Supplementary Material

Natural Selection in the Great Apes

Alexander Cagan, Christoph Theunert, Hafid Laayouni, Gabriel Santpere, Marc Pybus, Ferran Casals, Kay Prüfer, Arcadi Navarro, Tomas Marques-Bonet, Jaume Bertranpetit, Aida M. Andrés

Section 1: Sample processing

Gabriel Santpere, Alexander Cagan

1. Samples

1. Samples

The dataset we analyzed consists of whole-genome autosomal sequences from 83 individuals across all the major lineages of the *Hominidae* (with the exception of *G.b. beringei*) (Figure 1 & supplementary table S1). The dataset was originally presented in Prado-Martinez et al. (Prado-Martinez et al. 2013; SOM), where the SNP calling pipeline and filtering criteria are described in detail. All reads are mapped to the human reference genome (hg18). This approach has three main advantages. First, we take advantage of the extensive data-quality exploration and filtering performed in the original publication. Second, mapping to the human genome ensures that all species are mapped to a high-quality genome, avoiding the (hard to account for) biases that would result from mapping to genomes of low and different qualities. Third, the human genome has the most comprehensive annotation of gene coding regions, which is a central object of this study.

1.1. Filtering strategy

Genomic data is prone to many artifacts that can bias downstream analyses. To account for this we applied a comprehensive series of filters which are summarized in supplementary fig S1. To avoid errors introduced by miss-mapping caused by multi-copy sequences or structural variants among species, we restricted all analyses to genomic intervals with a unique mapping to the human genome. Using UCSC tracks we excluded from analysis all repetitive regions identified by RepeatMasker (~248 Mb), if repeat divergence were lower than 10%, and from Tandem Repeat Finder (~38 Mb). In both cases we only masked repeats longer than 80 bp, since repeats shorter than reads can be better mapped by their flanking non-repetitive sequence. We also masked segmental duplications (~154 Mb) and genomic gaps (~226 Mb). Finally, we also excluded structural variants detected in any of the great ape lineages (~334 Mb). These regions were detected in a previous study of structural variation in great apes which analysed this dataset (Sudmant et al., 2013).

Exclusion of regions that may contain unnanotated structural variants removes a potential source of false positives in our analyses. We expect filtering out structural variants to have a minor effect on our subsequent analyses. As our approach is empirical (except for the MK, where we compare synonymous and non-synonymous mutations and thus removing CNVs would have minimal effect) we do not attempt to estimate the proportion of the genome under natural selection. We expect the tails of the empirical

distributions of our selection tests to be enriched in true candidates of natural selection regardless of the removal of CNVs.

Furthermore, sites with depth of coverage (DP) < (mean read depth/8.0) and DP > (mean read depth * 3), were also removed. To maximize the number of sites to be analyzed we excluded multiple individuals with low coverage (supplementary tables S1 and S2). Some variable sites presented missing data introduced by an Allele Balance Filter used in Prado-Martinez et al. to account for the putative contribution of traces of genetic contamination. These positions were generally excluded particularly in analysis dependent on allele frequencies. Additionally, we also required positions to have at least 5x coverage in all individuals per species. To increase the number of callable sites we excluded low coverage individuals Only the resulting set of sites, which we termed 'callable sites', were used in further analysis. This resulted in a mean of 2,099 Mb of analyzable genome sequence per species (supplementary fig S1).

References

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. *Nature* **499**: 471–5.

Sudmant PH, Huddleston J, Catacchio CR, Malig M, Hillier LW, Baker C, Mohajeri K, Kondova I, Bontrop RE, Persengiev S, Antonacci F. 2013. Evolution and diversity of copy number variation in the great ape lineage. *Genome research* Sep 1;23(9):1373-82.

Section 2: HKA test

Alexander Cagan, Christoph Theunert, Aida M. Andrés

1. The HKA test

We aim to identify candidate targets of natural selection by analyzing the patterns of diversity and divergence in the genomes of each lineage. The neutral model predicts that the ratio of polymorphism to divergence will be approximately constant across the genome, and regions where this ratio is highly skewed in either direction represent potential deviation from neutrality due to selection. The Hudson-Kreitman-Aguade (HKA) test is a neutrality test that contrasts intra-population diversity to inter-populations divergence in a test locus, and compares it with putatively neutral loci (Hudson et al. 1987). By contrasting patterns of diversity and divergence in genomic windows without any differentiation of loci by their function (e.g. coding and non-coding regions), the HKA test is able to detect signatures of selection without an *a priori* assumption that selection will be in a particular type of loci (e.g. coding regions). The ratio of diversity to divergence also allows the HKA test to detect selection at a greater time scale than many SFS-based tests, which lose power once the selected locus has recovered to equilibrium levels of diversity. A previous study comparing the statistical power of selection tests found the HKA test to have the most power to detect positive selection (Zhai et al. 2009). At the genome scale we apply a test based on the HKA by analyzing windows across the genome to identify loci that are outliers of the empirical distribution, either because their ratio of polymorphisms to substitutions is too high or because it is too low when compared with the rest of the genome; these regions represent candidate targets of natural selection in the great apes. Because different types of selection affect differently the ratio of polymorphism to divergence, this test allows us to distinguish among different types of selection. Regions with the lowest HKA scores, where diversity levels are lowest relative to divergence, are candidates for being under positive or negative selection (diversity-reducing types of selection). Regions of the genome where the HKA score is highest, due to high levels of diversity relative to divergence, are suggestive of long-standing balancing selection. Regions of the genome in the middle part of the empirical distribution are expected to be either neutral or subject to varying levels of purifying selection.

When whole-genome data is unavailable, significance of the HKA test has been estimated based on neutral coalescent simulations, which allow us to predict the range of HKA scores expected under neutrality. Here, with availability of full genome data, and under the reasonable assumptions that the majority of the genome is evolving neutrally or under purifying selection, and that a small portion of the genome is evolving under positive and balancing selection, candidate targets of natural selection can be obtained from the tails of the empirical distribution. Regions of the genome with scores at either extremes of the HKA score empirical distribution have the most unusual ratios of polymorphism to divergence, and are thus potential targets of natural selection.

1. Overview of steps taken to avoid and account for possible artifacts

An unusual level of polymorphism may result from technical artifacts and thus not represent true signatures of natural selection. We address this problem by taking several steps to ensure signals are not the result of technical artifacts. Particular attention was taken to artifacts that may result in increased polymorphism, since the HKA is the only test in this study that aims to identify the signatures of balancing selection, but our filtering strategy should be effective in removing artifacts that may result both in high and in low levels of diversity. We present here a summary of the steps we followed to ensure that we detect true biological signals rather than technical artifacts; details can be found in the sections below.

1.1. Pre-test Filtering

The fact that not all populations are analyzed for exactly the same sites can also lead to unequal power for the HKA test. To reduce the effect of these possible biases, data was filtered using the following criteria: for each pair-wise HKA test sites were excluded if they had a read depth of < 5x in any individual in either of the two populations analysed.

1.2. Performing the HKA test

The test, based on the HKA, was calculated in 30kb windows with a 15kb overlap between windows across the genome. For each window and population, we computed the number of polymorphic sites and the number of fixed substitutions to the out-group. Fixed substitutions are positions where both populations are fixed homozygous for alternative alleles. Here, our HKA score for each window was calculated by dividing the number of polymorphic sites (diversity) by the number of fixed substitutions (divergence). Windows with less than 300 callable sites or with less than 6 informative sites (substitutions between populations or within-populations polymorphisms) were removed from the analysis as the scarcity of sites may introduce too much noise**.** H. sapiens was used as an out-group for all tests for consistency, and *P. troglodytes* was used as an out-group for analyzing *H. sapiens*. In *P. troglodytes*, the test was carried out at the sub-populations level, for each of the four *P. troglodytes* sub-populations separately and with *H. sapiens* as out-group. After performing the HKA test we confirm that the regions in the extreme tails of the HKA score distribution are not unusual in terms of their coverage, mapping quality, or the number of duplications they contain (which addresses possible mapping problems as well as gene conversion). We find no systematic bias in any of these categories (see Supplementary Materials HKA 1.3 and 1.4), showing that technical artifacts do not substantially contribute to the regions we detect as candidates of natural selection.

1.3. Coverage Bias

One possible source of unusual diversity levels is extremely high or low coverage in the analyzed window. To ensure that this technical aspect of the data is not substantially contributing to the extremes of the HKA distribution (so we detect true biological signals rather than technical artifacts), we tested whether extreme HKA scores are biased towards lower or higher average coverage than other regions of the genome. To do this, we divided the 30kb windows which are the output of the HKA test into three categories based on their HKA score: regions with HKA scores in the top 1%, bottom 1% and middle 98% of the empirical distribution.

We also ran permutations from each of the three categories to ensure that there is no bias that is being obscured because the middle 98% of the distribution has a much larger number of data points than the top and bottom 1% distributions. For each permutation, 50 regions were randomly selected from each of the distributions and their average coverage was calculated. This was repeated 1000 times with replacement.

These results show no evidence of differences in coverage among the groups of loci with average and with extreme HKA scores (supplementary fig S2 for an example with G.g. gorilla).

1.4. Mapping Quality Bias

To ensure that mapping problems did not affect our analysis, we calculated the average Mapping Quality score across all callable sites for the 30kb windows in the bottom 5%, top 5%, and middle 90% of the genomic distribution of HKA scores (see supplementary fig S3 for an example with G.g. gorilla). All regions have an average Mapping Quality Score within the range of 56-59, well above the mapping quality threshold of 25 used for the dataset, and with no skew towards lower Mapping Quality in the top tail of the distribution, suggesting no systematic bias towards low mapping scores in the extreme tails of the HKA distribution.

2. Percentage of windows containing exons analysis

To explore the results of the HKA test and to investigate the power of the test to detect regions under natural selection we tested if there was an enrichment of windows containing functional elements in the tails of our HKA score distribution. As most of the genome is believed to be non-functional but natural selection targets functional sites, we expect an enrichment of windows containing functional elements in the tails of the HKA empirical distribution.

2.1. Methods

To test this we first created two different lists of functional sites based on the GENCODE annotation of the *H. sapiens* genome. We made one list with the start and end coordinates of all exons classified as being from protein-coding genes. We made another list with the start and end coordinates of all exons classified as being from either RNA (mtRNA , miRNA, Mt_rRNA, misc_RNA, rRNA, snRNA, snoRNA) or non-protein-coding genes. This allowed us to additionally test whether there was a difference in the pattern of HKA scores between protein-coding compared to non-protein-coding regions.

For each list separately, and for each lineage, we annotated the 30kb HKA windows as overlapping (or not overlapping) with an exon by at least 1bp. We then defined different thresholds of the tail of the empirical distribution of the HKA score, and calculated what percentage of windows within each percentile overlap with exons.

220

Results

2.2. Protein-coding exons

225

The results for protein-coding exons show that there is a trend for the number of windows containing protein-coding exons to increase when moving from windows with high HKA scores to windows with low HKA scores, as long as the windows are not in the far tails of the empirical distribution (e.g. from the 99-5% bins, Figure 2 and supplementary table S3). As the lower HKA scores tend to be driven by reduction in 230 diversity rather than an excess of divergence (supplementary fig S4) this suggests that this trend of exonic enrichment towards lower HKA scores is most likely driven by the effect of purifying selection, which are on average stronger in exonic regions of the genome (McVicker et al. 2009).

It is interesting that at the very bottom tail of the HKA empirical distribution all lineages show a slight difference in trends, with the percentage of windows containing protein-coding exons not always 235 increasing further as we move further into the tail (depending on the populations at different points between the bottom 5% and the end of the bottom tail). There could be several reasons for this change, which is observed in all lineages. One explanation could be noise, as the number of regions decreases when moving further in the tail. An alternative explanation is that the most strongly selected targets of selection are non-coding variants in regulatory regions that are not close to genes. A third explanation is that some selective 240 sweeps targeting exonic regions extend far beyond the exons themselves, resulting in an excess of windows in the extreme bottom tail of the HKA empirical distribution that do not contain exons but have low diversity because they are part of a selective sweep on an exonic variant. This is not unexpected as fast selective sweeps can extend very long genomic regions because the short-term effects of recombination do not break the association across variants (Maynard Smith & Haigh, 1974)**.** In fact, we observe an unusually high

clustering of windows among those present in the bottom tail of the HKA empirical distribution (see Supplementary Materials HKA 4)

It is also interesting that in the extreme top tail of the HKA empirical distribution (candidate regions under balancing selection) we observe a change with respect to the overall trend described above. Specifically, we observe a drastic increase in the percentage of windows overlapping exons for the windows more strongly enriched in SNPs, in all the populations. This is unexpected under neutral evolution and under purifying selection, and is likely to be the result of balancing selection acting on or near these protein-coding exons. Also, the signature of long-term balancing selection is narrow due to the long-term effects of recombination (e.g. Kaplan et al. 1988; Charlesworth et al. 1997) and we do not expect several windows to show the signatures of one single event. This conclusion is further supported by the observation that in all lineages many of these genes are from the MHC region, a well-known target of balancing selection across vertebrates (Hughes et al. 1998). Therefore this analysis provides evidence to support the idea that long-term balancing selection is not prevalent across the genome (although we note that we may have limited power to detect its signatures in 30kb windows). Despite the small number of regions, many of them are shared across populations, emphasizing the important and conserved role that balancing selection plays in maintaining adaptive diversity, particularly with regards to immune function. Among the other genes in these windows with high HKA scores are novel candidates for being targets of long-standing balancing selection.

2.3. Non-protein-coding exons

The pattern we observe across the HKA empirical distribution for the percentage of windows overlapping non-protein-coding exons is strikingly different (supplementary fig S5 and supplementary table S4). There is no trend for an increasing percentage of windows overlapping exons as the HKA score decreases. Instead, the percentage of windows overlapping exons is flat along the middle 80% of its distribution, consistent with weaker influence of purifying selection. This may be due to weaker natural selection in non-protein-coding exons, a lower percentage of windows overlapping conserved elements (as non-protein-coding RNAs tend to be shorter than protein-coding ones), or a combination of both.

2.4. The percentage of protein-coding exons in different bins of the HKA empirical distribution correlates with effective population size

We observe some differences between populations in the patterns above, which may be explained by their differences in effective population size (Ne). The effective population size influences the effectiveness of natural selection (Lanfear et al. 2014). For example, with larger Ne slightly deleterious alleles are more effectively removed from the population through negative selection, and advantageous alleles are more likely to increase in frequency through positive selection. A consequence of negative selection is background selection, the removal of linked neutral variation along with truly deleterious alleles due to linkage. Therefore we expect a greater relative deficit in diversity in and around conserved regions in populations with larger effective population sizes as both positive and purifying selection act to lower diversity.

To test whether the differences observed between populations correlate with their effective population size we used two different estimates of Ne previously calculated from this data set (Prado-Martinez et al. 2013). The first estimate of Ne used is an estimate of long-term Ne calculated based on Watterson's estimator (supplementary table S3 & Prado-Martinez et al. 2013). The second estimate of Ne corresponds to the size of the population since the split with its closest population in the dataset. This was previously estimated using the method of PSMC (Li & Durbin 2011) (supplementary table 5 of Prado-Martinez et al. 2013) . We used two different estimates because they measure the effective population size during different evolutionary periods. We were interested in whether either recent or rather long-term Ne shows a stronger correlation with our power to detect the signatures of natural selection and/or with the efficacy of selection. Recent changes in population size can have dramatic effects on levels of genetic diversity, in turn affecting our power to detect localized reduction in diversity. If the differences we observe among populations are largely due to these power issues (higher power to detect local reduction of diversity in populations with a higher overall level of diversity) recent Ne should show the strongest correlation with the relative reduction in diversity around conserved sequences (protein-coding exons).

The results of a Pearson's correlation test show that the difference between populations in the percentage of windows containing protein-coding exons (E) in a given bin of the HKA empirical distribution significantly positively correlates with their estimated long-term Ne (Figure 2 and supplementary table S3)

for the 0.05% and 1-20% bins (0.1% bin marginally non-significant, p-value: 0.07). There is a significant negative correlation between E and estimated long-term Ne for the 55-99% bins. There is a significant positive correlation between E and an estimate of recent Ne for the 0.05-5% bins (Figure 2 and supplementary table S3). The estimate of recent Ne significantly correlates with E in fewer bins that the estimate of long-term Ne (4 compared to 13 bins) and the correlation with recent Ne is only stronger than the correlation with long-term Ne in the bottom 0.05-0.1% bins, suggesting that overall we observe the effects of the relatively long-term evolutionary history of each population, rather than merely differences in power due to the overall level of diversity. Nevertheless, putative differences in accuracy between the two Ne estimates may affect their comparison here.

The correlation notably changes from a positive correlation to a negative one when comparing the regions in the bottom and in the top half of the HKA empirical distribution. We observe a strong positive correlation for several bins in the bottom half of the HKA empirical distribution, which likely reflects the increased effect of negative selection lowering diversity in populations with historically large effective population sizes (possibly combined with our higher power to detect that reduction in samples with higher genetic diversity). This is particularly striking in the case of *P. paniscus*, which has among the lowest Ne of any of the lineages apart from *H. sapiens* and *P.t. verus* (according to the Ne estimate based on Watterson's estimator) and shows no trend for reduced diversity in regions overlapping protein-coding exons. Interestingly, *P. paniscus* showed little effect of purifying selection for loss-of-function variants, whose detection is not dependent on their levels of diversity (Prado-Martinez et al. 2013). On the other hand, populations with historically large effective populations sizes have a greater number of exons in regions with low diversity, likely reflecting the actions of purifying and background selection in removing deleterious and linked diversity in these regions respectively.

The strong negative correlation for most of the top half of the HKA empirical distribution also likely reflects less efficient purifying selection and weaker background selection in populations with low effective population sizes, combined with the fact that we may have higher power to detect unusual localized diversity peaks in populations with lower overall genetic diversity (due to higher Ne).

3. B-scores support the influence of positive and balancing selection in the tails of the HKA empirical distribution.

To further investigate the role of purifying and background selection in driving the different HKA scores observed between regions we tested whether regions with low HKA scores tend to be highly conserved. To do this we calculated the average B score per base of every genomic region for which we have an HKA score. The B score is a measure of the amount of background selection operating on a region, with a lower B score representing a higher level of background selection and being the same for a given position in all populations (McVicker et al. 2009). If purifying selection is shaping patterns of genetic diversity in the genome by removing deleterious variants, then we should observe that as the average HKA score of regions decreases (due to a reduction in diversity) the average B score of these regions also decreases. This general pattern is indeed what we observe (supplementary fig S6 and supplementary table S5) across the HKA empirical distribution if we exclude the tails (between the 5-99% range of the HKA empirical distribution), suggesting that the relative influence of purifying and background selection, which is greatly influenced by the effective population size, may play a role in shaping the relative levels of genetic diversity between the lineages studied.

We observe an upturn in the average B scores of regions in the 5% or lower tail of the HKA empirical distribution in all populations, showing that many of the regions with the lowest HKA scores do not have as low a B score as expected given their extreme reductions in diversity relative to the genome-wide distribution. This might be due to noise, as there are few of these regions relative to larger cutoffs in the HKA distribution. Alternatively, it could be due to strong positive selection acting on these regions, reducing their levels of diversity much more than predicted based on negative selection (B-score) alone. Finally, this pattern could be explained by the relatively large selective sweeps, as diversity is removed in unconserved regions that are close to the selected variants due to hitch-hiking. In any case, the data suggests that we cannot explain well the regions in tails taking into account only purifying selection, supporting the idea that positive selection has contributed significantly to the tails.

We were interested in investigating the effect of Ne in the correlation between a window's HKA score and its B-score. A Pearson's Correlation test comparing the correlation of the average B-scores per window for different cut-offs in the HKA empirical distribution with the two different estimates of Ne described above shows that the average B-score correlates significantly with Ne from the 1-10% tail for both estimates of Ne (supplementary fig S6 and supplementary table S5). The Ne values from Watterson's

estimator show a significant negative correlation with Ne from the 0.05-40% range of the HKA empirical distribution and a significant positive correlation from the 50-99% range. This suggests that effective populations size likely has a significant effect on the ability of the different populations to remove deleterious variants from conserved functional regions.

The 5% tail of the empirical distribution shows the strongest negative correlation with Ne (calculated using Watterson's estimator) with an R of -0.9 and a p-value of 0.0009. This supports the hypothesis that the Ne of a population has a very strong influence on its ability to remove variation in conserved regions through either purifying selection or selective sweeps due to positive selection, although differences in power among the populations may also play a role here.

4. Spatial clustering of windows supports the influence of balancing and positive selection in the top and bottom tails of the HKA empirical distribution.

Neutral evolutionary processes are expected to act uniformly across the genome while selection is considered to be locus-specific. If the windows we observe in the extreme tails of the HKA empirical distribution are the result of drift we would not expect to see increased clustering of these windows in terms of their spatial proximity in the genome relative to the clustering of regions from other parts of the HKA distribution. Therefore any evidence of an increase in spatial clustering of windows in the tails relative to what is observed across the HKA empirical distribution would be indicative of selection acting at specific loci in the genome greater than 30kb in size.

To investigate whether we find any such evidence of spatial clustering we calculated the average distance between all possible pairs of windows in the top and bottom 0.1 % of the HKA empirical distribution respectively for each population. This distance was only calculated between windows on the same chromosome. To determine whether this average distance showed evidence of increased clustered relative to the neutral expectation we compared it to the average distance between an equal number of regions randomly selected from the remaining 99.8% of the HKA empirical distribution. We observe an increase in spatial clustering in the 0.1% tails of the HKA empirical distribution relative to the central 99.8% for almost all populations (supplementary table S12). This suggests that the extreme tails of the HKA

distribution are enriched for regions that contain targets of selective processes. We compared the mean mapping quality and depth of coverage of sites in these regions to randomly sampled regions from the genome-wide distribution. We find no evidence that regions in the extreme tails of the HKA distribution are unusual in these regards (see Supplementary Materials HKA 1.3 and 1.4).

The only exception to this pattern is the 99.9% tail of the HKA empirical distribution in *P.t. schweinfurthii*, where the windows do not show an excess of spatial clustering relative to the central 99.8%. However closer inspection of the location of the windows in the 99.9% tail shows that there is a strong pattern of clustering at the chromosomal level, with 42 windows on chromosome 6. The chromosome with the 2nd largest number of windows is chromosome 9, which has only 16 windows in the 99% tail. The high number of windows on chromosome 6 in the 99.9% tail of the HKA empirical distribution is a feature shared across populations and is likely to primarily reflect long term balancing selection acting on the MHC region. As *P.t. schweinfurthii* has several chromosomes with only two or three windows in the 99.9% tail of the HKA empirical distribution it is possible that the large distance between these windows masks the signal of spatial clustering of windows under balancing selection in the MHC region and other potential cases of balancing selection.

The increased spatial clustering of windows in the 0.1% tail of the HKA empirical distribution relative to the general distribution is observed across all populations (supplementary table S12). This may be due to positive selection causing selective sweeps, which result in the loss of diversity in a wide region flanking the selected variant. This observation may explain the reduction in exonic regions in the extreme bottom tail of the HKA empirical distribution we report above (see Supplementary Materials HKA 2.4), if genic targets of selection are flanked by selective sweeps with similarly low levels of diversity and high levels of divergence. Alternatively this may be due to the strongest targets of selection being non-genic.

To investigate if these results were related to differences in Ne between the populations we ran a Pearson's Correlation test with two different measures of Ne described above (supplementary table S12). Although we observe no significant correlations with either estimate of Ne, the amount of spatial clustering in the bottom 0.1% of the HKA empirical distribution approaches significance with a p-value of 0.06 and an R score of -0.6 when using the long-term Ne values based on Watterson's estimator. This suggests that Ne may be influencing the strength and/or number of selective sweeps that occur in a population, with more and/or stronger selective sweeps occurring in populations with larger Ne. These results suggest that among

the *Hominidae* selective sweeps may be more prevalent in populations with larger long term Ne, although differences in power between species with different levels of diversity could also play a role.

5. Correlation of results with Tajima's D

To search for additional evidence that the candidate regions we identified as evolving under positive and balancing selection were true positives we calculated Tajima's D, another commonly used test for detecting positive and balancing selection, genome-wide for each species and performed permutation testing to explore Tajima's D signatures in the 1% extreme of the FWH (one-tail, see Supplementary Materials Section 3) and HKA (two-tails) empirical distributions. The results (supplementary table S107) show that windows in the tails of the FWH and HKA distribution are highly enriched for extreme values of Tajima's D (highly significant in 25 out of 27 comparisons). The Tajima's D test thus identifies similar outlier genes to the FWH and HKA tests and demonstrates that are results are broadly consistent with another test for selection.

6. Candidate selected regions and genes

For each population we produced two separate tables of the 200 protein coding genes with the strongest signal of positive or balancing selection as determined by the rank of the windows they overlap in the HKA empirical distribution. Genes in windows with the lowest HKA score in the empirical distribution are candidates for being under positive selection while genes in windows with the highest HKA scores are candidates for being under balancing selection. These tables can be found in supplementary table S6. For each gene we report the rank of the window it overlapped, its chromosomal location, the HKA score of the window, the p-value of the window based on its rank in the HKA empirical distribution, and further information about the gene including its Ensembl ID and gene name, the gene type (e.g. protein-coding, lincRNA etc) and how many base pairs of exons overlapped the window.

7. Functional annotation in candidate regions

We sought to identify whether the regions with signatures of positive selection in the HKA test are enriched for functional elements, such as protein coding genes. To identify any potential signals of functional enrichment among the regions with signatures of positive selection in the HKA test we calculated the percentage of windows in the 0.1% tail containing either functional annotation, protein-coding exons or non-protein-coding exons. Windows were annotated using GENCODE hg18 annotation (Harrow et al. 2012). Across all lineages values ranged from 62-90% for windows containing any type of functional annotation, 50-85% for protein coding exons (a subset of functional annotation) and 15-24% for non-protein coding exons.

To test whether these values were greater than expected for each lineage we performed random sampling of an equal number of windows as found in the 0.1% tail from across the genome-wide distribution of 30kb windows. From these regions we similarly calculated the percentage of regions containing functional annotation. This process was repeated 100 times for each lineage. The results of these random permutations were compared to the results in the 0.1% tail (supplementary figs S7-S15). We find that the proportion of candidate windows with signatures of positive selection that overlap functionally annotated elements or protein coding exons is significantly greater than the proportion that these annotation categories represent in the genome for the majority of lineages except *P. paniscus*, *P.t. verus* and *P. pygmaeus*. In contrast, no lineage showed an enrichment of non-protein coding genes in the 0.1% compared to sub sampling from across the genome These results suggest that the candidate regions do not simply reflect the tail of a neutral distribution and provides clues as to the targets of adaptive evolution.

8. Length bias of GO categories

As the HKA and FWH tests are performed using genomic windows there is a potential for variation in gene length to bias the results of category enrichment. Assuming a null hypothesis of no selection acting on the genome, then the genes in the extreme tail of the distribution may be overrepresented by long genes, because these cover more genomic windows and are therefore more likely to appear in the tail by chance. Any biological categories that are enriched for long genes may therefore show signals of significant enrichment due to this length bias. This potential effect of gene length on enrichment tests is rarely accounted for in window-based selection analyses.

To investigate if our results are influenced by this potential source of bias, we calculated, for each lineage, the mean length of the 200 genes in the bottom tail of the HKA empirical distribution. We compared this to the mean gene length of 200 genes in a random sample of genomic windows, 100 times. We observe no evidence of genes in the bottom tail of the HKA distribution being longer than the genomic mean based on the mean of the 100 means generated by this sub sampling procedure.

We repeated the process for the top tail of the HKA empirical distribution (candidate genes for balancing selection). For most lineages the mean gene length falls within the distribution of means obtained from random sampling. However for *P. paniscus* the mean length of genes in the top tail falls above the distribution of means obtained by random sampling. This trend may be partially driven by large intronic regions in these genes, where the absence of purifying selection permits diversity to accumulate.

We additionally explored the potential influence of length bias on our gene category enrichment test results by randomly sampling 200 windows from our genome-wide distribution using the process described above, 100 times. This process was repeated across lineages. For each pseudotest set we ran the GO enrichment analysis using FUNC (a software for biological enrichment analyses that can be run at the command line and is therefore capable of performing this number of tests) (Prüfer et al. 2007). The most frequently enriched GO category in these re-samplings is the 'biological process' category 'homophilic cell adhesion', which appears as significantly enriched in 6 out of the 100 re-samplings in *P. abelii* (supplementary tables S7-S11). Most other categories that appear as significantly enriched appear only once in a given population. In our actual results the GO category 'dendrite' appears as significantly enriched in three lineages (supplementary tables S16-S29. In the random sampling in two lineages this category is significantly enriched by chance in one of the 100 random sampling procedures and in one population it is significantly enriched in three of the 100 cases. The probability of this occurring by chance in all three lineages where we observe significant enrichment is thus extremely low. Therefore, these results suggest that gene length bias is unlikely to affect our enrichment of biological categories results (supplementary tables S7-S11).

References

Charlesworth B, Nordborg M, Charlesworth D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Ge*
510 *netical research* **70**: 155–174.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: The reference human genome annotation for the ENCODE project. *Genome Research* **22**: 1760–1774.

Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data.
515 *Genetics* **116**: 153–159.

Hughes AL, Yeager M. 1998. Natural selection at major histocompatibility complex loci of vertebrates. *An nual review of genetics* **32**: 415–435.

Kaplan NL, Darden T, Hudson RR. 1988. The coalescent process in models with selection. *Genetics* **120**: 819–829. http://www.genetics.org/content/120/3/819.short. H, Eyre-Walker A. 2014. Population size
520 and the rate of evolution. *Trends in ecology & evolution* **29**: 33–41.

Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favorable gene. *Genetical Research* **23**: 23–35.

McVicker G, Gordon D, Davis C, Green P. 2009. Widespread genomic signatures of natural selection in ho minid evolution. *PLoS Genetics* **5**.

525 Prüfer K, Muetzel B, Do H-H, Weiss G, Khaitovich P, Rahm E, Pääbo S, Lachmann M, Enard W. 2007. FUNC: a package for detecting significant associations between gene sets and ontological annota tions. *BMC bioinformatics* **8**: 41.

Zhai W, Nielsen R, Slatkin M. 2009. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. *Molecular Biology and Evolution* **26**: 273–283.

530

Section 3: FWH test

Hafid Laayouni, Marc Pybus, Ferran Casals, Jaume Bertranpetit

1. Fay Wu's H statistic calculation:

The Fay and Wu H statistic (FWH; Fay and Wu, 2000) measures departures from neutrality reflected in the difference between derived segregating sites at high-frequency and intermediate-frequency alleles. This statistic is especially robust to population demographic history, making it an improvement over Tajima's D in this regard (Tajima, 1989)). FWH takes advantage of ancestral information and is implemented in an overlapping window approach. The algorithm is run in windows of 30Kb with an offset of 15Kb. We validated the algorithm using simulations under neutral and selective scenarios with human demographic parameters (supplementary fig S19).

Windows with less than 300 callable sites were removed from the analysis to increase the quality of the data. Finally, an empirical p-value is calculated for each window for which we have a score thus generating a genome-wide ranking of scores.

2. Results

The Fay Wu's H algorithm was run on each lineage. In order to obtain a list of genes under positive selection for each lineage, windows were annotated following the same procedure as for the HKA pipeline (see Supplementary Materials HKA).

Using the empirical p-value of the FWH score, we first focus on the 1% extreme distribution of windows analyzed (approx. 1600 windows). Once annotated, for most lineages approximately 45% of these windows correspond to protein coding genes, 12% are lincRNA, 8% are pseudogene, 15% belong to other functional elements while 20% are not annotated as functional elements.

2.1 Genes with extreme footprints of selection

The supplementary table S74 lists the top candidate gene targets of positive selection for the lineages

analyzed. All genes are provided with a rank value for the FWH statistic. Ranks correspond to the empirical distribution of windows in the analyzed genome. Many genes involved in immune response appear among the top of genes putatively having evolved under positive selection (which we refer to as outliers with reference to the empirical distribution). Some of these genes also belong to the 1% extreme tail of many lineages analyzed. For example, *FER* (cytokine-mediated signaling pathway) is an outlier in *G.g. gorilla, P. abelii* and *P. pygmaeus* and in *P.t. ellioti*. *STAB2* (defense response to bacterium) is outlier in *P. paniscus, P.t. ellioti* and *P. abelii. HLA-DMA* (immunoglobulin mediated immune response) is outlier in *P.t. verus* and *P.t. schweinfurthii*. All the other genes involved in immune response appearing in the 200 top genes are outliers in two lineages or are specific to one lineage, a finding expected if there is stratification in the pathogenic environment among groups.

Many genes involved in neurobiological processes appear among the top genes putatively evolving under positive selection. *KCNIP4* (neuronal cell body and signal transduction) belong to this extreme 1% in *P. paniscus, P. t. troglodytes, P.t. schweinfurthii,* and *P.t. verus*. *ITGA8* (nervous system development; memory) belong to the 1% extreme tail of the distribution in *P. paniscus, P.t. verus, P.t. schweinfurthii* and *P.t. ellioti*. *FAM169B* (neuronal cell body and signal transduction) is an outlier in *P.t. troglodytes*. *GPR98* (neurological system process; sensory perception of sound) is an outlier in *P.t. schweinfurthii. NELL1* (nervous system development) is an outlier in *P. paniscus. FGF14* is also involved in the development of nervous system and is an outlier in *P. paniscus, G.g. gorilla* and *P.t. verus. NRG3* (member of the neuregulin gene family and implicated in susceptibility schizophrenia and schizoaffective disorder.) is an outlier in *P. abelii* and *P. pygmaeus*, *P. t. ellioti*, *P.t. schweinfurthii* and *P.t. troglodytes*. *NRXN3*, a gene encoding a member of a family of proteins that function in the nervous system as receptors and cell adhesion molecules shows up as extreme in three *Pan troglodytes* subspecies (*P.t. schweinfurthii, P.t. troglodytes, P.t. ellioti*) and in *P. pygmaeus.*

Interestingly, many genes involved in reproductive processes show up among the targets of positive selection; *LGR4* is involved in the development of male genitalia and is an outlier in the *P.t. troglodytes* lineages. *BARD1* is involved in spermatogenesis and is an outlier in *P.t. schweinfurthii* and *P.t. ellioti*. *HSD17B4* is involved in androgen and estrogen metabolic processes and is an outlier in *P. paniscus* and *P. abelii*. *RAD23B* is involved in spermatogenesis and is an outlier in *P. abelii*. *SPAM1* is involved in fusion of

sperm to egg plasma membrane and is an outlier in *G.g. gorilla*. *IQCJ-SCHIP1* is involved in female gonadal development and is an us outlier in *G.g. gorilla, P. paniscus, P. abelii,* and *P.t. schweinfurthii.* Worth noting is that *PCDH15*, a gene involved in adult walking behavior and visual perception, shows up as an outlier in *P.t. schweinfurthii* and *P.t. ellioti* and *P. paniscus.*

The 200 coding genes in the 1% tail of the FWH distribution, containing putative signals of positive selection, show considerable overlap across lineages (supplementary table S73). The average pair-wise intersection is 7.4% (ranging from 3% between *G.g. gorilla* and *P. abelii*, and between *P.t. troglodytes* and *P. pygmaeus* to 25% between *P.t. troglodytes* and *P.t. schweinfurthii*). This suggests that there are multiple cases of shared genic targets of positive selection between lineages.

References

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics, 155, 1405–1413.

Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D. Calibrating a coalescent simulation of human genome sequence variation. Genome Res. 2005 Nov;15(11):1576-83.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–95.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2):R14.

Section 4: MK test

Gabriel Santpere, Arcadi Navarro

1. Methods

1.1. Selection scan

We carry out a scan of natural selection in each of the studied lineages. To do so we focus on protein-coding genes and use the classical McDonald-Kreitman (MK) test (McDonald & Kreitman, 1991). This test compares the variation accumulated in a given species to the divergence between this and other species at two site classes (e-g. at synonymous versus non-synonymous sites). This first version of the test assumes that all non-synonymous changes are neutral, strongly advantageous or strongly deleterious. Considered this way, strongly deleterious mutation would have been eliminated by selection while strongly advantageous ones would have been fixed or contribute very little to polymorphism. Thus, existing diversity at non-synonymous sites is assumed to be mostly neutral. Another assumption of the test is that evolution occurred in a diploid panmictic scenario with stationary size. The MK test has been shown to be more powered to detect negative rather than positive selection using simulated *H. sapiens-P. troglodytes* genetic data (Zhai, Nielsen, & Slatkin, 2009).

1.2. Samples

For the MK test scan, we considered only species with at least five individuals (in the case of *P.t.verus*, we included the admixed Donald). We excluded individuals based on coverage to increase the proportion of the genome that could be evaluated (supplementary table S93 and Tables S1-S2).

1.3. Protein-coding sequences for selection scan

We extract coordinates for the coding part of all coding autosomal transcripts annotated in the RefSeq hg18, considering only unique identifiers (n=31,703, covering ~32Mb of genome).

We then intersect these CDS coordinates with the callable portion of the unique part of the genome that considers a position if all individuals used in the analyses show a minimum coverage of 5x. This leaves ~15.1Mb of coding sequence (collapsing overlapping transcripts).

For each species and coding sequence we count all present SNVs and substitutions (supplementary table S94). We require both SNVs and substitutions to have appeared after the split with the most recent common ancestral node with *H. sapiens* (in the case of *H. sapiens*, we take the ancestral node with *Pan*). We oriented variants using the reconstruction of variant ancestralities from (Prado-Martinez et al., 2013). For the *H. sapiens, Pan* and *Gorilla* genus we only consider variants that present a monomorphic ancestral node. For *Pongo*, we get only those variants that are different from a monomorphic ancestral node between *Gorilla* and *H. sapiens*, and from *Macaca mulatta*. Missing data was allowed in the case of counting a position with a SNP but no missing data was allowed when counting substitutions.

1.4. Variant annotation and MK test

We use the ANNOVAR software (http://www.openbioinformatics.org/annovar/) to annotate the synonymous (S)/non-synonymous (NS) effect of all SNPs and substitutions occurring in a transcript. Positions with more than two alleles were excluded.

We then count how many S and NS polymorphisms and substitutions we observe in each transcript. We then construct a contingency table and test for association with a one-tailed Fisher exact test. The test is performed only when a transcript shows three or more S changes (Ps+Ds) and three or more NS changes (Pn+Dn). For each transcript we obtain a p-value and an odds ratio (OR). For OR calculations, we added 0.5 to all cells if one of the cells was zero. In supplementary table S95 we present all transcripts evaluated for each species.

1.5. Estimates of rate of adaptive substitutions in the different primate species

We estimate the proportion of adaptive substitutions (α) and their rate (ωα) in all studied species, using an extension of the MK test (Eyre-Walker & Keightley, 2009) as implemented in the DFE-alpha software. This method compares the site frequency spectrum from neutral and selected sites and infers a distribution of fitness effects of new deleterious mutations by maximum likelihood, and models recent demographic changes, to emit an estimation of α and ωα. This more sophisticated method has limitations in the number of sites evaluated to gives proper estimates, and makes it unsuitable to study individual genes. To estimate the rate of protein adaptation in the different primate species we used a concatenated data-set of 3859 orthologous genes non-overlapping with any other described transcribed genomic elements in the human genome according to Refseq. 4fold synonymous sites were used as a neutral reference to estimate the proportion and rate of adaptive substitutions in 0fold non-synonymous sites using the DFE-alpha software. Unfolded site frequency spectra for each primate species were obtained using ancestral states in the closer node splitting each species with *H. sapiens*, as described for the classical MK test scan.

Correlation between Ne and various DFE-Alpha estimates was studied while attending to phylogenetic non-independence of the traits in the tree of the species analyzed using BayesTraitsV2, setting the methods to employ random-walk and maximum likelihood. Significance was assessed by comparing a model with a free correlation with another model with correlation fixed to 0, by means of a likelihood ratio statistic.

2. Results

2.1. Significant genes in MK scans

We tested for genes significant in the MK test at both tails indicating genes especially constrained (under strong purifying selection) and genes putatively under positive selection. In supplementary table S96 the number of genes significant in the MK test with $p < 0.05$ are shown. The MK test is more powered to detect instances of strong purifying selection (Zhai et al., 2009) as observed by the increased number of significant transcripts obtained. P-log columns indicate transcripts that are significant and at the same time possess an extreme ($p<0.05$) OR value. The complete list of significant transcripts can be found in supplementary table S97.

We compared our results with a previous MK genome-wide scan (Bustamante et al., 2005). We obtained from this study a list of genes described as positively selected (115) or negatively selected (215) between H. sapiens and P. troglodytes. Fifty positively selected genes in Bustamante et al. (2005) were also evaluated in our study and none of them gave a significant signal for positive selection with a p-value < 0.05. Of the 215 genes determined to be under strong purifying selection we had power to evaluate 147, and found four genes (P2RX7, TIAM1, COL12A1, CNGB3) that were also significant in our MK test. For the overlapping genes in both studies we performed a density plot of the logOR obtained for these genes in our analysis (supplementary fig S16). Clearly, genes with described strong purifying selection in Bustamante et al., 2005 are skewed to the left and separated from the overall distribution of logOR in our study. The distribution of logOR of genes with described positive selection appear bimodal with one peak skewed to the right and one other peak overlapping the average of the total gene-set. In general, the overlap between both studies is higher regarding purifying selection.

Several significant genes are related to nervous system function (supplementary table S98) and development. In particular *MCPH1*, *CASC5*, *PHGDH*, *FTO* and *NBN*, when mutated in *H. sapiens*, carry associated changes in brain size. Variants in other genes, such as *SETX*, *MTPAP*, *RNF213* or *VCAN* are related to neurodegenerative processes in *H. sapiens*. Interestingly, *SETX,* a gene involved in spinocerebellar ataxia and amyotrophic lateral sclerosis, has been also described to be a target of recent positive selection in CEU (Grossman et al., 2013). Another well-represented group of significant genes are related to the immune system (supplementary table S99). Strikingly, we find several genes that help in the defense against viruses, such as *ZC3HAV1*, *HIVEP1*, *MX1*. Finally, we obtained a list of other significant genes related to several other human disorders (supplementary table S100).

Lifespan and attributes of the senescence process are important divergent features among primates. Longevity in humans is notably enhanced compared to non-human primates even if they are in captivity; the oldest living human was 122 years old compared to the record of 74 years for chimpanzees (de Magalhães & Church, 2007). However, the genetic bases of these differences are yet to be revealed. Many comparative studies on ageing have been performed on Rhesus monkeys but much less is known about the ageing and cognitive decline in great apes, with chimpanzees being the most studied. We have identified ageing-related genes with signatures of positive selection particularly in *Pongo*, which present a maximum lifespan in captivity of around 59 years. For example we detected one gene related to ageing, *WRN*, with evidence of

positive selection in *P. pygmaeus* using the MK test. Mutations in *WRN* in humans can cause Werner syndrome which is a dramatic progeroid syndrome, i.e. causes premature ageing (Goto, 1997). Interestingly, previous studies reported possible adaptive acceleration in *WRN* since humans and chimpanzee diverged (Clark et al., 2003; de Magalhães & Church, 2007). We also found *NBN* under positive selection in *P. pygmaeus*, a gene also considered to be involved in progeroid syndromes in humans. Mutations in *NBN* cause a chromosomal instability syndrome called Nijmegen breakage syndrome (Martin & Oshima, 2000), that is accompanied by features of senescence. *P. abelii* shows also signals of positive selection in *ERCC5*. ERCC proteins play a role in DNA repair and have been linked to senescence. Mutations in *ERCC6* and *ERCC8* have been reported as causes of premature ageing in humans (de Magalhães & Church, 2007). Finally, *PPM1D* shows evidences of positive selection in *P. t. elliotti*, which has been shown to reduce longevity in *PPM1D*-null mice, specifically in males (Nannenga et al., 2006).

2.2. Functional enrichment analysis

We performed a functional enrichment analysis with WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/) to test for gene set enrichment in functional categories in *non-slimmed* Gene Ontology, KEGG pathways, Pathway Commons, Wikipathways, Disease (PharmGKB and GLAD4U) and PheWAS associated. In this case, according to a Bonferroni correction adjusted p-value < 0.05 and requiring at least two genes to be present in one category, we obtained several enriched categories (supplementary table S101). For instance, 'pancreatic function' appeared enriched in positively selected genes in *P.t. ellioti*, *P.t. schweinfurthii* and *P. pygmaeus*. *P. paniscus* selected genes showed an enrichment in 'potassium voltage-gated channel activity' and also in diseases of the nervous system such as 'Amyotrophic Lateral Sclerosis'. Gorilla selected genes are enriched in the 'glycoprotein metabolic process′ and 'glycosaminoglycan binding' categories, and also some of these same genes drove significance to the *cartilage disease* category.

Since the small number of significant transcripts obtained precludes a properly powered functional enrichment analyses we concatenated all genes belonging to each PANTHER pathway (Thomas et al. 2003)

and performed the MK test. In supplementary table S102 the p-values obtained for each of these pathways are shown. We observed that many pathways were under strong negative selection in many great ape species at the same time (p-value <= 0.05). In particular, 'Alzheimer's disease-presenilin pathway', 'Integrin signaling pathway' and 'Wnt signaling pathway' appeared significantly constrained in all lineages. The Integrin pathway includes the components involved in the downstream events triggered by the interaction of integrins with elements of the extracellular matrix, such as actin related genes and MAPKs. Actin and MAPK related KEGG pathways have also been reported to be enriched in genes under purifying selection between human and chimpanzees and between mouse and rats in previous studies (Serra et al. 2011). The Wnt pathway has been related to many important biological processes and may have a universal role in configuring the primary axis of animals (Nusse & Varmus, 2012). The Alzheimer's disease (AD) pathway includes genes involved in this human neurodegenerative disease. In agreement with our results, crucial AD genes such as *APP* and *MAPT* were reported to be conserved between human and chimpanzees (Hamilton, 2004; Holzer, Craxton, Jakes, Arendt, & Goedert, 2004; Rosen et al., 2008). This is an interesting finding because although the main neuropathological hallmarks of AD (i.e. Aβ and hyper-phosphorylated tau deposition) have also been observed in the ageing brain of chimpanzees (Gearing, Rebeck, Hyman, Tigges, & Mirra, 1994; Rosen et al., 2008), the complete AD clinical and neuropathology seems to be presented by humans only.

A few pathways showed a nominally significant p-value for positive selection: i.e. 'Plasminogen activating cascade' in *P.t. verus* and *P.t. schweinfurthii*, 'Axon guidance mediated by semaphorins' also in *P.t. schweinfurthii*, 'Glutamine glutamate conversion' in *P. paniscus*, 'Serine glycine biosynthesis' in *G.g. gorilla*, and finally the 'Thyrotropin-releasing hormone receptor signaling pathway', 'Formyltetrahydroformate biosynthesis' and the 'Alpha adrenergic receptor signaling pathway' in *P. abelii*. But only one pathways, 'Plasminogen activating cascade', appeared in more than one lineage: *P.t. verus* and *P.t. schweinfurthii.*

2.3. Correlation between rate of adaptive substitutions (ωα) and effective population size (Ne).

Proportion of adaptive substitutions (α) and rate of adaptive substitutions (ω(α)) were estimated using DFE-alpha by combining its estimated distribution of fitness effects (out of the species SNP data) and each species derived substitutions, considering modeled recent demography/sweeps effects. The

supplementary table S103 shows the estimates of α and jointly with estimates of synonymous and non-synonymous polymorphism (θs and θn) and the evolutionary rate (omega, ω) in non-synonymous 0fold sites. The values for Ne calculated from Watterson's θ, from (Prado-Martinez et al., 2013) are also indicated. We obtained in general low values of both α and ω(α), in agreement with previous estimates (Eyre-Walker, 2006; McManus et al., 2015). *P.t. schweinfurthii* presented the highest values probably as the result of including two of the six individuals (Vincent and Andromeda) with high levels of inbreeding.

According to the nearly neutral theory, selection efficiency depends on the effective population size, because the fate of a mutation is determined by the product NeS (Ohta, 1976; Tomoko Ohta, 2002)(Otha 1976,2002; Lynch and Connery 2003). We check this theory correlating the estimates of rate of adaptive substitutions with the estimates of effective population size for each primate species. Correlations were calculated while controlling for the phylogenetic non-independence using the generalized least square approach (Table S105) implemented in BayesTraitsV2 and using the random walk/maximum likelihood method. We obtained a measure of significance by comparing the likelihood of the model with free correlation value with a model with a correlation fixed to 0. The correlation is positive for alpha and omega(alpha) but non-significant. The observed rate of adaptive evolution in P. pygmaeus was poorly supported by bootstrap and fell close the 95% quantile; removing them together with P.t. schweinfurthii, that included inbreed individuals from a different geographical region and also excluding humans, that presented very different Ne between African and Non-African individuals as in (Prado-Martinez et al., 2013), we obtained a significant correlation between ω(α) and Ne (supplementary fig S16(C)). This all suggests that primate species with higher effective population size have a higher rate of adaptive evolution in proteins.

Estimated DFE for new mutations at 0-fold sites (supplementary table S104 and supplementary fig S18) shows that most mutations are deleterious at this class of sites with little variation among species (sites with predicted NeS>10 account for more than 65% in all primates). Additionally, we also consistently found a marginally significant negative correlation between proportions of neutral or nearly neutral sites with a long term Ne, indicating that the deleterious effect of mutations is greater in populations with larger Ne, illustrating a more efficient action of selection in the latter.

References

Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M. T., Glanowski, S., … Clark, A. G. (2005). Natural selection on protein-coding genes in the human genome. *Nature*, *437*(7062), 1153–7. doi:10.1038/nature04240

Clark, A. G., Glanowski, S., Nielsen, R., Thomas, P. D., Kejariwal, A., Todd, M. a, … Cargill, M. (2003). Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science (New York, N.Y.)*, *302*(5652), 1960–3. doi:10.1126/science.1088821

De Magalhães, J. P., & Church, G. M. (2007). Analyses of human-chimpanzee orthologous gene pairs to explore evolutionary hypotheses of ageing. *Mechanisms of Ageing and Development*, *128*, 355–364. doi:10.1016/j.mad.2007.03.004

Eyre-Walker, A. (2006). The genomic rate of adaptive evolution. *Trends in Ecology & Evolution*, *21*(10), 569–75. doi:10.1016/j.tree.2006.06.015

Gearing, M., Rebeck, G. W., Hyman, B. T., Tigges, J., & Mirra, S. S. (1994). Neuropathology and apolipoprotein E profile of aged chimpanzees: implications for Alzheimer disease. *Proc Natl Acad Sci U S A*, *91*(20), 9382–9386.

Goto, M. (1997). Hierarchical deterioration of body systems in Werner's syndrome: implications for normal ageing. *Mechanisms of Ageing and Development*, *98*(3), 239–54. Retrieved from

Hamilton, B. A. (2004). alpha-Synuclein A53T substitution associated with Parkinson disease also marks the divergence of Old World and New World primates. *Genomics*, *83*(4), 739–742.

Holzer, M., Craxton, M., Jakes, R., Arendt, T., & Goedert, M. (2004). Tau gene (MAPT) sequence variation among primates. *Gene*, *341*, 313–322.

Martin, G. M., & Oshima, J. (2000). Lessons from human progeroid syndromes. *Nature*, *408*(6809), 263–6. doi:10.1038/35041705

McManus KF, Kelley JL, Song S, Veeramah KR, Woerner AE, Stevison LS, Ryder OA, Ape Genome Project G, Kidd JM, Wall JD, et al. 2015. Inference of gorilla demographic and selective history from whole-genome sequence data. *Molecular biology and evolution* **32**: 600–12. http://mbe.oxfordjournals.org/content/32/3/600.short.

Nannenga, B., Lu, X., Dumble, M., Van Maanen, M., Nguyen, T.-A., Sutton, R., … Donehower, L. A. (2006). Augmented cancer resistance and DNA damage response phenotypes in PPM1D null mice. *Molecular Carcinogenesis*, *45*(8), 594–604. doi:10.1002/mc.20195

Nusse, R., & Varmus, H. (2012). Three decades of Wnts: a personal perspective on how a scientific field developed. *The EMBO Journal*, *31*(12), 2670–84. doi:10.1038/emboj.2012.146

Prado-Martinez, J., Sudmant, P. H., Kidd, J. M., Li, H., Kelley, J. L., Lorente-Galdos, B., … Marques-Bonet, T. (2013). Great ape genetic diversity and population history. *Nature*, *499*(7459), 471–475.

Rosen, R. F., Farberg, A. S., Gearing, M., Dooyema, J., Long, P. M., Anderson, D. C., … Walker, L. C. (2008). Tauopathy with paired helical filaments in an aged chimpanzee. *J Comp Neurol*, *509*(3), 259–270.

Serra, F., Arbiza, L., Dopazo, J., & Dopazo, H. (2011). Natural selection on functional modules, a genome-wide analysis. *PLoS Comput Biol*, *7*(3), e1001093.

Sidow, A. (1992). Diversification of the Wnt gene family on the ancestral lineage of vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(11), 5098–102.

Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=49236&tool=pmcentrez&rendertype=abstract

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome research* **13**: 2129–41.

Section 5: ELS test

Alexander Cagan, Christoph Theunert, Kay Prüfer, Aida M. Andrés

890

1. Extended Lineage Sorting (ELS) Test and Motivation

We aim to identify lineage-specific targets of natural selection by identifying regions of the genome
895 where variation falls external to an outgroup species and which are longer than expected under neutrality. For a detailed description of the Extended Lineage Sorting test see Supplementary Information 7 in Prüfer et al. (2012) and also SOM 13 in Green et al. (2010) where it was first used. Briefly, strong positive selection on a new mutation is expected to cause a selective sweep at that locus. This pattern is produced by the haplotype carrying the selected allele rising to fixation, removing neutral diversity in the region. The effect is
900 equivalent to a bottleneck of one individual around this locus, as all lineages in the population will coalesce at the time of the sweep. Over time variation in this region will recover due to new mutations and recombination. The new variation will be unique to this lineage. Thus, when comparing diversity in this region in relation to an outgroup, the variation in the outgroup will always fall outside the variation in the ingroup (external). When two sister species with a divergence time that is recent enough for the majority of
905 coalescent events to occur in the common ancestor (internal) are compared, the presence of large external regions can thus be used as a signal for detecting positive selection. The size of external regions is a product of the selection coefficient and the recombination rate. The stronger the selection and the lower the recombination rate, the longer we expect the external regions to be. By controlling for variation in recombination rate it should be possible to identify regions based on the strength of selection they have
910 experienced since their divergence with the species used as an outgroup. It is worth noting that purifying selection is also expected to result in shallow trees and external regions. Nevertheless, under purifying selection the external signal is not expected to extend to the same extent as under positive selection, and thus we expect the longest external regions in the genome to be highly enriched for targets of positive selection.

This method allows for the detection of regions that experienced positive selection at a time depth
915 not usually reached by other tests based on the departure of present-day intra-specific diversity from neutral

expectations, such as those based on haplotype length (where the signal of selection decays once the sweep has reached fixation) or the site frequency spectrum (where the signal fades as the locus reaches equilibrium). Therefore this method has the potential to detect lineage-specific positive selection with a greater time depth than many other selection tests which rely on selection being ongoing or very recent. In some species, this includes selection around the time of speciation, which could have played an important role in creating species-specific adaptive traits.

1.1. Pre-test Filtering

We applied a series of pre-test filtering steps to make the data as comparable as possible within and between populations. Variability in the number of individuals with genotype calls across sites results in unequal power to detect selection (for general pre-test filtering see Methods). For each comparison the dataset is further filtered to only include sites where at least one of the lineages is polymorphic, as these are the most informative sites for the test.

1.2. Performing the ELS Test

The ELS test requires two lineages as input. The test is most powerful when the divergence time of these lineages is recent enough that the majority of loci have a coalescence time in the common ancestor of the two lineages (internal) but where there has been sufficient time for diversity to recover after lineage-specific selection (external).

For each comparison, neutral coalescent simulations were performed using the program ms to infer the expected fraction of the genome for which the tested population falls inside the variation of the outgroup (internal) compared to outside this variation (external) (Hudson 2002). Simulations were performed using a uniform recombination rate. The simulations require demographic parameters of the lineages being compared, such as estimates of current effective population size for the test and outgroup lineages, mutation rate and generation time (parameters taken from Prado-Martinez et al., 2013; supplementary table S66). The results of the simulations were compared to the observed fraction of derived sites in the data. The simulations were generally found to closely match the observed fraction of derived sites (supplementary fig

S20). The simulations were run for the following comparisons: P. troglodytes-P. paniscus (outgroup), P. paniscus-P. troglodytes (outgroup), G.g. gorilla-G.b. graueri (outgroup), G.b. graueri-G.g. gorilla (outgroup), P. abelii-P. pygmaeus (outgroup), P. pygmaeus-P. abelii (outgroup). We also looked at the P. troglodytes sub-species P.t. ellioti-P.t. schweinfurthii (outgroup), P.t. schweinfurthii-P.t. ellioti (outgroup), P.t. troglodytes-P.t. ellioti (outgroup), P.t. verus-P.t. ellioti (outgroup). For comparisons with G.g. gorilla we used G.b. graueri although due to the low number of individuals (three) they were excluded from other analyses.

The simulations provided information regarding which species comparisons are adequate for this test, and which ones would be underpowered due to too high a percentage of their genome falling external to the outgroup under neutrality (supplementary table S67). For example, both P. paniscus and P. pygmaeus had over 75% of their genome falling external to their outgroup under neutral simulations, and thus the test was not run for these species. We also excluded G. b. graueri because of our inability to find adequate demographic parameters (the frequencies of derived to ancestral alleles in simulations did not closely match the data, suggesting that the test would produce inaccurate results, supplementary fig S11).

1.3. Hidden Markov Model

A hidden Markov model (HMM) is used to assign all SNP positions with a hidden state of *internal* or *external*. For a full account of the calculation of the emission and transition probabilities see Prüfer et al. (2012). The HMM uses all available individuals from the test lineage and one individual from the outgroup lineage. Due to the presence of multiple individuals in the outgroup populations, the HMM was run repeatedly on the test lineage, using a different individual from the outgroup lineage each time.

After running the HMM, the output is filtered to remove SNPs with a posterior probability < 0.8 for being internal or external. Neighboring SNPs that have been classified as external are merged to form external regions along the genome. Larger regions should be indicative of stronger positive selection. However, region size may also be influenced by recombination rate and low SNP density. To correct for this, regions are scored as a function of their size, the local recombination rate, and physical distance between adjacent SNPs. Specifically, each region is re-scored by calculating the 1Mb average human recombination

rate (rr) for the midpoint between any two adjacent SNPs in a region. The recombination map is taken from Kong et al (2002). Each pair of adjacent SNPs is assigned a value of 1000/rr if their physical distance exceeds this value. Otherwise the physical distance is given in base pairs. These values are then summed over all pairs of SNPs and multiplied by the average recombination rate of the entire region. This provides a SNP-corrected value of genetic distance per region.

For each lineage the results of the multiple runs of the HMM (each run obtained using a different outgroup individual) are combined to eliminate the effect of using a single outgroup individual, and thus reduce the number of false positives. This is done as follows: For each run of the HMM, the results are ranked by score as described above. The top 5% of external regions from each run are selected and are given a rank based on their score. Regions are only retained if they appear in the top 5% in all runs (thus, regardless of the outgroup individual used). To refine the detection of the signal, the regions are then trimmed to only the part of the region present in all runs. This list of regions is then ranked based on the cumulative total of the rank score from the multiple runs.

2. Results

The number of unusually long external regions in the 5% extreme of the score distribution varies for each species tested (supplementary tables S68-S70). We detect 27 external regions candidates to contain targets of selection in *P. troglogytes,* 11 regions in *G.g.* g*orilla*, and 26 regions in *P. abelii.* These regions contain 15 genes in *P. troglodytes,* 7 genes in *Gorilla* and 27 genes in *P. abelii.* The average external region size is 30,723bp for *P. troglogytes,* 29,875bp for *Gorilla* and 65,451bp for *P. abelii.* Many of these regions do not contain genes, suggesting that selection has been acting on both coding and non-coding variation.

References

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, et al. 2010. A draft sequence of the Neandertal genome. *Science (New York, NY)* **328**: 710–722.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics (Oxford, England)* **18**: 337–338.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, et al. 2002. A high-resolution recombination map of the human genome. *Nature genetics* **31**: 241–247.

Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, Koren S, Sutton G, Kodira C, Winer R, et al. 2012. The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**: 527–31.

Section 6: Subsampling analysis

Alexander Cagan, Christoph Theunert, Aida M. Andrés, Gabriel Santpere, Arcadi Navarro

Investigating the influence of sample size variation on results

The dataset analysed in this manuscript consists of population level whole-genome sequence data from multiple lineages. The number of individual genomes available varies between lineages, from three in the case of *G.b. graueri* to 12 for *P. paniscus and G.g gorilla*). This raises concerns that differences in sample size may influence our ability to detect signatures of selection in each lineage. This could lead to results that are due to differences in power rather than being biologically meaningful.

1. Methods

To assess the influence that variation in sample size has on our results we took a subsampling approach. For each of our selected tests we re-ran the entire analysis using only subsets of individuals. The overlap in the results obtained with these subsets could then be compared to the original results (using all individuals) to obtain a measure of how much variation in sample size is likely to be influencing our results.

For the HKA, FWH and MK tests we chose the *G.g. gorilla* and *P. paniscus* lineages for subsampling. We chose these lineages they have the largest sample sizes, providing the greatest opportunity to see the impact of sample size variation when subsampling. Coincedentally these lineages are also very different in their Ne (Figure 1), thus by subsampling in both lineages we may simulatenously explore whether differences in Ne may influence the effect of subsampling. For the ELS test we only use *G.g gorilla* for subsampling as *P. paniscus* is not included in the original ELS analysis.

1.1. HKA subsampling results

For the HKA we randomly selected four and eight individuals from G.g. gorilla and P. paniscus. This was repeated 100 times for each lineage and sample size respectively. For each of these subsamplings we re-ran the HKA analysis for chromosome one only and compared the results with the original results based on using all individuals. We analysed regions in the 0.001% and 0.01% tails for the positive and balancing selection tails respectively. For each region that occurred in a given tail on chromosome one in the original results we calculated the percentage of times it also appears in the tail in the subsamplings. So a region that remains in the tail in all 100 subsamplings has a percentage of 100%. The results are presented in supplementary figure S21.

The results suggest that for the HKA analyses variation in sample size does not have a major impact on the candidate positively selected regions. We observe that with the larger 0.01% tail there are a greater number of outlier regions that occur less frequently in the tail subsampling. However for both tails in *G.g Gorilla* and for the *P. paniscus* 0.01 tail for subsampling from either four or eight individuals the mean overlap of regions across subsamplings is almost 100%. The lowest mean overlap we observe is in the *P. paniscus* 0.01 tail when subsampling with eight individuals, where the mean overlap is 90%. It is unexpected that the overlap is lower when subsampling with eight individuals rather than four, however in both cases we consider the overall amount of overlap across subsamplings to be high. Therefore for the HKA positive selection tails we conclude that variation in sample size is not causing a strong bias in detecting putatively selected regions.

The results indicate that the balancing selection candidates for the HKA are much more influenced by sample size variation (supplementary figure S21). This may be due to the general rarity of balancing selection in the genome relative to positive selection. We expect only a small number of true targets of balancing selection on chromosome one, if any. Therefore the 'candidate' regions in the balancing selection tail on chromosome one may just be neutrally evolving regions, which may explain the low overlap between regions depending on sample size. Given that we detect the MHC region, a known target of balancing selection across a wide-range of vertebrates, in all our lineages despite their variation in sample size, we do not think that sample size variation is preventing us from detecting the strongest targets of balancing selection in the genome.

1.2 FWH subsampling results

We repeated the same subsampling analysis for FWH, though only for positive selection as this test does not detect signatures of balancing selection (supplementary figure S21). We observe that FWH is more sensitive to sample size variation than the HKA test for detecting candidate regions under positive selection. As expected, we observe a greater overlap with the original results when using the subsampling of eight individuals compared to when we use only four (supplementary figure S22(E-J)). We also observe higher overlap with the original results with the more stringent 0.001% tail compared to the 0.01% tail, suggesting that the strongest candidates of positive selection are relatively robust to subsampling. We also observe subsampling results in a higher mean percentage of overlaps with the original results with *P. paniscus* compared to *G.g. gorilla*□, although the ranges overlap.

The differing sensitivity to sample size variation between these two tests may be due to the different types of information that they use to detect signatures of selection. The HKA test is dependent on the ratio of substitutions to polymorphisms, without considering the allele frequency of polymorphisms. This should make it particularly robust to sample size variation. In contrast, methods that are more reliant on fluctuations in the allele frequency spectrum to detect selection, such as FWH, are more sensitive to variation in sample

size. Therefore, our results provide support for using the HKA test to detect selection in cases where sample size is low or unequal between populations.

1.3 ELS subsampling results

We performed a similar subsampling analysis for the ELS test, randomly subsamplinging four or eight randomly selected individuals from *G.g. gorilla* 100 times in both cases. We find that overlap of candidate selected regions between the original results and the subsamplings is generally low, with a mean overlap of ~20% for each putatively selected region across subsamplings. This may be due to the reliance of ELS on using simulations to infer parameters, which makes the test vulnerable to errors if the demographic model in the simulations is inappropriate.

1.4 HKA and Ne correlations subsampling results

We observe significant correlations between the long term Ne of lineages and the percentage of protein coding exons in several bins of the HKA empirical distribution (Supplementary Materials 2.2). To ensure that these correlations are not due to variance in sample size between lineages we re-ran these correlations after excluding lineages with either high or low sample-sizes. We re-ran the correlations using only populations with < 10 individuals (excluding *G.g gorilla, P. paniscus* and *P.t. ellioti*) and only populations with > 5 individuals (excluding *P. abelii, P. pygmaeus, P.t. troglodytes* and *Pt. Verus*). In both cases there were at least five lineages.

The results for both sub-sampled correlations are very similar to the original results with all lineages (supplementary table S106, supplementary fig S22). The subsampling excluding lineages with larger sample sizes has the most similar correlation scores to our original results (supplementary table S106). However the R values are very similar to the original results in both cases (supplementary figure S22). This suggests that the correlation between the long-term Ne of a lineage and the percentage of protein-coding exons in bins of the HKA empirical distribution, which we infer as a measure of the strength of background selection, is robust to variation in sample size between lineages.

Section 7: Targets of selection

Alexander Cagan, Christoph Theunert, Jaume Bertranpetit, Aida M. Andrés

1. Introduction

Here we provide further discussion of genes and pathways with signatures of selection identified by our analyses which were not presented in the main Discussion section.

1.1. Diet

All lineages of the *Hominidae* are omnivorous, with all lineages apart from *H. sapiens* having a preference for frugivory (Boyd & Silk, 1997). However, there is still considerable variation in diet between lineages (Uchida, 1996). Adaptations that maximize the extraction of energy from food are expected to be highly beneficial. As a result we might expect genes related to digestive processes to have been targeted by positive selection.

Human populations tend to consume high levels of starch rich foods compared to other members of the *Hominidae* (Hohmann et al. 2012), a trend exacerbated by the transition to agriculture and the wide-spread availability of starchy foods (Zohary et al. 2012). Previous work identified copy number variation in *AMY1* in humans as advantageous to increase the benefits from starchy foods (Perry et al. 2007). We provide additional evidence that genes related to starch metabolism have been targets of positive selection in humans. In *H. sapiens* the strongest category enrichment among the HKA candidate targets of positive selection is for the KEGG pathway 'starch and sucrose metabolism' (p-value=0.02), a signal not shared among any of the other ape lineages. This signal is driven by the genes *GAA*, *AMY2B* and *GUSB*. Incidentally, copy number increase of *AMY2B* in dogs is considered an adaptation to starch-rich diets that arose during cohabitation with humans (Axelsson et al. 2013). The evidence of positive selection on this gene in humans suggests that adaptation to a starch-rich diet may have a partially shared genetic basis between dogs and humans.

1.2 Anatomy

There is considerable anatomical variation between the different lineages of the *Hominidae*. One of the most clearly visible differences between lineages is in body size, with gorillas being the largest extant primate. The ELS test detects signatures of positive selection in *G.g. gorilla* in a region containing the gene *IGF2R*, which encodes insulin growth factor receptor 2. Mutations in this gene are associated with variation in body size traits in cattle (Berkowicz et al. 2012). Biological category enrichment analysis also shows significant enrichment on the KEGG pathway 'vascular smooth muscle contraction' (p-value=0.03). Changes in the vascular system, which regulates blood pressure, may have been necessary to cope with the changes in body size that must have occurred during evolution of the gorilla lineage. Therefore it is possible that selection on these genes are related to body size differences between the *Gorilla* subspecies, with *G.g. gorilla* considered to have less sexual dimorphism in body size and a smaller male body size than *G.g. beringei* (Taylor 1997).

The *Hominidae* lineages display a variety of anatomical adaptations to their various forms of quadrupedal and bipedal locomotion. *Pongo* are the only lineage in the *Hominidae* that regularly brachiate (a form of arboreal locomotion based on swinging through trees using only the arms), which likely involved some molecular adaptations. Interestingly, in *P. abelii* HKA candidate targets of positive selection are significantly enriched for genes in the GO molecular function category 'structural constituent of muscle' (p-value=0.01). One of the genes driving this signal is *NEB*, which encodes nebulin, a protein that helps to maintain the structural integrity of myofibrils in skeletal muscle. Deficits in nebulin result in a dramatic decrease in the force production capacity of skeletal muscle (Bang et al. 2006). This gene also shows a signature of positive selection from the HKA test in *P. abelii* and also *P.t. verus*, as well as a recent signature of positive selection in *P. abelii* from the FWH test.

Another feature differentiating the *Hominidae* lineages is the morphology and distribution of their body hair (Yesudian 2011). However, the genetic basis of such differences is not well understood. Genes related to 'hair follicle' development were identified as accelerated in the gorilla lineage in a recent study that used dN/dS ratios to compare protein coding sequences in humans, chimpanzees and gorillas (Scally et al. 2012). One of the genes they identified as contributing to this signal is *DSG4*. We find that this gene appears particularly constrained in all species of the genus *Pan* (MK results) but it shows signatures of positive selection in *G.g. gorilla*. This gene encodes the protein desmoglein 4, which plays an important role in the maintenance of hair follicle keratinocytes (Bazzi et al. 2006). Mutations in *DSG4* can cause hypotrichosis in humans, which is an abnormal condition of hair that affects its amount and results in atrophied hair follicles and shafts (Shimomura et al. 2006).

References

Axelsson E, Ratnakumar A, Arendt M-L, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**: 360–4.

Bazzi H, Getz A, Mahoney MG, Ishida-Yamamoto A, Langbein L, Wahl JK, Christiano AM. 2006. Desmoglein 4 is expressed in highly differentiated keratinocytes and trichocytes in human epidermis and hair follicle. *Differentiation* **74**: 129–140.

Berkowicz EW, Magee DA, Berry DP, Sikora KM, Howard DJ, Mullen MP, Evans RD, Spillane C, MacHugh DE. 2012. Single nucleotide polymorphisms in the imprinted bovine insulin-like growth factor 2 receptor gene (IGF2R) are associated with body size traits in Irish Holstein-Friesian cattle. *Animal genetics* **43**: 81–7.

Boyd R, Silk J. 1997. *How Humans Evolved*. W.W. Norton & Company, New York.

Hohmann G, Robbins MM, Boesch C. 2012. *Feeding Ecology in Apes and Other Primates*. Cambridge University Press.

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nature genetics* **39**: 1256–1260.

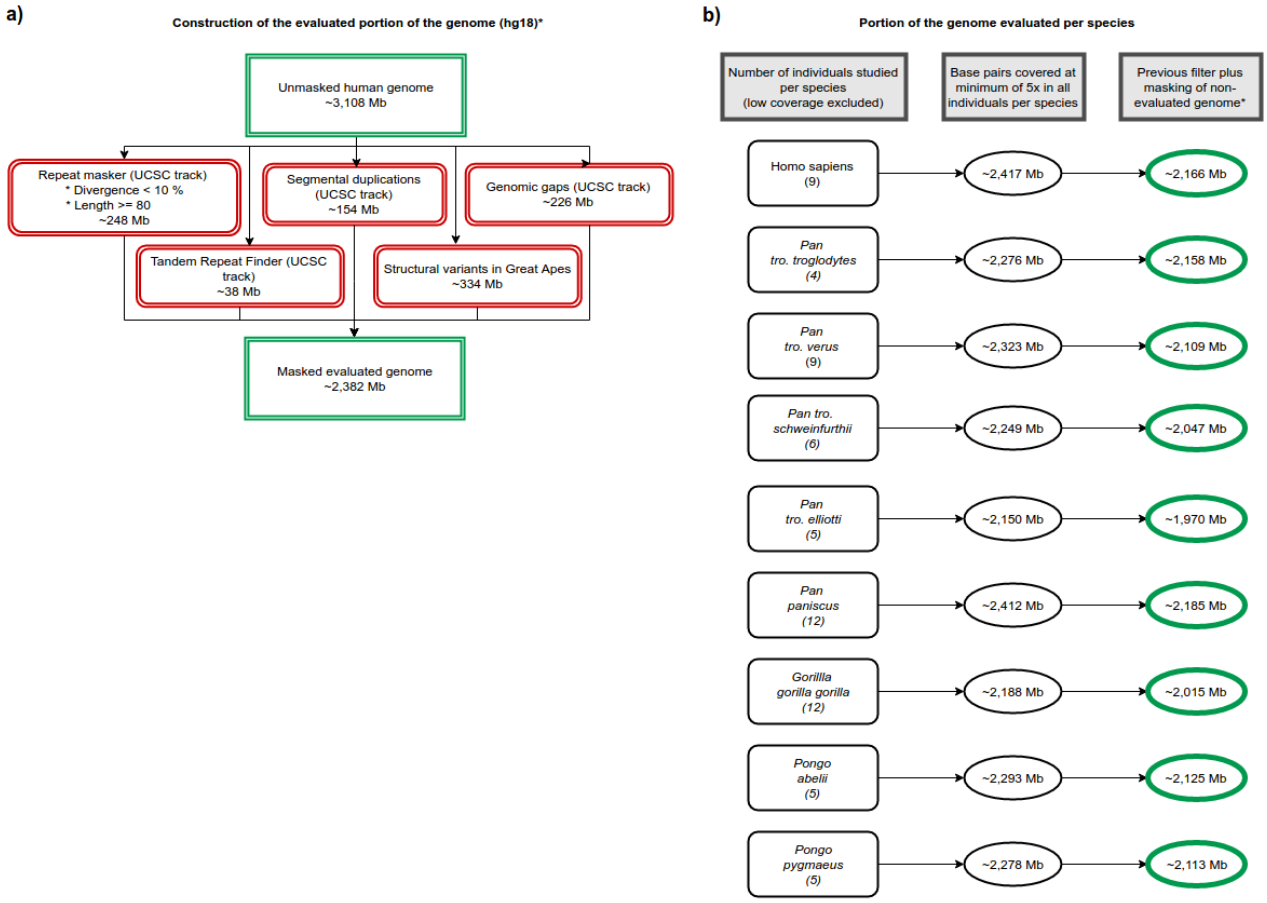
Scally A, Dutheil J, Hillier L. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**: 169–175.

Shimomura Y, Sakamoto F, Kariya N, Matsunaga K, Ito M. 2006. Mutations in the desmoglein 4 gene are associated with monilethrix-like congenital hypotrichosis. *The Journal of investigative dermatology* **126**: 1281–5.

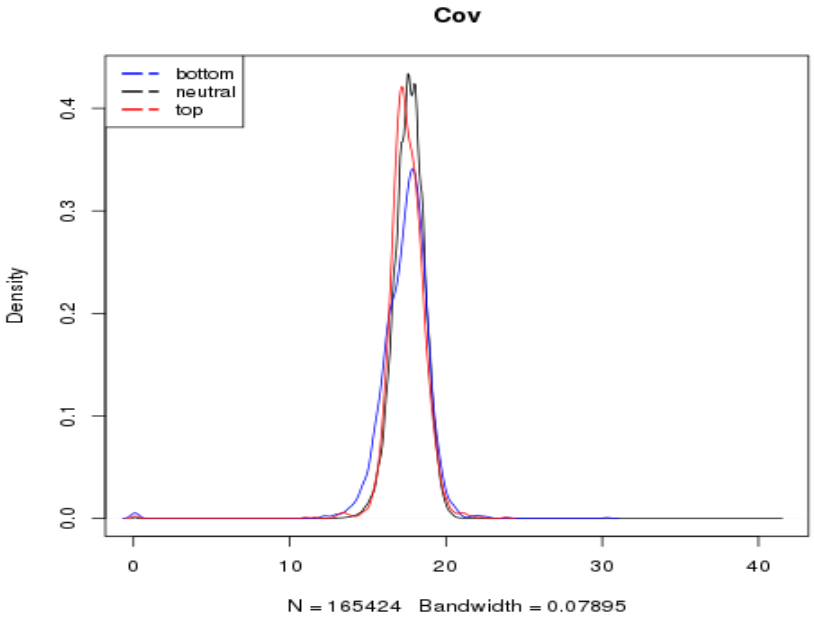
Taylor AB. 1997. Relative growth, ontogeny, and sexual dimorphism in gorilla (Gorilla gorilla gorilla and G. g. beringei): evolutionary and ecological considerations. *American journal of primatology* **43**: 1–31.

Uchida A. 1996. What we don't know about great ape variation. *Trends in ecology and evolution* **11**: 163–168.

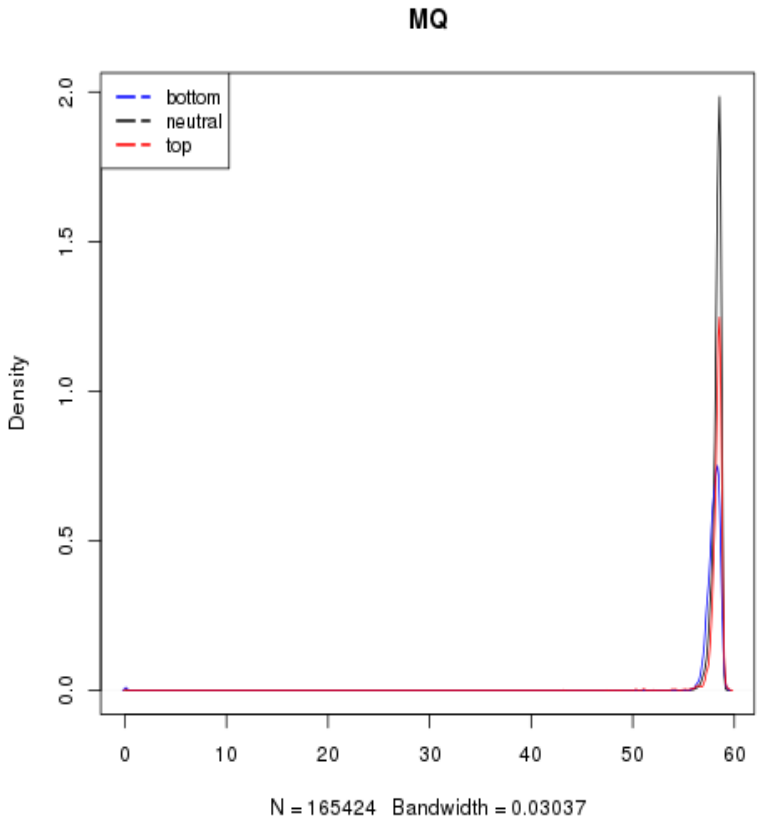
Zohary D, Hopf M, Weiss E. 2012. Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin. OUP Oxford.



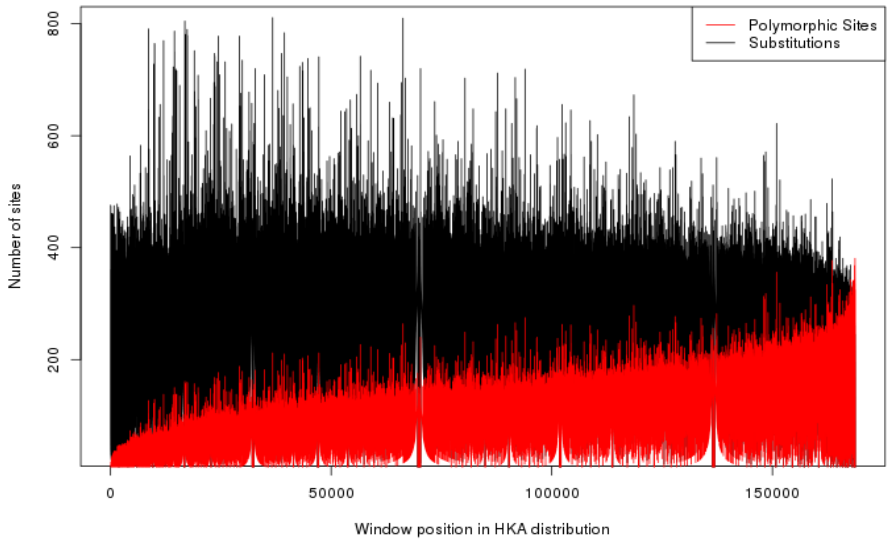
Supplementary Figure S1. Flowchart of filtering steps. A. Filtering steps that were applied across the entire dataset. Each red box represents a particular filtering step and the amount of the genome (Mb) that was excluded. Green boxes show the amount of the genome available for analysis before and after these filtering steps. **B.** Portion of the genome evaluated in each species, based on filtering of sites with < 5x coverage across all individuals per species



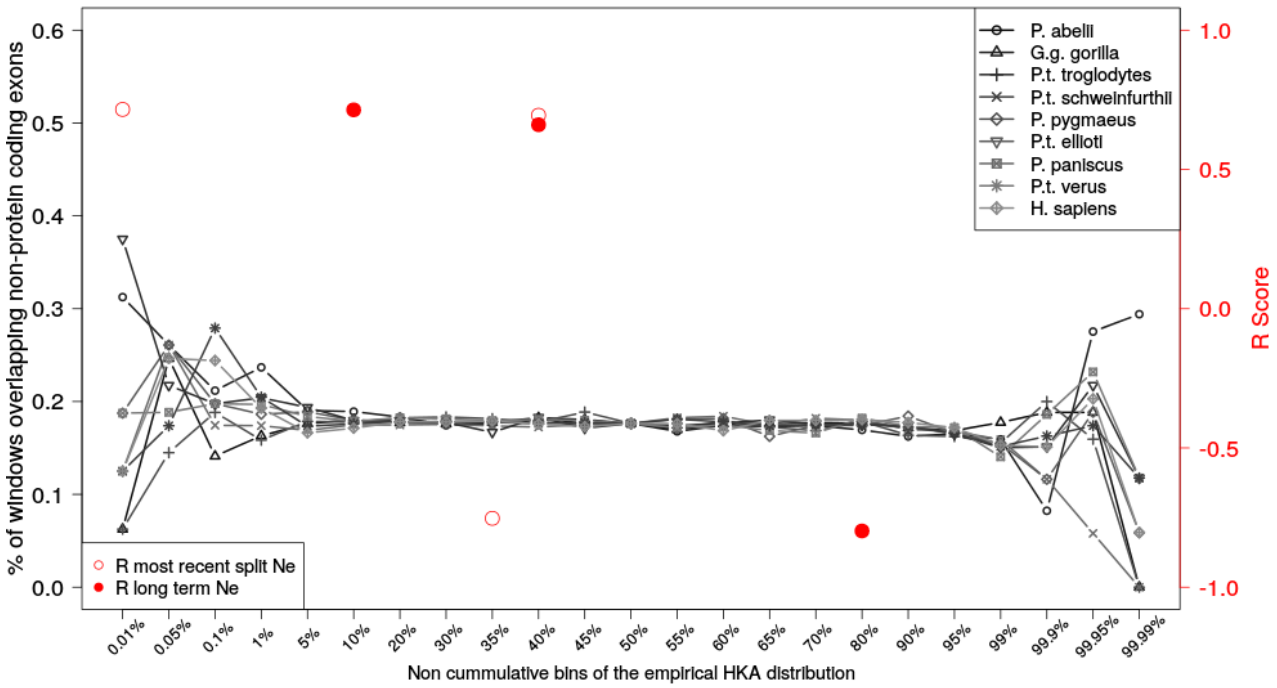
Supplementary Figure S2. Plot showing relationship between coverage and HKA score for G.g. *gorilla*. The X-axis shows average coverage. The Y-axis shows the frequency. Red lines represent results from the top 1% of the HKA score distribution, blue lines from the bottom 1% and black lines from the middle 98%. The vertical bars represent the mean coverage after 1000 permutations from each distribution.



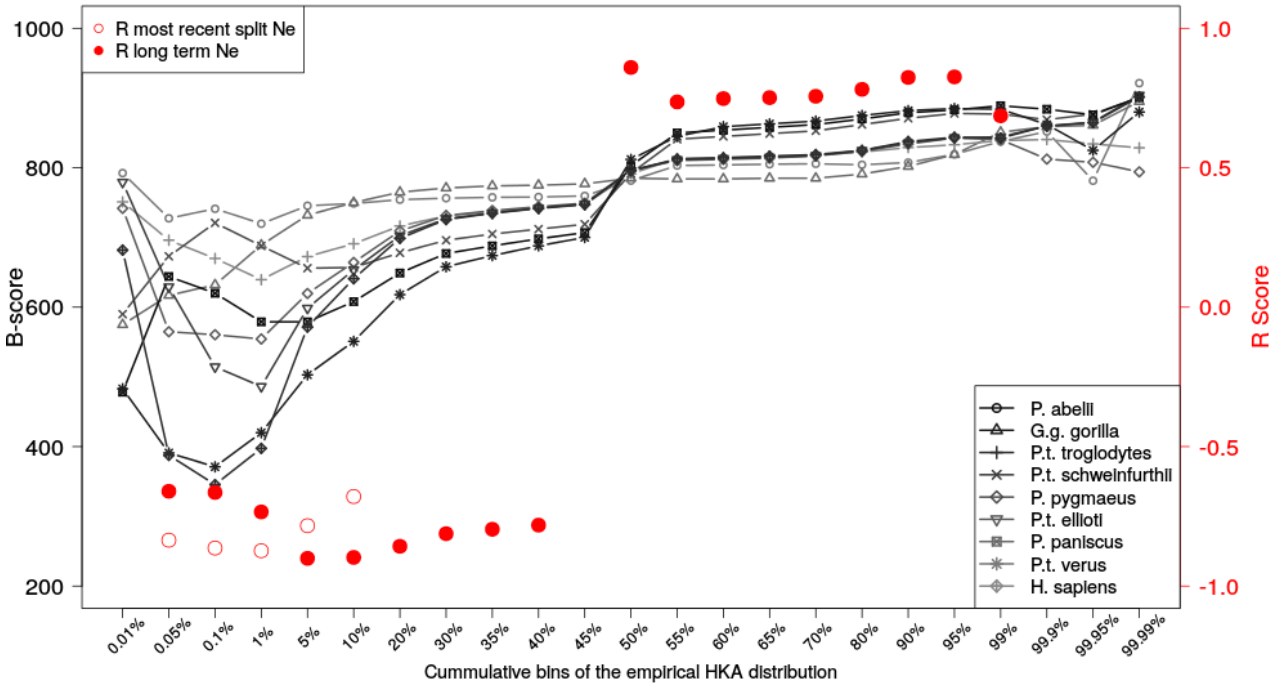
Supplementary Figure S3. Plot showing relationship between the Mapping Quality Score and HKA score for G.g. *gorilla*. The X-axis represents the mean Mapping Quality Score in each 30kb window. The Y-axis represents the frequency. Red lines represent results from the top 1% of the HKA score distribution, blue lines from the bottom 1% and black lines from the middle 98%. The vertical bars represent the mean coverage after 1000 permutations from each distribution.



Supplementary Figure S4. Number of of substitutions and polymorphic sites per window across the HKA empirical distribution for *P. troglodytes.* The Y-axis shows the number of sites in a window. The X-axis shows the position of windows in the HKA empirical distribution. The HKA score of windows increases along the X-axis.



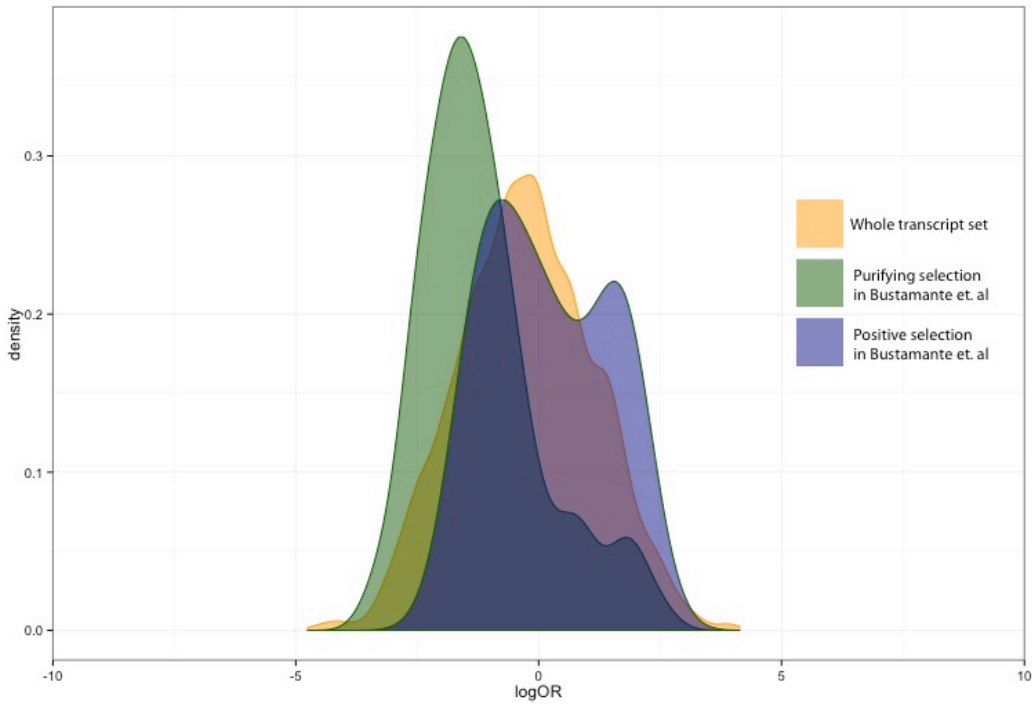
Supplementary Figure S5. Percentage of windows overlapping non-protein coding exons. Percentage of windows overlapping non-protein coding exons (>=1bp overlap exon with genomic window) for non-cumulative bins of the HKA empirical distribution (X-Axis). We plot this for each lineage as a shaded line. Furthermore, for each bin we measure the correlation between the % of windows overlapping non-protein coding exons and Ne among all lineages using a Pearson's correlation analysis. We do this separately with an estimate of short-term and long-term Ne, derived from PSMC and Watterson's estimator respectively (taken from Prado-Martinez *et al.* 2013). The right-side Y-axis shows the R score. R score is plotted with dashed lines. Only R values with significant P values (P <0.05) are shown.



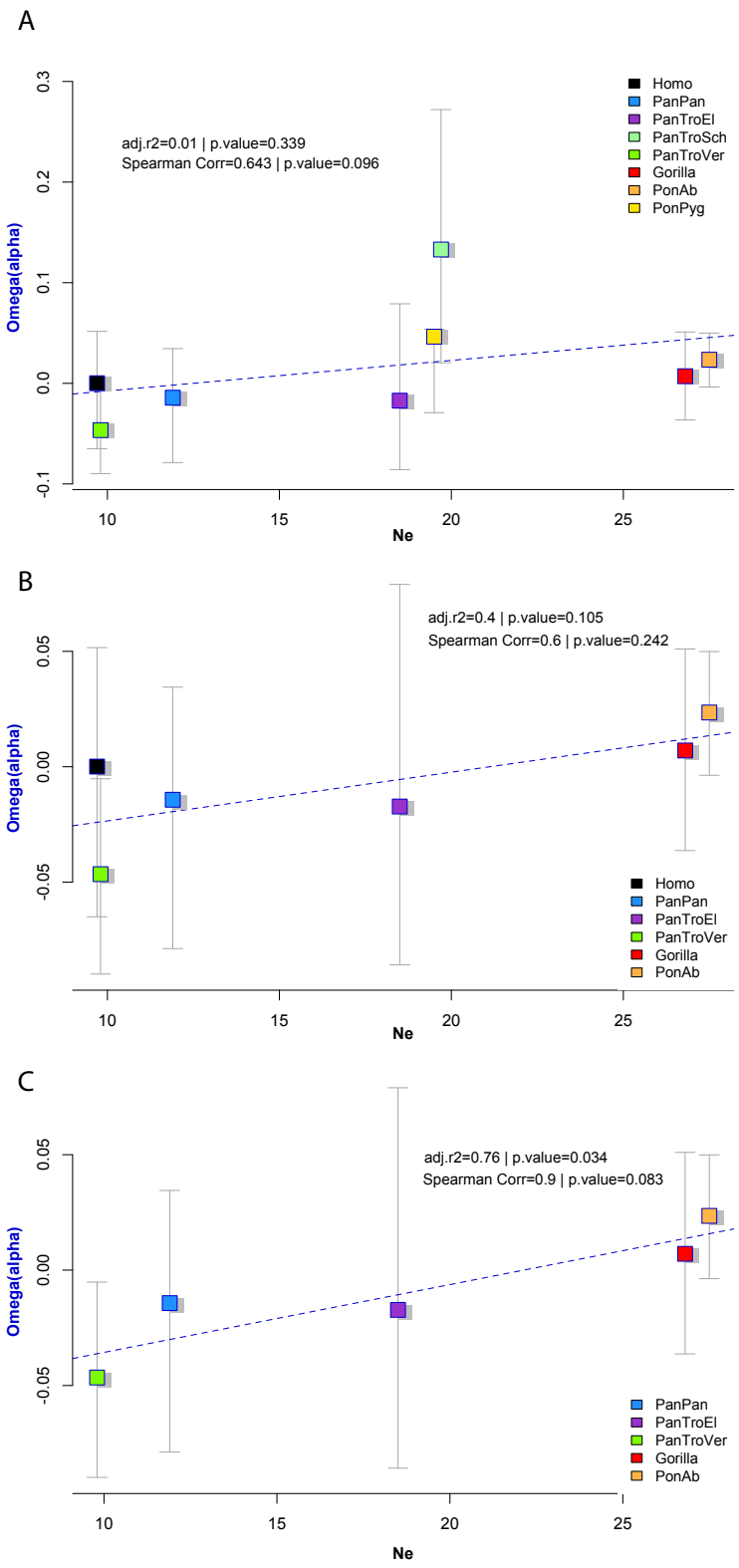
Supplementary Figure S6. Average B-score for windows with different cut-offs of the HKA empirical distribution. Average B-scores for each cumulative bin of the HKA empirical distribution (X-axis). B-scores were calculated according to McVicker *et al.* (2009). We plot this for each lineage as a shaded line. Furthermore, for each bin we measure the correlation between the average B-score and Ne among all lineages using a Pearson's correlation analysis. We do this separately with an estimate of short-term and long-term Ne, derived from PSMC and Watterson's estimator respectively (taken from Prado-Martinez *et al.* 2013). The right-side Y-axis shows the R score. R score is plotted with dashed lines. Stars indicate R scores with significant P values (P <0.05).

See file 'Cagan_S7_S15.pdf' for Supplementary Figures S7-S15

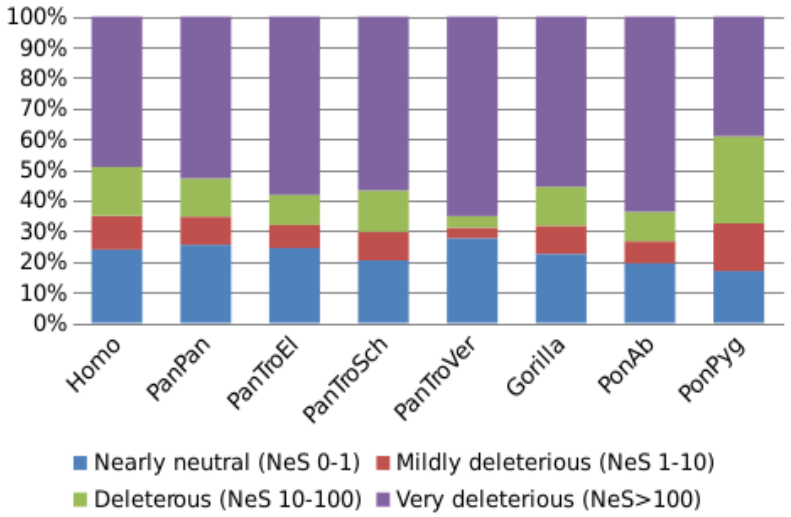
Supplementary Figures S7-S15. Annotation in HKA 0.1% tail compared to neutral sub-sampling. For each lineage we separately calculated the percentage of windows in the 0.1% tail of the HKA distribution containing any type of functional annotation, protein coding exons, or non-protein coding exons (see title of figure for category presented). These results were compared to 100 random sub-samplings of an equal number of windows from the genome-wide distribution. The value from the 0.1% tail is given as a horizontal line while the results from the neutral sub-samplings are given as a box-plot. P-values indicating whether the percentage of annotation in the 0.1% tail is significantly enriched relative to the genome-wide sub-sampling are presented below the X-axis.



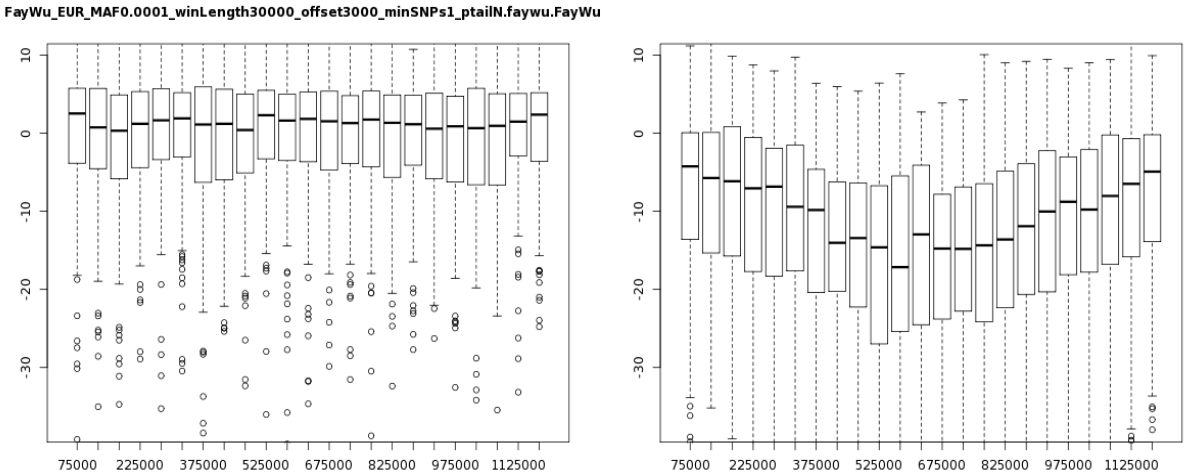
Supplementary Figure S16. DFE of deleterious mutations in 0-fold sites. Density plot of MK-test logOR for genes showing different signals of selection in Bustamante et al., 2005.



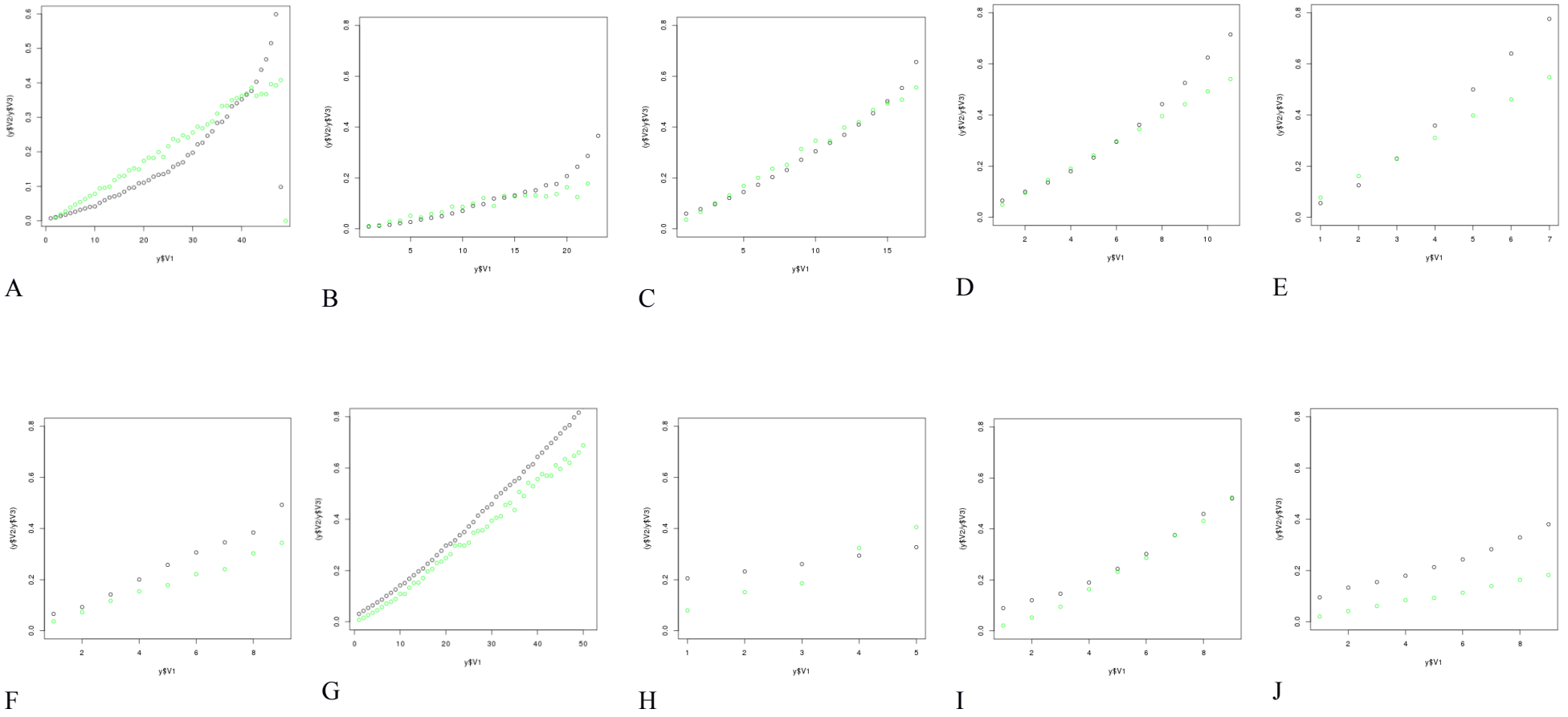
Supplementary Figure S17. Correlation between rate of adaptive substitutions (ω_a) and effective population size (Ne). using all species (A), excluding *P.t.schweinfurthii* and *P. abelii* (B), also excluding *H. sapiens* (C).



Supplementary Figure S18. Distribution of fitness effect of mutations at 0-fold sites.



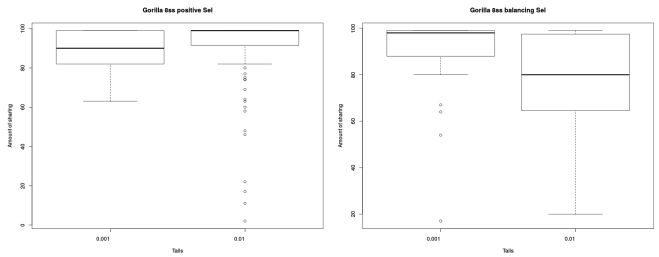
Supplementary Figure S19. **Validation of Fay and Wu's H algorithm with simulations.** Validation of FWH algorithm in a neutral simulation (left) and with a selective event in the central region (right). COSI simulator (Schaffner et 2005) was run under the best fit model (validated human demography) to generate 1000 realistic simulations under neutrality and under selection. Simulations with selection consisted in a selective sweep starting 500 generations ago with a selection coefficient of 0.022 located in the middle position of the 1.2Mb simulation. Both scenarios were then analyzed using Fay Wu's H algorithm. Plots were generated averaging the obtained scores in windows of 25Kb.



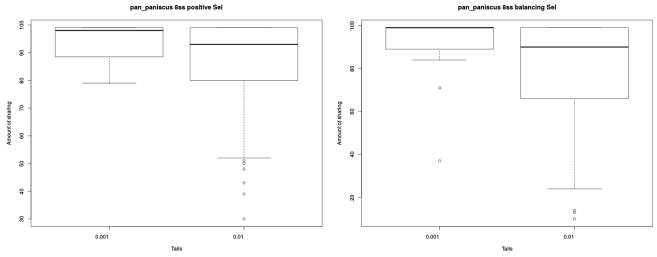
Supplementary Figure S20. Pairwise comparison between 2 analyzed lineages, **s***pecies 1 – species 2. Each graph shows the f*raction of sites where *species 1* is

derived in dependence of number of individuals in *species 2.* Simulated results in green, observed results in black.

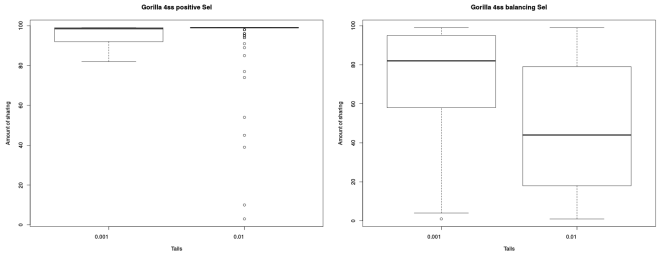
A. *P. troglodytes – P. paniscus.* **B.** *P. paniscus – P. troglodytes.* **C.** *P.t. ellioti – P. paniscus.* **D.** *P.t. schweinfurthii – P. paniscus.* **F.** *P.t. troglodytes – P. paniscus.* **F.** *P.t. verus – P. paniscus.* **G.** *G.g. gorilla – G.b. graueri.* **H.** *G.b. graueri – G.g. gorilla.* **I.** *P. abelii – P. pygmaeus.* **J.** *P. pygmaeus – P. abelii.*



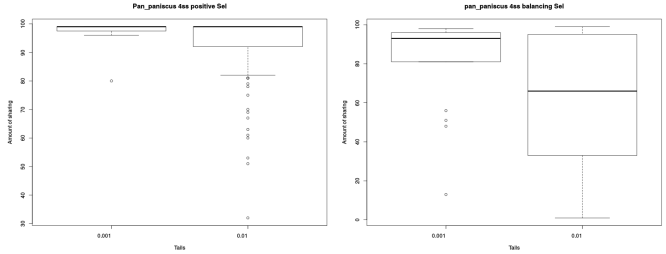
A



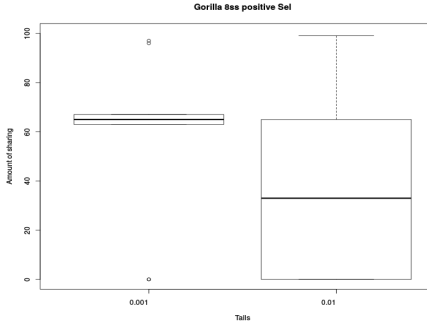
B



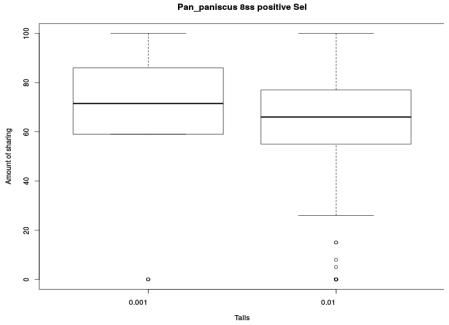
C



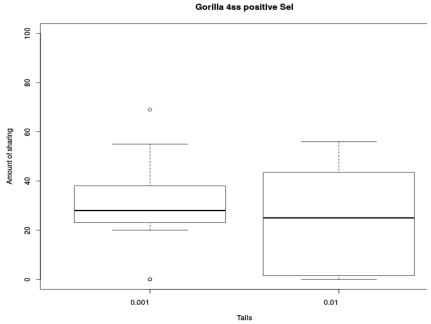
D


Gorilla 8ss positive Sel
Amount of sharing
0
20
40
60
80
100
0.001
0.01
Talls

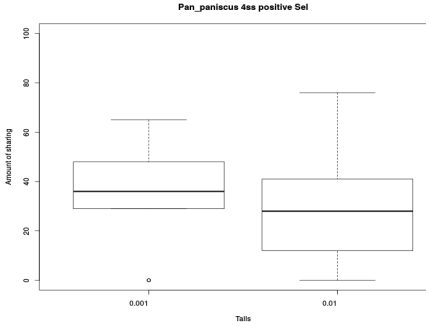
E


Pan_paniscus 8ss positive Sel
Amount of sharing
0
20
40
60
80
100
0.001
0.01
Talls

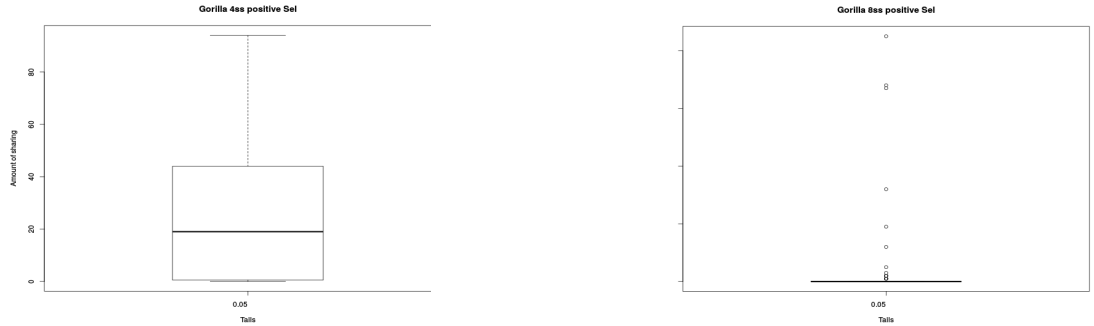
F


Gorilla 4ss positive Sel
Amount of sharing
0
20
40
60
80
100
0.001
0.01
Talls

G


Pan_paniscus 4ss positive Sel
Amount of sharing
0
20
40
60
80
100
0.001
0.01
Talls

H



I J

Supplementary Figure S21. Subsampling results for HKA, FWH and ELS with 4 and 8 individuals for *G.g. gorilla* and *P. paniscus*.

Box plots summarizing the amount of sharing of candidate regions from the original test results with the 100 subsampling analyses. For HKA test, the left-side plot presents results for the positive selection tail. The right-side plot presents results for the balancing selection tail. The X-axis shows the fraction of regions from the empirical distribution that are considered in the tail (far-left tail for positive selection, far-right for balancing selection). The Y-axis shows the percentage of overlap between regions from the original results compared to the 100 subsamplings. For FWH, Box plots summarizing the amount of sharing of candidate regions from the original test results on positive selection with the 100 subsampling analyses. For ELS, Box plots summarizing the amount of sharing of the top 5% of putatively selected regions from the original test results with the 100 subsampling analyses.

A. Subsampling results for HKA with 8 individuals for *G.g. gorilla*.

B. Subsampling results for HKA with 8 individuals for *P. paniscus*.

C. Subsampling results for HKA with 4 individuals for *G.g. gorilla*.

D. Subsampling results for HKA with 4 individuals for *P. paniscus*.

E. Subsampling results for FWH with 8 individuals for *G.g. gorilla*.

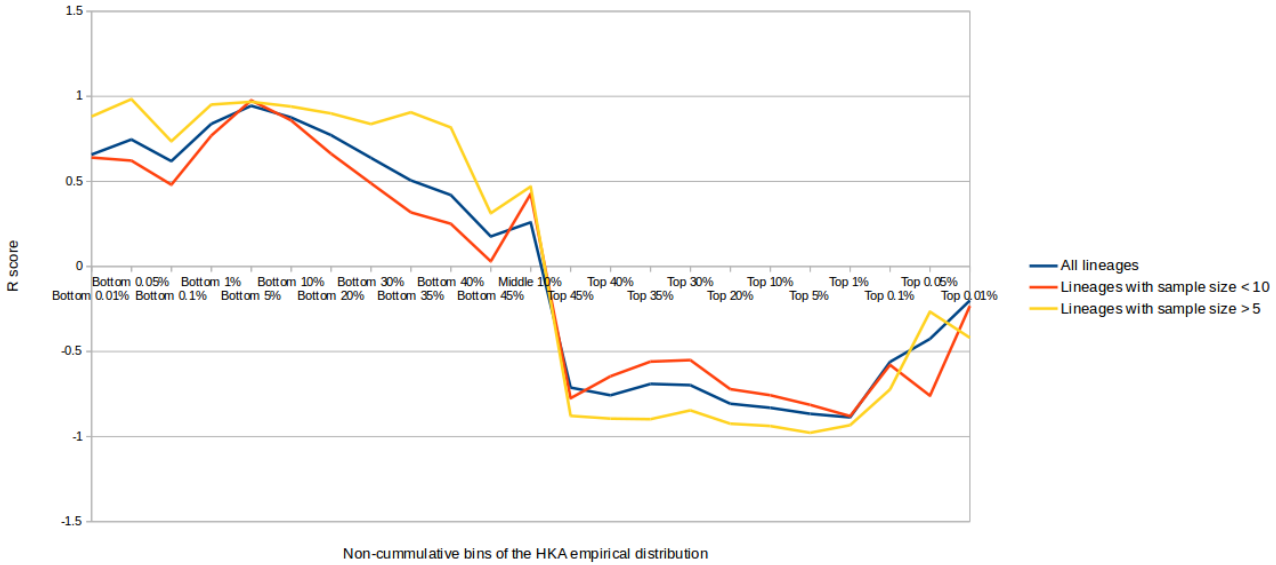
F. Subsampling results for FWH with 8 individuals for *P. paniscus*.

G. Subsampling results for FWH with 4 individuals for *G.g. gorilla*.

H. Subsampling results for FWH with 4 individuals for *P. paniscus*.

I. Subsampling results for ELS with 4 individuals.

J. Subsampling results for ELS with 8 individuals.



Supplementary Figure S22. Subsampling results for HKA Ne correlations. Correlations between Watterson's Ne estimates for each lineage and the percentage of regions overlapping protein-coding exons in non-cumulative bins of the HKA emipirical distribution. Results from using all lineages are plotted alongside results calculated using only lineages with a sample size < 10 or > 5. The X-axis shows the non-cumulative bins of the HKA empirical distribution. The Y-axis shows the Pearson correlation (R) between the E statistic and Ne within each HKA bin and across all lineages.