Main Title:

Closing the reference gap for Australo-Papuan rodents: a 431-mitogenome library for environmental DNA and species discovery

Running Title:

A mitogenome library for Australo-Papuan rodents

Authors:

Charlotte F. Drury[1,2,*], Anna J. MacDonald[2], Kevin C. Rowe[3], Craig Moritz[2], Emily Roycroft[1,2,*]

Author Affiliations:

1. School of Biological Sciences, Monash University, Clayton VIC, Australia
2. Research School of Biology, Australian National University, Acton ACT, Australia
3. Museums Victoria Research Institute, Melbourne VIC, Australia

* Corresponding authors:
Charlotte Drury: charlotte.drury@monash.edu; Emily Roycroft: emily.roycroft@monash.edu

Data availability:

Raw sequence data used in this project come from previously published studies (see Methods) and are available at the Bioplatforms Australia Data Portal (https://data.bioplatforms.com), along with corresponding metadata for all samples, as part of the Oz Mammals Genomics Initiative (doi:10.25953/cg5b-qd03). Sample numbers are listed in Supplementary Table S1. Final assembled mitogenomes and mitochondrial gene alignments have been deposited in 10.6084/m9.figshare.34964431. Code supporting analyses and the mitogenome assembly pipeline is available on GitHub (github.com/DruryCF/mtdna-australo-papuan-rodents).

Benefits Generated:

Benefits from this research accrue from the sharing of our assembled mitochondrial genome data on public databases as described above, making them available as an openly accessible resource for biodiversity monitoring, eDNA and species discovery.

Conflict of interest:

The authors declare that they have no conflicts of interest

Ethical Statement:

This study utilised publicly available sequencing datasets and data derived from museum specimens. As no live animals were involved, ethics approval was not required.

Abstract

As biodiversity declines at an unprecedented rate, rapid and scalable approaches to monitoring species are increasingly needed, particularly in poorly known and highly threatened groups. Environmental DNA (eDNA) is increasingly enabling high-throughput monitoring to inform conservation. However, species-level monitoring through eDNA relies on comprehensive reference databases and short, species-discriminating 'mini-barcode' loci, resources that are lacking for many taxa. The application of eDNA for species-level monitoring is therefore constrained by these gaps. Here, we use Australo-Papuan rodents (subfamily Murinae: tribe Hydromyini) as a model system to address these limitations, given their high species diversity, conservation concern, and distribution across geographically complex regions. Among animals, eDNA monitoring typically depends on mitochondrial markers. Leveraging taxonomically-verified museum specimens, we develop a 431-mitogenome reference library representing ~72% of described species diversity, and use this database to assess the species-level resolution of candidate mini-barcodes. We show that common *12S rRNA* mammalian eDNA markers cannot reliably discriminate between Hydromyini species (resolving 23-33% of tested species). In contrast, a Rodentia-specific *Cytb* marker resolves 66% of tested species, and two novel *ATP6* Hydromyini markers, presented here, resolves 93% through a multi-locus approach. This reference library represents a >40-fold increase in available complete mitogenomes for Australo-Papuan rodents, enabling eDNA monitoring in this highly threatened group. More broadly, we demonstrate the value of pairing comprehensive reference resources with empirical marker assessment to reveal where eDNA-based approaches may fail to distinguish species. Understanding these limitations is essential to realising the promise of eDNA as a reliable tool for species-level monitoring and conservation.

Keywords: Australia, Mitochondrial DNA (mtDNA), Environmental DNA (eDNA), metabarcoding, reference database, Hydromyini

1 Introduction

Biodiversity is declining globally at an unprecedented rate (Barnosky et al. 2011), driven by habitat loss, environmental change, invasive species, overexploitation, and other anthropogenic pressures. At the same time, much of global biodiversity remains undocumented, with an estimated 80-90% species undescribed (Mora et al. 2011). Effective monitoring of biodiversity is therefore increasingly important for detecting population declines, identifying threatened species, and informing conservation management (Lindenmayer and

Likens 2010; Stephenson et al. 2022). However, conventional biodiversity surveys can be time-consuming, expensive, and difficult to implement across large or remote areas, particularly for rare, elusive or cryptic species. Environmental DNA (eDNA) – DNA extracted from environmental sources such as water, soil, or air (Pawlowski et al. 2020) – is increasingly used for high-throughput biodiversity monitoring and species detection (e.g. Allen et al. 2023; Broadhurst et al. 2021; Gray et al. 2025; Kim et al. 2025; Kim et al. 2022; Leempoel et al. 2020; Sales et al. 2020; Ushio et al. 2017). By enabling the detection of multiple taxa from environmental samples without requiring direct observation or capture, eDNA can increase the efficiency and spatial coverage of biodiversity surveys. This approach also has considerable potential for species discovery, with candidate species subsequently validated using nuclear genomic data, particularly in regions where biodiversity remains incompletely documented.

Because of these advantages, the field of eDNA is currently transitioning from research into broadly relevant operational monitoring tools. As such, there is considerable ongoing effort to establish best practice guidelines and address methodological challenges across all stages of eDNA workflows (e.g. De Brauwer et al. 2023a; Hajibabaei et al. 2026; Loeza-Quintana et al. 2020; MacDonald and Sarre 2017; Theroux et al. 2025; Watts et al. 2026). For instance, the effectiveness of eDNA-based biodiversity assessments is strongly dependent on the availability of reference sequences from taxonomically verified specimens. Many taxonomic groups remain poorly represented in existing databases, with sequences often generated opportunistically rather than through systematic sampling (Chorlton 2024; Keck et al. 2023; but see De Brauwer et al. 2023b; Gold et al. 2021). Consequently, eDNA sequences may be incorrectly assigned or left unidentified, potentially leading to underestimation of biodiversity (Gold et al. 2021; Morey et al. 2024; De Santiago et al. 2025).

Furthermore, because eDNA is typically highly degraded, the loci that are most commonly targeted in eDNA sequencing projects are generally very short fragments (‘mini-barcodes’; ~100-200bp in length; Hajibabaei et al. 2006; Meusnier et al. 2008), that are flanked by conserved primer sites. These mini-barcodes have often been adopted based on convention (e.g. Leray et al. 2013; Roe and Sperling 2007; Ushio et al. 2017), based on sequence availability in reference databases for primer design (e.g. Galan et al. 2012), or for their ability to be amplified at broad taxonomic scales (e.g. Kelly et al. 2014; Riaz et al. 2011), rather than being specifically designed and validated for species-level identification within focal taxonomic groups (but see Miya et al. 2015). While such markers may be sufficient for characterising community composition at higher taxonomic levels, their capacity to discriminate among closely related species is likely to be highly variable and dependent on the focal taxonomic

group (e.g. Campbell et al. 2024; Gold et al. 2021; Kelly et al. 2014; Miya et al. 2015). This can be particularly important where eDNA is used to survey threatened species that co-occur with a close relative as insufficient mini-barcode resolution may prevent reliable discrimination of the target species (e.g. Polanowski et al. 2024; Wilcox et al. 2015). Developing comprehensive, taxonomically verified reference databases using barcode regions specifically evaluated for their capacity to distinguish focal species is therefore critical for ensuring the reliability of eDNA for species-level detection and biodiversity discovery (Gold et al. 2022; Schmid et al. 2025; Weigand et al. 2019).

The Australo-Papuan region, defined here as including Australia, New Guinea, and neighbouring islands, is a highly biodiverse region (Myers et al. 2000) that has suffered extensive recent biodiversity decline (Burbidge 2023; Leunufna 2026; Roycroft et al. 2021b; Woinarski et al. 2019), and harbours varied, complex and remote landscapes. Australian mammals have been especially impacted by recent biodiversity decline, with at least 40 species extinct since European colonisation in 1788 (Woinarski et al. 2019; Burbidge 2023), and a further 31% threatened with extinction (EPBC Act). Many mammals across New Guinea and surrounding islands are poorly known, with several species presumed extinct and many mammal species likely undocumented (see Flannery et al. 2026; Travouillon et al. 2026). Protecting these remaining species from extinction requires innovative and cost-effective approaches to biodiversity monitoring and species discovery, making the region poised to benefit substantially from the deployment of eDNA approaches.

Australo-Papuan rodents (subfamily Murinae) are among the region's most threatened mammal groups. In Australia, rodents make up 19% of terrestrial mammal diversity, but 41% of modern mammal extinctions (Roycroft et al. 2021b; Burbidge 2023). These modern extinctions are the result of several interacting and ongoing factors, including predation by feral cats and foxes, land-clearing and the introduction of novel diseases (Woinarski et al. 2015). There are two independent evolutionary radiations of rodents in the Australo-Papuan region, the Hydromyini (~160 species, with an 8.5-million-year evolutionary origin in the region; Roycroft et al. 2022), and the Rattini radiation (~24 species, with a 1.1-million-year origin in the region; Rowe et al. 2011). As a highly vulnerable and wide-ranging group, Australo-Papuan rodents would particularly benefit from eDNA approaches. However, to facilitate this, there is a need to generate comprehensive mitochondrial reference sequences for this group. The more diverse Hydromyini tribe is represented by only six complete mitochondrial genomes in public databases (GenBank, accessed 17 July 2026), representing ~4% of total Sahul hydromyine species diversity. Less than 50% of recognised Australo-Papuan rodent species are represented by partial mitochondrial DNA fragments, usually only one or two genes.

Available sequences are unevenly distributed among genera and geographic regions, with parts of New Guinea particularly underrepresented. This lack of available references currently severely limits the broad-scale application of eDNA for biodiversity detection and monitoring of Australo-Papuan rodents, and prevents both the diagnosis of unknown eDNA samples to species level and the evaluation of appropriate taxon-specific markers.

Here, we address these gaps by assembling and curating a taxonomically comprehensive 431-mitogenome reference library from Australo-Papuan rodent species from across Australia, New Guinea, Maluku and the Solomon Islands, linked to taxonomically verified museum specimens. We leverage this novel reference library to quantitatively assess the capacity of existing and novel mini-barcodes to diagnose and distinguish species, and make recommendations for their use in environmental DNA and species discovery.

2 Methods

2.1 Sampling and taxonomic completeness

Raw sequencing data were generated as part of the Oz Mammals Genomics Initiative, a collaborative project that aimed to generate genomic resources for all Australian terrestrial mammals (https://ozmammalsgenomics.com/; Eldridge et al. 2020). Sampling was targeted to cover all living and recently extinct Australo-Papuan rodent species, including taxonomically described species, distinct geographic lineages and putative species (see Roycroft et al. 2022), and included species from Australia, New Guinea (Papua New Guinea and West Papua), Solomon Islands, Maluku, and broader southeast Asian outgroups. Full methods of specimen sampling, DNA extraction and genomic library preparation are provided in Roycroft et al. (2020; 2021b; 2022). The purposes of these datasets were two-fold: 1) to reconstruct the evolutionary history of Australo-Papuan rodents based on reliable, single-copy, nuclear markers appropriate for phylogenomics, and 2) to generate by-catch mitochondrial genomes as resources for Australo-Papuan rodents, which are critical resources for species identification and discovery. To meet both these goals, genomic libraries were enriched for a targeted set of coding regions using an in-solution hybridisation capture approach. A total of 642 libraries (from 614 unique specimens) were enriched for ~1417 nuclear exons (1.27 Mb) and two mitochondrial genes (Cytochrome *b* and Cytochrome *c* oxidase I) using the bait design described in Roycroft et al. (2020; 2022), and a further 58 libraries (from a further 40 unique specimens) were enriched for whole exomes (54.3 Mb) as described in Roycroft et al. (2021a; 2021b). Although these datasets were enriched for primarily nuclear genomic regions, previous studies (e.g. Allio et al. 2020; Mulder et al. 2016; Picardi and Pesole 2012; Raposo

do Amaral et al. 2015) have had high success in assembling off-target mitochondrial genes from nuclear-enriched sequencing datasets, given the overrepresentation of the mitochondrial genome within cells relative to the nuclear genome in DNA libraries.

Among a total of 654 paired-end sequencing datasets from unique specimens (including both exon and whole exome enriched libraries; see Supplementary Table S1), at least 227 described species or previously known distinct phylogenetic lineages (Roycroft et al. 2022) are represented. These include representatives from all six Divisions within the Sahul Hydromyini, and the Australo-Papuan *Rattus* radiation within the Rattini tribe.

2.2 Mitochondrial genome assembly and curation

Raw sequencing reads were first deduplicated using the dedupe.sh module of BBMap (v39.06; Bushnell 2014) to remove PCR duplicates. We then used the paired-end option of Trimmomatic (v0.39; Bolger et al. 2014) to remove sequences from adapters, trim regions with an average quality score of less than 20 in a sliding window of 8 bases, and to filter out reads with a final trimmed length of less than 40 bases.

Taking all pairs of deduplicated, trimmed reads, we initially assessed three alternative mitochondrial genome assembly pipelines – MIRA (v4.0.2; Chevreux et al. 1999) through MITObim (v1.9.1; Hahn et al. 2013), MEGAHIT (Li et al. 2015) through MitoZ (v3.6; Meng et al. 2019), and NOVOPlasty (v4.3.5; Dierckxsens et al. 2017) – to determine which approach generated the most complete and high-quality final mitochondrial assemblies (see Supplementary S.1). From this assessment, we found that MITObim tended to assemble the most complete mitochondrial genomes (~94%; see Supplementary Fig. S1), compared to MitoZ (~71%) and NOVOPlasty (~51%). However, MITObim was unable to identify cases where contaminant sources were present within a sequencing library (e.g. low levels of cross-contamination between samples, potentially via index switching; Supplementary Fig. S2). In contrast, MitoZ was able to distinguish reads from multiple sources within a library, and assemble these as separate contigs that could then be manually curated and filtered.

Based on this assessment, we developed a two-stage assembly and quality control approach that leveraged the contrasting benefits of both MitoZ and MITObim (see Supplementary S.1; Fig. S3). This approach first used the MEGAHIT module of MitoZ to both assemble mitochondrial genomes and screen for contamination. Then, for samples with no evidence of cross-contamination present in a library, but where the mitogenome was incomplete after MitoZ assembly, we took samples downstream to MITObim to extend the completeness of the final assemblies where possible.

Following this two-step process, an initial 519 successfully assembled mitochondrial genomes and mitochondrial gene regions were aligned to the *Mus musculus* reference genome (Ref Seq: NC_005089.1) using MAFFT's (v7.526; Katoh and Standley 2013) addfragments alignment protocol (the control region was separately aligned using MAFFT's E-INS-i alignment protocol across all assembled mitogenomes) and alignments were manually curated to remove clear errors or mis-assembled regions using the Jalview (v2.11.4.0; Waterhouse et al. 2009) alignment viewer. Then we used a phylogenetically-informed approach to further curate the dataset, by excluding genes from the final dataset if they did not fall in their expected phylogenetic grouping, based on published phylogenetic relationships and taxonomic groupings of Australo-Papuan rodents (Roycroft et al. 2022). Given the intent of this dataset is as a reference set of mitochondrial genomes to facilitate the identification of species from eDNA and mitochondrial barcodes, we applied this conservative approach to remove taxonomically erroneous mitochondrial sequences, retaining a total of 431 mitogenomes for downstream analysis.

2.3 Assessing phylogenetic information content across the mitogenome

To quantify the distribution of phylogenetic information content across the mitochondrial genome with respect to species-level relationships in Australo-Papuan rodents, and assess the diagnostic capacity of barcode-sized regions for species-level diagnosis, we took adjacent, non-overlapping 200bp windows in silico across the curated, final mitochondrial alignment of 431 samples. We only considered protein coding and rRNA gene regions, as non-coding regions of the mitochondrial genome are typically not suitable regions for cross-species amplification given high levels of sequence divergence. Each 200bp window was taken starting from the 5' end of the alignment of each gene. We first subset each of these 200bp alignments to only include samples that had >80% nucleotide completeness across the window. Then, as we were interested in the ability of these barcode regions to accurately diagnose unknown sequences to species level, and not in their resolution within species, we further subset each alignment to only retain the single, most complete representative of each species. Finally, to prevent taxonomic bias between windows, we retained only species that were represented across all windows.

For each filtered 200bp alignment, we generated maximum likelihood phylogenetic trees with 1000 bootstrap replicates using IQ-TREE (v1.6.12; Nguyen et al. 2015; Hoang et al. 2018), with automated model selection using ModelFinder (Kalyaanamoorthy et al. 2017). As a proxy for phylogenetic informativeness, and as a direct measure in the ability of each 200bp window

to unequivocally support a particular phylogenetic topology, we took the mean tree-wide bootstrap score for each window and plotted these against mitogenomic position, to visualise the distribution of phylogenetic support across the mitogenome.

Windows with an average bootstrap score of ≥80 were selected as candidate loci for assessment as potential new barcode regions with high capacity for species-level diagnosis in Australo-Papuan rodents, and for comparison with existing, commonly used barcode regions in eDNA studies. As a control to test whether higher average bootstrap support was an appropriate metric with which to infer the species-level diagnostic capacity of a barcode region, we also included 11 windows of <80 average bootstrap support across the full complement <80 scores (~67-79). We further retrieved the IQ-TREE reported number of phylogenetically informative sites for these windows to test whether phylogenetically informative sites would be a more or similarly informative metric to infer the species-level diagnostic capacity of a barcode region.

2.4 Testing the capacity of new and existing mini-barcodes to diagnose species in Australian rodents

2.4.1 Assessing species-level diagnostic capacity with reciprocal monophyly

We then tested the capacity of mini-barcode length regions and commonly-used eDNA mini-barcodes to diagnose species in Australian rodents. For comparison in this testing, we included our selected 200bp windows (described above); centrally subset 100bp windows to contrast the resolution obtained with 200bp versus 100bp regions; and three published mini-barcodes designed to target broad to narrow taxonomic groups (i.e. subphylum, class, and order). We selected the most common vertebrate eDNA marker (12-V5; Riaz et al. 2011), a common mammal eDNA marker (MiMammal; Ushio et al. 2017), and a Rodentia-specific marker (Galan's *Cytb* barcode; Galan et al. 2012) for our analysis. As a point of comparison, we also assessed the standard full-length 658bp *COI* barcode, Folmer's region (Folmer et al. 1994; Hebert et al. 2003), noting that it is too long to be used in typical eDNA studies. We subsequently also compared these to two new *ATP6* mini-barcodes designed for Australo-Papuan rodents, the design and selection of which is described in Section 2.4.2 below. Window and barcode alignments were subset to all Hydromyini samples with ≥80% within-region completeness in our dataset. For each filtered alignment, we generated maximum likelihood phylogenetic trees with 1000 bootstrap replicates using IQ-TREE, with automated model selection using ModelFinder.

We assessed species-level diagnostic capacity of these regions based on a criterion of reciprocal monophyly. Reciprocal monophyly is defined as all lineages of species 'A' arriving at a common ancestor before meeting any lineage of species 'B' and vice versa (Kizirian and Donnelly 2004). Here, we define reciprocal monophyly specifically as cases when all lineages of a previously defined clade meet at a common ancestor in a given phylogenetic tree (described above) without encountering any other clade (monophyly) and when all lineages sister to the monophyletic clade are self-contained (reciprocity; i.e. no previously defined clade placed as sister to the focal clade occur elsewhere in the phylogeny; see Supplementary Fig. S4). As the sequence regions underlying each tree are very short, trees do not necessarily reflect true species relationships. As such, monophyly is used here to describe the reconstruction of a pre-defined taxonomic group as a single clade in a given phylogenetic tree and not intended as a statement on true species relationships. In our assessments, we used the criteria of reciprocal monophyly to indicate reliable species diagnosis, monophyly without reciprocity was used to indicate less-reliable species diagnosis, and cases where neither criterion were met to indicate unreliable species diagnosis. Clades were then classified as reciprocally monophyletic, monophyletic, and neither using a custom phylogeny analysis pipeline in R (v4.4.3; R Core Team 2025) that incorporates functions from the *ape* (v5.8.1; Paradis and Schliep 2019), *phytools* (v2.4.4; Revell 2012), and *TreeTools* (v1.13.1; Smith 2019) packages (see supplementary materials on GitHub). We chose a phylogenetic assessment method rather than a similarity-threshold based method to a) incorporate within-species diversity, and b) to avoid the problem of variable similarity-thresholds required for species-level diagnosis across different taxa and markers (Pappalardo et al. 2025). However, eDNA surveys typically utilise genetic-distance or sequence-similarity (e.g. Allen et al. 2023; Gray et al. 2025; Kim et al. 2025; Polanowski et al. 2024) metrics to assign species. Researchers applying these markers in distance- or similarity-based eDNA workflows should therefore evaluate their performance under the relevant assignment metrics, including establishing appropriate species-level discrimination thresholds for their focal taxa.

Clades for the purposes of reciprocal monophyly determination (see Supplementary Table S2) were based initially on species placement in a nuclear tree (Roycroft et al. 2022), and secondly based on placement in a full mitochondrial sequence tree. Samples from the diverse *Uromys caudimaculatus* species complex were excluded due to uncertainty surrounding species classifications. A known backcrossed *Pseudomys novaehollandiae* individual with *Pseudomys mimulus* mitochondrial genome and a known F1-hybrid (Roycroft et al. 2024) were also excluded at this stage (see Supplementary Fig. S5). Known geographic structure of *P. delicatulus* (Roycroft et al. 2024) was incorporated into the reciprocal monophyly analysis. Following these exclusions and considerations, we only assessed reciprocal monophyly for

clades that had at least three representatives retained within the window or barcode. Clades with two representatives were retained as comparison sequences only when they placed as sister taxa in a given window, and clades with one representative were retained only when they had ≥80 bootstrap branch support.

To confirm the validity of our reciprocal monophyly criterion and compare it to other metrics of phylogenetic information content, we ran pairwise linear models comparing the proportion of reciprocally monophyletic clades within each tested 200bp window to mean tree-wide bootstrap support and number of phylogenetically informative sites (see supplementary materials on GitHub). We also ran a generalised linear model with a quasibinomial distribution to test the relationship between likelihood of diagnosis by a mini-barcode length window (proportion of 200bp windows in which a clade meets the criteria of reciprocal monophyly) and time since divergence from closest relative. Finally, we tested the relationship between reciprocal monophyly and sequence length (200bp and 100bp windows) using a paired t-test.

2.4.2 Novel in-silico primer design

We used DeGenPrime (Fulghum et al. 2024) to screen for and design potential primers to amplify windows with the greatest species-level resolution among those tested (according to our reciprocal monophyly criterion, described below in *2.4.1*). We took all 200bp window alignments that could reliably diagnose ≥80% of tested clades according to our criteria (three windows, one each from *ND2*, *ND5*, and *ATP6*). We added a 40bp flanking region on either side of each window (i.e. 280bp total length) to maximise the potential for conserved primer binding regions, subset windows to unique within-species haplotypes to ensure equal consideration of within and between species diversity, and ran DeGenPrime with a minimum amplicon size of 100bp and a primer length of 18-25bp.

DeGenPrime identified nine candidate primer pairs across all three windows. We first aligned these to our complete mitochondrial database alignment with MAFFT's addfragments option to identify and exclude off-target primers. We then tested the species-level diagnostic capacity of the four retained on-target amplicons using our reciprocal monophyly criterion and identified two adjacent *ATP6* markers as the most promising candidates for species-level diagnosis in Australo-Papuan rodents. However, one of the candidate forward primers had high degeneracy (10 three-fold degenerate sites). To correct this, we added an additional 40bp to the forward primer end of the 200bp window (i.e 80bp flanking region), identifying an alternative primer set with acceptable degeneracy (two three-fold degenerate sites in the reverse primer) and a longer amplicon.

Finally, to evaluate the likelihood of off-target amplification we screened our two retained primer sets against GenBank's core_nt database using NCBI's Primer-BLAST (Ye et al. 2012). Both primer sets returned only murine rodents, near-exclusively from the Hydromyini (where available) and Rattini tribes (the singular exception is a West African rodent belonging to the Malacomyini tribe).

3 Results

3.1 Taxonomic representation and completeness of assembled mitochondrial genomes

Of a total set of 654 samples sequenced from nuclear-enriched exon capture and whole exome libraries, 431 yielded mitochondrial sequence data (66%), representing 192 species or distinct phylogenetic lineages, including at least seven extinct species (Fig 1; Supplementary Table S2; S3). Successfully assembled mitogenomes include representatives from nearly all extant genera in all Divisions within the Sahul Hydromyini (95%), and the Australo-Papuan *Rattus* radiation within the Rattini tribe (Fig 1.), and originated from lineages distributed in Australia (241 assemblies across 63 lineages), New Guinea (117 assemblies across 81 lineages), Solomon Islands (five assemblies and lineages), Maluku (11 assemblies and lineages), southeast Asia (six assemblies and lineages), and from widespread species (40 assemblies across 15 lineages) that span multiple geographic regions (Fig. 1). Only two Sahul Hydromyini genera, collectively represented by only three extant species, are not represented in this dataset; *Baiyankamys* (two species) and *Brassomys* (one species). The 34% of libraries that failed to yield any usable mitochondrial data tended to be samples that had also failed to generate nuclear data (e.g. Roycroft et al. 2022), though 19 libraries successfully generated usable mitochondrial data despite unsuccessful nuclear recovery.

Of the 431 samples from which we recovered mitochondrial DNA, 248 yielded complete mitochondrial genomes, with a further 48 at 90% completeness, and 39 at 70% completeness (Fig. 2). Of a total of 192 species or distinct lineages represented in the final dataset, 34 are represented by assemblies from at least three individuals, and a further 45 are represented by two individuals. Australian-distributed species (especially in the Pseudomys Division) tended to have the largest number of individuals represented (Fig. 1; Fig. 2), and also tended to have higher average completeness of assembled mitogenomes (Fig. 2; Fig. 3A). Of the 126 assemblies originating from the island of New Guinea, 114 originate from Papua New Guinea, and nine originate from West Papua (Fig. 2C), with three of uncertain origin. Most species with widespread distributions across New Guinea, therefore, are represented in our database by samples from eastern populations in Papua New Guinea (Fig. 2D). Of 431 samples with

successfully recovered mtDNA, 55 came from toe pad, dry skin or bone subsamples from historical museum material (Fig. 3B), representing 46 species, including 35 that would otherwise not have any mitochondrial reference sequences available.

Using Welch's ANOVA, we found a significant difference (Fig. 3A; $p = 0.0001$) in mitogenome completeness (proportion of identified nucleotides relative to *Mus musculus* RefSeq: NC_005089.1) between geographic regions (i.e., Australia; Indonesia; Papua New Guinea; and Solomon Islands and Bougainville Island), likely highlighting geographic patterns in DNA degradation of available specimens. Post-hoc Games-Howell pairwise comparisons revealed a significant difference ($p = 0.0002$) in mitogenome completeness between Australian ($\bar{x} = 0.72$, $\sigma = 0.41$) and Indonesian ($\bar{x} = 0.42$, $\sigma = 0.44$) distributed lineages, and a significant difference ($p = 0.006$) between Indonesian and Papua New Guinea ($\bar{x} = 0.66$, $\sigma = 0.40$) distributed lineages, with Indonesian distributed lineages having less complete mitogenomes overall than both Australian and Papua New Guinea distributed lineages. Wilcoxon's rank sum test revealed a significant difference (Fig. 3B; $p = 2.75 \times 10^{-37}$) in mitogenome completeness resulting from libraries prepared from historical ($\bar{x} = 0.21$, $\sigma = 0.33$) versus modern ($\bar{x} = 0.79$, $\sigma = 0.35$) specimens, with historical samples generating less complete assemblies overall. Mitochondrial assemblies with <70% nucleotide-completeness more often assembled two target-enriched mitochondrial loci (*COI* and *Cytb*), and their physically adjacent loci, compared to other regions of the genome (Fig. 3C).

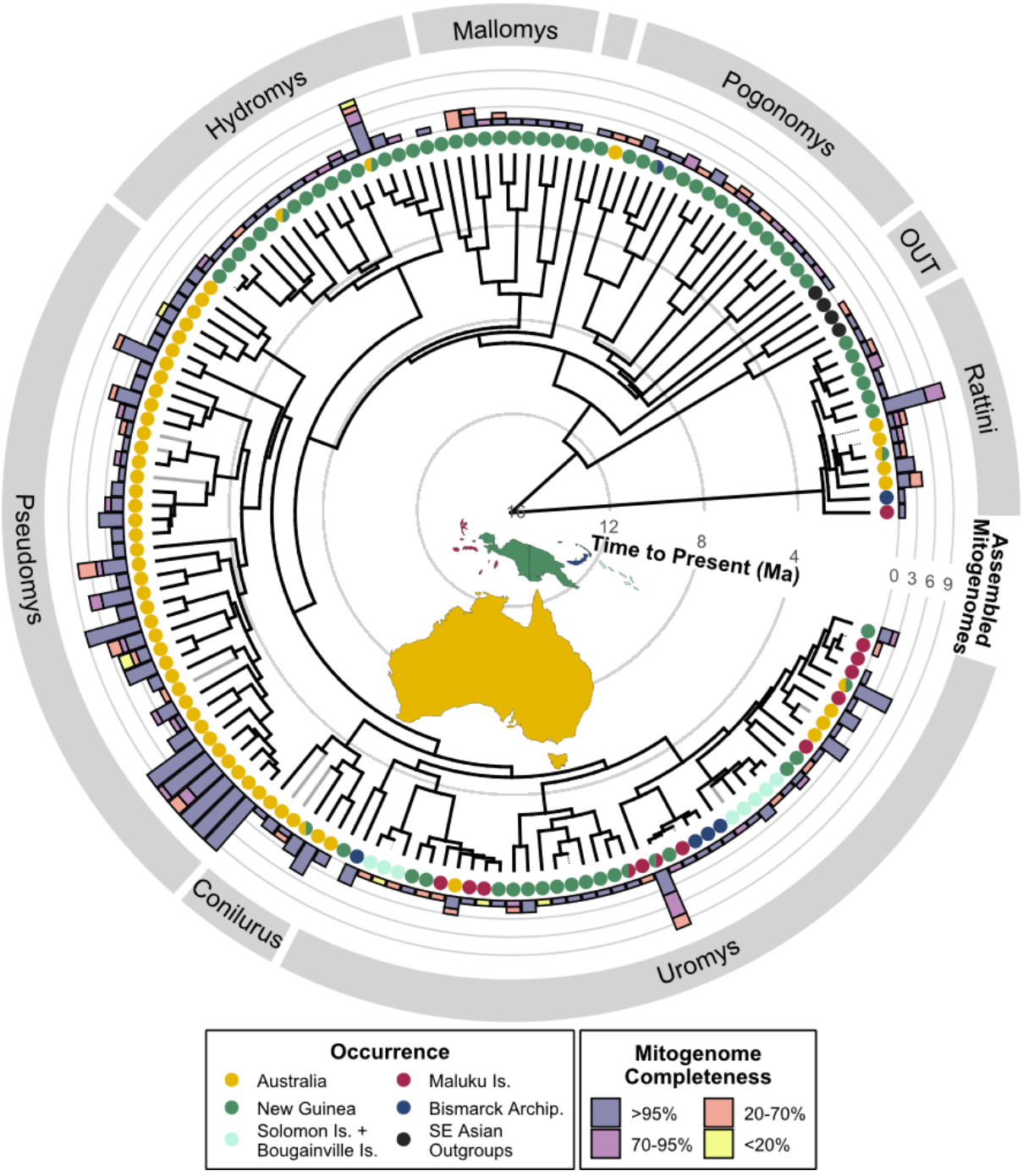


Figure 1. Phylogenetic representation of assembled mitochondrial genomes for Australo-Papuan rodents. Coloured bars show the count and completeness of assembled mitochondrial genomes per-species. Where there are more than 10 assemblies for a species in the dataset, only the 10 most complete assemblies are presented (refer to Supplementary Table S2 for full dataset). Coloured circles show their distribution across geographic regions, corresponding to the central map. In the Hydromyini tribe, clades are labelled by taxonomic Division (The unlabelled group is the *Coccymys* genus which is not assigned to a Division). The Australo-Papuan *Rattus* clade is labelled 'Rattini' and South-east Asian outgroups are labelled 'OUT'. The phylogenetic tree is based on nuclear exon-capture data (Roycroft et al.

2022). Dotted branches show species that were manually added to the nuclear tree. Grey branches represent extinct species. Map boundaries sourced from Natural Earth (South et al. 2024).

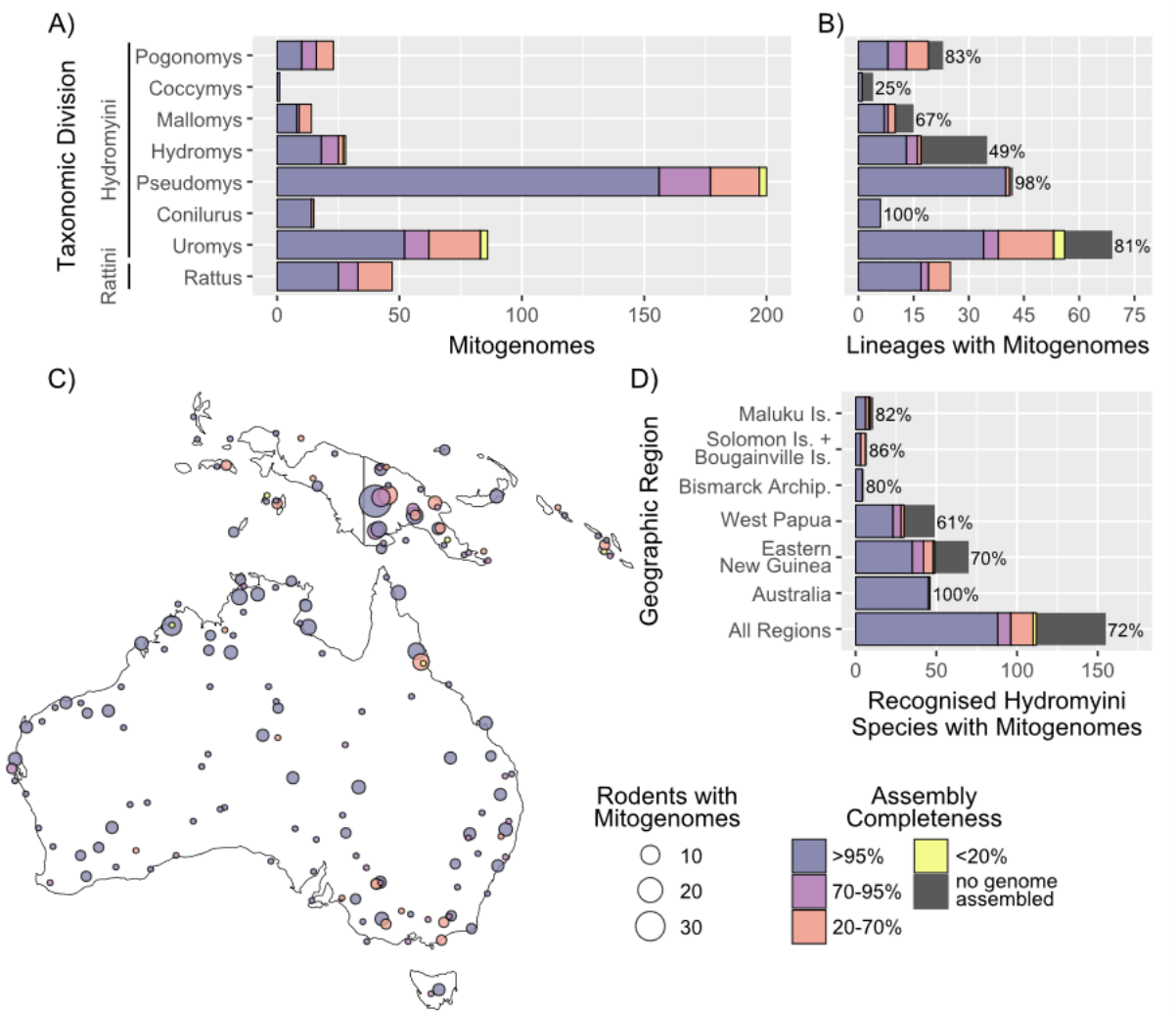


Figure 2. A summary of the completeness of assembled mitochondrial genomes across Australo-Papuan rodents. The count and completeness of each mitogenome is grouped by taxonomic Division, and shown A) by total number of assemblies; and B) by species or distinct phylogenetic lineage as per Roycroft et al. 2022 (hereafter lineages). In B), only the most complete assembly per lineage is shown. C) shows the geographic distribution of assembled mitogenomes. Points are clustered centrally within 100km, provided they fall within the same geographic region according to D). In C) only the 10 most complete assemblies per lineage, per region are presented. D) shows the number of recognised extant Hydromyini species occurring within each region that are represented by a mitochondrial genome in the dataset (with colour indicating the completeness of the most complete assembly). Multi-region species are included across all regions of occurrence regardless of specimen origin (e.g. a species occurring in Eastern New Guinea and West Papua represented only by specimens from

Eastern New Guinea is counted as represented in both regions). Dark grey bars (no genome assembled) indicate the gap between total recognised lineages (B) and species (D) diversity represented and not represented in the dataset, with percentage of total represented diversity indicated. Map boundaries in C) sourced from Natural Earth (South et al. 2024).

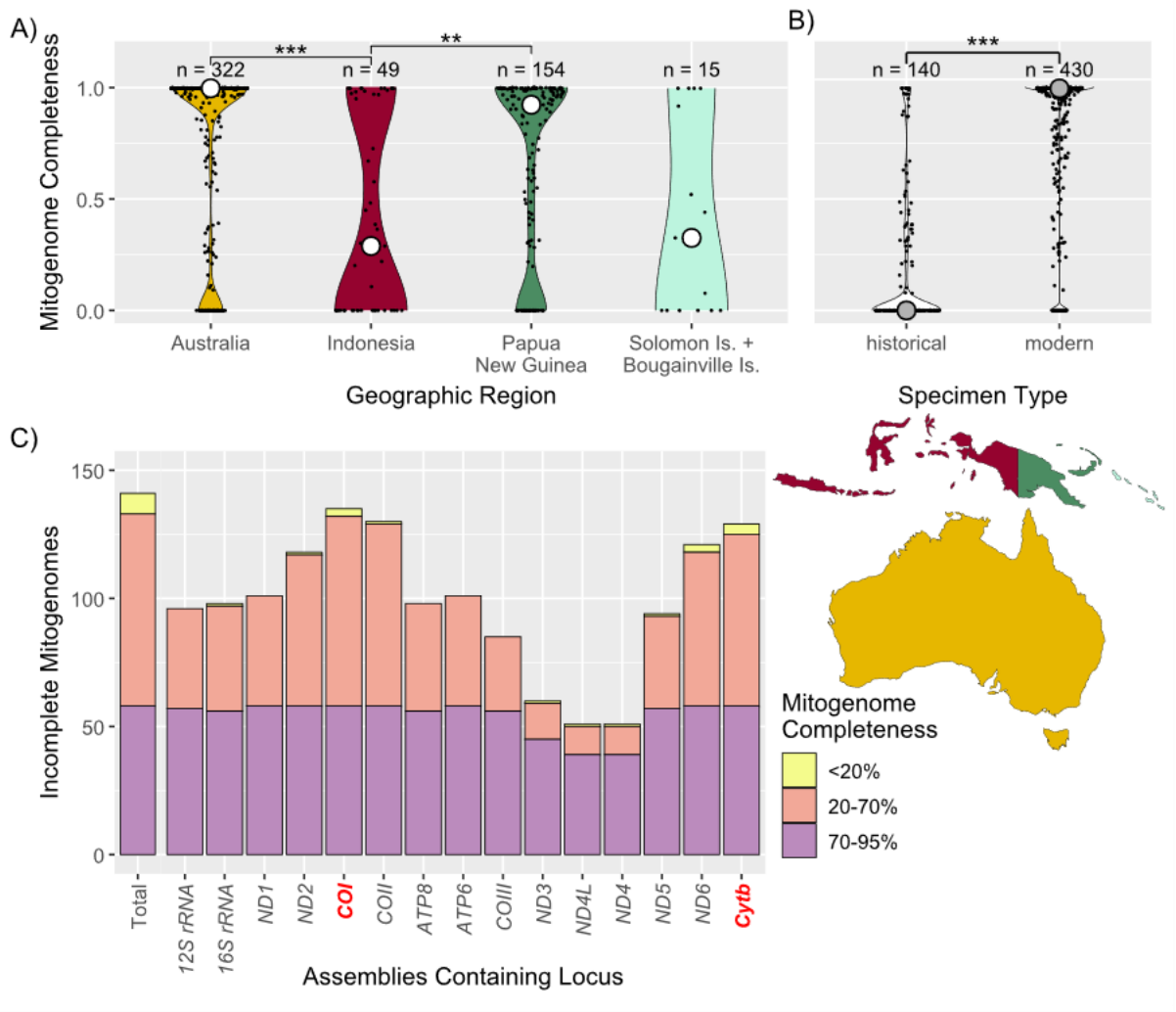


Figure 3. Comparison of the completeness of assembled mitochondrial genomes between A) geographic regions; B) historical and modern specimens; and C) loci. Significant differences in A) and B) are indicated by asterisks ('**' $p < 0.01$; '***' $p < 0.001$). In C), loci are arranged according to position in the Muridae mitochondrial genome (i.e. *12S rRNA* is adjacent to *16S rRNA*, etc.); only data from incomplete mitochondrial assemblies are shown (i.e. <95% complete), to demonstrate that the distribution of recovered regions across incomplete assemblies clusters around targeted loci (*COI* and *Cytb*; Roycroft et al. 2022). Map boundaries sourced from Natural Earth (South et al. 2024).

3.2 Phylogenetic information content across assembled mitochondrial genomes

Leveraging our new 431-mitogenome reference library to explore the most phylogenetically diagnostic regions for potential use in eDNA and species discovery of Australo-Papuan rodents, we first compared non-overlapping 200bp windows from across protein and rRNA encoding mitochondrial genes. Based on mean bootstrap scores, 200bp windows from NADH loci (ND2, ND4, ND5) had the highest phylogenetic information content overall, while *COI* and *Cytb* also contained several windows with similarly high phylogenetic information (Fig. 4). In contrast, rRNA loci had the lowest phylogenetic information content, with the exception of one hyper-informative 200bp window in the *12S rRNA* gene (Fig. 4). These results broadly align with the expected relative nucleotide substitution rates of mitochondrial loci (Pesole et al. 1999) where faster evolving regions are more phylogenetically informative. Though notably, *ATP6* and *ATP8*, the fastest evolving mitochondrial protein encoding genes, only show moderate phylogenetic information content - based on mean bootstrap scores - relative to other regions.

We then compared additional metrics to assess the diagnostic capacity of mitochondrial loci, including mean bootstrap score, the number of parsimony informative sites, and the proportion of reciprocally monophyletic clades across 200bp length windows of mitochondrial protein and rRNA encoding genes (Fig. 5). The proportion of reciprocally monophyletic clades had no significant correlation with mean bootstrap score or number of parsimony informative sites (Fig. 5A-B), while mean bootstrap score and number of parsimony informative sites had a significant positive correlation (Fig. 5C; $p = 0.004$, $R^2 = 0.27$). This is because average bootstrap and parsimony metrics are taxonomically naive, and do not account for cases where taxonomically different species have identical sequences (i.e. where windows could not distinguish between close relatives), while our reciprocal monophyly criterion is taxonomically aware and therefore accounts for such cases.

The proportion of reciprocally monophyletic windows across clades had a significant positive correlation ($p = 0.0063$, $R^2 = 0.21$) with time since divergence (Fig. 5D), indicating that the proportion of reciprocally monophyletic clades is a good proxy for the ability of barcodes to distinguish between close relatives. Among clades that diverged *circa* < 1 million years ago, there is greater variation in the diagnostic power of 200bp length windows. We then compared the proportion of reciprocal monophyly between 200bp and centrally subset 100bp windows to inform suitable length of potential barcodes for eDNA and species discovery. 200bp length windows resolve significantly more clades as reciprocally monophyletic than paired 100bp windows (Fig. 5E; $p = 4.4e\text{-}5$), indicating that longer mini-barcode designs should be prioritised

where possible. However, there were three cases where 100bp windows provided greater resolution (in terms of reciprocal monophyly) than the matched 200bp window, likely due to alignment uncertainty in the longer regions.

Following in-silico screening for potential primer sites in and around the top-scoring windows (based on proportion of clades resolved as reciprocally monophyletic), we identified two potential mini-barcodes for eDNA and species discovery in Sahul Hydromyni species (see Supplementary Table S4). We then investigated the phylogenetic informative content of commonly used and previously published vertebrate mitochondrial barcodes, as a point of comparison to these novel amplicons. A phylogenetic tree of the full-length 'barcode of life', Folmer's region (658bp amplicon from *COI*; Hebert et al. 2003) resolved 29 of 30 Australo-Papuan Hydromyini rodent clades (see Supplementary Table S2) that were eligible for assessment as reciprocally monophyletic in our tests (97%; Fig. 6; Supplementary Table S5). The vertebrate designed eDNA barcode 12S-V5 (108bp amplicon from *12S rRNA*; Riaz et al. 2011) resolved only six of 26 clades as reciprocally monophyletic (23%), with a further three as monophyletic (35%). The mammal designed eDNA barcode MiMammal (170bp amplicon from *12S rRNA*, adjacent to 12S-V5, adapted by Ushio et al. 2017 from the MiFish barcode developed by Miya et al. 2015) resolved nine of 27 clades as reciprocally monophyletic (33%) with a further five as monophyletic (42%). The Rodentia mini-barcode designed by Galan et al. (2012), hereafter called Galan's region (136bp amplicon from *Cytb*) resolved 19 of 29 clades as reciprocally monophyletic (66%) with a further three as monophyletic (76%). The two novel potential mini-barcodes from *ATP6* designed for the Hydromyini tribe presented here (ATP6_7962-163: 163bp amplicon; and adjacent ATP6_8067-183: 183bp amplicon) resolved 21 and 22 of 27 clades as reciprocally monophyletic, respectively (78%; 81%) with one additional clade as monophyletic each (81%; 85%). Collectively the two novel mini-barcodes resolve 25 clades as reciprocally monophyletic (93%; Fig. 6), nearly on par with Folmer's Region, which, at 658bp, is too long to be suitable for most eDNA studies.

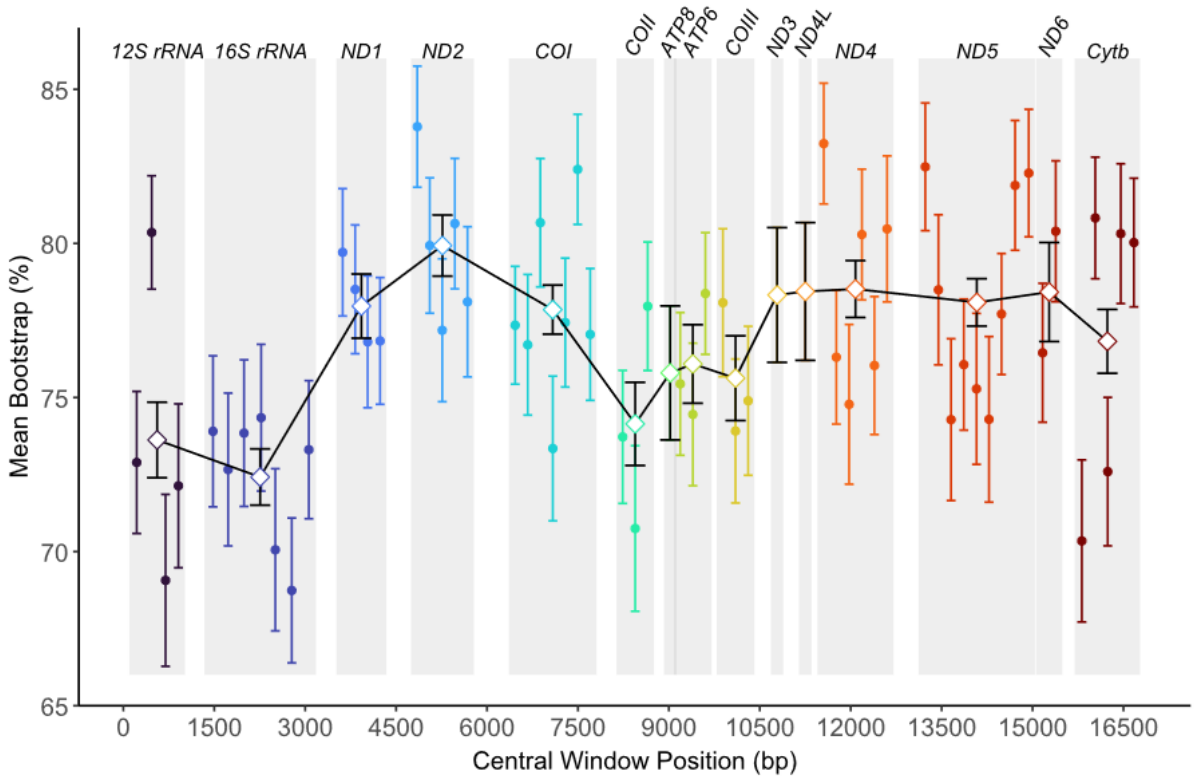


Figure 4. The distribution of phylogenetic information content across the mitochondrial genome of Australo-Papuan rodents. Average bootstrap score was calculated across all bipartitions in phylogenetic trees of 200bp, non-overlapping windows across each protein or rRNA encoding locus in the mitochondrial genome. Windows were taken from the 5' end of all loci apart from ND6 (encoded on the complimentary strand). Windows were controlled for taxonomy by subsetting to one representative from each species with at least 80% nucleotide-completeness across all windows. Error bars represent standard error of the mean.

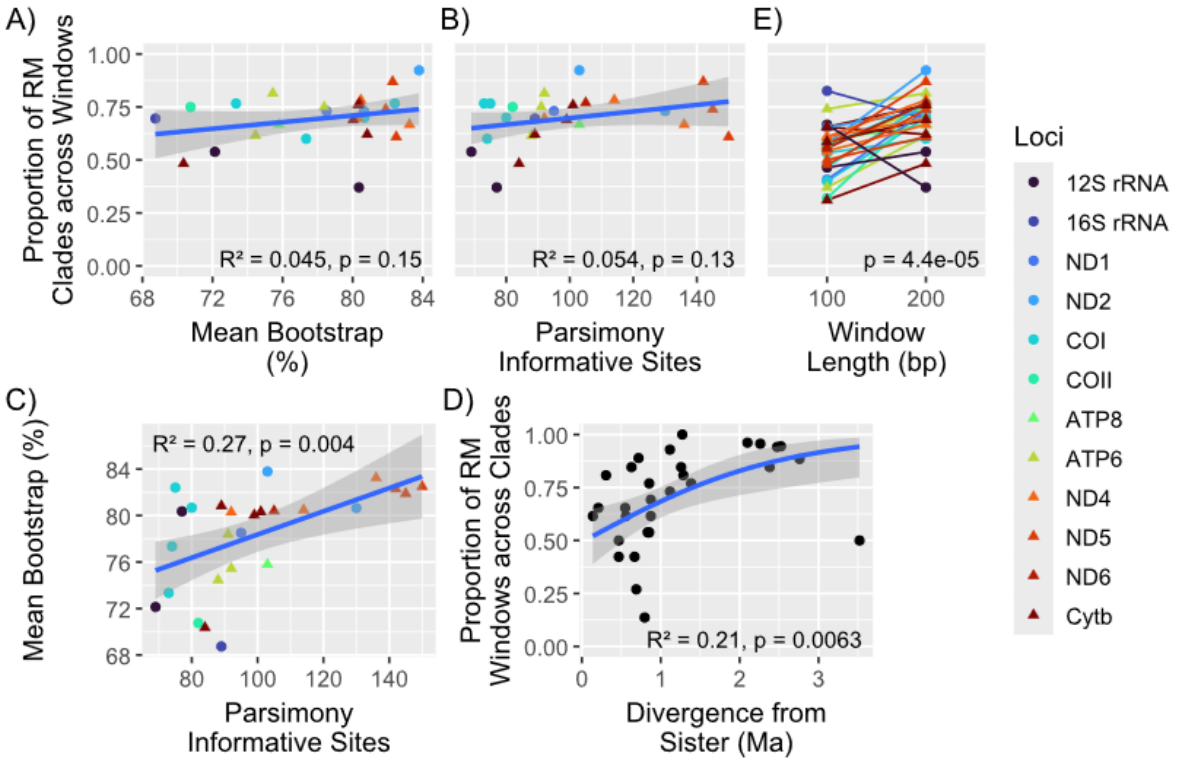


Figure 5. Reciprocal monophyly as a proxy for species-level diagnostic power of mitochondrial mini-barcodes for Australo-Papuan rodents. Panels A), B), and C) compare the phylogenetic information content of 200bp length mitochondrial windows as measured by reciprocal monophyly (RM) versus mean bootstrap score (A), reciprocal monophyly versus parsimony informative sites (B), and mean bootstrap score versus parsimony informative sites (C). Panel D) shows that reciprocal monophyly has a significant positive correlation with time since divergence from sister clade (according to a time-calibrated nuclear phylogeny, Roycroft et al. 2022), indicating its suitability as a proxy for the ability of mini-barcodes to distinguish between close relatives. Panel E) compares the proportion of reciprocally monophyletic clades in 200bp versus paired 100bp windows, indicating that 200bp windows hold significantly greater species-level diagnostic power than 100bp windows. The loci origin of windows are indicated by colour-shape combinations.

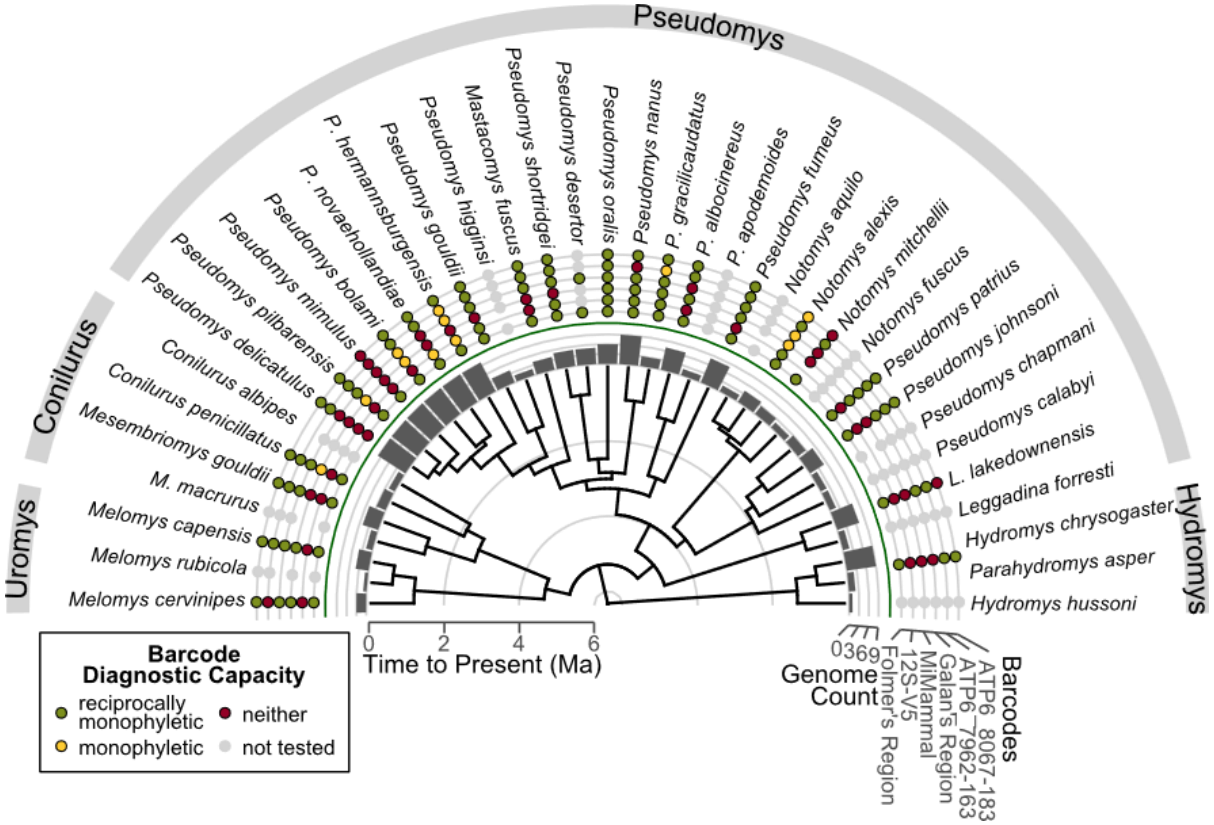


Figure 6. The species-level diagnostic capacity of established and novel barcodes across Australian-distributed Hydromyini rodents. Presented here are species with at least 3 representatives within a given barcode locus, and their close relatives. As a consequence of our sampling, this is primarily Australian-distributed lineages. Species with at least 3 representatives in a given locus are indicated by a coloured circle, those with <3 are indicated by a grey circle (not tested), and those not represented in a given region are indicated by the absence of a circle. Coloured circles indicate the ability of each of six barcodes to diagnose each species as either reciprocally monophyletic, monophyletic, or neither, given the other species represented. Barcode regions tested are Folmer's Region (658bp, Hebert et al. 2003), 12S-V5 (108bp, Riaz et al. 2011), MiMammal (170bp, Ushio et al. 2017), Galan's Region (136bp, Galan et al. 2012), and two novel barcodes designed for Sahul Hydromyini eDNA presented here: ATP6_7962-163 (163bp) and ATP6_8067-183 (183bp). The number of mitochondrial genomes for each species (capped at 10) are shown to indicate within-species representation across barcodes. The phylogenetic tree is based on nuclear exon-capture data (Roycroft et al. 2022) to illustrate time since species divergence.

4 Discussion

4.1 Museum specimens enable the generation of a comprehensive, species-level mitogenome reference database

Here we present a geographically comprehensive and taxonomically validated reference library of mitochondrial genomes for Australo-Papuan rodents, spanning 192 species or distinct lineages from 95% of Australo-Papuan rodent genera, and including all extant described Australian species. Prior to this study, only six species (i.e. ~4% of described hydromyine Australo-Papuan rodent diversity) were represented by a complete mitochondrial genome on GenBank. Here, we contribute mitogenomic resources for at least 50 species that were not previously represented in any online database, and expand significantly on the available sequence data for a further 67 species that were previously only represented by *Cytb*, *COI*, *COII*, and/or *16S rRNA*, and now are represented by at least one complete or nearly complete mitochondrial genome (See Supplementary Table S2; S3). Collectively, this represents a tenfold increase in represented taxonomic diversity, and >40-fold increase in publicly available, complete mitogenomes. The majority of our database is derived from vouchered museum specimens, with all mitochondrial sequences taxonomically verified and linked to corresponding nuclear sequences where available (Roycroft et al. 2022). Through this, our database facilitates stable taxonomic assignment of new tissue and eDNA samples, solving a problem that is pervasive across sequence databases that are generated *ad hoc* (Chorlton 2024; Keck et al. 2023), and follows established best-practice for development of eDNA reference databases (De Brauwer et al. 2023b; Guzik et al. 2025). Through connection to vouchered specimens, taxonomic assignment of these mitogenomes can be updated as taxonomy changes, ensuring their ongoing benefit and relevance to biodiversity research and species monitoring.

4.2 Mitochondrial references facilitate eDNA surveys for long undetected species and biodiversity discovery

Our dataset includes sequence data obtained from both historical and modern museum specimens, including mitochondrial data for seven extinct species, and several highly elusive or possibly extinct species. For example, the critically endangered, possibly extinct Guadalcanal Rat (*Uromys porculus*) of the Solomon Islands, last recorded in the 1880s (Helgen et al. 2016) is represented in our dataset by a nearly complete mitogenome (~98%). Other extinct Australian species in our reference database include the Lesser Stick-nest Rat (*Leporillus apicalis*), White-footed Rabbit Rat (*Conilurus albipes*), and the Short-tailed

Hopping-mouse (*Notomys amplus*), that have not been recorded since 1933, ca. 1860, and 1896, respectively (Burbidge and Woinarski 2024; Woinarski and Burbidge 2024; 2025). Surveys using eDNA are increasingly used to test for the presence of long undetected species, in some cases demonstrating the persistence of species thought to be extinct (e.g. Lopes et al. 2021; Mynhardt et al. 2024). Enabled by our taxonomically validated database of extant, recently extinct, and elusive rodent species, eDNA surveys can now be reliably undertaken in the Australo-Papuan region to test for the presence of long-undetected species.

Furthermore, our database will facilitate species discovery in regions where rodent biodiversity is currently underexplored and difficult to access, including West Papua and parts of the New Guinea highlands (Fisher et al. 2018; Flannery et al. 2026; Myers et al. 2000; Parsons et al. 2022; Travouillon et al. 2026). It is important to note that some geographic regions are less species-complete in our dataset, especially West Papua, and those species that are included are typically represented by incomplete mitogenomes. Some species from West Papua and/or Eastern New Guinea are currently not represented at all by mitochondrial sequences, including *Brassomys* species, *Uromys anak*, *Rattus arrogans*, and *Rattus arfakiensis*, or are only represented by a few loci (e.g. *Baiyankamys habbema* and *Coccymys ruemmleri* represented by partial *Cytb* sequences from prior studies (Rowe et al. 2019; Smissen and Rowe 2018), or *Rattus omlichodes* and *Uromys emmae* represented by ~20% of the mitogenome in our dataset). Our most species-complete region, Australia, now has all but one described extant rodent species represented by a complete mitochondrial genome, with the one exception (*Uromys hadrourus)* represented by sequences from four specimens that span ~77% of the mitogenome (See Supplementary S.2). This geographic bias is likely due to variation in specimen quality as, proportionally, non-Australian lineages, especially West Papua and other Indonesian lineages, were represented by more historical specimen derived raw sequence datasets than Australian lineages (See Supplementary Table S1).

Gaps in mitochondrial reference databases can pose limitations to eDNA surveys, as missing species may be misdiagnosed as close relatives, or unrepresented diversity may be incorrectly inferred as a novel species (Gold et al. 2021; Morey et al. 2024; De Santiago et al. 2025). In some cases, un- or underrepresented species may be identified to the genus level, however the usefulness of this assignment varies across taxa and regions. Thus, it is critical that future studies aim to address remaining reference gaps to improve species-level identification of unrepresented species. Despite this, our comprehensive database opens currently unused or underused avenues for biodiversity discovery and surveying across the Australo-Papuan region, including in West Papua. Furthermore, if best-practice guidelines on archiving metadata and raw sequence data is followed (Gold et al. 2022; Klymus et al. 2024; Takahashi

et al. 2024), eDNA data from past surveys can be reanalysed against more comprehensive reference databases to update biodiversity monitoring results.

4.3 Benefits of a systematic approach to the generation of mitochondrial reference sequences

Reference sequences available in online sequence repositories (i.e. Genbank) are rarely generated in a taxonomically systematic manner, but instead are typically generated *ad hoc* and collated from collective phylogenetic and taxonomic studies from many disparate research groups (Chorlton 2024; Keck et al. 2023; Thomson and Shaffer 2010). Thus, public sequence repositories are often biased toward certain species and loci due to unequal distribution of public interest and funds (Hinchliff et al. 2015; Šmíd 2022; Thomson and Shaffer 2010). The most prevalent loci in online databases are often driven by legacy, where sequencing effort is focused on previously proposed marker loci (e.g. Gold et al. 2021), rather than for their ability to discriminate between species in the group of interest. Meanwhile, elusive species and reference sequences from other loci remain underrepresented or absent, and recognition of these gaps requires dedicated and ongoing assessment (e.g. Šmíd 2022; Schmid et al. 2025; Weigand et al. 2019). Furthermore, *ad hoc* sequences are derived from various sources including field samples and captive animals which may not be identified to the species level (e.g. De Santiago et al. 2025; where novel sequences were only identified to the genus level), or which may become taxonomically uncertain without ongoing reference to source material (see 4.1). This can limit the taxonomic resolution or reliability of online databases as references for eDNA and biodiversity surveys, and further limits the ability to develop new primers and loci that are customised to the focal group under study.

In contrast, through a systematic approach to mitogenome reference generation, we have rapidly improved the species-level availability of mitochondrial sequences for Australo-Papuan rodents - particularly the Hydromyini tribe - and simultaneously highlighted the remaining gaps, overcoming these limitations of *ad hoc* reference generation. Other systematic reference generation studies (e.g. Folk et al. 2021; Gold et al. 2021) have similarly improved the species-level availability of sequence data in their focal taxonomic groups. Furthermore, using our comprehensive species-level database, we have assessed the capacity of existing markers to distinguish Australo-Papuan rodents at the species-level, and – when they were found to be insufficient (see 4.3) – proposed novel markers that may be more suitable. Critically, we were able to propose novel markers because we aimed to assemble complete mitochondrial genomes rather than marker-specific loci. In contrast, Gold et al. (2021) systematically generated only MiFish marker loci from voucher-linked specimens with the aim of assessing taxonomic resolution. They found that a recently diverged clade within their focal group could

not be identified to the species level by MiFish, and therefore additional sequencing effort would be required to facilitate species-level eDNA monitoring of this group. Therefore, frameworks and guidelines defining best-practice reference development for future eDNA studies (e.g. Guzik et al. 2025; Schmid et al. 2025; Weigand et al. 2019) should prioritise generation of complete mitochondrial genomes, where possible, for underrepresented taxa, rather than marker-specific loci, to ensure the ongoing benefit of sequencing effort.

One major barrier to this recommendation is funding, which can be partially alleviated by leveraging by-catch in previously developed sequencing datasets. Although the majority of sequencing libraries utilised here include target-enrichment for two mitochondrial loci, we have successfully recovered off-target loci from these datasets, and recovered complete mitochondrial sequences from datasets without targeted enrichment for any mitochondrial regions (also see Allio et al. 2020; Mulder et al. 2016; Picardi and Pesole 2012; Raposo do Amaral et al. 2015). Due to the high per-cell copy number of the mitochondrial genome, mitochondrial by-catch in existing sequence datasets is likely to be pervasive (Allio et al. 2020; Bogenhagen and Clayton 1974; Picardi and Pesole 2012), and these represent a potentially under-utilised source for rapidly improving the availability of mitochondrial sequences at low cost. Future studies aiming to develop mitochondrial references in a taxonomically systematic manner should consider mining existing sequence datasets first (e.g. via online short-read archives, including NCBI's Sequence Read Archive; SRA and the European Nucleotide Archive; ENA), so that novel sequencing effort can be directed toward filling remaining reference gaps. Overall, a systematic approach to reference generation - combined with leveraging mitochondrial by-catch of existing data - can rapidly provide a comprehensive species-level reference database, simultaneously assess the remaining gaps in that database, and – when paired with marker assessment and development – provide the tools and knowledge to use that database. Thus, overcoming several limitations of *ad hoc* reference generation including taxonomic bias, lack of awareness of existing gaps in reference databases, and underappreciated marker limitations.

4.4 Existing eDNA markers are insufficient for species-level diagnosis in Australo-Papuan rodents

Commonly used eDNA barcodes are often inadequate for species-level identification, particularly between closely related species (e.g. Kelly et al. 2014; Gold et al. 2021). Here, we have shown that the common vertebrate eDNA barcode 12S-V5 (Riaz et al. 2011) has low species-level diagnostic power for Australo-Papuan rodents, at least within the Hydromyini tribe, consistent with assessments in other taxa (e.g. Kelly et al. 2014). The MiMammal barcode (Ushio et al. 2017) – commonly used for mammalian eDNA – is similarly poor.

Notably, this mini-barcode was not originally designed around the evolutionary relationships of mammals, instead it was adapted from a marker designed to diagnose fish (MiFish; Miya et al. 2015) and therefore may be comparatively less suited to species-level diagnosis of mammals. Variation in marker resolution between taxonomic groups is also evident in full-length barcodes like Folmer's Region which is known to be less suitable for diagnosing species in certain taxa (e.g. in flies; Meier et al. 2006), despite high success rates across diverse taxa. Furthermore, both 12S-V5 and MiMammal were developed on existing *ad hoc* reference databases of large taxonomic groups with many target species not represented. Initial assessments of the diagnostic capacity of these barcodes at their development therefore likely underestimated the failure rate between recently diverged lineages, highlighting the importance of both understanding and addressing gaps in reference sequences when designing barcodes. In contrast, a Rodentia-specific *Cytb* mini-barcode (Galan et al. 2012) was able to diagnose twice as many Australo-Papuan rodents (using reciprocal monophyly) as MiMammal, and our Hydromyini-specific *ATP6* mini-barcodes were, together, able to diagnose a further ~31% of clades – though they have not yet been empirically tested with eDNA samples (See Supplementary Table S4). Existing eDNA markers, therefore, should not be assumed to provide species-level resolution in new taxonomic groups without quantitative assessment using a comprehensive species-level reference database (Campbell et al. 2024; De Brauwer et al. 2023a; MacDonald and Sarre 2017; Watts et al. 2026).

Mini-barcodes designed to target broader taxonomic groups (e.g. 12S-V5, MiMammal) may be less informative for biodiversity monitoring at fine taxonomic scale than those designed to target smaller groups (e.g. Galan's Region, or our *ATP6* markers). Focusing on smaller groups facilitates access to primer sites that are not conserved across broader taxonomy, thus allowing a greater choice of marker regions that may be more informative for species-level diagnosis. Alternatively, future studies developing markers for broader taxonomic groups could adopt a diagnosis-first approach – as we have done here (also see Govender et al. 2019; Shi et al. 2025) – rather than the standard primer-first approach (e.g. Hebert et al. 2003; Leray et al. 2013; Miya et al. 2015; Riaz et al. 2011). Through a diagnosis-first approach, potential barcode loci are first screened for species-level resolution and then for potential primer sites. This may identify highly diagnostic regions that could be amplified using degenerate primers or taxon-specific primers that are likely to be missed in a primer-first approach. Furthermore, a diagnosis-first approach may highlight a minimal set of markers that, together, achieve greater resolution than a single marker alone, as in our *ATP6* markers. Although the use of multiple markers increases survey costs, many studies utilise a multi-locus approach to achieve greater species-level resolution, either through a higher-taxonomy metabarcode combined with group-specific markers (e.g. Perry et al. 2026) or through multiple

within-group markers (e.g. Campbell et al. 2024; Miya et al. 2015; Polanowski et al. 2024) as we have proposed here, and such approaches are already recommended best-practice in some frameworks (e.g. De Brauwer et al. 2023a; Guzik et al. 2025; Watts et al. 2026). Overall, there is a tradeoff between the number of species that can be reliably amplified by a single primer set and the diagnostic capacity of the amplicon, particularly with short sequences like those necessary for eDNA, and future studies should consider this tradeoff when developing species-level metabarcodes.

4.5 Implications and recommendations for eDNA and conservation monitoring of Australo-Papuan rodents

In future eDNA research on Australo-Papuan rodents, it is crucial that researchers and conservation managers be aware of limitations where markers may fail to diagnose unknown sequences to the correct species, or where they may have limited taxonomic resolution. One key example elucidated in this database, as well as shown in a previous study (Roycroft et al. 2024) is mitochondrial introgression between the Eastern Delicate Mouse (*Pseudomys mimulus*) and the Pookila (*Pseudomys novaehollandiae*) in northern NSW (Supplementary Fig. S5). This mismatch between the mitochondrial genome and species identity may lead to misdiagnosis of unknown sequences to *P. mimulus* instead of *P. novaehollandiae* regardless of mitochondrial marker choice, and thus result in incorrect presence data. As the Pookila is a federally threatened species, and one of the Australian Government's top 100 priority species under the Threatened Species Strategy Action Plan, this example is particularly important for conservation management.

Researchers and managers also need to be aware of when a marker cannot reliably diagnose target species (See Fig. 6; Supplementary Table S5), and whether or not indistinguishable species co-occur within a survey region. Where markers cannot reliably discriminate certain species, geographic origin of the eDNA sample can inform probable identification where such species do not co-occur (e.g. Allen et al. 2023; Kim et al. 2025), although this also requires an understanding of the spatial distribution of eDNA across substrates (Gold et al. 2022; Watts et al. 2026). Researchers and managers should therefore consider whether close relatives co-occur within their survey region - or occur within range of eDNA distribution, and whether or not distinguishing these species is critical to answer their questions. If critical, researchers and managers should consider using multiple markers that are collectively capable of diagnosing target species, or alternatively using species-specific barcode assays if such markers are unavailable or inefficient. Here, we have identified two short *ATP6* markers that appear promising for Australo-Papuan rodents, although the primers still need to be evaluated on DNA samples (Supplementary Table S4) and researchers using alternative species

assignment workflows (e.g. genetic distance/similarity) will need to evaluate marker performance under the relevant metric. Alternatively, Galan's region is also a promising candidate, depending on specific research goals (e.g. surveys of diagnosable species or where diagnosis to a small set of candidate species is suitable).

More broadly, researchers and managers should be aware that there may be underappreciated marker limitations and reference gaps that prevent or limit eDNA surveys in other taxonomic groups. Without assessment of these limitations and addressing these gaps, current eDNA-based monitoring and biodiversity discovery efforts, globally, may be under-powered or misleading. Future studies should therefore focus on addressing mitochondrial reference gaps with whole mitogenome assembly in a taxonomically systematic manner - linked to vouchered museum specimens and nuclear sequencing where possible, providing the resources to assess existing markers and develop novel markers as required to facilitate and make recommendations for best practice biodiversity monitoring.

References

Allen, M. C., Kwait, R., Vastano, A., Kisurin, A., Zoccolo, I., Jaffe, B. D., Angle, J. C., Maslo, B., & Lockwood, J. L. (2023). Sampling environmental DNA from trees and soil to detect cryptic arboreal mammals. *Scientific Reports*, *13*(1), 180. https://doi.org/10.1038/s41598-023-27512-8

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., & Delsuc, F. (2020). MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour*, *20*(4), 892-905. https://doi.org/10.1111/1755-0998.13160

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., & Ferrer, E. A. (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, *471*(7336), 51-57. https://doi.org/10.1038/nature09678

Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome Res*, *14*(5), 988-995. https://doi.org/10.1101/gr.1865504

Bogenhagen, D., & Clayton, D. A. (1974). The Number of Mitochondrial Deoxyribonucleic Acid Genomes in Mouse L and Human HeLa Cells: QUANTITATIVE ISOLATION OF MITOCHONDRIAL DEOXYRIBONUCLEIC ACID. *Journal of Biological Chemistry*, *249*(24), 7991-7995. https://doi.org/10.1016/S0021-9258(19)42063-2

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Broadhurst, H. A., Gregory, L. M., Bleakley, E. K., Perkins, J. C., Lavin, J. V., Bolton, P., Browett, S. S., Howe, C. V., Singleton, N., Tansley, D., Sales, N. G., & McDevitt, A. D. (2021). Mapping differences in mammalian distributions and diversity using environmental DNA from rivers. *Sci Total Environ*, *801*, 149724. https://doi.org/10.1016/j.scitotenv.2021.149724

Burbidge, A. (2023). Australian terrestrial mammals: how many modern extinctions? *Australian Mammalogy*, *46*. https://doi.org/10.1071/AM23037

Burbidge, A. A., Woinarski, J.C.Z. (2024). Notomys amplus. *The IUCN Red List of Threatened Species*. https://doi.org/10.2305/IUCN.UK.2024-2.RLTS.T14861A254618139.en

Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner.

Campbell, C. D., MacDonald, A. J., & Sarre, S. D. (2024). Evaluation of genetic markers for the metabarcoding of Australian marsupials from predator scats. *Wildlife Research*, *51*(7). https://doi.org/10.1071/wr23134

Chevreux, B., Wetter, T., & Suhai, S. (1999). Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. *German Conference on Bioinformatics*.

Chorlton, S. D. (2024). Ten common issues with reference sequence databases and how to mitigate them. *Front Bioinform*, *4*, 1278228. https://doi.org/10.3389/fbinf.2024.1278228

De Brauwer, M., Clarke, L. J., Chariton, A., Cooper, M. K., de Bruyn, M., Furlan, E., MacDonald, A. J., Rourke, M. L., Sherman, C. D. H., Suter, L., Villacorta-Rath, C., Zaiko, A., & Trujillo-González, A. (2023a). Best practice guidelines for environmental DNA biomonitoring in Australia and New Zealand. *Environmental DNA*, *5*(3), 417-423. https://doi.org/https://doi.org/10.1002/edn3.395

De Brauwer, M., Deagle, B., Dunstan, P., & Berry, O. (2023b). *Integrating environmental DNA science into Australia's marine parks: a roadmap*. CSIRO.

De Santiago, A., Pereira, T. J., Ferrero, T. J., Barnes, N., Lallias, D., Creer, S., & Bik, H. M. (2025). Persistent gaps and errors in reference databases impede ecologically meaningful taxonomy assignments in 18S rRNA studies: a case study of terrestrial and marine nematodes. *Environ DNA*, *7*(2). https://doi.org/10.1002/edn3.70080

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*, *45*(4), e18. https://doi.org/10.1093/nar/gkw955

Eldridge, M. D. B., Deakin, J. E., MacDonald, A. J., Byrne, M., Fitzgerald, A., Johnson, R. N., Moritz, C., Palmer, S., & Young, A. (2020). The Oz Mammals Genomics (OMG) initiative: developing genomic resources for mammal conservation at a continental scale. *Australian Zoologist*, *40*(3), 505-509. https://doi.org/10.7882/az.2020.003

EPBC Act List of Threatened Fauna. (2026). Retrieved from https://www.environment.gov.au/cgi-bin/sprat/public/publicthreatenedlist.pl

Firke, S. (2023). *janitor: Simple Tools for Examining and Cleaning Dirty Data*. In (Version 2.2.0) https://CRAN.R-project.org/package=janitor

Fisher, M. A., Vinson, J. E., Gittleman, J. L., & Drake, J. M. (2018). The description and number of undiscovered mammal species. *Ecology and Evolution*, *8*(7), 3628-3635. https://doi.org/https://doi.org/10.1002/ece3.3724

Flannery, T. F., Aplin, K. P., Bocos, C., Koungoulos, L. G., & Helgen, K. M. (2026). Found alive after 6,000 years: modern records of an ‘extinct’ Papuan marsupial, Dactylonax kambuayai (Marsupialia: Petauridae), with a revision of the systematics and zoogeography of the genus Dactylonax. *Records of the Australian Museum*, *78*(1), 17–34. https://doi.org/10.3853/j.2201-4349.78.2026.3003

Folk, R. A., Kates, H. R., LaFrance, R., Soltis, D. E., Soltis, P. S., & Guralnick, R. P. (2021). High-throughput methods for efficiently building massive phylogenies from natural history collections. *Appl Plant Sci*, *9*(2), e11410. https://doi.org/10.1002/aps3.11410

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*, *3*(5), 294-299.

Fulghum, B., Tanker, S. H., & White, R. A., 3rd. (2024). DeGenPrime provides robust primer design and optimization unlocking the biosphere. *Bioinform Adv*, *4*(1), vbae044. https://doi.org/10.1093/bioadv/vbae044

Galan, M., Pagès, M., & Cosson, J. F. (2012). Next-generation sequencing for rodent barcoding: species identification from fresh, degraded and environmental samples. *PLoS One*, *7*(11), e48374. https://doi.org/10.1371/journal.pone.0048374

Garnier, S., Ross, N., Rudis, R., Camargo, A. P., Sciaini, M., & Scherer, C. (2023). *viridis(Lite) - Colorblind-Friendly Color Maps for R*. In (Version 0.4.2) https://sjmgarnier.github.io/viridis/

Gertz, E. M., Yu, Y. K., Agarwala, R., Schäffer, A. A., & Altschul, S. F. (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol*, *4*, 41. https://doi.org/10.1186/1741-7007-4-41

Gold, Z., Curd, E. E., Goodwin, K. D., Choi, E. S., Frable, B. W., Thompson, A. R., Walker, H. J., Jr., Burton, R. S., Kacev, D., Martz, L. D., & Barber, P. H. (2021). Improving metabarcoding taxonomic assignment: A case study of fishes in a large marine ecosystem. *Mol Ecol Resour*, *21*(7), 2546-2564. https://doi.org/10.1111/1755-0998.13450

Gold, Z., Wall, A. R., Schweizer, T. M., Pentcheff, N. D., Curd, E. E., Barber, P. H., Meyer, R. S., Wayne, R., Stolzenbach, K., Prickett, K., Luedy, J., & Wetzer, R. (2022). A manager's guide to using eDNA metabarcoding in marine ecosystems. *PeerJ*, *10*, e14071. https://doi.org/10.7717/peerj.14071

Gomes, J., Bagnaschi, E., Campos, I., David, M., Alves, L., Martins, J., Pina, J., López-García, A., & Orviz, P. (2018). Enabling rootless Linux Containers in multi-user environments: The udocker tool. *Computer Physics Communications*, *232*, 84-97. https://doi.org/https://doi.org/10.1016/j.cpc.2018.05.021

Govender, A., Groeneveld, J., Singh, S., & Willows-Munro, S. (2019). The design and testing of mini-barcode markers in marine lobsters. *PLoS One*, *14*(1), e0210492. https://doi.org/10.1371/journal.pone.0210492

Gray, E. L., Walker, F. M., Sanchez, D. E., Marriott, S. J., & Baker, A. M. (2025). A pilot study using eDNA collected from soil and active air samplers to detect terrestrial vertebrates in an open grassland habitat of central Queensland, Australia. *BMC Research Notes*, *18*(1), 237. https://doi.org/10.1186/s13104-025-07302-3

Guzik, M. T., Stringer, D. N., Thornhill, J., Coates, P. J., van der Heyde, M., Hillyer, M. J., White, N. E., Saccò, M., Beasley-Hall, P. G., Humphreys, W. F., Harvey, M. S., Huey, J. A., Wilson, N. G., Alexander, J., Humphreys, G., King, R. A., Cooper, S. J. B., Pinder, A., Perina, G., . . . Austin, A. D. (2025). What are the best practices for curating eDNA custom barcode reference libraries? A case study using Australian subterranean fauna. *Biological Journal of the Linnean Society*, *146*(1). https://doi.org/10.1093/biolinnean/blaf053

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. *Nucleic Acids Res*, *41*(13), e129. https://doi.org/10.1093/nar/gkt371

Hajibabaei, M., Leese, F., Abbott, C., Araki, H., Baird, D. J., Beja, P., Deiner, K., Doi, H., Fahner, N., Goodwin, K. D., Hunter, M. E., Kelly, R. P., Klymus, K., Kondoh, M., Mapholi, N. O., Minamoto, T., Pochon, X., Porter, T. M., Theroux, S., . . . Meissner, K. (2026). The International eDNA Standardization Task Force (iESTF): toward development of inclusive international standards for eDNA biodiversity monitoring. *Metabarcoding and Metagenomics*, *10*, e193334. https://doi.org/10.3897/mbmg.10.193334

Hajibabaei, M., Smith, M. A., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B., & Hebert, P. D. N. (2006). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, *6*(4), 959-964. https://doi.org/https://doi.org/10.1111/j.1471-8286.2006.01470.x

Hebert, P. D., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc Biol Sci*, *270*(1512), 313-321. https://doi.org/10.1098/rspb.2002.2218

Helgen, K., Leary, T., & Wright, D. (2016). Uromys porculus. *The IUCN Red List of Threatened Species*. https://doi.org/10.2305/IUCN.UK.2016-1.RLTS.T22805A22446647.en

Hinchliff, C. E., Smith, S. A., Allman, J. F., Burleigh, J. G., Chaudhary, R., Coghill, L. M., Crandall, K. A., Deng, J., Drew, B. T., Gazis, R., Gude, K., Hibbett, D. S., Katz, L. A., Laughinghouse, H. D. t., McTavish, E. J., Midford, P. E., Owen, C. L., Ree, R. H., Rees, J. A., . . . Cranston, K. A. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc Natl Acad Sci U S A*, *112*(41), 12764-12769. https://doi.org/10.1073/pnas.1423041112

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol*, *35*(2), 518-522. https://doi.org/10.1093/molbev/msx281

Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, *33*(6), 1635-1638. https://doi.org/10.1093/molbev/msw046

Jühling, F., Pütz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., & Stadler, P. F. (2012). Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. *Nucleic Acids Res*, *40*(7), 2833-2845. https://doi.org/10.1093/nar/gkr1131

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*, *14*(6), 587-589. https://doi.org/10.1038/nmeth.4285

Karlicki, M., Antonowicz, S., & Karnkowska, A. (2022). Tiara: deep learning-based classification system for eukaryotic sequences. *Bioinformatics*, *38*(2), 344-350. https://doi.org/10.1093/bioinformatics/btab672

Kassambara, A. (2025). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. In (Version 0.7.3) https://CRAN.R-project.org/package=rstatix

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, *30*(4), 772-780. https://doi.org/10.1093/molbev/mst010

Keck, F., Couton, M., & Altermatt, F. (2023). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. *Mol Ecol Resour*, *23*(4), 742-755. https://doi.org/10.1111/1755-0998.13746

Kelly, R. P., Port, J. A., Yamahara, K. M., & Crowder, L. B. (2014). Using environmental DNA to census marine fishes in a large mesocosm. *PLoS One*, *9*(1), e86175. https://doi.org/10.1371/journal.pone.0086175

Kim, G., Kang, Y., & Song, Y. (2025). Multi-primer-based environmental DNA survey for vertebrate biodiversity monitoring in protected island wetlands. *Global Ecology and Conservation*, *58*, e03474. https://doi.org/https://doi.org/10.1016/j.gecco.2025.e03474

Kim, W.-M., Lee, H.-J., & Song, W. (2022). Environmental DNA metabarcoding effectively monitors terrestrial species by using urban green spaces. *Urban Forestry & Urban Greening*, *78*, 127782. https://doi.org/https://doi.org/10.1016/j.ufug.2022.127782

Kizirian, D., & Donnelly, M. A. (2004). The criterion of reciprocal monophyly and classification of nested diversity at the species level. *Molecular Phylogenetics and Evolution*, *32*(3), 1072-1076. https://doi.org/https://doi.org/10.1016/j.ympev.2004.05.001

Klymus, K. E., Baker, J. D., Abbott, C. L., Brown, R. J., Craine, J. M., Gold, Z., Hunter, M. E., Johnson, M. D., Jones, D. N., Jungbluth, M. J., Jungbluth, S. P., Lor, Y., Maloy, A. P., Merkes, C. M., Noble, R., Patin, N. V., Sepulveda, A. J., Spear, S. F., Steele, J. A., . . . Theroux, S. (2024). The MIEM guidelines: Minimum information for reporting of environmental metabarcoding data. *Metabarcoding and Metagenomics*, *8*, e128689. https://doi.org/10.3897/mbmg.8.128689

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res*, *19*(9), 1639-1645. https://doi.org/10.1101/gr.092759.109

Leempoel, K., Hebert, T., & Hadly, E. A. (2020). A comparison of eDNA to camera trapping for assessment of terrestrial mammal diversity. *Proc Biol Sci*, *287*(1918), 20192353. https://doi.org/10.1098/rspb.2019.2353

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1), 34. https://doi.org/10.1186/1742-9994-10-34

Leunufna, S. (2026). Maluku between Blessing and Curse; Exploitation of Block Masela Oil and Gas Mining Field and the Implication. *European Journal of Applied Sciences*, *14*(01), 503-525. https://doi.org/10.14738/aivp.1401.19963

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & 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. https://doi.org/10.1093/bioinformatics/btv033

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, G. P. D. P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Lindenmayer, D. B., & Likens, G. E. (2010). The science and application of ecological monitoring. *Biological Conservation*, *143*(6), 1317-1328. https://doi.org/https://doi.org/10.1016/j.biocon.2010.02.013

Loeza-Quintana, T., Abbott, C. L., Heath, D. D., Bernatchez, L., & Hanner, R. H. (2020). Pathway to Increase Standards and Competency of eDNA Surveys (PISCeS)—Advancing collaboration and standardization efforts in the field of eDNA. *Environmental DNA*, *2*(3), 255-260. https://doi.org/https://doi.org/10.1002/edn3.112

Lopes, C. M., Baêta, D., Valentini, A., Lyra, M. L., Sabbag, A. F., Gasparini, J. L., Dejean, T., Haddad, C. F. B., & Zamudio, K. R. (2021). Lost and found: Frogs in a biodiversity hotspot rediscovered with environmental DNA. *Molecular Ecology*, *30*(13), 3289-3298. https://doi.org/https://doi.org/10.1111/mec.15594

965 Lüdecke, D. (2018). sjmisc: Data and Variable Transformation Functions. *Journal of Open*
966 *Source Software*, *3*, 754. https://doi.org/10.21105/joss.00754

967 MacDonald, A. J., & Sarre, S. D. (2017). A framework for developing and validating taxon-
968 specific primers for specimen identification from environmental DNA. *Molecular Ecology*
969 *Resources*, *17*(4), 708-720. https://doi.org/https://doi.org/10.1111/1755-0998.12618

970 Meier, R., Shiyang, K., Vaidya, G., & Ng, P. K. L. (2006). DNA Barcoding and Taxonomy in
971 Diptera: A Tale of High Intraspecific Variability and Low Identification Success. *Systematic*
972 *Biology*, *55*(5), 715-728. https://doi.org/10.1080/10635150600969864

973 Meng, G., Li, Y., Yang, C., & Liu, S. (2019). MitoZ: a toolkit for animal mitochondrial genome
974 assembly, annotation and visualization. *Nucleic Acids Res*, *47*(11), e63.
975 https://doi.org/10.1093/nar/gkz173

976 Meusnier, I., Singer, G. A. C., Landry, J.-F., Hickey, D. A., Hebert, P. D. N., & Hajibabaei, M.
977 (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, *9*(1), 214.
978 https://doi.org/10.1186/1471-2164-9-214

979 Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto,
980 S., Yamanaka, H., Araki, H., Kondoh, M., & Iwasaki, W. (2015). MiFish, a set of universal PCR
981 primers for metabarcoding environmental DNA from fishes: detection of more than 230
982 subtropical marine species. *Royal Society Open Science*, *2*(7), 150088.
983 https://doi.org/10.1098/rsos.150088

984 Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., & Worm, B. (2011). How many species are
985 there on Earth and in the ocean? *PLoS Biol*, *9*(8), e1001127.
986 https://doi.org/10.1371/journal.pbio.1001127

987 Morey, K. C., Myler, E., Hanner, R., & Tetreault, G. (2024). Taxonomic Blind Spots: A
988 Limitation of Environmental DNA Metabarcoding-Based Detection for Canadian Freshwater
989 Fishes. *Environmental DNA*, *6*(6), e70054. https://doi.org/https://doi.org/10.1002/edn3.70054

990 Mulder, K. P., Lourenço, A., Carneiro, M., & Velo-Antón, G. (2016). The complete
991 mitochondrial genome of Salamandra salamandra (Amphibia: Urodela: Salamandridae).
992 *Mitochondrial DNA Part B*, *1*(1), 880-882. https://doi.org/10.1080/23802359.2016.1253042

993 Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000).
994 Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853-858.
995 https://doi.org/10.1038/35002501

Mynhardt, S., Matthew, E., le Roux, J. P., Little, I., Bloomer, P., & Theron, C. (2024). Environmental DNA from soil reveals the presence of a “lost” Afrotherian species. *Biodiversity and Conservation*, *33*(1), 31-50. https://doi.org/10.1007/s10531-023-02728-2

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, *29*(22), 2933-2935. https://doi.org/10.1093/bioinformatics/btt509

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*, *32*(1), 268-274. https://doi.org/10.1093/molbev/msu300

Pappalardo, P., Hemmi, J. M., Machida, R. J., Leray, M., Collins, A. G., & Osborn, K. J. (2025). Taxon-specific BLAST percent identity thresholds for identification of unknown sequences using metabarcoding. *Methods in Ecology and Evolution*, *16*(10), 2380-2394. https://doi.org/https://doi.org/10.1111/2041-210X.70147

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526-528. https://doi.org/10.1093/bioinformatics/bty633

Parsons, D. J., Pelletier, T. A., Wieringa, J. G., Duckett, D. J., & Carstens, B. C. (2022). Analysis of biodiversity data suggests that mammal species are hidden in predictable places. *Proceedings of the National Academy of Sciences*, *119*(14), e2103400119. https://doi.org/10.1073/pnas.2103400119

Pawlowski, J., Apothéloz-Perret-Gentil, L., & Altermatt, F. (2020). Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Molecular Ecology*, *29*(22), 4258-4264. https://doi.org/https://doi.org/10.1111/mec.15643

Pedersen, T. L. (2025). *patchwork: The Composer of Plots*. In (Version 1.3.2) https://CRAN.R-project.org/package=patchwork

Perry, F., Shaw, J. D., Cassey, P., Coutts, A., Lewis, P. N., MacDonald, A. J., Polanowski, A. M., Stark, J. S., & Suter, L. (2026). Passive environmental DNA (eDNA) sampling is effective for monitoring vessel biofouling in the Southern Ocean. *Science of The Total Environment*, *1013*, 181230. https://doi.org/https://doi.org/10.1016/j.scitotenv.2025.181230

Pesole, G., Gissi, C., De Chirico, A., & Saccone, C. (1999). Nucleotide substitution rate of mammalian mitochondrial genomes. *J Mol Evol*, *48*(4), 427-434. https://doi.org/10.1007/pl00006487

Picardi, E., & Pesole, G. (2012). Mitochondrial genomes gleaned from human whole-exome sequencing. *Nature Methods*, *9*(6), 523-524. https://doi.org/10.1038/nmeth.2029

Polanowski, A. M., MacDonald, A. J., Double, M. C., Barrington, J. H. S., Burg, T. M., Wienecke, B., & McInnes, J. C. (2024). The Development of DNA Markers to Resolve Uncertainties of Seabird Bycatch Identification From Longline Fisheries in Australian Waters. *Ecology and Evolution*, *14*(11), e70568. https://doi.org/https://doi.org/10.1002/ece3.70568

R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. In R Foundation for Statistical Computing. https://www.R-project.org/

Raposo do Amaral, F., Neves, L., Resende, M., Mobili, F., Miyaki, C., Pellegrino, K., & Biondo, C. (2015). Ultraconserved Elements Sequencing as a Low-Cost Source of Complete Mitochondrial Genomes and Microsatellite Markers in Non- Model Amniotes. *PLoS One*, *10*. https://doi.org/10.1371/journal.pone.0138446

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217-223. https://doi.org/https://doi.org/10.1111/j.2041-210X.2011.00169.x

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Res*, *39*(21), e145. https://doi.org/10.1093/nar/gkr732

Roe, A. D., & Sperling, F. A. H. (2007). Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution*, *44*(1), 325-345. https://doi.org/https://doi.org/10.1016/j.ympev.2006.12.005

Rowe, K., Achmadi, A., Fabre, P.-H., Schenk, J., Steppan, S., & Esselstyn, J. (2019). Oceanic islands of Wallacea as a source for dispersal and diversification of murine rodents. *Journal of Biogeography*, *46*, 2752-2768. https://doi.org/10.1111/jbi.13720

Rowe, K., Aplin, K., Baverstock, P., & Moritz, C. (2011). Recent and Rapid Speciation with Limited Morphological Disparity in the Genus Rattus. *Systematic Biology*, *60*, 188-203. https://doi.org/10.1093/sysbio/syq092

Roycroft, E. J., Moussalli, A., & Rowe, K. C. (2020). Phylogenomics Uncovers Confidence and Conflict in the Rapid Radiation of Australo-Papuan Rodents. *Systematic Biology*, *69*(3), 431-444. https://doi.org/10.1093/sysbio/syz044

Roycroft, E., Achmadi, A., Callahan, C. M., Esselstyn, J. A., Good, J. M., Moussalli, A., & Rowe, K. C. (2021a). Molecular Evolution of Ecological Specialisation: Genomic Insights from

the Diversification of Murine Rodents. *Genome Biology and Evolution*, *13*(7). https://doi.org/10.1093/gbe/evab103

Roycroft, E., Fabre, P.-H., MacDonald, A. J., Moritz, C., Moussalli, A., & Rowe, K. C. (2022). New Guinea uplift opens ecological opportunity across a continent. *Current Biology*, *32*(19), 4215-4224.e4213. https://doi.org/10.1016/j.cub.2022.08.021

Roycroft, E., Ford, F., Ramm, T., Schembri, R., Breed, W. G., Burns, P. A., Rowe, K. C., & Moritz, C. (2024). Speciation across biomes: Rapid diversification with reproductive isolation in the Australian delicate mice. *Molecular Ecology*, *33*(7), e17301. https://doi.org/https://doi.org/10.1111/mec.17301

Roycroft, E., MacDonald, A. J., Moritz, C., Moussalli, A., Portela Miguez, R., & Rowe, K. C. (2021b). Museum genomics reveals the rapid decline and extinction of Australian rodents since European settlement. *Proceedings of the National Academy of Sciences*, *118*(27), e2021390118. https://doi.org/10.1073/pnas.2021390118

Sales, N. G., McKenzie, M. B., Drake, J., Harper, L. R., Browett, S. S., Coscia, I., Wangensteen, O. S., Baillie, C., Bryce, E., Dawson, D. A., Ochu, E., Hänfling, B., Lawson Handley, L., Mariani, S., Lambin, X., Sutherland, C., & McDevitt, A. D. (2020). Fishing for mammals: Landscape-level monitoring of terrestrial and semi-aquatic communities using eDNA from riverine systems. *Journal of Applied Ecology*, *57*(4), 707-716. https://doi.org/https://doi.org/10.1111/1365-2664.13592

Schmid, S., Straube, N., Albouy, C., Delling, B., Maclaine, J., Matschiner, M., Møller, P. R., Nocita, A., Palandačić, A., Rüber, L., Sonnewald, M., Alvarez, N., Manel, S., & Pellissier, L. (2025). Unlocking natural history collections to improve eDNA reference databases and biodiversity monitoring. *BioScience*, *75*(12), 1083-1095. https://doi.org/10.1093/biosci/biaf140

Shi, C., Guo, Y., Yao, L., Xu, Y., Zhou, J., & Hua, M. (2025). Development of a mitochondrial mini-barcode and its application in metabarcoding for identification of leech in traditional Chinese medicine. *Scientific Reports*, *15*(1), 1698. https://doi.org/10.1038/s41598-024-77913-6

Šmíd, J. (2022). Geographic and taxonomic biases in the vertebrate tree of life. *Journal of Biogeography*, *49*(12), 2120-2129. https://doi.org/https://doi.org/10.1111/jbi.14491

Smissen, P. J., & Rowe, K. C. (2018). Repeated biome transitions in the evolution of Australian rodents. *Molecular Phylogenetics and Evolution*, *128*, 182-191. https://doi.org/https://doi.org/10.1016/j.ympev.2018.07.015

Smith, M. R. (2019). *TreeTools: create, modify and analyse phylogenetic trees*. In Comprehensive R Archive Network.

South, A., Michael, S., & Massicotte, P. (2024). *rnaturalearthdata: World Vector Map Data from Natural Earth Used in 'rnaturalearth'*. In (Version 1.0.0) https://CRAN.R-project.org/package=rnaturalearthdata

Stephenson, P. J., Londoño-Murcia, M. C., Borges, P. A. V., Claassens, L., Frisch-Nwakanma, H., Ling, N., McMullan-Fisher, S., Meeuwig, J. J., Unter, K. M. M., Walls, J. L., Burfield, I. J., do Carmo Vieira Correa, D., Geller, G. N., Montenegro Paredes, I., Mubalama, L. K., Ntiamoa-Baidu, Y., Roesler, I., Rovero, F., Sharma, Y. P., . . . Fumagalli, L. (2022). Measuring the Impact of Conservation: The Growing Importance of Monitoring Fauna, Flora and Funga. *Diversity*, *14*(10), 824. https://www.mdpi.com/1424-2818/14/10/824

Takahashi, M., Frøslev, T. G., Jenkins, G., Laporte, M., Thalinger, B., Jerde, C., Rajbhandari, S., Jeppesen, T. S., Bissett, A., Deagle, B., Hahn, E., Pauperio, J., Newman, P., Klymus, K., & Berry, O. (2024). Best Practice for Publishing Environmental DNA (eDNA) Data According to FAIR Principles. *Biodiversity Information Science and Standards*, *8*, e137742. https://doi.org/10.3897/biss.8.137742

Theroux, S., Sepulveda, A., Abbott, C. L., Gold, Z., Watts, A. W., Hunter, M. E., Klymus, K. E., Hirsch, S. L., Craine, J. M., Jones, D. N., Brown, R. J., Steele, J. A., Takahashi, M., Noble, R. T., & Darling, J. A. (2025). What is eDNA method standardisation and why do we need it? *Metabarcoding and Metagenomics*, *9*, e132076. https://doi.org/10.3897/mbmg.9.132076

Thomson, R. C., & Shaffer, H. B. (2010). Rapid progress on the vertebrate tree of life. *BMC Biology*, *8*(1), 19. https://doi.org/10.1186/1741-7007-8-19

Travouillon, K. J., Bertozzi, T., MacDonald, A. J., Cooper, S. J. B., Flannery, T. F., & Helgen, K. M. (2026). A new species of bandicoot in the genus Peroryctes (Marsupialia: Peramelemorphia) from Kwiyawagi, West Baliem Valley of Papua Province, Indonesia. *Records of the Australian Museum*, *78*(1), 5–16. https://doi.org/10.3853/j.2201-4349.78.2026.3002

Ushio, M., Fukuda, H., Inoue, T., Makoto, K., Kishida, O., Sato, K., Murata, K., Nikaido, M., Sado, T., Sato, Y., Takeshita, M., Iwasaki, W., Yamanaka, H., Kondoh, M., & Miya, M. (2017). Environmental DNA enables detection of terrestrial mammals from forest pond water. *Mol Ecol Resour*, *17*(6), e63-e75. https://doi.org/10.1111/1755-0998.12690

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189-1191. https://doi.org/10.1093/bioinformatics/btp033

Watts, A. W., Gold, Z., Patin, N. V., Adams, N. G., Baker, J. D., El Baidouri, F., Grey, E., Holmes, A. E., Jacobson, K., Jungbluth, S., Kajita, T., Kiledal, A., Lemay, M. A., Montes, E., Muller-Karger, F. E., Miller, J., Ogburn, M. B., Pitz, K. J., Silliman, K., . . . Thompson, L. R. (2026). Guidance and best practices for species identification using eDNA metabarcoding – When do you call a cod a cod? *Metabarcoding and Metagenomics*, *10*, e174450. https://doi.org/10.3897/mbmg.10.174450

Weigand, H., Beermann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A. M., Willassen, E., Wyler, S. A., Bouchez, A., Borja, A., Čiamporová-Zaťovičová, Z., Ferreira, S., . . . Ekrem, T. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of The Total Environment*, *678*, 499-524. https://doi.org/https://doi.org/10.1016/j.scitotenv.2019.04.247

Wheeler, T. J., & Eddy, S. R. (2013). nhmmer: DNA homology search with profile HMMs. *Bioinformatics*, *29*(19), 2487-2489. https://doi.org/10.1093/bioinformatics/btt403

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*, 1686. https://doi.org/10.21105/joss.01686

Wilcox, T. M., Carim, K. J., McKelvey, K. S., Young, M. K., & Schwartz, M. K. (2015). The Dual Challenges of Generality and Specificity When Developing Environmental DNA Markers for Species and Subspecies of Oncorhynchus. *PLoS One*, *10*(11), e0142008. https://doi.org/10.1371/journal.pone.0142008

Woinarski, J. C. Z., Braby, M. F., Burbidge, A. A., Coates, D., Garnett, S. T., Fensham, R. J., Legge, S. M., McKenzie, N. L., Silcock, J. L., & Murphy, B. P. (2019). Reading the black book: The number, timing, distribution and causes of listed extinctions in Australia. *Biological Conservation*, *239*, 108261. https://doi.org/https://doi.org/10.1016/j.biocon.2019.108261

Woinarski, J. C. Z., Burbidge, A.A. (2024). Leporillus apicalis. *The IUCN Red List of Threatened Species*. https://doi.org/10.2305/IUCN.UK.2024-2.RLTS.T11633A254614193.en

Woinarski, J. C. Z., Burbidge, A.A. (2025). Conilurus albipes. *The IUCN Red List of Threatened Species*. https://doi.org/10.2305/IUCN.UK.2025-1.RLTS.T5223A254613685.en

Woinarski, J. C., Burbidge, A. A., & Harrison, P. L. (2015). Ongoing unraveling of a continental fauna: decline and extinction of Australian mammals since European settlement. *Proc Natl Acad Sci U S A*, *112*(15), 4531-4540. https://doi.org/10.1073/pnas.1417301112

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T. W., Li, Y., Xu, X., Wong, G. K., & Wang, J. (2014). SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, *30*(12), 1660-1666. https://doi.org/10.1093/bioinformatics/btu077

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*, 134. https://doi.org/10.1186/1471-2105-13-134

Acknowledgements:

The authors acknowledge and pay our respects to the Traditional Owners and Custodians of the lands on which we work, and of the lands from which specimens used in this study were collected. We acknowledge the contribution of the Oz Mammals Genomics Initiative consortium in the generation of data used in this publication (https://ozmammalsgenomics.com/consortium/). The initiative is supported by funding from Bioplatforms Australia through the Australian Government National Collaborative Research Infrastructure Strategy (NCRIS). We thank the Biomolecular Resource Facility at ANU for sequencing support, Niccy Aitken and Liz Broady for laboratory support. We are grateful to the curators and collections staff at the American Museum of Natural History; Australian Biological Tissue Collection; Australian Museum; Australian National Wildlife Collection; Bishop Museum; Field Museum of Natural History; Museum and Art Gallery of the Northern Territory; Museum Zoologicum Bogoriense; Museums Victoria; Natural History Museum Giacomo Doria; Natural History Museum London; Naturalis Biodiversity Center; Queen Victoria Museum & Art Gallery; Queensland Museum; South Australian Museum; Western Australian Museum and University of Kansas Biodiversity Institute for providing access to specimens, and to staff at University of Canberra; Southern Cross University; Charles Darwin University; Western Australian Government Department of Biodiversity, Conservation and Attractions and Zoos Victoria for the contribution of tissues. CFD was supported by the Commonwealth through an Australian Government Research Training Program Scholarship [DOI: https://doi.org/10.82133/C42F-K220] via Monash University, and ER was supported by funding from the Australian Research Council grant DE240100573.

Author Contributions:

CFD and ER designed the project, ER and KCR performed sampling and laboratory work, with support from AJM and CM, CFD processed and analysed the data, CFD and ER wrote the paper, with contributions from all authors.

Table S1. Summarised metadata for all included raw sequence datasets

See the excel workbook, sheet 1: Full sample metadata, on our linked FigShare repository

S.1 Comparison between three existing mitochondrial genome assembly pipelines

In order to determine the best approach to generate the most complete, high-quality mitochondrial genomes, we compared three alternative mitochondrial genome assembly pipelines: MIRA (v4.0.2; Chevreux et al. 1999) through MITObim (v1.9.1; Hahn et al. 2013), MEGAHIT (Dinghua et al. 2015) through MitoZ (v3.6, modified as per github.com/linzhi2013/MitoZ/issues/212; Meng et al. 2019), and NOVOPlasty (v4.3.5; Dierckxsens et al. 2017) using a test set of 28 short-read sequence datasets. We compared a total of 11 assembly options using alternative arguments offered by these assemblers. Raw reads were trimmed using Trimmomatic (v0.39; Bolger et al. 2014) following documentation recommendations where possible. Assembly methods are summarised as follows:

- MEGAHIT through MitoZ supplied with forward and reverse paired read files;
- MITObim's quick option supplied with concatenated forward, reverse, paired, and unpaired reads;
- MITObim's two-step option supplied with concatenated forward, reverse, paired, and unpaired reads; and
- eight NOVOPlasty options using all combinations of the following:
    - forward and reverse paired only read files versus forward and reverse concatenated paired and unpaired read files;
    - 21 versus 33 kmer length; and
    - use of a reference guide to help resolve repeat regions versus no reference guide

In MITObim and NOVOPlasty methods using a reference guide, the *Mus musculus* reference genome (Ref Seq: NC_005089.1) was used, MitoZ does not require a supplied reference. All methods were run on a High-performance computing system running Ubuntu 20.04 and allocated 8 CPUs. MitoZ and NOVOPlasty methods were allocated 30GB memory and MITObim methods were allocated 64GB memory. MitoZ's annotation module was used to annotate all resulting assembly files. The full pipeline can be found on our GitHub repository listed in our data availability statement.

We found that the two-step MITObim approach tended to assemble the most complete mitochondrial genomes ($\bar{x}$ = 94%, σ = 8.6%; Fig. S1), compared to MitoZ ($\bar{x}$ = 71%, σ = 32.8%;) and NOVOPlasty approaches ($\bar{x}$ = 51%, σ = 44.6%;). However, MITObim methods were unable to identify cases where contaminant sources were present within a sequencing library (e.g. low levels of cross-contamination between samples, potentially via index switching; e.g. Fig. S2). In contrast, MitoZ was able to distinguish reads from multiple sources within a library, and assemble these as separate contigs (e.g. Fig. S2) that could then be manually curated and filtered. NOVOPlasty also reports multiple contigs, however this was usually a conservative approach to integrating overlapping contigs from the same source rather than reporting potential contaminant reads.

We thus developed a two-stage assembly and quality control approach that leveraged the contrasting benefits of both MitoZ and two-step MITObim (Fig. S3). This approach first used the MEGAHIT module of MitoZ to both assemble mitochondrial genomes and screen for contamination. We investigated cases where MitoZ assembled multiple contigs for a single locus (note there are no known gene duplications in Muridae mitogenomes outside of the conserved two-copy tRNA-Leu and tRNA-Ser loci) for potential contamination by examining each sequence and their placement in a phylogenetic tree. Then, for samples with no evidence of cross-contamination present in a library, but where the mitogenome was incomplete after MitoZ assembly, we took samples downstream to MITObim to extend the completeness of the final assemblies where possible.

To screen for previously undetected contamination, we compared the new MITObim assembly to the MitoZ assembly and quantified the proportion of nucleotide differences across overlapping regions. We investigated cases where this proportion exceeded within-species and between-species thresholds – calculated using complete MitoZ assemblies from recently diverged species *Pseudomys shortridgei* and *Pseudomys desertor* – to determine whether we would retain the MitoZ or MITObim assembly downstream. Finally, alignments were manually curated to remove clear errors (e.g. NUMTs) or mis-assembled regions using the Jalview (v2.11.4.0; Waterhouse et al. 2009) alignment viewer. Then we used a phylogenetically-informed approach to further curate the dataset, by excluding any remaining genes from the final dataset if they did not fall in their expected phylogenetic grouping, based on published phylogenetic relationships and taxonomic groupings of Australo-Papuan rodents (Roycroft et al. 2022).

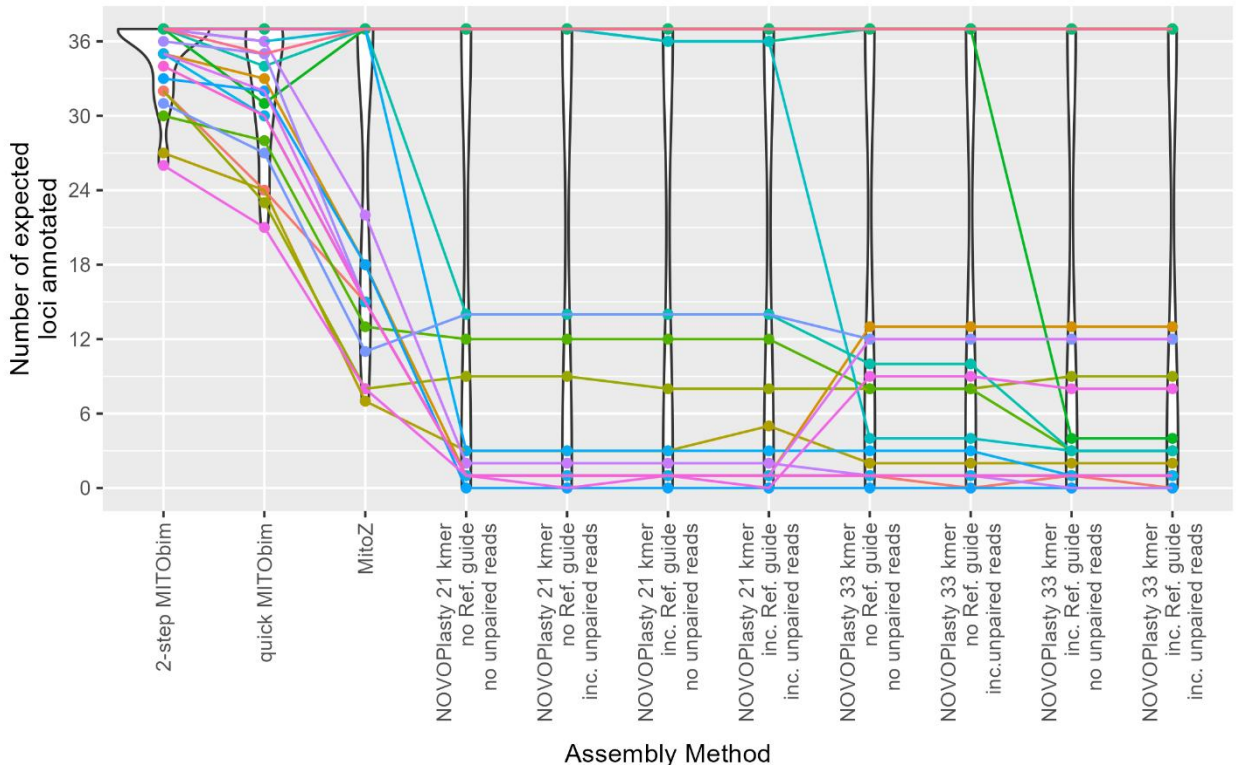


Figure S1. The two-step MITObim approach assembles the most loci-complete mitochondrial genomes of assessed assembly methods. Here we have used the number of expected loci that have been annotated as a proxy for genome completeness to control for when multiple contigs have been assembled for a single locus. Coloured lines connect assemblies generated from the same sequencing library.

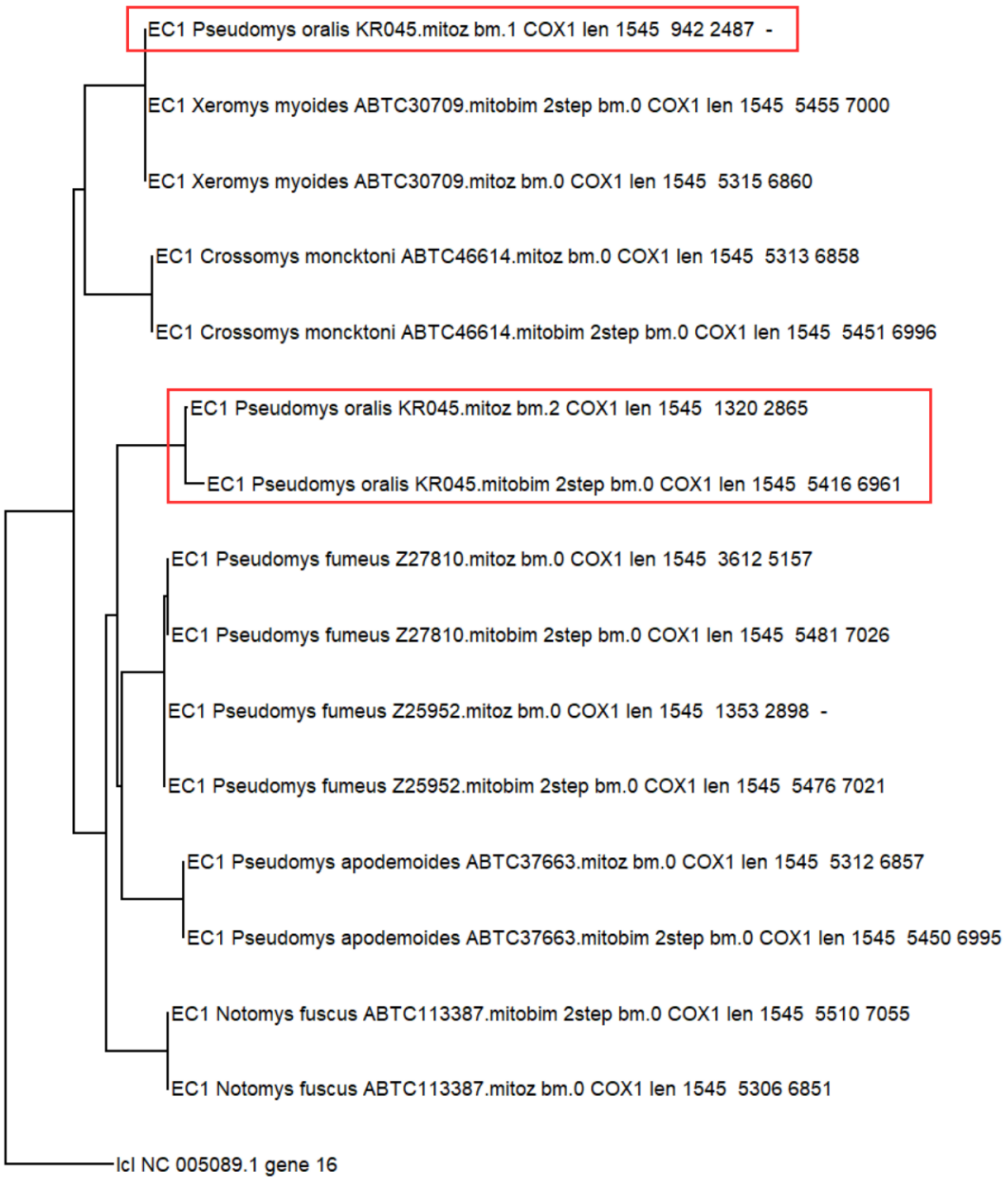


Figure S2. In cases of potential contamination, MitoZ assembles separate contigs for distinct DNA sources while MITObim assembles chimeric sequences. Here we show paired *COI* assemblies from MitoZ and MITObim, from seven samples of six different Hydromyini species. Highlighted in red are example *COI* sequences assembled and annotated from a single library, *Pseudomys oralis* (KR045). MitoZ has assembled two distinct and complete assemblies for *COI*, one with the expected phylogenetic placement and another indicating contamination. The corresponding MITObim *COI* assembly is sister to the correct MitoZ assembly, but the two sequences are not identical, indicating MITObim may have incorporated reads from the contaminant into its assembly. In contrast, all other MITObim and MitoZ *COI* pairs shown demonstrate identical or nearly identical sequences.

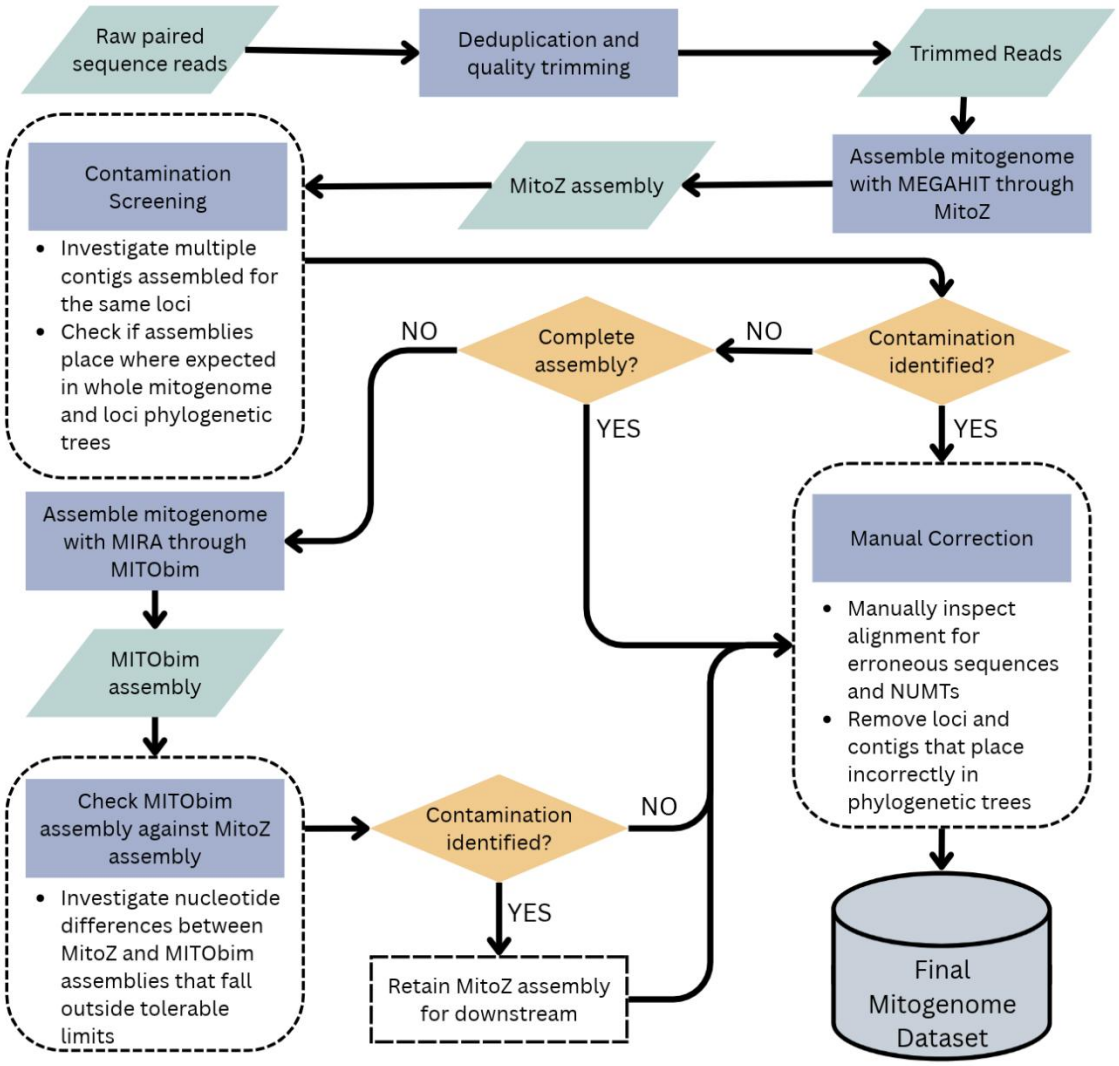


Figure S3. Dual pipeline mitochondrial assembly method flowchart.

Figure S4.

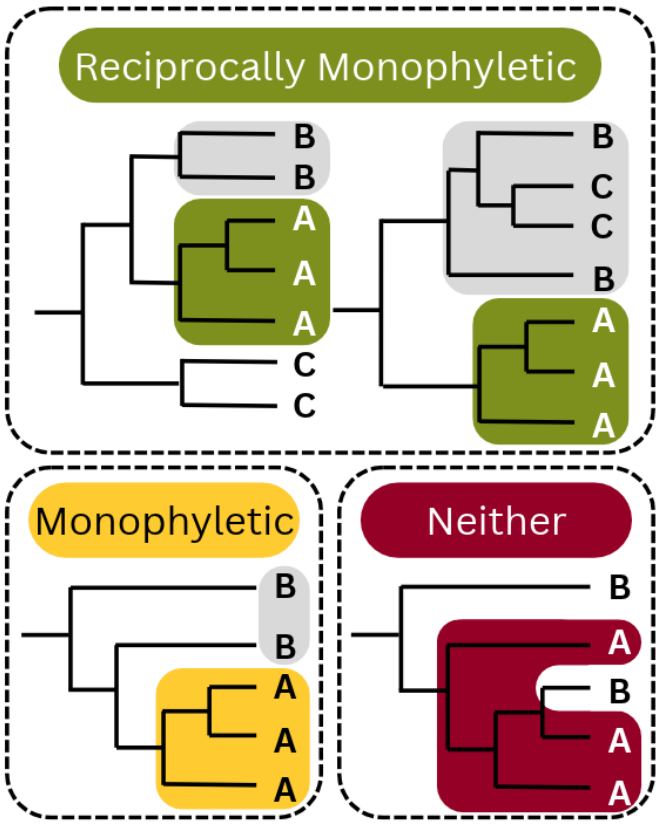


Figure S4. Reciprocal monophyly classifications of focal clade ‘A’. Demonstrated here are two examples of reciprocal monophyly according to our criterion compared with examples of our monophyletic and ‘neither’ classifications. In the top left example, the sister clade is a monophyletic clade of lineage ‘B’ and therefore ‘A’ is reciprocally monophyletic. In contrast, the sister clade in the top right example shows a non-monophyletic configuration of lineages ‘B’ and ‘C’. However, the clade contains all represented lineages of both ‘B’ and ‘C’ and therefore ‘A’ is still classified as reciprocally monophyletic. In the monophyly without reciprocity example (bottom left) the sister to the focal clade ‘A’ is a single representative of lineage ‘B’. As there is another representative of lineage ‘B’ elsewhere in the tree ‘A’ is not reciprocally monophyletic.

Table S2. Metadata and completeness statistics for mitogenome reference database

See the excel workbook, sheet 2: Mitogenome statistics, on our linked FigShare repository

Figure S5.

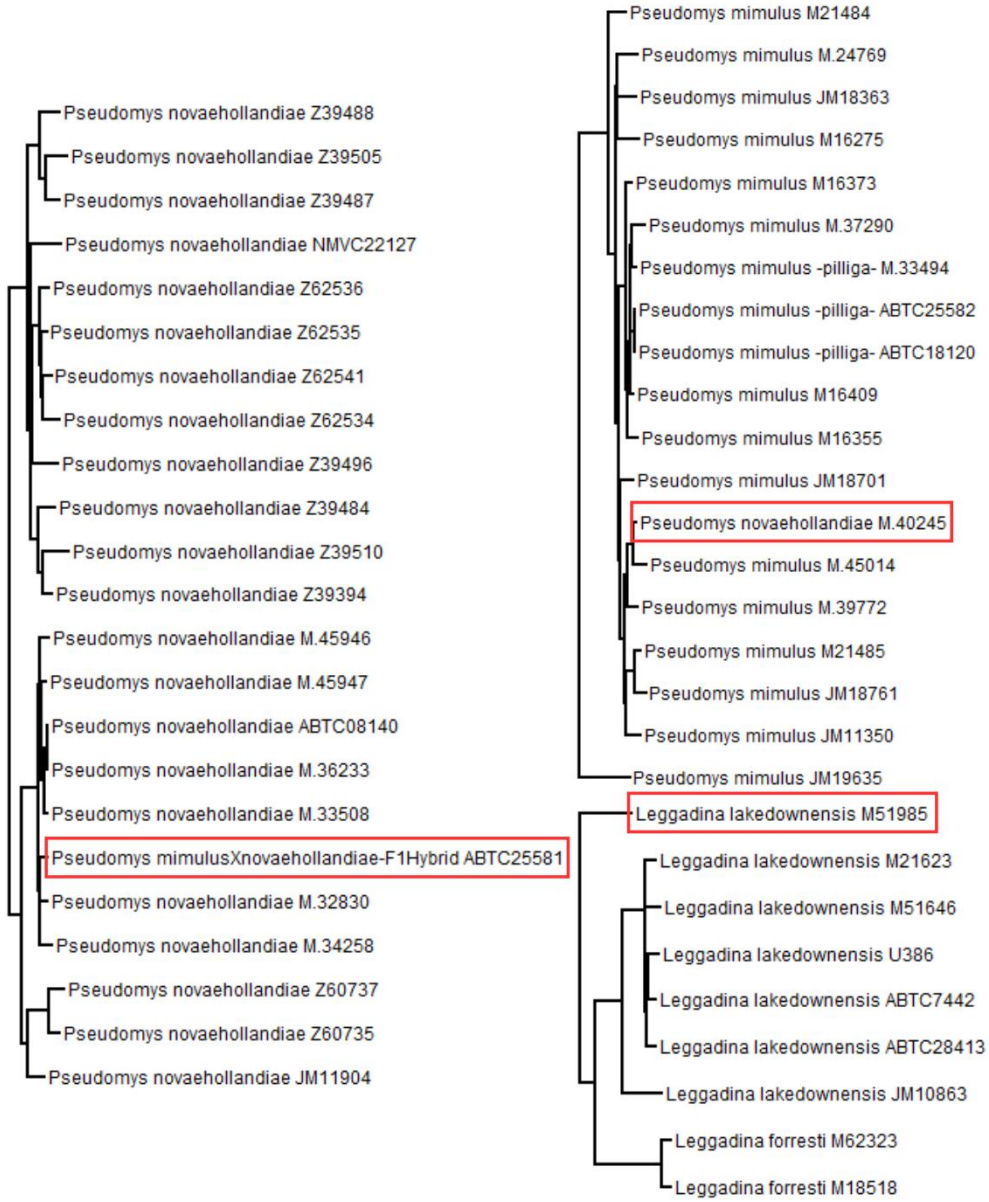


Figure S5. Mitochondrial introgression between *Pseudomys novaehollandiae* and *Pseudomys mimulus* and probable incomplete lineage sorting between *Leggadina lakedownensis* and *Leggadina forresti*. These phylogenies are subsets of a full tree generated from our complete mitogenome assembly database. Highlighted are a known F1 *P. novaehollandiae* and *P. mimulus* hybrid (Roycroft et al. 2024), a known *P. novaehollandiae* individual with a backcrossed *P. mimulus* mitogenome, and a *L. lakedownensis* individual from a distinct population that places as an outgroup to the remaining *Leggadina* clade. All three assemblies were excluded from window and barcode alignments and subsequent phylogenetic trees used for assessment of reciprocal monophyly.

Table S3. Loci completeness statistics for mitogenome reference database

See the excel workbook, sheet 3: Gene level statistics, on our linked FigShare repository

Table S4.

Table S4. Novel *ATP6* primers developed in-silico with DeGenPrime to target highly diagnostic mini-barcodes in Australo-Papuan rodents. Melting temperature is calculated by DeGenPrime. Alignments of unique within-species primer-site haplotypes are available as FASTA files on FigShare for visualisation of nucleotide mismatches.

Barcode Name	Primer Direction	Primer Sequence	Amplicon Length	Melting Temp (°C)
ATP6_7962-163	Forward	5' TTDCCTCTTTTATAACCCCA	163bp	44.62
	Reverse	5' AGACAATTATTAATGCTCAGG		46.55
ATP6_8067-183	Forward	5' CGTCTCCACTCATTCCAACAC	183bp	54.36
	Reverse	5' CTADGATTACDGCTCCTGCTCA		51.78

S.2 Development of a *Uromys hadrourus* assembly for eDNA through pooling two sequence libraries from distinct individuals

Uromys hadrourus is the only described rodent in Australia that, after development of our database, is still not represented by a complete mitochondrial genome. Following our assembly protocol, we assembled a combined total of ~33% of a *U. hadrourus* mitochondrial genome across two specimens (ABTC103222, ABTC103223). The two other *U. hadrourus* raw sequence datasets that were included in our pipeline (ABTC101163, M.12520) did not generate any usable mitogenome sequence. To reduce the impact of this gap on the efficiency of eDNA surveys in Australia, we pooled the reads of the two unsuccessful datasets to improve overall read depth and thus the likelihood of mitogenome sequence recovery. We successfully recovered mitochondrial DNA using this pooled dataset, and thus have a combined total of ~77% of the *U. hadrourus* mitochondrial genome across all four specimens. As this assembly is derived from two distinct individuals, it should only be used to inform eDNA surveys and not in any evolutionary studies. Furthermore, as the assembly is derived from two low-quality sequence datasets, our confidence in the accuracy of this assembly is lower than the others in our dataset, and this confidence should be considered in interpreting eDNA results.

The assembly is available in our FigShare repository. It has not been included in any analysis and is not included in our full mitogenome database statistics or alignment.

Table S5. Full reciprocal monophyly criteria results

See the excel workbook, sheet 4: Reciprocal monophyly data, on our linked FigShare repository

S.3 Additional software dependencies

MitoZ software dependencies:

Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome Res*, *14*(5), 988-995. https://doi.org/10.1101/gr.1865504

Gertz, E. M., Yu, Y. K., Agarwala, R., Schäffer, A. A., & Altschul, S. F. (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol*, *4*, 41. https://doi.org/10.1186/1741-7007-4-41

Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, *33*(6), 1635-1638. https://doi.org/10.1093/molbev/msw046

Jühling, F., Pütz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., & Stadler, P. F. (2012). Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. *Nucleic Acids Res*, *40*(7), 2833-2845. https://doi.org/10.1093/nar/gkr1131

Karlicki, M., Antonowicz, S., & Karnkowska, A. (2022). Tiara: deep learning-based classification system for eukaryotic sequences. *Bioinformatics*, *38*(2), 344-350. https://doi.org/10.1093/bioinformatics/btab672

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res*, *19*(9), 1639-1645. https://doi.org/10.1101/gr.092759.109

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, G. P. D. P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, *29*(22), 2933-2935. https://doi.org/10.1093/bioinformatics/btt509

Wheeler, T. J., & Eddy, S. R. (2013). nhmmer: DNA homology search with profile HMMs. *Bioinformatics*, *29*(19), 2487-2489. https://doi.org/10.1093/bioinformatics/btt403

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T. W., Li, Y., Xu, X., Wong, G. K., & Wang, J. (2014). SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, *30*(12), 1660-1666. https://doi.org/10.1093/bioinformatics/btu077

Assembly and Statistics Pipeline dependencies:

Firke, S. (2023). *janitor: Simple Tools for Examining and Cleaning Dirty Data*. In (Version 2.2.0) https://CRAN.R-project.org/package=janitor

Gomes, J., Bagnaschi, E., Campos, I., David, M., Alves, L., Martins, J., Pina, J., López-García, A., & Orviz, P. (2018). Enabling rootless Linux Containers in multi-user environments: The udocker tool. *Computer Physics Communications*, *232*, 84-97. https://doi.org/https://doi.org/10.1016/j.cpc.2018.05.021

Lüdecke, D. (2018). sjmisc: Data and Variable Transformation Functions. *Journal of Open Source Software*, *3*, 754. https://doi.org/10.21105/joss.00754

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*, 1686. https://doi.org/10.21105/joss.01686

R analysis and figure generation packages:

Garnier, S., Ross, N., Rudis, R., Camargo, A. P., Sciaini, M., & Scherer, C. (2023). *viridis(Lite) - Colorblind-Friendly Color Maps for R*. In (Version 0.4.2) https://sjmgarnier.github.io/viridis/

Kassambara, A. (2025). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. In (Version 0.7.3) https://CRAN.R-project.org/package=rstatix

Pedersen, T. L. (2025). *patchwork: The Composer of Plots*. In (Version 1.3.2) https://CRAN.R-project.org/package=patchwork

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org