J. Entomol. Res. Soc., 28(2): 89-107, 2026 *Research Article*
Doi: 10.51963/jers.v28i2.3006 *Online ISSN:2651-3579*

Phylogenetic Analysis of Korean Cicada (Insecta: Hemiptera) Species Based on Mitochondrial and Nuclear Loci

Hoa Quynh NGUYEN[#1,2a*] Sungsik KONG[#3,4b] Yoonhyuk BAE[5c]

Jiman HEO[2d] Jae-Yeon KANG[2,6e] Yikweon JANG[2,7*]

[1]University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology, No. 18 Hoang Quoc Viet Street, Nghia Do Ward, Hanoi, 100000, VIETNAM

[2]Interdisciplinary Program of Ecocreative, Ewha Womans University, Ewhayeodaegil-52, Seoul, 03760, REPUBLIC OF KOREA

[3]The Institute for Computational and Experimental Research in Mathematics, Brown University, 121 South Main Street, Providence, RI, 02903, USA

[4]RIKEN Interdisciplinary Theoretical and Mathematical Sciences (iTHEMS), 2-1 Hirosawa, Wako, Saitama, 351-0198, JAPAN

[5]Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, REPUBLIC OF KOREA

[6]National Institute of Ecology, 1210 Geumgang-ro, Maseo-myeon, Seocheon-gun, Chungcheongnam-do, 33657, REPUBLIC OF KOREA

[7]Department of Life Sciences and Division of Ecoscience, Ewha Womans University, Ewhayeodaegil-52, Seodaemun-gu, Seoul, 03760, REPUBLIC OF KOREA

e-mails: [a]nguyen-quynh.hoa@usth.edu.vn; [b]sungsik.kong@riken.jp; [c]gyyh0303@snu.ac.kr; [d]wlaksaoal@khu.ac.kr; [e]jaeyoni@nie.re.kr; [f]jangy@ewha.ac.kr

ORCID IDs: [a]0000-0002-9157-1647, [b]0000-0002-5882-2109, [c]0000-0002-5840-9277, [d]0000-0002-6154-8630, [e]0000-0002-3937-7812, [f]0000-0002-7213-7319

*Correspondence, # Equal contribution

ABSTRACT

Documenting species diversity and elucidating their evolutionary relationships are critical to understanding evolutionary processes such as adaptation and speciation. While 13 cicada species from 11 genera are known to inhabit in the Republic of Korea, their phylogenetic relationships remain poorly understood. In this study, we present phylogenies of 12 species found in the Republic of Korea along with other closely related species, using Maximum Likelihood, Bayesian Inference and Maximum Parsimony methods based on a mitochondrial cytochrome c oxidase locus, two nuclear loci (elongation factor 1-alpha and 18S rRNA), and a dataset composed of two or all loci concatenated. The estimated phylogenies were generally congruent with one another. More specifically, all species belonging to the subfamily

Nguyen, H. Q., Kong, S., Bae, Y., Heo, J., Kang, J-Y., & Jang, Y. (2026). Phylogenetic analysis of Korean Cicada (Insecta: Hemiptera) species based on mitochondrial and nuclear loci. *Journal of the Entomological Research Society*, 28(2),89-107.

Received: December 27, 2025 *Accepted: June 25, 2026*

Cicadettinae formed a monophyletic group in the tree inferred using the concatenated sequences. The positions of some species varied depending on the loci used, possibly due to a lack of phylogenetic signal in parts of the dataset. Our analysis fills the knowledge gap in the phylogenetic positions of three Korean cicada species, *Suisha coreana* Matsumura, 1927, *Leptosemia takanonis* Matsumura 1917 and *Tettigetta isshikii* Kato 1926, all of which were not previously investigated.

Keywords: cicada fauna, Korean Peninsula, *COI, EF1α*, 18S rRNA, concatenated data.

INTRODUCTION

The cicada fauna of the Korean Peninsula has been documented through various taxonomic studies. Early descriptions and subsequent systematic reviews have documented 13 species from 11 genera under seven tribes within two major subfamilies: Cicadinae and Cicadettinae (Lee, 2008). Recently, several taxonomic revisions have been made including the review of *Lyristes intermedius* and *Auritibicen intermedius* and the synonymy of *Oncotympana fuscata* with *Hyalessa maculaticollis* (Motschulsky, 1866) (Liu et al., 2018; Motschulsky, 1866; Nguyen et al., 2022; Puissant & Lee, 2016). The current cicada species occurring in The Republic of Korea include *Suisha coreana* Matsumura, 1927, *A. intermedius* Mori, 1931, *Cicadetta abscondita* Lee 2009, *C. montana* Scopoli, 1772, *Graptopsaltria nigrofuscata* de Motschulsky 1866, *Cryptotympana atrata* Fabricius, 1775, *Hyalessa maculaticollis* Motschulsky, 1866, *Kosemia yezoensis* Matsumura 1898, *Leptosemia takanonis* Matsumura 1917, *Meimuna mongolica* Distant, 1881, *M. opalifera* Walker, 1850, *Tettigetta isshikii* Kato 1926, and *Platypleura kaempferi* Fabricius, 1794.

Studies on cicada phylogeny have employed both transcriptomic and morphological data to infer evolutionary relationships. Transcriptome-based phylogenomics (Skinner et al., 2017) enabled the examination of ambiguous relationships within Auchenorrhyncha using extensive gene datasets, degeneracy coding, and gene conflict analysis. Meanwhile, morphological analyses (Moulds 2005), supported the division of Cicadidae into two families, Tettigarctidae and Cicadidae, and three subfamilies, with these groupings later corroborated by multi-gene molecular phylogenies (Marshall et al., 2018). Morphological traits, including male genitalia and external features, have been used to assess species boundaries, though subtle differences sometimes require integrating acoustic and genetic evidence (Lee, 2008). However, transcriptomic data remains limited for many cicada species, and morphological datasets often show conflicting signals when delimiting species (Nunes et al., 2014; Wiens & Penkrot, 2002). Other phylogenetic analyses have also been used to infer evolutionary relationships and delimit species boundaries in cicadas (Arensburger, Buckley, Simon, Moulds, & Holsinger, 2004; Cryan & Urban, 2012; Hill, Marshall, Moulds, & Simon, 2015; Marshall, Hill, Cooley, & Simon, 2011; Marshall et al., 2012; Marshall et al., 2016; Marshall et al., 2018; Popple, 2013; Price, Marshall, Barker, Simon, & Villet, 2019; Sueur, Vanderpool, Simon, Ouvrard, & Bourgoin, 2007).

Several attempts have been made for the revision of the Korean cicadas by reconstructing phylogenetic trees based on the mitochondrial cytochrome c oxidase gene (*COI*) (Lee, 2008; Lee, Choe, Lee, & Woo, 2002) and the mitochondrial genome

as the input for phylogenetic (or matrilineal) tree reconstruction (Nguyen et al., 2022). However, phylogenetic analysis based on a single mitochondrial locus can lead to biased evolutionary relationships (Castresana, 2007). The use of both mitochondrial and nuclear loci is crucial for an accurate depiction of evolutionary histories.

Previous molecular studies on Korean cicadas have either focused on *COI* or on the mitochondrial genome of a single genus *Hyalessa*, providing important but taxonomically restricted insights into Korean cicada phylogeny (Lee, 2008; Lee et al., 2002; Nguyen et al., 2022). At a broader scale, Marshall et al. (2018) reconstructed a global phylogeny of Cicadidae based on multiple loci, and proposed major revisions of tribal and subfamily classification, e.g., in which the genus *Leptosemia* was moved to the tribe Leptopsaltriini Moulton and the genus *Meimuna* was moved to the tribe Dundubiini Distant. Although this study covers most cicadid genera, of the genera in the Cicadidae subfamily, some species occurring in the Korean peninsula were missing, such as *S. coreana, L. takanonis* and *T. isshikii*. This omission is due to the lack of readily available sequences in public databases like GenBank.

Here, we provide the first multi-locus phylogenetic analysis of the cicada fauna of the Korean Peninsula that includes 12 of the 13 currently recognized species and all 11 genera. Using one mitochondrial locus (*COI*) and two nuclear loci (elongation factor 1-alpha (*EF1α*) and 18S rRNA), we reconstruct Maximum Likelihood (ML), Maximum Parsimony (MP) and Bayesian inference (BI) of phylogenetic trees for individual loci and concatenated datasets to (i) evaluate the relationships of Korean genera and tribes under the updated classification, (ii) place previously unsampled Korean species (*S. coreana*, *L. takanonis* and *T. isshikii*) within a broader phylogenetic framework, and (iii) qualitatively assess the degree of congruence among mitochondrial and nuclear markers for this regional fauna. These results fill important gaps left by previous Korean and global studies and provide a basis for future genomic-scale work for understanding the evolutionary relationship of East Asian cicadas.

MATERIAL AND METHODS

Sample collection

Fresh and dried specimens were sampled. Detailed information regarding sample collection is provided in Figure 1 and Table S1. Tissue samples from the limbs of each sample were used for DNA extraction using DNeasy® Blood and Tissues Kit (QIAGEN group, Hilden, Germany) following the instructions of the manufacturers. Although Lee et al. (2008) listed 13 cicada species from 11 genera in the Korean cicada fauna, we focus on 12 species because we were unable to sample *Cicadetta montana*. As an alternative, the inclusion of *C. abscondita*, a closely related species and putative sister to *C. montana* in the Far Eastern Palaearctic (Lee et al., 2002; Lee, 2008), could retain the representation of all 11 genera.

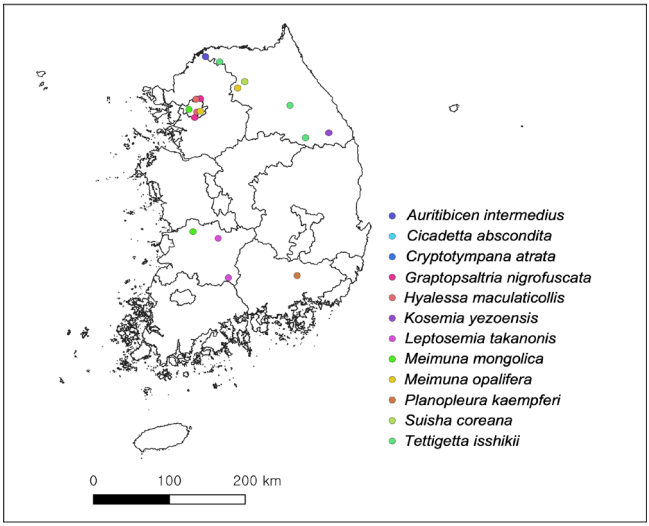

Figure 1. Map of sampling locations for 12 cicada species in the Korean Peninsula.

PCR amplification

Segments of three loci were chosen for polymerase chain reaction (PCR) amplification: 619 bp of *COI*, 558 bp of *EF1α* and 1797 bp of 18S rRNA. The sequences of the primers used for PCR amplification and their annealing temperature used for each locus are provided in Table 1. Note that for the 18S rRNA, two overlapping fragments were amplified, namely 18s1 and 18s2, and concatenated to form an overall sequence of this gene. In addition, the cicada samples were run along with a negative control of ultrapure water (Biosesang Inc., Gyeonggi-do, Republic of Korea) for each gene. Takara Ex Taq (Takara Korea Biomedical Inc., Seoul, Republic of Korea) was used for the amplification reaction. Specifically, a total of 25 μL amplified sample was comprised of 3 μL template DNA and 22 μL of master mix (0.125 μL of Takara Ex TaqTM 5 U/μL, 2.5 μL of 10× Buffer, 2.5 μL of dNTP Mix 2.5 mmol/L, 2 μL of MgCl2 25 mmol/L, 1 μL of each primer 10 mmol/L and ultrapure water (Biosesang Inc., Gyeonggi-do, Republic of Korea). Polymerase chain reaction amplification initiated by 1 min initial denaturation at 94°C, followed by 30 cycles of 30 s denaturation at 94°C, 1 min at the specific annealing temperature of each gene (Table 1), and 1 min elongation at 72°C, finally completed by 2 min terminal elongation at 72°C. Two microliters of each PCR product were loaded on 1.5% agarose gel and visualized using the loading dye MaestroSafe Nucleic Acid loading dye (MaestroGen Inc., Hsinchu, Taiwan) on an UltraSlim LED Illuminator (MaestroGen Inc., Hsinchu, Taiwan). Bidirectional sequencing of PCR products was performed by COSMO Genetech Company (COSMO Genetech Co., Ltd., Seoul, Republic of Korea). Among all species, only *COI* fragments of *A. intermedius* were not successfully amplified; therefore, this species was retained in the concatenated dataset of nuclear genes, but omitted in the concatenated dataset of all genes.

Table 1. List of primers and related information regarding their sequences and respective annealing temperatures used for the polymerase chain reaction (PCR) amplification used in this study. Note 18s1 and 18s2 are used to amplify two overlapping fragments of 18S rRNA.

Gene	Primer name	Annealing temperature	Primer sequence
EF1α	EF1-F650-mod	53°C	TGCTGCGGGTACTGGTGAAT (Arensburger et al., 2004)
	EF1-N-1419		ACACCAGTTTCAACTCTGCC (Arensburger et al., 2004)
18s1	18s1F	49°C	TACCTGGTTGATCCTGCCAGTAG (Giribet, Carranza, Baguñà, Riutort, & Ribera, 1996)
	18s5R		CTTGGCAAATGCTTTCGC (Giribet et al., 1996)
18s2	18s5F	49°C	GCGAAAGCATTTGCCAAGAA (Giribet et al., 1996)
	18s9R		GATCCTTCCGCAGGTTCACCTAC (Giribet et al., 1996)
COI	LCO1490	48°C	GGTCAACAAATCATAAAGATATTGG (Folmer, Black, Hoeh, Lutz, & Vrijenhoek, 1994)
	HCO2198		TAAACTTCAGGGTGACCAAAAAATCA (Folmer et al., 1994)

Sequencing chromatograms were manually inspected for base-calling quality and trimmed to remove low-quality ends in Geneious v9.1 (https://www.geneious.com). For *COI* and *EF1α*, sequences were translated into amino acids to screen for premature stop codons or frameshifts that would indicate pseudogenes. Preliminary alignments were generated using ClustalW (Thompson, Higgins, & Gibson, 1994) implemented in Geneious with manual inspection to remove ambiguously aligned regions. For each species, representative sequences (including the *L. takanonis* accession MN160132) were compared to the National Center for Biotechnology Information (NCBI) database using Basic Local Alignment Search Tool to verify that the closest matches corresponded to the expected taxa and to rule out obvious contamination. The annotated sequences of our samples were submitted to GenBank under the accession numbers listed in Table S1.

Phylogenetic analysis

Phylogenetic relationships among cicada species in the Republic of Korea were inferred using five multiple sequence alignment datasets where three of them contain a single mitochondrial or nuclear locus (i.e., *COI*, *EF1α*, and 18S rRNA), one that concatenates two nuclear loci, and one that concatenates all three loci. For the three cases that contain a single locus, we extended the alignment by downloading relevant and appropriate sequences available on GenBank to corroborate our species identification. Detailed information regarding the GenBank accession number for each sequence can be found in Table S2.

For each dataset, we performed a multiple sequence alignment using Multiple Alignment using Fast Fourier Transform (MAFFT) v7.526 (Katoh, Misawa, Kuma, & Miyata, 2002) with a local alignment strategy and an iterative refinement using eight threads for efficiency, keeping the original input order of sequences. For the single-loci alignments, we conducted ML, MP, and BI analyses of phylogenetic trees. More specifically, ML phylogenetic analyses were performed using IQ-TREE 2 v2.3.6 (Minh et al., 2020). The best-fitting nucleotide substitution model was automatically selected using ModelFinder (Kalyaanamoorthy, Minh, Wong, Von Haeseler, & Jermiin, 2017) implemented in the software according to the Bayesian information criterion (BIC).

Ultrafast bootstrap support (BS) was computed from 10,000 bootstrap replicates to determine the clade robustness (Hoang, Chernomor, Von Haeseler, Minh, & Vinh, 2018). Phylogenetic analysis under MP criterion was performed using Phylogenetic Analysis Using Parsimony *and other methods (PAUP*) v4.0a (Swofford, 2003). Heuristic searches were performed with 100 random sequence addition replicates and tree-bisection-reconnection branch swapping. Node support was evaluated using 1,000 bootstrap replicates with heuristic searches. Bootstrap trees were saved and summarized in a majority-rule consensus tree to assess the robustness of inferred clades. Bayesian inference was conducted using MrBayes v3.2.7 (Ronquist et al., 2012). The nucleotide substitution model that was selected in the ML analysis using the same dataset was applied, incorporating empirical base frequencies, a proportion of invariant sites, and gamma-distributed rate variation among sites. Two independent Markov chain Monte Carlo analyses were performed, each consisting of four chains with a temperature setting of 0.2 to improve mixing. Analyses were run for ten million generations, with trees sampled at regular intervals and convergence assessed through standard diagnostic measures. The first fifty thousand generations were discarded as burn-in, after which posterior parameter estimates were summarized and a majority-rule consensus tree was generated. For the concatenated datasets, we performed ML and BI analyses. We partitioned the alignment by gene, with each gene treated as a separate partition. For each gene, we applied substitution model parameters estimated above using ModelFinder implemented in IQ-TREE 2. For all BI analysis, we used Tracer v1.7.2 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018) to ensure convergence.

Higher-level molecular phylogenies of Cicadidae consistently recover Cicadettinae as a close relative of Cicadinae, whereas Tibicininae is more distantly related and often represented by long branches (Lukasik et al., 2018; Marshall et al., 2018). At the time of this study, *EF1α* sequences for Tibicininae species were not available in GenBank. Using Tibicininae as outgroups would therefore have required relying on a single nuclear marker and a mitochondrial locus and might have introduced artefactual long-branch attraction. For these reasons, and consistent with the topology of Marshall et al. (2018), we used *K. yezoensis*, *Cicadetta* spp., and *Tettigetta* spp. (all Cicadettinae) as outgroups to Cicadinae species. Personal communication with Dr. C. Simon further supported this choice. While we used *K. yezoensis* to root the tree, we also included multiple taxa from the sister group to be a part of our outgroup species, as this strategy has been shown to improve consistency (Luo et al., 2010).

RESULTS

Data acquisition

Phylogenetic trees were reconstructed for 37 cicada specimens, comprising 23 fresh and 14 dried samples. Amplification was largely successful for all three loci, with exceptions observed in two dried and one fresh samples of *A. intermedius* and one fresh sample of *C. abscondita* sample for the *COI* gene, one fresh and one dried

samples of *L. takanonis* for *EF1α*, and two dried samples of *H. maculaticollis* samples for 18S rRNA. We conducted a Fisher's exact test to test if there was any tentative impact of the preservation type of the specimen on PCR success. No significant differences were observed between the two preservation types, regardless of single gene testing or a combination of all three genes. Therefore, such minor failures in PCR amplification are unlikely to be attributable to preservation state or overall DNA quality. Instead, they may reflect the presence of PCR inhibitors co-extracted with DNA or stochastic amplification failure. All acquired sequences are deposited in GenBank (see Table S1 for accession numbers).

Phylogenetic relationships estimated from mitochondrial and nuclear markers

The multiple sequence alignment contained 51 sequences of 611 bp for *COI*, 47 sequences of 525 bp for *EF1α*, and 37 sequences of 1798 bp for 18S rRNA. TVM+F+G4, HKY+F+I, and K2P+I were selected as the best fit nucleotide substitution models for *COI, EF1α*, and 18S rRNA, respectively. Two concatenated alignments of two nuclear loci and all three loci were also created to enhance the phylogenetic resolution, in which the former contained 33 sequences of 2358 bp, whereas the latter contained 29 sequences of 2970 bp. No DNA sequences of the species *A. intermedius* were included due to the lack of *COI* sequence.

Figures 2-4 present the consensus phylogenies using the alignment that contains *COI*, two concatenated nuclear loci, and all three loci concatenated, respectively. For *COI*, three genealogies from the three analyses using different optimality criteria (i.e., MP, ML, and BI) mostly agreed with each other. Figure 2 shows the BI phylogeny using *COI* with node support showing posterior probabilities of the BI, BS of the ML and MP. The only difference among the three trees was that *C. abscondita* and *C. calliope* did not form a monophyletic group in the MP tree. All species formed a clade and the species that belong to the subfamily Cicadinae also formed a monophyletic group. The two species that belong to the Tribe Platypleurini (i.e., *P. kaempferi* and *S. coreana*) formed a monophyletic group, and this tribe was sister to the rest of the tribes. Three species in the tribe Cryptotympanini (*A. intermedius, A. japonicus,* and *C. atrata*) formed a monophyly with low BS (62/100 for ML and 66/100 for MP) but high posterior probability (0.86), forming a cherry with *L. takanonis*. Note the term "cherry" is standard in mathematical and theoretical phylogenetics and refers to a pair of leaves sharing the same parent node, resembling a pair or cherries on a stem. The two species that belongs to Polyneurini and Sonatini, *G. nigrofuscata* and *H. maculaticollis*, respectively, formed a monophyly, which subsequently shared the common ancestor with the two species of Dundubiini that occurs in the Republic of Korea (*M. mongolica* and *M. opalifera*), although the node supports were relatively low for BI (0.59) and ML (39/100). This clade shared the common ancestor with the rest of the species in Cicadinae; this sister relationship was supported in all node supports.

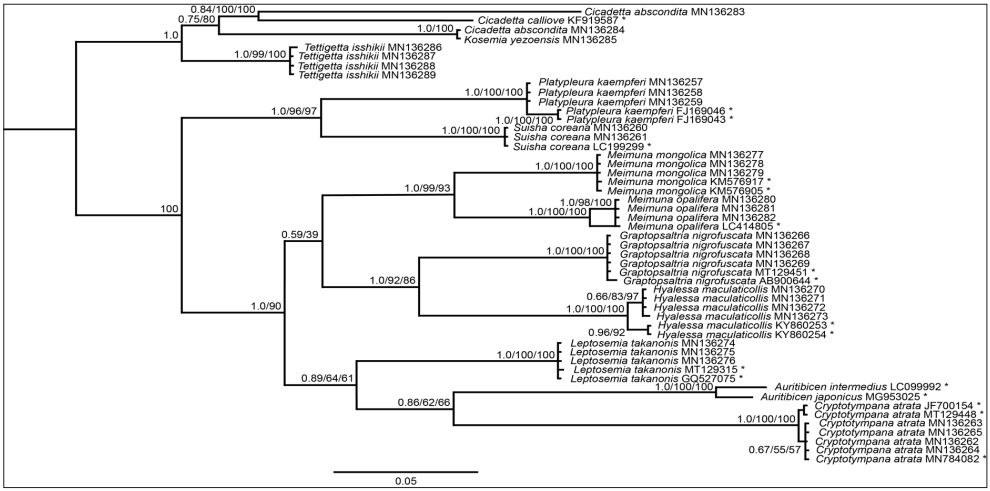

Figure 2. An extended majority-rule consensus Bayesian inference (BI) genealogy of the cicada species occurring in the Republic of Korea sampled in this study plus available GenBank sequences added, based on partial mitochondrial cytochrome c oxidase subunit 1 (COI) locus. Values at the internal nodes represent node support, shown as posterior probabilities from BI, bootstrap values from ML, and bootstrap values from MP analyses, presented in that order and separated by slashes. Each tip label includes the species name and a GenBank accession number; asterisks were added only for sequences retrieved from GenBank. The genealogy is polarized using the three species in Cicadettinae: *Kosemia yezeonsis*, *Cicadetta abscondita* and *Tetigetta isshikii*. The scale bar below the tree measures branch length in the number of substitutions per site.

All MP, ML, and BI phylogenies inferred using *EF1α* and 18S rRNA alignment are presented in Figures S3 to S8. While phylogenies inferred using either *EF1α* or 18S rRNA generally agreed with that of *COI* phylogeny (Figure 2), some discrepancies were observed. For instance, genus *Meimuna* failed to form a monophyletic group when the phylogeny was inferred using *EF1α*. Moreover, one terminal of *L. takanonis* (MN160132) formed a clade with the members of *H. maculaticollis*, suggesting potential misidentification during the sequencing process. In the 18S rRNA phylogeny, the genus *Meimuna* was resolved as a monophyletic group (Figures S6-S8), however, many parts of the tree were different from the tree in Figure 2, which may be attributable to limited phylogenetic signal for resolving species-level relationships.

Figures 3 and 4 show the BI phylogenies using the dataset composed of two nuclear loci (*EF1α* and 18S rRNA) and all three loci, respectively. The species relationships in these topologies are in general congruent with the *COI* phylogeny and with each other, except for *L. takanonis* (MN160132) which again formed a monophyletic group with *H. maculaticollis* as in the *EF1α* phylogeny. Interestingly, when all three loci were concatenated (Figure 4), the monophyly of *L. takanonis* and *H. maculaticollis* is no longer observed. Instead, *H. maculaticollis* forms a cherry with *G. nigrofuscata*, which is the sister clade of *L. takanonis*.

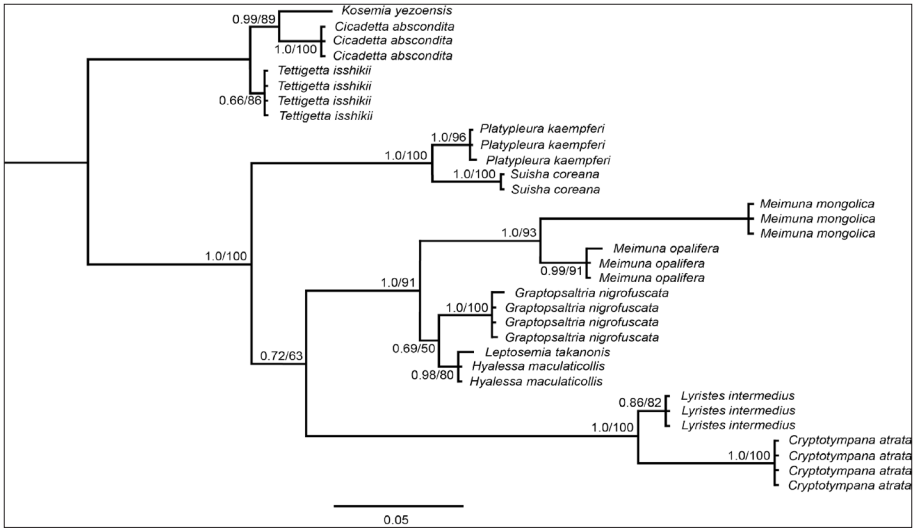

Figure 3. An extended majority-rule consensus BI genealogy of the 12 cicada species occurring in the Republic of Korea sampled in this study based on the concatenated sequences of two nuclear loci (*EF1α* and 18S rRNA). Values at the internal nodes represent node support, shown as posterior probabilities from BI, and bootstrap percentages from ML analyses, presented in that order and separated by a slash. The genealogy is polarized using the three species in Cicadettinae: *Kosemia yezeonsis*, *Cicadetta abscondita* and *Tetigetta isshikii*. The scale bar below the tree measures branch length in the number of substitutions per site. Separate BI, ML and MP trees for individual genes are available in Supplementary Materials S3-S8.

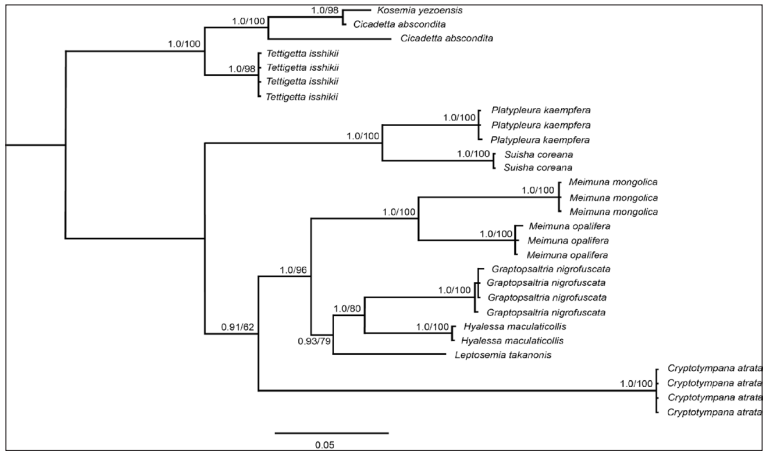

Figure 4. An extended majority-rule consensus BI genealogy of the 11 cicada species occurring in the Republic of Korea sampled in this study based on concatenated mitochondrial COI locus and two nuclear loci (EF1α and 18S rRNA). Values at the internal nodes represent node support, shown as posterior probabilities from BI, and bootstrap percentages from ML analyses, presented in that order and separated by a slash. The genealogy is polarized using the three species in Cicadettinae: *Kosemia yezeonsis*, *Cicadetta abscondita* and *Tetigetta isshikii*. The scale bar below the tree measures branch length in the number of substitutions per site.

DISCUSSION

In general, the reconstructed phylogenies agreed with previous studies on cicada fauna which also indicate a clear segregation between the two subfamilies Cicadinae and Cicadettinae. Comparing with the most updated phylogeny inferred using five genes (C*OI, COII, ARD1, EF1α* and 18S rRNA) from Marshall et al. (2018), the two species, *C. atrata* and *A. japonicus* belonged to the tribe Cryptotympanini, which was consistent with the relationship between *C. atrata* and *A. intermedius*, the sister species of *A. japonicus*, observed in both the nuclear-gene tree and the *COI* tree of this study. Moreover, the tribe Cryptotympanini was sister to all other cicada species in the family Cicadidae and this was supported with 1.0 posterior probability and 86/100 BS. This was also observed in our nuclear-gene phylogeny with high BS (99/100). In contrast, the mitochondrial tree and the concatenated-data tree somewhat exhibited different topologies. Further phylogenetic analysis including other cicada species and more genes should be conducted to better resolve the relationships between these tribes.

Our results have several implications for the tribal classification of Korean cicadas. First, in agreement with Marshall et al. (2018), *C. atrata* and *Auritibicen* species (*A. intermedius*, *A. japonicus*) form a monophyletic group that is sister to all other Cicadinae in our nuclear concatenated tree (Fig. 2). This strongly supports the placement of these genera into Cryptotympanini for the Korean fauna. Second, *M. mongolica* and *M. opalifera* form a clade that is closely associated with *G. nigrofuscata* (Polyneurini) and *H. maculaticollis* (Sonatini), consistent with the transfer of genus *Meimuna* to Dundubiini and with the broader relationships among tribes inferred by Marshall et al. (2018). Third, the two Korean representatives of Platypleurini, *P. kaempferi* and *S. coreana*, consistently form a strongly supported clade that is sister to the remaining Cicadinae in our *COI* tree and concatenated analyses, matching the tribal level placement and patterns of Platypleurini documented by Price et al. (2019). Overall, our multi-locus regional phylogeny strengthens the tribal assignments proposed in recent global studies and provides additional support for applying these revisions to the Korean cicada fauna.

One of the limitations of our study is the incomplete dataset as the sequence of *C. montana* sequences is absent. However, the inclusion of its close relative *C. abscondita* means that the genus *Cicadetta* is still represented, and we therefore expect minimal impact on the inferred relationships among the focal Korean tribes. In addition, the variation in relationships among tribes in the three phylogenies may suggest a lack of signals in some of the examined genes. The BS of the tree using concatenated sequences of all three genes was also high in almost all nodes (<0.9; except two internal nodes that had ML BS of 62/100 and 79/100) compared to those relying on a single mitochondrial gene or two nuclear genes. This observation reassures the gene tree discordance and therefore, the examination of various genes, like genomic data, could provide a comprehensive understanding of the evolution of targeted taxonomic groups.

The observed disagreements among our phylogenetic trees likely stem in part from differences in the amount and quality of signal present in the loci analyzed (i.e., marker choice and alignment length). Nuclear DNA markers in our dataset may

not contain sufficient phylogenetic signals to fully resolve the relationships, leading to topological incongruences relative to the mitochondrial DNA tree (Gaines, Hare, Beck, & Rosenbaum, 2005). This is not uncommon, as single or few nuclear loci often accumulate limited informative variation, and robust resolution typically emerges only when multiple independent markers are combined (Gaines et al., 2005). By contrast, *COI* genes, which are widely used as a standard DNA barcode for animal species identification with relatively high mutation rate and extensive reference data, often provide well-resolved distinctions among genera or species. In our case, the *COI* genealogy clearly resolved major lineages distinctly. However, reliance on a single mitochondrial marker like *COI* can be problematic for resolving deeper or more complex relationships. Mitochondrial gene trees represent the history of a single maternally inherited locus and may therefore provide a strong but potentially biased signal, particularly when compared with multi-locus nuclear datasets (Wang et al., 2018).

Beyond methodological consideration, biological processes can also drive discordance among gene trees. Mitochondrial and nuclear trees are often incongruent because gene trees can deviate from the underlying species history due to processes such as incomplete lineage sorting (ILS) or hybridization (Wang et al., 2018). For instance, ILS can lead to stochastic retention of ancestral alleles and the hybridization can horizontally inherit a portion of genetic material across species boundaries, both resulting in topologies that do not reflect the true species history. In short, phylogenetic inference based on a single gene (particularly mitochondrial locus) may fail to capture the full complexity of evolutionary history. These discrepancies in our results therefore likely reflect a combination of limited resolution in the nuclear dataset and biological processes, and highlight the importance of multi-locus approaches to obtain a more reliable cicada phylogeny.

Future research on Korean cicadas should directly address the limitations revealed in this study, namely limited loci, incomplete taxon coverage, and geographically sparse sampling. First, an extension of genomic-scale datasets in terms of locus number is necessary to resolve nodes that remain weakly supported or incongruent among *COI*, *EF1α*, and 18S rRNA. Importantly, large multi-locus datasets allow analysis of researchers to detect and account for evolutionary processes such as incomplete lineage sorting and hybridization, which are often undetectable or confounded in single-gene analyses (Wang et al., 2018). Gene tree discordance caused by incomplete lineage sorting or introgressive hybridization can be teased apart using coalescent-based frameworks applied to genome-scale data, helping to distinguish true species relationships from random gene tree noise (Wang et al., 2018). Second, expanding taxon sampling to include the unsampled *C. montana*, additional individuals of *Leptosemia* (including the problematic MN160132 lineage), and other under-represented tribes will help test whether current topological conflicts reflect missing taxa or true biological complexity. Third, broader geographic sampling across the ranges of widespread species such as *H. maculaticollis* and *Meimuna* spp. will capture intraspecific variation and reveal population structure or cryptic lineages that cannot be detected from the limited regional sampling used here. Together, genomic-scale phylogenomic analyses combined with improved taxon and geographic

coverage will provide a more robust framework for clarifying the remaining ambiguities in Korean cicada relationships and for linking patterns of gene tree discordance to underlying evolutionary processes. This integrative approach will help clarify any remaining ambiguities and improve our understanding of the evolutionary processes shaping the Korean cicada fauna.

FUNDING

This work was funded by the National Research Foundation of Korea (Grant Number: 2015R1A4A1041997). This work was partially funded by the National Science Foundation (NSF) under Grant DMS-1929284 while S.K. was in residence at the Institute for Computational and Experimental Research in Mathematics in Brown University, Providence, RI, USA, during the Theory, Methods, and Applications of Quantitative Phylogenomics program. This work was also partially supported by RIKEN iTHEMS.

ACKNOWLEDGMENTS

We thank Chris Simon for her advice on selecting appropriate outgroup species for rooting the reconstructed genealogies.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

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(4), 557-569.

Castresana, J. (2007). Topological variation in single-gene phylogenetic trees. *Genome biology,* 8(6), 216.

Cryan, J. R. & Urban, J. M. (2012). Higher-level phylogeny of the insect order Hemiptera: is Auchenorrhyncha really paraphyletic? *Systematic Entomology,* 37(1), 7-21.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology,* 294-9.

Gaines, C., Hare, M., Beck, S., & Rosenbaum, H. (2005). Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species. *Proceedings of the Royal Society B: Biological Sciences,* 272(1562), 533-542.

Giribet, G., Carranza, S., Baguñà, J., Riutort, M., & Ribera, C. (1996). First molecular evidence for the existence of a Tardigrada+Arthropoda clade. *Molecular biology and evolution,* 13(1), 76-84.

Hill, K. B., Marshall, D. C., Moulds, M. S., & Simon, C. (2015). Molecular phylogenetics, diversification, and systematics of *Tibicen* Latreille 1825 and allied cicadas of the tribe Cryptotympanini, with three new genera and emphasis on species from the USA and Canada (Hemiptera: Auchenorrhyncha: Cicadidae). *Zootaxa,* 3985(2), 219-251-219-251.

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(2), 518-522.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods,* 14(6), 587-589.

Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research,* 30(14), 3059-3066.

Lee, Y. J. (2008). Revised synonymic list of Cicadidae (Insecta: Hemiptera) from the Korean Peninsula, with the description of a new species and some taxonomic remarks. *Proceedings of the Biological Society of Washington,* 121(4), 445-467.

Lee, Y. J. (2017). Resurrection of *Auritibicen shikokuanus* (Kato, 1959) stat. rev. & comb. nov. (Hemiptera: Cicadidae: Cryptotympanini) from Ehime and Hiroshima of Japan. *Journal of Asia-Pacific Biodiversity,* 10(1), 59-64.

Lee, Y. J., Choe, H.-J., Lee, M.-L., & Woo, K.-S. (2002). A phylogenetic consideration of the Far Eastern Palaearctic species of the genus *Cicadetta* Kolenati (Homoptera, Cicadidae) based on the mitochondrial COI gene sequences. *Journal of Asia-Pacific Entomology,* 5(1), 3-11.

Liu, Y., Qiu, Y., Wang, X., Yang, H., Hayashi, M., & Wei, C. (2018). Morphological variation, genetic differentiation and phylogeography of the East Asia cicada *Hyalessa maculaticollis* (Hemiptera: Cicadidae). *Systematic Entomology,* 43(2), 308-329.

Luo A.-R., Zhang Y.-Z., Qiao H.-J., Shi W.-F., Murphy R. W., Zhu C.-D. 2010. Outgroup selection in tree reconstruction: a case study of the family Halictidae (Hymenoptera: Apoidea). Acta *Entomologica Sinica*, 53(2), 192-201

Marshall, D. C., Hill, K. B., 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(4), 482-502.

Marshall, D. C., Hill, K. B., Marske, K. A., Chambers, C., Buckley, T. R., & Simon, C. (2012). Limited, episodic diversification and contrasting phylogeography in a New Zealand cicada radiation. *BMC Evolutionary Biology,* 12(1), 177.

Marshall, D. C., Hill, K. B., Moulds, M., Vanderpool, D., Cooley, J. R., Mohagan, A. B., & Simon, C. (2016). Inflation of molecular clock rates and dates: molecular phylogenetics, biogeography, and diversification of a global cicada radiation from Australasia (Hemiptera: Cicadidae: Cicadettini). *Systematic biology,* 65(1), 16-34.

Marshall, D. C., Moulds, M., Hill, K. B., Price, B. W., Wade, E. J., Owen, C. L., Goemans, G., Marathe, K., Sarkar, V., Cooley, J. R., & Sanborn, A.F. (2018). A molecular phylogeny of the cicadas (Hemiptera: Cicadidae) with a review of tribe and subfamily classification. *Zootaxa,* 4424(1), 1-64.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution,* 37(5), 1530-1534.

Motschulsky, V. d. (1866). Catalogue des insects reçus du Japon. *Bulletinde la Societe imperiale des Naturalistes de Moscou,* 39(1), 163-200.

Moulds, M. (2005). An appraisal of the higher classification of cicadas (Hemiptera: Cicadoidea) with special reference to the Australian fauna. *Records-Australian Museum,* 57(3), 375.

Nguyen, H. Q., Ho, P.-T., Kong, S., Bae, Y., Pham, T. H., La, H. T., & Jang, Y. (2022). A time-calibrated mitogenomic phylogeny suggests that Korean *Hyalessa fuscata* is a bridge between Chinese and Japanese *H. maculaticollis*. *Journal of Genetics,* 102(1), 9.

Nunes, V. L., Mendes, R., Marabuto, E., Novais, B. M., Hertach, T., Quartau, J. A., Seabra, S. G., Paulo, O. S., & Simoes, P. C. (2014). Conflicting patterns of DNA barcoding and taxonomy in the cicada genus Tettigettalna from southern Europe (Hemiptera: Cicadidae). *Molecular Ecology Resources*, 14(1), 27-38.

Popple, L. W. (2013). A revision of the *Pauropsalta annulata* Goding & Froggatt species group (Hemiptera: Cicadidae) based on morphology, calling songs and ecology, with investigations into calling song structure, molecular phylogenetic relationships and a case of hybridisation between two subspecies. *Zootaxa,* 3730, 1-102.

Price, B. W., Marshall, D. C., Barker, N. P., Simon, C., & Villet, M. H. (2019). Out of Africa? A dated molecular phylogeny of the cicada tribe Platypleurini Schmidt (Hemiptera: Cicadidae), with a focus on African genera and the genus *Platypleura* Amyot & Audinet-Serville. *Systematic Entomology,* 44(4), 842-861.

Puissant, S. & Lee, Y. J. (2016). Description of a new species of the genus *Hyalessa* China (Hemiptera: Cicadidae: Sonatini) from Yunnan, China, with a key to the species of *Hyalessa* and a calling song analysis for two *Hyalessa* species. *Zootaxa,* 4114(4), 434-446.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology,* 67(5), 901-904.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology,* 61(3), 539-542.

Skinner, R., Dietrich, C. H., Walden, K. K., Podsiadlowski, L., Petersen, M., Robertson, H. M., & Johnson, K. P. (2017). *Phylogenomics of Auchenorrhyncha using transcriptomes.* Paper presented at the Entomology 2017.

Steenwyk, J. L., Li, Y., Zhou, X., Shen, X.-X., & Rokas, A. (2023). Incongruence in the phylogenomics era. *Nature Reviews Genetics,* 24(12), 834-850.

Sueur, J., Vanderpool, D., Simon, C., Ouvrard, D., & Bourgoin, T. (2007). Molecular phylogeny of the genus *Tibicina* (Hemiptera, Cicadidae): rapid radiation and acoustic behaviour. *Biological Journal of the Linnean Society,* 91(4), 611-626.

Swofford, D. L. (2003). PAUP* phylogenetic analysis using parsimony (* and other methods). Version 4. *http://paup.csit.fsu.edu/*.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research,* 22(22), 4673-4680.

Wang, K., Lenstra, J. A., Liu, L., Hu, Q., Ma, T., Qiu, Q., & Liu, J. (2018). Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Communications biology,* 1(1), 169.

Wiens, J. J. & Penkrot, T. A. (2002). Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (Sceloporus). *Systematic biology*, 51(1), 69-91.

Table S1. A sample list of 12 cicada species inhabiting the Korean Peninsula employed in this study. For each sample, sampling information and the accession numbers of the COI, EF1α, and 18S rRNA genes are also included.

Subfamily	Tribe	Subtribe	Genus	Species	Collection date	Type of specimen	Latitude	Longitude	COI (619 bp)	EF1a (558 bp)	18s (1797 bp)
Cicadinae	Platypleurini		*Platypleura*	*kaempferi*	13-Jun-16	Dry	35.429369 N	128.495564 E	MN136257	MN160112	MN147780
					13-Jun-16	Dry	35.429369 N	128.495564 E	MN136258	MN160113	MN147781
					13-Jun-16	Dry	35.429369 N	128.495564 E	MN136259	MN160114	MN147782
			Suisha	*coreana*	24-Sep-17	Fresh	37.885361 N	127.738992 E	MN136260	MN160115	MN147783
					24-Sep-17	Fresh	37.885361 N	127.738992 E	MN136261	MN160116	MN147784
	Cryptotympanini		*Auritibicen*	*intermedius*	15-Aug-16	Dry	38.20363 N	127.148575 E	-	MN160117	MN147785
					15-Aug-16	Dry	38.20363 N	127.148575 E	-	MN160118	MN147786
					14-Aug-17	Fresh	38.202108 N	127.149933 E	-	MN160119	MN147787
			Cryptotympana	*atrata*	2-Aug-14	Fresh	37.510082 N	127.076703 E	MN136262	MN160120	MN147788
					2-Aug-14	Fresh	37.510082 N	127.076703 E	MN136263	MN160121	MN147789
					2-Aug-14	Fresh	37.510082 N	127.076703 E	MN136264	MN160122	MN147790
					2-Aug-14	Fresh	37.510082 N	127.076703 E	MN136265	MN160123	MN147791
	Polyneurini		*Graptopsaltria*	*nigrofuscata*	8-Aug-16	Dry	37.668706 N	127.077208 E	MN136266	MN160124	MN147792
					7-Aug-14	Fresh	37.431089 N	126.992468 E	MN136267	MN160125	MN147793
					7-Aug-14	Fresh	37.431089 N	126.992468 E	MN136268	MN160126	MN147794
					27-Jul-17	Fresh	37.661647 N	127.007675 E	MN136269	MN160127	MN147795
	Sonatini		*Hyalessa*	*maculaticollis*	18-Jul-16	Fresh	37.502129 N	127.020451 E	MN136270	MN160128	-
					18-Jul-16	Fresh	37.502129 N	127.020451 E	MN136271	MN160129	-
					2-Aug-14	Fresh	37.510082 N	127.076703 E	MN136272	MN160130	MN147796
					27-Jul-17	Fresh	37.661647 N	127.007675 E	MN136273	MN160131	MN147797
	Leptopsaltriini	Cicadina	*Leptosemia*	*takanonis*	2-Jul-12	Dry	35.905471 N	127.345198 E	MN136274	-	MN147798
					9-Jul-12	Dry	35.905471 N	127.345198 E	MN136275	MN160132	MN147799
					Jul-17	Fresh	35.406497 N	127.494247 E	MN136276	-	MN147800
	Dundubiini	Cosmopsaltriina	*Meimuna*	*mongolica*	3-Aug-16	Fresh	37.534074 N	126.905903 E	MN136277	MN160133	MN147801
					22-Aug-16	Dry	35.988303 N	126.976125 E	MN136278	MN160134	MN147802
					22-Aug-16	Dry	35.988303 N	126.976125 E	MN136279	MN160135	MN147803
				opalifera	19-Aug-16	Dry	37.805631 N	127.634019 E	MN136280	MN160136	MN147804
					19-Aug-16	Dry	37.805631 N	127.634019 E	MN136281	MN160137	MN147805
					2-Aug-14	Fresh	37.510082 N	127.076703 E	MN136282	MN160138	MN147806
Cicadettinae	Cicadettini		*Cicadetta*	*abscondita*	28-May-17	Fresh	37.581786 N	128.415894 E	MN136283	MN160139	MN147807
					28-May-17	Fresh	37.581786 N	128.415894 E	MN136284	MN160140	MN147808
					2-Jun-18	Fresh	37.581786 N	128.415894 E	-	MN160141	MN147809
			Kosemia	*yezoensis*	7-Aug-18	Fresh	37.230417 N	128.989454 E	MN136285	MN160142	MN147810
			Tettigetta	*isshikii*	1-Jul-16	Dry	37.172185 N	128.642995 E	MN136286	MN160143	MN147811
					Jun-15	Dry	38.138259 N	127.362734 E	MN136287	MN160144	MN147812
					5-Jun-17	Fresh	37.581786 N	128.415894 E	MN136288	MN160145	MN147813
					5-Jun-17	Fresh	37.581786 N	128.415894 E	MN136289	MN160146	MN147814

Table S2. Accession numbers for sequences downloaded from GenBank for each gene.

Gene	Species	GenBank Accession Number
COI	*Cryptotympana atrata*	JF700154, MT129448, MN784082
	Cicadetta calliope	KF919587
	Auritibicen japonicus	MG953025
	Platypleura kaempferi	FJ169046, FJ169043
	Suisha coreana	LC199299
	Leptosemia takanonis	MT129315, GQ527075
	Meimuna mongolica	KM576905, KM576917
	Hyalessa maculaticollis	KY860253, KY860254
EF1α	*A. intermedius*	LC127302
	C. atrata	MN241616
	T. isshikii	KT713561
	Platypleura kaempferi	GU343834, GU343868
	M. opalifera	LC414933, LC414932
	M. oshimensis	LC414923
	L. takanonis	MT599177 (unverified*)
	H. maculaticollis	KY929069 (unverified*)
	Graptopsaltria nigrofuscata	MN241621, OR480543 (unverified*)
18S rRNA	*C. atrata*	MG953178
	M. opalifera	MG953132
	G. nigrofuscata	MG953116
	H. maculaticollis	MG953163

According to NCBI's standard review process, sequences are labeled as unverified when the accuracy of submitted sequence data or annotations cannot be confirmed by GenBank staff. These sequences begin with 'unverified:' in the definition line and are not included in BLAST databases until the issues are resolved.

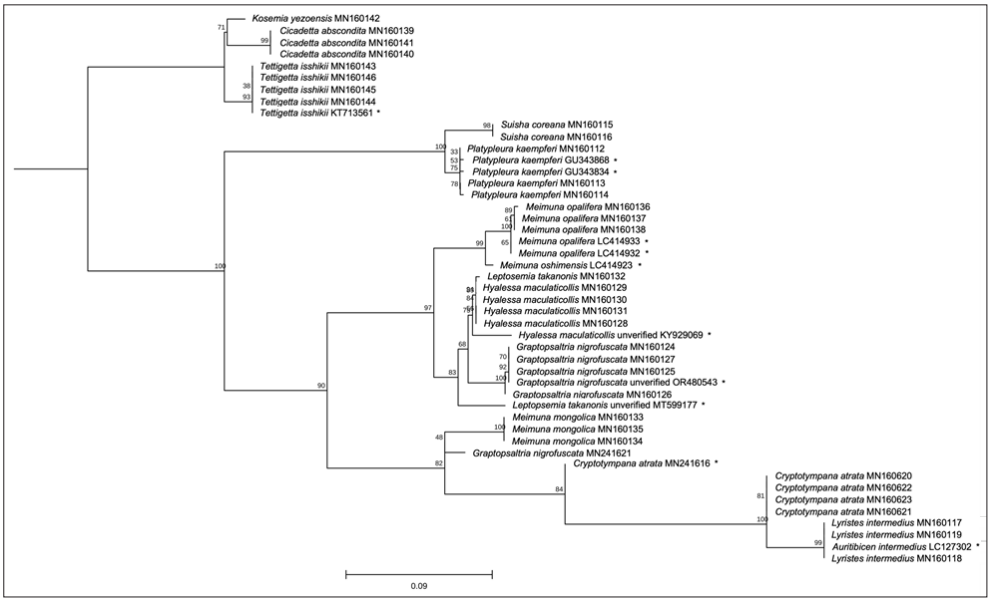

Figure S3. An extended majority-rule consensus maximum likelihood genealogy of the cicada species occurring in the Republic of Korea sampled in this study based on partial elongation factor 1-alpha (EF1α) locus. Values at the internal nodes represent node support estimated via 10,000 Ultrafast bootstrap replicates using IQ-TREE 2. Each tip label includes the species name and a GenBank accession number; an asterisk is added only for sequences retrieved from GenBank. The genealogy is polarized using an outgroup species *Kosemia yezoensis*. The scale bar below the tree measures branch length in the number of substitutions per site.

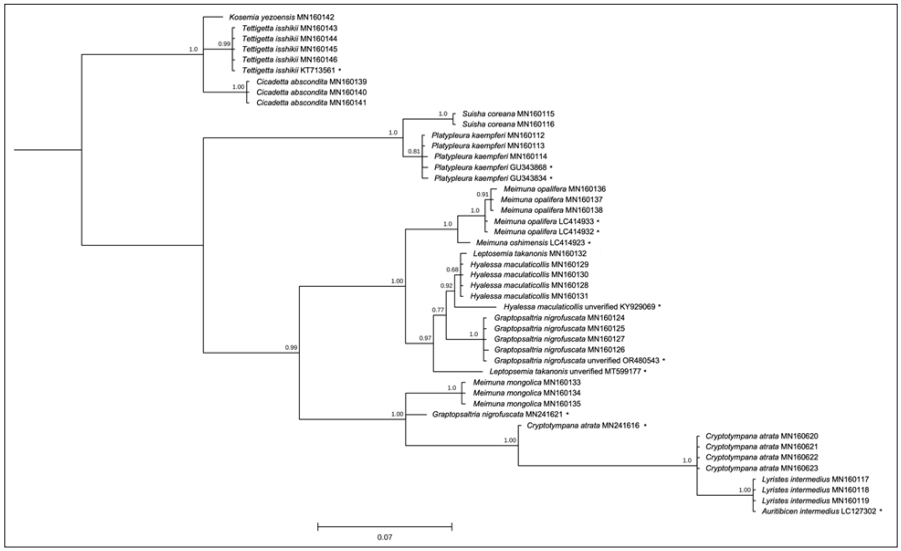

Figure S4. 50% majority rule consensus Bayesian inference genealogy of the cicada species occurring in the Republic of Korea sampled in this study plus available GenBank sequences added, based on partial elongation factor 1-alpha (EF1α) locus. Values at the internal nodes represent the posterior probability. Each tip label includes the species name and a GenBank accession number; an asterisk is added only for sequences retrieved from GenBank. The genealogy is polarized using an outgroup species *Kosemia yezoensis*. The scale bar below the tree measures branch length in the number of substitutions per site

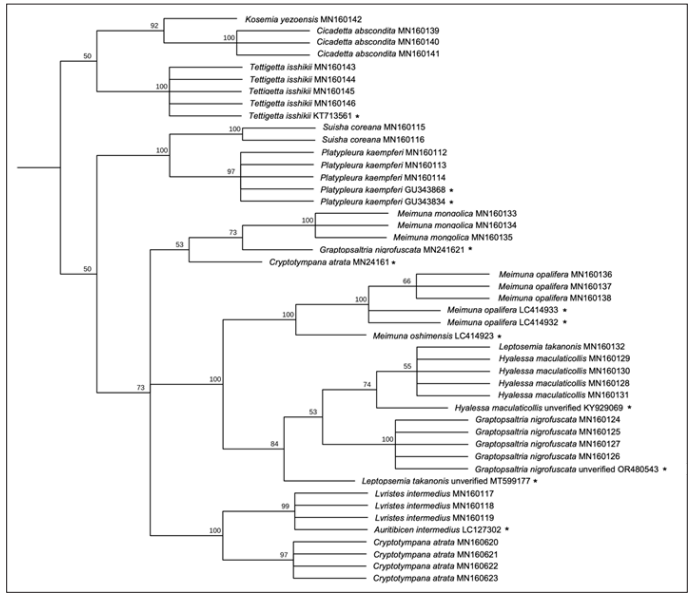

Figure S5. Strict consensus maximum parsimony genealogy of the cicada species occurring in the Republic of Korea sampled in this study plus available GenBank sequences added, based on partial elongation factor 1-alpha (EF1α) locus. Values at the internal nodes represent the nonparametric bootstrap support. Each tip label includes the species name and a GenBank accession number; an asterisk is added only for sequences retrieved from GenBank. The genealogy is polarized using an outgroup species *Kosemia yezoensis*.

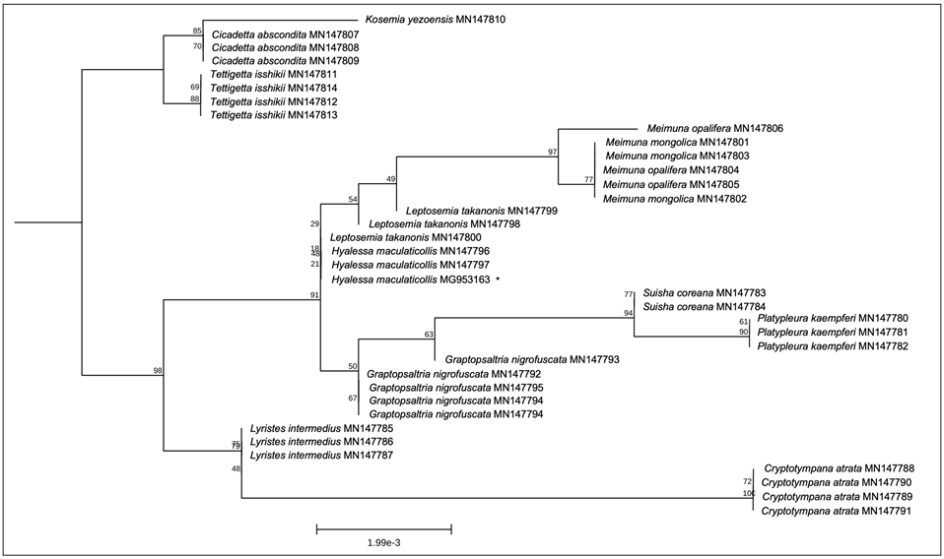

Figure S6. An extended majority-rule consensus maximum likelihood genealogy of the cicada species occurring in the Republic of Korea sampled in this study based on partial 18S rRNA locus. Values at the internal nodes represent node support estimated via 10,000 Ultrafast bootstrap replicates using IQ-TREE 2. Each tip label includes the species name and a GenBank accession number; an asterisk is added only for sequences retrieved from GenBank. The genealogy is polarized using an outgroup species *Kosemia yezoensis*. The scale bar below the tree measures branch length in the number of substitutions per site.

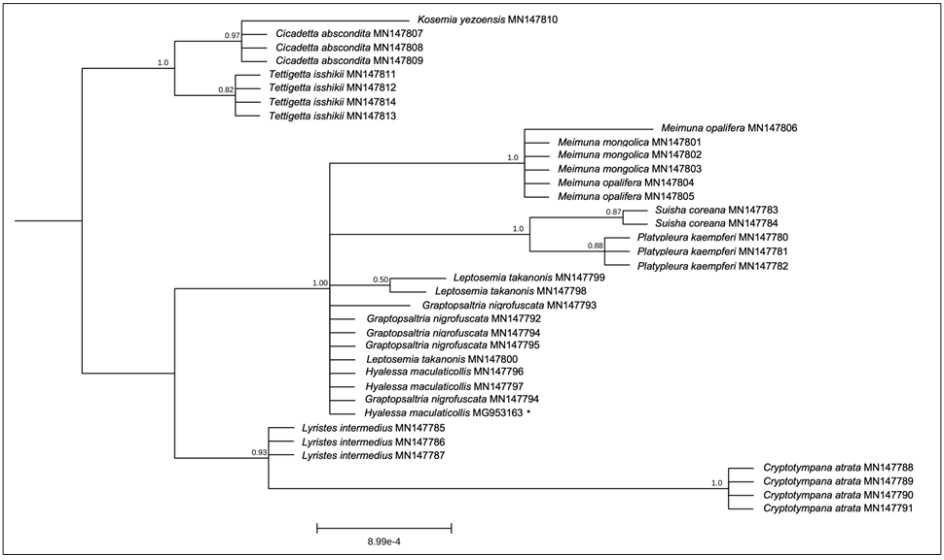

Figure S7. 50% majority rule consensus Bayesian inference genealogy of the cicada species occurring in the Republic of Korea sampled in this study plus available GenBank sequences added, based on partial 18S rRNA locus. Values at the internal nodes represent the posterior probability. Each tip label includes the species name and a GenBank accession number; an asterisk is added only for sequences retrieved from GenBank. The genealogy is polarized using an outgroup species *Kosemia yezoensis*. The scale bar below the tree measures branch length in the number of substitutions per site.

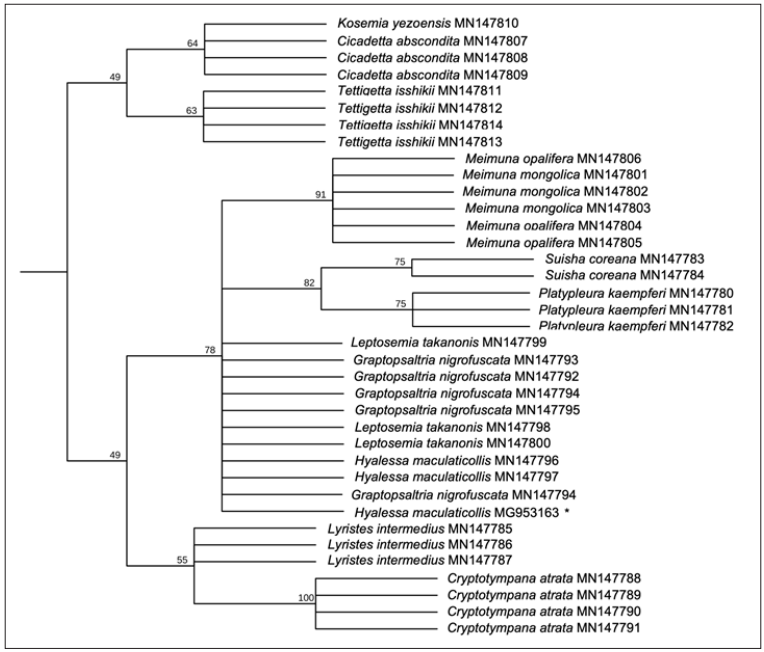

Figure S8. Strict consensus maximum parsimony genealogy of the cicada species occurring in the Republic of Korea sampled in this study plus available GenBank sequences added, based on partial 18S rRNA locus. Values at the internal nodes represent the nonparametric bootstrap support. Each tip label includes the species name and a GenBank accession number; an asterisk is added only for sequences retrieved from GenBank. The genealogy is polarized using an outgroup species *Kosemia yezoensis*.