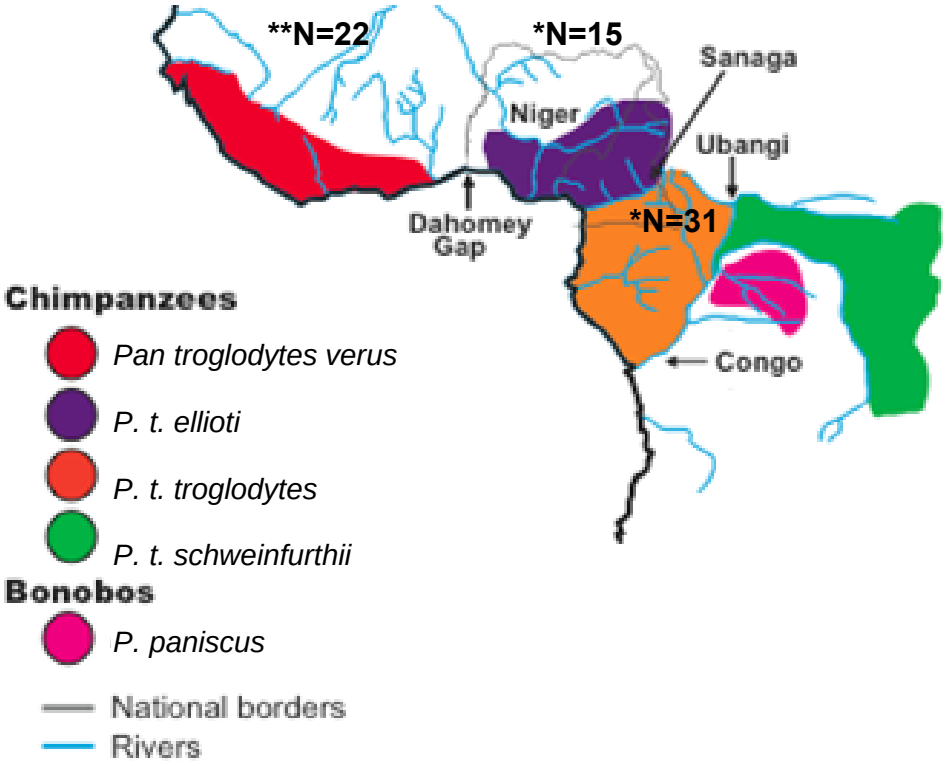
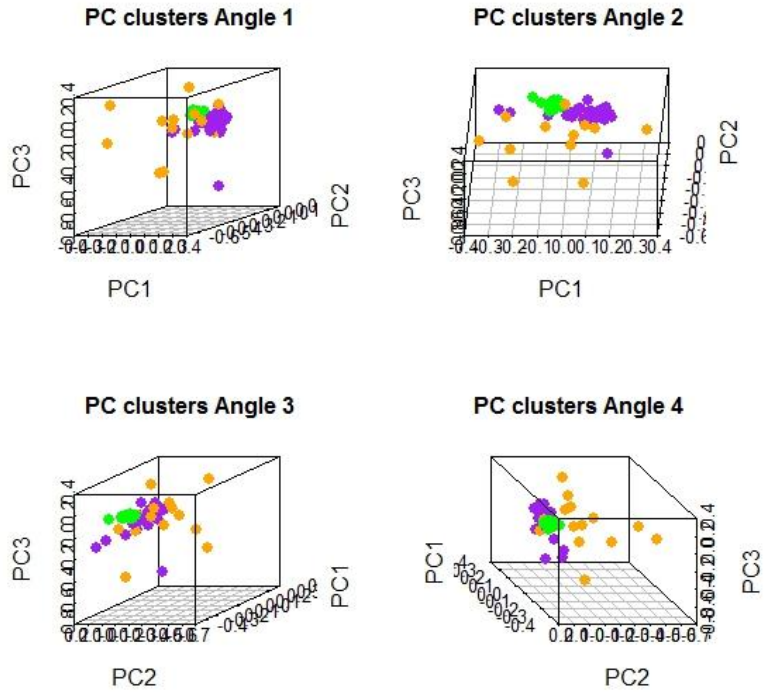
Sequence diversity of *Pan troglodytes* subspecies and the impact of *WFDC6* selective constraints in reproductive immunity.

Zélia Ferreira, Belen Hurle, Aida M. Andrés, Warren Kretzschmar, Jim Mullikin Praveen Cherukuri, Pedro Cruz, Mary Katherine Gonder, Anne Stone, Sarah Tishkoff, Willie Swanson, NISC Comparative Sequencing Program, Eric Green, Andrew G. Clark, and Susana Seixas.

Supplementary Figures

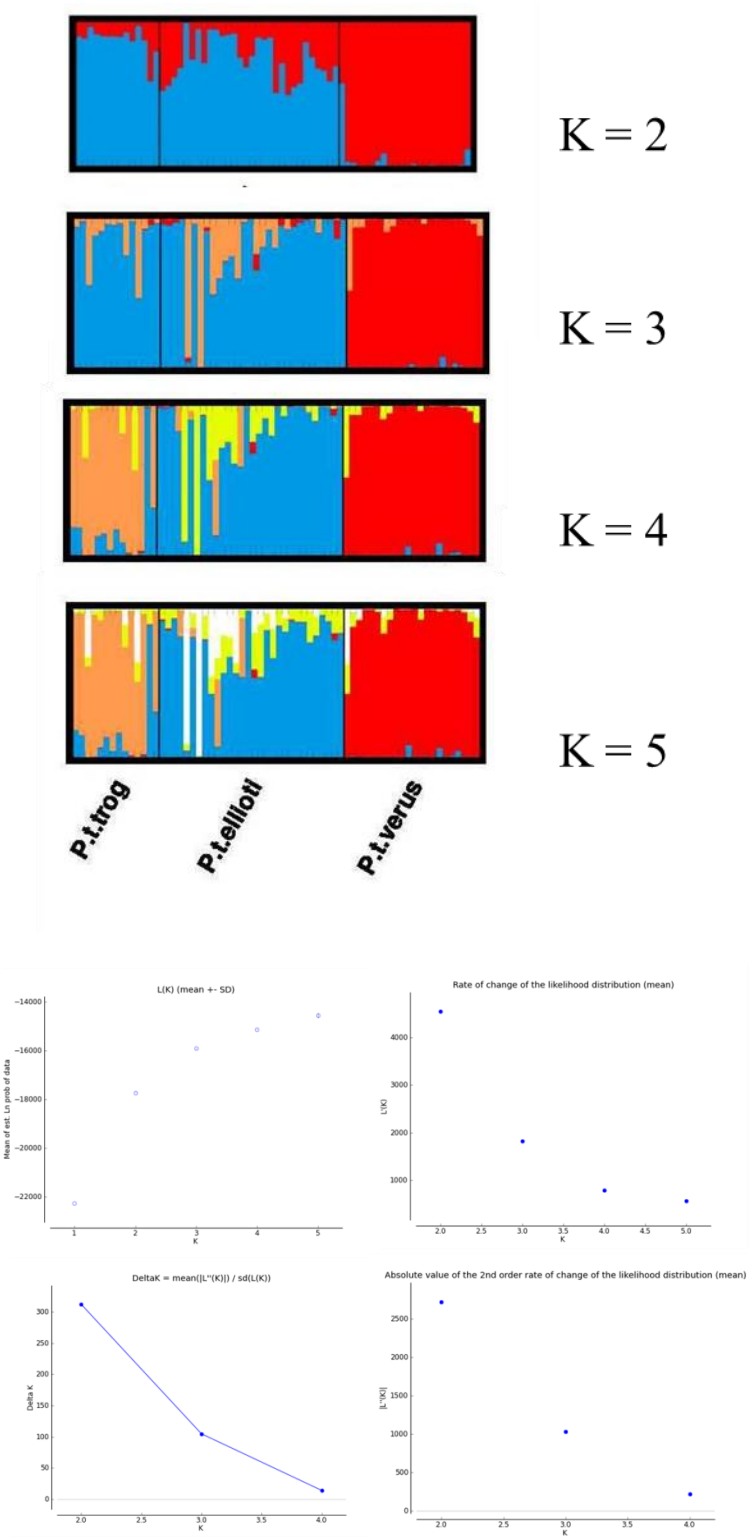


Supplementary Figure S1: Geographic distribution of *P. troglodytes* subspecies in Africa (adapted from Gonder et al. 2011 (Gonder, et al. 2011)).

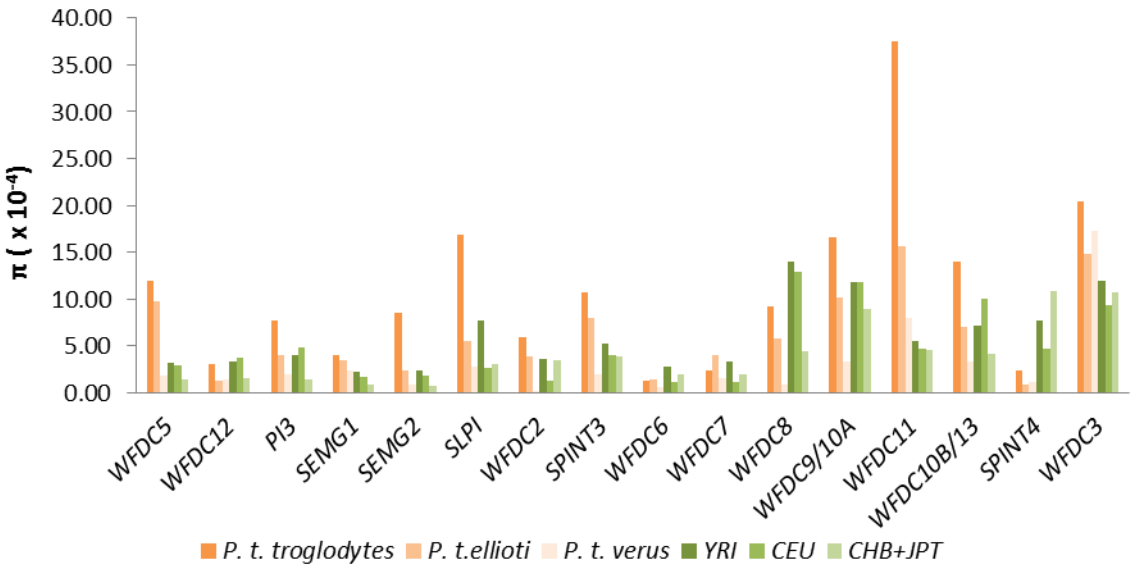


- *P. t. troglodytes*
- *P. t. ellioti*
- *P. t. verus*

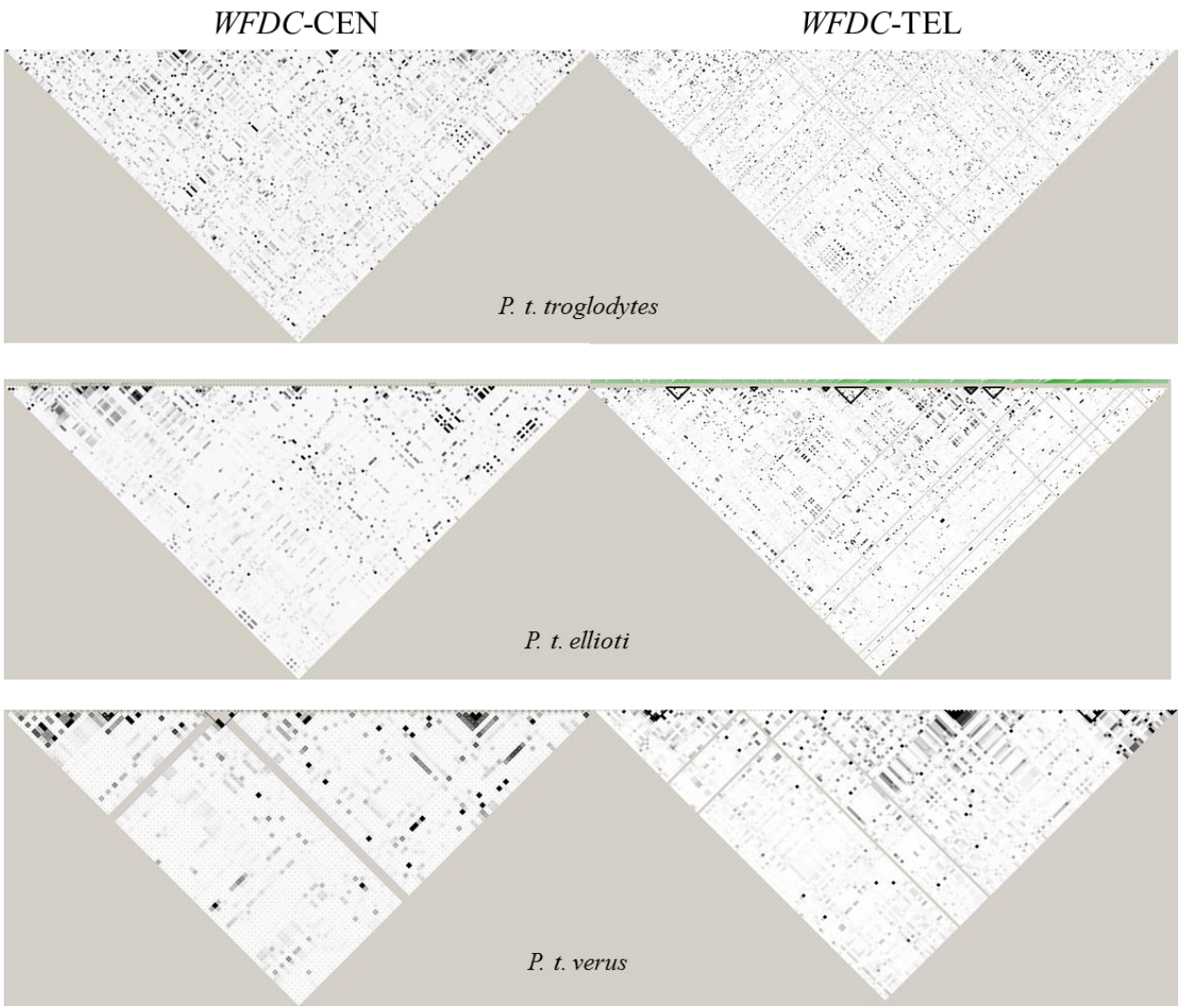
Supplementary Figure S2: PCA plots of 419 SNPs in the control regions. Eigen values were calculated in eigenstrat and plots were made using R (Patterson, et al. 2006).



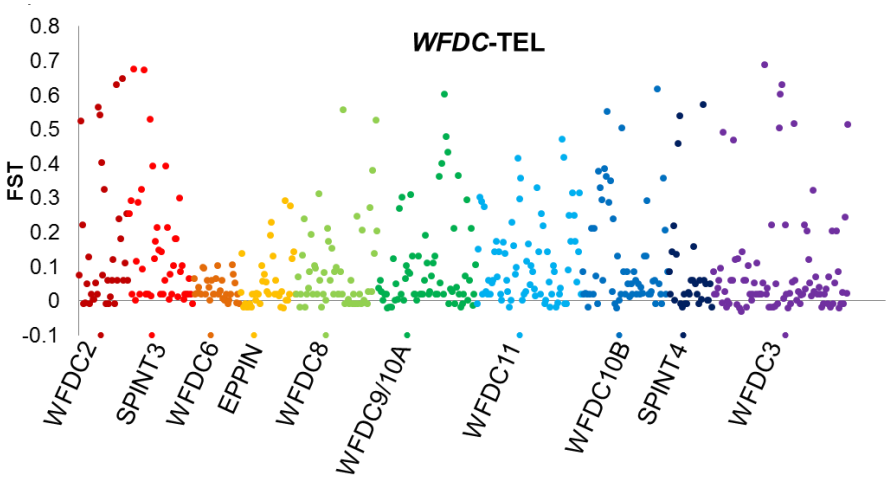
Supplementary Figure S3: A) Bar plot output from STRUCTURE; B) Likelihoods calculated from STRUCTURE outputs, using Harvester (http://taylor0.biology.ucla.edu/structureHarvester/) (Pritchard, et al. 2000).



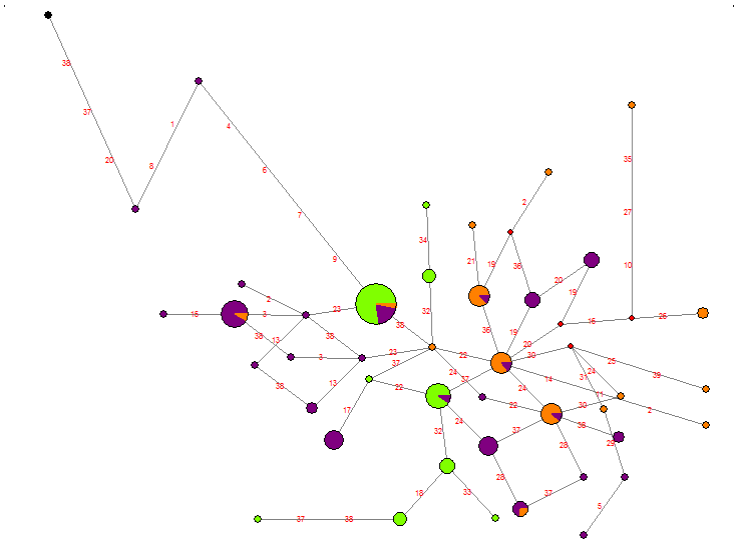
Supplementary Figure S4: Distribution of nucleotide diversity for all the *WFDC* genes in the sequenced chimpanzee subspecies and the corresponding regions in the 1000 genomes database (The 1000 Genomes Project Consortium 2010).



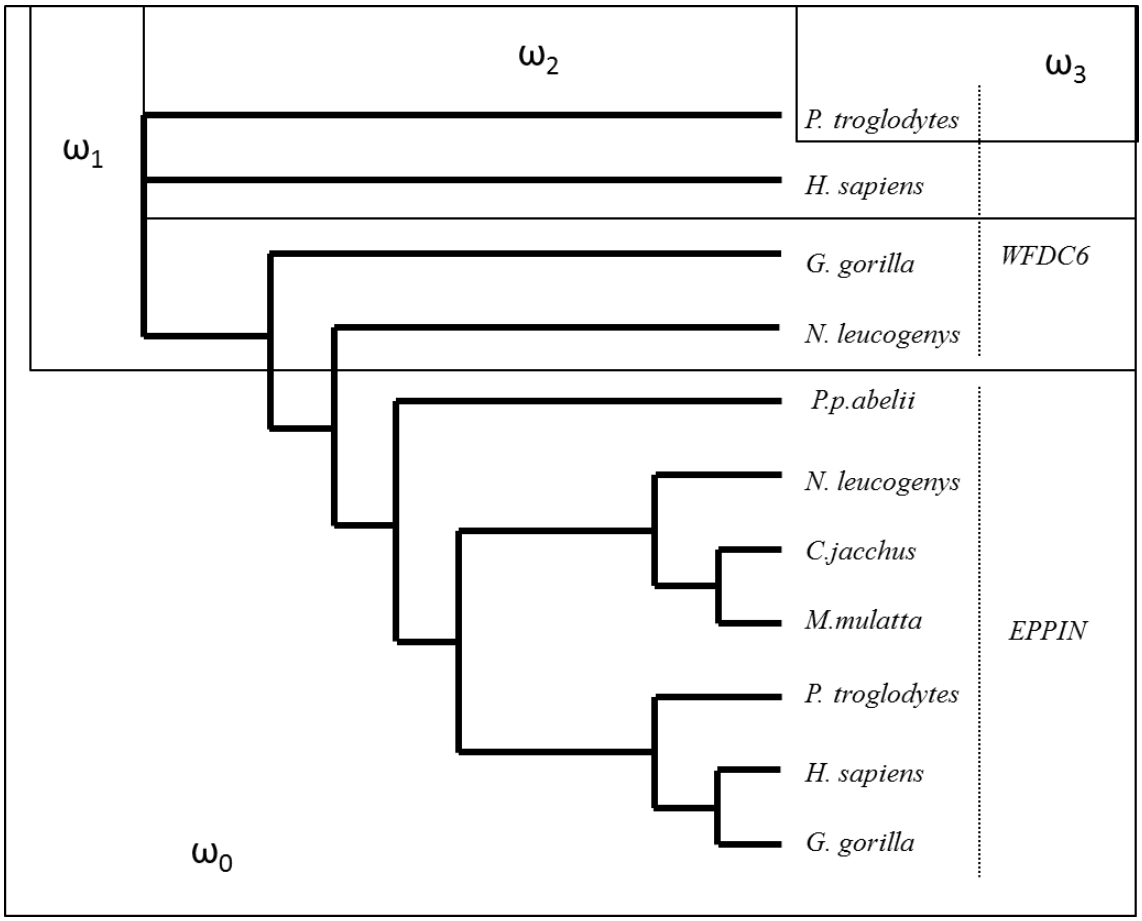
Supplementary Figure S5: LD plots from Haploview (r^2) (Barrett 2009; Barrett, et al. 2005).



Supplementary Figure S6: F_{ST} values for all the populations, plotted based on genomic position (Excoffier 2002).



Supplementary Figure S7: Inferred network haplotypes at the *EPPIN*. Each circle represents a unique haplotype, and its area is proportional to its frequency. Within each circle, *P. t. verus*, *P. t. ellioti*, and *P. t. troglodytes* are labeled in green, purple and orange, respectively. The mutations that differentiate each haplotype are shown along each branch.(Bandelt, et al. 1999).



Supplementary Figure S8: Phylogenetic analysis of *WFDC6* and *EPPIN* in primates, showing d_N/d_S ratios. The different branch models are represented as follows: ω_0, one ratio model; ω_1, two-ratio model; ω_2 the three-ratio model 1; and ω_3 the three-ratio model 2.

Supplementary Figures References

Bandelt HJ, Forster P, Rohl A 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37-48.
Barrett JC 2009. Haploview: Visualization and analysis of SNP genotype data. Cold Spring Harbor protocols 2009: pdb ip71. doi: 10.1101/pdb.ip71
Barrett JC, Fry B, Maller J, Daly MJ 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263-265. doi: 10.1093/bioinformatics/bth457
Excoffier L 2002. Human demographic history: refining the recent African origin model. Current opinion in genetics & development 12: 8.
Gonder MK, et al. 2011. Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations. PNAS 108: 12.
Patterson N, Price AL, Reich D 2006. Population structure and eigenanalysis. PLoS Genetics 2: e190. doi: 10.1371/journal.pgen.0020190
Pritchard JK, Stephens M, Donnelly P 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics 155: 14.
The 1000 Genomes Project Consortium 2010. A map of human genome variation from population-scale sequencing. Nature 467: 1061-1073. doi: 10.1038/nature09534