Triplicate parallel life cycle divergence despite gene flow in periodical cicadas

Tomochika Fujisawa 1, Takuya Koyama 1, Satoshi Kakishima 2,3, John R. Cooley 4,5, Chris Simon 5, Jin Yoshimura 2,6,7 & Teiji Sota 1

Periodical cicadas comprise three species groups containing three pairs of 13- and 17-year life cycle species showing parallel divergence, along with a more anciently diverged 13-year species (*Magicicda tredecim*). The mechanism and genetic basis of this parallel divergence is unknown. Here we use orthologous transcriptome sequences to explore the demographic processes and genomic evolution associated with parallel life cycle divergence. The three 13- and 17-year species pairs have similar demographic histories, and the two life cycles diverged 200,000–100,000 years ago. Interestingly, these life cycle differences have been maintained despite substantial gene flow between 13- and 17-year species within species groups, which is possible during co-emergences. Sequence divergence between 13- and 17-year species in each species group (excluding *M. tredecim*) is minimal, and we find no shared divergent single-nucleotide polymorphisms (SNPs) or loci associated with all instances of life cycle divergence. The two life cycles may be controlled by highly limited genomic differences.
Life history diversity is a remarkable feature of living organisms and underlies fundamental evolutionary questions. Periodical cicadas of the genus *Magicicada* are found only in the eastern United States and are well known for their unusual life history patterns, characterised by prolonged juvenile periods of 13 or 17 years, followed by synchronised mass emergence of adults within local populations. Only one cohort, or ‘brood’, of periodical cicadas emerges every 13 or 17 years in any given location. There are three co-occurring species groups of periodical cicadas, Decim, Decula and Cassini. Each has one species with a 17-year life cycle and one or two species with a 13-year cycle, and there are seven described species (four 13-year and three 17-year) in total. Among these species groups are extremely similar or indistinguishable in these characters; thus, the difference in life cycle length is one of the only diagnostic characters for their identification.

The three species groups are estimated to have diverged 3.9–2.5 million years ago (mya), and subsequent divergence of the present 13-year (mostly southern) and 17-year (mostly northern) life cycles has occurred in parallel in the three species groups during the Quaternary, except for the cycles has occurred in parallel in the three species groups during the 13-year (mostly southern) and 17-year (mostly northern) life cycle and there are seven described species (four 13-year and three 17-year) in total. Although the species groups clearly differ in morphology, male songs and female song preferences, the 13-year and 17-year species within each species group are extremely similar or indistinguishable in these characters; thus, the difference in life cycle length is one of the only diagnostic characters for their identification.

The genetic basis of life cycle length has not been studied for their identification. The hypothesis that all periodical cicadas possess monomorphic developmental plasticity and that this common plasticity underlies the switching of life cycle lengths triggered by environmental cues (e.g., a drastic change in temperature during juvenile development), followed by a genetic change in a life-cycle control locus (genetic accommodation), which enables a permanent life cycle shift.

In general, parallelism in adaptive character divergence among closely related species results from parallel mutation or selection, ancestral polymorphism with balancing selection, or adaptive introgression. In periodical cicadas, an ancestral polymorphism in life cycle length followed by collateral genetic evolution is considered the most parsimonious explanation for the parallel divergence and the formation of synchronous broods among three species groups, because multiple independent acquisitions of identical life cycles are unlikely. In addition, a hypothesis of life cycle switching via introgressive hybridisation of the putative 13-year allele from 13- to 17-year cicadas has been proposed. This hypothesis was used to explain the existence of two 13-year species in the Decim group, proposing that introgressive hybridisation from the preexisting 13-year species *M. neotredecim* to the 17-year *M. septendecim* produced the new 13-year species *M. neotredecim*. However, the hybrid origin hypothesis of *M. neotredecim* was rejected based on population genetic studies. The hybrid origin hypothesis of 13-year species is unlikely to be applicable to the Cassini and Decula groups, which have no early diverged 13-year species (unless hybridisation between species groups drove life-cycle switching from 17- to 13-year cycles).

**Fig. 1** Species and species groups of *Magicicada* periodical cicadas. **a** Seven *Magicicada* species in three species groups, with relationships and divergence times for four lineages estimated in our previous study. Individuals in photographs are all males. Photographs were taken by T. Sota. **b** Distribution ranges of 13- and 17-year cicadas. The dark pink area in the range of 13-year cicadas indicates the contact zone with *M. neotredecim* (northern species) and *M. tredecim* (southern species) in the Decim group.
To understand the process and genetic basis of the parallel life cycle divergence observed in periodical cicadas, we inferred the demographic histories of broods of three *Magicicada* species groups. We used reduced representation sequences from transcriptomes (mRNA sequences) because *Magicicada* genomes are likely large in size (>6 Gbp) as in other cicadas and the whole genomes have not yet been sequenced. We focused on three pairs of 13- and 17-year species (excluding *M. tredecim*, which diverged earlier). In addition, we surveyed the genes responsible for life cycle control by comparing divergent loci of 13- and 17-year species (excluding *M. tredecim*, which diverged earlier). In general, comparisons of populations with parallel character divergence can be an effective means for discovering diverged portions of the genome and genes responsible for the character divergence.

Our study reveals the historical process of the parallel life-cycle divergence in the three species groups. First we confirm the relationships of four major lineages (Cassini and Decula groups, and two lineages within Decim) and the absence of introgressive hybridisation among these four lineages. Then we estimate demographic histories in the three species groups and find that, in each group, 13-year broods are monophyletic, sister to or derived from 17-year broods. Interestingly, we find evidence of gene flow between the 13- and 17-year species in each species group. Finally we search for single-nucleotide polymorphisms (SNPs) or loci showing elevated divergence between life cycles, but do not find any divergent SNPs or loci shared among all 13- and 17-year species pairs, nor any evidence for parallel genomic divergence across all pairs. Thus, the genetic background of the life cycle divergence in periodical cicadas remains unclear.

**Results**

**Assembly of transcriptome sequences and orthologous loci.** We sequenced mRNA from head tissues of 28 individuals (Supplementary Data 1) from two representative 17-year broods (eastern and western broods, II and III, respectively) and the two major 13-year broods, XIX and XXIII (Fig. 2a). (Note that 12 broods...
with 17-year cycles and three broods with 13-year cycles currently exist). De novo assemblies of the RNAseq reads were generated separately for four distinct groups: Decim (M. septendecim, M. neotredecim); M. tredecim; Cassini (M. cassini, M. tredecaassini); and Decula (M. septendecula, M. tredecula). For each group, we obtained 76,519–90,287 contigs (length: 300–26,405 bp) with an average N50 length of 1476 bp (Table 1). Using these contigs, which 5270 were shared by all four groups (lineages) of Magicicada (Supplementary Fig. 1). Among the 5270 clusters, we identified 2636 loci, 99% had BLAST hits with e-values <1×10^{-5} in the RefSeq protein database (Supplementary Data 2). The average alignment length of the loci was 1627 bp, and the average nucleotide diversity (\( \pi \)) of the loci for all Magicicada sequences (\( n = 28 \)) was 0.0019 (range: 0–0.0238). The nucleotide diversity of the loci within the seven species (\( n = 4 \) for each species) was generally low, with a mean of 0.00045–0.00071 and a median of 0–0.00018 (Fig. 2c).

### Molecular phylogeny of periodical cicadas

To characterise the historical relationship of species groups and broods, we first reconstructed phylogenetic trees using concatenated sequence data from the orthologous loci. The concatenated alignment was ca. 4.3 Mb in length, with 18% missing sites, and it contained 18,243 informative sites. The maximum-likelihood tree reliably recovered the monophyly of three species groups and the two lineages within the Decim group (M. tredecim and the lineage containing M. neotredecim and M. septendecim), but the relationships among broods within the Decim group (excluding M. tredecim), the Cassini group, and the Decula group were unresolved (Fig. 3). We also applied a species-tree method (SVQuartets\(^a\)) to resolve the relationships among allochronically-separated broods, but it again poorly resolved the relationships among 13- and 17-year broods within each species group (Fig. 4a). In this tree, monophyly of the two 13-year broods was weakly supported in the Decim and Decula groups, whereas they were not monophyletic in the Cassini group.

### Lack of hybridisation between four major lineages

To reconstruct the process of life cycle divergence, we first tested whether introgressive hybridisation among the four lineages (i.e. M. tredecim and three paired 13-year and 17-year species) was involved in life-cycle divergence events using the ABBA-BABA test with the D-statistic\(^b\) for SNPs. In particular, we tested the possibility that the earliest-diverged 13-year species, M. tredecim, introduced the 13-year life cycle into another lineage of Decim, or the Cassini and Decular groups through hybridisation, but we found no evidence for introgressive hybridisation (Table 2). We also tested for hybridisation between M. neotredecim and the Cassini or Decula group, and between the Cassini and Decula groups, but found no positive evidence (Table 2). Thus, we excluded the possibility that introgressive hybridisation between species groups or between the distinct Decim lineages was involved in the life-cycle divergence process.

### Demographic histories within species groups

To further investigate the historical process of life-cycle divergence, we inferred the demographic histories of broods within species groups using the program fastsimcoal2\(^d\), which analyses the joint site frequency spectra of synonymous SNPs. We used only high-quality SNPs from loci for which we could reliably infer reading frames. We considered three alternative scenarios of the relationships among broods (scenarios S1–S3), which reflected the possible diversification of the broods (Supplementary Fig. 1). In addition, we included three alternative models with gene flow between broods under each scenario, because recent divergence alone may not explain the low nodal support on the brood phylogenies. The three models were no gene flow, all possible recent and past gene flow (between ancestral populations and between current populations), and possible recent gene flow (between current populations). Thus, a total of nine models were compared in each of three species groups (Supplementary Fig. 1). For the Decim group, we included only samples of M. septendecim and M. neotredecim because M. tredecim had clearly diverged from the two species and gene flow between M. tredecim and parapatric M. neotredecim is virtually absent as was shown in our previous study\(^e\) and the ABBA-BABA test in the previous section.

We selected the best models of brood diversification based on model comparison using Akaike information criterion (AIC) weights and bootstrap proportions (Table 3). In all species groups, models with recent gene flow exhibited better fit than did models with no gene flow and those with both past and recent gene flow (Table 3). The best-fit scenarios were monophyly of both life cycles in the Decim and Cassini groups and monophyly of the 13-year species in the Decula group (Table 3, Fig. 4b). Note that the likelihood difference between recent gene flow models and past/recent gene flow models was marginal; the former models were favoured in AIC-based model comparisons because they had fewer parameters.

### Table 1 Assembly of sequence reads for four groups (lineages) of Magicicada

| Species group | Species/broods\(^a\) | N\(^b\) | No. contigs | Total contig length, bp | Max contig length, bp | Contig N50, bp | GC (%) | gap site (%) |
|---------------|---------------------|-------|-------------|------------------------|----------------------|----------------|--------|-------------|
| Decim         | M. tredecim/ XIX, XXIII | 4     | 76,519      | 75,508,799             | 23,352               | 1511            | 34.5   | 0.0         |
| Decim         | M. septendecim II, III; M. neotredecim/ XIX, XXIII | 8     | 90,287      | 87,384,502             | 25,741               | 1474            | 34.4   | 0.0         |
| Cassini       | M. cassini/ II, III; M. tredecaassini/ XIX, XXIII | 8     | 86,587      | 83,074,541             | 26,405               | 1441            | 34.4   | 0.0         |
| Decula        | M. septendecula/ II, III; M. tredecula/ XIX, XXIII | 8     | 85,135      | 82,756,110             | 26,269               | 1478            | 34.5   | 0.0         |
| Total         |                     | 28    | 338,528     | 328,723,952            |                      |                |        |             |

\(^a\) Species/broods: M. tredecim/ XIX, XXIII; M. septendecim II, III; M. neotredecim/ XIX, XXIII; M. cassini/ II, III; M. tredecaassini/ XIX, XXIII; M. septendecula/ II, III; M. tredecula/ XIX, XXIII.

\(^b\) Number of individuals used in the assembly. Two individuals per brood per species.
The estimated divergence times of 13- and 17-year life cycles in the three groups were 197, 121 and 95 ky ago (ky) in the Decim, Cassini and Decula groups, respectively (Fig. 4b, Supplementary Table 1). These three divergences occurred between the Illinoian glacial period and the last glacial period. These divergence times are comparable to the times of the most recent common ancestor (tMRCA) for 13- and 17-year species pairs estimated in the maximum-likelihood tree, 213, 131 and 111 ky for the Decim, Cassini and Decula groups, respectively (Supplementary Table 1). Thus, the split of the two major 13-year broods likely occurred during the last glacial period (Fig. 4b).

The estimated effective population size ($N_e$) was consistent with the known biology of Magicicada (Fig. 4b, Supplementary Table 1). In the Cassini and Decula groups, $N_e$ was larger in 13-year broods than in 17-year broods, which generally reflects the widespread range of 13-year species in these groups. By contrast, in the Decim group containing $M.\ septendecim$ and $M.\ neotredecim$, $N_e$ of the 17-year broods ($M.\ septendecim$) was larger than that of the 13-year broods ($M.\ neotredecim$), which suggests a recent origin for the narrowly distributed 13-year species $M.\ neotredecim$. The current population sizes of 13-year and 17-year cicadas in each species group were larger than ancestral population sizes (except 17-year broods in the Decula group), which suggests recent population expansion associated with divergence of broods.

Estimated gene flow ($N_M$) between broods with the best models ranged from 0.01 to 23.8 migrants (individuals) per generation (Fig. 4c, Supplementary Table 1). For brood pairs in 13- or 17-year cicadas, gene flow was small between 17-year broods II and III, which are geographically separated, but 13-year brood pairs XIX and XXIII, which share lengthy boundaries, showed substantial gene flow (>1.0) in all species groups. For brood pairs between the 13- and 17-year cicadas, a substantial amount of gene flow (>1.0) was estimated to have occurred between all pairs in the Decim and Decula groups and between two of the four pairs in the Cassini group. Although the $N_M$ confidence intervals were wide in each instance, the lower confidence limits of the gene flow between adjacent broods III and XIX were higher than 1.0 in the Decim and Decula groups, as well as between broods III and XXIII in the Decim group. In the Decula group, the $N_M$ between broods II and XIX was greater than 1.0 despite the geographic separation of the samples, which indicates gene flow between eastern and western populations of brood XIX. In the Cassini group, the lower confidence limits of the gene flow between 13- and 17-year broods were low but non-zero (>0.003; Supplementary Table 1).

Genomic divergence between 13-year and 17-year cicadas. We measured genomic divergence between four 13- and 17-year species pairs of three groups (the three pairs detailed above and the $M.\ tredecim$/M. septendecim pair) using the fixation index $F_{st}$ for individual SNPs and loci (Fig. 5). In general, $F_{st}$ did not indicate divergence between 17- and 13-year species except in the anciently diverged pair, $M.\ tredecim$ and $M.\ septendecim$ (Fig. 5). At the locus level, Tajima’s $D$ values were generally negative, and only 16–22% of loci exhibited positive values (Fig. 5), which indicates that the loci were mainly under purifying selection, although population size expansion (Fig. 4b) may have also affected Tajima’s $D$. We also calculated $d_{xy}$ between 13- and 17-year species as an absolute measure of nucleotide divergence, but the values of $d_{xy}$ were strongly correlated with $\pi$ (Supplementary Fig. 2) and did not capture the divergence between species. Therefore, we used only $F_{st}$ for the analyses of shared outliers.

![Maximum-likelihood (ML) tree resulting from the concatenated sequences for individual cicadas. The tree was made ultrametric with the time of most recent common ancestor (tMRCA) of all Magicicada set to 3.89 million years ago (the 95% confidence interval of this node is shown by the broken line). Numerals above the branches are the bootstrap percentages resulting from the ML analysis. Estimated ages and 95% confidence intervals based on two calibrations using the upper and lower confidence limits of Magicicada tMRCA are shown for five major nodes (grey bars)](image-url)
One possible mechanism of parallel life-cycle divergence is parallel divergence at the same nucleotide sites or loci, which may be accompanied by divergence of linked genomic portions. Therefore, we searched for divergent SNPs or loci shared between the four pairs of 13- and 17-year species (six comparisons) using \( F_{st} \) values for each 13- and 17-year species pair (Fig. 6). We defined elevated \( F_{st} \) by simulating SNPs under the best demographic models inferred in the previous section and taking the 95% quantiles of simulated \( F_{st} \). Among 23,524 SNPs examined, 30 SNPs (0.1%) with elevated \( F_{st} \) (hereafter divergent SNPs) were shared by two species pairs (Fig. 6a, Supplementary Data 3), of which 27 SNPs were found in the within-Decim group comparison (i.e., \( M. \text{tredecim} \)/\( M. \text{septendecim} \) vs. \( M. \text{neotredecim} \)/\( M. \text{septendecim} \)). The other five comparisons yielded few or no shared divergent SNPs (Fig. 6a). The proportion of non-synonymous changes in the shared divergent SNPs was 0.47 (14/30) and did not significantly differ from genome wide proportion, 0.32 (\( P = 0.11 \), binomial test). At the locus level, we found 21 ‘divergent loci’ (0.7% of 2636 loci) with elevated \( F_{st} \) (Weir–Cockerham weighted \( F_{st} \))29, which were shared by two or
more species pairs (Fig. 6b, Supplementary Data 3). Further, we selected the maximum SNP $F_{st}$ for each locus as an alternative measure of locus-level divergence. We discovered 15 divergent loci (0.6%) with elevated maximum $F_{st}$ shared by two or three pairs (Fig. 6c, Supplementary Data 3). In the only divergent locus shared by three pairs (the exception being the Decula pair), the SNPs with maximum $F_{st}$ were located in different positions among the three pairs.

The above three analyses suggest that parallel genomic divergence associated with life-cycle divergence is uncommon. To clarify this, we conducted a permutation test to estimate the probability that the number of shared divergent SNPs or loci observed in each comparison were obtained by chance alone. We found non-random occurrence of outliers of that number only for the within-Decim comparison in the SNP $F_{st}$, none in the locus $F_{st}$, and only two comparisons involving the Decula or Cassini groups and $M. neotredecim$/M. septendecim in the locus maximum $F_{st}$ (the number of shared divergent SNPs or loci, $N_{shared}$, and the permutational $P$ values are given in the legend of Fig. 6a–c).

In the above outlier analyses, we obtained a total of 45 loci that exhibited elevated $F_{st}$ at SNP or locus level between pairs of 13- and 17-year species (Supplementary Data 4). The functional annotation of these genes did not indicate enrichment of any kind of gene function. The 2636 loci studied included 21 genes involved in pathways potentially related to life cycle control (circadian clock, insulin signalling, insect hormone biosynthesis, MAPK signalling, and phototransduction30–33, Supplementary Data 5). However, the genes involved in these pathways were not found in the shared divergent loci.

### Discussion

Our phylogenetic analysis using mRNA sequences clearly resolved the branching pattern of the four major periodical cicada clades, consistent with those of previous studies that used mitochondrial and genome-wide (restriction-site-associated DNA; RAD) markers6,20. However, neither mRNA and RAD sequence data resolved the relationships among broods within species groups despite the vast amounts of data, whereas the mitochondrial gene tree partly resolved phylogeographic (eastern, middle and western) patterns within species groups6. However, phylogenetic and population genetic inferences with mitochondrial gene sequences can be distorted by introgressive hybridisation.

---

**Table 2** Results of ABBA-BABA test with $D$-statistic for testing introgression hypotheses between 13-year species of different species groups or between 13-year species in the Decim group

| Taxon assigned (species/brood) | P1 | P2 | P3 | O | No. of sites | $D$ (STD) | $Z^b$ |
|------------------------------|----|----|----|---|-------------|---------|------|
| (a) $M. tredecim$ to $M. neotredecim$ | sdecim III | ndecim XIX | tdecim XIX | cassini II,III | 9225 | -0.0048 (0.0701) | 0.0680 |
| (b) $M. tredecim$ to $M. tredecassini$ | cassini III | tcassini XIX | tdecim XIX | O. villosa | 336928 | -0.0424 (0.0724) | 0.5860 |
| (c) $M. tredecim$ to $M. tredecula$ | sdecula III | tdecula XIX | tdecim XIX | O. villosa | 346099 | -0.027 (0.085) | 0.3180 |
| (d) $M. neotredecim$ to $M. tredecassini$ | cassini III | tcassini XIX | ndecim XIX | O. villosa | 336989 | -0.0376 (0.0733) | 0.5124 |
| (e) $M. neotredecim$ to $M. tredecula$ | sdecula III | tdecula XIX | tdecim XIX | O. villosa | 331558 | -0.0291 (0.0868) | 0.3354 |
| (f) $M. tredecassini$ to $M. tredecula$ | sdecula III | tdecula XIX | tdecim XIX | O. villosa | 336928 | -0.0097 (0.0849) | 0.1137 |
| (g) $M. tredecula$ to $M. tredecassini$ | sdecula III | tdecula XIX | tdecim XIX | O. villosa | 331759 | -0.038 (0.0899) | 0.4224 |

---

### Table 1:

| Taxon assigned (species/brood) | P1 | P2 | P3 | O | No. of sites | $D$ (STD) | $Z^b$ |
|------------------------------|----|----|----|---|-------------|---------|------|
| $M. tredecim$ | sdecim III | ndecim XIX | tdecim XIX | cassini II,III | 9225 | -0.0048 (0.0701) | 0.0680 |
| $M. neotredecim$ | cassini III | tcassini XIX | tdecim XIX | O. villosa | 336928 | -0.0424 (0.0724) | 0.5860 |
| $M. tredecassini$ | cassini III | tcassini XIX | tdecim XIX | O. villosa | 339643 | -0.0616 (0.0752) | 0.8193 |

---

1. P1, P2, P3 and O are operational taxonomic units (OTUs), in which P3 is the donor and P2 the recipient of introgression. The relationships among the OTUs are (((P1,P2),P3),O).
2. The maximum $Z$ score is 0.8193, which corresponds to $P=0.3579$ (two-sided probability). Thus, no $D$ values are significantly different from zero.
3. $D$ is the standard deviation of $D$ obtained by the bootstrap procedure.
4. The maximum $Z$ score is 0.8193, which corresponds to $P=0.3579$ (two-sided probability). Thus, no $D$ values are significantly different from zero.
5. $O$. villosa = Okanogano villosa
and incomplete lineage sorting of ancestral polymorphism\(^{34}\). Therefore, it was necessary to revisit the results of our previous study based mainly on mitochondrial data, especially to confirm the divergence process of broods with different life cycles.

Our demographic inference using a site frequency-based method provides new insights into the parallel divergence process of 13- and 17-year life cycles, revealing that the three species groups have nearly parallel demographic histories, with 13-year broods monophyletically diverged from 17-year broods in each species group. The Decim and Cassini groups share a common diversification pattern in which 13- and 17-year groups diversified first, whereas the Decula group had a slightly different history, as the 13-year group was derived from brood III (representing the western 17-year brood). However, the divergence time of the 13-year group from brood III is close to that of broods II and III in the Decula group; thus the differences in divergence patterns among the species groups may not be substantial. The present results differ from those found through mitochondrial phylogenetics\(^6\), in which 13-year broods in both the Decim and Cassini groups were found to have been derived from the western haplotype group, including brood III, whereas the origin of 13-year broods appears to have been polyphyletic in the Decula group. Our demographic inference also suggests population expansions following brood splits in each species group. This finding is consistent with the previous results using mitochondrial data that population expansions occurred after the last glacial period in the Decim and Cassini groups\(^6\) although the present study does not restrict the timing of population expansion to the post LGM except for Decula 13-year broods. We estimated that life cycle divergence (i.e., the split between 13- and 17-year species) in the Decim group (excluding M. neotredecim and M. septendecim) occurred at the beginning of the Illinoian glacial period (130–115 kya), and those of the Cassini and Decula groups during the Sangamon interglacial period (130–115 kya) or early in the last (Wisconsin) glacial period (115–12 kya). Although the confidence intervals for these estimated times are wide (between 247 and 19 kya overall), the present estimates are much older than the divergence times estimated using mitochondrial gene sequences.

### Table 3 Comparison of demographic models for the divergence of broods in the three species groups

| Divergence scenario and gene flow pattern | No. of parameters | Log likelihood | AIC | \(\Delta\)AIC | AIC weight | Bootstrap proportion of the model |
|------------------------------------------|-------------------|----------------|-----|----------------|-------------|---------------------------------|
| **Decim (M. neotredecim and M. septendecim)** | | | | | | |
| S1: (II, (XIX, XXIII)) | 9 | −9159.70 | 18,337,398 | 55.89 | 5.7E-13 | 0.00 |
| (1) No gene flow | | | | | | |
| (2) Past/Recent | 19 | −9129.30 | 18,296,592 | 15.08 | 0.0004 | 0.01 |
| (3) Recent only | 15 | −9125.76 | 18,281,512 | 0.00 | 0.7806 | 0.92 |
| S2: (II, (XXIII, (III, XIX))) | 9 | −9211.42 | 18,440,838 | 159.33 | 2.0E-35 | 0.00 |
| (1) No gene flow | | | | | | |
| (2) Past/Recent | 19 | −9128.76 | 18,295,514 | 14.00 | 0.0007 | 0.00 |
| (3) Recent only | 15 | −9129.98 | 18,289,966 | 8.45 | 0.0114 | 0.05 |
| S3: (II, (XIX, (XXIII))) | 9 | −9150.28 | 18,318,564 | 37.05 | 7.0E-09 | 0.00 |
| (1) No gene flow | | | | | | |
| (2) Past/Recent | 19 | −9127.59 | 18,293,174 | 11.66 | 0.0023 | 0.00 |
| (3) Recent only | 15 | −9127.10 | 18,284,19 | 2.68 | 0.2046 | 0.02 |
| **Cassini (M. tredecassini and M. cassini)** | | | | | | |
| S1: (II, (XIX, XXIII)) | 9 | −7977.40 | 15,972,792 | 360.83 | 2.7E-79 | 0.08 |
| (1) No gene flow | | | | | | |
| (2) Past/Recent | 19 | −7789.05 | 15,616,092 | 4.13 | 0.0770 | 0.02 |
| (3) Recent only | 15 | −7790.98 | 15,611,962 | 0.00 | 0.6068 | 0.75 |
| S2: (II, (XXIII, (III, XIX))) | 9 | −8199.22 | 16,416.44 | 804.48 | 1.2E-175 | 0.01 |
| (1) No gene flow | | | | | | |
| (2) Past/Recent | 19 | −7792.81 | 15,623,612 | 11.65 | 0.0018 | 0.00 |
| (3) Recent only | 15 | −7792.62 | 15,615,234 | 3.27 | 0.1182 | 0.08 |
| S3: (II, (XIX, (XXIII))) | 9 | −7849.35 | 15,716.70 | 104.73 | 1.1E-23 | 0.03 |
| (1) No gene flow | | | | | | |
| (2) Past/Recent | 19 | −7792.77 | 15,623,538 | 11.58 | 0.0019 | 0.00 |
| (3) Recent only | 15 | −7792.12 | 15,614,238 | 2.28 | 0.1944 | 0.03 |
| **Decula (M. tredenica and M. septendeculca)** | | | | | | |
| S1: (II, (XIX, XXIII)) | 9 | −6272.00 | 12,562,006 | 1095.14 | 5.9066E-239 | 0.00 |
| (1) No gene flow | | | | | | |
| (2) Past/Recent | 19 | −5717.92 | 11,473,838 | 6.97 | 0.0161 | 0.00 |
| (3) Recent only | 15 | −5718.67 | 11,467,338 | 0.47 | 0.2989 | 0.32 |
| S2: (II, (XXIII, (III, XIX))) | 9 | −6208.73 | 12,435,464 | 968.60 | 1.77E-121 | 0.00 |
| (1) No gene flow | | | | | | |
| (2) Past/Recent | 19 | −5718.51 | 11,475,01 | 8.14 | 0.0064 | 0.00 |
| (3) Recent only | 15 | −5718.66 | 11,467,314 | 0.45 | 0.3025 | 0.22 |
| S3: (II, (XIX, (XXIII))) | 9 | −6098.49 | 12,214,98 | 748.11 | 1.33E-163 | 0.00 |
| (1) No gene flow | | | | | | |
| (2) Past/Recent | 19 | −5719.41 | 11,476,816 | 9.95 | 0.0026 | 0.00 |
| (3) Recent only | 15 | −5718.43 | 11,466,868 | 0.00 | 0.3780 | 0.46 |

The best fit model among nine models (three divergence scenarios and three patterns of gene flow; Supplementary Fig. 1) was selected using the AIC method. Bold letters indicate the best fit models with smallest AIC.

AIC: Akaike Information Criterion. \(\Delta\)AIC: AIC difference from the best model.

---

ARTICLE COMMUNICATIONS BIOLOGY | DOI: 10.1038/s42003-018-0025-7 | www.nature.com/commsbio
in our previous study\textsuperscript{6}, suggesting divergence within the last 23 kya (i.e. after the last glacial maximum). Because estimated tMRCA for the species groups are similar between this study and our previous study (see Fig. 3 and Supplementary Table 2), the short divergence times revealed by mitochondrial data may reflect recent mitochondrial introgression between geographically adjacent broods. If our new inferences are correct, it follows that both life cycles have persisted in all species groups since at least the beginning of the last glacial period.

Notably, our demographic inference showed that gene flow has occurred between the 13- and 17-year species in each species group, particularly between species with geographically adjacent

\begin{figure}[h!]
\centering
\includegraphics[width=\textwidth]{genomic_sequence_divergence.png}
\caption{Genomic sequence divergence between 13- and 17-year species within species groups. Frequency distributions of $F_{ST}$ of individual SNPs (left), $F_{ST}$ of loci (middle) and Tajima’s $D$ (right) are shown for $M. \text{tredecim}$ vs. $M. \text{septendecim}$ (a), $M. \text{neotredecim}$ vs. $M. \text{septendecim}$ (b), $M. \text{tredecassini}$ vs. $M. \text{cassini}$ (c), and $M. \text{tredecula}$ vs. $M. \text{septendecula}$ (d). Median and mean values for SNP $F_{ST}$: a median = 0.000, mean = 0.135; b median = 0.000, mean = −0.007; c median = 0.000, mean = −0.002; d median = 0.000, mean = −0.007; for locus $F_{ST}$: a median = 0.1111, mean = 0.2146; b median = 0.0000, mean = −0.0012; c median = 0.0000, mean = 0.0055; d median = 0.0000, mean = −0.0049; for Tajima’s $D$: a median = −0.7995, mean = −0.6886; b median = −0.8015, mean = −0.7065; c median = −0.8015, mean = −0.7046; d median = −0.7806, mean = −0.5952}
\end{figure}
13- and 17-year broods. Although it is difficult to discriminate ‘diverged populations with gene flow’ from ‘recently diverged populations with no gene flow after the divergence,’ our model comparison showed that models with no gene flow had the lowest likelihoods compared to other models with gene flow (Table 2).

In addition, the SVDquartets tree (Fig. 4a) showed low resolution for brood relationships; this lack of resolution makes sense because SVDquartets is designed to accommodate cases where gene flow is absent and incomplete lineage sorting is the source of gene tree incongruence.

Gene flow between neighbouring populations of 13- and 17-year broods (species) may have occurred in the year of their co-emergence, every 221 (13×17) years or during occasional off-schedule emergences of smaller number of individuals (called ‘stragglers’). The 13- and 17-year cicadas within species groups do not show clear morphological or behavioural differentiation;
hence they could potentially hybridize.\textsuperscript{2,37} The finding of gene flow between sister 13- and 17-year species may be odd, because historical records indicate stability of the boundary between 13- and 17-year broods.\textsuperscript{38} It is possible that the synchronised life cycle among individuals of each brood has been strongly selected, and thus is stable in the face of occasional gene flow.\textsuperscript{7}

We initially hypothesised that the difference between the two life cycles was controlled by a locus that regulates juvenile development and an ancestral polymorphism at the locus may have caused the parallel life cycle divergence through collateral genetic evolution. Our comparison of orthologous gene sequences between 13- and 17-year species, however, has not provided any substantial clues resolving the genetic basis of life cycle divergence. We searched for shared SNPs or diverged loci among the four pairs of 13- and 17-year species (i.e., including \textit{M. tredecim}), which may be related to the regulation of life cycles. Such shared SNPs/loci would show elevated \( F_{ST} \) and deep divergence if ancestral polymorphisms were responsible for cycle shifts; alternatively, shared SNPs/loci with shallow divergence would be detected if independent mutations were responsible. However, we found no divergent SNPs or loci that were shared by all pairs. Even if life cycle loci exist, they may be undetectable in reduced representation sequences such as the RNAseq used in this study, likely because the responsible regions are small regardless of whether they are ancestral polymorphisms or independent mutations. We also found that non-random parallel genomic divergence (in terms of \( F_{ST} \)) has not occurred among the four pairs of 13- and 17-year species, which may be expected in the parallel evolution of alternative phenotypes in different lineages.\textsuperscript{12,23} If life cycle is controlled at multiple genetic levels rather than by a single mutation or a single diverged locus, any mutation in a group of genes within the same pathway could trigger a life cycle shift.\textsuperscript{7} However, the results of functional annotation for the divergent loci between 13- and 17-year species showed no evidence of enrichment for a particular pathway or gene function. Thus, we have no conclusive information on the genetic control of life cycles at present.

Considering that we did not observe definitive genomic differences between the two life cycles, a non-genetic explanation for life cycle differences based on life cycle plasticity may not be ruled out completely. In a non-genetic scenario, different life cycles may be maintained by a threshold response of nympths to climatic factors such as the cumulative temperature during growing seasons. In fact, the geographic life cycle boundary (Fig. 1b) is predictable by local temperature data.\textsuperscript{39} However, such an environmentally cued life cycle control may be unstable under fluctuating climatic conditions. In either case (i.e., genetic or non-genetic control of life cycle regulation), it would be necessary to conduct a thorough comparison of the whole genomic sequences between closely related 13- and 17-year species to fully explore the nature of life-cycle divergence in periodical cicadas.

**Methods**

**RNA preparation and sequencing.** We sampled 28 individuals from the seven known species of \textit{Magicicada} (Supplementary Data 1). Four 13-year species were sampled from brood XIX (2011) and XXIII (2015), and three 17-year species from brood II (2013) and III (2014) during their emergences (Fig. 2a, Supplementary Data 1). Total RNA was extracted from hemolymph using QIAgen RNAasy. Libraries for sequencing were constructed and sequenced using the Illumina HiSeq2000 platform. Quality-filtered raw reads were deposited at the DNA Data Bank of Japan (DDBJ), in the DDBJ Read Archive (DRA).

De novo assembly and SNP calling. The quality-filtered sequence reads were de novo assembled using the Trinity assembler version r2014071740 with the default parameter settings. Samples from the species groups (Decim, Cassini and Decula) were pooled, and consensus contigs of species groups were assembled. Within the Decim group, \textit{M. tredecim} samples were separately assembled because \textit{M. tredecim} is clearly diverged from the monophyletic group that includes \textit{M. septendecim} and \textit{M. neotredecim}.\textsuperscript{2,20} Thus, we obtained consensus assemblies for \textit{M. tredecim} and the remaining Decim (\textit{M. septendecim}/\textit{M. neotredecim}), Cassini (\textit{M. cassini}/\textit{M. tredecassini}), and Decula (\textit{M. septendecula}/\textit{M. tredecula}) species.

SNPs for each sample were called as follows. Reads of samples were mapped to the consensus contigs using bowtie2 version 4.1.2,\textsuperscript{21} and variants were called with the ‘map’ command in SAMtools version 1.0.2 and the ‘call’ command in bcftools version 1.2.\textsuperscript{22} The varfile was filtered by the ‘mpileup’ command in SAMtools version 1.0.2 and the ‘call’ command in bcftools version 1.2.\textsuperscript{22} This step implemented the likelihood model with multi-sample SNP calling. Only SNPs supported with coverage ≥3 and a quality score ≥20 were retained. These SNPs were inserted into the contigs using the bcftools ‘consensus’ command, with heterogeneous sites retained using UPAarf-style ambiguity coding. Bases with coverage <3 were masked with N, and terminal Ns were removed. Contigs shorter than 300 bp were filtered out, and the longest isoform for each trinity sequence cluster was selected for downstream orthology clustering.

**Orthology clustering.** The consensus contigs of the samples were clustered into putative orthologous groups (loci) following the approach of Yang and Smith.\textsuperscript{12} In brief, all BLASTN\textsuperscript{43} searches were conducted on all contigs from different species. For each gene, we calculated the percentage of hits with an E-value <1 × 10\(^{-5}\) and sequence identity >50% among hits. Then, these homologous sequence clusters were aligned using Mafft version 7.1237,\textsuperscript{23} and initial homologous trees were built using RaxML version 8.2.4.\textsuperscript{24} Orthologous gene sequences were obtained following the ‘monophyletic outlier’ criterion\textsuperscript{44}, i.e., keeping the largest subtree that consisted exclusively of ingroup samples without duplication and monophyletic outlier samples. Clustering was conducted using the phylogenomic dataset construction scripts available at https://bitbucket.org/ yangra/phylogenomic_dataset_construction. To obtain the final alignments, consensus contigs were replaced by contigs with SNPs, and the sequences were realigned using PRANK version 140034\textsuperscript{27} using the default parameters. We retained orthologous clusters containing ≥22 \textit{Magicolora} samples (>95% of samples) as a final data set. Clusters with overall genetic variation greater than 0.05 were removed as putative erroneous clusters. The longest cluster sequences were used for phylogenetic inference.

**Phylogenetic inference.** The maximum-likelihood (ML) phylogeny of individual samples was estimated using RAXML version 8.2.4.\textsuperscript{28} with the concatenated alignment. RaxML was run using the ‘rapid bootstrap analysis and search for best- scoring ML tree’ algorithm with a GTR-G model and 100 bootstrap replicates. To estimate divergence time, the ML tree was converted to an ultrametric tree using the LSD version 0.3beta\textsuperscript{50}, with a calibration time of 3.89 mya at the node of the most recent common ancestor of all \textit{Magicolora}.\textsuperscript{6} Confidence intervals of node ages were obtained by 1000 bootstrap analysis. To account for the uncertainty for the time of the \textit{Magicolora} MRCA, we also estimated divergence times of 3.08 and 4.69 mya, which were the lower and upper values of the 95% highest probability density interval. For each node, the confidence interval was determined as the oldest and youngest ages of 95% confidence intervals obtained from 1000 bootstrap replicates. A brood-level population tree was constructed using SVQuartets\textsuperscript{11} implemented in PAUP* version 4.0a147.\textsuperscript{51} All clusters were concatenated, and SVQuartets was run using the ‘species tree’ option with 100 bootstrap replicates.

**ABBA-BABA test.** We used the ABBA-BABA test with the D-statistics to test whether introgressive hybridisation has occurred between different 13-year species from different species groups or between distinct lineages of the same species group (i.e., \textit{M. tredecim} vs. \textit{M. neotredecim} in the Decim group). Under the assumption that population P1 and P2 are derived from population P3 and outgroup O, the ABBA-BABA test searches for evidence of hybridisation between P3 and P1 or P2 by comparing the frequencies of the site patterns ABBA and BABA. We set 17- and 13-year broods in the same species group as P1 and P2, respectively, and set one of 13-year broods from different species group as P3. An outgroup (O) was chosen from the closest available outgroup species. We tested for hybridisation of four pairs of 13-year species with all four combinations of broods, totalling 28 comparisons (Table 1). D-statistics were calculated by a modified version of PyRAD version 3.0.66,\textsuperscript{22} which accepts a fasta alignment as an input. The standard deviation of the D-statistic was obtained by a bootstrap resampling with 1000 replications.

**Demographic inference and model selection.** We conducted demographic inference and model comparison using a method based on the site frequency spectrum (SFS) implemented in the software 2D.\textsuperscript{52} Independent SFSs were selected from the alignments of clusters, and folded joint SFSs of four populations representing 17-year broods II and III and 13-year broods XIX and
were used in the model comparison (Supplementary Fig. 1). We also generated a time difference between all current and ancestral populations; and three models were employed the repeated random haplotype sampling (RRHS) strategy when phase differentiation when it is highly negatively correlated with nucleotide diversity. Outlier analysis for diverged genomic portions sequences, node heights for seven clades are obtained (Supplementary Table 2). In the non-linear model with the time-dependent evolutionary rate equation:

\[
\text{Rate}(t) = \mu \exp(-\lambda t) + k,
\]

where \(\mu\) is the instantaneous mutation rate, and \(\lambda\) is inversely proportional to the half-life of the rate decay, and \(k\) is a finite asymptotic evolutionary rate. As a result, we obtained estimates of these variables as \(\mu = 0.00894\), \(\lambda = 2.9185\) and \(k = 0.006849\) (per site per million generations). This value was used as the mutation rate in the demographic analysis.

We included the following three alternative scenarios in the model comparison, which are based on known phyleogeographic trees and the two life cycles:

**Scenario S1:** 13- and 17-year broods form monophyletic groups (II, III, (XIX, XXIII));

**Scenario S2:** geographically adjacent sampled broods form clades irrespective of their life cycles (II, (XXIII, (III, XIX)));

**Scenario S3:** 13-year species are monophyletic, and adjacent 17-year broods are closer to these (II, (III, (XIX, XXIII))).

To assess the effects of gene flow, we included three models of gene flow between broods under the three population divergence scenarios listed above. The three models were ‘no gene flow’; ‘past and recent gene flow’, where gene flow exists between all current and ancestral populations; and ‘recent gene flow only’, where gene flow only exists between current populations. In total, nine models were used in the model comparison (Supplementary Fig. 1).

We chose the best model using AIC values and AIC weights, calculated from composite likelihoods of the models, as recommended by Excoffier et al. In addition to model comparison with maximum likelihood inference, we performed bootstrap resampling of 100 replicates with Poisson approximation and recorded the bootstrap proportions, i.e., the proportions of replicates for which a given model was repeatedly chosen as the best model.

**Population genomic measures.** To characterise the within- and between-species genetic profiles of 17- and 13-year *Magicicada* species, population genomic measures were calculated for each orthologous cluster (locus). Genetic variation (\(\pi\)) and the number of segregating sites (\(S\)) within seven species were calculated. Tajima’s \(D^*\) was calculated for each locus. As a measure of nucleotide diversity, we used Tajima’s \(D^*\) within a locus as a measure of evolutionary rate. We also calculated the average number of pairwise differences, \(d_{xy}\), for each locus between 13- and 17-year species because this index is recommended as an absolute measure of population divergence.

Due to the sparse number of SNPs within loci and small sample sizes within populations, we were not able to reliably phase the genotypes. Therefore, we employed the repeated random haplotype sampling (RRHS) strategy when phase information was required. RRHS randomly assigns one of two possible genotypes at heterozygous sites. Thus, \(d_{xy}\) and Tajima’s \(D^*\) were repeatedly calculated with 100 RRHS replicates, and their averages were used as estimates.

**Outlier analysis for diverged genomic portions.** To detect diverged genomic portions associated with the divergences of 13- and 17-year species, we conducted outlier analyses of \(F_{st}\) for each SNP, \(F_{st}\) for each locus, and maximum \(F_{st}\) among all SNPs within each locus. \(F_{st}\) is an inappropriate measure of population differentiation when it is highly negatively correlated with nucleotide diversity. However, in our case, \(F_{st}\) was not correlated with mean nucleotide diversity at the locus level except for a weak negative correlation in the *M. tridentem/M. septen-decin pair* (Supplementary Fig. 2). Note that \(F_{st}\) and mean nucleotide diversity are expected to be uncorrelated with each other when demographic factors (e.g., gene flow, genetic drift) outweigh the effect of mutations, whereas a negative correlation is expected between these measures in the opposite situation. Meanwhile, the nucleotide divergence \(d_{xy}\), which is considered a more appropriate measure of population differentiation, was strongly positively correlated with nucleotide diversity and hence may lead to false discovery of elevated \(d_{xy}\) at regions with high nucleotide diversity (Supplementary Fig. 3). Thus, the use of \(F_{st}\) rather than \(d_{xy}\) was considered appropriate in the present case.

We defined the SNPs/loci with elevated \(F_{st}\) values as ‘divergent SNPs/loci’ to determine thresholds to define elevated \(F_{st}\), we used neither unlinked SNPs for the best-fitting demographic models selected above using fastsimcoal2 and calculated the number of shared divergent SNPs/loci responsible for the parallel life cycle divergence. The threshold value for the maximum \(F_{st}\) for a locus was determined by repeatedly taking a maximum of five \(F_{st}\) values of simulated SNPs to generate a distribution of maximum of \(F_{st}\) and obtaining the 95% quantile of this distribution. To determine the threshold to define elevated locus-level \(F_{st}\), we simulated linked sites of 2500 bp long for 5000 times (replicates) under the same demographic model. The weighted average of \(F_{st}\) values for SNPs in the linked sites was calculated each time, and the 95% quantile of the 5000 average \(F_{st}\) values was chosen as the threshold.

The number of divergent SNPs or loci with elevated \(F_{st}\) shared by two or more comparisons (i.e., ‘shared divergent SNPs/loci’) was considered an indicator of parallel divergence. The statistical significance of the numbers of shared divergent SNPs or loci was tested with permutation tests with 1000 replicates, which estimated the probability that the number of shared divergent SNPs or loci observed in each comparison were obtained by chance alone. For the shared divergent loci, functional annotations were made using DAVID Bioinformatics Resources 6.8.

**Data availability.** The raw sequence reads used in the present study are available from the DDBJ Read Archive (DRA) of the DNA Data Bank of Japan (DDBJ) (BioProject, PRJDB4567; BioSample, SAMD00047121). Other relevant data and input files used in the fastsimcoal2 runs are available via Figshare with the handle https://doi.org/10.5061/d8.5f471e126347.

Received: 20 September 2017 Accepted: 1 March 2018

Published online: 19 April 2018

**References**

1. Roff, D. A. *Life History Evolution* (Sinauer, Sunderland, MA, 2002).
2. Williams, K. S. & Simon, C. The ecology, behavior, and evolution of periodical cicadas. *Annu. Rev. Entomol.* 40, 269–295 (1995).
3. Alexander, R. D. & Moore, T. E. 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 (1962).
4. Marshall, D. C. & Cooley, J. R. Reproductive character displacement and speciation in periodical cicadas, with description of a new species, 13-year Magicicada neotredecim. *Evolution* 54, 1313 (2000).
5. Cooley, J. R., Simon, C., Marshall, D. C., Slon, K. & Ehrrhardt, C. Allochronic speciation, secondary contact, and reproductive character displacement in periodical cicadas (Hemiptera: Magicicada spp.): genetic, morphological, and behavioural evidence. *Mol. Ecol.* 10, 661–671 (2001).
6. Sota, T. et al. Independent divergence of 13- and 17 year life cycles among three periodical cicada lineages. *Proc. Natl. Acad. Sci. USA* 110, 6919–6924 (2013).
7. Lloyd, M. & Dybas, H. S. The periodical cicada problem II. Evolution. *Evolution* 20, 466–505 (1966).
8. Cox, R. T. & Carlton, C. E. Paleoclimatic influences in the evolution of periodical cicadas (Insecta: Homoptera: Cicadidae: Magicicada spp.). *Am. Midl. Nat.* 120, 183–193 (1988).
9. Yoshimura, J. The evolutionary origins of periodical cicadas during ice ages. *Am. Nat.* 149, 112–124 (1997).
10. Lloyd, M., Kritsky, G. & Simon, C. A simple Mendelian model for 13- and 17-year life cycles of periodical cicadas, with historical evidence of hybridization between them. *Evolution* 37, 1162–1180 (1983).
11. Cox, R. T. & Carlton, C. E. Evidence of genetic dominance of the 13-year life cycle in periodical cicadas (Homoptera: Cicadidae: Magicicada spp.). *Am. Midl. Nat.* 125, 63–74 (1991).
12. White, J. A. & Lloyd, M. Growth rates of 13- and 17-year periodical cicadas. *Am. Midl. Nat.* 94, 127–143 (1975).
14. Marshall, D. C., Cooley, J. R. & Hill, K. B. R. Developmental plasticity of life-cycle length in thirteen-year periodical cicadas (Hemiptera: Cicadidae). *Ann. Entomol. Soc. Am.* 104, 443–450 (2011).

15. West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford University Press, New York, 2003).

16. Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* 14, 751–764 (2013).

17. Elmer, K. R. & Meyer, A. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol. Evol.* 26, 298–306 (2011).

18. Cox, R. T. & Carlton, C. E. A comment on gene introgression versus an masse cycle switching in the evolution of 13-year and 17-year life cycles in periodical cicadas. *Evolution* 57, 428–432 (2003).

19. Simon, C. et al. 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, 1326–1336 (2000).

20. Young, M. & Kay, S. A. Time zones: a comparative genetics of cicadian clocks. *Nat. Rev. Genet.* 2, 702–715 (2001).

21. Kosol, V. Insect photoperiodic calendar and cicadian clock: independence, cooperation, or unity? *J. Insect Physiol.* 57, 538–556 (2011).

22. Yamanaka, N., Revitz, K. F. & O’Connor, M. B. Ec dysone control of developmental transitions: lessons from Drosophila research. *Annu. Rev. Genet.* 47, 497–516 (2013).

23. Nijhout, H. F. et al. The developmental control of size in insects. *Annu. Rev. Entomol.* 48, 443–450 (2003).

24. Marshall, D. C. Periodical cicada (Hemiptera: Cicadidae) life-cycle variations, the historical emergence record, and the geographic stability of brood distributions. *Ann. Entomol. Soc. Am.* 94, 386–399 (2001).

25. Cooley, J. R., Marshall, D. C., Simon, C., Neckermann, M. L. & Bunker, G. At the limits: habitat suitability modelling of northern 17-year periodical cicada extinctions (Hemiptera: Magicicada spp.). *Glob. Ecol. Biogeogr.* 22, 410–421 (2013).

26. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data. *Nat. Biotechnol.* 29, 644–652 (2011).

27. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359 (2012).

28. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping, and population genetic parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993 (2011).

29. Misof, B. et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767 (2014).
