# Mitochondrial Genomics Reveals Shared Phylogeographic Patterns and Demographic History among Three Periodical Cicada Species Groups

Zhenyong Du,[1] Hiroki Hasegawa,[2] John R. Cooley,[3] Chris Simon,[3] Jin Yoshimura,[4,5,6] Wanzhi Cai,[1] Teiji Sota,*[,2] and Hu Li*[,1]

[1]Department of Entomology and MOA Key Lab of Pest Monitoring and Green Management, College of Plant Protection, China Agricultural University, Beijing, China

[2]Department of Zoology, Graduate School of Science, Kyoto University, Sakyo, Kyoto, Japan

[3]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

[4]Graduate School of Science and Technology and Department of Mathematical and Systems Engineering, Shizuoka University, Hamamatsu, Japan

[5]Department of Environmental and Forest Biology, State University of New York College of Environmental Science and Forestry, Syracuse, NY

[6]Marine Biosystems Research Center, Chiba University, Kamogawa, Chiba, Japan

***Corresponding authors:** E-mails: tigerleecau@hotmail.com; sota@terra.zool.kyoto-u.ac.jp.

**Associate editor:** David Irwin

## Abstract

**The mass application of whole mitogenome (MG) sequencing has great potential for resolving complex phylogeographic patterns that cannot be resolved by partial mitogenomic sequences or nuclear markers. North American periodical cicadas (*Magicicada*) are well known for their periodical mass emergence at 17- and 13-year intervals in the north and south, respectively. *Magicicada* comprises three species groups, each containing one 17-year species and one or two 13-year species. Within each life cycle, single-aged cohorts, called broods, of periodical cicadas emerge in different years, and most broods contain members of all three species groups. There are 12 and three extant broods of 17- and 13-year cicadas, respectively. The phylogeographic relationships among the populations and broods within the species groups have not been clearly resolved. We analyzed 125 whole MG sequences from all broods and seven species within three species groups to ascertain the divergence history of the geographic and allochronic populations and their life cycles. Our mitogenomic phylogeny analysis clearly revealed that each of the three species groups had largely similar phylogeographic subdivisions (east, middle, and west) and demographic histories (rapid population expansion after the last glacial period). The mitogenomic phylogeny also partly resolved the brood diversification process, which could be explained by hypothetical temporary life cycle shifts, and showed that none of the 13- and 17-year species within the species groups was monophyletic, possibly due to gene flow between them. Our findings clearly reveal phylogeographic structures in the three *Magicicada* species groups, demonstrating the advantage of whole MG sequence data in phylogeographic studies.**

***Key words:*** **demographic history, life cycle polymorphism, *Magicicada*, mitogenome, parallel evolution, phylogeography.**

Article

## Introduction

Life history diversity underlies the tremendous biodiversity among organisms on our planet, and explaining how life history diversity arises is a challenge in evolutionary biology (Roff 2002). Periodical cicadas (*Magicicada* spp.) in the eastern United States are notable for possessing the longest juvenile period known among insects and for exhibiting more or less strict periodicity in adult emergence at 13- or 17-year intervals. These cicadas provide a unique opportunity for studying the evolution of extended periodical life cycles (Heliövaara et al. 1994; Karban 1997; Veller et al. 2015) and parallel adaptive evolution among closely related species (Sota et al. 2013; Koyama et al. 2016; Fujisawa et al. 2018). They also have been an important group for studying the roles of temporal isolation in speciation (Simon et al. 2000; Marshall and Cooley 2000; Cooley et al. 2001; Taylor and Friesen 2017) and of developmental plasticity in life cycle evolution (Martin and Simon 1990, Marshall et al. 2003; West-Eberhard 2003; Marshall et al. 2011, 2017).

The complex components of *Magicicada* life history are summarized in table 1. Three species groups are recognized (Decim, Cassini, and Decula), which all contain both 13- and 17-year species (Alexander and Moore 1962; Lloyd and Dybas 1966a; Simon 1988; Williams and Simon 1995). The ranges of

**Table 1.** Taxonomy, Life Cycles, Spatial/Temporal Distributions in the Eastern United States, Mitochondrial and Nuclear DNA Lineages, and Mitochondrial Phylogeographic Subdivisions of Periodical Cicadas (*Magicicada* spp.).

| Species Group | Species[a] | | Phenotype | Spatial/Temporal Distribution[b] | | Molecular Phylogeny[c] | |
|---|---|---|---|---|---|---|---|
| | | | Life Cycle Length | Regions in Eastern US | Emergence Year (Brood) | mt/nu DNA Major Lineage | Mitochondrial Phylogeographic Subdivision |
| Decim | *M. septendecim* | } | 17-Year | North | I–X, XIII–XIV | A | Ae, Am, Aw |
| | *M. neotredecim* | | 13-Year | Midwest | XIX, XXIII | A | Aw |
| | *M. tredecim* | | 13-Year | South | XIX, XXII–XXIII | B | B |
| Cassini | *M. cassini* | } | 17-Year | North | I–V, VIII–X, XIII–XIV | C | Ce, Cm, Cw |
| | *M. tredecassini* | | 13-Year | South | XIX, XXII–XXIII | C | Cw |
| Decula | *M. septendecula* | } | 17-Year | North | I–VI, VIII–X, XIII–XIV | D | De, Dm, Dw |
| | *M. tredecula* | | 13-Year | South | XIX, XXII–XXIII | D | (De), Dm, Dw |

NOTE.—"}" indicates a pair of species distinguishable only by different life cycle lengths. Recent emergence year of each brood: Brood I, 2012; Brood II, 2013; Brood III, 2014; Brood IV, 2015; Brood V, 2016; Brood VI, 2017; Brood VII, 2018; Brood VIII, 2019; Brood IX, 2020; Brood X, 2021; Brood XIII, 2024; Brood XIV, 2025 (17-year cicadas); Brood XIX, 2011; Brood XXII, 2014; Brood XXIII, 2015 (13-year cicadas).

[a]Alexander and Moore (1962) and Marshall and Cooley (2000).

[b]Simon (1988).

[c]Sota et al. (2013) and present study.

the two life cycles are separated parapatrically between generally northern (17-year) and generally southern (13-year) parts of the eastern US. The three species groups comprise seven species (with one 17-year species and one or two 13-year species in each group; Alexander and Moore 1962; Marshall and Cooley 2000). During an emergence year, a cohort or "brood" of adults of the same age appear in enormous numbers (Dybas and Lloyd 1974). The term brood is used to specify the emergence years (calendar years) of adult periodical cicadas. A brood generally contains all three species groups; hence, it is not a cohort of a single species. Currently, there are 12 broods of 17-year cicadas (I–X, XIII, and XIV) and three broods of 13-year cicadas (XIX, XXII, and XXIII; Simon 1988). In general, only a single brood occurs at one location. Therefore, different broods of the same life cycle are not only allochronic but also generally allopatric. The current broods are suggested to have arisen from an ancestral population by 4-year or, more rarely, 1-year shifts in emergence timing based on developmental plasticity (Lloyd and Dybas 1966b; Simon and Lloyd 1982; Martin and Simon 1988, 1990; Marshall et al. 2017; Cooley et al. 2018). Resolving the genetic relationships among different allochronic populations that occupy different areas and emerge as adults in different years (i.e., broods) is therefore essential for understanding the speciation process of periodical cicadas.

The phylogenetic relationships of *Magicicada* species groups, species, and broods have been studied most recently using various molecular markers, including mitochondrial genes (Sota et al. 2013), amplified fragment length polymorphisms (Sota et al. 2013), restriction site-associated DNA sequences (Koyama et al. 2016), and transcriptomic mRNA sequences (Fujisawa et al. 2018). Both mitochondrial and nuclear genes have clearly demonstrated early divergence of the three Decim species, with Cassini and Decula as sister groups. In the Decim group, the 13-year species *Magicicada tredecim* is sister to the other group containing *Magicicada septendecim* (17-year) and *Magicicada neotredecim* (13-year). However, no nuclear markers have yet resolved additional relationships such as those among broods or between 13- and 17-year species within species groups. Mitochondrial gene sequences from the *cox1-cox2* region showed phylogeographic subdivisions (e.g., east, middle, and west) within each species group, although the phylogeographic structure has remained ambiguous for the Decula group due to low sequence divergence. In addition, mitochondrial gene sequences have not resolved the relationships between 13- and 17-year species or broods of the same life cycles within the phylogenetic groups (Sota et al. 2013). Contrary to the results of phylogenetic analysis, the divergence history of life cycles within each species group inferred based on single-nucleotide polymorphisms (SNPs) from transcriptome sequence data has suggested that 13-year broods monophyletically diverged from 17-year broods 100–200 Ka and that gene flow (hybridization) occurred between 13- and 17-year cicadas in all pairs (except *M. tredecim*) of the three species groups (Fujisawa et al. 2018). With the exception of *M. tredecim*, life cycle cognate species (13- and 17-year species of the same species group; Cooley et al. 2001) are generally indistinguishable in morphology, mating signals, and behavior, suggesting an absence of intrinsic reproductive barriers. Thus, gene flow is possible; for example, during co-emergence along the border between 13- and 17-year broods (occurring once every 221 years for any given brood pair). If the inferred divergence history and hybridization between life cycles are correct, it is possible that mitochondrial introgression via hybridization may explain the unsorted phylogenetic relationships in the mitochondrial phylogeny.

The synchronized mass emergence of periodical cicadas at long intervals is considered advantageous in predator avoidance for all participating species and enhances reproductive success for each species (Lloyd and Dybas 1966a, 1966b; Yoshimura 1997). The two different life cycle lengths (i.e., 13 and 17 years) shared among the three species groups provide a unique case for studying genetic polymorphisms maintained in parallel among closely related species. More interestingly, the two life cycle lengths function to synchronize adult emergence of co-occurring species for a common fitness advantage. The divergence of the two life cycles may

have been related to adaptation to different climatic conditions (Cox and Carlton 1988; Yoshimura 1997); however, the genetic basis of the life cycle difference is still unknown (Fujisawa et al. 2018). In general, the control mechanisms, genetic backgrounds and evolutionary processes of long periodical life cycles in *Magicicada* and other organisms remain unclear.

To understand the process of population divergence in *Magicicada*, a more resolved phylogeny is needed. Given that genome-wide nuclear markers have provided little resolution to date (Koyama et al. 2016; Fujisawa et al. 2018), and that partial mitogenome (MG) sequences have provided appreciable resolution (Sota et al. 2013), one possible solution is to extend the latter to include whole MG sequence analysis. Compared with nuclear DNA, gene sequences involved in MGs have short coalescent times, unambiguous orthology, and generally rapid evolutionary rates in both vertebrates and invertebrates (Remi et al. 2017). These features make MGs crucial to the resolution of animal phylogeography (e.g., Batini et al. 2011; Bjork et al. 2011; Botero-Castro et al. 2013; Taylor et al. 2013). Whole MG sequences have much higher resolution than partial mitochondrial gene sequences, and recent single-species phylogeography studies have demonstrated the strong power of MG data in resolving the divergence patterns and histories of geographic populations (e.g., Morin et al. 2010; Ma et al. 2012; Hirase et al. 2016; Fields et al. 2018).

In this study, we applied next-generation sequencing to acquire whole MG sequences from 125 representative *Magicicada* samples selected based on our previous study (Sota et al. 2013), which included representatives from all broods and all seven species from three species groups. We aimed to resolve the phylogenetic relationships and phylogeographic structures of populations with different life cycles (13 or 17 years) and belonging to different broods, which were not clearly resolved in our previous study (Sota et al. 2013). We also estimated the divergence times and demographic histories to ascertain the impact of their population histories on the phylogeography of each species group. On the basis of the results of these analyses, we obtained evidence of shared phylogeographic structures among the three species, which could be expected from the evolution of adaptive synchronized mass emergence of cicadas from the three species groups. We also attempted to explain the brood divergence process within a phylogeographic group based on hypothetical brood shifting scenarios.

## Results

### Mitogenomic Features and Diversity

We obtained whole mitogenomic sequences with all 37 functional mitochondrial genes, except the control region, from 125 individuals of seven species and 15 broods of *Magicicada* (fig. 1; supplementary table S1, Supplementary Material online). These sequences differed in length among species groups: 14,434–14,436 bp for the Decim group (GC content, 23.6–23.7%); 14,554–14,864 bp for the Cassini group (GC content, 24.1%); and 14,453–14,454 bp for the Decula group (GC content, 23%). Length variation was found mainly in the noncoding regions between *trnY* and *trnC*. All of the sequences share the same gene arrangements with *Drosophila yakuba* (Clary et al. 1982).

We prepared two concatenated data sets for phylogenetic and population genetic analyses: the MG data set (total aligned length, 14,403 bp), which included sequences of the complete MG, except for the control region and the noncoding region between *trnY* and *trnC*; and the protein-coding gene (PCG) data set (total aligned length, 10,929 bp), which included sequences of all 13 PCGs.

On the basis of the MG data set, we detected 1,895 polymorphic sites (*S*) with 115 haplotypes (*H*) (table 2). Haplotype diversity (Hd) was 0.997, with a nucleotide diversity ($\pi$) of 0.0534 and average GC content of 23.6%. Among the species groups, the Decim group exhibited the highest genetic diversity ($\pi = 0.0062$), and the Decula group the lowest ($\pi = 0.0013$). Among species, *Magicicada cassini* exhibited the highest diversity, and *M. neotredecim* the lowest. Nucleotide diversity in each species was low ($\pi < 0.01$), and 13-year species had lower values of $\pi$ (0.0006–0.0012) compared with 17-year species (0.0013–0.0035). The sequence divergence among the three species groups, as measured by Nei's $D_A$ distance ($D_A$) and $F_{st}$, was substantial (table 3). Within species groups, however, sequence divergence was small between 13- and 17-year species ($D_A = 0.000006$–$0.000682$), except for that between *M. tredecim* and *M. septendecim* in the Decim group ($D_A = 0.014670$). $F_{st}$ between *M. tredecula* and *Magicicada septendecula* did not significantly differ from zero (table 3).

For nuclear genes, we obtained the combined 18S and 28S ribosomal RNA (rRNA) gene sequences for all specimens. However, we found no variation within each species group for this gene region (supplementary fig. S1, Supplementary Material online). Therefore, we used only mitogenomic sequences in the following analyses.

### Phylogenetic Relationships and Divergence Times

The maximum-likelihood (ML) and Bayesian inference (BI) trees as well as the time-calibrated BEAST tree resulting from the PCG and MG data sets showed similar topologies with high nodal supports in most of the pivotal nodes (fig. 2; supplementary figs. S2–S6, Supplementary Material online). We assigned major clade names according to Sota et al. (2013). The monophyly of the four lineages—two lineages of the Decim group (Decim A: all haplotypes of *M. septendecim* and *M. neotredecim*; Decim B: haplotypes of *M. tredecim*), the Cassini group (all haplotypes of *M. cassini* and *M. tredecassini*), and the Decula group (all haplotypes of *M. septendecula* and *M. tredecula*)—was well supported. In addition, three phylogeographic subdivisions were clearly revealed in Decim A (Ae, Am, and Aw), the Cassini group (Ce, Cm, and Cw), and the Decula group (De, Dm, and Dw) (fig. 2). The relationships among individual haplotypes are presented by median-joining networks (fig. 1), which clearly showed that the subdivisions within each species group correspond to geographic divisions and comprise haplotypes derived from a core haplotype or related haplotypes separated by small numbers of substitutions.

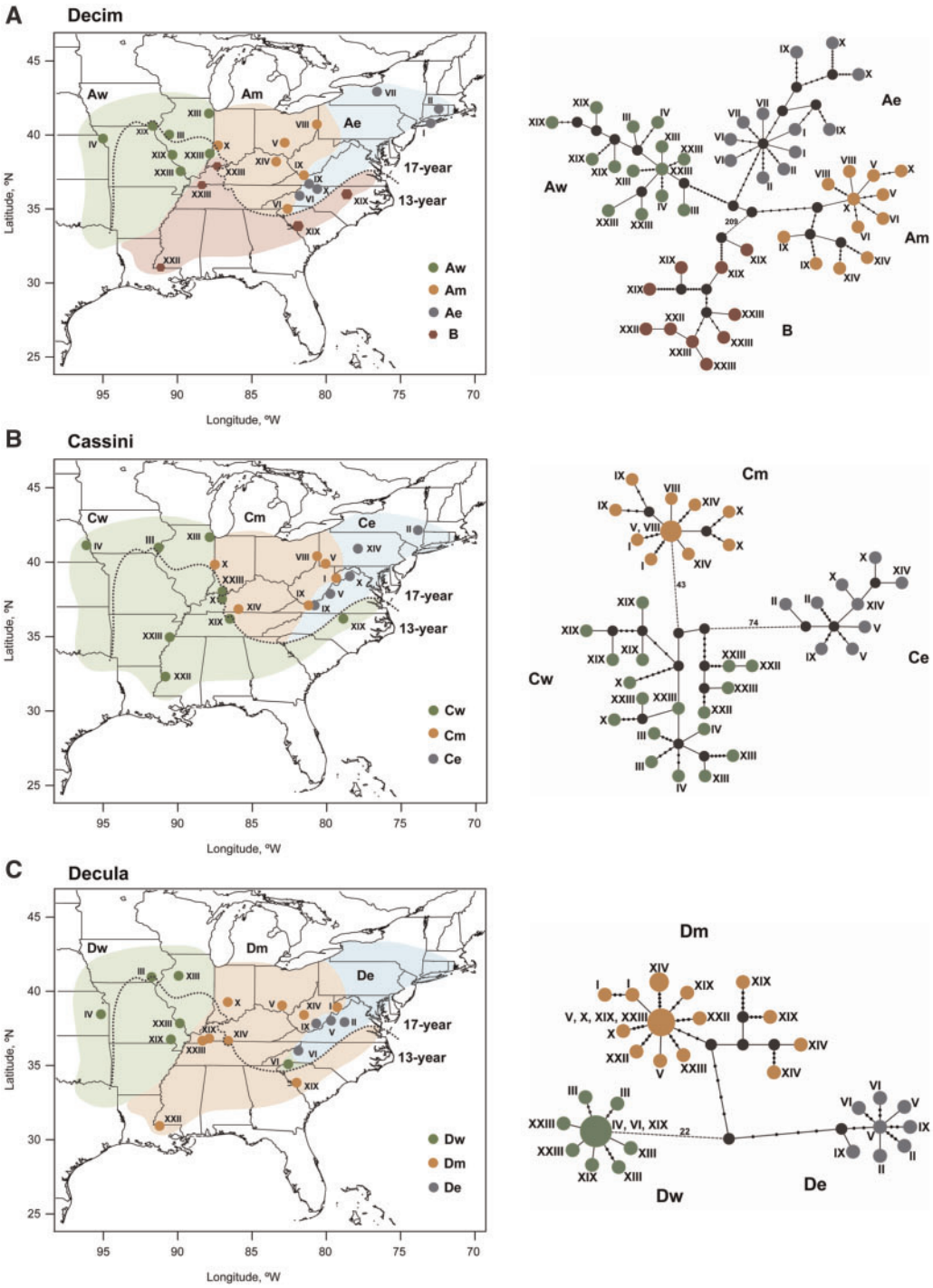


**Fig. 1.** Geographic distribution of samples and median-joining network of mitogenomic haplotypes in each species group. Different colors of the circles represent phylogeographic divisions. Brood numbers are indicated by Roman numerals next to the circles. Broken line indicates the boundary between 13- and 17-year cicadas. For the networks (right), colored circles represent haplotypes, and black circles represent missing haplotypes that were not observed. Solid lines between each of two linked haplotypes correspond to one mutation. Twenty mutations or less are represented by solid lines with small black dots. More than 20 mutations are indicated by dashed lines with numerals. The area of the circle is proportional to the number of haplotypes.

In Decim A, Aw included both 17-year broods (III, IV, and XIII; *M. septendecim*) and 13-year broods (XIX and XXIII; *M. neotredecim*), whereas Am and Ae contained only 17-year broods (V, VI, VIII, IX, X, and XIV in Am; I, II, VI, VII, IX, and X in Ae) (figs. 1–3). In the Cassini group, Cw included both 17-year broods (III, IV, X, and XIII) and 13-year broods (XIX, XXII, and XXIII), whereas Cm and Ce included only 17-year broods (I, V, VIII, IX, X, and XIV in Cm; II, V, IX, X, and XIV in Ce) (figs. 1, 2, and 4*A*). In the Decula group, Dw contained both 17-year broods (III, IV, VI, and XIII) and 13-year broods (XIX and XXIII), and Dm also included both 17-year broods (I, V, X, and XIV) and 13-year broods (XIX, XXII, and XXIII); however, De

**Table 2.** Genetic Diversity and Neutrality Test of the Four Major Lineages and Nine Phylogeographic Subdivisions of Periodical Cicadas.

| Group | *N* | *S* | *H* | Hd | $\pi$ | Fu' $F_s$ | Tajima's *D* |
|---|---|---|---|---|---|---|---|
| **Whole samples** | 125 | 1,895 | 115 | 0.997 | 0.0534 | 0.96 | 3.98 |
| **Decim group** | 48 | 324 | 48 | 1.000 | 0.0062 | −10.68** | 0.79 |
| **Cassini group** | 39 | 178 | 37 | 0.996 | 0.0033 | −7.67** | 0.47 |
| **Decula group** | 38 | 93 | 30 | 0.976 | 0.0013 | −6.56** | −0.61 |
| **Decim A group (13/17-y)** | 38 | 126 | 38 | 1.000 | 0.0014 | −22.69*** | −1.23 |
| ***M. tredecim* (13-y)** | 10 | 28 | 10 | 1.000 | 0.0006 | −4.06** | −0.78 |
| ***M. septendecim* (17-y)** | 30 | 105 | 30 | 1.000 | 0.0013 | −15.70*** | −1.08 |
| ***M. neotredecim* (13-y)** | 8 | 27 | 8 | 1.000 | 0.0006 | −2.70* | −1.28 |
| ***M. cassini* (17-y)** | 29 | 156 | 27 | 0.993 | 0.0035 | −3.16 | 1.01 |
| ***M. tredecassini* (13-y)** | 10 | 27 | 10 | 1.000 | 0.0007 | −3.69* | −0.12 |
| ***M. septendecula* (17-y)** | 26 | 73 | 21 | 0.975 | 0.0013 | −3.02 | −0.09 |
| ***M. tredecula* (13-y)** | 12 | 49 | 11 | 0.985 | 0.0012 | −1.25 | 0.45 |
| **Aw group (13/17-y)** | 14 | 39 | 14 | 1.000 | 0.0005 | −8.69*** | −1.94* |
| **Am group (17-y)** | 12 | 30 | 12 | 1.000 | 0.0005 | −6.61** | −1.43 |
| **Ae group (17-y)** | 12 | 41 | 12 | 1.000 | 0.0006 | −5.35** | −1.70 |
| **Cw group (13/17-y)** | 18 | 52 | 18 | 1.000 | 0.0007 | −10.71*** | −1.54 |
| **Cm group (17-y)** | 12 | 22 | 10 | 0.955 | 0.0003 | −4.60** | −2.03* |
| **Ce group (17-y)** | 9 | 15 | 9 | 1.000 | 0.0003 | −5.72** | −1.33 |
| **Dw group (13/17-y)** | 12 | 12 | 8 | 0.848 | 0.0001 | −4.02** | −2.09** |
| **Dm group (13/17-y)** | 18 | 37 | 14 | 0.954 | 0.0004 | −5.06** | −1.79 |
| **De group (17-y)** | 8 | 16 | 8 | 1.000 | 0.0003 | −4.61** | −1.81* |

NOTE.—*N*, sample size; *S*, number of polymorphic sites; *H*, number of haplotypes; $\pi$, nucleotide diversity; Hd, haplotype diversity.
*$P < 0.05$. **$P < 0.02$. ***$P < 0.001$.

**Table 3.** Sequence Divergence between Species Groups and Species.

| Comparison | $D_A$ | $F_{st}$ |
|---|---|---|
| **Decim vs. Cassini** | 0.082221 | 0.9492* |
| **Cassini vs. Decula** | 0.058879 | 0.9652* |
| **Decula vs. Decim** | 0.078477 | 0.9555* |
| **Decim A vs. Decim B (*M. tredecim*)** | 0.014650 | 0.9237* |
| ***M. tredecim* (13-y) vs. *M. neotredecim* (13-y)** | 0.015066 | 0.9646* |
| ***M. septendecim* (17-y) vs. *M. neotredecim* (13-y)** | 0.000682 | 0.3556* |
| ***M. septendecim* (17-y) vs. *M. tredecim* (13-y)** | 0.014670 | 0.9289* |
| ***M. tredecassini* (13-y)vs. *M. cassini* (17-y)** | 0.001390 | 0.3168* |
| ***M. tredecula* (13-y) vs. *M. septendecula* (17-y)** | 0.000006 | 0.0033$^{NS}$ |

$^{NS}P > 0.05$. *$P < 0.00001$.

contained only 17-year broods (II, V, VI, and IX) (figs. 1, 2, and 4*B*). The supports for nodes within phylogeographic subdivisions were generally low, and the relationships among broods were not resolved. Exceptionally, the brood relationships in lineage B (*M. tredecim*) were resolved with high node supports: Brood XIX was basal, from which Brood XXIII was derived followed by Brood XXII (figs. 2 and 3). The sequential mutation steps of haplotypes from Brood XIX to XXII are clearly shown by the median-joining network (fig. 1*A*).

Within the clades containing both 17- and 13-year broods (Aw, Cw, Dw, and Dm), different life cycles did not exhibit monophyly (figs. 3 and 4). Thus, the genetic differentiation between the two cycles, as measured by $F_{st}$, was low and did not differ significantly from zero (supplementary table S1, Supplementary Material online), with the exception of Cw, in which 17-year haplotypes were monophyletic except for two Brood X haplotypes (fig. 4*A*).

Our divergence time estimations (shown in figs. 2 and 3) suggest that the age of the most recent common ancestor (MRCA) of the Decim group was 0.42 Ma, which diverged into lineages A (*M. septendecim* and *M. neotredecim*) and B (*M. tredecim*). The MRCA age of lineage B was 18 Ka, whereas that of lineage A was 56 Ka. Within lineage A, the MRCA age of Aw and Am was 47 Ka, and those of Aw, Am, and Ae were 18, 19, and 23 Ka, respectively. In lineage B, Brood XXIII diverged from Brood XIX at 6.0 Ka, and Brood XXII diverged from Brood XXIII at 1.1 Ka.

The Cassini and Decula groups diverged from each other 2.23 Ma, in the early Pleistocene, but the MRCA times of the extant phylogeographic subdivisions in these species groups were 148 and 60 Ka, respectively, and thus more recent. In the Cassini group, Cw and Cm diverged 82 Ka, and the MRCA times of Cw, Cm, and Ce were 26, 11, and 14 Ka, respectively (fig. 4*A*). In the Decula group, De and Dm diverged 30 Ka, and the MRCA times of the Dw, Dm, and De groups were 10, 19, and 13 Ka, respectively (fig. 4*B*). Thus, all of the individual phylogeographic subdivisions and broods within the species groups diverged during the last glacial period or at the beginning of the Holocene.

## Evolution of Mitochondrial Gene Sequences

Under the one-ratio model, the $\omega$ (dN/dS) ratio was much smaller than unity for each of the four major groups, suggesting the action of negative (purifying) selection on mitochondrial genes (table 4). Comparisons of the one-ratio model with the two-ratio model based on the phylogeographic divergence scenario in each group did not support the latter model, suggesting a lack of changes in the selection regime in each group. These results, with $\omega$ much smaller than unity, suggest that the mitochondrial genes have been under purifying selection throughout their sequence divergence.

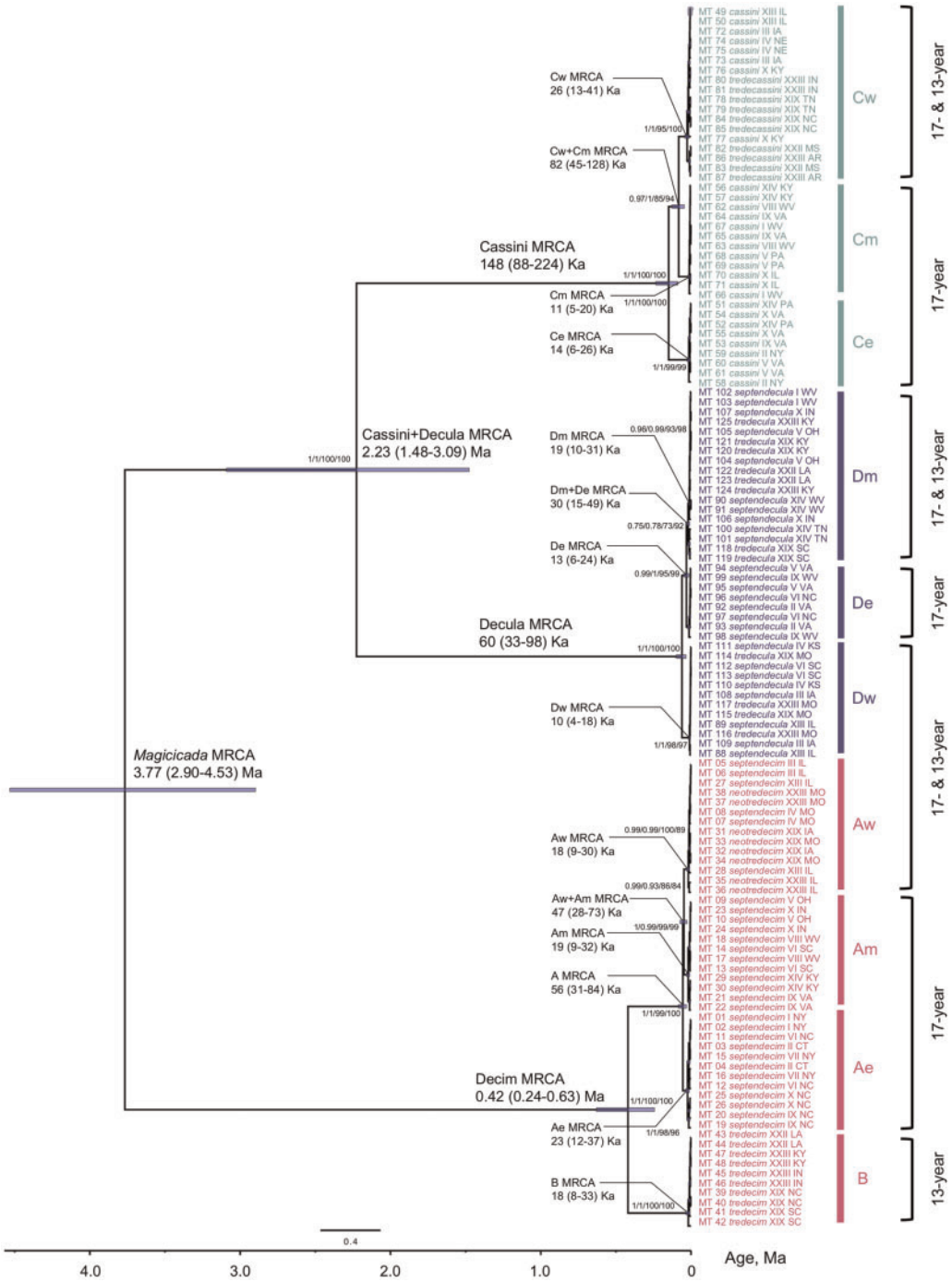


**Fig. 2.** Phylogenetic relationships of major nodes and divergence time based on the protein coding gene (PCG) data set. Bars in purple show 95% highest posterior density intervals. The four node supports shown are posterior probabilities of Bayesian inference and bootstrap percentages in maximum-likelihood (ML) analysis based on the PCG and mitogenome (MG) data sets. The OTU labels describe the sample number, species, brood number, and state of collection (supplementary table S1, Supplementary Material online). Phylogeographic subdivisions and life cycles are also shown.

## Demographic History

For the four distinct lineages (Decim A and B, and the Cassini and Decula groups) and clades within lineages, Fu's $F_s$ values were all significantly negative (table 2). Tajima's *D* did not show departure from zero (neutrality) for the four distinct lineages but was significantly negative for four of nine clades within Decim A, Cassini, and Decula (table 2). Given the higher sensitivity of Fu's $F_s$ values in detecting deviation from neutrality compared with Tajima's *D*, these results suggest that all lineages experienced population expansion. In the mismatch distribution analysis (fig. 5), observed and simulated values of neither sum of square deviations (*SSD*) nor Harpending's raggedness index (*r*) significantly differed from one another in any of the four groups (Decim A and B, Cassini, and Decula), indicating that the observed mismatch distributions fitted well with the simulated curves under the

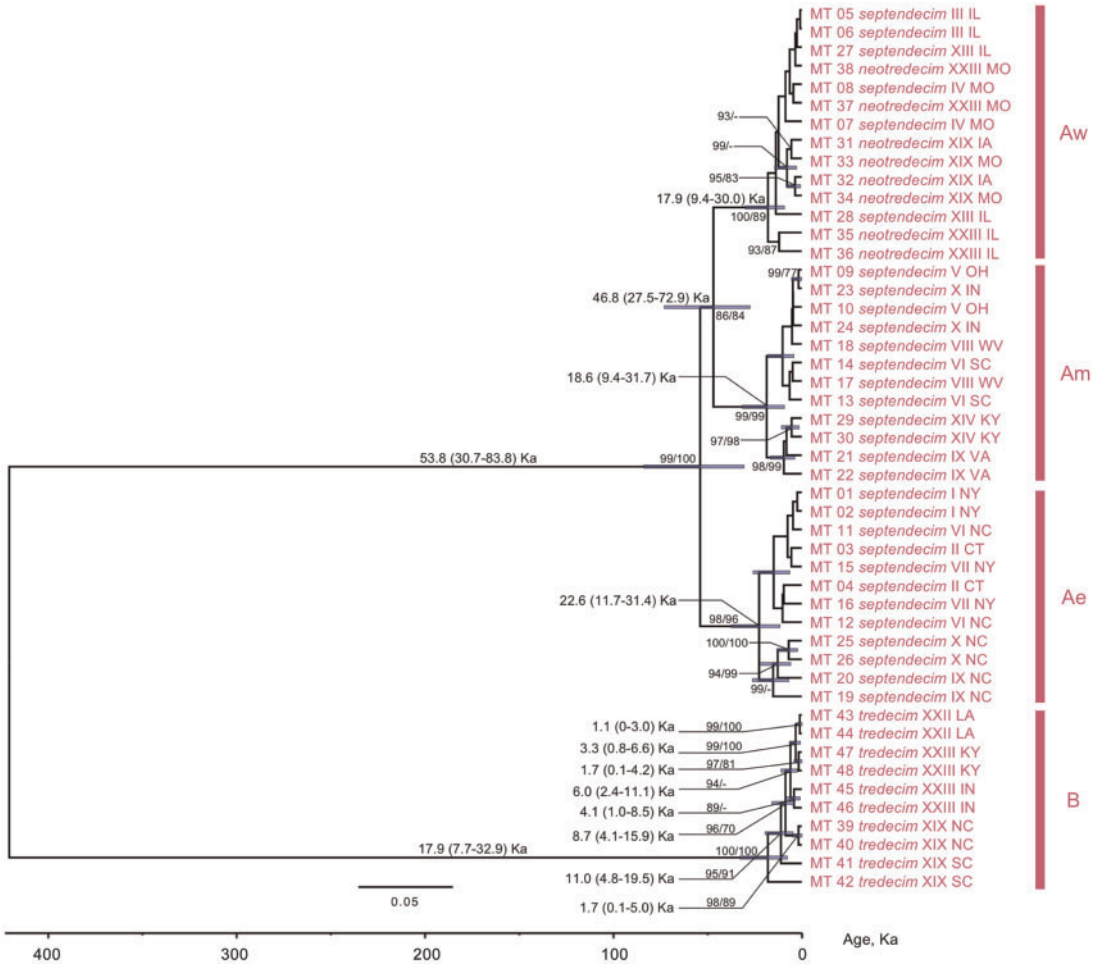


**Fig. 3.** Phylogenetic relationships in Decim. The two node supports shown are bootstrap percentages inferred from ML analysis (shown when > 70%) based on the PCG and MG data sets.

population expansion model. Lastly, Bayesian skyline plot (BSP) analysis revealed that the population sizes of Decim A, Cassini, and Decula all experienced a dramatic increase in effective population size from 10,000 years ago; that is, the beginning of the Holocene (fig. 5). However, the population size of Decim B (*M. tredecim*) was steadier, showing at most a slight increase during the Holocene.

## Discussion

Our data showed substantial genetic divergence among four major lineages (Decim A, Decim B, Cassini, and Decula) but rather small divergences between 13- and 17-year species within each of Decim A, Cassini, and Decula. These results are consistent with previous studies (Sota et al. 2013; Koyama et al. 2016; Fujisawa et al. 2018) and indicate that the periodical cicadas comprise four well-established species or species groups (i.e., *M. tredecim*, Decim A, Cassini, and Decula) and three pairs of 13- and 17-year populations, which have been recognized as different species by their allochronic and allopatric segregation despite the fact that they are indistinguishable morphologically and behaviorally (Alexander and Moore 1962; Marshall and Cooley 2000; Simon et al. 2000; Cooley et al. 2001). Our analysis revealed phylogeographic subdivisions within species groups that were not clearly resolved in previous studies. As the MG is a small haploid genome, differences in selection regimes on genes among lineages can result in gene trees that are not concordant with the true evolutionary history of the populations (Ballard and Whitlock 2004). However, because we found no evidence of a shift in selection regimes, we consider that the evolutionary pattern of the MG sequences were homogeneous among the different clades.

### Phylogeographic Subdivisions

Our phylogenetic data resolved three phylogeographic subdivisions (east, middle, and west) in each of the Decim A, Cassini, and Decula lineages with strong supports. All Decim B lineage cicadas (*M. tredecim*) are southern/central in distribution and do not show east, middle, or west subdivisions. In contrast to our previous study (Sota et al. 2013), our analysis revealed that the Decula group contains three phylogeographic subdivisions (east, middle, and west), comparable with those of the Decim and Cassini groups, and these subdivisions became separated from one another by a number of substitutions; our previous study detected only one substitution among the groups. The geographic barriers for dispersal imposed by the Appalachian Mountains and the Mississippi River may have contributed to the phylogeographic pattern

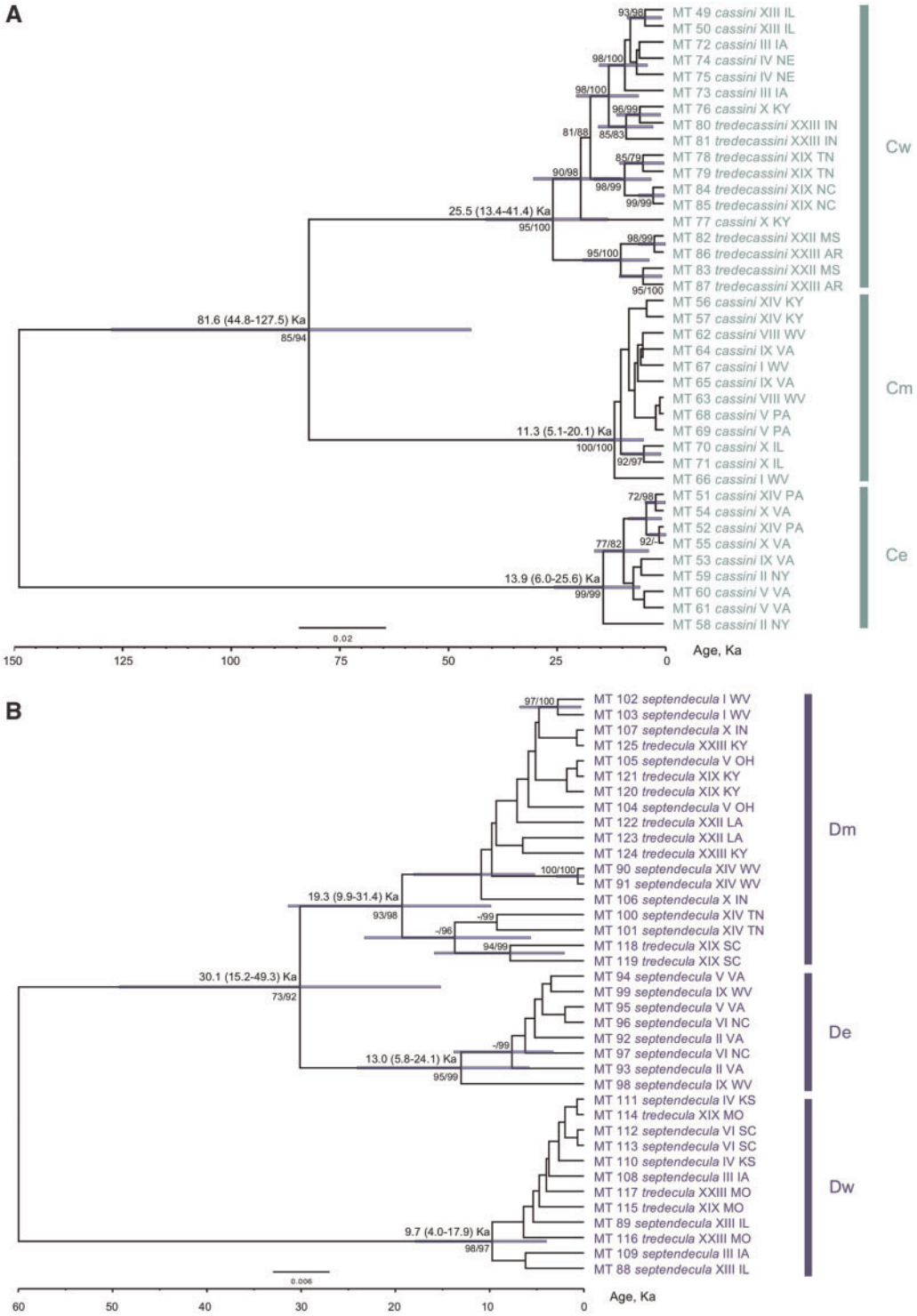


FIG. 4. Phylogenetic relationships in (*A*) Cassini and (*B*) Decula. The two node supports shown are bootstrap percentages inferred from ML analysis (shown when > 70%) based on the PCG and MG data sets.

of periodical cicadas, as in many other organisms in unglaciated eastern North America (Soltis et al. 2006). However, there are also differences in the phylogeographic subdivisions among species groups. The central and southern parts, the range of the Decim B lineage (13-cicadas, *M. tredecim*), are co-occupied by the western group Cw of the Cassini group and by the middle group Dm of the Decula group. The western phylogeographic subdivisions of the Decim group (Aw) and Decula group (Dw) also contain 13-year cicadas, although none of the eastern phylogeographic subdivision species group representatives (Ae, Ce, and De) includes cicadas with 13-year life cycles. (Note that our De group did not contain the 13-year *M. tredecula* species, unlike the previous De group in Sota et al. 2013, as a result of redefining the Decula phylogeographic subdivisions.) Thus, the central and western phylogeographic subdivisions differ from the east in

**Table 4.** Test for Heterogeneity in Selection Regime within Each Species Group.

| Group | $\omega$ Ratios (dN/dS) | | *P*-values of Likelihood Ratio Test |
|---|---|---|---|
| | One-ratio Model | Two-ratio Model (*B*/*F*) | |
| Decim | 0.114 | 0.081/0.149 | 0.082 |
| Decim A | 0.149 | 0.101/0.183 | 0.153 |
| Cassini | 0.075 | 0.098/0.049 | 0.078 |
| Decula | 0.118 | 0.110/0.123 | 0.819 |

NOTE.—*B*/*F* represents background branches with foreground branches.

hosting both 13-year and 17-year cicadas. These results suggest differences in the evolutionary history of 13-year cicadas among species groups.

The phylogenetic relationships of phylogeographic subdivisions within each species group are: (Ae, (Am, Aw)) in Decim A; (Ce, (Cm, Cw)) in Cassini; and ((De, Dm), Dw) in Decula. These relationships were not clearly resolved by Sota et al. (2013). The times of MRCAs for these phylogeographic subdivisions ranged between 26,000 and 10,000 years ago, from the last glacial period to the beginning of the Holocene. Because these phylogeographic subdivisions within the species groups originated 56–140 Ka (figs. 2 and 3), it is likely that the three subdivisions within the species groups have been derived from three different refugia during the last glacial period. We infer that rapid population expansion occurred in each species group except *M. tredecim* after the beginning of the Holocene. The timing of population expansion suggests that it was associated with rapid range expansion from refugia during the last glacial period, probably mainly by 17-year cicadas.

## Brood Diversification

Brood divergence promotes genetic differentiation in each species, because different broods are allochronically isolated from one another. However, a brood may not be permanent; brood shifts via occasional 4- or 1-year acceleration or deceleration of emergence appear to have resulted in brood diversity (Lloyd and Dybas 1966b; Simon and Lloyd 1982; Cooley et al. 2018). MG phylogenetic trees show that for each species, some broods occur in more than one phylogeographic subdivision and are substantially differentiated in mitochondrial DNA (see also Sota et al. 2013; Cooley et al. 2018). These polyphyletic broods include VI, IX, and X in Decim; V, IX, X, and XIV in Cassini; and V, VI, XIX, and XXIII in Decula. The polyphyly of broods may be attributed to phylogenetic divergence of long-persisting broods (i.e., one brood diverged into two phylogeographic subdivisions) and brood shifts that can result in the same broods forming independently in different phylogeographic regions. The polyphyly of broods may also be related to gene flow between 13- and 17-year cicadas.

Broods within each phylogeographic subdivision generally show shallow divergence, and it is difficult to clearly identify sister relationships among different broods. The only case for which we could unambiguously infer the history of brood diversification was in lineage B of the Decim group, which contains only *M. tredecim*. In this group, the most parsimonious reconstruction of brood divergence is that the ancestral brood is XIX; XXIII diverged from XIX via a 4-year shift, and XII diverged from XXIII via a 1-year shift. The robust topology of *M. tredecim* may have resulted from the more ancient origin of *M. tredecim* and the lack of hybridization with other species (Koyama et al. 2016).

## Divergence of 13- and 17-Year Life Cycles

The existence of three pairs of allochronic species with different prime number life cycles in *Magicicada* is enigmatic. It is questionable whether each of the paired, allochronic species that can be distinguished only by life cycle length (table 1) is different species according to the biological species concept. However, an interesting question of adaptive evolution is how the divergence of 13-year and 17-year life cycles occurred in response to selection pressures such as climatic change (Cox and Carlton 1988; Yoshimura 1997). The process of life cycle divergence is not clear. Except for the 13-year cicada *M. tredecim*, which diverged earlier (0.42 Ma; fig. 2) and likely has had a 13-year life cycle from the beginning, none of the 13-year and 17-year cicadas within species groups is monophyletic in any phylogenetic trees constructed using various markers (Sota et al. 2013; Koyama et al. 2016; Fujisawa et al. 2018). Only SNP-based analysis of demographic history (Fujisawa et al. 2018) has suggested that a single split resulting in the two life cycles occurred 200–100 Ka in each of the three species groups. Demographic inference suggests the occurrence of substantial gene flow between 13- and 17-year cicadas excluding *M. tredecim*, which might explain the polyphyly of life cycles in the phylogenetic trees (Fujisawa et al. 2018).

Our MG phylogeny reveals the polyphyly of 13- and 17-year life cycle species within each of four phylogeographic subdivisions that contain both life cycles (Aw, Cw, Dm, and Dw) with MRCA times of 18,000–26,000 years ago (the last glacial maximum), which are inconsistent with the divergence times of life cycles estimated by Fujisawa et al. (2018). If the life cycle divergence with gene flow scenario suggested by the demographic inference of Fujisawa et al. (2018) is correct, then the polyphyly of life cycle could be attributed to repeated gene flow between the 13- and 17-year cicadas (particularly involving the mitochondrial genes) after the last glacial maximum, as suggested by the above MRCA times. An alternative scenario for the life cycle divergence is a recent life cycle switching of 17-year cicadas to 13-year cicadas after the last glacial maximum. This scenario was proposed for the evolution of *M. neotredecim* (Martin and Simon 1988, 1990; Marshall and Cooley 2000; Simon et al. 2000), and the mechanism of life cycle switching was hypothesized to be the 4-year acceleration of the life cycle by developmental plasticity and subsequent genetic canalization of the altered life cycle (Cooley et al. 2001). This scenario may explain the polyphyly of life cycles in the mitogenomic phylogeny if life cycle switching has occurred in two or more 17-year cicada populations with different mitogenomic haplotypes. However, comparable evidence for parallel divergence of 13-year cicadas from

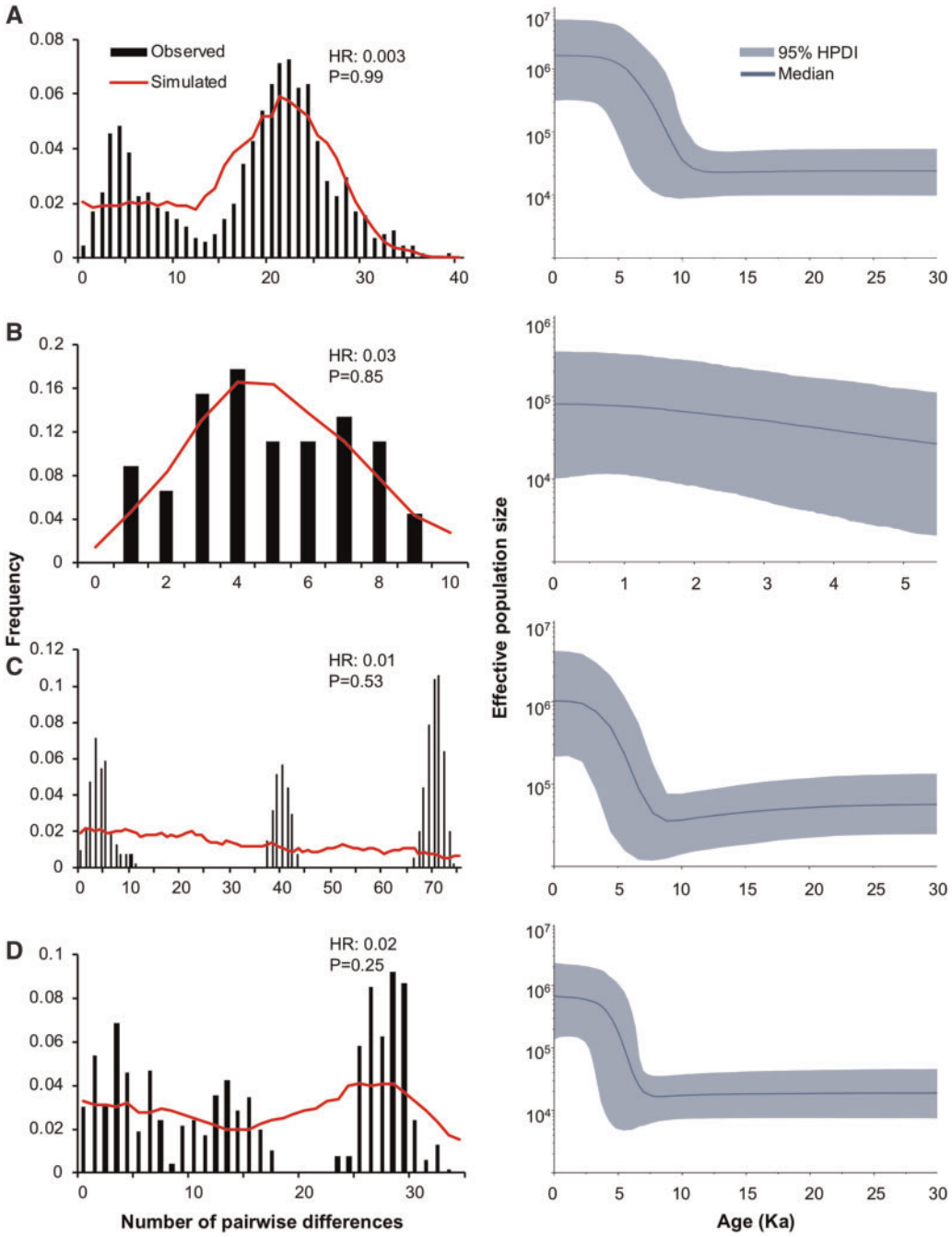


FIG. 5. Mismatch distributions (left) and Bayesian skyline plots (right). (*A*) Decim A, (*B*) Decim B (*Magicicada tredecim*), (*C*) Cassini, and (*D*) Decula were each calculated based on the PCG data set. The observed mismatch distribution is denoted by vertical bars, and the expected distribution under the population expansion model is represented by red lines. Harpending's raggedness indices are shown. For Bayesian skyline plots, the median estimated effective population sizes (middle lines) are enclosed within the 95% highest posterior density intervals (shaded areas).

different 17-year cicada lineages has not been obtained from nuclear markers due to the lack of definite genomic divergence associated with life cycle differences (Koyama et al. 2016; Fujisawa et al. 2018). To resolve life cycle divergence history during at least the past 200,000 years (the most ancient life cycle divergence time proposed by Fujisawa et al. 2018), an integrated analysis with more extensive MG data, combined with geographic context in terms of ecological niche modeling, may be useful, and it is essential to resolve the genetic basis of these life cycle differences.

## Implications for MG Phylogeography

In this study, we applied whole MG phylogenetics to a group of closely related species for the first time. The periodical cicada species groups that we studied have evolved many biological features in parallel (or shared as ancestral polymorphisms), and their phylogeographic patterns are mutually related, raising complex questions to be solved. Additionally, divergences in life cycle and spatial and temporal distribution appear to be recent and the divergence of their nuclear DNA sequences is essentially unresolved except

for the major lineages. Our research findings demonstrate that whole MG data are a powerful tool that can fill the gap of genetic information in such difficult cases. Recently developed sequencing technology allows the generation of hundreds of whole MG sequences in a cost-effective manner (Gillett et al. 2014; Tang et al. 2014; Crampton-Platt et al. 2016), facilitating the use of large whole MG data sets instead of a few gene sequences. It is important to appreciate the limitations of information obtainable from MG sequences; they represent only one nonrecombining haplotype sequence that reflects only the maternal lineage. We predict that combining whole MG sequence data with a collection of nuclear gene sequences (such as restriction site-associated DNA sequences), which are also obtainable from high-throughput sequencing, will be a common approach in the future.

## Conclusions

On the basis of the well-resolved phylogenetic relationships among MG haplotypes, we showed that the three species groups of *Magicicada* (except for *M. tredecim* in the Decim group, which diverged earlier) had similar east, middle and west phylogeographic subdivisions, which originated from three refugia during the last glacial period. The populations of the three phylogeographic divisions had similar demographic histories, in which rapid population expansion occurred after the last glacial period. This parallel divergence and shared demographic history was not clearly shown in previous studies based on partial mitochondrial gene sequences. We also revealed that broods within species groups are often polyphyletic in terms of their mitogenomic lineages, possibly as a result of geographic differentiation, brood shifts or gene flow between the 13- and 17-year cicadas. In terms of brood shifts, we found that the differentiation of three broods of *M. tredecim* could be explained by hypothetical temporary 4- and 1-year life cycle shifts. Lastly, each of the 13- and 17-year cicadas (species) within a species group was not monophyletic, possibly as a result of gene flow between the 13- and 17-year cicadas. These findings contribute to our understanding of the processes underlying the complex evolutionary history of *Magicicada* populations, which have occurred in parallel among the three lineages tied to one another, due to the great advantage of synchronous mass emergence of adult cicadas.

## Materials and Methods

### Sampling and Genomic DNA Extraction

We used 125 specimens from seven species, which were a subset of the specimens used in Sota et al. (2013) (fig. 1; supplementary table S2, Supplementary Material online). These specimens represented haplotypes of 15 broods based on the sequences of the mitochondrial *cox1-cox2* gene region. As a check, we used two replicate individuals for each sample site/brood/species, except for *M. cassini* BroodIX from Virginia. Total genomic DNA was extracted from the muscles of ethanol-fixed specimens using a DNeasy Blood and Tissue kit (QIAGEN).

### MG Sequencing and Assembly

A 350-bp paired-end library was constructed from each specimen and sequenced to obtain ~3 Gb of data using an Illumina Hiseq X Ten platform at BerryGenomics (China). Raw reads were trimmed of adapters using Trimmomatic (Bolger et al. 2014). De novo assemblies for each sample were generated using IDBA-UD (Peng et al. 2012). BLAST searches with at least 98% similarity were conducted to obtain best-fit mitochondrial scaffolds, using the available sequences of *cox1-cox2* as the "bait" sequence reference (Sota et al. 2013). To investigate the accuracy of the assembly, clean reads were mapped onto the obtained mitochondrial scaffold in each library using Geneious 10.1.3 (http://www.geneious.com/), allowing for up to 2% mismatches, a maximum gap size of 2 bp, and minimum overlap of 40 bp.

For each mitochondrial scaffold, annotation of transfer RNA (tRNA) genes was performed using MITOS webservers (Bernt et al. 2013). PCGs and rRNA genes were identified by alignment with homologous genes of the *M. tredecim* MG (GenBank accession number KM000130). Although we could not determine sequences for the control region due to the presence of long tandem repeats, we successfully obtained and annotated complete sequences from *trnI* to 12S rRNA for all samples, which were used in the analyses. We also attempted to assemble nuclear 18S and partial 28S rRNA gene sequences in each library, referring to the available sequences of 18S rRNA for *Magicicada* (Sota et al. 2013) and the most conserved region of the aligned 28S rRNA gene sequences in Cicadidae. All sequences thus obtained have been deposited at GenBank (supplementary table S2, Supplementary Material online).

### Phylogenetic Analyses

Sequences of the MG and those of 18S and 28S rRNA genes for the 125 periodical cicada samples were aligned using the MAFFT 7.0 online server with the G-INS-i strategy (Katoh and Standley 2013). PCG sequences were also aligned by individual genes based on codon-based multiple alignments using the MAFFT algorithm implemented in the TranslatorX online platform (Abascal et al. 2010). All alignments were verified in Geneious for quality, and sequences of the noncoding region between *trnY* and *trnC* were removed because of ambiguously aligned sites. We prepared two concatenated data sets for further analyses: 1) the MG data set, which included sequences of the complete MG except for the control region and noncoding region between *trnY* and *trnC*; and 2) the PCG data set, which included sequences of all 13 PCGs.

Phylogenetic trees were constructed based on the above two data sets by BI using MrBayes 3.2.2 (Ronquist and Huelsenbeck 2003) and the ML method using IQ-TREE 1.6.5 (Trifinopoulos et al. 2016). The optimal partitioning scheme and substitution models were selected by PartitionFinder2 (Lanfear et al. 2012). We created input configuration files containing different predefined partitions for each data set: 1) 43 gene partitions for the MG data set (39 codon positions for PCGs, two partitions for rRNAs, one partition for tRNAs, and one partition for noncoding regions); and 2) 39 ($13 \times 3$) codon partitions for the PCG data set. We used the "greedy"

algorithm with branch lengths estimated as "unlinked" and the Akaike information criterion to search for the best-fit partitioning schemes and substitution models (see supplementary table S3, Supplementary Material online). Bootstrap ML analyses with 1,000 replicates were performed using the ultrafast bootstrap approximation approach (Minh et al. 2013) implemented in IQ-TREE. For MrBayes analyses, two simultaneous Markov chain Monte Carlo (MCMC) runs of 2 million generations were conducted for the two data sets; trees were sampled every 1,000 generations, and the first 25% of data was discarded as burn-in. Stationarity was considered to be reached when the average standard deviation of split frequencies was below 0.01 (Huelsenbeck et al. 2001).

Median-joining networks (Bandelt et al. 1999) of mitogenomic haplotypes based on the MG data set were constructed separately for three species groups using Network 5 (www.fluxus-engineering.com).

### Genetic Diversity, Population Differentiation, and Selection Pressure

To assess how genetic diversity varied across geographic populations, the number of polymorphic sites ($S$), the number of haplotypes ($H$), nucleotide diversity ($\pi$), haplotype diversity (Hd), and the average number of nucleotide differences were calculated to estimate DNA polymorphism using DnaSP 5.10 (Librado and Rozas 2009). To check the differentiation levels in different groups and lineages, we used Arlequin 3.5 (Excoffier and Lischer 2010) to calculate $F_{st}$ and MEGA7 (Kumar et al. 2016) to calculate Nei's $D_A$ genetic distance.

We tested for heterogeneity in the evolutionary pattern of mitochondrial gene sequences among lineages as follows. The ratio ($\omega$) of nonsynonymous (dN) to synonymous substitutions (dS) is normally used to estimate the evolutionary rates or natural selection pressure on coding regions. We used the CodeML program of PAML 4 (Yang 2007) through the PAML-X graphical interface (Xu and Yang 2013) to calculate $\omega$ (dN/dS) and to test the potential divergences of selection pressure using the PCG data set. We performed this test to determine whether the selective force had changed following the divergence of the phylogeographic lineages in each species group based on the tree topology obtained in the phylogenetic analyses. We considered four cases: 1) divergence between the mitochondrial lineages A ($=$ Ae $+$ Am $+$ Aw) and B in the Decim group; 2) between Aw and Ae $+$ Am in the Decim group; 3) between Ce and Cm $+$ Cw in the Cassini group; and 4) between De $+$ Dm and Dw in the Decula group (foreground and background branches were assigned using these orders). We hypothesized that phylogeographic divergence may have been associated with changes in selection pressure due to an associated shift of climatic zones. Using the branch model of CodeML, we conducted the M0 (one-ratio) model to obtain a $\omega$ value common for all branches. Then, using the two-ratio model, we estimated two $\omega$ values for the foreground and background branches as defined above. The statistical difference between the one- and two-ratio models was tested using a likelihood ratio test with Chi-square statistics in PAML.

### Divergence Time Estimation

We performed divergence time estimation with BEAST 2.4.8 (Bouckaert et al. 2014) using the PCG data set. Although there are fossils of Cicadettinae, the subfamily containing *Magicicada* (Moulds 2018), these are too ancient (34–8 Ma) and cannot be applied directly to our phylogenetic tree for calibration. Therefore, we used the age previously estimated at the node of the MRCA of all *Magicicada*: 3.89 Ma (95% highest posterior density intervals: 3.08–4.69 Ma; Sota et al. 2013). This MRCA age and its credible intervals were obtained by applying a general time-dependent molecular clock of insect mitochondrial genes (Papadopoulou et al. 2010) to mitochondrial gene sequence data (see Sota et al. 2013 for details); they were also used for transcriptome phylogeny divergence time estimation (Fujisawa et al. 2018). We used the best predicted partitioning scheme (supplementary table S3, Supplementary Material online) and the birth-death process of the speciation model for the tree prior, with trees and clock models linked and the site models unlinked. We used the uncorrelated lognormal relaxed clock model based on comparison with the strict clock model and set the node age prior for the MRCA of *Magicicada* as a normal distribution (mean $=$ 3.89, standard deviation $=$ 0.41). We performed two independent MCMC runs for 300 million generations with tree sampling every 1,000 generations. Tracer 1.7 (Rambaut et al. 2018) was used to verify whether the MCMC runs had reached a stationary distribution based on the effective sample sizes (ESS) of each estimated parameter, where we required ESS for the posterior, prior, and tree likelihood to be at least 200. We heuristically removed the first 25% of data for each run as burn-in, and the resulting trees for each replicate were combined using LogCombiner 1.5.3. TreeAnnotator 1.5.3 was used to calculate the consensus tree and to annotate the divergence times.

### Demographic History

We applied multiple methods to explore the demographic history of different lineages, including the neutral test using Tajima's $D$ (Tajima 1989) and Fu's $F_s$ (Fu 1997) implemented in Arlequin. In addition, we performed a mismatch distribution analysis in Arlequin, with 1,000 bootstrap replicates to detect expansion through the linear fitting of observed and simulated curves with the significance of two parameters, *SSD* and $r$ (Excoffier and Lischer 2010). BSP analysis was also applied to the PCG data set using the best partitioning scheme for each group (supplementary table S3, Supplementary Material online), under a strict clock model with a clock rate of 0.0239, which was evaluated above in the time estimation analysis in BEAST using a strict clock model. BEAST analysis was conducted for 300 million generations with sampling trees every 1,000 generations. We determined the effective population size through time using Tracer, discarding the initial 10% of generations as burn-in.

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Acknowledgments

This study was supported by grants from the National Natural Science Foundation of China (nos. 31730086, 31420103902, and 31772498), JSPS KAKENHI (22255004, 26257405 to J.Y.), and SPIRITS at Kyoto University (to T.S.). C.S. and J.R.C. were supported by the National Science Foundation (NSF) Division of Environmental Biology (DEB; no. 1655891) and a University of Connecticut Vice President Research Excellence grant.

## References

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38(Suppl 2):W7–W13.

Alexander RD, Moore TE. 1962. The evolutionary relationships of 17-year and 13-year cicadas, and three new species (Homoptera, Cicadidae, *Magicicada*). *Misc Publ Mus Zool Univ Mich.* 121:1–59.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol Ecol.* 13(4):729–744.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 16(1):37–48.

Batini C, Lopes J, Behar DM, Calafell F, Jorde LB, van der Veen L, Quintana-Murci L, Spedini G, Destro-Bisol G, Comas D. 2011. Insights into the demographic history of African pygmies from complete mitochondrial genomes. *Mol Biol Evol.* 28(2):1099–1110.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol.* 69(2):313–319.

Bjork A, Liu W, Wertheim JO, Hahn BH, Worobey M. 2011. Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. *Mol Biol Evol.* 28(1):615–623.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

Botero-Castro F, Tilak MK, Justy F, Catzeflis F, Delsuc F, Douzery E. 2013. Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). *Mol Phylogenet Evol.* 69(3):728–739.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 10(4):e1003537.

Crampton-Platt A, Yu DW, Zhou X, Vogler AP. 2016. Mitochondrial metagenomics: letting the genes out of the bottle. *Gigascience* 5:15.

Clary DO, Goddard JM, Martin SC, Fauron CM, Wolstenholme DR. 1982. *Drosophila* mitochondrial DNA: a novel gene order. *Nucleic Acids Res.* 10(21):6619–6637.

Cooley JR, Simon C, Marshall DC, Slon K, Ehrhardt C. 2001. Allochronic speciation, secondary contact, and reproductive character displacement in periodical cicadas (Hemiptera: *magicicada* spp.): genetic, morphological, and behavioural evidence. *Mol Ecol.* 10(3):661–671.

Cooley JR, Arguedas N, Bonaros E, Bunker G, Chiswell SM, DeGiovine A, Edwards M, Hassanieh D, Haji D, Knox J, et al. 2018. The periodical cicada four-year acceleration hypothesis revisited and the polyphyletic nature of Brood V, including an updated crowd-source enhanced map (Hemiptera: Cicadidae: *Magicicada*). *PeerJ* 6:e5282.

Cox RT, Carlton CE. 1988. Paleoclimatic influences in the evolution of periodical cicadas (Insecta: Homoptera: Cicadidae: *Magicicada* spp.). *Am Midl Nat.* 120(1):183–193.

Dybas HS, Lloyd M. 1974. The habitats of 17-year periodical cicadas (Homoptera: Cicadidae: *Magicicada* spp.). *Ecol Monogr.* 44(3):279–324.

Excoffier L, Lischer H. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10(3):564–567.

Fields PD, Obbard DJ, McTaggart SJ, Galimov Y, Little TJ, Ebert D. 2018. Mitogenome phylogeographic analysis of a planktonic crustacean. *Mol Phylogenet Evol.* 129:138–148.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147(2):915–925.

Fujisawa T, Koyama T, Kakishima S, Cooley JR, Simon C, Jin Y, Sota T. 2018. Triplicate parallel life cycle divergence despite gene flow in periodical cicadas. *Commun Biol.* 1:26.

Gillett CPDT, Crampton-Platt A, Timmermans MJTN, Jordal BH, Emerson BC, Vogler AP. 2014. Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Mol Biol Evol.* 31(8):2223–2237.

Heliövaara K, Väisänen R, Simon C. 1994. Evolutionary ecology of periodical insects. *Trends Ecol Evol.* 9(12):475–480.

Hirase S, Takeshima H, Nishida M, Iwasaki W. 2016. Parallel mitogenome sequencing alleviates random rooting effect in phylogeography. *Genome Biol Evol.* 8(4):1267–1278.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294(5550):2310–2314.

Karban R. 1997. Evolution of prolonged development: a life table analysis for periodical cicadas. *Am Nat.* 150(4):446–461.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.

Koyama T, Ito H, Fujisawa T, Ikeda H, Kakishima S, Cooley JR, Simon C, Yoshimura J, Sota T. 2016. Genomic divergence and lack of introgressive hybridization between two 13-year periodical cicadas support life cycle switching in the face of climate change. *Mol Ecol.* 25(21):5543–5556.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 33(7):1870–1874.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol.* 29(6):1695–1701.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452.

Lloyd M, Dybas HS. 1966a. The periodical cicada problem. I. Population ecology. *Evolution* 20(2):133–149.

Lloyd M, Dybas HS. 1966b. The periodical cicada problem. II. Evolution. *Evolution* 20(4):466–505.

Ma C, Yang P, Jiang F, Chapuis MP, Shali Y, Sword GA, Kang L. 2012. Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust. *Mol Ecol.* 21(17):4344–4358.

Marshall DC, Cooley JR. 2000. Reproductive character displacement and speciation in periodical cicadas, with description of a new species, 13-year *Magicicada neotredecim*. *Evolution* 54(4):1313–1325.

Marshall DC, Cooley JR, Hill K. 2011. Developmental plasticity of life-cycle length in thirteen-year periodical cicadas (Hemiptera: Cicadidae). *Ann Entomol Soc Am.* 104(3):443–450.

Marshall DC, Cooley JR, Simon C. 2003. Holocene climate shifts, life cycle plasticity, and speciation in periodical cicadas: a reply to Cox and Carlton. *Evolution* 57(2):433–437.

Marshall DC, Hill KBR, Cooley JR. 2017. Multimodal life-cycle variation in 13- and 17-year cicadas (Hemiptera: Cicadidae: *Magicicada*). *J Kansas Entomol Soc.* 90(3):211–226.

Martin AP, Simon C. 1988. Anomalous distribution of mitochondrial and nuclear gene markers in periodical cicadas. *Nature* 336:247–249.

Martin AP, Simon C. 1990. Temporal variation in insect life cycles and its evolutionary significance: lessons from periodical cicadas. *BioScience* 40(5):359–367.

Minh BQ, Nguyen MAT, Haeseler AV. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 30(5):1188–1195.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L, et al. 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res.* 20(7):908–916.

Moulds MS. 2018. Cicada fossils (Cicadoidea: Tettigarctidae and Cicadidae) with a review of the named fossilized Cicadidae. *Zootaxa* 4438(3):443–470.

Papadopoulou A, Anastasiou I, Vogler AP. 2010. Revisiting insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Mol Biol Evol.* 27(7):1659–1672.

Peng Y, Leung HCM, Yiu SM, Chin F. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420–1428.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 67(5):901–904.

Remi A, Stefano D, Nicolas G, Benoit N. 2017. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol Biol Evol.* 34(11):2762–2772.

Roff DA. 2002. Life history evolution. Sunderland, MA: Sinauer.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572–1574.

Simon C. 1988. Evolution of 13- and 17-year periodical cicadas (Homoptera: Cicadidae: *Magicicada*). *Bull Entomol Soc Am.* 34:163–176.

Simon C, Lloyd M. 1982. Disjunct populations of periodical cicadas: relicts or evidence of polyphyly. *J N Y Entomol Soc.* 90(4):275–301.

Simon C, Tang J, Dalwadi S, Staley G, Deniega J, Unnasch TR. 2000. Genetic evidence for assortative mating between 13-year cicadas and sympatric "17-year cicadas with 13-year life cycles" provides support for allochronic speciation. *Evolution* 54(4):1326–1336.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS. 2006. Comparative phylogeography of unglaciated eastern North America. *Mol Ecol.* 15(14):4261–4293.

Sota T, Yamamoto S, Cooley JR, Hill KB, Simon C, Yoshimura J. 2013. Independent divergence of 13- and 17-y life cycles among three periodical cicada lineages. *Proc Natl Acad Sci U S A.* 110(17):6919–6924.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585–595.

Tang M, Tan M, Meng G, Yang S, Su X, Liu S, Song W, Li Y, Wu Q, Zhang A, et al. 2014. Multiplex sequencing of pooled mitochondrial genomes: a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Res.* 42(22):e166.

Taylor RS, Friesen VL. 2017. The role of allochrony in speciation. *Mol Ecol.* 26(13):3330–3342.

Taylor JE, Pacheco MA, Bacon DJ, Beg MA, Machado RL, Fairhurst RM, Herrera S, Kim JY, Menard D, Póvoa MM. 2013. The evolutionary history of *Plasmodium vivax* as inferred from mitochondrial genomes: parasite genetic diversity in the Americas. *Mol Biol Evol.* 30(9):2050–2064.

Trifinopoulos J, Nguyen LT, Haeseler AV, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44(W1):W232–W235.

Veller C, Nowak MA, Davis CC. 2015. Extended flowering intervals of bamboos evolved by discrete multiplication. *Ecol Lett.* 18(7):653–659.

West-Eberhard MJ. 2003. Developmental plasticity and evolution. New York: Oxford University Press.

Williams KS, Simon C. 1995. The ecology, behavior, and evolution of periodical cicadas. *Annu Rev Entomol.* 40(1):269–295.

Xu B, Yang Z. 2013. PAMLX: a graphical user interface for PAML. *Mol Biol Evol.* 30(12):2723–2724.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.

Yoshimura J. 1997. The evolutionary origins of periodical cicadas during ice ages. *Am Nat.* 149(1):112–124.