

## Supplementary Materials

Concatenated ScaA and TSA56 Surface Antigen Sequences Reflect Genome-Scale Phylogeny of

*Orientia tsutsugamushi*: An Analysis Including Two Genomes from Taiwan

Nicholas T. Minahan<sup>1</sup>, Tsai-Ying Yen<sup>2</sup>, Yue-Liang Leon Guo<sup>1,3</sup>, Pei-Yun Shu<sup>2</sup> and Kun-Hsien Tsai<sup>1,4,\*</sup>

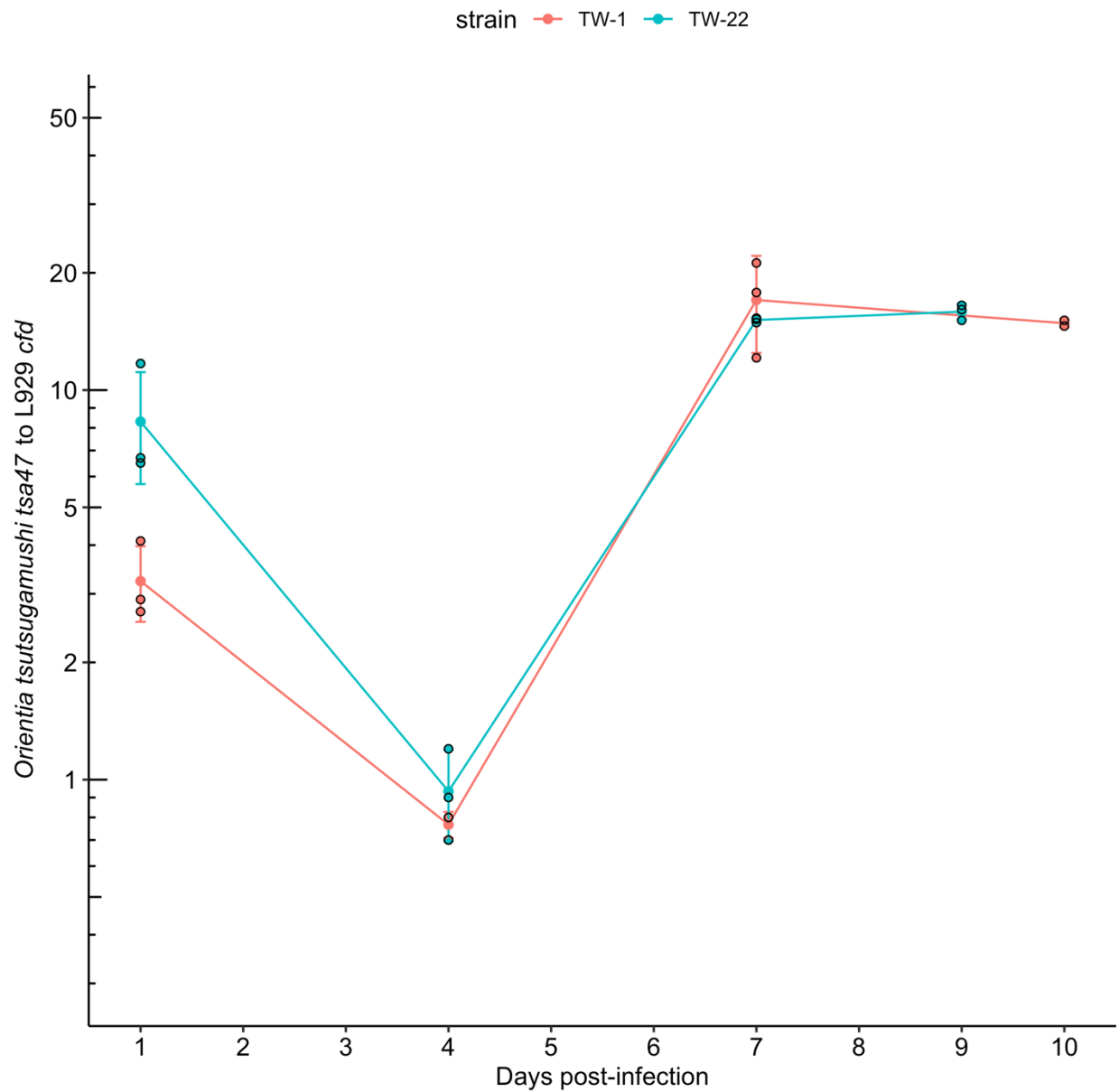
<sup>1</sup> Institute of Environmental and Occupational Health Sciences, College of Public Health, National Taiwan University, Taipei 100025, Taiwan; f05841021@ntu.edu.tw (N.T.M.); leonguo@ntu.edu.tw (Y.-L.L.G.)

<sup>2</sup> Centers for Diagnostics and Vaccine Development, Centers for Disease Control, Ministry of Health and Welfare, Taipei 115210, Taiwan; yenty@cdc.gov.tw (T.-Y.Y.); cdcpysu@gmail.com (P.-Y.S.)

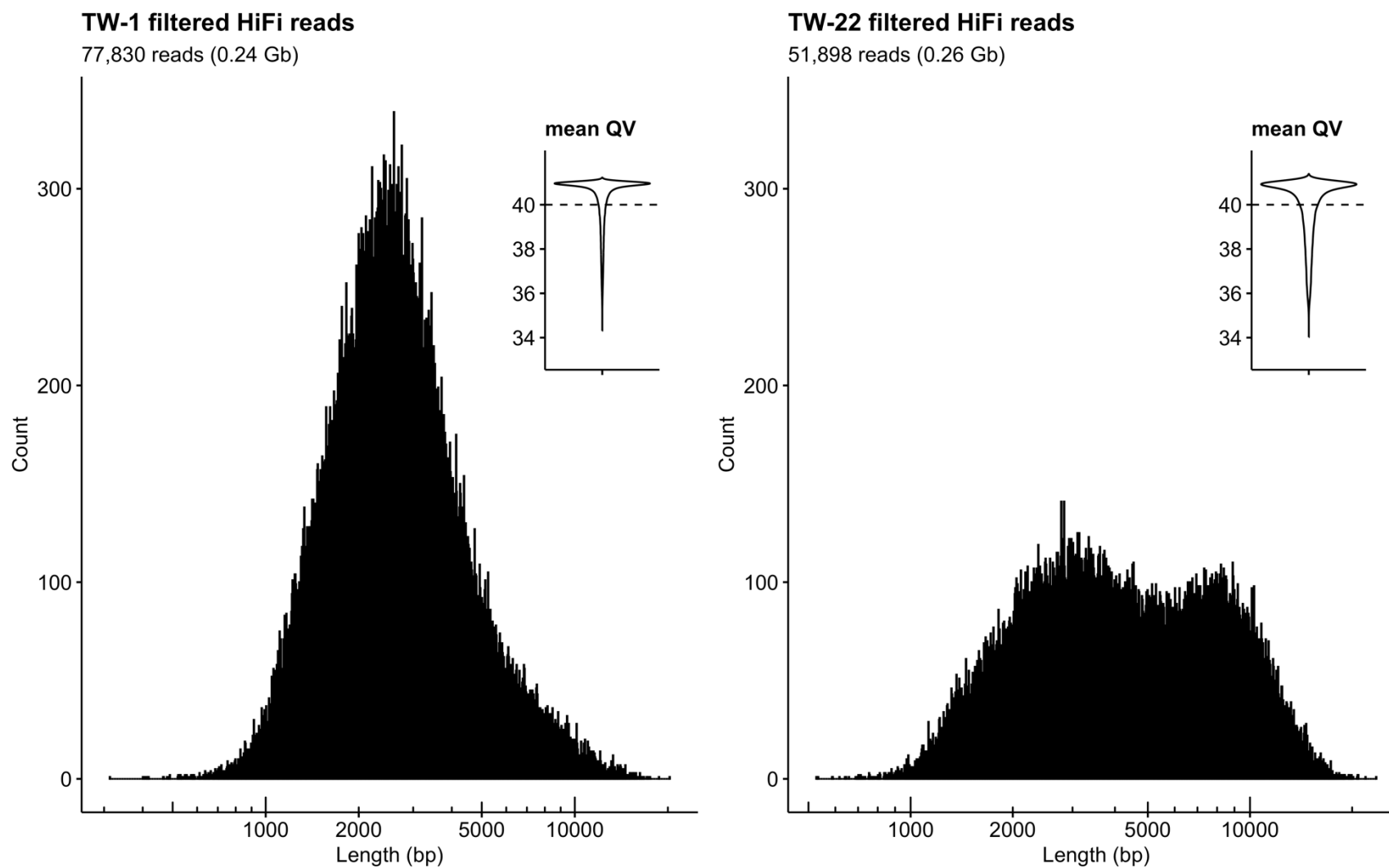
<sup>3</sup> Department of Environmental and Occupational Medicine, National Taiwan University (NTU) College of Medicine and NTU Hospital, Taipei 100025, Taiwan

<sup>4</sup> Global Health Program, College of Public Health, National Taiwan University, Taipei 100025, Taiwan

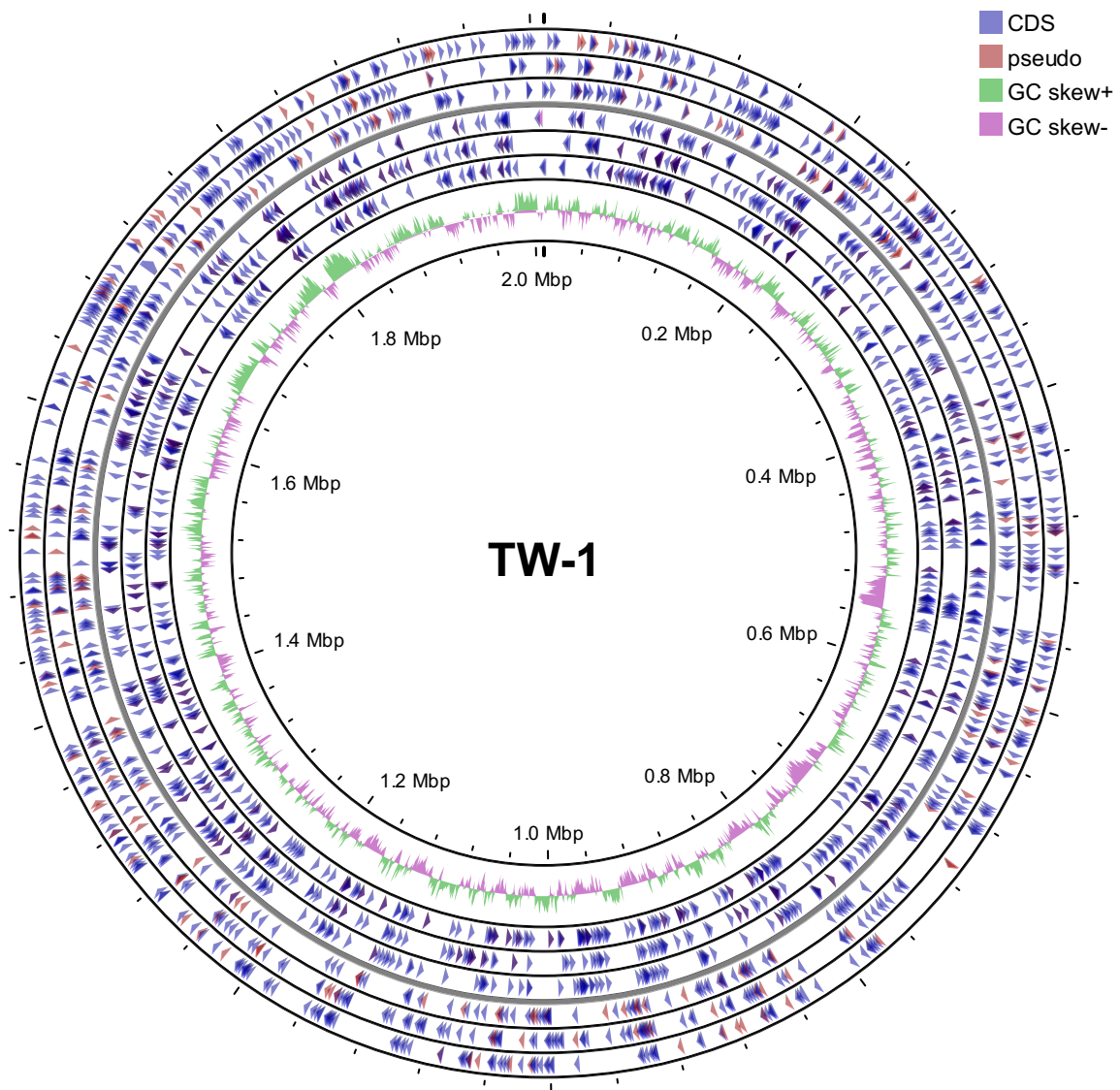
\* Correspondence: kunhtsai@ntu.edu.tw



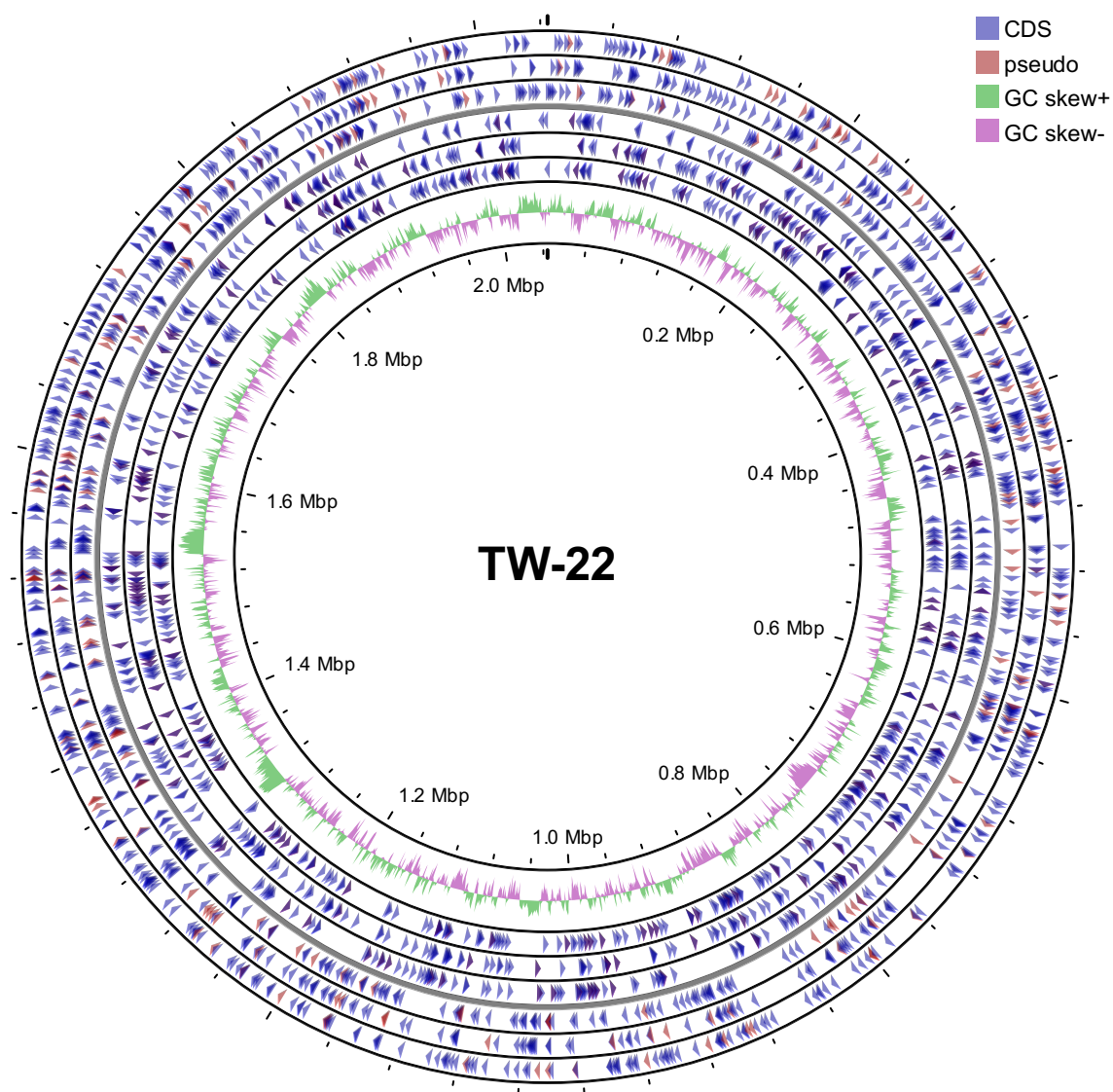
**Figure S1. Growth curves.** Mean ratio of *Orientia tsutsugamushi* (OT) to L929 cells in the supernatant of cultures (passage 8) of strains TW-1 and TW-22 based on single-copy targets. Each point represents mean values of triplicate measurements for an individual 75 cm<sup>2</sup> (T75) flask. Colored dots represent mean values, and error bars represent standard deviation.



**Figure S2. Summary of filtered HiFi reads used for de novo assembly.** Quality values (QVs) were capped at 41 (representing  $<1$  in 10,000 probability of error) using Biopython 1.81, and mean QVs were calculated using qckitfastq 1.16.0.



**Figure S3. Circular chromosome of *Orientia tsutsugamushi* strain TW-1.** CDS = protein-coding sequence; pseudo = pseudogene.



**Figure S4. Circular chromosome of *Orientia tsutsugamushi* strain TW-22.** CDS = protein-coding sequence; pseudo = pseudogene.

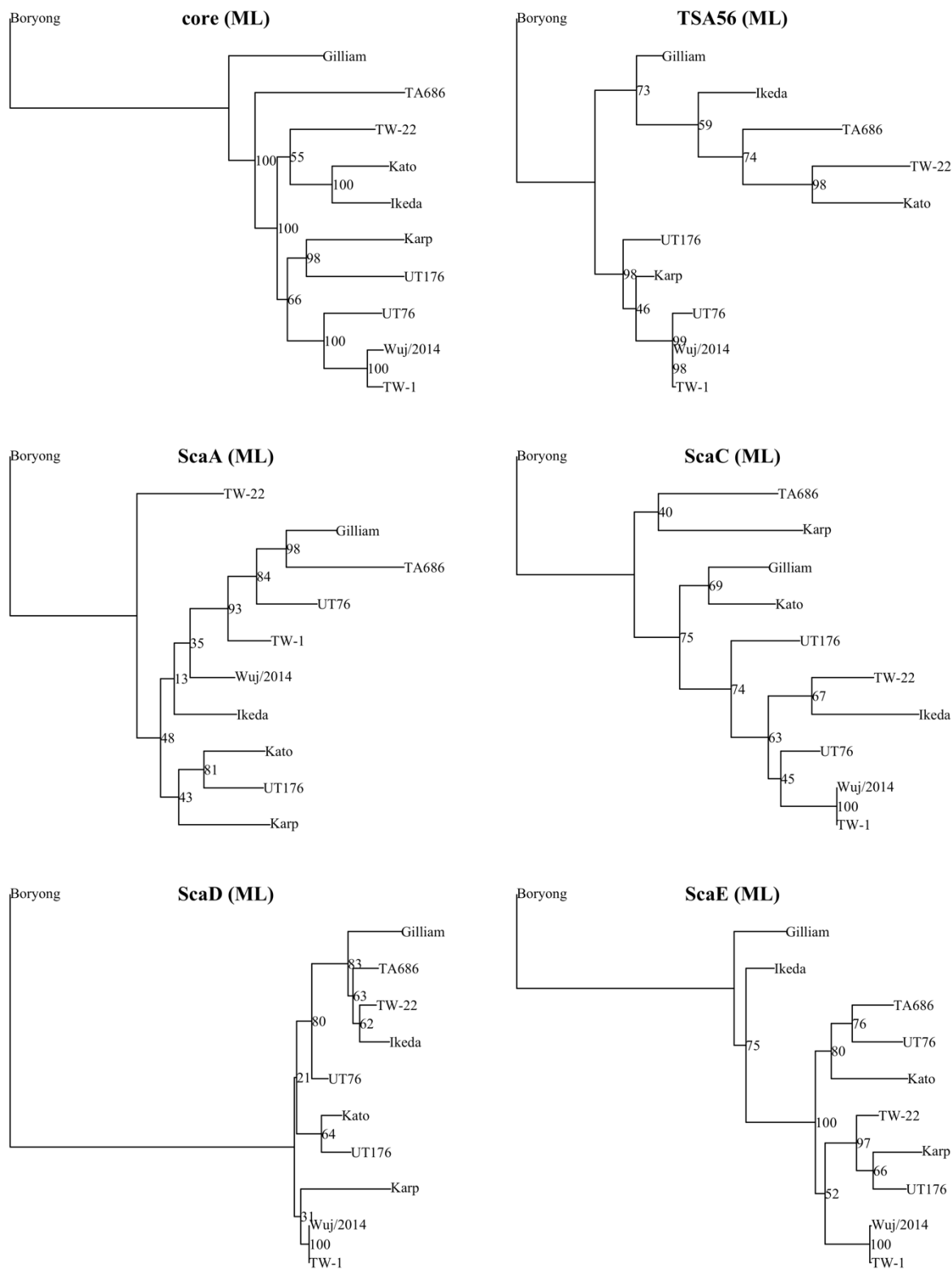
**Table S1. Pairwise amino acid alignments of *Orientia tsutsugamushi* surface antigens.**

Strain	TSA56	ScaA	ScaB	ScaC	ScaD	ScaE	ScaF
Boryong	<b>532 aa (reference)</b>	<b>1461 aa (reference)</b>	649 aa; <b>2 copies</b>	<b>526 aa (reference)</b>	<b>998 aa (reference)</b>	<b>760 aa (reference)</b>	<b>absent</b>
Ikeda	516 aa (532 aa; 18,540 bits; 72.74% id)	1508 aa (1532 aa; 57,020 bits; 75.52% id)	<b>absent</b>	520 aa (526 aa; 23,620 bits; 88.21% id)	834 aa (998 aa; 30,700 bits; 69.34% id)	753 aa (778 aa; 28,200 bits; 77.25% id)	<b>absent</b>
Gilliam	524 aa (532 aa; 21,860 bits; 82.89% id)	1518 aa (1529 aa; 58,610 bits; 77.17% id)	<b>absent</b>	517 aa (526 aa; 23,590 bits; 88.78% id)	745 aa (998 aa; 26,620 bits; 63.53% id)	757 aa (774 aa; 28,280 bits; 76.74% id)	<b>absent</b>
Karp	532 aa (534 aa; 22,650 bits; 84.08% id)	1514 aa (1529 aa; 56,810 bits; 75.67% id)	<b>absent</b>	517 aa (526 aa; 23,380 bits; 87.83% id)	988 aa (1000 aa; 35,330 bits; 73.70% id)	755 aa (772 aa; 26,510 bits; 72.54% id)	645 aa (645 aa; 32,750 bits; 99.38% id)
Kato	529 aa (544 aa; 17,630 bits; 68.93% id)	1447 aa (1470 aa; 59,320 bits; 79.52% id)	<b>absent</b>	520 aa (526 aa; 23,780 bits; 89.54% id)	872 aa (998 aa; 34,060 bits; 73.35% id)	747 aa (775 aa; 25,470 bits; 71.10% id)	<b>absent</b>
TA686	532 aa (536 aa; 19,540 bits; 75.37% id)	1446 aa (1464 aa; 57,020 bits; 77.66% id)	<b>absent</b>	517 aa (526 aa; 23,570 bits; 88.78% id)	777 aa (998 aa; 28,500 bits; 66.03% id)	756 aa (774 aa; 26,720 bits; 73.00% id)	<b>645 aa (reference)</b>
UT76	536 aa (543 aa; 21,380 bits; 81.22% id)	1526 aa (1531 aa; 57,350 bits; 75.11% id)	<b>absent</b>	517 aa (526 aa; 23,360 bits; 87.26% id)	1035 aa (1036 aa; 37,150 bits; 74.90% id)	751 aa (772 aa; 26,600 bits; 72.80% id)	<b>absent</b>
UT176	533 aa (534 aa; 22,630 bits; 83.52% id)	1508 aa (1533 aa; 56,530 bits; 75.21% id)	<b>absent</b>	520 aa (526 aa; 23,830 bits; 88.97% id)	872 aa (998 aa; 33,260 bits; 71.94% id)	747 aa (765 aa; 26,690 bits; 73.33% id)	<b>absent</b>
Wuj/2014	535 aa (542 aa; 21,570 bits; 81.00% id)	1524 aa (1544 aa; 58,630 bits; 77.07% id)	<b>absent</b>	523 aa <sup>†</sup> (526 aa; 23,630 bits; 88.02% id)	988 aa (1002 aa; 37,420 bits; 77.15% id)	749 aa (774 aa; 26,480 bits; 73.51% id)	<b>absent</b>
TW-1	535 aa (542 aa; 21,480 bits; 80.63% id)	1531 aa (1547 aa; 56,330 bits; 74.53% id)	<b>absent</b>	523 aa <sup>†</sup> (526 aa; 23,630 bits; 88.02% id)	941 aa (999 aa; 36,700 bits; 76.38% id)	749 aa (774 aa; 26,420 bits; 73.39% id)	<b>absent</b>
TW-22	524 aa (539 aa; 18,160 bits; 71.24% id)	1507 aa (1525 aa; 57,870 bits; 79.38% id)	<b>absent</b>	520 aa (526 aa; 23,900 bits; 89.16% id)	807 aa (998 aa; 29,650 bits; 68.04% id)	748 aa (772 aa; 26,180 bits; 72.80% id)	<b>absent</b>
	68.93-99.63% identity	71.32-89.16% identity	NA	86.35-100% identity	63.53-95.41% identity	71.10-99.73% identity	NA

aa = amino acids; id = identity; <sup>†</sup> = identical

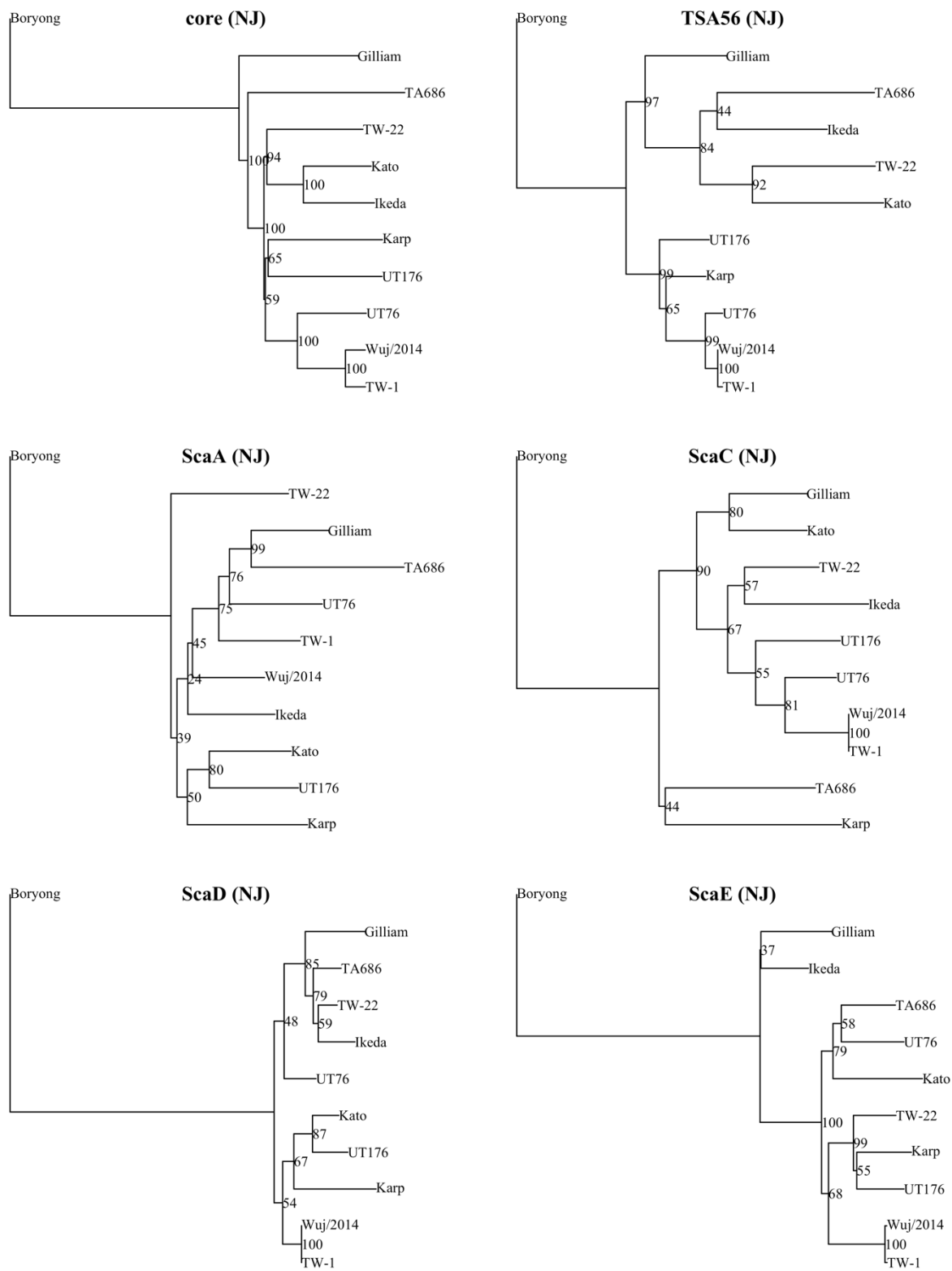
**Table S2. Summary of NCBI locus tags for surface antigens.**

	<b>Boryong</b> (AM494475)	<b>Ikeda</b> (AP008981)	<b>Gilliam</b> (LS398551)	<b>Karp</b> (LS398548)	<b>Kato</b> (LS398550)	<b>TA686</b> (LS398549)	<b>UT76</b> (LS398552)	<b>UT176</b> (LS398547)	<b>Wuj/2014</b> (CP044031)	<b>TW-1</b> (CP142421)	<b>TW-22</b> (CP142420)
<b>TSA56</b>	OTBS_0602	OTT_0945	GILLIAM_02257	KARP_02001	KATO_00907	TA686_02342	UT76HP_01865	UT176_00941	F0363_02480	QU601_001572	QU600_000493
<b>ScaA</b>	OTBS_0102	OTT_1215	GILLIAM_01513	KARP_00798	KATO_00211	TA686_00130	UT76HP_00328	UT176_00529	F0363_06280	QU601_000243	QU600_001479
<b>ScaB</b>	OTBS_0864 OTBS_2137										
<b>ScaC</b>	OTBS_1686	OTT_0460	GILLIAM_01556	KARP_01628	KATO_01770	TA686_01405	UT76HP_01782	UT176_01268	F0363_02065	QU601_001499	QU600_000816
<b>ScaD</b>	OTBS_1913	OTT_1741	GILLIAM_01380	KARP_00576	KATO_00608	TA686_00458	UT76HP_02170	UT176_00383	F0363_04340	QU601_001941	QU600_001734
<b>ScaE</b>	OTBS_2126	OTT_1766	GILLIAM_00783	KARP_00606	KATO_00640	TA686_00866	UT76HP_02139	UT176_00408	F0363_04165	QU601_001910	QU600_001689
<b>ScaF</b>				KARP_01610		TA686_01679					

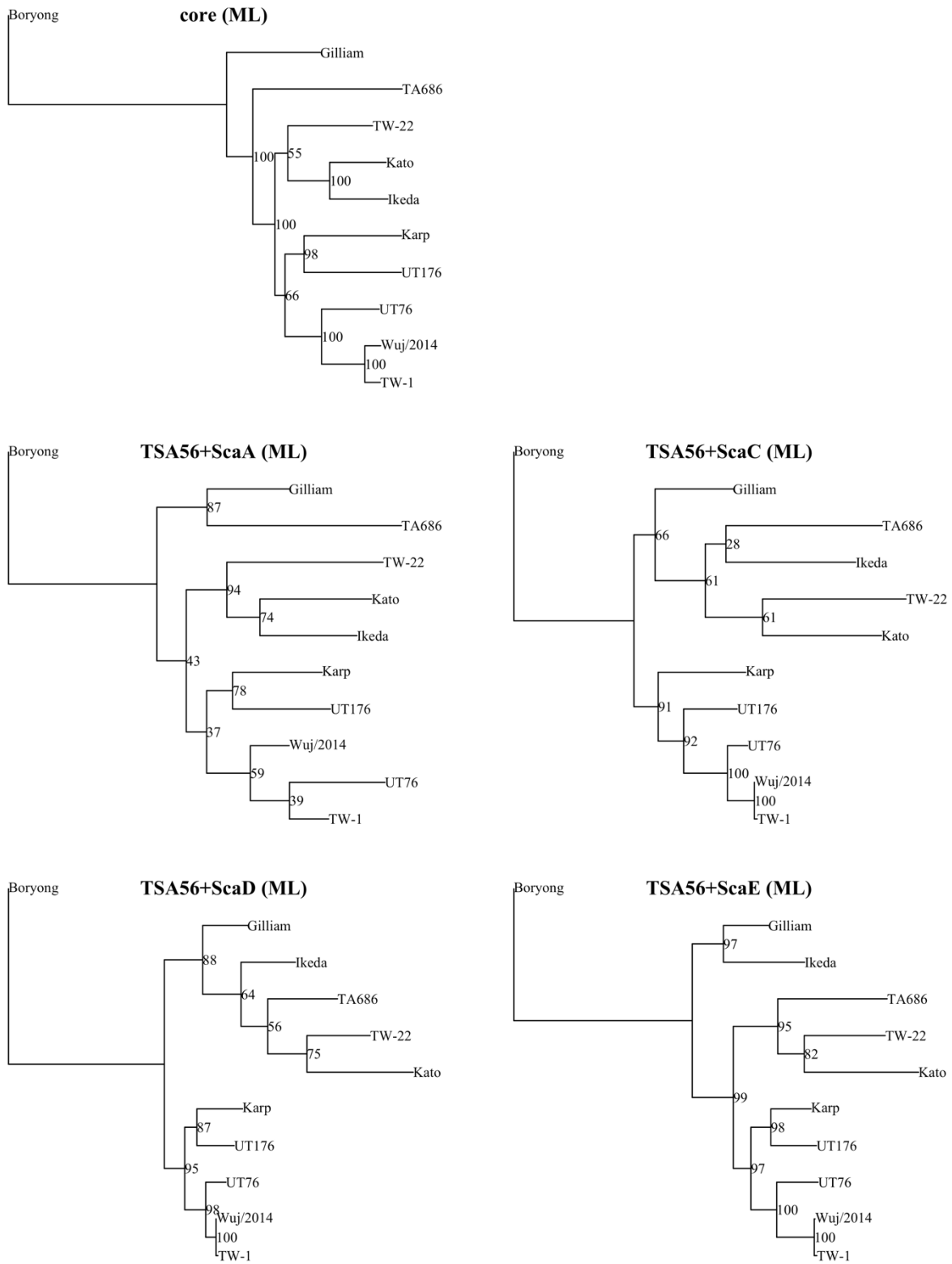


**Figure S5. Maximum likelihood phylogenetic trees based on individual amino acid sequences.**

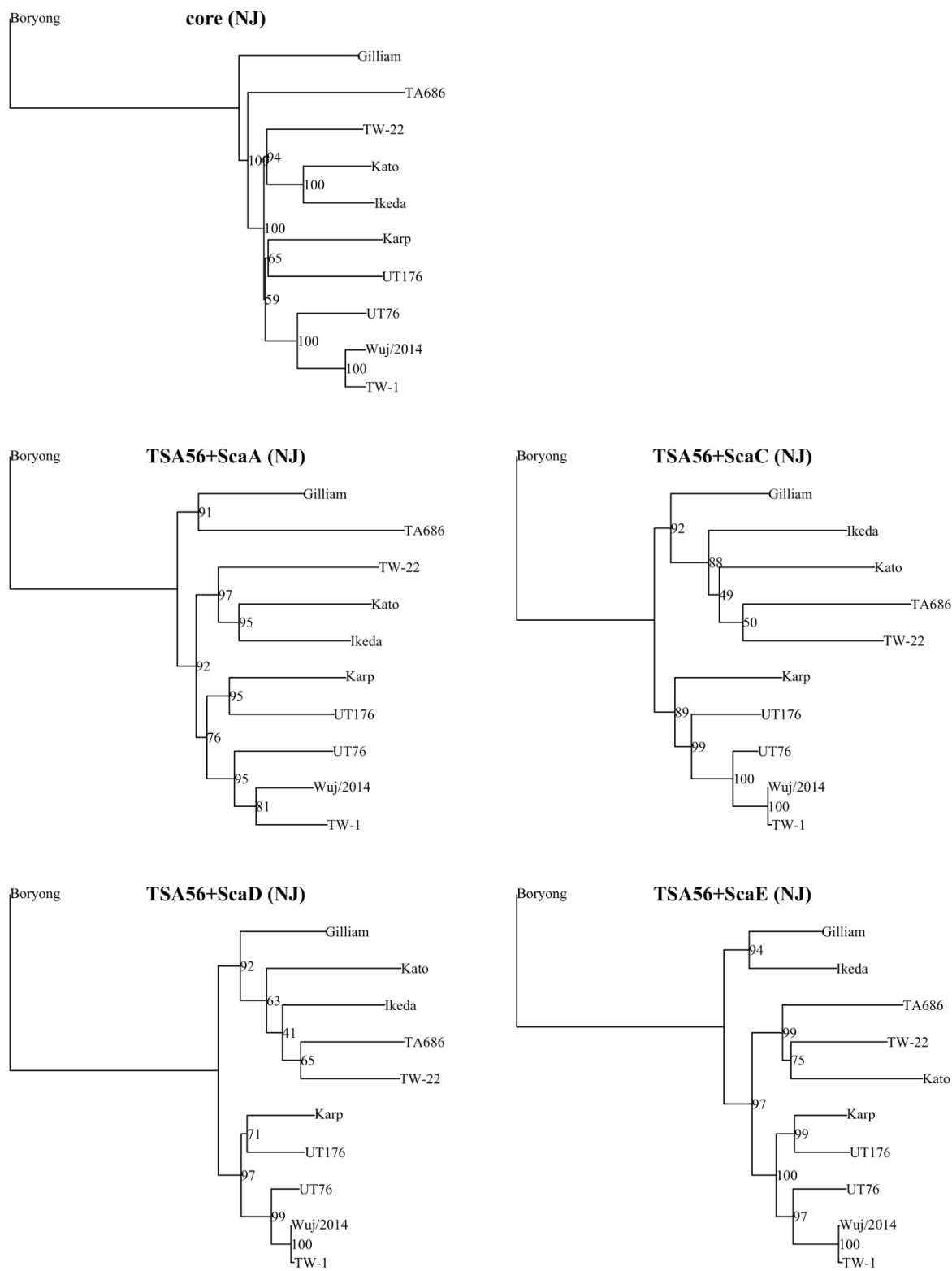




**Figure S6. Neighbor-joining phylogenetic trees based on individual amino acid sequences.**



**Figure S7. Maximum likelihood phylogenetic trees based on concatenated amino acid sequences.**



**Figure S8. Neighbor-joining phylogenetic trees based on concatenated amino acid sequences.**