The Maintenance of Deleterious Variation in Wild Chinese Rhesus Macaques

Camille Steux[1, 2] and Zachary A. Szpiech[1, 3, *]

[1]Department of Biology, Pennsylvania State University, USA

[2]Département de Biologie, École Normale Supérieure, PSL Université Paris, Paris, France

[3]Institute for Computational and Data Sciences, Pennsylvania State University, USA

[*]Author for Correspondence: Zachary A. Szpiech, Department of Biology, Pennsylvania State University, USA, zps5164@psu.edu

October 4, 2023

Abstract

Understanding how deleterious variation is shaped and maintained in natural populations is of importance for both medical research and conservation biology. Various demographic processes can influence these patterns. For example, population bottlenecks and inbreeding increase the likelihood of inheriting identical-by-descent haplotypes from a recent common ancestor, creating long tracts of homozygous genotypes called runs of homozygosity (ROH). Long ROH (e.g., those created as a result of inbreeding) have been associated with an accumulation of mildly deleterious homozygotes. Counter intuitively, positive selection can also maintain deleterious variants in a population by increasing the frequency of a damaging allele tightly linked to an advantageous mutation (i.e., via genetic hitchhiking).

Here we analyze whole genomes of 79 wild Chinese rhesus macaques across five subspecies and characterize patterns of deleterious variation with respect to ROH and signals of recent positive selection. We show that the fraction of deleterious homozygotes occurring in long

ROH is significantly higher for deleterious homozygotes than tolerated ones ($p = 3.4 \times 10^{-5}$) but not in short and medium ROH. This confirms that inbreeding, by generating these long tracts of homozygosity, is the main driver of the high burden of homozygous deleterious alleles in wild macaque populations. Furthermore, we show that small ROH have a lower proportion of homozygous loss-of-function variants than tolerated ones ($p = 0.05$), suggesting the purging of these variants through natural selection. Next, we identify regions putatively experiencing a recent or ongoing sweep and examine the presence of deleterious variants in these regions to find signals of genetic hitchhiking. We found 7 deleterious variants at high frequency in regions under selection containing genes involved with olfaction and other processes.

Significance Our results offer insights into the relationship between demographic and evolutionary processes and the maintenance of deleterious alleles in rhesus macaques. We demonstrate how inbreeding and positive selection can contribute to the maintenance of deleterious variants in wild non-human primate populations. This work also adds to our understanding of the evolutionary history of macaques and, as rhesus macaques are the most widely used non-human primate in biomedical research, may help inform the study of human diseases.

Key words deleterious variation, runs of homozygosity, inbreeding, natural selection, Chinese rhesus macaque.

Introduction

Investigating patterns of deleterious variation in both human and non-human populations is of crucial interest for understanding the consequences of these mutations on the phenotype and fitness of individuals, as well as for understanding how evolutionary processes such as selection or demography act on their distribution. On one hand, in humans, it is recognized that individuals carry a substantial load of deleterious heterozygotes, which are lethal in homozygous state (Kondrashov 1995; Morton et al. 1956). Additionally, individuals also carry thousands of mildly deleterious alleles, some of them have been linked to the occurrence of various complex diseases (Ohta 1973, 1992; Eyre-Walker and Keightley 1999; Consortium 2012; Keller et al. 2012; Fu et al. 2013; Lohmueller 2014). Furthermore, given that human populations have experienced different population histories (Keinan et al. 2007; Lohmueller et al. 2008), exploring the prevalence and impact of deleterious variation is particularly relevant in the context of medical research. On the other hand, in non-human species, genetic load can pose significant challenges, particularly in small and isolated endangered populations (Robinson et al. 2023; Leon-Apodaca et al. 2023). It has been shown that small and isolated populations can exhibit a higher proportion of damaging mutations, which can reduce individuals' reproductive fitness, populations' adaptive potential and extinction risk (Hedrick and Kalinowski 2000; Robinson et al. 2016, 2019; Bertorelle et al. 2022). Therefore, gaining insights in the maintenance of deleterious variation in human and non-human populations can potentially inform not only human medical research but also conservation efforts.

A well documented process influencing the maintenance of genetic load is the gathering of recessive deleterious variation in homozygous state via the phenomena of inbreeding (Charlesworth and Willis 2009; Szpiech et al. 2013; Zhang et al. 2015). Mating between close relatives generates long tracts of identical-by-descent chromosomal segments in the offspring, due to the presence of shared haplotypes in the parents inherited from the same recent ancestor (Woods et al. 2006; Ceballos et al. 2018). The increased availability of whole-genome sequencing data has allowed for the identification of these runs of homozygosity (ROH) (Broman and Weber 1999). Given the

relationship between ROH and autozygosity (McQuillan et al. 2008), studying patterns of ROH is useful for understanding inbreeding in populations (Keller et al. 2011). Studies of ROH in human populations show that ROH are ubiquitous, even in apparently outbred populations (Kirin et al. 2010; Pemberton et al. 2012; Ceballos et al. 2018), and ROH genome coverage can range from a few percent to 50% (Kirin et al. 2010; Ceballos et al. 2018; Pemberton et al. 2012; Szpiech et al. 2013; Yengo et al. 2019).

In humans, inbreeding can occur because of cultural preferences, some societies promoting consanguinity, whereas it can also result from periods of small population size population such as bottlenecks (Kirin et al. 2010; Pemberton et al. 2012; Ceballos et al. 2018; Laurent et al. 2023). Recent inbreeding generates long tracts of homozygosity. Over time, mutations and recombination events break up the underlying haplotypes, which can eventually result in shorter ROH (Szpiech et al. 2013; Ceballos et al. 2018). Therefore, the length of ROH in an individual can be used as a proxy to infer the time of the events that generated them (Ceballos et al. 2018), and the study of these regions can be informative about both long-term population demographic history and cultural practices (Pemberton et al. 2012; Ceballos et al. 2018; Pemberton and Szpiech 2018).

It is known that increased homozygosity can lead to a decrease in an offspring's fitness, a phenomenon called inbreeding depression (Woods et al. 2006; Darwin 2010). Studies in humans have shown that long ROH are a reservoir of homozygous deleterious variation, meaning that an individual with a high burden in ROH is likely to have a high burden in deleterious homozygotes as well (Szpiech et al. 2013; Pemberton and Szpiech 2018; Szpiech et al. 2019). Indeed, several studies have shown a significant association between Wright's inbreeding coefficient and some complex traits (Rudan et al. 2003), and several of these associations have been confirmed using inbreeding coefficient estimated from ROH, in humans (Woods et al. 2006; Verweij et al. 2014) and in other species as well like dogs (Mooney et al. 2021). In humans, a high burden of ROH has been shown to be associated with lower standing height and with increased risk in coronary artery disease, some cancers and some neurodegenerative disorders such as Alzheimer and Parkinson diseases (Verweij et al. 2014; Rudan et al. 2003). In non-humans, despite that some wild endangered populations

challenge this paradigm (Robinson et al. 2016; Benazzo et al. 2017; Robinson et al. 2019; Leon-Apodaca et al. 2023), several studies have shown that inbreeding reduces individuals reproduction and survival, and that inbreeding depression can increase species extinction risk (Brüniche-Olsen et al. 2018; Spielman et al. 2004).

Liu et al. (2018) have sequenced the whole genomes of wild Chinese rhesus macaques (*macaca mulatta*) from various locations in China and covering the five currently recognised sub-species, providing the opportunity to study the genomics of wild populations of this non-human primate. In this paper, we investigate patterns of deleterious variation in relation to runs of homozygosity and natural selection. In the first part, we follow Szpiech et al. (2013) and compute the fraction of deleterious, loss-of-function and tolerated variants falling within the different classes of ROH (small, medium and large). Then we identify regions putatively under positive selection in each of the five sub-species and identify within these regions deleterious variants at high frequency and the genes they are in. As *Macaca mulatta* is the most widely used non-human primate model in biomedical research (Phillips et al. 2014; Rogers 2022), our study may therefore inform the investigation of human diseases, as well as give insights in macaque's evolutionary history.

Results

Genomic Data Set

Our data was retrieved from Liu et al. (2018), who sequenced whole genomes of 79 wild Chinese rhesus macaques across the five subspecies. After lifting over to a more recent reference genome (Mmul_10, Warren et al. (2020)) and filtering and phasing (see Materials and Methods), our data set consisted in 5,713,999 single-nucleotide polymorphisms for which at least one of the 79 individuals had a non-reference allele. We predicted coding mutation effects with SIFT4G (Vaser et al. 2016), finding 17,477 variants with a high confidence call. Of these, 14,913 mutations were predicted to be tolerated, 2,385 mutations were predicted to be deleterious, and 179 mutations were

annotated as loss-of-function.

Runs of Homozygosity

ROH were identified using GARLIC (Szpiech et al. 2017) and classified into three different size classes: short ROH (class A, < 92,729 bp), medium ROH (class B) and long ROH (class C, > 441,568 bp). The mean length was 44,695 bp, 157,206 bp and 2.06 Mbp for class A, B and C respectively. The longest identified ROH across all individuals was 30.7 Mbp long and was found in *M. m. lasiotis* (individual C_rhe_50).

By definition, heterozygous genotypes should be absent in ROH across the genome. However, ROH calling methods (including GARLIC) do allow for a small number to appear within ROH calls, on the assumption that they are either likely genotyping errors or very recent mutations. We check the distribution of heterozygotes within ROH and non-ROH regions, expecting that heterozygotes will be distributed almost exclusively outside of ROH. Furthermore, as we expect short ROH to consist in older haplotypes with more mutations having had time to occur than in long ROH, we expect the proportion of heterozygotes falling in ROH to decrease as a function of longer (i.e., younger) ROH. For each individual, we counted the number of genotypes homozygous for the reference allele (0/0), homozygous for the alternate allele (1/1) and heterogyotes (0/1). We then mapped each variant onto ROH, and computed the proportion of each genotype falling outside or inside a specific class ROH.

The average genome-wide proportion of heterozygotes across all individuals is 16.80 % of the total number of polymorphic sites. The proportion of heterozygotes falling outside ROH is 16.70%, whereas the proportion of heterozygotes falling within ROH is equal to 0.05%. Furthermore, the proportion of heterozygotes falling in ROH drops as we move from short ROH (0.023%), to medium ROH (0.020%), to long ROH (0.009%).

To summarize the distribution of ROH, we first compute two statistics used in previous studies (Pemberton et al. 2012; Ceballos et al. 2018; Szpiech et al. 2013): the sum total length of ROH per individual (SROH) and the total number of ROH per individual (NROH). SROH and NROH are in-

formative about population size history and present levels of inbreeding (Ceballos et al. 2018). We expect small bottlenecked populations to show high values of SROH and NROH, whereas large (or admixed) populations are expected to carry fewer and smaller ROH (Ceballos et al. 2018). Inbred populations are expected to show many extremely long ROH, and thus spread out conspicuously far on the SROH axis. We observe in Figure 1 different patterns among the subspecies. Individuals from the same subspecies tend to cluster together, though there is some heterogeneity for *M. m. lasiotis* and *M. m. littoralis*. The subspecies *M. m. mulatta* exhibits short total lengths of ROH and small numbers of ROH, whereas *M. m. tcheliensis* presents the longest and most numerous ROH, followed by *M. m. brevicaudus*. *M. m. lasiotis* and *M. m. littoralis* present similar intermediate distributions.

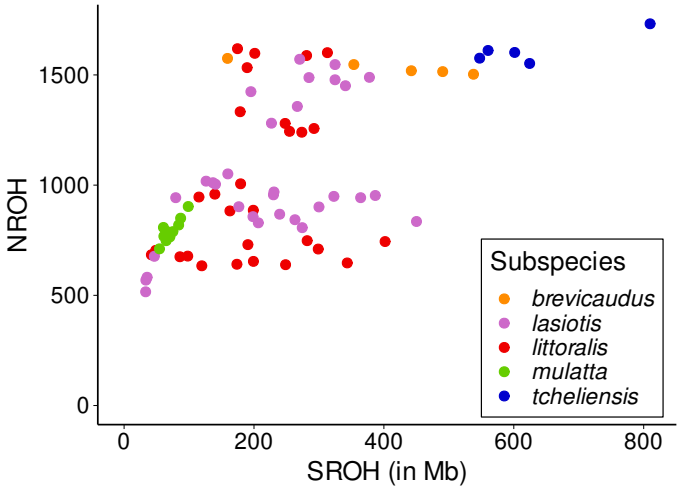

Figure 1: Total number of ROH (NROH) versus the total length of ROH (SROH) for each individual, colored by sub-species.

Next we consider total genomic ROH partitioned by size class. We calculate $G_{i,j}$, the total fraction of the genome covered by any ROH ($j = R$) or a specific class ($j \in \{A, B, C\}$) in individual i, as:

$$G_{i,j} = \frac{\text{total length of class } j \text{ ROH in individual } i}{\text{total length of the genome}}$$

Figure 2 shows the total fraction of the genome covered by ROH in each individual of our

sample, using a genome size of 2.9 Gb (Warren et al. 2020). ROH genome coverage ranges from a few percentages to around 28%, and the proportion of class A, class B and class C ROH varies across individuals. The subspecies *M. m. tcheliensis* and *M. m. brevicaudus* exhibit the individuals with the highest ROH genome coverage, suggesting a strong bottleneck or inbreeding in these subspecies, whereas individuals from *M. m. mulatta* show the lowest ROH genome coverage likely indicating a large effective population size.

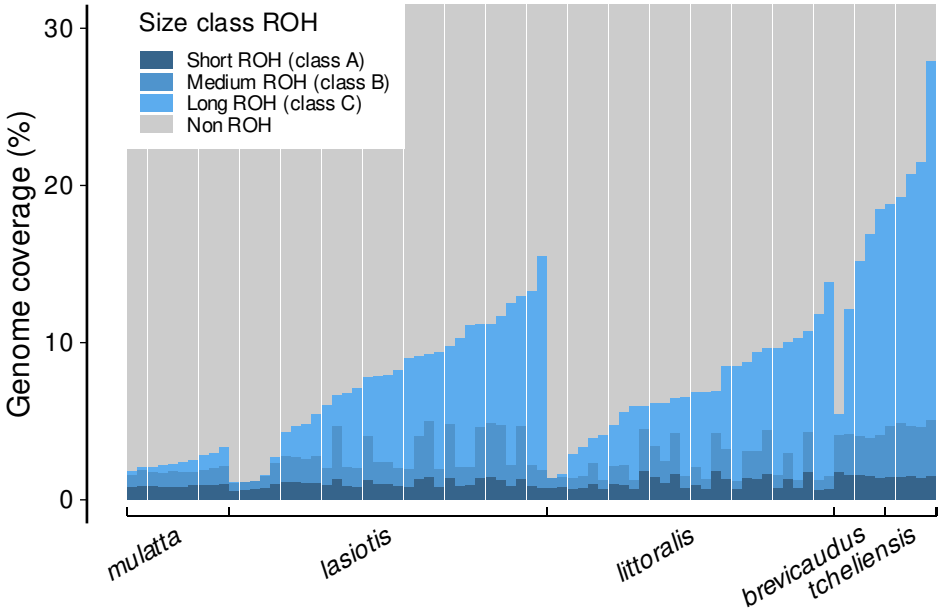

Figure 2: Fraction of the genome covered in ROH. Each shade of blue represent a size class ROH: class A in dark blue, class B in medium blue and class C in light blue. Each bar represents one individual.

Distribution of Deleterious and Tolerated Variation

Total distribution of deleterious variation

We used SIFT (Vaser et al. 2016) (see Material and Methods) to predict the deleteriousness of variants. Variants were predicted to be deleterious, tolerated, stop-gain, stop-loss or start-loss. We combined the variants predicted to be stop-gain, stop-loss and start-loss as loss of function, leaving us with three categories: deleterious, tolerated and loss of function. Figure 3A shows

that individuals carry between 23 and up to 64 deleterious alternate-allele homozygotes, with *M. m. mulatta* harboring the smallest number on average, followed by *M. m. lasiotis* and *M. m. littoralis*, *M. m. tcheliensis* and *M. m. brevicaudus*. This pattern is quiet close to the observed distribution of ROH across sub-species, which is consistent with the fact that we would expect the subspecies harbouring the highest length of ROH to have the highest number of homozygotes. We would expect the opposite trend for deleterious heterozygotes, and we see in Figure 3B that *M. m. tcheliensis* carries indeed the smallest number of deleterious heterozygotes (between 160 and 215), though the four other subspecies carry a similar number (between 190 and 280). Finally, the highest average genetic load (2*number of homozygotes + number of heterozygotes) (Figure 3C) is found in *M. m. brevicaudus*, whereas *M. m. tcheliensis* carries the smallest load on average. The highest number of deleterious alleles is found in *M. m. littoralis*, with 373 deleterious alleles.

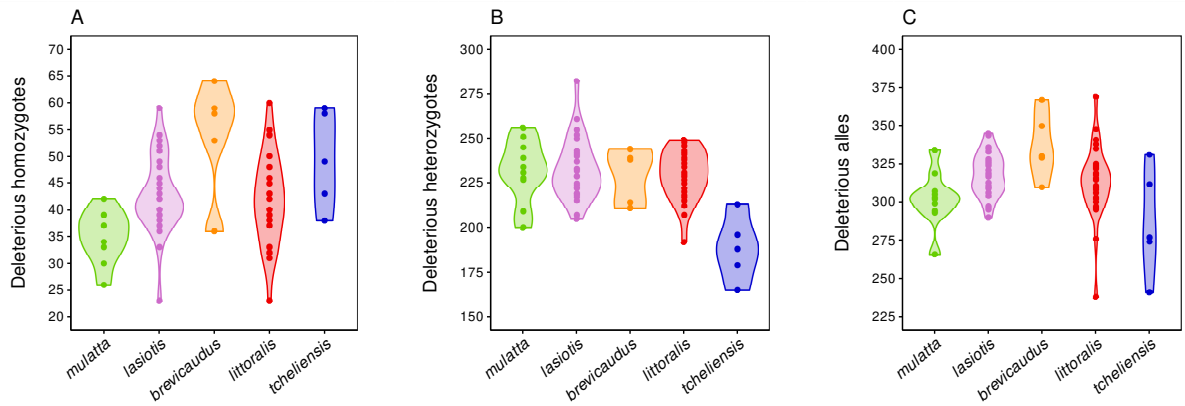

Figure 3: Distribution of the Number of (A) Deleterious Alternate-allele Homozygotes, (B) Deleterious Heterozygotes and (C) Deleterious Alleles predicted by SIFT (Vaser et al. 2016) for each sub-species.

Deleterious and Tolerated Homozygotes in Runs of Homozygosity

For each individual i, we computed $H^d_{i,j}$, which is the number of deleterious alternate-allele homozygotes that occur within a ROH of class j, where $j \in \{A, B, C, R\}$ indicates small (A), medium (B), long (C), or any (R) ROH. We similarly computed $H^t_{i,j}$ for tolerated alternate-allele homozygotes. Supplementary Figure S2 and Table S3 show that the number of deleterious or tolerated alternate homozygotes falling in ROH increases as we move from *M. m. mulatta* to *M. m. lasiotis*,

M. m. littoralis, *M. m. brevicaudus* and to *M. m. tcheliensis*. This pattern across supspecies is the same as the pattern observed in ROH (see Supplementary Figure S1), which is consistent with the fact that the higher the ROH genome coverage is, the higher is the probability for homozygotes to fall within these regions. Furthermore, we see that the number of homozygotes falling in ROH increases on average as we move from short, medium to long ROH, which was expected as homozygotes are more likely to fall within longer ROH.

For each individual i, we computed the fraction of deleterious homozygotes occurring within ROH type j as

$$F^d_{i,j} = \frac{H^d_{i,j}}{H^d_i},$$

where H^d_i is the total number of deleterious homozygotes genome-wide. We computed $F^t_{i,j}$ and H^t_i as above but for tolerated alternate-allele homozygotes. In Figure 4A, we observe a positive correlation between both the deleterious (Pearson r = 0.878) and the tolerated (Pearson r = 0.962) fraction of homozygotes falling into ROH and the fraction of the genome covered by ROH. These correlations were expected, as we can expect a higher number of homozygotes falling within ROH when the length of ROH in the genome increases. Furthermore, we observe a higher fraction of deleterious homozygotes falling within ROH than the fraction of tolerated homozygotes, suggesting a faster enrichment in deleterious homozygotes as the ROH genome coverage increases than in tolerated homozygotes.

To assess the statistical significance between the separate regressions for damaging and tolerated homozygotes, we fit the following linear model:

$$f_{i,R} = \beta_0 + \beta_1 G_{i,R} + \beta_2 D_i + \beta_3 G_{i,R} D_i + \varepsilon$$

where $f_{i,R}$ is a vector of length 158 containing, for all individuals, the fraction of genome-wide damaging homozygotes in any ROH region ($F^d_{i,R}$) and the fraction of genome-wide tolerated homozygotes in any ROH region ($F^t_{i,R}$). $G_{i,R}$ is the fraction of the genome covered by any size ROH in individual i, and D_i is an indicator variable taking the value 1 if the response variable is calculated

from predicted deleterious variation and 0 if the response variable is calculated from predicted to be tolerated variation. Thus, a statistically significant β_2 implies a difference in the intercepts of separate regression for damaging and tolerated homozygotes, and a statistically significant β_3 indicates a significant difference in slope.

We found that β_2 is not significant (p = 0.61225) but β_3 = 0.70608 is ($p = 4.9 \times 10^{-5}$). Therefore, there is no significant difference in the intercept of the two regressions, but a significant difference in slope. This means that there is a significant higher enrichment in deleterious homozygotes per unit increase in ROH as compared to tolerated homozygotes.

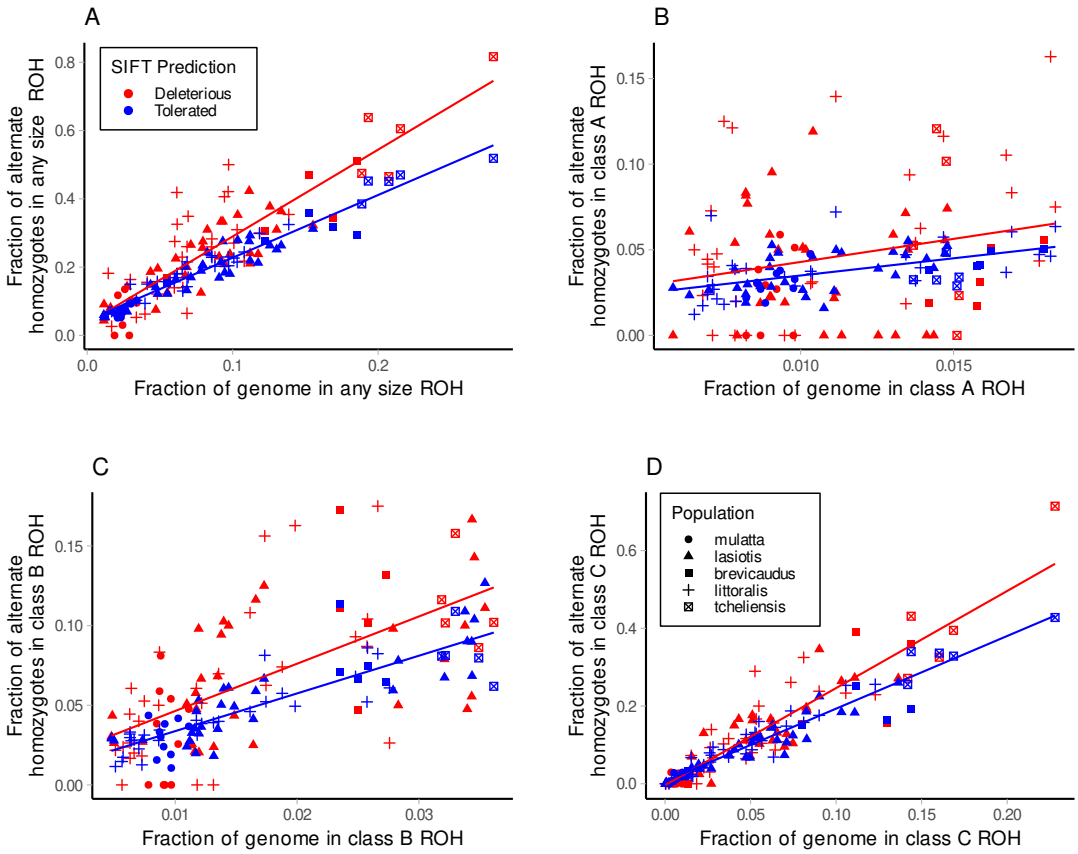

Figure 4: Fraction of genome-wide alternate homozygotes falling in ROH versus the fraction of the genome covered by ROH. Each point is an individual, blue points represent damaging homozygotes and red points represent tolerated homozygotes. (A) Any size ROH, (B) Short ROH, (C) Medium ROH and (D) Long ROH.

Figures 4B-D represent the fraction of deleterious and tolerated homozygotes falling within

short, medium, or long ROH, versus the fraction of genome covered by short, medium and long ROH, respectively. For each size class, we observe the expected positive correlation between the fraction of homozygotes occurring in ROH and the fraction of the genome covered by ROH (Pearson $r_{F^d_{i,A}-G_{i,A}}$ = 0.236, Pearson $r_{F^t_{i,A}-G_{i,A}}$ = 0.54, Pearson $r_{F^d_{i,B}-G_{i,B}}$ = 0.628, Pearson $r_{F^t_{i,B}-G_{i,B}}$ = 0.868, Pearson $r_{F^d_{i,C}-G_{i,C}}$ = 0.907 and Pearson $r_{F^t_{i,C}-G_{i,C}}$ = 0.957). Furthermore, for each class and especially for long ROH, the fraction of deleterious homozygotes falling in ROH seems higher than the fraction of tolerated homozygotes.

We then assessed the statistical difference between the two regression with the following linear model:

$$f_{i,j} = \beta_0 + \beta_1 G_{i,j} + \beta_2 D_i + \beta_3 G_{i,j} D_i + \varepsilon,$$

where the variables are defined as before but for $j \in \{A, B, C\}$ depending on which ROH size class we test. For small and medium ROH, there is no significant difference in intercept (p = 0.944 and p = 0.439 respectively) or slope (p = 0.604 and p = 0.181 respectively). However for long ROH, while there is no significant difference in intercept (p = 0.35), we found a significant difference in slope (β_3 = 0.62870, $p = 3.4 \times 10^{-5}$). This indicates a significant higher enrichment in deleterious homozygotes per unit increase in long ROH compared to tolerated homozygotes. Our results are consistent with the previous work of Szpiech et al. (2013), who also found a higher enrichment of deleterious homozygotes than tolerated ones in long ROH (Szpiech et al. 2013).

Loss of Function Variants

As a critical part of our work relies on predicting whether a variant is deleterious or tolerated, and while these sets of variants are likely to represent, in aggregate, mutations that are deleterious or tolerated, we wish to focus now on loss-of-function polymorphisms that are *a priori* likely to be deleterious, even highly deleterious. We identified 179 variants annotated as either stop-loss or stop-gain to include in our loss-of-function category (LOF). We compared these to our set of SIFT-predicted tolerated sites (n=14,913).

In Figure 5, we plot the distribution of LOF variants across the five sub-species. In our sample, individuals carry between 16 and 34 loss-of-function in heterezygous state, *M. m. tcheliensis* having the least (Figure 5A). More remarkably, individuals carry on average between 2 and 8 LOF variants in homozygous state (Figure 5B), one individuals found in *M. m. littoralis* having up to 12 LOF homozygotes. This is noteworthy given that LOF variants are likely to affect protein function and thus impact individuals fitness. Finally, the genetic load (2*number of homozygotes + number of heterozygotes) decreases as we move from *M. m. mulatta*, *M. m. lasiotis*, *M. m. brevicaudus*, *M. m. littoralis* and to *M. m. tcheliensis*.

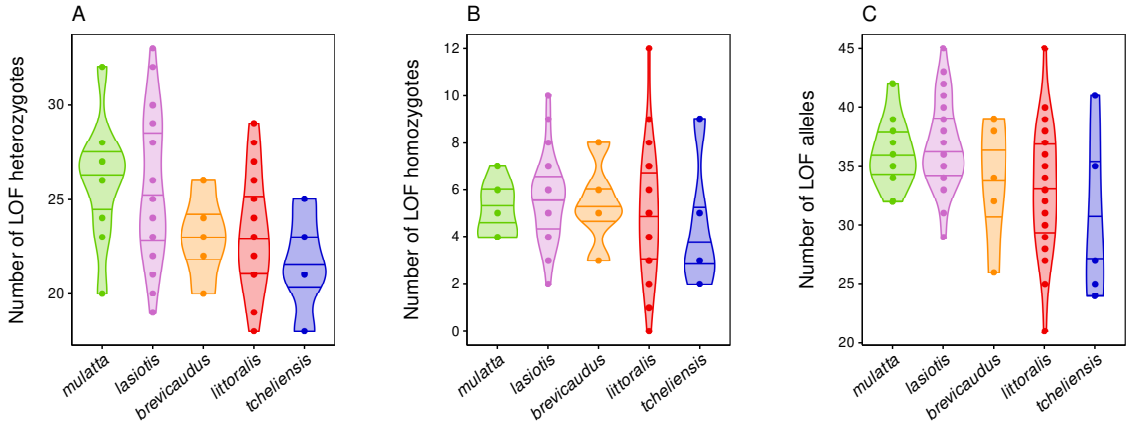

Figure 5: Distribution of the Number of (A) LOF heterozygotes, (B) LOF alternate homozygotes and (C) LOF alleles predicted by SIFT (Vaser et al. 2016) for each sub-species.

For both tolerated and loss-of-function variants, we counted the number of polymorphic sites homozygous for the alternate allele occurring inside ROH. Then, in the same way as done above, we computed, for each individual i, the fraction of tolerated and loss-of-function homozygous occurring within ROH, denoted $F^t_{i,j}$ and $F^l_{i,j}$ respectively, for any size ROH ($j = R$) and for each class ($j \in \{A, B, C\}$) (see Supplementary Tables S3).

Because our number of loss-of-function variants is much smaller than our number of tolerated ones, instead of plotting $F^t_{i,j}$ and $F^l_{i,j}$ against $G_{i,j}$, we partitioned our individuals into two groups based on their ROH genome coverage. For any size ROH and each class ROH, we classified individuals as Low-ROH if the fraction of their genome covered by ROH was lower than the average fraction across individuals, and classified as High-ROH if the fraction of their genome

covered by ROH was higher than the average across individuals. We then plotted the distribution of the fraction of loss of function and tolerated variants occurring in ROH for each category of individuals.

Figure 6 shows the fraction of tolerated and loss-of-function homozygotes falling outside or inside ROH regions, for any size ROH and each class ROH, and for individuals classified as having a low ROH genome coverage (Low ROH) and individuals classified as having a high ROH genome coverage (High ROH). Pairwise t-test were performed to assess the significance of the difference between loss of function variants and tolerated ones.

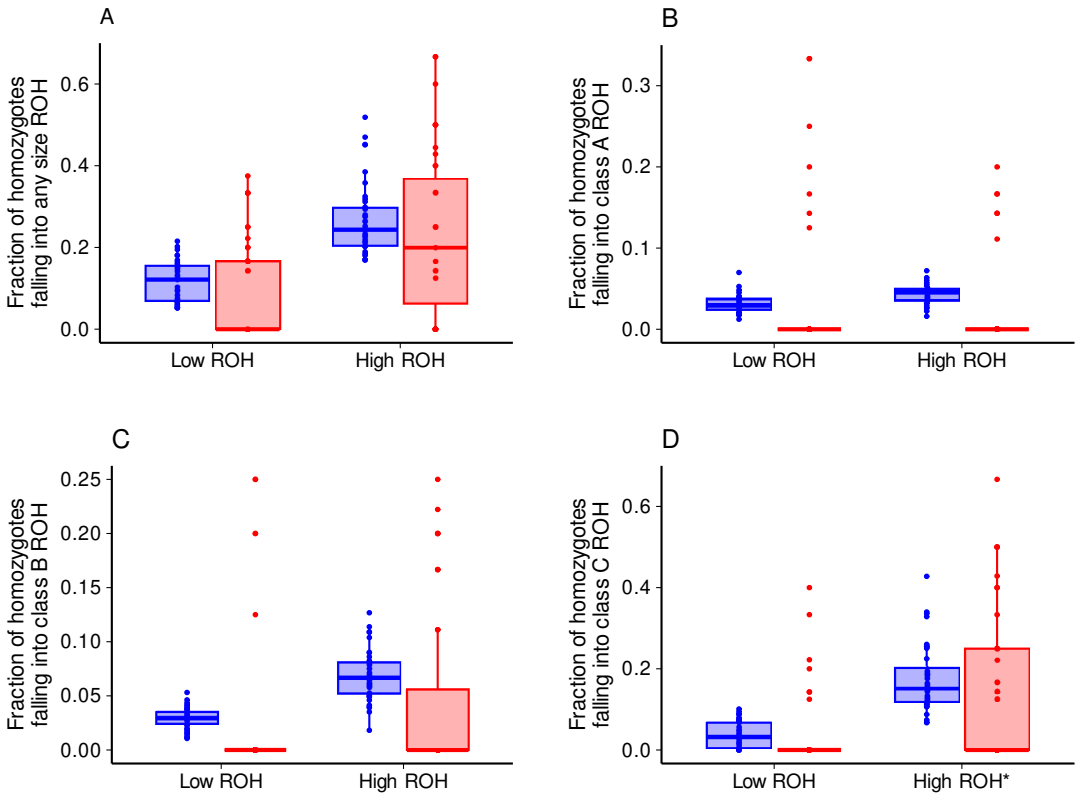

Figure 6: Fraction of genome-wide alternate homozygotes falling in ROH for individuals with a small ROH genome coverage (Low ROH) and individuals with a high ROH genome coverage (High ROH). The red boxplots refers to the LOF variants, and the blue bloxplots refers to the tolerated variants. The * indicates a p-value equal or lower than 0.05.

Whether we distinguish the size of ROH or not, we observe in Figure 6 a higher fraction of homozygotes within ROH in individuals categorized as having a high ROH genome coverage than

in individuals having a low ROH coverage. Once again, this was expected, as more variants are predicted to fall within ROH as the total length of ROH in the genome increases. For any size ROH, medium size ROH and long ROH, there is no significant difference between the fraction of loss of function homozygotes occurring in ROH and the fraction of tolerated variants, for Low-ROH and High-ROH individuals. We observe however a significantly lower fraction of loss-of-function variants in small ROH compared to the fraction of tolerated homozygotes in individuals that were classified as having a high fraction of their genome covered by small ROH (p=0.05).

Deleterious sites in regions under positive selection

The maintenance of mildly deleterious variation in populations can also be caused by positive natural selection, as regions linked to advantageous loci can increase in frequency in the population via genetic hitchhiking (Barton 2000; Charlesworth 2012). Therefore, we further investigated if some deleterious mutations could have been maintained by genetic hitchhiking. We identified regions under selection by using the nSL statistic (Ferrer-Admetlla et al. 2014) as implemented in selscan (Szpiech and Hernandez 2014) in each sub-species separately. Between 237 and 254 windows of 100 kb were identified to be under selection (see Supplementary Tables S5-9). We then partitioned deleterious variants in two sub-sets, the ones falling outside regions under selection and the ones falling inside, and computed the site frequency spectrum (SFS) on each of these sub-sets, to test for an excess of high-frequency deleterious sites (see Supplementary Figure S3). We identified 7 deleterious mutations at high frequency (>0.7) in regions under selection and across the 5 sub-species. Their position, the genes involved and their associated functions are summarized Table 1.

We found that both *M. m. mulatta* and *M. m. tcheliensis* harbour one private mutation at high frequency in selected regions, on chromosome 1 and 8 respectively, involving an olfactory receptor (gene OR11L1) in the case of *M. m. mulatta* and an oxygen regulated I (gene RP1) for *M. m. tcheliensis*. The three sub-species *M. m. mulatta*, *M. m. lasiotis* and *M. m. littoralis* share one high-frequency variant in selected regions on chromosome 14 at position 59,880,975, involved

with the gene OR56A1 that codes an olfactory receptor. Finally and interestingly, we found in *M. m. brevicaudus* four variants at high-frequency on chromosome 14 in a window of less than 1Mb, involved with the genes CEP126 (coding a DNA metabolism protein), BIRC2 (a protease inhibitor), and MMPI2 (a metalloprotease).

Table 1: High frequency deleterious sites in genomic regions identified to be under selection and the gene ontologies involved

Site position	**Subspecies**	**Gene**	**PANTHER Family/ Subfamily**	**PANTHER Protein Class**
chr1:1218571	*mulatta*	ENSMMUG-00000056211	NA	NA
		OR11L1	Olfactory receptor (PTHR24242:SF386)	G-protein coupled receptor (PC00021)
chr8:55470846	*tcheliensis*	RP1	Oxygen-regulated 1 (PTHR23005:SF4)	NA
chr14:59880975	*mulatta, lasiotis, littoralis*	OR56A1	Olfactory receptor 56A1 (PTHR26450:SF425)	Transmembrane signal receptor(PC00197)
chr14:95238067 chr14:95248391	*brevicaudus*	CEP126	Centrosomal protein of 126 KDA (PTHR31191:SF4)	DNA metabolism protein (PC00009)
chr14:95618606	*brevicaudus*	BIRC2	Baculoviral IAP repeat-containing 2 (PTHR10044:SF79)	Protease inhibitor (PC00191)
chr14:96147958	*brevicaudus*	MMP12	Macrophage metalloelastase (PTHR10201:SF267)	Metalloprotease (PC00153)

Discussion

Patterns of ROH

In our data set, that consisted of 79 Chinese rhesus macaques across the five sub-species, we identified patterns of ROH and showed that the sub-species carry different burden of ROH. *M. m. mulatta* carry the smallest burden in ROH, both in total length (SROH) and number (NROH), whereas *M. m. tcheliensis* carry the highest length and number of ROH, followed by *M. m. brevicaudus*. Amounts of ROH are intermediate for the two other sub-species, with individuals carrying heterogeneous length and number of ROH (though this variance might be an effect of sampling ?). These patterns are consistent what is known of the evolutionary history of Chinese rhesus macaques. Indeed, Chinese rhesus macaques are thought to have diverged from Indian rhesus macaques around 162 kya, colonizing the South of China while Indian rhesus macaques moving toward the West. Liu et al. (2018) inferred the demographic history of Chinese rhesus macaques and suggested a step-by-step divergence scenario, with sub-species diverging from the ancestral population successively in the following order: *M. m. mulatta* (that would have first diverged from the ancestral population around 126 kya), followed by *M. m. lasiotis* 104 kya, then *M. m. brevicaudus* 62 kya, and finally *M. m. littoralis* and *M. m. tcheliensis* diverging from each other 51 kya. Therefore, under this scenario, *M. m. mulatta* would have gone through only one bottleneck, whereas *M. m. tcheliensis* would have gone through four bottlenecks. As episodes of small population size promote homozygosity, this is consistent with the significant higher burden in ROH, especially small and medium size ROH, in *M. m. tcheliensis* compared to *M. m. mulatta*. The highest total length of long ROH in *M. m. tcheliensis* would more likely result from recent smaller population size, which is consistent with smallest estimated genetic diversity in this sub-species, compared to the other sub-species, especially *M. m. mulatta* which harbours the highest genetic diversity. Regarding the other three subspecies, given the demographic history inferred by Liu et al. (2018), we would have expected to observe shorter and less ROH for *M. m. lasiotis*, followed by *M. m. brevicaudus* and then *M. m. littoralis*, as they went through an increasing number of bottlenecks.

This lack of consistency could be due to the two individuals Liu *et al.*. 2018 added to their sample, as one belonged to *M. m. lasiotis* and the other to *M. m. littoralis*, or a different filtering our the data, as Liu et al. (2018) filtered their data in a less stringent way as we did. In any case, the wide range of total length of ROH across our data set provides some robustness to our analysis.

Long ROH are enriched in mildly deleterious homozygotes

In accordance with previous work (Szpiech et al. 2013, 2019; Pemberton and Szpiech 2018; Woods et al. 2006; Sams and Boyko 2019; Zhang et al. 2015), our results show that long ROH are enriched in deleterious homozygotes variation. Intuitively, when the fraction of a genome covered by ROH increases, the genome gets enriched in polymorphic sites in homozygous state. We have demonstrated here that this enrichment is higher for deleterious alleles than tolerated ones in long ROH. Because deleterious variants tend to be present in low frequency in a population, it is more likely for two deleterious alleles to get together in homozygous form by being inherited together from a common recent ancestor than by chance. This is also true for tolerated alleles, but to a lesser extant given that they are likely to be present in a population at a higher frequency. Thus, the probability that two tolerated alleles are paired together in an individual is higher even outside of ROH.

As long ROH are likely to result from mating between two closely related individuals, these results suggest that recent inbreeding is the process driving the enrichment in deleterious homozygotes in inbred populations. These observations corroborate what has been known for a long time as inbreeding depression (Darwin 2010; Woods et al. 2006). However, as was suggested by Szpiech et al. (2013), these results also imply that the damaging effect of inbreeding on fitness could be more subtle (Szpiech et al. 2013). By gathering recessive damaging alleles in homozygous form, inbreeding allows these alleles to affect fitness. Therefore, patterns of ROH are specifically important for recessive damaging alleles, because these alleles are not likely to be paired together outside of ROH. Furthermore, it suggests that recent inbreeding can shift the zygosy of damaging alleles in only a few generations. In conservation genetics, a classic example is the Isle Royale gray wolf population, highly inbred because of an isolated and small size population as well as an

extreme bottleneck, and which suffers from inbreeding depression due to the presence of a high number of deleterious homozygotes in their genomes (Räikkönen et al. 2009; Hedrick et al. 2014; Robinson et al. 2019).

Small ROH are impoverished in strongly deleterious homozygotes

In the second analyis, we found a significant lower fraction of loss-of-function homozygotes falling withing small ROH than the fraction of tolerated ones. This observation could suggest that, because loss of function variants are strongly deleterious, they might have been purged by natural selection. The fact that this is observed only in small ROH might be because, as these ROH were probably generated a long time ago, natural selection would have had the time to purge them, whereas in long and intermediate ROH, it has not. Furthermore, we do not find a significantly higher fraction of deleterious homozygotes in small ROH compared to tolerated ones, which could be because their effect on fitness is not strong enough for natural selection to purge them in the same amount of time as the strongly damaging alleles. Therefore, under the condition of being deleterious enough, inbreeding could allow the purging of these variants from our populations after some time. This has been observed in other genomic studies, for instance in another primate, the mountain gorilla (Xue et al. 2015), Indian tigers (Khan et al. 2021), Alpine ibex (Grossen et al. 2020) or other species (Ochoa and Gibbs 2021; Xie et al. 2022; Kleinman-Ruiz et al. 2022).

It is important to note that our results were obtained with a small sub-set of loss-of-function variants (n=179), which reduces the power of our analysis. Furthermore, if our case suggests a positive effect of inbreeding, this would be conditioned by several factors, such as that the population does not suffer from local extinction due to demographic stochasticity during the bottleneck, or that the overall fitness is not affected too much by the recessive mildly deleterious homozygotes not purged effectively (Robinson et al. 2023), or that the overall lack of genetic diversity does not weaken population adaptive potential facing future environmental changes.

Some deleterious variants are maintained via genetic hitchhiking

When looking for signs of high frequency deleterious variants that would be maintained by genetic hitchhiking (or, possibly, by "deleterious sweeps" (Johri et al. 2021), we found across the five sub-species several mutations that fell within regions under selection. Because variants classified by SIFT as deleterious are mutations that are likely to affect protein function (Vaser et al. 2016), these mutations could be either advantageous or detrimental for the individual fitness. In the case of *M. m. mulatta*, *M. m. lasiotis*, *M. m. littoralis* and *M. m. tcheliensis*, each of the variants we identified is alone on its respective chromosome. In the case of *M. m. brevicaudus* however, we identified on chromosome 14 four deleterious mutations in a region shorter than 1 Mb. It might be that these variants are under positive selection, or it could be also that some of them are detrimental and are maintained by linkage disequilibrium. Especially, BIRC2 was identified to be under positive selection by Liu et al. (2018) and could be the gene driving the maintenance of the other variants at a high frequency.

Significance for the use of rhesus macaques in medical research

Rhesus macaques are the most common non-human model in bio-medical research (Rogers 2022; Gibbs et al. 2007; Phillips et al. 2014). This species diverged from the ancestor of present humans around 25 million years ago, and is the closest relative to humans after apes (Roos et al. 2019; Gibbs et al. 2007). They cover the widest geographical range of all primates after humans, and show adaptability to many different environments making it one of the most ecologically successful primate species (Maestripieri and Hoffman 2012; Rogers 2022). Their close genetic and ecological proximity to humans are one of the two main features that explain their extended use as non-human primate model.

Our results suggest that the different evolutionary histories that the five sub-species of *macaca mulatta* have gone through, have led to differences in the distribution of ROH and thus deleterious variation between the sub-species, and also suggest that there is some variability between individuals of the same sub-species. These findings could potentially impact medical research, as they

imply that, in experimental design, primates from different backgrounds could potentially react differently to experimental tests. Therefore, we suggest that the precise geographic origin of primates used in experiments should be assessed beforehand, as it has previously been advanced by Vallender and Miller (2013). This is consistent with several studies in humans that have demonstrated the effect of ancestry on phenotypic response to environmental stress, predisposition to certain diseases and susceptibility to pharmaceutical drugs (Mak et al. 2018; Cheng et al. 2009; Torgerson et al. 2012).

Conclusion

Our results add understanding to how damaging alleles are transmitted and maintained in populations of wild Chinese rhesus macaques, and how evolutionary processes such as demography and natural selection can shape the distribution of deleterious variation. Recent inbreeding is an important driver in the gathering of many recessive deleterious allele in homozygous form, which is particularly relevant for the study of complex diseases. In particular, this highlights the importance for medical research to take into account individuals' ethnicity and genetic background when diagnosing diseases and treating patients. Furthermore, there is a need to deeper understand the effect of small size population on the maintenance or purging of strongly deleterious variants in small and isolated wild populations, in order to better implement conservation efforts, such as animal translocations. Finally, on a more general note, this study gives insights in Chinese rhesus macaques genome, in relation to its evolutionary history.

Material and Methods

Data preparation

79 wild-born Chinese rhesus macaques were sequenced using whole-genome sequencing by Liu et al. (2018) from 17 sites in China, across the five different sub-species of *Macaca mulatta*: *M. m.*

brevicaudus (n=5), *M. m. mulatta* (n=10), *M. m. littoralis* (n=28), *M. m. lasiotis* (n=31) and *M. m. tcheliensis* (n=5). Liu et al. (2018) called all multi-allelic polymorphic sites following GATK best practices using the reference genome rheMac8, and identified 52,534,348 passing polymorphic autosomal sites. For our analysis, we filtered multi-allelic SNPs and indels, and set to missing the genotypes with a genotype quality (GQ) lower than 20. Furthermore, we filtered the sites with more than 10% missing genotypes across the 79 individuals and were left with 7,001,681 variants. An updated reference genome assembly for *M. mulatta* was recently released, increasing the sequence contiguity 120-fold and providing a better understanding of the rhesus macaque genome (Warren et al. 2020). For this reason, we decided to lift over our variants on the reference genome Mmul_10, which was achieved using LiftOver from the University of California Santa Cruz (UCSC) Genome browser. Finally, our data set consisted in 5,713,999 variants.

Annotating variants with SIFT

The effect of variants on individual's phenotype was predicted using SIFT 4G (Vaser et al. 2016). SIFT (Sorting Intolerant From Tolerant) is an algorithm that classifies a variant as synonymous or non-synonymous (non-synonymous, stop-loss, stop-gain, stop-loss), and predicts whether the amino acid substitution is deleterious to protein function. No database was available for the reference genome Mmul_10 so we had to created it using the genome and gene annotations from UCSC Genome Browser, and the UniRef90 protein data base (Uniprot). We used the multi-transcript option, meaning that for one substitution, the effect on all transcripts were predicted. We filtered low-quality predictions, and when two or more transcripts of the same gene were given opposite predictions, we kept the worst prediction. Furthermore, SIFT does not provide a prediction for variants identified as stop-loss, stop-gain and start-loss and we therefore manually classified them as loss-of-function.

In total, SIFT annotated with enough confidence 17,477 variants across our 79 individuals (Figure 7). Among these, 8,761 variants were annotated as synonymous, 8,537 as non-synonymous, 165 as stop-gain, 14 as stop-loss and none were annotated as start-loss. 14,913 substitutions were

predicted to be tolerated and 2,385 were predicted to be deleterious. We classified the stop-loss and stop-gain variants as loss-of-function variants, and it concerned 179 sites in total.

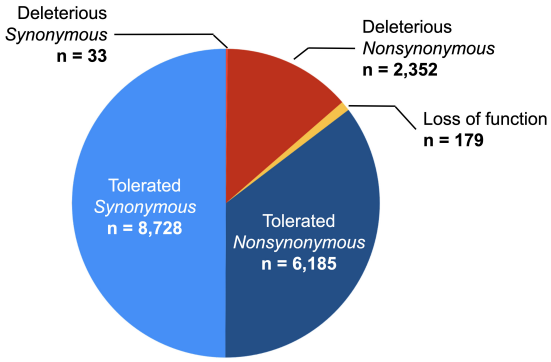

Figure 7: SIFT (Vaser et al. 2016) classification of our set of 5,713,999 after filtering. The loss-of-function category pools the variants predicted to be stop-loss, stop-gain and start-loss.

Calling ROH with GARLIC

ROH were called using the program GARLIC, a model-based approach that was implemented by Szpiech et al. (2017). In addition to perform ROH calling, GARLIC also classifies these ROH into classes according to their length (Szpiech et al. 2017). It relies on an ROH calling algorithm introduced by Pemberton et al. (2012) that uses the logarithm of the odds (LOD) score measure of autozygosity in a sliding-window along the genome (Szpiech et al. 2017; Pemberton et al. 2012). We let GARLIC choose the window size, starting with a window size of 100 nucleotides and increasing the size by 10 nucleotides, the default value. We let GARLIC choose for the number of classes into which classify ROH as well. Furthermore, we specified a number of re-sampling equal to 40 to infer alleles frequencies. Finally, the centromere positions for built Mmul_10 were retrieved from the UCSC genome browser. All other parameters were set to default values, and we called ROH on the 79 individuals (garlic flags –auto-winsize –auto-overlap-frac –winsize 100 –tped input.tped –tfam input.tfam –tgls input.tgls –gl-type GQ –resample 40 –centromere centromere_file.txt –out output_file.txt). For each individual, we obtained the list of ROH with their position on the genome (chromosome, beginning and end), the size class they were classified into, and their length.

Mapping variants onto ROH

In the first step of our analysis, we focused on synonymous and non-synonymous variants. We first split our subset of variants between tolerated and deleterious substitutions. Then we filtered the homozygotes for the reference allele and the heterozygotes. We ended up with a list, for each individual, of the positions of the damaging and tolerated alternate homozygous. We then used the intersect command provided by BEDTools (Quinlan and Hall 2010), that checked for each homozygotes if it was contained in an ROH. If it was, the size class of the ROH was given. For any size ROH and each size class, we counted the number of homozygotes falling inside and outside ROH, for both damaging and tolerated alleles. In the second part of our analysis, we focused on loss of function variants. Stop-loss and stop-gain substitutions have indeed a higher likelihood of being damaging than variants predicted to be deleterious, and we thought interesting to focus on this subset in particular to account for the robustness of our results. We therefore classified these variants as loss of function and compared their distribution to the distribution of tolerated substitutions, applying the same method as described above.

Identifying regions under selection

We first phased our data using the program SHAPEIT (Delaneau et al. 2008) to recover part of missing genotypes. For a lack of genetic map for Mmul_10, we assumed a recombination rate constant over the genome using the –rho option with a value of 0.00017043. This value was computed assuming a recombination rate per site per generation of 5.126×10^{-9} (Szpiech et al. 2021), and the average effective population size over the 5 sub-species of rhesus macaques using our estimates of genetic diversity (Supplementary Table S4) and assuming a mutation rate of 1.08×10^{-8} (Liu et al. 2018). We then identified regions under selection independently for each sub-species with the program selscan (Szpiech and Hernandez 2014) using the nSL statistics (Ferrer-Admetlla et al. 2014) (selscan flags –nsl –thread 8 –vcf file.vcf). We then normalized the scores using the genome-wide empirical background with selscan's norm function (norm flags –nsl –files file.vcf –bp-win) (Szpiech and Hernandez 2014). We kept the windows with a normalized nSL

score in the top 1%. Chromosomal regions under selection were then annotated with the Ensembl Genes database (Martin et al. 2023) using the UCSC Table Browser (Nassar et al. 2023; Karolchik et al. 2004).

Identifying high frequency deleterious variants in regions under selection

Deleterious variants present in the regions under selection were identified using BEDTools (Quinlan et al. 2010). We then computed the site-frequency-spectrium (SFS) on variants inside and outside these regions. We identified the sites at high frequency by keeping the alternate alleles present at a frequency equal or higher than 0.7 in each sub-species. These sites were annotated with the Ensembl database, and the genes ontologies were extracted with PANTHER (Thomas et al. 2022). To be able to compare the SFS of the different sub-species (see Figure S3), we re-sampled the SFS with the smallest sample size, 10 haploid individuals, using the following equation:

$$S'_j = \sum_{i=j}^{n-1} S_i \frac{\binom{i}{j}\binom{n-i}{h-j}}{\binom{n}{h}}$$

where S is the original SFS of size n and S_i the number of snps carried by i (haploid) individuals, S' is the re-sampled SFS of size h and S'_j the number of variants carried by j individuals, and with $h \leq n$.

Data Availability

Macaque whole genome VCFs are available at http://doi.org/10.5524/100484.

Acknowledgements

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM146926 (ZAS) and by start-up funds from the Pennsylvania State University's Department of Biology (ZAS). Computations for this research

were performed using the Pennsylvania State University's Institute for Computational Data Sciences' Roar supercomputer.

References

Barton NH. 2000. Genetic hitchhiking. Philos Trans R Soc Lond B Biol Sci. 355:1553–1562.

Benazzo A, et al.. 2017. Survival and divergence in a small group: The extraordinary genomic history of the endangered apennine brown bear stragglers. Proc Natl Acad Sci USA. 114:E9589–E9597.

Bertorelle G, et al.. 2022. Genetic load: genomic estimates and applications in non-model animals. Nat Rev Genet. 23:492–503.

Broman KW, Weber JL. 1999. Long homozygous chromosomal segments in reference families from the centre d'etude du polymorphisme humain. Am J Hum Genet. 65:1493–1500.

Brüniche-Olsen A, Kellner KF, Anderson CJ, DeWoody JA. 2018. Runs of homozygosity have utility in mammalian conservation and evolutionary studies. Conserv Genet. 19:1295–1307.

Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. 2018. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet. 19:220–234.

Charlesworth B. 2012. The effects of deleterious mutations on evolution at linked sites. Genetics. 190:5–22.

Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. Nat Rev Genet. 10:783–796.

Cheng CY, et al.. 2009. Admixture mapping of 15,280 african americans identifies obesity susceptibility loci on chromosomes 5 and x. PLOS Genet. 5:e1000490.

Consortium GP. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature. 491:56–65.

Darwin C. 2010. *The variation of animals and plants under domestication*, volume 2. Cambridge University Press.

Delaneau O, Coulonges C, Zagury JF. 2008. Shape-it: new rapid and accurate algorithm for haplotype inference. BMC Bioinformatics. 9:1–14.

Eyre-Walker A, Keightley PD. 1999. High genomic deleterious mutation rates in hominids. Nature. 397:344–347.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. Mol Biol Evol. 31:1275–1291.

Fu W, et al.. 2013. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature. 493:216–220.

Gibbs RA, et al.. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science. 316:222–234.

Grossen C, Guillaume F, Keller LF, Croll D. 2020. Purging of highly deleterious mutations through severe bottlenecks in alpine ibex. Nat Commun. 11:1001.

Hedrick PW, Kalinowski ST. 2000. Inbreeding depression in conservation biology. Annu Rev Ecol Syst. 31:139–162.

Hedrick PW, Peterson RO, Vucetich LM, Adams JR, Vucetich JA. 2014. Genetic rescue in isle royale wolves: genetic analysis and the collapse of the population. Conserv Genet. 15:1111–1121.

Johri P, Charlesworth B, Howell EK, Lynch M, Jensen JD. 2021. Revisiting the notion of deleterious sweeps. Genetics 219:iyab094.

Karolchik D, et al.. 2004. The ucsc table browser data retrieval tool. Nucleic Acids Res. 32:D493–D496.

Keinan A, Mullikin JC, Patterson N, Reich D. 2007. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in east asians than in europeans. Nat Genet. 39:1251–1255.

Keller MC, Visscher PM, Goddard ME. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 189:237–249.

Keller MC, et al.. 2012. Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. PLOS Genet. 8:e1002656.

Khan A, et al.. 2021. Genomic evidence for inbreeding depression and purging of deleterious genetic variation in indian tigers. Proc Natl Acad Sci USA. 118:e2023018118.

Kirin M, et al.. 2010. Genomic runs of homozygosity record population history and consanguinity. PLOS ONE. 5:e13996.

Kleinman-Ruiz D, et al.. 2022. Purging of deleterious burden in the endangered iberian lynx. Proc Natl Acad Sci USA. 119:e2110614119.

Kondrashov AS. 1995. Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? J Theor Biol. 175:583–594.

Laurent R, et al.. 2023. A genetic and linguistic analysis of the admixture histories of the islands of cabo verde. eLife 12:e79827.

Leon-Apodaca AV, et al.. 2023. Genomic consequences of isolation and inbreeding in an island dingo population. bioRxiv. p. 2023.09.15.557950.

Liu Z, et al.. 2018. Population genomics of wild chinese rhesus macaques reveals a dynamic demographic history and local adaptation, with implications for biomedical research. GigaScience. 7:giy106.

Lohmueller KE. 2014. The distribution of deleterious genetic variation in human populations. Curr Opin Genet Dev. 29:139–146.

Lohmueller KE, et al.. 2008. Proportionally more deleterious genetic variation in european than in african populations. Nature. 451:994–997.

Maestripieri D, Hoffman CL. 2012. Behavior and social dynamics of rhesus macaques on cayo santiago. In: *Bones, genetics, and behavior of rhesus macaques*, Springer, pp. 247–262.

Mak AC, et al.. 2018. Whole-genome sequencing of pharmacogenetic drug response in racially diverse children with asthma. Am J Respir Crit Care Med. 197:1552–1564.

Martin FJ, et al.. 2023. Ensembl 2023. Nucleic Acids Res. 51:D933–D941.

McQuillan R, et al.. 2008. Runs of homozygosity in european populations. Am J Hum Genet. 83:359–372.

Mooney JA, Yohannes A, Lohmueller KE. 2021. The impact of identity by descent on fitness and disease in dogs. Proc Natl Acad Sci USA. 118.

Morton NE, Crow JF, Muller HJ. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. Proc Natl Acad Sci USA. 42:855–863.

Nassar LR, et al.. 2023. The ucsc genome browser database: 2023 update. Nucleic Acids Res. 51:D1188–D1195.

Ochoa A, Gibbs HL. 2021. Genomic signatures of inbreeding and mutation load in a threatened rattlesnake. Mol Ecol. 30:5454–5469.

Ohta T. 1973. Slightly deleterious mutant substitutions in evolution. Nature. 246:96–98.

Ohta T. 1992. The nearly neutral theory of molecular evolution. Annu Rev Ecol Syst. 23:263–286.

Pemberton TJ, Szpiech ZA. 2018. Relationship between deleterious variation, genomic autozygosity, and disease risk: insights from the 1000 genomes project. Am J Hum Genet. 102:658–675.

Pemberton TJ, et al.. 2012. Genomic patterns of homozygosity in worldwide human populations. Am J Hum Genet. 91:275–292.

Phillips KA, et al.. 2014. Why primate models matter. Am J Primatol. 76:801–827.

Quinlan AR, Hall IM. 2010. Bedtools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26:841–842.

Räikkönen J, Vucetich JA, Peterson RO, Nelson MP. 2009. Congenital bone deformities and the inbred wolves (canis lupus) of isle royale. Biol Conserv. 142:1025–1031.

Robinson J, Kyriazis CC, Yuan SC, Lohmueller KE. 2023. Deleterious variation in natural populations and implications for conservation genetics. Annu Rev Anim Biosci. 11:93–114.

Robinson JA, et al.. 2016. Genomic flatlining in the endangered island fox. Curr Biol. 26:1183–1189.

Robinson JA, et al.. 2019. Genomic signatures of extensive inbreeding in isle royale wolves, a population on the threshold of extinction. Sci Adv. 5:eaau0757.

Rogers J. 2022. Genomic resources for rhesus macaques (macaca mulatta). Mamm Genome. 33:91–99.

Roos C, Kothe M, Alba DM, Delson E, Zinner D. 2019. The radiation of macaques out of africa: Evidence from mitogenome divergence times and the fossil record. J Hum Evol. 133:114–132.

Rudan I, et al.. 2003. Inbreeding and risk of late onset complex disease. J Med Genet. 40:925–932.

Sams AJ, Boyko AR. 2019. Fine-scale resolution of runs of homozygosity reveal patterns of inbreeding and substantial overlap with recessive disease genotypes in domestic dogs. G3: Genes Genomes Genet. 9:117–123.

Spielman D, Brook BW, Frankham R. 2004. Most species are not driven to extinction before genetic factors impact them. Proc Natl Acad Sci USA. 101:15261–15264.

Szpiech ZA, Blant A, Pemberton TJ. 2017. Garlic: genomic autozygosity regions likelihood-based inference and classification. Bioinformatics. 33:2059–2062.

Szpiech ZA, Hernandez RD. 2014. selscan: an efficient multithreaded program to perform ehh-based scans for positive selection. Mol Biol Evol. 31:2824–2827.

Szpiech ZA, Novak TE, Bailey NP, Stevison LS. 2021. Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. Evol lett. 5:408–421.

Szpiech ZA, et al.. 2013. Long runs of homozygosity are enriched for deleterious variation. Am J Hum Genet. 93:90–102.

Szpiech ZA, et al.. 2019. Ancestry-dependent enrichment of deleterious homozygotes in runs of homozygosity. Am J Hum Genet. 105:747–762.

Thomas PD, et al.. 2022. Panther: Making genome-scale phylogenetics accessible to all. Protein Sci. 31:8–22.

Torgerson DG, et al.. 2012. Case-control admixture mapping in latino populations enriches for known asthma-associated genes. J Allergy Clin Immunol. 130:76–82.

Uniprot. 2021. Uniprot: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49:D480–D489.

Vallender EJ, Miller GM. 2013. Nonhuman primate models in the genomic era: a paradigm shift. ILAR J. 54:154–165.

Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. 2016. Sift missense predictions for genomes. Nat Protoc. 11:1–9.

Verweij KJ, et al.. 2014. The association of genotype-based inbreeding coefficient with a range of physical and psychological human traits. PLOS ONE. 9:e103102.

Warren WC, et al.. 2020. Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility. Science. 370:eabc6617.

Woods CG, et al.. 2006. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. Am J Hum Genet. 78:889–896.

Xie HX, et al.. 2022. Ancient demographics determine the effectiveness of genetic purging in endangered lizards. Mol Biol Evol. 39:msab359.

Xue Y, et al.. 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science. 348:242–245.

Yengo L, Wray NR, Visscher PM. 2019. Extreme inbreeding in a european ancestry sample from the contemporary uk population. Nat Commun. 10:3719.

Zhang Q, Guldbrandtsen B, Bosse M, Lund MS, Sahana G. 2015. Runs of homozygosity and distribution of functional variants in the cattle genome. BMC Genom. 16:1–16.