The impact of deleterious mutations on the transition to meiotic sex and the structure of the germline

Marco Colnaghi

University College London

Research Degree: COMPLEX PhD

UCL Centre for Mathematics and Physics in the Life Sciences and Experimental Biology

UCL Department of Genetics, Evolution and Environment

I, Marco Colnaghi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

To all my teachers – past, present, and future.

Table of Contents

Acknowledgements

It would be impossible to start this section with someone else than Prof. Nick Lane and Prof. Andrew Pomiankowski - words are not enough to express my gratitude to them. During the last 5 years they have not only been terrific supervisors, but also great mentors and teachers. Without their support, stimulating dialogues, and many insights and suggestions, this work would have never existed. What little I understand about biology and how to do science, I mostly owe it to them. My gratitude also goes to Dr. Alessia Annibale and Prof. Ton Coolen from King's College Department of Mathematics, to Prof. Giuseppe Carini from the University of Messina, and to all of my other teachers and mentors for their patient and inspiring guidance. I am extremely thankful to bgc3 and the UCL Department of Genetics, Evolution and Environment for their financial support throughout my PhD, and to Prof. Franjo Weissing and Prof. Chris Barnes, who took the time to read and review this thesis and provided me with insightful comments and suggestions (besides a very stimulating Viva Voce examination!).

Huge thanks to all colleagues and comrades of CoMPLEX and GEE, especially Raquel Nunes Palmeira, Hanadi Rammu, Enrique Rodriguez, Stuart Harrison, Florencia Camus, and to all the other members of the Lane and Pomiankowski groups, for the insightful discussions, their friendship, and the joy of working together in such a stimulating environment, alongside brilliant and creative minds. My gratitude also goes to all the members of the Seed-Vault Society (whose names, having sworn a vow of secrecy, I will have to pass on in silence) and to all my friends and colleagues from Sicily, especially Roberto Bertilone, Andrea Oteri, my brother Davide, 'Ciccio' Patti, and the late Antonio Anastasi, who will be sorely missed.

As I write these words, I cannot help but think of my first initiation as a scientist – a visit to the lab where my mother, a researcher in Veterinary Microbiology, cultured a plethora of microbes. I still remember the thrill, as a young boy, of looking into a microscope for the first time. I also remember the wide array of books on various scientific topics – from astronomy to geology to dinosaurs – that my father once bought me. These are just two small examples of how, throughout my life, my parents always did their best to support my curiosity, creativity, intellectual growth, and scientific interests. Heartfelt thanks to them both. I am also grateful to my grandmother, Amelia, and to my late grandfather, Cristofaro, for their guidance and support.

Among my numerous mentors and teachers throughout my life, all of which have my profound gratitude, I expressly wish to thank those who helped me achieve a better understanding of Buddhist philosophy, which has had a deep influence on my life, core values, and way of thinking. In particular, I wish to thank His Holiness the 14th Dalai Lama, Geshe Tenzin Namdak, Ven. Fabienne Pradelle, Lama Zangmo, and Scott Snibbe, for their wisdom, kindness, and inspiring presence.

Finally, I wish to thank Sajda, my partner and now fiancée, from the very bottom of my heart. In the last 6 years, she has supported me in countless ways, helping me to grow intellectually, emotionally, and spiritually, in my personal and professional life – I do not know where I would be without her love, trust, care, and companionship. My deepest thanks go to her, for being my side throughout the whole of my PhD, for sharing every joy and every hardship I met along the road, and for loving me deeply and sincerely through all the ups and downs, regardless of good or bad fortune.

Abstract

The accumulation of deleterious mutations predicted by Muller's ratchet – the progressive increase in mutation load caused by genetic drift – can cause the extinction of asexual populations and is considered one of the forces behind the maintenance of sex, the evolution of sex chromosomes, and the loss of genetic information in inversions. Here, I investigate the extent to which this process has influenced, constrained, and shaped the evolution of life. Using theoretical and computational models, I demonstrate that the need for increased purifying selection played a key role in major evolutionary transitions, focussing on the origin of meiotic sex and the evolution of the female germline.

Early models of the origin of sex have generally focussed on the transition from asexual to sexual lifestyle. It is now universally accepted that prokaryotes undergo homologous recombination via lateral gene transfer (LGT), which can prevent the mutational meltdown predicted by Muller's ratchet. Here, I investigate the origin of sex as part of the transition from prokaryotes to eukaryotes. I develop a theoretical model to investigate the impact of the increase in genome size and density of genomic repeats that took place during eukaryogenesis. My results indicate that these conditions led to the failure of LGT, generating a strong selective pressure for the origin of meiotic sex. But while meiotic sex can facilitate purifying selection on nuclear genes, it cannot prevent the accumulation of mutations in mitochondrial DNA (mtDNA). I demonstrate that the need to preserve mtDNA against Muller's ratchet caused the evolution of tight mechanisms for mitochondrial quality control, shaping the evolution of the female germline in metazoans.

This theoretical framework can be applied to a wide range of biological processes, including bacterial evolution, genome streamlining in organelles and endosymbionts, and the evolution of a two-step meiosis.

Impact Statement

In this thesis, I develop novel theoretical frameworks centred around the idea that deleterious mutations constitute a creative evolutionary force and drive adaptation and innovation. The hypotheses and theoretical models presented here constitute an advance in our understanding of the origin of meiotic sex and transmission of mitochondrial mutations and can potentially open up new avenues of research.

The origin of sex has been long considered the 'Queen of problems in evolutionary biology'. In Chapters 2 and 3 of this thesis, I develop a novel approach to tackle this problem, presenting a new framework to understand the transition from prokaryotes to eukaryotes. The idea that the failure of lateral gene transfer (LGT) in preserving genomic integrity led to the evolution of sexual reproduction can inform our understanding of eukaryotic evolution and the origin of meiotic sex. The results presented are relevant to evolutionary biologists, especially those interested in the emergence of the eukaryotes, and will potentially shape the field's understanding of the origins of meiotic sex.

To test the hypothesis presented in Chapters 2 and 3, I developed a novel theoretical model of the accumulation of mutation (Muller's ratchet) in populations undergoing lateral gene transfer (LGT). This model broadens the scope of previous theoretical studies by considering variables that had previously been neglected (genome size, recombination length, genetic repeat density, ectopic recombination). This development allows Muller's ratchet models to be applied to the study of ancestral eukaryotic evolution and several new areas. In Chapter 5, I discuss new avenues of research in which this theoretical framework can yield promising results, such as the evolution of prokaryotic genomes, the process of genome streamlining in endosymbionts and organelles, and the evolution of 2-step meiosis.

Chapter 2 has been published in 2020 (Colnaghi *et al*., 2020), while Chapter 3 is currently being prepared for publication. The contents of these chapters have already been the focus of several presentations during national and international conferences and symposia, both by the author and by his collaborators, and has been included in academic curricula, informing the understanding of both current and future academics.

Chapter 4 elucidates complex features of the female germline and demonstrates the evolutionary need for organelle-level selection to prevent the build-up of mitochondrial mutations. This work will provide a new focus for research on mitochondrial inheritance and its relation to mitochondrial diseases and common conditions including diabetes, cancer, and dementia. My findings that selection for mitochondrial quality dictates the major developmental stages of the female germline will be of fundamental interest to biologists and may be important for emerging technologies for editing mitochondrial genes. While there has been abundant theoretical and empirical work on germline bottlenecks and follicular atresia, many papers are mutually contradictory and just add to the confusion over how selection against mitochondrial mutations operates. My analysis provides the field with a clear understanding of how different selective processes operate, clarifying the dynamics of transmission of mitochondrial mutations and addressing long-standing paradoxes such as the over-proliferation and massive loss of germ cells.

Introduction

Vicissitudes of fortune, which spares neither man nor the proudest of his works, which buries empires and cities in a common grave…

(E. Gibbon, 'The History of the Decline and Fall of the Roman Empire', Chapter LXXI)

Mutations have been defined as 'the raw materials of evolution', the source of variation upon which natural selection operates. However, only a small proportion of them are beneficial. The vast majority have either a negligible or negative impact on the fitness of their carrier. The accumulation of deleterious mutations is an evolutionary force that any life form must deal with in order to preserve the integrity of its genetic information and ensure its transmission to future generations. In this thesis, I argue that this force played a major role in the evolution of life. With the use of theoretical models, I evaluate the extent to which the need for increased purifying selection against mutation accumulation impacted the evolution of eukaryotes, with a specific focus on the origin of meiotic sex and the evolution of purifying selection of mitochondrial DNA.

In chapter 1, I introduce the process known as 'Muller's ratchet', one of the main theoretical frameworks to study the accumulation of mutations in asexual populations. In the absence of strong selection and/or in the presence of strong genetic drift, mildly deleterious mutations can progressively accumulate in asexual populations because of random fluctuations in the number of offspring, ultimately causing extinction (mutational meltdown). This process is considered to be one of the driving forces behind the emergence of sexual recombination. Throughout the chapter, I review the main theoretical models of Muller's ratchet, discussing the extent to which this process impacts evolution. In addition, I

review the existing experimental evidence in order to evaluate under which conditions Muller's ratchet takes place in real populations and which mechanisms allow the persistence of asexual lineages.

Homologous recombination (HR)	The genetic exchange between two similar or identical DNA molecules, facilitating DNA repair. In eukaryotes, HR takes places between homologous chromosomes during meiosis (meiotic recombination) and, less occasionally, during mitosis (mitotic recombination). In prokaryotes, HR is achieved through a form of LGT known as transformation (Ambur *et al.*, 2016).
Ectopic recombination	The genetic exchange between two DNA molecules with low or no sequence homology. In prokaryotes undergoing LGT, the presence of genetic repeats can facilitate ectopic recombination, resulting in the loss of genetic information (Siguier *et al*., 2014; Vandecraen *et al*., 2017).
Meiotic sex	The process of sexual cell fusion followed by homologous pairing and recombination. Given the near-universality of meiotic genes in eukaryotes, meiotic sex likely evolved in early stages of eukaryotic evolution, before the divergence of the main eukaryotic lineages (Schurko and Logsdon, 2008; Speijer *et al.*, 2015)
Lateral gene transfer (LGT)	Exchange of genetic material through para-sexual processes which involve the transfer of DNA from one organism to another. In prokaryotes, the main mechanisms of LGT are conjugation (plasmid-mediated), transduction (phage-mediated) and transformation (controlled by the cell and leading to HR of environmental DNA) (Ambur *et al.*, 2016; Vos *et al.*, 2015). Throughout the thesis, when discussing the benefits of LGT, I mainly refer to the latter.
Muller's Ratchet	The progressive and inevitable accumulation of deleterious mutations predicted to take place in populations which do not undergo homologous recombination (Muller, 1964).
Asexual population	A population which does not undergo any sexual (meiotic sex) or para-sexual (homologous recombination via transformation) processes, and is therefore susceptible to the accumulation of deleterious mutations via Muller's ratchet (Muller, 1964). This term is used interchangeably with 'non-recombining population'.

Table 1.1. Definition of key terms.

Muller's Ratchet has long been considered to be one of the driving forces behind the emergence of sexual recombination, which enhances the action of purifying selection and

counters the increase in mutation load. Early theoretical studies on the origin of sex considered the transition from asexuality to meiotic recombination. But homologous recombination is not a eukaryotic prerogative: prokaryotes can achieve it through para-sexual mechanisms of Lateral Gene Transfer (LGT), such as transformation. The selective pressures that caused the failure of LGT and led to the evolution of meiosis are still poorly understood. In Chapter 2, I tackle this question by investigating the unique conditions that led to the origin of the first eukaryotes. Early eukaryotes underwent a massive genome size expansion via gene duplication and endosymbiotic gene transfer. This process was concomitant with a sharp increase in the density of repeat sequences, such as gene families and the selfish genetic elements that gave rise to eukaryotic introns. Using a theoretical and computational model, I demonstrate that Muller's ratchet is strongly dependent on genome size and recombination length, two variables that were neglected by previous theoretical models. My results indicate that genome size expansion cannot take place without a concomitant increase in recombination rate per base-pair.

In chapter 3, I broaden this theoretical model to include the presence of genomic repeats, and the possibility of deletions, insertions, and gene loading via LGT. I proceed to demonstrate that, in the presence of genomic repeats, increasing the rate of recombination leads to a catastrophic loss of genetic information because of deletions. Together, these results indicate that under the conditions of eukaryogenesis LGT could not prevent the accumulation of mutations, exposing the first eukaryotes to the risk of mutational meltdown. The increase in genome size and repeat density during eukaryogenesis caused the failure of LGT as a viable mechanism to prevent the accumulation of deleterious mutations, resulting in a strong selective pressure for the evolution of meiotic sex.

While elucidating the selective advantages of meiotic sex compared to LGT, the models presented in chapters 2 and 3 do not explicitly address the transition from the latter to the former. Obviously, natural selection does not operate with foresight, and the fact that meiosis corresponds to a higher fitness state than LGT does not automatically ensure its evolution. Future studies will address this question by a) investigating under which conditions can meiosis spread and reach fixation in a population undergoing LGT, and b) explicitly modelling the intermediate steps that led to the evolution of meiosis (e.g., the transition to cell fusion, ploidy cycles, and mitosis), evaluating their fitness benefits under ratchet conditions.

Given the uniparental inheritance of organelles without recombination, mitochondrial and chloroplast genomes are subjected to Muller's ratchet. Meiotic sex can purge deleterious mutations from the nuclear genome, but not from mitochondrial DNA (mtDNA). This makes mitochondrial genomes uniquely vulnerable to Muller's ratchet. Purifying selection must take place – either at the individual, cell, or organelle level – in order to prevent the fixation of deleterious mutations and the mutational meltdown of mitochondrial genomes. In chapter 4, I evaluate the extent to which different levels of selection (individual, cell, and organelle level) affect the accumulation of deleterious mutations in mammalian mtDNA. With the use of numerical simulations, I demonstrate that, for a realistic parameter range, individual and cell level selection is less effective than selection at the organelle level in preventing the deterioration of mitochondrial genomes. Understanding Muller's ratchet dynamics in mitochondrial genomes can be crucial to understand the evolution of the female germline.

Finally, in chapter 5 I discuss the relevance of these theoretical frameworks to other systems, such as viral and bacterial evolution, the origin of genetic inheritance and Eigen's

paradox, the process of genome streamlining in organelles and endosymbionts, and the origin of two-steps meiosis, highlighting unanswered questions that constitute possible avenues for future research.

Chapter 1 – The Evolutionary Impact of Muller's Ratchet

[A]n asexual population incorporates a kind of ratchet mechanism, such as it can never get to contain, in any of its lines, a load of mutation smaller than that already existing in its at present least loaded lines. (Muller, 1964)

1. Introduction

The first attempts to understand the advantages of sexual reproduction date back to Weisman (1887), who argued that sex provided 'a source of individual variability furnishing material for the operation of natural selection'. Both Fisher (1930) and Muller (1932) subsequently claimed that recombination is maintained by natural selection because it increases the probability of fixation of beneficial mutations. The Fisher-Muller hypothesis became the source of a controversy among evolutionary biologists, who tried to evaluate under which conditions can recombination aid the fixation of beneficial mutants (Crow and Kimura, 1965; 1969; Eshel and Feldman, 1970; Kimura and Ohta, 1971; Maynard Smith, 1968; 1971). It is now well-established that beneficial mutations are rare, and most mutations have a neutral or negative impact on fitness (Eyre-Walker and Keightley, 2007). The role of deleterious mutations in the evolution and maintenance of sexual reproduction was first postulated by Muller (1964), who shifted the focus away from beneficial mutations and suggested that deleterious mutations might have an impact on evolution. He predicted that asexual populations subjected to genetic drift would inevitably accumulate deleterious mutations, a process later known as Muller's ratchet (Felsenstein, 1974; Muller, 1964).

In an infinite asexual population, neglecting the probability of beneficial mutations, the equilibrium fitness distribution is reached when the opposite forces of deleterious mutations and purifying selection are in balance. The equilibrium frequency of each line (characterised by a different mutation load) depends on the relative strength of the mutation rate and strength of selection (Haigh, 1978). In finite populations, however, genetic drift can cause the least-loaded line to go extinct because of random fluctuations in the number of offspring. In the absence of recombination, the least-loaded line cannot be restored: non-recombining populations cannot reduce their mutation burden below that of their least-loaded line. This process, illustrated in **Figure 1.1**, generates a ratchet mechanism which leads to the progressive accumulation of deleterious mutation (Muller, 1964). In the 'little-noticed passage' quoted at the beginning of this chapter, 'Muller introduced what may be the most quantitatively important evolutionary effect of recombination' (Felsenstein, 1974): the idea that asexual populations cannot escape a progressive, ratchet-like increase in their mutation load. In addition to asexual populations, Muller's ratchet has been shown to take place in highly selfing populations (Heller and Maynard Smith, 1978) and in regions of the genome where recombination is restricted, such as in Y chromosomes (Charlesworth, 1978) and mitochondrial DNA (Loewe, 2006; Lynch, 1996).

Theoretical models of Muller's ratchet are based on the approximation that beneficial mutations and back-mutations are so infrequent as to be negligible. As a consequence, the least-loaded line, once extinguished, can never be replenished, causing the progressive and inevitable increase of mutational burden. This assumption approximates a population where the ratchet is too severe for back-mutations or compensatory beneficial mutations (which arise at a much lower rate than deleterious ones) to halt it. Under certain conditions (e.g., strong genetic drift, small population size, or high

mutation rate) Muller's ratchet does take place despite possible compensatory mechanisms. But, as it will be discussed in the following paragraphs, it is hard to evaluate exactly when these conditions are met.

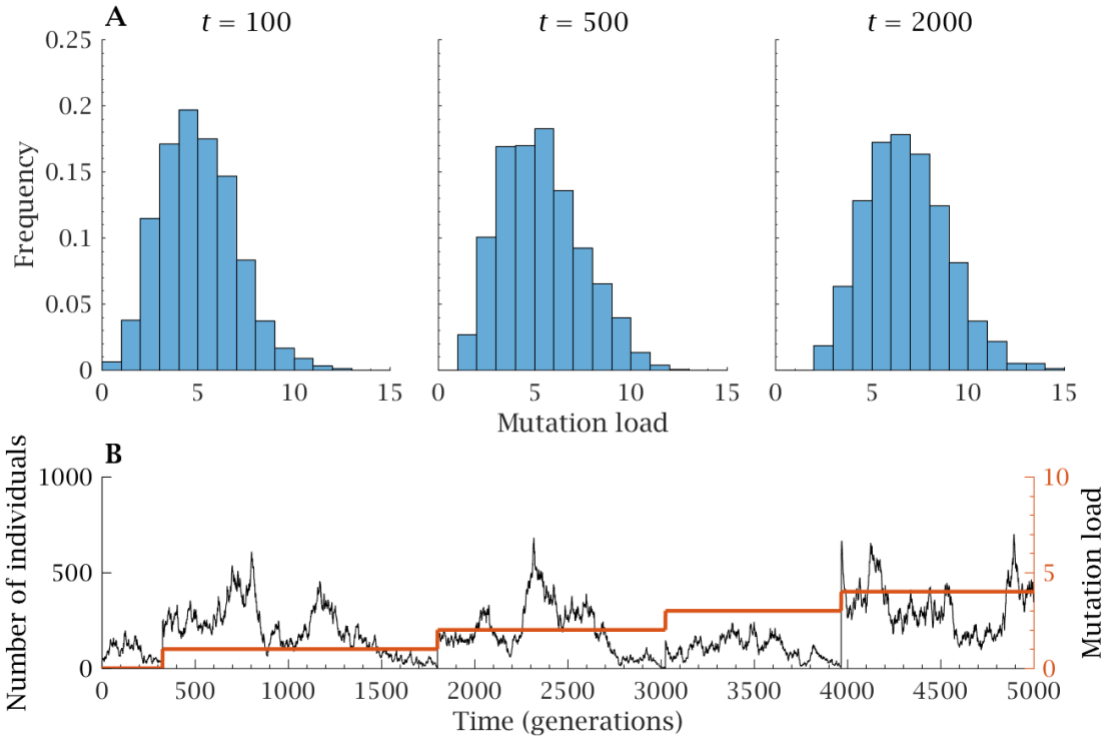

Figure 1.1 | Illustration of Muller's ratchet, using the computational model described in Chapter 2**.** (**A**) Distribution of mutation loads in a non-recombining population subjected to Muller's ratchet. After 500 generations, the mutation-free class has gone extinct and cannot be recovered, causing the whole distribution to shift towards higher mutation loads. After 2000 generations (upper right panel), the class of individual with 1 mutation is extinct – the ratchet has irreversibly 'clicked' again. (**B**) Number of individuals (left y-axis) and mutation load (right y-axis) in the least loaded class (LLC). The extinction of the LLC causes an irreversible increase in the minimum mutation load of the population. The initial average mutation load of the population is $\overline{m_0} = 0.05$. Other parameters: $N = 10{,}000$, $U = 0.04$, $s = 0.01$.

2. Theoretical Models

Muller's ratchet was first described quantitatively by Haigh (1978). In order to evaluate the impact of population size (N), strength of selection (s) and mutation rate (U) on ratchet

dynamics, Haigh calculated the distribution of mutations in the absence of genetic drift. Using a Fisher-Wright model, he showed that the number of individuals in the least-loaded class (LLC) at mutation-selection balance is $n_0 = Ne^{-U/s}$ (**Figure 1.2**) (Haigh, 1978). When genetic drift is taken into account, random fluctuations in the number of offspring can lead to the extinction of the least-loaded class (a 'click' of the ratchet), causing the distribution to shift towards higher mutation loads. The rate at which mutations accumulate depends on whether n_0 is small (fast ratchet) or large (slow ratchet). This, in turn, depends on population size and the relative ratio between U and s. Small values of U increase the number of individuals in the LLC, while smaller population size or weaker selection make a population more vulnerable to Muller's ratchet.

Haigh's model relies on a number of approximations, among which the assumption that finite asexual populations reach the same equilibrium distribution as infinite populations, and it is only then that finite-size effects (stochastic fluctuations in offspring number) lead to the extinction of the least loaded-class. Using Monte Carlo simulations, Gessler (1995) found that this is not necessarily the case: many asexual populations never reach Haigh's equilibrium distribution, if the time between clicks of the ratchet is smaller than the time necessary to transition to equilibrium. As a consequence, the variance of a finite population can be smaller than the variance of an infinite one as evaluated by Haigh (1978). This introduces another force driving Muller's ratchet in addition to the loss of the fittest class due to genetic drift (Gessler, 1995), which becomes stronger with higher mutation rates and decreases in the presence of strong selection (Neher and Shraiman, 2012).

Haigh's model also assumes an infinite number of loci, each with the same additive effect of mutations on fitness. Under this assumption, it is possible for the ratchet to click

without any marked increase in the frequency of any single mutant, leading to an underestimation of the rate of fixation of deleterious mutations (Charlesworth and Charlesworth, 1997). While Charlesworth *et al*. (1993) suggested that the accumulation of mutations through Muller's ratchet can proceed without a substantial increase in the number of fixed deleterious mutations, Higgs and Woodcock (1995) found that the loss of the fittest class coincided with fixation events. A further analysis by Charlesworth and Charlesworth (1997) confirmed that the advance of the ratchet corresponds to the fixation of a deleterious mutation at a single locus in haploid asexual populations. Following the extinction of the mutation-free class, genetic drift facilitates the fixation of a deleterious mutation in the new LLC, from which it then spreads throughout the rest of the population (Charlesworth and Charlesworth, 1997). Until this new LLC goes extinct, it is not possible for a second mutation to reach fixation. Ratchet dynamics and the fixation of mutants are intimately connected and, in the absence of recombination, the fixation rate of deleterious mutations is significantly higher than in sexual populations (Charlesworth and Charlesworth, 1997). The models developed in Chapters 2 and 3 relax the original assumption and assume a finite number of loci, explicitly including genome size as a model parameter and leading to a more accurate estimate of the rate of fixation.

A further assumption of Haigh's model (as well as the majority of the models developed in this thesis) is the focus on mutations with a fixed effect on fitness. This is, obviously, an oversimplification of biological reality. Studies on mutational robustness indicate that living organisms have evolved to tolerate reasonable loads of deleterious mutations without any sharp fitness decline (Wagner, 2013), potentially resulting in a higher tolerance to mutation accumulation. At the same time, this effect might actually exacerbate Muller's ratchet by 'masking deleterious mutations from natural selection and therefore

expediting their accumulation. Relaxing this assumption and investigating how gene networks are affected by Muller's ratchet would provide us with a more accurate picture of how mutations accumulate in living organisms, and should therefore be the object of future modelling work.

The evolutionary impact of Muller's ratchet is tightly linked to the speed of the ratchet; but in spite of more than four decades of theoretical efforts, the exact rate at which the ratchet 'clicks' is still hard to compute. Following Haigh (1978), numerous theoretical attempts have been made to estimate the speed of the ratchet and the average extinction time of the least-loaded class (Butcher, 1995; Charlesworth and Charlesworth, 1997; 2000; Etheridge *et al*., 2009; Gessler, 1995; Jain, 2008; Kondrashov, 1994; Neher and Shraiman, 2012; Stephan *et al*., 1993).

Using a diffusion approximation, Stephan *et al.* (1993) showed that when $n_0 > 1$ the time between 'clicks' of the ratchet decreases linearly with n_0 for intermediate values of N and s, and exponentially for large values of N and s (Stephan *et al.*, 1993). This indicates that small populations are extremely vulnerable to Muller's ratchet, while large populations with strong selection are virtually immune from it. For constant values of n_0, the speed of the ratchet is dictated by the ratio between the mutation rate and strength of selection, U/s (Gordo and Charlesworth, 2000b). A more recent study identified the quantity $\gamma = NU/(Ns \log(NU))$ as the key parameter in determining the rate of mutation accumulation, which is of the order $N^{\gamma-1}U^{\gamma}$ when $1 > \gamma > 0.5$, and negligible when $\gamma < 0.5$ (Etheridge *et al.*, 2009).

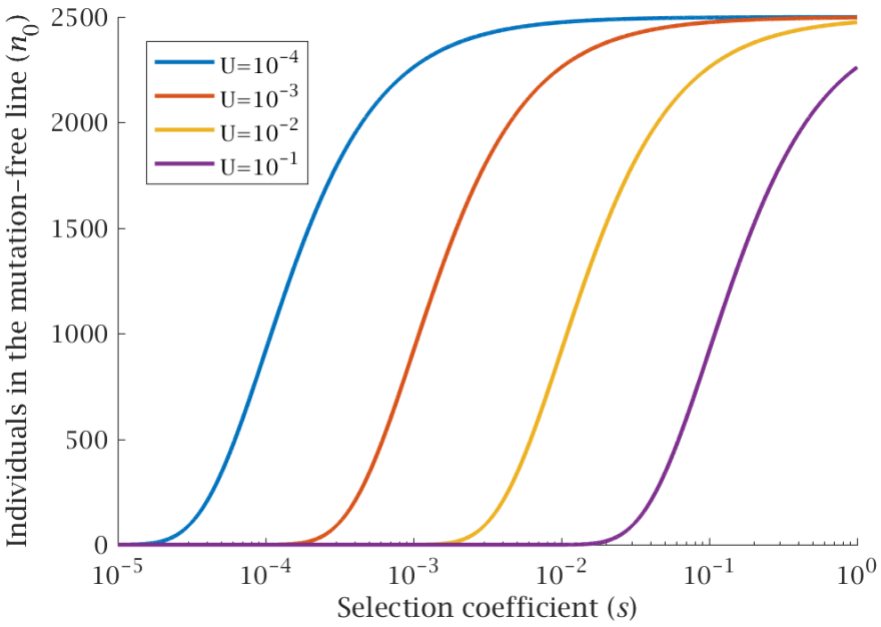

Figure 1.2 | Number of mutation-free individuals. The number of individuals in the mutation-free line is shown as a function of selection coefficient (s) and genome-wide mutation rate (U). Following Haigh (1978), the number of individuals in the mutation-free line at equilibrium is given by $n_0 = Ne^{-U/s}$. Stronger selection increases the number of individuals in the mutation-free line, while a higher mutation rate decreases it, making the population more vulnerable to stochastic fluctuations in the number of offspring and exposing it to Muller's ratchet. $N = 2{,}500$.

The interplay between different dynamics makes even the simplest models for Muller's ratchet hard to solve, and an exact solution for the average time between ratchet clicks is still lacking. Theoretical models indicate the presence of 4 different ratchet regimes, depending on the strength of selection: extremely weak, weak, intermediate, and high (Charlesworth and Charlesworth, 2000; Loewe, 2006). In the presence of extremely weak selection, the Hill-Robertson effect (the interference between linked alleles, which inhibits the elimination deleterious mutations and the spread of beneficial ones) is the driving force behind mutation accumulation in asexual populations; given the small fitness penalty per mutation, this leads to a long extinction time (Charlesworth and Charlesworth, 2000; Loewe, 2006). Weak selection leads to a quasi-deterministic ratchet regime, as described in Gessler

(1995). The small fitness loss per mutations makes it possible for back-mutations to occasionally restore the LLC, slowing down Muller's ratchet. Intermediate values of s lead to a stochastic Muller's ratchet, with a higher extinction probability compared to the previous scenarios. At this point there is a trade-off, as higher values of s favour the elimination of deleterious mutations, but also cause a faster decline in fitness when mutations do accumulate. As a consequence, extinction time remains approximately constant as s increases, until selection becomes too strong for mutations to accumulate (Charlesworth and Charlesworth, 2000; Loewe, 2006). In a strong selection regime, Muller's ratchet does not take place, as purifying selection keeps the population free from mutations. The above analysis is based on the approximation that all genes are under the same strength of selection. This is obviously not the case in real populations, and little is known about the extent to which ratchet dynamics are influenced by the distribution of fitness effects per mutation across the genome. In Chapter 2, I will address this question by using theoretical models to evaluate the rate of mutational decay in different areas of the genomes subjected to different strengths of selection.

In diploid asexual populations, ratchet dynamics are complicated by dominance. Small values of the dominance coefficient hinder the fixation of mutations at a locus which has already acquired a deleterious mutation in the other haploid set (Charlesworth and Charlesworth, 1997). Within the class of individuals with two mutations, a small dominance coefficient means that homozygotes for deleterious mutations have a selective disadvantage relative to individuals that are heterozygotes for two mutations at two different loci. This case is not dissimilar to that of synergistic epistasis, where the fitness penalty of deleterious mutations increases with the mutation load (Butcher, 1995; Charlesworth *et al.*, 1993; Jain, 2008; Kondrashov, 1994). While most ratchet models

generally assume that mutations act independently from each other, causing an equal decline in fitness, the introduction of synergistic epistasis can potentially retard (Charlesworth *et al.*, 1993) and even halt (Jain, 2008; Kondrashov, 1994) Muller's ratchet. This, however, is only true under the assumption that all mutations have an equal impact on fitness. When mutations are modelled as having a continuous distribution of fitness effects (which is likely the case in real populations) their accumulation proceeds even in the presence of synergistic epistasis (Butcher, 1995).

Despite the difficulties in evaluating the exact time between clicks of the ratchet, theoretical models provide us with a qualitative understanding about which conditions would likely favour the accumulation of mutations in real populations. A higher mutation rate causes faster accumulation of mutations, leading to an exponential decrease of the equilibrium number of individuals in the LLC (Haigh, 1978). Large population size limits the magnitude of genetic drift and enhances fitness variance, retarding the accumulation of mutations (Gessler, 1995; Haigh, 1978). If the fitness penalty per mutation is small, fixation of slightly deleterious mutations can take place at a high rate without causing a noticeable decline in fitness (Charlesworth and Charlesworth, 2000; Loewe, 2006). On the other hand, strong selection will prevent the fixation of mutation (think of the extreme case of a lethal mutation). Intermediate values of strength of selection are likely to increase the risk of mutational meltdown, ultimately leading to extinction (Charlesworth and Charlesworth, 2000; Lynch *et al*., 1993). In light of these considerations, we can expect Muller's ratchet to take place in populations subjected to high mutation rates and small effective population size, for example due to frequent bottlenecks. Changes in the selective landscape might also promote mutation accumulations in genes under weak selection. Asexual species affected by Muller's ratchet should display a higher ratio between non-synonymous and synonymous

nucleotide substitutions (dN/dS) than their sexual relatives. Low values of the dN/dS ratio are indicative of the presence of purifying selection, while high values are correlated with weak selection, genetic drift, and a high rate of mutation accumulation. In addition, the loss of the LLC drives the fixation of deleterious alleles, promoting the gradual erosion of genes to pseudogenes (Bergstrom and Pritchard, 1998; Charlesworth and Charlesworth, 1997; Higgs and Woodcock, 1995). Another typical signature of Muller's ratchet can be found in the pattern of neutral variation. In presence of tight linkage, strong selection at one locus eliminates variation at other loci. As a consequence, populations subjected to the ratchet are predicted to undergo a marked reduction in genetic diversity compared to a ratchet-free population (Gordo *et al*., 2002).

3. Muller's ratchet in experimental and natural populations

3.1. Laboratory experiments with microbes

Given their high mutation rates and quick replication, viruses provide an ideal system to study Muller's ratchet. Under experimental conditions that maximise genetic drift through frequent bottlenecks, both RNA (Andersson and Hughes, 1996; Chao, 1990; de la Iglesia and Elena, 2007; Duarte *et al*., 1992) and DNA viruses (Jaramillo *et al*., 2013) accumulate mutations at a higher rate than populations not subjected to bottlenecks; this process is also accompanied by a fitness decline (Andersson and Hughes, 1996; Chao, 1990; de la Iglesia and Elena, 2007; Duarte *et al.*, 1992; Jaramillo *et al.*, 2013). These experimental results are consistent with the prediction that small effective population size facilitates Muller's ratchet by increasing the magnitude of stochastic fluctuations between generations, increasing the probability of loss of the least-loaded class (Haigh, 1978) and fixation of deleterious alleles (Charlesworth and Charlesworth, 1997; Higgs and Woodcock,

1995). Under analogous conditions, Muller's ratchet has been shown to operate in yeast (Zeyl *et al*., 2001) as well as in several bacterial species, including *Salmonella typhimurium* (Andersson and Hughes, 1996) and *Escherichia coli* (Funchain *et al*., 2000; Tenaillon *et al*., 2016), reducing fitness and causing the progressive increase in mutation load. ~1500 generations are sufficient to introduce significant fitness loss in mutator strains of *E. coli*, with an average of 24-30 inactivated genes per cell and frequent loss of key metabolic pathways (Funchain *et al.*, 2000). In contrast, only 3% of wild-type lineages exhibited any type of detectable mutations after ~1000 generations (Funchain *et al.*, 2000). Similarly, lineages of *E. coli* subjected to stringent bottlenecks exhibit significant fitness loss compared to the ancestral lines (Tenaillon *et al.*, 2016). Frequent bottlenecks have been shown to drive genome reduction in *Salmonella enterica*, accompanied by the loss of metabolic and regulatory pathways (Nilsson *et al*., 2005).

Evidence of Muller's ratchet has also been observed in strains of *S. cerevisiae* with a high mutation rate, subjected to repeated bottlenecks of ~250 cells. After 175 transfers (~2900 generations), 2 of 6 experimental populations experienced a sharp decline in fitness, followed by extinction (Zeyl *et al.*, 2001). By contrast, wildtype strains with a normal mutation rate did not experience any fitness decline (Zeyl *et al.*, 2001). These experiments confirm that small effective population size causes non-recombining populations to undergo progressive loss of fitness because of the accumulation of deleterious mutations. In agreement with theoretical predictions, stringent bottlenecks maximise the severity of the ratchet, which is further aggravated by high mutation rates. But numerous experimental observations of Muller's ratchet are in organisms with a high mutation rate, either virus (Chao, 1990; de la Iglesia and Elena, 2007; Duarte *et al.*, 1992; Jaramillo *et al.*, 2013; Novella and Ebendick-Corpus, 2004) or mutator strains of *E. coli* and yeast (Funchain *et al.*, 2000;

Tenaillon *et al.*, 2016; Zeyl *et al.*, 2001). How often are these conditions met in nature? The following paragraphs discuss the evidence of Muller's ratchet in natural populations that have transitioned to an asexual lifestyle, including apomictic plants, parasites, endosymbionts, and organelles.

3.2. Muller's ratchet in plants

In eukaryotes, most instances of obligate asexuality are evolutionarily recent, indicating that mutation accumulation can lead to the premature extinction of asexual lineages (Bell, 1982; Butlin, 2002; Maynard Smith, 1978). In plants, apomixis (asexuality) is frequently observed among hybrids and/or polyploid species (Koltunow and Grossniklaus, 2003; Richards, 2003). Polyploidy and hybridisation mask the accumulation of deleterious mutation and make the transition to asexuality advantageous in the short term (Gerstein and Otto, 2009), but in the long run it makes apomictic lineages more vulnerable to Muller's ratchet. Apomictic lineages accumulate mutations faster than sexual ones and are more susceptible to the loss of beneficial alleles (Lovell *et al*., 2017). Occasional facultative sexual reproduction, however, can halt the ratchet and prevent the increase in mutation load (Hojsgaard and Hörandl, 2015). For example, the genome of *Ranunculus auricomus*, which possesses both sexual and facultative asexual lineages, does not show any signature of excess deleterious mutation accumulation in the asexual lineages (Pellino *et al*., 2013). One possible reason is that apomixis evolved relatively recently in *Ranunculus* (~70,000 years ago), and this timescale is too small for the effects of Muller's ratchet to be detectable. An alternative explanation is that facultative sexuality (~30% of sexually formed seeds) favours the elimination of deleterious mutations, preventing genome-wide mutation accumulation (Hodač *et al*., 2019; Hojsgaard and Hörandl, 2015). These results are in agreement with

theoretical models of Muller's ratchet in sexual populations. Bell (1988) suggested that a small amount of recombination is necessary to prevent Muller's ratchet; similar results were reproduced by Charlesworth *et al.* (1993), who showed that recombination, even at very low frequency, retards the progress of the ratchet in a random-mating, diploid population.

The transition to apomixis based upon vegetative reproduction can provide a short-term selective advantage by amplifying the fittest genotype, but in the long-term it exposes a population to the risk of mutational meltdown (Klekowski, 2003; Lovell *et al.*, 2017). In gymnosperms (non-flowering vascular plants such as gingko, cycads, and conifers), the severity of the ratchet is thought to be mitigated by cell-level selection arising from the competition between cells in a single meristem, the unit from which new plants arise (Klekowski *et al*., 1985). But this mechanism is less effective in angiosperms (flowering plants) and pteridophytes (non-seed forming vascular plants, e.g., ferns, horsetails and lycopods) due to the specific cellular organisation in apical meristems (Klekowski, 2003). Populations of the clonally reproducing ferns, such as *Matteuccia struthiopteris* or *Pteridium aquilinum*, sometimes display extremely high frequencies of recessive lethal mutations (Klekowski, 2003). Similarly, *Osmunda regalis*, the royal fern, has become endangered in Europe because of its extreme sensitivity to climatic stress, likely because of the presence of deleterious mutations (Zenkteler, 1999). Long-lived trees constitute another system of particular interest to the study of Muller's ratchet in plants. Although trees do not sequester a germline, they can live for thousands of years without showing any strong signature of genetic ageing (Burian *et al*., 2016). They achieve this by the early sequestering of axillary meristems, which grow out to form new branches and are generally maintained in a quiescent state as to minimise the mutational input (Burian *et al.*, 2016). In addition, by restricting the number of germ-cell divisions, numerous plant species are able to minimise

the transmission of deleterious mutations; this mechanism is analogous to the early segregation of the germline in animals (Lanfear, 2018).

3.3. Asexual populations and evolutionary scandals

Asexually reproducing taxa usually correspond to small phylogenetic branches distributed among sexual taxa, indicating short evolutionary lifetimes (Bell, 1982; Butlin, 2002). But it would be incorrect to assume that all asexual organisms are evolutionarily short-lived. The clade of bdelloid rotifers (Welch and Meselson, 2000) and the ostracod family Darwinulidae (Rossetti and Martens, 1998), seem to have persisted for millions of generations in the absence of sexual reproduction (Martens *et al*., 2003). Rather than undermining our understanding of ratchet dynamics, the existence of such 'evolutionary scandals' indicates the presence of alternative mechanisms to prevent the accumulation of deleterious mutations. For example, as I will discuss more extensively in the following paragraphs, bdelloid rotifers are thought to undergo lateral gene transfer (LGT) (Eyres *et al*., 2015; Gladyshev *et al*., 2008), which can limit the detrimental effects of the ratchet (Colnaghi *et al.*, 2020; Takeuchi *et al*., 2014).

The unisexual fish Amazon molly (*Poecilia formosa*) is the result of a hybridisation event that took place between 40,000 to 100,000 years ago (Schartl *et al*., 1995b). Loewe and Lamatsch (2008) have shown that, for a wide range of parameter values, the expected extinction time is much lower than the estimated age of this species. Paternal leakage of undamaged DNA from the sexual sister species has been suggested as a possible mechanism employed by *P. formosa* to prevent Muller's ratchet (Schartl *et al*., 1995a). Leakage of paternal DNA could load genes that have been lost because of the ratchet and restore the least-loaded genotype after it has disappeared because of stochastic fluctuations.

Alternatively, the accumulation of mutations could have been slowed down by low levels of mitotic recombination, which (like meiotic recombination) has the potential to diminish Muller's ratchet by generating variance and promoting purifying selection (Loewe and Lamatsch, 2008; Rasch *et al*., 1982). A similar process is thought to take place in asexual lineages of *Daphnia* (Omilian *et al*., 2006).

An alternative way to avoid the accumulation of deleterious mutations is through a reduction of the genome-wide mutation rate. This can be observed in asexual species like *Paramecium tetraurelia* (Sung *et al*., 2012) and the fairy-ring fungus *Marasmius oreades* (Hiltunen *et al*., 2019), which show a mutation rate of the order of 10^{-11} and 10^{-12} per bp per generation, respectively. The low evolutionary rate observed in Darwinulid ostracods (Schön *et al*., 2003) might also indicate a low genome-wide mutation rate. In bdelloid rotifers, the rate of synonymous substitutions is not significantly lower than in their sexual relatives (Welch and Meselson, 2001), and phylogenetic analysis indicates the rapid accumulation of deleterious mutations after the emergence of asexuality (Butlin, 2002; Welch and Meselson, 2001). The internal branch leading to bdelloid rotifers is significantly longer in phylogenetic trees based on non-synonymous than on synonymous substitution (Butlin, 2002; Welch and Meselson, 2001). A potential explanation provided by Butlin (2002) is that, shortly after the transition to asexuality, bdelloid rotifers accumulated mildly deleterious mutations (in agreement with Muller's hypothesis). What then prevented the mutational meltdown of bdelloidea?

Butlin (2002) suggested that the fixation of numerous mildly deleterious mutation could have potentially increased the fitness cost associated with every new mutation, leading to stronger purifying selection and preventing further clicks of the ratchet. This hypothesis is compatible with theoretical models of Muller's ratchet which include

synergistic epistasis, which enhances the deleterious effects of concomitantly occurring mutations (i.e., the higher the mutation load, the higher the fitness effect of each mutation) (Jain, 2008; Kondrashov, 1994). As discussed in Chapter 1, the introduction of synergistic epistasis is shown to reduce ratchet severity, but only in theoretical models where mutations are assumed to have the same fitness effect (Jain, 2008; Kondrashov, 1994). Synergistic epistasis does not retard Muller's ratchet if mutations are modelled as having a continuous distribution of fitness effects, which is likely the case in real populations (Butcher, 1995). In addition, the extent to which the assumption of synergistic epistasis is valid in real populations is a matter of debate, given the conflicting experimental evidence for synergistic epistasis (de Visser and Elena, 2007; Puchta *et al*., 2016).

Experimental evidence indicates that bdelloid rotifers may also exchange genes through lateral gene transfer (LGT), in ways similar to those used by bacteria to acquire new genetic material from their environment (Eyres *et al.*, 2015; Gladyshev *et al.*, 2008). While the presence of synergistic epistasis between deleterious mutations could potentially explain phylogenetic data, the evidence for lateral gene transfer in bdelloid rotifers offers a more plausible explanation for their long evolutionary lifespan (Eyres *et al.*, 2015; Gladyshev *et al.*, 2008). LGT likely takes place during the process of DSB repair associated with the cycle of desiccation and re-hydration experienced by bdelloid rotifers (Gladyshev and Meselson, 2008; Lapinski and Tunnacliffe, 2003), as species which undergo frequent desiccation experience higher rates of LGT (Eyres *et al.*, 2015). While most signature of LGT in bdelloid rotifers are from foreign (i.e. non-metazoan) genes, this process likely involves the homologous recombination of DNA fragments from related individuals (Eyres *et al.*, 2015; Gladyshev *et al.*, 2008), providing a way of reducing the severity of the ratchet. The presence of LGT makes bdelloid rotifers much less of an evolutionary scandal than a

confirmation of theoretical evidence that even a small degree of recombination can halt Muller's ratchet (Bell, 1988; Charlesworth *et al.*, 1993).

3.4. Endosymbiont parasite genome streamlining

Endosymbiont parasites constitute excellent systems to study Muller's ratchet in natural populations, as they often experience stringent bottlenecks, and many do not undergo systematic recombination. Moreover, these largely asexual parasites cannot escape the ratchet by decreasing their mutation rate, because they require higher mutation rate to respond sufficiently quickly to their hosts' adaptation. This phenomenon, known as the Red Queen hypothesis, can explain why most parasitic Protists tend to be sexual or to retain para-sexual mechanisms (Van Valen, 1977; Vergara *et al*., 2014).

Endosymbiotic parasites typically undergo genome streamlining, allowing them to resist more successfully to Muller's ratchet. After the transition to an obligate endosymbiotic lifestyle, many genes that were previously essential become redundant and are therefore subjected to weaker selection. In agreement with the predictions of Muller's ratchet theory, these genes (now under weak selection) progressively accumulate mutations and degrade to pseudogenes, before being eliminated through the fixation of deletions (Bobay and Ochman, 2017; Moran and Mira, 2001). As noted before, similar dynamics have been reproduced in experimental conditions, showing that frequent bottlenecking induce rapid genome size reduction in bacteria (Nilsson *et al.*, 2005). This reduction in genome size counters Muller's ratchet by limiting the genome-wide mutation rate of endosymbionts, without reducing their point mutation rate. In other words, it allows them to maintain adaptability and respond promptly to compensatory mutations in their host, while at the same time minimising the mutational target.

In light of these observations, it is not surprising that obligate symbionts typically display much smaller genomes than their free-living relatives (McCutcheon and Moran, 2012). For example, *Sulcia muelleri*, *Nasuia deltacephalinicola*, and *Candidatus Zinderia insecticola*, bacterial symbionts of sap-feeding insects, possess genomes as small as 191kb, 112kb, and 208kb, respectively (Bennett and Moran, 2013; McCutcheon and Moran, 2010). Other examples of symbionts which underwent massive genome streamlining include psyllid symbiont *Carsonella ruddii* (159bp) (Nakabachi *et al*., 2006); *Hodgkinia cicadicola*, found in cicadas (140kb) (McCutcheon *et al*., 2009); and *Candidatus* Tremblaya princeps, a symbiont of the citrus mealybug *Planococcus citri* (138kb) (López-Madrigal *et al*., 2011). These organisms all possess genomes smaller than 300kb, half to one-quarter of the genome of *Mycoplasma genitalium*, the free-living organism with the smallest genome (McCutcheon and Moran, 2012). Incidentally, this is comparable with the size of largest known viral genomes, both in terms of size and number of genes (Fischer *et al*., 2010; Raoult *et al*., 2004).

Analyses of 16S rDNA reveals that endosymbiotic bacteria exhibit a higher ratio between non-synonymous and synonymous nucleotide changes (dN/dS) than their free-living relatives, consistent with higher rates of mutation fixation (Lambert and Moran, 1998; Moran, 1996). Fungi-like *Microsporidia* parasites have the smallest known eukaryotic genome and possess extremely streamlined ribosomes, with an unusually high rate of fixed mutations – the signature of Muller's ratchet (Melnikov *et al*., 2018). Mycoplasma-related endobacteria (MRE), one of the most ancient lineages of endosymbionts, have been associated with ancient fungal lineages *Glomeromycota* and *Mucoromycotina* since early Palaeozoic. Thanks to their ability to undergo recombination, they have been able to maintain their genetic integrity in the face of Muller's ratchet (Naito and Pawlowska, 2016).

Nevertheless, MRE possess more streamlined genomes than their free-living relatives and exhibit a higher rate of fixation of slightly deleterious mutations, as inferred from the ratio between nonsynonymous and synonymous substitution (dN/dS) (Naito and Pawlowska, 2016). This supports the hypothesis that vertical transmission, because of frequent bottlenecking and strong genetic drift, exposes endosymbionts to faster rates of mutation accumulation than horizontally transmitted bacteria. By undergoing genome streamlining, MRE have been able maximise the effectiveness of recombination and survived Muller's ratchet over an extensive evolutionary period.

3.5. Organelles

Because of their direct impact on cell fitness and the possibility of nuclear compensation, mitochondria and chloroplast constitute a unique kind of endosymbionts. Similarly to other endosymbionts, however, their population structure is characterized by (relatively) small population size, and their uniparental inheritance without systematic recombination makes them vulnerable to the accumulation of deleterious mutations because of genetic drift. Unsurprisingly, signatures of Muller's ratchet, such as a 25-fold increase in nucleotide substitution between mammalian mitochondrial tRNA and nuclear tRNA, can be observed in the mtDNA of several animal taxa (Lynch, 1996). Both mammal and *Drosophila* mtDNA displays an excess of non-synonymous nucleotide polymorphisms compared to a neutral model, indicating the fixation of slightly deleterious mutations because of genetic drift (Ballard and Kreitman, 1994; Nachman *et al*., 1996; Rand and Kann, 1996). The signature of Muller's ratchet is also revealed by the analysis of mutations in *Caenorhabditis briggsae* mitochondrial genomes (Howe and Denver, 2008). Numerous lineages of *C. briggsae* display a high frequency of nonsynonymous substitutions in protein-coding genes and deletions,

whose proportion is negatively correlated with fecundity (Howe and Denver, 2008). A compensatory mutation was also observed, which restricts the number of mitochondria carrying deleterious deletions (Howe and Denver, 2008). This suggests that the negative consequences of Muller's ratchet in mitochondrial genomes can be partially offset by compensatory mutations. Other compensatory mechanisms (i.e., selective processes at the level of the individual, the cell, or the organelle) that prevent the build-up of deleterious mitochondrial mutations will be discussed in more detail in Chapter 4.

Despite a population structure which makes mitochondria vulnerable to the ratchet, mtDNA shows a strong signal of purifying selection (Fonseca *et al*., 2008) and evidence of adaptive change (James *et al.*, 2016). Moreover, selective processes are observed that curtail the transmission of severe mitochondrial mutations (Fan *et al.*, 2008; Hill *et al.*, 2014; Stewart *et al.*, 2008). This evidence indicates that (beyond the extreme shrinkage of mitochondrial genomes, which will be discussed in Chapter 5) mechanisms have evolved that limit the transmission of severe mitochondrial mutations and prevent the mutational meltdown of organellar genomes. Without such mechanisms for purifying selection, mtDNA would be extremely susceptible to the accumulation of deleterious mutation, potentially causing the extinction of animal species (Loewe, 2006; Loewe and Lamatsch, 2008). In Chapter 4, I discuss how the interplay between different levels of selection (individual, cell, and organelle) prevents the build-up of mitochondrial mutations and facilitates their elimination, preventing the mutational decay of mitochondrial genomes. I analyse these dynamics with the use of a mathematical model, showing that organelle-level selection is necessary to prevent Muller's ratchet in mitochondrial genomes. My modelling work clarifies the role of enigmatic germline processes whose purpose has been, so far, poorly understood.

4. Conclusion

The need to prevent the accumulation of deleterious mutation is a major evolutionary force, likely responsible for the origin of sexual reproduction and the evolution of prokaryotic genomes. As predicted by Muller (1964), the lack of recombination exposes asexual populations to the progressive accumulation of deleterious mutations. This effect has been investigated by numerous theoretical studies, which predict increasingly high mutation accumulation rates in asexual populations subjected to strong genetic drift. While the exact speed of the ratchet is hard to evaluate with precision, a high mutation rate and small population size increase the rate of mutation accumulation and fixation, leading to a high probability of extinction through mutational meltdown. Under these conditions, asexual populations are predicted to undergo a progressive loss of genetic information, as revealed by higher dN/dS rates, lower genetic diversity, genome shrinkage in endosymbionts, and higher frequency of pseudogenes compared to recombining populations. The majority of theoretical models discussed throughout this chapter follow Haigh (1978) in assuming the presence of an infinite number of loci, the absence of synergistic epistasis, and a constant strength of selection across the whole genome. In the following chapters, I will relax some of these restrictive assumptions, investigating how finite genome size and variable strength of selection across the genome impact the dynamics of accumulation and fixation of deleterious mutations.

Theoretical predictions of Muller's ratchet models are consistent with observations in natural populations subjected to strong genetic drift due to small population size and/or stringent bottlenecks, such as endosymbionts and parasites. The theoretical model that will be presented in Chapters 2 and 3 indicates that a possible strategy to avert Muller's ratchet

is to reduce the overall mutation rate by genome streamlining, a process observed in asexual parasites and endosymbionts, including organelles. As an alternative, low rates of recombination (either via sporadic sexual reproduction or via lateral gene transfer) are generally able to prevent a mutational meltdown.

As most theoretical models assume an infinite number of loci, the relationship between genome size and the minimum rate of recombination necessary to prevent Muller's ratchet is still poorly understood. Moreover, only a few theoretical studies have analysed the impact of recombination via lateral gene transfer (LGT) on mutation accumulation in prokaryotes (Levin and Cornejo, 2009; Takeuchi *et al.*, 2014; Wylie *et al.*, 2010). None of these considered the role of genome size, recombination length, and ectopic recombination caused by the presence of genetic repeats. In chapters 2 and 3, I discuss in more depth the relationship between genome size and ratchet dynamics in the presence of LGT. Using a theoretical model and numerical simulations, I show that genome size is inversely correlated with LGT's ability to eliminate deleterious mutations and, consequently, with the time between clicks of the ratchet. Following this theoretical analysis, I discuss the role that an increase in Muller's ratchet severity due to genome size expansion and the increase of repeat sequences played in the evolution of eukaryotes and the origin of meiotic sex. In the special case of mitochondrial DNA, the great impact of mutations on individual and cell fitness gives other levels of selection (cell and individual) the potential to affect ratchet dynamics (Rand, 2001). In Chapter 4, I discuss how the need for purifying selection in mitochondrial genomes is met through complicated developmental processes in the female germline and develop a theoretical model to study the interplay between different levels of selection.

The results presented in the following chapters are but a reminder of the grandeur of evolution. Solving the inevitable problem of deleterious mutations was the catalyst of adaptation and evolutionary change, a creative impulse that led to the evolution of sex and, with it, to the vibrancy, complexity, and beauty of the living world we see around us – none of which could exist without sexual reproduction.

Chapter 2 – Genome Expansion in Early Eukaryotes drove the Transition from Lateral Gene Transfer to Meiotic Sex[1]

1. Summary

Prokaryotes generally reproduce clonally but can also acquire new genetic material via lateral gene transfer (LGT). Recombination of environmental DNA can prevent the accumulation of deleterious mutations predicted by Muller's ratchet for asexual populations, but LGT was abandoned by ancestral eukaryotes in favor of sexual reproduction. Understanding the limitations of LGT provides insight into this evolutionary transition. Here I develop a theoretical model of a haploid population undergoing LGT which includes two new parameters, genome size and recombination length, neglected by previous theoretical models. The greater complexity of eukaryotes is linked with larger genomes, and I demonstrate that the benefit of LGT declines rapidly with genome size. The degeneration of larger genomes can only be resisted by increases in recombination length to the same order as genome size, as occurs in meiosis. These results can explain the strong selective pressure towards the evolution of sexual cell fusion and reciprocal recombination during early eukaryotic evolution – the origin of meiotic sex.

2. Introduction

Understanding the origin and maintenance of sex in the face of multiple costs was long considered the 'Queen of problems in evolutionary biology' (Bell, 1982). Sexual reproduction breaks up advantageous combinations of alleles, halves the number of genes

[1]A revised version of this chapter was published as Colnaghi, M., Lane, N., and Pomiankowski, A. (2020). Genome expansion in early eukaryotes drove the transition from lateral gene transfer to meiotic sex. Elife *9*, e58873.

transmitted to the offspring, and is less efficient and energetically more costly than asexual reproduction (Bell, 1982; Otto, 2009; Otto and Lenormand, 2002). In spite of these disadvantages sex is a universal feature of eukaryotic life. The presence of common molecular machinery, widespread among all eukaryotic lineages, is a strong indication that the Last Eukaryotic Common Ancestor (LECA) was already a fully sexual organism (Schurko and Logsdon, 2008; Speijer *et al.*, 2015). Meiotic genes are commonly found in putative asexual eukaryotes, including Amoebozoa (Hofstatter *et al.*, 2018; Lahr *et al.*, 2011), Diplomonads (Ramesh *et al.*, 2005), Choanoflagellates (Carr *et al.*, 2010) and even early diverging lineages such as *Trichonomas vaginalis* (Malik *et al.*, 2008). Eukaryotic asexuality is not ancestral but a secondarily evolved state. The selective pressures that gave rise to the origin of meiotic sex must therefore be understood in the context of early eukaryotic evolution.

Phylogenomic analysis shows that eukaryotes arose from the endosymbiosis between an archaeal host and a bacterial endosymbiont, the ancestor of mitochondria (Martin *et al.*, 2015; Williams *et al.*, 2013; Zaremba-Niedzwiedzka *et al.*, 2017). The presence of energy-producing endosymbionts allowed the first eukaryotes to escape the bioenergetic constraints that limit the genome size and cellular complexity of prokaryotes (Lane, 2014; 2020). Extra energetic availability came with the evolutionary challenge of the coexistence of two different genomes within the same organism. As with other endosymbioses, the symbiont genome underwent a massive reduction, with the loss of many redundant gene functions (López-Madrigal and Gil, 2017; Timmis *et al.*, 2004). Alongside this, symbiont release of DNA into the host's cytosol caused the repeated transfer of genes to the host genome, many of which were retained, contributing to the massive genome size expansion

during early eukaryotic evolution (Lane, 2011; Martin and Koonin, 2006; Timmis *et al.*, 2004).

Both the host and the endosymbiont, like modern archaea and proteobacteria, are likely to have been capable of transformation – the uptake of exogenous DNA from the environment followed by homologous recombination (Ambur *et al.*, 2016; Bernstein and Bernstein, 2013; Vos *et al.*, 2015). This process involves the acquisition of foreign DNA, the recognition of homologous sequences and recombination, and therefore presents striking similarities with meiosis in Eukaryotes. The *Rad51*/*Dcm1* gene family, which plays a central role in meiosis, has high protein sequence similarity with *RecA*, which promotes homologous search and recombination in prokaryotes (Johnston *et al.*, 2014; Lin *et al.*, 2006). It has been suggested that *RecA* was acquired by the archaeal ancestor of eukaryotes via endosymbiosis from its bacterial endosymbiont (Lin *et al.*, 2006). Alternatively, the *Rad51*/*Dcm1* family could have evolved from archaeal homologs of *RadA* (Seitz *et al.*, 1998). Regardless, the presence of this common molecular machinery and the striking similarities between these processes suggest that meiosis evolved from bacterial transformation (Bernstein and Bernstein, 2013; Schurko and Logsdon, 2008). But the selective pressures that determined this transition are still poorly understood.

Historically, the main focus of the literature on the origin and the maintenance of sex has been the comparison of sexual and clonal populations, or the spread of modifiers that increase the frequency of recombination (Bell, 1982; Otto, 2009). Recombination can eliminate the linkage between beneficial and deleterious alleles due to Hill-Robertson effects (Barton and Otto, 2005; Felsenstein and Yokoyama, 1976), increase adaptability in rapidly changing (Gandon and Otto, 2007; Hamilton, 1980; Jokela *et al.*, 2009) or spatially heterogeneous (Lenormand and Otto, 2000; Pylkov *et al.*, 1998) environments, and prevent

the accumulation of deleterious mutations predicted by Muller's ratchet for asexual populations (Haigh, 1978; Muller, 1964). These benefits of recombination outweigh the multiple costs of sexual reproduction and explain the rarity of asexual eukaryotes (Otto, 2009). But remarkably, they provide us with virtually no understanding of why bacterial transformation was abandoned in favour of reciprocal meiotic recombination. The real question is not why sex is better than clonal reproduction, but why did meiotic sex evolve from prokaryotic transformation?

Lateral Gene Transfer (LGT) has been recognised as a major force shaping prokaryotic genomes (Lapierre and Gogarten, 2009; Ochman *et al.*, 2000; Vos *et al.*, 2015). Unlike conjugation and transduction, mediated respectively by plasmids and phages, transformation is the only LGT mechanism to be exclusively encoded by genes present on prokaryotic chromosomes (Ambur *et al.*, 2016), and is maintained by natural selection because it provides benefits analogous to those of sexual recombination (Levin and Cornejo, 2009; Takeuchi *et al.*, 2014; Vos *et al.*, 2015; Wylie *et al.*, 2010). Recombination via transformation favours adaptation by breaking down disadvantageous combinations of alleles (Levin and Cornejo, 2009; Wylie *et al.*, 2010) and preventing the fixation of deleterious mutations (Levin and Cornejo, 2009; Takeuchi *et al.*, 2014). Some theoretical studies (Redfield, 1988; Redfield *et al.*, 1997) suggest that transformation is only advantageous in presence of strong positive epistasis, a condition rarely met by extant prokaryotes. But more recent modelling work shows that transformation facilitates the elimination of deleterious mutations and prevents Muller's ratchet, even when the genes acquired from the environment have a slightly higher mutation load than the population undergoing LGT (Levin and Cornejo, 2009; Takeuchi *et al.*, 2014). As transformation provides similar advantages as meiotic sex, why did the first eukaryotes forsake one for the other?

How did the unique conditions at the origin of eukaryotic life give rise to the selective pressures that determined this transition? In particular, is it possible that the massive expansion in genome size in early eukaryotes created the conditions for the evolution of a more systematic way of achieving recombination?

Very little is known about the relation between genome size and the accumulation of deleterious mutations in populations undergoing transformation, as previous models either do not consider it explicitly (Levin and Cornejo, 2009; Wylie *et al.*, 2010) or treat it as a constant parameter (Takeuchi *et al.*, 2014). In order to evaluate the impact of genome size and recombination rate on the dynamics of accumulation of mutation, I develop a new theoretical and computational model of Muller's ratchet in a population of haploid individuals, undergoing homologous recombination of genetic material via transformation. For simplicity, I refer to this process simply as 'LGT'. I do not consider genetic exchange facilitated by plasmids, phages, or other selfish genetic elements (Lapierre and Gogarten, 2009; Ochman *et al.*, 2000; Vos *et al.*, 2015).

The population is subjected to variable rates of mutation and recombination due to LGT. The model includes two new parameters, genome size and recombination length, which have not been taken into account by previous theoretical studies (Levin and Cornejo, 2009; Takeuchi *et al.*, 2014; Wylie *et al.*, 2010). I evaluate the effects of different selective landscapes, either uniform across the genome, or split between core and accessory genomes. The severity of the ratchet is evaluated using standard approaches for measuring the rate of fixation of deleterious mutations and the expected extinction time of the fittest class (Gordo and Charlesworth, 2000a; Haigh, 1978; Takeuchi *et al.*, 2014). I suggest that systematic recombination across the entire bacterial genomes was a necessary

development to preserve the integrity of the larger genomes that arose with the emergence of eukaryotes, giving a compelling explanation for the origin of meiotic sex.

3. Methods

I use a Fisher-Wright process with discrete generations to model the evolution of a population of N haploid individuals, subject to a rate of deleterious mutation μ per locus per generation, with LGT at a rate λ. The genome of an individual j is described by a state vector $\vec{z}^{(j)} = (z_1, \dots, z_g)$, where g is the number of loci. Each locus i can accumulate a number of mutations (0, 1, 2,…). The components $z_i^{(j)}$ are the number of deleterious mutations at the i-th locus of the j-th individual. This allows me to keep track of the number of mutants in an individual and the distribution of mutations at each locus in the population. I define fixation of a mutant at a locus when the least-loaded class (LLC) at that locus is lost. As I neglect back-mutation, fixation of a mutant is permanent.

Table 2.1 Parameters and variables

N	*population size*
μ	*mutation rate per locus per generation*
g	*genome size (number of loci)*
U	*genome-wide mutation rate*
s	*strength of selection against deleterious mutations*
λ	*LGT rate*
L	*recombination length (number of loci)*

I consider two different mutational regimes: constant mutation rate per locus and constant genome-wide mutation rate. In the first scenario, the genome-wide mutation rate $U = \mu g$ is calculated as the product between the mutation rate per locus per generation and the number of loci (I assume that the mutation rate per locus is constant across the whole genome). Under this assumption, the severity of the ratchet increases exponentially with genome size, and this effect is can only partially be offset by an increase in population size (SI; **Figure 2.S1**). I also investigate the effect of a constant genome-wide mutation rate U for genomes of different size. This allows me to study the intrinsic limitations of LGT and distinguish them from the increase in the ratchet severity associated with larger g in the presence of a constant mutation rate per bp (see **Appendix A** and **Figure 2.S1**).

I introduce a new parameter *L*, the number of contiguous genes acquired with each LGT recombination event (i.e., the size of imported DNA), which has not been considered by previous theoretical studies (Levin and Cornejo, 2009; Takeuchi *et al.*, 2014; Wylie *et al.*, 2010). In order to avoid unnecessary complexity, I ignore the probability of ectopic recombination (whose effects will be analysed in great detail in the following chapter), and assume that DNA strands present in the environment (eDNA pool) are only stable for one generation before decaying irreversibly.

In the first part of this study, I assume that all mutations are mildly deleterious. Each mutation at a locus and across loci causes the same decrease in individual fitness s. Following previous studies of Muller's ratchet (Gordo and Charlesworth, 2000a; Haigh, 1978; Takeuchi *et al.*, 2014), I choose a multiplicative function to model the fitness of an individual carrying m mutations, given by the formula $w_m = (1 - s)^m$ (i.e., no epistasis). In the second part of the study, I investigate more complex distributions of strength of selection across the genome. In particular, I differentiate between a strongly selected core

genome and accessory genome under weaker selection. Which genes belong to the core and to the accessory genome is determined by random sampling. The fitness of an individual that carries m_i mutations at locus i is given by $w(t) = \prod_{i=1}^{g}(1 - s_i)^{m_i}$, where $s_i = 0.005$ if locus i belongs to the core genome and $s_i = 0.001$ otherwise.

Each generation, the life history of the population follows the following processes (**Figure 2.1**). The new generation is obtained by sampling N individuals at random, with replacement, from the old population. The probability of reproduction is proportional to the individual fitness w_m. Each individual acquires $n^{(j)}$ new deleterious mutations, where $n^{(j)}$ is a random variable drawn from a Poisson distribution with mean U. Then the old generation dies, and their DNA forms the genetic pool from which the new generation acquires exogenous DNA (eDNA) for recombination. The number of mutations and their position in the genome are randomly determined. Each individual has a probability λ of undergoing LGT. For each individual that undergoes LGT, a random donor is selected from the previous generation and L contiguous loci are randomly selected from its genome. The genome is assumed to be circular, so locus g is contiguous with locus 1. The corresponding components of the state vector of the recipient become equal to those of the donor. This can lead both to an increase or a decrease in the mutation load of the recipient.

The model described above has been implemented using MATLAB. Random sampling is achieved using the function 'randsample'. All random processes described above have been simulated using MATLAB's functions 'rand' and 'random'. The code used for the simulations has been uploaded on GitHub (https://github.com/MarcoColnaghi1990/Colnaghi-Lane-Pomiankowski-2020).

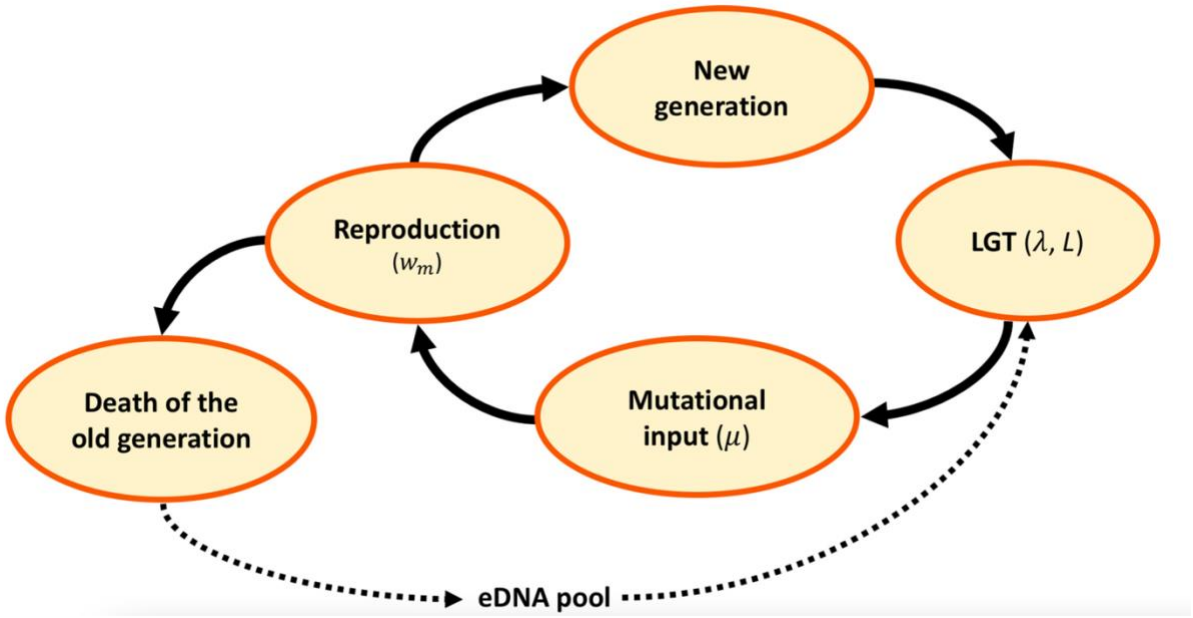

Figure 2.1 | Model dynamics. After the birth of a new generation, eDNA is acquired from the environment and randomly recombined with recombination length L, at a rate λ per genome. Following LGT, new mutations are randomly introduced at a rate μ per locus. A new generation is then sampled at random, in proportion to reproductive fitness w_m. The old generation dies and its DNA is released, constituting the eDNA pool for the new generation.

Simulations are started from a population free of mutation and run for 10,000 generations, with 50 replicates for a given set of parameter values. Two measures, time of extinction of the least-loaded class (LLC) (T_{ext}) and the time of irreversible[2] loss of the LLC ($\Delta m/\Delta t$), have been used to assess the effect of the ratchet (Gordo and Charlesworth, 2000a; Haigh, 1978; Takeuchi *et al.*, 2014). After recombination, I calculate the number of individuals in the least-loaded class (LLC), the set of individuals whose genomes are mutation free: $\vec{z}^{(j)} = (0,0,\ldots,0)$. When a mutant reaches fixation at a particular locus, the mutation-free class is irreversibly lost, and I mark this as T_{ext} (extinction time of the LLC).

[2] In populations undergoing LGT, the loss of the least-loaded class can be reversed by the acquisition of environmental DNA, unless a mutant reaches fixation at a particular locus and the wild-type allele is not present anymore in the environmental gene pool.

T_{ext} gives an estimate of the time that a population can remain free of mutations. The second measure is the genome-wide rate of fixation $\Delta m/\Delta t$, which is calculated as the ratio between the total number of fixed mutations at the end of the simulation, and the duration of the simulation itself (i.e., 10,000 generations). The rate of fixation per single locus is the ratio between the genome-wide rate of fixation and genome size g. $\Delta m/\Delta t$ is a measure of the rate of accumulation of mutations.

4. Results

4.1. Constant mutation rate per locus

Genome size increases the severity of the ratchet, measured by T_{ext}, the expected extinction time of the LLC (**Figure 2.2**). Large genomes gain *de novo* mutations at a faster rate than small ones, leading to a decline in LLC extinction time, as there are more independent loci that can possibly fix for the mutant (**Figure 2.2**, no LGT). LGT reduces the severity of the ratchet and increases the expected LLC extinction time (**Figure 2.2**), making the population less vulnerable to stochastic fluctuations in population size. The beneficial effect is more evident as recombination length (L) and LGT rate (λ) increase (**Figure 2.2**). However, as genome size (g) increases the expected extinction time plummets, rapidly approaching that of a clonal population with a larger genome, both in the presence of high ($\lambda = 0.1$, **Figure 2.2A**) and low ($\lambda = 0.01$, Figure 2.2B) LGT rates. The sole exception is when recombination length is of the same order as the magnitude of genome size ($L = 0.2g$, **Figure 2.2**). Only under this condition can increases in genome size be tolerated without a drastic decline in T_{ext}.

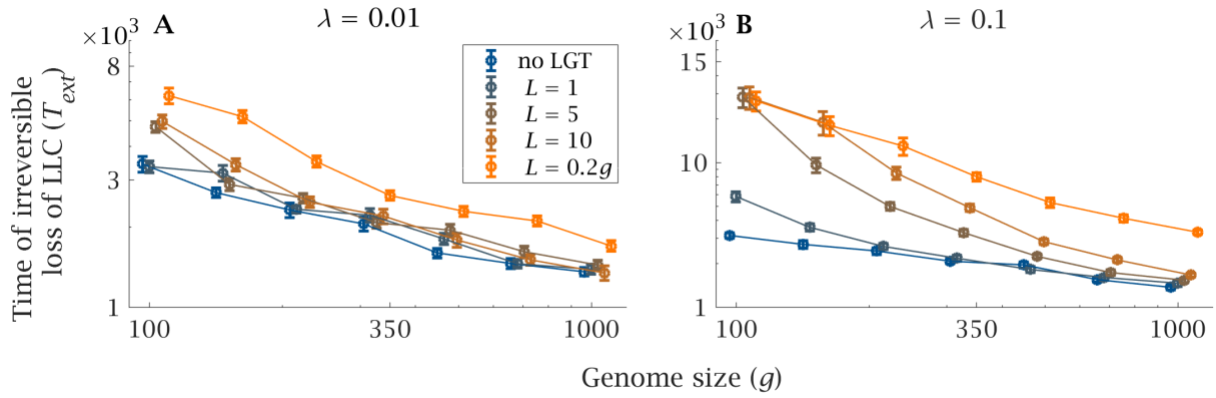

Figure 2.2 | Impact of LGT and genome size on the ratchet. The mean extinction time (generations) of the Least-Loaded Class (T_{ext}) is shown as a function of genome size (g) for various recombination lengths (L), in the presence of (**A**) low ($\lambda = 0.01$) and (**B**) high ($\lambda = 0.1$) LGT rates. The blue lines show the extinction time when there is no LGT, and is the same in (**A**) and (**B**). Parameters: $s = 10^{-3}$, $N = 5 \times 10^3$, $\mu = 10^{-4}$, $U = \mu \times g$. Data points and error bars show, respectively, the average and the standard deviation over 50 independent simulations.

In a small genome, LGT reduces the speed at which mutations accumulate in a population, counteracting the ratchet effect both in the presence of a high ($\lambda = 0.1$) or low ($\lambda = 0.01$) LGT rate (Figure 2.3). This effect is more pronounced with higher LGT rates (λ) and longer recombination lengths (L). But even in presence of LGT, large genomes are subjected to higher rates of accumulation, comparable to those of a purely clonal population (**Figure 2.3**). Only when recombination length approaches the same order of magnitude as genome size ($L = 0.2g$) and occurs at high frequency ($\lambda = 0.1$) can LGT sufficiently repress mutation accumulation in large genomes (**Figure 2.3B and 4D**).

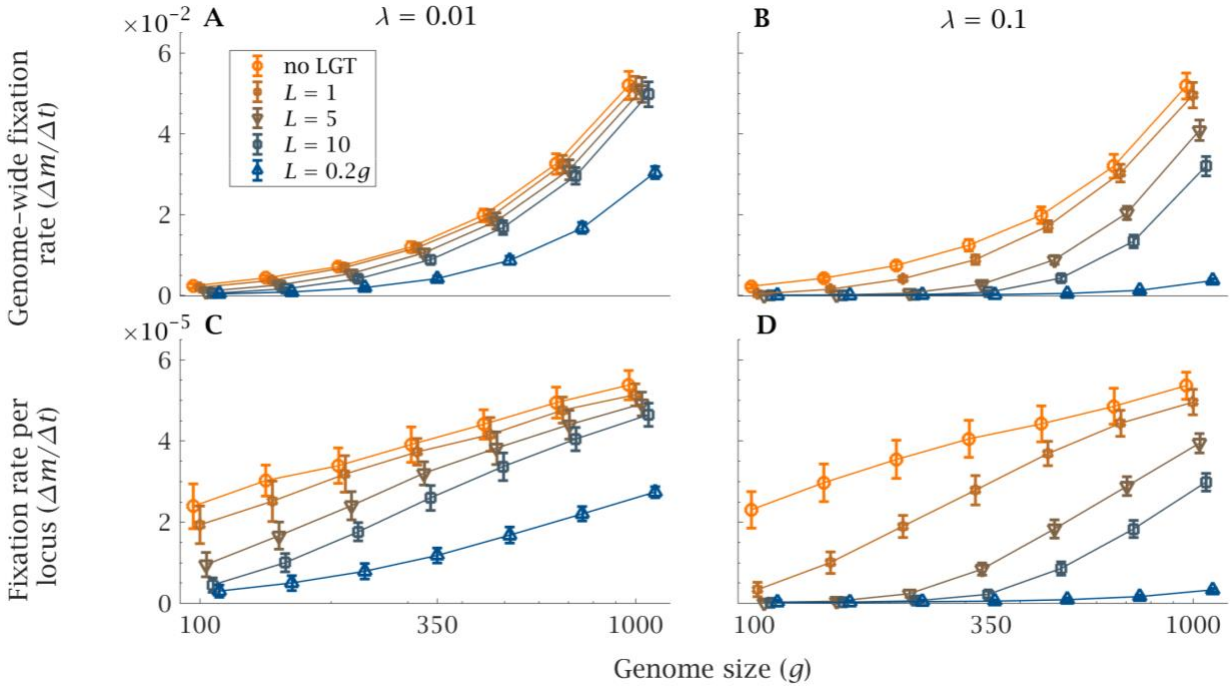

Figure 2.3 | Impact of LGT and genome size on the rate of accumulation of mutation. The mean genome-wide rate of fixation of deleterious mutations per generation, calculated over a time interval $t = 10^5$ generations, as a function of genome size (g) for various recombination lengths (L). This is shown for (**A**) low ($\lambda = 0.01$) and (**B**) high ($\lambda = 0.1$) LGT rates. Similarly, the rate of fixation of deleterious mutation per locus per generation is shown, again for (**C**) low ($\lambda = 0.01$) and (**D**) high ($\lambda = 0.1$) LGT rates. As genome size increases, LGT becomes less effective in reducing the mutational burden of a population. An increase in recombination length improves the efficiency of LGT in preventing the accumulation of mutations, but this beneficial effect declines rapidly with genome size. Only if recombination length is of the same order of magnitude as genome size ($L = 0.2g$) and the rate of LGT is high ($\lambda = 10^{-1}$) can large genomes be maintained in a mutation-free state. Parameters: $s = 10^{-3}$, $N = 10^4$, $\mu = 10^{-4}$, $U = \mu \times g$. Data points and error bars show, respectively, the average and the standard deviation over 50 independent simulations.

4.2. Constant genome-wide mutation rate

The limitations of LGT in large genomes become evident in simulations with a constant genome-wide mutation rate in which I increase genome size but constrain the genome-wide mutation rate (*U*) to a constant value. In this case, changes in the efficacy of LGT cannot be due to changes in the rate at which new mutations arise in a genome.

With a low LGT rate ($\lambda = 0.01$), recombination of short fragments ($L = 1$) is not enough to halt the mutational advance and provides virtually no benefits compared to clonal reproduction, with either a low (**Figure 2.4A**) or high genome-wide mutation rate (**Figure 2.4B**). With a high LGT rate ($\lambda = 0.1$, **Figure 2.4B and D**), larger values of L (i.e., $L \geq 5$) can drastically reduce the rate of accumulation of mutations in small genomes, but are not able to counter Muller's ratchet in larger genomes. As in the previous section, this effect is less evident with a low LGT rate (**Figure 2.4A** and **4C**). However, when the recombination length approaches the same order of magnitude as genome size ($L = 0.2g$), the accumulation of mutations is retarded regardless of genome size, both in the presence of a low ($U = 0.01$, **Figure 2.4A** and **B**) and high genome-wide mutation rates ($U = 0.1$, **Figure 2.4C and D**). In presence of a high LGT rate ($\lambda = 0.1$), recombination of such long fragments ($L = 0.2g$) effectively prevents any increase in mutation load (**Figure 2.4C** and **4D**). It is even possible to see a decline in mutation accumulation as genome size increases (**Figure 2.4D**). But this is for the artefactual reason that for a fixed genome wide mutation rate, the *per gene* mutation rate declines with g. This results in a net benefit as genome size increases under the condition that $L = 0.2g$ where the recombination length remains a constant fraction of g. Even though unrealistic, constraining the genome-wide mutation rate reveals that LGT becomes less efficient *per se* as genome size increases.

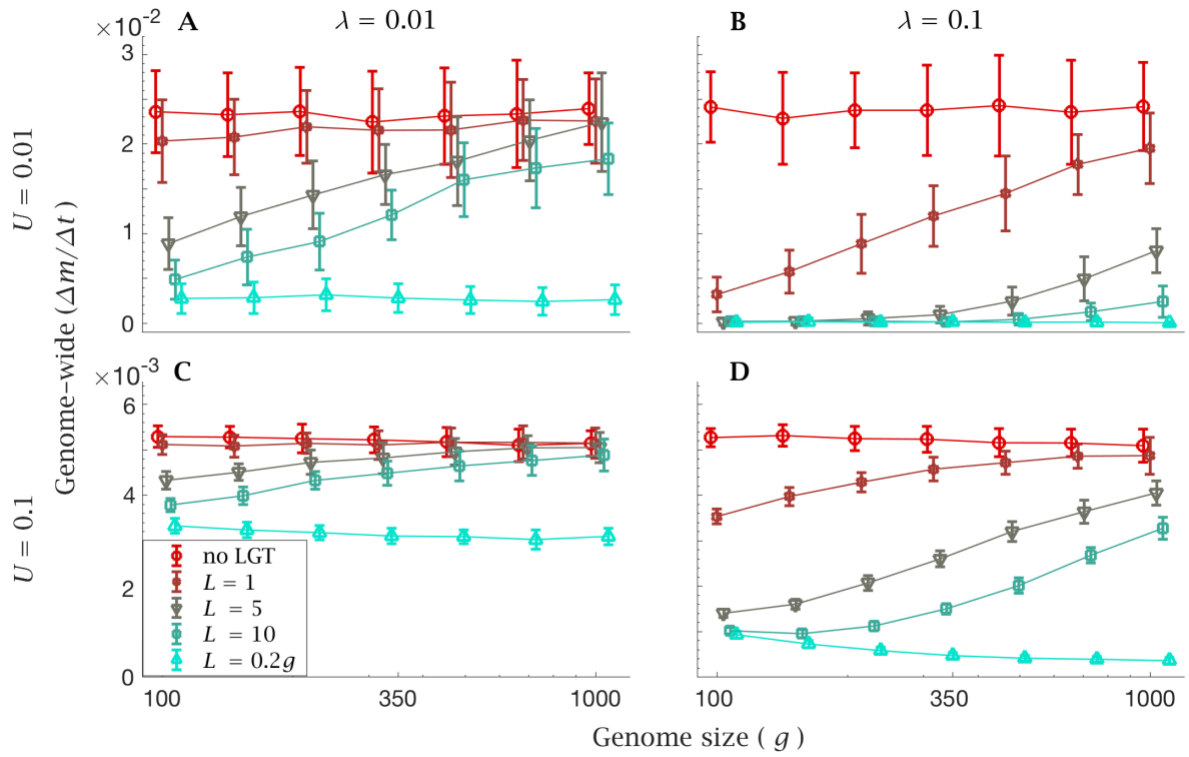

Figure 2.4 | Rate of fixation of mutations with constant genome-wide mutation rate. The mean genome-wide rate of fixation of deleterious mutations per generation. This is shown for a low rate genome-wide mutation rate $U = 0.01$ for (**A**) low ($\lambda = 0.01$) and (**B**) high ($\lambda = 0.1$) LGT rates, and a high genome-wide mutation rate $U = 0.1$ for (**C**) low ($\lambda = 0.01$) and (**D**) high ($\lambda = 0.1$) LGT rates, across a range of recombination lengths (L). The fixation rate was calculated over a time interval of $t = 10^5$ generations, as a function of genome size. LGT is not able to prevent the accumulation of mutation in large genomes, except for a recombination length on the order of the whole genome ($L = 0.2g$). Parameters: $s = 10^{-3}$, $N = 5 \times 10^3$. Data points and error bars show, respectively, the average and the standard deviation over 50 independent simulations. The different data points have been slightly off-set in order to prevent overlap between errorbars.

4.3. Non-uniform strength of selection

Different loci in the genome are typically under different strengths of selection. In order to capture this fact in the model, I consider the core and accessory genomes differently. The size of the core genome is fixed ($g_c = 50$), while the accessory genome size varies as the genome expands. The core loci are under strong selection ($s = 0.005$) and the accessory

loci are under weak selection ($s = 0.001$). Core and accessory loci are randomly distributed in the genome. Under this selection regime, mutations preferentially accumulate in the accessory genome, where the strength of selection is lower, while the core genome accumulates mutations at a relatively slow rate (**Figure 2.5**).

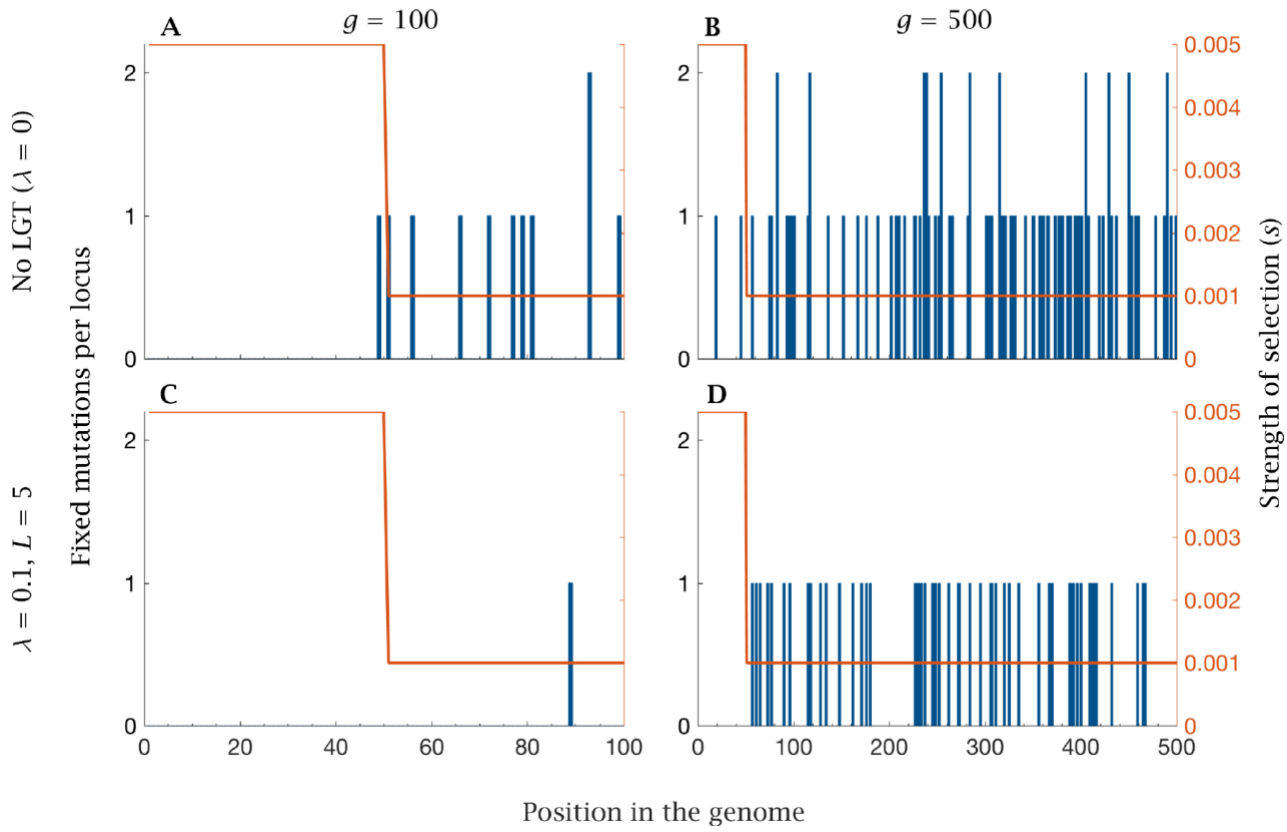

Figure 2.5 | Fixation of mutations in the core and accessory genome. Fixed mutations in the core and accessory genome after $t = 10^5$ generations for no LGT with (**A**) small ($g = 100$) and (**B**) intermediate ($g = 500$) genome size, and for LGT ($L = 5, \lambda = 0.1$) with (**C**) small ($g = 100$) and (**D**) intermediate ($g = 500$) genome size. Mutations preferentially accumulate in the accessory genome under weaker selection ($s = 0.001$), while the strongly selected core genome ($s = 0.005$) accumulates few or no mutations. The rate of fixation increases with genome size, while the benefits of LGT decline with genome size. Parameters: $N = 10^4$, $\mu = 10^{-4}$, $U = \mu \times g$.

Genome size expansion results in more severe ratchet effects, with a marked increase in the rate of mutations reaching fixation in the regions of the genome that are under weaker selection, alongside a moderate increase in core genome mutation fixation rate (**Figure 2.6**). LGT is effective in reducing the mutational burden, both in the accessory and in the core genome; but this beneficial effect is less evident in large genomes than in small ones (**Figure 2.6**). Recombination across the whole genome ($L = 0.2g$) completely eliminates fixation in the core genome, regardless of genome size, and markedly reduces the fixation rate in the accessory genome, facilitating genomic expansion (**Figure 2.6**).

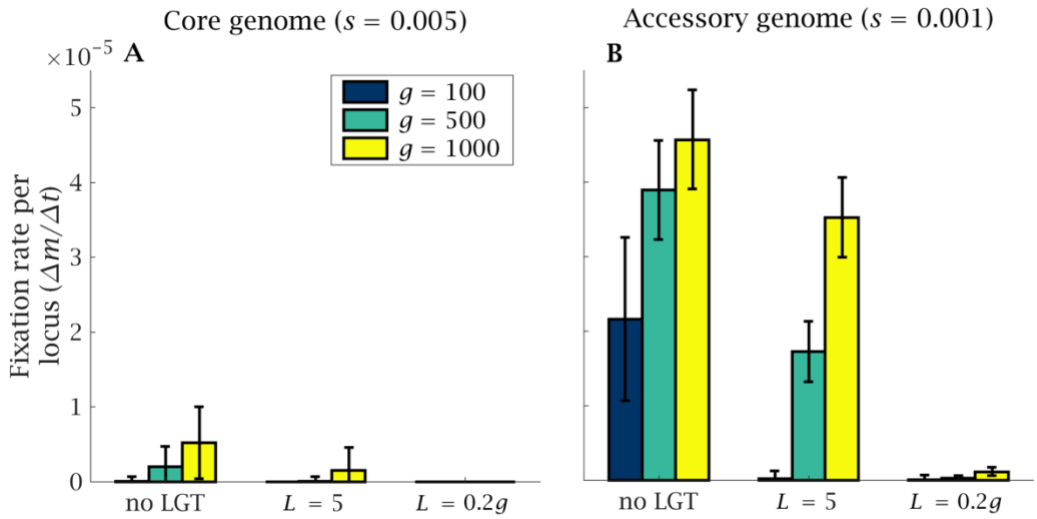

Figure 2.6 | Rate of fixation of mutations in the core and accessory genome. The fixation of mutants in the (**A**) core and in the (**B**) accessory genome is shown after $t = 10^5$ generations, normalised by genome size. A higher number of mutations accumulate in the accessory genome that is under weaker selection. Genome size expansion increases the severity of the ratchet and the number of fixed mutations in the core and accessory genome. The introduction of LGT considerably reduces the mutational burden. Parameters: $\lambda = 10^{-1}, \mathrm{L} = 5, N = 10^4, \mu = 10^{-4}, U = \mu \times g$, $s_{core} = 0.005$ and $s_{acc} = 0.001$. Bars and error bars show, respectively, the average and the standard deviation over 50 independent simulations.

5. Discussion

Asexual organisms are well known to be vulnerable to the effects of drift, which reduces the genetic variation within a population, causing the progressive and inescapable accumulation of deleterious mutations known as Muller's ratchet (Haigh, 1978; Muller, 1964; Otto, 2009). In eukaryotes, sexual recombination counters the effects of genetic drift and restores genetic variance, increasing the effectiveness of purifying selection and preventing mutational meltdown (Otto, 2009). In prokaryotes, sexual fusion does not occur. But the exchange of genetic material does occur through transformation, the lateral gene transfer (LGT) and recombination of environmental DNA (eDNA). Meiotic recombination likely arose from bacterial transformation. Understanding the reasons why this transition occurred during early eukaryotic evolution are critical to a rigorous understanding of the Queen of problems in evolutionary biology, the origin of sex. Sex did not arise from cloning, as tacitly assumed in the classic theoretical literature, but from prokaryotic transformation. To elucidate this transition, I examined the effectiveness of LGT at countering the dynamics of Muller's ratchet, to understand where and why LGT becomes ineffective at maintaining genome integrity, necessitating the transition to sexual reproduction in early eukaryotes.

I assessed the effect of LGT on the severity of the ratchet using the expected extinction time of the least-loaded class and the rate of fixation of deleterious mutations. Unlike previous modelling work, I included genome size as a model parameter in order to investigate the impact of genome size, as opposed to keeping it fixed (e.g. 100 loci (Takeuchi *et al.*, 2014)). Genome size is plainly important in relation to the evolution of eukaryotes, which have expanded considerably in almost every measure of genome size (e.g. DNA content, number of protein-coding genes, size of genes, number of gene families, regulatory DNA content; Lane and Martin, 2010}. Considering gene number in the model reveals a

strong inverse relationship between genome size and the benefits of LGT. In small genomes, LGT is effective at preventing Muller's ratchet, with long extinction times (**Figure 2.2**) and low rates of mutation accumulation (**Figure 2.4**). These results mirror those from modelling assuming a single genome size ($g = 100$), which show that LGT can halt Muller's ratchet even if the environmental DNA has a higher mutation load than the population (Takeuchi *et al.*, 2014). However, those results show that large genomes limit the efficiency of LGT, increasing the overall input of mutations to the genome. Larger genomes increase the severity of the ratchet leading to shorter extinction times and faster rates of mutation accumulation (**Figure 2.2–3**). In order to avoid unnecessary complexity, the model does not take into account other possible benefits of LGT. In particular, a number of other theoretical studies indicate that LGT can increase the rate of adaptive evolution (Iles et al., 2003; Levin and Cornejo, 2009; Wylie *et al.*, 2010). In light of my results, it would be reasonable to expect these adaptive benefits decline as genome size increases.

Here I deliberately fixed genome size to evaluate the efficacy of LGT in the face of mutation accumulation. It would be illuminating to model the spread of modifiers, both of recombination length and the rate of LGT, given different genome size. As in the literature on organisms with meiotic sex (Barton and Otto, 2005; Iles *et al.*, 2003), the frequency of modifiers of LGT induced recombination will spread by hitchhiking as they generate variance in fitness, dependent on population size and the number of loci. This approach could be extended to ask how the spread of modifiers of LGT lead to changes in genome size. This could be attempted within a general model in which there is a fitness advantage of acquiring new genes through LGT balanced by the rate of deletion (Sela *et al.*, 2016). The role of recombination length and the rate of LGT in genome size expansion and the frequency of gene loss needs to be clarified, and will be addressed in the following chapter.

Under the assumption of a constant mutation rate per locus, theoretical results predict that the severity of the ratchet (measured as the equilibrium number of individuals in the least-loaded class) increases exponentially with genome size (Haigh, 1978) (see also Appendix A; **Figure 2.S1**). In addition to this effect – which could potentially be offset by a lower mutation rate per locus – large genomes are penalized by the decreased effectiveness of LGT at reducing the mutation load. This result holds true regardless of the assumption that large genomes present a greater mutational target than smaller ones. Even with a constant genome-wide mutation rate, LGT becomes less effective at purging large genomes from deleterious mutations, showing that this effect is not due to a higher mutation rate, but to an intrinsic limitation of LGT (**Figure 2.4**).

The increased potency of the ratchet as genome size increases is ameliorated by an increase in the rate of LGT (λ; **Figure 2.2–3**). Is this a viable option for prokaryotic species to enable them to expand genome size? In a number of species, LGT has been estimated as being the same magnitude (or higher) as the rate of nucleotide substitution, including *B. cereus* (Hao and Golding, 2006), *Streptococcus* (Marri *et al.*, 2006), *Corynebacterium* (Marri *et al.*, 2007), and *Pseudomonas syringae* (Nowell *et al.*, 2014). Rates are highly variable among species (Croucher *et al.*, 2012; Johnston *et al.*, 2015; Vos *et al.*, 2015). Competence for transformation can be induced by a range of environmental stressors including DNA damage, high cell density and limited nutrient availability (Bernstein and Bernstein, 2013). But LGT rates are constrained by eDNA availability, which depends on the amount of DNA in the environment and the degree of sequence homology (Croucher *et al.*, 2012; Vos *et al.*, 2015).

The model predicts that higher LGT rates will strengthen purifying selection and favour the elimination of mutants. This result is compatible with the strong correlation

observed between the number of horizontally transferred genes and genome size across a range of prokaryotes (Fuchsman *et al*., 2017; Jain *et al*., 2003). However, it is not clear to what extent the rate of LGT can be modified. The results presented here indicate that larger bacterial genomes are more likely to be sustained by higher rates of LGT, but the benefits of LGT as actually practiced by bacteria – the non-reciprocal uptake of small pieces of DNA comprising one or a few genes – are unlikely to sustain eukaryotic-sized genomes. In short, LGT as actually practised by bacteria cannot prevent the degeneration of larger genomes.

Importantly, the model shows that the benefits of LGT also increase with recombination length (L; **Figure 2.2–3**). In gram positive bacteria, recombination of large eDNA sequences is the exception rather than the rule (Croucher *et al.*, 2012; Mell *et al*., 2014). Experimental work indicates that the distribution of eDNA length acquired is skewed towards short fragments, with a third of transformation events less than 1kb, a median around 2-6kb and range extending up to ~50kb (Croucher *et al.*, 2012), although other studies have reported a larger median and range for transfer sizes (Hiller *et al*., 2010). Recombination of short DNA sequences appears to be an evolved state in *Streptococcus pneumoniae,* as the dedicated system cleaves eDNA into smaller fragments before recombination takes place (Claverys *et al*., 2009). Given that loci are around 1kb, with short intergenic regions, this represents the potential for several genes to be transferred in a single LGT event (Mira *et al*., 2001; Moran, 2002). There are several potential reasons for focus on small genomic pieces in LGT recombination. Cleavage of eDNA into smaller sequences increases the likelihood of homologous recombination, while the acquisition of long sequences can be associated with loss of genetic information (Croucher *et al.*, 2012) and can potentially disrupt regulatory and physiological networks (Baltrus, 2013). It has also been suggested that the small size of recombination fragments is a mechanism for

preventing the spread of mobile genetic elements (Croucher *et al*., 2016). On the other hand, gram-negative bacteria do not cleave eDNA on import, but their ability to acquire eDNA sequences >50kb is limited by physical constraints (Mell *et al.*, 2014). The high variability of LGT size suggest that there is flexibility and the potential for evolutionary change. But there is no evidence that larger genome size is accompanied by a higher recombination length. Bacteria do not appear to load large pieces of chromosome via LGT (i.e., >10% of a genome), although in principle it should be possible for them to do so.

As for purely asexual populations (Haigh, 1978), the strength of selection plays a critical role in determining the rate of mutation accumulation, with regions of the genome under strong selection accumulating mutations at a low rate (**Figure 2.5–6**). The ratchet effect is mainly observed in the accessory genome, with mutations accumulating preferentially in loci under weak selection (**Figure 2.5–6**). The model predicts that genome size expansion can occur in populations under strong purifying selection (e.g., due to a larger effective population size). Strong selection decreases the rate of genetic information loss, allowing the acquisition of new genetic content without an attendant increase in mutation fixation. This prediction is in agreement with the positive correlation observed between genome size and the rate of non-synonymous to synonymous substitutions (dN/dS) in bacteria (Bobay and Ochman, 2018; Novichkov *et al*., 2009). However, organisms under similar selective pressures often display a broad range of genome sizes (Novichkov *et al.*, 2009), indicating that other factors, including mutation rate and LGT, have a strong impact on prokaryotic genome size. Under high mutation rate and weak selective pressure, genome size expansion is disfavoured.

Eukaryotes, including simple unicellular organisms, typically possess much larger genomes than prokaryotes (Elliott and Gregory, 2015; Koonin, 2009a). Eukaryotic genome

size expansion was favoured by the acquisition of an endosymbiont, which evolved into the mitochondrion. This released bioenergetic constraints on cell size and allowed the evolution of genetic and morphological complexity (Lane, 2014; 2020; Lane and Martin, 2010). The endosymbionts underwent gene loss, a frequently observed process in extant endosymbiotic relationships (López-Madrigal and Gil, 2017), and transferred multiple genes to the host, enriching the host's genome size with genes of proto-mitochondrial origin (Martin *et al.*, 2015; Timmis *et al.*, 2004). The reduction of energetic constraints on genome size probably also facilitated gene (and even whole genome) duplications (Tria *et al*., 2021), leading to several thousand new gene families in LECA (Koonin *et al*., 2004), as well as lower selective pressure for gene loss after acquisition of novel gene functions by LGT (Szollosi *et al*., 2006). The acquisition of endosymbiotic DNA is also thought to have allowed the spread of mobile genetic elements in the host cell's genome, contributing to the increase in genome size and likely increasing the mutation rate (Martin and Koonin, 2006; Rogozin *et al*., 2012; Timmis *et al.*, 2004). The model explicitly considers an increase in the number of selected loci, but the benefits of LGT would decline in a similar fashion if genome size expansion proceeded through the acquisition of non-selected "junk" DNA. The presence of non-coding sequence would reduce the chance that a locus under selection undergoes recombination in a single LGT event (corresponding to smaller λ in the model), thus increasing the severity of the ratchet. The same would follow if increases in genome size followed from the expansion of mobile elements within the genome, although this would have introduced complexities relating to repeated sequences and ectopic recombination. These factors will be discussed in more detail in the following chapter.

Unprecedented genome size expansion brought the first eukaryotes under the threat of mutational accumulation, creating the need for stronger purifying selection in

order to keep the expanded genetic content free from mutations. The results presented here offer a possible explanation of why this process drove the transition from LGT to meiotic recombination at the origin of sex. In small prokaryotic genomes, LGT provides sufficient benefits to maintain genome integrity without incurring the multiple costs associated with sexual reproduction. But LGT fails to prevent the accumulation of deleterious mutations in larger genomes, promoting the loss of genetic information and therefore constraining genome size. The model shows that genome size expansion is only possible through a proportional increase in recombination length. I considered a recombination length $L = 0.2g$, which is equivalent to 500 genes for a species with genome size of 2,500 genes – two orders of magnitude above the average estimated eDNA length in extant bacteria (Croucher *et al.*, 2012). Recombination events of this magnitude are unknown among prokaryotes, possibly because of physical constraints on eDNA acquisition. Limiting factors likely include the restricted length of eDNA, uptake kinetics and the absence of an alignment mechanism for large eDNA strands (Baltrus, 2013; Croucher *et al.*, 2016; Thomas and Nielsen, 2005).

The requirement for a longer recombination length L cannot be achieved by LGT, which must therefore have failed to maintain a mutation-free genome, generating a strong selective pressure towards the evolution of a new mechanism of inheritance with the loss of energetic constraints on genome size. However, this magnitude of L is easily achievable via meiotic sex. The transition from LGT to meiotic sex involves the evolution of cell fusion, the transition from circular to linear chromosomes, whole-chromosome alignment and homologous recombination (Goodenough and Heitman, 2014; Lane, 2011). I have not explicitly modelled the details of this process or considered the order in which these factors arose (see Chapter 5 for a more detailed discussion of how my hypotheses fit different

eukaryogenesis scenarios). These aspects are crucial to our understanding of the origin of meiotic sex and will be addressed in a series of future publications. Nonetheless, the results presented in this chapter show that eukaryotes had to increase the magnitude of recombination length beyond the limits of LGT in order to permit the expansion in genetic complexity without the attendant increase in mutational burden. Eukaryotes had to abandon LGT in order to increase recombination length and maintain a large genome. Sex was forced upon us.

6. Conclusion

The benefits of LGT in maintaining genome integrity decline rapidly with genome size, making large genomes vulnerable to the accumulation of mutations. This effect constrains genome size in prokaryotes, and becomes even more severe with small population sizes and high mutation rates. These constraints can be partially overcome by increases in LGT rate and recombination length (**Figure 2.3–4**). But only recombination across the whole genome can wholly overcome these constraints. With the massive genome expansion at the origin of eukaryotes, the evolution of meiosis allowed homologous recombination across the whole genome, and not only across a limited region spanning little more than a few loci, as in LGT. The endosymbiosis that gave rise to the first eukaryotes led to the frequent transfer of genes from the endosymbiont to the host, resulting in a large expansion in genome size, likely coupled to high mutation rates. The model shows that these conditions wrought the failure of LGT in preventing Muller's ratchet. The resulting selective pressure promoted the evolution of sexual cell fusion and meiosis, maximizing recombination length and protecting eukaryotic genomes from excessive mutational burden. LGT in prokaryotes gave way to

meiotic sex in eukaryotes because only sex can sustain the expansion in genome size that underpins all eukaryotic complexity.

Chapter 3 – Repeat sequences limit the effectiveness of LGT and favoured the evolution of meiotic sex in early eukaryotes[3]

1. Summary

In eukaryotes, meiotic recombination prevents the accumulation of deleterious mutations predicted by Muller's ratchet. In prokaryotes, recombination of environmental DNA via Lateral Gene Transfer (LGT) provides similar benefits, but this effect is strongly constrained by genome size. As discussed in Chapter 2, recombination of long DNA sequences (as in meiosis) can halt mutation accumulation in large (eukaryotic-sized) genomes. But rather than taking up and recombining long DNA sequences, eukaryotes evolved sexual cell fusion, followed by homologous pairing and recombination. Long recombination events are also rarely observed in bacteria, which display a distribution of recombination length skewed towards short DNA fragments, indicating the presence on constraints on LGT. Repeat sequences, such as gene families and selfish genetic elements, might be crucial to understand these constraints and their role in the evolution of meiosis. In the presence of repeats, ectopic recombination can lead to the loss of genetic information through deletion, introducing an additional cost to the LGT of long DNA sequences. Here I show that in the presence of genetic repeats, the benefits of LGT are rapidly offset by increasingly high rates of genetic information loss. These results offer new insight into the origin of meiotic sex. Early eukaryotes underwent an unprecedented genome size expansion, with a concurrent rise in repeat sequences through the spread of introns and the creation of ~3000 new gene families. The need for purifying selection imposed by such large genome could not be met

[3] A revised version of this chapter is currently being prepared for publication.

by an increased LGT rate or recombination length without incurring in a catastrophic loss of genetic information. Homologous pairing followed by recombination across whole chromosomes allows a conspicuous increase in recombination length, while eliminating the risk of information loss associated with LGT in the presence of repeat sequences.

2. Introduction

The origin of sex has been long considered 'the queen of problems in evolutionary biology' (Bell, 1982). The universal prevalence of sexual reproduction among eukaryotes indicates that meiotic sex evolved early, before the divergence of the first eukaryotic clades (Schurko and Logsdon, 2008; Speijer *et al.*, 2015). It did so alongside all the distinguishing features of eukaryotic complexity, including mitochondria, a dynamic cytoskeleton, linear chromosomes, the nucleus and an endomembrane system (Dacks *et al*., 2016; Koumandou *et al*., 2013). But the selective reasons behind the evolution of meiosis are still poorly understood. A traditional approach in the literature is to evaluate the advantages sexual reproduction by contrasting clonal and sexual populations (Bell, 1982; Otto, 2009). While this approach has been successful in illuminating the widespread maintenance of sex in eukaryotes, it does not directly address the selective pressures that led to the origin of meiotic sexual reproduction in the context of eukaryogenesis, the evolutionary transition that gave rise to eukaryotic life (Maynard Smith and Szathmáry, 1995; Szathmáry, 2015).

The presence of common molecular machinery suggests that meiosis evolved from transformation (one of the major routes of genetic exchange via LGT in prokaryotes, which involves the acquisition of environmental DNA, followed by recombination into the host genome (Ambur *et al.*, 2016; Johnston *et al.*, 2014)) during the transition to eukaryotic organisation (Bernstein and Bernstein, 2013; Mirzaghaderi and Hörandl, 2016; Schurko and

Logsdon, 2008). Some key genetic elements involved in meiotic sex were repurposed in early eukaryotes from genes present in archaea and bacteria (Lin *et al.*, 2006). In particular, prokaryotes possess homologues of the cardinal molecular machinery for meiotic sex, including proteins of the *SMC* gene family of ATPases necessary for chromosome cohesion and condensation (Hirano, 2005); actin and tubulin, required for daughter cell separation and the movement of chromosomes, respectively (Erickson, 2007); and the *Rad51/Dcm1* gene family, which plays a central role in meiosis and has high protein sequence similarity with *RecA*, responsible for homologous search and recombination during bacterial transformation (Johnston *et al.*, 2014; Lin *et al.*, 2006). Besides sharing central molecular machinery with meiosis, transformation also provides similar benefits.

Transformation is one of the major routes of genetic exchange via LGT in prokaryotes, which involves the acquisition of environmental DNA, followed by recombination into the host genome, and has been studied intensively in bacteria (Ambur *et al.*, 2016; Johnston *et al.*, 2014). Archaea also undergo transformation, although the mechanisms involved are less well understood (Fonseca *et al*., 2020; Wagner *et al*., 2017). In several species, uptake is limited by specific recognition sequences, favouring sequences from close relatives (de Vries and Wackernagel, 2002; Frye *et al*., 2013; Hülter and Wackernagel, 2008; Mell *et al*., 2012). Unlike other forms of LGT (conjugation by plasmids and transduction by phage), transformation is completely regulated and controlled by the host cell, through genes exclusively encoded on the host chromosome (Seitz and Blokesch, 2013).

By allowing genetic exchange between lineages, transformation provides similar benefits to those of meiotic recombination, favouring adaptation through the acquisition of new genes (Gogarten and Townsend, 2005; McInerney *et al*., 2017; Pál *et al*., 2005; Polz *et*

al., 2013). In several species, the acquisition of sequence from close relative is favoured by specific recognition sequences, supporting the idea that transformation provides not only adaptive, but also conservative benefits (de Vries and Wackernagel, 2002; Frye *et al.*, 2013; Hülter and Wackernagel, 2008; Mell *et al.*, 2012). In experimental evolution, recombination via LGT, mediated by transformation, can accelerate adaptation by reducing selective interference (Baltrus *et al*., 2008; Engelmoer *et al*., 2013). Populations of *Streptococcus pneumoniae* undergoing transformation also accumulate deleterious mutations at a lower rate than non-recombining populations, and are less likely to evolve mutator strains (Engelmoer *et al.*, 2013). This evidence is in agreement with theoretical models indicating that transformation can restore genes that have been disrupted through mutation or deletion (Levin and Cornejo, 2009), prevent the loss of genes under weak selection (van Dijk et al., 2020), and halt Muller's ratchet in prokaryotes (Colnaghi *et al.*, 2020; Takeuchi *et al.*, 2014). Given that LGT provides very similar benefits to those of meiotic sex, it is still hard to understand why early eukaryotes abandoned the former for the latter.

In the previous chapter, I suggested that ancestral genome size expansion was pivotal in determining the failure of LGT and drove the evolution of meiosis (Colnaghi *et al.*, 2020). Eukaryotes originated from the endosymbiosis between a close relative of modern alpha-protobacteria and an archaeal host related to the Asgard clade, both of which were likely to undergo LGT via transformation (Martin *et al.*, 2015; Zaremba-Niedzwiedzka *et al.*, 2017). During the early stages of eukaryotic evolution, the archaeal host's genome was enriched with genes of endosymbiotic origin, whilst the endosymbiont genome became greatly reduced (Martin *et al.*, 2015; Timmis *et al.*, 2004). The extra energetic availability provided by the proto-mitochondrial endosymbiont released bioenergetic constraints over prokaryotic cell and genome size (Lane, 2020; Lane and Martin, 2010). But this extra

energetic availability came with the cost of maintaining a larger genome, as the majority of bacterial genes supporting the function of the proto-mitochondrial endosymbiont were transferred to the host genome, leading to a massive increase in genome size (Koonin *et al.*, 2004; Lane and Martin, 2010; Timmis *et al.*, 2004; Vosseberg *et al*., 2021). In addition to the import of bacterial coding sequences, host genome size expansion reflected an increase in repeat sequence density. This occurred through gene duplication and divergence to enable a range of novel functions, which is estimated to have doubled gene number and led to the acquisition of ~3000 new gene families in LECA (Koonin *et al.*, 2004; Makarova *et al*., 2005; Vosseberg *et al.*, 2021). In addition, repeat sequences are thought to have expanded early in the proto-eukaryotic cell, potentially leading to gene loss through ectopic recombination. Group II self-splicing selfish mobile elements with a bacterial endosymbiotic origin are thought to have spread widely through the host genome, giving rise to introns in early eukaryotes and promoting the evolution of the nuclear membrane (Koonin, 2006; Martin and Koonin, 2006; Rogozin *et al.*, 2012). These and related transposable elements are present in many bacterial species, almost always at low copy numbers (<10 per genome) (Lambowitz and Belfort, 2015). But in the proto-eukaryote cell, they increased in a more uninhibited manner, perhaps exploiting the nonhomologous end-joining mechanism of DNA repair found throughout eukaryotes (Lee *et al*., 2018), reaching a density comparable to that seen in modern eukaryote species (Rogozin *et al.*, 2012).

Increased genome size is likely to have caused the failure of LGT in preventing the accumulation of deleterious mutations, as the severity of Muller's ratchet increases exponentially with genome size and the benefits of LGT are inversely correlated with it (Colnaghi *et al.*, 2020). As discussed in the previous chapter, LGT helps to purge deleterious mutations (Takeuchi *et al.*, 2014), but becomes less effective as genome size increases

(Colnaghi *et al.*, 2020). The results presented in Chapter 2 indicate that LGT can resist mutation accumulation in larger genomes through increases in recombination length, the mean size of DNA picked up from the environment and recombined into the host genome (Colnaghi *et al.*, 2020); similarly, meiotic sex allows genetic exchange across whole chromosomes, increasing the magnitude of recombination. Recombination of large DNA sequences via LGT should provide similar benefits to those of meiosis, but is rarely observed in extant prokaryotes. Experimental work indicates that the distribution of recombination length in bacteria is skewed towards short eDNA sequences, with a median length that encompasses at most just few genes (Claverys *et al.*, 2009; Croucher *et al.*, 2012; Hiller *et al.*, 2010). In addition, bacteria typically cleave environmental DNA reducing recombination length (Croucher *et al.*, 2012), and no bacterial species regularly load large pieces of DNA. Another route to a larger genome is an increase in the rate of LGT, but again there are strong constraints on the rates of uptake and recombination of environmental DNA, which are limited by eDNA availability and sequence homology (Croucher *et al*. 2012; Vos *et al*. 2015). These observations suggest that LGT becomes increasingly ineffective with genome size expansion, requiring the evolution of a different method of genetic exchange. But the forces limiting the efficacy of LGT are still poorly understood.

Here, I hypothesize that ectopic recombination caused by repeat sequences constrains recombination length and LGT frequency, and that the dramatic increase in repeat density was one of the reasons why eukaryotes were forced to abandon LGT in favour of meiosis. In order to study whether repeats constrain LGT frequency and recombination length, I develop a computational model of mutation and selection in a population of haploid individuals undergoing homologous recombination of genetic material via transformation (which, for simplicity, I will refer to as 'LGT') in the presence of genetic

repeats. The model presented in this chapter highlights a trade-off between the benefits of LGT, greater genetic variance leading to increased purifying selection, and its cost, the loss of genetic information because of ectopic recombination. I postulate that the evolution of meiotic sex was driven by the need for increased purifying selection in the expanding genomes of the early eukaryotes, which could not be met by an increase in recombination length or LGT rate alone. The modelling shows that the limitations of LGT in the presence of repeat sequences is a crucial factor in understanding the transition from prokaryotic transformation to meiotic sex.

3. Methods

I use a Fisher-Wright process with discrete generations to model the evolution of a population of N haploid individuals, subjected to a deleterious point mutation rate of μ per locus per generation and undergoing transformation (LGT) at a rate λ. Model dynamics are analogous to those described in the previous chapter (**Figure 2.1**). Every individual begins with a genome composed of g unique protein-coding genes, each of which can exist either in a wildtype or deleterious mutant state. Interspersed in the genome is a repeated sequence at a density ρ per protein-coding gene. The initial positions of the repeats are randomly sampled from a uniform distribution.

The new generation is obtained by sampling N individuals, with replacement, from the old population. The probability of reproduction is proportional to the individual fitness. Following previous theoretical studies (Colnaghi *et al.*, 2020; Gordo and Charlesworth, 2000b; Haigh, 1978; Takeuchi *et al.*, 2014) I assume no epistatic interactions and measure fitness as a multiplicative function:

$$w_i(t) = (1 - s)^{g - n_i(t)}, \quad (3.1)$$

where $n_i(t)$ is the number of different functional protein-coding genes possessed by an individual i at time t. In order to avoid unnecessary complexity, I neglect the fitness effect of gene duplication, and only consider whether there is at least one functional copy of each protein-coding gene.

Table 3.1: Model parameters and variables

μ	Deleterious point mutation rate per locus
U	Genome-wide deleterious point mutation rate
λ	LGT rate
L	Recombination length
N	Population size
t_{max}	Length of simulations (generations)
ρ	Initial repeat density
g	Number of protein-coding genes
ΔM_{del}	Gene loss rate per generation due to deletions
ΔM_{mut}	Gene loss rate per generation due to point mutation
ΔM	Total gene loss rate (mutations plus deletions)

Once a new generation is formed, the old generation dies and their DNA forms the genetic pool from which the new generation acquires environmental DNA (eDNA) for recombination. I assume that eDNA strands are only stable for one generation before decaying irreversibly. Individuals of the new generation undergo homologous

recombination via LGT with a probability λ. For successful recombination, homology is required at the terminal regions (spanning one locus each) of the eDNA sequence and the host genome. Homology can be either to protein-coding genes or repeats (**Figure 3.1A–B**). The genome is assumed to be circular, so locus g is contiguous with locus 1.

For each individual that undergoes LGT, a sequence of eDNA of length L is randomly sampled from the eDNA pool. For successful recombination, homology is required at the last elements of the eDNA sequence and the host genome. Homology can be either to protein-coding genes or repeats (**Figure 3.1A–B**). These dynamics follow the experimental evidence that recombination of nonhomologous DNA can take place fairly easily in the presence of homologous flanking sequences, but not in their absence (de Vries and Wackernagel, 2002; Hülter and Wackernagel, 2008). Integration of foreign DNA is estimated to be at least 10^9 times less likely of that of homologous DNA (Hülter and Wackernagel, 2008) and therefore neglected in the model. In addition to LGT favoured by flanking sequence homology, homologous recombination is studied by placing a further restriction for sequence homology throughout the whole eDNA. After the eDNA sequence is sampled from the eDNA pool, one of the its two terminal regions (spanning one locus each) is randomly selected, and a homologous locus x_b is found in the host genome to which the eDNA sequence attaches. If multiple homologous loci are present, one site is selected at random. Homologous loci x_e to the other end of the eDNA sequence are found in the recipient genome. For each x_e a recombination probability is generated according to a Gaussian distribution,

$$\pi(d) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{d-L}{\sigma}\right)^2} \qquad (3.2)$$

where d is the distance (number of loci) between x_b and x_e, and the standard deviation is $\sigma = \sqrt{L}$. This reflects the assumption that longer eDNA sequences lead to increased variance in the distribution. One of the x_e loci is randomly selected, with weights proportional to $\pi(d)$, to be the other terminal region of recombination. With successful pairing, all the elements between x_b and x_e (included) are substituted by the recombining eDNA sequence. If there is no match to either x_b or x_e, recombination with the eDNA sequence is not possible and no genetic exchange takes place. In addition, to model homologous recombination, I randomly select two loci x_b and x_e, as for LGT. In the presence of full sequence homology (i.e., if the two sequences contain exactly the same genes, either as wild-type or mutant alleles) recombination successfully takes place; the sequence between x_b and x_e is excised and replaced by the eDNA sequence. Otherwise, no genetic exchange takes place.

After LGT, each individual acquires m new deleterious mutations, where m is a random variable drawn from a Poisson distribution with mean U. The genome wide mutation rate $U = \mu g'$, where g' is the number of wildtype protein-coding genes, as we assume that mutated genes cannot be mutated again. The position of the particular locus or loci in the genome that mutate is then randomly determined. For simplicity, the possibility of back mutation in protein-coding genes and mutations within repeat sequences is neglected.

The evolutionary process is studied with a population initially free of mutation, over 5,000 generations, with 50 replicates for a given set of parameter values. For each replicate, I evaluate the gene-loss rate per generation from deletions ($\Delta M_d/\Delta t$) and mutations ($\Delta M_m/\Delta t$) as the sum of the number of fixed deletions and mutations respectively, divided

by the number of generations. Similarly, the total gene loss rate per generation ($\Delta M/\Delta t$) is calculated as the sum of these two components.

The model described above has been implemented using MATLAB. Random sampling is achieved using the function 'randsample'. All random processes described above have been simulated using MATLAB's functions 'rand' and 'random'. The code used for the simulations has been uploaded on GitHub (https://github.com/MarcoColnaghi1990/LGT-repeat-sequences).

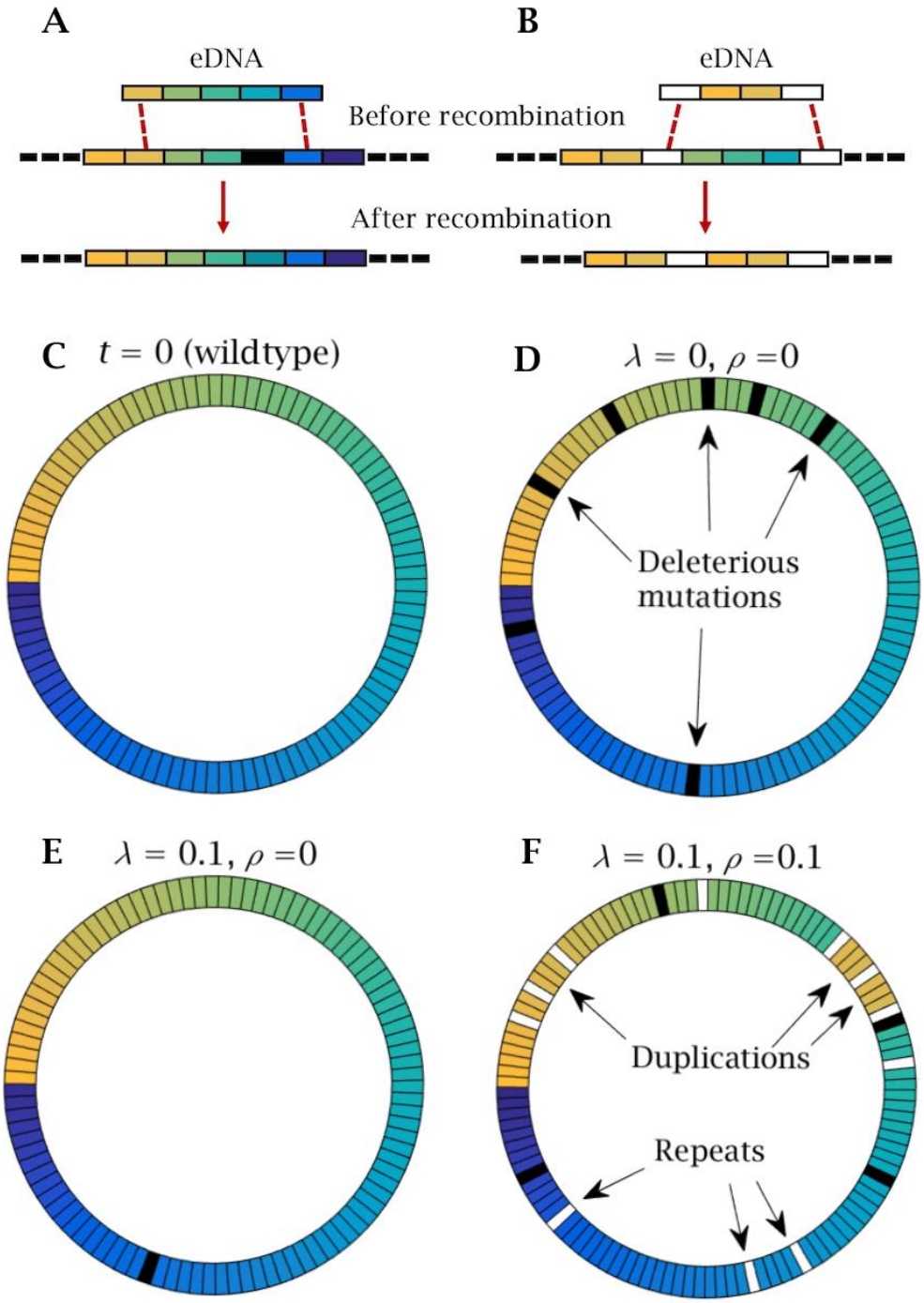

Figure 3.1 | Illustration of the model dynamics.

In LGT recombination of eDNA occurs through the matching of end sequences. It can lead either to **(A)** homologous recombination with the replacement of deleterious mutations (in black) by the wildtype allele, or **(B)** ectopic recombination caused by the presence of repeats (in white) resulting in the loss or duplication of genes. **(C)** The wildtype genome is represented as a linear series of genes indicated by different colours. **(D)** The wildtype genome is subject to mutation pressure resulting in

the accumulation of deleterious alleles by Muller's ratchet. **(E)** LGT ($\lambda = 0.1$) allows homologous recombination, increases genetic variation and favours the elimination of deleterious mutations. **(F)** In the presence of repeats ($\rho = 0.1$) the possibility of ectopic recombination limits this benefit and leads to gene deletions and duplications. Other simulation parameters: $t = 5{,}000, g = 100, N = 2{,}500, L = 5, U = 0.003$.

4. Results

The mean mutation load of the whole population and the number of mutations in the least-loaded class (LLC), reflect the interplay between genetic drift and Muller's ratchet (**Figure 3.2**). In a clonal population ($\lambda = 0$), random fluctuations can cause the LLC to go extinct (a 'click' of the ratchet) determining an ever-increasing mutation load baseline (**Figure 3.2A**). In the absence of LGT, the fittest class cannot be restored, and the increase in mutation load is irreversible. The introduction of LGT ($\lambda = 0.1, L = 5$) favours the elimination of mutations by increasing genetic variation and strengthening purifying selection (**Figure 3.2B**). LGT permits the reversal of Muller's ratchet and the reduction in mutation number in the LLC. In the presence of a high repeat density ($\rho = 0.1$), the benefit of LGT comes at the price of a high number of deletions due to ectopic recombination, making LGT less obviously useful (**Figure 3.2C**).

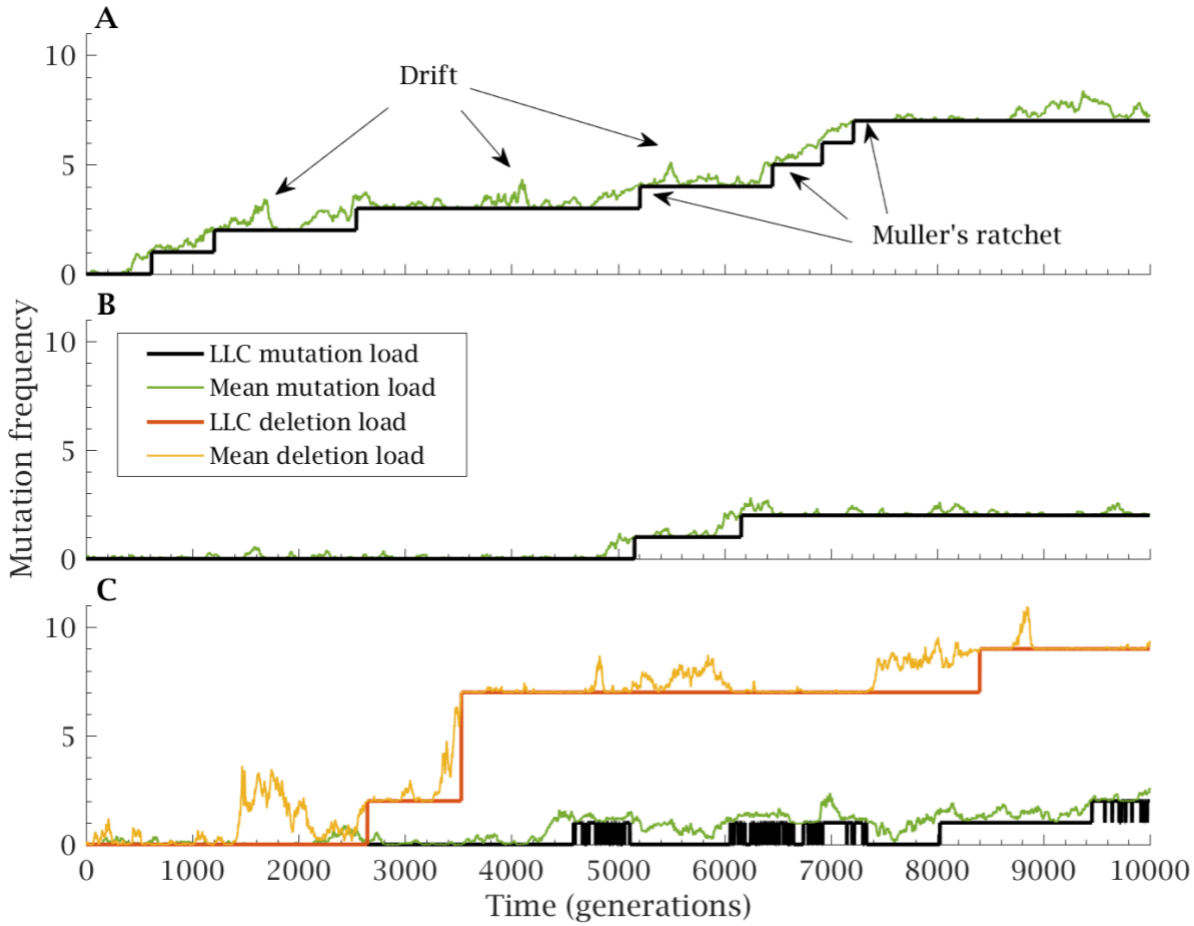

Figure 3.2 | Impact of LGT on Muller's ratchet dynamics. (A) Mean mutation load (green line) and least loaded class (LLC) mutation load (black line) of a repeats-free ($\rho = 0$) clonal population that does not undergo LGT ($\lambda = 0$), across $t_{max} = 10{,}000$ generations. Random fluctuations due to genetic drift lead to irreversible increases of the LLC mutation load (Muller's ratchet). **(B)** Mean mutation load (green line) and LLC mutation load (black line) of a repeats-free ($\rho = 0$) population undergoing LGT ($\lambda = 0.1$, $L = 5$). Recombination via LGT increases purifying selection, leading to a lower mutation load and a less severe ratchet. **(C)** Mean mutation (green line) and deletion (yellow line) loads, and LLC mutation (black) and deletion (red) loads of a population undergoing LGT ($\lambda = 0.1$, $L = 5$) in presence of repeats ($\rho = 0.1$). While LGT allows the ratchet to be reversed and reduces mutation accumulation, ectopic recombination due to the presence of repeats leads to a high rate of deletion. Other simulation parameters: $g = 100$, $N = 2{,}500$, $U = 0.003$.

Repeat density strongly influences the benefit of LGT. If repeat density is low ($\rho = 10^{-2}$), increasing LGT (λ) is beneficial and reduces the total gene loss rate (**Figure 3.3**). But as repeat density rises ($\rho = 0.5 \times 10^{-2}$) this benefit is eroded, and higher levels of LGT are

not advantageous (**Figure 3.3**). At high levels of repeats ($\rho = 10^{-1}$), LGT is always detrimental and elevates total gene loss (**Figure 3.3**).

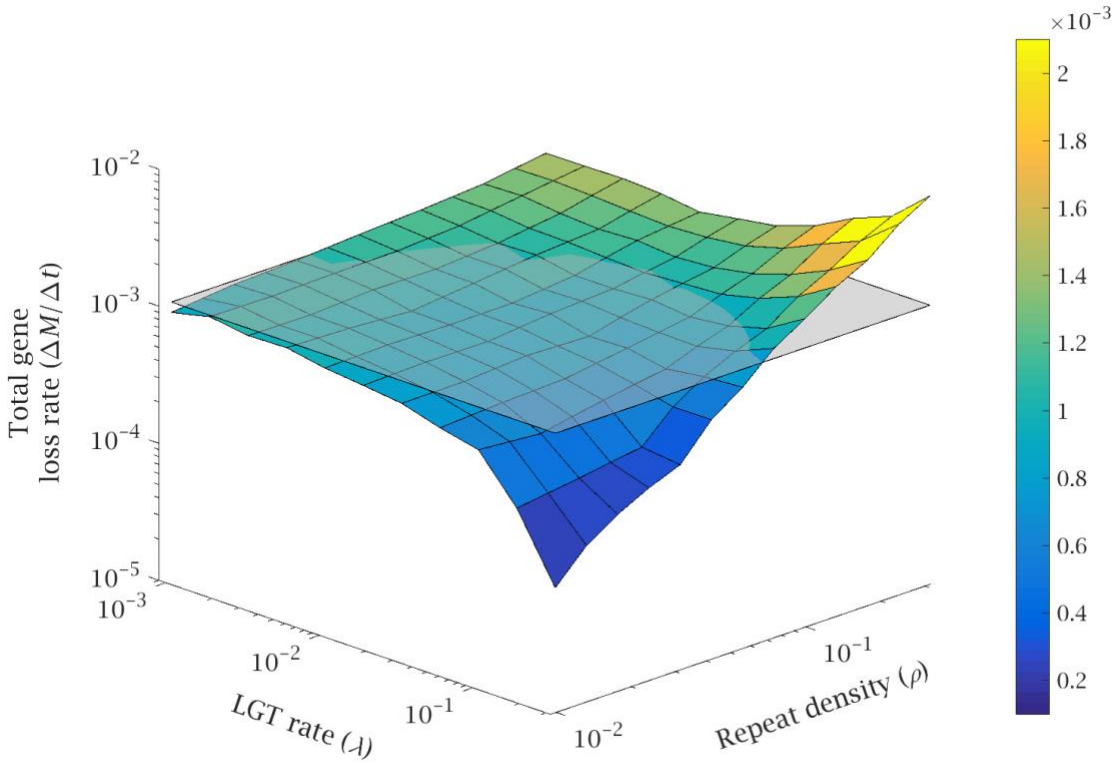

Figure 3.3 | Genomic repeats cause the failure of LGT.
The total gene loss rate ΔM (through both deletions and deleterious point mutations) for a range of initial repeat densities (ρ) and LGT rates (λ). For comparison, the grey plane shows the total gene loss rate in a repeats-free population not undergoing LGT (null model). Other parameters: $g = 100$, $N = 2{,}500$, $t_{max} = 5{,}000$, $L = 10$, $\mu = 3 \times 10^{-5}$. Each data point has been obtained by averaging over the outcome of 100 independent simulations.

This is because a higher density of genomic repeats increases the likelihood of ectopic recombination, leading to an exponential increase in gene loss through deletions with repeat density (**Figure 3.4A**). In addition, high repeat density limits the effectiveness of LGT in purging deleterious mutations (**Figure 3.4B**). In sum, LGT ceases to have a beneficial effect beyond a threshold repeat density (**Figure 3.4C**).

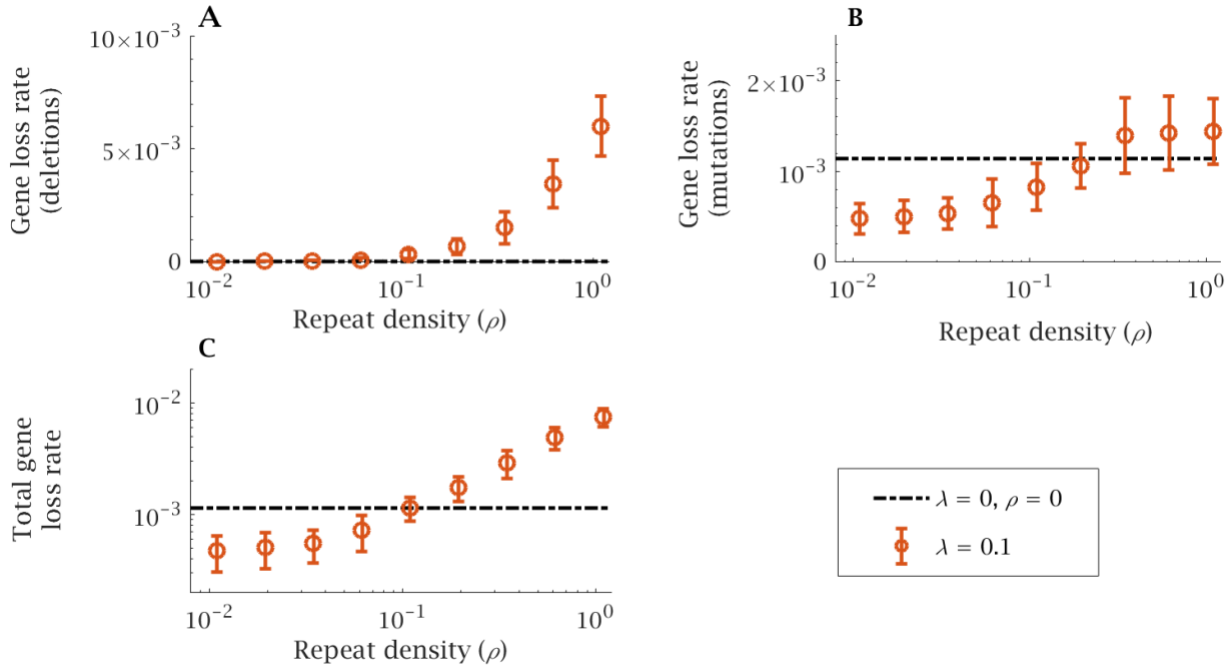

Figure 3.4 | Effect of repeat density.

The gene loss rate caused by (**A**) deletions ($\Delta M_d/\Delta t$), (**B**) mutations ($\Delta M_m/\Delta t$) and (**C**) the total gene loss rate ($\Delta M/\Delta t$) given different values of repeat density (ρ). Simulations were carried out with a high rate of LGT ($\lambda = 0.1$), reporting the mean and standard deviation over 100 runs. The dotted line represents the null case of mutation accumulation when there is no LGT and no repeats. Other simulation parameters: $g = 100$, $N = 2{,}500$, $\mu = 3 \times 10^{-5}$, $t_{max} = 5{,}000$, $L = 10$. Data points and error bars show, respectively, the average and the standard deviation over 100 independent simulations.

The lower efficiency of LGT at removing deleterious mutations arises because repeats make homologous recombination less likely and increase the frequency of ectopic recombination. This can be seen by adding a requirement for homology in the middle of the eDNA as well as homology at the ends. This eliminates ectopic recombination, and mutation accumulation then closely follows the case without repeats, showing an exponential decline with LGT rate (**Figure 3.5A**). In contrast, with normal recombination based on end homology alone, there is a monotonic decline in mutation accumulation as the rate of LGT increases

(**Figure 3.5A**). In addition, deletions mean that sometimes one or both ends of the eDNA lack homology to any genomic sequence, increasing the probability that no recombination takes place (**Figure 3.5B**). These effects combine to reduce the rate of homologous recombination as repeat density increases, limiting LGT's ability to purge deleterious mutations (**Figure 3.5B**).

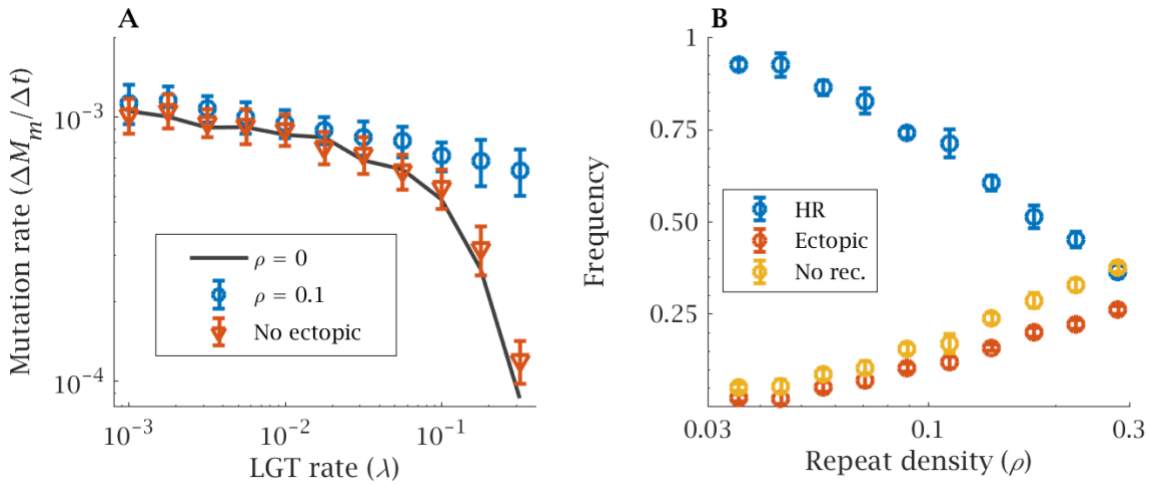

Figure 3.5. Ectopic recombination.

(**A**) Rate of mutation accumulation in the presence of repeats ($\rho = 0.1$). The comparison is shown between a null model without repeats ($\rho = 0$, black line), a genome with high repeat density ($\rho = 0.1$, blue circles) and a genome with high repeat density but no ectopic recombination due to the requirement for homology in the middle of the eDNA (red triangles). (**B**) Frequency of homologous recombination (HR), ectopic recombination and failure to recombine as a function of repeat density. Rate and frequency were calculated over $t_{max} = 5{,}000$ generations. Other parameters: $g = 100$, $N = 2{,}500$, $L = 10$, $U = 0.003$ and $\lambda = 0.1$ (in (**B**)). Data points and error bars show, respectively, the average and the standard deviation over 100 independent simulations.

Genome size has a marked impact on the rate of accumulation of mutations (**Figure 3.6**). At low repeat density ($\rho = 0.01$), the purging of deleterious mutations by LGT of average-sized DNA sequences ($L = 10$) is higher in smaller genomes; this is because the probability that a stretch of eDNA matches to a particular deleterious mutation decreases

with genome size (**Figure 3.6A**). In small genomes ($g = 100$), higher rates of LGT are beneficial, lowering the total rate of gene loss. As genome size increases ($g = 1{,}000$), higher rates of LGT are required in order to prevent the ratchet, and LGT provides little or no advantage in large genomes ($g = 5{,}000$; **Figure 3.6A**). At high repeat density ($\rho = 0.1$), even small genomes benefit little from LGT, and increases in LGT frequency are detrimental (**Figure 3.6B**). This effect is evident even when assuming a constant genome-wide mutation rate ($U = 0.003$) rather than a constant mutation rate per bp, which would further penalise large genomes.

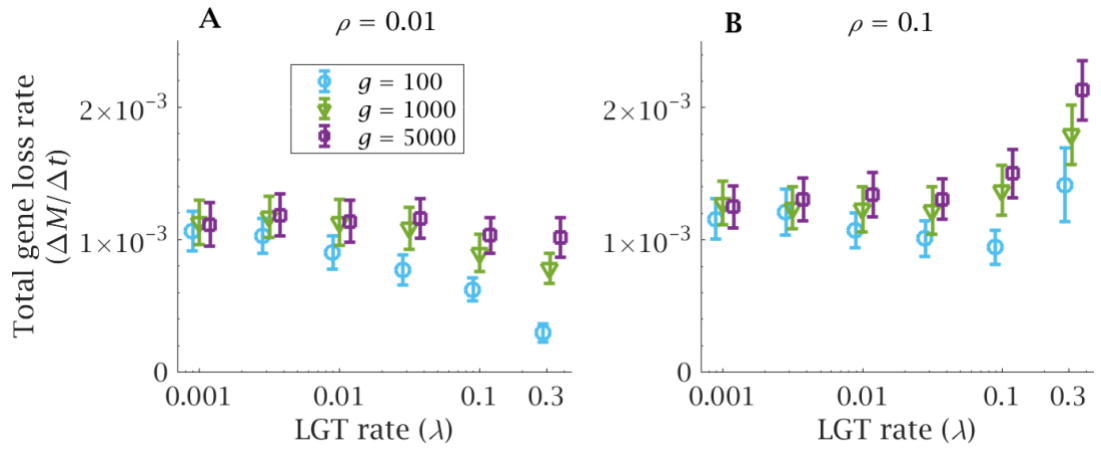

Figure 3.6 | Genome size. Impact of genome size on gene loss rate under a constant genome-wide mutation rate ($U = 0.003$) against the LGT rate (λ). (**A**) At a low repeat density ($\rho = 0.01$) and (**B**) high repeat density ($\rho = 0.1$), in small ($g = 100$, blue), medium ($g = 1{,}000$, green) and large genomes ($g = 5{,}000$, purple). Other parameters: $L = 10$, $N = 2{,}500$, $t_{max} = 5.000$, $U = 0.003$. Data points and error bars show, respectively, the average and the standard deviation over 100 independent simulations.

In the previous chapter, I have shown that increasing L maximises the benefits of LGT in repeat-free genomes (Colnaghi *et al.*, 2020). In particular, if recombination length scales proportionally to genome size, LGT can prevent the mutational meltdown of large genomes. These results hold true when repeat density is low (~ 0.01; **Figure 3.7B**), but

increase in the recombination length (L) is no longer a solution in large genomes with numerous repeats. Repeats lessen the reduction in mutation accumulation ($\Delta M_m/\Delta t$) as L increases, due to LGT becoming less efficient (**Figure 3.7A**). But the greater effect is felt through a strong increase in the frequency of deletions ($\Delta M_d/\Delta t$) as L increases, which pushes up the total gene loss rate ($\Delta M/\Delta t$, **Figure 3.7A**). This is because the magnitude of deletions (and therefore the number of loss genes) is proportional to recombination length. The variance of the binding function is also dependent on L, making large deletions more frequent as recombination length increase. These effects are magnified as repeat density (ρ) increases (**Figure 3.7B**). LGT is beneficial, but is unable to resist gene loss as repeat density increases because of the rising effect of deletions, which dominate with higher L (**Figure 3.7B**). A higher L is better, but can't resist high repeat density ($\rho = 0.1$; **Figure 3.7B**). In the previous chapter, I have shown that homologous recombination, with recombination length proportional to genome size and the requirement for sequence homology in the middle of the eDNA, allows a reduction of the mutation load regardless of genome size. Here, I show that the same holds true in the presence of repeats, as homologous recombination allows an increase in recombination length ($L = 0.1g$) without incurring the associated increase in gene deletions. As a consequence, it is able to lower the total gene loss rate even in the presence of a high repeat density (**Figure 3.7B**).

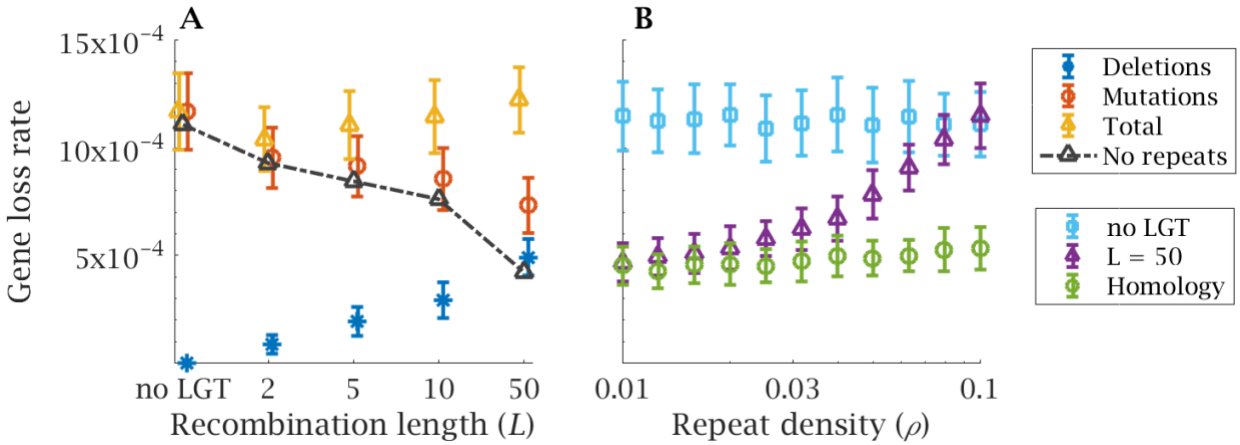

Figure 3.7 | Recombination length and the advantage of homologous recombination.

(**A**) Gene loss rate as a function of recombination length (L) in a moderate size genome ($g = 500$) with high repeat density ($\rho = 0.1$). The effect of recombination length (L) on mutation accumulation ($\Delta M_m/\Delta t$, red circles), the frequency of deletions ($\Delta M_d/\Delta t$, blue stars) and total gene loss rate ($\Delta M/\Delta t$, yellow triangles) are compared to the rate in a genome without repeats (black triangles). (**B**) Total gene loss rate as a function of repeat density (ρ) given homologous recombination (green circles) with sequence homology in the middle of the eDNA and length proportional to genome size ($L = 0.1g$) is compared to LGT with low ($L = 5$) and high ($L = 0.1g$) recombination length, and the null model in the absence of LGT (light blue squares). Parameter values: $g = 500$, $N = 2{,}500$, $t_{max} = 5{,}000$, $\lambda = 0.1$, $U = 0.003$. Data points and error bars show, respectively, the average and the standard deviation over 100 independent simulations.

5. Discussion

Organisms lacking a normal sexual process are often portrayed as destined to mutation accumulation through Muller's ratchet, on an inevitable decline towards extinction through mutational meltdown (Haigh, 1978; Lynch *et al.*, 1993; Muller, 1964). This view emanates from a eukaryotic perspective seeking to explain the maintenance of sexual reproduction in the face of the two-fold cost of males and other costs relating to meiosis, finding a mate and cell fusion (Bell, 1982; Otto, 2009; Otto and Lenormand, 2002). Prokaryotes, both bacteria and Archaea, lack meiotic sex and typically reproduce through asexual division, but they nonetheless have a number of mechanisms of genetic exchange (Gogarten and Townsend,

2005; Ochman *et al.*, 2000; Polz *et al.*, 2013; Vos *et al.*, 2015) . In particular, lateral gene transfer (LGT) caused by transformation allows competent cells to pick-up eDNA released from related lineages and to recombine it into the host genome. In accord with previous studies (Colnaghi *et al.*, 2020; Levin and Cornejo, 2009; Takeuchi *et al.*, 2014), the model presented here shows that LGT generates genetic variation and strengthens purifying selection, reducing the rate of mutation accumulation, performing a similar function to meiotic sex in eukaryotes. LGT permits the reversal of Muller's ratchet and the reduction in mutation number in the least loaded class, which would otherwise be irreversible (**Figure 3.2**). This leads to a simple question: why did the first eukaryotes, derived from a symbiotic association of prokaryotic cells, largely abandon LGT and replace it with meiotic sex?

The analysis here shows that the advantages of LGT are undermined by the presence of genomic repeats (**Figure 3.3**). The reason for this is that repeat sequences open the opportunity for ectopic recombination, resulting in loss of genomic sequence. The modelling shows that gene loss increases exponentially with the density of repeats in the host genome (**Figure 3.4A**). Repeats also make LGT less effective at purging deleterious mutations (**Figure 3.5A**) by reducing the rate of homologous recombination (**Figure 3.5B**). In addition, as deletions segregate with mutations at the same time, they cause a form of Hill-Robertson interference (Hill and Robertson, 1966; Roze and Barton, 2006), lowering the effectiveness of selection (**Figure 3.5A-B**). A higher deletion rate, lower homologous recombination rate and weaker selection contribute to increased loss of genetic information as repeat density increases (**Figures 3**). These combined effects depend on higher LGT rates (λ), which are advantageous in the absence of repeats (**Figure 3.3**; Chapter 2), but become disadvantageous above a critical repeat density (**Figures 4C**).

In the previous chapter, it was demonstrated that the effectiveness of LGT declines with genome size expansion in early eukaryotes (Colnaghi *et al.*, 2020). As genome size rises, the homologous recombination rate *per locus* falls, simply because the probability that an eDNA piece matches to a particular locus is inversely proportional to genome size. A solution to this is to increase the rate of LGT. This is beneficial in small genomes, but has rather weaker benefits as genome size increases (**Figure 3.6A**). The presence of repeats exaggerates this problem, as the same density of repeats in a large genome multiplies the opportunity for ectopic recombination (**Figure 3.5B**). The net effect of repeats is that increasing the rate of LGT pushes up total gene loss (through mutation and deletion) especially in populations with larger genomes (**Figure 3.6C**). Higher rates of transformation are no longer a solution.

As discussed in Chapter 2, an alternative solution for a larger genome is to increase recombination length. Picking up larger pieces of eDNA (L) leads to a higher recombination rate *per locus* and thereby facilitates the elimination of deleterious mutations (Colnaghi *et al.*, 2020). But these results again depend on the absence of repeats. Larger recombination length is associated with higher deletion probability in genomes with repeats (**Figure 3.7A**). At high repeat density ($\rho = 0.1$), the enhanced elimination of deleterious mutations is outweighed by more deletions, coupled with the lower efficiency of purifying selection, causing the total rate of genetic information loss to increase with recombination length (**Figure 3.7A**). Once again, a higher repeat density magnifies this deficit ($\rho = 0.01 - 0.1$), resulting in the total gene loss of a population using long eDNA sequences (10% of total genome size) to rapidly approach that of a population in the absence of LGT (**Figure 3.7B**).

These results help to explain why the distribution of recombination length in extant prokaryotes is skewed towards shorter sequences (Croucher *et al.*, 2012) and why gram-

positive bacteria cleave eDNA sequences before recombination (Mell *et al.*, 2014). As the presence of repeats can undermine the stability of bacterial genomes, causing to deletions, inversions, and whole-genome rearrangements (Siguier *et al*., 2014; Vandecraen *et al*., 2017), prokaryotes typically display low numbers of transposable elements or gene families (Lambowitz and Belfort, 2015; Puigbò *et al*., 2014). Together with low repeat density, LGT of short DNA sequences permits homologous recombination that repairs mutations whilst greatly reducing the chances of deletion through ectopic recombination (**Figure 3.7A**). This works well in small genomes but fails with large genomes infected with a high density of repeats (**Figure 3.7B**).

The acquisition of new genes through duplications and endosymbiotic gene transfer was a crucial step in the evolution of eukaryotic complexity. It also made the first eukaryotes vulnerable to the accumulation of mutations caused by Muller's ratchet, because of the greater mutational target and decreasing benefits of LGT in large genomes (**Figure 3.6**). The process of genomic expansion came at the price of a high repeat density, which prevented the first eukaryotes from achieving stronger purifying selection through an increase in LGT rates or recombination length. In order to support an expanded genome rich in repeat sequences, the first eukaryotes had to abandon LGT for meiotic sex: the transition to cell fusion and the requirement for homologous pairing across whole chromosomes was the only way they could retain the benefits of recombination without losing genetic information through LGT in the presence of repeats (**Figure 3.7B**). I simulate this homology matching in the LGT model by adding a requirement for homology in the middle of the eDNA as well as homology at the ends. This eliminates ectopic recombination, and mutation accumulation then closely follows the case without repeats, showing an exponential decline with LGT rate (**Figure 3.5A**). In contrast, with normal recombination based on end homology

only, the exponential increase in the rate of deletional information loss (**Figure 3.4A**) leads to a catastrophic loss of genetic information (**Figure 3.7**).

One of the main assumptions of the model is the existence of only one type of repeat sequence at high density. This is, obviously, an approximation of a much more complex scenario, with numerous repeats present at different densities. Frequent gene duplications in early eukaryotes contributed to the increase in repeat density, and it is estimated that the average copy number per gene in LECA was around 1.8 (Vosseberg *et al.*, 2021). In order to avoid unnecessary complexity, I only simulated the presence of one type of repeat at high density and began every simulation with only a single copy of each other gene. As a consequence, the number of repeat sequences introduced by gene duplication is small compared with the initial number of repeats. Their impact on the total deletion rate is negligible, and I did not explicitly evaluate their population dynamics. The results of this model can be easily generalised to include the presence of multiple repeats at lower density.

The main source of repeated sequences in prokaryotic genomes is the spread of transposable elements (TEs) and other selfish genetic elements, which can promote their own spread and reduce host's fitness through gene function disruption or gene inactivation (Touchon and Rocha, 2007; Wagner, 2006). As the focus of this study is on recombination and genetic information loss, I did not explicitly model the fitness cost of a high repeat density nor the advantage of a smaller genome. The only cost arising from a high repeat density is that associated with weaker purifying selection and a higher rate of deletion. Including the additional fitness penalty of high repeat density and large genome size is likely to increase the costs associated with large genomes, strengthening the conclusions of this study. In addition, the density of intron-derived sequences in ancestral eukaryotic genomes

is estimated to be as high as 80%, making the choice of $\rho = 0.1$ used in most simulations a conservative one (Koonin, 2009b).

I have not explicitly modelled crossing over, but assumed the recombination of a single, long eDNA sequence. Having multiple points of recombination would likely reduce selective interference and enhance the benefits of recombination. I will evaluate the magnitude of this effect and other potential benefits of crossing over in future theoretical studies.

In prokaryotes, TEs have a major impact on genome plasticity by increasing the rate of gene duplication, deletions, insertions and whole-genome rearrangements (Siguier *et al.*, 2014; Vandecraen *et al.*, 2017). They are generally found at low density despite their ability to promote their own spread, suggesting presence of counter-veiling selective pressures (Lambowitz and Belfort, 2015; Wagner, 2006). According to previous theoretical studies, the impact of repeats is inversely proportional to the number of different repeated sequences, and several repeats at low density are less impactful than a single repeat at very high density (Owen, Lane and Pomiankowski; unpublished data). The relevance of these dynamics should be evaluated more rigorously through further modelling studies.

As the model presented here focuses on the conservative benefits of LGT, I do not explicitly model the possibility of beneficial recombination events, e.g., those which contribute to the acquisition of novel genes. This is in part because the acquisition of eDNA via transformation is, in most instances, tightly regulated and requires a high degree of sequence homology (de Vries and Wackernagel, 2002; Frye *et al.*, 2013; Hülter and Wackernagel, 2008; Mell *et al.*, 2012). While LGT via plasmids is the main source of acquisition of novel genes, transformation is mainly limited to sequences from more closely related lineages. In the model, gene loading can occur through the acquisition of genes lost

through deletion. However, in the presence of a high repeat density, this effect is negligible and cannot stop the loss of genes because of deletions.

I assume that point mutations do not significantly affect the probability of homologous recombination and used a simple assumption about homology requirements for successful recombination, i.e., homology of the terminal gene sequence of eDNA. I also adopted a simple function to determine where the trailing end of eDNA binds (see **Eq(3.2)**) which penalises extremely large deletions, in agreement with experimental data showing they are rare (Croucher *et al.*, 2012; Mell *et al.*, 2014). To my knowledge, no data is available on the distribution of recombination length as a function of repeats density, and I deliberately ignored this complexity to avoid over complication of the modelling. Further empirical investigations can contribute to an understanding of the generality and accuracy of the assumptions made.

While the evolution of meiotic recombination prevented the accumulation of mutations in nuclear genomes, the endosymbiont was subjected to extremely stringent bottlenecks and limited scope for recombination. As discussed in Chapter 1, both theoretical models (Haigh, 1978; Muller, 1964) and experimental evolution with different bacterial species (Andersson and Hughes, 1996; Funchain *et al.*, 2000; Tenaillon *et al.*, 2016) indicate that these conditions aggravate the severity of Muller's ratchet. The integration of mitochondrial genes into the host's genome allowed the endosymbiont to undergo genome streamlining (Khachane *et al*., 2007; Martin *et al.*, 2015; Timmis *et al.*, 2004) – a potential way of escaping mutation accumulation by reducing genome size (Colnaghi *et al.*, 2020). A similar process of genome shrinkage can be observed in extant endosymbionts (López-Madrigal and Gil, 2017) and under experimental conditions (Nilsson *et al.*, 2005). In extant eukaryotes, the number of mitochondrial protein-coding genes is generally lower in taxa

with higher mitochondrial mutation rates (Allio *et al*., 2017; Burger *et al*., 2003; Fauron *et al*., 2004; Lavrov and Pett, 2016; Nabholz *et al*., 2009). Transfer of genes to the host allowed the endosymbiont to minimise its genome-wide mutation rate, limiting the rate of fixation of deleterious mutations (Colnaghi *et al.*, 2020). These dynamics will be discussed in greater detail in Chapter 5. In metazoans, subjected to higher mitochondrial rates than other organisms, genome streamlining is not enough to prevent Muller's ratchet (Loewe, 2006), and other mechanisms of purifying selection have evolved to curtail the transmission of deleterious mitochondrial mutations (Colnaghi *et al*., 2021), as I discuss in Chapter 4.

The transfer of endosymbiotic genes to the host's genomes is thought to have also led to the spread of group II self-splicing introns, an additional source of genomic repeats (Cavalier-Smith, 1991; Koonin, 2006; Martin and Koonin, 2006; Rogozin *et al.*, 2012). In this study, I have not explicitly modelled the spread of selfish genetic elements; instead, I assumed that the frequency of repeated sequences can only increase through insertions during LGT. This assumption is plainly conservative, as the presence of selfish genetic elements would further promote the increase in repeated sequences density. This, in turn, would make LGT even more detrimental, strengthening the selective pressure for the evolution of meiotic sex.

The selective advantages provided by meiotic sex, however, do not automatically ensure its evolution. The model developed here and in the previous chapter compares two states (LGT and sexual reproduction), rather than addressing the more complex problem of the transition from one to the other. In spite of the need to minimise ectopic recombination end ensure purifying selection in large eukaryotic genomes, determining the selective advantages of homolog pairing, the evolution of meiosis could have in principle been difficult or even unlikely. Nevertheless, the following argument yields some theoretical

support to the ideas presented in this chapter and the previous. In the presence of tight linkage (as in populations subjected to Muller's ratchet), modifiers that increase the rate of recombination are predicted to spread, not only because of their benefits at the population level, but also because they directly contribute to increasing the fitness of the offspring of individuals carrying the modifier (Otto, 2021; Roze, 2014; Iles et al., 2003). Early eukaryogenesis was characterised by a sharp increase in the rate of mutation accumulation, mainly due to genome size expansion (Chapter 2). In the presence of a high repeat density, a high frequency of LGT is detrimental (Chapter 3), potentially favouring a decrease in recombination rate. This would lead to even higher mutation loads through Muller's ratchet, tight genetic linkage, and strong selective interference: exactly the conditions under which sexual recombination is predicted to spread and reach fixation in a population (Otto, 2021; Roze, 2014; Iles et al., 2003). Obviously, the problem of the transition from LGT to sex is complicated by a number of likely intermediate steps – cell fusion, ploidy cycles, mitosis. In future publications, I will elucidate in greater details how each of these steps provides some selective advantage in the presence of strong selective interference and evaluate which conditions could have led to their evolution during eukaryogenesis.

6. Conclusion

Early eukaryotes evolved through a massive genome size expansion with a sharp rise in repeated sequence density (Lambowitz and Belfort, 2015; Lane, 2020; Martin *et al.*, 2015; Vosseberg *et al.*, 2021). Such genome size expansion cannot take place without a concomitant increase in the severity of Muller's ratchet, and an increase in recombination frequency and length is necessary to prevent the mutational degradation of early eukaryotic genomes (Colnaghi *et al.*, 2020). But in the presence of repeats, increasing recombination

length or LGT frequency is not a viable way of countering mutation accumulation. Higher rates of LGT or longer eDNA length amplify the loss of genetic information caused by ectopic recombination and curtail the capacity of LGT to prevent mutation accumulation. The increase in recombination length and frequency must be coupled with the requirement for homology throughout the whole eDNA sequence, as in meiotic recombination. The need for increased purifying selection arising from genome expansion in early eukaryotes could not be met by LGT, because of the concomitant increase in genomic repeats. This selective pressure promoted the evolution of cell fusion and whole-chromosomes alignment followed by homologous recombination (Goodenough and Heitman, 2014; Lane, 2011) – in other words, meiotic sex. Meiotic sex allows recombination to take place over a considerable fraction of each chromosome, promoting an increase in L of a magnitude that can prevent the decay of eukaryotic genomes predicted by Muller's ratchet. At the same time, the transition to linear chromosomes and whole-chromosomes alignment minimises the risk of ectopic recombination and the associated loss of genetic information. Crossing over also combines aligning whole chromosomes with multiple points of recombination, with the added benefit of maximising recombination length while reducing selective interference.

The origin of meiosis is tightly linked to the process of eukaryogenesis. Its evolution was crucial for the survival and evolution of early eukaryotes, as it allowed them to maintain an expanded and repeat-rich genome in the face of crippling mutation pressure. Without it, complex life as we know it could have not survived its inception.

Chapter 4 – The Need for High-Quality Oocyte Mitochondria at Extreme Ploidy dictates Mammalian Germline Development[4]

1. Summary

Selection against deleterious mitochondrial mutations is facilitated by germline processes, lowering the risk of genetic diseases. How selection works is disputed: experimental data are conflicting and previous modelling work has not clarified the issues. Here I develop computational and evolutionary models that compare the outcome of selection at the level of individuals, cells and mitochondria. Using realistic *de novo* mutation rates and germline development parameters from mouse and humans, the evolutionary model predicts the observed prevalence of mitochondrial mutations and diseases in human populations. I show the importance of organelle-level selection, seen in the selective pooling of mitochondria into the Balbiani body, in achieving high-quality mitochondria at extreme ploidy in mature oocytes. Alternative mechanisms debated in the literature, bottlenecks and follicular atresia, are unlikely to account for the clinical data, because neither process effectively eliminates mitochondrial mutations under realistic conditions. These findings explain the major features of female germline architecture, notably the longstanding paradox of over-proliferation of primordial germ cells followed by massive loss. The near-universality of these processes across animal taxa makes sense in light of the need to maintain mitochondrial quality at extreme ploidy in mature oocytes, in the absence of sex and recombination.

[4]A revised version of this chapter has been published as Colnaghi, M., Pomiankowski, A., and Lane, N. (2021). The need for high-quality oocyte mitochondria at extreme ploidy dictates mammalian germline development. Elife *10*, e69344.

2. Introduction

In mammals, mitochondrial gene sequences diverge at 10-30 times the mean rate of nuclear genes (Allio *et al.*, 2017; Lynch *et al*., 2006). This difference is typically ascribed to a faster underlying mutation rate and limited scope for purifying selection on mitochondrial genes, given uniparental inheritance, negligible recombination and high ploidy (Rand, 2001). At face value, weak selection against mitochondrial mutations might seem to be consistent with the high prevalence of mitochondrial mutations (~1 in 200) (Elliott *et al*., 2008) and diseases (~1 in 5000 births) (Schaefer *et al*., 2008) in human populations. But it is not consistent with the strong signal of purifying selection (Fonseca *et al.*, 2008), evidence of adaptive change (James *et al.*, 2016) and codon bias (Yang and Nielsen, 2008) in mitochondrial genes, nor with the low transmission rate of severe mitochondrial mutations between generations (Fan *et al.*, 2008; Hill *et al.*, 2014; Stewart *et al.*, 2008). Despite the high rate of sequence divergence, female germline processes apparently facilitate selection against mitochondrial mutations, but the mechanisms are disputed and poorly understood (Burr *et al*., 2018).

Here I develop computational and evolutionary models that compare three hypotheses of germline mitochondrial inheritance and selection: (i) selection at the individual level, facilitated by mitochondrial bottlenecks; (ii) selection at the cell level through follicular atresia, which weeds out primordial germ cells with high mutation load; and (iii) selection at the organelle level through selective transfer of mitochondria into oogonia during development. These modes of selection are not mutually exclusive nor the only ones that operate (a full examination is provided in the Discussion). The objective here is to clearly delineate the major forces involved and the effectiveness of each in controlling mutation accumulation. This approach incorporates important factors neglected in earlier

work, in particular the input of *de novo* mitochondrial mutations and their segregation over multiple rounds of germ-cell division. This provides a realistic model of mutation, segregation and selection allowing the three hypotheses to be tested against the observed levels of mitochondrial mutation and fitness across a variety of species with an emphasis on the detailed clinical reports in human populations (Elliott *et al.*, 2008; Schaefer *et al.*, 2008).

The idea that mitochondrial mutations are winnowed through a tight germline bottleneck is pervasive in the literature and has long been held to explain sharp changes in mutation load between generations (Floros *et al*., 2018; Johnston *et al.*, 2015; Stewart and Chinnery, 2015). The exact size of the bottleneck is unclear, with estimates from mouse (Cao *et al*., 2009; Cree *et al*., 2008; Wai *et al*., 2008) and human studies (Floros *et al.*, 2018; Guo *et al*., 2013; Li *et al*., 2016; Rebolledo-Jaramillo *et al*., 2014) spanning two orders of magnitude. Bottlenecks generate variance in mutation loads among the resulting germ cells, and tighter bottlenecks produce greater variance, offering scope for selection against mitochondrial mutations at the level of the individual (Bergstrom and Pritchard, 1998; Hadjivasiliou *et al*., 2013; Roze *et al*., 2005). The problem with this line of thinking is that it ignores two other forces. First, gametes are produced through multiple rounds of cell division, leading to repeated rounds of mitochondrial segregation, which in itself generates considerable variance (Radzvilavicius *et al*., 2017). Second, bottlenecks induce greater input of *de novo* mutations as more rounds of mitochondrial replication are required to regenerate the extreme ploidy of mitochondrial DNA in mature oocytes. By applying realistic segregation dynamics and mutational input, I evaluate the impact of these forces on the value of bottleneck size on individual fitness.

Follicular atresia is another force widely considered to be critical in maintaining oocyte quality (Chu *et al*., 2014; Haig, 2016; Krakauer and Mira, 1999) . In humans

(Albamonte *et al*., 2008; Kaipia and Hsueh, 1997), the number of germ cells declines dramatically in the foetus between mid-gestation (~20 weeks in humans) when there are 7-8 million oocytes, to late gestation when at least two thirds of these are lost, leaving a reserve of 1-2 million at birth (Townson and Combelles, 2012). Oocyte loss continues throughout the life of an individual, eventually leading to the depletion of the ovarian pool and loss of reproductive function at menopause (Cummins, 2004; Galimov *et al*., 2014; Suganuma *et al*., 1993). Similar loss of female germ cells before sexual maturity is evident in mice and several other animal species (Morita and Tilly, 1999; Nezis *et al*., 2000; Rodrigues *et al*., 2009; Saidapur, 1978). This attrition has historically been ascribed to cell death during oocyte maturation (Perez *et al*., 2000; Tilly, 2001), but more recent findings implicate the apoptotic loss of 'nurse cells' during the genesis of primary oocytes (Lei and Spradling, 2016). In either case, differential oocyte loss offers scope for between-cell selection. However, the basis for between-cell selection has long been questioned, on the grounds that it seems unlikely that 70–80% of oocytes have low fitness as a result of mitochondrial mutations (Perez *et al.*, 2000). I therefore test whether selection against oocytes with higher loads of mitochondrial mutations during follicular atresia is capable of giving rise to the distribution of mutations observed.

A more recent interpretation of germ-cell loss links it to the formation of the Balbiani body, a prominent feature of the humans (Hertig, 1968; Motta *et al*., 2000) and mouse (Kloc *et al*., 2004; Lei and Spradling, 2016; Reunov *et al*., 2019; Tworzydlo *et al*., 2016) female germline, as well as a range of other vertebrates and invertebrates, with a range of terminology (e.g. fusome, mitochondrial cloud) (Heasman *et al*., 1984; Kloc *et al.*, 2004; Larkman, 1984; Lei and Spradling, 2016; Reunov *et al.*, 2019; Tworzydlo *et al.*, 2016). In the mouse, proliferating germ cells typically form clusters of 5-8 cells that establish cytoplasmic

bridges (Lei and Spradling, 2016; Pepling *et al*., 2007). It is thought that around half the mitochondria from each nurse cell are streamed into the Balbiani body of the primary oocyte, through an active cytoskeletal process that depends in part on the membrane potential of discrete mitochondria (Bilinski *et al*., 2017; Zhou *et al*., 2010). This offers scope for purifying selection through the preferential exclusion of dysfunctional mitochondria. The remaining nurse cells, now denuded of half their mitochondria, undergo apoptosis (Lei and Spradling, 2016). Selective transfer and pooling of mitochondria from interconnected cells may occur in other vertebrate and invertebrate systems. I consider the consequence of different strengths of selection at the level of mitochondrial function in the production of the Balbiani body.

To systematically distinguish between the predictions of these three different hypotheses, under a range of reasonable parameter values, I use a computational model to evaluate the patterns of mutation load generated over a single generation in each case. I then use an evolutionary model to generate equilibrium levels and compare the predictions to the prevalence of mutations and disease from human studies. Our results show that selection at the organelle level through the pooling of high-quality mitochondria into the Balbiani body is a more potent force than germline bottlenecks and follicular atresia and must play a key role in the maintenance of mitochondrial function in the face of pervasive mutational pressure. This analysis also pleasingly clarifies the longstanding paradox of germ-cell over-proliferation followed by massive loss which is a widely conserved feature of the female germline in animal taxa.

3. Methods

3.1. Computational Model

I use a computational model to follow the distribution of mitochondrial mutations in the female germline over a single generation from zygote to a new set of mature oocytes, as set out in the developmental history given in the main text (**Figure 4.1A**). The initial state of the system is a zygote containing $M_0 = 2^{19} = 524{,}288$ copies of mtDNA (Reynier *et al*., 2001), of which m_0 carry a deleterious mutation. Three specific models are considered: bottleneck, follicular atresia and cytoplasmic transfer. A list of terms and parameter values is given in **Table 4.1**, which also apply in the evolutionary model considered below.

During early embryonic development, there is no mtDNA replication. The number of cells doubles at each time step. The existing population of mutant and wildtype mtDNA undergoes random segregation into daughter cells according to a binomial distribution – each mtDNA copy has a 50% probability of being assigned to either daughter cell. During this process, the average number of mtDNA copies per cell halves at each time step. There is no mutational input, as I only consider mutations that arise due to replication errors.

The early embryonic period lasts for the first 12 cell divisions. A group of 32 cells is selected at random to form the primordial germ cells (PGC). The PGCs then undergo proliferation for a further 18 rounds of cell division, until the maximum number of germ cells is reached, $N_{max} = 32 \times 2^{18} = 8{,}388{,}608$. This value is close to the average reported in the literature (Albamonte *et al.*, 2008).

Parameters and variables	**Symbols and values**
Maximum number of germ cells	$N_{max} = 8{,}388{,}608$
mtDNA number in mature oocytes	$M_0 = 2^{19} = 524{,}288$
Minimum mtDNA ploidy	B

Final number of germ cells	$N_{max}/8 = 1{,}048{,}576$
Initial mutation load	m_0
Mutation rate per bp per cell division	μ
Strength of epistatic interactions	ξ
Transfer probability of mutant mtDNA	p_{mut}
Transfer probability of wildtype mtDNA	p_{wt}
Human mitochondrial genome size	$g = 16{,}569$bp (Palca, 1990)

Table 4.1

Mitochondrial (i.e. mtDNA) replication resumes after cell division 12, at the point of PGC determination. At this point, cells have an average of 128 mtDNA copies. At each following time step, the number of mtDNA copies doubles prior to random segregation into daughter cells. This means that the average number of mtDNA copies per cell is kept constant. New mutations are introduced as errors in mitochondrial replication. During the replication process, the new replica of each wildtype mtDNA copy has a probability of mutation μ/bp. The genome wide mutation rate $U = g \times \mu$ is calculated as genome size ($g = 16{,}569$ bp (Palca, 1990)) multiplied by μ. This estimate assumes each site contributes equally to selective effects and ignores many subtleties relating to mutation probability and within-cell maintenance processes, but should give a reasonable order of magnitude estimate of the target size of mutational input per cell division. Given n wildtype and m mutant mtDNAs, the number of new mutants Δm resulting from replication errors is obtained by sampling at random from a binomial distribution with n trials with probability U. After replication and mutation and prior to segregation the total number of wildtype and

mutant mtDNAs is $2n - \Delta m$ and $2m + \Delta m$ respectively. Back mutation to wildtype is not permitted.

At the end of PGC proliferation, the N_{max} oogonia undergo random cell death, leaving $N_{max}/8 = 1{,}048{,}576$ primary oocytes. This is achieved by sampling the surviving cells at random with uniform weights (i.e., every cell has an equal probability of survival). The primary oocytes do not undergo further cell division or mitochondrial replication during the quiescent period (this is not explicitly modelled). At puberty, oocyte maturation commences. The number of mitochondria per cell is brought back to the original value $M_0 = 2^{19}$ through 12 rounds of replication without cell division. I assume that the number of mtDNA copies doubles at each time step. This introduces new deleterious mutations, which again are randomly drawn from a binomial distribution (as described above).

I consider three specific models in the main text with modifications to the base model described above.

The first model adds a bottleneck stage at the time of PGC determination (**Figure 4.2**). As before, 32 cells are selected at cell division 12 to form the PGCs. These go through b extra rounds of cell division without mtDNA replication. This reduces the mean number of mtDNA copies per cell to $\bar{B} = 128 * (0.5)^b$. The mtDNA replication commences at cell division 14. The PGCs then proliferate as before to produce oogonia that undergo random cell death to produce primary oocytes. The primary oocytes have a reduced number of mtDNA copies, and so must undergo $12 + b$ extra rounds of mtDNA replication in order to regain the original value M_0 mitochondria in mature oocytes. Note that this is an extreme model of the bottleneck, where mtDNA copy number is kept low throughout the period of PGC proliferation, and so maximises the benefit derived from the increase in segregational variation caused by the bottleneck.

For the model of the bottleneck, I allow selection dependent on individual fitness in relation to their mutation load m among mature oocytes, according to the fitness function $f(m) = 1 - (\frac{m}{M})^5$ (**Figure 4.2C**). The concave shape of this function accounts for the fact that mitochondrial mutations typically have a detrimental effect on individual fitness only for loads >60%. Changes to the power exponent make little qualitative difference to the outcome of this model (data not shown).

A second model considers non-random death during the cull of oogonia as these cells transition to being primary follicles (**Figure 4.3**). Selection in this case is applied at the cell level. Cell fitness is expressed as $f(m) = 1 - (\frac{m}{M})^{\xi}$, where m is the number of mutant mitochondria. The parameter ξ determines the strength of epistatic interactions (**Figure 4.3B**). As in other models, the number of cells is reduced from $N_{max} = 8{,}388{,}608$ to $N_{max}/8 = 1{,}048{,}576$. This is achieved by sampling without replacement the surviving cells at random, with weights proportional to cell fitness (i.e., every cell has a probability of survival proportional to its fitness).

The third model considers that the oogonia are organised in cysts of 8 cells each. These are the descendants of a single cell (*i.e.* three cell divisions prior). One cell is randomly designated as the primary oocyte using the MATLAB function randsample. The Balbiani body of the primary oocyte contains a proportion f of the mtDNA copies of all cells in the cyst. The mitochondria that join the Balbiani body are sampled at random without replacement from each cell with different weights for wildtype (p_{wt}) and mutant (p_{mut}). After mitochondrial transfer to the Balbiani body, nurse cells undergo apoptosis (i.e. all cells except the one designated as the primary oocyte), reducing the total number of oocytes to $N_{max}/8 = 1{,}048{,}576$.

3.2. Evolutionary Model

In order to calculate the equilibrium distribution of a population undergoing the developmental dynamics mentioned in the previous section, I develop an analytical model for the distribution of mitochondrial mutations in an infinite population, with non-overlapping generations. As it was not possible to find an analytical solution, I solved the equations through numerical iterations. The system converges to a unique equilibrium state, independent of the initial conditions.

The state of the system is described by the vector $p(t) = \{p_0(t), \dots, p_{M(t)}(t)\}$, where $M(t)$ is the number of mtDNA copies per cell at time t. The elements $p_m(t)$ are the frequency of mutation load $m(t)/M(t)$ at time t. The evolution of the system is determined by a set of transition matrices whose elements are the transition probabilities between states. To avoid unnecessary complexity in the evolutionary model, I assume that fluctuations in mitochondrial number per cell due to segregation are negligible (*i.e.* in contrast to the computational model which allows binomial segregation at each division). Therefore, the mtDNA number per cell is constant across the whole population of cells at every time step. That is, during early embryonic development (when there is no mtDNA replication), after t cell divisions, the total number of mitochondria per cell is $M^{(t)} = 2^{-t}M_0$. Then, during PGC proliferation, the total number of mitochondria per cell is constant. Finally, during oocyte maturation, the number of mitochondria per cell exactly doubles with each mtDNA replication cycle. To aid in calculations, I also set the initial number of mtDNA copies to be proportional to the bottleneck size, *i.e.* $M_0 = 2^{12} \times B$. As the mtDNA number per cell halves at each cell division during early embryonic development, setting M_0 this way allows the mtDNA number to remain an integer. This is

important for the modelling procedure, because the dimension of the transition matrixes (which is determined by the mtDNA number) must be an integer.

During early embryonic development, when mitochondrial replication is not active, changes in frequency arise purely from the process of segregation. Let $W^{(t)}$ be a $M^{(t)} + 1 \times M^{(t)} + 1$ square matrix, whose elements W_{mn} represent the transition probabilities from a state with m to a state with n mutants:

$$W_{mn}^{(t)} = \binom{m}{n}\binom{M^{(t-1)} - m}{M^{(t)} - n} \Big/ \binom{M^{(t-1)}}{M^{(t)}} \tag{4.1}$$

These matrix elements model the probability of transitioning from a state with m mutants and $M - m$ wildtype to a state with n mutants and $M - n$ wild type via the segregation of $2M$ mitochondria into two daughter cells with M mitochondria each.

After t cell divisions, the average number of mutants per cell is $\overline{m} = 2^{-t} m_0$, and the variance is $Var(t) = \frac{1}{4}[Var(t-1) + 2^{-t} m_0]$. The state of the system is updated as $\vec{p}^{(1)} = \left(\prod_t W^{(t)}\right) \times \vec{p}^{(0)}$.

During PGC proliferation, new mutations are introduced at a rate $U = \mu \times g$. The transition coefficient Q_{mn} from a state with m to a state with n mutants results from the combined effects of replication, mutation and segregation:

$$\begin{aligned} Q_{mn} &= \sum_k \binom{M-n}{k-n} U^{k-n} (1-U)^{M-k} \binom{k}{m}\binom{M-k}{M-m} \Big/ \binom{2M}{M} \\ &= \sum_k \binom{M-n}{k-n} U^{k-n} (1-U)^{M-k} a_{k,m} \end{aligned} \tag{4.2}$$

The coefficient $a_{k,m} = \binom{k}{m}\binom{M-k}{M-m}/\binom{2M}{M}$ models the probability of transitioning from a state with k mutants and $M-k$ wildtype to a state with m mutants and $M-m$ wild type via the segregation of $2M$ mitochondria into two daughter cells with M mitochondria each; the remaining part of the equation models the probability of reaching a state with k mutant mitochondria through replication and mutation of M mitochondria, of which n are mutant (this corresponds to the probability of introducing $k-n$ new mutations). The system is updated $\vec{p}^{(2)} = Q^q \times \vec{p}^{(1)}$, across q rounds of PGC cell division. I then apply particular processes to capture the effects of the bottleneck, follicular atresia and cytoplasmic transfer.

As before, I model the bottleneck as b extra rounds of segregation before the onset of mtDNA replication, following **Eq(4.1)** with $q+b$ cell divisions. This has no effect on the mean mutational number but increases mutational variance between the resulting PGCs. The transition between oogonia and primary oocytes occurs at random, and so does not alter the frequency distribution of mutants. Finally, during oocyte maturation, the mtDNA content of each cell doubles at every time step until the initial ploidy M_0 is restored. The transition matrix G_{mn} is analogous to the first term of **Eq(4.2)**, incorporating replication and mutation, but without segregation (last term of **Eq(4.2)**):

$$G_{mn}^{(t)} = \binom{M^{(t)} - m}{n - m} U^{n-m}(1-U)^{M-k} \tag{4.3}$$

Eq(4.3) models the probability of transitioning from a state with m to a state with n mutants, which is equivalent to the probability that exactly $n - m$ out of $M^{(t)} - m$ wildtype acquire a deleterious mutation. As the bottleneck reduces mtDNA copy number per cell, there is the need for b extra rounds of replication of mtDNA during oocyte maturation. Hence, the transition coefficient G is applied $b + 12$ times in the bottleneck model, to restore the number of mtDNA copies per oocytes to the original ploidy level M_0: $\vec{p}^{(3)} = \left(\prod_t G^{(t)}\right) \times \vec{p}^{(2)}$.

At the end of the maturation phase, for the bottleneck model, selection is applied on individual fitness using a vector w whose elements w_m are equal to the corresponding fitness: $w_m = f(m) = 1 - \left(\frac{m}{M}\right)^5$. This causes a change in the population mutation load as the system is updated to:

$$\vec{p}^{(3)} = (I\vec{w})\, \vec{p}^{(2)} / \vec{w}^T\, \vec{p}^{(2)} \tag{4.4}$$

where I is the identity matrix.

In the model of follicular atresia, an extra step is included to reflect selection that operates when the population of oogonia are culled to produce the primary oocytes. This causes a change in the population mutation load analogous to that described in **Eq(4.4)**, but using the cell fitness function $w_m = f(m) = 1 - \left(\frac{m}{M}\right)^{\xi}$ instead. This determines the shift in mutation loads that arises from fitness-dependent culling of oogonia. The transition coefficient **Eq(4.3)** for oocyte maturation is then applied 12 times in the follicular atresia model, to restore the original level of ploidy.

Finally, in order to model cytoplasmic transfer, a different process is used in the production of primary oocytes. A set of 8 clonally derived cells is selected. The mutation levels of each cell in the cyst is obtained by applying **Eq(4.3)** three times. Then, 50% of the mitochondria in each cell are pooled into the Balbiani body of the primary oocyte. The probability for a cell with m mutants to contribute n mutants to the Balbiani body is given by:

$$C_{mn} = \binom{m}{n} p_{mut}^{n} \binom{M-m}{M/2\ -n} p_{wt}^{M/2-n} / N \tag{4.5}$$

which gives the number of permutations of n mutant and $M/2\ -n$ wildtype mtDNA copies, weighted by the probability of transfer $p_{mut}$and p_t respectively, and divided by a normalisation constant N. As the primary oocyte contains half of mitochondria from 8 cells, it needs to undergo 2 fewer rounds of replication during oocyte maturation. Hence only 10 rounds of replication following **Eq(4.3)** are carried out in this case to restore the original level of ploidy.

For all three models (bottleneck, follicular atresia and cytoplasmic transfer), the frequency distribution of mutation loads after these steps is used as the starting point for the next generation.

The processes described above are iterated until the Kullback-Leibler divergence (a theoretical measure of how two probability distributions differ from each other (Kullback and Leibler, 1951)) between the new and the old distribution is smaller than a threshold $\eta = 10^{-9}$. I then assume that the system has reached a stationary state, *e.g.* without significant changes in the overall distribution of mutation loads between generations (mutation-selection balance).

	Frequency	Reference
Mitochondrial mutations	$\alpha_1 = 5.4 \times 10^{-3}$	Elliott *et al.*, 2008
Mitochondrial diseases	$\alpha_2 = 9.2 \times 10^{-6}$	Schaefer *at al.,* 2008

Table 4.2: Occurrence of mitochondrial diseases and mutations in the general population.

In order to compare the prediction of the model with the clinical data (**Table 4.2**), I use the equilibrium distribution to calculate the fraction of the population which carries a detectable load of mitochondrial mutations but does not manifest any detrimental phenotype (α_1) and the fraction of individuals affected by mitochondrial disease (α_2) using a threshold of 60% mutation load to discriminate between carrier and disease status. Individuals are assumed to be mutation free beyond the detection threshold of 2% (Elliott *et al.*, 2008).

The accuracy of the model is evaluated as the logarithm of the probability of reproducing clinical data by sampling the theoretical distribution at random. This is calculated as follows: let X_1 be the number of healthy individuals with detectable mutation load, and X_2 be the number of individuals affected by mitochondrial diseases; N_1 and N_2 the total number of individuals in the two trials; α_1 and α_2 the probability of observing, respectively, a healthy individual with detectable mutation load and an individual affected by mitochondrial disease, according to the prediction of the model. The log-likelihood of observing X_1 and X_2 by random sampling the theoretical distribution is given by:

$$\log\left(\lambda_{tot}(\mu, M, \dots) = log\left[\prod_{i=1}^{2} p\left(X_i \middle| \alpha_i(\mu, M, \dots)\right)\right]\right.$$

$$= \sum_{i=1}^{2} \log\left[\binom{N_i}{X_i} \alpha_i^{N_i}(1-\alpha_i)^{N_i - X_i}\right] \tag{4.6}$$

$$= \sum_{i=1}^{2} log\binom{N_i}{X_i} + N_i \log(\alpha_i) + (N_i - X_i) log(1-\alpha_i)$$

3.3. Estimation of the deleterious mutation rate

The parameter values for the deleterious mutation rate I investigate reflect data collected from a number of species. Estimates of mtDNA point mutation rates in the crustacean *Daphnia pulex* range between 1.37×10^{-7} and 2.28×10^{-7} per site per generation (Xu *et al*., 2012). Assuming this rate applies to humans and there are ~ 20 cell divisions before oocyte maturation, leads to a range between 0.68×10^{-8} and 1.14×10^{-8} per site, per cell division. Analysis of *Caenorhabditis elegans* mtDNA leads to a similar estimate of $\sim 1.6 \times 10^{-7}$ per site, per generation (Denver *et al*., 2000), which corresponds to a rate of 0.8×10^{-8} per site, per cell division. For *Drosophila melanoganster*, the mtDNA mutation rate yields an estimate of 6.2×10^{-8} per site, per generation, and hence $\sim 0.31 \times 10^{-8}$ per site, per cell division (Haag-Liautard *et al*., 2008). Finally, analysis of human mtDNA point mutation rates give a mutation rate of 0.0043 per genome per generation (Sigurdardottir *et al.*, 2000), corresponding to $\sim 1.3 \times 10^{-8}$ mutations per site, per cell division.

These values do not take into account the presence of a number of processes likely to remove mutants and is therefore a conservative estimate. The loss of mutations would mean that the actual mutation rate is higher than the estimates above. But unlike nuclear

rates, the compact structure of mtDNA where intergenic sequences are absent or limited to a few bases, means that the rate of point mutations is probably not much higher than the rate of deleterious mutations. Therefore, for this study I consider a broad interval of possible deleterious mutation rates, labelled as low (10^{-9}), standard (10^{-8}) and high (10^{-7}).

The model described above has been implemented using MATLAB. Random sampling is achieved using the function 'randsample'. All random processes described above have been simulated using MATLAB's functions 'rand' and 'random'. The code used for the simulations can be found on GitHub (https://github.com/MarcoColnaghi1990/colnaghi-pomiankowski-lane-elife-2021).

4. Results

4.1. Computational model

The computational model follows the distribution of mitochondrial mutations across a single generation, using model parameters derived from human data (Albamonte *et al.*, 2008) (**Figure 4.1**). The zygote is assumed to have around half a million copies of mitochondrial DNA (exact number 2^{19}), which are randomly partitioned to the daughter cells at each cell division. The pattern of segregation is in agreement with recent evidence for actin-mediated mixing of mitochondria within cells during mitosis leading to random segregation (Moore et al., 2021). I assume independent segregation of mitochondria with one mtDNA per mitochondrion, and do not consider complications that might arise from the packaging of multiple mtDNA copies per mitochondrion (Floros *et al.*, 2018). This assumption is supported by evidence that mitochondrial networks fragment into multiple

smaller structures at cell division (Park and Cho, 2012; Taguchi *et al*., 2007) that probably contain only one or a few mtDNAs.

Mitochondrial replication is not active during early embryo development (Dumollard *et al*., 2007), so the mean mitochondrial number per cell approximately halves with each division (**Figure 4.1B**). In humans, after 12 cell divisions a random group of 32 cells form the primordial germ cells (PGC) (Extavour and Akam, 2003), which in the model corresponds to a mean of 128 mitochondria per PGC. Mitochondrial replication resumes at this point (Albamonte *et al.*, 2008; Dumollard *et al.*, 2007). Each mtDNA doubles prior to cell division. With probability μ, one of the daughter mitochondria acquires a new deleterious mutation through a copying error. I consider μ in the range 10^{-9} to 10^{-8} to 10^{-7} per base pair per cell division (designated low, standard and high respectively), consistent with the range of estimates for the female germline, and assume no back mutations (see Methods). Point mutations during replication are the dominant form of mutation in mtDNA, so I do not consider damage from other sources such as oxidative damage (Stewart and Larsson, 2014). Mitotic proliferation of PGCs gives rise to ~8 million oogonia, which are reduced to ~1 million primary oocytes during late gestation (**Figure 4.1B**) (Albamonte *et al.*, 2008; Dumollard *et al.*, 2007). Proliferation is followed by a quiescent phase during which the mitochondria in primary oocytes are not actively replicated. Mutations accumulate far more slowly during this phase, which persists over decades in humans (Allen and de Paula, 2013; Dumollard *et al.*, 2007). For simplicity, I assume no mutational input during this period (not marked in **Figure 4.1B**). At puberty, the primary oocytes mature through clonal amplification of mitochondria back to the extreme ploidy in mature oocytes (~500,000 copies; **Figure 4.1B**) (Radzvilavicius *et al*., 2016). The same copying error mutation rate μ is applied during this process.

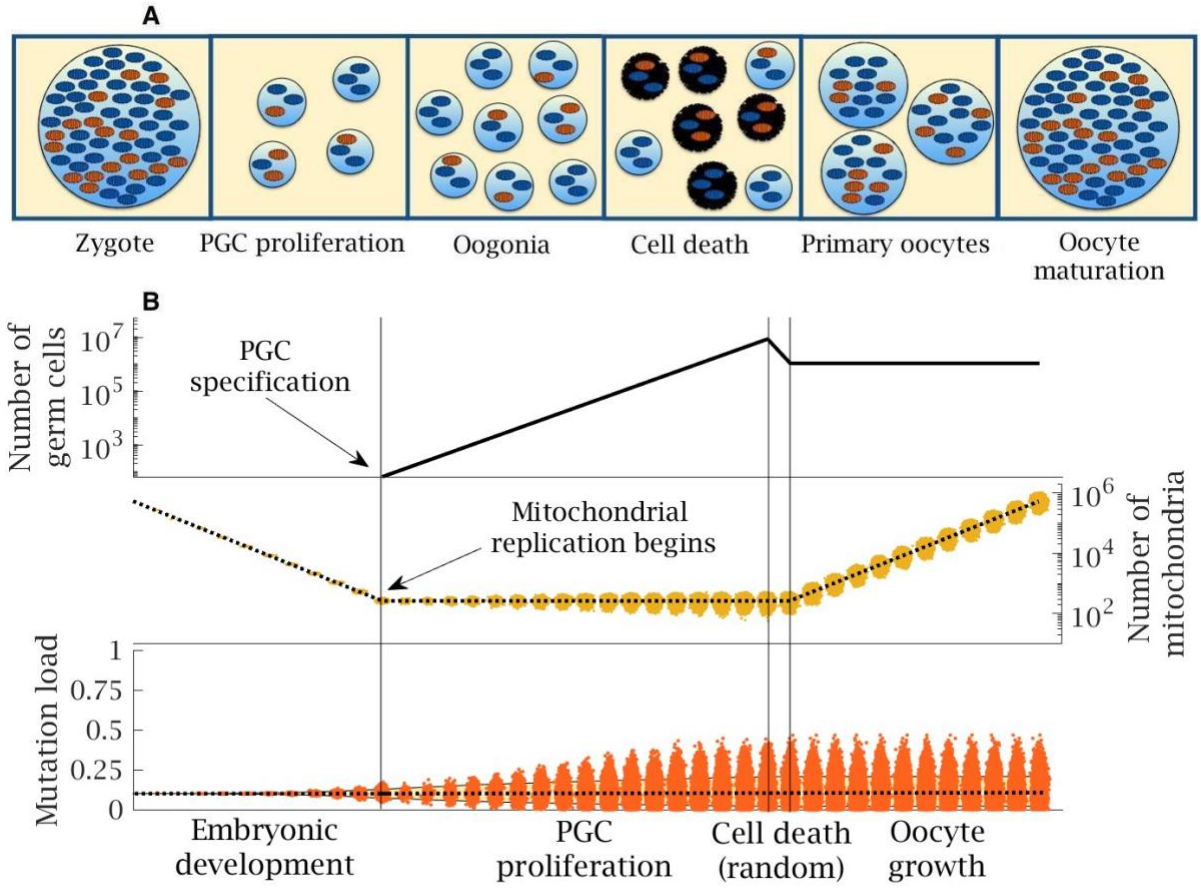

Figure 4.1. **Stages in female germline development.** (**A**) Timeline of human oocyte development showing the main stages modelled, with wildtype (blue) and mutant mitochondria (orange). (**B**) Numerical simulation of the base model. Top panel: number of germ cells from specification of the 32 primordial germ cells (PGCs) after 12 cell divisions; proliferation to form 8 million oogonia; random cell death reducing to 1 million primary oocytes; quiescent period (not shown) and finally oocyte maturation at puberty. Middle panel: copy number of mitochondria (i.e. mtDNA); from zygote with ~500,000 copies, which are partitioned at cell division during early embryo development until replication begins (first vertical line) during PGC proliferation; copy number is amplified during oocyte maturation back to ~500,000 copies; dotted line shows the mean mitochondria copy number, with the distribution across oocytes shown in yellow. Note, skew reflects the log-scale. Bottom panel: mean (dotted line) and distribution of mutation load through development. The yellow shaded area shows the 90% quantile. Other parameter values $\mu = 10^{-8}$, $m_0 = 0.1$.

I consider three different forms of selection on mitochondria: selection at the level of individuals, cells, or mitochondria. I apply selection at the level of individuals on the

zygotic mutation load. Selection at the level of cells or mitochondria is applied during culling at late gestation when primary oocytes are produced. Each of these processes can be captured by modifications of the computational model, allowing easy comparison between them. In order to distinguish between different levels of selection, the model extends earlier work that considered segregational variation of a fixed burden of existing mutations (Bergstrom and Pritchard, 1998; Hadjivasiliou *et al.*, 2013; Johnston *et al.*, 2015; Roze *et al.*, 2005) but neglected the input of new mutations during PGC proliferation and oocyte maturation, as well as the loss of germ cells during late gestation. The analysis here shows the importance of considering these additional processes governing the population of mitochondria in germline development.

The effect of a bottleneck was assessed in the model by allowing b extra rounds of cell division without mitochondrial replication during early embryonic development (e.g., two extra rounds shown in **Figure 4.2A**). Each additional cell division leads to an average reduction of $(0.5)^b$ mitochondria in PGCs compared to the base model. For simplicity I then hold mitochondrial numbers at this lower value through the period of PGC proliferation. This assumption gives greater impact to the bottleneck and is consistent with some views (Cao *et al.*, 2009). Tighter bottlenecks at this early developmental stage generate greater segregational variance in mutation load between cells (**Figure 4.2B**). This increase in variance persists and is enhanced through PGC proliferation to the production of primary oocytes and ultimately in mature oocytes (**Figure 4.2B**). The bottleneck not only creates a wider spread of mutation number per cell, but also the possibility that cells can be mutation free even when initiated from a zygote that contains significant numbers of mutations (**Figure 4.2B**). Bottlenecks in themselves do not change the mean mutation load, as they occur before the start of mitochondrial replication (i.e. at PGC specification; **Figure 4.2B**)

(Dumollard *et al.*, 2007). But oocyte maturation requires the expansion of mitochondrial number back to half a million. Cells starting with lower numbers must therefore undergo more rounds of mitochondrial replication, and hence will accumulate more *de novo* mutations. So, the mean mitochondrial mutation load in mature oocytes increases with tighter bottleneck size, albeit this effect is small with standard mutation rates ($\mu = 10^{-8}$; **Figure 4.2B**). Nonetheless, the tension between variance and mean determines the overall selective consequence of the bottleneck.

The advantage that the bottleneck brings depends on how selection acts against the mutation load carried by an individual. Based on the observed dependence of mitochondrial diseases on mutation load (Kopinski *et al*., 2019; Rossignol *et al*., 2003; Wallace and Chalkia, 2013), in which more serious phenotypes typically manifest only at high mutation loads of >60 % (Kopinski *et al.*, 2019; Rossignol *et al.*, 2003; Wallace and Chalkia, 2013), it is thought that individual fitness is defined by a concave fitness function, indicative of negative epistasis (**Figure 4.2C**). This assumes that each additional mitochondrial mutation causes a greater reduction in fitness beyond that expected from independent effects. In other words, low mutation loads have a relatively trivial fitness effect, whereas higher mutation loads produce a steeper decline in fitness.

The change in mutation load (Δm) over a single generation after individual selection was measured against 5 mean bottleneck sizes ($\bar{B} = 128, 64, 32, 16, 8$), for three initial mutation loads (m_0) and three mutation rates (μ). The bottleneck shows an ambiguous relationship with fitness, dependent on the inherited mutation load (m_0). For the standard mutation rate ($\mu = 10^{-8}$) there is always an increase in mutation load in individuals who inherit low or medium mutation loads ($m_0 = 0.001, 0.01$; **Figure 4.2D**). This increase in load ($\Delta m > 0$) becomes more deleterious with a tighter bottleneck (**Figure 4.2D**). The bottleneck

only confers a benefit ($\Delta m < 0$) among individuals who inherit a high mutation load ($m_0 = 0.1$; **Figure 4.2D**), where the advantage of greater variance outweighs the increase in *de novo* mutation load. If the mutation rate is lower ($\mu = 10^{-9}$), bottlenecks have little effect except when severe, where they again cause an increase in mutation number in individuals with low or medium mutation loads ($m_0 = 0.001, 0.01$; **Figure 4.S1A**). In individuals with high mutation load ($m_0 = 0.1$) only tighter bottlenecks ($\bar{B} = 16, 8$) are beneficial (**Figure 4.S1A**). If the mutation rate is higher ($\mu = 10^{-7}$) the pattern is more extreme, with the accumulation of *de novo* mutations except in individuals with high inherited mutation loads ($m_0 = 0.1$) at the tightest bottleneck size ($\bar{B} = 8$) (**Figure 4.S1B**).

In sum: even though bottlenecks generate greater variance, they impose the need for additional rounds of mitochondrial replication during oocyte maturation, resulting in greater *de novo* mutational input. This makes tight bottlenecks advantageous only for rare individuals who inherit high mutation loads, but not for the great majority of the population where the prevalence of mitochondrial mutations is below the limits of detectability, between 0.001 and 0.01 (Elliott *et al.*, 2008; Floros *et al.*, 2018).

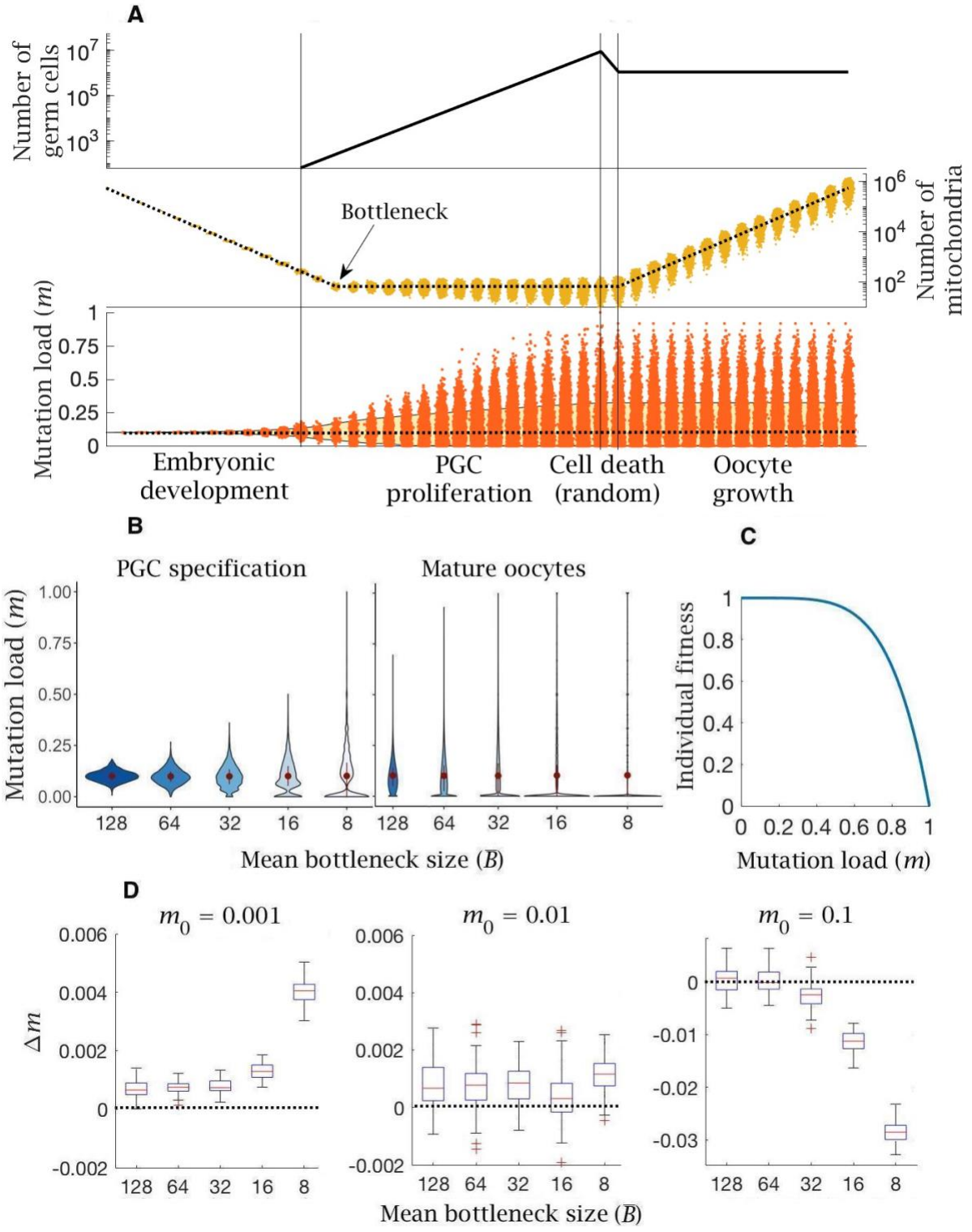

Figure 4.2. **Model of germline bottleneck and individual selection.** (**A**) A bottleneck with two extra rounds of cell division without replication (cell division 13 and 14; after the first vertical line), reducing mitochondria copy number per PGC (by a quarter on average). Two extra rounds of mitochondrial replication are required to regenerate the copy number in mature oocytes. Compared to the base model (**Figure 4.1**), mean mutation load (dotted line, bottom panel) is slightly higher and variation in load is substantially greater (yellow shaded area, 90% quantile). Parameter values $\mu =$

10^{-8}, $m_0 = 0.1$. (**B**) Violin plots of the distribution of mutations (mean ± SD shown in red) at two developmental stages, PGC specification and mature oocytes, given 5 mean bottleneck sizes ($\bar{B}$) when $m_0 = 0.1$. (**C**) Strength of selection on individual fitness, with a concave fitness function based on clinical data from mitochondrial diseases. (**D**) Change in mutation load (Δm) across a single generation for three initial mutation loads (m_0), given 5 mean bottleneck sizes ($\bar{B}$), showing the median (red line) and distribution (box plot IQR with min/max whiskers and outliers) over 10 independent simulations.

In the analysis of bottlenecks above, the culling of ~8 million oogonia to ~1 million primary oocytes at the end of PGC proliferation was assumed to be a random process (**Figure 4.2A**). This loss has a minimal effect on the mean and variance of mitochondrial mutations in germ cells, given the large numbers involved (and no effect at all when averaged over a population). However, the loss of ~80% of oocytes via follicular atresia during late gestation has long been puzzling and could arguably reflect selection against cells with higher mutation loads.

To analyse follicular atresia, cell-level selection was applied to oogonia at the end of PGC proliferation (**Figure 4.3A**). PGCs vary in mutation frequency due to both the random segregation of mutants during the multiple cell divisions of proliferation and the chance input of new mutations during mtDNA replication. In principle, I assume that between-cell selection is governed by a negative epistatic fitness function (**Figure 4.3B**) similar to that thought to apply at the individual level, and vary selection from linear ($\xi = 1$), weak ($\xi = 2$) to strong epistasis ($\xi = 5$). Positive epistasis ($\xi < 1$), whereby a single point mutation produces a steep loss of fitness, but additional mutations have less impact (i.e. mutations are less deleterious in combination), seems biologically improbable, so I do not consider it here.

The effect of cell selection during follicular atresia was calculated as the change in mutation frequency for individuals carrying different mutation loads (m_0) over a single generation, given standard values for *de novo* mutations ($\mu = 10^{-8}$) and bottleneck size ($\bar{B} = 128$). Under strong negative epistasis ($\xi = 5$), only the few cells with very high mutation loads (generated by segregation) are eliminated. Cell-level selection does not reduce mutation load, even for individuals with a high initial frequency of mutations ($m_0 = 0.1$; **Figure 4.3C**). Cell-level selection is more effective with weak epistasis ($\xi = 2$) or linear selection ($\xi = 1$) as this makes cells with lower mutation loads more 'visible' to selection, and has a greater benefit in individuals carrying higher initial mutation loads (**Figure 4.3C**). However, in individuals who inherit low or medium mutation load ($m_0 = 0.001, 0.01$) cell selection offers a minimal constraint against mutation input. The only case in which cell selection produces a benefit is with high mutation load ($m_0 = 0.1$) under linear selection ($\xi = 1$) (**Figure 4.3C**). This pattern holds for a lower mutation rate ($\mu = 10^{-9}$; **Figure 4.S2A**), while there is no benefit at all at a higher mutation rate ($\mu = 10^{-7}$; **Figure 4.S2B**).

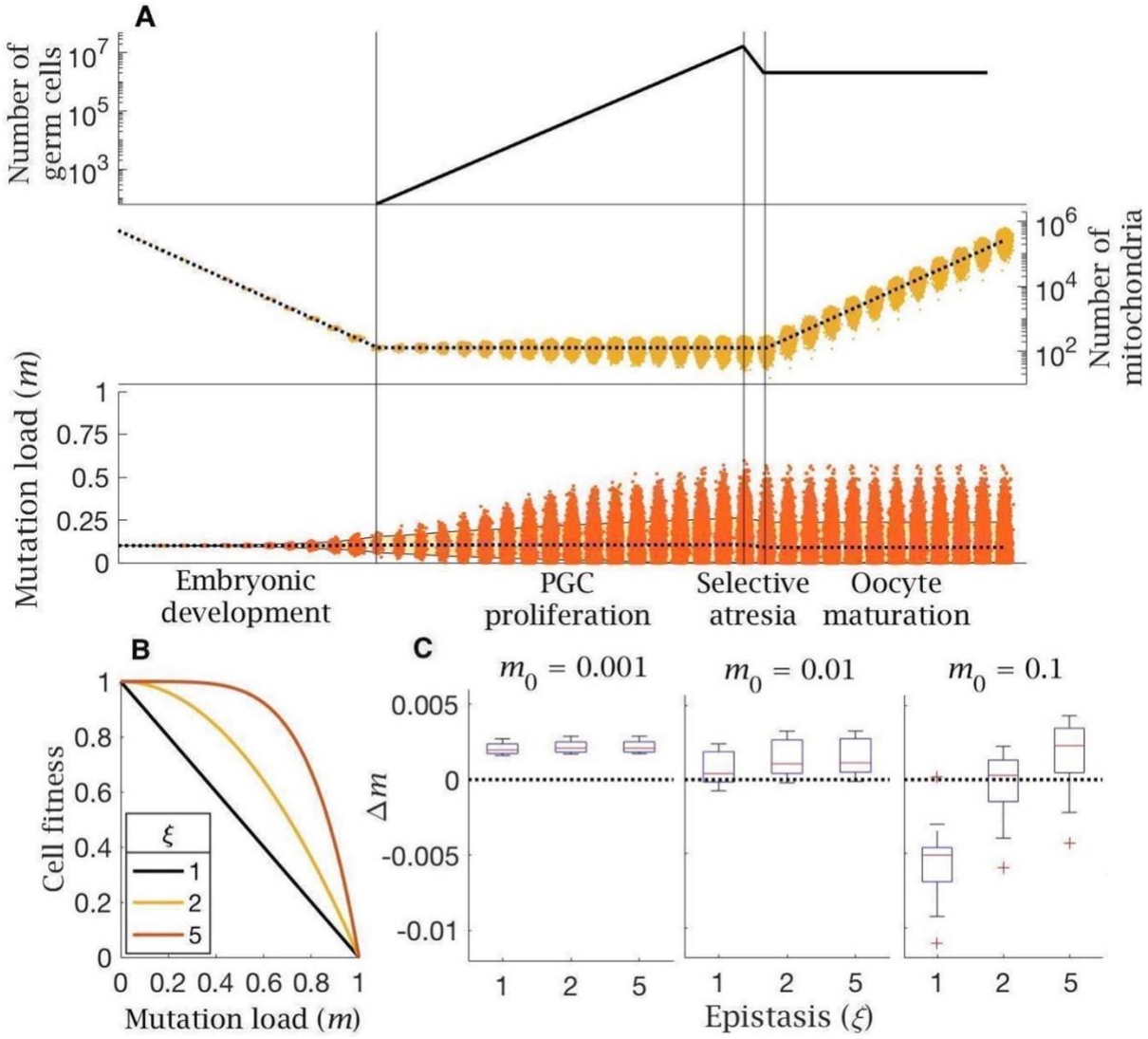

Figure 4.3. Model of follicular atresia and cell selection. (**A**) After PGC proliferation, follicular atresia occurs through selective apoptosis of oogonia. (**B**) Cell fitness is assumed to be linear ($\xi = 1$) or follow negative epistasis ($\xi = 2, 5$) in which mutations are more deleterious in combination. (**C**) Change in mutation load, Δm, across a single generation after cell selection, at an intermediate mutation rate ($\mu = 10^{-8}$), for individuals with low ($m_0 = 0.001$), medium ($m_0 = 0.01$) and high ($m_0 = 0.1$) initial mutation loads, for variable levels of epistasis (median (red line) and distribution (box plot IQR with min/max whiskers and outliers)) over 10 independent simulations.

An alternative interpretation of atresia in late gestation lies in the formation of the Balbiani Body. I model the developmental process giving rise to the Balbiani body by assuming that cysts of 8 oogonia form at the end of PGC proliferation (**Figure 4.4A**). In the model, cells

within a cyst are derived from a common ancestor (i.e. via 3 consecutive cell divisions). At the 8-cell stage, intercellular bridges form between the oogonia. These allow cytoplasmic transfer of a proportion of mitochondria (f) from each cell to join the Balbiani body of the single cell destined to become the primary oocyte (**Figure 4.4A**). The mitochondria that undergo cytoplasmic transfer are sampled at random (without replacement), with different weights for wildtype (p_{wt}) and mutant (p_{mut}) mitochondria, until a fraction f of them have moved to the Balbiani body. The oogonia that donate their cytoplasm to the primary oocyte are now defined as nurse cells, and undergo programmed cell death – atresia (**Figure 4.4A**).

The model shows that two benefits accrue from cytoplasmic transfer. The first benefit of mitochondrial transfer into the Balbiani body is that pooling increases the number of mitochondria in primary oocytes. As the proportion of mitochondria transferred increases towards the estimated rate of f = 50% (Lei and Spradling, 2016), the number of mitochondria in primary oocytes increases 4-fold. Pooling therefore cuts the number of rounds of replication needed to reach the extreme ploidy required by mature oocytes, which decreases the input of new mutations from replication errors during oocyte maturation. This benefit accrues whatever the initial mutation load, and more dramatically with a higher mutation rate (**Figure 4.S3**).

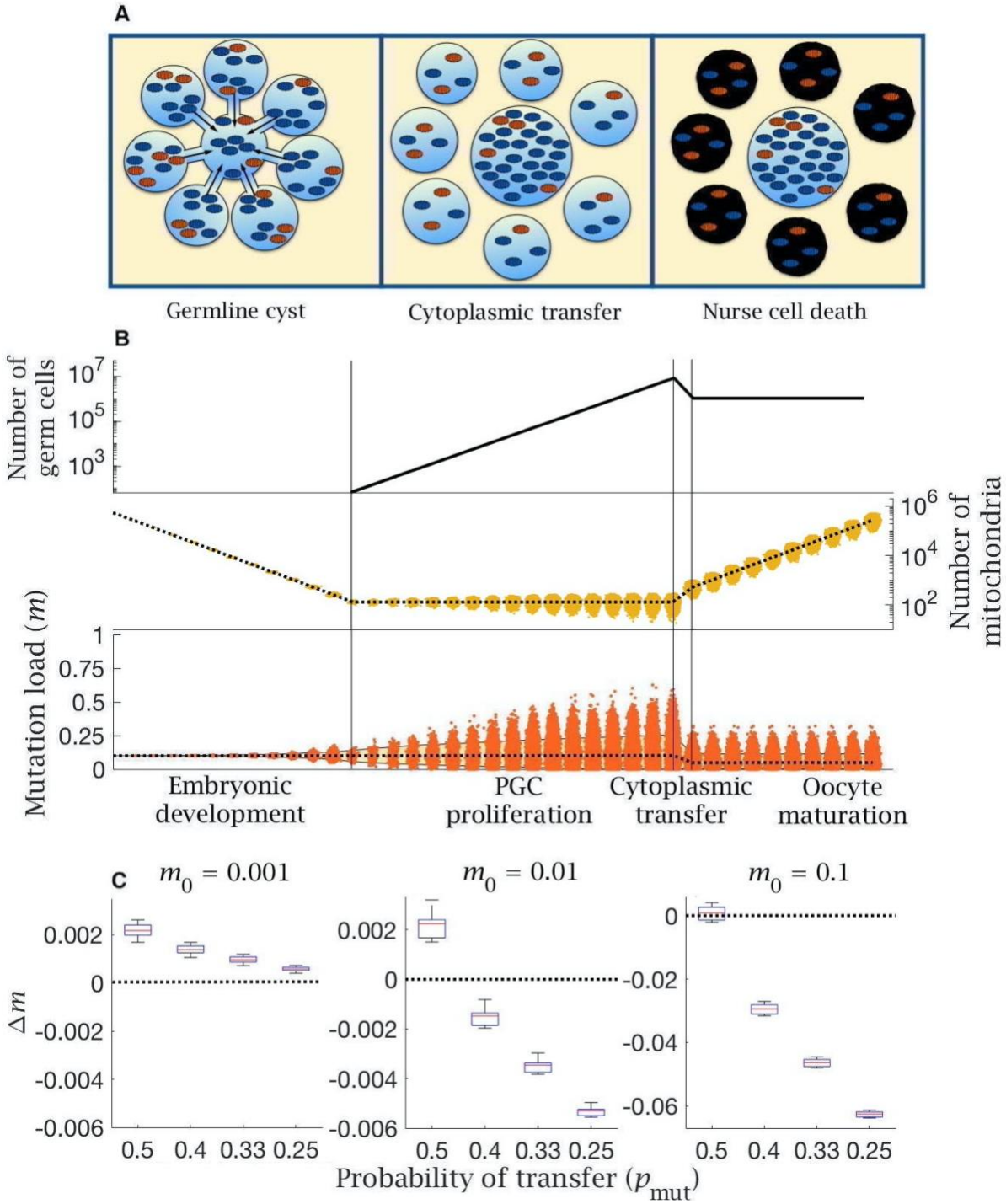

Figure 4.4. Model of cytoplasmic transfer and mitochondria selection. (**A**) Cytoplasmic bridges form among oogonia in the germline cyst, leading to selective transfer of wild-type mitochondria (blue) to the primary oocyte, leaving mutant mitochondria (red) in nurse cells that then undergo apoptosis. (**B**) Cytoplasmic transfer which selectively pools $f = 50\%$ of mtDNA from 8 germline cyst cells into a single primary oocyte causes a large increase in the number of mitochondria (middle panel) and a large reduction in the mean (dotted line, bottom panel) and distribution of mutation load (yellow shaded area shows the 90%

quantile, bottom panel), which persists during oocyte maturation. Pooling of mtDNA requires two fewer rounds of mtDNA replication to regenerate copy number in mature oocytes. Parameter values $\mu = 10^{-8}$, $m_0 = 0.1$. (**C**) Change in mutation load (Δm) across a single generation (median (red line) and distribution (box plot IQR with min/max whiskers and outliers)) over 10 independent simulations, for individuals with low ($m_0 = 0.001$), medium ($m_0 = 0.01$) and high ($m_0 = 0.1$) initial mutation loads, with variable strengths of selective transfer (p_{mut}). Parameter value $\mu = 10^{-8}$.

The second benefit arises from selective transfer of mitochondria. Preferential exclusion of mutant mitochondria ($p_{wt} > p_{mut}$), as suggested by experimental evidence (Bilinski *et al.*, 2017; Chen *et al*., 2020; Lei and Spradling, 2016), lowers the mutation load in primordial oocytes (**Figure 4.4B**). The difference between p_{wt} and p_{mut} determines the extent to which the mutation load is reduced, with stronger exclusion of mutant mitochondria (lower p_{mut}) reducing the number of mutations when the inherited load is medium or high ($m_0 = 0.01, 0.1$), albeit with a negligible effect at low initial mutation load ($m_0 = 0.001$; **Figure 4.4C**). The same effect is seen with lower and higher mutation rates (**Figure 4.S4**). Nurse cells retain a higher fraction of mutant mitochondria but undergo apoptosis, removing mutants from the pool of germ cells, and explaining the need for an extreme loss of germ cells during late gestation. This effect acts in concert with pooling leading to a reduction in both the mean and variance of mitochondria mutation load in the cells destined to develop into mature oocytes. (**Figure 4.4B**).

4.2. Evolutionary model

The computational model discussed above gives an indication of the effectiveness of selection at the level of individuals, cells or mitochondria in eliminating mitochondrial mutations across a single generation. To address the long-term balance of mutation

accumulation versus selection over many generations, I developed an evolutionary model. This assesses the effectiveness of the three representations of germline development in explaining the observed prevalence of mitochondrial mutation load and disease in human populations (see Methods). This mathematical model evaluates long-term evolutionary change in an infinite population with non-overlapping generations and is implemented using a number of approximations, which greatly reduce the model complexity (see Methods).

By iterating the patterns of germline inheritance and selection, the equilibrium mutation distribution was calculated across a range of mutation rates and bottleneck sizes. The accuracy of the three models was then assessed as the likelihood of reproducing the observed levels of mitochondrial mutations in the human population (**Figure 4.5**). Specifically, I used estimated values of approximately 1/5000 for mitochondrial disease (>60% mutant), 1/200 for carriers of mitochondrial mutants (2-60% mutant) and hence 99.5% of individuals are 'mutation free' (i.e. carry <2% mutants, the threshold for detection in these estimates of mutation frequency) (Elliott *et al.*, 2008; Schaefer *et al.*, 2008). Recent deep-sequencing estimates using a mutation detection threshold of >1% (Floros *et al.*, 2018), show that a minor allele frequency of 1-2% is relatively common in selected human PGCs, but this does not alter earlier population-level estimates of the proportion of carriers not suffering from overt mitochondrial disease, defined as a 2-60% mutation load used here. A high likelihood indicates good agreement between model predictions and clinical data. The broader the parameter range that produces good agreement between data and model predictions, the larger the overall yellow/orange area in **Figure 4.5**. This area is indicative of the likelihood that a given level of selection (between-individual, between-cell, and organelle-level) might play a prominent role in determining the level of transmission of mitochondrial mutations observed in human populations.

Likelihood heatmaps confirm that selection at the level of individuals or cells alone do not readily approximate the clinical data whatever the bottleneck size (**Figure 4.5A-B**).

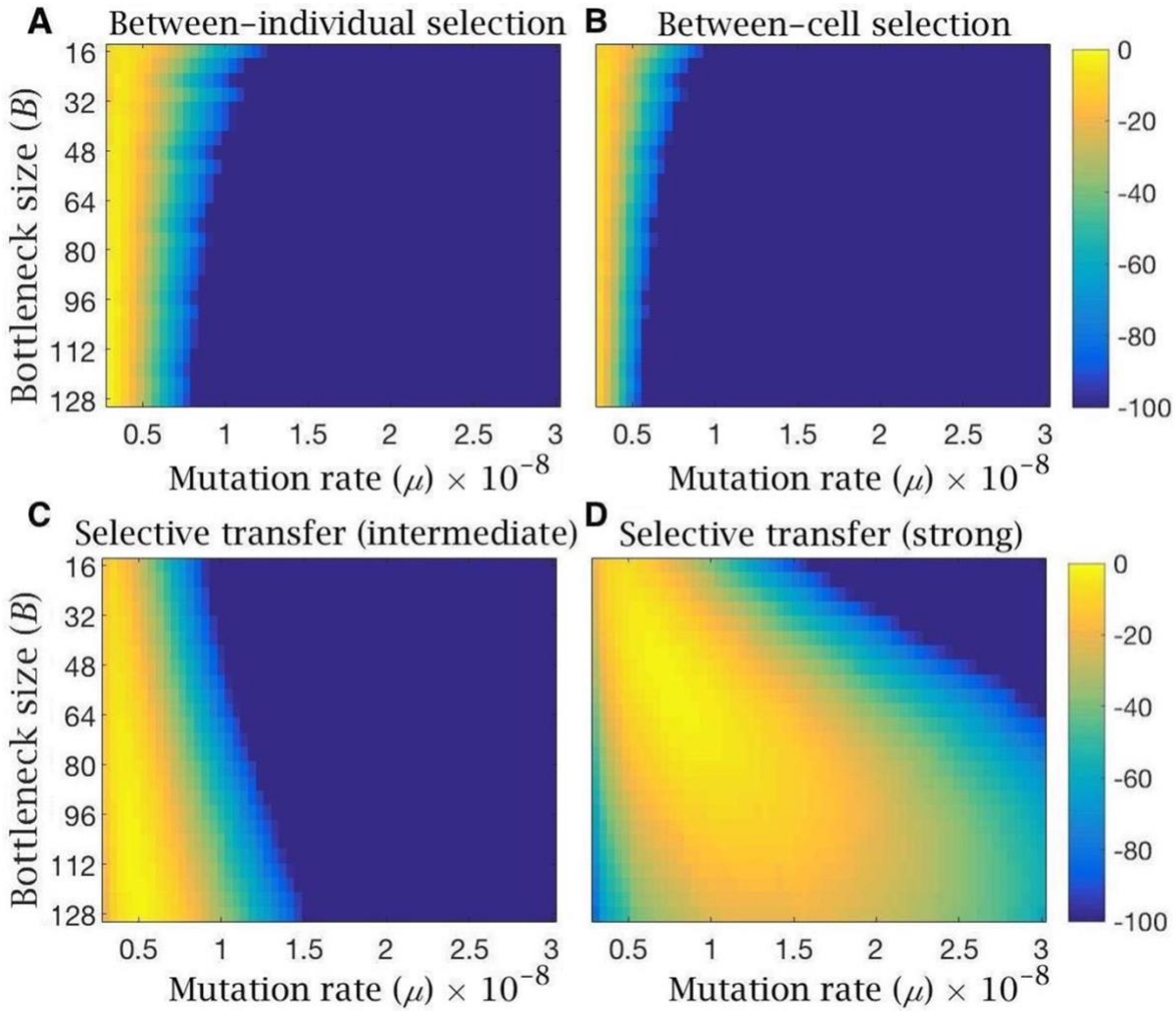

Figure 4.5. Heatmaps showing log-likelihood of reproducing the observed mutation load and disease frequency in humans, for equilibrium conditions under the evolutionary model with (**A**) bottleneck and selection on individuals, (**B**) follicular atresia and selection on cells ($\xi = 5$), (**C**) cytoplasmic transfer with intermediate ($p_{mut} = 0.33, p_{wt} = 0.67$) or (**D**) strong ($p_{mut} = 0.25, p_{wt} = 0.75$) selective transfer of wildtype mitochondria. Yellow depicts high likelihood, indicating that the model is in good agreement with clinical data; blue indicates a poor agreement between the model and clinical data. All models are shown for variable bottleneck size (the minimum mitochondria population size at which replication commences) and variable mutation rates. The overall yellow/orange area is proportional to the width of the parameter range that leads to good agreement between clinical data and simulations.

Only at a mutation rate $\mu < 5 \times 10^{-9}$ do these forms of selection explain the observed mutation load and disease frequency in humans at high likelihood, especially when using tighter bottlenecks (**Figure 4.5A**). These limitations do not apply to the preferential transfer of wildtype mitochondria into the Balbiani body (**Figure 4.5C-D**). Even intermediate levels of selection against the transfer of mutant mitochondria into the Balbiani body ($p_{mut} = 0.33$, $p_{wt} = 0.67$) generates a high log-likelihood of reproducing the clinical data at the standard mutation rates ($\mu = 10^{-8}$) and bottleneck sizes (> 100 mitochondria per cell) (**Figure 4.3C**). Stronger selection on transfer probabilities ($p_{mut} = 0.25$, $p_{wt} = 0.75$) can account for the clinical pattern under a wide range of bottleneck sizes and mutation rates (**Figure 4.5D**).

5. Discussion

How selection operates on mitochondria has long been controversial. At the heart of this problem is the paradox that mtDNA accumulates mutations faster than nuclear genes, yet there is evidence that mtDNA is under strong purifying selection. Mitochondrial mutations accumulate through Muller's ratchet, as mtDNA is exclusively maternally inherited, and does not undergo recombination through meiosis (Rand, 2001). In addition, mitochondrial genes are highly polyploid, which obscures the relationship between genotype and phenotype, hindering the effectiveness of selection on individuals. Despite these constraints, deleterious mitochondrial mutations seem to be eliminated effectively (Fan *et al.*, 2008; Fonseca *et al.*, 2008; Hill *et al.*, 2014; James *et al.*, 2016; Stewart *et al.*, 2008; Yang and Nielsen, 2008), facilitated by female germline processes that have long been mysterious. These include: the excess proliferation of primordial germ cells (PGCs) (Pepling, 2006); the germline mitochondrial bottleneck (when mitochondrial numbers are reduced to

a disputed minimum in PGCs) (Floros *et al.*, 2018; Johnston *et al.*, 2015; Stewart and Chinnery, 2015); the formation of the Balbiani body in primary oocytes (Bilinski *et al.*, 2017; Lei and Spradling, 2016); the atretic loss of 70-80% of germ cells during late gestation (Tilly, 2001; Townson and Combelles, 2012); the extended oocyte quiescence until puberty or later (during which time mitochondrial activity and replication is suppressed) (Allen and de Paula, 2013; de Paula *et al*., 2013); and the generation of around half a million copies of mtDNA in mature oocytes (Radzvilavicius *et al.*, 2016). The key question is how do these processes facilitate the maintenance of mitochondrial quality over generations?

In this study, I introduced a computational model that considers these germline processes from the perspective of mitochondrial proliferation, segregation and selection, using realistic estimates of parameter values, drawn from the human literature (Albamonte *et al.*, 2008; Dumollard *et al.*, 2007). Most work to date (Cao *et al.*, 2009; Cree *et al.*, 2008; Johnston *et al.*, 2015; Stewart and Chinnery, 2015; Wai *et al.*, 2008) has focused on the mitochondrial bottleneck as a means of generating variation in mitochondrial content between oocytes and by extension zygotes (**Figure 4.2B**), providing the opportunity for selection to act on individuals in the following generation. These studies have been unable to reconcile serious differences in experimental estimates of mitochondrial numbers during PGC proliferation, inciting inconclusive debates over the tightness of the bottleneck (Cao *et al.*, 2009; Cree *et al.*, 2008; Johnston *et al.*, 2015; Stewart and Chinnery, 2015; Wai *et al.*, 2008). More significantly, this earlier work neglects an important germline feature, the introduction of *de novo* mitochondrial mutations produced by copying errors (Arbeithuber et al., 2020) rather than damage by reactive oxygen species (Stewart and Larsson, 2014; Trifunovic et al., 2005). These accumulate during PGC proliferation and, equally importantly, during the mass-production of mtDNAs in the mature oocyte. Tighter bottlenecks are

disadvantageous as they impose the need for more rounds of mitochondrial replication which means a greater input of *de novo* mutations. Our modelling shows that for most individuals the mean mutation load shows little meaningful change (**Figure 4.2D**), regardless of whether the mutation rate is set low or high (**Figure 4.S1**), and in fact increases with tighter bottleneck size (**Figure 4.2D**). Most individuals have low mutation loads (~99.5% in human populations (Elliott *et al.*, 2008; Schaefer *et al.*, 2008)), and for them, the normal process of repeated segregation during cell division generates sufficient variance in itself. Any marginal increase in variance caused by bottlenecks is more than offset by increased mutational input. Tighter bottlenecks only benefit individuals who already carry high mutation loads (i.e. $m_0 \geq 0.1$, **Figure 4.2D**). For them, there is benefit in further reductions in bottleneck size as this increases the fraction of mature oocytes with significantly reduced mutation load (**Figure 4.2D**). In the modelling, I assumed that the bottleneck size was maintained across the period of PGC proliferation. Some studies have found that from a low number in early development, copy number increases 5-10 fold to production of the oogonia (Cree *et al.*, 2008; Wai *et al.*, 2008). This would lessen the effect of the bottleneck in general as it would have less effect on segregation.

These results show that the popular idea that a germline mitochondrial bottleneck facilitates selection against mitochondrial mutations is misconstrued. The value of a bottleneck depends on the unforeseen trade-off between increasing genetic variance and mutation input. In fact, the reduction in mitochondrial copy numbers from zygote to primordial germ cells should be thought of as the reestablishment of a typical copy number at the start of cellular differentiation, which commences after multiple cell divisions *without* mtDNA replication. What counts as a bottleneck are the 'extra' rounds of cell division reducing mitochondrial number below the 'normal' number, and the incremental increase

in variance this induces. Most critically, the bottleneck needs to be understood in relation to oogamy, the massively exaggerated mitochondrial content of the female gamete. This is a characteristic of metazoan gametogenesis (Radzvilavicius *et al.*, 2016). Previous work has shown it is beneficial in animals with mutually interdependent organ systems (Radzvilavicius *et al.*, 2016). The extreme ploidy in the zygote allows early rounds of cell division to occur without mitochondrial replication, and hence without *de novo* mutational input. These initial cell divisions generate little between-cell differences, as segregational variance is weak when numbers are high and mitochondria segregate randomly during mitosis (Moore *et al.*, 2021) (e.g. **Fig 1B** before PGC specification). So at the point of cellular differentiation (~12 cell divisions) there is homogeneity in the mutation load among the different organ systems and no one system is likely to fail, which would massively lower the fitness of the whole organism (Radzvilavicius *et al.*, 2016). This contrasts with organisms that have modular growth, such as plants and morphologically simple metazoa (sponges, corals, placozoa), which neither sequester a recognizable germline early in development (although recent work challenges this view (Lanfear, 2018)), nor have oocytes with massively expanded mitochondrial numbers (Radzvilavicius *et al.*, 2016).

Follicular atresia is another female germline feature examined in the model, in which there is over-proliferation of PGCs followed by ~80% loss early in development, before oocyte maturation (Tilly, 2001; Townson and Combelles, 2012). This massive reduction in germ cell number has long been enigmatic. It is unlikely to be random, yet does not obviously serve a selective function, as it seems unlikely that such a high proportion of germ cells could have low fitness (Chu *et al.*, 2014; Haig, 2016; Krakauer and Mira, 1999). The model confirms this intuition. Selection among PGCs at the end of the period of proliferation has little effect in significantly reducing mutation load (**Figure 4.3C**). Assuming

a concave fitness function (**Figure 4.3B**), which seems reasonable by extension from the severity of mitochondrial diseases (Rossignol *et al.*, 2003; Wallace and Chalkia, 2013), between-cell selection is ineffective, as it only eliminates PGCs with very high mutation loads. This has little effect in constraining the build-up of mutations in less loaded cells. Linear selection does better, even if it seems unrealistic, as it will act against a broader range of mutational states. But as with bottlenecks, it is only beneficial in individuals already carrying significant mutation loads (i.e. $m_0 \geq 0.1$, **Figure 4.3C**). These results show that cell-level selection produces little measurable reduction in mutation load and so is unlikely to be the main mechanism behind follicular atresia.

A more recent explanation of PGC loss relates to the formation of the Balbiani body in primary oocytes (Kloc *et al.*, 2004; Lei and Spradling, 2016). In many metazoa, including clams (Reunov *et al.*, 2019), insects (Cox and Spradling, 2003; Tworzydlo *et al.*, 2016), mice (Pepling *et al.*, 2007) and probably humans (Kloc *et al.*, 2004), the over-proliferation of PGCs culminates in their organization into germline cysts of multiple oogonia connected by cytoplasmic bridges (Cox and Spradling, 2003; Lei and Spradling, 2016; Pepling *et al.*, 2007). These connections allow the transfer of mitochondria and other cytoplasmic constituents by active attachment to microtubules, into what becomes the primary oocyte (Lei and Spradling, 2016). The surrounding oogonia that transferred their mitochondria, now termed nurse cells, die by apoptosis (Lei and Spradling, 2016). The plethora of terms should not mask the key point that nurse cell death accounts for a considerable fraction of the germ cell loss usually ascribed to follicular atresia. I modelled selective mitochondrial transfer into the Balbiani body, perhaps in part reflecting membrane potential (Bilinski *et al.*, 2017; Tworzydlo *et al.*, 2016). This achieves two complementary benefits: it purges mutations and pools high-quality mitochondria in a single cell. If the germline cyst is composed of eight

cells that contribute half of their mitochondria to the Balbiani body, then the primary oocyte gains four times as many mitochondria which have passed through quality control. This also cuts the need for additional rounds of mtDNA copying, and so reducing the input of *de novo* mutations. Selective transfer and pooling lowers the mutation load across a wide range of mutation rates and inherited loads (**Figure 4.4C, Figure 4.S3-4**). This process differs from mitophagy, the main route used in somatic cells for maintaining mitochondrial quality (Kim *et al*., 2007; Twig *et al*., 2008), as it not only removes mutant mitochondria, but crucially also increases mitochondrial numbers, a key requirement for prospective gametes. The requirement for pooling of mitochondria to lower the mutation load from copying errors also aligns with experimental observations of active spindle-associated mitochondrial migration to the generative oocyte in the formation of polar bodies during meiosis I of oogenesis (Dalton and Carroll, 2013). I predict that selection for mitochondrial quality occurs during this process (i.e. polar bodies retain mutant mitochondria) but have not dealt with that explicitly in the model.

These insights depend in part on the parameter values used in the modelling, many of which are uncertain. I have examined variation around the most representative values drawn from the literature (Allio *et al.*, 2017; Floros *et al.*, 2018; Sigurdardottir et al., 2000; Stewart and Chinnery, 2015), and aimed to be conservative wherever possible. I considered mutation rates across two orders of magnitude, around 10^{-8} per bp as the standard (Sigurdardottir *et al.*, 2000) and a similar range of bottleneck sizes ($\bar{B} = 8 - 128$). Strong selective pooling of mitochondria into the Balbiani body predicts the observed prevalence of mitochondrial mutations and diseases in human populations (Elliott *et al.*, 2008; Schaefer *et al.*, 2008) under a wide range of mutation rates and bottleneck sizes (**Figure 4.5**). Selection at the level of individuals or cells are much more constrained explanations, although the

model does not rule out some role for these processes (**Figure 4.5**). In general, higher mutation rates (10^{-7} per base pair) strengthen the conclusions discussed here (**Figures S1-S4**) whereas the lowest mutation rates are more commensurate with weaker forms of evolutionary constraint generated by selection on individuals or cells. Plainly, weaker selection approximates best to clinical data when the mutation input tends towards zero (**Figure 4.5**). However, such low mutation rates are not consistent with the 10-30-fold faster evolution rates of mtDNA compared with nuclear genes (Allio *et al.*, 2017; Lynch *et al.*, 2006), or with the strong signatures of purifying (Fonseca *et al.*, 2008) and adaptive (James *et al.*, 2016) selection on mitochondrial genes. In the modelling, I ignored the contribution of oxidative damage caused by reactive oxygen species. While this source of mutation is likely low compared with copying errors (Arbeithuber *et al.*, 2020; Stewart and Larsson, 2014), oxidative mutations may accumulate over female reproductive lifespans (Trifunovic *et al.*, 2005), perhaps contributing to the timing of the menopause (Shoubridge and Wai, 2007). As primary oocytes contain ~6000 mitochondria (Shoubridge and Wai, 2007), expansion up to ~500,000 copies in the mature oocyte will amplify any mutations acquired during oocyte arrest at prophase I, potentially over decades (Arbeithuber *et al.*, 2020). The metabolic quiescence of oocytes can best be understood in light of the need to repress mitochondrial mutation accumulation during the extended period before reproduction (Allen and de Paula, 2013; de Paula *et al.*, 2013).

In this chapter, I have addressed a simple paradox at the heart of mitochondrial inheritance. Like Gibbon's *Decline and Fall of the Roman Empire*, mitochondrial DNA is often portrayed as being in continuous and implacable decline through Muller's ratchet (Rand, 2001); yet like the Empire, which endured for another millennium, mitochondrial DNA has persisted and has been at the heart of eukaryotic cell function for over a billion years (Lane,

2005). Strong evidence for purifying and adaptive selection implies that the female germline facilitates selection for mitochondrial quality, but the mechanisms have remained elusive. I have modelled segregation and selection of mitochondrial DNA at each stage of germline development, and shown that direct selection for mitochondrial function during transfer into the Balbiani body is the most likely explanation of the observed prevalence of mitochondrial mutations and diseases in human populations. More remarkably, this mitochondria-centric model elucidates the complexities of the female germline. It explains why mature oocytes are crammed with mitochondria (Radzvilavicius *et al.*, 2016), whereas sperm mitochondria are typically destroyed, giving rise to two sexes (Radzvilavicius *et al.*, 2017); why germ cells over-proliferate during early germline development; why oogonia organize themselves into germline cysts, forming the Balbiani body; why the majority of germ cells then perish by apoptosis as nurse cells; why primary oocytes enter metabolic quiescence, sometimes for decades; and even why polar bodies channel most of their mitochondria into a single mature oocyte. The need for mitochondrial quality extends to somatic cells, as mitochondria activity is crucial to cellular, tissue and organ functioning in the adult organism (Carelli *et al*., 2015; Diot *et al*., 2016; Pereira *et al*., 2021). Some of the approaches adopted here need to be applied to development and whether specific processes have evolved to maintain mitochondria where their function is more critically related to somatic fitness (Buss, 1987; Radzvilavicius *et al.*, 2016). Most fundamentally, this perspective challenges the claim that complex multicellularity requires passage through a single-celled, haploid stage to constrain the emergence of lower-level, selfish genetic elements (Buss, 1987; Maynard Smith and Szathmáry, 1995). This is true for nuclear genes in oocytes, whose quality is maintained by sexual exchange and recombination (Maynard Smith and Szathmáry, 1995), but is not the case for mitochondria, which are generally

transmitted uniparentally, without sexual exchange or recombination. In animals, the oocyte cytoplasm is not derived from a single cell, but instead requires the selective pooling of mitochondrial DNA from clusters of progenitor cells, which together generate high-quality mitochondria at extreme ploidy in mature gametes.

Chapter 5 – Concluding Remarks and Future Challenges

1. Introduction

In the previous chapters, I have demonstrated the role of mutation accumulation in constraining and shaping the evolution of life, with a focus on the evolution of sexual reproduction in ancestral eukaryotes and selective processes for mitochondrial quality control in the female germline. Since the great majority of mutations are deleterious, all forms of life must prevent their accumulation in order to avoid extinction (Haigh, 1978; Muller, 1964). During eukaryogenesis, as discussed in Chapter 2 and 3, the need for increased purifying selection led ancestral eukaryotes to abandon LGT in favour of meiotic sex (Colnaghi *et al.*, 2020). Similarly, the results presented in Chapter 4 indicate that uniparental inheritance and negligible recombination in mitochondrial genomes prompted the evolution of filtering mechanisms in the female germline to minimise the transmission of deleterious mitochondrial haplotypes (Colnaghi *et al.*, 2021).

The theoretical models presented in Chapter 2 and 3 clarify the selective pressures, arising from genome size expansion and high genetic repeat density, that drove the evolution of meiotic sex. This model implicitly assumes that the acquisition of mitochondria initiated the increase in genome size and repeat density that led to the origin of eukaryotes (Lane, 2007; 2014). Nevertheless, the theoretical arguments presented are generally valid and can be applied to other timelines of eukaryogenesis, such as 'mitochondria-late' models in which the acquisition of the bacterial symbiont occurred during the last stages of eukaryotic evolution (Ettema, 2016; Pittis and Gabaldón, 2016). In addition to eukaryogenesis and germline evolution, the mathematical and computational approaches developed can be used to model several other evolutionary systems, including the evolution

of bacterial genomes, the process of genome streamlining in organelles and endosymbionts, and the evolution of a two-step meiosis. In the following paragraphs, I will outline possible modelling approaches to solve these as yet unanswered evolutionary questions.

2. Eukaryogenesis timeline and the origin of sex

The lack of evolutionary intermediates and the difficulties associated with estimating the exact order for the acquisition of key eukaryotic traits (e.g., mitochondria, the nucleus, a dynamic cytoskeleton, endomembrane system, and linear chromosomes) make it difficult to establish a precise timeline of eukaryogenesis (Dacks *et al.*, 2016; Koumandou *et al.*, 2013). One of the main hypotheses advanced in the literature is that the acquisition of the mitochondrial symbiont favoured the transition from prokaryotes to eukaryotes, as the extra energetic availability supported the expansion of the host's genome (Lane, 2014; 2020). This process took place through frequent gene duplication and the integration of endosymbiotic genes into the host's genome (Tria *et al.*, 2021), but the exact timing of endosymbiosis is debated. According to the 'mitochondria early' view, the acquisition of new genetic material becomes possible only after the acquisition of the bacterial endosymbiont eliminated the bioenergetic constraints that prevent genome size expansion in prokaryotes (Lane, 2007; 2014). An alternative hypothesis ('mitochondria late') is that mitochondria were acquired towards the end of eukaryotic evolution by an organism that already possessed a cytoskeleton, an active endomembrane system, and was fully capable of phagocytosis (Ettema, 2016; Pittis and Gabaldón, 2016).

If the acquisition of mitochondria were the apex of eukaryogenesis, rather than its catalyst, we should expect evolutionary intermediates between prokaryotes and eukaryotes to be found in nature. But they are not. Prokaryotes with typically eukaryotic features

(linear chromosomes, giant size, a dynamic cytoskeleton, a nucleus-like structure, an endomembrane system, endosymbionts) are not uncommon, but none of them possess more than a few of the traits that characterise eukaryotic complexity (Lane, 2014). Even the largest bacteria do not come anywhere near the morphological complexity associated with all eukaryotic life forms (Angert, 2012; Griese *et al*., 2011). A recent phylogenetic analysis of the timing of duplications supports a 'mitochondria early-ish' hypothesis, according to which a series of duplication events promoted the amplification of cell trafficking and structural genes, leading to the evolution of phagocytosis (Vosseberg *et al.*, 2021). This in turn favoured the acquisition of endosymbionts, which supported the process of genome size expansion through duplications and gene transfer to the host genome (Vosseberg *et al.*, 2021). The idea that an initial phase of genome expansion through duplications preceded the acquisition of the bacterial symbiont is hard to justify, given that duplications are so rarely observed in extant prokaryotes; on the contrary, they generally display extremely streamlined genomes and a strong tendency to reduce, rather than increase, their genome size (Mira *et al.*, 2001; Treangen and Rocha, 2011). In addition, this hypothesis does not readily explain why most duplication events in the Last Eukaryotic Common Ancestor (LECA) involve genes of bacterial (endosymbiotic), not archaeal, origin (Tria *et al*., 2021). A different phylogenetic approach, based on the analysis of duplication events, supports the idea that mitochondria were acquired early (Tria *et al.*, 2021). Most duplication events that can be traced back to LECA involve bacterial-derived genes of endosymbiotic origin, rather than archaeal-derived or eukaryote-specific genes (Tria *et al.*, 2021). This suggests the integration of bacterial genes in the archaeal host's genome following the repeated lysis of endosymbionts, in line with a mitochondria-early scenario.

Besides these uncertainties around the time of acquisition of the mitochondrion, the precise timing of the origin of meiosis is also hard to evaluate (Dacks *et al.*, 2016; Koumandou *et al.*, 2013). Did it arise in the first eukaryotes, facilitating the evolution of further complexity? Or did it arise during late eukaryotic evolution because of the evolutionary necessity of purging large genomes from the accumulation of deleterious mutations? The models presented in this thesis do not strictly rely on a single timeline for eukaryogenesis nor on the specific time of acquisition of different eukaryotic traits. They can be applied to explain the evolution of meiosis under different eukaryogenesis scenarios. In a 'mitochondria-early' scenario, the acquisition of endosymbionts ramps up the host's genome size because of both endosymbiotic genome transfer (which promotes genome size expansion and duplication) and the extra energy availability (which removes bioenergetic constraints on genome size). The presence of the endosymbiont reduces the selective pressure for genome streamlining, allowing the host to expand its genome and consequently exposing the first eukaryotes to the threat of mutational meltdown because of the limitations of LGT discussed in Chapter 2 and 3 (Colnaghi *et al.*, 2020; Lane, 2014). A large genome, potentially combined with a higher mutation rate due to ROS (Speijer, 2015), accumulates mutations at a faster rate and curtails LGT's efficiency in promoting purifying selection, demanding an increased recombination rate (Colnaghi *et al.*, 2020). But the early eukaryotes' genomes accumulated an increasingly high density of repeated sequences because of duplications and the spread of transposable elements of endosymbiotic origin. The results presented in Chapter 3 indicate that, under these conditions, LGT cannot prevent eukaryotic-sized genomes from crumbling under the accumulation of deleterious point mutations without severe loss of genetic information through deletions. The only way out is the evolution of a system of genetic exchange that requires chromosome alignment

and homologous recombination across whole chromosomes, minimising the likelihood of ectopic recombination – in other words, meiotic sex.

The predictions of this model are also compatible with the 'mitochondria-late' hypothesis, according to which genome size expansion proceeded through gene duplication events before the acquisition of the mitochondrial symbiont (Ettema, 2016; Pittis and Gabaldón, 2016; Vosseberg *et al.*, 2021). Under these assumptions, the observations discussed in Chapters 2 and 3 are equally valid. In this scenario, an initial increase in genetic repeats through gene duplication is followed by the acquisition of the endosymbiont, which caused a further rise in repeat density. This increasingly high density of repeats led to a strong selective pressure against LGT, while the expansion in genome size called for an increase in purifying selection. As in the previous scenario, under these conditions LGT is not able to prevent the accumulation of mutations in the genome of the first eukaryotes, generating a strong selective pressure towards the evolution of meiosis.

3. Bacterial genome evolution and endosymbiont genome streamlining

The results presented in Chapter 2 and 3 show that in the absence of meiotic sex, genome size expansion cannot take place without a concomitant aggravation of Muller's ratchet. In light of this observation, one can postulate the existence of a critical genome size (depending on mutation rate, recombination rate, strength of selection, repeat density) above which any further expansion in genome size is balanced by gene inactivation and deletion. A further increase in genome size cannot take place unless accompanied by one of the following: a decrease of the genome-wide mutation rate, an increase in the frequency of recombination, or stronger selection (e.g., because of a larger population size). A previous study proposed that prokaryotic genome size is the result of a balance between the

selective advantage of acquiring new genes and the tendency towards deletion (Sela *et al.*, 2016). While providing a compelling explanation, the authors neglect the cost of maintaining a large genome in the face of mutation accumulation, the decreasing benefits of LGT with genome size, and the additional costs associated with the presence of genetic repeats (see Chapter 2 and 3). The trade-off between the advantages of a larger genome (which increases adaptivity) and the reduced scope for purifying selection via LGT is likely to result in constraints on the bacterial genome depending on population size, LGT and mutation rate. While meiotic sex and mitochondria allow eukaryotes to expand their genomes almost indefinitely, prokaryotes rely on a different strategy to ensure adaptability and evolvability: they have streamlined and extremely plastic genomes, with a collective 'pan-genome' from which they can rapidly acquire genetic information via LGT in order to respond effectively to environmental changes (McInerney *et al.*, 2017; Pál *et al.*, 2005; Wiedenbeck and Cohan, 2011). The model presented in Chapters 2 and 3 can be developed to study the constraints on prokaryotic genome size, the evolution of prokaryotic genomes, and the process of genome streamlining in endosymbiotic parasites.

If genome size expansion cannot take place without a concomitant decrease in mutation rate or increase in strength of selection/recombination rate, it is reasonable to assume the opposite to be true as well: that a lower recombination rate and/or strength of selection (e.g., because of smaller population size), or a higher mutation rate, would cause a decrease of the maximum genome size that a population can sustain. As discussed in Chapter 1, many of these conditions (reduced scope for recombination, weaker selection, smaller population size) are typical of the lifestyle of endosymbiotic parasites. The transition to an endosymbiotic lifestyle implies a decrease in the strength of selection over a large majority of the genome, since numerous genes necessary for survival in a normal

environment become superfluous after endosymbiosis. This, in turn, causes the progressive erosion of genes under weak selection, which are eventually lost through deletions (Bobay and Ochman, 2017; Mira *et al.*, 2001).

Endosymbionts that undergo genome streamlining are often thought to gain a selective advantage due to faster reproduction (Andersson and Kurland, 1998; Cavalier-Smith, 1987; Selosse *et al*., 2001; Vellai and Vida, 1999). In extant prokaryotes, however, growth rate does not show any significant correlation with genome size (Bergthorsson and Ochman, 1998; Mira *et al.*, 2001). On the contrary, some endosymbiont bacteria with extremely streamlined genomes, such as leprosy bacillus *Mycobacterium leprae*, are among the most slowly reproducing prokaryotes, with a doubling time of over two weeks (Cole *et al*., 2001). It has therefore been suggested that other factors, such as ribosomal translational efficiency and tRNA availability, are more important than genome size in controlling bacterial growth rate (Emilsson and Kurland, 1990; Kurland and Ehrenberg, 1987; Mikkola and Kurland, 1991). Nevertheless, sharp genome size reduction can be consistently observed in numerous microorganisms subjected to constant and stable environments, suggesting that predictable environmental conditions favour the process of gene loss (see, for example, Tenaillon *et al.* 2016 and Lee and Marx, 2012).

An alternative view is that deletion bias is responsible for genome size reduction in endosymbionts (Mira *et al.*, 2001). This hypothesis is compatible with the idea that Muller's ratchet is responsible for the progressive degradation and inactivation of genes under weak selection, which are eventually eliminated through deletion (Bobay and Ochman, 2017; Silva *et al*., 2001). A signature of Muller's ratchet is indeed found in endosymbiotic bacteria, which display a higher ratio of non-synonymous to synonymous substitutions (dN/dS) relative to closely related free-living species (Lambert and Moran, 1998; Moran, 1996). This

is indicative of the fixation of mutations due to ratchet mechanisms (Charlesworth and Charlesworth, 1997; Higgs and Woodcock, 1995). The theoretical models of Chapters 2 and 3 can be developed further to study endosymbiotic parasites and improve our understanding of how Muller's ratchet promotes genome streamlining. One qualitative prediction of the model is the presence of a correlation between severe bottlenecks, high mutation rates and extreme genome size reduction in endosymbionts. By providing quantitative predictions of optimal genome size, given effective population size, duplication time and mutation rate, it would be possible to test this hypothesis against experimental data, allowing us to determine more accurately the forces that govern the evolution endosymbiotic genomes.

4. Mutation accumulation and organelle gene transfer to the nucleus

A specific application of this model is the process of transfer of mitochondrial genomes to the nucleus. The bacterial symbiont that gave rise to mitochondria is thought to have possessed, like extant alpha-proteobacteria, around 1,500 protein-coding genes (Gray *et al*., 2001; Martin and Müller, 2007; Roger *et al*., 2017). After endosymbiosis, a large number of these genes were integrated into the host's genome, while the mitochondrial genome underwent a progressive shrinkage; a process that continued throughout eukaryotic evolution. As a result, different eukaryotic taxa display a range of mitochondrial genome size, structure, and number of genes (Burger *et al.*, 2003). The number of mitochondrial genes spans displays a 20-fold variation, ranging from nearly 100 genes in jakobid flagellates to only 5 in *Plasmodium* spp. (Burger *et al.*, 2003; Lang et al., 1997).

As nuclear genomes undergo systematic recombination through meiosis and are (at least in metazoans) subjected to lower mutation rates, they accumulate mutations at a

lower rate than mtDNA (Allio *et al.*, 2017; Lynch *et al.*, 2006). Allen and Raven (1996) suggested that the driving force behind organellar gene transfer to the nucleus is the selective advantage provided by recombination and lower mutation rates. The genes that are retained by the organelles are the ones whose expression needs to be controlled locally (Allen, 1993; 2015). According to the 'CoRR' (Colocation for Redox Regulation of gene expression) hypothesis, the need for tight regulation of gene products near the organellar membrane determines the retention of genes in mitochondria and chloroplasts (Allen, 1993; 2015). Further modelling efforts are necessary to quantitatively understand what gives rise to the wide distribution of mitochondrial gene number across eukaryotes.

A possible approach would involve modelling Muller's ratchet in the mitochondrial genomes of individuals whose fitness depends on mitochondrial mutation load: as the transfer of mitochondrial genes to the nucleus would reduce the mutational target offered by mitochondrial genomes, the severity of the ratchet would be inversely correlated with the number of transferred genes. Organisms with smaller mitochondrial genomes would accumulate mitochondrial mutations at a lower rate, making the transfer of genes to the nucleus advantageous. This selective advantage would increase with the severity of mitochondrial mutation rate, leading to the prediction of an inverse correlation between the number of mitochondrial genes and mitochondrial mutation rates. Indeed metazoa, who are subjected to higher mitochondrial mutation rates than most other eukaryotes, possess particularly small and compact mitochondrial genomes (Allio *et al.*, 2017; Burger *et al.*, 2003; Lavrov and Pett, 2016) and very large numbers of mtDNA copies in mature oocytes (Radzvilavicius *et al.*, 2016). In addition to reducing the number of mitochondrial genomes, metazoa also rely on purifying selection (Colnaghi *et al.*, 2021), as discussed in Chapter 4, and extreme mtDNA ploidy in mature oocytes (Radzvilavicius *et al.*, 2016), to

limit the transmission of deleterious mitochondrial haplotypes. The opposite trend can be observed in higher plants, which typically possess extremely large mitochondrial genomes (Burger *et al.*, 2003; Fauron *et al.*, 2004; Ward et al., 1981).

Unicellular eukaryotes with low mtDNA ploidy generally possess small mitochondrial genomes, suggesting that reduced mtDNA ploidy might also favour genome size reduction. For example, the phytoflagellate *Chlamydomonas reinhardii*, a red algae, possesses a mitochondrial genome of only 15.8 kbp (Michaelis *et al*., 1990), and a proportionally small number of mitochondria, oscillating between 7 and 50 copies per cell (Blank *et al*., 1980). Reducing mtDNA ploidy increases segregational variance (see Chapter 4), favouring selection at the level of the cell, but at the same time it increases the severity of Muller's ratchet because of smaller population size. As discussed in Chapter 4, the presence of negative epistasis among mitochondrial mutations can make selective processes at the cell level less effective, favouring the build-up of deleterious mutations. The interplay between mitochondrial genome size and mtDNA ploidy is another aspect that could be elucidated by future theoretical studies.

Together, these observations point to a correlation between mitochondrial genome size and the severity of Muller's ratchet caused by either high mutation rates or low effective population size. The theoretical models developed in this thesis could be expanded to study this process in a more rigorous, quantitative way, and tested against experimental data. It seems important to point out, however, a fundamental assumption of the models outlined above: that deleterious mutations are detrimental for both the mitochondrion and its host. This approximation overlooks the complications arising from intragenomic conflict between the mitochondrion and the host (i.e., the presence of mutations that increase mitochondrial

fitness at the expenses of cell fitness), which could also generate a selective pressure for the streamlining of mitochondrial genomes (Havird *et al.*, 2019; Dobler *et al.*, 2014).

5. Why 2-step meiosis?

In addition to investigating the transfer of organellar genomes to the nucleus, modelling Muller's ratchet in mitochondrial genomes might provide a novel approach to understanding the evolution of meiosis. In the majority of eukaryotes, meiosis is a 2-step process consisting of duplication followed by two cell divisions, generating haploid gametes (Lenormand *et al*., 2016). The pervasive nature of 2-step meiosis and the conserved kit of meiotic genes across eukaryotes suggests that it arose only once, during early eukaryotic evolution (Ramesh *et al.*, 2005; Schurko and Logsdon, 2008). If the purpose of meiosis is to generate haploid gametes, what need is there of duplication followed by an additional cell division (Hurst, 1993; Lenormand *et al.*, 2016; Wilkins and Holliday, 2009)? In other words, why a 2-step meiosis with duplication and two cell divisions instead of a single cell division? Further modelling work focussed on the unique conditions of early eukaryogenesis and the need to prevent mutation accumulation might provide a novel and promising approach to tackle this evolutionary puzzle. In Chapters 1 and 4, I have discussed the need for increased purifying selection in mitochondrial genomes arising from uniparental inheritance and limited scope for recombination. Numerous mechanisms have evolved in metazoa to curtail the transmission of deleterious mitochondrial haplotypes, including the segregation of the germline and the presence of organelle-level selection (Colnaghi *et al.*, 2021; Kim *et al.*, 2007; Lieber *et al*., 2019; Radzvilavicius *et al.*, 2016) . Analogously, plants limit the amount of cell division in both germ cells and stem cells, limiting the build-up of mitochondrial mutations (Groot and Laux, 2016; Lanfear, 2018). The process of selection against

mitochondrial mutations is facilitated by segregation, which increases variance and promotes cell-level selection (Colnaghi *et al.*, 2021; Radzvilavicius *et al.*, 2016).

The need to prevent the build-up of deleterious mutations in endosymbiotic genomes is not exclusive of extant eukaryotes: it is an evolutionary imperative the first eukaryotes must have dealt with as well. Considering that early mitochondria possessed much larger genomes, it is also reasonable to assume that they were present in smaller numbers, as observed in extant unicellular eukaryotes. As discussed in the previous chapters, these two factors – large genome and small population size, with limited scope for recombination – are predicted to precipitate the severity of Muller's ratchet, generating strong pressure for purifying selection (Colnaghi *et al.*, 2020). To which extent could this be achieved through a combination of segregational variance and cell-level selection? One of the potential evolutionary advantages of 2-step meiosis is that it generates greater variance in mitochondrial mutation load among daughter cells than a single cell division. If Muller's ratchet were taking place in early mitochondrial genomes, the greater variance generated by two rounds of mitochondrial segregation would confer selective advantage by favouring the reduction of mitochondrial mutation load. Could the need for increased segregational variance have driven the evolution of 2-step meiosis in LECA? Further modelling work is necessary to answer this question. This could be achieved by integrating the model presented in Chapters 2 and 3 with a 2-level selection process (organelle and cell level) and analysing under which conditions can 2-step meiosis prevent mutational meltdown, while 1-step meiosis cannot. If the parameter range under which this takes place is sufficiently wide, this might shed some light on the ancestral origin of 2-step meiosis.

6. Conclusion

All composite things are impermanent… The world is afflicted by death and decay.

But the wise do not grieve, having realised the nature of the world.

– Siddhārtha Gautama

The single most important and revolutionary concept introduced by Charles Darwin in his '*On the Origin of Species*' is the absence of fixed, permanent categories in nature (Darwin, 1859). Species change, evolve, transmute into other species; life is not something fixed, part of a pre-determined natural hierarchy. A great deal of effort in evolutionary biology is devoted to understanding how change takes place, how beneficial traits arise and spread, how life can explore the multitude of extremely diverse developmental, morphological, and bioenergetic landscapes observed in nature. But adaptive change, innovation and novelty could not exist without their opposite: the inevitable tendency towards decay and dissolution. The yang of evolutionary creativity is balanced by the yin of deleterious mutations, which cause an inevitable tendency towards the loss of genetic information. Mutational decay is the price to be paid for adaptive change, and every form of life is under the evolutionary imperative to contrast this process of decay or perish. The tendency towards innovation and adaptation is only a small aspect of life, which is essentially a conservative process against decay and entropy. This is true for cells and organisms, which must constantly adapt to changing external conditions in order to maintain their homeostatic balance, and for species, which must adapt to constantly changing fitness landscapes without losing essential genetic information.

Obviously, deleterious mutations cannot be avoided – they are the price to be paid for evolutionary creativity and adaptive change. But their accumulation, nevertheless, must be prevented in order to ensure the persistence of a species. Throughout this thesis, I have

given numerous examples of how the evolutionary imperative of purifying selection has shaped the evolution of life. In particular, I have analysed its role in the evolutionary transition from prokaryotes to eukaryotes and the origin of meiotic sex, and in the evolution of germline processes that limit the transmission of deleterious mitochondrial haplotypes. I have also highlighted possible avenues of future research in various areas, including the evolution of prokaryotic genomes, the transfer of organelle genes to the nucleus, and the evolution of a two-step meiosis. The list is by no means exhaustive: other examples of evolutionary dynamics affected by the need for purifying selection can be found throughout the history of life, demonstrating that life's tendency towards change and transformation is as inevitable as mutation, death, and decay. In fact, the evolutionary imperative of transcending death and decay is one of the greatest sources of adaptive innovation, the mud out of which the lotus of evolutionary creativity can flower. The origin of sex, the presence of two sexes, the germline, and perhaps in the end even complex life itself – everything colourful and beautiful in the world – can be seen as arising from the need to maintain quality in a world that inevitably tends towards decay and impermanence.

During the evolution of the first living organisms, the need to ensure the transmission of beneficial variants and prevent their decay must have arisen together with genetic information itself. Eigen's paradox, the apparent impossibility for faithful replication to arise in a system of self-replicating nucleotides in the presence of copying errors, presents numerous analogies with the informational decay predicted by Muller's ratchet. This paradox arises from the Manfred Eigen's estimate that, given the error rate of self-replicating nucleotide polymers in absence of error-repair mechanisms, the maximum length of a successfully replicating polymer in a hypothetic RNA-world would be approximately 100 base pairs – at least one order of magnitude smaller than the size of a

molecule capable of encoding an error-correction enzyme (Eigen, 1971). As life evolves towards greater complexity, the need arises for the evolution of molecular machinery that can ensure higher replication fidelity, minimise copying error, and promote the elimination of deleterious mutations across multiple levels of selection (indeed, the introduction of multi-level selection offers a possible way out of Eigen's paradox, as in Eigen's hypercycle model or Szathmáry's stochastic replicator; see, for example, Maynard Smith and Szathmáry, 1995).

Muller's ratchet is also thought to play a crucial role in shaping the evolution of viruses and bacteria, determining the structure and size of their genomes. At the other end of the evolutionary spectrum, the build-up of mutations in somatic cells has a profound impact on human health, being potentially involved in ageing, cancer, and mitochondrial disease. Obviously, it was impossible to explore all those different avenues in the present work: some of them have just been briefly hinted at in these last chapters, others have been passed on in almost complete silence.

The theoretical ideas discussed throughout this dissertation can potentially offer novel approaches of looking at these evolutionary questions and advance our understanding of how deleterious mutations impact and shape the evolution of life. It is my sincere hope that any advances in this field will, ultimately, contribute to human welfare in general and to the flourishing of our society in harmony with the natural environment. In the same way mutational decay led to the flourishing of living complexity, may the disruptive processes in our society also give rise to a better, more meaningful future.

Bibliography

Albamonte, M.S., Willis, M.A., Albamonte, M.I., Jensen, F., Espinosa, M.B., and Vitullo, A.D. (2008). The developing human ovary : immunohistochemical analysis of germ-cell-specific VASA protein, BCL-2 / BAX expression balance and apoptosis. Hum Reprod *23*, 1895-1901.

Allen, J.F. (1993). Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. J Theor Biol *165*, 609-631.

Allen, J.F. (2015). Why chloroplasts and mitochondria retain their own genomes and genetic systems: colocation for redox regulation of gene expression. Proc Natl Acad Sci U S A *112*, 10231-10238.

Allen, J.F., and de Paula, W. (2013). Mitochondrial genome function and maternal inheritance. Biochem Soc Trans *41*, 1298-1304.

Allen, J.F., and Raven, J. (1996). Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. J Mol Evol *42*, 482-492.

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

Ambur, O.H., Engelstädter, J., Johnsen, P.J., Miller, E.L., and Rozen, D.E. (2016). Steady at the wheel: conservative sex and the benefits of bacterial transformation. Phil Trans R Soc B *371*, 20150528.

Andersson, D.I., and Hughes, D. (1996). Muller's ratchet decreases fitness of a DNA-based microbe. Proc Natl Acad Sci U S A *93*, 906-907.

Andersson, S.G., and Kurland, C.G. (1998). Reductive evolution of resident genomes. Trends Microbiol *6*, 263-268.

Angert, E.R. (2012). DNA replication and genomic architecture of very large bacteria. Annu Rev Microbiol *66*, 197-212.

Arbeithuber, B., Hester, J., Cremona, M.A., Stoler, N., Zaidi, A., Higgins, B., Anthony, K., Chiaromonte, F., Diaz, F.J., and Makova, K.D. (2020). Age-related accumulation of de novo mitochondrial mutations in mammalian oocytes and somatic tissues. PLOS Biol *18*, e3000745. 10.1371/journal.pbio.3000745.

Ballard, J.W., and Kreitman, M. (1994). Unraveling selection in the mitochondrial genome of Drosophila. Genetics *138*, 757-772.

Baltrus, D.A. (2013). Exploring the costs of horizontal gene transfer. Trends Ecol Evol *28*, 489-495.

Baltrus, D.A., Guillemin, K., and Phillips, P.C. (2008). Natural transformation increases the rate of adaptation in the human pathogen Helicobacter pylori. Evolution: International Journal of Organic Evolution *62*, 39-49.

Barton, N.H., and Otto, S.P. (2005). Evolution of recombination due to random drift. Genetics *169*, 2353-2370.

Bell, G. (1982). The masterpiece of nature: the evolution and genetics of sexuality (Routledge).

Bell, G. (1988). Recombination and the immortality of the germ line. J Evol Biol *1*, 67-82.

Bennett, G.M., and Moran, N.A. (2013). Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. Genome Biol Evol *5*, 1675-1688.

Bergstrom, C., and Pritchard, J. (1998). Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. Genetics *149*, 2135-2146.

Bergthorsson, U., and Ochman, H. (1998). Distribution of chromosome length variation in natural isolates of Escherichia coli. Mol Biol Evol *15*, 6-16.

Bernstein, H., and Bernstein, C. (2013). Evolutionary origin and adaptive function of meiosis. Meiosis *1*, 56557.

Bilinski, S.M., Kloc, M., and Tworzydlo, W. (2017). Selection of mitochondria in female germline cells: is Balbiani body implicated in this process? J Assist Reprod Genet *34*, 1405-1412.

Blank, R., Hauptmann, E., and Arnold, C.G. (1980). Variability of mitochondrial population in Chlamydomonas reinhardii. Planta *150*, 236-241.

Bobay, L.M., and Ochman, H. (2017). The evolution of bacterial genome architecture. Front Genet *8*, 72.

Bobay, L.M., and Ochman, H. (2018). Factors driving effective population size and pan-genome evolution in bacteria. BMC Evol Bio *18*, 1-12.

Burger, G., Gray, M.W., and Lang, B.F. (2003). Mitochondrial genomes: anything goes. Trends Genet *19*, 709-716.

Burian, A., De Reuille, P.B., and Kuhlemeier, C. (2016). Patterns of stem cell divisions contribute to plant longevity. Curr Biol *26*, 1385-1394.

Burr, S.P., Pezet, M., and Chinnery, P.F. (2018). Mitochondrial DNA heteroplasmy and purifying selection in the mammalian female germ line. Dev Growth Differ *60*, 21-32.
Buss, L. (1987). The Evolution of Individuality (Princeton University Press).
Butcher, D. (1995). Muller's ratchet, epistasis and mutation effects. Genetics *141*, 431-437.
Butlin, R. (2002). The costs and benefits of sex: new insights from old asexual lineages. Nat Rev Genet *3*, 311-317.
Cao, L., Shitara, H., Sugimoto, M., Hayashi, J., Abe, K., and 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 Genet *5*, e1000756.
Carelli, V., Maresca, A., Caporali, L., Trifunov, S., Zanna, C., and Rugolo, M. (2015). Mitochondria: biogenesis and mitophagy balance in segregation and clonal expansion of mitochondrial DNA mutations. Int J Biochem Cell Biol *63*, 21-24.
Carr, M., Leadbeater, B.S., and Baldauf, S.L. (2010). Conserved meiotic genes point to sex in the choanoflagellates. J Eukaryot Microbiol *57*, 56-62.
Cavalier-Smith, T. (1991). Intron phylogeny: a new hypothesis. Trends Genet *7*, 145-148.
Cavalier-Smith, T. (1987). The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies. Ann N Y Acad Sci *503*, 55-71.
Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. Nature *348*, 454-455.
Charlesworth, and Charlesworth (1997). Rapid fixation of deleterious alleles can be caused by Muller's ratchet. Genet Res *70*, 63-73.
Charlesworth, and Charlesworth (2000). The degeneration of Y chromosomes. Phil Trans R Soc B *355*, 1563-1572.
Charlesworth, Morgan, and Charlesworth (1993). Mutation accumulation in finite populations. J Hered *84*, 321-325.
Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. Proc Natl Acad Sci U S A *75*, 5618-5622.
Chen, Z., Wang, Z.H., Zhang, G., Bleck, C.K.E., Chung, D.J., Madison, G.P., Lindberg, E., Combs, C., Balaban, R.S., and Xu, H. (2020). Mitochondrial DNA segregation and replication restrict the transmission of detrimental mutation. J Cell Biol *219*. 10.1083/jcb.201905160.
Chu, H.P., Liao, Y., Novak, J.S., Hu, Z., Merkin, J.J., Shymkiv, Y., Braeckman, B.P., Dorovkov, M.V., Nguyen, A., Clifford, P.M., and al., e. (2014). Germline quality control : eEF2K stands guard to eliminate defective oocytes. Dev Cell *28*, 561-572.

Claverys, J.P., Martin, B., and Polard, P. (2009). The genetic transformation machinery: composition, localization, and mechanism. FEMS Microbiol Rev *33*, 643-656.
Cole, S., Eiglmeier, K., Parkhill, J., James, K., Thomson, N., Wheeler, P., Honore, N., Garnier, T., Churcher, C., and Harris, D. (2001). Massive gene decay in the leprosy bacillus. Nature *409*, 1007-1011.
Colnaghi, M., Lane, N., and Pomiankowski, A. (2020). Genome expansion in early eukaryotes drove the transition from lateral gene transfer to meiotic sex. Elife *9*, e58873.
Colnaghi, M., Pomiankowski, A., and Lane, N. (2021). The need for high-quality oocyte mitochondria at extreme ploidy dictates mammalian germline development. Elife *10*, e69344.
Cox, R.T., and Spradling, A.C. (2003). A Balbiani body and the fusome mediate mitochondrial inheritance during *Drosophila* oogenesis. Development *130*, 1579. 10.1242/dev.00365.
Cree, L.M., Samuels, D.C., de Sousa Lopes, S.C., Rajasimha, H.K., Wonnapinij, P., Mann, J.R., Dahl, H., H., and Chinnery, P.F. (2008). A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. Nat Genet *40*, 249-254.
Croucher, N.J., Harris, S.R., Barquist, L., Parkhill, J., and Bentley, S.D. (2012). A high-resolution view of genome-wide pneumococcal transformation. PLoS Pathog *8*, e1002745.
Croucher, N.J., Mostowy, R., Wymant, C., Turner, P., Bentley, S.D., and Fraser, C. (2016). Horizontal DNA transfer mechanisms of bacteria as weapons of intragenomic conflict. PLoS Biol *14*, e1002394.
Crow, J.F., and Kimura, M. (1965). Evolution in sexual and asexual populations. Am Nat *99*, 439-450.
Crow, J.F., and Kimura, M. (1969). Evolution in sexual and asexual populations: a reply. Am Nat *103*, 89-91.
Cummins, J.M. (2004). The role of mitochondria in the establishment of oocyte functional competence. Eur J Obstet Gynecol Reprod Biol *115*, S23-29.
Dacks, J.B., Field, M.C., Buick, R., Eme, L., Gribaldo, S., Roger, A.J., Brochier-Armanet, C., and Devos, D.P. (2016). The changing view of eukaryogenesis–fossils, cells, lineages and how they all come together. J Cell Sci *129*, 3695-3703.
Dalton, C.M., and Carroll, J. (2013). Biased inheritance of mitochondria during asymmetric cell division in the mouse oocyte. J Cell Sci *126*, 2955-2964. 10.1242/jcs.128744.
Darwin, C.R. (1859). On the Origin of Species.

de la Iglesia, F., and Elena, S.F. (2007). Fitness declines in Tobacco etch virus upon serial bottleneck transfers. J Virol *81*, 4941-4947.

de Paula, W.B.M., Lucas, C.H., Agip, A.N.A., Vizcay-Barrena, G., and Allen, J.F. (2013). Energy, ageing, fidelity and sex: oocyte mitochondrial DNA as a protected genetic template. Philos Trans R Soc B *368*, 20120263. 10.1098/rstb.2012.0263.

de Visser, J.A.G., and Elena, S.F. (2007). The evolution of sex: empirical insights into the roles of epistasis and drift. Nat Rev Genet *8*, 139-149.

de Vries, J., and Wackernagel, W. (2002). Integration of foreign DNA during natural transformation of Acinetobacter sp. by homology-facilitated illegitimate recombination. Proc Natl Acad Sci U S A *99*, 2094-2099.

Denver, D.R., Morris, K., Lynch, M., Vassilieva, L.L., and Thomas, W.K. (2000). High Direct Estimate of the Mutation Rate in the Mitochondrial Genome of *Caenorhabditis elegans*. Science *289*, 2342. 10.1126/science.289.5488.2342.

Diot, A., Morten, K., and Poulton, J. (2016). Mitophagy plays a central role in mitochondrial ageing. Mammal Genome *27*, 381-395.

Dobler, R., Rogell, B., Budar, F., and Dowling, D. K. (2014). A meta-analysis of the strength and nature of cytoplasmic genetic effects. J Evo Biol, 27(10), 2021-2034.

Duarte, E., Clarke, D., Moya, A., Domingo, E., and Holland, J. (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. Proc Natl Acad Sci U S A *89*, 6015-6019.

Dumollard, R., Duchen, M., and Carroll, J. (2007). The role of mitochondrial function in the oocyte and embryo. Curr Top Dev Biol *77*, 21-49.

Eigen, M. Selforganization of matter and the evolution of biological macromolecules. (1971). Naturwissenschaften *58*, 465–523.

Elliott, H.R., Samuels, D.C., Eden, J.A., Relton, C.L., and Chinnery, P.F. (2008). Pathogenic mitochondrial DNA mutations are common in the general population. Am J Hum Genet *83*, 254-260.

Elliott, T.A., and Gregory, T.R. (2015). What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philos Trans R Soc B *370*, 20140331.

Emilsson, V., and Kurland, C.G. (1990). Growth rate dependence of transfer RNA abundance in Escherichia coli. EMBO J *9*, 4359-4366.

Engelmoer, D.J., Donaldson, I., and Rozen, D.E. (2013). Conservative sex and the benefits of transformation in Streptococcus pneumoniae. PLoS Pathog *9*, e1003758.
Erickson, H.P. (2007). Evolution of the cytoskeleton. Bioessays *29*, 668-677.
Eshel, I., and Feldman, M.W. (1970). On the evolutionary effect of recombination. Theor Popul Biol *1*, 88-100.
Etheridge, A., Pfaffelhuber, P., and Wakolbinger, A. (2009). How often does the ratchet click? Facts, heuristics, asymptotics. Trends Stochastic Anal *353*, 365-390.
Ettema, T.J. (2016). Mitochondria in the second act. Nature *531*, 39-40.
Extavour, C.G., and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. Development *130*, 5869-5884.
Eyre-Walker, A., and Keightley, P.D. (2007). The distribution of fitness effects of new mutations. Nature Rev Genet *8*, 610-618.
Eyres, I., Boschetti, C., Crisp, A., Smith, T.P., Fontaneto, D., Tunnacliffe, A., and Barraclough, T.G. (2015). Horizontal gene transfer in bdelloid rotifers is ancient, ongoing and more frequent in species from desiccating habitats. BMC Biol *13*, 1-17.
Fan, W., Waymire, K.G., Narula, N., Li, P., Rocher, C., Coskun, P.E., Vannan, M.A., Narula, J., Macgregor, G.R., and Wallace, D.C. (2008). A mouse model of mitochondrial disease reveals germline selection against severe mtDNA mutations. Science *319*, 958-963.
Fauron, C., Allen, J., Clifton, S., and Newton, K. (2004). Plant mitochondrial genomes. In Molecular biology and biotechnology of plant organelles, (Springer), pp. 151-177.
Felsenstein, J. (1974). The evolutionary advantage of recombination. Genetics *78*, 737-756.
Felsenstein, J., and Yokoyama, S. (1976). The evolutionary advantage of recombination. II. Individual selection for recombination. Genetics *83*, 845-859.
Fischer, M.G., Allen, M.J., Wilson, W.H., and Suttle, C.A. (2010). Giant virus with a remarkable complement of genes infects marine zooplankton. Proc Natl Acad Sci U S A *107*, 19508-19513.
Floros, V.I., Pyle, A., Dietmann, S., Wei, W., Tang, W.C.W., Irie, N., Payne, B., Capalbo, A., Noli, L., Coxhead, J., et al. (2018). Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos. Nat Cell Biol *20*, 144-151. 10.1038/s41556-017-0017-8.
Fonseca, D.R., Halim, M.F.A., Holten, M.P., and Costa, K.C. (2020). Type IV-like pili facilitate transformation in naturally competent archaea. J Bacteriol *202*, e00355-00320.

Fonseca, R.R., Johnson, W.E., Brien, S.J.O., Ramos, M.J., and Antunes, A. (2008). The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics *9*, 119.
Frye, S.A., Nilsen, M., Tønjum, T., and Ambur, O.H. (2013). Dialects of the DNA uptake sequence in Neisseriaceae. PLoS Genet *9*, e1003458.
Fuchsman, C.A., Collins, R.E., Rocap, G., and Brazelton, W.J. (2017). Effect of the environment on horizontal gene transfer between bacteria and archaea. PeerJ *5*, e3865.
Funchain, P., Yeung, A., Lee Stewart, J., Lin, R., Slupska, M.M., and Miller, J.H. (2000). The consequences of growth of a mutator strain of Escherichia coli as measured by loss of function among multiple gene targets and loss of fitness. Genetics *154*, 959-970.
Galimov, E.R., Chernyak, B.V., Sidorenko, A.S., Tereshkova, A.V., and Chumakov, P.M. (2014). Prooxidant properties of p66shc are mediated by mitochondria in human cells. PloS 1 *9*, e86521-e86521. 10.1371/journal.pone.0086521.
Gandon, S., and Otto, S.P. (2007). The evolution of sex and recombination in response to abiotic or coevolutionary fluctuations in epistasis. Genetics *175*, 1835-1853.
Gerstein, A.C., and Otto, S.P. (2009). Ploidy and the causes of genomic evolution. J Hered *100*, 571-581.
Gessler, D.D.G. (1995). The constraints of finite size in asexual populations and the rate of the ratchet. Genet Res *66*, 241-253.
Gladyshev, Meselson, M., and Arkhipova, I. (2008). Massive horizontal gene transfer in bdelloid rotifers. Science *320*, 1210-1213.
Gladyshev, E., and Meselson, M. (2008). Extreme resistance of bdelloid rotifers to ionizing radiation. Proc Natl Acad Sci U S A *105*, 5139-5144.
Gogarten, J.P., and Townsend, J.P. (2005). Horizontal gene transfer, genome innovation and evolution. Nature Rev Microbiol *3*, 679-687.
Goodenough, U., and Heitman, J. (2014). Origins of eukaryotic sexual reproduction. Cold Spring Harbor Perspect Biol *6*, a016154.
Gordo, I., and Charlesworth, B. (2000a). The degeneration of asexual haploid populations and the speed of Muller's ratchet. Genetics *154*, 1379-1387.
Gordo, I., and Charlesworth, B. (2000b). On the speed of Muller's ratchet. Genetics *156*, 2137-2140.
Gordo, I., Navarro, A., and Charlesworth, B. (2002). Muller's ratchet and the pattern of variation at a neutral locus. Genetics *161*, 835-848.

Gray, M.W., Burger, G., and Lang, B.F. (2001). The origin and early evolution of mitochondria. Genome Biol *2*, 1-5.

Griese, M., Lange, C., and Soppa, J. (2011). Ploidy in cyanobacteria. FEMS Microbiol Lett *323*, 124-131.

Groot, E.P., and Laux, T. (2016). Ageing: how do long-lived plants escape mutational meltdown? Curr Biol *26*, R530-R532.

Guo, Y., Li, C.-I., Sheng, Q., Winther, J.F., Cai, Q., Boice, J.D., and Shyr, Y. (2013). Very low-level heteroplasmy mtDNA variations are inherited in humans. J Genet Genomics *40*, 607-615. 10.1016/j.jgg.2013.10.003.

Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B., and Keightley, P.D. (2008). Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. PLoS Biol *6*, e204. 10.1371/journal.pbio.0060204.

Hadjivasiliou, Z., Lane, N., Seymour, R.M., and Pomiankowski, A. (2013). Dynamics of mitochondrial inheritance in the evolution of binary mating types and two sexes. Proc R Soc B Biol Sci *280*, 20131920.

Haig, D. (2016). Intracellular evolution of mitochondrial DNA (mtDNA) and the tragedy of the cytoplasmic commons. BioEssays *38*, 549-555.

Haigh, J. (1978). The accumulation of deleterious genes in a population—Muller's ratchet. Theor Popul Biol *14*, 251-267.

Hamilton, W.D. (1980). Sex versus non-sex versus parasite. Oikos, 282-290.

Hao, W., and Golding, G.B. (2006). The fate of laterally transferred genes: life in the fast lane to adaptation or death. Genome Res *16*, 636-643.

Havird, J. C., Forsythe, E. S., Williams, A. M., Werren, J. H., Dowling, D. K., and Sloan, D. B. (2019). Selfish mitonuclear conflict. Curr Biol, 29(11), R496-R511.

Heasman, J., Quarmby, J., and Wylie, C.C. (1984). The mitochondrial cloud of Xenopus oocytes: The source of germinal granule material. Dev Biol *105*, 458-469.

Heller, R., and Maynard Smith, J. (1978). Does Muller's ratchet work with selfing? Genet Res *32*, 289-293.

Hertig, A.T. (1968). The primary human oocyte: some observations on the fine structure of Balbiani's vitelline body and the origin of the annulate lamellae. Am J Anat *122*, 107-137.

Higgs, P.G., and Woodcock, G. (1995). The accumulation of mutations in asexual populations and the structure of genealogical trees in the presence of selection. Journal Math Biol *33*, 677-702.

Hill, J.H., Chen, Z., and Xu, H. (2014). Selective propagation of functional mitochondrial DNA during oogenesis restricts the transmission of a deleterious mitochondrial variant. Nat Genet *46*, 389-392.

Hill, W.G., and Robertson, A. (1966). The effect of linkage on limits to artificial selection. Genet Res *8*, 269-294.

Hiller, N.L., Ahmed, A., Powell, E., Martin, D.P., Eutsey, R., Earl, J., Janto, B., Boissy, R.J., Hogg, J., and Barbadora, K. (2010). Generation of genic diversity among Streptococcus pneumoniae strains via horizontal gene transfer during a chronic polyclonal pediatric infection. PLoS Pathog *6*, e1001108.

Hiltunen, M., Grudzinska-Sterno, M., Wallerman, O., Ryberg, M., and Johannesson, H. (2019). Maintenance of high genome integrity over vegetative growth in the fairy-ring mushroom Marasmius oreades. Curr Biol *29*, 2758-2765. e2756.

Hirano, T. (2005). SMC proteins and chromosome mechanics: from bacteria to humans. Phil Trans R Soc B *360*, 507-514.

Hodač, L., Klatt, S., Hojsgaard, D., Sharbel, T.F., and Hörandl, E. (2019). A little bit of sex prevents mutation accumulation even in apomictic polyploid plants. BMC Evol Biol *19*, 1-11.

Hofstatter, P.G., Brown, M.W., and Lahr, D.J. (2018). Comparative genomics supports sex and meiosis in diverse Amoebozoa. Genome Biol Evol *10*, 3118-3128.

Hojsgaard, D., and Hörandl, E. (2015). A little bit of sex matters for genome evolution in asexual plants. Front Plant Sci *6*, 82.

Howe, D.K., and Denver, D.R. (2008). Muller's Ratchet and compensatory mutation in Caenorhabditis briggsae mitochondrial genome evolution. BMC Evol Bio *8*, 1-13.

Hülter, N., and Wackernagel, W. (2008). Double illegitimate recombination events integrate DNA segments through two different mechanisms during natural transformation of Acinetobacter baylyi. Mol Microbiol *67*, 984-995.

Hurst, L.D. (1993). Drunken walk of the diploid. Nature *365*, 206-207.

Iles, M.M., Walters, K., and Cannings, C. (2003). Recombination can evolve in large finite populations given selection on sufficient loci. Genetics *165*, 2249-2258.

Jain, K. (2008). Loss of least-loaded class in asexual populations due to drift and epistasis. Genetics *179*, 2125-2134.

Jain, R., Rivera, M.C., Moore, J.E., and Lake, J.A. (2003). Horizontal gene transfer accelerates genome innovation and evolution. Mol Biol Evol *20*, 1598-1602.

James, J.E., Piganeau, G., and Eyre-Walker, A. (2016). The rate of adaptive evolution in animal mitochondria. Mol Biol Evol *25*, 67-78.

Jaramillo, N., Domingo, E., Muñoz-Egea, M.C., Tabares, E., and Gadea, I. (2013). Evidence of Muller's ratchet in herpes simplex virus type 1. J Gen Virol *94*, 366-375.

Johnston, C., Martin, B., Fichant, G., Polard, P., and Claverys, J.-P. (2014). Bacterial transformation: distribution, shared mechanisms and divergent control. Nature Rev Microbiol *12*, 181-196.

Johnston, I.G., Burgstaller, J.P., Havlicek, V., Kolbe, T., Rulicke, T., Brem, G., Poulton, J., and Jones, N.S. (2015). Stochastic modelling, bayesian inference, and new in vivo measurements elucidate the debated mtDNA bottleneck mechanism. eLife *4*, e07464.

Jokela, J., Dybdahl, M.F., and Lively, C.M. (2009). The maintenance of sex, clonal dynamics, and host-parasite coevolution in a mixed population of sexual and asexual snails. Am Nat *174*, S43-S53.

Kaipia, A., and Hsueh, A.J.W. (1997). Regulation of ovarian follicle atresia. Annu Rev Physiol *59*, 349-363. 10.1146/annurev.physiol.59.1.349.

Khachane, A.N., Timmis, K.N., and Martins dos Santos, V.A. (2007). Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. Mol Biol Evol *24*, 449-456.

Kim, I., Rodriguez-Enriquez, S., and Lemasters, J.J. (2007). Selective degradation of mitochondria by mitophagy. Arch Biochem Biophys *462*, 245-253.

Kimura, M., and Ohta, T. (1971). On the rate of molecular evolution. J Mol Evol *1*, 1-17.

Klekowski, E.J. (2003). Plant clonality, mutation, diplontic selection and mutational meltdown. Biol J Linn Soc *79*, 61-67.

Klekowski, E.J., Kazarinova-Fukshansky, N., and Mohr, H. (1985). Shoot apical meristems and mutation: stratified meristems and angiosperm evolution. Am J Bot *72*, 1788-1800.

Kloc, M., Bilinski, S., and Etkin, L.D. (2004). The Balbiani body and germ cell determinants: 150 years later. Curr Top Dev Biol *59*, 1-36.

Koltunow, A.M., and Grossniklaus, U. (2003). Apomixis: a developmental perspective. Annu Rev Plant Biol *54*, 547-574.

Kondrashov, A.S. (1994). Muller's ratchet under epistatic selection. Genetics *136*, 1469-1473.

Koonin, E.V. (2006). The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol Direct *1*, 1-23.

Koonin, E.V. (2009a). Evolution of genome architecture. Int J Biochem Cell Biol *41*, 298-306.

Koonin, E.V. (2009b). Intron-dominated genomes of early ancestors of eukaryotes. J Hered *100*, 618-623.

Koonin, E.V., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., and Rao, B.S. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol *5*, 1-28.

Kopinski, P.K., Janssen, K.A., Schaefer, P.M., Trefely, S., Perry, C.E., Potluri, P., Tintos-Hernandez, J.A., Singh, L.N., Karch, K.R., Campbell, S.L., et al. (2019). Regulation of nuclear epigenome by mitochondrial DNA heteroplasmy. Proc Natl Acad Sci U S A *116*, 16028-16035. 10.1073/pnas.1906896116.

Koumandou, V.L., Wickstead, B., Ginger, M.L., Van Der Giezen, M., Dacks, J.B., and Field, M.C. (2013). Molecular paleontology and complexity in the last eukaryotic common ancestor. Crit Rev Biochem Mol Biol *48*, 373-396.

Krakauer, D.C., and Mira, A. (1999). Mitochondria and germ-cell death. Nature *400*, 125-126.

Kullback, S., and Leibler, R.A. (1951). On Information and Sufficiency. Ann Math Statist *22*, 79-86. 10.1214/aoms/1177729694.

Kurland, C., and Ehrenberg, M. (1987). Growth-optimizing accuracy of gene expression. Annu Rev Biophys *16*, 291-317.

Lahr, D.J., Parfrey, L.W., Mitchell, E.A., Katz, L.A., and Lara, E. (2011). The chastity of amoebae: re-evaluating evidence for sex in amoeboid organisms. Proc R Soc B Biol Sci *278*, 2081-2090.

Lambert, J.D., and Moran, N.A. (1998). Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. Proc Natl Acad Sci U S A *95*, 4458-4462.

Lambowitz, A.M., and Belfort, M. (2015). Mobile bacterial group II introns at the crux of eukaryotic evolution. Microbiol Spectrum *3*, 3.1. 04.

Lane, N. (2005). Power, Sex, Suicide: Mitochondria and the Meaning of Life (Oxford University Press).

Lane, N. (2007). Mitochondria: key to complexity. In Origin of mitochondria and hydrogenosomes, (Springer), pp. 13-38.

Lane, N. (2011). Energetics and genetics across the prokaryote-eukaryote divide. Biol Direct *6*, 1-31.

Lane, N. (2014). Bioenergetic constraints on the evolution of complex life. Cold Spring Harbor Perspect Biol *6*, a015982.

Lane, N. (2020). How energy flow shapes cell evolution. Curr Biol *30*, R471-R476.

Lane, N., and Martin, W. (2010). The energetics of genome complexity. Nature *467*, 929-934.

Lanfear, L. (2018). Do plants have a segregated germline? PLoS Biol *16*, e2005439.

Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., and Gray, M.W. (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature *387*, 493-497.

Lapierre, P., and Gogarten, J.P. (2009). Estimating the size of the bacterial pan-genome. Trends Genet *25*, 107-110.

Lapinski, J., and Tunnacliffe, A. (2003). Anhydrobiosis without trehalose in bdelloid rotifers. FEBS Lett *553*, 387-390.

Larkman, A.U. (1984). The fine structure of mitochondria and the mitochondrial cloud during oogenesis on the sea anemone Actinia. Tissue and Cell *16* (*3*), 393-404.

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

Lee M.C., and Marx C.J. (2012). Repeated, Selection-Driven Genome Reduction of Accessory Genes in Experimental Populations. PLOS Genetics 8(5): e1002651.

Lee, G., Sherer, N.A., Kim, N.H., Rajic, E., Kaur, D., Urriola, N., Martini, K.M., Xue, C., Goldenfeld, N., and Kuhlman, T.E. (2018). Testing the retroelement invasion hypothesis for the emergence of the ancestral eukaryotic cell. Proc Natl Acad Sci U S A *115*, 12465-12470.

Lei, L., and Spradling, A.C. (2016). Mouse oocytes differentiate through organelle enrichment from sister cyst germ cells. Science *252*, 95-99.
Lenormand, T., Engelstädter, J., Johnston, S.E., Wijnker, E., and Haag, C.R. (2016). Evolutionary mysteries in meiosis. Phil Trans R Soc B *371*, 20160001.
Lenormand, T., and Otto, S.P. (2000). The evolution of recombination in a heterogeneous environment. Genetics *156*, 423-438.
Levin, B.R., and Cornejo, O.E. (2009). The population and evolutionary dynamics of homologous gene recombination in bacteria. PLoS Genet *5*, e1000601.
Li, M., Rothwell, R., Vermaat, M., Wachsmuth, M., Schröder, R., Laros, J.F.J., van Oven, M., de Bakker, P.I.W., Bovenberg, J.A., van Duijn, C.M., et al. (2016). Transmission of human mtDNA heteroplasmy in the Genome of the Netherlands families: support for a variable-size bottleneck. Genome Res *26*, 417-426. 10.1101/gr.203216.115.
Lieber, T., Jeedigunta, S.P., Palozzi, J.M., Lehmann, R., and Hurd, T.R. (2019). Mitochondrial fragmentation drives selective removal of deleterious mtDNA in the germline. Nature *570*, 380-384. 10.1038/s41586-019-1213-4.
Lin, Z., Kong, H., Nei, M., and Ma, H. (2006). Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. Proc Natl Acad Sci U S A *103*, 10328-10333.
Loewe, L. (2006). Quantifying the genomic decay paradox due to Muller's ratchet in human mitochondrial DNA. Genet Res *87*, 133-159.
Loewe, L., and Lamatsch, D.K. (2008). Quantifying the threat of extinction from Muller's ratchet in the diploid Amazon molly (Poecilia formosa). BMC Evol Bio *8*, 1-20.
López-Madrigal, S., and Gil, R. (2017). Et tu, brute? not even intracellular mutualistic symbionts escape horizontal gene transfer. Genes *8*, 247.
López-Madrigal, S., Latorre, A., Porcar, M., Moya, A., and Gil, R. (2011). Complete genome sequence of “Candidatus Tremblaya princeps” strain PCVAL, an intriguing translational machine below the living-cell status. Am Soc Microbiol.
Lovell, J.T., Williamson, R.J., Wright, S.I., McKay, J.K., and Sharbel, T.F. (2017). Mutation accumulation in an asexual relative of Arabidopsis. PLoS Genet *13*, e1006550.
Lynch, M. (1996). Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol Biol Evol *13*, 209-220.

Lynch, M., Bürger, R., Butcher, D., and Gabriel, W. (1993). The mutational meltdown in asexual populations. J Hered *84*, 339-344.

Lynch, M., Koskella, B., and Schaack, S. (2006). Mutation pressure and the evolution of organelle genomic architecture. Science *311*, 1727-1739.

Makarova, K.S., Wolf, Y.I., Mekhedov, S.L., Mirkin, B.G., and Koonin, E.V. (2005). Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. Nucleic Acids Res *33*, 4626-4638.

Malik, S.-B., Pightling, A.W., Stefaniak, L.M., Schurko, A.M., and Logsdon Jr, J.M. (2008). An expanded inventory of conserved meiotic genes provides evidence for sex in Trichomonas vaginalis. PloS 1 *3*, e2879.

Marri, P.R., Hao, W., and Golding, G.B. (2006). Gene gain and gene loss in streptococcus: is it driven by habitat? Mol Biol Evol *23*, 2379-2391.

Marri, P.R., Hao, W., and Golding, G.B. (2007). The role of laterally transferred genes in adaptive evolution. BMC Evol Bio *7*, 1-14.

Martens, K., Rossetti, G., and Horne, D.J. (2003). How ancient are ancient asexuals? Proc R Soc B Biol Sci *270*, 723-729. doi:10.1098/rspb.2002.2270.

Martin, W., Garg, S., and Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. Phil Trans R Soc B *370*, 20140330.

Martin, W., and Koonin, E.V. (2006). Introns and the origin of nucleus–cytosol compartmentalization. Nature *440*, 41-45.

Martin, W.F., and Müller, M. (2007). Origin of mitochondria and hydrogenosomes (Springer).

Maynard Smith, J. (1968). Evolution in sexual and asexual populations. Am Nat *102*, 469-473.

Maynard Smith, J. (1971). What use is sex? J Theor Biol *30*, 319-335.

Maynard Smith, J. (1978). The evolution of sex (Cambridge University Press Cambridge).

Maynard Smith, J., and Szathmáry, E. (1995). The Major Transitions in Evolution (Oxford University Press).

McCutcheon, J.P., McDonald, B.R., and Moran, N.A. (2009). Origin of an alternative genetic code in the extremely small and GC–rich genome of a bacterial symbiont. PLoS Genet *5*, e1000565.

McCutcheon, J.P., and Moran, N.A. (2010). Functional Convergence in Reduced Genomes of Bacterial Symbionts Spanning 200 My of Evolution. Genome Biol Evol *2*, 708-718. 10.1093/gbe/evq055.

McCutcheon, J.P., and Moran, N.A. (2012). Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol *10*, 13-26.

McInerney, J.O., McNally, A., and O'connell, M.J. (2017). Why prokaryotes have pangenomes. Nat Microbiol *2*, 1-5.

Mell, J.C., Hall, I.M., and Redfield, R.J. (2012). Defining the DNA uptake specificity of naturally competent Haemophilus influenzae cells. Nucleic Acids Res *40*, 8536-8549.

Mell, J.C., Lee, J.Y., Firme, M., Sinha, S., and Redfield, R.J. (2014). Extensive cotransformation of natural variation into chromosomes of naturally competent Haemophilus influenzae. G3 *4*, 717-731.

Melnikov, S.V., Manakongtreecheep, K., Rivera, K.D., Makarenko, A., Pappin, D.J., and Söll, D. (2018). Muller's ratchet and ribosome degeneration in the obligate intracellular parasites microsporidia. Int J Mol Sci *19*, 4125.

Michaelis, G., Vahrenholz, C., and Pratje, E. (1990). Mitochondrial DNA of Chlamydomonas reinhardtii: the gene for apocytochrome b and the complete functional map of the 15.8 kb DNA. MGG *223*, 211-216.

Mikkola, R., and Kurland, C. (1991). Is there a unique ribosome phenotype for naturally occurring Escherichia coli? Biochimie *73*, 1061-1066.

Mira, A., Ochman, H., and Moran, N.A. (2001). Deletional bias and the evolution of bacterial genomes. Trends Genet *17*, 589-596.

Mirzaghaderi, G., and Hörandl, E. (2016). The evolution of meiotic sex and its alternatives. Proc R Soc B Biol Sci *283*, 20161221.

Moore, A.S., Coscia, S.M., Simpson, C.L., Ortega, F.E., Wait, E.C., Heddleston, J.M., Nirschl, J.J., Obara, C.J., Guedes-Dias, P., Boecker, C.A., et al. (2021). Actin cables and comet tails organize mitochondrial networks in mitosis. Nature *591*, 659-664. 10.1038/s41586-021-03309-5.

Moran, N.A. (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci U S A *93*, 2873-2878.

Moran, N.A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. Cell *108*, 583-586.

Moran, N.A., and Mira, A. (2001). The process of genome shrinkage in the obligate symbiont Buchnera aphidicola. Genome Biol *2*, Research0054. 10.1186/gb-2001-2-12-research0054.
Morita, Y., and Tilly, J.L. (1999). Oocyte apoptosis: like sand through an hourglass. Dev Biol *213*, 1-17.
Motta, P.M., Nottola, S.A., Makabe, S., and Heyn, R. (2000). Mitochondrial morphology in human fetal and adult female germ cells. Hum Reprod *15*, 129-147. 10.1093/humrep/15.suppl_2.129.
Muller, H.J. (1964). The relation of recombination to mutational advance. Mutat Res *1*, 2-9.
Nabholz, B., Glémin, S., and Galtier, N. (2009). The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. BMC Evol Bio *9*, 1-13.
Nachman, M.W., Brown, W.M., Stoneking, M., and Aquadro, C.F. (1996). Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics *142*, 953-963.
Naito, M., and Pawlowska, T.E. (2016). Defying Muller's Ratchet: Ancient heritable endobacteria escape extinction through retention of recombination and genome plasticity. mBio *7*, e02057-02015.
Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H.E., Moran, N.A., and Hattori, M. (2006). The 160-kilobase genome of the bacterial endosymbiont Carsonella. Science *314*, 267-267.
Neher, R.A., and Shraiman, B.I. (2012). Fluctuations of fitness distributions and the rate of Muller's ratchet. Genetics *191*, 1283-1293.
Nezis, I.P., Stravopodis, D.J., Papassideri, I., Robert-Nicoud, M., and Margaritis, L.H. (2000). Stage-specific apoptotic patterns during Drosophila oogenesis. Eur J Cell Biol *79*, 610-620.
Nilsson, A., Koskiniemi, S., Eriksson, S., Kugelberg, E., Hinton, J., and Andersson, D.I. (2005). Bacterial genome size reduction by experimental evolution. Proc Natl Acad Sci U S A *102*, 12112-12116.
Novella, I., and Ebendick-Corpus, B. (2004). Molecular basis of fitness loss and fitness recovery in vesicular stomatitis virus. J Mol Biol *342*, 1423-1430.
Novichkov, P.S., Wolf, Y.I., Dubchak, I., and Koonin, E.V. (2009). Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. J Bacteriol *191*, 65-73.

Nowell, R.W., Green, S., Laue, B.E., and Sharp, P.M. (2014). The extent of genome flux and its role in the differentiation of bacterial lineages. Genome Biol Evol *6*, 1514-1529.
Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. Nature *405*, 299-304.
Omilian, A.R., Cristescu, M.E., Dudycha, J.L., and Lynch, M. (2006). Ameiotic recombination in asexual lineages of Daphnia. Proc Natl Acad Sci U S A *103*, 18638-18643.
Otto, S.P. (2009). The evolutionary enigma of sex. Am Nat *174*, S1-S14.
Otto, S.P., and Lenormand, T. (2002). Resolving the paradox of sex and recombination. Nat Rev Genet *3*, 252-261.
Pál, C., Papp, B., and Lercher, M.J. (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat Genet *37*, 1372-1375.
Palca, J. (1990). The other human genome. Science *249*, 1104-1105.
Park, Y.Y., and Cho, H. (2012). Mitofusin 1 is degraded at G2/M phase through ubiquitylation by MARCH5. Cell Div *7*, 25. 10.1186/1747-1028-7-25.
Pellino, M., Hojsgaard, D., Schmutzer, T., Scholz, U., Hörandl, E., Vogel, H., and Sharbel, T.F. (2013). Asexual genome evolution in the apomictic R anunculus auricomus complex: examining the effects of hybridization and mutation accumulation. Mol Ecol *22*, 5908-5921.
Pepling, M.E. (2006). From primordial germ cell to primordial follicle: mammalian female germ cell development. Genesis *44*, 622-632. 10.1002/dvg.20258.
Pepling, M.E., Wilhelm, J.E., Hara, A.L., Gephardt, G.W., and Spradling, A.C. (2007). Mouse oocytes within germ cell cysts and primordial follicles contain a Balbiani body. Proc Natl Acad Sci U S A *104*, 187. 10.1073/pnas.0609923104.
Pereira, C.V., Gitschlag, B.L., and Patel, M.R. (2021). Cellular mechanisms of mtDNA heteroplasmy dynamics. Crit Rev Biochem Mol Biology, 1-16.
Perez, G.I., Trbovich, A.M., Gosden, R.G., and Tilly, J.L. (2000). Mitochondria and the death of oocytes. Nature *403*, 500-501. 10.1038/35000651.
Pittis, A.A., and Gabaldón, T. (2016). Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature *531*, 101-104.
Polz, M.F., Alm, E.J., and Hanage, W.P. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet *29*, 170-175.
Puchta, O., Cseke, B., Czaja, H., Tollervey, D., Sanguinetti, G., and Kudla, G. (2016). Network of epistatic interactions within a yeast snoRNA. Science *352*, 840-844.

Puigbò, P., Lobkovsky, A.E., Kristensen, D.M., Wolf, Y.I., and Koonin, E.V. (2014). Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes. BMC Biol *12*, 1-19.

Pylkov, K.V., Zhivotovsky, L.A., and Feldman, M.W. (1998). Migration versus mutation in the evolution of recombination under multilocus selection. Genet Res *71*, 247-256.

Radzvilavicius, A.L., Hadjivasiliou, Z., Pomiankowski, A., and Lane, N. (2016). Selection for mitochondrial quality drives evolution of the germline. PLoS Biol *14*, e2000410.

Radzvilavicius, A.L., Lane, N., and Pomiankowski, A. (2017). Sexual conflict explains the extraordinary diversity of mechanisms regulating mitochondrial inheritance. BMC Biol *14* (*94*).

Ramesh, M.A., Malik, S.-B., and Logsdon Jr, J.M. (2005). A phylogenomic inventory of meiotic genes: evidence for sex in Giardia and an early eukaryotic origin of meiosis. Curr Biol *15*, 185-191.

Rand, D.M. (2001). The units of selection on mitochondrial DNA. Annu Rev Ecol Sys *32*, 415-448.

Rand, D.M., and Kann, L.M. (1996). Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol Biol Evol *13*, 735-748.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J.-M. (2004). The 1.2-megabase genome sequence of Mimivirus. Science *306*, 1344-1350.

Rasch, E., Monaco, P.J., and Balsano, J.S. (1982). Cytophotometric and autoradiographic evidence for functional apomixis in a gynogenetic fish, Poecilia formosa and its related, triploid unisexuals. Histochemistry *73*, 515-533.

Rebolledo-Jaramillo, B., Su, M.S.-W., Stoler, N., McElhoe, J.A., Dickins, B., Blankenberg, D., Korneliussen, T.S., Chiaromonte, F., Nielsen, R., Holland, M.M., et al. (2014). Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A *111*, 15474. 10.1073/pnas.1409328111.

Redfield, R.J. (1988). Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all? Genetics *119*, 213-221.

Redfield, R.J., Schrag, M.R., and Dean, A.M. (1997). The evolution of bacterial transformation: sex with poor relations. Genetics *146*, 27-38.

Reunov, A., Alexandrova, Y., Reunova, Y., Komkova, A., and Milani, L. (2019). Germ plasm provides clues on meiosis: the concerted action of germ plasm granules and mitochondria in gametogenesis of the clam Ruditapes philippinarum. Zygote *27*, 25-35. 10.1017/S0967199418000588.

Reynier, P., May-Panloup, P., Chretien, M., Morgan, C., Jean, M., Savagner, F., Barriere, P., and Malthiery, Y. (2001). Mitochondrial DNA content affects the fertilizability of human oocytes. Mol Hum Repr *7*, 425-429.

Richards, A. (2003). Apomixis in flowering plants: an overview. Phil Trans R Soc B *358*, 1085-1093.

Rodrigues, P., Limback, D., McGinnis, L.K., Plancha, C.E., and Albertini, D.F. (2009). Multiple mechanisms of germ cell loss in the perinatal mouse ovary. Reproduction *137*, 709-720.

Roger, A.J., Muñoz-Gómez, S.A., and Kamikawa, R. (2017). The origin and diversification of mitochondria. Curr Biol *27*, R1177-R1192.

Rogozin, I.B., Carmel, L., Csuros, M., and Koonin, E.V. (2012). Origin and evolution of spliceosomal introns. Biol Direct *7*, 1-28.

Rossetti, G., and Martens, K. (1998). Taxonomic revision of the Recent and Holocene representatives of the Family Darwinulidae (Crustacea, Ostracoda), with a description of three new genera. Bulletin van het Koninklijk Belgisch Instituut voor Natuurwetenschappen, Biologie *68*, 55-110.

Rossignol, R., Faustin, B., Rocher, C., Malgat, M., Mazat, J.P., and Letellier, T. (2003). Mitochondrial threshold effects. Biochem J *370*, 751-762.

Roze D. (2014). Selection for sex in finite populations. J Evol Biol. *27*:1304–1322

Roze, D., and Barton, N.H. (2006). The Hill–Robertson effect and the evolution of recombination. Genetics *173*, 1793-1811.

Roze, D., Rousset, F., and Michalakis, Y. (2005). Germline bottlenecks, biparental inheritance and selection on mitochondrial variants: a two-level selection model. Genetics *170*, 1385-1399. 10.1534/genetics.104.039495.

Saidapur, S.K. (1978). Follicular atresia in the ovaries of nonmammalian vertebrates. Int Rev Cytol *54*, 225-244.

Schaefer, A.M., Blakely, E.L., He, L., Whittaker, R.G., Taylor, R.W., Chinnery, P.F., and Turnbul, D.M. (2008). Prevalence of mitochondrial DNA disease in adults. Ann Neurol *63*, 35-39.

Schartl, M., Nanda, I., Schlupp, I., Wilde, B., Epplen, J.T., Schmid, M., and Parzefall, J. (1995a). Incorporation of subgenomic amounts of DNA as compensation for mutational load in a gynogenetic fish. Nature *373*, 68-71.

Schartl, M., Wilde, B., Schlupp, I., and Parzefall, J. (1995b). Evolutionary origin of a parthenoform, the Amazon molly Poecilia formosa, on the basis of a molecular genealogy. Evolution *49*, 827-835.

Schön, I., Martens, K., Van Doninck, K., and Butlin, R.K. (2003). Evolution in the slow lane: molecular rates of evolution in sexual and asexual ostracods (Crustacea: Ostracoda). Biol J Linn Soc *79*, 93-100.

Schurko, A.M., and Logsdon, J.M. (2008). Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex. Bioessays *30*, 579-589.

Seitz, E.M., Brockman, J.P., Sandler, S.J., Clark, A.J., and Kowalczykowski, S.C. (1998). RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. Genes Dev *12*, 1248-1253.

Seitz, P., and Blokesch, M. (2013). Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. FEMS Microbiol Rev *37*, 336-363.

Sela, I., Wolf, Y.I., and Koonin, E.V. (2016). Theory of prokaryotic genome evolution. Proc Natl Acad Sci U S A *113*, 11399-11407.

Selosse, M.-A., Albert, B., and Godelle, B. (2001). Reducing the genome size of organelles favours gene transfer to the nucleus. Trends Ecol Evol *16*, 135-141.

Shoubridge, E.A., and Wai, T. (2007). Mitochondrial DNA and the mammalian oocyte. Curr Top Dev Biol *77*, 87-111. 10.1016/s0070-2153(06)77004-1.

Siguier, P., Gourbeyre, E., and Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev *38*, 865-891.

Sigurdardottir, S., Helgason, A., Gulcher, J.R., and Donnelly, P. (2000). The mutation rate in the human mtDNA control region. Am J Hum Genet *66*, 1599-1609.

Silva, F.J., Latorre, A., and Moya, A. (2001). Genome size reduction through multiple events of gene disintegration in Buchnera APS. Trends Genet *17*, 615-618. 10.1016/s0168-9525(01)02483-0.

Speijer, D. (2015). Birth of the eukaryotes by a set of reactive innovations: new insights force us to relinquish gradual models. Bioessays *37*, 1268-1276.

Speijer, D., Lukeš, J., and Eliáš, M. (2015). Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. Proc Natl Acad Sci U S A *112*, 8827-8834.

Stephan, W., Chao, L., and Smale, J.G. (1993). The advance of Muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory. Genet Res *61*, 225-231.

Stewart, J.B., and Chinnery, P.F. (2015). The dynamics of mitochondrial DNA heteroplasmy : implications for human health and disease. Nat Rev Genet *16*, 530-542.

Stewart, J.B., Freyer, C., Elson, J.L., Wredenberg, A., Cansu, Z., Trufunovic, A., and N., L. (2008). Strong purifying selection in transmission of mammalian mitochondrial DNA. PLoS Biol *6*, 63-71.

Stewart, J.B., and Larsson, N.-G. (2014). Keeping mtDNA in shape between generations. PLoS Genet *10*, e1004670-e1004670. 10.1371/journal.pgen.1004670.

Suganuma, N., Kitagawa, T., Nawa, A., and Tomoda, Y. (1993). Human ovarian aging and mitochondrial DNA deletion. Horm Res *39* (*1*), 16-21. 10.1159/000182752.

Sung, W., Tucker, A.E., Doak, T.G., Choi, E., Thomas, W.K., and Lynch, M. (2012). Extraordinary genome stability in the ciliate Paramecium tetraurelia. Proc Natl Acad Sci U S A *109*, 19339-19344.

Szathmáry, E. (2015). Toward major evolutionary transitions theory 2.0. Proc Natl Acad Sci U S A *112*, 10104-10111.

Szollosi, G.J., Derényi, I., and Vellai, T. (2006). The maintenance of sex in bacteria is ensured by its potential to reload genes. Genetics *174*, 2173-2180.

Taguchi, N., Ishihara, N., Jofuku, A., Oka, T., and Mihara, K. (2007). Mitotic phosphorylation of dynamin-related GTPase Drp1 participates in mitochondrial fission. J Biol Chem *282*, 11521-11529. 10.1074/jbc.M607279200.

Takeuchi, N., Kaneko, K., and Koonin, E.V. (2014). Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: benefit of DNA from dead cells and population subdivision. G3 *4*, 325-339.

Tenaillon, O., Barrick, J.E., Ribeck, N., Deatherage, D.E., Blanchard, J.L., Dasgupta, A., Wu, G.C., Wielgoss, S., Cruveiller, S., and Médigue, C. (2016). Tempo and mode of genome evolution in a 50,000-generation experiment. Nature *536*, 165-170.

Thomas, C.M., and Nielsen, K.M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol *3*, 711-721.

Tilly, J.L. (2001). Commuting the death sentence: how oocytes strive to survive. Nat Rev Mol Cell Biol *2*, 838-848.
Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet *5*, 123-135.
Touchon, M., and Rocha, E.P. (2007). Causes of insertion sequences abundance in prokaryotic genomes. Mol Biol Evol *24*, 969-981.
Townson, D.H., and Combelles, C.M.H. (2012). Ovarian Follicular Atresia. (InTech).
Treangen, T.J., and Rocha, E.P. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. PLoS Genet *7*, e1001284.
Tria, F.D., Brueckner, J., Skejo, J., Xavier, J.C., Kapust, N., Knopp, M., Wimmer, J.L., Nagies, F.S., Zimorski, V., and Gould, S.B. (2021). Gene duplications trace mitochondria to the onset of eukaryote complexity. Genome Biol Evol *13*, evab055.
Trifunovic, A., Hansson, A., Wredenberg, A., Rovio, A.T., Dufour, E., Khvorostov, I., Spelbrink, J.N., Wibom, R., Jacobs, H.T., and Larsson, N.G. (2005). Somatic mtDNA mutations cause aging phenotypes without affecting reactive oxygen species production. Proc Natl Acad Sci U S A *102*, 17993-17998. 10.1073/pnas.0508886102.
Twig, G., Elorza, A., Molina, A.J.A., Mohamed, H., Wikstrom, J.D., Walzer, G., Stiles, L., Haigh, S.E., Katz, S., Las, G., et al. (2008). Fission and selective fusion govern mitochondrial segregation and elimination by autophagy. EMBO J *27*, 433-446. 10.1038/sj.emboj.7601963.
Tworzydlo, W., Kisiel, E., Jankowska, W., Witwicka, A., and Bilinski, S.M. (2016). Exclusion of dysfunctional mitochondria from Balbiani body during early oogenesis of Thermobia. Cell Tissue Res *366*, 191-201. 10.1007/s00441-016-2414-x.
van Dijk, B., Hogeweg, P., Doekes, H.M., and Takeuchi, N. (2020). Slightly beneficial genes are retained by bacteria evolving DNA uptake despite selfish elements. Elife *9*, e56801.
Van Valen, L. (1977). The red queen. Am Nat *111*, 809-810.
Vandecraen, J., Chandler, M., Aertsen, A., and Van Houdt, R. (2017). The impact of insertion sequences on bacterial genome plasticity and adaptability. Crit Rev Microbiol *43*, 709-730.
Vellai, T., and Vida, G. (1999). The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. Proc R Soc B Biol Sci *266*, 1571-1577.
Vergara, D., Jokela, J., and Lively, C.M. (2014). Infection dynamics in coexisting sexual and asexual host populations: support for the Red Queen hypothesis. Am Nat *184*, S22-S30.

Vos, M., Hesselman, M.C., Te Beek, T.A., van Passel, M.W., and Eyre-Walker, A. (2015). Rates of lateral gene transfer in prokaryotes: high but why? Trends Microbiol *23*, 598-605.
Vosseberg, J., van Hooff, J.J., Marcet-Houben, M., van Vlimmeren, A., van Wijk, L.M., Gabaldón, T., and Snel, B. (2021). Timing the origin of eukaryotic cellular complexity with ancient duplications. Nat Ecol Evol *5*, 92-100.
Wagner, A. (2006). Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. Mol Biol Evol *23*, 723-733.
Wagner, A., Whitaker, R.J., Krause, D.J., Heilers, J.-H., Van Wolferen, M., Van Der Does, C., and Albers, S.-V. (2017). Mechanisms of gene flow in archaea. Nat Rev Microbiol *15*, 492-501.
Wai, T., Teoli, D., and Shoubridge, E.A. (2008). The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. Nat Genet *40*, 1484-1488.
Wallace, D.C., and Chalkia, D. (2013). Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. Cold Spring Harb Perspect Biol *5*, a021220.
Ward, B.L., Anderson, R.S., and Bendich, A.J. (1981). The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). Cell *25*, 793-803.
Welch, D.B.M., and Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science *288*, 1211-1215.
Welch, D.B.M., and Meselson, M. (2001). Rates of nucleotide substitution in sexual and anciently asexual rotifers. Proc Natl Acad Sci U S A *98*, 6720-6724.
Wiedenbeck, J., and Cohan, F.M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev *35*, 957-976.
Wilkins, A.S., and Holliday, R. (2009). The evolution of meiosis from mitosis. Genetics *181*, 3-12.
Williams, T.A., Foster, P.G., Cox, C.J., and Embley, T.M. (2013). An archaeal origin of eukaryotes supports only two primary domains of life. Nature *504*, 231-236.
Wylie, C.S., Trout, A.D., Kessler, D.A., and Levine, H. (2010). Optimal strategy for competence differentiation in bacteria. PLoS Genet *6*, e1001108.
Xu, S., Schaack, S., Seyfert, A., Choi, E., Lynch, M., and Cristescu, M.E. (2012). High mutation rates in the mitochondrial genomes of *Daphnia pulex*. Mol Biol Evol *29*, 763-769. 10.1093/molbev/msr243.

Yang, Z., and Nielsen, R. (2008). Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. Mol Biol Evol *25*, 568-579.
Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., and Kjeldsen, K.U. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature *541*, 353-358.
Zenkteler, E. (1999). Sporophytic lethality in lowland populations of Osmunda regalis L. in Poland. Acta Biol Cracov.
Zeyl, C., Mizesko, M., and De Visser, J.A.G.M. (2001). Mutational meltdown in laboratory yeast populations. Evolution *55*, 909-917.
Zhou, R.R., Wang, B., Wang, J., Schatten, H., and Zhang, Y.Z. (2010). Is the mitochondrial cloud the selection machinery for preferentially transmitting wild-type mtDNA between generations? Rewinding Müller's ratchet efficiently. Curr Genet *56*, 101-107.

Appendix A – Supplementary Material for Chapter 2

In absence of LGT, previous theoretical results (Haigh, 1978) have shown that, at equilibrium, the number of individuals in the least-loaded class (LLC) is $n_0 = Ne^{-U/s}$. Without recombination and back-mutation, the loss of the LLC is an irreversible process – a 'click' of the ratchet. The magnitude of n_0 determines the likelihood that the least-loaded class becomes extinct because of stochastic fluctuations (i.e. genetic drift). High values of n_0 increase the expected time of extinction of the LLC, whereas small values make the LLC more vulnerable to stochastic fluctuations (Muller, 1964; Haigh, 1978;). Therefore n_0 is a good indication of the speed of the ratchet (Haigh, 1978). Expressing the genome wide mutation rate U as $\mu \times g$ allows the equilibrium number of individuals in the LLC to be rewritten $n_0 = Ne^{-\mu g/s}$. The speed of the ratchet scales exponentially with genome size and mutation rate, and is negatively correlated with the strength of selection. Crucially, the impact of genome size is much stronger than that of population size (**Figure 2.S1**). For example, a 2-fold increase in genome size can increase the speed of the ratchet by several orders of magnitude, whereas even a 10-fold reduction in population size has a rather meagre effect, except at low values (**Figure 2.S1**). Therefore, any increase in genome size must be balanced by a proportional increase in strength of selection in order to avoid a drastic reduction of n_0.

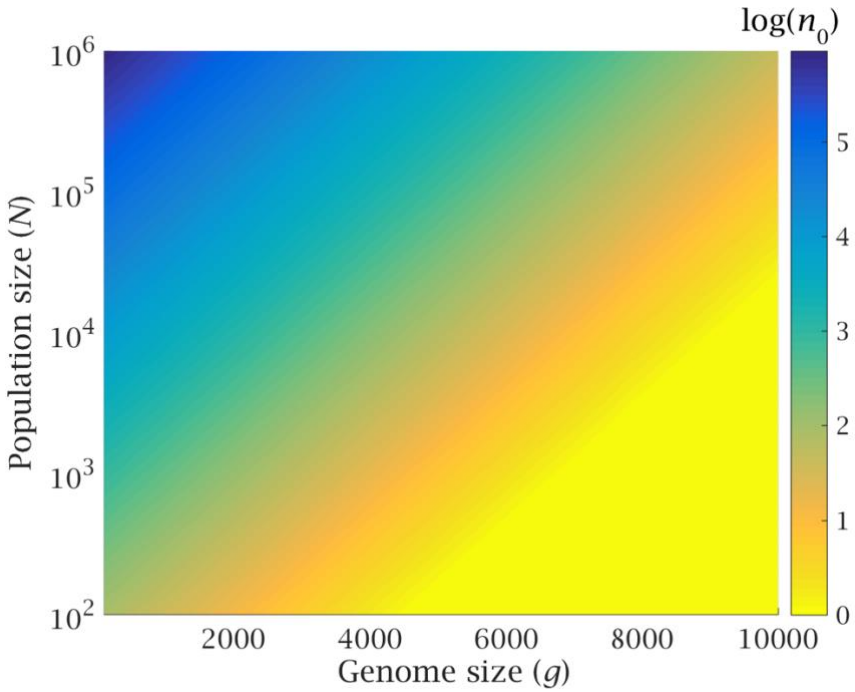

Figure 2.S1 | Genome size and population size determine n_0. The equilibrium number of individuals in the least-loaded class (n_0) is shown as a function of genome size (number of loci) and population size, with constant mutation rate $\mu = 10^{-5}$ and constant strength of selection $s = 10^{-3}$.

Appendix B – Supplementary Material for Chapter 4

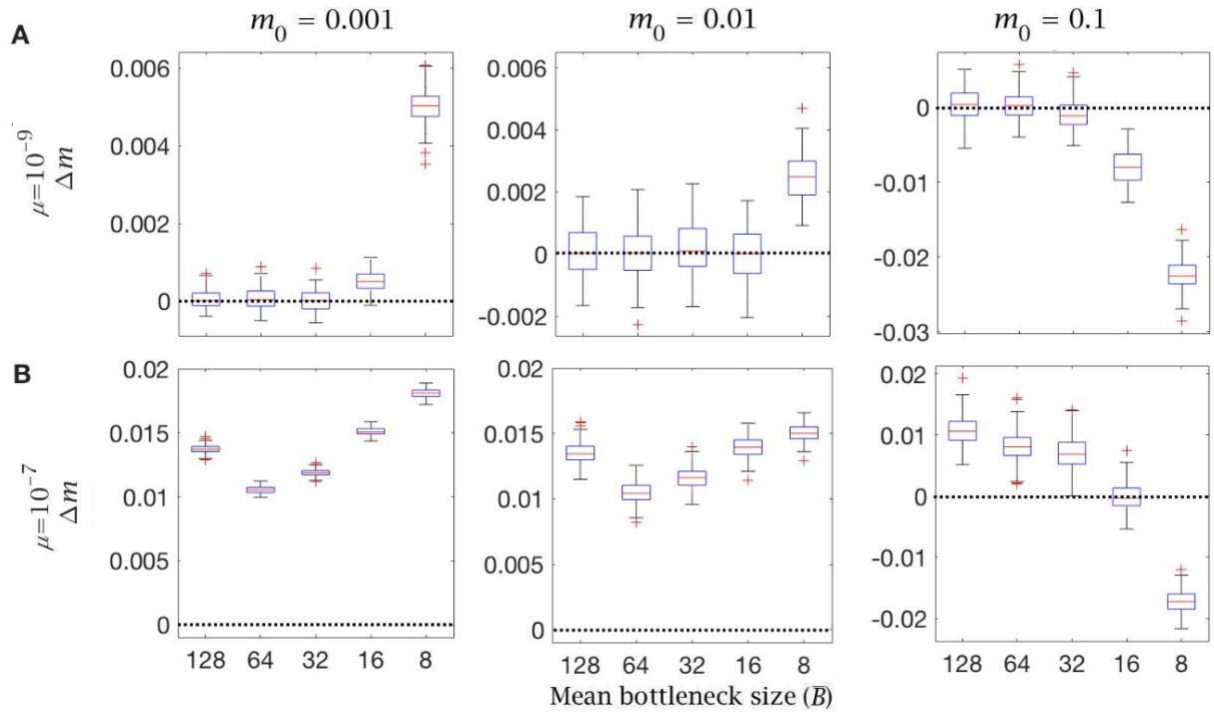

Figure 4.S1. Change in mutation load (Δm) across a single generation after individual selection with variable mean bottleneck size ($\bar{B}$). This is shown with (**A**) low ($\mu = 10^{-9}$) and (**B**) high ($\mu = 10^{-7}$) mutation rate, for individuals with low ($m_0 = 0.001$), medium ($m_0 = 0.01$) and high ($m_0 = 0.1$) initial mutation loads. Box plots show the median (red line) and distribution (box plot IQR with min/max whiskers and outliers).

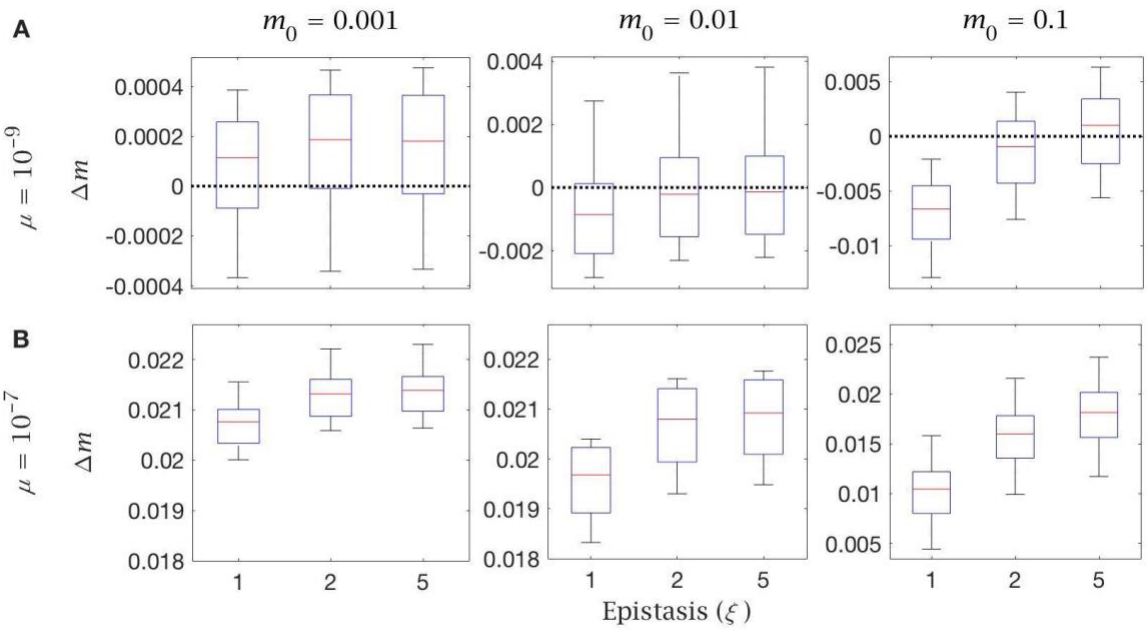

Figure 4.S2. Change in mutation load (Δm) across a single generation after cell selection with variable levels of epistasis (ξ). This is shown with (**A**) low ($\mu = 10^{-9}$) and (**B**) high ($\mu = 10^{-7}$) mutation rate, for individuals with low ($m_0 = 0.001$), medium ($m_0 = 0.01$) and high ($m_0 = 0.1$) initial mutation loads. Box plots show the median (red line) and distribution (box plot IQR with min/max whiskers and outliers).

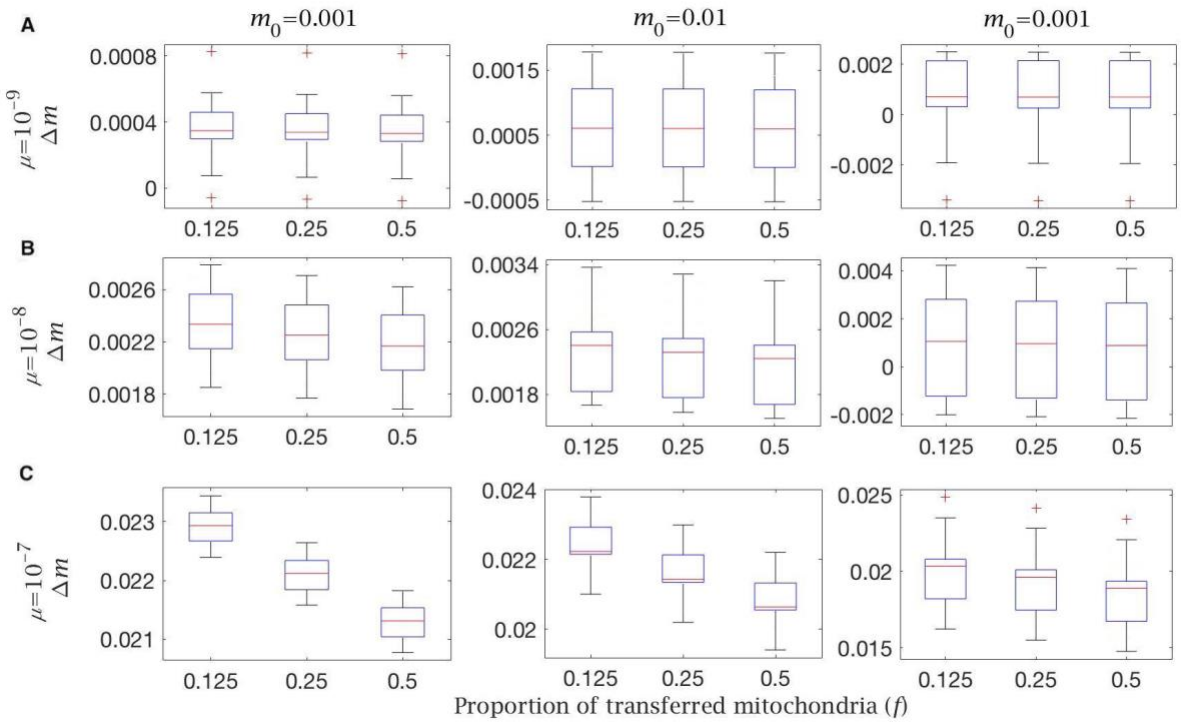

Figure 4.S3. Change in mutation load (Δm) across a single generation, given a variable proportion of transferred mitochondria (f) to the Balbiani body when transfer is non-selective ($p_{mut} = p_{wt} = 0.5$). This is shown with (**A**) low ($\mu = 10^{-9}$), (**B**), standard ($\mu = 10^{-8}$) and (**C**) high ($\mu = 10^{-7}$) mutation rate, for individuals with low ($m_0 = 0.001$), medium ($m_0 = 0.01$) and high ($m_0 = 0.1$) initial mutation loads. Box plots show the median (red line) and distribution (box plot IQR with min/max whiskers and outliers).

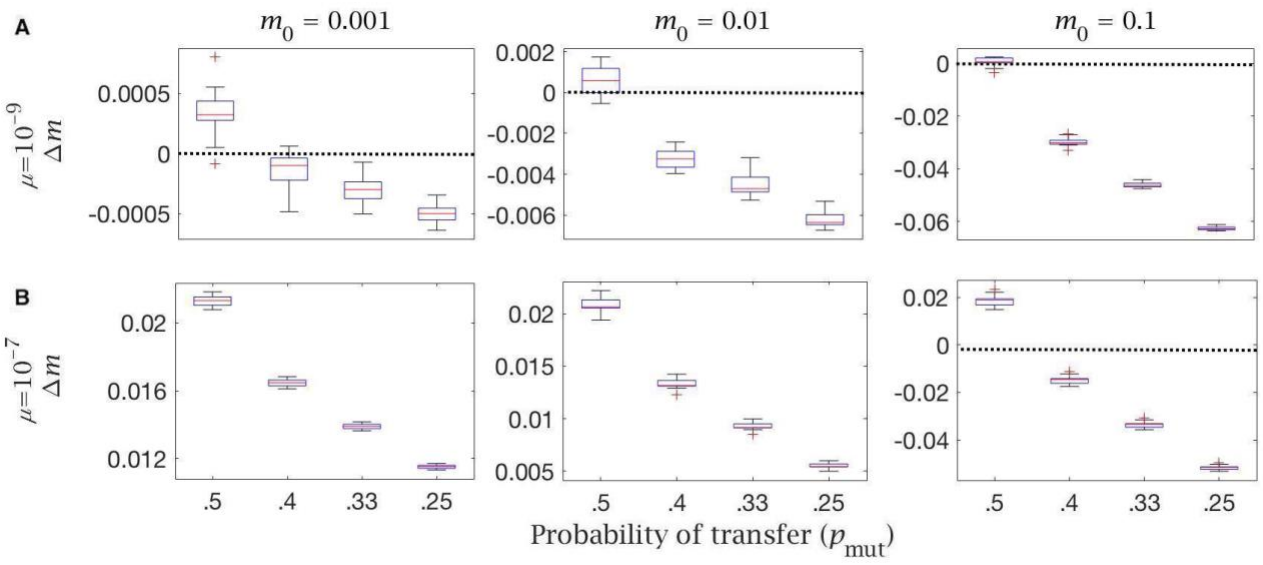

Figure 4.S4. Change in mutation load (Δm) across a single generation, for individuals undergoing cytoplasmic transfer the Balbiani body with variable strength of selection, given a fixed probability of transfer of wildtype mitochondria ($p_{wt} = 0.5$) and a decreasing probability of transfer of mutant mitochondria (p_{mut}). Note the null case is when ($p_{mut} = p_{wt} = 0.5$). This is shown with (**A**) low ($\mu = 10^{-9}$) and (**B**) high ($\mu = 10^{-7}$) mutation rate, for individuals with low ($m_0 = 0.001$), medium ($m_0 = 0.01$) and high ($m_0 = 0.1$) initial mutation loads, and a fixed proportion of transferred mitochondria ($f = 0.5$). Box plots show the median (red line) and distribution (box plot IQR with min/max whiskers and outliers).