Supplemental Material

The Evolution of Dark Matter in the Mitogenome of Seed Beetles

Ahmed Sayadi, Elina Immonen, Christian Tellgren-Roth, Göran Arnqvist

LIST OF CONTENT

1. SUPPLEMENTAL MATERIALS AND METHODS

We used PacBio sequencing technology to sequence and assemble the mitogenomes of *Callosobruchus maculatus*, *C. chinensis*, *C. analis* and *Acanthoscelides obtectus*. As this new technology delivered reads ranging from 7 kbp to >25 kbp, many reads covered much (or even all) of the mitogenome making the assembly of even repeat-rich mitogenomes such as these very straight-forward. Illumina short-read sequencing were then used for (i) DNA sequencing and guided assembly of the mitogenomes of an additional three populations of *C. maculatus* and (ii) RNA sequencing of replicated samples from our reference population of *C. maculatus* to study variation in mtDNA transcription.

PacBio sequencing, assembly and annotation

The *de novo* assemblies of the mitogenomes of the four seed beetle species are part of an ongoing comparative genomics study in this group. For *C. maculatus*, we used a standard reference line from the "South India" population, for *C. chinensis* and *C. analis* we used lines ("Leicester") provided by Robert H. Smith, University of Leicester, and for *A. obtectus* we used the "Belgrade" line provided by Biljana Stojković, University of Belgrade. The four lines had all been subjected to at least six consecutive generations of inbreeding though single-pair full-sib matings, to remove mtDNA variation and reduce nuclear heterozygosity, prior to extractions. DNA was extracted using the Genomic-tip 20/G kit, following the manufacturer's protocol (Qiagen). Ten individual live males were used for each extraction/sample.

For the *Callosobruchus* samples, genomic DNA was sheared into 10 kbp fragments using a Genemachines HydroShear Instrument (Digilab, Marlborough, MA, USA). SMRTbells were constructed according to the manufacturer's instructions (Pacific Biosciences, Menlo Park, CA, USA). SMRTbells for the three *Callosobruchus* species were sequenced on a Pacific Biosciences RSII sequencer according to the manufacturer's instructions with 4 hours movie-time. In total, we sequenced 79 SMRT-cells for *C. maculatus*, 56 SMRT-cells for *C. analis* and 53 SMRT-cells for *C. chinensis*. Reads >3kbp from a single SMRT-cell for each species were selected and were sufficient to assemble the mitogenomes, using the SMRT-analysis HGAP3 assembly pipeline.

Average coverage for mtDNA reads in the single SMRT-cell used for the assembly was 45x for *C. maculatus*, 80x for *C. analis* and 67x for *C. chinensis*.

The *Acanthoscelides obtectus* sample was instead sheared into 25 kbp fragments, using the Megaruptor system (Diagenode, Seraing, Belgium). SMRTbells were constructed according to the manufacturer's instructions and were sequenced on a PacBio Sequel instrument, according to the manufacturer's instructions with 10 hours movie-time. We sequenced a total of 21 SMRT-cells. Here, reads >23,673bp from a single SMRT-cell were used to assemble the mitogenome, using the SMRT-analysis HGAP4 assembly pipeline. Coverage was 11x. In effect, thus, this mitogenome was assembled from 11 congruent reads that basically spanned the entire circular mitogenome. We subsequently mapped all reads longer >15,000bp from the same cell to the mitogenome, which yielded an even coverage across the entire circular mitogenome at an average of 48x.

The assembled genomes were annotated using DOGMA (Wyman et al. 2004) and MITOS (Bernt et al. 2013), using default parameter settings, and were finally curated manually. For the comparative analysis, the mitogenomes were first aligned using ClustalW and MAFFT (Larking et al. 2002; Katoh et al. 2002).

Illumina DNA sequencing, assembly

To assess within-species variation in the mitogenome of *C. maculatus*, we prepared samples for Illumina sequencing from three additional populations with a geographic origin different from our “South India” reference population: California, Yemen and Brazil. From each of the three populations, we extracted samples of DNA from two different isofemale lines. A salt-ethanol precipitation protocol was used to extract high-quality DNA from our beetles. Beetles were first gently macerated and placed in preparation buffer (100 mM NaCl, 10 mM Tris-HCl, pH = 8.0, 0.5% SDS) together with proteinase K, vortexed and incubated at 50^{O}C overnight. Samples were then frozen overnight. To precipitate DNA, we added saturated NaCl several times before adding 95% ethanol, and we spun the DNA into a pellet. The DNA pellet was suspended in TE

buffer (pH = 7.6). DNA quality and quantity was assessed using NanoDrop, Qubit and Bioanalyzer, followed by fragment length assessment on an agarose gel.

Sequencing libraries were prepared using the TruSeq PCRfree DNA library preparation kit. The six libraries (3 populations × 2 isofemale lines) were then subjected to cluster generation and 125 cycles paired-end sequencing in 3 lanes using the HiSeq2500 system and v4 sequencing chemistry.

In total, we sequenced on average 150 million reads for each library. For each library, we used five percent of the sequenced reads (resulting in a mitogenome coverage of approximately 300X) and fed these to the MITObim V 1.8 algorithm (Hahn et al. 2013) and the MIRA V 4.0.2 (Chevreux et al. 1999) assembler to perform a guided assembly, using the assembled *C. maculatus* ("South India") mitogenome as a reference. For each library, several mitogenomes were assembled using different parameter settings. All obtained assemblies provided a circular mitogenome with a size similar to the reference genome and the two independent assemblies from each population (i.e., one from each isofemale line) were congruent. The final assemblies were then aligned using ClustalW and MAFFT, and manually curated to obtain the annotated mitogenome for each population.

Illumina RNA sequencing and analyses of transcription

Using the "South-India" standard reference strain, we prepared samples from larvae, pupae and adults for RNA sequencing. We refer to Sayadi et al. (2016) and Immonen et al. (2017) for a full description of these RNA sequencing and assembly efforts. Briefly, we prepared 11 types of samples, from larvae (6 individuals per sample), pupae (2 pupae per sample) and an adult mix (2 males and 2 females per sample). The other 8 sample types were all from adult beetles, either from mated individuals or virgins, from males or females and from the head and thorax (the somatic tissues) or from the abdomen (the reproductive tissues) (2 × 2 × 2 = 8). Three replicates of these latter types of samples were prepared, resulting in a total of 27 samples. Total RNA was extracted using RNeasy Mini Kit (Qiagen), following the manufacturer's protocol.

DNase digestion was applied using DNase I (RNase-Free DNase set by Qiagen). The RNA quality and quantity was assessed and affirmed using NanoDrop, Qubit and Bioanalyzer.

The RNA-seq libraries were prepared using the Illumina TruSeq stranded mRNA sample preparation kit according to the manufacturer's guidelines. Poly(A) containing RNA was enriched from total RNA using poly(T) oligo attached magnetic beads, after which mRNA was fragmented and reverse transcribed to first strand cDNA using random primers. The cDNA fragments were ligated to adapters and purified cDNA libraries enriched with PCR. All sequencing was performed using Illumina HiSeq 2500 sequencing technology producing 100 bp length paired-end reads. The 27 libraries were sequenced in three lanes.

Due to the lack of a reference genome, the transcriptome was assembled *de novo*, using all libraries. In total, more than 492 million read pairs were sequenced and used to generate a reference transcriptome with the Trinity assembler (Grabherr et al. 2011; Haas et al. 2013). The *de novo* assembly has been described in detail in Sayadi et al. (2016) and we refer to this source for information on transcripts. Data is available from NCBI, both as raw sequence reads (PRJNA308906) and as the assembled transcriptome (PRJNA309272).

All assembled transcripts were blasted against the mtDNA PCGs of the *C. maculatus* "South-India" mitogenome. We obtained at least one good hit to a transcript for each of the 13 genes, with some variation in length. The assembled transcripts were often somewhat longer, and in a few cases somewhat shorter, compared to the PCGs. One transcript (TR32651|c0_g1_i1) encapsulated two adjacent genes (ATP8 and ATP6). Variation in transcript length may in part be due to the Trinity assembler, which strives to elongate transcripts. We note here that assembled transcripts of seven mtDNA PCGs genes (COX1, COX2, COX3, ATP6/8, NAD5, NAD4, COB) showed clear evidence of a poly(A) tail.

To obtain a more accurate measure of mtDNA gene expression, we calculated expression levels by mapping all raw reads from each sample back to a transcriptome containing (1) all Trinity

assembled transcripts except the ones that mapped to a mtDNA PCG and (2) the predicted transcripts of the 13 mtDNA PCGs (ATP8 and ATP6 being joined). To quantify transcript abundance, we then used the RSEM package through the Trinity pipeline. The number of reads mapped to each transcript (read counts) was then normalized by transcript length to report the expression level value as 'fragments per kilobase transcript length per million fragments mapped' (FPKM). We note that this procedure produced FPKM values for all mtDNA PCGs that were very similar to measures of transcript abundance as estimated using the more conventional approach of mapping reads to assembled transcripts only.

Variation in transcription of mtDNA PCGs across the 24 samples from adults was analyzed using a three-way multivariate analysis of variance, where the FPKM values for all genes was treated as the response variable matrix and mating status, sex and tissue were factorial variables. This model provides omnibus tests of the effects of our factors on gene expression and so unnecessitates corrections for multiple testing. To better characterize effects, variation in expression of single PCGs was then analyzed using analyses of variance.

Analysis of selection and structure

To assess haplotype diversity we used PHYLIP (Felsenstein 2005) and DnaSP v. 5.10.01 (Librado and Rozas 2009) and to test for the molecular signatures of selection on PCGs in the mitogenome we used and the CodeML package in PAMLX v. 1.3.1 (Yang 2007; Xu B, Yang Z 2013), based on default parameter settings. These tests were based on the phylogenies of Tuda et al. (2006) and Kergoat et al. (2005) for the four species and a star phylogeny for the four populations of *C. maculatus*. We first fitted M0 models, which fit a single value of ω for each gene. We then tested for variable ω among sites within genes, by comparing M0 models with M3 models (which fits several discrete ω) by likelihood ratio tests for each gene.

To search the mitogenomes for repeat units and characterize tandem repeat motifs, we used Tandem Repeats Finder V 4.09 (Benson 1999).

2. SUPPLEMENTAL RESULTS

The seven mitogenomes assembled and analyzed here are deposited at GenBank under accession numbers KY856743, KY856744, KY856745, KY942060, KY942061, KY942062 and MF925724.

Start and stop codons

All PCGs started with typical start (ATN) codons, except for Cox1 which started with an (AAT) codon. In our study, PCGs start and stop codons, were selected based on the alignment of the translated amino acid sequences of the four beetle mitogenomes and on the minimization of intergenic spaces and gene overlaps. The Cox1 start codon position has remained a controversial topic in many studies of insect mitogenomes (Sheffield et al. 2008). Many have encountered difficulties with identifying a typical (ATN) start codon at the beginning of the open reading frame, without having a large gene overlap or a large intergenic spacer. In our study, we encountered a similar problem. By aligning the region that include tRNAy and Cox1 of the four mito-genomes, (AAT) appeared to be the most correct start codon for Cox1, as in several other beetles' mitogenomes (Richards et el. 2008; Sheffield et al. 2008; Zhu et al. 2012; Li et al. 2015). For *C. maculatus*, *C. analis*, *C. chinensis* and *A. obtectus* the intergenic spacer between tRNAy and Cox1 was 9 bp, 1 bp, 1 bp and 1 bp, respectively.

With regards to stop codons, eight out of the thirteen PCGs used (TAA) as a stop codon, while Nad1 had a (TAG) stop codon. The three remaining genes, Cox2, Nad5 and Nad4, showed (T) as an incomplete stop codon, except for Nad4 in *C. analis* which terminated with a proper (TAG) stop codon. Partial stop codons have also been reported in other beetles studied and they are common in invertebrates (Clary and Wolstenholme 1985; Li et al. 2007; Sheffield et al. 2008). It has been suggested that incomplete termination codons are transformed to a typical (TAA) stop codon after a posttranscriptional polyadenylation step (Ojala et al. 1981). In seed beetles, the number of incomplete termination codons is actually quite low compared to other beetles (Liu et al. 2014; Fang et al. 2015).

The comparison of all PCGs for the four mitogenomes showed a conserved pattern in terms of start, stop codon and gene size (SI Table 1). The lowest sequence identity recorded was 74.84 % between *C. chinensis* and *C. analis* for atp8. The remaining genes showed high pairwise sequence identity, in most cases considerably higher than 80 %.

Control region

The control region (CR) is a large non-coding region, typically very AT rich. It plays an essential role in the initiation of the transcription and replication, and is therefore termed the control region. In our study, the CR was positioned between rRNAs and tRNA-I in all mitogenomes. The length of the control region was 1,031 bp in *C. maculatus*, 1,024 bp in *C.analis*, 1,230 bp in *C. chinensis* and 1,306 bp in *A. obtectus*. The AT content in the CR was 79.22 % in *C. maculatus*, 82.5 % in *C.analis*, 81.45 % in *C. chinensis* and 83.2 % in *A. obtectus*, compared to the coding regions where AT content was 74.9 % in *C. maculatus*, 75.86 % in *C.analis*, 76.04 % in *C. chinensis* and 75.7 % in *A. obtectus* (SI Table 2). Comparing the four CRs at their sequence level, we found that *C. maculatus* share 67.23 % sequence identity with *C. analis*, 61.96 % sequence identity with *C. chinensis* and 54.85% sequence identity with *A. obtectus*. Finally, *C. analis* only share 63.52 % sequence identity with *C. chinensis* and 56.25% sequence identity with *A. obtectus*. These low sequence identities illustrate the fact that the CRs show relatively rapid evolution (Zhang and Hewitt 1997).

A tandem repeats search of the CRs, using the default cut-off of 50 as the minimum alignment score to report, revealed no significant TRs in any of the control regions. Lowering the cut-off to 40, we did uncover a short TR of 11 bp in *C. maculatus*, represented in 2 full copies, and a partial copy of 9 pb in length. In *C. analis,* we also identified a small TR of 10 bp, repeated 2.5 times. In *C. chinensis*, we discovered a TR of 13 bp repeated 2 times. In *A. obtectus*, a short TR of 2 bp repeated 19 times was present (SI Table 3). No sequence similarity was found between the TRs, which is unsurprising given the low sequence identity between the four CRs. CR length is usually highly variable between insect species, and this variation is mainly due to the

differences in the copy number of TRs (Mardulyn et al. 2003). This was not the case in these seed beetles.

Intergenic spacers and gene overlap

In the four mitogenomes, short intergenic spacers of variable size occurred over the entire mitogenome (SI Table 1, SI Figure 2). The unique feature of the seed beetle mitogenomes is the presence of two long intergenic spacers (LIGSs), by far the longest ever assembled. The first (LIGS1) is located between Nad2 and tRNAw. The second (LIGS2) lies between tRNAq and Nad1 (Figure 1). The length of these LIGSs varied between the four mitogenomes. LIGS1 has a length of 2,067 bp in *C. maculatus*, 3,341 bp in *C.analis*, 6,456 bp in *C. chinensis* and 114 bp in *A. obtectus.* The length of LIGS2 is 7,009 bp in *C. maculatus*, 5,641 bp in *C.analis*, 1,996 bp in *C. chinensis* and 10,408 bp in *A. obtectus*. Pairwise comparisons of these regions within and between the four mitogenomes did not result in any significant sequence similarity. Another key facet of these LIGSs is the high AT content. It is, in fact, even higher than the AT content of the AT-rich control region (SI Table 2).

The other intergenic spacers were short, not more than 93 bp in length, spread all over the mitogenome, and showed no obvious consistency in size over the four mitogenomes (SI Figure 2). For example, a spacer of 93 bp in length is located between tRNAs and tRNAq in *C. maculatus*, but this spacer is 24 bp in *C. analis* and 18 bp in *C. chinensis*. In terms of TRs in short spacers, we found only one TR in the spacer between tRNAe and tRNAf in *C. analis*. The length of this spacer is 38 bp, holding a TR of 2 bases (AT) repeated 15 times (SI Table 3). In total, we counted 13 short spacers in *C. maculatus*, 15 in *C. analis*, 13 in *C. chninensis* and 13 in *A. obtectus*. Comparing the combined length of all short spacers, excluding LIGS1 and LIGS2, *C. maculatus* has the highest sum. In total 243 bp represents short intergenic spacers in *C. maculatus*, against 162 bp in *C. analis,* 135 bp in *C. chinensis* and 142 bp in *A. obtectus. Acanthscelides obtectus* has the longest LIGS1 and LIGS2, they both sum to 10,522 bp, followed by 9,074 bp in *C. maculatus*, versus 8,981 bp in *C. analis*, and 8,450 bp in *C. chinensis*.

Few genes overlapped in the four mitogenomes. Only 9 gene overlaps occurred in *C. maculatus* and *C. analis,* totaling 31 bp in length. This number is somewhat higher in *C. chinensis*, were we found 10 gene overlaps totaling 32 bp in length. In *A. obtectus* the number was higher yet, with 11 gene overlaps totaling 33 bp in length.

LIGSs and tandem repeats

The entire mitogenomes were subjected to tandem repeat (TR) searches. This showed that TRs occurred almost exclusively in the LIGSs. They were in fact riddled with TRs, with the exception of LIGS2 in *C. analis* which was only partly formed by small TRs (less than 35 pb in length). A few other small TRs were also found, in some short intergenic spacers and in a few genes (e.g. Nad1, Nad2, Nad4, Nad5 and Nad6) (SI Table 3). Pairwise comparison of all TRs of the four mitogenomes did not reveal any significant sequence similarity of TR motifs, which is expected given the low sequence identity of the LIGSs.

The longest predicted TR is 372 bp in *C. maculatus*, and is represented by almost 3 full copies. The first copy is partly shared between Nad2 and LIGS1. The tandem repeat starts from the last 258 bases of Nad2 and ends in the first third part of LIGS1 in *C. maculatus*. The remainder of LIGS1 is almost fully formed by other smaller TRs. In *C. analis*, LIGS1 contain a TR of 262 bp in nine full copies and a partial copy of 196 bp that covers 81 % of LIGS1. Three TRs were found in LIGS1 of *C. chinensis*, covering 93 % of the sequence. The first covered in total 2834 bp, comprising 27.6 copies of a 103 bp TR. The second covered the second half of LIGS1, with 60.8 copies of 52 bp. The last is a short TR (AT), iterated 16 times (SI Table 3).

A similar pattern was also seen in LIGS2, with TRs covering almost all of the LIGSs, except for *C. analis*. In this species, only a few TRs were identified covering some 5% of LIGS2. For *C. maculatus*, four TRs were predicted in LIGS2. The longest is 164 bp in length, repeated 30.4 times. The second TR is 81 bp and is iterated 18.4 times. The last two TRs are short tandem repeats with 2 bases (TA and AT); (TA) was repeated 51 times, and (AT) was repeated 23.5 times. In total, TRs covered 95 % of LIGS2. For *C. chinensis*, LIGS2 contained 2 TRs; one long TR

of 209 bp, covering 1501 bp, in 7.2 copies. The second is a short TR (AT), iterated 15 times. In *A. obtectus*, TRs were detected only in LIGS2 which was composed mainly by three blocks of TR arrays. The first block contains a TR of 155 bp repeated 7 times. The next block is formed by a 90 bp TR, iterated 60.8 times. Finally the last TR is 51 bp long and repeated 70.4 times.

Importantly, we found no evidence for conserved sequence blocks in TR motifs across the LIGs. This was obvious when inspecting the results of multiple sequence alignments of the LIGSs of the four species as well as from efforts to align all extracted TRs, using both ClustalW and MAFFT, and explicit searches for conserved blocks identified none. For example, alignments of the entire LIGSs of the four species yielded no identical blocks >9 bp for either LIGS1 or LIGS2 and the vast majority of identical block were only a few bp long. Given the high AT content of the LIGSs, this low degree of identity is consistent with a random expectation. For example, sequence identities of aligned sequences for LIGS1 (LIGS2) were 0.275 (0.235) for *C. analis* and *C. chinensis*, 0.304 (0.395) for *C. analis* and *C. maculatus* and 0.223 (0.160) for *C. chinensis* and *C. maculatus*.

Blasting the LIGSs sequences in NCBI, using default settings, yielded no hits. However, when including also LIGS regions of low compositional complexity, a few notable and biologically relevant partial hits appeared. In *C. maculatus*, the initial ≈150 bp of LIGS2, dominated by a short TR composed of 2 bases (TA) repeated 51 times (SI Table 3), mapped well against the CR of the mitogenome of several other insects (e.g., KF385868.1; KR703583.1; EU871947.1; KJ101608.1) (all E-values < 1×10^{-36}). Further, the terminal ≈100 bp part of LIGS1 in *C. chinensis*, also dominated by short TRs composed of 2 base repeats (AT) (SI Table 3), also mapped well against the CR of the mitogenome of other insects (e.g., KT876896.1; KP995260.1; AB242844.1; HQ335349.1) (all E-values < 2×10^{-14}). Hence, a few short blocks of TRs of the LIGSs do show a significant sequence similarity to the mtDNA CR of other insects.

Apart from the TRs found in the LIGSs, we also identified a few TRs in coding regions. Relatively short TRs were also present in a few genes; Nad1, Nad2, Nad4, Nad5 and Nad6. The longest is 20 bp. They were represented in a low number of copies and with a low alignment score, and they were not shared between the four mitogenomes (SI Table 3).

Within-species variation

The mitogenomes of the three populations of *C. maculatus* (Brazil, California, Yemen) all showed the same organization as the reference mitogenome (South India) and start and stop codons were conserved, but the genomes exhibited some variation in total size (SI Table 4). Compared to South India (25,011 bp), Yemen had the largest mitogenome (25,069 bp) followed by California (25,026 bp) and Brazil (24,947 bp) (SI Table 2).

The sequence identity of the LIGSs in the four populations was high and the consensus pattern of TRs in LIGS1 and LIGS2 was very similar (SI Table 5). No new tandem repeats was discovered in LIGS1. LIGS2 was formed mainly by 4 TRs. The first one is an (AT) TR that covered the first ~100 bp of LIGS2 in South India but was replaced by another longer TR of 91 bp repeated twice in Brazil and California. Yemen showed an even longer TR version with 2 copies of 127 pb. Another TR of 34 bp with 2 copies just adjacent to the previous, was present only in California and Yemen. The rest of TRs were identical in all four populations and covered the same regions. The 2 TRs found only in Yemen and California makes them the most repeat-rich of the four populations.

3. SUPPLEMENTAL LITERATURE CITED

Benson, G. (1999) Tandem Repeats Finder: a program to analyse DNA sequences. Nucleic Acids Res. *27*, 573–578.

Chaisson, M.J., and Tesler, G. (2012) Mapping single molecule sequencing reads using Basic Local Alignment with Successive Refinement (BLASR): Theory and Application. BMC Bioinformatics *13*, 238.

Chevreux, B., Wetter, T., and Suhai, S. (1999) Genome sequence assembly using trace signals and additional sequence information. German Conf. Bioinf. *99*, 45-56.

Clary, D.O., and Wolstenholme, D.R. (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. *22*, 252–271.

Fang, J., Qian, L., Xu, M., Yang, X., Wang, B., and An, Y. (2015) The complete nucleotide sequence of the mitochondrial genome of the Asian longhorn beetle, *Anoplophora glabripennis* (Coleoptera: Cerambycidae). Mitochondrial DNA *1394*, 1–2.

Felsenstein, J. (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotech. *29*, 644-652.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols *8*, 1494-1512.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A., McWilliam, H., et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup. (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, W., Yang, X., Qian, L., An, Y., and Fang, J. (2015) The complete mitochondrial genome of the citrus long-horned beetle, *Anoplophora chinensis* (Coleoptera: Cerambycidae). Mitochondrial DNA 1736, 1–3.

Li, X., Ogoh, K., Ohba, N., Liang, X., and Ohmiya, Y. (2007) Mitochondrial genomes of two luminous beetles, *Rhagophthalmus lufengensis* and *R. ohbai* (Arthropoda, Insecta, Coleoptera). Gene *392*, 196–205.

Liu, Q.N., Bian, D.D., Jiang, S.H., Li, Z.X., Ge, B.M., Xuan, F.J., et al. (2014) The complete mitochondrial genome of the red flour beetle, *Tribolium castaneum* (Coleoptera: Tenebrionidae). Mitochondrial DNA 1394, 1–3.

Mardulyn, P., Termonia, A., and Milinkovitch, M.C. (2003) Structure and evolution of the mitochondrial control region of leaf beetles (Coleoptera: Chrysomelidae): a hierarchical analysis of nucleotide sequence variation. J. Mol. Evol. *56*, 38–45.

Ojala, D., Montoya, J., and Attardi, G. (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature *290*, 470–474.

Richards, S., Gibbs, R., Weinstock, G.M., Brown, S.J., Denell, R., et al. (2008) The genome of the model beetle and pest *Tribolium castaneum*. Nature *452*, 949–55.

Sheffield, N.C., Song, H., Cameron, S.L., and Whiting, M.F. (2008) A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. Mol. Biol. Evol. *25*, 2499–2509.

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics *14*, 178-192.

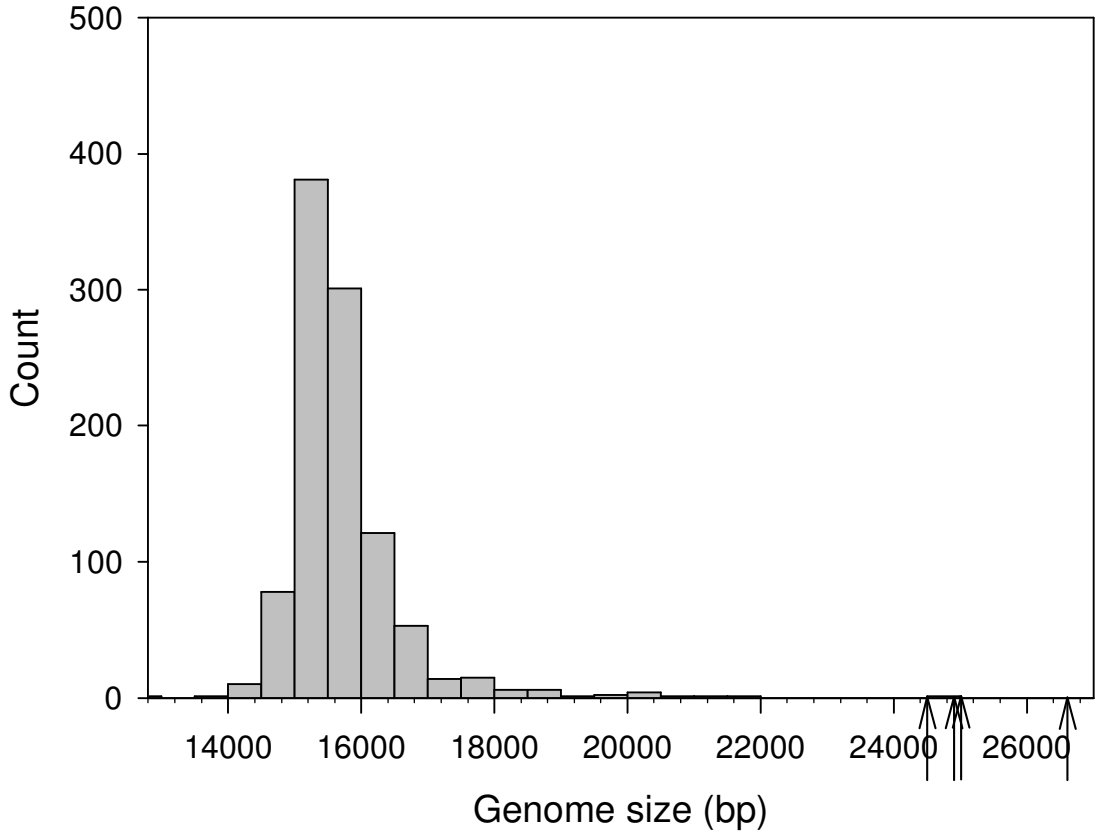
Wyman, S.K., Jansen, R.K., and Boore, J.L. (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252–3255.

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

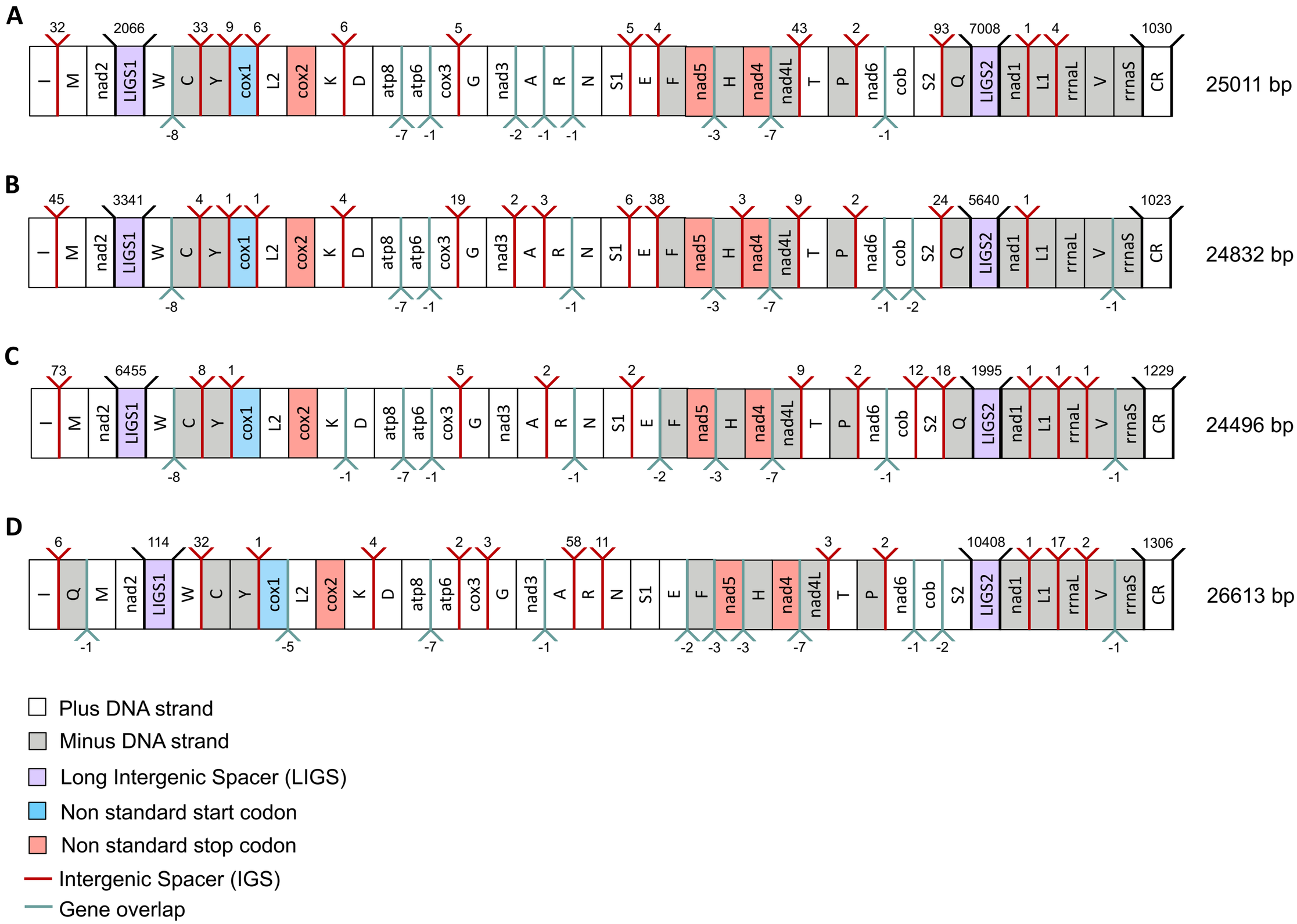
Yang, Z. (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586-1591.

Zhu, J.Y., Zhao, N., and Yang, B. (2012) Global transcriptome profiling of the pine shoot beetle, *Tomicus yunnanensis* (Coleoptera: Scolytinae). PLoS One *7*, 1–12.

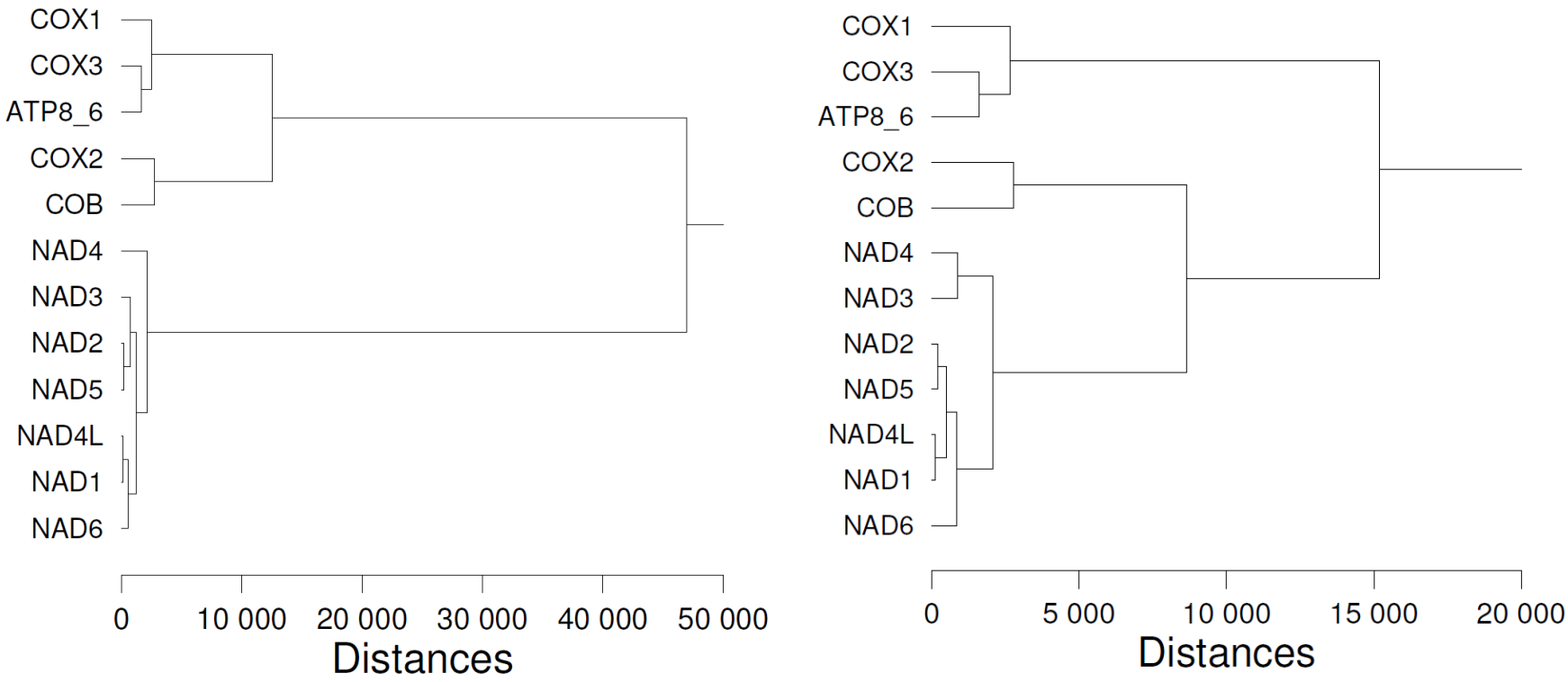
Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31*, 3406-3415.



SI Figure 1. Frequency distribution of all assembled insect mitogenomes (N=998) by size, retrieved from NCBI 2016-12-08 (excluding Phthiraptera which have very small mitogenomes). The arrows represent the four species studied here (three *Callosobruchus* species and *Acanthoscelides obtectus*).

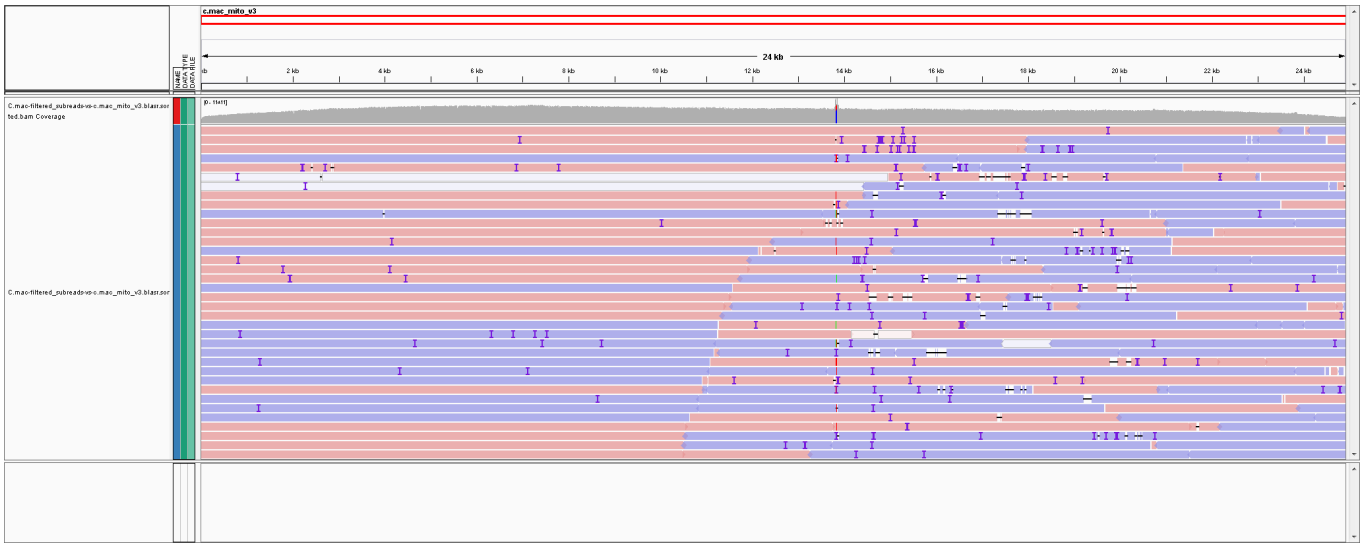


SI Figure 2. Mitogenome gene organization. (**A**) *C. maculatus,* (**B**) *C. analis,* (**C**) *C. chinensis,* (**D**) *A. obtectus*. The circular mitogenomes are here linearized for illustration. Genes lengths are not scaled.



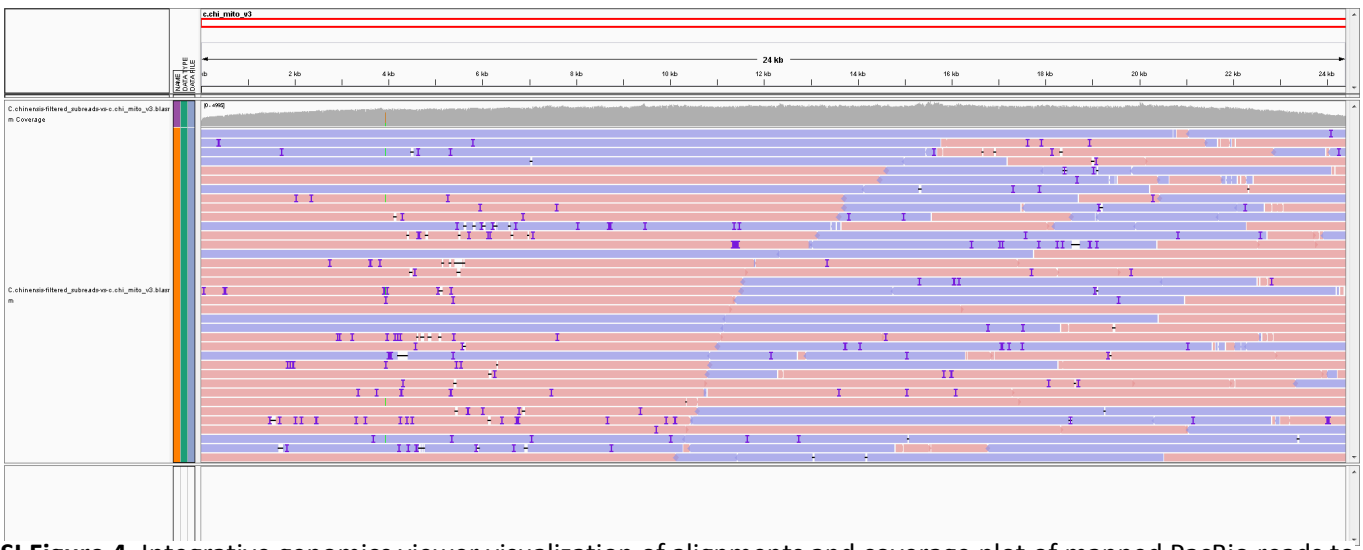
SI Figure 3. Hierarchical cluster analyses of covariation in mRNA transcript abundance (FPKM values) across all samples of *C. maculatus*, based on the Euclidean distance matrix and either Ward's (left) or complete (right) linkage. We note here that abundance of all mtDNA genes were correlated across samples. A principal component analysis (PCA; based on the correlation matrix) of within-species variation in abundance among the 12 genes across the 27 samples yielded a first PC which accounted for 90.8 % of the total variance in gene expression. Yet, the expression of NAD genes covaried relative to other genes. See Figure 2 for expression levels of all genes.

A



B

C



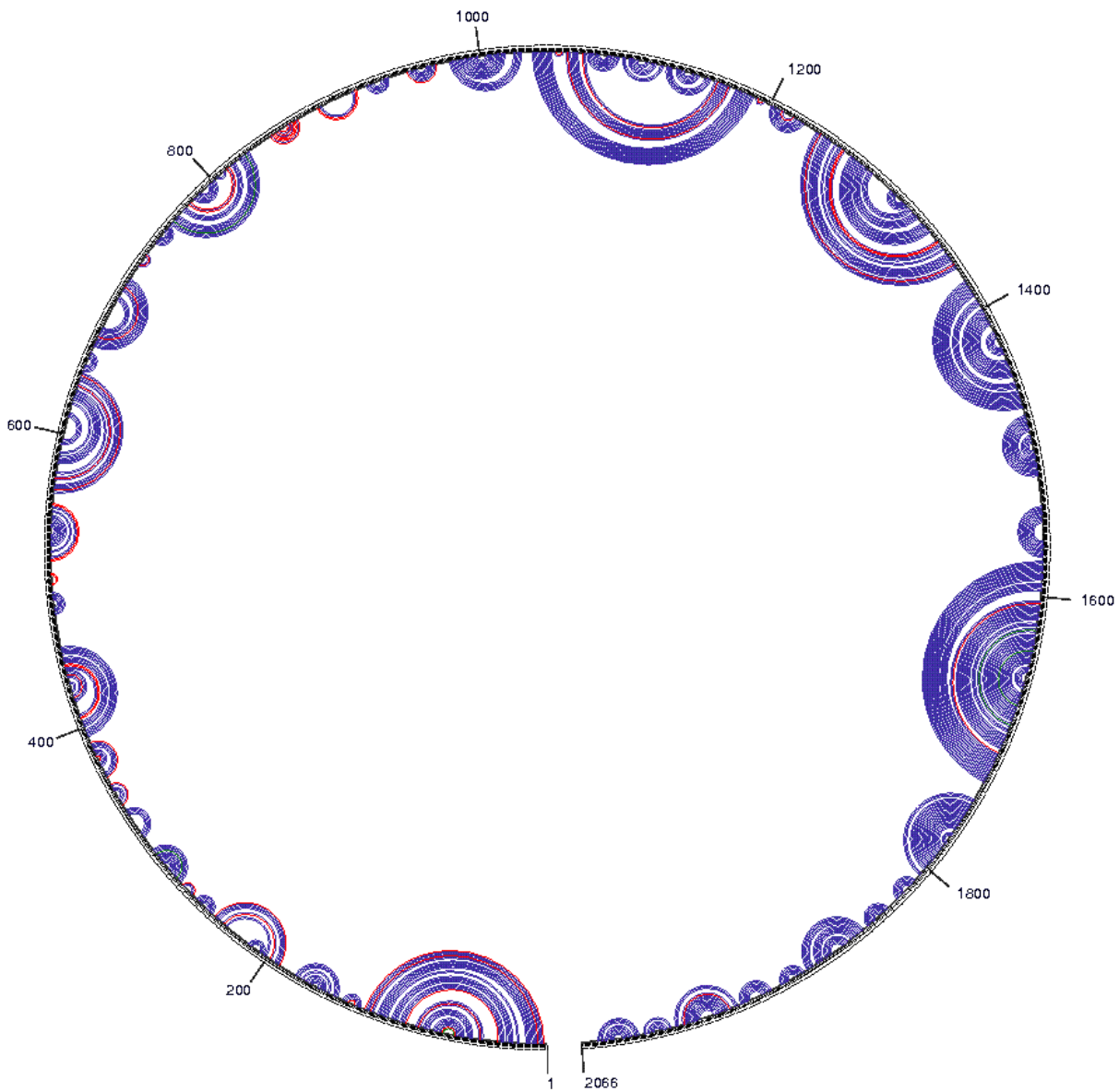
SI Figure 4. Integrative genomics viewer visualization of alignments and coverage plot of mapped PacBio reads to mitogenomes of (A) *C. maculatus,* (B) *C. analis*, and (C) *C. chinensis*. First panel shows the mitogenome as a linear representation. The second panel represents the coverage. The third panel represents a small subset of aligned reads. Forward reads are indicated in red, reverse reads in blue. PacBio reads were mapped back to the mitogenomes using Blasr (Chaisson and Tesler 2012), sorted and indexed using Samtools v. 1.4 (Li et al. 2009) and visualized using IGV software v 3.0 (Thorvaldsdóttir et al. 2013). Note that (1) coverage is uniform and high across the entire mitogenome and (2) many reads span a large part of the mitogenome.

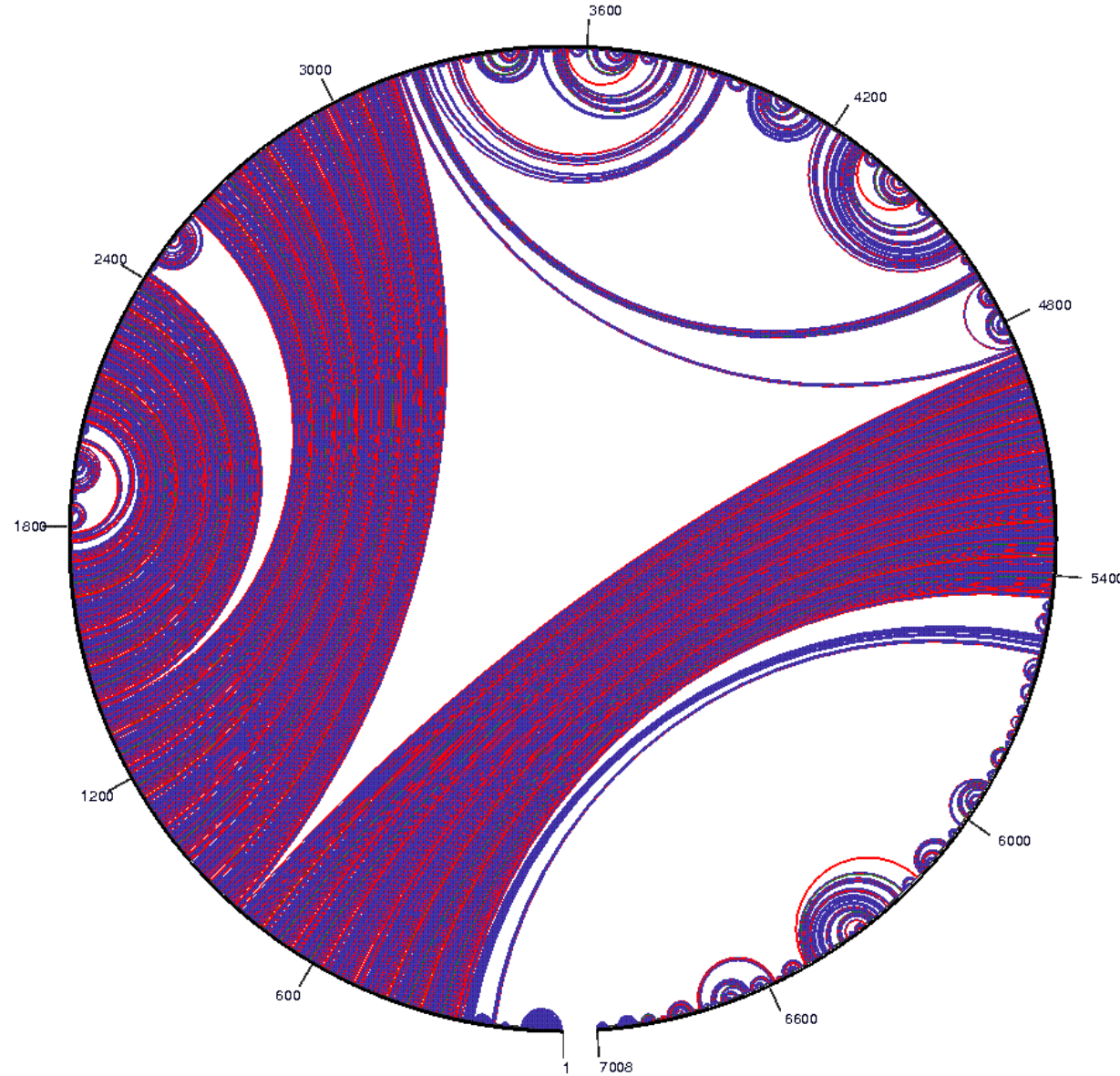
SI Figure 5. Predicted DNA folding of the LIGSs of the four seed beetle species *C. maculatus*, *C. analis*, *C. chinensis* and *A. obtectus*. Predictions were generated in Mfold [S21], using default parameters apart from folding temperature which was set equal to the mean temperature in June for the geographic center of origin for the four species; *C. maculatus*: 30^{o}C (Nigeria, West Africa); *C. analis*: 29^{o}C (India, South Asia); *C. chinensis:* 22^{o}C (Eastern China, East Asia); *A. obtectus*: 28^{o}C (Cancun, Mexico). All predictions represent the most stable secondary structure (that minimizing ΔG).

All panels show both LIGS1 (left) and LIGS2 (right). Panels A, C, E and G are circular structure plots, where paired bases are interconnected by arcs. Here, G-C pairings are drawn in red, A-T pairings in blue and G-T pairings in green. Panels B, D, F and H illustrate the same folding as the previous panels, but show the topology of the predicted secondary structure. As would be expected for A-T rich repeat arrays, the LIGSs are predicted to form multiple hairpin loops. Yet, the predicted secondary structure is in several cases very striking indeed, forming multiple markedly extended hairpin loops and stacks (e.g., LIGS2 of *C. maculatus*, LIGS1 of *C. chinensis*).

Panel A

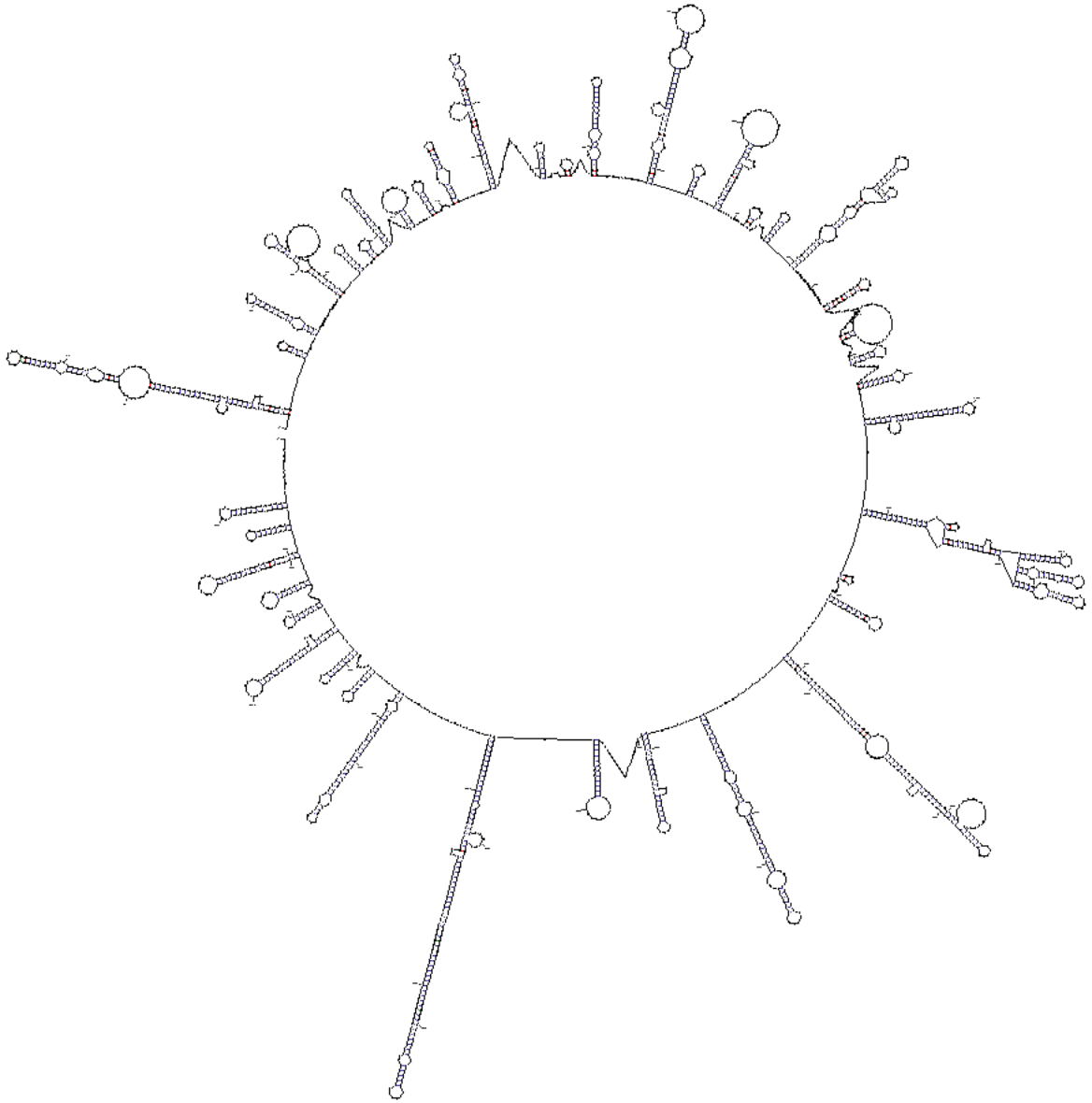
C. maculatus

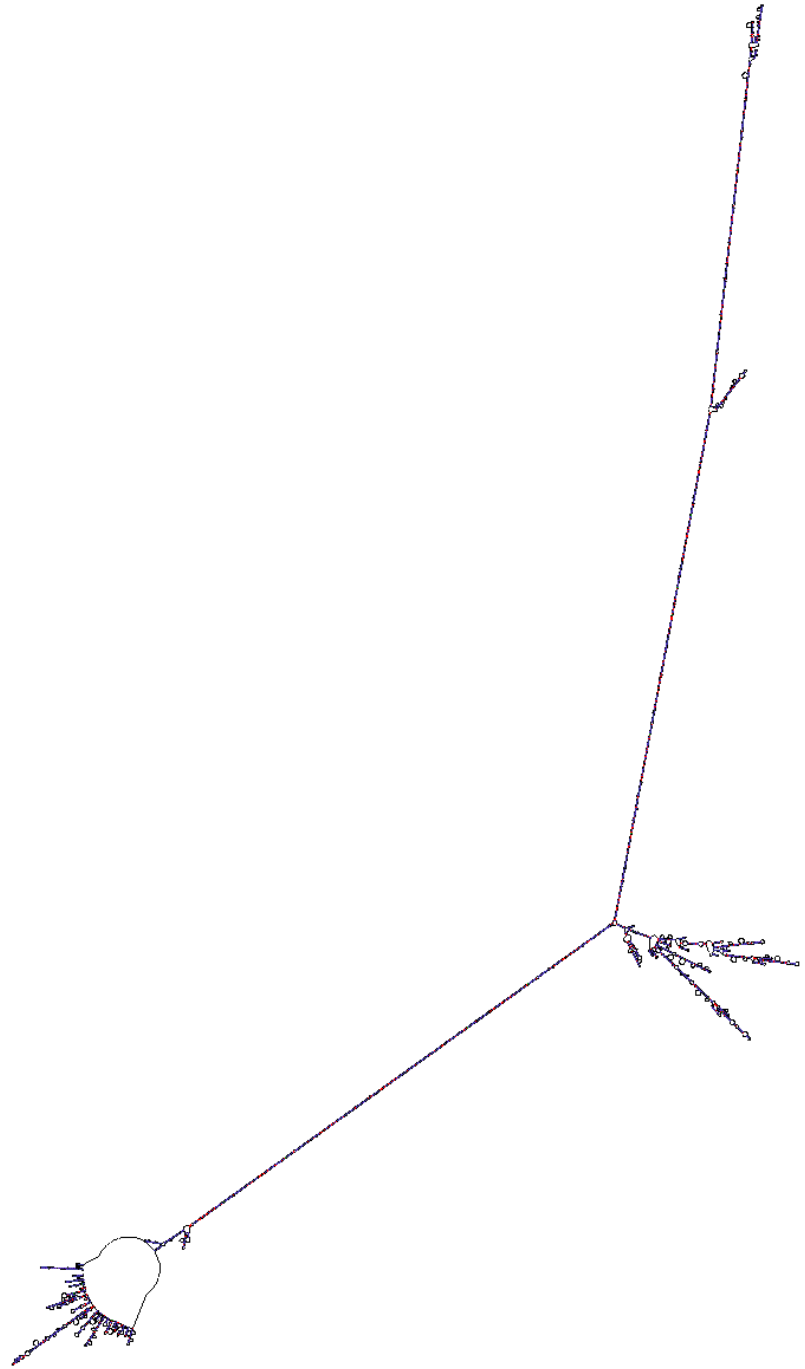




Panel B

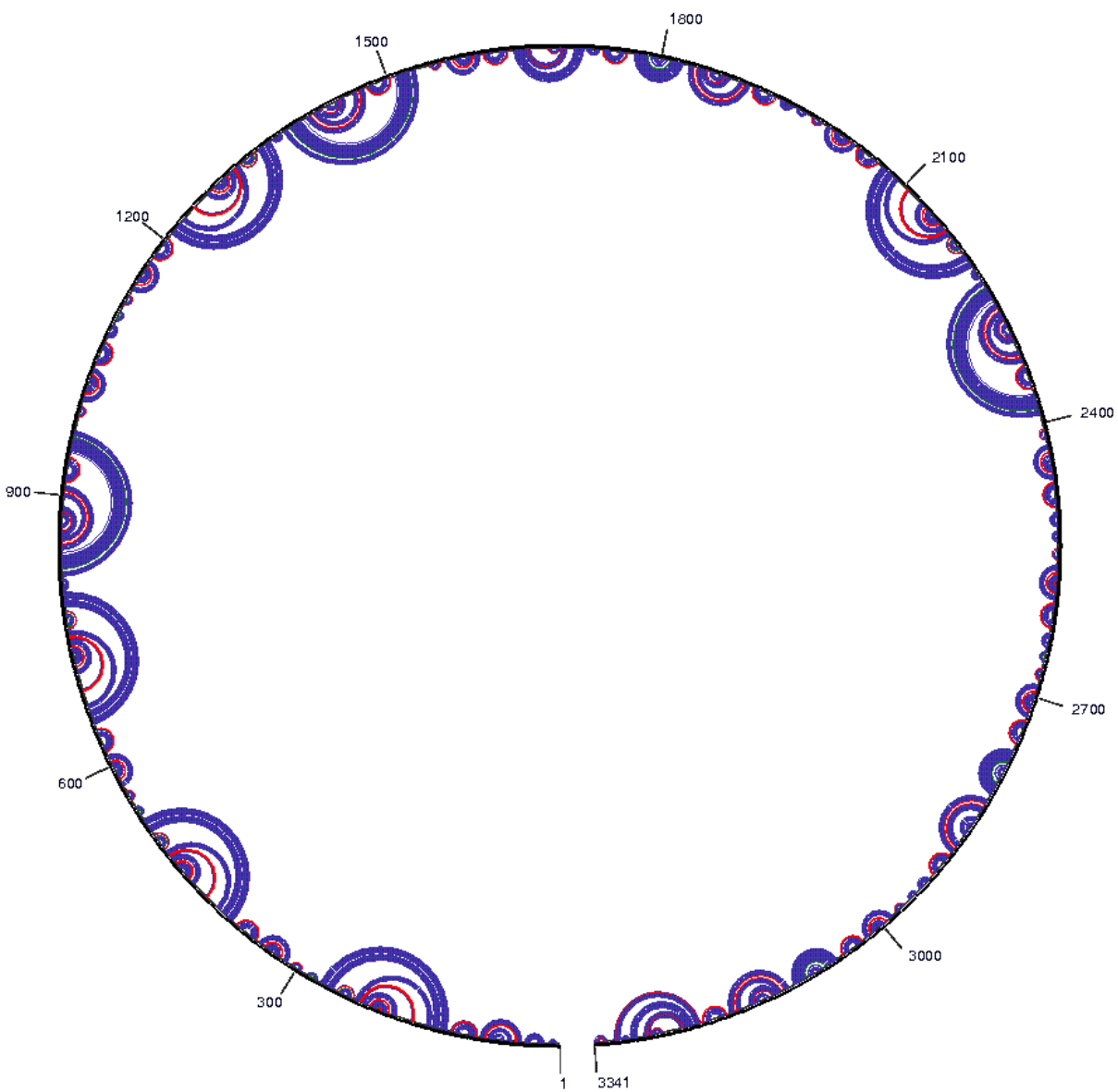
C. maculatus

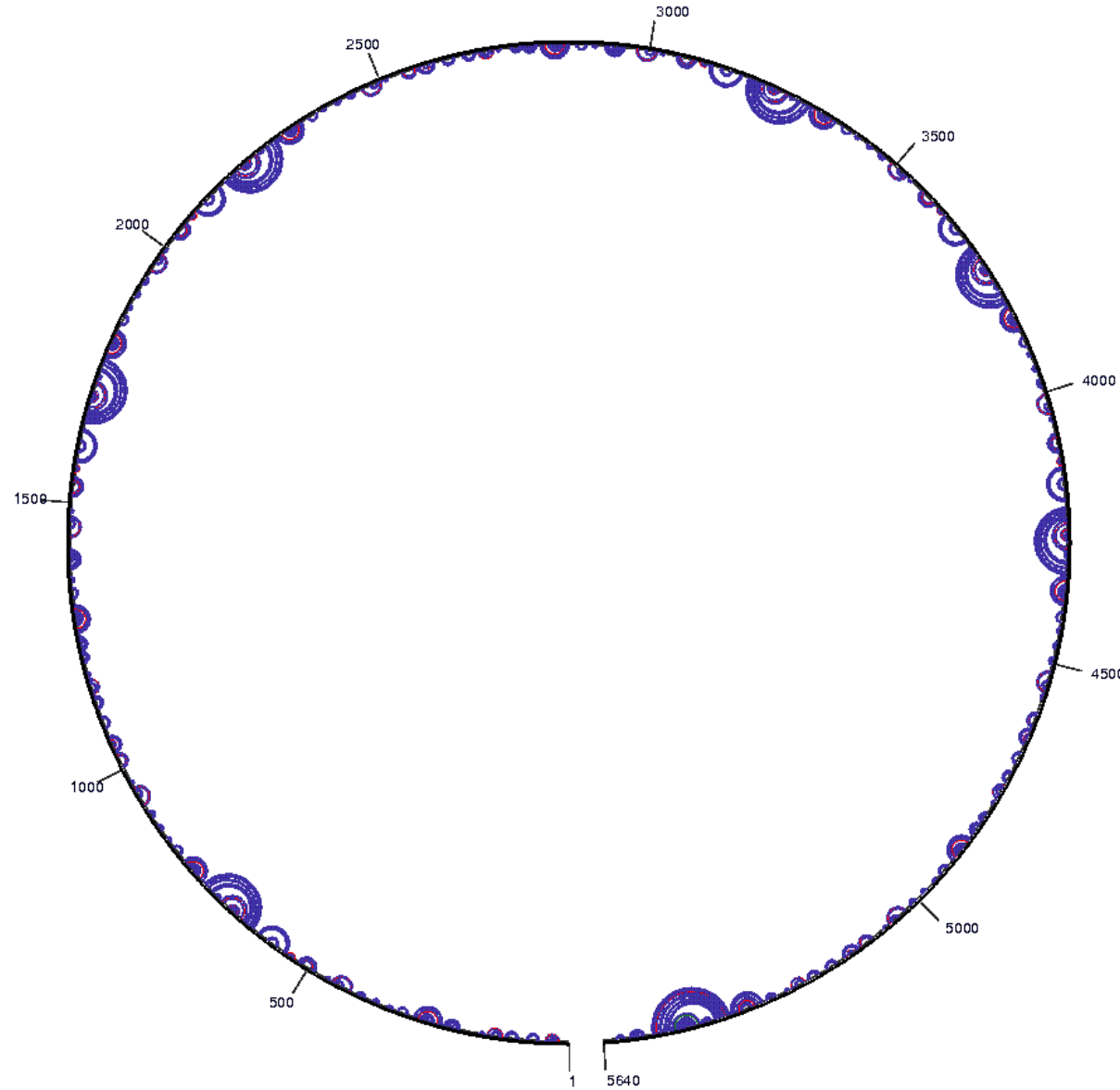




Panel C

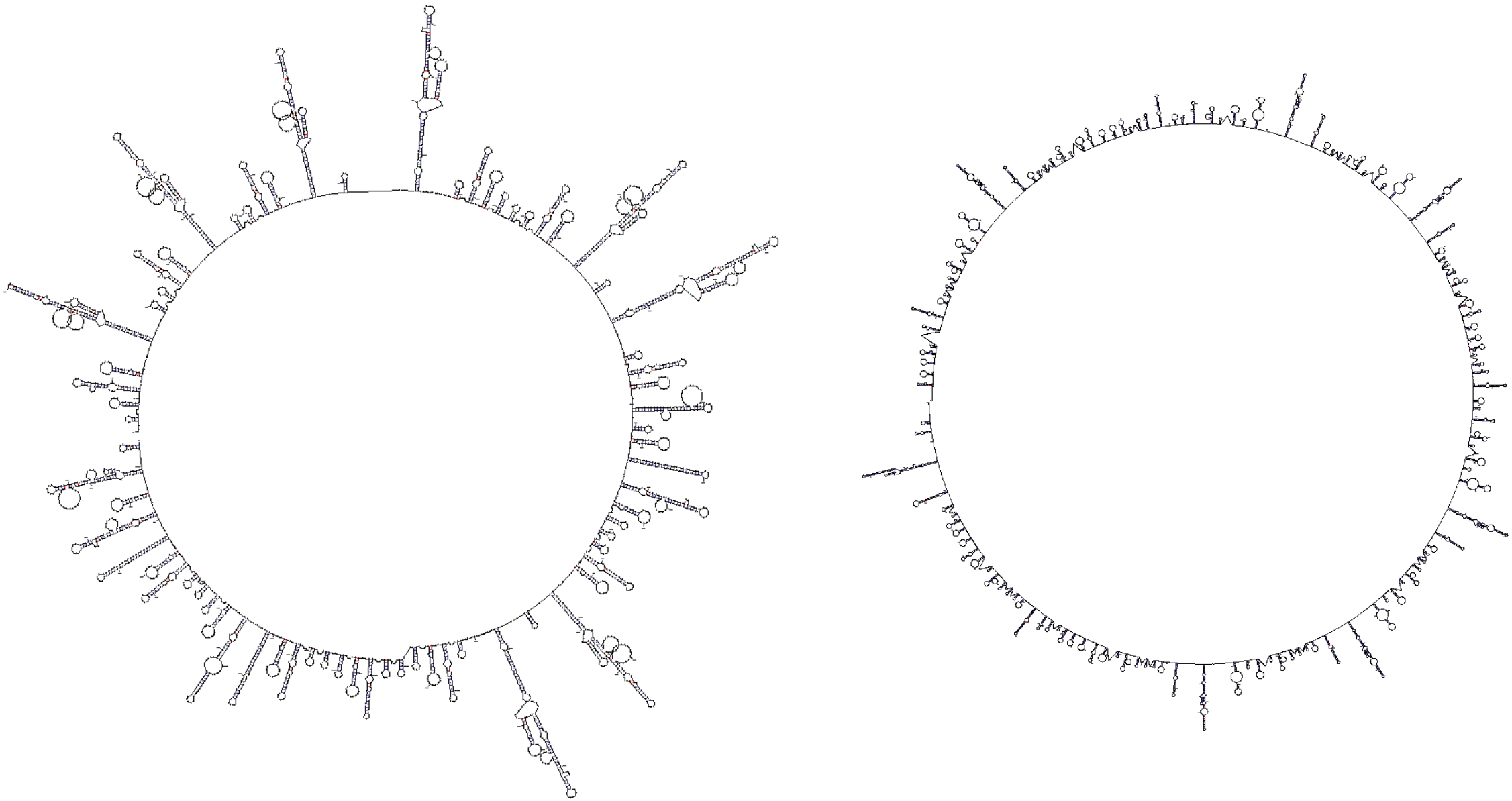
C. analis





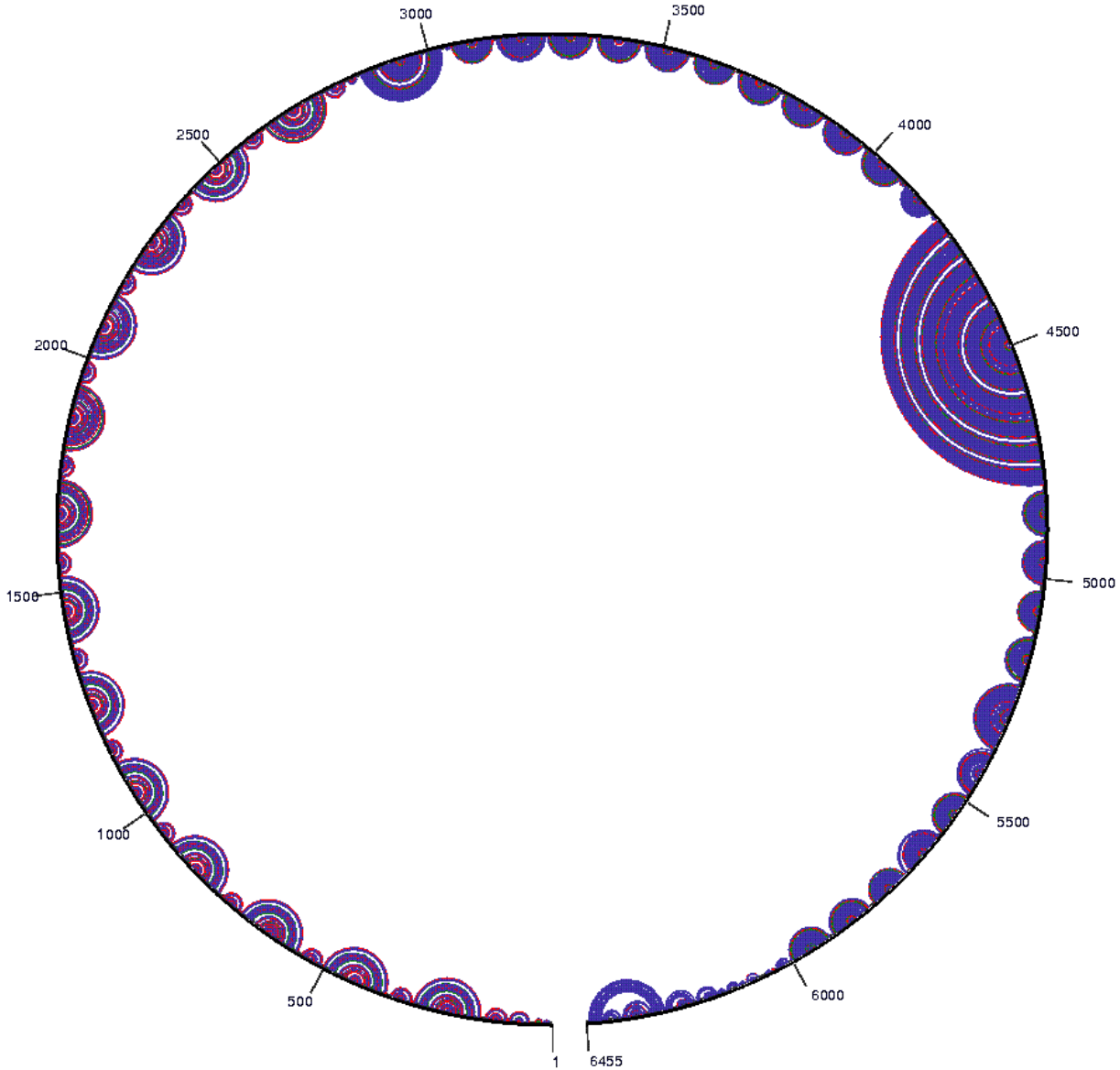
Panel D

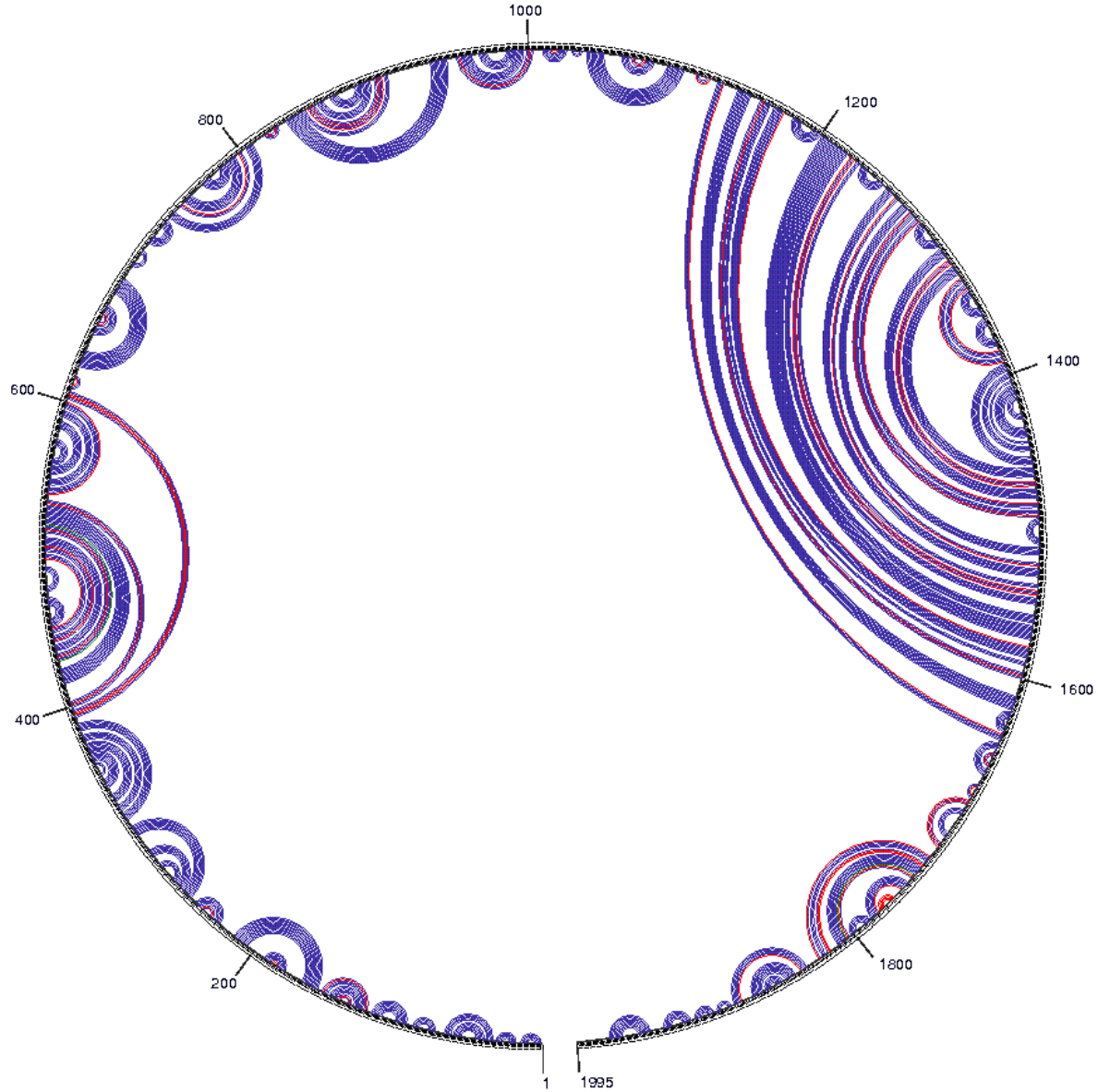
C. analis



Panel E

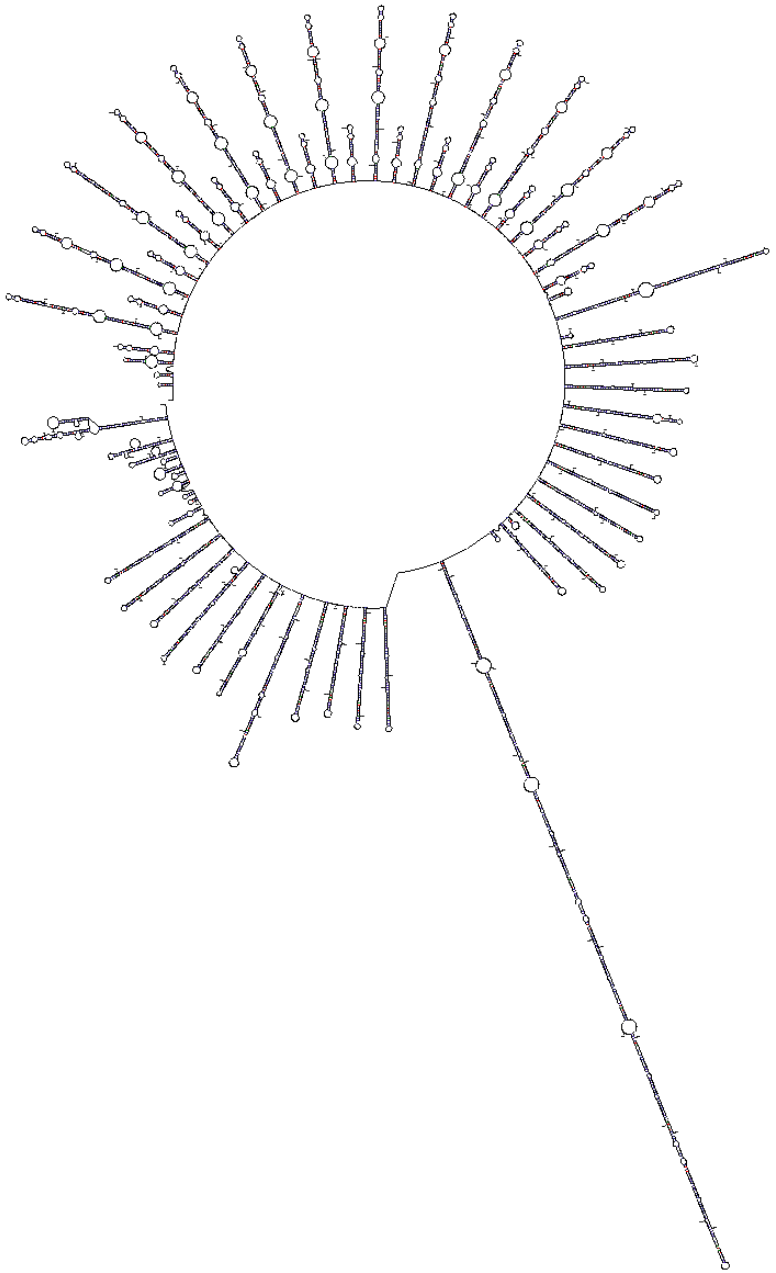
C. chinensis

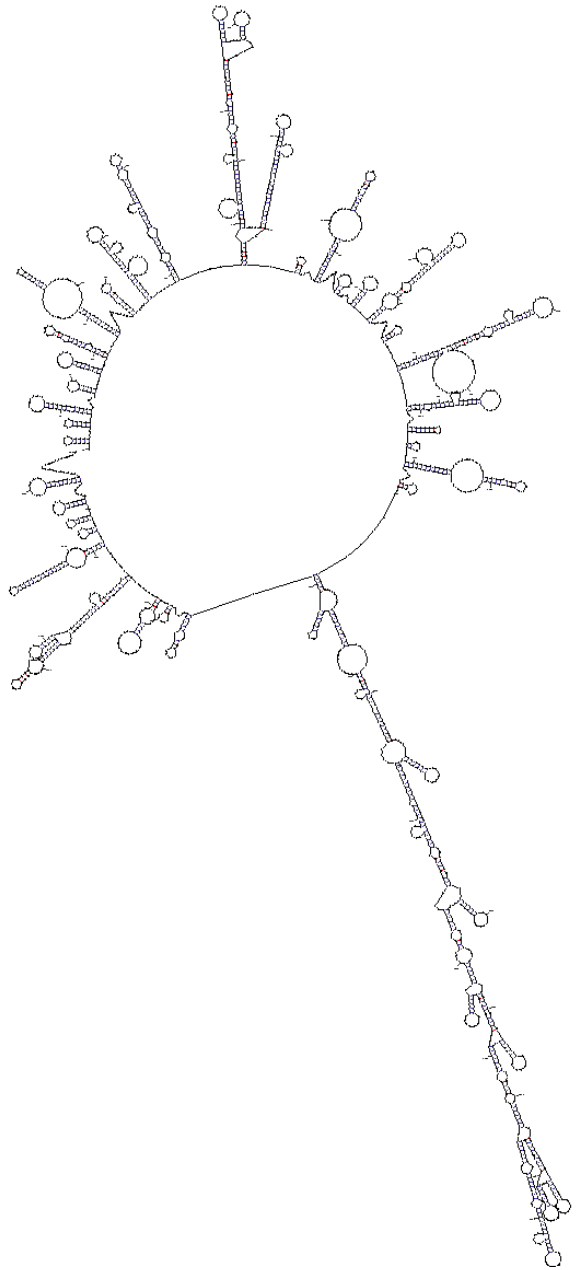




Panel F

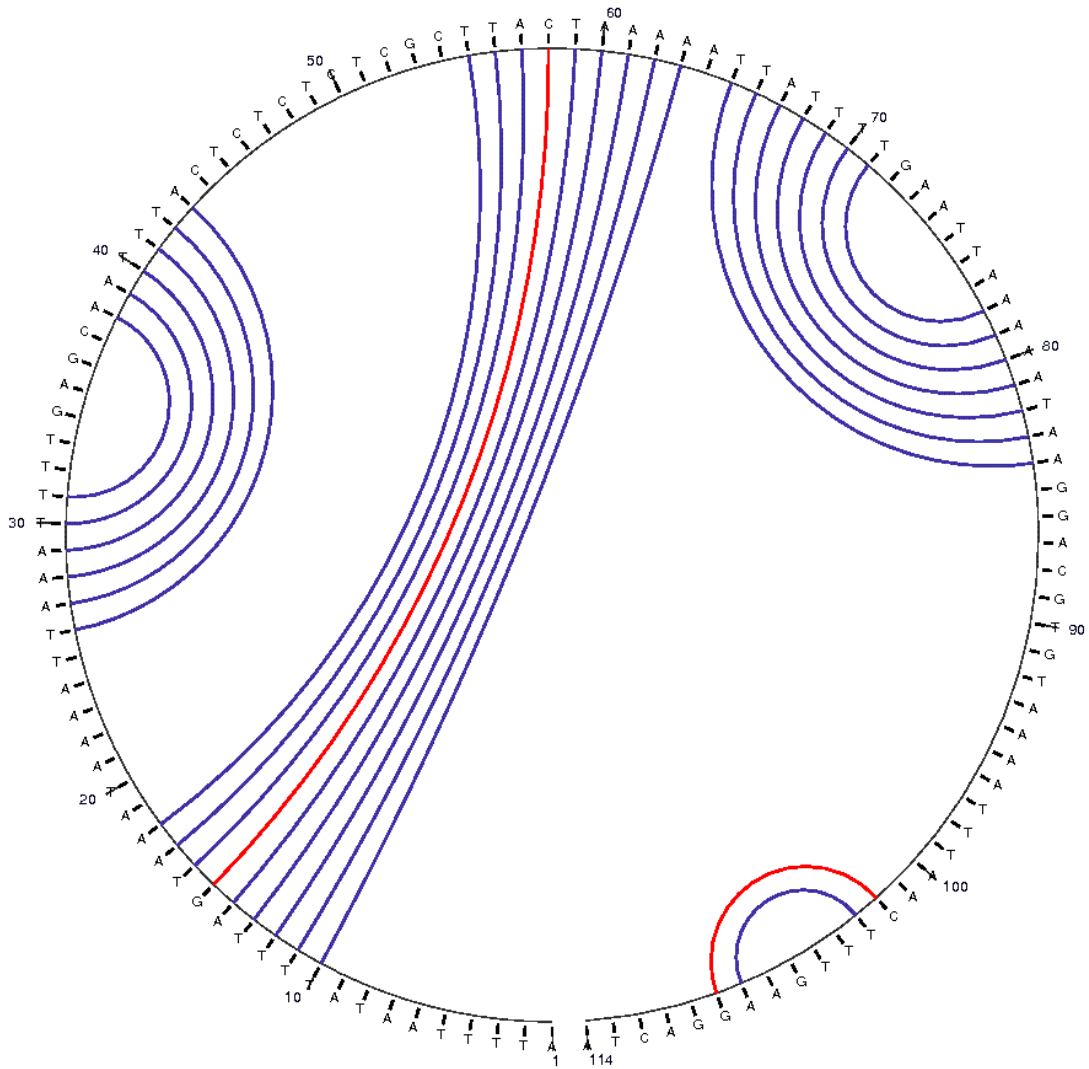
C. chinensis

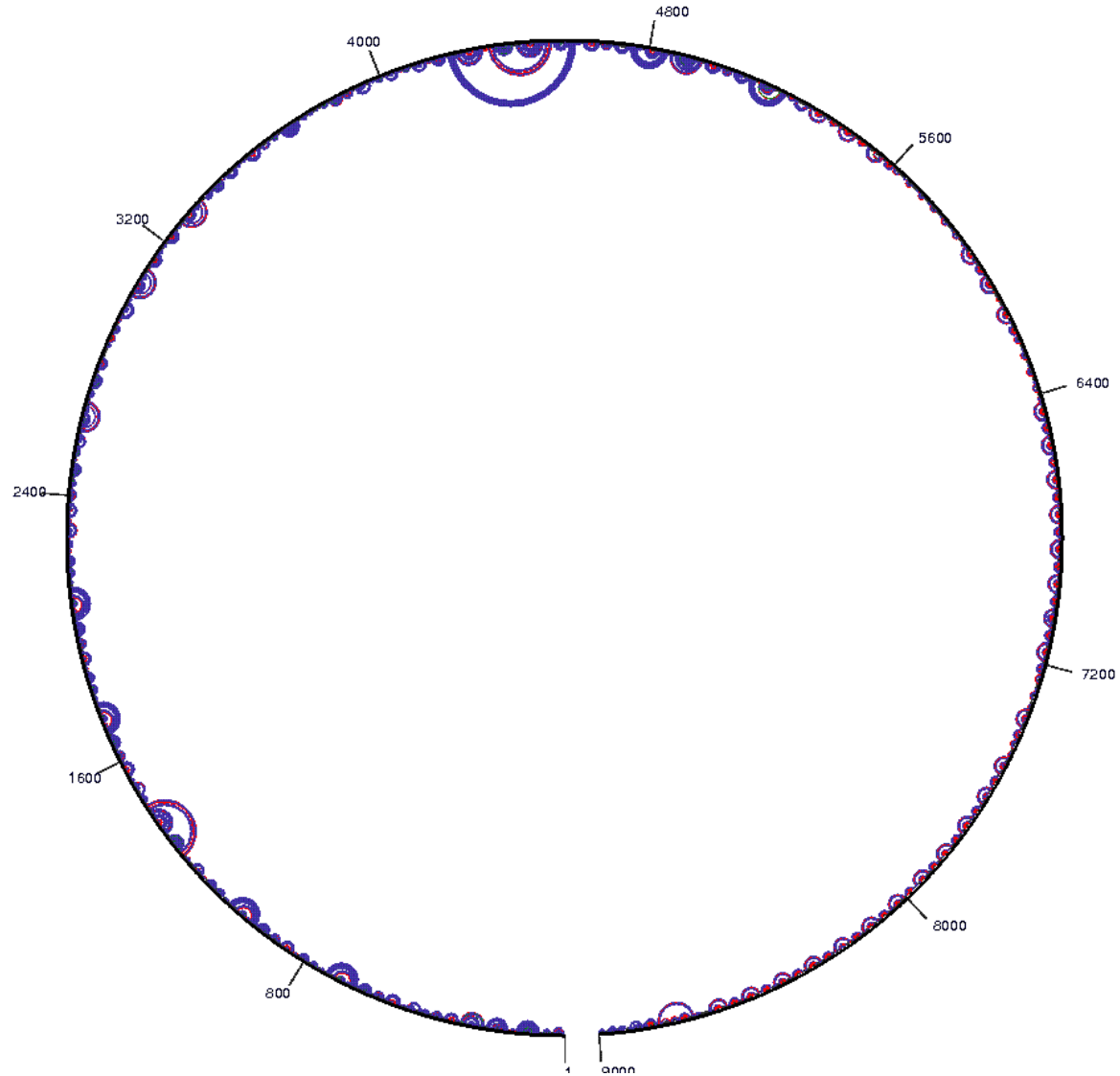




Panel G

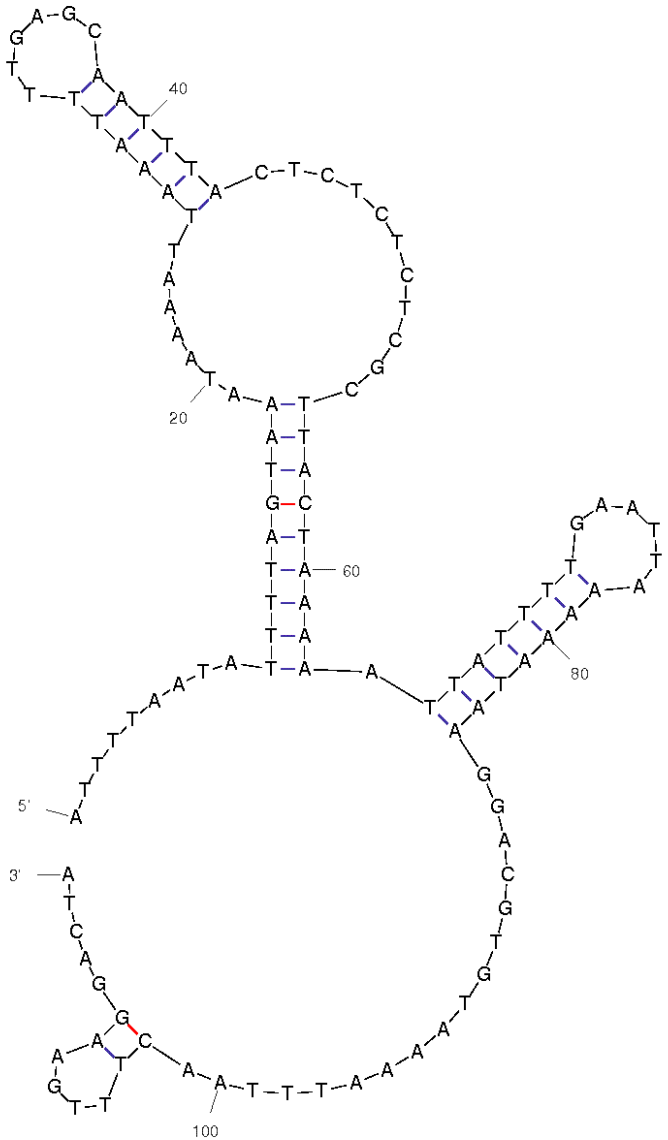
A. obtectus

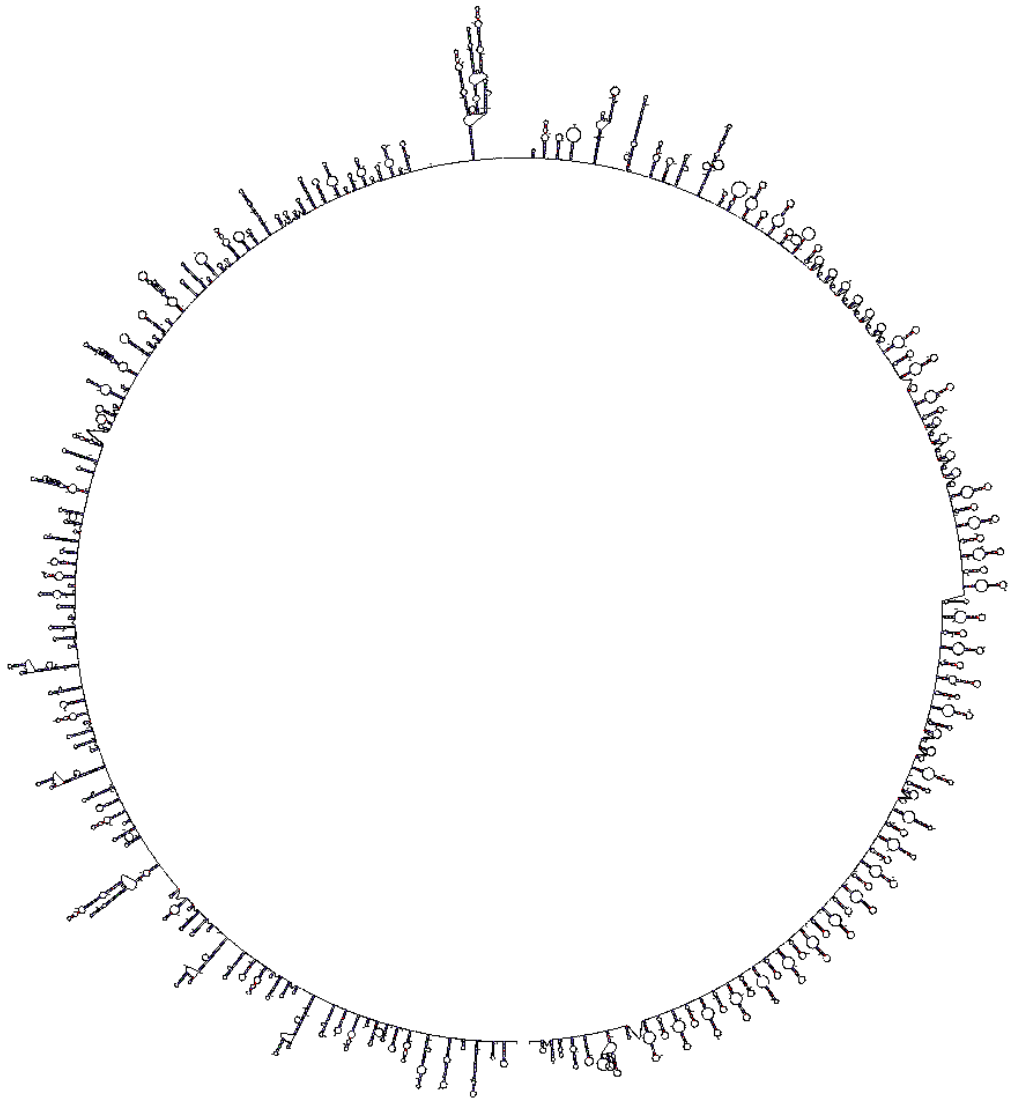


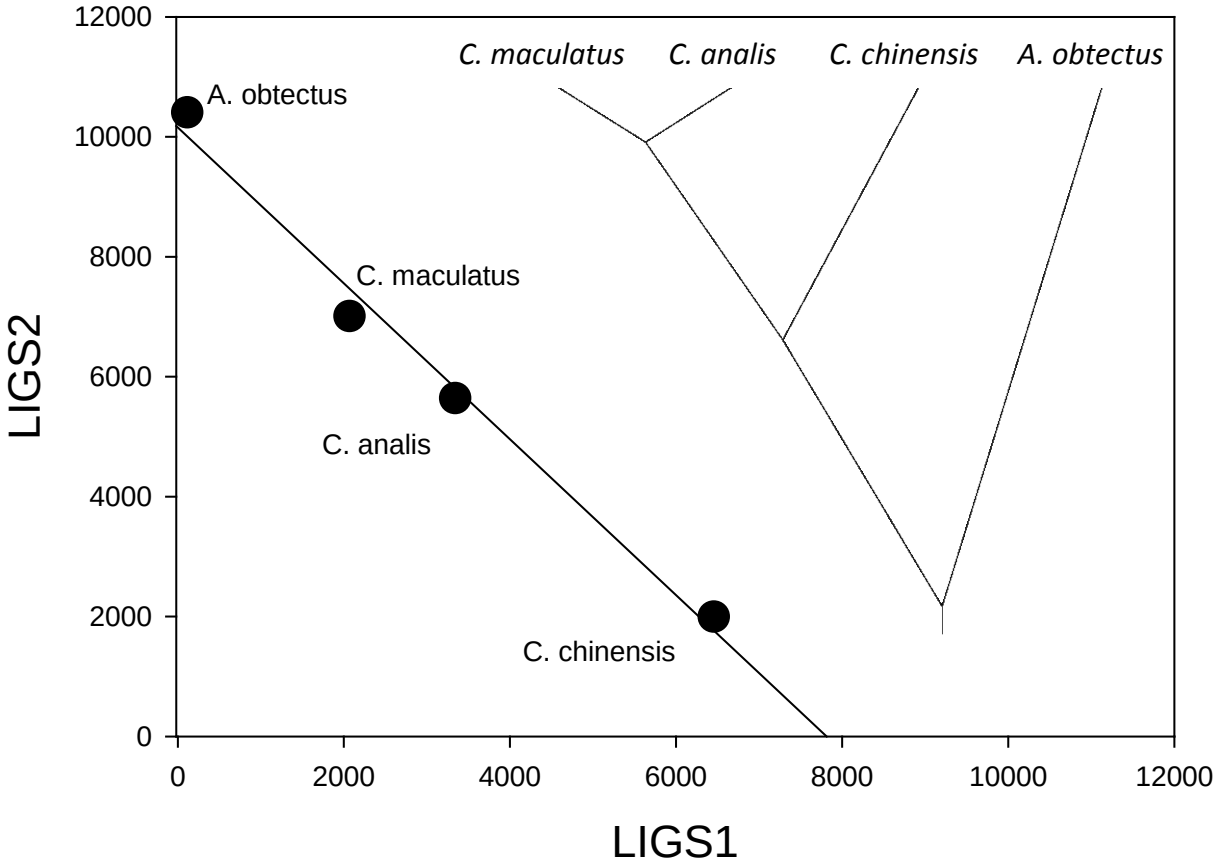


Panel H

A. obtectus







SI Figure 6. Plot of the relative length (bp) of the two LIGSs in the four seed beetle species studied. The size of LIGS1 and LIGS2 show a strong negative correlated evolution in seed beetles (Phylogenetic Least-Squares Regression; r = -0.99, P = 0.006). Moreover, the relatedness of the four species (phylogeny inserted) implies that the concerted evolution of LIGS1 and LIGS2 must have been bi-directional.

SI Table 1. Mitogenome maps of the four seed beetle species *C. maculatus*, *C. analis*, *C. chinensis* and *A. obtectus*.

Gene	Direction	Position				Anticodon	Start codon	Stop codon
		C. mac	***C. ana***	***C. chi***	***A. obt***			
trnI(gat)	+	1-66 (+32)	1-65 (+45)	1-65 (+73)	1-67 (+6)	GAT		
trnQ(ttg)	-				74-143 (-1)	TTG		
trnM(cat)	+	99-166 (0)	111-178 (0)	139-207 (0)	143-211 (0)	CAT		
nad2	+	167-1180 (0)	179-1192 (0)	208-1230 (0)	212-1222 (0)		ATC / ATT	TAA
LGIS1		1181-3246 (0)	1193-4533 (0)	1231-7685 (0)	1223-1336 (0)			
trnW(tca)	+	3247-3315 (-8)	4534-4601 (-8)	7686-7754 (-8)	1337-1405 (+32)	TCA		
trnC(gca)	-	3308-3376 (+33)	4594-4664 (+4)	7747-7811 (+8)	1438-1501 (0)	GCA		
trnY(gta)	-	3410-3475 (+9)	4669-4735 (+1)	7820-7886 (+1)	1502-1568 (+1)	GTA		
cox1	+	3485-5017 (+6)	4737-6269 (+1)	7888-9420 (0)	1570-3105 (-5)		AAT	TAA
trnL2(taa)	+	5024-5088 (0)	6271-6335 (0)	9421-9485 (0)	3101-3165 (0)	TAA		
cox2	+	5089-5776 (0)	6336-7023 (0)	9486-10173 (0)	3166-3853 (0)		ATT / ATC	T
trnK(ttt)	+	5777-5846 (+6)	7024-7094 (+4)	10174-10243 (-1)	3854-3923 (+4)	TTT		
trnD(gtc)	+	5853-5920 (0)	7099-7165 (0)	10243-10308 (0)	3928-3992 (0)	GTC		
atp8	+	5921-6079 (-7)	7166-7324 (-7)	10309-10467 (-7)	3993-4148 (-7)		ATT / ATA	TAA
atp6	+	6073-6747 (-1)	7318-7992 (-1)	10461-11132 (-1)	4142-4816 (+2)		ATG	TAA
cox3	+	6747-7535 (+5)	7992-8780 (+19)	11132-11920 (+5)	4819-5607 (+3)		ATG	TAA
trnG(tcc)	+	7541-7605 (0)	8800-8866 (0)	11926-11992 (0)	5611-5677 (0)	TCC		
nad3	+	7606-7962 (-2)	8867-9220 (+2)	11993-12349 (0)	5678-6031 (-1)		ATA / ATT / ATG	TAG / TAA
trnA(tgc)	+	7961-8026 (-1)	9223-9287 (+3)	12350-12414 (+2)	6031-6098 (+58)	TGC		
trnR(tcg)	+	8026-8092 (-1)	9291-9358 (-1)	12417-12482 (-1)	6157-6224 (+11)	TCG		
trnN(gtt)	+	8092-8156 (0)	9358-9422 (0)	12482-12547 (0)	6236-6300 (0)	GTT		
trnS1(tct)	+	8157-8222 (+5)	9423-9488 (+6)	12548-12614 (+2)	6301-6367 (0)	TCT		
trnE(ttc)	+	8228-8293 (+4)	9495-9559 (+38)	12617-12681 (-2)	6368-6436 (-2)	TTC		
trnF(gaa)	-	8298-8364 (0)	9598-9664 (0)	12680-12745 (0)	6435-6502 (-3)	GAA		
nad5	-	8365-10078 (-3)	9665-11378 (-3)	12746-14459 (-3)	6500-8213 (-3)		ATT / ATA	T
trnH(gtg)	-	10076-10141 (0)	11376-11445 (+3)	14457-14521 (0)	8211-8276 (0)	GTG		
nad4	-	10142-11471 (-7)	11449-12780 (-7)	14522-15851 (-7)	8277-9606 (-7)		ATG	T / TAG

nad4l	-	11465-11743 (+43)	12774-13052 (+9)	15845-16123 (+9)	9600-9884 (+3)		ATG	TAA
trnT(tgt)	+	11787-11850 (0)	13062-13127 (0)	16133-16197 (0)	9888-9953 (0)	TGT		
trnP(tgg)	-	11851-11917 (+2)	13128-13193 (+2)	16198-16263 (+2)	9954-10019 (+2)	TGG		
nad6	+	11920-12429 (-1)	13196-13702 (-1)	16266-16775 (-1)	10022-10528 (-1)		ATC / ATT	TAA
cob	+	12429-13565 (0)	13702-14838 (-2)	16775-17914 (+12)	10528-11670 (-2)		ATG	TAA / TAG
trnS2(tga)	+	13566-13633 (+93)	14837-14903 (+24)	17927-17994 (+18)	11669-11735 (0)	TGA		
trnQ(ttg)	-	13727-13795 (0)	14928-14996 (0)	18013-18081 (0)		TTG		
LIGS2		13796-20803 (0)	14997-20636 (0)	18082-20076 (0)	11736-22143 (0)			
nad1	-	20804-21754 (+1)	20637-21581 (+1)	20077-21027 (+1)	22144-23094 (+1)		TTG	TAA / TAG
trnL1(tag)	-	21756-21821 (+4)	21583-21647 (0)	21029-21093 (+1)	23096-23161 (+17)	TAG		
rrnL	-	21826-23136 (0)	21648-22956 (0)	21095-22416 (+1)	23179-24462 (+2)			
trnV(tac)	-	23137-23205 (0)	22957-23027 (-1)	22418-22486 (-1)	24465-24533 (-1)	TAC		
rrnS	-	23206-23981 (0)	23027-23809 (0)	22486-23267 (0)	24533-25307 (0)			
CR		23982-25011 (0)	23810-24832 (0)	23268-24496 (0)	25308-26613 (0)			
mt DNA		25011	24832	24496	26613			

SI Table 2. AT content by regions in the mitogenomes of *A. obtectus*, *C. analis, C. chinensis* and the four populations of *C. maculatus*.

	Control region		LIGS1		LIGS2		Coding regions		mt DNA	
	Size	**AT%**	**Size**	**AT%**	**Size**	**AT%**	**Size**	**AT%**	**Size**	**AT%**
A .obt	1306	83.2	114	79.8	10408	80.5	14676	75.7	26613	78
C.ana	1023	82.5	3341	81.83	5640	85.69	14699	75.86	24832	79.21
C.chi	1229	81.45	6455	79.35	1995	86.72	14713	76.04	24496	77.98
***C.mac* SI**	1030	79.22	2066	87.37	7008	76.76	14694	74.9	25011	76.69
***C.mac* Bra**	1032	78.97	2060	86.51	6947	75.67	11136	73.55	24947	76.33
***C.mac* Cal**	1032	78.97	2068	87.33	7017	75.2	11136	73.52	25026	76.25
***C.mac* Yem**	1031	79.05	2072	87.55	7059	76.53	11136	73.42	25069	76.73

SI Table 3. List of tandem repeat units found in the mitogenomes of *C. maculatus, C. analis, C. chinensis and A. obtectus.* Default score for minimum alignment to report repeats was 50.

C. mac:

	Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)	Gene
1	922--2010	372	2.9	372	88	1	1493	36	12	5	45	1.66	Nad2-LIGS1
2	2133--2542	147	2.8	144	91	4	650	46	5	4	44	1.45	LIGS1
3	2595--2627	12	2.5	13	90	9	50	42	0	0	57	0.98	LIGS1
4	2560--2655	32	3.1	31	70	15	83	47	1	2	48	1.2	LIGS1
5	2644--2737	18	5.4	17	77	11	100	51	1	1	46	1.15	LIGS1
6	2560--2683	60	2.1	59	80	8	135	48	0	2	48	1.2	LIGS1
7	2640--2737	35	2.7	35	82	11	110	52	1	1	45	1.14	LIGS1

8	2760--2895	70	1.9	72	87	3	204	52	0	3	42	1.24	LIGS1
9	2906--3175	97	3	86	78	16	273	51	4	0	43	1.27	LIGS1
10	3012--3122	21	5	21	69	20	64	48	3	0	47	1.19	LIGS1
11	3006--3086	24	3.4	24	75	11	85	49	3	0	46	1.19	LIGS1
12	3006--3058	23	2.2	25	90	3	67	49	1	0	49	1.12	LIGS1
13	13791--13892	2	51	2	90	0	159	49	3	0	47	1.2	LIGS2
14	14035--15539	82	18.4	82	98	0	2869	39	11	9	39	1.74	LIGS2
15	15552--20538	164	30.4	164	96	0	7944	37	12	12	37	1.81	LIGS2
16	20714--20759	2	23.5	2	87	12	69	47	2	0	50	1.13	LIGS2
17	23887--23926	15	2.6	15	84	4	53	47	5	2	45	1.38	rRNAs

C. ana:

	Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)	Gene
1	187--224	15	2.5	15	91	8	60	36	0	0	63	0.95	Nad2
2	1267--1992	131	5.5	131	99	0	1416	37	6	13	42	1.72	LIGS1
3	1267--3973	262	9.8	262	76	15	2594	38	6	12	42	1.68	LIGS1
4	2090--2553	131	3.5	131	99	0	910	37	6	14	42	1.71	LIGS1
5	1791--4503	430	6.3	430	87	7	3668	39	6	11	42	1.66	LIGS1
6	2651--2982	130	2.5	131	98	0	639	37	6	14	42	1.71	LIGS1
7	3080--3412	131	2.5	131	98	0	639	37	6	13	42	1.71	LIGS1
8	3510--3973	131	3.5	131	98	0	892	37	6	13	42	1.72	LIGS1

9	3518--4347	262	3	262	85	12	934	39	6	11	42	1.66	LIGS1
10	3772--4503	299	2.4	299	98	0	1428	39	5	11	43	1.65	LIGS1
11	9560--9589	2	15	2	100	0	60	50	0	0	50	1	IGS
12	12436--12460	11	2.3	11	100	0	50	84	0	0	16	0.63	Nad4
13	16200--16255	26	2.2	25	80	3	58	58	0	0	41	0.98	LIGS2
14	17739--17794	26	2.2	25	80	3	58	58	0	0	41	0.98	LIGS2
15	19792--19847	26	2.2	25	80	3	58	58	0	0	41	0.98	LIGS2
16	20411--20463	12	4.8	12	76	19	62	67	1	0	30	1.01	LIGS2
17	20411--20452	20	2.1	20	95	0	75	69	2	0	28	1.01	LIGS2
18	20476--20534	16	3.5	16	69	13	55	47	5	1	45	1.34	LIGS2

19	20450--20531	35	2.3	35	97	0	155	48	4	1	45	1.31	LIGS2

C. chi:

	Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)	Gene
1	1247--4079	103	27.6	103	99	0	5574	39	9	19	31	1.84	LIGS1
2	4077--7236	52	60.8	52	93	0	5116	40	5	8	45	1.59	LIGS1
3	7406--7436	2	16	2	93	6	55	48	0	0	51	1	LIGS1 end
4	12740--12785	18	2.3	20	78	10	51	58	4	2	34	1.3	tRNAF-Nad5
5	14422--14468	13	3.4	13	81	16	58	55	0	0	44	0.99	Nad5-start tRNAH
6	15507--15531	11	2.3	11	100	0	50	84	0	0	16	0.63	Nad4

7	18185--19683	53	28.7	53	86	4	1877	43	7	4	43	1.53	LIGS2
8	18239--18333	32	3.3	32	64	29	86	40	8	6	45	1.6	LIGS2
9	18185--19684	209	7.2	209	90	3	2211	43	7	4	43	1.53	LIGS2
10	19072--19166	32	3.3	32	64	29	86	40	8	6	45	1.6	LIGS2
11	19928--19957	2	15	2	92	0	51	50	3	0	46	1.18	LIGS2
12	20087--20130	20	2.2	20	83	0	52	65	2	0	31	1.05	Nad1
13	22020--22091	16	4.6	16	69	17	53	54	2	0	43	1.15	rRNAl

A. obt:	Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	**A**	**C**	**G**	**T**	Entropy (0-2)	Gene
1	7691--7725	18	1.9	18	94	0	61	48	0	2	48	1.16	NAD5
2	10129--10153	11	2.3	11	100	0	50	84	8	0	8	0.79	Nad6
3	12684--19	78	13.9	78	93	1	1786	31	8	3	56	1.44	LIGS2

	13768												
4	12684--13768	155	7	156	95	1	1804	31	8	3	56	1.44	LIGS2
5	13249--13298	13	3.5	14	78	10	57	36	2	2	60	1.2	LIGS2
6	13328--13378	13	3.6	14	78	10	59	37	1	1	58	1.2	LIGS2
7	13483--13533	13	3.6	14	78	10	59	37	1	1	58	1.2	LIGS2
8	13639--13688	13	3.5	14	78	10	57	36	2	2	60	1.2	LIGS2
9	14799--14846	19	2.5	19	77	12	53	33	2	2	62	1.18	LIGS2
10	15147--15194	19	2.5	19	77	12	53	33	2	2	62	1.18	LIGS2
11	15856--15903	19	2.5	19	77	12	53	33	2	2	62	1.18	LIGS2
12	16204--16251	19	2.5	19	77	12	53	33	2	2	62	1.18	LIGS2
13	16254--16314	32	1.9	32	83	12	81	44	1	4	49	1.33	LIGS2
14	17762--17810	20	2.5	20	76	6	55	32	2	2	63	1.17	LIGS2
15	17813--17871	31	1.9	31	86	6	84	44	1	5	49	1.34	LIGS2
16	17859--18270	46	9.2	46	61	21	137	37	4	6	51	1.47	LIGS2
17	18819--18961	46	3.1	46	67	16	143	38	2	3	55	1.32	LIGS2
18	18846--18992	46	3.2	46	67	16	133	38	4	4	53	1.39	LIGS2

19	13825--19185	90	60.8	90	83	8	6501	35	6	6	51	1.53	LIGS2
20	19285--22871	51	70.4	51	98	0	6770	25	21	11	41	1.87	LIGS2-Nad1
21	22967--23021	20	2.8	20	81	10	76	54	1	1	41	1.21	Nad1
22	25285--25326	21	2	22	90	4	68	50	2	0	47	1.14	rRNAs-CR

SI Table 4. Mitogenome maps of *C. maculatus* from three different populations (Brazil, California and Yemen).

Gene	Direction	Position			Anticodon	Start codon	Stop codon
		C. mac Bra	*C. mac Cal*	*C. mac Yem*		*C. mac Bra* *C. mac Cal* *C. mac Yem*	*C. mac Bra* *C. mac Cal* *C. mac Yem*
trnI (gat)	+	1-66 (32)	1-66 (32)	1-66 (32)	GAT		
trnM (cat)	+	99-166 (0)	99-166 (0)	99-166 (0)	CAT		
nad2	+	167-1180 (0)	167-1180 (0)	167-1180 (0)		ATC	TAA
LIGS1		1181-3241 (0)	1181-3249 (0)	1181-3253 (0)			
trnW (tca)	+	3242-3310 (-8)	3250-3318 (-8)	3254-3322 (-8)	TCA		
trnC (gca)	-	3303-3371 (33)	3311-3379 (33)	3315-3383 (33)	GCA		
trnY (gta)	-	3405-3470 (9)	3413-3478 (9)	3417-3482 (9)	GTA		
cox1	+	3480-5012 (6)	3488-5020 (6)	3492-5024 (6)		AAT	TAA
trnL2 (taa)	+	5019-5083 (0)	5027-5091 (0)	5031-5095 (0)	TAA		
cox2	+	5084-5771 (0)	5092-5779 (0)	5096-5783 (0)		ATT	T
trnK (ttt)	+	5772-5841 (6)	5780-5849 (6)	5784-5853 (5)	TTT		
trnD (gtc)	+	5848-5914 (0)	5856-5923 (0)	5859-5926 (0)	GTC		
atp8	+	5915-6073 (-7)	5924-6082 (-7)	5927-6085 (-7)		ATT	TAA
atp6	+	6067-6741 (-1)	6076-6750 (-1)	6079-6753 (-1)		ATG	TAA
cox3	+	6741-7529 (5)	6750-7538 (5)	6753-7541 (5)		ATG	TAA
trnG (tcc)	+	7535-7599 (0)	7544-7608 (0)	7547-7611 (0)	TCC		
nad3-0	+	7600-7956 (-2)	7609-7965 (-2)	7612-7968 (-2)		ATA	TAG
trnA (tgc)	+	7955-8020 (-1)	7964-8029 (-1)	7967-8032 (-1)	TGC		
trnR (tcg)	+	8020-8086 (-1)	8029-8095 (-1)	8032-8098 (-1)	TCG		
trnN (gtt)	+	8086-8150 (0)	8095-8159 (0)	8098-8162 (0)	GTT		
trnS1 (tct)	+	8151-8216 (5)	8160-8225 (5)	8163-8228 (5)	TCT		
trnE (ttc)	+	8222-8287 (4)	8231-8296 (4)	8234-8299 (4)	TTC		
trnF (gaa)	-	8292-8358 (0)	8301-8367 (0)	8304-8370 (0)	GAA		
nad5-0	-	8359-10072 (-3)	8368-10081 (-3)	8371-10084 (-3)		ATT	T

trnH (gtg)	-	10070-10135 (0)	10079-10144 (0)	10082-10147 (0)	GTG		
nad4	-	10136-11465 (-7)	10145-11474 (-7)	10148-11477 (-7)		ATG	T
nad4l-0	-	11459-11737 (43)	11468-11746 (43)	11471-11749 (43)		ATG	TAA
trnT (tgt)	+	11781-11844 (0)	11790-11853 (0)	11793-11856 (0)	TGT		
trnP (tgg)	-	11845-11911 (2)	11854-11920 (2)	11857-11923 (2)	TGG		
nad6	+	11914-12423 (-1)	11923-12432 (-1)	11926-12435 (-1)		ATC	TAA
cob	+	12423-13559 (0)	12432-13568 (0)	12435-13571 (0)		ATG	TAA
trnS2 (tga)	+	13560-13627 (93)	13569-13636 (93)	13572-13639 (92)	TGA		
trnQ (ttg)	-	13721-13789 (0)	13730-13798 (0)	13732-13800 (0)	TTG		
LIGS2		13790-20737 (0)	13799-20816 (0)	13801-20860 (0)			
nad1	-	20738-21688 (1)	20817-21767 (1)	20861-21811 (1)		TTG	TAA
trnL1 (tag)	-	21690-21755 (4)	21769-21834 (4)	21813-21878 (4)	TAG		
rrnL	-	21760-23070(0)	21839-23149(0)	21883-23193(0)			
trnV (tac)	-	23071-23139(0)	23150-23218(0)	23194-23262(0)	TAC		
rrnS	-	23140-23915(0)	23219-23994(0)	23263-24038(0)			
CR		23916-24947(0)	23995-25026(0)	24039-25069(0)			
mt DNA		1-24947	1-25026	1-25069			

SI Table 5. List of tandem repeat units found in in the mitogenomes of three different in *C. maculatus* populations (Brazil, California and Yemen). Default score for minimum aLIGSnment to report repeats was 50.

C. maculatus Brazil

	Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)	Gene
1	1054--1985	374	2.5	374	95	1	1622	36	12	6	44	1.68	Nad2-LIGS1
2	2104--2552	146	3.1	146	91	3	672	45	5	5	43	1.5	LIGS1
3	2600--2632	12	2.5	13	90	9	50	42	0	0	57	0.98	LIGS1
4	2610--2666	11	5.1	12	75	16	59	45	0	3	50	1.18	LIGS1
5	2649--2742	18	5.4	17	76	11	91	51	2	1	45	1.2	LIGS1
6	2600--2684	21	4	20	76	11	71	47	0	2	50	1.14	LIGS1
7	2647--2742	35	2.8	33	80	10	102	52	2	1	44	1.19	LIGS1
8	2765--2900	70	1.9	72	89	3	213	54	0	2	42	1.18	LIGS1
9	2952--3173	94	2.5	92	84	10	259	51	4	0	42	1.27	LIGS1
10	13760--13947	93	2	93	100	0	376	40	9	2	47	1.48	end tRNAQ-LIGS2
11	14045--15554	82	18.4	82	99	0	2993	37	12	10	39	1.78	LIGS2
12	15541--15720	81	2.3	80	86	3	256	40	15	10	34	1.8	LIGS2
13	15567--15738	81	2.1	81	93	4	296	37	13	13	35	1.85	LIGS2
14	15673--15803	65	2	65	96	0	244	32	10	16	39	1.83	LIGS2
15	15738--16216	83	5.8	83	90	3	788	34	11	14	39	1.82	LIGS2
16	15886--15921	18	2	18	100	0	72	22	11	5	61	1.5	LIGS2
17	15738--20467	164	29.3	164	89	6	6499	36	12	13	38	1.82	LIGS2
18	16151--16281	65	2	65	100	0	262	32	10	16	39	1.83	LIGS2
19	16216--19234	83	36.8	83	88	5	4365	36	12	13	37	1.82	LIGS2
20	16546--16689	65	2.2	65	96	2	263	32	11	14	41	1.81	LIGS2
21	18177--18208	16	2	16	100	0	64	25	6	12	56	1.59	LIGS2
22	18179--18215	16	2.5	15	91	8	58	29	2	10	56	1.47	LIGS2
23	19169--19299	65	2	65	98	0	253	33	10	16	39	1.83	LIGS2

24	19234--19546	83	3.8	83	91	1	522	36	12	13	38	1.82	LIGS2
25	19481--19611	65	2	65	92	0	217	34	10	16	38	1.83	LIGS2
26	19629--19759	65	2	65	97	2	246	32	10	16	39	1.83	LIGS2
27	19694--20087	83	4.8	83	89	3	602	36	12	13	37	1.82	LIGS2
28	20022--20152	65	2	65	92	0	217	32	9	16	40	1.82	LIGS2
29	20077--20467	167	2.3	166	92	2	669	35	10	14	39	1.81	LIGS2
30	20087--20467	83	4.6	83	92	3	649	35	10	14	39	1.81	LIGS2
31	20640--20693	2	27.5	2	89	10	85	48	1	0	50	1.11	LIGS2
32	22913--22966	24	2.2	24	77	9	56	53	18	0	27	1.45	rRNAl
33	23821--23860	15	2.6	15	88	4	62	50	5	0	45	1.23	rRNAS

C. maculatus California

	Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)	Gene
1	936--1986	374	2.8	374	89	1	1507	36	12	5	45	1.66	Nad2-LIGS1
2	2241--2295	10	5.6	10	77	18	53	45	1	1	50	1.22	LIGS1
3	2131--2557	149	2.9	149	92	2	682	46	5	5	43	1.48	LIGS1
4	2605--2637	12	2.5	13	90	9	50	42	0	0	57	0.98	LIGS1
5	2615--2671	11	5.1	12	75	16	59	45	0	3	50	1.18	LIGS1
6	2654--2747	18	5.4	17	76	11	91	51	2	1	45	1.2	LIGS1
7	2605--2689	21	4	20	76	11	71	47	0	2	50	1.14	LIGS1
8	2652--2747	35	2.8	33	80	10	102	52	2	1	44	1.19	LIGS1
9	2770--2905	70	1.9	72	89	3	213	54	0	2	42	1.18	LIGS1
10	2845--3052	111	1.9	112	86	4	303	50	1	0	46	1.19	LIGS1
11	2916--3178	90	3	86	83	10	313	51	4	0	43	1.26	LIGS1
12	2976--3124	21	6.8	22	71	17	89	51	4	0	44	1.2	LIGS1
13	3001--3054	24	2.2	25	81	15	69	48	1	0	50	1.11	LIGS1
14	2979--3124	45	3.3	45	77	12	160	51	4	0	44	1.2	LIGS1
15	2961--3222	81	3.1	81	80	8	211	51	4	1	43	1.28	LIGS1
16	13758--13958	91	2.2	91	95	2	363	40	9	2	47	1.49	LIGS2

17	13958--14021	34	1.9	32	81	12	76	48	4	0	46	1.23	LIGS2
18	13957--14028	34	2.2	32	78	9	74	50	4	0	45	1.21	LIGS2
19	14050--15561	82	18.4	82	98	0	2943	38	12	10	38	1.77	LIGS2
20	14050--15561	493	3.1	493	99	0	2954	38	12	10	38	1.77	LIGS2
21	15574--20543	164	30.4	164	96	0	8148	36	12	13	37	1.83	LIGS2
22	15574--20543	409	12.1	409	95	1	8077	36	12	13	37	1.83	LIGS2
23	20733--20772	2	20.5	2	85	14	57	47	2	0	50	1.14	LIGS2
24	23900--23939	15	2.6	15	88	4	62	50	5	0	45	1.23	rRNAS

C. maculatus Yemen

	Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0-2)	Gene
1	922--2015	375	2.9	375	87	2	1498	35	12	6	45	1.68	Nad2-LIGS1
2	2246--2300	11	5.6	10	79	14	53	47	1	1	49	1.23	LIGS1
3	2135--2557	148	2.9	147	91	2	656	46	4	4	43	1.45	LIGS1
4	2605--2637	12	2.5	13	90	9	50	42	0	0	57	0.98	LIGS1
5	2615--2671	12	5.1	12	71	16	50	45	1	3	49	1.29	LIGS1
6	2654--2747	18	5.3	18	79	11	106	52	1	0	46	1.07	LIGS1
7	2652--2747	35	2.8	33	80	10	111	53	1	0	45	1.07	LIGS1
8	2770--2905	70	1.9	72	87	3	204	53	0	2	42	1.21	LIGS1
9	2845--3056	111	1.9	115	86	5	306	50	1	0	46	1.15	LIGS1
10	3022--3129	21	5	21	69	18	69	49	3	0	47	1.19	LIGS1
11	2961--3182	81	2.6	82	85	10	275	51	4	0	43	1.25	LIGS1
12	3054--3226	81	2.1	81	78	4	179	50	4	1	42	1.33	LIGS1
13	13796--13827	2	16	2	100	0	64	50	0	0	50	1	end tRNAQ-LIGS2
14	13914--13954	2	21	2	95	5	75	51	0	0	48	1	LIGS2
15	13769--14018	127	2	128	98	0	484	43	6	0	49	1.34	LIGS2
16	14007--14070	34	1.9	32	81	12	76	48	4	0	46	1.23	LIGS2
17	14006--14077	34	2.2	32	78	9	74	50	4	0	45	1.21	LIGS2
18	14100--15604	82	18.4	82	97	0	2772	40	12	9	38	1.75	LIGS2

19	14100--15604	245	6.1	246	96	0	2779	40	12	9	38	1.75	LIGS2
20	15617--20601	81	60.9	81	93	2	8100	36	12	13	37	1.82	LIGS2
21	15617--20601	164	30.4	164	95	1	8253	36	12	13	37	1.82	LIGS2
22	20777--20816	2	20.5	2	85	14	57	47	2	0	50	1.14	LIGS2
23	23944--23983	15	2.6	15	88	4	62	50	5	0	45	1.23	rRNAS

SI Table 6. Estimates of the relative rates of non-synonymous to synonymous substitutions (ω) and nucleotide diversities in the coding regions of the seed beetles mitogenomes. Given are analyses of the three congeneric *Callosobruchus* species, all four species and the four populations of *C. maculatus*. *P*-values refer to LLR tests of a single versus three distinct values of ω within genes (i.e., a model M0 – M3 comparison).

Gene	**ω (model M0)**	**SE**	**P**	**No codons**	π_S	π_N	π_N/π_S
Interspecific variation (three species):							
ATP6	0.055	0.011	0.000	224	0.511	0.066	0.129
ATP8	0.380	0.149	0.590	52	0.349	0.203	0.582
COB	0.048	0.007	0.000	378	0.598	0.068	0.114
COX1	0.017	0.003	0.006	510	0.616	0.031	0.050
COX2	0.038	0.007	0.003	229	0.592	0.052	0.088
COX3	0.048	0.009	0.038	262	0.536	0.054	0.101
NAD1	0.060	0.010	0.000	316	0.460	0.061	0.133
NAD2	0.134	0.016	0.000	337	0.452	0.111	0.246
NAD3	0.123	0.027	0.000	118	0.578	0.124	0.215
NAD4	0.092	0.012	0.000	443	0.462	0.076	0.165
NAD4L	0.126	0.039	0.981	92	0.429	0.092	0.214
NAD5	0.105	0.012	0.000	571	0.448	0.097	0.217
NAD6	0.102	0.019	0.000	169	0.589	0.146	0.248
*Interspecific variation (four species):**							
ATP6	0.072	0.010	0.000	224	0.533	0.093	0.175
ATP8	0.379	0.118	0.001	52	0.387	0.227	0.588
COB	0.055	0.006	0.000	378	0.605	0.076	0.125
COX1	0.025	0.003	0.000	510	0.617	0.042	0.067
COX2	0.050	0.011	0.000	229	0.580	0.064	0.110
COX3	0.060	0.008	0.000	262	0.582	0.076	0.130
NAD2	0.172	0.016	0.000	337	0.443	0.163	0.368
NAD3	0.119	0.020	0.000	118	0.590	0.137	0.233
NAD6	0.124	0.019	0.000	169	0.574	0.207	0.361

Intraspecific variation (four populations):						
ATP6	0.038	0.043	224	0.042	0.001	0.024
ATP8	0.000	0.547	52	0.050	-	-
COB	0.042	0.025	378	0.071	0.001	0.014
COX1	0.020	0.015	510	0.062	0.001	0.016
COX2	0.038	0.028	229	0.062	0.002	0.032
COX3	0.039	0.041	262	0.031	0.001	0.032
NAD1	0.048	0.035	316	0.039	0.001	0.026
NAD2	0.179	0.134	337	0.017	0.002	0.118
NAD3	0.000	1.387	118	0.036	-	-
NAD4	0.041	0.029	443	0.044	0.001	0.023
NAD4L	0.000	0.722	93	0.035	-	-
NAD5	0.138	0.049	571	0.040	0.003	0.075
NAD6	0.000	3.184	169	0.014	-	-

* For four genes, efforts to align the sequences of all four species failed to produce a single, unambiguous and fully reliable alignment. These genes were thus excluded from this part of the analysis.

SI Table 7. Analyses of variance of variation in transcript abundance (FPKM) of the 13 mtDNA PCGs across sexes, tissue (head/thorax vs. abdomen) and mating status (i.e., virgin vs. mated). The effect sizes of single factors are here given as *F*–ratios for all genes as well as for a test of the entire model (bottom). Sex and tissue type, as well as their interaction, both had major effects on mtDNA transcript abundance. The most differentially abundant genes included COX1, COX3 and COB, while the least differentially expressed genes included NAD2, NAD4L and NAD6. Critical *F*–ratio for P<0.001 is $F_{7,17}$ = 6.22.

Source ($F_{1,17}$)	**NAD2**	**COX1**	**COX2**	**ATP8/6**	**COX3**	**NAD3**	**NAD5**	**NAD4**	**NAD4L**	**NAD6**	**COB**	**NAD1**
Sex	15.6	429.4	237.0	241.6	343.0	160.5	187.4	227.4	61.0	38.0	432.5	119.3
Tissue	71.4	1201.9	387.5	795.8	744.4	188.3	393.8	652.6	137.5	21.8	843.8	229.0
Mating	2.6	6.4	5.5	5.1	11.5	13.1	9.3	10.9	2.2	0.9	9.2	4.4
Mating × Sex	0.6	1.6	5.3	2.4	4.2	0.8	2.1	0.5	1.1	4.2	1.8	1.2
Tissue × Sex	0.8	44.7	13.8	26.1	23.9	3.7	11.5	28.6	7.5	1.5	23.5	7.5
Mating × Tissue	4.0	5.4	4.9	4.1	8.3	4.2	7.5	11.5	3.8	0.1	6.4	5.1
Entire model ($F_{7,17}$)	53.4	832.3	422.7	499.1	615.0	337.4	430.7	496.2	157.1	43.0	783.6	283.1