Mitochondrion 17 (2014) 126–131

Contents lists available at ScienceDirect

Mitochondrion

journal homepage: www.elsevier.com/locate/mito

mtDNA sequence, phylogeny and evolution of laboratory mice

Junfeng Zheng [a], Yan Chen [a], Fengyuan Deng [a], Renliang Huang [a], Frank Petersen [b], Saleh Ibrahim [c], Xinhua Yu [a,b,*]

[a] The Medical College of Xiamen University, Xiamen University, 361005 Xiamen, China
[b] Priority Area Asthma and Allergy, Research Center Borstel, 23845 Borstel, Germany, Airway Research Center North (ARCN), UGMLC, Member of the German Center for Lung Research (DZL)
[c] Department of Dermatology, University of Lübeck, 23538 Lübeck, Germany

ARTICLE INFO

Article history:
Received 6 March 2014
received in revised form 8 July 2014
accepted 9 July 2014
Available online 16 July 2014

Keywords:
mtDNA
Outbred stock
Inbred strains
Phylogenetic history
Relaxed purifying selection

ABSTRACT

Laboratory mice are important tools for biomedical research. Aiming to investigate the phylogeny and evolution of laboratory mice, we investigated the mtDNA sequences of classic inbred strains, classic outbred stocks and wild-derived inbred strains. Our results showed that the most classic outbred stocks and classic inbred strains are descended from a single mtDNA ancestor. The phylogenic analysis supports the topology of *M. m. castaneus*/*M. m. domesticus* as sister subspecies, and the divergence time between the two sister subspecies and *M. m. musculus* was 493,000 (435,000–557,000) years ago. Furthermore, the mtDNA polymorphisms accumulated in the last 100 years in the laboratory mice are under a relaxed purifying selection.

1. Introduction

The house mouse (*Musculus musculus*) originated from a central population from Asia and subsequently evolved into different subspecies and colonized in different regions around the world (Guenet and Bonhomme, 2003; Prager et al., 1996). There are three major *Mus musculus* subspecies, namely *M. m. domesticus*, *M. m. castaneus*, and *M. m. musculus*. The phylogenetic history of the three *M. mus.* subspecies has three possible topologies, including *M. m. musculus*/*M. m. domesticus*, *M. m. musculus*/*M. m. castaneus*, and *M. m. castaneus*/*M. m. domesticus* as sister subspecies respectively. Recently, White et al. investigated the nuclear genome of the three *M. musculus* subspecies representative strains. Their results suggested that the most likely topology is *M. m. musculus*/*M. m. castaneus* as the sister subspecies, followed by the possibility of *M. m. castaneus*/*M. m. domesticus* as sister subspecies (White et al., 2009). However, investigating the mtDNA of *Mus mus.* subspecies using rat mtDNA as outgroup was unable to detect which topology represents the phylogenetic history (Goios et al., 2007).

Mice are the most widely used animal in biomedical research. Laboratory mice have been established in the last 100 years from fancy mice and wild-caught house mice. A genetic study revealed that laboratory mice have a mixed genome of all three major *M. musculus* subspecies (Yang et al., 2007). The laboratory mice consist of inbred strains and outbred stocks. The origin of mouse inbred strains and outbred stocks has been extensively documented by Beck et al. and Chia et al. in 2000 and in 2005, respectively (Beck et al., 2000; Chia et al., 2005). Inbred mouse strains contain classic inbred strains (CIS), which either originated from Castel's mice, Swiss mice or Asia mice, the wild-derived inbred strains (WIS), which were developed from wild-caught mice (Beck et al., 2000). Previous studies revealed that most CIS were descended from a single mtDNA ancestor (Goios et al., 2007; Yu et al., 2009). In regard to the classic outbred stocks (COS), most COS originated from 9 Swiss mice, including 2 males and 7 females (Chia et al., 2005; Yalcin et al., 2010).

The mitochondrial genome (mtDNA) is a closed, circular, double-stranded and strictly maternally transmitted DNA (Wallace, 1999). Due to exposure to high oxidative damage, the lack of protective histones and an efficient DNA repair system, the mtDNA mutation rate is much higher than that of the nuclear genome (Richter et al., 1988; Wallace, 1999). Furthermore, given the essential role of mitochondria in cells, mtDNA is sensitive to natural selection. Therefore, the mtDNA is widely used to study phylogenic history and evolution. In this study we aimed to investigate the mtDNA phylogeny and evolution of laboratory mice. In our study we i) determined the female origin of the laboratory COS, ii) analyzed the phylogenetic history of the laboratory mouse, and iii) characterized the natural selection on laboratory mouse mtDNA.

* Corresponding author at: The Medical College of Xiamen University, Xiamen University, 361005 Xiamen, China. Tel.: +86 5922188621; fax: +86 5922182736.
E-mail address: xinhuayu@fz-borstel.de (X. Yu).

http://dx.doi.org/10.1016/j.mito.2014.07.006
1567-7249/

2. Materials and methods

2.1. DNA and sequencing

The three outbred stocks including Slac:ICR, Slac:KM and Slac:CF-1 were obtained from SLAC Laboratory Animals (Shanghai, China). The DNA of the three outbred stocks was prepared from the liver. Mice were sacrificed in a CO_2 chamber, and livers of the mice were used for the DNA preparation. The institutional ethics committee of the Medical College of Xiamen University approved this study. The DNA of the PANCEVO/EiJ strain was obtained from the Jackson Laboratory. We sequenced the entire mtDNA of three mice per outbred stock and one mouse of the PANCEVO/EiJ inbred strain. The mtDNA was sequenced using the method that was described previously (Yu et al., 2009). Briefly, we amplified 14 overlapping fragments of 1.3–2.5 kb covering the entire mtDNA molecule. The PCR products were purified and then directly sequenced. The mtDNA sequences of the three COS and the PANCEVO/EiJ strain are shown in supplementary text 1. The sequence data from this study have been submitted to GenBank under accession numbers KF937873–KF937876.

2.2. Phylogenetic analysis

Sequences of mouse mtDNAs of 8 representative mouse strains, including C57BL/6J, WSB/EiJ, NZB/BiNJ, MiP, MOLF/EiJ, PWD/PhJ, CAST/EiJ and PANCEVO/EiJ were aligned using the multiple alignment algorithm in the CLUSTALW program with default settings. The mtDNA of C57BL/6J was used to represent all CLS mice. The aligned dataset was subjected to an ML analysis. The rate variation among sites was modeled using a gamma distribution with four rate categories in the HKY model as suggested in the program Mega 5. We reconstructed an ML tree, and a rapid bootstrap analysis with 1000 replications was simultaneously performed to calculate the robustness of each branch of the resultant tree. All ambiguous positions were removed for each sequence pair, and there were 16,318 positions in the final dataset. The evolutionary history was inferred using the maximum likelihood method (Tamura et al., 2011).

2.3. Coalescence time estimation

To investigate the relative timings of major cladogenetic events in the genus *Mus*, a strict molecular-clock Bayesian method that was implemented in the program MCMCtree in the PAML v.4.4 package was used for the dating analysis (Yang, 1997). The best-scoring Maximum Likelihood (ML) tree from the mtDNA dataset was used for the divergence time estimation and assumed a divergence time between *Musculus spicilegus* and *Mus musculus* of 1.6 (1.5–1.7) million years (Lundrigan et al., 2002; Suzuki et al., 2004). ML analysis estimates of branch lengths were obtained using the BASEML program in the PAML package under the HKY + Γ substitution model with the gamma prior set at 0.5 (Hasegawa et al., 1985). The global molecular clock model was used to specify the priors of rates among internal nodes (clock = 1 in the MCMCtree program). The overall substitution rate was set at G and used a strict molecular clock assumption with 1.5–1.7 million years constraint to root age. The parameters of the birth–death process for tree generation with species sampling were fixed at $\lambda = \mu = 1$ and $\rho = 0$.

All time constraints were provided in a unit of 1 MA (i.e., 1 = 1 MA) because some of the model components in the Bayesian analysis are scale variant, and the node ages should fall between 0.01 and 10. An MCMC approximation with a burn-in period of 15,000 cycles was obtained, and every 100 cycles was taken to create 600,000 samples. To diagnose the possible failure of the Markov chains to converge to their stationary distribution, we performed two replicate MCMC runs with two different random seeds for each analysis. MCMC samples from the two runs were combined after verifying the distributions of parameter values using the program Tracer 1.5 (available from http://tree.bio.ed.ac.uk/software/tracer/website). The number of samples was large enough to reach effective sample sizes (ESSs > 200) for all parameters that were estimated in this study.

2.4. Replacement substitution frequency

The replacement substitution frequency (RF) was calculated as the ratio between the number of non-synonymous polymorphisms and the number of synonymous polymorphisms (NS/S). The theoretical value of RF was calculated as the ratio between the number of possible *non-syn* sites and the number of possible *syn* sites using the DnaSP software (Rozas and Rozas, 1999).

2.5. Stem substitution frequency analysis

The stem substitution frequency (SF) was calculated as the ratio between the number of substitutions on the stem of RNAs and the number of substitutions on the loop of RNAs. The published secondary structures for tRNAs (Helm et al., 2000) and rRNAs (Springer and Douzery, 1996) were used to define the stem and loop structures.

2.6. Statistics

The comparison for the RF was performed using the chi-square test or Fisher's exact test. The comparison for the conservation index was performed using Student's *t*-test. A *P*-value of less than 0.05 was considered significant.

3. Results

3.1. mtDNA sequence of the classic outbred stocks

According to the documentation of the origin of the COS and CIS, most COS and some CIS can be traced back to 2 male and 7 female Swiss mice (Beck et al., 2000; Chia et al., 2005) which are called Swiss COS and Swiss CIS, respectively. Since all sequenced mtDNA of the Swiss CIS including ALR/LtJ, ALS/LtJ, BUB/BnJ, FVB/NJ, MA/MyJ, NOR/LtJ, NOD/ltJ, SJL/J and SWR/J strains are descended from a single mtDNA ancestor (Yu et al., 2009), we speculated that the mtDNA of Swiss COS descended from the same mtDNA ancestor. To verify this hypothesis, we sequenced the mtDNA of two representative Swiss COS, including ICR stock which is widely used in Europe and America and KM stock which is widely used in China. As expected, the mtDNA sequences of the two COS are identical to that of the AKR/J inbred strain, with the exception of the nt9821 A-repeat polymorphism (Table 1). We also sequenced a widely used non-Swiss COS, CF-1 stock which is thought to be not descended from Swiss mice (Chia et al., 2005). Unexpectedly, the mtDNA of the CF-1 stock is completely identical as the mtDNA of AKR/J strain (Table 1). Therefore, this result demonstrates that the mtDNA of most CIS and COS are descended from a single mtDNA ancestor.

Table 1
Sequence polymorphism between classic inbred strains and classic outbred stocks.

Polymorphism	Gene	C57BL/6 J	AKR/J	CF1	ICR	KM
nt9461 T/C	ND3 Met-Met	T	C	C	C	C
nt9821 A repeat	tRNA-Arg	8A	9A	9A	8A	10A/11A

Note: The mtDNA sequence of AKR/J strain (AB042432) and of C57BL/6J strain (NC_005089) were retrieved from genebank. The mtDNA sequence of the CF1, ICR and KM outbred stocks were sequenced in this study.