

Estimating the rate and spectrum of mitochondrial mutation in *Daphnia* using
next-generation sequencing

A Thesis
Presented to
The Division of Mathematics and Natural Sciences
Reed College

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Arts

Fenner Macrae

May 2016

Approved for the Division
(Biology)

Sarah Schaack

Acknowledgements

When I first heard about Reed's thesis requirement as a high school student, I remember how completely impossible and overwhelming it all sounded. This feeling has persisted throughout all of my time at Reed, and even now, as I am looking at a complete draft of my thesis, I am suspicious of its materiality. I feel grateful to the innumerable people who have propelled me through this experience.

To Sarah Schaack, who endured my numerous stress-induced, inarticulate monologues about mitochondrial inheritance, and always somehow managed to cut to the point, or point me in the right direction. She has been an incredible thesis advisor, and I am very thankful for all of her guidance.

To my readers, Derek Applewhite, Adam Groce and Tamara Metz.

To Michael Pitts and Todd Schlenke, who I had the great fortune to work with during my time at Reed. Michael approaches research with a relentless curiosity, and a keen eye for detail, which is commonly encouraged in science but extremely rare to come across in practice. I have tried to emulate him in all of my work at Reed, and I expect I will continue to do so. I have Todd to thank for my first and only foray into wet-lab biology - he taught me much of what I know about genetics, and everything I know about fruit flies.

To my friends, Maddie Cantrell, Ivan Nyby and Roman Meza, who despite not going to Reed, have put up with my missed calls, late messages, and generally poor social skills with considerable patience. I love you all, and I am immensely grateful that you are still willing to associate with me.

To Kathi Schwaiger, who has been wonderful and thoughtful throughout this entire year, despite the fact that I spent much of her thesis whining about organic chemistry. My time at Reed has been worth it, if only to have met her.

To my father and my stepmother for their love and support, and the Hello Kitty emblazoned care packages.

And to my mother, for whom I have everything to thank, always.

Table of Contents

Introduction	1
Mitochondrial Diversity and Evolutionary History	1
Mitochondrial Origins	1
Phagocytic Models	2
Fusion Models	3
Present Diversity	5
Mitochondrial Physiology	6
Organelle Structure	6
Mitochondrial Function	8
Mitochondrial DNA	9
Peculiarities of Mitochondrial Inheritance	11
Maternal Inheritance	11
Heteroplasmy	12
Recombination	14
The Mitochondrial Genetic Bottleneck	15
Estimating the Mitochondrial Mutation Rate	18
The Molecular Clock	18
The Mutation Rate as Trait Under Selection	19
Dynamics of Mitochondrial Mutation	20
Direct Estimates of the Mutation Rate	21

The Current Study	22
Materials and Methods	25
Mutation Accumulation	25
Sequencing	26
Sequence Processing	26
Read Filtering	26
nuMT Removal	26
Variant Calling	30
Statistics	30
Mutation Rate Calculations	30
Mutation Rate Comparisons	31
Detection of Shifts in Allele Frequency	31
Results and Discussion	33
Detected Mutations	33
Mutation Rate Estimates	37
Minor Allele Frequencies	40
Shifts in Allele Frequency	40
nuMT Analysis	41
Potential Sources of Bias and Error	44
Future Directions	45
Appendix	47
Appendix 1: Pipeline Commands	47
Appendix 2: Filtered Mutations	49
Appendix 3: Analysis Scripts	52
Mutation Rate Calculations	52
Allele Frequency Shift Significance Testing	53

Works Cited	55
-----------------------	----

List of Tables

1	MA line information	25
2	List of detected variants	36
3	Mutation Statistics	37
4	Table of filtered variants	49

List of Figures

1	Comparison of the dominant theories for the origin of mitochondria	4
2	The current model of mitochondrial structure	7
3	Representative micrographs of fragmented and connected mitochondrial network topologies	8
4	Mitochondrial genome structure.	10
5	Two competing models of the mitochondrial genetic bottleneck	16
6	Relationship between the mitochondrial and nuclear base substitution rate	23
7	Cyclical parthenogenesis in natural populations of <i>Daphnia pulex</i>	24
8	Sequence processing pipeline	27
9	Example nuMT Data showing spikes in read depth	29
10	Read depth across the mitochondrial genome	34
11	Locations of detected variants on the mitochondrial genome	35
12	Effect of allele frequency cutoff on mutation rate estimates	39
13	Minor allele frequency distributions.	41
14	Shifts in heteroplasmic allele frequency	42
15	BLAST results for detected nuMTs	44

Abstract

Mutation is the ultimate source of genetic variation, and thus is a driving force of evolution. Despite this fact, rates of spontaneous mutation have only been directly estimated in a few organisms, and little is known about their patterns of variation across taxa. Estimating rates of mutation in the mitochondrial genome is of particular importance, due to the widespread usage of mitochondrial genes as phylogenetic markers. In the present study, we detected mutations in next-generation sequencing data from a set of *Daphnia pulex* mutation accumulation lines. The majority of mutations were low-frequency heteroplasmic insertions and deletion mutations (indels) at homopolymeric tracts, consistent with previous examinations of mitochondrial mutations, but several base substitutions were also detected. We estimated the base-substitution rate to be 22.7×10^{-8} , and the rate of insertions and deletions (indels) to be 8.0×10^{-8} , resulting in an overall mutation rate of 32.3×10^{-8} . This differs significantly from a previous estimate in *D. pulex*, which found an increased rate of indels, but a decreased rate of base-substitutions. We suspect that differences in the error profiles of next-generation sequencing and Sanger sequencing may partially explain this discrepancy. This suggests that caution should be taken in comparing the results of mutation-accumulation experiments which utilize differing sequencing methodologies.

Dedication

To Ellie and Feo.

Introduction

Mutation is the ultimate source of genetic variation, and thus, is a driving force of evolution, similar to natural selection or genetic drift. However, the mutation rate of an organism is governed by the molecular machinery of DNA replication and mutation repair, and thus, is itself a trait under selection. Indeed, mutation rates appear to be highly variable between species, suggesting them to be a product of the specific evolutionary history of an organism (Lynch, 2010). Mutation rates in the mitochondrial and nuclear genomes often differ markedly, for example, the mitochondrial mutation rate is thought to be on average 20 times higher than the nuclear rate in mammals, but 20 times lower in plants (Lynch et al., 2006). In recent years, direct estimates of mitochondrial mutation rate have been made in a small sample of model organisms, providing an initial glimpse of the patterns of variance across taxa. In the following sections, I outline the evolutionary history and present diversity of mitochondria, examine the peculiarities of mitochondrial inheritance which influence the mitochondrial mutation rate, and explore the potential reasons for its deviation from the nuclear rate.

Mitochondrial Diversity and Evolutionary History

Mitochondrial Origins

The origin of mitochondria represents an important example of lateral genetic exchange between the bacterial and eukaryotic domains early in their diversification. The discovery that mitochondria possess their own genome isolated from the nuclear genome provided the first strong indication of their independent origin (Nass & Nass, 1963). Later, sequence analysis showed that genes within the mitochondrial genome, and nuclear genes involved in mitochondrial function are homologous to modern α -proteobacteria while the remainder of the nuclear genome is archaeal in origin (Doolittle & Brown, 1994; Iwabe et al., 1989). This suggests that the

membrane-bound nucleus, a defining feature of eukaryotes, evolved in the archaeal lineage (Embley & Martin, 2006). Endosymbiotic theory, the dominant explanation for the origin of mitochondria, proposes that certain membrane-bound organelles are derived from symbiotic interactions between single-celled eukaryotes and bacteria (Margulis, 1970; Martin et al., 2015). Under endosymbiotic theory, the bacterial ancestor of mitochondria came to reside inside the ancestor of eukaryotes as a symbiont. This symbiont freely replicated within its host, while providing it with some selective advantage. The precise nature of the interaction between the bacterial ancestor of mitochondria and the ancestor of eukaryotes, which eventually lead to mitochondrial residence in eukaryotic cells is heavily debated (Martin et al., 2015). Explanations for this problem can be divided into 'phagocytic', or 'fusion' models, based on the mechanism of internalization.

Phagocytic Models

The first application of endosymbiotic theory to mitochondria suggested that they are derived from the phagocytosis of a bacterium by an early eukaryote (Figure 1A) (Margulis, 1970). This hypothesis predicts that the eukaryote was a phagotroph, which already possessed a compartmentalized cellular structure and a complex cytoskeletal system (Cavalier-Smith, 2002; Archibald, 2015). Proponents claim that the flexible outer membrane and sequestered nucleus of eukaryotic cells are prerequisites for phagocytosis (Cavalier-Smith, 2002). However, the assumption that phagocytosis is required for internalization has been severely challenged by the discovery of γ -proteobacterial symbionts inside of the β -proteobacterial endosymbionts of mealybugs (von Dohlen et al., 2001). Additionally, recent phylogenetic analyses suggest that phagosomes arose independently in multiple lineages relatively late in eukaryotic diversification (Yutin et al., 2009).

Phagocytic explanations have fallen out of favor in recent years due to a lack of empirical support. The discovery of eukaryotic cells which lack mitochondria would provide strong evidence for phagocytic hypotheses, but such evidence has been lacking. Until recently, one such group of amitochondriate eukaryotes, named Archezoa, were thought to exist (Cavalier-Smith, 1987). Archezoa consisted of unicellular eukaryotes, such as microsporidia and diplomonads, and were hypothesized to represent an intermediate link between eukaryotes and Archaea. However, these organisms were later shown to contain organelles derived from mitochondria but which had subsequently lost their genome (Embley et al., 2003; Tovar, 2007). Since then, no

amitochondriate eukaryotes have been documented, suggesting that if the origin of the mitochondria followed that of the nucleus, it was shortly thereafter. Perhaps even more damning, it has been recently demonstrated that the high energy requirements of the eukaryotic cell could not even be met by an amitochondriate organism (Lane & Martin, 2010). Many of the assumptions of phagocytic models now seem misguided, leading researchers to pursue alternative explanations.

Fusion Models

Fusion models propose that mitochondria stem from an extracellular symbiotic relationship between archaeal and bacterial cells. Besides the insistence on phagotrophy, phagocytic hypotheses are often vague concerning the metabolic capabilities of first eukaryote, generally supposing it to be an aerobic heterotroph (Margulis, 1970; Cavalier-Smith, 2002). One problem with this assumption is that it does not explain how the endosymbiont initially benefited its host. While the endosymbiont may have utilized a more efficient means of aerobic ATP production, it is unclear why it would release ATP into its extracellular environment to be utilized by its host. Under the hydrogen hypothesis, the dominant fusion model, the mitochondria and the nucleus both originated from the association of an archaeal methanogen and a hydrogen-producing bacterium (Figure 1B; Martin & Mller, 1998). The methanogen fed off the by-products of the bacterium's metabolism, becoming dependent on its proximity, eventually fusing with it to form a single cell. These byproducts would have been freely released into the environment, providing an obvious means for benefiting the host upon internalization. Gratifyingly, associations of this kind are abundant in nature (Bryant et al., 1967; Broers et al., 1990; Finlay et al., 1993).

A number of fusion-based theories which diverge from the hydrogen hypothesis on the identities of the symbiotic partners have been proposed. Models which replace the hydrogen-producing bacterium with one which metabolized oxygen (Vellai & Vida, 1999), or sulfur compounds (Searcy, 1992) have been forwarded. The existence of hydrogenosomes, a class of anaerobic mitochondria found in diverse eukaryotes, which produce hydrogen as a metabolic byproduct, favors the hydrogen hypothesis above these alternatives. While the hydrogen hypothesis is convincing from a metabolic perspective, phylogenetic data has not been supportive. Multiple independent phylogenomic analyses have recently suggested that eukaryotes arose within the archaeal TACK superphylum (Guy & Ettema, 2011; Williams et al.,

2012; Guy et al., 2014), which does not contain methanogens. Consequently, the eocyte (Lake et al., 1984; Cox et al., 2008) and pHAT (Martijn & Ettema, 2013) hypotheses, which are more consistent with this tree have been advocated. The identification of the archaeal progenitor of eukaryotes would likely put this debate to rest, but extensive horizontal gene transfer among prokaryotes has made this difficult to determine in practice (Archibald, 2015).

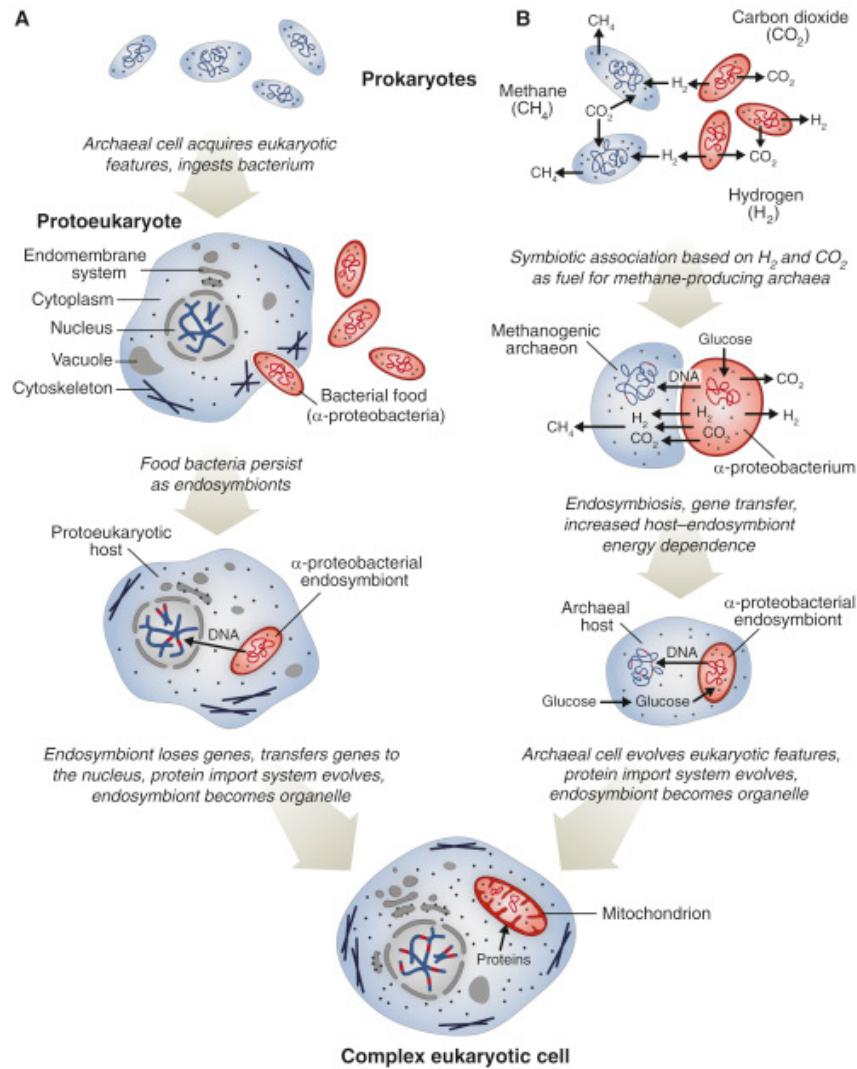


Figure 1: Comparison of the dominant theories for the origin of mitochondria. (A) In phagocytic models, a phagotrophic proto-eukaryote internalizes an α -proteobacterium, which evades degradation and replicates inside its host. (B) In fusion models, an archaeon and an α -proteobacterium, become associated through the symbiotic exchange of metabolic byproducts and eventually unite to form a single cell. Figure sourced from Archibald, 2015.

Present Diversity

Since their ancient origin, mitochondria have come to exhibit remarkable functional diversity among eukaryotic lineages. Proto-mitochondria were likely present in the last eukaryotic common ancestor (Archibald, 2015), which is estimated to have lived between 1.87-1.68 bya (Parfrey et al., 2011). Accordingly, mitochondria-related organelles (MROs) have now been documented in every eukaryotic group (Baldauf, 2003; Shiflett & Johnson, 2010). MROs can be broadly divided into three classes, classical mitochondria, hydrogenosomes and mitosomes (Shiflett & Johnson, 2010). Classical mitochondria are present in animals, plants and many unicellular eukaryotes. They are primarily involved in aerobic ATP production, but also play important roles in apoptosis and regulation of membrane potentials. Hydrogenosomes have thus far been documented in anaerobic species of trichomonads (Lindmark & Müller, 1973), ciliates (Yarlett et al., 1981) and chytridiomycete fungi (Yarlett et al., 1986). They produce ATP like classical mitochondria, but rely on substrate-level, rather than oxidative, phosphorylation, generating molecular hydrogen as a byproduct (Müller, 1994). Mitosomes which are found amongst parasitic unicellular taxa such as microsporidia (Williams et al., 2002), and entamoeba (Mai et al., 1999), on the other hand, completely lack ATP production. In general, only classical mitochondria possess their own genome, whereas mitosomes and hydrogenosomes depend entirely on genes encoded in the nucleus (but see Boxma et al., 2005 and Wawrzyniak et al., 2008 for exceptions to this rule). Despite their widely divergent function, it is now commonly accepted that all MROs are derived from the same endosymbiotic event (Embley et al., 2003; Shiflett & Johnson, 2010). The only features which appear to be conserved across all MROs are their double membrane structure and their role in the assembly of iron-sulfur (Fe-S) clusters (Lill & Kispal, 2000; Shiflett & Johnson, 2010).

The existence of these highly divergent mitochondrial subtypes, in addition to intermediate forms, emphasizes the dynamism of mitochondrial evolution. Hydrogenosomes and mitosomes are not monophyletic, rather, they are scattered across a diverse set of eukaryotic groups, suggesting numerous parallel evolution events (Shiflett & Johnson, 2010). The distribution of hydrogenosomes and mitosomes appears to be related to a set of commonly observed environmental niches. The conversion of classical mitochondria to hydrogenosomes is presumably an adaptation to hypoxic environments where aerobic respiration is impossible (Hackstein et al., 1999). Mitosomes appear to result from conditions where an alternative source

of energy eliminates the need for mitochondrial ATP production. Microsporidia, for example, have evolved ATP translocases which allow them to source ATP from their hosts, while *Entamoeba histolytica* has shifted ATP production to the cytosol (Mller et al., 2012). Intermediate forms of MROs have also been described, such as those found in *Nyctotherus ovalis* (Boxma et al., 2005) and *Blastocystis hominis* (Stechmann et al., 2008) which utilize components of both hydrogenosomal and classical mitochondrial metabolism.

The independent origin and subsequent diversification of mitochondria should be taken into account when considering their mutation rate. Despite being almost entirely integrated into the eukaryotic cell, the mitochondrial genome is still subject to different selective pressures than the nuclear genome. While the mitochondria may act to benefit its host, evolution still favors those phenotypes which are most successful in replicating their DNA. As will be discussed in later sections, mitochondrial evolution is influenced by the separate, and sometimes conflicting, selective pressures on mitochondrial and host function. Similarly, the diversity of mitochondrial function should be taken into account when considering how the mutation rate varies across taxa. In the case of *N. ovalis*, for example, which metabolizes anaerobically but retains a genome, a high mitochondrial mutation rate might be expected due to relaxed selective pressure on genes involved in aerobic respiration. More generally, the mitochondrial mutation rate may be influenced by the importance of the mitochondrial gene function to the specific organism in question.

Mitochondrial Physiology

Organelle Structure

With the evident diversity of mitochondria in mind, we now shift our attention to organization and function of mitochondria in the animal lineage. Perhaps the dominant feature of the mitochondrion is its double-membrane structure, in which the outer and inner membranes enclose an intermembrane space (Figure 2). The inner membrane has a much larger surface area than the outer membrane, creating numerous membrane pockets named cristae. The typical textbook illustration of cristae represents them simply as a random folding of the inner membrane, but modern electron tomography has shown this to be false (Perkins et al., 1997; Mannella, 2006). Instead cristae are tubular shaped and can merge to form parallel flat plates with large surface area, which interface with the intermembrane at

small passages called cristae junctions (Perkins et al., 1997; Perkins & Frey, 2000). The cristal membrane houses the majority of respiratory complexes, with a small proportion instead localized to the inner membrane (Gilkerson et al., 2003; Logan, 2006). The intercristal space, formed between cristal folds, houses the mitochondrial DNA (mtDNA) and associated transcription and translation machinery (Figure 2A). mtDNA molecules are packaged inside protective protein coats called nucleoids (Bogenhagen, 2012), which associate tightly with the inner membrane (Brown et al., 2011). Estimates from human cells suggest that mitochondria possess an average of three nucleoids, with the majority containing only one (Satoh & Kuroiwa, 1991).

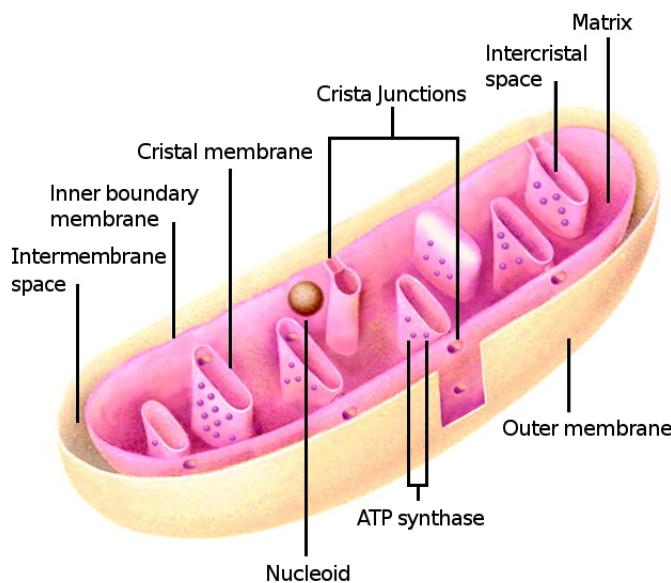


Figure 2: The current model of mitochondrial structure. Figure modified from Perkins & Frey, 2000.

While there is a tendency to depict mitochondria as isolated static organelles, in many eukaryotes they instead form a complex reticulum, which coordinates to meet metabolic requirements. Mitochondria are trafficked along both microtubules and actin filaments within the cell (Hollenbeck & Saxton, 2005), allowing them to localize to energy demand. The importance of efficient mitochondrial transport is especially evident in neurons, which can require mitochondria to be transported multiple centimeters from the soma to the high ATP demands of the synapse (Schwarz, 2013). Mitochondria are capable of fission, the division of a mitochondrion into two functional organelles, and fusion, the merging of two mitochondria into a single

organelle (Youle & Bliek, 2012). Fission is required to supply dividing cells with a mitochondrial population and to replace damaged mitochondria in mature cells. Fusion enables mitochondria to exchange proteins and genetic material. This appears to enable the mitochondrial network to compensate for dysfunctional mitochondria by sharing gene products between organelles (Nakada et al., 2001). Fusion and fission may also have a higher level function in altering the connectedness of the mitochondrial network. *In vivo* imaging of mitochondrial networks has shown them to be in a constant state of fusion and fission, switching between highly connected and fragmented topologies (Figure 3; Jakobs et al., 2003; Rafelski, 2013).

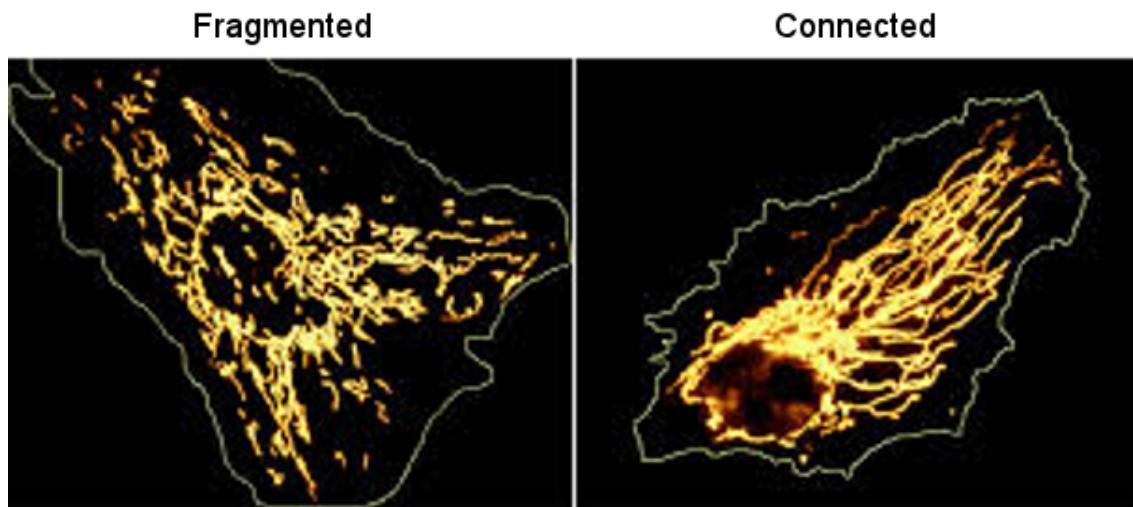


Figure 3: Representative micrographs of fragmented and connected mitochondrial network topologies. Both cells are expressing pEYFP-mitochondrial vector to label mitochondria. Left cell is an Indian Muntjac deer skin fibroblast and right is bovine pulmonary artery endothelial cell (BPAE). Figure modified from Rafelski, 2013, with images originally sourced from the Nikon MicroscopyU website (<http://www.microscopyu.com>).

Mitochondrial Function

ATP synthesis by aerobic respiration is the most fundamental role of mitochondria in animal cells. Aerobic respiration facilitates the conversion of diverse nutrient sources into usable energy for endergonic cellular processes. The majority of aerobic respiratory processes, such as β -oxidation, the citric acid cycle and oxidative phosphorylation (OXPHOS) take place at mitochondria. OXPHOS, which occurs at cristae, is responsible for most of the ATP produced by aerobic respiration. Succinate, NADH (reduced nicotinamide adenine dinucleotide) and

FADH₂ (flavin adenine dinucleotide), produced by the citric acid cycle and β -oxidation, provide input to the OXPHOS system, which produces ATP from ADP (adenosine diphosphate) and phosphate. OXPHOS consists of a series of redox reactions at mitochondrial complexes I-IV, which pass electrons to progressively more electronegative compounds. The resultant free energy is utilized to pump protons from the mitochondrial matrix into the intermembrane space, producing an electrochemical gradient. Finally, ATP synthase harnesses this gradient to phosphorylate ADP to ATP.

In addition to their traditional role in metabolism, mitochondria also act as important regulators of stress response and programmed cell death. Mitochondria are thought to be a crucial node in signaling pathways which can both positively and negatively regulate cell death (Galluzzi et al., 2012). Proteins tethered to the outer mitochondrial membrane have been shown to negatively regulate apoptosis (Hengartner et al., 1992). Conversely, the dispersion of mitochondrial proteins into the cytoplasm via mitochondrial outer membrane permeabilization is a powerful pro-apoptotic signal (Zamzami et al., 1995; Tait & Green, 2010). Additionally, mitochondria have been implicated in cytoprotective responses to disruptions of homeostasis, such as nutrient starvation (Gomes et al., 2011) and viral infection (Arnoult et al., 2011).

Mitochondrial DNA

mtDNA are small, circular, double-stranded genomes which exist in a haploid state in eukaryotic cells (Figure 4A) Unlike the eukaryotic nuclear genome, but similar to bacterial genomes, mtDNA is almost entirely coding. In animals, mtDNA are approximately 15-20 kb, consisting of a remarkably conserved set of 37 genes (Figure 4B). Thirteen of these genes are protein-coding, encompassing many of the core proteins in the OXPHOS system. Of the rest, 2 are rRNAs and 22 are tRNAs utilized specifically in the translation of mitochondrial proteins. The genic region of mtDNA is transcribed continuously into a single polycistronic molecule, which is subsequently cleaved. Additionally, mtDNA possess an approximately 1 kb untranscribed region named the control region, or D-loop, which contains the origin of replication. The genes found in mtDNA represent a tiny fraction of the approximately 1000 genes whose products localize to the mitochondria (Pagliarini et al., 2008). The majority of these genes are thought to have originally resided in the mitochondrial genome, but were later transferred to the nucleus in a process called endosymbiotic gene transfer

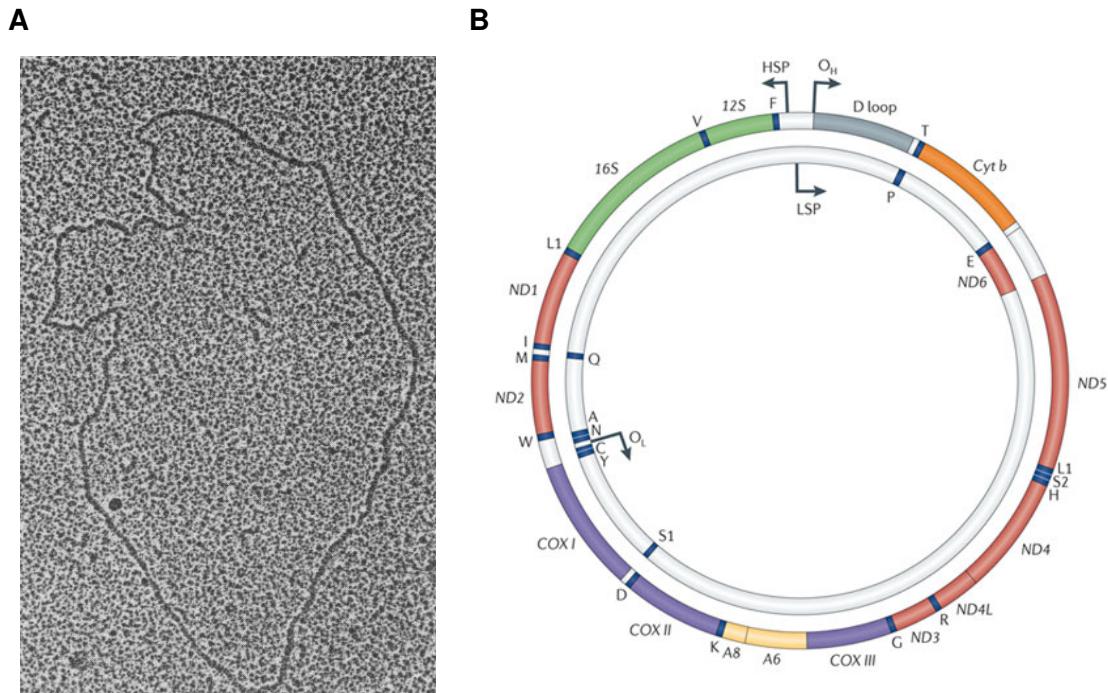


Figure 4: Mitochondrial genome structure. (A) Electron micrograph of purified mtDNA (B) The contents of the mitochondrial genome. Shown are the components of the mitochondrial complex I (red), complex III (orange), complex IV (purple), ATP synthase (yellow), ribosomal RNAs (green), and tRNAs (blue). Origins of replication are shown for the heavy strand (O_H) and the light strand (O_L), as well as the promoters of transcription for the heavy strand (HSP) and the light strand (LSP). Figures adapted from Nass, 1966 (A) and Schon et al., 2012 (B).

(Timmis et al., 2004). This process has contributed to the reduced state of modern mtDNA.

The quantification of mtDNA apportionment into the cell, mitochondria or nucleoid has been the subject of much research in recent years. The mtDNA copy number per cell is massively variable, with estimates ranging between 10-10⁴ copies depending upon cell type and developmental stage (Shuster et al., 1988; White et al., 2008). During mammalian oogenesis, cellular populations of mtDNA can expand from just 10 copies to 100,000-200,000 (Cao et al., 2007; Shoubridge & Wai, 2007; White et al., 2008). Populations subsequently shrink to the approximately 10³ copies which are present in most cells during embryogenesis. In recent years, mtDNA copy number has taken on clinical significance, with a number of studies which suggest that mtDNA depletion is associated with various cancer subtypes (Thyagarajan et al., 2012, 2013; Zhou et al., 2014; El-Hattab & Scaglia, 2013). This association does not appear to be present in all tissues however, indicating that some patterns of mtDNA copy number variation may be tissue dependent. Unfortunately, data on regional

variation in mtDNA populations is lacking, though one preliminary study in mice suggested that differences in mtDNA copy number between brain regions are broadly consistent with the incidence of certain cancers (Fuke et al., 2011). On a coarse scale at least, mtDNA copy number variation certainly occurs, for example in the case of mammalian red blood cells which completely lack mitochondria and contain only a few mtDNA (Shuster et al., 1988). Further variation likely also occurs at the level of the nucleoid and the mitochondria. Nucleoids can contain a variable number of mtDNA with the majority containing only a single copy (Kukat et al., 2011) but may be capable of containing up to ten (Legros et al., 2004; Iborra et al., 2004; Brown et al., 2011). The functional significance of this variation is not understood, but may be relevant to partitioning of mtDNA during cell division and the population genetic environment faced by mtDNA in cells (Birky, 2001).

Peculiarities of Mitochondrial Inheritance

Maternal Inheritance

Animal mtDNA is predominantly propagated by uniparental inheritance with offspring receiving their mtDNA haplotype from their mother. Only one exception has been documented in bivalve mollusks, in which males receive mtDNA from both parents while females receive it only from their mother (Zouros et al., 1994). However, most animals likely do not exhibit strict maternal mtDNA inheritance, instead paternal DNA is passed to offspring in small, and generally inconsequential, quantities. Many examples of 'paternal leakage' have now been documented in both intraspecific and interspecific crosses (Kvist et al., 2003; Meusel & Moritz, 1993; Zhao et al., 2004; Wolff et al., 2013; Robison et al., 2015). In one particularly interesting case, a patient was found with mitochondrial myopathy stemming from a mtDNA mutation inherited from his father (Schwartz & Vissing, 2002). It is possible that some degree of paternal leakage is ubiquitous in animals, generating phenotypic consequences which are difficult to detect.

Specialized mechanisms for purging paternal mtDNA from offspring have been described in many species (Sato & Sato, 2013). In some mammals the outer mitochondrial membrane is ubiquitinated in sperm cells, which marks them for degradation upon fertilization (Sutovsky, 2003). In tunicates, sperm mitochondria are prevented from entering the egg entirely (Ursprung & Schabtach, 1965). In the rare cases where paternal mtDNA manages to bypass these obstacles, it subsequently

becomes diluted into insignificance by maternal mtDNA (Wolff & Gemmell, 2008). The evolution of such barriers indicates that maternal inheritance is being maintained by selection, but its advantages over biparental inheritance are unclear. One popular hypothesis posits that maternal inheritance prevents selfish mitochondrial evolution (Hastings, 1992). Biparental inheritance would produce intracellular competition between mtDNA haplotypes. In such cases, selective pressure of replicative efficiency would be increased, perhaps at the expense of organelle function. Uniparental inheritance might then have evolved to prevent such maladaptive competition, by promoting genetic homogeneity. An alternative hypothesis stems from the observation that mutagenic reactive oxygen species (ROS) are given off during OXPHOS. During the race to the egg, sperm are highly metabolically active and might accumulate mtDNA damage (Allen, 1996). The elimination of paternal mtDNA would then act to prevent deleterious mutations from entering the germline.

Heteroplasmy

Even within uniparental inheritance, the high copy number of mtDNA means that all copies need not be genetically identical. This state is known as heteroplasmy, the coexistence of distinct mtDNA haplotypes within an organism. The opposite state, in which all mtDNA have the same haplotype, is known as homoplasmy. Heteroplasmy can be produced either by mutation, or by paternal leakage (White et al., 2008). All mitochondrial mutations must initially arise in a heteroplasmic state and subsequently become fixed or lost, similar to a mutant allele within a population. Due to its rarity, paternal leakage is likely a minor source of heteroplasmy in typical populations, but may be important in hybrid zones where barriers against paternal mtDNA appear to break down (Kaneda et al., 1995; Dokianakis & Ladoukakis, 2014; Radojii et al., 2015).

Initially, mtDNA populations were thought to exist predominantly in homoplasmic states. Early sequencing efforts concluded that mtDNA populations were genetically identical with only a few rare exceptions (Monnat & Loeb, 1985). These results were surprising to researchers at the time because traditional population genetic theory predicts that heteroplasmy should be extremely common. Due to large mtDNA population sizes, organismal selection on individual alleles is expected to be weak, meaning that even strongly deleterious haplotypes could theoretically be maintained for generations. Additionally, neutral heteroplasmic alleles might persist for even longer periods before becoming fixed or eliminated by genetic drift. The apparent

rarity of heteroplasmy in the face of these arguments suggested the presence of specific cellular mechanisms for maintaining homoplasmy. Researchers speculated that a combination of nonrandom segregation and random genetic drift might together act to purge heteroplasmic states (Birky, 2001).

In recent years, modern sequencing technologies have lead to more precise estimates of heteroplasmy, suggesting it to be far more common than previously supposed. Occurrences of heteroplasmy have now been documented across a broad range of taxa (Solignac et al., 1983; Mat et al., 2007; Paduan & Ribolla, 2008; Mjelle et al., 2008; White et al., 2008; Magnacca & Brown, 2010; Payne et al., 2013; Robison et al., 2015). One recent sequencing study in humans found that 88 % of people demonstrated some degree of heteroplasmy, with the majority exhibiting multiple heteroplasmic sites (Naue et al., 2015). Even this is likely to be an underestimate because heteroplasmies at frequencies <5% could not be reliably detected. It has been shown that heteroplasmies at frequencies as low as 0.15 % are inherited in humans (Guo et al., 2013), suggesting that low-frequency heteroplasmy may be ubiquitous.

Heteroplasmy can cause alterations in mitochondrial phenotype, and diseases stemming from heteroplasmic alleles are commonly observed. Interestingly, metabolic defects related to heteroplasmic mutations tend to be more severe than those stemming from homoplasmic mutations (Stewart & Chinnery, 2015). The phenotypic effect of a particular heteroplasmic allele is generally proportional to its frequency. Often, the frequency of a deleterious heteroplasmic allele must pass a certain threshold before disease symptoms can be observed (Taylor & Turnbull, 2005). The threshold for a specific mitochondrial disease is presumably related to the ability of wild-type mtDNA to compensate for dysfunctional variants. This threshold effect is also dependent on the variable metabolic requirements of different cell types, leading to tissue-specific mitochondrial disorders (Taylor & Turnbull, 2005). Heteroplasmy can also be deleterious in cases of nuclear-mitochondrial incompatibility. Cytoplasmic hybrid mice with a mixture of two neutral heteroplasmic variants are less fit than their homoplasmic counterparts (Sharpley et al., 2012), likely due to incongruous nuclear and mitochondrial genotypes. Due to the importance of nuclear-mitochondrial cohesion, heteroplasmy is probably detrimental in most cases.

Negative effects of heteroplasmy may be partially offset by the ability to adjust mitochondrial populations to the specific requirements of different tissues. Multiple independent studies have found that the frequency of certain heteroplasmic variants is correlated with tissue type (Samuels et al., 2013; Li et al., 2015; Naue et al., 2015). Certain loci are found to be heteroplasmic at similar frequencies in the same tissue

between unrelated individuals. Interestingly, these alleles do not appear to originate from the germline, instead they are likely the result of recurrent somatic mutation. Such tissue-specific sites are clustered at regions which are thought to be involved in mtDNA replication, suggesting a functional role (Samuels et al., 2013). Variations in heteroplasmy frequency at certain sites may represent a mechanism for catering mitochondrial function to the specific requirements of each tissue.

Recombination

Mitochondrial recombination is common in plants and fungi (Barr et al., 2005), but its absence in animal mitochondria was generally agreed upon until recently (Wolstenholme, 1992). The discovery of recombination machinery in mitochondrial extracts cast doubt on this assumption (Thyagarajan et al., 1996). Recombination has since been documented in a number of hybrids and species with unique mitochondrial inheritance systems (Zouros et al., 1992; Gantenbein et al., 2005; Guo et al., 2006; Ciborowski et al., 2007), but these results may not be generalizable. More promising cases of intraspecific recombination have been found in humans and lizards (Kraytsberg et al., 2004; Ujvari et al., 2007), and examination of published sequence data suggests that low levels of recombination may be common (Ladoukakis & Zouros, 2001; Piganeau et al., 2004; Tsaousis et al., 2005). However, a recent study which followed artificially produced heteroplasmic mice for 50 generations found no evidence of recombination (Hagstrom et al., 2014). Furthermore, the same study showed that many previous cases of recombination may have been false positives due to jumping PCR errors. Thus, definitive evidence for mitochondrial recombination is still lacking. It may be that recombination is completely absent in typical animal mitochondria, but it is also possible that it occurs only extremely infrequently, or under certain conditions.

The question of mitochondrial recombination is of particular relevance in deleterious mutation accumulation in mtDNA. Uniparentally inherited genomes without recombination seemingly have no mechanism for purging deleterious mutations upon fixation. Continuous mutation accumulation should therefore lead to a progressive loss of fitness, and eventual extinction. Selection can slow down this effect, but eventually even the most fit haplotype will be lost due to genetic drift (Muller, 1964). This paradox of mitochondrial evolution is known as Muller's ratchet. However, models find that even low recombination rates are capable of reversing Muller's ratchet (Charlesworth et al., 1993). This suggests that mitochondrial

recombination may be of profound importance to mitochondrial evolution, even if it is extremely infrequent.

It is possible that recombination at mitochondria is a more targeted process than has been assumed. One study of cytoplasmic mice found low levels of recombination, but only in mtDNA which contained deletions (Sato et al., 2005). Recombination might then be specific to deletions, or highly deleterious mutations more generally. This hypothesis was tested in a recent report which generated *Drosophila* heteroplasmic for two deleterious mtDNA haplotypes which could be rescued by recombination (Ma & O'Farrell, 2015). Populations typically went extinct after several generations, but a single population recovered and was found to possess recombinant mtDNA. This effect was present only in cases of severe selective pressure, suggesting that recombination may only become relevant under extreme conditions. Further research is required to determine if this represents a general phenomenon, but a selection-dependent mechanism for mitochondrial recombination may explain some contradictory findings in the literature.

The Mitochondrial Genetic Bottleneck

Levels of heteroplasmy often fluctuate markedly between mother and offspring. Early studies which tracked heteroplasmy intergenerationally in *Chlamydomonas*, yeast, mice, and dairy cows, found that offspring returned to homoplasmy within just a few generations (Sager & Ramanis, 1968; Birky & Strausberg, 1977; Olivo et al., 1983; Ashley et al., 1989; Jenuth et al., 1996). This effect bears a close resemblance to the rapid fixation of alleles caused by genetic drift in small populations, leading researchers to postulate that mtDNA populations become greatly diminished during some early stage in development (Bergstrom & Pritchard, 1998). This hypothesized stage of reduced mtDNA copy number is referred to as the mitochondrial genetic bottleneck.

Intracellular mtDNA copy number is incredibly dynamic during oogenesis, making it difficult to discern the precise timing and mechanism of the bottleneck. Several studies in mice have found that mtDNA populations are significantly reduced in the earliest emerging germ cells of the developing female embryo (Jenuth et al., 1996; Shoubridge & Wai, 2007; Cree et al., 2008; Wai et al., 2008). Specifically, they find that during the initial mitotic production of primordial germ cells (PGCs) immediately post-implantation, mtDNA populations shrink to approximately 10^2 copies per cell (Wai et al., 2008). Population genetic models suggest that

unequal partitioning into daughter cells during the production of PGCs, along with the consequent amplification of mtDNA in mature oocytes, are sufficient to produce the observed generational heteroplasmy fluctuations (Figure 5A; Cree et al., 2008). However, one research group has consistently found no evidence of reduced mtDNA copy number during oogenesis using similar methodology (Cao et al., 2007, 2009). Instead, they argue that the bottleneck is produced by the replication of a subpopulation of mtDNA during oocyte maturation (Figure 5B). At present, these contradictions have yet to be satisfactorily explained, and no consensus has been reached on the precise mechanism of the mitochondrial bottleneck.

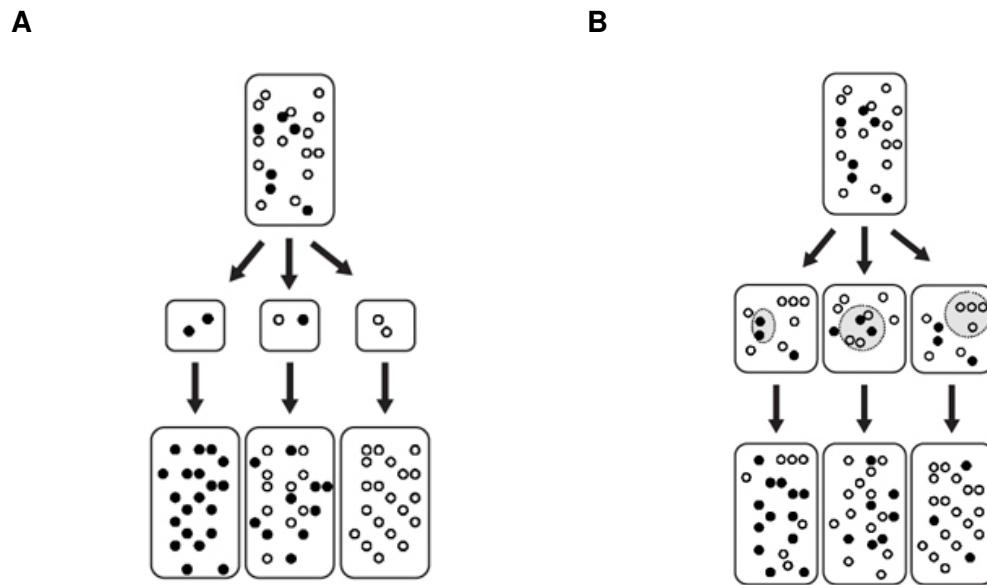


Figure 5: Two competing models of the mitochondrial genetic bottleneck. (A) Populations of mtDNA become significantly reduced during oogenesis, leading to rapid haplotype fixation through genetic drift. (B) A subpopulation of mtDNA are selectively amplified during oocyte development, causing a decrease in heteroplasmy without a reduction in mtDNA copy number. Figure modified from Cao et al., 2007.

A recent study in rhesus macaques provides evidence for an additional mitochondrial bottleneck present during somatic development. Heteroplasmic oocytes were produced by cytosolic fusion, fertilized and transferred into recipient mothers, and mtDNA haplotypes were tracked in both germline and somatic tissues across various developmental stages (Lee et al., 2012). Interestingly, mtDNA rapidly segregated into homoplasmic haplotypes during fetal development, such that most somatic tissues were over 90% homoplasmic 105 days post embryo transfer. This suggests that the somatic bottleneck acts to produce homogeneous mtDNA populations. In contrast, in germline tissues, mtDNA haplotypes were extremely

variable, with heteroplasmy levels ranging continuously from 3.7-99.2 %. This is consistent with previous research which suggests that the germline bottleneck produces large, seemingly stochastic shifts in heteroplasmy.

The majority of research on mitochondrial bottlenecks is confined to mammals, but some degree of bottlenecking appears to be widely conserved among animal lineages. Typically, the bottleneck is measured indirectly, by calculating the mitochondrial effective population size (mN_e), which is proportional to variability in heteroplasmy across generations, and inversely related to bottleneck size (Lynch & Blanchard, 1998). Estimates of mN_e suggest that the bottleneck is considerably variable between lineages, ranging from 2-10 copies in mammals to, up to several hundred copies in some insects (Solignac et al., 1984; Rand & Harrison, 1986; Ashley et al., 1989; Jenuth et al., 1996; Marchington et al., 1997; Xu et al., 2011; Wolff et al., 2011). Even within mammals, mN_e has been shown to vary by an order of magnitude (Piganeau & Eyre-Walker, 2009). These results show that the bottleneck is broadly conserved, but they also imply that the underlying mechanism may vary between lineages. In particular, the large disparity in mN_e estimates from vertebrates and invertebrates suggests the presence of distinct bottlenecking systems.

Despite its broad conservation, a convincing argument for the function of the bottleneck in mitochondrial inheritance has yet to be presented. For some time the dominant explanation has been that the bottleneck acts to expose low frequency deleterious mutations to purifying selection (Bergstrom & Pritchard, 1998). The bottleneck can cause strong founder effects, in which previously low frequency alleles suddenly spike in frequency. In this case, deleterious alleles are no longer shielded from selection by accompanying functional alleles, and are thus more likely to be eliminated. Under this theory, the bottleneck is disadvantageous to individual organisms, but is advantageous at the population level. Recently, an alternative hypothesis based on *in vivo* observation of mtDNA during oogenesis has gained favor. It has been shown that dividing mitochondria accumulate at a structure known as the Balbiani body in the oocytes of many organisms (Kloc et al., 1996, 2004). In parallel, two recent studies in *Drosophila* have suggested that intense selection on mitochondrial function occurs during oogenesis (Ma et al., 2014; Hill et al., 2014). These results lead some researchers to speculate that the Balbiani body is a selective apparatus for ensuring that functional mtDNA haplotypes are passed to offspring (Cox & Spradling, 2003; Zhou et al., 2010; Stewart & Larsson, 2014). This hypothesis is promising, but evidence from species other than *Drosophila* will be necessary before it can be applied more broadly.

Estimating the Mitochondrial Mutation Rate

The Molecular Clock

Historically, mutations have been thought to accumulate in the genome linearly across time and at a similar rate among species. Estimates of this universal mutation rate are deemed the 'molecular clock' (Zuckerkandl & Pauling, 1965) and have been used widely in molecular dating and phylogenetics (Hasegawa et al., 1985; Mercer & Roth, 2003; Weir & Schluter, 2007). The molecular clock hypothesis received substantial support from the neutral theory of molecular evolution, proposed by Motoo Kimura (1983). Under neutral theory, the majority of mutations are thought to have negligible effects on organismal function, meaning that purifying selection has minimal impact on the mutation rate. Consequently, the fixation of novel mutations is governed exclusively by genetic drift. The mutation rate is then a function of the effective population size of an organism, which is assumed to be relatively stable across evolutionary time.

Some of the core assumptions of the original molecular clock hypothesis have become increasingly controversial in recent years. The idea that mutation accumulation is linear across time has been undermined by the observation that molecular clocks systematically overestimate the dates of recent (1-2 Myr) events, in comparison to those occurring at more ancient timescales (Ho et al., 2005; Ho & Larson, 2006). The most plausible explanation for this phenomenon is that many mutations are detected in modern populations are later subjected to purifying selection, leading to a reduced set of observed mutations at ancestral timescales. This directly contradicts Kimura's neutral theory, which predicts that selection against mutations should be minimal. A second difficulty for the molecular clock hypothesis is presented by the great volume of evidence which now suggests that rates of mutation are not constant across taxa (Welch & Bromham, 2005; Lynch, 2010). Mutation rates have been found to vary across species (Baer et al., 2007; Elango et al., 2009; Halligan & Keightley, 2009), and even between different genotypes of the same species (Haag-Liautard et al., 2007; Ness et al., 2015). Additionally, evidence suggests that mutation rates are not constant across the genome, because certain features, such as highly repetitive regions are prone to errors in replication or mutation repair (Jeffreys et al., 1988; Smith et al., 2002; Supek & Lehner, 2015). It is clear that ignoring the influence of selection on genome evolution can no longer be justified. Sophisticated statistical techniques have arisen to account for these complications in

modern molecular clock estimates (see Ho & Duchne, 2014 for a review), but the assumption of minimal interspecific variability in mutation rates is still commonly utilized.

The Mutation Rate as Trait Under Selection

An alternate model of the mutation rate conceptualizes it as a trait, which reflects the individual evolutionary history of an organism. Conceptually, rates of mutation accumulation are mechanistically determined by a combination of the error-rate in DNA replication, the efficiency of mutation repair, and the presence of mutagens in the genomic environment (Baer et al., 2007). Mutations in polymerases and repair genes can cause large alterations in the mutation rate, as demonstrated in many cancers (Umar & Kunkel, 1996). Moreover, different species utilize distinct families of repair (Eisen & Hanawalt, 1999) and replication proteins (Filee et al., 2002) which possess their own characteristic error rates. Selection on the function and the relative levels of expression of these genes is presumably capable of producing a large diversity of mutation rates.

The mutation rate of an organism results from the interacting forces of selection and genetic drift. Mutations are thought to be mildly deleterious on average (Keightley & Lynch, 2003), and so selection will predominantly act to purge mutations from a population. Things become more complicated when recombination is minimized, as in mtDNA, because mutator alleles can become genetically linked to beneficial alleles on the same DNA molecule. In this case, the typical role of selection may become reversed, and instead act to increase deleterious allele frequency (Sniegowski et al., 2000). In most cases genetic drift will also favor the elimination of novel mutations, because low frequency alleles are particularly vulnerable to stochastic elimination. The power of genetic drift is most pronounced in small populations, where allele frequencies can quickly shift across generations. However, genetic drift also represents the only means by which selectively neutral and deleterious alleles can become fixed in a population. Therefore, selection and drift are often thought of as opposing forces in the production of the mutation rate (Sniegowski et al., 2000; Lynch, 2011).

Interspecific and intraspecific variability in the mutation rate can be explained by differing dynamics of selection and drift. The question of why the mutation rate has not evolved to zero, despite the selective pressure against mutations has puzzled biologists for almost a century (Sturtevant, 1937). The most intuitive explanation

is that some level of mutation is required to adapt to changing environmental conditions. Species which are unable to generate genetic variation through mutation will inevitably be driven to extinction as they are confronted with novel evolutionary challenges. It is also possible that at a certain point the associated metabolic cost of perfect fidelity in DNA replication outweighs the selective benefit of reducing mutation further (Baer et al., 2007). This hypothesis has not been widely tested, but there is convincing evidence that it at least occurs in retroviruses, in which rapid replication is of paramount importance, and so mutation repair is minimal (Furi et al., 2005; Lloyd et al., 2014). Finally, it has been argued that the selective benefit of antimutator alleles becomes insignificant when the mutation rate is already low, at which point genetic drift prevents further rate reduction (Lynch, 2011). Under each of these explanations, the relative fitness consequences of mutation and the power of genetic drift will together drive the individual mutation rate of an organism. The characteristic mutation rate of a species can therefore be viewed as an equilibrium, which is produced by the relative power of selection and drift within its evolutionary history (Kondrashov, 1995).

Dynamics of Mitochondrial Mutation

The mitochondrial mutation rate is of particular importance to molecular dating and phylogenetics, which frequently utilize mtDNA genes as markers. Many high profile molecular dating estimates, such as the African origin of humans (Maca-Meyer et al., 2001), and the first emergence of humans with *Hominidae* (Hasegawa et al., 1985) were achieved through analysis of mtDNA. Many traits of mtDNA, such as its conserved gene set, ease of amplification, and minimal recombination, make it an especially useful marker (Avise et al., 1987). However, there is little evidence that the mtDNA mutation rate is any more clock-like than that of the nucleus. One large-scale comparison of mitochondrial mutation rates found that they varied by as much as two orders of magnitude among mammals alone (Nabholz et al., 2008). Similarly, the mitochondrial bottleneck size has also been shown to vary significantly between taxa (Piganeau & Eyre-Walker, 2009)

In fact, many lines of evidence suggest that selection-drift dynamics are even more complex and unpredictable in mtDNA than the nuclear genome. The mitochondrial bottleneck presents an obvious complicating factor. The bottleneck means that mtDNA possess their own insulated population upon which genetic drift can act. In addition, mtDNA are also subject to drift at the level of the organismal population

because haplotypes can drift to extinction through elimination of the maternal line. To complicate matters further, mitochondrial bottleneck size is regulated by the nuclear genome (as evidenced by disregulation in disease; Clay Montier et al., 2009), meaning that the power of drift is partially controlled by selection on nuclear genes. Mitochondrial DNA replication and repair machinery is also encoded by the nucleus, providing a further mechanism for nuclear evolution to influence mtDNA mutation. Some data suggests that selection on mtDNA and nuclear genes may actually produce opposing pressures on mitochondrial evolution. Selective pressure on mitochondrial regulatory genes in the nuclear genome will act to increase metabolic fitness. On the other hand, selection will favor fast replication at the level of mtDNA. This competition has been demonstrated in yeast *petite* mtDNA mutants, which are non-functional but possess a replicative advantage over wild-type haplotypes (de Zamaroczy et al., 1981; Taylor et al., 2002). Much more work will be required to determine the relative importance of each of these aspects of mitochondrial inheritance on the mutation rate, but it at least seems clear that the assumption of minimal selection on mtDNA is ungrounded.

Direct Estimates of the Mutation Rate

Criticism of the molecular clock hypothesis will bear little fruit unless practical methods for molecular dating, which take mutational variability into account can be produced. Producing estimates of the mutation rate and spectrum across a broad range of taxa represents the most promising strategy for improving evolutionary models. The accuracy of dating estimates will undoubtedly be improved by using individualized mutation rate data from the species in question. Obtaining mutation rate estimates from multiple closely related species provides bounds on the range of variation between taxa. This in turn will be valuable in creating bounds on the range of mutation rate variation across time. Clearly, this will require that a large number of mutation rate estimates be produced. Fortunately, the increased availability of next-generation sequencing (NGS) technologies has significantly eased the burden of directly estimating mutation rates in recent years.

Mutation accumulation (MA) experiments are the most frequently used approach for directly estimating mutation rates (Halligan & Keightley, 2009). In MA studies, lines derived from a common ancestral genotype are propagated for many generations by single progeny descent in order to accumulate spontaneous mutations. This protocol is expected to produce a relatively unbiased view of the mutation spectrum

of an organism, with minimal interference from selection. In nuclear MA studies, all mutations except those which are lethal or cause sterility are expected to be passed to the next generation. Selection is more difficult to eliminate in mitochondrial MA experiments, due to the impossibility of preventing intracellular competition. However, the mutations detected in mitochondrial MA studies do tend to exhibit a much higher proportion of nonsynonymous mutations, in comparison to those detected by indirect phylogenetic analysis (see Thomas & Wilson, 1991 and Denver et al., 2000 for example). This indicates that the MA procedure does at least reduce the impact of selection on mtDNA.

Direct estimates of the mitochondrial mutation rate have now been produced in a small selection of model organisms, providing a preliminary view of intraspecific and interspecific variation. In total, mitochondrial MA experiments have been carried out in *Caenorhabditis elegans* (Denver et al., 2000), *Caenorhabditis briggsae* (Howe et al., 2010) *Drosophila melanogaster* (Haag-Liautard et al., 2008; Keightley et al., 2009), *Daphnia pulex* (Xu et al., 2012), *Saccharomyces cerevisiae* (Lynch et al., 2008), *Dictyostelium discoideum* (Saxer et al., 2012), *Paramecium tetraurelia* (Sung et al., 2012) and *Pristionchus pacificus* (Molnar et al., 2011). In species for which both mitochondrial and nuclear mutation data is available, the mitochondrial mutation rate has tended to be elevated in comparison to the nuclear rate by 10-100x (Figure 6). However, significant variability is noticeable even within this small sampling. For example, in *D. pulex* the ratio of the nuclear:mitochondrial base substitution rates is approximately 1:5, while in *P. tetraurelia* this ratio is 1:3587. Even within the *Caenorhabditis* genus, rates of large deletion mutations have been found to differ by 10x (Denver et al., 2000; Howe et al., 2010). These divergences provide strong support for the importance of generating species-specific mutation rate estimates, and indicate that more data is necessary to fully understand the dynamics of variation.

The Current Study

Here, we present an analysis of the mitochondrial mutation spectrum in *D. pulex*. Our data was sourced from an MA experiment published recently, which used NGS to quantify the nuclear mutation rate (Keith et al., 2015). One benefit of NGS is that reads are produced from both the mitochondrial and nuclear genomes, allowing concurrent analysis of both mutation rates with identical experimental conditions.

The unique reproductive cycle of *Daphnia* species make them an interesting target for MA studies. *D. pulex* typically exhibit a cyclical parthenogenetic life cycle, in

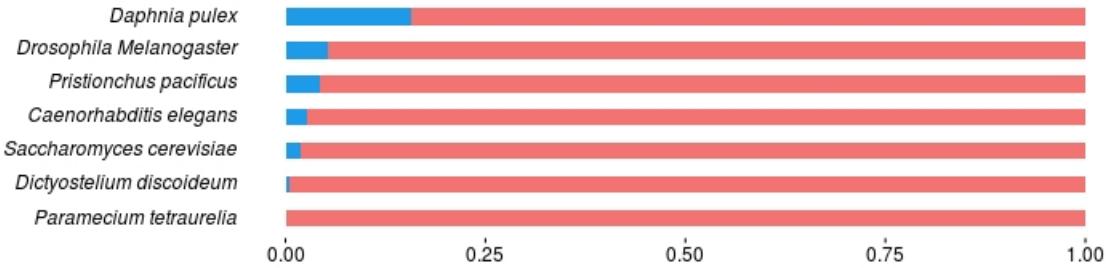


Figure 6: Relationship between the mitochondrial and nuclear base substitution rate. Bars show the ratio of direct estimates of the base substitution mutation rate of the nuclear (blue), and mitochondrial (red) genomes. Mutation rates were averaged across strains in cases where separate estimates were provided. All species for which direct mutation rate estimates have been made for both genomes are included. Data sourced from Denver et al. (2000), Lynch et al. (2008), Haag-Liautard et al. (2008), Denver et al. (2009), Keightley et al. (2009), Molnar et al. (2011), Saxer et al. (2012), Sung et al. (2012), Xu et al. (2012), Weller et al. (2014), Keith et al. (2015).

which females alternate between periods of sexual and asexual reproduction (Figure 7; Ebert, 2005). Nonetheless, *D. pulex* can be propagated entirely clonally in the lab, which makes single progeny descent trivial to implement. Interestingly, some natural *D. pulex* populations have transitioned to an obligately asexual reproductive cycle (Innes & Hebert, 1988). The progression from intermittently sexual, to entirely clonal reproduction has been hypothesized to be associated with accelerated mutation accumulation (Normark & Moran, 2000). The lack of recombination in asexual organisms is expected to produce strong linkage between the nuclear and mitochondrial genomes, which should theoretically weaken the power of selection to eliminate deleterious mutation. This effect has been validated in asexual freshwater snail (Neiman et al., 2010), and stick insect populations (Henry et al., 2012). A previous mitochondrial MA experiment in *D. pulex*, which utilized Sanger sequencing, found a non-significant marginal increase in mutation rate in obligately asexual genotypes (Xu et al., 2012). Sanger sequencing has a low sensitivity to low-frequency mutations, so it is possible that NGS will have more power to approach this question. Additionally, this provides a useful opportunity to compare mutation rate estimates produced by differing sequencing technologies, which are known to exhibit characteristic types of error (Nakamura et al., 2011).

This study will make up part of a large ongoing MA project in the *Daphnia* genus. NGS sequencing of MA lines from three natural isolates of *D. magna* sourced from Germany, Israel and Finland is expected to be completed this year. That project will make use of the results of this study, along with the previous mitochondrial (Xu

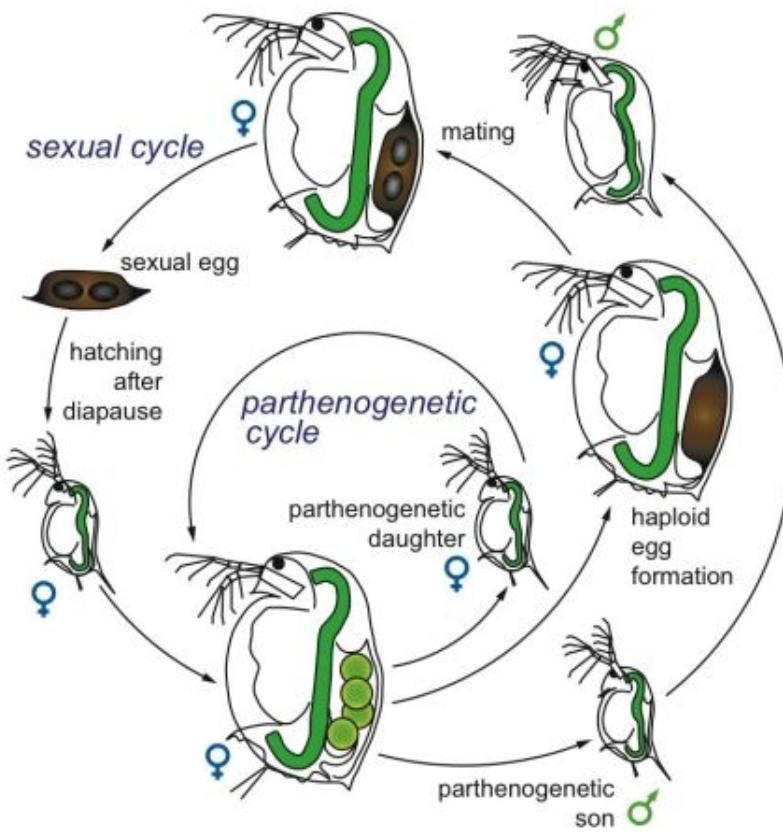


Figure 7: Cyclical parthenogenesis in natural populations of *Daphnia pulex*. Cyclical parthenogenetic *Daphnia* are capable of alternating between sexual and asexual reproduction depending upon environmental conditions. Sourced from Ebert, 2005 without modification.

et al., 2012) and nuclear (Keith et al., 2015) MA experiments in *D. pulex*. This will represent the largest volume of MA data from a single genus produced to date, and will enable many previously unavailable analyses of mutation dynamics for the first time.

Materials and Methods

The sequencing data for the present study was obtained from a recent MA experiment in *D. pulex* (Keith et al., 2015). MA line maintenance, sample preparation and sequencing was handled by the lab of M. Lynch at Indiana University, Bloomington. Subsequent sequence data processing and analysis was carried out during the course of this thesis.

Mutation Accumulation

The *D. pulex* lines consisted of two natural isolates, one obligately asexual genotype was sourced from Linwood Lake, Ontario (ASEX), and one cyclically parthenogenetic genotype from Slimy Log Pond, Oregon (CYC) . Separate isogenic populations were initiated from each founder individual by clonal reproduction. Sublines were produced from each population and propagated by single-progeny descent. Sublines were maintained for an average of 170 generations, and 85 generations for the ASEX, and CYC genotypes respectively (see Keith et al., 2015 for complete MA protocol; Table 1). Individuals from previous generations were saved in order to restore lineages in cases of extinction or experimental error.

Table 1: MA line information. Sample ID and number of generations of propagation are shown for each MA line.

Isolate	Sample ID	Generations
Linwood Lake, ON	ASEX-2	230
	ASEX-12	91
	ASEX-41	164
	ASEX-48	195
Slimy Log Pond, OR	CYC-17	90
	CYC-97	86
	CYC-237	80

Sequencing

Sequencing data was obtained from 4 randomly chosen ASEX sublines (n=2 replicates per subline), and 3 randomly chosen CYC sublines (n=1 replicates per subline). In order to extract sufficient DNA for sequencing, a single female from each chosen subline was allowed to clonally reproduce to form a population of hundreds of individuals. Whole genome 100 bp paired-end Illumina libraries were prepared, and submitted for sequencing at the Center for Genomics and Bioinformatics at Indiana University (CYC) and the Beijing Genomics Institute (ASEX).

Sequence Processing

The pipeline used for preprocessing and alignment of Illumina short reads is shown in Figure 8. See Appendix 1 for precise tools and parameters used during sequence processing. Unless specified otherwise, all read mapping was completed with Bowtie2 using the very-sensitive-local setting (Langmead & Salzberg, 2012) followed by local realignment around indels with the Genome Analysis Toolkit (GATK) IndelRealigner (McKenna et al., 2010).

Read Filtering

Before mapping, Illumina adapter sequences, and low quality bases (Phred < 30) were trimmed from reads using Trimmomatic (Bolger et al., 2014). Reads from each sample were separately mapped to the published *D. pulex* mitochondrial genome reference sequence (Accession: AF117817; Crease, 1999). PCR duplicates were removed with Picard Tools MarkDuplicates (<http://picard.sourceforge.net>). Consensus sequences were produced for each sample based on the dominant allele at each locus (allele frequency > 50 %) as observed in the preliminary alignments. Reads from each sample were then remapped to their respective consensus sequences. To mitigate the risk of sequence contamination, we excluded all reads with > 1 mismatch to the consensus sequence from further analysis (see Potential Sources of Bias and Error for a discussion of the validity of this criterion).

nuMT Removal

Nuclear mitochondrial pseudogenes (nuMTs) are sequences in the nuclear genome, which are homologous to mitochondrial genes, and derived from endosymbiotic

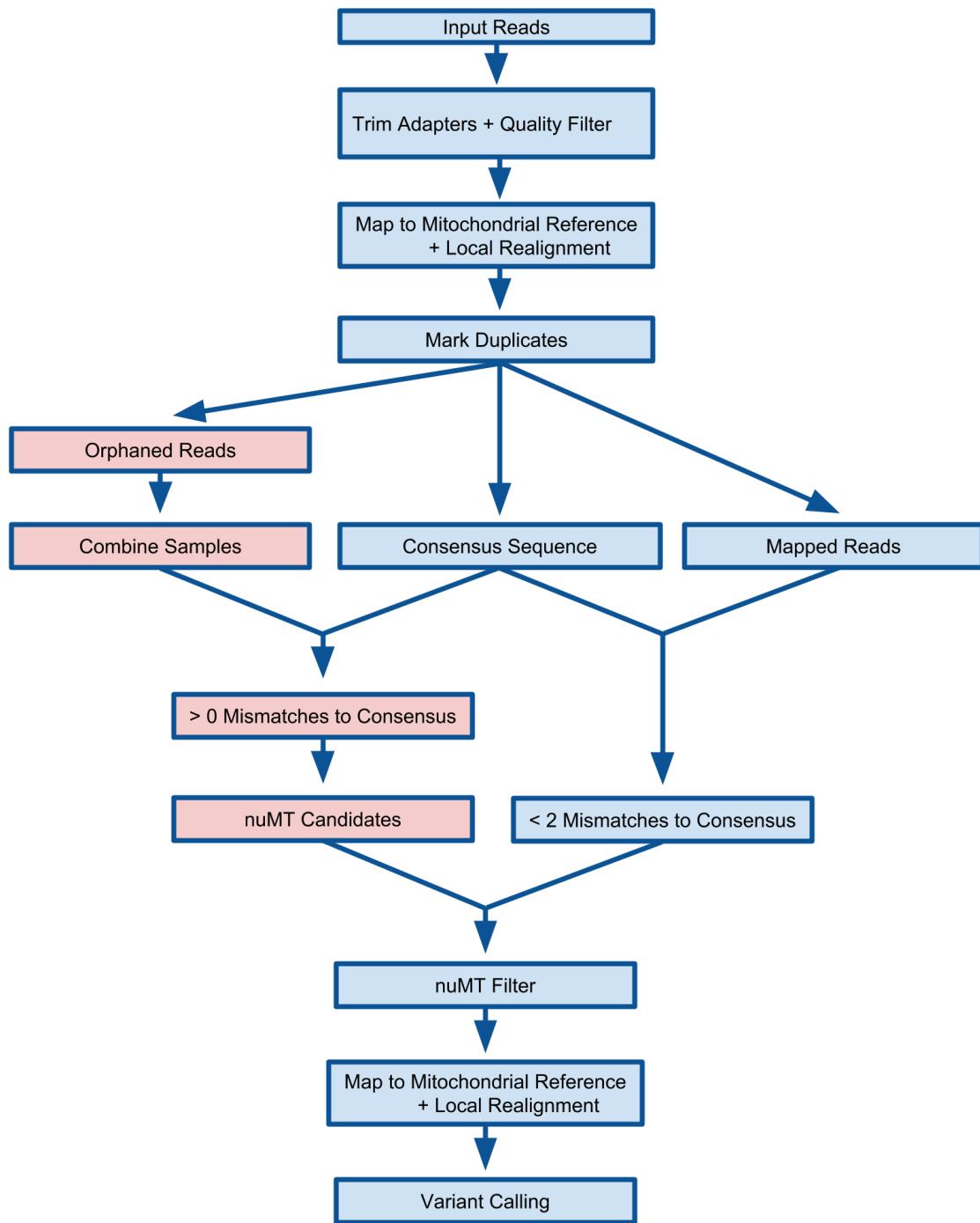


Figure 8: Sequence processing pipeline. Red boxes represent steps of the nuMT removal process. See Appendix 1 for exact utilities and commands used.

gene transfer (Bensasson et al., 2001). These sequences are capable of mapping to the mitochondrial genome and can cause errors in heteroplasmy detection when they accumulate polymorphisms (Parr et al., 2006). Reads from the small and multi-copy mitochondrial genome are generally sequenced at an average read depth of 1000 - 5000x, as compared to the 10 - 50x which is typical for the nuclear genome. This means that nuMTs are unlikely to cause errors in the detection of high frequency mutations, but become problematic for calling low-frequency heteroplasmic mutations.

Typically, nuMTs are filtered by removing all reads which map to both the mitochondrial and nuclear genome (Saxer et al., 2012; Guo et al., 2013; Ye et al., 2014). This approach has the potential to introduce substantial error into subsequent mutation calling if the nuclear genome sequence is low-quality. The rise of NGS technologies has substantially increased the pace of genome assembly, but it has done so at the cost of some accuracy. One comparison of human genome assemblies built with Sanger sequencing and NGS found that only 56 % of genes were completely represented in the NGS assembly (Alkan et al., 2011). Additionally, contaminant DNA introduced during sample preparation are routinely discovered in published genome assemblies (Percudani, 2013; Strong et al., 2014; Koutsovoulos et al., 2015). Without stringent controls, true mtDNA sequence can also become incorporated into nuclear assemblies, which would produce false positive nuMTs (Hlaing et al., 2009; Haridas et al., 2011). Oppositely, in some nuclear genome assemblies, mitochondrial reads are excluded by mapping to the mitochondrial genome, which may unintentionally cause true nuMTs to be excluded from the assembly.

Even in the case that the nuclear assembly is completely error-free, removing nuMTs by mapping to the nuclear genome has the potential to cause large biases in the allele frequency of detected variants. For example, consider the case in which a heteroplasmic mutation occurs in a region of the mitochondria, which matches a nuMT sequence in the nuclear genome. Removing all reads which perfectly map to the nuclear genome would exclude all of the mitochondrial reads with the ancestral allele, but not those with the novel allele. This would cause the frequency of the heteroplasmic allele to become substantially inflated, perhaps even to homoplasy.

To avoid these issues, we generated candidate nuMT sequences *de novo* using unpaired read alignments (orphaned reads). Typically, read pairs will map to the mitochondrial genome together, suggesting that they are true mitochondrial reads, or neither will map, suggesting that they are from the nuclear genome, or contaminant DNA. However, in cases of structural variation, such as gene

duplications, translocations or large insertions and deletions (indels), orphaned reads tend to accumulate. This observation is frequently used in structural variant detection, and has been implemented in several popular algorithms (Medvedev et al., 2010; Jiang et al., 2012; Koboldt et al., 2012; Schrder et al., 2014). A similar approach can be used for the detection of nuMTs since they are analogous to interchromosomally transferred genes.

After the preliminary mapping to the mitochondrial genome, orphaned reads were extracted and pooled across samples, to create a single set of reads from each genotype. For subsequent variant detection, the only nuMT reads which could potentially lead to false positives, are those which differ from the ancestral genotype at one or more bases. To isolate these reads, we generated a consensus sequence for each sample and mapped potential nuMT reads to each of the consensus sequences from their respective genotypes. All nuMT reads which mapped to any of these sequences with 0 mismatches were excluded from further analysis. Potential nuMTs were evidenced by regions with spikes in read depth produced by orphaned reads (see Figure 9 for example). To remove any reads outside of these spikes, which may correspond to true mitochondrial reads, we excluded all alignments below a minimum depth of 5x. This resulted in a library of nuMT sequences which could potentially be mistaken for mutations in downstream analysis. A consensus nuMT sequence was produced, and preprocessed reads from the previous section were mapped against this consensus. All reads which aligned with 0 mismatches to the nuMT consensus were excluded from further analysis.

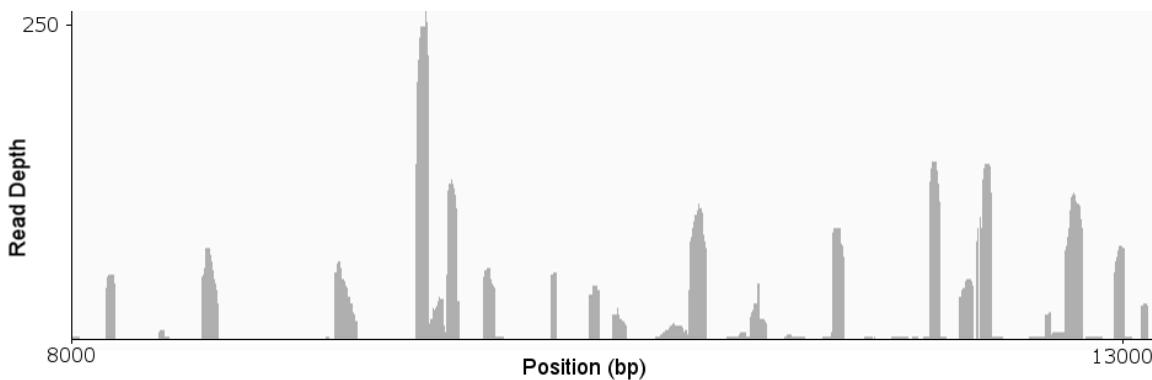


Figure 9: Example nuMT data showing spikes in read depth. Candidate nuMT reads were mapped to the mitochondrial genome. Shown is a representative region of the CYC nuMT alignment. Figure produced from Integrative Genomics Viewer (IGV) output (Robinson et al., 2011).

Variant Calling

The reads which survived sequence preprocessing were remapped to the *D. pulex* mitochondrial reference sequence for variant calling. An initial list of candidate base substitutions and indels were produced using LoFreq (Wilm et al., 2012). Unlike the majority of published variant callers, LoFreq calls mutations by modeling the sample-specific sequencing error rate, and makes no assumptions regarding ploidy. This makes it well suited for mitochondrial mutation detection, because heteroplasmic allele frequencies follow a continuous distribution, and are not clustered by ploidy as in the nuclear genome. LoFreq’s default variant filtering protocol was used, which automatically filters for strand bias (see Guo et al., 2012b for the importance of strand bias filtration). The resultant variant sets included all sites which differed from the mitochondrial reference sequence, including those which predate the MA experiment. To remove these polymorphisms, VCF files were pairwise intersected using LoFreq’s vcfset operation, which works similarly to the BedTools operation (Quinlan & Hall, 2010) but is allele-specific. In the case of the ASEX genotype, variant sets were also intersected between replicates. As a final precaution, we manually inspected each sample’s alignment to ensure that variants were truly unique, and had not simply fallen below the significance threshold in other samples. Variants were annotated with predicted gene effects using SnpEff (Cingolani et al., 2012).

Statistics

See Appendix 3 for the R scripts used for mutation rate and allele frequency calculations which are described in the following sections.

Mutation Rate Calculations

Mutation rates were calculated for each genotype following the protocol described in Haag-Liautard et al. (2008). Under the assumption that the MA protocol minimizes the impact of selection, mutation frequencies are produced entirely by genetic drift. In this case, the observed set of detected variant allele frequencies is directly proportional to the mutation rate, under neutral theory (Kimura, 1983). The mutation rate (μ) can then be calculated using the following equation:

$$\mu = \sum_i D_i / (n gl)$$

Where for each genotype, D is the set of detected mutation frequencies, n is the number of nucleotides surveyed (always 15333 in our case), g is the average number of MA generations and l is the number of MA lines. This equation was also used to calculate the mutation rate for base substitutions (μ_{bs}) and the mutation rate for indels (μ_{ind}) separately. As noted previously, the assumption of negligible selection is almost certainly not valid for mitochondrial mutation, but is necessary to estimate the mutation rate. Confidence intervals for mutation rates were produced by non-parametric bootstrapping (10,000 replicates) as described in Xu et al. (2012).

Mutation Rate Comparisons

We used a permutation approach to compare μ estimates between genotypes and to results from previous studies. Mutations were permuted between the compared variant pools using R's 'sample' function and μ was calculated for each permutation (10,000 replicates). Statistical significance was determined by whether the observed difference in μ fell outside the 95 % confidence intervals for the permuted sample distribution. To compare ratios of μ_{bs} to μ_{ind} between studies, variant pools were permuted as above, but μ_{bs}/μ_{ind} was calculated at each permutation, rather than μ .

Detection of Shifts in Allele Frequency

A freely available script (<https://github.com/genome/bam-readcount>) was used to produce allele frequency counts for each locus in every sample. The major allele was then extracted for each locus and the samples with the minimum and maximum major allele frequency at that locus were detected. A Fisher's exact test was carried out for every locus to test for a difference between the minimum and maximum allele frequency. The resultant p-values were corrected for multiple comparisons by Bonferroni correction.

Results and Discussion

Sequences were successfully mapped to the mitochondrial genome, and mutations were called for each line. Reads were initially mapped to the mitochondrial genome to an average coverage of 1325x ($sd = 263$) and 2379x ($sd = 367$) in the CYC and ASEX datasets respectively (Figure 10). After read filtering, coverage was reduced to 906x ($sd = 189$) and 1201x ($sd = 294$) respectively. The larger losses in coverage for the CYC data were caused by the input reads having lower average Phred quality scores. Coverage was generally consistent across the genome, with the notable exception of the D-loop region (14645-15333 bp), which exhibited substantially reduced read depth in both datasets. This can be attributed to the long stretches of repetitive sequence in this region, which are known to cause alignment (Ekblom et al., 2014), and sequencing difficulties (Jiang et al., 2015).

Detected Mutations

A total of 18 *de novo* mutation events were detected across the two genotypes (Table 2; Figure 11). Of these, 7 were base substitutions, 3 were insertions, and 8 were deletions. Within the base substitutions, 5 were transitions (purine \leftrightarrow purine or pyrimidine \leftrightarrow pyrimidine) and 2 were transversions (purine \leftrightarrow pyrimidine), resulting in an overall Ts/Tv ratio (Transitions/Transversions) of 2.5. No clear bias in the frequency of A/T \leftrightarrow G/C mutations were observed. We found a 1:1 ratio of non-synonymous to synonymous (dN/dS) base substitutions, and several high frequency mutations within gene-coding regions. Two particularly interesting examples were a missense mutation in the NADH dehydrogenase 2 gene (ND2; position 1000) and a base substitution in the large ribosomal RNA gene (position 12949), both of which were homoplasmic. Together with the high rate of transversions, this suggests that the effects of natural selection were minimized during MA propagation. All detected indels were 1-2 bp in length and occurred at homopolymeric tracts with runs of 5-10 identical nucleotides. Such mutations are commonly observed

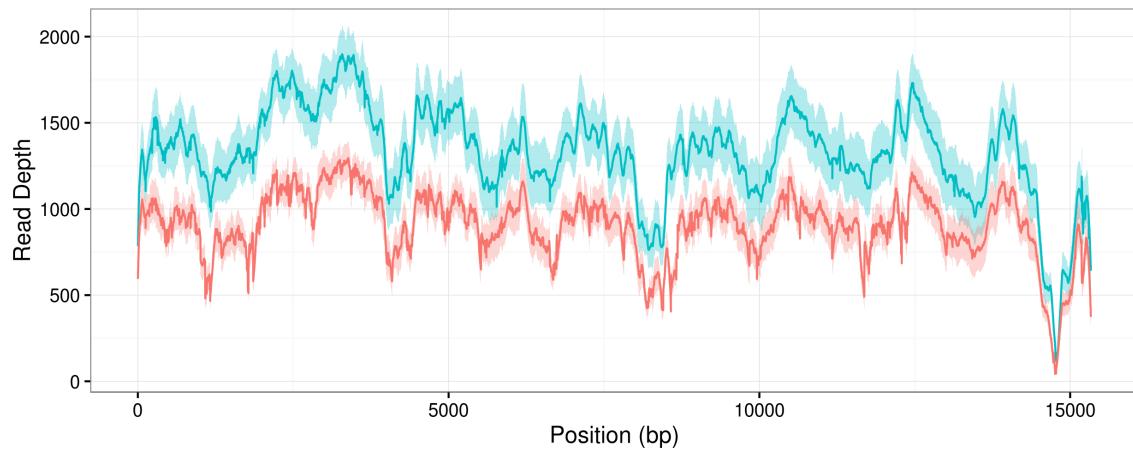
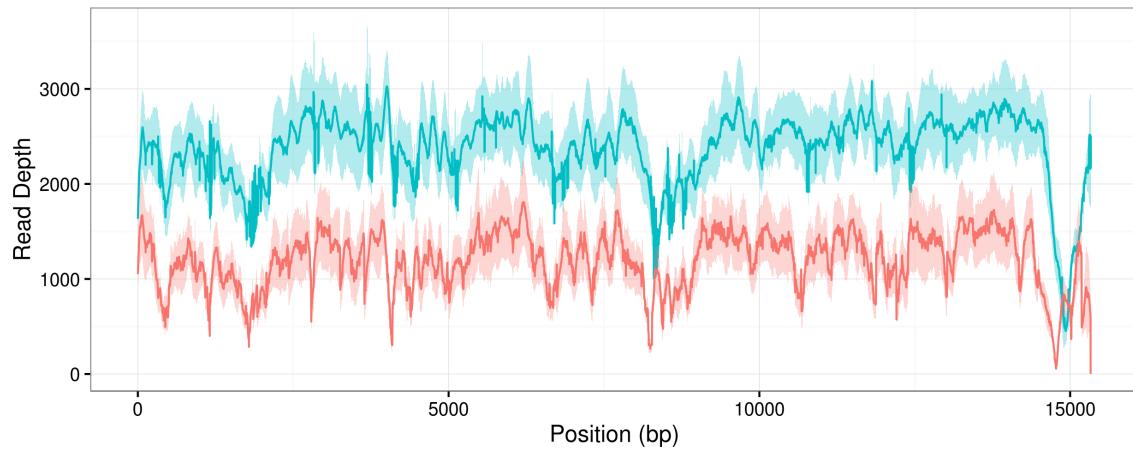
A**B**

Figure 10: Read depth across the mitochondrial genome. Read depth at each locus was averaged across samples from the ASEX (A) and CYC (B) genotypes. Blue line shows depth before read filtering, and red line shows depth post-filtering. Ribbons show standard error of the mean.

in mitochondrial MA experiments (Haag-Liautard et al., 2008; Seyfert et al., 2008; Howe et al., 2010; Xu et al., 2012), and are thought to stem from DNA polymerase slippage during replication (Viguera et al., 2001).

Variants were classified as homoplasmic if their allele frequency was >99 %, otherwise they were considered heteroplasmic. Only two of the detected variants were categorized as homoplasmic using this rule, the rest ranged in allele frequency from 1-95 % (Table 2). Furthermore, the majority of heteroplasmy were present at extremely low frequencies (1-5 %). This is consistent with a neutral model of selection, in which the majority of mutations are expected to segregate at low frequency, with

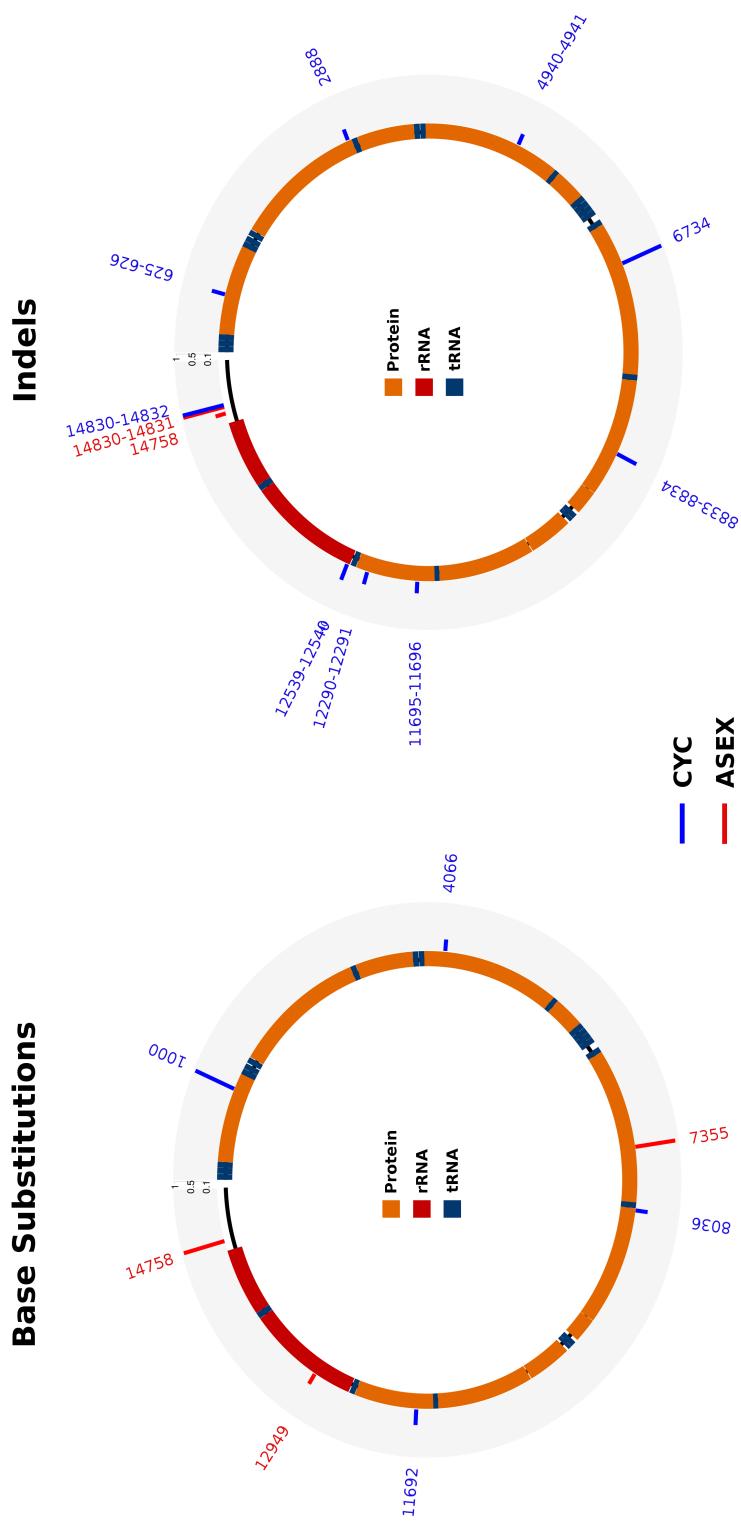


Figure 11: Locations of detected variants on the mitochondrial genome. Labels represent locations on the *D. pullex* mitochondrial reference sequence (Accession: AF117817). Tick heights represent allele frequency.

Table 2: List of detected variants. Nucleotide sequences and predicted gene effects were derived from the forward strand of the *D. pulex* mitochondrial reference sequence (Crease, 1999).

ID	Position (bp)	Type	Mutation	Freq.	Region	Effect	Context
CYC-97	625-626	Deletion	CT→C	0.01	ND2	Frameshift	AAGAATGCC T TTTTTTT
CYC-237	1000	Base Sub	T→C	0.99	ND2	Missense	GCTTTTAG C TGGGGGGG
CYC-17	2888	Insertion	T→T(C)	0.02	Cox1	Frameshift	CATTCTTT C CCCCCCCC
CYC-237	4066	Base Sub	C→T	0.02	ATP6	Missense	TGTTTGCC T TTCCCTGA
CYC-237	4940-4941	Deletion	AT→A	0.01	Cox3	Frameshift	TTTGTCTCA A TTTTTTG
CYC-17	6734	Insertion	T→T(C)	0.02	ND5	Frameshift	ATCGAAGTT C CCCCCC
ASEX-41	7355	Base Sub	G→A	0.09	ND5	Synonymous	CAAGCAGAA A AGGGAAT
CYC-237	8036	Base Sub	A→G	0.07	t-RNA		GGGATAGAG C AGGAAAA
CYC-237	8833-8834	Deletion	TA→T	0.02	ND4	Frameshift	CCCTGCCCT A AAAAAAA
CYC-17	11692	Base Sub	C→T	0.13	ND1	Synonymous	CAATCACCT I CCCAAAAA
CYC-97	11695-11696	Deletion	CA→C	0.01	ND1	Frameshift	TCACCCCCC A AAAAAAAT
CYC-237	12290-12291	Deletion	AC→A	0.01	ND1	Frameshift	GCTGAAGAACCCCCCCC
CYC-237	12539-12540	Deletion	TC→T	0.02	I-rRNA		TAAGAGCTT C CCCCTAT
ASEX-41	12949	Base Sub	T→A	1	I-rRNA		AACGCATT A AGCTTTT
ASEX-12	14758	Base Sub	T→A	0.95	D-Loop		CGAGGGGA A CCCCCCCC
ASEX-12	14758	Insertion	T→(A)C ¹	0.1	D-Loop		CGAGGGGAACCCCCCCC
ASEX-41	14830-14831	Deletion	AT→A	0.89	D-Loop		TCTTTTTA T TTTTTT
CYC-237	14830-14832	Deletion	ATT→A	0.04	D-Loop		TCTTTTTA T TTTTTT

¹ This insertion was found to occur only in reads which also showed the T→A base substitution at position 14758.

only a few drifting to fixation (Kimura, 1983). However, these results contrast with the majority of previous mitochondrial MA experiments, which primarily report mutations at frequencies above 20 % (Denver et al., 2000; Haag-Liautard et al., 2008; Howe et al., 2010; Molnar et al., 2011; Xu et al., 2012). This discrepancy is also evident between our results and those of the previous MA study in *D. pulex* (Xu et al., 2012), indicating that the disagreement stems from methodological differences, rather than the species in question. The most plausible explanation for the discrepancy is that our use of NGS, as opposed to Sanger sequencing, allowed the detection of lower frequency mutations. Outside of MA experiments, NGS has successfully been used to detect heteroplasmies at frequencies as low as 0.5 % (Guo et al., 2012a; Li et al., 2015). To date, only a handful of mitochondrial MA experiments have been published which utilize NGS. Studies in *D. melanogaster* (Keightley et al., 2009) and *P. tetraurelia* (Sung et al., 2012), utilized higher minimum allele frequency cutoffs the present study (5 % and 10 % respectively), making comparison difficult. However, an experiment in the amoeba *Dictyostelium discoideum* mirrored our results, finding that all mutations except one occurred at frequencies below 10 % (Saxer et al., 2012).

In addition to differences in low-frequency heteroplasmy detection, our results diverge from the findings of the previous mitochondrial MA study in *D.pulex* in several respects (Xu et al., 2012), suggesting that more data may be required to accurately characterize the mutation spectrum in this system (Table 3). One striking aspect of the results of that study was the observation of many recurrent mutations at two sites in the D-loop. In total, they found 6 homoplasmic indels at position 14766 and an additional 6 indels at position 14840, representing over half of all observed variants. Given the apparent hypermutability of these sites, it is surprising that we found no evidence of mutations at these positions in our data, especially given that our MA lines are derived from the same isolates as used in their study. We do however, observe several mutations within the same homopolymeric tracts (positions 14758, 14830), suggesting that these may be hypermutable tracts rather than single bases. Our estimates of indel:base substitution, and dN/dS also differ quite markedly from the previous estimates (Table 3), suggesting that a greater sample of variants may be required to converge on the true ratios.

Table 3: Mutation statistics. Comparison of mutation statistics gathered in the present study, with estimates produced by the previous mitochondrial MA experiment in *D.pulex* (Xu et al., 2012). Indel:bs = ratio of indel mutations to base substitutions, dN:dS = ratio of non-synonymous to synonymous base substitutions.

										$\times 10^{-8}$
	Reproduction	Lines	Gen.	Indel:bs	dN:dS	Ts:Tv	$\frac{GC \rightarrow AT}{AT \rightarrow GC}$	μ_{bs}	μ_{ind}	μ
Present Study	Asexual	4	170	2:3	0:1	1:2	1:0	19.6	9.5	29.1
	Cyc. Parth.	3	85	9:4	2:1	1:3	2:2	31.1	4.1	35.2
			Total	1.57:1	1:1	0.4:1	1.5:1			
Xu et al. (2012)	Asexual	37	100	8:3	0:1	2:1	1:1	4.3	13	17.3
	Cyc. Parth.	82	61	9:3	3:0	1:2	2:1	2	11.7	13.7
			Total	2.83:1	3:1	1:1	1.5:1			

Mutation Rate Estimates

For the CYC genotype, μ_{bs} was estimated to be 31.1×10^{-8} and μ_{ind} was estimated to be 4.1×10^{-8} resulting in a combined μ estimate of 35.2×10^{-8} (CI = 12.2 - 45.9 x 10^{-8} ; Table 3). In the ASEX genotype, μ_{bs} was estimated to be 19.6×10^{-8} and μ_{ind} was estimated to be 9.5×10^{-8} , leading to a combined μ of 29.1×10^{-8} (CI = 0 - 64.1

$\times 10^{-8}$; Table 3). No difference in μ , was observed between genotypes ($p=0.99$), as determined by permutation of mutation rate estimates (see Materials and Methods). Thus, in agreement with the results of Xu et al. (2012), we find no evidence for an effect of reproductive strategy on mutation accumulation in *D. pulex*. Given the extreme similarity of the calculated mutation rates, variants from the ASEX and CYC genotypes were combined for further analysis. The mutation rate estimates across all lines were as follows: $\mu_{bs} = 22.7 \times 10^{-8}$, $\mu_{ind} = 8.0 \times 10^{-8}$ and $\mu = 32.3 \times 10^{-8}$ (CI = 6.5 - 52.9 $\times 10^{-8}$). Using the same permutation technique as previously, no statistical difference was found between μ_{bs} and μ_{ind} across all lines ($p=0.22$).

The mitochondrial mutation rates calculated here are the highest of all direct estimates to date. Previous estimates of μ have ranged from 0.68×10^{-8} (*D. discoideum*) to approximately 16×10^{-8} (*D. pulex* and *C. elegans*; Denver et al., 2000; Xu et al., 2012; Saxer et al., 2012). Given that the previous estimate in *D. pulex* was already elevated in comparison to other species, it is perhaps not surprising that our estimate is extreme. Additionally, the difference in mutation rates between the present study and the previous was not significant ($p > 0.999$). Nonetheless, it is possible that we are simply lacking in statistical power to robustly test for differences between our estimates. One possible cause of the slightly elevated μ in our study is the inclusion of heteroplasmic variants with lower allele frequency. To test this possibility, we recalculated μ using sets of variants with different minimum allele frequency cutoffs (Figure 12). Our μ estimates remained high under all tested cutoffs, suggesting that low frequency heteroplasmy were not responsible for the disparity. We conclude that the slight disagreement in μ estimates is due to sampling error, and that both estimates are consistent in reporting an elevated mutation rate in *D. pulex*.

However, the relationship between μ_{bs} and μ_{ind} reported here is inconsistent with the previous estimate. The ratio of $\mu_{bs}:\mu_{ind}$ in our results is 2.83:1, while in Xu et al. (2012), the ratio is 0.28:1. Permutation testing revealed that this difference in proportions is statistically significant ($p < 0.001$). Again, we suggest that this discrepancy can be partially attributed to differences in sequencing technology. The error profile of NGS have been the subject of considerable examination in recent years (Nakamura et al., 2011; Wang et al., 2012; Wall et al., 2014; Schirmer et al., 2015). Reports have typically found an elevated error rate, and decreasing sensitivity to indels in NGS compared to Sanger sequencing. Our estimate of μ_{ind} is slightly decreased in our results compared to Xu et al.'s (2012), suggesting that some true indels may have been missed by our analysis. However, this difference is insufficient to explain the complete reversal in μ_{bs} and μ_{ind} which we observe here. Much less

attention has been focused upon rates of error in Sanger sequencing, but false-positives are generally considered to be significantly decreased in comparison to NGS (Shendure & Ji, 2008). Some data suggests however, that rates of false-negatives may be a larger concern. In one recent screening for clinically relevant mutations in Dravet syndrome, NGS detected 28 true mutations which were missed by Sanger sequencing, whereas only 1 mutation was missed by NGS but detected by Sanger sequencing (Djmi et al., 2016). The majority of the false-negatives (19/28) were attributed to human-error, frequently due to mutations being missed when examining Sanger sequencing traces by eye. Indels are easier to detect than base-substitutions in trace files, because indels cause frameshifts which effect all proceeding peaks. We estimate that 24 homoplasmic base-substitutions would need to have been missed in the previous study to produce the μ_{bs} estimate which we report. This would represent an astonishing rate of error, and so we suspect that bias from Sanger sequencing alone is also not sufficient to resolve this disagreement. One final possibility is that an extremely high number of false-positive base-substitutions are present in our data. This could be resolved by validating these mutations with Sanger sequencing. Regardless of the exact cause of the disagreement, it does suggest that considerable caution should be taken when comparing the results of MA studies which utilize different sequencing methodologies.

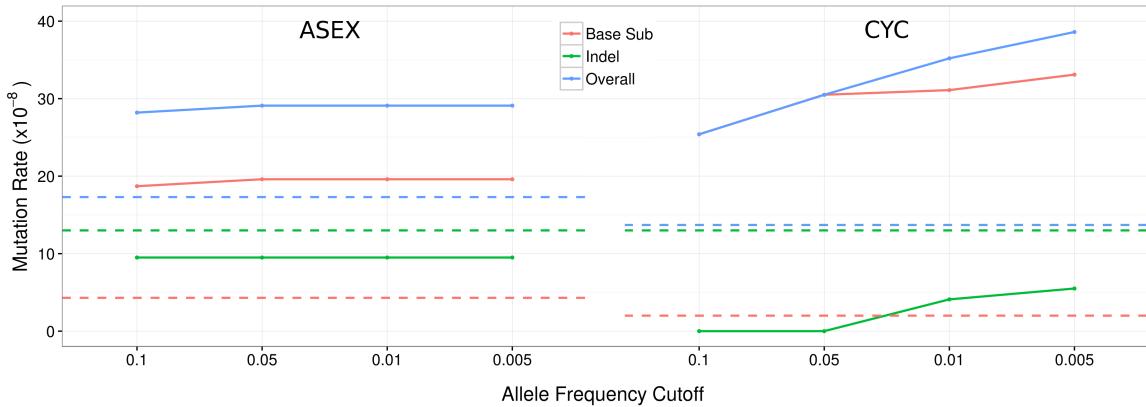


Figure 12: Effect of allele frequency cutoff on mutation rate estimates. Points and solid lines represent mutation rates calculated in the present study. Dashed lines represent mutation rate estimates from Xu et al. (2012). Mutation rates were calculated separately using different sets of variants depending upon allele frequency. For the 0.005 cutoff, variants which were flagged by LoFreq as statistically significant, but which were removed by minimum allele frequency filtering, were included.

Minor Allele Frequencies

Despite the similarity in μ between genotypes, many more low frequency heteroplasmic mutations were found in the CYC genotype than in ASEX. This observation was validated by examining the distributions of minor allele frequencies (MAF) in each genotype (Figure 13). The MAF distribution from the CYC data was markedly broader and its peak was right-shifted in comparison to the ASEX distribution. A Kolmogorov-Smirnov test confirmed that the distributions were distinct ($p < 0.001$). This effect was also present in the unfiltered read data ($p < 0.001$), indicating that it was not an artifact produced by sequence processing. This disparity could conceivably be produced either by a substantial difference in the rate of sequencing error, or by a true difference in the underlying mutation dynamics. DNA extraction and sequencing was carried out separately for each genotype, so it is plausible that the sample quality was not equivalent. On the other hand, the sequence quality cutoffs which we used were exceedingly strict (Phred > 30) and the average read quality was almost identical after filtering. Additionally, LoFreq's variant calling model is extremely resilient to sequencing quality (Wilm et al., 2012; Huang et al., 2015), so sequencing error is unlikely to have caused an enrichment of low-frequency variants in its call set. Nevertheless, it is impossible to statistically differentiate between sequencing error and low frequency mutation with complete confidence, and so validation of low-frequency variants with Sanger sequencing would be beneficial.

Shifts in Allele Frequency

We detected two heteroplasmic alleles which had shifted in allele frequency in one or more of the lines during the course of the MA experiment (Figure 14). In the ASEX genotype, we discovered that a single base deletion (position 14831) which was present at frequencies in the range of 0-4.4% in three of the samples had shifted to a frequency of 90 % in ASEX-41 ($p < 0.001$, after Bonferroni correction; Figure 14A). In the CYC genotype, a G→T heteroplasmy (position 5488) had shifted from the range of 8.1-8.7 %, up to 11.8 % in CYC-97 ($p = 0.02$; Figure 14B). A relatively large (9.2 %) shift in the frequency of a deletion (position 14759) was also observed in the CYC sample, but was non-significant after Bonferroni correction ($p = 0.09$). These results highlight the advantage NGS in MA studies, since these shifts may have been mis-characterized as *de novo* mutations with Sanger sequencing due to insufficient sensitivity.

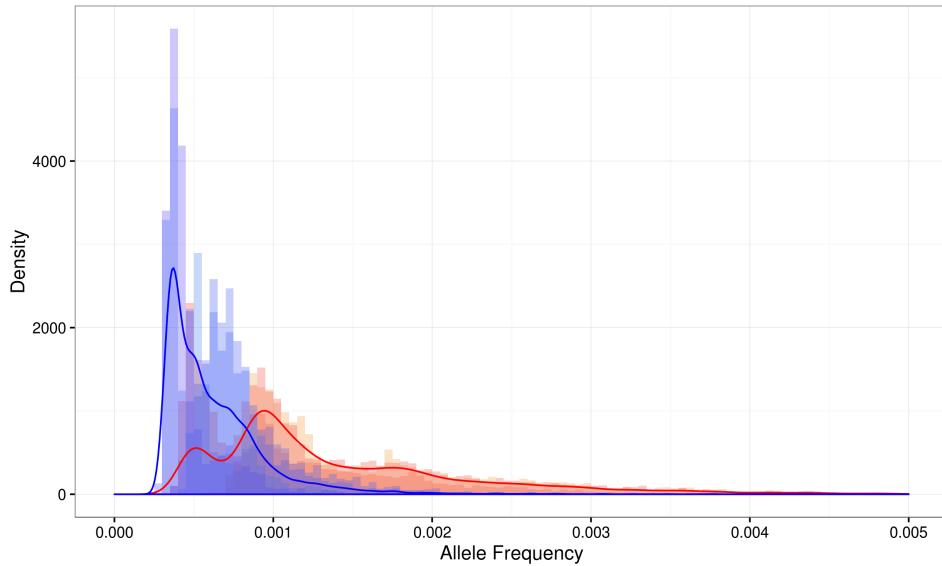


Figure 13: Minor allele frequency distributions. Distributions of non-zero minor allele frequencies are shown for the CYC (red), and ASEX (blue) genotypes. Histograms show single sample data. Lines show kernel density distribution for each genotype.

We also observed a number of heteroplasmies with stable allele frequencies across samples, particularly in the CYC genotype. These are hard to explain within the context of the extremely narrow bottleneck ($mN_e = 5-10$) which has been estimated for *D. pulex* (Xu et al., 2012). Under this bottleneck size we should expect all heteroplasmic alleles to become fixed within only a few generations, and the chances of maintaining allele frequency in multiple lines over the course of an MA study are extremely slim. One possibility is that these heteroplasmies are caused by recurrent somatic mutations at hypermutable loci. This hypothesis is supported by the fact that 14/19 of the stable heteroplasmies were low frequency indels at homopolymeric tracts. Weak selection against these indels might prevent their growing in frequency beyond a certain threshold, potentially explaining their stability.

nuMT Analysis

A library of candidate nuMTs was produced from orphaned reads which mapped to the mitochondrial genome. Alignment of candidate nuMTs to the mitochondrial genome produced regions with large abrupt increases in read depth, as would be expected for displaced sequences (Figure 9). As an initial validation of the procedure, we tested whether the pool of detected nuMTs were similar between the CYC and

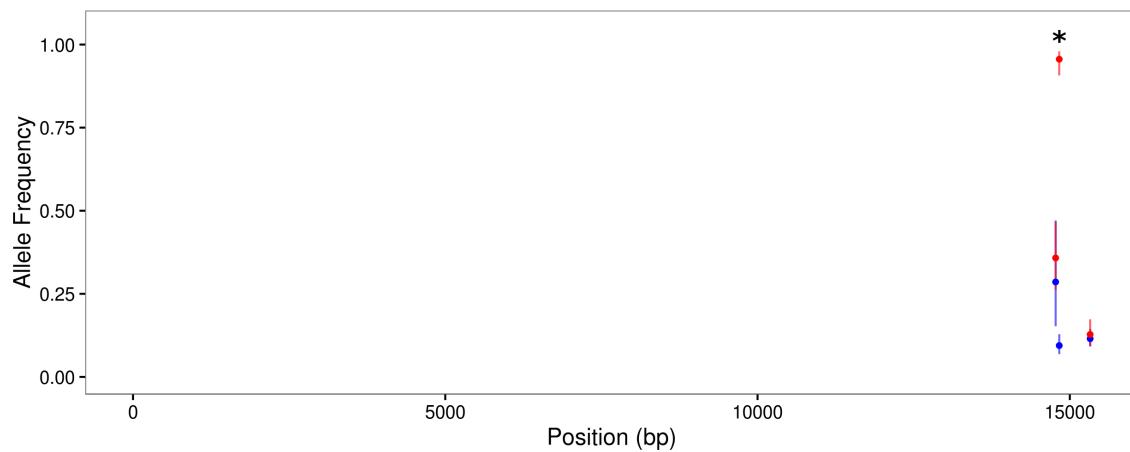
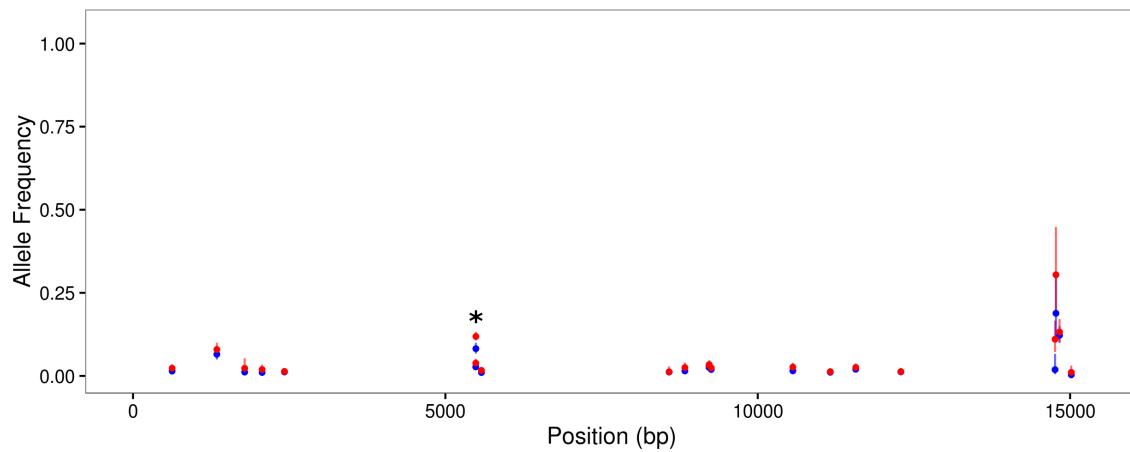
A**B**

Figure 14: Shifts in heteroplasmic allele frequency. Points represent the minimum (blue) and maximum (red) sample allele frequencies at a given locus. Lines show 95 % confidence intervals. Asterisks indicate statistically significant shifts in allele frequency. (A) ASEX genotype. (B) CYC genotype.

ASEX genotypes. If they are true nuMTs we expect them to be conserved across genotypes, since the majority are presumably the result of ancient endosymbiotic gene transfer. We generated BED files with coverage information from the nuMT alignments to the mitochondrial genome for each genotype, and then tested whether the overlap in coverage was greater than we would expect by chance, using BEDTools Fisher utility (Quinlan & Hall, 2010). We found a statistically significant correlation between the nuMT libraries between genotypes ($p < 0.001$). One worrying possibility is that these high coverage regions could be produced by indel-associated alignment

errors, rather than nuMTs. To test this hypothesis, we repeated the same overlap analysis using BED files generated from indel locations in our variant sets. We found no association between the location of indels and the locations of candidate nuMTs ($p > 0.999$ for both genotypes). Genotypic consistency alone is insufficient to be confident that we have captured true nuMTs, since many other types of structural variation could potentially cause correlated density in unpaired mappings. As a further validation, we mapped the pool of nuMTs to the *D. pulex* nuclear reference sequence (Accession: gca_000187875; Colbourne et al., 2011) using NCBI's BLAST utility (Altschul et al., 1990). If the entire mitochondrial genome sequence is mapped to the nuclear genome, a total of 60 confident hits are produced. In comparison, the ASEX nuMT library produced 50 hits, and the CYC library produced 69, indicating that we were successful in recapitulating the nuMT population (Figure 15). The fact that the CYC library had more hits to the nuclear genome than the entire mitochondrial genome, suggests that this approach may be sensitive to ancestral nuMTs which have become significantly diverged from the modern mtDNA sequence. The nuMT libraries also contained many sequences which did not map to the nuclear genome, which may represent nuMTs which are absent from the nuclear genome assembly, or other sources of contamination.

A total of 18 variants were removed after mapping against the libraries of potential nuMTs. All of these variants were also flagged either by our mismatch filter or our allele frequency filter (see Appendix 2 for list of filtered variants). Thus the nuMT removal protocol did not affect our final set of variant calls, but would have had a large impact on calls in the absence of other hard filters. This is consistent with the expectation that nuMT reads are likely to be present at lower depth, and be more diverged from the reference sequence, than mitochondrial reads. Interestingly all of the flagged variants were clustered between approximately 10,000-14,500 bp, which our BLAST results suggest is the site of two long continuous stretches of nuMT sequence (Figure 15). This indicates that some caution should be taken in interpreting low-frequency variants within this region in *Daphnia*. Unfortunately this area also contains much of the hypermutable D-loop region, which is of considerable importance to estimates of mitochondrial mutation. Candidate nuMT reads in this region tended to contain clusters of mismatches to the mitochondrial reference, and lead to a large number of high confidence variant calls in the absence of additional filtering. This emphasizes the continued importance of sensible hard filtering criteria, even with the rise of statistically sophisticated variant calling algorithms.

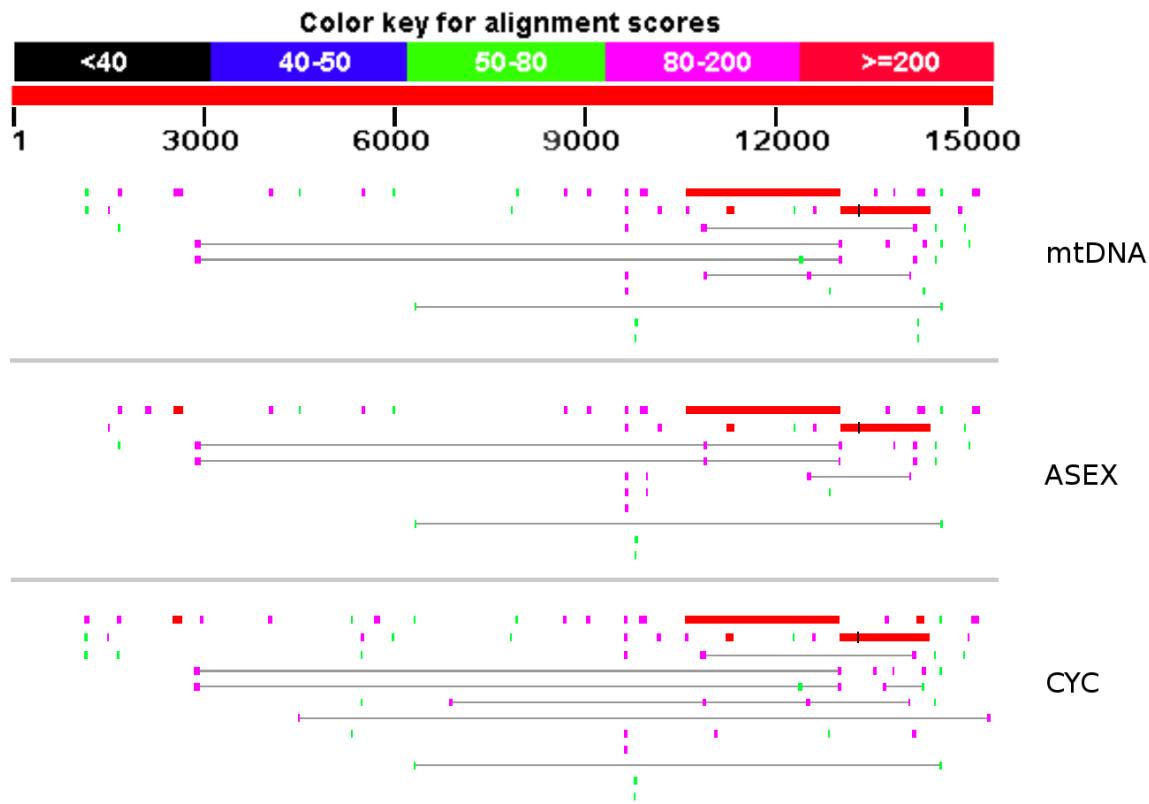


Figure 15: BLAST results for detected nuMTs. Pools of detected nuMTs were mapped to the *D. pulex* nuclear reference assembly (Accession: gca_000187875) with NCBI's BLAST utility (Altschul et al., 1990). Top panel (mtDNA) shows the result of a BLAST search of the entire mtDNA reference sequence (Accession: AF117817). Bottom two panels show the result of BLAST search using detected nuMT pools for each genotype.

Potential Sources of Bias and Error

While using hard filtering criteria can be beneficial for eliminating false positives, it can also lead to bias if the cutoffs are misjudged. If our data contains a large undetected source of error, then even more stringent mismatch and allele frequency filters might have been necessary. On the other hand, overly strict criteria may cause erroneous rejection of true positive variants. LoFreq has been shown to be capable of detecting variants at frequencies below 1 % (Wilm et al., 2012; Huang et al., 2015), and it predicted many variants at frequencies between 0.4 - 1 % with high confidence in our dataset (see Appendix). We chose to reject these variants primarily in order to mitigate the possibility of undetected nuMTs influencing our results. However, if we overestimated the danger of nuMTs, this may have caused us to reject many true variants. Filtering out reads with multiple mismatches could also represent a

considerable source of error. It is not unusual for mutations to be clustered together at highly variable sites (Schrider et al., 2011), and so some true positives were likely eliminated by this filter.

On balance, we erred on the side of the caution with our filtering criteria, meaning that our mutation rate estimates are likely to be downwardly biased. In general, the results of MA experiments should be interpreted as lower bounds on the true rate of mutation in a species or genotype. Somatic mosaicism means that mitochondrial mutations can theoretically exist at proportions as low as one molecule in the total population of mtDNA in an organism. Consequently, regardless of the sensitivity of variant detection, missing true variants is inevitable. If the mutation rate was calculated with the assumption that all MAFs represented true mutations, our estimates jumped into the range of 10^{-6} mutations per nucleotide per generation. Presumably the true rate of mutation falls somewhere between these two bounds, but exact estimation may be impossible within an MA framework.

Future Directions

The results of this thesis will provide a preliminary framework, and a useful comparison, for larger scale analysis of mutation in the *Daphnia* genus. The primary aim of that project will be the quantification of μ in the nuclear genome, but mtDNA reads will also be captured during sequencing. The methodology for analysis of that mitochondrial mutation data will be informed by the sequencing processing and variant calling pipeline produced in the present study. The substantial number of mutations present in our data suggests that a rich set of mitochondrial mutations will be detected in the *D. magna* dataset. This will allow a complex set of comparisons of the mutation dynamics between species, lines, genetic backgrounds and environments which have not been possible in previous studies. The present study will aid in the additional useful comparison of mutation patterns across closely related species of *Daphnia*, and thus further our understanding of the mutational process in animals more generally.

Appendix

Appendix 1: Pipeline Commands

Trim adapters: Trimmomatic (Bolger et al., 2014)

```
trimmomatic PE -threads 48 -phred64 <fwd.paired.fq>
<rev.paired.fq> <fwd.paired.trim.fq> <fwd.unpaired.trim>
<rev.paired.trim> <rev.unpaired.trim.fq> ILLUMINACLIP:
<adapter_file.fa>:2:30:10:5
```

Mapping: Bowtie2 (Langmead & Salzberg, 2012)

```
bowtie2 --very-sensitive-local -x <bowtie_index> -p "46" "-1"
<fwd.fq> "-2" <rev.fq> -S <output.sam>
```

Mark Duplicates: Picard (<http://picard.sourceforge.net>)

```
picard MarkDuplicates INPUT=<in.bam> OUTPUT=<out.bam>
```

Local Realignment Around Indels: GATK (McKenna et al., 2010)

```
gatk -T IndelRealigner -R <reference.fa> -I <in.bam>
-targetIntervals <intervals_file> -o <out.bam>
```

Build Consensus Sequence: GATK

```
gatk -T FastaAlternateReferenceMaker -R <ref.fa> -o
<consensus.fa> --variant
```

Quality Trimming: Trimmomatic

```
trimmomatic PE -threads 48 -phred64 <fwd.paired.fq>
<rev.paired.fq> <fwd.paired.trim.fq> <fwd.unpaired.trim>
<rev.paired.trim> <rev.unpaired.trim.fq> SLIDINGWINDOW:4:30
```

Mismatch Filtering: Bamtools (Barnett et al., 2011)

```
bamtools filter -in <in.bam> -out <out.bam> -script
.../filter_json/filter.json
```

```
filter.json:  
{  
  "tag" : "NM:<2",  
  "isMapped" : "true"  
}
```

Mutation Calling: LoFreq (Wilm et al., 2012)

```
lofreq indelqual --dindel -f <reference.fa> -o <out.bam> <in.bam>  
lofreq call-parallel --pp-threads 46 -f <consensus.fa> -o <out.vcf>  
--call-indels
```

Appendix 2: Filtered Mutations

Table 4: Table of filtered variants. AF: Variant fell below minimum allele frequency of 0.01. MM: Variant was no longer called if reads were removed with > 1 mismatch to the consensus sequence. nuMT: Variant matched a sequence from the nuMT pool. See Materials and Methods for filtering details.

ID	Position (bp)	Reference Allele	Novel Allele	Filter	Allele Frequency
ASEX-12	7583	G	A	AF	0.007199
CYC-17	1076	AG	A	AF	0.005291
CYC-17	3855	T	C	AF	0.003685
CYC-17	5239	CT	C	AF	0.006107
CYC-17	10967	T	C	AF	0.00431
CYC-17	11115	AT	A	AF	0.003774
CYC-17	11160	TG	T	AF	0.006186
CYC-17	12364	GA	G	AF	0.002738
CYC-17	13562	CA	C	AF	0.004286
CYC-17	14275	A	G	AF	0.004018
CYC-237	3361	T	C	AF	0.006272
CYC-237	4226	AT	A	AF	0.007018
CYC-237	6214	T	C	AF	0.00504
CYC-237	6284	TA	T	AF	0.007098
CYC-237	11719	A	G	AF	0.005725
CYC-237	13255	T	C	AF	0.007235
CYC-237	13814	TA	T	AF	0.004866
CYC-97	994	T	C	AF	0.007203
CYC-97	2312	T	C	AF	0.00729
CYC-97	3724	A	G	AF	0.006012
CYC-97	5655	T	C	AF	0.00689
CYC-97	5980	AT	A	AF	0.004433
CYC-97	7069	AG	A	AF	0.004435
CYC-97	9317	A	G	AF	0.005307
CYC-97	11229	A	G	AF	0.006151
CYC-97	11354	CT	C	AF	0.00572
CYC-97	12144	TA	T	AF	0.009198
CYC-97	13008	T	C	AF	0.009677
CYC-97	13952	A	G	AF	0.006261
CYC-97	14335	GT	G	AF	0.0053
ASEX-12	12569	A	C	MM	0.006021
ASEX-12	12577	T	C	MM	0.006021
ASEX-12	12799	T	C	MM	0.006084

Continued on next page...

ID	Position (bp)	Reference Allele	Novel Allele	Filter	Allele Frequency
ASEX-12	12838	T	C	MM	0.00758
ASEX-12	12871	T	C	MM	0.007718
ASEX-12	12909	A	G	MM	0.008984
ASEX-12	13354	T	G	MM	0.007233
ASEX-12	14039	A	G	MM	0.010101
ASEX-12	14257	C	A	MM	0.008505
ASEX-2	11266	T	G	MM	0.010817
ASEX-41	11160	T	TG	MM	0.010645
ASEX-41	14068	T	A	MM	0.011212
ASEX-41	14107	C	T	MM	0.011494
ASEX-48	6226	CT	C	MM	0.004577
ASEX-48	10978	A	G	MM	0.009281
ASEX-48	12183	G	A	MM	0.013052
ASEX-48	12187	C	T	MM	0.011905
ASEX-48	12869	C	T	MM	0.013906
ASEX-48	13946	G	A	MM	0.007526
ASEX-48	14008	G	A	MM	0.009825
ASEX-48	14302	C	T	MM	0.019444
ASEX-48	14534	CA	C	MM	0.013405
CYC-17	1943	A	T	MM	0.02618
CYC-17	3445	A	C	MM	0.018974
CYC-17	3648	C	A	MM	0.015636
CYC-17	5488	G	T	MM	0.027624
CYC-17	6452	A	C	MM	0.017248
CYC-17	6687	A	C	MM	0.020536
CYC-17	7343	T	G	MM	0.01874
CYC-17	7413	T	G	MM	0.009881
CYC-17	8078	A	C	MM	0.014574
CYC-17	8703	A	C	MM	0.021674
CYC-17	9731	ACCCAAAGCT	A	MM	0.008557
CYC-17	10882	T	A	MM	0.007739
CYC-17	10894	C	T	MM	0.007273
CYC-17	10897	C	T	MM	0.006877
CYC-17	10903	G	A	MM	0.006905
CYC-17	10912	A	C	MM	0.007249
CYC-17	11081	C	T	MM	0.009059
CYC-17	11083	G	A	MM	0.00873
CYC-17	11086	T	C	MM	0.009141
CYC-17	11095	A	G	MM	0.011272

Continued on next page...

ID	Position (bp)	Reference Allele	Novel Allele	Filter	Allele Frequency
CYC-17	11670	G	A	MM	0.005891
CYC-17	11800	C	T	MM	0.006556
CYC-17	11956	T	C	MM	0.00717
CYC-17	11998	G	A	MM	0.005922
CYC-17	12238	G	A	MM	0.008028
CYC-17	12241	A	G	MM	0.008766
CYC-17	12342	A	G	MM	0.006472
CYC-17	12369	A	C	MM	0.014388
CYC-17	12385	A	C	MM	0.010593
CYC-237	7506	TC	T	MM	0.002892
CYC-237	10940	T	C	MM	0.007496
CYC-237	10951	C	T	MM	0.008403
CYC-237	10978	A	G	MM	0.014264
CYC-237	10981	T	C	MM	0.012668
CYC-237	11658	C	T	MM	0.008892
CYC-237	11673	CA	C	MM	0.007366
CYC-237	11910	A	C	MM	0.006793
CYC-237	11964	A	G	MM	0.006689
CYC-237	12157	A	G	MM	0.007082
CYC-237	14092	CT	C	MM	0.003108
CYC-97	11725	A	G	MM	0.007968
CYC-97	11967	G	A	MM	0.008621
ASEX-12	10762	T	C	nuMT	0.008596
ASEX-12	13792	G	A	nuMT	0.00681
ASEX-41	10849	G	A	nuMT	0.01
ASEX-41	11066	A	G	nuMT	0.009217
ASEX-41	11071	A	G	nuMT	0.009881
ASEX-41	11074	C	T	nuMT	0.009186
ASEX-41	11081	C	T	nuMT	0.00791
ASEX-41	11083	G	A	nuMT	0.007895
ASEX-41	12776	C	T	nuMT	0.006559
ASEX-41	13854	G	T	nuMT	0.00732
ASEX-48	11255	T	C	nuMT	0.015484
ASEX-48	14521	G	A	nuMT	0.015837
CYC-17	10630	C	T	nuMT	0.005514
CYC-17	10672	A	G	nuMT	0.006159
CYC-17	10885	C	T	nuMT	0.006858
CYC-17	10889	T	C	nuMT	0.009948
CYC-17	13118	A	G	nuMT	0.007965
CYC-237	12931	T	C	nuMT	0.009517

Appendix 3: Analysis Scripts

All scripts are written in R (R Core Team, 2014).

Mutation Rate Calculations

```

library(boot)
library(perm)

# Simple mutation rate calculation.
mut_rate <- function(muts, gen, nuc) {
  summed = sum(muts)
  summed/(gen*nuc)
}

# Indexed version of mutation rate calculation for bootstrapping confidence intervals
mut_rate_indexed <- function(data, indices, gen, nuc) {
  d = data[indices]
  summed = sum(data[indices])
  return((summed)/(gen*nuc))
}

# Bootstrap mutation rate estimate (10000 replicates)
bootstrap_CI <- function(data, gen=gen, nuc) {
  bt = boot(data, statistic=mut_rate_indexed, R=10000, gen=gen, nuc=nuc)
  boot.ci(bt)$basic
}

# Compare mutation rates using permutation
permute_diffs <- function(data1, data2, gen1, gen2, nuc1, nuc2) {
  grouped.dat = data.frame(af=c(data1, data2), group=as.numeric(c(rep(1, each=length(data1)), rep(2, each=length(data2)))))
  
  # Actual difference in mutation rate estimates
  t0 = mut_rate(grouped.dat[which(grouped.dat$group==1), "af"], nuc=nuc1, gen=gen1) - mut_rate(grouped.dat[which(grouped.dat$group==2), "af"], nuc=nuc2, gen=gen2)
  
  b5 = numeric(10000)
  for (i in 1:10000) {
    # Permute group labels
    grouped.dat$group = sample(grouped.dat$group, length(grouped.dat$group))
    # Recalculate Mutation Rate
    b5[i] = mut_rate(grouped.dat[which(grouped.dat$group==1), "af"], nuc=nuc1, gen=gen1) - mut_rate(grouped.dat[which(grouped.dat$group==2), "af"], nuc=nuc2, gen=gen2)
  }
  mean(abs(b5) > abs(t0))
}

# Compare indel:bs mutation rate ratios by permuting variants
permute_bs_ind <- function(data1, data2, type1, type2, gen1, gen2, nuc1, nuc2) {
  grouped.dat = data.frame(af=c(data1, data2), type=c(type1, type2), group=as.numeric(c(rep(1, each=length(data1)), rep(2, each=length(data2)))))
  
  # Compute base substitution mutation rates
  bs1 = mut_rate(grouped.dat[which((grouped.dat$group==1) & (grouped.dat$type=="bs")), "af"], nuc=nuc1, gen=gen1)
  bs2 = mut_rate(grouped.dat[which((grouped.dat$group==2) & (grouped.dat$type=="bs")), "af"], nuc=nuc2, gen=gen2)
  
  # Compute indel mutation rates
  ind1 = mut_rate(grouped.dat[which((grouped.dat$group==1) & (grouped.dat$type=="ind")), "af"], nuc=nuc1, gen=gen1)
  ind2 = mut_rate(grouped.dat[which((grouped.dat$group==2) & (grouped.dat$type=="ind")), "af"], nuc=nuc2, gen=gen2)
  
  # Calculate observed difference in indel:bs
  t0 = (bs1/ind1) - (bs2/ind2)
  tp = numeric(10000)
  
  for (i in 1:10000) {
    grouped.dat$group = sample(grouped.dat$group, length(grouped.dat$group))
    bs1 = mut_rate(grouped.dat[which((grouped.dat$group==1) & (grouped.dat$type=="bs")), "af"], nuc=nuc1, gen=gen1)
    bs2 = mut_rate(grouped.dat[which((grouped.dat$group==2) & (grouped.dat$type=="bs")), "af"], nuc=nuc2, gen=gen2)
    ind1 = mut_rate(grouped.dat[which((grouped.dat$group==1) & (grouped.dat$type=="ind")), "af"], nuc=nuc1, gen=gen1)
    ind2 = mut_rate(grouped.dat[which((grouped.dat$group==2) & (grouped.dat$type=="ind")), "af"], nuc=nuc2, gen=gen2)
    tp[i] = (bs1/ind1) - (bs2/ind2)
  }
  mean(abs(tp) > abs(t0))
}

```

Allele Frequency Shift Significance Testing

```

# Determine the frequency of the major allele at every locus and test for significant differences between
genotypes
# in_list: a list of output files from bam-readcount (https://github.com/genome/bam-readcount) converted to
# tables.

prop_test_across_genome <- function(in_list) {

  #Output data frame, containing locus position (pos), minimum and maximum frequency between genotypes (minfreq, maxfreq), p-value for difference (p), and confidence intervals(conf1,2)
  pvals = data.frame(pos=numeric(0), minfreq=numeric(0), maxfreq=numeric(0), p=numeric(0), conf1=numeric(0),
  , conf2=numeric(0))

  # For every locus
  for (i in 1:nrow(in_list[[1]])) {
    maj_alls = c()
    # Determine major allele for each genotype
    for (j in 1:length(in_list)) {
      A = as.numeric(strsplit(in_list[[j]][i, "A"], ":"[[1]][2]))
      C = as.numeric(strsplit(in_list[[j]][i, "C"], ":"[[1]][2]))
      G = as.numeric(strsplit(in_list[[j]][i, "G"], ":"[[1]][2]))
      T = as.numeric(strsplit(in_list[[j]][i, "T"], ":"[[1]][2]))
      max_id = which.max(c(A,C,G,T))
      if (max_id==1) {
        maj_alls[j] = "A"
      } else if (max_id==2) {
        maj_alls[j] = "C"
      } else if (max_id==3) {
        maj_alls[j] = "G"
      } else if (max_id==4) {
        maj_alls[j] = "T"
      }
    }
    # Major allele = the consensus between genotypes
    ma = names(sort(table(maj_alls),decreasing=TRUE)[1])

    maj_all_freqs = c()
    depths = c()

    # Determine major allele freq for each genotype
    for (k in 1:length(in_list)) {
      maj_all_freqs[k] = as.numeric(strsplit(in_list[[k]][i,ma], ":"[[1]][2]))
      depths[k] = as.numeric(in_list[[k]][i, "depth"])
    }

    het = which(maj_all_freqs/depths < 0.99)
    maj_all_freqs = maj_all_freqs[het]
    depths = depths[het]

    # Significance testing
    if (length(maj_all_freqs) > 1 ) {
      max_index = which.max(maj_all_freqs/depths)
      min_index = which.min(maj_all_freqs/depths)
      Max = maj_all_freqs[max_index]
      Min = maj_all_freqs[min_index]
      Maxfail = depths[max_index] - Max
      Minfail = depths[min_index] - Min
      prop_matrix = matrix(c(Max, Min, Maxfail, Minfail), ncol=2)
      pvals[i, "minfreq"] = Min/depths[min_index]
      pvals[i, "maxfreq"] = Max/depths[max_index]
      pvals[i, "p"] = fisher.test(prop_matrix, simulate.p.value=TRUE)$p.value
      pvals[i, "conf1"] = fisher.test(prop_matrix, simulate.p.value=TRUE, conf.int=TRUE)$conf.int[1]
      pvals[i, "conf2"] = fisher.test(prop_matrix, simulate.p.value=TRUE, conf.int=TRUE)$conf.int[2]
    }
    pvals[i,"pos"] = in_list[[j]][i,"pos"]
    padj = pvals[complete.cases(pvals),]
    padj$p = p.adjust(padj$p,method="bonferroni")
  }
  return(padj)
}

```


Works Cited

- Alkan, C., Sajjadian, S., & Eichler, E. E. (2011). Limitations of next-generation genome sequence assembly. *Nature Methods*, 8(1), 61–65.
- Allen, J. F. (1996). Separate Sexes and the Mitochondrial Theory of Ageing. *Journal of Theoretical Biology*, 180(2), 135–140.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.
- Archibald, J. M. (2015). Endosymbiosis and Eukaryotic Cell Evolution. *Current Biology*, 25(19), 911–921.
- Arnoult, D., Soares, F., Tattoli, I., & Girardin, S. E. (2011). Mitochondria in innate immunity. *EMBO Reports*, 12(9), 901–910.
- Ashley, M. V., Laipis, P. J., & Hauswirth, W. W. (1989). Rapid segregation of heteroplasmic bovine mitochondria. *Nucleic Acids Research*, 17(18), 7325–7331.
- Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., & Saunders, N. C. (1987). Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics*, 18(1), 489–522.
- Baer, C. F., Miyamoto, M. M., & Denver, D. R. (2007). Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics*, 8(8), 619–631.
- Baldauf, S. L. (2003). The Deep Roots of Eukaryotes. *Science*, 300(5626), 1703–1706.
- Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strmberg, M. P., & Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, 27(12), 1691–1692.

- Barr, C. M., Neiman, M., & Taylor, D. R. (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *The New Phytologist*, 168(1), 39–50.
- Bensasson, D., Zhang, D.-X., Hartl, D. L., & Hewitt, G. M. (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*, 16(6), 314–321.
- Bergstrom, C. T., & Pritchard, J. (1998). Germline Bottlenecks and the Evolutionary Maintenance of Mitochondrial Genomes. *Genetics*, 149(4), 2135–2146.
- Birky, W. (2001). The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models. *Annual Review of Genetics*, 35(1), 125–148.
- Birky, W., & Strausberg, R. (1977). Vegetative segregation of mitochondria in yeast: Estimating parameters using a random model. *Molecular and General Genetics*, 158(3), 251–261.
- Bogenhagen, D. F. (2012). Mitochondrial DNA nucleoid structure. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819(910), 914–920.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.
- Boxma, B., de Graaf, R. M., van der Staay, G. W. M., van Alen, T. A., Ricard, G., Gabaldn, T., van Hoek, A. H. A. M., Moon-van der Staay, S. Y., Koopman, W. J. H., van Hellemond, J. J., Tielens, A. G. M., Friedrich, T., Veenhuis, M., Huynen, M. A., & Hackstein, J. H. P. (2005). An anaerobic mitochondrion that produces hydrogen. *Nature*, 434(7029), 74–79.
- Broers, C. A., Stumm, C. K., Vogels, G. D., & Brugerolle, G. (1990). Psalteriomonas lanterna gen. nov., sp. nov., a free-living amoeboflagellate isolated from freshwater anaerobic sediments. *European Journal of Protistology*, 25(4), 369–380.
- Brown, T. A., Tkachuk, A. N., Shtengel, G., Kopek, B. G., Bogenhagen, D. F., Hess, H. F., & Clayton, D. A. (2011). Superresolution Fluorescence Imaging of Mitochondrial Nucleoids Reveals Their Spatial Range, Limits, and Membrane Interaction . *Molecular and Cellular Biology*, 31(24), 4994–5010.
- Bryant, M. P., Wolin, E. A., Wolin, M. J., & Wolfe, R. S. (1967). Methanobacillus omelianskii, a symbiotic association of two species of bacteria. *Archiv für Mikrobiologie*, 59(1-3), 20–31.

- Cao, L., Shitara, H., Horii, T., Nagao, Y., Imai, H., Abe, K., Hara, T., Hayashi, J.-I., & Yonekawa, H. (2007). The mitochondrial bottleneck occurs without reduction of mtDNA content in female mouse germ cells. *Nature Genetics*, 39(3), 386–390.
- Cao, L., Shitara, H., Sugimoto, M., Hayashi, J.-I., Abe, K., & Yonekawa, H. (2009). New evidence confirms that the mitochondrial bottleneck is generated without reduction of mitochondrial DNA content in early primordial germ cells of mice. *PLoS genetics*, 5(12), e1000756.
- Cavalier-Smith, T. (1987). Eukaryotes with no mitochondria. *Nature*, 326(6111), 332–333.
- Cavalier-Smith, T. (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *International Journal of Systematic and Evolutionary Microbiology*, 52(2), 297–354.
- Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The Effect of Deleterious Mutations on Neutral Molecular Variation. *Genetics*, 134(4), 1289–1303.
- Ciborowski, K. L., Consuegra, S., Leniz, C. G. d., Beaumont, M. A., Wang, J., & Jordan, W. C. (2007). Rare and fleeting: an example of interspecific recombination in animal mitochondrial DNA. *Biology Letters*, 3(5), 554–557.
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80–92.
- Clay Montier, L. L., Deng, J. J., & Bai, Y. (2009). Number matters: control of mammalian mitochondrial DNA copy number. *Journal of Genetics and Genomics*, 36(3), 125–131.
- Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., Tokishita, S., Aerts, A., Arnold, G. J., Basu, M. K., Bauer, D. J., Cceres, C. E., Carmel, L., Casola, C., Choi, J.-H., Detter, J. C., Dong, Q., Dusheyko, S., Eads, B. D., Frhlich, T., Geiler-Samerotte, K. A., Gerlach, D., Hatcher, P., Jogdeo, S., Krijgveld, J., Kriventseva, E. V., Klitz, D., Laforsch, C., Lindquist, E., Lopez, J., Manak, J. R., Muller, J., Pangilinan, J., Patwardhan, R. P., Pitluck, S., Pritham, E. J., Rechtsteiner, A., Rho, M., Rogozin, I. B., Sakarya, O., Salamov, A.,

- Schaack, S., Shapiro, H., Shiga, Y., Skalitzky, C., Smith, Z., Souvorov, A., Sung, W., Tang, Z., Tsuchiya, D., Tu, H., Vos, H., Wang, M., Wolf, Y. I., Yamagata, H., Yamada, T., Ye, Y., Shaw, J. R., Andrews, J., Crease, T. J., Tang, H., Lucas, S. M., Robertson, H. M., Bork, P., Koonin, E. V., Zdobnov, E. M., Grigoriev, I. V., Lynch, M., & Boore, J. L. (2011). The Ecoresponsive Genome of Daphnia pulex. *Science*, 331(6017), 555–561.
- Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R., & Embley, T. M. (2008). The archaeabacterial origin of eukaryotes. *Proceedings of the National Academy of Sciences*, 105(51), 20356–20361.
- Cox, R. T., & Spradling, A. C. (2003). A Balbiani body and the fusome mediate mitochondrial inheritance during *Drosophila* oogenesis. *Development (Cambridge, England)*, 130(8), 1579–1590.
- Crease, T. J. (1999). The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene*, 233(12), 89–99.
- Cree, L. M., Samuels, D. C., de Sousa Lopes, S. C., Rajasimha, H. K., Wonnapiñij, P., Mann, J. R., Dahl, H.-H. M., & Chinnery, P. F. (2008). A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nature Genetics*, 40(2), 249–254.
- de Zamaroczy, M., Marotta, R., Faugeron-Fonty, G., Goursot, R., Mangin, M., Baldacci, G., & Bernardi, G. (1981). The origins of replication of the yeast mitochondrial genome and the phenomenon of suppressivity. *Nature*, 292(5818), 75–78.
- Denver, D. R., Dolan, P. C., Wilhelm, L. J., Sung, W., Lucas-Lled, J. I., Howe, D. K., Lewis, S. C., Okamoto, K., Thomas, W. K., Lynch, M., & Baer, C. F. (2009). A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proceedings of the National Academy of Sciences*, 106(38), 16310–16314.
- Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L., & Thomas, W. K. (2000). High Direct Estimate of the Mutation Rate in the Mitochondrial Genome of *Caenorhabditis elegans*. *Science*, 289(5488), 2342–2344.
- Djmi, T., Weckhuysen, S., von Spiczak, S., Carvill, G. L., Jaehn, J., Anttonen, A.-K., Brilstra, E., Caglayan, H. S., de Kovel, C. G., Depienne, C., Gaily, E., Gennaro, E., Giraldez, B. G., Gormley, P., Guerrero-Lpez, R., Guerrini, R., Hmlinen,

- E., Hartmann, C., Hernandez-Hernandez, L., Hjalgrim, H., Koeleman, B. P. C., Leguern, E., Lehesjoki, A.-E., Lemke, J. R., Leu, C., Marini, C., McMahon, J. M., Mei, D., Mller, R. S., Muhle, H., Myers, C. T., Nava, C., Serratosa, J. M., Sisodiya, S. M., Stephani, U., Striano, P., van Kempen, M. J. A., Verbeek, N. E., Usluer, S., Zara, F., Palotie, A., Mefford, H. C., Scheffer, I. E., De Jonghe, P., Helbig, I., Suls, A., & EuroEPINOMICS-RES Dravet working group (2016). Pitfalls in genetic testing: the story of missed SCN1a mutations. *Molecular Genetics & Genomic Medicine*, (pp. n/a–n/a).
- Dokianakis, E., & Ladoukakis, E. D. (2014). Different degree of paternal mtDNA leakage between male and female progeny in interspecific *Drosophila* crosses. *Ecology and Evolution*, 4(13), 2633–2641.
- Doolittle, W. F., & Brown, J. R. (1994). Tempo, mode, the progenote, and the universal root. *Proceedings of the National Academy of Sciences*, 91(15), 6721–6728.
- Ebert, D. (2005). *Ecology, Epidemiology, and Evolution of Parasitism in Daphnia*. National Center for Biotechnology Information (US).
- Eisen, J. A., & Hanawalt, P. C. (1999). A phylogenomic study of DNA repair genes, proteins, and processes. *Mutation Research*, 435(3), 171–213.
- Ekblom, R., Smeds, L., & Ellegren, H. (2014). Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. *BMC Genomics*, 15, 467.
- El-Hattab, A. W., & Scaglia, F. (2013). Mitochondrial DNA depletion syndromes: review and updates of genetic basis, manifestations, and therapeutic options. *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, 10(2), 186–198.
- Elango, N., Lee, J., Peng, Z., Loh, Y.-H. E., & Yi, S. V. (2009). Evolutionary rate variation in Old World monkeys. *Biology Letters*, 5(3), 405–408.
- Embley, M., Giezen, M. v. d., Horner, D. S., Dyal, P. L., & Foster, P. (2003). Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 358(1429), 191–203.

- Embley, T. M., & Martin, W. (2006). Eukaryotic evolution, changes and challenges. *Nature*, 440(7084), 623–630.
- Filee, J., Forterre, P., Sen-Lin, T., & Laurent, J. (2002). Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. *Journal of Molecular Evolution*, 54(6), 763–773.
- Finlay, B. J., Embley, T. M., & Fenchel, T. (1993). A new polymorphic methanogen, closely related to *Methanocorpusculum parvum*, living in stable symbiosis within the anaerobic ciliate *Trimyema* sp. *Journal of General Microbiology*, 139(2), 371–378.
- Fuke, S., Kubota-Sakashita, M., Kasahara, T., Shigeyoshi, Y., & Kato, T. (2011). Regional variation in mitochondrial DNA copy number in mouse brain. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1807(3), 270–274.
- Furi, V., Moya, A., & Sanjun, R. (2005). The cost of replication fidelity in an RNA virus. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 10233–10237.
- Galluzzi, L., Kepp, O., & Kroemer, G. (2012). Mitochondria: master regulators of danger signalling. *Nature Reviews Molecular Cell Biology*, 13(12), 780–788.
- Gantenbein, B., Fet, V., Gantenbein-Ritter, I. A., & Balloux, F. (2005). Evidence for recombination in scorpion mitochondrial DNA (Scorpiones: Buthidae). *Proceedings of the Royal Society B: Biological Sciences*, 272(1564), 697–704.
- Gilkerson, R. W., Selker, J. M. L., & Capaldi, R. A. (2003). The cristal membrane of mitochondria is the principal site of oxidative phosphorylation. *FEBS letters*, 546(2-3), 355–358.
- Gomes, L. C., Benedetto, G. D., & Scorrano, L. (2011). During autophagy mitochondria elongate, are spared from degradation and sustain cell viability. *Nature Cell Biology*, 13(5), 589–598.
- Guo, X., Liu, S., & Liu, Y. (2006). Evidence for Recombination of Mitochondrial DNA in Triploid Crucian Carp. *Genetics*, 172(3), 1745–1749.
- Guo, Y., Cai, Q., Samuels, D. C., Ye, F., Long, J., Li, C.-I., Winther, J. F., Tawn, E. J., Stovall, M., Lhteenmki, P., Malila, N., Levy, S., Shaffer, C., Shyr, Y., Shu, X.-O., & Boice, J. D. (2012a). The use of next generation sequencing technology

- to study the effect of radiation therapy on mitochondrial DNA mutation. *Mutation Research*, 744(2), 154–160.
- Guo, Y., Li, C.-I., Sheng, Q., Winther, J. F., Cai, Q., Boice, J. D., & Shyr, Y. (2013). Very Low-Level Heteroplasmy mtDNA Variations Are Inherited in Humans. *Journal of Genetics and Genomics*, 40(12), 607–615.
- Guo, Y., Li, J., Li, C.-I., Long, J., Samuels, D. C., & Shyr, Y. (2012b). The effect of strand bias in Illumina short-read sequencing data. *BMC genomics*, 13, 666.
- Guy, L., & Ettema, T. J. G. (2011). The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends in Microbiology*, 19(12), 580–587.
- Guy, L., Saw, J. H., & Ettema, T. J. G. (2014). The archaeal legacy of eukaryotes: a phylogenomic perspective. *Cold Spring Harbor Perspectives in Biology*, 6(10), a016022.
- Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B., & Keightley, P. D. (2008). Direct Estimation of the Mitochondrial DNA Mutation Rate in *Drosophila melanogaster*. *PLoS Biology*, 6(8).
- Haag-Liautard, C., Dorris, M., Maside, X., Macaskill, S., Halligan, D. L., Charlesworth, B., & Keightley, P. D. (2007). Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature*, 445(7123), 82–85.
- Hackstein, J. H. P., Akhmanova, A., Boxma, B., Harhangi, H. R., & Voncken, F. G. J. (1999). Hydrogenosomes: eukaryotic adaptations to anaerobic environments. *Trends in Microbiology*, 7(11), 441–447.
- Hagstrom, E., Freyer, C., Battersby, B. J., Stewart, J. B., & Larsson, N.-G. (2014). No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline. *Nucleic Acids Research*, 42(2), 1111–1116.
- Halligan, D. L., & Keightley, P. D. (2009). Spontaneous Mutation Accumulation Studies in Evolutionary Genetics. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 151–172.
- Haridas, S., Breuill, C., Bohlmann, J., & Hsiang, T. (2011). A biologist's guide to de novo genome assembly using next-generation sequence data: A test with fungal genomes. *Journal of Microbiological Methods*, 86(3), 368–375.

- Hasegawa, M., Kishino, H., & Yano, T.-a. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2), 160–174.
- Hastings, I. M. (1992). Population genetic aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. *Genetics Research*, 59(03), 215–225.
- Hengartner, M. O., Ellis, R., & Horvitz, R. (1992). *Caenorhabditis elegans* gene ced-9 protects cells from programmed cell death. *Nature*, 356(6369), 494–499.
- Henry, L., Schwander, T., & Crespi, B. J. (2012). deleterious Mutation Accumulation in Asexual Timema Stick Insects. *Molecular Biology and Evolution*, 29(1), 401–408.
- Hill, J. H., Chen, Z., & Xu, H. (2014). Selective propagation of functional mitochondrial DNA during oogenesis restricts the transmission of a deleterious mitochondrial variant. *Nature Genetics*, 46(4), 389–392.
- Hlaing, T., Tun-Lin, W., Somboon, P., Socheat, D., Setha, T., Min, S., Chang, M. S., & Walton, C. (2009). Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. *BMC genetics*, 10, 11.
- Ho, S. Y. W., & Duchne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, 23(24), 5947–5965.
- Ho, S. Y. W., & Larson, G. (2006). Molecular clocks: when timesare a-changin'. *Trends in Genetics*, 22(2), 79–83.
- Ho, S. Y. W., Phillips, M. J., Cooper, A., & Drummond, A. J. (2005). Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times. *Molecular Biology and Evolution*, 22(7), 1561–1568.
- Hollenbeck, P. J., & Saxton, W. M. (2005). The axonal transport of mitochondria. *Journal of cell science*, 118(Pt 23), 5411–5419.
- Howe, D. K., Baer, C. F., & Denver, D. R. (2010). High Rate of Large Deletions in *Caenorhabditis briggsae* Mitochondrial Genome Mutation Processes. *Genome Biology and Evolution*, 2, 29–38.
- Huang, H. W., Mullikin, J. C., & Hansen, N. F. (2015). Evaluation of variant detection software for pooled next-generation sequence data. *BMC Bioinformatics*, 16(1).

- Iborra, F. J., Kimura, H., & Cook, P. R. (2004). The functional organization of mitochondrial genomes in human cells. *BMC Biology*, 2, 9.
- Innes, D. J., & Hebert, P. D. N. (1988). The Origin and Genetic Basis of Obligate Parthenogenesis in *Daphnia pulex*. *Evolution*, 42(5), 1024–1035.
- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., & Miyata, T. (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of the National Academy of Sciences of the United States of America*, 86(23), 9355–9359.
- Jakobs, S., Schauss, A. C., & Hell, S. W. (2003). Photoconversion of matrix targeted GFP enables analysis of continuity and intermixing of the mitochondrial lumen. *FEBS Letters*, 554(1-2), 194–200.
- Jeffreys, A. J., Royle, N. J., Wilson, V., & Wong, Z. (1988). Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature*, 332(6161), 278–281.
- Jenuth, J. P., Peterson, A. C., Fu, K., & Shoubridge, E. A. (1996). Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nature Genetics*, 14(2), 146–151.
- Jiang, Y., Turinsky, A. L., & Brudno, M. (2015). The missing indels: an estimate of indel variation in a human genome and analysis of factors that impede detection. *Nucleic Acids Research*, (p. gkv677).
- Jiang, Y., Wang, Y., & Brudno, M. (2012). PRISM: Pair read informed split read mapping for base-pair level detection of insertion, deletion and structural variants. *Bioinformatics*, (p. bts484).
- Kaneda, H., Hayashi, J., Takahama, S., Taya, C., Lindahl, K. F., & Yonekawa, H. (1995). Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 92(10), 4542–4546.
- Keightley, P. D., & Lynch, M. (2003). Toward a Realistic Model of Mutations Affecting Fitness. *Evolution*, 57(3), 683–685.
- Keightley, P. D., Trivedi, U., Thomson, M., Oliver, F., Kumar, S., & Blaxter, M. L. (2009). Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Research*, 19(7), 1195–1201.

- Keith, N., Tucker, A. E., Jackson, C. E., Sung, W., Lled, J. I. L., Schrider, D. R., Schaack, S., Dudycha, J. L., Ackerman, M., Younge, A. J., Shaw, J. R., & Lynch, M. (2015). High mutational rates of large-scale duplication and deletion in *Daphnia pulex*. *Genome Research*.
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- Kloc, M., Bilinski, S., & Etkin, L. D. (2004). The Balbiani body and germ cell determinants: 150 years later. *Current Topics in Developmental Biology*, 59, 1–36.
- Kloc, M., Larabell, C., & Etkin, L. D. (1996). Elaboration of the messenger transport organizer pathway for localization of RNA to the vegetal cortex of *Xenopus* oocytes. *Developmental Biology*, 180(1), 119–130.
- Koboldt, D. C., Larson, D. E., Chen, K., Ding, L., & Wilson, R. K. (2012). Massively Parallel Sequencing Approaches for Characterization of Structural Variation. *Methods in molecular biology (Clifton, N.J.)*, 838, 369–384.
- Kondrashov, A. S. (1995). Modifiers of mutation-selection balance: general approach and the evolution of mutation rates. *Genetical Research*, 66(01), 53.
- Koutsovoulos, G., Kumar, S., Laetsch, D. R., Stevens, L., Daub, J., Conlon, C., Maroon, H., Thomas, F., Aboobaker, A., & Blaxter, M. (2015). The genome of the tardigrade *Hypsibius dujardini*. *bioRxiv*, (p. 033464).
- Kraytsberg, Y., Schwartz, M., Brown, T. A., Ebralidse, K., Kunz, W. S., Clayton, D. A., Vissing, J., & Khrapko, K. (2004). Recombination of Human Mitochondrial DNA. *Science*, 304(5673), 981–981.
- Kukat, C., Wurm, C. A., Sphr, H., Falkenberg, M., Larsson, N.-G., & Jakobs, S. (2011). Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *Proceedings of the National Academy of Sciences*, 108(33), 13534–13539.
- Kvist, L., Martens, J., Nazarenko, A. A., & Orell, M. (2003). Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Molecular Biology and Evolution*, 20(2), 243–247.
- Ladoukakis, E. D., & Zouros, E. (2001). Recombination in Animal Mitochondrial DNA: Evidence from Published Sequences. *Molecular Biology and Evolution*, 18(11), 2127–2131.

- Lake, J. A., Henderson, E., Oakes, M., & Clark, M. W. (1984). Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proceedings of the National Academy of Sciences*, 81(12), 3786–3790.
- Lane, N., & Martin, W. (2010). The energetics of genome complexity. *Nature*, 467(7318), 929–934.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359.
- Lee, H.-S., Ma, H., Juanes, R. C., Tachibana, M., Sparman, M., Woodward, J., Ramsey, C., Xu, J., Kang, E.-J., Amato, P., Mair, G., Steinborn, R., & Mitalipov, S. (2012). Rapid mitochondrial DNA segregation in primate preimplantation embryos precedes somatic and germline bottleneck. *Cell reports*, 1(5), 506–515.
- Legros, F., Malka, F., Frachon, P., Lombs, A., & Rojo, M. (2004). Organization and dynamics of human mitochondrial DNA. *Journal of Cell Science*, 117(13), 2653–2662.
- Li, M., Schrder, R., Ni, S., Madea, B., & Stoneking, M. (2015). Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proceedings of the National Academy of Sciences*, 112(8), 2491–2496.
- Lill, R., & Kispal, G. (2000). Maturation of cellular FeS proteins: an essential function of mitochondria. *Trends in Biochemical Sciences*, 25(8), 352–356.
- Lindmark, D. G., & Mller, M. (1973). Hydrogenosome, a Cytoplasmic Organelle of the Anaerobic Flagellate *Tritrichomonas foetus*, and Its Role in Pyruvate Metabolism. *Journal of Biological Chemistry*, 248(22), 7724–7728.
- Lloyd, S. B., Kent, S. J., & Winnall, W. R. (2014). The high cost of fidelity. *AIDS research and human retroviruses*, 30(1), 8–16.
- Logan, D. C. (2006). The mitochondrial compartment. *Journal of Experimental Botany*, 57(6), 1225–1243.
- Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics*, 26(8), 345–352.
- Lynch, M. (2011). The Lower Bound to the Evolution of Mutation Rates. *Genome Biology and Evolution*, (p. evr066).

- Lynch, M., & Blanchard, J. L. (1998). Deleterious mutation accumulation in organelle genomes. *Genetica*, 102-103(1-6), 29–39.
- Lynch, M., Koskella, B., & Schaack, S. (2006). Mutation Pressure and the Evolution of Organelle Genomic Architecture. *Science*, 311(5768), 1727–1730.
- Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., Dickinson, W. J., Okamoto, K., Kulkarni, S., Hartl, D. L., & Thomas, W. K. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences*, 105(27), 9272–9277.
- Ma, H., & O'Farrell, P. H. (2015). Selections that isolate recombinant mitochondrial genomes in animals. *eLife*, 4, e07247.
- Ma, H., Xu, H., & O'Farrell, P. H. (2014). Transmission of mitochondrial mutations and action of purifying selection in *Drosophila*. *Nature genetics*, 46(4), 393–397.
- Maca-Meyer, N., Gonzlez, A. M., Larruga, J. M., Flores, C., & Cabrera, V. M. (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics*, 2, 13.
- Magnacca, K. N., & Brown, M. J. F. (2010). Tissue segregation of mitochondrial haplotypes in heteroplasmic Hawaiian bees: implications for DNA barcoding. *Molecular Ecology Resources*, 10(1), 60–68.
- Mai, Z., Ghosh, S., Frisardi, M., Rosenthal, B., Rogers, R., & Samuelson, J. (1999). Hsp60 Is Targeted to a Cryptic Mitochondrion-Derived Organelle (Crypton) in the Microaerophilic Protozoan Parasite *Entamoeba histolytica*. *Molecular and Cellular Biology*, 19(3), 2198–2205.
- Mannella, C. A. (2006). Structure and dynamics of the mitochondrial inner membrane cristae. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1763(56), 542–548.
- Marchington, D. R., Hartshorne, G. M., Barlow, D., & Poulton, J. (1997). Homopolymeric tract heteroplasmy in mtDNA from tissues and single oocytes: support for a genetic bottleneck. *American Journal of Human Genetics*, 60(2), 408–416.
- Margulis, L. (1970). *Origin of Eukaryotic Cells*. Yale University Press.

- Martijn, J., & Ettema, T. J. G. (2013). From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. *Biochemical Society Transactions*, 41(1), 451–457.
- Martin, W., & Mller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature*, 392(6671), 37–41.
- Martin, W. F., Garg, S., & Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. *Phil. Trans. R. Soc. B*, 370(1678), 20140330.
- Mat, M. L., Di Rocco, F., Zambelli, A., & Vidal-Rioja, L. (2007). Mitochondrial heteroplasmy in Control Region DNA of South American camelids. *Small Ruminant Research*, 71(13), 123–129.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303.
- Medvedev, P., Fiume, M., Dzamba, M., Smith, T., & Brudno, M. (2010). Detecting copy number variation with mated short reads. *Genome Research*, 20(11), 1613–1622.
- Mercer, J. M., & Roth, V. L. (2003). The Effects of Cenozoic Global Change on Squirrel Phylogeny. *Science*, 299(5612), 1568–1572.
- Meusel, M. S., & Moritz, R. F. A. (1993). Transfer of paternal mitochondrial DNA during fertilization of honeybee (*Apis mellifera* L.) eggs. *Current Genetics*, 24(6), 539–543.
- Mjelle, K. A., Karlsen, B. O., Jrgensen, T. E., Moum, T., & Johansen, S. D. (2008). Halibut mitochondrial genomes contain extensive heteroplasmic tandem repeat arrays involved in DNA recombination. *BMC Genomics*, 9(1), 10.
- Molnar, R. I., Bartelmes, G., Dinkelacker, I., Witte, H., & Sommer, R. J. (2011). Mutation Rates and Intraspecific Divergence of the Mitochondrial Genome of *Pristionchus pacificus*. *Molecular Biology and Evolution*, 28(8), 2317–2326.
- Monnat, R. J., & Loeb, L. A. (1985). Nucleotide sequence preservation of human mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 82(9), 2895–2899.

- Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1), 2–9.
- Muller, M. (1994). The hydrogenosome. *Journal of general microbiology*, 139(12), 2879–89.
- Mller, M., Mentel, M., Hellemond, J. J. v., Henze, K., Woehle, C., Gould, S. B., Yu, R.-Y., Giezen, M. v. d., Tielens, A. G. M., & Martin, W. F. (2012). Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes. *Microbiology and Molecular Biology Reviews*, 76(2), 444–495.
- Nabholz, B., Glmin, S., & Galtier, N. (2008). Strong Variations of Mitochondrial Mutation Rate across Mammals—the Longevity Hypothesis. *Molecular Biology and Evolution*, 25(1), 120–130.
- Nakada, K., Inoue, K., Ono, T., Isobe, K., Ogura, A., Goto, Y. I., Nonaka, I., & Hayashi, J. I. (2001). Inter-mitochondrial complementation: Mitochondria-specific system preventing mice from expression of disease phenotypes by mutant mtDNA. *Nature Medicine*, 7(8), 934–940.
- Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M. C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, M., Ogasawara, N., & Kanaya, S. (2011). Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Research*, (p. gkr344).
- Nass, M. M. (1966). The circularity of mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 56(4), 1215–1222.
- Nass, M. M., & Nass, S. (1963). Intramitochondrial Fibers With DNA Characteristics. I. Fixation and Electron Staining Reactions. *The Journal of Cell Biology*, 19, 593–611.
- Naue, J., Hrter, S., Snger, T., Strobl, C., Hatzler-Grubwieser, P., Parson, W., & Lutz-Bonengel, S. (2015). Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. *Mitochondrion*, 20, 82–94.
- Neiman, M., Hehman, G., Miller, J. T., Logsdon, J. M., & Taylor, D. R. (2010). Accelerated Mutation Accumulation in Asexual Lineages of a Freshwater Snail. *Molecular Biology and Evolution*, 27(4), 954–963.

- Ness, R. W., Morgan, A. D., Vasanthakrishnan, R. B., Colegrave, N., & Keightley, P. D. (2015). Extensive de novo mutation rate variation between individuals and across the genome of *Chlamydomonas reinhardtii*. *Genome Research*, (p. gr.191494.115).
- Normark, B. B., & Moran, N. A. (2000). Testing for the accumulation of deleterious mutations in asexual eukaryote genomes using molecular sequences. *Journal of Natural History*, 34(9), 1719–1729.
- Olivo, P. D., Van de Walle, M. J., Laipis, P. J., & Hauswirth, W. W. (1983). Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. *Nature*, 306(5941), 400–402.
- Paduan, K. D. S., & Ribolla, P. E. M. (2008). Mitochondrial DNA Polymorphism and Heteroplasmy in Populations of *Aedes aegypti* in Brazil. *Journal of Medical Entomology*, 45(1), 59–67.
- Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S.-E., Walford, G. A., Sugiana, C., Boneh, A., Chen, W. K., Hill, D. E., Vidal, M., Evans, J. G., Thorburn, D. R., Carr, S. A., & Mootha, V. K. (2008). A mitochondrial protein compendium elucidates complex I disease biology. *Cell*, 134(1), 112–123.
- Parfrey, L. W., Lahr, D. J. G., Knoll, A. H., & Katz, L. A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences*, 108(33), 13624–13629.
- Parr, R. L., Maki, J., Reguly, B., Dakubo, G. D., Aguirre, A., Wittock, R., Robinson, K., Jakupciak, J. P., & Thayer, R. E. (2006). The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. *BMC Genomics*, 7, 185.
- Payne, B. A. I., Wilson, I. J., Yu-Wai-Man, P., Coxhead, J., Deehan, D., Horvath, R., Taylor, R. W., Samuels, D. C., Santibanez-Koref, M., & Chinnery, P. F. (2013). Universal heteroplasmy of human mitochondrial DNA. *Human Molecular Genetics*, 22(2), 384–390.
- Percudani, R. (2013). A Microbial Metagenome (*Leucobacter* sp.) in *Caenorhabditis* Whole Genome Sequences. *Bioinformatics and Biology Insights*, 7, 55–72.
- Perkins, G., Renken, C., Martone, M. E., Young, S. J., Ellisman, M., & Frey, T. (1997). Electron tomography of neuronal mitochondria: three-dimensional

- structure and organization of cristae and membrane contacts. *Journal of Structural Biology*, 119(3), 260–272.
- Perkins, G. A., & Frey, T. G. (2000). Recent structural insight into mitochondria gained by microscopy. *Micron*, 31(1), 97–111.
- Piganeau, G., & Eyre-Walker, A. (2009). Evidence for Variation in the Effective Population Size of Animal Mitochondrial DNA. *PLoS ONE*, 4(2).
- Piganeau, G., Gardner, M., & Eyre-Walker, A. (2004). A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution*, 21(12), 2319–2325.
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), 841–842.
- R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN 3-900051-07-0.
- Radojii, J. M., Krizmani, I., Kasapidis, P., & Zouros, E. (2015). Extensive mitochondrial heteroplasmy in hybrid water frog (*Pelophylax* spp.) populations from Southeast Europe. *Ecology and Evolution*, 5(20), 4529–4541.
- Rafelski, S. M. (2013). Mitochondrial network morphology: building an integrative, geometrical view. *BMC Biology*, 11, 71.
- Rand, D. M., & Harrison, R. G. (1986). Mitochondrial DNA Transmission Genetics in Crickets. *Genetics*, 114(3), 955–970.
- Robinson, J. T., Thorvaldsdttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24–26.
- Robison, G. A., Balvin, O., Schal, C., Vargo, E. L., & Booth, W. (2015). Extensive Mitochondrial Heteroplasmy in Natural Populations of a Resurging Human Pest, the Bed Bug (Hemiptera: Cimicidae). *Journal of Medical Entomology*, (p. tju055).
- Sager, R., & Ramanis, Z. (1968). The Pattern of Segregation of Cytoplasmic Genes in Chlamydomonas. *Proceedings of the National Academy of Sciences of the United States of America*, 61(1), 324–331.

- Samuels, D. C., Li, C., Li, B., Song, Z., Torstenson, E., Boyd Clay, H., Rokas, A., Thornton-Wells, T. A., Moore, J. H., Hughes, T. M., Hoffman, R. D., Haines, J. L., Murdock, D. G., Mortlock, D. P., & Williams, S. M. (2013). Recurrent Tissue-Specific mtDNA Mutations Are Common in Humans. *PLoS Genet*, 9(11), e1003929.
- Sato, A., Nakada, K., Akimoto, M., Ishikawa, K., Ono, T., Shitara, H., Yonekawa, H., & Hayashi, J.-I. (2005). Rare creation of recombinant mtDNA haplotypes in mammalian tissues. *Proceedings of the National Academy of Sciences*, 102(17), 6057–6062.
- Sato, M., & Sato, K. (2013). Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1833(8), 1979–1984.
- Satoh, M., & Kuroiwa, T. (1991). Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Experimental Cell Research*, 196(1), 137–140.
- Saxer, G., Havlak, P., Fox, S. A., Quance, M. A., Gupta, S., Fofanov, Y., Strassmann, J. E., & Queller, D. C. (2012). Whole Genome Sequencing of Mutation Accumulation Lines Reveals a Low Mutation Rate in the Social Amoeba *Dictyostelium discoideum*. *PLoS ONE*, 7(10), e46759.
- Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research*, (p. gku1341).
- Schon, E. A., DiMauro, S., & Hirano, M. (2012). Human mitochondrial DNA: roles of inherited and somatic mutations. *Nature Reviews Genetics*, 13(12), 878–890.
- Schrider, D. R., Hourmozdi, J. N., & Hahn, M. W. (2011). Pervasive multinucleotide mutational events in eukaryotes. *Current biology: CB*, 21(12), 1051–1054.
- Schrider, J., Hsu, A., Boyle, S. E., Macintyre, G., Cmero, M., Tothill, R. W., Johnstone, R. W., Shackleton, M., & Papenfuss, A. T. (2014). Socrates: identification of genomic rearrangements in tumour genomes by re-aligning soft clipped reads. *Bioinformatics*, 30(8), 1064–1072.
- Schwartz, M., & Vissing, J. (2002). Paternal Inheritance of Mitochondrial DNA. *New England Journal of Medicine*, 347(8), 576–580.

- Schwarz, T. L. (2013). Mitochondrial Trafficking in Neurons. *Cold Spring Harbor Perspectives in Biology*, 5(6), a011304.
- Searcy, D. (1992). Origins of mitochondria and chloroplasts from sulphur-based symbioses. In *The origin and evolution of the cell*, (pp. 47–78). Singapore: World Scientific.
- Seyfert, A. L., Cristescu, M. E. A., Frisse, L., Schaack, S., Thomas, W. K., & Lynch, M. (2008). The Rate and Spectrum of Microsatellite Mutation in *Caenorhabditis elegans* and *Daphnia pulex*. *Genetics*, 178(4), 2113–2121.
- Sharpley, M., Marciniak, C., Eckel-Mahan, K., McManus, M., Crimi, M., Waymire, K., Lin, C., Masubuchi, S., Friend, N., Koike, M., Chalkia, D., MacGregor, G., Sassone-Corsi, P., & Wallace, D. (2012). Heteroplasmy of Mouse mtDNA Is Genetically Unstable and Results in Altered Behavior and Cognition. *Cell*, 151(2), 333–343.
- Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), 1135–1145.
- Shiflett, A., & Johnson, P. J. (2010). Mitochondrion-related Organelles in Parasitic Eukaryotes. *Annual review of microbiology*, 64, 409–429.
- Shoubridge, E. A., & Wai, T. (2007). Mitochondrial DNA and the Mammalian Oocyte. vol. 77 of *The Mitochondrion in the Germline and Early Development*, (pp. 87–111). Academic Press.
- Shuster, R. C., Rubenstein, A. J., & Wallace, D. C. (1988). Mitochondrial DNA in anucleate human blood cells. *Biochemical and Biophysical Research Communications*, 155(3), 1360–1365.
- Smith, N. G. C., Webster, M. T., & Ellegren, H. (2002). Deterministic Mutation Rate Variation in the Human Genome. *Genome Research*, 12(9), 1350–1356.
- Sniegowski, P. D., Gerrish, P. J., Johnson, T., & Shaver, A. (2000). The evolution of mutation rates: separating causes from consequences. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 22(12), 1057–1066.
- Solignac, M., Gnermont, J., Monnerot, M., & Mounolou, J.-C. (1984). Genetics of mitochondria in *Drosophila*: mtDNA inheritance in heteroplasmic strains of *D. mauritiana*. *Molecular and General Genetics MGG*, 197(2), 183–188.

- Solignac, M., Monnerot, M., & Mounolou, J. C. (1983). Mitochondrial DNA heteroplasmy in *Drosophila mauritiana*. *Proceedings of the National Academy of Sciences of the United States of America*, 80(22), 6942–6946.
- Stechmann, A., Hamblin, K., Prez-Brocal, V., Gaston, D., Richmond, G., van der Giezen, M., Clark, C., & Roger, A. J. (2008). Organelles in Blastocystis that Blur the Distinction between Mitochondria and Hydrogenosomes. *Current Biology*, 18(8), 580–585.
- Stewart, J. B., & Chinnery, P. F. (2015). The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nature Reviews Genetics*, 16(9), 530–542.
- Stewart, J. B., & Larsson, N.-G. (2014). Keeping mtDNA in Shape between Generations. *PLoS Genetics*, 10(10).
- Strong, M. J., Xu, G., Morici, L., Splinter Bon-Durant, S., Baddoo, M., Lin, Z., Fewell, C., Taylor, C. M., & Flemington, E. K. (2014). Microbial Contamination in Next Generation Sequencing: Implications for Sequence-Based Analysis of Clinical Samples. *PLoS Pathogens*, 10(11).
- Sturtevant, A. H. (1937). On the Effects of Selection on Mutation Rate. *The Quarterly Review of Biology*, 12(4), 464–467.
- Sung, W., Tucker, A. E., Doak, T. G., Choi, E., Thomas, W. K., & Lynch, M. (2012). Extraordinary genome stability in the ciliate *Paramecium tetraurelia*. *Proceedings of the National Academy of Sciences*, 109(47), 19339–19344.
- Supek, F., & Lehner, B. (2015). Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature*, 521(7550), 81–84.
- Sutovsky, P. (2003). Ubiquitin-dependent proteolysis in mammalian spermatogenesis, fertilization, and sperm quality control: killing three birds with one stone. *Microscopy Research and Technique*, 61(1), 88–102.
- Tait, S. W. G., & Green, D. R. (2010). Mitochondria and cell death: outer membrane permeabilization and beyond. *Nature Reviews Molecular Cell Biology*, 11(9), 621–632.
- Taylor, D. R., Zeyl, C., & Cooke, E. (2002). Conflicting levels of selection in the accumulation of mitochondrial defects in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, 99(6), 3690–3694.

- Taylor, R. W., & Turnbull, D. M. (2005). Mitochondrial DNA mutations in human disease. *Nature reviews. Genetics*, 6(5), 389–402.
- Thomas, W. K., & Wilson, A. C. (1991). Mode and Tempo of Molecular Evolution in the Nematode *Caenorhabditis*: Cytochrome Oxidase II and Calmodulin Sequences. *Genetics*, 128(2), 269–279.
- Thyagarajan, B., Padua, R. A., & Campbell, C. (1996). Mammalian mitochondria possess homologous DNA recombination activity. *The Journal of Biological Chemistry*, 271(44), 27536–27543.
- Thyagarajan, B., Wang, R., Barcelo, H., Koh, W.-P., & Yuan, J.-M. (2012). Mitochondrial Copy Number is Associated with Colorectal Cancer Risk. *Cancer Epidemiology Biomarkers & Prevention*, 21(9), 1574–1581.
- Thyagarajan, B., Wang, R., Nelson, H., Barcelo, H., Koh, W.-P., & Yuan, J.-M. (2013). Mitochondrial DNA copy number is associated with breast cancer risk. *PloS One*, 8(6), e65968.
- Timmis, J. N., Ayliffe, M. A., Huang, C. Y., & Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics*, 5(2), 123–135.
- Tovar, J. (2007). Mitosomes of Parasitic Protozoa: Biology and Evolutionary Significance. In D. W. F. Martin, & D. M. Müller (Eds.), *Origin of Mitochondria and Hydrogenosomes*, (pp. 277–300). Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-38502-8_11.
- Tsaousis, A. D., Martin, D. P., Ladoukakis, E. D., Posada, D., & Zouros, E. (2005). Widespread recombination in published animal mtDNA sequences. *Molecular Biology and Evolution*, 22(4), 925–933.
- Ujvari, B., Dowton, M., & Madsen, T. (2007). Mitochondrial DNA recombination in a free-ranging Australian lizard. *Biology Letters*, 3(2), 189–192.
- Umar, A., & Kunkel, T. A. (1996). DNA-replication fidelity, mismatch repair and genome instability in cancer cells. In *EJB Reviews 1996*, no. 1996 in EJB Reviews, (pp. 163–173). Springer Berlin Heidelberg.
- Ursprung, H., & Schabtach, E. (1965). Fertilization in tunicates: Loss of the paternal mitochondrion prior to sperm entry. *Journal of Experimental Zoology*, 159(3), 379–383.

- Vellai, T., & Vida, G. (1999). The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. *Proceedings of the Royal Society B: Biological Sciences*, 266(1428), 1571–1577.
- Viguera, E., Canceill, D., & Ehrlich, S. (2001). Replication slippage involves DNA polymerase pausing and dissociation. *The EMBO Journal*, 20(10), 2587–2595.
- von Dohlen, C. D., Kohler, S., Alsop, S. T., & McManus, W. R. (2001). Mealybug -proteobacterial endosymbionts contain -proteobacterial symbionts. *Nature*, 412(6845), 433–436.
- Wai, T., Teoli, D., & Shoubridge, E. A. (2008). The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. *Nature Genetics*, 40(12), 1484–1488.
- Wall, J. D., Tang, L. F., Zerbe, B., Kvale, M. N., Kwok, P.-Y., Schaefer, C., & Risch, N. (2014). Estimating genotype error rates from high-coverage next-generation sequence data. *Genome Research*, (p. gr.168393.113).
- Wang, X., Blades, N., Ding, J., Sultana, R., & Parmigiani, G. (2012). Estimation of sequencing error rates in short reads. *BMC Bioinformatics*, 13, 185.
- Wawrzyniak, I., Roussel, M., Diogon, M., Couloux, A., Texier, C., Tan, K. S. W., Vivars, C. P., Delbac, F., Wincker, P., & El Alaoui, H. (2008). Complete circular DNA in the mitochondria-like organelles of *Blastocystis hominis*. *International Journal for Parasitology*, 38(12), 1377–1382.
- Weir, J. T., & Schlüter, D. (2007). The Latitudinal Gradient in Recent Speciation and Extinction Rates of Birds and Mammals. *Science*, 315(5818), 1574–1576.
- Welch, J. J., & Bromham, L. (2005). Molecular dating when rates vary. *Trends in Ecology & Evolution*, 20(6), 320–327.
- Weller, A. M., Rdelsperger, C., Eberhardt, G., Molnar, R. I., & Sommer, R. J. (2014). Opposing forces of A/T-biased mutations and G/C-biased gene conversions shape the genome of the nematode *Pristionchus pacificus*. *Genetics*, 196(4), 1145–1152.
- White, D. J., Wolff, J. N., Pierson, M., & Gemmell, N. J. (2008). Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, 17(23), 4925–4942.

- Williams, B. A. P., Hirt, R. P., Lucocq, J. M., & Embley, T. M. (2002). A mitochondrial remnant in the microsporidian Trachipleistophora hominis. *Nature*, *418*(6900), 865–869.
- Williams, T. A., Foster, P. G., Nye, T. M. W., Cox, C. J., & Embley, T. M. (2012). A congruent phylogenomic signal places eukaryotes within the Archaea. *Proceedings of the Royal Society of London B: Biological Sciences*, (p. rspb20121795).
- Wilm, A., Aw, P. P. K., Bertrand, D., Yeo, G. H. T., Ong, S. H., Wong, C. H., Khor, C. C., Petric, R., Hibberd, M. L., & Nagarajan, N. (2012). LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*, *40*(22), 11189–11201.
- Wolff, J. N., & Gemmell, N. J. (2008). Lost in the zygote: the dilution of paternal mtDNA upon fertilization. *Heredity*, *101*(5), 429–434.
- Wolff, J. N., Nafisina, M., Sutovsky, P., & Ballard, J. W. O. (2013). Paternal transmission of mitochondrial DNA as an integral part of mitochondrial inheritance in metapopulations of *Drosophila simulans*. *Heredity*, *110*(1), 57–62.
- Wolff, J. N., White, D. J., Woodhams, M., White, H. E., & Gemmell, N. J. (2011). The Strength and Timing of the Mitochondrial Bottleneck in Salmon Suggests a Conserved Mechanism in Vertebrates. *PLoS ONE*, *6*(5).
- Wolstenholme, D. R. (1992). Animal mitochondrial DNA: structure and evolution. *International Review of Cytology*, *141*, 173–216.
- Xu, S., Omilian, A. R., & Cristescu, M. E. (2011). High Rate of Large-Scale Hemizygous Deletions in Asexually Propagating *Daphnia*: Implications for the Evolution of Sex. *Molecular Biology and Evolution*, *28*(1), 335–342.
- Xu, S., Schaack, S., Seyfert, A., Choi, E., Lynch, M., & Cristescu, M. E. (2012). High Mutation Rates in the Mitochondrial Genomes of *Daphnia pulex*. *Molecular Biology and Evolution*, *29*(2), 763–769.
- Yarlett, N., Hann, A. C., Lloyd, D., & Williams, A. (1981). Hydrogenosomes in the rumen protozoon *Dasytricha ruminantium* Schuberg. *Biochemical Journal*, *200*(2), 365–372.

- Yarlett, N., Orpin, C. G., Munn, E. A., Yarlett, N. C., & Greenwood, C. A. (1986). Hydrogenosomes in the rumen fungus *Neocallimastix patriciarum*. *Biochemical Journal*, 236(3), 729–739.
- Ye, F., Samuels, D. C., Clark, T., & Guo, Y. (2014). High-throughput sequencing in mitochondrial DNA research. *Mitochondrion*, 17, 157–163.
- Youle, R. J., & Bliek, A. M. v. d. (2012). Mitochondrial Fission, Fusion, and Stress. *Science*, 337(6098), 1062–1065.
- Yutin, N., Wolf, M. Y., Wolf, Y. I., & Koonin, E. V. (2009). The origins of phagocytosis and eukaryogenesis. *Biology Direct*, 4(1), 9.
- Zamzami, N., Marchetti, P., Castedo, M., Zanin, C., Vayssire, J. L., Petit, P. X., & Kroemer, G. (1995). Reduction in mitochondrial potential constitutes an early irreversible step of programmed lymphocyte death in vivo. *The Journal of Experimental Medicine*, 181(5), 1661–1672.
- Zhao, X., Li, N., Guo, W., Hu, X., Liu, Z., Gong, G., Wang, A., Feng, J., & Wu, C. (2004). Further evidence for paternal inheritance of mitochondrial DNA in the sheep (*Ovis aries*). *Heredity*, 93(4), 399–403.
- Zhou, R. R., Wang, B., Wang, J., Schatten, H., & Zhang, Y. Z. (2010). Is the mitochondrial cloud the selection machinery for preferentially transmitting wild-type mtDNA between generations? Rewinding Muller's ratchet efficiently. *Current Genetics*, 56(2), 101–107.
- Zhou, W., Zhu, M., Gui, M., Huang, L., Long, Z., Wang, L., Chen, H., Yin, Y., Jiang, X., Dai, Y., Tang, Y., He, L., & Zhong, K. (2014). Peripheral blood mitochondrial DNA copy number is associated with prostate cancer risk and tumor burden. *PloS One*, 9(10), e109470.
- Zouros, E., Freeman, K. R., Ball, A. O., & Pogson, G. H. (1992). Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. *Nature*, 359(6394), 412–414.
- Zouros, E., Oberhauser Ball, A., Saavedra, C., & Freeman, K. R. (1994). An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proceedings of the National Academy of Sciences of the United States of America*, 91(16), 7463–7467.

- Zuckerkandl, E., & Pauling, L. (1965). Evolutionary Divergence and Convergence in Proteins. In H. J. Vogel (Ed.), *Evolving Genes and Proteins*, (pp. 97–166). Academic Press.