Single-molecule Sequencing of an Animal Mitochondrial Genome Reveals Chloroplast-like Architecture and Repeat-mediated Recombination

Joel Sharbrough [1],*,[1] Laura Bankers [2] Emily Cook [1] Peter D. Fields [3] Joseph Jalinsky [2] Kyle E. McElroy [2,4] Maurine Neiman [2] John M. LogsdonJr. [2] and Jeffrey L. Boore [5]

[1]Department of Biology, New Mexico Institute of Mining and Technology, Socorro, NM 87801

[2]Department of Biology, University of Iowa, Iowa City, IA

[3]Zoologisches Institut, University of Basel, Basel, Switzerland

[4]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, IA

[5]Phenome Health and Institute for Systems Biology, Seattle, WA

***Corresponding author:** E-mail: joel.sharbrough@nmt.edu.

Associate editor: Aida Ouangraoua

Abstract

Recent advances in long-read sequencing technology have allowed for single-molecule sequencing of entire mitochondrial genomes, opening the door for direct investigation of the mitochondrial genome architecture and recombination. We used PacBio sequencing to reassemble mitochondrial genomes from two species of New Zealand freshwater snails, *Potamopyrgus antipodarum* and *Potamopyrgus estuarinus*. These assemblies revealed a ~1.7 kb structure within the mitochondrial genomes of both species that was previously undetected by an assembly of short reads and likely corresponding to a large noncoding region commonly present in the mitochondrial genomes. The overall architecture of these *Potamopyrgus* mitochondrial genomes is reminiscent of the chloroplast genomes of land plants, harboring a large single-copy (LSC) region and a small single-copy (SSC) region separated by a pair of inverted repeats (IRa and IRb). Individual sequencing reads that spanned across the *Potamopyrgus* IRa-SSC-IRb structure revealed the occurrence of a "flip-flop" recombination. We also detected evidence for two distinct IR haplotypes and recombination between them in wild-caught *P. estuarinus*, as well as extensive intermolecular recombination between single-nucleotide polymorphisms in the LSC region. The chloroplast-like architecture and repeat-mediated mitochondrial recombination we describe here raise fundamental questions regarding the origins and commonness of inverted repeats in cytoplasmic genomes and their role in mitochondrial genome evolution.

Key words: **flip-flop recombination, heteroplasmy, mitochondrial DNA, PacBio sequencing, repeat-mediated recombination.**

Article

Introduction

Wide taxonomic sampling enabled by the genomic era has revealed remarkable diversity in the content, architecture, and biology of mitochondrial genomes (Mower et al. 2012; Lavrov and Pett 2016; Johri et al. 2019). The mitochondrial genomes of bilaterian animals have been particularly well-studied. With few (but notable) exceptions, these genomes are of small size (~16–17 kb) (Boore 1999), encode only a tiny fraction of the genes necessary to carry out tasks performed within mitochondria (Timmis et al. 2004), engage in complex co-evolution with their nuclear-encoded interacting partners (Rand et al. 2004; Osada and Akashi 2011; van der Sluis et al. 2015; Sloan, Warren, et al. 2018), have relatively high mutation rates (Wolfe et al. 1987; Denver et al. 2000) and seemingly reduced levels of homologous recombination (Smith and Keeling 2015), and are predominantly maternally inherited (Camus et al. 2022).

Bilaterian animal mitochondrial DNAs (mtDNA) are typically very compact, with few intergenic nucleotides, except for a single noncoding region ranging in size from several dozen to several thousand nucleotides (Ghiselli et al. 2021). In some cases, this noncoding region also contains a third DNA strand that causes a displacement of base pairing in this region, the so-called "D-loop" structure (Brown et al. 2005). Hereafter, we will refer to this noncoding region as the "D-loop region" regardless of whether a D-loop structure has been observed for any particular taxon. The D-loop region has been shown to contain regulatory elements for initiation, termination of transcription, and replication in several taxa (Sbisà et al. 1997). This structure and function are presumed to exist across animals, although obvious similarities in primary sequence or potential secondary structures are not readily identifiable across distantly related taxa. This apparent lack of

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

homology is probably related to the fact that portions of the D-loop region are often hypervariable and can contain repeated sequence elements (Wakeley 1993; Satoh et al. 2016; Formenti et al. 2021). Repetitive structures in the mitochondrial genomes can serve as hotspots for recombination (e.g., Maréchal and Brisson 2010; Davila et al. 2011), leading to the intriguing possibility that repeat elements within the D-loop may allow for extensive intra- and inter-molecular recombination.

The variability and repetitive nature of the D-loop region also likely contribute to why it has been especially difficult to amplify by polymerase chain reaction (PCR) in many animal whole-mitochondrial genome sequencing efforts. Challenges in PCR amplification of the D-loop could be due to its highly biased base composition, the presence of secondary structures that are difficult for the polymerase to read through, or the response of the polymerase, itself a prokaryotic DNA polymerase, to signals for terminating replication of the mtDNA. Barriers to sequencing posed by the D-loop region could also explain why many animal mitochondrial genomes, even though listed as "complete" and "circular" in GenBank, do not report the D-loop region sequence (e.g., Nilsson et al. 2004; Cook et al. 2005; Arnason et al. 2008; Neiman et al. 2010; Sharbrough et al. 2018).

Almost all mitochondrial genome assemblies have relied on bioinformatically assembling sequencing reads generated *via* Sanger sequencing (~700 nucleotides [nt] in length) or next-generation sequencing (NGS; 100–250 nt in length) from preparations of whole DNA or that are enriched for mtDNA. This strategy often fails to successfully assemble repeated elements of a length larger than any individual read that cannot find a unique flanking sequence on each side. This problem is exacerbated by assembly via de Bruijn graphs, which provide computationally tractable assembly of millions of short reads (e.g., Bankevich et al. 2012), but also typically break assemblies at complex structural features (Pop 2009) such as those common in the mitochondrial D-loop regions.

The so-called "third-generation sequencing" reads (i.e., PacBio and Oxford Nanopore) that can span many kilobases offer the tantalizing opportunity to uncover previously undetectable genomic architectural features. The application of these long-read technologies has already led to major advances in the identification of disease-causing variants (Merker et al. 2018), the generation of complete microbial genome assemblies (Koren and Phillippy 2015), the detection of rare mutations and DNA damage (Sloan, Broz, et al. 2018), the production of assembly free transcriptomes (Minio et al. 2019), the description of the complex patterns of mitochondrial recombination in plant mitochondria (Shearman et al. 2016), and a truly complete human genome sequence with each chromosome in a single, uninterrupted contig from one telomere to the other (Nurk et al. 2022).

The New Zealand freshwater snail *Potamopyrgus antipodarum* (fig. 1) is a powerful model for a number of ecological and evolutionary questions, including the maintenance of sex (Lively 1987), consequences of polyploidy (Neiman et al. 2011), host–parasite dynamics (Jokela and Lively 1995; Gibson et al. 2016; Bankers et al. 2017), biology of invasive organisms (Levri and Clark 2015), and mitonuclear coevolution (Neiman et al. 2010; Paczesniak et al. 2013; Sharbrough et al. 2017, 2018; Greimann et al. 2020). To generate the whole-genome resources needed to investigate these fundamental topics, we used a combination of Illumina HiSeq (DNA and RNA), MiSeq, and PacBio long-read technologies to sequence total cellular DNA from an inbred (~25 generations) sexual lineage of *P. antipodarum* and wild-caught specimens of its close relative (Haase 2008, our data), *P. estuarinus*. Here, we used a subset of these sequencing reads to reassemble the mitochondrial genomes of both species. Our complete and circular mitochondrial assemblies report a novel ~1.7 kb repeat structure not detected in previous whole-mitochondrial genome sequencing efforts (Neiman et al. 2010; Sharbrough et al. 2018) associated with intra- and inter-molecular recombination. Long PacBio reads from *P. antipodarum* were able to detect two distinct orientations of the 1.7 kb repeat at a relatively high abundance, reminiscent of a similar observation first made in chloroplast genomes (Palmer 1983) and potentially implicating flip-flop recombination as a mechanism of intramolecular mitochondrial recombination. We also observed high levels of single-nucleotide polymorphism (SNP) diversity in wild-caught *P. estuarinus* samples and numerous recombinant molecules in pooled-MiSeq data, implying that heteroplasmy occurs with high frequency in the population. Indeed, the abundance of recombinant molecules in such a small sample of individuals leads us to speculate that *Potamopyrgus* exhibits high degrees of paternal leakage or even biparental inheritance of mitochondrial DNA. Together, these observations provide critical steps toward understanding the biology of this fascinating natural system, with potentially broad implications for mitochondrial biology, inheritance, and evolution in animals.

Results and Discussion

Mitochondrial Genome Architecture in Potamopyrgus Resembles Chloroplast genomes

Despite very high coverage (>800×, see below), our first efforts to assemble complete and circular *Potamopyrgus* mitochondrial genomes using Illumina HiSeq paired-end reads with SPAdes v. 3.13.0 (Bankevich et al. 2012) were unsuccessful, with the mitochondrial contigs from the de Bruijn assembly graphs forming "barbell" (*P. antipodarum*) and "lollipop" (*P. estuarinus*) structures (supplementary fig. S1, Supplementary Material online) rather than simple circles. Additional attempts to polish and circularize the assembly with Pilon v 1.23 (Walker et al. 2014) were unsuccessful, but we confirmed that the region preventing circularization corresponded exactly to the barbell and lollipop structural anomalies. Using PacBio reads from both species (acquired as a part of the ongoing *P. antipodarum* and *P. estuarinus* nuclear genome sequencing projects), we

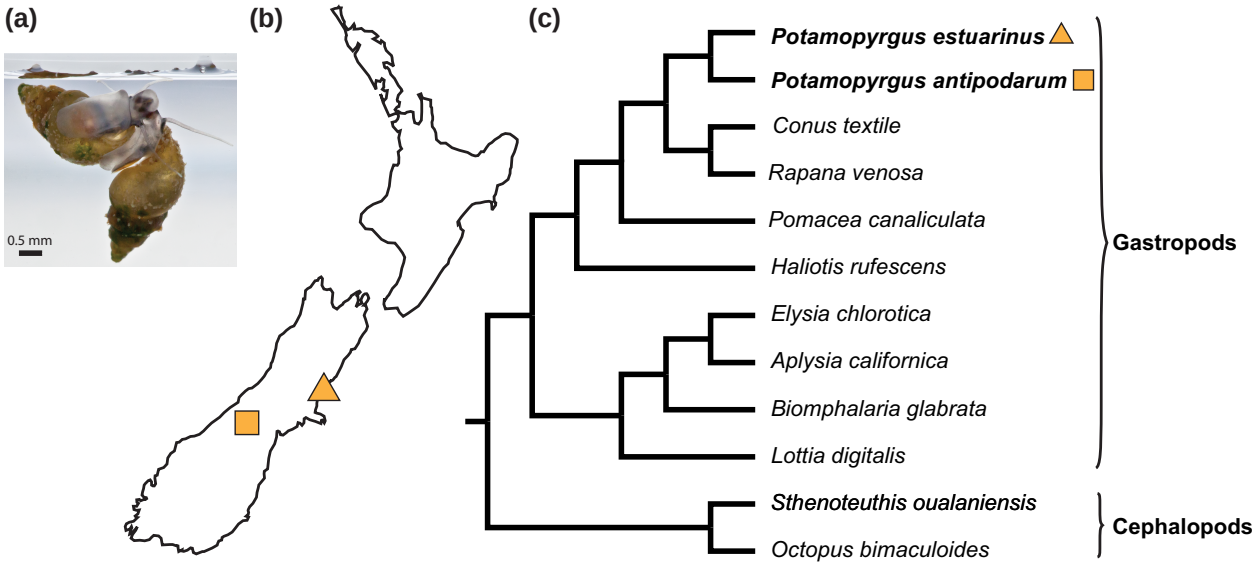

Fig. 1. *Potamopyrgus antipodarum* and *Potamopyrgus estuarinus* are prosobranch snails native to New Zealand. (*a*) *Potamopyrgus antipodarum* is ~4–6 mm in length. (*b*) Map of New Zealand depicting sampling locations of the snails used in this study, and a scaled image of *P. estuarinus*. (*c*) Cladogram of selected mollusk species depicting *Potamopyrgus* within the *Caenogastropoda*.

discovered a previously undetected structural element of ~1.7 kb in length. In both species, the structure consists of a pair of inverted repeats (IRs) interrupted by a small single-copy (SSC) region. In the mtDNAs of *P. antipodarum*, IRa and IRb are identical in sequence, and the intervening SSC consists of two dinucleotide repeats ($TA_{[x54]}$ and $TC_{[x337]}$). In contrast, we detected 15 high-confidence SNPs (i.e., present in both PacBio and MiSeq pool-seq data sets) in the IRs of wild-caught *P. estaurinus*, all of which can be observed in both IRs, and the intervening SSC consists solely of a $TA_{[x334]}$ dinucleotide repeat.

Because the $TA_{[x334]}$ dinucleotide repeat that forms the entirety of the SSC in *P. estuarinus* is a palindrome, its orientation relative to the LSC is impossible to determine. However, the *P. antipodarum* SSC does have identifiable directionality because it is made up of a series of two dinucleotide repeats ($TA_{[x54]}$ and $TC_{[x337]}$) that, when reversed, read $GA_{[x337]}$ and $TA_{[x54]}$, respectively. Notably, SPAdes correctly predicted these interspecific differences in SSC content as a barbell genome architecture for *P. antipodarum* and a lollipop genome architecture for *P. estuarinus*. We used a combination of Illumina MiSeq paired-end reads (IRs) and PacBio Circular Consensus Sequencing Reads (dinucleotide repeats) to insert the structural elements into each respective assembly, resulting in a closed, circular assembly in each case.

Both de novo assemblies exhibit consistent short-read mapping depth throughout the LSC (*P. antipodarum* mean coverage depth [±SD] = 879.83 [±95.17]; *P. estuarinus* mean coverage depth [±SD] = 5220.69 [±317.94]), and gene order is identical between the two species (fig. 2). Similarly, consistent coverage is obtained from mapping PacBio reads to the mitochondrial assemblies (supplementary fig. S2*a* and *b*, Supplementary Material online). When the nuclear genome assembly (BioProject ID: PRJNA717745) is excluded, coverage in the SSC spikes to >700,000× (supplementary fig. S2*c* and *d*, Supplementary Material online), likely as a result of simple-sequence repeats mapping here that are part of the nuclear genome. When the nuclear genome is included as a part of the reference during read mapping, the short-read coverage of the SSC declines dramatically (fig. 2). Despite the decline in coverage, we are confident that this IR-SSC-IR structure is indeed a part of the mitochondrial genome because (1) numerous PacBio subreads from each species (78 from *P. antipodarum* and 42 from *P. estuarinus*) span across the entire ~1.7 kb structure in a unique sequence in the LSC region on each side and (2) there are no reads in any of our HiSeq, MiSeq, or PacBio data sets that span between the two ends of the LSC region that also do not contain the IR-SSC-IR element. To confirm that the reads used in our assembly were mitochondrial in origin, we also mapped PacBio reads to the entire assembly, including both the nuclear and mitochondrial genomes. The vast majority of reads that have primary mapping locations in the mitochondrial genome (i.e., the reads used in our assembly) did not map elsewhere in the nuclear genome (*P. antipodarum*: 269/278 reads—96.8%; *P. estuarinus*: 637/644 reads—98.9%), and all of the reads with significant alignments elsewhere in the nuclear assembly have lower alignment scores in the nuclear genome than in the mitochondrial genome (supplementary fig. S3, Supplementary Material online).

We also confirmed the presence of the IRs by applying a custom-designed PCR approach using a single outward-facing primer paired with either a primer matching the top strand of the 5′ flanking sequence (as oriented in fig. 2) or the bottom strand of the 3′ flanking sequence

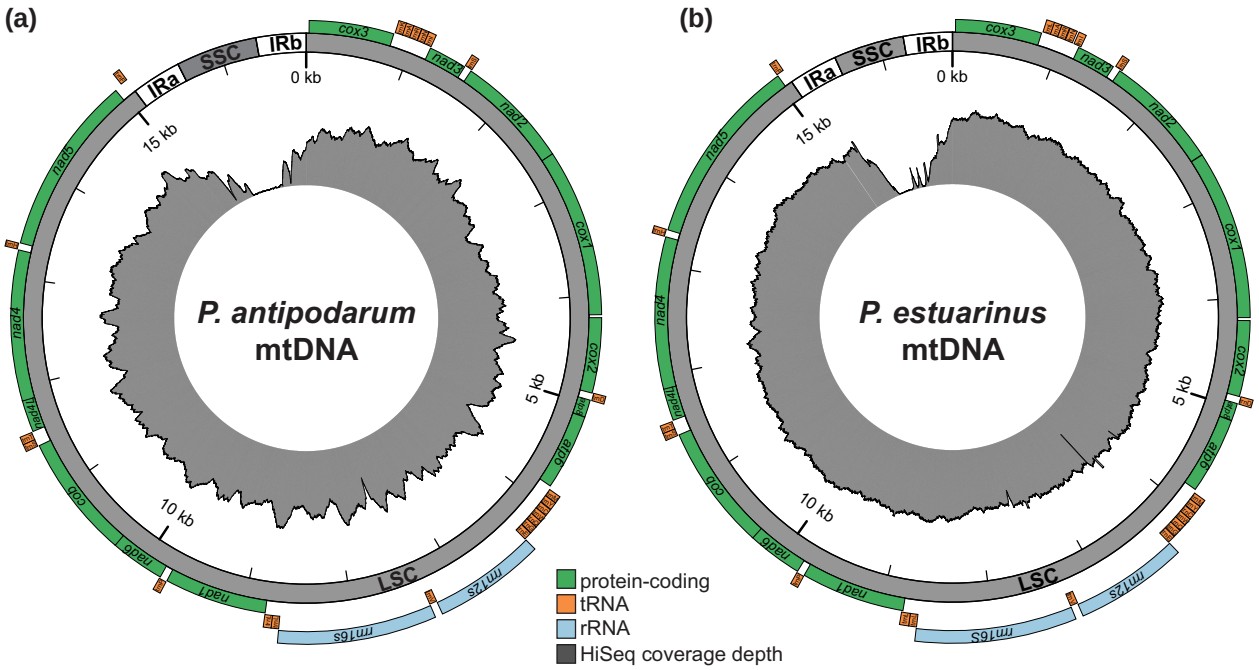

Fig. 2. *Potamopyrgus* mitochondrial genome architecture. De novo mitochondrial genome assemblies depicted for (*a*) *Potamopyrgus antipodarum* and (*b*) *Potamopyrgus estuarinus*. Protein-coding (green), tRNA (orange), and rRNA (light blue) genes are arranged along a mitochondrial genome architecture consisting of a large single-copy (LSC) region, two inverted repeats (IRa and IRb), and a SSC region comprised of dinucleotide repeats. Illumina HiSeq depth-of-coverage is depicted in the innermost circle in gray. The drops in coverage associated with the IRs likely reflect recombination breakpoints, whereas the low coverage in the SSC is likely due to low complexity reads mapping to non-mitochondrial contigs in the nuclear genome assembly.

(supplementary table S1, Supplementary Material online). That is, the primer designed to hybridize to the top strand of IRa (allowing polymerization toward *trnF* and *nad5*) can also anneal to the bottom strand of IRb (allowing polymerization toward *cox3*). The target amplicon spanning the *trnF*-IRa junction is designed to be 518 bp in length, whereas the amplicon targeting the IRb-*cox3* junction is designed to be 601 bp (supplementary fig. S4*a*, Supplementary Material online). Both products could be amplified from both species (supplementary fig. S4*b*, Supplementary Material online). We also attempted to amplify the SSC region using the reverse complement of the IR primer (i.e., anneals to the bottom strand of IRa and the top strand of IRb, potentially allowing for amplification of the SSC region). However, despite multiple attempts, we were unable to recover an amplification product. It is possible that the high AT content in the SSC does not permit efficient amplification (consistent with previous unsuccessful efforts to sequence the SSC region with Sanger reads). Still, our PCR efforts directly confirm the presence of an IR independent of any sequencing data or library preparation artifacts.

Sliding-window analyses of patterns of substitution in *P. antipodarum* versus *P. estuarinus* indicate that IRs exhibit high rates of sequence evolution relative to the rest of the genome (fig. 3). The SSC also appears to evolve rapidly, as evidenced by the presence of a second dinucleotide repeat in *P. antipodarum* versus only a single dinucleotide repeat in *P. estuarinus*. We also were able to determine that the IR is expressed from RNA sequencing reads.

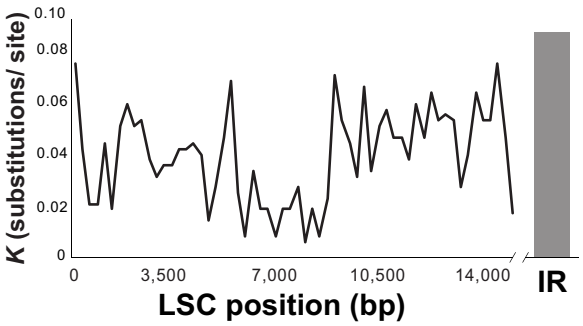

Fig. 3. Sliding-window analysis of the substitution rate (*K*) in the LSC (black line) compared with the single window of the inverted repeat (gray bar). Window size was 500 bp, with a 100-bp step size.

BLAST searches of the transcribed sequence did not have any significant hits, potentially indicating that the IR produces a novel noncoding RNA (e.g., Pozzi et al. 2017). A more detailed survey of mitochondrial genome architectural diversity in mollusks and a broader sample of bilaterians may benefit tremendously from the wide availability of long-read sequencing data.

Except for the IR-SSC-IR structure (supplementary fig. S5, Supplementary Material online), we did not find any sequence differences between our assembly and the previous Sanger sequence-generated assembly for the mitochondrial genome of *P. antipodarum* (GenBank Acc. No.: MG979468.1). Evidently, 23 bp of the IRs nearest the

LSC boundaries (46 bp in total) were used to circularize the old assembly. Other efforts to assemble this region in different *P. antipodarum* lineages (GQ996415-GQ996433, Neiman et al. 2010; MG979458-MG979470, Sharbrough et al. 2018) took similar approaches.

The IR-SSC-IR architecture has profound implications for mitochondrial genome biology, particularly, regarding mtDNA replication and repair. Indeed, IR-SSC-IR architectures are prone to forming hairpin or cruciform secondary structures, with the former implicated in DNA polymerase stalling (Voineagu et al. 2008). The predicted structure of IR-SSC-IR is somewhat reminiscent of the structure formed during vertebrate mtDNA replication at the light-chain replication origin (OriL), though the IR-SSC-IR is substantially longer (OriL ~30 bp vs. IR-SSC-IR ~1,700 bp). In vertebrates, mtDNA replication generally involves a strand-displacement mechanism in which leading-strand synthesis proceeds until encountering OriL, a ~30-bp region that forms a stem-loop structure when single-stranded. Lagging-strand synthesis is then primed starting in the loop of the single-stranded stem-loop structure (Fusté et al. 2010; Bannwarth et al. 2012). Whether this genomic structural feature is involved in mtDNA replication represents an open question in *P. antipodarum*. However, the symmetrical orientation of the IRs raises the question of whether replication could proceed in both directions (either *via* displacement synthesis (Brown et al. 2005) or *via* strand-coupled synthesis (Jõers and Jacobs 2013). Because mtDNA replication accuracy likely plays a central role in driving the notably high rate of mtDNA evolution in bilaterians, additional investigation into the diversity and distribution of mtDNA replication mechanisms will be critical for understanding mitochondrial genome biology and evolution.

Repeat-associated Recombination in Potamopyrgus Mitochondrial Genomes

Our observation that *Potamopyrgus* mitochondrial genome architecture resembles that found in the chloroplast genomes of land plants (Mower 2018) suggests that intramolecular recombination between the IRs may lead to "flip-flopping" of the SSC, as originally suggested for the chloroplast genomes by (Palmer 1983). Although the SSC of *P. estuarinus* is itself an IR (i.e., $TA_{[x334]}$), and therefore, impossible to use to determine directionality, the SSC of *P. antipodarum* (i.e., $TA_{[x54]}$-$TC_{[x337]}$) has definitive directionality. If flip-flopping is occurring, both orientations of the SSC should be detectable among PacBio reads spanning the region (fig. 4*a*). Consistent with this expectation, among 131 PacBio subreads that mapped unequivocally to both the LSC and the SSC in *P. antipodarum*, 77 subreads (58.8%) exhibit forward orientation of the SSC, and 54 subreads (41.2%) exhibit the reverse orientation (fig. 4*c*). We also found that among 20 polymerase reads with multiple subreads mapping to both the LSC and the SSC, consecutive subreads invariably exhibit opposing orientations of the SSC after being oriented relative to the LSC (supplementary fig. S6, Supplementary Material online). We interpret this invariable alternation of SSC orientation across consecutive subreads as evidence that the different strands of the mitochondrial genome have different SSC orientations, potentially indicating the presence of mitochondrial dimers. Several potential molecular mechanisms may explain this observation (see additional discussion below), all of which require some form of intra- or intermolecular recombination. Together, these data provide strong and direct evidence of flip-flopping of the SSC relative to the LSC regions by frequent intramolecular recombination between the IRs. These data also explain why assembling an animal mitochondrial genome with only Illumina HiSeq data was so challenging.

Our long-read sequencing data provide useful glimpses into recombination mechanisms that could generate these patterns of flip-flopping. In particular, reads containing forward-oriented SSCs (i.e., matching the assembly presented herein) are not significantly more common than reads containing the reverse-oriented SSC ($\chi^2 = 2.02$, $P = 0.155$). This pattern is consistent with IR-mediated recombination equilibrium dynamics. Alternatively, the strict alternation of SSC orientations observed in *P. antipodarum* would seem to conflict with flip-flopped orientations resulting from equilibrium dynamics of IR-mediated recombination. This observation leads us to speculate that mitochondrial dimers, joined at the IRs, might explain the strand-specific nature of flip-flopping. To assess whether we had captured mitochondrial dimers in our data set, we identified 854 subreads from *P. antipodarum* that mapped to the mitochondrial assembly but that were longer than the length of the assembly. Of these 854 subreads, nine (1.1%) subreads (originating from seven distinct polymerase reads) from *P. antipodarum* appear to capture an additional mtDNA molecule (supplementary fig. S7, Supplementary Material online). There was one subread that bridged across the SSC multiple times, and for which the orientations were in the same direction in that case (i.e., m54138_171231_122620/22610103/17646_34950 TA-TC). Together, these data point toward IR-mediated recombination, resulting in the constant flip-flopping of the SSC. This same phenomenon is commonly observed in plant chloroplast genomes that harbor an IR (Wang and Lanfear 2019).

Mitochondrial genomes, unlike the chloroplast genomes that reside inside organelles that rarely fuse (Schattat et al. 2012), often encounter one another inside the cell during the process of mitochondrial fusion (Chan 2006). Indeed, mitochondrial fusion and subsequent recombination among mtDNA molecules are thought to be a primary mechanism for maintaining mtDNA integrity in the face of oxidative damage (Osman et al. 2015). The widespread prevalence of mitochondrial fusion in eukaryotes combined with our observation of extensive flip-flop recombination in *P. antipodarum* mitochondrial genomes led us to speculate that *Potamopyrgus* mitochondrial genomes might also experience inter-molecular mitochondrial recombination associated with the IRs. We did not find any polymorphisms within the IRs with which we might

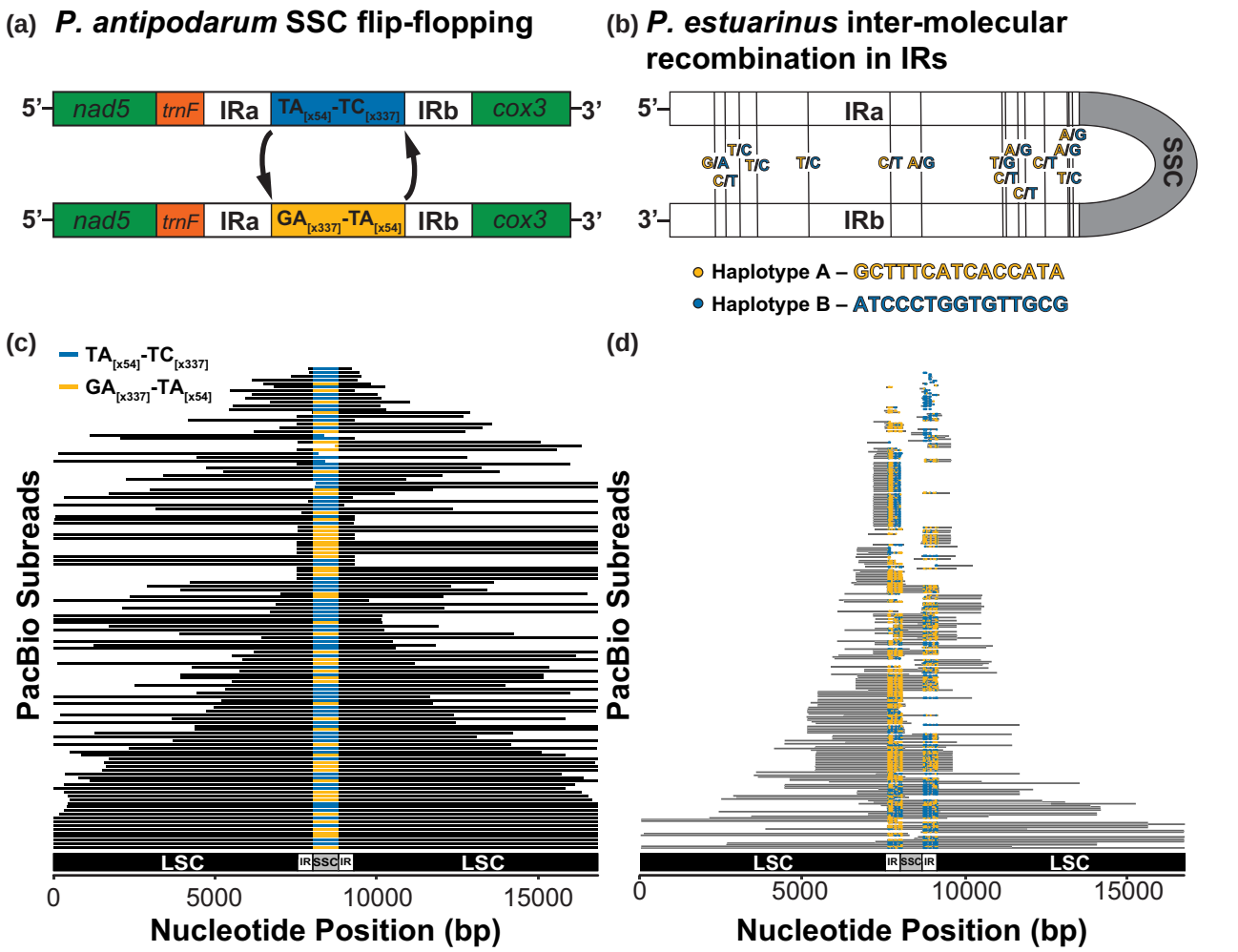

Fig. 4. Repeat-mediated recombination in mitochondrial genomes of *Potamopyrgus*. (*a*) Flip-flop recombination results in forward (top) and reverse (bottom) orientations of the SSC relative to the LSC in *Potamopyrgus antipodarum*. (*b*) Two IR haplotypes exist in *Potamopyrgus estuarinus* mitochondrial reads, suggestive of inter-molecular recombination. (*c*) Forward (blue) and reverse (orange) orientations of the SSC relative to the LSC exist at high frequencies among PacBio subreads from *P. antipodarum*. (*d*) Sites of inter-molecular recombination in PacBio subreads from *P. estuarinus*.

Table 1. Inverted Repeat Haplotype Frequencies in *Potamopyrgus estuarinus* PacBio Pool-seq.

IRa position	IRb position	Haplotype A allele	Haplotype B allele	A-A[a]	A-B[b]	B-A[c]	B-B[d]
7603	9098	G	A	27 (47%)	7 (12%)	0 (0%)	23 (40%)
7615	9086	C	T	23 (44%)	13 (25%)	0 (0%)	16 (31%)
7631	9070	T	C	29 (64%)	5 (11%)	0 (0%)	11 (24%)
7651	9050	T	C	34 (61%)	8 (14%)	1 (2%)	13 (23%)
7709	8992	T	C	37 (66%)	7 (13%)	0 (0%)	12 (21%)
7802	8899	C	T	21 (35%)	18 (30%)	1 (2%)	20 (33%)
7837	8864	A	G	26 (43%)	12 (20%)	0 (0%)	22 (37%)
7929	8772	T	G	35 (48%)	13 (18%)	2 (3%)	23 (32%)
7933	8768	C	T	38 (54%)	10 (14%)	0 (0%)	23 (32%)
7947	8754	A	G	31 (47%)	8 (12%)	0 (0%)	27 (41%)
7955	8746	C	T	22 (32%)	7 (10%)	0 (0%)	39 (57%)
7977	8724	C	T	32 (44%)	20 (27%)	0 (0%)	21 (29%)
8003	8698	A	G	29 (48%)	16 (27%)	0 (0%)	15 (25%)
8005	8696	T	C	29 (48%)	17 (28%)	0 (0%)	15 (25%)
8009	8692	A	G	30 (48%)	18 (27%)	0 (0%)	18 (27%)

[a]Number (percentage) of reads homozygous for the Haplotype A allele across the IR.
[b]Number (percentage) of reads homozygous for the Haplotype A allele in IRa and the Haplotype B allele in IRb.
[c]Number (percentage) of reads homozygous for the Haplotype B allele in IRa and the Haplotype A allele in IRb.
[d]Number (percentage) of reads homozygous for the Haplotype B allele across the IR.

map inter-molecular recombination, likely because the lineage we used for our reference genome sequencing has been inbred for more than 20 generations. Fortunately, *P. estuarinus* has no such inbreeding history, and all *P. estuarinus* individuals sequenced during this project were collected from the wild in New Zealand.

To test for inter-molecular recombination in *P. estuarinus* mitochondrial genomes, we first identify 15 high-frequency SNPs present in both Illumina MiSeq and PacBio read data sets generated from pools of multiple wild-caught *P. estuarinus* females (fig. 4*b*). We originally suspected these SNPs to represent differences among IRs rather than any form of heteroplasmy, particularly, in light of our hypothesis that this region represents the hypervariable D-loop. However, most PacBio reads that could span across both IRs harbor the same allele at both nucleotide positions (table 1), indicating that the SNPs most likely come from different mitochondrial haplotypes rather than representing the accumulation of divergence between IRs. We use these same PacBio reads to phase the SNPs into two distinct haplotypes, A and B (fig. 4*b*). Although Haplotype A appears to be more common among this read pool than Haplotype B (table 1), there are many individual reads that harbor SNPs from both A and B haplotypes (fig. 4*d*). In sum, although there is little variation between the IR copies within a single mtDNA molecule, there are many combinations of alleles among the distinct mtDNA molecules. This pattern implies that both intra- and inter-molecular recombination contribute to the distribution of genetic diversity in *Potamopyrgus* mtDNA.

To gain a broader perspective of the recombinational activity in *Potamopyrgus* mtDNA beyond the IRs, we also identified SNPs present in the LSC from the pooled *P. estuarinus* MiSeq data. In total, we identified 264 sites that are variable within the pooled sequencing (supplementary table S2, Supplementary Material online, fig. 5). After mapping MiSeq reads to the LSC, we combined paired-end reads into super reads using mapping information and quantified recombination between all pairs of SNP sites in the LSC. For each pair of SNP positions, we determined the haplotype of each read covering both sites and quantified the number of reads supporting each of the four combinations of alleles (i.e., REF-REF, REF-ALT, ALT-REF, and ALT-ALT). There were 4,697 SNP pairs with at least a single read covering both positions and 2,701 SNP pairs with >90× coverage (~5-fold higher than expected nuclear coverage assuming mean insert size ≥600 bp). In this case, the presence of more than two haplotypic classes for any given SNP pair was used as evidence of the activity of inter-molecular recombination. Among the 4,697 SNP pairs, we found 367 pairs with >90 reads supporting the third-most-common haplotypic class. We even found 32 SNP pairs with >180 reads covering the third-most-common haplotypic class. There were fewer pairs in which the fourth-most-common haplotypic class exceeded the 5× threshold expected of the nuclear genome, as only two pairs were found in which the fourth-most-common haplotypic class met or exceeded the

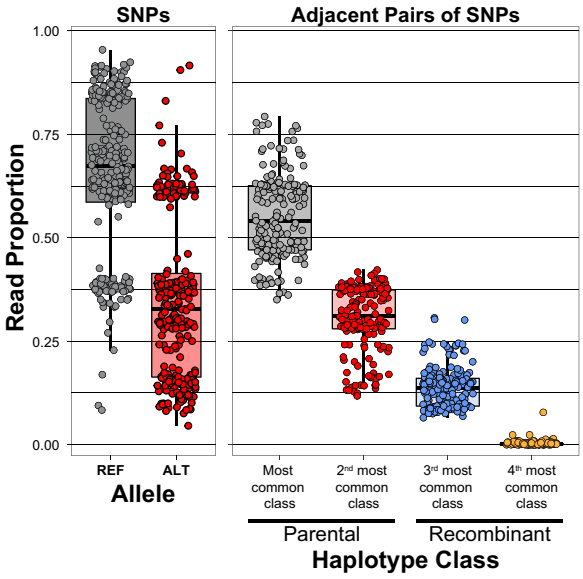

Fig. 5. Polymorphism and recombination in *Potamopyrgus estuarinus* mitochondrial genomes. Left: Allele frequencies for individual SNPs. Minimum read depth at SNP positions is 349*x*, with at least 47× coverage over minor alleles at all positions. Right: Haplotype class frequencies for pairs of SNPs measured by Illumina MiSeq read depth. The presence of more than two haplotype classes for any given pair of SNPs implies recombination among the molecules sequenced.

90× coverage threshold. When only considering the 263 independent pairs of SNPs (i.e., adjacent SNPs), we found that 139 pairs (52.9%) exhibit >90× coverage of the third-most-common haplotypic class (fig. 5, supplementary table S3, Supplementary Material online). Conversely, only 125 (46.4%) of the adjacent SNP pairs exhibited <90× sequencing coverage over the third-most-common haplotype class, indicating that recombination between adjacent SNPs was more likely than not. One of the adjacent SNP pairs (10,573 [C/T]—10,630 [T/C]) exhibited relatively high coverage of all four haplotype classes, with 365 C-T reads, 78 T-T reads, 243 C-C reads, and 312 T-C reads. In sum, the high frequency of SNP pairs with >2 haplotypic classes at coverage levels far exceeding those expected by nuclear-encoded mitochondrial sequences (numts) or that can be explained by PCR or sequencing errors supports the presence of inter-molecular recombination in *P. estuarinus* individuals.

The presence of inter-molecular recombination documented above implies extensive heteroplasmy within individual cells, as mtDNA molecules must co-exist inside the same cell (indeed, the same mitochondrion) in order to recombine. Although these data do not provide direct evidence regarding the mode of mitochondrial inheritance, we speculate that the apparent large degree of heteroplasmy required to explain these data is likely to have originated from multiple parents (Camus et al. 2022). Perhaps even more surprising is the extensive mitochondrial diversity maintained in *P. estuarinus* populations in the face of recombination. These snails were collected

within meters from one another, yet hundreds of mitochondrial SNPs can be recovered from only eight individuals. The mechanism(s) whereby such high levels of mitochondrial diversity are maintained in this system represent an important area of future investigation.

Mitochondrial recombination has been detected in a wide diversity of plant, fungi, and even animal taxa (Mita et al. 1990; Kajander et al. 2000; Ladoukakis and Zouros 2001; Piganeau et al. 2004; Ciborowski et al. 2007; Leducq et al. 2017; Dahal et al. 2018). Indeed, plant geneticists have long recognized the importance of recombination for mitochondrial genome architecture and evolution (Lonsdale et al. 1988; Sloan 2013). A recent study even reported a history of extensive genetic recombination in the massive mitochondrial genomes of populations of the angiosperm *Silene* (Wu and Sloan 2019). More commonly, plant mitochondrial genomes undergo extensive repeat-mediated recombination. So called "substoichiometric shifting" (Woloszynska 2009; Maréchal and Brisson 2010; Arrieta-Montiel and Mackenzie 2011; Davila et al. 2011) results in multiple distinct mitochondrial isoforms existing within the same cell (André et al. 1992; Gualberto and Newton 2017). Fungal mitochondrial genomes also appear to experience relatively frequent homologous recombination (Wu et al. 2015; Brankovics et al. 2017; Leducq et al. 2017; Wolters et al. 2018).

In contrast, evidence for mitochondrial recombination in animals is relatively scarce (Chen 2013) and has primarily been observed in comparative and population genetic data sets (Goddard et al. 2006; Guo et al. 2006; Fukami et al. 2007; Ujvari et al. 2007; Shao and Barker 2011), with a couple of notable exceptions. One prominent example is provided by doubly uniparentally inherited (DUI) mitochondrial genomes in some mollusk taxa, which can facilitate recombination (Ladoukakis and Zouros 2001; Stewart et al. 2009). Mitochondrial genomes in human and mouse heart tissues have been found to form recombination-associated structures during mitochondrial genome replication (Pohjoismäki et al. 2009; Herbers et al. 2019), and intra- and inter-molecular recombination between distinct mitochondrial molecules in heteroplasmic individuals can be induced through oxidative stress in *Drosophila* (Ma and O'Farrell 2015) and using restriction enzymes in mice (Bacman et al. 2009). Recent reports of heteroplasmy in humans (Rebolledo-Jaramillo et al. 2014; Luo et al. 2018) underscore the potential importance of mitochondrial recombination to human health outcomes.

Repeat-mediated recombination in plant mitochondrial genomes is strongly associated with DNA repair via *MSH1* and is often cited as a primary cause of their relatively low rate of molecular evolution (Drouin et al. 2008; Davila et al. 2011; Wu et al. 2020). In contrast, the high levels of nucleotide diversity and high rates of sequence evolution in the *Potamopyrgus* IRs suggest that recombination is not acting to reduce the rate of sequence evolution. The recombination we observed in the LSC and the IR has two primary consequences. First, recombination between distinct mtDNA molecules can reset Muller's ratchet (Muller 1964; Neiman and Taylor 2009), which is otherwise expected to lead to the eventual collapse of asexual lineages (Gabriel et al. 1993). Second, recombination between distinct mtDNA molecules breaks down linkage disequilibrium between mutations in the mitochondrial genome, countering the Hill–Robertson effect (Hill and Robertson 1966) by allowing these mutations to be selected (positively or negatively) independently from their mitochondrial genomic background. Under the presumption that mitochondrial genomes lack recombination, both Muller's ratchet and the Hill–Robertson effect have long been assumed to generate harmful mutation accumulation in the mitochondrial genomes (Neiman and Taylor 2009). However, recent work in nematodes (Konrad et al. 2017) and other animals (Allio et al. 2017) indicates that the effective population size of mtDNA may be much larger than previously thought. Our discovery of inter-molecular recombination in these mollusks and the continued existence of separately inherited nuclear and mitochondrial genomes after more than a billion years of intra-cellular co-evolution are consistent with the hypothesis that mitochondrial recombination might facilitate escape from a seemingly inevitable mutational meltdown.

Conclusion

We document a novel and complex genomic structural feature of the mitochondrial genomes of *Potamopyrgus* snails. These native New Zealand prosobranch snails are well-studied because of the status of *P. antipodarum* as a model system for sexual reproduction, host–parasite co-evolution, and as a worldwide invader. Interest in how the consequences of reproductive mode polymorphism might influence mitochondrial evolution and the maintenance of sex has motivated surveys revealing extensive genetic variation in mtDNA (Neiman and Lively 2004; Neiman et al. 2010; Paczesniak et al. 2013; Sharbrough et al. 2018) as well as a phenotypic variation for mitochondrial function (Sharbrough et al. 2017; Greimann et al. 2020) in *P. antipodarum*. Although previous efforts to sequence *P. antipodarum* mitochondrial genomes have provided valuable data on this front (Neiman et al. 2010; Sharbrough et al. 2018), all have been restricted to short sequencing reads or PCR-amplified segments, and none have been able to assemble the entire circular structure.

Using a combination of short (Illumina HiSeq, Illumina MiSeq) and long (PacBio) reads obtained from the initial stages of sequencing the whole nuclear genome, we successfully reassembled the entire *P. antipodarum* and *P. estuarinus* mitochondrial genomes de novo. In addition to confirming the presence of the previously reported sequence (Neiman et al. 2010; Sharbrough et al. 2018), these assemblies revealed a pair of IRs interrupted by a repetitive single-copy region that were not recognized in the earlier studies. The IRs appear to evolve very rapidly (in contrast to the slow rates of molecular evolution seen in chloroplast IRs (Perry and Wolfe 2002), and RNAseq data revealed the presence of a novel transcript, ~450 bp in

length, with no open reading frames or identified homology. The nature of this genomic feature and its clear resemblance to the genomic architecture of plant chloroplast genomes (Smith and Keeling 2015) suggests that this component of *Potamopyrgus* mitochondrial genomes may be a site for intra- and inter-molecular repeat-mediated recombination. PacBio reads confirmed this hypothesis and indicated that intra-molecular recombination that resulted in "flip-flopping" (Palmer 1983) of the SSC region relative to the large single-copy (LSC) region is very frequent and that inter-molecular recombination between distinct mtDNA molecules is common when present within the same cell. Together, these findings necessitate a careful re-examination of bilaterian mitochondrial genomes, with special emphasis on repeat-mediated mitochondrial recombination. More broadly, the potential for the widespread presence of recombination in bilaterian animal mitochondrial genomes that this result suggests will enable important new insights into mitochondrial biology and evolution and raise important questions: Are hypervariable regions like D-loops recombination hotspots in mitochondrial genomes? Does mitochondrial recombination stave off mutational meltdown? Is the IR-SSC-IR genome architecture a cause or consequence of mitochondrial recombination?

Materials and Methods

DNA Sequence Data

DNA from *P. antipodarum* and *P. estuarinus* were sequenced as part of the *Potamopyrgus* nuclear genomes project using Illumina HiSeq, Illumina MiSeq, and PacBio SMRT sequencing (PRJNA717745). DNA extraction details and library preparation details are fully described elsewhere. Briefly, we extracted DNA from Alex Yellow (a sexual *P. antipodarum* lineage inbred for 20+ years) and field-collected *P. estuarinus* using a guanidinium-thiocyanate-based phenol-chloroform DNA extraction protocol (Sharbrough et al. 2018; McElroy et al. 2021), cleaned DNA extractions with Zymo's Clean and Concentrate Kit (Zymo Research, Irvine, CA, USA), and resuspended DNA in 30–100 µL of T-low-E buffer (10 mM Tris pH 8.0, 0.1 mM EDTA). We sequenced 2 × 100 nt paired-end reads on an Illumina Hi-Seq 2500 and 2 × 250 nt paired-end reads on an Illumina MiSeq. To obtain high-molecular-weight DNA for PacBio sequencing, we extracted DNA from pools of Alex Yellow ($N = 20$) and *P. estuarinus* ($N = 20$) individuals with a user-developed protocol for Qiagen genomic tip DNA extraction for insects (Genomic tip 100/G) that uses a gravity column for high-molecular-weight extractions. The Pacbio sequencing library construction followed the PacBio 20-kb protocol in both cases. These libraries were sequenced in nine SMRT cells of a Sequel for Alex Yellow and 10 SMRT cells for *P. estuarinus*. All PacBio sequencing was performed at the Arizona Genomics Institute (Tuscon, AZ, USA), which provided all subread BAM files and associated SMRT movie metadata. PacBio Circular Consensus Sequencing (CCS) reads were obtained using ccs v3.4.0 with default settings (https://ccs.how/), requiring a minimum of three subread passes for CCS generation.

Mitochondrial Genome Assembly

We assembled the paired-end HiSeq reads de novo using SPAdes v. 3.13.0 (Bankevich et al. 2012) separately for Alex Yellow and *P. estuarinus* with default settings. We then extracted mitochondrial contigs from the assembled contigs of each species using blastn v. 2.7.1 + against the published mitochondrial genome from Alex Yellow (MG979468.1), setting the *e*-value threshold to 10^{-5} and max_target_seqs to 10,000. Significant blast hits were manually stitched together according to the published Alex Yellow genome assembly (Neiman et al. 2010). We used Pilon v. 1.23 (Walker et al. 2014) to polish the mitochondrial assemblies with lineage-specific paired-end reads.

At this stage, both the Alex Yellow and *P. estuarinus* mitochondrial assemblies displayed a single discontinuity (i.e., position preventing circularization), leading us to hypothesize that a repeat element was causing the assembly to break at position 12,481 relative to the published Alex Yellow mitochondrial reference genome. We did not detect any other differences between our newly generated assembly and the previously published assembly (MG979468.1). Further investigation of the discontinuity revealed that the published Alex Yellow mitochondrial reference genome contained a 44-bp palindromic repeat at this site, potentially explaining the anomaly. However, paired-end Illumina reads could not span the structure, despite being of adequate length to do so. To further investigate this discrepancy, we used Bandage v. 0.8.1 (Wick et al. 2015) to visualize the assembly graph (supplementary fig. S1, Supplementary Material online), which revealed a clear signature of IRs. As a result, we inferred that single-molecule PacBio data would be able to resolve the apparent discontinuity in these short-read assemblies.

First, we reoriented our short-read assemblies to place the discontinuity at the center of the Alex Yellow and *P. estuarinus* assemblies. We then mapped PacBio subreads from each species to these re-oriented mitochondrial assemblies using blasr v. 5.3.2-06c9543 (Chaisson and Tesler 2012). To ensure that we removed any reads that might be of nuclear origin, we also mapped to the nuclear genome in concert with the re-oriented mitochondrial assemblies with minimap2 v 2.15-r905 and excluded any reads with primary mapping locations in the nuclear contigs (supplementary fig. S3, Supplementary Material online). We collected reads that spanned the discontinuity by identifying reads with primary mapping locations in the mtDNA (i.e., do not map to the rest of the nuclear assembly) and with significant alignments on both sides of the discontinuity (i.e., 5′ and 3′ relative to the discontinuity). We mapped only those spanning reads to themselves using minimap2 v 2.15-r905 (Li 2018) and assembled these

mapped reads using miniasm v 0.3-r179 (Li 2016). We aligned the resulting long-read assembly to the mitochondrial assemblies of both species, which both contained a ~1.7-kb previously undetected region that coincided exactly with the discontinuity. A dot plot of the structures revealed that their 5′ and 3′ flanks were IRs, with an interior region comprised of dinucleotide repeats. In Alex Yellow, there were two distinct dinucleotide repeats (AT and CT), whereas only the AT repeat was present in *P. estuarinus*. Therefore, we inserted this IR-dinucleotide-IR structure into each of our respective short-read assemblies.

To confidently assemble the IRs, we mapped MiSeq reads to the mitochondrial genome assemblies and collected all reads that mapped within 100 bp of the 5′ end of the structure and within 100 bp of the 3′ end of the structure. We discarded reads whose alignments did not extend *beyond* the boundaries of the structure to ensure that we only included reads of clear mitochondrial origin. We assembled these MiSeq reads for the 5′ IR and the 3′ IR separately using SPAdes. In Alex Yellow, both 5′ and 3′ assemblies yielded a single identical contig that captured the entire length of the IR (i.e., spanned from the 5′ end to the dinucleotide repeat and from the 3′ end to the dinucleotide repeat). However, in *P. estuarinus*, 5′ and 3′ IR assemblies contained two distinct contigs. From these results, we infer that the IRs contain 15 single-nucleotide differences (see fig. 4*b*), which are later verified with PacBio CCS reads from *P. estuarinus* obtained from distinct sample collections and DNA extractions. We also tested whether the IRs were expressed using previously obtained RNAseq from both species by mapping RNAseq reads to the mitochondrial genome.

Finally, to assemble the dinucleotide repeats, we aligned PacBio CCS reads to the HiSeq-PacBio-MiSeq assemblies for each species using blasr and obtained consensus lengths for the two dinucleotide repeats in Alex Yellow ($TA_{[x54]}$-$TC_{[x337]}$) and the single dinucleotide repeat in *P. estuarinus* ($TA_{[x334]}$). We validated our assemblies using (1) long single-molecule PacBio reads that could span the insertion and the artificially introduced assembly break at the 5′ and 3′ ends of the linearized assemblies, and (2) comparing coverage across the entire assembly using short reads.

PCR Confirmation of Inverted Repeats

We designed a custom PCR approach to directly test the IR hypothesis using four experiments (supplementary table S1, Supplementary Material online). All PCR experiments were performed in 25 µL volumes, using 5 ng DNA, 2.5 µL 10X iTaq Buffer (Bio-Rad, Hercules, CA, USA), 1.5 mM $MgCl_2$, 200 µM for each dNTP, 0.1 µM of each primer, and 1 U of iTaq DNA polymerase (Bio-Rad). The first experiment was a positive control to amplify a 240 bp region in the gene-dense LSC. The second experiment is designed to amplify a 518-bp fragment of the left flank of the IR using a forward primer complementary to the bottom strand of *trnF* and a reverse primer that is complementary to the top strand of Ira (as oriented in fig. 2). For the third experiment, we used this same IR primer as a forward primer in which it had to be complementary to the bottom strand of IRb in order to amplify a 601-bp product in concert with a reverse primer designed to complement the top strand of *cox3*. Thus, the IR primer can be used to amplify two completely different regions of the mitochondrial genome, at the predicted sizes (supplementary fig. S4*b*, Supplementary Material online). Finally, in an attempt to amplify the SSC region, we designed a final primer, which was the reverse complement of the IR primer. In this experiment, we used only a single primer that was predicted to be complementary to both the bottom strand of IRa and the top strand of IRb and was expected to produce a ~1,100-bp product. We used the same conditions for each PCR experiment: 95 °C for a 3-min denaturation, then cycled the following three steps 40×: (1) 30 s at 95 °C, (2) 30 s at 56 °C for primer annealing, and (3) 1.5 min at 72 °C for extension. We had a final extension time of 6 min and then evaluated the success of the PCR reactions on 1% sodium borate-agarose gel. Because the fourth experiment did not yield any positive results, we also tried a 6-min extension time for each cycle, using a LongAmp polymerase (New England Biolabs, Ipswich, MA, USA), which also did not yield any bands. Gels containing 3.5 µl of SYBR Safe stain (Invitrogen, Waltham, MA, USA) were run at 300 V for 15 min and visualized under blue or UV light.

Identification of Recombinant Mitochondrial Molecules

The resemblance of certain elements of the *Potamopyrgus* mitochondrial genome structure to that of the chloroplast genomes in most land plants (Smith and Keeling 2015) led us to hypothesize that the IRs could mediate intramolecular recombination. Because the SSC in Alex Yellow comprises two distinct dinucleotide repeats (i.e., TA and TC), reads that span the length of the IR-SSC-IR structure can readily be evaluated for recombination. If recombination exists in these mitochondrial genomes, both TA-TC and GA-TA-SSC conformations should be observable among reads that span the structure. In *P. estuarinus*, the orientation of the SSC is not useful for diagnosing recombination because the SSC is identical in either direction. Therefore, we had to rely on 15 single-nucleotide differences between IRa and IRb (inferred from MiSeq reads and confirmed in PacBio CCS reads) to quantify the rates and patterns of mitochondrial recombination in *P. estuarinus*.

We mapped PacBio subreads and CCS reads from each species to our final assemblies and quantified IR/SSC orientations in reads that (1) spanned the entire structure and (2) in all reads that overlapped with the IR-SSC-IR structure, but that also mapped unambiguously to one side of the LSC or the other. We visualized the orientations of these reads using CIRCOS v. 0.69-5 (Krzywinski et al. 2009) and the overall read frequencies in R v4.1.2. We

tested whether the forward orientation of the SSC was more common than the reverse orientation using a χ^2 goodness-of-fit test.

Recombination among IRs in chloroplast genomes is known to reduce the rate of sequence evolution relative to rates in the single-copy regions (Perry and Wolfe 2002). Therefore, we aligned IRs and LSCs separately and compared sequence divergence between *P. antipodarum* and *P. estuarinus* across the two regions, removing any sites containing polymorphism within *P. antipodarum* from the analysis. We excluded the SSC from this analysis because its repetitive nature makes for uncertain dubious base calls and alignments.

Recombination between SNP pairs in the LSC was investigated using MiSeq reads. We mapped MiSeq reads to the *P. estuarinus* assembly and used mapping information to convert paired-end reads into super reads using a custom python script (superReads.py). We realigned super reads to the mitochondrial assembly and converted the resulting SAM file to a multiple sequence alignment using a custom python script (sam2Fasta.py). Haplotype calls from individual reads were obtained from the multiple sequence alignment and quantified using the recombinantReads.py python script. All python scripts used in this study are available at (https://github.com/jsharbrough/potamoMitoGenomes).

Supplementary material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

The authors thank Gus Waneka, Dan Sloan, and Sam Ward for comments on the manuscript, discussions about mechanisms of recombination, and statistical design. The authors thank Cindy Toll, Katelyn Larkin, and Peter Wilton for their contributions to the *Potamopyrgus* genome project. They acknowledge funding from the National Science Foundation (NSF-MCB 1122176, NSF-DEB 1753851), NM-INBRE, and the New Mexico Institute of Mining and Technology. Some of this work utilized resources from the University of Colorado Boulder Research Computing Group (NSF awards ACI-1532235 and ACI-1532236).

Data Availability

All sequence data (genomic and transcriptomic) supporting the mitochondrial assemblies are available in GenBank under BioProject PRJNA717745 (SRA Accessions: SRR22947758-SRR22947762). Mitochondrial assemblies are available in GenBank (Accession Nos. OQ161205, OQ161206). Mitochondrial genome assemblies and annotations, sequencing reads, and transcriptome assemblies are also available on FigShare (*P. antipodarum*: https://doi.org/10.6084/m9.figshare.20471784; *P. estuarinus*: https://doi.org/10.6084/m9.figshare.20474262).

References

Allio R, Donega S, Galtier N, Nabholz B. 2017. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol Biol Evol.* **34**:2762–2772.

André C, Levy A, Walbot V. 1992. Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet.* **8**: 128–132.

Arnason U, Adegoke JA, Gullberg A, Harley EH, Janke A, Kullberg M. 2008. Mitogenomic relationships of placental mammals and molecular estimates of their divergences. *Gene* **421**:37–51.

Arrieta-Montiel MP, Mackenzie SA. 2011. Plant mitochondrial genomes and recombination. In: Kempken F, editor. *Plant mitochondria*. New York (NY): Springer. p. 65–82.

Bacman SR, Williams SL, Moraes CT. 2009. Intra- and inter-molecular recombination of mitochondrial DNA after *in vivo* induction of multiple double-strand breaks. *Nucleic Acids Res.* **37**:4218–4226.

Bankers L, Fields P, McElroy KE, Boore JL, Logsdon JM Jr, Neiman M. 2017. Genomic evidence for population-specific responses to coevolving parasites in a New Zealand freshwater snail. *Mol Ecol.* **26**: 3663–3675.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* **19**:455–477.

Bannwarth S, Figueroa A, Fragaki K, Destroismaisons L, Lacas-Gervais S, Lespinasse F, Vandenbos F, Pradelli LA, Ricci J-E, Rötig A, *et al.* 2012. The human *MSH5* (MutS homolog 5) protein localizes to mitochondria and protects the mitochondrial genome from oxidative damage. *Mitochondrion* **12**:654–665.

Boore JL. 1999. Animal mitochondrial genomes. *Nucl Acids Res.* **27**: 1767–1780.

Brankovics B, van Dam P, Rep M, de Hoog GS, van der Lee TAJ, Waalwijk C, van Diepeningen AD. 2017. Mitochondrial genomes reveal recombination in the presumed asexual *Fusarium oxysporum* species complex. *BMC Genomics*. **18**:735.

Brown TA, Cecconi C, Tkachuk AN, Bustamante C, Clayton DA. 2005. Replication of mitochondrial DNA occurs by strand displacement with alternative light-strand origins, not via a strand-coupled mechanism. *Genes Dev.* **19**:2466–2476.

Camus MF, Alexander-Lawrie B, Sharbrough J, Hurst GDD. 2022. Inheritance through the cytoplasm. *Heredity (Edinb).* **129**: 31–43.

Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics*. **13**:238.

Chan DC. 2006. Mitochondrial fusion and fission in mammals. *Annu Rev Cell Dev Biol.* **22**:79–99.

Chen XJ. 2013. Mechanism of homologous recombination and implications for aging-related deletions in mitochondrial DNA. *Microbiol Mol Biol Rev.* **77**:476–496.

Ciborowski KL, Consuegra S, García de Leániz C, Beaumont MA, Wang J, Jordan WC. 2007. Rare and fleeting: an example of interspecific recombination in animal mitochondrial DNA. *Biol Lett.* **3**: 554–557.

Cook CE, Yue Q, Akam M. 2005. Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. *Proc Biol Sci.* **272**:1295–1304.

Dahal S, Dubey S, Raghavan SC. 2018. Homologous recombination-mediated repair of DNA double-strand breaks operates in mammalian mitochondria. *Cell Mol Life Sci.* **75**:1641–1655.

Davila JI, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, Shedge V, Xu Y-Z, Weigel D, Mackenzie SA. 2011. Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*. *BMC Biol.* **9**:64.

Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289**:2342–2344.

Drouin G, Daoud H, Xia J. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol Phylogenet Evol.* **49**:827–831.

Formenti G, Rhie A, Balacco J, Haase B, Mountcastle J, Fedrigo O, Brown S, Capodiferro MR, Al-Ajli FO, Ambrosini R, *et al.* 2021. Complete vertebrate mitogenomes reveal widespread repeats and gene duplications. *Genome Biol.* **22**:120.

Fukami H, Chen CA, Chiou C-Y, Knowlton N. 2007. Novel group-I introns encoding a putative homing endonuclease in the mitochondrial *cox1* gene of *Scleractinian* corals. *J Mol Evol.* **64**: 591–600.

Fusté JM, Wanrooij S, Jemt E, Granycome CE, Cluett TJ, Shi Y, Atanassova N, Holt IJ, Gustafsson CM, Falkenberg M. 2010. Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Mol Cell.* **37**:67–78.

Gabriel W, Lynch M, Bürger R. 1993. Muller's ratchet and mutational meltdowns. *Evolution* **47**:1744–1757.

Ghiselli F, Gomes-Dos-Santos A, Adema CM, Lopes-Lima M, Sharbrough J, Boore JL. 2021. Molluscan mitochondrial genomes break the rules. *Philos Trans R Soc Lond B Biol Sci.* **376**:20200159.

Gibson AK, Xu JY, Lively CM. 2016. Within-population covariation between sexual reproduction and susceptibility to local parasites. *Evolution* **70**:2049–2060.

Goddard MR, Leigh J, Roger AJ, Pemberton AJ. 2006. Invasion and persistence of a selfish gene in the *Cnidaria*. *PLoS One.* **1**:e3.

Greimann ES, Ward SF, Woodell JD, Hennessey S, Kline MR, Moreno JA, Peters M, Cruise JL, Montooth KL, Neiman M, *et al.* 2020. Phenotypic variation in mitochondria-related performance traits across New Zealand snail populations. *Integr Comp Biol.* **60**: 275–287.

Gualberto JM, Newton KJ. 2017. Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Ann Rev Plant Biol.* **68**: 225–252.

Guo X, Liu S, Liu Y. 2006. Evidence for recombination of mitochondrial DNA in triploid crucian carp. *Genetics* **172**:1745–1749.

Haase M. 2008. The radiation of hydrobiid gastropods in New Zealand: a revision including the description of new species based on morphology and mtDNA sequence information. *System Biodivers.* **6**:99–159.

Herbers E, Kekäläinen NJ, Hangas A, Pohjoismäki JL, Goffart S. 2019. Tissue specific differences in mitochondrial DNA maintenance and expression. *Mitochondrion* **44**:85–92.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet Res.* **8**:269–294.

Jõers P, Jacobs HT. 2013. Analysis of replication intermediates indicates that *Drosophila melanogaster* mitochondrial DNA replicates by a strand-coupled theta mechanism. *PLoS One.* **8**:e53249.

Johri P, Marinov GK, Doak TG, Lynch M. 2019. Population genetics of *Paramecium* mitochondrial genomes: recombination, mutation spectrum, and efficacy of selection. *Genome Biol Evol.* **11**: 1398–1416.

Jokela J, Lively CM. 1995. Parasites, sex, and early reproduction in a mixed population of freshwater snails. *Evolution* **49**:1268–1271.

Kajander OA, Rovio AT, Majamaa K, Poulton J, Spelbrink JN, Holt IJ, Karhunen PJ, Jacobs HT. 2000. Human mtDNA sublimons resemble rearranged mitochondrial genomes found in pathological states. *Hum Mol Genet.* **9**:2821–2835.

Konrad A, Thompson O, Waterston RH, Moerman DG, Keightley PD, Bergthorsson U, Katju V. 2017. Mitochondrial mutation rate, spectrum and heteroplasmy in *Caenorhabditis elegans* spontaneous mutation accumulation lines of differing population size. *Mol Biol Evol.* **34**:1319–1334.

Koren S, Phillippy AM. 2015. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr Opin Microbiol.* **23**:110–120.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**:1639–1645.

Ladoukakis ED, Zouros E. 2001. Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Mol Biol Evol.* **18**:1168–1175.

Lavrov DV, Pett W. 2016. Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. *Genome Biol Evol.* **8**:2896–2913.

Leducq J-B, Henault M, Charron G, Nielly-Thibault L, Terrat Y, Fiumera HL, Shapiro BJ, Landry CR. 2017. Mitochondrial recombination and introgression during speciation by hybridization. *Mol Biol Evol.* **34**:1947–1959.

Levri EP, Clark TJ. 2015. Behavior in invasive New Zealand mud snails (*Potamopyrgus antipodarum*) is related to source population. *Biol Invasions.* **17**:497–506.

Li H. 2016. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics* **32**:2103–2110.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100.

Lively CM. 1987. Evidence from a New Zealand snail for the maintenance of sex by parasitism. *Nature* **328**:519–521.

Lonsdale DM, Brears T, Hodge TP, Melville SE, Rottmann WH, Leaver CJ, Lonsdale DM. 1988. The plant mitochondrial genome: homologous recombination as a mechanism for generating heterogeneity. *Philos Trans R Soc Lond B Biol Sci.* **319**:149–163.

Luo S, Valencia CA, Zhang J, Lee N-C, Slone J, Gui B, Wang X, Li Z, Dell S, Brown J, *et al.* 2018. Biparental inheritance of mitochondrial DNA in humans. *Proc Natl Acad Sci U S A.* **115**:13039–13044.

Ma H, O'Farrell PH. 2015. Selections that isolate recombinant mitochondrial genomes in animals. *Elife* **4**:e07247.

Maréchal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. *New Phytol.* **186**:299–317.

McElroy KE, Müller S, Lamatsch D, Bankers L, Fields PD, Jalinsky JR, Sharbrough J, Boore JL, Logsdon JM, Neiman M. 2021. Asexuality associated with marked genomic expansion of tandemly repeated rRNA and histone genes. *Mol Biol Evol.* **38**:3581–3592.

Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, Waggott D, Utiramerur S, Hou Y, Smith KS, *et al.* 2018. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet Med.* **20**:159–163.

Minio A, Massonnet M, Figueroa-Balderas R, Vondras AM, Blanco-Ulate B, Cantu D. 2019. Iso-Seq allows genome-independent transcriptome profiling of grape berry development. *G3* **9**: 755–767.

Mita S, Rizzuto R, Moraes CT, Shanske S, Arnaudo E, Fabrizi GM, Koga Y, DiMauro S, Schon EA. 1990. Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA. *Nucleic Acids Res.* **18**: 561–567.

Mower V. 2018. Structural diversity among plastid genomes of land plants. *Adv Bot Res.* **85**:263–292.

Mower JP, Sloan DB, Alverson AJ. 2012. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ, editors. *Plant genome diversity volume 1: plant genomes, their residents, and their evolutionary dynamics*. Vienna: Springer. p. 123–144.

Muller HJ. 1964. The relation of recombination to mutational advance. *Mutat Res.* **106**:2–9.

Neiman M, Hehman G, Miller JT, Logsdon JM Jr, Taylor DR. 2010. Accelerated mutation accumulation in asexual lineages of a freshwater snail. *Mol Biol Evol.* **27**:954–963.

Neiman M, Lively CM. 2004. Pleistocene glaciation is implicated in the phylogeographical structure of *Potamopyrgus antipodarum*, a New Zealand snail. *Mol Ecol.* **13**:3085–3098.

Neiman M, Paczesniak D, Soper DM, Baldwin AT, Hehman G. 2011. Wide variation in ploidy level and genome size in a New Zealand freshwater snail with coexisting sexual and asexual lineages. *Evolution* **65**:3202–3216.

Neiman M, Taylor DR. 2009. The causes of mutation accumulation in mitochondrial genomes. *Proc Biol Sci.* **276**:1201–1209.

Nilsson MA, Arnason U, Spencer PBS, Janke A. 2004. Marsupial relationships and a timeline for marsupial radiation in South Gondwana. *Gene* **340**:189–196.

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, *et al.* 2022. The complete sequence of a human genome. *Science* **376**:44–53.

Osada N, Akashi H. 2011. Mitochondrial–nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome C oxidase complex. *Mol Biol Evol.* **29**:337–346.

Osman C, Noriega TR, Okreglak V, Fung JC, Walter P. 2015. Integrity of the yeast mitochondrial genome, but not its distribution and inheritance, relies on mitochondrial fission and fusion. *Proc Natl Acad Sci U S A.* **112**:E947–E956.

Paczesniak D, Jokela J, Larkin K, Neiman M. 2013. Discordance between nuclear and mitochondrial genomes in sexual and asexual lineages of the freshwater snail *Potamopyrgus antipodarum*. *Mol Ecol.* **22**:4695–4710.

Palmer JD. 1983. Chloroplast DNA exists in two orientations. *Nature* **301**:92–93.

Perry AS, Wolfe KH. 2002. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J Mol Evol.* **55**:501–508.

Piganeau G, Gardner M, Eyre-Walker A. 2004. A broad survey of recombination in animal mitochondria. *Mol Biol Evol.* **21**: 2319–2325.

Pohjoismäki JLO, Goffart S, Tyynismaa H, Willcox S, Ide T, Kang D, Suomalainen A, Karhunen PJ, Griffith JD, Holt IJ, *et al.* 2009. Human heart mitochondrial DNA is organized in complex catenated networks containing abundant four-way junctions and replication forks. *J Biol Chem.* **284**:21446–21457.

Pop M. 2009. Genome assembly reborn: recent computational challenges. *Brief Bioinform.* **10**:354–366.

Pozzi A, Plazzi F, Milani L, Ghiselli F, Passamonti M. 2017. SmithRNAs: could mitochondria "bend" nuclear regulation? *Mol Biol Evol.* **34**: 1960–1973.

Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol.* **19**:645–653.

Rebolledo-Jaramillo B, Su MS-W, Stoler N, McElhoe JA, Dickins B, Blankenberg D, Korneliussen TS, Chiaromonte F, Nielsen R, Holland MM, *et al.* 2014. Maternal age effect and severe germline bottleneck in the inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A.* **111**:15474–15479.

Satoh TP, Miya M, Mabuchi K, Nishida M. 2016. Structure and variation of the mitochondrial genome of fishes. *BMC Genomics.* **17**:719.

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C. 1997. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene* **205**:125–140.

Schattat MH, Griffiths S, Mathur N, Barton K, Wozny MR, Dunn N, Greenwood JS, Mathur J. 2012. Differential coloring reveals that plastids do not form networks for exchanging macromolecules. *Plant Cell.* **24**:1465–1477.

Shao R, Barker SC. 2011. Chimeric mitochondrial minichromosomes of the human body louse, *Pediculus humanus*: evidence for homologous and non-homologous recombination. *Gene* **473**:36–43.

Sharbrough J, Cruise JL, Beetch M, Enright NM, Neiman M. 2017. Genetic variation for mitochondrial function in the New Zealand freshwater snail *Potamopyrgus antipodarum*. *J Hered.* **108**:759–768.

Sharbrough J, Luse M, Boore JL, Logsdon JM Jr, Neiman M. 2018. Radical amino acid mutations persist longer in the absence of sex. *Evolution* **72**:808–824.

Shearman JR, Sonthirod C, Naktang C, Pootakham W, Yoocha T, Sangsrakru D, Jomchai N, Tragoonrung S, Tangphatsornruang S. 2016. The two chromosomes of the mitochondrial genome of a sugarcane cultivar: assembly and recombination analysis using long PacBio reads. *Sci Rep.* **6**:31533.

Sloan DB. 2013. One ring to rule them all? Genome sequencing provides new insights into the "master circle" model of plant mitochondrial DNA structure. *New Phytol.* **200**:978–985.

Sloan DB, Broz AK, Sharbrough J, Wu Z. 2018. Detecting rare mutations and DNA damage with sequencing-based methods. *Trends Biotechnol.* **36**:729–740.

Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, Havird JC. 2018. Cytonuclear integration and co-evolution. *Nat Rev Genet.* **19**:635–648.

Smith DR, Keeling PJ. 2015. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proc Natl Acad Sci U S A.* **112**:10177–10184.

Stewart DT, Breton S, Blier PU, Hoeh WR. 2009. Masculinization events and doubly uniparental inheritance of mitochondrial DNA: a model for understanding the evolutionary dynamics of gender-associated mtDNA in mussels. In: Pontarotti P, editor. *Evolutionary biology*. Berlin (Germany): Springer. p. 163–173.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* **5**:123–135.

Ujvari B, Dowton M, Madsen T. 2007. Mitochondrial DNA recombination in a free-ranging Australian lizard. *Biol Lett.* **3**:189–192.

van der Sluis EO, Bauerschmitt H, Becker T, Mielke T, Frauenfeld J, Berninghausen O, Neupert W, Herrmann JM, Beckmann R. 2015. Parallel structural evolution of mitochondrial ribosomes and OXPHOS complexes. *Genome Biol Evol.* **7**:1235–1251.

Voineagu I, Narayanan V, Lobachev KS, Mirkin SM. 2008. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc Natl Acad Sci U S A.* **105**:9936–9941.

Wakeley J. 1993. Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J Mol Evol.* **37**:613–623.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* **9**:e112963.

Wang W, Lanfear R. 2019. Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants. *Genome Biol Evol.* **11**:3372–3381.

Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**: 3350–3352.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A.* **84**:9054–9058.

Woloszynska M. 2009. Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—though this be madness, yet there's Method in't. *J Exp Bot.* **61**:657–671.

Wolters JF, Charron G, Gaspary A, Landry CR, Fiumera AC, Fiumera HL. 2018. Mitochondrial recombination reveals mito-mito epistasis in yeast. *Genetics* **209**:307–319.

Wu B, Buljic A, Hao W. 2015. Extensive horizontal transfer and homologous recombination generate highly chimeric mitochondrial genomes in yeast. *Mol Biol Evol.* **32**:2559–2570.

Wu Z, Sloan DB. 2019. Recombination and intraspecific polymorphism for the presence and absence of entire chromosomes in mitochondrial genomes. *Heredity (Edinb).* **122**:647–659.

Wu Z, Waneka G, Broz AK, King CR, Sloan DB. 2020. *MSH1* is required for maintenance of the low mutation rates in plant mitochondrial and plastid genomes. *Proc Natl Acad Sci U S A.* **117**: 16448–16455.