Supplemental Figures

Fig. S1

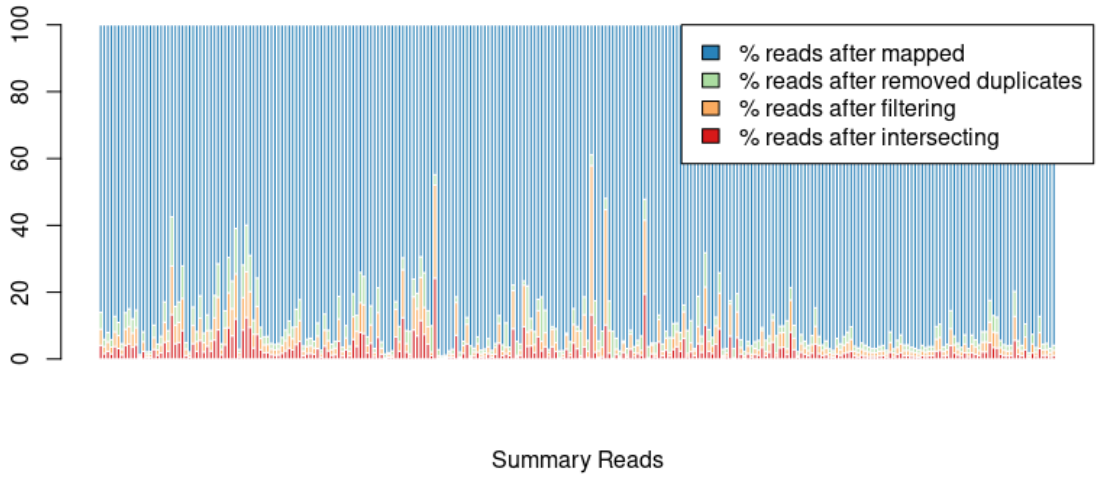

Fig. S1: Percentage of reads between each step in our pipeline: blue is the final amount left after mapping, removing duplicates, filtering and intersecting. Image created with R [82,83].

Fig. S2

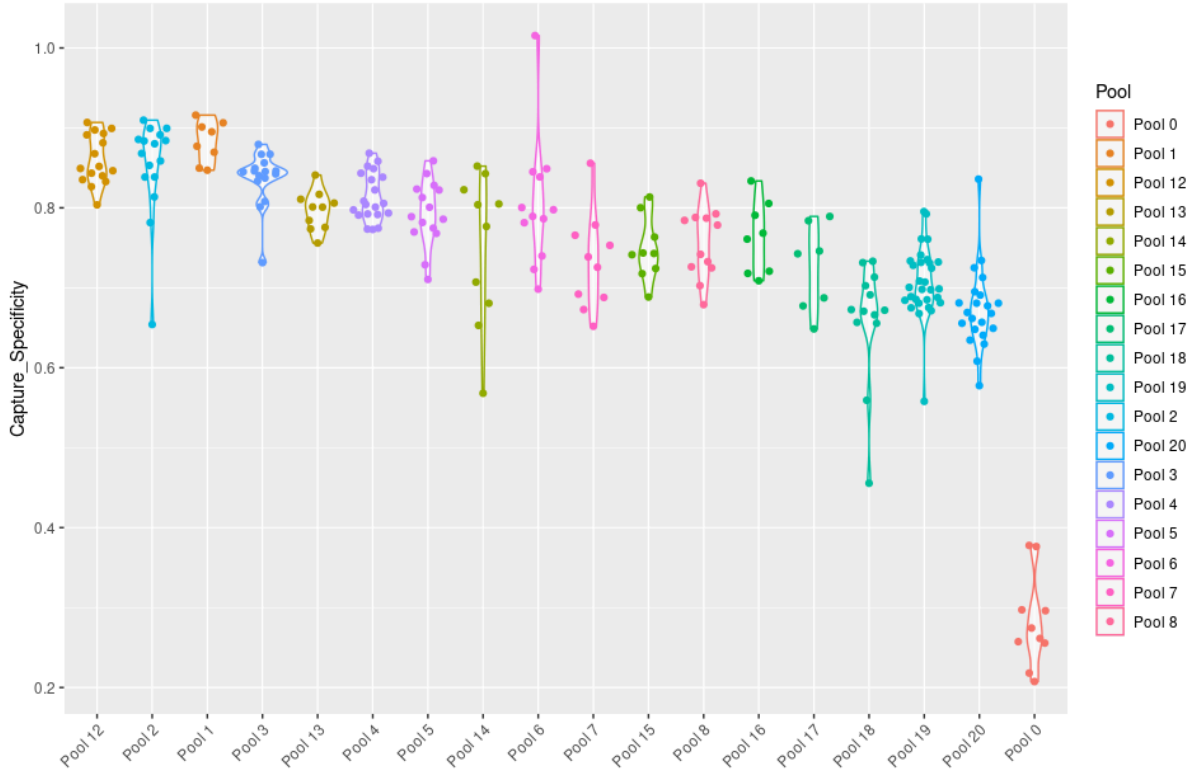

Fig. S2: Capture specificity of each pool, calculated by dividing reliable reads on-target by reliable reads. Image created with R [82,83].

Fig. S3

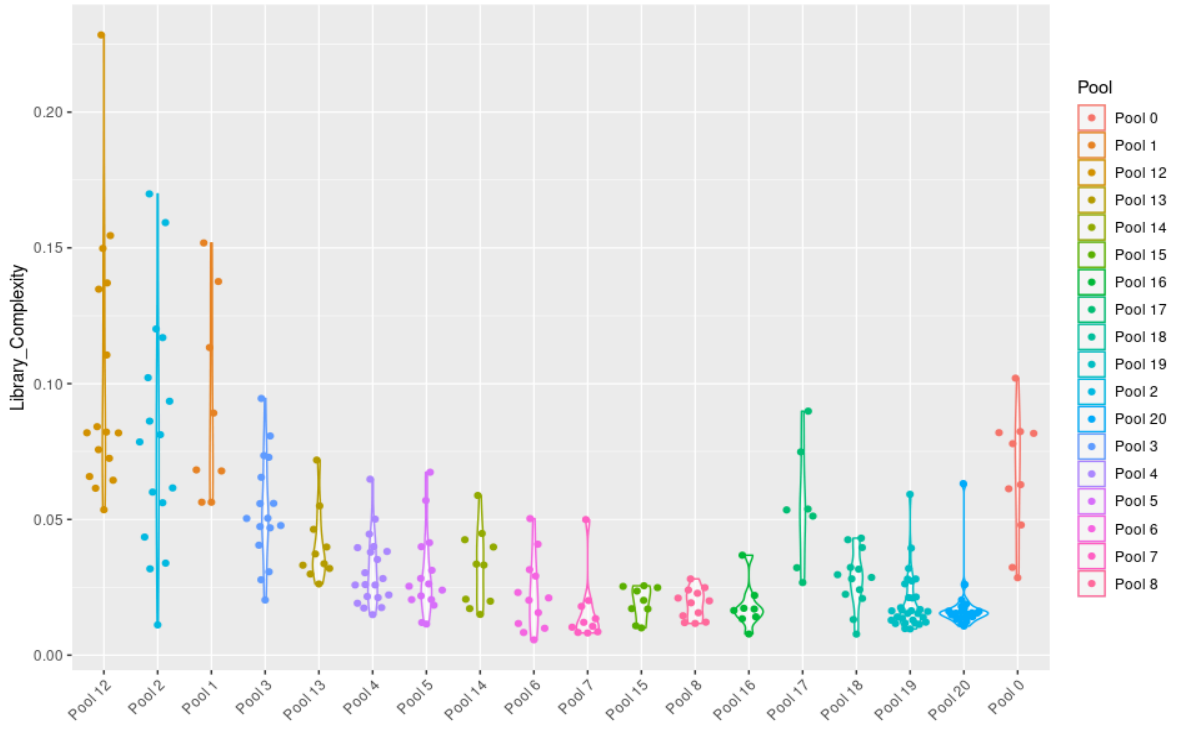

Fig. S3: Library complexity of each pool, calculated by dividing reliable reads by mapped reads. Image created with R [82,83].

Fig. S4

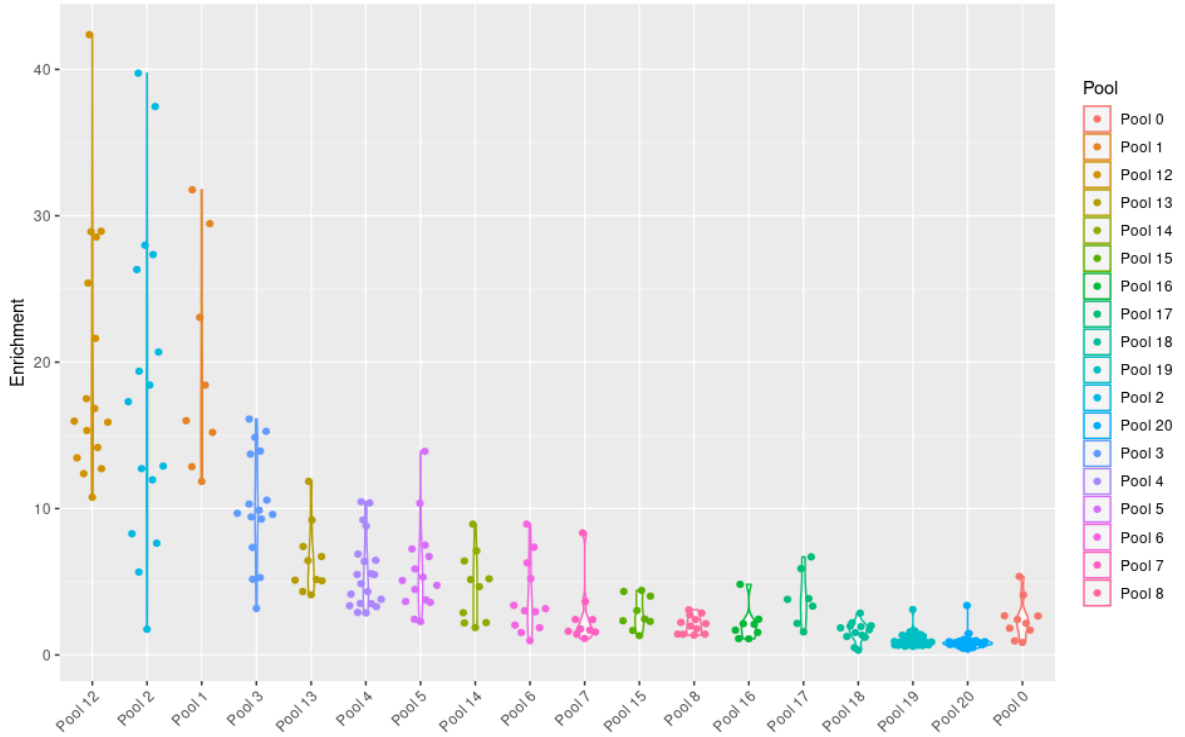

Fig. S4: Enrichment performance of each pool, calculated by the following: (reliable reads on-target/by mapped reads)/(chr21 reads/genome size). Image created with R [82,83].

Fig. S5

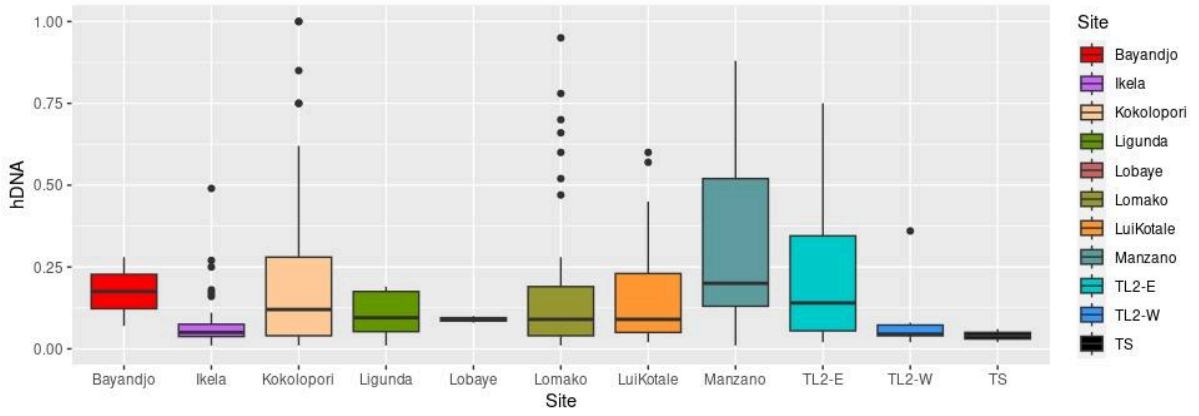

Fig. S5: Boxplot of percentage host DNA in each site prior to sequencing. Image created with R [82,83].

Fig. S6

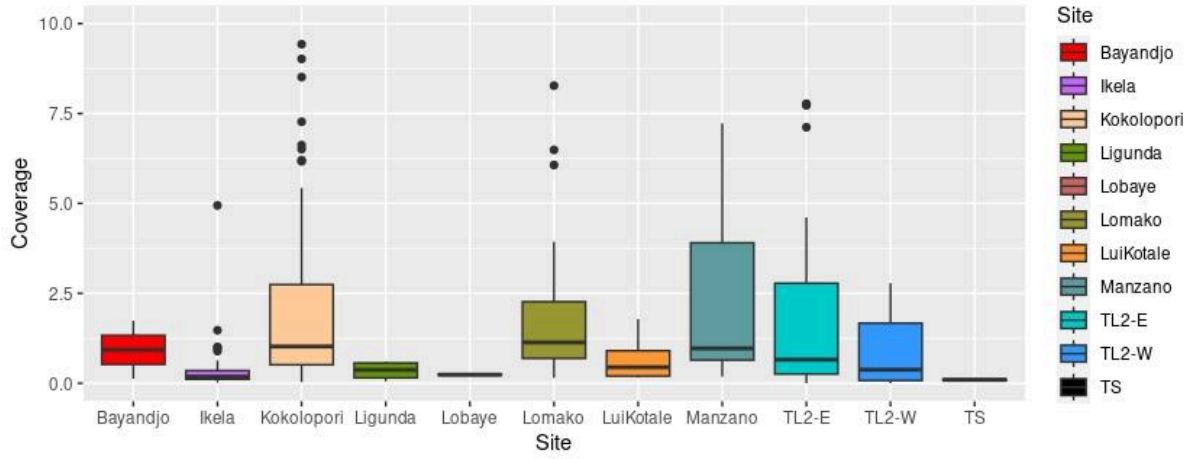

Fig. S6: Boxplot of depth of coverage in each site after filtering at ≥0.1x and 1% human contamination. Image created with R [82,83].

Fig. S7

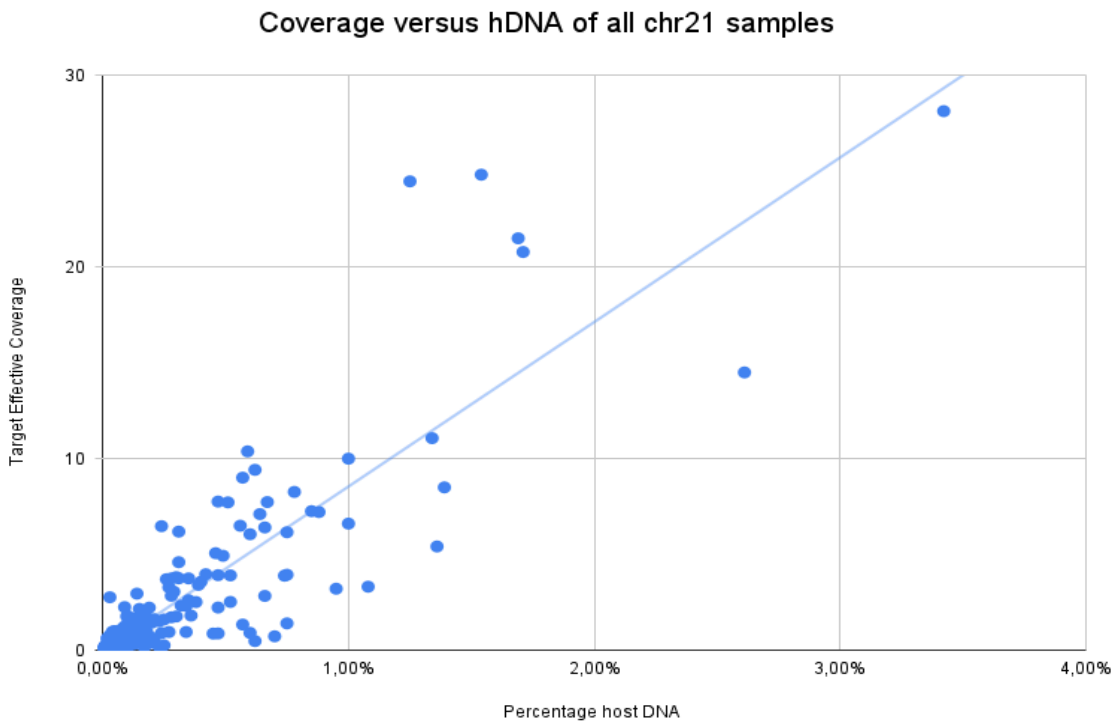

Fig. S7: Scatterplot of depth of coverage versus host DNA per sample. Image created with Microsoft Excel [95].

Fig. S8

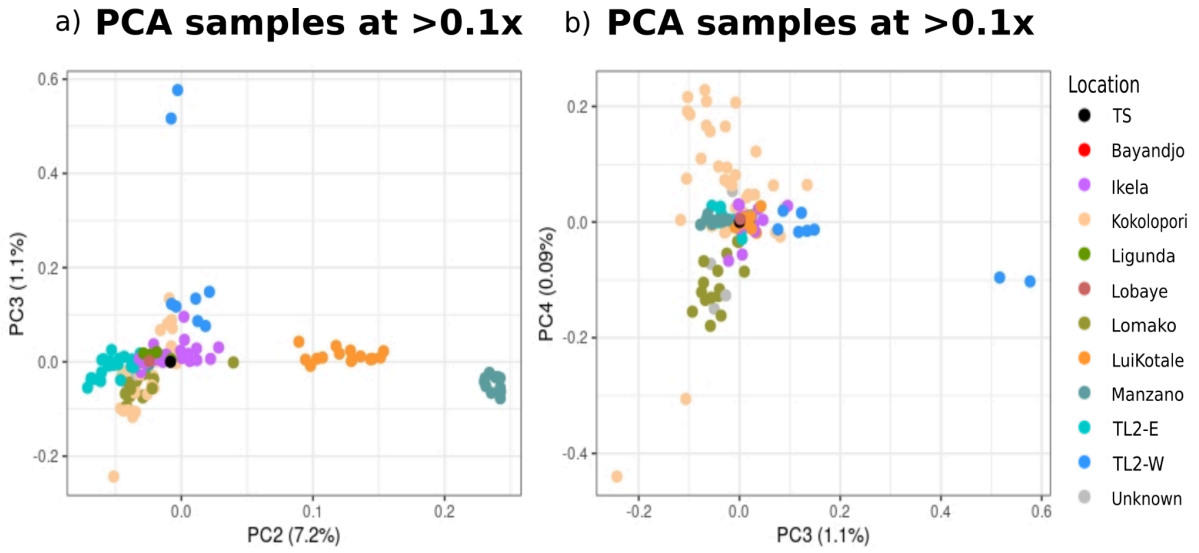

Fig. S8: Principal Component Analysis: a) components 2 and 3 using PCAngsd of all samples filtered at ≥0.1x & 1% human contamination; b) components 3 and 4 using PCAngsd of all samples filtered at ≥0.1x & 1% human contamination (Meisner & Albrechtsen 2018). Some samples in grey were mislabelled but have been identified as Lomako and Kokolopori. Image created with R [82,83].

Fig. S9

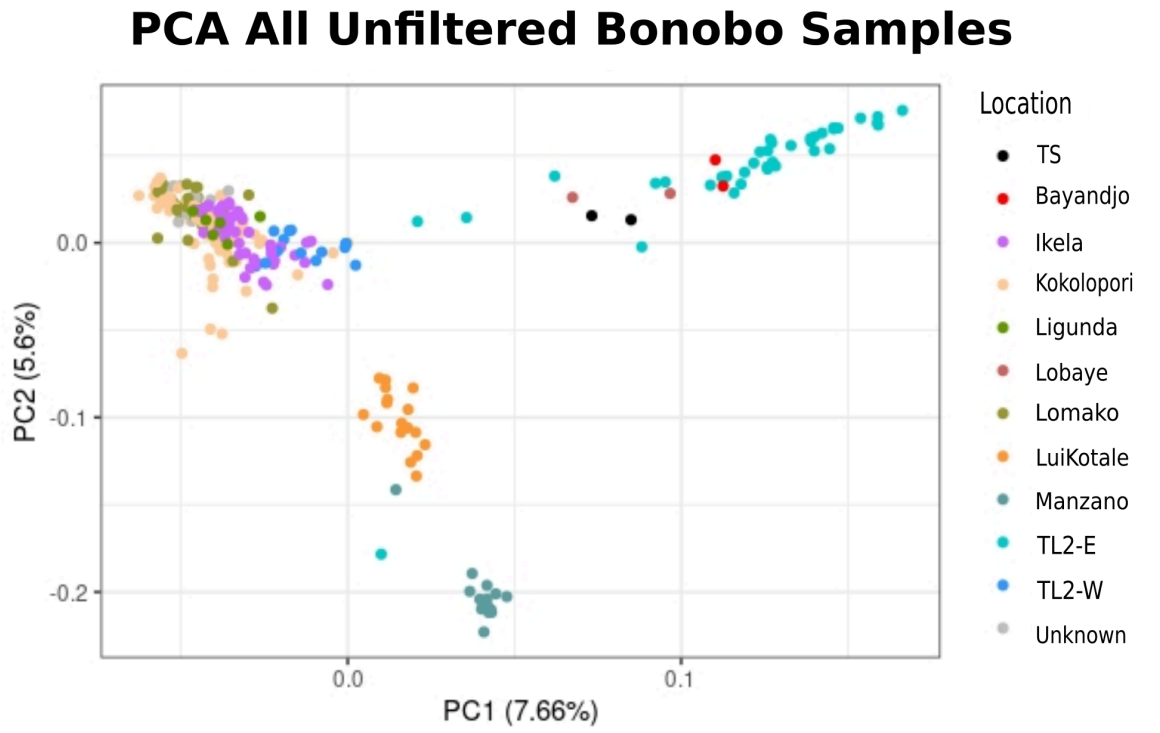

Fig. S9: Principal Component Analysis using PCAngsd (Meisner & Albrechtsen 2018) of all samples unfiltered. Some samples in grey were mislabelled but have been identified as Lomako and Kokolopori. Image created with R [82,83].

Fig. S10

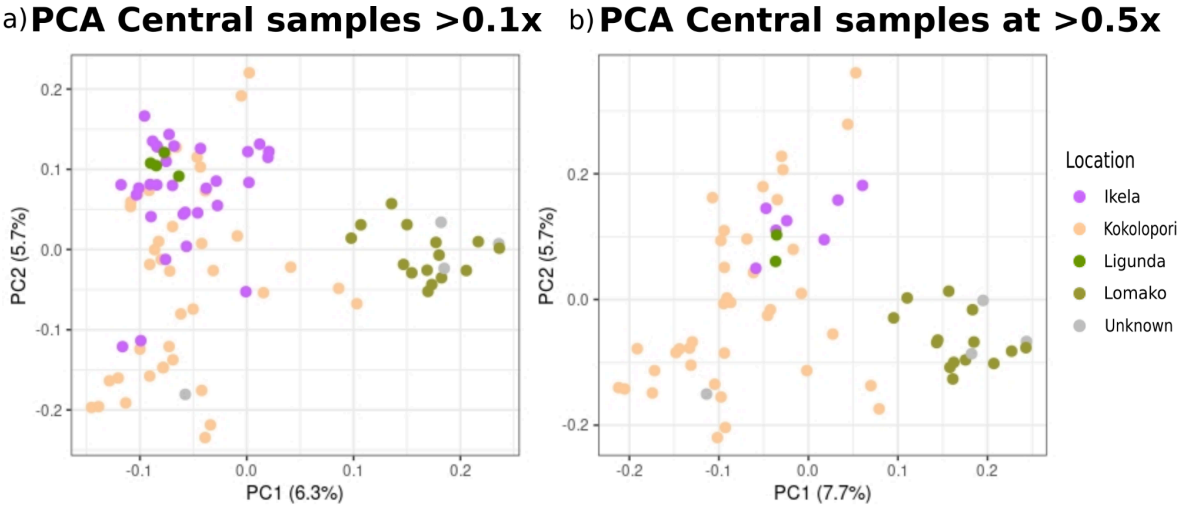

Fig. S10: Principal Component Analysis using PCAngsd (Meisner & Albrechtsen 2018): a) North-Central bonobo samples without TL2-W and filtered at ≥0.1x & 1% human contamination; b) North-Central bonobo samples without TL2-W and filtered at ≥0.5x & 1% human contamination. Central samples here are from Ikela, Kokolopori, Ligunda, Lomako and samples that were mislabelled but have been identified as Lomako and Kokolopori. Image created with R [82,83].

Fig. S11

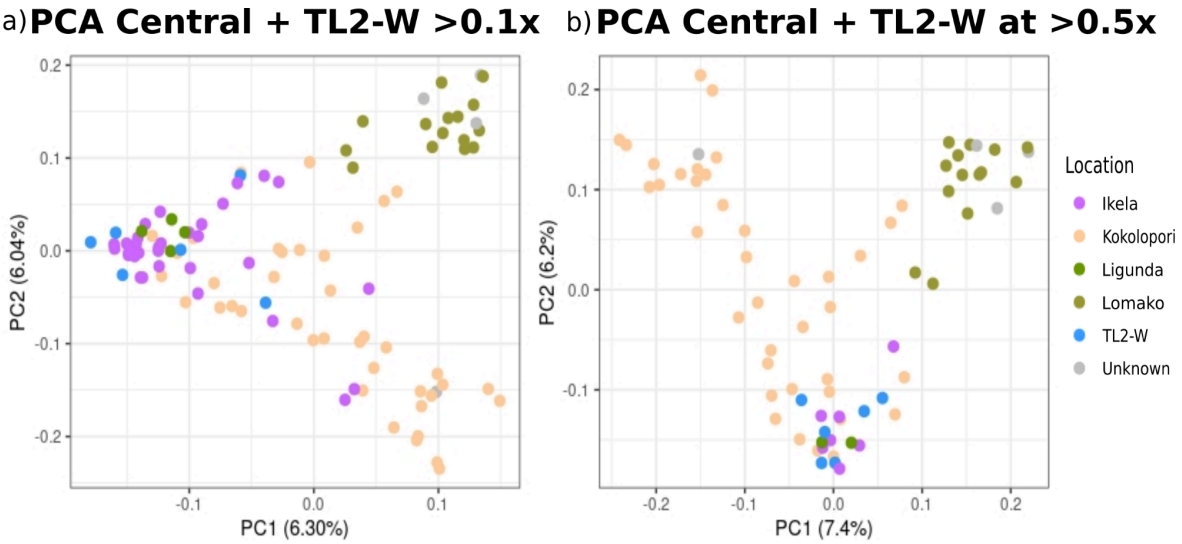

Fig. S11: Principal Component Analysis using PCAngsd (Meisner & Albrechtsen 2018): a) North-Central bonobo samples and TL2-W, and filtered at ≥0.1x & 1% human contamination; b) North-Central bonobo samples and TL2-W, and filtered at ≥0.5x & 1% human contamination. Samples are from TL2-W, Ikela, Kokolopori, Ligunda, Lomako and samples that were mislabelled but have been identified as Lomako and Kokolopori. Image created with R [82,83].

Fig. S12

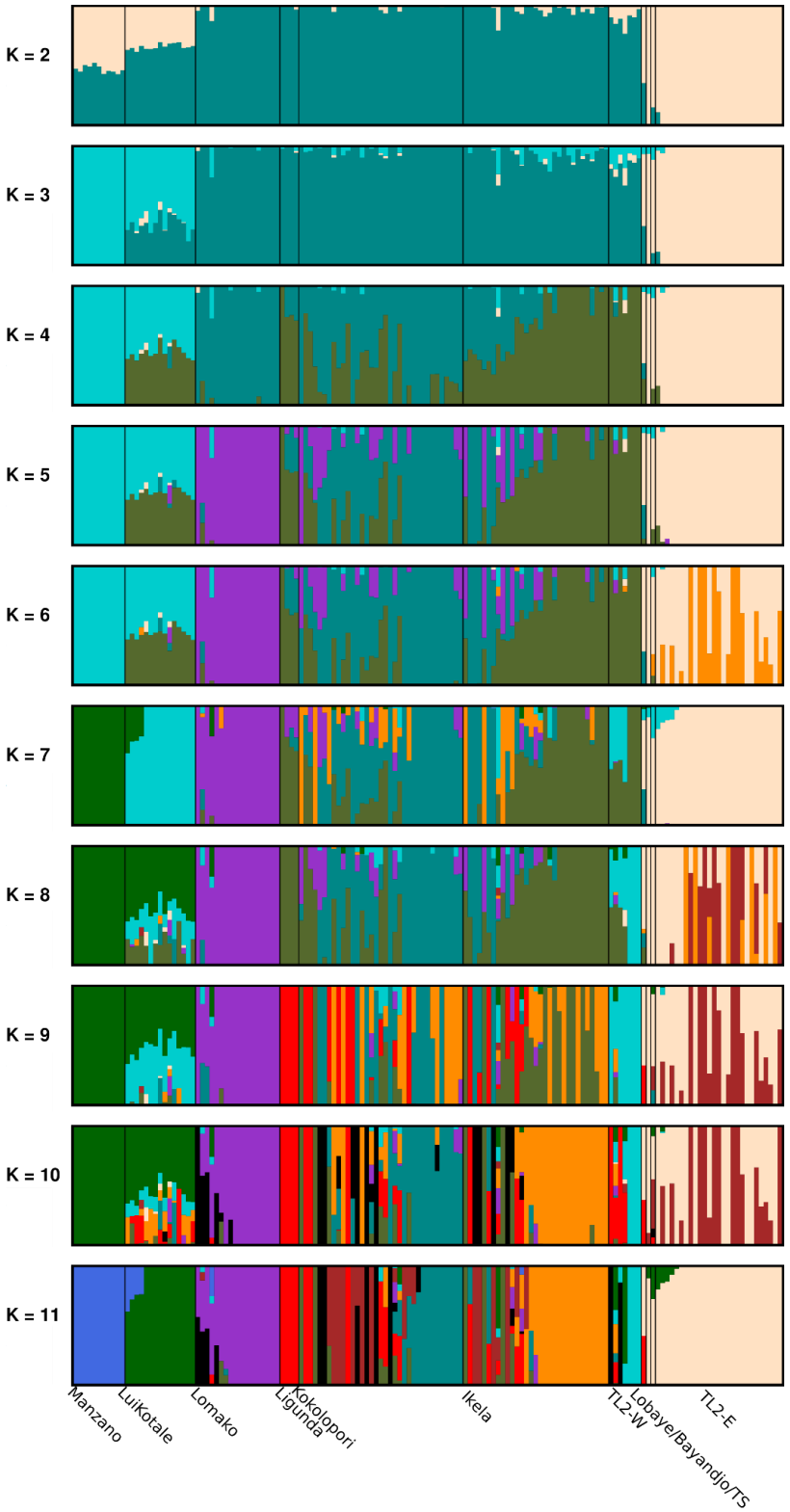

Fig. S12: Bar plot of admixture values per individual grouped per site at ≥ 0.1x depth of coverage of first run from K2 to K11 without whole genome samples. Image created with pong [52].

Fig. S13

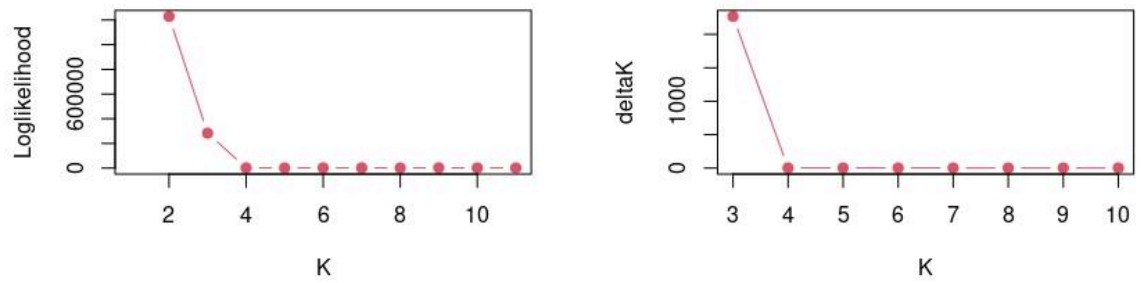

Fig. S13: NgsAdmix plot for inferred K: left is the inferred K calculated from the normalised mean of likelihoods from all Ks; right is the inferred K calculated from the normalised mean of differences from likelihoods of Ks. Image created with R [82,83].

Fig. S14

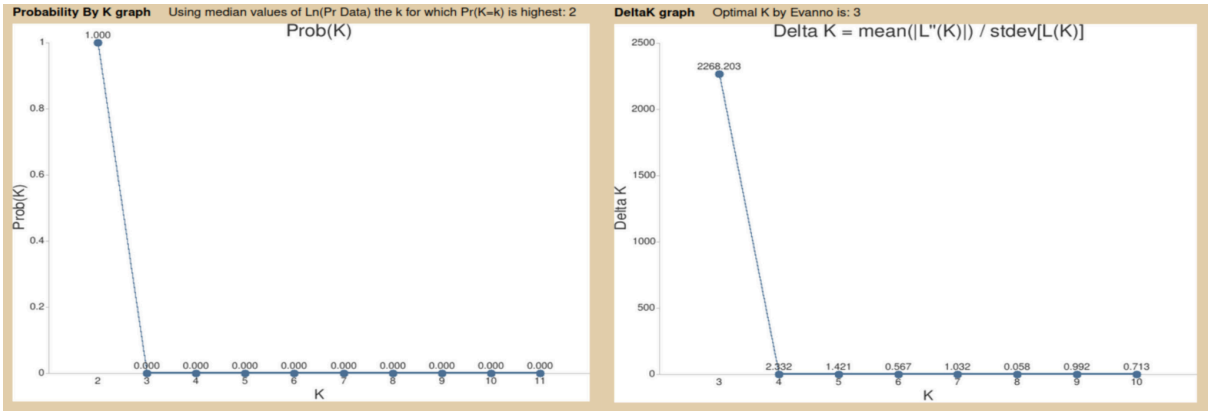

Fig. S14: Clumpak plot for inferred K: left is the inferred K calculated from the normalised mean from all Ks; right is the inferred K calculated from the normalised mean of differences from likelihood Ks. Output image created by Clumpak (https://tau.evolseq.net/clumpak/index.html).

Fig. S15

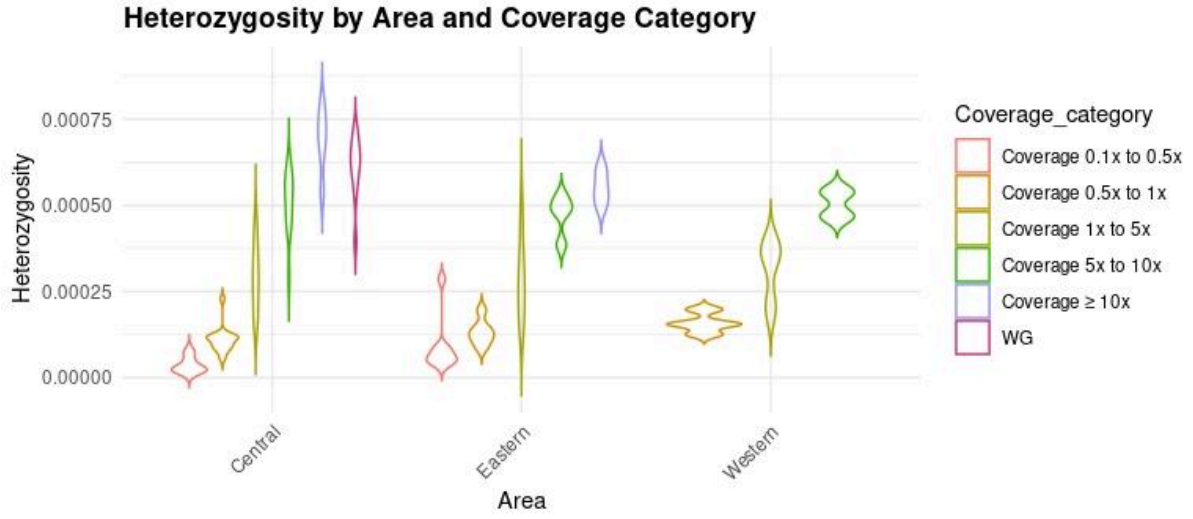

Fig. S15: Violin plot of heterozygosity levels per area at different depths of coverages with whole genome bonobo samples (WG) from the Prado et al. 2013 paper. Image created with R [82,83].

Fig. S16

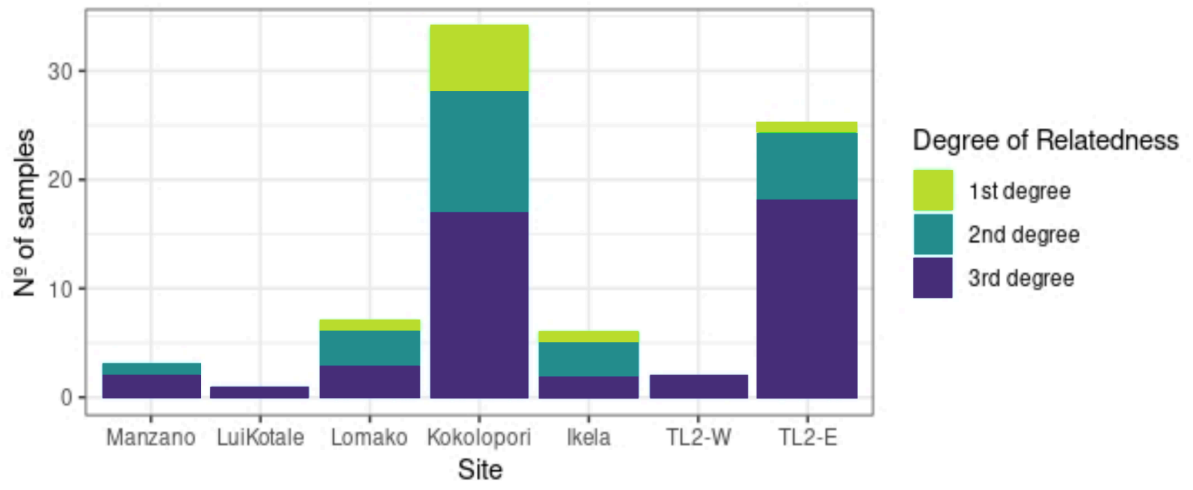

Fig. S16: Bar plot of total number of related samples for each site at 0.3x depth of coverage or more. Sample size of first degree for each site: Lomako (n=1); Kokolopori (n=6); Ikela (n=1); TL2-E (n=1). Image created with R [82,83].

Fig. S17

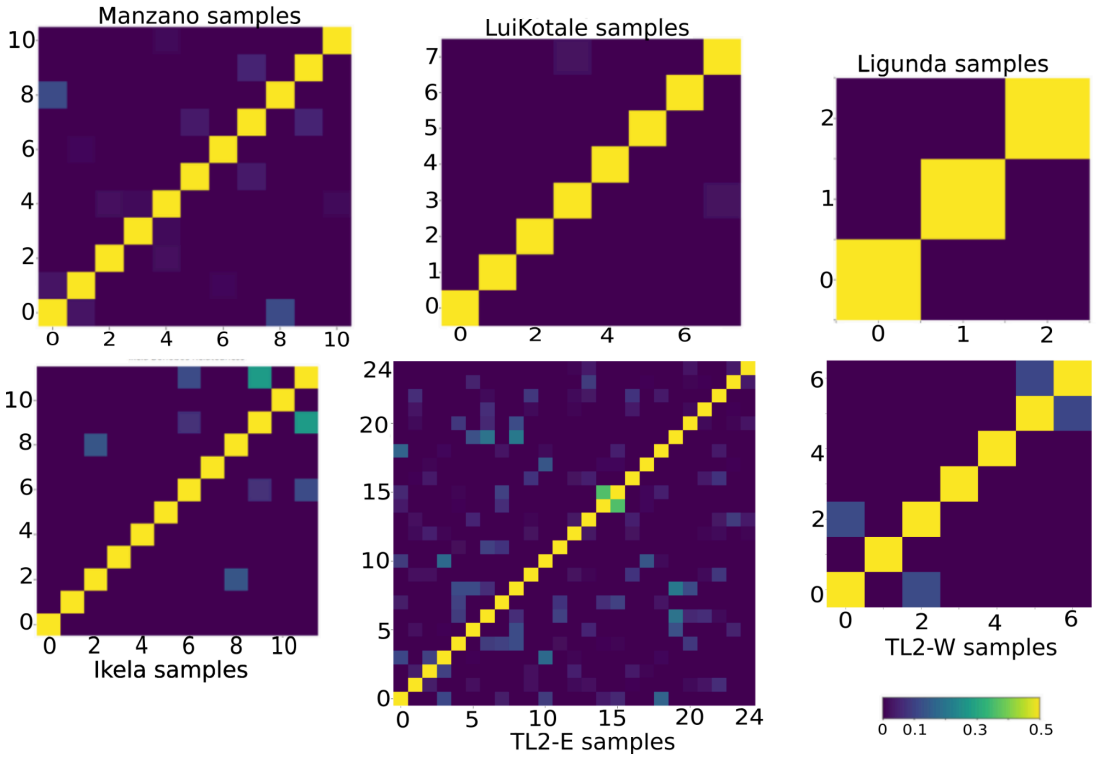

Fig. S17: Relatedness-site matrices made using NgsRelate v.2 (Korneliussen et al. 2015) of all bonobo samples filtered, QC-passed and with depth of coverage above 0.3x. Axis represents the number of samples and the gradient is level of relatedness. Closer to yellow is a higher level whilst yellow represents the same individual. Images were created with R [82,83].

Fig. S18

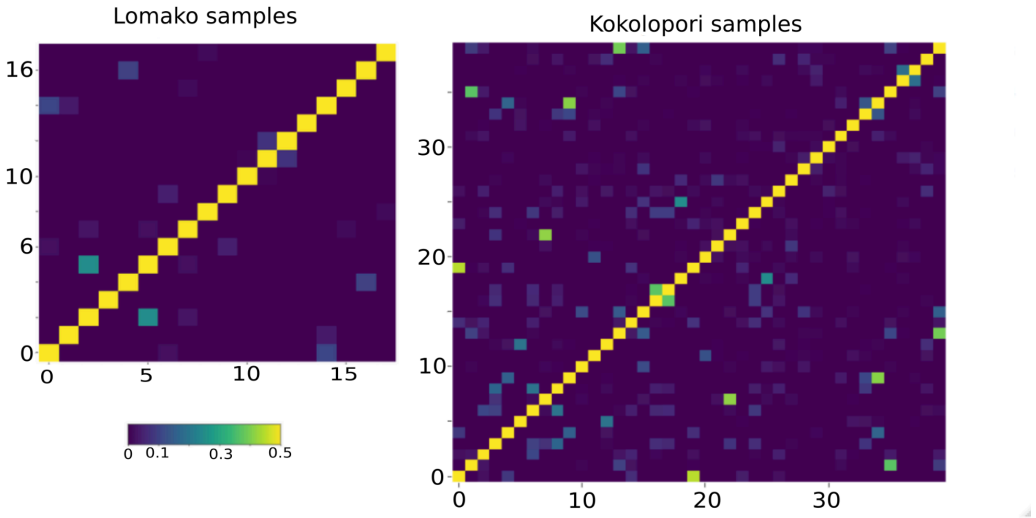

Fig. S18: Site matrices of LuiKotale and Kokolopori samples made using NgsRelate v.2 (Korneliussen et al. 2015) of all bonobo samples filtered, QC-passed and with depth of coverage above 0.3x. Axis represents the number of samples and the gradient is level of relatedness. Closer to yellow is a higher level whilst yellow represents the same individual. Images were created with R [82,83].

Fig. S19

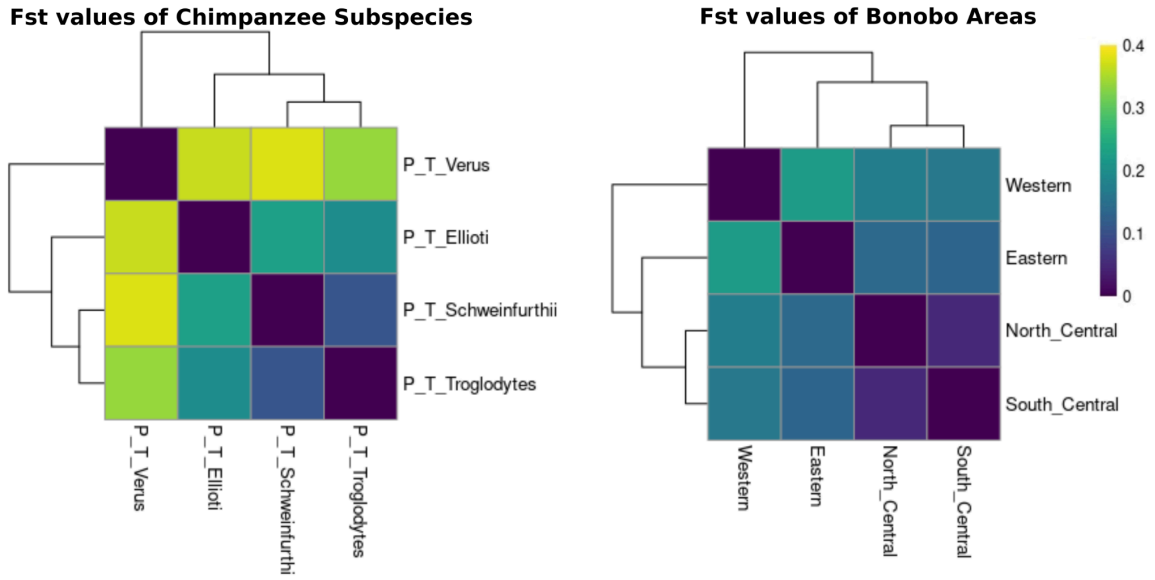

Fig. S19: Heat-maps of F_{ST} indexes: left are values for chimpanzee subspecies; right are values for bonobo areas. Barriers such as rivers were not taken into account as we use large groups of sites and subspecies. Bonobo areas are here defined as: Western = Manzano samples, Eastern = TL2-E samples, North Central = Lomako and Kokolopori samples, South Central = Ikela and TL2-W samples. Darker colouring indicates lower genetic differentiation whilst the lighter yellow colouring indicates higher genetic variation. Image created with R [82,83].

Fig. S20

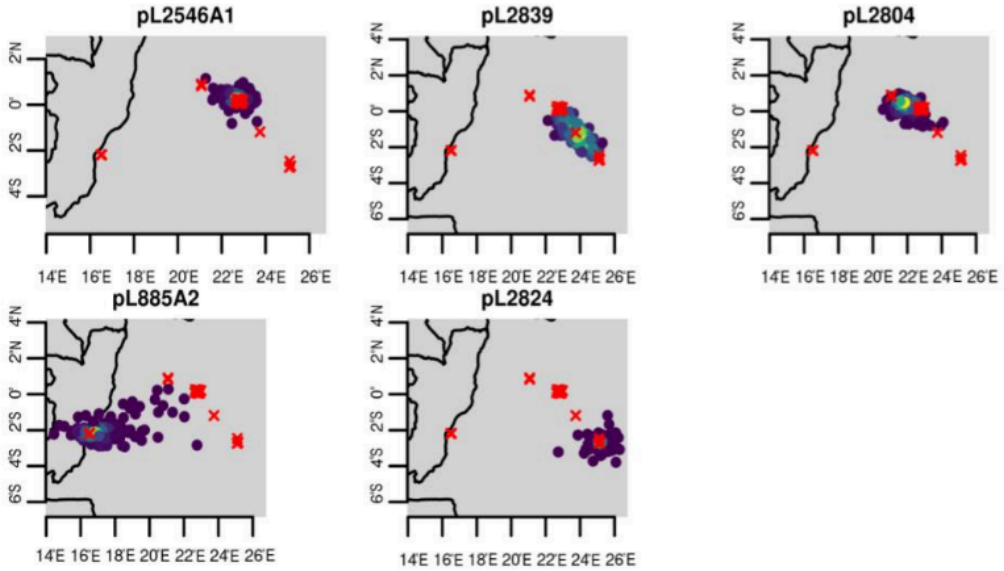

Fig. S20: Output test plots of Locator data with real coordinates used for reference (red crosses) and predicted location (circles, with colour grading according to number of coinciding predictions - lighter ones being the most probable and darker least). Samples' real origin is: pL2546A1 - Kokolopori; pL2839 - TL2-W; pL2804 - Lomako; pL885A2 - Manzano; pL2824 - TL2-E. Image edited with R [82,83].

Fig. S21

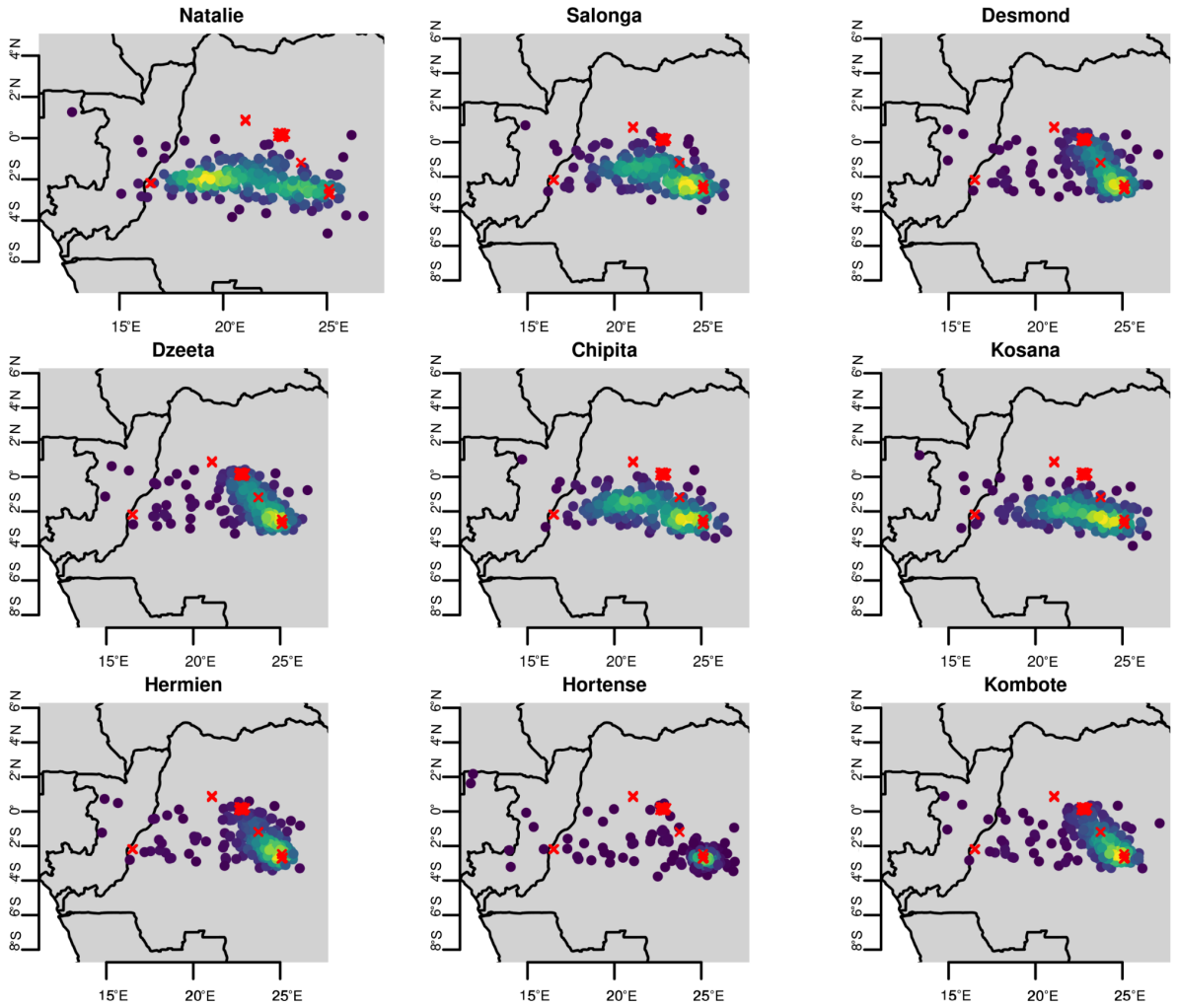

Fig. S21: Output plots of Locator data with real data coordinates used for reference (red crosses) and predicted location (circles, with colour grading according to number of coinciding predictions - lighter ones being the most probable and darker the least).

Fig. S22

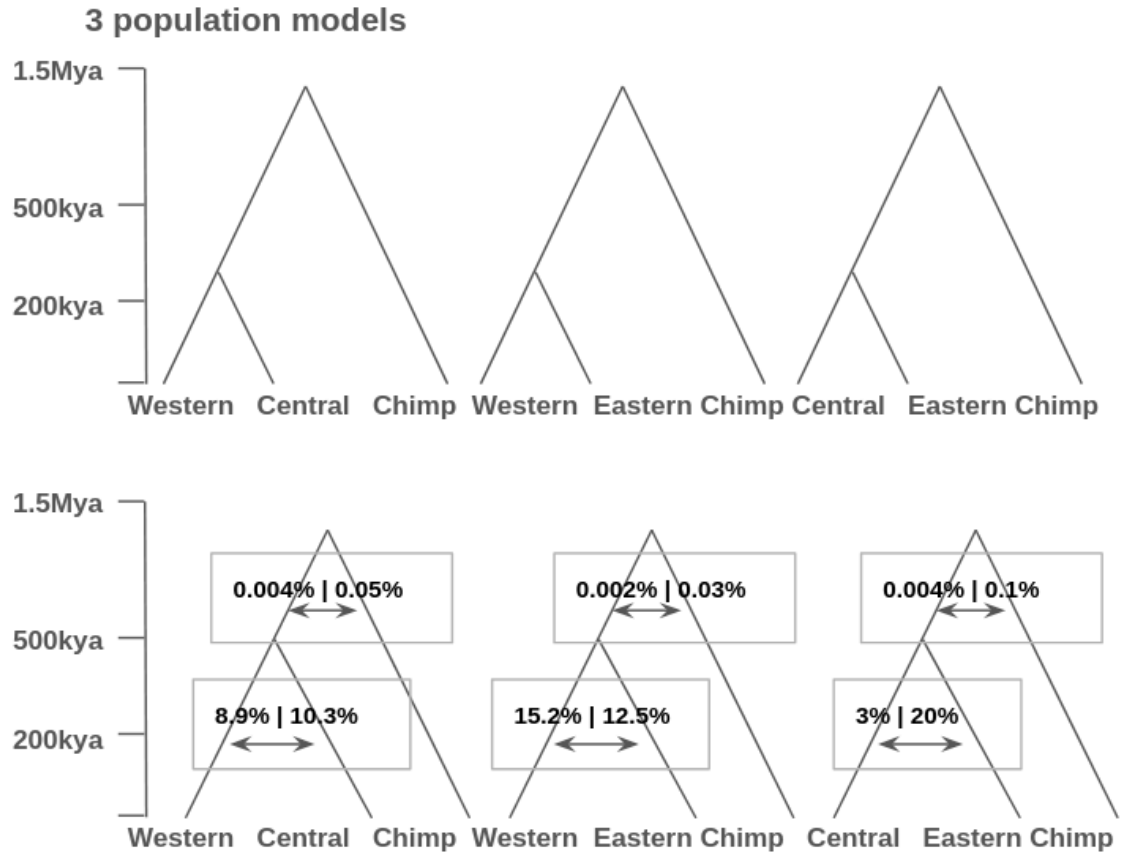

Fig. S22: A schematic representation of the G-PhoCS analyses. Each analysis included two bonobo populations and a Western chimpanzee lineage as an outgroup. We considered three models without migration (top) and three models with migration (bottom). Models with migration included four directional migration bands, as depicted in the figure. The scaled inferred divergence times are depicted by the height of the corresponding split in the tree (see y-axis). Inferred migration rates in % are written on the arrows, e.g. 0.004% from bonobos to chimpanzees and 8.9% from western to central bonobos on the WeCeChimp model on the bottom left. See Table S4 for the estimated values of these parameters.

Fig. S23

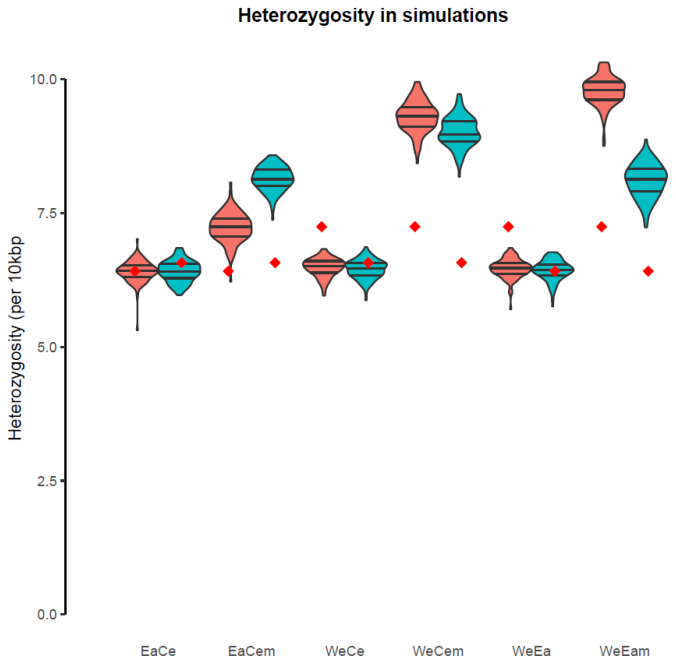

Fig. S23: Heterozygosity (per 10,000 bp) for single simulated individuals (100 replicates) given the parameters inferred by G-PhoCS on the X-axis. Combinations of two populations from the model used with or without migrations are marked on the Y-axis: such as EaCe (Eastern and Central without migration) and EaCem (Eastern and Central with migration). Distribution in red violin plots represents the values from the first population and in green the second population, with real data shown as red dots.