

Organism	Number of raw Illumina reads	Number of Illumina reads after filtering	Number of reads after MLST
<i>Salmonella</i> Agona	100	100	100
<i>Salmonella</i> Anatum	100	100	100
<i>Salmonella</i> Braenderup	100	100	100
<i>Salmonella</i> Derby	100	100	100
<i>Salmonella</i> Dublin	100	100	100
<i>Salmonella</i> Enteritidis	100	99	99
<i>Salmonella</i> Hadar	100	100	100
<i>Salmonella</i> Heidelberg	100	100	100
<i>Salmonella</i> Infantis	100	100	100
<i>Salmonella</i> Javiana	100	100	100
<i>Salmonella</i> Johannesburg	100	100	100
<i>Salmonella</i> Kentucky	100	100	100
<i>Salmonella</i> Mbandaka	100	100	100
<i>Salmonella</i> Montevideo	100	100	100
<i>Salmonella</i> Muenchen	100	96	96
<i>Salmonella</i> Newport	100	98	98
<i>Salmonella</i> Schwarzengrund	100	100	100
<i>Salmonella</i> Senftenberg	100	100	100
<i>Salmonella</i> Thompson	100	100	100
<i>Salmonella</i> Typhimurium	100	100	100
<i>Acinetobacter baumannii</i>	1000	901	343
<i>Campylobacter jejuni</i>	1000	996	995
<i>Clostridium difficile</i>	1000	839	833
<i>Enterococcus faecium</i>	1000	915	913
<i>Escherichia coli</i>	1000	895	895
<i>Haemophilus influenzae</i>	1000	951	949
<i>Helicobacter pylori</i>	1000	874	874
<i>Klebsiella pneumoniae</i>	1000	920	918
<i>Listeria monocytogenes</i>	1000	998	996
<i>Mycobacterium tuberculosis</i>	1000	809	790
<i>Neisseria gonorrhoeae</i>	1000	915	915
<i>Pseudomonas aeruginosa</i>	1000	956	954
<i>Staphylococcus aureus</i>	1000	998	998
<i>Streptococcus pneumoniae</i>	1000	940	940
<i>Enterococcus faecalis</i>	100	92	92
<i>Enterococcus faecium</i>	100	92	92
<i>Helicobacter pylori</i>	100	81	81

Table S4. Number of reads used for calculating percentage of agreement between mlst and stringMLST for 23 *Salmonella enterica* serovars and 14 phylogenetic divergent bacterial pathogens.

The datasets used to calculate percentage of agreement contained 100 or 1,000 paired-end raw Illumina reads downloaded from NCBI-SRA (first column in the Table). Before running mlst, a few steps were required, such as quality trimming and adapter clipping, *de novo* assembly and assembly filtering. After the filtering steps were completed, a fraction of all genomes were discarded from the mlst analyses. This number is shown in the second column in the Table. Additionally, mlst infrequently adds empty columns in its results, and these records were removed before conducting the final statistical analyses to calculate the percentage of agreement between programs (third column in the Table). All downloaded genomes were freely available at NCBI-SRA.