**Supplementary material for**

**Mice as Stowaways? Colonization history of Danish striped field mice. Biology Letters.**

1#Liselotte Wesley Andersen, 2#Magnus Jacobsen, 3Christina Vedel-Smith, 3Thomas Secher Jensen

1Institute of Bioscience-Kalø, Aarhus University, Grenåvej 14, DK-8410, Denmark

2Institute of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000, Denmark

3Natural History Museum, Aarhus, Wilhelm Meyers Allé 10, DK-8000, Denmark

**#**both authors contributed equally

**Supplementary 1**. To avoid numts, the complete mitochondrial genome was amplified using three long fragments of ~6kb (Richly & Leister, 2004). The PCR’s were carried out in a total volume of 20 µl using GOTaq Long PCR master mix (Promega) using the thermal profile: 950 C for 3 min, 40 cycles of 940 C for 45 s, annealing temperature 550 C (primerAA1 and AA2, Table 1) for 45 s, extension of 720 C for 6 min and finally extension of 720 C for 10 min. Annealing temperature for primer AA3 was 600 C with identical extension time. PCR fragments were purified and concentrations measured using a Qubit 2.0 Fluorimeter. The fragments representing each individual were subsequently fragmented and individual libraries built using NEBNext® Fast DNA Fragmentation & Library Prep Set for Ion Torrent and the Ion express barcodes (1-48) for multiplexing. Finally, sequencing was conducted using Ion express template 400 chemistry (paired end technology) on a 318 chip. All kits were used according to the manufacturer recommendations

Primers used to amplify the complete mitochondrion of the striped field mouse *Apodemus agrarius*. For Poland, Germany, Filskov, Lolland and Falster it was necessary to use all primers to amplify all regions successfully as mutations were observed in the some of the primers.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | Sequence |  | Base pairs |
| mtDNA primer 1 | | AA1F | GTTAGACATAACTCTATTACTACTT | | ~ 5775 |
|  |  | AA1R | CGGGAGCTCCAATCATAAGTGGTAC | |  |
|  |  | AA1F2 | TTTATGAATGCTTGTTAGAC |  | ~ 5775 |
|  |  | AA1R2 | GCATAAGTGGTACGAGTCAG |  |  |
| mtDNA primer 2 | | AA2F | AGTAGGGACTGCATTGAGCATTCTG | | ~5790 |
|  |  | AA2R | CTTGCTACTAACCAGCAAGTTGCT | |  |
|  |  | AA2F2 | GCATTGAGCATTCTGATCCGAG |  | ~5790 |
|  |  | AA2R2 | GCTACTAACCAGCAAGTTGCTATA |  |  |
| mtDNA primer 3 | | AA3F | GCCTAACTTCATCACTTCTAT | | ~5775 |
|  |  | AA3R | AGCAAGAGATGGTGAGGTAGA | |  |
|  |  | AA3F2 | TTGCTCACGGCCTAACTTCAT |  | ~5775 |

**References**

Richly E, Leister D. 2004 NUMTs in Sequenced Eukaryotic Genomes. *Mol. Biol. Evol*. **21**, 1081-1084. (doi: 10.1093/molbev/msh110)

**Supplementary 3a**. Sample size (N) HD (haplotype diversity), π (nucleotide diversity), sd (standard deviation) and Fu’s *Fs* (Fu, 1997) for all analysed geographical regions using the complete mitogenome dataset. Bold= p < 0.05. Analyses were conducted using DnaSP and ARLEQUIN (Librado & Rozas, 2009; Excoffier & Lischer, 2010 ).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| Complete mtDNA | |  |  |  |  |  |
|  | Estonia | Poland | Germany | Filskov | Lolland | Falster |
| N | 3 | 12 | 17 | 22 | 17 | 14 |
| HD | 1 | 0.97 | 0.985 | 0.983 | 0.993 | 0.890 |
| sd | 0.272 | 0.044 | 0.025 | 0.018 | 0.023 | 0.081 |
| π | 0.0023 | 0.0027 | 0.0034 | 0.0009 | 0.0008 | 0.0003 |
| sd | 0.0007 | 0.0002 | 0.0003 | 0.0004 | 0.00008 | 0.00004 |
| Fu's *Fs* | 2.65 | 2.05 | -1.45 | -3.21 | **-4.02** | -3.03 |

**Supplementary 3b**. Phylogenetic relationships amongst the mitogenome sequences were analyzed in BEAST v1.8.2 (Drummond & Rambaut, 2007). Given low genetic variability, we used the entire mitogenome sequence for analysis. The best fit substitution model (HKY+I+G) was identified via the Akaike information criterion with eight gamma categories using JMODELTEST version 2.11 (Posada, 2008). A constant clock evolutionary rate was chosen following the rate heterogeneity tests in TRACER (Rambaut *et al.*, 2014), using an uncorrelated log-normal clock. Given that the standard deviation of the uncorrelated lognormal relaxed clock (ucld.stdev) was abutting against zero (95%HPD of ucld.stdev = 1.76×10-5- 0.25) a strict molecular clock could not be rejected (page 12-13: <http://www.molecularevolution.org/molevolfiles/beast/BEAST14_MANUAL-7-6-07.pdf>) and subsequent analyses were conducted using a strict clock. Subsequently, we tested the fit of three different tree priors to our data and estimated the Bayes factor in order to choose the prior that fitted the data best. In all pairwise comparisons constant size showed the best fit (Bayes factor >3.5; see table below). The final Markov chain Monte Carlo sample obtained was based on a run for 50000000 generations, with genealogies sampled every 5000 generations with 10% discarded as burn-in. Convergence was assessed by ESS and by conducting an additional run, which showed a very similar result. A maximum clade credibility tree with mean heights for branches was estimated in TREEANNOTATOR (Drummond *et al*., 2007) using an initial burn-in of 10%. The tree was visualized and edited in FIGTREE v1.3.1 (Andre Rambaut, University of Edinburgh, <http://tree.bio.ed.ac.uk/software/figtree/>) (Figure 2 in the main paper).

Bayes factor values for the pairwise comparisons of the three tested priors. The values correspond to the fit of the rightmost prior compared with either of the two other priors tested, as denoted in the three following columns.

|  |  |  |  |
| --- | --- | --- | --- |
| Prior tested | Exponential growth | Constant size | Expansion growth |
| Exponential growth | - | 0.285 | 1.413 |
| Constant size | 3.503 | - | 4.951 |
| Expansion growth | 0.708 | 0.202 | - |

Further, IMa was run without migration (m=0) for all pairwise comparisons that included Danish populations. Priors on population sizes (θ) were set to equal values and wide priors were chosen that included the entire likelihood distribution (values of q =40-100 before unscaling by mutation rate). Divergence time (t) was set to 100, which roughly corresponded to 18000 yrs using the mutation rate of 3.2×10-7 sub/site/yr and thus included divergence after the Last Glacial Maximum (LGM). For the last analysis, comparing all Danish individuals with German and Polish individuals (Denmark vs. mainland plot in Fig 1C and Table 2 in the main text), analyses both with and without migration were conducted. Population sizes (q) and divergence time were set to 100 while migration (m) was set to 0 and 10. A linear heating scheme with 10 MCMC chains was used to explore the parameter space in all analyses (-f1 –n 10 –g1 0.05). We used an initial burn-in of 1000000 followed by 100000 chain steps. Sampling was carried out every 10th steps according to default settings. Convergence was assessed by ESS and by conducting an additional run, which showed a very similar result.

**References**

Drummond AJ, Rambaut A. 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol*. **7**. (doi: 10.1186/1471-2148-7-214)

Excoffier L, Lischer HEL. 2010 Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* **10**, 564- 567. (doi: 10.1111/j.1755-0998.2010.02847.x)

Fu YX. 1997 Statistical neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925.

Librado P, Rozas J. 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452. (doi:10.1093/bioinformatics/btp187)

Posada D. 2008 jModelTest: Phylogenetic model averaging. *Mol. Biol.Evol.* **25**, 1253–1256. (doi: 10.1093/molbev/msn083)

Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014 Tracer v1.6, Available from <http://beast.bio.ed.ac.uk/Tracer>

**Supplementary 4**. Positive selection acting on the 13 coding mitochondrial genes was tested using the MEME (Mixed Effects Model of Evolution) (Murrel *et al.* , 2012a) and FUBAR (Fast Unbiased Bayesian AppRoximation) (Murrel *et al.* , 2012b) codon-based selection test implemented in HyPHY (DataMonkey server, (<http://www.datamonkey.org/dataupload.php>)). These methods test for episodic selection (MEME) or diversifying (positive) and negative selection (FUBAR) acting on individual codons. Default settings were used. Substitutions models were estimated in HyPhy and significance threshold set as either P>0.05 (MEME) or Bayes factor >10 (FUBAR). For these analyses we used the full gene sequences from the 86 mitogenome sequenced individuals along with full gene sequences from the additional 22 partial sequenced mitogenomes when available.

Overall, strong negative selection was observed as expected (see Supplementary 4 table below). Moreover episodic selection was observed only within single individuals suggesting relaxed purifying selection (e.g. Ho *et al*., 2005; Jacobsen *et al.*, 2012; 2016). One site in the cytochrome b gene (CYTB) showed evidence for diversifying positive selection (codon 232), however, as the mutation occurring here lead to replacement of amino acids with extremely similar physiochemical properties (Isoleucine (I) to Valine (V)), possible relaxed purifying selection cannot be rejected.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Information about the results of the MEME and FUBAR tests for all 13 coding genes. The analyses were  conducted in HYPhy using all full gene sequences generated in this study, including the 5 reference sequences | | | | | | | | |
| from Korean individuals as outgroups | | | | | | | | |
| **Gene** | **Number of sequences** | **Number of different haplotypes** | **Best substitution model** | **MEME\*** | **FUBAR#** | |
| positive sites | positive sites | negative sites |
| COX1 | 106 | 40 | HKY85 | 0 | 0 | 14 |
| COX2 | 106 | 19 | HKY85 | 0 | 0 | 8 |
| COX3 | 108 | 28 | HKY85 | 1 | 0 | 6 |
| CYTB | 102 | 41 | HKY85 | 1 | 1 | 14 |
| ND1 | 109 | 26 | HKY85 | 1 | 0 | 6 |
| ND2 | 106 | 38 | HKY85 | 1 | 0 | 10 |
| ND3 | 109 | 21 | HKY85 | 0 | 0 | 5 |
| ND4 | 98 | 55 | HKY85 | 0 | 0 | 9 |
| ND4L | 107 | 15 | HKY85 | 0 | 0 | 3 |
| ND5 | 105 | 43 | HKY85 | 0 | 0 | 20 |
| ND6 | 100 | 32 | HKY85 | 1 | 0 | 4 |
| \* All significant amino acid changes found by the MEME method only occurred in one individual | | | | | | |  |  |
| # The one site detected to be under possible positive selection by the FUBAR test was a change at codon 232 from Isoleucine (I) to Valine (V) | | | | | | | | |

**References**

Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005 Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol*. **22**, 1561–1568. (doi: 10.1093/molbev/msi145)

Jacobsen MW, Hansen MM, Orlando L, Bekkevold D, Bernatchez L, Willerslev E, Gilbert MTP. 2012 Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus* spp.). *Mol. Ecol.* **21**, 2727-2742. (doi: 10.1111/j.1365-294X.2012.05561.x)

Jacobsen MW, da FonsecaRR, Bernatchez L, HansenM. 2016 Comparative analysis of complete mitochondrial genomes suggests that relaxed purifying selection is driving high nonsynonymous evolutionary rate of the NADH2 gene in whitefish (*Coregonus* ssp.). *Mol. Phyl. Evol.* **95**, 161-170. (doi: 10.1016/j.ympev.2015.11.008)

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. 2012a Detecting individual sites subject to episodic diversifying selection. *Plos Genetics* **8**, e1002764. (doi: 0.1371/journal.pgen.1002764)

Murrell B, Moola S, Mabona A, Mabona A, Weighill T, Sheward D, Pond SLK, Scheffler K.2013b FUBAR: A Fast, unconstrained Bayesian appRoximation for inferring selection. *Mol. Biol. Evol.* **30**, 1196-1205. (doi: 10.1093/molbev/mst030)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | |
| Gene | Length (bp)1 | NH | HD | s.d. | θ | μRELATIVE2 | μESTIMATED3 | |
| 3.2×10-7 sub/site/yr | 1.5×10-7 sub/site/yr |
| COX1 | 1542 | 32 | 0.830 | 0.037 | 0.00664 | 0.778 | 2.49×10-7 | 1.167×10-7 |
| COX2 | 681 | 15 | 0.699 | 0.044 | 0.00810 | 0.948 | 3.034×10-7 | 1.422×10-7 |
| COX3 | 783 | 18 | 0.690 | 0.051 | 0.00710 | 0.831 | 2.659×10-7 | 1.247×10-7 |
| CYTB | 1143 | 30 | 0.873 | 0.024 | 0.01175 | 1.376 | 4.403×10-7 | 2.064×10-7 |
| ATP6 | 678 | 27 | 0.770 | 0.047 | 0.01279 | 1.498 | 4.794×10-7 | 2.247×10-7 |
| ATP8 | 201 | 6 | 0.128 | 0.048 | 0.00502 | 0.588 | 1.882×10-7 | 0.882×10-7 |
| ND1 | 954 | 19 | 0.653 | 0.057 | 0.00909 | 1.064 | 3.405×10-7 | 1.596×10-7 |
| ND2 | 1035 | 30 | 0.741 | 0.051 | 0.01140 | 1.335 | 4.272×10-7 | 2.003×10-7 |
| ND3 | 345 | 18 | 0.678 | 0.055 | 0.01141 | 1.336 | 4.275×10-7 | 2.004×10-7 |
| ND4 | 1377 | 31 | 0.818 | 0.041 | 0.01018 | 1.192 | 3.814×10-7 | 1.788×10-7 |
| ND4L | 294 | 14 | 0.391 | 0.066 | 0.01338 | 1.567 | 5.014×10-7 | 2.351×10-7 |
| ND5 | 1821 | 31 | 0.834 | 0.038 | 0.00935 | 1.095 | 3.504×10-7 | 1.643×10-7 |
| ND6 | 516 | 26 | 0.855 | 0.031 | 0.01452 | 1.700 | 5.44×10-7 | 2.550×10-7 |
| 12SrRNA | 955 | 17 | 0.692 | 0.051 | 0.00474 | 0.555 | 1.776×10-7 | 0.833×10-7 |
| 16SrRNA | 1573 | 24 | 0.706 | 0.054 | 0.00416 | 0,487 | 1.558×10-7 | 0.731×10-7 |
| Control region | 862 | 31 | 0.825 | 0.037 | 0.01043 | 1.221 | 3.907×10-7 | 1.832×10-7 |
| D-loop (partial)4 | 529 | 19 | 0.627 | 0.060 | 0.01053 | 1.233 | 3.946×10-7 | 1.850×10-7 |
| Small data set | 4502 | 42 | 0.886 | 0.031 | 0.00723 | 0.847 | 2.71×10-7 | 1.271×10-7 |
| Mitogenome | 16273 | 73 | 0.994 | 0.003 | 0.00854 | 1.000 | 3.2×10-7 | 1.500×10-7 |
| 1 For coding genes without the stop codon | | | |  |  |  |  |  |
| 2 Relative to the full mitogenome sequence | | | |  |  |  |  |  |
| 3 Estimated using a mitogenome mutation rate of either 3.2×10-7 or 1.5×10-7 sub/site/yr | | | | | | | |  |
| 4 Includes the partial D-loop region used for analysing mutation rate in *Apodemus argenteus* and *A. speciosus* (Suzuki *et. al.*, 2015) | | | | | | | | |

**Supplementary 5**. Information about length in bases, number of haplotypes (NH), haplotype diversity (HD), theta (θ) and the relative mutation rate (μRELATIVE) for different regions of the mitogenome. Data of the 91 full mitogenomes were used (including the 5 Korean sequences). The relative mutation rate is calculated relative to the full mitogenome while the estimated substitution rates is estimated based on a mitogenome mutation rate of either 3.2×10-7 or 1.5 ×10-7 sub/site/yr as estimated by Hardouin & Tautz (2013). For more information, see the main paper.

**References**

Hardouin E & Tautz D. 2013 Increased mitochondrial mutation frequency after an island colonization: positive selection or accumulation of slightly deleterious mutations. *Biol. Let.* **9**, 20121123. (doi: 10.1098/rsbl.2012.1123)

Suzuki Y, Tomozawa M, Koizumi Y, Tsuchiya K, Suzuki H. 2015 Estimating the molecular evolutionary rates of mitochondrial genes referring to Quaternary ice age events with inferred population expansions and dispersals in Japanese *Apodemus*. *BMC Evol. Biol.* **15**, 187. (doi: 10.1186/s12862-015-0463-5)

**Supplementary 6**. Figure showing a neighbour joining phylogenetic tree conducted on the partial mitogenome dataset in MEGA v5 using 100 bootstraps. The additional individual from Falster associated with Clade 2 in Figure 2 and discussed in the main paper can be found below (Fa81, fifth from the bottom)



**Supplementary 7a**. Probability of divergence time estimated in IMa using the partial mitogenome dataset. Black lines represents estimates unscaled by a mutation rate of 2.709×10-7 and grey dotted lines a mutation rate of 1.270×10-7 sub/site/yr. Both rates are estimated based on a full mitogenome mutation rate 3.2×10-7 and 1.5×10-7 sub/site/yr (see Suppl. 5 and main text for details)



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Supplementary 7b**. IMa results from partial mitogenome dataset. All values represent unscaled estimates calibrated from 2010. Values in brackets denote 90% High Point Density (HPD) intervals. | | | | | | | |
| mutation rate# | Pop 1 | Pop 2 | Ne population 1 | Ne population 2 | Ne ancestral population | Divergence time BP (yrs) | Divergence time calibrated from 2010 |
| 2.709×10-7 sub/site/yr | Filskov | Falster | 20612 (4556-276206) | 1953 (217-20612) | 8028 (3255-18877) | 317 (21-968) | 1693 AD (1989 AD - 1042 AD) |
| Filskov | Lolland | 21029 (6794-92594) | 22841 (7959-83665) | 7571 (2782-18700) | 551 (178-998) | 1459 AD (1832 AD - 1012 AD) |
| Falster | Lolland | 2395 (392-13109) | 20686 (8231-68765) | 7186 (2831-17899) | 500 (112-1103) | 1510 AD (1898 AD - 907 AD) |
| Germany | Falster | 53735 (19007-146657) | 2581 (704-8213) | 31209 (12437-65937) | 1454 (222-2342) | 556 AD (1788 AD - 332 BC) |
| Germany | Lolland | 63461 (31595-103478) | 11612 (5671-22954) | 22414 (9992-47798) | 1746 (959-2632) | 264 AD (1051 AD - 622 BC) |
| Germany | Filskov | 532802 (29236-107926) | 112295 (6284-23771) | 21039 (7924-50001) | 2299 (1320-2249) | 289 BC (700 AD - 239 BC) |
| Poland | Lolland | 29070 (13948-67395) | 13166 (6909-24377) | 1955 (130-64006) | 4933 (3099-7512) | 2923 BC (1089 BC - 5502 BC) |
| Poland | Falster | 31998 (11937-81349) | 3711 (903-8727) | 10332 (1304-56272) | 2831 (419-4358) | 821 BC (1591 AD - 2348 BC) |
| Poland | Filskov | 28770 (8267-114970) | 82679 (3417-17306) | 31636 (15542-66469) | 865 (321-1705) | 1145 AD (1689 AD - 305 AD) |
|  |  |  |  |  |  |  |  |
| 1.270×10-7 sub/site/yr | Filskov | Falster | 43968 (9719-589166) | 4165 (463-43968) | 17124 (6942-40265) | 676 (44-2064) | 1334 AD (1966 AD - 54 BC) |
| Filskov | Lolland | 44857 (14492-197510) | 48722 (16977-178463) | 16149 (5935-39888) | 1174 (380-2128) | 836 AD (1630 AD - 118 BC) |
| Falster | Lolland | 5109 (836-27961) | 44125 (17557-146681) | 15328 (6038-38180) | 1066 (238-2353) | 944 AD (1772 AD - 343 BC) |
| Germany | Falster | 114621 (40543-312830) | 5506 (1502-17518) | 66570 (26528-140648) | 3102 (473-4995) | 1092 BC (1537 AD - 2985 BC) |
| Germany | Lolland | 135367 (67395-220725) | 24769 (12097-48962) | 47810 (21313-101957) | 3725 (2046-5614) | 1715 BC (36 BC - 3604 BC) |
| Germany | Filskov | 113650 (62362-230214) | 26227 (13405-50705) | 44877 (16902-106656) | 4904 (2795-4798) | 2894 BC (785 BC - 2788 BC) |
| Poland | Lolland | 62008 (29753-143759) | 28084 (14737-51998) | 4171 (278-136529) | 10523 (6610-16023) | 8513 BC (4600 - 14013 BC) |
| Poland | Falster | 68254 (25462-173524) | 7917 (1926-18615) | 22038 (2782-120033) | 6039 (893-9295) | 4029 BC (1117 AD - 7285 BC) |
| Poland | Filskov | 61369 (17635-245240) | 17635 (7289-36915) | 67482 (33153-141783) | 1845 (685-3637) | 165 AD (1325 AD - 1627 BC) |
| # Mutation rates are calculated from a full mitogenome mutation rate of either 3.2×10-7 or 1.5×10-7 sub/site/yr. | | | | | | |  |