

SUPPLEMENTARY ANALYSES

Analysis of population structure

We used the program *structure* to further test for evidence of subdivision within each species. For each species, we randomly selected 200 SNPs from variable four-fold degenerate synonymous positions and tested a strict model of no admixture and a model allowing individuals to descend from multiple populations. For each model we considered between one to three populations (K), performed three replicates per K, and used a burn-in period of 100,000 iterations of the Markov Chain Monte Carlo (MCMC) followed by 1,000,000 MCMC iterations per replicate. All replicates were checked for convergence and repeated as necessary. For gorillas, we used an extended burnin length of 500,000 iterations to improve convergence of parameter estimates from the MCMC. The posterior probability of each K was calculated using Bayes' Rule assuming a uniform prior for K. The results of these analyses are summarized in supplemental table 5.

For Yorubans, we found that a model assuming two populations was most likely when allowing for admixture between groups. However, the difference in likelihoods between K=1 and K=2 or 3 was small. Moreover, several metrics commonly used to identify the optimal K value indicated a lack of true population structure. First, the proportion of admixture from each putative cluster was equally partitioned within and across individuals. Second, α was consistently large and > 1 , implying all individuals in the sample were extensively admixed. Finally, a model assuming no admixture unambiguously rejected subdivision. The K=2 admixture models also provided a marginally significantly better fit for chimpanzees and gorillas, with the same pattern of shared ancestry across all individuals. Collectively, these observations suggest that mild departures from the underlying model (e.g., linkage among sites, Hardy-Weinberg disequilibrium, genotyping error, etc) are leading to a spurious overestimation of K in these three species.

We did find compelling evidence for substructure (K=2) within both the gorilla and chimpanzee samples when considering a model of no admixture. For chimpanzees, this was due to four individuals that we previously determined have a significant portion of ancestry from adjacent and closely related eastern chimpanzee populations (*Pan troglodytes schweinfurthii*; Fischer et al. 2011). All four individuals carry mtDNA genotypes that group within eastern chimpanzees and three of these individuals were

confiscated in the Democratic Republic of Congo, where both eastern and central chimpanzees can occur (Fischer et al. 2011). In gorillas, a single individual from the Cross River area fell out as distinct. A second individual potentially from this region unambiguously grouped with other western gorillas.

We found a strong and consistent support for three distinct populations within bonobos (supplemental table S5; supplemental fig. S3). To verify this result, we reanalyzed the data incorporating 126 additional SNPs (326 total) across a broader range of possible populations ($K = 1-5$; supplemental table 6). In all cases, we found a consistent signature of substructure for both models with and without admixture. Interestingly, admixture between the clusters appears fairly low with significant allele sharing largely restricted to four individuals of mixed ancestry (supplemental fig. 3). Comparison between our results and a phylogeny of complete mtDNA genomes from the same individuals (Fischer et al. 2011) revealed that at least one of the partitions mostly overlaps with the deepest split within the mtDNA genealogy.

Lineage-specific evolution of SFPs

The neutrality index (NI_{TG}) shows high variance when levels of variation and/or divergence are low. To further explore the possible relationship between mating system and evolution of the male ejaculate we calculated the direction of selection (DoS) for reproductive genes for chimpanzees and bonobos as a group (*Pan*) and gorillas and humans as a group (*G-H*). Similar to the NI , this statistic incorporates both polymorphism and divergence and takes on positive values for genes showing an excess of nonsynonymous substitutions and negative values for excess polymorphism but may be more informative when variation is low (Stoletzki and Eyre-Walker 2011). Interestingly, estimates of DoS were only weakly correlated between *Pan* and *H-G* (supplemental fig. 5; Spearman's rank correlation $\rho = 0.15$, $P = 0.03$). A significant excess of genes showed negative values for both species (FET $P < 0.05$). One possible signature of shifts in the form and intensity of selection would be genes for which the sign of DoS changes between the two groups. 43 such genes showed a shift towards positive selection in *Pan*, compared to 43 genes in *H-G*. Thus, there was no apparent underlying directionality in DoS with respect to mating system.