Increased mitochondrial mutation frequency after an island colonization: positive selection or accumulation of slightly deleterious mutations?

Emilie A. Hardouin† and Diethard Tautz

Max Planck Institute for Evolutionary Biology, Ploëhn 24306, Germany

Island colonizations are excellent models for studying early processes of evolution. We found in a previous study on mice that had colonized the sub-Antarctic Kerguelen Archipelago about 200 years ago that they were derived from a single founder lineage and that this showed an unexpectedly large number of new mutations in the mitochondrial D-loop. To assess whether positive selection has played a role in the emergence of these variants, we have obtained 16 full mitochondrial genome sequences from these mice. For comparison, we have compiled 57 mitochondrial genome sequences from laboratory inbred lines that became established about 100 years ago, also starting from a single founder lineage. We find that the island mice and the laboratory lines show very similar mutation frequencies and patterns. None of the patterns in the Kerguelen mice provides evidence for positive selection. We conclude that nearly neutral evolutionary processes that assume the presence of slightly deleterious variants can fully explain the patterns. This supports the notion of time-dependency of molecular evolution and provides a new calibration point. Based on the observed mutation frequency, we calculate an average evolutionary rate of 0.23 substitutions per site per Myr for the earliest time frame of divergence, which is about six times higher than the long-term rate of 0.037 substitutions per site per Myr.

1. Introduction

Mitochondrial sequence comparisons are broadly used for reconstructing phylogeographic patterns and for dating splitting events between species and populations [1]. However, it has been suggested that there is a time dependence of substitution rates that can influence the estimation of the age of splitting events [2]. More recently separated populations show generally higher rates, while rates between long separated populations (greater than 1 Myr) converge to a lower level. The shape of the transition from high rates to constant rates is insufficiently understood so far, and it has therefore been proposed that more data from recently split populations with good calibration will be required to explore this phenomenon further [2].

Another unresolved question concerns the impact of positive selection on mitochondrial diversity. Bazin et al. [3] inferred from a large comparative analysis of mitochondrial sequences that positive selection is expected to occur frequently. Positive selection has indeed been observed in very recently differentiated populations, in particular for ND2 and ND6 in the context of high-altitude adaptations [4–6]. In a previous study, we found that mouse (Mus musculus domesticus) populations that had colonized sub-Antarctic islands less than 200 years ago showed a high number of substitutions in mitochondrial D-loop sequences [7]. This raised the question of whether positive
selection could explain this pattern, because the mice on sub-Antarctic islands may have had to adapt to a feral lifestyle in a colder climate with new food resources. To test this, we explore the possibility here that the D-loop variants correlate with mutations elsewhere in the mitochondrial genome that may have been fixed by positive selection. Full mitochondrial genome sequencing from these animals was used to assess possible patterns of selection and to address the rate question.

2. Material and methods
Mitochondrial genomes were sequenced from 16 mice using primers described in Stewart et al. [8] and in electronic supplementary material, table S1 (sequences are deposited under GenBank accession nos JX945964 to JX945979). Mice were collected as described in Chapuis et al. [9] (locations are depicted in electronic supplementary material, figure S1). DNA extractions and sequencing reactions were performed as described in Hardouin et al. [7]. Mitochondrial genomes from inbred laboratory lines were retrieved from GenBank (accession nos in electronic supplementary material, table S2). Since both the island mice and the laboratory mice are derived from a single, albeit different, mitochondrial haplotype [7,10], we estimated mutation frequencies with respect to the total number of nucleotides sequenced by counting the mutations in comparison to the consensus sequence, in a procedure analogous to the estimation of mutation frequencies from mutation–accumulation lines [8].

3. Results
Sixteen full mitochondrial genomes from the mice representing the first invasion wave on the Kerguelen Archipelago (Grande Terre and associated islands—electronic supplementary material, figure S1) [7] were sequenced. All 13 previously identified animals that had shown one or more new mutations in the D-loop were included, plus three representing the sequence of the invasion haplotype in the D-loop region. We found a total of seven additional single-step point mutations, four in coding regions leading to amino acid changes, two synonymous ones and one in tRNAVal (table 1). In addition, tRNAsP shows a variable number of As within a stretch of As in two animals, representing a known polymorphic region [8]. The ND6 mutation in one individual showed signs of heteroplasmy (table 1) with about 40 per cent of the peak size representing the original C at this position. In addition, the A-stretch in tRNAsP in the same animal could not be clearly sequenced, also suggesting the presence of at least two variants.

Mutations in the full mitochondrial genome sequences occurred at the same frequency as in our previous survey of D-loop sequences [7]. In that survey, we found eight point mutations in 258 540 nt sequenced (310 animals \( \times 834 \) nt). In the present study, the seven mutations found outside of the D-loop were from a total of 247 456 nt sequenced (16 animals \( \times 15 466 \) nt—D-loop exempted) \( (p = 1.0, \) Fisher’s exact test, two-tailed). Taken together, these data imply a mutation frequency of \( 3.0 \times 10^{-5} \) per nucleotide for the Kerguelen mice.

Because mouse inbred laboratory lines were also derived from only a single mitochondrial variant starting about 100 years ago [10], they can serve as an excellent comparison for the Kerguelen colonization. Among 57 completely sequenced genomes from laboratory lines, we found 30-point mutations (table 2), implying a mutation frequency of \( 3.2 \times 10^{-5} \) per nucleotide. Sixteen mutations were non-synonymous, seven synonymous and seven occurred in the tRNA genes. However, six of these belong to a single unusual haplotype (see the electronic supplementary material, table S2). In addition, one indel was found in a tRNA and the tRNAsP showed the same variable stretch of As as in Kerguelen. No point mutations were found in the D-loop sequences of the laboratory mice.

4. Discussion
The full sequencing of the Kerguelen mice mitochondria carrying a D-loop mutation yielded no convincing evidence for a possibly adaptive mutation that could have caused a selective sweep (table 1). The two haplotypes with the C at position 15 524 are only linked to a synonymous mutation. The two haplotypes with the C at position 15 597 are linked to a coding mutation in ND4, but this was also found independent of the D-loop mutation on the same island, i.e. they must have arisen independently of each other. Of the four haplotypes carrying the T at position 15 577, only one is linked to a coding mutation in COII and the two haplotypes carrying the C at 15 548 are not linked to a second mutation.

When one can exclude positive selection, the mutational pattern should be compatible with the nearly neutral model of evolution. This model assumes that the polymorphism pattern reflects mostly the primary mutation rate and implies the presence of mutations with weak negative effects in the population, because these would be lost over longer evolutionary time [11]. If this model applies, we should see approximately the same pattern in the laboratory lines, since these were also derived from a single mitochondrial haplotype [10]. Although the laboratory lines were established only 100 years ago, they are likely to have gone through more generations than the animals in the wild and are thus comparable with Kerguelen mice that started to expand no more than about 200 years ago [12]. The ratio of 30 point mutations in 57 sequences from laboratory lines is not significantly different from our finding of seven point mutations in 16 sequences from Kerguelen mice (table 2) (note that the D-loop mutations were not included in this calculation, because they served to pre-select the genomes that were fully sequenced and are therefore not representative). Similarly, the respective ratios of coding to synonymous changes and as the transition–transversion ratios in the wild populations and laboratory lines are not significantly different from each other (table 2). However, the fact that no D-loop mutations were found in the laboratory lines is unusual, because the frequency of D-loop mutations in Kerguelen is not different from the frequency found in the rest of the mitochondrial genome. This may still be a sampling effect, given that only 57 laboratory individuals were surveyed compared with 310 Kerguelen mice \( (p = 0.61, \) Fisher’s exact test, two-tailed), but an as yet unknown effect cannot be excluded either.

The higher number of non-synonymous (total of 20) than synonymous (total of 9) mutations in the whole dataset (table 2) corresponds to the fact that there are also more than twice as many non-synonymous sites (8603) as synonymous ones (2794) in the laboratory lines. These ratios are
Table 1. Mutations found in the full mitochondrial sequences of the Kerguelen mice—positions that differ from the consensus sequence are shown. Positions and outgroup sequences are taken from the UCSC genome browser (mm9). n.a., not applicable; nc, non-coding (silent site).

<table>
<thead>
<tr>
<th>position</th>
<th>tRNA^Val</th>
<th>ND2</th>
<th>ND2</th>
<th>COI1</th>
<th>COI1</th>
<th>tRNA^Arg</th>
<th>ND4</th>
<th>ND6</th>
<th>CR</th>
<th>CR</th>
<th>CR</th>
<th>CR</th>
<th>CR</th>
<th>CR</th>
<th>CR</th>
<th>CR</th>
<th>CR</th>
<th>CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>consensus</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Port Gouvreux</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guillou Island</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isthme Bas 06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>La cabane à Jacky 0902</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0902</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>La cabane à Jacky 0912</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0912</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>La cabane à Jacky 0922</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0922</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mayes 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mayes 34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mayes 46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Port-Aux- Français 0933</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Français 0939</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Port-Jeanne d'Arc 01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Port-Jeanne d'Arc 03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Port-Jeanne d'Arc 0903</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Port-Jeanne d'Arc 0907</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Port-Jeanne d'Arc 0908</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
not significantly different from each other ($p = 0.39$, Fisher's exact test) suggesting little germline-level selection against non-synonymous mutations. Still, one should nonetheless expect that there should be selection against strongly deleterious non-synonymous mutations in the germline, as found in mitochondrial mutator lines [8]. Indeed in the Kerguelen mice, we find only one non-conservative amino acid change (as judged by the Blossum64 matrix score—table 1), namely the D→Y change at position 13743. But this was found in only one animal and this is also the only one that is heteroplasmic. All mutations that we find in Kerguelen represent amino acid variants that can also be found in at least one of the outgroups at the respective positions (table 1), i.e. should in principle be tolerable. Hence, germline selection against deleterious mutations may have also occurred in Kerguelen, but not to a degree that has led to a significant bias against non-synonymous changes.

Evolutionary rate estimates for calculating divergence times are usually expressed in substitutions per million years. However, scaling this to very recent divergence comparisons requires considering the respective generation times. When we convert the mutation frequency for the Kerguelen mice into a substitution rate, we would estimate 0.15 substitutions per site per Myr ($3 \times 10^{-5}$ mutations per site in the 200 years since colonization). But it may be possible that generation times in the Kerguelen population have been longer than in most other mouse populations, owing to adverse climatic conditions. If we take the laboratory line rate, which was estimated to have had 2.4 generations per year [10], we would get 0.32 substitutions per site per Myr. Hence, it seems safe to estimate an intermediate value 0.23 substitutions per site per Myr for recently separated wild populations of mice. The long-term evolutionary rate for the full mitochondrial genome, which is a combination of mutation rate and fixation rate, was estimated to be 0.037 substitutions per site per Myr [10], i.e. about six times lower than the short-term rate.

### Table 2. Comparison of mutations in the Kerguelen mice versus laboratory inbred strains.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>point mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kerguelen</td>
</tr>
<tr>
<td></td>
<td>genotypes</td>
</tr>
<tr>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td>coding</td>
<td>4</td>
</tr>
<tr>
<td>synonymous</td>
<td>2</td>
</tr>
<tr>
<td>structural</td>
<td>1</td>
</tr>
<tr>
<td>RNA</td>
<td>5</td>
</tr>
<tr>
<td>transitions</td>
<td>2</td>
</tr>
<tr>
<td>transversions</td>
<td>2</td>
</tr>
</tbody>
</table>

$^a$Six of the seven mutations occur in a single haplotype only.

$^b$Genomes sequenced versus point mutations, Fisher's exact test (two-tailed).

$^c$Coding versus synonymous changes, Fisher's exact test (two-tailed).

$^d$Transitions versus transversions, Fisher's exact test (two-tailed).
calculated above, which includes the slightly deleterious mutations that would be lost over time.

Our results thus support the notion of a time-dependency of molecular evolution [2]. In their review, Ho et al. [2] noted that there is a paucity of reliable calibrations for time frames that are between very short (e.g. documented pedigrees or laboratory mutation–accumulation lines) or very long (e.g. palaeontological evidence) time frames. Given the well-established timing of our dataset, we can provide here such a calibration point with 0.23 substitutions per site per Myr for the few-hundred-years time frame. Full mitochondrial genome comparisons are not yet available from natural populations that have split a few thousand years ago. However, studying the D-loop, Rajabi-Maham et al. [13] found rates of 0.4 substitutions/site/Myr for post-glacially separated populations and 0.06–0.1 substitutions/site/Myr for the subspecies divergence, i.e. also an up to sixfold difference. Hence, these rates appear to confirm the notion of a decline of mitochondrial molecular divergence rates over time [2], but determining the shape of the decline will still require full mitochondrial genome sequences from populations that have split in the several-thousand-year range.

We thank J.-L. Chapuis and the Institut Paul-Emile Victor, program no. 136 for providing the mouse samples, and we thank H. Buhtz and C. Burghardt for expert technical support. We thank the editor for helping to improve the manuscript. The study was funded through institutional funds of the Max Planck Society to D.T.

References


