Bacterial genomics: from sequencing one genome to thousands of genomes

The *Streptococcus agalactiae* case

Philippe Glaser

Institut Pasteur

cnrs
dépasser les frontières

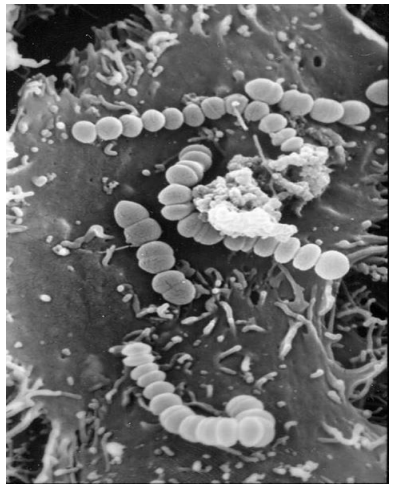
Institut Pasteur
de Tunis

Outline

- *Streptococcus agalactiae* or group B streptococcus
- Analysing1 genome
  - Sequencing methods
  - Bio-informatatics
  - What we learned
- Analysing 7 genomes
- Analysing 230 genomes
- Analysing 1500 genomes
- Conclusions and perspectives

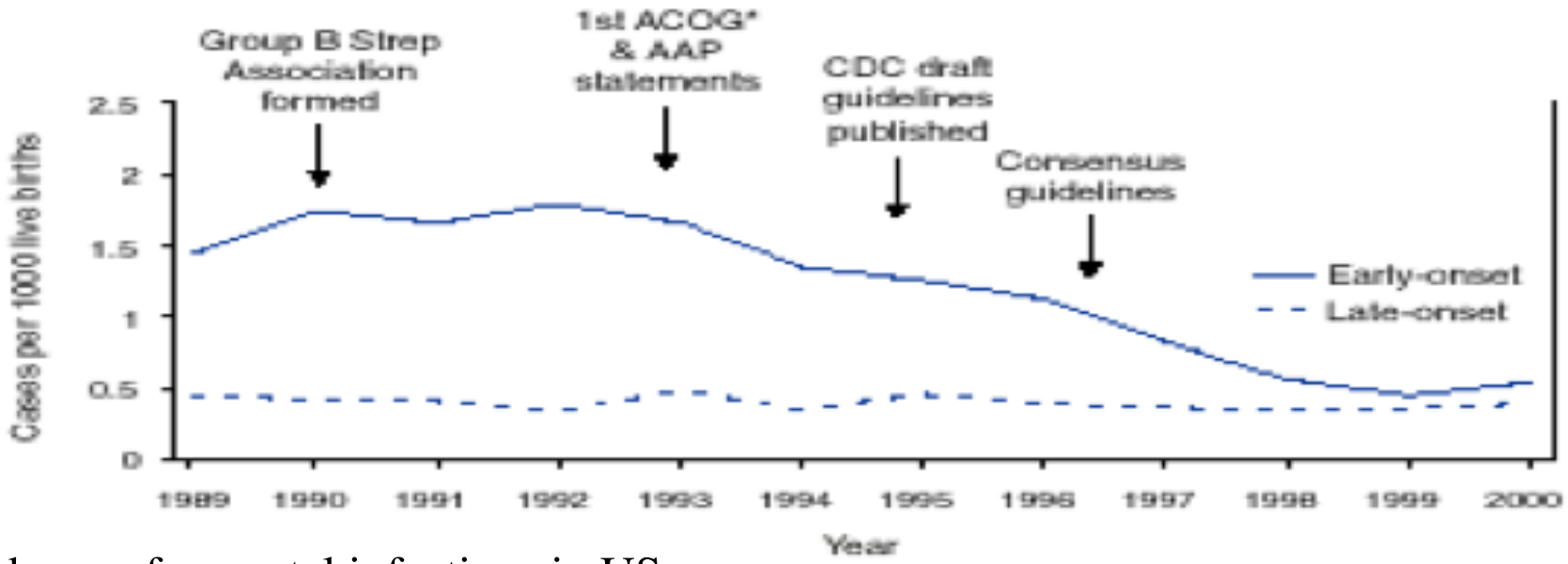
S. agalactiae or GBS

- Opportunistic pathogen
- Human
  - Commensal of the digestive or urinary tract
  - Leading cause of neonatal infections
  - Emerged during the 1960s–1970s
  - Risk for immuno-compromised adults and elderly



- Broad host spectrum in animals
  - Udder infections in bovines and camels
  - Invasive diseases in fish

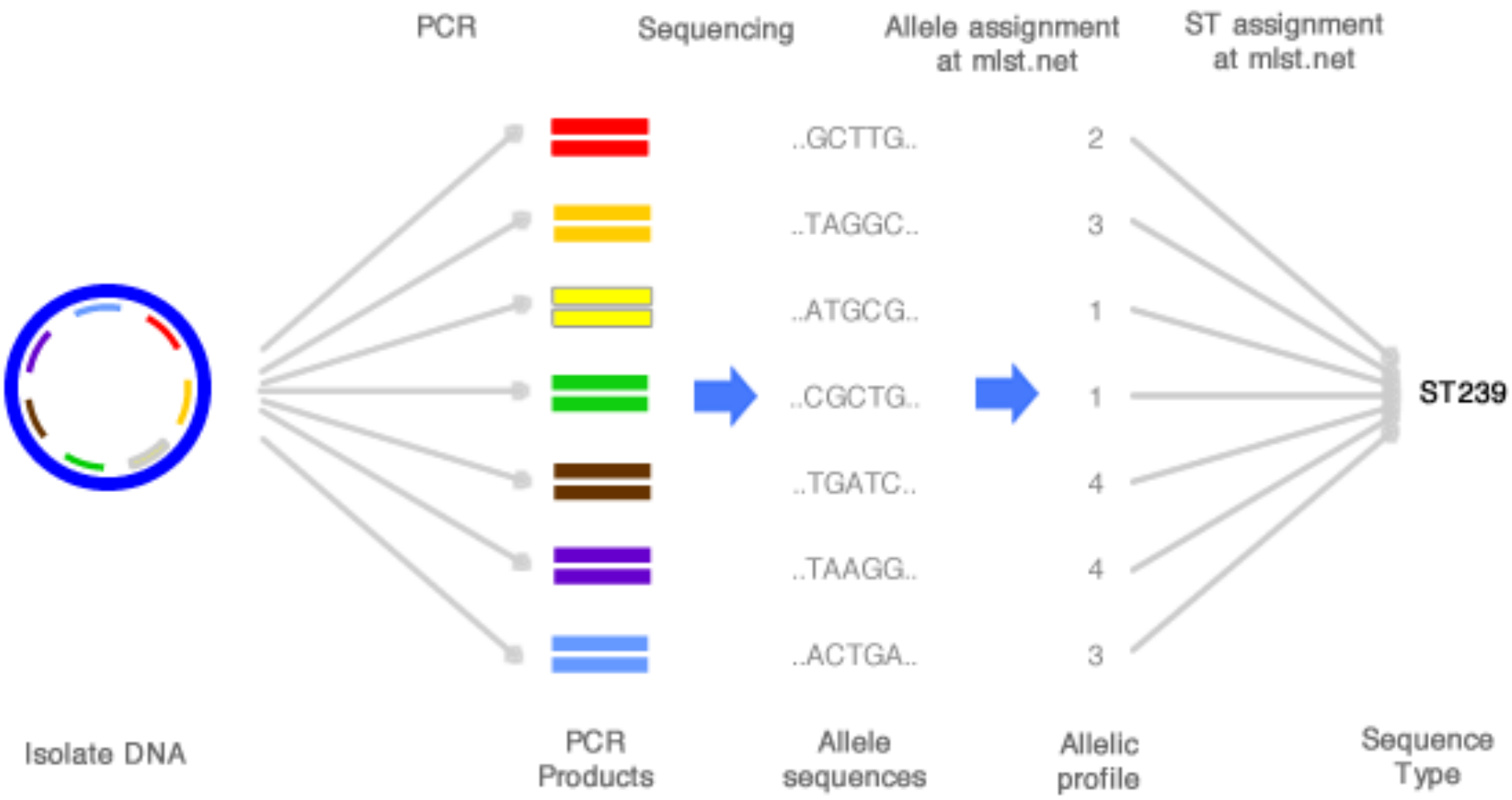
Recent epidemiology of GBS infections



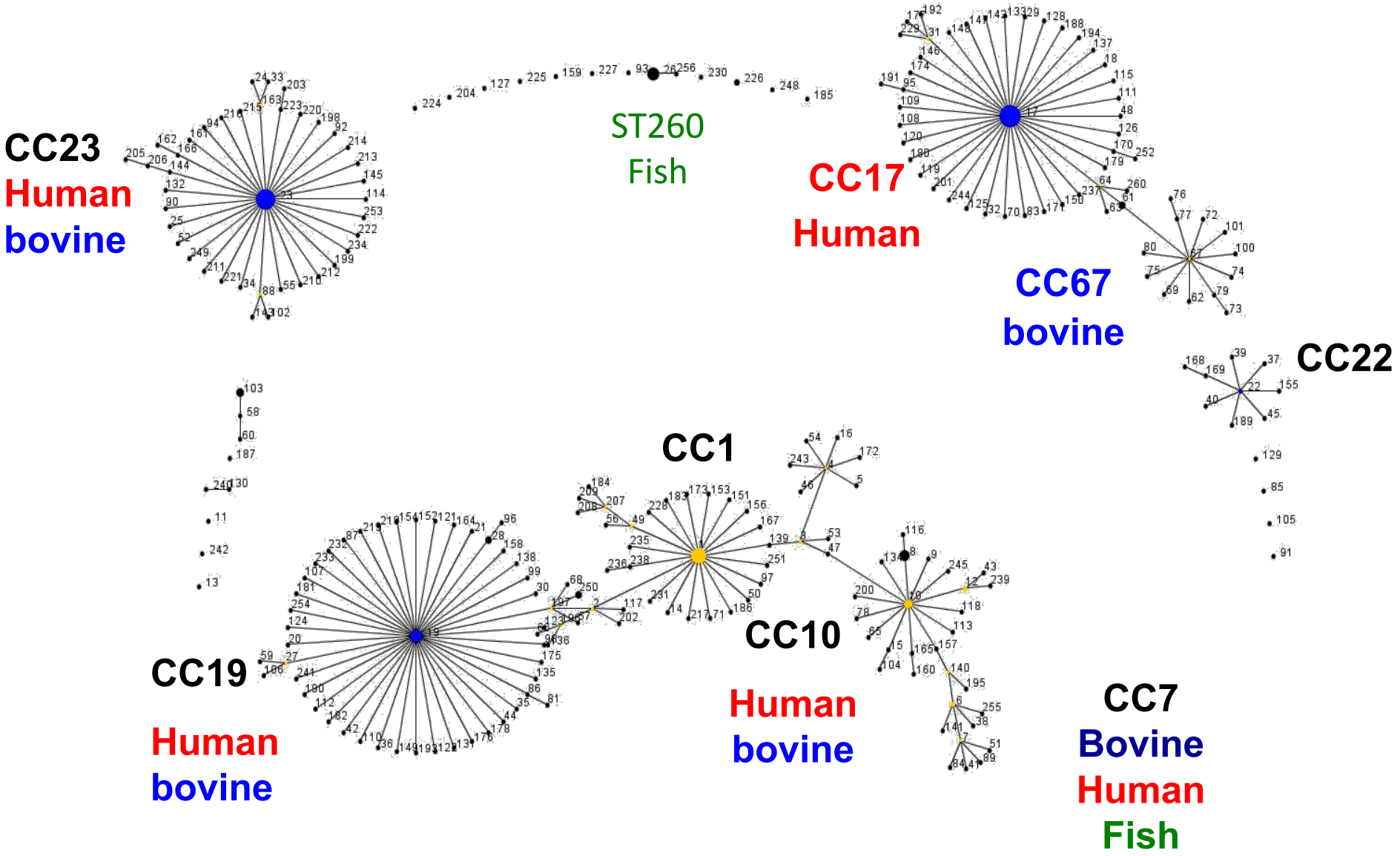
Incidence of neonatal infections in US

- Emergence of the disease in the 60s – 70s
- Relation between human and animal strains
- Transition from commensal to pathogen

Population structure of GBS based on Multi Locus Sequence Typing (MLST)



Population structure of GBS based on Multi Locus Sequence Typing (MLST)



Sanger sequencing

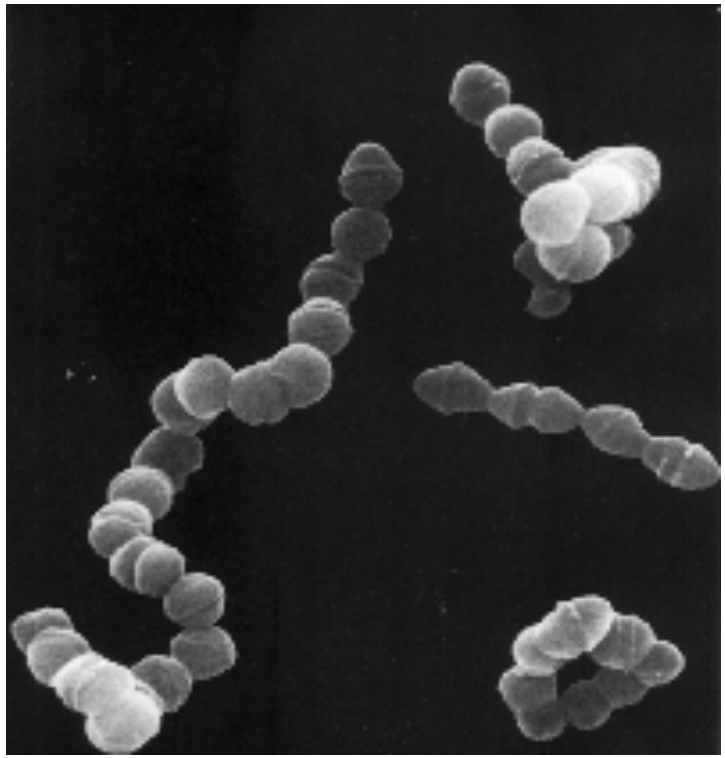
F. Sanger, et al. Cambridge UK
1977 - PNAS 74; Nature 256 ΦX174

ACGTGGGCTAAGTGCGTATGCATGCGTGCT
TGCACCCGATTCACGCATACGT
TGCACCCGATTCACGCATACG
TGCACCCGATTCACGCATAC
TGCACCCGATTCACGCATA
TGCACCCGATTCACGCAT
TGCACCCGATTCACGCA
TGCACCCGATTCACGC
TGCACCCGATTCACG

→ DNA copy

First genome sequenced: ΦX174 – 5386 bases

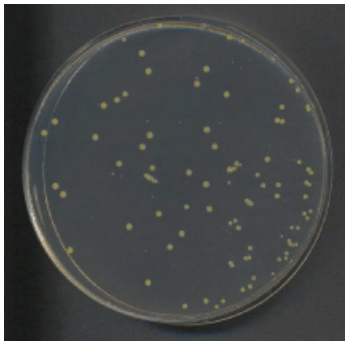
Sequencing one genome

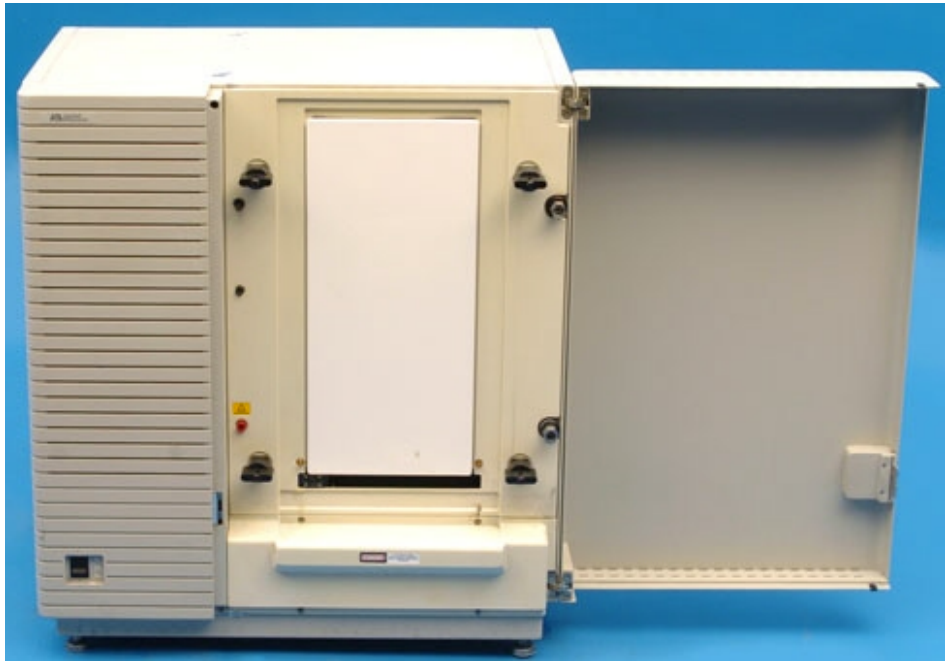


The reference strain NEM316

Whole genome shotgun sequencing

- Library construction in plasmid vectors
- Picking clones, minipreps (5000 / Mb)…
- Sequencing clones from both ends (700 bases)
- Assembly (phrap)
- Gap closure and polishing
- Expert annotation





Annotation framework – CAAT box

length = 434 AA Contig55 from 86155 to 87459 (p) ORF sequence come from 86149 to 87459 wide nucleotide sequence come from 85649 to 87659 GC% = 34 3rd=34 **SignalP NN+ HMM+**	**secY - *gbs0078*** **Unknown** similar to preprotein translocase SecY *1.6 Protein secretion*

BEST-BLASTP:

```
>Strep_pyogenes.aa 95%    Identities = 377/434 (86%), Positives = 417/434 (95%)

     AAK33202.1 secY putative preprotein translocase (73724,75028)          Length = 434

>nrprot 95%    Identities = 377/434 (86%), Positives = 417/434 (95%)

     ref|NP_268480.1| (NC_002737) putative preprotein translocase [Streptococcus          pyogenes] [Streptococcus pyogenes M1 GAS] ref|NP_606397.1| (NC_003485) putative pre

>Strep_Pneu_R6.aa 84%    Identities = 311/436 (71%), Positives = 366/436 (83%), Gaps = 2/436 (0%)

     secY, Multispanning membrane protein, translocator of proteins          205609:206919 forward MW:47353          Length = 436
```

USER COMMENT :

= Empty =

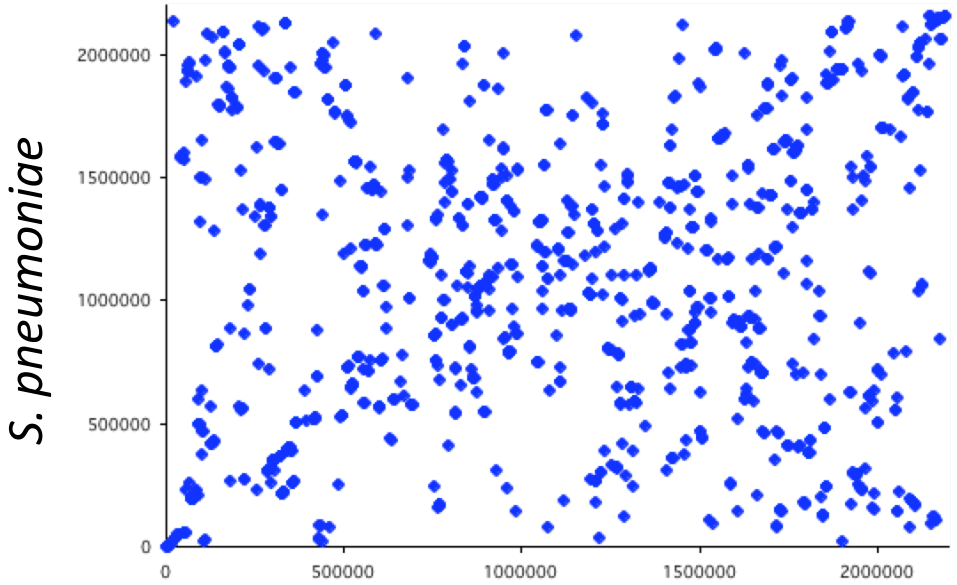
IPF Field				
MW-IP	BDBH	SEQCDS	BEST-BLASTP	EMBL_NAME
COMMENT_SEQCDS	RBS	SIGNALP	V_CLASS	V_PRODUCT
TERMINATOR	V_SNAME	SEQAA	SEQ_V_CDS	SEQNUCL
SEQNUCLARGE	V_RBS	V_NOTE	OLD_FIELDS	

IPF Results			
IPF Viewer	HMM_Pfam	IPF_mapping	auto-blastp
toppred SEQ_V_CDS	toppred SEQAA	blastp	gm graph ORF=500 on frame 3
gm txt ORF=500 on frame 3			

Analysing one genome

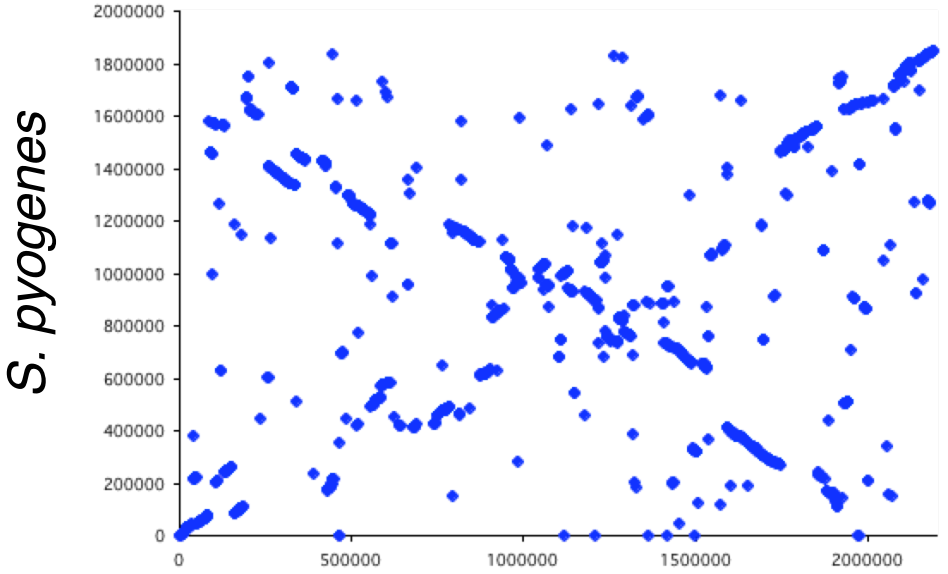
- Genome/species specificities
  - Abundant surface protein encoding genes
  - Regulatory systems
  - Metabolic reconstruction (auxotrophies, rich carbohydrate metabolism …)
- Genome comparison
  - Orthologs identifications (Bidirectional Blast best Hits)
- Phylogenomics

Genome comparison



S. agalactiae

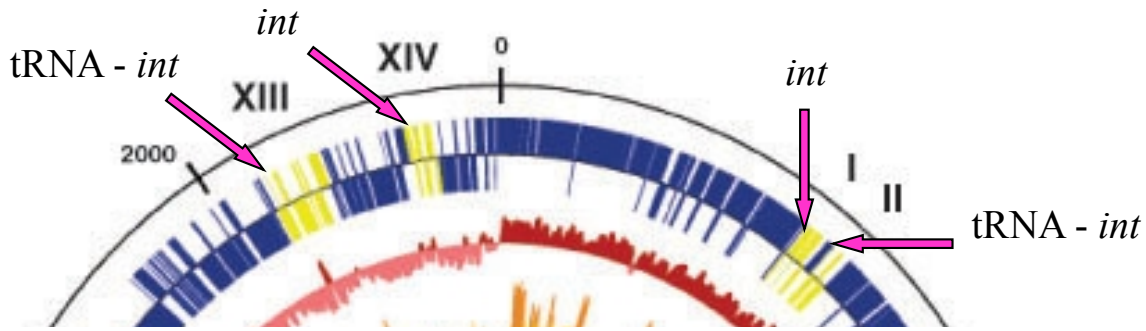
(1141 pairs of orthologous genes)



S. agalactiae

(1170 pairs of orthologous genes)

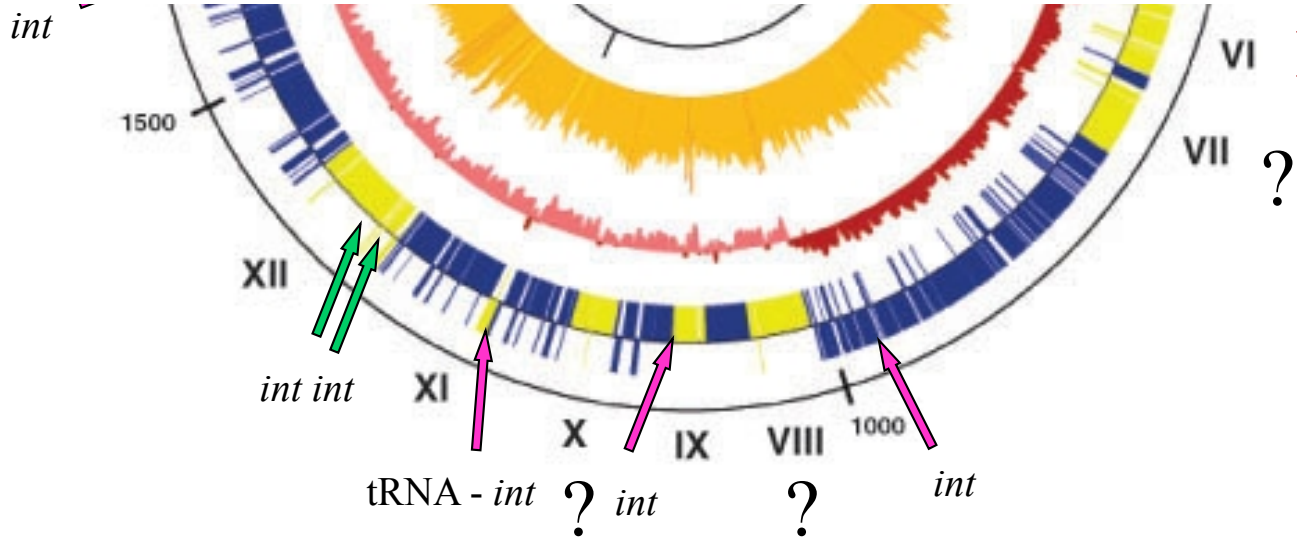
14 genomic islands - 9 associated with integrases



Molecular Microbiology (2009) 71(4), 948–959 ■

doi:10.1111/j.1365-2958.2008.06579.x
First published online 13 January 2009

Atypical association of DDE transposition with conjugation specifies a new family of mobile elements



Island?

? ? ?

Discoveries from One genome

- The abundance of integrative and conjugative elements (ICE).
- The discovery of a new family of ICEs
  - Use a DDE transposase and not an integrase
  - Specificity of insertion upstream promoters

Molecular Microbiology (2002) **45**(6), 1499–1513

Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease

Glaser et al. 2002

Comparison of multiple isolates: the pan genome concept

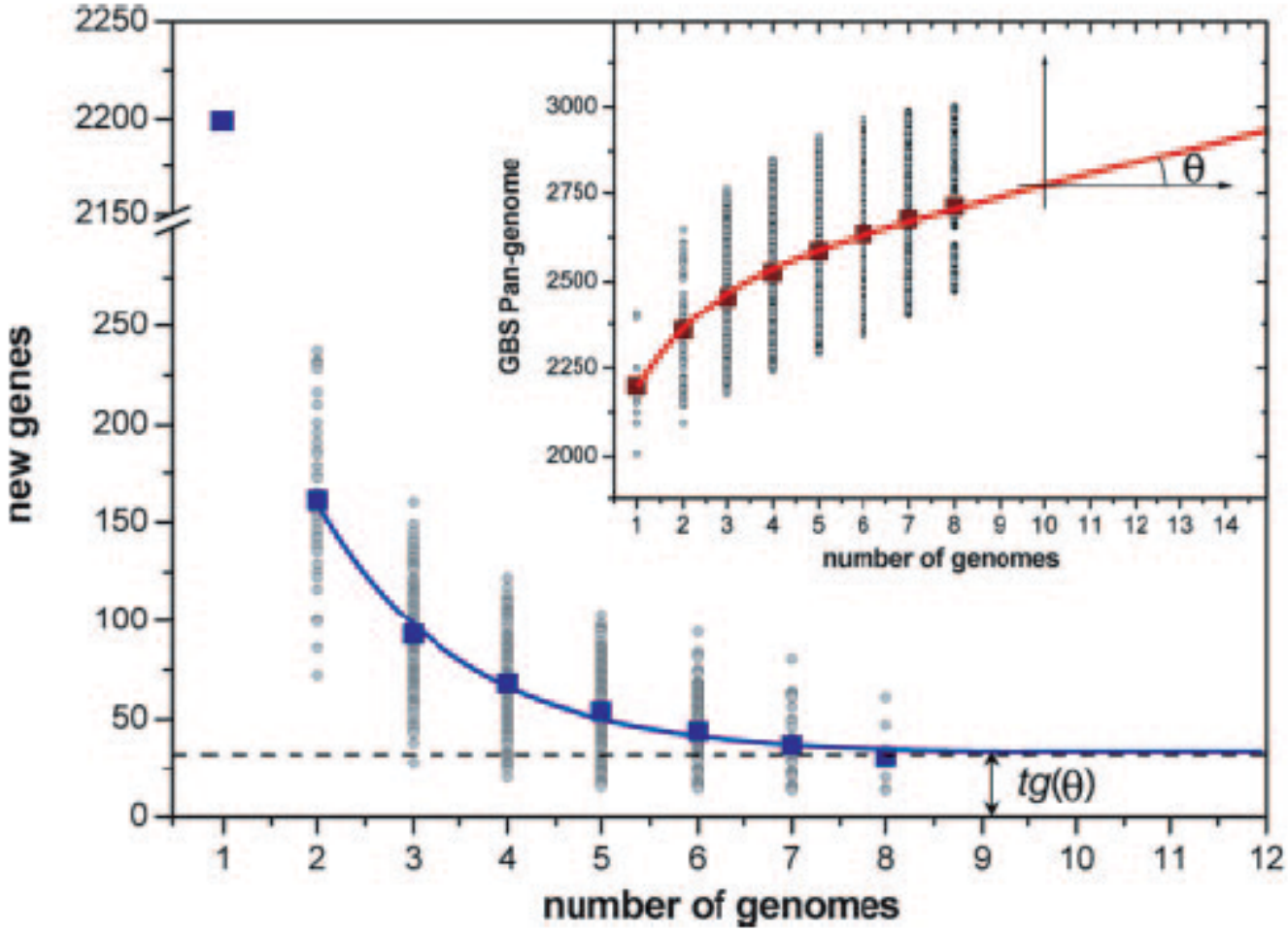
Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome"

Hervé Tettelin[a,b], Vega Masignani[b,c], Michael J. Cieslewicz[b,d,e], Claudio Donati[c], Duccio Medini[c], Naomi L. Ward[a,f], Samuel V. Angiuoli[a], Jonathan Crabtree[a], Amanda L. Jones[g], A. Scott Durkin[a], Robert T. DeBoy[a], Tanja M. Davidsen[a], Marirosa Mora[c], Maria Scarselli[c], Immaculada Margarit y Ros[c], Jeremy D. Peterson[a], Christopher R. Hauser[a], Jaideep P. Sundaram[a], William C. Nelson[a], Ramana Madupu[a], Lauren M. Brinkac[a], Robert J. Dodson[a], Mary J. Rosovitz[a], Steven A. Sullivan[a], Sean C. Daugherty[a], Daniel H. Haft[a], Jeremy Selengut[a], Michelle L. Gwinn[a], Liwei Zhou[a], Nikhat Zafar[a], Hoda Khouri[a], Diana Radune[a], George Dimitrov[a], Kisha Watkins[a], Kevin J. B. O'Connor[h], Shannon Smith[i], Teresa R. Utterback[i], Owen White[a], Craig E. Rubens[g], Guido Grandi[c], Lawrence C. Madoff[a,j], Dennis L. Kasper[a,j], John L. Telford[c], Michael R. Wessels[d,e], Rino Rappuoli[c,k,l], and Claire M. Fraser[a,b,k,m]

Methods

- 8 sequenced genomes: 2 from the internet, 5 draft genome sequence and 1 complete (ST1, ST6, ST7, ST17, ST19, ST23)
- Annotation
- Search for conserved genes (>50% identities, >50 length)
  - Core genome: genes shared by all isolates
  - Strain specific genes
  - Pan genome: all the genes
  - Statistical analysis to extrapolate

Open vs. closed genome



⇒*Streptococcus agalactiae* genome is open

⇒*Bacillus anthracis* genome is closed

Tettelin et a. PNAS 2005

A second analysis of these 8 genomes

Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*

Mathieu Brochet*, Christophe Rusniok†, Elisabeth Couvé*, Shaynoor Dramsi‡, Claire Poyart§, Patrick Trieu-Cuot‡, Frank Kunst*, and Philippe Glaser*¶

PNAS 2008

$\Rightarrow$ Impact of recombination of Bacterial evolution

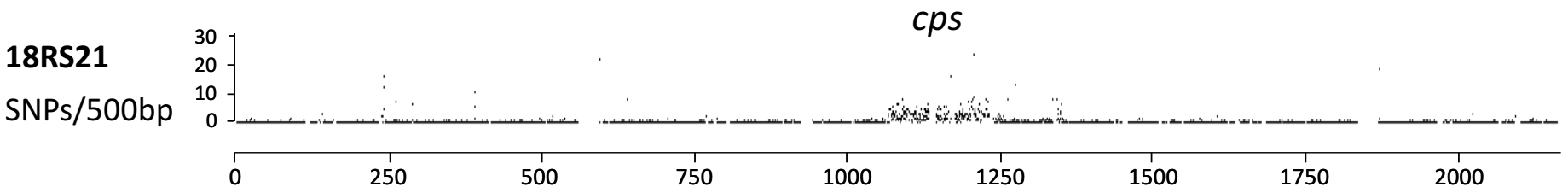
A simple method to identify large genetic exchanges

1. Each genome is fragmented in 500bp fragments
2. BLASTn comparison against an other genome sequence
3. For each comparison, the number of SNPs is counted
4. Analysis of the distribution of SNPs along the chromosome

$\Rightarrow$ Todays other tools to detect recombination: Gubbins, Bratnextgen.

2603 V/R versus 18RS21 :

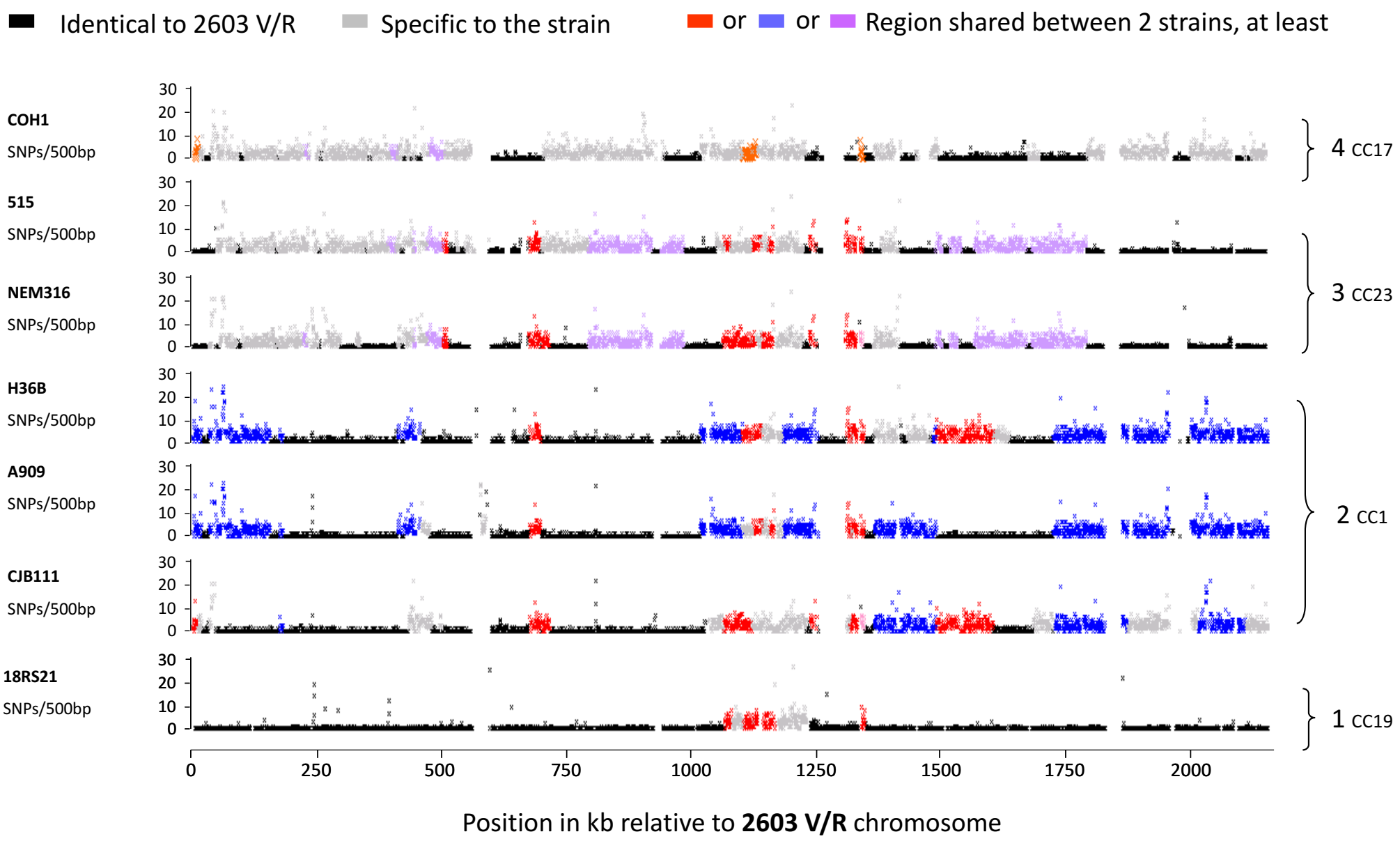
diversification by recombination



Position in kb relative to **2603 V/R** chromosome

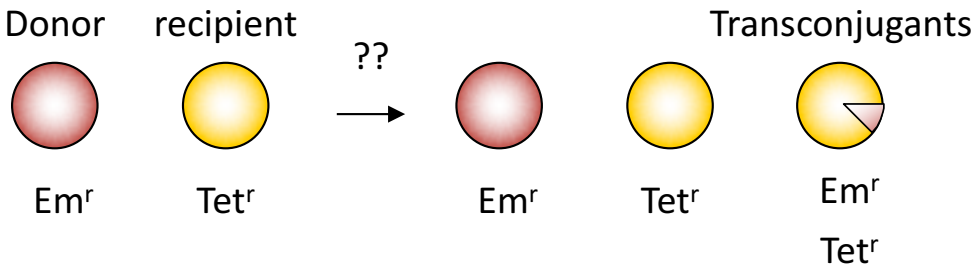
- 18RS21 ST19 (CC 19) serotype II
- 2603 V/R ST110 (CC 19) serotype V

Polymorphisms among the 8 genome sequences



Genetic exchanges under laboratory conditions

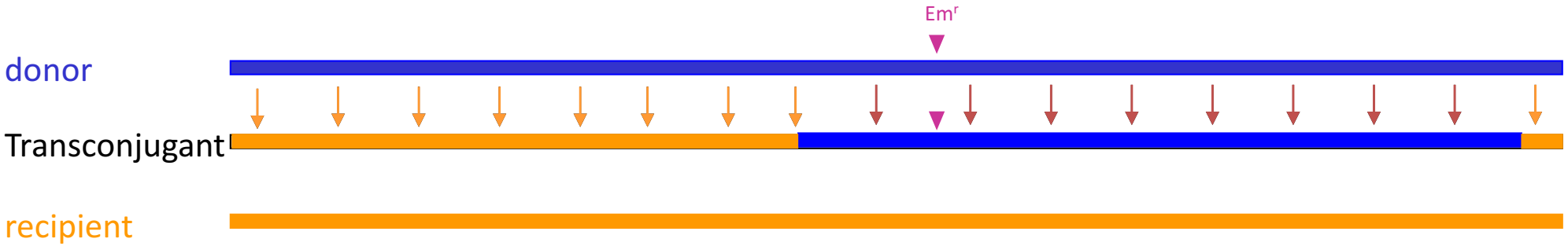
Conjugation essays



Essays on 30 Erm insertions scattered along the chromosome

$\Rightarrow$transfers of the Erm cassette observed in 20 cases

Characterization of the ErmR clones by DNA sequencing

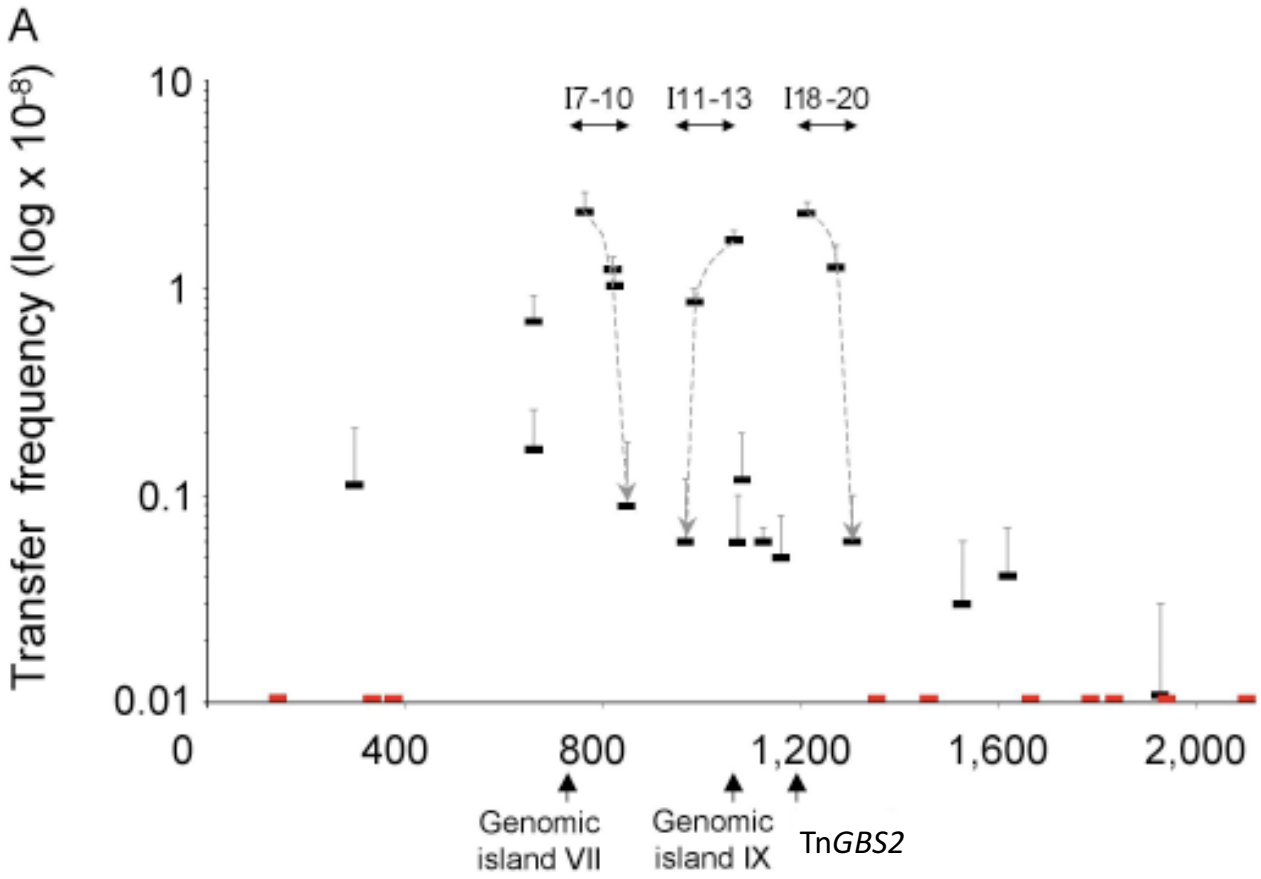


⇒The ErmR marker is inserted at the same location as in the donor

⇒A single fragment is transferred into recipient cell

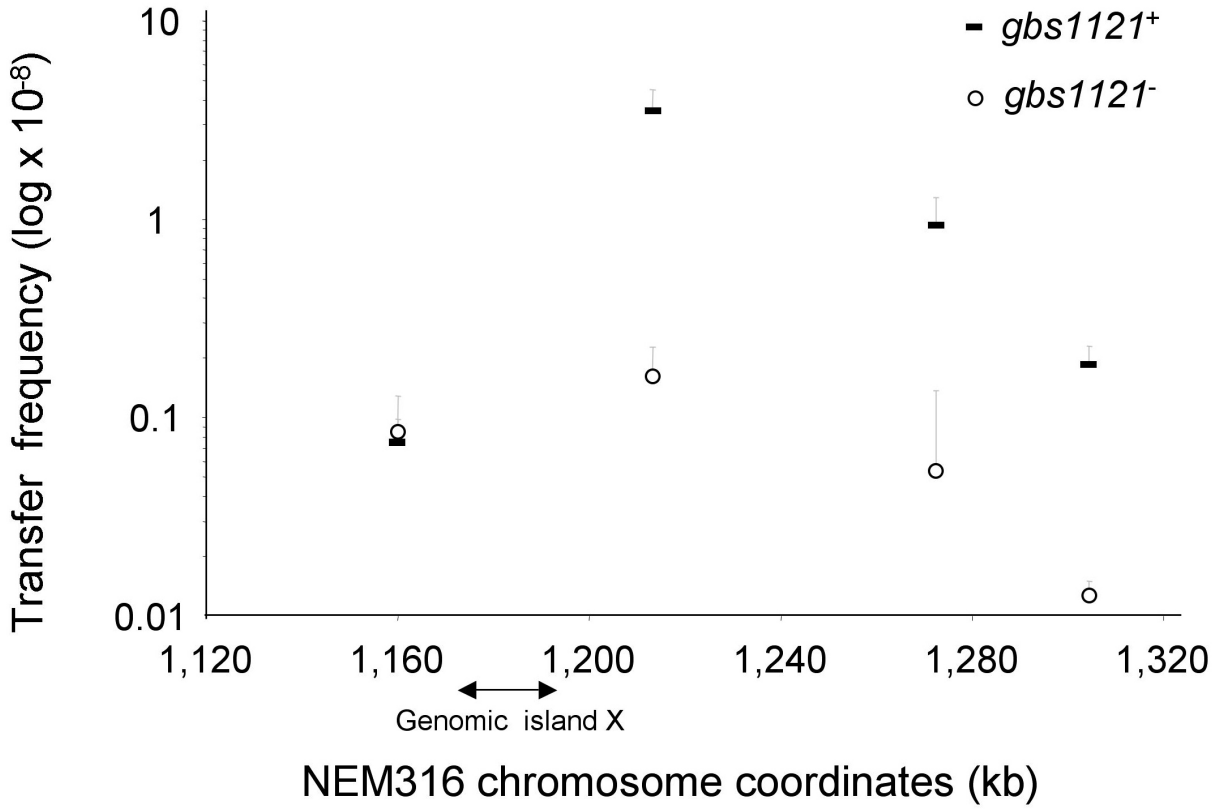
⇒Fragments of up to 332 kb could be transferred

Variable transfer frequencies along the genome



Role of MGE in genetic exchange

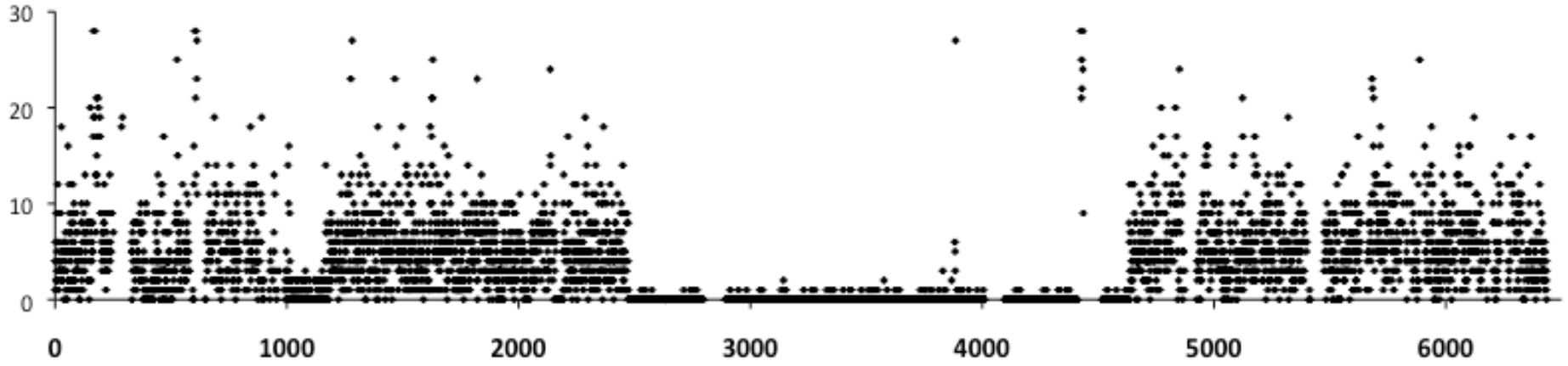
Effect of the inactivation of TnGBS2 relaxase gene



⇒ Hfr type mechanism of DNA transfer

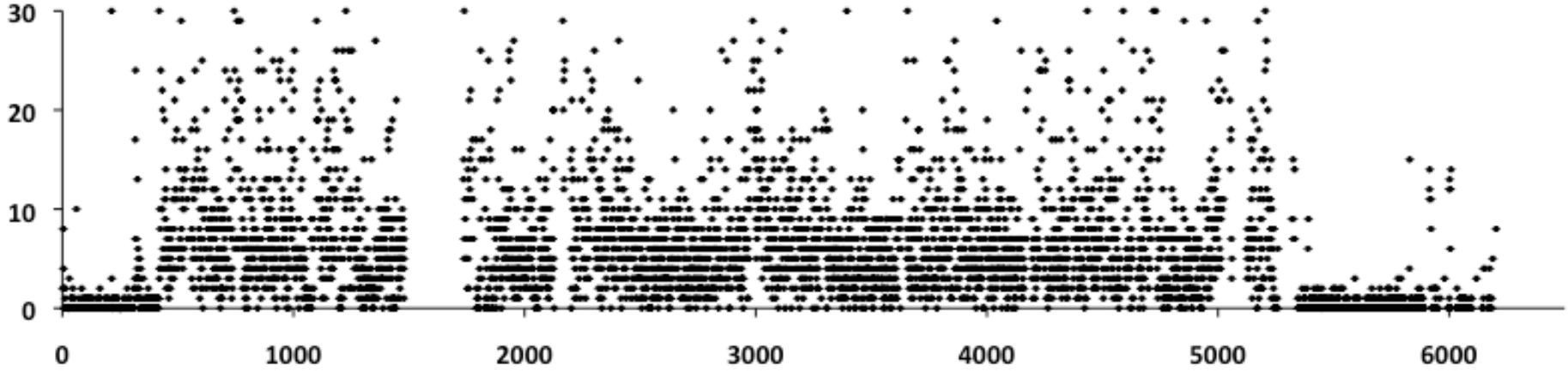
Conclusions from 8 genomes

- *S. agalactiae* shows an open pan genome and diverse MGEs
- *S. agalactiae* genomes are shaped by the transfer of large DNA regions
- This transfer result of an Hfr type mechanism mediated by inserted MGE (Tn*GBS*)
- Is this specific to *S. agalactiae* or was the role of conjugation underestimated?



Enterococcus faecalis, V583 versus TX104

⇒exchange of more than 1000 kb



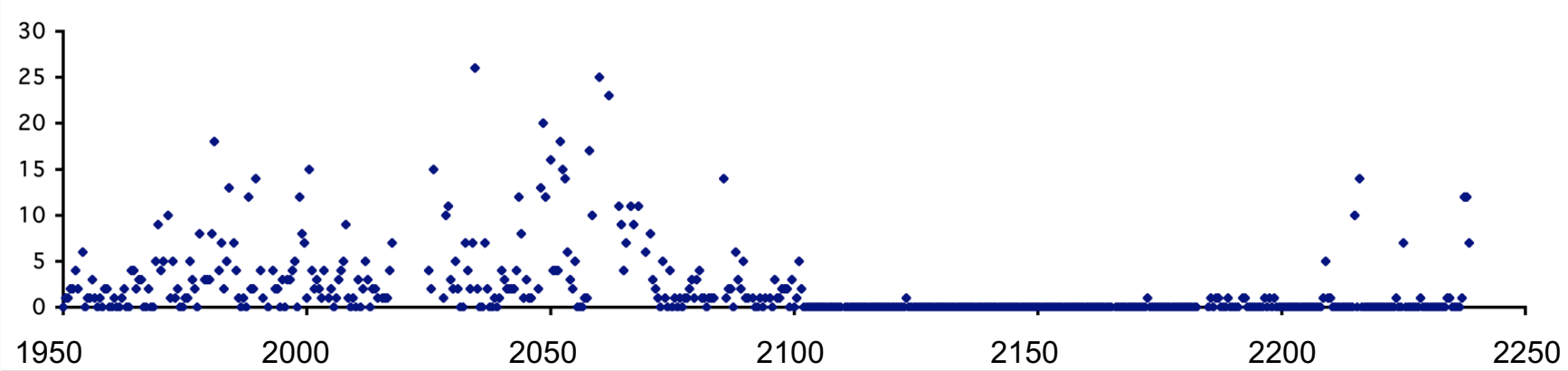
Staphylococcus aureus, MRSA252 versus 0582

⇒exchange of more than 500 kb

Escherichia coli UTI89 - 536

Identification of a 135 kb transfered region

Nb of SNPs / 500 nt



Position *E. coli* UTI89 (kb)

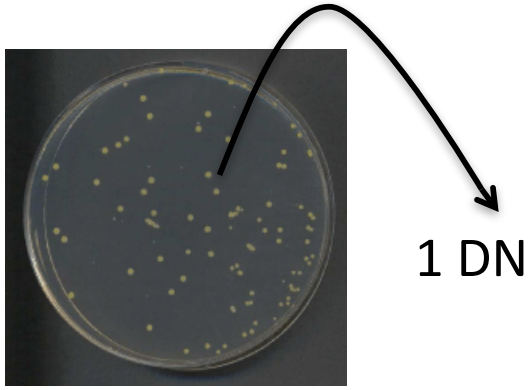
⇒Pathogenicity island : *irp* locus and transfer of adjacent regions including the *cob*, *his* … operons

Conclusion

Conjugation in bacterial para sexuality is more important than previously thought

New sequencing technologies

Sanger sequencing

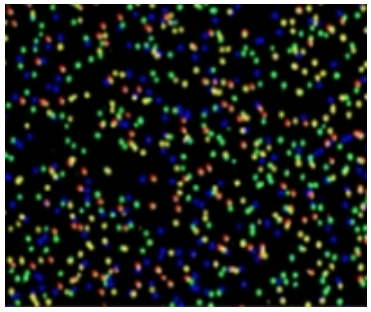


1 DNA

1 sequencing reaction

96 sequences of 1 kb

Next generation sequencing (second)

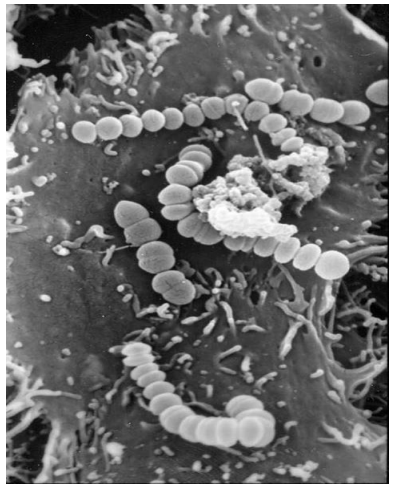


N amplified DNA

N sequencing reactions up to 300 bases

S. agalactiae or group B streptococcus (GBS)

- Opportunistic pathogen
- Human
  - Commensal of the digestive or urinary tract
  - Leading cause of neonatal infections
  - Emerged during the 1960s–1970s
  - Risk for immuno-compromised adults and elderly



- Broad host spectrum in animals
  - Udder infections in bovines and camels
  - Invasive diseases in fish - outbreaks

Group B streptococcus in humans

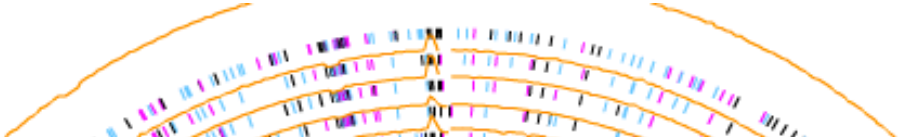
- Objective
  - To confirm the observed emergence of GBS infections at the bacterial population level
  - To get clues on the reasons for this emergence
- Strains:
  - Sanger Centre:
    - 92 European strains – DEVANI consortium (7 countries)
    - 24 Australian strains
  - Institut Pasteur:
    - 112 isolates: 27 from Africa, carriage and clinical isolates, 9 bovine strains.
    - 12 strains isolated between 1953 and 1961

Ad hoc sampling – MLST based diversity
Meta data: geographical origin, year of isolation, carriage / disease ….

Methods

- *de novo* assembly and gene content analysis
- mapping against reference genomes to identify variants (SNPs, short indels)
- Pseudo sequence alignment of concatenated variable positions (SNPs) (recombination filtered)
- Maximum likelihood phylogeny
- Bayesian analysis for dating the most recent common ancestors of lineages
- Characterization of the resistome and of the mobilome

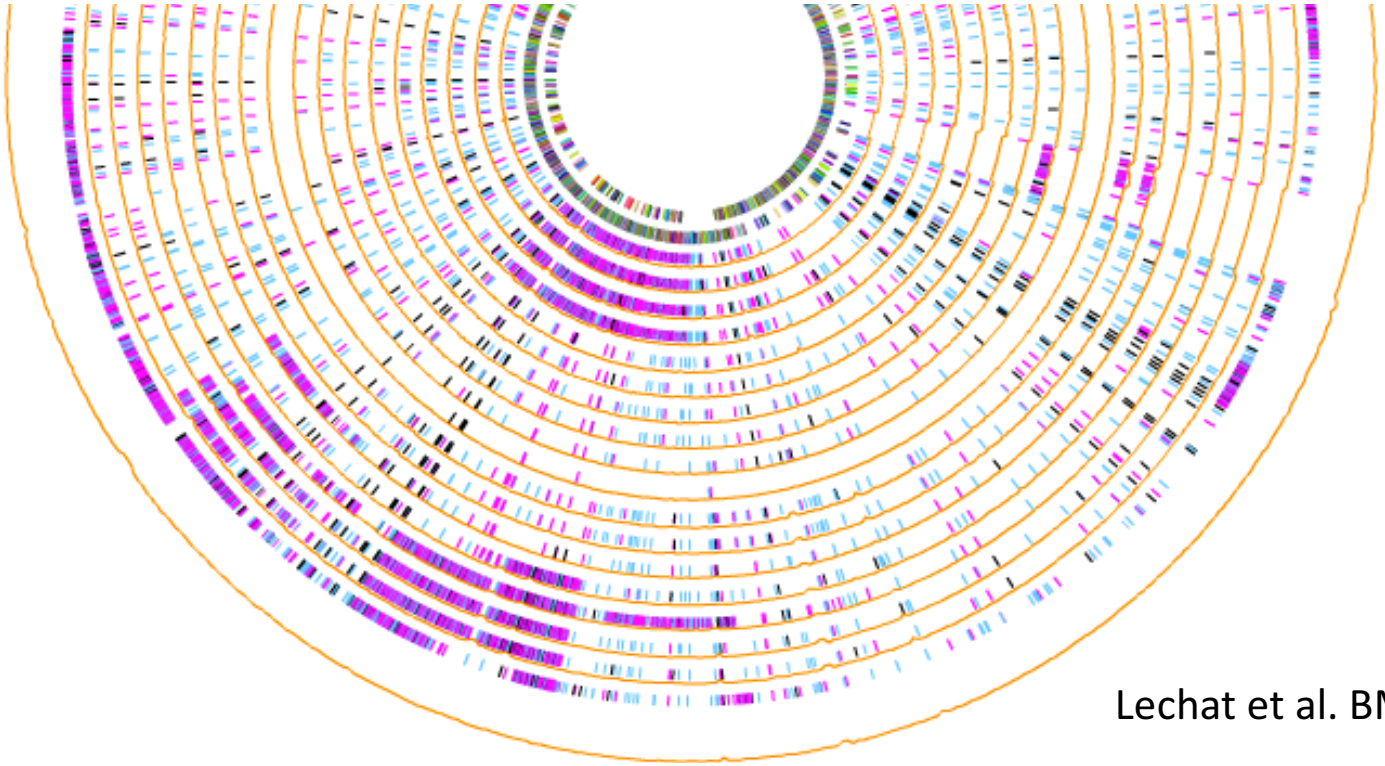
Circular view of CC10 strains



⇒Diversification occurs mostly through recombination

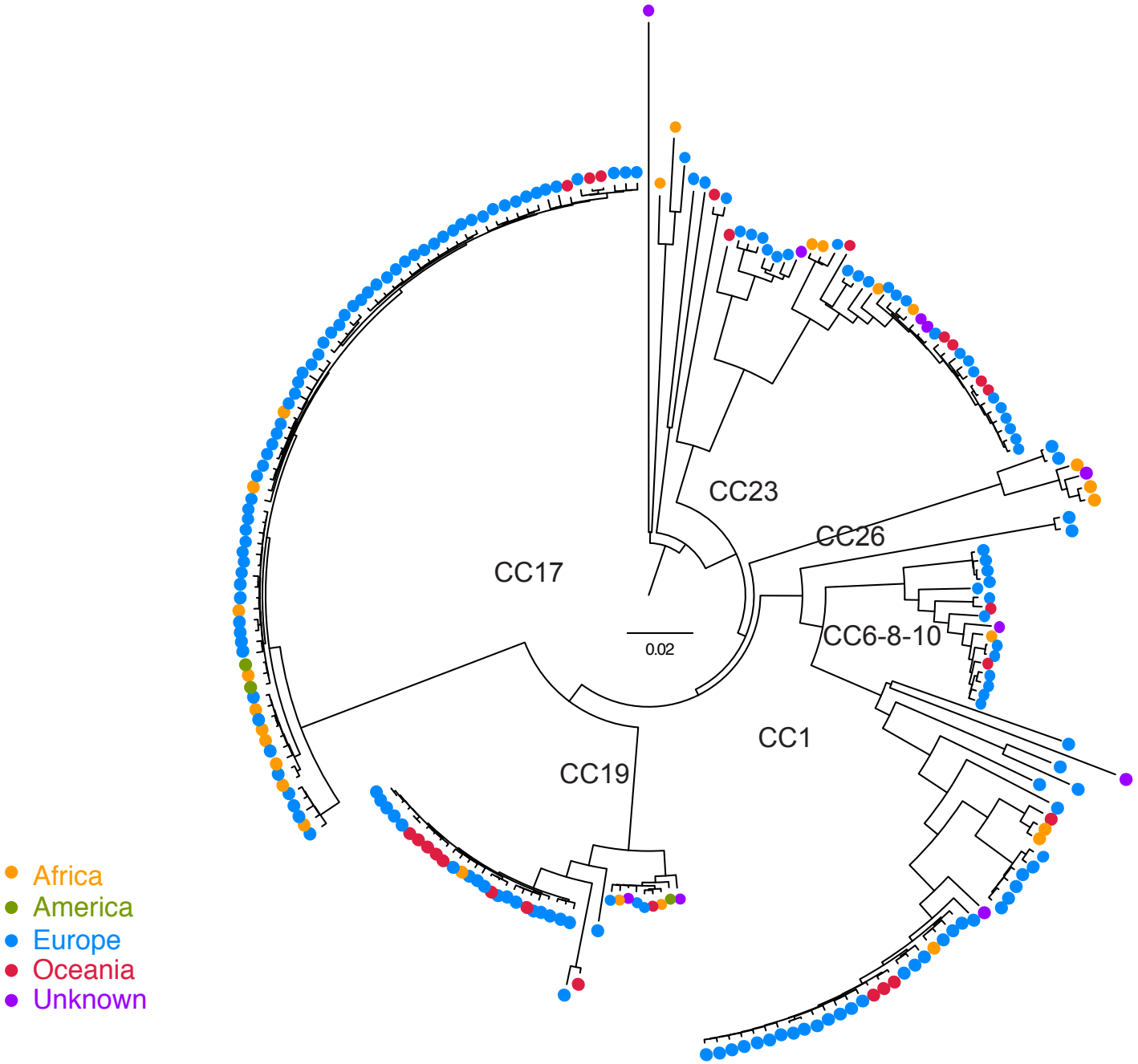
⇒The tree is not representative of the phylogeny of the isolates

⇒Phylogeny without considering recombined regions on each clonal complex

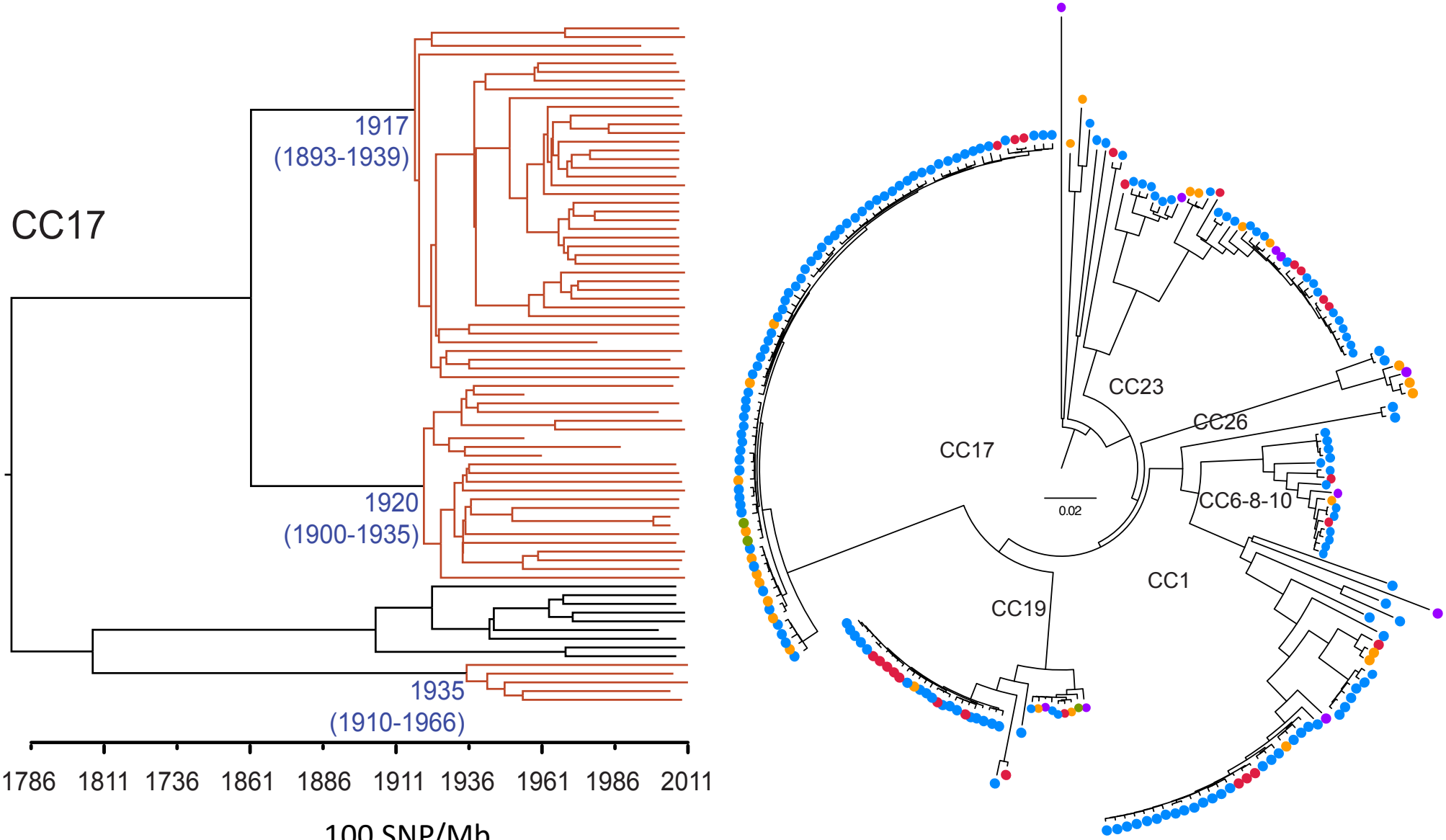


Lechat et al. BMC Bioinfo. 2013

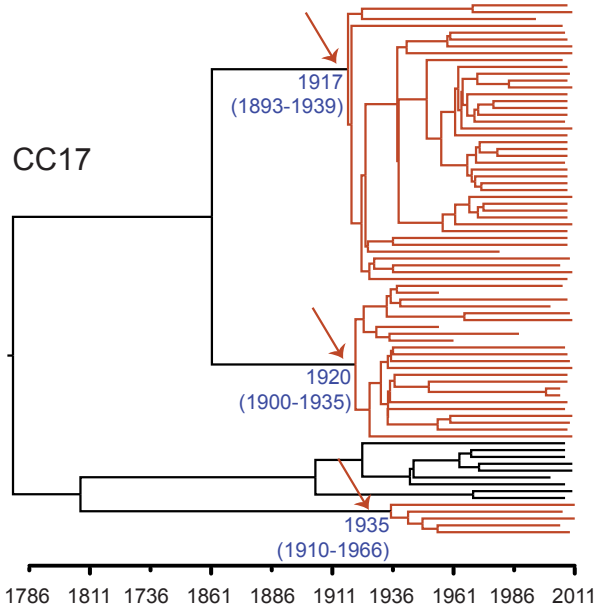
Phylogenetic relationships between the 230 Isolates

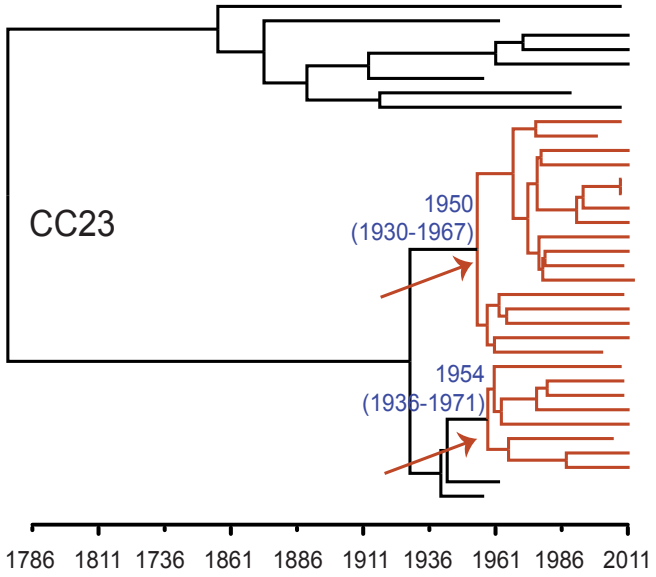


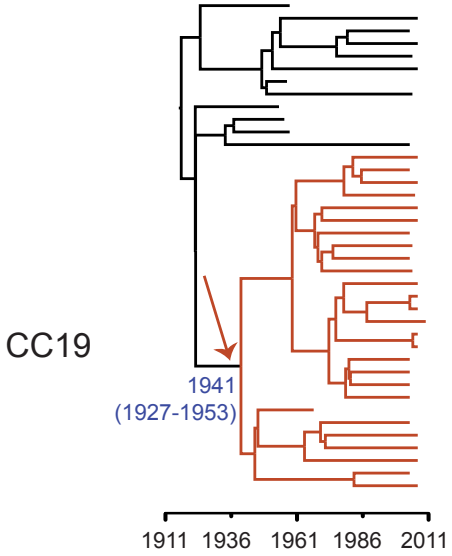
Time frame for the emergence of clones and complexes

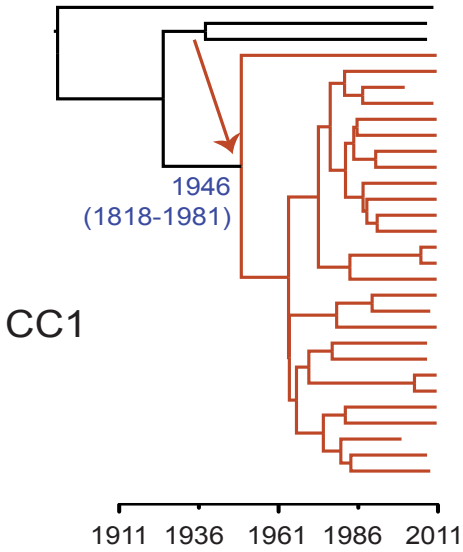


Bayesian analysis









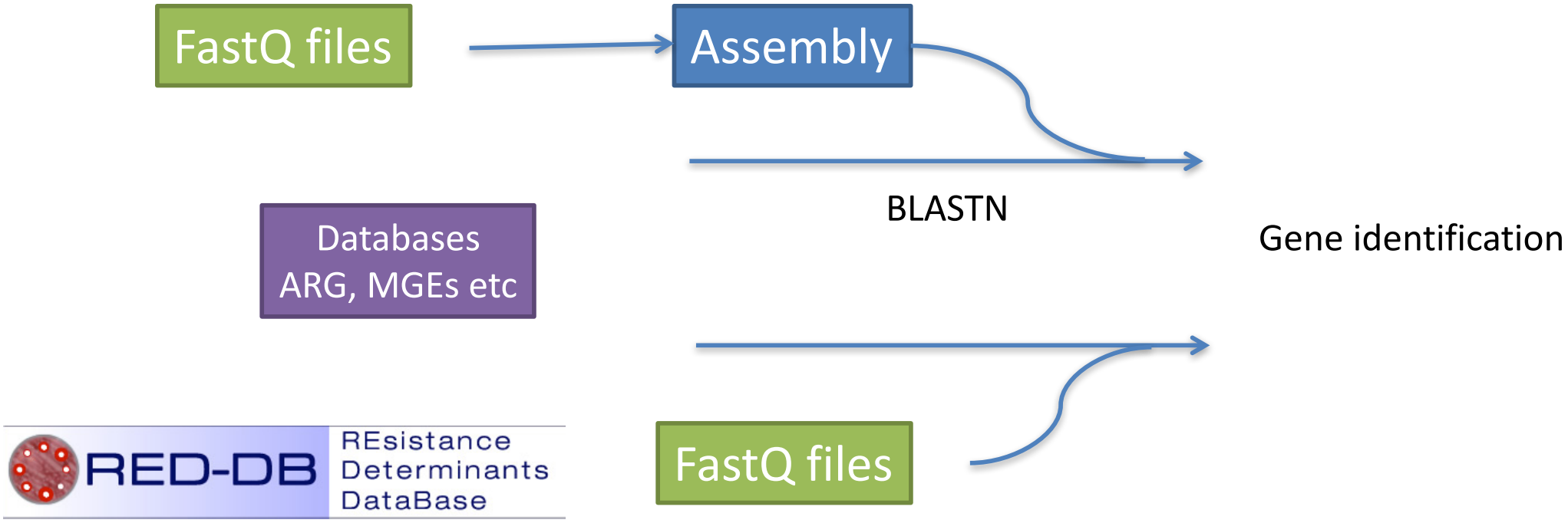
- Similar timeframe for the origin of the 4 major clonal complex
- Timing for the emergence of major clones correspond to the emergence of GBS infections.

Antibiotic resistance in GBS

In all studies:

- More than 80 % of human isolates are tetracycline resistant

Search for antibiotic ressitance genes and mobile genetic elements



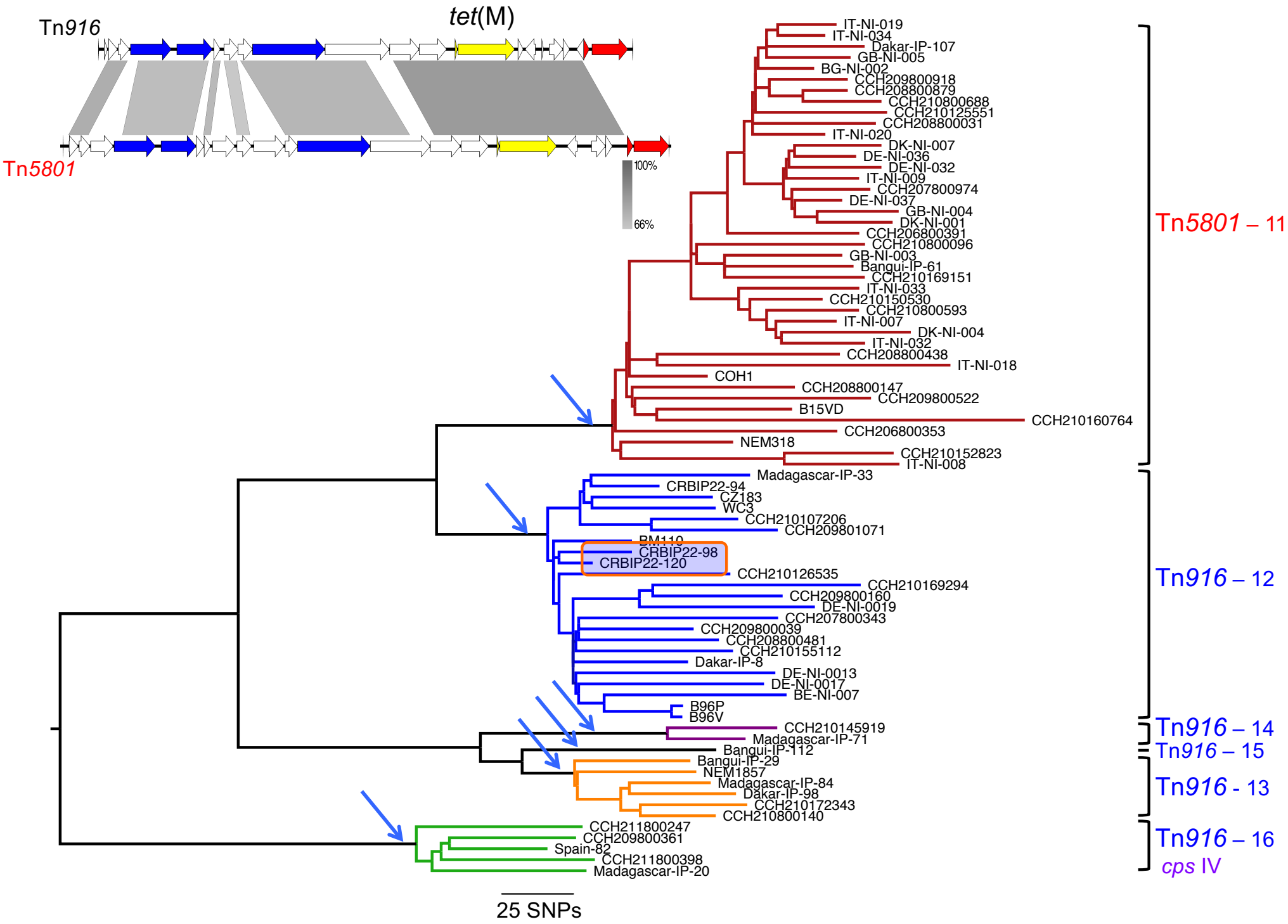
Antibiotic resistance in GBS

In all studies:

- More than 80 % of human isolates are tetracycline resistant

In the 230 analysed genomes

- 190 isolates express the *tet*(M) gene and 11 the *tet*(O) gene


Tn916
tet(M)
Tn5801
100%
66%
IT-NI-019
IT-NI-034
Dakar-IP-107
GB-NI-005
BG-NI-002
CCH209800918
CCH208800879
CCH210800688
CCH210125551
CCH208800031
IT-NI-020
DK-NI-007
DE-NI-036
DE-NI-032
IT-NI-009
CCH207800974
DE-NI-037
GB-NI-004
DK-NI-001
CCH206800391
CCH210800096
GB-NI-003
Bangui-IP-61
CCH210169151
IT-NI-033
CCH210150530
CCH210800593
IT-NI-007
DK-NI-004
IT-NI-032
CCH208800438
IT-NI-018
COH1
CCH208800147
CCH209800522
B15VD
CCH210160764
CCH206800353
NEM318
CCH210152823
IT-NI-008
Madagascar-IP-33
CRBIP22-94
CZ183
WC3
CCH210107206
CCH209801071
BM110
CRBIP22-98
CRBIP22-120
CCH210126535
CCH210169294
CCH209800160
DE-NI-0019
CCH207800343
CCH209800039
CCH208800481
CCH210155112
Dakar-IP-8
DE-NI-0013
DE-NI-0017
BE-NI-007
B96P
B96V
CCH210145919
Madagascar-IP-71
Bangui-IP-112
Bangui-IP-29
NEM1857
Madagascar-IP-84
Dakar-IP-98
CCH210172343
CCH210800140
CCH211800247
CCH209800361
Spain-82
CCH211800398
Madagascar-IP-20
25 SNPs
Tn5801 – 11
Tn916 – 12
Tn916 – 14
Tn916 – 15
Tn916 - 13
Tn916 – 16
cps IV

Few facts

- Tetracycline a broad spectrum antibiotic extensively used starting in the 1950s
- The high rate of tetracycline resistance is specific to human strains compared to bovine strains (but not the use of tetracycline)
- Tetracycline is today rarely used but tetracycline resistance remains at a high rate.

Proposed scenario for the emergence of GBS neonatal infections

Before 1950: a diverse population of GBS teracycline sensitive (unknown)

1950: Extensive use of tetracycline

1. Selection of TcR isolates by gain of mobile genetic elements
2. Created a niche by eliminating TcS GBS and by altering the gut microbiota
3. Among TcR clones selection of those with higher colonization and dissemination properties
4. Worldwide dissemination of few TcR clones with higher virulence potential

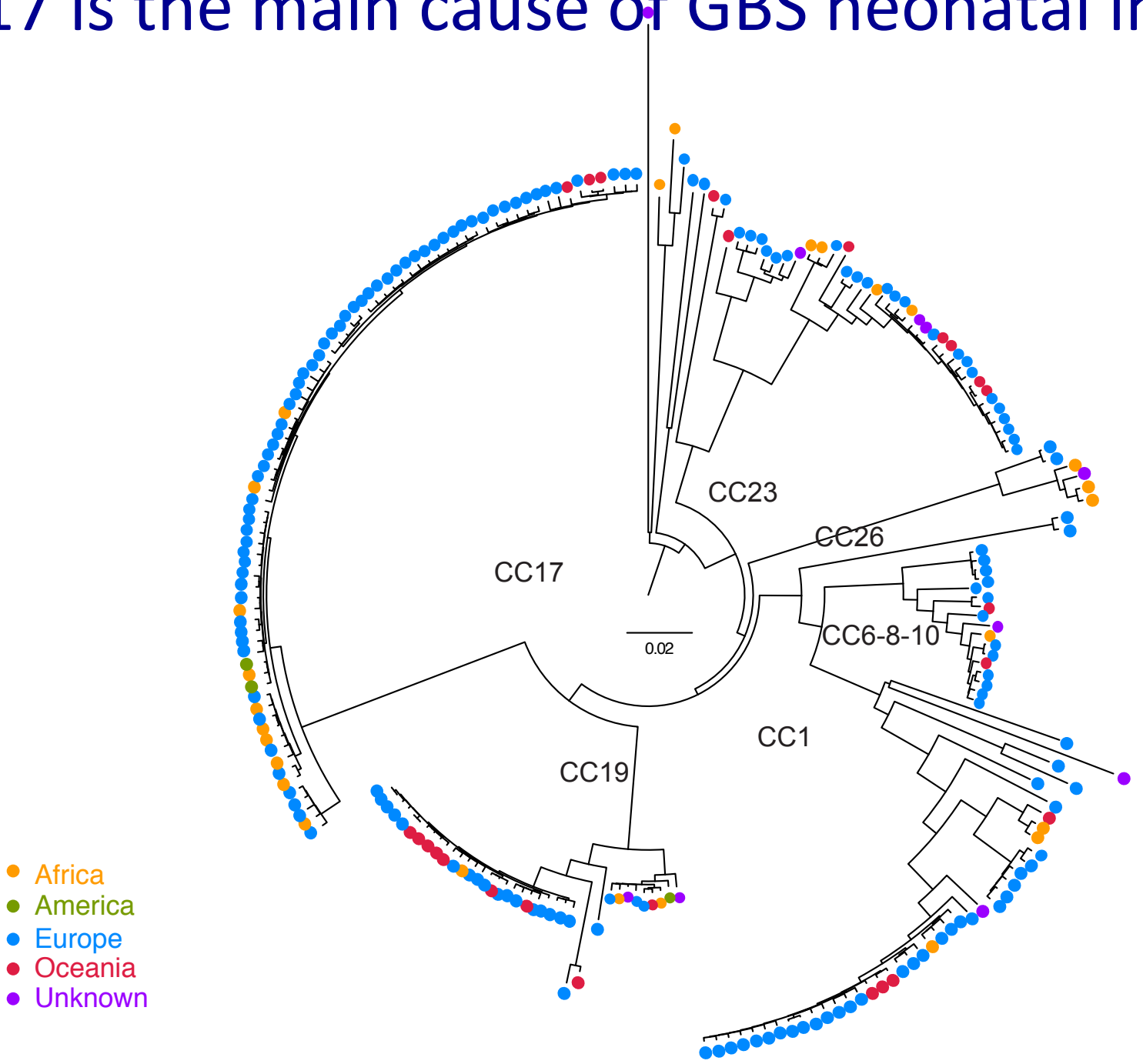
(Da Cunha *et al*. Nature Communication 2014)

Major pandemic antibiotic resistant clones

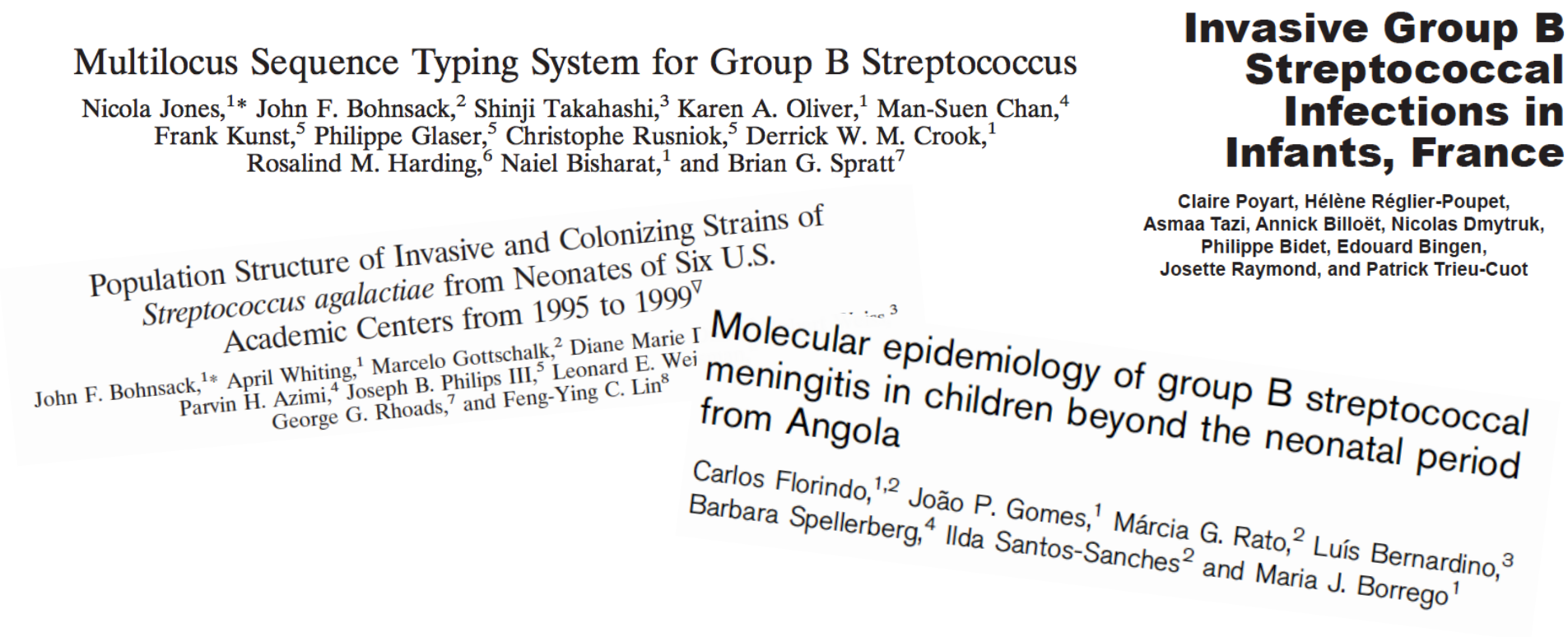
- *Staphylococcus aureus* USA300 methicilin-R
- *Clostridium difficile* (027/BI/NAP1) fluoroquinolones-R
- *Escherichia coli* ST131 fluoroquinolone-R
- Klebsiella pneumoniae ST258 (MDR)

These clones are considered as highly virulent

CC17 is the main cause of GBS neonatal infections



CC17 is the main cause of GBS neonatal infections



Multilocus Sequence Typing System for Group B Streptococcus

Nicola Jones,[1]* John F. Bohnsack,[2] Shinji Takahashi,[3] Karen A. Oliver,[1] Man-Suen Chan,[4] Frank Kunst,[5] Philippe Glaser,[5] Christophe Rusniok,[5] Derrick W. M. Crook,[1] Rosalind M. Harding,[6] Naiel Bisharat,[1] and Brian G. Spratt[7]

Invasive Group B Streptococcal Infections in Infants, France

Claire Poyart, Hélène Réglier-Poupet, Asmaa Tazi, Annick Billoët, Nicolas Dmytruk, Philippe Bidet, Edouard Bingen, Josette Raymond, and Patrick Trieu-Cuot

Population Structure of Invasive and Colonizing Strains of *Streptococcus agalactiae* from Neonates of Six U.S. Academic Centers from 1995 to 1999[▽]

John F. Bohnsack,[1]* April Whiting,[1] Marcelo Gottschalk,[2] Diane Marie I
Parvin H. Azimi,[4] Joseph B. Philips III,[5] Leonard E. Wei
George G. Rhoads,[7] and Feng-Ying C. Lin[8]

Molecular epidemiology of group B streptococcal meningitis in children beyond the neonatal period from Angola

Carlos Florindo,[1,2] João P. Gomes,[1] Márcia G. Rato,[2] Luís Bernardino,[3] Barbara Spellerberg,[4] Ilda Santos-Sanches[2] and Maria J. Borrego[1]

Objective: to identify loci important for the adaptation of GBS CC17 to the human host, and differences between carriage and disease-associated strains

CC17 strains currently available

	CC17
New strains (CIP/CNR-Strep/Angola)	**45**
Almeida, *et al*. (2015) J. Bacteriology	25
Campisi, *et al*. (2016) Frontiers in Microbiology	14
Da Cunha, *et al*. (2014) Nature Communications	79
Rosini, *et al*. (2015) PLOS One	18
Seale, *et al*. (2016) Nature Microbiology	333
Teatero, *et al*. (2016) Scientific Reports	92
Other (NCBI)	20

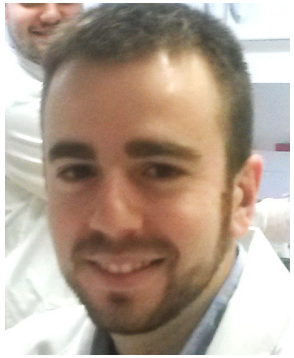
Total 626

Compare to:

CC1 (423 genomes)

CC19 (241 genomes)

CC23 (314 genomes)

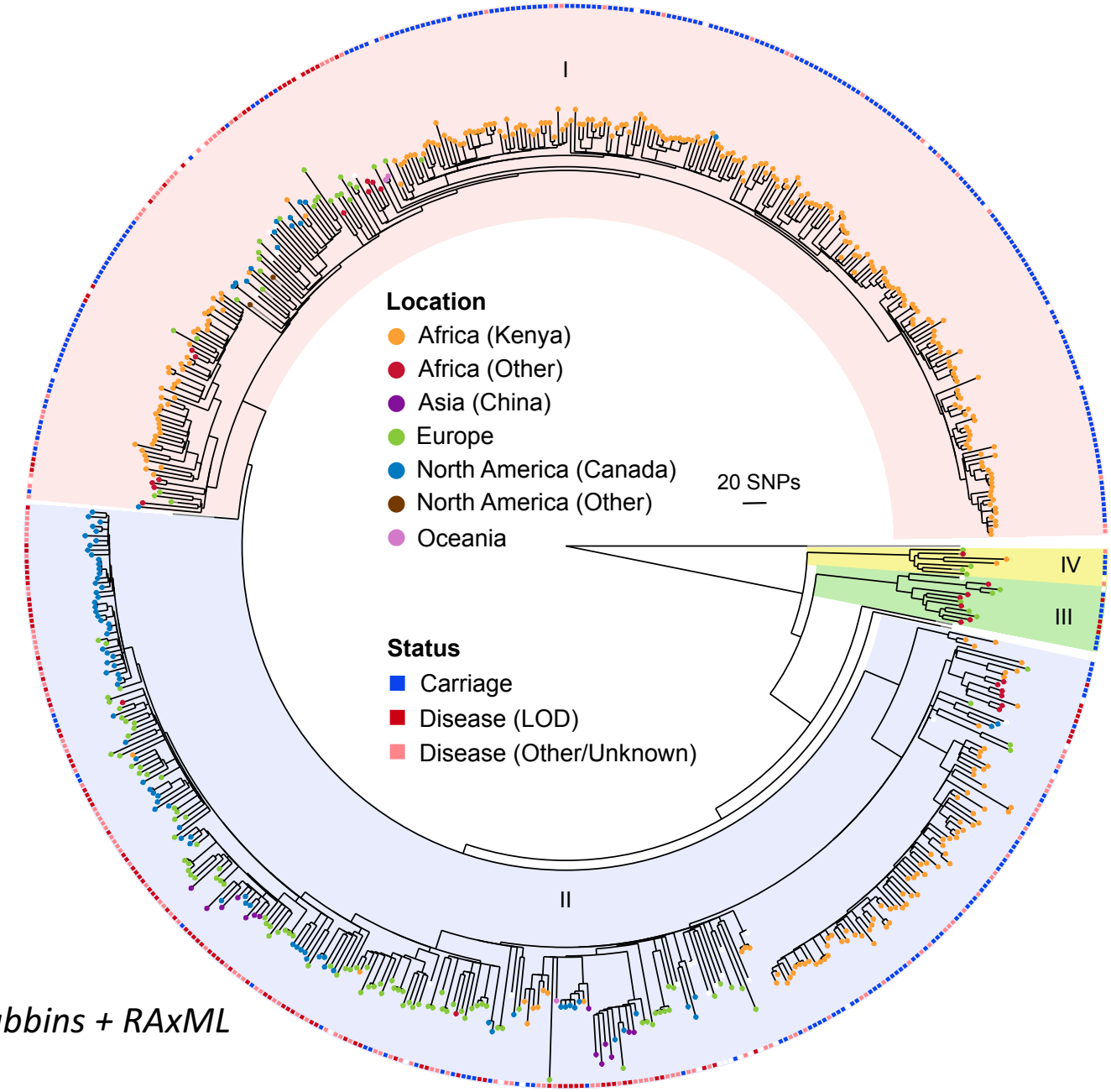


Alexandre Almeida

Core-genome phylogeny of CC17

Reference strain: COH1

12 584 SNPs

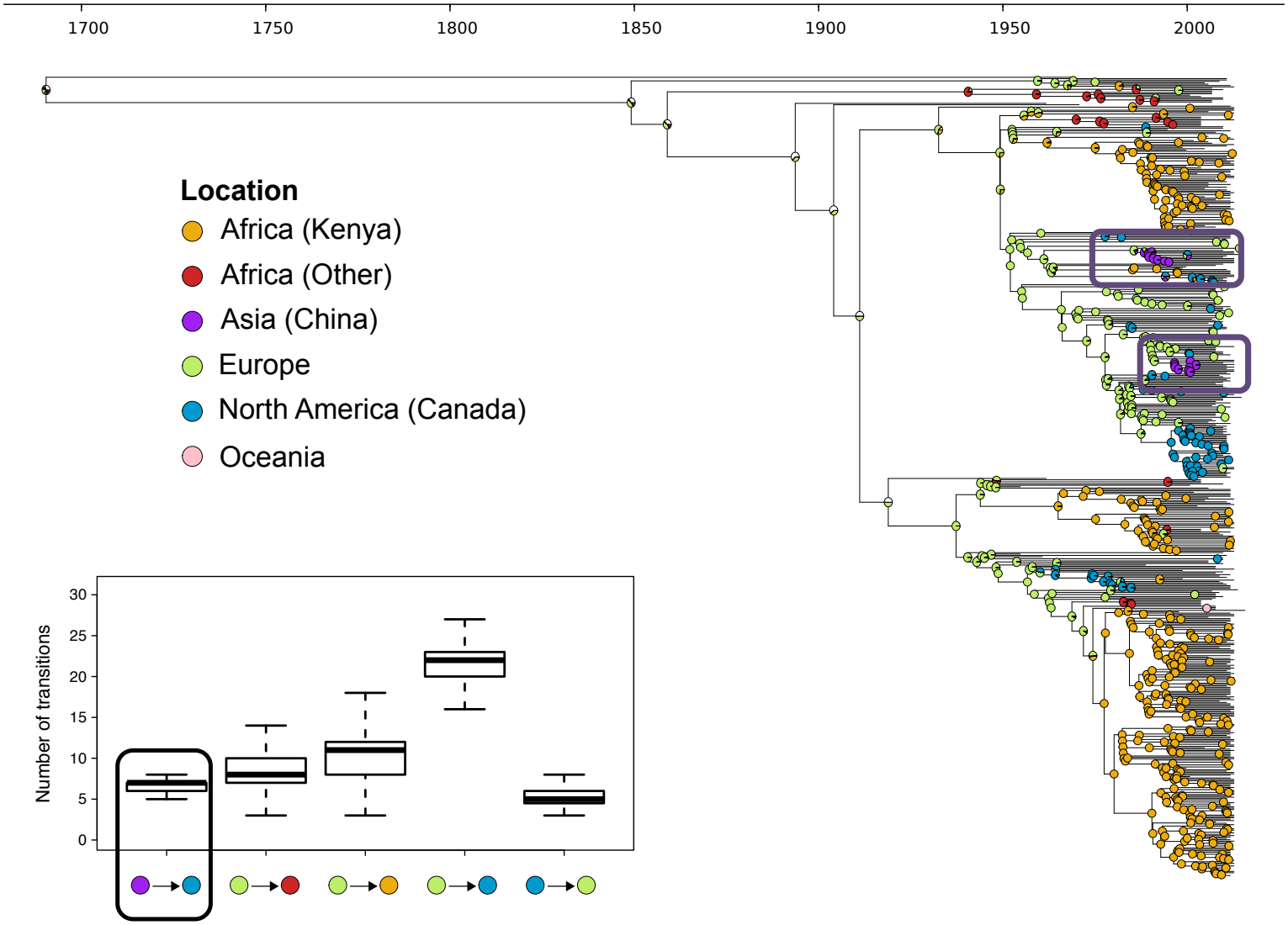


Signs of intercontinental transmission

Disease strains are found in all four main clades

Tools: *BWA + GATK + LS-BSR + Gubbins + RAxML*

Phylogeographic analysis of CC17



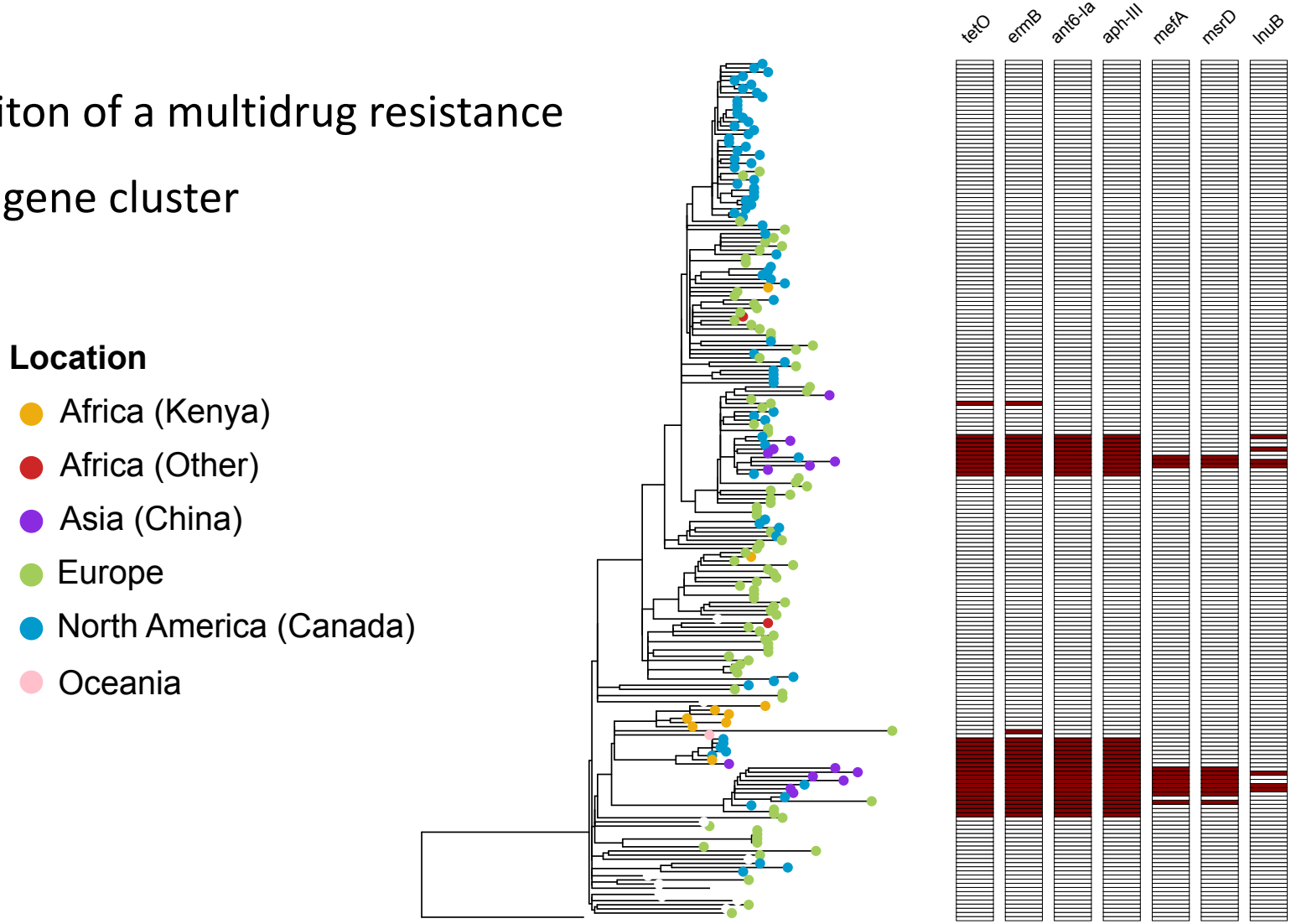
Multiple intercontinental transmission events occurred since the 1950s

Tools: *BEAST + phytools (R package)*

Intercontinental transmission of MDR strains

Acquisiton of a multidrug resistance

(MDR) gene cluster



MDR strains were recently transmitted from China to Canada

Genomic insights into the evolution and pathogenesis of CC17

General evolution of CC17

Mutational bias

dN/dS

Detect mutations under parallel and adaptive evolution

Disease versus carriage

Gene-by-gene mutation frequency

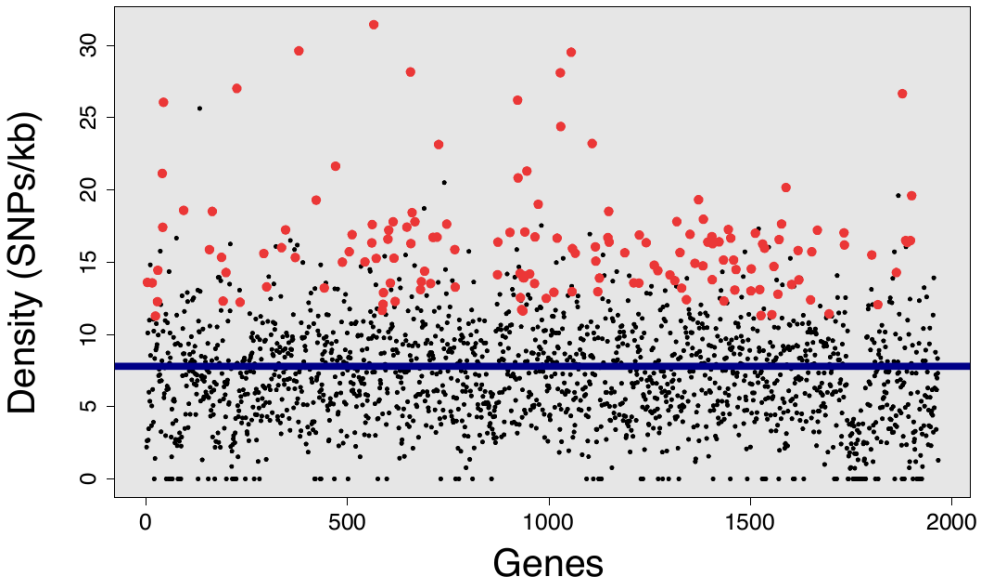
Linear correlation

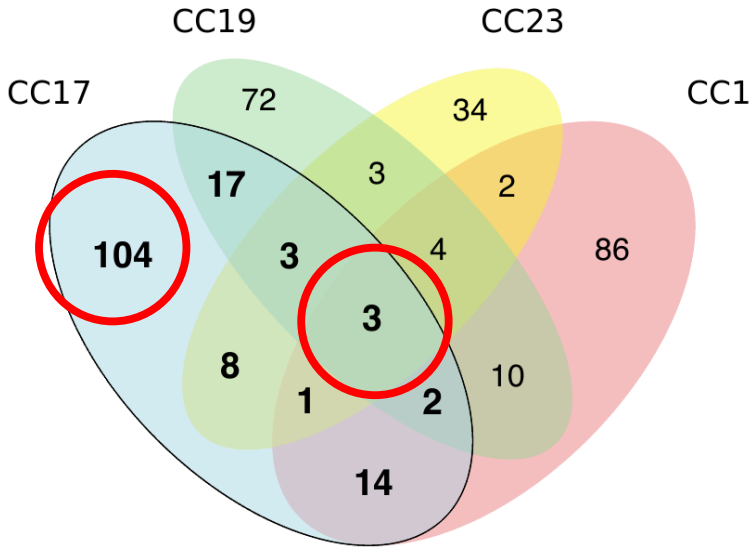
Detect mutations related to disease or carriage

Parallel evolution

152 genes

($P < 0.05$, Poisson test)





Only three genes were significantly mutated in the four different CCs

104 genes (68%) are under parallel evolution solely among CC17 strains

Parallel and adaptive evolution

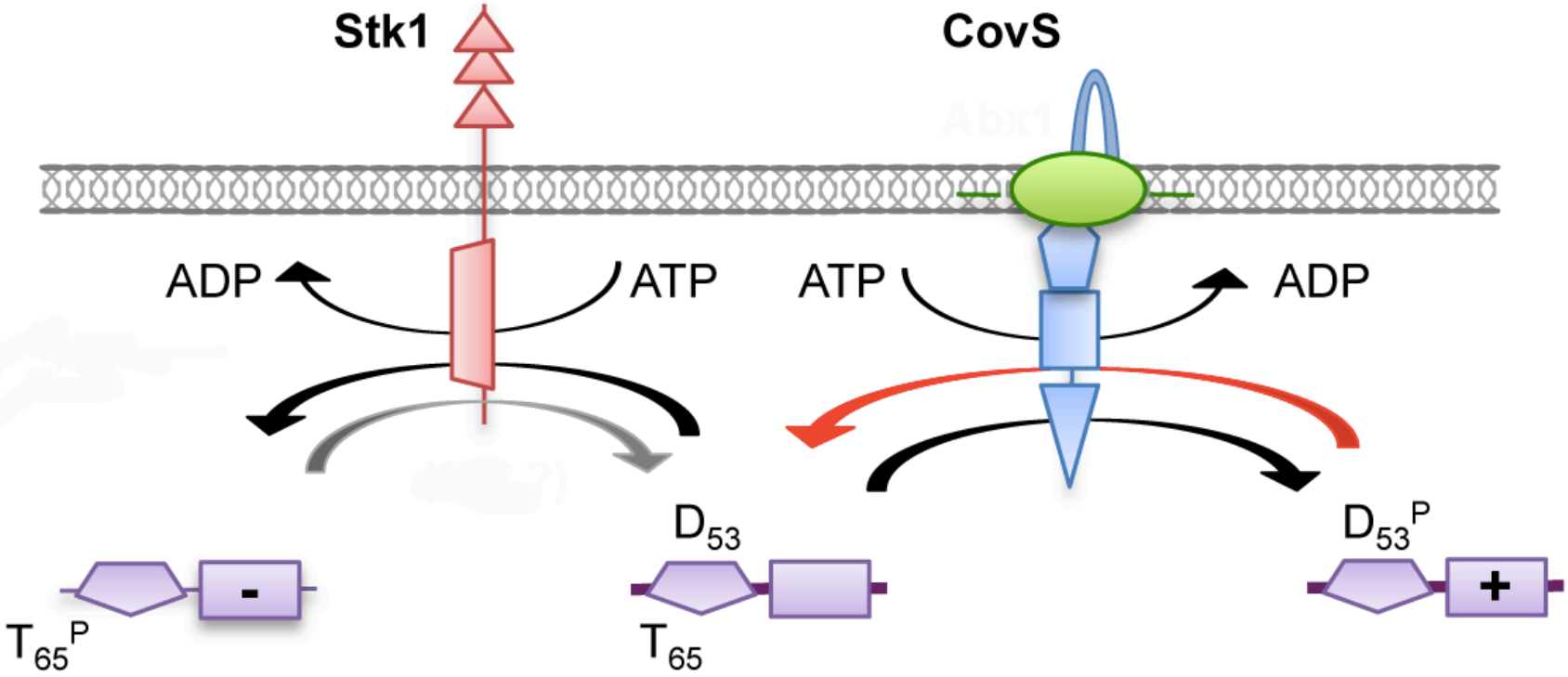
Parallel evolution in virulence-associated genes

	Locus	Product	NS	S	Isolates	CC1	CC19	CC23
regulation	GBSCOH1_RS01850	serine/threonine protein kinase **Stk1**	19	7	39	N	Y	N
	GBSCOH1_RS07705	two-component sensor histidine kinase **CovS**	21	5	30	Y	Y	N
adhesion	GBSCOH1_RS05180	fibronectin-binding protein **FbsA**	9	5	21	N	N	N
	GBSCOH1_RS06600	cell wall anchor **Srr2**	26	29	364			
immune evasion	GBSCOH1_RS05745	beta-1,4-galactosyltransferase **CpsG**	12	2	18	N	N	N
	GBSCOH1_RS05755	galactosyl transferase **CpsE**	16	2	27	Y	N	N
	GBSCOH1_RS05760	tyrosine-protein kinase **CpsD**	10	1	19	N	Y	N
	GBSCOH1_RS03250	**PI-1** class C sortase	12	3	22	N	N	N
	GBSCOH1_RS03255	**PI-1** class C sortase	8	7	41	N	N	N
resistance to CAMPs	GBSCOH1_RS08490	D-alanyl-lipoteichoic acid biosynthesis protein **DltD**	15	2	21	N	N	N

Adaptive evolution (genes under positive selection)

Locus	Product	NS	S	Isolates	CC1	CC19	CC23
GBSCOH1_RS00315	Zoocin A	13	0	18	N	N	N
GBSCOH1_RS03260	**PI-1** ancillary protein 1	20	0	352	N	N	N
GBSCOH1_RS08985	Streptococcal histidine triad protein	19	0	42	N	N	N

Both CovS and Stk1 regulate CovR



CovR inactive | **CovR inactive** | **CovR active**

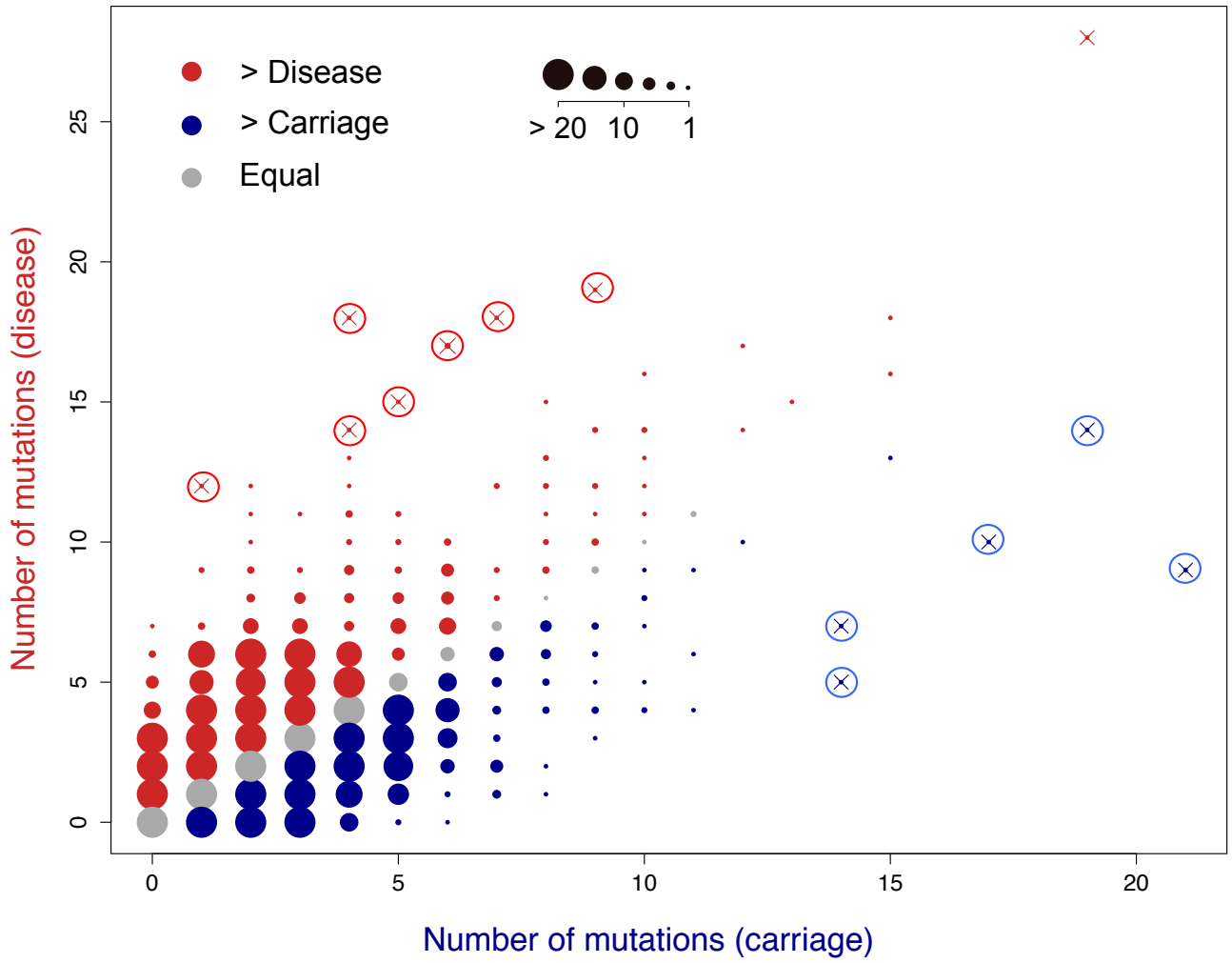
++ Virulence factors | **+ Virulence factors** | **- Virulence factors**

Firon, et al. 2013

Mutations in disease versus carriage

6771 mutations in disease strains (280 isolates)

5818 mutations in carriage strains (306 isolates)



Mutations in 14 genes were significantly skewed towards either carriage or disease

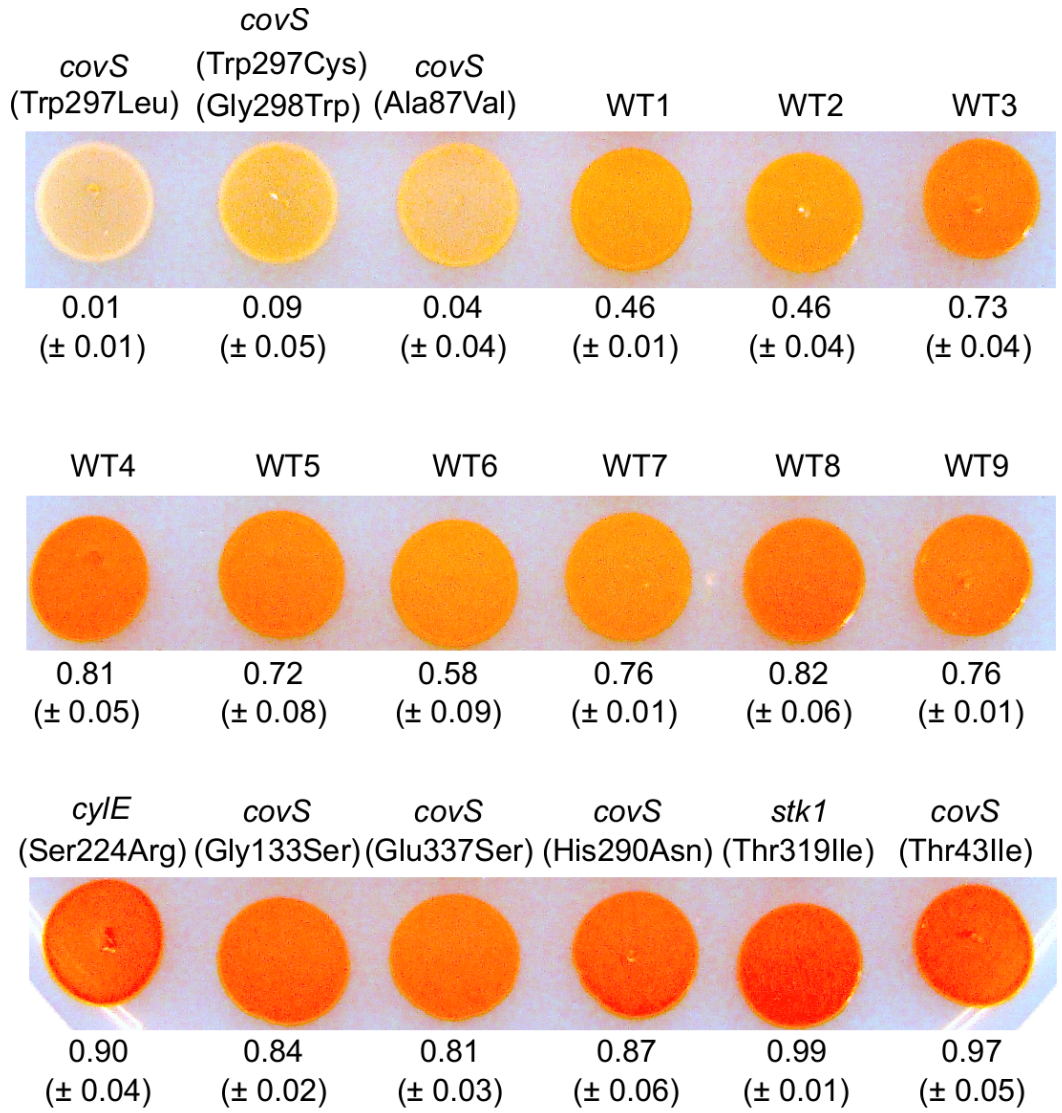
Mutations in disease versus carriage

Genes with most mutations in disease-associated strains

Locus	Product	Disease	Carriage
GBSCOH1_RS00285	Phosphoribosylformylglycinamidine synthase	19	9
GBSCOH1_RS01850	Serine/threonine protein kinase Stk1	17	6
GBSCOH1_RS04925	23S rRNA methyltransferase	12	1
GBSCOH1_RS04975	ABC transporter permease	18	4
GBSCOH1_RS05095	Cell division protein FtsK	28	19
GBSCOH1_RS06600	Cell wall anchor Srr2	18	7
GBSCOH1_RS07650	Amidase	14	4
GBSCOH1_RS07705	Two-component sensor histidine kinase CovS	15	5
GBSCOH1_RS08845	DNA polymerase III subunit alpha	17	6

Mutations affecting CovRS (*stk1* and *covS*) are recurrently selected during disease

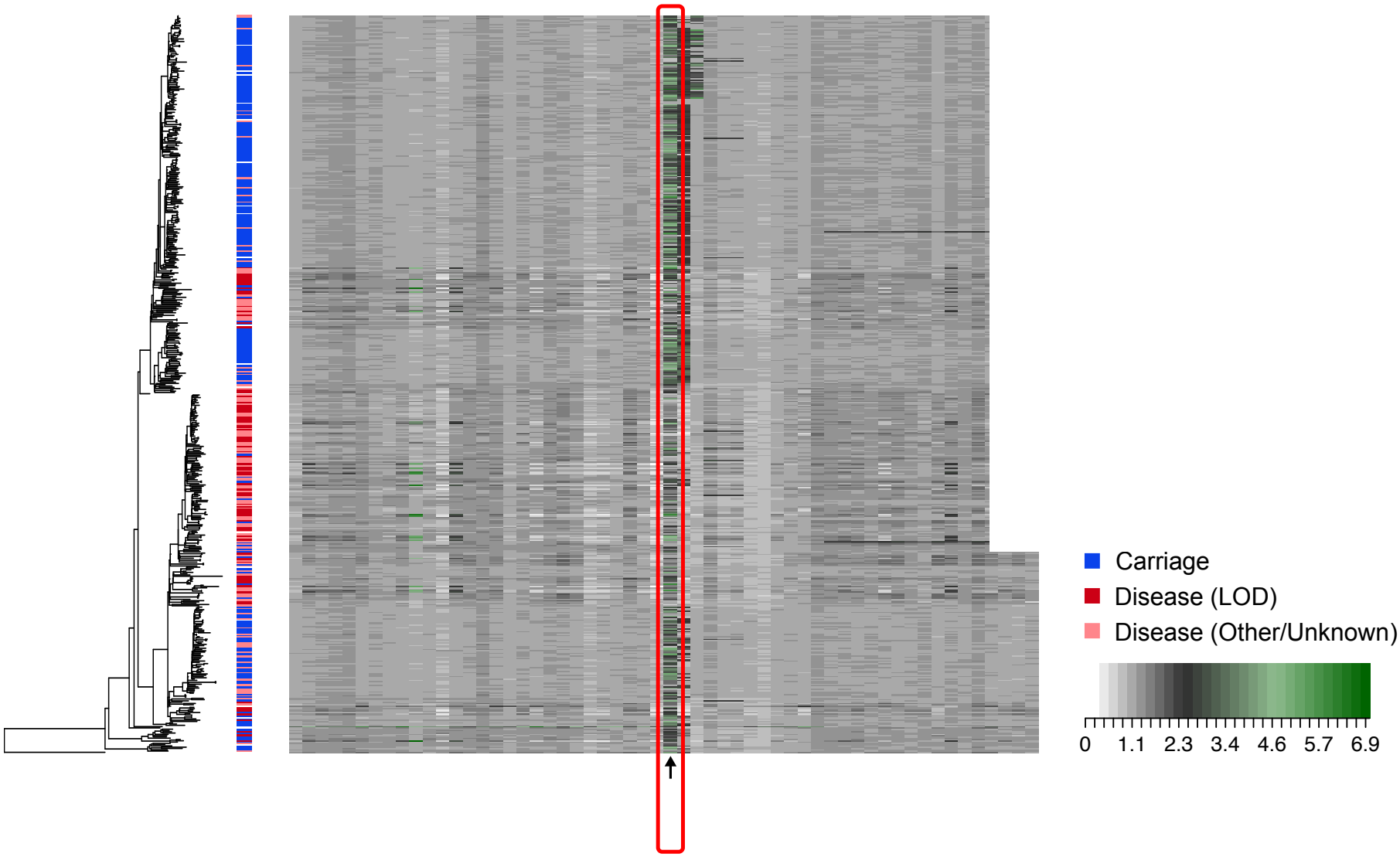
Phenotypic impact of covS mutations



Pigmentation and hemolysis are repressed by CovR

Gene Copy Number Variation (CNV)

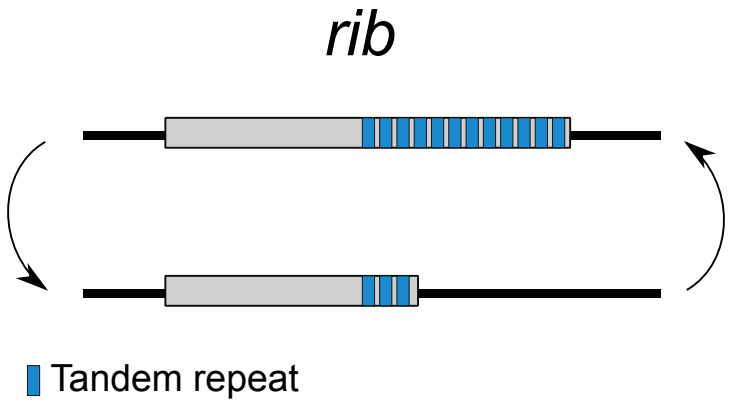
59 functionally relevant genes with the most variation in sequencing coverage



Tools: *CNOGpro (R package)*

Gene Copy Number Variation (CNV)

The gene encoding the **surface protein Rib** displays the most variable sequencing coverage

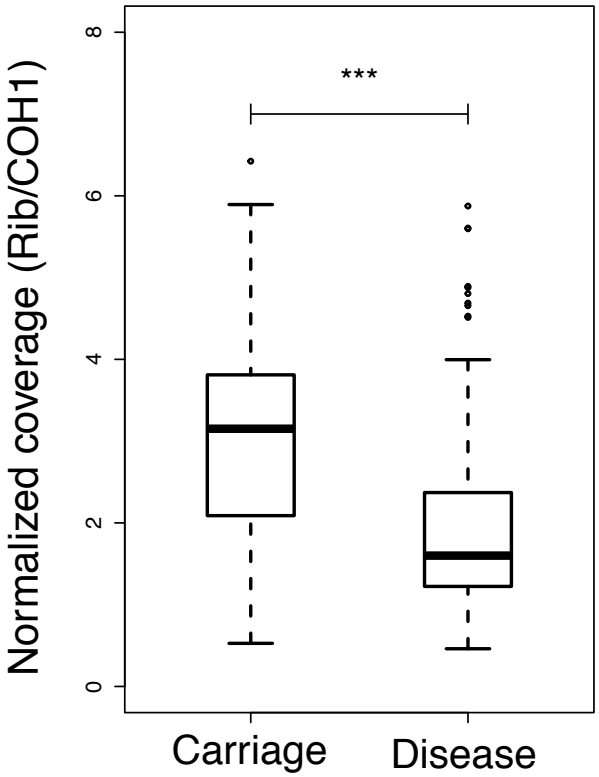


Immunogenicity

High

Low

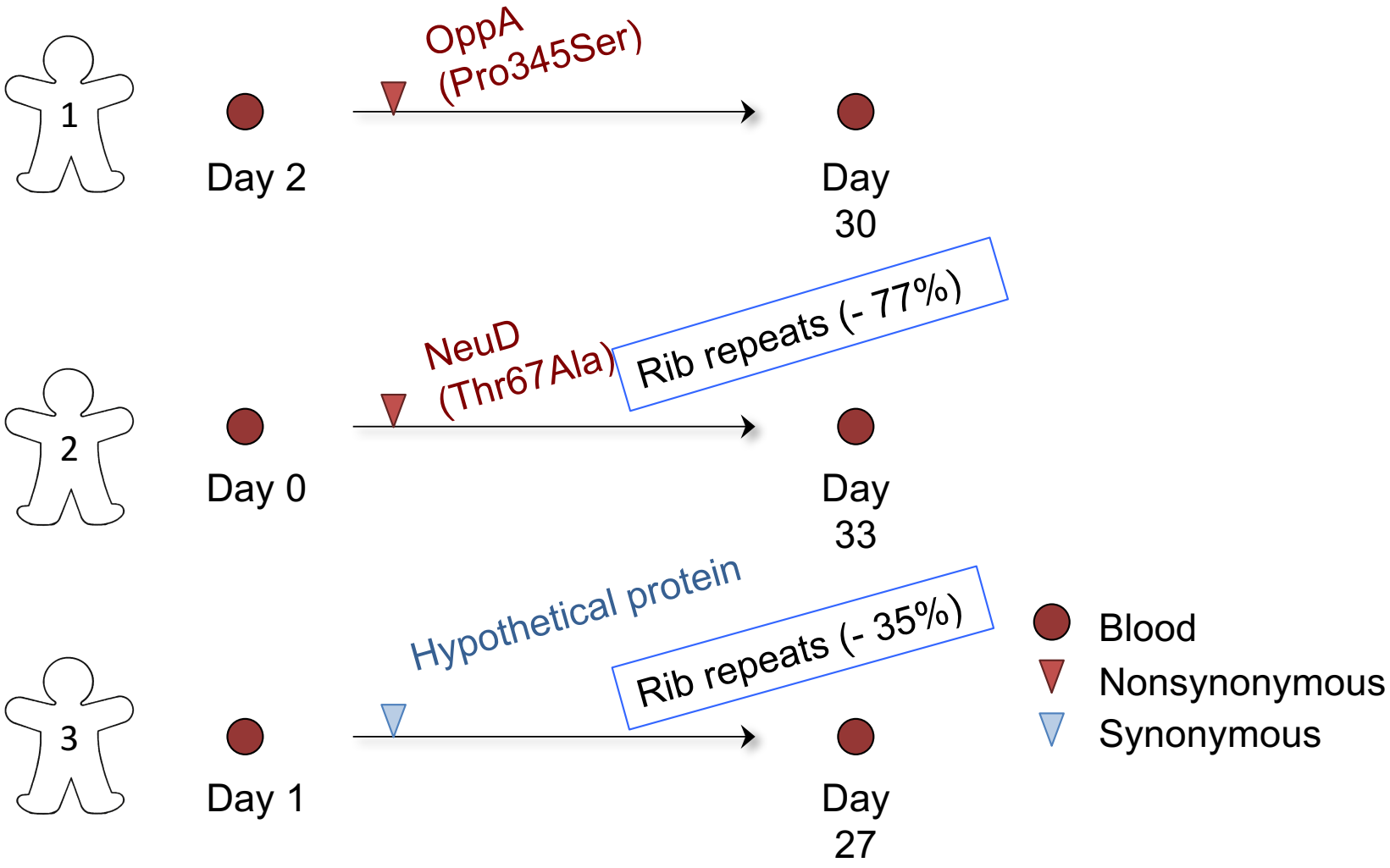
Madoff, et al. 1996



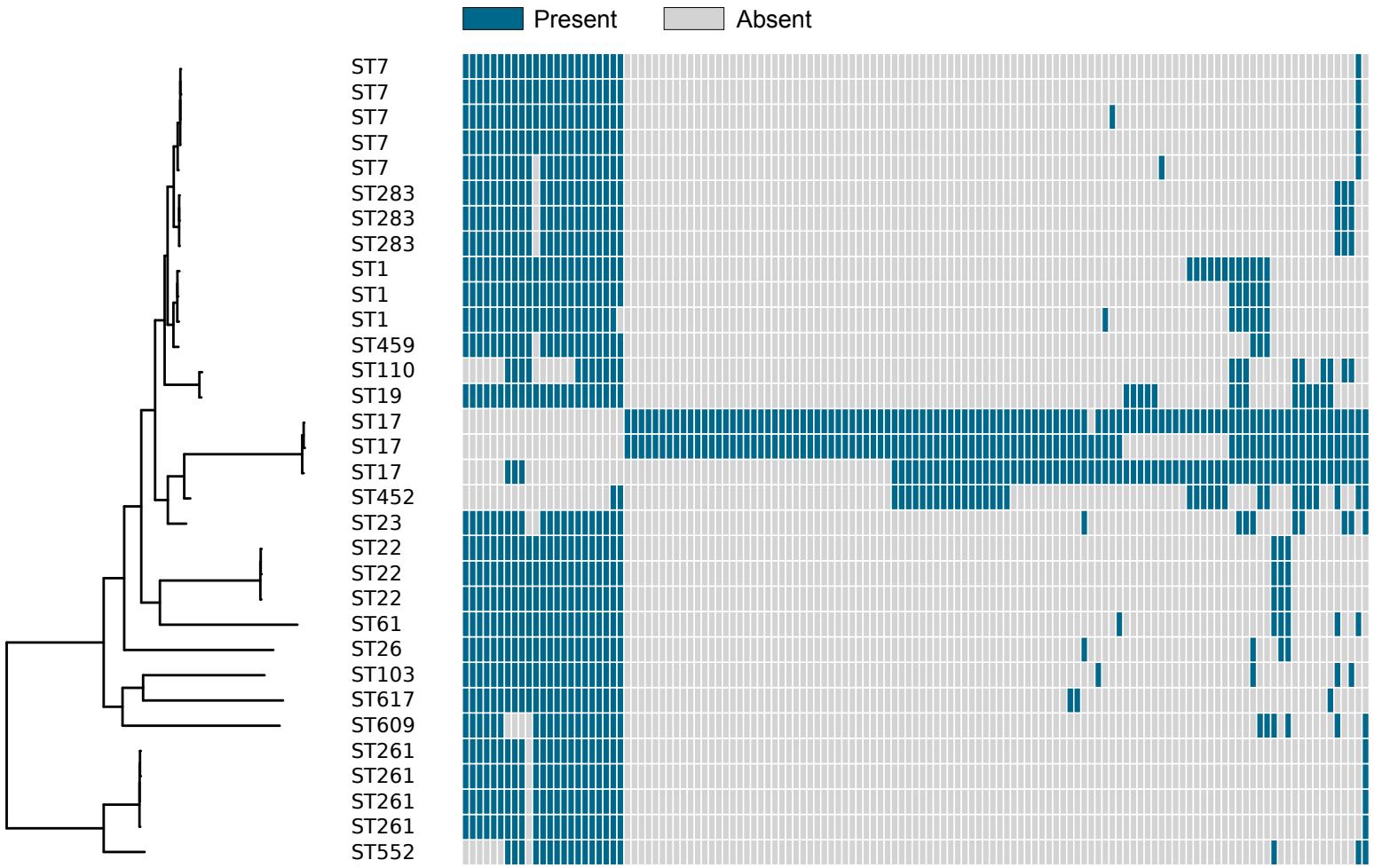
Reduction in the number of repeats in Rib is selected during disease

Tools: *CNOGpro (R package)*

Recurrent GBS infections in three newborns



Genes specific to ST17



Surface proteins: SvgA, Pilus 2, the Rib protein, Srr2

Uncharacterized surface proteins, lipoproteins, permease, kinase

Conclusions

- CC17 strains are frequently **transmitted worldwide**
- Disease and carriage strains display different mutation patterns, and modifications within the **CovRS system** are associated with **disease**
- Pathways related to **metabolism**, **cell adhesion** (FbsA/Srr2), **regulation** (CovRS and Stk1) and **immune evasion** (CPS and Rib) are under evolutionary pressures in the CC17 population
- Candidate genes that have not been previously implicated in virulence are **promising targets for future investigation**

Host specificity and host adaptation

- Broad host spectrum in animals
  - Udder infections in bovines and camels
  - Invasive diseases in fish (*S. difficile*)

Truite Fario

Rosinski-Chupin *et al. BMC Genomics* 2013, **14**:252
http://www.biomedcentral.com/1471-2164/14/252

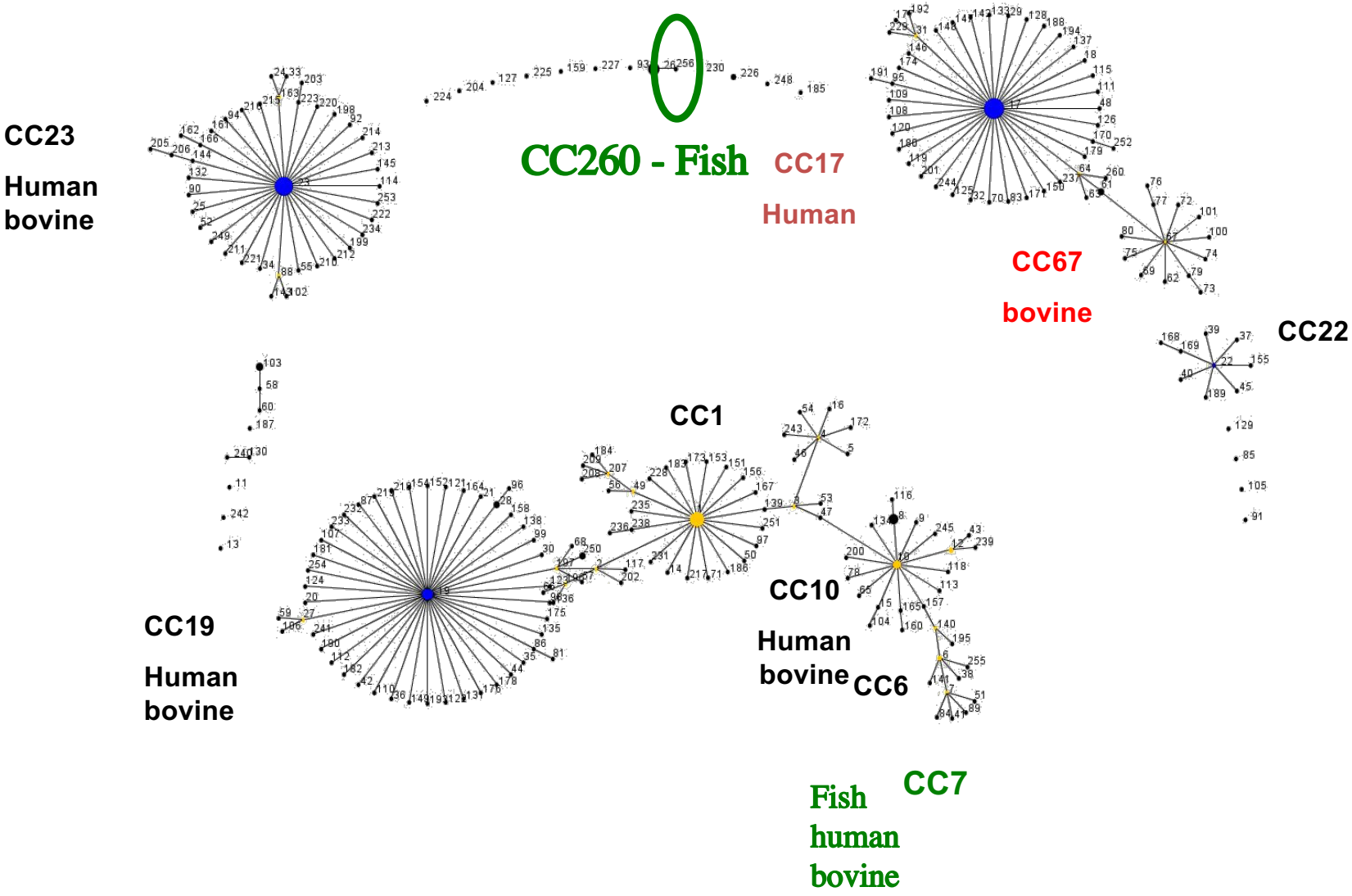
BMC Genomics

RESEARCH ARTICLE Open Access

Reductive evolution in *Streptococcus agalactiae* and the emergence of a host adapted lineage

Isabelle Rosinski-Chupin[1,2*], Elisabeth Sauvage[1,2], Barbara Mairey[3], Sophie Mangenot[3], Laurence Ma[4], Violette Da Cunha[1,2], Christophe Rusniok[2,5], Christiane Bouchier[4], Valérie Barbe[3] and Philippe Glaser[1,2]

Fish isolates group in two main clusters



Sequencing of seven isolates

- 2-22 ST261 *S. difficile* strain, tilapia Israel 1984
- SS1218 ST261 frog, Louisiana
- SS1219 ST260 frog, Brazil 1982
- 90-503 ST260 hybrid striped bass, Louisiana (1990)
- 05-108A ST260 Tilapia Honduras (2005)

- CF01173 ST7 Trout farm England
- SS1014 ST7 Fish USA

The ST7 strains

Nb of SNP

Comparisons with human strains

	A909	H36B	CF01173 UK	SS1014 USA
A909		3835	**389**	3484
H36B			3625	**689**
SS1173				**3514**
SS1014				

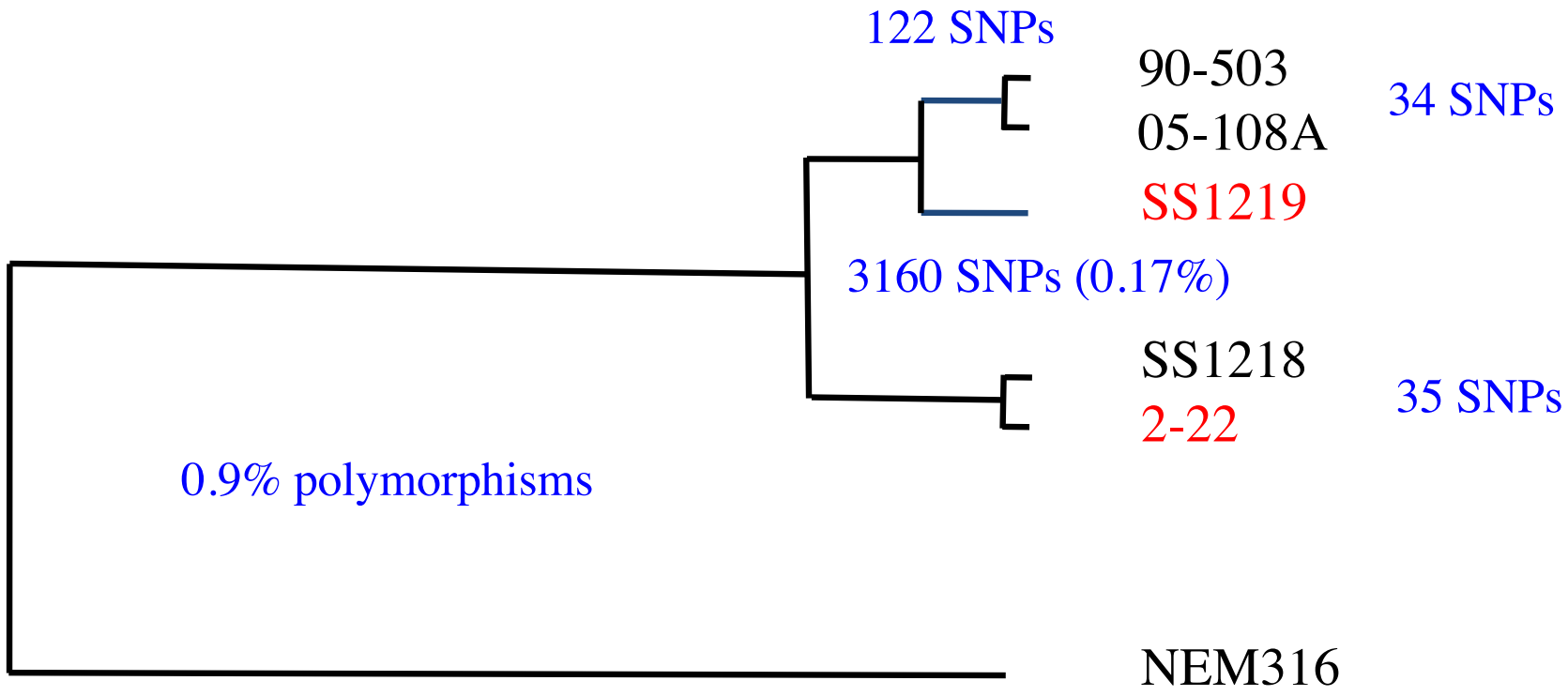
⇒ST7 strains from fish are closely related to human strains

⇒These strains adapted independently to the fish host

CC 260 A specific lineage of fish pathogen

- Also described as *S. difficile*
- Specific phenotypes: thermosensitive, reduce capacity to utilize diverse C-sources, different cell envelop property
- Highly pathogenic and transmissible: cause epidemics in fish farms; LD50 : 100
- Shown as a serotype 1b GBS
- Isolated in different parts of the world: Israel, USA, Brazil (warm water)

Fish strains belong to a distinct lineage compared to human strains

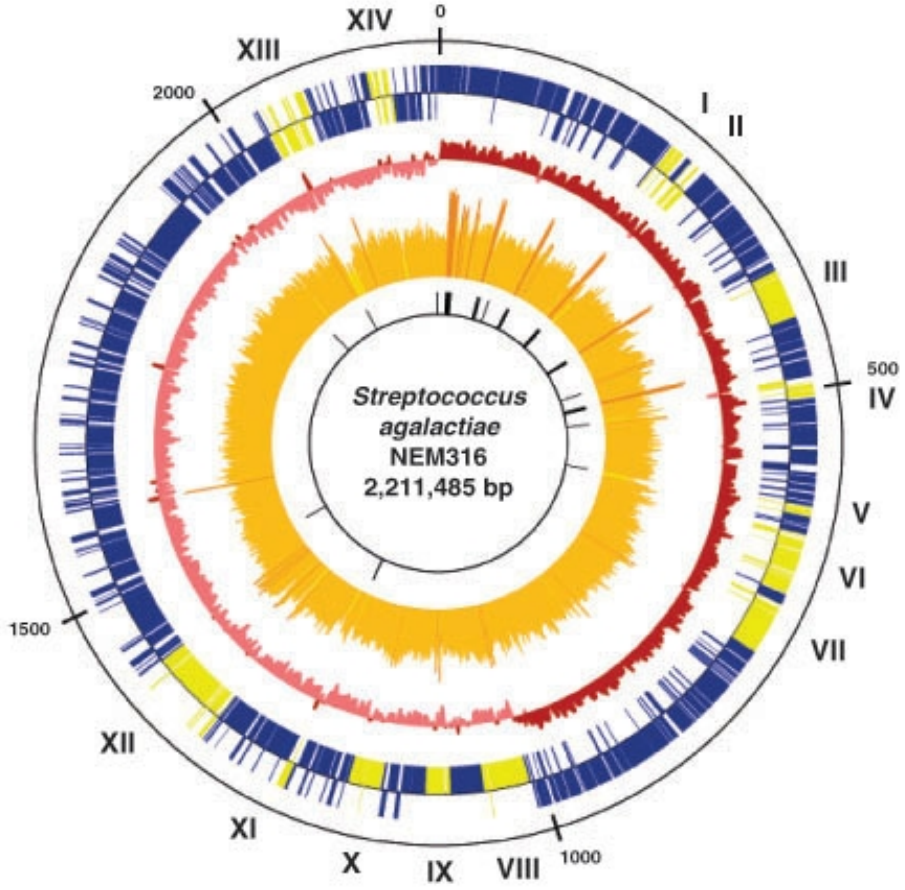


⇒recent spreading of the SS1219 – 90503 cluster

⇒This lineages is anciant

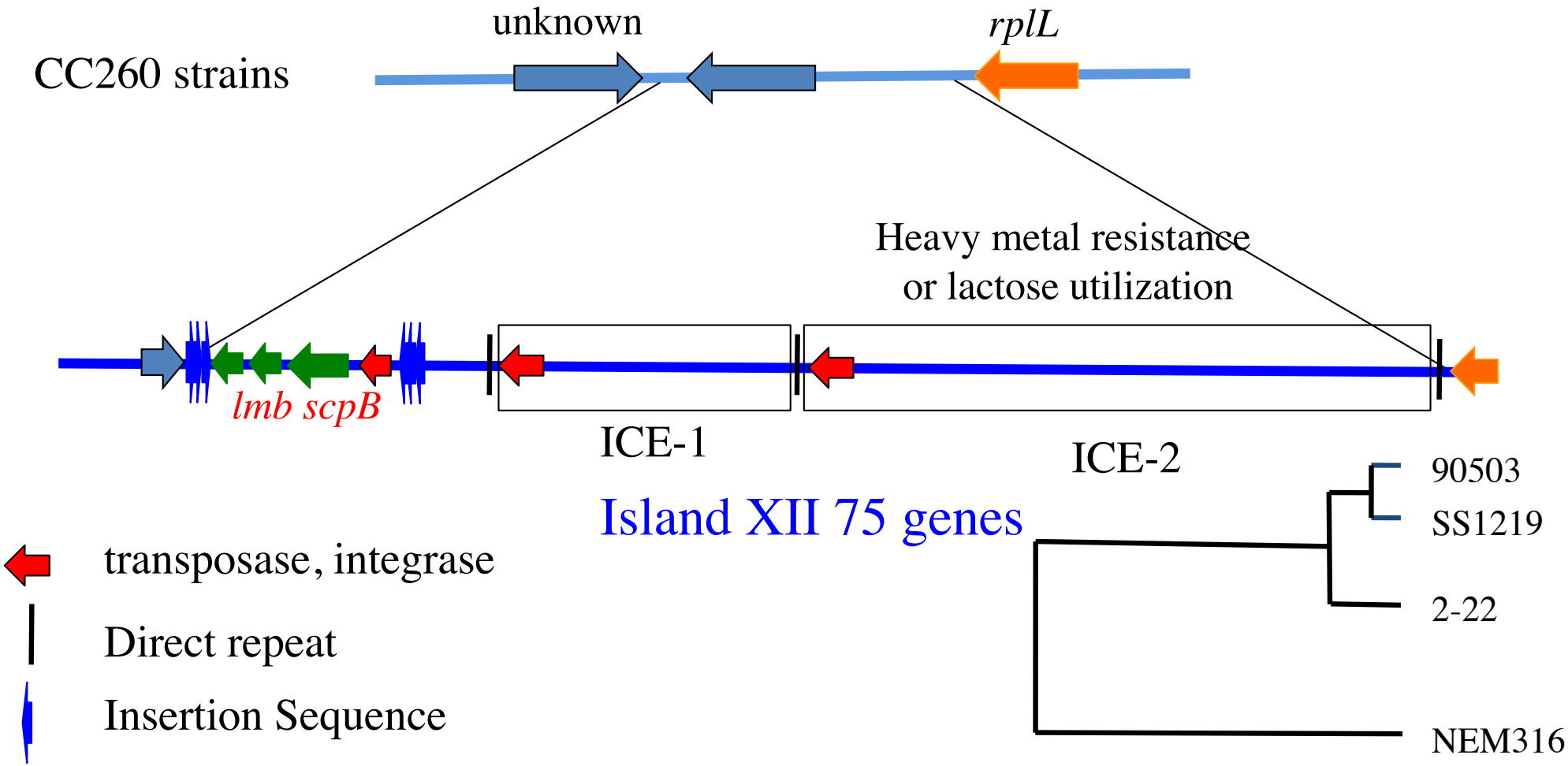
⇒It is distantly related from human and bovine strains

Size: 2-22: 1834 kb - ss1219: 1845 kb

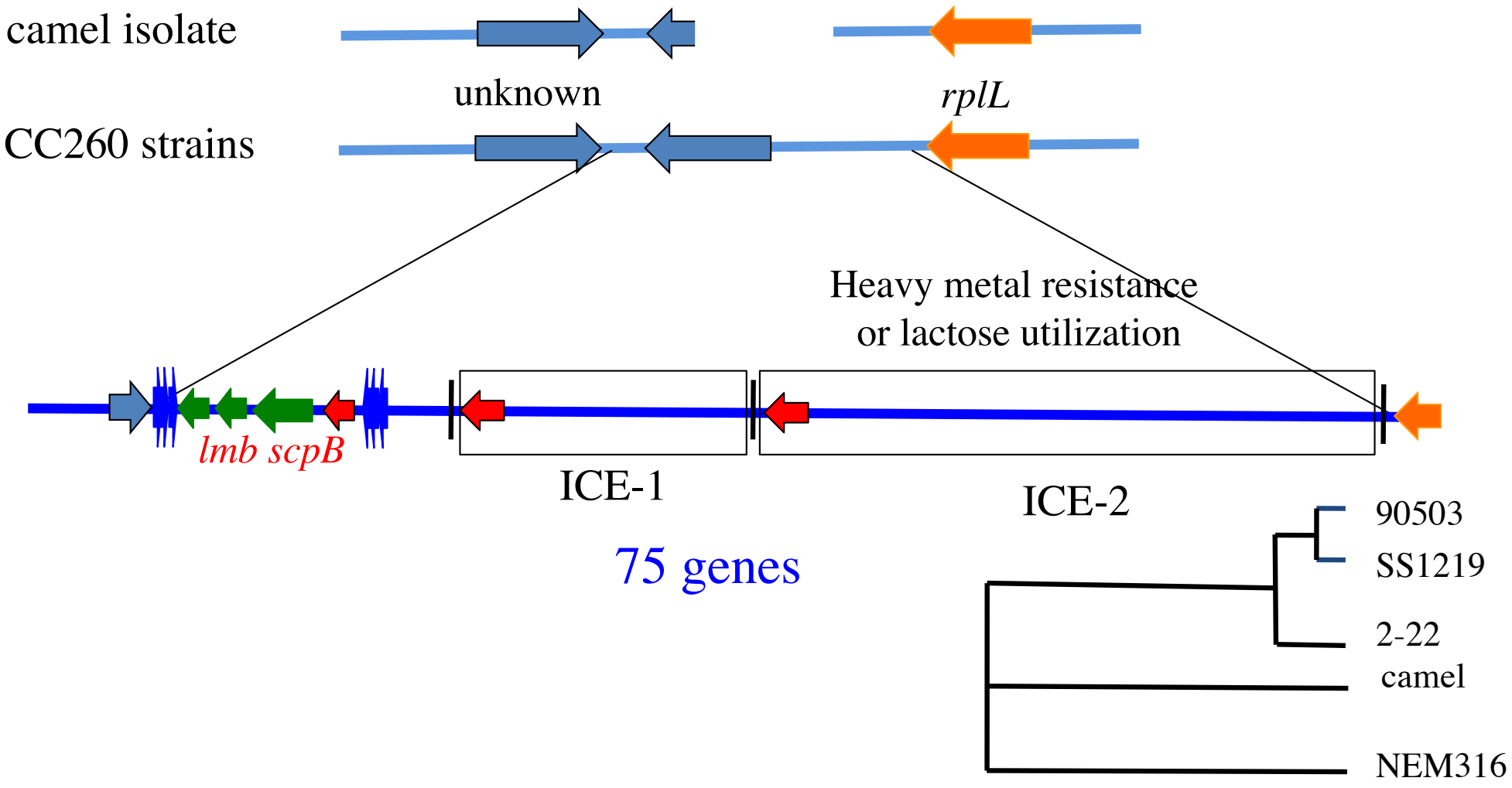


⇒Three types of events: gene loss, pseudogene, in frame deletion

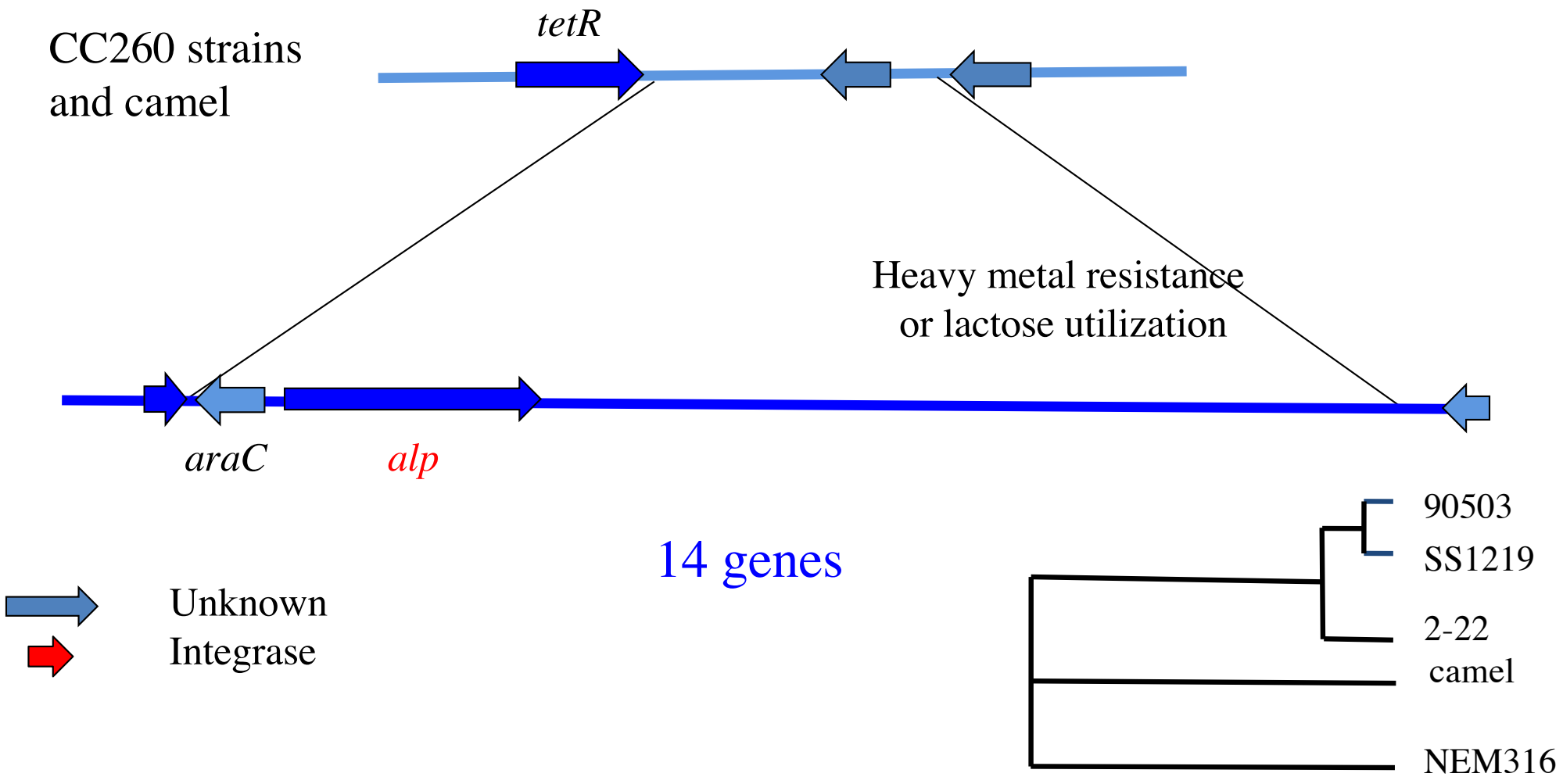
Most islands are missing in the CC260 fish strains
Deletion in CC260 or insertion after divergence?



Island XII likely inserted after the divergence of the CC260 and ST270 (camel) strains



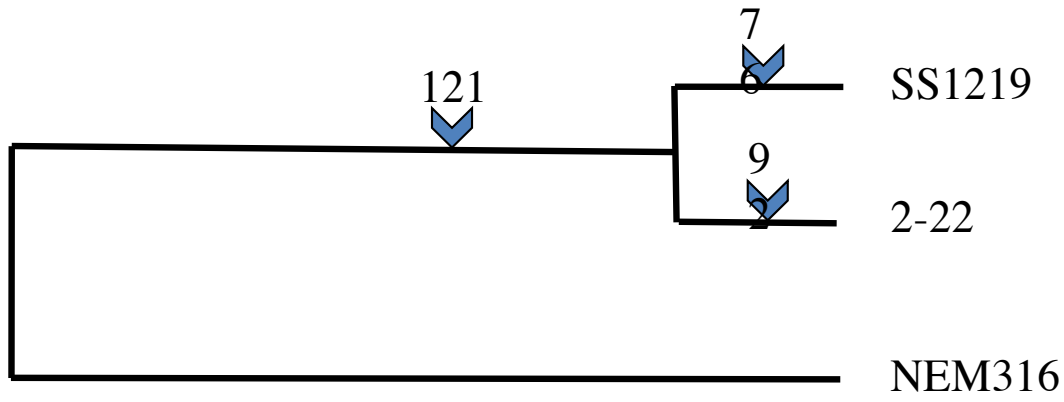
The *alp* virulence gene was inserted after the divergence of the CC260 and ST270 (camel) strains



Conclusion:
Some virulence factors were acquired by the human lineages after the divergence from some animal lineages

Gene decay

- Core genome among human GBS: c.a. 1800 genes
- 213 genes are pseudogenes in 2-22
- 197 genes are pseudogenes in SS1219
- 121 genes are inactivated in both strains



⇒ Affect metabolic functions (C sources), energetic, surface components and regulatory functions

⇒ Genes from the same operon could be inactivated differentially in the two strains

⇒ Ongoing process

Most virulence or fitness factors are missing

- Agglutinin receptor
- Fibronectin-binding proteins fbsA; fbsB
- Lmb, ScpB
- The two pilus loci
- Most gene encoding surface proteins are missing or mutated (30/38); several are shortened
- Cytolysin
- Glucuronyl hydrolase
- Other important functions: *htrA*, *pta*, *camp* etc.

Deletion of repetitions in repeated proteins

serin rich protein

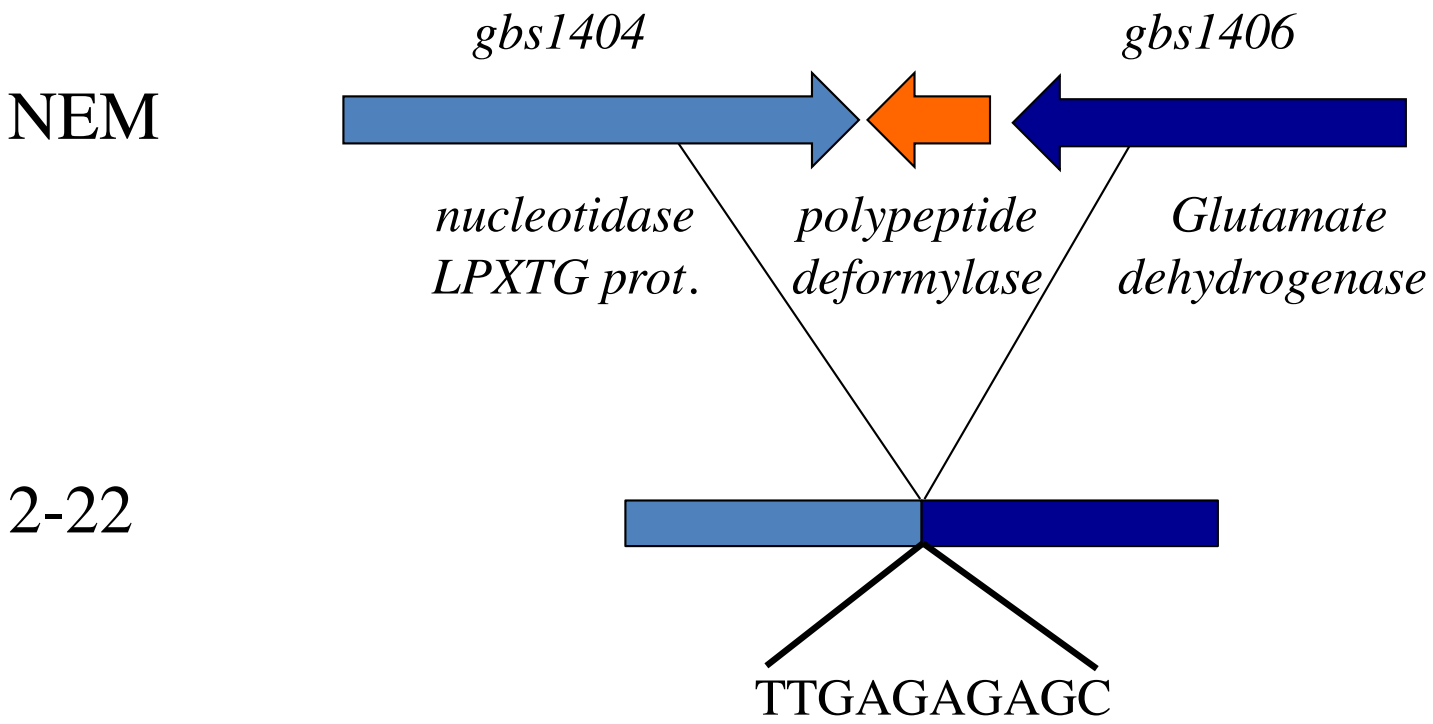
```
2-22  NFLSESASTSASTSASMSASTSASTSASTSASTSASTSSSSSVTSN
NEM   KLLSESASTSASTSA     310 x SX   SASTSA-TSSSSSVTSN
```

Fibrinogen binding protein A

```
2-22  NQSQGNV   LERRQRDVENKSQV          GQLIEKNPLFSK
NEM   NQSQGNV   LERRQRDVENKSQV  17x     GQLIEKNPLFSK
```

Deletion of one rDNA operon

Mechanism of deletions



In frame internal deletions

```
Query: 481     ttctcttgatgggagaa------------------ttggagaacactttgaaagctc
               ||||||||||||||||                    ||||||||||||||||||||||
Sbjct: 2108538 ttctcttgatgggagaaccatctctgatgggagaattggagaacactttgaaagctc


Query:     121 WVQKEEEIVGLQSKEMSLMELEKQDRIHKLL------LMGELENTLKAQYPHLSIAQS
               WVQKEEEIVGLQSKEMSLMELEKQDRIHKLL      LMGELENTLKAQYPHLSIAQS
Sbjct:     121 WVQKEEEIVGLQSKEMSLMELEKQDRIHKLLLMGEPSLMGELENTLKAQYPHLSIAQS
```

⇒about 70 cases in each strain

⇒What is the functional impact ?

Ongoing evolution of these genes

```
Query: 1      MNEIKCPHCGTAFAINESEYHQLLEQIRGDAFDKEVSERLEKERLILGEQAKNQLQEVVV 60
              MNEIKCPHCGTAFAINESEYHQLLEQIRGDAFDKEVSERLEKERLILGEQAKNQLQEVVV
Sbjct: 43355 MNEIKCPHCGTAFAINESEYHQLLEQIRGDAFDKEVSERLEKERLILGEQAKNQLQEVVV 43176

Query: 61     EKDKEIAKLQYKVKQFLIEKDNLLKDNEYQLAEQLNQKDMMLRDLENQIDRLRLEHENSL 120
              EKD                   NLLKDNEYQLAEQLNQKDMMLRDLENQIDRLRLEHENSL
Sbjct: 43175 EKD------------------NLLKDNEYQLAEQLNQKDMMLRDLENQIDRLRLEHENSL 43050

Query: 121    QEALTKVERERDAIQNQLHIQEKEKDLALASVKSDYEVQLKAANEQVEFYKNFKAQQSTK 180
              QEALTKVERERDAIQNQL IQEKEKDLA  SVKSDYEVQLKAANEQVEFYKNFKAQQSTK
Sbjct: 43049 QEALTKVERERDAIQNQLLIQEKEKDLA--SVKSDYEVQLKAANEQVEFYKNFKAQQSTK 42876

Query: 181    AVGESLEHYAETEFNKVRHLAFPNAYFEKDNTLSSRGSKGDFIYREKDENDLEFLSIMFE 240
              AVGESLEHYAETEFNKVRHLAFPNAYFEKDNTLSSRGSKGDF+YREKD+NDLEFLSIMFE
Sbjct: 42875 AVGESLEHYAETEFNKVRHLAFPNAYFEKDNTLSSRGSKGDFVYREKDKNDLEFLSIMFE 42696

Query: 241    MKNESDDTIKKHKNEDFFKELDKDRREKSCEYAVLVTMLEADNDYYNTGIVDVSHKYPKM 300
              MKNESDDTIKKHKNEDFFKELDKDRREKSCEYAVLVTMLEADNDYYNTGIVDV+HKYP
Sbjct: 42695 MKNESDDTIKKHKNEDFFKELDKDRREKSCEYAVLVTMLEADNDYYNTGIVDVNHKYPNA 42516

Query: 301    YVIRPQFFIQLIGILRNAALNTLKYKQELALMKEQNIDITHFEEDLDIFKNAFAKNYNSA 360
              +                                                 AKNYNSA
Sbjct: 42515 F-------------------------------------------------AKNYNSA 42492

Query: 361    SKNFQKAIDEIDKSIKRMEAVKAALTTSENQLRLANNKLDDVSVKKLTRKNPTMKAKFDA 420
              SKNFQKAIDEIDKSIKRMEAVKAALTTSENQLRLANNKLDDVSVKKLTRKNPTMKAKFDA
Sbjct: 42491 SKNFQKAIDEIDKSIKRMEAVKAALTTSENQLRLANNKLDDVSVKKLTRKNPTMKAKFDA 42312

Query: 421    LKD 423
              LKD
Sbjct: 42311 LKD 42303
```

NEM316 /SS1219

```
MNEIKCPHCGTAFAINESEYHQLLEQIRGDAFDKEVSERLEKERLILGEQAKNQLQEVVV 60
MNEIKCPHCGTAFAINESEYHQLLEQIRGDAFDKEVSERLEKERLILGEQAKNQLQEVVV
MNEIKCPHCGTAFAINESEYHQLLEQIRGDAFDKEVSERLEKERLILGEQAKNQLQEVVV 574456

EKDKEIAKLQYKVKQFLIEKDNLLKDNEYQLAEQLNQKDMMLRDLENQIDRLRLEHENSL 120
EKDKEIAKLQYKVKQFLIEKDNLLKDNEYQLAEQLNQKDMMLRDLENQIDRLRLEHENSL
EKDKEIAKLQYKVKQFLIEKDNLLKDNEYQLAEQLNQKDMMLRDLENQIDRLRLEHENSL 574276

QEALTKVERERDAIQNQLHIQEKEKDLALASVKSDYEVQLKAANEQVEFYKNFKAQQSTK 180
QEALTKVERERDAIQNQL IQEKEKDLA  SVKSDYEVQLKAANEQVEFYKNFKAQQSTK
QEALTKVERERDAIQNQLLIQEKEKDLA--SVKSDYEVQLKAANEQVEFYKNFKAQQSTK 574102

AVGESLEHYAETEFNKVRHLAFPNAYFEKDNTLSSRGSKGDFIYREKDENDLEFLSIMFE 240
AVGESLEHYAETEFNKVRHLAFPNAYFEKDNTLSSRGSKGDF+YREKD+NDLEFLSIMFE
AVGESLEHYAETEFNKVRHLAFPNAYFEKDNTLSSRGSKGDFVYREKDKNDLEFLSIMFE 573922

MKNESDDTIKKHKNEDFFKELDKDRREKSCEYAVLVTMLEADNDYYNTGIVDVSHKYPKM 300
MKNESDDTIKKHKNEDFFKELDKDRREKSCEYAVLVTMLEADNDYYNTGIVDVSHKYP
MKNESDDTIKKHKNEDFFKELDKDRREKSCEYAVLVTMLEADNDYYNTGIVDVSHKYP-- 573748

YVIRPQFFIQLIGILRNAALNTLKYKQELALMKEQNIDITHFEEDLDIFKNAFAKNYNSA 360
                                                 NAFAKNYNSA
-------------------------------------------------NAFAKNYNSA 573718

SKNFQKAIDEIDKSIKRMEAVKAALTTSENQLRLANNKLDDVSVKKLTRKNPTMKAKFDA 420
SKNFQKAIDEIDKSIKRMEAVKAALTTSENQLRLANNKLDDVSVKKLTRKNPTMKAKFDA
SKNFQKAIDEIDKSIKRMEAVKAALTTSENQLRLANNKLDDVSVKKLTRKNPTMKAKFDA 573538

LKD 423
LKD
LKD 573529
```

NEM316 / 2-22

summary

- Highly virulent in a specific host
- Probably not a commensal ?
- Loss of most virulence or fitness factors
- No specific virulence factor
- The capsule is the major virulence factor
- Phenotypic isolation (Ts), host adaptation is irreversible