Parallel independent evolution of pathogenicity within the genus *Yersinia*

Sandra Reuter[a,b,1], Thomas R. Connor[b,c,1], Lars Barquist[b], Danielle Walker[b], Theresa Feltwell[b], Simon R. Harris[b], Maria Fookes[b], Miquette E. Hall[a], Nicola K. Petty[b,d], Thilo M. Fuchs[e], Jukka Corander[f], Muriel Dufour[g], Tamara Ringwood[h], Cyril Savin[i], Christiane Bouchier[j], Liliane Martin[i], Minna Miettinen[f], Mikhail Shubin[f], Julia M. Riehm[k], Riikka Laukkanen-Ninios[l], Leila M. Sihvonen[m], Anja Siitonen[m], Mikael Skurnik[n], Juliana Pfrimer Falcão[o], Hiroshi Fukushima[p], Holger C. Scholz[k], Michael B. Prentice[h], Brendan W. Wren[q], Julian Parkhill[b], Elisabeth Carniel[i], Mark Achtman[r,s], Alan McNally[a], and Nicholas R. Thomson[b,q,2]

[a]Pathogen Research Group, Nottingham Trent University, Nottingham NG11 8NS, United Kingdom; [b]Pathogen Genomics, Wellcome Trust Sanger Institute, Cambridge CB10 1SA, United Kingdom; [c]Cardiff University School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales, United Kingdom; [d]The ithree institute, University of Technology, Sydney, NSW 2007, Australia; [e]Zentralinstitut für Ernährungs- und Lebensmittelforschung, Technische Universität München, D-85350 Freising, Germany; [f]Department of Mathematics and Statistics, and [l]Department of Food Hygiene and Environmental Health, Faculty of Veterinary Medicine, University of Helsinki, FIN-00014 Helsinki, Finland; [g]Institute of Environmental Science and Research, Wallaceville, Upper Hutt 5140, New Zealand; [h]Department of Microbiology and [r]Environmental Research Institute, University College Cork, Cork, Ireland; [i]*Yersinia* Research Unit, [j]Genomics Platform, Institut Pasteur, 75724 Paris, France; [k]Department of Bacteriology, Bundeswehr Institute of Microbiology, D-80937 Munich, Germany; [m]Bacteriology Unit, National Institute for Health and Welfare (THL), FIN-00271 Helsinki, Finland; [n]Department of Bacteriology and Immunology, Haartman Institute, University of Helsinki and Helsinki University Central Hospital Laboratory Diagnostics, FIN-00014 Helsinki, Finland; [o]Brazilian Reference Center on *Yersinia* spp. other than *Y. pestis*, Faculdade de Ciências Farmacêuticas de Ribeirão Preto–Universidade São Paulo, Ribeirão Preto, CEP 14040-903, São Paulo, Brazil; [p]Shimane Prefectural Institute of Public Health and Environmental Science, Matsue, Shimane 699-0122, Japan; [q]Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom; and [s]Warwick Medical School, University of Warwick, Coventry CV4 7AL, United Kingdom

Edited by Ralph R. Isberg, Howard Hughes Medical Institute and Tufts University School of Medicine, Boston, MA, and approved March 21, 2014 (received for review October 2, 2013)

The genus *Yersinia* has been used as a model system to study pathogen evolution. Using whole-genome sequencing of all *Yersinia* species, we delineate the gene complement of the whole genus and define patterns of virulence evolution. Multiple distinct ecological specializations appear to have split pathogenic strains from environmental, nonpathogenic lineages. This split demonstrates that contrary to hypotheses that all pathogenic *Yersinia* species share a recent common pathogenic ancestor, they have evolved independently but followed parallel evolutionary paths in acquiring the same virulence determinants as well as becoming progressively more limited metabolically. Shared virulence determinants are limited to the virulence plasmid pYV and the attachment invasion locus *ail*. These acquisitions, together with genomic variations in metabolic pathways, have resulted in the parallel emergence of related pathogens displaying an increasingly specialized lifestyle with a spectrum of virulence potential, an emerging theme in the evolution of other important human pathogens.

genomics metabolic streamlining | pathoadaptation | *Enterobacteriaceae*

Bacterial species are defined on the basis of phenotypic characteristics, such as cellular morphology and biochemical characteristics, as well as DNA-DNA hybridization and 16S rRNA comparison. Using high-throughput whole-genome approaches we can now move beyond classic methods and develop population frameworks to reconstruct accurate inter- and intraspecies relationships and gain insights into the complex patterns of gene flux that define different taxonomic groups.

Bacterial whole-genome sequencing has revealed enormous heterogeneity in gene content, even between members of the same species. From a bacterial perspective the acquisition of new genes provides the flexibility to adapt and exploit novel niches and opportunities. From a human perspective, integration of genes by bacteria has been directly linked to the emergence of new pathogenic clones, often from formerly harmless lineages (1, 2). In addition to gene gain, gene loss is also strongly associated with host restriction in acutely pathogenic bacterial species, such as *Yersinia pestis* and *Salmonella enterica* serovars, including *Salmonella* Typhi (3–5), where gene loss can remove functions unnecessary in the new niche (6). These specialist pathogens show a much higher frequency of functional gene loss than closely related host generalist pathogens, such as *Yersinia pseudotuberculosis* (7).

Previous *Yersinia* genome studies (8, 9) have examined the evolution of pathogenicity by comparing strains from a selection of species or species subtypes within the genus, limiting our understanding of the evolutionary context of individual species. The majority of the *Yersinia* species are found in the environment and do not cause disease in mammals. Three species are known as human pathogens: the plague bacillus *Y. pestis* and the enteropathogens *Yersinia enterocolitica* and *Y. pseudotuberculosis*.

Significance

Our past understanding of pathogen evolution has been fragmented because of tendencies to study human clinical isolates. To understand the evolutionary trends of pathogenic bacteria though, we need the context of their nonpathogenic relatives. Our unique and detailed dataset allows description of the parallel evolution of two key human pathogens: the causative agents of plague and *Yersinia* diarrhea. The analysis reveals an emerging pattern where few virulence-related functions are found in all pathogenic lineages, representing key "foothold" moments that mark the emergence of these pathogens. Functional gene loss and metabolic streamlining are equally complementing the evolution of *Yersinia* across the pathogenic spectrum.

Author contributions: S.R., T.R.C., B.W.W., J.P., M.A., A.M., and N.R.T. designed research; S.R., T.R.C., L.B., D.W., T.F., S.R.H., M.F., M.E.H., N.K.P., J.C., M.M., M. Shubin, and N.R.T. performed research; T.M.F., M.D., T.R., C.S., C.B., L.M., J.M.R., R.L.-N., L.M.S., A.S., J.P.F., H.F., and H.C.S. were involved in isolate collection and typing; S.R., T.R.C., L.B., D.W., S.R.H., M.F., M.E.H., N.K.P., J.C., M.M., M. Shubin, and N.R.T. analyzed data; and S.R., T.R.C., L.B., S.R.H., T.M.F., M. Skurnik, M.B.P., B.W.W., J.P., E.C., M.A., A.M., and N.R.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the European Nucleotide Archive (ENA study nos. PRJEB2116, PRJEB2117) GenBank and SRA numbers are given in Table S1.

[1]S.R. and T.R.C. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: nrt@sanger.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1317161111/-/DCSupplemental.

Only *Y. pestis* and *Y. pseudotuberculosis* have been studied extensively in a phylogenetic context (7, 9, 10). In our study, we present a global analysis of diversity in multiple isolates representing all current species of the *Yersinia* genus to examine the evolution of bacterial pathogens in the context of the entire genus, encompassing both genomic features and metabolic signatures. It has been previously suggested that the pathogenic *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica* share a recent common ancestor to the exclusion of the nonpathogenic species (11–13). Contrary to this, we show conclusively that the human pathogenic *Yersinia* lineages have evolved independently. Early ecological separation is likely to have split the most acutely pathogenic *Yersinia* strains from the environmental species and nonpathogenic *Y. enterocolitica*. In human pathogenic lineages this was then followed by a common pattern of acquisition of similar pathogenicity determinants, namely the virulence plasmid pYV and the invasin *ail*, followed by the acquisition of further determinants within lineages along with functional gene loss and reduced metabolic flexibility. This parallel evolution in distinct pathogenic *Yersinia* species may represent a paradigm of the emergence of important human pathogens from nonpathogenic bacterial species.

Results

Defining the Structure of the Genus *Yersinia*. To investigate the evolution of pathogenicity within the genus *Yersinia* we selected a representative sample of strains composed of *Y. pseudotuberculosis* and *Y. pestis*, as well as the understudied *Y. enterocolitica*. In addition, a diverse set of 94 isolates representing all of the other 15 *Yersinia* species were also sequenced. In total, 241 *Yersinia* strains were analyzed, all of which had been typed by traditional biochemical and serological methods (Table S1).

Because high diversity within the population precluded the use of a read mapping-based approach, we used a set of core genes common to all strains to produce a genus-wide phylogeny. The resulting phylogenetic tree (Fig. 1 and Fig. S1) shows a wide diversity of clearly defined lineages within the genus, with clustering of isolates at the tips of long branches signifying very ancient common ancestry. To subdivide the population by patterns of shared sequence similarity we used the program Bayesian Analysis of Population Structure (BAPS) (14). BAPS resolved the genus into 14 species clusters (SC) (Fig. 1). Routine *Yersinia* species identification is largely based on limited biochemical data. Superimposing this information onto the phylogenetic tree revealed the lack of resolution provided by biochemical tests, and emphasized the need to use modern molecular methods in classifying bacterial species. *Yersinia frederiksenii* is known to be heterogeneous and *Yersinia massiliensis* was recently subspeciated from "atypical" *Y. frederiksenii* (15). As can be seen in Fig. 1, isolates described traditionally as *Y. frederiksenii* can be found in species complexes SC8, SC9, and SC14, with the latter also containing both *Yersinia intermedia* and *Y. massiliensis* isolates. On the other hand, the three species *Yersinia aleksiciae*, *Yersinia bercovieri*, and *Yersinia mollaretii* form a single, heterogeneous species complex SC13 (Fig. 1). In line with previous findings, the pathogenic lineages of *Y. pestis* and *Y. pseudotuberculosis* form a tight group (16), with the nonpathogenic *Yersinia similis* appearing basal to these species on the tree (17) (Fig. 1). The other human pathogenic species, *Y. enterocolitica*, forms a pair of distinct complexes in the tree, SC6 composed of the pathogenic biotype (BT) 1B and the nonpathogenic BT 1A, and SC7 composed of the other pathogenic BTs 2–5 (Fig. 1).

The remaining species occupy intermediate branching positions between the *Y. pseudotuberculosis* and *Y. enterocolitica* SCs

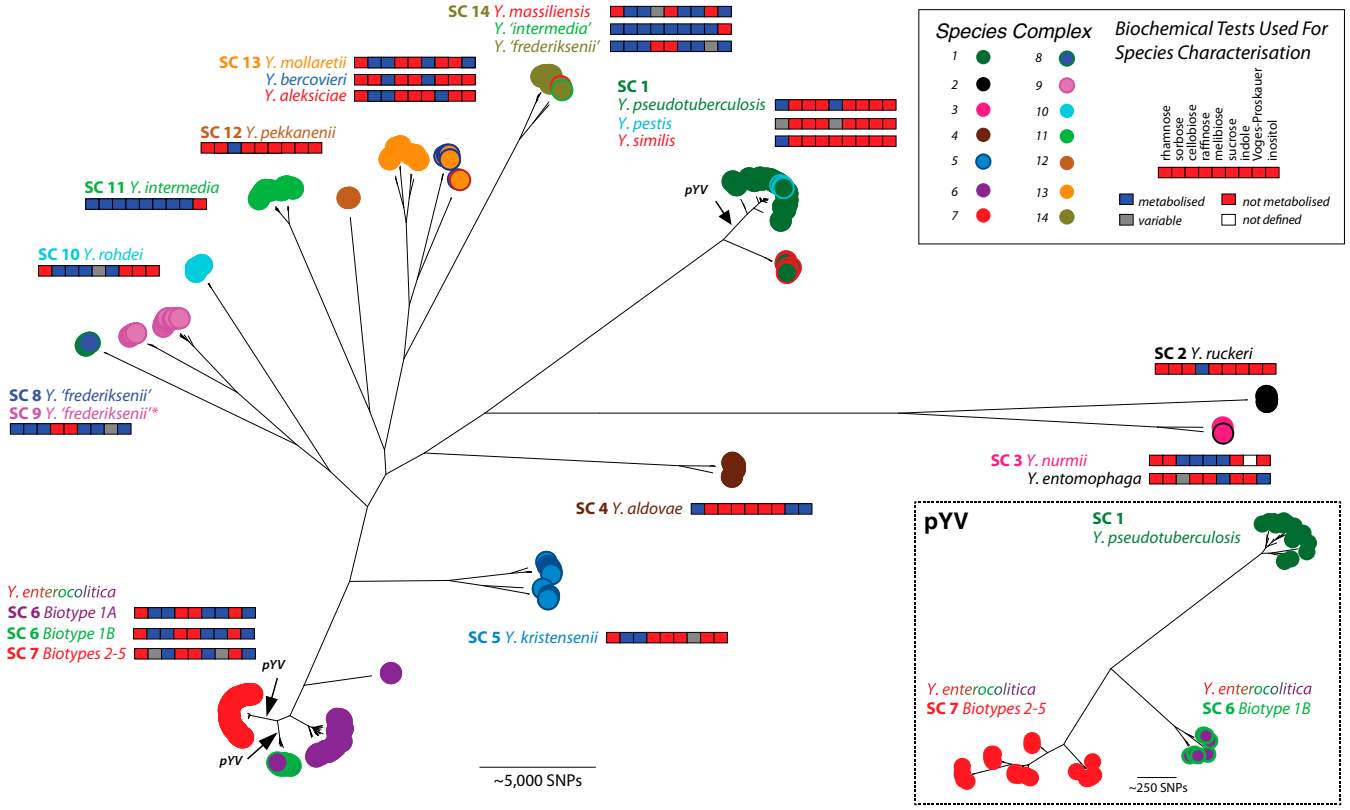

Fig. 1. The phylogeny of the genus *Yersinia* and the virulence plasmid pYV. Maximum-likelihood phylogenetic tree of the genus *Yersinia* based on the concatenated sequence of 84 housekeeping genes. Current species assignments based on biochemical typing (36) (color circle borders) are contrasted with the species complexes (colored circles) as allocated by BAPS. The SC for the *Y. frederiksenii* type strain (8) is depicted by an asterisk. Arrows show the independent acquisition events of the virulence plasmid pYV. The pYV plasmid tree for pathogenic *Y. enterocolitica* and *Y. pseudotuberculosis* samples is shown.

with the fish pathogen, *Yersinia ruckeri*, falling on a distinct divergent branch (SC2) (Fig. 1). This branch is also occupied by *Yersinia entomophaga* and *Yersinia nurmii*, two newly characterized species grouped within a single species cluster, SC3.

Plotting the Distribution of Known Virulence Functions Across the Genus. Numerous studies have defined important virulence determinants in *Y. pseudotuberculosis*/*pestis* and *Y. enterocolitica* (18–23). Analysis of these virulence genes has formed a central narrative in our understanding of the evolution of these pathogens. However, athough these determinants have been well characterized in human pathogenic lineages, their origin and distribution across other members of the genus remain unclear or incomplete.

From the phylogenetic distribution of known pathogenicity determinants (Fig. 2 and Dataset S1) it is evident that the distribution of virulence-related genes falls into three broad categories: (group I) genes represented in all lineages of the genus, (group II) genes selectively gained or lost, wholly or partially, by entire lineages, and (group III) those peculiar to single isolates.

Virulence-related genes present in all lineages (group I) include the Flag-1 flagella biosynthesis cluster and the global virulence regulator RovA (Fig. 2), which controls the expression of multiple key virulence factors (23, 24). RovA has been suggested as a promiscuous ancestral regulon into which related as well as unique functions were incorporated multiple times in different lineages, as indicated by the little overlap of regulatory networks between *Y. enterocolitica* and *Y. pseudotuberculosis* (23).

Selectively acquired or lost genes in group II include a chromosomal type 3 secretion system (T3SS) recently described in *Y. enterocolitica* (25–27). This system is most similar to the ancestral *Salmonella* pathogenicity island 2 (SPI-2) encoded T3SS and we propose to call it *Yersinia* genus T3SS (*Ygt*). T3SS have been associated with important roles in bacterial disease because they allow direct injection of T3SS effector proteins into eukaryotic cells. Although there is clear evidence for the presence of *Ygt* in all *Yersinia* lineages, it appears to be in the process of being lost in the highly pathogenic species (Fig. 2). It is also clear that the loss of *Ygt* is always coincident with the acquisition of an alternative T3SS: either chromosomal *Ysa* or the virulence plasmid, pYV, bearing the *Yop* T3SS (see below).

Group II also includes metabolic pathways including the cobalamin (vitamin B_{12}) biosynthetic operon (*cob*), 1,2-propanediol utilization pathway (*pdu*), tetrathionate respiration genes (*ttr*), hydrogenase complexes (*hyd4* and *hyd2*), and cellulose biosynthesis pathway (*cel*). Across the genus, these dispersed operons are absent from a distinct branch of the *Yersinia* phylogenetic tree encompassing SC1–3 (*Y. pestis*, *Y. pseudotuberculosis*, *Y. similis*, *Y. ruckeri*, *Y. nurmii*, and *Y. entomophaga*), except for *ttr*, which has been independently deleted in SC1 and -2, and is only maintained by SC3 (Figs. 1 and 2). The distinct distribution of these functions across the genus is likely to be evidence of early ecological specialization in *Yersinia* lineages. *Salmonella*, like *Yersinia*, synthesizes cobalamin only under anaerobiosis (28, 29) and is able to degrade 1,2-propanediol or ethanolamine by cobalamin-dependent enzymes using tetrathionate as a terminal electron acceptor, making the endogenous production of cobalamine essential for growth (30). Tetrathionate was previously thought to be important for environmental survival, but has now been shown to be produced in the vertebrate inflamed gut following the host's response to *Salmonella* Typhimurium infection, whereby the SPI-1 and -2 encoded T3SS are essential for stimulating the inflammatory response (31, 32). This widely distributed metabolic capability of *Yersinia* SCs 4–14 could provide a competitive growth advantage similar to that observed in *S.* Typhimurium over the largely fermentative gut microbiota by using naturally occurring carbon sources that are not readily fermented.

Only two acquisitions were evident on the tree as delimiters between pathogenic lineages and their most closely related "environmental" or nonpathogenic sister taxa (Fig. 2, group III, and Dataset S1). These "foothold moments" are the acquisition of the chromosomal attachment and invasion locus *ail* (19) (Fig. 2) and the plasmid pYV. Plasmid pYV encodes the *Ysc-Yop* T3SS apparatus, Yop-secreted protein effectors and chaperones (33), and is required for optimal virulence in all human pathogenic *Yersinia*. Reconstructing the phylogenies of pYV (Fig. 1) showed far fewer mutation events than would be predicted from the genus tree, but that it segregated with the host lineages. Also, unlike any other pYV lineage, pYV from BT 1B strains (SC6) possess a distinct origin of replication (34, 35). The most parsimonious explanation for the isolated presence of highly similar plasmids at extremes of the core phylogenetic tree is that recent independent acquisition of distinct versions of pYV has occurred in both the *Y. enterocolitica* lineages SC6 and -7, as well as the *Y. pseudotuberculosis*/*pestis* lineage SC1, contrary to existing hypotheses (11–13).

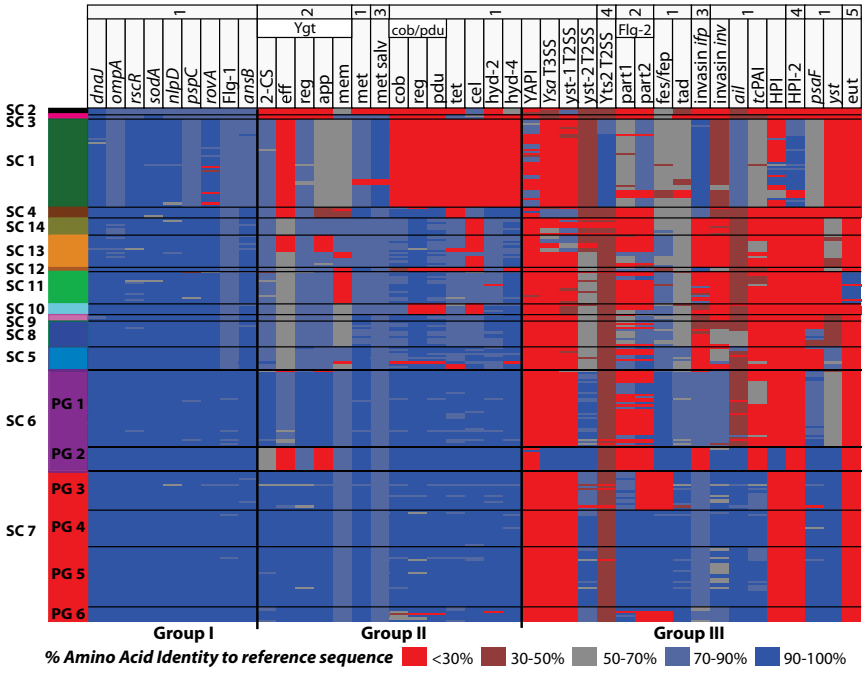

Fig. 2. Distribution of pathogenicity determinants across the genus. Three broad groups of pathogenicity determinants present in the genus are highlighted (see text and Dataset S1). Heatmap colors are based upon average amino acid identity either of single genes or as an average of the amino acid identity across the genes in the operon, as indicated. The percent identities were identified using BLAST searches of the assembled genomes. The corresponding species complexes (Fig. 1) and PGs (Fig. 3) are highlighted. Comparator sequences used: 1, *Y. enterocolitica* (YE8081); 2, *Y. enterocolitica* (YE212/02); 3, *Y. pseudotuberculosis* (IP32953); 4, *Y. pestis* (CO92); 5, *Y. intermedia* (ATCC29909). Gene names are given for pathogencity determinants, operons are labeled with their names. Abbreviations: Flg, flagella cluster; *Ygt*, *Yersinia* genus T3SS; 2-CS, two component system; eff, effectors; reg, regulator; app, apparatus; mem, membrane proteins; met (salv), methionine (salvage); YAPI, *Yersinia* adhesion pathogenicity island; T2SS, type 2 secretion system (general secretion pathway); fes/fep, siderophore operon; *tc*PAI, toxin complex pathogenicity island; HPI, high-pathogenicity island.

Evolution of Pathogenic Lineages Within *Y. enterocolitica*. Traditionally the *Y. enterocolitica* species cluster has been divided into six biotypes BT 1A, BT 1B, and BT 2–5 (36), which are further subdivided into various serotypes (Fig. 3 and Table S1). Biotyping is based on a number of biochemical tests that assesses the metabolization or acid production of a panel of metabolites (Fig. 3). The biotypes show differing pathogenic potential in the mouse model, with BTs 1A, BT 1B, and BTs 2–5 being described as non-, high-, and low-pathogenic lineages, respectively (11). Examining the phylogenetic relationships of the species using SNPs called against the reference BT 1B strain 8081 (37) revealed six distinct lineages (Fig. 3). BT 1A and 1B form discrete clusters, whereas BTs 2–5 consist of four closely related but distinct lineages, confirmed by BAPS clustering (Fig. 3). For the low-pathogenic *Y. enterocolitica* lineages, the phylogeny is largely congruent with serotype (Fig. 3). In subsequent analyses the six *Y. enterocolitica* lineages will be referred to as phylogroups (PG) 1–6.

The extent of divergence within the *Y. enterocolitica* SC is striking. PGs 3–6 display tight clusters with short terminal branches compared with PGs 1 and 2. One possible explanation for the low genetic diversity within PGs 3–6 is that they are the product of one or more recent population bottlenecks or population expansions. Although there is no discernible clustering of PGs 3–6 according to host origin, the strong correlation between the phylogenetic signal and serotype in this group may suggest that their early evolution was dominated by ecological specialization.

The Evolution of *Y. enterocolitica* Is Marked by a Reduction in Metabolic Capacity and Functional Gene Loss. It is evident from the results of the 12 classic biotyping reactions (Fig. 3) that metabolic flexibility differs between the diverse PGs 1 and 2 and the nondiverse PGs 3–6. This difference in metabolic flexibility does not appear to result from markedly different distributions of the functional genes possessed by each phylogroup representative (Fig. S2*A*), but it does correlate with the accumulation of pseudogenes and the expansion of insertion sequence elements (Fig. S2*B* and Table S2). The absolute numbers of pseudogenes varies significantly (χ^2 test P value 5.5E-59) between the different *Y. enterocolitica* lineages with the highest number present in strains in PGs 3 and 6 and the lowest in PG 1 strains. However, there is no significant difference in the class distribution between phylogroups, suggesting a general loss of functions across all classes (Fig. S2*B*). In agreement with the observed link between genetic and metabolic decay, PG 6 strains, with their high numbers of pseudogenes, have not been isolated from livestock and are reported to be host restricted to hares (38, 39).

To determine whether the signal seen from biotyping extended to the whole metabolic capacity of the biotypes, we selected representative strains from each of the six *Y. enterocolitica* phylogroups and tested their ability to metabolize a wide range of different carbon, nitrogen, sulfur, and phosphate sources on a phenotypic microarray (Biolog). The resulting signal values correspond to the intensity of utilization for different nutrient sources (Fig. S3 and Dataset S2). Variability in metabolic flexibility is evident across the 380 metabolic tests, with PG 1 strains showing the broadest metabolic capacity and the PG 6 strains exhibiting the most extreme loss of capacity. Principal component analysis (Fig. S4) using the phenotypic signals showed a clear separation between PGs 1, 2–5, and 6 along the first component, accounting for ~60% of the variance in our phenotypic data matrix. We interpret this component as capturing a gradient of metabolic restriction, from the broadly active environmental PG 1 strains to the host-restricted PG 6 strain. We believe that the lack of a specific metabolic signal together with the observation that most of the mutations in the PGs 3–6 are derived from single insertion/deletion events suggests that the genome decay seen in PGs 3–6 is the result of recent evolutionary events. This finding is also consistent with the short evolutionary distance observed in the branch leading to PGs 3–6 (Fig. 3).

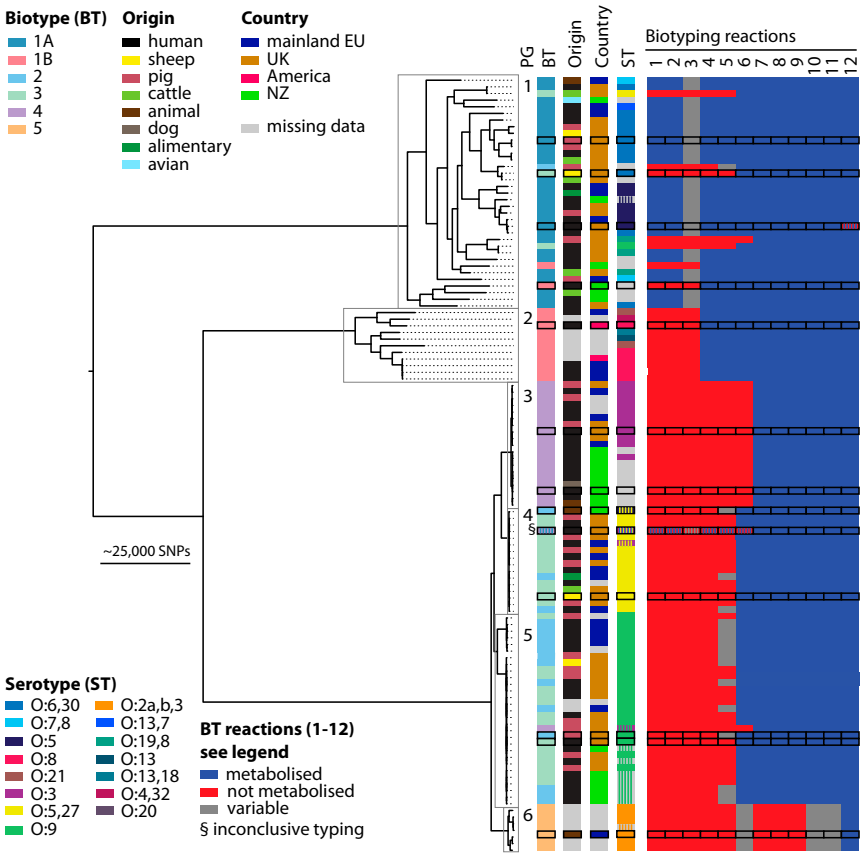

Fig. 3. The phylogeny of *Y. enterocolitica*. Maximum-likelihood phylogeny for the species based on SNPs across the whole genome excluding laterally acquired elements and phages. Lineages are characterized by biotype, biological origin, country of origin, and serotype. The biochemical tests used for biotyping (36) of *Y. enterocolitica* are presented for each strain. Shaded boxes highlight results of genomic serotype analysis differing from in vitro results. Biotyping reactions are: 1, salicin acid production; 2, pyrazinamidase activity; 3, esculin hydrolysis; 4, lipase activity; 5, indole production; 6, xylose acid production; 7, trehalose acid production; 8, sorbose acid production; 9, nitrate reduction; 10, ornithine decarboxylase; 11, Voges-Proskauer test; 12, inositol acid production.

MICROBIOLOGY

Like the phylogenetic analyses, the phenotypic results emphasize the parallelism between the two pathogenic species clusters. It has been shown that combined plasmid gain and chromosomal gene loss was a predominant driving mechanism in the evolution of the highly specialized, lifestyle-restricted clone *Y. pestis*, which lacks the metabolic repertoire of the wider *Y. pseudotuberculosis* species cluster. The reduction of metabolic flexibility through gene loss from the ubiquitous plasmid-negative PG 1 strains to the lifestyle-restricted PG 6 is particularly striking, and appears to show a similar pattern to that observed in the gradient from *Y. similis* to *Y. pseudotuberculosis* and *Y. pestis*. This appears to be a case of parallel evolution of virulence within a bacterial genus.

Discussion

The data presented here bring together our often fragmented view of bacterial pathogen evolution. Through whole-genome sequencing of over 200 genomes we have provided a robust framework to redefine the species clusters of this genus, mapping genetic traits across its full diversity and scoring the significance of pathogenicity genes based on their presence or absence. By looking at functions gained and lost from specific lineages it is clear that there are metabolic functions, such as the ability to make endogenous vitamin B_{12} or anaerobically respire tetrathionate that differentiate SC1 and -2 from all other *Yersinia* lineages (Fig. 1). The wide phylogenetic distribution of these functions suggests they are ancestral, important for growth in a range of niches, and have been lost rather than gained in specific lineages.

It has been previously subject to debate whether human pathogenic *Yersinia* shared a common pathogenic ancestor (11–13). In fact, the pathogenic lineages occupy positions at diametrically opposite ends of the *Yersinia* genus tree. From studies looking at *Y. enterocolitica* pathophysiology we know that infections are characterized by inflammation of the gut and sometimes of the mesenteric lymph nodes. It is also clear that inflammation requires pYV, as without it there is no inflammation and pYV-cured *Y. enterocolitica* are rapidly eliminated from the gut (40). In *Y. pestis*, pYV has been shown to have an early anti-inflammatory effect preceding the inflammatory response of the host. Although not demonstrated in *Y. enterocolitica*, it may be speculated that a similar early anti-inflammatory effect is necessary for *Y. enterocolitica* to establish itself at the site of infection before the onset of the host inflammatory response (41), Either way, it is clear that pYV is a key factor allowing *Y. enterocolitica* to persist in the mammalian gut lumen long enough for the host to mount an inflammatory response. In the course of such an inflammatory response tetrathionate is produced, and can be used to gain a metabolic advantage over the resident gut microbiota (32). This finding may partly explain why the acquisition of pYV represented an apparent foothold moment in the evolution of these pathogens, allowing them to exploit a new niche.

The dispensability of metabolic functions in *Y. pestis* can be explained by adoption of a lifestyle bypassing the gut infection phase, and for *Y. pseudotuberculosis* similar losses are likely to be explained by its greater invasiveness and occupation of alternative niches away from the mammalian gut. Because even the nonpathogenic members of SC 1–3, such as *Y. similis*, have lost these functions, this is evidence of an early change in niche before the evolution of the other human pathogenic *Yersinia* species in SC1. This finding may also explain why the evolution of pathogens in SC 1 is markedly different from *Y. enterocolitica* SC 6–7 in terms of gene gain. *Y. pestis* is characterized by the acquisition of an array of mostly plasmid borne virulence functions, which apart from pYV are found nowhere else in the genus.

Despite the differences in metabolic pathways there is considerable symmetry between the two branches leading to the pathogenic *Yersinia*; both have closely related nonpathogenic relatives and have independently acquired pYV. In addition, species clusters within these lineages show signs of having passed through an evolutionary bottleneck. Although this process has been well described for *Y. pestis*, it is evident from these data that much of the diversity has been similarly removed in *Y. enterocolitica* PG 3–6 populations. Large-scale functional gene loss through pseudogene formation and insertion sequence element expansion is evident in both lineages generally, but especially in the most extreme host-restricted PG 6 strains and *Y. pestis* (3). Looking outside of the *Yersiniae*, these signatures have also been shown to occur in recently emerged pathogens, such as *S.* Typhi (4), where they too are considered to be indicative of a recent evolutionary bottleneck and a change in lifestyle or niche (42). The patterns of niche adaptation and metabolic streamlining seen in PG 3–6 strains, unlike *Y. pestis*, are likely to be the consequence of transition from environmental ubiquity to specialization in enteric infection of animals rather than adapting to a new lifestyle including an insect vector.

What is significant about *Yersinia* is that until now we have associated widespread functional gene loss with acutely pathogenic lineages. Although this is true of *Y. pestis*, it is not completely true for *Y. enterocolitica*, where the degraded genomes are found in PG 3–6 strains. These latter strains show limited pathogenicity in mouse models and, although not the most acute pathogens, are the most successful lineages in terms of disease causation. These series of events that appear as common evolutionary paths in the most distant *Yersinia* lineages, and across the *Enterobacteriaceae*, are markers of the emergence of lineages with an increasingly restricted lifestyle. These highly adapted organisms can cause a spectrum of disease and have emerged independently, on multiple occasions, out of an environmental background to become successful human and animal pathogens. Thus, these dramatic genetic changes we have demonstrated in the genus *Yersinia* form a paradigm rather than a unique combination of chance events, and appear to underlie the emergence of pathogenic enterobacteria.

Materials and Methods

Sequencing and Assembly. The genomes of six *Y. enterocolitica* strains (YE12/03, YE56/03, YE53/03, YE212/02, YE149/02, YE3094/96) were sequenced and completed to an Improved High Quality Draft (43) standard using multiple sequencing technologies (ABI3730 automated sequencers 2× coverage, 454/Roche GS20/FLX 30× coverage, and Illumina GAII 60× coverage). End-sequenced PCR products were used to close gaps and scaffold large repeat regions. The assembly was corrected using ICORN (44).

Sequence for the worldwide collection was generated to a minimum of 20-fold coverage using the Illumina sequencing platform (GAII instruments; 6–12 samples per lane of sequencing, 200- to 300-bp fragments, 76-bp paired-end reads). This collection included 27 *Y. pseudotuberculosis* and 118 *Y. enterocolitica* strains, as well as further 78 isolates of environmental species, plus previously published genomes (Table S1). The genome accession numbers can be found in Table S1.

Phylogenetic Analysis. To determine the phylogenetic relationship of the genus *Yersinia* housekeeping genes present, all *Yersinia* sequenced were identified based on maximum of 25% SNP divergence between *Y. pestis* and *Y. enterocolitica*. The genes were extracted from de novo velvet assemblies (45) and a concatenated sequence of the chosen 84 genes encompassing 99,724 bases (Table S3) was used to reconstruct the phylogeny of the chromosome. The plasmid alignment was generated by mapping sequenced reads against pYVe8081 (46). See Dataset S3 for single copy core genes used for reconstruction of pangenome genus phylogeny of 50 representative isolates (Fig. S1).

The concatenated sequences for the chromosome and pYV plasmid were both used to construct an unrooted phylogeny using RAxML with a general time-reversible evolutionary model and γ-correction for among site rate variation. Species complexes were defined using the BAPS program (14) run for three independent iterations with upper population sizes of 25, 30, and 35, to obtain the optimal clustering of the population (47).

The *Y. enterocolitica* SNP analysis was based on whole-genome alignment produced by mapping using the YE8081 genome as reference as described previously (46). Mobile genetic elements were removed, and the resulting SNP alignment was analyzed using RAxML (see above). Hierarchical BAPS clustering was performed to define the population structure.

Search for Genes and Operons Related to Virulence and Pathogenicity. Virulence genes were identified using BLAST searches of the assembled genomes

(Dataset S1). Heatmap colors are based upon average amino acid identity either of a gene, where a single gene is examined, or as average of the amino acid identity across the genes in the operon, as indicated.

Phenotypic Microarray Experiment and Analysis. Bacterial strains were cultured on Luria Bertani agar plates at 25 °C. Inoculation and preparation was as manufacturer instructions (Biolog). PM1, 2A (both carbon), PM3B (nitrogen), and PM4A (phosphorus, sulfur) were chosen. L-Cystine (homo-dimer of L-cysteine, 12.5 μM) was added as a reduced sulfur source to all plates, and sodium pyruvate (2M) was added to PM3B and 4A. Plates were incubated under aerobic conditions for 48 h at 28 °C in the OmniLog Incubator/Reader, taking readings every 15 min. Three biological replicates were conducted for each experiment. Data were exported and analyzed in R with datapoints being transformed in to signal values, as described previously (Dataset S2) (48). Each plate showed a clear bimodal distribution of signal values, and normal distributions were fitted to the two peaks to characterize metabolic activity ("off" and "on"). Log-odds scores were calculated for each data point showing how likely the "on" distribution was for each nutrient source. A conservative cut-off of four times more likely to be "on" than "off" was used for further analysis. The limma R/Bioconductor package was used to examine differential metabolism whereby all strains were compared with YE8081c, and the Benjamini-Hochberg corrected *P* values were used to determine statistical significance of differences controlling for a false-discovery rate of 10%.

ACKNOWLEDGMENTS. We thank Martin Cormican (University College Galway) and Dr. Brenda Murphy (University College Dublin), as well as Peter Roggentin, Henry Derschum, and Herbert Nattermann for providing strains for sequencing. This study was funded in part by Wellcome Trust Sanger Institute Grant 098051; Nottingham Trent University Vice Chancellor scholarships (to S.R. and M.E.H.); Academy of Finland Grant 1114075 (to M. Skurnik); and Finnish Ministry of Agriculture and Forestry Grant 4850/501/2004 (to A.S.). M.A. was funded by Science Foundation of Ireland Grant 05/FE1/B882.

1. Rasko DA, et al. (2011) Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med* 365(8):709–717.
2. Bennett JS, et al. (2010) Independent evolution of the core and accessory gene sets in the genus *Neisseria*: Insights gained from the genome of *Neisseria lactamica* isolate 020-06. *BMC Genomics* 11:652.
3. Parkhill J, et al. (2001) Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 413(6855):523–527.
4. Parkhill J, et al. (2001) Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 413(6858):848–852.
5. Bliven KA, Maurelli AT (2012) Antivirulence genes: Insights into pathogen evolution through gene loss. *Infect Immun* 80(12):4061–4070.
6. Sokurenko EV, Hasty DL, Dykhuizen DE (1999) Pathoadaptive mutations: Gene loss and variation in bacterial pathogens. *Trends Microbiol* 7(5):191–195.
7. Chain PSG, et al. (2004) Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* 101(38):13826–13831.
8. Chen PE, et al. (2010) Genomic characterization of the *Yersinia* genus. *Genome Biol* 11(1):R1.
9. Achtman M, et al. (2004) Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc Natl Acad Sci USA* 101(51):17837–17842.
10. Morelli G, et al. (2010) *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* 42(12):1140–1143.
11. Wren BW (2003) The *yersiniae*—A model genus to study the rapid evolution of bacterial pathogens. *Nat Rev Microbiol* 1(1):55–64.
12. Carniel E (2003) Evolution of pathogenic *Yersinia*, some lights in the dark. *Adv Exp Med Biol* 529:3–12.
13. Carniel E (2002) Plasmids and pathogenicity islands of *Yersinia*. *Curr Top Microbiol Immunol* 264(1):89–108.
14. Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. *Math Biosci* 205(1):19–31.
15. Merhej V, Adékambi T, Pagnier I, Raoult D, Drancourt M (2008) *Yersinia massiliensis* sp. nov., isolated from fresh water. *Int J Syst Evol Microbiol* 58(Pt 4):779–784.
16. Kotetishvili M, et al. (2005) Multilocus sequence typing for studying genetic relationships among *Yersinia* species. *J Clin Microbiol* 43(6):2674–2684.
17. Laukkanen-Ninios R, et al. (2011) Population structure of the *Yersinia pseudotuberculosis* complex according to multilocus sequence typing. *Environ Microbiol* 13(12): 3114–3127.
18. Revell PA, Miller VL (2001) *Yersinia* virulence: More than a plasmid. *FEMS Microbiol Lett* 205(2):159–164.
19. Miller VL, Falkow S (1988) Evidence for two genetic loci in *Yersinia enterocolitica* that can promote invasion of epithelial cells. *Infect Immun* 56(5):1242–1248.
20. Foultier B, Troisfontaines P, Müller S, Opperdoes FR, Cornelis GR (2002) Characterization of the *ysa* pathogenicity locus in the chromosome of *Yersinia enterocolitica* and phylogeny analysis of type III secretion systems. *J Mol Evol* 55(1):37–51.
21. Collyn F, Billault A, Mullet C, Simonet M, Marceau M (2004) YAPI, a new *Yersinia pseudotuberculosis* pathogenicity island. *Infect Immun* 72(8):4784–4790.
22. Carniel E, Guilvout I, Prentice MB (1996) Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia enterocolitica*. *J Bacteriol* 178(23): 6743–6751.
23. Cathelyn JS, Ellison DW, Hinchliffe SJ, Wren BW, Miller VL (2007) 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(1):189–205.
24. Dube PH, Handley SA, Revell PA, Miller VL (2003) The *rovA* mutant of *Yersinia enterocolitica* displays differential degrees of virulence depending on the route of infection. *Infect Immun* 71(6):3512–3520.
25. Wang X, et al. (2011) 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(4):1251–1259.
26. Batzilla J, Antonenka U, Höper D, Heesemann J, Rakin A (2011) *Yersinia enterocolitica palearctica* serobiotype O:3/4—A successful group of emerging zoonotic pathogens. *BMC Genomics* 12:348.
27. Fuchs TM, Brandt K, Starke M, Rattei T (2011) Shotgun sequencing of *Yersinia enterocolitica* strain W22703 (biotype 2, serotype O:9): Genomic evidence for oscillation between invertebrates and mammals. *BMC Genomics* 12:168.
28. Prentice MB, et al. (2003) Cobalamin synthesis in *Yersinia enterocolitica* 8081. Functional aspects of a putative metabolic island. *Adv Exp Med Biol* 529:43–46.
29. Jeter RM, Olivera BM, Roth JR (1984) *Salmonella* Typhimurium synthesizes cobalamin (vitamin B12) de novo under anaerobic growth conditions. *J Bacteriol* 159(1):206–213.
30. Price-Carter M, Tingey J, Bobik TA, Roth JR (2001) The alternative electron acceptor tetrathionate supports B12-dependent anaerobic growth of *Salmonella enterica* serovar Typhimurium on ethanolamine or 1,2-propanediol. *J Bacteriol* 183(8):2463–2475.
31. Barrett EL, Clark MA (1987) Tetrathionate reduction and production of hydrogen sulfide from thiosulfate. *Microbiol Rev* 51(2):192–205.
32. Winter SE, et al. (2010) Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* 467(7314):426–429.
33. Cornelis GR (2006) The type III secretion injectisome. *Nat Rev Microbiol* 4(11):811–825.
34. Portnoy DA, Falkow S (1981) Virulence-associated plasmids from *Yersinia enterocolitica* and *Yersinia pestis*. *J Bacteriol* 148(3):877–883.
35. Snellings NJ, Popek M, Lindler LE (2001) Complete DNA sequence of *Yersinia enterocolitica* serotype 0:8 low-calcium-response plasmid reveals a new virulence plasmid-associated replicon. *Infect Immun* 69(7):4627–4638.
36. Wauters G, Kandolo K, Janssens M (1987) Revised biogrouping scheme of *Yersinia enterocolitica*. *Contrib Microbiol Immunol* 9:14–21.
37. Thomson NR, et al. (2006) The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. *PLoS Genet* 2(12):e206.
38. Swaminathan B, Harmon MC, Mehlman IJ (1982) *Yersinia enterocolitica*. *J Appl Bacteriol* 52(2):151–183.
39. Wuthe HH, Aleksić S (1997) [*Yersinia enterocolitica* serovar 2a, wb, 3:b,c biovar 5 in hares and sheep]. *Berl Munch Tierarztl Wochenschr* 110(5):176–177. German.
40. Lian CJ, Hwang WS, Kelly JK, Pai CH (1987) Invasiveness of *Yersinia enterocolitica* lacking the virulence plasmid: An in-vivo study. *J Med Microbiol* 24(3):219–226.
41. Schmid Y, et al. (2004) *Yersinia enterocolitica* adhesin A induces production of interleukin-8 in epithelial cells. *Infect Immun* 72(12):6780–6789.
42. Andersson DI, Hughes D (1996) Muller's ratchet decreases fitness of a DNA-based microbe. *Proc Natl Acad Sci USA* 93(2):906–907.
43. Chain PS, et al.; Genomic Standards Consortium Human Microbiome Project Jumpstart Consortium (2009) Genomics. Genome project standards in a new era of sequencing. *Science* 326(5950):236–237.
44. Otto TD, Sanders M, Berriman M, Newbold C (2010) Iterative orrection of Reference Nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics* 26(14):1704–1707.
45. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829.
46. Harris SR, et al. (2010) Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327(5964):469–474.
47. Cheng L, Connor TR, Sirén J, Aanensen DM, Corander J (2013) Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol Biol Evol* 30(5): 1224–1228.
48. Homann OR, Cai H, Becker JM, Lindquist SL (2005) Harnessing natural diversity to probe metabolic pathways. *PLoS Genet* 1(6):e80.