1 **Population level mitogenomics of the long-lived Greater Mouse-eared bat, *Myotis myotis*,**

2 **reveals dynamic heteroplasmy and challenges the Free Radical Theory of Ageing**

3 David Jebb[1], Nicole M. Foley[1], Conor V. Whelan[1], Frédéric Touzalin[3], Sebastien J.

4 Puechmaille[1,2], Emma C. Teeling[1]*

5 [1]School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4,

6 Ireland

7 [2]Applied Zoology and Conservation, Zoological Institute, Greifswald University, Greifswald,

8 Germany

9 [3]Laboratoire Evolution et Diversité Biologique, Université Toulouse 3, Paul Sabatier 31062

10 Toulouse Cedex 09, France.

11 *Corresponding author: emma.teeling@ucd.ie

12

13

14

15

16

17

18

19   **Abstract**

20   **Background**

21   Bats are the only mammals capable of true, powered flight, which drives an extremely high

22   metabolic rate. The "Free Radical Theory of Ageing" (FTRA) posits that a high metabolic rate

23   leads to mitochondrial heteroplasmy and the progressive ageing phenotype. Contrary to

24   this, bats are the longest lived order of mammals despite their small size and high metabolic

25   rate. To investigate if bats exhibit increased mitochondrial heteroplasmy with age as

26   predicted by the FRTA, we performed targeted, deep sequencing of mitogenomes and

27   measured point heteroplasmy in wild, long lived *Myotis myotis* as they age.

28   **Results**

29   In total, blood was sampled from 195 individuals, aged between <1 and at 6+ years old, and

30   whole mitochondria were sequenced, with a subset sampled over multiple years. The

31   majority of heteroplasmies, 77.6%, were at a frequency below 5%. Oxidative mutations

32   were not the primary source of heteroplasmies and present in only a small number of

33   individuals, likely representing local oxidative stress events. There was a significant positive

34   correlation between age and heteroplasmy, with a rate of increase of 0.13 sites per year.

35   Longitudinal data from recaptured individuals show heteroplasmy is dynamic, and does not

36   increase uniformly over time.

37   **Conclusions**

38   We show that bats do not suffer from the predicted, inevitable increase in heteroplasmy

39   which underscores the FRTA. Most heteroplasmies were at low frequency and are primarily

40   transitions. Heteroplasmy increased with age, however how this contributes to ageing is

41 unclear, as heteroplasmy was dynamic, questioning its presumed role as a primary driver of

42 ageing.

Keywords

44 Longevity, Mitochondria, Bats, Mutation, Oxidative stress

45

Background

47 The mitochondrion in mammals harbours a ~16.5kb, circular chromosome, or mitogenome.

48 Across vertebrates the mitogenome shows a conserved gene content and synteny,

49 containing 13 protein coding genes, 2 ribosomal RNA genes, 22 transfer RNA genes and a

50 control region. Protein coding genes encode essential subunits of the electron transport

51 chain (ETC) complexes, while the tRNA and rRNA genes encode components of the

52 mitochondrial translation machinery. The mitogenome is maternally inherited, is non-

53 recombining and has a higher mutation rate than the nuclear genome [1, 2]. These

54 properties have made it an attractive study locus for phylogenetics, population genetics and

55 forensics. The mitochondria have also been the focus of ageing research since the 1950s due

56 to their central role in the "Free Radical Theory of Ageing" [3].

57 The Free Radical Theory of Ageing (FRTA) states, through normal mitochondrial function,

58 reactive oxygen species (ROS) are generated which damage biomolecules leading to

59 mitochondrial dysfunction [3]. ROS are generated when electrons leak from the ETC and

60 reduce oxygen in the absence of hydrogen cations. The mitogenome was once thought to

61 be particularly susceptible to oxidative damage due to its lack of protective histones and

62 close proximity to the ETC, though it is now known that the mtDNA is not naked. Rather, the

3

63   mtDNA and the TFAM protein compact to form DNA-protein aggregates, called nucleoids,

64   which in mammals typically contain a single mitogenome [2, 4, 5]. These oxidative

65   mutations become hardcoded in the mitogenome, leading to production of mutated ETC

66   components, and mitochondrial dysfunction. Increased mitochondrial dysfunction gives rise

67   to greater ROS production, which cause greater dysfunction in a "vicious cycle". This gradual

68   accumulation of heteroplasmy and dysfunctional mitochondria is thought to give rise to the

69   familiar progressive, ageing phenotype [5–7].

70   Mitochondrial heteroplasmy is the presence of multiple non-identical mitogenomes in a

71   single individual [8]. Mitogenomes may vary in length or at single nucleotides, referred to as

72   length and point heteroplasmy respectively. Heteroplasmy plays an important role in

73   mitochondrial disease, cancer and ageing [7, 9–11]. Once thought to be extremely rare in

74   human populations, recent next generation sequencing efforts have found extensive

75   heteroplasmy in healthy individuals [12]. These heteroplasmies are predominantly at low

76   frequency, enriched for pathogenic mutations, and show tissue and allele specific patterns

77   [13–15]. As the mitochondrial genome is maternally inherited, studies of mother-child pairs

78   have shown that maternal age at conception is positively correlated with levels of

79   heteroplasmy [16]. More generally, levels of heteroplasmy have been shown to increase

80   with age in humans, and contribute to debilitating age related diseases, as well as

81   developmental disorders such as autism [10, 13, 15, 17].

82   The correlated increase in low level heteroplasmies with age in humans seems to support

83   the FRTA. However, human studies have shown a lack of oxidative transversions but instead

84   show transitions are the primary source of mtDNA mutations [13, 15, 18]. Apart from

85   humans, little is known about low level heteroplasmy in other mammals. Studies involving

4

86   mutant mouse models have been used to test the relationship between of mitochondrial

87   mutation and age. POLG[-/-] mice express a mitochondrial polymerase with no proof reading

88   activity [19, 20] and so accumulate mitochondrial mutations faster than a wild type mouse.

89   These mice show a strong progeroid phenotype, supporting a role for mitochondrial

90   mutations in ageing [20, 21]. Mammalian studies have been restricted to the control region,

91   often in repeat containing loci, and have not investigated any potential relationship with age

92   [22–29]. Moreover, few comparative studies have taken advantage of recent advances in

93   next generation sequencing technologies. Recently, Rensch et al., (2016) [1] used publically

94   available ChIP-Seq data from 16 species to evaluate patterns of heteroplasmy across the

95   vertebrate phylogeny and concluded that divergent species show similar patterns in

96   frequency and location of heteroplasmy. This study discovered relatively few

97   heteroplasmies, likely due to the high frequency cut off used, and few individuals available

98   for some species. They did not investigate change in heteroplasmy levels with age. To date,

99   next generation sequencing methods have not been used to investigate the relationship

100 between heteroplasmy and age outside of humans.

101 Chiroptera, the bats, display exceptional longevity for their body size and metabolic rate.

102 Bats are the longest lived order of mammals, living almost ten times longer than expected

103 given their body size [30, 31]. The longest lived bat is *Myotis brandtii* (weight ~7g), which

104 was first captured as an adult and recaptured 41 years later showing negligible signs of

105 senesce [32, 33]. Interestingly, bats are the only mammals capable of true powered flight

106 [34, 35] This energy intensive form of locomotion requires an extremely high metabolic rate.

107 During flight the metabolic rate in bats is up to 3 times higher than that of a terrestrial

108 mammal of the same size during exercise [36]. Bats can increase oxygen consumption 20-30

5

109   fold during flight, and *M. velifer* has been shown to increase oxygen consumption rate-130

110   fold [36, 37]. According to the FRTA, increased oxygen consumption as a result of flight

111   should lead to increased ROS production, thereby accelerating the ageing process. However,

112   bats seem to defy the FRTA [30, 38]. *M. lucifugus* has been shown to produce less ROS per

113   unit of oxygen consumed compared to mammals of a similar size, suggesting *Myotis* bats

114   have increased "mitochondrial efficiency" and have potentially evolved some mechanism to

115   avoid or manage flight-induced oxidative stress [39].

116   Heteroplasmy has previously been studied in the long lived *Myotis* genus. Petri et al., (1996)

117   [23] studied length and point heteroplasmy in the control region of the *M. myotis*

118   mitogenome. They found high levels of diversity within individuals at a tandem repeat

119   between the tRNA Proline and the conserved sequence blocks of the control region.

120   Sequence diversity at this locus within individuals was similar to diversity between colonies

121   from Germany and Portugal. The length of the repeated array also varied at this locus

122   between and within individuals, with approximately 40% of individuals showing at least two

123   different size arrays. Kerth et al.,[40] studied the homologous repeat array in *M. bechsteinii*

124   from Germany and exploited the high levels of sequence and length diversity in this array to

125   study "microgeographic" population structuring in these bats. This locus has been suggested

126   to possibly regulate mitochondrial transcription due to homology with termination-

127   associated sequences leading to a complex secondary structure [29, 41]. However, this

128   repeat locus is found only in the Vespertilioninae and Myotinae subfamilies [41, 42], and so

129   cannot apply generally to exceptional longevity across the chiropteran phylogeny.

130   Heteroplasmy has also been studied outside the Vespertilionidae. The distantly related

131   *Rhinolophus sinicus* has was found to exhibit point and length heteroplasmy. As in the

132   studies in vespertilionid bats, heteroplasmy was observed at a tandem repeat in the control

133   region using PCR and cloning based techniques [22]. So far, no study has used next

134   generation sequencing to investigate low frequency heteroplasmy across the whole

135   mitogenome in a long lived bat species, or tested if heteroplasmy shows any association

136   with age in bats.

137   To elucidate if bats exhibit increased mitochondrial heteroplasmy with age as predicted by

138   the FRTA, we performed targeted, deep sequencing of mitogenomes and measured point

139   heteroplasmy in wild, long lived *M. myotis* as they age. We report, for the first time,

140   population level characterization of heteroplasmy based on whole mitogenomes from 195

141   individuals of known age of the long lived bat, *M. myotis*. Polymerase error likely explains

142   the majority of mutations, with oxidative transversions at a lower rate and concentrated in

143   only a few individuals. Most heteroplasmies are at low frequency, less than 5%. We also

144   sampled and sequenced individuals over multiple years, to assess changes in levels of

145   heteroplasmy at the individual level through time. These unique samples revealed the

146   dynamic nature of heteroplasmy, possibly as a consequence of transient oxidative stress.

147   Contrary to the FRTA, we found that *M. myotis* do not suffer widespread oxidative damage

148   due to flight, instead exhibit acute and transient oxidative stress. We suggest enhanced

149   mitochondrial quality control mechanisms, which may repair or remove damaged

150   mitochondria, possibly contributing to the exceptional longevity in *M. myotis.*

151   **Results**

152   **Sequencing and Quality Control**

153   252 samples were sequenced obtaining a total of 72,111,278 read pairs prior to quality

154   control. After adapter trimming and stringent quality filtering of reads, 56,608,559 (78.5%)

155   of read pairs were retained. Stringent quality control (QC) resulted in loss of more than 20%

156   of our raw data. 20 samples were deemed unreliable after testing the effect of PCR

157   duplicate removal. Average coverage of samples ranged from 142 to 7800X. 4 samples with

158   average coverage below 1000X were removed. 12 which failed initial QC tests were re-

159   sequenced, and the same quality control steps as before were performed. 4 re-sequenced

160   samples were retained after quality control. A total of 232 samples were retained for further

161   downstream analysis. The average coverage of the 232 QC passed samples was 3837X,

162   ranging from 1355 to 7800X.

163   **Sample Demography**

164   As the individuals have been tagged with passive integrated transponder (PIT) tags since

165   2010, it was possible to identify which samples belonged to a unique individual, and know

166   the age for each individual. For individuals caught first as an adult only a minimum bound

167   for the age was known. The 232 QC passed samples included 88 juveniles (<1 years old), and

168   143 adults (>1 years old) and one individual to which age could not be reliably assigned

169   (Table 1). The 232 samples represent 195 unique individuals, with 21 individuals sampled

170   two, three or four times over consecutive years. All juveniles are 0 years old, with adult

171   samples ranging from 1 years old to at least 7 (hence denoted as 7+). 167 unique individuals

172   aged between 0-5 and a "6+" cohort were used to investigate the relationship between

173   heteroplasmy and age.

174

175

176

Table 1: Sample demography

Age	Number of samples, N
Juvenile Samples (0 years old)	88
Adult Samples (>1 year old)	143
1	26
2	22
3	25
4	24
5	3
6+	13
1+,2+,3+,4+,5+,7+	30
No age assigned	1

Table 1: Number of samples assigned to each age cohort

177

178

Table 2: Analysed Datasets

Dataset Title	Description	Number of samples, N
Sequenced Samples	Total number of samples sequenced	252
QC Passed	Samples which passed quality control	232
Unique Individuals	Unique individual bats, recaptures removed	195
Primary	Unique individuals assigned to age cohorts used in model fitting	167
Recaptures	21 individuals caught and sequenced at least twice between 2013 and 2016	56

Table 2: Description of and number of samples in each datasets used for downstream analyses

9

179 **Simulation Results**

180 9 read sets were generated using GemSIM [43], ranging from 50 to 25,000X coverage. Each

181 read set was analysed using one of the three callers LoFreq, VarScan or FreeBayes,

182 producing 27 variant call sets, which were given a score between 0 and 1 defined as

183 (Power*Accuracy)*(1-False Positive Rate) (described in full in Material and Methods, Variant

184 Simulation). LoFreq was the best scoring caller, as shown in Supp. Figure 1, and used to call

185 heteroplasmies in all empirical datasets. FreeBayes consistently had the highest power,

186 detecting more than 85% of true variants even at coverages as low 100X, however,

187 FreeBayes also exhibited a false positive rate >5% at all coverages. LoFreq and VarScan had

188 extremely low error rates, with each detecting only one false positive in the 27 variant call

189 sets. LoFreq had higher power and accuracy than VarScan at all coverage levels (see

190 Additional File 1).

191 **Frequency and distribution of heteroplasmies in *M. myotis***

192 From 232 samples, 254 heteroplasmies, across 143 sites in mitogenome, were discovered

193 with a minor allele frequency (MAF) greater than 1%, Figure 1A. The frequency of minor

194 alleles was strongly skewed toward low frequency variants. Figure 1B shows a histogram of

195 number of heteroplasmies by MAF bin for all 254 heteroplasmies. The vast majority, 77.6%,

196 of heteroplasmies are below 5% MAF.

197 210 heteroplasmies were collected from unique individuals, using the most recent sample

198 from recaptured individuals. The difference between the frequency distribution of coding

199 and non-coding heteroplasmies was approaching significance (Kolmogorov-Smirnov (KS)

200 Test, p=0.054) with non-coding heteroplasmies tending to be at lower frequencies. Closer

201 investigation of coding heteroplasmies found no significant difference in the frequency

10

202   distribution of nonsynonymous and synonymous heteroplasmies (KS Test, 0.082), however

203   nonsynonymous heteroplasmies were at lower frequencies than synonymous (Mann-

204   Whitney U-test, p=0.021). This also drove the difference observed between coding and non-

205   coding heteroplasmies, as a significant difference was found between the frequency

206   distributions of synonymous and non-coding heteroplasmies, but not between

207   nonsynonymous and non-coding heteroplasmies (KS Test, p=0.006 and p=0.374

208   respectively) (Supp. Figure 2A). There was also a difference in the frequency of

209   heteroplasmies depending on the codon position with heteroplasmies at the second codon

210   position at a lower frequency than position 1 or 3, which showed no difference to each

211   other (KS Test, p=0.004, p=0.005, p=0.411 respectively) (Supp. Figure 2B). To further explore

212   the effect of heteroplasmies, all nonsynonymous heteroplasmies, excluding nonsense

213   mutations, were scored as "Deleterious" or "Neutral". Of the 62 nonsynonymous sites

214   scored, 42 were found to be deleterious and 20 neutral. Deleterious heteroplasmies were at

215   significantly lower frequency than neutral heteroplasmies (KS Test, p=0.012).

216   Heteroplasmic sites were spread relatively evenly across the mitogenome. The number of

217   heteroplasmic *versus* homoplasmic sites were used to construct a contingency table for

218   each mitochondrial partition. Dividing the mitogenome into tRNA, rRNA, protein coding and

219   non-coding regions, showed enrichment for heteroplasmic sites in tRNA genes (p=0.0005,

220   Chi Square Test). Analysis of individual genes revealed *ND2*, *trnaE* and *trnaN* genes were

221   enriched for heteroplasmic sites (p<0.05, Chi Square Test). The majority of sites, 72.7%,

222   were only observed as heteroplasmic in one individual (Figure 1A), hereafter called "private"

223   heteroplasmies as opposed to "shared". Private and shared heteroplasmies showed

224   significantly different frequency spectrums (KS Test, p=0.005) with private heteroplasmies at

225   lower average frequency suggesting these were predominantly somatic mutations. Private

226   and shared heteroplasmies showed similar proportions of synonymous and nonsynonymous

227   mutations and no bias toward coding versus non-coding sites (Chi Square test, p>0.1 in all

228   comparisons).

229   **Transitions are the primary source of heteroplasmies**

230   210 heteroplasmies from 195 unique individuals were classified into 4 classes, 1 transition

231   class (G:C↔A:T) and 3 transversion classes (G:C↔T:A, A:T↔T:A and G:C↔C:G), as shown

232   in Figure 1C. Transitions were the primary class of heteroplasmy at 60.6%. This was followed

233   by G:C↔T:A transversions at 31.9%, which are characteristic of oxidative stress events. The

234   remaining 7.5% of mutations were A:T↔T:A transversions. *M. myotis* showed a small

235   transition to transversion (Ts/Tv) ratio, with a Ts/Tv of 1.8. Only 38 samples (16.4%) showed

236   evidence of oxidative transversions, of which 33 were unique individuals. The 5 samples

237   containing the highest number of oxidative transversions accounted for 37% of these

238   transversions (30 of 81).

239   **Correlation between heteroplasmy and age**

240   To test if there was an association with age we fitted a generalised linear model with age as

241   the only explanatory variable, and the count of heteroplasmic sites as the negative binomial

242   distributed response variable. Two samples within the 6+ cohort were found to be unduly

243   influential points using Cook's Statistic (Supp. Figure 3A and B). As such they were removed

244   and the model fitted again. With these two samples removed no significant association with

245   age was found (ANOVA, p=0.1), as shown in Figure 2. Further, these two samples contained

246   primarily oxidative mutations, one of which contains the highest number of oxidative

247   mutations in our data. To investigate any bias that may be introduced by choosing the most

12

248   recent sample from recaptured individuals, we produced 1000 permutations in which the

249   sample from recaptured individuals of usable age was chosen at random, to produce a

250   similar dataset of 165 unique individuals. 19.8% of samplings found a significant association

251   between age and number of heteroplasmies (Supp. Figure 4A), with an average slope 0.0969

252   and ranging from 0.0815 to 0.1237 (Supp. Figure 4B).

253   **Longitudinal analysis of heteroplasmy**

254   From the 27 individuals sampled more than once, we were able to obtain samples from 21

255   individuals from at least two time-points, which passed quality control. 3 individuals were

256   sampled at 4 time-points, 8 individuals sampled at 3 time-points and 10 individuals at 2

257   time-points, individuals sampled at 3 or 4 time-points are shown in Figure 3. The mean

258   change in number of heteroplasmies between two consecutive years across individuals was

259   0.303. Of the 33 consecutive intervals (e.g. 2013 to 2014 or 2014 to 2015 etc.) 15 showed no

260   change, 10 showed a decrease and 8 showed an increase (Supp. Figure 5). Not all

261   heteroplasmies were shared between time-points. 14.7% of heteroplasmies were observed

262   in at least two time-points, with 13% shared between all time-points for an individual.

263   Shared sites ranged in frequency from 43.36% to as low as 1.02%. Some individuals were

264   extremely consistent over all time points. For example individual 000715B9C8, shown in

265   Figure 3, initially showed 2 heteroplasmies, but subsequently lost one. The remaining

266   heteroplasmy was at a low frequency but was found at all 4 time-points and had an

267   extremely consistent frequency, range$_{maf}$ 1.02 – 1.83%. However some individuals were

268   more dynamic. These "spikes" in the number of heteroplasmy was normally followed by a

269   marked decrease the following year. Two bats, 000702E6B0 and 000702FF51 (also shown in

270   Figure 3), showed spikes and subsequent loss, though they retained a subset of the

13

271   mutations acquired during the spike. Four individuals showing spikes in the number of

272   heteroplasmy, with an increase of 4 or more sites between two subsequent years, were

273   previously found to have some of the highest levels of oxidative mutations. Interestingly,

274   bat 000702EA84 showed a spike in heteroplasmy with little recovery the following year.

275   Instead, this bat may have experienced oxidative stress events during two consecutive field

276   seasons.

Discussion

278   Comparative genomics has emerged as a powerful new tool in ageing research [30, 44].

279   However, though mitochondrial dysfunction and heteroplasmy have been long implicated in

280   ageing, there remains a paucity of data from non-human or non-model organisms to

281   associate increased heteroplasmy as a mechanism underpinning the progressive, ageing

282   phenotype. To this end, for the first time we have deep sequenced, to an average depth of

283   >3500X, whole mitogenomes from 195 *M. myotis* individuals, and investigated the

284   association between heteroplasmy and age in these long-lived bats. After outlier removal,

285   we found no evidence for an increase in heteroplasmy with age. We also found little to no

286   evidence to support the Free Radical Theory of Ageing, finding no chronic increase in

287   oxidative mutations with age. Through unique, longitudinal sampling we found bats may

288   experience local oxidative stress events followed by removal of the majority of mutations.

289   Nonsynonymous, protein coding mutations were also at a significantly lower frequency than

290   synonymous, and those with predicted deleterious effects were lower again. Together our

291   results suggest *M. myotis* can remove deleterious mitochondrial DNA mutations, preventing

292   their expansion, to maintain mitochondrial homeostasis, and possibly promoting longevity.

293   The FRTA posits oxidative mutations in the mitogenome should drive mitochondrial

294   dysfunction [3]. However, many studies have shown a dearth of oxidative mutations in

295   mitochondrial DNA, which has led some to question this central tenet of the FRTA [45]. In

296   *M. myotis,* transitions, not oxidative transversions, are the primary source of mutation.

297   Transitions are likely due to errors during DNA replication or spontaneous deamination of

298   nucleosides, rather than damage accrued by oxidative stress [4, 7]. Still, the proportion of

299   heteroplasmies owing to oxidative transversions in *M. myotis* mitochondrial DNA is much

300   higher than reported in humans [13, 15]. However, the majority of these oxidative

301   mutations were concentrated in only a few individuals. These likely represent local oxidative

302   stress events, rather than a population wide phenomenon. The longitudinal data supports

303   this interpretation as two individuals exhibiting high numbers of oxidative transversions,

304   and some of the highest levels of heteroplasmy, were also sampled the following year

305   (Individual 000702E6B0 and 000702FF51, Figure 3). They had cleared or repaired the

306   majority of the oxidative damage, though each retained at least one of the heteroplasmies

307   from the previous year. We suggest *M. myotis* experience local, transient oxidative stress

308   followed by return to homeostasis and removal or repair of oxidative lesions.

309   As oxidative damage is concentrated, and transient, within a small proportion of individuals,

310   it is unlikely to be the ultimate driver of ageing, as posited in the FRTA. It also seems unlikely

311   that flight is the cause of this oxidative stress, as it should then be ubiquitous across the

312   population rather than concentrated in few individuals. There are multiple other

313   explanations for the transient oxidative damage observed in only some individuals, including

314   infection and recent reproductive effort [46–48].

315   In recent years, the role of ROS as essential signalling molecules in immune cells has become

316   clear. *In vitro*, macrophages recruit mitochondria to the phagosome where promotion of

317   H_2O_2production contributes to the control of *Salmonella* infection [48]. Activation of the

318   NLRP3 inflammasome has also been shown to promote ROS production [47]. It has been

319   suggested that bats have a constitutively active innate immune system, possibly explaining

320   their unique ability to act as reservoirs for diverse viruses, such as SARS, which can be

321   pathogenic in humans and other mammals [49–52]. The constitutive expression of

322   interferon or the presence of circulating viruses in bats may lead to acute levels of ROS

323   production in bats enabling them to quickly mediate infection [53, 54]. The long lived bat

324   species, *Corollia perspicillata*, has also been shown to experience increased oxidative stress

325   during simulated bacterial infection [55]. Potentially, individuals showing increased

326   oxidative damage were experiencing immune related oxidative stress at the time of

327   sampling, although this remains to be tested. This may also suggest that spikes in oxidative

328   transversions are also a blood specific phenomenon; future studies incorporating different

329   tissues may shed light on the role and effect of ROS in the bat immune system.

330   A second possibility is reproductive effort. It has been shown that reproductive effort

331   increases oxidative stress in free living animals, though not unequivocally [46]. As the *M.*

332   *myotis* sampled over the course of the study were caught at maternity roosts, most females

333   will have recently given birth to a pup, and may still be lactating. Under the disposable soma

334   theory of ageing, reallocation of resources from self-maintenance to reproductive effort

335   leads to increased damage of somatic tissue during reproductive phases, with unrepaired

336   damage accumulating with time [46, 56]. The spikes in oxidative damage followed by

337   recovery as observed in our longitudinal sampling, showed while most oxidative damage

338   observed during the spike were eventually lost or removed, some mutations persisted, and

339   could be detected the following year. Further longitudinal sampling, or sampling at different

340   life history phases may help elucidate the cause of oxidative stress in these bats.

341   The relationship between age and heteroplasmy was unclear in bats. A positive correlation

342   with age seen in our data was driven by two samples. These samples were clear outliers

343   with a large effect on our regression. Removal of these samples also removed the

344   association with age. The outliers were both in the oldest cohort and had two of the highest

345   levels of heteroplasmy, primarily due to oxidative mutations. It seems likely these

346   individuals may have been experiencing acute oxidative stress, and may have removed the

347   majority of these oxidative mutations in following years, as seen in other individuals.

348   However, this remains speculative as we have yet to re-sample these two individuals.

349   Further, the observed dynamic nature of heteroplasmy questions the direct contribution of

350   heteroplasmy to ageing in this species. However, as our study population has only been

351   tagged since 2010 we have only sampled a small portion of the life span of these bats, which

352   can live up to 37.1 years. It is possible accumulation of heteroplasmy is not detectable over

353   this time scale. However, if bats aged at a rate predicted by their body size we would expect

354   to see accumulation of heteroplasmy even over this time scale.

355   The question remains, how do bats maintain their efficient mitochondria, and how might

356   this lead to an increased lifespan? Our data shows no support for flight induced oxidative

357   damage accumulating in *M. myotis*. However, oxidative stress is observed in numerous

358   individuals, and those with samples from subsequent time-points show removal of the

359   majority of mutations, suggesting mitochondrial stress followed by repair or removal of

360   damaged mitochondria. Further, nonsynonymous mutations were at significantly lower

17

361   frequency than synonymous, suggesting purifying selection acting on these heteroplasmies.

362   Nonsynonymous sites predicted to be deleterious were even lower in frequency again than

363   those with little or no functional impact. This suggests *M. myotis* can prevent expansion of

364   nonsynonymous mutations, which may contribute to the maintenance of mitochondrial

365   function with age.

366   There are multiple mechanisms by which *M. myotis* may remove and prevent expansion of

367   mitochondrial mutations. One such mechanism is through direct repair of DNA damage.

368   DNA repair genes and pathways have been shown to be evolving or expressed differently in

369   bats compared to other mammals [57–59], however, many of these pathways are not active

370   in the mitochondria. Base excision repair (BER) is the primary method of DNA repair in

371   mitochondria. While double strand break repair operates in mitochondria, the mechanisms

372   have not been fully characterised [4]. Another mechanism by which *M. myotis* may remove

373   oxidative damage is through mitophagy. Mitophagy is a form of autophagy through which

374   cells remove dysfunctional mitochondria. Inhibition of autophagy, and so mitophagy, by

375   knockdown of autophagosome genes has also been shown to lead to increased ROS

376   production [45]. Mitochondria derived ROS have also been shown to promote autophagy

377   during starvation, through oxidation and inactivation of the ATG4 protein [60]. Long lived

378   bats have also been shown to have increased macroautophagy activity in comparison to a

379   short lived, phylogenetically close sister taxa. Interestingly, other homeostatic mechanisms

380   such as chaperone protein levels and proteasome activity, showed no difference in long and

381   short lived bats [61]. This suggests long-lived bats rely on autophagy for maintaining

382   homeostasis. Indeed, previous studies have shown that autophagy genes are upregulated in

383   the blood of *M. myotis* compared to other mammals [59].

18

384   Potentially flight may induce selection for efficient mitochondria in bats [62]. *M. lucifugus*

385   juveniles produced more H_2O_2 than adults across three tissues; however, this difference was

386   most striking in pre-fledged juveniles. Cloning of the Hyper Variable Region I of the *M.*

387   *lucifugus* D-loop found less variation in adults compared to juveniles, suggesting

388   convergence on a single mitotype within individuals with age [62]. This may be due to intra-

389   individual selection for efficient mitotypes due to the high energy demands of flight [62].

390   Endurance training has also been shown to induce mitochondrial health in mutator mice

391   completely rescuing the progeroid phenotype. Mice experiencing an endurance exercise

392   regime showed wild type lifespan, increased aerobic capacity, and indirect evidence for

393   decreased mtDNA mutations [19]. Coupled with our results, this suggests bats may possess

394   stringent mitochondrial quality control to prevent accumulation and expansion of

395   deleterious mutations, maintain mitochondrial function leading to their extraordinary

396   longevity.

397   **Conclusions**

398   Overall, we found little support for the long standing "Free Radical Theory of Ageing" in the

399   exceptionally long-lived bat, *M. myotis*. Instead, oxidative mutations were not the primary

400   source of mutations and most seemed to be transient, arising from local oxidative stress

401   events before disappearing. We did not find that accumulation of heteroplasmic sites in the

402   mitogenome increases with age in these bats, however, we have only analysed a small

403   portion of the lifespan of this species. Unique longitudinal sampling revealed the dynamic

404   nature of heteroplasmy, with some bats gaining or losing up to 6 heteroplasmies between

405   consecutive years. These drastic shifts in heteroplasmy suggest increases in heteroplasmy

406   may not only be tolerable but removable in *M. myotis*. We propose that stringent

19

407 mitochondrial quality control mechanisms and intra-individual selection may drive

408 mitochondrial health potentially contributing to the exceptional longevity of *M. myotis*.

409 **Methods**

410 **Capture and Sampling**

411 All procedures were carried out in accordance with the ethical guidelines and permits

412 delivered by University College Dublin and the Préfet du Morbihan respectively. *M. myotis*

413 were sampled in western France in 2013, 2014, 2015 and 2016 (detailed in Table 3) as

414 described in Huang et al (2016).

Table 3: Colony location and sample numbers				
Colony	**Latitude**	**Longitude**	**No. Samples**	
Béganne	N47°35'	W2°14'	112	
Noyal-Muzillac	N47°35	W2°27	20	417
Férel	N47°28'	W2°20'	78	418
La Roche-Bernard	N47°31'	W2°17'	32	419
Limerzel	N47°38'	W2°21'	10	420

415

416

Table 3: Latitude and longitude of 5 sampling colonies, and the total number of sequenced samples originating from each.

421 **DNA Extraction**

422 Prior to DNA extraction all blood samples had RNA extracted using the RNAzol BD for blood

423 kit (catalogue number RB 192, Molecular Research Centre, Inc.) using the manufacturers

424 protocol with minor modification as per [63]. Post extraction of the RNA containing layer,

425 the phenol phase/interphase was placed at -80℃. DNA was extracted from the phenol

426 phase/interphase using the protocol detailed in the RNAzol BD kit manual with minor

427 modifications. Briefly, for each original volume of blood, the DNA extraction buffer was

428 prepared with 1 volume of 4 M guanidine thiocynate solution, 0.1 volumes of 3 M sodium

429 hydroxide and 0.005 volumes of polyacryl carrier. Samples were allowed to thaw in the DNA

430 extraction buffer and then were vortexed vigorously for 30 s. Samples were incubated at

431 room temperature (RT) for 10 minutes, then vortexed for 30s. 0.2 volumes of chloroform

432 was added to each sample, briefly vortexed, and centrifuged at 12000rpm for 20 mins at RT.

433 The upper fraction was extracted, and extracted once more with an equal volume of

434 chloroform where necessary, to produce a clear, aqueous DNA phase. DNA was precipitated

435 overnight at -20℃ in an equal volume of isopropanol. DNA was pelleted by centrifugation at

436 12000rpm for 20 mins at 4℃. DNA pellets were washed twice with 75% ethanol and then

437 resuspended in 30µl nuclease free water.

438 **mtDNA Enrichment and Sequencing**

439 The whole mitogenome of *M. myotis* was amplified and sequenced using the primers, and

440 sequencing protocols previously outlined in Jebb et al., (2017)[64]. Briefly, samples were

441 enriched for mitochondrial sequences using long range PCR, in two overlapping ~10kb

442 fragments. Both amplicons were successfully amplified for 252 samples. The two amplicons

443 for each of the 252 samples were purified by vacuum filtration (MilliporeHTS Filtration Plate,

444 cat. no. MAVM0960R), quantified using a nanodrop spectrophotometer and then pooled in

445 equimolar amounts. Pooled amplicons were used to produce sequencing libraries using the

446 Nextera XT preparation kit as per the manufacturer's instructions. Varying numbers of

447 samples were multiplexed and loaded onto an Illumina MiSeq.

448 **Quality Control**

449   Raw reads were initially processed using a stringent quality control pipeline prior to variant

450   calling. Samples were trimmed of Illumina and Nextera adapter sequences using Cutadapt

451   (v1.8.3) [65] and The NGS QC Toolkit [66] was used to filter out reads for which 80% of bases

452   had quality scores less than Q30. Due to extreme sequence depths at small target regions in

453   resequencing studies, it is possible that inserts will begin and end at the same position due

454   to chance, called sampling coincidence. Reads from such inserts will be falsely identified as

455   PCR duplicates. However, not removing true PCR duplicates may affect the accuracy of

456   variant calling. In order to remove any potential bias introduced by duplicate removal or

457   retention, we ran the heteroplasmy detection pipeline with duplicates present and

458   removed. As some samples showed extremely high levels of variation we used the

459   interquartile range to identify and remove outliers. An outlier was defined as any value for

460   which the count value was 2 interquartile ranges below the first quartile or above the third

461   quartile. The mean and standard deviation of these filtered counts was then calculated. A

462   cut-off value of two standard deviations from the mean was rounded down to the nearest

463   integer value. Samples with a difference in counts between the two treatments, which was

464   greater than this cut-off were deemed unreliable and removed from downstream analysis.

465   Finally, average coverage was calculated for each sample using the depth command in

466   samtools (v0.1.19) [67]. As all variant callers exhibited power below 80% at coverages below

467   1000X (Fig. 2), samples with an average coverage below this threshold were removed.

468   **Heteroplasmy Detection Pipeline**

469   The first and last 500 bp of the mitogenome were copied to the opposite ends, to extend

470   the reference and account for circularity of the mitogenome. Reads were mapped against

471   the extended *M. myotis* mitogenome using the BWA-MEM [68] algorithm. Prior to variant

22

472   calling BAM files were processed using Picard Tools (v1.90) [69] and the Genome Analysis

473   Tool Kit (GATK) [70]. Briefly, SAM files were sorted and read groups added using Picard

474   Tools. SAM files were then converted to BAM files and indexed. Prior to base quality

475   recalibration a set of raw variants was produced using the chosen low frequency variant

476   caller, either LoFreq Star [71] or VarScan 2 [72]. These raw variants were used as the

477   "known sites" necessary for the BQSR walker in the GATK. As no databases of variants exist

478   for most non-model organisms, these raw variants provide an *ad hoc* solution. Final

479   heteroplasmies were called on the recalibrated data again using the caller of choice. A third

480   variant caller, FreeBayes [73], does not require recalibrated or realigned data, and so these

481   steps were omitted when FreeBayes was used. For the empirical data, variants with a minor

482   allele frequency greater than 1% and at sites with greater than 1000X coverage were

483   retained. These variants are then annotated, removing variants in primer binding sites and

484   the two Myotinae repeats in the control region. An illustration of the bioinformatic

485   workflow for a single sample from quality processing through to annotated variants is

486   depicted in Supp. Figure 6. GNU Parallel (v20170122) [74] was used to parallelise the

487   mapping and processing of SAM and BAM files, markedly decreasing computational time.

488   The number of samples to be run in parallel can be defined by the user.

489   **Variant Simulation**

490   In order to gauge the sensitivity, accuracy and false positive rate of our bioinformatic

491   pipeline, known variants and Illumina sequence data were generated *in-silico* using the

492   GemSIM package (v 1.5) [43]. Sequence reads in FASTQ format from a reference individual

493   (MMY104, used to generate the reference mitogenome for this species) were mapped to

494   the previously published *M. myotis* mitogenome [64] using the BWA-MEM algorithm,

495   producing a SAM file which was used to produce a general error model using the GemErr.py

496   module. The reference mitogenome was used to generate haplotype information for variant

497   simulation with the GemHap.py module. In total 500 known variants were simulated, in sets

498   of 100 at frequencies of 1%, 4%, 7%, 10% and 15%. The reference mitogenome, error profile

499   and haplotype information were used to generate 250bp paired end Illumina reads.

500   Datasets were simulated at 50X, 100X, 250X, 500X, 1000X, 2500X, 5000X, 10,000X and

501   25,000X coverage. Predicted variant positions were checked against the known set. Power

502   (proportion of true sites called), false positive rate (proportion of erroneous sites called) and

503   accuracy (proportion of true sites with predicted frequency within ±0.01 of the true

504   frequency) were calculated on each dataset for each caller, FreeBayes, LoFreq and VarScan2

505   (see Heteroplasmy Detection Pipeline). Finally a score between 0 and 1 was assigned to

506   each caller defined as (Power*Accuracy)*(1-False Positive Rate). All variants called from real

507   samples used the best performing caller.

508   **Characterising heteroplasmy in *M. myotis***

509   All filtered and annotated heteroplasmies were collected and manually inspected.

510   Heteroplasmies were binned into several overlapping classes: protein-coding, synonymous,

511   nonsynonymous, non-protein-coding, tRNA, rRNA, non-coding, and by each mitogenome

512   feature. Nonsynonymous mutations were further binned for possible effect as "Deleterious"

513   or "Neutral" based on the PROVEAN algorithm as implemented on the dedicated web

514   server. Default cut-offs were used such that a mutation was binned as "Deleterious" if the

515   computed PROVEAN Score was greater than 2.5 or less than -2.5 [75, 76]. Frequency

516   distributions for each bin were compared using a two tailed non-parametric Kolmogorov-

517   Smirnov Test. The number of unique heteroplasmic sites in a bin was tested for enrichment

518 against the remainder of mitogenome sites analysed (length of mitogenome – primer

519 regions – number of sites cover by bin), using Pearson's Chi-Square test.

520 **Correlation between heteroplasmy and age**

521 We constructed a primary dataset of n=167 unique individuals of known age and an oldest

522 cohort, divided into cohorts from 0 to 6+, described in Table 1 and 2, taking the most recent

523 sample of an individual had the bat been sampled multiple times. This dataset was used to

524 test for any significant correlation between number of heteroplasmic sites and age by fitting

525 a generalised liner model (GLM) in R (v3.3.3) using the "MASS" package [77] with age as the

526 sole explanatory variable and number of heteroplasmic sites as the response variable

527 modelled under a negative binomial distribution. A negative binomial distribution was used

528 due the presence of overdispersion in the data making a Poisson distribution unsuitable.

529 The "boot" package [78, 79] in R (v3.3.3) was used to diagnose samples which may be

530 unduly influential points using the Cook Statistic [80]. Influential points were removed and

531 the model was fitted again. To investigate any bias that may be introduced by manually

532 selecting a sample from recaptured individuals, we performed 1000 samplings in which the

533 sample from recaptured individuals of usable age was chosen at random, to produce 1000

534 datasets (of n=165).

535 **Longitudinal Analysis**

536 *M. myotis* show natal philopatry, returning to the same roost or broad area to give birth

537 each year. As such, it was possible to recapture the same individuals almost every year. 27

538 individuals with samples from at least two time-points were sequenced; samples that

539 passed quality control were retained, for a final set of 56 samples from 21 individuals

540 (Tables 2 and 4). We calculated the mean change between any two consecutive years over

541   all individuals, excluding 2 which were sampled non-consecutively. We also calculated the

542   proportion of shared heteroplasmies between any two time-points, and those shared

543   between all time-points.

Table 4: Recaptured individuals and associated samples

Number of times sampled	Number of Individuals	Number of Samples
4	3	12
3	8	24
2	10	20

Table 4: Number of individuals sampled 4, 3 or 2 times between 2013 and 2016. The final number of samples shown is for QC passed samples. Only these samples were used in recapture analysis.

544

Declarations

Ethics approval and consent to participate

547   All procedures were carried out in accordance with ethics approval issued by University

548   College Dublin to Prof. Emma Teeling (AERC_1338_Teeling) and in accordance with Arrêté

549   préfectoral permits issued by the Préfet de Morbihan to Frédéric Touzalin and Sébastien J.

550   Puechmaille.

Consent for publication

552   Not applicable

Availability of data and material

554   Annotated variants and sample information used for statistical analyses are available within

555   Additional File 1. The scripts and binaries for all dependencies needed to run the 3ML

556   pipeline are available at https://github.com/jebbd/3ML. Sequence data generated during

557   the course of the project from real and simulated data are available from the authors on

558   request.

Competing interests

560   The authors declare no competing interests.

Authors' contributions

566   E.C.T and D.J devised the study. All authors participated in the capture and sampling of *M.*

567   *myotis* bats. D.J and C.V.W extracted DNA and prepared samples for sequencing. All

568   sequence data was analysed by D.J. D.J was responsible for statistical analyses. D.J is

569   responsible for the figures presented throughout. The manuscript was written by E.C.T. and

570   D.J. with input from all authors.

27

Figure 1: Characteristics of heteroplasmy in *M. myotis*. A) A circus plot of the mitogenome of *M. myotis*. In the outer ring histogram, each bar is one of the 143 unique sites which were heteroplasmic in the population, with the height representing the number of individuals sharing the site. The middle ring is a schematic of the mitogenome generated with OGDRAW [81]. The inner most ring is a heat plot showing the average minor allele frequency at each heteroplasmic site, with low frequencies in red and high frequencies in green. **B)** Histogram of minor allele frequencies of 254 heteroplasmies discovered. The vast majority of heteroplasmies were at a frequency below 5%. **C)** Bar chart showing the proportion of mutations attributable to one of four transition or transversion classes. The transition transversion ratio (Ts/Tv) for *M. myotis* was 1.8. Oxidative mutations, G:C⇔T:A transversions, were at a higher level than reported previously in humans.

Figure 2: Heteroplasmy is not significantly associated with age in *Myotis myotis.* Boxplots of the distribution of heteroplasmy counts in each age cohort used for age analysis. No significant association was found between age and heteroplasmy using a negative binomial generalized linear model, after the removal unduly influential points.

Figure 3: Plots of heteroplasmy count against year sampled for 11 individuals. Heteroplasmy counts through time for 11 individual bats which were sampled 3 or 4 times. Numbers above each plot indicate PIT number used to identify each individual. Some individuals were extremely consistent while others increased sharply, always followed by a decrease where subsequent samples were available.

Supplementary Figure 1: Results from simulated datasets using different variant callers. Nine read sets were generated *in silico* using GemSIM and the *M. myotis* mitochondrial genome, containing 500 known variants. The heteroplasmy detection pipeline was run on

each dataset three times, using a different variant caller, with or without INDEL realignment and base quality recalibration as appropriate. The 27 resulting variant sets were compared to the known set for power, accuracy and false positive rate. A score was given to each set as (Power*Accuracy)*(1-False Positive Rate), and plotted against the expected coverage for the set. LoFreq was the best performing caller and was used for all variant calling on real data.

Supplementary Figure 2: Frequency distributions and comparisons for different classes of heteroplasmies. A) Density plots of minor allele frequencies for noncoding, nonsynonymous coding and synonymous coding mutations. **B)** Density plots of minor allele frequencies for heteroplasmies at the first, second and third codon position.

Supplementary Figure 3: Cook's Statistic predicting influential points in the Primary dataset. A) A plot of Cook's statistics versus the standardised leverages for each point. The horizontal line is at 8/(n-2p) and the vertical line is at 2p/(n-2p) where n is the number of observations and p is the number of parameters estimated. Points above the horizontal line have a high influence on the model, and those to the right of the vertical line have a high leverage. The two outliers previously shown in Figure 4 are again indicated in red, with both above the horizontal and right of the vertical lines. **B)** A plot of the Cook's statistics for each observation, order as they appear in the dataset. Again the two highly influential observations are coloured in red.

Supplementary Figure 4: Test for the effect of sample selection on model fitting. Recaptured individuals of known age had more than one sample available for use in regression analyses. In the primary dataset the most recent, and thus oldest, sample from each recaptured was chosen. To ensure no bias was introduced by this selection, 1000

29

datasets were generated, randomly choosing which sample to use from the recaptured individuals, and then fitting the negative binomial model as before and estimating the significance and size of the age as the sole fixed effect. **A)** shows a histogram of p-values from 1000 models. 19.8% of models predicted a significant association between age and heteroplasmy **B)** shows a density plot of the slopes estimated from the 198 significant models, with a slope of 0 indicated by the red dotted line. The mean estimate was 0.0969 sites per year.

Supplementary Figure 5: Change in the number of heteroplasmy between two consecutive years. Bar plot showing the number of times a change in heteroplasmy was observed, with respect to the size of the change, between two consecutive years. The mean change between any two years was 0.32, with median value of 0, though there were multiple observations of gains/loss of 5 or more heteroplasmies over a year.

Supplementary Figure 6: Overview of Heteroplasmy Detection Pipeline. Diagram depicting the bioinformatic workflow used in this project. Processes are depicted as rectangles, those in blue are mandatory, while those in grey are optional within the framework of the pipeline (duplicate removal and variant annotation). Documents are represented by rectangles with a wavy base. Those in green are input provided by the user (reference genome, annotations and sequence data), while the yellow documents are generated by default during analyses (extended reference, high quality reads, and variant calls) and grey documents can be generated but are optional (annotated variants). The pipeline can be parallelised, concurrently running a user defined number of jobs.

Bibliography

1. Rensch T, Villar D, Horvath J, Odom DT, Flicek P. Mitochondrial heteroplasmy in vertebrates using ChIP-sequencing data. Genome Biol. 2016;17:139.

2. Gustafsson CM, Falkenberg M, Larsson N-G. Maintenance and Expression of Mammalian Mitochondrial DNA. Annu Rev Biochem. 2016;85 March:133–60.

3. Harman D. Aging: a theory based on free radical and radiation chemistry. J Gerontol. 1956;11:298–300.

4. Alexeyev M, Shokolenko I, Wilson G, LeDoux S. The Maintenance of Mitochondrial DNA Integrity - Critical Analysis and Update. Cold Spring Harb Perspect Biol. 2013;5:a012641--a012641.

5. Lagouge M, Larsson N-G. The role of mitochondrial DNA mutations and free radicals in disease and ageing. J Intern Med. 2013;273:529–43.

6. Barja G. Free radicals and aging. Trends Neurosci. 2004;27:595–600.

7. Larsson NG. Somatic mitochondrial DNA mutations in mammalian aging. Annu Rev Biochem. 2010;79:683–706.

8. Lane N. The problem with mixing mitochondria. Cell. 2012;151:246–8.

9. Casoli T, Spazzafumo L, Di Stefano G, Conti F. Role of diffuse low-level heteroplasmy of mitochondrial DNA in Alzheimer's disease neurodegeneration. Front Aging Neurosci. 2015;7 July:142.

10. Wang Y, Picard M, Gu Z. Genetic Evidence for Elevated Pathogenicity of Mitochondrial

DNA Heteroplasmy in Autism Spectrum Disorder. PLoS Genet. 2016;12:1–24.

11. He Y, Wu J, Dressman DC, Iacobuzio-donahue C, Sanford D, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumor cells. Nature. 2010;464:610–4.

12. Payne BAI, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, et al. Universal heteroplasmy of human mitochondrial DNA. Hum Mol Genet. 2013;22:384–90.

13. Li M, Schröder R, Ni S, Madea B, Stoneking M. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. Proc Natl Acad Sci U S A. 2015;112:2491–6.

14. Ye K, Lu J, Ma F, Keinan A, Gu Z. Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. Proc Natl Acad Sci U S A. 2014;111:10654–9.

15. Ding J, Sidore C, Butler TJ, Wing MK, Qian Y, Meirelles O, et al. Assessing Mitochondrial DNA Variation and Copy Number in Lymphocytes of ~2,000 Sardinians Using Tailored Sequencing Analysis Tools. PLoS Genet. 2015;11:e1005306.

16. Rebolledo-Jaramillo B, Su MS-W, Stoler N, McElhoe J a, Dickins B, Blankenberg D, et al. Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A. 2014;111:15474–9.

17. Keogh MJ, Chinnery PF. Mitochondrial DNA mutations in neurodegeneration. Biochim Biophys Acta (BBA)-Bioenergetics. 2015;1847:1401–11.

18. Kauppila JHK, Stewart JB. Mitochondrial DNA: Radically free of free-radical driven mutations. Biochim Biophys Acta (BBA)-Bioenergetics. 2015;1847:1354–61.

19. Safdar A, Bourgeois JM, Ogborn DI, Little JP, Hettinga BP, Akhtar M, et al. Endurance exercise rescues progeroid aging and induces systemic mitochondrial rejuvenation in mtDNA mutator mice. Proc Natl Acad Sci U S A. 2011;108:4135–40.

20. Ross JM, Stewart JB, Hagström E, Brené S, Mourier A, Coppotelli G, et al. Germline mitochondrial DNA mutations aggravate ageing and can impair brain development. Nature. 2013;501:412–5.

21. Fan W, Waymire KG, Narula N, Li P, Rocher C, Coskun PE, et al. A Mouse Model of Mitochondrial Disease Reveals Germline Selection Against Severe mtDNA Mutations. Sci . 2008;319:958–62.

22. Mao X, Dong J, Hua P, He G, Zhang S, Rossiter SJ. Heteroplasmy and ancient translocation of mitochondrial DNA to the nucleus in the Chinese Horseshoe Bat (Rhinolophus sinicus) complex. PLoS One. 2014;9:e98035.

23. Petri B, Haeseler A von, Paabo S. Extreme Sequence Heteroplasmy in Bat Mitochondrial DNA. Biol Chem. 1996;377:661–7.

24. He X, Chen X, Zhang W, Pu Y, Song S, Han J, et al. High occurrence of length heteroplasmy in domestic Bactrian camel (*Camelus bactrianus*). Mitochondrial DNA Part A. 2016;0:1–3.

25. Zhao YJ, Huang YF, others. Sheep mitochondrial heteroplasmy arises from tandem motifs and unspecific PCR amplification. Mitochondrial DNA Part A. 2016;0:1–5.

26. Fumagalli L, Taberlet P, Favre L, Hausser J. Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. Mol Biol Evol. 1996;13:31.

27. Cook CE, Wang Y, Sensabaugh G. A Mitochondrial Control Region and CytochromebPhylogeny of Sika Deer (Cervus nippon) and Report of Tandem Repeats in the Control Region. Mol Phylogenet Evol. 1999;12:47–56.

28. Hoelzel AR, Lopez J V, Dover GA, O'Brien SJ. Rapid evolution of a heteroplasmic repetitive sequence in the mitochondrial DNA control region of carnivores. J Mol Evol. 1994;39:191–9.

29. Lunt DH, Whipple LE, Hyman BC. Mitochondrial DNA variable number tandem repeats (VNTRs): Utility and problems in molecular ecology. Mol Ecol. 1998;7:1441–55.

30. Austad SN. Methusaleh's Zoo: how nature provides us with clues for extending human health span. J Comp Pathol. 2010;142 Suppl:S10-21.

31. Brunet-Rossinni AK, Austad SN. Ageing studies on bats: a review. Biogerontology. 2004;5:211–22.

32. Seim I, Fang X, Xiong Z, Lobanov A V, Huang Z, Ma S, et al. Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. Nat Commun. 2013;4:2212.

33. Podlutsky AJ, Khritankov AM, Ovodov ND, Austad SN. A New Field Record for Bat Longevity. 2005;60:1366–8.

34. Teeling EC, Madsen O, Van Den Bussche RA, de Jong WW, Stanhope MJ, Springer MS. Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. Proc Natl Acad Sci. 2002;99:1431–6.

35. Teeling EC, Springer MS, Madsen O, Bates P, O'brien SJ, Murphy WJ. A molecular

phylogeny for bats illuminates biogeography and the fossil record. Science. 2005;307:580–4.

36. Thomas SP, Suthers RA. The physiology and energetics of bat flight. J Exp Biol. 1972;57:317–35.

37. Maina JN. What it takes to fly: the structural and functional respiratory refinements in birds and bats. J Exp Biol. 2000;203:3045–64.

38. Munshi-South J, Wilkinson GS. Bats and birds: Exceptional longevity despite high metabolic rates. Ageing Res Rev. 2009;9:12–9.

39. Brunet-Rossinni AK. Reduced free-radical production and extreme longevity in the little brown bat (Myotis lucifugus) versus two non-flying mammals. Mech Ageing Dev. 2004;125:11–20.

40. Kerth G, Mayer F, Konig B. Mitochondrial DNA (mtDNA) reveals that female Bechstein ' s bats live in closed societies. Mol Ecol. 2000;9:793–800.

41. Wilkinson GS, Mayer F, Gerald K, Petri B. Evolution of Repeated Sequence Arrays in the D-Loop Region of Bat Mitochondrial DNA. Genetics. 1997;146:1035–48.

42. Hoofer SR, van den Bussche R a. Molecular phylogenetics of the chiropteran family Vespertilionidae. Acta Chiropterologica. 2003;5:1–63.

43. McElroy KE, Luciani F, Thomas T. GemSIM: general, error-model based simulator of next-generation sequencing data. BMC Genomics. 2012;13:74. doi:10.1186/1471-2164-13-74.

44. de Magalhães JP, Kean M. Endless paces of degeneration--applying comparative genomics to study evolution's moulding of longevity. EMBO Rep. 2013;14:661–2.

45. Kauppila TES, Kauppila JHK, Larsson N-G. Mammalian Mitochondria and Aging: An Update. Cell Metab. 2016;:1–15.

46. Speakman JR, Garratt M. Oxidative stress as a cost of reproduction: Beyond the simplistic trade-off model. BioEssays. 2014;36:93–106.

47. Nathan C, Cunningham-Bussel A. Beyond oxidative stress: an immunologist's guide to reactive oxygen species. Nat Rev Immunol. 2013;13:349–61.

48. McGettrick AF, O'Neill L a J. How Metabolism Generates Signals during Innate Immunity and Inflammation. J Biol Chem. 2013;288:22893–8.

49. Li W. Bats Are Natural Reservoirs of SARS-Like Coronaviruses. Science. 2005;310:676–9.

50. Wang LF, Walker PJ, Poon LLM. Mass extinctions, biodiversity and mitochondrial function: Are bats "special" as reservoirs for emerging viruses? Curr Opin Virol. 2011;1:649–57.

51. Calisher CH, Childs JE, Field HE, Holmes K V., Schountz T. Bats: Important reservoir hosts of emerging viruses. Clin Microbiol Rev. 2006;19:531–45.

52. Luis AD, Hayman DTS, O'Shea TJ, Cryan PM, Gilbert AT, Pulliam JRC, et al. A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? Proc R Soc B Biol Sci. 2013;280: 20122753.

53. Zhou P, Tachedjian M, Wynne JW, Boyd V, Cui J, Smith I, et al. Contraction of the type I IFN locus and unusual constitutive expression of IFN-α in bats. Proc Natl Acad Sci. 2016;113:2696–701.

54. Sena L, Chandel N. Physiological roles of mitochondrial reactive oxygen species. Mol Cell.

2012;48:158–67.

55. Schneeberger K, Czirjak GA, Voigt CC. Inflammatory challenge increases measures of oxidative stress in a free-ranging, long-lived mammal. J Exp Biol. 2013;216:4514–9.

56. Garratt M, Mcardle F, Stockley P, Vasilaki A, Beynon RJ, Jackson MJ, et al. Tissue-dependent changes in oxidative damage with male reproductive effort in house mice. Funct Ecol. 2012;26:423–33.

57. Zhang G, Cowled C, Shi Z, Huang Z, Bishop-Lilly K a, Fang X, et al. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. Science. 2013;339:456–60.

58. Morgan CC, Mc Cartney AM, Donoghue MT, Loughran NB, Spillane C, Teeling EC, et al. Molecular adaptation of telomere associated genes in mammals. BMC Evol Biol. 2013;13:251.

59. Huang Z, Jebb D, Teeling EC. Blood miRNomes and transcriptomes reveal novel longevity mechanisms in the long-lived bat, Myotis myotis. BMC Genomics. 2016;17:906.

60. Scherz-Shouval R, Shvets E, Fass E, Shorer H, Gil L, Elazar Z. Reactive oxygen species are essential for autophagy and specifically regulate the activity of Atg4. EMBO J. 2007;26:1749–60.

61. Pride H, Yu Z, Sunchu B, Mochnick J, Coles A, Zhang Y, et al. Long-lived species have improved proteostasis compared to phylogenetically-related shorter-lived species. Biochem Biophys Res Commun. 2015;457:669–75.

62. Brunet-Rossinni AK. Testing the Free Radical Theory of Aging in Bats. Ann N Y Acad Sci.

2004;1019:506–8.

63. Huang Z, Gallot A, Lao NT, Puechmaille SJ, Foley NM, Jebb D, et al. A non-lethal sampling method to obtain, generate and assemble whole-blood transcriptomes from small, wild mammals. Mol Ecol Resour. 2016;16:150–62.

64. Jebb D, Foley NM, Puechmaille SJ, Teeling EC. The complete mitochondrial genome of the Greater Mouse-Eared bat, Myotis myotis (Chiroptera: Vespertilionidae). Mitochondrial DNA. 2017;28:347–9.

65. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17:10.

66. Patel RK, Jain M. NGS QC toolkit: A toolkit for quality control of next generation sequencing data. PLoS One. 2012;7.

67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

68. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. doi:arXiv:1303.3997 [q-bio.GN].

69. Picard tools version 1.90. http://broadinstitute.github.io/picard. citeulike-article-id:12519005.

70. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.

71. Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, et al. LoFreq: a sequence-

quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012;40:11189–201.

72. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568–76.

73. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012;:9. http://arxiv.org/abs/1207.3907.

74. Tange O. GNU Parallel - The Command-Line Power Tool. ;login USENIX Mag. 2011;36:42–7.

75. Choi Y, Chan AP. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31:2745–7.

76. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS One. 2012;7.

77. Venables WN, Ripley BD. Modern Applied Statistics with S. Fourth. New York: Springer; 2002.

78. Canty A, Ripley BD. boot: Bootstrap R (S-Plus) Functions. 2017.

79. Davison AC, Hinkley D V. Bootstrap Methods and Their Applications. Cambridge: Cambridge University Press; 1997.

80. Cook RD. Influential Observations and Outliers in Linear Regression. Technometrics. 1977;19:15–8.

81. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet. 2007;52:267–74.

Additional Files

Name: Additional File 1

Format: Microsoft Excel File (.xlsx)

Sheet 1: Simulation_results – Power, accuracy, false positive rate and score used to determine best variant caller from simulated datasets at different coverages

Sheet 2: Filtered_heteroplasmies – 254 annotated heteroplasmies discovered in this study

Sheet 3: Heteroplasmies_all_samples – Information for 232 samples including RFID, number of heteroplasmic sites and whether or not the sample was part of the primary dataset.

Sheet 4: Primary_dataset – Age, RFID and number of heteroplasmic sites for 167 samples used in the "Primary" dataset.
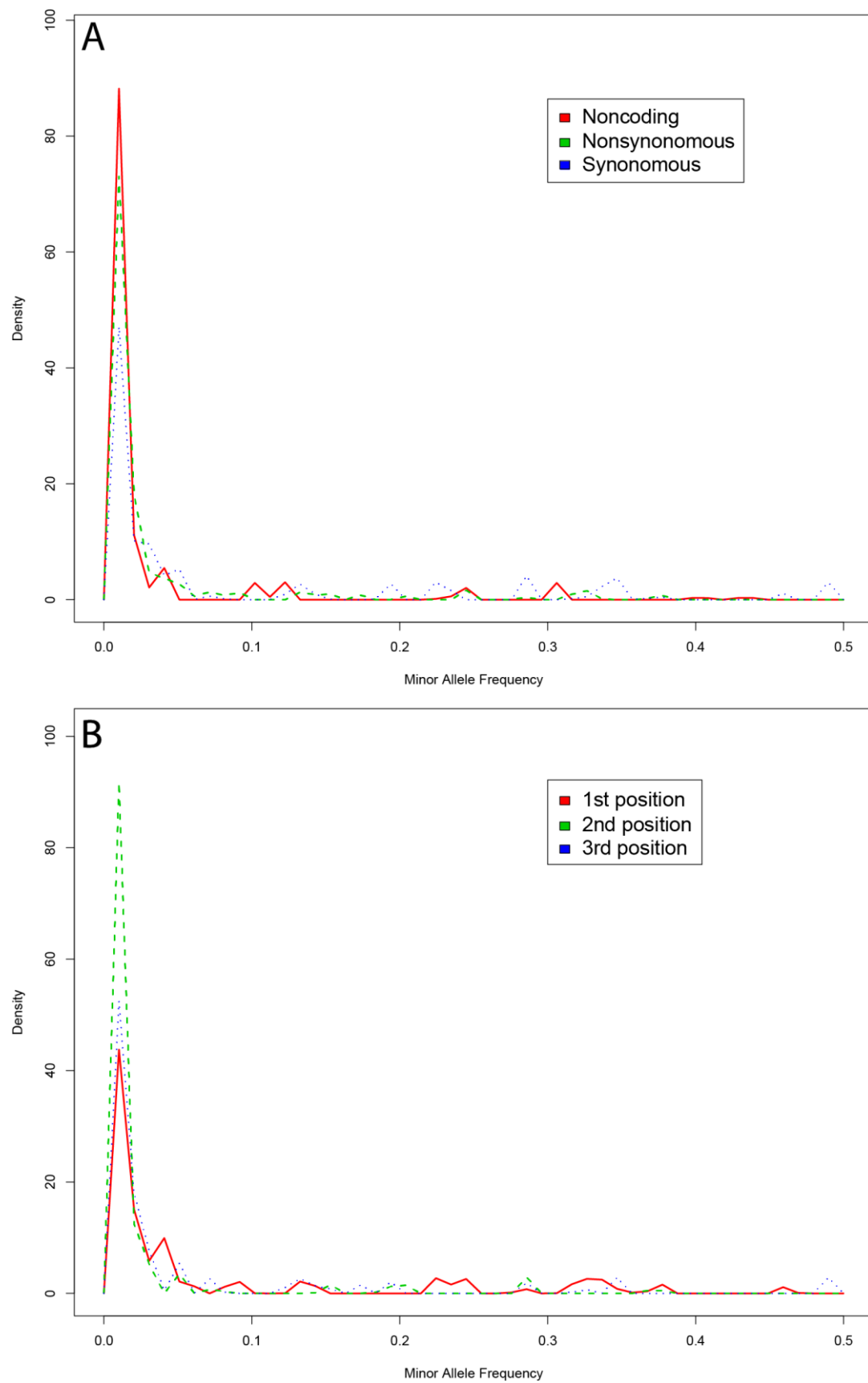
A

B

C

Figure 1

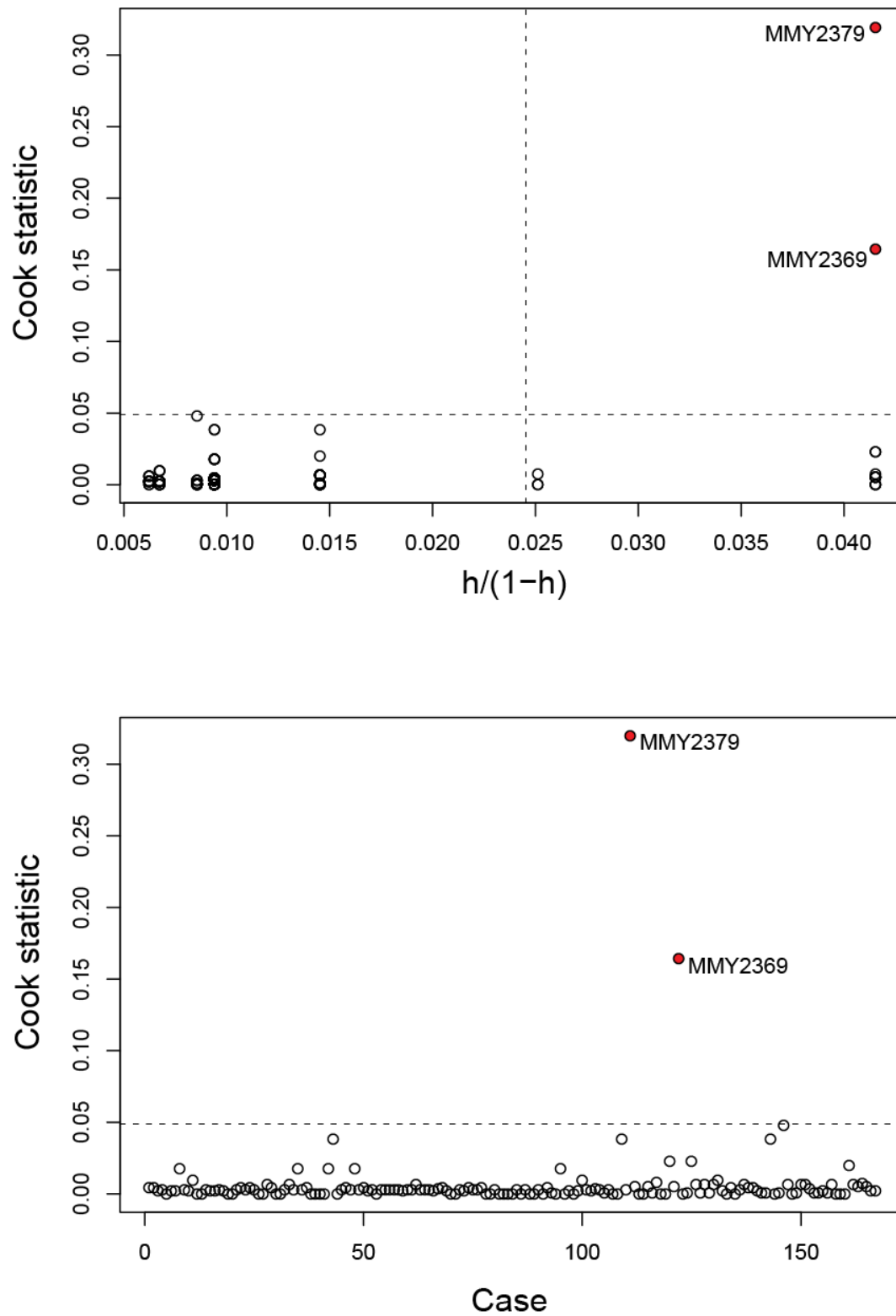N = 165 unique individuals
p = 0.1, Neg Binom GLM
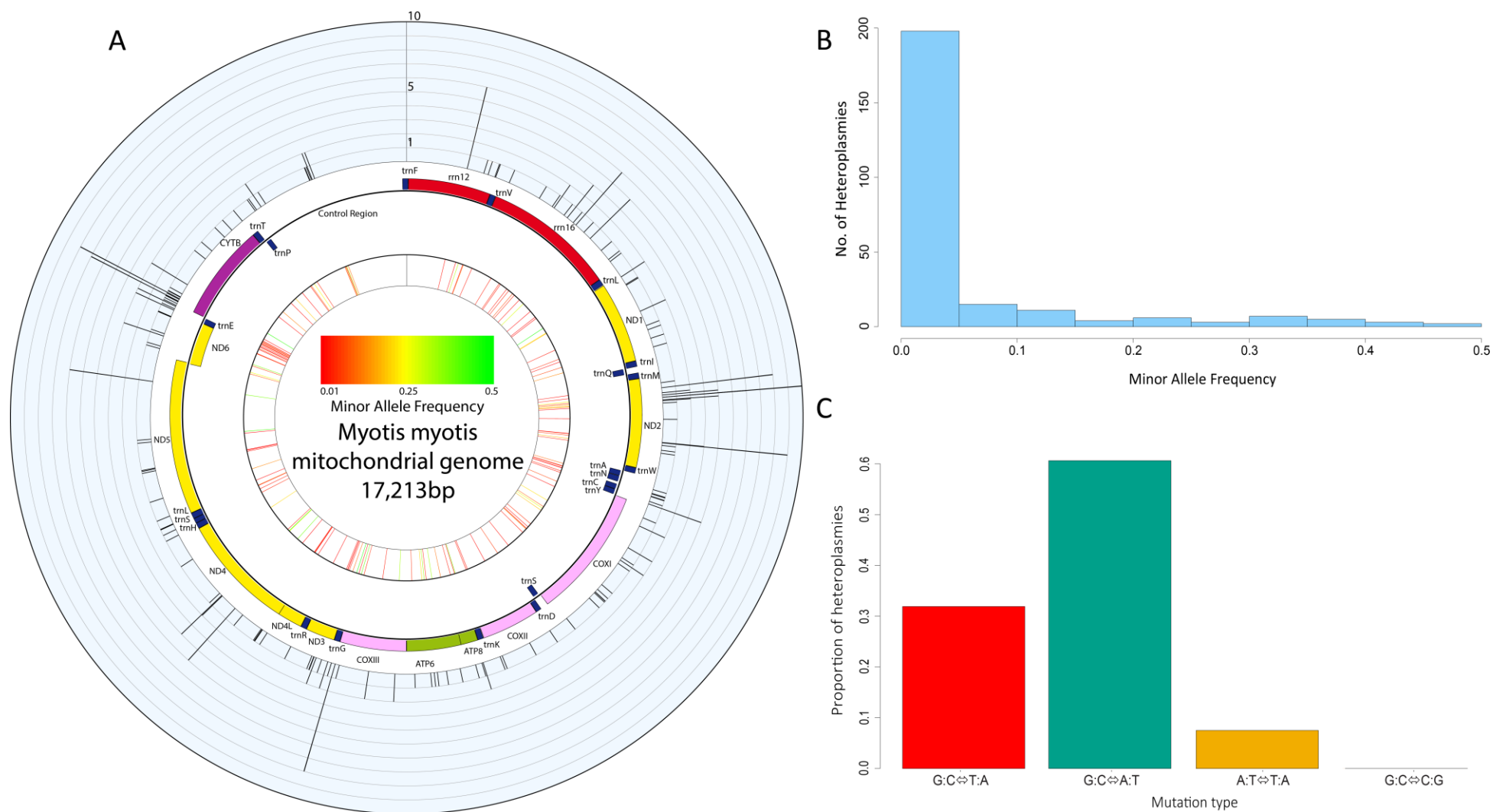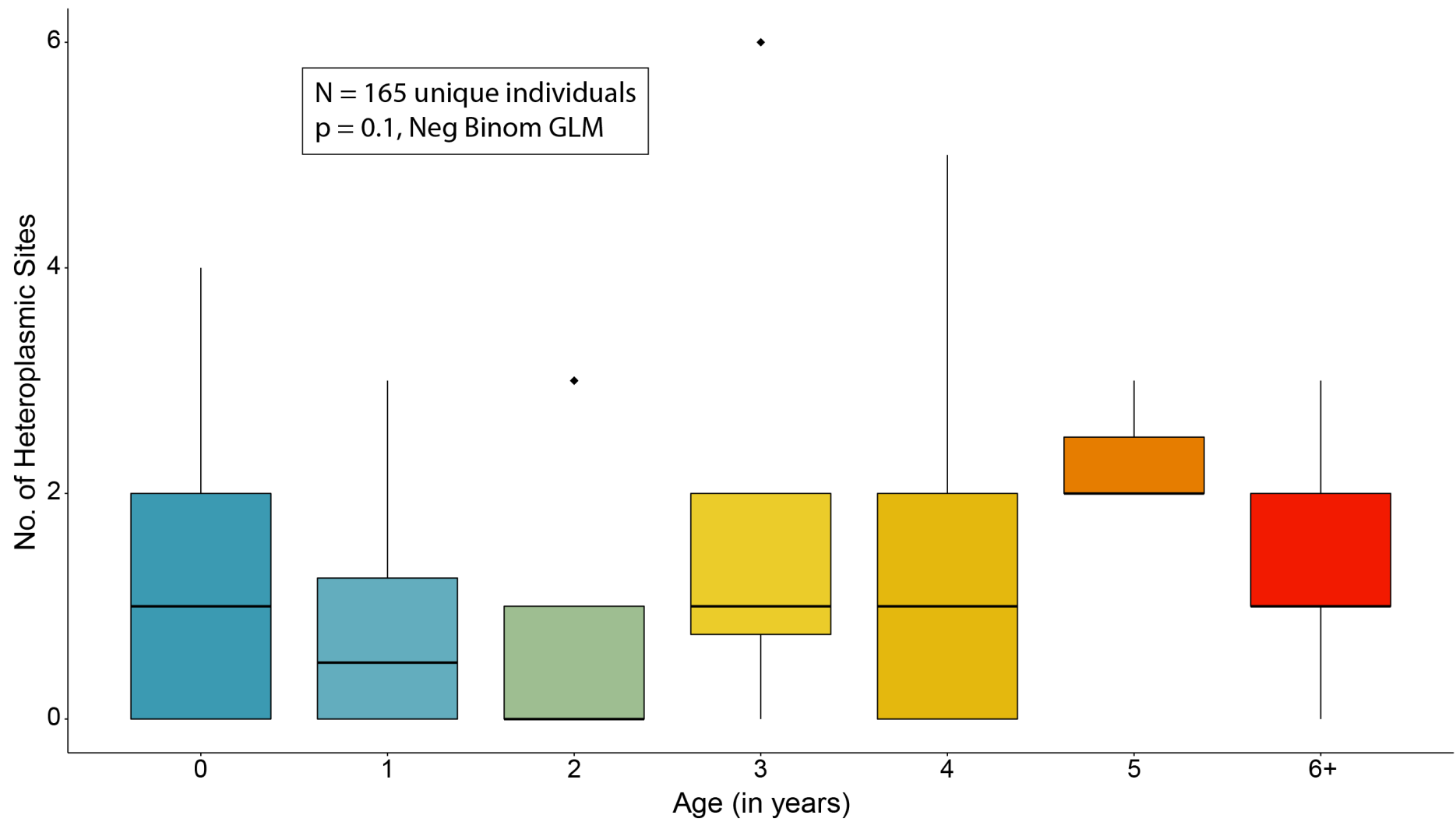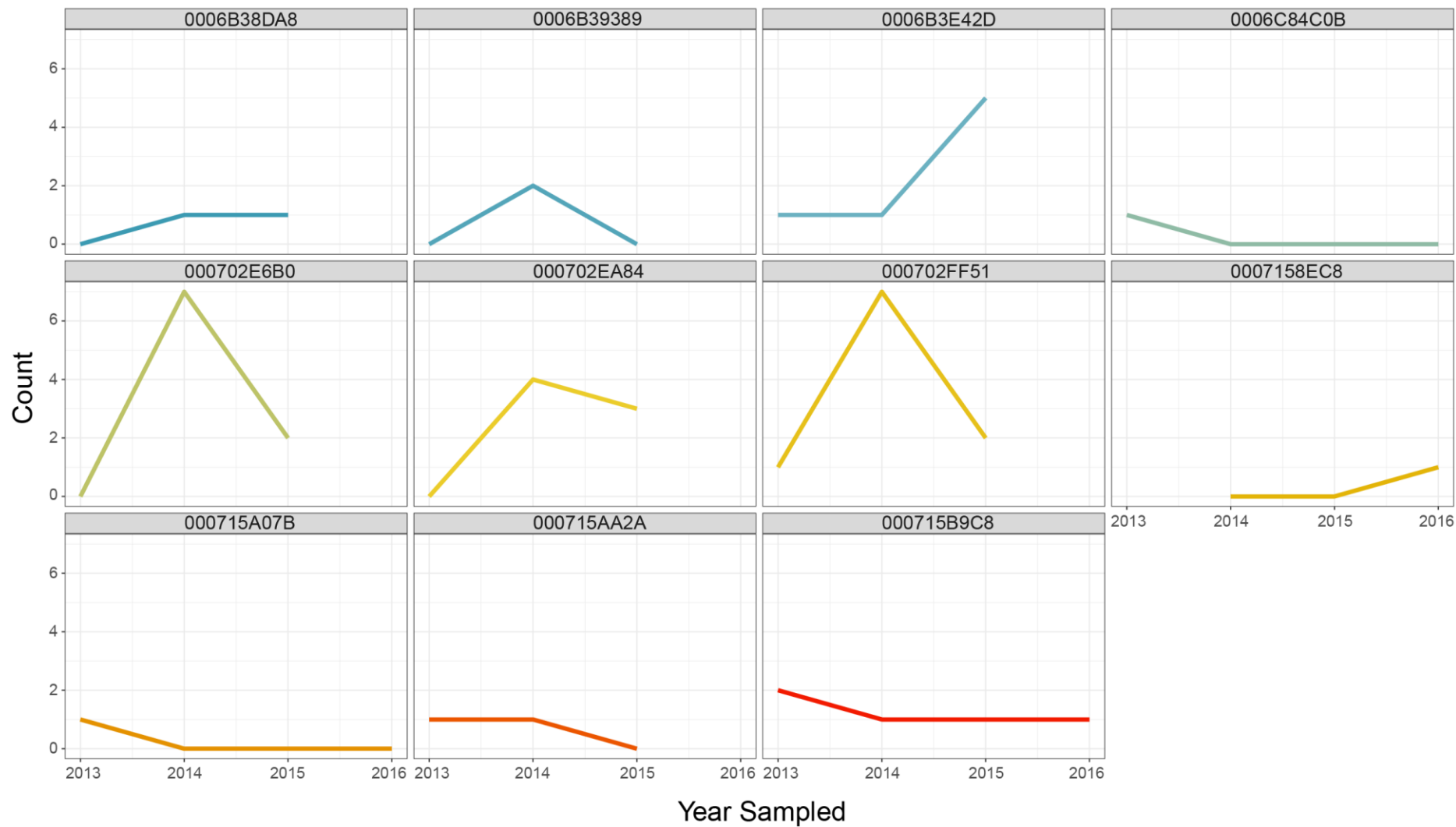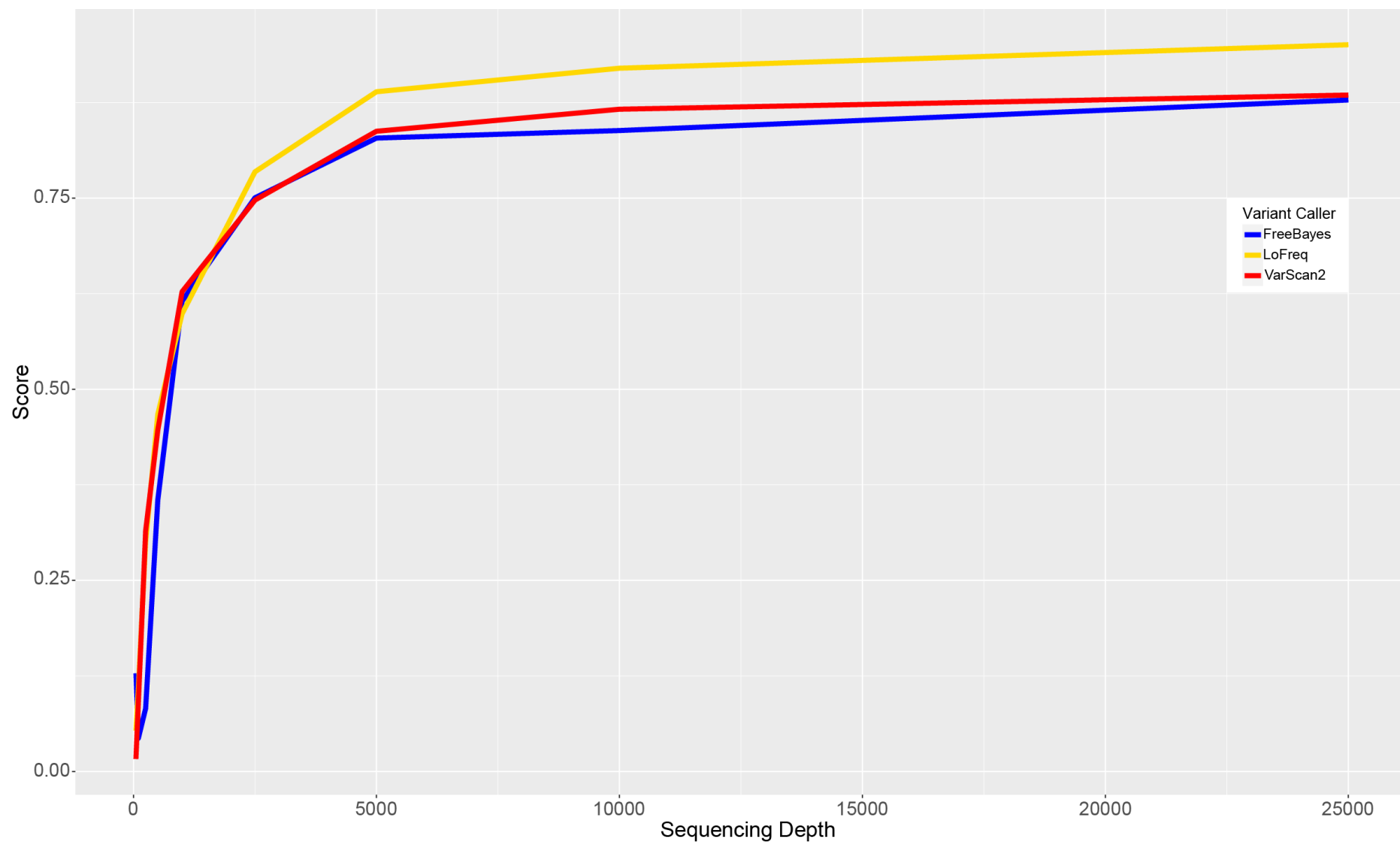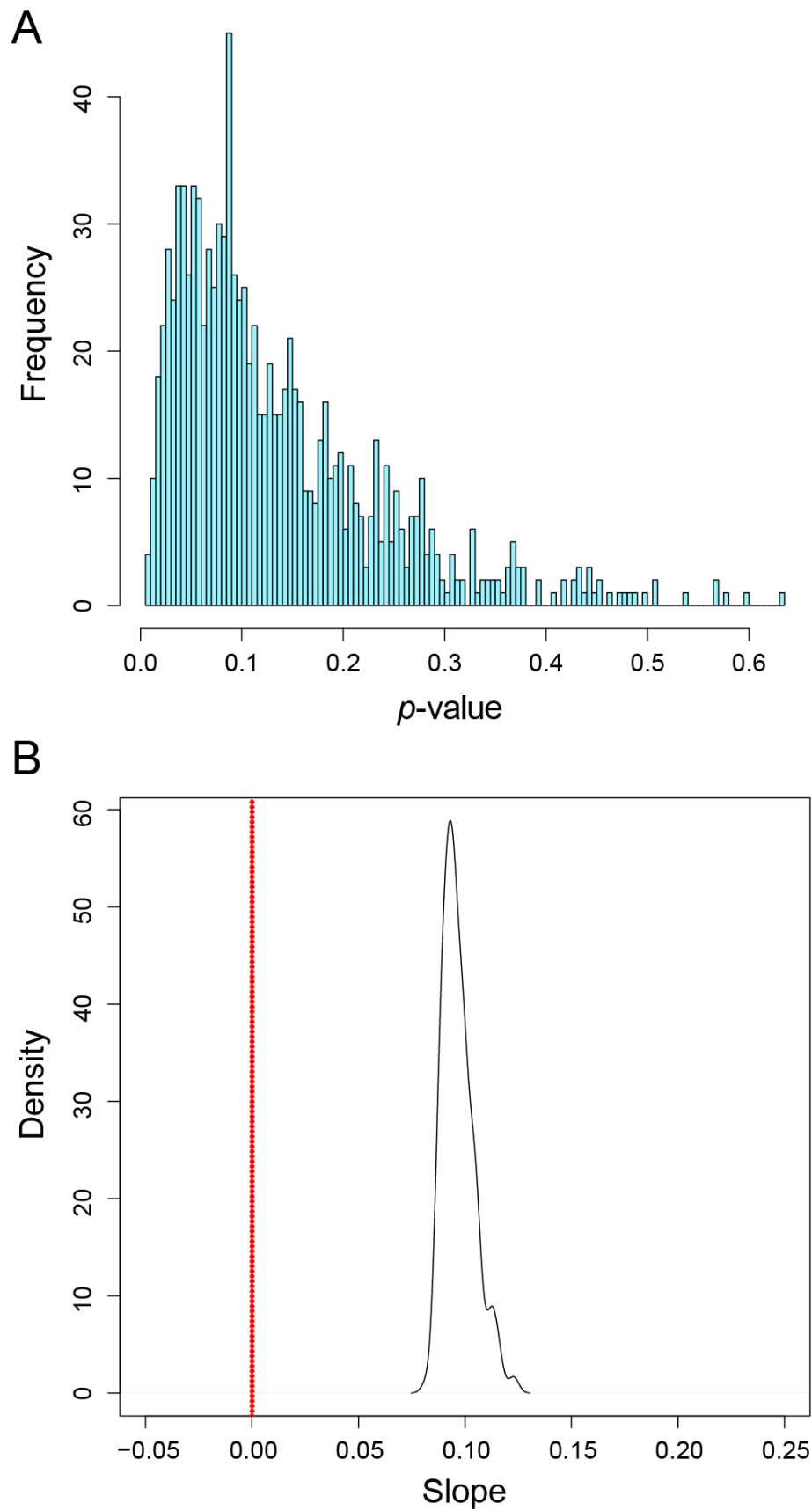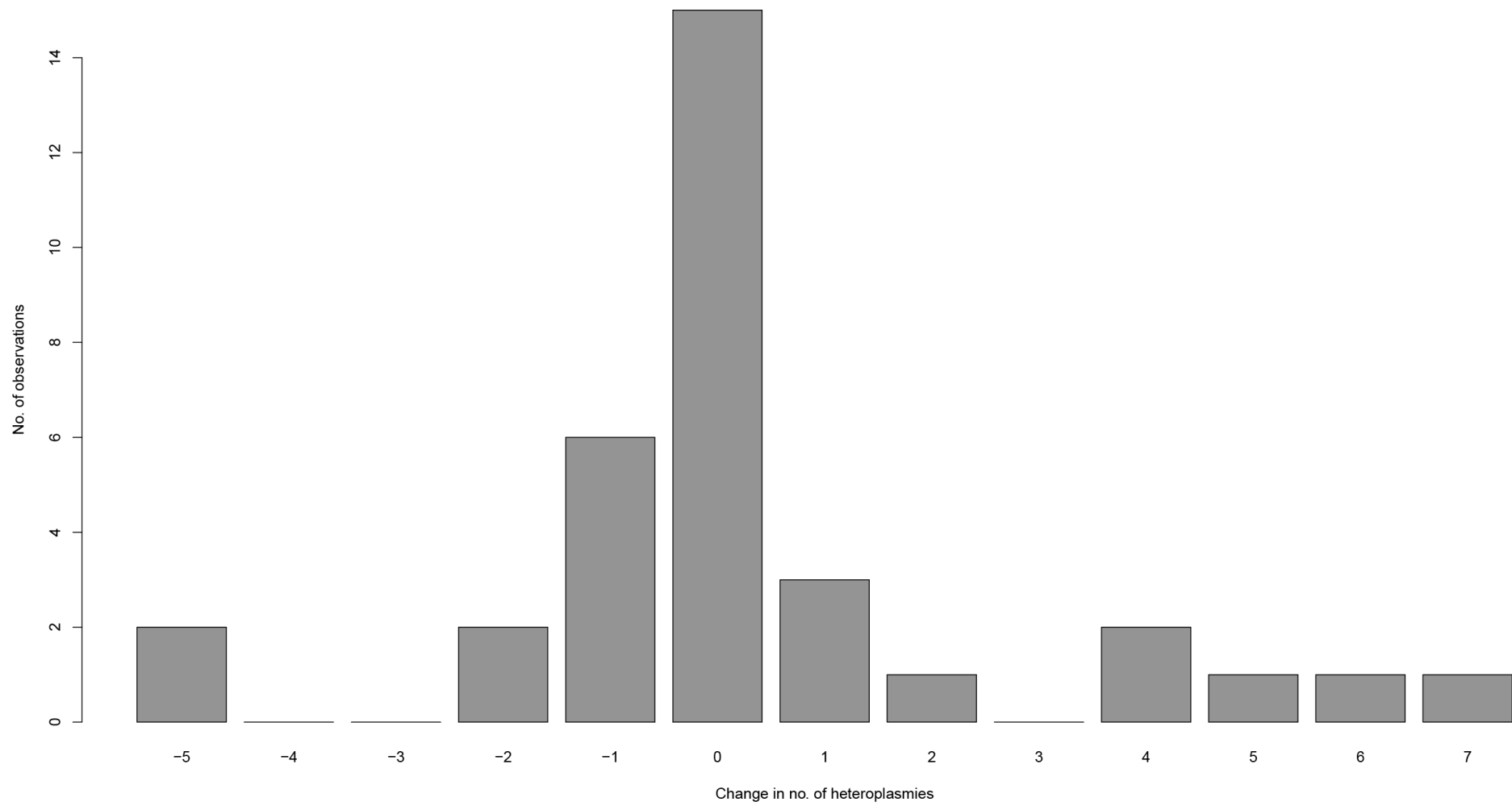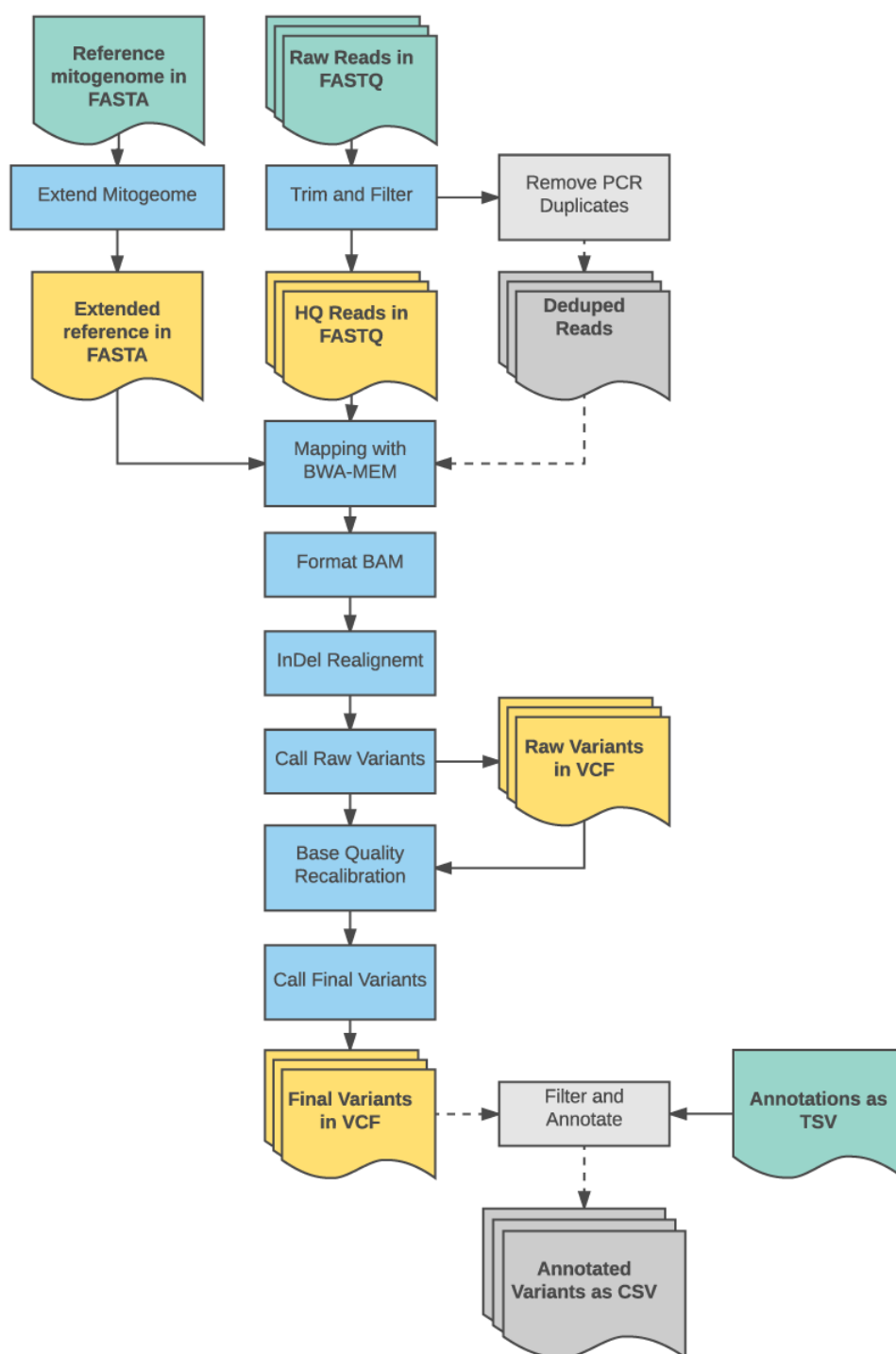
Figure 2

Figure 3

Supp. Figure 1

Supp. Figure 2

45

Supp. Figure 3

Supp. Figure 4

Supp. Figure 5

Supp. Figure 6