

1 **'Add, stir and reduce': The *Yersinia* as model bacteria for the evolution of**
2 **mammalian pathogens**

3

4 Alan McNally ^{1*}, Nicholas R Thomson^{2,3}, Sandra Reuter^{3,4} & Brendan W Wren²

5

6

7 ¹Pathogen Research Group, Nottingham Trent University, Clifton Lane,

8 Nottingham NG11 8NS

9 ²Department of Pathogen Molecular Biology, London School of Hygiene and

10 Tropical Medicine, Keppel Street, London, WC1E 7HT

11 ³Pathogen Genomics, Wellcome Trust Sanger Institute, Hinxton, Cambs CB10 1SA

12 ⁴ Department of Medicine, University of Cambridge, Box 157 Addenbrooke's

13 Hospital, Hills Road, Cambridge CB2 0QQ

14

15 **Key Points:**

16 • The evolution of mammalian pathogenesis in the *Yersinia* genus has
17 occurred in a parallel fashion via a balanced mixture of gene gain and gene loss
18 events

19 • Only by sequencing pathogenic and non pathogenic representatives
20 positioned across an entire bacterial genus can such observations be made

21 • The parallel nature of the evolution of pathogenesis extends beyond
22 *Yersinia* and into other enteric pathogens, notably the salmonellae

23 • Gene loss events lead to niche restriction due to a reduction in metabolic
24 flexibility, often seen in the more acutely pathogenic members of a lineage

- 1 • Potential negative impacts of gene gain events are offset by the
- 2 transcriptional silencing or fine control of these acquired elements by ancestral
- 3 regulons such as RovA and H-NS.

1 **Abstract**

2 In the study of molecular microbiology and bacterial genetics, pathogenic species
3 of the *Yersinia* genus have been pillars for research aimed at understanding how
4 bacteria evolve into mammalian pathogens. The advent of large-scale population
5 genomic studies has hugely accelerated progress in this field, and the pathogenic
6 *Yersinia* species have re-emerged as model organisms to help shape our
7 understanding of the evolutionary processes involved in pathogenesis. In this
8 review, we highlight how microbial genomic studies of the yersinia have
9 revealed distinct features marking the evolutionary path towards pathogenesis,
10 which are changing our understanding of pathogen evolution. As these features
11 are also found in the genomes of other members of the *Enterobacteriaceae*, they
12 provide a blueprint for the evolution of enteropathogenic bacteria.

13

1 **Introduction**

2 The ever-expanding microbial genome sequence dataset, combined with an
3 increased understanding of the pathogenesis and ecology of bacterial pathogens,
4 is illuminating our view of the dynamics that have shaped the evolution of
5 bacterial pathogens. It is now evident that many bacterial pathogens infect
6 humans only incidentally and often produce virulence factors that are active
7 against non-mammalian adversaries as diverse as insects, protozoa, nematodes,
8 predatory bacteria and bacteriophage¹⁻³. These hosts provide a considerable
9 driving force in the evolution of bacterial pathogens enabling us to re-evaluate
10 human—pathogen interactions in the light of bacterial ecology and evolution
11 (the eco-evo perspective). A consequence of the eco-evo perspective is that it is
12 not surprising that genes encoding ‘virulence factors’ are found in pathogens and
13 non-pathogens colonizing diverse mammalian and non-mammalian hosts^{4,5}.
14 Conversely, as bacteria evolve to inhabit new hosts, or new niches within the
15 same hosts, using genomics we can see remnants of genes and gene clusters that
16 are thought to have provided an adaptive advantage in the past, but that are now
17 non-functional and represent evolutionary baggage.

18 Thus, many bacteria seem to be pre-armed with virulence determinants as they
19 traverse new environments and niches, and their evolution is not solely
20 dependent on mammalian/human–pathogen interactions^{6,7}. The yersiniae
21 represent a genus for which the eco-evo perspective is highly pertinent and for
22 which all 18 species in the genus have recently been fully sequenced,
23 metabolically phenotyped and compared⁸. The genus includes both non-
24 pathogenic and pathogenic species, and a diversity of pathogenesis phenotypes
25 so it represents a key model for understanding the forces at play in the evolution

1 of pathogenic bacteria. Our current understanding of the *Yersinia* encompasses a
2 long-term view of how broad host range pathogenic species such as *Yersinia*
3 *enterocolitica* evolve over millions of years from a non-pathogenic ancestor⁹. It
4 also encompasses short term evolution examining how recent evolutionary
5 bottlenecks have led to rapid emergence of high-pathogenic clones from their
6 ancestral lineage such as the emergence of *Yersinia pestis* from *Yersinia*
7 *pseudotuberculosis*¹⁰.

8 The evolution of pathogenic *Yersinia* spp. shows remarkable parallels with
9 members of the genus *Salmonella*, which encompasses the broad host range self-
10 limiting diarrheagenic *Salmonella enterica* subsp. *enterica* serovar Typhimurium
11 (*S. Typhimurium*) as well as those that are host restricted and acutely
12 pathogenic, such as *Salmonella enterica* subsp. *enterica* serovar Typhi (*S. Typhi*).
13 This has also made the pathogenic *Yersinia* species organisms of choice for much
14 of the pioneering work performed understanding the mechanisms of bacterial
15 pathogenesis through classical genetics. Studies using *Y. enterocolitica* have been
16 at the forefront of developing our understanding of how bacteria invade host
17 epithelial cells¹¹. Studies of the pathogenic *Yersinia* also forged our
18 understanding of the role of plasmids in bacterial virulence^{12,13}, the identification
19 and characterization of Type III secretion systems and their role in host
20 subversion¹⁴. This in turn led to advances in our understanding of how bacteria
21 evade host responses during infection¹⁵⁻¹⁷.

22 In this Review, we examine how the eco-evo perspective, informed by whole-
23 genome sequencing and by population genomic and ecological studies, has
24 challenged previous hypotheses on the evolution of mammalian pathogenesis in
25 *Yersinia* spp. Furthermore, we consider how these studies are revealing common

1 principles in the emergence of pathogenesis among other *Enterobacteriaceae*.
2 Finally, we show how considering all members of a genus in population genomic
3 studies, and not just pathogenic species, can provide greater insight and
4 resolution for understanding of the processes underpinning the evolution of
5 mammalian pathogenesis.

6

7 **The yersiniae**

8 The *Yersinia* is a genus of Gram-negative bacteria belonging to the
9 *Enterobacteriaceae*. The genus nomenclature is based on classical systematics
10 and biochemical speciation methods, resulting in a highly diverse group of
11 bacteria with 18 defined species¹⁸⁻²⁶. The best characterized of these are the
12 three species that cause disease in humans: *Y. enterocolitica* and *Y.*
13 *pseudotuberculosis* are zoonotic pathogens that cause self-limiting gastroenteritis
14 in humans, whereas *Y. pestis* is a rodent and flea pathogen that is occasionally
15 transmitted to humans and is the causative agent of plague²⁷. The remaining 15
16 known species of this genus are commonly found in soil and water environments
17 and are generally considered to be non-pathogenic to humans, though *Yersinia*
18 *ruckeri* is the causative agent of red mouth disease in salmonids²¹ and *Yersinia*
19 *entomophoga* has insecticidal activity²⁴. Recent phylogenomic analysis of the
20 entire *Yersinia* genus allowed for an accurate, whole genome SNP-based,
21 assessment of its population structure⁸, showing there were 14 distinct species
22 complexes within the genus. This robust sequenced based taxonomy disagreed
23 with the standing taxonomic description of the genus, based largely on
24 biochemical differences. It also showed that the mammalian pathogenic *Y.*
25 *enterocolitica* species and *Y. pseudotuberculosis* species complex (also containing

1 *Y. pestis* and *Y. similis*) fell at opposite ends of the *Yersinia* phylogenetic spectrum
2 (Box 1) and did not branch together as was previously thought.

3 Early microbial population genetics studies showed that *Y. pestis* is a recently
4 emerged clone of the enteropathogen *Y. pseudotuberculosis*, yet has a markedly
5 different lifestyle and causes a much more acute disease^{28,29}. The evolutionary
6 narrative tells how *Y. pestis* evolved from the gastrointestinal pathogen, in an
7 evolutionary blink of an eye (estimated to be between 2,000 and 10,000 years
8 ago^{28,30}) through the processes of gene gain, gene loss and genome
9 rearrangements. This finding was further explored with early microbial
10 comparative genomics studies. These revealed that *Y. pestis* had emerged from *Y.*
11 *pseudotuberculosis* through a mixture of both gene gain events, in the form of
12 acquisition of species-specific plasmids, bacteriophage, integrons and genomic
13 islands, as well as multiple complete or partial gene loss events leading to a
14 reduction in genomic flexibility and metabolic streamlining. All of this is
15 associated with its change in lifestyle and ultimate niche restriction^{10,31-33}. The
16 relationship of *Y. enterocolitica* to the *Y. pseudotuberculosis*—*Y. pestis* complex
17 was until recently poorly defined. In 2006, a comparison of the available
18 genomes from all three species suggested that they shared a distant but common
19 pathogenic ancestor which had acquired key pathogenicity determinants
20 including the essential virulence plasmid pYV, encoding Ysc the prototype type
21 III secretion system (T3SS), before splitting into distinct lineages^{9,27}. However,
22 the availability of just a single *Y. enterocolitica* genome sequence at this time
23 limited the interpretations that could be made. Further genomes were
24 subsequently published for *Y. enterocolitica* in 2011 and 2013³⁴⁻³⁷, as well as
25 singular genomes for type strains of environmental species in 2010³⁸ but our

1 understanding of the evolution of this whole genus and where *Y. enterocolitica*
2 fitted into the evolutionary scale of *Yersinia* species was still limited by the lack
3 of comprehensive whole genome data. Nonetheless, even with this limited
4 genomic data, a combination of gene gain and gene loss in the three human
5 pathogenic *Yersinia* species was evident and highlighted the complexity of
6 evolution of mammalian pathogenicity in these bacteria, showing that pathogens
7 can evolve by a process of “add DNA, stir, and reduce”²⁷.

8 The study of the evolution of pathogenesis in the genus also suggested an
9 inexorably distinct relationship between loss of metabolic function by genome
10 decay, reduction in ecological flexibility, and the evolution towards niche-
11 restricted highly pathogenic lineages²⁷. As population genomic studies have
12 become an established tool in microbiology, the level of resolution at which we
13 can study these evolutionary events has improved considerably. By genome
14 sequencing across entire bacterial genera it is possible to determine the key
15 acquisition and loss events in the evolution of bacterial pathogenesis. It is also
16 possible to shed new light on the evolutionary forces and patterns underpinning
17 these processes, and identify their common features across bacterial genera.

18

19 **The role of gene gain**

20 The evolutionary path from environmental or commensal bacterium to human
21 pathogen has long been associated with the gain of mobile genetic elements via
22 horizontal gene transfer, through phage transduction, plasmid uptake, natural
23 transformation and DNA conjugation. In the *Yersinia* genus there seems to have
24 been ‘foothold moments’ characterized by acquisitions that, looking back across
25 the genus phylogeny appear as pivotal steps in the emergence of mammalian

1 pathogenic lineages. The most important of these seems to have been the
2 acquisition of the family of pYV virulence plasmids encoding the Ysc prototype
3 T3SS^{8,12,13,39}. The Ysc T3SS delivers Yop effector proteins directly into host cells
4 upon contact, resulting in the cessation of actin polymerization for endocytosis
5 (which promotes bacterial uptake) and in the suppression of host-cell
6 transcription⁴⁰. Following contact with host macrophages, the activities of Yop
7 effectors silence the innate immune response, resulting in the down-regulation
8 of pro-inflammatory cytokine production and allowing proliferation of the
9 infection⁴¹⁻⁴³, a process termed the “*Yersinia* deadly kiss”.

10 The advantage of sequencing broadly across a genus and deep within each
11 species is that the origins and distribution of all genes described as *Yersinia* spp.
12 virulence factors, or otherwise, in the published literature can be ascertained⁸.
13 Such an analysis confirmed that pYV was one of only two virulence-associated
14 genetic loci that were uniquely present in all three *Yersinia* lineages commonly
15 associated with mammalian disease (pYV is absent from *Yersinia* species non-
16 pathogenic to mammals). The other was the chromosomally encoded attachment
17 and invasion locus *ail* which has a role in the attachment to, and invasion of,
18 mammalian epithelial cells as well as in resistance to killing by the host
19 complement system⁴⁴⁻⁴⁶. The pivotal role of pYV in the emergence of mammalian
20 pathogenesis is highlighted by the revelation that, contrary to accepted
21 wisdom^{9,27,47}, closely related versions of this plasmid were acquired
22 independently on at least three occasions in the *Yersinia* genus; twice by the *Y.*
23 *enterocolitica* and once by the *Y. pseudotuberculosis*—*Y. pestis* lineage⁸, and that
24 *Y. enterocolitica* did not share a common pathogenic ancestor with *Y.*
25 *pseudotuberculosis* as previously thought⁸. As such, multiple independent *ail* and

1 pYV acquisitions were key gene gain events that marked the parallel
2 independent evolution of mammalian pathogenesis in the *Yersinia* genus. This is
3 surprising if one considers the large number of virulence-associated genes that
4 have been reported in *Yersinia* based on published classic bacterial genetics
5 experiments. This includes key virulence factors such as Invasin⁴⁸, the Fes/Fep
6 iron acquisition system⁴⁹, the Tad type IVb pilus⁵⁰ and Myf fimbriae⁵¹, and the
7 Yst heat stable toxin⁵², all of which are widely distributed across pathogenic and
8 non-pathogenic members of the genus. This could be explained through the eco-
9 evo perspective whereby the environmental yersiniae may have acquired and
10 maintained some virulence factors to combat predatory protozoa, nematodes,
11 predatory bacteria, and bacteriophage or even to colonise plants. This highlights
12 the benefits of studying the accessory gene pool of pathogenic bacteria in the
13 wider context of the genera they belong to, a strategy that has yet to be widely
14 applied to other enteric pathogens. Indeed in order to confirm the eco-evo
15 hypothesis there is also a benefit and need to study the role of virulence genes in
16 these non-pathogenic isolates to allow us to fully determine the true ecology of
17 bacterial species.

18 In addition to the crucial gains of the pYV plasmid and *ail* locus in all three
19 human pathogenic *Yersinia* spp., a number of other acquired loci have been
20 maintained by clonal expansion in individual *Yersinia* species complexes and
21 phylogroups (PG). Perhaps the most striking and formative of these is the
22 acquisition of the pFra (pMT1) and pPla (pPCP1) plasmids of *Y. pestis*, which
23 encode the murine toxin Ymt and the F1 capsular protein and the plasminogen
24 activator Pla. The acquisition of these two plasmids is a significant event in the
25 move towards efficient insect vector-mammalian transmission of *Y. pestis*,⁵³ and

1 Pla has recently been shown to be pivotal for causing the fulminant pneumonic
2 infection unique to *Y. pestis* within the *Yersinia* genus⁵⁴. Also in this group of
3 acquired elements is the High Pathogenicity Island (HPI). HPI is a large
4 integrative-conjugative element capable of excision and transfer via the use of
5 self encoded integrase and excision factors⁵⁵. As such the HPI is often found in a
6 range of bacterial genera including extra-intestinal pathogenic *E. coli*⁵⁶. The HPI
7 encodes the siderophore yersiniabactin that sequesters iron in the host^{56,57},
8 imports zinc into bacterial cells⁵⁸, and assists in reducing the respiratory
9 oxidative burst response of immune cells⁵⁹. Within *Yersinia* spp., HPI containing
10 a fully functioning yersiniabactin is found only in strains belonging to the *Y.*
11 *pseudotuberculosis*—*Y. pestis* species complex, where its presence is isolate
12 dependant, and in the highly virulent phylogroup 2 lineage of the *Y.*
13 *enterocolitica* species complex, where it is uniformly present. The key remaining
14 gene acquisition events that have occurred in the pathogenic members of the
15 genus involve the Ysa T3SS by phylogroup 2 *Y. enterocolitica*, and Hms by the *Y.*
16 *pseudotuberculosis*—*Y. pestis* species complex. The Ysa T3SS is located in a
17 region of the genome known as the plasticity zone⁶⁰, first identified in the
18 archetypal phylogroup 2 strain 8081⁹. The plasticity zone is a region of the *Y.*
19 *enterocolitica* genome that shows signs of multiple independent, and phylogroup
20 specific acquisitions and deletions⁹. Ysa T3SS has been shown to secrete a large
21 number of chromosomally encoded T3SS effector proteins⁶¹ as well as the Yop
22 T3SS effector proteins of the Ysc system carried on pYV⁶². Although mutagenesis
23 studies suggest *ysa* plays only a minor role in mammalian pathogenesis⁶¹,
24 consistent with the eco-evo perspective, *ysa* mutants were shown to be
25 attenuated in their ability to survive inside cultured *Drosophila melanogaster*

1 cells suggesting *ysa* is more important outside of the mammalian host for
2 phylogroup 2 *Y. enterocolitica*⁶³. The Hms locus is located on the chromosome
3 and encodes a hemin storage locus that is unique to the *Y. pseudotuberculosis*-*Y.*
4 *pestis* lineage and is responsible for its ability to form pigmented colonies on
5 congo-red agar⁶⁴. Hms is important for *Y. pestis* pathogenesis as it is responsible
6 for formation of biofilms, required for the blockage of the flea foregut, which
7 facilitates its subsequent transmission to mammals through reflux of infected
8 blood from the infected feeding flea back onto the bite site⁶⁵.

9

10 **Regulatory control**

11 Although the acquisition of mobile genetic elements has been a key process in
12 the emergence of *Yersinia* spp. pathogenesis, loci are more likely to be stably
13 maintained in a population if there is a minimal fitness cost to their integration
14 into the genome of the host bacterium, or a high selective pressure for them to
15 be maintained. To achieve this, many acquired elements fall under strict
16 transcriptional and/or translational control to minimise the fitness costs
17 associated with their acquisition. Across the *Enterobacteriaceae* many acquired
18 genetic elements are regulated by H-NS, a nucleoid associated, histone-like
19 protein that binds to A-T rich regions of DNA silencing transcriptional activity⁶⁶.
20 Once under the strict negative control of H-NS, the acquired elements are then
21 integrated into existing transcriptional activating regulons within the bacterial
22 cell. These function in an antagonistic fashion with H-NS to tightly control the
23 transcriptional expression of the new genes. In the case of *Yersinia* spp., the
24 chromosomal *rovA/ymoA* regulon has been suggested to fulfil this antagonistic
25 role to H-NS^{8,67}. The *rovA* gene is universally present across the genus⁸ and

1 encodes a member of the MarR/SlyA family of global transcriptional regulators⁶⁷.
2 It is known to antagonise H-NS silencing of the invasin protein gene, *invA*, the
3 product of which is involved in the early stages of *Yersinia* spp. attachment to
4 cells and internalization during infection. The binding of RovA to the promoter
5 region masks the H-NS binding site, triggering transcription⁶⁸. Genes *rovA/ymoA*
6 are thought to encode a promiscuous ancestral regulon into which the
7 expression of acquired genes were incorporated multiple times in the different
8 lineages (FIG 1). This is thought to explain why RovA regulates different gene
9 sets in *Y. enterocolitica* and *Y. pseudotuberculosis/pestis*^{67,69}: RovA was shown to
10 transcriptionally activate 73 genes in *Y. pestis* and 63 genes in *Y. enterocolitica*,
11 but only three genes were common to both species⁶⁷. The majority of RovA
12 controlled genes are lineage-specific virulence factors such as Invasin or Myf
13 fimbriae, or lineage specific prophage or metabolic genes⁶⁷. The *rovA/ymoA*
14 regulon also controls the expression of genes encoded on the pYV plasmid,
15 highlighting its importance as a global regulator of virulence.

16 The expression of the Yop proteins encoded on pYV is down-regulated at a
17 transcriptional level by LcrF (an ortholog of HN-S, encoded on the pYV plasmid),
18 in response to the mammalian environment. Recent work has identified that this
19 regulation is controlled at a hierarchical level by YmoA, which functions as a
20 thermo-sensing switch to turn off *lcrF* transcription, and therefore Yop
21 translation and secretion at low temperatures⁷⁰, and by Hfq and other small
22 regulatory RNAs conserved in the *Yersinia* genus⁷¹. This thermo-sensing switch
23 is a tandem process which relies on weak binding of YmoA directly to the
24 promoter region of *lcrF* at low temperatures and the formation of a double stem
25 loop in the intergenic region resulting in steric hindrance of RNA polymerase

1 binding and formation of transcripts^{70,72}. Overall the importance of the
2 RovA/YmoA system for the evolution of pathogenesis in *Yersinia* is evident from
3 these data: At temperatures below 37°C virulence factors, essential for
4 mammalian pathogenesis, are not expressed because RovA/YmoA is inactive and
5 they are repressed by H-NS. However flagella biosynthesis genes (in the case of
6 the enteropathogenic species) or insect transmission loci such as *hms* (in the
7 case of *Y. pestis*) are expressed at high levels^{73,74}. Upon a shift to mammalian
8 body temperatures essential acquired virulence factors such as those on pYV, *ail*
9 and *invA* are activated by RovA/YmoA. Other environmental triggers are also
10 known to influence this regulatory cascade such as levels of exogenous calcium
11 ions⁷⁵ or levels of exogenous magnesium ions via the PhoPQ two component
12 regulatory system⁷⁶. Indeed PhoPQ has been shown to have an overarching
13 control on levels of RovA and as such a key role in virulence phenotypes in all 3
14 human pathogenic species⁷⁷⁻⁷⁹.

15 The presence of an ancestral promiscuous regulatory system controlling
16 expression of acquired genes involved in virulence draws striking parallels with
17 *Salmonella* spp. The key acquisition event in the evolution of pathogenesis in the
18 *Salmonella* genus was *Salmonella* pathogenicity island-1 (SPI-1). This
19 pathogenicity island encodes a T3SS and its cognate secreted effector proteins
20 that remodel the actin cytoskeleton of infected cells and force uptake of attached
21 bacteria into endocytic vacuoles. As with pYV, the acquisition of SPI-1 was a
22 landmark event in the evolution of the *Salmonella* genus, being present in both
23 species: *S. enterica* and *Salmonella bongori*⁸⁰. SPI-1 is under negative
24 transcriptional control by H-NS. Similar to *Yersinia*, H-NS-mediated repression of
25 SPI-1 functions is alleviated by the concerted action of a number of

1 transcriptional regulators including Hild/HilA, SlyA which is an ortholog of
2 RovA, and the two-component regulatory system PhoPQ⁸¹. Recent work showed
3 that SPI-1 was readily lost in *Salmonella* H-NS null mutants, Showing that H-NS
4 silencing of SPI-1 was likely to have been an essential factor in SPI-1 becoming
5 fixed in all *Salmonella* spp. and so the evolution of pathogenesis in members of
6 this genus too⁸².

7 Therefore it seems that an emerging theme in the evolution of pathogenesis in
8 enteric bacteria is not only the pivotal acquisition of one or two key genetic loci,
9 but also the co-ordinated transcriptional control of those acquired loci by H-NS
10 and other species-specific promiscuous global regulators.

11

12 **Role of gene loss**

13 Although the importance of gene gain in the evolution of bacterial pathogenicity
14 is intuitive and well documented, the role of gene loss and host adaptation has
15 been shown to be central in the emergence of bacterial pathogens⁸³. The
16 majority of work looking at the evolution of pathogenesis in *Yersinia* spp. has
17 focused on the evolutionary events that led to the emergence of *Y. pestis* from *Y.*
18 *pseudotuberculosis*²⁸. Although there were important acquisition events, the
19 most striking observation of early comparative genomic analyses was the
20 functional loss of a large number of genes associated with virulence and
21 metabolism in *Y. pestis*¹⁰. This functional loss applied to about 10% of the genes
22 in the genome, a considerably larger proportion of the genome compared to gene
23 gain events. The gene attrition was largely attributable to the expansion of four
24 major classes of IS elements (*IS1541*, *IS100*, *IS285*, *IS1661*) leading to insertional
25 gene inactivation. Other processes of gene loss included deletion of genes and

1 loci and SNP-based pseudogene formation^{32,33,84,85}. Many of the genes that were
2 functionally lost encode virulence factors such as invasin and YadA, and
3 metabolic loci involved in motility, dicarboxylic amino-acid metabolism, and
4 uracil transport^{10,27}. These are required for successful colonization of the
5 mammalian gastrointestinal tract, a niche no longer frequently utilised by *Y.*
6 *pestis* as it emerged from the intestinal pathogenic *Y. pseudotuberculosis* to
7 colonize the mid-gut of fleas and, in mammals, to become a more systemic
8 pathogen invading the lymphatic system²⁷.

9 The importance of gene loss in the evolution of *Y. pestis* has recently been
10 exemplified by studies investigating key mutations in *Y. pestis* that mark its
11 evolutionary adaptation to flea-borne transmission in mammals. Using a
12 combination of comparative genomics and classical bacterial genetics, three key
13 loss-of-function mutations were found, all of which increased biofilm formation
14 in the flea foregut in a cyclic-di-GMP dependant manner⁶. The key mutations
15 were in two genes encoding EAL-domain phosphodiesterases (PDE), and the
16 *rcsA* gene that is a component of the *rcs* regulatory network. The *rcs* system
17 transcriptionally represses the *hms* locus and biofilm formation. PDE are
18 enzymes that degrade cyclic-di-GMP, a bacterial signalling molecule that
19 enhances transcription of *hms* and biofilm formation⁶. The functional loss of the
20 PDE enzymes and the regulator repressing transcription in a *Y.*
21 *pseudotuberculosis*-ancestor led to the emergence of a lineage with enhanced
22 ability to form biofilms in the foregut of fleas and a resulting shift in pathogenic
23 lifestyle of *Y. pestis*⁶. The relatively recent emergence of *Y. pestis*²⁸ is consistent
24 with the SNP-based pseudogenes, which have swept to fixation through the *Y.*
25 *pestis* population but have not yet been fully deleted^{30,86}. Also in *Y. pestis* is a

1 mutation in the *ureD* gene that leads to functional loss of urease. In *Y.*
2 *pseudotuberculosis* it has been shown that the functional *ureD* encoding urease is
3 associated with significant oral toxicity in the flea⁸⁷, hence the *ureD* mutation
4 may play a contributory role in the emergence of *Y. pestis*^{6,87}(Fig 2). However,
5 perhaps significantly the *ureD* mutation appears to be a reversible phase-
6 variable mutation and maybe a contingency gene required in the mammalian
7 host or soil dwelling stage, rather than the insect vector, part of its life cycle⁸⁸.
8 This provides *Y. pestis* the flexibility of retaining urease expression, as and when
9 required.

10 Genomic analysis of *Y. enterocolitica* showed that similar gene loss and metabolic
11 restriction also mark the evolution of certain subtypes of this species⁸. As
12 discussed above and shown in Figure 3 *Y. enterocolitica* can be subdivided into
13 the non-pathogenic phylogroup PG 1, which is an ancestral lineage, the highly
14 pathogenic PG 2, and PGs 3-5 which show limited pathogenicity in mouse models
15 but are the most successful lineages in terms of disease causation and are most
16 commonly isolated from livestock and human clinical cases. PG 6 is a rare lineage
17 which has only ever been isolated from wild hares^{89,90}. The evolutionary descent
18 from host-generalists (PGs 3-5) to the host-restricted PG 6, similar to the
19 emergence of *Y. pestis*, was accompanied in *Y. enterocolitica* by the expansion of a
20 single IS element, IS1667, as well as functional loss of a large number of
21 metabolic genes. The impact of the functional loss of metabolic loci is
22 particularly pronounced in PG 6 and is thought to reflect its extreme niche
23 restriction to hares. Interestingly the list of metabolic pathways inactivated in PG
24 6 is similar in part to those lost in *Y. pestis* (cobalamin biosynthesis, tetrathionate
25 respiration and the hydrogenase-4 operon responsible for the use of molecular

1 hydrogen in anaerobic respiration). These mutational observations are directly
2 supported by high throughput metabolic phenotyping of PG 6⁸.

3 Looking at the functions that were lost as different lineages of *Y. enterocolitica*
4 diverged and adapted to new niches, one of the earliest events is likely to have
5 been the decay of the *Yersinia* genus T3SS, YGT. YGT is an ancestral T3SS present
6 across the whole genus and is distinct from the Ysc and Ysa T3SSs acquired in
7 pathogenic lineages. YGT has yet to be fully characterized but it shares a high
8 level of sequence identity with the T3SS encoded on the SPI-2 genomic island in
9 *Salmonella enterica*, which was a key acquisition required for intracellular
10 survival⁹¹. The YGT present in non-pathogenic PG 1 *Y. enterocolitica* seems to
11 contain a full complement of effector protein encoding genes compared with the
12 SPI-2 T3SS⁸. In the low-pathogenic PGs 3-6 the YGT seems to have been
13 inactivated either by single base pair mutations or deletion of the secretion
14 apparatus genes, whereas almost all traces of the YGT have been deleted in the
15 highly pathogenic PG 2 isolates⁸. As an example of how dynamic the evolution of
16 pathogenesis is through gene gain and loss, the loss of YGT in the high-
17 pathogenic PG 2 *Y. enterocolitica* is, conceptually at least, offset by the gain of the
18 Ysa T3SS⁶⁰. A full functional characterisation of YGT is now required to
19 determine if the selective maintenance of YGT in PG 1 *Y. enterocolitica* is a
20 classical example of eco-evo principal, where the T3SS could be conferring a
21 protective benefit in a non-mammalian host environment.

22 IS mediated gene inactivation is also evident in the low-pathogenic *Y.*
23 *enterocolitica* PG 3 lineage (comprising the classically entitled serotype O:3
24 strains)^{92,93}. Molecular analysis revealed a *IS1667* had inserted into the *invA*
25 invasin promoter region, resulting in the formation of an additional RovA

1 binding site and a concomitant increase in expression of *invA*⁹⁴. Comparative
2 analysis with the other low-pathogenic phylogroups of *Y. enterocolitica* suggest
3 this is the primary event leading to the enhanced virulence properties of PG 3
4 isolates compared to PGs 2, 4, and 5⁹³. As a result of the increased expression of
5 invasion PG 3 isolates can colonise porcine tissues more effectively than other
6 phylogroups, and as result PG 3 have become the predominant isolates taken
7 from human clinical cases in Europe over the past 10 years and also dominate
8 veterinary epidemiological surveys⁹⁵.

9

10 **Gene loss in other *Enterobacteriaceae***

11 Functional gene loss in metabolic loci has been a key event in the *Yersinia* genus
12 that mediated the transition from ubiquitous host-generalist lineage to niche
13 restricted and host-adapted pathogen on at least two occasions. Once again this
14 observation has also been made in *Salmonella* species. *S. Typhi*⁹⁶ shows many
15 similarities with *Y. pestis*, it is niche restricted (in this case to the human host)
16 and causes an acute systemic disease known as typhoid. Similarly *S. Gallinarum*
17 and *S. Pullorum* also cause a typhoid-like disease but are restricted to galliforme
18 birds⁹⁷, whereas *S. Typhimurium* is a host-generalist causing an inflammatory
19 intestinal infection of mammals⁹⁸. Comparative genomic studies of these
20 salmonellae have shown that functional gene loss in loci involved in anaerobic
21 metabolism (more specifically functional loss in the *cob/pdu* and *ttr* operons)
22 was a marker of the transition from intestinal pathogen, like *S. Typhimurium*, to
23 invasive pathogens such as *S. Typhi*, *S. Gallinarum* and *S. Pullorum*⁹⁹.

24 In *S. Typhimurium* under anaerobic conditions, the *cob* genes encode the ability
25 to synthesis cobalamin (or vitamin B12). Cobalamin is required as a cofactor for

1 several enzymes including the first enzyme in the 1, 2-propanediol utilisation
2 pathway (encoded by *pdu* operon). 1,2-Propanediol is a by-product of
3 fermentative growth on rhamnose and fucose¹⁰⁰, common constituents of plant
4 cell walls and intestinal epithelial cells (and therefore the gut)¹⁰¹. The anaerobic
5 degradation of 1,2-propanediol requires tetrathionate for use as a terminal
6 electron acceptor, facilitated by the products of the *ttr* genes¹⁰². Unlike 1,2-
7 propanediol, tetrathionate was only previously known to be present in soils not
8 in the mammalian gut. However, it has been recently shown that tetrathionate is
9 produced naturally during an inflammatory response to a *Salmonella* infection. It
10 is under such conditions that the *cob*, *pdu* and *ttr* functions combine to provide a
11 competitive growth advantage for *S. Typhimurium*, by allowing it to outgrow the
12 natural and largely fermentative gut microbiota by using naturally occurring
13 carbon sources such as 1,2-propanediol that are not readily fermented^{103,104}
14 (FIG. 4). As the invasive typhoidal *Salmonella* have evolved away from an
15 intestinal lifestyle, functions in *cob/pdu/ttr* operons have been sequentially lost
16 and so represent markers of their change in niche and a movement away from
17 causing intestinal disease to a more systemic infection cycle¹⁰⁵.

18 Although the *cob/pdu* and *ttr* loci are present and intact in *Y. enterocolitica* and
19 in the majority of environmental *Yersinia* species¹⁰⁶, they are absent from the *Y.*
20 *pseudotuberculosis*—*Y. pestis* complex. As mentioned above, tetrathionate is
21 produced in the vertebrate inflamed gut following the host's response to *S.*
22 *Typhimurium* infection, whereby the SPI-1 and -2 encoded T3SS are essential for
23 stimulating this inflammatory response. Studies of *Y. enterocolitica* pathogenesis
24 have shown that infections are characterized by inflammation of the gut¹⁰⁷.
25 Moreover, inflammation appears to require pYV¹⁰⁸. It is possible that by inducing

1 a pYV-associated inflammatory response during intestinal infection¹⁰⁹, and by
2 expression of the *cob/pdu* and *ttr* loci, *Y. enterocolitica* is able, like *S.*
3 *Typhimurium*, to gain a metabolic advantage over the resident gut microbiota by
4 using tetrathionate and naturally occurring carbon sources to respire under
5 these conditions. The fact that the *cob/pdu* and *ttr* loci are lost completely in *Y.*
6 *pseudotuberculosis* – *Y. pestis* may explain why these pathogens are incapable of
7 causing the inflammatory intestinal infection observed in *Y. enterocolitica*,
8 despite containing pYV.

9

10 **Summary and concluding remarks**

11 For decades the pathogenic yersiniae have been at the vanguard of
12 understanding bacterial pathogenesis. Examples include the role of T3SSs^{42,110}
13 and understanding the molecular mechanisms by which intestinal pathogens
14 invade the intestinal epithelium⁴⁴. Work on the pathogenic *Yersinia* species also
15 drove the understanding of the importance of plasmids in conferring virulence to
16 pathogenic bacteria^{12,39}. The yersiniae are also at the forefront in developing our
17 understanding of the evolution of mammalian pathogenesis. Furthermore,
18 *Yersinia* spp. genomics has also been at the frontline of palaeomicrobiology, with
19 the near complete reconstruction and interpretation of the Black Death and
20 Justinian plague *Y. pestis* genomes from ancient DNA studies on historical disease
21 episodes^{86,111}. Additionally the microevolution of *Y. pestis* is beginning to be
22 unravelled and the distinction between ancestral and modern *Y. pestis* has been
23 made ¹¹². And very recent advances have identified that *Y. pestis* was a human
24 pathogen of bronze-age humans, but had not yet acquired Ymt meaning it
25 evolved the ability to infect fleas after becoming a human pathogen ¹¹³.

1 The position of *Yersinia* as a model genus for pathogenesis has been further
2 cemented as it is the first multispecies bacterial genus for which all species have
3 been sequenced and placed into phylogenomic context⁸. By taking this approach
4 coupled with high throughput metabolic analyses using phenotype microarrays
5 it was possible to trace the evolutionary origins of the species, contextualise
6 them and reveal the parallel evolutionary paths of virulence in humans⁸ (FIG. 5).
7 The genus *Yersinia* evolved from environmental ancestors. The most outlying
8 species are the fish pathogenic *Y. ruckeri* (SC 2), and newly classified lineages *Y.*
9 *nurmii* and *Y. entomophaga* (SC 3). Previously, it had been debated whether *Y.*
10 *ruckeri* was truly part of the genus *Yersinia*²¹, however this has now been
11 strengthened by the recognition of this novel lineage. The human pathogenic
12 species clusters emerged independently following the common themes of 'add,
13 stir, and reduce'. *Y. pestis* acquired its unique virulence plasmids, an expansion of
14 several IS elements led to widespread genome rearrangements, and considerable
15 gene loss can be observed. However, in the *Y. enterocolitica* lineages, these
16 themes are weighted differently in the various phylogroups. PG 2 added the
17 high-pathogenicity island, and type 2 and 3 secretions systems, and on the other
18 hand lost the second flagella cluster and the genus T3SS YGT. PGs 3-6 show
19 genome rearrangements following the expansion of IS1667. These phylogroups
20 also show general loss of metabolic flexibility, and in the case of PG 6 an extreme
21 loss of metabolic capability.

22 Comparative analysis of complete genome sequence data sets of all
23 representative yersiniae species coupled with the eco-evo perspective of
24 bacteria has provided a greater understanding of the key evolutionary steps that
25 led to the emergence of successful mammalian pathogens. We need to move

1 away from our biased human-centric view of bacteria and consider more the
2 environmental context, habitat and niches of these organisms.

3 Finally, the *Yersinia* genus data shows that a genus-wide sequencing approach
4 leads to new information on the distribution of virulence-associated genes and
5 on which of these genes are genuine virulence factors that are appropriate for
6 novel diagnostic and pathogen-targeting research. However, even with the large
7 number of *Yersinia* spp. isolates sequenced to date, the origin and reservoir of
8 mobile genetic elements, such as pYV, remain obscure. It may be that we have
9 only sequenced the tip of the bacterial iceberg and that future large genome and
10 microbiome studies coupled with high throughput phenotypic methods will be
11 even more revealing. This will provide a web of bacterial life, the inter-
12 relatedness of bacteria, no doubt blurring species boundaries. The genome data
13 together with the eco-evo perspective will form the basis of how and why
14 bacteria evolve to be more pathogenic with the potential to avert future threats
15 posed by old adversaries such as the emergence of antimicrobial resistance.

16

17 **References**

18

- 19 1. Fuchs, T., Bresolin, G., Marcinowski, L., Schachtner, J. & Scherer, S.
20 Insecticidal genes of *Yersinia* spp.: taxonomical distribution, contribution
21 to toxicity towards *Manduca sexta* and *Galleria mellonella*, and evolution.
22 *BMC Microbiol.* **8**, 214 (2008).
23
- 24 2. Lithgow, K. V *et al.* A general protein O-glycosylation system within the
25 *Burkholderia cepacia* complex is involved in motility and virulence. *Mol.*
26 *Microbiol.* **92**, 116–137 (2014).
27
- 28 3. Pukklay, P. *et al.* Involvement of EnvZ-OmpR two-component system in
29 virulence control of *Escherichia coli* in *Drosophila melanogaster*. *Biochem.*
30 *Biophys. Res. Commun.* **438**, 306–311 (2013).

- 1
- 2 4. Kumar, P. & Viridi, J. S. Identification and distribution of putative virulence
3 genes in clinical strains of *Yersinia enterocolitica* biovar 1A by suppression
4 subtractive hybridization. *J. Appl. Microbiol.* **113**, 1263–1272 (2012).
5
- 6 5. Korea, C.-G., Badouraly, R., Prevost, M.-C., Ghigo, J.-M. & Beloin, C.
7 *Escherichia coli* K-12 possesses multiple cryptic but functional chaperone-
8 usher fimbriae with distinct surface specificities. *Environ. Microbiol.* **12**,
9 1957–1977 (2010).
10
- 11 6. Sun, Y.-C., Jarrett, C. O., Bosio, C. F. & Hinnebusch, B. J. Retracing the
12 evolutionary path that led to flea-borne transmission of *Yersinia pestis*. *Cell*
13 *Host Microbe* **15**, 578–586 (2014).
- 14 Key recent paper on the importance of *ure* mutations in the emergence of *Y.*
15 *pestis*
- 16 7. Viana, D. *et al.* A single natural nucleotide mutation alters bacterial
17 pathogen host tropism. *Nat. Genet.* **47**, 361–366 (2015).
18
- 19 8. Reuter, S. *et al.* Parallel independent evolution of pathogenicity within the
20 genus *Yersinia*. *Proc Natl Acad Sci U S A* **111**, 6768–6773 (2014).
- 21 First ever population genomic analysis of an entire bacterial genus, and key
22 *Yersinia* evolution paper
- 23 9. Thomson, N. R. *et al.* The Complete Genome Sequence and Comparative
24 Genome Analysis of the High Pathogenicity *Yersinia enterocolitica* Strain
25 8081. *PLoS Genet.* **2**, e206 (2006).
26
- 27 10. Chain, P. S. G. *et al.* Insights into the evolution of *Yersinia pestis* through
28 whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc Natl*
29 *Acad Sci U S A* **101**, 13826–13831 (2004).
30
- 31 11. Isberg, R. R., Voorhis, D. L. & Falkow, S. Identification of invasins: A protein
32 that allows enteric bacteria to penetrate cultured mammalian cells. *Cell*
33 **50**, 769–778 (1987).
34
- 35 12. Portnoy, D. A. & Falkow, S. Virulence-associated plasmids from *Yersinia*
36 *enterocolitica* and *Yersinia pestis*. *J. Bacteriol.* **148**, 877–883 (1981).
37
- 38 13. Portnoy, D. A., Moseley, S. L. & Falkow, S. Characterization of plasmids and
39 plasmid-associated determinants of *Yersinia enterocolitica* pathogenesis.
40 *Infect. Immun.* **31**, 775–782 (1981).

- 1
- 2 14. Rosqvist, R., Hakansson, S., Forsberg, A. & Wolf-Watz, H. Functional
3 conservation of the secretion and translocation machinery for virulence
4 proteins of yersiniae, salmonellae and shigellae. *EMBO J.* **14**, 4187–4195
5 (1995).
6
- 7 15. Yao, T., Mecsas, J., Healy, J. I., Falkow, S. & Chien, Y. Suppression of T and B
8 lymphocyte activation by a *Yersinia pseudotuberculosis* virulence factor,
9 yopH. *J. Exp. Med.* **190**, 1343–1350 (1999).
10
- 11 16. Monack, D. M., Mecsas, J., Bouley, D. & Falkow, S. *Yersinia*-induced
12 apoptosis in vivo aids in the establishment of a systemic infection of mice.
13 *J. Exp. Med.* **188**, 2127–2137 (1998).
14
- 15 17. Boland, A. & Cornelis, G. R. Role of YopP in suppression of tumor necrosis
16 factor alpha release by macrophages during *Yersinia* infection. *Infect*
17 *Immun* **66**, 1878–1884 (1998).
18
- 19 18. Carniel, E. Evolution of pathogenic *Yersinia*, some lights in the dark. *Adv.*
20 *Exp. Med. Biol.* **529**, 3–12 (2003).
21
- 22 19. Sprague, L. D. & Neubauer, H. *Yersinia aleksiciae* sp. nov. *Int. J. Syst. Evol.*
23 *Microbiol.* **55**, 831–835 (2005).
24
- 25 20. Sprague, L. D., Scholz, H. C., Amann, S., Busse, H.-J. & Neubauer, H. *Yersinia*
26 *similis* sp. nov. *Int. J. Syst. Evol. Microbiol.* **58**, 952–958 (2008).
27
- 28 21. Sulakvelidze, A. *Yersiniae* other than *Y. enterocolitica*, *Y.*
29 *pseudotuberculosis*, and *Y. pestis*: the ignored species. *Microbes Infect.* **2**,
30 497–513 (2000).
31
- 32 22. Murros-Kontiainen, A. *et al.* *Yersinia nurmii* sp. nov. *Int J Syst Evol Micro* **61**,
33 2368–2372 (2011).
34
- 35 23. Murros-Kontiainen, A. *et al.* *Yersinia pekkanenii* sp. nov. *Int J Syst Evol*
36 *Microbiol* **61**, 2363–2367 (2011).
37
- 38 24. Hurst, M. R. H., Becher, S. A., Young, S. D., Nelson, T. L. & Glare, T. R. *Yersinia*
39 *entomophaga* sp. nov., isolated from the New Zealand grass grub *Costelytra*
40 *zealandica*. *Int. J. Syst. Evol. Microbiol.* **61**, 844–849 (2011).
41

- 1 25. Merhej, V., Adekambi, T., Pagnier, I., Raoult, D. & Drancourt, M. *Yersinia*
2 *massiliensis* sp. nov., isolated from fresh water. *Int. J. Syst. Evol. Microbiol.*
3 **58**, 779–784 (2008).
4
- 5 26. Savin, C. *et al.* The *Yersinia pseudotuberculosis* complex: characterization
6 and delineation of a new species, *Yersinia wautersii*. *Int. J. Med. Microbiol.*
7 **304**, 452–463 (2014).
8
- 9 27. Wren, B. W. The Yersiniae - a model genus to study the rapid evolution of
10 bacterial pathogens. *Nat Rev Micro* **1**, 55–64 (2003).
11
- 12 28. Achtman, M. *et al.* *Yersinia pestis*, the cause of plague, is a recently emerged
13 clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A* **96**, 14043–
14 14048 (1999).
- 15 The initial paper that showed *Y. pestis* was a recently emerged clone of *Y.*
16 *pseudotuberculosis*
- 17 29. Achtman, M. *et al.* Microevolution and history of the plague bacillus,
18 *Yersinia pestis*. *Proc Natl Acad Sci U S A* **101**, 17837–17842 (2004).
19
- 20 30. Bos, K. I. *et al.* *Yersinia pestis*: new evidence for an old infection. *PLoS One*
21 **7**, e49803 (2012).
22
- 23 31. Chain, P. S. *et al.* Complete genome sequence of *Yersinia pestis* strains
24 Antiqua and Nepal516: evidence of gene reduction in an emerging
25 pathogen. *J Bacteriol* **188**, 4453–4463 (2006).
26
- 27 32. Parkhill, J. *et al.* Genome sequence of *Yersinia pestis*, the causative agent of
28 plague. *Nature* **413**, 523–527 (2001).
29
- 30 33. Deng, W. *et al.* Genome Sequence of *Yersinia pestis* KIM. *J. Bacteriol.* **184**,
31 4601–4611 (2002).
32
- 33 34. Savin, C. *et al.* Draft Genome Sequence of a Clinical Strain of *Yersinia*
34 *enterocolitica* (IP10393) of Bioserotype 4/O:3 from France. *Genome*
35 *Announc.* **1**, e00150-12 (2013).
36
- 37 35. Batzilla, J., Hoper, D., Antonenka, U., Heesemann, J. & Rakin, A. Complete
38 genome sequence of *Yersinia enterocolitica* subsp. *palaearctica* serogroup
39 O:3. *J. Bacteriol.* **193**, 2067 (2011).
40

- 1 36. Fuchs, T. M., Brandt, K., Starke, M. & Rattei, T. Shotgun sequencing of
2 *Yersinia enterocolitica* strain W22703 (biotype 2, serotype O:9): genomic
3 evidence for oscillation between invertebrates and mammals. *BMC*
4 *Genomics* **12**, 168 (2011).
5
- 6 37. Wang, X. *et al.* Complete genome sequence of a *Yersinia enterocolitica* 'Old
7 World' (3/O:9) strain and comparison with the 'New World' (1B/O:8)
8 strain. *J. Clin. Microbiol.* **49**, 1251 (2011).
9 38. Chen, P. *et al.* Genomic characterization of the *Yersinia* genus. *Genome Biol.*
10 **11**, R1 (2010).
11
- 12 39. Portnoy, D. A., Wolf-Watz, H., Bolin, I., Beeder, A. B. & Falkow, S.
13 Characterization of common virulence plasmids in *Yersinia* species and
14 their role in the expression of outer membrane proteins. *Infect. Immun.* **43**,
15 108–114 (1984).
16
- 17 40. Cornelis, G. R. & Wolf-Watz, H. The *Yersinia* Yop virulon: a bacterial system
18 for subverting eukaryotic cells. *Mol. Microbiol.* **23**, 861–867 (1997).
19
- 20 41. Grosdent, N., Maridonneau-Parini, I., Sory, M.-P. & Cornelis, G. R. Role of
21 Yops and Adhesins in Resistance of *Yersinia enterocolitica* to Phagocytosis.
22 *Infect. Immun.* **70**, 4165–4176 (2002).
23
- 24 42. Cornelis, G. R. & Wolf-Watz, H. The *Yersinia* Yop virulon: a bacterial system
25 for subverting eukaryotic cells. *Mol. Microbiol.* **23**, 861–867 (1997).
26
- 27 43. Bohn, E. *et al.* Gene expression patterns of epithelial cells modulated by
28 pathogenicity factors of *Yersinia enterocolitica*. *Cell. Microbiol.* **6**, 129–141
29 (2004).
30
- 31 44. Miller, V. L. & Falkow, S. Evidence for two genetic loci in *Yersinia*
32 *enterocolitica* that can promote invasion of epithelial cells. *Infect. Immun.*
33 **56**, 1242–1248 (1988).
34
- 35 45. Pierson, D. E. & Falkow, S. The *ail* gene of *Yersinia enterocolitica* has a role
36 in the ability of the organism to survive serum killing. *Infect. Immun.* **61**,
37 1846–1852 (1993).
38
- 39 46. Biedzka-Sarek, M., Venho, R. & Skurnik, M. Role of YadA, Ail, and
40 Lipopolysaccharide in Serum Resistance of *Yersinia enterocolitica* Serotype
41 O:3. *Infect. Immun.* **73**, 2232–2244 (2005).
42

- 1 47. Carniel, E. Plasmids and pathogenicity islands of *Yersinia*. *Curr. Top.*
2 *Microbiol. Immunol.* **264**, 89–108 (2002).
3
- 4 48. Pepe, J. C. & Miller, V. L. *Yersinia enterocolitica* invasin: A primary role in
5 the initiation of infection. *Proc Natl Acad Sci U S A* **90**, 6473–6477 (1993).
6
- 7 49. Schubert, S., Fischer, D. & Heesemann, J. Ferric enterochelin transport in
8 *Yersinia enterocolitica*: molecular and evolutionary aspects. *J. Bacteriol.*
9 **181**, 6387–6395 (1999).
10
- 11 50. Tomich, M., Planet, P. J. & Figurski, D. H. The *tad* locus: postcards from the
12 widespread colonization island. *Nat. Rev. Microbiol.* **5**, 363–375 (2007).
13
- 14 51. Iriarte, M. *et al.* The Myf fibrillae of *Yersinia enterocolitica*. *Mol. Microbiol.*
15 **9**, 507–520 (1993).
16
- 17 52. Delor, I. & Cornelis, G. R. Role of *Yersinia enterocolitica* Yst toxin in
18 experimental infection of young rabbits. *Infect Immun* **60**, 4269–4277
19 (1992).
20
- 21 53. Hinnebusch, B. J. *et al.* Role of *Yersinia* Murine Toxin in Survival of *Yersinia*
22 *pestis* in the Midgut of the Flea Vector. *Science (80)*. **296**, 733 (2002).
23
- 24 54. Zimble, D. L., Schroeder, J. A., Eddy, J. L. & Lathem, W. W. Early emergence
25 of *Yersinia pestis* as a severe respiratory pathogen. *Nat. Commun.* **6**, 7487
26 (2015).
- 27 Recent publication showing key role of Pla in the evolution of the unique
28 virulence of *Y. pestis*
- 29 55. Lesic, B. *et al.* Excision of the high-pathogenicity island of *Yersinia*
30 *pseudotuberculosis* requires the combined actions of its cognate integrase
31 and Hef, a new recombination directionality factor. *Mol. Microbiol.* **52**,
32 1337–1348 (2004).
33
- 34 56. Schubert, S., Picard, B., Gouriou, S., Heesemann, J. & Denamur, E. *Yersinia*
35 high-pathogenicity island contributes to virulence in *Escherichia coli*
36 causing extraintestinal infections. *Infect Immun* **70**, 5335–5337 (2002).
37
- 38 57. Carniel, E., Guilvout, I. & Prentice, M. Characterization of a large
39 chromosomal 'high-pathogenicity island' in biotype 1B *Yersinia*
40 *enterocolitica*. *J. Bacteriol.* **178**, 6743–6751 (1996).
41

- 1 58. Bobrov, A. G. *et al.* The *Yersinia pestis* siderophore, yersiniabactin, and the
2 ZnuABC system both contribute to zinc acquisition and the development of
3 lethal septicaemic plague in mice. *Mol. Microbiol.* **93**, 759–775 (2014).
4
- 5 59. Paauw, A., Leverstein-van Hall, M. A., van Kessel, K. P. M., Verhoef, J. & Fluit,
6 A. C. Yersiniabactin reduces the respiratory oxidative stress response of
7 innate immune cells. *PLoS One* **4**, e8240 (2009).
8
- 9 60. Haller, J. C., Carlson, S., Pederson, K. J. & Pierson, D. E. A chromosomally
10 encoded type III secretion pathway in *Yersinia enterocolitica* is important
11 in virulence. *Mol. Microbiol.* **36**, 1436–1446 (2000).
12
- 13 61. Matsumoto, H. & Young, G. M. Proteomic and functional analysis of the
14 suite of Ysp proteins exported by the Ysa type III secretion system of
15 *Yersinia enterocolitica* Biovar 1B. *Mol. Microbiol.* **59**, 689–706 (2006).
16
- 17 62. Young, B. M. & Young, G. M. Evidence for Targeting of Yop Effectors by the
18 Chromosomally Encoded Ysa Type III Secretion System of *Yersinia*
19 *enterocolitica*. *J. Bacteriol.* **184**, 5563–5571 (2002).
20
- 21 63. Walker, K. A., Maltez, V. I., Hall, J. D., Vitko, N. P. & Miller, V. L. A phenotype
22 at last: essential role for the *Yersinia enterocolitica* Ysa type III secretion
23 system in a *Drosophila melanogaster* S2 cell model. *Infect. Immun.* **81**,
24 2478–2487 (2013).
25
- 26 64. Perry, R. D., Pendrak, M. L. & Schuetze, P. Identification and cloning of a
27 hemin storage locus involved in the pigmentation phenotype of *Yersinia*
28 *pestis*. *J. Bacteriol.* **172**, 5929–5937 (1990).
29
- 30 65. Hinnebusch, B. J., Perry, R. D. & Schwan, T. G. Role of the *Yersinia pestis*
31 hemin storage (hms) locus in the transmission of plague by fleas. *Science*
32 **273**, 367–370 (1996).
33
- 34 66. Schroder, O. & Wagner, R. The bacterial regulatory protein H-NS--a
35 versatile modulator of nucleic acid structures. *Biol. Chem.* **383**, 945–960
36 (2002).
37
- 38 67. Cathelyn, J. S., Ellison, D. W., Hinchliffe, S. J., Wren, B. W. & Miller, V. L. The
39 RovA regulons of *Yersinia enterocolitica* and *Yersinia pestis* are distinct:
40 evidence that many RovA-regulated genes were acquired more recently
41 than the core genome. *Mol. Microbiol.* **66**, 189–205 (2007).

- 1 Publication showing the marked difference in the RovA regulons of the
2 pathogenic *Yersinia*
- 3 68. Ellison, D. W., Lawrenz, M. B. & Miller, V. L. Invasin and beyond: regulation
4 of *Yersinia* virulence by RovA. *Trends Microbiol.* **12**, 296–300 (2004).
5
- 6 69. Bucker, R., Heroven, A. K., Becker, J., Dersch, P. & Wittmann, C. The
7 pyruvate-tricarboxylic acid cycle node: a focal point of virulence control in
8 the enteric pathogen *Yersinia pseudotuberculosis*. *J. Biol. Chem.* **289**,
9 30114–30132 (2014).
- 10 The only published version of the *Y. pseudotuberculosis* RovA regulon
- 11 70. Bohme, K. *et al.* Concerted actions of a thermo-labile regulator and a
12 unique intergenic RNA thermosensor control *Yersinia* virulence. *PLoS*
13 *Pathog.* **8**, e1002518 (2012).
- 14 A comprehensive analysis of the mechanism of RovA/YmoA regulation
- 15 71. Koo, J. T., Alleyne, T. M., Schiano, C. A., Jafari, N. & Lathem, W. W. Global
16 discovery of small RNAs in *Yersinia pseudotuberculosis* identifies *Yersinia*-
17 specific small, noncoding RNAs required for virulence. *Proc. Natl. Acad. Sci.*
18 *U. S. A.* **108**, E709–17 (2011).
19
- 20 72. Quade, N. *et al.* Structural basis for intrinsic thermosensing by the master
21 virulence regulator RovA of *Yersinia*. *J. Biol. Chem.* **287**, 35796–35803
22 (2012).
23
- 24 73. Kapatral, V. *et al.* Gene array analysis of *Yersinia enterocolitica* FlhD and
25 FlhC: regulation of enzymes affecting synthesis and degradation of
26 carbamoylphosphate. *Microbiol* **150**, 2289–2300 (2004).
27
- 28 74. Perry, R. D. *et al.* Temperature regulation of the hemin storage (Hms+)
29 phenotype of *Yersinia pestis* is posttranscriptional. *J. Bacteriol.* **186**, 1638–
30 1647 (2004).
31
- 32 75. Fälker, S., Schmidt, M. A. & Heusipp, G. Altered Ca²⁺ Regulation of Yop
33 Secretion in *Yersinia enterocolitica* after DNA Adenine Methyltransferase
34 Overproduction Is Mediated by Clp-Dependent Degradation of LcrG. *J.*
35 *Bacteriol.* **188**, 7072–7081 (2006).
36
- 37 76. Grabenstein, J. P., Marceau, M., Pujol, C., Simonet, M. & Bliska, J. B. The
38 Response Regulator PhoP of *Yersinia pseudotuberculosis* Is Important for
39 Replication in Macrophages and for Virulence. *Infect. Immun.* **72**, 4973–
40 4984 (2004).

- 1
- 2 77. Rebeil, R. *et al.* Induction of the *Yersinia pestis* PhoP-PhoQ regulatory
3 system in the flea and its role in producing a transmissible infection. *J.*
4 *Bacteriol.* **195**, 1920–1930 (2013).
5
- 6 78. Pisano, F. *et al.* Influence of PhoP and intra-species variations on virulence
7 of *Yersinia pseudotuberculosis* during the natural oral infection route. *PLoS*
8 *One* **9**, e103541 (2014).
9
- 10 79. Nuss, A. M. *et al.* A direct link between the global regulator PhoP and the
11 Csr regulon in *Y. pseudotuberculosis* through the small regulatory RNA
12 CsrC. *RNA Biol.* **11**, 580–593 (2014).
13
- 14 80. Fookes, M. *et al.* *Salmonella bongori* provides insights into the evolution of
15 the Salmonellae. *PLoS Pathog.* **7**, e1002191 (2011).
- 16 Key publication highlighting the evolution of pathogenesis in *Salmonella*
- 17 81. Navarre, W. W. *et al.* Co-regulation of *Salmonella enterica* genes required
18 for virulence and resistance to antimicrobial peptides by SlyA and
19 PhoP/PhoQ. *Mol. Microbiol.* **56**, 492–508 (2005).
20
- 21 82. Ali, S. S. *et al.* Silencing by H-NS potentiated the evolution of *Salmonella*.
22 *PLoS Pathog.* **10**, e1004500 (2014).
- 23 Key publication showing the importance of H-NS in the evolution of SPI mediated
24 pathogenesis in *Salmonella*
- 25 83. Pallen, M. J. & Wren, B. W. Bacterial pathogenomics. *Nature* **449**, 835–842
26 (2007).
27
- 28 84. Morelli, G. *et al.* *Yersinia pestis* genome sequencing identifies patterns of
29 global phylogenetic diversity. *Nat. Genet.* **42**, 1140–3 (2010).
- 30 Comprehensive analysis of the pattern on fine scale and ongoing recent
31 evolution in *Y. pestis*
- 32 85. Song, Y. *et al.* Complete genome sequence of *Yersinia pestis* strain 91001,
33 an isolate avirulent to humans. *DNA Res* **11**, 179–197 (2004).
34
- 35 86. Bos, K. I. *et al.* A draft genome of *Yersinia pestis* from victims of the Black
36 Death. *Nature* **478**, 506–510 (2011).
37
- 38 87. Chouikha, I. & Hinnebusch, B. J. Silencing urease: a key evolutionary step

- 1 that facilitated the adaptation of *Yersinia pestis* to the flea-borne
2 transmission route. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18709–18714 (2014).
3
- 4 88. Sebbane, F., Devalckenaere, A., Foulon, J., Carniel, E. & Simonet, M.
5 Silencing and reactivation of urease in *Yersinia pestis* is determined by one
6 G residue at a specific position in the *ureD* gene. *Infect. Immun.* **69**, 170–
7 176 (2001).
8
- 9 89. Aleksic, S., Wuthe, H.H. *Yersinia enterocolitica* serovar 2a, wb, 3:b,c biovar
10 5 in hares and sheep. *Berl Munch Tierarztl Wochenschr* **110**, 176 (1997).
11
- 12 90. Swaminathan, B., Harmon, M. C. & Mehlman, I. J. *Yersinia enterocolitica*. *J.*
13 *Appl. Bacteriol.* **52**, 151–183 (1982).
14
- 15 91. Shea, J. E., Hensel, M., Gleeson, C. & Holden, D. W. Identification of a
16 virulence locus encoding a second type III secretion system in *Salmonella*
17 Typhimurium. *Proc Natl Acad Sci U S A* **93**, 2593–2597 (1996).
18
- 19 92. Schaake, J. *et al.* Essential role of invasin for colonization and persistence
20 of *Yersinia enterocolitica* in its natural reservoir host, the pig. *Infect.*
21 *Immun.* **82**, 960–969 (2014).
22
- 23 93. Schaake, J. *et al.* Human and animal isolates of *Yersinia enterocolitica* show
24 significant serotype-specific colonization and host-specific immune
25 defense properties. *Infect. Immun.* **81**, 4013–4025 (2013).
26
- 27 94. Uliczka, F. *et al.* Unique cell adhesion and invasion properties of *Yersinia*
28 *enterocolitica* O:3, the most frequent cause of human Yersiniosis. *PLoS*
29 *Pathog.* **7**, e1002117 (2011).
30
- 31 95. Valentin-Weigand, P. Heesemann, J. Dersch, P. Unique virulence properties
32 of *Yersinia enterocolitica* O:3 - An emerging zoonotic pathogen using pigs
33 as preferred reservoir host. *Int. J. Med. Microbiol.* **14**, 00090–3 (2014).
34
- 35 96. Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant
36 *Salmonella enterica* serovar Typhi CT18. *Nature* **413**, 848–852 (2001).
37
- 38 97. Thomson, N. R. *et al.* Comparative genome analysis of *Salmonella*
39 Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into
40 evolutionary and host adaptation pathways. *Genome Res.* **18**, 1624–1637
41 (2008).
42

- 1 98. McClelland, M. *et al.* Complete genome sequence of *Salmonella enterica*
2 serovar Typhimurium LT2. *Nature* **413**, 852–856 (2001).
3
- 4 99. Nuccio, S.-P. & Baumler, A. J. Comparative analysis of *Salmonella* genomes
5 identifies a metabolic network for escalating growth in the inflamed gut.
6 *MBio* **5**, e00929–14 (2014).
7
- 8 100. Badia, J., Ros, J. & Aguilar, J. Fermentation mechanism of fucose and
9 rhamnose in *Salmonella* Typhimurium and *Klebsiella pneumoniae*. *J.*
10 *Bacteriol.* **161**, 435–437 (1985).
11
- 12 101. Bry, L., Falk, P. G., Midtvedt, T. & Gordon, J. I. A model of host-microbial
13 interactions in an open mammalian ecosystem. *Science* **273**, 1380–1383
14 (1996).
15
- 16 102. Price-Carter, M., Tingey, J., Bobik, T. A. & Roth, J. R. The alternative electron
17 acceptor tetrathionate supports B12-dependent anaerobic growth of
18 *Salmonella enterica* serovar typhimurium on ethanolamine or 1,2-
19 propanediol. *J. Bacteriol.* **183**, 2463–2475 (2001).
20
- 21 103. Winter, S. E. *et al.* Gut inflammation provides a respiratory electron
22 acceptor for *Salmonella*. *Nature* **467**, 426–429 (2010).
- 23 Key account of the role of cob/pdu/ttr in inflammation and intestinal
24 colonisation by *Salmonella*
- 25 104. Thiennimitr, P. *et al.* Intestinal inflammation allows *Salmonella* to use
26 ethanolamine to compete with the microbiota. *Proc. Natl. Acad. Sci. U. S. A.*
27 **108**, 17480–17485 (2011).
28
- 29 105. Langridge, G. C. *et al.* Patterns of genome evolution that have accompanied
30 host adaptation in *Salmonella*. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 863–868
31 (2015).
32
- 33 106. Prentice, M. B. *et al.* Cobalamin synthesis in *Yersinia enterocolitica* 8081.
34 Functional aspects of a putative metabolic island. *Adv. Exp. Med. Biol.* **529**,
35 43–46 (2003).
36
- 37 107. Rabson, A. R., Hallett, A. F. & Koornhof, H. J. Generalized *Yersinia*
38 *enterocolitica* infection. *J. Infect. Dis.* **131**, 447–451 (1975).
39
- 40 108. Lian, C. J., Hwang, W. S., Kelly, J. K. & Pai, C. H. Invasiveness of *Yersinia*
41 *enterocolitica* lacking the virulence plasmid: an in-vivo study. *J. Med.*

- 1 *Microbiol.* **24**, 219–226 (1987).
2
- 3 109. Buret, A., O’Loughlin, E. V, Curtis, G. H. & Gall, D. G. Effect of acute *Yersinia*
4 *enterocolitica* infection on small intestinal ultrastructure. *Gastroenterology*
5 **98**, 1401–1407 (1990).
6
- 7 110. Tardy, F. *et al.* *Yersinia enterocolitica* type III secretion-translocation
8 system: channel formation by secreted Yops. *Embo J.* **18**, 6793–6799
9 (1999).
10
- 11 111. Wagner, D. M. *et al.* *Yersinia pestis* and the plague of Justinian 541-543 AD:
12 a genomic analysis. *Lancet. Infect. Dis.* **14**, 319–326 (2014).
13
- 14 112. Cui, Y. *et al.* Historical variations in mutation rate in an epidemic pathogen,
15 *Yersinia pestis*. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 577–582 (2013).
16
- 17 113. Rasmussen, S. *et al.* Early divergent strains of *Y. pestis* in Eurasia 5000 years
18 ago. *Cell.* **163**, 1-12 (2015).
19
- 20 114. Hall, M. *et al.* Use of whole-genus genome sequence data to develop a
21 multilocus sequence typing tool that accurately identifies *Yersinia* isolates
22 to the species and subspecies levels. *J. Clin. Microbiol.* **53**, 35–42 (2015).
23
- 24 115. Bottone, E. J. *Yersinia enterocolitica*: overview and epidemiologic
25 correlates. *Microbes Infect.* **1**, 323–333 (1999).
26
- 27 116. Barnes, P. D., Bergman, M. A., Meccas, J. & Isberg, R. R. *Yersinia*
28 *pseudotuberculosis* disseminates directly from a replicating bacterial pool
29 in the intestine. *J. Exp. Med.* **203**, 1591–1601 (2006).
30
- 31 117. Bottone, E. J. *Yersinia enterocolitica*: the charisma continues. *Clin.*
32 *Microbiol. Rev.* **10**, 257–276 (1997).
33
- 34 118. Sebbane, F., Jarrett, C. O., Gardner, D., Long, D. & Hinnebusch, B. J. Role of
35 the *Yersinia pestis* plasminogen activator in the incidence of distinct
36 septicemic and bubonic forms of flea-borne plague. *Proc. Natl. Acad. Sci. U.*
37 *S. A.* **103**, 5526–5530 (2006).
38
- 39 119. Sha, J. *et al.* Characterization of an F1 deletion mutant of *Yersinia pestis*
40 CO92, pathogenic role of F1 antigen in bubonic and pneumonic plague, and
41 evaluation of sensitivity and specificity of F1 antigen capture-based
42 dipsticks. *J. Clin. Microbiol.* **49**, 1708–1715 (2011).

43

1 **Box 1: Phylogenomics of *Yersinia* species: approaches and major findings.**

2 Traditionally, bacteria are speciated using limited biochemical tests that are
3 based on the phenotypic expression or non-expression of a trait. This can lead to
4 problems in standardising methods, and the potential to mutate the trait can
5 impact on the phenotype. Genomic information is less ambiguous, and variations
6 within genes under neutral selection can be used to infer evolutionary
7 relationships, providing a robust framework that is independent of phenotypic
8 variation. For the genus *Yersinia*, 84 housekeeping genes were identified, that
9 showed between 10-25% SNP divergence between *Y. pestis* and *Y. enterocolitica*⁸.
10 These criteria meet those of housekeeping genes used for the design of multi-
11 locus sequence typing (MLST) schemes¹¹⁴, so that the accumulation of variation
12 reflects the history of the genus, and a maximum likelihood phylogeny can be
13 estimated. Furthermore, Bayesian analysis of the population structure was
14 employed to identify defined species clusters (SC⁸). This means that statistical
15 probabilities are calculated to describe the variation hidden within the
16 population, and does not only consider the sequence of the 84 genes as a whole
17 but as separate gene entities. Some of the species clusters thus described contain
18 multiple species as defined by classical techniques, such as SCs 13 and 14. On the
19 other hand, other species like *Y. frederiksenii*, that show heterogeneity in their
20 phenotype, are distributed over several SCs (SCs 8, 9, and 14). This highlights the
21 limitations of biochemical tests that fail to accurately reflect the relationships
22 between the different species. The analysis also emphasizes the position of the
23 pathogenic species. The human pathogenic species are positioned at
24 diametrically opposite ends of the tree, with the fish pathogen *Y. ruckeri* (green)
25 forming a third outlying SC. Even though no classical root of the tree is given, *Y.*

1 *ruckeri* might be considered an out group since its position within the genus has
2 previously been under debate. The position of *Y. pseudotuberculosis*—*Y. pestis*
3 and *Y. enterocolitica* nevertheless supports independent evolution of their
4 pathogenic potential.

5 It must be noted however, that speciation might involve more than purely
6 genomic evidence, as *Y. pestis* and *Y. pseudotuberculosis* are highly similar on a
7 genomic level, yet exhibit markedly different mechanisms of disease as well as
8 niche preference and life style.

9

10 **Box 2: Mechanisms of pathogenesis in *Yersinia* spp.**

11 The three human-pathogenic species of *Yersinia* share a key genetic component
12 in their ability to cause disease, a large virulence plasmid named pYV (plasmid of
13 *Yersinia* Virulence). Though commonly thought of as a single entity, pYV is
14 actually highly variable between species and strains of species, containing
15 different origin of replications and showing variable genetic architecture^{8,12}.

16 What is common amongst them all is a large genetic locus encoding for the Ysc
17 type III secretion system (T3SS), responsible for the targeted delivery of Yop
18 effector proteins into host cells. The Yops are the key virulence factor in all three
19 of the pathogenic species. Upon contact with host macrophages Yops are injected
20 into the host cells, resulting in silencing of the pro-inflammatory cytokine
21 response and apoptotic death of the infected macrophages⁴⁰.

22 *Y. enterocolitica*/*Y. pseudotuberculosis* – are zoonotic infectious agents which
23 cause inflammatory intestinal disease associated with ingestion via consumption
24 of undercooked or contaminated pork products and vegetables¹¹⁵. The bacteria
25 enter the small intestine where they attach to the intestinal epithelium.

1 Numerous virulence factors have been implicated in this including the Myf, YadA,
2 and Ail adhesins as well as flagella¹⁸. The invasin protein then triggers
3 internalisation of the bacteria, which translocate across the epithelium⁴⁸.
4 Bacteria replicating in the intestinal tract can also disseminate to lymphatic
5 tissues via infection and silencing of phagocytic immune cells¹¹⁶. As a result, this
6 infection is often diagnosed as pseudo-appendicitis, and can lead to
7 septicaemia¹¹⁷.

8 *Y. pestis* – unlike *Y. enterocolitica* and *Y. pseudotuberculosis*, does not have an
9 intestinal phase to its infection-cycle. *Y. pestis* host reservoir is rodent fleas,
10 where it forms a biofilm in their foregut⁶⁵. When the flea next takes a blood meal
11 the biofilm is regurgitated into the host prey. If injected into the bloodstream the
12 bacteria immediately encounter phagocytic immune cells triggering activation of
13 the Ysc system and host-cell silencing, leading to septicaemic plague¹¹⁸. If
14 injected into the extravascular part of the dermis, the bacteria travel to the
15 lymph nodes where an acute infection is established leading to the formation of
16 painful swellings at the lymph nodes, or buboes¹¹⁸. This is known as bubonic
17 plague. If untreated the infection will spill back into the bloodstream with
18 bacteria travelling to the lungs, where pneumonic plague is established. *Y. pestis*
19 has acquired additional plasmids, pPla and pFra which play pivotal roles in the
20 insect-infection and systemic-infection phase of the organism^{53,54,119}.

21

22

23 Figure 1: Diagrammatic comparison of the RovA/YmoA regulons of *Y.*
24 *enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*. Genes activated by RovA are
25 shown in green, and genes repressed by RovA are shown in red. Genes in bold

1 share common RovA regulatory patterns between *Y. pseudotuberculosis* and *Y.*
2 *pestis*. Also shown are the loci activated and repressed by RovA in a conserved
3 fashion. This shows that the key acquired pYV plasmid (denoted by the *lcrF*
4 gene) is the only locus showing conserved RovA repression. The invasion
5 encoding gene *inv* is also shown as activated by RovA in a conserved fashion,
6 however the asterisk indicates that this is not entirely conserved, as *inv* is a
7 pseudogene in *Y. pestis*. The lack of common activated and repressed genes in the
8 3 species by a conserved regulatory mechanism highlights the key role played by
9 the RovA/YmoA systems in controlling expression of lineage defining loci.

10

11 Figure 2: Gene loss in the emergence of *Y. pestis* from *Y. pseudotuberculosis*.
12 Deletion events in genes encoding motility and the metabolic ability to colonise
13 the mammalian intestinal tract in ancestral *Y. pseudotuberculosis* occurred as a
14 result of a shift away from that environment towards an insect-vector phase of
15 the life-cycle. As a result there was selection for deletion of genes encoding
16 insect-toxins, including SNP-based reversible inactivation of the *ureD* gene,
17 encoding urease, which has high oral toxicity in fleas. For transmission from the
18 flea to occur *Y. pestis* must form biofilms in the flea foregut. SNP based
19 pseudogene inactivation of *rcaA* resulted in increased expression of the *hms*
20 operon which encodes biofilm formation. Similarly SNP based pseudogene
21 inactivation occurred in a gene encoding a phosphodiesterase. The product of
22 this gene degrades cyclic-di-GMP, a molecule that acts as a transcriptional
23 activator of *hms* and biofilm formation.

24

1 Figure 3: Gene gain and loss in the *Y. enterocolitica* species complex. The diagram
2 shows how the species is organised into phylogenetically distinct phylogroups,
3 which parallel LPS structure and serotype. PG 1 is most similar to the ancestral *Y.*
4 *enterocolitica*. The emergence of pathogenicity in the species is marked by the
5 acquisition of pYV, and in PG 2 the Ysa T3SS. The YGT T3SS is lost from all the
6 pathogenic phylogroups. The other significant event is the acquisition of
7 serotype specific LPS operons in each of the pathogenic phylogroups. PG 3 has
8 become the dominant isolate from pig reservoirs and human disease cases, and
9 has a deletion in the Flag-2 secondary flagella genes as well as an insertion in the
10 invasins promoter, which increases transcription of this key virulence factor.

11

12 Figure 4: Enteropathogenic *Salmonella* and *Yersinia* species use *cob/pdu* and *ttr*
13 loci to outcompete normal intestinal flora during mammalian infection. The *cob*
14 genes encode cobalamin biosynthesis. Cobalamin activates enzymes encoded by
15 *pdu* that degrade or ethanolamine. This process requires the use of tetrathionate
16 as a terminal electron acceptor in anaerobic respiration, the respiration of which
17 is encoded by the *ttr* genes. The enteropathogenic *Yersinia* and *Salmonella*
18 species induce inflammation during infection. This inflammation leads to over-
19 production of ethanolamine by the intestinal epithelium, which creates a hostile
20 environment for gastrointestinal commensal bacteria, but a favorable growth
21 environment for the enteropathogens. This leads to the enteropathogens
22 outcompeting the normal flora and overgrowth and colonization by the
23 pathogens.

24

1 Figure 5: A diagrammatic overview of the evolution of pathogenesis in the
2 *Yersinia* genus, highlighting key gene gain and gene loss events in the formation
3 of each of the pathogenic lineages. Acquisition of pYV occurs independently on
4 two occasions within *Y. enterocolitica* PG 2 and PGs 3-6, as well as additionally in
5 a *Y. pseudotuberculosis* ancestor. These two human-pathogenic lineages have
6 emerged on independent occasions from environmental generalists following
7 similar themes of gene gain, gene loss, metabolic reduction, and genome
8 rearrangements.

9

10 **Glossary**

11 Compensatory mutation – Mutations that occur to offset detrimental gene gain,
12 loss, or mutation events in independent parts of the genome.

13 Conjugation - The transfer of DNA – usually plasmids – between organisms via
14 direct cell-to-cell contact or a bridge between cells.

15 Cyclic-di-GMP – Secondary messenger molecule used in bacterial signal
16 transduction to modulate gene expression in response to environmental
17 perturbations.

18 Eco-evo perspective – A perspective in which organisms are evaluated broadly in
19 the light of evolution and ecology, rather than narrow constraints of their
20 behaviour in the laboratory or in human infection

21 F1 capsular protein – Protein antigen found on the surface of pathogenic *Yersinia*
22 thought to modulate targeting of bacteria to sites of infection.

23 Genomic islands - Large genetic regions that are part of the accessory gene pool.

24 They form the horizontally acquired part of a genome encoding one of more
25 functional groups of genes. They are frequently associated with tRNA genes and

1 are flanked by repeat structures, and contain mobility genes coding for
2 integrases or transposases required for chromosomal integration and excision.

3 Horizontal gene transfer - The transfer of DNA, frequently cassettes of genes,
4 between organisms. This process is in contrast to vertical gene transfer, which is
5 much more common and occurs when genetic material is passed from parent to
6 offspring or, more generally, from ancestor to descendent.

7 Integrons - A cassette of genes encoding a site-specific recombinase/integrase, a
8 recombination recognition site, and a promoter. Often found in conjunction with
9 other genes such as antibiotic resistance genes.

10 IS element - The simplest type of transposable element in a bacterium. Insertion
11 Sequence elements encode only the gene required for its own transposition, and
12 is flanked by repeats.

13 LEE pathogenicity island - Mobile genetic element encoding for a T3SS found in
14 enteropathogenic and enterohaemorrhagic *E. coli*. Injects effector molecule into
15 host intestinal epithelial cells resulting in actin rearrangement and formation of
16 characteristic attaching and effacing lesions.

17 MarR/SlyA - family of transcriptional regulators found in bacteria. Generally act
18 as activators of transcription by alleviating HN-S mediated repression.

19 Natural transformation - Direct uptake of DNA from the environment and
20 incorporation of this genetic material into the chromosome by competent cells.

21 Palaeomicrobiology - Study of ancient infectious disease outbreaks by
22 recovering nucleic acid sequences from ancient human remains.

23 Phenotypic microarray analysis - High-throughput, automated assays that
24 determine the ability of bacterial strains to metabolise metabolic substrates in
25 parallel.

1 Phylogenomic analysis – The use of whole genome sequences to create
2 phylogenetic trees and infer high-resolution evolutionary patterns. This is in
3 contrast to using phylogenetic markers such as 16S.

4 Pla – Plasminogen activator protein found in *Y. pestis* and encoded on a *Y. pestis*
5 specific plasmid. Protease required for pneumonic infection.

6 Promiscuous (regulon) – Promiscuous regulators assume transcriptional control
7 of large numbers of genes (regulons) that do not come under fine-scale
8 environmental control. Examples are HN-S, IHF, and FIS.

9 Salmonids - Relating to salmon and trout fish.

10 Sideophore - Compound enabling highly efficient capture of exogenous iron.

11 Transduction - The transfer of DNA – frequently cassettes of genes – between
12 organisms with the help of phages.

13 Two-component regulatory system - Bacterial sensor-kinase systems composed
14 of an outer membrane sensor, which autophosphorylates in response to a
15 specific environmental stimulus. This then leads to phosphorylation of a
16 response regulator that up and down regulates expression of operons.

17 Type III secretion system (T3SS) – A mechanism by which bacteria export
18 proteins. Often responsible for the translocation of bacterial effector proteins
19 from pathogenic or symbiotic bacteria directly into the cytoplasm of their
20 eukaryotic host, where these proteins subvert eukaryotic-cell functions to the
21 advantage of the bacterium.

22 Verotoxin – Shiga-like toxin produced by verotoxigenic *E. coli*. AB₅ toxin that
23 targets cells in the renal glomeruli leading to kidney damage.

24 Ymt – *Yersinia murine* toxin. First characterised as a determinant of lethality in
25 mice, now known to play a crucial role in the ability of *Y. pestis* to survive in fleas.

- 1 Yops – Yersinia outer proteins are a set of effector proteins secreted by the Ysc
- 2 T3SS found on the pYV plasmid in pathogenic *Yersinia* spp. Yops are injected into
- 3 phagocytic cells where they silence the production of pro-inflammatory cytokines
- 4 and induce apoptosis in the infected cell.
- 5

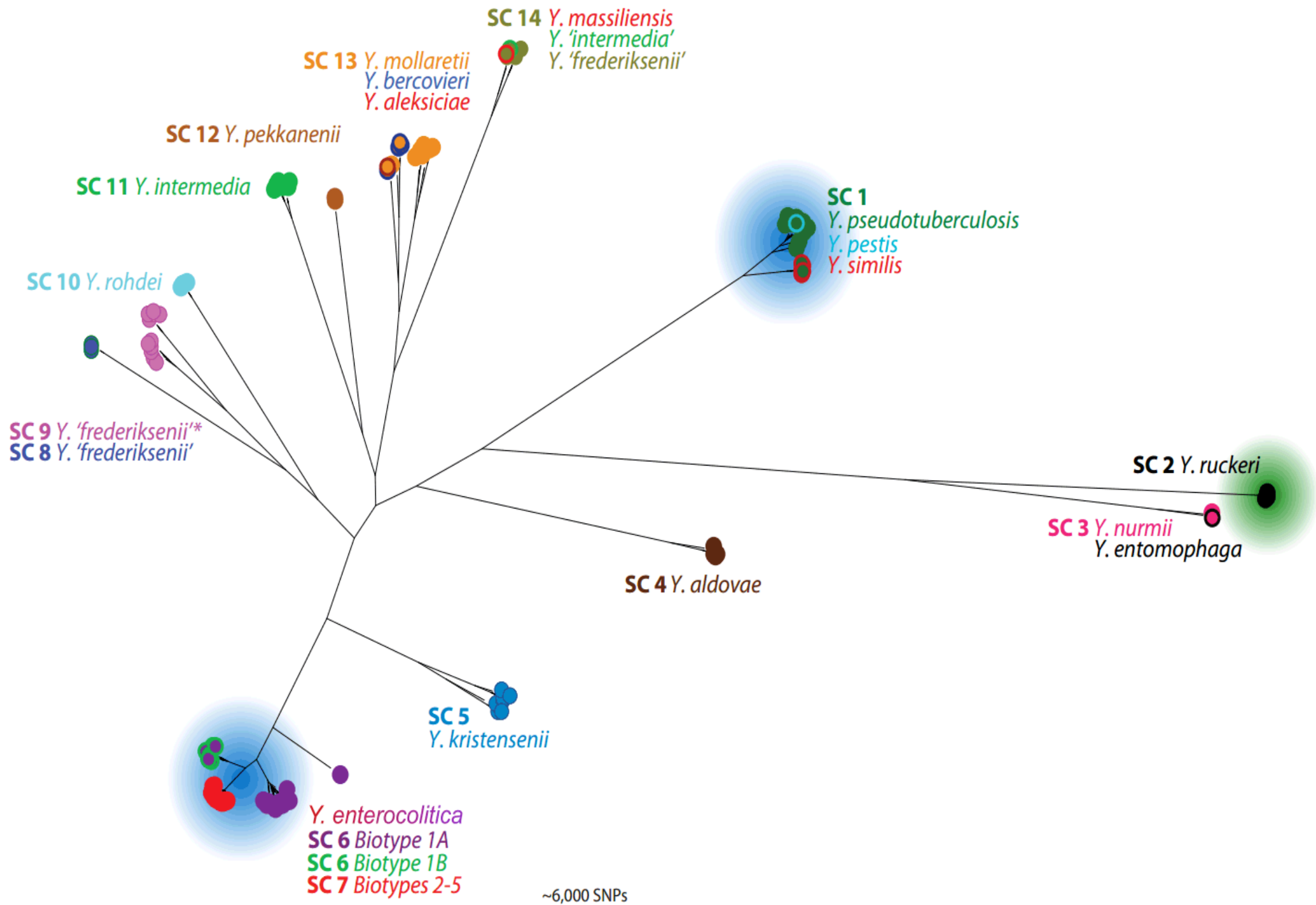
1 **Author Biographies**

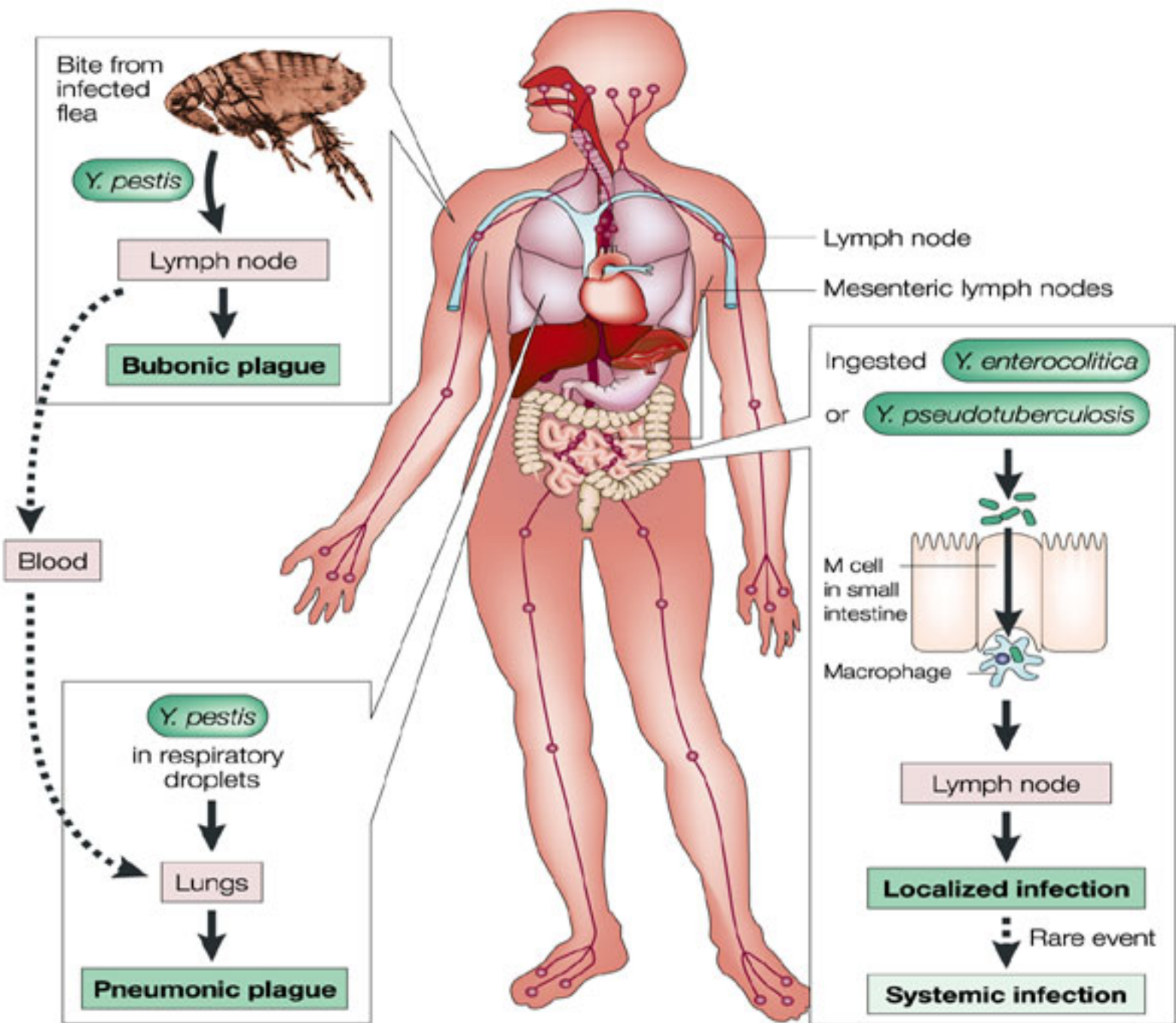
2 Alan McNally is a reader in Microbial Genomics at Nottingham Trent University.
3 He started working on *Yersinia* in 2003 as a post-doc in the lab of Dianne Newell
4 at the veterinary laboratories agency in Surrey. Upon moving to NTU he
5 continued his work on *Yersinia*, and now has his own group working on the
6 evolution of pathogenesis in *Yersinia*, and the emergence of drug resistant
7 lineages of extra-intestinal pathogenic *E. coli*.

8 Nick Thomson is a faculty member at the Wellcome Trust Sanger Institute. He is
9 a microbiologist and bioinformatician and is interested in bacterial evolution and
10 spread with a focus on sexually transmitted and diarrhoeal diseases. He is also
11 Professor of Bacterial Genomics and Evolution at the London School of Hygiene
12 and Tropical Medicine.

13 Sandra Reuter is a postdoctoral research associate at the University of
14 Cambridge. She undertook her PhD with Alan McNally and Nick Thomson on
15 *Yersinia* whilst at Nottingham Trent University, and continued some of this
16 research at the Wellcome Trust Sanger Institute. She is now working on hospital-
17 acquired infections, mainly methicillin-resistant *Staphylococcus aureus* (MRSA),
18 and antimicrobial resistance.

19 Brendan Wren is Professor of Microbial Pathogenesis at the London School of
20 Hygiene and Tropical Medicine since 1999. His research group exploits a range
21 of post genome research strategies to gain a comprehensive understanding of
22 how pathogens function and how they interact with their respective hosts.
23 Current research focuses on glycosylation in bacterial pathogens and developing
24 a “glycotoolbox” for glycoengineering, comparative phylogenomics and the
25 evolution of bacterial virulence and mechanisms of bacterial pathogenesis.





Y. enterocolitica species complex

