Received: 30 March 2023 | Accepted: 13 October 2023

DOI: 10.1111/syen.12613

Systematic Entomology

ORIGINAL ARTICLE

Phylogenomics improves the phylogenetic resolution and provides strong evidence of mito-nuclear discordance in two genera of a New Zealand cicada (Hemiptera: Cicadidae) species radiation

Mark Stukel[1] | Alexandra E. Porczak[1,2] | Eric R. L. Gordon[1] | Jason Vailionis[1,3] | Diler Haji[1,4] | Thomas R. Buckley[5] | Alan R. Lemmon[6] | Emily Moriarty Lemmon[7] | Chris Simon[1]

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut, USA

[2]Department of Physiology and Neurobiology, University of Connecticut, Storrs, Connecticut, USA

[3]Department of Biological Sciences, University of Rhode Island, Kingston, Rhode Island, USA

[4]Department of Integrative Biology, University of California, Berkeley, Berkeley, California, USA

[5]Manaaki Whenua–Landcare Research, Auckland, New Zealand

[6]Department of Scientific Computing, Florida State University, Tallahassee, Florida, USA

[7]Department of Biological Science, Florida State University, Tallahassee, Florida, USA

Correspondence
Mark Stukel, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA.
Email: mark.stukel@uconn.edu

Funding information
Fulbright Association; Linnean Society of London; New Zealand MBIE, Grant/Award Number: CO9X2201; Systematics Association; University of Connecticut; National Science Foundation, Grant/Award Numbers: DEB 16-55891, DEB 09-55849, DEB 07-20664, DEB 04-22386

Abstract

Rapid species radiations present difficulties for phylogenetic reconstruction due to lack of phylogenetic information and processes such as deep coalescence/incomplete lineage sorting and hybridization. Phylogenomic data can overcome some of these difficulties. In this study, we use anchored hybrid enrichment (AHE) nuclear phylogenomic data and mitochondrial genomes recovered from AHE bycatch with several concatenated and coalescent approaches to reconstruct the poorly resolved radiation of the New Zealand cicada species in the genera *Kikihia* Dugdale and *Maoricicada* Dugdale. Compared with previous studies using only three to five Sanger-sequenced genes, we find increased resolution across our phylogenies, but several branches remain unresolved due to topological conflict among genes. Some nodes that are strongly supported by traditional support measures like bootstraps and posterior probabilities still show significant gene and site concordance conflict. In addition, we find strong mito-nuclear discordance; likely the result of interspecific hybridization events in the evolutionary history of *Kikihia* and *Maoricicada*.

KEYWORDS
anchored hybrid enrichment, coalescence, concordance factors, hybridization, mito-nuclear discordance, phylogenomics, species radiations

INTRODUCTION

In an age where the planet's biodiversity is decreasing, some important questions in biology are how is biodiversity generated, how does it spread and what happens when formerly isolated lineages come into secondary contact? Evolutionary histories of species provide insight into the origin of biodiversity and the processes that shape it. One phenomenon of biodiversity generation, rapid species

radiations, makes evolutionary relationships extremely difficult to reconstruct because they take place in relatively short periods of time. The branches separating groups of taxa are short (i.e., contain little information), and descendent taxa are more likely to have experienced random inheritance of ancestral polymorphisms. This process, called incomplete lineage sorting (ILS), leads to individual gene trees that often disagree with each other and the underlying species tree (Degnan & Rosenberg, 2006; Maddison, 1997; Maddison & Knowles, 2006). In addition, rapid radiations may also involve extensive hybridization (Mallet et al., 2016), either through speciation with more or less continuous gene flow (Marsden et al., 2011; Morales et al., 2017; Nosil, 2008) or speciation involving divergence followed by range expansion and later reinforced by secondary contact between formerly isolated, related lineages (Butlin et al., 2008). Both ILS and hybridization consequently make the inference of species relationships in rapid radiations difficult.

Phylogenetic work using Sanger-sequenced genes has largely been unsuccessful in resolving phylogenies of rapid radiations due to the lack of phylogenetic information and the processes of ILS and hybridization described above (Meusemann et al., 2010; Simon et al., 2009; Whitfield & Kjer, 2008). However, the dramatic decrease in cost and increase in throughput of DNA sequencing, combined with target-capture methods that allow for the recovery of hundreds of orthologous genes across specimens, has led to great improvement in resolving rapid radiations (Brunke et al., 2021; Giarla & Esselstyn, 2015; Irisarri et al., 2018; Pyron et al., 2014). Phylogenomic methods using these genome-scale data can take advantage of the increased information and use many individual gene trees to reconstruct the underlying species tree to combat the confounding effect of ILS.

The colonization of New Zealand (NZ) by cicadas and their resulting diversification is an excellent example of an adaptive radiation (Fleming, 1975). NZ has experienced over 80 million years of geographic isolation from other larger landforms and is known for its large number of endemic species (Buckley & Simon, 2007; Marske & Boyer, 2022; Shepherd et al., 2022). Three of the five endemic cicada genera found in NZ, *Maoricicada* Dugdale, *Kikihia* Dugdale and *Rhodopsalta* Dugdale descend from a single colonization event, possibly from New Caledonia (Figure 1; Arensburger, Buckley, et al., 2004; Arensburger, Simon, & Holsinger, 2004; Buckley et al., 2002). As typical of cicada species, these genera each have a distinct, species-specific male song, which is a useful tool for distinguishing species (Dugdale & Fleming, 1978).

Maoricicada

The NZ cicada genus *Maoricicada* contains five lowland and nine alpine or subalpine species (one of which contains four subspecies). Radiation of the alpine group was stimulated by a period of mountain building (Buckley & Simon, 2007), which began during the Miocene with a bout of increased uplift during the Pliocene about 5 million years ago and continues to this day (Craw et al. 2003; King 2000). Pleistocene glaciation and accompanying climate fluctuations were also an important force, especially on the South Island (Buckley & Simon, 2007).

Kikihia

The genus *Kikihia* has around 30 species, approximately half of which are currently undescribed (Marshall et al., 2008). These species are found in a variety of habitat types across the NZ islands, with some groups specialized for forest understory, others for evergreen forests and shrubs and still others for dry grasslands and scrub. *Kikihia* species are found in a wide variety of habitats with most occurring from sea level to mid elevations, and a few species occurring in subalpine habitats. Previous research on *Kikihia* has found that the majority of species originated during a species radiation 3–5 million years ago, during a period of climate deterioration following the Pliocene mountain-building (Arensburger, Buckley, et al., 2004; Arensburger, Simon, & Holsinger, 2004; Marshall et al., 2008).

Rhodopsalta

In contrast to *Maoricicada* and *Kikihia*, the genus *Rhodopsalta* has no evidence of aspecies radiation and has only three species: *R. microdora* (Hudson), *R. leptomera* (Myers) and *R. cruentata* (Fabricius). Based on current *Rhodopsalta* species distributions and experimental work on NZ cicada temperature regulation, it appears that this genus has a lower tolerance to cool temperatures than its related genera, which would have forced them into more northern lowland glacial refugia (Bator et al., 2021; Heath et al., 2022). The extant species diverged in the Pliocene and Pleistocene, roughly 5 and 2 Ma with one species restricted to coastal North Island, and the other two found on both islands in wetter versus drier lowland areas, respectively. Clades within these species show some phylogeographic structure (Bator et al., 2021).

Lack of resolution within species groups

Previous work on NZ cicadas has found evidence of both recent and ancient hybridization based on cicada courtship songs, morphology, and mitochondrial and nuclear phylogenies. In *Kikihia*, there is evidence for at least 11 hybrid zones involving 20 different species pairs that come into secondary contact (Marshall et al., 2008, 2011; Wade, 2014). Several clades of *Kikihia* species have populations where the mitochondrial haplotype conflicts with the species-specific song, indicating past mitochondrial introgression, and putative F1 hybrids have been identified in contact zones between closely related species based on intermediate song characters, mtDNA and microsatellites, indicating ongoing hybridization (Banker et al., 2017; Marshall et al., 2011; Wade, 2014). There is no evidence for current hybridization in *Maoricicada* or *Rhodopsalta*, but discordance between mitochondrial and nuclear phylogenies, combined with information from song structures, provides evidence of ancient hybridization in *Maoricicada* (Buckley et al., 2006).

Because of low phylogenetic information, ILS, and the abovementioned hybridization, previous Simon lab studies using

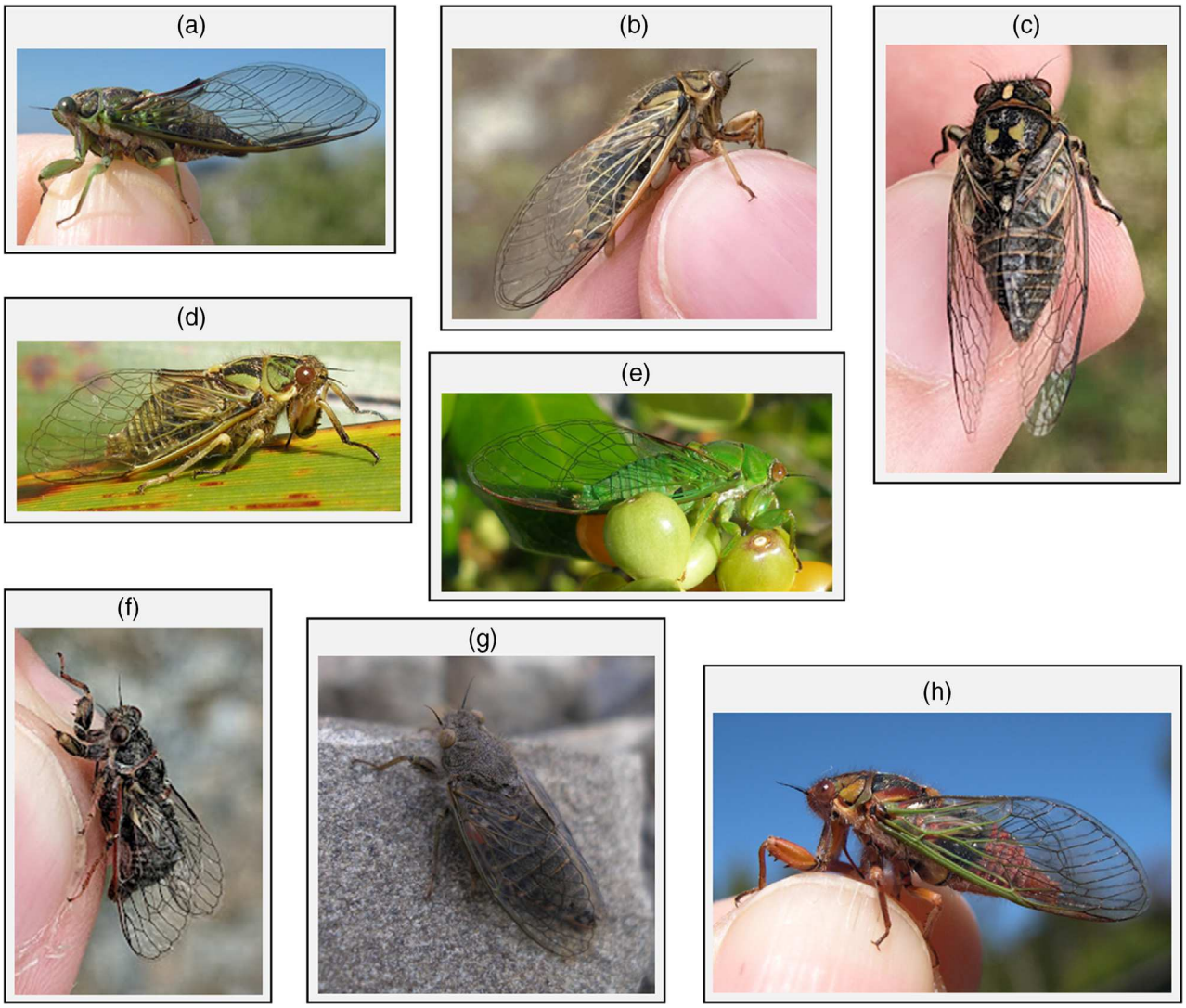


FIGURE 1 Representatives of the cicada species used in this study. (a) Shade singer *Kikihia scutellaris*. (b) Grass cicada *K. angusta*. (c) Scrub cicada *K.* 'acoustica'. (d) Grass cicada *K. muta*. (e) Green foliage cicada *K. ochrina*. (f) Lowland *Maoricicada campbelli*. (g) Subalpine *M. mangu*. (h) *Rhodopsalta microdora*. Photo credits: (a), (b), (d), (e), (g) and (h): C. Simon; (c) and (f): M. Stukel.

Sanger-sequenced loci have not been able to fully resolve the phylogeny of this NZ cicada radiation. In *Maoricicada*, the main alpine radiation is overall poorly resolved, and many species show highly unstable positions depending on different analyses (Buckley et al., 2006; Buckley & Simon, 2007). In *Kikihia*, mtDNA analyses have identified two forest cicadas (the 'shade singers') that are sequential sister species to the remainder of the genus and a large radiation of four species groups largely distinguished by habitat type: evergreen forest and shrub cicadas (the 'Cutora group'), dry scrub cicadas (the 'Rosea group'), grass and shrub cicadas (the 'Muta group') and a group containing cicadas from all three previous habitat types (the 'Westlandica group') (Marshall et al., 2008). However, analyses using nuclear as well as mitochondrial genes have not been able to confirm the monophyly of these groups and the relationships among them (Banker et al., 2017).

Due to the confounding effects of ILS and hybridization, phylogenomic data involving hundreds of loci as well as complete mitochondrial genomes are likely required to resolve the relationships of these cicadas. MtDNA is less likely to be affected by ILS because of its faster rate of coalescence, and mitochondrial-nuclear phylogenetic discordance is a good indicator of past hybridization. In this study, we take anchored hybrid enrichment (AHE) capture data (Lemmon et al., 2012) and mitochondrial genomes assembled from AHE bycatch and use a variety of concatenated and coalescent species tree methods to resolve the relationships among these cicadas.

METHODS

Cicada specimens for this study were collected and identified by song and morphology (Bator et al., 2021; Buckley et al., 2006; Dugdale & Fleming, 1978; Fleming, 1984; Lane, 1984, 1995; Marshall et al., 2008, 2011) over the course of 20+ years of field trips,

TABLE 1 List of *Kikihia*, *Maoricicada* and *Rhodopsalta* specimens, number of AHE loci recovered and amount of mtDNA recovered.

Species	Name in tree	Simon lab code	Lemmon code	Latitude	Longitude	AHE loci recovered	mtDNA recovered (bp)
K. 'acoustica'	K._'acoustica'_MK	99.NZ.MK.LOH.01	I27897	−44.2373	169.8228	493	11,864
K. angusta	K._angusta_SC	08.NZ.SC.IDN.12	I6885	−44.7817	170.8701	501	13,907
K. 'aotea-east'	K._'aotea_east'_HB	02.NZ.HB.POR.08	I6887	−40.3081	176.6659	502	9971
K. 'aotea-west'	K._'aotea_west'_TO	02.NZ.TO.RCG.14	I6889	−39.1919	175.5317	505	7777
K. 'astragali'	K._ 'astragali'_NN_1	14.NZ.NN.STE.02	I6891	−40.7940	172.4601	483	9199
K. 'astragali'	K._ 'astragali'_NN_2	14.NZ.NN.STE.01	I27898	−40.7940	172.4601	505	13,282
K. 'balaena'	K._ 'balaena'_KA	02.NZ.KA.WBS.01	I27899	−42.4869	173.2018	495	11,064
K. cauta	K._cauta_WN	99.NZ.99–14	I6893	−40.9472	175.1133	502	8682
K. convicta	K._convicta_NF	03.NF.NF.NFI.05	I27900	−29.0333	167.9500	495	13,675
K. cutora cumberi	K._cutora_cumberi_TO_1	02.NZ.TO.RCG.03	I6839	−39.1919	175.5317	505	7407
K. cutora cumberi	K._cutora_cumberi_TO_2	02.NZ.TO.RCG.25	I27901	−39.1919	175.5317	482	13,837
K. cutora cutora	K._cutora_cutora_ND	02.NZ.ND.MAU.09	I27902	−36.1137	174.3008	497	7189
K. cutora exulis	K._cutora_exulis_KE	98.NZ.KE.RAO.46	I27940	−29.2457	177.9273	491	9532
K. dugdalei	K._dugdalei_BP	02.NZ.BP.CRE.1	I27903	−38.1542	176.2640	478	14,084
K. 'flemingi'	K._'flemingi'_NN	11.NZ.NN.LGU.02	I6895	−41.5990	172.8817	496	13,819
K. horologium	K._horologium_NC	14.NZ.NC.NIG.05	I6897	−42.9100	171.5798	505	11,588
K. laneorum	K._laneorum_BP	14.NZ.BP.KMS.01	I6899	−37.8711	175.9296	494	6671
K. longula	K._longula_CH	K.longula1997	I27904	−44.0000	176.0000	489	10,783
K. 'murihikua'	K._'murihikua'_OL	08.NZ.OL.RSC.01	I6901	−44.5020	168.7825	499	14,156
K. 'muta east'	K._'muta_east'_HB	02.NZ.HB.OCB.01	I6903	−39.7429	177.0107	505	14,330
K. muta muta	K._muta_muta_NC	12.NZ.NC.BAL.01	I6905	−42.8676	172.7716	502	6593
K. muta muta	K._muta_muta_WI	01.NZ.WI.FER.03	I9971	−40.2299	175.5716	505	7903
K. 'muta tuta'	K._'muta_tuta'_NC	12.NZ.NC.WAI.14	I6843	−42.6804	172.9553	506	12,729
K. 'nelsonensis'	K._nelsonensis_MB	05.NZ.MB.TFL.01	I27905	−41.3027	173.5767	491	13,940
K. 'nelsonensis'	K._nelsonensis_NN	14.NZ.NN.JDH.01	I6907	−41.2558	173.3106	504	12,613
K. ochrina	K._ochrina_AK	14.NZ.AK.PNL.04	I27906	−36.7861	175.0209	474	8123
K. paxillulae	K._paxillulae_KA	18.NZ.KA.KAI.3	I27907	−42.4196	173.6876	497	13,472
K. 'peninsularis'	K._'peninsularis'_MC	03.NZ.MC.BPN.02	I6914	−43.8200	172.7750	480	10,524
K. rosea	K._rosea_MC	11.NZ.MC.CHZ.02	I9972	−43.1928	171.7251	506	9279
K. scutellaris	K._scutellaris_WN	14.NZ.WN.NEV.03	I9973	−41.3020	174.8292	499	8236
K. subalpina	K._subalpina_WN	01.NZ.WN.RIM.01	I6916	−41.1146	175.2321	502	13,971
K. 'tasmani'	K._tasmani_NN	02.NZ.NN.COR.21	I6918	−41.1071	172.6921	509	13,976
K. 'tuta'	K._'tuta'_MB	05.NZ.MB.TFL.03	I27908	−41.3027	173.5767	493	12,574

TABLE 1 (Continued)

Species	Name in tree	Simon lab code	Lemmon code	Latitude	Longitude	AHE loci recovered	mtDNA recovered (bp)
K. 'tuta'	K._'tuta'_NN_1	08.NZ.NN.BES.17	I9904	−41.2845	173.1530	511	0
K. 'tuta'	K._'tuta'_NN_2	02.NZ.NN.TTA.35	I27909	−40.5495	172.7216	490	10,595
K. 'westlandica north'	K._'westlandica_north'_BR	12.NZ.BR.MMK.17	I6911	−41.7995	172.3140	502	13,244
K. 'westlandica north'	K._'westlandica_north'_NN	11.NZ.NN.KOB.10	I6909	−41.1108	172.1026	501	5289
K. 'westlandica south'	K._'westlandica_south'_WD	11.NZ.WD.OKT.33	I6849	−43.2258	170.1565	496	6365
M. campbelli	M._campbelli_MB	04.NZ.MB.ISF.03	I16243	−42.0994	173.1440	475	13,763
M. campbelli	M._campbelli_OL	02.NZ.OL.FRL.04	I16251	−45.3975	168.5936	499	13,816
M. campbelli	M._campbelli_SC	06.NZ.SC.BOU.05	I27910	−43.7846	171.1849	497	10,212
M. campbelli	M._campbelli_TO	01.NZ.TO.MAN.04	I16240	−39.1463	175.5808	483	12,153
M. 'cassiomelans'	M._'cassiomelans'_MC_1	12.NZ.MC.HUT.03	I16259	−43.4957	171.5363	495	13,730
M. 'cassiomelans'	M._'cassiomelans'_MC_2	12.NZ.MC.HUT.01	I27911	−43.4957	171.5363	482	13,071
M. cassiope	M._cassiope_NN	14.NZ.NN.STE.04	I16252	−40.7940	172.4601	496	13,948
M. cassiope	M._cassiope_TO	01.NZ.TO.BRR.01	I27912	−39.2336	175.5447	491	10,609
M. clamitans	M._clamitans_CO	02.NZ.CO.AWT.14	I27914	−44.7817	170.3198	495	13,196
M. clamitans	M._clamitans_MK	03.NZ.MK.STH.02	I16249	−44.4613	170.2861	483	12,248
M. clamitans	M._clamitans_SC	03.NZ.SC.HPS.03	I27913	−44.3338	170.5871	496	11,572
M. 'false alticola'	M._'false_alticola'_NN	Rainbow_Skifield_Nelson	I16231	−41.8698	172.8594	474	14,117
M. 'false alticola'	M._'false_alticola'_MB	14.NZ.MB.RST.01	I27929	−41.8698	172.8594	497	10,098
M. hamiltoni	M._hamiltoni_BR	02.NZ.BR.MRV.02	I16257	−42.3803	172.3146	496	13,045
M. hamiltoni	M._hamiltoni_WA	17.NZ.WA.THR.01	I27915	−41.0808	175.3656	485	14,122
M. iolanthe	M._iolanthe_TO	02.NZ.TO.TAS.03	I16255	−38.6959	176.1631	484	13,959
M. iolanthe	M._iolanthe_WA	02.NZ.WA.BUL.02	I27916	−41.3233	175.3042	494	9512
M. lindsayi	M._lindsayi_KA	06.NZ.KA.CAM.04	I27917	−41.8646	173.6774	497	9962
M. lindsayi	M._lindsayi_NC	02.NZ.NC.NCH.01	I16254	−42.8065	173.2743	494	14,341
M. 'mangu awakino'	M._'mangu_awakino'_CO	17.NZ.CO.AWC.03	I27918	−44.7824	170.3227	499	13,700
M. mangu mangu	M._mangu_mangu_MC_1	10.NZ.MC.HTL.01	I27919	−43.5423	171.5364	495	13,717
M. mangu mangu	M._mangu_mangu_MC_2	10.NZ.MC.HUT.06	I16258	−43.4957	171.5363	495	13,981
M. mangu multicostata	M._mangu_multicostata_MB_1	04.NZ.MB.ISO.32	I16241	−42.1286	173.0819	481	11,636
M. mangu multicostata	M._mangu_multicostata_MB_2	04.NZ.MB.ISF.02	I27941	−42.0994	173.1440	500	10,357
M. myersi	M._myersi_WN	02.NZ.WN.ORO.03	I27920	−41.4123	174.9042	496	13,278
M. nigra frigida	M._nigra_frigida_CO	02.NZ.CO.REM.02	I16253	−45.0554	168.8141	494	13,104
M. nigra frigida	M._nigra_frigida_OL	02.NZ.OL.TLC.03	I27921	−44.6327	168.8792	495	10,294

(Continues)

TABLE 1 (Continued)

Species	Name in tree	Simon lab code	Lemmon code	Latitude	Longitude	AHE loci recovered	mtDNA recovered (bp)
M. nigra nigra	M._nigra_nigra_NC_1	03.NZ.NC.NIG.01	I16247	−42.9100	171.5798	476	13,556
M. nigra nigra	M._nigra_nigra_NC_2	03.NZ.NC.TBS.01	I27922	−42.9122	171.5690	499	13,834
M. oromelaena	M._oromelaena_MK	02.NZ.MK.LTM.30	I16250	−43.6601	170.1779	491	9741
M. oromelaena	M._oromelaena_NC	02.NZ.NC.TBS.01	I27923	−42.9122	171.5690	498	12,232
M. otagoensis maceweni	M._otagoensis_maceweni_SL_1	XX.NZ.SL.TAK.01	I27925	−45.7000	167.9000	498	12,264
M. otagoensis maceweni	M._otagoensis_maceweni_SL_2	05.NZ.SL.CLA.07	I27924	−45.6203	167.9522	500	13,588
M. otagoensis otagoensis	M._otagoensis_otagoensis_CO_1	02.NZ.CO.RSR.04	I16246	−45.0401	168.8015	484	11,728
M. otagoensis otagoensis	M._otagoensis_otagoensis_CO_2	05.NZ.CO.WBG.02	I27926	−45.4469	169.2499	498	13,905
M. phaeoptera	M._phaeoptera_CO	02.NZ.CO.AWT.13	I27927	−44.7817	170.3198	495	12,615
M. phaeoptera	M._phaeoptera_MK	03.NZ.MK.RHS.02	I16245	−43.8240	170.6595	491	12,967
M. tenuis	M._tenuis_BR	03.NZ.BR.MMU.01	I27928	−41.7309	172.4989	493	13,578
M. tenuis	M._tenuis_MB	02.NZ.MB.PAT.03	I16248	−41.5895	173.2988	486	13,049
R. cruentata	R._cruentata_GB	03.NZ.GB.MAR.01	I16244	−38.8382	177.8949	495	11,732
R. leptomera	R._leptomera_HB	06.NZ.HB.POR.12	I16260	−40.3081	176.6659	499	14,068
R. microdora	R._microdora_MC	02.NZ.MC.LCR.02	I16256	−43.6427	172.4781	497	14,169

Note: Two-letter area code following the names in the second column follows Crosby et al. (1998). Names in quotes are informal names of currently undescribed species.
Abbreviation: AHE, anchored hybrid enrichment.

opportunistically as weather permitted (Table 1). This was part of an effort to collect representatives of all NZ cicada species throughout their ranges to study the complete history of all five NZ cicada genera, including the two genera not included in this study that descend from a separate colonization event. Specimens were collected by hand or with insect nets and stored in 95% ethanol. Ethanol specimens were refrigerated in the field and kept frozen between −20 and −80°C in the Simon lab at the University of Connecticut. The specimens selected for sequencing in this study represent all known described and undescribed *Kikihia* species and subspecies, all known *Rhodopsalta* species, and all known described and undescribed *Maoricicada* species excepting *M. mangu celer* Dugdale & Fleming, *M. m. gourlayi* Dugdale & Fleming, and *M. alticola* Dugdale & Fleming. The two specimens identified as *M.* 'false alticola' were initially identified as *M. alticola* on collection, but on subsequent examination were deemed unlikely to be *M. alticola* as originally described. Cicada genomic DNA was extracted from leg muscle tissue using the Qiagen DNeasy Tissue kit (Qiagen, Valencia, California, USA) following the manufacturer's instructions. We prepared Illumina libraries from nuclear genomic DNA following Prum et al. (2015), which were then sequenced using AHE at the Center for Anchored Phylogenomics, Florida State University (Lemmon et al., 2012). In brief, we sonicated DNA to a fragment size of 125–325 bp using a Covaris ultrasonicator. We then added universal Illumina adaptors with 8 bp indexes and pooled the libraries for sequencing. We used the probe set targeting Auchenorryncha based on the set developed by Dietrich et al. (2017) described in Simon et al. (2019). Rather than use the data processing pipeline developed by the Center for Anchored Phylogenomics (Granados Mendoza et al., 2020), we chose to develop our own custom pipeline for better control over the loci flanking regions and finer control over paralog detection and recovery. Specimen vouchers are currently deposited in the University of Connecticut Biodiversity Research Collection.

Data processing of AHE loci

We chose to divide our sequence data into a *Kikihia* dataset and a *Maoricicada* + *Rhodopsalta* dataset for our data processing steps and subsequent phylogenetic analyses. We chose to separate the datasets in this way because we found in our initial investigations that around 5% of the targeted loci were recovered only in the *Kikihia* specimens, and because we expected different amounts of hybridization in *Kikihia* and *Maoricicada* based on the results of previous studies.

The raw Illumina sequence data were assembled using the following steps. First, we assessed the quality of the AHE raw reads using FastQC (Andrews, 2010). Next, we trimmed the Illumina adaptors using Trimmomatic (Bolger et al., 2014 and merged the paired reads using BBMerge (Bushnell et al., 2017). Finally, we used SPAdes v.3.12.0 to assemble the AHE-trimmed merged reads for each cicada sample (Bankevich et al., 2012).

We built our nuclear AHE datasets from the cicada assemblies with the following pipeline (pipeline scripts available here https://github.com/markstukel/Simon-target-capture-pipeline). First, we searched for the targeted loci in each cicada assembly with BLAST using the cicada assembly as the database, the loci probe sequences as queries and an e-value cutoff of 1e-10 (Camacho et al., 2009). We compiled the results from each cicada assembly by locus, used ALiBaSeq (Knyshov et al., 2021) to extract all sequences without flanking regions above our e-value cutoff from each assembly for each locus and aligned the sequences using the MAFFT v.7 E-INS-I algorithm (Katoh & Standley, 2013). We used CDhit v.4.6.8 to cluster the sequences for each locus alignment using a 97% cluster threshold, which removed redundant sequences and closed in-paralogs (Fu et al., 2012).

Since there was still the possibility that some of the sequences in our locus alignments were paralogs of the targeted loci, we used UPhO to further cluster the sequences in each locus alignment into putative ortholog groups (Ballesteros & Hormiga, 2016). If UPhO gave multiple orthogroups for a locus (i.e., paralogs were detected in the original locus alignment), our next step depended on the size of each orthogroup. If an orthogroup contained fewer than 50% of the taxa in the dataset, all the sequences in that orthogroup were discarded from the alignment and the BLAST result files, whereas orthogroups containing at least 50% of the taxa were retained. If a locus had more than one orthogroup retained, the locus alignment was split into new renamed alignments based on the retained orthogroups, and the original BLAST result file was modified according to the new alignment names. This process of clustering and splitting loci based on orthogroups, therefore, allowed us to not only confidently filter out paralogs from our targeted loci but also to increase the number of loci in our datasets through the process of splitting loci with multiple retained orthogroups. To confirm that no paralogs were missed by the UPhO step, we BLASTed the new alignments against reference genomes from the related taxa *Nilaparvata lugens* (Stål) and *Laodelphax striatellus* (Fallén). We chose these genomes as no Cicadamorpha reference genomes were available and they were the closest Auchenorrhyncha genomes available at the time of data processing. We confirmed that all sequences from each locus were mapped to the same genomic location using BEDtools (Quinlan & Hall, 2010). Any sequences that did not map to the same location were determined to be paralogs and removed from the assembly BLAST result files modified earlier. We then used ALiBaSeq a second time to extract a new set of alignments with 300 bp flanking regions using the modified BLAST result files that were cleaned of paralogs over the previous steps. After checking the new, paralog-split loci alignments for overlaps in the flanking regions and manually merging any loci that overlapped, we used HMMCleaner to remove regions that were misaligned or misassembled (Franco et al., 2019). Finally, we used a custom script to trim the ends of the alignments until 75% of the taxa in the alignment were present. These trimmed alignments were used for the nuclear phylogenetic analyses.

Assembly of mitochondrial genomes from AHE bycatch

To assemble mitochondrial genomes, we used off-target AHE capture data from the SPAdes assemblies above (Haji et al., 2022;

Lemmon et al., 2012; Simon et al., 2019). We identified mitochondrial contigs in the assemblies with a partial *Kosemia yezoensis* (Matsumura) [Cicadamorpha: Cicadidae: Cicadettinae: Cicadettini] reference mitochondrial genome—Genbank MG737723.1 (Łukasik et al., 2019). We used BWA v0.7.5a (Li & Durbin, 2009) for a second processing step and reassembled with SPAdes. We used MAFFT to align the contigs for each cicada assembly with the reference genome and used Geneious v10.1.3 to merge contigs into a single genome sequence and manually removed any misassembled regions (Kearse et al., 2012). We then used MITObim v1.9.1 to extend the contigs with unassembled sequence reads (Hahn et al., 2013) and again aligned with MAFFT before manual editing in Geneious.

Nuclear phylogenomic analyses

Phylogenetic analysis scripts are available on Dryad: https://doi.org/doi:10.5061/dryad.t1g1jwt7v. We constructed a concatenated maximum-likelihood phylogeny for the nuclear AHE data using IQ-Tree (Minh, Schmidt, et al., 2020). The data were partitioned by gene, and partitions were merged using the -m TESTMERGE setting in ModelFinder, which re-implements the PartitionFinder algorithm (Chernomor et al., 2016; Kalyaanamoorthy et al., 2017; Lanfear et al., 2017). As the intron-exon boundaries in the AHE loci were not known, we did not partition the loci by codon position. We ran IQ-Tree with 1000 ultrafast bootstrap (UFB) replicates and 1000 Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-aLRT) replicates (Hoang et al., 2018).

To account for ILS, we used two different multispecies coalescent methods. The first method we used was a gene tree approach in ASTRAL-III (Zhang et al., 2018). We inferred gene trees for all loci using IQ-Tree with model selection using ModelFinder and with 1000 SH-aLRT replicates. To reduce error and increase branch support values in the species tree due to arbitrarily resolved branches in the gene trees, we collapsed gene tree branches with very low support as recommended by the ASTRAL-III manual. We chose 0% SH-aLRT support as our threshold to collapse branches instead of an arbitrary bootstrap threshold (e.g., 10% bootstrap support) because 0% SH-aLRT support indicates that the branch has no difference in likelihood from an alternatively resolved branch (Anisimova & Gascuel, 2006; Simmons & Gatesy, 2021). After collapsing low-supported branches using this threshold, we inputted the gene trees into ASTRAL-III with default settings. We also constructed gene trees in RAxML under GTR-G model for all loci, collapsed branches with the same support threshold as above, and used them as input of ASTRAL to compare the two gene-tree inference methods (Stamatakis, 2014).

The second coalescent approach we used was SVDQuartets as implemented in PAUP* (Chifman & Kubatko, 2014; Wilgenbusch & Swofford, 2003). As SVDQuartets infers a species tree using individual site patterns from the alignment instead of input gene trees, it is not affected by potential gene tree inference error or choice of threshold for collapsing low-supported gene tree branches. However, SVDQuartets was originally developed for single nucleotide polymorphism (SNP) data and as a result assumes sites are independent, which does not strictly hold for gene loci sequences. The advantages and shortcomings of ASTRAL and SVDQuartets each allow us to compare species tree inference between the two methods. We used the concatenated alignment partitioned by gene as input and ran the analysis with 100 bootstrap replicates under the multilocus bootstrap setting.

Because phylogenetic datasets are known to have dramatically inflated branch supports under traditional bootstrap methods, we used gene and site concordance factors (gCF and sCF) as an additional tool to evaluate branch support (Minh et al., 2020). Using the gene trees generated using IQ-Tree above and the concatenated alignment, we calculated the gCF and sCF values for the concatenated, ASTRAL and SVDQuartets tree topologies in IQ-Tree. Finally, we calculated internode certainty (IC) values for the concatenated, ASTRAL and SVDQuartets trees using the program QuartetScores (Zhou et al., 2020). This program uses a quartet-based method for computing IC values instead of a bipartition-based method, making it better-suited for datasets containing partial gene trees like our own. QuartetScores returns three quartet-based IC measures: the lowest quartet IC, the quadripartition IC and the extended quadripartion IC.

Mitochondrial phylogenomic analyses

We conducted a concatenated maximum likelihood analysis for the mitochondrial genomes obtained from the AHE bycatch. We used the following partitioning scheme: combined first and second codon position of each protein-coding gene, third position of each protein coding gene, 12S rRNA, 16S rRNA and a single partition for all tRNAs. We used RAxML on the CIPRES web server under the GTRCAT model with 1000 bootstrap replicates as well as IQ-Tree with 1000 UFB replicates. Due to low branch support values in the concatenated mitochondrial genome trees, we performed follow-up analyses to rule out the presence of conflicting information among mitochondrial genes due to chimerism in the bycatch assemblies. We generated mitochondrial gene trees for each protein-coding gene and rRNA and for all tRNAs combined using IQ-Tree for comparison with the concatenated maximum likelihood tree, partitioning the protein-coding genes by codon. We also constructed a coalescent tree of the mitochondrial genomes using SVDQuartets as implemented in PAUP* partitioned by gene with 100 multilocus bootstraps for comparison with the concatenated maximum likelihood tree.

Nuclear and mitochondrial combined phylogenomic analysis

To assess the effects of combining the nuclear AHE loci with the mitochondrial genome sequences on the phylogeny and for better comparison to previously published NZ cicada phylogenies, we concatenated the nuclear and mitochondrial data for both the *Kikihia* and *Maoricicada* + *Rhodopsalta* datasets. We performed maximum likelihood analyses in IQ-Tree using the same procedure described above for the nuclear-only data.

Reanalysis of Maoricicada nuclear genes

For a more accurate comparison to our nuclear phylogenomic analyses, we reanalyzed the *Maoricicada* Sanger-sequence data from Buckley et al. (2006) to exclude the mtDNA sequence and include only the nuclear genes. We downloaded the period, calmodulin and EF-1 alpha sequences from all Buckley et al. (2006) specimens from GenBank, aligned them in Geneious and performed concatenated maximum likelihood phylogenetic inference in IQ-Tree. We used the -m TESTMERGE setting in ModelFinder to determine the appropriate substitution model and merge partitions, and we assessed the support using 200 non-parametric bootstrap replicates.

RESULTS

AHE pipeline

The custom AHE pipeline recovered 516 AHE loci for the *Kikihia* dataset and 505 loci for the *Maoricicada* + *Rhodopsalta* dataset, out of 583 locus probes. Thirty of the loci in the *Kikihia* and 32 of the loci in the *Maoricicada* dataset were paralogs of the targeted loci recovered by the pipeline. The number of loci recovered from each specimen ranged from 474 to 511 (mean 497) for the *Kikihia* dataset and 474 to 500 (mean 492) for the *Maoricicada* + *Rhodopsalta* dataset (Table 1). The read quality as assessed by FastQC and the number of recovered AHE loci were not substantially different between the oldest and youngest collected specimens. The final trimmed AHE loci in the *Kikihia* dataset ranged in length from 297 to 2228 bp with a total alignment length of 324,861 bp, whereas those from the *Maoricicada* + *Rhodopsalta* dataset ranged from 219 to 2345 bp with a total alignment length of 357,379 bp. We recovered partial mitochondrial genomes from 37 of 38 *Kikihia* taxa with a total alignment length of 14,393 bp, and from all 43 *Maoricicada* and *Rhodopsalta* taxa with a total alignment length of 14,631 bp. The mitochondrial genomes ranged from 0.6% to 64% missing data for *Kikihia* (mean 25%) and 1.7% to 35% (mean 13%) for *Maoricicada* and *Rhodopsalta* (Table 1). The mitochondrial genomes are deposited in GenBank under the accession numbers OR413940-OR413978 and OR459868-OR459906. The mitochondrial genome recovered for *K. laneorum* Fleming (Simon lab code 14.NZ.BP.KMS.01) was too incomplete to be deposited in GenBank but is included on Dryad. Raw sequence reads are available on SRA (BioProject PRJNA1015981, BioSample accessions SAMN37366775-SAMN37366854).

Phylogenetic analyses

Our phylogenetic analyses generally improve the support for and resolution of species relationships compared with previous Sanger-based studies, as we expected from having more extensive sequence data. However, there were areas of the phylogeny that remain poorly resolved, even with the increase of data. Furthermore, there appears to be considerable gene and site conflict in our data for many branches in the phylogeny. We explore the results for each dataset in detail below.

Phylogenetic analyses—*Kikihia*

The three nuclear phylogenetic analyses (concatenated ML, ASTRAL and SVDQuartets) largely agreed with one another in topology (Figures 2, S1 and S2). All three nuclear analyses had very strong bootstrap, posterior probability (PP), UFB or SH-aLRT support for most relationships, with only a few branches in each tree having weak support. We found support for the monophyly of three of the four species groups of the main radiation identified by Marshall et al. (2008): the Rosea group, the Cutora group and the Muta group. All three nuclear analyses found the mitochondrially defined Westlandica group to be polyphyletic, with the two *K.* 'westlandica' species grouping with the Muta group. The main differences among the three nuclear analyses were in the placement of *K. angusta* (Walker), *K.* 'murihikua' and *K.* 'tasmani'. Unlike the three nuclear trees, the mitochondrial genome tree found all four species groups to be monophyletic, with the Westlandica group sister to the other three groups, and the Muta group sister to the Rosea group and the Cutora + *K. horologium* Fleming group (Figure 3).

Despite having strong support in traditional branch support metrics such as bootstrap percentage or PP, the three nuclear analyses had much lower gCF and sCF support. Across all three nuclear trees, most branches had gCF values under 10%, indicating that very few individual gene trees contained the inferred branch relationships (Supporting Information S1). The sCF values in all three nuclear trees, while higher than the gCF values, were similarly low, with most branches having values under 50% (Supporting Information S1). Given that sCF values of 33% indicate equal support for the three possible quartet arrangements around a branch, this indicates that most branches in the nuclear trees were only slightly favoured by the nucleotide site patterns over the alternative arrangements. Similarly, the ASTRAL tree had several branches with equal or nearly equal quartet frequency support (Figure 2 and Supporting Information S1). The three quartet-based IC measures we obtained for each nuclear tree were very close to zero for all but a few branches. No branch in any of the nuclear trees had IC values near 100%, and all negative values were only weakly negative (Figures 2, S1 and S2).

The combined nuclear AHE loci and mitochondrial genome tree were very similar in topology to the nuclear-only concatenated tree, with only a few small differences (Figure S6). The UFB and SH-aLRT branch supports for the combined nuclear + mitochondrial tree were also extremely similar to those of the nuclear-only concatenated tree.

Phylogenetic analyses—Maoricicada + Rhodopsalta

As with the *Kikihia* dataset, the three nuclear phylogenetic analyses largely agree with one another in topology (Figures 4, S4 and S5). *Maoricicada hamiltoni* (Myers) is recovered as sister to all other

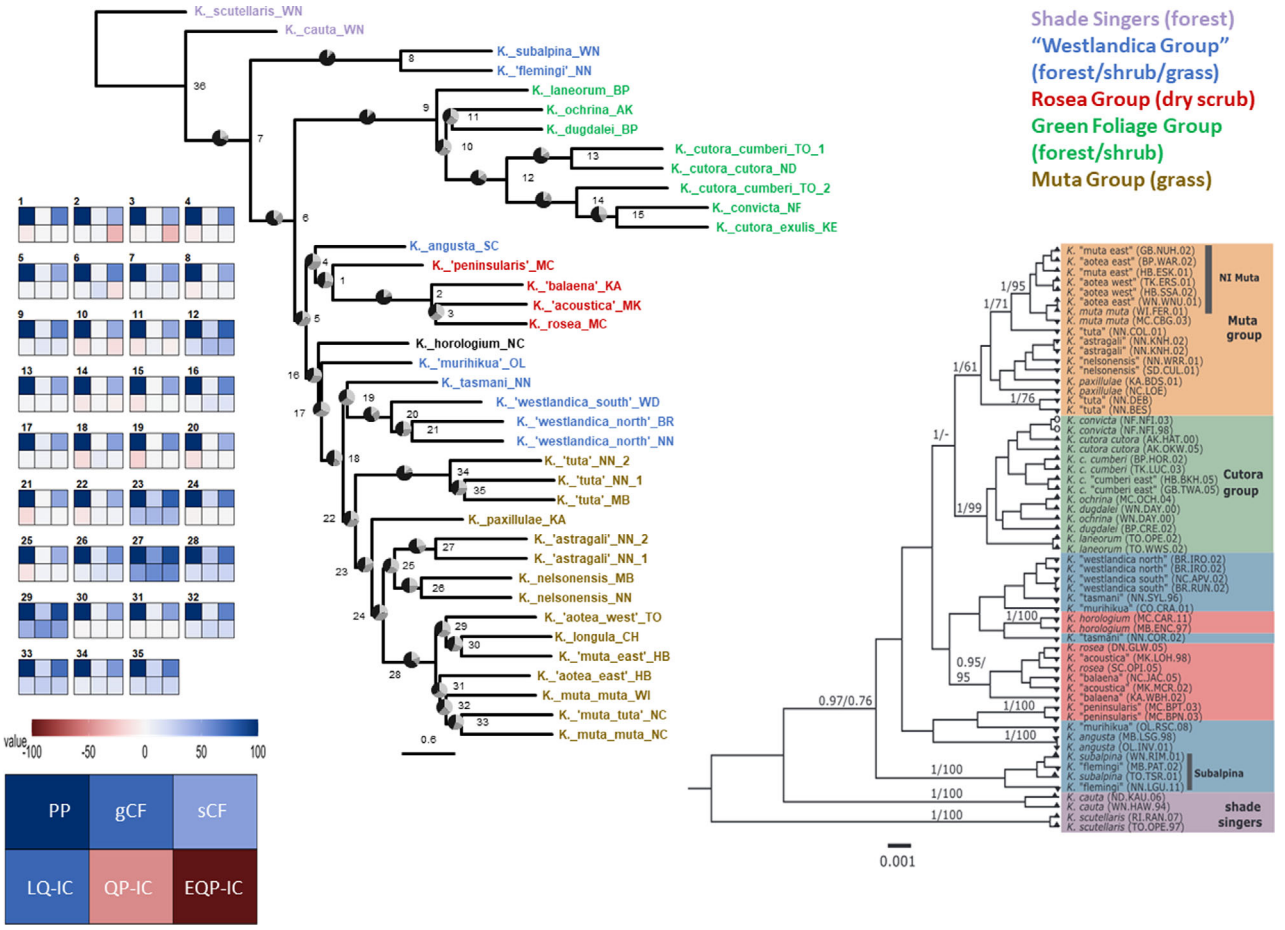


FIGURE 2 Left: *Kikihia* ASTRAL nuclear phylogeny built from 516 anchored hybrid enrichment (AHE) nuclear gene trees. Pie charts on branches display ASTRAL quartet frequencies (highest quartet frequency and the two alternative frequencies). Heatmaps for corresponding node numbers display local posterior probability (PP), gene concordance factor (gCF), site concordance factor (sCF), lowest quartet internode certainty (LQ-IC), quadripartition internode certainty (QP-IC) and extended quadripartition internode certainty (EQP-IC). Branch lengths are in coalescent units, except for terminal branches which are unscaled. Inset right: 5-gene nuclear phylogeny reproduced from Banker et al. (2017) for comparison. Shapes on branch tips refer to whether the specimen is from the North Island (upwards triangle), South Island (downwards triangle), or offshore island (circle). Refer to figure caption in the original paper for details.

Maoricicada, and the other lowland *Maoricicada* species (*M. myersi* (Fleming), *M. lindseyi* (Myers), *M. iolanthe* (Hudson), and *M. campbelli* (Myers)) are recovered as sister to the alpine radiation. The alpine radiation itself is largely well-resolved (high bootstrap support), with the notable exception of the base of the radiation: *M. cassiope* (Hudson) + *M. tenuis* Dugdale & Fleming, *M. mangu* (White), and the rest of the alpine species together form a polytomy. The specimens identified as *M.* 'cassiomelans' were suspected to be possible hybrids of *M. cassiope*, *M. oromelaena* (Myers) and *M. clamitans* Dugdale & Fleming when initially collected, but based on their consistent placement among the *M. oromelaena* specimens in all analyses, we treat them as *M. oromelaena* elsewhere in the text. The only difference between the concatenated tree and the ASTRAL tree is the placement of *M.* 'false alticola', and the only difference between the ASTRAL and SVDQuartets trees is the placement of *M. iolanthe*.

Like the *Kikihia* results, the three *Maoricicada* nuclear analyses had low gCF and sCF support (Supporting Information S1). While a smaller proportion of branches had very low support values, the branches with high gCF and sCF tended to be branches uniting individuals from the same species. For branches showing relationships among species, across all three nuclear trees, most branches again had gCF values under 10% and sCF values under 50% (Supporting Information S1). This indicates that for relationships among species, there was again very little support in the gene trees or the site patterns. Similar to the *Kikihia* analysis, the *Maoricicada* ASTRAL tree had several branches with equal or nearly equal quartet frequency support (Figure 5 and Supporting Information S1). The three quartet-based IC measures we obtained for each nuclear tree were very close to zero for all but a few branches among species (Figure 4 and Supporting Information S1).

The mitochondrial genome tree is less well-resolved than the nuclear trees and features some notable conflicts with the nuclear tree (Figure 5). First, *M. hamiltoni* is now sister to the lowland species *M. myersi* and *M. lindsayi* (Myers) instead of the rest of the genus as a whole. Second, the *M. campbelli* + *M. iolanthe* clade is now nested within the alpine radiation instead of grouping with the other lowland

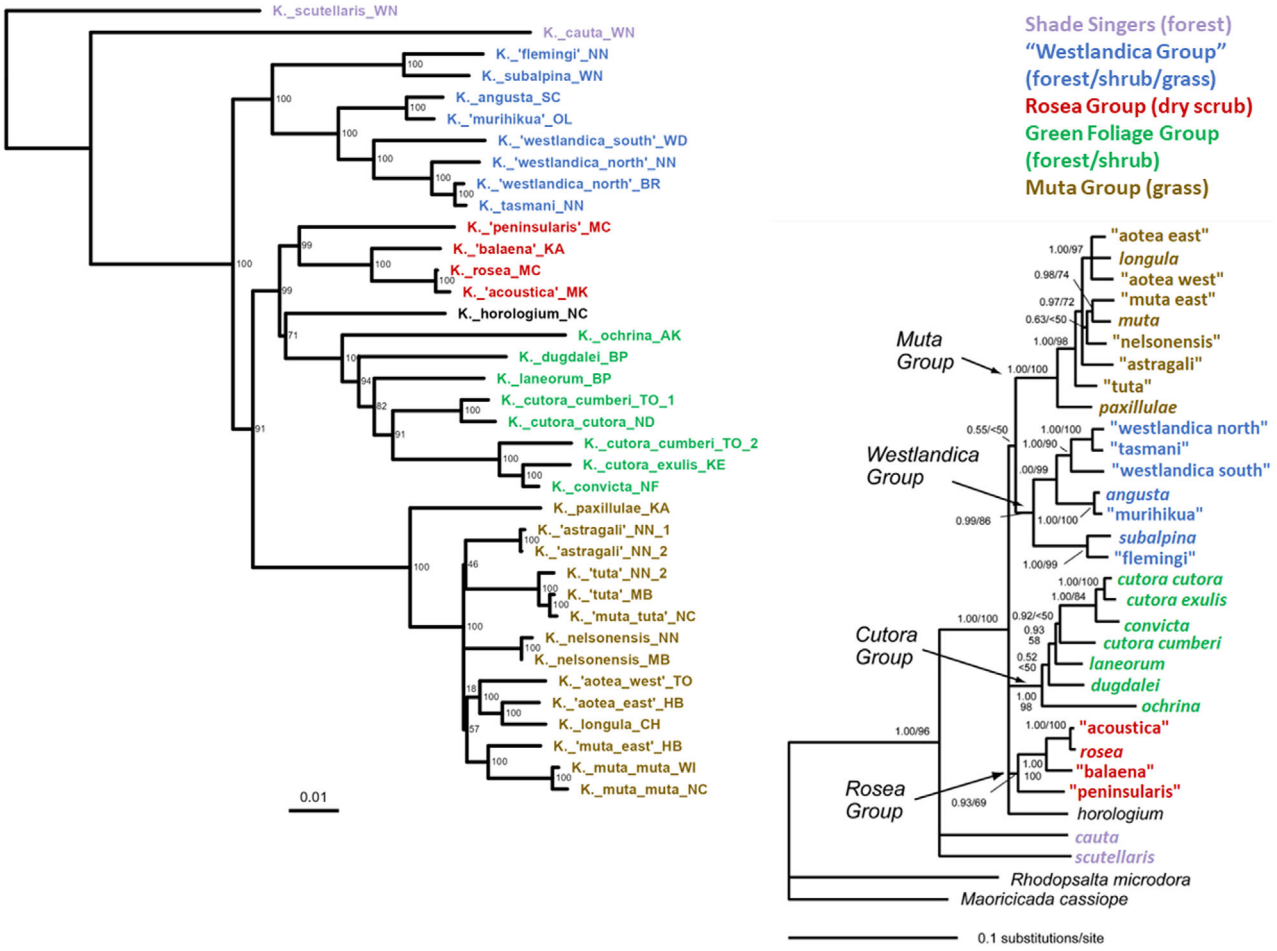


FIGURE 3 Left: *Kikihia* mitochondrial genome phylogeny built from anchored hybrid enrichment (AHE) bycatch. Branch lengths are in substitutions/site. Node supports are ultrafast bootstrap support (UFB). Inset right: mitochondrial phylogeny based on 2152 bp of mtDNA sequence modified from Marshall et al. (2008) for comparison.

species. Finally, a specimen of *M. mangu* collected from the Awakino Ski Field no longer groups with the other *M. mangu* specimens, instead connecting to a polytomy elsewhere in the alpine radiation.

Despite the differences in topology between the nuclear trees and the mitochondrial genome tree, when we combined the nuclear AHE loci and mitochondrial genomes into a single concatenated alignment, the resulting maximum likelihood tree had an identical topology to the nuclear-only concatenated tree (Figure S6). The UFB and SH-aLRT branch supports were also nearly identical between the nuclear-only and nuclear + mitochondrial concatenated trees.

Reanalysis of Maoricicada nuclear genes

The total alignment length for the three nuclear genes from Buckley et al. (2006) was 3705 bp. Partitionfinder merged all three nuclear genes into a single partition under an HKY + G model. The maximum likelihood tree has a quite different topology from that of the *Maoricicada* nuclear phylogenomic trees and mitochondrial genome tree (Figure 4 inset). However, this is less surprising in light of the fact that the three-gene tree has low bootstrap support for most of the contentious branches.

DISCUSSION

Improving phylogenetic resolution using genomic data

The phylogenetic trees inferred in this study are overall well-resolved based on traditional measures of branch support such as UFB, SH-aLRT, ASTRAL PP and ML bootstraps. This marks a dramatic improvement compared with previous Sanger-based trees due to the vast increase in phylogenetic information in genomic data. Our *Kikihia* phylogenomic dataset resolves five deep-level polytomies and over 10 shallow-level polytomies compared with the most recently published Sanger-based tree (Banker et al., 2017), whereas our *Maoricicada* phylogenomic dataset resolves at least four polytomies spread throughout the tree compared with the previous Sanger tree (Buckley et al., 2006). The older *Maoricicada* tree had fewer polytomies to resolve but featured more contentious relationships compared with the phylogenomic trees inferred in this study.

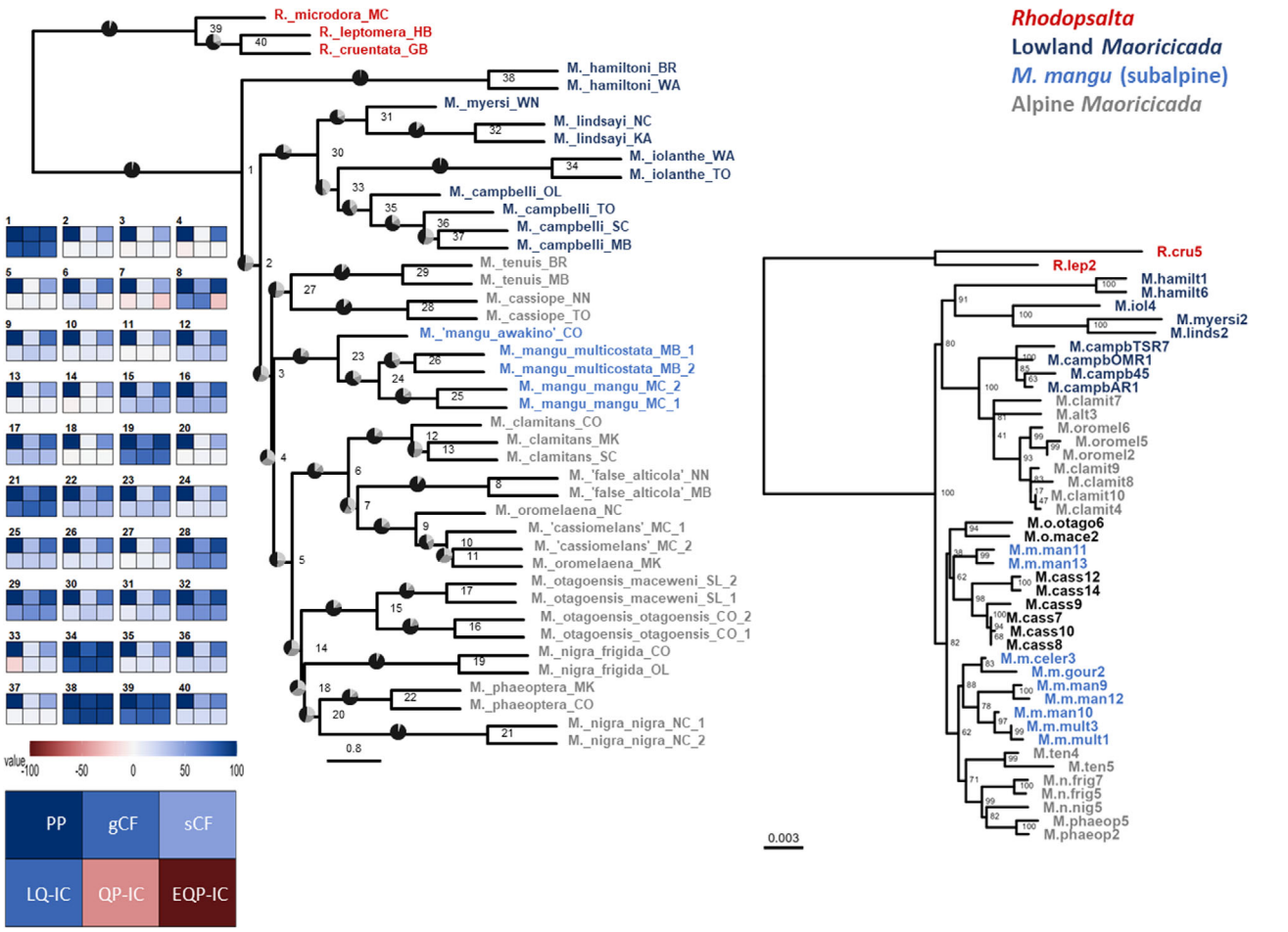


FIGURE 4 Left: *Maoricicada* ASTRAL nuclear phylogeny built from 505 anchored hybrid enrichment (AHE) nuclear gene trees. Pie charts on branches display ASTRAL quartet frequencies (highest quartet frequency and the two alternative frequencies). Heatmaps for corresponding node numbers display local posterior probability (PP), gene concordance factor (gCF), site concordance factor (sCF), lowest quartet internode certainty (LQ-IC), quadripartition internode certainty (QP-IC), and extended quadripartition internode certainty (EQP-IC). Branch lengths are in coalescent units, except for terminal branches which are unscaled. Inset right: 3-gene nuclear phylogeny using sequence data from Buckley et al. (2006) for comparison.

The differences in patterns of evolution between *Maoricicada* and *Kikihia* appear to play a role in the performance of phylogenomic versus Sanger data in resolving species relationships. The genus *Kikihia* has more species and is characterized by an initial radiation of several well-defined species groups conforming to ecological habitats that then themselves diversified over time. The genus *Maoricicada* in contrast has fewer species and is characterized by a small clade of lowland species that is sister to a single rapid radiation of alpine taxa; sister to both of those clades is a single lowland species found on both the North and South Island. As expected, the genus *Maoricicada*, without well-defined species groups within the radiation, had more drastic changes in the placement of taxa with the increase in data compared with the genus *Kikihia* that possesses well-defined species groups. This suggests that when Sanger-sequence data have low information for resolving a species radiation, the presence of well-defined species groups within the radiation constrains the placement of species by breaking up the radiation into smaller sections. In other words, as previous authors have noted, some trees are easier than others. Clades with fewer bursts of evolution, and well-spaced cladogenetic events are easier to resolve (Whitfield & Lockhart, 2007), and in this paper, we present examples from shallower levels of resolution.

Although our phylogenomic data dramatically improved the resolution of these NZ cicada species radiations, they do reveal some limitations. Despite the alignment length and the large number of genes sequenced, for both genera, some relationships have low support. This indicates that, for some species radiations, genomic data from conserved targeted loci are not enough to resolve some relationships. In addition, even within 'highly supported' nodes, there are varying levels of support and conflict within the data. One way this is apparent is through comparing concatenated maximum likelihood analyses to coalescent-based species tree analyses such as ASTRAL and SVDQuartets, which reveal small differences in topology. Even within nodes that are supported by all analysis types, newer phylogenomic measures of support and conflict such as concordance factors and IC reveal large amounts of hidden conflict within our datasets.

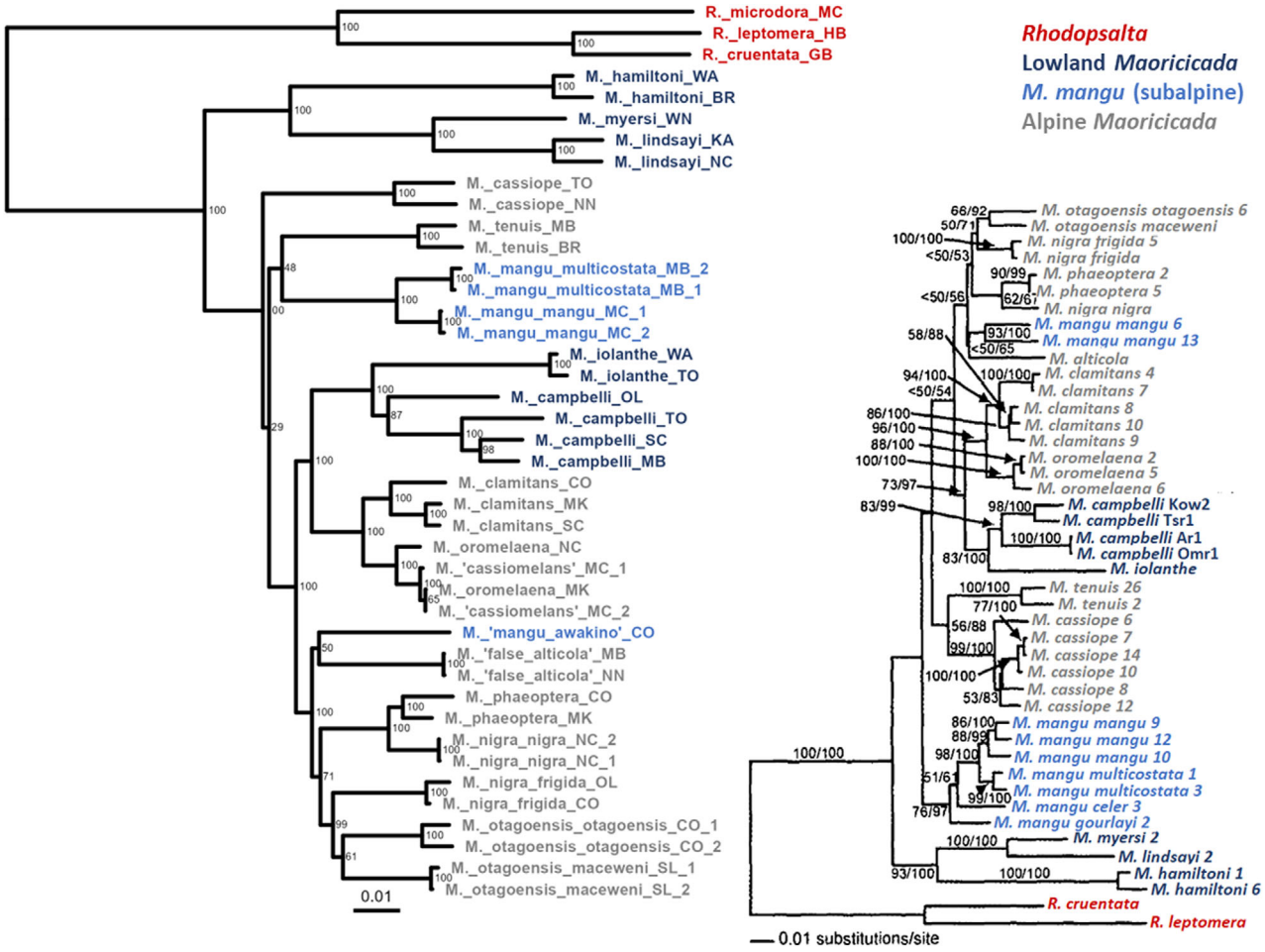


FIGURE 5 Left: *Maoricicada* mitochondrial genome phylogeny built from anchored hybrid enrichment (AHE) bycatch. Branch lengths are in substitutions/site. Node supports are ultrafast bootstrap support (UFB). Inset right: mitochondrial phylogeny based on 2274 bp of mtDNA sequence modified from Buckley et al. (2006) for comparison.

Improving phylogenetic resolution—*Kikihia*

The best-resolved *Kikihia* nuclear phylogeny previously published is the one inferred by Banker et al. (2017) using five nuclear gene segments. Regardless of which phylogenetic analyses were used, Banker et al. (2017) found that many relationships were unresolved. The phylogenomic results presented here agree with the Banker et al. (2017) phylogeny in the monophyly of the Muta and Cutora groups and the non-monophyly of the artifactual 'Westlandica group'. The Westlandica group is polyphyletic and artifactual rather than reflecting the true species tree because it results from past mitochondrial capture among hybridizing species (Marshall et al., 2008, 2011). Unsurprisingly, the phylogenomic tree disagrees with many of the poorly supported relationships in the Banker et al. (2017) phylogeny. Instead of being weakly supported as the sister to the Muta group, the Cutora group is now strongly supported as the sister group to a clade containing the Rosea and Muta groups. Furthermore, the *K.* 'westlandica north' and *K.* 'westlandica south' species are now found to be sister to the Muta group with strong support instead of grouping with the Rosea group with weak support. This placement is closer to Dugdale and Fleming's intuitive *Kikihia* phylogeny based on song, ecology and morphology (redrawn in Banker et al., 2017), because they considered *K.* 'westlandica north' and *K.* 'westlandica south' to be part of the species *K. muta* (Fabricius). In addition, the phylogenomic results presented here strongly recover *K. subalpina* (Hudson) and *K.* 'flemingi' as a clade sister to all *Kikihia* except the 'shade singers' *K. cauta* (Myers) and *K. scutellaris* (Walker), a relationship that is also recovered by Banker et al. (2017), although with weak support.

In addition to improving the resolution of the relationships among the main species groups, our phylogenomic results also improve the resolution of species relationships within the species groups. In the five-gene phylogeny from Banker et al. (2017), there is no strong support for species relationships within any main species group. In contrast, the phylogenomic trees show strong support for nearly all relationships within the species groups. Within the Cutora group, the only mitochondrial species group strongly supported by Banker et al.'s (2017) nuclear gene phylogeny, *K. cutora cutora* (Walker), *K. c. cumberi* Fleming, the *K. c. exulis* (Hudson) and *K. convicta* (Distant) all form a strongly supported clade with *K. c. cumberi* strongly supported as polyphyletic, indicating that the definitions of these species and

subspecies need revision. Within the Rosea group, *K.* 'peninsularis' is recovered as a sister to the other members, which is a relationship that Banker et al. (2017) did not recover. However, *K. rosea* (Walker) and the other Rosea group members *K.* 'balaena', and *K.* 'acoustica' form a polytomy. Within the Muta group, the 'North Island Muta' subclade, the only well-resolved subclade identified by Banker et al. (2017), is once again recovered, this time including a sister-group relationship between *K.* 'muta east' and the Chatham Island species *K. longula* (Hudson). The relationships within the 'North Island Muta' are recovered here as in Banker et al., except for *K.* 'aotea-east', which is part of a polytomy that includes *K. muta*, *K.* 'aotea-west' and *K. longula* + 'muta east'.

The mitochondrial genome tree for *Kikihia* also improves the resolution of species relationships compared with the COI + COII-based tree of Marshall et al. (2008). This Sanger-based tree notably recovers the four clades described in detail above (the Muta group, Cutora group, Rosea group and Westlandica group) as a polytomy, with the Muta and Westlandica groups weakly supported as sister to one another. In contrast, the mitochondrial genome tree recovers the Westlandica group sister to the other species groups with strong support. The mitochondrial genome tree does recover the same species relationships within the four clades with very high support except for in a few instances. While we recover the Cutora group species *K. ochrina* (Walker), *K. dugdalei* Fleming and *K. laneorum* branching off in sequence at the base of the Cutora group, the relationships among *K. cutora cutora*, *K. c. cumberi*, *K. c. exulis* and *K. convicta* are slightly different from the Sanger-based tree of Marshall et al. (2008). Marshall et al. recovered *K. cutora cutora*, and *K. c. exulis* as a sister group to *K. convicta*, with *K. cutora cumberi* as sister to the previous three. In contrast, our mitochondrial genome tree recovers the Norfolk Island species *K. convicta* and the Kermadec Islands species *K. cutora exulis* as a sister group and *K. cutora cumberi* as polyphyletic, which are the relationships recovered in the nuclear phylogenomic trees. This suggests that Norfolk Island was colonized from the Kermadec Islands or vice versa, rather than two independent colonizations of these remote northern islands from North Island, NZ.

Improving phylogenetic resolution—Maoricicada + Rhodopsalta

The best-resolved *Maoricicada* phylogeny previously published is the phylogeny published by Buckley et al. (2006), based on three nuclear genes (period, calmodulin, and EF-1 alpha) and a segment of mtDNA. The species relationships in this Sanger-based phylogeny are overall poorly resolved, especially within the alpine radiation. For the current study, we combined a large, nearly complete mitochondrial genome dataset with our nuclear AHE data for a concatenated maximum likelihood analysis. The topology was identical to our AHE nuclear-only concatenated tree. We used this combined nuclear and mitochondrial phylogenomic tree for comparison to the Buckley et al. (2006) combined tree. The Buckley et al. (2006) combined phylogeny recovers the lowland species *M. hamiltoni* as sister to other lowland species *M. myersi* and *M. lindsayi* with moderate support, whereas our combined phylogenomic tree recovers *M. hamiltoni* as sister to all other *Maoricicada*. The position of *M. hamiltoni* in the Buckley et al. (2006) tree instead agrees with the *Maoricicada* mitochondrial genome tree presented in our study; this is not surprising since the Buckley et al. (2006) combined dataset informative sites were dominated by mitochondrial data. Our combined phylogenomic tree recovers the lowland species *M. campbelli* and *M. iolanthe* in a group with their fellow lowland species *M. myersi* and *M. lindsayi*. The Buckley et al. (2006) combined phylogeny places them in a large polytomy containing the alpine species. This placement is also congruent with our mitochondrial genome tree, which places *M. campbelli* and *M. iolanthe* within the alpine radiation sister to *M. clamitans* and *M. oromelaena*. The Buckley et al. (2006) combined tree does not group the *M. mangu* specimens from Awakino Ski Field and Hakataramea Pass with the other *M. mangu* specimens; it instead shows them branching within the large alpine polytomy. Our mitochondrial genome tree agrees with the Buckley et al. combined tree's placement of the Awakino Ski Field *M. mangu* specimen. We lacked a Hakataramea *M. mangu* specimen for our AHE analysis, but our combined phylogenomic tree disagrees with Buckley et al.'s (2006) combined tree and groups the Awakino Ski Field *M. mangu* specimen with the other *M. mangu* specimens. This suggests that the Awakino Ski Field mtDNA was the result of gene flow from another species (see 'Evidence for hybridization—*Maoricicada*' below).

To explore whether the signal from the mtDNA was conflicting with the signal from the three nuclear genes, we reanalyzed the data from Buckley et al. (2006) to infer a concatenated ML tree using only the nuclear genes. This tree is better resolved than the published Buckley et al. (2006) combined tree, indicating that there might have been some mito-nuclear discordance in the original Buckley et al. (2006) dataset. Despite using only nuclear genes, this reanalyzed tree has many differences from the nuclear phylogenomic trees we present in this study. Interestingly, the reanalyzed tree still does not place *M. hamiltoni* as sister to all other *Maoricicada*, but instead continues to place it within the lowland clade with *M. myersi*, *M. lindsayi*, and *M. iolanthe*. Instead of recovering a monophyletic alpine radiation, the three-gene reanalyzed tree moves the alpine species *M. clamitans*, *M. oromelaena* and *M. alticola* into a clade sister to the lowland species *M. campbelli*. The resulting grouping of *M. campbelli*, *M. clamitans*, *M. oromelaena* and 'false *M. alticola*' is recovered as sister to the clade of the other lowland species. These relationships in the three-gene reanalyzed tree provide us with two interesting observations. First, they seem to form a sort of compromise between our nuclear phylogenomic trees and our mitochondrial genome tree: *M. campbelli* is being pulled away from the other lowland species towards the alpine *M. oromelaena* and *M. clamitans*, and the alpine *M. oromelaena* and *M. clamitans* are being pulled away from their alpine relatives and towards *M. campbelli*'s lowland relatives. This suggests that there is some signal within the nuclear genes period, calmodulin and EF-1 alpha that agrees with the mitochondrial genome topology. Second, *M. campbelli* and *M. iolanthe* are separated in the three-gene reanalyzed tree, whereas in our nuclear phylogenomic trees and our

mitochondrial genome tree, they are always recovered as sister to one another. This provides additional support for the hypothesis in Buckley et al. (2006) that the relationship between *M. campbelli* and *M. iolanthe* was shaped by ancient hybridization.

In contrast to the strong disagreement between the Sanger-based nuclear tree of Buckley et al. and our *Maoricicada* nuclear phylogenomic trees, there is much more agreement between Buckley et al.'s Sanger-based mitochondrial tree and our *Maoricicada* mitochondrial genome tree. Ignoring branches with poor support in both trees, not surprisingly the mitochondrial genome tree is similar to the Sanger-based mtDNA tree with higher bootstrap support. The few branches in the mitochondrial genome tree with low support correspond to areas with low support in the Buckley et al. (2006) tree, which suggests that these splits in the *Maoricicada* radiation happened too quickly for the mitochondrial genome to have coherent signal. In the literature, this is termed a 'hard polytomy' (Maddison, 1989).

Comparison of analysis types

The three main phylogenetic analyses we performed on the nuclear genomic data yielded slightly different results. These differences can largely be attributed to the differences in approach among concatenated maximum likelihood, ASTRAL and SVDQuartets. ASTRAL and SVDQuartets are both coalescent approaches, explicitly allowing different parts of the genome to have different evolutionary histories because of deep coalescence (ILS), whereas concatenated ML assumes all parts of the genome share the same evolutionary history (Kubatko & Degnan, 2007; Maddison, 1997; Rokas et al., 2003; Salichos & Rokas, 2013). Because of this, phylogenomics using concatenated ML can be sensitive to a handful of rogue genes with wildly different signals than the rest of the sampled loci (Brown & Thomson, 2017; Jiang et al., 2020; Shen et al., 2017, 2021; Walker et al., 2018). Furthermore, ASTRAL and SVDQuartets approach the multispecies coalescent differently: ASTRAL takes fixed gene tree topologies as input, assuming that the species relationships within them are accurate and that all sites within a gene share the same history, whereas SVDQuartets ignores gene boundaries and examines individual site patterns within the total alignment. It should be noted, however, that both coalescent approaches only accommodate ILS and not hybridization.

Comparison of nuclear analysis types—*Kikihia*

The main differences between the different nuclear analysis types in the *Kikihia* trees are between the concatenated maximum likelihood nuclear tree and the two coalescent nuclear trees. The concatenated tree places *K. angusta* and *K.* 'murihikua' as sister species outside the combined Westlandica + Muta group, whereas ASTRAL and SVDQuartets both move *K. angusta* away from *K.* 'murihikua' and place it just outside the Rosea group. This placement is opposite from Dugdale and Fleming's intuitive phylogeny (based on morphology and ecology) where *K.* 'murihikua' is closely related to the Rosea group and *K. angusta* is closely related to the grass singers of the Muta group. The swapped positions of these two species compared with the natural inference from habitat, song and morphology are peculiar. Given that there is disagreement between the concatenated and coalescent trees in the grouping of *K. angusta* and *K.* 'murihikua', it is likely that these two species have hybridized in the past, a hypothesis that was also proposed in previous studies (Banker et al., 2017; Marshall et al., 2008, 2011). Since the coalescent methods used in this study do not accommodate hybridization, the swapped position of *K. angusta* and *K.* 'murihikua' in the ASTRAL and SVDQuartets trees compared to the intuitive phylogeny may be an artefact. Additional analyses using methods that accommodate hybridization will shed light on the placement of these two species.

The concatenated tree places *K.* 'tasmani' sister to the combined Westlandica + Muta group, whereas the ASTRAL and SVDQuartets trees place *K.* 'tasmani' as sister to *K.* 'westlandica north' and *K.* 'westlandica south' with lower support. The only differences between the ASTRAL and SVDQuartets trees are the placement of *K. horologium* with respect to *K.* 'murihikua' and the placement of *K.* 'aotea-east', but these differences are not well supported in the SVDQuartets tree.

Comparison of analysis types—*Maoricicada*

In the concatenated tree, *M.* 'false alticola' is sister to *M. oromelaena* (including the proposed-hybrid species *M.* 'cassiomelans'), whereas in the ASTRAL tree, *M. clamitans* is sister to *M. oromelaena*, with *M.* 'false alticola' sister to the two of them. In the SVDQuartets tree, *M. iolanthe* is sister to a clade containing *M. myersi*, *M. lindsayi* and *M. campbelli*, whereas in the ASTRAL tree, *M. iolanthe* is sister to *M. campbelli* only. However, the placement in the SVDQuartets tree is not well-supported. Buckley et al. (2006) found incongruence in the placement of *M. iolanthe* similar to our findings here, with the main disagreement being whether *M. iolanthe* grouped with *M. myersi* and *M. lindsayi* or if it grouped with *M. campbelli*, and attributed this disagreement to hybridization. It is interesting that our two coalescent analyses found similar shifting in the placement of *M. iolanthe*, and further work using methods that accommodate hybridization will be necessary to investigate this issue.

Evidence of conflict in datasets

The persistently low gCF, sCF and quartet-based IC values demonstrate that our NZ cicada nuclear phylogenomic datasets contain large amounts of conflict. GCF, sCF and IC offer a complement to traditional measures of branch support such as bootstraps or PP in that they tend not to increase as more loci or sites are added to the dataset (Minh et al. 2020; Salichos & Rokas, 2013). Since many of these branches with low gCF, sCF and IC values have 100% bootstrap or PP, our results further support the observation that high bootstrap

support can mask large amounts of conflict in phylogenomic datasets (Salichos & Rokas, 2013).

We have three explanations for the large amount of conflict in our nuclear datasets: two biological and one artifactual. One biological explanation is that the conflict reflects the signal from deep coalescence/ILS (Degnan & Rosenberg, 2006; Maddison, 1997). Alternatively, the conflict among genes or sites could be the result of different genes or sites sharing different histories from hybridization and/or introgression events. The coalescent methods used in our study are only able to model ILS as the source of gene tree or site conflict, so other methods are needed for exploring hybridization and/or introgression. A third explanation for the source of the conflict is gene tree estimation error caused by low information content or the presence of recombination breakpoints within gene sequences (Gatesy & Springer, 2014; Roch & Warnow, 2015; Simmons & Gatesy, 2021). It is likely that our datasets have some gene tree estimation error, because the Auchenorryncha AHE probe set we used was designed for much deeper phylogenetic scales than the genus-level phylogenies inferred in this study (Dietrich et al., 2017). For shallower phylogenetic scales, probe sets incorporating more rapidly evolving loci can increase gene tree resolution (Banker et al., 2020). However, the genomic resources to produce such a probe set for NZ cicadas were unavailable.

Gene tree estimation error only affects some of our measures of conflict in our nuclear datasets, and during our analyses, we made efforts to reduce incorrect gene tree resolution. Before using them in phylogenetic analyses, we collapsed all the branches in our gene trees with SH-aLRT support values of 0% following the recommendations of Simmons and Gatesy (2021). Not only did we build our ASTRAL species trees from gene trees that had such branches collapsed but we also used these branch-collapsed gene trees as input for determining gCF and IC values. While this means the gCF and IC values we obtained are free from the effect of the most egregious estimation errors, an SH-aLRT support value of 0% is still a very low threshold for determining if a branch is dubiously resolved, meaning that gene tree estimation error is likely still a cause of the conflict in our datasets. However, gene tree estimation error is only a factor for measures of conflict that use gene trees (gCF and IC); it does not affect sCF values as they are a measure of support or conflict from the individual site patterns.

Mito-nuclear discordance as evidence for hybridization

There are several instances of discordance between the mitochondrial and nuclear tree topologies for both NZ cicada datasets. While these discordance events may be the result of ILS (treating the mitochondrial tree as a gene tree within the overall species tree), there is a strong possibility that some of this discordance is the result of hybridization. We present this evidence for hybridization for NZ cicadas below.

Evidence for hybridization—*Maoricicada*

The first major difference in topology between the mitochondrial and nuclear trees for *Maoricicada* is the placement of *M. hamiltoni*. In the nuclear trees, *M. hamiltoni* is strongly supported as the sister to all other *Maoricicada*, whereas in the mitochondrial tree, this species forms a clade with *M. myersi* and *M. lindsayi*. These three lowland species may group together in the mitochondrial tree due to ancient hybridization or ILS, and resolving these two processes is difficult (Holder et al., 2001). Additional analyses are required to distinguish between ILS and hybridization for the placement of *M. hamiltoni*.

A second major example of discordance between the mitochondrial and nuclear trees is the placement of *M. campbelli* and *M. iolanthe*. Fleming (1971) placed *M. iolanthe* as sister to all other *Maoricicada*. Contrary to Fleming (1971), in our AHE nuclear trees, both *M. campbelli* and *M. iolanthe* group with the lowland species *M. myersi* and *M. lindsayi*. In the mitochondrial genome tree, they nest within the alpine radiation. *M. iolanthe* and *M. campbelli* are primarily lowland to mid-elevation species and are similar in genitalia phenotype to the other lowland species, but *M. campbelli* is distributed across a wider elevation range, meaning it is possible for it to have captured an alpine-type mitochondrial genome through hybridization with an alpine species and then to have shared it with *M. iolanthe*. Buckley et al. (2001) using only mitochondrial DNA proposed ILS or early splitting of an alpine lineage with retained ancestral phenotype as an explanation for the anomalous placement of *M. campbelli* and *M. iolanthe* with the alpine species in their mitochondrial tree; however, they concluded based on Shimodaira–Hasegawa tests that there was 'a surprising amount of ambiguity in the phylogenetic placement' of these two species. Buckley et al. (2006), using additional mitochondrial and nuclear genes, tested lineage sorting versus hybridization as possible explanations of *M. campbelli* and *M. iolanthe* as sister taxa and concluded that hybridization was best supported. The consistent placement of *M. campbelli* and *M. iolanthe* with the other lowland species in all our AHE nuclear analyses also appears to rule out the retained ancestral phenotype hypothesis.

Finally, we also observe discordance between the mitochondrial and nuclear trees in the placement of an *M. mangu* individual from the Awakino Ski Field. Buckley et al. (2006) noted that their *M. mangu* sample from this locality did not group with the other *M. mangu* subspecies specimens in their mitochondrial tree and in one of their three nuclear gene trees; our mitochondrial genome tree confirms this previous finding. The Awakino Ski Field location is unusual, as it is a southern range extension of *M. mangu*. Dugdale and Fleming (1978) originally thought that the species was not found south of the Waitaki River. The placement of this individual in our mitochondrial tree, along with Buckley et al.'s (2006) finding, suggests that the Awakino Ski Field population of *M. mangu* acquired its mitochondrion from introgression with a now-extinct species of *Maoricicada*, that is, a 'ghost lineage' (Ai et al., 2015; Huang et al., 2014; Norell, 1993; Zhang et al., 2019). In addition, all the nuclear analyses place this individual at the base of *M. mangu* instead of with the individuals of *M. m. mangu*, with which the Awakino individual shares its song and

morphology. Buckley et al. (2006) suggest that a specimen of *M. mangu* collected from Hakatarmea Pass, just north of the Waitaki River, also has the ghost-lineage mtDNA of the Awakino SF individual. Further investigation of the nuclear gene data may find additional signal for this ancient hybridization event.

Evidence for hybridization—*Kikihia*

The strongest evidence for hybridization in *Kikihia* comes from the presence of the Westlandica group in the mitochondrial tree. The group contains *K. subalpina* and *K.* 'flemingi' and is positioned in the same location as those two species in the nuclear trees, but it also contains five other species that are found elsewhere in the nuclear trees: *K. angusta*, *K.* 'murihikua', *K.* 'tasmani', *K.* 'westlandica north' and *K.* 'westlandica south'. The position of these species in the nuclear trees compared to the mitochondrial trees strongly suggests mitochondrial capture as the explanation for the discordance. Earlier phylogenetic hypotheses based on ecology and morphology (Fleming, 1975, 1984) did not group these species in a single clade either. We hypothesize that a Muta-group cicada related to *K. angusta*, *K.* 'westlandica north' and *K.* 'westlandica south' captured mitochondria from the ancestor of *K. subalpina* and *K.* 'flemingi' via hybridization. Based on the differences in topology among the nuclear trees explored above, it appears that there was a mitochondrial capture event between *K. angusta* and *K.* 'murihikua'. The exact relationships of *K.* 'murihikua', *K.* 'tasmani', *K.* 'westlandica north' and *K.* 'westlandica south' are not well-resolved in the nuclear trees, but it is possible that another mitochondrial capture event united the ancestor of *K.* 'murihikua' with the ancestor of one or more of the other three. Alternatively, this grouping of cicadas could represent the true species relationships or be the result of ILS/deep coalescence.

Our hypothesis that the Westlandica group found in the mitochondrial tree is an artefact of hybridization is supported by the geography and ecology of the constituent species. Unlike the other *Kikihia* species groups, the Westlandica group is neither monophyletic in the nuclear phylogenomic trees nor associated with a single habitat type; instead, it contains species from several disparate habitat types. The group contains grass and tussock specialists (*K.* 'westlandica north' and *K.* 'westlandica south'; *K. angusta*), a wet shrub specialist (*K.* 'tasmani'), forest and forest edge specialists (*K.* 'flemingi', *K. subalpina*), and a dry scrub specialist (*K.* 'murihikua'). All Westlandica group species are exclusively found on the South Island, except for *K. subalpina*, which is restricted to the North Island. Its yet-to-be-described sister species, *K.* 'flemingi' (Marshall et al., 2009), is currently sympatric with all the other South Island species, and *K. angusta* and *K.* 'murihikua' have a region of sympatry on the South Island. The species *K.* 'westlandica north' and *K.* 'westlandica south' have a contact zone with one another in the Punakaiki and Springs Junction regions, and *K.* 'tasmani' is in contact with *K.* 'westlandica north' in NW Nelson (Marshall et al., 2008; Shepherd et al., 2022). However, these are the current-day distributions of these cicadas; a molecular clock analysis suggests that the common ancestor of the Westlandica group mitochondrial lineage is well over 2 million years old (Marshall et al., 2008). While there has been some work using climate modelling to reconstruct past ranges of *K.* 'westlandica north' and *K.* 'westlandica south' throughout the Pleistocene glaciation periods (Wade, 2014), there has been no such range reconstruction for the other Westlandica group species.

CONCLUSION

In this study, we use a phylogenomic dataset of over 500 genomic loci to improve phylogenetic resolution of two NZ cicada species radiations. Given rapid radiations in these groups, some relationships remained unresolved. Our two principal study genera, *Maoricicada* and *Kikihia* exhibited different degrees of difficulty of resolution. *Maoricicada* had many short internodes, and *Kikihia* had fewer short internodes. Both genera show evidence of conflict in our phylogenomic datasets and exhibit mito-nuclear discordance, suggesting a history of hybridization. *Kikihia* is likely to have been more frequently influenced by hybridization than *Maoricicada*. Future work involving statistical introgression tests and phylogenetic network methods is planned to test hypotheses of hybridization and to understand the causes of phylogenetic uncertainty at recalcitrant nodes in the phylogeny.

AUTHOR CONTRIBUTIONS

Mark Stukel: Conceptualization; investigation; writing – original draft; writing – review and editing; methodology; visualization; formal analysis; data curation; software. **Alexandra E. Porczak:** Investigation; writing – original draft; methodology; writing – review and editing; formal analysis; software. **Eric R. L. Gordon:** Investigation; methodology; software. **Jason Vailionis:** Investigation; software; methodology; writing – review and editing. **Diler Haji:** Investigation; writing – review and editing. **Thomas R. Buckley:** Resources; writing – review and editing. **Alan R. Lemmon:** Resources; investigation. **Emily Moriarty Lemmon:** Resources; investigation. **Chris Simon:** Writing – review and editing; resources; funding acquisition; supervision; conceptualization; project administration.

ACKNOWLEDGEMENTS

We would like to thank David Marshall and Kathy Hill for specimen collection, insight on the biology and evolution of NZ cicadas, and feedback on preliminary results; John Dugdale and Charles Fleming for foundational work in developing this system; and Elizabeth Jockusch and Paul Lewis for helpful comments during the design and early analysis phase. We also wish to thank various field biologists from NZ Department of Conservation Te Papa Atawhai for local knowledge and help with Māori contacts, three anonymous reviewers for helpful suggestions in the review stage, and many good colleagues and local New Zealanders too numerous to name for all manner of assistance. Computational resources for analyses were provided by the University of Connecticut Computational Biology Core.

FUNDING INFORMATION

This research was supported by National Science Foundation DEB 04-22386, DEB 07-20664, DEB 09-55849, and DEB 16-55891; the New Zealand MBIE CO9X2201; the LinnéSys: Systematics Research Fund; the Fulbright Foundation; and grants from the University of Connecticut.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data and analysis scripts for this manuscript are available on Dryad: https://doi.org/doi:10.5061/dryad.t1g1jwt7v. Mitochondrial genomes are uploaded to GenBank at accession numbers OR413940-OR413978 and OR459868-OR459906. Raw sequence reads are available on SRA (BioProject PRJNA1015981, BioSample accessions SAMN37366775-SAMN37366854). Paralogy detection pipeline is available on GitHub: https://github.com/markstukel/Simon-target-capture-pipeline.

ETHICS STATEMENT

No research ethics approval required. Specimens were collected with the following permits and permissions. The New Zealand Department of Conservation (Te Papa Atawhai) provided permits from 1996-present. NZ DoC permit 12 March 2018–28 February 2022, Authorisation number: 63973-RES. NZ DoC permit 1 May 2010–30 April 2015, NM-25129-FAU. From 1996 to 2010, NZ DoC permits coordinated by John Dugdale and Ian Millar cited in the acknowledgements of Arensburger, Buckley, et al., 2004; Arensburger, Simon, & Holsinger, 2004; and Buckley et al., 2002; Hill et al., 2009; Marshall et al., 2011, 2009, 2008. NZ DOC local district officials obtained permissions from local Māori Iwi and provided collecting information and access to locked gates (especially Mike Morisee, Faith Barber and Bert Borger). Local landowners provided access and collecting permissions on private land (Peter and Hugh Northcote, Vicki and Chris Parr, and Doug Chiswell). Fred Howe and Marge Jowett provided Norfolk Island National Park permits.

ORCID

Mark Stukel https://orcid.org/0000-0002-4774-7991
Thomas R. Buckley https://orcid.org/0000-0002-3076-4234

REFERENCES

Ai, H., Fang, X., Yang, B., Huang, Z., Chen, H., Mao, L. et al. (2015) Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nature Genetics*, 47, 217–225.

Andrews S. (2010) *FastQC: a quality control tool for high throughput sequence data*, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Anisimova, M. & Gascuel, O. (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Systematic Biology*, 55, 539–552.

Arensburger, P., Buckley, T.R., Simon, C., Moulds, M. & Holsinger, K.E. (2004) Biogeography and phylogeny of the New Zealand cicada genera (Hemiptera: Cicadidae) based on nuclear and mitochondrial DNA data. *Journal of Biogeography*, 31, 557–569.

Arensburger, P., Simon, C. & Holsinger, K. (2004) Evolution and phylogeny of the New Zealand cicada genus Kikihia Dugdale (Homoptera: Auchenorrhyncha: Cicadidae) with special reference to the origin of the Kermadec and Norfolk Islands' species. *Journal of Biogeography*, 31, 1769–1783.

Ballesteros, J.A. & Hormiga, G. (2016) A new orthology assessment method for phylogenomic data: unrooted phylogenetic orthology. *Molecular Biology and Evolution*, 33, 2117–2134.

Banker, S.E., Lemmon, A.R., Hassinger, A.B., Dye, M., Holland, S.D., Kortyna, M.L. et al. (2020) Hierarchical hybrid enrichment: multitiered genomic data collection across evolutionary scales, with application to chorus frogs (Pseudacris). *Systematic Biology*, 69, 756–773.

Banker, S.E., Wade, E.J. & Simon, C. (2017) The confounding effects of hybridization on phylogenetic estimation in the New Zealand cicada genus Kikihia. *Molecular Phylogenetics and Evolution*, 116, 172–181.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19, 455–477.

Bator, J., Marshall, D.C., Hill, K.B.R., Cooley, J.R., Leston, A. & Simon, C. (2021) Phylogeography of the endemic red-tailed cicadas of New Zealand (Hemiptera: Cicadidae: Rhodopsalta), and molecular, morphological and bioacoustical confirmation of the existence of Hudson's Rhodopsalta microdora. *Zoological Journal of the Linnean Society*, 195, 1219–1244.

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120.

Brown, J.M. & Thomson, R.C. (2017) Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. *Systematic Biology*, 66, 517–530.

Brunke, A.J., Hansen, A.K., Salnitska, M., Kypke, J.L., Predeus, A.V., Escalona, H. et al. (2021) The limits of Quediini at last (Staphylinidae: Staphylininae): a rove beetle mega-radiation resolved by comprehensive sampling and anchored phylogenomics. *Systematic Entomology*, 46, 396–421.

Buckley, T.R., Arensburger, P., Simon, C. & Chambers, G.K. (2002) Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. *Systematic Biology*, 51, 4–18.

Buckley, T.R., Cordeiro, M., Marshall, D.C. & Simon, C. (2006) Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (Maoricicada Dugdale). *Systematic Biology*, 55, 411–425.

Buckley, T.R. & Simon, C. (2007) Evolutionary radiation of the cicada genus Maoricicada Dugdale (Hemiptera: Cicadoidea) and the origins of the New Zealand alpine biota. *Biological Journal of the Linnean Society*, 91, 419–435.

Buckley, T.R., Simon, C., Shimodaira, H. & Chambers, G.K. (2001) Evaluating hypotheses on the origin and evolution of the New Zealand alpine cicadas (Maoricicada) using multiple-comparison tests of tree topology. *Molecular Biology and Evolution*, 18, 223–234.

Bushnell, B., Rood, J. & Singer, E. (2017) BBMerge—accurate paired shotgun read merging via overlap. *PLoS One*, 12, e0185056.

Butlin, R.K., Galindo, J. & Grahame, J.W. (2008) Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 2997–3007.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 1–9.

Chernomor, O., von Haeseler, A. & Minh, B.Q. (2016) Terrace aware data structure for Phylogenomic inference from Supermatrices. *Systematic Biology*, 65, 997–1008.

Chifman, J. & Kubatko, L. (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30, 3317–3324.

Craw, D., Nelson, E., & Koons, P.O. (2003) Structure and topographic evolution of the Main Divide in the Landsborough-Hopkins area of the Southern Alps, New Zealand. *New Zealand Journal of Geology and Geophysics*, 46, 553–562.

Crosby, T.K., Dugdale, J.S. & Watt, J.C. (1998) Area codes for recording specimen localities in the New Zealand subregion. *New Zealand Journal of Zoology*, 25, 175–183.

Degnan, J.H. & Rosenberg, N.A. (2006) Discordance of species trees with their most likely gene trees. *PLOS Genetics*, 2, e68.

Dietrich, C.H., Allen, J.M., Lemmon, A.R., Lemmon, E.M., Takiya, D.M., Evangelista, O. et al. (2017) Anchored hybrid enrichment-based phylogenomics of leafhoppers and treehoppers (Hemiptera: Cicadomorpha: Membracoidea). *Insect Systematics and Diversity*, 1, 57–72.

Dugdale, J.S. & Fleming, C.A. (1978) New Zealand cicadas of the genus Maoricicada (Homoptera: Tibicinidae). *New Zealand Journal of Zoology*, 5, 295–340.

Fleming, C.A. (1971) new species of cicada from rock fans in southern Wellington, with a review of three species with similar songs and habitat. *New Zealand Journal of Science*, 14, 443–479.

Fleming, C.A. (1975) Adaptive radiation in New Zealand cicadas. *Proceedings of the American Philosophical Society*, 119, 298–306.

Fleming, C.A. (1984) The cicada genus Kikihia Dugdale (Hemiptera, Homoptera). Part. 1: the New Zealand green foliage cicadas. *Records-National Museum of New Zealand*, 2, 191–206.

Franco, A.D., Poujol, R., Baurain, D. & Philippe, H. (2019) Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evolutionary Biology*, 19, 1–17.

Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28, 3150–3152.

Gatesy, J. & Springer, M.S. (2014) Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution*, 80, 231–266.

Giarla, T.C. & Esselstyn, J.A. (2015) The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology*, 64, 727–740.

Granados Mendoza, C., Jost, M., Hágsater, E., Magallón, S., van den Berg, C., Lemmon, E.M. et al. (2020) Target nuclear and off-target plastid hybrid enrichment data inform a range of evolutionary depths in the orchid genus Epidendrum. *Frontiers in Plant Science*, 10, 1761.

Hahn, C., Bachmann, L. & Chevreux, B. (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*, 41, e129.

Haji, D., Vailionis, J., Stukel, M., Gordon, E., Lemmon, E.M., Lemmon, A.R. et al. (2022) Lack of host phylogenetic structure in the gut bacterial communities of New Zealand cicadas and their interspecific hybrids. *Scientific Reports*, 12, 20559.

Heath, J.E., Heath, M.S. & Sanborn, A.F. (2022) Cold cicadas and hot rocks: thermal responses and thermoregulation in some New Zealand cicadas (Insecta: Hemiptera: Cicadidae: Cicadettinae: Cicadettini). *Journal of Thermal Biology*, 107, 103273.

Hill, K.B.R., Simon, C., Marshall, D.C., & Chambers, G.K. (2009) Surviving glacial ages within the Biotic Gap: Phylogeography of the New Zealand cicada Maoricicada campbelli. *Journal of Biogeography*, 36, 675–692.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35, 518–522.

Holder, M.T., Anderson, J.A. & Holloway, A.K. (2001) Difficulties in detecting hybridization. *Systematic Biology*, 50, 978–982.

Huang, D.I., Hefer, C.A., Kolosova, N., Douglas, C.J. & Cronk, Q.C.B. (2014) Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, Populus balsamifera and P. trichocarpa (Salicaceae). *New Phytologist*, 204, 693–703.

Irisarri, I., Singh, P., Koblmüller, S., Torres-Dowdall, J., Henning, F., Franchini, P. et al. (2018) Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nature Communications*, 9, 3159.

Jiang, X., Edwards, S.V. & Liu, L. (2020) The multispecies coalescent model outperforms concatenation across diverse phylogenomic data sets. *Systematic Biology*, 69, 795–812.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S. et al. (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28, 1647–1649.

King, P.R. (2000) Tectonic reconstructions of New Zealand: 40 Ma to the Present. *New Zealand Journal of Geology and Geophysics*, 43, 611–638.

Knyshov, A., Gordon, E.R.L. & Weirauch, C. (2021) New alignment-based sequence extraction software (ALiBaSeq) and its utility for deep level phylogenetics. *PeerJ*, 9, e11019.

Kubatko, L.S. & Degnan, J.H. (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, 56, 17–24.

Lane D.H. (1984) *An enquiry into suspected hybridization in zones of overlap involving species of the genus Kikihia*, Victoria University of Wellington, Wellington, New Zealand.

Lane, D.H. (1995) The recognition concept of species applied in an analysis of putative hybridization in New Zealand cicadas of the genus Kikihia (Insecta: Hemiptera: Tibicinidae). In: Lambert, D.M. & Spencer, H.G. (Eds.) *Speciation and the recognition concept: theory and application*. Baltimore, MD: Johns Hopkins University Press, pp. 367–421.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2017) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34, 772–773.

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, 61, 727–744.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, 25, 1754–1760.

Łukasik, P., Chong, R.A., Nazario, K., Matsuura, Y., Bublitz, D.A.C., Campbell, M.A. et al. (2019) One hundred mitochondrial genomes of cicadas. *Journal of Heredity*., 110, 247–256.

Maddison, W. (1989) Reconstructing character evolution on polytomous cladograms. *Cladistics*, 5, 365–377.

Maddison, W.P. (1997) Gene trees in species trees. *Systematic Biology*, 46, 523–536.

Maddison, W.P. & Knowles, L.L. (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, 55, 21–30.

Mallet, J., Besansky, N. & Hahn, M.W. (2016) How reticulated are species? *BioEssays*, 38, 140–149.

Marsden, C.D., Lee, Y., Nieman, C.C., Sanford, M.R., Dinis, J., Martins, C. et al. (2011) Asymmetric introgression between the M and S forms of the malaria vector, Anopheles gambiae, maintains divergence despite extensive hybridization. *Molecular Ecology*, 20, 4983–4994.

Marshall, D.C., Hill, K.B.R., Cooley, J.R. & Simon, C. (2011) Hybridization, mitochondrial DNA phylogeography, and prediction of the early stages of reproductive isolation: lessons from New Zealand cicadas (Genus Kikihia). *Systematic Biology*, 60, 482–502.

Marshall, D.C., Hill, K.B.R., Fontaine, K.M., Buckley, T.R. & Simon, C. (2009) Glacial refugia in a maritime temperate climate: cicada (Kikihia subalpina) mtDNA phylogeography in New Zealand. *Molecular Ecology*, 18, 1995–2009.

Marshall, D.C., Slon, K., Cooley, J.R., Hill, K.B.R. & Simon, C. (2008) Steady Plio-Pleistocene diversification and a 2-million-year sympatry threshold in a New Zealand cicada radiation. *Molecular Phylogenetics and Evolution*, 48, 1054–1066.

Marske, K.A. & Boyer, S.L. (2022) Phylogeography reveals the complex impact of the last glacial maximum on New Zealand's terrestrial biota. *Journal of the Royal Society of New Zealand*, 54, 8–29.
Meusemann, K., von Reumont, B.M., Simon, S., Roeding, F., Strauss, S., Kück, P. et al. (2010) A phylogenomic approach to resolve the arthropod tree of life. *Molecular Biology and Evolution*, 27, 2451–2464.
Minh, B.Q., Hahn, M.W. & Lanfear, R. (2020) New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution*, 37, 2727–2733.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37, 1530–1534.
Morales, A.E., Jackson, N.D., Dewey, T.A., O'Meara, B.C. & Carstens, B.C. (2017) Speciation with gene flow in north American Myotis bats. *Systematic Biology*, 66, 440–452.
Norell, M.A. (1993) Tree-based approaches to understanding history; comments on ranks, rules and the quality of the fossil record. *American Journal of Science*, 293, 407–417.
Nosil, P. (2008) Speciation with gene flow could be common. *Molecular Ecology*, 17, 2103–2106.
Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M. et al. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, 526, 569–573.
Pyron, R.A., Hendry, C.R., Chou, V.M., Lemmon, E.M., Lemmon, A.R. & Burbrink, F.T. (2014) Effectiveness of phylogenomic data and coalescent species-tree methods for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes: Caenophidia). *Molecular Phylogenetics and Evolution*, 81, 221–231.
Quinlan, A.R. & Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842.
Roch, S. & Warnow, T. (2015) On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. *Systematic Biology*, 64, 663–676.
Rokas, A., Williams, B.L., King, N. & Carroll, S.B. (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, 425, 798–804.
Salichos, L. & Rokas, A. (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, 497, 327–331.
Shen, X.-X., Hittinger, C.T. & Rokas, A. (2017) Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology & Evolution*, 1, 126.
Shen, X.-X., Steenwyk, J.L. & Rokas, A. (2021) Dissecting incongruence between concatenation- and quartet-based approaches in phylogenomic data. *Systematic Biology*, 70, 997–1014.
Shepherd, L., Simon, C., Langton-Myers, S. & Morgan-Richards, M. (2022) Insights into Aotearoa New Zealand's biogeographic history provided by the study of natural hybrid zones. *Journal of the Royal Society of New Zealand*, 54, 55–74.
Simmons, M.P. & Gatesy, J. (2021) Collapsing dubiously resolved gene-tree branches in phylogenomic coalescent analyses. *Molecular Phylogenetics and Evolution*, 158, 107092.
Simon, C., Gordon, E.R.L., Moulds, M.S., Cole, J.A., Haji, D., Lemmon, A.R. et al. (2019) Off-target capture data, endosymbiont genes and morphology reveal a relict lineage that is sister to all other singing cicadas. *Biological Journal of the Linnean Society*, 128, 865–886.
Simon, S., Strauss, S., von Haeseler, A. & Hadrys, H. (2009) A phylogenomic approach to resolve the basal Pterygote divergence. *Molecular Biology and Evolution*, 26, 2719–2730.
Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.
Wade E.J. (2014) *Species and hybridization: understanding the exchange of nuclear and mitochondrial DNA in song-delimited cicada species complexes*, University of Connecticut, Storrs, CT, USA.
Walker, J.F., Brown, J.W. & Smith, S.A. (2018) Analyzing contentious relationships and outlier genes in phylogenomics. *Systematic Biology*, 67, 916–924.
Whitfield, J.B. & Kjer, K.M. (2008) Ancient rapid radiations of insects: challenges for phylogenetic analysis. *Annual Review of Entomology*, 53, 449–472.
Whitfield, J.B. & Lockhart, P.J. (2007) Deciphering ancient rapid radiations. *Trends in Ecology & Evolution.*, 22, 258–265.
Wilgenbusch, J.C. & Swofford, D. (2003) Inferring evolutionary trees with PAUP*. *Current Protocols in Bioinformatics*, 6.4.1–6.4.28.
Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19, 153.
Zhang, D., Tang, L., Cheng, Y., Hao, Y., Xiong, Y., Song, G. et al. (2019) "Ghost Introgression" as a cause of deep mitochondrial divergence in a bird species complex. *Molecular Biology and Evolution*, 36, 2375–2386.
Zhou, X., Lutteropp, S., Czech, L., Stamatakis, A., Looz, M.V. & Rokas, A. (2020) Quartet-based computations of internode certainty provide robust measures of phylogenetic incongruence. *Systematic Biology*, 69, 308–324.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1. Supporting Information.

Figure S1. Left: *Kikihia* concatenated ML nuclear phylogenybuilt from 516 AHE nuclear genes. Heatmaps for corresponding node numbers display SH-aLRTsupport, ultrafast bootstrap support (UFB), gene concordance factor (gCF), site concordance factor (sCF), lowest quartet internode certainty (LQ-IC), quadripartition internode certainty (QP-IC), and extended quadripartition internode certainty (EQP-IC). Branch lengths are in substitutions/site. Inset right: 5-gene nuclear phylogeny reproduced from Banker et al. (2017) for comparison. Shapes on branch tips refer to whether the specimen is from the North Island (upwards triangle), South Island (downwards triangle), or offshore island (circle). Refer to figure caption in original paper for details.

Figure S2. Left: *Kikihia* SVDQuartetsnuclear phylogenybuilt from 516 AHE nuclear genes. Heatmaps for corresponding node numbers display bootstrap support (BP), gene concordance factor (gCF), site concordance factor (sCF), lowest quartet internode certainty (LQ-IC), quadripartition internode certainty (QP-IC), and extended quadripartition internode certainty (EQP-IC). Branch lengths are unscaled. Inset right: 5-gene nuclear phylogeny reproduced from Banker et al. (2017) for comparison. Shapes on branch tips refer to whether the specimen is from the North Island (upwards triangle), South Island (downwards triangle), or offshore island (circle). Refer to figure caption in original paper for details.

Figure S3. Left: *Kikihia* concatenated ML nuclear+mtDNAphylogenybuilt from mitochondrial genomes and 516 nuclear genes. Upper support values are SH-aLRT values; support values underneath are ultrafast bootstrap percentages (UFB). Branch lengths are in substitutions/site. Inset right: phylogeny combining 2152 bp of mtDNAsequence and 1545 bp EF-1 alpha nuclear sequence reproduced from Marshall et al. (2008) for comparison.

Figure S4. Left: *Maoricicada* concatenated ML nuclear phylogeny built from 505 AHE nuclear genes. Heatmaps for corresponding node numbers display SH-aLRT support, ultrafast bootstrap support (UFB), gene concordance factor (gCF), site concordance factor (sCF), lowest quartet internode certainty (LQ-IC), quadripartition internode certainty (QP-IC), and extended quadripartition internode certainty (EQP-IC). Branch lengths are in substitutions/site. Inset right: 3-gene nuclear phylogeny using sequence data from Buckley et al. (2006) for comparison.

Figure S5. Left: *Maoricicada* SVDQuartets nuclear phylogeny built from 505 AHE nuclear genes. Heatmaps for corresponding node numbers display bootstrap support (BP), gene concordance factor (gCF), site concordance factor (sCF), lowest quartet internode certainty (LQ-IC), quadripartition internode certainty (QP-IC), and extended quadripartition internode certainty (EQP-IC). Branch lengths are unscaled. Inset right: 3-gene nuclear phylogeny using sequence data from Buckley et al. (2006) for comparison.

Figure S6. Left: *Maoricicada* concatenated ML nuclear+mtDNA phylogeny built from mitochondrial genomes and 505 AHE nuclear genes. Upper supports are SH-aLRTvalues; supports underneath are ultrafast bootstrap percentages (UFB). Branch lengths are in substitutions/site. Inset right: phylogeny combining 3 nuclear genes and 1 mitochondrial gene reproduced from Buckley et al. (2006) for comparison.

Figure S7. Histograms of gene concordance factors (gCF, top) and site concordance factors (sCF, bottom) for branches in *Kikihia* ASTRAL (left), Concatenated (center), and SVDQuartets (right) nuclear phylogenies. Vertical bar in sCF histograms represents 33% sCF, which indicates equal support for all three quartet arrangements.

Figure S8. Histograms of gene concordance factors (gCF, top) and site concordance factors (sCF, bottom) for branches in *Maoricicada* ASTRAL (left), Concatenated (center), and SVDQuartets (right) nuclear phylogenies. Vertical bar in sCF histograms represents 33% sCF, which indicates equal support for all three quartet arrangements.

Figure S9. *Kikihia* ASTRAL phylogeny displaying A: local posterior probability, B: gene concordance factors, and C: site concordance factors. Branch lengths are in coalescent units. Terminal branches are unscaled.

Figure S10. *Kikihia* ASTRAL phylogeny displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC). Branch lengths are in coalescent units. Terminal branches are unscaled.

Figure S11. *Kikihia* ASTRAL phylogeny displaying A: the highest branch quartet support frequency, B: the first alternative branch quartet support frequency, and C: the second alternative branch quartet support frequency. Branch lengths are in coalescent units. Terminal branches are unscaled.

Figure S12. *Kikihia* concatenated ML phylogeny displaying A: SH-aLRT branch supports, and B: Ultrafast bootstrap branch supports (UFB). Branch lengths are in substitutions/site.

Figure S13. *Kikihia* concatenated ML phylogeny displaying A: gene concordance factors, and B: site concordance factors. Branch lengths are in substitutions/site.

Figure S14. *Kikihia* concatenated ML phylogeny displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC). Branch lengths are in substitutions/site.

Figure S15. *Kikihia* SVDQuartets cladogram displaying A: bootstrap percentage, B: gene concordance factors, and C: site concordance factors.

Figure S16. *Kikihia* SVDQuartets cladogram displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC).

Figure S17. *Maoricicada* ASTRAL phylogeny displaying A: local posterior probability, B: gene concordance factors, and C: site concordance factors. Branch lengths are in coalescent units. Terminal branches are unscaled.

Figure S18. *Maoricicada* ASTRAL phylogeny displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC). Branch lengths are in coalescent units. Terminal branches are unscaled.

Figure S19. *Maoricicada* ASTRAL phylogeny displaying A: the highest branch quartet support frequency, B: the first alternative branch quartet support frequency, and C: the second alternative branch quartet support frequency. Branch lengths are in coalescent units. Terminal branches are unscaled.

Figure S20. *Maoricicada* concatenated ML phylogeny displaying A: SH-aLRT branch supports, and B: Ultrafast bootstrap branch supports (UFB). Branch lengths are in substitutions/site.

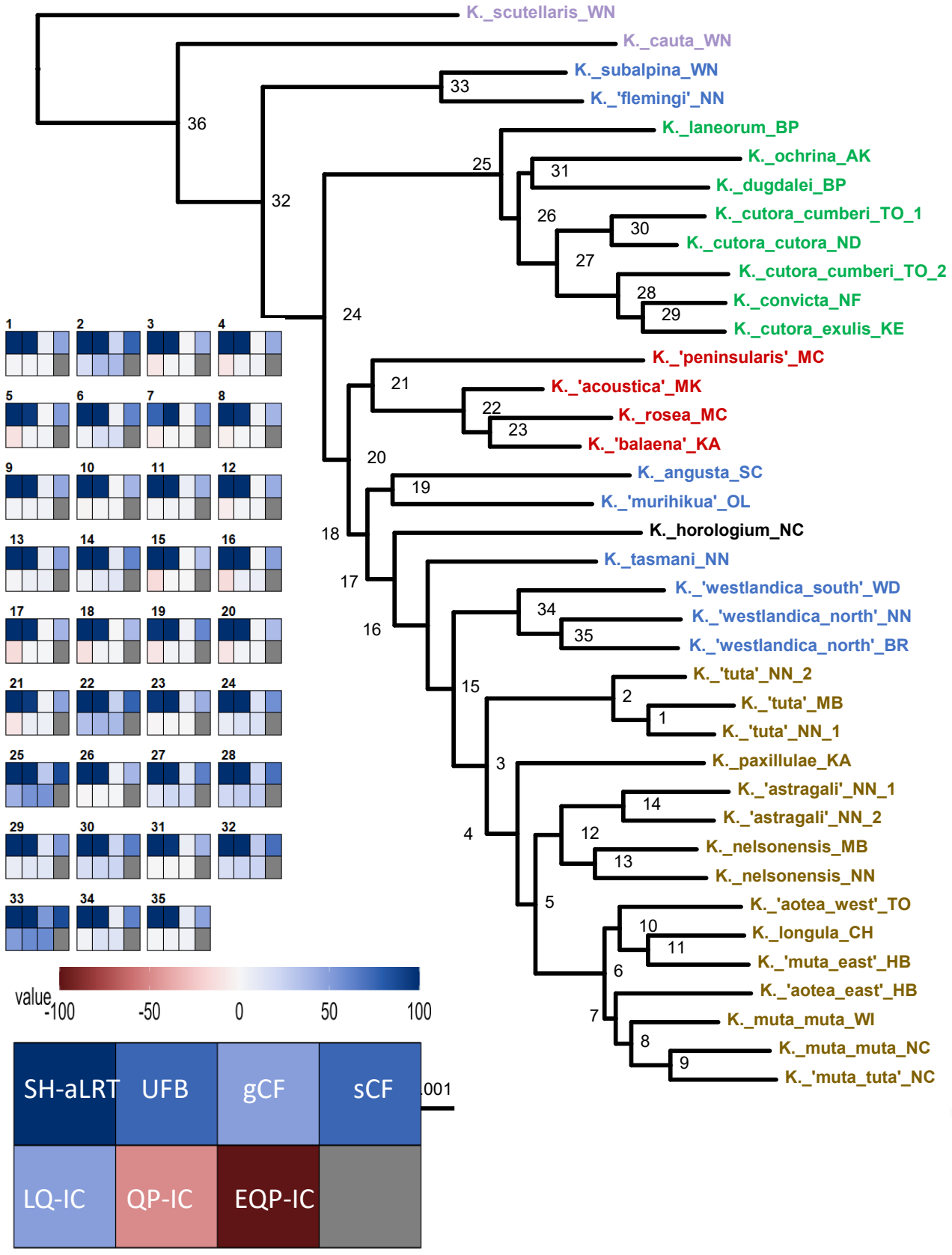
Figure S21. *Maoricicada* concatenated ML phylogeny displaying A: gene concordance factors, and B: site concordance factors. Branch lengths are in substitutions/site.

Figure S22. *Maoricicada* concatenated ML phylogeny displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC). Branch lengths are in substitutions/site.

Figure S23. *Maoricicada* SVDQuartets cladogram displaying A: bootstrap percentage, B: gene concordance factors, and C: site concordance factors.

Figure S24. *Maoricicada* SVDQuartets cladogram displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC).

How to cite this article: Stukel, M., Porczak, A.E., Gordon, E.R.L., Vailionis, J., Haji, D., Buckley, T.R. et al. (2023) Phylogenomics improves the phylogenetic resolution and provides strong evidence of mito-nuclear discordance in two genera of a New Zealand cicada (Hemiptera: Cicadidae) species radiation. *Systematic Entomology*, 1–21. Available from: https://doi.org/10.1111/syen.12613



Shade Singers (forest)
"Westlandica Group" (forest/shrub/grass)
Rosea Group (dry scrub)
Green Foliage Group (forest/shrub)
Muta Group (grass)

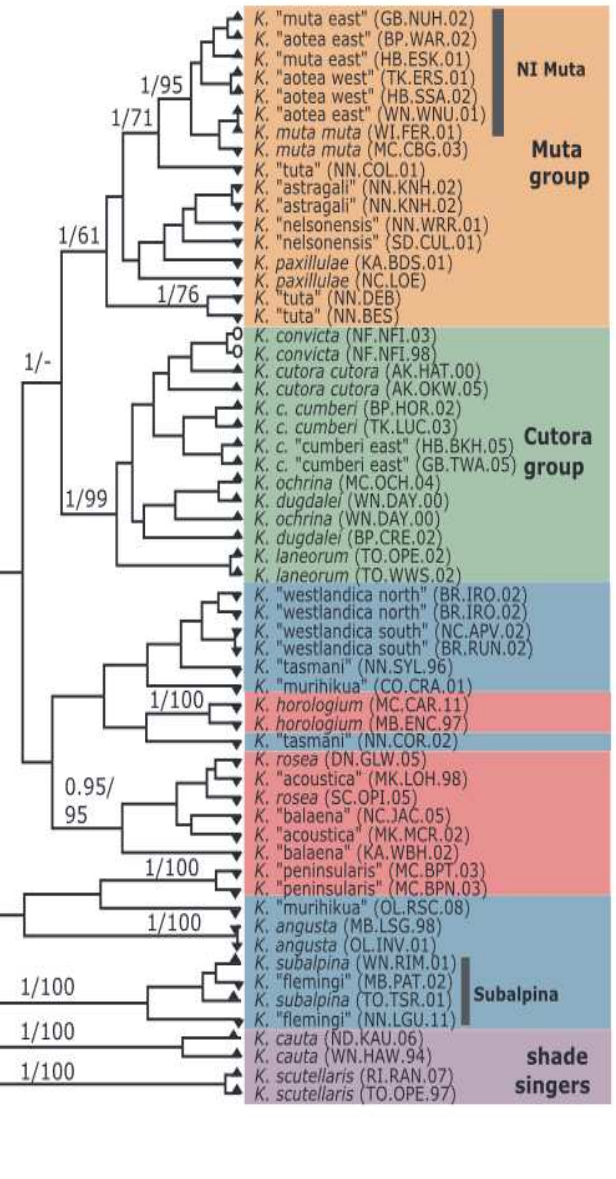
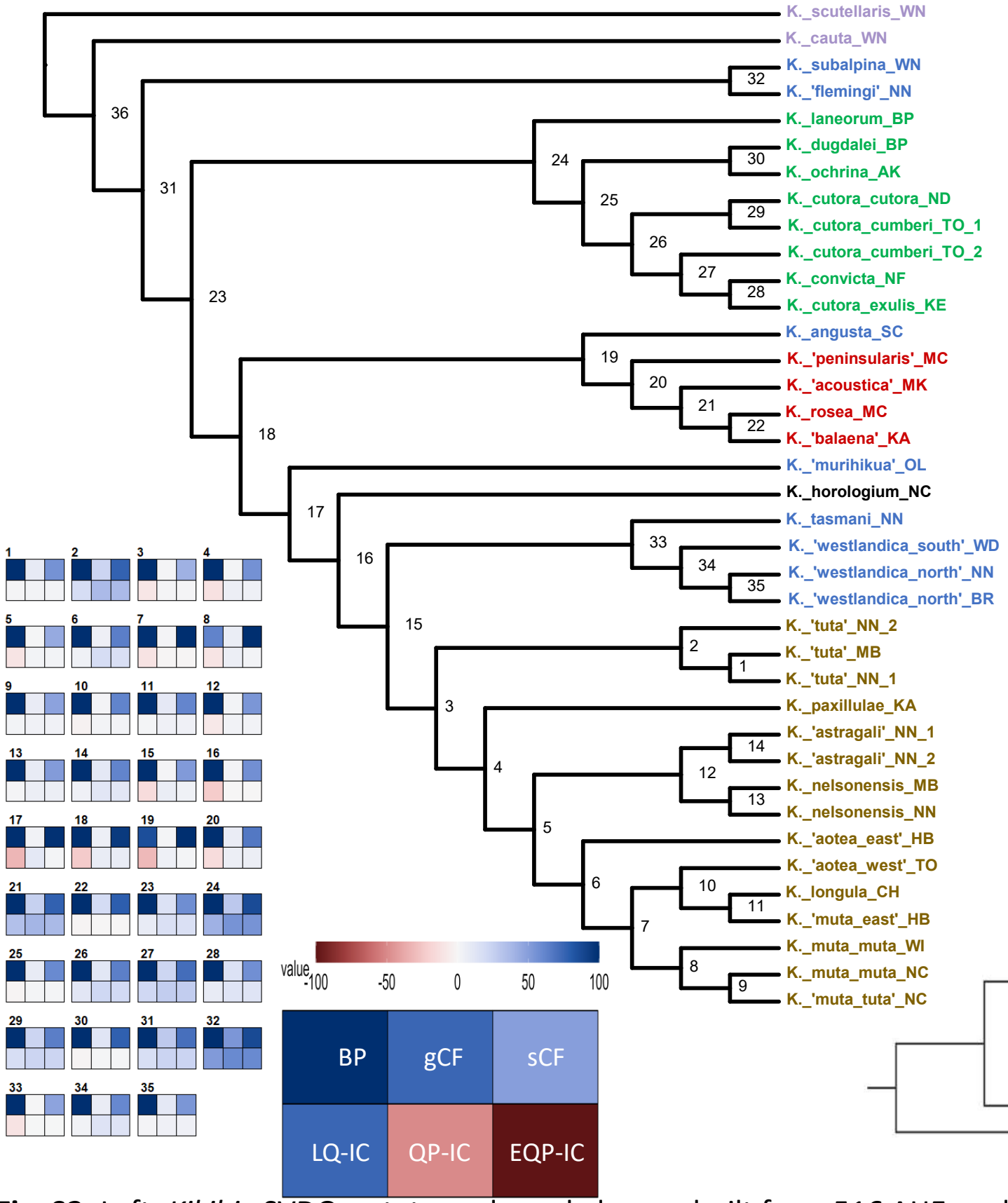


Fig. S1: Left: *Kikihia* concatenated ML nuclear phylogeny built from 516 AHE nuclear genes. Heatmaps for corresponding node numbers display SH-aLRT support, ultrafast bootstrap support (UFB), gene concordance factor (gCF), site concordance factor (sCF), lowest quartet internode certainty (LQ-IC), quadripartition internode certainty (QP-IC), and extended quadripartition internode certainty (EQP-IC). Branch lengths are in substitutions/site. Inset right: 5-gene nuclear phylogeny reproduced from Banker et al. (2017) for comparison. Shapes on branch tips refer to whether the specimen is from the North Island (upwards triangle), South Island (downwards triangle), or offshore island (circle). Refer to figure caption in original paper for details.



Shade Singers (forest)
"Westlandica Group" (forest/shrub/grass)
Rosea Group (dry scrub)
Green Foliage Group (forest/shrub)
Muta Group (grass)

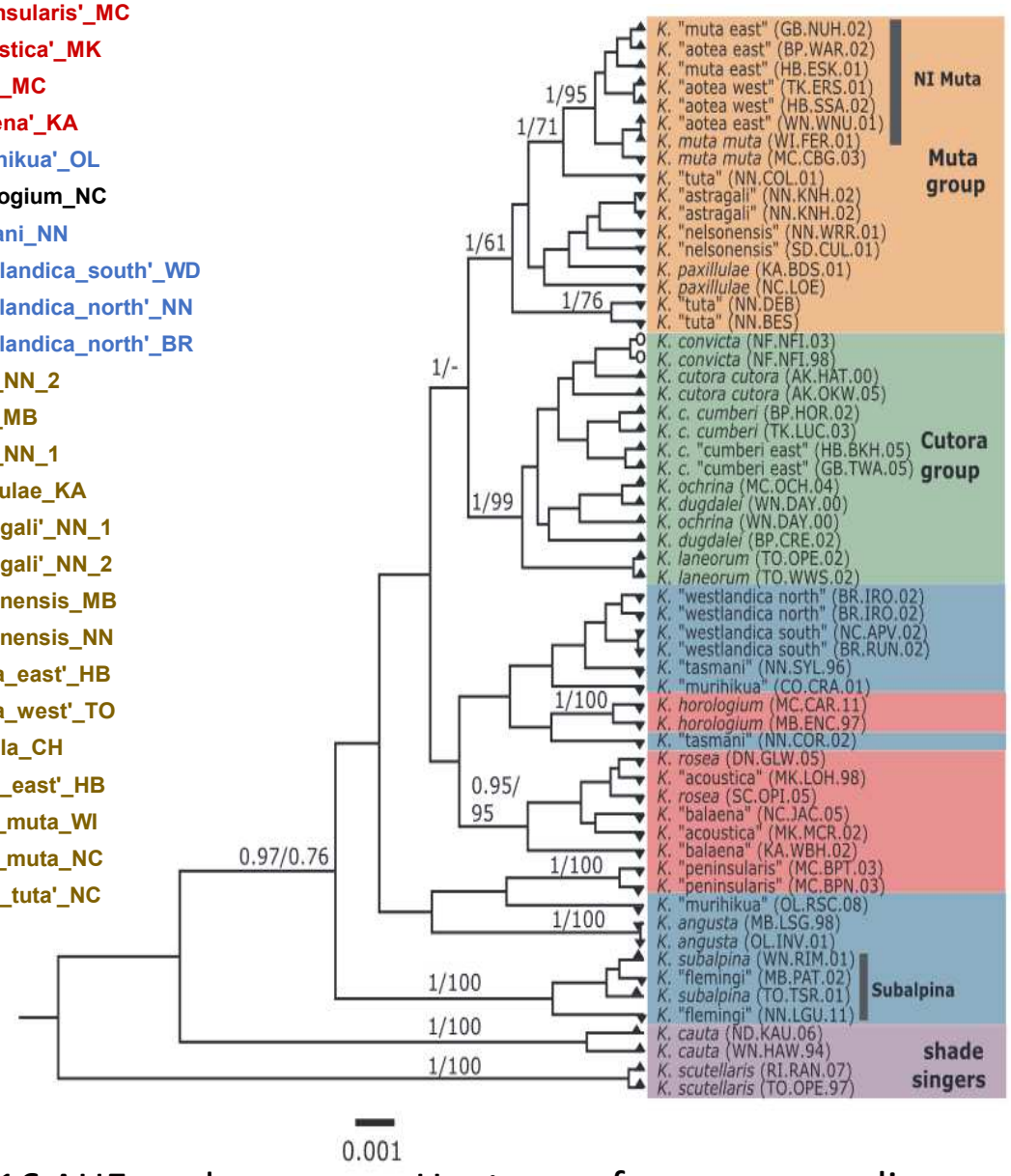
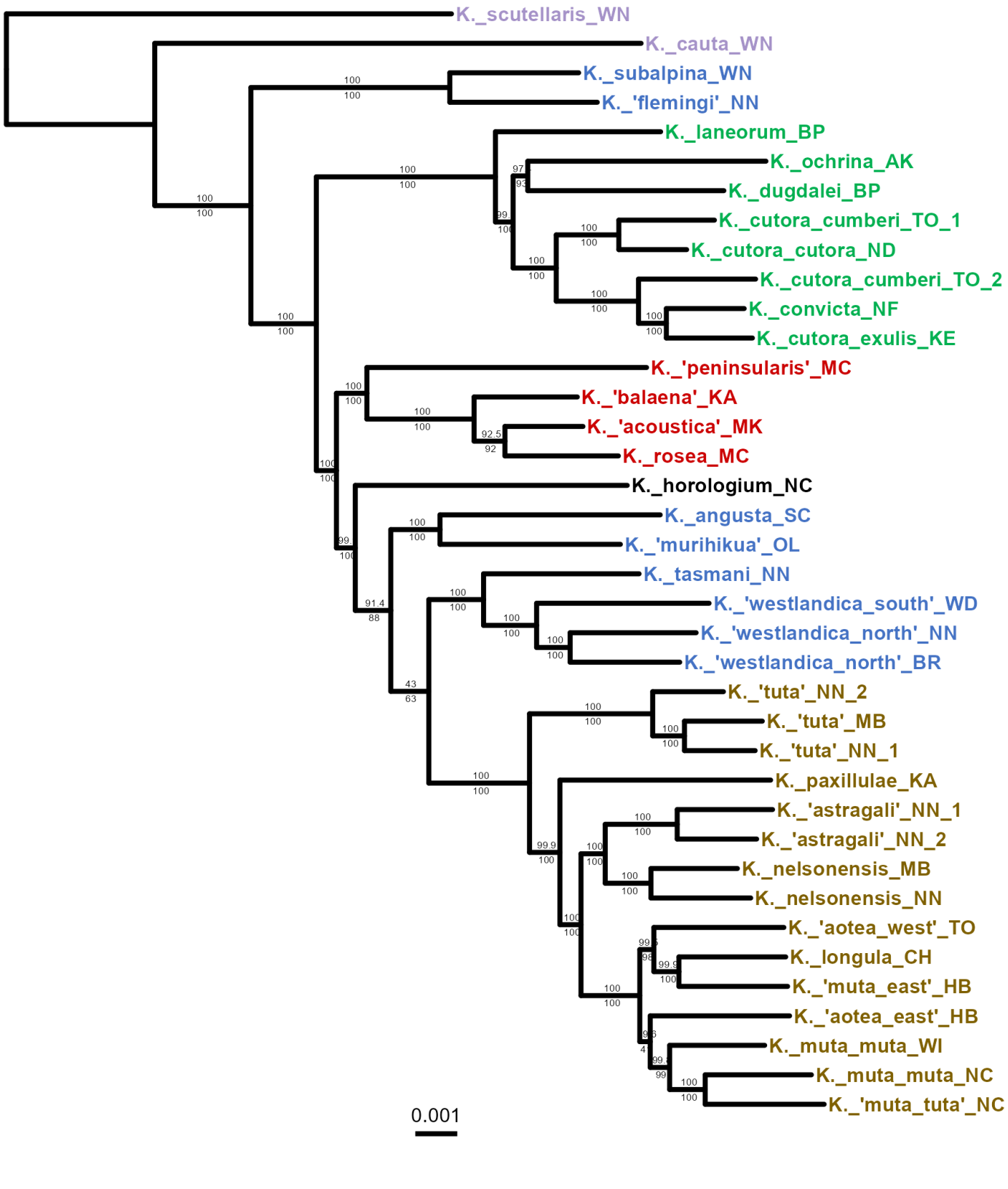


Fig. S2: Left: *Kikihia* SVDQuartets nuclear phylogeny built from 516 AHE nuclear genes. Heatmaps for corresponding node numbers display bootstrap support (BP), gene concordance factor (gCF), site concordance factor (sCF), lowest quartet internode certainty (LQ-IC), quadripartition internode certainty (QP-IC), and extended quadripartition internode certainty (EQP-IC). Branch lengths are unscaled. Inset right: 5-gene nuclear phylogeny reproduced from Banker et al. (2017) for comparison. Shapes on branch tips refer to whether the specimen is from the North Island (upwards triangle), South Island (downwards triangle), or offshore island (circle). Refer to figure caption in original paper for details.



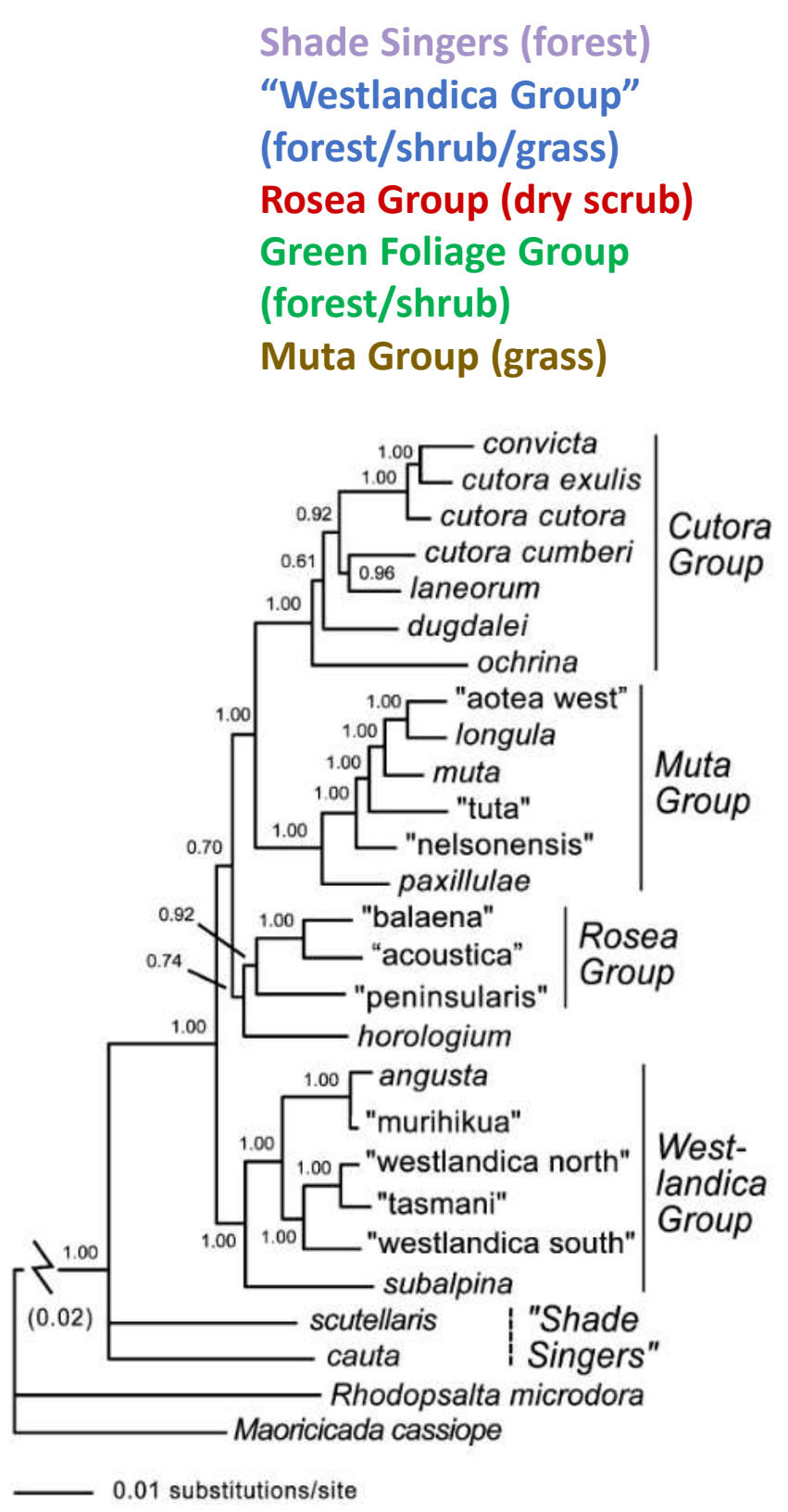


Fig. S3: Left: *Kikihia* concatenated ML nuclear+mtDNA phylogeny built from mitochondrial genomes and 516 nuclear genes. Upper support values are SH-aLRT values; support values underneath are ultrafast bootstrap percentages (UFB). Branch lengths are in substitutions/site. Inset right: phylogeny combining 2152 bp of mtDNA sequence and 1545 bp EF-1 alpha nuclear sequence reproduced from Marshall et al. (2008) for comparison.

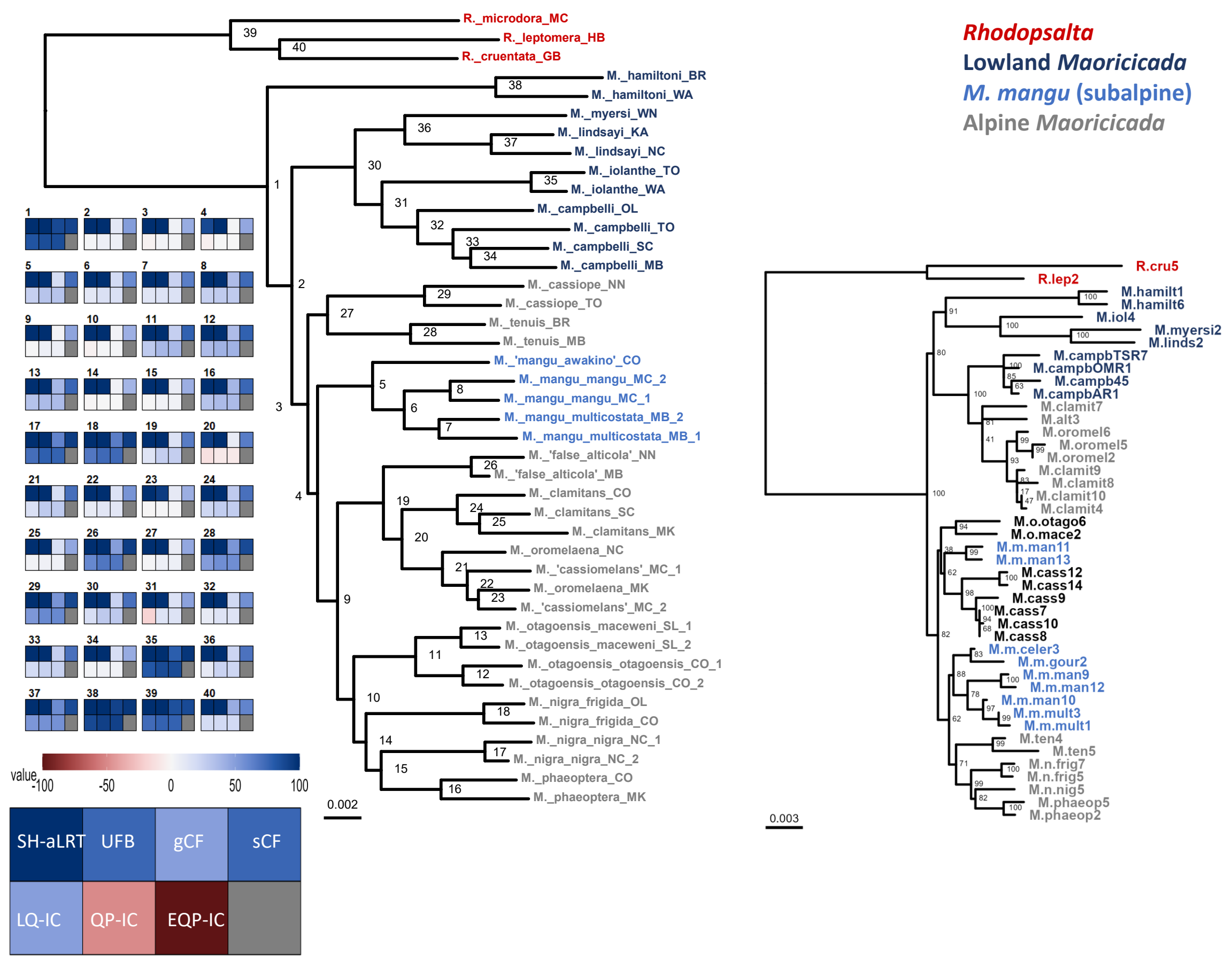
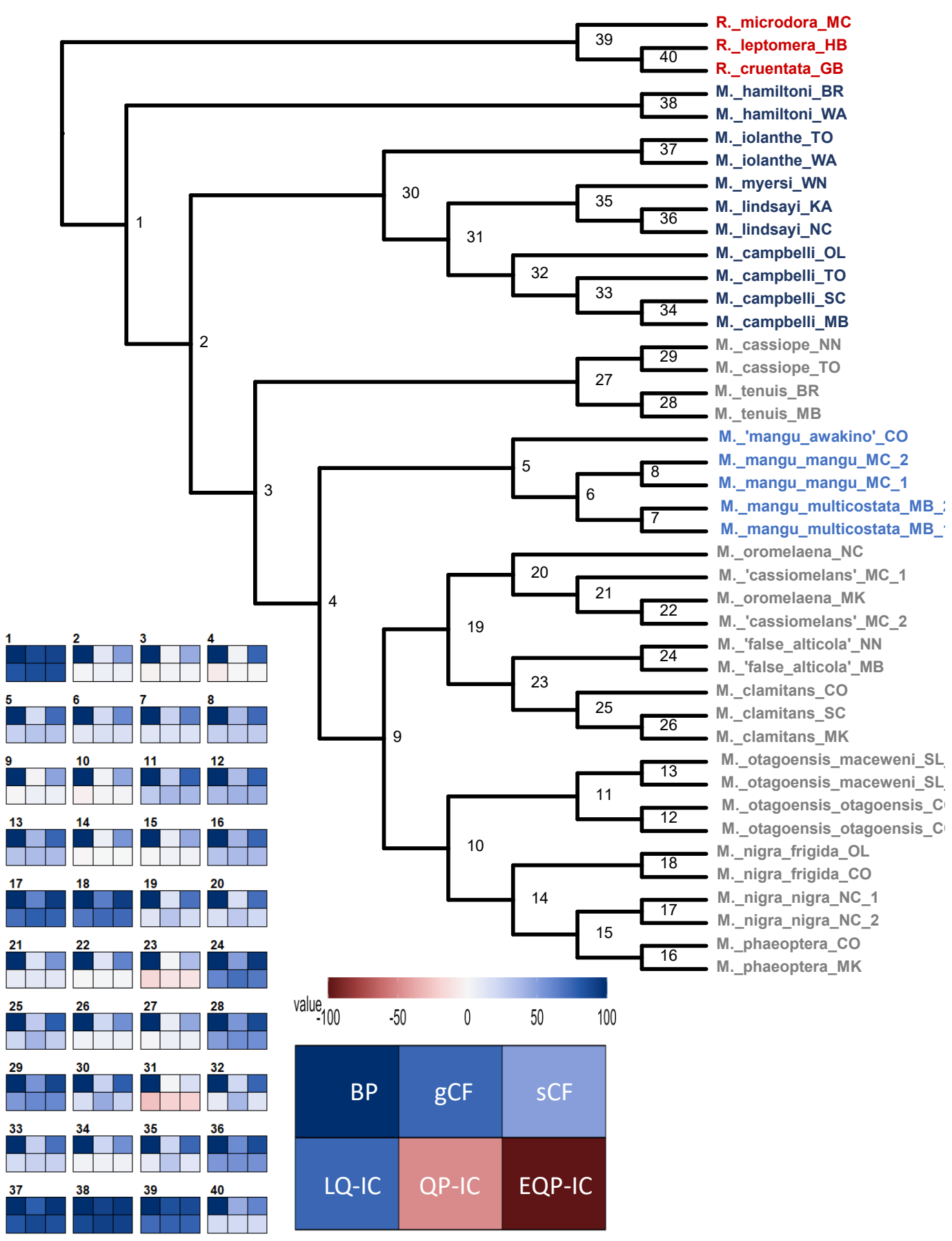


Fig. S4: Left: *Maoricicada* concatenated ML nuclear phylogeny built from 505 AHE nuclear genes. Heatmaps for corresponding node numbers display SH-aLRT support, ultrafast bootstrap support (UFB), gene concordance factor (gCF), site concordance factor (sCF), lowest quartet internode certainty (LQ-IC), quadripartition internode certainty (QP-IC), and extended quadripartition internode certainty (EQP-IC). Branch lengths are in substitutions/site. Inset right: 3-gene nuclear phylogeny using sequence data from Buckley et al. (2006) for comparison.



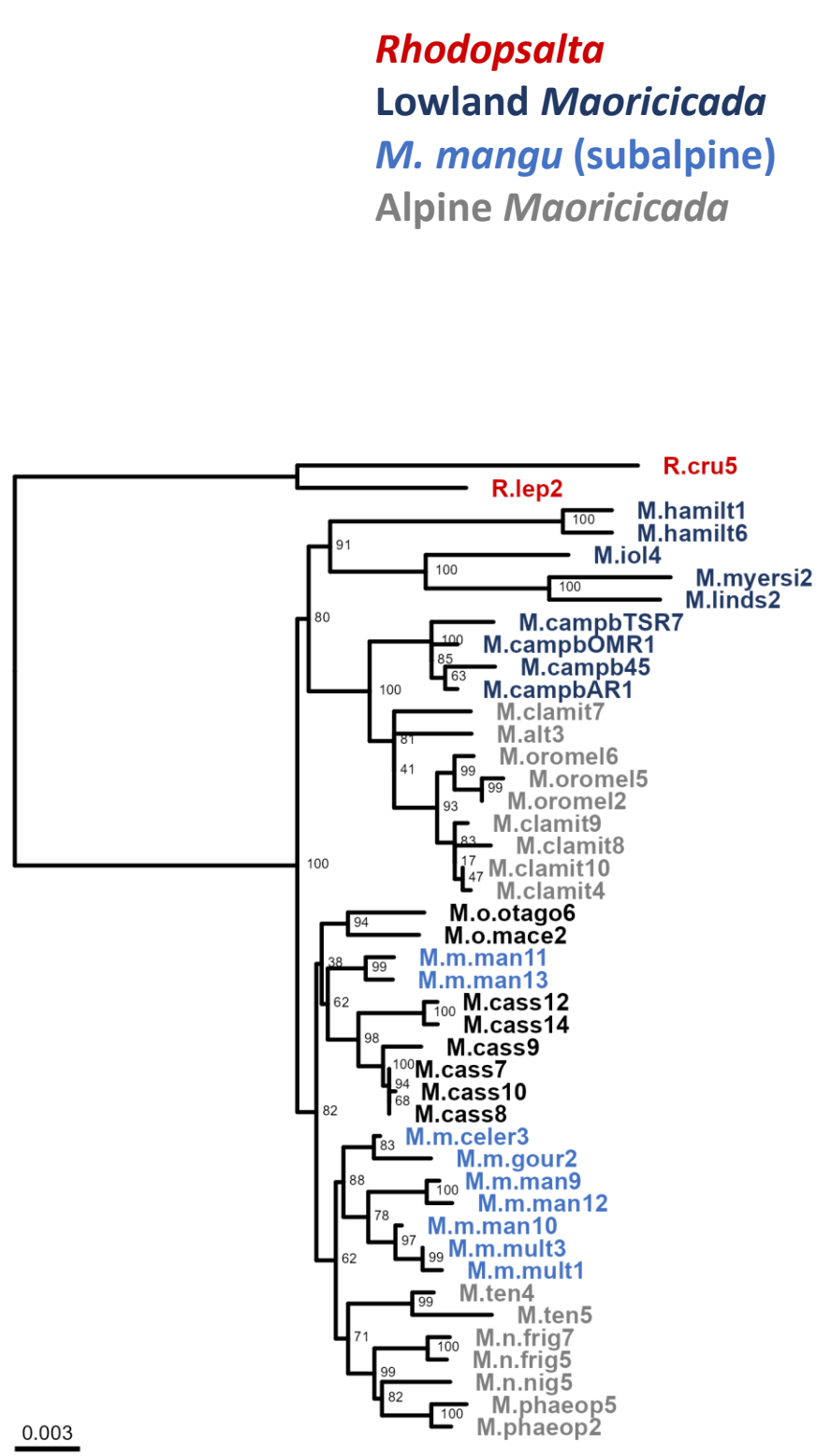
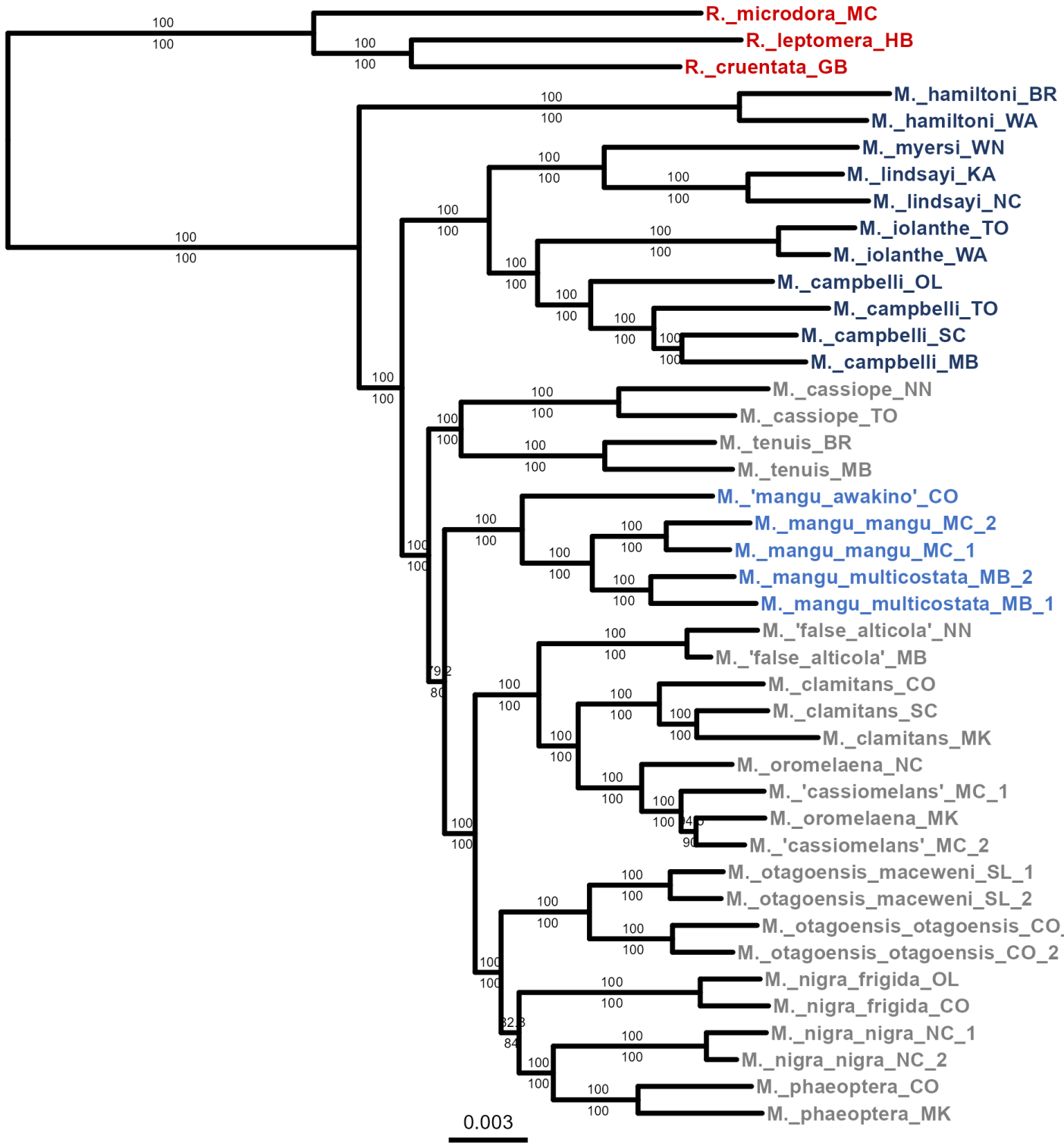


Fig. S5: Left: *Maoricicada* SVDQuartets nuclear phylogeny built from 505 AHE nuclear genes. Heatmaps for corresponding node numbers display bootstrap support (BP), gene concordance factor (gCF), site concordance factor (sCF), lowest quartet internode certainty (LQ-IC), quadripartition internode certainty (QP-IC), and extended quadripartition internode certainty (EQP-IC). Branch lengths are unscaled. Inset right: 3-gene nuclear phylogeny using sequence data from Buckley et al. (2006) for comparison.



Rhodopsalta

Lowland ***Maoricicada***

M. mangu **(subalpine)**

Alpine ***Maoricicada***

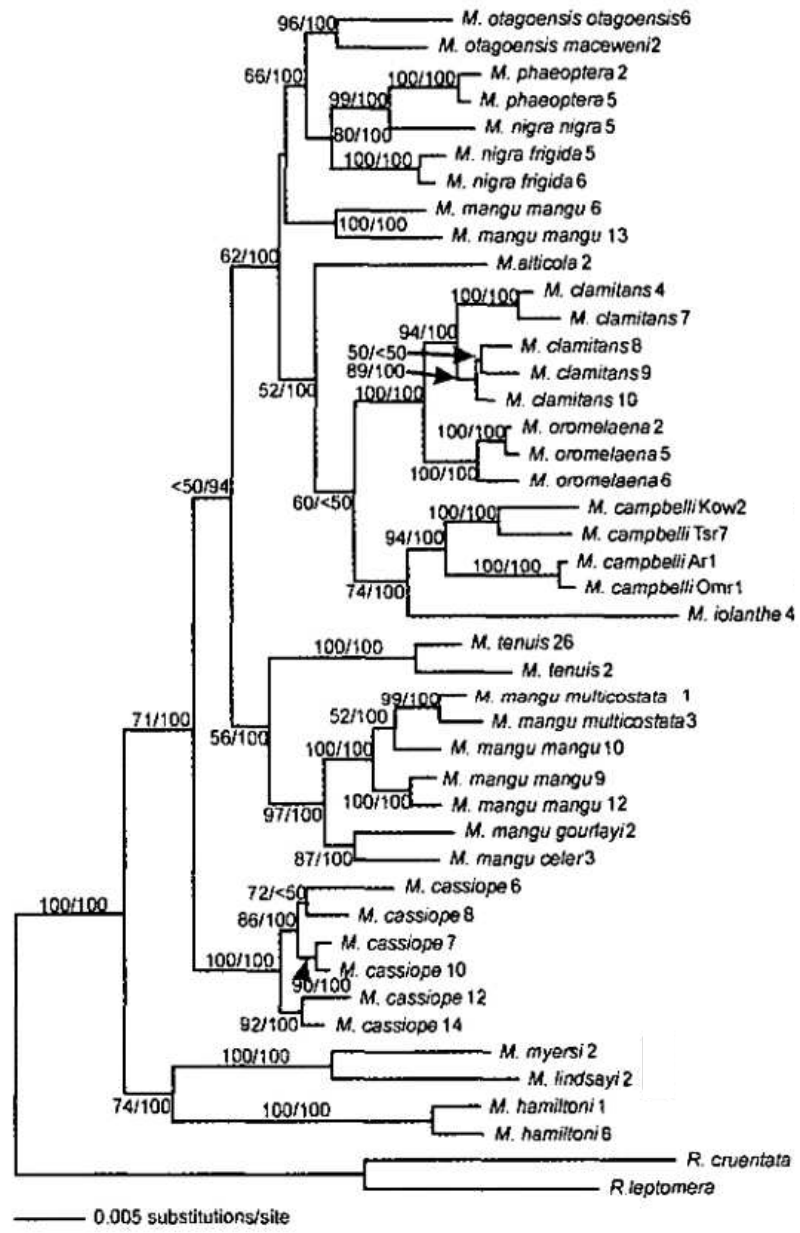


Fig. S6: Left: *Maoricicada* concatenated ML nuclear+mtDNA phylogeny built from mitochondrial genomes and 505 AHE nuclear genes. Upper supports are SH-aLRT values; supports underneath are ultrafast bootstrap percentages (UFB). Branch lengths are in substitutions/site. Inset right: phylogeny combining 3 nuclear genes and 1 mitochondrial gene reproduced from Buckley et al. (2006) for comparison.

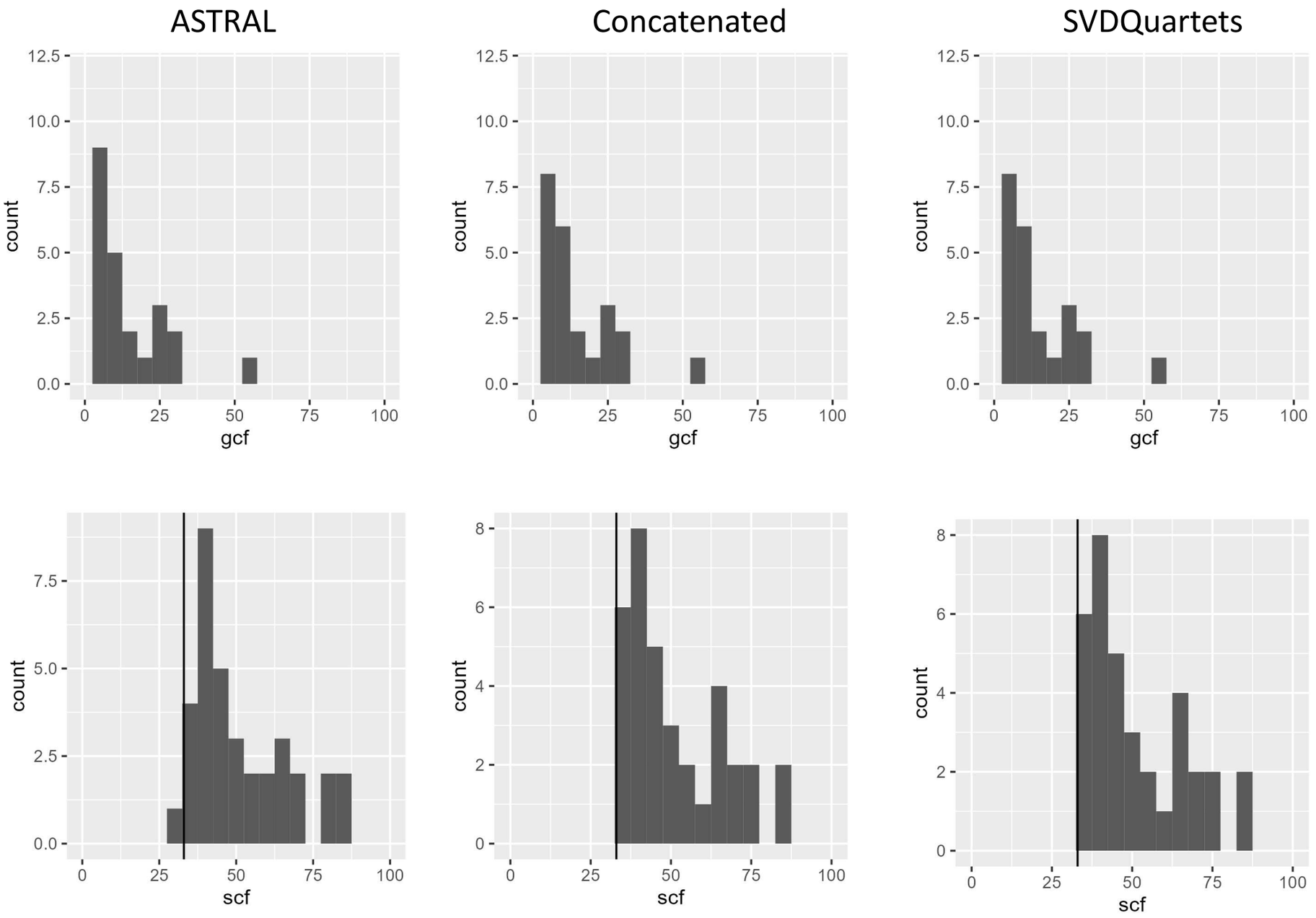


Fig. S7: Histograms of gene concordance factors (gCF, top) and site concordance factors (sCF, bottom) for branches in *Kikihia* ASTRAL (left), Concatenated (center), and SVDQuartets (right) nuclear phylogenies. Vertical bar in sCF histograms represents 33% sCF, which indicates equal support for all three quartet arrangements.

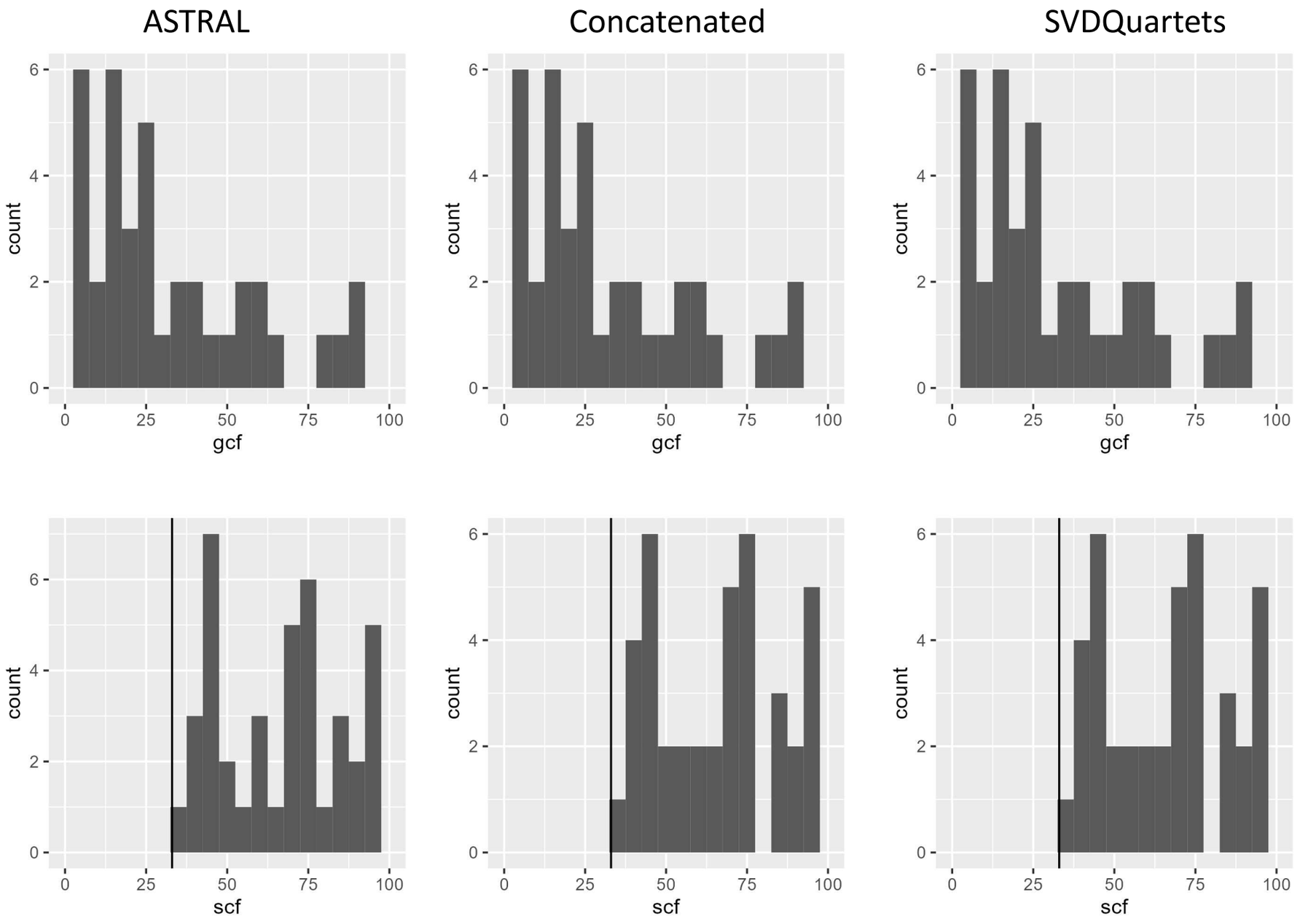


Fig. S8: Histograms of gene concordance factors (gCF, top) and site concordance factors (sCF, bottom) for branches in *Maoricicada* ASTRAL (left), Concatenated (center), and SVDQuartets (right) nuclear phylogenies. Vertical bar in sCF histograms represents 33% sCF, which indicates equal support for all three quartet arrangements.

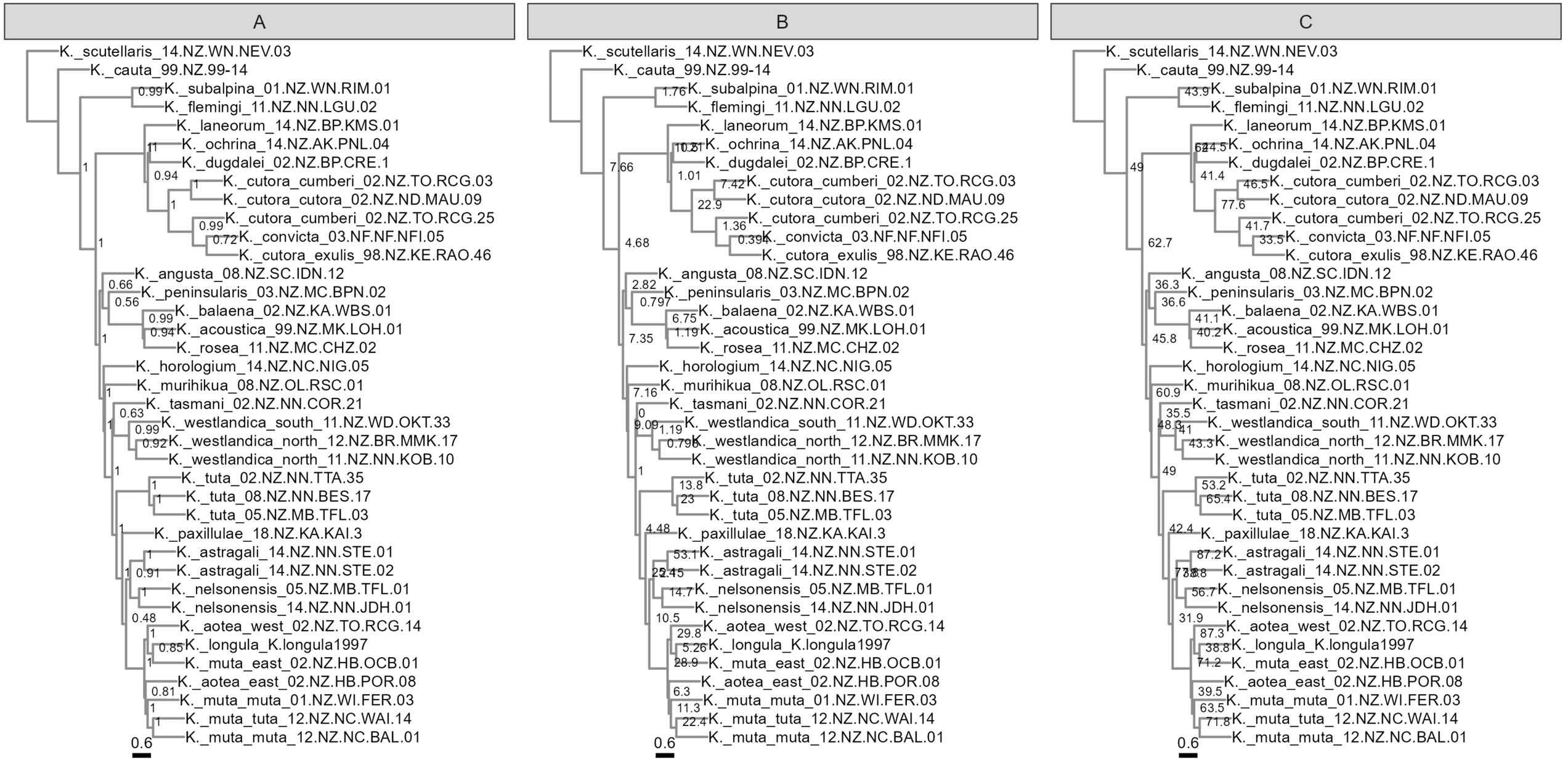


Fig. S9: *Kikihia* ASTRAL phylogeny displaying A: local posterior probability, B: gene concordance factors, and C: site concordance factors. Branch lengths are in coalescent units. Terminal branches are unscaled.

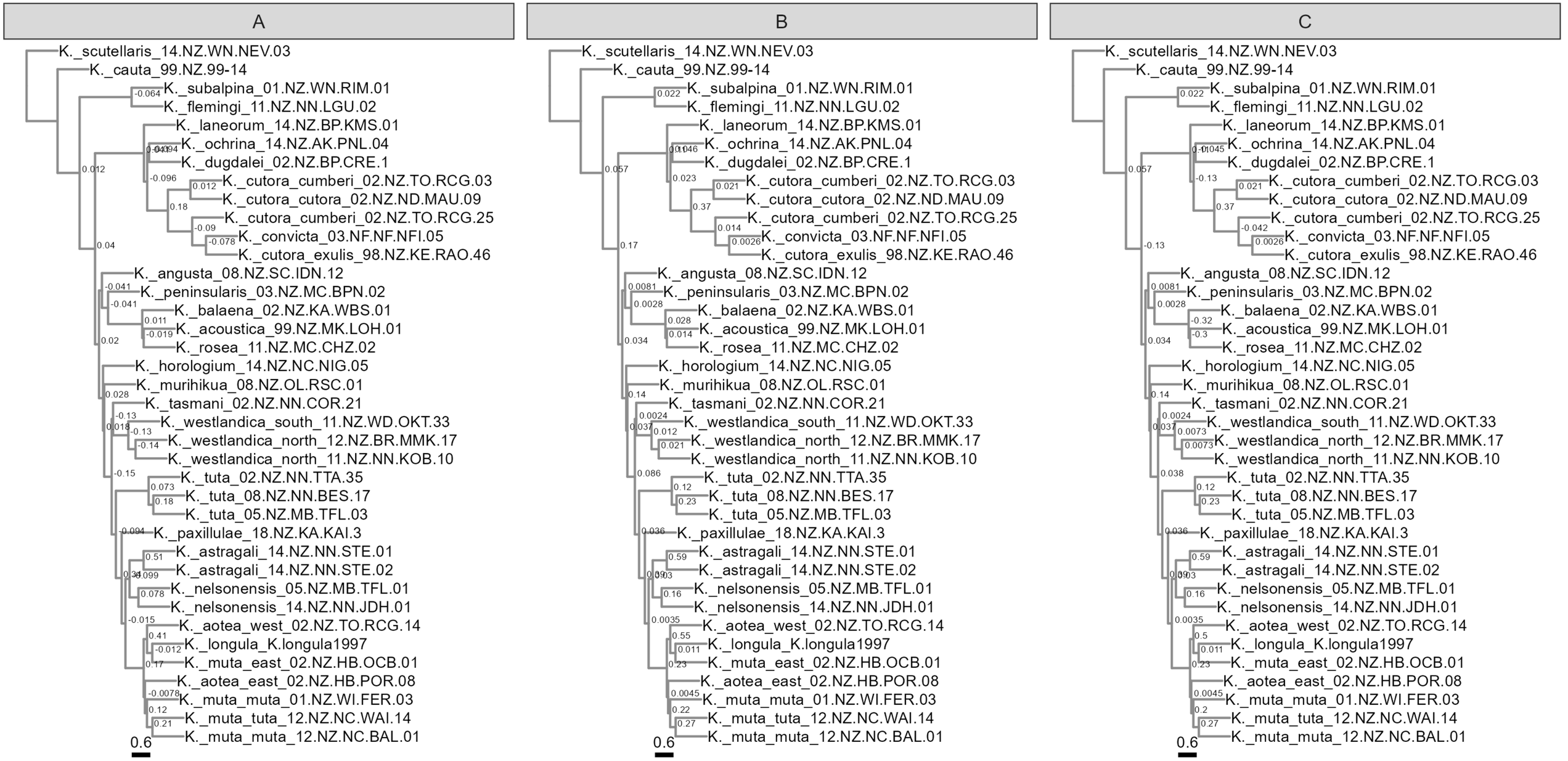


Fig. S10: *Kikihia* ASTRAL phylogeny displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC). Branch lengths are in coalescent units. Terminal branches are unscaled.

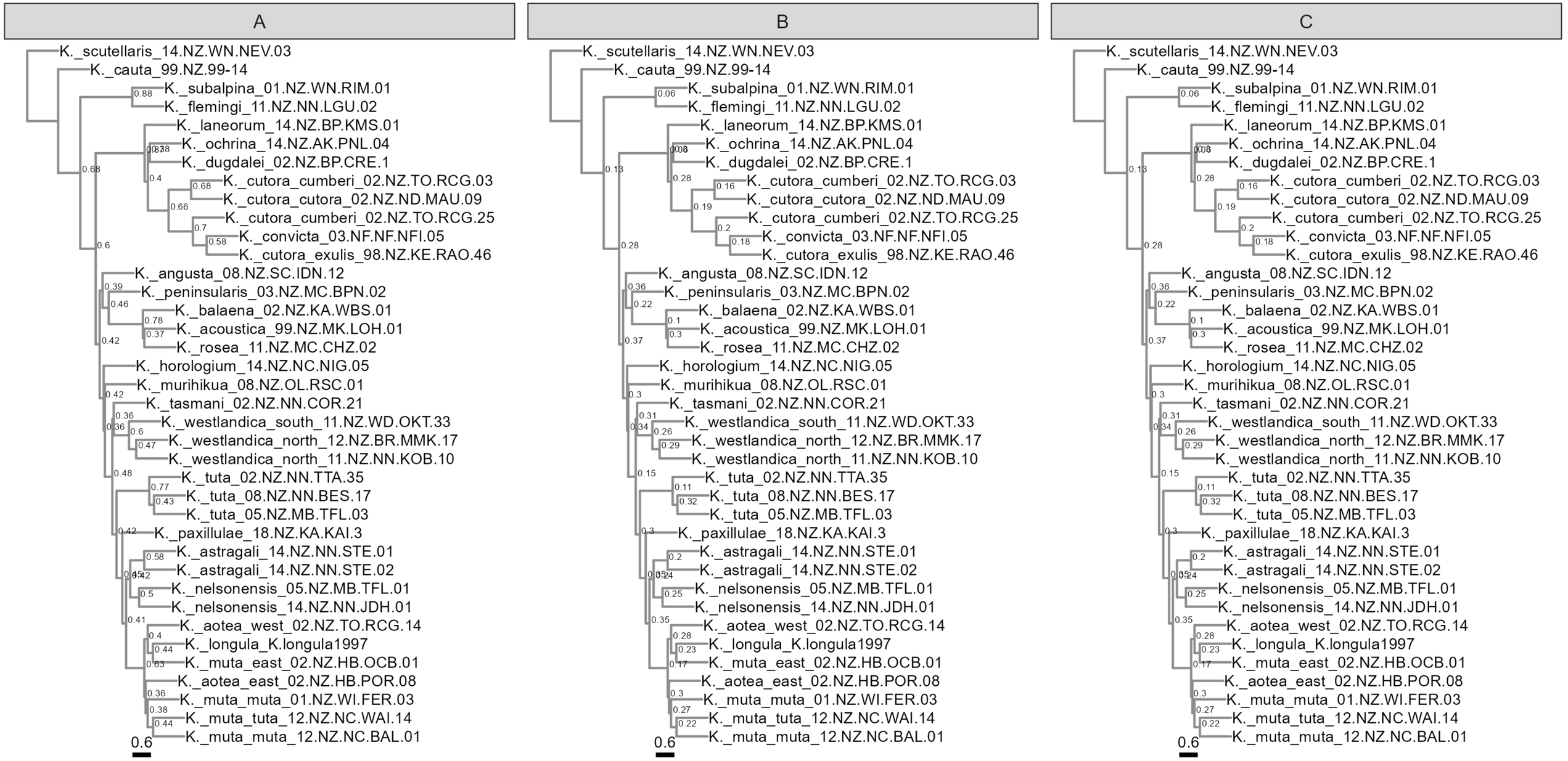


Fig. S11: *Kikihia* ASTRAL phylogeny displaying A: the highest branch quartet support frequency, B: the first alternative branch quartet support frequency, and C: the second alternative branch quartet support frequency. Branch lengths are in coalescent units. Terminal branches are unscaled.

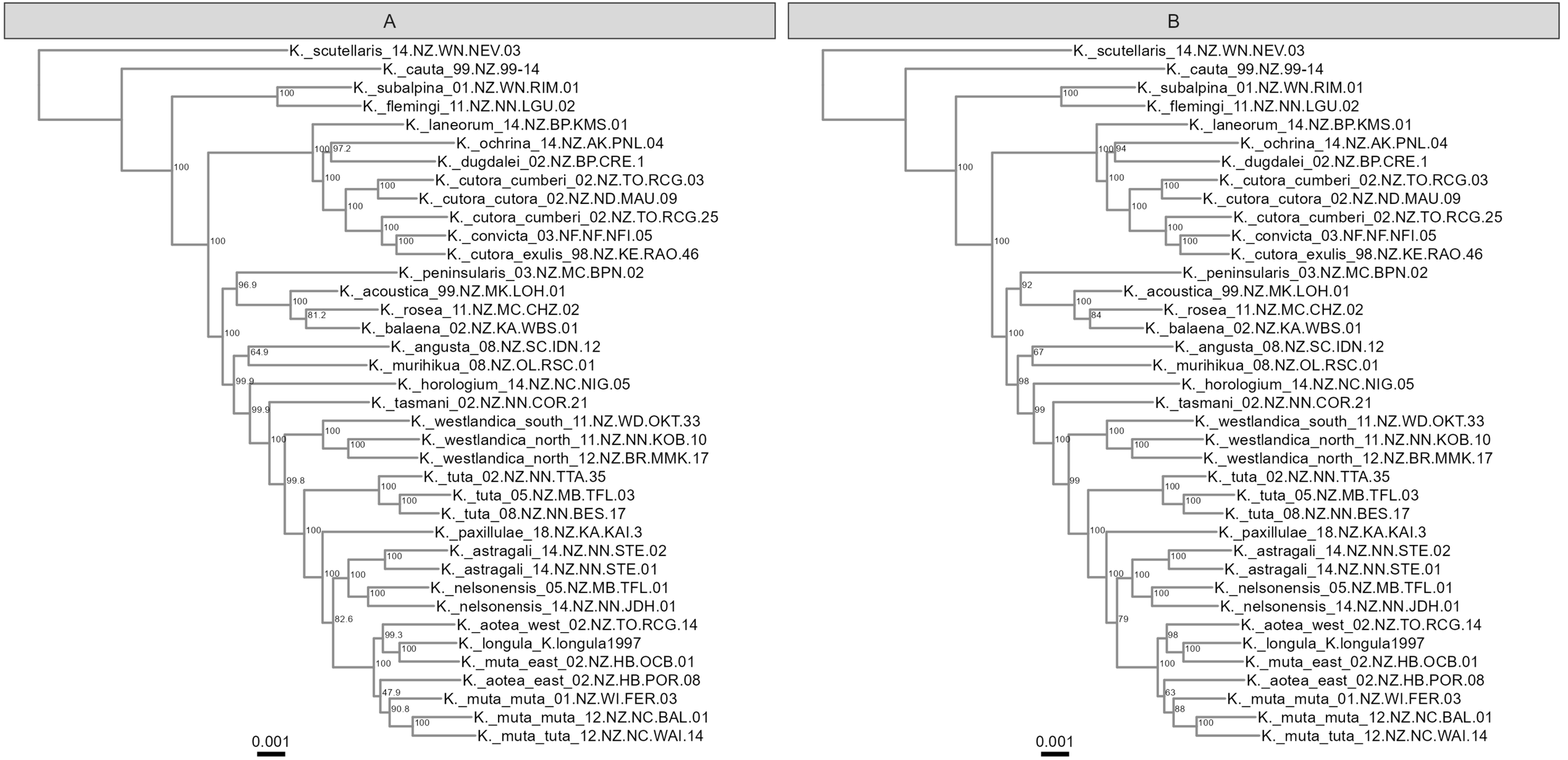


Fig. S12: *Kikihia* concatenated ML phylogeny displaying A: SH-aLRT branch supports, and B: Ultrafast bootstrap branch supports (UFB). Branch lengths are in substitutions/site.

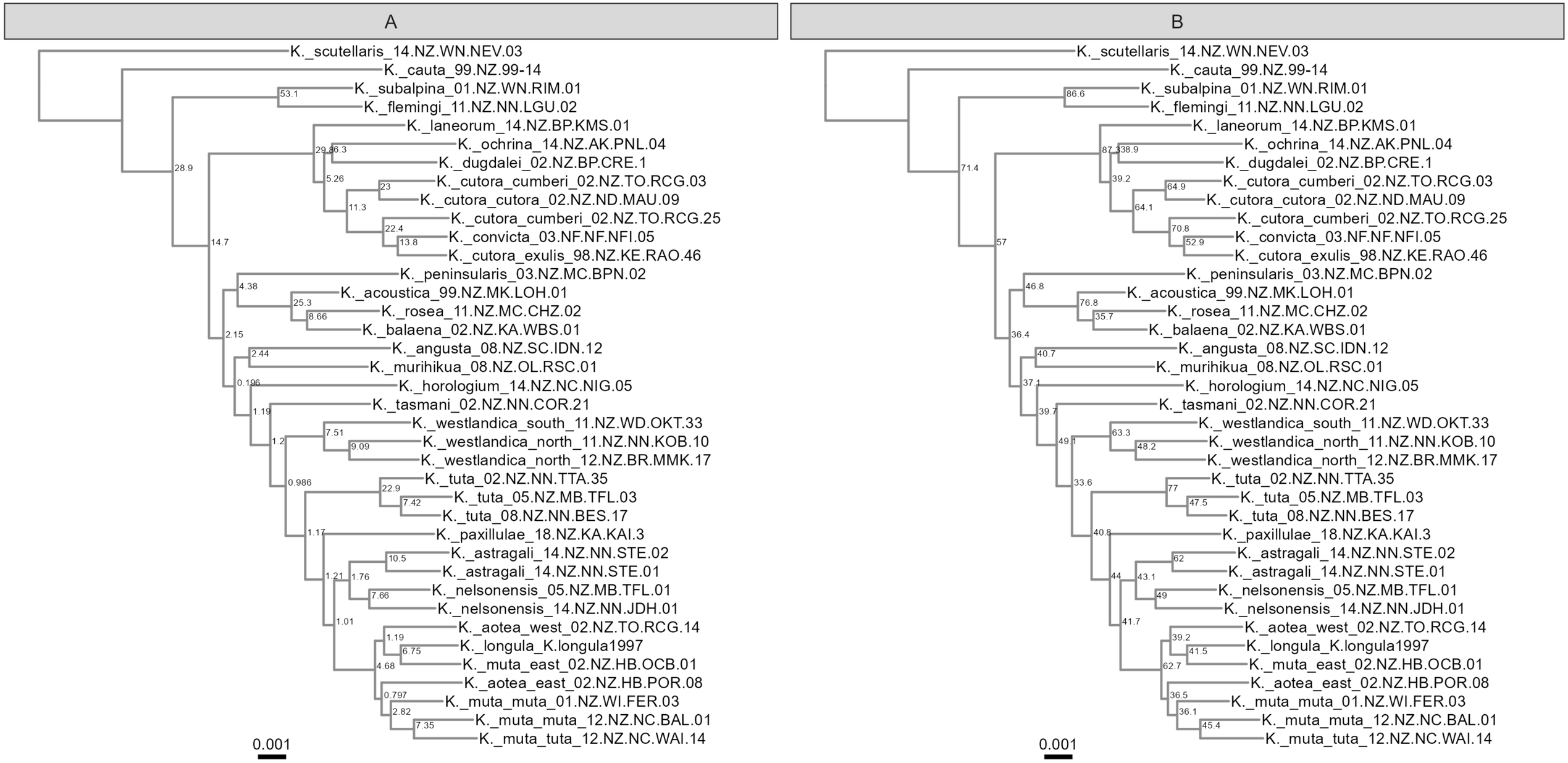


Fig. S13: *Kikihia* concatenated ML phylogeny displaying A: gene concordance factors, and B: site concordance factors. Branch lengths are in substitutions/site.

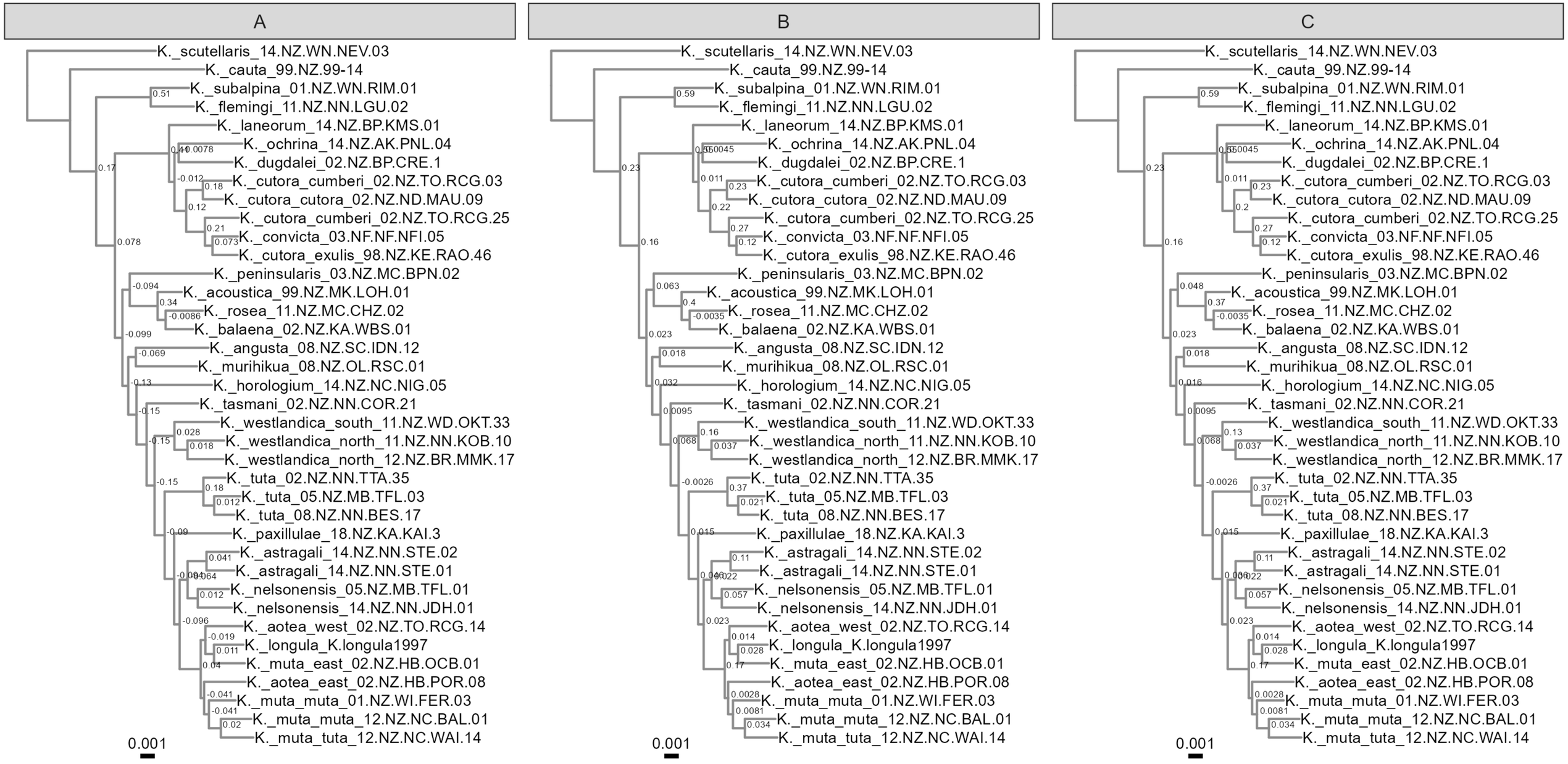


Fig. S14: *Kikihia* concatenated ML phylogeny displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC). Branch lengths are in substitutions/site.

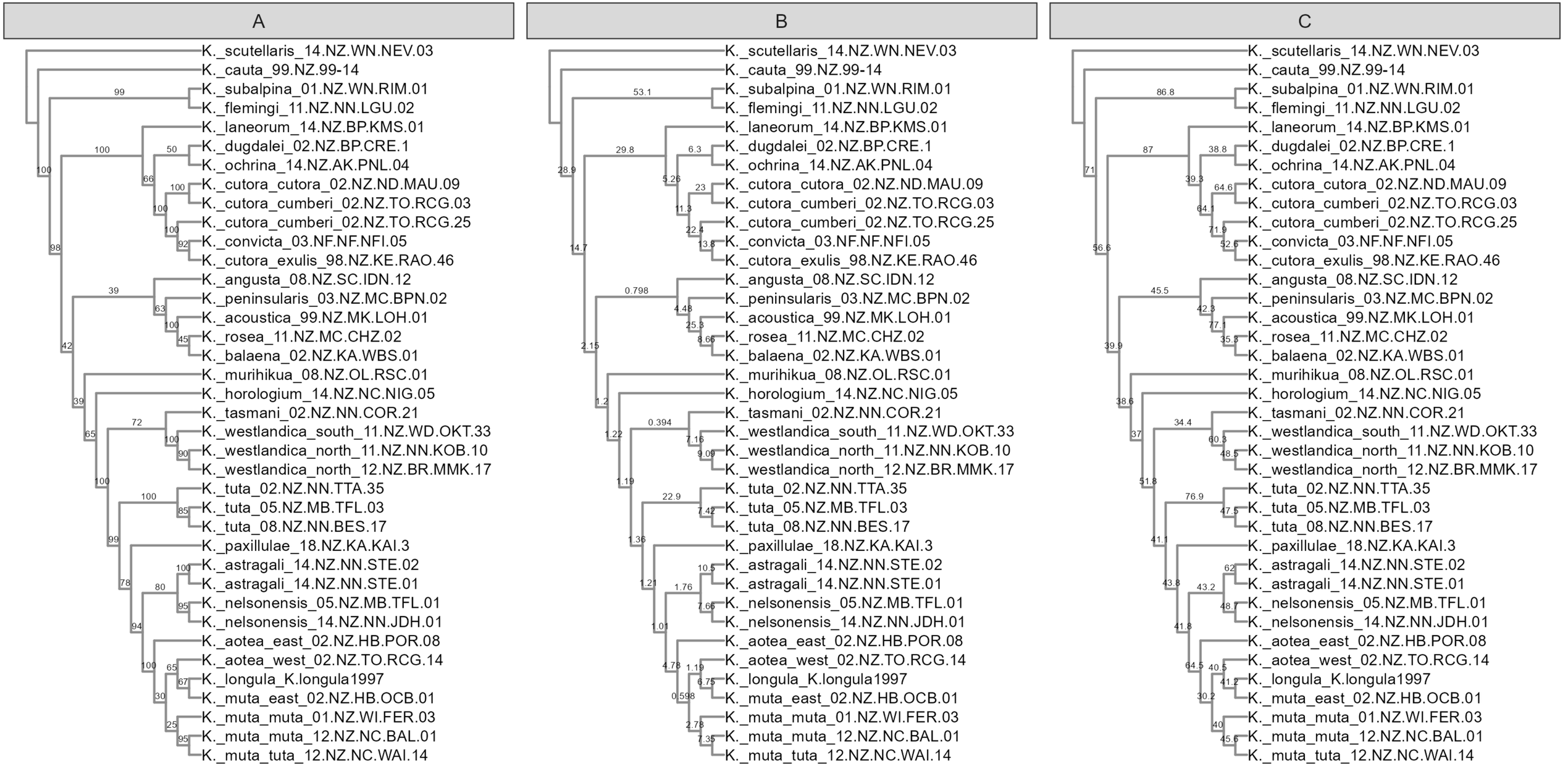


Fig. S15: *Kikihia* SVDQuartets cladogram displaying A: bootstrap percentage, B: gene concordance factors, and C: site concordance factors.

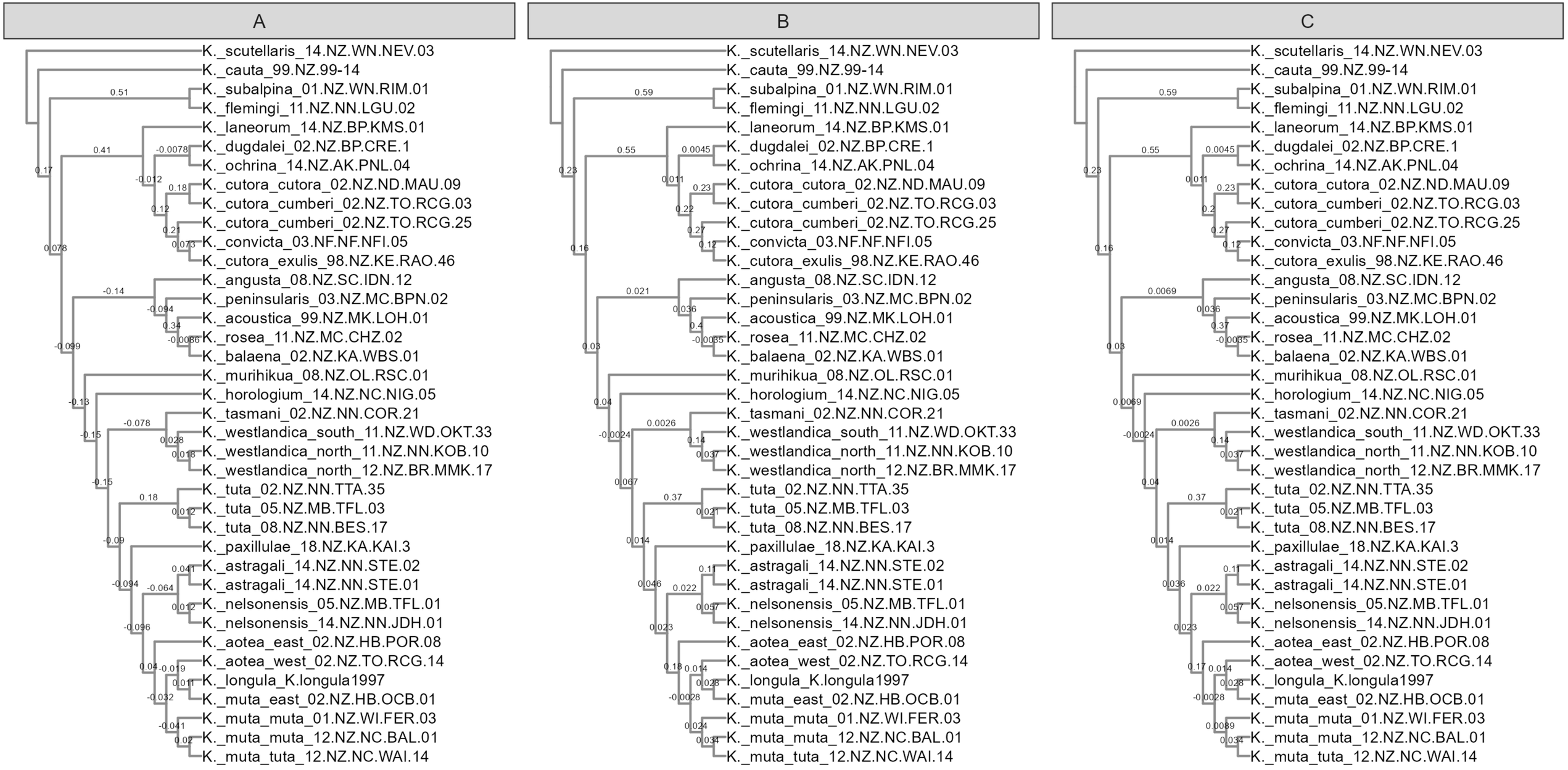


Fig. S16: *Kikihia* SVDQuartets cladogram displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC).

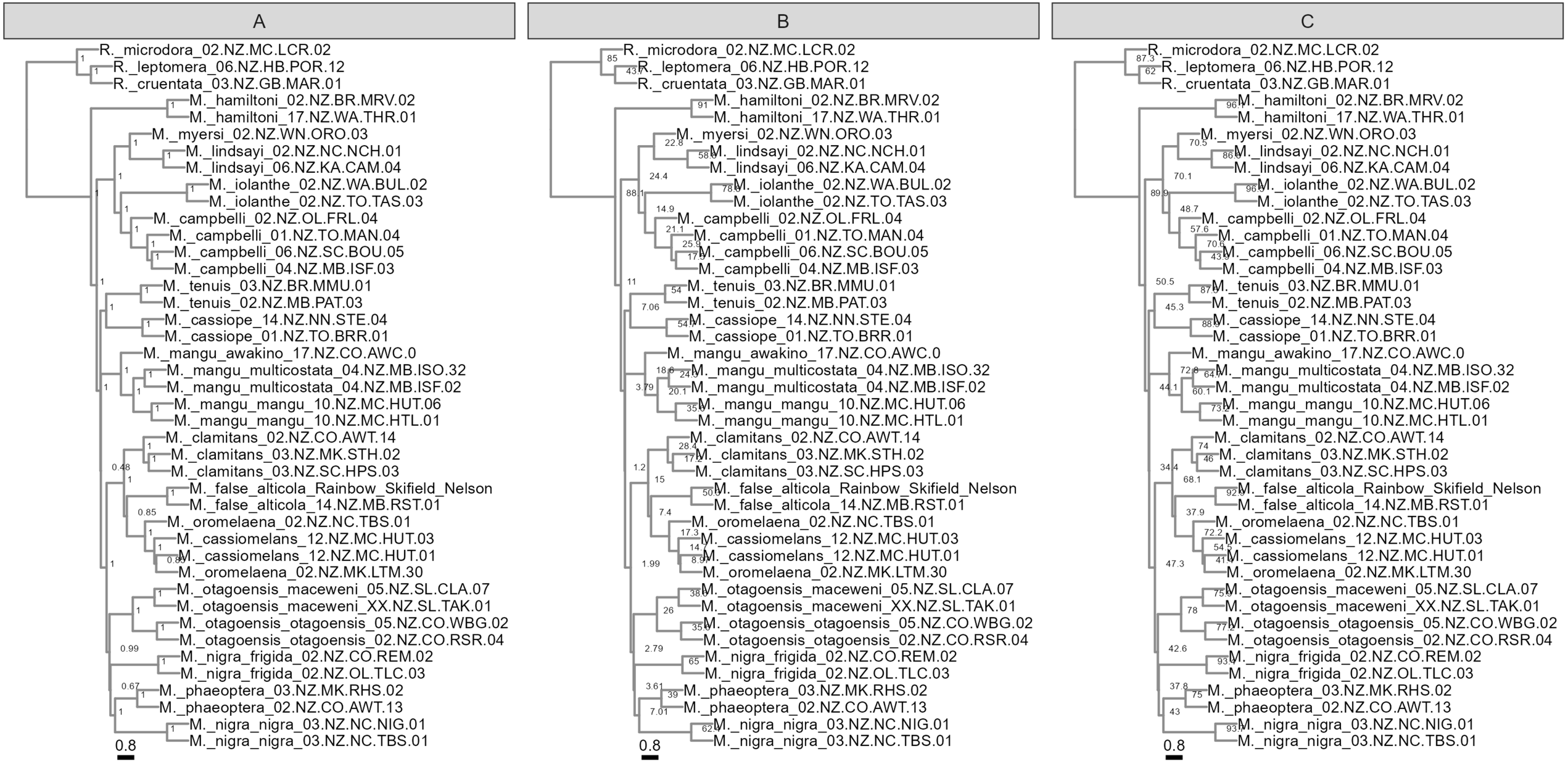


Fig. S17: *Maoricicada* ASTRAL phylogeny displaying A: local posterior probability, B: gene concordance factors, and C: site concordance factors. Branch lengths are in coalescent units. Terminal branches are unscaled.

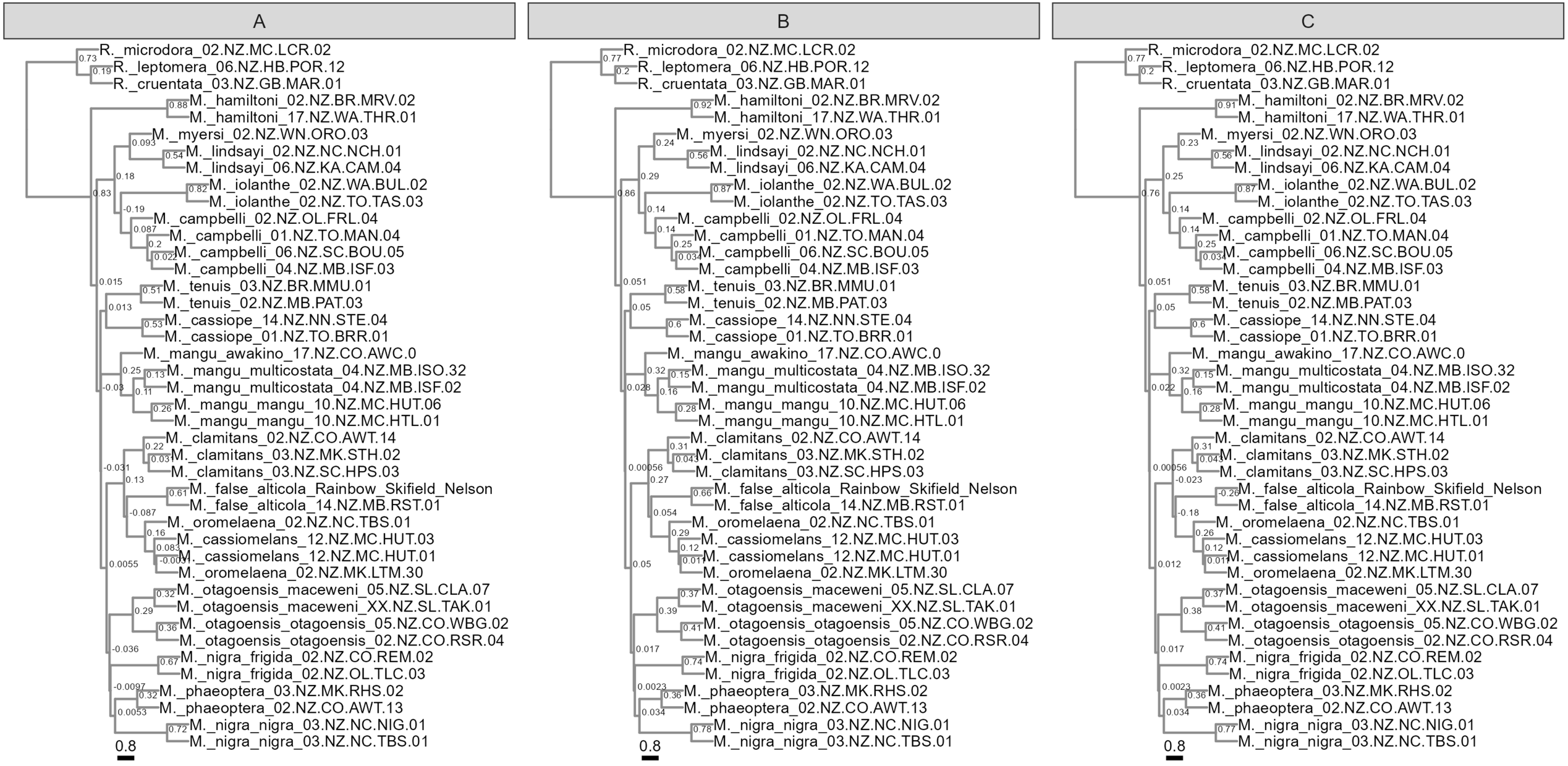


Fig. S18: *Maoricicada* ASTRAL phylogeny displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC). Branch lengths are in coalescent units. Terminal branches are unscaled.

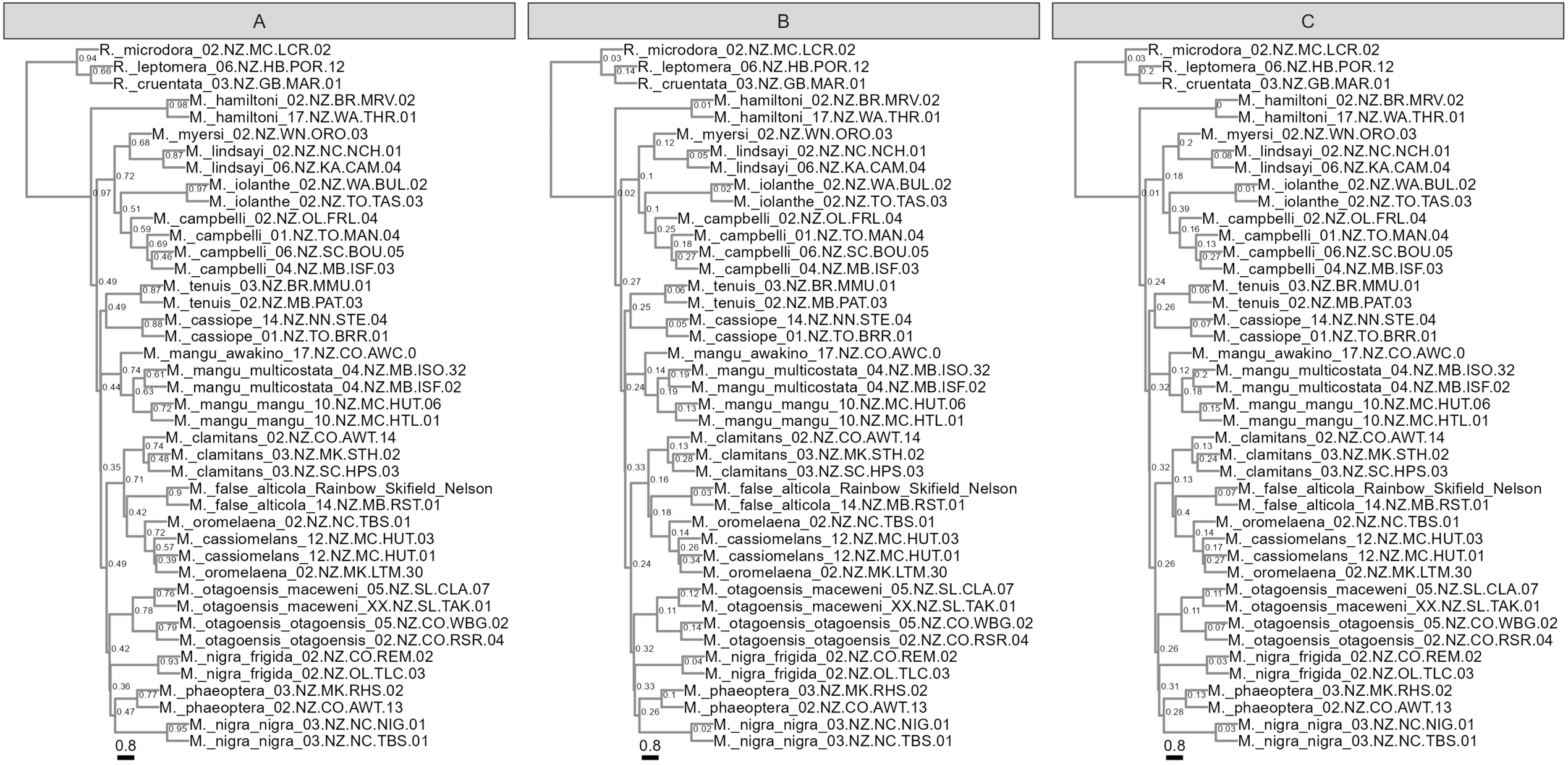


Fig. S19: *Maoricicada* ASTRAL phylogeny displaying A: the highest branch quartet support frequency, B: the first alternative branch quartet support frequency, and C: the second alternative branch quartet support frequency. Branch lengths are in coalescent units. Terminal branches are unscaled.

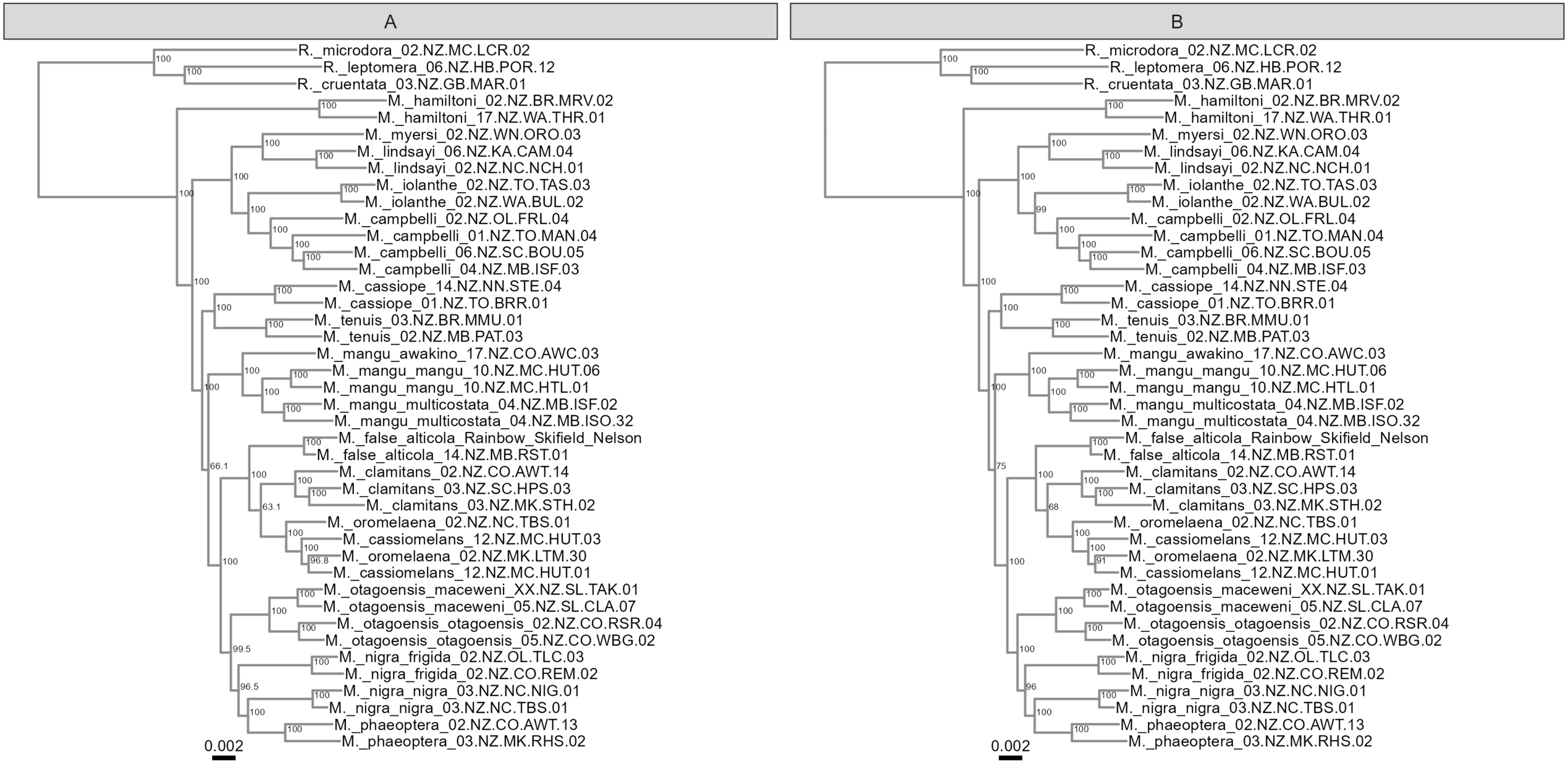


Fig. S20: *Maoricicada* concatenated ML phylogeny displaying A: SH-aLRT branch supports, and B: Ultrafast bootstrap branch supports (UFB). Branch lengths are in substitutions/site.

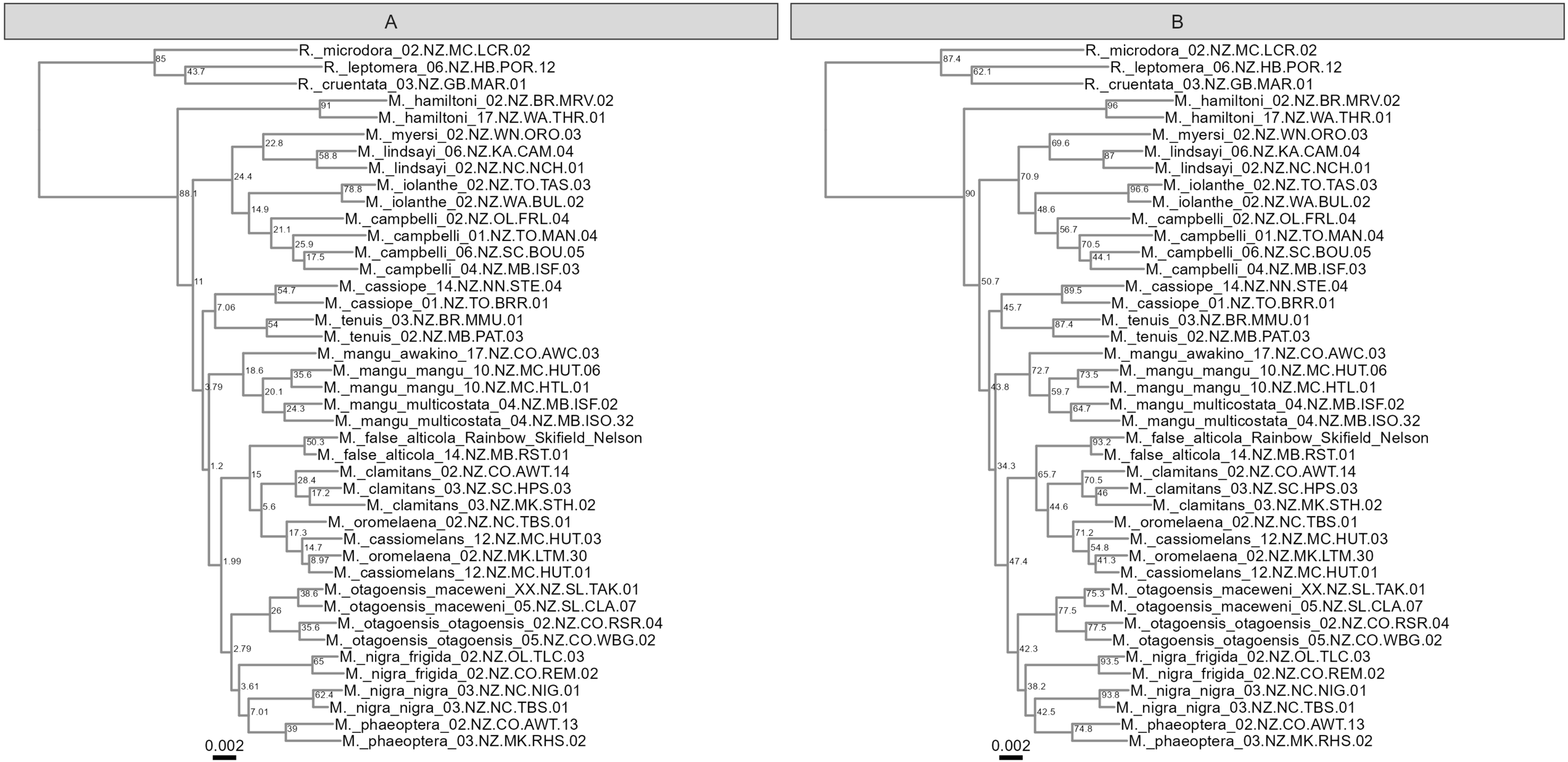


Fig. S21: *Maoricicada* concatenated ML phylogeny displaying A: gene concordance factors, and B: site concordance factors. Branch lengths are in substitutions/site.

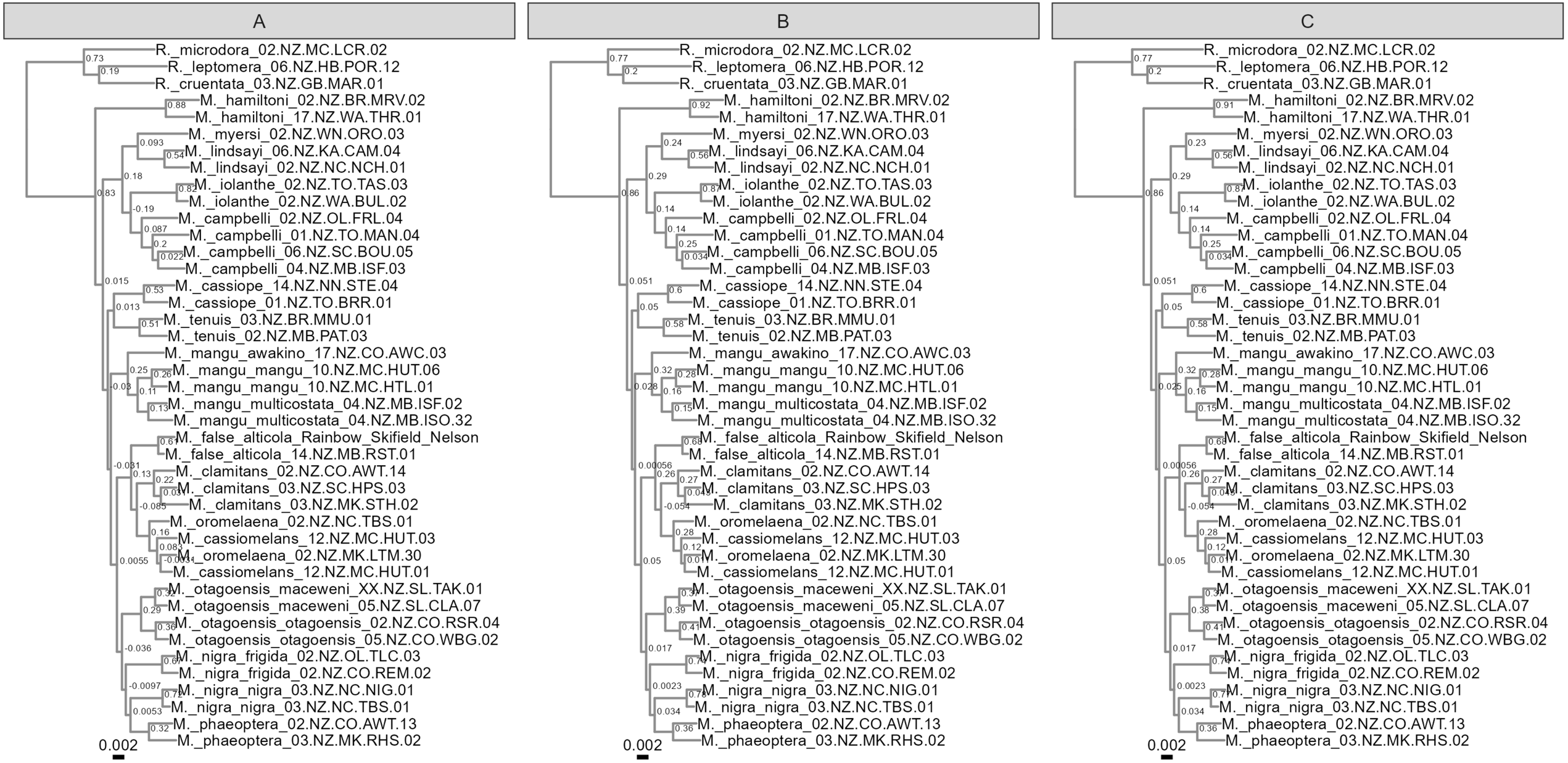


Fig. S22: *Maoricicada* concatenated ML phylogeny displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC). Branch lengths are in substitutions/site.

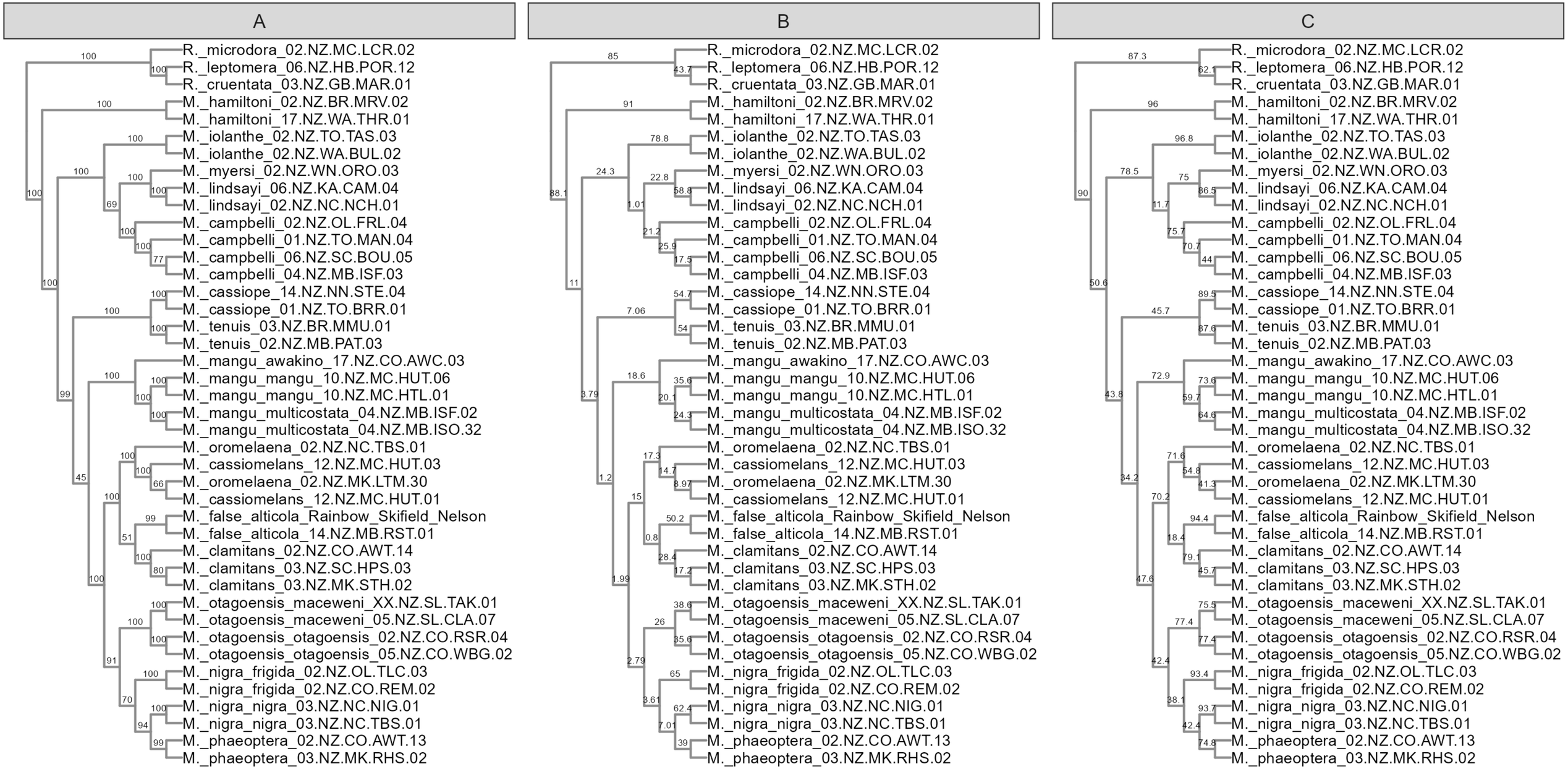


Fig. S23: *Maoricicada* SVDQuartets cladogram displaying A: bootstrap percentage, B: gene concordance factors, and C: site concordance factors.

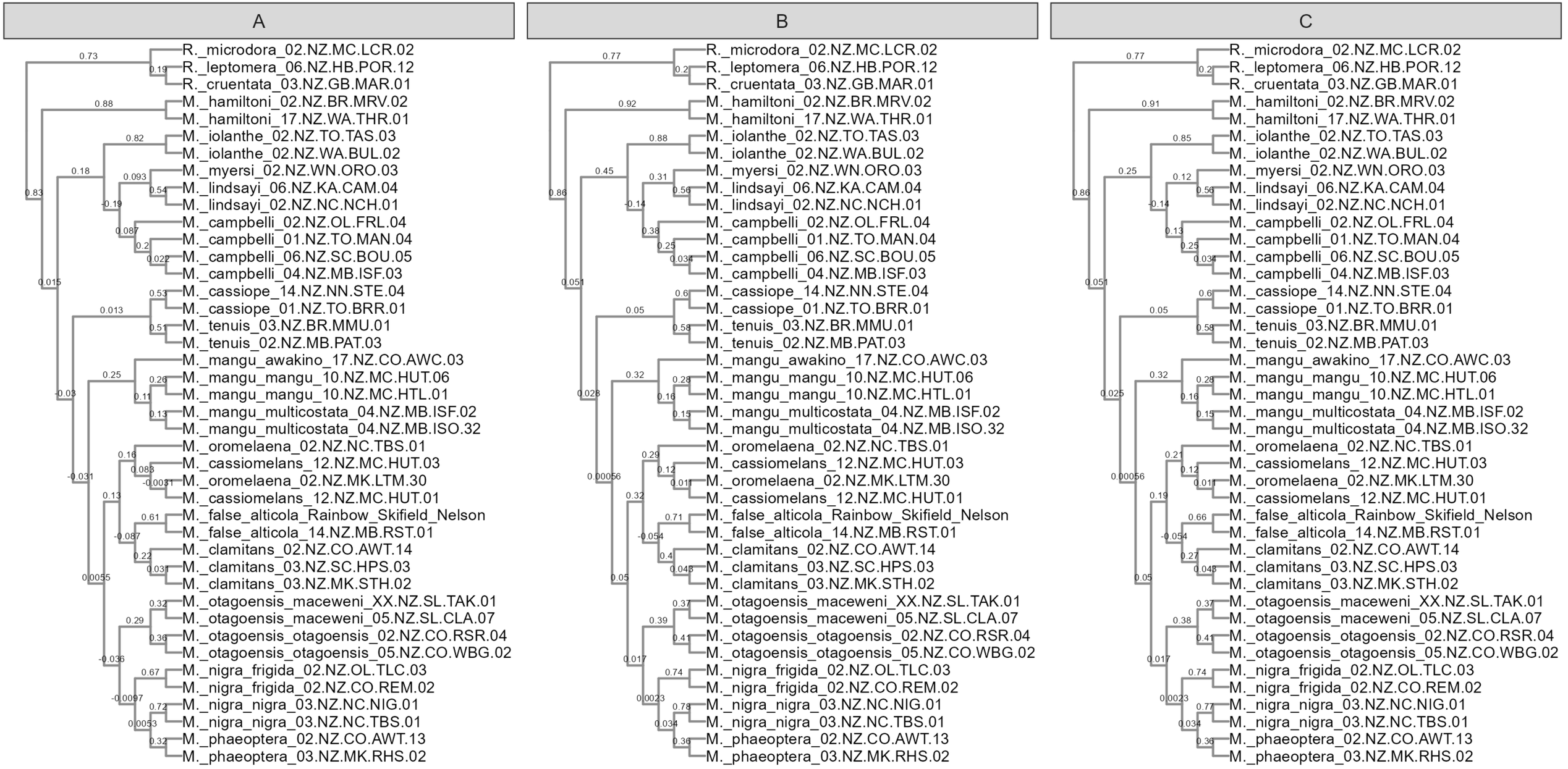


Fig. S24: *Maoricicada* SVDQuartets cladogram displaying A: Lowest Quartet Internode Certainty (LQ-IC), B: Quadripartition Internode Certainty (QP-IC), and C: Extended Quadripartition Internode Certainty (EQP-IC).