

Phylogeny and Divergence of the 100 Most Common *Salmonella* Serovars Available in the NCBI Pathogen Detection Database

Figure Legends for Supplementary Figures 96 – 99

[±] For the analysis of antigenic formula diversification events, the maximum likelihood phylogenetic trees of *Salmonella* Enteritidis and Infantis were re-run using 301 reference serovars used in this study (available at GitHub: https://github.com/ly276/USDA_Salmonella_Phylogeny_Project/tree/main/Reference_Isolates).

Supplementary Figure 96. *S. Enteritidis* [±]

Antigenic formula diversification comparisons of the eight lineages of *S. Enteritidis* (i.e., Enteritidis A-H) with its corresponding closely-related serovars. Closely-related serovars share the same most recent common ancestor as the *S. Enteritidis* lineages. Differences in somatic (O) and (H) flagellar antigens between closely-related serovars and *S. Enteritidis* lineages are shown in red. Bootstrap values have been provided for the ancestor nodes. Antigenic formula comparisons were made if the bootstrap value between Enteritidis lineages and closely-related serovar was >0.7. Closely-related serovars and their antigenic formulas are shown explicitly. Brown and gray triangles represent genomes in the target serovar (i.e., *S. Enteritidis*) and the non-Enteritidis (distantly-related) serovars, respectively. Enteritidis S1 and S2 are stand-alone singletons that do not cluster with any Enteritidis lineages.

^α The lineage is paraphyletic.

^β The lineage is monophyletic.

* *S. Gallinarum* clusters within Enteritidis A. However, the clustering could not be explicitly shown in the tree due to the figure's size.

Not analyzed because it is a stand-alone singleton.

Commented [BY1]: Needs to replace with Typhimurium. Also check text.

The interpretation for the antigenic formula: underlined O factors () – determined by phage conversion; the genome should be lysogenized with the specific converting phage; curly brackets ({ }) – O antigens represented in curly brackets cannot coexist with the other antigens in curly brackets; square brackets ([]) – O or H antigens that are present or absent with no relation to phage conversion; brackets (()) – O or H antigens that are weakly agglutinable.

Supplementary Figure 97. *S. Newport*

Antigenic formula diversification comparisons of the four lineages of *S. Newport* (i.e., Newport A-D) with its corresponding closely-related serovars. Closely-related serovars share the same most recent common ancestor as the *S. Newport* lineages. Differences in somatic (O) and (H) flagellar antigens between closely-related serovars and *S. Newport* lineages are shown in red. Bootstrap values have been provided for the ancestor nodes. Antigenic formula comparisons were made if the bootstrap value between Newport lineages and closely-related serovar was >0.7. Closely-related serovars and their antigenic formulas are shown explicitly. Green and gray triangles represent genomes in the target serovar (i.e., Newport) and the non-Newport (distantly-related) serovars, respectively.

^β The lineage is monophyletic.

The interpretation for the antigenic formula: underlined O factors () – determined by phage conversion; the genome should be lysogenized with the specific converting phage; curly brackets ({ }) – O antigens represented in curly brackets cannot coexist with the other antigens in curly brackets; square brackets ([]) – O or H antigens that are present or absent with no relation to phage conversion; brackets (()) – O or H antigens that are weakly agglutinable.

Supplementary Figure 98. *S. I 4,[5],12:i:-*

Antigenic formula diversification comparisons of the four lineages of *S. I 4,[5],12:i:-* (i.e., I 4,[5],12:i:- A-D) with its corresponding closely-related serovars. Closely-related serovars share the same most recent common ancestor as the *S. I 4,[5],12:i:-* lineages. Differences in somatic (O) and (H) flagellar antigens between closely-related serovars and *S. I 4,[5],12:i:-* lineages are shown in red. Bootstrap values have been provided for the ancestor nodes. Antigenic formula comparisons were made if the bootstrap value between I 4,[5],12:i:- lineages and closely-related serovar was >0.7. Closely-related serovars and their antigenic formulas are shown explicitly. Purple and gray triangles represent genomes in the target serovar (i.e., *S. I 4,[5],12:i:-*) and the non-I 4,[5],12:i:- (distantly-related) serovars, respectively. I 4,[5],12:i:- S1 stand-alone singleton that does not cluster with any I 4,[5],12:i:- lineages.

^a The lineage is paraphyletic.

^b The lineage is monophyletic.

* Typhimurium clusters within 4,[5],12:i:- A. However, the clustering could not be explicitly shown in the tree due to the figure's size. In addition, as the comparison for the antigen divergence was made for *S. Typhimurium* and *S. I 4,[5],12:i:-* in Figure 4, this comparison was not reanalyzed here.

Not analyzed because it is: either i) a stand-alone singleton (i.e., I 4,[5],12:i:- S1), or ii) monophyletic lineages that share an MRCA with many serovars (i.e., I 4,[5],12:i:- B and D). The interpretation for the antigenic formula: underlined O factors () – determined by phage conversion; the genome should be lysogenized with the specific converting phage; curly brackets ({ }) – O antigens represented in curly brackets cannot coexist with the other antigens in curly

brackets; square brackets ([]) – O or H antigens that are present or absent with no relation to phage conversion; brackets (()) – O or H antigens that are weakly agglutinable.

Supplementary Figure 99. *S. Infantis* [±]

Antigenic formula diversification comparisons of the two lineages of *S. Infantis* (i.e., *Infantis* A-B) with its corresponding closely-related serovars. Closely-related serovars share the same most recent common ancestor as the *S. Infantis* lineages. Differences in somatic (O) and (H) flagellar antigens between closely-related serovars and *S. Infantis* lineages are shown in red. Bootstrap values have been provided for the ancestor nodes. Antigenic formula comparisons were made if the bootstrap value between *Infantis* lineages and closely-related serovar was >0.7. Closely-related serovars and their antigenic formulas are shown explicitly. Pink and gray triangles represent genomes in the target serovar (i.e., *Infantis*) and the non-*Infantis* (distantly-related) serovars, respectively. *Infantis* S1 stand-alone singleton that does not cluster with any *Infantis* lineages.

^a The lineage is paraphyletic.

^β The lineage is monophyletic.

[#] Not analyzed because it is a stand-alone singleton (i.e., *Infantis* S1).

The interpretation for the antigenic formula: underlined O factors () – determined by phage conversion; the genome should be lysogenized with the specific converting phage; curly brackets ({ }) – O antigens represented in curly brackets cannot coexist with the other antigens in curly brackets; square brackets ([]) – O or H antigens that are present or absent with no relation to phage conversion; brackets (()) – O or H antigens that are weakly agglutinable.