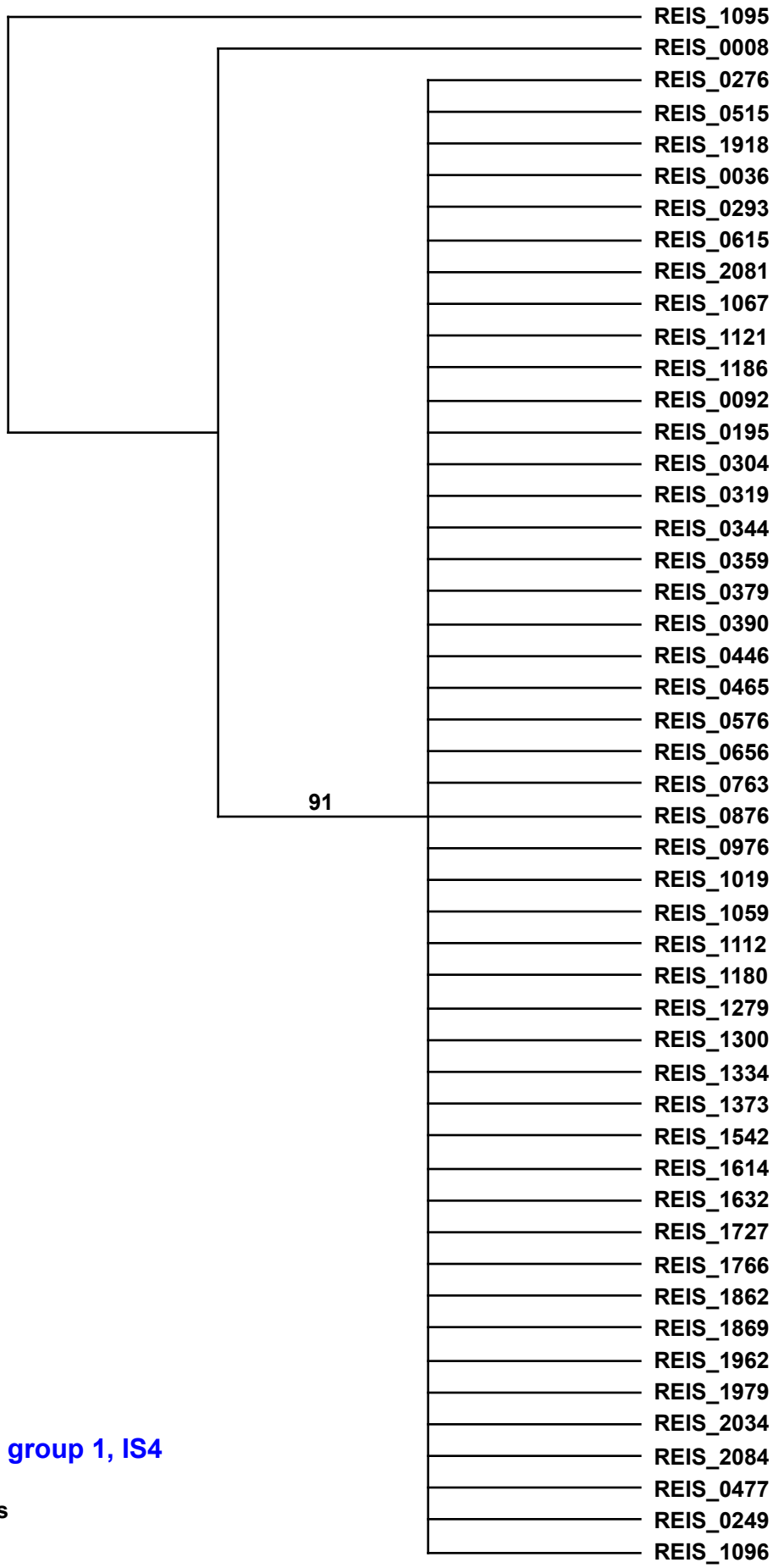
Fig. S8. Compilation of the most frequently occurring transposases in the REIS genome. (A) IS4 (49 sequences). (B) IS200-TNP17 (86 sequences). (C) IS4-TNP11 (54 sequences). (D) rve-COG3335 (43 sequences). For each analysis, top blastp subjects with significant alignments to the REIS queries were retrieved, with sequences less than half of the size of the query discarded. Remaining subjects were aligned using MUSCLE v3.6 [1, 2] (default parameters). Phylogenetic trees were estimated in PAUP* v4.0b10 (Altivec) under parsimony, implementing 500 random sequence additions, saving 100 trees per replication [3]. Majority rule consensus trees were constructed for each for analysis.

1. Edgar, R.C., *MUSCLE: a multiple sequence alignment method with reduced time and space complexity.* BMC Bioinformatics, 2004. **5**: p. 113.
2. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
3. Swofford, D., *PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4 ed.* , 1999, Sinauer: Sunderland, MA.

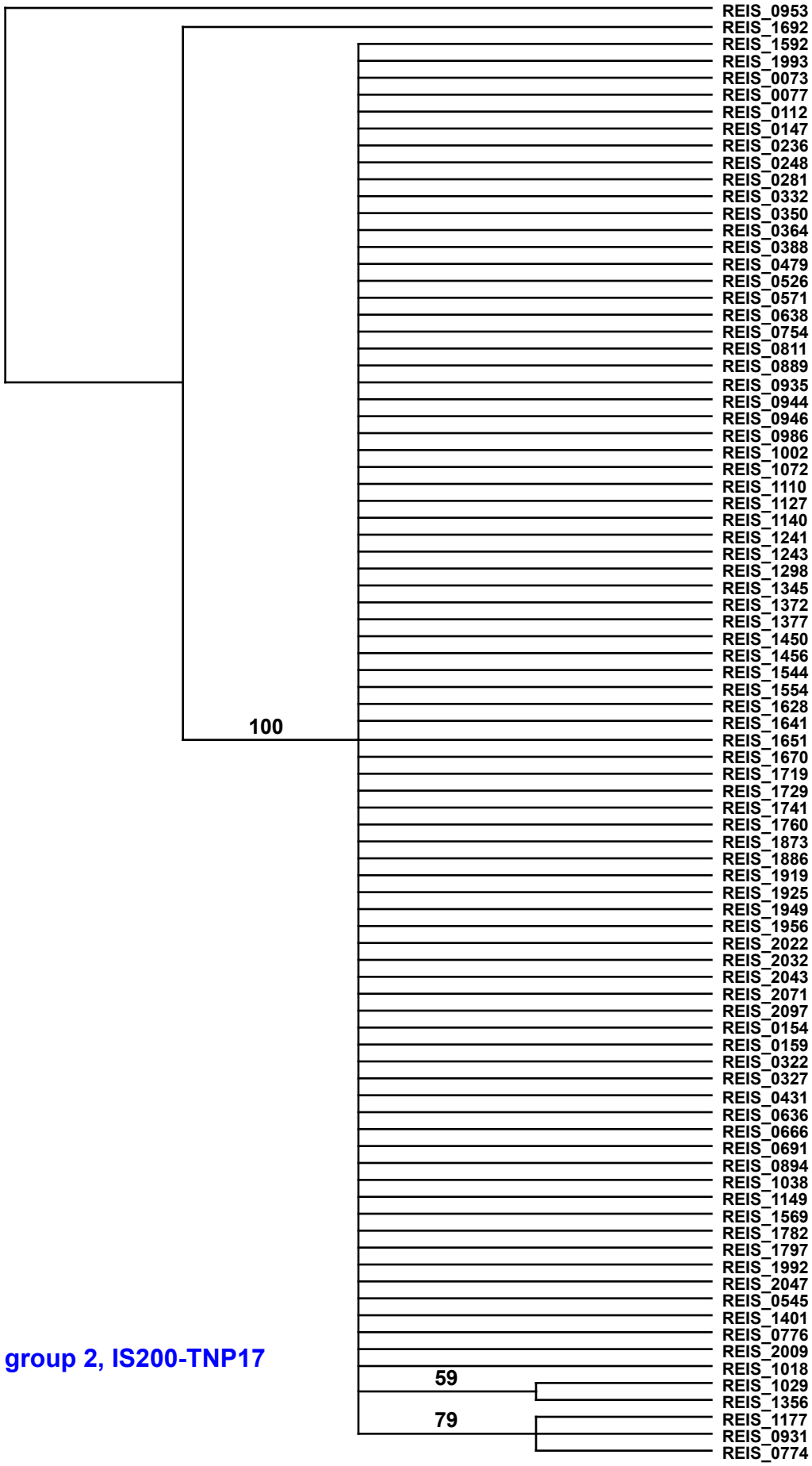
A



Transposase group 1, IS4

score: 111 steps
no. trees: 110

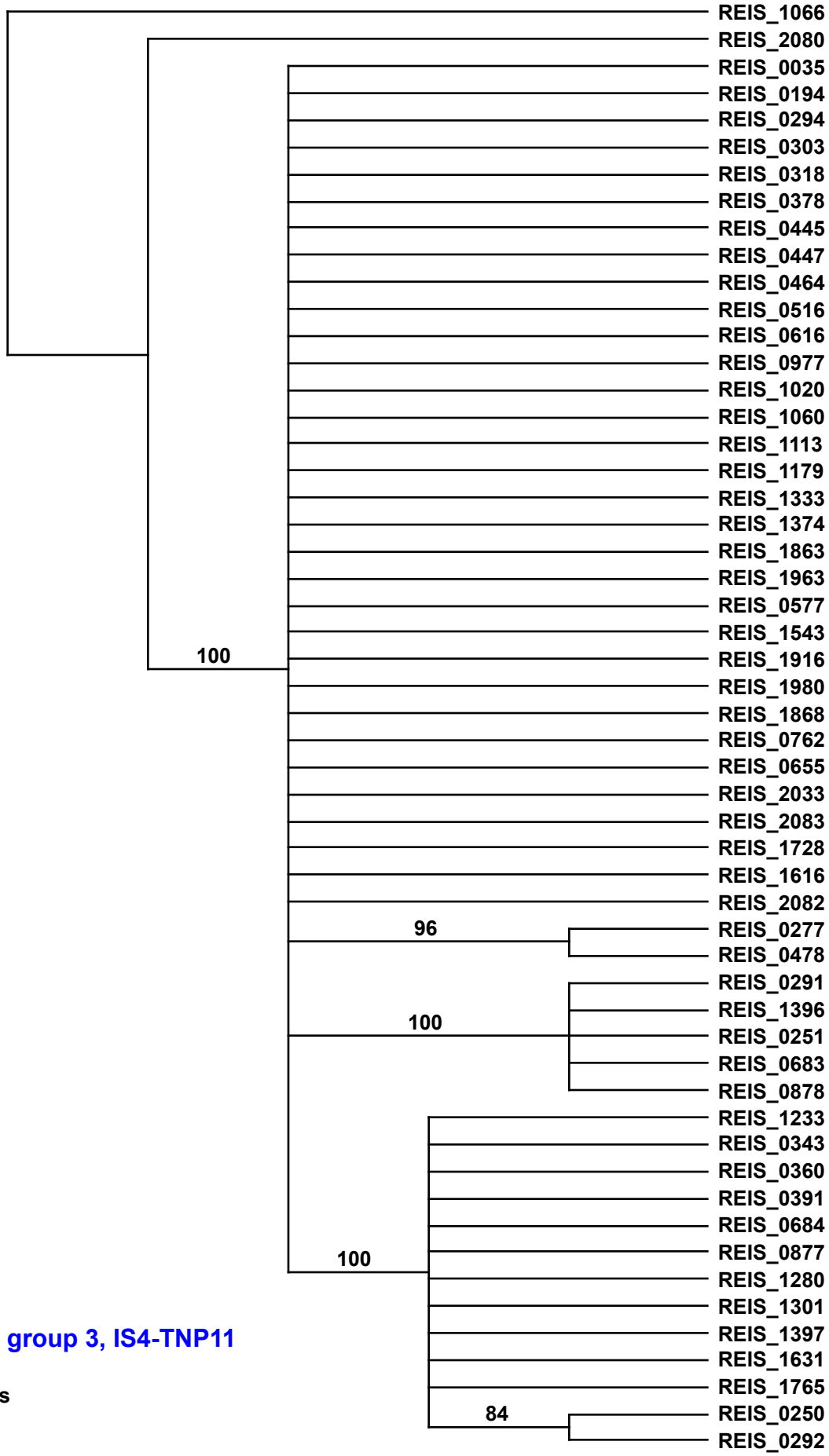
B



Transposase group 2, IS200-TNP17

score: 20 steps
no. trees: 290

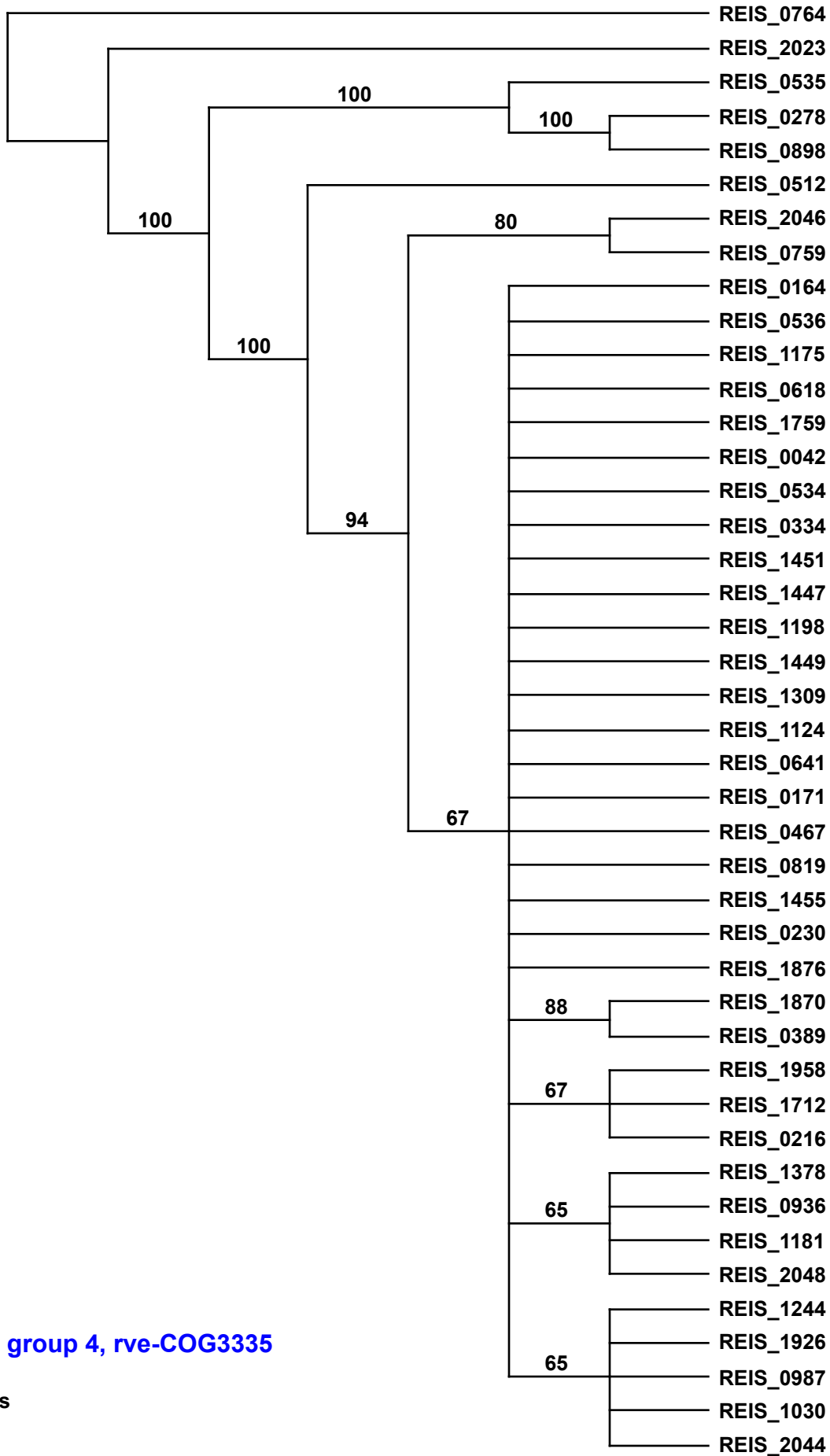
C



Transposase group 3, IS4-TNP11

score: 129 steps
no. trees: 1000

D



Transposase group 4, rve-COG3335

score: 565 steps
no. trees: 690

Fig. S9. Distribution of 317 transposases and inactive derivatives across other Rickettsiaceae genomes. Green numbers depict number of REIS proteins in each orthologous group (OG). Genome codes are as follows: Bg = *Orientia tsutsugamushi* str. Boryong, Ik = *O. tsutsugamushi* str. Ikeda, Br = *R. bellii* str. RML369-C, Bo = *R. bellii* str. OSU 85-389, Ca = *R. canadensis* str. McKiel, Fe = *R. felis* str. URRWXCal2, Ak = *R. akari* str. Hartford, Ma = *R. massiliae* str. MTU5, Ri = *R. rickettsii* str. Sheila Smith, Rw = *R. rickettsii* str. Iowa, Co = *R. conorii* str. Malish 7, Si = *R. sibirica* str. 246, and *R. africae* str. ESF-5. Note: No REIS transposase (TNP) homologs were detected in TG rickettsiae genomes. Some putative TNP sequences were identified by blastp searches using all full length REIS TNPs as queries. Distribution of TNPs across other Rickettsiaceae genomes was determined from OG clustering [1] (see Fig. S3).

1. Li, L., C.J. Stoeckert, Jr., and D.S. Roos, *OrthoMCL: identification of ortholog groups for eukaryotic genomes.* Genome Res, 2003. **13**(9): p. 2178-89.

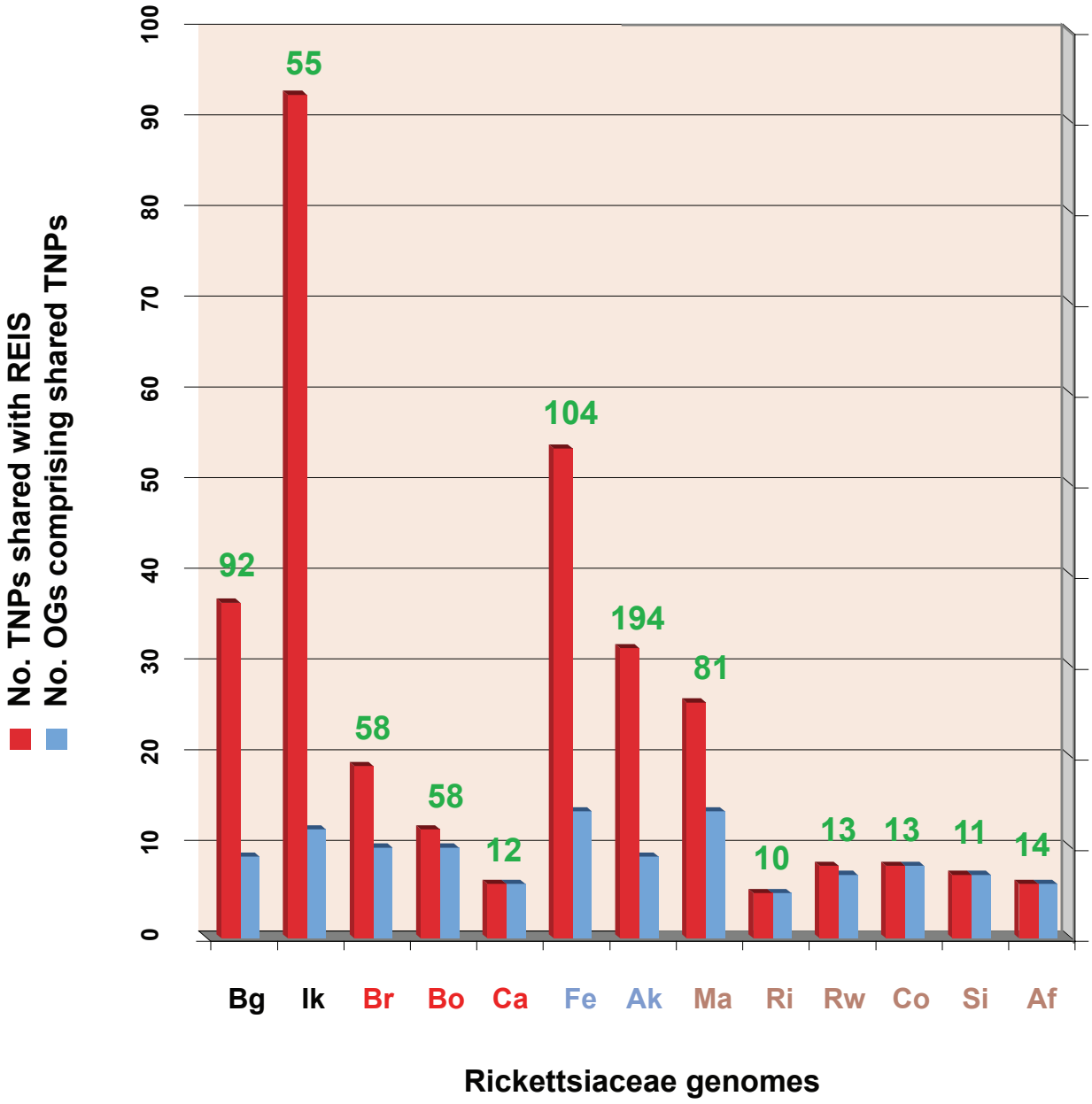
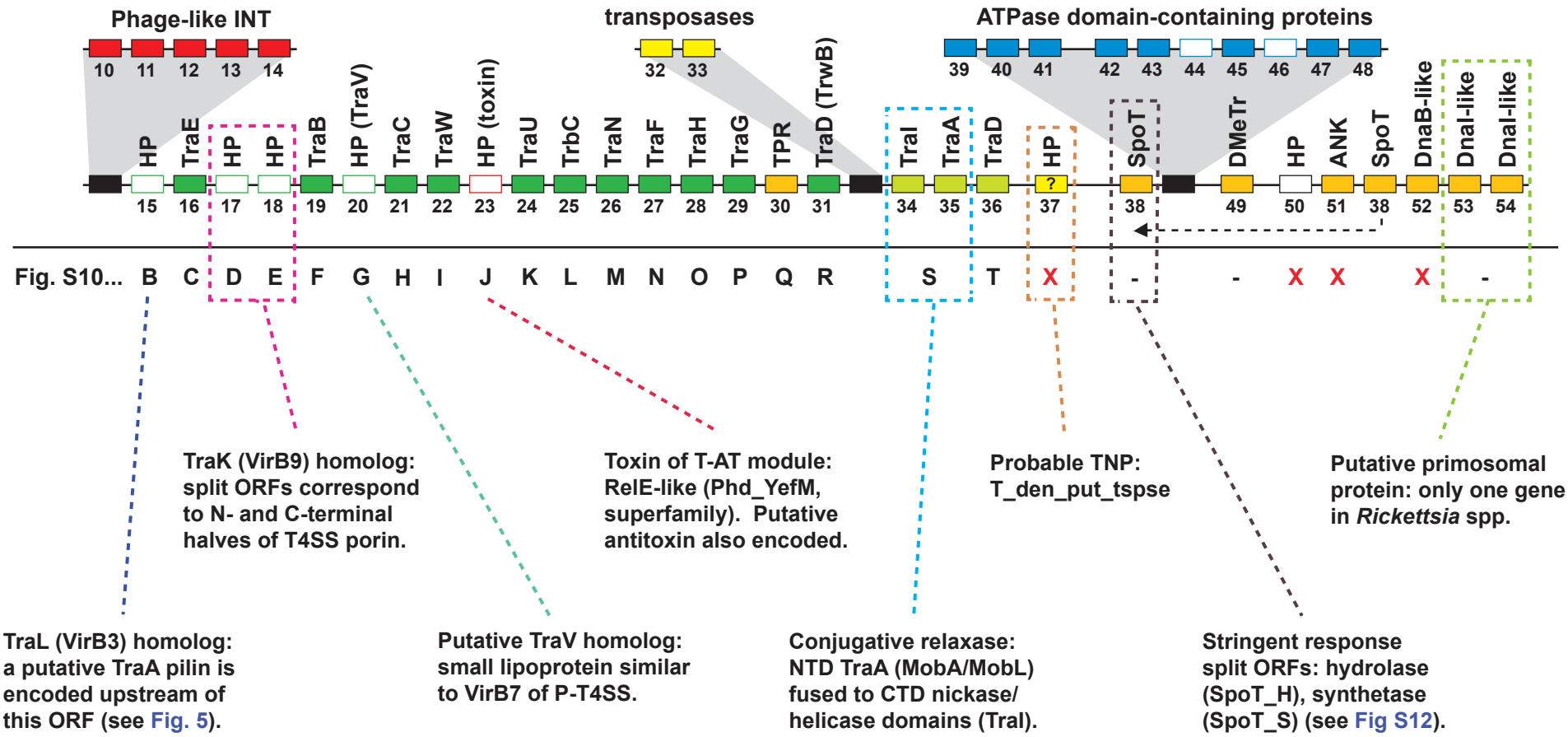

No. TNPs shared with REIS
No. OGs comprising shared TNPs
0
10
20
30
40
50
60
70
80
90
100
92
55
58
58
12
104
194
81
10
13
13
11
14
Bg
Ik
Br
Bo
Ca
Fe
Ak
Ma
Ri
Rw
Co
Si
Af
Rickettsiaceae genomes

Fig. S10. Characteristics of RAGEs encoded within Rickettsiaceae genomes. (A) Schema at top (above line) depicts the reconstructed *Orientia tsutsugamushi* amplified genetic element (OtAGE), with OtAGE components numbered 10-54 [1]. Coloring of RAGE components is explained at bottom. Important amendments to the OtAGE gene annotation, as well as major differences between the RAGE of *Rickettsia* spp. and *O. tsutsugamushi* are within dashed boxes, with descriptions provided. Letters below line in schema depict additional supplementary information for the F-T4SS genes (B-T, below). Hyphens denote RAGE components present across both *Rickettsia* spp. and *O. tsutsugamushi* for which phylogenies were not estimated. Red Xs denote RAGE components unique to the OtAGE. NOTE: the putative T_den_put_tspse transposase is present elsewhere in the REIS genome; similarly, the replicative DNA helicase (DnaB-like ORF) is present as a single copy gene in all *Rickettsia* genomes but is not a part of the RAGE. For comparison, the major RAGEs of *Rickettsia* spp. are shown in Fig. 5 of the text. (B-T) Individual phylogenetic trees estimated from the F-T4SS proteins encoded within the RAGEs. For each analysis, all blastp subjects (retrieved from searching within the Rickettsiales database, taxid:766) with significant alignments to the OtAGE queries were downloaded from NCBI and aligned with MUSCLE v3.6 using default parameters [2, 3] Phylogenetic trees were estimated in PAUP* v4.0b10 (Altivec) under parsimony [4]. Majority rule consensus trees were constructed for analyses generating multiple equally parsimonious trees, otherwise a phylogram of the single best tree is shown. The protein names associated with the RAGEs illustrated in Fig. 5 are colored blue, with additional REIS proteins from less conserved RAGEs colored orange. The highly similar REIS RAGE-Be and the RAGEs of *R. bellii* are encircled in a dashed red line in each of the estimated phylogenies. The tree estimated from the concatenation of these 19 datasets is shown in Fig. 5C.

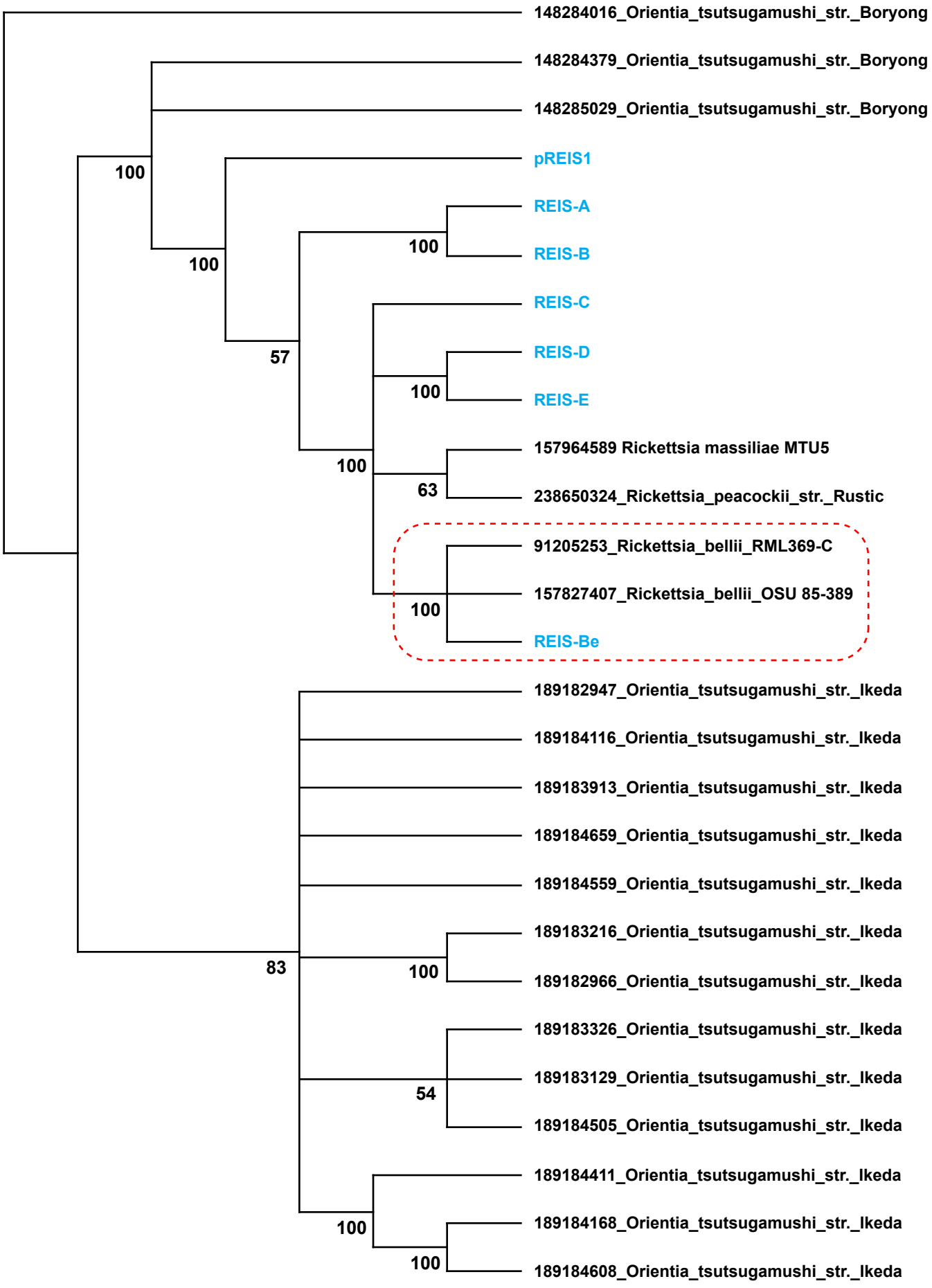
1. Nakayama, K., et al., *The Whole-genome sequencing of the obligate intracellular bacterium Orientia tsutsugamushi revealed massive gene amplification during reductive genome evolution.* DNA Res, 2008. **15**(4): p. 185-99.
2. Edgar, R.C., *MUSCLE: a multiple sequence alignment method with reduced time and space complexity.* BMC Bioinformatics, 2004. **5**: p. 113.
3. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
4. Swofford, D., *PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4 ed.* , 1999, Sinauer: Sunderland, MA.

A

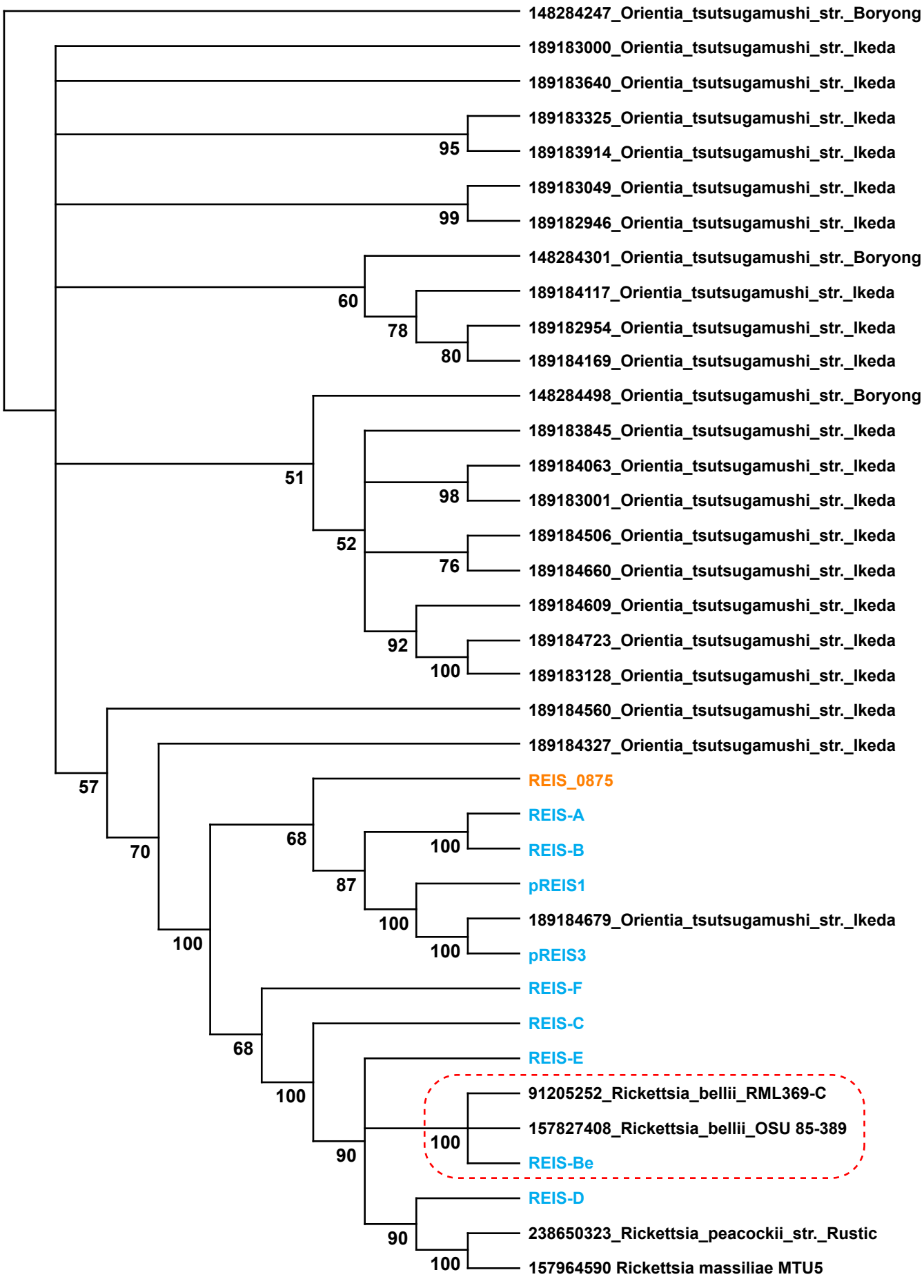


- Cassette-like component
- Integrase (phage-like)
- Transposase (TNP_31, YhgA-like)
- ORF with His kinase domain (HKD)
- HKD associated HP
- F-T4SS (*F* plasmid-like)
- F-T4SS (proposed here)
- Toxin of T-AT system (proposed here)
- F-T4SS (*Ti* plasmid-like)
- ? Probable transposase (T_den_put_tpse)
- HP protein (absent in *Rickettsia* RAGE)
- Conserved proteins associated with RAGE: ankyrin-repeat domain containing ORFs (ANK), tetratricopeptide repeat domain containing ORFs (TPR), DNA methyltransferase (type 12), stringent response regulators (SpoT), DnaB replicative helicases, and DnaI-like primosomal proteins

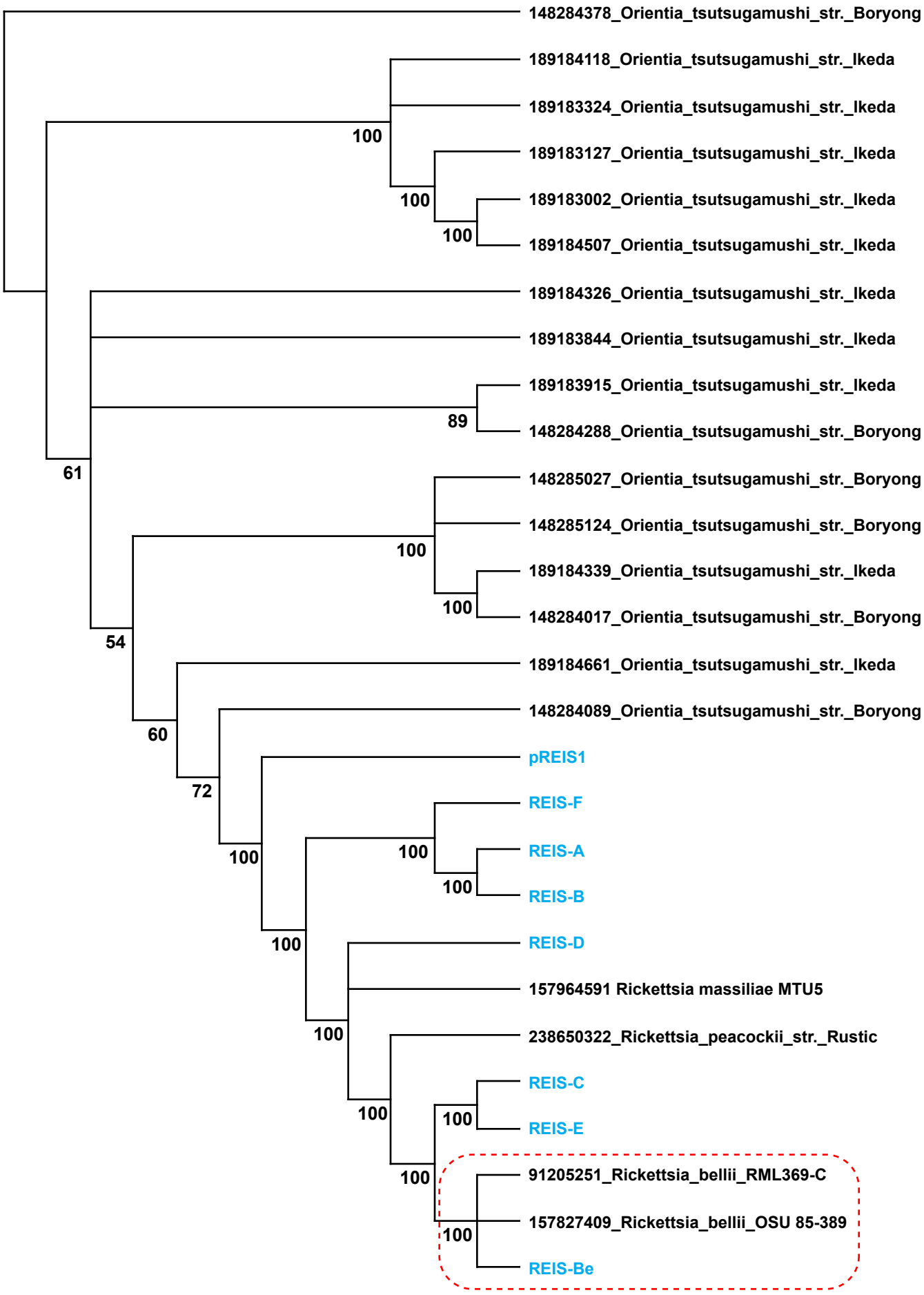
B **HP_15, putative TraL homolog (VirB3 in P-T4SS)**



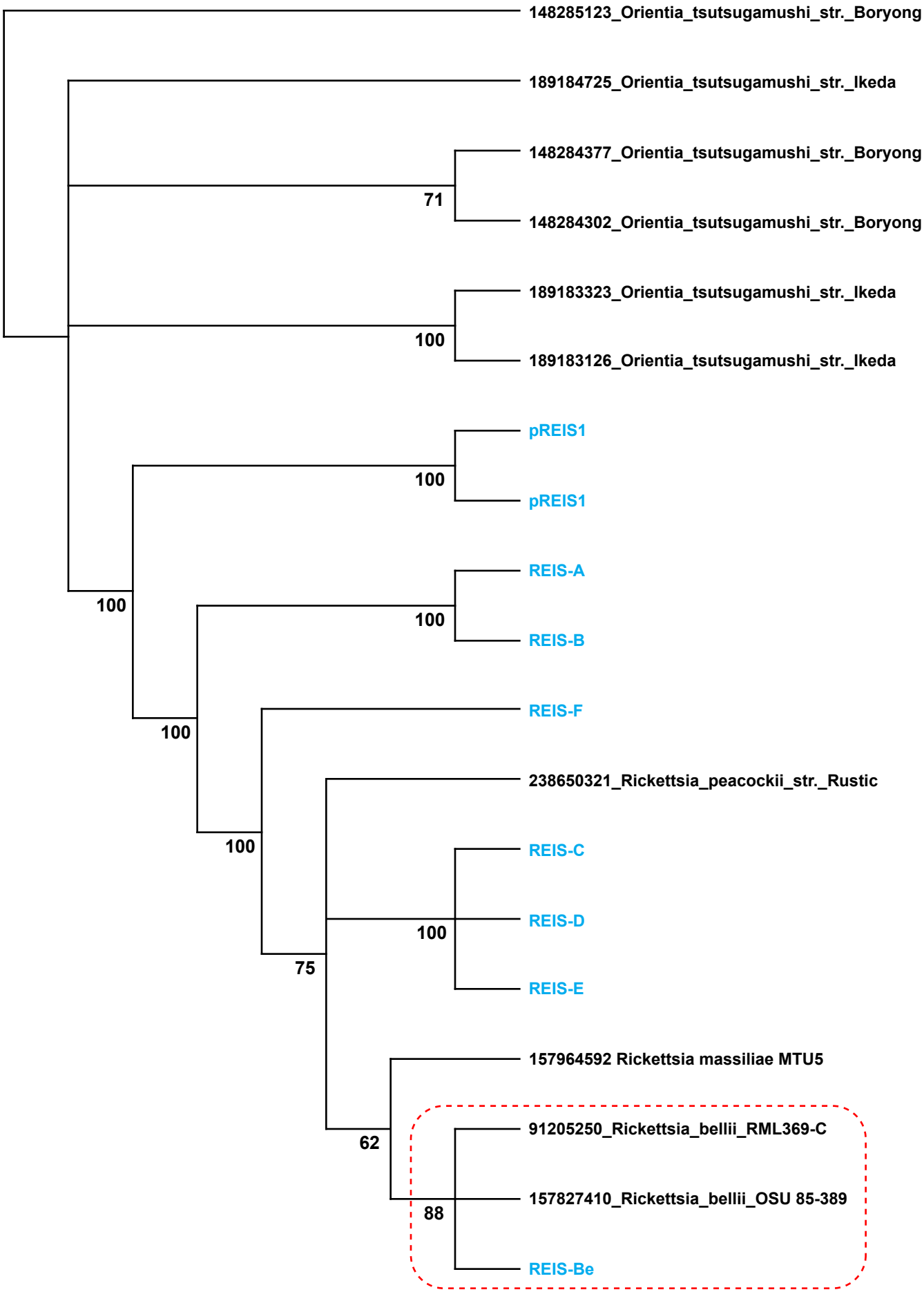
C **TraE (VirB5 in P-T4SS)**



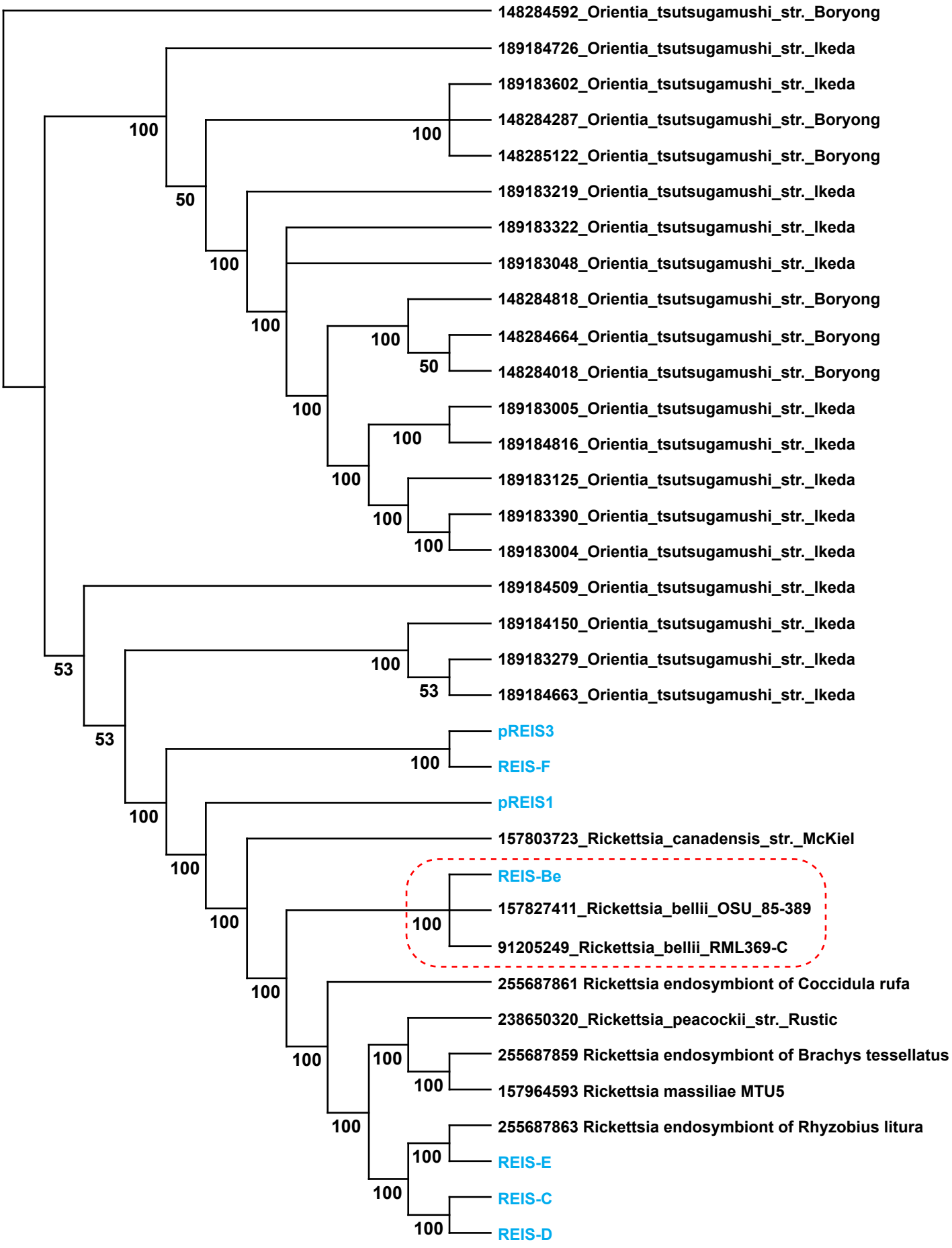
D **HP-17, putative N-terminal half of TraK (VirB9 of P-T4SS)**



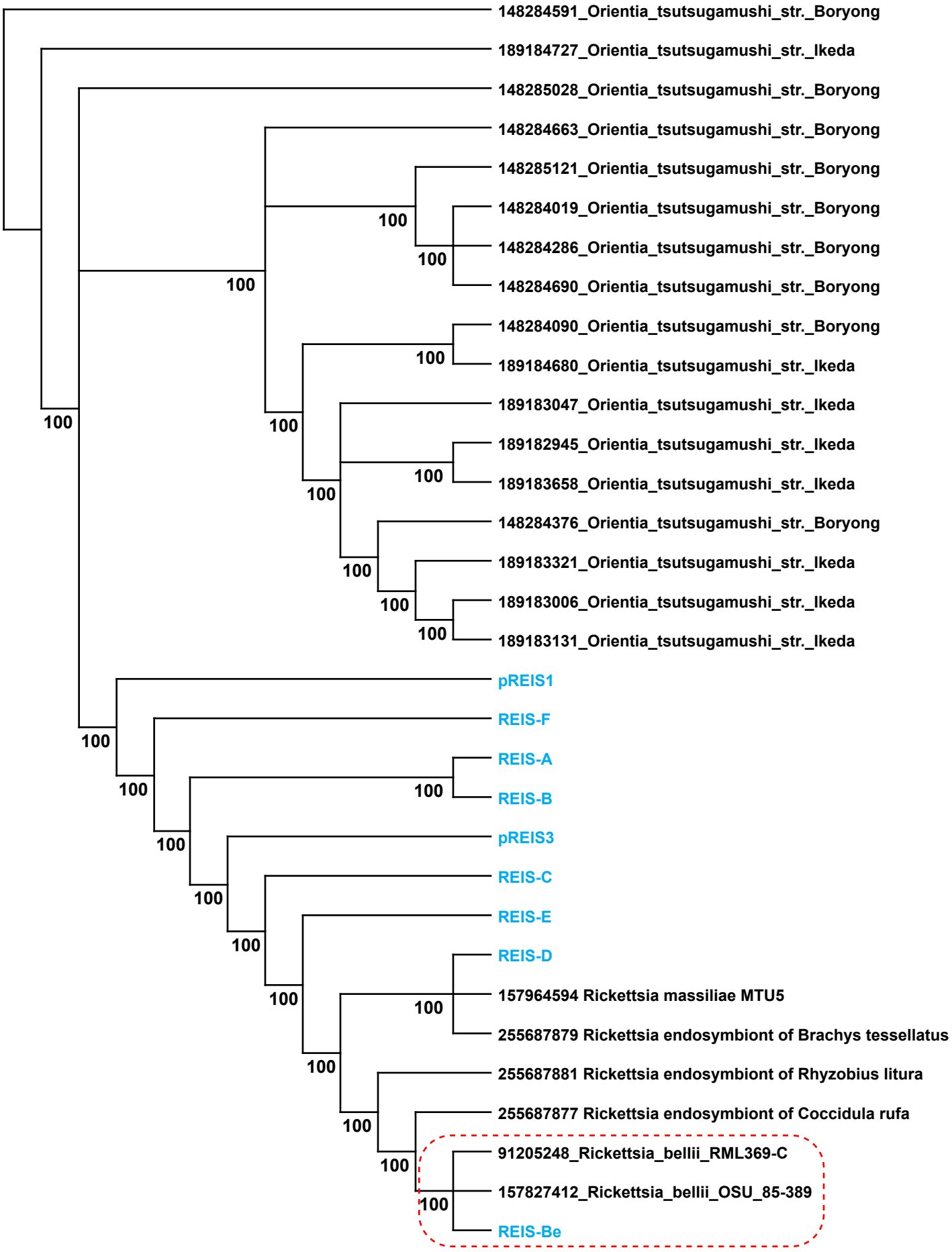
E **HP-18, putative C-terminal half of TraK (VirB9 of P-T4SS)**



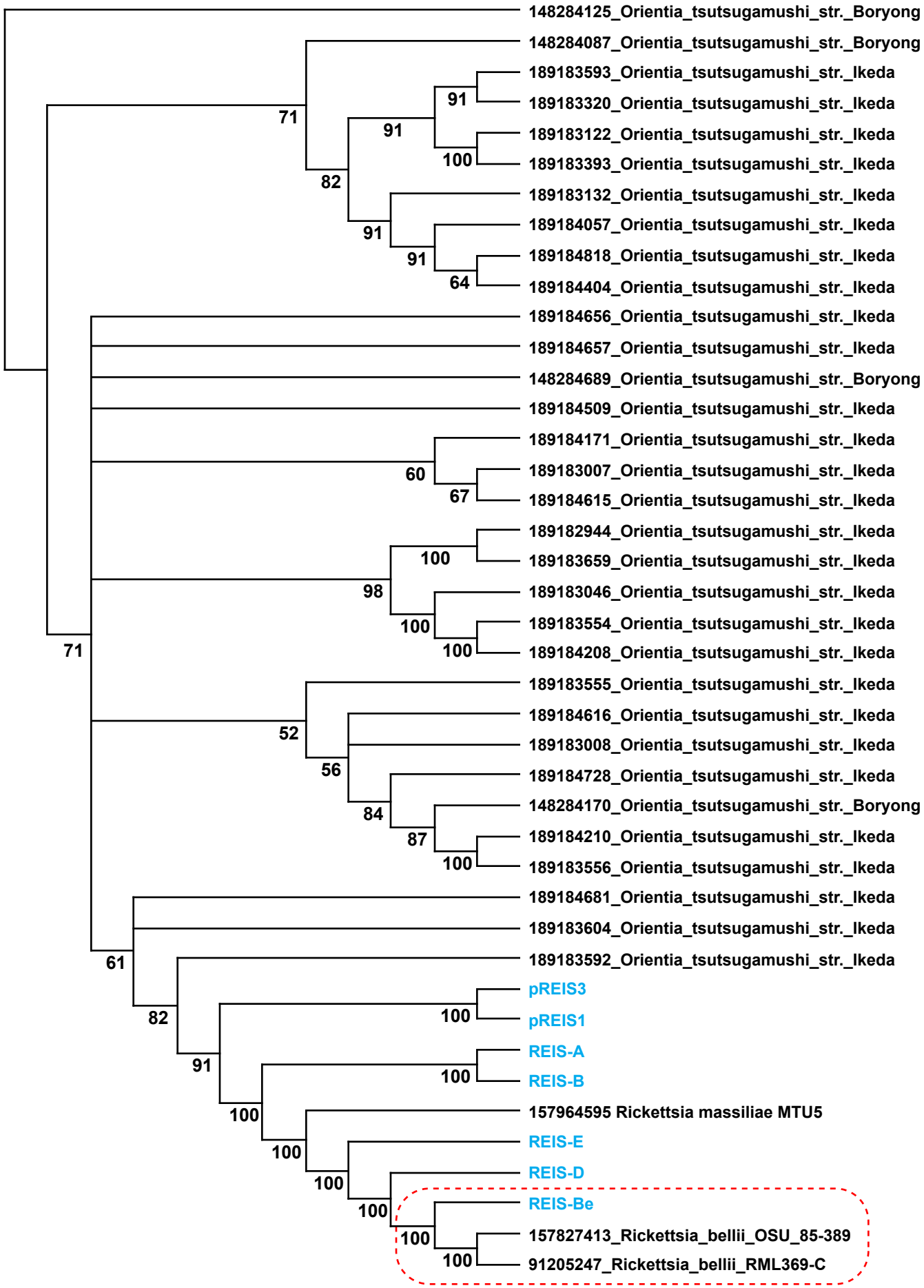
F **TraB (VirB10 in P-T4SS)**



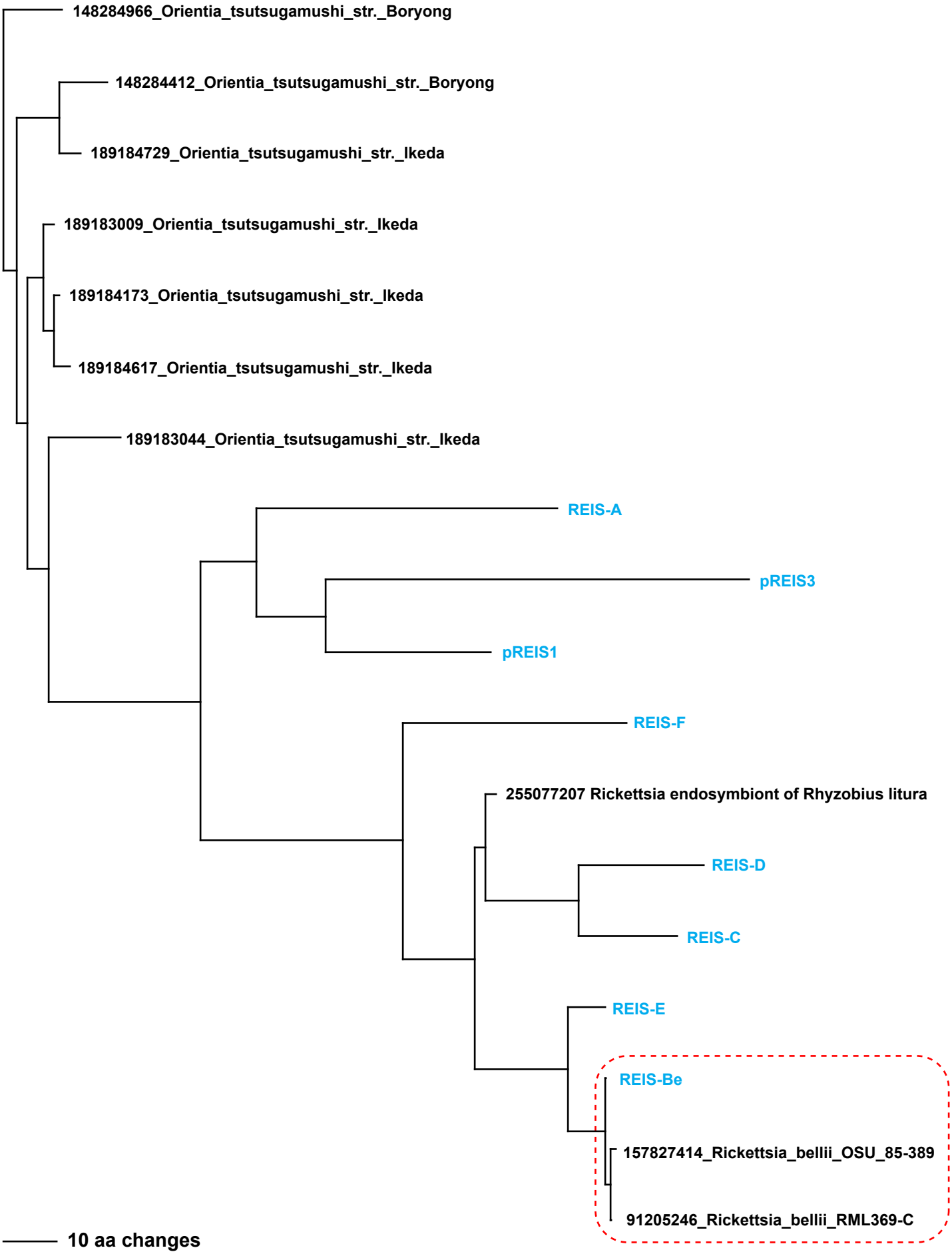
G HP_20, putative TraV homolog (VirB7 in P-T4SS)



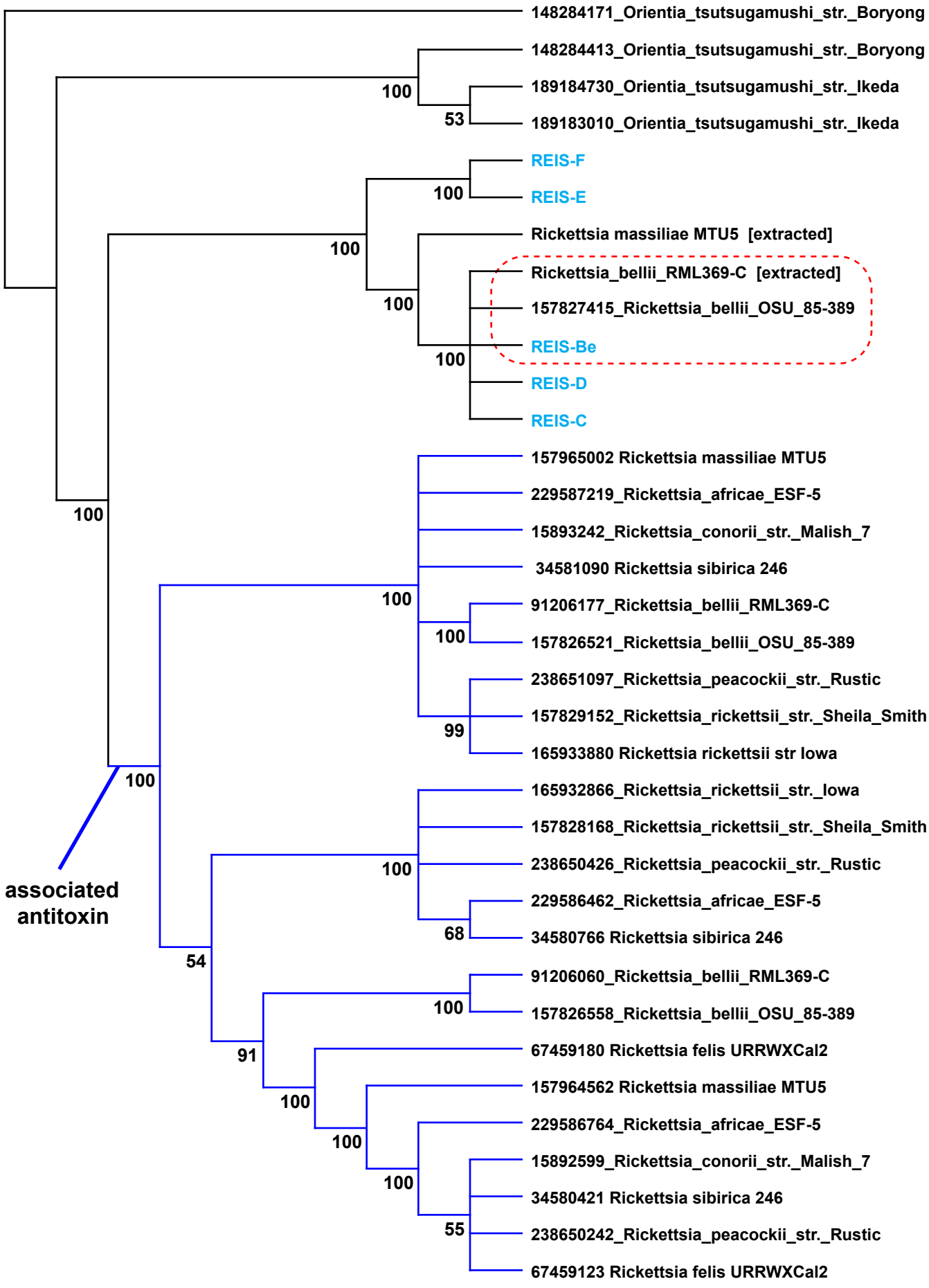
H TraC (VirB4 in P-T4SS)



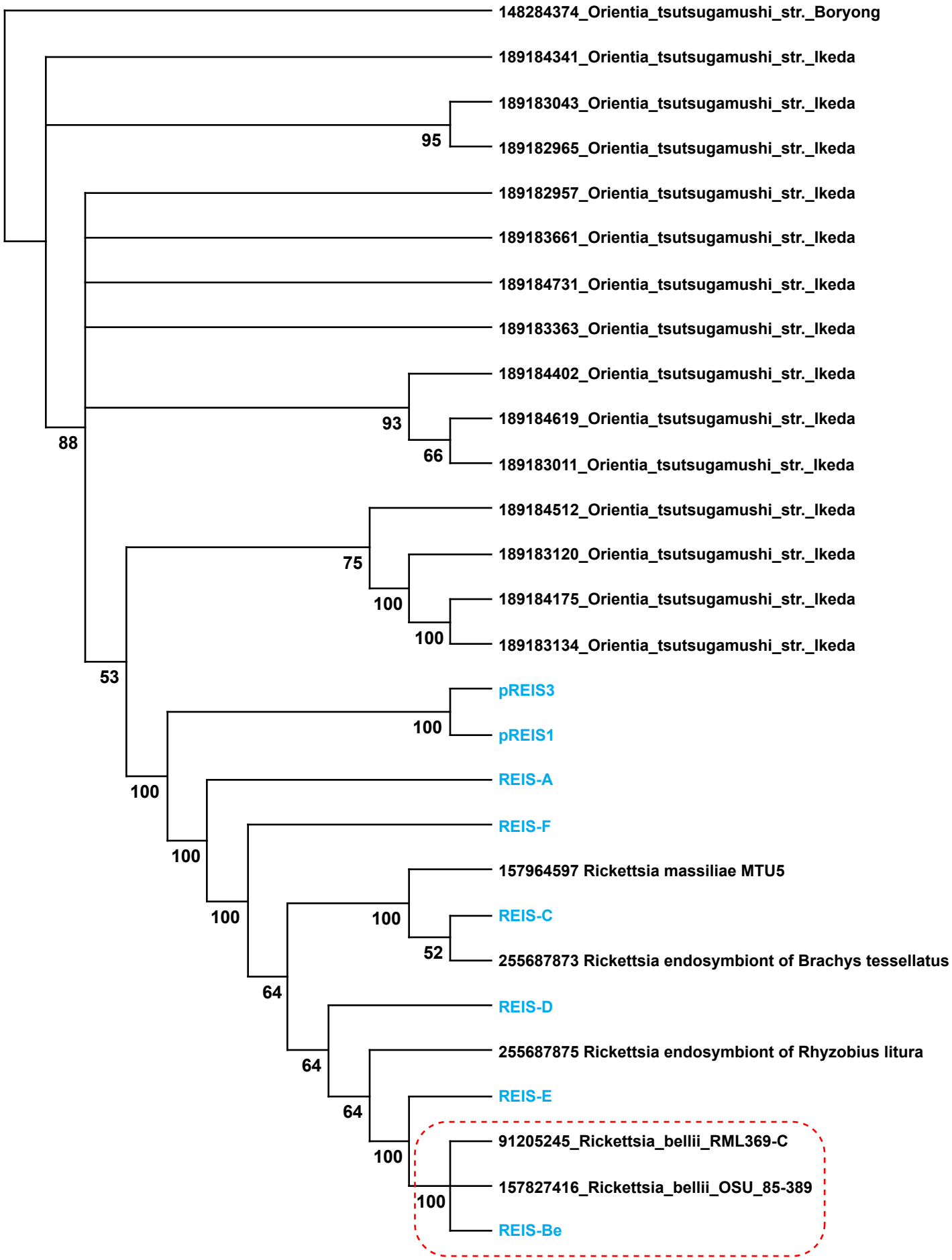
I TraW (no homolog in P-T4SS)



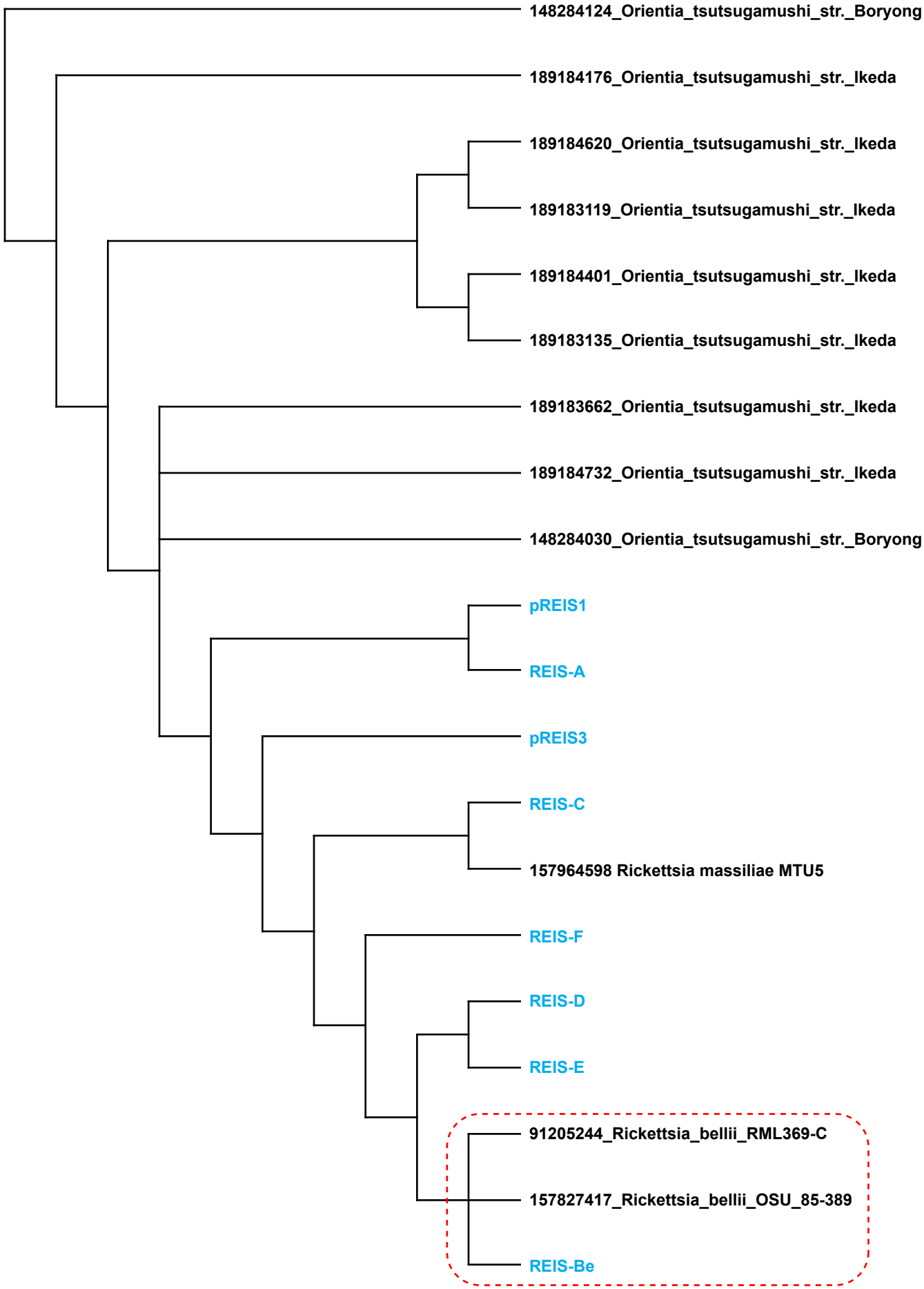
J Toxin of T-AT module (RelE-like); not a part of T4SSs



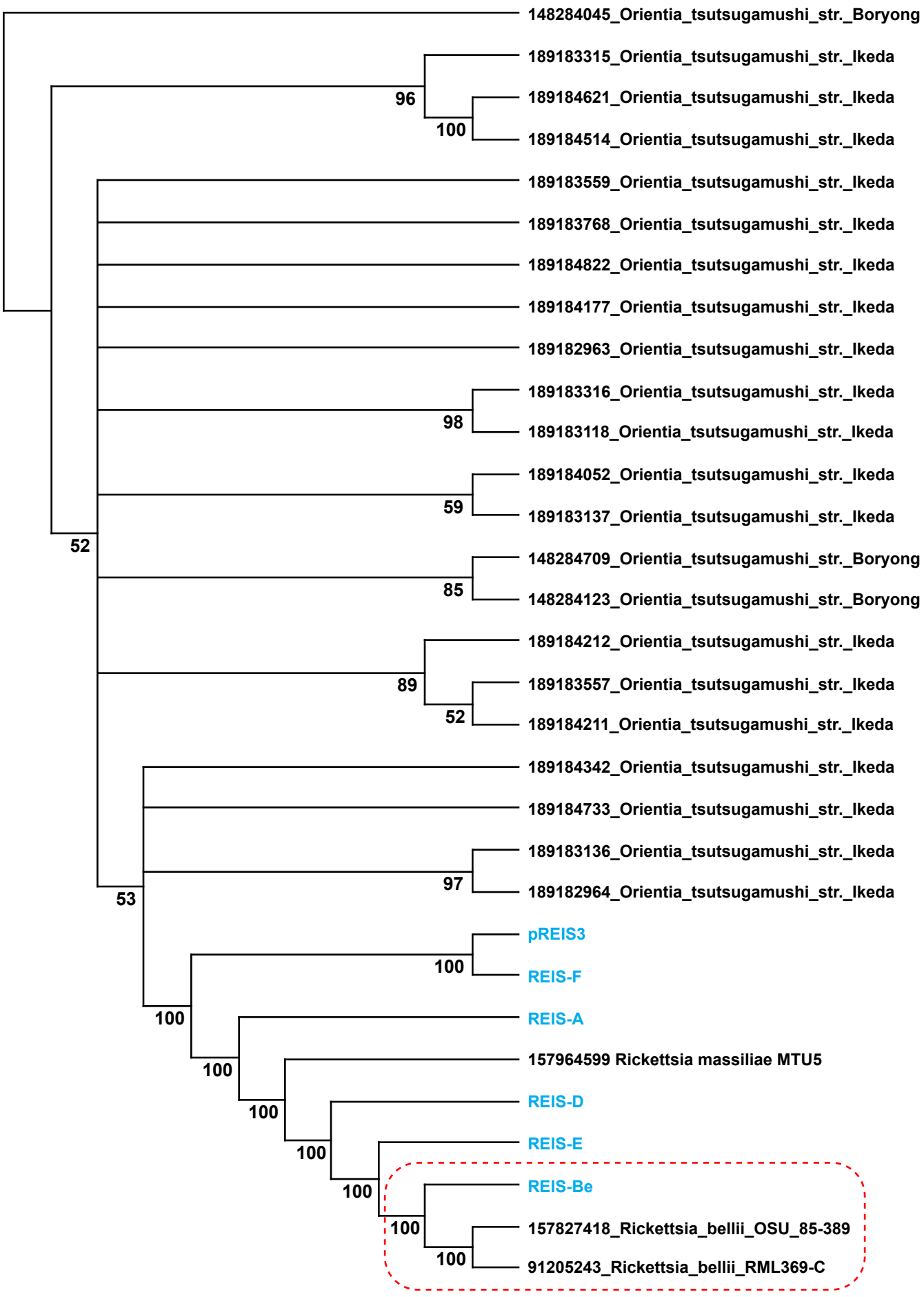
K **TraU (no homolog in P-T4SS)**



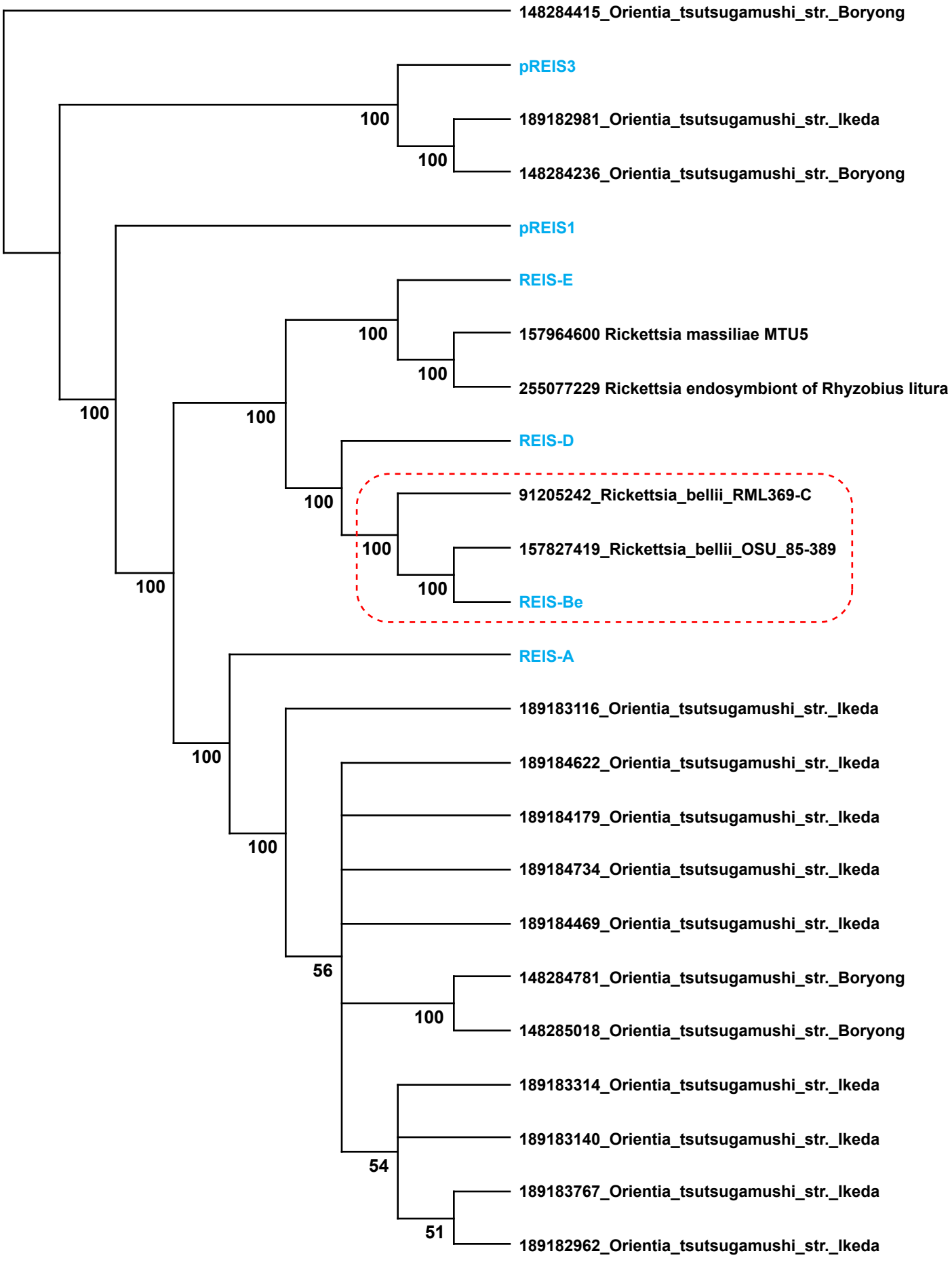
L TrbC (no homolog in P-T4SS)



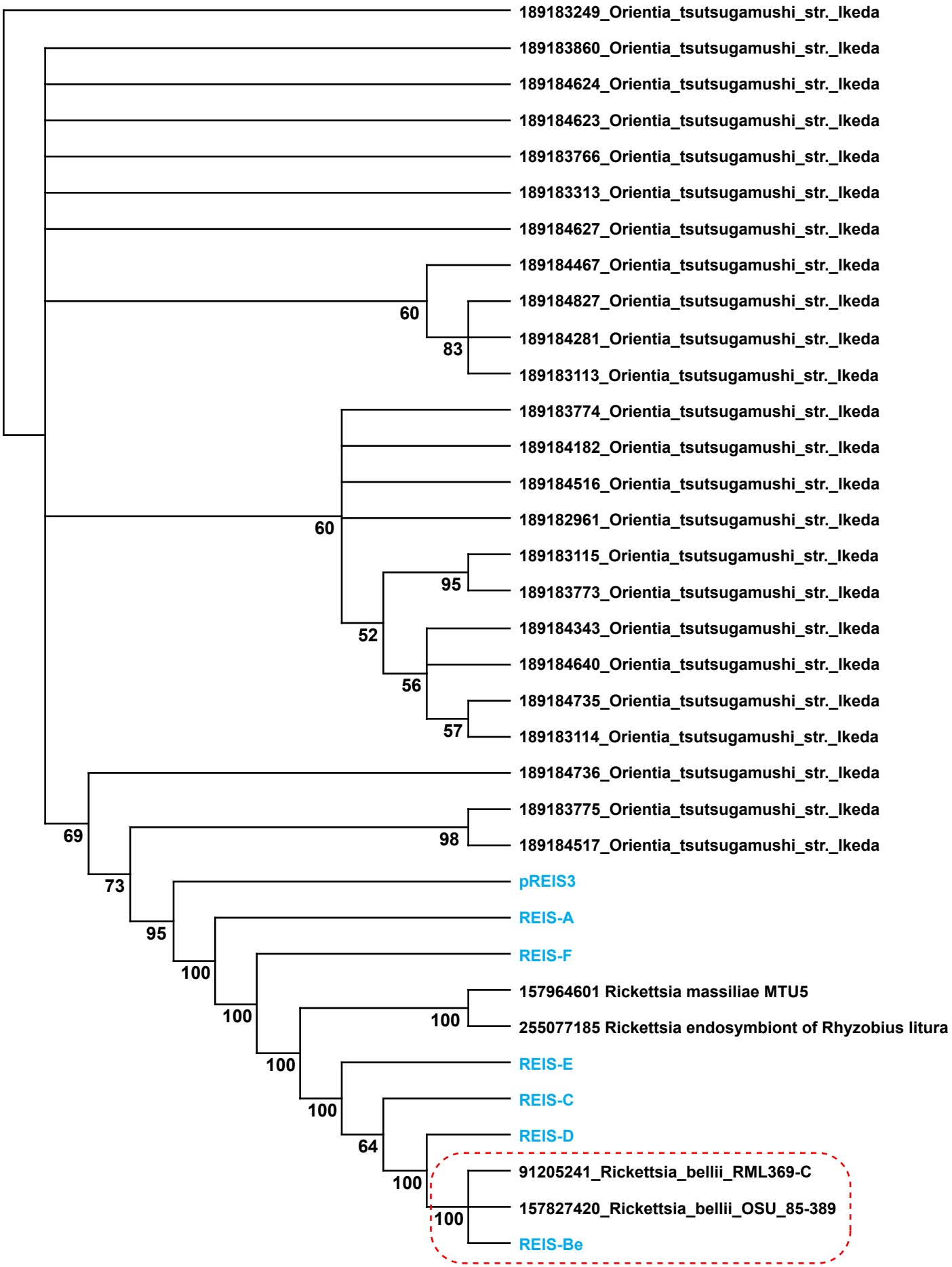
M **TraN (no homolog in P-T4SS)**



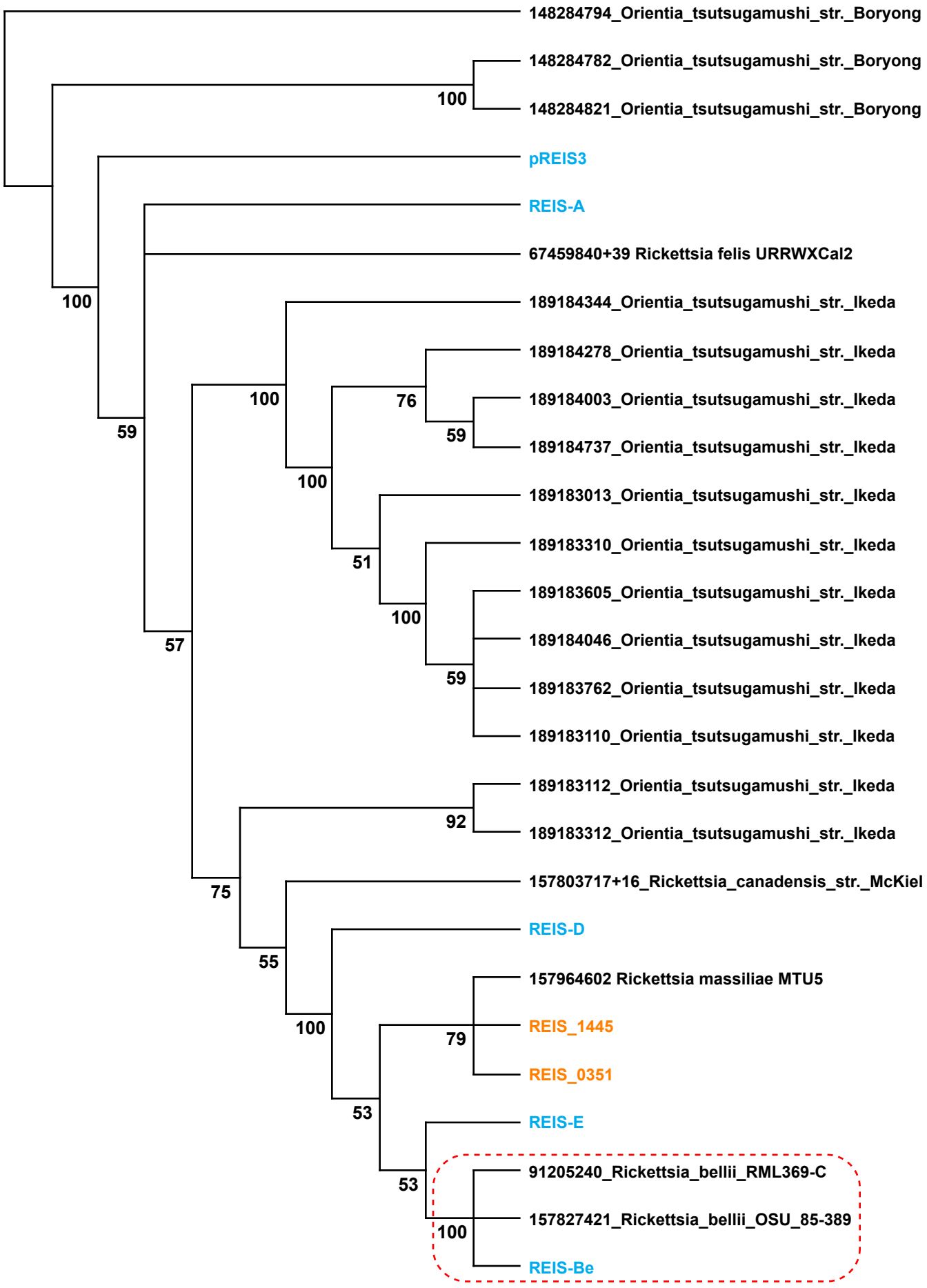
N **TraF (no homolog in P-T4SS)**



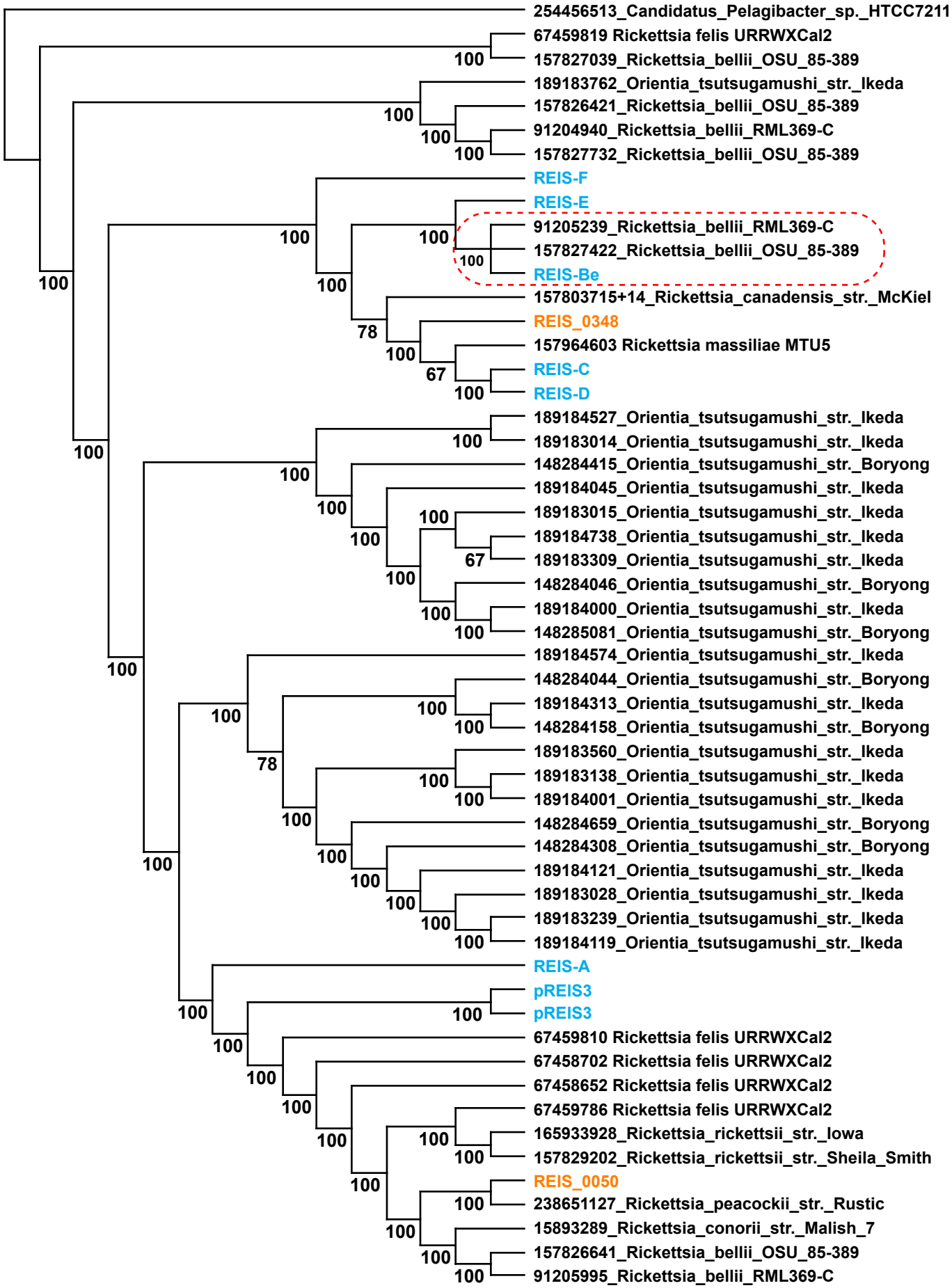
O **TraH (no homolog in P-T4SS)**



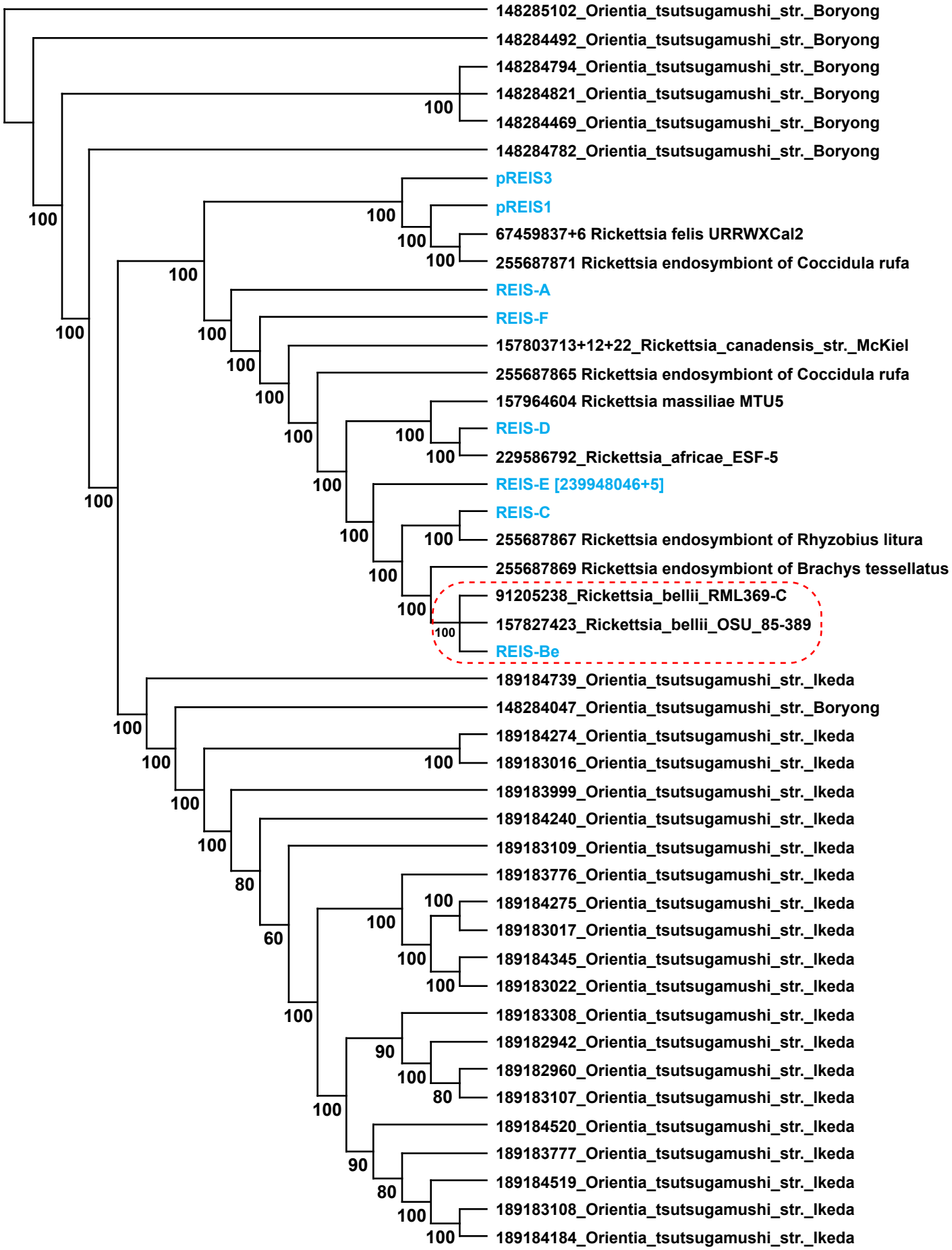
P TraG (putative NTD homolog of VirB6, CTD homolog of VirB8 in P-T4SS)



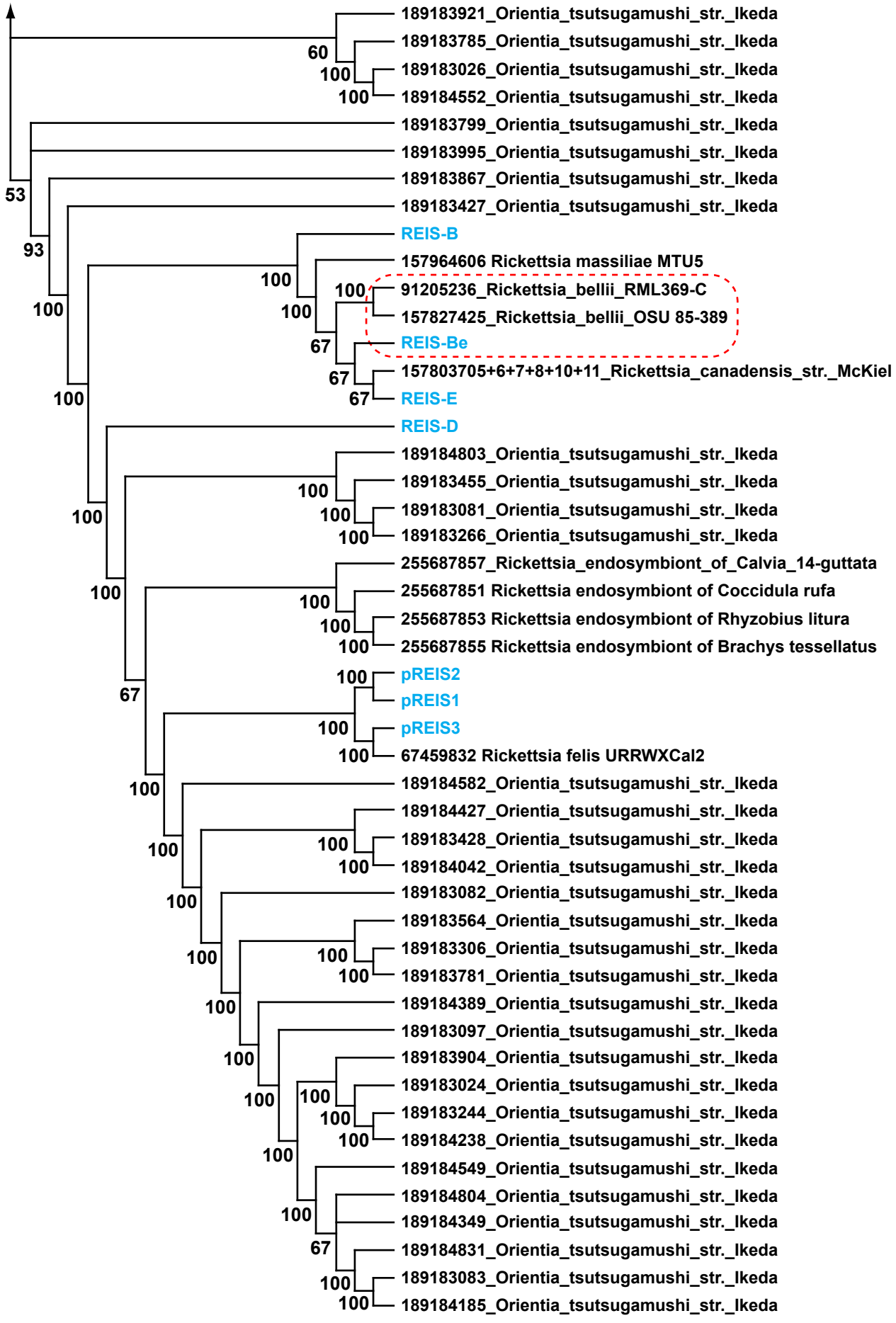
Q TPR-domain containing protein; not a part of T4SSs



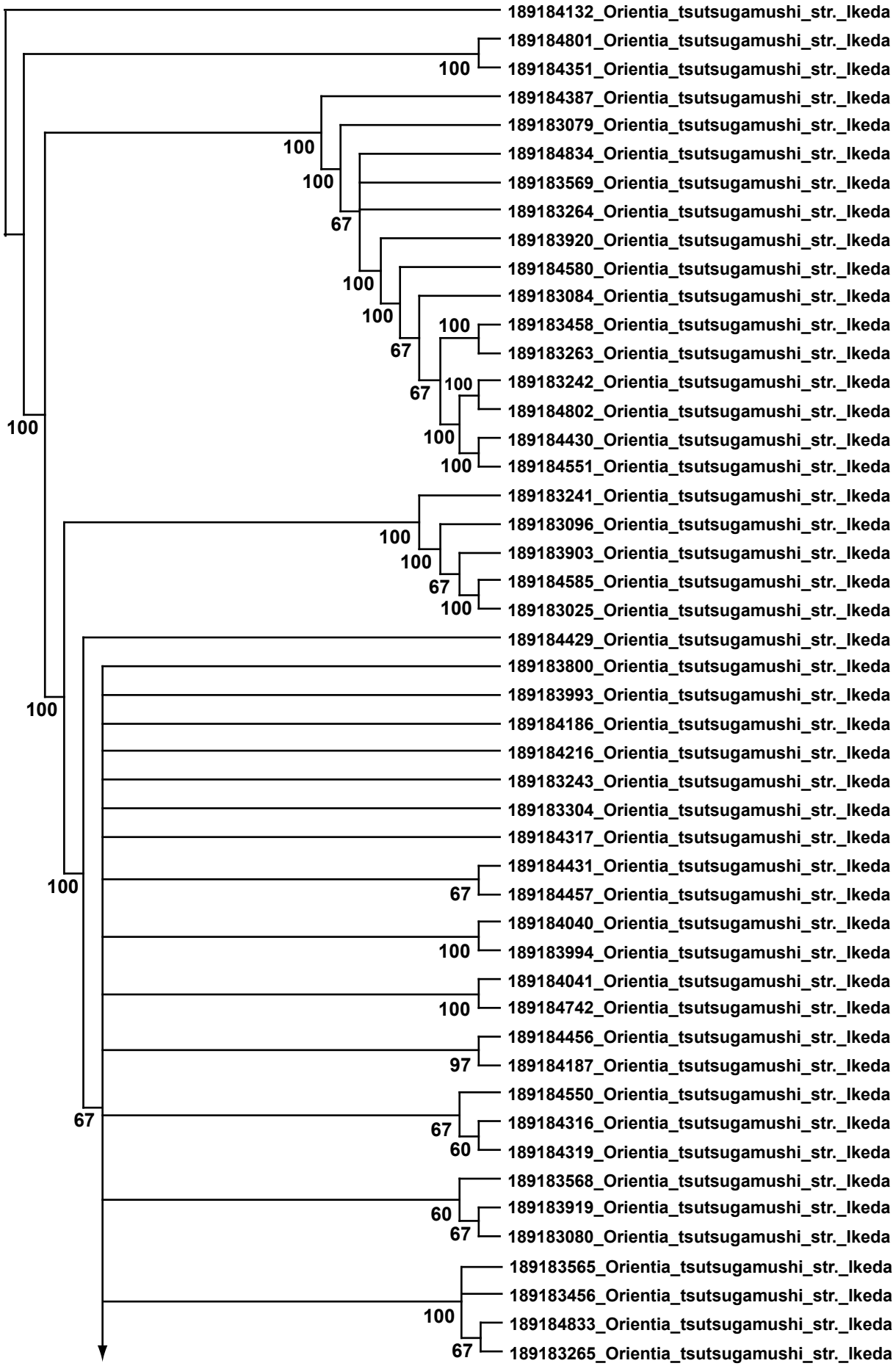
R TraD (VirD4 in P-T4SS)



S (cont'd.) TraAI (relaxase with MobA/MobL, helicase, and nickase domains)



S TraAI (relaxase with MobA/MobL, helicase, and nickase domains)



T TraD, Ti-like: conjugative transfer protein (F-T4SS)

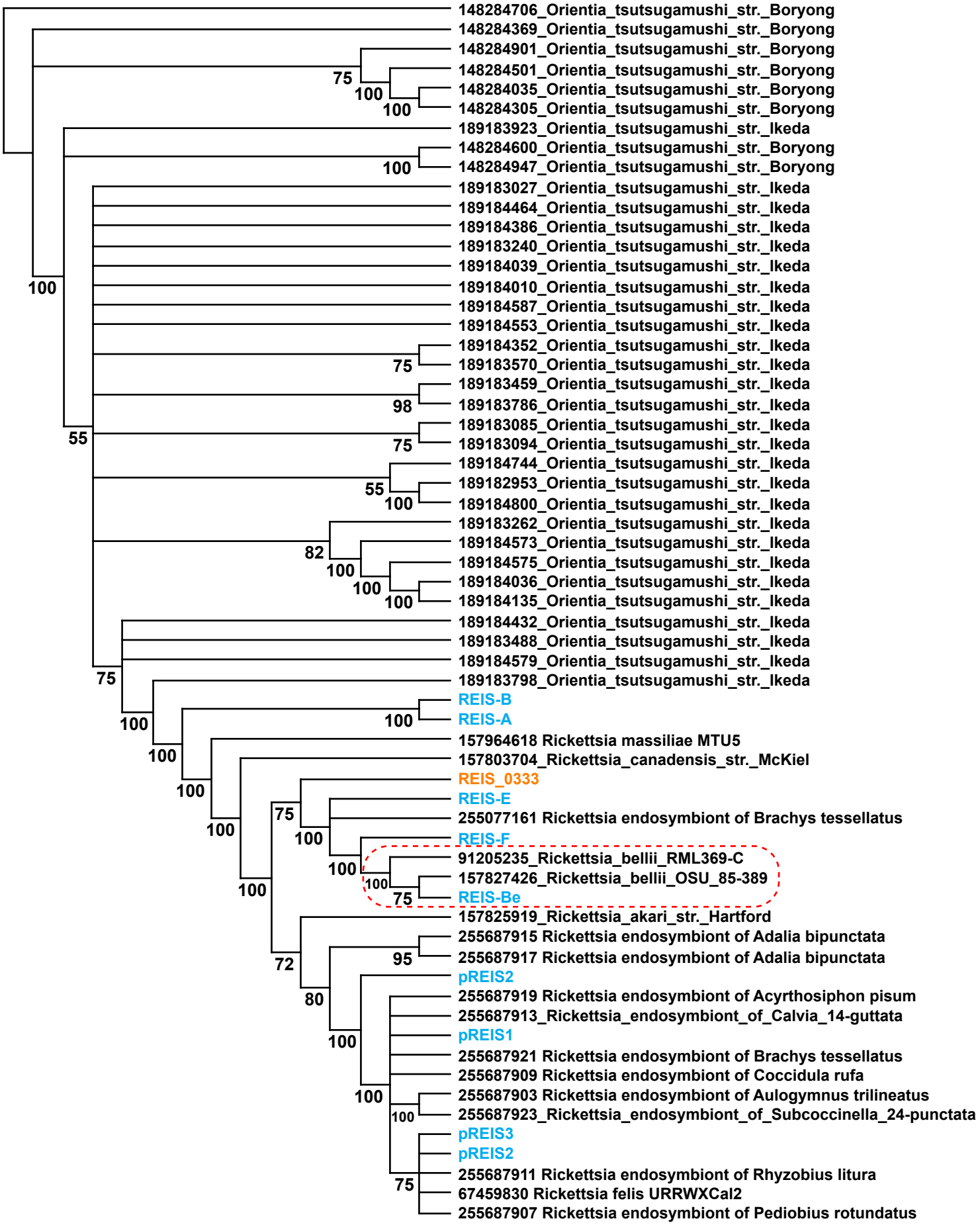
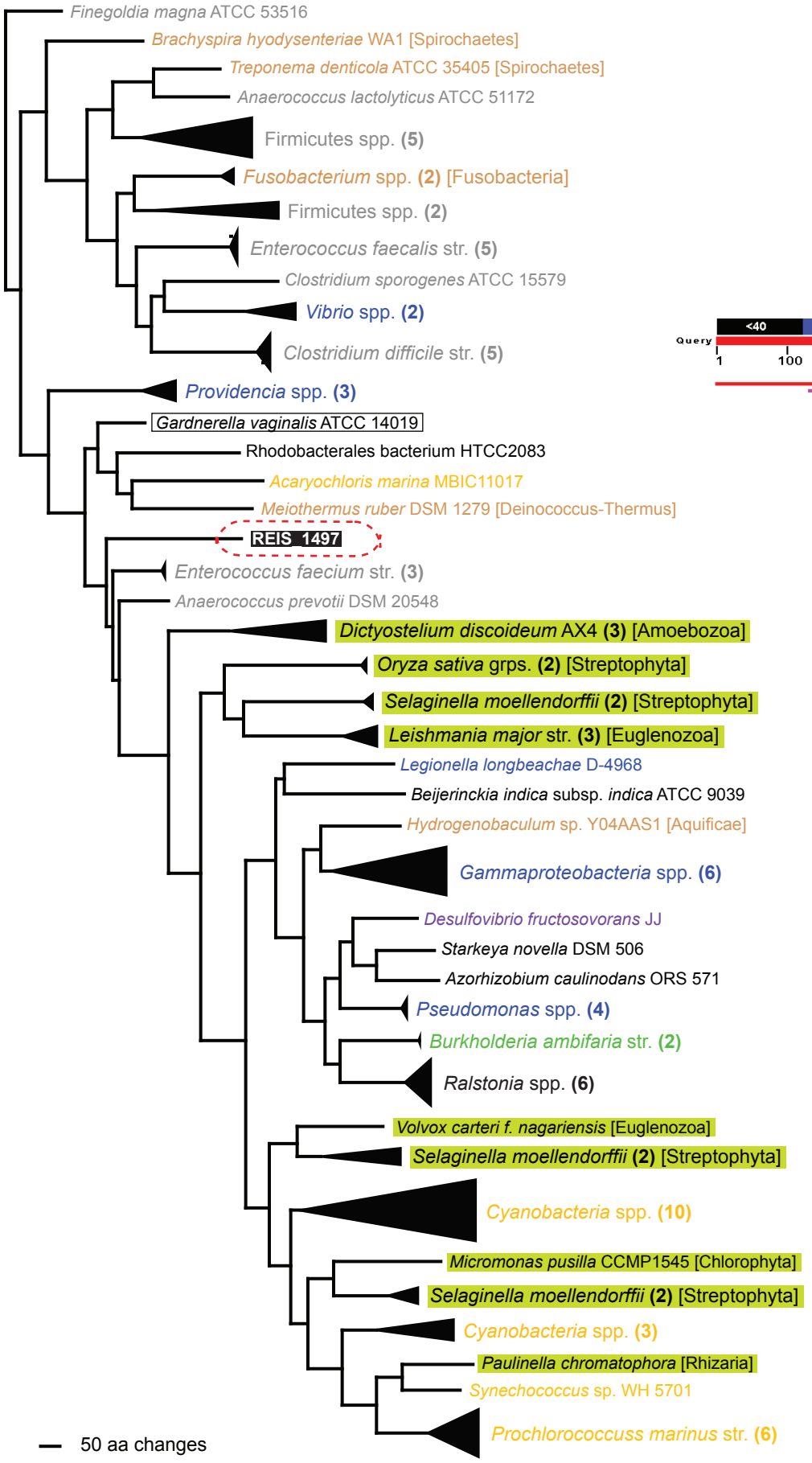


Fig. S11. Characteristics of the non-aminoglycoside antibiotic biosynthesis genes encoded within the RAGE-A piggyback region of the REIS genome. Three ORFs depicted in Fig. 6 are further illustrated here (ABC-like transporter, Tlc2-like transporter, LuxR-like transcription regulator). See Fig. S2 for analysis of the GlpT transporters that includes REIS_1488. See Fig. 6 for analysis of the ATP-binding multidrug resistance transporter MdlB, which includes REIS_1494. For each analysis, top blastp subjects (cut-off of 100) with significant alignments to the REIS queries were downloaded from NCBI and aligned with MUSCLE v3.6 using default parameters [1, 2]. Phylogenies were estimated in PAUP* v4.0b10 (Altivec) under parsimony [3]. Analyses of ABC-like transporter (COG4178) and Tlc transporters implemented 500 heuristic searches, saving 100 trees per replication. Analysis of REIS_1499 implemented an exhaustive search. Majority rule consensus trees were constructed for analyses generating multiple equally parsimonious trees. (A) Analysis of REIS_1497, ABC-like transporter (COG4178). Schema of top blastp subjects shows limited similarity of the N-terminal sequence with other organisms. C-terminal similarity with other bacteria is across the predicted ABC-like ATPase domain. Inset shows predicted TMS regions [4]. A single phylogram is shown, one of 15 equally parsimonious trees of score 16343. All nodes shown were recovered in 100% of the trees. (B) Analysis of REIS_1498, a Tlc2-like transporter. Characterized Tlc proteins and their associated substrates are provided [5]. A single phylogram is shown, one of 83 equally parsimonious trees of score 13470. All nodes shown were recovered in 100% of the trees. *Rickettsia* proteins Tlc1-5 are enclosed in red dashed box on the phylogram, with distribution within each clade depicted in inset at right. (C) Analysis of REIS_1499, LuxR-like transcription regulator. Top left, schema of the larger ORFs analyzed that contain a C-terminal LuxR-like domain. Top right, top blastp hits retrieved using REIS_1499 as a query. The sequences having homology outside of the LuxR-like domain were selected for alignment and phylogeny estimation. Middle, full protein alignment of selected LuxR-like domain-containing proteins. Asterisks below alignment denote invariant residues. Region shaded orange corresponds to the predicted LuxR-like domain. Bottom, phylogeny estimation of 11 LuxR-like domain-containing proteins. An exhaustive search resulted in one tree of score 1022. Branch support (>50%) from 1000 bootstrap pseudoreplications is shown on the tree.

1. Edgar, R.C., *MUSCLE: a multiple sequence alignment method with reduced time and space complexity.* BMC Bioinformatics, 2004. **5**: p. 113.
2. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
3. Swofford, D., *PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4 ed.* , 1999, Sinauer: Sunderland, MA.
4. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* J Mol Biol, 2001. **305**(3): p. 567-80.
5. Audia, J.P. and H.H. Winkler, *Study of the five Rickettsia prowazekii proteins annotated as ATP/ADP translocases (Tlc): Only Tlc1 transports ATP/ADP, while Tlc4 and Tlc5 transport other ribonucleotides.* J Bacteriol, 2006. **188**(17): p. 6261-8.

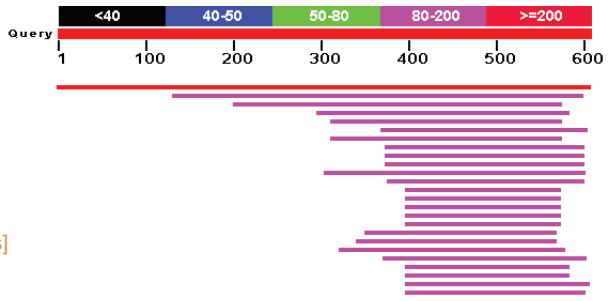
A

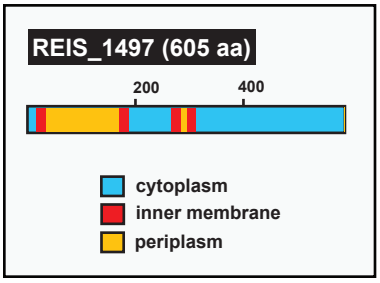


ABC (COG4178)

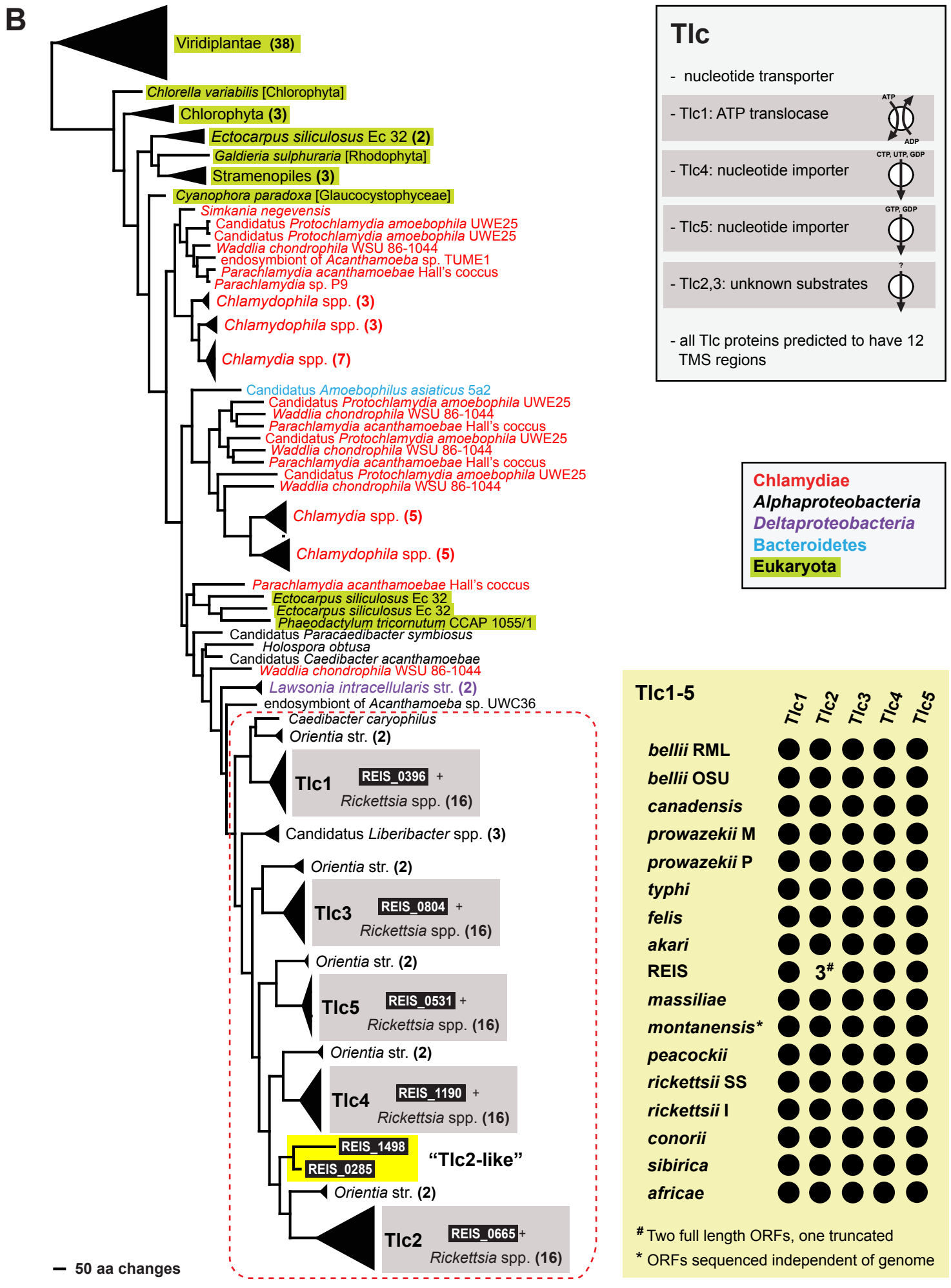
- ABC-type uncharacterized transport system, permease and ATPase components

- other possible models include: MdlB and SunT.

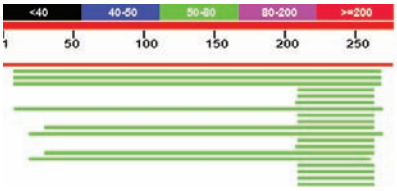




Actinobacteria
Firmicutes
Cyanobacteria
Alphaproteobacteria
Deltaproteobacteria
Gammaproteobacteria
Betaproteobacteria
other bacteria
Eukaryota


B
Viridiplantae (38)
Chlorella variabilis [Chlorophyta]
Chlorophyta (3)
Ectocarpus siliculosus Ec 32 (2)
Galdieria sulphuraria [Rhodophyta]
Stramenopiles (3)
Cyanophora paradoxa [Glaucocystophyceae]
Simkania negevensis
Candidatus Protochlamydia amoebophila UWE25
Candidatus Protochlamydia amoebophila UWE25
Waddlia chondrophila WSU 86-1044
endosymbiont of Acanthamoeba sp. TUME1
Parachlamydia acanthamoebae Hall's coccus
Parachlamydia sp. P9
Chlamydophila spp. (3)
Chlamydophila spp. (3)
Chlamydia spp. (7)
Candidatus Amoebophilus asiaticus 5a2
Candidatus Protochlamydia amoebophila UWE25
Waddlia chondrophila WSU 86-1044
Parachlamydia acanthamoebae Hall's coccus
Candidatus Protochlamydia amoebophila UWE25
Waddlia chondrophila WSU 86-1044
Parachlamydia acanthamoebae Hall's coccus
Candidatus Protochlamydia amoebophila UWE25
Waddlia chondrophila WSU 86-1044
Chlamydia spp. (5)
Chlamydophila spp. (5)
Parachlamydia acanthamoebae Hall's coccus
Ectocarpus siliculosus Ec 32
Ectocarpus siliculosus Ec 32
Phaeodactylum tricornutum CCAP 1055/1
Candidatus Paracaedibacter symbiosus
Holospora obtusa
Candidatus Caedibacter acanthamoebae
Waddlia chondrophila WSU 86-1044
Lawsonia intracellularis str. (2)
endosymbiont of Acanthamoeba sp. UWC36
Caedibacter caryophilus
Orientia str. (2)
Tlc1
REIS_0396 +
Rickettsia spp. (16)
Candidatus Liberibacter spp. (3)
Orientia str. (2)
Tlc3
REIS_0804 +
Rickettsia spp. (16)
Orientia str. (2)
Tlc5
REIS_0531 +
Rickettsia spp. (16)
Orientia str. (2)
Tlc4
REIS_1190 +
Rickettsia spp. (16)
REIS_1498
REIS_0285
"Tlc2-like"
Orientia str. (2)
Tlc2
REIS_0665 +
Rickettsia spp. (16)
50 aa changes
Tlc
- nucleotide transporter
- Tlc1: ATP translocase
ATP
ADP
- Tlc4: nucleotide importer
CTP, UTP, GDP
- Tlc5: nucleotide importer
GTP, GDP
- Tlc2,3: unknown substrates
?
- all Tlc proteins predicted to have 12 TMS regions
Chlamydiae
Alphaproteobacteria
Deltaproteobacteria
Bacteroidetes
Eukaryota
Tlc1-5
Tlc1
Tlc2
Tlc3
Tlc4
Tlc5
bellii RML
bellii OSU
canadensis
prowazekii M
prowazekii P
typhi
felis
akari
REIS
3#
massiliae
montanensis*
peacockii
rickettsii SS
rickettsii I
conorii
sibirica
africae
Two full length ORFs, one truncated
* ORFs sequenced independent of genome

C




```
Parachlamydia acanthamoebae str. Halls coccus   ----------MLRKI-TSSPYH------NKMLQFAAPLNDH   24
Legionella drancourtii str. LLAP12              ------MTKLTQLDENPLSYQDKYA----DSVSDMVYPFIDE   32
REIS_1499                                       ------MSTLFNIPKDHSTFIYH------KDVEDIVRPLREL   30
Rickettsia bellii str. OSU 85-389               ------MLDMLNLTET-NALAYQYYLSTQKRLKTIYDHLIS-   34
Rickettsia bellii str. RML369-C                 ---------MLNLTET-NALAYQYYLSTQKRLKTIYDHLIS-   31
Legionella pneumophila str. Lens                ---------MYELNDT-PYFIYS------NLIQEYCSPLFNL   26
Ralstonia solanacearum str. UW551               ---------MNKTPEVIFT----------GARDICQPFFNE   22
Ralstonia solanacearum str. PSI07               --------MMKKTPEIIFT----------GARDICQPFFNE   23
Legionella pneumophila subsp. pneumophila str. Philadelphia 1  MPMSLKREVMPNSSNVLRHMENHFALTTCEDIKNIMQPLLIK   42
Legionella pneumophila str. 130b                ---------MPNSSNVLRHMENHLALTTCEDIKNIMQPLLAK   33
Legionella pneumophila str. Paris               ---------MPNSSNVLRHMENHFALTTCEDIKNIMQPLLIK   33

PaHc   LGINHFWYYRITFSGHYSYIGTN--SAWNEFCF-ENQMNSHFPCLRHPKT---------LKSGISLMKASEDPSYQNVLKTAWEKFNIN  101
Ld     LKITFFSHVRVWHTGIFKSLMTE--TSLSRFYI----EQKYPIRFSLGKNI-------LLNSGCYLLQHLPDLSSQQDLNLTRAMFNLD  108
REIS   LNINYFCFSRIYDNGKHIALTTN--PAWFLRYY----NNRYFLKEAISKTGIVGN---FYPIVWDYVIDIDKLSYQAMLDTR--NHNFA  108
RbO    LGVPFFGYIKIFKDGSYLPLISNITTEFMQAYFSIIKNQGFSATTVINKTINTKYNYVFFPTEIEHY-DKRKDPIMNLMYDFNIWKNML  122
RbR    LGVPFFGYIKIFKDGSYLPLISNITTEFIQAYFSIIKNQGFSATTVINKTINTKYNYVFFPTEVEHY-DKRKDPIMNLMYDFNIWKNML  119
LpL    TPIKGFTFCRTYGDKTAFLLTSE--YQFHIECIQHEKKISKLNMLRTPEYTQAGY---FLNSPA------LDGKYSPFYESTAQKYGYG  104
RsU    TGLNSFSYSRLFMDGTRCEVWSD--AGAFEHTF---HKARYIVGAYTPQYFGVRERYSILDKKVETYPAHLRNRYRMQLADQREYFDYD  106
RsP    TGLNSFSYSRLFMDGTRTEVWSD--AEAFEHTF---HKARYIVGAYTPQYFGIRERYSILDKKVETYPAYLRTRYRSQLADQREYFDYD  107
LppP   HGMTVFNYYRIYFDSTVIRLSTD--QKWTEHYF----KMNYLETLTIPES--------YLKKPLNYY-LWLTDDCPLMLQDAAINFNMA  116
Lp130  HGMTVFNYYRIYFDSTVIRLSTD--QKWTEHYF----KMNYLETLTIPES--------YLKKTLNYY-LWLTDDCPLMLQDAAINFDMA  107
LpPa   HGMTVFNYYRIYFDSTVIRLSTD--QKWTEHYF----KMNYLETLTIPES--------YLKKPLNYY-LWLTDDCPLMLQDAAINFDMA  107
            *

PaHc   FNINLQKITPEGIEAFG-FATRFNDLK-AEERLINDLPVLRNFIKIFKEKNEKLISMLSE-YQI-------H-LPSKFGPKFYEQPKTI  179
Ld     HFIYLVDKQTKWDDLFI-LATKPENDV-FVNLVFNNLDFIKQGLHRYKYNARELLEKSPG-ISYSTDRFIDH-EQQCQYKIQMDRVDLQ  193
REIS   HLFSLKYISMKYIDLYS-MSVPEEEKQ-SNEKYLQYYEIIKKFTLYFQSAAASIINKAEK-NPISVNNLTEF-QST--EDDIISKLDQF  191
RbO    GIYKL--INSEFIECYM-FSMEGSAIQ-AMNFYLNNTQLLEYGIDYFDVKAKDLIDTTDK-TKL------AY-FKQKLNFNILDQKELI  199
RbR    GIYKL--INSEFIECYM-FSMEGSAIQ-AMNFYLNNTQLLEYGIDYFDVKAKDLIDTTDK-TKL------AY-FKQKLNFNILDQKELI  196
LpL    FSFTIVRKNENHCDYFH-FIGNRHQKD-MNNYYLHNKWLLDKFCDYFLERAAKVLAYSEKENRFTLTNTAEHPFPLKIHSSALDQ-FLK  190
RsU    HCFAILNQDKDFCEYFI-FYAPRSNVM-ALNFYLNNFDRLENFSRYFLQTADNLIEQANR-HRI----HSAV-HPT-PSTATVES----  182
RsP    HCFSILNQTKDFCEYFI-FYAPRSNVM-AINFYLNNFDRLENFSDYFLQAADNLIEQADQ-YRI----HSAF-NHT-LDAAKVGA----  183
LppP   NGISIAKINHDSIEYFC-FASTRDNTSIVNNFYLNNLDVLEQYSLYFKDQFNSIISQFEK-NKI----ILPY-NATCCQSNKINK----  194
Lp130  NGISIAKINHDSIEYFC-FASTRDNTSIVNNFYLNNLDVLEQYSLYFKDQFNSIISQFEK-NKI----ILPY-NATCCQSNKINK----  185
LpPa   NGISIAKINHDSIEYFC-FASTRDNTSIVNNFYLNNLDVLEQYSLYFKDQFNSIISQFEK-NKI----ILPY-NATCCQSNKINK----  185

PaHc   AFPTQHQLFMQKIGGQPAFSLTPREKDILKFVSNGYPASYIAEQLQLSQRTVENYIATIKCKLSCSSKVELIQEAQKITSIECDSI---  265
Ld     -------KILVPLN-DKEISISRQEYHCLGFLIQNHSIKEIAQLMNLSPRTVETYLNNLKNKLHCQNNSEL------LGFIHKLYLF--  266
REIS   EQQIKVLQLNLNLN-GFIIKLTITETKYLKALALGMSSKSIAKKFNVSTRTVETHINNLKSKLGVYCKSEIAFKL--NSQGVNFNMV--  275
RbO    LQQSKQFSLNAEVKTGKYNKLSYRETECIYYLSLGNSMKEIGSILNLSPRTVEHHLNSIKQKTGLLYKSQLVKEFIKNAPSALFKDIRT  288
RbR    LQQSKQFSLNAEVKTGKYNKLSYRETECIYYLSLGNSMKEIGSILNLSPRTVEHHLNSIKQKTGLLYKSQLVKEFIKNAPSALFKDIRT  285
LpL    DYHFEQFGLKPVC------LPKRQIECAQLLVRGKTAEEIAIILNLSKRTVEHYINILKRKLNAHNKGLLITKLLHQYPDMFESIL--  270
RsU    -------TVQRALP-DEKRAFTPREDDVAKLLITGATAKEIGNTLGISHRTVESHLEHMKQKLGCMKKSMLVKAL--LAQQHAHYLMNR  261
RsP    -------VAQRVFP-DEKKTFTPREDDVARLLITGATAKEIGNTLGISHRTVESHLEHMKQKLGCTKKSMLVKEL--LARQHAHY----  262
LppP   -------ELKRSL------LSPRQKDCAKLLLQGMSYKEIGKILQLSARTVETHVNQLKTKLGCDNKAELILQL--NG------IITR  261
Lp130  -------EIKRSL------LSPRQKDCAKLLLQGMSYKEIGKILQLSARTVETHVNQLKTKLGCDNKAELILQL--NG------IITR  252
LpPa   -------EIKRSL------LSPRQKDCAKLLLQGMSYKEIGKILQLSARTVETHVNQLKTKLGCDNKAELILQL--NG------MITR  252
                                          *      * ****      * *
```

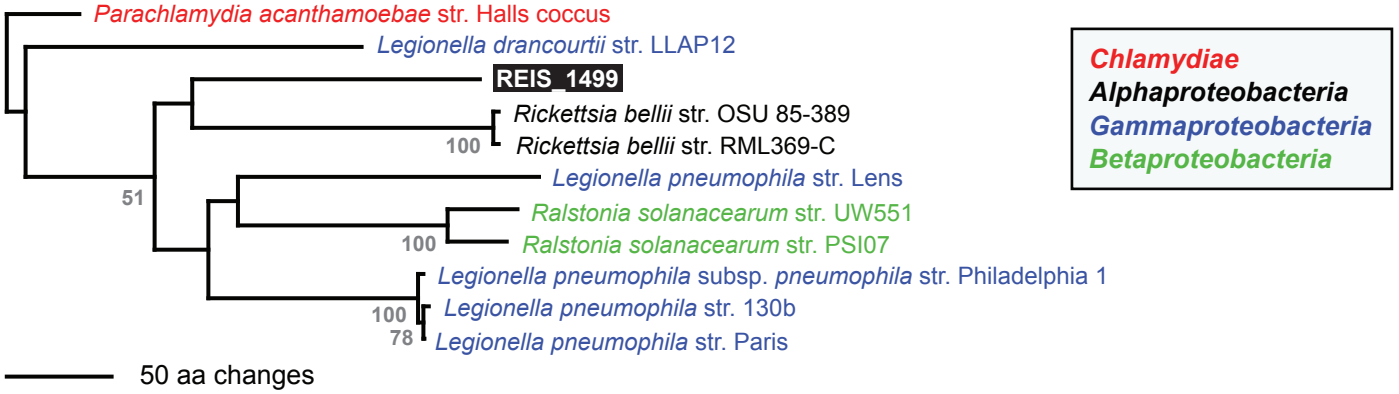
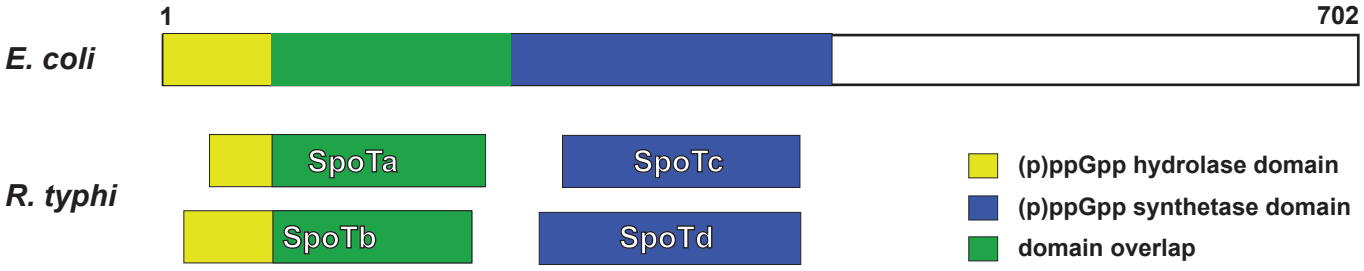


Fig. S12. Compilation and bioinformatics characterization of SpoT-like ORFs across sequenced *Rickettsia* genomes. (A) Schema comparing the SpoT protein of *E. coli* [1] to the rickettsial SpoT-like ORFs. As previously characterized, the *R. typhi* proteins corresponding to the (p)ppGpp hydrolase (SpoTa, SpoTb) and (p)ppGpp synthetase (SpoTc, SpoTd) domains are shown [2]. (B) Compilation of SpoTa/SpoTb-like (left) and SpoTc/SpoTd-like (right) ORFs from all sequenced genomes and plasmids of *Rickettsia* spp. SpoTa/SpoTb-like hydrolases were retrieved from the NCBI *Rickettsia* database (taxid:766) searching with RT0302 (SpoTa). SpoTc/SpoTd-like synthetases were retrieved similarly using RT0700 (SpoTd) as a query. Lengths (aa) for each sequence are provided in parentheses. Split ORFs were merged. Compiled hydrolase and synthetase ORFs were independently aligned with MUSCLE v3.6 using default parameters [3, 4].

1. Gentry, D.R. and M. Cashel, *Mutational analysis of the Escherichia coli spoT gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation.* Mol Microbiol, 1996. **19**(6): p. 1373-84.
2. McLeod, M.P., et al., *Complete genome sequence of Rickettsia typhi and comparison with sequences of other rickettsiae.* J Bacteriol, 2004. **186**(17): p. 5842-55.
3. Edgar, R.C., *MUSCLE: a multiple sequence alignment method with reduced time and space complexity.* BMC Bioinformatics, 2004. **5**: p. 113.
4. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.

A



B

Genome	SpoTa/SpoTb-like (hydrolase) [A]	SpoTc/SpoTd-like (synthetase) [B]
Br	RBE_0806 (213), RBE_0564 (253), RBE_1034 (243), RBE_0550 (234), RBE_0552 (234), RBE_0593 (149), RBE_0417 (248)	RBE_0183 (206), RBE_0419 (215), RBE_0592 (141)
Bo	A1I_03225 (253), A1I_05160 (213), A1I_03145 (218), A1I_02185 (190), A1I_03160 (234), A1I_03380 (147), A1I_05655 (248)	A1I_06955 (201), A1I_05635 (215), A1I_03375 (140)
Ca	A1E_03980 (235), A1E_02160 (215), A1E_02605 (235)	A1E_04680 (201), A1E_01625 (126)
Pr	RP312 [C], RP624 [C]	RP705 [C], RP625 [C]
P2	rpr22_CDS306 (235), rpr22_CDS604 (250)	rpr22_CDS683 (70), rpr22_CDS605 (134)
Ty	RT0302 (235), RT0615 (250)	RT0700 (191), RT0616 (113)
Ak	A1C_02330 (164), A1C_05910 (44) + A1C_05905 (111)	A1C_05540 (201), A1C_04920 (113)
Fe	RF_0508 (239), RF_0884 (223), RF_0450 (233), RF_0438 (243), RF_0953 (229), RF_0440 (238), RF_0384 (235), RF_0305 (140), RF_0968 (204), RF_0380 (167), RF_0188 (56) + RF_0187 (172)	RF_0204 (201), RF_0410 (134), RF_0381 (1179) [D]
REIS	REIS_1055 (235), REIS_0540 (244), REIS_1482 (240), REIS_0867 (248), REIS_1557 (217), REIS_1799 (151), REIS_2260 (197), REIS_2099 (54) + REIS_2100 (155), REIS_1079 (103), ORF10326 (122), ORF10359 (245)	REIS_0058 (201), REIS_0864 (215), REIS_1891 (139)
Ma	RMA_0435 (235), RMA_0995 (234), RMA_0753 (219)	RMA_1119 (201), RMA_0996 (139), RMA_p04 (511) [E]
Cam		pRAM18_00005 (1061) [E]
Mo	CAC33635 (235), CAC33640 (145)	CAC33644 (178), CAC33641 (113)
Pe	RPR_04450 (235), RPR_04180 (162)	RPR_01320 (201)
Ri	A1G_02420 (235), A1G_02150 (165), A1G_04900 (109),A1G_06100 (99), A1G_05420 (39)	A1G_06035 (201)
Rw	Rrlowa_0453 (165), Rrlowa_1056 (109), Rrlowa_1309 (99), Rrlowa_1168 (39)	Rrlowa_1296 (178)
Co	RC0426 (235), RC0374 (177), RC0888 (150), RC1098 (99) + RC1099 (57)	
Si	ZP_00142126 (235), ZP_00142169 (174) ZP_00142959 (148), ZP_00142767 (99)	
Af	RAF_ORF0395 (235), RAF_ORF0652 (170), RAF_ORF0349 (103), RAF_ORF0805 (89)	RAF_ORF0991 (201)

[A] Blastp subjects recovered searching with RT0302 (SpoTa) against the '*Rickettsia*' database. Red depicts split ORFs.

[B] Blastp subjects recovered searching with RT0700 (SpoTd) against the '*Rickettsia*' database.

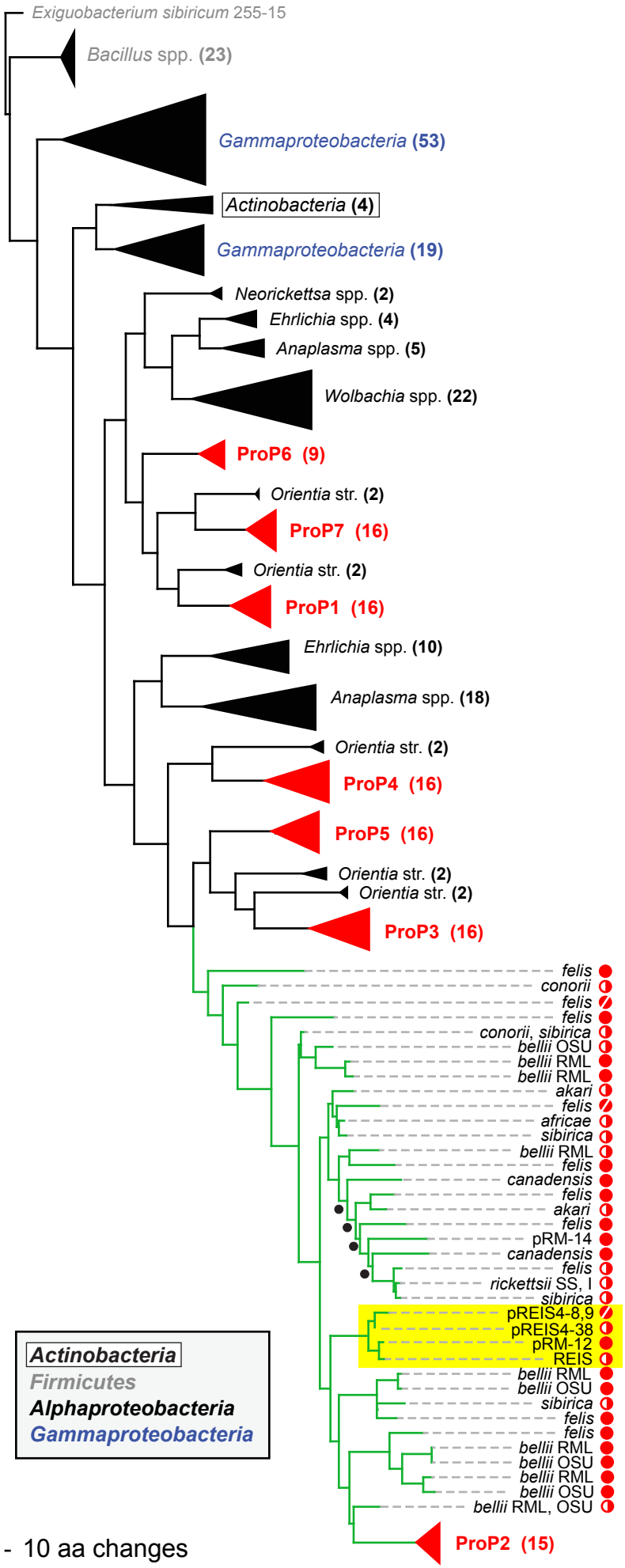
[C] Pseudogenes not annotated as functional ORFs.

[D] GTP pyrophosphosynthetase domain fused to ankyrin repeat domains.

[E] "*Candidatus* Rickettsia amblyommii" str. AaR/SC multidomain protein with GTP pyrophosphosynthetase, leucine-rich repeat (LRR), and Ran GTPase-activating protein (RanGAP) domains. RMA_p04 has the GTP pyrophosphosynthetase domain split from the LRR and RanGAP domains (RMA_p03).

Fig. S13. Compilation and phylogeny estimation of proline betaine transporters (ProP) encoded within *Rickettsia* genomes. A total of 143 putative ProP sequences were retrieved using *R. typhi* ProP4 (YP_067622) as a query against the '*Rickettsia*' database (taxid:780). Split ORFs were merged resulting in a total of 134 ProP sequences. These *Rickettsia* sequences were combined with 171 blastp subjects acquired using the same query sequence against the NCBI non-redundant protein database excluding taxid:780. All 305 ProP sequences were combined in one file and aligned with MUSCLE v3.6 using default parameters [1, 2]. A phylogeny was estimated in PAUP* v4.0b10 (Altivec) under parsimony [3]. A heuristic search was implemented employing 250 random sequence additions, saving 10 trees per replication. A majority rule consensus tree was constructed for 20 equally parsimonious trees of tree score 13290. One phylogram is presented. All nodes shown on the tree (black branches) were present in all 20 trees, with collapsed clades (open and filled triangles) containing nodes not present in all trees. The inset at top illustrates the seven conserved ProP families of *Rickettsia* spp. (ProP1-7). The portion of the tree depicting the less conserved cluster of *Rickettsia* ProP sequences (green branches) is illustrated by a phylogram with nodes not recovered in all equally parsimonious trees depicted with black dots. The clade comprised of plasmid encoded ProP ORFs is highlighted yellow. The partial and complete ProP sequences from this less conserved group were tallied and combined with the seven conserved families to provide totals for each *Rickettsia* genome (gray inset at bottom).

1. Edgar, R.C., *MUSCLE: a multiple sequence alignment method with reduced time and space complexity.* BMC Bioinformatics, 2004. **5**: p. 113.
2. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
3. Swofford, D., *PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4 ed.* , 1999, Sinauer: Sunderland, MA.



- 10 aa changes

ProP

- Proton symporter involved in osmoregulation
- osmolytes include proline, glycine betaine, stachydrine, pipecolic acid, ectoine, and taurine

ProP1-7

	1	2	3	4	5	6	7
***bellii* RML**	●		●	●	●	●	●
***bellii* OSU**	●		●	●	●	●	●
canadensis	●	●	●	●	●		●
***prowazekii* M**	●	●	●	●	●	●	●
***prowazekii* P**	●	●	●	●	●	●	●
typhi	●	●	●	●	●	●	●
felis	●	●	●	●	●	●	●
akari	●	●	●	●	●	◑	●
REIS	●	●	●	◉	◉		●
massiliae	●	●	●	●	●		●
montanensis*	?	●	?	?	?	?	?
peacockii	●	●	●	●	●		●
***rickettsii* SS**	●	●	●	●	●		●
***rickettsii* I**	●	●	●	●	●		●
conorii	●	●	●	●	●	●	●
sibirica	●	●	●	●	●	◪	●
africae	●	●	●	●	●		●

* ORF sequenced independent of genome

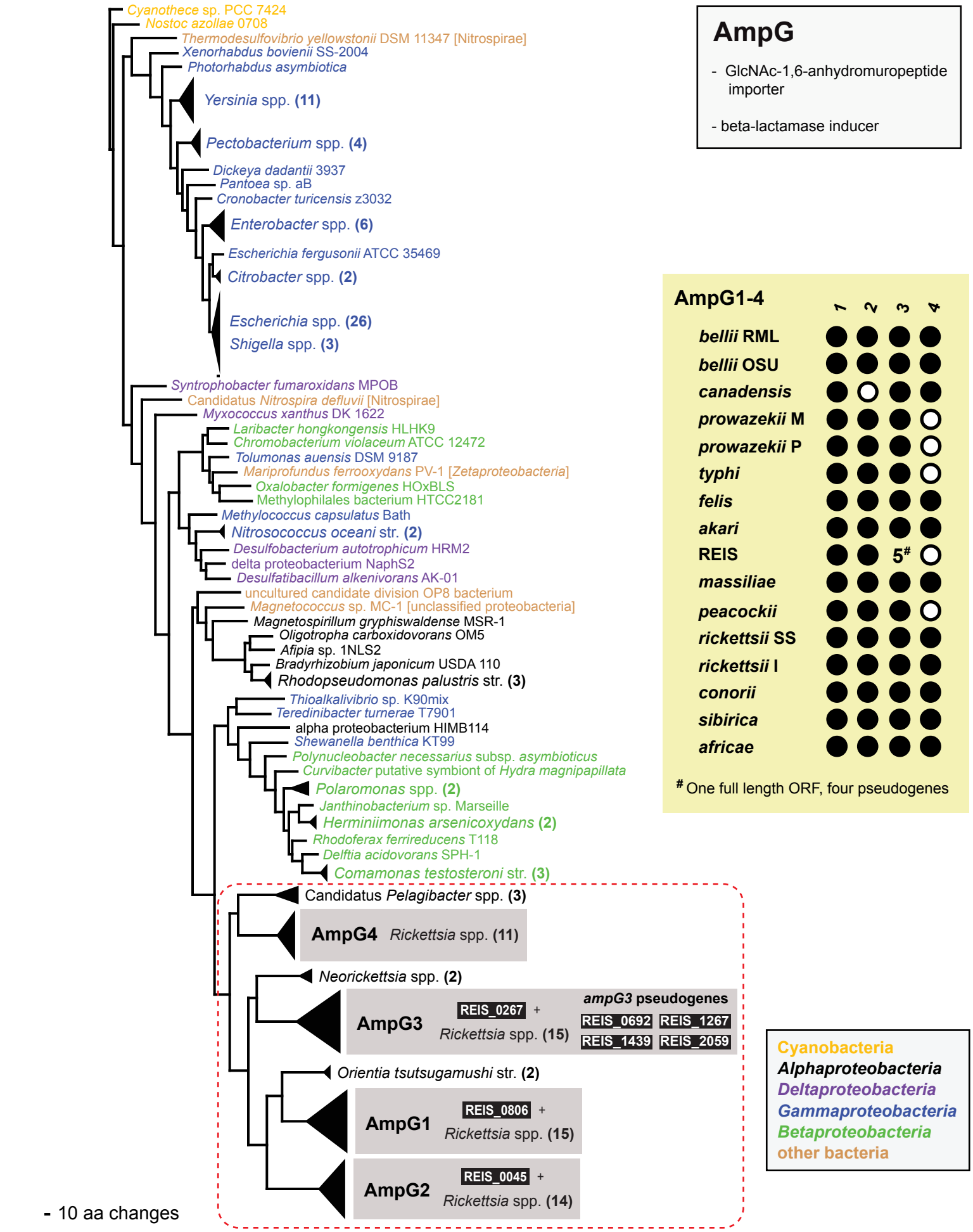
● **full length ORF**
◉ **full length ORF, multicopy**
◪ **split ORF**
◑ **ORF fragment**

Total ProP

	●	◪	◑	TOT
***bellii* RML**	11	0	2	13
***bellii* OSU**	9	0	2	11
canadensis	8	0	0	8
***prowazekii* M**	7	0	0	7
***prowazekii* P**	7	0	0	7
typhi	7	0	0	7
felis	14	2	1	17
akari	6	0	3	9
REIS	5	0	1	6
pREIS4	0	1	1	2
pRM	2	0	0	2
massiliae	6	0	0	6
montanensis	1	0	0	1
peacockii	6	0	0	6
***rickettsii* SS**	6	0	1	7
***rickettsii* I**	6	0	1	7
conorii	7	0	2	9
sibirica	6	1	4	11
africae	6	0	1	7

Fig. S14. Phylogeny estimation of AmpG proteins encoded within *Rickettsia* genomes. A total of 62 AmpG sequences were retrieved using *R. typhi* AmpG1 (YP_067419) as a query against the Rickettsiales database (taxid:766). These Rickettsiales sequences were combined with 105 blastp subjects acquired using the same query sequence against the NCBI non-redundant protein database excluding taxid:766. The total 167 AmpG sequences were aligned with MUSCLE v3.6 using default parameters [1, 2]. A phylogeny was estimated in PAUP* v4.0b10 (Altivec) under parsimony [3]. A heuristic search was implemented employing 250 random sequence additions, saving 10 trees per replication. A majority rule consensus tree was constructed for 40 equally parsimonious trees of tree score 8675. All nodes shown on the phylogram were present in all 40 trees, with collapsed clades (open and filled triangles) containing nodes not present in all trees. Inset at right shows the distribution of four conserved AmpG protein families (AmpG1-4) across *Rickettsia* genomes.

1. Edgar, R.C., *MUSCLE: a multiple sequence alignment method with reduced time and space complexity*. BMC Bioinformatics, 2004. **5**: p. 113.
2. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
3. Swofford, D., *PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4 ed.* , 1999, Sinauer: Sunderland, MA.


AmpG
- GlcNAc-1,6-anhydromuropeptide importer
- beta-lactamase inducer
AmpG1-4
1 2 3 4
bellii RML
bellii OSU
canadensis
prowazekii M
prowazekii P
typhi
felis
akari
REIS
massiliae
peacockii
rickettsii SS
rickettsii I
conorii
sibirica
africae
5#
One full length ORF, four pseudogenes
Cyanothece sp. PCC 7424
Nostoc azollae 0708
Thermodesulfovibrio yellowstonii DSM 11347 [Nitrospirae]
Xenorhabdus bovienii SS-2004
Photorhabdus asymbiotica
Yersinia spp. (11)
Pectobacterium spp. (4)
Dickeya dadantii 3937
Pantoea sp. aB
Cronobacter turicensis z3032
Enterobacter spp. (6)
Escherichia fergusonii ATCC 35469
Citrobacter spp. (2)
Escherichia spp. (26)
Shigella spp. (3)
Syntrophobacter fumaroxidans MPOB
Candidatus Nitrospira defluvii [Nitrospirae]
Myxococcus xanthus DK 1622
Laribacter hongkongensis HLHK9
Chromobacterium violaceum ATCC 12472
Tolumonas auensis DSM 9187
Mariprofundus ferrooxydans PV-1 [Zetaproteobacteria]
Oxalobacter formigenes HOxBLS
Methylophilales bacterium HTCC2181
Methylococcus capsulatus Bath
Nitrosococcus oceani str. (2)
Desulfobacterium autotrophicum HRM2
delta proteobacterium NaphS2
Desulfatibacillum alkenivorans AK-01
uncultured candidate division OP8 bacterium
Magnetococcus sp. MC-1 [unclassified proteobacteria]
Magnetospirillum gryphiswaldense MSR-1
Oligotropha carboxidovorans OM5
Afipia sp. 1NLS2
Bradyrhizobium japonicum USDA 110
Rhodopseudomonas palustris str. (3)
Thioalkalivibrio sp. K90mix
Teredinibacter turnerae T7901
alpha proteobacterium HIMB114
Shewanella benthica KT99
Polynucleobacter necessarius subsp. asymbioticus
Curvibacter putative symbiont of Hydra magnipapillata
Polaromonas spp. (2)
Janthinobacterium sp. Marseille
Herminiimonas arsenicoxydans (2)
Rhodoferax ferrireducens T118
Delftia acidovorans SPH-1
Comamonas testosteroni str. (3)
Candidatus Pelagibacter spp. (3)
AmpG4 Rickettsia spp. (11)
Neorickettsia spp. (2)
AmpG3
REIS_0267 + Rickettsia spp. (15)
ampG3 pseudogenes
REIS_0692 REIS_1267
REIS_1439 REIS_2059
Orientia tsutsugamushi str. (2)
AmpG1
REIS_0806 + Rickettsia spp. (15)
AmpG2
REIS_0045 + Rickettsia spp. (14)
- 10 aa changes
Cyanobacteria
Alphaproteobacteria
Deltaproteobacteria
Gammaproteobacteria
Betaproteobacteria
other bacteria