Selection on Mitochondrial Variants Occurs between and within Individuals in an Expanding Invasion

Lee A. Rollins*[,1,2], Andrew P. Woolnough[3,4], Benjamin G. Fanson[1], Michelle L. Cummins[5], Tamsyn M. Crowley[5,6], Alan N. Wilton[‡,7], Ron Sinclair[8], Ashleigh Butler[1], and William B. Sherwin[2,9]

[1]School of Life and Environmental Sciences, Centre for Integrative Ecology, Deakin University, Geelong, Vic, Australia

[2]Evolution & Ecology Research Centre; Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW, Australia

[3]Biosecurity Branch, Department of Economic Development, Jobs, Transport and Resources, Attwood, Vic, Australia

[4]Vertebrate Pest Research Section, Department of Agriculture and Food Western Australia, Forrestfield, WA, Australia

[5]School of Medicine, Deakin University, Geelong, Vic, Australia

[6]Biosecurity Flagship, CSIRO Australian Animal Health Laboratory, Geelong, Vic, Australia

[7]School of Biotechnology and Biomolecular Science, University of New South Wales, Sydney, NSW, Australia

[8]Biosecurity SA, Adelaide, SA, Australia

[9]Murdoch University Cetacean Research Unit, Murdoch University, Murdoch, WA, Australia

[‡]Author included posthumously.

***Corresponding author:** E-mail: lee.rollins@deakin.edu.au.

Associate editor: Joshua Akey

Abstract

Mitochondria are critical for life, yet their underlying evolutionary biology is poorly understood. In particular, little is known about interaction between two levels of evolution: between individuals and within individuals (competition between cells, mitochondria or mitochondrial DNA molecules). Rapid evolution is suspected to occur frequently in mitochondrial DNA, whose maternal inheritance predisposes advantageous mutations to sweep rapidly though populations. Rapid evolution is also predicted in response to changed selection regimes after species invasion or removal of pathogens or competitors. Here, using empirical and simulated data from a model invasive bird species, we provide the first demonstration of rapid selection on the mitochondrial genome within individuals in the wild. Further, we show differences in mitochondrial DNA copy number associated with competing genetic variants, which may provide a mechanism for selection. We provide evidence for three rarely documented phenomena: selection associated with mitochondrial DNA abundance, selection on the mitochondrial control region, and contemporary selection during invasion.

Key words: **heteroplasmy, mitochondrial DNA copy number, rapid evolution, invasive species.**

Introduction

Selection on the mitochondrial genome is postulated to be strong and rapid, and therefore to constrain its use in genetic analyses (e.g., in phylogenetics, Rand 1996). The identification of mitochondrial selection remains uncommon but has been documented in a number of taxa (Grossman et al. 2004; Castoe et al. 2008; Shen et al. 2010; Finch et al. 2014). Likewise, although rapid adaption during successful invasion has been hypothesized, there is little evidence (Lee 2002, but see Rollins et al. 2013), possibly because of the difficulty of catching such rapid processes. In the haploid mitochondrial genome, selection is likely to be rapid (Hauswirth and Laipis 1982; Hill et al. 2014) because disadvantageous variants cannot persist in heterozygous form. However, this rapidity can be turned to the investigator's advantage when studying consecutive generations at a wild invasion front because such rapid changes can be observed in a relatively short study. Here, we use an invasive system to show how selection acts on multiple levels to rapidly increase the proportion of a new mutation. These findings have important implications for our understanding of evolution in changing environments and of the genetic mechanisms underpinning mitochondrial diseases in humans.

Shifts in allele frequencies can result from selection of beneficial changes, but may also be due to dispersal. Immigrants and residents may interbreed, resulting in "admixture" of their respective genes pools. Allele frequency shifts are unlikely to occur when a single unique individual enters a large resident population, because a low frequency variant will most often be lost to genetic drift. However, frequency shifts may occur in small populations or when large numbers of immigrants arrive that share a variant not previously present in the resident population. Many examples are available in the literature of anthropogenic admixture in managed populations, including byproducts of aquaculture (e.g., Arnaud-Haond et al. 2004), reintroductions of endangered species

Article

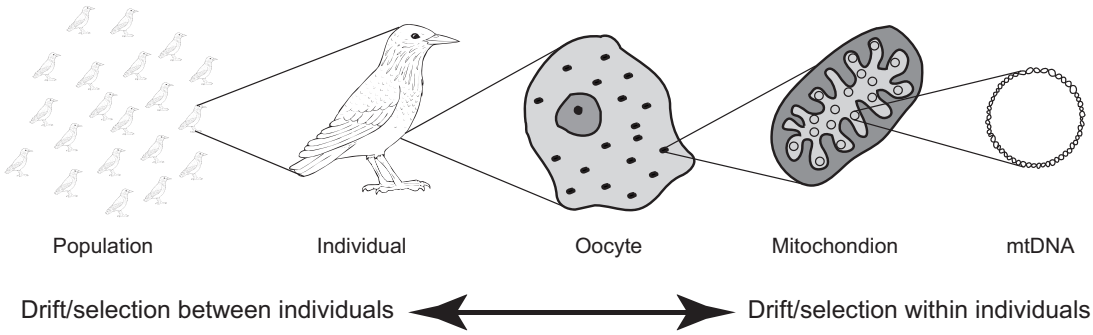

FIG. 1. Levels at which evolution on mtDNA may occur. Drift and selection occur within populations (far left) but also within individuals (acting on populations of mtDNA molecules) when two mitochondrial variants coexist (Phillips et al. 2015). Starling illustration by P. Bolton.

(Jacobsen et al. 2008), and alien introductions (Kolbe et al. 2008). Temporal genetic changes caused by admixture should be detectable across the genome. Therefore, if temporal changes occur at one locus, it would be expected that the same changes could be detected at other loci because all alleles from the dispersers would be linked by population history.

Variant proportions within populations may also shift due to random genetic changes between generations (drift) (Klopfstein et al. 2006; Hallatschek et al. 2007; Phillips et al. 2010) or selection or a combination of these forces. Additionally, for mitochondrial DNA (mtDNA), whose inheritance does not follow Mendelian genetics, there are multiple levels at which drift and selection need to be considered: At the population level, but also at the level of replication of cells, mitochondria and mtDNA molecules during the creation of gametes in the germ line in each generation (Thrailkill et al. 1980; Taylor et al. 2002). When a new point mutation arises in the mitochondrial genome, it can result in "heteroplasmy" (two different mtDNA variant "haplotypes" in a single mitochondrion), instead of the usual "homoplasmy" (only one haplotype). If a new mutation occurs in a germ cell, it may be passed to offspring, and the fate of that mutation in subsequent generations depends on a number of factors, including any competitive advantage it might have, plus random changes whose rate depends upon the number of germ line generations per animal generation and the effective number of mitochondrial molecules passed on to progeny (Chesser 1998). Rapid shifts in mtDNA haplotype proportions within lineages were first reported by Hauswirth and Laipis (1982). The speed of such shifts has been attributed to a mitochondrial bottleneck during oogenesis (Ashley et al. 1989) and the size of this bottleneck has been estimated for a number of species using population genetic theory (Hauswirth and Laipis 1982; Rand and Harrison 1986; Jenuth et al. 1996).

There are several possible mechanisms of within-individual drift or selection, whereby one mtDNA haplotype might increase at the expense of the other within a single heteroplasmic individual: between cells with different mtDNA haplotypes, between organelles with different mtDNA haplotypes (within each cell), and between alternative haplotype molecules (within each organelle; fig. 1). Experimental evidence suggests within-individual selection occurs on mitochondrial variants within oocytes; for example, an experimentally introduced mitochondrial mutation impairing oxidative phosphorylation was selectively purged from mouse germ lines whereas a less severe mutation persisted (Fan et al. 2008). The purging of this deleterious mutation may result from selective replication of mtDNA (Shoubridge and Wai 2008). Finally, although it applies more to free-living cells than those competing within a single individual, it is known that in heteroplasmic yeast there are two possible advantages to a given haplotype (a replication advantage, plus some other advantage in transmission to daughter cells) either of which might involve interaction with the nuclear genome (Taylor et al. 2002).

Here we explain patterns of mitochondrial variation in the wild and, for the first time, simultaneously investigate the combined effects of drift and selection within and between multicellular individuals. Theory of random genetic drift within populations (Birky et al. 1983; Chesser 1998; Wonnapinij et al. 2008) has rarely been combined with that for drift occurring within multicellular individuals, which can occur due to the mitochondrial bottleneck (Ashley et al. 1989; Chesser 1998), although there has been one model of selection within and between free-living yeast cells (Taylor et al. 2002). We explicitly test for the presence of temporal changes in mitochondrial haplotype proportions at the western expansion front of the common starling (*Sturnus vulgaris*) invasion in Australia. This expansion front was first identified in 2001 and data from nuclear genes suggest that although this population appears to receive low levels of immigrants from the east, it remains genetically distinct (Rollins et al. 2009). We assess differentiation of haplotypes and their relative abundances in heteroplasmic individuals. As possible explanations of the observed patterns, we evaluate the relative importance of admixture at the population level, and drift and selection at between-individual and within-individual levels.

Results

Sequencing

We Sanger sequenced a 942-bp portion of the mitochondrial control region using DNA extracted from starling liver. Across all years, five variant "haplotypes" were detected in the 181 adults sequenced (haplotype name/proportion: A/0.40, F/0.10, G/0.18, H/0.35, L/0.01; see Rollins et al. [2011] for

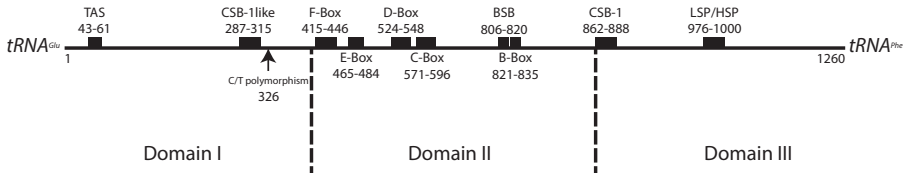

Fig. 2. Map of the mitochondrial control region of *Sturnus vulgaris*. The T/C polymorphism distinguishing haplotypes H and G is shown at position 326 of the control region, near the 3′-end of the CSB1-like sequence motif.

haplotype sequences). Haplotype proportions showed a distinct temporal trend at the expansion front, where a significant change was seen in two of the haplotypes that differed from one another by a single base pair (thymine in haplotype H and cytosine in haplotype G; Rollins et al. [2011] previously concluded that haplotype G was ancestral to haplotype H). This polymorphism is located near the 3′-end of the CSB1-like sequence (fig. 2). Over a 5-year period, the distribution of haplotypes changed due to the change in proportion of adults carrying haplotypes H and G ($Z = 3.48$, $P < 0.001$); haplotype H increased from 17% to 47%, whereas haplotype G decreased from 42% to 14% (fig. 3).

Haplotypes H and G differed at a single nucleotide (T/C, respectively) and 31% of individuals carrying haplotype G or H were heteroplasmic for these two haplotypes (Rollins et al. 2011). We created an empirical heteroplasmy distribution based on relative peak heights of T/C in electropherograms from these individuals. Figure 4A shows, for all G/H haplotype adults used in this study, the estimated number of H haplotype mitochondria per cell during the genetic bottleneck in gametogenesis, if the total number of mitochondria per oocyte is 200 (Cree et al. 2008). We validated this method of estimating heteroplasmy levels by next-generation sequencing the mitochondrial genome of a subset of 24 individuals having various levels of heteroplasmy (details below). For haplotypes whose frequency appeared to be changing (i.e., G and H; fig. 3), we calculated a coefficient of apparent selection between individuals. The selection coefficient against G relative to H (s) was 0.66 (fig. 3).

Using next-generation sequencing, we analyzed full mitochondrial genomes of multiple individuals of both haplotypes (G and H), to investigate the possibility that the mitochondrial control region polymorphism has its effect through complete linkage to polymorphisms elsewhere in the nonrecombining mitochondrial genome. The latter can be ruled out because in full genomes of 24 individuals carrying either G or H haplotypes (or both), there was no sequence variation other than the control region polymorphism that distinguishes these two haplotypes. Further, haplotype proportions within heteroplasmic individuals were calculated from these data and were significantly correlated with proportions estimated by peak-height measurement for the same 24 individuals (supplementary fig. S1, Supplementary Material online; $r = 0.98$, $P < 0.001$).

Admixture Analysis

To investigate the possibility that admixture caused the temporal changes in mitochondrial control region haplotype

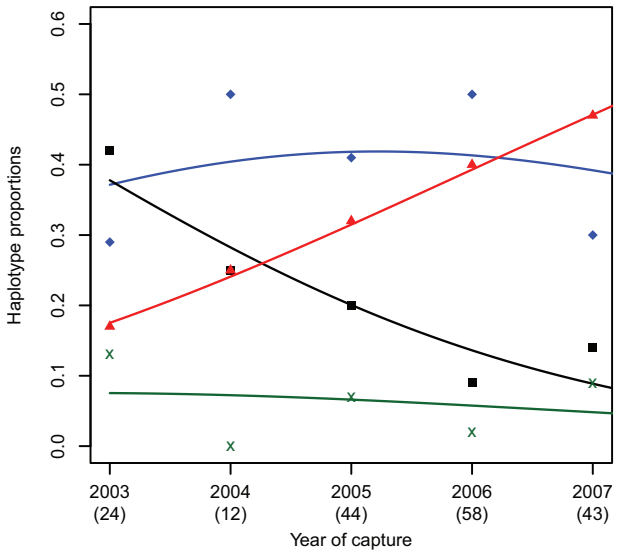

Fig. 3. Temporal variation in adult starling (*Sturnus vulgaris*) haplotype frequency on the expansion front. The 181 individuals sampled carried haplotypes H (triangles), G (squares), A (diamonds), or other (crosses). Annual sample sizes are shown below year. Multinomial logistic regression found significant divergence between G and H haplotypes ($Z = 3.48$, $P < 0.001$; selection coefficient against G relative to H, $s = 0.66$).

proportions, we studied a suite of 11 microsatellites in Hardy–Weinberg equilibrium that were capable of resolving population structure in Australian starlings (Rollins et al. 2009). If the increase in the H haplotype resulted from interbreeding with novel immigrants to the expansion front, we should see a corresponding difference at nuclear genes, because dispersal affects all genes simultaneously. However, we found no evidence for this. Although STRUCTURE (Pritchard et al. 2000; Falush et al. 2003) analysis of microsatellite data indicated that two genetic groups were present at the front (fig. 5A), there was no association between nuclear genetic group membership and the presence or absence of the H haplotype (fig. 5B; $Z = 0.564$, $P = 0.575$). Further, admixture in populations having overlapping generations, such as starlings, should result in extreme fluctuations of nuclear allele proportions for 10–20 generations (Ryman 1997). In contrast to this expectation, over the period of the study (2003–2007) the nuclear microsatellite data set showed no significant changes in allele frequencies (supplementary table S1, Supplementary Material online). Thus, it appears unlikely that the observed temporal changes of mitochondrial control region haplotype proportions (fig. 3) were due to admixture.

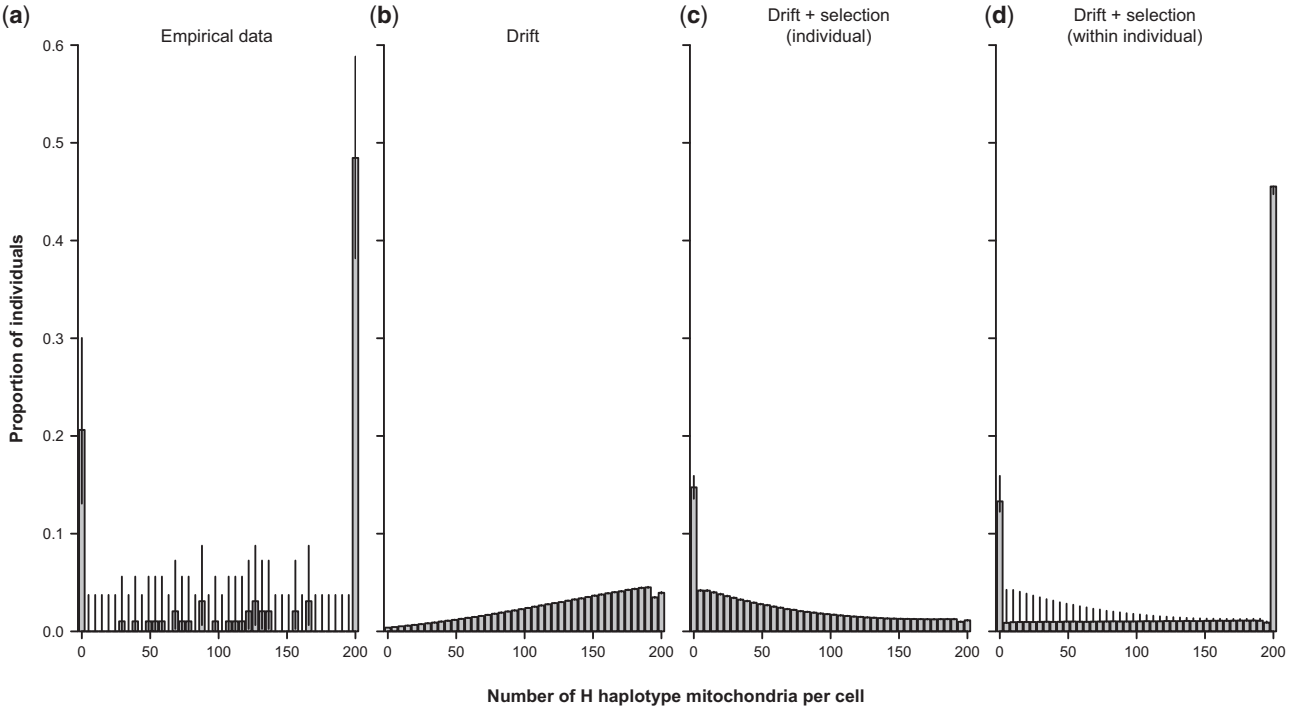

Fig. 4. Heteroplasmy distributions. Proportion of individuals having a given number of H haplotype mitochondria (*hz*) at the mitochondrial bottleneck (*mz*) during gametogenesis, assuming *mz* = 200. (*A*) Empirical data (97 individuals carrying G, H, or G/H). (*B*) Drift simulations (*hz* = 150). (*C*) Simulations of drift+selection on the individual only (selective disadvantage for individuals not carrying H was *s* = 1.0, advantage for cells/organelles/molecules carrying H was *c* = 1.00, and *hz* = 1). (*D*) Simulations of drift+selection at two levels: Between-individuals and within-individuals (*s* = 0.1, *c* = 1.01, *hz* = 1). Simulated data shown are those of each type that best fit empirical data (fig. 6 and supplementary fig. S3, Supplementary Material online). Error bars (whiskers) show 95% CI. These were based on data from iterations, for the simulation data. For the empirical data, 95% CI are binomial expected confidence limits, based on the number of observations of that frequency bin, and the total number of sequences, using the website http://statpages.org/confint.html, last accessed January 2, 2016.

Genetic Drift Simulations

Random genetic drift is accelerated by low effective population size (Halliburton 2004), as found in range expansions (Austerlitz et al. 1997; Edmonds et al. 2004), so we investigated the possibility that changes in G and H haplotype proportions were due to the random effects of drift. We used simulations accounting for genetic drift between individuals, and within individuals. In the expansion front population, heteroplasmy was observed in 39% of individuals characterized as predominantly G and 23% of individuals characterized as predominantly H. Here, we asked whether the patterns observed in empirical data (fig. 4*A*) could be explained by random genetic drift under either of two conditions: 1) If the mutation to the H haplotype occurred after founders arrived at the expansion front; or 2) if one of the founders arrived already carrying heteroplasmic DNA. To address the second condition, we ask what level of initial heteroplasmy would be required for drift to produce the patterns observed in empirical data.

Although we used scenarios with extreme values for all of the input parameters, designed to favor rapid genetic drift, it was not possible to achieve any similarity to the empirical patterns of homoplasmy and heteroplasmy when we simulated a single mutant mitochondrion (*hz* = 1) of the 200 mitochondria in a germ cell (oocyte, *mz* = 200) of a single founder. By the final generation, the mutant mitochondrial variant was lost to drift 16% of the time and reached fixation (i.e., 100% of molecules were mutant) within any individual in only 7% of these simulations (supplementary fig. S2, Supplementary Material online). Further, 0.2% of individuals in these simulations were homoplasmic for the mutant haplotype as compared with 70% in the empirical data set; thus, the overall fit of the drift simulations with *hz* = 1 was poor (fig. 6, Akaike Information Coefficient [AIC] = 985.8; supplementary fig. S2, Supplementary Material online).

We also investigated whether the observed patterns could be explained by drift scenarios commencing with *hz* > 1 (i.e., one of the founders arrived already carrying heteroplasmic DNA). We conducted simulations where *hz* = 25, 50, 75, 100, 150, or 199. The overall fit of the drift simulations with *hz* > 1 was poor (AIC ranged from 507.7 to 742.3). These simulations did not reach levels of heteroplasmy similar to those in empirical data seen in figure 4*A*; moreover, the simulated distribution of heteroplasmic individuals differed from empirical distributions (fig. 4*B* and supplementary fig. S2, Supplementary Material online). In the drift simulation that had the best fit according to AIC value (*hz* = 150, AIC = 507.7), the proportion of individuals that were homoplasmic for the H haplotype was 4%, whereas in empirical data, 48% of individuals were homoplasmic for the H haplotype (empirical data: fig. 4*A*; simulated data: figs. 4*B* and 6; supplementary

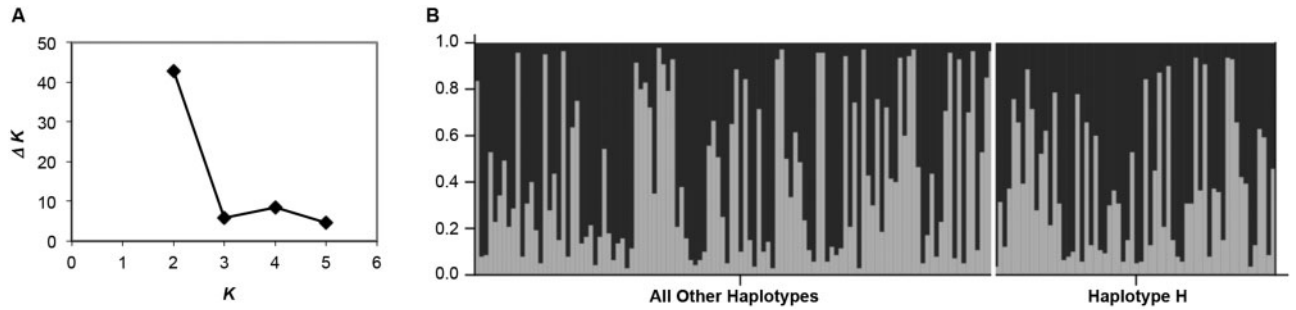

FIG. 5. Admixture analysis. Includes adults sampled between 2003 and 2007 (176 individuals). (*A*) STRUCTURE analysis of nuclear microsatellite data indicates two genetic groups are present (maximum value of ΔK when $K = 2$ genetic groups). (*B*) For each individual, the chance of membership of each of the two nuclear groupings is shown in black and gray. No evidence of nuclear group structure was seen when individuals were separated according to whether they carry mitochondrial haplotype H.

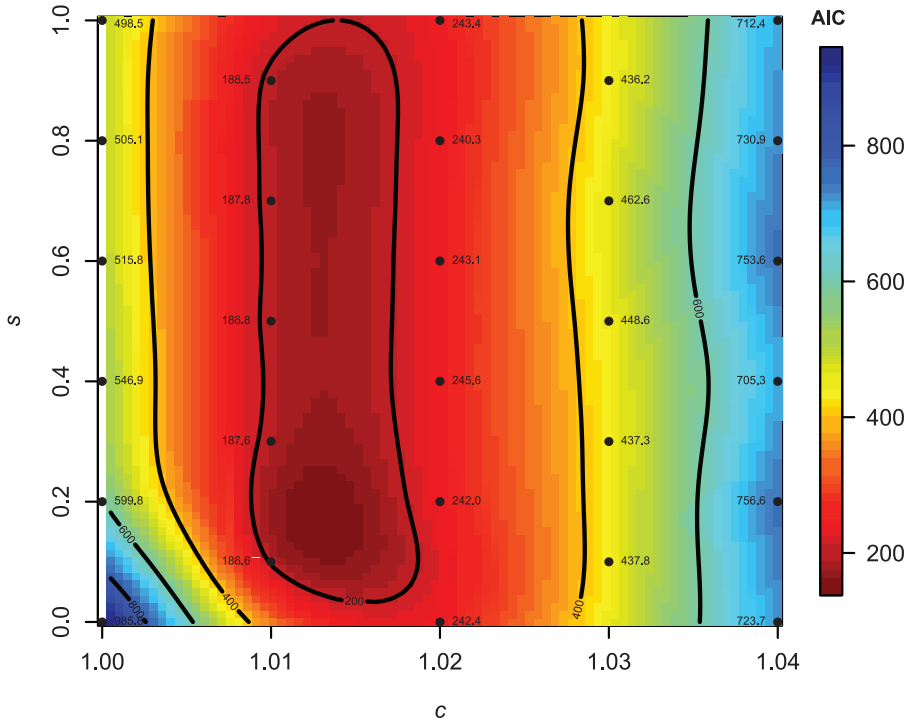

FIG. 6. Simulation model fits with $hz = 1$. Within-individual selection (c) on the horizontal axis, selection on the individual (s) on the vertical axis. Best-fit models are represented by red (lower AIC values). For each simulated selection scenario (indicated by black dots), AIC values are given. Drift alone (i.e., no selection; $s = 0$, $c = 1$) is shown at the bottom left corner (AIC = 985.8).

fig. S2, Supplementary Material online). In the best fitting drift simulation ($hz = 150$), the level of heteroplasmy (percentage of heteroplasmic individuals, P_{Ih}) was much larger than in the empirical data (supplementary fig. S2, Supplementary Material online; $hz = 150$, $P_{Ih} = 95.7$ for simulation data; $P_{Ih} = 30.9$ for empirical data).

Selection Simulations

We next investigated addition of selection in favor of the new mutant, at two different levels of organization: between- and within-individuals. For between-individual selection, each ancestral individual was assigned a chance s ($0 \leq s \leq 1$, $s = 0$ for no selection) that it would be excluded from the reproductive pool due to infertility or early death. For within-individual selection, c was the relative advantage of mutant molecules over ancestral molecules ($0 \leq c \leq \infty$, $c = 1$ for no selection).

To allow comparability, we kept all parameters (other than selection) identical between drift and drift+selection simulations. Simulations only including selection on the individual resulted in poor approximations of empirical data (figs. 4C and 6; AIC ranged from 498.5 to 599.8). Interestingly, adding low levels of selection within the germ line of each individual for haplotype H ($c = 1.01$, i.e., 1% advantage) resulted in simulations which best matched the empirical data; this was true for all levels of selection against individuals carrying other haplotypes between $s = 0.1$ and $s = 0.9$ (empirical data: fig. 4A; simulated data: figs. 4D and 6; $c = 1.01$, AIC ranged from 186.6 to 188.5). This relationship disappeared rapidly when selection within individuals was increased further (fig. 6 and supplementary fig. S3, Supplementary Material online; AIC ranged from 240.3 to 756.6 when $c > 1.01$) and heteroplasmy became negligible.

Copy Number Analysis

To investigate the hypothesis that density of mitochondria or mtDNA molecules differed between haplotypes G and H, we compared the relative mtDNA copy number per cell in individuals carrying these haplotypes, using quantitative polymerase chain reaction (PCR) of three single-copy nuclear and three mitochondrial genes. For each individual, we analyzed the mean cycle threshold (Ct) value as a measure of abundance of mitochondrial genome copies, relative to nuclear genome copies. Individuals carrying the H haplotype had higher Ct values for mitochondrial genes than those carrying the G haplotype (difference in Ct value = 1.72 ± 0.56, t_{11} = 3.06, P = 0.01; fig. 7). The higher Ct value implied that there are fewer mtDNA copies per cell in the H haplotype individuals.

To investigate potential functional differences between these two haplotypes further, we used the software mfold (Zuker 2003; http://www.bioinfo.rpi.edu/applications/mfold, last accessed January 2, 2016) to predict secondary structure of RNA for each haplotype (supplementary fig. S4, Supplementary Material online). For haplotype G, mfold returned a single secondary structure (ΔG = −4.40). For haplotype H, two secondary structures were returned: one was the same as for haplotype G and the other had a similar free energy value (ΔG = −3.40).

Discussion

The directed changes in the proportion of adults carrying haplotype H (fig. 3) likely indicate selection. However, it can be difficult to distinguish selection from population admixture or from drift, in which proportions of selectively neutral haplotypes fluctuate unpredictably due to the randomness of transmission (Ballard and Kreitman 1995; Hallatschek et al. 2007). Admixture seems unlikely, because neither of its two predictions were satisfied: there was no association between microsatellite variation and mitochondrial haplotype (fig. 5); and there was no extreme (or even minor) fluctuation of nuclear allele proportions (Ryman 1997) (supplementary table S1, Supplementary Material online). Further, it appears that there are no likely sources of haplotype H for admixture, because this haplotype was found only in this invasion front population and not in other starling populations sampled across Australia (Rollins et al. 2011).

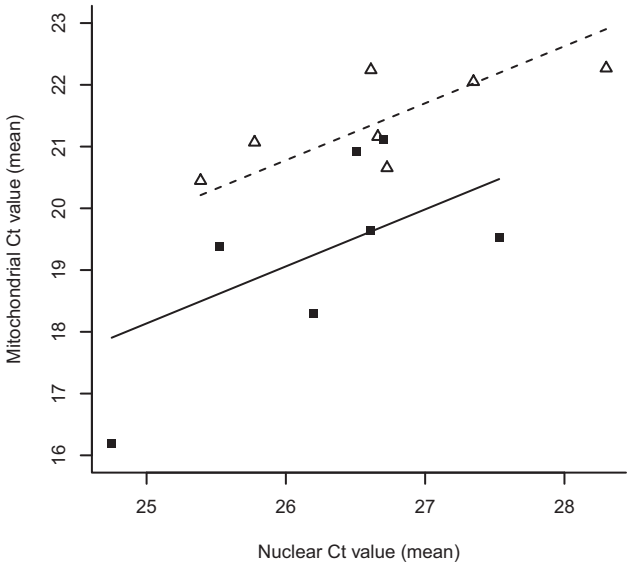

FIG. 7. Mitochondrial relative copy number, as cycle threshold (Ct). For each individual, Ct values are shown (averaged over three nuclear and three mitochondrial loci). Haplotype H individuals are denoted by open triangles (dashed line); haplotype G individuals are denoted by filled squares (solid line). ANCOVA indicated that individuals carrying the H haplotype had higher Ct values for mitochondrial genes (relative to nuclear genes) than those carrying the G haplotype (difference in Ct value = 1.72 ± 0.56, t_{11} = 3.06, P = 0.01). As expected, mean mitochondrial Ct values increased with mean nuclear Ct values (slope = 0.92 ± 0.32, t_{11} = 2.91, P = 0.01).

Random genetic drift is accelerated by low effective population size (Halliburton 2004), as found in range expansions (Austerlitz et al. 1997; Edmonds et al. 2004). However, drift cannot explain the data adequately, irrespective of whether the mutation to the H haplotype occurred after founders arrived at the expansion front or one of the founders arrived already carrying heteroplasmic DNA (empirical data: fig. 4; simulated data: figs. 4*B* and 6; supplementary fig. S2, Supplementary Material online). These simulations used extreme values for all of the input parameters, designed to favor rapid genetic drift. Addition of selection in favor of the new mutant showed much improved fit between simulation results and the empirical data.

Simulations only including selection due to competition between individuals resulted in poor approximations of empirical data (empirical data: fig. 4*A*; simulated data: figs. 4*C* and 6). The best fit occurred when there was selection at both levels of organization: Between-individuals and within-individuals (empirical data: fig. 4*A*; simulated data: figs. 4*D* and 6). Our results are supported by other findings that mitochondrial variants are not selectively neutral (Ballard and Whitlock 2004; Katewa and Ballard 2007).

The evidence that the patterns in our empirical data could be caused by selection, and the failure of other explanations such as drift and admixture, indicate that it is plausible that starling populations at the expansion front in Western Australia (WA) may be rapidly evolving in response to a novel environment. The increase in haplotype H proportions of 22% per generation (fig. 3) is consistent with strong selection (s = 0.66 against G relative to H, on the assumption that all selection was acting between individuals—eq. 2 in Materials and Methods). The rate of change identified here is not unprecedented for nuclear or mitochondrial genes. For mtDNA, MacRae and Anderson (1988) found changes of the same order of magnitude (but see Singh and Hale 1990; Kambhampati et al. 1992). For nuclear genes, it also appears that very strong selection can occur over brief intervals, especially in perturbed systems, such as during invasions. Endler (1986) reviewed 144 studies of perturbed populations in which selection coefficients were measured, 23% of which were equal to or higher than the value estimated in the current study. In a study of microevolutionary rates, Hendry and Kinnison (1999) concluded that although rapid evolution may occur, such intense selection in wild populations must

be limited to short time spans, in order to generate the lower rates of change observed over longer evolutionary time spans. This model of evolution is supported by empirical studies of Capricorn silvereye (*Zosterops lateralis*) demonstrating rapid evolution following colonization of a new environment, most likely due to directional selection followed by a stabilization of morphological traits (Clegg et al. 2002, 2008). In fact, in a review of 47 studies documenting rapid evolution, Reznick and Ghalambor (2001) found that the study populations fell into two categories: colonization of new environments or adaptation to heterogeneous environments.

Early research on heteroplasmy indicated that novel mitochondrial mutations often were rapidly fixed within individuals (Hauswirth and Laipis 1982), which has been attributed to the effects of drift during the mitochondrial bottleneck (Ashley et al. 1989). More recent evidence supports a role for selection (Fan et al. 2008) and modelling suggests that the direction of selection may be different between and within cells (Taylor et al. 2002). Both the relative strength of selection across these levels and the population size (on each level) can affect the speed of loss or fixation of a novel variant (Taylor et al. 2002). Further, heteroplasmy per se has been associated with reduced fitness in mice (Sharpley et al. 2012), indicating that there may be a fitness advantage to rapid fixation of a single haplotype, and selection against heteroplasmy may explain the evolution of uniparental inheritance of mitochondria (Christie et al. 2015). If selection against heteroplasmy exists, we should expect to see few heteroplasmic individuals and a rapid shift in allele frequencies within populations following a novel, advantageous mutation. In the present study, we are unable to assess whether and how quickly this novel mutation would have been driven to fixation, because our sampling only encompassed approximately 1.5 starling generations and the population has since been eradicated (this species is considered to be an agricultural pest in Western Australia). It is clear that heteroplasmy exists in some individuals, so we do know that selection against heteroplasmy is not strong enough to fix a haplotype within an individual in a single generation. Why might the within-individual selection be relatively weak? It is possible that our single modelled coefficient of selection (c) is actually composed of selection at several different levels within individuals (fig. 1); furthermore, it is possible that these selection effects might oppose one another (Taylor et al 2002), giving rise to the apparent low overall level of within-individual selection, and thus the currently widespread heteroplasmy.

mtDNA is normally uniparentally inherited in animals, but some exceptions exist (Barr et al. 2005). There is no evidence to support biparental inheritance in the individuals included in this study: double peaks were only seen at one polymorphic position in sequence electropherograms and only for the population studied here when other populations were examined (Rollins et al. 2011). Because its uniparental inheritance limits recombination, mtDNA may be susceptible to build up of deleterious mutations ("Muller's ratchet"; Howe and Denver 2008). However, the novel mutant that we found would be expected to vanish rapidly if it had low fitness, which is the opposite to the rapid increase we observed.

It is interesting to consider how the alteration to the H haplotype might cause a selective difference. Selection has only rarely been demonstrated to act on the mitochondrial control region (Moore et al. 2010; Ebner et al. 2011), whose functions are initiation of replication and transcription of the entire mitochondrial genome. However, point mutations in the control region have previously been linked to athletic performance (Murakami et al. 2002; Mikami et al. 2013). It is possible that the mitochondrial control region polymorphism described here has an effect on metabolic efficiency either directly or through complete linkage to polymorphisms elsewhere in the nonrecombining mitochondrial genome. However, the latter can be ruled out because in full genomes of 24 individuals carrying either G or H haplotypes (or both), there was no sequence variation other than the control region polymorphism that distinguishes these two haplotypes.

Thus any selection appears to be a direct result of the single base change in the control region, which might lead to altered numbers of mitochondria per cell. Altered numbers of mitochondria or mtDNA molecules could be due to either direct effects on replication, or because altered transcription changes the metabolic efficiency of the mitochondrion that carries the molecule, and the relative metabolic efficiency of mitochondria with particular haplotypes can secondarily affect their relative rates of replication (Wai et al. 2008; Hill et al. 2014), possibly due to interactions with the nuclear genome. There could also be a between-individual effect: for example, an altered density of mitochondria in cells of haplotype H starlings may alter flight and dispersal ability.

Indeed, it does appear that there is a difference in mtDNA copy-number per cell in individuals of the two different haplotypes. The higher Ct value of H haplotype individuals implied that there are fewer mtDNA copies per cell in those individuals than in G haplotype individuals (fig. 7). It is remarkable that a single base-pair change in the mitochondrial control region could have a physiological outcome, and also that a mutant which is associated with decreased copy number appears to be at a selective advantage. mtDNA replication and transcription are linked; transcribed 7S RNAs serve as primers for DNA replication (Clayton 2000; Jemt et al. 2015). The D-Loop is subsequently formed and is situated between the conserved sequence block (CSB1) and a termination-associated sequence (TAS) near the 3′-end of the control region heavy strand (Clayton 2000). A sequence similar to CSB1 (CSB1-like) has been identified in a number of avian species (Zou et al. 2015), which lack other CSB motifs (CSBII and CSBIII) (Shadel and Clayton 1997). The polymorphism we have identified is located near the CSB1-like sequence motif (fig. 2) and may result in altered RNA secondary structure (supplementary fig. S4, Supplementary Material online). The role of the CSB1-like sequence motif, if any, is unclear but because other CSB motifs are involved in mtDNA replication (Shadel and Clayton 1997), it is possible that the proximity of the polymorphism to the CSB1-like motif may be responsible for the differences in replication we identified.

Interactions between the mitochondrial and nuclear genomes (mitonuclear interactions) may also affect replication efficiency. The nuclear genome codes for proteins required for mtDNA transcription and replication (Shadel and Clayton 1997). Mitonuclear interactions are well-characterized (reviewed in Dowling et al. 2008) and effects of these interactions on mtDNA copy number have been demonstrated (Ellison and Burton 2010). Therefore, it is possible that mitonuclear interactions have contributed to the phenotypic difference between the G and H haplotypes identified here.

Other than investigating potential mitonuclear interactions, several future directions may be useful. Experiments involving starlings carrying haplotypes G and H could enable the investigation of fitness at both the individual and mitochondrial levels. In this context, reproductive success could be measured to determine whether carrying haplotype H has a fitness advantage. Studying mitochondrial function (e.g., respiratory capacity) across haplotypes may support the presence of selection acting on this locus. Unfortunately, the latter requires fresh tissue and those used in the present study were frozen. This is further complicated because the population where this study was focused has been extirpated as a result of control efforts, though perhaps our findings will stimulate investigations of these effects in other species.

Here we have shown that patterns in our empirical data cannot be explained by random drift, nor by selection that is solely between individuals—there must also be selection at the within-individual level between molecules, mitochondria or cells with the H haplotype versus other haplotypes. The possibility of within-individual selection due to mitochondrial control region variation may require a substantial reconsideration of the general understanding of the evolution of this molecule. Our work is a rare, detailed example of selection on a newly arisen mtDNA haplotype, in a perturbed situation (recent invasion). Rapid adaptation during successful invasion has been postulated, but with little evidence (Lee 2002). In the future, temporal sampling of invasive populations will help to assess rapid evolution in novel habitats. This would also lead to an improved understanding of adaptation of native species, and will improve our ability to anticipate and manage invasions and understand adaptation to other changes such as climate change. Finally, these conclusions provide a novel and valuable contribution to our understanding of the evolution of the mtDNA, which is a central focus of studies of human mitochondrial disease.

Materials and Methods

Sampling was conducted annually from 2003 to 2007 using starlings captured by invasive species management agencies for destruction. Adults (181) and juveniles (98) were sampled from Munglinup (Western Australia), on the front of the starling invasion (annual samples sizes listed in fig. 3 and supplementary fig. S5, Supplementary Material online). This population has recently been extirpated. All work described here was approved by the University of New South Wales Animal Ethics Committee (Project #05/11A).

Sequencing

DNA was extracted from liver tissue using a PureGene Tissue Extraction Kit (Qiagen) and a 942-bp segment of the mitochondrial control region was Sanger sequenced and analyzed following Rollins et al. (2011). Haplotype proportions were calculated across years and a multinomial regression of haplotype proportion against time was conducted with juveniles included (supplementary fig. S5, Supplementary Material online) and excluded (fig. 3). If particular haplotypes enhanced survival, the effect would be far more evident in the adult portion of the population, which had survived the approximately 70% mortality (Coulson 1960) in their first winter. This appeared to be the case, so only results for adults are discussed from here forward.

As previously reported (Rollins et al. 2011), haplotypes H and G differed at a single nucleotide (T/C respectively) and individuals carrying either haplotype G or H were sometimes heteroplasmic for these two haplotypes. We assessed the proportions of homoplasmic individuals and the proportion of different haplotypes within each heteroplasmic individual. To do this, we measured peak heights of thymine and cytosine at the sequence position distinguishing haplotypes H and G, respectively, using electropherograms produced from Sanger sequencing (supplementary fig. S6, Supplementary Material online). For each individual shown in figure 3, the empirical proportion of heteroplasmy was generated using the relative height of the mutant peak. Mutant occurrence was calculated as:

$$\text{proportion of } H = A_T / (A_T + A_C), \quad (1)$$

where A is the amplitude of the peak for thymine or cytosine. For plotting, this was converted to numbers of mutant mitochondria by multiplying by the value taken to be the total at the mitochondrial bottleneck (below, $mz = 200$). We validated this method of estimating heteroplasmy levels by next-generation sequencing the mitochondrial genome of a subset of 24 individuals having various levels of heteroplasmy (details below). The next-generation sequencing allowed us to make an independent estimate of proportions of heteroplasmy, by counting the number of times each haplotype sequence was present for each individual (Tang and Huang 2010).

For calculation of between-individual selection coefficients, electropherograms of individuals carrying either G or H haplotype were scrutinized at the diagnostic nucleotide position and heteroplasmic individuals were characterized as having the haplotype represented by the highest peak (supplementary fig. S6, Supplementary Material online). For haplotypes whose frequency appeared to be changing (i.e., G and H; fig. 3), we calculated a coefficient of apparent selection between individuals (s—selection against the disadvantageous haplotype G relative to H), using equation (2) (from http://www.apsnet.org/edcenter/advanced/topics/PopGenetics/Pages/NaturalSelection.aspx, last accessed January 2, 2016) (The American Phytopathological Society 2014):

$$s = \frac{(q_1 - q_0)}{q_1(1 - q_0)}, \quad (2)$$

where q_0 is the starting proportion of haplotype H and q_1 is the proportion of haplotype H after one starling generation of selection (estimated at 2.8 years, see below). Haplotype proportions q_0 and q_1 were estimated using a linear regression of empirical estimates of haplotype proportions of adults against time, calculated in MICROSOFT EXCEL 2003.

Next-Generation Sequencing

We used next-generation sequencing to validate the method used to create the empirical heteroplasmy distribution (fig. 4A), which was based on ratio of peak heights of thymine and cytosine seen in electropherograms from Sanger sequencing. Paired-end sequencing was conducted on 24 samples that were individually barcoded, using Nextera XT library preparation kits and the Illumina MiSeq platform. Data were analyzed using CLC Genomics Workbench v 7.0.3. Demultiplexed Illumina files were imported and trimmed to include only sequences greater than 30 bp, having Phred scores greater than 30. For each sample, we recorded the number of MiSeq reads having a thymine versus a cytosine in the sequence position defining haplotypes H and G. These results were compared with peak height data measured using ImageJ v 1.47v (http://imagej.nih.gov/ij/, last accessed January 2, 2016) at the same position in Sanger sequencing electropherograms for each of these 24 individuals.

The next-generation sequence data also allowed us to determine whether the G and H haplotypes might be linked to other polymorphisms of the mitochondrial genome, which could potentially be responsible for the results we observed. Using CLC Genomics Workbench, overlapping sequence pairs were merged and de novo assemblies were created using default parameters for each of a subset of eight samples (including those that were haplotypes G, H, and G/H). The longest contig was chosen from each of these eight assemblies (~16 kb in length) and these contigs were aligned to create a reference mitochondrial genome. A BLAST (Basic Local Alignment Search Tool) search of this reference sequence aligned with 91% similarity to three congeners (*S. cineraceus*, HQ896037.1; *S. sericeus*, HM849900.1; *S. nigricollis*; JQ003191.1). Each of the 24 data sets was mapped back to the reference and consensus sequences were extracted from each data set and aligned. Full mitochondrial genome sequences for haplotypes G and H were constructed (GenBank accession number: KT946691, KT946692). No polymorphism was identified other than the single base change that distinguishes haplotypes G and H.

Admixture Analysis

If H haplotype individuals were more recent immigrants to the population, they should also show differentiation at nuclear genes. To investigate this, a suite of 11 nuclear microsatellites was amplified following Rollins et al. (2009) using individuals shown in figure 3, barring five individuals that were excluded from this analysis due to amplification failure. STRUCTURE v 2.3 (Pritchard et al. 2000; Falush et al. 2003) was used to determine the number of nuclear genetic groups present following Evanno et al. (2005) and using simulation conditions described in Rollins et al. (2009). Individuals were also separated into two groups depending on whether they carried Haplotype H or did not. We compared the mean nuclear ancestry proportions for individuals from each of these two mitochondrial groups (Z test).

The microsatellite data set was also used to test for nuclear genetic differentiation between years (measured as pairwise F_{ST} values, adjusted using Benjamini and Hochberg's [1995] false discovery rate correction) because simulations have shown that admixture of novel nuclear genotypes may result in extreme nuclear allele proportions fluctuations for several generations, in species having overlapping generations (Ryman 1997).

Genetic Drift Simulations

Simulations were conducted in MATLAB v R2013a (code available at http://hdl.handle.net/10536/DRO/DU:30079101, last accessed January 2, 2016) to determine whether the empirical haplotype distribution could be explained by genetic drift acting on individuals in the population and/or within individuals, assuming no dispersal and no selection. For most parameters used in these simulations, estimates in the literature span a range of values, but most have not been estimated in avian systems. Below, for each parameter, we have chosen values from the end of the range that would maximize the chance that drift could explain our results without any selection.

Simulations tracked the fate of each individual in each generation. Populations were capped at 1,000 individuals to reduce computation time; it is unlikely that the empirical population in our study exceeded this value. We calculated the number of offspring for each individual from a Poisson distribution with a mean of the population growth rate (λ). Cabe's (1998) estimate of population growth rate was based on expanding introduced starling populations in North America and data from Kessel (1957), and ranged from λ = 1.1 to 1.628. We used the upper bound of this estimate in our simulations. Each simulation was run over an estimated number of animal generations since introduction. Generation time is the average age of females at time of birth of their offspring; a 22-year study of starlings in their native range (Feare and Forrester 2002) gave an estimate of 2.8 years. Thus, the number of generations between colonization and the end of our sampling period (2007) was either two (since first reports of starlings at our study site in 2000–2001) or 12 (since the original incursion of starlings in WA in 1971). Unless otherwise stated, we used 12 generations, a generous estimate that maximizes the chance that drift could explain the change of haplotype proportions. We assumed the founding number of individuals to be 5, representing four females carrying each of the four haplotypes found in our study site, plus a single male. This represents the minimum possible number of founders, which also maximizes the chance that drift could explain our empirical results.

The number of mitochondria per gonad cell (m = 1,500) and the size of the mitochondrial bottleneck in oocytes (mz = 200) have been debated (Carling et al. 2011; Wallace and

Chalkia 2013). Our estimate (mz = 200) was taken from estimates over the first 14 days postfertilization in mice (Jenuth et al. 1996), confirmed more recently using quantitative PCR and simulations by Cree et al. (2008) and supported in other taxa using population genetic theory (Hauswirth and Laipis 1982; Rand and Harrison 1986; Jenuth et al. 1996).

For the initial generation of individuals, we investigated seven scenarios for the number of mutant mitochondria (hz) present in oocytes of any heteroplasmic individuals. First, we simulated a new mutation arising in the study population the oocyte of a single founder (i.e., number of mutant mitochondria hz = 1 out of the total number mz = 200 mitochondria in the mitochondrial bottleneck). We also simulated the fate of a pre-existing mutation H, occurring in all cells of a single heteroplasmic haplotype H carrier in the founding group for this population, whose hz was equal to 25, 50, 75, 100, 150, or 199 out of a possible 200 (mz) in the oocyte. For birds, there are no available estimates of the number of female germ cell generations per animal generation (ga). Estimates are available for insects and mammals, the highest of which is 50 (Ashley et al. 1989). We have used a value of ga = 60 to maximize the chance that drift could explain the change of haplotype proportions.

The number of offspring per individual was sampled from a Poisson distribution, with mean λ. To speed up the process of simulating every cell generation in every heteroplasmic individual, the standard deviation of the proportion of mutant mitochondria per cell in an offspring after ga cell generations was calculated by equation (3) modified from Wright (1942):

$$\sigma_h = \sqrt{ph * (1 - ph) * (1 - (1 - 1/m)^g)}, \quad (3)$$

where $g = ga - 1$ is the number of generations during the transmission between generations of cells destined to become oocytes, each with number of mitochondria m = 1,500; and $ph = hz/mz$ = proportion of mutant type H mitochondria present at the start of the mitochondrial bottleneck (i.e., in transmission to the ovum). For each offspring, the actual proportion of mutant mitochondria per cell after ga cell generations was then calculated by sampling from a normal distribution with mean ph and standard deviation σ_h; values outside the range 0:1 were truncated to zero or unity. This proportion was then used to stochastically determine the number of mutant mitochondria, out of the mz = 200 present at the final generation (the mitochondrial bottleneck in the oocyte). As for other choices in the simulation process, equation (3) and the truncation would somewhat exaggerate the effects of drift.

Individual iterations ended when either 12 generations were completed, or earlier if the mutation was lost to drift (i.e., 0% mutant mitochondria) or became fixed (i.e., 100% mutant mitochondria) in the population. When the final generation was reached, we recorded the distribution of heteroplasmy in the population during the bottleneck and the level of heteroplasmy (percentage of heteroplasmic individuals, P_{lh}). In our analyses, we only included iterations that ended with at least as many individuals as were "known to be alive" (KTBA = 136) in the empirical population (KTBA = 136). KTBA was estimated using the following formula:

$$\text{KTBA} = B \times \frac{G}{Y}, \quad (4)$$

where B is the total number of different adults sampled; G, generation time (2.8 years, see above); and Y, the number of years over which B was calculated. Simulation results stabilized after 1,000 iterations (data not shown). To be conservative, for each scenario we analyzed 3,000 iterations fitting the criterion: total number of individuals between 136 and 1,000 at the final generation.

We further examined the importance of drift, by investigating the possibility that there were less than 12 generations of individuals since founding of the population, which would decrease the chance that drift could explain the empirical results. Trials with shorter time since founding produced a poorer fit between empirical and simulated results (data not shown) so are not discussed further.

Selection Simulations

We investigated the effects of selection between- and within-individuals. First, we modified the genetic drift simulation to include a relative disadvantage (s) to offspring with only ancestral mitochondria versus those with some mutant mitochondria. Each ancestral individual had a chance s ($0 \leq s \leq 1$, s = 0 for no selection) that it would be excluded from the array of offspring from which the next generation of breeders was chosen at random. Next, to apply selection within individuals, the proportion of mutant mitochondria per cell after the ga cell generations was increased by "c^g", where c is the relative advantage of mutant over ancestral haplotypes ($0 \leq c \leq \infty$, c = 1 for no selection). Note that this selective advantage could relate to competition between molecules within mitochondria, or mitochondria within cells, or cells within the germ line—the model does not discriminate between these possibilities.

We ran simulations for each selection scenario using a checkerboard matrix with between-individual selection (s) values from 0 to 1 in 0.1 (10%) increments, and value of within-individual selection (c) from 1 to 1.04 in 0.01 (1%) increments. All other parameter values were taken from the genetic drift simulations and we used hz = 1 (a new mutation arising in the study population in a single founder). Where all iterations of a particular selection scenario reached fixation for a single haplotype (i.e., no heteroplasmy; c = 1.04), we truncated the matrix by not investigating more extreme combinations of s and c.

Analysis of Simulation Data

We tested whether simulation scenarios fitted the empirical data, based on proportions of homoplasmic and heteroplasmic individuals in the population, plus the proportion of different haplotypes within each heteroplasmic individual. We used the AIC to assess the fit between empirical and simulation data. For each iteration, simulation data were binned into one of 42 bins of number of mutant mitochondria (0, 1–5,

6–10, . . . , 196–199, 200) and the proportion of individuals in each bin was calculated. We then calculated the average proportion for each bin across the 3,000 iterations for each simulation scenario. Using these proportion estimates, we then calculated the likelihood of obtaining the empirical data, assuming a multinomial distribution. Likelihoods were converted into an AIC estimate (Burnham and Anderson 2002). To visualize patterns in AIC across the *s* and *c* grid, we performed a thin-plate spline (fitted using generalized cross validation) with the "Tps()" function in the "fields" package in R (V3.1).

These comparisons assume that trends over generations in relative abundance of G and H in the sampled liver tissue would follow trends in the oocytes and the ova (which we simulated). We know of no cases where these trends might be sufficiently different to invalidate our approach.

Copy Number Analysis

We wished to identify whether the G and H haplotypes had different copy numbers that might correlate with any selective differences. Because it is not possible to know the number of cells present in each sample, using an absolute measure of mitochondrial genome copies would be inaccurate. Nuclear single-copy genes should be present once in each cell and thus their abundance can be used as a baseline against which to estimate relative mitochondrial genome copy number. As any single-copy gene can be used for this purpose, we chose genes that had published sequences and were putatively single-copy genes (see supplementary table S2, Supplementary Material online for details). Using Primer 3 (Rozen and Skaletsky 2000), we designed primers for three nuclear genes (*C-MYC*, *RDP1*, and *ZENK*) and three mitochondrial loci equally spaced around the mitochondrial genome (*ATP6*, *CR*, and *ND2*) using the full starling mitochondrial genome sequence produced here. These primers were used to quantify relative mitochondrial copy number in seven individuals carrying haplotype H and seven individuals carrying haplotype G. Quantitative PCR was conducted on a Bio-Rad CFX96 instrument using SYBR green (Bio-Rad Laboratories, USA). A master mix of each primer pair contained 1× SYBR green, 400 nM of each primer, and nuclease free water. Reactions were carried out in 96-well reaction plates (Bio-Rad Laboratories). Each well contained 23 µl of master mix and 2 µl of DNA template (5 ng approximately). Template-free controls were included for all primer pair combinations (nuclease free water added in place of DNA). The cycling parameters were performed as follows: 95 °C for 3 min, 40 cycles of 95 °C 10 s, 60 °C for 30 s, and 72 °C for 30 s. Raw cycle threshold (Ct) values were used for analysis; these values are inversely proportional to the amount of nucleic acid in the sample. The quantitative PCR reaction was directly followed by melt curve analysis (95 °C for 1 min and melt curve analysis 60–95 °C with increments of 0.5 °C for 30 s) to ensure that a single product was being amplified, thus ensuring specificity.

We performed an analysis of covariance (ANCOVA) in which haplotype was the fixed factor, mean raw Ct value of nuclear genes was the covariate and mean raw Ct value of mitochondrial genes was the response variable. To test for difference in slopes, we initially included the interaction between haplotype and mean Ct value of the nuclear genes, but removed the interaction from the final model because it was not significant ($P = 0.28$). The final model met normality and homoscedasticity. We ensured that Ct values across genes within genomes were consistent (correlation analyses between the Ct value of each gene and the mean Ct value for the genome; correlation coefficients ranged from 0.93 to 1.0).

We further investigated potential functional differences between these haplotypes using mfold (Zuker 2003) to predict secondary structure of RNA. For each haplotype, we included 50 bp on each side of the polymorphism distinguishing haplotypes H and G. Other than choosing the "circular RNA" option, all other values were left as defaults. Free energy values (ΔG) for each potential RNA secondary structure were compared.

Data Availability

The genome sequence is available (GenBank accession number: KT946691, KT946692) and raw Illumina Mi-Seq read data are available under BioProject PRJNA214891. Microsatellite data and simulation code are available at http://hdl.handle.net/10536/DRO/DU:30079101 (DOI: 10.4225/16/56298CC03A59F).

Supplementary Material

Supplementary tables S1 and S2 and supplementary figures S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Bill Ballard for useful discussions about this manuscript and suggestions regarding copy number analysis; Rob Brooks, Kate Buchanan, Damian Dowling, John Endler, Simon Griffith, and Rick Shine for comments that improved this manuscript; Peri Bolton (http://peribolton.weebly.com/) for the starling illustration; Mark Richardson for assistance with accessioning sequence data; assistance from the Department of Agriculture and Food Western Australia (DAFWA) with sample collections; the Ramaciotti Centre for Genomics for assistance with Sanger and Illumina sequencing and microsatellite genotyping. This work was funded by the Australian Research Council (LP#0455776 with contributions from DAFWA and Biosecurity SA, Department of Primary Industries and Regions South Australia); and Deakin University for fellowship support for L.A.R. and B.G.F.

References

Arnaud-Haond S, Vonau V, Bonhomme F, Boudry P, Blanc F, Prou J, Seaman T, Goyard E. 2004. Spatio-temporal variation in the genetic composition of wild populations of pearl oyster (*Pinctada margaritifera cumingii*) in French Polynesia following 10 years of juvenile translocation. *Mol Ecol.* 13:2001–2007.

Ashley MV, Laipis PJ, Hauswirth WW. 1989. Rapid segregation of heteroplasmic bovine mitochondria. *Nucleic Acids Res.* 17:7325–7331.
Austerlitz F, Jung-Muller B, Godelle B, Gouyon P-H. 1997. Evolution of coalescence times, genetic diversity and structure during colonization. *Theor Popul Biol.* 51:148–164.
Ballard JWO, Kreitman M. 1995. Is mitochondrial DNA a strictly neutral marker? *Trends Ecol Evol.* 10:485–488.
Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol Ecol.* 13:729–744.
Barr CM, Neiman M, Taylor DR. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytol.* 168:39–50.
Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol.* 57:289–300.
Birky CW Jr, Maruyama T, Fuerst P. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* 103:513–527.
Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer Science & Business Media.
Cabe PR. 1998. The effects of founding bottlenecks on genetic variation in the European starling (*Sturnus vulgaris*) in North America. *Heredity* 80:519–525.
Carling PJ, Cree LM, Chinnery PF. 2011. The implications of mitochondrial DNA copy number regulation during embryogenesis. *Mitochondrion* 11:686–692.
Castoe TA, Jiang ZJ, Gu W, Wang ZO, Pollock DD. 2008. Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS One* 3:e2201.
Chesser RK. 1998. Heteroplasmy and organelle gene dynamics. *Genetics* 150:1309–1327.
Christie JR, Schaerf TM, Beekman M. 2015. Selection against heteroplasmy explains the evolution of uniparental inheritance of mitochondria. PLoS Genet. 11(4): e1005112.
Clayton DA. 2000. Transcription and replication of mitochondrial DNA. *Hum Reprod.* 15:11–17.
Clegg SM, Degnan SM, Moritz C, et al. 2002. Microevolution in island forms: the roles of drift and directional selection in morphological divergence of a passerine bird. *Evolution* 56:2090–2099.
Clegg SM, Frentiu FD, Kikkawa J, Tavecchia G, Owens IPF. 2008. 4000 years of phenotypic change in an island bird: heterogeneity of selection over three microevolutionary timescales. *Evolution* 62:2393–2410.
Coulson JC. 1960. A study of the mortality of the starling based on ringing recoveries. *J Anim Ecol.* 29:251–271.
Cree LM, Samuels DC, de Sousa Lopes SC, et al. 2008. A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nat Genet.* 40:249–254.
Dowling DK, Friberg U, Lindell J. 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol Evol.* 23:546–554.
Ebner S, Lang R, Mueller EE, et al. 2011. Mitochondrial haplogroups, control region polymorphisms and malignant melanoma: a study in middle European Caucasians. *PLoS One* 6:e27192.
Edmonds CA, Lillie AS, Cavalli-Sforza LL. 2004. Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci U S A.* 101:975–979.
Ellison C, Burton R. 2010. Cytonuclear conflict in interpopulation hybrids: the role of RNA polymerase in mtDNA transcription and replication. *J Evol Biol.* 23:528–538.
Endler JA. 1986. Natural selection in the wild. Princeton (NJ): Princeton University Press.
Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol.* 14:2611–2620.
Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.
Fan W, Waymire KG, Narula N, et al. 2008. A mouse model of mitochondrial disease reveals germline selection against severe mtDNA mutations. *Science* 319:958–962.
Feare CJ, Forrester GJ. 2002. The dynamics of a suburban nestbox breeding colony of starlings *Sturnus vulgaris*. In: Crick HPQ, Robinson RA, Appleton GF, Clark NA, Rickard AD, editors. Investigations into the causes of the decline of starlings and house sparrows in Great Britain. Bristol: DEFRA. p. 73–90.
Finch TM, Zhao N, Korkin D, Frederick KH, Eggert LS. 2014. Evidence of positive selection in mitochondrial complexes I and V of the African elephant. *PLoS One* 9:e92587.
Grossman LI, Wildman DE, Schmidt TR, Goodman M. 2004. Accelerated evolution of the electron transport chain in anthropoid primates. *Trends Genet.* 20:578–585.
Hallatschek O, Hersen P, Ramanathan S, Nelson DR. 2007. Genetic drift at expanding frontiers promotes gene segregation. *Proc Natl Acad Sci U S A.* 104:19926–19930.
Halliburton R. 2004. Introduction to population genetics. Upper Saddle River (NJ): Pearson Education, Inc.
Hauswirth WW, Laipis PJ. 1982. Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. *Proc Natl Acad Sci U S A.* 79:4686–4690.
Hendry AP, Kinnison MT. 1999. Perspective: The pace of modern life: measuring rates of contemporary microevolution. *Evolution* 53:1637–1653.
Hill JH, Chen Z, Xu H. 2014. Selective propagation of functional mitochondrial DNA during oogenesis restricts the transmission of a deleterious mitochondrial variant. *Nat Genet.* 46:389–392.
Howe DK, Denver DR. 2008. Muller's Ratchet and compensatory mutation in *Caenorhabditis briggsae* mitochondrial genome evolution. *BMC Evol Biol.* 8:62.
Jacobsen F, Nesje M, Bachmann L, Lifjeld J. 2008. Significant genetic admixture after reintroduction of peregrine falcon (*Falco peregrinus*) in Southern Scandinavia. *Conserv Genet.* 9:581–591.
Jemt E, Persson Ö, Shi Y, et al. 2015. Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element. *Nucleic Acids Res.* doi: 10.1093/nar/gkv804.
Jenuth JP, Peterson AC, Fu K, Shoubridge EA. 1996. Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nat Genet.* 14:146–151.
Kambhampati S, Rai KS, Verleye DM. 1992. Frequencies of mitochondrial-DNA haplotypes in laboratory cage populations of the mosquito, *Aedes-albopicus*. *Genetics* 132:205–209.
Katewa SD, Ballard JWO. 2007. Sympatric *Drosophila simulans* flies with distinct mtDNA show difference in mitochondrial respiration and electron transport. *Insect Biochem Mol Biol.* 37:213–222.
Kessel B. 1957. A study of the breeding biology of the European starling (*Sturnus vulgaris* L.) in North America. *Am Midl Nat.* 58:257–331.
Klopfstein S, Currat M, Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol.* 23:482–490.
Kolbe JJ, Larson A, Losos JB, de Queiroz K. 2008. Admixture determines genetic diversity and population differentiation in the biological invasion of a lizard species. *Biol Lett.* 4:434–437.
Lee CE. 2002. Evolutionary genetics of invasive species. *Trends Ecol Evol.* 17:386–391.
MacRae AF, Anderson WW. 1988. Evidence for non-neutrality of mitochondrial DNA haplotypes in *Drosophila pseudoobscura*. *Genetics* 120:485–494.
Mikami E, Fuku N, Takahashi H, et al. 2013. Polymorphisms in the control region of mitochondrial DNA associated with elite Japanese athlete status. *Scand J Med Sci Sports.* 23:593–599.
Moore AZ, Biggs ML, Matteini A, et al. 2010. Polymorphisms in the mitochondrial DNA control region and frailty in older adults. *PLoS One* 5:e11069.
Murakami H, Ota A, Simojo H, et al. 2002. Polymorphisms in control region of mtDNA relates to individual differences in endurance capacity or trainability. *Jpn J Physiol.* 52:247–256.

Phillips BL, Brown GP, Shine R. 2010. Evolutionarily accelerated invasions: the rate of dispersal evolves upwards during the range advance of cane toads. *J Evol Biol.* 23:2595–2601.

Phillips WS, Coleman-Hulbert AL, Weiss ES, et al. 2015. Selfish mitochondrial DNA proliferates and diversifies in small, but not large, experimental populations of *Caenorhabditis briggsae*. *Genome Biol Evol.* 7:2023–2037.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

Rand DM. 1996. Neutrality tests of molecular markers and the connection between DNA polymorphism, demography, and conservation biology. *Conserv Biol.* 10:665–671.

Rand DM, Harrison RG. 1986. Mitochondrial DNA transmission genetics in crickets. *Genetics* 114:955–970.

Reznick DN, Ghalambor CK. 2001. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* 112–113:183–198.

Rollins L, Woolnough AP, Sinclair RG, Mooney NJ, Sherwin WB. 2011. Mitochondrial DNA offers unique insights into invasion history of the common starling. *Mol Ecol.* 20:2307–2317.

Rollins LA, Moles AT, Lam S, et al. 2013. High genetic diversity is not essential for successful introduction. *Ecol Evol.* 3:4501–4517.

Rollins LA, Woolnough AP, Wilton AN, Sinclair RG, Sherwin WB. 2009. Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. *Mol Ecol.* 18:1560–1573.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: methods in molecular biology. Totowa (NJ): Humana Press. p. 365–386.

Ryman N. 1997. Minimizing adverse effects of fish culture: understanding the genetics of populations with overlapping generations. *ICES J Mar Sci.* 54:1149–1159.

Shadel GS, Clayton DA. 1997. Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem.* 66:409–435.

Sharpley MS, Marciniak C, Eckel-Mahan K, et al. 2012. Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition. *Cell* 151:333–343.

Shen Y-Y, Liang L, Zhu Z-H, et al. 2010. Adaptive evolution of energy metabolism genes and the origin of flight in bats. *Proc Natl Acad Sci U S A.* 107:8666–8671.

Shoubridge EA, Wai T. 2008. Sidestepping mutational meltdown. *Science* 319:914–915.

Singh RS, Hale LR. 1990. Are mitochondrial DNA variants selectively non-neutral? *Genetics* 124:995–997.

Tang S, Huang T. 2010. Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system. *BioTechniques* 48:287–296.

Taylor DR, Zeyl C, Cooke E. 2002. Conflicting levels of selection in the accumulation of mitochondrial defects in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A.* 99:3690–3694.

The American Phytopathological Society. 2014. Natural selection in plant pathosystems. Available from: http://www.apsnet.org/edcenter/advanced/topics/PopGenetics/Pages/NaturalSelection.aspx.

Thrailkill KM, Birky CW Jr, Luckemann G, Wolf K. 1980. Intracellular population genetics: evidence for random genetic drift of mitochondrial allele frequencies in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Genetics* 96:237–262.

Wai T, Teoli D, Shoubridge EA. 2008. The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. *Nat Genet.* 40:1484–1488.

Wallace DC, Chalkia D. 2013. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harb Perspect Biol.* 5: a021220.

Wonnapinij P, Chinnery PF, Samuels DC. 2008. The distribution of mitochondrial DNA heteroplasmy due to random genetic drift. *Am J Hum Genet.* 83:582–593.

Wright S. 1942. Statistical genetics and evolution. *Bull Am Math Soc.* 48:223–246.

Zou Y, Jing M-D, Bi X-X, Zhang T, Huang L. 2015. The complete mitochondrial genome sequence of the little egret (*Egretta garzetta*). *Genet Mol Biol.* 38:162–172.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406–3415.