



Swiss Federal Institute of Technology Zurich

Seminar for  
Statistics

Departement of Mathematics

---

Master Thesis

Spring 2015

---

Nicolas Bennett

The title of my thesis  
which should be split on  
several lines if it is too long

---

Submission Date: August 25th 2015

---

Co-Adviser: Anna Drewek  
Adviser: Prof. Dr. Peter Bühlmann



---

[B]ecause we tend to reward others when they do well and punish them when they do badly, and because there is regression to the mean, it is part of the human condition that we are statistically punished for rewarding others and rewarded for punishing them — [Kahneman \(2012\)](#).



---

## Preface

---

First words and acknowledgements. And we add a lot of text to make sure that it spans more than one line, as otherwise it may not show up.



---

## **Abstract**

---

This example thesis briefly shows the main features of our thesis style, and how to use it for your purposes.



---

# Contents

---

Preface	iii
Abstract	v
Contents	vii
List of Figures	ix
List of Tables	xi
List of Abbreviations and Notation	xiii
<b>1 Introduction</b>	<b>1</b>
<b>2 Biological Background</b>	<b>7</b>
2.1 Microbial Host-Cell Infection . . . . .	7
2.1.1 Viral Infection Mechanisms . . . . .	8
2.1.2 Bacterial Entry Mechanisms . . . . .	11
2.1.3 Actin-based Intracellular Motility . . . . .	13
2.2 Select Bacterial Pathogens . . . . .	14
2.2.1 <i>Bartonella Henselae</i> . . . . .	14
2.2.2 <i>Brucella Abortus</i> . . . . .	16
2.2.3 <i>Listeria Monocytogenes</i> . . . . .	18
2.2.4 <i>Salmonella Typhimurium</i> . . . . .	21
2.2.5 <i>Shigella Flexneri</i> . . . . .	24
2.3 Select Viral Pathogens . . . . .	27
2.3.1 Adenoviruses . . . . .	27
2.3.2 Rhinoviruses . . . . .	29
2.3.3 Vaccinia . . . . .	31

## CONTENTS

---

2.4 RNA Interference . . . . .	35
2.4.1 Molecular Mechanism . . . . .	36
2.4.2 Biological Function . . . . .	38
2.4.3 Applications . . . . .	39
<b>3 Mathematical Background</b>	<b>45</b>
<b>4 Data</b>	<b>47</b>
<b>A R Package singleCellFeatures</b>	<b>49</b>
<b>B Yet another appendix....</b>	<b>53</b>
B.1 Description . . . . .	53
B.2 Tables . . . . .	53
<b>Bibliography</b>	<b>55</b>
<b>Epilogue</b>	<b>58</b>

---

## List of Figures

---

1.1	Relative frequencies of death causes in 2012 by World Bank income groups . . . . .	3
1.2	Relative frequencies deadly infectious diseases for 2012 by World Bank income groups . . . . .	3
2.1	A generalized view of the steps necessary for viral–cellular membrane fusion . . . . .	9
2.2	Zipper and trigger mechanisms for bacterial host-cell entry . . . . .	12
2.3	Replication cycle of <i>Vaccinia viruses</i> for both intracellular mature and extracellular enveloped virions. . . . .	15
2.4	Replication cycle of <i>Vaccinia viruses</i> for both intracellular mature and extracellular enveloped virions. . . . .	17
2.5	Capsid structure and genome of rhinoviruses. . . . .	20
2.6	Capsid structure and genome of rhinoviruses. . . . .	22
2.7	Capsid structure and genome of rhinoviruses. . . . .	25
2.8	Capsid structure and genome of rhinoviruses. . . . .	28
2.9	Capsid structure and genome of rhinoviruses. . . . .	31
2.10	Replication cycle of <i>Vaccinia viruses</i> for both intracellular mature and extracellular enveloped virions. . . . .	34
2.11	The three major pathways of RNA interference. . . . .	36
4.1	Decision tree for adenovirus infection scoring. . . . .	47
4.2	Decision tree for adenovirus infection scoring. . . . .	48
A.1	Visualization of cell colony edge detection by 2D binning. . . . .	51



---

## **List of Tables**

---

2.1	The Baltimore classification scheme for viruses . . . . .	10
2.2	A selection of RNAi based drugs in clinical trials. . . . .	42
B.1	Test results . . . . .	53



---

## Abbreviations and Notation

---

**AAV** adeno-associated virus.

**Ago** Argonaute.

**AIDS** acquired immune deficiency syndrome.

**AMR** antimicrobial resistance.

**Arp2/3** actin-related-protein 2 and 3.

**ATR** acid tolerance response.

**CAP** community-acquired pneumonia.

**CAR** coxsackievirus and adenovirus receptor.

**CCP** clathrin-coated pit.

**Cdc42** cell division control protein 42 homolog.

**CDHR3** cadherin-related family member 3.

**COPII** coat protein complex II, involved in anterograde endoplasmic reticulum (ER)–Golgi transport.

**CSD** Cat Scratch Disease.

**DBP** DNA-binding protein.

**DGCR8** DiGeorge syndrome critical region gene 8.

**ER** endoplasmic reticulum.

**ERAD** endoplasmic-reticulum-associated protein degradation.

**ERGIC** ER-Golgi intermediate compartment.

**EV** extracellular virion.

---

**LIST OF ABBREVIATIONS AND NOTATION**

---

- GDP** guanosine diphosphate.
- GEF** guanine nucleotide exchange factors.
- GFP** green fluorescent protein.
- GTP** guanosine triphosphate.
- HDT** host directed therapeutics.
- HIV** human immunodeficiency virus.
- HTS** high throughput screening.
- I(1,3,4,5,6)P<sub>5</sub>** inositol 1,3,4,5,6-pentakisphosphate.
- ICAM-1** intercellular adhesion molecule 1.
- Ics** *Shigella* intracellular spread protein.
- ICU** intensive care unit.
- Ipa** *Shigella* invasion plasmid antigen.
- LOF** loss-of-function.
- LPS** lipopolysaccharide.
- M cells** microfold cells.
- MID** Argonaute middle domain.
- miRNA** microRNA.
- mRNA** messenger RNA.
- MV** mature virion.
- N-WASP** neural Wiskott-Aldrich syndrome protein.
- NLS** nuclear localization signal.
- NPC** nuclear pore complex.
- OTE** off-target effects.
- PACT** protein activator of protein kinase PKR.
- PAZ** PIWI-Argonaute-Zwille domain.
- PI(4,5)P<sub>2</sub>** phosphatidylinositol 4,5-bisphosphate.
- PI(5)P** phosphatidylinositol 5-phosphate.
- PIWI** P-element-induced whipmy testes domain.
- PKR** RNA-dependent protein kinase pathway.
- pre-microRNA (miRNA)** precursor miRNA.
- pri-miRNA** primary miRNA.

- pTP** precursor terminal protein.
- Rac1** Ras-related C3 botulinum toxin substrate 1.
- RDRC** RNA-dependent RNA polymerase complex.
- RhoG** Ras homology growth-related.
- RISC** RNA-induced silencing complex.
- RITS** RNA-induced transcriptional silencing complex.
- SARS** severe acute respiratory syndrome.
- SCV** *Salmonella*-containing vacuole.
- shRNA** short hairpin RNA.
- SIF** *Salmonella*-induced filaments.
- Sip** *Salmonella* invasion protein.
- siRNA** short interfering RNA.
- Sop** *Salmonella* outer protein.
- SptP** secreted effector protein.
- T3SS** type III secretion system.
- T4SS** type IV secretion system.
- TIR** toll-interleukin-1 receptor.
- Tir** translocated intimin receptor.
- TLR** toll-like receptor.
- TRBP** TRA RNA-binding protein.
- VEGF** vascular endothelial growth factor.
- VLDLR** very low-density lipoprotein receptor.
- WHO** World Health Organization.



## Chapter 1

---

### Introduction

---

Infectious diseases have played an undeniably important role in human history. With human populations becoming sufficiently aggregated to sustain direct life cycle viral and bacterial infectants around 2000 BC, devastating invasions of a growing number of pathogens started to occur ([Dobson and Carper, 1996](#)).

One of the earliest well documented incidence of a large-scale epidemic is known as the Plague of Athens. Starting in 430 BC and lasting roughly three years, a highly infectious disease killed 75'000 to 100'000 people or 25% of Athen's population. This catastrophic event is attributed either to smallpox, a viral infection with *Variola major* or typhus, caused by *Rickettsia* bacteria ([Littman, 2009](#)).

The bacterium *Yersinia pestis* caused three major plague pandemics in the early and late middle ages, as well as in the late 19th century. Originating in northern Africa in 523 AD and spreading around the Mediterranean basin throughout the years 541–546, the Plague of Justinian is assumed to have killed up to half of the population of affected areas. The effect on cities was disproportionately severe. In Constantinople, for example, an estimated 230'000 people out of 375'000 lost their lives to the disease ([Treadgold, 1997](#)). Returning in the years 1347–1351, known today as the Black Death, a plague pandemic again wiped out around half of Europe's population. Death toll estimates range from 15 to 23.5 million ([Zietz and Dunkelberg, 2004](#)). Leaving behind a grim cultural heritage, this catastrophe had a lasting effect on economic and social structures in Europe. The third large-scale outbreak started around 1855 in southern China and quickly spread to Japan, Taiwan and India again wreaking havoc on the affected population.

Bringing diseases such as smallpox, measles (an infection with the *Measles virus*) and typhus to the Americas during the European invasion of the New World had grave repercussions for the indigenous population, carrying no nat-

## 1. INTRODUCTION

ural resistance towards the newly introduced pathogens. It is estimated that the population of present day Mexico fell from 20 million to 1.6 million over the course of the 16th century due to multiple disease epidemics, critically contributing to the successful colonization of the new continents ([Dobson and Carper, 1996](#)).

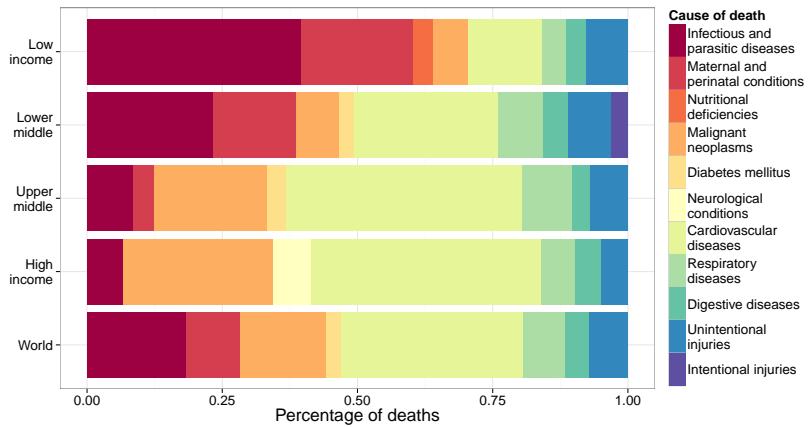
Cholera and influenza are further contagious diseases with high mortality rates, responsible for global epidemics. *Vibrio cholerae*, a bacterium which causes infections of the intestine, became widespread in the early 19th century and caused seven pandemics since, the last of which only started in 1961. Antibacterial treatment of sewage and purification of drinking water greatly help to prevent and contain spreading of the disease but in areas with inadequate sanitation, such as Haiti after the 2010 earthquake, it remains a pathogen difficult to control. The influenza virus causes seasonal epidemics characterized by low lethality rates among people with intact immune systems<sup>1</sup>. Irregularly occurring influenza pandemics, initiated by zoonosis of new virus strains, against which no natural immunity exists, however, are accompanied by much higher lethality rates. The most significant such event is known today as the Spanish flu pandemic of 1918, costing the lives of 50–100 million, nearly half of which were young, healthy adults ([Taubenberger and Morens, 2006](#)).

In addition to diseases plaguing humanity for centuries, new ones continually emerge. Human immunodeficiency virus (HIV) is believed to have transferred from non-human primates in the early 20th century and the recent outbreaks of severe acute respiratory syndrome (SARS) and swine flu serve as vivid illustrations of such occurrences.

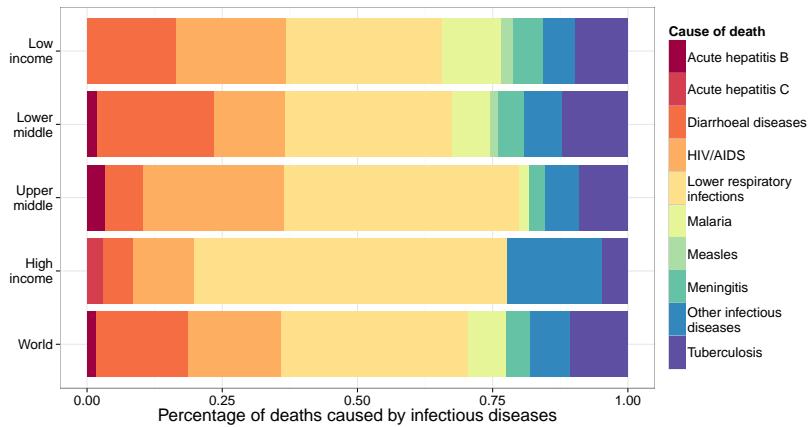
Despite development of means to treat and prevent many previously devastating diseases, infectious pathogens remain a serious threat to global health. In 2012, an estimated total of 58.3 million people died (20.1% in high, 29.4% in upper-middle, 36.5% in lower-middle and 14% in low income countries). Figure 1.1 partitions the total death count into World Bank income groups and causes. In low income countries, infective diseases are the most prevalent cause of death (39.6%), followed by maternal and perinatal complications with substantial margin (20.8%). In lower middle income countries, cardiovascular conditions catch up (26.5%), but are still almost matched in frequency by infectious diseases (23.3%). In upper middle (8.5%) and high income countries (6.7%), the importance of infectious disease while weakened remains accountable for a significant number of deaths. Globally, infectious diseases are the second most frequent cause of death (18.3%), even more prevalent than all forms of cancer combined (15.8%) and only preceded by cardiovascular diseases (33.7%).

Focusing only on deaths caused by infectious disease, lower respiratory in-

<sup>1</sup>In spite of low lethality, these seasonal epidemics still incur significant economic damages. The [World Health Organization \(2003\)](#) estimates annual health care costs and loss of productivity due to influenza at US \$71–176 billion for the United States of America alone.



**Figure 1.1:** Relative frequencies of death causes in 2012 by World Bank income groups. Binning is based on Gross National Income (GNI) per capita and the thresholds are \$1045 or less for low income, \$1046 to \$4125 for lower-middle, \$4126 to \$12745 for upper-middle and \$12746 or more for high income economies. The data was obtained from the [World Health Organization \(2012\)](#).



**Figure 1.2:** Relative frequencies of deadly infectious diseases for 2012 by World Bank income groups. Binning is based on Gross National Income (GNI; see figure 1.1). The data was obtained from the [World Health Organization \(2012\)](#).

## 1. INTRODUCTION

---

fections are most frequent (for each income region individually, low to high: 28.7%, 30.8%, 43.5% and 57.7% as well as worldwide: 34.5%; cf. figure 1.2). Diarrhoeal diseases and HIV/AIDS are the next most common worldwide (16.9% and 17.3%, respectively) where diarrhea is more prevalent in lower income regions (16.6% and 21.4% versus 7% and 5.6%), while HIV/AIDS plays a major role irrespective of income region (low to high: 20.4%, 13.3%, 26.2% and 11.3%).

Dealing with highly virulent pathogens and preventing their spreading requires a multi-pronged approach. First and foremost, etiology and routes of transmission have to be understood. Knowledge of vectors and natural reservoirs is of great importance as a first line of defense. In the case of plague, for example, insecticides killing fleas were successfully used as a prophylactic measure, as well as controlling rat populations. Sanitary precautions including purification of drinking water, cooking foods well and the usage of disinfectants prevent initial infection, while measures such as sewage treatment, hand washing and wearing face masks help limiting spread among humans. Vaccination is the most important preventive measure. Exposing the immune system to a foreign antigen in a controlled manner artificially induces immunity. Among the great successes of widespread vaccination efforts is the global eradication of smallpox through a coordinated initiative lead by the World Health Organization in the 1970's.

Post-infection therapies include symptomatic treatments, as well as anti-infective drugs. Anti-bacterial or anti-viral agents exploit differences in proteomes between host and pathogen to selectively disable the invader with minimal toxicity to the host. This approach has been tremendously successful throughout most of the 20th century, leading to widespread application and prompting development of resistance towards the commonly used compounds. Adding to the severity of the problem is a lack of discovery of new drugs. No new class of anti-bacterial agents has been found since 1987, causing big pharmaceutical companies to withdraw from the area. The remaining research is mainly focused on improving on existing drugs, leading to a weak product pipeline, especially for the treatment of gram-negative bacteria (Silver, 2011). In its first global study on anti-microbial resistance, the [World Health Organization \(2014\)](#) notes:

Antimicrobial resistance (AMR) within a wide range of infectious agents is a growing public health threat of broad concern . . . A post-antibiotic era—in which common infections and minor injuries can kill—far from being an apocalyptic fantasy, is instead a very real possibility for the 21st century.

An alternative to pathogen directed search lies in targeting the set of host proteins necessary for infection. Many intracellular parasites subvert cellular functions to gain entry via host-mediated processes such as endocytosis. Upon entry, they move to a suitable niche and rely on host resources for proliferation. Challenges include evading host-cell defense mechanisms, generating sufficient space for replication, nutrient acquisition and keeping the host alive

---

as long as possible, most of which require complex interactions between invader and host-based mechanisms. Finally, exiting the host cell again requires the parasite to successfully insert itself into existing signaling pathways ([Leirão et al., 2004](#)).

Host directed therapeutics (HDT) offer an escape from the conundrum of wanting to kill but not wanting to select for the surviving microbial parasites. The major challenge under this regimen is finding infectome components that are nonessential for cell survival, as orthogonality of host and infectant can no longer be exploited. Proving feasibility of the approach, [Czyż et al.](#) screened a library of 640 compounds already approved by the United States Food and Drug Administration (USFDA or FDA) for inducing resistance to four intracellular pathogens (*Coxiella burnetii*, *Legionella pneumophila*, *Brucella abortus*, and *Rickettsia conorii*). They found multiple drugs, not classified as antibiotics, that successfully inhibited intracellular bacterial growth while entailing only limited toxicity to host cells (THP-1). [Prussia et al.](#) review the usage of genome-wide screens to study host-pathogen interactions (for HIV and influenza) which in turn serve as basis for rational identification of drug targets for novel host-directed antivirals.

A detailed understanding of the human infectome is of crucial importance to the development of HDT and may even benefit the development of new antimicrobial agents. Feasibility of systematic loss of function screens using RNA interference methodology offers a unique opportunity to investigate complex cellular networks, making this an ideal tool for laying groundwork in combating infectious diseases. Of great importance, however, is ensuring reproducibility and comparability of such datasets, as well as ready availability to the scientific community.



## Chapter 2

---

# Biological Background

---

In order to better understand infectious diseases from a cell biological standpoint, this chapter reviews the current state of knowledge surrounding both bacterial and viral entry mechanisms. A sweeping overview of epidemiology and pathogenesis for several specific bacterial (*Bartonella henselae*, *Brucella abortus*, *Listeria monocytogenes*, *Salmonella enterica* and *Shigella flexneri*), as well as viral parasites (adenoviruses, rhinoviruses and *Vaccinia virus*) is given and the chapter concludes with a look at RNA interference as this mechanism is a cornerstone of genome-wide knockdown experiments.

### 2.1 Microbial Host-Cell Infection

Multi-layered keratinized skin is impenetrable for almost all microbial parasites. Instead they either require breaches such as cuts, scratches, puncture wounds and arthropod bites, or environmental interfaces which offer less impervious protection. Examples include respiratory, gastrointestinal and urogenital tracts, which all contain segments where only a single layer of epithelial cells has to be overcome. Although often protected by chemical defense mechanisms (acidity of the stomach and urogenital tract, as well as microbicidal factors in mucous secretions in the respiratory tract and small intestine), combined with frequent flushing (urination, peristalsis and the coordinated beating of cilia), some microbes have adapted to survive these hostile environments.

For extracellular pathogens to successfully colonize epithelial linings, they must avoid being removed by cleansing mechanisms of the host. Many bacteria accomplish this by expressing adhesins, protein complexes that recognize and bind to specific host-cell receptors, providing host and tissue tropism. Bacterial pili serve to extend reach and penetrate mucous secretions and therefore often carry adhesins. Enteropathogenic *Escherichia coli* have extended this scheme by injecting their own receptor protein Tir (translocated intimin receptor) through

## 2. BIOLOGICAL BACKGROUND

the T3SS (type III secretion system) into the host cell to which it then attaches. This has the additional advantage that the intracellular domain of Tir can be used to modify host cell behavior ([Alberts et al., 2008](#)).

The outside of many epithelial barriers is covered in natural bacterial flora and crossing over into sterile cavities has the advantage of not having to compete with organisms well accustomed to that particular niche. Furthermore, intracellular pathogens are no longer accessible to antibodies and phagocytic cells and have a nutrient rich environment at their disposal. Mechanisms for entering host cells are described in the following sections.

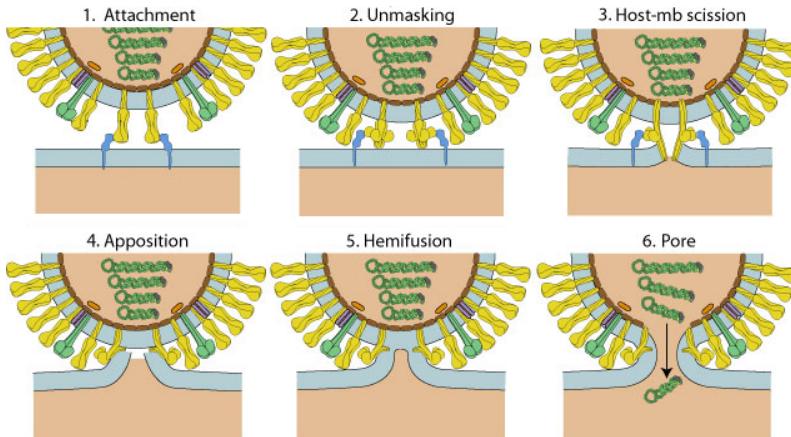
### **2.1.1 Viral Infection Mechanisms**

The first step of any viral entering sequence is binding to the target surface. This can be mediated by attachment factors which simply serve to concentrate the virions on the cell membrane or by virus receptors, which additionally act as communicators between host and pathogen. Common attachment factors include glycosaminoglycan chains and sialic acids and are comparatively unspecific. Glycoprotein spikes on enveloped and capsid proteins of non-enveloped viruses provide host specificity by binding cellular receptors. Mostly these cellular receptors serve other purposes which are exploited for infection. Binding affinity for individual interactions may be weak but aggregation of multiple interactions provide virtually irreversible avidity ([Smith, 2012](#)).

**Viral import.** For viral cell entry, different strategies exist. Enveloped viruses can either directly fuse with the plasma membrane (e.g. HIV) or be endocytosed by the host cell (e.g. influenza), while non-enveloped viruses either create a pore and directly inject their genome into the cytosol (e.g. polio virus) or are endocytosed (e.g. adenovirus). Endocytosis has major advantages over alternative strategies. Reaching its replicatory niche within the host cell is a difficult task for a microorganism having no means of locomotion and hijacking the endocytic system solves this problem elegantly. Furthermore, maturation of endosomes provides precise environmental cues to the invader for triggering uncoating and release. Both fusion with the cell membrane and injection of viral material into the cytosol leaves back traces of infection to be detected by the immune system. Being completely engulfed by the host, however, the intruders leave back no telltale traces. Additionally, lytic membrane penetration techniques are not as problematic to the host if only applied inside an endosome as opposed to the plasma membrane.

Endocytic viruses trigger uptake either in a receptor mediated fashion (clathrin, caveolin or lipid raft dependent) or via non-specific macropinocytosis. The clathrin pathway is most widely used, for example by rhinoviruses, some adenoviruses, and coronaviruses and presents with characteristic invaginations, termed clathrin-coated pits (CCPs). The inwards facing pockets are subsequently pinched off by the membrane scission proteins dynamin-1 and dynamin-

## 2.1. Microbial Host-Cell Infection



**Figure 2.1:** A generalized view of the steps necessary for viral–cellular membrane fusion. In pre-fusion conformation, fusion proteins have their hydrophobic fusion moieties tucked away. Upon attachment (1), they are unmasked (2), interact with the target membrane and ultimately penetrate it. Conformational change in fusion proteins induces membrane scission (3) and forces the two bilayers into close proximity (4), yielding a state of hemifusion (5). Finally a fusion pore is formed (6), stabilized by the post-fusion conformation which is lower in energy than the pre-fusion state. Adapted from [Hulo et al. \(2011\)](#)

2, releasing clathrin-coated vesicles into the cytosol. Caveolin-mediated endocytosis is thought to be a more tightly regulated and low capacity pathway but is nevertheless exploited by several virus species, including picornaviruses and some retroviruses. Caveola formation is lipid raft dependent and recruitment of caveolin yields 50–70 nm flask-shaped pockets which are closed off by dynamin action. A lipid raft dependent, caveolin independent pathway has been described in simian virus 40 infection, but remains poorly understood. Lastly, larger virions such as poxviruses or herpesviruses initiate macropinocytosis, a mechanism typically employed by the cell for non-specific uptake of extracellular particles. An actin dependent membrane ruffling leads to formation of a lamellipodium which folds back onto the plasma membrane, enclosing an extraluminal volume and thus creating a macropinosome.

Upon endocytic uptake, viral pathogens need to uncoat and eject their genetic material into the cytosol, as soon as their replicatory niche is reached. Escape timing is a critical issue, as late endosomes finally turn into lysosomes, capable of digesting their contents. Many enveloped viruses employ fusion mechanisms, which can be classified as type I or type II. For both types, increasing acidity associated with endosome maturation, initiates membrane fusion. Type I fusion proteins are forced into a metastable conformation prior to being added to the viral envelope and low pH triggers a conformational

## 2. BIOLOGICAL BACKGROUND

---

**Table 2.1:** The Baltimore classification scheme is based on diversity of genetic system that have evolved in viruses. For each group, a selection of virus families capable of infecting humans, is provided, along with whether the virions are enveloped and the location of their replicatory niche. The data was obtained from [Hulo et al. \(2011\)](#)

	Genome based class	Examples	Enveloped	Replication site
DNA viruses	Group I: dsDNA	<i>Adenoviridae</i>	no	nucleus
		<i>Poxviridae</i>	yes	cytoplasm
	Group II: ssDNA(+)	<i>Parvovirinae</i>	no	nucleus
		<i>Anelloviridae</i>	no	nucleus
	Group III: dsRNA	<i>Reoviridae</i>	no	cytoplasm
	Group IV: ssRNA(+)	<i>Coronaviridae</i>	yes	cytoplasm
		<i>Picornaviridae</i>	no	cytoplasm
RNA viruses		<i>Hepeviridae</i>	no	cytoplasm
Group V: ssRNA(-)	<i>Filoviridae</i>	yes	cytoplasm	
	<i>Paramyxoviridae</i>	yes	cytoplasm	
Group VI: ssRNA(+)-RT	<i>Orthoretrovirinae</i>	yes	nucleus	
Group VII: dsDNA-RT	<i>Hepadnaviridae</i>	yes	nucleus	

change to a state of lower energy. The energy released is used to force the two membranes close together resulting in their fusion (see figure 2.1). In type II fusion proteins, the critical transformation is not a conformational change but one in quaternary structure.

Non enveloped viruses cannot fuse with host membranes and have developed alternative approaches such as lysis (e.g. adenovirus) or ejecting their genome through pore-forming complexes (e.g. reovirus). Polyomaviruses need to pass through the ER because they rely on ER localized proteins to uncoat their capsid. For export from the ER into the cytosol, they exploit the endoplasmic-reticulum-associated protein degradation (ERAD) pathway, which serves as export mechanism for misfolded proteins from the endoplasmic reticulum to be degraded by proteasomes.

**Replication.** In contrast to larger intracellular parasites that carry the genetic information required for sustaining their own metabolism and replication, viral pathogens typically rely almost exclusively on host machinery. Furthermore, viruses have developed strategies for interfering with host transcription and translation in order to promote synthesis of viral proteins at the expense of host gene expression. Even modulation of the host cell cycle is not uncommon, as some DNA viruses (including adenoviruses) are able to trigger a G1 to S phase transition, yielding an increased concentration of active DNA polymerase, while other species are capable of inducing a G2/M arrest, which again provides an optimized environment for those viruses. Further virus-host

## 2.1. Microbial Host-Cell Infection

---

interactions include regulation of apoptosis, immune response modulation and interferon signaling.

The remarkable diversity of genomic systems employed by viruses is captured by a classification system devised by [Baltimore \(1971\)](#). Table 2.1 lists the 7 types of viral genomes alongside examples of human viruses for each group, as well as whether those viruses are enveloped and where they replicate. Consequently, requirements for replication, transcription and translation vary widely among the different groups of viruses and due to the resulting mechanistic heterogeneity, viral propagation is not further explored within this general section. An excellent overview is provided by the online database of [Hulo et al.](#).

**Viral export.** The final stage of the viral life-cycle is concerned with virion assembly and exiting the host cell. Again, many strategies exist. Some nuclear replicating viruses (such as polyomaviruses) assemble their capsid proteins within the nucleus, requiring their structural proteins to target the nucleus via nuclear localization signals (NLSs) and leave the nucleus by disrupting the nuclear envelope, while others have their genome exported via nuclear pores and assemble progeny virions in the cytoplasm. Some cytoplasmic viruses (including poxviruses) replicate within special structures called viral factories or viroplasms, which increase efficiency of assembly and packaging and provides protection from host defense mechanisms. Other cytoplasmic viruses localize to organelles such as the ER (e.g. flaviviruses) where they are assembled and enter the secretory pathway via the Golgi apparatus. For intracellular motility, large virions such as poxviruses or herpesviruses have to rely on microtubule dependent transport whereas particles smaller than 20 nm can freely diffuse within the cytosol.

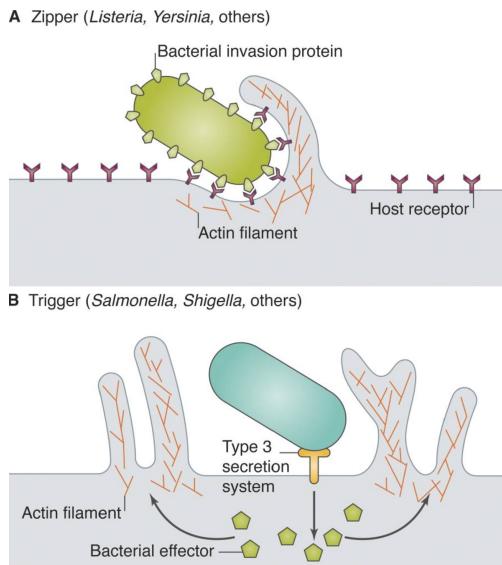
Once the host cell resources are depleted and replication is completed, progeny virions trigger their release. Viral shedding may occur via cell lysis, apoptosis, exocytosis or virion budding. Most non-enveloped and few enveloped viruses disrupt the plasma membrane with lytic viroproteins leading to cell death and release of cytoplasmic contents. While many viruses inhibit apoptosis, typically employed as a host defense measure, some (including hepeviruses and lentiviruses) have been implicated in exploiting this mechanism for expulsion and possibly subsequent infection of macrophages. Exocytosis and virion budding are two release strategies that are non-lethal to the host cell. Enveloped viruses acquire host-derived membrane either within the cell, typically at ER or Golgi exit sites or directly from the plasma membrane. In the latter case, envelopment coincides with host exit, whereas in the former case, virions are expelled via fusion of exocytic vesicles with the plasma membrane.

### 2.1.2 Bacterial Entry Mechanisms

Due to the much larger size of bacterial pathogens, endocytosis is not a feasible mechanism for entry. Phagocytosis, however can deal with uptake of

## 2. BIOLOGICAL BACKGROUND

---



**Figure 2.2:** Both the zipper (A) and trigger mechanisms (B) are actin dependent and lead to phagocytosis by usually non-phagocytic host cells. Zippering bacteria display an invasion protein on their surface that recruits actin filaments via a host receptor, while triggering bacteria inject an effector into the host cytosol via a type III secretion system leading to uptake. ([Haglund and Welch, 2011](#))

particles this large. While phagocytosis is a function usually only available to macrophages, some bacteria have evolved mechanisms of inducing phagocytosis in other cell types. Species explicitly targeting macrophages, such as *Mycobacterium tuberculosis* and *Legionella pneumophila*, have to be able to escape phagosomes or deal with resisting digestion.

Two recurring patterns for inducing phagocytosis in non-phagocytic cells have been described as the zipper mechanism, encountered in *Yersinia pseudotuberculosis* and *Listeria monocytogenes* and the trigger mechanism used by *Salmonella enterica* and *Shigella flexneri*. Not all entry strategies can be assigned to these two classes and several additional, unrelated pathways have been investigated.

**Zipper Mechanism.** The first step for zippering bacteria is binding to target cell receptors by expressing adhesins. *Y. pseudotuberculosis* displays invasins on its cell surface, capable of interacting with  $\beta_1$  integrins, while *L. monocytogenes* uses internalin A, a protein that binds E-cadherin. In both settings, a downstream signaling cascade leads to actin polymerization via recruitment of an Arp2/3 (actin-related-protein 2 and 3) complex and Rac1 (Ras-related C3

## 2.1. Microbial Host-Cell Infection

---

botulinum toxin substrate 1), yielding a phagocytic cup.

Cadherin and integrins are usually involved in anchoring cell junctions and the host cell is fooled into thinking a neighboring cell is initiating formation of such a junction. Responding by recruiting actin at the site of bacterial attachment to cover the surface of mistaken invasion proteins leads to engulfment and phagocytosis to the bacterium. This process incurs only modest cytoskeletal rearrangements.

**Trigger Mechanism.** In case of the trigger mechanism, bacterial T3SS weakly adhere to target cell receptors (CD44 for *Shigella*) and effector molecules are injected into the host cytosol through a pore formed by Sip (*Salmonella* invasion protein) or Ipa (*Shigella* invasion plasmid antigen) proteins. These induce major actin rearrangements that result in localized ruffling of the plasma membrane and subsequent swallowing of the bacterium.

Early in *Shigella* entry, VirA, secreted through the T3SS causes local destabilization of microtubules by binding to  $\alpha/\beta$ -tubulin heterodimers. This in turn stimulates Rac1 activity, Cdc42 (cell division control protein 42 homolog) recruitment and subsequent Arp2/3 activation leading to protrusions formed by actin filaments. IpaC recruits the Src tyrosine kinase further enhancing actin dynamics. Upon closing of the phagocytic pocket, the T3SS-secreted protein IpaA binds to vinculin and induces actin depolymerization.

*Salmonella* inject the proteins Sop (*Salmonella* outer protein)E and SptP (secreted effector protein), an activator/inhibitor pair for the GTPase complex Rac1/Cdc42, into the host cell. First, guanine nucleotide exchange factors (GEF) activity of SopE induces massive actin-rearrangements leading to membrane ruffling and facilitating phagocytosis, followed by GAP activity of SptP, restoring the inactive guanosine diphosphate (GDP) state of Rac1/Cdc42 and leading to actin depolymerization. SopE is degraded more rapidly than SptP, enabling the invading pathogen to reversibly control the pathways exploited for entering.

**Other Entry Pathways.** In addition to trigger and zipper type uptake, other atypical mechanisms exist. Host cell entry of *Brucella abortus*, for example, has been described as invasome mediated (Dehio, 2005). In this actin dependent process, bacteria aggregate on the cell surface and trigger their engulfment by injecting bacterial effectors into the host via type IV secretion system (T4SS). The internalized structure is called an invasome. Actin-independent uptake albeit rare, is possible as evidenced by *Campylobacter jejuni*, which have evolved a microtubule dependent invasion strategy (Kopecko et al., 2001).

### 2.1.3 Actin-based Intracellular Motility

As for large viruses, free diffusion within the cytoplasm is not possible for bacteria.

## 2. BIOLOGICAL BACKGROUND

---

### **2.2 Select Bacterial Pathogens**

A total of 5 bacterial pathogens were selected for study within the InfectX RTD project by SystemsX. This section shortly describes each organism in terms of microbiological features, pathogenesis, epidemiology and diseases caused in humans. For each organism, a chapter of [Rolain and Raoult \(2006\)](#) serves as basis and is augmented by one or two review article referenced in the first paragraph of each section.

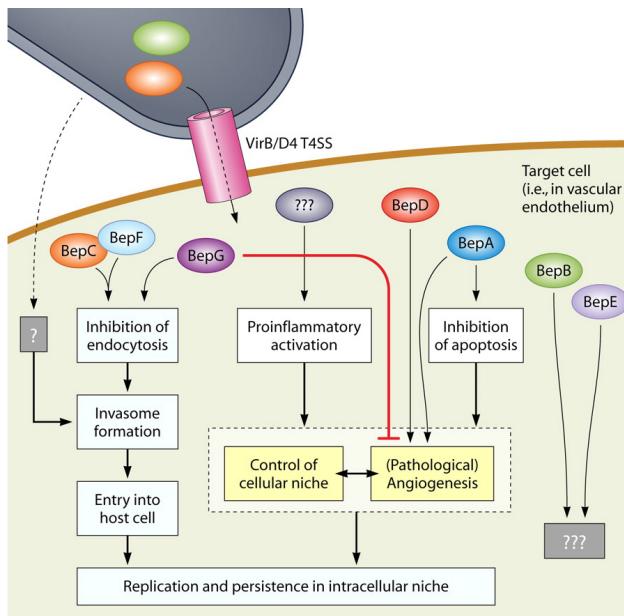
#### **2.2.1 *Bartonella Henselae***

*Bartonella henselae* is a short, rod shaped, unflagellated proteobacterium, phylogenetically closely related to the genus *Brucella*, presenting 94.4% 16S rRNA gene sequence homology, compared with *Brucella abortus*. The Gram-negative bacillus is a facultative anaerobic, intracellular parasite and was first described in 1992. Relatively harmless for healthy humans, infections can become life threatening in immunocompromised patients, making the species an important opportunistic pathogen ([Anderson and Neuman, 1997](#)).

**Diseases.** In immunocompetent humans, infection with *B. henselae* can lead to a condition known as Cat Scratch Disease (CSD). As the name suggests, most patients report being in contact with a cat and transmission often occurs through scratches and bites. Affecting primarily children and young adults (80% are 21 or younger), the self limiting infection typically presents itself with lymphadenopathy. Most patients remain afebrile and do not report feeling ill, with low-grade fever and malaise shown in roughly 30% of the cases. Recovery from uncomplicated CSD usually takes 2 to 6 months and requires no specific treatment.

Possible complications include Parinaud's oculoglandular syndrome (granulomatous conjunctivitis in one eye and parotid lymphadenitis on the same side), splenomegaly and hepatic or splenic abscesses, accompanied by fever, weight loss, fatigue and malaise. In 1 to 7% of the cases, the disease spreads to the central nervous system, leading to encephalopathy, but recovery is usually rapid (within several weeks).

Infections with *B. henselae* tend to have more severe consequences for immunocompromised patients, such as bacillary angiomatosis, bacteremia and endocarditis. Acquired immune deficiency syndrome (AIDS) patients suffering from CSD usually experience severe, progressive disease with infection spreading systematically and without appropriate treatment, fatal outcome. *Bartonella spp.* are the only prokaryotes known to be able to induce angiogenic tumors such as bacillary angiomas, which may involve skin, respiratory or gastrointestinal epithelia, heart, liver, spleen, bone marrow, muscles, or lymph nodes. Bacteremia may lead to inflamed heart valves, usually requiring endocarditic patients to have heart valve replacement surgery.



**Figure 2.3:** Capsid proteins VP1 through VP4 form a pseudo  $T = 3$  icosahedral coat, roughly 30 nm in diameter around the RNA genome (A), which is monopartite, linear, 7.2 kp long and encodes 11 proteins (Harms and Dehio, 2012).

**Pathogenesis.** *B. henselae* are capable of intracellular growth in both epithelial cells, and erythrocytes. Bundle forming type IV pili are essential for binding to target cells, making them important virulence factors. Internalization into red blood cells may be spectrin mediated and the bacterial protein deformin may also be involved. Endothelial receptors involved include intercellular adhesion molecule 1 (ICAM-1) uptake is either via endocytosis or by a unique cellular structure termed an invasome.

Vasculogenesis is induced through an increased production in vascular endothelial growth factor (VEGF) by infected host cells. Currently it is poorly understood how the pathogen provokes overexpression but the mechanism has been determined to be protease sensitive.

**Epidemiology.** The role of cats and in particular kittens, as reservoirs to *B. henselae* has been firmly established. Infected felines are asymptomatic and show no signs of illness. Cat fleas (*Ctenocephalides felis*) serve as vectors to spread the bacteria among cats and have also been suspected of infecting humans. The main path of transmission to humans however, is through scratches

## **2. BIOLOGICAL BACKGROUND**

---

and bites by infected cats. *B. henselae* has also been found in ticks and tick bites prior to contraction of CSD have been reported.

In the United States, 24000 cases of CSD are reported yearly, yielding 2000 hospital admissions with an estimated health care cost of \$12 million. Children are more likely to be affected (80%) and incidence is higher in males (60%). The seasonal pattern (occurrences higher in fall/winter) is attributed to cat mating patterns, as well as pet acquisition fluctuations.

### **2.2.2 *Brucella Abortus***

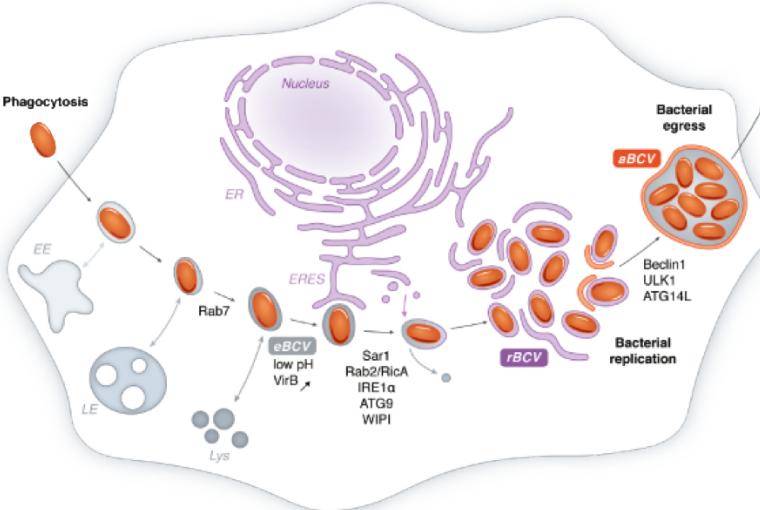
The Danish physician David Bang first isolated *Brucella abortus* in 1895 from cystic cattle tissue, investigating a contagious disease causing abortions in cows. *B. abortus* are small, unflagellated proteobacteria with a cell wall consisting of an outer layer of lipopolysaccharide (LPS) (9 nm) and an inner layer of muramyl mucopeptide complexes (3–5 nm). The Gram-negative cocobacilli appear to have evolved from free-living, soil-dwelling species and are closely related to other human pathogens such as *Bartonella* spp., based on 16S rRNA sequences. *Brucella* species were investigated for possible use as warfare agents in the mid 20th century by several armed forces. ([Atluri et al., 2011](#))

**Diseases.** Brucellosis is a human disease caused by several pathogenic *Brucella* species, most importantly *B. abortus*, *B. melitensis*, *B. canis* and *B. suis*. Onset may be acute or insidious and due to protean symptoms, diagnosis based on clinical presentation alone is difficult. The febrile disease is generalized and may involve many parts of the body, including nervous, skeletal, gastrointestinal, cardiovascular, respiratory and genitourinary systems. Furthermore, as the bacteria spread to other reservoir hosts via their reproductive systems, persistence of infection is crucial to the pathogen and it comes as no surprise that brucellosis can manifest as a chronic disease in humans too.

Fever is the most consistent sign of *Brucella* infection and depending on what specific organs are affected, further symptoms include asthenia, anorexia, nausea, malaise, arthritis, hepatomegaly, splenomegaly, epididymo-orchitis in males, and pulmonary manifestations such as bronchitis or pneumonia. A rare complication (less than 2%), albeit the most lethal, is infective endocarditis. Invasion of the nervous system occurs in less than 5% of cases and often results in meningitis or meningoencephalitis with good prognosis under antimicrobial treatment.

**Pathogenesis.** Host entry happens primarily via the digestive system but is also possible through the respiratory tract or skin lesions. On the gastrointestinal route, *Brucella* spp. target Peyer's patches (lymphoid nodules localized towards the end of the small intestine) and must therefore pass through acidic conditions in the stomach. This is facilitated by expression of two ureases capable of hydrolyzing urea and producing a protective bicarbonate buffering

## 2.2. Select Bacterial Pathogens



**Figure 2.4:** Capsid proteins VP1 through VP4 form a pseudo  $T = 3$  icosahedral coat, roughly 30 nm in diameter around the RNA genome (A), which is monopartite, linear, 7.2 kp long and encodes 11 proteins (Celli, 2015).

system. When entering through the respiratory system, *B. abortus* target alveolar macrophages which serve as access point to the lymphatic system therefore facilitating systematic spread.

In order to persist at systemic sites, both active and passive mechanisms for evading the immune system are in place. LPS of the outer cell wall disguises the bacteria from toll-like receptors (TLRs) and expression of two proteins containing toll-interleukin-1 receptor (TIR) domains actively interferes with TLR signaling.

Uptake by macrophages happens via phagocytosis, which is either triggered by nonopsonized bacteria through a lipid raft mediated mechanism or by opsonization. Although opsonin marked bacteria are 10-fold more likely to be ingested, the number of pathogens reaching their replicatory niche within the host cell is higher for nonopsonized bacteria. Maturation of early endosomes into lysosomes is important for successful infection as preventing acidification (through addition of baflomycin A) or fusion with lysosomes (through suppression of the late-endosomal GTPase Rab7) interferes with bacterial replication. Nonopsonized bacteria finally become associated with rough ER and begin replication in ER derived vacuoles within the ER-Golgi intermediate compartment (ERGIC). Blocking the small GTPase Sar1 inhibits intracellular

## 2. BIOLOGICAL BACKGROUND

---

replication by preventing acquisition of coat protein complex II, involved in anterograde ER–Golgi transport (COPII) by ER-exit vesicles and the small GTPase Rab2, involved in ER–cis-Golgi traffic, is required for maximal proliferation.

Despite multiplying intra-cellularly in high numbers, host cells are kept alive and are even able to replicate despite infection. Furthermore *Brucella* species are able to interfere with apoptosis, maintaining their replicatory niche, protected from immune response.

**Epidemiology.** Preferred natural reservoir species for *B. abortus* are cattle (*Bos taurus* and *Bos indicus*) and almost all parts of the world are affected. The disease exists in both domestic and wild animals and is most prevalent in Mediterranean countries, North Africa, throughout the Middle East, India, Central Asia, as well as South and Central America. Zoonosis most often occurs through ingestion of unpasteurized milk products but airborne transmission is also possible, putting professionals involved in animal husbandry at risk. Vertical transmission among reservoir hosts can occur through lactation and horizontal transmission is facilitated by mating and placental discharge associated with aborted gestation. Human-to-human transmission is rare (but has been suspected to be possible via sexual intercourse), making humans dead-end hosts. As opposed to *Bartonella*, immunodeficient patients do not seem to be especially susceptible towards *Brucella* infections.

Worldwide, an estimated 500000 new cases of brucellosis occur annually, making it one of the most prevalent zoonoses. Although usually susceptible to combined antibiotic therapies of at least two agents (usually a tetracycline antibiotic combined with an aminoglycoside or rifampin), untreated brucellosis leads to a high degree of morbidity, leading to being classified a neglected zoonosis by the World Health Organization (WHO).

### 2.2.3 *Listeria Monocytogenes*

The short, Gram-positive bacilli are non-sporeforming facultative anaerobes, capable of growing in a wide temperature range (0–50°C) and in many different environments. Flagellation is temperature dependent with flagellin being expressed and assembled into peritrichous flagella around 20–25°C but not at 37°C. First described in 1924 by Murray after isolation from lymph glands of diseased laboratory animals, the pathogen was found to also infect humans four years later. For much of the time since, listeriosis was considered a rare zoonotic disease and did not receive much attention. It was not until the 1980s, when several food-borne listeria outbreaks caused a shift in interest towards the pathogen, which has since become a well studied facultatively intracellular parasite (Farber and Peterkin, 1991).

**Diseases.** Maternal and neonatal listeriosis accounts for almost half of all infections. Listeriosis in pregnancy typically manifests in bacteraemia and

## 2.2. Select Bacterial Pathogens

---

presents as a self-limiting febrile disease with flu-like symptoms. Many cases, however are asymptomatic and the first sign of infection is abortion or neonatal listeriosis. Maternal infection does not necessarily carry over to the fetus, especially if proper chemotherapy is administered. Perinatal incidences are divided into early and late onset ( $>5$  days after parturition), with former cases typically resulting in septicemia and latter cases in meningitis. While in early onset cases the predominant route of transmission is transplacental, the situation is less clear in late onset cases. Both the maternal genital tract during child birth and environmental sources have been implicated. Despite antibiotic treatment, overall mortality rates of 30–40% are typical and prognosis for early onset disease is worse, as it is often associated with preterm birth and advanced stage of infection.

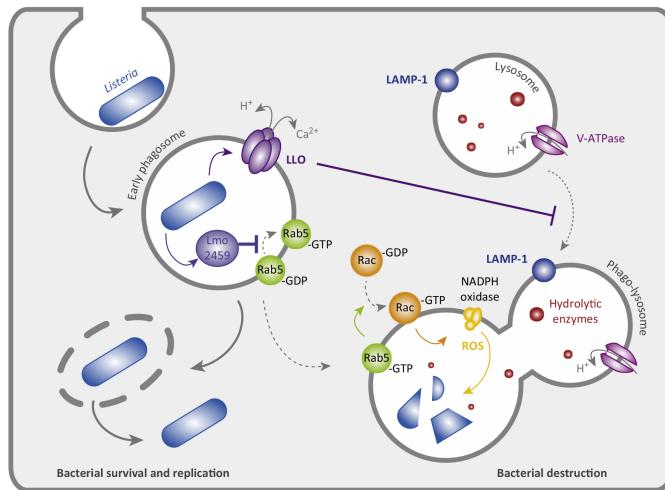
Among adults, most cases of listeriosis occur in T-cell deficient individuals. HIV infection, for example, increases incidence 150–300 fold compared to general population control groups. Predisposing conditions include lymphoreticular neoplasms, deliberate immunosuppression (e.g. antirejection treatment after organ transplants), alcoholism and diabetes mellitus. Despite increased susceptibility caused by immunodeficiency, roughly 30% of all infections affect immunocompetent individuals. In healthy adults, consumption of food contaminated with *L. Monocytogenes* can either lead to self-limiting febrile gastroenteritis with short incubation time (<24 h) or invasive listeriosis with much longer incubation periods (3–4 weeks). The systemic form of infection often manifests as bacteraemia or as a neurological infection, but can also involve endocarditis and spread to other parts of the body. Central nervous system involvement occurs in as much as 75% of cases and either presents as meningitis or encephalitis. Mortality rates of 35–45% have been reported for listeriosis in adults.

**Pathogenesis.** The predominant entry path for *L. Monocytogenes* into the human body is via the gastrointestinal tract, where Peyer's patches are targeted. The bacteria can induce cellular uptake (see section 2.1.2), even by non-phagocytic host cells via expression of cell-surface associated interanlins. Upon cell entry, the phagosomal membrane is lysed, mediated by haemolysin secretion, and actin polymerization (see section 2.1.3) provides means of intracellular movement. Listerial haemolysin (listeriolysin) is acid activated with an optimum around pH 5.5, which is reached in late endosomes and membrane lysis is caused by the formation of large transmembrane pores. Growth and replication occur in the cytoplasm and adjacent cells can be entered by pushing against the plasma membrane and forming a pseudopod-like structure which in turn is taken up the neighbor. The resulting double-membrane vacuole can be escaped by cytolytic.

In an extracellular setting, haemolysins serve to rupture erythrocytes in order to generate free iron, a limiting growth factor. Furthermore the nonspecific immune system has to be evaded and haemolysins have also been shown to be cytotoxic towards leukocytes. Additionally, expression of bacterial super-

## 2. BIOLOGICAL BACKGROUND

---



**Figure 2.5:** Capsid proteins VP1 through VP4 form a pseudo  $T = 3$  icosahedral coat, roughly 30 nm in diameter around the RNA genome (A), which is monopartite, linear, 7.2 kp long and encodes 11 proteins (B). Figures adapted from [Lebreton et al. \(2015\)](#).

oxide dismutase mitigates the effect of free superoxide radicals which play an important role in killing phagocytized bacteria.

**Epidemiology.** Incidence of listeriosis has initially been increasing since its recognition as food-borne disease but effects of awareness and diagnostic methods are unclear. While typically long incubation periods do pose difficulties for clinical diagnosis, the number of susceptible individuals is on the rise and certain aspects modern processing and handling of foods may be beneficial for growth of *L. Monocytogenes*. Disease rates of 2–15 cases per million population per year have been reported and listeriosis is among the leading case of lethal food-borne pathogen infections. Most recent data, however suggests, that incidence is decreasing again.

Due to its non-fastidious lifestyle, *L. Monocytogenes* has been isolated from a wire array of ecological niches, including soil, sewage and water (both fresh water bodies and estuaries). High persistence (up to 4 years) in soil samples is problematic when contaminated manure is used as fertilizer and biofilm formation poses challenges for eradication from food processing plants. Additionally, the ability to grow in refrigerated foods and resistance towards heat treatment such as pasteurization warrant alertness and special preventive care. Many different foods have been implicated in listeriosis outbreaks, including vegetables (potatoes, radishes and celery), seafood (shrimp, crabmeat and smoked fish),

## 2.2. Select Bacterial Pathogens

---

dairy products (soft cheese, pasteurized and unpasteurized milk) and meats (poultry, various types of sausages and pâté).

Despite high prevalence in food (studies have found 20–80% of meat product samples and 1–10% of dairy product samples contaminated with *L. Monocytogenes*), comparatively few successful infections occur. The bacteria are ingested frequently in small doses and stool sample examinations suggest that 10–70% of investigated populations could be intestinal carriers. While the minimum infectious dose has not been settled definitively, approximations range from  $10^6$  to  $10^9$  bacteria.

### 2.2.4 *Salmonella* Typhimurium

*Salmonella* are non-sporulating, Gram-negative bacilli, belonging to the family Enterobacteriaceae. The motile bacteria are able to produce peritrichous flagella and diameters span 0.7 to 1.5  $\mu\text{m}$  while typical lengths range from 2 to 5  $\mu\text{m}$ . They are closely related to the genus *Escherichia*, showing only 15% chromosomal sequence disparity. Currently, two distinct species, *S. bon-gori* and *S. enterica*, within the genus *Salmonella* are recognized, both of which are pathogenic towards a wide array of hosts. *S. enterica* is further divided into 6 subspecies, the most relevant of which for human and domestic animal hosts being *enterica*. A large number of serovars (more than 2500) for *S. enterica* subsp. *enterica* have been characterized and due to an originally mistaken classification as separate species, some serovars are designated with shortened names. *S. Typhimurium*, therefore is shorthand for *S. enterica* subsp. *enterica*, serovar Typhimurium and to emphasize that Typhimurium is not a species description it is not italicized.

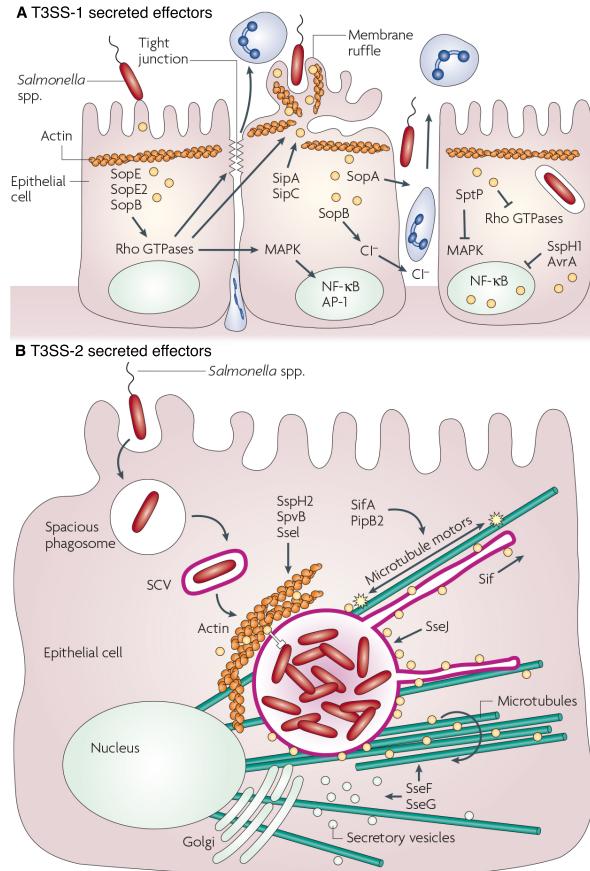
The first description of the genus *Salmonella* dates back to an investigation into swine fever led by Salmon and Smith in 1885. The newly isolated bacterium was wrongly proposed as the etiological agent, as the disease later turned out to be caused by a virus ([Fàbrega and Vila, 2013](#)).

**Diseases.** Two distinct disease patterns are associated with *Salmonella* spp. infections, typhoid fever and salmonellosis. While the former is exclusively caused by the serovars Typhi and Paratyphi, the latter is associated with several serovars, the most frequent being Enteritidis and Typhimurium, accounting for 65% and 12% of cases worldwide. The current and following sections will not discuss typhoid fever.

Salmonellosis is a diarrheal disease with a short incubation period of 6–24 h, followed by nausea, vomiting, loose or liquid bowel movements, abdominal cramps and fever. Clinical features are similar to those of dysentery and other gastroenteric disease and can include bloody and mucosal stool. In most cases, the infection is self-limiting and symptoms fade away within 4 to 7 days. The most common complication is bacteraemia, which presents in 5% of cases and is more likely to develop in children, especially if malnourished,

## 2. BIOLOGICAL BACKGROUND

---



**Figure 2.6:** Capsid proteins VP1 through VP4 form a pseudo  $T = 3$  icosahedral coat, roughly 30 nm in diameter around the RNA genome (A), which is monopartite, linear, 7.2 kp long and encodes 11 proteins (B). Figures adapted from [Haraga et al. \(2008\)](#).

and immunocompromised individuals. Further manifestations of invasive infections include meningitis, osteomyelitis, cholangitis, pneumonia and endocarditis. While mortality in immunocompetent hosts in developed countries is as low as 0.1%, it can increase to 77% for HIV positive patients in undeveloped countries.

## 2.2. Select Bacterial Pathogens

---

**Pathogenesis.** In order to reach the small intestine, ingested *S. Typhimurium* first are exposed to the hostile environment of the stomach. A set of proteins, summarized as acid tolerance response (ATR) helps mediate the acidic conditions and improves survival rates. The reamining bacteria subsequently reacht the small bowel and target its epithelial cells, with preference towards micro-fold cells (M cells). Flagellar motility enables penetration of intestinal mucus secretions and improves the chance of reaching the intestinal walls where adhesion can be established. Fimbriae are important attachment factors, capable of interaction with host-cell laminin and fibronectin and provide an initial platform for pathogen induced phagocytosis via trigger mechanism (see section 2.1.2).

*S. Typhimurium* utilize two separate T3SS systems for host-cell colonization, the first of which (T3SS-1) mediates invasion. In addition to strengthening initial interactions attaching the pathogen to its target, the needle like structure serves as delivery mechanisms, capable of injecting effector proteins, including SopE, SopE2, SopB (also known as SigD) and SipA. SopE serves as GEF and activates Rac1 and Cdc42 via GDP/GTP exchange, which in turn initiate cytoskeletal rearrangements. SopE2 is a further GEF, highly homologous to SopE and it is assumed to fill in for SopE in *Salmonella* strains missing the SopE gene. SopB is a phosphatase, capable of hydrolyzing various substrates, including inositol 1,3,4,5,6-pentakisphosphate ( $I(1,3,4,5,6)P_5$ ), which has been linked to intestinal fluid secretion causing diarrhea and phosphatidylinositol 4,5-bisphosphate ( $PI(4,5)P_2$ ), which is involved with membrane rigidity. Finally, the actin binding protein SipA stimulates stimulates actin polymerization and promotes membrane ruffling.

Upon engulfment, maturation of the phagosome and fusion with lysosomes have to be prevented. Unlike other pathogens that escape the digestive mechanisms of phagocytic vesicles by moving to the cytosol, *Salmonella* replicate inside phagosome-derived *Salmonella*-containing vacuoles (SCVs). In this setting, the second translocation complex (T3SS-2) is activated, and used to secrete effectors, including SsaB and SifA, capable of interacting with vesicular trafficking mechanisms and guiding the SCV away from its original degradation pathway. Furthermore T3SS-2 translocated effectors induce actin polymerization events, which guide the SCV toward a perinuclear position. A last step in creating the intracellular niche needed for replication, is formation of *Salmonella*-induced filamentss (SIFs), extending outwards from the SCV. T3SS-2 secretes effectors such as SifA, PipB2, SseF and SseG, which mediate the microtubule dependent processes by bundling and accumulating microtubules and regulating microtubule motor function.

**Epidemiology.** The global disease burden caused by nontyphoidal salmonellosis is estimated at 90 million cases per year and 150000 deaths. Incidence rates are highest in East and Southeast Asia (up to 4000 cases per 100000 population per year) and both developed and undeveloped countries are affected (incidence rates in Africa are estimated at 320 while estimates for Europe are

## 2. BIOLOGICAL BACKGROUND

around 690 cases per 100000 population per year). An estimated 80.3 million or 86% of reported cases are food borne ([Majowicz et al., 2010](#)).

In order to control *Salmonella* outbreaks, preventive measures in food production and processing is of major importance. This starts with disease containment in domestic animals, such as vaccination of chickens, enforcing hygiene standards in manufacturing and distribution facilities and ends with proper preparation, exploiting heat sensitivity of the organisms. As the main route of transmission is fecal-oral, good sanitary infrastructure, treatment of sewage and water processing are crucial prerequisites in combating outbreaks of salmonellosis.

### **2.2.5 *Shigella Flexneri***

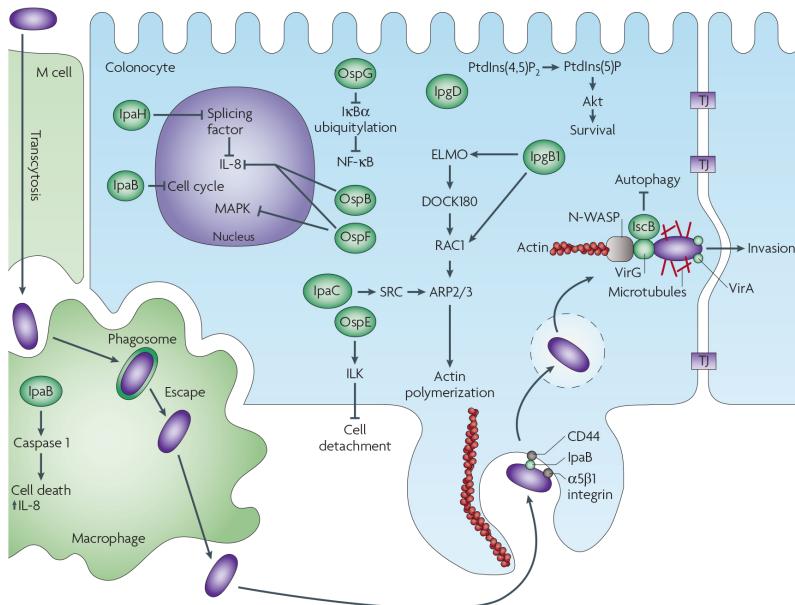
Shigellae are small, non-sporeforming, Gram-negative bacilli and belong to the family *Enterobacteriaceae*, along with *Escherichia*, *Salmonella* and *Yersinia*. While flagellar genes are present and their expression is observed under certain conditions, the bacteria are usually described as non-motile and unflagellated. The facultative intracellular parasite shows strong specificity towards human hosts where it typically infects the lower gastrointestinal tract.

*Shigella dysenteriae* was identified as the etiological agent of dysentery by Shiga in 1897 during an epidemic in Japan with 91000 reported cases and a >20% mortality rate. *S. Flexneri* was first described by Flexner in 1900, while investigating diseases endemic to the Philippines. Recent genetic studies suggest, that *Shigella* spp. belong to the species *Escherichia coli*, rather than forming a distinct genus, as only marginal sequence divergence (1.5%) between *E. coli* and *S. Flexneri* was found ([Schroeder and Hilbi, 2008](#)).

**Diseases.** Bacillary dysentery is an acute infection of the intestine. Mild cases of the disease are self-limiting and afebrile with diarrhea and possibly vomiting as the only symptoms. In as little as 24 hours after onset, bowel movements usually begin to normalize and the condition is resolved within a few days. More severe cases are accompanied by strong abdominal cramps, fever and watery diarrhea containing blood and mucus, indicative of injury to the intestinal epithelium. While still usually self limiting in healthy individuals, recovery takes 10–14 days and relapses are possible. In immunocompromised patients, young children, especially if malnourished, and elderly individuals, life threatening complications including bacteraemia, renal failure, intestinal perforation and toxic megacolon are more frequent. Involvement of the central nervous system and respiratory tract is rare.

Administration of antibiotics is not recommended in mild to moderate cases as the disease can usually be overcome by the immune system and AMR in Shigellae is becoming an increasing concern. Oral rehydration therapy is the most effective treatment, helping the body replenish liquids and salts lost due

## 2.2. Select Bacterial Pathogens



**Figure 2.7:** Capsid proteins VP1 through VP4 form a pseudo  $T = 3$  icosahedral coat, roughly 30 nm in diameter around the RNA genome (A), which is monopartite, linear, 7.2 kp long and encodes 11 proteins (B). Figures adapted from [Croxen and Finlay \(2010\)](#).

to diarrhea. For severe infections, the use of antibiotics can become necessary and testing for resistance patterns, if possible, is advised.

**Pathogenesis.** Main targets of *S. Flexneri* are mucosa of the distal ileum and colon where they enter epithelial cells from the basolateral side. For initial crossing over from the apical side, action of M cells at Peyer's patches is exploited. These specialized enterocytes constantly sample antigens from the gut lumen and pass them to intraepithelially located dendritic cells and lymphocytes. Upon uptake by basolaterally located macrophages, *S. Flexneri* survive digestive action of the phagosomal vacuole by disrupting the surrounding membrane and initiating host-cell apoptosis, mediated by the bacterial effector protein IpaB, capable of acting on a caspase 1 regulated apoptotic pathway. The bacteria are subsequently released into the sub-mucosa, where they induce phagocytosis by normal epithelial cells via trigger mechanism (see section 2.1.2).

Initial contact between pathogen and target host cell is mediated by cellular  $\alpha_5\beta_1$  integrins and binding of IpaB to CD44 receptors may initiate first actin

## 2. BIOLOGICAL BACKGROUND

rearrangements, preparing the cell for uptake. Subsequent injection of at least 6 effector proteins into the epithelial cytosol through the T3SS triggers engulfment of the bacterium in an actin dependent process, involving the small GTPases Rac1 and Cdc42, which recruit the actin nucleation complex Arp2/3. IpaC, a component of the translocation complex, is involved in stimulating Rac1 and Cdc42 GTPase activity through an unknown mechanism and the secreted effector proteins IpgB1, IpgB2, IpgD and IpaA facilitate actin polymerization. IpgD, an inositol phosphate phosphatase, catalyzes the hydrolysis of PI(4,5)P<sub>2</sub> to PI(5)P which promotes disassociation of cytoskeleton and membrane, increasing actin availability and making the plasma membrane more susceptibility to manipulation. The mechanism of action of IpgB2 remains to be resolved, while IpgB1 is assumed to mimic activated RhoG (Ras homology growth-related), a small GTPase, located upstream in the signaling pathway of Rac1. Finally, binding of IpaA to vinculin induces depolymerization of actin, which is assumed to be important for closing of the phagocytic cup.

The resulting phagocytic vacuole has to be escaped before maturation progresses, which is accomplished in a non-acid dependent fashion, by the translocator proteins IpaB, IpaC and IpaD via membrane lysis. With release into the epithelial cytosol, the replicatory niche of *S. Flexneri* is reached. Actin mediated intracellular motility enables intercellular dissemination (see section 2.1.3) and targeting of epithelial tight junctions initiates break-down of the epithelial barrier, providing more pathogens with access to the basolateral side of the gut lining.

Host-cell actin utilization, is driven by activity of two bacterial proteins. VirA is secreted at the forward facing end of the rod shaped bacilli, which promotes degradation of tubulin structures and therefore clearing a path through the dense network of microtubules and surface bound Ics (*Shigella* intracellular spread protein)A facilitates actin polymerization at the opposite end. Both Arp2/3 and neural Wiskott-Aldrich syndrome protein (N-WASP) are involved in actin nucleation, the directed nature of which provides the driving force for locomotion.

**Epidemiology.** Estimates by the WHO peg the disease burden caused by *Shigella* spp. at 80 million annual cases worldwide, leading to 700000 deaths. Developing countries are disproportionately affected, representing 99% of all cases, as are children less than 5 years old, accounting for 70% of cases and 60% of deaths. In developed countries, incidence rates of 1–2 per 100000 population are typical and Shigellae are common causes of Traveler's diarrhea.

The predominant route of transmission is fecal–oral, highlighting the importance of sanitary precautions for infection control. Proper treatment of fecal matter is important for preventing contamination of drinking water and inhibiting spreading by disease vectors such as house flies. During acute phases, diseased individuals excrete pathogens in large numbers and as few as 100–200 organisms are sufficient of causing a new infection.

## 2.3 Select Viral Pathogens

In addition to the previous 5 bacterial pathogens, 3 viruses were selected for study within the InfectX RTD project by SystemsX. This section shortly describes each pathogen in terms of physical features, pathogenesis, epidemiology and diseases caused in humans. For each section, a chapter of [Craighead \(2000\)](#) serves as basis.

### 2.3.1 Adenoviruses

The family *Adenoviridae* encompasses 5 genera of non-enveloped, medium sized (90 nm diameter) viruses, capable of infecting a broad range of vertebrate hosts. The capsid is of  $T = 25$  icosahedral symmetry, composed of 720 hexons arranged as 240 trimers which form the triangular facets and 12 penton capsomeres located at the vertices. A homotrimeric fiber glycoprotein protrudes from each vertex, attached to the penton base via interactions of its N-terminal domains and ending in a globular, C-terminal knob. The genome is present as double stranded DNA (Baltimore group I), is non-segmented, linear, 35–35 kb long and encodes 40 proteins.

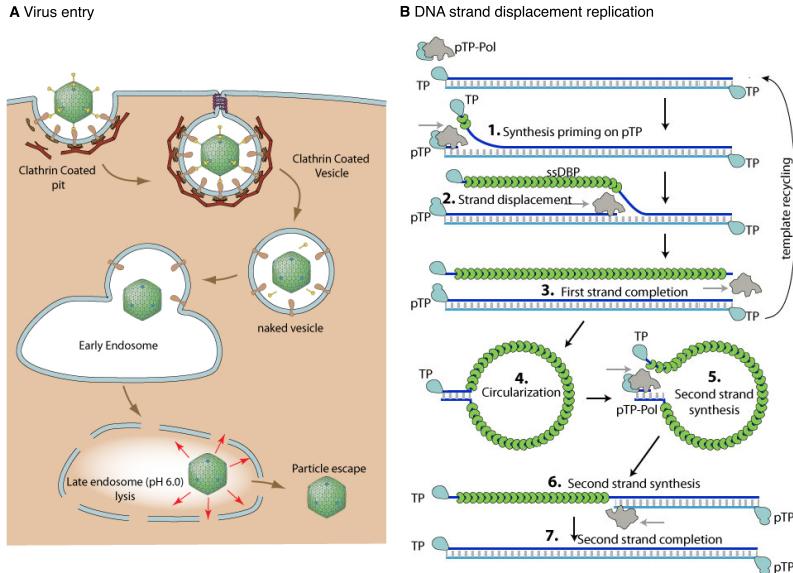
Adenoviruses were first isolated from human adenoid tissue cultures by Rowe in 1953 and their study led to the discovery of alternative splicing in 1977, a commonly encountered phenomenon among eukaryotes. Currently, 57 serovars are recognized as pathogenic towards humans, all belong to the genus *Mastadenovirus* and are classified into 6 species, labeled A through G. The following sections are mostly concerned with *Human adenovirus C* ([Lenaerts et al., 2008](#)).

**Diseases.** In immunocompetent individuals, adenoviruses seldom cause more than transient disease with many infections even occurring subclinically and fatal outcome being very uncommon. Symptomatic cases usually manifest as respiratory tract infections or conjunctivitis and less frequently as hemorrhagic cystitis, nephritis or gastroenteritis. Infections of the oropharynx can spread to the lower respiratory tract, causing bronchitis, bronchiolitis or pneumonia, which can become chronic, leading to desquamated epithelial tissue and long-term damage to the respiratory mucosa. Heart failure and central nervous system involvement can occur in severe cases. Ocular infections range from mild, short-term follicular conjunctivitis to highly contagious keratoconjunctivitis with possible long-term damage to the cornea. *Human adenovirus C* serotypes are mostly associated with respiratory diseases but have also been implicated in eye infections.

Invasive forms of disease are mostly limited to immunocompromised individuals, including transplant recipients, HIV infected, hereditary immunodeficient and cancer patients subject to chemotherapy. In addition to the lungs, a wide variety of organs has been reported to be affected, such as partoid glands, liver, gall bladder, colon, brain and kidney and pathological changes range from perivascular cuffing to parenchymal necrosis.

## 2. BIOLOGICAL BACKGROUND

---



**Figure 2.8:** Capsid proteins VP1 through VP4 form a pseudo  $T = 3$  icosahedral coat, roughly 30 nm in diameter around the RNA genome (A), which is monopartite, linear, 7.2 kp long and encodes 11 proteins (B). Figures adapted from [Hulo et al. \(2011\)](#).

**Pathogenesis.** Initial attachment of virions is mediated by interactions between the globular knobs of viral fiber proteins and target cell CARs (coxsackievirus and adenovirus receptors). In *Human adenovirus C* infections, cellular heparan sulfate proteoglycans serve as additional attachment factors, reinforcing adhesion. Subsequent binding of penton bases to  $\alpha,\beta$ -integrin receptors induces clathrin-mediated endocytosis and leads to loss of viral fiber proteins.

The adenoviral replication cycle is divided into early, intermediate and late stages, with each seeing expression of a typical set of genes. Upon engulfment by the host cell and triggered by endosomal acidification, hexon bound protein VI disassociates from the capsid structure and induces lysis of the endosomal membrane. The remainder of the now cytosolic virion is shuttled to the nucleus by microtubular transport where viral protein IX recruits kinesin thereby procuring capsid disruption. Nuclear penetration is mediated by core protein VII and occurs at nuclear pore complexes (NPCs), leading to transcription of early viral genes by host RNA polymerase II. The resulting proteins manipulate various cellular processes, such as evasion of host immune response (E3), cell cycle (E1A), apoptosis (E1B and E4) and messenger RNA (mRNA) transport (E4), while also promoting viral DNA replication (E2). A virally encoded DNA

## 2.3. Select Viral Pathogens

polymerase replicates the genome by DNA strand displacement in a unique protein primed fashion involving precursor terminal protein (pTP) acting as primer and DNA-binding protein (DBP), as well as several host proteins. With onset of replication, translation of late genes by host RNA polymerase II is initiated, leading to the production of structural proteins and proteins required for virion assembly. Encapsidation occurs in the nucleus and progeny virions are released by lysis of the host cell.

**Epidemiology.** Adenoviruses are endemic and ubiquitous, causing an estimated 2–5% of all respiratory infections. Distribution is worldwide and incidence higher in the first half of the year. Children are frequently infected as serological studies show that by the age of 5 years some 50% present antibodies towards the most common species, including *Human adenovirus C*. On the order of 1 in every  $10^7$  lymphoid cells in the oropharynx harbor fully assembled latent state virions, making most humans latent carriers. Transmission can both occur via respiratory droplet or fecal-oral routes and virions are very stable towards chemical and physical agents, surviving for long periods outside their host.

Adenovirus outbreaks are a common cause of pneumonia in militaries, leading to the development of live, oral vaccines in the 1960's by the U.S. Army. The only supplier however ceased production and as of 1999, vaccinations could no longer be administered, resulting in reemerging incidence. In the first 5 years after loss of the vaccine, 110000 cases of febrile respiratory illness were reported, of which an estimated 90% are considered preventable. By October 2011, new vaccine again was available and has been administered to new recruits since.

### **2.3.2 Rhinoviruses**

Investigations into the etiological agent of the common cold within the British Common Cold Research Unit led to the discovery of rhinoviruses in 1953. Initially classified as a separate genus of the family *Picornaviridae*, recent genomic evidence led to a revised taxonomy, recognizing three species of rhinoviruses (A through C) as members of the genus *Enterovirus*. Over 100 distinct serotypes have been isolated from humans, 74 belong to species A, 25 to B and the newly identified species C, currently under active study, may encompass an additional 55 types.

Rhinovirus virions are small (30 nm in diameter), non-enveloped, with a pseudo  $T = 3$  icosahedral capsid, consisting of 4 different polypeptides (VP1, VP2, VP3 and VP4) surrounding the RNA genome. There are 60 copies of each structural protein and VP1-3 face towards the outside and are responsible for antigenic diversity, while VP4 faces inwards and anchors the RNA core to the capsid. A canyon formed by VP1 and VP3 serves as receptor binding site. The viral genome consists of monopartite, linear, single stranded, positive sense RNA,

## **2. BIOLOGICAL BACKGROUND**

---

roughly 7.2 kb in length and encodes a single polyprotein, which cleaved by viral proteases yields 11 proteins. Instead of a methylated 5' cap, the RNA genome is terminated by a viral protein (VPg) at its 5' end ([Jacobs et al., 2013](#)).

**Diseases.** Over half and up to two thirds of all cold-like illnesses are thought to be caused by rhinoviruses. In addition, asymptomatic infections, especially in children are very common with rates among children under 4 years ranging from 12 to 32%. These surprisingly high numbers may to some extent result from virus persistence in hosts that have recovered in addition to disease that has not broken out. In adults, rates of asymptomatic carriage are significantly lower, reported at 0–2%.

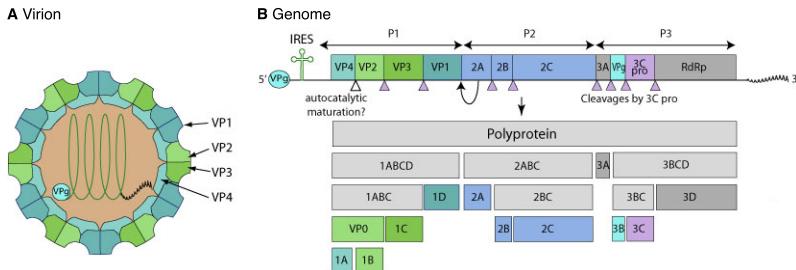
In immunocompetent individuals, symptomatic disease typically manifests as upper respiratory infection with clinical syndromes associated with common cold, including rhinorrhea, nasal congestion, sore throat, headache and possibly fever and general malaise. Disease is self-limited, incubation periods are between 12 and 72 h and within 7 to 14 days, symptoms wear off. A common complication is acute otitis media, which has been reported to happen in up to 30% of cases in early childhood. In 41–45% of investigated middle ear infections, rhinoviruses were detected. Further cavities that are frequently infected are the paranasal sinuses. Nose blowing has been suggested as mechanism for spreading the virus and causing rhinosinusitis.

Initially thought to primarily cause benign upper respiratory infections, recent data clearly implicates rhinoviruses in more severe lower respiratory infections. Croup, bronchiolitis and community-acquired pneumonia (CAP) have been associated with rhinovirus infections and studies have shown that in 12, 14 and 18–26% of respective cases, rhinoviruses were present. Furthermore, a study of children admitted to intensive care units (ICUs) with lower respiratory tract infections detected no other pathogens in 49% of the patients. Immunocompromised individuals are predisposed to contracting more serious forms of disease, with morbidity and mortality comparable to that of pandemic H1N1 influenza.

While not typically associated with cytopathogenic effects on epithelia of the upper respiratory tract, cell damage to lung tissue, especially among children, has been documented. Thus, rhinoviruses are linked to exacerbations of chronic pulmonary diseases such as asthma, chronic obstructive pulmonary disease and cystic fibrosis.

**Pathogenesis.** Members of rhinovirus species A and B are divided into two group according to their host cell receptors. Members of the major group form interactions with ICAM-1 while minor group types (including serotype 1a) associate with very low-density lipoprotein receptors (VLDLRs). Attachment of species C have very recently been identified as induced by cadherin-related family member 3 (CDHR3). Upon receptor mediated endocytosis, pH dependent conformational changes in capsid structure exposes VP4 which has pore-

## 2.3. Select Viral Pathogens



**Figure 2.9:** Capsid proteins VP1 through VP4 form a pseudo  $T = 3$  icosahedral coat, roughly 30 nm in diameter around the RNA genome (A), which is monopartite, linear, 7.2 kp long and encodes 11 proteins (B). Figures adapted from [Hulo et al. \(2011\)](#).

forming properties and facilitates release of viral genomic material into the cytosol.

Host cell ribosomes readily translate the released positive-sense RNA into polyprotein, which is cleaved in *cis* by 2A and 3Cpro, yielding preproteins P1, P2 and P3 (see figure 2.9). P1 is digested into structural capsid proteins while P2 and P3 are further processed to produce the replication machinery. Viral RNA-dependent RNA polymerase synthesizes a complementary, negative-sense RNA strand, primed by VPg, which in turn serves as template for many copies of the viral genome. These can be both translated into more protein and in a later stage of replication also be packaged into progeny virions. Final cell export is mediated by host-cell membrane lysis.

**Epidemiology.** Despite most infections with rhinoviruses only resulting in mild disease, economic impact both due to health care costs and loss of productivity is considerable. This is primarily owed to the overwhelming prevalence of these pathogens. Respiratory illnesses attributed to rhinoviral infections occur throughout the world and all year round with peak incidences in early fall and spring. Vaccination efforts so far have been futile, mainly because of the large number of serotypes and lack of individual epidemiological data.

Due to acid-sensitivity, fecal-oral infection is unlikely most person-to-person transmission occurs through aerosols and contact spread either direct or via fomites. Extra-host survival times of hours to days have been reported for indoor environments and up to 2 h on undisturbed skin.

### 2.3.3 Vaccinia

*Vaccinia virus* is a species within the genus *Orthopoxvirus*, alongside the exceptionally virulent *Variola virus*, the etiological agent of smallpox. Immunological similarities between the two species allows for cross-protection, which coupled

## 2. BIOLOGICAL BACKGROUND

with the large discrepancy in pathogenicity presents a fortunate opportunity for artificially inducing acquired immunity. This was recognized by Jenner in 1798, while investigating the zoonosis of *Cowpox virus* and subsequent susceptibility towards contraction of smallpox. The origins of vaccinia are unknown. It has been speculated to have derived from cowpox or smallpox, to be a hybrid of both and of being the prototype orthopoxvirus, as well as descending from an extinct ancestor.

The virions are large and brick shaped, measuring 200 by 250 by 340 nm and exist as both mature virion (MV) and extracellular virion (EV). Structurally they are unusually complex. The linear, double-stranded DNA genome, roughly 200 kb long, is encased in an S-shaped, tube-like nucleocapsid with an outside diameter of 50 nm, a cavity of 10 nm and an overall length of 250 nm. Additionally, 47 viral proteins, of which 16 are involved in early mRNA synthesis and 28 have no known enzymatic function, are packed into a core structure. The core wall consists of two layers, the palisade (outer) layer which is 17 nm thick and an inner smooth layer, measuring 8 nm across. Centered on the top and bottom faces, two proteinaceous lateral bodies separate core from the surface membrane, forming the virion core into a biconcave disc with dumbbell shaped vertical cross sections. The lipidic surface membrane also consist of two layers, the outermost measuring 9 nm and the innermost domain is 5 nm thick. EV virions are additionally wrapped in a membrane derived from Golgi cisternae ([Marennikova et al., 2005](#); [Condit et al., 2006](#)).

**Diseases.** While infection with variola causes severe disease manifesting in skin lesions covering the whole body, accompanied by 20–40% mortality rates, the closely related *Vaccinia virus* is far less invasive. Virulence of the latter pathogen is so low that it has been routinely used as live vaccine against the former. Owing to the massive effort put into the worldwide fight against smallpox led by the WHO in the late 1960's and throughout the 1970's, the disease was found to be eliminated by 1980. At the center of the smallpox eradication program was the administration of freeze-dried, calf lymph derived vaccinia with a bifurcated needle, by multiple pricking of the skin. Towards the end of the initiative, 200 million vaccinations were performed annually.

The predominant reaction to vaccinia inoculation is localized, self-limited disease. After an incubation period of 3–4 days, a papule with a sunken center develops at the site of infection, accompanied by erythema and swelling. Over the following days the papule increases in size and liquid begins to accumulate within. Fever may develop around days 7–10, possibly followed by malaise, headache and anorexia. Local lymphadenopathy is frequently encountered and days 8–10 typically mark the beginning of involution of the papule, which subsequently dries out and forms a scab.

Of great concern for routine vaccination procedures are complications which inevitably occur in a small numbers. Atypical side effects develop in roughly 1 per mill of cases and potentially life threatening reactions usually manifest

## 2.3. Select Viral Pathogens

as neurological (477.4 cases and 47.0 fatalities per 1 million) or cutaneous disease (278.4 cases and 0.2 fatalities per 1 million). Predisposing conditions for central nervous system involvement are not known and despite decades of inquiry into this pathology, it remains poorly understood. Symptoms range from febrile seizures to severe encephalitis, but so far no neuropathological characteristics have been identified. Complications affecting the skin and mucous membranes are classified as progressive vaccinia, eczema vaccinatum and generalized vaccinia and disease severity decreases in this order. Predisposing conditions for progressive and generalized vaccinia are immunodeficiencies while a history of eczema is a risk factor for eczema vaccinatum.

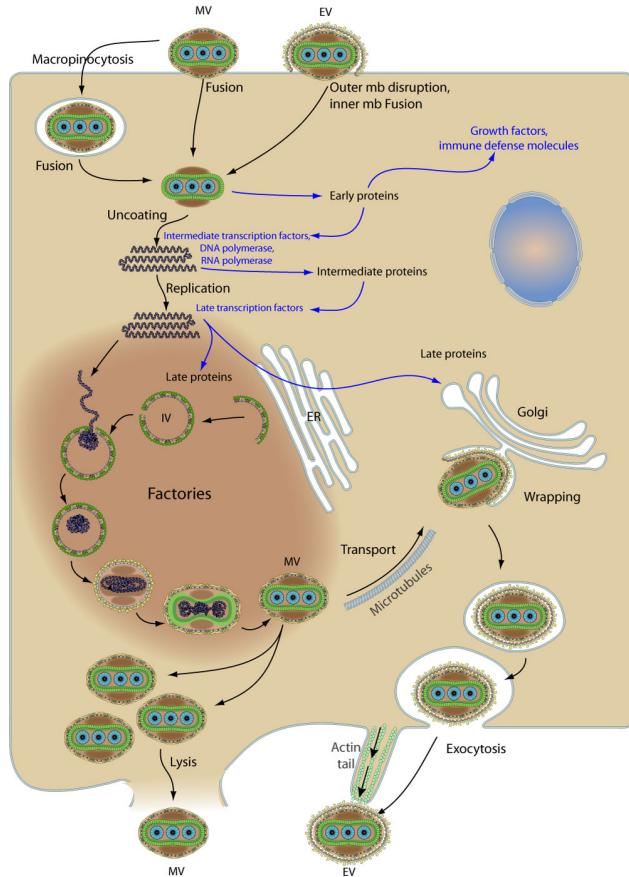
**Pathogenesis.** Initial attachment is mediated by interaction between viral proteins and cellular heparan sulfate chains. For cell entry, various strategies have been reported, depending on the strain under investigation. WR strains induce macropinocytosis and proteins A25/A26 act as fusion suppressors that only lift their embargo under acidic conditions encountered in a maturing endosome, while other strains, such as Copenhagen, present no A25/A26 on their outer membrane and fuse directly with the target plasma membrane. Due to the additional membrane of EVs, a differing entry mechanism needs to be employed. In a currently not well understood fashion, the outer membrane is lost by non-fusogenic disruption, followed by fusion of the inner virion membrane. All pathways lead to cytosolic localization of virions devoid of their envelope.

Members of the *Poxviridae* family are special among Baltimore group I viruses in that their genome encodes all necessary replicatory proteins, allowing for cytosolic localization. Replication is temporally tightly regulated and consists of distinct phases of early, intermediate and late gene expression. Each class of genes encodes factors capable of initiating the succeeding stage, providing transcription level regulation. Uncoating of the core structure releases early proteins, including RNA polymerases and enzymes for mRNA processing which start transcribing early genes. At least 50 different products, such as DNA replicatory enzymes, additional RNA polymerase, mRNA processing machinery, host defense molecules and intermediate gene transcription factors, have been identified and account for 25% of the viral coding capacity. Early gene transcripts are detectable 20 minutes after cell entry and reach their productive peak within 100 minutes of infection.

Expression of intermediate genes is initiated by accumulation of intermediate transcription factors and the onset of DNA replication. Only 7 products of this phase are known, which functionally are mostly concerned with host defense, DNA/RNA metabolism and commencement of the final phase. Beginning 100 minutes after infection, intermediate gene transcription reaches its peak at 120 minutes and is thereafter superseded by late gene transcription, beginning 140 minutes after cell entry. Products of the final phase comprise a large number of genes (up to 75% of the vaccinia genome) and include enzymes needed for initiating replication (RNA polymerase and modification proteins), early transcription factors and structural proteins, as well as virion assembly machinery.

## 2. BIOLOGICAL BACKGROUND

---



**Figure 2.10:** Capsid proteins VP1 through VP4 form a pseudo  $T = 3$  icosahedral coat, roughly 30 nm in diameter around the RNA genome (A), which is monopartite, linear, 7.2 kp long and encodes 11 proteins ([Hulo et al., 2011](#)).

DNA replication not only serves for progeny virions, but also to increase the concentration of templates used for gene expression. Both DNA synthesis and virion assembly occur within factories, readily visualized by electron microscopy as electron dense cytoplasmic inclusion bodies. Owing to the complex virion structure DNA packaging and virion assembly is an involved procedure with is incompletely understood.

**Epidemiology.** It is unknown if a natural reservoir of vaccinia exists. It has been speculated that the virus is maintained only within research laboratories and vaccination production facilities, while others have implicated some rodent species as possible reservoir hosts. Small scale zoonotic outbreaks of vaccinia have been documented in Brazil and it was initially suspected that these were linked to vaccine that had escaped into the environment. Recent phylogenetic studies however were able to rule out this explanation but were unable to shed further light into underlying epidemiological mechanisms.

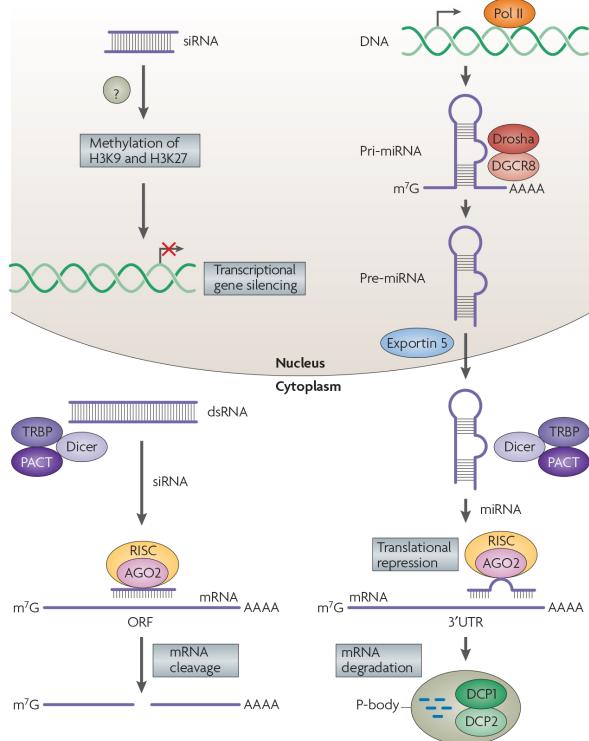
While transmission from vaccinees to unvaccinated individuals is rare, direct contact transmission is possible and such occurrences have been documented. Special care has to be taken to avoid direct contact between recently vaccinated and individuals predisposed towards developing complications.

### 2.4 RNA Interference

First described only two decades ago, regulation of gene expression by short strands of RNA has become an indispensable tool to both experimental biology and bioinformatics. Recognizing the importance of applications made possible by this discovery, the 2006 Nobel prize in Physiology or Medicine was awarded to Fire and Mello who studied RNA interference in the nematode worm *Caenorhabditis elegans* and published their findings in 1998. Building on studies by Guo and Kemphues, who showed that sense RNA, as well as antisense RNA was capable of suppressing gene expression, Fire, Mello and coworkers found that double stranded RNA was at least ten-fold more effective as silencing agent than individual single stranded fragments. Further investigations showed that several gene regulatory processes, previously thought to be unrelated, were in fact manifestations of RNA interference and that the underlying mechanism was conserved in many if not most eukaryotic organisms ([Hannon, 2002](#)).

The RNAi pathway can take as input two separate types of RNA molecules, miRNA and short interfering RNA (siRNA), of differing origins. While miRNAs are endogenous and purposefully employed in post-transcriptional regulation of gene expression, siRNAs are exogenous synthetic or viral inducers of gene suppression, in which case, RNA interference can be viewed as an immune response to foreign genetic material. Parsimony-based phylogenetic analysis of involved genes suggests that the key components to an RNAi system were already present in the last common ancestor of eukaryotes and were subsequently lost or extensivley simplified in some protists. The original function of RNAi is hypothesized to be that of a defense mechanism against genomic parasites as indicated by the extent of its conservation, whereas miRNA-directed silencing most probably was introduced at a later point in evolution ([Cerutti and Casas-Mollano, 2006](#)).

## 2. BIOLOGICAL BACKGROUND



**Figure 2.11:** RNA interference comprises of three distinct mechanisms that yield control over gene expression. Exogenous double-stranded RNA are processed into siRNA fragments that both act inside the nucleus as transcriptional silencing agents and in the cytoplasm, post-transcriptionally cleaving mRNA strands. Endogenous miRNA is synthesized by RNA polymerase, originates from the nucleus in processed form and mediates milder translational repression (Kim and Rossi, 2007).

### 2.4.1 Molecular Mechanism

RNA interference refers to three separate mechanisms for regulation of gene expression by small RNAs, as visualized by figure 2.11. While siRNAs are involved both in transcriptional and post-transcriptional gene silencing, the miRNA pathway is focused on translational repression. The severity of action on the targeted genes once again highlights the differing purposes of the siRNA and miRNA pathways, one tasked with inhibitory and the other with regulatory measures (Wilson and Doudna, 2013; Kim and Rossi, 2007; Carthew and Sontheimer, 2009).

## 2.4. RNA Interference

---

**Translational repression by miRNA.** The biogenesis of miRNA occurs in the nucleus and is initiated by RNA polymerase II transcription of long (>1000 nt) primary miRNA (pri-miRNA) segments, characterized by double-stranded hairpin loops with single-stranded 5'- and 3'-terminal overhangs which are polyadenylated and capped. Subsequent processing by the microprocessor complex consisting of the RNase III family enzyme Drosha and DGCR8 (DiGeorge syndrome critical region gene 8) yields ~60–70 nt stem-loop structured precursor miRNA (pre-miRNA) fragments. DGCR8 recognizes pri-miRNAs by the junction of stem and single-stranded overhang and helps positioning the substrate for endonucleolytic cleavage by Drosha at a site ~11 nt from the junction. Nuclear export is mediated by the transport facilitator exportin-5 and is Ran-GTP dependent.

In the cytoplasm, pri-miRNAs are targeted by Dicer and the associated dsRNA binding proteins TRBP (TRA RNA-binding protein) and PACT (protein activator of protein kinase PKR) for further processing into 21–25 nt dsRNA strands with 2 nt overhangs at the 3' termini and phosphate groups at each of the recessed 5' ends. The mature miRNAs are loaded onto Argonaute (Ago) by Dicer, which leads to the formation of RNA-induced silencing complex (RISC). Concomitantly with RISC-loading, one of the two RNA stands is selected as guide strand whereas its complement (the passenger strand) is ejected and degraded. Thermodynamic asymmetry between the two strands is exploited in this step and the strand with the less stable 5' end is preferred. As opposed to strand separation in siRNAs, the passenger strand is not cleaved but rather unwound by helicase activity, facilitated by imperfect sequence alignment.

Finally, active RISC, exposing the Argonaute-bound guide strand, interacts with the 3' untranslated region of mRNA targets and directs translational repression and mRNA degradation. Sequence homology between mRNA and the miRNA seed sequence (the first 2–6 or 2–8 nt from the 5' end) is critical while mismatched nucleotides towards the 3' end of the miRNA are readily tolerated. The extent of base-pairing influences the subsequent mechanism of silencing, ranging from direct target cleavage (perfect match) over deadenylation (followed by degradation), to nonendonucleolytic translational repression (imperfect match).

**Post-transcriptional gene silencing by siRNA.** Precursors to siRNA are long, linear, perfectly base-paired double stranded sequences of RNA, typically of exogenous origin either introduced directly into the cytoplasm, or taken up from the environment. A complex consisting of Dicer and several RNA-binding proteins are responsible for trimming down dsRNA fragments to the appropriate size for loading onto Ago2. Of the four Argonaute family members in humans, capable of associating with miRNA, only Ago2 seems to be involved with siRNA. Furthermore, Ago2 is the only mammalian Argonaute bearing endonucleolytic functionality and therefore capable of directly cleaving targeted mRNA.

## **2. BIOLOGICAL BACKGROUND**

---

Strand selection is again based on differences in stability of base-pairing at the 5' termini with the weaker interacting end being favored as guide strand. Accuracy of discrimination can be low, leading to incorporation of both strands with equal frequency. In contrast to miRNA loading, the passenger strand is not merely separated but directly cleaved by Ago2 and the differing treatment seems to only depend on perfect strand complementarity given in siRNA and absent in miRNA. Upon RNA incorporation, RISC is formed and activated by cleavage of the passenger strand. The 3' guide strand end is bound by the Argonaute's PIWI-Argonaute-Zwille domain (PAZ), while the 5' end interacts with Argonaute middle domain (MID), closely located to the catalytic RNase H-like P-element-induced whimp testes domain (PIWI).

Post-transcriptional gene silencing is accomplished by endonucleolysis of perfectly matching mRNA precisely at the phosphodiester linkage between bases 10 and 11 relative to the 5' terminus of the siRNA guide strand. Following cleavage, the target dissociates, freeing RISC for further catalysis, and the mRNA fragments are degraded by cellular exonucleases. Imperfectly matched mRNA may be targeted, much like it is the case with miRNA, leading to siRNA off-target effects which are of great practical importance.

**Transcriptional gene silencing by siRNA.** In addition to post transcriptional action of siRNA, nuclear inhibition of gene transcription has been described in many eukaryotes. Diced siRNA fragments are transported into the nucleus where they are assembled with a group of proteins, including Ago1, to form RNA-induced transcriptional silencing complex (RITS). Currently only incompletely understood, the siRNA guide strand is thought to recognize RNA transcripts as they emerge from RNA polymerase II, followed by recruitment of factors that enable covalent modifications of nearby histones. Methylation of lysines 9 and 27 in H3 by histone methyltransferases leads to chromatin compaction and heterochromatin formation. RITS has also been shown to induce direct methylation of DNA, repressing gene expression even further.

Contributing to the potency of RNA interference, engagement of RITS with nascent transcripts activates the RNA-dependent RNA polymerase complex (RDRC), capable of generating secondary siRNA fragments and therefore amplifying silencing capabilities. The role of this reinforcement mechanism has been firmly established in many eukaryotic RNAi systems with the notable exceptions of vertebrates and insects. Whether a similar system exists in these organisms remains an open question.

### **2.4.2 Biological Function**

The mechanisms of RNA interference have most probably evolved in order to protect against foreign genetic material such as parasitic DNA sequences or viral RNA. Transposable elements (transposons) are DNA sequences that are mobile within the genome, can make up a significant fraction of eukaryotic

genomes and are typically considered non-coding. Transposition is mediated by transposases, enzymes often encoded within the transposons themselves, that act on specific sequences at the transposon ends and cause unspecific insertion into new target sites. Retrotransposons move via an RNA intermediate which is reverse transcribed to DNA and inserted, while DNA transposons employ a cut and paste mechanism. Retroviruses therefore can be viewed as transposons and in general, transposable elements are a form of selfish DNA that often incur deleterious effects.

RNAi is an important regulatory force to transposon activity, both by processing transcripts of retrotransposons, thereby reducing their concentration and eliciting epigenetic modifications, as well as transcriptional inhibition via heterochromatin formation. The importance of keeping transposable elements in check is highlighted by their prevalence, with roughly half of the human genome being thought to derive thereof.

Antiviral mechanisms are particularly important in organisms lacking an adaptive immune system as found in vertebrates and exploiting the orthogonality of most genomic systems to double-stranded RNA puts RNA interference in a powerful position. Corroborating this notion is the observation that, in *Drosophila melanogaster*, three key proteins of the RNAi pathway (Dicer-2, Ago2 and R2D2, a protein involved in RISC loading) are among the top 3% in terms of genetic instability. Furthermore, miRNA pathway paralogs to these three genes (Dicer-1, Ago1 and R3D1), not being involved in immune response, evolve at a much slower rate ([Obbard et al., 2009](#)).

Although small RNA-guided, Ago-dependent up-regulation of gene expression (termed RNA activation or RNAa) has been described, most regulation of gene expression by miRNA is of inhibitory nature. This widespread mechanism, consisting of >1000 miRNA sequences (as much as 5% of the human genome) controls at least 30% of human genes and is responsible for vital processes including cell growth, tissue differentiation and cell proliferation.

### 2.4.3 Applications

In *C. elegans*, RNA interference is especially powerful, making it a popular model organism for RNAi. Not only is delivery efficiently possible simply by feeding the nematodes with bacteria such as *E. coli* that carry the desired dsRNA, but the resulting gene silencing effects are hereditary. Moreover, RNAi response is not stoichiometric but catalytic, is amplified in a feedback loop and in many organisms, systemic spread has been documented.

The initial burst of excitement surrounding possible applications was somewhat moderated by difficulties in applying RNAi to mammalian systems. At first it seemed impossible to use this technology in somatic cells as the introduction of dsRNA is typically met with overwhelming non-specific responses, including RNA-dependent protein kinase pathway (PKR), which leads to arrest of translation and apoptosis. This issue was shown to be overcome by

## 2. BIOLOGICAL BACKGROUND

exclusively using siRNAs duplexes of 21–23 nt with 2-nt 3' overhangs that mimic Dicer products and are too short for inducing PKR. Mammalian RNAi response, however is transient, lacking amplification and spreading mechanisms documented in other organisms (mainly plants and worms) and delivery, especially *in vivo* remains problematic. A further issue that continues to be an actively researched area of interest is that of off-target effects (OTE), which considerably complicate the interpretation of phenotypic data.

**Gene knockdown studies.** Large-scale loss-of-function (LOF) and modifier or synthetic lethality screens are readily possible by means of RNAi based high throughput screening (HTS). Such experiments usually proceed by arraying libraries of gene specific siRNAs onto microtiter plates (96 and 384 well formats are common), followed by the addition of liquid cell cultures. After an appropriate transfection time, the cells may be subjected to an additional treatment, such as exposure to drugs or pathogens (modifier screen) or LOF phenotypes can be investigated directly. Assay readout is performed via optical measurements such as detection of fluorescence or luminescence signals or by microscopic imaging (confocal or wide-field).

Transcriptional reporters, fluorescent dyes that detect enzymatic activity and protein-modification specific antibodies have been employed in plate reader-based investigations which yield a single numerical readout per well. This quantitative approach is contrasted with microscopy based assay read-outs that are able to capture spatial information on antibody stained proteins, fluorescently labeled cellular structures and green fluorescent protein (GFP) expression, yielding much more data per well. Significant challenges incurred by automated high-content imaging have successfully been addressed by computational image analysis software.

A multitude of sources of technical and biological noise contribute to serious problems in interpretability and comparability of observed data. Common to all high throughput approaches, errors arise from difficulties guaranteeing equal conditions in a large number of parallel experiments. Liquid dispensing errors, temporal disparities caused by bottleneck resources such as imaging equipment and spatial discrepancies, for example inhomogeneous temperature distribution over the plate, are only a few issues that come to mind. Biological sources of error include OTEs, varying potency of reagents (both the knockdown strength and time required to achieve optimal knockdown may be affected) and obscuration of assay phenotype by the knockdown phenotype (e.g. cell death). Furthermore, incorrect gene models lead to errors in library design and detection may be hampered by weak phenotypes. Replicates, although expensive in large-scale experiments and control wells embedded in every assay plate are indispensable measures in order to assess reproducibility of the data ([Echeverri and Perrimon, 2006](#); [Perrimon and Mathey-Prevot, 2007](#)).

Despite being a young technology, RNA interference has already proven itself as an invaluable tool and has yielded many insights with significant impact

## 2.4. RNA Interference

---

on various fields. A review by Mohr et al. (2010) lists some of these findings which lead to refined understanding of cell proliferations, cancer biology, cell cycle regulation, mitochondrial diseases, signal transduction, RNA biology and pathogen response.

**Biotechnological applications.** Intercellular, systemic spread of RNAi response in plants and even its heredity over several generations have been documented and it comes as no surprise that the technology is investigated for possible utilization in crop improvement. Removal of plant endotoxins by targeting genes of toxin biosynthesis has been accomplished, leading to the production of decaffeinated coffee plants (knockdown of theobromine synthase), tobacco with reduced concentration of carcinogenic compounds (inhibition of nicotine demethylase activity) and edible cotton seeds (reduction of delta-cadinene synthase leads to low levels of gossypol, a toxic terpenoid), which are naturally rich in dietary protein.

In addition to investigations with consumer health in mind, improvements in environmental resistance have been studied in many organisms. Susceptibility to bacterial and viral pathogen infection has been reduced in rice, bean, barley and lemon, while fungal resistance has been increased in potato, tobacco and wheat. Successful RNAi application as insecticide has been shown in cotton and maize and even improved resistance to parasitic weeds could be demonstrated in transgenic tomato plants. Despite these achievements, concerns over environmental issues and adverse health effects have so far prevented RNAi based genetic modifications from exiting experimental stages (Saurabh et al., 2014).

**Therapeutic potential.** Great promise lies in therapeutic application of RNAi as theoretically, any gene can be targeted, yielding unparalleled flexibility not encountered with typical small molecule drugs. An initial obstacle to harnessing this power in humans is the issue of delivery. Systemic spread of naked siRNAs is hampered by kidney filtration, phagocytic uptake and degradation by serum RNases. Movement across capillary walls is not readily possible for molecules larger than 5 nm and phagocytes patrol the extracellular matrix, ingesting foreign genetic material. Furthermore, polyanionic macromolecules do not easily penetrate hydrophobic cellular membranes.

Topical or local administration offers advantages, including increased bioavailability and reduced side effects and is therefore preferred for treatment of eye, skin and mucosal diseases, as well as localized tumors. If targets are not localized or difficult to reach, injection into the bloodstream may provide a mode of systemic application. Chemical modification of the RNA backbone (2'-O-methylation or 2'-fluorination of ribose) has been shown to provide resistance to RNase and covalent attachment to cholesterol promotes cellular uptake. Encapsulation of siRNA in liposomes and cationic polymers are further proven

## 2. BIOLOGICAL BACKGROUND

---

**Table 2.2:** A non exhaustive list of RNAi based drugs that currently are in clinical trials. The data was obtained from ?

Company	Disease	Delivery system	Status
Alnylam	(TTR)-Mediated Amyloidosis	siRNA-GalNAc conjugate	Phase III recruiting
Alnylam	Antitrypsin Deficiency Liver Disease	Liposome	Phase II recruiting
Alnylam	Acute Intermittent Porphyria	siRNA-GalNAc conjugate	Phase I recruiting
Silenseed	Advanced pancreatic cancer	Polymer (LODER)	Phase III planned
Sylentis	Ocular pain	Naked siRNA	Phase II completed
Sylentis	Open angle glaucoma	Naked siRNA	Phase II recruiting
Tacere	Chronic hepatitis C	AAV vector	Phase II recruiting
Tekmira	Cancers	liposome	Phase I active

techniques for improving extracellular stability, stimulate endocytosis and facilitate endosomal escape.

Some sort of target selectivity is desirable on order to avoid high dosages, associated concerns of toxicity and potential OTEs in non-target tissue. Coupling of siRNA reagents to antibodies specific for HIV envelope glycoproteins has been successfully employed for selectively entering infected cells, while aptamers (oligonucleotides specifically engineered for binding a given target), carrying siRNAs have also been shown to be capable homing mechanisms.

Viral delivery of of RNAi inducing agents presents an alternative technique to the above and in case of retroviral transport vessels, short hairpin RNAs (shRNAs) that are reverse transcribed and integrated into the host genome, provide stable expression and prolonged RNAi activity. Adenoviral and adeno-associated virus (AAV) vectors have also been employed, resulting in a more transient response. While health concerns associated with perpetuity of gene therapy no longer apply, repeated administrations may prove problematic, triggering strong immune response and thereby limiting therapeutic potential.

Currently, multiple siRNA based therapeutics are in clinical trials, including stage III studies (see table 2.2). Due to the unprecedented pace at which RNAi technology went from discovery to development of applications, much uncertainty remains surrounding long term effects of exposure to such drugs. Chronic diseases including hepatitis C or HIV infections require life-long treatments and consequences of repetitively triggering RNAi response has not been thoroughly studied (unwanted changes to chromatin structure for example, are one area of concern). Apart from safety issues, several implementation aspects

---

#### 2.4. RNA Interference

require further study. While neurodegenerative diseases have successfully been treated in mouse models via direct injection into the brain, this is not easily feasible in human patients and currently no delivery vehicle capable of crossing the blood-brain barrier. ([Kim and Rossi, 2007](#); [Whitehead et al., 2009](#)),



## Chapter 3

---

# Mathematical Background

---

Modeling the relationship among variables is one of the most important applications of statistical theory. The study of regression analysis (and the closely related notion of correlation) started to form towards the end of the 19th century with Sir Francis Galton's study of height heredity in humans and his observation of regression towards the mean. Over the next few years, Udny Yule and Karl Pearson cast the developed concepts into precise mathematical formulation, in turn building on work performed by Adrien-Marie Legendre and Carl Friedrich Gauss who developed the method of least squares almost a century earlier ([Allen, 1997](#)).

A multiple linear regression model can be written in matrix-vector form as

$$y = X\beta + \varepsilon \quad (3.1)$$

where  $y \in \mathbb{R}^n$  is the vector of observations on the dependent variable, the design matrix  $X \in \mathbb{R}^{n \times p}$  contains data on the independent variables,  $\beta \in \mathbb{R}^p$  is the  $p$ -dimensional parameter vector and the error term  $\varepsilon \in \mathbb{R}^n$  captures effects not modeled by the regressors. Without loss of generality, all variables are assumed to be expressed as deviations from their means and measured on the same scale.

In order to find unknown coefficients  $\beta_i$ , the ordinary least squares estimator minimizes the residual sum of squares, the squared differences between observed responses and their predictions according to the linear model.

$$\hat{\beta} = \arg \min_{\beta} \|y - X\beta\|^2 \quad (3.2a)$$

$$= (X^T X)^{-1} X^T y \quad (3.2b)$$

Some assumptions are typically associated with linear regression models that yield desirable attributes for the estimates. None of these restrictions are im-

### 3. MATHEMATICAL BACKGROUND

---

posed on the explanatory variables; they can be continuous or discrete and combined as well as transformed arbitrarily. Furthermore, in practice, it is irrelevant whether the covariates are treated as random variables or as deterministic constants. With exception of the field of econometrics it appears that the majority of literature adheres to the latter interpretation and therefore, statements will not explicitly be conditional on covariate values.

**Linearity.** The relationship between dependent and independent variables should be linear (after suitable transformations) and individual effects additive. If this cannot be satisfied, a linear model is not suitable.

**Full rank.** For the matrix  $X^T X$  to be invertible, it has to have full rank  $p$ . Therefore  $n \leq p$  and all covariates must be linearly independent.

**Exogeneity.** All independent variables should be known exactly i.e. contain no measurement or observation errors as only the mean squared error of the dependent variable is minimized. Additionally, all important causal factors have to be included in the model. Exogeneity implies  $\mathbf{E}[\varepsilon_i] = 0 \ \forall i$ , as well no correlation between regressors and error terms (Hayashi, 2000).

**Spherical errors.** This includes both homoscedasticity or constant error variance:  $\mathbf{E}[\varepsilon_i^2] = \sigma^2 \ \forall i$  and uncorrelated errors  $\mathbf{E}[\varepsilon_i \varepsilon_j] = 0 \ \forall i \neq j$ . These two conditions can be written more compactly as  $\text{Var}[\varepsilon] = \sigma^2 I_{n \times n}$ .

**Normality.** For the estimated coefficients to gain some additional desirable characteristics, it can be required that the errors  $\varepsilon_i$  be jointly normally distributed. Together with the above restrictions on expectation and variance, this yields  $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_{n \times n})$ .

Violations of these assumptions have varying consequences. In case of perfect multicollinearity, the ordinary least squares estimator  $\hat{\beta}$  as defined in (3.2b) does not exist. Recovering such a situation is possible by using a generalized matrix inverse (for example the Moore–Penrose pseudoinverse) or employing a regularization scheme such as ridge regression.

Omitting a variable that is both correlated with dependent variables and has an effect on the response (a nonzero true coefficient) will introduce bias in the parameters. The method of instrumental variables can help to produce an unbiased estimator.

The assumption of spherical errors ensures that the least squares estimator is the best linear unbiased estimator in the sense that it has minimal variance among all linear unbiased estimators. Heteroscedasticity and autocorrelation do not cause coefficient estimates to be biased but can introduce bias in OLS estimates of variance, causing inaccurate standard errors. A generalized least squares estimator (for example weighted least squares)

## Chapter 4

---

# Data

---

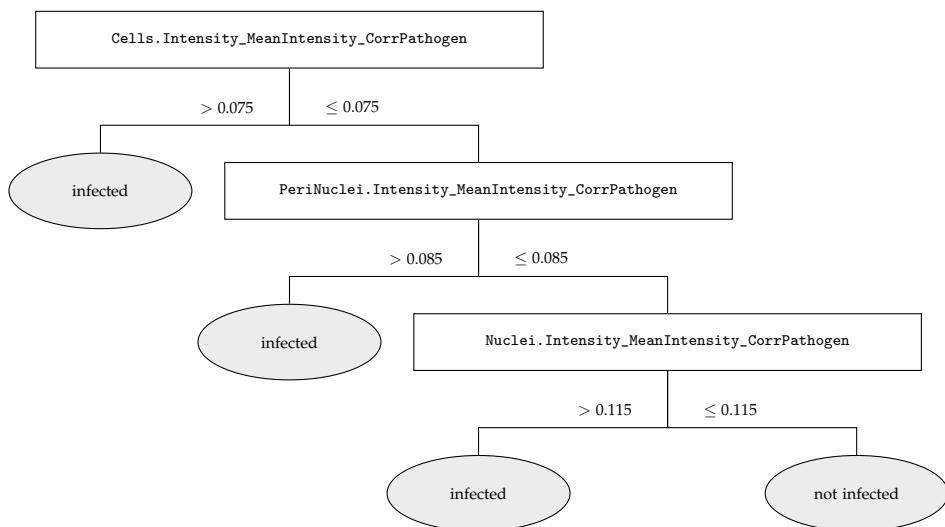
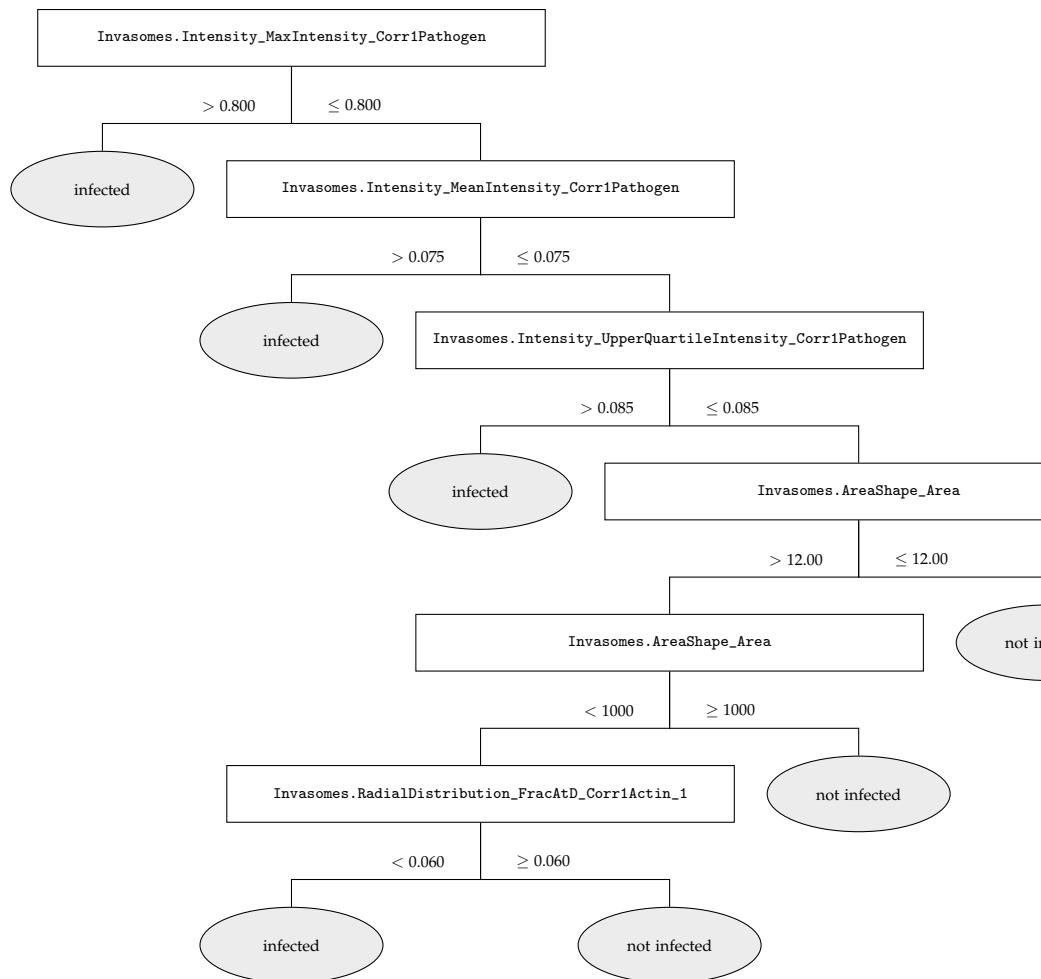


Figure 4.1: Decision tree for adenovirus infection scoring.

#### 4. DATA

---



**Figure 4.2:** Decision tree for adenovirus infection scoring.

## Appendix A

---

### R Package singleCellFeatures

---

Additional material. For example long mathematical derivations could be given in the appendix. Or you could include part of your code that is needed in printed form. You can add several Appendices to your thesis (as you can include several chapters in the main part of your work).

```
facetBorder <- function(x, y, img, facet) {
  facet.size <- img / facet
  # calculate facets (2d binning)
  x.bin <- ceiling(x / facet.size[1])
  y.bin <- ceiling(y / facet.size[2])
  # initialize empty grid/border matrices
  grid <- matrix(0, facet[2], facet[1])
  # calculate col-major grid index for each object
  index <- y.bin + (x.bin - 1) * facet[2]
  # summarize as counts
  counts <- table(index)
  # fill grid with counts
  grid[as.numeric(names(counts))] <- counts
  grid.res <- grid
  grid <- grid > 0
  # extend grid with a frame of ones
  grid.ext <- rbind(rep(1, (facet[1] + 2)),
    cbind(rep(1, facet[2]), grid, rep(1, facet[2])),
    rep(1, (facet[1] + 2)))
  # set up stencil
  row <- rep(rep(1:facet[2], facet[1]))
  col <- rep(1:facet[1], each=facet[2])
  colP1 <- col + 1
  colM1 <- col - 1
  rowP1 <- row + 1
  rowP2 <- row + 2
  nrowP <- facet[2] + 2
  stencil <- cbind(row + colM1 * nrowP, # northwest neighbor
    row + col * nrowP, # north neighbor
```

## A. R PACKAGE SINGLECELLFEATURES

---

```

    row + colP1 * nrowP, # northeast neighbor
    rowP1 + colP1 * nrowP, # west neighbor
    rowP2 + colP1 * nrowP, # east neighbor
    rowP2 + col * nrowP, # southwest neighbor
    rowP2 + colM1 * nrowP, # south neighbor
    rowP1 + colM1 * nrowP) # southeast neighbor
# apply stencil row-wise to grid
border <- apply(stencil, 1, function(ind, mat) {
  return(sum(mat[ind]))
}, as.numeric(grid.ext))
# map col-major object index to border array
return(border[index])
}

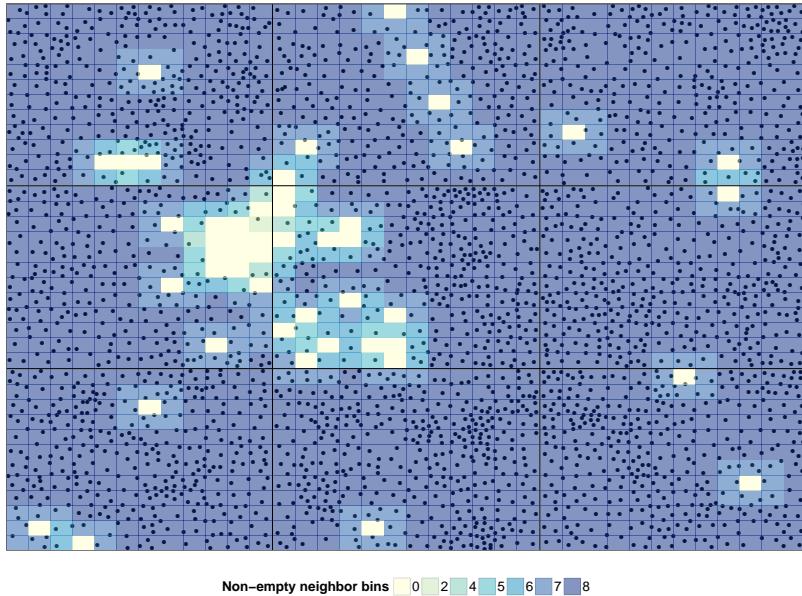
```

```

edgepos <- function(x, y, img, n) {
  empty <- logical()
  xst <- img[1] / n
  yst <- img[2] / n
  sgrid <- matrix(0, nrow=n, ncol=n)
  for (i in 1:n) {
    for (j in 1:n) {
      ispos <- (x > (i - 1) * xst) & (x <= (i * xst)) &
        (y > (j - 1) * yst) & (y <= (j * yst))
      sgrid[i, j] <- sum(ispos)
    }
  }

  for (i in 1:n) {
    for (j in 1:n) {
      ispos <- (x > (i - 1) * xst) & (x <= (i * xst)) &
        (y > (j - 1) * yst) & (y <= (j * yst))
      isempty <- F
      if ((i > 1) && (j > 1) && (sgrid[i - 1, j - 1] == 0))
        isempty <- T
      else if ((i > 1) && (sgrid[i - 1, j] == 0))
        isempty <- T
      else if ((i > 1) && (j < n) && (sgrid[i - 1, j + 1] == 0))
        isempty <- T
      else if ((j > 1) && (sgrid[i, j - 1] == 0))
        isempty <- T
      else if ((j < n) && (sgrid[i, j + 1] == 0))
        isempty <- T
      else if ((i < n) && (j > 1) && (sgrid[i + 1, j - 1] == 0))
        isempty <- T
      else if ((i < n) && (sgrid[i + 1, j] == 0))
        isempty <- T
      else if ((i < n) && (j < n) && (sgrid[i + 1, j + 1] == 0))
        isempty <- T
      empty[ispos] <- isempty
    }
  }
  return(empty)
}

```



**Figure A.1:** Cell colony edges are detected by 2D binning of cell center locations. Dots represent cell centers within the well H6 of plate J107-2C. Each of the nine images is segmented into 12 horizontal and 12 vertical sections yielding 144 tiles (1296 bins for the entire well). The tiles are colored according to the number of non-empty neighboring bins.

Times for my version are mean 7.902 ms (with sd 2.744 ms) and total runtime for a plate is 3.034 s, while theirs runs with mean 529.369 ms (sd 22.81 ms) and for a plate 203.278 s.



## Appendix B

---

### Yet another appendix....

---

#### B.1 Description

**Something** details.

**Something else** other definition.

#### B.2 Tables

Refer to Table B.1 to see a left justified table with caption on top.

Table B.1: Results.	
Student	Grade
Marie	6
Alain	5.5
Josette	4.5
Pierre	5



---

## Bibliography

---

- Alberts, B., J. Wilson, and T. Hunt (2008). *Molecular biology of the cell*. New York: Garland Science.
- Allen, M. P. (1997). The origins and uses of regression analysis. In *Understanding Regression Analysis*, pp. 1–5. Springer US.
- Anderson, B. and M. Neuman (1997, apr.). *Bartonella* spp. as emerging human pathogens. *Clin. Microbiol. Rev.* 10(2), 203–219.
- Atluri, V. L., M. N. Xavier, M. F. de Jong, A. B. den Hartigh, and R. M. Tsolis (2011, sep.). Interactions of the Human Pathogenic *Brucella* Species with Their Hosts. *Annual Review of Microbiology* 65(1), 523–541.
- Baltimore, D. (1971, sep.). Expression of animal virus genomes. *Microbiol. Mol. Biol. Rev.* 35(3), 235–241.
- Carthew, R. W. and E. J. Sontheimer (2009, feb.). Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136(4), 642–655.
- Celli, J. (2015, jul.). The changing nature of the *Brucella*-containing vacuole. *Cellular Microbiology* 17(7), 951–958.
- Cerutti, H. and J. A. Casas-Mollano (2006, aug.). On the origin and functions of RNA-mediated silencing: from protists to man. *Current genetics* 50(2), 81–99.
- Condit, R. C., N. Moussatche, and P. Traktman (2006, jan.). In a nutshell: structure and assembly of the vaccinia virion. *Advances in virus research* 66, 31–124.
- Craighead, J. E. (2000). *Pathology and Pathogenesis of Human Viral Disease*. San Diego: Academic Press.
- Croxen, M. A. and B. B. Finlay (2010, jan.). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Micro* 8(1), 26–38.

## BIBLIOGRAPHY

---

- Czyż, D. M., L.-P. Potluri, N. Jain-Gupta, S. P. Riley, J. J. Martinez, T. L. Steck, S. Crosson, H. A. Shuman, and J. E. Gabay (2014, jan.). Host-directed antimicrobial drugs with broad-spectrum efficacy against intracellular bacterial pathogens. *mBio* 5(4), e01534–14.
- Dehio, C. (2005, aug.). Bartonella-host-cell interactions and vascular tumour formation. *Nat Rev Micro* 3(8), 621–631.
- Dobson, A. P. and E. R. Carper (1996, feb.). Infectious Diseases and Human Population History. *BioScience* 46(2), 115–126 CR – Copyright © 1996 Oxford Universi.
- Echeverri, C. J. and N. Perrimon (2006, may.). High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet* 7(5), 373–384.
- Fàbrega, A. and J. Vila (2013, apr.). Salmonella enterica serovar Typhimurium skills to succeed in the host: virulence and regulation. *Clinical microbiology reviews* 26(2), 308–41.
- Farber, J. M. and P. I. Peterkin (1991, sep.). Listeria monocytogenes, a food-borne pathogen. *Microbiological Reviews* 55(3), 476–511.
- Haglund, C. M. and M. D. Welch (2011, oct.). Pathogens and polymers: Microbe–host interactions illuminate the cytoskeleton. *The Journal of Cell Biology* 195(1), 7–17.
- Hannon, G. J. (2002, jul.). RNA interference. *Nature* 418(6894), 244–251.
- Haraga, A., M. B. Ohlson, and S. I. Miller (2008, jan.). Salmonellae interplay with host cells. *Nat Rev Micro* 6(1), 53–66.
- Harms, A. and C. Dehio (2012, jan.). Intruders below the radar: molecular pathogenesis of *Bartonella* spp. *Clinical microbiology reviews* 25(1), 42–78.
- Hawn, T. R., J. A. Shah, and D. Kalman (2015, mar.). New tricks for old dogs: countering antibiotic resistance in tuberculosis with host-directed therapeutics. *Immunological Reviews* 264(1), 344–362.
- Hayashi, F. (2000). *Econometrics*. Princeton, NJ [u.a.]: Princeton Univ. Press.
- Hulo, C., E. de Castro, P. Masson, L. Bougueleret, A. Bairoch, I. Xenarios, and P. Le Mercier (2011, jan.). ViralZone: a knowledge resource to understand virus diversity. *Nucleic acids research* 39(Database issue), D576–82.
- Jacobs, S. E., D. M. Lamson, K. St George, and T. J. Walsh (2013, jan.). Human rhinoviruses. *Clinical microbiology reviews* 26(1), 135–62.
- Kahneman, D. (2012, mar.). Quotation: Kahneman on Contingencies. *Journal of the Experimental Analysis of Behavior* 97(2), 182.
- Kim, D. H. and J. J. Rossi (2007, mar.). Strategies for silencing human disease using RNA interference. *Nat Rev Genet* 8(3), 173–184.

---

## Bibliography

- Kopecko, D. J., L. Hu, and K. J. M. Zaal (2001, aug.). *Campylobacter jejuni – microtubule-dependent invasion.* *Trends in Microbiology* 9(8), 389–396.
- Lebreton, A., F. Stavru, and P. Cossart (2015, jun.). *Organelle targeting during bacterial infection: insights from Listeria.* *Trends in cell biology* 25(6), 330–8.
- Leirião, P., C. D. Rodrigues, S. S. Albuquerque, and M. M. Mota (2004, dec.). *Survival of protozoan intracellular parasites in host cells.* *EMBO Reports* 5(12), 1142–1147.
- Lenaerts, L., E. De Clercq, and L. Naesens (2008, nov.). *Clinical features and treatment of adenovirus infections.* *Reviews in Medical Virology* 18(6), 357–374.
- Littman, R. J. (2009, oct.). *The Plague of Athens: Epidemiology and Paleopathology.* *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine* 76(5), 456–467.
- Majowicz, S. E., J. Musto, E. Scallan, F. J. Angulo, M. Kirk, S. J. O'Brien, T. F. Jones, A. Fazil, and R. M. Hoekstra (2010, mar.). *The global burden of non-typhoidal Salmonella gastroenteritis.* *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 50(6), 882–9.
- Marennikova, S. S., R. C. Condit, and R. W. Moyer (2005). *Vaccinia Virus.* In *Orthopoxviruses Pathogenic for Humans SE - 3*, pp. 19–87. Springer US.
- Mohr, S., C. Bakal, and N. Perrimon (2010). *Genomic Screening with RNAi: Results and Challenges.* *Annual review of biochemistry* 79, 37–64.
- Obbard, D. J., K. H. J. Gordon, A. H. Buck, and F. M. Jiggins (2009, jan.). *The evolution of RNAi as a defence against viruses and transposable elements.* *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364(1513), 99–115.
- Perrimon, N. and B. Mathey-Prevot (2007, jan.). *Applications of High-Throughput RNA Interference Screens to Problems in Cell and Developmental Biology.* *Genetics* 175(1), 7–16.
- Prussia, A., P. Thepchatri, J. P. Snyder, and R. K. Plemper (2011). *Systematic Approaches towards the Development of Host-Directed Antiviral Therapeutics.* *International Journal of Molecular Sciences* 12(6), 4027.
- Rolain, J. M. and D. Raoult (2006). *Bartonella spp.* In *Principles and Practice of Clinical Bacteriology*, pp. 295–304. John Wiley & Sons, Ltd.
- Saurabh, S., A. Vidyarthi, and D. Prasad (2014). *RNA interference: concept to reality in crop improvement.* *Planta* 239(3), 543–564.
- Schroeder, G. N. and H. Hilbi (2008, jan.). *Molecular pathogenesis of Shigella spp.: controlling host cell signaling, invasion, and death by type III secretion.* *Clinical microbiology reviews* 21(1), 134–56.

## BIBLIOGRAPHY

---

- Silver, L. L. (2011, jan.). Challenges of Antibacterial Discovery. *Clinical Microbiology Reviews* 24(1), 71–109.
- Smith, A. (2012). Cellular Biochemistry II: Cellular Infection.
- Taubenberger, J. and D. M. Morens (2006). 1918 Influenza: the Mother of All Pandemics. *Emerging Infectious Diseases Volume* 12(1).
- Treadgold, W. (1997, jun.). *A History of the Byzantine State and Society*. Stanford: Stanford University Press.
- Whitehead, K. A., R. Langer, and D. G. Anderson (2009, feb.). Knocking down barriers: advances in siRNA delivery. *Nat Rev Drug Discov* 8(2), 129–138.
- Wilson, R. C. and J. A. Doudna (2013, may.). Molecular Mechanisms of RNA Interference. *Annual Review of Biophysics* 42(1), 217–239.
- World Health Organization (2003). Fact sheet No. 211: Influenza.
- World Health Organization (2012). Mortality and global health estimates: Causes of death.
- World Health Organization (2014). Antimicrobial resistance: global report on surveillance 2014. Technical report, World Health Organization.
- Zietz, B. P. and H. Dunkelberg (2004). The history of the plague and the research on the causative agent Yersinia pestis. *International Journal of Hygiene and Environmental Health* 207(2), 165–178.

---

## **Epilogue**

---

A few final words. Test 2



## Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

---

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

**Title of work** (in block letters):

**Authored by** (in block letters):

*For papers written by groups the names of all authors are required.*

**Name(s):**

**First name(s):**

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the '[Citation etiquette](#)' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

**Place, date**

**Signature(s)**

*For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.*