

# Comparative Analysis of *Salmonella* Genomes Identifies a Metabolic Network for Escalating Growth in the Inflamed Gut

Sean-Paul Nuccio, Andreas J. Bäumlner

Department of Medical Microbiology and Immunology, School of Medicine, University of California at Davis, Davis, California, USA

**ABSTRACT** The *Salmonella* genus comprises a group of pathogens associated with illnesses ranging from gastroenteritis to typhoid fever. We performed an *in silico* analysis of comparatively reannotated *Salmonella* genomes to identify genomic signatures indicative of disease potential. By removing numerous annotation inconsistencies and inaccuracies, the process of reannotation identified a network of 469 genes involved in central anaerobic metabolism, which was intact in genomes of gastrointestinal pathogens but degrading in genomes of extraintestinal pathogens. This large network contained pathways that enable gastrointestinal pathogens to utilize inflammation-derived nutrients as well as many of the biochemical reactions used for the enrichment and biochemical discrimination of *Salmonella* serovars. Thus, comparative genome analysis identifies a metabolic network that provides clues about the strategies for nutrient acquisition and utilization that are characteristic of gastrointestinal pathogens.

**IMPORTANCE** While some *Salmonella* serovars cause infections that remain localized to the gut, others disseminate throughout the body. Here, we compared *Salmonella* genomes to identify characteristics that distinguish gastrointestinal from extraintestinal pathogens. We identified a large metabolic network that is functional in gastrointestinal pathogens but decaying in extraintestinal pathogens. While taxonomists have used traits from this network empirically for many decades for the enrichment and biochemical discrimination of *Salmonella* serovars, our findings suggest that it is part of a “business plan” for growth in the inflamed gastrointestinal tract. By identifying a large metabolic network characteristic of *Salmonella* serovars associated with gastroenteritis, our *in silico* analysis provides a blueprint for potential strategies to utilize inflammation-derived nutrients and edge out competing gut microbes.

Received 10 February 2014 Accepted 12 February 2014 Published 18 March 2014

**Citation** Nuccio S, Bäumlner AJ. 2014. Comparative analysis of *Salmonella* genomes identifies a metabolic network for escalating growth in the inflamed gut. *mBio* 5(2):e00929-14. doi:10.1128/mBio.00929-14.

**Editor** B. Brett Finlay, The University of British Columbia

**Copyright** © 2014 Nuccio and Bäumlner This is an open-access article distributed under the terms of the [Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license](#), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Andreas J. Bäumlner, [ajbaumlner@ucdavis.edu](mailto:ajbaumlner@ucdavis.edu).

Among the foremost insights sought at the dawn of the genomic era was the information held within pathogen genomes. In the ensuing years, elevated genome degradation has surfaced as a common trait among diverse subsets of bacteria exhibiting relatively specialized lifestyles and pathogenicity, including members of the genera *Coxiella*, *Mycobacterium*, *Salmonella*, *Shigella*, and *Yersinia* (1–6). Nevertheless, specific connections between genome degradation and major alterations to pathogen behavior remain elusive.

As a model pathogen and worldwide scourge of both humans and animals, *Salmonella* is an important focus of novel research into the myriad aspects of pathogenesis, from the basic physiology of bacteria to the function of the host’s immune system. Based on their pathogenic potential, members of the species *Salmonella enterica* are often divided into those causing typhoid fever or paratyphoid fever in humans, termed typhoidal *Salmonella* serovars, and those associated with a localized gastroenteritis in immunocompetent individuals, termed nontyphoidal *Salmonella* serovars. However, the properties that distinguish *Salmonella* serovars associated with a localized gastroenteritis from those causing disseminated infections remain poorly understood.

Advances in high-throughput sequencing make genomic com-

parison an increasingly powerful tool for identifying features that might explain differences in the disease potential of *Salmonella* serovars. Even so, the process of genome annotation can produce a considerable number of errors, an outcome which is enhanced by an overreliance on automation. Furthermore, genomes available for comparison are annotated using different methods, and the sequences are increasingly left unfinished, limiting the power of comparative genome analysis.

Here, we performed a manually curated comparative reannotation of orthologs from 15 completed *S. enterica* genomes to identify genomic signatures that distinguish pathogens causing different disease presentations. Our analysis suggests that removal of annotation inconsistencies and inaccuracies through the annotation normalization process markedly enhanced the resolution of comparative genome analysis, thereby enabling us to identify a previously hidden genetic fingerprint that distinguishes pathogens associated with gastroenteritis from those causing disseminated disease.

## RESULTS

**Comparative reannotation of 15 *Salmonella* genomes.** Fifteen completed *S. enterica* genomes, comprising all serovars with a gap-

less chromosome assembly available from NCBI at the time this work was initiated, were included in the analysis (see Fig. S1A in the supplemental material). *S. enterica* serovar Paratyphi B is a polyphyletic lineage containing pathogens associated with paratyphoid fever as well as members of the variety Java, which are associated with gastroenteritis (7). The *S. Paratyphi B* genomic sequence included in our analysis originated from strain SPB7, a representative of the variety Java. Thus, our collection contained 5 genomes representing typhoidal serovars, including *S. enterica* serovar Typhi (strains CT18 and Ty2), *S. Paratyphi A* (strains ATCC 9150 and AKU 12601), and *S. Paratyphi C*. The remaining 10 genomes represented nontyphoidal serovars.

A roadblock encountered early during our analysis was that the different methods used for annotating available genomes, along with a considerable number of inaccuracies detected in some annotations, rendered any direct comparison of the degraded (i.e., hypothetically disrupted or deleted) content between genomes imprecise. We thus performed a comparative reannotation of ortholog data from all 15 genomes (see Table S1 in the supplemental material), identified deletions (see Table S2), and compiled the degraded content in each genome (see Table S3). To reflect their putatively disrupted state, we will refer to loci previously called “pseudogenes” instead as hypothetically disrupted coding DNA sequences (HDCs); as the literal meaning of “pseudogene” is “false gene,” as in “without function,” and as it is often ambiguously employed to denote genes of hypothetical or validated disrupted status, we suggest that its usage be reserved for labeling loci where loss of all known function has been empirically demonstrated (e.g., the *fepE* pseudogene of *S. Typhi* [8]).

It was possible to automate only a portion of the reannotation process, which made this task time-consuming. However, the necessity to perform this onerous *in silico* analysis was validated by the identification of marked changes in the degraded content for each genome (see Table S4 in the supplemental material). For example, our reannotation of 15 *S. enterica* genomes identified a total of 1,004 new HDCs, while a total of 471 entries, which had been annotated as “pseudogenes” previously, were found to be intact hypothetical coding DNA sequences (CDSs).

**A genomic signature distinguishes two *Salmonella* pathovars.** Surprisingly, our analysis of comparatively reannotated *S. enterica* genomes did not provide compelling support for a classification into typhoidal and nontyphoidal serovars. Degradation of only three genes, *fhuE*, *fliB*, and *STM4065*, was unique to and present in all analyzed typhoidal serovars (see Table S3 in the supplemental material). Furthermore, degradation of the *wca* gene cluster, which encodes colanic acid biosynthesis, was common and unique to genomes of typhoidal serovars.

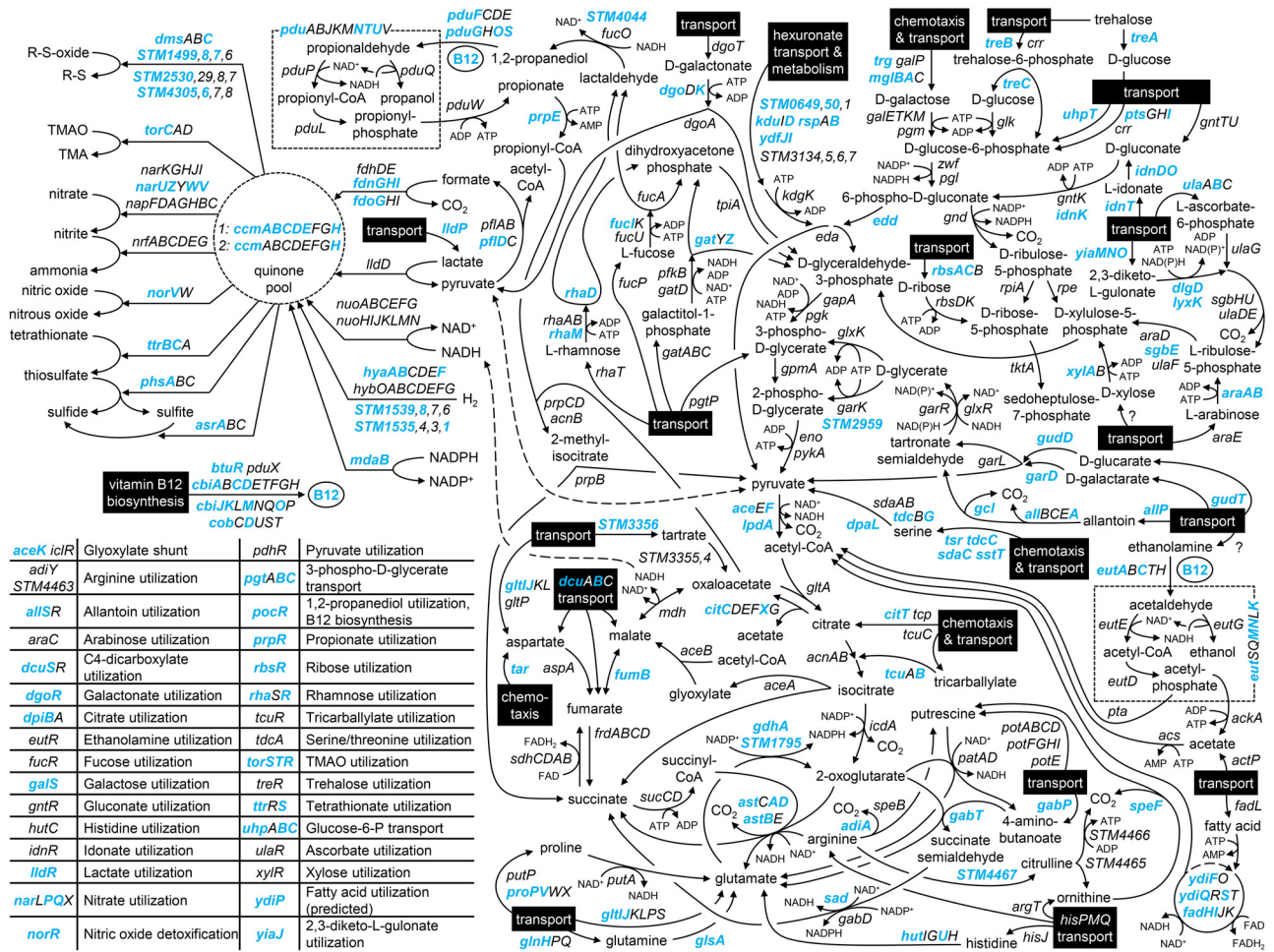
However, analysis of the degraded content in each genome suggested that *S. enterica* serovars could be divided into one group carrying a low number of HDCs (on average 66 HDCs per genome) and a second group with a high number of HDCs (on average 246 HDCs per genome) (see Fig. S1B and Table S4 in the supplemental material). The latter group, which we will refer to as the “extraintestinal pathovar,” was formed by host-adapted serovars associated exclusively with disseminated infections in their respective human or animal reservoirs. Genomes exhibiting the HDC signature of the extraintestinal pathovar included those of *S. enterica* serovar Choleraesuis, which is associated with bacteremia in pigs, *S. enterica* serovar Dublin, a cause of bacteremia in cattle,

*S. enterica* serovar Gallinarum, the causative agent of fowl typhoid in poultry, as well as all typhoidal *Salmonella* serovars incorporated in our analysis (i.e., *S. Paratyphi A*, *S. Paratyphi C*, and *S. Typhi*). Genomes characterized by a low number of HDCs belonged to *S. enterica* serovar Agona, *S. enterica* serovar Enteritidis, *S. enterica* serovar Heidelberg, *S. enterica* serovar Newport, *S. enterica* serovar Schwarzengrund, *S. enterica* serovar Typhimurium, and *S. Paratyphi B*. We will refer to the latter group as the “gastrointestinal pathovar,” because all of its members exhibit a broad host range and are associated with gastroenteritis in at least some host species. It should be noted that certain members of the gastrointestinal pathovar are also able to cause extraintestinal infections in certain hosts. For example, *S. Typhimurium* is associated with bacteremia in mice; however, the pathogen causes a localized gastroenteritis in cattle and in immunocompetent humans. Thus, we refer to this group as the gastrointestinal pathovar, because the ability to cause gastroenteritis in at least some host species presumably places genes necessary for this lifestyle under selection.

Several genomic signatures supporting a distinction between a gastrointestinal pathovar and an extraintestinal pathovar were detected in our analysis. Analysis of CDSs that were frequently degraded ( $n \geq 4$ ) in members of one group but rarely ( $n \leq 1$ ) in members of the other supported a classification into two pathovars but provided little functional insights (see Table S5 in the supplemental material). Analysis of genes involved in virulence revealed that genomes representing the extraintestinal pathovar exhibited more instances of degraded genes encoding type III secreted effector proteins, fimbrial adhesins, and functions related to motility and chemotaxis than did genomes representing the gastrointestinal pathovar (see Table S6 and Fig. S2), which was consistent with a previous report (6). Fimbriae, motility, and chemotaxis are required for intestinal colonization (9–11) but are dispensable for survival in host tissue (12, 13), which may explain why these functions are maintained in the gastrointestinal pathovar but undergo degradation in the extraintestinal pathovar.

The most striking result of our *in silico* analysis of comparatively reannotated *Salmonella* genomes was the identification of a large metabolic network composed of 469 CDSs, 167 of which were uniquely degraded in one or more genomes of the extraintestinal pathovar (Fig. 1; see also Table S7 in the supplemental material). The total number of HDCs and deleted CDSs belonging to this metabolic network, not counting duplicate instances from strains belonging to the same serovar, was 224 for all genomes representing the extraintestinal pathovar, compared to only 13 for all genomes representing the gastrointestinal pathovar (a ratio of 17.23). Statistical analysis revealed that a ratio of 17.23 is approximately 9 standard deviations away from the average ratio obtained when the degraded content is determined for randomly populated groups of 469 CDSs from each genome ( $P \sim 0$ ).

While the statistically overrepresented degradation of metabolic genes identified here provided compelling support for distinguishing an extraintestinal pathovar from a gastrointestinal pathovar, such a classification was not backed by previous genome annotations. Using published annotations, analysis of the 469 CDSs belonging to the metabolic network depicted in Fig. 1 revealed 169 degraded CDSs in the extraintestinal pathovar compared to 46 in the gastrointestinal pathovar. The resulting ratio of 3.67 was not significantly different ( $P = 0.17$ ) from the ratio observed in randomly selected groups of 469 CDSs from each genome, which explains why a previous analysis of these *Salmonella*



**FIG 1** Central anaerobic metabolism of the gastrointestinal pathovar. Black text denotes genes unaffected by degradation in the extraintestinal pathovar, while blue text denotes genes putatively affected by disruptions or deletions in the extraintestinal pathovar. Due to space restrictions, not all intermediates, products, cofactors, or stoichiometries are shown for every reaction; the production of carbon dioxide and the involvement of nucleoside polyphosphate, vitamin B<sub>12</sub>, or adenine dinucleotide cofactors are always shown. The table displays genes whose products regulate processes involved in central anaerobic metabolism.

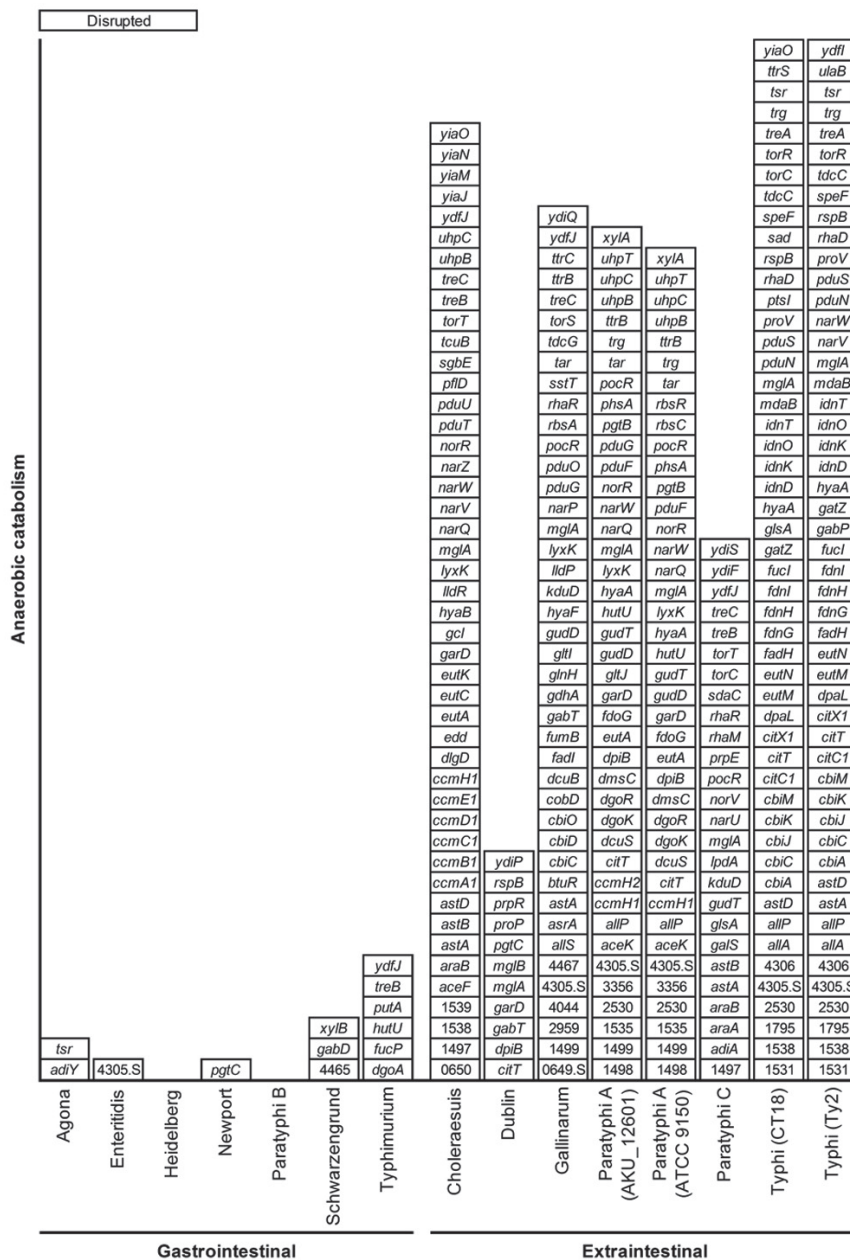
genomes did not identify this large metabolic network (14). Thus, until now, the fact that a network of 469 CDSs involved in central anaerobic metabolism is degrading in the genomes of the extraintestinal pathovar has remained hidden behind the statistical noise generated by inconsistencies and inaccuracies in previous genome annotations.

**A large metabolic network containing functions for the utilization of inflammation-derived nutrients is degrading in the extraintestinal pathovar.** The metabolic network emerging from our analysis includes many functions previously shown to be important for anaerobic growth in the intestinal lumen during gastroenteritis. *S. Typhimurium*, a member of the gastrointestinal pathovar, uses its type III secretion systems encoded by *Salmonella* pathogenicity island 1 (SPI1) and SPI2 to trigger acute intestinal inflammation (15). A by-product of the ensuing inflammatory host response is the generation of the terminal electron acceptors nitrate and tetrathionate, the presence of which boosts luminal growth of the pathogen by anaerobic respiration (16, 17). Our analysis identified these pathways along with several additional functions related to anaerobic respiration, which involves the transfer of electrons from a donor, such as formate, lactate, or

hydrogen (H<sub>2</sub>), through the quinone pool to an acceptor, such as nitrate, tetrathionate, nitrite, S-oxides, N-oxides, nitric oxide, thiosulfate, or sulfite (Fig. 1). Formate, lactate, and hydrogen are fermentation end products generated by obligate anaerobic microbial communities inhabiting the distal gut (18, 19), and microbiota-derived hydrogen has recently been shown to fuel growth of *S. Typhimurium* in the lumen of the large bowel (20).

The presence of alternative electron acceptors, such as tetrathionate, enables *S. Typhimurium* to grow on other nonfermentable carbon sources, such as ethanolamine, which is produced by microbial degradation of the abundant phospholipid phosphatidylethanolamine in the distal gut (21). Genomes representing the extraintestinal pathovar exhibited degradation of CDSs involved in ethanolamine utilization (*eut* genes), as well as in the biosynthesis of vitamin B<sub>12</sub> (*cbi* and *cob* genes), a cofactor produced under anaerobic condition, which is required for ethanolamine utilization (22) (Fig. 2).

Vitamin B<sub>12</sub> is also necessary for the utilization of 1,2-propanediol, a catabolite produced by microbes fermenting fucose or rhamnose. Expression of *S. Typhimurium* proteins involved in sugar catabolism is increased in the intestinal lumen in a



**FIG 2** Degradation of central anaerobic metabolism. Boxes contain the names of all hypothetically disrupted or deleted coding DNA sequences (CDSs) involved in central anaerobic metabolism for each genome analyzed. Entries with numbers represent abbreviated STM locus tags (e.g., 4308 = *STM4308*).

mouse colitis model (23). Furthermore, communities of obligate anaerobic bacteria in the distal gut liberate host mucus-derived monosaccharides, such as fucose, which leads to increased expression of *S. Typhimurium* genes involved in the degradation of fucose (*fuc* genes) and its fermentation product 1,2-propanediol (*pdu* genes) in the intestinal lumen of mice monoassociated with *Bacteroides thetaiotaomicron* compared to germfree mice (24). Our analysis identified substantial degradation in the extraintestinal pathovar across a large network of genes involved in the uptake and catabolism of various monosaccharides, which included the *fuc* and *pdu* genes (Fig. 1).

Besides pathways that have surfaced previously in studies on

luminal growth of *S. Typhimurium* during colitis, our network identified several new functions that likely contribute to the central anaerobic metabolism of the gastrointestinal pathovar. For instance, degradation of CDSs involved in anaerobic  $\beta$ -oxidation of fatty acids was overrepresented in genomes representing the extraintestinal pathovar. This pathway, which is distinct from the aerobic  $\beta$ -oxidation pathway for fatty acid degradation, is encoded by the *ydiFO*, *ydiQRST*, and *fadHIJK* genes and requires the presence of an alternative electron acceptor, such as nitrate, *S*-oxides, or *N*-oxides (23). Interestingly, short-chain fatty acids accumulate in the lumen of the distal gut when communities of obligate anaerobic bacteria break down and ferment complex carbohydrates, while nitrate is generated in this environment as a by-product of the inflammatory host response (17), which is elicited when *S. Typhimurium* deploys the type III secretion systems encoded by SPI1 and SPI2 (15).

All *Salmonella* genomes exhibited very little degradation of CDSs involved in central metabolic functions required under aerobic conditions, likely because these traits are essential for bacterial growth in host tissue (25); for example, the genes involved in the glyoxylate cycle, an anaerobic variant of the aerobic tricarboxylic acid cycle, remained intact, presumably because their functions are also required for the aerobic version of this pathway. However, degradation of CDSs involved in the uptake of compounds from the environment that can replenish intermediates in the glyoxylate cycle, such as citrate, tartrate, tricarballylate, serine, and aspartate, was overrepresented in genomes representing the extraintestinal pathovar (Fig. 1). Furthermore, CDSs required for anaplerotic reactions that fill the gap between 2-oxoglutarate and succinate in the anaerobic glyoxylate cycle were commonly degraded in genomes representing the extraintestinal pathovar. These anaplerotic reactions are not required under aerobic conditions, because *SucA* and *SucB* convert 2-oxoglutarate into succinyl-coenzyme A (CoA) within the tricarboxylic acid cycle.

Finally, genomes representing the extraintestinal pathovar exhibited degradation of regulators for a variety of anaerobic processes, including anaerobic respiration (*narPQ*, *norR*, *torSTR*, *ttrS*), the consequent anaerobic degradation of fermentation products and fatty acids (*lldR*, *pocR*, *prpR*, and *ydiP*), carbohydrate catabolism (*dgoR*, *galS*, *rbsR*, *rhaR*, *uhpBC*, *yiaJ*), and functions related to the aerobic glyoxylate cycle (*aceK*, *dcuS*, and *dpiB*) (Fig. 1 and 2).

## DISCUSSION

The large metabolic network identified in our analysis (Fig. 1) contained many of the biochemical reactions taxonomists and clinical laboratories use to isolate and discriminate *Salmonella* serovars. For example, growth in broth containing tetrathionate has been in use since 1923 as a method to enrich for *Salmonella* serovars in samples containing other microbes (26). This initial enrichment culture is followed by detecting the production of sulfide on iron or bismuth-containing selective agar, such as triple sugar iron agar slants developed in 1917 (27) or bismuth sulfite agar plates developed in 1923 (28). While these metabolic traits have been used empirically for many decades to isolate *Salmonella* serovars, our analysis suggests they are part of a large metabolic network that defines the gastrointestinal pathovar. Since the vast majority of the more than 2,500 *S. enterica* serovars is associated with gastroenteritis in immunocompetent humans, it might be unsurprising that these functions are often considered to be characteristic of the entire *S. enterica* species, despite the fact that they are degrading in genomes of a few specialists belonging to the extraintestinal pathovar.

Degradation in the extraintestinal pathovar of functions involved in anaerobic central metabolism (Fig. 1) is used empirically to distinguish pathogens associated with paratyphoid fever from closely related organisms that cannot be differentiated by serotyping but cause gastroenteritis in humans. One example is *S. Paratyphi* B variety Java, a pathogen associated with human gastroenteritis, which has the same antigen formula (1,4[5],0.12:b:1,2) as *S. Paratyphi* B, a cause of paratyphoid fever. The ability to ferment tartrate is used empirically to distinguish these two pathogens biochemically (7). While *S. Paratyphi* B variety Java isolates can ferment tartrate, this pathway that contributes to the metabolic network identified in our analysis is disrupted by a nucleotide transition from G to A within the ATG start codon of *STM3356* in *S. Paratyphi* B isolates from patients with paratyphoid fever (29). A second example is *S. enterica* serovar Sendai, a cause of paratyphoid fever, which has the same antigen formula (1,9,12:a:1,5) as *S. enterica* serovar Miami, a cause of human gastroenteritis. Both pathogens can be distinguished biochemically, because isolates of *S. Miami* can ferment citrate, while *S. Sendai* isolates are negative for this reaction within the anaerobic central metabolism (30).

From the perspective of serovars among *S. enterica*, our analysis of comparatively reannotated genomes represents the broadest in-depth examination of *Salmonella* genome degradation to date. In this regard, the monophyletic origin and high similarity of *S. Typhi* isolates (31), coupled with the polyphyletic, host-isolated history of the extraintestinal pathovar (see Fig. S1 in the supplemental material) (32) and our inclusion of a similar broad assortment of gastrointestinal serovars (see Fig. S1), suggest that our data set is suitably diverse. These considerations, together with the exceedingly low probability that central anaerobic metabolism degradation arose stochastically in all analyzed members of the extraintestinal pathovar, as well as the similar unlikelihood that the difference in said degradation among the pathovars is an artifact arising from the specific 15 genomes we analyzed, give us confidence that our observations will hold true as more strains and serovars are sequenced; indeed, we expect that expanding the number of genomes analyzed will bring even more subtle, potentially host-specific degradative patterns to prominence.

Still, many forms of genome alteration exist that are, at present,

more difficult to postulate the effects of through *in silico* analysis alone. Such instances include the identification and adaptive roles of novel hypomorphic alleles arising from missense mutations (e.g., the E211 allele of *pmrA* in extraintestinal *S. Paratyphi* B) (33), the outcome of mutation within *cis*-acting regulatory elements, the polarity of indels located within known or putative operons, and the influence of regulator acquisition through horizontal gene transfer (e.g., regulon alterations made by TviA of *S. Typhi*) (34, 35). On this front, empirical analysis is essential to facilitating their identification and rationalization. The necessity for experimental analysis is compellingly illustrated by the example of the *fepE* gene, which encodes a regulator of very long O-antigen chain (>100 repeat units) assembly (36), a surface structure conferring bile resistance in *S. Typhimurium* (37). In the *S. Typhi* genome, the *fepE* open reading frame is disrupted by a stop codon (2), resulting in loss of very long O-antigen chains (8). Interestingly, this loss of very long O-antigen chains maximizes immune evasion mediated by the virulence-associated (Vi) capsular polysaccharide of *S. Typhi* (38). Thus, the consequences of pseudogene formation can be complex, illustrating the need to follow up *in silico* studies with an experimental analysis.

Nevertheless, putting the degradative genomic signatures we detected by *in silico* analysis of comparatively reannotated *S. enterica* genomes into the context of the existing body of work on the biology of these pathogens supports a model that distinguishes two pathovars, each exploiting a different host niche for transmission. Members of the gastrointestinal pathovar use their virulence factors to rapidly induce acute intestinal inflammation (15) and to exploit the ensuing changes in the environment by boosting their luminal growth using a large metabolic network involved in central anaerobic metabolism (Fig. 1) (11, 16, 17, 21, 24). The resulting luminal bloom of members of the gastrointestinal pathovar enhances their transmission by the fecal-oral route (39).

In contrast, *S. Typhi*, a member of the extraintestinal pathovar, initially suppresses intestinal inflammation (38, 40, 41) and causes a disseminated infection known as typhoid fever. A small fraction (approximately 4%) of individuals that recover from typhoid fever develop chronic gallbladder carriage and are the main reservoir for transmission of typhoid fever (42). While other members of the extraintestinal pathovar also cause disseminated infections, some exploit different organs for transmission, such as the ovaries in the case of *S. Gallinarum* (43) or the udder in the case of *S. Dublin* (44). Nevertheless, in each case, the organism's transmission is facilitated by dissemination followed by chronic persistence in host tissue, a microaerobic environment (25), thereby rendering genes required for anaerobic growth in the distal gut dispensable to the extraintestinal pathovar. Our analysis shows that the resulting degradation of functions involved in central anaerobic metabolism is an experiment of nature that produced a prominent genetic fingerprint characteristic of genomes representing the extraintestinal pathovar. By identifying functions degrading in genomes of the extraintestinal pathovar, our study defined a large metabolic network that likely epitomizes the "winning strategy" employed by members of the gastrointestinal pathovar to edge out competing microbes in the lumen of the inflamed gut, thereby enhancing their transmission.

## MATERIALS AND METHODS

**Comparative reannotation.** For each analyzed genome (see the list at the top of Table S1 in the supplemental material) (2, 6, 14, 45–50), we gath-

ered all CDS and pseudo-CDS information by parsing NCBI GenBank records. We then obtained UniProt KnowledgeBase (51) records for these loci by cross-referencing Entrez GeneIDs (52) and parsed them for gene names, functional annotations, and associated COG (53), PFAM (54), and TIGRFAM (55) protein domains. To normalize ortholog annotations, we took one CDS at a time from the index as a reference and located its orthologs in the other genomes, blinding initial reference choices to gene function and biasing it to the least degraded manually curated genomes (*S. Typhimurium* LT2, *S. Enteritidis* P125109).

To annotate orthologs, we wrote custom scripts to analyze reference sequence alignments made to subject genomes with *blastn* and *tblastn* via NCBI's Web application programming interface (API) (56). In brief, our script parsed and collated BLAST results, we manually confirmed contextually accurate alignments, and then the script integrated coordinates and sequence information from both BLAST methods to locate the bounds of the reference gene in the subject genome; if an aligned start or stop codon was not located, we manually inspected the region. The script then analyzed alignments for insertions, deletions, premature stop codons, frame-shifts, and changes to the start codon. We define an HDC to be an orthologous locus with  $\geq 10$  codons disrupted by the aforementioned mutations relative to a reference CDS. An alignment in the same genomic context with  $\geq 90\%$  amino acid identity, excluding gaps and truncations, was our initial cutoff for orthology. Granted that any such cutoffs are arbitrary, we postulated that larger open reading frame alterations to highly similar CDSs would be more likely to signal disrupted function; therefore, our size cutoff was chosen to avoid noise in the form of smaller, potentially nondisruptive events (e.g., truncations of a single codon). In this regard, our disruption size cutoff is effectively less than or equal to all previous cutoffs among the genomes analyzed, as evidenced by the at most two instances per genome (see Table S4 in the supplemental material, "Now Unclear" column) of previous pseudogene calls bearing a potential disruption that did not meet our size cutoff. Nevertheless, all sub-cutoff events are labeled "Unclear" in the supplemental tables should the reader desire to consider them.

Next, if the majority annotation did not match that of the reference, we investigated the reference and switched it with an ortholog's annotation if appropriate. Prior to selecting a new reference, our script removed any locus tags from the index that were associated with identified orthologs. Table S1 in the supplemental material contains data collected on each ortholog, with the genome of LT2 serving as a scaffold for ordering entries and with episomal data placed at the end of the list. The Table S1 legend describes the data and provides associated cutoffs.

To preclude analyzing potentially overannotated genome content, we discarded CDSs  $\leq 75$  codons from the potential reference index unless they bore an annotated function, informative homology, or a protein domain. References found within prophage or mobile genetic elements were compared only for orthologs with similar regions located in the same genomic context. As the expression of integrases and transposition-related genes is not known to immediately impact the pathobiology of *Salmonella* serovars, we did not meticulously investigate these entries or mark them as intact or disrupted; we identified these loci using the IS-Finder database (57) and CD-Search (58). Regarding previously annotated pseudo-CDSs that did not associate with intact references, we checked for disruptions relative to nonorthologous references and then checked for orthologs, discarding small fragments and loci that were disrupted in all analyzed strains, as their differential role in genome degradation was unclear at this juncture.

**Deletions and truncations.** To identify disruptive lesions, we located remnants of reference loci from Table S1 in the supplemental material and of RNA genes as an indicator that a gene or region was present and subsequently truncated or deleted. Table S2 in the supplemental material contains a list of alignment gaps within, and extending outside, at least one locus and that we propose to be disruptive (see Table S2 for definitions and cutoffs; Table S1 data contains intragenic indels). In brief, we wrote scripts and used manual curation to systematically compare par-

tially overlapping segments of *S. Typhimurium* LT2 against all other analyzed genomes, utilizing the megablast algorithm of *blastn* via the BLAST Web API (56) with a high-scoring alignment pair cutoff of 80% identity, and then catalogued alignment gaps residing within the same genomic context. We then compared regions in the same context that were missing from LT2 and filtered out highly mosaic regions and dissimilar prophage insertions in the same context from further examination. Our script identified gap intersections with reference locus coordinates and calculated disruptions, which we then manually curated and swapped with other regions to serve as a reference when the original reference appeared to be affected, updating Table S1 references as necessary.

We marked missing regions without a flanking remnant as absent. If an absent region from one strain resided completely within a proposed deletion in another strain, we marked that section of the deletion as absent. When reference DNA was plausibly not present (e.g., mobile element insertion) prior to a proposed deletion having occurred, or when stepwise intermediate genotypes were unavailable to resolve multiple instances having occurred, we marked the region as absent and marked the disrupted border gene(s) as truncated.

**CDS groupings.** To identify pathways involved in central anaerobic metabolism, we examined primary literature, associated entries in the *Kyoto Encyclopedia of Genes and Genomes* (59), and *Escherichia coli* K-12 ortholog entries in the BioCyc database (60). To index genes involved in other aspects of pathogenesis, we used protein domains to identify chaperone-usher fimbrial gene clusters (61), obtained the identities of type III secretion system effectors primarily from reference 62, and utilized the *S. Typhimurium* FlhDC regulon (63) to populate our list of motility and chemotaxis CDSs.

To calculate the probability of the observed extraintestinal-to-gastrointestinal pathovar ratio of total degradation in the central anaerobic metabolism group (3.67 before reannotation, 17.23 after) having occurred at random, we generated 250 random groups of 469 reference loci present or once present in  $\geq 10$  of the analyzed genomes; multiple hits for a reference locus within a serovar were tallied only once. From this data set, we log-transformed the ratios and computed the mean (0.482) and standard deviation (0.088) of the random group ratios and then used a quantile-quantile plot to confirm that the log-transformed random ratios closely fit a normal distribution (trendline of  $y = 0.9945x + 6 \times 10^{-16}$ ,  $R^2 = 0.9902$ ). With these values, we computed the *z* scores (before = 0.945, after = 8.598) and one-tailed *P* values (0.172,  $\sim 0$ ) for the log-transformed observed ratios (0.565, 1.236).

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00929-14/-DCSupplemental>.

Figure S1, PDF file, 0.4 MB.  
Figure S2, PDF file, 3.1 MB.  
Table S1, XLSX file, 3.3 MB.  
Table S2, XLSX file, 0.1 MB.  
Table S3, XLSX file, 0.5 MB.  
Table S4, XLSX file, 0.1 MB.  
Table S5, PDF file, 0.1 MB.  
Table S6, XLSX file, 0.1 MB.  
Table S7, XLSX file, 0.1 MB.

## ACKNOWLEDGMENTS

We are grateful to Renée M. Tsois for helpful suggestions on the manuscript.

We acknowledge support by Public Health Service grants AI044170 and AI096528 to A.J.B. and by the Floyd and Mary Schwall Fellowship in Medical Research and Public Health Training Grant AI060555 to S.-P.N.

## REFERENCES

- Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, Sebahia M, James KD, Churcher C, Mungall KL, Baker S, Basham D, Bentley SD, Brooks K, Cerdeño-Tárraga AM, Chillingworth T,

- Cronin A, Davies RM, Davis P, Dougan G, Feltwell T, Hamlin N, Holroyd S, Jagels K, Karlyshev AV, Leather S, Moule S, Oyston PC, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG. 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 413:523–527. <http://dx.doi.org/10.1038/35097083>.
2. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connor P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, O'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG. 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 413:848–852. <http://dx.doi.org/10.1038/35101607>.
  3. Vissa VD, Brennan PJ. 2001. The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set. *Genome Biol.* 2:reviews1023–reviews1023.8. <http://dx.doi.org/10.1186/gb-2001-2-8-reviews1023>.
  4. Jin Q, Yuan Z, Xu J, Wang Y, Shen Y, Lu W, Wang J, Liu H, Yang J, Yang F, Zhang X, Zhang J, Yang G, Wu H, Qu D, Dong J, Sun L, Xue Y, Zhao A, Gao Y, Zhu J, Kan B, Ding K, Chen S, Cheng H, Yao Z, He B, Chen R, Ma D, Qiang B, Wen Y, Hou Y, Yu J. 2002. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K-12 and O157. *Nucleic Acids Res.* 30:4432–4441. <http://dx.doi.org/10.1093/nar/gkf566>.
  5. Seshadri R, Paulsen IT, Eisen JA, Read TD, Nelson KE, Nelson WC, Ward NL, Tettelin H, Davidsen TM, Beanan MJ, Deboy RT, Daugherty SC, Brinkac LM, Madupu R, Dodson RJ, Khouri HM, Lee KH, Carty HA, Scanlan D, Heinzen RA, Thompson HA, Samuel JE, Fraser CM, Heidelberg JF. 2003. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc. Natl. Acad. Sci. U. S. A.* 100:5455–5460. <http://dx.doi.org/10.1073/pnas.0931379100>.
  6. McClelland M, Sanderson KE, Clifton SW, Latreille P, Porwollik S, Sabo A, Meyer R, Bieri T, Ozersky P, McLellan M, Harkins CR, Wang C, Nguyen C, Berghoff A, Elliott G, Kohlberg S, Strong C, Du F, Carter J, Kremizki C, Layman D, Leonard S, Sun H, Fulton L, Nash W, Miner T, Minx P, Delehaunty K, Fronick C, Magrini V, Nhan M, Warren W, Florea L, Spieth J, Wilson RK. 2004. Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of *Salmonella enterica* that cause typhoid. *Nat. Genet.* 36:1268–1274. <http://dx.doi.org/10.1038/ng1470>.
  7. Kauffmann F. 1955. Differential diagnosis and pathogenicity of *Salmonella* Java and *Salmonella* Paratyphi B. *Z. Hyg. Infektionskr.* 141:546–550. (In German.) <http://dx.doi.org/10.1007/BF02156850>.
  8. Bravo D, Silva C, Carter JA, Hoare A, Alvarez SA, Blondel CJ, Zaldivar M, Valvano MA, Contreras I. 2008. Growth-phase regulation of lipopolysaccharide O-antigen chain length influences serum resistance in serovars of *Salmonella*. *J. Med. Microbiol.* 57:938–946. <http://dx.doi.org/10.1099/jmm.0.47848-0>.
  9. Weening EH, Barker JD, Laarakker MC, Humphries AD, Tsolis RM, Bäuml AJ. 2005. The *Salmonella enterica* serotype Typhimurium *lpf*, *bcf*, *stb*, *stc*, *std*, and *sth* fimbrial operons are required for intestinal persistence in mice. *Infect. Immun.* 73:3358–3366. <http://dx.doi.org/10.1128/IAI.73.6.3358-3366.2005>.
  10. Stecher B, Barthel M, Schlumberger MC, Haberli L, Rabsch W, Kremer M, Hardt WD. 2008. Motility allows *S. Typhimurium* to benefit from the mucosal defence. *Cell. Microbiol.* 10:1166–1180. <http://dx.doi.org/10.1111/j.1462-5822.2008.01118.x>.
  11. Rivera-Chávez F, Winter SE, Lopez CA, Xavier MN, Winter MG, Nuccio SP, Russell JM, Laughlin RC, Lawhon SD, Sterzenbach T, Bevins CL, Tsolis RM, Harshey R, Adams LG, Bäuml AJ. 2013. *Salmonella* uses energy taxis to benefit from intestinal inflammation. *PLoS Pathog.* 9:e1003267. <http://dx.doi.org/10.1371/journal.ppat.1003267>.
  12. Lockman HA, Curtiss R, III. 1990. *Salmonella* Typhimurium mutants lacking flagella or motility remain virulent in BALB/c mice. *Infect. Immun.* 58:137–143.
  13. Lockman HA, Curtiss R, III. 1992. Virulence of non-type 1-fimbriated and nonfimbriated nonflagellated *Salmonella* Typhimurium mutants in murine typhoid fever. *Infect. Immun.* 60:491–496.
  14. Fricke WF, Mammel MK, McDermott PF, Tartera C, White DG, Leclerc JE, Ravel J, Cebula TA. 2011. Comparative genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sub-lineage evolution. *J. Bacteriol.* 193:3556–3568. <http://dx.doi.org/10.1128/JB.00297-11>.
  15. Tsolis RM, Adams LG, Ficht TA, Bäuml AJ. 1999. Contribution of *Salmonella* Typhimurium virulence factors to diarrheal disease in calves. *Infect. Immun.* 67:4879–4885.
  16. Winter SE, Thiennimitr P, Winter MG, Butler BP, Huseby DL, Crawford RW, Russell JM, Bevins CL, Adams LG, Tsolis RM, Roth JR, Bäuml AJ. 2010. Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* 467:426–429. <http://dx.doi.org/10.1038/nature09415>.
  17. Lopez CA, Winter SE, Rivera-Chávez F, Xavier MN, Poon V, Nuccio SP, Tsolis RM, Bäuml AJ. 2012. Phage-mediated acquisition of a type III secreted effector protein boosts growth of salmonella by nitrate respiration. *mBio* 3(3):e00143-12. <http://dx.doi.org/10.1128/mBio.00143-12>.
  18. Fischbach MA, Sonnenburg JL. 2011. Eating for two: how metabolism establishes interspecies interactions in the gut. *Cell Host Microbe* 10:336–347. <http://dx.doi.org/10.1016/j.chom.2011.10.002>.
  19. Maier RJ, Olczak A, Maier S, Soni S, Gunn J. 2004. Respiratory hydrogen use by *Salmonella enterica* serovar Typhimurium is essential for virulence. *Infect. Immun.* 72:6294–6299. <http://dx.doi.org/10.1128/IAI.72.11.6294-6299.2004>.
  20. Maier L, Vyas R, Cordova CD, Lindsay H, Schmidt TS, Brugiroux S, Periaswamy B, Bauer R, Sturm A, Schreiber F, von Mering C, Robinson MD, Stecher B, Hardt WD. 2013. Microbiota-derived hydrogen fuels *Salmonella* Typhimurium invasion of the gut ecosystem. *Cell Host Microbe* 14:641–651. <http://dx.doi.org/10.1016/j.chom.2013.11.002>.
  21. Thiennimitr P, Winter SE, Winter MG, Xavier MN, Tolstikov V, Huseby DL, Sterzenbach T, Tsolis RM, Roth JR, Bäuml AJ. 2011. Intestinal inflammation allows *Salmonella* to use ethanolamine to compete with the microbiota. *Proc. Natl. Acad. Sci. U. S. A.* 108:17480–17485. <http://dx.doi.org/10.1073/pnas.1107857108>.
  22. 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.* 183:2463–2475. <http://dx.doi.org/10.1128/JB.183.8.2463-2475.2001>.
  23. Deatherage Kaiser BL, Li J, Sanford JA, Kim YM, Kronewitter SR, Jones MB, Peterson CT, Peterson SN, Frank BC, Purvine SO, Brown JN, Metz TO, Smith RD, Heffron F, Adkins JN. 2013. A multi-omic view of host-pathogen-commensal interplay in *Salmonella*-mediated intestinal infection. *PLoS One* 8:e67155. <http://dx.doi.org/10.1371/journal.pone.0067155>.
  24. Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, Gopinath S, Naidu N, Choudhury B, Weimer BC, Monack DM, Sonnenburg JL. 2013. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* 502:96–99. <http://dx.doi.org/10.1038/502S96a>.
  25. Craig M, Sadik AY, Golubeva YA, Tidhar A, Schlauch JM. 2013. Twin-arginine translocation system (*tat*) mutants of *Salmonella* are attenuated due to envelope defects, not respiratory defects. *Mol. Microbiol.* 89:887–902. <http://dx.doi.org/10.1111/mmi.12318>.
  26. Muller L. 1923. Un nouveau milieu d'enrichissement pour la recherche du Bacille Typhique at Paratyphique. *Comp. Rend. Soc. Biol.* 89:434–437.
  27. Krumwiede C, Kohn LA. 1917. A triple-sugar modification of the Russell double-sugar medium. *J. Med. Res.* 37:225–227.
  28. Wilson WJ. 1923. Reduction of sulphites by certain bacteria in media containing a fermentable carbohydrate and metallic salts. *J. Hyg.* 21:392–398. <http://dx.doi.org/10.1017/S0022172400031594>.
  29. Malorny B, Bunge C, Helmuth R. 2003. Discrimination of d-tartrate-fermenting and -nonfermenting *Salmonella enterica* subsp. *enterica* isolates by genotypic and phenotypic methods. *J. Clin. Microbiol.* 41:4292–4297. <http://dx.doi.org/10.1128/JCM.41.9.4292-4297.2003>.
  30. Edwards PR, Moran AB. 1945. *Salmonella* cultures which resemble the Sendai type. *J. Bacteriol.* 50:257–260.
  31. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, Goodhead I, Rance R, Baker S, Maskell DJ, Wain J, Dolecek C, Achtman M, Dougan G. 2008. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nat. Genet.* 40:987–993. <http://dx.doi.org/10.1038/ng.195>.
  32. Uzzau S, Brown DJ, Wallis T, Rubino S, Leori G, Bernard S, Casadesús J, Platt DJ, Olsen JE. 2000. Host adapted serotypes of *Salmonella enterica*. *Epidemiol. Infect.* 125:229–255. <http://dx.doi.org/10.1017/S0950268899004379>.

33. Chen HD, Jewett MW, Groisman EA. 2012. An allele of an ancestral transcription factor dependent on a horizontally acquired gene product. *PLoS Genet.* 8:e1003060. <http://dx.doi.org/10.1371/journal.pgen.1003060>.
34. Winter SE, Winter MG, Thiennimitr P, Gerriets VA, Nuccio SP, Rüssmann H, Bäumlér AJ. 2009. The TviA auxiliary protein renders the *Salmonella enterica* serotype Typhi RcsB regulon responsive to changes in osmolarity. *Mol. Microbiol.* 74:175–193. <http://dx.doi.org/10.1111/j.1365-2958.2009.06859.x>.
35. Winter SE, Winter MG, Godinez I, Yang HJ, Russmann H, Andrews-Polymenis HL, Baumlér AJ. 2010. A rapid change in virulence gene expression during the transition from the intestinal lumen into tissue promotes systemic dissemination of *Salmonella*. *PLoS Pathog.* 6:e1001060. <http://dx.doi.org/10.1371/journal.ppat.1001060>.
36. Murray GL, Attridge SR, Morona R. 2003. Regulation of *Salmonella* Typhimurium lipopolysaccharide O antigen chain length is required for virulence; identification of FepE as a second Wzz. *Mol. Microbiol.* 47:1395–1406. <http://dx.doi.org/10.1046/j.1365-2958.2003.03383.x>.
37. Crawford RW, Keestra AM, Winter SE, Xavier MN, Tsois RM, Tolstikov V, Bäumlér AJ. 2012. Very long O-antigen chains enhance fitness during *Salmonella*-induced colitis by increasing bile resistance. *PLoS Pathog.* 8:e1002918. <http://dx.doi.org/10.1371/journal.ppat.1002918>.
38. Crawford RW, Wangdi T, Spees AM, Xavier MN, Tsois RM, Bäumlér AJ. 2013. Loss of very-long O-antigen chains optimizes capsule-mediated immune evasion by *Salmonella enterica* serovar Typhi. *mBio* 4(4):e00232-13. <http://dx.doi.org/10.1128/mBio.00232-13>.
39. Lawley TD, Bouley DM, Hoy YE, Gerke C, Relman DA, Monack DM. 2008. Host transmission of *Salmonella enterica* serovar Typhimurium is controlled by virulence factors and indigenous intestinal microbiota. *Infect. Immun.* 76:403–416. <http://dx.doi.org/10.1128/IAI.01189-07>.
40. Raffatellu M, Santos RL, Chessa D, Wilson RP, Winter SE, Rossetti CA, Lawhon SD, Chu H, Lau T, Bevins CL, Adams LG, Bäumlér AJ. 2007. The capsule encoding the *viaB* locus reduces interleukin-17 expression and mucosal innate responses in the bovine intestinal mucosa during infection with *Salmonella enterica* serotype Typhi. *Infect. Immun.* 75:4342–4350. <http://dx.doi.org/10.1128/IAI.01571-06>.
41. Haneda T, Winter SE, Butler BP, Wilson RP, Tükel C, Winter MG, Godinez I, Tsois RM, Bäumlér AJ. 2009. The capsule-encoding *viaB* locus reduces intestinal inflammation by a *Salmonella* pathogenicity island 1-independent mechanism. *Infect. Immun.* 77:2932–2942. <http://dx.doi.org/10.1128/IAI.00172-09>.
42. Stone WS. 1912. The medical of chronic typhoid infection (typhoid bacillus carriers). *Am. J. Med. Sci.* 143:544–557. <http://dx.doi.org/10.1097/0000441-191204000-00009>.
43. Shivaprasad HL. 2000. Fowl typhoid and pullorum disease. *Rev. Sci. Tech.* 19:405–424.
44. Nielsen LR, Schukken YH, Gröhn YT, Ersbøll AK. 2004. *Salmonella* Dublin infection in dairy cattle: risk factors for becoming a carrier. *Prev. Vet. Med.* 65:47–62. <http://dx.doi.org/10.1016/j.prevetmed.2004.06.010>.
45. Thomson NR, Clayton DJ, Windhorst D, Vernikos G, Davidson S, Churcher C, Quail MA, Stevens M, Jones MA, Watson M, Barron A, Layton A, Pickard D, Kingsley RA, Bignell A, Clark L, Harris B, Ormond D, Abdallah Z, Brooks K, Cherevach I, Chillingworth T, Woodward J, Norberczak H, Lord A, Arrowsmith C, Jagels K, Moule S, Mungall K, Sanders M, Whitehead S, Chabalgoity JA, Maskell D, Humphrey T, Roberts M, Barrow PA, Dougan G, Parkhill J. 2008. 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. <http://dx.doi.org/10.1101/gr.077404.108>.
46. Deng W, Liou SR, Plunkett G, III, Mayhew GF, Rose DJ, Burland V, Kodoyianni V, Schwartz DC, Blattner FR. 2003. Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J. Bacteriol.* 185:2330–2337. <http://dx.doi.org/10.1128/JB.185.7.2330-2337.2003>.
47. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller W, Stoneking T, Nhan M, Waterston R, Wilson RK. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 413:852–856. <http://dx.doi.org/10.1038/35101614>.
48. Holt KE, Thomson NR, Wain J, Langridge GC, Hasan R, Bhutta ZA, Quail MA, Norberczak H, Walker D, Simmonds M, White B, Bason N, Mungall K, Dougan G, Parkhill J. 2009. Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars Paratyphi A and Typhi. *BMC Genomics* 10:36. <http://dx.doi.org/10.1186/1471-2164-10-S3-S36>.
49. Chiu CH, Su LH, Chu C. 2004. *Salmonella enterica* serotype Choleraesuis: epidemiology, pathogenesis, clinical disease, and treatment. *Clin. Microbiol. Rev.* 17:311–322. <http://dx.doi.org/10.1128/CMR.17.2.311-322.2004>.
50. Liu WQ, Feng Y, Wang Y, Zou QH, Chen F, Guo JT, Peng YH, Jin Y, Li YG, Hu SN, Johnston RN, Liu GR, Liu SL. 2009. *Salmonella* Paratyphi C: genetic divergence from *Salmonella* Choleraesuis and pathogenic convergence with *Salmonella* Typhi. *PLoS One* 4:e4510. <http://dx.doi.org/10.1371/journal.pone.0004510>.
51. Magrane M, UniProt Consortium. 2011. UniProt Knowledgebase: a hub of integrated protein data. Database 2011:bar009. doi:10.1093/database/bar009. PubMed.
52. Maglott D, Ostell J, Pruitt KD, Tatusova T. 2005. Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res.* 33:D54–D58. <http://dx.doi.org/10.1093/nar/gni052>.
53. Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, Arnold R, Rattei T, Letunic I, Doerks T, Jensen LJ, von Mering C, Bork P. 2012. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* 40:D284–D289. <http://dx.doi.org/10.1093/nar/gkr1060>.
54. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301. <http://dx.doi.org/10.1093/nar/gkr1065>.
55. Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families. *Nucleic Acids Res.* 31:371–373. <http://dx.doi.org/10.1093/nar/gkg128>.
56. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402. <http://dx.doi.org/10.1093/nar/25.17.3389>.
57. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34:D32–D36. <http://dx.doi.org/10.1093/nar/gkj014>.
58. Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* 32:W327–W331. <http://dx.doi.org/10.1093/nar/gkh454>.
59. Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28:27–30. <http://dx.doi.org/10.1093/nar/28.7.e27>.
60. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muñoz-Rascado L, Ong Q, Paley S, Schröder I, Shearer AG, Subhraveti P, Travers M, Weerasinghe D, Weiss V, Collado-Vides J, Gunsalus RP, Paulsen I, Karp PD. 2013. EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* 41:D605–D612. <http://dx.doi.org/10.1093/nar/gks1027>.
61. Nuccio SP, Bäumlér AJ. 2007. Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiol. Mol. Biol. Rev.* 71:551–575. <http://dx.doi.org/10.1128/MMBR.00014-07>.
62. Haraga A, Ohlson MB, Miller SI. 2008. Salmonellae interplay with host cells. *Nat. Rev. Microbiol.* 6:53–66. <http://dx.doi.org/10.1038/nrmicro1788>.
63. Frye J, Karlinsey JE, Felise HR, Marzolf B, Dowidar N, McClelland M, Hughes KT. 2006. Identification of new flagellar genes of *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* 188:2233–2243. <http://dx.doi.org/10.1128/JB.188.6.2233-2243.2006>.