**Table S1. Genome assembly information of two strains of *Toxoplasma gondii1***

|  |  |  |
| --- | --- | --- |
|  | ***Toxoplasma gondii* RH-88** | ***Toxoplasma gondii* ME49** |
| **Identifier for assembly** | [GCA\_013099955.1](https://www.ncbi.nlm.nih.gov/assembly/GCA_013099955.1) | [GCA\_000006565.2](https://www.ncbi.nlm.nih.gov/assembly/GCA_000006565.2) |
| **Assembly level** | chromosome | chromosome |
| **Size (Mb)** | 67.06 | 65.67 |
| **Number of chromosomes and plasmids** | 14 | 15 |
| **Number of scaffolds** | 197 | 2277 |
| **Release date** | 05/15/2020 | 08/02/2013 |

1. The information about genome assembly is from NCBI (<https://www.ncbi.nlm.nih.gov/genome/browse/#!/eukaryotes/30/>)

**Table S2. Genome annotation comparison of two *Toxoplasma gondii* strains**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Species and strains** | **Software** | **Genes** | **CDS** | **CDS with function** | **CDS without function** | **rRNA** | **tRNA** | **Pseudogenes** | **Genes with same start and end position(%)1** | **Genes with same start position(%)** | **Similarity score2** |
| *Toxoplasma gondii* RH88 | Ref\_annotation**3** | 8654 | 8316 | 4078 (49.04%) | 4238 (50.96%) | 96 | 242 | / | / | / | / |
|  | Companion\_web4 | / | / | / | / | / | / | / | / | / | / |
|  | Companion\_cl | 11003 | 10455 | 2317 (22.16%) | 8138 (77.84%) | 262 | 268 | 1674 | 4371 (50.51%) | 4624 (53.41) | 47.05% |
|  | GeneSAS\_genemarkES | 6714 | 6714 | / | / | / | / | / | 2037 (23.53%) | 3147 (36.36%) | 40.96% |
|  | GAL | 33964 | 33964 | 27073 (79.71%) | 6891 (20.29%) | / | / | / | 602 (6.96%) | 2207 (25.50%) | 10.35% |
|  | GAAP | 26415 | 26415 | / | / | / | / | / | 502 (5.80%) | 1940 (22.42%) | 11.06% |
| *Toxoplasma gondii* ME49 | Ref\_annotation | 8925 | 8292 | 4008 (48.34%) | 4284 (51.66%) | 424 | 183 | / | / | / | / |
|  | Companion\_web | 4996 | 4488 | 2441 (54.39%) | 508 (11.32%) | 301 | 193 | 2162 | 1639 (18.36%) | 1976 (22.14%) | 28.39% |
|  | Companion\_cl | 11297 | 10520 | 2151 (20.45%) | 8369 (79.55%) | 566 | 191 | 1109 | 1067 (12.06%) | 1488 (16.67%) | 14.72% |
|  | GeneSAS\_genemarkES5 | / | / | / | / | / | / | / | / | / | / |
|  | GAL | 34288 | 34288 | 27368 (79.82%) | 6920 (20.18%) | / | / | / | 116 (1.30% | 796 (8.92%) | 3.68% |
|  | GAAP | 26204 | 26204 | / | / | / | / | / | 96 (1.08%) | 545 (6.11%) | 3.10% |

1. Percentage is equal to (Genes with same start and end position / Ref\_annotation genes )\*100

2. SimilarityScore = ((Genes with same start position) / (Totalx + Totalz)) \* 2 \* 100. Totalx and Totalz are the total numbers of genes in the software annotation and reference annotation. The function is from BEACON (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4539851/)

3. ref\_annotation means the reference annotation. It is from NCBI RefSeq.

4. The dataset size of strain RH88 is bigger than the maximum size of the uploaded dataset in Companion webserver, so Companion webserver can not annotate the genome of RH88.

5. GeneSAS can not annotate the genome of strain ME49. Because GeneSAS requires assembly dataset must contain less than 25000 total sequences and more than 50% of the sequences must be over 2500 bases in length. The dataset of ME49 violates the rule.