Salmonella enterica serovar Typhimurium from Wild Birds in the United States Represent Distinct Lineages Defined by Bird Type

Running title: *S.* Typhimurium lineages from U.S. wild birds

Yezhi Fu,[a] Nkuchia M. M'ikanatha,[b] Jeffrey M. Lorch,[c] David S. Blehert,[c] Brenda Berlowski-Zier,[c] Chris A. Whitehouse,[d] Shaoting Li,[e] Xiangyu Deng,[e] Jared C. Smith,[f] Nikki W. Shariat,[f] Erin M. Nawrocki,[a] Edward G. Dudley[a,g,*]

[a]Department of Food Science, The Pennsylvania State University, University Park, PA, U.S.

[b]Pennsylvania Department of Health, Harrisburg, PA, U.S.

[c]U.S. Geological Survey - National Wildlife Health Center, Madison, WI, U.S.

[d]Center for Veterinary Medicine, U.S. Food and Drug Administration, Laurel, MD, U.S.

[e]Center for Food Safety, Department of Food Science and Technology, University of Georgia, Griffin, GA, U.S.

[f]Department of Population Health, Poultry Diagnostic and Research Center, University of Georgia, Athens, GA, U.S.

[g]*E. coli* Reference Center, The Pennsylvania State University, University Park, PA, U.S.

*Corresponding author:

427 Rodney A. Erickson Food Science Building University Park, PA 16802, USA. Email: egd100@psu.edu

ABSTRACT *Salmonella enterica* serovar Typhimurium is typically considered a host generalist, however certain strains are associated with specific hosts and show genetic features of host adaptation. Here, we sequenced 131 *S.* Typhimurium strains from wild birds collected in 30 U.S. states during 1978-2019. We found that isolates from broad taxonomic host groups including passerine birds, water birds (Aequornithes), and larids (gulls and terns) represented three distinct lineages and certain *S.* Typhimurium CRISPR types presented in individual lineages. We also showed that lineages formed by wild bird isolates differed from most strains originating from domestic animal sources, and genomes from these lineages substantially improved source attribution of Typhimurium genomes to wild birds by a machine learning classifier. Furthermore, virulence gene signatures that differentiated *S.* Typhimurium from passerines, water birds, and larids were detected. Passerine isolates tended to lack *S.* Typhimurium-specific virulence plasmids. Isolates from the passerine, water bird, and larid lineages had close genetic relatedness with human clinical isolates, including those from a 2021 U.S. outbreak linked to passerine birds. These observations indicate that *S.* Typhimurium from wild birds in the U.S. are likely host-adapted, and the representative genomic dataset examined in this study can improve source prediction and facilitate outbreak investigation.

IMPORTANCE Within-host evolution of *S.* Typhimurium may lead to pathovars adapted to specific hosts. Here, we report the emergence of disparate avian *S.* Typhimurium lineages with distinct virulence gene signatures. The findings highlight the importance of wild birds as a reservoir for *S.* Typhimurium and contribute to our understanding of the genetic diversity of *S.* Typhimurium from wild birds. Our study indicates that *S.* Typhimurium may have undergone

adaptive evolution within wild birds in the U.S. The representative *S.* Typhimurium genomes from wild birds, together with the virulence gene signatures identified in these bird isolates, are valuable for *S.* Typhimurium source attribution and epidemiological surveillance.

68 **INTRODUCTION**

69 *Salmonella enterica* is a major zoonotic and foodborne pathogen, and one of the leading bacterial
70 causes of foodborne illnesses in the U.S. The Centers for Disease Control and Prevention (CDC)
71 estimates that *Salmonella enterica* is responsible for approximately 1.35 million illnesses, 26,500
72 hospitalizations, and 420 deaths each year in the U.S.(1). Food products such as eggs, livestock,
73 and poultry meat are the main sources for most of these illnesses (1). Source attribution studies
74 have documented that food-producing animals such as pigs, cattle, and chicken are the main
75 reservoirs for *Salmonella* bacteria that infect humans (2). Although transmission of non-typhoidal
76 *Salmonella* to humans through food and food-producing animals is well described, the
77 contribution of nonfood sources such as wild birds is less defined. Nonfood-producing animals
78 are less explored in most source-attribution studies mainly due to a lack of representative and
79 routinely collected data on these animals (3).

80 There have been several reports suggesting transmission of *S. enterica* between wild birds and
81 humans in recent years. For instance, possible interspecies transmission of *S.* Enteritidis among
82 gulls, poultry, and humans has been reported in Chile (4). Isolates of *S.* Hvittingfoss from a 2016
83 multi-state outbreak originating from tainted cantaloupes in Australia closely matched *S.*
84 Hvittingfoss strains isolated from bar-tailed godwits (5). Although there is no direct evidence
85 suggesting that the above-mentioned outbreaks originated from charadriiform birds (shorebirds
86 and larids), it highlights the potential role of charadriiform birds in dissemination of *S. enterica*
87 involved in human infections. Investigations focused on passerine birds (songbirds) identified a
88 total of 337 human infections with passerine-associated *S.* Typhimurium DT40, DT56(v) and

89 DT160 isolates during 2000-2010 (6). In 2021, an *S.* Typhimurium outbreak linked to passerines
90 has led to 29 illnesses and 14 hospitalizations in over 12 U.S. states (7).
91 As wild birds may serve as a reservoir of *S. enterica* that cause human salmonellosis, it is
92 important to characterize isolates from wild birds so that strains from human cases can be properly
93 traced to the appropriate source and potential transmission routes can be identified. A public portal
94 integrating pathogen genomic sequences from various sources and geographic regions can
95 facilitate source attribution studies with large datasets in real-time. Although the NCBI Pathogen
96 Detection (PD) provides such a data portal for traceback investigation of an outbreak
97 (https://www.ncbi.nlm.nih.gov/pathogens/), the sequencing data are skewed towards isolates from
98 clinical samples, foods, or food-producing animals (i.e., livestock and poultry). The limited data
99 from nonfood sources such as wild birds may hinder source attribution during an outbreak
100 investigation, thus affecting subsequent prevention and mitigation measures. Besides direct
101 sequence comparisons, machine learning algorithms such as Random Forest (RF) classifier have
102 shown promise for predicting *Salmonella* host specificity (8) or zoonotic sources (9). Zhang et al.
103 reported a Random Forest classifier for genomic source prediction of *S.* Typhimurium in the U.S.
104 that showed a marked difference in prediction accuracies between food-producing animals
105 (prediction accuracy ~90%) and wild birds (prediction accuracy ~50%). The difference was
106 presumably due to a lack of representative and available wild bird isolates (9). Therefore, it is
107 necessary to include *S. enterica* isolates beyond food or food-producing animals in a public data
108 portal such as PD to help better resolve complex *Salmonella* epidemiology.
109 In this study, we sequenced 131 *S.* Typhimurium strains from wild birds isolated by the U.S.
110 Geological Survey - National Wildlife Health Center. Strains were isolated between 1978-2019

111 and originated from birds collected in 30 U.S. states. We used these data to test the following
112 hypotheses: 1) Strains would be phylogenetically segregated by bird host; 2) Additional *S.*
113 Typhimurium genomes from wild bird isolates would improve the performance of machine
114 learning-based source attribution; 3) Allelic variation in virulence genes would contribute to *S.*
115 Typhimurium host specificity; and 4) Inclusion of these data in the PD database would better
116 define cross-species transmission of *S.* Typhimurium between wild birds, humans, and other hosts.
117

118 **RESULTS**

119 **Distinct *S.* Typhimurium lineages are carried by passerine birds, water birds, and larids.**
120 Metadata (isolate name, SRA accession number, collection year, state, source/host species,
121 CRISPR type, and classic 7-gene MLST sequence type) for the 131 isolates are listed in **Table**
122 **S1**. Based on single nucleotide polymorphism (SNP) phylogenetic analyses, isolates from wild
123 birds formed three major lineages (**Fig. 1A**), each supported by robust bootstrap values of 100%.
124 The largest branch included 42 isolates, all of which originated from passerine birds (order
125 Passeriformes) (e.g., cardinal, finches, and sparrows). The second major lineage included 40
126 isolates predominantly originating from water birds of the clade Aequornithes (e.g., cormorants,
127 pelicans, and herons). This lineage also included occasional strains isolated from other birds that
128 commonly share the same aquatic habitats or prey/scavenge on the carcasses of the above-
129 mentioned species: terns (n = 2), gull (n = 1), grebe (n = 1), goose (n = 1), and bald eagle (n = 1).
130 The third major *S.* Typhimurium lineage consisted of 32 strains primarily isolated from larids
131 (terns and gulls; order Charadriiformes). This lineage also included isolates originating from a
132 plover (also a member of Charadriiformes) and an owl. Subclades formed by *S.* Typhimurium

133 strains within each of the three major lineages did not show clear patterns of host specificity (e.g.,
134 strains isolated from terns and gulls did not form unique subclades but rather were interspersed
135 within the core larid lineage). Isolates from raptors (e.g., owls, bald eagles, and hawks) did not
136 cluster together, and are highlighted in red in **Fig. 1A**. Based on Bayesian phylogenetic inference
137 (**Fig. 1B**), we were able to estimate that the most recent common ancestor (MRCA) of the water
138 bird lineage was from around 1915 [95% highest posterior distribution (HPD): 1902-1926]. The
139 MRCAs of the passerine lineage and larid lineage were from approximately 1941 (95% HPD:
140 1934-1947) and 1938 (95% HPD: 1929-1947), respectively. The passerine lineage and the larid
141 lineage were estimated to have split from each other around 162 years ago (median: 1857; 95%
142 HPD: 1841-1873).
143
144 **Defined *S.* Typhimurium lineages contain disparate CRISPR types.**
145 We further classified isolates by two established *Salmonella* typing methods, i.e., classic 7-gene
146 multilocus sequence typing (MLST) and CRISPR (Clustered Regularly Interspaced Short
147 Palindromic Repeats) typing, to determine the sequence type (ST) and *S.* Typhimurium CRISPR
148 type (STCT) of our isolates. Six STs were detected and 93.9% of the isolates (123/131) belonged
149 to ST19 (n = 90) and ST99 (n = 33). All the ST99 isolates belonged to the water bird lineage.
150 However, ST did not always distinguish between the lineages. Specifically, ST19 isolates were
151 dispersed among the larid and passerine bird lineages (**Fig. 1A** and **Table S1**). On the other hand,
152 CRISPR analysis identified 37 STCTs within the 131 isolates (**Fig. 1A** and **Table S1**), several of
153 which only appeared in certain lineages. For examples, STCT10 (n = 12), STCT11 (n = 14), and
154 STCT34 (n = 10) were unique to the passerine bird lineage; STCT18 (n = 24), STCT19 (n = 5),

155 and STCT20 (n = 6) only occurred in the water bird lineage; and STCT7 (n = 9), STCT8 (n = 10),

156 and STCT36 (n = 6) were exclusive to the larid lineage.

157

158 **Genetic relatedness of U.S. wild bird isolates with isolates from other reservoirs and**

159 **geographic regions.**

160 We examined whether wild bird isolates of *S*. Typhimurium were related to those from other host

161 sources and geographic regions. **Fig. 2** and **Table S2** show that wild bird isolates were related to

162 other isolates in the PD database from diverse sources (e.g., humans, water, fish/shellfish, horses,

163 cats, and food) and different geographic regions. The PD SNP Tree Viewer showed that our

164 isolates from the passerine lineage, water bird lineage, and larid lineage formed six, four, and four

165 SNP clusters with other isolates in the large database, respectively (**Table S2**). Each SNP cluster

166 contains isolates whose genomes are within 50-SNP distance of each other. We summarized the

167 source niches and geographic regions of all the isolates in the SNP clusters formed by different *S.*

168 Typhimurium lineages, and the results are represented in **Fig. 2**. As shown in **Fig. 2**, **A** and **B**,

169 isolates from both the water bird and larid lineages had close genetic relatedness to isolates from

170 water, fish/shellfish, etc. In contrast, isolates of the passerine lineage were not related to

171 fish/shellfish isolates in the PD database; rather these isolates were more closely related with

172 strains originating from horses and cats (**Fig. 2C**). Notably, isolates from the three major wild bird

173 lineages were genetically similar to a number of clinical isolates from humans. The numbers of

174 clinical isolates clustered with isolates from water bird, larid, and passerine lineages in the PD

175 database were 37, 84, and 495, respectively. Interestingly, isolates from food-producing animals

176 such as livestock and poultry rarely clustered with wild bird isolates. Only two isolates from

177 livestock were related to passerine-derived isolates, while none was closely associated with water
178 bird or larid isolates (**Fig. 2**, **A** to **C**). Geographically, isolates in both the water bird and larid
179 lineages were closely related to strains originating from outside of the U.S. Isolates from Canada,
180 Chile, Mexico, UK, Japan, and Australia clustered with isolates from the water bird and larid
181 lineages from the U.S. (**Fig. 2**, **D** and **E**). Isolates from the passerine lineage only demonstrated
182 close genetic relatedness with strains isolated from sources within North America (**Fig. 2F**).
183

184 **Inclusion of the 131 *S.* Typhimurium isolates from wild birds improves source prediction**
185 **accuracy of a Random Forest classifier.**

186 Given the distinct clustering of wild bird isolates, we next explored whether our dataset would
187 improve a previously published Random Forest (RF) classifier for source prediction of *S.*
188 Typhimurium strains. With the original training dataset (195 bovine isolates, 338 porcine isolates,
189 440 poultry isolates, and 68 wild bird isolates), the RF classifier had a prediction accuracy of
190 64.62%, 88.76%, 91.14%, and 52.94% for bovine, porcine, poultry, and wild birds, respectively,
191 and an out-of-bag prediction accuracy of 82.9%. After adding our 131 wild bird isolates into the
192 training dataset, the out-of-bag prediction accuracy was increased to 84.3% (**Table 1**). Notably,
193 there was a substantial increase in prediction accuracy for wild birds from 52.94% to 83.42%. In
194 contrast, the prediction accuracies for bovine and poultry remained largely unaffected. To explain
195 the prediction results using phylogeny, a SNP tree was built that included 1,604 isolates from
196 different sources [171 human clinical isolates, 261 isolates from food and other sources, and 1,172
197 isolates used for the RF classifier training (i.e., 195 bovine isolates, 338 porcine isolates, 440
198 poultry isolates, and 199 wild bird isolates in which 131 were from this study)]. These different

199 sourced isolates excluding the 131 wild bird isolates were characterized in a previous *S.*
200 Typhimurium source attribution study (9) The wild bird isolates formed three lineages that did not
201 overlap with those formed by other animal-sourced isolates (**Fig. 3**), which corresponded to the
202 source prediction performance of the RF classifier that the number of wild bird isolates falsely
203 attributed to porcine, bovine, or poultry sources only increased from 32 to 33 after adding our bird
204 isolates into the training dataset (**Table 1**). It should be noted that isolates from bovines were more
205 dispersed and overlapped with other sourced isolates in the phylogenetic tree (**Fig. 3**). This was
206 consistent with the source prediction performance of the RF classifier that 37.44% (73/195) of the
207 bovine isolates were assigned to porcine, poultry, or wild bird sources (**Table 1**).
208
209 **Genetic signatures related to virulence differ between strains originating from different**
210 **types of wild birds.**
211 We next took a more focused approach to examine whether presence of antimicrobial resistance
212 genes, virulence genes and/or plasmids were associated with the major wild bird lineages. All 131
213 wild bird *S.* Typhimurium strains were predicted to be sensitive to major types of antimicrobials
214 (beta-lactam, tetracycline, phenicol, fluoroquinolone, etc.) (**Table S3**). These isolates carried
215 common virulence determinants found in *S.* Typhimurium. These included fimbrial/nonfimbrial
216 adherence determinants, genes promoting bacterial survival in activated macrophage, type three
217 secretion system (TTSS) encoded by *Salmonella* pathogenicity island I and II (SPI-1 and SPI-2),
218 and genes associated with magnesium uptake, serum resistance, stress adaptation, and toxin
219 production (**Table S3**).

220 Isolates from the water bird, larid, and passerine lineages each had defining mutations that may
221 abolish the function of specific virulence genes. The *fimC* gene (full length: 708 bp) encoding the
222 type 1 pilus chaperone FimC had a single base-pair deletion at position 87 in all isolates (100%;
223 42/42) belonging to the passerine lineage. Only two isolates (i.e., PSU-2849 and PSU-3252) from
224 passerine birds were predicted to encode the full length *fimC*, however these isolates were
225 divergent from the core passerine lineage. Conversely, all the isolates (100%; 72/72) from the
226 water bird and larid lineages possessed an intact *fimC* (**Fig. 4A**). On the other hand, all the isolates
227 (100%; 40/40) in the water bird lineage had a single base-pair insertion in the *safB* gene (full
228 length: 714 bp; insertion at position 601) encoding the Saf pilus chaperone SafB, a 60 base-pair
229 deletion in the *sthC* gene (full length: 2538 bp; deletion at 481-540) encoding the Sth pilus usher
230 protein SthC, and a single base-pair insertion and substitution in the *sseL* gene (*Salmonella*
231 secreted effector L; full length: 962 bp; insertion at position 474, substitution at position 823).
232 These three genes occurred in all the isolates from passerine and larid lineages (**Fig. 4A**). However,
233 there was a substitution at position 867 of the *sseL* gene in isolates from the larid lineage that led
234 to a premature stop codon (**Fig. S1**). The organization of the fimbrial operons containing *fimC*,
235 *safB*, and *sthC* in isolates from the three major wild bird lineages, and reference isolate *S.*
236 Typhimurium LT2 are shown in **Fig. 4**, **B** to **D**. Gene *sseL* was monocistronic and its alignment
237 within different bird isolates is shown in **Fig. S1**. Overall, the mutations (frameshift and premature
238 stop codon) of these virulence genes resulted in pseudogene accumulation that may inactivate
239 their functions (**Fig. 4**, **B** to **D**).
240 Because of known concurrence, we conducted plasmid profiling to examine the correlation
241 between certain virulence/resistance genes and plasmids in different wild bird lineages. As shown

242 in **Fig. 4E**, four plasmids of the incompatibility groups ColpVC, IncFIB(S), IncFII(S), and IncI1
243 were present in these isolates. None of the plasmids carried antimicrobial resistance genes.
244 However, three plasmids [IncI1, IncFIB(S), and IncFII(S)] carried virulence genes. IncI1 was
245 present in a single isolate (PSU-2832); this plasmid carried the type IV pili genes (*pilUQRS*).
246 IncFIB(S) and IncFII(S) carried virulence genes including *pefABCD* (plasmid encoded fimbriae),
247 *mig-5* (macrophage induced gene), *ric* (resistance to complement killing), and *spvB* (*Salmonella*
248 plasmid virulence). We found that 100% (72/72) of isolates from both water bird and larid lineages
249 carried these two virulence plasmids, while only 33.3% (14/42) of isolates from the passerine
250 lineage had these two plasmids. Isolates from passerine birds without the two plasmids also lacked
251 the corresponding virulence genes (i.e., *pefABCD*, *mig-5*, *rck*, and *spvB*) (**Fig. 4E**).
252

253 **DISCUSSION**

254 In this study, we explored the genetic diversity of 131 *S.* Typhimurium isolates collected from
255 wild birds over four decades. Whole genome sequence phylogeny revealed three major lineages
256 of *S.* Typhimurium largely defined by bird host taxonomic group; these lineages possessed certain
257 STCTs and virulence gene signatures. Potential transmission of strains between the three lineages
258 and non-avian host species was variable as was the extent of their geographic distribution. Our
259 study also shows the wild bird *S.* Typhimurium lineages did not overlap with major lineages
260 formed by domestic animal-sourced isolates. The addition of these wild bird isolates into a training
261 dataset improved source prediction accuracy of a RF classifier among bovine, porcine, poultry,
262 and wild birds.

263 Different host physiologies, habitats, and life histories may explain the divergence of *S.*
264 Typhimurium lineages infecting different groups of wild birds. The three major lineages of *S.*
265 Typhimurium in our study each corresponded to taxonomically disparate host groups. The water
266 bird lineage was primarily associated with birds in the clade Aequornithes (e.g., cormorants,
267 pelicans, herons), the larid lineage with birds in the order Charadriiformes (e.g., gulls, terns,
268 plover), and the passerine lineage with birds in the order Passeriformes (e.g., finches, sparrows).
269 Correspondence of these lineages to bird taxonomic group suggests a degree of host adaptation
270 by *S.* Typhimurium infecting wild birds, similar to what has been described in strains infecting
271 sea turtles (10) and pigeons (11). Some evidence suggests that certain *S.* Typhimurium strains
272 maintained within passerine birds in the U.K. are host-adapted (12, 13). However, passerine-
273 adapted *S.* Typhimurium strains in the U.K. belong to two definitive phage types (DT), i.e., DT40
274 and DT56(v) (the STCTs are 745, 7743, or 9520) (14, 15) and were apparently unrelated to the
275 passerine isolates in this study, which mostly have the STCTs of 10, 11, and 34 (**Fig. 1A**). To the
276 best of our knowledge, there are no readily available systematic studies reporting *S.* Typhimurium
277 strains adapted to water birds or larids. Although the three lineages we describe herein were
278 associated with particular bird groups at broad host taxonomic levels, we did not observe evidence
279 of further specialization occurring at finer taxonomic scales. Specifically, *S.* Typhimurium strains
280 isolated from the same host species did not form subclades within their major lineages (e.g.,
281 cormorant isolates did not form a subclade unique from pelican isolates within the water bird
282 lineage). However, we sampled a relatively small number of wild bird hosts relative to the avian
283 diversity that exists in North America. Specifically, our sampling was biased toward species which
284 regularly experience outbreaks of salmonellosis such as cormorants, gulls, and finches (16).

285 Outbreaks may be more prevalent in these bird species because colonial nesting (cormorants,
286 pelicans, herons, gulls, and terns) and congregation at feeding and water sites such as bird feeders
287 and bird baths (songbirds) facilitate transmission. Thus, a greater diversity of host-adapted
288 lineages of *S*. Typhimurium may be discovered by sampling additional avian groups that harbor
289 cryptic infections.

290 Spatial patterns, including habitat partitioning by different bird hosts do not fully explain the *S.*
291 Typhimurium phylogenetic patterns we observed. For example, cormorants, pelicans, and gulls
292 often occupy the same habitats and occur in close proximity to one another. Yet, strains from these
293 birds tended to fall into two distinct lineages based on bird phylogeny, suggesting factors such as
294 host physiology may be the key factor driving *S.* Typhimurium host adaptation. Notably, the
295 passerine lineage is sister to the larid lineage, and both are more distantly related to the water bird
296 lineage despite water birds and larids being much more closely related than either group is to the
297 passerines (17). The Bayesian phylogenetic inference indicates the three lineages formed
298 sometime after 1900. This implies that host adaptation occurred well after the divergence of the
299 avian host groups. Similarly, a lineage of *S*. Typhimurium adapted to sea turtles across the Pacific
300 Ocean evolved from a MRCA only a few decades ago (10). Taken together, these findings suggest
301 that spillover of *S*. Typhimurium into wildlife and subsequent host adaptation may be a relatively
302 recent phenomenon, likely driven by anthropogenic influences.

303 While the presence of three major lineages suggested host adaptation by bird-infecting strains of
304 *S.* Typhimurium, there was evidence that strains within a lineage could infect bird species outside
305 of its typical host group. Infections in aberrant hosts usually corresponded with hosts from
306 different groups that share similar environments to one another. For example, *S.* Typhimurium

307 strains in the water bird lineage were occasionally isolated from other birds (i.e., terns, gulls, grebe,
308 and goose) that share aquatic habitats with cormorants, pelicans, and herons. Interestingly, strains
309 of *S.* Typhimurium isolated from raptors (e.g., eagles, hawks, and owls) were phylogenetically
310 dispersed. The diversity of strains harbored by raptors may be due to acquisition of *S.*
311 Typhimurium through consumption of other birds and animals that may be infected with the
312 bacterium rather than possessing their own unique strains.
313 While *S.* Typhimurium strains clustered by host species, there was little evidence of geographic
314 patterns based on collection locations within the U.S. (**Fig. 1A**). This may be explained by the
315 high mobility of wild birds, which travel long distances and across political boundaries. Isolates
316 from the three major lineages did, however, cluster with isolates from different global regions in
317 the PD database. Strains from the larid lineage were genetically similar to isolates in the database
318 originating from South America, Oceania, and Europe. This is consistent with the long-distance
319 migratory capacity for terns and gulls which makes them capable of spreading certain strain of *S.*
320 Typhimurium across the world. Isolates from the water bird lineage exhibited a similar geographic
321 pattern. In contrast, isolates from the passerine lineage did not closely match isolates outside of
322 North America, which is consistent with the more typical intracontinental migratory patterns for
323 the passerine species. It should be noted that the PD database is heavily skewed towards isolates
324 from the U.S. Isolates from other geographic regions are underrepresented since many countries
325 do not sequence isolates or actively submit data to the NCBI PD database for surveillance purposes.
326 Additional sampling in other parts of the world may reveal that all three major lineages are more
327 widely distributed.

328 Host adaptation of *S.* Typhimurium in various groups of wild birds may affect their transmission
329 among different hosts or sources. For example, isolates from the water bird and larid lineages
330 clustered with isolates from aquatic animals (fish/shellfish), which is understandable given their
331 overlap in habitats. However, no isolates from aquatic animals were closely related with isolates
332 from the passerine lineage. Instead, the passerine lineage was more likely to contain *S.*
333 Typhimurium strains isolated from horses and cats. Studies have reported that cats can acquire *S.*
334 Typhimurium by catching infected songbirds or spending time around bird feeders (18, 19).
335 However, few studies explored the potential spread of *S.* Typhimurium between wild birds and
336 horses, which may be exposed to strains of the passerine lineage in open pastures or stables that
337 are also occupied by avian species such as finches and house sparrows.
338 Interestingly, isolates from the three wild bird lineages were rarely related to isolates from
339 livestock and poultry, but were frequently associated with clinical isolates from human beings
340 (**Fig. 2** and **Table S2**). Our addition of isolates to the database will facilitate attribution of *S.*
341 Typhimurium strains from wild birds recovered from other hosts or substrates to the appropriate
342 source. Although the exact transmission route and direction are difficult to determine, studies such
343 as this are important in raising awareness of wild birds as a reservoir of *S.* Typhimurium that may
344 contribute to preventable cases of human salmonellosis. For example, several studies have
345 reported the concurrence of *S.* Typhimurium outbreaks in both passerine birds and humans in the
346 U.K., U.S., and Sweden during late winter or early spring (6, 19, 20), and cats are considered to
347 be a possible intermediate vehicle for *S.* Typhimurium transmission between passerine birds and
348 human beings during outbreaks (19, 21, 22). This is consistent with our observation that passerine
349 bird isolates cluster with isolates from cats in the PD database (**Fig. 2C**). In 2021, an *S.*

350 Typhimurium outbreak that caused at least 29 illnesses and 14 hospitalizations in the U.S. was
351 traced back to isolates from songbirds (7). In fact, clinical isolates from this outbreak formed a
352 SNP cluster with passerine isolates in our study (**Fig. 5**; three pine siskin isolates: PSU-3338,
353 PSU-3367, and PSU-3392; two redpoll isolates: PSU-3337; and one crossbill isolate: PSU-3353,
354 highlighted in red). These bird isolates belonged to STCT10 and STCT11. Our passerine isolates
355 differed from clinical isolates by <10 SNPs, suggesting a recent common ancestor (23). None of
356 our passerine isolates were collected earlier than the clinical isolates from the 2021 outbreak (**Fig.**
357 **5**), and although we cannot say with certainty, it is plausible that strains transmitted from passerine
358 birds to humans were responsible for the outbreak. Humans may get infected via direct contact
359 with songbirds, contaminated bird feeders, or infected cats; or there may be additional unknown
360 sources not represented in the PD database. To further elucidate routes of transmission, collecting
361 contemporaneous *S*. Typhimurium isolates from humans, wild birds, and environmental samples
362 through outbreak investigations is necessary. Such efforts offer insights to guide preventative
363 measures to reduce the health risk for future outbreaks.
364 Inclusion of the wild bird isolates in the PD database not only facilitates outbreak investigation,
365 but also improves source attribution. The RF classifier in this study performed much better to
366 predict isolates from wild birds with the addition of our 131 isolates to a training dataset from
367 porcine (n = 338), bovine (n = 195), poultry (n = 440), and wild birds (n = 68) (**Table 1**). The
368 significant increase in wild bird isolates from 68 to 199 may contribute to the improved prediction
369 accuracy for wild bird sources. It should be noted, however, that prediction accuracy not only
370 depends on sampling intensity but also presence of representative isolates from different sources.
371 Even though the count of isolates from bovine sources (n = 195) is almost equal to that from wild

372 birds (n = 199), prediction accuracy for bovine sources (62.56%) is lower than that for wild bird
373 sources (83.42%). The high prediction accuracy for wild birds may also be attributed to consistent
374 presence of genetic signatures such as pseudogene accumulation in fimbrial genes which may help
375 discriminate wild bird isolates from other sourced isolates. This is consistent with the observation
376 that wild bird isolates form three unique lineages, while bovine isolates are more phylogenetically
377 dispersed in the SNP phylogenetic tree (**Fig. 3**). Overall, an improvement in prediction accuracy
378 is crucial to help identify root sources of *S.* Typhimurium outbreaks, especially when the outbreak
379 is linked to nonfood sources.

380 It is well documented that genome degradation contributes to *Salmonella* host adaptation (24, 25,
381 26, 27, 28, 29). Our study shows pseudogene accumulation in specific chromosomal fimbrial
382 genes in certain wild bird *S.* Typhimurium lineages (i.e., *fimC* in passerine lineage, *safB* and *sthC*
383 in water bird lineage). The inactivation of these genes due to frameshift and premature stop codons
384 may help *S.* Typhimurium adapt to a particular bird host. Furthermore, passerine bird isolates
385 tended to lose the plasmid carrying virulence genes (i.e., *pefABCD*, *mig-5*, *rck*, and *spvB*). These
386 virulence genes are important for *S.* Typhimurium survival and replication in human and mouse
387 macrophages (30, 31, 32), which may indicate isolates from the passerine lineage are more host-
388 adapted to birds than those from water bird and larid lineages.

389 To conclude, the findings in this study highlight the importance of wild birds as a reservoir for *S.*
390 Typhimurium. Source attribution can be more precise if we include more representative wild bird
391 isolates into *S.* Typhimurium databases. In addition, the distinct lineages defined by bird type,
392 together with the presence of virulence gene signatures, imply host adaptation of *S.* Typhimurium
393 to wild birds. Considering wild birds may spread *S.* Typhimurium in a very different manner than

394 domestic animals, alternative management strategies will be required to prevent their transmission
395 to human beings. While wild birds serve as potential reservoirs of *S.* Typhimurium, these species
396 also provide critical ecological and economic services to humans and are of great cultural
397 significance. Activities aimed at preventing zoonotic transmission should focus on strategies that
398 benefit both wild birds and humans alike.
399

400 **MATERIALS AND METHODS**

401 **Bacterial strains.**

402 *Salmonella* Typhimurium strains (*n* = 131) from wild birds were isolated by the U.S. Geological
403 Survey - National Wildlife Health Center between 1978 and 2019 from 30 states in the U.S. and
404 stored in cryogenic vials at -80 °C. All strains were confirmed as *S.* Typhimurium by traditional
405 *Salmonella* serotyping and reconfirmed by SeqSero2 v1.2.1 (33) from newly generated whole
406 genome sequencing data (see below).
407

408 **DNA extraction and whole genome sequencing.**

409 For DNA extraction, each isolate was streaked onto xylose lysine deoxycholate (XLD) agar plates
410 and incubated for 18 h at 37 °C. A single colony was then picked, transferred to Luria-Bertani
411 broth (LB) and cultured overnight at 37 °C with continuous agitation (250 rpm). Genomic DNA
412 was extracted using the Qiagen DNeasy® Blood & Tissue kit (Qiagen, Valencia, CA) following
413 the manufacturer's instructions. Genomic DNA purity was confirmed via an A260/A280
414 measurement (target ≥1.8) using NanoDrop™ One (Thermo Scientific™, DE), and DNA
415 concentration was quantified using Qubit® 3.0 (Thermo Fisher Scientific Inc., MA) fluorometer.

416 Isolated genomic DNA was stored at -20 °C before use. For whole genome sequencing (WGS),
417 genomic DNA was adjusted to 0.2 ng/µL. DNA library was then prepared using the Nextera XT
418 DNA Library Prep Kit (Illumina Inc., San Diego, CA). The resulting library was normalized using
419 quantitation-based procedure and pooled together at equal volume. The pooled library (600 µL)
420 was denatured and sequenced on an Illumina MiSeq sequencer (Illumina Inc., San Diego, CA)
421 using a MiSeq reagent v3 kit, with 500 (2 x 250) cycles.
422
423 **Quality assessment for raw reads.**
424 After sequencing, the quality of Illumina paired-end reads of all isolates were assessed using the
425 MicroRunQC workflow in GalaxyTrakr v2 (34). Raw reads passing quality control thresholds (i.e.,
426 average coverage >30, mean quality score >30, number of contigs <400, total assembly length
427 between 4.4-5.1 Mb) were submitted to the NCBI under BioProject PRJNA357723. Metadata
428 information and Sequence Read Archive (SRA) accession numbers of these isolates are listed in
429 **Table S1**.
430
431 **Phylogenetic analysis.**
432 For genetic relatedness comparison between the 131 isolates, raw reads were uploaded to
433 Enterobase (35). *S.* Typhimurium strain LT2 (RefSeq NC_003197.1) served as the reference
434 genome and a maximum-likelihood (ML) phylogenetic tree was created using the SNP project in
435 Enterobase based on 5,839 SNPs in the core genomic regions of the 131 wild bird *S.* Typhimurium
436 isolates. The ML phylogenetic tree was re-constructed by MEGA X v10.1.8 using the Tamura-
437 Nei model and 500 bootstrap replicates (36). The SNP phylogenetic tree was visualized and

438 annotated using the Interactive Tree of Life (iTOL v6; https://itol.embl.de). CRISPR arrays were
439 identified using CRISPRviz (37). Spacers were aligned and unique arrays were given a unique
440 allelic identifier as described by Shariat et al (38). *S.* Typhimurium CRISPR Type (STCT) was
441 then determined by the unique combination of CRISPR1 and CRISPR2 alleles. In addition,
442 sequence type (ST) of these *S.* Typhimurium isolates was identified using 7-gene (*aroC*, *dnaN*,
443 *hemD*, *hisD*, *purE*, *sucA* and *thrA*) multilocus sequence typing (MLST) at Enterobase (35). STCTs
444 and STs were annotated in the SNP phylogenetic tree.
445
446 **Bayesian phylogenetic inference.**
447 The evolutionary history of the 131 isolates was inferred by Bayesian phylogenetic analysis.
448 Snippy (Galaxy v4.5.0) (https://github.com/tseemann/snippy) was used to generate a full
449 alignment and find SNPs between the reference genome LT2 (RefSeq NC_003197.1) and the
450 genomes of wild bird isolates, and Snippy-core (Galaxy v4.5.0)
451 (https://github.com/tseemann/snippy) was used to convert the Snippy outputs into a core SNP
452 alignment. The resultant multiple sequence alignment (5,251 variant sites) was used to construct
453 a time-scale Bayesian phylogenetic tree by BEAUti v2.6.5 and BEAST2 v2.6.5 (39). The
454 parameters were set as followings: Prior assumption-coalescent Bayesian skyline; Clock model-
455 relaxed clock log normal; Markov chain Monte Carlo (MCMC): chain length-100 million, storing
456 every 1,000 generations. A maximum clade credibility tree was created using TreeAnnotator
457 v2.6.4 with burnin percentage of 10% and node height option of median height (39). Finally, the
458 tree was visualized using FigTree v1.4.4 (https://github.com/rambaut/figtree/releases).
459

460 **Genetic relatedness between wild bird isolates and other isolates.**

461 The genetic relatedness between wild bird isolates and other isolates from environmental sources,
462 food, or human hosts were inferred by using the NCBI Pathogen Detection
463 https://www.ncbi.nlm.nih.gov/pathogens/. After uploading the raw reads of the 131 isolates to
464 NCBI, NCBI Pathogen Detection assembled, annotated, and clustered the newly sequenced isolate
465 genomes to other closely related isolates in the database. Clustering involved two steps: First,
466 related isolates were clustered based on whole genome MLST (wgMLST) scheme of *Salmonella*
467 with a 25-allele cut-off; Once wgMLST clusters were created, SNPs are called by aligning
468 assemblies against a reference genome chosen from each wgMLST cluster of closely related
469 isolates, and SNP clusters and phylogenetic trees were inferred. A SNP cluster was classified as a
470 cluster of isolates where each isolate was less than or equal to 50 SNPs distant from other members
471 of the cluster (40). Individual phylogenetic trees for each SNP cluster, together with the metadata
472 information of isolates in the same cluster (**Table S2**), were used to examine relationship between
473 wild bird isolates and other isolates.

474

475 **Random Forest-based source attribution.**

476 A machine learning Random Forest (RF) classifier was used for source attribution of *S.*
477 Typhimurium to bovine, porcine, poultry and wild bird sources. The RF classifier was built using
478 the method described by Zhang et al (9). Briefly, a set of 3,102 genetic features, including 1,850
479 SNPs, 147 indels, and 1,105 accessory gene, were used for the RF classifier. The randomForest
480 package (v4.6-12) of R was used to build the RF classifier with the "ntree" argument set to be
481 1,000 and other parameters being default. The original training data (195 bovine isolates, 338

482 porcine isolates, 440 poultry isolates, and 68 wild bird isolates) and the training data with the
483 addition of the 131 isolates from wild birds were applied for classifier training. Tables of
484 confusion were generated to describe the performance of the classifier. To explain the prediction
485 performance of the classifier after adding our bird isolates into the training dataset, a maximum-
486 likelihood tree based on core genome alignment of 1,604 *S.* Typhimurium isolates from different
487 sources was constructed using Parsnp v1.5.0 (41). The 1,604 *S.* Typhimurium isolates included
488 171 human clinical isolates, 83 food isolates, 178 isolates from miscellaneous sources, and 1,172
489 isolates used for the RF classifier training (i.e., 195 bovine isolates, 338 porcine isolates, 440
490 poultry isolates, and 199 wild bird isolates in which 131 isolates were from this study).
491
492 **AMR, virulence, and plasmid profiling.**
493 Raw reads of each isolate were *de novo* assembled using Shovill with Trimmomatic on (Galaxy
494 v1.0.4) (42). To identify the antimicrobial resistance (AMR) genes present in the isolates, each
495 draft genome assembly was searched against ResFinder database and CARD database using
496 ResFinder 4.1 (43) and Resistance Gene Identifier (RGI) v5.2.0 (44). AMR genes that passed the
497 default threshold values for ResFinder (≥90% nucleotide identity and ≥60% coverage) and RGI
498 (≥95% nucleotide identity) were considered to be present in the isolate. Virulence factors were
499 predicted by aligning the draft genome assembly for each isolate against the Virulence Factor
500 database using VFanalyzer (45). Virulence factors that passed the default threshold values for
501 VFanalyzer (≥90% amino acid identity and ≥50% coverage) were considered to be present in the
502 isolate. If a virulence factor was absent in the isolate identified by VFanalyzer, the virulence gene
503 encoding the specific virulence factor was then compared to its reference gene from strain *S.*

504 Typhimurium LT2 by BLAST to confirm the mutation type (e.g., insertion, deletion, or
505 substitution). Plasmid replicon sequences were identified by comparing the draft genome
506 assembly against the PlasmidFinder database using PlasmidFinder 2.1 with default search settings
507 of ≥95% nucleotide identity and ≥60% coverage (46).
508
509 **Data availability.**
510 Sequence data of the 131 *S.* Typhimurium strains are deposited in the NCBI Sequence Read
511 Archive (https://www.ncbi.nlm.nih.gov/sra) under BioProject PRJNA357723. Accession
512 numbers are available in **Table S1**. Strains that form SNP clusters with wild bird isolates
513 examined in this study are provided in **Table S2** and publicly available in the NCBI Pathogen
514 Detection (https://www.ncbi.nlm.nih.gov/pathogens/).
515
516 **ACKNOWLEDGEMENTS**
517 We thank Drs. Heather Tate and Patrick McDermott for editing this manuscript prior to
518 publication.
519 This work is supported by the U.S. Food and Drug Administration under Grant No. 1 U19
520 FD007114-01 (to E.G.D.), U.S. Department of Agriculture under Grant No. PEN4522 (to
521 E.G.D.), and Penn State College of Agricultural Sciences.
522 The views expressed in this article are those of the authors and do not necessarily reflect the
523 official policy of the Department of Health and Human Services, the U.S. Food and Drug
524 Administration, or the U.S. Government. Any use of trade, firm, or product names is for
525 descriptive purposes only and does not imply endorsement by the U.S. Government.

Author contributions: Y.F. whole genome sequenced the wild bird isolates, designed the study, performed the majority bioinformatics analysis of the data, interpreted the data, and wrote the draft manuscript. N.M.M., J.M.L., D.S.B., B.B-Z., C.A.W., and E.G.D. assisted in collection of the isolates and contributed to interpretation of data and manuscript revision. S.L. and X.D. participated in Random Forest-based source attribution, interpretation of data, and manuscript revision. J.C.S. and N.W.S. participated in *S.* Typhimurium CRISPR typing, interpretation of data, and manuscript revision. E.M.N contributed to identification of virulence gene mutation, interpretation of data, and manuscript revision. E.G.D. additionally designed and supervised the study.

The authors declare no competing interests.

548 **REFERENCES**

549 1. CDC, Centers for Disease Control and Prevention. 2020. *Salmonella*.
550 https://www.cdc.gov/salmonella/index.html
551 2. Pires SM, Vieira AR, Hald T, Cole D. 2014. Source attribution of human salmonellosis:
552 an overview of methods and estimates. *Foodborne Pathog Dis* 11:667-676.
553 https://doi.org/10.1089/fpd.2014.1744
554 3. Mughini-Gras L, Franz E, van Pelt W. 2018. New paradigms for *Salmonella* source
555 attribution based on microbial subtyping. *Food Microbiol* 71:60-67.
556 https://doi.org/10.1016/j.fm.2017.03.002
557 4. Toro M, Retamal P, Ayers S, Barreto M, Allard M, Brown EW, Gonzalez-Escalona N.
558 2016. Whole-genome sequencing analysis of *Salmonella enterica* serovar Enteritidis
559 isolates in Chile provides insights into possible transmission between gulls, poultry, and
560 humans. *Appl Environ Microbiol* 82:6223-6232. https://doi.org/10.1128/AEM.01760-16
561 5. Smith HG, Bean DC, Hawkey J, Clarke RH, Loyn R, Larkins JA, Hassell C, Valcanis M,
562 Pitchers W, Greenhill AR. 2020. *Salmonella enterica* serovar Hvittingfoss in Bar-Tailed
563 Godwits (*Limosa lapponica*) from Roebuck Bay, Northwestern Australia. *Appl Environ*
564 *Microbiol* 86:e01312-20. https://doi.org/10.1128/AEM.01312-20
565 6. Lawson B, De Pinna E, Horton RA, Macgregor SK, John SK, Chantrey J, Duff JP,
566 Kirkwood JK, Simpson VR, Robinson RA, Wain J, Cunningham AA. 2014.
567 Epidemiological evidence that garden birds are a source of human salmonellosis in
568 England and Wales. *PloS One* 9:e88968. https://doi.org/10.1371/journal.pone.0088968

569 7. CDC, Centers for Disease Control and Prevention. 2021. *Salmonella* outbreak linked to
570 wild songbirds. https://www.cdc.gov/salmonella/typhimurium-04-21/index.html
571 8. Lupolova N, Dallman TJ, Holden NJ, Gally DL. 2017. Patchy promiscuity: machine
572 learning applied to predict the host specificity of *Salmonella enterica* and *Escherichia coli*.
573 *Microb Genom* 3:e000135. https://doi.org/10.1099/mgen.0.000135
574 9. Zhang S, Li S, Gu W, den Bakker H, Boxrud D, Taylor A, Roe C, Driebe E, Engelthaler
575 DM, Allard M, Brown E, McDermott P, Zhao S, Bruce BB, Trees E, Fields PI, Deng X.
576 2019. Zoonotic source attribution of *Salmonella enterica* serotype Typhimurium using
577 genomic surveillance data, United States. *Emerg Infect Dis* 25:82.
578 10.3201/eid2501.180835
579 10. Work TM, Dagenais J, Stacy BA, Ladner JT, Lorch JM, Balazs GH, Barquero-Calvo E,
580 Berlowski-Zier BM, Breeden R, Corrales-Gómez N, Gonzalez-Barrientos R, Harris HS,
581 Hernández-Mora G, Herrera-Ulloa Á, Hesami S, Jones TT, Morales JA, Norton TM,
582 Rameyer RA, Taylor DR, Waltzek TB. 2019. A novel host-adapted strain of *Salmonella*
583 Typhimurium causes renal disease in olive ridley turtles (*Lepidochelys olivacea*) in the
584 Pacific. *Sci Rep* 9:1-13. https://doi.org/10.1038/s41598-019-45752-5
585 11. Rabsch W, Andrews HL, Kingsley RA, Prager R, Tschäpe H, Adams LG, Bäumler AJ.
586 2002. *Salmonella enterica* serotype Typhimurium and its host-adapted variants. *Infect.*
587 *Immun.* 70:2249-2255. https://doi.org/10.1128/IAI.70.5.2249-2255.2002
588 12. Hughes LA, Shopland S, Wigley P, Bradon H, Leatherbarrow AH, Williams NJ, Bennett
589 M, de Pinna E, Lawson B, Cunningham AA, Chantrey J. 2008. Characterization of

590 *Salmonella enterica* serotype Typhimurium isolates from wild birds in northern England
591 from 2005–2006. *BMC Vet Res* 4:1-10. https://doi.org/10.1186/1746-6148-4-4
592 13. Mather AE, Lawson B, de Pinna E, Wigley P, Parkhill J, Thomson NR, Page AJ, Holmes
593 MA, Paterson GK. 2016. Genomic analysis of *Salmonella enterica* serovar Typhimurium
594 from wild passerines in England and Wales. *Appl Environ Microbiol* 82:6728-6735.
595 https://doi.org/10.1128/AEM.01660-16
596 14. Lawson B, Hughes LA, Peters T, de Pinna E, John SK, Macgregor SK, Cunningham AA.
597 2011. Pulsed-field gel electrophoresis supports the presence of host-adapted *Salmonella*
598 *enterica* subsp. *enterica* serovar Typhimurium strains in the British garden bird population.
599 *Appl Environ Microbiol* 77:8139-8144. https://doi.org/10.1128/AEM.00131-11
600 15. Mohammed M. 2017. Phage typing or CRISPR typing for epidemiological surveillance of
601 *Salmonella* Typhimurium? *BMC Res Notes* **10**:1-7. https://doi.org/10.1186/s13104-017-
602 2878-0
603 16. Hall AJ, Saito EK. 2008. Avian wildlife mortality events due to salmonellosis in the United
604 States, 1985–2004. *J Wildl Dis* 44:585-593. https://doi.org/10.7589/0090-3558-44.3.585
605 17. Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR.
606 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA
607 sequencing. *Nature* 526:569-573. https://doi.org/10.1038/nature15697
608 18. Taylor DJ, Philbey AW. 2010. *Salmonella* infections in garden birds and cats in a domestic
609 environment. *Vet Rec* 167:26-27. https://doi.org/10.1136/vr.c3156
610 19. Söderlund R, Jernberg C, Trönnberg L, Pääjärvi A, Ågren E, Lahti E. 2019. Linked
611 seasonal outbreaks of *Salmonella* Typhimurium among passerine birds, domestic cats and

612 humans, Sweden, 2009 to 2016. *Euro Surveill* 24:900074. https://doi.org/10.2807/1560-
613 7917.ES.2019.24.34.1900074
614 20. Hernandez SM, Keel K, Sanchez S, Trees E, Gerner-Smidt P, Adams JK, Cheng Y, Ray
615 A, Martin G, Presotto A, Ruder MG, Brown J, Blehert DS, Cottrell W, Maurer JJ. 2012.
616 Epidemiology of a *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain
617 associated with a songbird outbreak. *Appl Environ Microbiol* 78:7290-7298.
618 https://doi.org/10.1128/AEM.01408-12
619 21. Tauni MA, Osterlund A. 2000. Outbreak of *Salmonella* Typhimurium in cats and humans
620 associated with infection in wild birds. *J Small Anim Pract* 41:339-41.
621 https://doi.org/10.1111/j.1748-5827.2000.tb03214.x
622 22. Hauser E, Hühn S, Junker E, Jaber M, Schroeter A, Helmuth R, Rabsch W, Winterhoff N,
623 Malorny B. 2009. Characterization of a phenotypic monophasic variant belonging to
624 *Salmonella enterica* subsp. *enterica* serovar Typhimurium from wild birds and its possible
625 transmission to cats and humans. *Berl Munch Tierarztl Wochenschr* 122:169–177. (In
626 German.)
627 23. Besser JM, Carleton HA, Trees E, Stroika SG, Hise K, Wise M, Gerner-Smidt P. 2019.
628 Interpretation of whole-genome sequencing for enteric disease surveillance and outbreak
629 investigation. *Foodborne Pathog Dis* 16:504-512. https://doi.org/10.1089/fpd.2019.2650
630 24. Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, Parsons BN, Seth-Smith
631 HMB, Barquist L, Stedman A, Humphrey T, Wigley P, Peters SE, Maskell DJ, Corander
632 J, Chabalgoity JA, Barrow P, Parkhill J, Dougan G, Thomson NR. 2015. Patterns of

633 genome evolution that have accompanied host adaptation in *Salmonella*. *Proc Natl Acad*
634 *Sci USA* 112:863-868. https://doi.org/10.1073/pnas.1416707112
635 25. Yue M, Han X, De Masi L, Zhu C, Ma X, Zhang J, Wu R, Schmieder R, Kaushik RS,
636 Fraser GP, Zhao S, McDermott PF, Weill FX, Mainil JG, Arze C, Fricke WF, Edwards
637 RA, Brisson D, Zhang NR, Rankin SC, Schifferli DM. 2015. Allelic variation contributes
638 to bacterial host specificity. *Nat Commun* 6:1-11. https://doi.org/10.1038/ncomms9754
639 26. Klemm EJ, Gkrania-Klotsas E, Hadfield J, Forbester JL, Harris SR, Hale C, Heath JN,
640 Wileman T, Clare S, Kane L, Goulding D, Otto TD, Kay S, Doffinger R, Cooke FJ,
641 Carmichael A, Lever AML, Parkhill J, MacLennan CA, Kumararatne D, Dougan G,
642 Kingsley RA. 2016. Emergence of host-adapted *Salmonella* Enteritidis through rapid
643 evolution in an immunocompromised host. *Nat Microbiol* 1:1-6.
644 https://doi.org/10.1038/nmicrobiol.2015.23
645 27. Wheeler NE, Gardner PP, Barquist L. 2018. Machine learning identifies signatures of host
646 adaptation in the bacterial pathogen *Salmonella enterica*. *PLoS Genet* 14:e1007333.
647 https://doi.org/10.1371/journal.pgen.1007333
648 28. Van Puyvelde S, Pickard D, Vandelannoote K, Heinz E, Barbé B, de Block T, Clare S,
649 Coomber EL, Harcourt K, Sridhar S, Lees EA, Wheeler NE, Klemm EJ, Kuijpers L,
650 Kalonji LM, Phoba MF, Falay D, Ngbonda D, Lunguya O, Jacobs J, Dougan G,
651 Deborggraeve S. 2019. An African *Salmonella* Typhimurium ST313 sublineage with
652 extensive drug-resistance and signatures of host adaptation. *Nat Commun* 10:1-12.
653 https://doi.org/10.1038/s41467-019-11844-z

654 29. Pulford CV, Perez-Sepulveda BM, Canals R, Bevington JA, Bengtsson RJ, Wenner N,
655 Rodwell EV, Kumwenda B, Zhu X, Bennett RJ, Stenhouse GE, De Silva PM, Webster HJ,
656 Bengoechea JA, Dumigan A, Tran-Dien A, Prakash R, Banda HC, Alufandika L,
657 Mautanga MP, Bowers-Barnard A, Beliavskaia AY, Predeus AV, Rowe WPM, Darby AC,
658 Hall N, Weill FX, Gordon MA, Feasey NA, Baker KS, Hinton JCD. 2021. Stepwise
659 evolution of *Salmonella* Typhimurium ST313 causing bloodstream infection in Africa. *Nat*
660 *Microbiol* 6:327-338. https://doi.org/10.1038/s41564-020-00836-1
661 30. Fàbrega A, Vila J. 2013. *Salmonella enterica* serovar Typhimurium skills to succeed in
662 the host: virulence and regulation. *Clin Microbiol Rev* 26:308-341.
663 https://doi.org/10.1128/CMR.00066-12
664 31. Dos Santos AM, Ferrari RG, Conte-Junior CA. 2019. Virulence factors in *Salmonella*
665 Typhimurium: the sagacity of a bacterium. *Curr Microbiol* 76:762-773.
666 https://doi.org/10.1007/s00284-018-1510-4
667 32. Cohen E, Azriel S, Auster O, Gal A, Zitronblat C, Mikhlin S, Scharte F, Hensel M, Rahav
668 G, Gal-Mor O. 2021. Pathoadaptation of the passerine-associated *Salmonella enterica*
669 *serovar* Typhimurium lineage to the avian host. *PLoS Pathog* 17:e1009451.
670 https://doi.org/10.1371/journal.ppat.1009451
671 33. Zhang S, den Bakker HC, Li S, Chen J, Dinsmore BA, Lane C, Lauer AC, Fields PI, Deng
672 X. 2019. SeqSero2: rapid and improved *Salmonella* serotype determination using whole-
673 genome sequencing data. *Appl Environ Microbiol* 85:e01746-19.
674 https://doi.org/10.1128/AEM.01746-19

675 34. Timme RE, Wolfgang WJ, Balkey M, Venkata SLG, Randolph R, Allard M, Strain E.
676 2020. Optimizing open data to support one health: best practices to ensure interoperability
677 of genomic data from bacterial pathogens. *One Health Outlook* 2:1-11.
678 https://doi.org/10.1186/s42522-020-00026-3
679 35. Zhou Z, Alikhan NF, Mohamed K, Fan Y, the Agama Study Group, Achtman M. 2020.
680 The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia*
681 *pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res* 30:138-152.
682 https://doi.org/10.1101/gr.251678.119
683 36. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary
684 genetics analysis across computing platforms. *Molecular Biol Evol* 35:1547-1549.
685 https://doi.org/10.1093/molbev/msy096
686 37. Nethery MA, Barrangou R. 2019. CRISPR Visualizer: rapid identification and
687 visualization of CRISPR loci via an automated high-throughput processing pipeline. *RNA*
688 *Biol* 16:577-584. https://doi.org/10.1080/15476286.2018.1493332
689 38. Shariat N, Sandt CH, DiMarzio MJ, Barrangou R, Dudley EG. 2013. CRISPR-MVLST
690 subtyping of *Salmonella enterica* subsp. *enterica* serovars Typhimurium and Heidelberg
691 and application in identifying outbreak isolates. *BMC Microbiol* 13:1-17.
692 https://doi.org/10.1186/1471-2180-13-254
693 39. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A,
694 Heled J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie
695 HA, du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH,
696 Xie D, Zhang C, Stadler T, Drummond AJ. 2019. BEAST 2.5: An advanced software

697 platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 15:e1006650.
698 https://doi.org/10.1371/journal.pcbi.1006650
699 40. Sanaa M, Pouillot R, Vega FG, Strain E, Van Doren JM. 2019. GenomeGraphR: a user-
700 friendly open-source web application for foodborne pathogen whole genome sequencing
701 data integration, analysis, and visualization. *PloS One* 14:e0213039.
702 https://doi.org/10.1371/journal.pone.0213039
703 41. Cherry JL. 2017. A practical exact maximum compatibility algorithm for reconstruction
704 of recent evolutionary history. *BMC Bioinform.* 18:1-12. https://doi.org/10.1186/s12859-
705 017-1520-4
706 42. Seemann T. 2017. Shovill: Faster SPAdes assembly of Illumina reads.
707 https://github.com/tseemann/shovill
708 43. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe
709 RL, Rebelo AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G,
710 Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H,
711 Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek K,
712 Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM. 2020.
713 ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother*
714 75:3491-3500. https://doi.org/10.1093/jac/dkaa345
715 44. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W,
716 Nguyen ALV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE,
717 Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva
718 A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire

719 F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG.
720 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic
721 resistance database. *Nucleic Acids Res* 48:D517-D525. https://doi.org/10.1093/nar/gkz935
722 45. Liu B, Zheng DD, Jin Q, Chen LH, Yang J. 2019. VFDB 2019: a comparative
723 pathogenomic platform with an interactive web interface. *Nucleic Acids Res* 47:D687-
724 D692. https://doi.org/10.1093/nar/gky1080
725 46. Carattoli A, Zankari E, García-Fernández A, Larsen MV, Lund O, Villa L, Aarestrup FM,
726 Hasman H. 2014. *In silico* detection and typing of plasmids using PlasmidFinder and
727 plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58:3895-3903.
728 https://doi.org/10.1128/AAC.02412-14
729
730
731
732
733
734
735
736
737
738
739
740

741 **Table 1** Confusion matrix showing the prediction accuracy of a Random Forest classifier for

742 source attribution of *S.* Typhimurium to bovine, porcine, poultry, and wild bird sources

Original training dataset					
Out-of-bag prediction accuracy: 82.9%					
Predicted class / Actual class	Bovine	Porcine	Poultry	Wild bird	Class accuracy
Bovine	126	38	29	2	64.62%
Porcine	24	300	10	4	88.76%
Poultry	24	11	401	4	91.14%
Wild bird	10	8	14	36	52.94%
Original training dataset with addition of the 131 wild bird isolates					
Out-of-bag prediction accuracy: 84.3%					
Predicted class / Actual class	Bovine	Porcine	Poultry	Wild bird	Class accuracy
Bovine	122	41	28	4	62.56%
Porcine	26	300	8	4	88.76%
Poultry	24	12	400	4	90.91%
Wild bird	12	7	14	166	83.42%

743

744

745

746

747

748

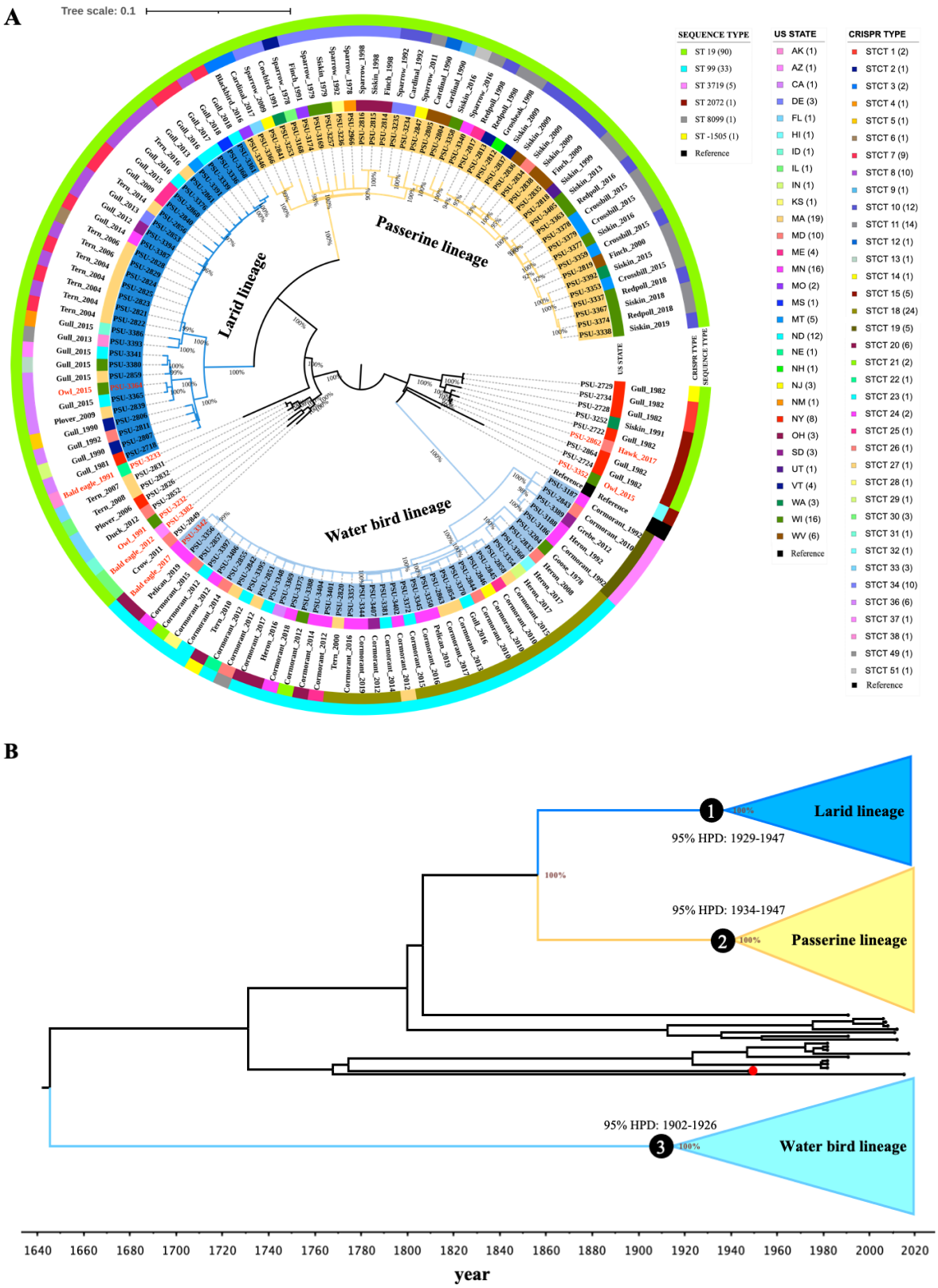


FIG 1 Phylogenetic analyses reveal distinct *S.* Typhimurium lineages in passerine birds, water birds, and larids. (A) Maximum-likelihood phylogenetic tree of 131 *S.* Typhimurium isolates from wild birds collected between 1978-2019 in 30 U.S. states. The tree is visualized using iTOL. Three

752 large lineages are defined by bird types, i.e., passerine lineage (yellow), larid lineage (blue), and
753 water bird lineage (light blue). Bootstrap values are displayed as percentage on tree branch. The
754 labels at tips represent the isolate name, the inner color strip represents the U.S. state from which
755 the sample originated, the outmost color strips represent the *S.* Typhimurium CRISPR type and
756 MLST sequence type, and the text labels between the color strips represent the bird host and
757 isolation year. The legend fields at the right of the tree represent the *S.* Typhimurium MLST
758 sequence type (number of isolates), U.S. state name (number of isolates), and the *S.* Typhimurium
759 CRISPR type (number of isolates). The labels highlighted in red represent isolates from raptors.
760 (B) Time-scaled Bayesian phylogenetic tree of 131 *S.* Typhimurium isolates from wild birds
761 collected between 1978-2019 in 30 U.S. states. The tree is visualized using FigTree v1.4.4.
762 Numbers in black circles at the nodes represent the 95% highest posterior probability density
763 (HPD) for the times of most recent common ancestor for the water bird lineage (light blue), larid
764 lineage (blue), and passerine lineage (yellow). The red circle at the tree tip represents the reference
765 strain LT2 (isolation year: 1948). Posterior probability values are displayed as percentage on tree
766 nodes of the three lineages.
767
768
769
770
771
772
773

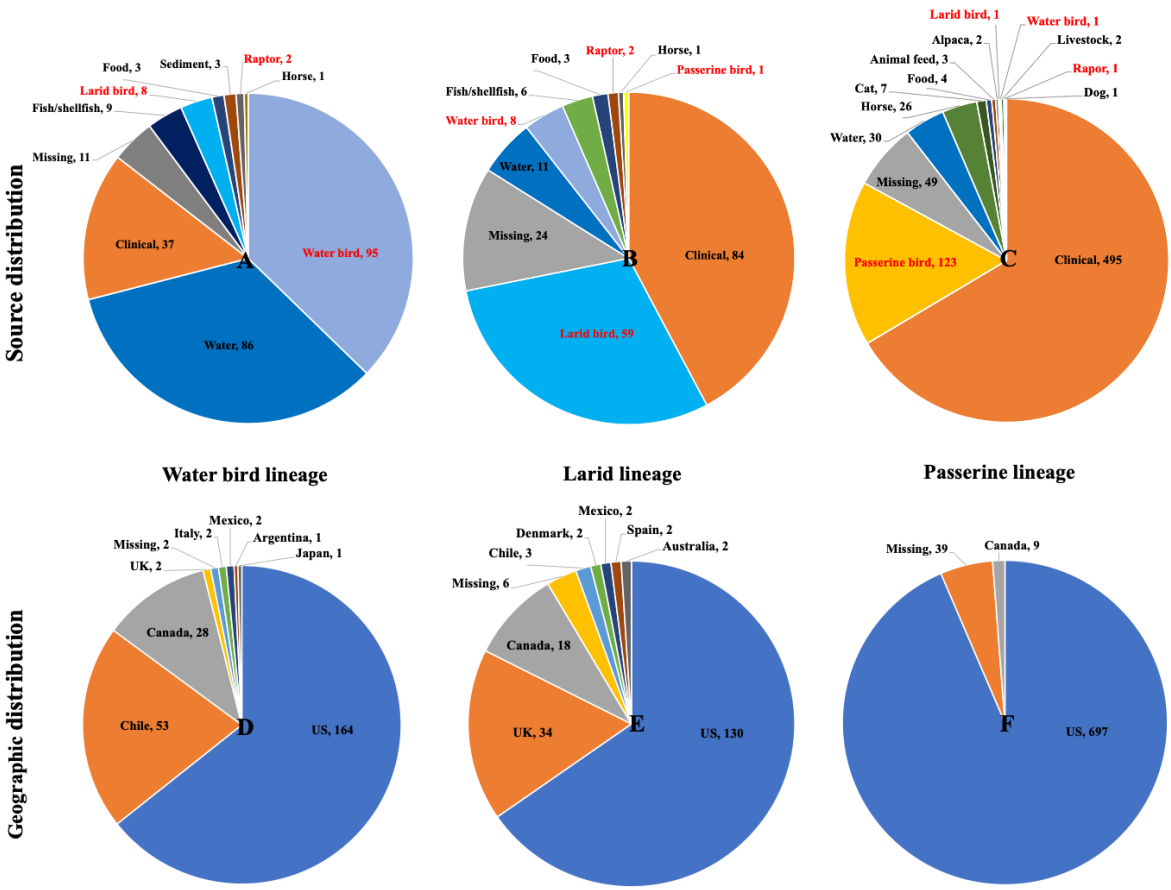
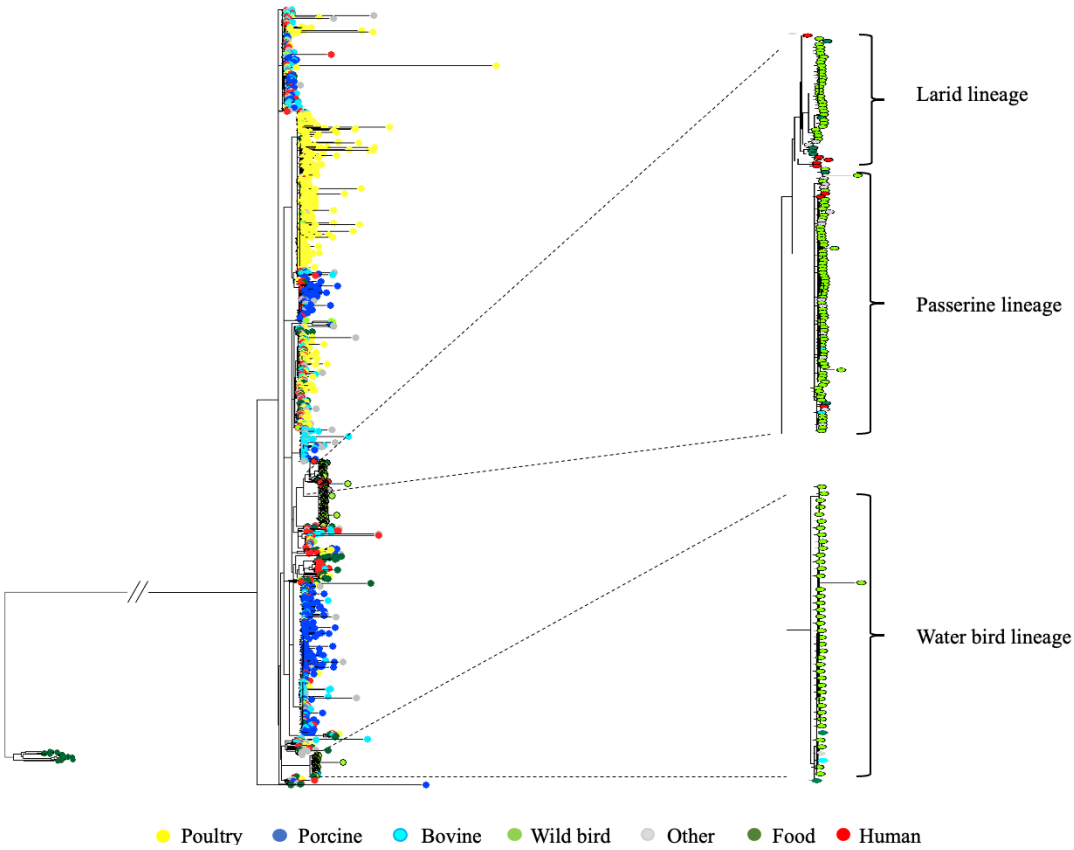


FIG 2 Distribution of source niches and geographic regions for the isolates in the Pathogen Detection (PD) database that cluster with *S*. Typhimurium isolates from different wild bird lineages. (A-C) Source distribution of isolates in the PD database clustering with isolates from water bird lineage, larid lineage, and passerine lineage. (D-F) Geographic distribution of isolates in the PD database clustering with isolates from water bird lineage, larid lineage, and passerine lineage. The number following the source niche or geographic region represents the number of isolates from that specific source or geographic region in the PD database. Complete data used to generate the pie charts are given in **Table S2**.



785 **FIG 3** Maximum-likelihood phylogenetic tree of 1,604 *S.* Typhimurium isolates from different
786 sources for zoonotic source prediction. The maximum-likelihood phylogenetic tree for zoonotic
787 source prediction includes *S.* Typhimurium isolates from seven source classes: (1) bovine (*n* =
788 195), including isolates from cattle, beef, and raw milk; (2) porcine (*n* = 338), including isolates
789 from pigs and pork; (3) poultry (*n* = 440), including isolates from chickens, turkeys, ducks, and
790 their eggs; (4) wild bird (*n* = 199), including 131 isolates from this study; (5) human (*n* = 171),
791 including human clinical isolates; (6) food (*n* = 83), including seafood, plant-based food, and other
792 ready-to-eat and/or processed food; (7) other (*n* = 178), including any isolates not belonging to
793 the aforementioned classes. The larid, passerine, and water bird lineages formed by the *S.*
794 Typhimurium isolates from wild birds are enlarged and highlighted in the phylogenetic tree.
795

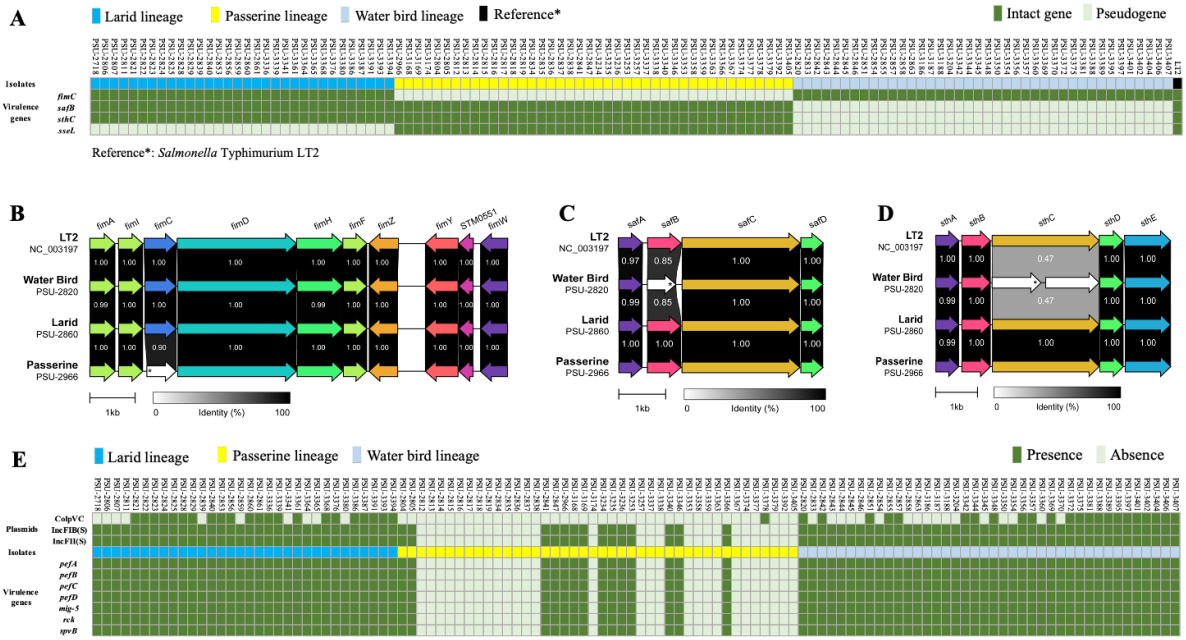


FIG 4 Chromosomal and plasmid-associated virulence gene signatures that differ among the larid, passerine, and water bird lineages. (A) Distribution of chromosomal virulence gene signatures (*fimC*, *safB*, *sthC*, and *sseL*) among the larid, passerine, and water bird lineages. (B-D) Organization of the fimbrial operons containing *fimC*, *safB*, and *sthC* in isolates from the larid, passerine, and water bird lineages, and reference strain *S.* Typhimurium LT2. (E) Distribution of plasmids and plasmid virulence genes among the larid, passerine, and water bird lineages. In the heatmaps (A) and (E): Light green = pseudogene in (A), and absence of the gene or plasmid in (E); Dark green = intact gene in (A), and presence of the gene or plasmid in (E); Blue = isolates from larid lineage; Yellow = isolates from passerine lineage; Light blue = isolates from water bird lineage; Dark = reference strain (*S.* Typhimurium LT2). Isolate names are represented on the x axis, virulence genes and plasmid names are represented on the y axis. Complete data used to generate the heatmaps is given in **Table S3**. In the fimbrial operons (B), (C), and (D), the pseudogenes are marked as white arrows, and the locations of indels in the pseudogenes are represented by asterisks.

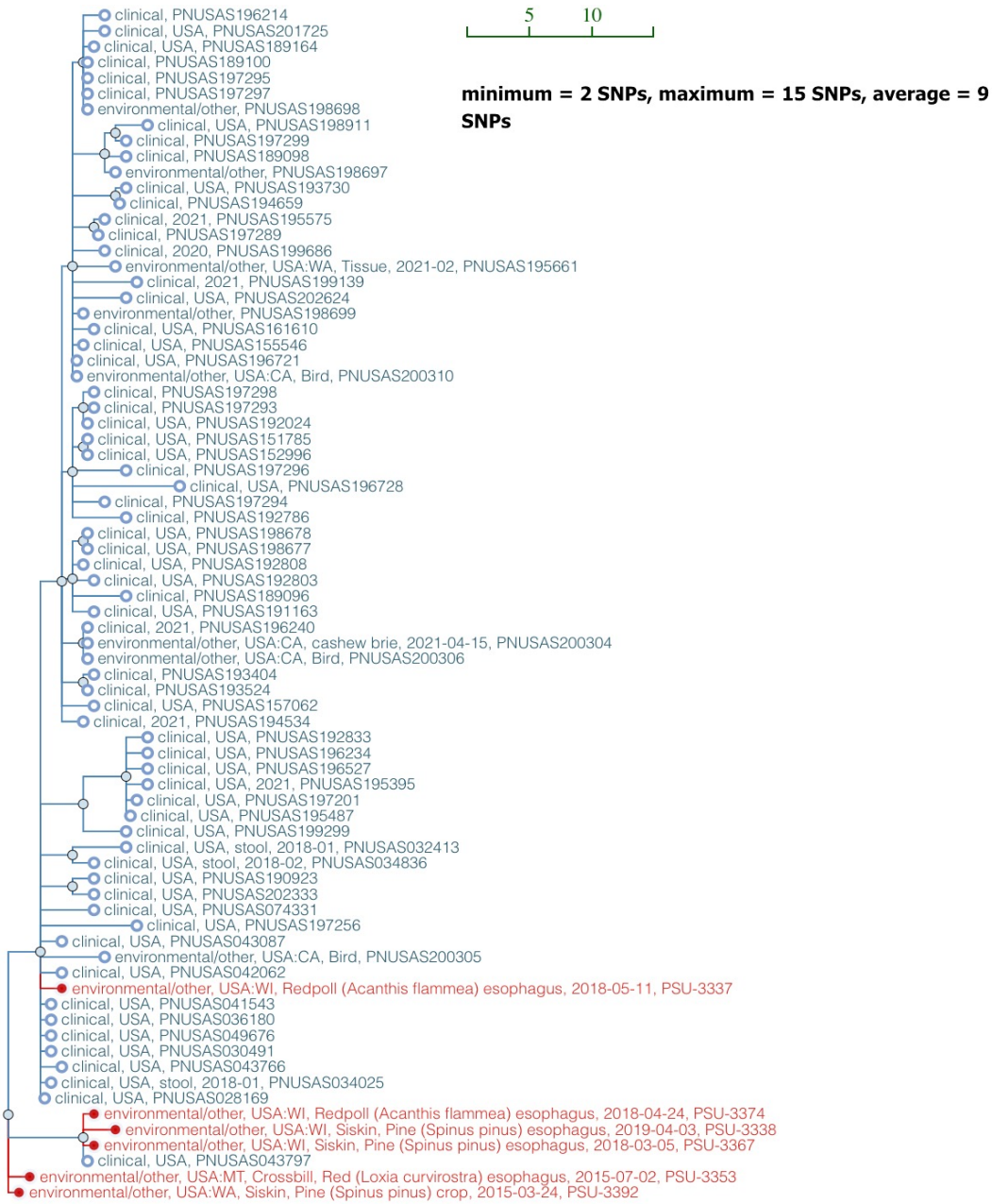


FIG 5 Close genetic relatedness between passerine isolates from this study and clinical isolates from a 2021 salmonellosis outbreak in the U.S. Single nucleotide polymorphism (SNP) tree is automatically generated in the NCBI Pathogen Detection database when sequencing data are submitted to the database (SNP cluster and accession link: PDS000013825.81). The scale bar represents the unit SNP distance. The average SNP distance for all isolates in the tree is nine. *S.* Typhimurium isolates in the SNP tree from this study are highlighted in red.

817 **Legends for Supplementary Materials**

818 **FIG S1** Comparison of SseL from different wild bird isolates. Multiple sequence alignment of
819 SseL amino acid sequences from the larid, passerine, and water bird lineages. SseL from *S.*
820 Typhimurium LT2 served as a reference.

821

822 **Table S1** Metadata information (isolate name, SRA accession number, isolation year, state, bird
823 host, CRISPR type, and 7-gene MLST sequence type) of the 131 *S.* Typhimurium isolates from
824 wild birds collected between 1978-2019 in 30 U.S. states.

825

826 **Table S2** SNP clusters generated by the NCBI Pathogen Detection when relating the isolates from
827 larid, passerine, and water bird lineages to other isolates in the database.

828

829 **Table S3** *In silico* antimicrobial resistance, virulence, and plasmid profiles of the 131 *S.*
830 Typhimurium isolates from wild birds collected between 1978-2019 in 30 U.S. states.

831

832

833

834

835

836

837

838