**Table S1.** Properties and features of previously published RB51 sequence by Ma *et al.* (2014), as compared to the RB51 sequence presented in this manuscript.

|  |  |  |
| --- | --- | --- |
| **Feature** | **Properties of RB51 Sequence by Ma *et al*. (2014)** | **Properties of RB51 Sequence described in this manuscript** |
| Number of Scaffolds | 36 | 2 |
| Number of Contigs | 150 | 2 |
| Number of Nonsense SNPs identified | 77 (position not reported)a | 1 in a pseudogene |
| Nonsynonymous SNPs reported | 41 (position not reported with exception of G 🡪 A transition reported in *capD*)a | 5 (all positions reported); *capD*, *rpoB*, and 3 novel locations (not including pseudogenes or putative genes) |
| Indels reported | 8 (only two positions reported; genes BAB2\_0582 and BAB1\_0055)a | 8 (all locations reported, including *narJ* indel) |

aAuthors indicate that the gene sequence was compared with the *B. abortus* 2308 sequence as reported in Tsolis *et al.* (2002).

Supplementary Reference:

Tsolis, R. M., 2002 Comparative genome analysis of the alpha-proteobacteria: relationships between plant and animal pathogens and host specificity. Proc. Natl. Acad. Sci. 99:12503-12505.

**Figure S1.** Alignment of the *wboA* gene sequence from *B. abortus* 2308 (reference strain; sequence AF1077681, *Brucella abortus* *wboA* gene, complete cds, 2114 nt in length) to the *wboA* gene sequence from NZ\_AQIE00000000.1.



**Table S2. Comparison of *Brucella abortus* RB51 genome features with other *Brucella abortus* biovar-1 genomes.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **RB51** | **2308** | **9-941** | **S19** |
| **Chr I** | **Chr II** | **Chr I** | **Chr II** | **Chr I** | **Chr II** | **Chr I** | **Chr II** |
| Refseq # | Pending | Pending | NC\_007618 | NC\_007624 | NC\_006932 | NC\_006933 | NC\_010742 | NC\_010740 |
| GenBank | Pending | Pending | AM040264 | AM040265 | AE017223 | AE017224 | CP000887 | CP000888 |
| Size | 2,122,197 | 1,161,994 | 2,121,359 | 1,156,948 | 2,124,241 | 1,162,204 | 2,122,487 | 1,161,449 |
| Genes | 2148 | 1112 | 2236 | 1182 | 2199 | 1156 | 2056 | 1077 |
| Coding | 2098 | 1094 | 2000 | 1034 | 2029 | 1055 | 1967 | 1033 |
| Pseudo | 188 | 132 | 186 | 130 | 123 | 84 | 39 | 25 |
| G + C | 57.2 | 57.3 | 57.2 | 57.3 | 57.2 | 57.3 | 57.2 | 57.3 |
| RNA operons | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |
| tRNA | 41 | 14 | 41 | 14 | 41 | 14 | 41 | 14 |
| Ref | This study | Chain *et al*. 2005 | Halling *et al*. 2005 | Crasta *et al.* 2008 |

Supplementary References:

Chain, P.S., D. J. Comerci, M.E. Tolmasky, F. W. Larimer, S. A. Malfatti, *et al*., 2005 Whole-genome analyses of speciation events in pathogenic Brucellae. Infect. Immun. 73: 8531-8361.

Crasta, O. R., O. Folkerts, Z. Fei, S. P. Mane, C. Evans, *et al*., 2008 Genome sequence of *Brucella abortus* vaccine strain S19 compared to virulent strains yields candidate virulence genes. PLoS One. 3: e2193.

Halling, S. M., B. D. Peterson-Burch, B. J. Bricker, R. L. Zuerner, Z. Qing, *et al*., 2005 Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*. J. Bacteriol. 187: 2715-2726.

**Table S3**. LPS synthesis gene mutations identified in RB51, analyzed in the context of *B. abortus* 2308.

|  |  |  |  |
| --- | --- | --- | --- |
| **Gene** | **RB51 Mutation (compared to 16M)** | ***B. melitensis* Gene Name** | **Amino Acid Identity in *B. abortus* 2308 Sequence** |
| *wboB* | L116S, R324Q | BMEI0997 | S, Q |
| *wboA* | Insertion | BMEI0998 | **Insertion not present** |
| *wa-1326* | S93L | BMEI1326 | L |
| *wbkE* | None | BMEI1393 | - |
| *manAAog* | None | BMEI1394 | - |
| *manBOag* | L381F | BMEI1395 | F |
| *manCOag* | D132G, F133L, L273F | BMEI1396 | G, L, F |
| *wbkA* | N349K | BMEI1404 | K |
| *gmd* | None | BMEI1413 | - |
| *per* | None | BMEI1414 | - |
| *wzm* | L152P | BMEI1415 | P |
| *wzt* | D171A | BMEI1416 | A |
| *wbkB* | None | BMEI417 | - |
| *wbkC* | None | BMEI1418 | - |
| *wbkF* | None | BMEI1426 | - |
| *wbkD* | T88I, V434I, E559K, K607E | BMEI1427 | I, I, **E**, E |
| *pgm* | H251R | BMEI1886 | R |
| wa (0053) | None | BMEII0053 | - |
| *manBcore* | None | BMEII0899 | - |

**Figure S2. BLAST sequence alignment between CapD (query) and the C-terminal domain of PgIF (subject).** The highly significant E-value (3e-82) and the relatively high percent ID (41%) indicate homology between the C-terminal regions of CapD and PgIF. Homology between CapD and PgIF extends to the full-length protein (Figure S4).

Query 262 RNVEIDDLLGRSPVPPDTTLLREVVEGRRIMITGAGGSIGSQLCLTIAQWNPAAIVLFES 321

 R++ I+DLL R P D + + ++ + ++++GAGG+IGS+LC ++ +++ +

Sbjct 2 RDISIEDLLARKPKDLDDSAVAAFLKDKVVLVSGAGGTIGSELCKQCIKFGAKHLIMVDH 61

Query 322 SEFALYQIDRQLRQFASCTVVPVLGSVRDRACVEKPIRDHSIDTVYHCAAYKHVPLVERN 381

 SE+ LY+I+ L + + P+L S+ D+ +++ ++ + + + H AAYKHVPL E+N

Sbjct 62 SEYNLYKINDDLNLYKE-KITPILLSILDKQSLDEVLKTYKPELILHAAAYKHVPLCEQN 120

Query 382 PLVGIFNNVFGTLEVAEAALNTDVERMVLISSDKAVRPTNVMGATKRWAELVIYYYGRLA 441

 P + NN+ GT + ++A V + V+IS+DKAVRPTN+MG TKR EL Y ++

Sbjct 121 PHSAVINNILGTKILCDSAKENKVAKFVMISTDKAVRPTNIMGCTKRVCEL---YTLSMS 177

Query 442 EQAGKKKAFYSVRFGNVLGSNGSVVPLFREQIANGGPVTLTHEDMTRYFMSIKEAAELIV 501

 ++ + VRFGNVLGS+GSV+P F+ QIAN P+TLTH D+ RYFM + EA +L++

Sbjct 178 DENFE---VACVRFGNVLGSSGSVIPKFKAQIANNEPLTLTHPDIVRYFMLVAEAVQLVL 234

Query 502 QSGAIAQSGDTVLLEMGEPVKIRDLAENMILLAGLTVRNEENPQGDIAIEVTGIREGEKM 561

 Q+GAIA+ G+ +L+MG+PVKI DLA+ M+LL+ RN D+ I++TG+R+GEK+

Sbjct 235 QAGAIAKGGELFVLDMGKPVKIIDLAKKMLLLSN---RN------DLEIKITGLRKGEKL 285

Query 562 YEELFYDPSLAQRTRHPKIMRAPQ 585

 YEEL D + A +T++ I A

Sbjct 286 YEELLIDENDA-KTQYESIFVAKN 308

Glu 559



Loop 2

Loop 1

**Figure S3.**  Superposition of the C-terminal domains of PgIF (5BJU) (orange) and CapD (CapD-C-term model) (green). Only the alpha carbons are shown for clarity. Significant deviations occur only in two loop regions (blue), which are located far from the active site residues of PgIF (not shown) and E559 of CapD, and are unlikely to affect function.

## **Figure S4.** BLASTsequence alignment between CapD (query) and PgIF (subject). The alignment covers 93% of the CapD sequence, and the identities (overall sequence ID is 33%) are concentrated in the C-terminal region.

Query 5 YPWQRLALLPRISKQIILVLSDCLLLLASAYLAFVVRFG----FVFVPNLAQLFLILIAP 60

 Y +RLA + SD +L+L S YLAF +RF +F + +IL+

Sbjct 4 YKSKRLAFF---------LTSDIVLILLSVYLAFSLRFSGDIPSIFYHGMMVSAIILL-- 52

Query 61 LLAIPVFIRFGLYRAIIRYLAERAIWSIFQATAVAALFWVALVFLMELYGSTGLPRSVPL 120

 +L + F +Y+ R+ + IF A +A + + + + + PRS +

Sbjct 53 VLKLSFLFVFRIYKVAWRFFSLNEARKIFIALLLAEFCFFLIFYFFSDFFNP-FPRSAIV 111

Query 121 LYWLLSTVFISASRFGAKWLLR-TAEHDKRYTSSALIIGIGEPARQLATALRSHSDTLV- 178

 + ++LS +FI R + L+ K + +++G A L + S L

Sbjct 112 IDFVLSYMFIGTLRISKRMLVDFKPSRMKEEETPCIVVGATSKALHLLKGAKEGSLGLFP 171

Query 179 VGFIDPAGQLAGMDIIGLRVYRTEEIPSLIENYGIKQVVVSEPALEQKERQEFARLLGRL 238

 VG +D +L G V E+I S +E G+K +++ + E++E +L L

Sbjct 172 VGVVDARKELIGTYCDKFIVEEKEKIKSYVEQ-GVKTAIIAL----RLEQEELKKLFEEL 226

Query 239 PVNTRILPPIADLTAGKYLVSALRNVEIDDLLGRSPVPPDTTLLREVVEGRRIMITGAGG 298

 + I D+ + + R++ I+DLL R P D + + ++ + ++++GAGG

Sbjct 227 -----VAYGICDVKIFSFTRNEARDISIEDLLARKPKDLDDSAVAAFLKDKVVLVSGAGG 281

Query 299 SIGSQLCLTIAQWNPAAIVLFESSEFALYQIDRQLRQFASCTVVPVLGSVRDRACVEKPI 358

 +IGS+LC ++ +++ + SE+ LY+I+ L + + P+L S+ D+ +++ +

Sbjct 282 TIGSELCKQCIKFGAKHLIMVDHSEYNLYKINDDLNLYKE-KITPILLSILDKQSLDEVL 340

Query 359 RDHSIDTVYHCAAYKHVPLVERNPLVGIFNNVFGTLEVAEAALNTDVERMVLISSDKAVR 418

 + + + + H AAYKHVPL E+NP + NN+ GT + ++A V + V+IS+DKAVR

Sbjct 341 KTYKPELILHAAAYKHVPLCEQNPHSAVINNILGTKILCDSAKENKVAKFVMISTDKAVR 400

Query 419 PTNVMGATKRWAELVIYYYGRLAEQAGKKKAFYSVRFGNVLGSNGSVVPLFREQIANGGP 478

 PTN+MG TKR EL Y ++++ + VRFGNVLGS+GSV+P F+ QIAN P

Sbjct 401 PTNIMGCTKRVCEL---YTLSMSDENFE---VACVRFGNVLGSSGSVIPKFKAQIANNEP 454

Query 479 VTLTHEDMTRYFMSIKEAAELIVQSGAIAQSGDTVLLEMGEPVKIRDLAENMILLAGLTV 538

 +TLTH D+ RYFM + EA +L++Q+GAIA+ G+ +L+MG+PVKI DLA+ M+LL+

Sbjct 455 LTLTHPDIVRYFMLVAEAVQLVLQAGAIAKGGELFVLDMGKPVKIIDLAKKMLLLSN--- 511

Query 539 RNEENPQGDIAIEVTGIREGEKMYEELFYDPSLAQRTRHPKIMRAPQ 585

 RN D+ I++TG+R+GEK+YEEL D + A +T++ I A

Sbjct 512 RN------DLEIKITGLRKGEKLYEELLIDENDA-KTQYESIFVAKN 551



Glu 559

Glu 257

**Figure S5.** The structural alignment of CapD-C-term (green) with CapE of *S. aureus* (orange). In CapE, the Glu 257 Ala mutation inhibits enzyme activity. The close superposition between E559 of CapD and E257 of CapE (PDB ID 3vvb.pdb) suggests that Glu 559 is also important for enzyme activity in CapD, and the Glu 559 Lys mutation is likely to inactivate CapD. Only alpha carbons are shown for clarity.