *P. putida*. The most thoroughly studied PAI is the Hrp region of the phytopathogenic *P. syringae*.

#### 3.1 The Hrp PAI of *P. syringae*

*P. syringae* elicits a wide variety of symptoms in plants, including blights (rapid death of tissue), leaf spots, and galls. The species is divided into pathogenic variants (pathovars) which differ by host range. The genetic basis of pathogenicity includes global regulators (KITTEN et al. 1998; RICH et al. 1994), virulence factors such as phytotoxins (BENDER et al. 1999) and the *hrp/hrc* cluster (HE 1998; GALAN and COLLMER 1999). The *hrp* region (*hypersensitive response* and *pathogenicity*) is probably conserved among necrosis-causing gram-negative plant pathogens and has been sequenced in three *P. syringae* strains (ALFANO et al. 2000), *Erwinia amylovora* Ea321, *Xanthomonas campestris* pv. *vesicatoria* (*Xcv*) 85–10 and *Ralstonia solanacearum* GMI1000 (ALFANO and COLLMER 1997). In nonpermissive plants the proteins encoded by the *hrp* region elicit the hypersensitive response (HR), which is characterized by a rapid (~24h) induction of programmed cell death in plant cells that are in contact with the pathogen. In contrast, in permissive hosts, the bacteria continue to grow for several days before producing visible disease symptoms (HE 1998; GALAN and COLLMER 1999). The ability to produce either of these reactions in plants is mediated by the *hrp/hrc* genes that encode a type III protein secretion pathway that translocates Avr (*avirulence*) and Hop (*Hrp-dependent outer protein*) effector proteins across the plant cell wall and into plant cells. Several genes in the *hrp/hrc* cluster are involved in the transcriptional regulation of the *hrp/hrc* regulon that is induced

following infiltration into host tissue (COLLMER et al. 2000; PENALOZA-VAZQUEZ et al. 2000).

Comparative sequencing in three *P. syringae* isolates revealed that the *hrp/hrc* gene cluster resides at the center of the ~50kb Hrp PAI, flanked by two further distinct loci that make different contributions to pathogenicity (ALFANO et al. 2000; COLLMER et al. 2000). These are an exchangeable effector locus (EEL) and a conserved effector locus (CEL) in a tripartite mosaic Hrp PAI. The Hrp PAI is linked to a *tRNA<sup>Leu</sup>* gene found also in *P. aeruginosa* but without linkage to Hrp system genes. The EEL is a locus with low G + C content that contains a strain-specific set of ORFs which are probably involved in fine-tuning the parasitic fitness of *P. syringae* strains with various plant hosts. Moreover, the EELs contain sequences homologous to IS elements, transposases, phage integrase genes, and plasmids. The hypervariable EEL terminates at the 3' end of *tRNA<sup>Leu</sup>*, which is the apparent integration site of the Hrp PAI in the ancestral *Pseudomonas* genome. Hence, the EEL shows the same genetic organization as the mobile DNA islands in *P. aeruginosa* (see above). In contrast to the EEL, whose apparent instability suggests ongoing rapid evolution at this locus, the CEL sequences at the right border of the *hrp/hrc* genes are essential for pathogenicity, are conserved in the species, have a G + C content comparable to that of the *hrp/hrc* genes, and contain no sequences similar to known mobile genetic elements. In summary, the Hrp-mediated pathogenicity in *P. syringae* is dependent on a mosaic of species-, pathovar- and strain-specific genes.

### 3.2 Islands of Catabolic Versatility in *Pseudomonas*

The catabolic versatility of *Pseudomonas* is exploited for the bioremediation of environmental pollutants. Biphenyl-utilizing bacteria which co-metabolize polychlorinated biphenyls (PCB) have been isolated from numerous genera, including *Pseudomonas* (ABRAMOWICZ 1990). The gene clusters encoding biphenyl-PCB degradation (termed *bph*) are present in bacterial chromosomes, plasmids, and transposons. In *P. putida* strain KF715 the *bph* and *sal* gene clusters that encode the biphenyl and salicylate metabolic pathways are located in a 90-kb chromosomal region termed the *bph-sal* element, that is highly prone to deletion and easily transferred to other strains (NISHI et al. 2000). When the bacteria were grown in nutrient-rich medium, most cells lost either the *bph* genes (~40kb deletion) or both *bph* and *sal* genes (70kb deletion). According to Southern analysis, the common deletion site is at or close to an IS-like sequence similar to IS5. The *bph-sal* element behaves like conjugative transposons (SCOTT 1992) which excise themselves from the genome in which they are integrated, transfer themselves by conjugation into a recipient cell, and integrate into the recipient's genome.

Two further conjugative catabolic transposons are currently known, Tn4371 (MERLIN et al. 1999) and the *clc* element of *Pseudomonas* sp. strain B13 (RAVATN et al. 1998a,b). Tn4371 has a modular structure: It carries an enteric phage-like integration system and RP4/Ti-like conjugation genes and encodes enzymes for the

degradation of biphenyl and 4-chlorobiphenyl compounds into benzoate and 4-chlorobenzoate derivatives like those found in *Pseudomonas* sp. strain KKS102 (MERLIN et al. 1999).

The *clc* element of *Pseudomonas* sp. strain B13 (RAVATN et al. 1998a,b) resembles a pathogenicity island. The sewage isolate B13 utilizes 3-chlorobenzoate as its sole carbon and energy source (DORN et al. 1974), whereby the substrate is first oxidized to chlorocatechols and then converted to 3-oxoadipate by enzymes encoded by the *clc* gene cluster. These genes are located on the self-transmissible 105-kb *clc* element that integrated into the chromosome at an 18-bp *attB* site within the 3' end of a *tRNA<sup>Gly</sup>* gene (RAVATN et al. 1998b). An ORF (*int-B13*) coding for an integrase of the bacteriophage P4 family starts ~200bp from the junction between the *clc* element's right junction and the *tRNA<sup>Gly</sup>* gene. Int-B13 was responsible for site-specific recombination between the *clc* element's *attP* site and the chromosomal *attB* site. In the presence of chlorobenzene the *clc* element was transferred from strain B13 to *P. putida* F1, whereby multiple copies were targeted in tandem into the recipient chromosome at the 3' ends of its *tRNA<sup>Gly</sup>* genes (RAVATN et al. 1998a). The *clc* element is the first example of a "degradation island" that contains genes encoding the degradation of xenobiotics.

## 4 Conclusions

PAIs and PAI-like structures have so far been detected in several species of the genus *Pseudomonas*. However, only the type species *P. aeruginosa* has been studied in sufficient detail to allow some general conclusions about the role of PAI-like structures for genomic diversity. *P. aeruginosa* has a 5- to 7-Mb circular chromosome (SCHMIDT et al. 1996). Although the contig of common genes is conserved in almost all *P. aeruginosa* strains (HEUER et al. 1998) and its sequence diversity of about 0.3% is one order of magnitude lower than in comparable housekeeping genes of enterobacteria (KIEWITZ and TÜMMLER 2000), 1- to 200-kb PAI-like blocks of clone- or strain-specific DNA interrupt the conserved gene contig and give rise to a mosaic genome structure (RÖMLING et al. 1997). The 15%–30% portion of intra- and interclonal genomic diversity is not equally distributed but clusters in hypervariable regions of the chromosome (RÖMLING et al. 1995; HEUER et al. 1998). In contrast to the findings in enterobacteria, the presence or absence of a specific PAI-like DNA block is not associated with a peculiar habitat or a disease phenotype (RÖMLING et al. 1997; KIEWITZ and TÜMMLER 2000). The 3' ends of tRNA genes (*att* sites) seem to be the major integration sites for these DNA blocks (KIEWITZ et al. 2000). Both phages and plasmids are incorporated at the *att* sites into the chromosome (KIEWITZ et al. 2000). The promiscuous use of typical phage attachment sites by conjugative genetic elements could be one of the major mechanisms used by *P. aeruginosa* to generate its genome structure: The contig of conserved genes is interrupted by islands of clone- and strain-specific DNA. Sec-

ondary events such as nucleotide substitutions or the truncation of the *att* site and/or the insertion of further elements (transposons, integrons, IS elements, etc.) irreversibly fix the initially mobile island in the chromosome.

*Acknowledgements.* Work in the authors' laboratory has been supported by grants from the *Deutsche Forschungsgemeinschaft*, the *Christiane-Herzog-Stiftung* and the *CF Selbsthilfe Bundesverband e.V.*

## References

- Abramowicz DA (1990) Aerobic and anaerobic biodegradation of PCBs: a review. *Crit Rev Biotechnol* 10:241–251
- Alfano JR, Collmer A (1997) The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *J Bacteriol* 179:5655–5662
- Alfano JR, Charkowski AO, Deng W-L, Badel JL, Petnicki-Ocwieja T, van Dijk K, Collmer A (2000) The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *Proc Natl Acad Sci USA* 97:4856–4861
- Bender CL, Alarcon-Chaidez F, Gross DC (1999) *Pseudomonas syringae* phytotoxins: mode of action, regulation and biosynthesis by peptide and polyketide synthetases. *Microbiol Mol Biol Rev* 63:266–292
- Campa M, Bendinelli M, Friedman H (1993) *Pseudomonas aeruginosa* as an opportunistic pathogen. Plenum, New York
- Coburn J, Frank DW (1999) Macrophages and epithelial cells respond differently to the *Pseudomonas aeruginosa* type III secretion system. *Infect Immun* 67:3151–3154
- Collmer A, Badel JL, Charkowski AO, Deng W-L, Fouts DE, Ramos AR, Rehm AH, Anderson DM, Schneewind O, van Dijk K, Alfano JR (2000) *Pseudomonas syringae* Hrp type III secretion system and effector proteins. *Proc Natl Acad Sci USA* 97:8770–8777
- Dacheux D, Toussaint B, Richard M, Brocier G, Croize J, Attree I (2000) *Pseudomonas aeruginosa* cystic fibrosis isolates induce rapid, type III secretion-dependent, but ExoU-independent, oncosis of macrophages and polymorphonuclear neutrophils. *Infect Immun* 68:2916–2924
- Dorn E, Hellwig M, Reineke W, Knackmuss H (1974) Isolation and characterization of a 3-chlorobenzoate degrading pseudomonad. *Arch Microbiol* 99:61–70
- Finck-Barbançon V, Goranson J, Zhu L, Sawa T, Wiener-Kronish JP, Fleiszig SMJ, Wu C, Mende-Mueller L, Frank DW (1997) ExoU expression by *Pseudomonas aeruginosa* correlates with acute cytotoxicity and epithelial injury. *Mol Microbiol* 25:547–557
- Frank DW (1997) The exoenzyme S regulon of *Pseudomonas aeruginosa*. *Mol Microbiol* 26:621–629
- Fyfe JAM, Harris G, Govan JRW (1984) Revised pyocin typing method for *Pseudomonas aeruginosa*. *J Clin Microbiol* 20:47–50
- Galan JE, Collmer A (1999) Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* 284:1322–1328
- Hacker J, Blum-Oehler G, Mühldorfer I, Tschape H (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* 23:1089–1097
- Hauser AR, Engel JN (1999) *Pseudomonas aeruginosa* induces type-III-secretion-mediated apoptosis of macrophages and epithelial cells. *Infect Immun* 67:5530–5537
- Hancock REW, Mutharia LM, Chan L, Darveau RP, Speert DP, Pier GB (1983) *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis: a class of serum-sensitive, nontypable strains deficient in lipopolysaccharide O side chains. *Infect Immun* 42:170–177
- He SY (1998) Type III protein secretion systems in plant and animal pathogenic bacteria. *Annu Rev Phytopathol* 36:363–392
- Heuer T, Bürger C, Maaß G, Tümmler B (1998) Cloning of prokaryotic genomes in yeast artificial chromosomes: application to the population genetics of *Pseudomonas aeruginosa*. *Electrophoresis* 19:486–494

- Kiewitz C, Tümmler B (2000) Sequence diversity of *Pseudomonas aeruginosa*: impact on population structure and genome evolution. *J Bacteriol* 182:3125–3135
- Kiewitz C, Larbig K, Klockgether J, Weinel C, Tümmler B (2000) Monitoring genome evolution ex vivo: reversible chromosomal integration of a 106-kb plasmid at two tRNA<sup>Lys</sup> gene loci in sequential *Pseudomonas aeruginosa* airway isolates. *Microbiology* 146:2365–2373
- Kitten T, Kinscherf TG, McEvoy JL, Willis DK (1998) A newly identified regulator is required for virulence and toxin production in *Pseudomonas syringae*. *Mol Microbiol* 28:917–929
- Merlin C, Springael D, Toussaint A (1999) Tn4371: a modular structure encoding a phage-like integrase, a *Pseudomonas*-like catabolic pathway, and RP4-like transfer functions. *Plasmid* 41:40–54
- Nishi A, Tominaga K, Furukawa K (2000) A 90-kilobase conjugative chromosomal element coding for biphenyl and salicylate catabolism in *Pseudomonas putida* KF715. *J Bacteriol* 182:1949–1955
- OECD (1997) Consensus document on information used in the assessment of environmental applications involving *Pseudomonas*. Series on Harmonization of Regulatory Oversight in Biotechnology, no. 6. OECD, Paris
- Pederson KJ, Vallis AJ, Aktories K, Frank DW, Barbieri JT (1999) The amino-terminal domain of *Pseudomonas aeruginosa* ExoS disrupts actin filaments via small-molecular-weight GTP-binding proteins. *Mol Microbiol* 32:393–401
- Penaloza-Vazquez A, Preston GM, Collmer A, Bender CL (2000) Regulatory interactions between the Hrp type III protein secretion system and coronatine biosynthesis in *Pseudomonas syringae* pv. *tomato* DC3000. *Microbiology* 146:2447–2456
- Ravatn R, Studer S, Springael D, Zehnder AJ, van der Meer JR (1998a) Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. Strain B13. *J Bacteriol* 180:4360–4369
- Ravatn R, Studer S, Zehnder AJ, van der Meer JR (1998b) Int-B13, an unusual site-specific recombinase of the bacteriophage P4 integrase family, is responsible for chromosomal insertion of the 105-kilobase clc element of *Pseudomonas* sp. strain B13. *J Bacteriol* 180:5505–5514
- Rich JJ, Kinscherf TG, Kitten T, Willis DK (1994) Genetic evidence that the *gacA* gene encodes the cognate response regulator for the *lemA* sensor in *Pseudomonas syringae*. *J Bacteriol* 176:7468–7475
- Römling U, Wingender J, Müller H, Tümmler B (1994) A major *Pseudomonas aeruginosa* clone common to patients and aquatic habitats. *Appl Environ Microbiol* 60:1734–1738
- Römling U, Greipel J, Tümmler B (1995) Gradient of genomic diversity in the *Pseudomonas aeruginosa* chromosome. *Mol Microbiol* 17:323–332
- Römling U, Schmidt KD, Tümmler B (1997) Large genome rearrangements discovered by the detailed analysis of 21 *Pseudomonas aeruginosa* clone C isolates found in environment and disease habitats. *J Mol Biol* 271:386–404
- Schmidt KD, Tümmler B, Römling U (1996) Comparative genome mapping of *Pseudomonas aeruginosa* PAO with *P. aeruginosa* C, which belongs to a major clone in cystic fibrosis patients and aquatic habitats. *J Bacteriol* 178:85–93
- Scott JR (1992) Sex and the single circle: conjugative transposon. *J Bacteriol* 174:6005–6010
- Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FSL, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltby L, Tolentino E, Westbroek-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK-S, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock REW, Lory S, Olson MV (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406:959–964

# Subject Index

## A

### A + T

- progenitor 152

- ratios 152

A<sub>1</sub>-B<sub>5</sub> conformation 57

accessory genetic elements 161

- insertion sequence elements 161

- plasmids 161

- prophages 161

- transposons 161

*Achromobacter* spp. 177, 180–184

- *A. piechaudii* 180

- *A. xylosoxidans* 180–181

- *Achromobacter-Bordetella* group 181

aerobactin operon 57, 59–60

*afa*-8 19

- duplication 19

- EDL933 20

- genome 20

- P4 integrase 19

AIDA-1 27

*Alcaligenes* sp. 177, 180–181

- *Al. faecalis* 180

antibiotic resistance 90, 96, 103

*attL* 53

*attR* 53

autotransporters 56

## B

### bacterial

- conjugation 7–8

- - - conjugative plasmid 8

- - - vector 8

- genomes of gene serving, molecular evolution 11

bacteriophage 18, 24–27

- integrase 19, 24, 26

- and *sopE* 25

barriers against DNA acquisition 9

- restriction/modification systems 9, 11

biological evolution 2–3

black holes, EIEC 44

## *Bordetella* sp., phylogenetic relationships

and virulence evolution 177–194

- *Achromobacter-Bordetella* group 181

- *B. avium* 179

- *B. bronchiseptica* 177–179, 183–189

- - PAI 186

- - - representational difference analysis (RDA) 185

- *B. hinzii* 179

- *B. holmesii* 179, 183

- *B. parapertussis* 177–179, 183–189

- *B. pertussis* 177–179, 181, 183–189

- *B. petrii* 180–181

- *B. trematum* 179, 181

- BvgAS system 189

- evolutionary impact of virulence gene regulation on host adaption 187–189

- genomic aspects of *B.* evolution 184–187

- phylogenetic relationship of host-associated *B.* spp. with *B. petrii* and others 180–184

- 16S rDNA 180

- taxonomy 177–180

- virulence

- - - control system, evolution of 192

- - - gene expresseion 187

*Buchnera* sp. 169

- symbiosis 169

BvgAS system 189

BvgS system 189

## C

*cadA* 25, 61

capsular determinants, genomic structure

137–152

- capsule

- - diversity

- - - of *E. coli* capsules 138–145

- - - in *Streptococcus pneumoniae* 149

- - - gene clusters

- - - in *Neisseria meningitidis* 146–148

- - - in other gram-positive pathogens 151

- - - in *Streptococcus pneumoniae* 146–148

- capsular determinants, genomic structure  
 - - genetics of *E. coli*  
 - - - group 1 and 4 capsules 140–142  
 - - - group 2 capsules 142–144  
 - - - group 3 capsules 144  
 - localization of the group 2 and 3 capsule gene clusters on the *E. coli* chromosome 145  
 - transposition and recombination of *Streptococcus pneumoniae* 150–151
- catabolic versatility of *Pseudomonas* 208
- CEL (conserved effector locus) 208
- chain length, O-antigen 110, 113
- Choleraesuis* serovar 79
- Citrobacter sp.* 102, 170
- clic* element 208
- clone- or strain-specific DNA 209
- CNF1 (cytotoxic necrotizing factory type 1)  
 19, 20
- colicin  
 - colicin V 57  
 - immunity 60
- conjugative transposons 208
- cryptic mobility genes 57
- CT/LT-like toxin family 56
- curli 61
- cytotoxic necrotizing factory type 1 (CNF1)  
 19, 20
- D**
- DNA  
 - 16S rDNA 180, 183  
 - acquisition 7–11  
 - - barriers against DNA acquisition 9  
 - - gene vector 7  
 - - genetic stability 7  
 - - horizontal gene transfer 8–9  
 - - in small steps 9  
 - clone- or strain-specific 209  
 - inversion 6  
 - secondary crossing-over sites 6  
 - rearrangement 4–7, 10, 11  
 - - general recombination (see recombination) 7  
 - - novel combinations 10  
 - - site-specific recombination (see there) 6–7  
 - - transposition of mobile genetic elements (see there) 5–6  
 - repair systems 11  
 - *Salmonella*-specific 81  
 - transformation 7–8  
 - - competence 8  
 - - DNA uptake 8
- dsdC* 53
- DT204 82
- DT204c 82
- DT49 82
- E**
- EAEC (enteroaggregative *E. coli*) 32, 41–43, 162  
 - high-pathogenic island (HPI) 42  
 - *she* PAI 43
- EDL933 145
- EEL (exchangable effector locus) 208
- EHEC (enterohaemorrhagic *E. coli*) 22, 26, 38–40, 160  
 - genome, EHEC strain EDL933 20, 22, 26, 27  
 - locus of enterocyte effacement (see LEE) 39  
 - O-island 39–40
- EIEC (enteroinvasive *E. coli*) 32, 43–44, 162
- black holes 44
- *Shigella* pathogenicity island 2 (SHI-2) 43–44
- enterobacteria 157–171  
 - enterobacter cloacae 102  
 - evolution 158–160  
 - gene pools 168–170  
 - genome plasticity 157–171
- enterotoxin 57
- EPEC (enteropathogenic *E. coli*) 24, 32–38, 162  
 - E2348/69 38  
 - EspC PAI 37–38  
 - locus of enterocyte effacement (see LEE) 33–37
- Escherichia coli* (*E. coli*) 15–28, 31–45, 68, 102–103, 110, 138, 157–159  
 - capsules, diversity of 138–146  
 - - genetics of *E. coli*  
 - - - group 1 and 4 capsules 140–142  
 - - - group 2 capsules 142–144  
 - - - group 3 capsules 144  
 - diarrhea 32  
 - - localization of the group 2 and 3 capsule gene clusters on the *E. coli* chromosome 145  
 - enteroaggregative (see EAEC) 32, 41–43, 162  
 - enterohaemorrhagic (see EHEC) 22, 26, 32, 38–40, 160  
 - enteroinvasive (see EIEC) 32, 43–44, 162  
 - enteropathogenic (see EPEC) 24, 32–38, 162  
 - enterotoxigenic (see ETEC) 32, 40–41, 162  
 - extraintestinal *E. coli*, pathogenicity island of 165–166

- flexible and core gene pool of *E. coli* 160–161
- gene pools of pathogenic *E. coli* 162–165
- intestinal *E. coli*, mobile genetic elements of 166–168
- K-12 genome 17, 22, 73, 145
- lipid A/core genes of 122
- LPS core synthesis genes 126–129
- meningitis-causing 16, 20, 166
- MLEE (multilocus enzyme electrophoresis) 159
- MLST (multilocus sequence typing) 159
- model organism 160–168
- nomenclature 110
- O-antigen genes 113–114
  - - *E. coli* O157 113
  - - *E. coli* O111 113
  - - *E. coli* O55 113
- pathotypes 157, 161–162
- pyelonephritis 20
- ribotyping 159
- sepsis-causing 16, 19
- *Shigella* vs. *E. coli* 51
- uropathogenic (see UPEC) 16–28, 162
- virulence determinants 158
- esg** 61
- ETEC** (enterotoxigenic *E. coli*) 32, 40–41, 162
  - Tia-PAI 41
  - toxicogenic invasion loci 41
- evolution evolutionary**
  - distances by molecular clock 9
  - genes 2–3, 11–12
    - - fine-tuned 12
    - - functional genomics 2
    - - proteomics 2
    - - to second-order selection 11–12
      - - - fine-tuning 12
    - strategy, molecular evolution 7
  - tree, molecular evolution 10–11
    - - horizontal connectors 10
- exoenzyme S (ExoS) regulation** 202–204
  - PAI 204
- F**
- F1 antigen** 93
- far-east scarlet-like fever** 95
- flexible and core gene pool of *E. coli*** 160–161
- flipase** 53
- fusions** 6, 10
  - gene 6, 10
  - operation 6, 10
- G**
- Gallinarum serovar** 80
- gene**
  - fusions 6, 10
- regulation, *Bordetella* sp. 187–189
- transfer
  - - horizontal (see *there*) 7–9
  - - lateral 7
  - - vector 7
  - - plasmid 7
  - - viral 7
- generation of genetic variants, quality differences** 9–10
- local sequence changes 9
- natural selection 9
- generators of occasional genetic variations** 11
  - DNA acquisition 11
  - DNA rearrangement 11
- genetic variation** 2, 4–10, 12
  - availability of 12
  - keeping the frequencies 11
  - natural strategies for the generation of 4–10
    - - DNA acquisition (see *there*) 7–10
    - - DNA rearrangement (see *there*) 4–7
    - - generation of genetic variants, quality differences (see *there*) 9–10
    - - local sequence changes (see *there*) 4
- genome/genomic** 20–27
  - *afa-8* genome 20
  - *Bordetella* evolution, genomic aspects 184–187
  - bacterial genomes of gene serving 11
  - comparison of whole genome contents 164
  - comparative genome analysis 171
  - degeneration 171
  - EHEC strain EDL933 20, 22, 26, 27
  - enterobacteria genome plasticity 157–171
  - evolution, *P. aeruginosa* 205
    - - clone K 205
    - - integrase 205
    - - natural populations 205
    - - temperate bacteriophages 205
  - intraclonal genome diversity 204
  - islands 161
    - K-12 genome, *E. coli* 17, 22, 73, 145
  - PAO1 genome, *P. aeruginosa* 202
  - prokaryotic genomes, evolution of 1–13
  - pseudogenes 169
  - size reduction 169
    - *Typhimurium* serovar 69–71
    - UPEC strain CFT073 24–26
  - glucotransferase 54
  - gram-positive pathogens 151–152
- gtr** 52
  - *gtrA* 52
  - *gtrB* 52

**H**

- Haemophilus influenzae* 138  
 hemolysin (*hly*) 16–19, 25, 27  
 horizontal gene transfer 8–9, 163  
 – bacterial conjugation (*see there*) 8  
 – DNA transformation (*see there*) 8  
 – virus-mediated transduction (*see there*) 8–9  
 host adaption 76–82  
*HPI* (high pathogenicity island) 18, 24, 42, 90, 94–97, 100–104  
 – *E. coli* 42  
 – *Yersinia* 18, 24  
*hrp* (hypersensitive response and pathogenicity) region 207  
 – *hrp/hrc* cluster 207–208  
*Hrp PAI* of *P. syringae* 207  
 phylogenetic relationships and virulence evolution in *Bordetella* 177–194

**I**

- infectious diseases, emerging 13  
 interleukin (IL)  
 – IL-1 $\beta$  50  
 – IL-18 50  
 intraclonal genome diversity 204  
 invasion plasmid antigens (*see ipa*) 50  
*ipa* (invasion plasmid antigens) 50  
*IpAB* 50  
 iron 97, 99  
*IS1* 59  
*IS1358* and the origins of the Ogawa form of the *Vibrio cholerae* O1-antigen 129  
*IS2* 59  
*IS256* 152  
*IS3* 59  
*IS600* 53, 59  
*IS629* 55, 59

**K**

- K* island 27  
*K-1* capsule (*kps*) 16, 20, 24, 139  
 – blood isolate 16  
 – meningitis 16, 166  
 – UPEC CFT073 16  
*K-12* 17, 22, 27, 67, 73  
 – *E. coli* 17, 22, 73, 145  
 – genome 22, 27  
 – (*Kdo*)<sub>2</sub>-lipid A synthesis genes in 123  
 – LPS core synthesis genes in 124  
 (*Kdo*)<sub>2</sub>-lipid A synthesis genes in *K-12* and LT2 123  
*Klebsiella pneumoniae* 102, 119, 141–144, 170

**L**

- LDC* (lysine decarboxylase) 44, 61  
*LEE* (locus of enterocyte effacement) 23–24, 27, 33–37, 51  
 – EHEC 38–40  
 – EPEC 36–38  
 – regulation 36–37  
 – – LER 37  
 – – PER (plasmid-encoded regulator) 36  
 – structure 33–36  
 – type III secretion system 35  
*lipid A/core* genes 110, 122–125  
 – of *E. coli* and *S. enterica* 122  
 – function 125  
 – G + C content of 125  
 – (*Kdo*)<sub>2</sub>-lipid A synthesis genes in *K-12* and LT2 123  
 – location 125  
 – LPS core synthesis in *K-12* 124  
 – synthesis of lipid A/core precursors 123  
 – variation in LPS core synthesis genes 111, 125  
*LPS* (lipopolysaccharide) core synthesis genes 109–130  
 – biosynthesis and genetics of 114  
 – *ddhABCD* 111  
 – in *E. coli* and *S. enterica* 126–129  
 – IS1358 and the origins of the Ogawa form of the *Vibrio cholerae* O1-antigen 129  
 – in *K-12* 124  
 – *manB* 116  
 – *manABC* 111  
 – *manc* 116  
 – O-antigen genes in *Yersinia spp.* 127  
 – variation in 125  
 – of *Yersinia enterocolitica* 127  
 – of *Yersinia pestis* 129  
 – of *Yersinia pseudotuberculosis* 129  
 living conditions 3  
 local sequence changes in genetic variation 4, 9  
 – chemical  
 – – instability 4  
 – – mutagens 4  
 – improvement of available biological functions 9  
 – radiation mutagens 4  
 – repair systems 4  
 – replication  
 – – infidelities 4  
 – – slippage 4  
 – tautomeric forms 4  
 locus of enterocyte effacement (*see LEE*) 23–24, 27, 33–37, 51  
*LPS-specific loci*, genomic organization 109–130

- biosynthesis and genetics 111–114
- nomenclature
- - *E. coli* 110
- - LPS biosynthetic genes 111
- - *Salmonella* 110
- structure 111
- LT2, (*Kdo*)<sub>2</sub>-lipid A synthesis genes in 123
- lysine decarboxylase (LDC) 44
  
- M**
- mannose-resistant hemagglutination (MRHA) 16
- meningitis
- *E. coli*, meningitis-causing 16, 20, 166
- strain C5 20
- strain EV36 20
- Mesorhizobium loti* 25
- MG1655 160
- mgtBC* 75
- minimal gene set 168
- mobile genetic elements, transposition of (*see* transposition) 5–6
- molecular evolution 2, 10–13
- bacterial genomes of gene serving 11
- emerging infectious diseases 13
- fine-tuned evolution genes 12
- molecular clock, evolutionary distances 9
- pathogenicity 13
- updated view 10–13
- - evolution genes to second-order selection 11–12
- - evolutionary tree (*see there*) 10–11
- - theory of molecular evolution (*see there*) 12–13
- MRHA (mannose-resistant hemagglutination) 16
- murine toxine 93, 94
- mutation, definitions and the effects of 3–4
- selective advantages 3
- selective disadvantages 3
- mxl* 50
  
- N**
- natural selection 2, 6, 9, 12
- Neisseria*
- *N. gonorrhoeae* 148
- *N. meningitidis* 138, 146–148
- - capsule gene clusters 146–148
- - serogroup B capsules 138
- nongenetic elements 12
- environmental mutagens 12
- random encounter 12
- structural flexibility 12
- novel ORFs 60
  
- O**
- O-antigen genes 109, 114–115, 118–119
- chain length 111, 113
- *E. coli* and *S. enterica* 114
- forms and structure 110–111, 122
- gene cluster 114–116, 120
- - G + C content 122
- - *manB* 116
- - *manABC* 111
- - *manC* 116
- - PA1 121
- interspecies movement 119
- intraspecies movement 118
- locus, origins of the polymorphism in 120
- - new forms 120
- - transfer 120
- - transposable elements 120–121
- variation 111–115
- Wzz protein 113
- in *Yersinia spp.* 127
- O-island 24, 27, 39–40
- OM (outer membrane) 110–111
- operation fusions 6, 10
- ORFs
- k-12 specific 165
- novel 60
- in region E 147
- oriC* 204
  
- P**
- P pilus/P pili 16, 18, 21
- P4 100–103
- PAI (pathogenicity island) 16–17, 24, 28–27, 37–38, 41, 43, 97, 109, 160
- *B. bronchiseptica* 186
- characteristics 17
- deletion 18–19, 23
- description of 17
- discovery 16–17
- duplication 19–20
- EspC PAI 37–38
- extraintestinal *E. coli* 165
- hypervariable loci 26
- name 18
- nomenclature 18, 27–28
- phage and PAI evolution 24–25
- *she* PAI 43
- of *Shigella* 49–62
- Tia-PAI 41
- in uropathogenic *E. coli* 18–19
- PAO1 genome, *P. aeruginosa* 202
- PAO1 islands 202
- pap* 19
- pathogenicity island (*see* PAI)
- PER (plasmid-encoded regulator) 36

- pesticin 94, 95  
 phage and PAI evolution 24–25  
 – *sopE* 25  
 phages 209  
 – transduction, phage-mediated 7  
*pheV* gene 27, 146  
*phnAB* 204  
*Photorhabdus* 170  
 – luminescens 157  
*pic* 54–56  
*pilA* 204  
 type 1 pilus 21, 27  
 plasminogen activator 94, 95  
 prokaryotic genomes, evolution of 1–13  
*Pseudomonas spp.*, PAI and PAI-like structures 201–210  
 – catabolic versatility of *Pseudomonas* 208  
 – cluster 208  
 – conserved effector locus (CEL) 208  
 – exchangable effector locus (EEL) 208  
 – islands of catabolic versatility in *Pseudomonas* 208  
 – Hrp PAI of 207  
 – origin of PAI-like structures 205  
 – *P. aeruginosa* 201–202  
 – – clone C 204  
 – – exoenzyme S regulon 202  
 – – genomic evolution (*see there*) 205  
 – – islands of clone- or strain-specific DNA 204  
 – – PAI-like structure 202  
 – – PAO1 genome 202  
 – *P. putida* 201  
 – *P. syringae* 201  
 – – Hrp PAI of 207  
 pyelonephritis, *E. coli* 20  
*pYV* 91, 95, 103
- R**  
*rdgB* 27  
 rDNA, 16S 180  
 recombination  
 – general 7  
 – – sequence homologies 7  
 – – SOS induction 7  
 – homologous 7  
 – repair 7  
 repair 7  
 restriction/modification systems 9, 11  
*rrnA* 204  
*rrnC* 204
- S**  
 16S rDNA 180, 183  
 16S rRNA 184
- Salmonella* 17, 67–83, 102  
 – aggregation, *Salmonella*-induced 69  
 – comparision with *E. coli* K-12 68  
 – DNA, *Salmonella*-specific 81  
 – during infection 69  
 – nomenclature 110  
 – pathogenicity island 24–25, 27, 68  
 – – islands 1 through 5 68–71  
 – *S. bongori* 76  
 – *S. enterica* 76, 78, 110, 126–129, 157  
 – – lipid A/core genes of 122  
 – – LPS core synthesis genes 126–129  
 – – O-antigen genes 113–114  
 – – O30/*E. coli* O157 113  
 – – O35/*E. coli* O111 113  
 – – O50/*E. coli* O55 113  
 – – virulence determinants 158  
 – *S. serovars* 67–83  
 – – horizontal gene transfer and host adaption 76–82  
 – – horizontally acquired virulence determinants of 68–76  
 scarlet-like fever, far-east 95  
 second-order selection, evolution genes to 11–12  
*selC* locus 24, 35, 39 pp, 58  
 sepsis  
 – sepsis strain AL862 19–20  
 – sepsis-causing *E. coli* 16, 19  
 sequence homologies 7  
 serine protease 56  
 serovar  
 – *Choleraesuis* 79  
 – *Gallinarum* 80  
 – *Salmonella* (*see Salmonella* serovar) 67–83  
 – *Typhimurium* (*see Typhimurium* serovar) 68–82  
*set1A* 54–55  
*set1B* 54  
*Sf II* 54  
*Sf X* 54  
*sf104* 55  
*sfiA* 55  
*ShET1* 57  
 Shiga toxin 167  
*Shigella* (*S. spp.*) 24–25, 50, 157  
 – and dysentery 50  
 – black holes 61  
 – infections 57  
 – serotype  
 – – 2a serotype 59  
 – – 5a serotype 59  
 – – Y serotype 52  
 – *S. island* 24  
 – *S. pathogenicity island (SAI)* 44, 49–62

- - PAI-2 44
- Shi-O 52–54, 61
- - gene products 53–54
- organization 53–54
- Shi-1 51, 54–57, 61
- - gene products 55
- - identification 55
- - organization and characterization 55
- Shi-2 51, 57–61
- - aerobactin operon 59–60
- - gene products 59–60
- - identification 58
- - organization 58–59
- virulence determinants 158
- vs. *E. coli* 51
- shigellosis 50
- siderophore 97, 99
- sigA* (Shigella IgA-like protease homolog) 54–56
- site-specific recombination, DNA rearrangement 6–7
- consensus DNA sequences 6
- flip-flop systems 6
- slrP* 83
- sopB* 70
- sopE1* 70
- sopE2* 70
- SOS induction 7
- spa* 50
- SPATE 56
- SPI-1, SPI-2, SPI-3, SPI-4 and SPI-5 69
- spvB* gene 75
- sspH1* and *sspH2* 81, 83
- Staphylococcus aureus* 151
- Streptococcus*
  - *S. agalactiae* 151
  - *S. pneumoniae* 138, 148–151
  - - capsule diversity 148–149
  - - capsule gene clusters 149–150
  - - transposition and recombination 150–151
- Streptomyces coelicolor* 144
- symbiosis island 25
  
- T**
- theory of molecular evolution 12–13
  - genetic variations 12
  - natural selection 12
- Tn4371 208
- transduction
  - general 9
  - phage-mediated 7
  - specialized 9
  - virus-mediated (*see there*) 8–9
- transposition of mobile genetic elements, DNA rearrangement 5–6
  
- insertion target 5
- IS (inserted sequence)
  - - dimers 6
  - - elements 5
  - rates of transposition 6
- tRNA (transfer RNAs) 17–20, 23–26, 68, 163, 205
- integrases 26
- type III system 35, 51
  - exoenzyme S (ExoS) regulation 202–204
  - secretion system 35, 91, 92, 202
- Typhimurium* serovar, horizontally acquired virulence determinants of 68–82
- chromosome 69
- DT204 82
- epidemiology studies 82
- genome 69–71
  - - LT-2 72
- horizontal gene transfer and host adaptation 76–82
- horizontally acquired virulence determinants 68–76
- host range 78–81
- pathogenicity islands 68
- pathogenicity islet, phage, and plasmid-encoded virulence determinants 73
- potential candidates 71–73
- *S. entericar* serovar *Typhimurium* 71
  
- U**
- UPD-GlcDH 151
- UPEC (*uropathogenic Escherichia coli*) 16–28, 162
  - K1 capsule (*kps*), UPEC CFT073 16
  - strain CFT073 19, 24–26
  - strain J96 18–19
- uropathogenic *E. coli* (*see UPEC*) 16–28, 162
  
- V**
- variation generator 6, 11
- Vibrio cholerae* O1-antigen 121, 129
- IS1358 and the origins of the Ogawa form of 129
- virulence evolution in *Bordetella* 177–194
- virulence
  - genetic determinants 50–51
  - plasmid 51
- virus-mediated transduction 8–9
  - bacteriophages 8
  - general transduction 9
  - specialized transduction 9
  
- W**
- Wzz* protein 113

**X**

xenobiotics, degradation 209

**Y**

*Yersinia spp.* 157–158

– high pathogenicity island (HPI) 18, 24

– LPS in *Yersinia, E. coli/S. enterica* 130

– O-antigen genes in *Yersinia spp.* 127

– virulence determinants 158

– *Y. enterocolitica* 127, 157, 158

– *Y. pestis* 127, 129, 157, 158

– *Y. pseudotuberculosis* 120, 127, 129, 157, 158

*Yersiniabactin* 97–100

yop 91–93

## Subject Index for Part II

### A

*Acanthamoeba* 64, 69

actin

- reorganisation, *Hp* strains carrying the *cag*-PAI 9–10

- tail 110, 112

activator

- PrfA 112, 113, 115, 117–119

- transcriptional 112

adaptation, metabolic 121

adaptive gene pool reservoir and cassette genesis 178–180

adenine-specific methyltransferases 50

adherence properties of *Hp* 16–17

- *omp* proteins 16

adhesin 132

- hemolysin/adhesin genes 132

aerosols, *L. pneumophila* 65

aggregation substance 85–87

*Agrobacterium* 128, 140

- *A. radiobacter* 140

AIDS 155, 159–160

*Ajellomyces* 151, 161

aminoglycoside resistance 177, 180

aminotransferase 50

ancestor, common 116–117

anamorph 151

anthrax 96

- toxin subunits 96

antibiotic

- R-plasmids 170

- resistance 79, 85–89, 169–171, 178

- - aminoglycoside 178, 180

- - bleomycin 84

- - chloramphenicol 169, 171, 178

- - conjugative transposons 170–171

- - development 168–170

- - disseminations 170–171

- - erythromycin 84, 89

- - gentamicin 89

- - glycopeptides 89

- - kanamycin-tobramycin 84, 88

- -  $\beta$ -lactams 85, 180

- - macrolide-lincosamide-streptogramin B 88

- - methicillin 84

- - penicillin 89

- - penicillinase 169

- - phosphinotricin 178

- - spectinomycin 89

- - streptomycin 84, 169, 171

- - streptothrinicin 178

- - sulfamethoxazole 171

- - sulfonamides 169

- - tetracycline 169, 171

- - tetracycline-minocycline 88–89

- - trimethoprim 171

- - vancomycin 179

- - vancomycin-teicoplanin 170

*aprV2* genes 48

*aprV5* genes 48

aquatic

- environments 65

- habitats 61, 65

*AraC* family 130, 141

ascospore 151

ascus 151

*askA* gene 55

asomycete 151, 157–160, 162

asomycota 149–151, 160

aspergillosis 158

*Aspergillus* 157–159

ATP-dependent helicases 50

*att1* 173, 174, 177, 182

*attC* 173, 174, 176, 179, 182

auxin 140

*avPpiB* 132

*avr* gene/proteins 130–138

- *avrBs2* 135

- *avrBs3* 135

- *avrBs3/pthA* family 138

- *avrE* 133

- *avrPphA* 137

- *avrPphC* 137

- *avrPphF* 137

- *avrPpiA1* 137

**avr gene/proteins**

- *avrPto* 133
- *avrRxxv* 135
- *P. syringae* *pv. phaseolicola* 137
- avrRxxv/yopJ* homolog 132

**B***Bacillaceae* 95

- Bacillus spp.* 95–105, 120–121
  - *B. anthracis* 96, 99–100, 120–121
  - *B. cereus* 84–85, 95–105, 120–121
  - - - *evolution of* 103–105
  - - - *genome structure* 100–101
  - - - *PAIs (see there)* 101–103
  - - - *strains* 105
  - - - *virulence factors* 97–100
  - *B. subtilis* 96, 120–121
  - *B. thuringiensis* 96, 99
  - *B. thuringiensis israelensis* 101

**bacteria/bacterial**

- evolution 182
- facultative intracellular 110, 118, 121
- plant pathogens, PAIs and virulence plasmids 127–142
  - - *hrp* PAIs (*see there*) 129–135

**bacteriocin** 84–86**bacteriophage(s)** 49–51, 79–85, 88–89, 116, 173, 183

- lysogenic conversion to virulence in gram-positive pathogens 80–84

*Bacteroides* 171

- basidomycete 159, 162
- basidomycota 149–150, 160

*bcrl* 104

- β-exotoxin 98
- β-hemolysin 83–84

**β-lactams** 85, 180**biopesticide** 96**blastoconidia** 151–152*Blastomyces dermatitidis* 150–151, 157, 160, 161*bprV* genes 48*Bradyrhizobium japonicum* 130*Brenneria* 128*Brucella suis* 135*Burkholderia spp.* 128–130

- *B. cepacia* 128

*B. pseudomallei* 130**C****C5a peptidase** 82*cagA* 9, 11–12

- structural variation 11–12

*cag-PAI*

- *Hp* (*see also Hp*) 3, 6–13

*- - phenotypes of *Hp* strains carrying the cag-PAI (*see there*)* 9

- - role of *cag-PAI* in colonisation 13

*- - structural variation*

- - - of the *cagA* 11–12

- - - of the *cag-PAI* 11

- - type IV secretion system 8–9

*- hrp* 140**cancer** 150, 155*Candida spp.* 154–157

- *C. albicans* 155–157

- medical relevance 155

*Candidatus phytoplasma* 128**candidiasis** 155, 158**capsule** 98

- formation 100

**cassette genesis** 138, 174–175, 178–180**catokinin precursor** 140, 141**cell**

- anchor cell wall 114, 117, 118, 120

- death, programmed 129

**chloramphenicol resistance** 169, 171, 178**cholera**

- toxin 181

- *V. cholerae* super-integron 175–177

**chromista** 149–150**chromosome/chromosomal**

- R chromosomes 153, 156

- rearrangements within the *cps* locus, meningococcal capsule 36

- sex chromosomes 152

- translocation 156

**class II transposons** 101*Clavibacter* 128

- *C. michiganensis* 128

*Clostridium botulinum* 80*Coccidioides immitis* 150, 157–159**codon usage** 115**commensal** 117**conjugation/conjugative** 85–89

- *L. pneumophila* 66, 68–69, 73

**- plasmids**

- - encoding virulence traits 85–87

- - extrachromosomally 96

- transposons 87–89

**coronatine** 138*Corynebacterium diphtheriae* 80–81**corynephage-β** 80–82*Coxiella burnetii* 68*cps* 35–36**crown gall** 128, 140*cry* 97

- toxins and proteins 97

*crylA* genes 101*crylC* genes 102*crylD* genes 102

- cryotolerance 109–110
- cryptococcosis 158–159
- Cryptococcus* 150, 157–159
- CS 176
- CTX  $\Phi$  173, 181, 183
- CUG codon 155
- cya* 103
- Cyt 97
- cytolysin 120
  - listeriolysin O 111
- D**
- DEAH family 50
- deletion, spontaneous 118
- dermatophytes 157, 160, 161
- deuteromycetes 155
- deuteromycota 151
- Dichelobacter nodosus* 47–57
  - bacteriophage (*see there*) 49–51
  - disease 47
  - elastase 49
  - genetic manipulation 56
  - genomic islands 48–49
    - – *aprV2* genes 48
    - – *aprV5* genes 48
    - – *bprV* genes 48
  - pathogenesis 47
  - proteases 48
  - – thermostability 49
  - *ssrA* 51–52, 56
  - rRNA operons 47
  - tRNA genes 51
  - type IV fimbriae 48
  - *vap* locus (*see there*) 48, 50, 52–56
  - *vrl* locus/region (*see there*) 49–52
- dinucleotide, relative abundance 119
- diphtheria toxin 80–81
- DNA
  - array 157
  - mitochondrial 152
  - r-DNA 152, 153, 156, 163
  - repetitive 152, 154, 163
  - ribosomal 156
  - t-DNA 140
  - topoisomerase 103
  - uptake signal 25–26, 30
- dot*
  - *icm/dot* (*see there*) 65–70
  - secretion system 68
- dspEF* locus 132
- E**
- E. faecalis* 86–89
- E. faecium* 88
- EdTx 98
- effector
  - locus 133
  - proteins 131
- elastase, *D. nodosus* 49
- emetic toxin 96
- eml* 66
- emm* 82
- EndA 82
- endosomal/lysosomal markers 67
- Enterobacteriaceae* 128, 131
- enterococcal cytolysin 86–88
- enterotoxin
  - enterotoxin A 80, 83
  - enterotoxin E 83
  - non-hemolytic enterotoxin (NHE) 99
- enzyme electrophoresis, multilocus 25, 29
- Epidermophyton* 160–161
- epithelial cells 67
- Erwinia* 128–132
  - *E. amylovora* 130–131
  - *E. chrysanthemi* 131–132
  - *E. herbicola* 132, 140
  - *E. stewartii* 132
- erythrogenic
  - exotoxin A 80–83
  - exotoxin C 80–83
- erythromycin resistance 84, 89
- Escherichia coli* 71
- Esp 87
- EspC 172
- EST (expressed sequence tags) 162
- eucosyltransferase genes, slipped strand mispairing regulates expression 14–15
- exfoliative toxin A 80, 83
- $\beta$ -exotoxin 98
- F**
- fas* loci 141
- fire blight 128, 132
- fitness 121
- FK-506-binding proteins 65
- flaA* 66, 71–72
  - alternative sigma-28 factor 71
- fluconazole 155–156
- food
  - contamination 109, 120
  - food-borne infections 96
- frgA* gene 71–72
- fungi
  - *F. imperfecti* 151
  - genome structure of pathogenic fungi 149–163
    - – *Aspergillus* 157–159
    - – *Candida* species 154–157
    - – *Cryptococcus* 150, 157–159
    - – fungal plant pathogens 160–163
    - – obligate pathogenic fungi 160

## fungi

- - *Pneumocystis carinii* 150, 151, 157–158, 160
- - *Saccharomyces cerevisiae* (*see there*) 151–154

fur 71

## G

## G + C content

- *hrp* 132, 134, 137
- *Legionella spp.* 62, 68–69, 73
- *Listeria* 115, 119

## gene

- cassette 138, 174–175
- density 158–159
- anactivation 155
- gene-for-gene interactions 130
- loss 154
- order 157
- pool 25
- - pool reservoir 178–180
- transfer 6–8, 154
- - horizontal 154
- - - in *Listeria* 115, 117, 119, 121
- - in *Hp*, natural transformation competence 6–8
- - in *Neisseria*, horizontal gene transfer 24–25, 28–29

## genetic

- diversity in *Hp* 3–6
- - macro-diversity 6
- - micro-diversity 4–6
- - slipped strand mispairing 14–15
- manipulation 152
- methods, *D. nodosus* 48
- recombination in *L. spp.* 63
- wealth of SI cassette reservoirs 178–179

## genome/genomic

- database 152
- functional 157
- islands of *Dichelobacter nodosus* (*see there*) 47–57
- mitochondrial 153, 160
- renewal 154
- gentamicin resistance 89
- gerX* 103
- glpA* gene 55–56
- glutamine 50
- glycopeptides resistance 89
- glycosidase 50
- gonococci and *N. lactamica* 29–30
- gram-positive pathogens, phages and other mobile virulence elements 79–89

## H

- haemolysin BL (HBL) 99
- harpin genes 132
  - *hrpN* 132
  - *hrpW* 132
- Hartmannella vermiformis* 64, 67
- HBL (haemolysin BL) 99
- heat-stable
  - enterotoxin 181
  - toxine gene 176
- helicases, ATP-dependent 50
- Helicobacter pylori* (*see Hp*) 1–17, 140
- hemagglutinin, mannose-fucose-resistant (*mrhA*) 176, 181
- hemolysin 132
  - $\beta$ -hemolysin 83–84
  - hemolysin/adhesin genes 132
- Histoplasma capsulatum* 150, 157, 160–161
- HIV 150, 155
- HL-60 derived macrophages 67
- hlb* 84
- homoplasy 27
- hop proteins 131
- hopPsyA* 133, 137
- horizontal gene transfer 82–83, 89
  - in *L. pneumophila* 72
  - in *Neisseria* 24–25, 28–29
- host adaption and *Hp* 14–17
  - LPS structure 14–16
  - slipped strand mispairing regulates expression of eucosyltransferase genes 14–15
- hot water systems, *L. pneumophila* 61–62, 73
- housekeeping genes 100
- Hp* (*Helicobacter pylori*) 1–17, 140
  - adherence properties 16–17
  - genetic diversity 3–6
    - - macro-diversity in *Hp* 6
    - - micro-diversity in *Hp* 4–6
    - - mutator species 4–5
    - - - panmictic species 5–6
    - and host adaption 14–17
    - - LPS structure, variation of (*see there*) 14–15
    - - variation in adherence properties of *Hp* 14–16
  - interaction of *Hp* with professional phagocytes 10–11
  - natural transformation competence and gene transfer in *Hp* 6–8
  - phenotypes of *Hp* strains carrying the *cag-PAI* (*see there*) 9
  - signal transduction pathways induced by *Hp* 10
- hpa1* 135
- hpa2* 135

- hph<sup>a</sup>* gene 72
- HPI 172
- hrc* 130
  - *hrp/hrc* cluster 130
- hrp* PAIs 129–135
  - *avr* genes (*see there*) 132–138
  - avirulence gene PAIs 136–138
  - effectors 132
  - elicitors 132
  - G + C content 132, 134, 137
  - genes for synthesis of plant hormones 140–141
  - genome sequencing 132, 134, 141
  - *hrp/hrc* cluster 130
  - - of *R. solanacearum* 134–135
  - - of *X. spp.* 134–135
  - *hrpF* genes 135
  - *hrpN* genes 132
  - *hrpW* genes 132
  - hypersensitive response (HR) 129
  - phytotoxin gene clusters 138–139
  - plant pathogenic enterobacteria 131–133
  - *P. syringae* *avr* genes 137–138
  - *R. fascians* 141
  - type III protein delivery system 129–131
  - virulence plasmids 129
- HrpL 130
- hsvG* 132
- htpB* 71
- hyaluronidase 81, 83
- hybrid, interspecific 154
- hylP* 83
- hylP2* 83
- hypersensitive response (HR) 129
  
- I**
- IAA 140
- icm/dot* (intracellular multiplication/defect in organelle trafficking) 65–70
  - conjugation 66
  - type IV secretion system 67, 68–70
  - type IVB secretion system 66
- ICS (inverse core-site) 174, 176
- IgA protease, *Neisseria* 32–33
  - *Inl* AB 113
  - *Inl* loci 113
- in vivo expression technology 156
- insecticidal crystal toxins 96
- insertion sequences (*see SI*) 101, 170, 177–183
- int1* (integrase) 173, 174, 182
- intA*, *intB* and *intC* element, *vap* locus 52, 54–55
- integrase(s) 82, 84, 88
  - *hrp* 131, 132
- integrons 170–184
  - can integrons be considered pathogenicity islands? 180–181
  - model for evolution 183–184
  - and resistance gene flow 173–178
    - - an ancient evolutionary apparatus that is prominent among proteobacterial genomes 177–178
    - - *V. cholerae* super-integron 175–177
  - Tn7 175
  - Tn21 175
  - and transposons on the evolution of resistance and virulence 167–184
    - - impact transposons in resistance and virulence gene flow (*see also* transposons) 170–173
  - and virulence gene flow 180–182
- internalin islets and islands, *Listeria* 113–115
- secreted 114
- intracellular multiplication/defect in organelle trafficking (*see icm/dot*) 65–70
- introns 158, 160
- inversion 36
- in vivo expression technology 156
- i-sequence IS1301, meningococcal capsule 37–38
- iraA* gene 72
- iraAB<sup>a</sup>* gene 72
- iraB* gene 72
- IS elements 82, 84, 87, 89
  - IS4 101
  - IS6 101
  - IS21 101
  - IS231 101
  - IS232 101
  - IS240 101
  - IS253 54
  - IS256 89
  - IS257 89
  - IS1216 89
- ispGH* mutant strain 67
  
- K**
- kanamycin-tobramycin resistance 84, 88
- karyotype 153, 156, 159
- Klebsiella spp.* 180
  
- L**
- L. monocytogenes* 88
- β-lactams 85, 180
- leafy gall 128, 141
- LEE 172
- lef* 103
- Legionella spp.* 61–73
  - *Acanthamoeba* 64, 69
  - aerosols 65
  - conjugation 66, 68–69, 73

- Legionella spp.*
- dormancy state 64
  - G + C content 62, 68, 69
  - genetic recombination 63, 73
  - genome 62
  - genomic islands 69
  - guinea pigs 65
  - *Hartmannella* 64, 67
  - horizontal gene transfer 72, 73
  - hot water systems 61–62, 73
  - invasion into human monocytes 64
  - *ligA* gene 64, 66, 67, 71–72
  - *L. dumoffii* 69
  - *L. gormanii* 69
  - *L. longbeachae* 69
  - *L. micdadei* 69
  - *L. pneumophila* 61–69
  - *lsp* gene (*see there*) 64, 66, 67
  - macrophages 67
  - mice 64
  - mobile genetic elements 69
  - *Naegleria* 64
  - opportunistic pathogen 61
  - PAIs (pathogenicity islands) 68–71
  - person-to-person spread 65, 73
  - phase variation 69
  - protozoa 64
  - *rpoS* gene 64, 66, 67
  - specific genes 71–72
  - survival and dual host system 64–65
  - taxonomy and evolution 62–64
  - transcriptional regulator 69
  - transmission 65
  - type IV secretion systems 68–69
  - viable but nonculturable (VBNC) 64
  - virulence and fitness factors (*see there*) 65–68
- Legionella*-like amoebal pathogens (LLAPs) 62–63
- legionnaires' disease 61, 73
- LeTx 98
- leucinerich repeats (LRR) 113, 118, 120, 134
- leukocidin 80, 83
- ligA* gene 64, 66, 67, 71–72
- linear plasmids 100
- lipase 83
- LIPI-1 (PrfA-dependent virulence gene cluster) 111–113
- LIPI-2 111
- Listeria spp.*, PAIs 111–122
- evolution of virulence in *L.* 115–121
  - G + C content 115, 119
  - internalin islets and islands 113–115
  - *L. ivanovii* 111
  - *L. monocytogenes* 111
  - *L. seeligeri* 113
- LIPI-1 (PrfA-dependent virulence gene cluster) 111–113
  - perspectives 121
  - listeriolysin O 111
  - listeriosis 110
  - LLAPs (*Legionella*-like amoebal pathogens) 62–63
  - loop, autoregulatory 112
  - LPS
    - gene locus 72
    - structure, variation of, *Hp* 14–15
    - – *Hp* and molecular mimicry 14
    - – slipped strand mispairing regulates expression of fucosyltransferase genes 14–15  - lsp* gene 64, 66, 67
  - type II protein secretion system 66
  - LTR 154, 157
  - lucA* gene 71
  - lucC* gene 71
  - lvh* operon 67, 68, 70
  - type IV secretion system 67, 68, 70
  - lysogeny 80
- M**
- M protein 82
- macrolide-lincosamide-streptogramin B resistance 88
- macrophage
- HL-60 derived 67
  - infectivity potentiator (*see mip*) gene 62–63, 65–66, 71
  - macrophage-specific infectivity loci (*see mil*) 65–67
- Magnaporthe grisea* 162
- major repeat sequence (MRS) 156
- mannose-fucose-resistant hemagglutinin (*mrhA*) 176, 181
- mating 154, 156, 162
- megaplasmid 134
- meningococci/meningococcal
- capsule, *Neisseria* 35–38
  - – chromosomal rearrangements within the *cps* locus 36
  - – inversion sequence IS1301 37–38
  - clonal groupings and genetic isolation 27–29
  - population structure 27
- metabolic adaption 121
- methicillin resistance 84
- methyltransferase, adenin-specific 50
- mga* 82
- microsatellites 156
- Microsporum* 160–161
- mil* (macrophage-specific infectivity loci) 65–67

- *milA* 66, 67
  - *mil* loci 66
  - *mil* mutant strains 67
  - mip* (macrophage infectivity potentiator) gene 62–63, 65–66, 71
    - homodimeric *mip* protein 65
    - phylogenetic dendrogram 63
  - mobile genetic elements
    - *D. nodosus* 52, 57
    - *L. pneumophila* 69
  - molecular mimicry and *Hp* 14
    - LPS structure (*see there*) 14–15
    - eucosyltransferase genes 14–15
  - monocytes, invasion to human monocytes, *L. pneumophila* 64
  - mosaic gene 25
  - motility, intracellular 112
  - mrhA* (mannose-fucose-resistant hemagglutinin) 176, 181
  - MRIs 174–177, 179, 182–183
  - multilocus
    - enzyme electrophoresis 25, 29
    - sequence typing 25, 27, 30
  - mutation, pathoadaptive 117
  - mutator species, *Hp* 4–5
  - Mycobacterium* spp. 180
  - mycotoxin 158
- N**
- N-acetyl-D-mannosamine-phosphate 35
  - Naegleria* 64
  - necI* 139
  - Nectria haematococca* 162
  - Neisseria* spp. 23–38
    - evolution of pathogenicity 31–38
    - - IgA protease 32–33
    - - meningococcal capsule 35–38
    - - *opc* genes (*see there*) 31–32
    - - porins 33–35
    - horizontal gene transfer 24–25, 28–29
    - *N. lactamica* 29–30
    - population structure 26–30
      - - clonal groupings of meningococci and genetic isolation 27–29
      - - gonococci and *N. lactamica* 29–30
      - - meningococci 27
  - neurotoxin
    - neurotoxin C1 80
    - neurotoxin D 80
  - NHE (non-hemolytic enterotoxin) 99
  - Nitrosomonas* 177
  - non-hemolytic enterotoxin (NHE) 99
- O**
- omp* proteins 16
    - *ompS* gene 71
- oomycota 149–150, 160
  - opc* gene, *Neisseria* 25, 31–32
    - *opcA* 31–32
  - opine 140
  - opportunistic pathogen, *L. pneumophila* 61
- P**
- pagA* 103
  - PAI (pathogenicity island)
    - *B. anthracis* 103
    - *B. thuringiensis* 101–102
    - of bacterial plant pathogens 127–142
    - *cag*-PAI (*see there*) 3, 6–13
    - integrons and transposons on evolution of resistance and virulence (*see there*) 172
    - *Legionella* spp. 68–71
    - *Listeria* spp. (*see there*) 111–122
    - phage and other mobile virulence elements in gram-positive pathogens 82, 84, 87
  - pAM $\beta$ 1 family 104
  - panmictic species, *Hp* 4–5
  - panmixis 27–28
  - Pantoea* 128
    - *P. agglomerans* 132
    - *P. stewartii* 132
  - Paracoccidioides brasiliensis* 150, 157, 160–161
  - pathogenic fungi, genome structure (*see fungi*) 149–163
  - pathogenicity island (*see PAI*)
  - Pectobacterium* 128
  - pedestal formation, *Hp* strains carrying the *cag*-PAI 9–10
  - penicillin
    - binding proteins 25
    - resistance 89
  - penicillinase resistance 169
  - phage 80–83, 171, 181, 183
    - conversion 80
      - - double 83
      - - negative 83
      - - positive 83
      - - triple 83
    - and other mobile virulence elements in gram-positive pathogens 79–89
  - phagocytosis
    - resistance, interaction of *Hp* with professional phagocytes 10–11
    - triggered 110, 113
  - phaseolotoxin 138
  - phenotypes of *Hp* strains carrying the *cag*-PAI
    - 9
      - actin reorganisation and pedestal formation 9–10
      - chemokine induction 9–10
      - interaction of *Hp* with professional phagocytes 10–11

- phenotypes of *Hp* strains carrying the *cag*-PAI  
 - signal transduction 10  
 pheromone 85  
*pheV* 132  
 phosphinotricin resistance 177  
 phospholipase 112, 120, 132  
 phylogeny 120  
*Phytophthora* 162  
 phytoplasma, *Candidatus* 128  
 phytotoxin 132  
 - gene clusters, *hrp* 138–139  
 Pierce's disease 128  
 pigment 159  
*pil* (pilus)  
 - biogenesis 65  
 - *pilB* 66–67, 71  
 - *pilC* 66–67  
 - *pilD* 66–67  
 plant pathogens, bacterial (*see there*) 127–142  
 plasmid 49–50  
 - killer system 54  
 plasmids 79, 84  
 - conjugative 79  
 - pheromone-responsive, conjugative 85–86, 89  
*PlcR*-regulated genes 99  
 ploidy 154, 162  
*pmi* (protozoan and macrophage infectivity) 65–66  
*Pneumocystis carinii* 150, 151, 157–158, 160  
 pneumophila (*see also Legionella spp.*) 61–73  
*pnpA* gene 56  
*pom* (pneumophila-specific outer membrane protein) 71–72  
 Pontiac fever 61  
*popA* 134  
*popB* 134  
*popC* 134–135  
 porins, *Neisseria* 33–35  
 pPATH 132  
 PPIase 65  
*PrhA* 134  
 promoter 182  
 proteases 48, 55  
 - thermostability 49, 55  
 protective antigen 98  
 proteobacteria 128  
 - γ-subgroup 62  
 proteobacterial genomes 177–178  
 proteome 152, 163  
 - database 152  
 protozoa/protozoan 149–151, 160  
 - bacteriovorous 116  
 - *L. pneumophila* 64  
 - and macrophage infectivity (*pmi*) 65–66  
*Providencia* 180  
*prp* gene 65–66  
 pseudogene 32, 33  
 pseudohyphae 152  
*Pseudomonas* 128, 177  
 - *P. aeruginosa* 128  
 - *P. fluorescens* 130  
 - *P. syringae* 130, 131, 133–134, 136–138, 140  
 - - *P. pv. phaseolicola* 137  
 - - *P. pisi* 137  
 - - *P. syringae* 133–134  
 - - *P. tomato* 133–134  
 - *P. Xanthomonas* 177  
*pthG* 132  
*pX01* 101, 103  
*pX02* 98, 101
- R**
- R chromosomes 153, 156  
*Ralstonia* 128  
 - *R. solanacearum* 130, 131  
 r-DNA 152, 153, 156, 163  
 rearrangement 157  
 recombination 113, 117–118, 154, 157, 163  
 - suppression 162  
 repeat  
 - element 30  
 - - AT-rich elements 30  
 - - Correia elements 30  
 - - RS-elements 30  
 - leucine-rich 113, 118, 120  
 - long-terminal 154  
 reporter gene 156  
 representational difference analysis 25  
 resistance 155  
 - antibiotic (*see there*) 79, 85, 87, 89, 168–171, 178  
 - integrons and resistance gene flow (*see also integrons*) 173–178  
 - and virulence, evolution of 167–184  
 restriction-modification system 26, 28–29  
 retroelement 152, 154, 156  
 retrotransposon 152, 154, 157  
*Rhizobium sp.* NGR234 130  
*Rhodococcus* 128  
 - *R. equi* 84–85  
 - *R. fascians* 141  
 rice blast 162  
 RNA  
 - rRNA (*see there*)  
 - ribosomal 151  
 - tRNA genes (*see there*) 50, 117, 131–132, 137  
 R-plasmids 170  
*rpoS* gene 64, 66, 67  
 rRNA  
 - operons *D. nodosus* 47  
 - 16S rRNA, *L. pneumophila* 62

- S**
- Saccharomyces cerevisiae* 151–154
    - apathogenic vs. pathogenic 154
    - Yeast genome 152–154
  - SaPI1 84, 172
  - SaPI2 84
  - SCCmec 84
  - scpA* 82
  - sequence
    - annotation 152
    - typing, multilocus 25, 27, 30
  - Serratia* 180
  - serum resistance 31, 34
  - sex chromosomes 152
  - Shewanella* 177
  - SI (super integrons) 101, 170, 177–183
    - SI cassette reservoirs, genetic wealth of 178–179
    - using SIs to track bacterial evolution in real time 182
  - sialic acid 35
  - signal transduction pathways induced by *Hp* 10
  - Sinorhizobium fredii* 130
  - size variation 153
  - slipped strand mispairing
    - in *Hp* 14–15
    - regulates expression of eucosyltransferase genes 14–15
    - in *Neisseria* 25, 31
  - smut disease 162
  - sodB* 71
  - sodC* gene 72
  - soil, habitat 109, 121
  - sour skin 128
  - southern bacterial wilt 128
  - speA* and *speC* 82
  - SpeA and SpeC 82
  - spectinomycin resistance 89
  - sphingomyelinase 115, 117, 119
  - SPI-3 172
  - Spiroplasma* 128
    - *Sp. citri* 128
  - split graph analysis 34
  - sporulation 154
  - ssrA* 51–52, 56
  - staphylococcal phages 83
    - 80 $\alpha$  84
    - FRI1106-1 83
    - $\Phi$ 13 84
    - phi-ZM-1 83
  - Staphylococcus aureus* 80, 83–84, 88–89
  - staphylokinase 80, 83
  - Stewart's* wilt 128
  - streptococcal
    - mitogenic factor 82
    - phages 81–82
  - – 370.1 82
  - – 370.2 82
  - – 370.3 82
  - – 370.4 82
  - – CS112 82
  - – T12 82
  - Streptococcus spp.*
    - *S. pneumoniae* 85, 88
    - – capsule production 85
    - *S. pyogenes* 80–83
    - – SF370 81–82
  - streptodornase 82
  - Streptomyces spp.* 128, 139, 180
    - *S. acidiscabies* 139
    - *S. lividans* 139
    - *S. scabies* 139
  - streptomycin resistance 84, 169, 171, 177
  - sulfamethoxazole resistance 171
  - sulfonamides resistance 169
  - superantigen 81–82, 84
  - synthesis of plant hormones, genes for 140–141
  - syringomycin 138
  - syringopeptin 138
- T**
- tabtoxin 138
  - t-DNA (transferred DNA) 140
  - telemorph 151, 154–155, 157, 159
  - tetM* 88
  - tetracycline
    - resistance 169, 171
    - tetracycline-minocycline resistance 88–89
  - Ti plasmid 140
  - tRNA genes, *D. nodosus* 50
  - Tn3 family 102
  - Tn7 175
  - Tn21 175
  - Tn4430 102
  - tox* 80–81
  - toxin co-regulated pilus (TCP) 181
  - tra/trb* genes 68, 140
  - transcriptional
    - activator, *D. nodosus* 50
    - factor, *L. pneumophila* 71
  - transcriptome 163
  - transduction 80, 82, 84, 89
  - transformation 156
  - transmission, *L. pneumophila* 65
  - transplant 150, 155
  - transposable elements 101
  - transposase(s) 82, 87
    - *hrp* gene 131
  - transposition 156, 157, 162
  - transposons 79, 81, 84, 87–89, 156, 170–180
    - classes of 170

- transposons
- conjugative 170–171
  - impact transposons in resistance and virulence gene flow 170–173
  - - adaptive gene pool reservoir and cassette genesis 178–180
  - - transposons and antibiotic resistance dissemination 170–171
  - - transposon encoded functions 172–173
  - - transposons and virulence determinant propagation 172
  - and integrons on the evolution of resistance and virulence (*see also* integrons) 167–184
  - *Listeria* 116
  - Tn552 89
  - Tn554 84
  - Tn916 88–89
  - Tn917 88
  - Tn1545 88
  - Tn1547 89
  - Tn1549 89
  - Tn4001 89
  - Tn5251 88
  - Tn5252 88
  - Tn5253 88
  - Tn5381 89
  - Tn5382 89
  - Tn5384 89
  - Tn5385 89
- Trichophyton* 160, 161
- trimethoprim resistance 171
- tRNA genes
- *D. nodosus* 50
  - *hrp* genes 131–132, 137
  - *Listeria* 117
- TSST-1 84
- tst* 84
- ty 152–154, 157
- type I protein secretion, *hrp* 129
- type II
- protein secretion, *hrp* 129
  - secretion system, *L. pneumophila* 66–67
- type III protein secretion, *hrp* 129
- type IV
- fimbriae, *D. nodosus* 48
  - secretion system
  - - *Hp* 8–9
    - - - chemokine induction 9–10
    - - - phagocytosis resistance 10–11
    - - - signal transduction pathways induced by *Hp* 10
  - - *hrp* 140
  - - *L. pneumophila* 67–70, 73
  - - - *icm/dot* 67, 68
  - - - *lvh* operon 67, 68
- type IVB secretion system, *L. pneumophila* 66
- U
- U937 cells 67
- human macrophagal 67
- Ustilago hordei* 162
- V
- V. cholerae* super-integron 175–177
- vanB* 89
- vanB2* 89
- vancomycin
- resistance 179
  - vancomycin-teicoplanin resistance 170
- vap* locus 48, 50, 52–56
- bacteriophage 54
  - *intA* element 52, 56–57
  - *intB* element 54, 57
  - *intC* element 55–57
  - insertion 55
  - integrase 54–57
  - IS/253 54
  - mobile genetic elements 52, 57
  - plasmid killer system 54
  - plasmids 52–54
  - protease 55, 57
  - recombination 52, 54
  - site-specific recombinase 54
  - virulence 56
- variation
- allelic 156
  - genomic 156
- VCRs 176, 177, 179
- Vibrio* 177, 180
- Vibrionaceae* 177
- Vip
- proteins 97
  - toxins 97
- vir* 137, 140
- virulence
- factor(s) 79–80, 83–86, 89, 97–100, 159
    - - *B. anthracis* 99–100
    - - *B. cereus* 99
    - - *B. thuringiensis* 99
    - - and fitness factors, *L. pneumophila* 65–68
    - - - replication in protozoa and human macrophages 65–67
    - - - replication in protozoa or human macrophages 67–68
    - - species-specific 97–98
    - gene flow 180–182
    - phages and other mobile virulence elements in gram-positive pathogens 79–89
    - plasmids, *hrp* 129
    - and resistance, evolution of 167–184
    - - impact transposons in resistance and virulence gene flow (*see also* transposons) 170–173

- VPI  $\Phi$  173, 181, 183
- vrl* locus/region 49–52, 54, 56–57
  - bacteriophage 49–51
  - genetic methods 52
  - integrase 50–51, 56–57
  - plasmid 49–50
  - recombination 50–51
  - site-specific recombinase 50
  - *trans*-translation 51
  - virulence 47, 49, 52, 56
- W**
- witches' broom 128
- X**
- Xanthomonas* spp. 128, 177, 180
  - *avrBs3/pthA* family 138
  - *hrp/hrc* cluster 134
  - *Pseudomonas/Xanthomonas* 177
  - *X. axonopodis* pv. *citri* 135
- Y**
- Y' element 152
- Yeast genome 152–154
- yellows
- Yersinia* spp. 130
- Z**
- zygomycota 149–150

# Current Topics in Microbiology and Immunology

Volumes published since 1989 (and still available)

Vol. 220: **Rauscher, Frank J. III; Vogt, Peter K. (Eds.):** Chromosomal Translocations and Oncogenic Transcription Factors. 1997. 28 figs. XI, 166 pp. ISBN 3-540-61402-8

Vol. 221: **Kastan, Michael B. (Ed.):** Genetic Instability and Tumorigenesis. 1997. 12 figs. VII, 180 pp. ISBN 3-540-61518-0

Vol. 222: **Olding, Lars B. (Ed.):** Reproductive Immunology. 1997. 17 figs. XII, 219 pp. ISBN 3-540-61888-0

Vol. 223: **Tracy, S.; Chapman, N. M.; Mahy, B. W. J. (Eds.):** The Coxsackie B Viruses. 1997. 37 figs. VIII, 336 pp. ISBN 3-540-62390-6

Vol. 224: **Potter, Michael; Melchers, Fritz (Eds.):** C-Myc in B-Cell Neoplasia. 1997. 94 figs. XII, 291 pp. ISBN 3-540-62892-4

Vol. 225: **Vogt, Peter K.; Mahan, Michael J. (Eds.):** Bacterial Infection: Close Encounters at the Host Pathogen Interface. 1998. 15 figs. IX, 169 pp. ISBN 3-540-63260-3

Vol. 226: **Koprowski, Hilary; Weiner, David B. (Eds.):** DNA Vaccination/Genetic Vaccination. 1998. 31 figs. XVIII, 198 pp. ISBN 3-540-63392-8

Vol. 227: **Vogt, Peter K.; Reed, Steven I. (Eds.):** Cyclin Dependent Kinase (CDK) Inhibitors. 1998. 15 figs. XII, 169 pp. ISBN 3-540-63429-0

Vol. 228: **Pawson, Anthony I. (Ed.):** Protein Modules in Signal Transduction. 1998. 42 figs. IX, 368 pp. ISBN 3-540-63396-0

Vol. 229: **Kelsoe, Garnett; Flajnik, Martin (Eds.):** Somatic Diversification of Immune Responses. 1998. 38 figs. IX, 221 pp. ISBN 3-540-63608-0

Vol. 230: **Kärre, Klas; Colonna, Marco (Eds.):** Specificity, Function, and Development of NK Cells. 1998. 22 figs. IX, 248 pp. ISBN 3-540-63941-1

Vol. 231: **Holzmann, Bernhard; Wagner, Hermann (Eds.):** Leukocyte Integrins in the Immune System and Malignant Disease. 1998. 40 figs. XIII, 189 pp. ISBN 3-540-63609-9

Vol. 232: **Whitton, J. Lindsay (Ed.):** Antigen Presentation. 1998. 11 figs. IX, 244 pp. ISBN 3-540-63813-X

Vol. 233/I: **Tyler, Kenneth L.; Oldstone, Michael B. A. (Eds.):** Reoviruses I. 1998. 29 figs. XVIII, 223 pp. ISBN 3-540-63946-2

Vol. 233/II: **Tyler, Kenneth L.; Oldstone, Michael B. A. (Eds.):** Reoviruses II. 1998. 45 figs. XVI, 187 pp. ISBN 3-540-63947-0

Vol. 234: **Frankel, Arthur E. (Ed.):** Clinical Applications of Immunotoxins. 1999. 16 figs. IX, 122 pp. ISBN 3-540-64097-5

Vol. 235: **Klenk, Hans-Dieter (Ed.):** Marburg and Ebola Viruses. 1999. 34 figs. XI, 225 pp. ISBN 3-540-64729-5

Vol. 236: **Kraehenbuhl, Jean-Pierre; Neutra, Marian R. (Eds.):** Defense of Mucosal Surfaces: Pathogenesis, Immunity and Vaccines. 1999. 30 figs. IX, 296 pp. ISBN 3-540-64730-9

Vol. 237: **Claesson-Welsh, Lena (Ed.):** Vascular Growth Factors and Angiogenesis. 1999. 36 figs. X, 189 pp. ISBN 3-540-64731-7

Vol. 238: **Coffman, Robert L.; Romagnani, Sergio (Eds.):** Redirection of Th1 and Th2 Responses. 1999. 6 figs. IX, 148 pp. ISBN 3-540-65048-2

Vol. 239: **Vogt, Peter K.; Jackson, Andrew O. (Eds.):** Satellites and Defective Viral RNAs. 1999. 39 figs. XVI, 179 pp. ISBN 3-540-65049-0

Vol. 240: **Hammond, John; McGarvey, Peter; Yusibov, Vidadi (Eds.):** Plant Biotechnology. 1999. 12 figs. XII, 196 pp. ISBN 3-540-65104-7

- Vol. 241: **Westblom, Tore U.; Czinn, Steven J.; Nedrud, John G. (Eds.):** Gastro-duodenal Disease and Helicobacter pylori. 1999. 35 figs. XI, 313 pp. ISBN 3-540-65084-9
- Vol. 242: **Hagedorn, Curt H.; Rice, Charles M. (Eds.):** The Hepatitis C Viruses. 2000. 47 figs. IX, 379 pp. ISBN 3-540-65358-9
- Vol. 243: **Famulok, Michael; Winnacker, Ernst-L.; Wong, Chi-Huey (Eds.):** Combinatorial Chemistry in Biology. 1999. 48 figs. IX, 189 pp. ISBN 3-540-65704-5
- Vol. 244: **Daëron, Marc; Vivier, Eric (Eds.):** Immunoreceptor Tyrosine-Based Inhibition Motifs. 1999. 20 figs. VIII, 179 pp. ISBN 3-540-65789-4
- Vol. 245/I: **Justement, Louis B.; Siminovitch, Katherine A. (Eds.):** Signal Transduction and the Coordination of B Lymphocyte Development and Function I. 2000. 22 figs. XVI, 274 pp. ISBN 3-540-66002-X
- Vol. 245/II: **Justement, Louis B.; Siminovitch, Katherine A. (Eds.):** Signal Transduction and the Coordination of B Lymphocyte Development and Function II. 2000. 13 figs. XV, 172 pp. ISBN 3-540-66003-8
- Vol. 246: **Melchers, Fritz; Potter, Michael (Eds.):** Mechanisms of B Cell Neoplasia 1998. 1999. 111 figs. XXIX, 415 pp. ISBN 3-540-65759-2
- Vol. 247: **Wagner, Hermann (Ed.):** Immunobiology of Bacterial CpG-DNA. 2000. 34 figs. IX, 246 pp. ISBN 3-540-66400-9
- Vol. 248: **du Pasquier, Louis; Litman, Gary W. (Eds.):** Origin and Evolution of the Vertebrate Immune System. 2000. 81 figs. IX, 324 pp. ISBN 3-540-66414-9
- Vol. 249: **Jones, Peter A.; Vogt, Peter K. (Eds.):** DNA Methylation and Cancer. 2000. 16 figs. IX, 169 pp. ISBN 3-540-66608-7
- Vol. 250: **Aktories, Klaus; Wilkins, Tracy, D. (Eds.):** Clostridium difficile. 2000. 20 figs. IX, 143 pp. ISBN 3-540-67291-5
- Vol. 251: **Melchers, Fritz (Ed.):** Lymphoid Organogenesis. 2000. 62 figs. XII, 215 pp. ISBN 3-540-67569-8
- Vol. 252: **Potter, Michael; Melchers, Fritz (Eds.):** B1 Lymphocytes in B Cell Neoplasia. 2000. XIII, 326 pp. ISBN 3-540-67567-1
- Vol. 253: **Gosztonyi, Georg (Ed.):** The Mechanisms of Neuronal Damage in Virus Infections of the Nervous System. 2001. approx. XVI, 270 pp. ISBN 3-540-67617-1
- Vol. 254: **Privalsky, Martin L. (Ed.):** Transcriptional Corepressors. 2001. 25 figs. XIV, 190 pp. ISBN 3-540-67569-8
- Vol. 255: **Hirai, Kanji (Ed.):** Marek's Disease. 2001. 22 figs. XII, 294 pp. ISBN 3-540-67798-4
- Vol. 256: **Schmaljohn, Connie S.; Nichol, Stuart T. (Eds.):** Hantaviruses . 2001, 24 figs. XI, 196 pp. ISBN 3-540-41045-7
- Vol. 257: **van der Goot, Gisou (Ed.):** Pore-Forming Toxins. 2001. 19 figs. IX, 166 pp. ISBN 3-540-41386-3
- Vol. 258: **Takada, Kenzo (Ed.):** Epstein-Barr Virus and Human Cancer. 2001. 38 figs. IX, 233 pp. ISBN 3-540-41506-8
- Vol. 259: **Hauber, Joachim; Vogt, Peter K. (Eds.):** Nuclear Export of Viral RNAs. 2001. 19 figs. IX, 131 pp. ISBN 3-540-41278-6
- Vol. 260: **Burton, Didier R. (Ed.):** Antibodies in Viral Infection. 2001. 51 figs. IX, 309 pp. ISBN 3-540-41611-0
- Vol. 261: **Trono, Didier (Ed.):** Lentiviral Vectors. 2002. 32 figs. X, 258 pp. ISBN 3-540-42190-4
- Vol. 262: **Oldstone, Michael B.A. (Ed.):** Arenaviruses I. 2002, 30 figs. XVIII, 197 pp. ISBN 3-540-42244-7
- Vol. 263: **Oldstone, Michael B. A. (Ed.):** Arenaviruses II. 2002, 49 figs. XVIII, 268 pp. ISBN 3-540-42705-8