Independent divergence of 13- and 17-y life cycles among three periodical cicada lineages

Teiji Sota,a,b Satoshi Yamamotob, John R. Cooleyc, Kathy B. R. Hilla, Chris Simon,c and Jin Yoshimurad,e,f

aDepartment of Zoology, Graduate School of Science, Kyoto University, Sakyō, Kyoto 606-8502, Japan; bGraduate School of Global Environmental Studies, Kyoto University, Sakyō, Kyoto 606-8501, Japan; cDepartment of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06268-3043; dDepartment of Systems Engineering, Shizuoka University, Hamamatsu 432-8561, Japan; eMarine Biosystems Research Center, Chiba University, Kamogawa 299-5502, Chiba, Japan; and fDepartment of Environmental and Forest Biology, State University of New York College of Environmental Science and Forestry, Syracuse, NY 13210

Edited by Mary Jane West-Eberhard, Smithsonian Tropical Research Institute, Ciudad Universitaria, Costa Rica, and approved February 22, 2013 (received for review November 16, 2012)

The evolution of 13- and 17-y periodical cicadas (Magicicada) is enigmatic because at any given location, up to three distinct species groups (Decim, Cassini, Decula) with synchronized life cycles are involved. Each species group is divided into one 13- and one 17-y species with the exception of the Decim group, which contains two 13-y species—13-y species are Magicicada tredecim, Magicicada neotredecim, Magicicada tredecassini, and Magicicada tredecula; and 17-y species are Magicicada septendecim, Magicicada cassini, and Magicicada septendecula. Here we show that the divergence leading to the present 13- and 17-y populations differs considerably among the species groups despite the fact that each group exhibits strikingly similar phylogeographic patterning. The earliest divergence of extant lineages occurred ~4 Mya with one branch forming the Decim species group and the other subsequently splitting 2.5 Mya to form the Cassini and Decula species groups. The earliest split of extant lineages into 13- and 17-y life cycles occurred in the Decim lineage 0.5 Mya. All three species groups experienced at least one episode of life cycle divergence since the last glacial maximum. We hypothesize that despite independent origins, the three species groups achieved their current overlapping distributions because life-cycle synchronization of invading congeners to a dominant resident population enabled escape from predation and population persistence. The repeated life-cycle divergences supported by our data suggest the presence of a common genetic basis for the two life cycles in the three species groups.

Periodical cicadas (Magicicada) in the eastern United States represent one of the most spectacular life history and population phenomena in nature (1–10). These periodical cicadas spend most of their lives (13 y in the south, 17 y in the north) as underground juveniles except for a brief 2- to 4-wk period when adults emerge simultaneously in massive numbers. With few exceptions, at any given location, all of the periodical cicadas share the same life cycle and emerge on the same schedule, forming a single-year class referred to as a “brood.” Surprisingly, each brood consists of multiple species from three species groups (Decim, Cassini, Decula).

These three groups were considered to have diverged from each other allopatrically and to have later become sympatric and formed 13- and 17-y life cycles (2). The prolonged, prime-numbered life cycles were hypothesized to have evolved in response to Pleistocene climatic cooling (9, 11) to avoid the adverse effect of low population density on mating success (9, 12, 13). Another view hypothesized that the long synchronized life cycles evolved in association with the predator avoidance strategy (2, 4, 8) and that this took place before both the glacial periods and the split of the three species groups (10) based on approximate genetic distances among species groups (8). To test these hypotheses, phylogenetic information about the relationships of species, broods, and populations is essential. However, phylogenetic studies of Magicicada have been largely limited to the Decim group (6, 14–18), and only rough divergence times among species have been inferred (8, 10). Thus, a comprehensive molecular phylogeny covering all of the extant broods, their phylogeography, and divergence time has been lacking until now.

We conducted molecular phylogenetic and population genetic analyses by using nuclear and mitochondrial DNA markers for samples collected over a 30-y period (1978–2008). These samples represent all 15 extant broods and all known species (Table S1).

Results and Discussion

Phylogenetic analyses and divergence time estimation based on four nuclear and three mitochondrial genes (Table S2) clearly reveal the monophyly of each of the three species groups (Decim, Cassini, and Decula) and the sister relationship between Cassini and Decula (Fig. 1). A Bayesian relaxed clock analysis shows that the three species groups diverged 3.9 Mya. Initially the Decim group diverged from the ancestor of Cassini + Decula, then Cassini and Decula separated 2.5 Mya (Fig. 1). The mitochondrial genealogy further shows divergence associated with regions, and partly with life cycles (Fig. 2, Figs. S1 and S2, and Table S3). We distinguished four mitochondrial haplotype groups in Decim and three each in Cassini and Decula (Fig. 2 A–C). The geographic distributions of mitochondrial haplotype groups within each species group show similar divisions among eastern, middle, and western regions (Fig. 2 D–F). In Decim, there is also a major divergence between northern and southern groups corresponding to formally distinguished mitochondrial lineages A and B, respectively (15). Group A is divided into three groups, Ae, Am, and Aw, which occur in the eastern, middle, and western parts of the United States east of the Great Plains, respectively. Populations of Ae and Am exhibit a 17-y cycle (Magicicada septendecim); those of Aw both 17-y (M. septendecim) and 13-y (Magicicada neotredecim) cycles; and those of B show only a 13-y cycle (Magicicada tredecula). At the boundary of Ae and Am, a few populations possess both Ae and Am haplotypes. The Cassini group consists of three haplotype groups, Ce, Cm, and Cw, again occurring in the eastern, middle, and western regions, respectively (Fig. 2). Populations with Ce and Cm show 17-y cycles only (Magicicada cassini), whereas those of Cw show both 17- and 13-y cycles (M. cassini and...
Lastly, the Decula group shows the least divergence of mitochondrial haplotypes, but yet there are differences in haplotype among the eastern (De), middle (Dm), and western (Dw) regions (Fig. 2). Each haplotype group is associated with both 17- and 13-y cycles (Magicicada septendecula and Magicicada tredecima, respectively), suggesting that life cycle divergence occurred independently in the three regions.

The boundaries of haplotype groups roughly coincide among the three species groups (Fig. 2 D–F). In each species group, an isolation-by-distance pattern was detected by a Mantel test (Decim, Mantel’s $r = 0.223$, $P = 0.0007$; Cassini, $r = 0.349$, $P = 0.0001$; Decula, $r = 0.377$, $P = 0.0001$), suggesting that contiguous range expansion resulted in the present distribution. Parts of broods of each species group were associated with two or three haplotype groups and occurred in different geographic regions, suggesting multiple origins of the same broods even within a single species (Fig. 2 and Fig. S2). In Decim, broods VI, IX, X, and XIV are found in Ae and Am, and broods XIX and XXIII are found in Aw and B. In Cassini, broods V and IX are found in Ce and Cm, and broods X and XIV are in Cc, Cm, and Cw. In Decula, broods V and XIV are found in De and Dm, broods VI and XXIII in Dc and Dw, and broods XIX in De, Dm, and Dw. Thus, the broods of multiple lineages are shared among the three species groups and show similar geographic patterns. The divergence pattern of Magicicada tredecassini, respectively).
the two life cycles differs among the three species groups in relation to haplotype groups (Fig. 2). Although genetic differentiation between 13-y *M. tredecim* and 17-y *M. septendecim* in Decim are evident, naturally, because these belong to different haplotype groups [Table S4; analyses of molecular variance (AMOVA)], significant differentiation between 13-y *M. neotredecim* and 17-y *M. septendecim* within the haplotype group Aw was also detected by an AMOVA (Table S4; $F_{CT} = 0.234$; $df = 1$, $P < 0.001$). However, there was not significant differentiation between 13- and 17-y cicadas in Cassini haplotype group Cw or in the whole Decula group (Table S4; Cassini Cw: $F_{CT} = 0.068$, $df = 1$, $P > 0.05$; Decula: $F_{CT} = -0.039$, $df = 1$, $P > 0.05$).

The analysis of amplified fragment length polymorphism (AFLP) markers show results generally consistent with the mitochondrial data results (Fig. 3). Each species group showed a significant differentiation in terms of the fixation index among groups defined by life cycle and regions associated with mitochondrial haplotype groups (Decim, $F_{ST} = 0.0241$, $P = 0.0000$; Cassini, $F_{ST} = 0.0099$, $P = 0.0000$; Decula, $F_{ST} = 0.0109$, $P = 0.0040$). In Decim, the 13-y *M. tredecim* (haplotype group B) was significantly differentiated from other groups as indicated by pairwise $F_{ST}$ values, whereas another 13-y group *M. neotredecim* (haplotype group Aw) and 17-y *M. septendecim* groups were not differentiated from one another except for one case (Fig. 3). In Cassini, the closest relationship between 13-y *M. tredecassini* and 17-y *M. cassini* within the haplotype group Cw was revealed (Fig. 3). In Decula, the six groups divided by life cycle and haplotype group showed relationships different from those inferred by using mitochondrial data; 13-y *M. tredecim* in haplotype groups De and Dm were closely related to each other, as were 17-y *M. septendecim* in these haplotype groups (Fig. 3). Note, however, that haplotype groups De and Dm show only 1-bp difference in the mitochondrial sequence (Fig. 2C).

The estimated age using a Bayesian relaxed clock analysis of the most recent common ancestor (MRCA) of all Decim mitochondrial haplotypes is 0.53 Mya, after which time it separated into clades A (Ac+Am+Aw) and B (Fig. 1). The MRCA of each haplotype group is more recent (0.16–0.08 Mya). The MRCA of the Decim Aw group containing 13-y (*M. neotredecim*) and 17-y (*M. septendecim*) cicadas is 0.12 Mya (Fig. 1). In the Cassini group, the MRCA of all mitochondrial haplotype groups is traced back to 0.32 Mya, and the MRCA of Cw back to 0.16 Mya. Lastly, mitochondrial haplotypes of the Decula group had their MRCA 0.23 Mya. The divergence time of 13- and 17-y cicadas within the same mitochondrial lineage was estimated together with the gene flow between them using the isolation-with-migration approach (Fig. 4 and Fig. S3). In Cassini Cw, divergence time was ~23 kya, whereas in Decim Aw, divergence time was ~10 kya. In Decula, the divergence time was not clearly estimated (close to zero) due to the lack of haplotype differences between *M. septendecim* and *M. tredecim*, so their life cycle divergence must have occurred quite recently.

The Bayesian skyline plot (BSP) based on mitochondrial gene sequence data (Fig. 5) as well as that based on both mitochondrial and nuclear gene sequence data (Fig. S4) revealed that the population sizes of both the Decim and the Cassini groups were relatively small during the last glacial period, probably due to a population bottleneck, but increased markedly after 10,000 y ago following the abrupt rise of temperature in the Holocene (19). By contrast, there was no increase in population size for the Decula group. This difference in the demographic history corresponds with the fact that except in some southern populations, today Decula is generally rare compared with Decim and Cassini (8).

Our results are broadly consistent with the previous idea that an ancestor of all *Magicicada* diverged into three species allopatrically, and later, the three became sympatric and each species
The divergence of 13- and 17-y cicadas was asynchronous among species groups, reflecting the genetic basis of these life cycles and shifts in the origin of the species groups, not a recent life cycle shift of M. neotredecim. Our results corroborate the idea that life cycle plasticity has been a creative force in the evolution of Magicicada species, suggesting that the genetic basis of life cycles and shifts is the same across species groups.

**Materials and Methods**

**Sampling.** Adults of Magicicada were collected from 27 states in the United States from 1978 to 2008 (Table S1) and stored in freezers. Total genomic DNA was extracted from leg muscles of each ethanol-fixed specimen by using the Wizard Genomic DNA Purification Kit (Promega). For Decim, Cassini, and Decula, 332, 238, and 165 individuals, respectively, were used. Specimens are kept in the Department of Zoology, Kyoto University, as vouchers and for future analyses of DNA. We used the sister genera of Magicicada (27) as outgroup taxa, specifically, Tryella crassa, Tryella graminea, Tryella burrilli, Aleeta curvicosta, and Aleeta curvicosta.

**Fig. 3.** Differentiation in AFLP loci among groups defined by life cycle and mitochondrial haplotype group (geographic region). The relationships among groups are depicted by the unrooted neighbor-joining tree based on pairwise Nei’s distances. Node supports are bootstrap percentages (shown when >50%). Pairwise Fst, and P values among groups in each species group are given in inset tables. Fst values in bold letters are significant at α = 0.05 after controlling for the false-positive rate.

**Fig. 4.** Divergence times between 17- and 13-y cicadas in three groups inferred by the isolation-with-migration approach. (A) Between M. septendecim and M. neotredecim in Decim Aw. (B) Between M. cassini and M. tredecassini in Cassini Cw. (C) Between M. tredecula and M. septendecula in Decula.
OTUs (114 unique sequences of Magicicada and 5 outgroup sequences) was analyzed. The optimal partitioning scheme consisted of three partitions: (i) COI first + tRNA(Leu) + COI first + TrN (Tamura–Nei) + G; (ii) COI second + COII second (TrN + I); and (iii) COII third + COII third (TrN + I). For each sequence dataset, a partitioned maximum-likelihood analysis was conducted by using Treefinder version October 2008 (29). Bootstrapping analysis with 1,000 replications was performed also by using Treefinder using the ML tree as the starting tree.

Haplotype Network and Population Genetic Analyses of Mitochondrial Gene Sequences. The relationships among mitochondrial haplotypes were assessed by constructing statistical parsimony networks with a 95% connection limit using TCS version 1.21 (30). Haplotype groups within each species group were assigned based on monophyly or the segregation from other groups with two or more missing haplotypes. Mantel tests for the isolation-by-distance trend in pairwise genetic distance between populations were conducted by using the R package (31). AMOVA for genetic differentiation by haplotype groups, broods, and life cycles were performed by using Arlequin version 3.11 (32).

Divergence Time Estimation. Estimation of divergence times was conducted with all nuclear and mitochondrial DNA sequence data by using BEAST version 1.7.2 (33). No fossil or geographic evidence for calibrating nodes is available for the Magicicada phylogeny. Therefore, we attempt to determine a plausible range of age for the node of the MRCA of Magicicada by using a recently proposed evolutionary rate for insect mitochondrial genes (ref. 34; Fig. 3), which was estimated by compiling sequence divergence rates at various time spans based on sequence divergence corrected for rate heterogeneity among sites using a gamma distribution and fitting them to a time-dependent rate equation (35): rate of sequence divergence (%) per million year = 17.256 exp[–1.1571] + 2.0968. We can estimate divergence times for a given sequence divergence with this rate by using equation 7 in ref. 35, which relates sequence divergence to divergence time. For the mitochondrial sequences of Magicicada, the optimal substitution model was GTR + G based on the Akaike Information Criterion according to MrModelltest version 2.3 (36). To determine the node age for the MRCA of Magicicada, the GTR + G corrected sequence divergence between Decim and Cassini + Decula, 0.2531 (SD = 0.0223), was converted to 4.16 My by using the above time-dependent sequence divergence rate. In addition to this estimation with BEAST, the node age prior was set as a normal distribution function with mean = 4.16 and SD = 0.3943 (proportional to original SD). Ref. 34 also provides another time-dependent clock based on uncorrected P sequence divergence from different datasets. Using this clock equation and an uncorrected sequence divergence of 0.0844 ± 0.0034 (SD) for Magicicada MRCA, we obtain 3.57 ± 0.1442 Mya. Because this alternative estimate is included well in the range (3.4–4.9 Mya for 95% HPD interval) of the previous estimate (assuming a normal distribution with mean = 4.16 and SD = 0.3943), our age prior included the age estimation based on a conservative substitution rate without accounting for rate heterogeneity among sites. In the BEAST analysis, data partitioning and substitution models followed the previous phylogenetic analysis. Both substitution model and clock parameter were unlinked among partitions, but the tree was linked. We tested runs with the strict clock model and the uncorrelated lognormal relaxed-clock model, and compared these by Bayes factor for marginal likelihood using TRACER version 1.5 after the runs. The BEAST Markov chain Monte Carlo run was conducted for 100 million generations sampling every 10,000th generation. The initial 1 million generations were discarded as burn-in when we obtained the maximum clade credibility tree. Based on Bayes factor, we selected uncorrelated lognormal relaxed-clock model.

Phylogenetic Analysis. We conducted a simultaneous analysis of five gene sequences with 64 operational taxonomic units (OTUs) using an optimal data-partitioning scheme (28). The nuclear gene sequence data were divided into nine subsets: 185 and Cal (intron); four of EFl-a (first, second, and third position, and intron); and three codon positions of Wg. Mitochondrial gene sequence data were divided into seven subsets: three codon positions of COI; tRNA; and three codon positions of COII. These 16 subsets were used in a heuristic search to identify the best partitioning scheme and evolutionary model for each partition according to the Bayesian information criterion by using the program PartitionFinder version 1.0.1 (28). The best scheme consisted of six partitions: (i) 185 + EFl-a second + Wg second [Substitution model: JC (Jukes–Cantor)]; (ii) COI first + tRNA(Leu) + COI first + Cal + EFl-a intron [HKY (Hasegawa–Kishino–Yano) + G]; (iii) Wg third (TVM (transversion model) + G); (iv) EFl-a first + Wg first [F81 (Felsenstein 1981)]; (v) COI second + COII second + EFl-a third (HKY + I); and (vi) COII third + COII third (HKY + I). For mitochondrial gene sequences, a larger data set with 119

Fig. 5. BSPs showing demographic histories of the three Magicicada species groups, Decim (A), Cassini (B), and Decula (C), based on mitochondrial gene sequence data. Thick solid curves indicate mean effective population sizes, and gray areas the 95% HPD limits. Dotted lines indicate the beginning of the Holocene (elevated air temperatures).

Gene Sequence Analysis. We sequenced partial gene regions of nuclear 18S rRNA (18S; 808-bp sequence), wingless (Wg; 376-bp exon sequence), and Elongation Factor 1-alpha (EFl-a exon and intron sequence; 1,302–1,303-bp for Magicicada; 1,270–1,616-bp for the outgroup) and Calmodulin (Cal; 385–394-bp intron sequence) gene and a 1,832-bp mitochondrial DNA region encompassing partial sequences of cytochrome oxidase subunit I and II genes (COI, COII) and a sequence of tRNA(Leu) in between. Primers used for PCR amplification and direct sequencing are given in Table S2. Direct sequencing of PCR products used an ABI3130xl sequencer (Applied Biosystems). When sequence data for COII sequence were affected by putative nuclear mitochondrial DNA (numt), we used specific internal primers (Table S2). Samples of each species group representing different broods and regions were sequenced for all gene regions. The samples consisted of 25 from the Decim group, 18 from the Cassini group, and 19 from the Decula group. Two outgroup specimens (T. crassa and A. curvicosta) were also sequenced. For the mitochondrial DNA sequences, were obtained for 329, 238, and 165 specimens of Decim, Cassini, and Decula, respectively (Table S1). All nuclear gene sequences and all mitochondrial haplotype sequences detected have been deposited in DNA Data Bank of Japan database (accession nos.: 185, AB740543–AB740606, Cal, AB740607–AB740670; EFl-a, AB740671–AB740734, Wg, AB740735–AB740798; COII, AB740799–AB740917; see also Table S3 for COII haplotypes).

AFLP Analysis. AFLP analysis (37) was performed for 237, 95, and 96 specimens of Decim, Cassini, and Decula, respectively, encompassing all broods from each species group (Table S1). We used a plant mapping kit (Applied Biosystems) and six primer combinations for selective amplification (EcoRI/MseI: ACT/CGT, AGG/ACG, AGG/ACG, ACA/CAC, AGG/CTA, and ACC/CGT). The amplified fragments were electrophoresed on an ABI 3130xl sequencer and binary-coded by using GeneMapper version 4.0 (Applied Biosystems). To ensure high reliability of analyzed AFLP loci, every specimen was genotyped twice, and a total of 440 loci with >90% repeatability was used in the analysis. The differentiation in AFLP loci among groups defined by life cycle and regions associated with mitochondrial haplotype groups within each species group was assessed by using pairwise Fst using AFLP-surf version 1.0 (38). The significance of Fst was tested by 1,000 bootstrap replications. We controlled the false-positive rate by using the Benjamini and Hochberg method (39) to determine statistical significance of multiple pairwise Fst values between groups. We also obtained pairwise Nei’s distances between groups

Sota et al.