

Idiosyncratic genome degradation in a bacterial endosymbiont of periodical cicadas

Matthew A. Campbell^a, Piotr Łukasik^a, Chris Simon^b, and John P. McCutcheon^{a,1}

^a Division of Biological Sciences, University of Montana, Missoula, MT, USA

^b Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA

¹ Correspondence: john.mccutcheon@umontana.edu

Summary

When a free-living bacterium transitions to a host-beneficial endosymbiotic lifestyle, it almost invariably loses a large fraction of its genome [1, 2]. The resulting small genomes often become unusually stable in size, structure, and coding capacity [3-5]. *Candidatus* *Hodgkinia* *cicadicola* (*Hodgkinia*), a bacterial endosymbiont of cicadas, sometimes exemplifies this genomic stability. The *Hodgkinia* genome has remained completely co-linear in some cicadas diverged by tens of millions of years [6, 7]. But in the long-lived periodical cicada *Magicicada tredecim*, the *Hodgkinia* genome has split into dozens of tiny, gene-sparse genomic circles that sometimes reside in distinct *Hodgkinia* cells [8]. Previous data suggested that other *Magicicada* species harbor similarly complex *Hodgkinia* populations, but the timing, number of origins, and outcomes of the splitting process were unknown. Here, by sequencing *Hodgkinia* metagenomes from the remaining six *Magicicada* species and two sister species, we show that all *Magicicada* species harbor *Hodgkinia* populations of at least twenty genomic circles each. We find little synteny among the 256 *Hodgkinia* circles analyzed except between the most closely related species. Individual gene phylogenies show that *Hodgkinia* first split in the common ancestor of *Magicicada* and its closest known relatives, but that most splitting has occurred within *Magicicada* and has given rise to highly variable *Hodgkinia* gene dosages between cicada species. These data show that *Hodgkinia* genome degradation has proceeded down different paths in different *Magicicada* species, and support a model of genomic degradation that is stochastic in outcome and likely nonadaptive for the host. These patterns mirror the genomic instability seen in some mitochondria.

Keywords: *Hodgkinia cicadicola*, *Magicicada*, endosymbiosis, nonadaptive evolution, organelle genomes, mutation, periodical cicadas, levels of selection

Results

Like all sap-feeding insects, cicadas depend on specialized endosymbiotic microorganisms for supplementation of their nutrient-poor diet of plant sap [9-11]. One of these microbes, an alphaproteobacterium called *Hodgkinia*, is associated with all reported cicadas [10]. As is typical for bacterial endosymbiont genomes, *Hodgkinia*'s genome is extremely reduced (~150kb at its largest in some cicadas), rendering it completely dependent on its cicada host and partner bacterial endosymbiont *Sulcia* for basic biological function [4]. Despite a very high rate of sequence evolution [8], *Hodgkinia* genomes can be remarkably stable in structure and gene content, changing little in cicadas diverged by tens of millions of years [7, 12]. In other cicadas, *Hodgkinia* has evolved into two or more distinct but co-dependent genomic and cellular lineages that are present in individual hosts, which have undergone reciprocal gene inactivation and loss [7]. We refer to this unusual process as “splitting”. We have shown that in the long-lived periodical cicada *Magicicada tredecim*, *Hodgkinia* has split into dozens of small distinct genomic circles encoding few recognizable genes [8]. Here we analyze the *Hodgkinia* metagenomes of the six remaining *Magicicada* species in order to understand the timing and outcome of this splitting process across the cicada genus.

***Hodgkinia* is complex in all *Magicicada* species.** Our new sequencing data confirm [8] that *Hodgkinia* is comprised of many distinct genomic circles in all species of *Magicicada* (Fig. 1, Tables 1, S1). We refer to individual circular-mapping *Hodgkinia* genomic contigs as ‘circles’

because, though we know some reside in distinct cells [8], we currently do not know whether most of these molecules are chromosomes that share the same cell or genomes representing different cell types [8]. We refer to the total complement of *Hodgkinia* contigs assembled in a single species of cicada—whether closed into circles or not—as that species’ *Hodgkinia* Genome Complex (HGC). The smallest HGC is found in *M. tredecula* and consists of at least 153 contigs totaling 1.20 megabases (Mb) of DNA, while the largest is from *M. neotredecim* and consists of 215 contigs totaling 1.58 Mb (Table 1). In each *Magicicada* species, we identified between 26 and 42 contigs with large-insert mate pair data suggesting they were circular DNA molecules. We were able to fully close these contigs into circular molecules in at least 20 instances in all cicada species (Fig. 1, Table 1). Contigs with mate pair, PCR, and/or Sanger sequence data supporting their circularity were considered putative circles if they were not fully closed (Fig. 1). The combination of confirmed and putative circular molecules comprised between 51.4% and 72.5% of the total DNA in each HGC (the remaining contigs lack end-joining data; Fig. 1). Individual completed circles range in size from 0.69kb to 70.5kb, contain a maximum of 27 genes, a minimum of 1 gene (with a single exception, a circle encoding only a single pseudogene), and span as much as a 653-fold range of sequencing coverage (Table S1). There is an even higher range in coverages for contigs that did not assemble into circular molecules: we find that contigs from *Magicicada* HGCs span at least a 2,500-fold difference in average sequencing coverage in each *Magicicada* species (Table 1).

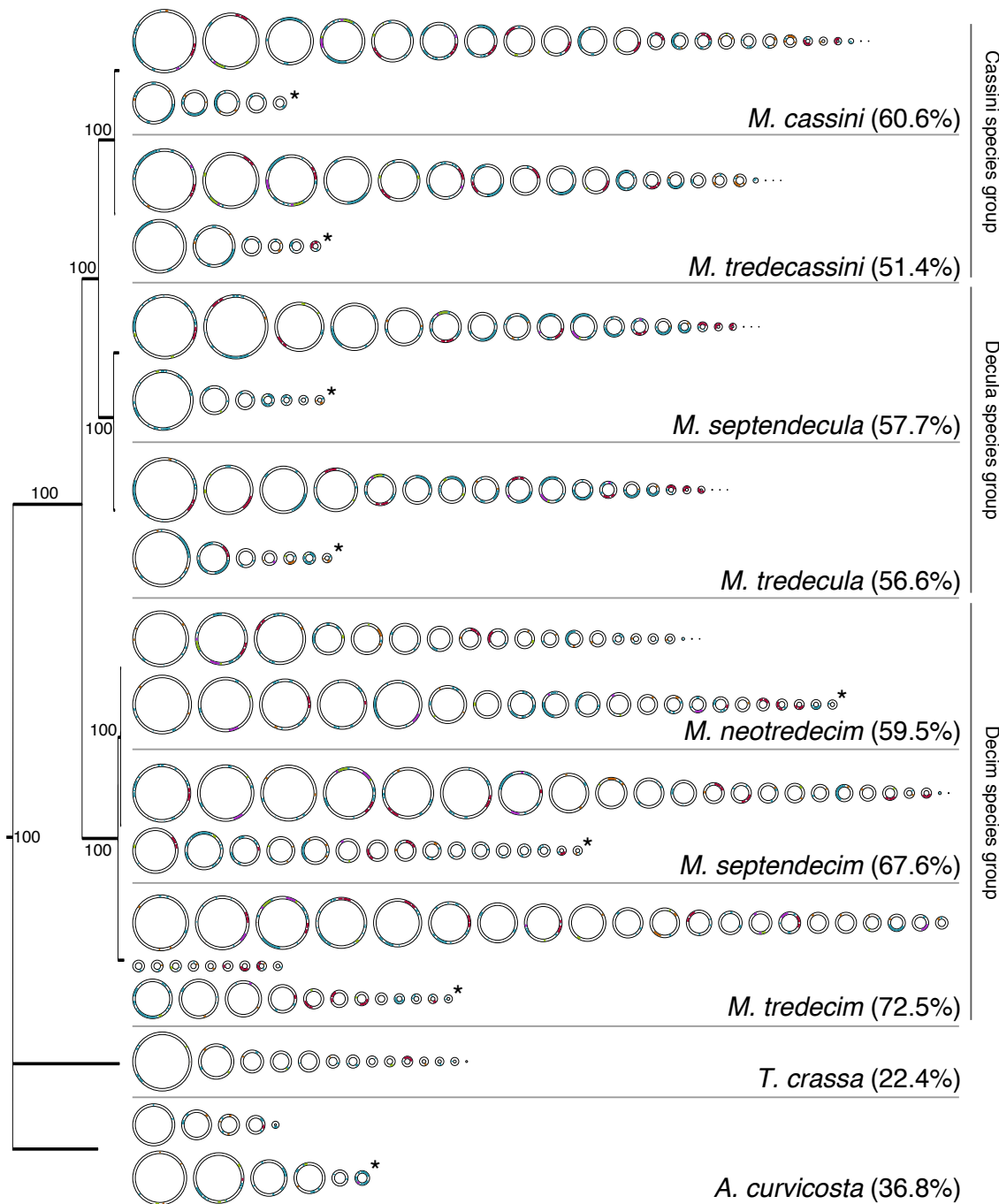


Figure 1: *Hodgkinia* genomic complexity in all study species.

Left: Phylogeny of the cicada species used in this study, based on all 13 protein-coding and both ribosomal genes from the mitochondrial genomes. The cicada *Diceroprocta semicineta* was used to root the tree, but was not included in the figure. Bootstrap support values are shown on each resolved node. Right: Diagrams representing the confirmed and putatively circular molecules of the HGC in all study species. Rows with an asterisk at the end represent putative circular molecules. On each circle, red regions represent rRNA genes, green represents histidine synthesis genes, orange represents cobalamin synthesis genes, purple represents methionine synthesis genes, blue represents all other genes, and white space represents noncoding DNA. Values in parentheses indicate the proportion of total *Hodgkinia* DNA from each cicada species represented by the circular molecules. The three species groups are annotated next to the species labels.

We identified between 135 and 145 of the 186 unique protein- and RNA-coding genes annotated as functional in at least one of *D. semicineta* and *T. ulnaria* in each *Magicicada* HGC [7, 12]. In all species, several additional genes were identified as truncated fragments or obvious pseudogenes. Because of the very low coverage of some contigs and the extremely rapid rate of *Hodgkinia* sequence evolution [8], it is likely that many of the remaining genes are present but were either not fully assembled or are not recognizable due to their low sequence similarity to other annotated *Hodgkinia* genes.

We also sequenced *Hodgkinia* from two cicada species that are closely related to *Magicicada* [13, 14] (Fig. 1, Table 1). The HGCs from Australian cicada species *Aleeta curvicosta* and *Tryella crassa* are similar in many ways to HGCs from *Magicicada*, but somewhat less complex. However, we generated less sequencing data for these species (~8.7 gigabases (Gb) for *A. curvicosta*, ~1.5 Gb for *T. crassa*, compared with an average of ~30.5 Gb per species of *Magicicada*) and so this relative simplicity may be due to sequencing effort. We therefore primarily focus our analyses on the *Hodgkinia* genomes from *Magicicada* species. Nevertheless, data from these outgroup species allowed us to infer whether *Hodgkinia* lineage splitting began in

| Species | Species abbreviation | Number of contigs | Total HGC size (Mb) | Total # of circles | Cumulative size of circles (Mb) | Unique genes | Total genes | Fold-coverage difference |
|------------------------|----------------------|-------------------|---------------------|--------------------|---------------------------------|--------------|-------------|--------------------------|
| <i>M. cassini</i> | MAGCAS | 118 | 1.27 | 29 | 0.77 | 142 | 306 | 6,376 |
| <i>M. tredecassini</i> | MAGTCS | 201 | 1.42 | 26 | 0.73 | 145 | 316 | 5,494 |
| <i>M. septendecula</i> | MAGSDC | 119 | 1.22 | 27 | 0.71 | 139 | 297 | 4,827 |
| <i>M. tredecula</i> | MAGTDC | 153 | 1.20 | 27 | 0.68 | 140 | 318 | 3,189 |
| <i>M. neotredecim</i> | MAGNEO | 215 | 1.68 | 41 | 1.00 | 139 | 333 | 3,379 |
| <i>M. septendecim</i> | MAGSEP | 166 | 1.64 | 39 | 1.11 | 137 | 314 | 5,723 |
| <i>M. tredecim</i> | MAGTRE | 118 | 1.53 | 42 | 1.11 | 135 | 305 | 2,500 |
| <i>T. crassa</i> | TRYCRA | 106 | 1.16 | 14 | 0.26 | 137 | 203 | 947 |
| <i>A. curvicosta</i> | ALECUR | 138 | 0.95 | 11 | 0.35 | 136 | 199 | 830 |

Table 1. Summary statistics for all HGCs described in this work. Total HGC size is a sum of all *Hodgkinia* contigs no matter whether they are closed into circular molecules or not. The number of unique genes found in other *Hodgkinia* genomes range from 168-183.

Magicicada or whether independent *Hodgkinia* lineages existed before *Magicicada* diverged from its common ancestor.

The origin of *Hodgkinia* lineage splitting predates the diversification of the genus *Magicicada*. To determine whether *Hodgkinia* lineage splitting started in the *Magicicada* genus or predated its origin, we reconstructed phylogenetic trees for *Hodgkinia* genes present in multiple copies in at least one of the two outgroup species. If splitting began independently in each genus, we would expect phylogenetic trees inferred from individual *Hodgkinia* genes to be monophyletic within each cicada genus (Fig. S1A). However, if splitting predated the divergence of *Magicicada* from *Tryella/Aleeta*, then gene trees should show two or more strongly supported monophyletic clades, each consisting of copies of genes from *Magicicada* along with *Tryella* and *Aleeta* (Figs. S1B-D). Though we see some cases where gene trees form monophyletic groupings within cicada genera (Fig. S1A), we also find several instances where gene phylogenies reveal two (Fig. S1B-C) or three (Fig. S1D) well-supported clades that group *Magicicada* genes with at least one gene copy from *Tryella* or *Aleeta*. It is possible to see both patterns because not all redundant genes from split lineages are retained in the new lineages [7]. Overall these patterns show that lineage splitting in *Hodgkinia* began before the *Magicicada*, *Tryella*, and *Aleeta* diverged from one another. We estimate that the last common ancestor of these genera had a minimum of three *Hodgkinia* lineages (Fig. S2A-C), similar to the complexity of *Hodgkinia* in the cicada *Tettigades undata* [7].

***Hodgkinia* lineage splitting is ongoing in *Magicicada* species groups.**

Having found evidence that *Hodgkinia* splitting had started prior to the divergence of *Magicicada* from its ancestor with *Tryella* and *Aleeta*, we tried to assess whether most circular molecules were formed prior to the diversification of *Magicicada* and were conserved throughout the genus, or if lineage splitting is a process that has been ongoing since the origin of *Magicicada*. If most lineage splitting occurred in the ancestor to all *Magicicada*, then most gene copies should have representatives in all extant cicada species. If most lineage splitting occurred within species groups, then many gene copies should be unique to species groups and not shared throughout the genus. Gene phylogenies generated with six representative *Hodgkinia* genes show multiple but

relatively few well-supported clades with representatives of all three *Magicicada* species groups (Fig. S2). In some cases, we identified only one gene copy per HGC, with gene phylogenies that recapitulate host phylogeny (Fig. S2A). This pattern suggests that there was a single copy of the gene in the last common ancestor of *Magicicada*, and that the genomic circle it was on may not have undergone splits. In other cases, the single ancestral gene copy co-diversified with hosts, but also underwent splits in some species groups (Fig. S2B). Trees of other genes form between two and five highly supported clades that include copies from all species groups (Fig. S2C-F), showing evidence for variable amount of additional splitting after species groups diverged (Fig. S2G). These phylogenies show that a minimum of five distinct *Hodgkinia* lineages existed in the last common ancestor of *Magicicada*.

These data suggest that most of the splitting we see in Fig. 1 happened after *Magicicada* started to diversify. If this is true, we expect that the similarity of HGCs should diminish as a function of cicada phylogenetic distance. In comparing extant circular molecules between cicada species, we find few clearly homologous circles with identical gene sets conserved in all *Magicicada* species. Because comparative genomic methods are generally based on sequence similarity and synteny comparisons, and we found little obvious synteny to compare, we developed a metric based on the Jaccard Index [15] to quantify the similarity in gene content of the finished circles between cicada species. We call this metric the *Hodgkinia* Similarity Index (HSI, Fig. 2). We calculate the HSI as follows, for hypothetical circular molecules A and B (**Equation 1**):

$$HSI = \frac{Genes\ in\ A \cap Genes\ in\ B}{Genes\ in\ A \cup Genes\ in\ B} \times \frac{Length\ (in\ bp)\ of\ smaller\ of\ A\ and\ B}{Length\ (in\ bp)\ of\ larger\ of\ A\ and\ B}$$

Briefly, a finished circular molecule of one cicada species is compared to a circular molecule of another cicada species. We calculate the Jaccard Index of the two gene sets (the intersection of gene sets divided by the union, the left half of Eq. 1) and multiply that by the ratio of the length of the smaller circle divided by the length of the larger one (right half of Eq. 1). We calculate this pairwise value for all circles of a

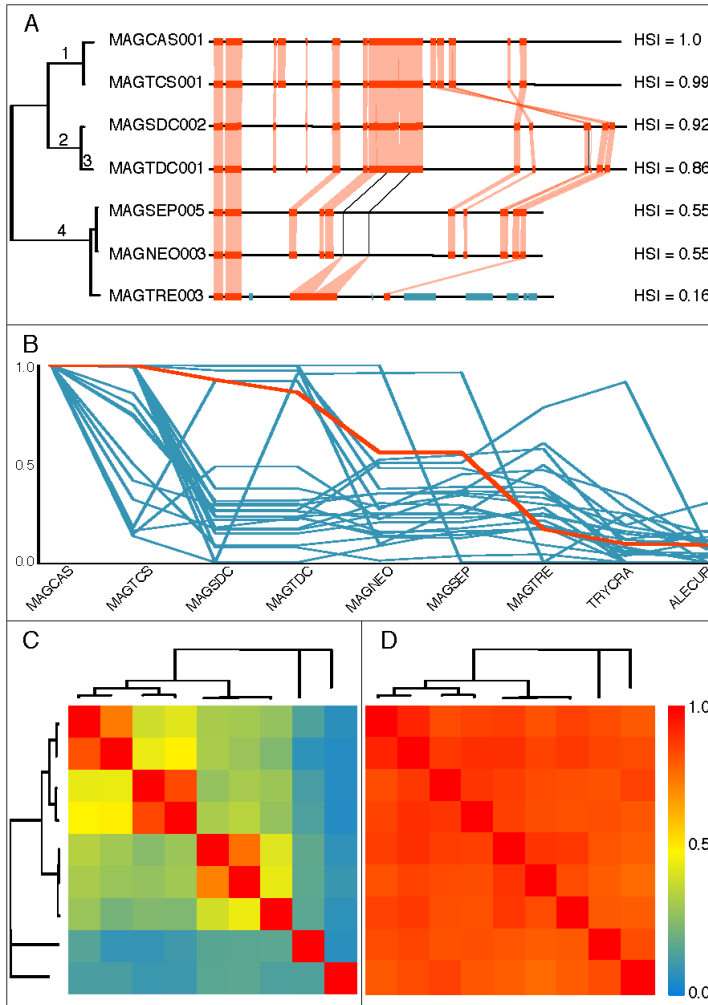


Figure 2: HSI scores for individual circles and HGCs.

(A) Illustration of synteny conservation between homologous circular molecules. Shown is the reference circle MAGCAS001 and the circle most similar to it from all other *Magicicada* species (abbreviations are taken from Table 1). Horizontal black lines represent the genome backbone, orange boxes are genes shared between a given circular molecule and MAGCAS001. Blue bars represent genes present in a given circular molecule but not present on MAGCAS001. Shaded vertical lines show gene homologs present on different circles, black lines connect putative homologs over gaps in some genomes. HSI scores between the reference and all circular molecules are shown on the right. Numbers on the phylogenies represent inferred mutational events on the respective lineage: (1) genome rearrangement, (2) and (3) individual gene loss events, (4) loss of five genes. The exact branch on which (4) occurred is ambiguous. Three contigs from *M. tredecim* seem to be homologous to the reference circle when joined together, but we could not close them to a single circle so they were not included in the HSI analysis. (B) Distribution of all HSI scores for *M. cassini*. The x-axis shows each species *M. cassini* was compared with, and the y-axis shows the HSI score. The bold orange line represents the circles shown in (A). (C) Heatmap showing pairwise average HSI scores between all species. (D) Heatmap showing pairwise average Jaccard Index of the whole *Hodgkinia* gene set in each species. In both B and C, a score of one indicates that the two species are identical, and zero indicates that they share no genes in common. All trees are taken from Figure 1.

species and report the average HSI score between those two cicada species. We then repeat this for all pairwise comparisons of cicada species. An HSI value of one indicates the two circles have the same functional genes and are the same length, whereas a value of zero indicates they share no common genes. Because the circles have on average very low coding densities and have apparently undergone rearrangements in some cases (Fig. 2A), this metric does not take gene co-linearity into account. It is also not (necessarily) a true measure of homology since it does not distinguish between conservation of an ancestral circle and convergent evolution to a similar state. Rather it is a rough metric to score the overall similarity of HGCs between cicada species in the absence of much calculable similarity (Fig. 2B). It is also a conservative metric, since there will undoubtedly be homologous circular molecules that were not completely assembled and thus not calculated in the HSI.

We find a strong phylogenetic signal in HSI scores, where HGCs between species pairs (*M. cassini*–*M. tredecassini*, *M. septendecula*–*M. tredecula*, and *M. septendecim*–*M. neotredicim*) are highly similar to one another (0.80 HSI on average, Fig. 2C). This is expected given that each of these species pairs are estimated to have diverged from each other less than 50 thousand years each ago [13]. The HSI scores degrade quickly with increased phylogenetic distance, however. Pairwise comparisons between *M. tredecim*–*M. neotredicim* and *M. tredecim*–*M. septendecim* (500 thousand years diverged [13]), Cassini species with Decula species (2.5 million years diverged [13]), and Cassini and Decula species with Decim species (4 million years diverged [13]) give average HSI scores of 0.43, 0.46, and 0.29, respectively. This lack of similarity is remarkable given that the HSI between the single *Hodgkinia* genomes of *Diceroprocta semicincta* and *Tettigades ulnaria*, which diverged more than 60 million years ago [16-19], is 0.88.

Our combined phylogenetic and HSI analyses suggest that splitting began in the ancestor of *Magicicada*, *Tryella*, and *Aleeta* (into 2-3 circles), continued somewhat in the ancestor of all *Magicicada* (into at least 5-6 circles), but that splitting accelerated dramatically (into 20+ circles) after *Magicicada* began diversifying.

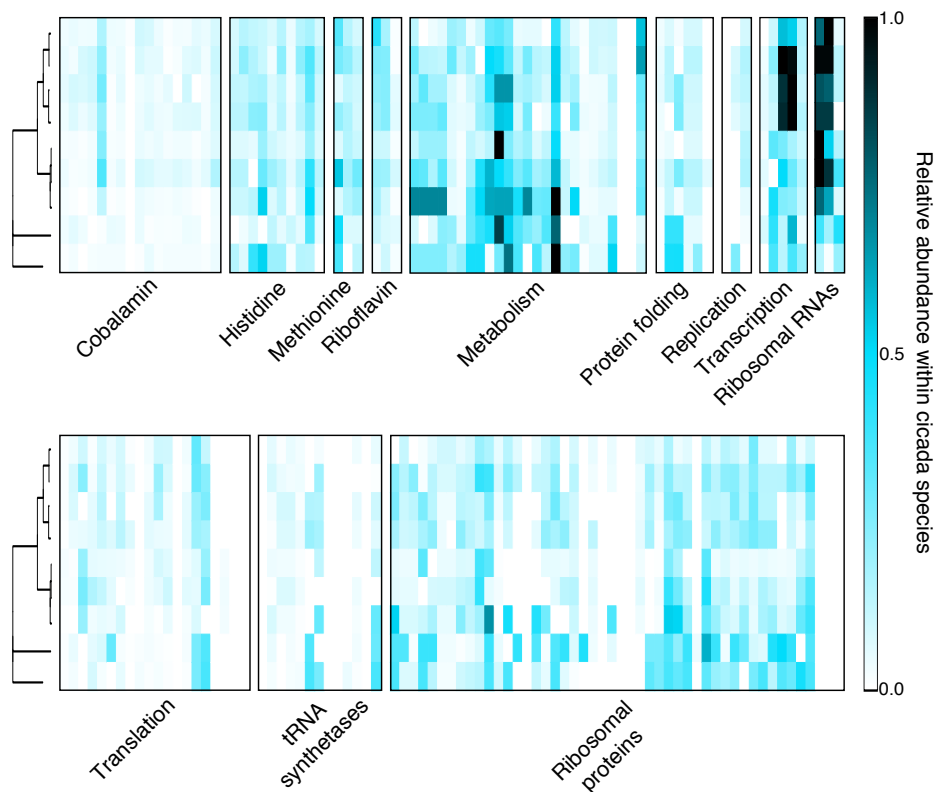


Figure 3: Relative gene abundance in all study species. Heatmaps showing the relative abundance of each gene in each species, ordered by gene category. A value of one (black) indicates the most abundant gene in that species, and zero (white) indicates that gene is absent in that species. Columns that are completely white represent genes that were not annotated in any species, so have either been lost, are present on broken contigs, or are present on contigs that did not otherwise assemble in our experiments. Trees are taken from Figure 1.

***Hodgkinia*'s overall function mostly remains intact.** The long-term stability of endosymbiont genomes is often attributed to the importance of their function to host survival [3, 20, 21]. Since *Hodgkinia* is clearly experiencing dramatic genomic instability, we wanted to test whether the complete ancestral *Hodgkinia* gene set was retained in HGCs in different *Magicicada* species. To directly compare gene complements between *Hodgkinia* HGCs, and to be consistent with the HSI, we calculated the Jaccard Index of each gene set for all pairwise comparisons of all *Magicicada* species. Similar to the HSI, a score of 1 would indicate that two cicada species have identical *Hodgkinia* gene sets, and a score of 0 would indicate that no genes are shared. We find that HGC gene sets within closely related species pairs are very similar (0.90 on average, Fig. 2D). Pairwise comparisons between *M. tredecim*-*M. neotredicim* and *M. tredecim*-*M. septendecim* (0.86), Cassini species with Decula species (0.87), and Cassini and Decula species with Decim species (0.86) also remain very similar, in contrast to the HSI scores calculated for these comparisons (compare Fig. 2C to 2D). These data indicate that while the patterns of *Hodgkinia* genome fragmentation is different in divergent *Magicicada* species, the overall set of retained genes is similar. For a sense of scale, the same analysis for cicadas diverged for dozens of millions of years [16-19], such as *Magicicada* and *D. semicincta*, *Magicicada* and *T. ulnaria*, and *D. semicincta* and *T. ulnaria* gives values of 0.82, 0.77, and 0.92, respectively. We note again that not all *Hodgkinia* genes present in *Magicicada* may have fully assembled due to the complexity of the dataset, so the true values for *Magicicada* HGCs may be higher than what we report here.

Lineage splitting leads to different gene dosages. It seemed possible that lineage splitting in *Hodgkinia* might be beneficial for the host, perhaps as a mechanism to control the dosage of *Hodgkinia* genes.

Under this hypothesis, we would expect similar gene dosages in comparisons of various *Magicicada* species. To calculate gene dosage in an HGC, we summed the average coverage of all contigs on which a given functional gene is found, scaled to the most abundant gene for each species. We find that the relative abundances of genes are similar within species groups (cicadas diverged less than 50 thousand years ago [13]), but not between species groups (Fig. 3). Principal coordinates analysis of relative gene abundances of all genes present in any species clusters the Decula and Cassini species groups together, *M. neotredicim* and *M. septendecim* together, and the remaining species – including *M. tredecim* separated from *M. neotredicim-septendecim* by only 0.5 Mya [13] – separately (Fig. S3A). This can be more clearly seen when only considering genes annotated in all species (Fig. S3B). This grouping is qualitatively similar to the HSI results, and suggests there is not the convergent pattern of gene dosage outcomes that might be expected if the host was dictating the process. Rather the gene dosage outcomes are stochastic and thus only similar in comparisons between very closely related cicadas.

Discussion

Many endosymbioses consist of two or more partners that are strictly reliant on one another for survival. The eukaryotic cell is now completely intertwined with and dependent upon its mitochondria (with one known exception [22]), and mitochondria cannot survive outside of their host cell. Similarly, cicadas require *Hodgkinia* for survival, and it is unlikely that *Hodgkinia* could survive outside of cicada cells. Despite these obligate host-symbiont co-dependencies, each partner can experience selection and drift independently of the other, so their evolutionary trajectories are not inevitably aligned and may directly

conflict with one another [23-32]. Although the engulfed partner is capable of exerting selfish tendencies in some cases [33-35], there exist several mechanisms for the host to constrain the evolution of its symbionts [36-39]. In bacterial endosymbioses, this host-level constraint is often reflected in the genomic stasis of the bacterial partner. Endosymbiont genomes can remain stable in gene content and structure for tens [3], hundreds [12], or even thousands [5, 40] of millions of years, and we interpret this stability as a reflection of host-dominated evolution for the preservation of endosymbiont function.

However, secondary genome instability subsequent to this stasis is now recognized as relatively common, especially in mitochondria [41, 42]. Mitochondrial genomic instability manifests both as genome reduction [43, 44] that sometimes leads to outright genome loss [45-49], and genome fragmentation [50-53] that sometimes leads to massive genome expansion with little obvious functional change [54-57]. We suggest that what unites these starkly different outcomes is a shift away from the host-driven constraint of the endosymbiont genome towards (sometimes temporary) symbiont-driven instability. In cases of mitochondrial reduction and loss, the host ecology changes such that the function of the organelle is no longer needed and therefore no longer under selective constraint from the host [46-48]. For example, many eukaryotes that live in anaerobic environments no longer require the oxidative respiratory function of their mitochondria, so the genes for this process are free to be lost [44]. The forces promoting mitochondrial genome fragmentation and expansion are less clear, but these expansions sometimes seem to be associated with increases in mitochondrial mutation rates [55] and have been hypothesized to result from less efficacious host-level selection against slightly deleterious symbiont mutations [57, 58].

Depending on whether one takes a *Hodgkinia*- or cicada-centric perspective, the outcomes we report here could be interpreted either as a genome reductive or genome expansive process [7, 8]. From the perspective of individual *Hodgkinia* lineages, each circular molecule in *Magicicada* gets smaller after a split, eventually resulting in circles less than half the size and encoding fewer than ~30% of the genes that were present on the already tiny ancestral *Hodgkinia* genome (Table S1)[6, 7]. This reductive process likely reflects the deletional mutation bias of bacteria [59], which in part explains the extremely small size of bacterial endosymbiont genomes in general [1]. In *Hodgkinia*, newly split lineages resemble a gene duplication event [7] that often results in one gene copy being pseudogenized and eventually lost entirely. *Hodgkinia*'s splitting and deletion process leads to individual circular molecules that resemble the extremely degraded genomes of mitochondria found in some eukaryotes. The idiosyncratic nature of these circles in closely related cicada species (Fig. 2C) is consistent with stochastic mutational loss and suggests a process with no particular goal or end point. But an important difference between cases of mitochondrial genome reduction and *Hodgkinia* from *Magicicada* is that the host ecology has not changed such that *Hodgkinia*'s functions are no longer required. The massive gene loss on individual *Hodgkinia* circles is likely only tolerable because, from the host's perspective, the combined HGCs seem to have retained *Hodgkinia*'s overall nutritional contribution to the symbiosis (Fig. 2D). From the host perspective, this splitting and genome reductive process results in a combined *Hodgkinia* "genome" size over an order of magnitude larger than the ancestral single genome (Table 1).

In our view, the most interesting parallel to what we report here for *Hodgkinia* can be found in the mitochondrial genomes of the angiosperm genus *Silene* [55, 60]. Like many plants, some *Silene*

mitochondrial genomes consist of a single "master circle" with multiple "subcircles" that arise primarily from recombination [61]. Other *Silene* species, though, have experienced dramatic increases in mitochondrial mutation rates, which seem to be accompanied by the expansion to dozens of enormous mitochondrial chromosomes [55]. These mitochondrial chromosomes, some encoding few or no detectable genes, can be rapidly lost or gained in closely related *Silene* lineages [60]. Like *Hodgkinia*, this diversity of genome expansion outcomes in closely related plant hosts is not accompanied by any detectable increase in functional capacity. We previously hypothesized that the increased complexity of *Hodgkinia* in *Magicicada* results from a similar increased effective mutation rate in *Hodgkinia* [8], with a conceptual modification related to lifecycle changes of the host cicada. While *Hodgkinia*'s inherent mutation rate may not be different in various cicada hosts, longer host lifecycles such as the 13- or 17-year lifecycle of *Magicicada* [62] may allow more symbiont generations and thus more *Hodgkinia* mutations per host lifecycle. We hypothesize that this increase in effective mutation rate enables *Hodgkinia*'s lineage splitting process and eventually results in stochastic differences between HGCs from different species groups at the level of genome structure (Fig. 2C). While *Hodgkinia* genes are (mostly) maintained in all HGCs, they are now present at wildly different abundances in different cicada species groups (Fig. 3). We hypothesize that lineage splitting and changes in gene dosages are either maladaptive or neutral for the host. The cicada does not benefit from *Hodgkinia* degeneration but must tolerate it because the cicada is wholly dependent on *Hodgkinia* for survival.

Materials and Methods

DNA extraction. Bacteriomes were dissected from a single male of *T. crassa*, a single female of *A. curvicauda* and *M. tredecim*, and two females of the remaining species. DNA was extracted from all dissected bacteriomes using a DNeasy Blood and Tissue kit (Qiagen catalog #69506). Extracted DNA was stored at -20C.

Library preparation and sequencing. Genomic DNA from *M. tredecim* was sheared to an average fragment size of 550 base pairs using a Covaris E220. Sheared DNA was made into a sequencing library using the NEBNext Ultra DNA Library Prep Kit for Illumina (catalog #E7370S), according to the standard protocol. The library was sequenced at the University of Montana Genomics Core on a MiSeq benchtop sequencer with a v3 600 cycle kit.

Genomic DNA from *A. curvicauda* was sheared to an average size of 480 base pairs using a Covaris E220. Sheared DNA was made into a sequencing library using a TruSeq PCR-free kit (Illumina) and sequenced as ~1/12 of a multiplexed lane at NGX Bio in San Francisco, CA using a HiSeq 2500 Rapid SBS kit (Illumina).

Genomic DNA from *T. crassa* was sheared to an average of 570 base pairs using a Covaris E220. Sheared DNA was made into a sequencing library using the NEBNext Ultra DNA Library Prep Kit for Illumina (catalog #E7370S), according to the standard protocol. The library was sequenced as ~1/4 of a multiplexed lane at the University of Montana Genomics Core on a MiSeq benchtop sequencer with a v3 600 cycle kit.

Genomic DNA from *M. neotredecim*, *septendecim*, *tredecassini*, *cassini*, *tredecula*, and *septendecula* was sheared to an average of 500 base pairs using a Covaris E220. Sheared DNA was made into a sequencing library using the NEBNext Ultra DNA Library Prep Kit for Illumina (catalog #E7370S), according to the standard protocol. Libraries were sequenced on two lanes on a HiSeq 2500 with 250 cycles

at the Johns Hopkins School of Medicine Genetic Resources Core Facility.

Genomic DNA from *M. neotredecim*, *septendecim*, *tredecassini*, *cassini*, *tredecula*, and *septendecula* was used for making libraries with a Nextera Mate Pair Sample Prep Kit (Illumina), according to the standard protocol. These libraries were sequenced on a single lane on a HiSeq 2500 with 100 cycles at the Case Western Reserve University Genomics Core Facility.

Genome assembly and annotation. Raw reads were quality filtered using Trimmomatic version 0.35 [63]. Remaining reads were further filtered using fastq_quality_filter from FASTX version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/).

Assembly of the filtered reads was done using Spades version 3.6.2 [64], using kmer sizes 127, 151, 191, and 291, both individually as well as combined together. Putative *Hodgkinia* contigs were identified with TBLASTN 2.2.31+ [65] with an E-value cutoff of 10e-5 using previously annotated *Hodgkinia* genes as the query. To remove redundant contigs, all putatively *Hodgkinia* contigs were queried against themselves, and any contig $\geq 97\%$ identical to another over $\geq 80\%$ of its length was considered redundant and removed. Any contigs with BLASTN E-values less than 10e-10 to the mitochondrial genome were also removed. Coverage of individual contigs was calculated by the total coverage at each base, divided by the length of the contig. Completely assembled *Hodgkinia* circles were identified based on sequence overlap on both ends of the contig. To identify putative circular contigs, filtered paired end and mate pair reads were mapped back to the assembly using BWA version 0.7.12-r1039 [66] with default parameters. Contigs were considered putatively circular if more than five read pairs mapped with one mate mapping in the first 10% of the contig, while its mate mapped in the last 10% of the contig. Putatively circular contigs were then closed when possible by PCR and Sanger sequencing.

Annotation of the *Hodgkinia* circles was done using a custom Python pipeline based around the Jackhammer module of HMMER v. 3.1b2 [67], RNAmmer 1.2 [68], and Aragorn v1.2.36 [69]. Occasionally RNAmmer misannotated the 23S rRNA gene, so barnap 0.6 [70] was used for corrections. The completely closed *Hodgkinia* circles were then checked manually for any long open reading frames that could contain missing genes.

Phylogenetic analysis. Host phylogeny was reconstructed using RAXML v. 8.2.0 [71] based on manually inspected alignments of 15 mitochondrial genes (13 protein-coding and two rRNA) of the total length of 12744 bp, divided into four partitions corresponding to three codon positions and to rRNA genes. Rapid bootstrapping (100 replicates) was used to estimate node support.

To construct individual gene phylogenies, homologous nucleotide sequences were translated into amino acids and aligned using mafft v. 7.221 [74]. Visually inspected alignments were analyzed using RAXML v. 8.2.4 [71] using a PROTGAMMAWAG model of amino acid substitution and 100 bootstrap replicates. Trees were rooted using *Aleeta-Tryella* as outgroups (whenever they formed a single monophyletic clade), or alternatively on the longest branch separating well-supported clades that included species from all or most hosts in a comparison.

Comparative *Hodgkinia* genome analysis. To compare the homology of HGC circles between cicada species, a *Hodgkinia* Similarity Index (HSI) score was calculated for all pairwise comparisons of all circles, as explained in Results. The pair with the highest HSI score was kept for each circle.

To determine relative coverage of all *Hodgkinia* genes, the coverage of all *Hodgkinia* genes was summed based on the coverage of the contig on which it was annotated. These abundance values were then normalized based on the most abundant gene. Principal coordinates analysis was done using the R package Vegan 2.4-3 [72].

End Matter

Author Contributions and Notes

Conceptualization, M.A.C. and J.P.M.; Methodology, M.A.C. and P.L.; Formal Analysis, M.A.C.; Investigation, M.A.C.; Resources, C.S. and J.P.M.; Data Curation, M.A.C. and C.S.; Writing – Original Draft, M.A.C.; Writing – Review & Editing, M.A.C., P.L., C.S., and J.P.M.; Visualization, M.A.C.; Supervision, J.P.M.; Funding Acquisition, J.P.M.

The authors declare no conflict of interest.

Acknowledgments

We thank all members of the McCutcheon lab for helpful discussion and comments, and Scott Miller for suggesting the use of the Jaccard Index. Funding for the sequencing and analysis was supported by National Science Foundation grants IOS-1256680 and IOS-1553529, and National Aeronautics and Space Administration Astrobiology Institute Award NNA15BB04A. Funding for cicada collecting was provided by NSF DEB-09-55849.

References

1. McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nature Rev Microbiol* 10, 13–26.
2. Bennett, G. M., and Moran, N. A. (2015). Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10169–10176.
3. Tamas, I. (2002). 50 Million Years of Genomic Stasis in Endosymbiotic Bacteria. *Science* 296, 2376–2379.
4. McCutcheon, J. P. (2010). The bacterial essence of tiny symbiont genomes. *Curr Opin Microbiol* 13, 73–78.
5. Boore, J. L. (1999). Animal mitochondrial genomes. *Nucl. Acids Res.* 27, 1767–1780.
6. McCutcheon, J. P., McDonald, B. R., and Moran, N. A. (2009). Origin of an Alternative Genetic Code in the Extremely Small and GC-Rich Genome of a Bacterial Symbiont. *PLOS Genet* 5, e1000565.
7. Van Leuven, J. T., Meister, R. C., Simon, C., and McCutcheon, J. P. (2014). Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. *Cell* 158, 1270–1280.
8. Campbell, M. A., Van Leuven, J. T., Meister, R. C., Carey, K. M., Simon, C., and McCutcheon, J. P. (2015). Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont *Hodgkinia*. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10192–10199.
9. White, J., and Strehl, C. E. (1978). Xylem feeding by periodical cicada nymphs on tree roots. *Ecol Entomol* 3, 323–327.

10. Buchner, P. (1965). Endosymbiosis of animals with plant microorganisms (John Wiley & Sons).
11. Zientz, E., Dandekar, T., and Gross, R. (2004). Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiol. Mol. Biol. Rev.* 68, 745–770.
12. McCutcheon, J. P., McDonald, B. R., and Moran, N. A. (2009). Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15394–15399.
13. Sota, T., Yamamoto, S., Cooley, J. R., Hill, K. B. R., Simon, C., and 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, 6919–6924.
14. Moulds, M. S. (2003). An appraisal of the cicadas of the genus *Abricta* Stal and allied genera (Hemiptera: Auchenorrhyncha: Cicadidae). *Records of the Australian Museum.*
15. Jaccard, P. (1901). Etude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull. Soc. Vaudoise Sci. Nat.* 37, 547–579.
16. Cooper, K. W. (1941). *Davisia* bearcreekensis Cooper, a new cicada from the Paleocene, with a brief review of the fossil Cicadidae. *Am. J. Sci.* 239, 286–304.
17. Poinar, G., Jr, and Kritsky, G. (2012). Morphological conservatism in the foreleg structure of cicada hatchlings, *Burmaticicada proteran* gen., n. sp. in Burmese amber, *Dominicicada youngin* gen., n. sp. in Dominican amber and the extant *Magicicada septendecim* (L.) (Hemiptera: Cicadidae). *Historical Biology* 24, 461–466.
18. Poinar, G., Jr, Kritsky, G., and Brown, A. (2012). *Minyscapheus dominicanus* n. gen., n. sp. (Hemiptera: Cicadidae), a fossil cicada in Dominican amber. *Historical Biology* 103, 1–5.
19. Marshall, D. C., Hill, K. B. R., Moulds, M., Vanderpool, D., Cooley, J. R., Mohagan, A. B., and Simon, C. (2016). Inflation of Molecular Clock Rates and Dates: Molecular Phylogenetics, Biogeography, and Diversification of a Global Cicada Radiation from Australasia (Hemiptera: Cicadidae: Cicadellini). *Syst. Biol.* 65, 16–34.
20. Wernegreen, J. J. (2002). Genome evolution in bacterial endosymbionts of insects. *Nat. Rev. Genet.* 3, 850–861.
21. Patiño-Navarrete, R., Moya, A., Latorre, A., and Peretó, J. (2013). Comparative Genomics of *Blattabacterium cuenoti*: The Frozen Legacy of an Ancient Endosymbiont Genome. *Genome Biol. Evol.* 5, 351–361.
22. Karnkowska, A., Vacek, V., Zubáčová, Z., Treitl, S. C., Petrželková, R., Eme, L., Novák, L., Žárský, V., Barlow, L. D., Herman, E. K., et al. (2016). A Eukaryote without a Mitochondrial Organelle. *Curr. Biol.* 26, 1274–1284.
23. Bennett, G. M., McCutcheon, J. P., MacDonald, B. R., Romanovicz, D., and Moran, N. A. (2014). Differential Genome Evolution Between Companion Symbionts in an Insect-Bacterial Symbiosis. *mBio* 5, e01697–14–e01697–14.
24. Keeling, P. J., and McCutcheon, J. P. (2017). Endosymbiosis: The feeling is not mutual. *J. Theor. Biol.*
25. Eberhard, W. G. (1990). Evolution in bacterial plasmids and levels of selection. *Q. Rev. Biol.* 65, 3–22.
26. Otto, S. P., and Orive, M. E. (1995). Evolutionary consequences of mutation and selection within an individual. *Genetics* 141, 1173–1187.
27. Okasha, S. (2005). Multilevel Selection and the Major Transitions in Evolution. *Philosophy of Science* 72, 1013–1025.
28. Maynard Smith, J. (1964). Group selection and kin selection. *Nature* 201, 1145–1147.
29. Hurst, L. D., Atlan, A., and Bengtsson, B. O. (2015). Genetic Conflicts. *Q. Rev. Biol.* 71, 317–364.
30. Kiers, E. T., and West, S. A. (2015). Evolving new organisms via symbiosis. *Science* 348, 392–394.
31. West, S. A., Fisher, R. M., Gardner, A., and Kiers, E. T. (2015). Major evolutionary transitions in individuality. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10112–10119.
32. Sachs, J. L., Skophammer, R. G., and Regus, J. U. (2011). Evolutionary transitions in bacterial symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 108 Suppl 2, 10800–10807.
33. Taylor, D. R., Zeyl, C., and Cooke, E. (2002). Conflicting levels of selection in the accumulation of mitochondrial defects in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3690–3694.
34. Bastiaans, E., Aanen, D. K., Debets, A. J. M., Hoekstra, R. F., Lestrade, B., and Maas, M. F. P. M. (2014). Regular bottlenecks and restrictions to somatic fusion prevent the accumulation of mitochondrial defects in *Neurospora*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130448–20130448.
35. Hansong Ma, P. H. O. (2016). Selfish drive can trump function when animal mitochondrial genomes compete. *Nat. Genet.* 48, 798–802.
36. Bergstrom, C. T., and Pritchard, J. (1998). Germline Bottlenecks and the Evolutionary Maintenance of Mitochondrial Genomes. *Genetics* 149, 2135–2146.
37. Rispe, C., and Moran, N. A. (2000). Accumulation of Deleterious Mutations in Endosymbionts: Muller's Ratchet with Two Levels of Selection. *The American Naturalist* 156, 425–441.
38. Leigh, E. G. (1983). When does the good of the group override the advantage of the individual? *Proc. Natl. Acad. Sci. U.S.A.* 80, 2985–2989.
39. Maynard Smith, J. (1976). Group selection. *Q. Rev. Biol.* 51, 277–283.
40. Adams, K. (2003). Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* 29, 380–395.
41. Smith, D. R., and Keeling, P. J. (2015). Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10177–10184.
42. Burger, G., Gray, M. W., and Franz Lang, B. (2003). Mitochondrial genomes: anything goes. *Trends Genet.* 19, 709–716.
43. Turmel, M., Lemieux, C., Burger, G., Lang, B. F., Otis, C., Plante, I., and Gray, M. W. (1999). The Complete Mitochondrial DNA Sequences of *Nephroselmis olivacea* and *Pedinomonas minor*: Two Radically Different Evolutionary Patterns within Green Algae. *The Plant Cell Online* 11, 1717–1729.
44. Conway, D. J., Fanello, C., Lloyd, J. M., Al-Joubori, B. M. A. S., Baloch, A. H., Somanath, S. D., Roper, C., Oduola, A. M. J., Mulder, B., Povoas, M. M., et al. (2000). Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol. Biochem. Parasitol.* 111, 163–171.
45. Mai, Z., Ghosh, S., Frisardi, M., Rosenthal, B., Rogers, R., and Samuelson, J. (1999). Hsp60 Is Targeted to a Cryptic Mitochondrion-Derived Organelle (“Crypton”) in the Microaerophilic Protozoan Parasite *Entamoeba histolytica*. *Mol. Cell. Biol.* 19, 2198–2205.
46. Hackstein, J. H. P., Akhmanova, A., Boxma, B., Harhangi, H. R., and Voncken, F. G. J. (1999). Hydrogenosomes: eukaryotic adaptations to anaerobic environments. *Trends Microbiol.* 7, 441–447.
47. van der Giezen, M. (2009). Hydrogenosomes and Mitosomes: Conservation and Evolution of Functions. *J. Eukaryot. Microbiol.* 56, 221–231.
48. Embley, M., van der Giezen, M., Horner, D. S., Dyal, P. L., and Foster, P. (2003). Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 191–203.
49. Tovar, J., León-Avila, G., Sánchez, L. B., Sutak, R., Tachezy, J., van der Giezen, M., Hernández, M., Müller, M., and Lucocq, J. M.

- (2003). Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature* 426, 172–176.
50. Shao, R., Kirkness, E. F., and Barker, S. C. (2009). The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. *Genome Res.* 19, 904–912.
51. Shao, R., Zhu, X.-Q., Barker, S. C., and Herd, K. (2012). Evolution of Extensively Fragmented Mitochondrial Genomes in the Lice of Humans. *Genome Biol Evol* 4, 1088–1101.
52. Shao, R., Li, H., Barker, S. C., and Song, S. (2017). The Mitochondrial Genome of the Guanaco Louse, *Microthoracius praelongiceps*: Insights into the Ancestral Mitochondrial Karyotype of Sucking Lice (Anoplura, Insecta). *Genome Biol Evol* 9, 431–445.
53. Vlček, Č., Marande, W., Teijeiro, S., Lukeš, J., and Burger, G. (2011). Systematically fragmented genes in a multipartite mitochondrial genome. *Nucl. Acids Res.* 39, 979–988.
54. Sloan, D. B. (2013). One ring to rule them all? Genome sequencing provides new insights into the “master circle” model of plant mitochondrial DNA structure. *New Phytologist* 200, 978–985.
55. Sloan, D. B., Alverson, A. J., Chuckalovcak, J. P., Wu, M., McCauley, D. E., Palmer, J. D., and Taylor, D. R. (2012). Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. *PLOS Biol* 10, e1001241.
56. Alverson, A. J., Rice, D. W., Dickinson, S., Barry, K., and Palmer, J. D. (2011). Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell* 23, 2499–2513.
57. Rice, D. W., Alverson, A. J., Richardson, A. O., Young, G. J., Sanchez-Puerta, M. V., Munzinger, J., Barry, K., Boore, J. L., Zhang, Y., dePamphilis, C. W., et al. (2013). Horizontal Transfer of Entire Genomes via Mitochondrial Fusion in the Angiosperm *Amborella*. *Science* 342, 1468–1473.
58. Lynch, M., Koskella, B., and Schaack, S. (2006). Mutation Pressure and the Evolution of Organelle Genomic Architecture. *Science* 311, 1727–1730.
59. Mira, A., Ochman, H., and Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet* 17, 589–596.
60. Wu, Z., Cuthbert, J. M., Taylor, D. R., and Sloan, D. B. (2015). The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10185–10191.
61. Palmer, J. D., and Shields, C. R. (1984). Tripartite structure of the *Brassica campestris* mitochondrial genome. *Nature*.
62. Williams, K. S., and Simon, C. (1995). The ecology, behavior, and evolution of periodical cicadas. *Annu Rev Entomol* 40, 269–295.
63. Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
64. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* 19, 455–477.
65. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 2009 10:1 10, 421.
66. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
67. Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucl. Acids Res.* 39, gkr367–W37.
68. Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucl. Acids Res.* 35, 3100–3108.
69. Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucl. Acids Res.* 32, 11–16.
70. Seemann, T. barnmap. Available at: <https://github.com/tseemann/barnmap>.
71. Stamatakis, A. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
72. Oksanen, J., Kindt, R., Legendre, P., and O'Hara, B. (2007). The vegan package. *Community ecology Package* 10, 631–637.
73. R Core Team R: A Language and Environment for Statistical Computing. Available at: <https://www.R-project.org>.
74. Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30, 772–780.

Supplemental Information

Figure S1: Phylogenetic trees showing ancient lineage splitting. Related to Figure 1.

Maximum likelihood phylogenies of selected genes of *Hodgkinia* from *Magicicada* (black) and outgroups - *Aleeta curvicosta* (green) and *Tryella crassa* (red), based on amino acid sequences of all copies from these nine species. Bootstrap support values >90% are shown, nodes with support <70% were collapsed. Trees are rooted on longest branches. Tree A show patterns consistent with the presence of only a single gene copy in the last common ancestor, and independent splits in some of the derived host clades. Trees B-D show patterns consistent with the presence of at least two gene copies in the last common ancestor of *Magicicada* and at least one of the outgroups.

Figure S2: Phylogenetic trees showing numbers of distinct gene copies in the ancestor of *Magicicada*. Related to Figure 1.

(A-F) Maximum likelihood phylogenies of selected genes of *Hodgkinia* from *Magicicada*, based on amino acid sequences of all copies from seven species. Bootstrap support values >90% are shown, nodes with support <70% were collapsed. Gene copies from Cassini, Decula and Decim species groups are indicated with green, blue and pink backgrounds, respectively. Solid maroon boxes indicate monophyletic clades that consist of gene copies from all species groups, and which are likely derived from a gene copy present in the last common ancestor. Dashed maroon boxes with question marks indicate monophyletic clades which may have derived from a gene copy or copies present in the common ancestor, but which do not include sequences from all species groups: it is possible that in the Decim species group, that gene copy was lost. Arrowheads indicate gene split events that took place after a species group diverged. Trees A-B are rooted using outgroups - *Aleeta* and *Tryella*; in other trees, the root has been placed on the longest branch between well-supported clades that included species from all or most hosts. (G) Numbers of distinct copies of selected genes in the last common ancestor of *Magicicada* and in extant species groups, estimated based on trees in panels A-F.

Figure S3: Principal coordinates chart of relative gene abundance. Related to Figure 3.

Charts from the principal coordinates analysis of relative gene abundance from Figure 3. A) All genes present in any species. B) Only genes present in all species. Species abbreviations are taken from Table 1.

Figure S2.

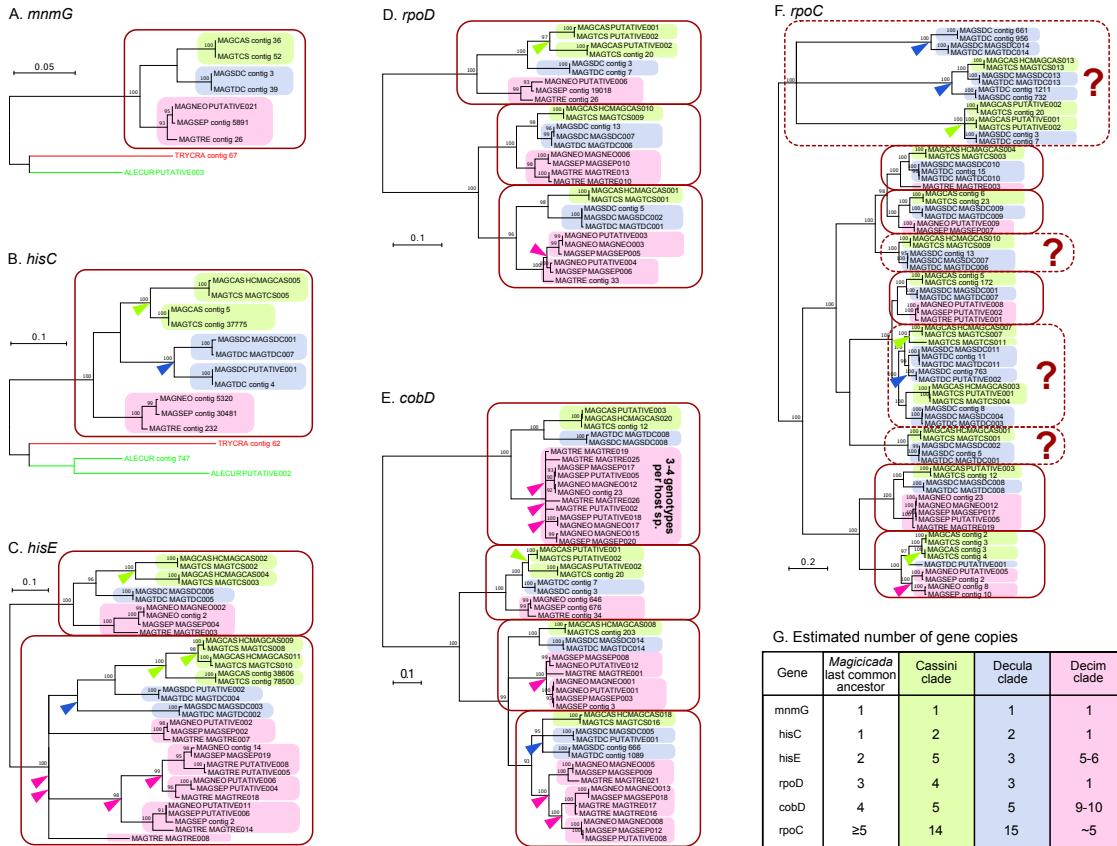
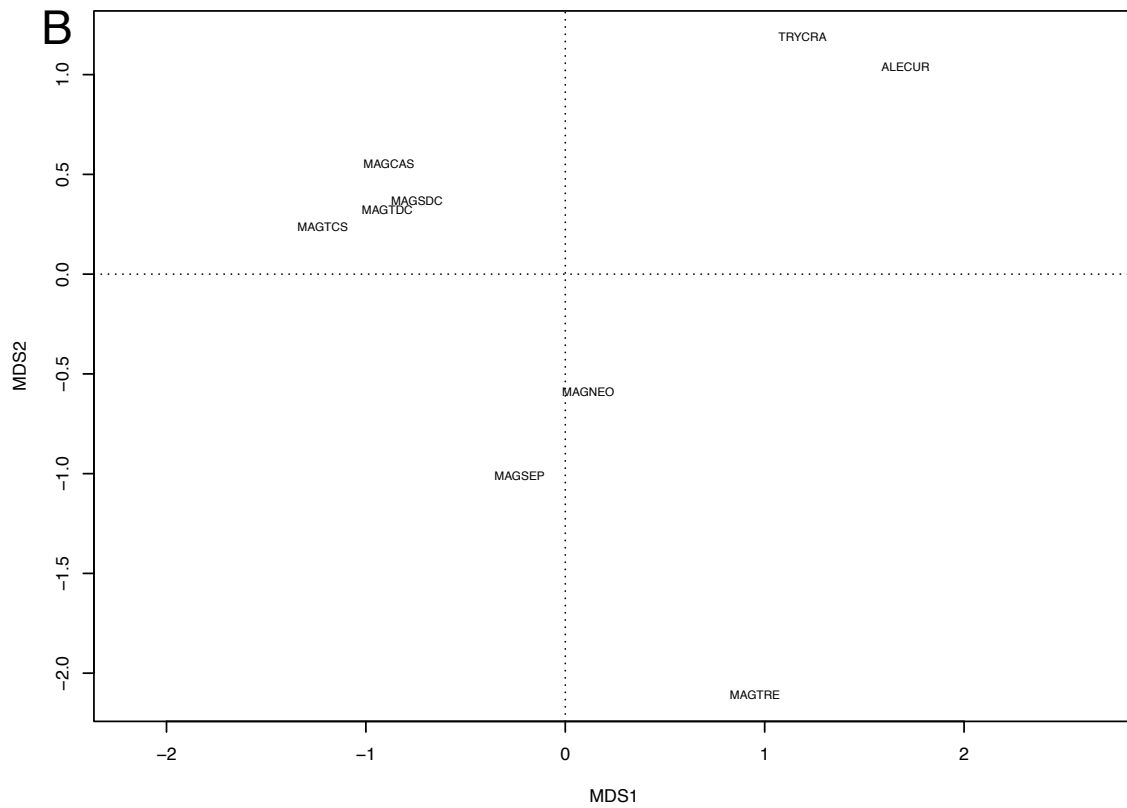
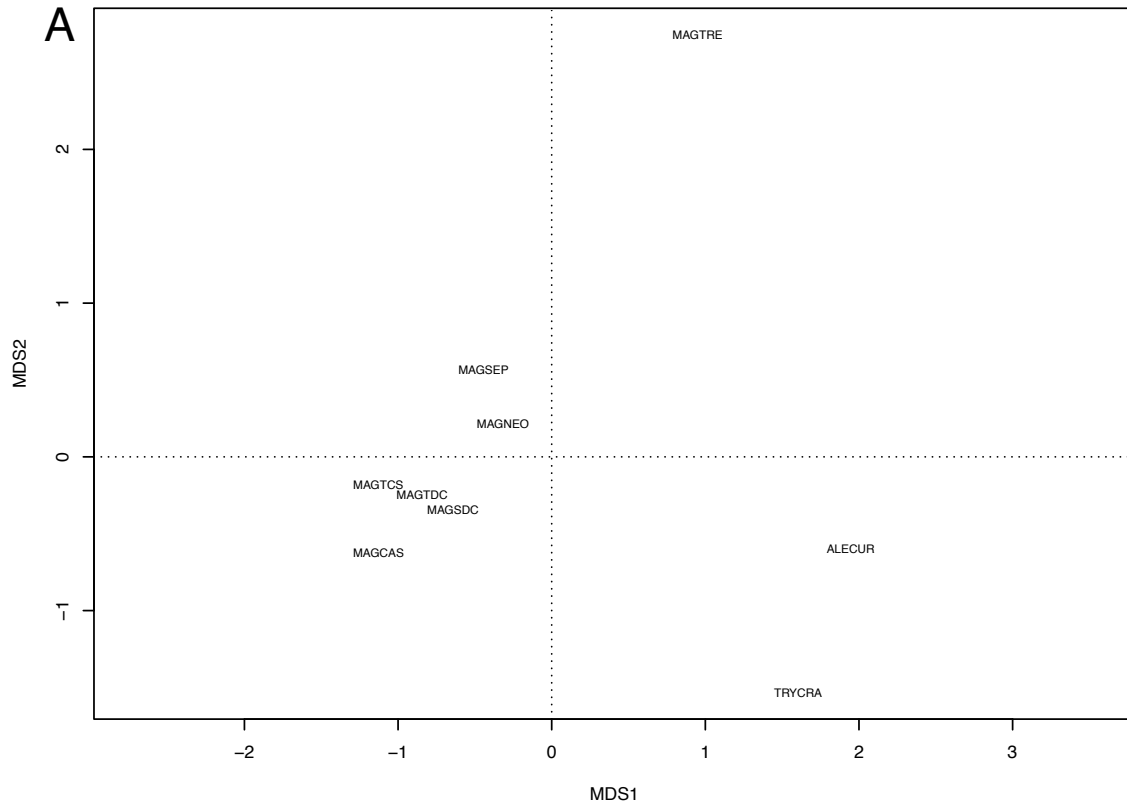


Figure S3.



| Species | Scaffold | Scaffold length | Number of genes | Coding density (%) | GC content (%) | Average coverage |
|------------------------|-------------|-----------------|-----------------|--------------------|----------------|------------------|
| <i>M. cassini</i> | MAGCAS001 | 69473 | 17 | 31 | 34 | 2535 |
| <i>M. cassini</i> | MAGCAS002 | 61653 | 13 | 20.7 | 29.9 | 1625 |
| <i>M. cassini</i> | MAGCAS003 | 52116 | 6 | 21.5 | 27.1 | 1263 |
| <i>M. cassini</i> | MAGCAS004 | 48995 | 18 | 43.6 | 31.9 | 2219 |
| <i>M. cassini</i> | MAGCAS005 | 45702 | 9 | 25.2 | 31.4 | 515 |
| <i>M. cassini</i> | MAGCAS006 | 41268 | 14 | 39.2 | 32.1 | 580 |
| <i>M. cassini</i> | MAGCAS007 | 35613 | 8 | 48.5 | 29.9 | 216 |
| <i>M. cassini</i> | MAGCAS008 | 33827 | 5 | 21.9 | 38.6 | 682 |
| <i>M. cassini</i> | MAGCAS009 | 32518 | 4 | 17.6 | 35.3 | 158 |
| <i>M. cassini</i> | MAGCAS010 | 31635 | 4 | 31.7 | 33.6 | 509 |
| <i>M. cassini</i> | MAGCAS011 | 29608 | 4 | 18 | 35.2 | 673 |
| <i>M. cassini</i> | MAGCAS012 | 18833 | 4 | 29.9 | 38.1 | 502 |
| <i>M. cassini</i> | MAGCAS013 | 18274 | 4 | 52.1 | 36.4 | 944 |
| <i>M. cassini</i> | MAGCAS014 | 18203 | 6 | 45.2 | 36.4 | 2231 |
| <i>M. cassini</i> | MAGCAS015 | 17469 | 3 | 12.9 | 36.5 | 1141 |
| <i>M. cassini</i> | MAGCAS016 | 16309 | 4 | 15.3 | 28.3 | 182 |
| <i>M. cassini</i> | MAGCAS017 | 15578 | 3 | 18.9 | 36.1 | 666 |
| <i>M. cassini</i> | MAGCAS018 | 13935 | 4 | 40.8 | 34.6 | 1032 |
| <i>M. cassini</i> | MAGCAS019 | 9976 | 4 | 59.9 | 45.2 | 689 |
| <i>M. cassini</i> | MAGCAS020 | 8941 | 1 | 10.7 | 23.1 | 553 |
| <i>M. cassini</i> | MAGCAS021 | 8091 | 4 | 62.2 | 45.3 | 1119 |
| <i>M. cassini</i> | MAGCAS022 | 5678 | 3 | 41.4 | 34 | 544 |
| <i>M. cassini</i> | MAGCAS023 | 1057 | 1 | 60.1 | 29.7 | 2642 |
| <i>M. cassini</i> | MAGCAS024 | 748 | 1 | 62.4 | 32.5 | 101 |
| <i>M. cassini</i> | PUTATIVE001 | 46023 | 11 | 34.3 | 29.7 | 870 |
| <i>M. cassini</i> | PUTATIVE002 | 28503 | 6 | 39.4 | 31.3 | 305 |
| <i>M. cassini</i> | PUTATIVE003 | 27789 | 6 | 37.5 | 25.4 | 836 |
| <i>M. cassini</i> | PUTATIVE004 | 21863 | 2 | 9.3 | 31 | 2590 |
| <i>M. cassini</i> | PUTATIVE006 | 14618 | 1 | 10.2 | 31.5 | 204 |
| <i>M. tredecassini</i> | MAGTCS001 | 69472 | 17 | 30.7 | 34 | 1450 |
| <i>M. tredecassini</i> | MAGTCS002 | 61641 | 13 | 20.8 | 29.9 | 1646 |
| <i>M. tredecassini</i> | MAGTCS003 | 56215 | 21 | 47.9 | 33.1 | 2524 |
| <i>M. tredecassini</i> | MAGTCS004 | 52125 | 6 | 21.6 | 27 | 753 |
| <i>M. tredecassini</i> | MAGTCS005 | 45699 | 10 | 28 | 31.5 | 340 |
| <i>M. tredecassini</i> | MAGTCS006 | 41278 | 14 | 39.1 | 32.2 | 80 |
| <i>M. tredecassini</i> | MAGTCS007 | 35614 | 8 | 48.5 | 29.9 | 129 |
| <i>M. tredecassini</i> | MAGTCS008 | 32520 | 3 | 15.6 | 35.4 | 502 |
| <i>M. tredecassini</i> | MAGTCS009 | 31636 | 4 | 31.7 | 33.6 | 237 |

| | | | | | | |
|------------------------|-------------|-------|----|------|------|------|
| <i>M. tredecassini</i> | MAGTCS010 | 29317 | 4 | 18.2 | 35 | 339 |
| <i>M. tredecassini</i> | MAGTCS011 | 22416 | 17 | 73.3 | 31.1 | 4356 |
| <i>M. tredecassini</i> | MAGTCS012 | 18834 | 4 | 29.9 | 38.2 | 190 |
| <i>M. tredecassini</i> | MAGTCS013 | 18271 | 4 | 52.1 | 36.4 | 500 |
| <i>M. tredecassini</i> | MAGTCS014 | 16306 | 3 | 10.7 | 28.4 | 183 |
| <i>M. tredecassini</i> | MAGTCS015 | 15579 | 3 | 18.9 | 36 | 549 |
| <i>M. tredecassini</i> | MAGTCS016 | 13934 | 4 | 40.8 | 34.5 | 996 |
| <i>M. tredecassini</i> | MAGTCS017 | 5680 | 3 | 40.7 | 33.9 | 172 |
| <i>M. tredecassini</i> | MAGTCS018 | 1057 | 1 | 68.6 | 29.8 | 1424 |
| <i>M. tredecassini</i> | MAGTCS019 | 748 | 1 | 63.2 | 32.5 | 83 |
| <i>M. tredecassini</i> | MAGTCS020 | 694 | 1 | 94.1 | 25.4 | 54 |
| <i>M. tredecassini</i> | PUTATIVE001 | 58824 | 6 | 19.1 | 27.1 | 738 |
| <i>M. tredecassini</i> | PUTATIVE002 | 46026 | 10 | 33 | 29.7 | 806 |
| <i>M. tredecassini</i> | PUTATIVE003 | 21389 | 2 | 9.5 | 30.9 | 1839 |
| <i>M. tredecassini</i> | PUTATIVE005 | 15824 | 3 | 18.6 | 35.9 | 552 |
| <i>M. tredecassini</i> | PUTATIVE006 | 15349 | 3 | 17.9 | 27.7 | 2623 |
| <i>M. tredecassini</i> | PUTATIVE007 | 11443 | 3 | 45 | 38.5 | 1237 |
| <i>M. septendecula</i> | MAGSDC001 | 70504 | 27 | 36.7 | 28.6 | 2344 |
| <i>M. septendecula</i> | MAGSDC002 | 70413 | 16 | 28.5 | 31.9 | 290 |
| <i>M. septendecula</i> | MAGSDC003 | 53932 | 6 | 14.8 | 25.9 | 664 |
| <i>M. septendecula</i> | MAGSDC004 | 51941 | 6 | 21.4 | 24 | 311 |
| <i>M. septendecula</i> | MAGSDC005 | 41633 | 5 | 11.5 | 28.6 | 1476 |
| <i>M. septendecula</i> | MAGSDC006 | 34671 | 17 | 43.4 | 32.8 | 2072 |
| <i>M. septendecula</i> | MAGSDC007 | 31673 | 4 | 31.7 | 32.4 | 196 |
| <i>M. septendecula</i> | MAGSDC008 | 29204 | 7 | 29.4 | 27.7 | 1293 |
| <i>M. septendecula</i> | MAGSDC009 | 29090 | 7 | 51 | 34 | 448 |
| <i>M. septendecula</i> | MAGSDC010 | 29000 | 7 | 50.8 | 31.9 | 1642 |
| <i>M. septendecula</i> | MAGSDC011 | 22533 | 9 | 48.5 | 30.1 | 1820 |
| <i>M. septendecula</i> | MAGSDC012 | 18788 | 7 | 46 | 37.6 | 495 |
| <i>M. septendecula</i> | MAGSDC013 | 17780 | 4 | 51.2 | 34.4 | 174 |
| <i>M. septendecula</i> | MAGSDC014 | 13841 | 3 | 60.8 | 30.6 | 388 |
| <i>M. septendecula</i> | MAGSDC015 | 10837 | 4 | 53.8 | 46.7 | 1520 |
| <i>M. septendecula</i> | MAGSDC016 | 8802 | 2 | 47.8 | 43 | 493 |
| <i>M. septendecula</i> | MAGSDC017 | 7860 | 2 | 53.7 | 47.3 | 331 |
| <i>M. septendecula</i> | MAGSDC018 | 1107 | 0 | 0 | 27.4 | 1637 |
| <i>M. septendecula</i> | MAGSDC019 | 748 | 1 | 63.2 | 29.4 | 72 |
| <i>M. septendecula</i> | MAGSDC020 | 692 | 1 | 93.9 | 26.3 | 17 |
| <i>M. septendecula</i> | PUTATIVE001 | 66259 | 20 | 26.9 | 28.7 | 456 |
| <i>M. septendecula</i> | PUTATIVE002 | 31705 | 6 | 12.8 | 31.2 | 1220 |
| <i>M. septendecula</i> | PUTATIVE003 | 20842 | 4 | 10.3 | 28.2 | 532 |

| | | | | | | |
|------------------------|-------------|-------|----|------|------|------|
| <i>M. septendecula</i> | PUTATIVE005 | 14087 | 8 | 63.8 | 24.7 | 1702 |
| <i>M. septendecula</i> | PUTATIVE006 | 11820 | 9 | 36.6 | 29.1 | 4455 |
| <i>M. septendecula</i> | PUTATIVE008 | 10372 | 3 | 11.3 | 31.4 | 914 |
| <i>M. septendecula</i> | PUTATIVE009 | 10281 | 3 | 28.6 | 37.5 | 1636 |
| <i>M. tredecula</i> | MAGTDC001 | 70424 | 15 | 29.3 | 31.9 | 465 |
| <i>M. tredecula</i> | MAGTDC002 | 54074 | 6 | 20 | 25.9 | 182 |
| <i>M. tredecula</i> | MAGTDC003 | 51873 | 6 | 21.4 | 24 | 154 |
| <i>M. tredecula</i> | MAGTDC004 | 47884 | 11 | 20.1 | 32.7 | 735 |
| <i>M. tredecula</i> | MAGTDC005 | 34671 | 17 | 43.4 | 32.8 | 1702 |
| <i>M. tredecula</i> | MAGTDC006 | 31664 | 4 | 31.7 | 32.4 | 479 |
| <i>M. tredecula</i> | MAGTDC007 | 29603 | 12 | 44.9 | 28.7 | 2787 |
| <i>M. tredecula</i> | MAGTDC008 | 29201 | 7 | 29.3 | 27.8 | 500 |
| <i>M. tredecula</i> | MAGTDC009 | 29085 | 7 | 51.1 | 34 | 480 |
| <i>M. tredecula</i> | MAGTDC010 | 29002 | 7 | 50.8 | 31.9 | 574 |
| <i>M. tredecula</i> | MAGTDC011 | 22533 | 10 | 57.5 | 30.1 | 571 |
| <i>M. tredecula</i> | MAGTDC012 | 18787 | 7 | 46.1 | 37.6 | 570 |
| <i>M. tredecula</i> | MAGTDC013 | 17780 | 4 | 51.2 | 34.4 | 131 |
| <i>M. tredecula</i> | MAGTDC014 | 13840 | 3 | 61.2 | 30.6 | 251 |
| <i>M. tredecula</i> | MAGTDC015 | 10839 | 4 | 53.8 | 46.8 | 1096 |
| <i>M. tredecula</i> | MAGTDC016 | 8803 | 2 | 47.8 | 43 | 640 |
| <i>M. tredecula</i> | MAGTDC017 | 7866 | 2 | 53.6 | 47.3 | 633 |
| <i>M. tredecula</i> | MAGTDC018 | 1108 | 1 | 65.4 | 27.8 | 514 |
| <i>M. tredecula</i> | MAGTDC019 | 748 | 1 | 58.8 | 29.3 | 27 |
| <i>M. tredecula</i> | MAGTDC020 | 692 | 1 | 93.9 | 26.3 | 16 |
| <i>M. tredecula</i> | PUTATIVE001 | 63118 | 12 | 24.1 | 30.4 | 856 |
| <i>M. tredecula</i> | PUTATIVE002 | 35545 | 17 | 63.7 | 30.7 | 2467 |
| <i>M. tredecula</i> | PUTATIVE003 | 20858 | 4 | 10.3 | 28.1 | 335 |
| <i>M. tredecula</i> | PUTATIVE004 | 16213 | 2 | 10.2 | 30.3 | 780 |
| <i>M. tredecula</i> | PUTATIVE005 | 13731 | 3 | 33.7 | 39.1 | 459 |
| <i>M. tredecula</i> | PUTATIVE006 | 13664 | 8 | 65.8 | 24.5 | 734 |
| <i>M. tredecula</i> | PUTATIVE007 | 10286 | 3 | 28.6 | 37.5 | 809 |
| <i>M. neotredecim</i> | MAGNEO001 | 61256 | 7 | 9 | 24.9 | 134 |
| <i>M. neotredecim</i> | MAGNEO002 | 57172 | 21 | 38.7 | 29.9 | 257 |
| <i>M. neotredecim</i> | MAGNEO003 | 56251 | 12 | 21.2 | 29.6 | 180 |
| <i>M. neotredecim</i> | MAGNEO004 | 35088 | 11 | 27.4 | 23.7 | 249 |
| <i>M. neotredecim</i> | MAGNEO005 | 34472 | 4 | 16.5 | 31 | 366 |
| <i>M. neotredecim</i> | MAGNEO006 | 34053 | 3 | 7.4 | 29.9 | 187 |
| <i>M. neotredecim</i> | MAGNEO007 | 27670 | 4 | 9.1 | 28.1 | 120 |
| <i>M. neotredecim</i> | MAGNEO008 | 23724 | 5 | 23.2 | 32.3 | 155 |
| <i>M. neotredecim</i> | MAGNEO009 | 22209 | 4 | 32.2 | 35.8 | 161 |

| | | | | | | |
|-----------------------|-------------|-------|----|------|------|------|
| <i>M. neutredecim</i> | MAGNEO010 | 21502 | 2 | 9 | 28.4 | 103 |
| <i>M. neutredecim</i> | MAGNEO011 | 19133 | 2 | 8 | 32.5 | 148 |
| <i>M. neutredecim</i> | MAGNEO012 | 18932 | 4 | 48.6 | 24.4 | 147 |
| <i>M. neutredecim</i> | MAGNEO013 | 17725 | 1 | 5.3 | 27.2 | 173 |
| <i>M. neutredecim</i> | MAGNEO014 | 12466 | 4 | 21.4 | 26.6 | 164 |
| <i>M. neutredecim</i> | MAGNEO015 | 11399 | 2 | 12.5 | 21.9 | 81 |
| <i>M. neutredecim</i> | MAGNEO016 | 11218 | 1 | 3.8 | 24.7 | 64 |
| <i>M. neutredecim</i> | MAGNEO017 | 10677 | 2 | 13.4 | 22 | 530 |
| <i>M. neutredecim</i> | MAGNEO018 | 3222 | 2 | 57.4 | 25.2 | 152 |
| <i>M. neutredecim</i> | MAGNEO019 | 1100 | 1 | 66.2 | 26.7 | 715 |
| <i>M. neutredecim</i> | MAGNEO020 | 748 | 1 | 60 | 30.9 | 81 |
| <i>M. neutredecim</i> | PUTATIVE001 | 64450 | 7 | 8.6 | 24.9 | 126 |
| <i>M. neutredecim</i> | PUTATIVE002 | 59782 | 6 | 12.6 | 23.3 | 53 |
| <i>M. neutredecim</i> | PUTATIVE003 | 55986 | 12 | 20.9 | 29.5 | 183 |
| <i>M. neutredecim</i> | PUTATIVE004 | 54126 | 8 | 12.8 | 29.1 | 155 |
| <i>M. neutredecim</i> | PUTATIVE005 | 53050 | 9 | 25.9 | 29.6 | 347 |
| <i>M. neutredecim</i> | PUTATIVE006 | 41569 | 7 | 12.8 | 27.5 | 184 |
| <i>M. neutredecim</i> | PUTATIVE007 | 30380 | 1 | 4.3 | 24 | 55 |
| <i>M. neutredecim</i> | PUTATIVE008 | 30105 | 13 | 41.7 | 25.2 | 363 |
| <i>M. neutredecim</i> | PUTATIVE009 | 28933 | 6 | 38.1 | 31.2 | 306 |
| <i>M. neutredecim</i> | PUTATIVE010 | 27031 | 10 | 30.3 | 25.9 | 402 |
| <i>M. neutredecim</i> | PUTATIVE011 | 26196 | 3 | 8 | 27.6 | 270 |
| <i>M. neutredecim</i> | PUTATIVE012 | 21868 | 1 | 4.2 | 26.4 | 5 |
| <i>M. neutredecim</i> | PUTATIVE013 | 20101 | 4 | 21.2 | 34 | 256 |
| <i>M. neutredecim</i> | PUTATIVE014 | 18183 | 5 | 29.6 | 27.6 | 257 |
| <i>M. neutredecim</i> | PUTATIVE015 | 17100 | 6 | 32.7 | 27.5 | 748 |
| <i>M. neutredecim</i> | PUTATIVE016 | 15929 | 1 | 5.7 | 27.9 | 289 |
| <i>M. neutredecim</i> | PUTATIVE018 | 14006 | 3 | 30 | 29.5 | 211 |
| <i>M. neutredecim</i> | PUTATIVE019 | 12258 | 3 | 43.6 | 32.8 | 449 |
| <i>M. neutredecim</i> | PUTATIVE020 | 10927 | 4 | 39.4 | 30.8 | 688 |
| <i>M. neutredecim</i> | PUTATIVE021 | 10855 | 4 | 41.6 | 32.4 | 270 |
| <i>M. neutredecim</i> | PUTATIVE022 | 10248 | 2 | 8.4 | 25.9 | 166 |
| <i>M. septendecim</i> | MAGSEP001 | 63626 | 22 | 27 | 27.2 | 724 |
| <i>M. septendecim</i> | MAGSEP002 | 61848 | 7 | 13.7 | 23.5 | 425 |
| <i>M. septendecim</i> | MAGSEP003 | 61285 | 7 | 8.8 | 24.9 | 534 |
| <i>M. septendecim</i> | MAGSEP004 | 57162 | 20 | 37.6 | 29.9 | 2101 |
| <i>M. septendecim</i> | MAGSEP005 | 56298 | 12 | 21.2 | 29.5 | 1285 |
| <i>M. septendecim</i> | MAGSEP006 | 56278 | 9 | 17.5 | 29.4 | 176 |
| <i>M. septendecim</i> | MAGSEP007 | 48066 | 12 | 37.1 | 30.2 | 863 |
| <i>M. septendecim</i> | MAGSEP008 | 43440 | 3 | 7.2 | 25.4 | 73 |

| | | | | | | |
|-----------------------|-------------|-------|----|------|------|------|
| <i>M. septendecim</i> | MAGSEP009 | 34489 | 4 | 16.5 | 30.9 | 781 |
| <i>M. septendecim</i> | MAGSEP010 | 33034 | 3 | 7.7 | 30.1 | 425 |
| <i>M. septendecim</i> | MAGSEP011 | 28244 | 4 | 8.9 | 28.3 | 487 |
| <i>M. septendecim</i> | MAGSEP012 | 22960 | 5 | 24.3 | 32 | 208 |
| <i>M. septendecim</i> | MAGSEP013 | 22204 | 4 | 25 | 35.8 | 348 |
| <i>M. septendecim</i> | MAGSEP014 | 21903 | 3 | 12.5 | 28.6 | 835 |
| <i>M. septendecim</i> | MAGSEP015 | 21496 | 2 | 9 | 28.4 | 440 |
| <i>M. septendecim</i> | MAGSEP016 | 19133 | 2 | 8 | 32.4 | 351 |
| <i>M. septendecim</i> | MAGSEP017 | 18932 | 4 | 48.6 | 24.3 | 350 |
| <i>M. septendecim</i> | MAGSEP018 | 17726 | 1 | 5.3 | 27.2 | 231 |
| <i>M. septendecim</i> | MAGSEP019 | 15979 | 3 | 30.4 | 35.8 | 697 |
| <i>M. septendecim</i> | MAGSEP020 | 11397 | 2 | 12.5 | 22 | 223 |
| <i>M. septendecim</i> | MAGSEP021 | 11166 | 2 | 38.1 | 42.4 | 385 |
| <i>M. septendecim</i> | MAGSEP022 | 3224 | 2 | 55.3 | 25.1 | 54 |
| <i>M. septendecim</i> | MAGSEP023 | 1099 | 1 | 66.2 | 26.8 | 753 |
| <i>M. septendecim</i> | PUTATIVE001 | 49706 | 4 | 11.8 | 25.7 | 360 |
| <i>M. septendecim</i> | PUTATIVE002 | 41910 | 19 | 46.4 | 26.2 | 1773 |
| <i>M. septendecim</i> | PUTATIVE003 | 33062 | 11 | 30.7 | 27.5 | 1566 |
| <i>M. septendecim</i> | PUTATIVE004 | 30610 | 5 | 10.2 | 26.1 | 780 |
| <i>M. septendecim</i> | PUTATIVE005 | 30280 | 6 | 35.3 | 23.4 | 289 |
| <i>M. septendecim</i> | PUTATIVE006 | 26244 | 3 | 7.9 | 27.5 | 978 |
| <i>M. septendecim</i> | PUTATIVE007 | 23391 | 5 | 18.7 | 28 | 2022 |
| <i>M. septendecim</i> | PUTATIVE008 | 23205 | 5 | 23.9 | 32 | 215 |
| <i>M. septendecim</i> | PUTATIVE009 | 20105 | 4 | 21.2 | 34 | 856 |
| <i>M. septendecim</i> | PUTATIVE010 | 18957 | 4 | 9.9 | 22.9 | 684 |
| <i>M. septendecim</i> | PUTATIVE011 | 18955 | 3 | 7.3 | 22.9 | 216 |
| <i>M. septendecim</i> | PUTATIVE013 | 15938 | 3 | 6.3 | 24.9 | 1209 |
| <i>M. septendecim</i> | PUTATIVE014 | 14676 | 3 | 10.7 | 24.5 | 351 |
| <i>M. septendecim</i> | PUTATIVE015 | 13624 | 6 | 20.6 | 26.7 | 836 |
| <i>M. septendecim</i> | PUTATIVE017 | 10118 | 3 | 36.1 | 32.2 | 736 |
| <i>M. septendecim</i> | PUTATIVE018 | 10107 | 2 | 14.1 | 21.8 | 857 |
| <i>M. tredecim</i> | MAGTRE001 | 61247 | 7 | 9 | 24.7 | 384 |
| <i>M. tredecim</i> | MAGTRE002 | 59026 | 9 | 16.2 | 30.4 | 504 |
| <i>M. tredecim</i> | MAGTRE003 | 58092 | 21 | 45.1 | 29.8 | 1392 |
| <i>M. tredecim</i> | MAGTRE004 | 55954 | 19 | 29.2 | 27.8 | 1419 |
| <i>M. tredecim</i> | MAGTRE005 | 52863 | 12 | 22.2 | 26.3 | 925 |
| <i>M. tredecim</i> | MAGTRE006 | 46330 | 19 | 32 | 27.5 | 1043 |
| <i>M. tredecim</i> | MAGTRE007 | 43811 | 5 | 8.3 | 22.9 | 233 |
| <i>M. tredecim</i> | MAGTRE008 | 41390 | 10 | 21.4 | 32.3 | 503 |
| <i>M. tredecim</i> | MAGTRE009 | 40594 | 2 | 5 | 28.3 | 116 |

| | | | | | | |
|--------------------|-------------|-------|----|------|------|------|
| <i>M. tredecim</i> | MAGTRE010 | 33207 | 3 | 7.6 | 30 | 155 |
| <i>M. tredecim</i> | MAGTRE011 | 32598 | 4 | 19.1 | 32.6 | 291 |
| <i>M. tredecim</i> | MAGTRE012 | 27904 | 6 | 21.7 | 35.7 | 401 |
| <i>M. tredecim</i> | MAGTRE013 | 25773 | 3 | 9.7 | 29.4 | 184 |
| <i>M. tredecim</i> | MAGTRE014 | 25212 | 4 | 14.3 | 27.4 | 421 |
| <i>M. tredecim</i> | MAGTRE015 | 24039 | 9 | 44.5 | 30.6 | 425 |
| <i>M. tredecim</i> | MAGTRE016 | 23178 | 1 | 4 | 25.6 | 57 |
| <i>M. tredecim</i> | MAGTRE017 | 23156 | 1 | 4 | 25.7 | 19 |
| <i>M. tredecim</i> | MAGTRE018 | 19535 | 3 | 8.6 | 31.7 | 262 |
| <i>M. tredecim</i> | MAGTRE019 | 19044 | 3 | 45.1 | 24.2 | 144 |
| <i>M. tredecim</i> | MAGTRE020 | 18064 | 2 | 25.5 | 24 | 256 |
| <i>M. tredecim</i> | MAGTRE021 | 13768 | 2 | 12.9 | 27.1 | 184 |
| <i>M. tredecim</i> | MAGTRE022 | 12894 | 1 | 7.2 | 32.2 | 157 |
| <i>M. tredecim</i> | MAGTRE023 | 12893 | 2 | 11.8 | 32.4 | 115 |
| <i>M. tredecim</i> | MAGTRE024 | 12600 | 3 | 16.1 | 25.7 | 1136 |
| <i>M. tredecim</i> | MAGTRE025 | 11445 | 2 | 12.4 | 21.6 | 175 |
| <i>M. tredecim</i> | MAGTRE026 | 11402 | 1 | 7.6 | 21.5 | 229 |
| <i>M. tredecim</i> | MAGTRE027 | 11191 | 2 | 16 | 33.7 | 223 |
| <i>M. tredecim</i> | MAGTRE028 | 11158 | 2 | 38.1 | 42.6 | 344 |
| <i>M. tredecim</i> | MAGTRE029 | 10693 | 5 | 54.2 | 40.2 | 99 |
| <i>M. tredecim</i> | MAGTRE030 | 9940 | 1 | 7.4 | 30.3 | 84 |
| <i>M. tredecim</i> | PUTATIVE001 | 43403 | 20 | 46 | 26 | 1480 |
| <i>M. tredecim</i> | PUTATIVE002 | 43134 | 5 | 7.5 | 28.5 | 184 |
| <i>M. tredecim</i> | PUTATIVE003 | 39854 | 3 | 4.9 | 31.3 | 150 |
| <i>M. tredecim</i> | PUTATIVE004 | 31153 | 3 | 11.8 | 28.3 | 220 |
| <i>M. tredecim</i> | PUTATIVE005 | 22075 | 4 | 22.2 | 24.4 | 91 |
| <i>M. tredecim</i> | PUTATIVE006 | 19049 | 2 | 21.7 | 27.9 | 236 |
| <i>M. tredecim</i> | PUTATIVE008 | 15164 | 4 | 33.6 | 35.7 | 84 |
| <i>M. tredecim</i> | PUTATIVE009 | 13153 | 0 | 0 | 26.5 | 192 |
| <i>M. tredecim</i> | PUTATIVE010 | 11299 | 8 | 54.4 | 24.5 | 4373 |
| <i>M. tredecim</i> | PUTATIVE011 | 11004 | 3 | 16.5 | 34.5 | 237 |
| <i>M. tredecim</i> | PUTATIVE012 | 10250 | 3 | 27.6 | 39.5 | 151 |
| <i>M. tredecim</i> | PUTATIVE013 | 8821 | 2 | 9.5 | 25.4 | 1121 |
| <i>T. crassa</i> | TRYCRA001 | 64739 | 8 | 8.3 | 20.9 | 140 |
| <i>T. crassa</i> | TRYCRA002 | 38749 | 4 | 11.4 | 21.1 | 68 |
| <i>T. crassa</i> | TRYCRA003 | 25142 | 3 | 7.6 | 24.2 | 76 |
| <i>T. crassa</i> | TRYCRA004 | 23146 | 4 | 7.1 | 20.4 | 51 |
| <i>T. crassa</i> | TRYCRA005 | 23137 | 1 | 1.8 | 20.3 | 121 |
| <i>T. crassa</i> | TRYCRA006 | 14624 | 2 | 10.1 | 20.6 | 58 |
| <i>T. crassa</i> | TRYCRA007 | 14492 | 2 | 11.8 | 15.9 | 89 |

| | | | | | | |
|----------------------|-------------|-------|----|------|------|-----|
| <i>T. crassa</i> | TRYCRA008 | 12260 | 1 | 2.9 | 28.8 | 38 |
| <i>T. crassa</i> | TRYCRA009 | 11775 | 1 | 10.9 | 19.8 | 14 |
| <i>T. crassa</i> | TRYCRA010 | 11711 | 3 | 38.9 | 44.1 | 264 |
| <i>T. crassa</i> | TRYCRA011 | 9758 | 1 | 9.6 | 14.4 | 39 |
| <i>T. crassa</i> | TRYCRA012 | 9446 | 1 | 9 | 13.8 | 43 |
| <i>T. crassa</i> | TRYCRA013 | 8867 | 2 | 10.6 | 14.4 | 93 |
| <i>T. crassa</i> | TRYCRA014 | 2151 | 2 | 88.3 | 22.5 | 20 |
| <i>A. curvicosta</i> | ALECUR001 | 46134 | 2 | 2.6 | 24 | 15 |
| <i>A. curvicosta</i> | ALECUR002 | 32844 | 3 | 12.2 | 32.6 | 16 |
| <i>A. curvicosta</i> | ALECUR003 | 23214 | 5 | 17.1 | 28.1 | 41 |
| <i>A. curvicosta</i> | ALECUR004 | 20530 | 3 | 19.6 | 33.1 | 17 |
| <i>A. curvicosta</i> | ALECUR005 | 7857 | 1 | 24.3 | 35.9 | 7 |
| <i>A. curvicosta</i> | PUTATIVE001 | 59192 | 5 | 6.1 | 28.1 | 41 |
| <i>A. curvicosta</i> | PUTATIVE002 | 55340 | 14 | 21.7 | 24.9 | 246 |
| <i>A. curvicosta</i> | PUTATIVE003 | 39946 | 9 | 18.4 | 30.1 | 18 |
| <i>A. curvicosta</i> | PUTATIVE004 | 34099 | 6 | 19.9 | 28.3 | 36 |
| <i>A. curvicosta</i> | PUTATIVE005 | 17753 | 2 | 7.1 | 35.5 | 43 |
| <i>A. curvicosta</i> | PUTATIVE006 | 16143 | 10 | 75.1 | 32.7 | 648 |