Letter

mtDNA phylogeny and evolution of laboratory mouse strains

Ana Goios,1,2,6 Luísa Pereira,1,3 Molly Bogue,4 Vincent Macaulay,5 and António Amorim1,2

1Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), 4200-465 Porto, Portugal; 2Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal; 3Medical Faculty, University of Porto, 4200-319 Porto, Portugal; 4The Jackson Laboratory, Bar Harbor, Maine 04609, USA; 5Department of Statistics, University of Glasgow, G12 8QQ Glasgow, United Kingdom

Inbred mouse strains have been maintained for more than 100 years, and they are thought to be a mixture of four different mouse subspecies. Although genealogies have been established, female inbred mouse phylogenies remain unexplored. By a phylogenetic analysis of newly generated complete mitochondrial DNA sequence data in 16 strains, we show here that all common inbred strains descend from the same Mus musculus domesticus female wild ancestor, and suggest that they present a different mitochondrial evolutionary process than their wild relatives with a faster accumulation of replacement substitutions. Our data complement forthcoming results on resequencing of a group of priority strains, and they follow recent efforts of the Mouse Phenome Project to collect and make publicly available information on various strains.

[Supplemental material is available online at www.genome.org. The sequence data from this study have been submitted to GenBank under accession nos. EF108330–EF108345.]

Long before the beginning of mouse genetics, humans in Eastern Asia were already breeding mice in order to obtain animals with different coat colors. Modern mouse genetics, however, did not start until the early 20th century with William Castle’s studies on inheritance. Most of his mice derived from collections of mice fanciers, and they were the ancestors of many inbred strains that are still used today (Rader 2004).

Mouse strains are known to have mixtures of various ancestral genomes from different Mus musculus (house mouse) subspecies (for review, see Yoshihki and Moriwaki 2006). Different molecular markers suggest that the main contributors are M. musculus musculus (Bishop et al. 1985), M. musculus domesticus (Yonekawa et al. 1982; Sakai et al. 2005), and, to a lesser extent, M. musculus castaneus (Sakai et al. 2005). One other subspecies that is usually considered to have contributed is M. musculus molossinus (Sakai et al. 2005), although this itself is supposed to be a hybrid between M. m. musculus and M. m. castaneus (Lundrigan et al. 2002; Wade et al. 2002).

More than 450 mouse inbred strains have been established since the first—DBA/2 (dilute, brown, non-agouti)—was developed by Castle’s student Clarence Cook Little in what would become The Jackson Laboratory (Beck et al. 2000; Rader 2004). Furthermore, in different laboratories worldwide, many sub-strains of each strain have also been maintained.

A mouse strain is defined as inbred if the animals have been crossed brother × sister for ≥20 consecutive generations and individuals of the strain can be traced back to a single ancestor pair at the 20th or subsequent generation (Eppig et al. 2005). Theoretical studies indicate that, at this time, ≥98.6% of loci should be homozygous, but many strains have been bred for >150 generations, which makes them homozygous at virtually every locus (Beck et al. 2000).

Beck et al. (2000) extensively documented inbred mouse genealogies, suggesting that independent inbreeding processes occurred in at least three regions of the globe: (1) Castle’s mice (Group B) and C57-related strains (Group E) originated from Abbie Lathrop’s stocks in the United States; (2) Swiss mice (Group A) derived from mice from Switzerland; and (3) strains derived from colonies from China and Japan (Group C). Little is known, however, of mitochondrial DNA (maternal) phylogenies of these strains. So far, there has been only one study on complete mtDNA sequences from different inbred strains, with the purpose of revising the complete mouse mtDNA reference sequence (Bayona-Bafaluy et al. 2003). The study of mtDNA phylogenies has the potential to elucidate the matrilineal lineage of common inbred strains and its relationship to the main subspecific lineages.

Mammalian mitochondrial DNA (mtDNA) is a circular double-stranded molecule that encodes 13 genes of the respiratory chain. Defects in these molecules have been associated with a variety of disorders that may affect different tissues in different ways (for review, see Wallace 1999). A number of mouse models have recently been developed that have clarified how mutations in mtDNA are transmitted. Examples include transmiotochondrial mice carrying heteroplasmic point mutations (Sligh et al. 2000), heteroplasmic mice with mtDNAs from variants characteristic of two different strains (Battersby and Shoubridge 2001), and mice with homoplasmic replacement of endogenous mtDNA (McKenzie et al. 2004). With this recent increasing interest in mtDNA from mice models, it becomes important to know the mtDNA sequence in each of the different inbred strains.

The analysis of complete mtDNA sequences of inbred mice is also useful for the establishment of mutation/substitution rates. Given that, in principle, these restricted animal populat-
Goios et al.

Results

mtDNA diversity

We have sequenced the complete mtDNA of 32 mice belonging to 16 inbred strains (two of each strain): 12 common inbred strains (129S1/SvImJ, A/J, AKR/J, BALB/cByJ, B6TBR T+J, C3H/HeJ, C57BL/6J, DBA/2J, FVB/NJ, KK/HJ, NOD/LtJ, NZW/LacJ) and five wild-derived inbred strains, each representing a different subspecies (CAST/EiJ, MOLF/EiJ, PWD/PhJ, and WSB/EiJ). Out of the 16 strains sequenced we obtained 12 different haplotypes. For all inbred strains, including the wild-derived, sequencing of the second animal of the same strain always confirmed the first sequence (haplotype) obtained.

We combined our data with 13 complete sequences available online, giving a total of 20 different haplotypes consisting of (1) 11 common inbred strains plus (2) two more divergent (NZB and MilP) common inbred haplotypes, (3) two cell lines, and (4) five wild-derived inbred strains’ haplotypes. The mtDNAs of all 11 common inbred strains were very similar. Only 15 substitutions were detected in the 16,299 mtDNA nucleotides, 14 inside genes and one in a ribosomal RNA (Supplemental Table 1). No substitutions were observed in the control region. Insertions in transfer RNA genes were also observed in sites previously reported as highly polymorphic (Bayona-Bafaluy et al. 2003; Supplemental Table 1). This similarity between common inbred strains has already been reported for some of these strains (Bayona-Bafaluy et al. 2003). However, we have extended the number of strains and included mice from different groups (Beck et al. 2000): A (Swiss mice), B (Castle’s mice), C (strains derived from China and Japan), and E (C57-related strains).

For strains C3H and C57BL/6, we obtained different haplotypes from the published ones. C3H/HeJ differs from C3H/He (Kiebish and Seyfried 2005) at position 8889, and the two C57BL/6 sequences differ at position 11,780. Based on the data of Kiebish and Seyfried (2005) and on a personal communication with J.A. Enríquez (Universidad de Zaragoza, Spain), we concluded that the substitution at 11,780 was probably an artifact, and will therefore be ignored, while that at 8889 was likely real. All other common inbred strains that we could compare with published data matched the reported sequences.

The four strains derived from wild mice that we sequenced are highly divergent from each other, and three of them also diverge from the reference C57BL/6 sequence: CAST (M. m. castaneus) differs from the reference in 377 positions, PWD (M. m. musculus) and MOLF differ (M. m. molossinus) in 386 positions. The MOLO strain (M. m. molossinus) mtDNA that was already reported (Akimoto et al. 2005) differed from MOLF in 15 sites. WSB, the M. m. domesticus representative, shows only 10 differences from the reference.

Phylogenetic analyses

We evaluated the possible branching order through a preliminary network analysis (data not shown) and tried to assess which was the first split within the wild-derived mice by maximum likelihood analysis. Since the closest outgroup is the rather divergent Rattus norvegicus, we did not include the control region in this first tree. The maximum likelihood values were not different enough to determine which of the three most divergent groups separated first, and therefore we left this deepest split unresolved with the following three branches diverging (Fig. 1): (1) Mus musculus castaneus-derived strain, (2) M. m. musculus and M. m. molossinus representatives, and (3) a group that includes the M. m. domesticus representative and all the common inbred strains, here represented by the C57BL/6 branch. Apart from NZB and MilP, all common inbred strains clustered together with the M. m. domesticus representative.

We estimated coalescence times for these branches assuming a divergence time between Mus and Rattus of 12 million yr (palaeontological data from Jacobs et al. 1989; Jaeger et al. 1986 in Michaux et al. 2002; Suzuki et al. 2004). The results date the split between the three main groups at 371,000 ± 91,300 yr ago, consistent with published mtDNA data that place the divergence between Mus musculus subspecies between 0.1 Mya (based on

![Figure 1. Maximum likelihood phylogenetic tree of five wild-derived and three common inbred strains. Rattus norvegicus was used as an outgroup.](image-url)
mtDNA phylogeny and evolution of lab mouse strains

Table 2. Divergence times for the major groups of strains

<table>
<thead>
<tr>
<th>Strains</th>
<th>Divergence times (yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rattus/Mus</td>
<td>12,000,000</td>
</tr>
<tr>
<td>Mus musculus subspecies</td>
<td>371,000 ± 91,000</td>
</tr>
<tr>
<td>NZB, MilP/WSB, common inbred</td>
<td>75,800 ± 14,500</td>
</tr>
<tr>
<td>WSB/common inbred</td>
<td>8900 ± 5200</td>
</tr>
<tr>
<td>AKR, BTBR, DBA/KK</td>
<td>3000 ± 1400</td>
</tr>
<tr>
<td>AKR, BTBR, DBA/C57BL/6</td>
<td>3000 ± 1400</td>
</tr>
<tr>
<td>C3H/He/C3H/HeJ</td>
<td>2200 ± 1400</td>
</tr>
<tr>
<td>NZW, SAMR1/SAMP</td>
<td>1800 ± 1500</td>
</tr>
<tr>
<td>SAMP1/SAMP8</td>
<td>490 ± 120</td>
</tr>
</tbody>
</table>

major subspecies, and so we present a tree with a tricotomy at the base (Fig. 1): (1) M. m. castaneus, (2) M. m. musculus, and (3) M. m. domesticus. Not surprisingly, this ambiguity had already been unresolved in previous analyses with mitochondrial genes Cyt b and 12S (Lundrigan et al. 2002), while analyses with combined nuclear and mitochondrial genes disagree on which subspecies was the first to diverge (Lundrigan et al. 2002; Suzuki et al. 2004).

Our results from complete mtDNA sequences agree with others based on particular genes (Lundrigan et al. 2002) in placing M. m. molossinus closer to M. m. musculus than to other subspecies. This confirms not only that M. m. molossinus is a hybrid but also that its mtDNA shares a recent common ancestor with M. m. musculus.

A single female origin for lab mice strains

The extreme similarity among mtDNAs of the different common inbred strains is well illustrated by the small number of polymorphisms that were found (Fig. 2; Supplemental Table 1). The accumulation of polymorphisms agrees with published genealogies (Beck et al. 2000) and the history (Eppig et al. 2005) of the strains, particularly when individual groups (Swiss mice, Castle’s mice, strains derived from China and Japan, and C57-related strains) of strains are considered. Small deviations from the published genealogical trees can generally be justified by outbreeding somewhere back in the history of the strain (each strain is discussed in detail in the Supplemental data).

Genealogies do not suggest any relationship between the different groups (except for C57-related mice and Castle’s mice, which both descend from Abbie Lathrop’s stocks). Our results show, however, that all the common inbred strains appear clustered in the network, irrespectively of the group to which they belong. This finding suggests a common origin for all analyzed groups of common inbred strains. Furthermore, the fact that, in the phylogenetic tree, common inbred strains appear clustered with the M. m. domesticus-derived strain, and far from representatives of other subspecies, shows that the common mitochondrially cytochrome B) and 0.5 Mya (based on mitochondrial and nuclear genes combined) (Suzuki et al. 2004).

Based on the same assumption, we obtained an overall substitution rate for the mouse coding mtDNA of 3.7 × 10⁻⁸ substitutions per site per yr. This value is rather similar to the 1.26 × 10⁻⁸ substitutions per site per yr obtained by Mishmar et al. (2003) for the human mtDNA coding region. A higher substitution rate of 1.1 × 10⁻⁷ substitutions per site per yr is observed for third codon positions, while the second codon positions present the lowest substitution rate of 5.1 × 10⁻⁸ substitutions per site per yr (Table 1).

We produced a second tree (not shown) with complete mtDNA data from all common inbred strains, cell lines, and wild-derived mice, but excluding the rat. This analysis confirmed that all common inbred mice group with the M. m. domesticus representative. Based on the coalescence time for the main subspecies of Mus calculated in the above analysis, we obtained similar substitution rates to those derived from the previous group of strains (Table 1). Moreover, we could estimate a substitution rate for the control region of 5.6 × 10⁻⁸ per site per yr, which is about three to four times higher than the substitution rate of the first codon positions and eight to 11 times higher than that of the second codon positions. Third codon positions, however, remain the ones with a highest substitution rate, which is twice the value obtained for the control region.

We estimated the divergence time for these common inbred mice and obtained older times than expected. The analysis of the common inbred strains (excluding the most divergent strains NZB and MilP) was dated at 3000 ± 1400 yr ago (Table 2), although it is documented that the first inbred strain was established only ~100 yr ago (Rader 2004). Another example is the case of the SAMP strains that were not separated before 1968 (Takeda et al. 1997), though our estimate for their coalescence time was 490 ± 120 yr ago.

The mtDNA sequences of the common inbred mice show a star-like network (Fig. 2). The domesticus-derived WSB is the furthest away from the center. Except for the SAM strains, all the strains that belong to groups A (Swiss mice), B (Castle’s mice), and E (C57-related strains) are in the central groups or less than one polymorphism apart. Four inbred strains are two or more polymorphisms apart from the central groups: KK belongs to group C (strains derived from colonies from China and Japan), cell-line LA9, and the SAM strains. Although these belong to group B, they were crossed with mice from an unknown strain in 1968 (Beck et al. 2000).

Discussion

The maximum likelihood analysis of the wild-derived strains did not permit the establishment of the first split among the three

Table 1. Substitution rates for the different positions of the mtDNA molecule, calculated based on different groups of mouse strains

<table>
<thead>
<tr>
<th>Substitution rates (substitutions per site per yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rat, wild-derived, NZB, MilP, CS7BL/6</td>
</tr>
<tr>
<td>Codon position 1</td>
</tr>
<tr>
<td>Codon position 2</td>
</tr>
<tr>
<td>Codon position 3</td>
</tr>
<tr>
<td>tRNAs</td>
</tr>
<tr>
<td>rRNAs</td>
</tr>
<tr>
<td>Control region</td>
</tr>
<tr>
<td>Overall</td>
</tr>
</tbody>
</table>
NZW shares the same haplotype as SAMR1. Female ancestor (Mizutani et al. 2001) of SAM strains, by showing. Moreover, our results suggest a hypothesis about the unknown aging process (Chinnery et al. 2002; Chomyn and Attardi 2003). mtDNA mutations have been proposed to be implicated in the could be related to their senescence-accelerated phenotype, since have accumulated more substitutions than any other strain. This SAMP strains, although established only 30 yr ago, appear to be the origin of all three groups; the maintenance of separate colonies in different regions would have resulted in the few differences that are now observed. The most divergent inbred strains were NZB and MilP as already reported by Bayona-Bafaluy et al. (2003). Different reasons may explain the divergence of these two strains: the MilP strain is not a true common inbred strain, as it descends from a wild female mouse caught in Italy (Loveland et al. 1990 in Bayona-Bafaluy et al. 2003), and NZB, although belonging to group B (Castle's mice), did not originate from Abbie Lathrop’s stocks, but from unidentified “European & U.S. stocks,” and it was outbred during its development (Beck et al. 2000).

Although few polymorphisms were observed between the different common inbred strains, a couple of comments are worth making: (1) A polymorphism has occurred in the split between two substrains of C3H. C3H/He and C3H/HeJ can therefore be distinguished by the nucleotide at position 8889. (2) SAMP strains, although established only 30 yr ago, appear to have accumulated more substitutions than any other strain. This could be related to their senescence-accelerated phenotype, since mtDNA mutations have been proposed to be implicated in the aging process (Chinnery et al. 2002; Chomyn and Attardi 2003). Moreover, our results suggest a hypothesis about the unknown female ancestor (Mizutani et al. 2001) of SAM strains, by showing that NZW shares the same haplotype as SAMR1.

Different mtDNA evolution in wild and inbred strains

The substitution rates calculated based on wild and inbred mice data vary depending on the position. Third codon positions, unlike second positions, show a higher substitution rate probably because degeneracy of the genetic code makes these positions less prone to selective pressure: A change in a third position base will probably not replace the coded amino acid. Intriguingly, when analyzing the substitutions that occur in the inbred strains (Supplemental Table 1), we observe a low frequency (4/14) of substitutions involving third codon positions, and an even lower number (3/14) of synonymous variants. The same is observed for the WSB strain. However, this could be a consequence of performing the comparisons relative to an inbred strain (C57BL/6). Most (13) of the 15 substitutions found among inbred strains are transitions, and the overall transition/transversion rates were higher for all positions in the analysis with common inbred strains than in the one including the rat (data not shown). This higher accumulation of transversions in the wild was expected given the very old Mus/Rattus divergence when compared with the inbred strains.

Although data on RFLPs have placed the divergence inside the Mus musculus species 1–2 million yr ago (Yonekawa et al. 1981), long before the ~370,000 yr ago that we obtained, our estimated coalescence times for the wild-derived strains fall within values obtained more recently with other markers (Suzuki et al. 2004). The estimates of coalescence times of the inbred strains, however, were much higher than the known divergence of the strains. It is known that the common inbred strains were first established at the beginning of the 20th century, but our estimate for coalescence of the inbred strains is ~3000 yr ago. Even if, as is suggested by our network, all the common inbred strains (Castle’s mice, Swiss mice, strains derived from colonies from China and Japan, and C57-related strains) had the same ancestor earlier than the 20th century, in Japanese fancy mice (17th century, Rader 2004), values obtained in our estimates would still be very high. Furthermore, the divergence times for the splits between the different inbred strains are also much older in our estimates than what is documented.

Taken together, these estimates show a faster mtDNA evolution in inbred than in wild mice with a higher accumulation of replacement substitutions. Three hypotheses may account for these discrepancies: (1) the calibration point that we used (12 million yr for the Mus/Rattus divergence) is too distant relative to the young splits that we are trying to date; (2) mtDNA evolution in inbred strains, with such controlled conditions and restricted crosses, occurs faster than in wild environments; or (3) there are differences in generation times of inbred and wild strains.

The issue of the calibration points has been under debate in recent publications, and it has been suggested that substitution rates extrapolated across different time scales result in invalid date estimates (Ho and Larson 2006). In our case, using a 12-million-yr calibration point for a 100-yr-old estimate might be inappropriate. Since it was not possible to find a more suitable calibration point (paleontological data after the Mus/Rattus split are absent), we tried to overcome this limitation by assuming that our estimate for the Mus musculus divergence (~370,000 yr
The substitution rate in inbred strains can be influenced by several factors, including: (1) a sequencing error and/or contamination, (2) the strain was misidentified, or (3) the strain following conclusions: (1) there was a sequencing error and/or contamination that we obtained from the wild strain data are of the same order of magnitude as the values obtained for human mtDNA may suggest that our estimates for the substitution rates are not too far from the truth.

Therefore, a higher substitution rate in inbred mice may still stand, despite this rough time estimate. Curiously, mice cell lines included in our analyses have accumulated a similar number of substitutions to that observed for mouse inbred strains, although it had been reported that relaxation of negative selection resulted in a high substitution rate in HeLa cell lines, when compared with phylogenetically derived divergence of mtDNAs (Hennstadt et al. 2002). Evolutionary rates of inbred mouse mtDNA appear to be closer to cell lines than to wild mice, suggesting that laboratory controlled conditions with strict bottlenecks may also account for the higher substitution rate.

A higher substitution rate could also be apparent if generation times of wild mice are longer than in inbred strains. It is possible that mice breeders cross mice before they would mate in the wild. For C57BL/6J mice it is documented (Genetic Background Resource Manual 2006, The Jackson Laboratory, http://jaxmice.jax.org/literature/manuals/mouse_genetics_resource_manual.pdf) that ~120 generations have passed in the period from 1920 to 1970, resulting in 2.4 generations per yr, but other inbred strains may have shorter generation times. Wild mice are commonly assumed to have about two generations per yr. This difference could explain a slightly higher substitution rate, but it does not explain the high frequency of replacement substitutions.

Final remarks

Although the history of inbred mice reports at least three geographically separate origins for inbred mice, our data permit us to conclude that all common inbred strains that are now being resequenced originate from the same female ancestor, a M. m. domesticus mouse. Phylogenetic analyses suggest that two different processes of mtDNA evolution may occur depending on the environment: one in wild mice that results in a substitution rate similar to the one obtained for human mtDNA, and the other in inbred strains. Assuming from paleontological data (Jacobs et al. 1989; Jaeger et al. 1986 in Michaux et al. 2002; Suzuki et al. 2004) that the Mus/Rattus divergence occurred 12 million yr ago, we determined substitution rates and coalescence times for all the branches in the tree. We performed a separate maximum likelihood analysis excluding the rat sequence and including all common inbred strains, and considering the whole molecule (including the control region). Assuming that the divergence time of the Mus musculus subspecies calculated in the previous tree was correct, we recalculated substitution rates and coalescence times separately for this tree and compared the results. The two separate phylogenetic analyses allowed us to compare results obtained with and without the common inbred strains and to reduce the amount of difference in evolutionary scale of the strains used in each case.

We also constructed a maximum parsimony phylogenetic network including data from the complete mtDNAs from all common inbred strains and the M. m. domesticus representative (WSB). This summarizes all sequence data and illustrates the evolutionary relationships among the common inbred strains.
treme lack of variation among the different common inbred strains.

**Accession numbers**

The published complete mtDNA sequences used in the analyses are available in GenBank, accession nos. AB042432, AB042523, AB042524, AB042809, AJ489607, AJ512208, AY172335, AY139599, L07095, and L07096 (Bayona-Balafuy et al. 2003, AY466499 (Moreno-Loshuertos et al. 2006), AY675564 (Akimoto et al. 2005), and DQ106412 (Kiebisch and Seyfried 2005). The results have been submitted to the Mouse Phenome Database (http://www.jax.org/phenome) as part of the project MPD: 202 (Goios et al. 2006).

**Acknowledgments**

This work was partially supported by Fundação para a Ciência e a Tecnologia through a research grant to A.G. (SFRH/BD/16518/2002) and by “Programa Operacional Ciência, e Inovação 2010” (POCI 2010), VI Programa-Quadro (2002–2006).

**References**


