**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.

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 comparison 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-
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, fluE, 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 ≥ 4) in members of one group but rarely (n ≤ 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 ~ 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
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 (H₂), 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 B₁₂ (cbi and cob genes), a cofactor produced under anaerobic condition, which is required for ethanolamine utilization (22) (Fig. 2).

Vitamin B₁₂ is also necessary for the utilization of 1,2-propanediol, a catabolite produced by microbes fermenting fu- cose or rhamnose. Expression of S. Typhimurium proteins involved in sugar catabolism is increased in the intestinal lumen in a
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 of 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 β-oxidation of fatty acids was overrepresented in genomes representing the extraintestinal pathovar. This pathway, which is distinct from the aerobic β-oxidation pathway for fatty acid degradation, is encoded by the *ydhFO, 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 (Fig. 1).

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).
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 tetraionate 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 sulfate 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 pathogen 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 supplementary material) (32) and our inclusion of a similar broad assortment of gastrointestinal serovars (see Fig. S1), suggest that our data set is suitable 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 likelihood 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 supplementary material) (2, 6, 14, 45–50), we gath-
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 ≥10 codons disrupted by the aforementioned mutations relative to a reference CDS. An alignment in the same genomic context with ≥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 ≤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 partially 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 ≥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 $y = 0.9945x + 6 	imes 10^{-16}$, $R^2 = 0.9902$). With these values, we computed the $z$ scores (before = 0.945, after = 8.598) and one-tailed $P$ values (0.172, −0.0) 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. Tsolis for helpful suggestions on the manuscript.

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

**REFERENCES**

1. Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, Sebaihia M, James KD, Churcher C, Mungall KL, Baker S, Basham D, Bentley SD, Brooks K, Cerdeño-Tarraga AM, Chillingworth T,
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, Thienemann P, Gerretts VA, Nuccio SP, Rüßmann H, Bäumler AJ. 2009. The TviA auxiliary protein renders the Salmonella enterica serovar Typhimurium RcaB regulon responsive to changes in osmolality. 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 H, Bäumler AJ. 2010. A rapid change in virulence gene expression during Salmonella-induced colitis by increasing bile resistance. PLoS Pathog. 6:e1000160. http://dx.doi.org/10.1371/journal.ppat.1000160.

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.1365-2958.2003.03585.x.

37. Crawford RW, Keestra AM, Winter SE, Xavier MN, Tsolis RM, Tolstikov V, Bäumler 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, Tsolis RM, Bäumler 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. Lowley TD, Boyle DM, Hoy YE, Gerke C, Reiman 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, Chen HD, Jewett MW, Groisman EA. 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/10.1111/j.1365-2958.2003.06859.x.

41. Nuccio and Bäumler

42. 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/10.1111/j.1365-2958.2003.03585.x.

43. Crawford RW, Keestra AM, Winter SE, Xavier MN, Tsolis RM, Tolstikov V, Bäumler 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.

44. Crawford RW, Wangdi T, Spees AM, Xavier MN, Tsolis RM, Bäumler 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.

45. Lowley TD, Boyle DM, Hoy YE, Gerke C, Reiman 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.

46. Raffatellu M, Santos RL, Chessa D, Wilson RP, Winter SE, Rossetti CA, Chen HD, Jewett MW, Groisman EA. 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/10.1111/j.1365-2958.2003.06859.x.

47. Nuccio and Bäumler

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, Thienemann P, Gerretts VA, Nuccio SP, Rüßmann H, Bäumler AJ. 2009. The TviA auxiliary protein renders the Salmonella enterica serovar Typhimurium RcaB regulon responsive to changes in osmolality. 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 H, Bäumler AJ. 2010. A rapid change in virulence gene expression during Salmonella-induced colitis by increasing bile resistance. PLoS Pathog. 6:e1000160. http://dx.doi.org/10.1371/journal.ppat.1000160.

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/10.1111/j.1365-2958.2003.03585.x.

37. Crawford RW, Keestra AM, Winter SE, Xavier MN, Tsolis RM, Tolstikov V, Bäumler 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, Tsolis RM, Bäumler 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. Lowley TD, Boyle DM, Hoy YE, Gerke C, Reiman 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, Chen HD, Jewett MW, Groisman EA. 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/10.1111/j.1365-2958.2003.06859.x.

41. Nuccio and Bäumler

42. 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/10.1111/j.1365-2958.2003.03585.x.