'Add, stir and reduce'

McNally, Alan; Thomson, Nicholas R; Reuter, Sandra; Wren, Brendan W

DOI:
10.1038/nrmicro.2015.29

Document Version
Peer reviewed version

Citation for published version (Harvard):
McNally, A, Thomson, NR, Reuter, S & Wren, BW 2016, ''Add, stir and reduce': Yersinia spp. as model bacteria for pathogen evolution', *Nature Reviews Microbiology*, vol. 14, no. 3, pp. 177-190.
https://doi.org/10.1038/nrmicro.2015.29

‘Add, stir and reduce’: The Yersiniae as model bacteria for the evolution of mammalian pathogens

Alan McNally [1*], Nicholas R Thomson[2,3], Sandra Reuter[3,4] & Brendan W Wren[2]

[1]Pathogen Research Group, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS

[2]Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT

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

[4]University of Cambridge, Department of Medicine, Box 157 Addenbrooke’s Hospital, Hills Road, Cambridge CB2 0QQ

Abstract

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

Introduction

The ever-expanding microbial genome sequence data set, combined with an increased understanding of the pathogenesis and ecology of bacterial pathogens, is illuminating our view of the dynamics that have shaped the evolution of bacterial pathogens. It is now evident that many bacterial pathogens infect humans only incidentally and often produce virulence factors that are active against non-mammalian adversaries as diverse as insects, protozoa, nematodes, predatory bacteria and bacteriophage[1–3]. These hosts provide a considerable driving force in the evolution of bacterial pathogens enabling us to re-evaluate human—pathogen interactions in the light of bacterial ecology and evolution (the eco-evo perspective). A consequence of the eco–evo perspective is that it is not surprising that genes encoding 'virulence factors' are found in pathogens and non-pathogens colonizing diverse mammalian and non-mammalian hosts[4,5]. Conversely, as bacteria evolve to inhabit new hosts, or new niches within the same hosts, using genomics we can see remnants of genes and gene clusters that are thought to have provided an adaptive advantage in the past, but that are now non-functional and represent evolutionary baggage.

Thus, many bacteria seem to be pre-armed with virulence determinants as they traverse new environments and niches, and their evolution is not solely dependent on mammalian/human–pathogen interactions[6,7]. The yersiniae represent a genus for which the eco-evo perspective is highly pertinent and for which all 18 species in the genus have recently been fully sequenced, metabolically phenotyped and compared[8]. In addition, this genus includes both non-pathogenic and pathogenic species, and a diversity of pathogenesis phenotypes so it represents a key model for understanding the forces at play in

the evolution of pathogenic bacteria. Our current understanding of the *Yersinia* encompasses a long-term view of how broad host range pathogenic species such as *Yersinia enterocolitica* evolve over millions of years from a non-pathogenic ancestor[9]. It also encompasses short term evolution examining how recent evolutionary bottlenecks have led to rapid emergence of high-pathogenic clones from their ancestral lineage such as the emergence of *Yersinia pestis* from *Yersinia pseudotuberculosis*[10]. As a model for evolution of pathogenesis, the evolution of pathogenic *Yersinia* spp. shows remarkable parallels with members of the genus *Salmonella,* which encompasses the broad host range self-limiting diarrheagenic *Salmonalle enterica* subsp. *enterica* serovar Typhimurium as well as those that are host restricted and acutely pathogenic, such as *Salmonella enterica* subsp. *enterica* serovar Typhi.

This has also made the pathogenic *Yersinia* species the organisms of choice for much of the pioneering work performed understanding the mechanisms of bacterial pathogenesis through classical genetics. Studies using *Y. enterocolitica* have been at the forefront of developing our understanding of how bacteria invade host epithelial cells[11]. Studies of the pathogenic *Yersinia* also forged our understanding of the role of plasmids in bacterial virulence[12,13], the identification and characterization of Type III secretion systems and their role in host subversion[14]. This in turn led to advances in our understanding of how bacteria evade host responses during infection[15–17].

In this Review, we examine how the eco-evo perspective, informed by whole-genome sequencing and by population genomic and ecological studies, has challenged previous hypotheses on the evolution of mammalian pathogenesis in

Yersinia spp. Furthermore, we consider how these studies are revealing common principles in the emergence of pathogenesis among other *Enterobacteriaceae*. Finally, we show how considering all members of a genus in population genomic studies, and not just pathogenic species, can provide greater insight and resolution for understanding of the processes underpinning the evolution of mammalian pathogenesis.

The yersiniae

The *Yersinia* is a genus of Gram-negative bacteria belonging to the *Enterobacteriaceae*. The genus nomenclature is based on classical systematics and biochemical speciation methods, resulting in a highly diverse group of bacteria with 18 defined species[18–26]. The best characterized of these are the three species that cause disease in humans: *Y. enterocolitica* and *Y. pseudotuberculosis* are zoonotic pathogens that cause self-limiting gastroenteritis in humans, whereas *Y. pestis* is a rodent and flea pathogen that is occasionally transmitted to humans and the causative agent of bubonic and pneumonic plague[27]. The remaining 15 known species of this genus are commonly found in soil and water environments and are considered to be non-pathogenic, with the exception of *Yersinia ruckeri*, the causative agent of red mouth disease in salmonids[21]. Recent phylogenomic analysis of the entire *Yersinia* genus allowed for an accurate, whole genome SNP-based, assessment of the population structure of the genus[8]. This analysis identified 14 distinct species complexes within the genus with the mammalian pathogenic *Y. enterocolitica* species and *Y. pseudotuberculosis* species complex (also containing *Y. pestis* and *Y. similis*) at opposite ends of the *Yersinia* phylogenic spectrum (Box 1).

The observed diversity of pathogenesis within the *Yersinia*, from avirulent to highly virulent, has provided a useful framework for the study of pathogenic mechanisms and of the evolution of virulence in Gram-negative bacteria. Early microbial population genetics studies showed that *Y. pestis* is a recently emerged clone of the enteropathogen *Y. pseudotuberculosis,* yet has a markedly different lifestyle and causes a much more acute disease[28,29]. The evolutionary narrative tells how *Y. pestis* evolved from the gastrointestinal pathogen, in an evolutionary blink of an eye (estimated to be between 2,000 and 10,000 years ago[28,30]) through the processes of gene gain, gene loss and genome rearrangements. This finding was further explored with early microbial comparative genomics studies. These revealed that *Y. pestis* had emerged from *Y. pseudotuberculosis* through a mixture of both gene gain events, in the form of acquisition of species-specific plasmids, bacteriophage, integrons and genomic islands, as well as multiple complete or partial gene loss events leading to a reduction in genomic flexibility and metabolic streamlining. All of this is associated with its change in lifestyle and ultimate niche restriction[10,31–33]. The relationship of *Y. enterocolitica* to the *Y. pseudotuberculosis*—*Y. pestis* complex was until recently poorly defined. In 2006, a comparison of the available genomes from all three species suggested that they shared a distant but common pathogenic ancestor which had acquired key pathogenicity determinants including the essential virulence plasmid pYV, encoding the Ysc prototype type III secretion system (T3SS), before splitting into distinct lineages[9,27]. However, the availability of just a single *Y. enterocolitica* genome sequence at this time limited the interpretations that could be made. Further genomes were subsequently published for *Y. enterocolitica* in 2011 and 2013[34–37], as well as singular genomes for type strains of environmental species

in 2010[38] but our understanding of the evolution of this whole genus and where *Y. enterocolitica* fitted into the evolutionary scale of *Yersinia* species was still limited by the lack of comprehensive whole genome data. Nonetheless, even with this limited genomic data, a combination of gene gain and gene loss in the three human pathogenic *Yersinia* species was evident and highlighted the complexity of evolution of mammalian pathogenicity in these bacteria, showing that pathogens can evolve by a process of "add DNA, stir, and reduce"[27].

The study of the evolution of pathogenesis in the genus also suggested an inexorable distinct relationship between loss of metabolic function by genome decay, reduction in ecological flexibility, and the evolution towards niche-restricted highly pathogenic lineages[27]. As population genomic studies have become an established tool in microbiology, the level of resolution at which we can study these evolutionary events has improved considerably. By genome sequencing across entire bacterial genera it is possible to determine the real key acquisition and loss events in the evolution of bacterial pathogenesis. It is also possible to shed new light on the evolutionary forces and patterns underpinning these processes, and identify their common features across bacterial genera.

The role of gene gain

The evolutionary path from environmental or commensal bacterium to human pathogen has long been associated with the gain of mobile genetic elements via horizontal gene transfer, through phage transduction, plasmid uptake, natural transformation and DNA conjugation. Events such as the acquisition of the locus of enterocyte effacement (LEE) pathogenicity island and the verotoxin-encoding prophages were critical events in the genesis of pathogenic lineages of *Escherichia coli* such as the enterohaemorrhagic *E. coli* (EHEC) serotype

O157:H7[39]. Similarly the acquisition of toxin-encoding bacteriophage in bacteria such as enterotoxigenic *E. coli* (ETEC), *Vibrio cholerae, Clostridium botulinum*, and *Clostridium tetani* are classic examples of gene gain that promoted mammalian pathogenesis[40].

In the *Yersinia* genus there seems to have been 'foothold moments' characterized by acquisitions that, looking back across the genus phylogeny appear as pivotal steps in the emergence of mammalian pathogenic lineages. The most important of these seems to have been the acquisition of the family of pYV virulence plasmids encoding the Ysc prototype T3SS[8,12,13,41]. The Ysc T3SS delivers Yop effector proteins directly into host cells upon contact, resulting in the cessation of actin polymerization for endocytosis (which promotes bacterial uptake) and in the suppression of host-cell transcription[42]. Following contact with host macrophages, the activities of Yop effectors silence the innate immune response, resulting in the down-regulation of pro-inflammatory cytokine production and allowing proliferation of the infection[43–45], a process termed the "*Yersinia* deadly kiss".

The advantage of sequencing broadly across a genus and deep within the species is that the origins and distribution of all genes described as *Yersinia* spp. virulence factors, or otherwise, in the published literature can be ascertained[8] (A strategy that has yet to be applied to other enteric pathogens). Such an analysis confirmed that pYV was one of only two virulence-associated genetic loci that were uniquely present in all three *Yersinia* lineages commonly associated with mammalian disease (pYV is absent from non-pathogenic species). The other was the chromosomally encoded attachment and invasion locus *ail* which has a role in the attachment to, and invasion of, mammalian epithelial cells as well as in

resistance to killing by the host complement system[46–48]. The pivotal role of pYV in the emergence of mammalian pathogenesis is highlighted by the revelation that, contrary to accepted wisdom[9,27,49], closely related versions of this plasmid were acquired independently on at least three occasions in the *Yersinia* genus; twice by the *Y. enterocolitica* and once by the *Y. pseudotuberculosis—Y. pestis* lineage[8], and that *Y. enterocolitica* did not share a common pathogenic ancestor with *Y. pseudotuberculosis* as previously thought[8]. As such, multiple independent *ail* and pYV acquisitions were key gene gain events that marked the parallel independent evolution of mammalian pathogens in the *Yersinia* genus. This is surprising if one considers the large number of virulence-associated genes that have been reported in *Yersinia* based on published classic bacterial genetics experiments. This includes key virulence factors such as Invasin[50], the Fes/Fep iron acquisition system[51], the Tad type IVb pilus[52] and Myf fimbriae[53], and the Yst heat stable toxin[54], all of which are widely distributed across pathogenic and non-pathogenic members of the genus. This could be explained through the eco-evo perspective whereby the environmental yersiniae may have acquired and maintained some virulence factors to combat predatory protozoa, nematodes, predatory bacteria, and bacteriophage or even to colonise plants. This highlights the benefits of studying the accessory gene pool of pathogenic bacteria in the wider context of the genera they belong to, a strategy that has yet to be applied to other enteric pathogens. Indeed in order to confirm the eco-evo hypothesis there is also a benefit and need to studying the role of virulence genes in these non-pathogenic isolates to allow us to fully determine the true ecology of bacterial species.

In addition to the crucial gains of the pYV plasmid and *ail* locus in all three human pathogenic *Yersinia* spp., a number of other acquired loci have been maintained by clonal expansion in individual *Yersinia* species complexes and phylogroups (PG). These include the well-characterized pFra (pMT1) and pPla (pPCP1) plasmids of *Y. pestis*, which encode the murine toxin Ymt and the F1 capsular protein and the plasminogen activator Pla (which has recently been shown to be pivotal for causing fulminant pneumonic infection[55]) respectively. Also in this group of acquired elements is the High Pathogenicity Island (HPI). HPI is a large integrative-conjugative element capable of excision and transfer via the use of self encoded integrase and excision factors[56]. As such the HPI is often found in species beyond *Yersinia* including extra-intestinal pathogenic *E. coli*[57]. The HPI encodes the siderophore yersiniabactin that sequesters iron in the host[57,58]. Within *Yersinia* spp., HPI containing a fully functioning yersiniabactin is found only in strains belonging to the *Y. pseudotuberculosis—Y. pestis* species complex, where its presence is isolate dependant, and in the highly virulent phylogroup 2 lineage of the *Y. enterocolitica* species complex, where it is uniformly present. The remaining gene acquisition events that have occurred in the pathogenic members of the genus involve the Ysa T3SS by phylogroup 2 *Y. enterocolitica*, and Hms by the *Y. pseudotuberculosis—Y. pestis* species complex. The Ysa T3SS is located in a region of the genome known as the plasticity zone[59], first identified in the archetypal phylogroup 2 strain 8081[9]. The plasticity zone is a region of the *Y. enterocolitica* genome that shows signs of multiple independent, and phylogroup specific acquisitions[9]. Ysa T3SS has been shown to secrete a large number of chromosomally encoded T3SS effector proteins[60] as well as the Yop T3SS effector proteins of the Ysc system carried on pYV[61].

Although mutagenesis studies suggest *ysa* plays only a minor role in mammalian pathogenesis[60], consistent with the eco-evo perspective *ysa* mutants were shown to be attenuated in their ability to survive inside cultured *Drosphilla melanogaster* cells suggesting *ysa* is more important outside of the mammalian host for phylogroup 2 *Y. enterocolitica*[62]. The Hms locus is chromosomally encoded and encodes a hemin storage locus that is unique to the *Y. pseudotuberculosis-Y. pestis* lineage and is responsible for its ability to form pigmented colonies on congo-red agar[63]. Hms is vital for *Y. pestis* pathogenesis as it is responsible for formation of biofilms, required for the blockage of the flea foregut, which facilitates its subsequent transmission to mammals through reflux of infected blood from the infected feeding flea back onto the bite site[64].

Regulatory control

Although the acquisition of mobile genetic elements has been a key process in the emergence of *Yersinia* spp. pathogenesis, loci are only stably maintained in a population if there is a minimal fitness cost to their integration into the genome of the host bacterium, or a high selective pressure for them to be maintained. To achieve this, many acquired elements fall under strict transcriptional and/or translational control to minimise the fitness costs associated with their acquisition. Across the *Enterobacteriaceae* many acquired genetic elements are regulated by H-NS, a nucleoid associated, histone-like protein that binds to A-T rich regions of DNA silencing transcriptional activity[65]. Once under the strict negative control of H-NS, the acquired elements are then integrated into existing transcriptional activating regulons within the bacterial cell. These function in an antagonistic fashion with H-NS to tightly control the transcriptional expression

of the new genes. In the case of *Yersinia* spp., the chromosomal *rovA/ymoA* regulon has been suggested to fulfil this antagonistic role to H-NS[8,66]. RovA is universally present across the whole genus[8] and is a member of the MarR/SlyA family of global transcriptional regulators[66]. It is known to antagonise H-NS silencing of the invasin protein gene, *invA*, which is involved in the early stages of *Yersinia* spp. attachment to cells and internalization during infection. The binding of RovA to the promoter region masks the H-NS binding site, triggering transcription[67]. Genes *rovA/ymoA* are thought to encode a promiscuous ancestral regulon into which the expression of acquired genes were incorporated multiple times in the different lineages. This is thought to explain why the RovA regulon includes different gene sets in *Y. enterocolitica* and *Y. pseudotuberculosis*[66]. RovA was shown to transcriptionally activate 73 genes in *Y. pestis* and 63 genes in *Y. enterocolitica*, but only three genes were common to both species[66]. The majority of RovA controlled genes are lineage-specific virulence factors such as Invasin or Myf fimbriae, or lineage specific prophage or metabolic genes[66]. The *rovA/ymoA* regulon also controls the expression of genes encoded on the pYV plasmid, highlighting its importance as a global regulator of virulence.

The expression of the Yop proteins encoded on pYV is down-regulated at a transcriptional level by LcrF (an ortholog of HN-S encoded by the plasmid), in response to temperature. Recent work has identified that this regulation is controlled at a hierarchical level by YmoA, which functions as a thermo-sensing switch to turn off *lcrF* transcription, and therefore Yop translation and secretion at low temperatures[68]. This thermo-sensing switch is a tandem process which relies on weak binding of YmoA directly to the promoter region of *lcrF* at low temperatures and the formation of a double stem loop in the intergenic region

resulting in stearic hindrance of RNA polymerase binding and formation of transcripts[68,69]. Overall the importance of the RovA/YmoA system for the evolution of pathogenesis in *Yersinia* is evident from these data: At temperatures below 37^{O}C virulence factors, essential for mammalian pathogenesis, are not expressed because RovA/YmoA is inactive and they are repressed by H-NS. However flagella biosynthesis genes (in the case of the enteropathogenic species) or insect survival loci including *hms* (in the case of *Y. pestis)* are expressed at high levels[70,71]. Upon a shift to mammalian body temperatures essential acquired virulence factors such as those on pYV, *ail* and *invA* are activated by RovA/YmoA (FIG. 1). Other environmental triggers are also known to influence this regulatory cascade such as levels of exogenous calcium ions[72] or levels of exogenous magnesium ions via the PhoPQ two component regulatory system[73].

The presence of an ancestral promiscuous regulatory system controlling expression of acquired genes involved in virulence draws striking parallels with *Salmonella* spp. The key acquisition event in the evolution of pathogenesis in the *Salmonella* genus was *Salmonella* pathogenicity island-1 (SPI-1). This pathogenicity island encodes a T3SS and its cognate secreted effector proteins that remodel the actin cytoskeleton of infected cells and force uptake of attached bacteria into endocytic vacuoles. As with pYV, the acquisition of SPI-1 was a landmark event in the evolution of the *Salmonella* genus, being present in both true species: *S. enterica* and *Salmonella bongori*[74]. SPI-1 is under negative transcriptional control by H-NS. Similar to *Yersinia*, H-NS-mediated repression of SPI-1 functions is alleviated by the concerted action of a number of transcriptional regulators including HilD/HilA, the two-component regulatory

system PhoPQ[75] and SlyA, an ortholog of RovA. Recent work showed that SPI-1 was readily lost in *Salmonella* H-NS null mutants, Showing that H-NS silencing of SPI-1 was likely to have been an essential factor in SPI-1 becoming fixed in all *Salmonella* spp. and so the evolution of pathogenesis in members of this genus too[76].

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

Role of gene loss

Although the importance of gene gain in the evolution of bacterial pathogenicity is intuitive and well documented, the role of gene loss and host adaptation has been shown to be equally as important in the emergence of bacterial pathogens[77]. The majority of work looking at the evolution of pathogenesis in *Yersinia* spp. has focused on the evolutionary events that led to the emergence of *Y. pestis* from *Y. pseudotuberculosis*[28]. Although there were important acquisition events, the most striking observation of early comparative genomic analyses was the functional loss of a large number of genes associated with virulence and metabolism in *Y. pestis*[10]. This functional loss applied to about 10% of the genes in the genome, a considerably larger proportion of the genome compared to gene gain events. The gene attrition was largely attributable to the expansion of four major classes of IS elements (IS*1541*, IS*100*, IS*285*, IS*1661*) leading to insertional gene inactivation. Other processes of gene loss included deletion of genes and loci and SNP-based pseudogene formation[32,33,78,79]. Many of the genes that were

functionally lost encode virulence factors such as the *tc*PAI insext toxin locus and hemolysin homologues and metabolic loci involved in motility, dicarboxylic amino-acid metabolism, and uracil transport[10,27]. These are required for successful colonization of the mammalian gastrointestinal tract, a niche no longer required by *Y. pestis* as it emerged from the intestinal pathogenic *Y. pseudotuberculosis* to colonize the mid-gut of fleas and, in mammals, to become a more systemic pathogen invading the lymphatic system[27].

The importance of gene loss in the evolution of *Y. pestis* has recently been exemplified by studies investigating key mutations in *Y. pestis* that mark its evolutionary adaptation to flea-borne transmission in mammals. Using a combination of comparative genomics and classical bacterial genetics, three key loss-of-function mutations were found, all of which increased biofilm formation in the flea foregut in a cyclic-di-GMP dependant manner[6]. The key mutations were in two genes encoding EAL-domain phosphodiesterases (PDE), and the *rcsA* gene that is a component of the *rcs* regulatory network. The *rcs* system transcriptionally represses the *hms* locus and biofilm formation. PDE are enzymes that degrade cyclic-di-GMP, a bacterial signalling molecule that enhances transcription of *hms* and biofilm formation[6]. The functional loss of the PDE enzymes and the regulator repressing transcription in a *Y. pseudotuberculosis*-ancestor led to the emergence of a lineage with enhanced ability to form biofilms in the foregut of fleas and a resulting shift in pathogenic lifestyle of *Y. pestis*[6]. The relatively recent emergence of *Y. pestis*[28] is consistent with the SNP-based pseudogenes having been swept to fixation through the *Y. pestis* population but not yet having been deleted in part or full[30,80]. Also in *Y. pestis* is a mutation in the *ureD* gene that leads to functional loss of urease. In *Y.*

pseudotuberculosis it has been shown that the functional *ureD* encoding urease is associated with significant oral toxicity in the flea[81], hence the *ureD* mutation may play a contributory role in the emergence of *Y. pestis*[6,81](Fig 2). However, perhaps significantly the *ureD* mutation appears to be a reversible phase-variable mutation and maybe a contingency gene required in the mammalian host, rather than the insect vector, part of its life cycle. This provides *Y. pestis* the flexibility of retaining urease expression, as and when required.

Genomic analysis of *Y. enterocolitica* showed that similar gene loss and metabolic restriction also mark the evolution of certain subtypes of this species[8]. As discussed above and shown in Figure 3 *Y. enterocolitica* can be subdivided into the non-pathogenic phylogroup PG 1, which is an ancestral lineage, the highly pathogenic PG 2, and PGs 3-5 which show limited pathogenicity in mouse models but are the most successful lineages in terms of disease causation and are most commonly isolated from livestock and human clinical cases. PG 6 is a rare lineage which has only ever been isolated from wild hares[82,83]. The evolutionary descent from host-generalists (PGs 3-5) to the host-restricted PG 6, like the emergence of *Y. pestis*, was accompanied in *Y. enterocolitica* by the expansion of a single IS element, IS*1667*, as well as functional loss of a large number of metabolic genes**.** The impact of the functional loss of metabolic loci is particularly pronounced in PG 6 and is thought to reflect its extreme niche restriction to hares. Interestingly the list of metabolic pathways inactivated in PG 6 is similar in part to those lost in *Y. pestis* (cobalamin biosynthesis, tetrathionate respiration, hydrogenase-4). These mutational observations are directly supported by high throughput metabolic phenotyping of PG 6[8].

Looking at the functions that were lost as different lineages of *Y. enterocolitica* diverged and adapted to new niches, one of the earliest events is likely to have been the decay of the *Yersinia* genus T3SS, YGT. YGT is an ancestral T3SS present across the whole genus and is distinct from the Ysc and Ysa T3SSs acquired in pathogenic lineages. YGT has yet to be fully characterized but it shares a high level of sequence identity with the T3SS encoded on the SPI-2 genomic island in *Salmonella enterica*, a key T3SS involved in intracellular survival[84]. The YGT present in non-pathogenic PG 1 *Y. enterocolitica* seems to contain a full complement of effector protein encoding genes compared with the SPI-2 T3SS[8]. In the low-pathogenic PGs 3-6 the YGT seems to have been inactivated either by SNP-based pseudogene formation or deletion of the secretion apparatus genes, whereas almost all traces of the YGT have been deleted in the highly pathogenic PG 2 isolates[8]. As an example of how dynamic the evolution of pathogenesis is through gene gain and loss, the loss of YGT in the high-pathogenic PG 2 *Y. enterocolitica* is, conceptually at least, offset by the gain of the Ysa T3SS[59]. A full functional characterisation of YGT is now required to determine if the selective maintenance of YGT in PG 1 *Y. enterocolitica* is a classical example of eco-evo principal, where the T3SS is conferring protection to the organism in a non-mammalian host environment.

Another important role played by IS expansion in *Y. enterocolitica* is observed in the low-pathogenic PG 3 lineage (comprising the classically entitled serotype O:3 strains)[85,86]. Molecular analysis revealed a IS*1667* had inserted into the *invA* invasin promoter region, resulting in an additional recognition site for binding of RovA and a concomitant increase in expression of *invA*[87]. Comparative analysis with the other low-pathogenic phylogroups of *Y. enterocolitica* suggest this is the

primary event leading to the enhanced virulence properties of PG 3 isolates compared to PGs 2, 4, and 5[86]. As a result of the increased expression of invasion PG 3 isolates can colonise porcine tissues more effectively than other phylogroups, and as result PG 3 has become the predominant isolates taken from human clinical cases in Europe over the past 10 years and also dominate veterinary epidemiological surveys[88].

Gene loss in other *Enterobacteriaceae*

Functional gene loss in metabolic loci has been a key event in the *Yersinia* genus that mediated the transition from ubiquitous host-generalist lineage to niche restricted and host-adapted pathogen on at least two occasions. Once again this observation has also been made in *Salmonella* species. *S.* Typhi[89] shows many similarities with *Y. pestis,* it is niche restricted (in this case to the human host) and causes an acute systemic disease known as typhoid. Similarly *S.* Gallinarum and *S.* Pullorum also cause a typhoid-like disease but are restricted to galliforme birds[90], whereas *S.* Typhimurium is a host-generalist causing an inflammatory intestinal infection of mammals[91]. Comparative genomic studies of these *Salmonella* spp. have shown that functional gene loss in loci involved in anaerobic metabolism (more specifically functional loss in the *cob/pdu* and *ttr* operons) was a marker of the transition from intestinal pathogen like *S.* Typhimurium to invasive pathogens such as *S.* Typhi, Gallinarum and Pullorum[92]. In *S.* Typhimurium under anaerobic conditions, the *cob* genes encode the ability to synthesis cobalamin (or vitamin B12). Cobalamin is required as a cofactor for several enzymes including the first enzyme in the 1, 2-propanediol utilisation pathway (encoded by *pdu* operon). 1,2-Propanediol is a by-product of

fermentative growth on rhamnose and fucose[93], common constituents of plant cell walls and intestinal epithelial cells (and therefore the gut)[94]. The anaerobic degradation of 1,2-propanediol requires tetrathionate for use as a terminal electron acceptor, facilitated by the products of the *ttr* genes[95]. Unlike 1,2-propanediol , tetrathionate was only previously known to be present in soils not in the mammalian gut. However, it has been recently shown that tetrathionate is produced naturally during an inflammatory response to a *Salmonella* infection. It is under such conditions that the *cob*, *pdu* and *ttr* functions combine to provide a competitive growth advantage for *S*. Typhimurium, by allowing it to outgrow the natural largely fermentative gut microbiota by using naturally occurring carbon sources, 1,2-propanediol, that are not readily fermented[96,97] (FIG. 4). As the invasive typhoidal *Salmonella* have evolved away from an intestinal lifestyle, functions in *cob/pdu/ttr* operons have been sequentially lost and so represent markers of their change in niche and a movement away from causing intestinal disease to a more systemic infection cycle[98].

Although the *cob/pdu* and *ttr* loci are present and intact in *Y. enterocolitica* and in the majority of environmental *Yersinia* species[99], they are absent from the *Y. pseudotuberculosis*—*Y. pestis* complex. As mentioned above, tetrathionate is produced in the vertebrate inflamed gut following the host's response to *S.* Typhimurium infection, whereby the SPI-1 and -2 encoded T3SS are essential for stimulating this inflammatory response. Studies of *Y. enterocolitica* pathogenesis have shown that infections are characterized by inflammation of the gut[100]. Moreover, inflammation appears to require pYV[101]. It is possible that by inducing a pYV-associated inflammatory response during intestinal infection[102], and by expression of the *cob/pdu* and *ttr* loci, *Y. enterocolitica* is able, like *S.*

Typhimurium, to gain a metabolic advantage over the resident gut microbiota by using tetrathionate and naturally occurring carbon sources to respire under these conditions. The fact that the *cob/pdu* and *ttr* loci are lost completely in *Y. pseudotuberculosis* – *Y. pestis* may explain why these pathogens are incapable of causing the inflammatory intestinal infection observed in *Y. enterocolitica,* despite containing pYV.

Summary and concluding remarks

For decades the pathogenic yersiniae have been at the vanguard of understanding bacterial pathogenesis. Examples include the role of T3SSs[44,103] and understanding the molecular mechanisms by which intestinal pathogens invade the intestinal epithelium[46]. Work on the pathogenic *Yersinia* species also drove the understanding of the importance of plasmids in conferring virulence to pathogenic bacteria[12,41]. The yersiniae are also at the forefront in developing our understanding of the evolution of mammalian pathogenesis. Furthermore, *Yersinia* spp. genomics has also been at the frontline of palaeomicrobiology, with the near complete reconstruction and interpretation of the Black Death and Justinian plague *Y. pestis* genomes from ancient DNA studies on historical disease episodes[80,104].

The position of *Yersinia* as a model genus for pathogenesis has been further cemented as it is the first multispecies bacterial genus for which all species have been sequenced and placed into phylogenomic context[8]. By taking this approach coupled with high throughput metabolic analyses using phenotype microarrays it was possible to trace the evolutionary origins of the species, contextualise them and reveal the parallel evolutionary paths of virulence in humans[8] (FIG. 5). The genus *Yersinia* evolved from environmental ancestors. The most outlying

species are the fish pathogenic *Y. ruckeri* (SC 2), and newly classified lineages *Y. nurmii* and *Y. entomophaga* (SC 3). Previously, it had been debated whether *Y. ruckeri* was truly part of the genus *Yersinia*[21], however this has now been strengthened by the recognition of this novel lineage. The human pathogenic species clusters emerged independently following the common themes of 'add, stir, and reduce'. *Y. pestis* acquired its unique virulence plasmids, an expansion of several IS elements led to widespread genome rearrangements, and considerable gene loss can be observed. However, in the *Y. enterocolitica* lineages, these themes are weighted differently in the various phylogroups. PG 2 added the high-pathogenicity island, and type 2 and 3 secretions systems, and on the other hand lost the second flagella cluster and the genus T3SS YGT. PGs 3-6 contain the toxin complex pathogenicity island, and show genome rearrangements following the expansion of IS*1667*. These phylogroups also show general loss of metabolic flexibility, and in the case of PG 6 an extreme loss of metabolic capability.

Comparative analysis of complete genome sequence data sets of all representative yersiniae species coupled with the eco-evo perspective of bacteria has provided a greater understanding of the key evolutionary steps that led to the emergence of successful mammalian pathogens. We need to move away from our biased human-centric view of bacteria and consider more the environmental context, habitat and niches of these organisms**.**

Finally, the *Yersinia* genus data shows that a genus-wide sequencing approach leads to new information on the distribution of virulence-associated genes and on which of these genes are genuine virulence factors that are appropriate for novel diagnostic and pathogen-targeting research. However, even with the large number of *Yersinia* spp. isolates sequenced to date, the origin and reservoir of

mobile genetic elements, such as pYV, remain obscure. It may be that we have only sequenced the tip of the bacterial iceberg and that future large genome and microbiome studies coupled with high throughput phenotypic methods will be even more revealing. This will provide a web of bacterial life, the inter-relatedness of bacteria, no doubt blurring species boundaries. The genome data together with the eco-evo perspective will form the basis of how and why bacteria evolve to be more pathogenic with the potential to avert future threats posed by old adversaries such as the emergence of antimicrobial resistance. .

References

1. Fuchs, T., Bresolin, G., Marcinowski, L., Schachtner, J. & Scherer, S. Insecticidal genes of Yersinia spp.: taxonomical distribution, contribution to toxicity towards Manduca sexta and Galleria mellonella, and evolution. *BMC Microbiol.* **8,** 214 (2008).

2. Lithgow, K. V *et al.* A general protein O-glycosylation system within the Burkholderia cepacia complex is involved in motility and virulence. *Mol. Microbiol.* **92,** 116–137 (2014).

3. Pukklay, P. *et al.* Involvement of EnvZ-OmpR two-component system in virulence control of Escherichia coli in Drosophila melanogaster. *Biochem. Biophys. Res. Commun.* **438,** 306–311 (2013).

4. Kumar, P. & Virdi, J. S. Identification and distribution of putative virulence genes in clinical strains of Yersinia enterocolitica biovar 1A by suppression subtractive hybridization. *J. Appl. Microbiol.* **113,** 1263–1272 (2012).

5. Korea, C.-G., Badouraly, R., Prevost, M.-C., Ghigo, J.-M. & Beloin, C. Escherichia coli K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. *Environ. Microbiol.* **12,** 1957–1977 (2010).

6. Sun, Y.-C., Jarrett, C. O., Bosio, C. F. & Hinnebusch, B. J. Retracing the evolutionary path that led to flea-borne transmission of Yersinia pestis. *Cell Host Microbe* **15,** 578–586 (2014).

7. Viana, D. *et al.* A single natural nucleotide mutation alters bacterial pathogen host tropism. *Nat. Genet.* **47,** 361–366 (2015).

8. Reuter, S. *et al.* Parallel independent evolution of pathogenicity within the genus Yersinia. *Proc Natl Acad Sci U S A* **111,** 6768–6773 (2014).

9. Thomson, N. R. *et al.* The Complete Genome Sequence and Comparative Genome Analysis of the High Pathogenicity Yersinia enterocolitica Strain 8081. *PLoS Genet.* **2,** e206 (2006).

10. Chain, P. S. G. *et al.* Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. *PNAS* **101,** 13826–13831 (2004).

11. Isberg, R. R., Voorhis, D. L. & Falkow, S. Identification of invasin: A protein that allows enteric bacteria to penetrate cultured mammalian cells. *Cell* **50,** 769–778 (1987).

12. Portnoy, D. A. & Falkow, S. Virulence-associated plasmids from Yersinia enterocolitica and Yersinia pestis. *J. Bacteriol.* **148,** 877–883 (1981).

13. Portnoy, D. A., Moseley, S. L. & Falkow, S. Characterization of plasmids and plasmid-associated determinants of Yersinia enterocolitica pathogenesis. *Infect. Immun.* **31,** 775–782 (1981).

14. Rosqvist, R., Hakansson, S., Forsberg, A. & Wolf-Watz, H. Functional conservation of the secretion and translocation machinery for virulence proteins of yersiniae, salmonellae and shigellae. *EMBO J.* **14,** 4187–4195 (1995).

15. Yao, T., Mecsas, J., Healy, J. I., Falkow, S. & Chien, Y. Suppression of T and B lymphocyte activation by a Yersinia pseudotuberculosis virulence factor, yopH. *J. Exp. Med.* **190,** 1343–1350 (1999).

16. Monack, D. M., Mecsas, J., Bouley, D. & Falkow, S. Yersinia-induced apoptosis in vivo aids in the establishment of a systemic infection of mice. *J. Exp. Med.* **188,** 2127–2137 (1998).

17. Boland, A. & Cornelis, G. R. Role of YopP in suppression of tumor necrosis factor alpha release by macrophages during Yersinia infection. *Infect Immun* **66,** 1878–1884 (1998).

18. Carniel, E. Evolution of pathogenic Yersinia, some lights in the dark. *Adv. Exp. Med. Biol.* **529,** 3–12 (2003).

19. Sprague, L. D. & Neubauer, H. Yersinia aleksiciae sp. nov. *Int. J. Syst. Evol. Microbiol.* **55,** 831–835 (2005).

20. Sprague, L. D., Scholz, H. C., Amann, S., Busse, H.-J. & Neubauer, H. Yersinia similis sp. nov. *Int. J. Syst. Evol. Microbiol.* **58,** 952–958 (2008).

21. Sulakvelidze, A. Yersiniae other than Y. enterocolitica, Y. pseudotuberculosis, and Y. pestis: the ignored species. *Microbes Infect.* **2,** 497–513 (2000).

22. Murros-Kontiainen, A. *et al.* Yersinia nurmii sp. nov. *Int J Syst Evol Micro* **61,** 2368–2372 (2011).

23. Murros-Kontiainen, A. *et al.* Yersinia pekkanenii sp. nov. *Int J Syst Evol Microbiol* **61,** 2363–2367 (2011).

24. Hurst, M. R. H., Becher, S. A., Young, S. D., Nelson, T. L. & Glare, T. R. Yersinia entomophaga sp. nov., isolated from the New Zealand grass grub Costelytra zealandica. *Int. J. Syst. Evol. Microbiol.* **61,** 844–849 (2011).

25. Merhej, V., Adekambi, T., Pagnier, I., Raoult, D. & Drancourt, M. Yersinia massiliensis sp. nov., isolated from fresh water. *Int. J. Syst. Evol. Microbiol.* **58,** 779–784 (2008).

26. Savin, C. *et al.* The Yersinia pseudotuberculosis complex: characterization and delineation of a new species, Yersinia wautersii. *Int. J. Med. Microbiol.* **304,** 452–463 (2014).

27. Wren, B. W. The Yersiniae - a model genus to study the rapid evolution of bacterial pathogens. *Nat Rev Micro* **1,** 55–64 (2003).

28. Achtman, M. *et al.* Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. *Proc Natl Acad Sci U S A* **96,** 14043–14048 (1999).

29. Achtman, M. *et al.* Microevolution and history of the plague bacillus, Yersinia pestis. *PNAS* **101,** 17837–17842 (2004).

30. Bos, K. I. *et al.* Yersinia pestis: new evidence for an old infection. *PLoS One* **7,** e49803 (2012).

31. Chain, P. S. *et al.* Complete genome sequence of Yersinia pestis strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *J Bacteriol* **188,** 4453–4463 (2006).

32. Parkhill, J. *et al.* Genome sequence of Yersinia pestis, the causative agent of plague. *Nature* **413,** 523–527 (2001).

33. Deng, W. *et al.* Genome Sequence of Yersinia pestis KIM. *J. Bacteriol.* **184,** 4601–4611 (2002).

34. Savin, C. *et al.* Draft Genome Sequence of a Clinical Strain of Yersinia enterocolitica (IP10393) of Bioserotype 4/O:3 from France. *Genome Announc.* **1,** (2013).

35. Batzilla, J., Hoper, D., Antonenka, U., Heesemann, J. & Rakin, A. Complete genome sequence of Yersinia enterocolitica subsp. palearctica serogroup O:3. *J. Bacteriol.* **193,** 2067 (2011).

36. Fuchs, T. M., Brandt, K., Starke, M. & Rattei, T. Shotgun sequencing of Yersinia enterocolitica strain W22703 (biotype 2, serotype O:9): genomic evidence for oscillation between invertebrates and mammals. *BMC Genomics* **12,** 168 (2011).

37. Wang X Li Y, J. H. R. Y. Z. Z. W. S. K. B. X. J. W. L. Complete genome sequence of a Yersinia enterocolitica 'Old World' (3/O:9) strain and comparison with the 'New World' (1B/O:8) strain. *J. Clin. Microbiol.* **49,** 1251 (2011).

38. Chen, P. *et al.* Genomic characterization of the Yersinia genus. *Genome Biol.* **11,** R1 (2010).

39. Perna, N. T. *et al.* Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. *Nature* **409,** 529–533 (2001).

40. Brussow, H., Canchaya, C. & Hardt, W.-D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* **68,** 560–602, table of contents (2004).

41. Portnoy, D. A., Wolf-Watz, H., Bolin, I., Beeder, A. B. & Falkow, S. Characterization of common virulence plasmids in Yersinia species and their role in the expression of outer membrane proteins. *Infect. Immun.* **43,** 108–114 (1984).

42. Cornelis, G. R. & Wolf-Watz, H. The Yersinia Yop virulon: a bacterial system for subverting eukaryotic cells. *Mol. Microbiol.* **23,** 861–867 (1997).

43. Grosdent, N., Maridonneau-Parini, I., Sory, M.-P. & Cornelis, G. R. Role of Yops and Adhesins in Resistance of Yersinia enterocolitica to Phagocytosis. *Infect. Immun.* **70,** 4165–4176 (2002).

44. Cornelis, G. R. & Wolf-Watz, H. The Yersinia Yop virulon: a bacterial system for subverting eukaryotic cells. *Mol. Microbiol.* **23,** 861–867 (1997).

45. Bohn, E. *et al.* Gene expression patterns of epithelial cells modulated by pathogenicity factors of Yersinia enterocolitica. *Cell. Microbiol.* **6,** 129–141 (2004).

46. Miller, V. L. & Falkow, S. Evidence for two genetic loci in Yersinia enterocolitica that can promote invasion of epithelial cells. *Infect. Immun.* **56,** 1242–1248 (1988).

47. Pierson, D. E. & Falkow, S. The ail gene of Yersinia enterocolitica has a role in the ability of the organism to survive serum killing. *Infect. Immun.* **61,** 1846–1852 (1993).

48. Biedzka-Sarek, M., Venho, R. & Skurnik, M. Role of YadA, Ail, and Lipopolysaccharide in Serum Resistance of Yersinia enterocolitica Serotype O:3. *Infect. Immun.* **73,** 2232–2244 (2005).

49. Carniel, E. Plasmids and pathogenicity islands of Yersinia. *Curr. Top. Microbiol. Immunol.* **264,** 89–108 (2002).

50. Pepe, J. C. & Miller, V. L. Yersinia enterocolitica invasin: A primary role in the initiation of infection. *PNAS* **90,** 6473–6477 (1993).

51. Schubert, S., Fischer, D. & Heesemann, J. Ferric enterochelin transport in Yersinia enterocolitica: molecular and evolutionary aspects. *J. Bacteriol.* **181,** 6387–6395 (1999).

52. Tomich, M., Planet, P. J. & Figurski, D. H. The tad locus: postcards from the widespread colonization island. *Nat. Rev. Microbiol.* **5,** 363–375 (2007).

53. Iriarte, M. *et al.* The Myf fibrillae of Yersinia enterocolitica. *Mol. Microbiol.* **9,** 507–520 (1993).

54. Delor, I. & Cornelis, G. R. Role of Yersinia enterocolitica Yst toxin in experimental infection of young rabbits. *Infect Immun* **60,** 4269–4277 (1992).

55. Zimbler, D. L., Schroeder, J. A., Eddy, J. L. & Lathem, W. W. Early emergence of Yersinia pestis as a severe respiratory pathogen. *Nat. Commun.* **6,** 7487 (2015).

56. Lesic, B. *et al.* Excision of the high-pathogenicity island of Yersinia pseudotuberculosis requires the combined actions of its cognate integrase and Hef, a new recombination directionality factor. *Mol. Microbiol.* **52,** 1337–1348 (2004).

57. Schubert, S., Picard, B., Gouriou, S., Heesemann, J. & Denamur, E. Yersinia high-pathogenicity island contributes to virulence in Escherichia coli causing extraintestinal infections. *Infect Immun* **70,** 5335–5337 (2002).

58. Carniel, E., Guilvout, I. & Prentice, M. Characterization of a large chromosomal 'high-pathogenicity island' in biotype 1B Yersinia enterocolitica. *J. Bacteriol.* **178,** 6743–6751 (1996).

59. Haller, J. C., Carlson, S., Pederson, K. J. & Pierson, D. E. A chromosomally encoded type III secretion pathway in Yersinia enterocolitica is important in virulence. *Mol. Microbiol.* **36,** 1436–1446 (2000).

60. Matsumoto, H. & Young, G. M. Proteomic and functional analysis of the suite of Ysp proteins exported by the Ysa type III secretion system of Yersinia enterocolitica Biovar 1B. *Mol. Microbiol.* **59,** 689–706 (2006).

61. Young, B. M. & Young, G. M. Evidence for Targeting of Yop Effectors by the Chromosomally Encoded Ysa Type III Secretion System of Yersinia enterocolitica. *J. Bacteriol.* **184,** 5563–5571 (2002).

62. Walker, K. A., Maltez, V. I., Hall, J. D., Vitko, N. P. & Miller, V. L. A phenotype at last: essential role for the Yersinia enterocolitica Ysa type III secretion system in a Drosophila melanogaster S2 cell model. *Infect. Immun.* **81,** 2478–2487 (2013).

63. Perry, R. D., Pendrak, M. L. & Schuetze, P. Identification and cloning of a hemin storage locus involved in the pigmentation phenotype of Yersinia pestis. *J. Bacteriol.* **172,** 5929–5937 (1990).

64. Hinnebusch, B. J., Perry, R. D. & Schwan, T. G. Role of the Yersinia pestis hemin storage (hms) locus in the transmission of plague by fleas. *Science* **273,** 367–370 (1996).

65. Schroder, O. & Wagner, R. The bacterial regulatory protein H-NS--a versatile modulator of nucleic acid structures. *Biol. Chem.* **383,** 945–960 (2002).

66. Cathelyn, J. S., Ellison, D. W., Hinchliffe, S. J., Wren, B. W. & Miller, V. L. The RovA regulons of Yersinia enterocolitica and Yersinia pestis are distinct: evidence that many RovA-regulated genes were acquired more recently than the core genome. *Mol. Microbiol.* **66,** 189–205 (2007).

67. Ellison, D. W., Lawrenz, M. B. & Miller, V. L. Invasin and beyond: regulation of Yersinia virulence by RovA. *Trends Microbiol.* **12,** 296–300 (2004).

68. Bohme, K. *et al.* Concerted actions of a thermo-labile regulator and a unique intergenic RNA thermosensor control Yersinia virulence. *PLoS Pathog.* **8,** e1002518 (2012).

69. Quade, N. *et al.* Structural basis for intrinsic thermosensing by the master virulence regulator RovA of Yersinia. *J. Biol. Chem.* **287,** 35796–35803 (2012).

70. Kapatral, V. *et al.* Gene array analysis of Yersinia enterocolitica FlhD and FlhC: regulation of enzymes affecting synthesis and degradation of carbamoylphosphate. *Microbiology* **150,** 2289–2300 (2004).

71. Perry, R. D. *et al.* Temperature regulation of the hemin storage (Hms+) phenotype of Yersinia pestis is posttranscriptional. *J. Bacteriol.* **186,** 1638–1647 (2004).

72. Fälker, S., Schmidt, M. A. & Heusipp, G. Altered Ca2+ Regulation of Yop Secretion in Yersinia enterocolitica after DNA Adenine Methyltransferase Overproduction Is Mediated by Clp-Dependent Degradation of LcrG. *J. Bacteriol.* **188,** 7072–7081 (2006).

73. Grabenstein, J. P., Marceau, M., Pujol, C., Simonet, M. & Bliska, J. B. The Response Regulator PhoP of Yersinia pseudotuberculosis Is Important for Replication in Macrophages and for Virulence. *Infect. Immun.* **72,** 4973–4984 (2004).

74. Fookes, M. *et al.* Salmonella bongori provides insights into the evolution of the Salmonellae. *PLoS Pathog.* **7,** e1002191 (2011).

75. Navarre, W. W. *et al.* Co-regulation of Salmonella enterica genes required for virulence and resistance to antimicrobial peptides by SlyA and PhoP/PhoQ. *Mol. Microbiol.* **56,** 492–508 (2005).

76. Ali, S. S. *et al.* Silencing by H-NS potentiated the evolution of Salmonella. *PLoS Pathog.* **10,** e1004500 (2014).

77. Pallen, M. J. & Wren, B. W. Bacterial pathogenomics. *Nature* **449,** 835–842 (2007).

78. Morelli, G. *et al.* Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* (2010). at <http://dx.doi.org/10.1038/ng.705;>

79. Song, Y. *et al.* Complete genome sequence of Yersinia pestis strain 91001, an isolate avirulent to humans. *DNA Res* **11,** 179–197 (2004).

80. Bos, K. I. *et al.* A draft genome of Yersinia pestis from victims of the Black Death. *Nature* **478,** 506–510 (2011).

81. Chouikha, I. & Hinnebusch, B. J. Silencing urease: a key evolutionary step that facilitated the adaptation of Yersinia pestis to the flea-borne transmission route. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 18709–18714 (2014).

82. HH, A. S. W. Yersinia enterocolitica serovar 2a, wb, 3:b,c biovar 5 in hares and sheep. *Berl Munch Tierarztl Wochenschr* **110,** 176 (1997).

83. Swaminathan, B., Harmon, M. C. & Mehlman, I. J. Yersinia enterocolitica. *J. Appl. Bacteriol.* **52,** 151–183 (1982).

84. Shea, J. E., Hensel, M., Gleeson, C. & Holden, D. W. Identification of a virulence locus encoding a second type III secretion system in Salmonella typhimurium. *PNAS* **93,** 2593–2597 (1996).

85. Schaake, J. *et al.* Essential role of invasin for colonization and persistence of Yersinia enterocolitica in its natural reservoir host, the pig. *Infect. Immun.* **82,** 960–969 (2014).

86. Schaake, J. *et al.* Human and animal isolates of Yersinia enterocolitica show significant serotype-specific colonization and host-specific immune defense properties. *Infect. Immun.* **81,** 4013–4025 (2013).

87. Uliczka, F. *et al.* Unique cell adhesion and invasion properties of Yersinia enterocolitica O:3, the most frequent cause of human Yersiniosis. *PLoS Pathog.* **7,** e1002117 (2011).

88. Valentin-Weigand, P. Heesemann, J. Dersch, P. Unique virulence properties of Yersinia enterocolitica O:3 - An emerging zoonotic pathogen using pigs as preferred reservoir host. *Int. J. Med. Microbiol.* **14,** 00090–3 (2014).

89. Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. *Nature* **413,** 848–852 (2001).

90. Thomson, N. R. *et al.* Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res.* **18,** 1624–1637 (2008).

91. McClelland, M. *et al.* Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature* **413,** 852–856 (2001).

92. Nuccio, S.-P. & Baumler, A. J. Comparative analysis of Salmonella genomes identifies a metabolic network for escalating growth in the inflamed gut. *MBio* **5,** e00929–14 (2014).

93. Badia, J., Ros, J. & Aguilar, J. Fermentation mechanism of fucose and rhamnose in Salmonella typhimurium and Klebsiella pneumoniae. *J. Bacteriol.* **161,** 435–437 (1985).

94. Bry, L., Falk, P. G., Midtvedt, T. & Gordon, J. I. A model of host-microbial interactions in an open mammalian ecosystem. *Science* **273,** 1380–1383 (1996).

95. Price-Carter, M., Tingey, J., Bobik, T. A. & Roth, J. R. The alternative electron acceptor tetrathionate supports B12-dependent anaerobic growth of Salmonella enterica serovar typhimurium on ethanolamine or 1,2-propanediol. *J. Bacteriol.* **183,** 2463–2475 (2001).

96. Winter, S. E. *et al.* Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature* **467,** 426–429 (2010).

97. Thiennimitr, P. *et al.* Intestinal inflammation allows Salmonella to use ethanolamine to compete with the microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 17480–17485 (2011).

98. Langridge, G. C. *et al.* Patterns of genome evolution that have accompanied host adaptation in Salmonella. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 863–868 (2015).

99. Prentice, M. B. *et al.* Cobalamin synthesis in Yersinia enterocolitica 8081. Functional aspects of a putative metabolic island. *Adv. Exp. Med. Biol.* **529,** 43–46 (2003).

100. Rabson, A. R., Hallett, A. F. & Koornhof, H. J. Generalized Yersinia enterocolitica infection. *J. Infect. Dis.* **131,** 447–451 (1975).

101. Lian, C. J., Hwang, W. S., Kelly, J. K. & Pai, C. H. Invasiveness of Yersinia enterocolitica lacking the virulence plasmid: an in-vivo study. *J. Med. Microbiol.* **24,** 219–226 (1987).

102. Buret, A., O'Loughlin, E. V, Curtis, G. H. & Gall, D. G. Effect of acute Yersinia enterocolitica infection on small intestinal ultrastructure. *Gastroenterology* **98,** 1401–1407 (1990).

103. Tardy, F. *et al.* Yersinia enterocolitica type III secretion-translocation system: channel formation by secreted Yops. *Embo J.* **18,** 6793–6799 (1999).

104. Wagner, D. M. *et al.* Yersinia pestis and the plague of Justinian 541-543 AD: a genomic analysis. *Lancet. Infect. Dis.* **14,** 319–326 (2014).

105. Hall, M. *et al.* Use of whole-genus genome sequence data to develop a multilocus sequence typing tool that accurately identifies *Yersinia* isolates to the species and subspecies levels. *J. Clin. Microbiol.* **53,** 35–42 (2015).

106. Bottone, E. J. Yersinia enterocolitica: overview and epidemiologic correlates. *Microbes Infect.* **1,** 323–333 (1999).

107. Bottone, E. J. Yersinia enterocolitica: the charisma continues. *Clin. Microbiol. Rev.* **10,** 257–276 (1997).

108. Revell, P. A. & Miller, V. L. Yersinia virulence: more than a plasmid. *FEMS Microbiol. Lett.* **205,** 159–164 (2001).

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

Traditionally, bacteria are speciated using limited biochemical tests that are based on the phenotypic expression or non-expression of a trait. This can lead to problems in standardising methods, and the potential to mutate the trait can impact on the phenotype. Genomic information is less ambiguous, and variations within genes under neutral selection can be used to infer evolutionary relationships, providing a robust framework that is independent of phenotypic variation. For the genus *Yersinia*, 84 housekeeping genes were identified, that showed between 10-25% SNP divergence between *Y. pestis* and *Y. enterocolitica*[8]. These criteria meet those of housekeeping genes used for the design of multi-locus sequence typing (MLST) schemes[105], so that the accumulation of variation reflects the history of the genus, and a maximum likelihood phylogeny can be estimated. Furthermore, Bayesian analysis of the population structure was employed to identify defined species clusters (SC[8]). This means that statistical probabilities are calculated to describe the variation hidden within the population, and does not only consider the sequence of the 84 genes as a whole but as separate gene entities. Some of the species clusters thus described contain multiple species as defined by classical techniques, such as SCs 13 and 14. On the other hand, other species like *Y. frederiksenii*, that show heterogeneity in their phenotype, are distributed over several SCs (SCs 8, 9, and 14). This highlights the limitations of biochemical tests that fail to accurately reflect the relationships between the different species. The analysis also emphasizes the position of the pathogenic species (highlighted in blue and green). The human pathogenic species (blue) are positioned at diametrically opposite ends of the tree, with the fish pathogen *Y. ruckeri* (green) forming a third outlying SC. Even though no

classical root of the tree is given, *Y. ruckeri* might be considered an out group since its position within the genus has previously been under debate. The position of *Y. pseudotuberculosis—Y. pestis* and *Y. enterocolitica* nevertheless supports independent evolution of their pathogenic potential.

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

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

The three human-pathogenic species of *Yersinia* share a key genetic component in their ability to cause disease, a large virulence plasmid named pYV (plasmid of *Yersinia* Virulence). Though commonly thought of as a single entity, pYV is actually highly variable between species and strains of species, containing different origin of replications and showing variable genetic architecture[8,12]. What is common amongst them all is a large genetic locus encoding for the Ysc type III secretion system (T3SS), responsible for the targeted delivery of Yop effector proteins into host cells. The Yops are the key virulence factor in all three of the pathogenic species. Upon contact with host macrophages Yops are injected into the host cells, resulting in silencing of the pro-inflammatory cytokine response and apoptotic death of the infected macrophages[42].

Y. enterocolitica/Y. pseudotuberculosis – are zoonotic infectious agents which cause inflammatory intestinal disease associated with ingestion via consumption of undercooked or contaminated poultry products and vegetables[106]. The bacteria enter the small intestine where they attach to the intestinal epithelium.

Numerous virulence factors have been implicated in this including the Myf, YadA, and Ail adhesins as well as flagella[18]. The invasin protein then triggers internalisation of the bacteria, which translocate across the epithelium[50] where they encounter macrophages, which they silence and hitch-hike upon to be carried to the lymph nodes where a localised infection occurs. As a result, this infection is often diagnosed as pseudo-appendicitis, and can lead to septicaemia[107].

Y. pestis – unlike *Y. enterocolitica* and *Y. pseudotuberculosis*, does not have an intestinal phase to its infection-cycle. *Y. pestis* host reservoir is rodent fleas, where it forms a biofilm in their foregut[64]. When the flea next takes a blood meal the biofilm is regurgitated directly into the bloodstream of the host prey. As a result the bacteria immediately encounter phagocytic immune cells triggering activation of the Ysc system and host-cell silencing. The bacteria then travel to the lymph nodes where an acute infection is established leading to the formation of painful swellings at the lymph nodes, or buboes. This is known as bubonic plague. If untreated the infection will spill back into the bloodstream with bacteria travelling to the lungs, where pneumonic plague is established. *Y. pestis* has acquired additional plasmids, pPLA and pFRA which play pivotal roles in the insect-infection and systemic-infection phase of the organism[108].

Figure 1: Diagrammatic comparison of HN-S transcriptional silencing of key acquired virulence genes in *Yersinia* and *Salmonella*. Both bacteria contain key T3SS involved in mammalian pathogenesis that are on acquired mobile genetic elements. In *Yersinia* this is the Ysc T3SS found on the pYV virulence plasmid. In

Salmonella this is the SPI-1 T3SS. In both bacteria HN-S acts as a transcriptional repressor under non-mammalian environmental conditions. In *Yersinia* the proxy environmental signal for being within a mammalian host is temperatures at or above 37^{0}C, whereby RovA and YmoA are activated and de-repress HN-S silencing by stearic hindrance of HN-S binding sites at promoter regions. In *Salmonella* the proxy signal is elevated Mg^{2+} concentration signalling the endocytic vacuole environment. This triggers expression of HilA and SlyA that act to de-repress HN-S and trigger expression of SPI-1.

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

Figure 3: Gene gain and loss in the *Y. enterocolitica* species complex. The diagram shows how the species is organised into phylogenetically distinct phylogroups,

which parallel LPS structure and serotype. PG 1 is most similar to the ancestral *Y. enterocolitica*. The emergence of pathogenicity in the species is marked by the acquisition of pYV, and in PG 2 the Ysa T3SS. The YGT T3SS is lost from all the pathogenic phylogroups. The other significant event is the acquisition of serotype specific LPS operons in each of the pathogenic phylogroups. PG 3 has become the dominant isolate from pig reservoirs and human disease cases, and has a deletion in the Flag-2 secondary flagella genes as well as an insertion in the invasin promoter, which increases transcription of this key virulence factor.

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

Figure 5: A diagrammatic overview of the evolution of pathogenesis in the *Yersinia* genus, highlighting key gene gain and gene loss events in the formation of each of the pathogenic lineages. Acquisition of pYV occurs independently on

two occasions within *Y. enterocolitica* PG 2 and PGs 3-6, as well as additionally in a *Y. pseudotuberculosis* ancestor. These two human-pathogenic lineages have emerged on independent occasions from environmental generalists following similar themes of gene gain, gene loss, metabolic reduction, and genome rearrangements.

Glossary

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

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

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

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

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

Genomic islands - Large genetic regions that are part of the accessory gene pool. They form the horizontally acquired part of a genome encoding one of more functional groups of genes. They are frequently associated with tRNA genes and are flanked by repeat structures, and contain mobility genes coding for integrases or transposases required for chromosomal integration and excision.

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

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

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

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

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

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

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

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

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

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

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

Salmonids - Relating to salmon and trout fish.

Sideophore - Compound enabling highly efficient capture of exogenous iron.

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

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

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

Verotoxin – Shiga-like toxin produced by verotoxigenic *E. coli*. AB_5 toxin that targets cells in the renal glomeruli leading to kidney damage.

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

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