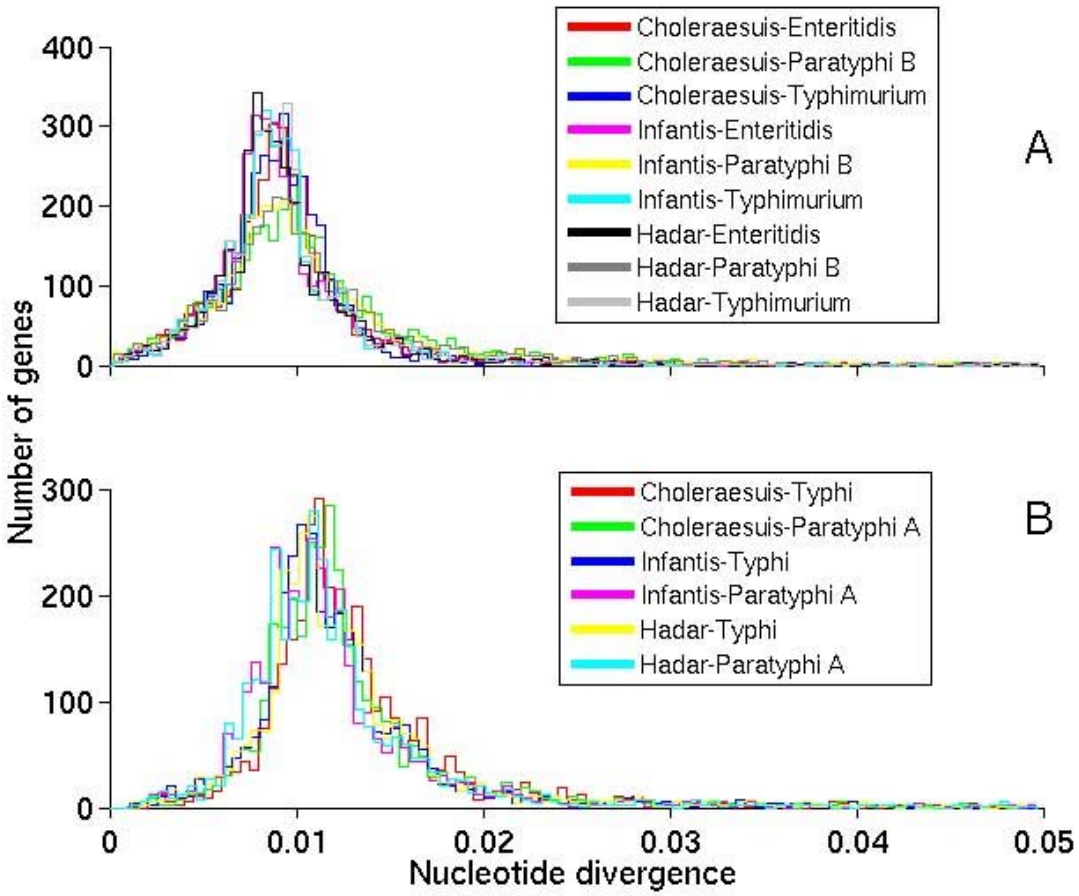
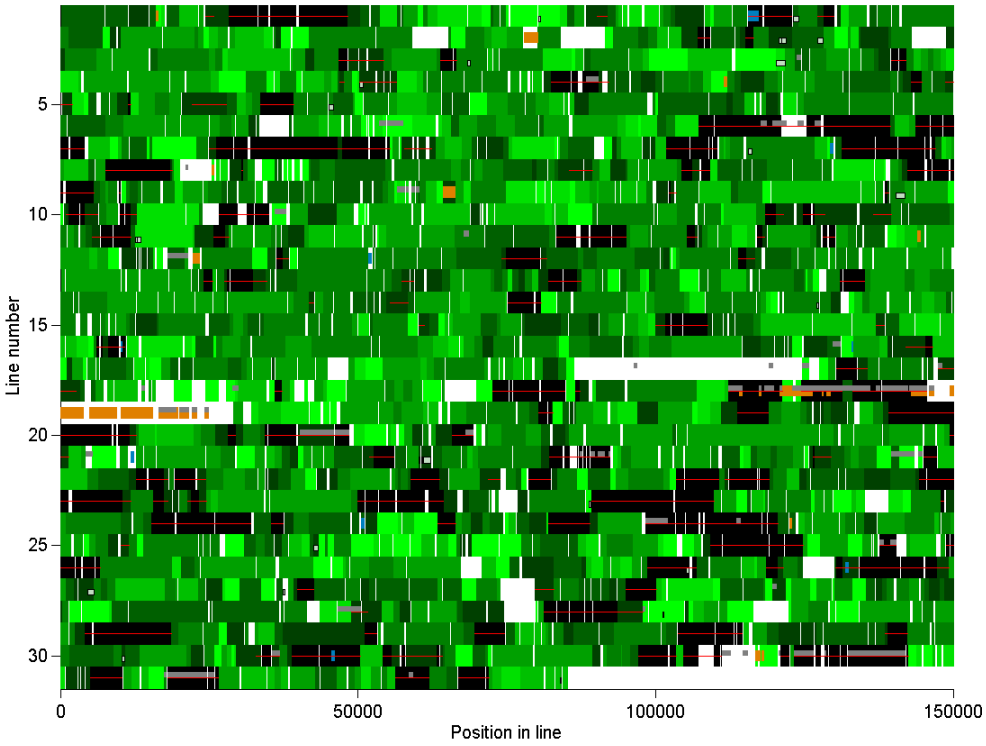
Supplemental Online Material



Supplemental Figure S1: Equivalent to Fig. 1A and B but for three additional genomes.



Supplemental Figure S2: Divergeome of Paratyphi A in comparison with Typhi. Details as for Fig. 2.

Supplemental Table S1

List of functional classes of genes. Shows the number of genes found for each category (Column D) in the Typhi genome (Column E) and the proportion of genes in the low divergence regions (Column F). Columns A-C show the classification code of each category according to the prokaryotic protein classification scheme used within the Sanger Institute Pathogen Sequencing Unit.

Supplemental Table S2

List of all genes for which either the Paratyphi A or Typhi genome possesses less than 0.3% nucleotide divergence from the Enteritidis, Typhimurium, Paratyphi B, Choleraesuis, Hadar, Infantis or Dublin genomes as estimated by the changepoint model. The table is divided in three parts: the first two contain the ribosomal and non-ribosomal genes for which Paratyphi A and Typhi themselves differ by less than 0.3% and which are therefore unlikely to represent imports. The third part represents the genes for which Paratyphi A and Typhi are more than 0.3% diverged, and have been grouped into contiguous blocks that represent putative imports from one source. The size of most of these imports (Column H) is too small for the import to be statistically credible. The location of these genes on Fig. 2 is indicated in Columns I and J.

Supplemental Table S3

For each of the 7 genes in the MLST scheme, the table lists serovars that have 3 or less nucleotide differences to Typhi or Paratyphi A in the MLST fragment (Column C and D). 6 of the 7 genes are in high divergence regions on the comparison of Typhi and Paratyphi

A, the exception being *hisD*. The number of differences between Typhi and Paratyphi A as well as the length of the gene are shown in Column E and F respectively. Dublin, Enteritidis, Typhimurium, Hadar, Infantis, Choleraesuis and Paratyphi B are often amongst the most closely related serovars. Thus, there are no other serovars which represent obvious candidates to have contributed 3-6 million base pairs to Typhi and Paratyphi A as required by the divergence hypothesis based on this MLST data.

Supplemental Table S4

List of all genes annotated in Typhi, with the distance to Paratyphi A as estimated by our change-point model (Column E) and the position of the gene on Fig. 2 (Columns G and H).