Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Two deeply divergent mitochondrial genomes in a single Euphaedusa aculus individual: a case of deep mitochondrial coalescence

Zhi-Jie Xu

zjxmolls@outlook.com

Donghua University https://orcid.org/0009-0003-3855-0921

Research Article

Keywords: mitochondrial heteroplasmy, deep coalescence, incomplete lineage sorting, DNA barcoding, Clausiliidae, Euphaedusa aculus

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11736846/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Two deeply divergent mitochondrial genomes in a single *Euphaedusa aculus* individual: a case of deep mitochondrial coalescence

Zhi-Jie Xu[1]

[1]**College of Biological Science and Medical Engineering, Donghua University, Shanghai, China**

ABSTRACT

We report two complete mitochondrial genomes coexisting in a single individual of *Euphaedusa aculus* (Benson, 1842), the type species of *Euphaedusa*, from Xixi Wetland, Hangzhou, China. The two genomes differ by 5.15% at COI and 11.0% at 16S and occur at a ratio of 42:58. We designate them lineage α (the published haplotype, PP485172) and lineage β (the entity deposited in INSDC under the name "*E. planostriata*"). Read-level analyses exclude NUMT origin, cross-sample contamination and assembly artefact. An independent 45S rDNA anchor and the H3–5S–H4 repeat family place the mitochondrial pool within the range of a single-lineage individual. New reference material of true *E. planostriata* and public barcodes place α and β at 0% ITS divergence from each other and 2.4–3.5% from *E. planostriata*, and show both lineages across four provinces, including one Nanjing locality where both were collected on the same day. The pattern is most simply interpreted as deep coalescence — incomplete lineage sorting — rather than recent introgression. Because a single specimen can carry two divergent mitogenomes, mitochondrial barcodes alone can misassign species boundaries. The core observation rests on one individual; population frequencies remain to be estimated.

Keywords: mitochondrial heteroplasmy; deep coalescence; incomplete lineage sorting; DNA barcoding; Clausiliidae; *Euphaedusa aculus*

INTRODUCTION

Multilocus systematic work on Clausiliidae, particularly in East Asian and Mediterranean taxa, has repeatedly uncovered mito-nuclear incongruence, including non-monophyly of morphologically defined genera and parallel evolution of shell characters (Motochin et al., 2017; Hwang et al., 2022; Mason et al., 2020; Chiba, 2005; Sauer and Hausdorf, 2012). Such discordance is usually attributed to mitochondrial introgression or incomplete lineage sorting (ILS). Under the multispecies coalescent (MSC), ILS is a major source of gene-tree discordance: viewed backwards in time, gene lineages coalesce in ancestral populations, and some coalescences predate the speciation events that separate the sampled species; stochastic sorting of ancestral polymorphisms can leave gene trees conflicting with the species tree. Recent arthropod genomic studies show that ILS and introgression can obscure species boundaries and cryptic diversity (Trillo and Ortego, 2026; Wang et al., 2026, 2025; Greenwood et al., 2025; Zhao et al., 2025; Ntatsopoulos et al., 2026; Noguerales and Emerson, 2025).

Deep coalescence is the same process viewed backwards in time: sampled gene lineages coalesce in an ancestral population, before the forward-time split of the descendant species. When such coalescence is deep, divergent mitochondrial genomes can persist in a current population. Within a single individual, this appears as heteroplasmy or coexistence of divergent mitogenomes. Across populations, the same process can leave divergent mitochondrial lineages segregating at high frequency, where barcode-based species delimitation may mistake them for cryptic species or for recent introgression. Such variation is well documented among individuals of freshwater snails — for example, the D and N mitotypes of *Physa acuta* (David et al., 2022) — but two complete mitochondrial genomes within one individual are rarely observed. The best-documented cases are the doubly uniparental inheritance system of bivalves (Zouros, 2013; Ghiselli et al., 2021), where the phenomenon is developmentally programmed, and a single tuatara in which two mitogenomes 10.4% apart were found at a 1:7 ratio (Macey et al., 2021). No comparable case has been reported from gastropods.

Here we report such a case in *Euphaedusa aculus*, the type species of *Euphaedusa*. A published genome-skimming library of voucher Eacu201601 (Xixi Wetland, Hangzhou) contains two complete mitochondrial genomes, α and β, which differ by 5.15% at COI and 11.0% at 16S and occur at 42:58. α corresponds to the published mitogenome PP485172. β is the mitogenome previously deposited under the name "*E. planostriata*" from a different individual from Nanchang (NC_056972; SRR14226860); reference material shows that true *E. planostriata* is a distinct taxon, so this individual is a misidentified *E. aculus* carrying β. To our knowledge, this is the first documented case in a gastropod of two complete, deeply divergent mitogenomes coexisting within one individual.

We (i) validate their coexistence at the read level and quantify their ratio; (ii) test alternative explanations, including doubly uniparental inheritance, mitochondrial introgression, NUMTs and cross-sample contamination; (iii) place the mitochondrial pool against an independent 45S rDNA anchor and the H3–5S–H4 repeat family; (iv) compare the two lineages with reference material of true *E. planostriata* and with public barcodes; and (v) reassign public mitochondrial records to α or β and express their divergence relative to the shallowest congeneric species-pair distances. These analyses describe a case of deep coalescence: two mitochondrial lineages whose common ancestor predates the sampled population and which still coexist in one individual. They also illustrate why mitochondrial barcodes can misassign species boundaries when such lineages coexist.

MATERIALS AND METHODS

2.1 Specimens, sequencing data and new reference sequences

The primary dataset is the published whole-genome sequencing library of *Euphaedusa aculus* voucher Eacu201601 (Liang et al., 2024) (SRA: SRR28286884; BioSample SAMN40374704; leg. 2016-01, foot tissue, Xixi Wetland, Hangzhou, Zhejiang, China; MiSeq, PCR-amplified library, $2\times$ $\sim$150 bp cleaned reads; 13.3 M read pairs, 1.93 Gbp). A second library, voucher 20-NCU-SOUY-ZY01 (SRR14226860; leg. 2019-07, damp rocks, Nanchang, Jiangxi; NovaSeq 6000, RANDOM-selection library, 34.1 M read pairs, 5.11 Gbp), submitted as "*Euphaedusa planostriata*", was re-analysed as a comparative individual. Reads were downloaded from ENA over HTTPS; public barcodes and reference loci were retrieved from INSDC through NCBI Entrez and HTTP (accessions in Table S1).

New reference sequences of *E. planostriata* were generated from a specimen collected in 2024 by Zi-An Qiu at Jilongshan Forest Park, Ma'anshan, Anhui, China (currently held as a private voucher; public deposition pending). A 485-bp fragment of mitochondrial 16S rRNA was amplified with universal primers 16sAR/16sBR (Palumbi et al., 1991) and an 815-bp nuclear fragment spanning the ITS region with primers LSU-1/LSU-3 (Wade and Mordan, 2000). Both strands were Sanger-sequenced and assembled in Vector NTI 6.0 ContigExpress.

2.2 Mitogenome assembly and annotation

Haplotypes were discovered by de novo local assembly: raw reads of Eacu201601 were quality-filtered with fastp 0.23.2 (Chen et al., 2018) using `--length_required 50` and down-sampled to the first 6.67 M read pairs per file (`--reads_to_process 6,666,666`; 13.32 M reads / 1.99 Gbp retained), then assembled with MEGAHIT v1.2.9 (Li et al., 2015) using `--k-list 43,71,99`. Putative mitochondrial contigs were identified by tiara 1.0.3 sequence classification (Karlicki et al., 2022) and HMMER 3.1b2 profile searches (Finn et al., 2011), yielding four mitochondrial contigs (29,697 bp) that resolve into two haplotypes. All read-level quantifications used this uniform sub-sample.

To close both haplotypes into circular molecules, a supplementary targeted re-assembly was performed: mitochondrial reads were recalled by dual-reference mapping (BWA-MEM v0.7.17 against PP485172.1 and NC_056972; strictly paired mapped reads extracted with samtools 1.15.1 (Li et al., 2009) and name-deduplicated; 5,424 pairs), re-assembled with SPAdes v3.15.4 (Bankevich et al., 2012) with `--sc` for the assembly graph and `--isolate` for contigs, and the resulting assembly graph was walked in Bandage 0.8.1 (Wick et al., 2015). Node order followed the two published genomes; lineage at graph forks was decided by node read depth (α $\sim$20$\times$ vs β $\sim$34$\times$), and control-region closure followed single graph edges (α: 1157- $\rightarrow$ 2086-; β: 1988- $\rightarrow$ 2024- $\rightarrow$ 21-). Neither final ring contains reference sequence (α 15,417 bp; β 15,110 bp). For the NCU library, reads were adapter- and quality-trimmed with fastp 0.23.2 (PE adapter auto-detection, poly-G trimming as enabled by default for two-colour chemistry).

Protein-coding genes were annotated by tBLASTn (BLAST+ 2.17.0; Camacho et al., 2009; `-db_gencode 5`) against the 13 proteins of PP485172.1; rRNA and tRNA genes were annotated by BLASTn against

reference *Euphaedusa*/*Albinaria* sequences, with boundaries verified manually against alignments. tRNA genes were cross-checked with ARWEN 1.2 (Laslett and Canbäck, 2008). Linear gene maps of the two annotated rings (Fig. 1D) were drawn with gbdraw 0.14.2 (Kawato, 2026).

2.3 Read-level validation of coexistence and lineage ratio

Cleaned reads were aligned to a dual reference (α mitogenome + β mitogenome) with BWA-MEM v0.7.17 (Li, 2013) and MAPQ $\geq$ 30; per-reference depth profiles were inspected for uniformity and gaps. The lineage ratio was estimated two ways: (i) per-reference read counts; (ii) allele depths at 1,133 lineage-diagnostic fixed SNPs (extracted from whole-genome α–β alignments), piled up by CIGAR-aware traversal of mapped reads. The two estimates were 0.58 and 0.579 $\pm$ 0.105; we report 42:58 as the consensus estimate. Coverage uniformity, insert-size consistency, edit distance (NM $\leq$ 2) and dual-reference re-mapping of the NCU library provided the read-level diagnostics.

2.4 Intraindividual haplotype structure

Allele frequencies at two homopolymer tracts (ND4, ND6) and at all 28 A/T runs $\geq$7 bp were quantified by three-allele 31-mer junction probes (full-length, -1, $+1$), counting reads carrying $\geq$5 probe hits in both read directions.

2.5 Mitochondrial copy number from nuclear repeat-family anchors

Nuclear repeat-family anchors were assembled from the Eacu library with GeneMiner 2 (Yu et al., 2026) using `MainFilterNew -m 4` and `main_assembler -ka 39 -sb -1`, with *Tauphaedusa*/*Orcula* reference seeds. The 45S rDNA cluster (18S / 5.8S–ITS2–28S) is physically independent of the 5S and H3–H4 arrays. In Phaedusinae, 5S rDNA occurs both in the canonical H3>–H4> tandem repeat and in a chimeric <H3–5S–H4> repeat (Xu, 2026), so 5S and H3–H4 were treated as a combined repeat family rather than as independent anchors. Short-insert read pairs confirmed that the 45S sub-loci are co-arrayed (18S–5.8S–ITS2–28S; 94/72 pairs in Eacu, 246/210 in NCU), whereas 5S–H3 pairs were rare (1 in Eacu, 2 in NCU), consistent with the combined repeat unit being longer than the $\sim$217 bp insert. Relative copy number was estimated as mean k-mer depth per locus (25-mer, both libraries) and as MAPQ-filtered per-reference read depth (BWA). Ratios were computed within each library and compared between Eacu201601 and the NCU individual.

2.6 Phylogenetic placement and taxonomic triangulation

Pairwise distances (uncorrected p, local alignment; MAFFT, Katoh and Standley, 2013, for multi-sequence cross-checks) were computed among α, β, *E. planostriata* (new 16S), and congeneric reference loci (Motochin et al., 2017) cytb–ND4L–ND1 fragments; ITS cluster; *Tauphaedusa* and *Albinaria* outgroups). Relative divergence depth is expressed as the ratio of α–β distance to the shallowest congeneric species-pair distance per marker. Maximum-likelihood phylogenies were inferred with IQ-TREE v3.0.1 (Wong et al., 2026) under automatic model selection (BIC) with 1,000 ultrafast bootstrap replicates: a COI-barcode window (TIM2+F+I), the cytb–ND4L–ND1 block as three charsets (K3Pu+F+G4), the 16S conserved core (HKY+F), and the nuclear ITS cluster (TPM3u+F); trees are rooted on *E. steetzneri* / *Albinaria* outgroups (ITS panel on the most divergent congeners) and reported in Fig. 3. Nuclear ITS clusters of both the Hangzhou (mixed) and Nanchang (β-only) individuals were assembled de novo from raw reads by reference-guided k-mer filtering and assembly (GeneMiner 2; *Motochin* ITS references; k = 39).

2.7 Geographic synthesis

All INSDC records labelled "*E. aculus*" or "*E. planostriata*" were re-assigned to lineages by diagnostic k-mers and pairwise distance, and combined with collection metadata (BioSample fields) into a four-province presence table.

2.8 Data availability

All analysis scripts, intermediate tables and probe sets, the annotated mitogenomes, the SPAdes assembly graph and contigs, and the tree alignments are provided in the analysis archive `aculus_mito_analysis.tar.gz` (file manifest and commands in `ANALYSIS_COMMANDS.md`; SHA-256 in the accompanying `.sha256` file); archived at Zenodo (DOI: `10.5281/zenodo.23759884`). All new sequences reported here — the two re-closed mitogenomes of Eacu201601, the six nuclear copy-number anchor consensus sequences,

the nuclear ITS cluster assemblies of both sequenced individuals (Eacu201601 and NCU ZY01), and the new 16S and ITS sequences of *E. planostriata* — are provided in the sequence appendix of this preprint (Appendix S1, file `released_sequences.fas`).

2.9 Nuclear heterozygosity and phasing

To test whether the two mitochondrial lineages could reflect a mixture of two individuals, nuclear reads were mapped to the annotated exon set of *Albinaria teres* (GCA_964271275.2; Ensembl 2025_09; 28,875 transcripts, 33.9 Mb) with BWA-MEM v0.7.17 (Li, 2013) using `-Y` (soft clipping). Duplicates were removed with `samtools markdup` after `fixmate`, and reads with MAPQ < 20 were discarded. Per-site allele counts (A/C/G/T) were obtained with pysam 0.23.1 requiring base quality ≥ 25 and depth ≥ 20; alleles supported by $\geq 10\%$ of reads were counted. SNPs were called with bcftools 1.19 (`mpileup -q 20 -Q 20`; `call -mv`), and heterozygous sites were retained when QUAL ≥ 30, depth 20–200, and both alleles had ≥ 5 supporting reads. *A. teres* is the closest clausiliid with an annotated exon set; only conserved exons map (5.8% of the reference in Eacu201601, 2.8% in NCU), and all analyses are restricted to these sites. Because the reference is a different species, fixed differences appear homozygous and do not contribute to the heterozygosity estimate; allele frequencies are read directly from A/C/G/T counts rather than from reference-based genotypes.

Heterozygous SNPs were phased across read pairs. For each pair of heterozygous sites within 140 bp on the same transcript, reads spanning both sites (base quality ≥ 20, MAPQ ≥ 20) were used to count two-site haplotypes. A haplotype was accepted if supported by $\geq 25\%$ of spanning reads and ≥ 6 reads; pairs with < 10 spanning reads were discarded. A single diploid individual yields one or two haplotypes per pair (up to four with WGD/homeologs), whereas a mixture of two individuals can yield three or four at single-copy loci.

RESULTS

3.1 Two complete, deeply divergent mitogenomes in one individual

MEGAHIT local assembly recovered four mitochondrial contigs resolving into two haplotypes (Fig. 1A); graph-guided closure (Methods 2.2) yielded two circular mitogenomes of 15,417 bp (α) and 15,110 bp (β), collinear, with shared breakpoints and a control region closed on single graph edges in both (Fig. 1A). The two genomes differ by 5.15% at COI, 7.9% across a cytb–ND4L–ND1 window, and 11.0% at 16S. The consensus sequence of each genome contains all 13 protein-coding genes, 2 rRNAs and 22 tRNAs; two homopolymer sites in α carry length variants (§3.2).

Both genomes are covered uniformly and without gaps (mean depth $37\times$ / $59\times$; $\geq 5\times$ over 99.6–100% of both genomes), reads carrying lineage-diagnostic SNPs show NM ≤ 2 to their own reference, and the two rings are equal-length and collinear with a shared gene order and a single-genome copy-number budget (§3.3); these observations exclude index hopping, cross-mapping and NUMT origin, respectively. The lineage ratio is 42:58 (β fraction 0.58, per-reference counts; 0.579 ± 0.105 across 626 mixed-coverage diagnostic sites by allele depths) — two channels converge (Fig. 1C). Residual under-representation of β reads at single-lineage sites is consistent with soft-clipping of divergent reads.

3.2 Length variation at two homopolymer tracts in α

Two A/T tracts in α — ND4 (position 10,987) and ND6 (position 2,456) — show length variants in the Eacu reads: 28–36% of α-lineage reads carry the full-length allele and 64–72% carry a one-base contraction. Across the other 26 A/T runs ≥ 7 bp, including four runs of the same length (9–10 bp), the total length-variant fraction (deletions plus insertions) is $\leq 8.3\%$ (Fig. S1; Table S4). The effect is therefore specific to these two tracts rather than a general property of the PCR-amplified library; a tract-local artefact remains possible, and we treat the variants as exploratory (read-level views in Fig. S2).

3.3 The coexistence fits a normal mitochondrial copy-number budget

Relative to the independent 45S rDNA anchor, the mitochondrial pool of the mixed individual is similar to that of a single-lineage individual: mt:45S = 0.98 (Eacu) vs 1.01 (NCU) (k-mer depth ratios; per-reference read-depth estimates concur). The combined 5S/H3–H4 repeat family (Methods 2.5) gives mt:H3–H4 = 1.17 vs 0.97 and mt:5S = 1:28 vs 1:24, both matching the 45S result. The 5S:45S ratio, a protocol-insensitive cross-library control, is 28 vs 25. The 45S anchor therefore supports a combined mitochondrial copy number equivalent to the single-genome norm. The NCU individual carries no detectable α signal

above the 0.13% k-mer detection threshold — a β-only individual — so loss of one lineage occurs without any copy-number signature.

3.4 The two lineages are intraspecific at the ITS locus; "planostriata" is a third taxon

New 16S and ITS sequences from *E. planostriata* (Ma'anshan, Anhui) triangulate the system (Figs 2A, C). At 16S, true *E. planostriata* is at similar distances from both lineages (14.0% / 14.6%) and farther from each than they are from each other (11.0%): the tree is (*E. planostriata*, (α, β)). In the nuclear ITS region, *E. planostriata* differs from *E. aculus* by 2.4–3.5% — the same magnitude as *E. spinula* vs *E. aculus* (3.3%) — whereas α and β are identical (0%); a de novo ITS assembly of the β-only Nanchang individual is 99.9% identical to the Hangzhou assembly, consistent with the β lineage carrying the same nuclear rDNA at this locus. The two coexisting lineages are conspecific at the ITS locus, and the name *E. planostriata* — attached in INSDC to the β lineage (NC_056972, PX432505) — belongs to a third, distinct taxon defined here by 16S and ITS. A third "*E. aculus*" barcode clade ($\sim$22%, OR681015/OR523411) lacks locality data and falls outside the genus; it is excluded from the geographic synthesis.

3.5 Both lineages across four provinces

Assigning all public records by diagnostic SNPs (Table S1; Fig. 2B) yields: α-lineage — Guangxi (expert-identified AG670, cytb 97.7% to α); β-lineage — Nanchang (β-only individual, this study); both lineages — Hangzhou (intraindividual, 42:58) and one Nanjing locality where three individuals collected on the same day comprise two β and one α (MW378789/797/858). The two lineages thus occur among individuals of one population and, in one individual, together, across $\geq$1,200 km of the species' range; the Nanchang sample contains only β.

3.6 Relative divergence depth: 0.8$\times$ the shallowest congeneric distance in mitochondria, 0$\times$ in the nucleus

Expressed against congeneric reference points (Fig. 2C): α–β cytb-region distance (7.9%) is 0.33$\times$ the *E. spinula–E. aculus* distance (23.8%) and 3.4$\times$ the intralineage geographic divergence of α over 1,200 km (2.3%); α–β 16S distance (11.0%) is 0.79$\times$ the *E. planostriata–E. aculus* distance (14.0%). In the nucleus the corresponding figure is zero (0% vs 2.4–3.5% species-pair minimum). The mitochondrial lineages are thus $\sim$80% as divergent as the shallowest documented species pair in the genus, whereas no divergence is detected at the ITS locus. Within one snail, two mitochondrial genomes differ by 3.4$\times$ more than the same lineage differs at cytb across 1,200 km.

3.7 Nuclear phasing shows no second contributor

In Eacu201601, 164,274 reference positions were covered at $\geq$20$\times$ (mean 242$\times$) and 9,633 heterozygous site pairs were phaseable; in NCU, 105,674 positions (mean 320$\times$) and 6,660 pairs. With a 25% haplotype threshold, the distributions were nearly identical: one haplotype at 52.9% vs 54.0% of pairs, two at 41.8% vs 40.0%, and three at 3.3% vs 3.5%; no pair yielded four haplotypes (Table S5). At the 1,488 pairs where NCU showed exactly two haplotypes — single-copy, diploid anchors — Eacu201601 showed two haplotypes at 62.6%, one at 31.6%, and three at 5.0%, with no four-haplotype pair.

A 42:58 mixture of two individuals would produce three or four haplotypes at single-copy pairs where the donors differ, and would shift heterozygous allele fractions toward 0.42/0.58. Neither signature was observed: the mean alternative-allele fraction was 0.471 in Eacu201601 and 0.497 in NCU, and the two- and three-allele fractions were not elevated (4.1% vs 5.6%; 0.22% vs 0.30%). The nuclear data therefore show no evidence of a second contributor and are consistent with a single diploid individual.

DISCUSSION

4.1 Deep coalescence rather than recent introgression

Two processes can produce deeply divergent mitochondrial lineages within a species: introgression and incomplete lineage sorting. They make different predictions. Introgression is a capture process: the immigrant mitochondrial genome replaces the resident one through backcrossing and is not expected to coexist with it in the same individual. By contrast, long-standing ancestral polymorphism persists as coexisting lineages. In our individual, the two lineages coexist at 42:58; the β lineage is closer to α (11.0% at 16S) than to any named congener, including true *E. planostriata* (14.6%); and the spatial pattern is a mosaic rather than a cline. These observations are consistent with deep coalescence, not recent

replacement. We cannot exclude ancient introgression from an unsampled donor; genome-wide nuclear statistics (e.g., ABBA–BABA) on additional specimens would test this.

A two-individual mixture would also leave a nuclear signature. At the mitochondrial ratio of 42:58, sites where the two donors differ should show allele fractions of 0.42/0.58, and single-copy heterozygous pairs should resolve into three or four haplotypes. Neither was observed: Eacu201601's allele-frequency spectrum and phasing profile were indistinguishable from the single-individual NCU baseline, and its multiallelic fraction was, if anything, lower (§3.7). We therefore interpret the two mitochondrial lineages as coexisting within one individual. The test cannot exclude a second contributor below the detection limit, or a donor too similar at the assayed exons to leave a detectable phase signature.

4.2 Persistence of the two lineages

The persistence of two mitochondrial lineages across generations does not require a special mechanism. The probability of losing a lineage at frequency p through a germ-line bottleneck of n effective molecules is $(1-p)^n$; for a wide bottleneck, loss is unlikely, and the geographic mosaic (β only in Nanchang, both in Hangzhou) is consistent with stochastic loss in some populations.

4.3 Implications for barcoding and observability

The coexistence is not only a curiosity; it affects how mitochondrial barcodes are interpreted. Standard barcoding treats a mixed chromatogram as a failed reaction, and reference-guided assembly collapses minor haplotypes into a consensus. In our case, the published consensus carries the full-length allele at both tracts, whereas the raw reads show a mixture; whether this reflects biology or the calling and assembly pipeline cannot be determined from the public record. De novo assembly resolved the two mitogenomes because their divergence is $\sim$5%; at 1–2% they would merge into a chimeric consensus. The condition is therefore largely invisible to standard workflows.

When divergent lineages coexist within a specimen, the unit of barcoding — one specimen, one haplotype — is not well defined. Sequencing the same individual twice can assign it to different entities (our Sanger-versus-reads discrepancy at ND6: 39% versus 72% minor allele). We suggest nuclear loci as the primary delimitation substrate, population-level sampling, and treating mixed chromatograms as signal rather than failure. The INSDC records filed under "*E. planostriata*" (NC_056972, PX432505) are *E. aculus* lineage β; true *E. planostriata* should be circumscribed from type-region material. Automated BIN assignment can propagate such errors.

4.4 The state that lineage sorting erodes

In this system, the frequencies of the coexisting mitochondrial lineages are read directly from sequencing reads rather than inferred from gene-tree topology. This makes the state that lineage sorting erodes — deep coalescence — observable as a standing condition, at least in principle. The case also shows that the coalescent state variables can be measured in a natural population, not only in experimental evolution.

4.5 Limitations

Three limitations bound the inference. First, all read-level evidence comes from Eacu201601, and the geographic pattern rests on public single-locus records of heterogeneous provenance; population frequencies remain unknown. Second, voucher Eacu201601 is held in the China Jiliang University laboratory collection and the 2024 Anhui specimen is a private voucher; both are available for re-examination on request. Third, the nuclear heterozygosity and phasing analysis relies on a cross-species exon reference (*A. teres*), compares a PCR-amplified library (Eacu201601) with a PCR-free library (NCU), and cannot exclude a second contributor below the detection limit; WGD/homeologs allow up to four haplotypes per individual.

TABLES

Table S2: 1,133 diagnostic SNPs, allele depths per library (`copynum/Eacu.fm_ad.v2.tsv`). Table S3: copy-number raw counts (`copynum/*.kmer.tsv`). Table S4: long-run allele frequencies (`copynum/longrun_het.tsv`).

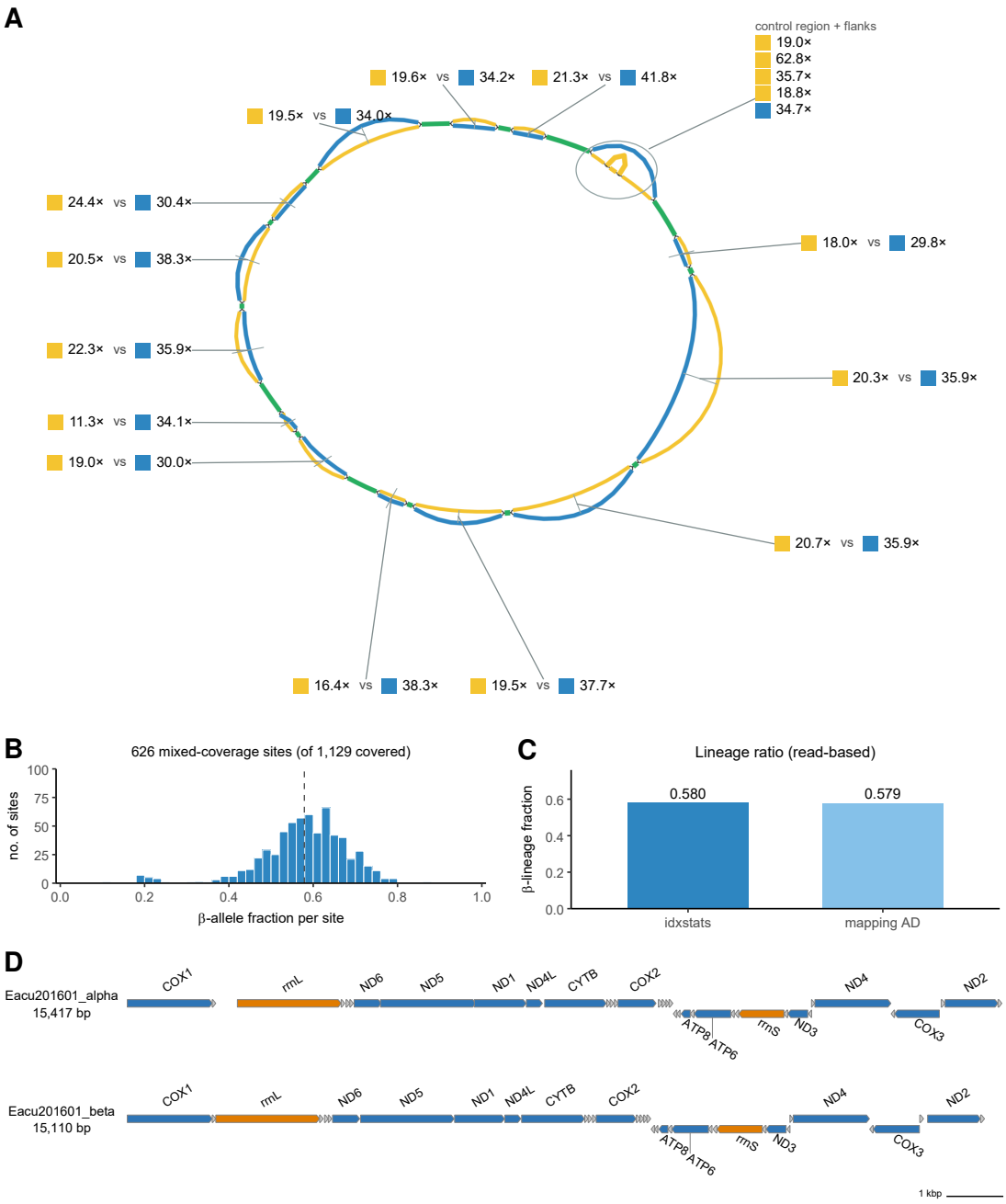


Figure 1. Two complete, deeply divergent mitochondrial genomes coexisting within one individual of *Euphaedusa aculus*. **(A)** SPAdes assembly graph (Bandage rendering) of mitochondrial reads recalled from the Eacu201601 library by dual-reference mapping. Depth-coloured nodes resolve into two closed circular haplotypes: yellow, α-lineage nodes (21); blue, β-lineage nodes (16); green, nodes shared by both lineages (7); grey, low-depth non-mitochondrial nodes (7). At graph forks, lineage was assigned by node depth (α $\sim$20$\times$ vs β $\sim$34$\times$); control-region closure followed single graph edges in both rings (α 15,417 bp; β 15,110 bp). **(B)** β-allele fraction across 1,133 lineage-diagnostic SNP sites: the 626 mixed-coverage sites form a unimodal distribution centred at 0.579 $\pm$ 0.105, consistent with a single mixed pool, while the remaining covered sites show single-lineage coverage. **(C)** The α:β ratio estimated from two read-based channels — per-reference read counts (0.580) and CIGAR-aware allele depths at diagnostic sites (0.579 $\pm$ 0.105) — converging on 42:58. **(D)** Gene maps of the two circular mitogenomes, linearized at the COI start and drawn in a common orientation (β reverse-complemented); both genomes carry the full complement of 13 protein-coding genes, 2 rRNAs and 22 tRNAs. Rows are the α (Eacu201601_alpha) and β (Eacu201601_beta) haplotypes; colours denote protein-coding genes (blue), rRNAs (orange) and tRNAs (grey).

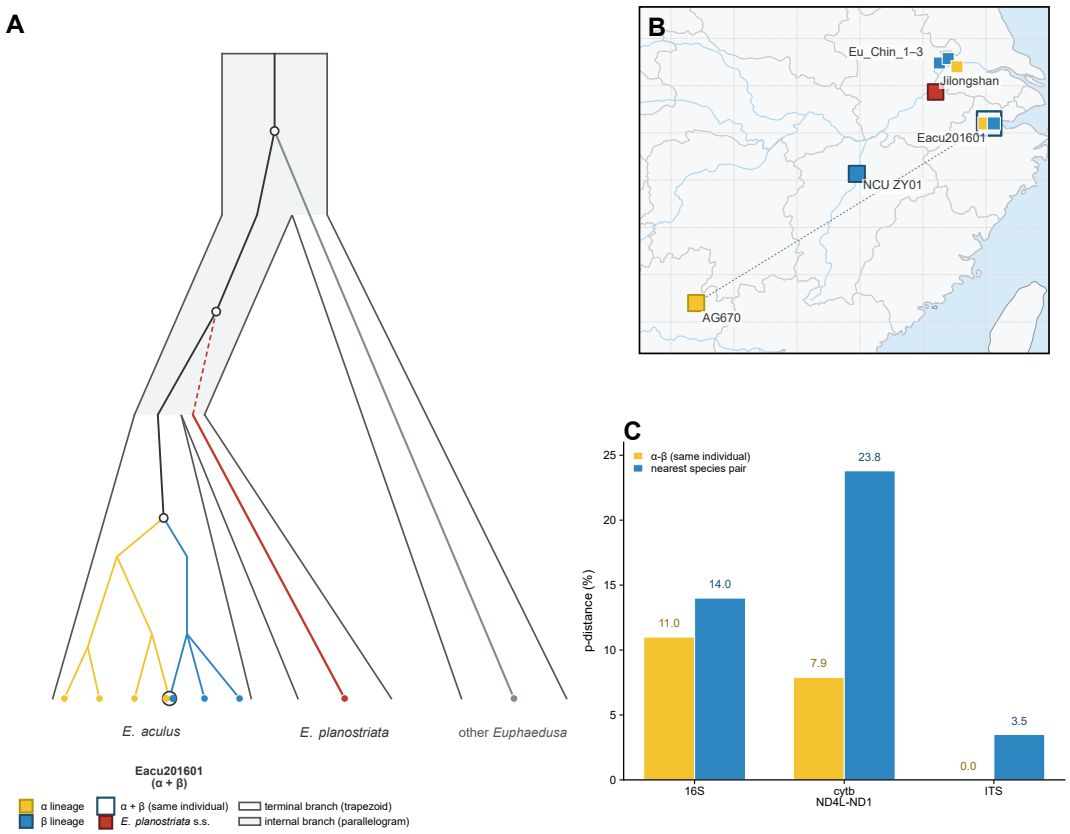


Figure 2. Lineage identity and geography of every public record. **(A)** Schematic summary of lineage assignment, drawn as a multispecies-coalescent hypothesis; inferred trees are shown in Fig. 3. Terminal branches are trapezoids and internal branches are parallelograms; the upper width of a terminal branch is its share of the parent branch. The α and β lineages coalesce deep within *E. aculus*, before the *E. aculus–E. planostriata* split, and the outgroup split is older. The half-filled tip marks the single specimen Eacu201601 carrying both lineages (42:58). All INSDC records submitted as "*E. aculus*" or "*E. planostriata*" were reassigned by diagnostic SNPs and pairwise distance. The two YS records ("*E. aculus*", no locality; OR681015/OR523411) are excluded. "*E. planostriata*" in INSDC (NC_056972, PX432505) is lineage β of *E. aculus*, whereas true *E. planostriata* s.s. (Jilongshan, Anhui; red) is a third, distinct taxon outside both lineages (16S + ITS). **(B)** Sampling geography: an α-only locality (Guangxi, expert-identified voucher AG670, ∼1,200 km from Hangzhou; dotted line), the mixed individual (Hangzhou; open double square), one Nanjing locality where three individuals collected the same day comprise 2 β + 1 α, a β-only individual (Nanchang ZY01; α undetectable by read mapping), and true *E. planostriata* s.s. from Anhui. Voucher identifiers are labelled directly on the map. **(C)** Divergence yardstick for 16S, the cytb–ND4L–ND1 block and the nuclear ITS cluster: α–β p-distance (gold) versus the nearest congeneric species-pair distance (blue). Values are 11.0 vs 14.0% (16S), 7.9 vs 23.8% (cytb), and 0.0 vs 3.5% (ITS); α–β reaches 0.79× the *planostriata–aculus* distance at 16S, 0.33× the *spinula–aculus* distance at cytb, and 0% at ITS. The within-α Hangzhou–Guangxi cytb control (2.3% on the same 1,393-bp block) is 3.4× lower than α–β.

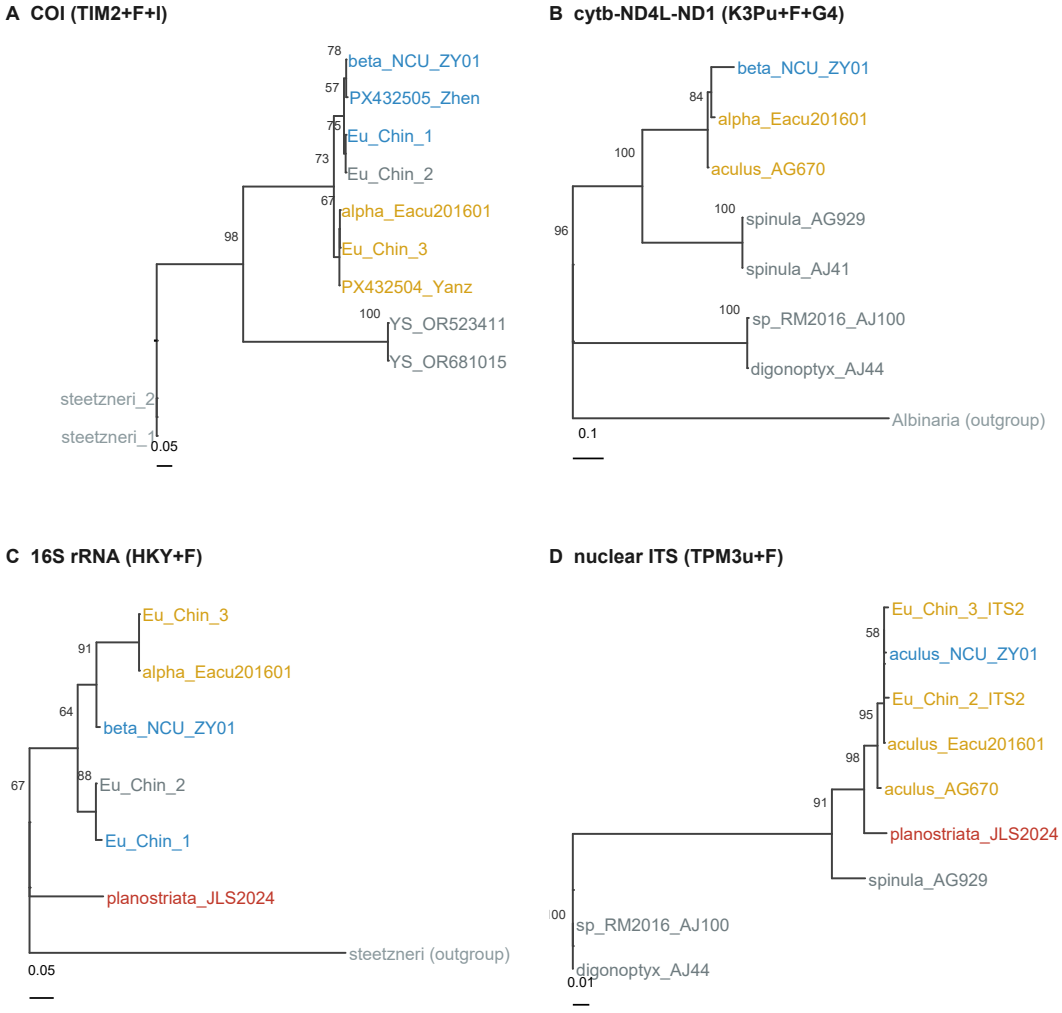


Figure 3. Maximum-likelihood phylogenies of the four marker matrices (IQ-TREE v3.0.1, BIC model selection, 1,000 ultrafast bootstraps; values $\geq$50 shown; tips coloured by lineage as in Fig. 1; scale bars, substitutions per site). **(A)** COI barcode window including the full COI CDS of both lineages (TIM2+F+I): the Nanjing COI barcodes MW378797 and MW378858 fall within lineage β, whereas MW378789 falls within lineage α; the "*planostriata*" barcode PX432505 falls within β and PX432504 (Yanz) within α; the YS records fall outside the genus. **(B)** Partitioned cytb–ND4L–ND1 block (K3Pu+F+G4): lineage α, lineage β and the expert-identified Guangxi voucher AG670 (2.2% from α) against congeners at 27–35% divergence; AG670's branch is effectively zero-length and its position among $\{\alpha, \beta\}$ is likelihood-ambiguous (two groupings within $\sim$3 log-likelihood units; site-pattern and distance evidence favour (α, AG670)). **(C)** 16S conserved core, including the three fragmentary Nanjing 16S records (262–264 bp; HKY+F+G4): the (*E. planostriata* s.s., (α, β)) geometry of §3.4 is retained (91); Eu_Chin_3 falls on the α tip (zero divergence over its span), while the Eu_Chin_1/2 pair attaches within the aculus clade at low support (64) — their barcode window sits in the most conserved segment of 16S (local α–β divergence 1.9% vs 7.7% gene-wide core), so this window does not resolve lineage membership, which rests on COI and genome-wide best hits (Table S1). **(D)** Nuclear ITS cluster, rooted on the most divergent congeners (TPM3u+F): all aculus-accession tips — the Hangzhou mixed individual, the Guangxi voucher AG670, the β-only Nanchang individual (de novo ITS assembly, 99.9% identical to the Hangzhou assembly), and ITS2 sequences of both Nanjing entities — collapse to a near-zero-divergence crown above the *E. planostriata* s.s. split (2.4–3.5%), irrespective of mitochondrial lineage.

Record(s)	Assigned here	Evidence	Locality
PP485172.1; SRR28286884	$\alpha + \beta$, 42:58 (intraindividual)	read depth + diagnostic SNPs	Hangzhou, Zhejiang
NC_056972.1; SRR14226860	β only (α undetectable by read mapping); nuclear ITS = aculus pool (99.9%)	read depth + ITS assembly	Nanchang, Jiangxi
LC171974; LC167924; LC168274	α	cytb 97.7% to α	Pingle, Guangxi
MW378789	α ($p < 0.001$)	COI	Nanjing, Jiangsu[1]
MW378797; MW378858	β	COI	Nanjing, Jiangsu[1]
MW378922/930/984	Eu_Chin_3 = α (100% to α), Eu_Chin_1/2 = β (95.4% to β vs 93.8% to α)[1]	16S window (low support)	Nanjing, Jiangsu
Eu_Chin_2/3 ITS2	aculus nuclear pool (both mt lineages)	ITS2	Nanjing, Jiangsu
PX432504	α (0.3%)	COI	not recorded
PX432505	β (0.6%)	COI	not recorded
OR681015; OR523411	excluded: outside the genus (22–23%)	COI	not recorded
this study, Appendix S1	third taxon, outside α/β (16S + ITS)	Sanger	Ma'anshan, Anhui
LC172009/038/039/026; LC168309/341/328	reference taxa (correctly labelled)	—	—
OP494302/298/308; NC_001761	outgroups (correctly labelled)	—	—

Table S1. INSDC records re-assigned to mitochondrial lineages of *Euphaedusa aculus*, and reference taxa. Full version with all columns (including data type and original submission) in `tables/TableS1.tsv`. [1] Same site and date (2018-04-25; 32.0583N 118.796E): two β + one α individual; 16S assignments by genome-wide best hit, concordant with COI. The 263-bp 16S window lies in the most conserved segment of 16S (local α–β divergence 1.9% vs 7.7% gene-wide) and therefore attaches within the aculus clade at low support (64) in Fig. 3C; lineage membership rests on COI and genome-wide best hits.

Threshold	Sample	1	2	3	4	≥5
10%	Eacu201601	1,080	3,345	3,856	1,221	84
10%	NCU	904	2,325	2,627	707	69
20%	Eacu201601	3,341	4,882	1,254	34	0
20%	NCU	2,433	3,280	813	22	0
25%	Eacu201601	5,101	4,024	322	0	0
25%	NCU	3,599	2,665	234	0	0

Table S5. Sensitivity of phased haplotype counts to the haplotype-frequency threshold. Shown are the numbers of heterozygous site pairs resolving into 1, 2, 3, 4 or ≥5 accepted haplotypes. A haplotype was required to be supported by ≥6 reads and by the stated fraction of spanning reads; pairs with no accepted haplotype are omitted.

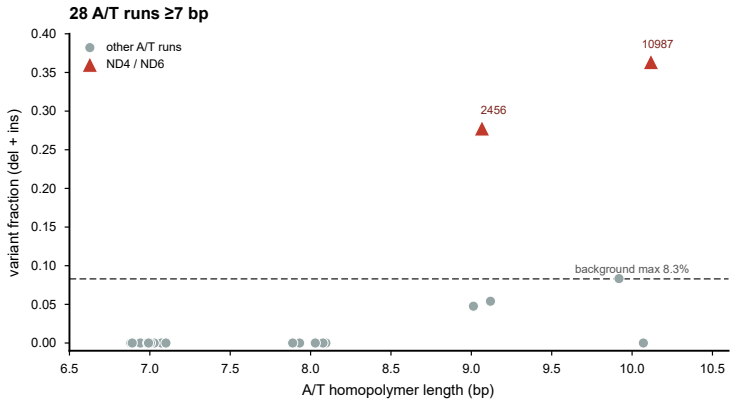


Figure S1. Length-variant fraction at 28 A/T homopolymer runs ≥7 bp in lineage α. Grey dots, other runs; red triangles, ND4 (position 10,987) and ND6 (2,456); dashed line, background maximum of 8.3% (deletions plus insertions). The two outliers reach 28–36% full-length and 64–72% contracted reads. Because the Eacu library is PCR-amplified and the published consensus carries full-length alleles at both tracts, biological heteroplasmy, PCR slippage and assembly/curation cannot be distinguished from these data; the variants are exploratory.

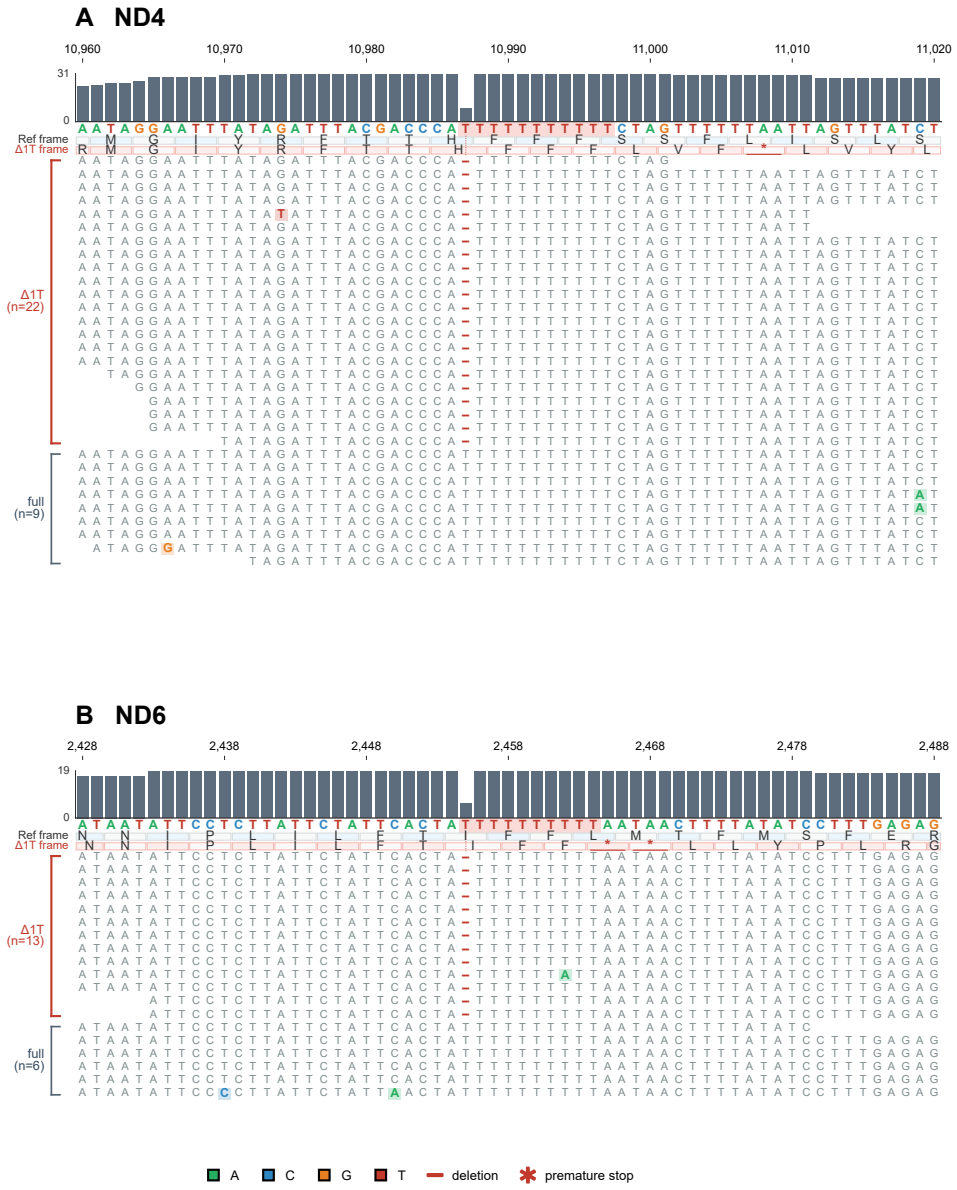


Figure S2. Read-level views of the two homopolymer length variants in lineage α. **(A)** ND4 T tract (position 10,987 in PP485172.1), remapped against the full-length-allele α reference. The upper track shows the reference reading frame; the lower red track shows the -1T frame, with the first premature stop (codon 244). Deletion-bearing reads (-1T) outnumber full-length reads 22:9. **(B)** Same for the ND6 T tract (position 2,455); 13:6 deletion-bearing versus full-length reads, with the first premature stop at codon 99.

REFERENCES

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., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5):455–477.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1):421.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17):i884–i890.

Chiba, S. (2005). Appearance of morphological novelty in a hybrid zone between two species of land snails. *Evolution*, 59(8):1712–1720.

David, P., Degletagne, C., Saclier, N., Jennan, A., Jarne, P., Plénet, S., Konecny, L., François, C., Guéguen, L., Garcia, N., Lefébure, T., and Luquet, E. (2022). Extreme mitochondrial DNA divergence underlies genetic conflict over sex determination. *Current Biology*, 32(10):2325–2333.e6.

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, 39(suppl_2):W29–W37.

Ghiselli, F., Iannello, M., Piccinini, G., and Milani, L. (2021). Bivalve molluscs as model systems for studying mitochondrial biology. *Integrative and Comparative Biology*, 61(5):1699–1714.

Greenwood, M. P., Capblancq, T., Wahlberg, N., and Després, L. (2025). Whole genome data confirm pervasive gene discordance in the evolutionary history of Coenonympha (Nymphalidae) butterflies. *Molecular Phylogenetics and Evolution*, 202:108222.

Hwang, C.-C., Ger, M.-J., and Wu, S.-P. (2022). Within-island diversification in the land snail genus Formosana (gastropoda, clausiliidae) in Taiwan. *Zoologica Scripta*, 51(5):562–588.

Karlicki, M., Antonowicz, S., and Karnkowska, A. (2022). Tiara: deep learning-based classification system for eukaryotic sequences in metagenomic data. *Bioinformatics*, 38(2):344–350.

Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4):772–780.

Kawato, S. (2026). gbdraw: a genome diagram generator for microbes and organelles. *bioRxiv*. Preprint, posted 9 April 2026.

Laslett, D. and Canbäck, B. (2008). ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics*, 24(2):172–175.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31(10):1674–1676.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

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 format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Liang, Q.-C., Tao, K.-X., Qiao, D.-D., Wang, P., Yin, H.-F., and Yang, Q.-Q. (2024). The nearly complete mitochondrial genome of the door snail Euphaedusa aculus (Stylommatophora: Clausiliidae) and phylogenetic analysis. *Mitochondrial DNA Part B*, 9(11):1581–1585.

Macey, J. R., Pabinger, S., Barbieri, C. G., Buring, E. S., Gonzalez, V. L., Mulcahy, D. G., DeMeo, D. P., Urban, L., Hime, P. M., Prost, S., Elliott, A. N., and Gemmell, N. J. (2021). Evidence of two deeply divergent co-existing mitochondrial genomes in the tuatara reveals an extremely complex genomic organization. *Communications Biology*, 4(1):116.

Mason, K., Fehér, Z., Bamberger, S., Reier, S., Szekeres, M., Sattmann, H., Kruckenhauser, L., De Mattia, W., and Haring, E. (2020). New insights into and limitations of the molecular phylogeny in the taxon-rich land snail genus Montenegrina (mollusca: Gastropoda: Clausiliidae). *Journal of Zoological Systematics and Evolutionary Research*, 58(3):662–690.

Motochin, R., Wang, M., and Ueshima, R. (2017). Molecular phylogeny, frequent parallel evolution, and new system of Japanese clausiliid land snails (gastropoda: Stylommatophora). *Zoological Journal of the Linnean Society*, 181(4):795–845.

Noguerales, V. and Emerson, B. C. (2025). Arthropod mtDNA paraphyly: a case study of introgressive origin. *Journal of Evolutionary Biology*, 38(2):272–283.

Ntatsopoulos, K., Meramveliotakis, E., Anastasiou, I., and Papadopoulou, A. (2026). Mito-nuclear discordance in a genus of darkling beetles (Coleoptera: Tenebrionidae: *Dailognatha*) reveals both incomplete lineage sorting and introgression driven by complex island palaeogeography. *Evolutionary Journal of the Linnean Society*, 5(1):kzag026.

Palumbi, S. R., Martin, A., Romano, S., McMillan, W. O., Stice, L., and Grabowski, G. (1991). *The Simple Fool's Guide to PCR*. University of Hawaii. 16sAR/16sBR primers.

Sauer, J. and Hausdorf, B. (2012). A comparison of DNA-based methods for delimiting species in a Cretan land snail radiation reveals shortcomings of exclusively molecular taxonomy. *Cladistics*, 28(3):300–316.

Trillo, M. and Ortego, J. (2026). Genomic-based species delimitation reveals cryptic diversity and semi-permeable species boundaries in a *Chorthippus* (Orthoptera: Acrididae) grasshopper complex from Greece. *Systematic Entomology*, 51(2):1–19.

Wade, C. M. and Mordan, P. B. (2000). Evolution within the gastropod molluscs: using the ribosomal RNA gene-cluster as an indicator of phylogenetic relationships. *Journal of Molluscan Studies*, 66(4):565–570.

Wang, Y., Zhu, X., Dong, X., Wang, K., Yang, R., Zheng, C., and Bu, W. (2025). Species boundaries and phylogenetics of *Aelia* stink bugs (Hemiptera: Pentatomidae) in China, with notes on mito-nuclear discordance. *Zoologica Scripta*, 54(2):196–212.

Wang, Y., Zhu, X., Tang, Z., Wang, K., Fu, S., Zheng, C., and Bu, W. (2026). The species delimitation challenge resulting from mito-nuclear discordance: a case of Macroscytus burrower bugs from China (Hemiptera: Heteroptera: Cydnidae). *Molecular Phylogenetics and Evolution*, 220:108591.

Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*, 31(20):3350–3352.

Wong, T. K. F., Ly-Trong, N., Ren, H., Demotte, P., Baños, H., Roger, A. J., Susko, E., Bielow, C., De Maio, N., Goldman, N., Hahn, M. W., Dos Reis, M., Vinh, L. S., Huttley, G., Lanfear, R., and Minh, B. Q. (2026). IQ-TREE 3: phylogenomic inference software using complex evolutionary models. *Molecular Biology and Evolution*, 43(5):msag117.

Xu, Z.-J. (2026). Two coexisting H3–H4 histone gene architectures in Phaedusinae land snails: tandem and divergent repeat units, architecture-stratified intergenic spacers, and 5S rRNA insertions. Companion manuscript, in preparation.

Yu, X.-Y., Tang, Z.-Z., Zhang, Z., Song, Y.-X., He, H., Shi, Y., Hou, J.-Q., and Yu, Y. (2026). GeneMiner2: accurate and automated recovery of genes from genome-skimming data. *Molecular Ecology Resources*, 26(2):e70111.

Zhao, T., Zhu, P., Liu, Q., Ma, L., Xu, Y., Lü, L., Duan, Y., Song, F., Tian, L., Cai, W., and Li, H. (2025). Mitonuclear discordance of beetles shaped by incomplete lineage sorting and introgression under loose interaction mechanism. *Molecular Biology and Evolution*, 42(11):msaf291.

Zouros, E. (2013). Biparental inheritance through uniparental transmission: the doubly uniparental inheritance (DUI) of mitochondrial DNA. *Evolutionary Biology*, 40(1):1–31.