The Journal of Infectious Diseases

MAJOR ARTICLE

Tracing Origins of the *Salmonella* Bareilly Strain Causing a Food-borne Outbreak in the United States

Maria Hoffmann,[1,3] Yan Luo,[2] Steven R. Monday,[1] Narjol Gonzalez-Escalona,[1] Andrea R. Ottesen,[1] Tim Muruvanda,[1] Charles Wang,[1] George Kastanis,[1] Christine Keys,[1] Daniel Janies,[5] Izzet F. Senturk,[6] Umit V. Catalyurek,[6] Hua Wang,[1] Thomas S. Hammack,[1] William J. Wolfgang,[7] Dianna Schoonmaker-Bopp,[7] Alvina Chu,[4] Robert Myers,[4] Julie Haendiges,[4] Peter S. Evans,[1] Jianghong Meng,[3] Errol A. Strain,[2] Marc W. Allard,[1] and Eric W. Brown[1]

[1]Division of Microbiology, Office of Regulatory Science, Center for Food Safety and Nutrition, and [2]Division of Public Health and Biostatistics, Office of Food Defense, Communication and Emergency Response, Center for Food Safety and Nutrition, US Food and Drug Administration, College Park, and [3]Department of Nutrition & Food Science and Joint Institute for Food Safety & Applied Nutrition, University of Maryland, College Park, and [4]Maryland Department of Health and Mental Hygiene, Baltimore; [5]Department of Bioinformatics and Genomics, University North Carolina at Charlotte; [6]Department of Biomedical Informatics, Ohio State University, Columbus; and [7]New York State Department of Health, Wadsworth Center, Albany

(See the editorial commentary by Dunn on pages 499–501.)

Background. Using a novel combination of whole-genome sequencing (WGS) analysis and geographic metadata, we traced the origins of *Salmonella* Bareilly isolates collected in 2012 during a widespread food-borne outbreak in the United States associated with scraped tuna imported from India.

Methods. Using next-generation sequencing, we sequenced the complete genome of 100 *Salmonella* Bareilly isolates obtained from patients who consumed contaminated product, from natural sources, and from unrelated historically and geographically disparate foods. Pathogen genomes were linked to geography by projecting the phylogeny on a virtual globe and produced a transmission network.

Results. Phylogenetic analysis of WGS data revealed a common origin for outbreak strains, indicating that patients in Maryland and New York were infected from sources originating at a facility in India.

Conclusions. These data represent the first report fully integrating WGS analysis with geographic mapping and a novel use of transmission networks. Results showed that WGS vastly improves our ability to delimit the scope and source of bacterial food-borne contamination events. Furthermore, these findings reinforce the extraordinary utility that WGS brings to global outbreak investigation as a greatly enhanced approach to protecting the human food supply chain as well as public health in general.

Keywords. salmonellosis; geographic information systems; next generation sequencing; single nucleotide polymorphism; traceback.

The global food network poses challenges to traceback of pathogens causing food-borne illnesses; a single meal can contain ingredients from multiple regions, and food products are shipped around the world. For example, public health authorities investigated a multistate outbreak of *Salmonella enterica* subsp. *enterica* serovar Bareilly and *Salmonella* Nchanga infections with illness onset dates between 1 January and 7 July 2012. Of a total of 425 cases reported in 28 states and the District of Colombia, 410 (96.5%) were found to be caused by *Salmonella* Bareilly infection [1]. Collaborative investigations by state, local, and federal public health and regulatory agencies linked this outbreak to a frozen, raw yellowfin tuna product ("tuna scrape"). This product had been marketed as Nakaochi Scrape by an Indian corporation and used to make spicy tuna sushi for restaurants and grocery stores [1].

Received 21 January 2015; accepted 1 April 2015; published online 20 May 2015.

Correspondence: M. Hoffmann, 5100 Paint Branch Parkway, College Park, MD 20740 (maria.hoffman@fda.hhs.gov).

The Journal of Infectious Diseases® 2016;213:502–8

DOI: 10.1093/infdis/jiv297

First identified in India in 1928 [2], *Salmonella* Bareilly is known for its wide host range [3–5] and has been among the top 20 most frequently isolated serovars in clinical cases of salmonellosis [6] in the United States. Some strains are clonal and can be detected in numerous sites throughout Southeast Asia. These features make epidemiologic traceback difficult with conventional tools such as pulsed-field gel electrophoresis (PFGE) [7], which does not provide the level of resolution necessary to distinguish outbreak strains from clonally related *Salmonella* Bareilly isolates previously obtained from other sources.

Diagnostic capabilities have been greatly enhanced with the development and increasing deployment of whole-genome sequencing (WGS), which is now being applied frequently as a molecular epidemiologic tool to assist in investigations of disease outbreaks [8–10]. The accessibility of WGS for pathogen analysis has made it possible to sequence multiple isolates, construct phylogenies showing the extent to which outbreak and environmental isolates are related to each other, and then use those phylogenies to corroborate epidemiologic data [11–13].

Advances in both WGS and geographic information systems software now enable researchers to integrate phylogenetic trees (based on single-nucleotide polymorphism [SNP] data) with

the metadata associated with isolates (eg, latitude and longitude coordinates, and, where possible, date of collection) [14]. In a retrospective investigation, we demonstrate the value of WGS to delimit and source track an outbreak global in scale. By merging comparative genomic analysis with geographic mapping tools, we were able to generate a transmission graph that revealed where *Salmonella* Bareilly pathogens may have originated. Global location and the unique mutations of the serovar were then projected onto a virtual globe. These new technologies enabled traceback to the contamination source with a high degree of certainty and allowed us to better understand the *Salmonella* Bareilly pathogen transmission network.

MATERIALS AND METHODS

Bacterial Isolates

Our data set consisted of 100 *Salmonella* Bareilly isolates obtained from clinical, food, feed, and environmental sources (Supplementary Table 1). The isolates were chosen based on similarity or dissimilarity by PFGE to *Salmonella* Bareilly isolated from the outbreak in 2012. Of these isolates, 41 were collected during the investigation of the tuna scrape-associated outbreak in 2012. The New York Department of Health and Maryland Department of Health and Mental Hygiene provided 29 clinical isolates. Environmental isolates from raw tuna fish were collected during US Food and Drug Administration (FDA) inspections. Historical food strains of *Salmonella* Bareilly gathered from different countries between 1968 and 2012 were isolated from various food sources by the FDA both during routine inspections and as part of compliance actions related to earlier contamination events. Two additional *Salmonella* Bareilly clinical isolates were analyzed for comparison; these were collected from different outbreaks in Maryland during 2011.

DNA Preparation, PFGE, and Genome Sequencing

PFGE was performed according to the PulseNet protocol of the Centers for Disease Control and Prevention (http://www.cdc.gov/pulsenet/pathogens/index.html). Genomic DNA from each strain was isolated from overnight cultures using the DNeasy Blood & Tissue kit (Qiagen). A single isolate, *Salmonella* Bareilly CFSAN000189, was sequenced on the Pacific Biosciences (PacBio) *RS* II Sequencer and assembled as described elsewhere [15–17]. Most of the isolates (96 of 105) were sequenced using the Illumina MiSeq platform. Five were sequenced using an Ion Torrent (PGM) sequencing system (Life Technologies) with 200–base pair readings. The Illumina and Ion Torrent readings were assembled de novo using CLC Genomics Workbench software (version 6; CLC bio). Five isolates were shotgun sequenced with the Genome Sequencer FLX 454 Life Sciences (Roche). De novo assemblies were performed using the Newbler software package (version 2.6; Roche). All assembled genomes were annotated using the National Center for Biotechnology Information Prokaryotic Genome Automatic Annotation Pipeline [18].

Phylogenetic Analyses and Novel Concept of Transmission Networks

We used the closed *Salmonella* Bareilly genome as a reference, mapping raw readings from draft genomes to this reference genome and then building a SNP matrix from which a maximum likelihood (ML) tree was constructed. In this use case, we used our SNP tree for *Salmonella* Bareilly to generate a transmission graph and applied *betweenness centrality* (BC) on the generated graph to evaluate importance of isolate locales. BC is an indicator of the mathematical centeredness of a node (eg, place) within a larger network, taking into account the shortest paths from all points in the network that pass through that node. In the transmission graph, a node represents a geographic place of isolation for a pathogen, and the directed edge between 2 nodes represents a historical transmission link for the corresponding places of isolation. The number of transmissions is considered in determining the edge weights.

The BC is used to determine how many times a node appears on a shortest path between every pair of nodes. If the number of shortest paths between a pair of nodes $s - t$ is σ_{st}, and node v exists $\sigma_{st}(v)$ times in these shortest paths, then the BC of node v can be expressed as

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}.$$

Details regarding DNA sequencing, comparative and phylogenetic genome analyses, and geographic mapping and visualization are provided in the Supplementary Materials.

RESULTS

PFGE of the Outbreak Isolates

PFGE, using 2 restriction enzymes (*Xba*I and *Bln*I), was performed for all 41 isolates collected from the 2012 tuna outbreak. The 41 isolates produced *Xba*I and *Bln*I PFGE patterns JAPX01.0042 and JF6A26.0076, respectively. One (CFSAN000755) of the 2 clinical *Salmonella* Bareilly isolates (CFSAN000755 and CFSAN000754) collected from different outbreaks in Maryland in 2011 also shared the same PFGE pattern as the 2012 outbreak. Most notably, a total of 7 historical isolates unrelated to the outbreak were indistinguishable from the outbreak *Xba*I pattern.

Comparative Phylogenetic Genome Analysis

Our work provides 100 new draft genomes of *Salmonella* Bareilly strains, including the first publicly available, fully closed genome of this serovar. Variable SNPs, extracted from the WGS data, were subjected to comparative genomic analysis, yielding a single ML tree of 100 isolates (Figure 1). Data from this phylogenetic analysis provided critical information for effectively delineating members of the serovar Bareilly. First, the resultant ML tree partitioned *Salmonella* Bareilly into 4 distinct lineages (100% bootstrap support) separated by >10 000 SNPs and at least hundreds, if not thousands, of unique SNPs common to

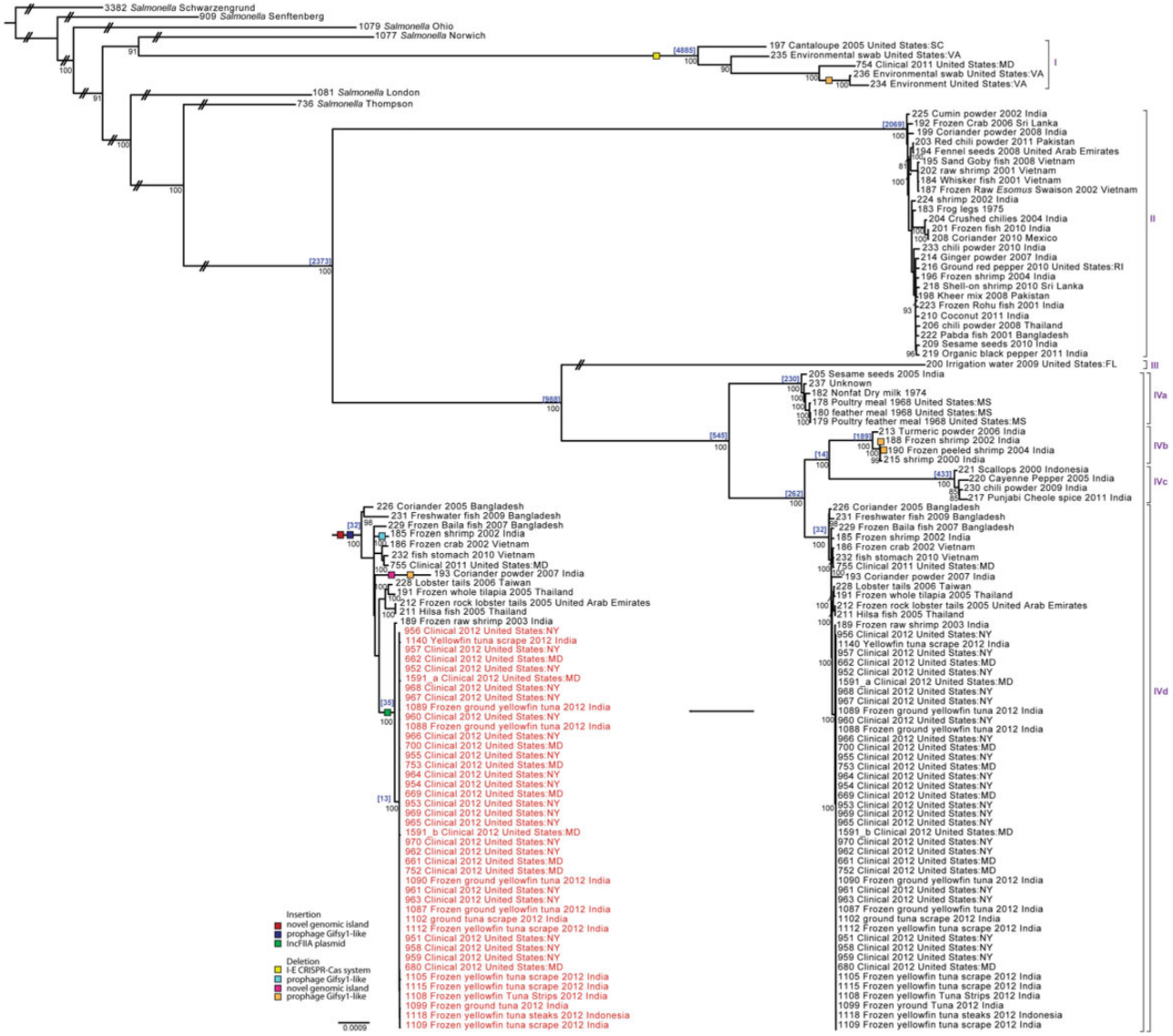

Figure 1. Maximum likelihood (ML) tree based on single-nucleotide polymorphism (SNP) analysis of 100 *Salmonella* Bareilly isolates. 106 597 variable SNPs, with 69 590 being informative, were found using SAMtools software (version 0.1.18) [20] followed by a custom pipeline. The ML tree was generated using GARLI software (version 2.0) [21] under the GTR + Γ model of nucleotide evolution and visualized using Figtree software (version 1.3.1). Parameter space was searched for the best tree, with simultaneous estimation for model parameters performed using a ML search. The best tree was identified from 200 runs on the nonbootstrapped data set. The numbers of unambiguous substitutions that mapped to the tree only once and are >0 are given above each node in blue. Measures of clade confidence are reported below each node in the form of bootstrap values (1000 iterations). Bootstrap values <70% were not shown. The tree was rooted with 6 outgroups including *Salmonella* Schwarzengrund CFSAN003382, *Salmonella* Ohio CFSAN001079, *Salmonella* Norwich CFSAN001077, *Salmonella* London CFSAN001081, *Salmonella* Thompson CFSAN00736, and *Salmonella* Senftenberg CFSAN000909. The taxa of source for each isolate, geographic location, and date were mapped onto the tree; prophage, insertion sequence; and plasmid observations are also depicted. Taxa shown in red were isolated during the outbreak investigation.

only an individual lineage. Lineage IV comprises 68 isolates and is by far the most comprehensive lineage. It is further divided into 4 sublineages (100% bootstrap support) separated by >1000 SNPs. Second, lineage I, which includes the 5 *Salmonella* Bareilly isolates, obtained from the East Coast of the United States, is more closely related to the outgroups than to any members of *Salmonella* Bareilly lineages II, III, and IV. Lineage I is also noteworthy in that was found to be devoid of the *cas* gene cluster, present in isolates from lineages II, III, and IV.

Third, *Salmonella* Bareilly isolates tended to cluster by geography in that those isolates derived in proximity to each other were also closely related in the tree. For example, isolates from lineages I and III were collected from environmental and clinical samples from the United States, whereas most isolates from lineages II and IV originate in Asia. Fourth, as was expected, isolates with matching PFGE profiles (ie, *Xba*I) are more closely related to each other in the tree than to *Salmonella* Bareilly isolates retaining disparate PFGE profiles. In this case, although

tens of thousands of SNPs could distinguish among the major lineages, <120 SNPs varied within any clonal PFGE pattern. Finally, *Salmonella* Bareilly isolates associated with the implicated food production facility and the clinical isolates from the 2012 outbreak clustered together (100% bootstrap support) by a mean distance of only 1 SNP (range, 0–6 SNPs). However, these isolates were phylogenetically distinct (with a mean of 96 SNPs differences) from the other 7 isolates that shared a common *Xba*I pattern. Of note, the reference genome from *Salmonella* Bareilly isolate CFSAN000189 was closest to the outbreak isolates, differing by a mean of only 20 SNPs and carrying the same IncFIIA plasmid. Pairwise SNP variation between and within the 4 lineages and the tree generated using the core SNP matrix (Supplementary Table 2 and Supplementary Figure 1) further support the phylogenetic partitions described above.

Geographic Mapping and the Novel Use of Transmission Networks for Genetic Data

To visualize the evolution and geographic transmission of *Salmonella* Bareilly taxon within and among different lineages, we used SUPRAMAP [19] to project the phylogenetic tree, derived from our SNP matrix, onto a virtual globe provided by Google Earth. (The KML file, suitable for interactive viewing in a browser after Google Earth plugin has been downloaded and installed, is available at http://minipointmap.herokuapp.com/mainpage/dispEarth.) The resulting map includes isolate-specific nodes that could be selected to obtain information about the isolation date, source, common ancestor(s), and mutations that were reconstructed along each branch.

To better understand the patterns of location and dissemination of *Salmonella* Bareilly, we used a novel concept to generate a transmission graph by using the ML tree shown in Figure 1 and applying BC on that generated graph (Figure 2). This analysis showed that India has the highest betweenness score, which provides a measure of its importance in the overall transmission network. India has high transmission frequency with several places, such as Bangladesh, Vietnam, Thailand, Pakistan, Sri Lanka, and the United States, suggesting that *Salmonella* Bareilly isolates from these countries often originated in India and are transported to these locations through contaminated food. Within the United States, the network also shows transmission frequencies with incoming and outgoing connections between South Carolina, Virginia, and Maryland. That is, outbreaks originating in any of these states have a higher potential of crossing into states that share betweenness. Moreover, akin to outbreak spread among states with shared betweenness, a moderate degree of betweenness and transmission frequency among Thailand, United Arab Emirates, Pakistan, and Vietnam were also noted. In contrast, the BC and transmission frequency are fairly low for all states in the United States, as well as for Mexico, Indonesia, and Taiwan.

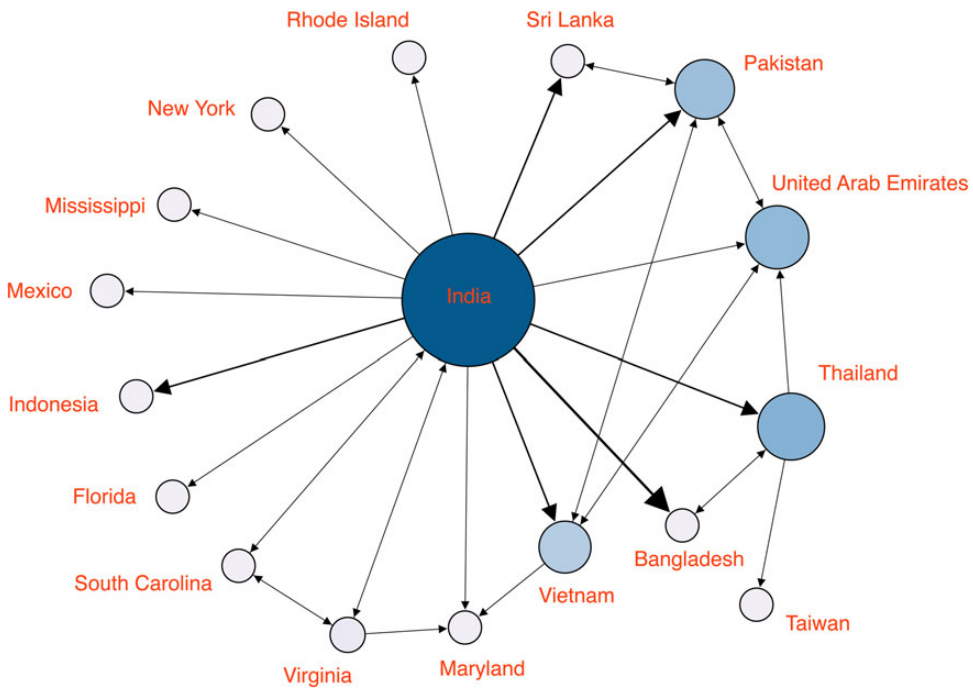

Figure 2. Transmission graph based on the *Salmonella* Bareilly isolates used in this study. The graph was generated by using our pangenome tree for *Salmonella* Bareilly and applying betweenness centrality on the generated graph (method described in Supplementary Materials). A node represents geographic location of isolation for *Salmonella* Bareilly isolates. The concept of distance between places is related to historical transmission links as indicated by character evolution on a phylogeny. The size and darkness of the spheres represent the relative importance of geographic places in spreading a pathogen, in terms of the betweenness metric, and the thickness of the lines and size of the arrows represent the frequency of historical transmissions.

Relevant SNP Differences and Variable Core Genes Among *Salmonella* Bareilly

Supplementary Table 3 lists variable genes having unique nucleotide differences that defined specific clades and their SNP change. Sublineage IVd, which clustered together isolates with 5 *Xba*I patterns (JAPX01.0100, JAPX01.0084, JAPX01.0649, JAPX01.0158, and JAPX01.0042) has 32 unique SNPs, 16 nonsynonymous, located on 27 genes. These SNPs may be suitable for identifying other strains carrying the lineage IVd strain profile. This is important given that sublineage IVd includes clinical isolates obtained from the United States and, as such, other isolates carrying traits identical to those of the clinical isolates that may also pose certain threats to public health. Thirteen unique SNPs, 9 nonsynonymous, were found on 10 genes and are present in only the outbreak isolates. These SNPs were most often located on core genes, such as the 2 nonsynonymous SNPs found on carbohydrate kinase and sulfate/proton transporter genes.

Of 1571 core genes detected, 1524 vary among 100 *Salmonella* Bareilly isolates. These core genes, along with the number of SNPs, haplotypes, and haplotype diversity are listed in Supplementary Table 4. In addition, a histogram comparing isolates, based on the number of SNPs according to the number of core genes (Supplementary Figure 2*A*) and the haplotype diversity is provided according to the number of core genes (Supplementary Figure 2*B*). A remarkably large number of mutation "hotspot" genes, for example 50S ribosomal subunit L7/12 (143 SNPs), bifunctional phosphoribosylaminoimidazo (117 SNPs), glycine dehydrogenase (116 SNPs), fimbrial outer membrane usher *staF* (97 SNPs), and nitrite reductase subunit *nirD* (93 SNPs) were found to have an enormous number of SNPs. In contrast, haplotype diversity was very low, with only a few types (2–12) noted among the 100 isolates reported here.

DISCUSSION

Using Genomics to Identify the Source of an Outbreak

Public health authorities were informed in 2012 of an outbreak involving *Salmonella* Bareilly that resulted from consumption of contaminated product containing raw yellowfin tuna. WGS analyses separated clinical isolates and the isolates collected from the potential contamination sources by FDA inspectors from other clonal *Salmonella* Bareilly isolates that had previously been indistinguishable by PFGE using *Xba*I restriction. All of the *Salmonella* Bareilly strains associated with the 2012 outbreak could be traced to a fishery facility in India. These results confirm that the clinical and tuna scrape isolates associated with the 2012 outbreak did, in fact, originate from the same source and are distinguished from other strains grouped with them by PFGE. SNP data easily segregated the 2012 outbreak isolates from the clinical isolate CFSAN000755 (same *Xba*I PFGE pattern as the 2012 outbreak), obtained in 2011, with a mean difference of 117 SNPs, whereas only 1–6 SNP differences were found between the 2012 outbreak isolates.

The results verify these epidemiologic data and indicate that patients in the United States became infected with *Salmonella* Bareilly from tuna scrape imported from a fishery in India for making spicy tuna sushi. Remarkably, one Maryland resident, who spent the entire incubation period vacationing in Thailand, became clinically ill on his return to Maryland. The strain isolated from this patient, CFSAN001591, clustered together with the other outbreak isolates and had the same 13 SNPs unique to this particular outbreak-associated group. This finding supports the belief that the patient also became ill because of ingesting contaminated tuna that presumably was imported by the same fishery into Thailand at about the same time. With divergence of only a few SNPs, the closest neighbor to the outbreak clade was CFSAN000189, obtained in 2003. Remarkably, this strain originated from another site in India located approximately 8 km from the implicated tuna facility, a fact that clearly documents the source-tracking capability of WGS.

We obtained 2 clinical *Salmonella* Bareilly isolates (CFSAN000754, and CFSAN000755) from the Maryland Department of Health and Mental Hygiene from 2011 outbreaks. Isolate CFSAN000754 clustered together with environmental/food isolates from the East Coast in lineage I and was only distantly related to isolate CFSAN000755, which instead clustered together with food isolates obtained from Asia in sublineage IVd. These 2 isolates are associated with 2 separate outbreaks, and it is noteworthy that WGS was able to distinguish and assign them to their appropriate home lineages. Because no Asian-derived isolates cluster in lineage I, one patient most likely became infected with *Salmonella* Bareilly after consuming a food product from the East Coast, and the other was most likely infected by eating imported food from Asia (similar to the cases associated with the 2012 outbreak isolates). However, a vehicle was never established, making it impossible to rule out a nonfood source for these infections, given potential reservoirs associated with *Salmonella* Bareilly that include reptiles and other environmental sources [22]. Moreover, owing to the limited number of isolates in lineage I, more sampling would be required to further support the hypothesis that these two patients became ill from ingesting contaminated food. Nevertheless, our analyses demonstrate that additional sequence variation present in the genome can reveal the geographically distinct origins of isolates. Collectively, these facts illustrate the strength of emerging high-throughput DNA sequencing technology for characterizing the dynamics of agent dissemination, as well as providing rapid traceback of clinical to food and environmental sources during a food-borne contamination event.

Geographic Mapping and Visualization

We introduce here several new strategies that offer both efficient and effective means by which to better present WGS data to public health investigators. By projecting the phylogeny onto a globe using SUPRAMAP, an interactive map can be produced

for exploring connections between outbreak cases and the mutations that occur along each branch in the tree. Transmission network analysis of contact investigation data has been used to elucidate the nationwide transmission dynamics of *Mycobacterium tuberculosis* [23]. In the current study, a combination of phylogenetic character evolution derived from WGS data and geographic metadata was used to create a transmission network. Such networks could be used to identify geographic locations important for historical transmission and spread of *Salmonella* Bareilly and can enable food producers and public health officials to target interventions at places where they may have the most impact for improving food safety. Our findings demonstrate that India had the highest betweenness score and high transmission frequency to several countries and has therefore probably been central to several transmission events associated with this particular pathogen.

Public health agencies concerned with *Salmonella* Bareilly contamination could use this information to determine most effective intervention points to minimize or eliminate outbreak risk. Mitigating actions in individual states with low betweenness values may alleviate a single outbreak at the local level but will do little to alleviate the problem at a global level. Rather, contamination events could be mediated at locations that exhibit high levels of betweenness, which should go far to disrupt and ultimately break the entire transmission network. Moreover, countries sharing betweenness may retain infection "bridges," spreading *Salmonella* Bareilly from one country to another. Although this possibility should be considered, it is important to remember that these predictions are based on a low sample number. Larger studies with more isolates are needed to fully understand the role of these countries in *Salmonella* Bareilly spread with respect to this outbreak. Nonetheless, transmission networks analysis performed in conjunction with the global visualization of a *Salmonella* outbreak phylogeny offers practical benefits for determining transmission patterns across geographic regions.

Phylogenetic Analyses

Only a few mutation hot-spot genes were found to have a disproportional number of SNPs, and they might serve as targets useful for rapidly subtyping serovar Bareilly. Although these genes have several SNPs, the haplotype diversity is fairly low, indicating there are only a few unique sequences among the servoar. Therefore, some of these genes, involved in a variety of different functions, such as metabolic processes, DNA replication and repair, cell division, transcription, and virulence, might be useful for subtyping the serovar.

In addition, tree topology and its associated SNP distances strongly suggest that *Salmonella* Bareilly formed a paraphyletic group. Lineage I, comprising 5 isolates, had a mean SNP distance of 36 018 to the remaining 95 isolates, more distant than even some of the outgroups used in the analysis. Comparing lineage I isolates with those of lineages II, III and IV, it becomes apparent that a unique ancestor common to all of these isolates is lacking. This observation suggests that *Salmonella* Bareilly evolved at least twice in independent events giving rise to the paraphyletic structure observed here. Interestingly, since WGS has become more widely used, such phenomena have also been seen in other *Salmonella* serovars, including *Salmonella* Newport and *Salmonella* Saintpaul [24, 25].

The study also demonstrated that all isolates in lineage I are devoid of the 8 *cas* genes (*cas3-cse1-cse2-cas7-cas5-cas6e-cas1-cas2*) comprising and encoding the Type I-E CRISPR-Cas system that forms adaptive immune systems in bacteria to combat phages, plasmids, or other mobile foreign DNA [26]. Because CRISPR-Cas loci inhibit the bacteria from acquiring plasmids or free DNA (which could carry genes that might enhance the survival of these bacteria) [27], lineage I *Salmonella* Bareilly isolates may have lost their Type I-E CRISPR-Cas system as a consequence of the East Coast environment. Such loss could provide selective advantages because these strains are able to attain and carry the horizontally transmitted genetic elements necessary for *Salmonella* Bareilly to persist in this particular niche.

Next-generation WGS of food-borne pathogens is revealing a bounty of genotypic information about the adaptive fitness and persistence of these organisms in their natural environment, the contribution of these data for delimiting the scope and traceback of bacterial outbreaks associated with the food supply cannot be overstated. In the current study case, our data strongly suggest that WGS combined with geographic metadata clearly improves source tracking and surveillance, which is critically important for protection of the public health. Given the important and new capabilities proffered by this technology, the FDA Foods Program has established the GenomeTrakr WGS network and database to provide more draft genomes to support future outbreak investigations. Currently, the database includes >12 000 food-borne pathogens that are publicly available and can be accessed by researchers and public health officials for real-time comparisons [28, 29]. Such tools further support more widespread deployments of WGS technology in tracing global food-borne contamination events back to their source.

Supplementary Data

Supplementary materials are available at http://jid.oxfordjournals.org. Consisting of data provided by the author to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the author, so questions or comments should be addressed to the author.

Notes

Acknowledgments. We thank Ruth Timme and Justin Payne for submitting the genomes to National Center for Biotechnology Information, and we thank Lili Fox Vélez for editorial support.

Financial support. This work was supported by an appointment of MH by the Joint Institute for Food Safety and Applied Nutrition, University of Maryland, College Park (appointment to M. H.) and the Centers for Disease Control and Prevention, National Center for Infectious Disease, Epidemiology and Laboratory Capacity for Infectious Diseases Cooperative Agreement (grant U50/CCU223671 to W. J. W. and D. S. B.). The efforts for the geographic mapping and the novel concept of transmission network were funded by the Defense Threat Reduction Agency (contract HDTRA1-14-C-0007).

Potential conflicts of interest. All authors: No potential conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. **2012**. Multistate outbreak of *Salmonella* Bareilly and *Salmonella* Nchanga infections associated with a raw scraped ground tuna product (final update). http://www.cdc.gov/salmonella/bareilly-04-12/index.html?s_cid=cs_654. Accessed 11 June 2015.
2. Bridges RF, Scott WM. A new organism causing paratyphoid fever in India. J Roy Army Med Corps **1931**; 56:241–9.
3. Saxena SN, Jayasheela M, Mago ML, John PC, Kumari N, Sharma NC. *Salmonella* serotypes in India, 1982–83. Indian J Pathol Microbiol **1988**; 31:286–97.
4. Cleary P, Browning L, Coia J, et al. A foodborne outbreak of *Salmonella* Bareilly in the United Kingdom, 2010. Euro Surveill **2010**; 15:pii:19732.
5. Sharma NC, John PC, Mago ML, Saxena SN. Phage-typing scheme of *Salmonella* bareilly based on lysogeny. Antonie Van Leeuwenhoek **1984**; 50:275–9.
6. Centers for Disease Control and Prevention. National Enteric Disease Surveillance: *Salmonella* annual report. Atlanta, GA: Centers for Disease Control and Prevention, **2011**. http://www.cdc.gov/ncezid/dfwed/PDFs/salmonella-annual-report-2011-508c.pdf.
7. den Bakker HC, Allard MW, Bopp D, et al. Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar enteritidis. Emerg Infect Dis **2014**; 20:1306–14.
8. Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med **2014**; 370:2408–17.
9. Eyre DW, Wilcox MH, Walker AS. Diverse sources of *C. difficile* infection. N Engl J Med **2014**; 370:183–4.
10. Chin CS, Sorenson J, Harris JB, et al. The origin of the Haitian cholera outbreak strain. N Engl J Med **2011**; 364:33–42.
11. Lienau EK, Strain E, Wang C, et al. Identification of a salmonellosis outbreak by means of molecular sequencing. N Engl J Med **2011**; 364:981–2.
12. Hoffmann M, Zhao S, Pettengill J, et al. Comparative genomic analysis and virulence differences in closely related *Salmonella enterica* serotype Heidelberg isolates from humans, retail meats, and animals. Genome Biol Evol **2014**; 6:1046–68.
13. Salter SJ. The food-borne identity. Nat Rev Microbiol **2014**; 12:533.
14. Janies DA, Treseder T, Alexandrov B, et al. The Supramap project: linking pathogen genomes with geography to fight emergent infectious diseases. Cladistics **2011**; 27:61–6.
15. Hoffmann M, Muruvanda T, Allard MW, et al. Complete genome sequence of a multidrug-resistant *Salmonella enterica* serovar Typhimurium var. 5- strain isolated from chicken breast. Genome Announc **2013**; 1:e01068-13.
16. Hoffmann M, Muruvanda T, Pirone C, et al. First fully closed genome sequence of *Salmonella enterica* subsp. enterica serovar Cubana associated with a food-borne outbreak. Genome Announc **2014**; 2:e01112–4.
17. Chin CS, Alexander DH, Marks P, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **2013**; 10:563–9.
18. Klimke W, Agarwala R, Badretdin A, et al. The National Center for Biotechnology Information's Protein Clusters Database. Nucleic Acids Res **2009**; 37:D216–23.
19. Janies DA, Pomeroy LW, Aaronson JM, et al. Analysis and visualization of H7 influenza using genomic, evolutionary and geographic information in a modular web service. Cladistics **2012**; 28:483–8.
20. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics **2009**; 25:2078–9.
21. Zwickl DJ. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD dissertation, The University of Texas at Austin. **2006**.
22. Mermin J, Hutwagner L, Vugia D, et al. Reptiles, amphibians, and human *Salmonella* infection: a population-based, case-control study. Clin Infect Dis **2004**; 38(suppl 3):S253–61.
23. Andre M, Ijaz K, Tillinghast JD, et al. Transmission network analysis to complement routine tuberculosis contact investigations. Am J Public Health **2007**; 97:470–7.
24. Cao G, Meng J, Strain E, et al. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PloS One **2013**; 8:e55687.
25. Fricke WF, Mammel MK, McDermott PF, et al. Comparative genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sublineage evolution. J Bacteriol **2011**; 193:3556–68.
26. Makarova KS, Haft DH, Barrangou R, et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol **2011**; 9:467–77.
27. Jiang W, Maniv I, Arain F, Wang Y, Levin BR, Marraffini LA. Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. PLoS Genet **2013**; 9:e1003844.
28. US Food and Drug Administration. Whole Genome Sequencing Program (WGS), **2014**. http://www.fda.gov/Food/FoodScienceResearch/WholeGenomeSequencingProgramWGS/default.htm. Accessed 11 June 2015.
29. US Food and Drug Administration. FDA investigates presence of *Listeria* in some Hispanic-style cheeses, **2014**. http://www.fda.gov/Food/RecallsOutbreaksEmergencies/Outbreaks/ucm386726.htm. Accessed 11 June 2015.