**Supplementary Table 1: List of 46 Bacteria genomes**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Bacterial Species Name** | **Size (Mb)** | **Genes** | **Reference genome** | **RefSeq** |
| **PATHOGENIC** | | | | |
| *Mycobacterium tuberculosis* | 4.4 | 4,008 | [ASM19595v2](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000195955.2/) | GCF\_000195955.2 |
| *Bacillus anthracis* | 5.5 | 5,757 | [ASM844v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000008445.1/) | GCF\_000008445.1 |
| *Bartonella tribocorum* | 2.6 | 2,332 | [ASM19643v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000196435.1/) | GCF\_000196435.1 |
| *Bordetella pertussis* | 4.1 | 3,975 | [ASM400897v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_004008975.1/) | GCF\_004008975.1 |
| *Bartonella rochalimae* | 1.5 | 1,333 | [Bart\_roch\_BMGH\_V1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000706645.1/) | GCF\_000706645.1 |
| *Bartonella henselae* | 1.9 | 1,628 | [PRJEB4971\_assembly\_1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000612965.1/) | GCF\_000612965.1 |
| *Escherichia coli* | 5.6 | 5,417 | [ASM886v2](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000008865.2/) | GCF\_000008865.2 |
| *Brucella ovis* | 3.3 | 3,187 | [ASM1684v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000016845.1/) | GCF\_000016845.1 |
| *Campylobacter coli* | 1.7 | 1,775 | [ASM973039v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_009730395.1/) | GCF\_009730395.1 |
| *Campylobacter concisus* | 1.8 | 1,907 | [ASM129846v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001298465.1/) | GCF\_001298465.1 |
| *Campylobacter curvus* | 2.0 | 1,975 | [ASM1337212v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_013372125.1/) | GCF\_013372125.1 |
| *Campylobacter jejuni* | 1.6 | 1,668 | [ASM908v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000009085.1/) | GCF\_000009085.1 |
| *Campylobacter rectus* | 2.6 | 2,502 | [ASM480379v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_004803795.1/) | GCF\_004803795.1 |
| *Clostridioides difficile* | 4.1 | 3,716 | [ASM1888508v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_018885085.1/) | GCF\_018885085.1 |
| *Corynebacterium jeikeium* | 2.5 | 2,182 | [ASM660v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000006605.1/) | GCF\_000006605.1 |
| *Bordetella bronchiseptica* | 5.3 | 5,104 | ASM155905v2 | GCF\_001559055.1 |
| *Granulibacter bethesdensis* | 2.7 | 2,439 | ASM188944v1 | GCF\_001889445.1 |
| *Prevotella nigrescens* | 2.9 | 2,449 | [ASM1812782v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_018127825.1/) | GCF\_018127825.1 |
| *Pseudomonas aeruginosa* | 6.4 | 5,697 | [ASM676v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000006765.1/) | GCF\_000006765.1 |
| *Renibacterium salmoninarum* | 3.2 | 3,266 | [ASM1888v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000018885.1/) | GCF\_000018885.1 |
| *Salmonella enterica* | 5.0 | 4,717 | [ASM694v2](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000006945.2/) | GCF\_000006945.2 |
| *Shigella flexneri* | 4.8 | 4,788 | [ASM692v2](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000006925.2/) | GCF\_000006925.2 |
| *Staphylococcus aureus* | 2.8 | 2,872 | [ASM1342v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000013425.1/) | GCF\_000013425.1 |
| *Streptococcus iniae* | 2.1 | 2,002 | ASM30091v1 | GCF\_000300915.1 |
| *Streptococcus sobrinus* | 2.1 | 2,069 | [ASM317297v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_003172975.1/) | GCF\_003172975.1 |
| *Yersinia pestis* | 4.7 | 4,234 | ASM22297v1 | GCF\_000222975.1 |
| *Haemophilus influenzae* | 1.8 | 1,836 | [ASM93157v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000931575.1/) | GCF\_000931575.1 |
| *Helicobacter pylori* | 1.6 | 1,544 | [ASM1782153v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_017821535.1/) | GCF\_017821535.1 |
| *Lactococcus garvieae* | 2.1 | 2,110 | [ASM1602669v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_016026695.1/) | GCF\_016026695.1 |
| *Stenotrophomonas maltophilia* | 4.5 | 4,105 | 44087\_C01 | GCF\_900475405.1 |
| *Citrobacter amalonaticus* | 5.1 | 4,884 | [ASM155893v2](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001558935.2/) | GCF\_001558935.2 |
| *Clostridium\_autoethanogenum* | 4.4 | 4,044 | ASM148472v1 | GCF\_001484725. |
| *Providencia alcalifaciens* | 4.0 | 3,721 | [ASM239350v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_002393505.1/) | GCF\_002393505.1 |
| *streptococcus agalactiae* | 2.1 | 2,076 | ASM155203v1 | GCF\_001552035.1\_ |
| *Streptococcus suis* | 2.2 | 2,136 | ASM2674v1 | GCF\_000026745.1 |
| **NON-PATHOGENIC** | | | | |
| *Fannyhessea vaginae* | 1.5 | 1,203 | [58174\_B01](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_900445305.1/) | GCF\_900445305.1 |
| *Cryobacterium soli* | 4.4 | 4,088 | ASM361103v1 | GCF\_003611035.1 |
| *Cutibacterium acnes* | 4.78 | 2,513 | [ASM37670v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000376705.1/) | GCF\_000376705.1 |
| *Xanthomonas hortorum* | 5.2 | 4,507 | [ASM96121v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000961215.1/) | GCF\_000961215.1 |
| *Xanthomonas vesicatoria* | 5.3 | 4,604 | [ASM190872v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001908725.1/) | GCF\_001908725.1 |
| *Lysinibacillus irui* | 4.7 | 5,104 | ASM2887747v1 | GCF\_028877475.1 |
| *Streptococcus oralis* | 1.9 | 1,868 | [46338\_H01](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_900637025.1/) | GCF\_900637025.1 |
| *Citrobacter amalonaticus* | 4.7 | 4,381 | ASM440394v1 | GCF\_004403945.1 |
| *Vibrio harveyi* | 5.9 | 5,371 | [ASM77011v2](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000770115.1/) | GCF\_000770115.1 |
| *Prochlorococcus marinus* | 1.7 | 1,924 | [ASM1566v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000015665.1/) | GCF\_000015665.1 |
| *Rhodococcus aetherivorans* | 6.6 | 5,960 | [ASM2572297v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_025722975.1/) | GCF\_025722975.1 |
| *Ruminococcus bovis* | 2.4 | 2,329 | [ASM560113v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_005601135.1/) | GCF\_005601135.1 |