

Supplementary Materials

Supplementary Materials and Methods

Detailed information on samples

We sequenced DNA samples of two *M. arctoides* and two *M. thibetana* individuals, all of wild origin. The Marc_R02 (*M. arctoides*) sample comes from Qiaojia County, Zhaotong City, Yunnan Province, China. The Marc_R19 (*M. arctoides*) sample comes from Guangxi Zhuang Autonomous Region, China. The Mthi_HT1 (*M. thibetana*) sample comes from Mount Huang, Huangshan City, Anhui Province, China. The Mthi_R25 (*M. thibetana*) sample comes from Guangxi Zhuang Autonomous Region, China.

We downloaded whole-genome sequencing data of 10 *Macaca* individuals, including seven species of *Macaca*, and one individual of *P. anubis*. The Mass (*M. assamensis*, NCBI accession No. SRR2981114) sample is likely a wild-born individual from Yunnan Province, China (Fan et al., 2018). The Mfas_1 (*M. fascicularis*, NCBI accession No. SRR8194877) sample is a captive-bred individual from China (Zhang et al., 2020), but its origin is unknown. The Mfas_Mau (*M. fascicularis*, NCBI accession No. ERS629711) sample is wild caught from Mauritius (Osada et al., 2015). The Mfus (*M. fuscata*, NCBI accession No. DRR002234) sample comes from Japan, but whether it is captive bred or wild is not known. The samples Mmul_Chi and Mmul_Ind (*M. mulatta*, NCBI accession Nos. SRR1944102 and SRR1952166) are of Chinese and Indian origin and show no evidence of admixture between regional populations (Xue et al., 2016). The origin of the Mnem_1 (*M. nemestrina*, NCBI accession No. SAMN03121842) sample, which was used for reference genome assembly, is unknown. The sample Mnem_2 (*M. nemestrina*, NCBI accession No. SRR5947292) is a

captive individual, with a geographical origin in Borneo (Evans et al., 2017). The Mnig (*M. nigra*, NCBI accession No. SRR5947294) and Mton (*M. tonkeana*, NCBI accession No. SRR5947293) samples are captive individuals originally from Sulawesi (Evans et al., 2017). The sample Panu (*P. anubis*, NCBI accession No. SRR8723580) is a pedigree captive Olive baboon (Robinson et al., 2019).

REFERENCES

- Evans BJ, Tosi AJ, Zeng K, Dushoff J, Corvelo A, Melnick DJ. 2017. Speciation over the edge: gene flow among non-human primate species across a formidable biogeographic barrier. *Royal Society Open Science*, **4**(10): 170351.
- Fan ZX, Zhou AB, Osada N, Yu JQ, Jiang J, Li P, et al. 2018. Ancient hybridization and admixture in macaques (genus *Macaca*) inferred from whole genome sequences. *Molecular Phylogenetics and Evolution*, **127**: 376–386.
- Osada N, Hettiarachchi N, Babarinde IA, Saitou N, Blancher A. 2015. Whole-genome sequencing of six mauritian cynomolgus macaques (*Macaca fascicularis*) reveals a genome-wide pattern of polymorphisms under extreme population bottleneck. *Genome Biology and Evolution*, **7**(3): 821–830.
- Robinson JA, Belsare S, Birnbaum S, Newman DE, Chan J, Glenn JP, et al. 2019. Analysis of 100 high-coverage genomes from a pedigreed captive baboon colony. *Genome Research*, **29**(5): 848–856.
- Xue C, Raveendran M, Harris RA, Fawcett GL, Liu XM, White S, et al. 2016. The population genomics of rhesus macaques (*Macaca mulatta*) based on whole-genome

sequences. *Genome Research*, **26**(12): 1651–1662.

Zhang WH, Aida T, Del Rosario RCH, Wilde JJ, Ding CH, Zhang XH, et al. 2020. Multiplex precise base editing in cynomolgus monkeys. *Nature Communications*, **11**(1): 2325.

Supplementary Tables

Supplementary Table S1. Constraints used for divergence time estimation

Prior distribution	Type of distribution	Constrained sample
normal(0,3.0254,0.2971)	crown	Mfas_1,Mfas_Mau,Mfus,Mmul_Chi,Mmul_Ind
normal(0,3.6134,0.3259)	crown	Marc_R02,Marc_R19,Mass,Mthi_HT1,Mthi_R25,Mfas_1,Mfas_Mau,Mfus,Mmul_Chi,Mmul_Ind
normal(0,3.9137,0.3456)	crown	Marc_R02,Marc_R19,Mass,Mthi_HT1,Mthi_R25,Mfas_1,Mfas_Mau,Mfus,Mmul_Chi,Mmul_Ind,Mnem_1,Mnem_2,Mnig,Mton
normal(0,7.3651,0.6116)	crown	Marc_R02,Marc_R19,Mass,Mthi_HT1,Mthi_R25,Mfas_1,Mfas_Mau,Mfus,Mmul_Chi,Mmul_Ind,Mnem_1,Mnem_2,Mnig,Mton,Panu

Supplementary Table S2. Sequencing and mapping statistics.

Sample ID	No. of reads	% mapped	No. duplicates	Coverage	Average insert size	Average read length
Marc_R02	1 003 208 808	0.9844	153 349 242	36.71	273.7	143
Marc_R19	736 772 746	0.9855	100 688 780	27.69	275.0	144
Mass	1 229 949 814	0.9793	20 061 338	43.04	286.4	118
Mfas_1	606 676 306	0.9915	134 920 438	20.60	283.9	143
Mfas_Mau	614 723 808	0.9825	17 277 032	17.80	467.2	98
Mfus	1 144 372 682	0.9674	104 949 180	28.38	430.1	92
Mmul_Chi	1 351 313 894	0.9834	6 627 086	37.47	163.2	92
Mmul_Ind	1 228 917 312	0.9901	15 723 200	34.71	208.1	93
Mnem_1	1 426 181 284	0.9486	16 775 230	38.81	324.0	94
Mnem_2	904 403 456	0.9808	56 510 430	36.73	354.5	144
Mnig	873 189 626	0.9800	58 687 344	35.34	356.3	144
Mthi_HT1	606 734 860	0.9825	67 441 268	23.05	260.2	142
Mthi_R25	747 766 878	0.9820	97 909 288	28.17	271.1	144
Mton	872 715 672	0.9812	53 681 256	35.53	354.1	144
Panu	885 913 086	0.9302	93 559 830	30.55	411.5	139

For individual samples, short ID, number of reads, mean insert size, average read length, and statistics after mapping to *M. mulatta* genome are shown.

Supplementary Table S3. Occurrence of repetitive elements in *M. mulatta* genome.

Family	Count	Length (bp)	Genome %
SINEs	1 783 119	412 631 037	12.75
LINEs	980 172	613 018 502	18.94
LTR	523 702	263 391 003	8.14
DNA	430 306	103 916 729	3.21
Unclassified	4 379	783 382	0.02
Small RNA	14 244	1 511 494	0.05
Satellites	178 741	239 774 799	7.41
Simple repeats	769 569	39 188 424	1.21
Low complexity	107 082	6 697 921	0.21

*Total length of *M. mulatta* genome: 3 236 224 332 bp (3 142 094 260 bp excluding N). Repeats are grouped by their respective repeat family. Table shows counts, combined lengths, and percentage of genome repetitive sequences.

Supplementary Table S4. Majority-rule consensus analysis of 26 633 individual

SNV-fragment trees. Only splits occurring >10% are shown. Samples in order: 1. Panu, 2.

Marc_R02, 3. Marc_R19, 4. Mass, 5. Mthi_HT1, 6. Mthi_R25, 7. Mnig, 8. Mton, 9. Mnem_1,

10. Mnem_2, 11. Mfas_Mau, 12. Mmul_Chi, 13. Mfus, 14. Mfas_1, 15. Mmul_Ind.

Sets included in consensus tree	Count	Frequency
.***** *****	26 633	1.0000
.....**	26 086	0.9795
.....****	25 730	0.9661
.....**	25 055	0.9408
.**	23 415	0.8792
.*****	22 904	0.8600
.....**.*	20 730	0.7784
.....*****	19 274	0.7237
...***	18 577	0.6975
....**	17 490	0.6567
.....*.*	14 247	0.5349
.*****	12 771	0.4795
.....****	9 984	0.3749
Sets not included in consensus tree		
.....*.*	9 959	0.3739
.....*****	7 861	0.2952
.....**	6 210	0.2332
.....*.*	4 962	0.1863
...*.*	4 901	0.1840
.....***.*	3 700	0.1389
...**	2 928	0.1099

Table summarizes results from CONSENSE analysis. Ranking is according to number of occurrences of splits. Only splits occurring more frequently than 10% are shown. In each vertical, column dot (.) and asterisk (*) represent one individual and its split into respective group (. or *). For example: row two (.....**.....) has species 7 and 8 (samples of Mnig and Mton) as most frequent split against all others, splitting from others with 26 086 occurrences.

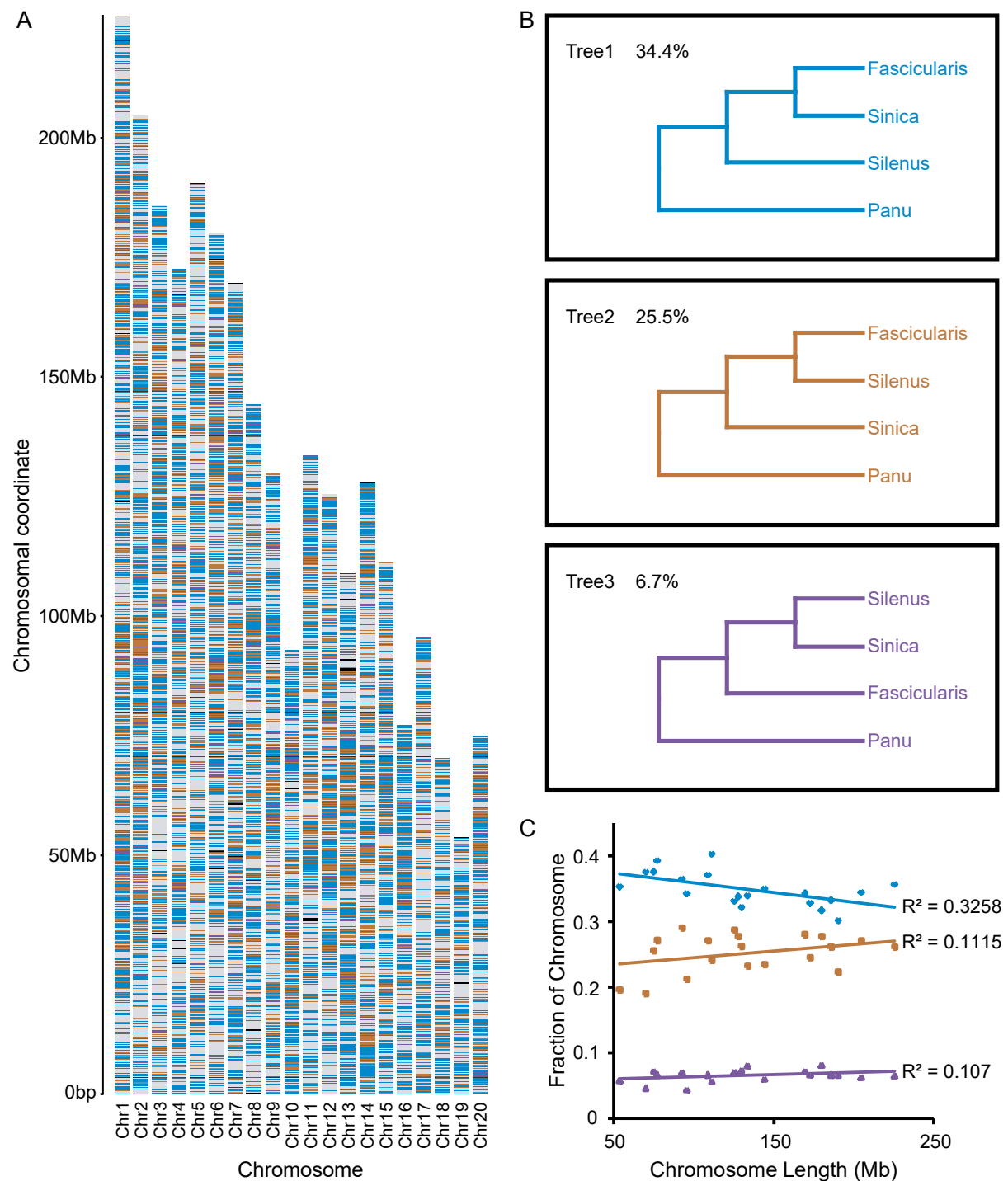
Supplementary Table S5. *D*-statistic results at individual level. Outgroup is Panu. Taxa with significant introgression signals are highlighted in yellow. In each row, the adjusted *P* value less than 0.001 indicates that the *D* value is significantly greater than 0, which indicates that gene flow exists between P2 and P3.

Supplementary Table S6. *D*_{FOIL}-statistic results. Outgroup is Panu. Taxa with significant introgression signals are highlighted in yellow. The directions of gene flows are shown in the "introgression" column. For example, "12→3" in the "introgression" column of row 290 indicates that the direction of gene flow is from the ancestral lineage of Mthi_HT1 (sp1) and Mthi_R25 (sp2) to Mnem_2 (sp3).

Supplementary Table S7. Results of KEEG and GO-based enrichment analysis of genes located in the top 5% of windows of *M. fascicularis*-*M. nemestrina* species pair. Only genes with known names were used in the enrichment analysis. We used adjusted *P* value of 0.05 as significance threshold.

Supplementary Tables S5-S7 are listed as separate files due to their large size.

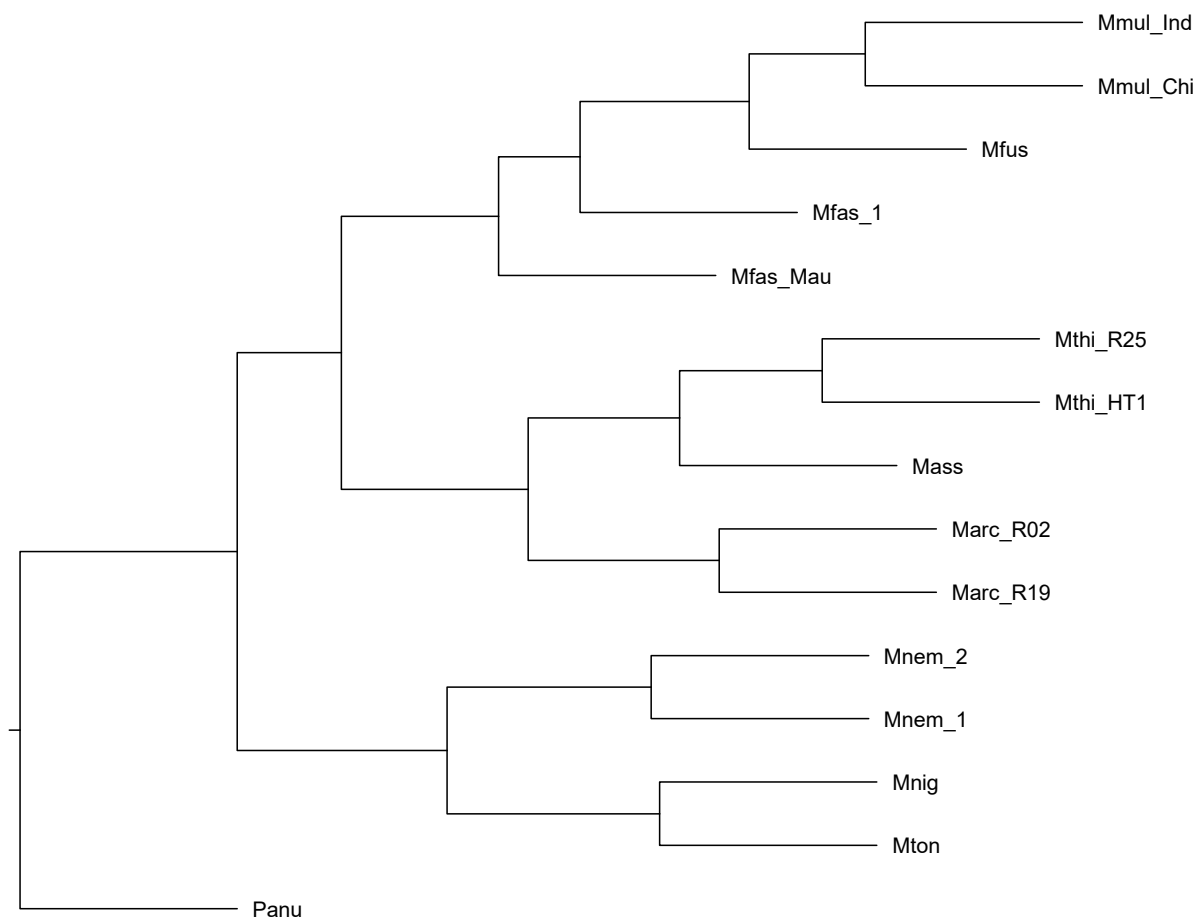
Supplementary Figures



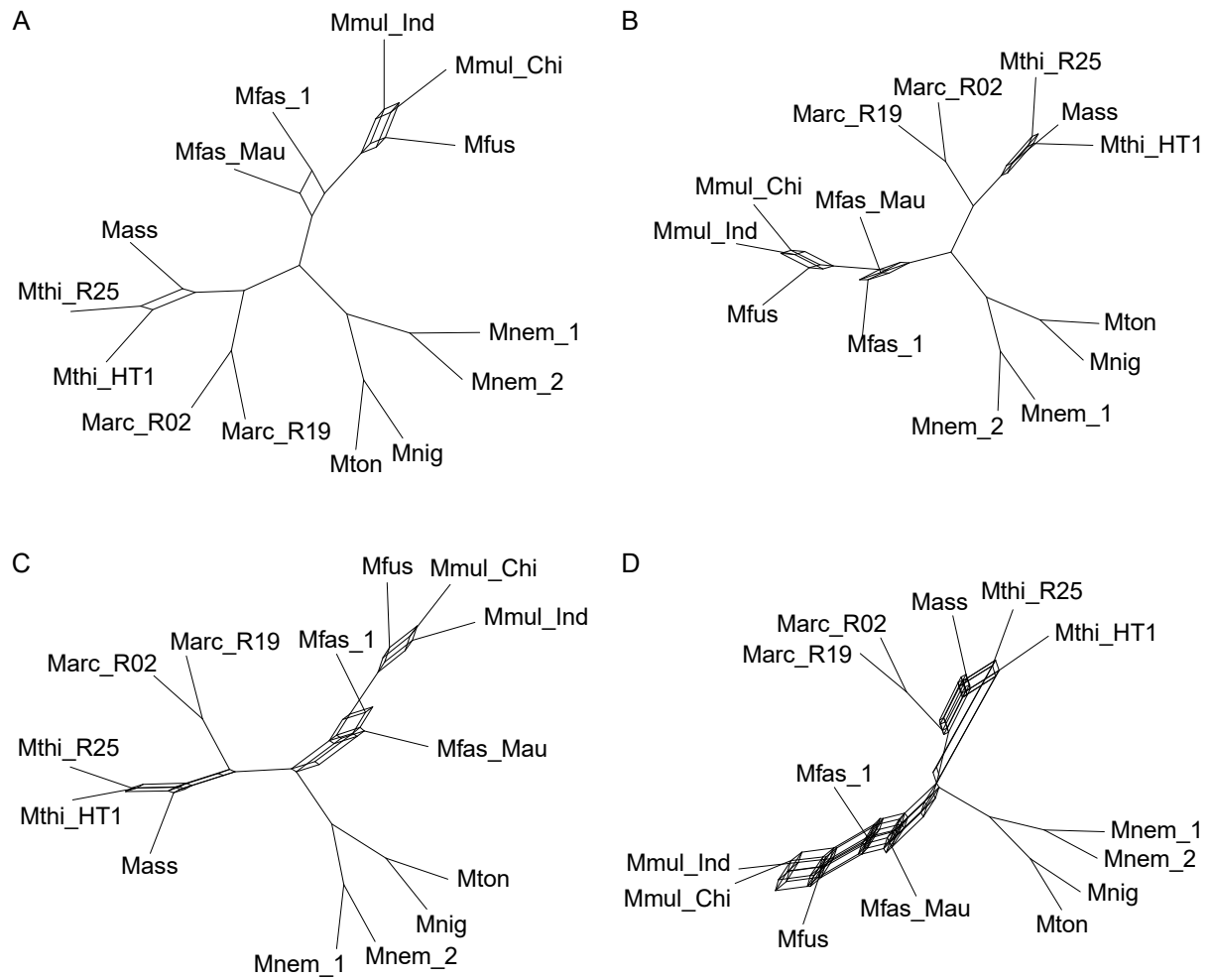
Supplementary Figure S1. Local evolutionary history of macaques at species group level.

(A) Each bar represents a chromosome, in terms of *M. mulatta* genome. Colored bands represent tree topologies of each 100 kb window. Colors correspond to topologies in (B). (B) Three most common trees. Values at top left corner are percentage of all 100 kb windows that recover that topology. Colors correspond to colored bands in (A), with gray and black regions showing other topologies and missing data. (C) Relationship of tree frequency and

chromosome size. Colors correspond to topologies in (B). Tree1 shows a negative relationship with chromosome size, whereas tree2 shows a positive relationship.

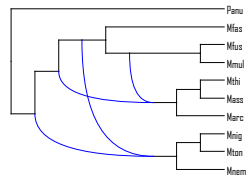


Supplementary Figure S2. Majority-rule consensus tree from 26 633 SNV-fragment trees calculated with CONSENSE in PHYLIP package. Topology is congruent with coalescent species tree generated by ASTRAL and MP-EST.

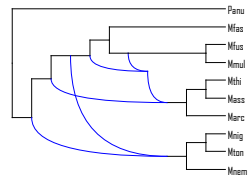


Supplementary Figure S3. Consensus networks of macaques from 26 633 SNV-fragment trees at different gene tree thresholds to form an edge. (A) 14% threshold. (B) 11% threshold. (C) 8% threshold. (D) 5% threshold.

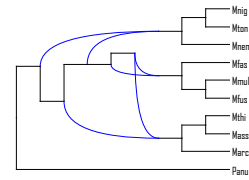
[1]Tree1



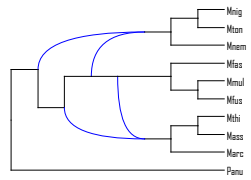
[2]Tree2



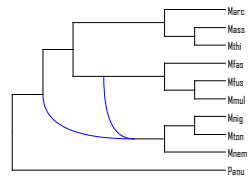
[3]Tree3



[4]Tree4

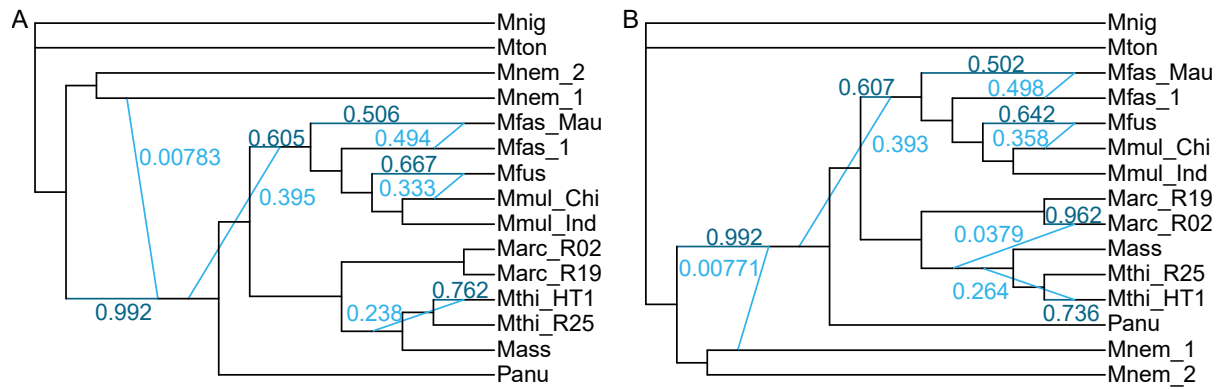


[5]Tree5

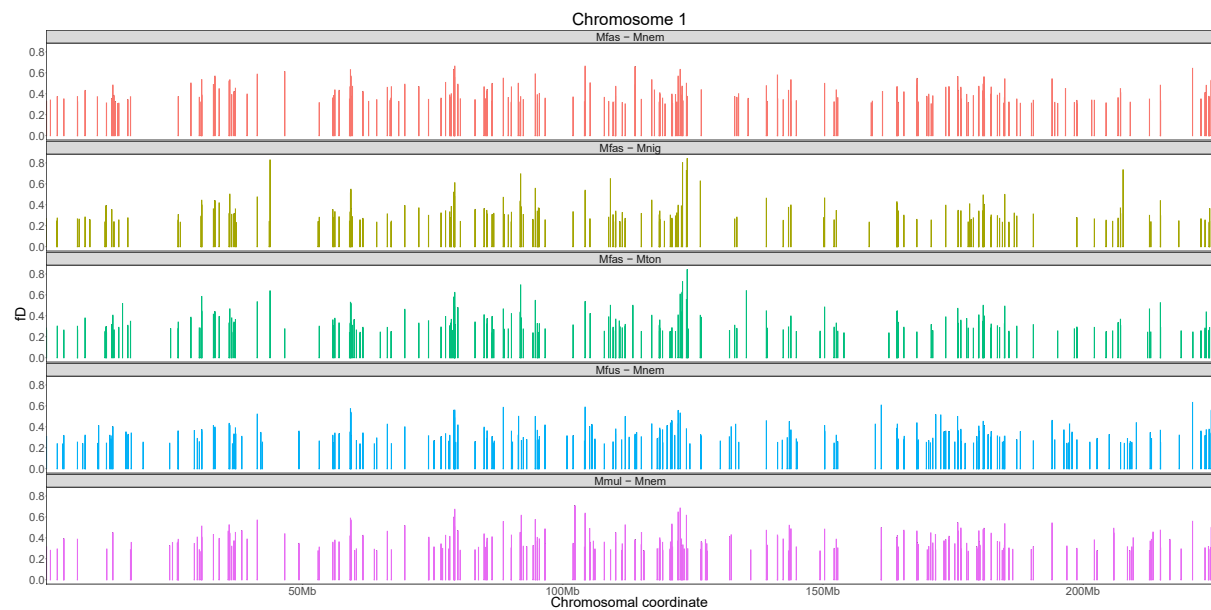


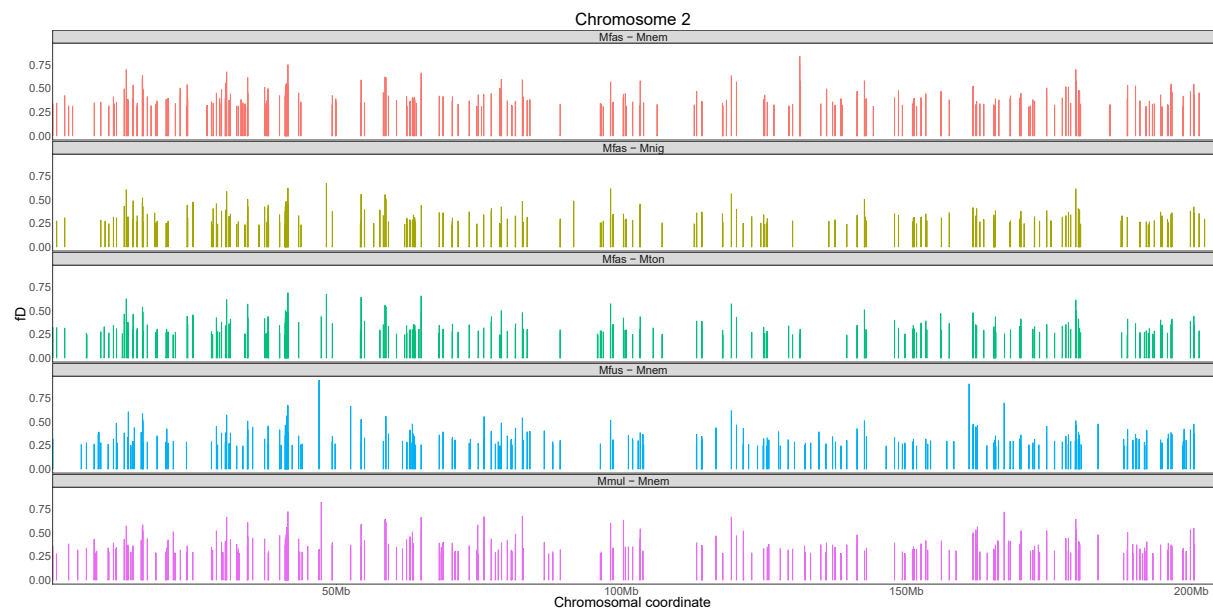
Supplementary Figure S4. Five best phylogenetic networks calculated with PhyloNet at species level.

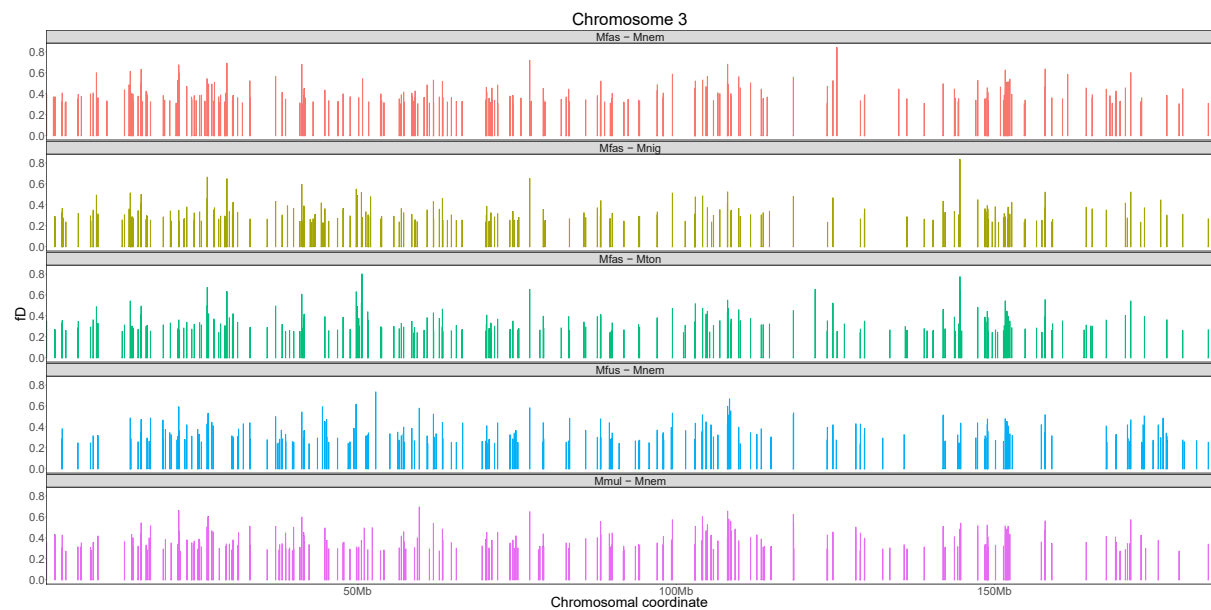
Supplementary Figure S5. Five best phylogenetic networks calculated with PhyloNet at individual level.

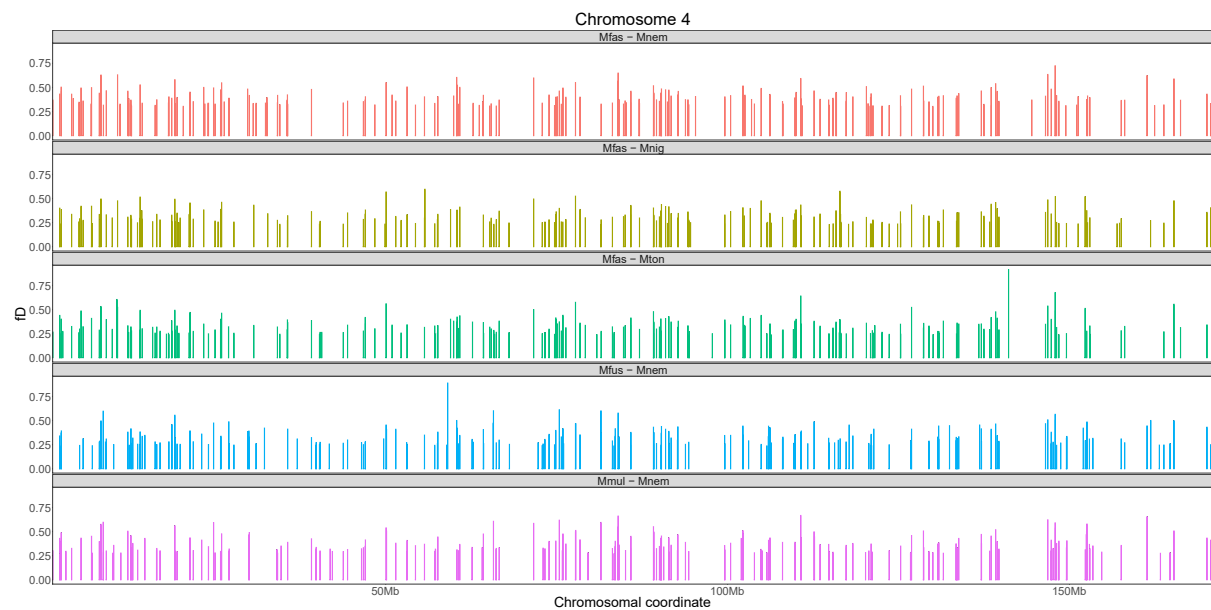


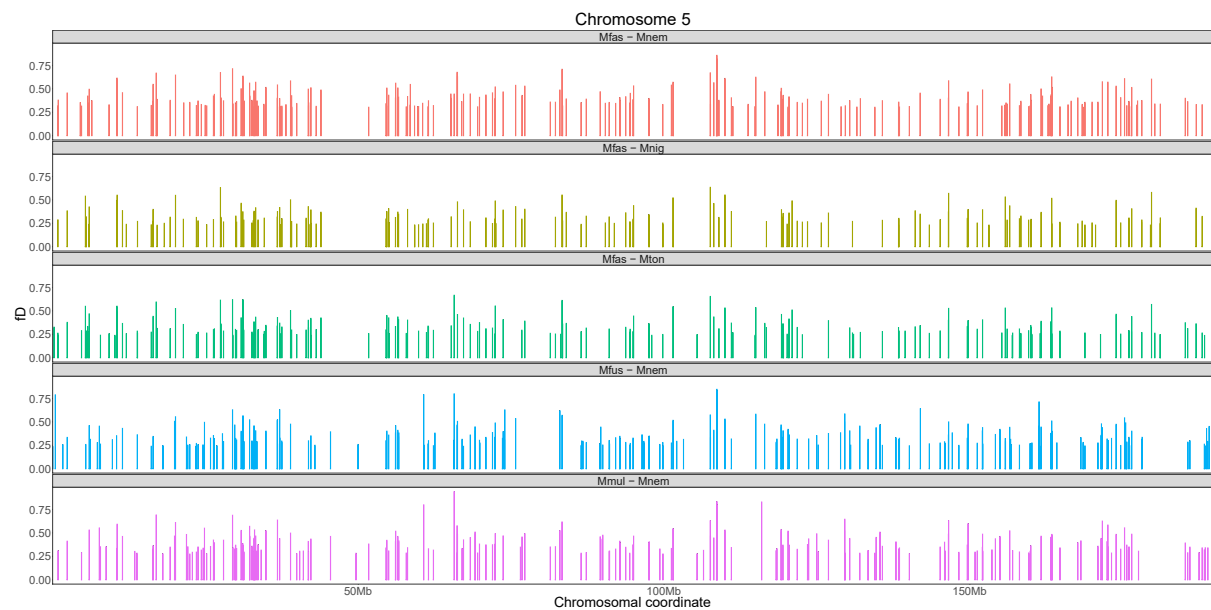
Supplementary Figure S6. Networks of macaques with (A) 5 and (B) 6 reticulations inferred by PhyloNetworks based on 26 633 SNV-fragments. Hybrid edges are annotated with their inheritance values, which measure proportion of genes inherited via gene flow.

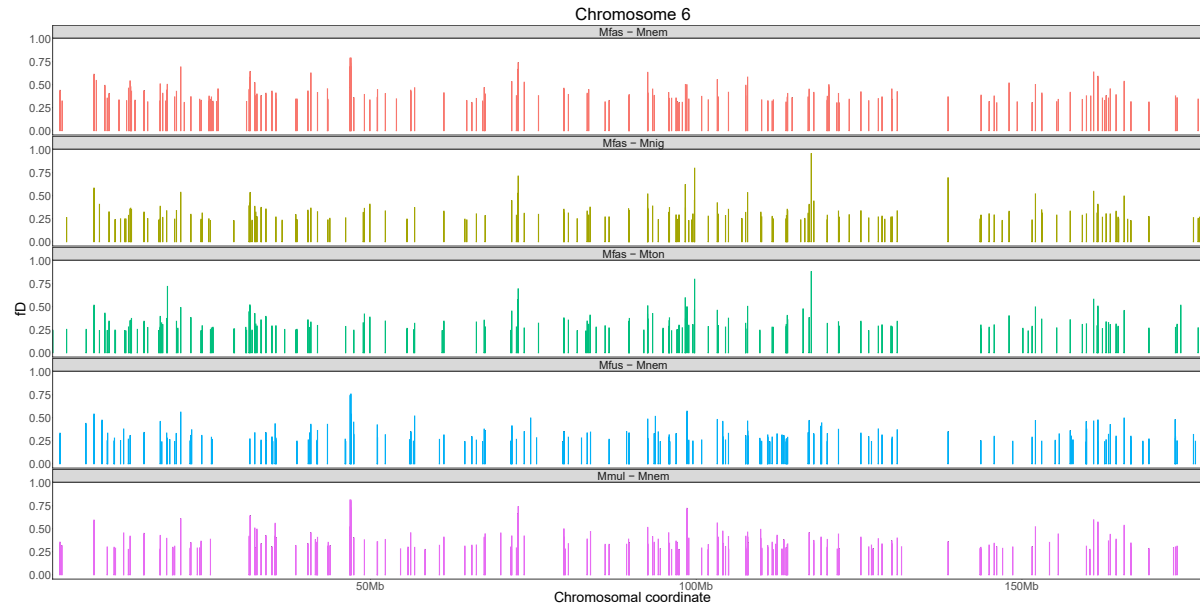


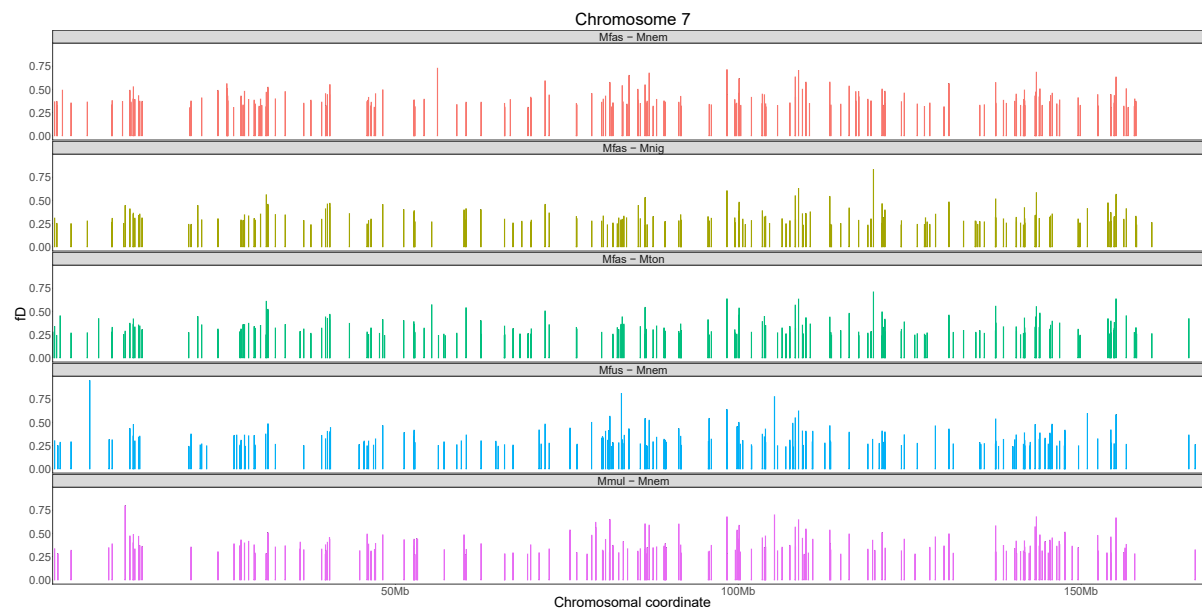


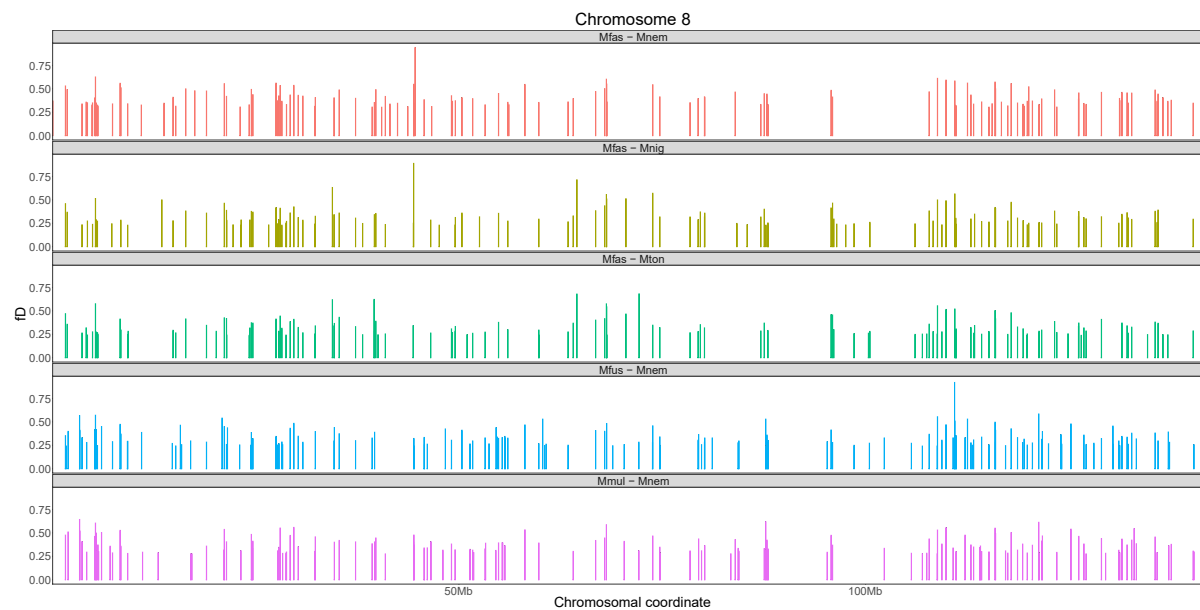


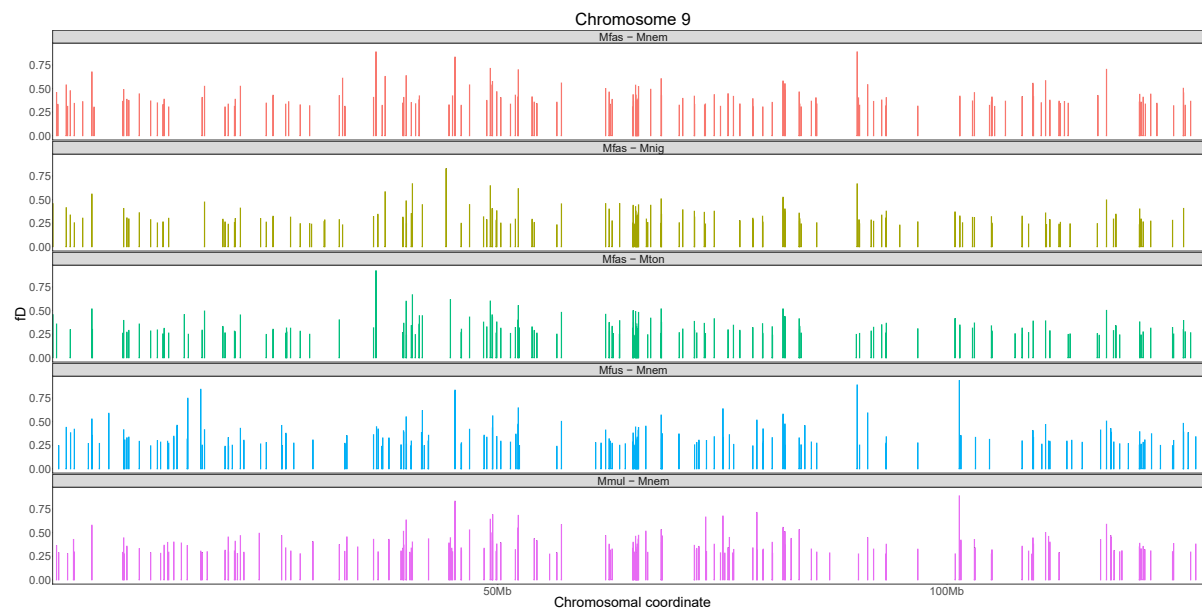


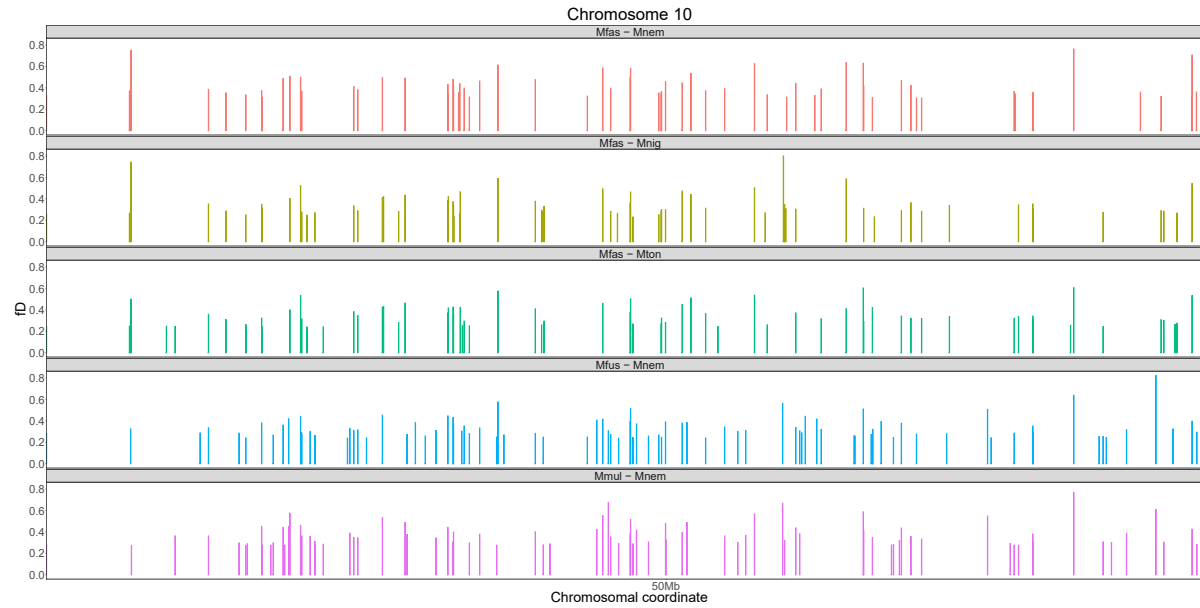


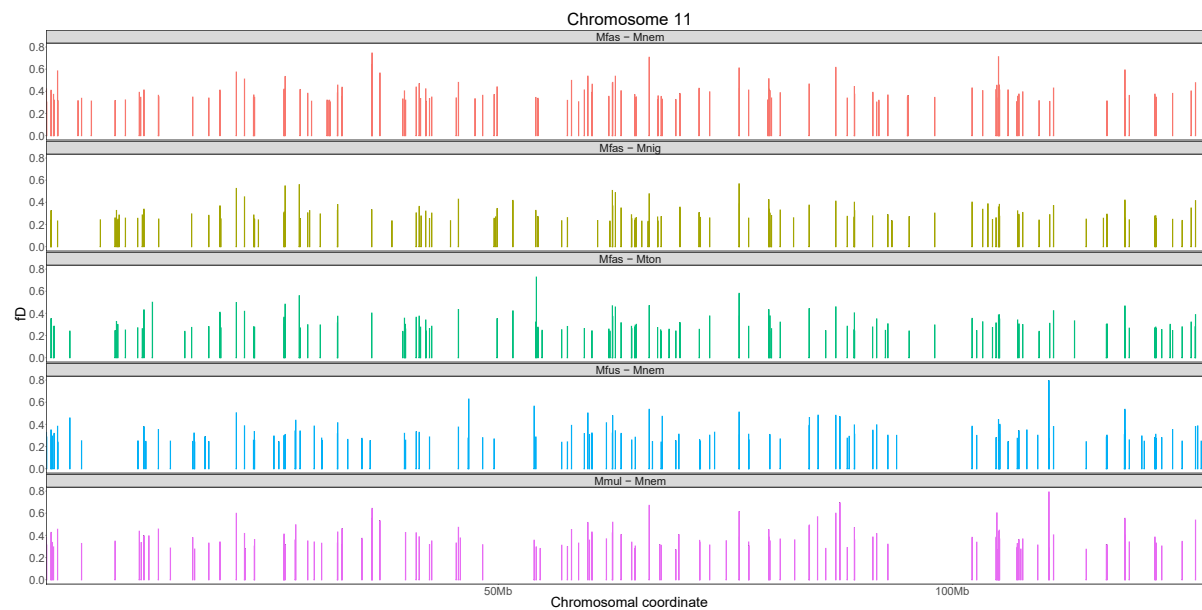


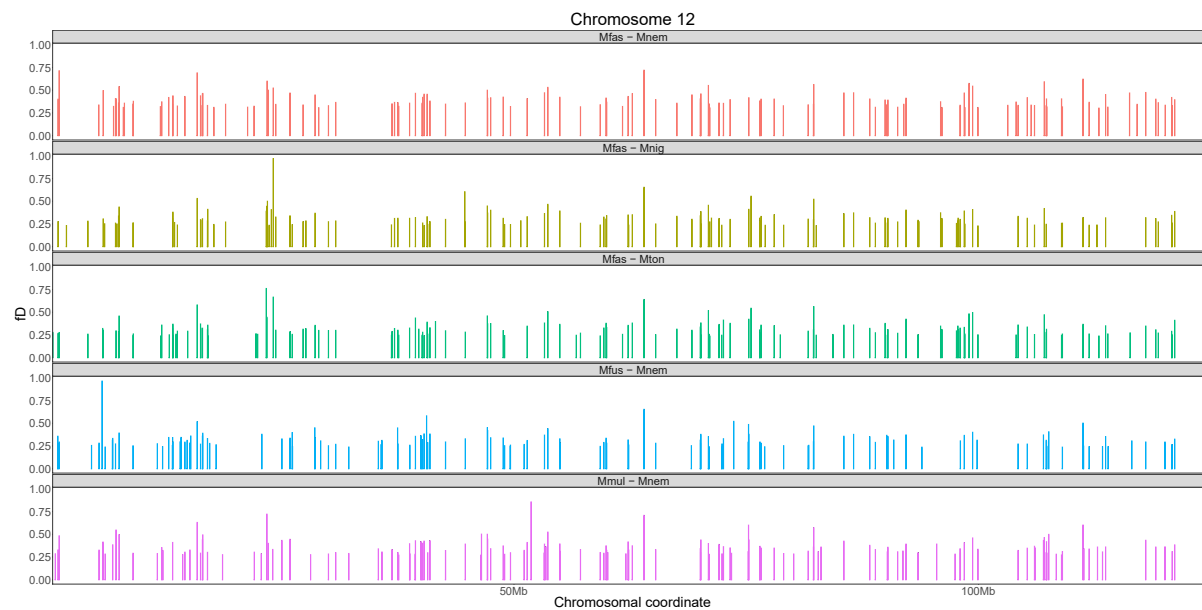


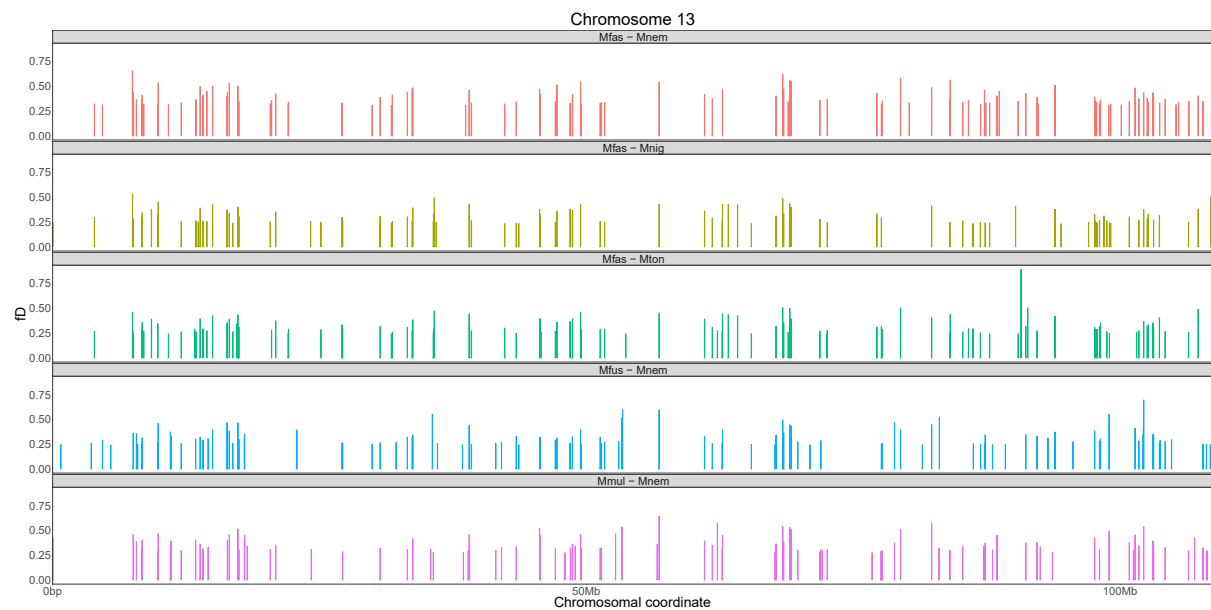


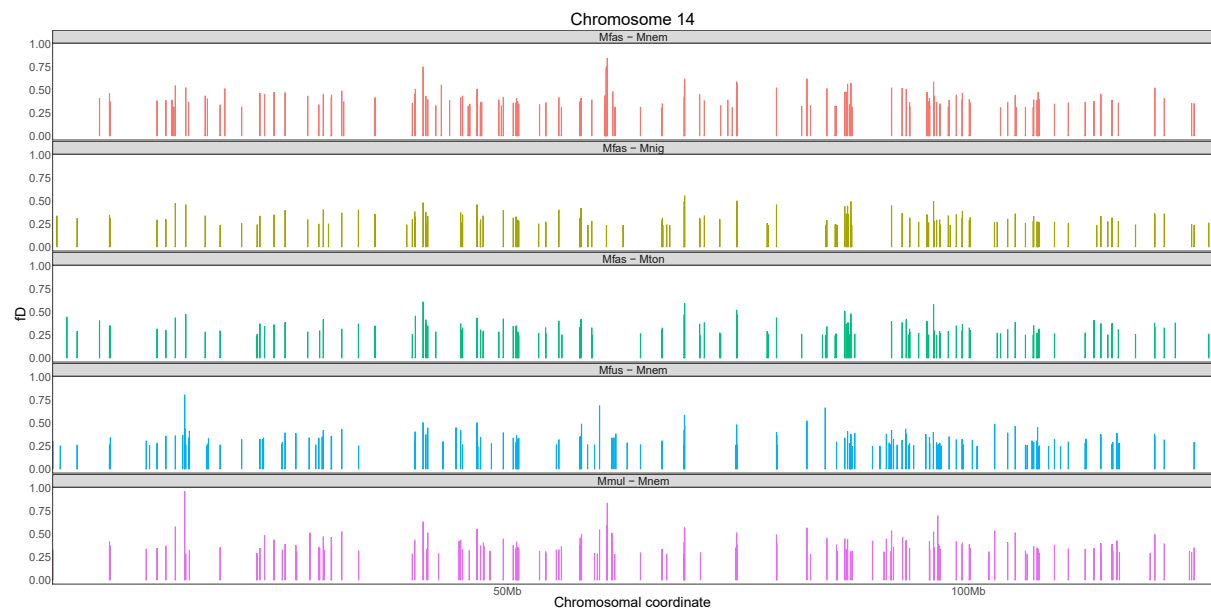


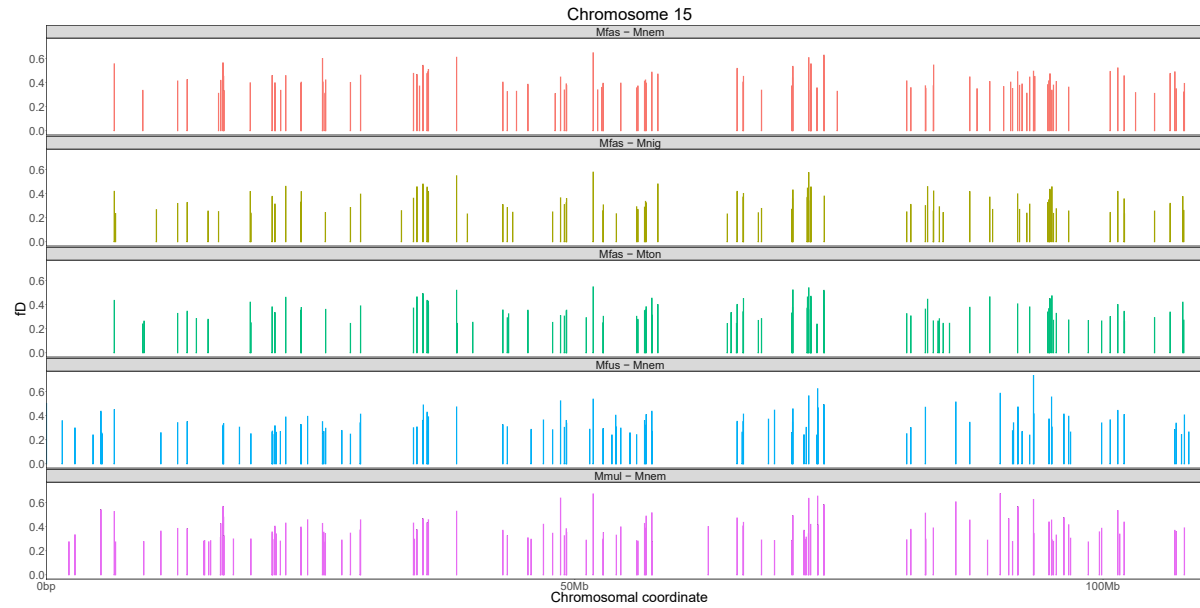


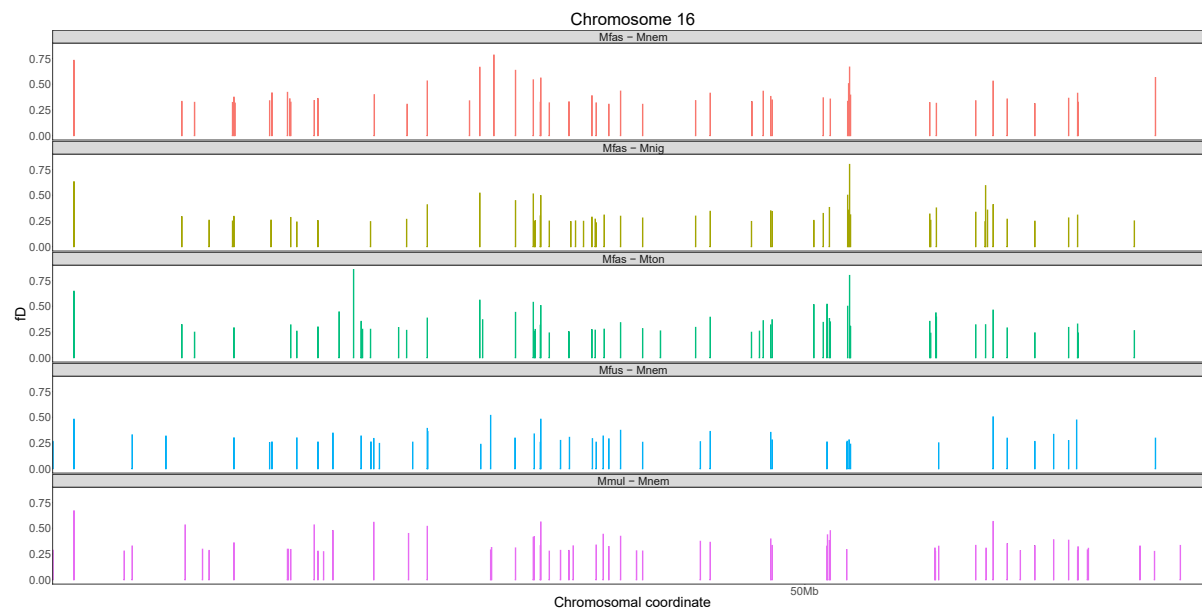


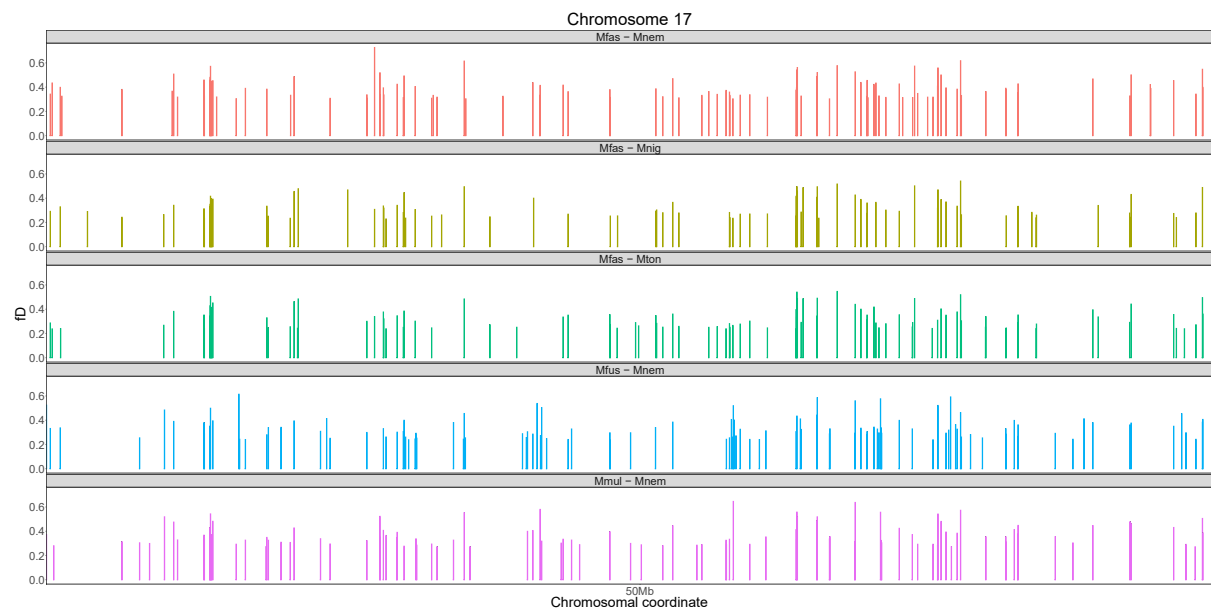


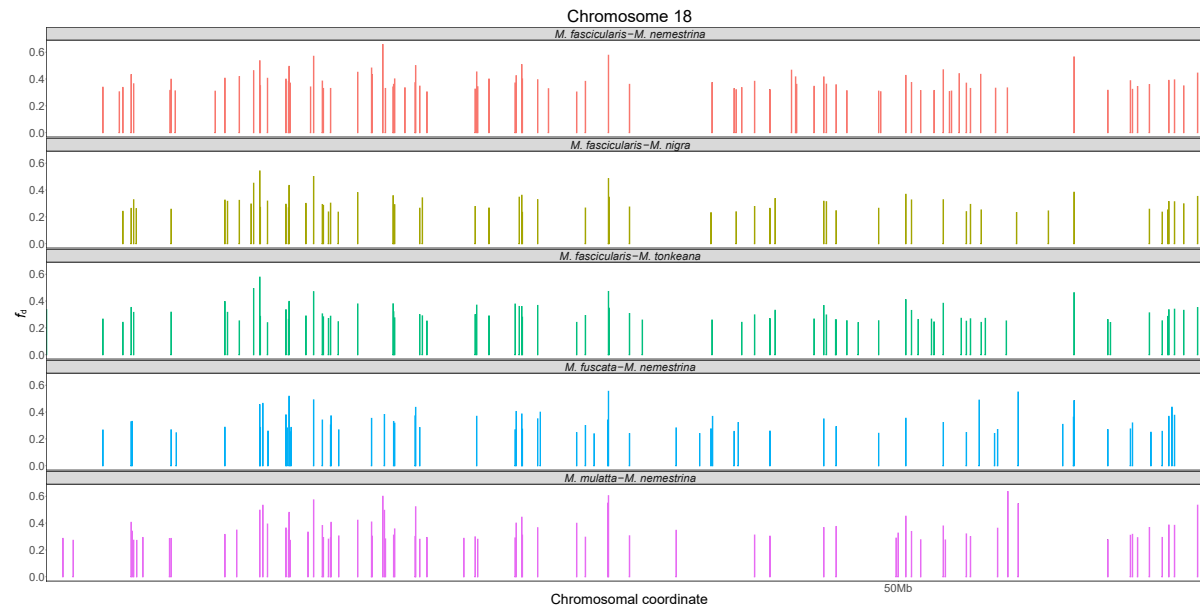


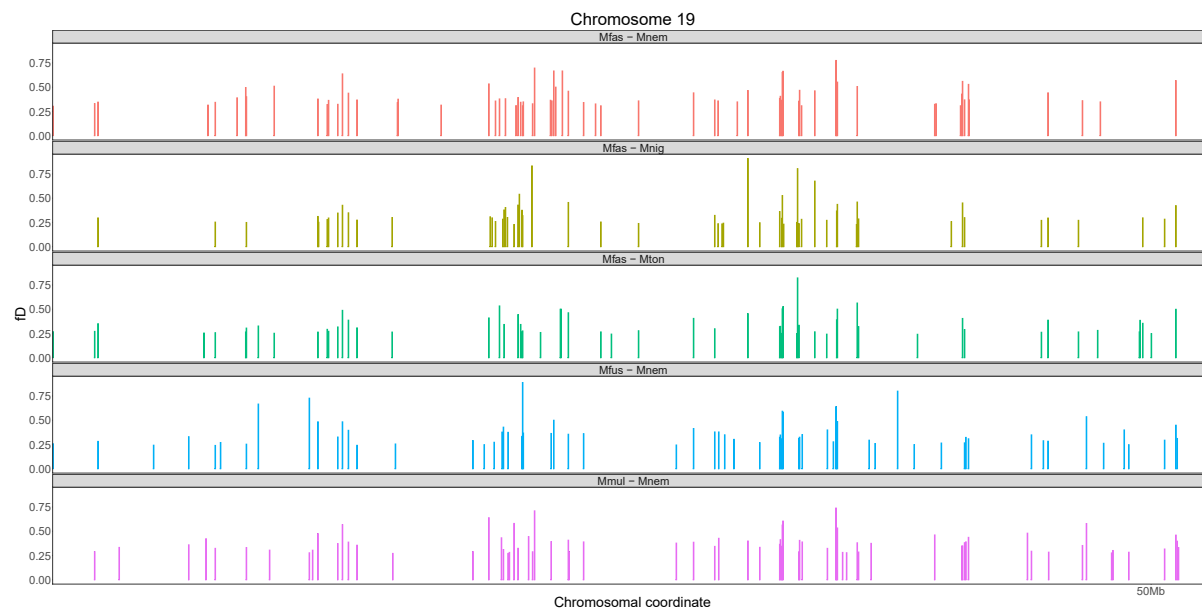


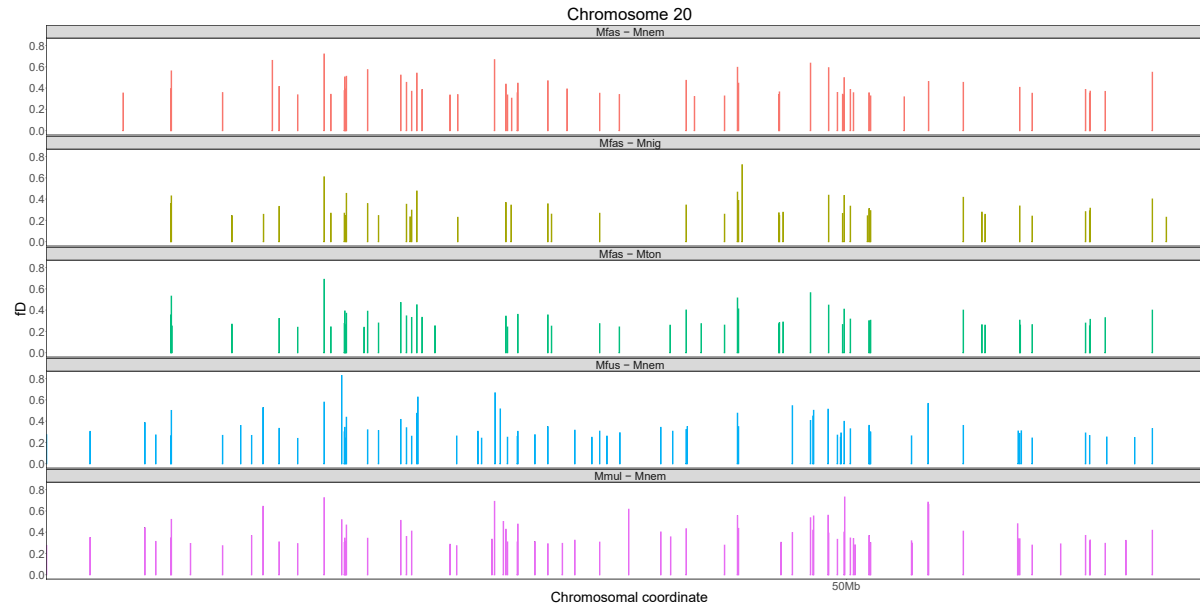




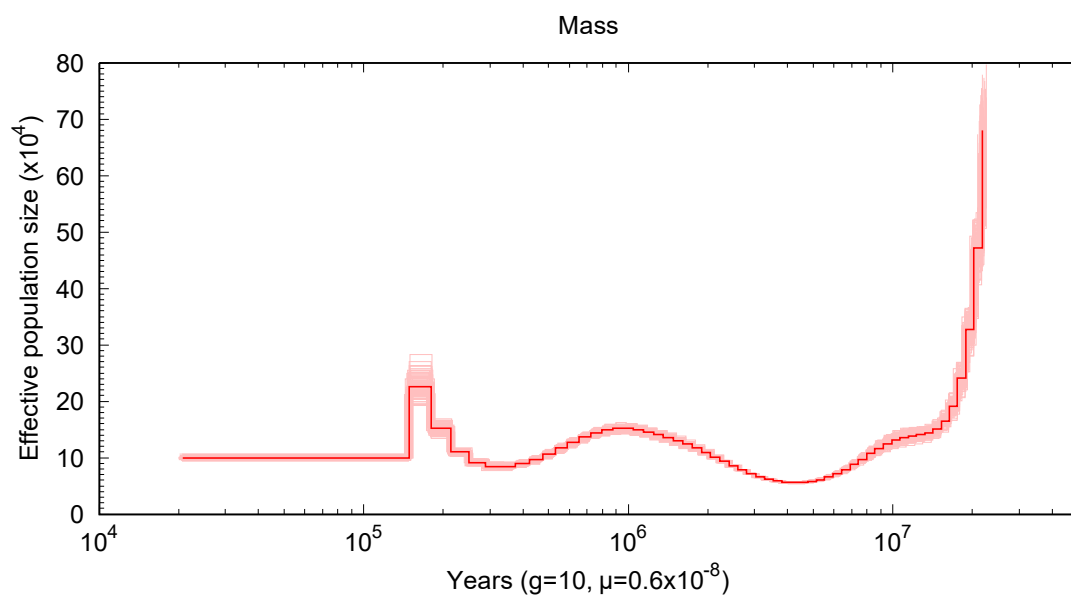
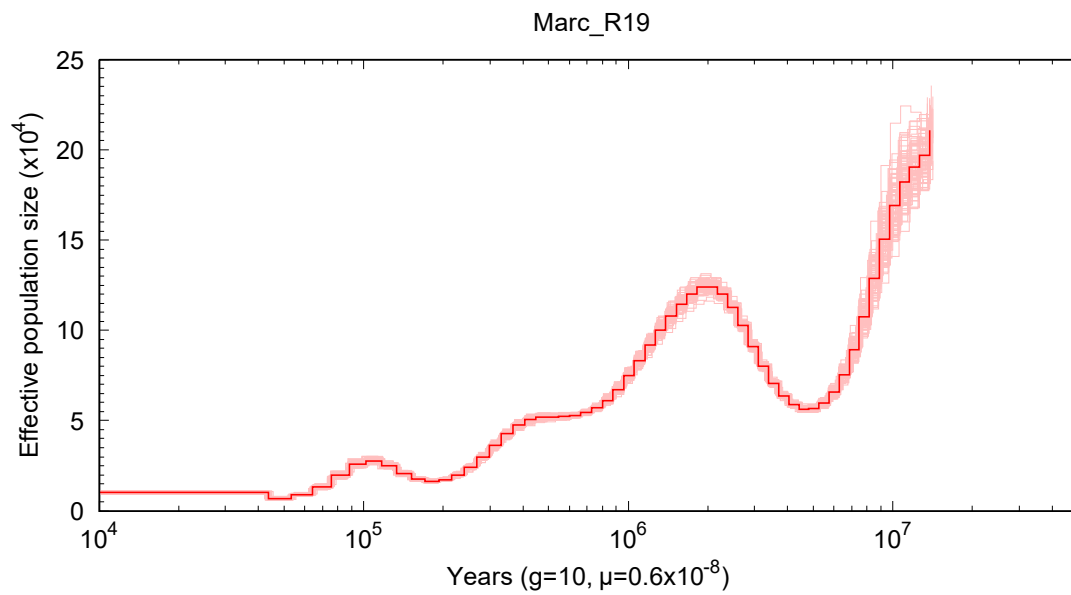
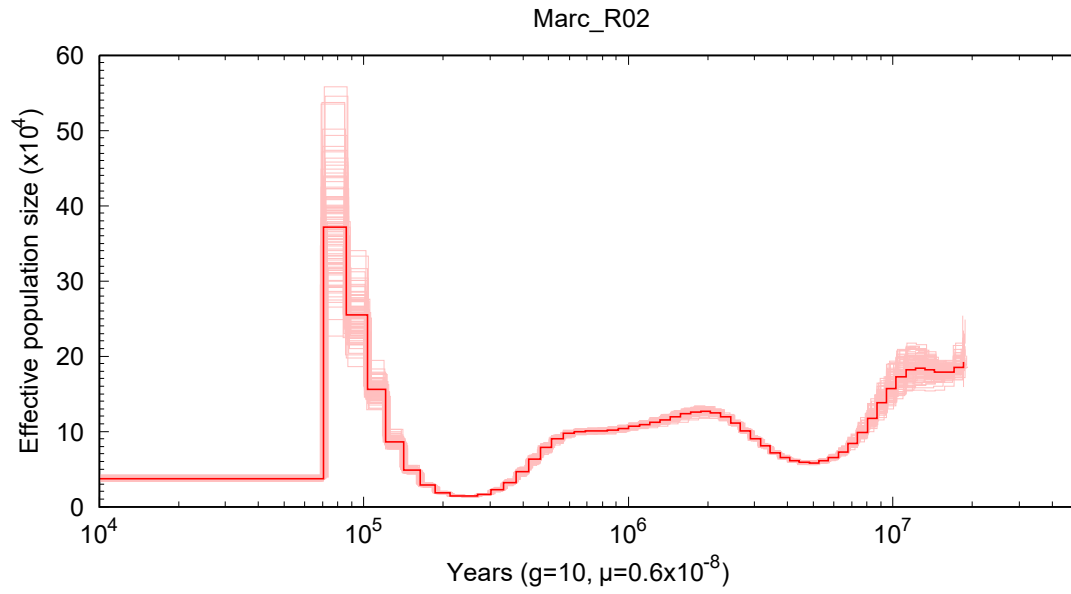


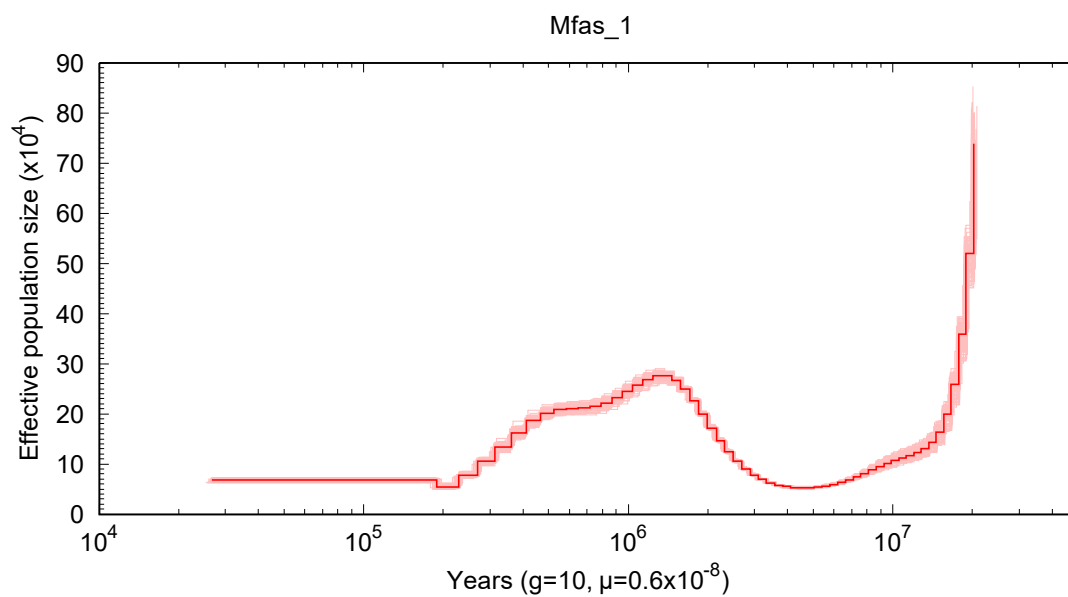
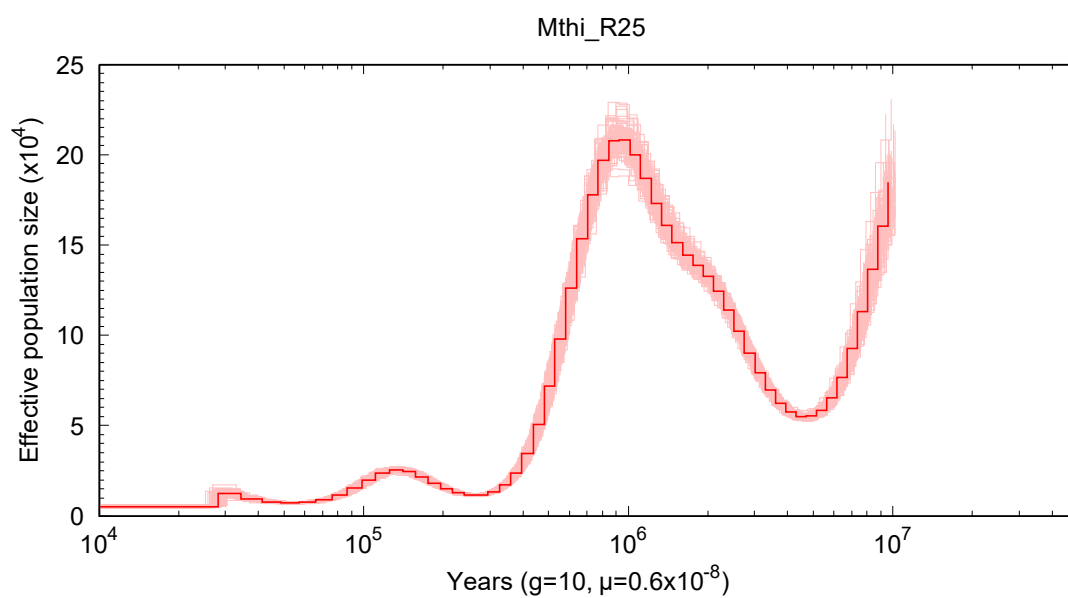
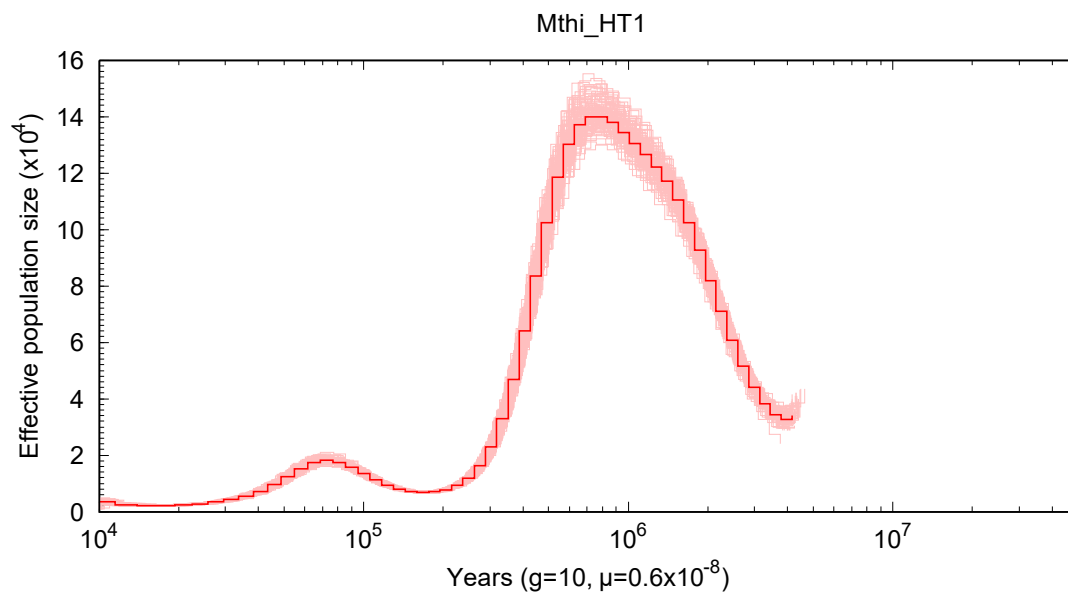


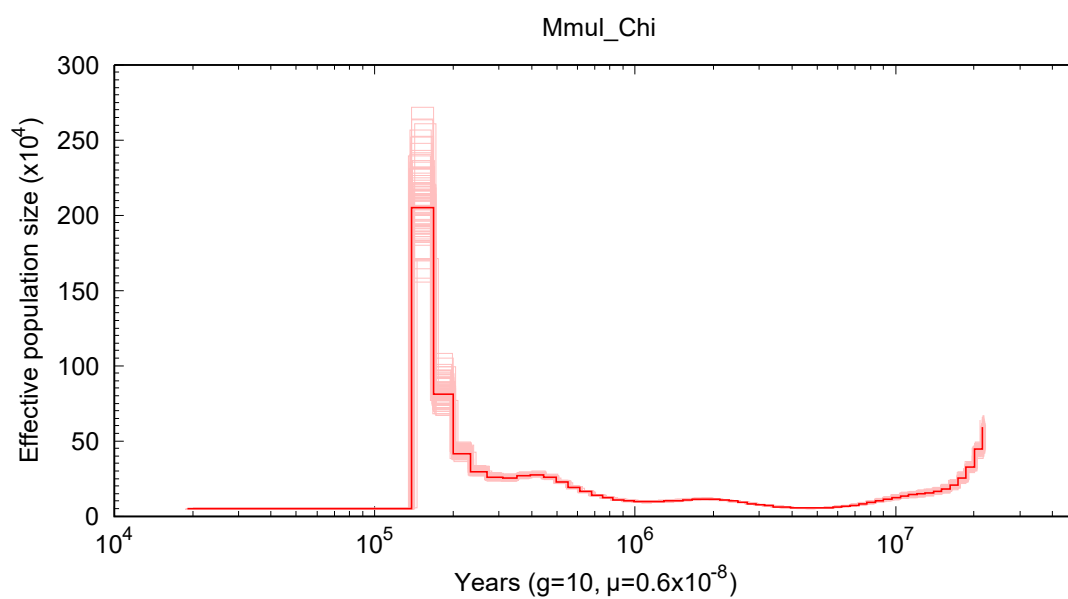
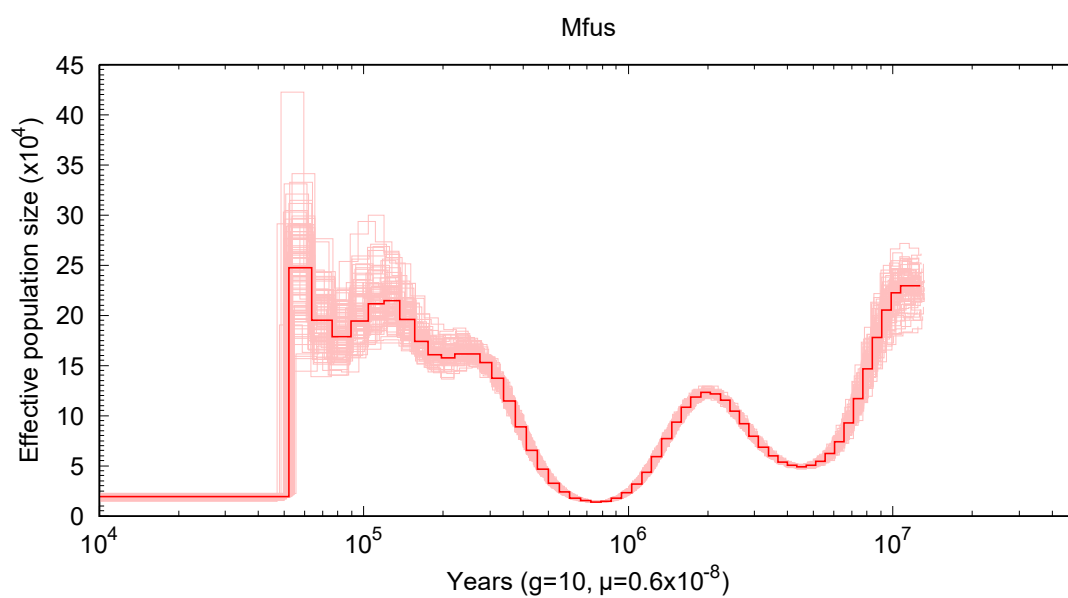
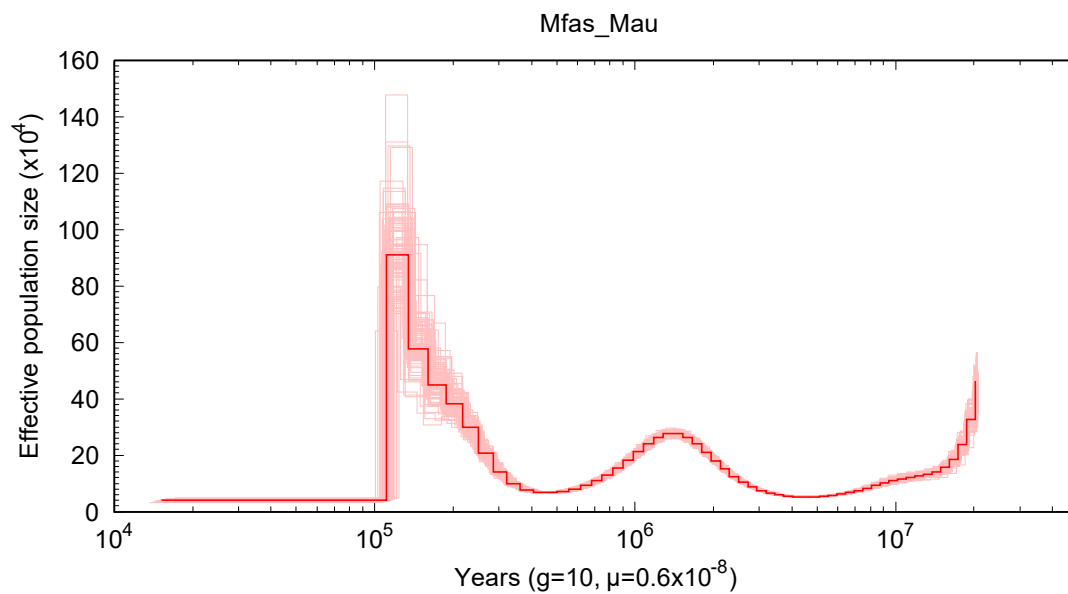


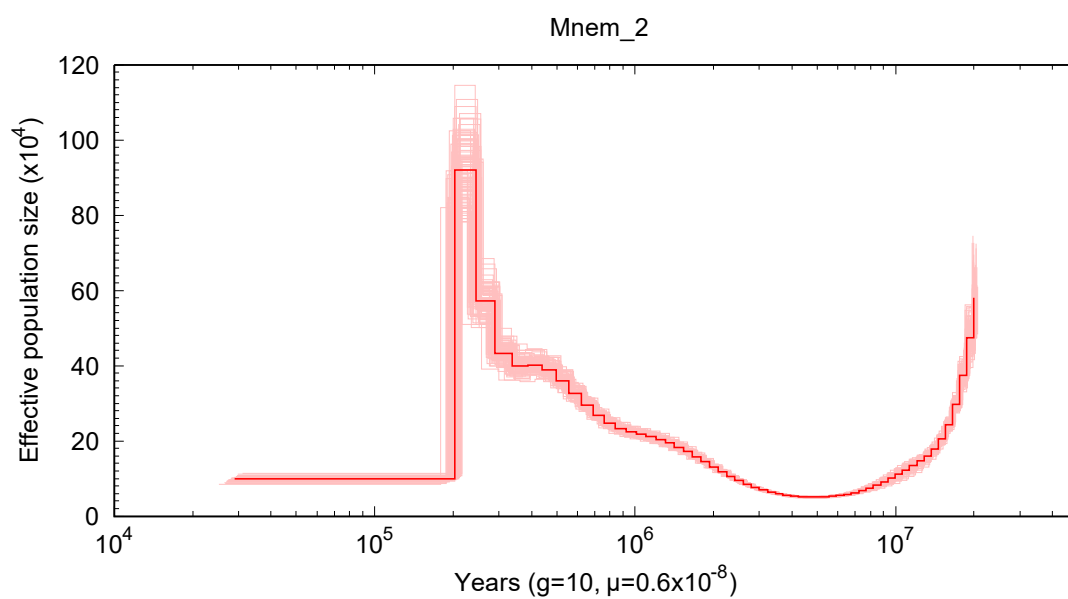
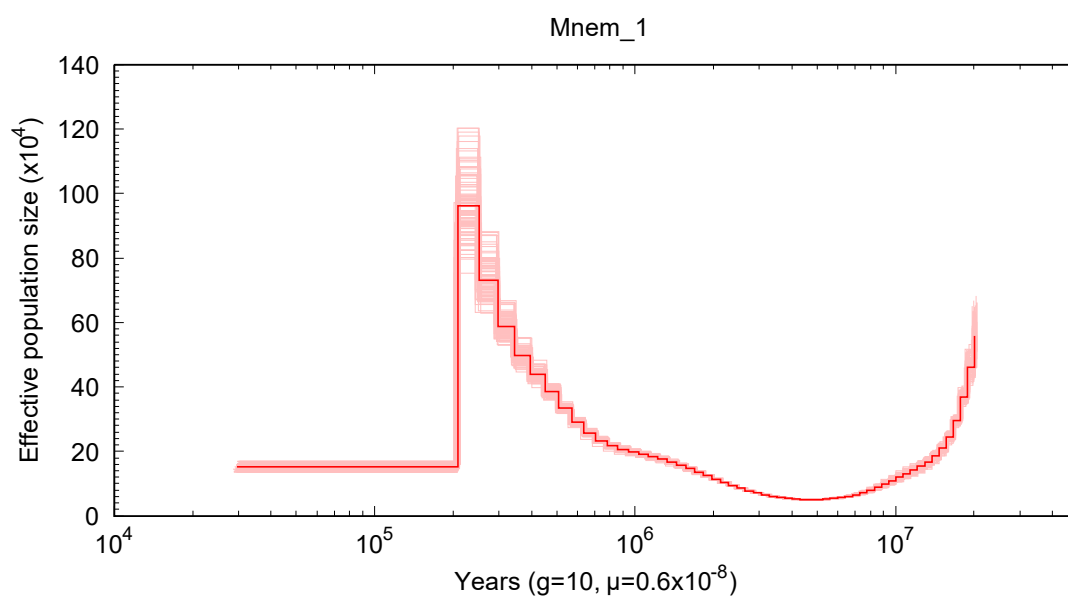
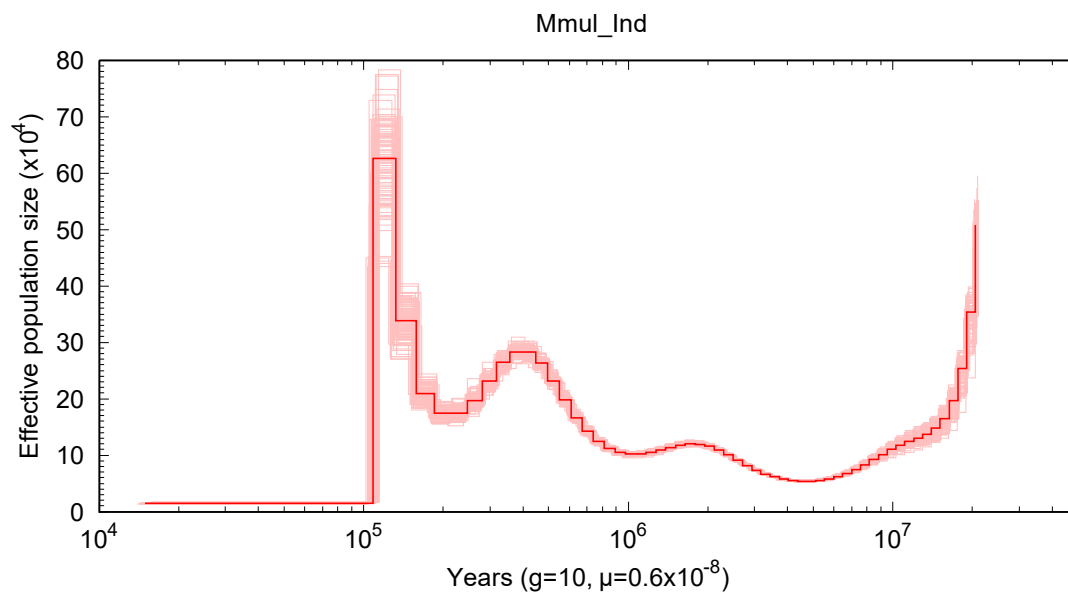


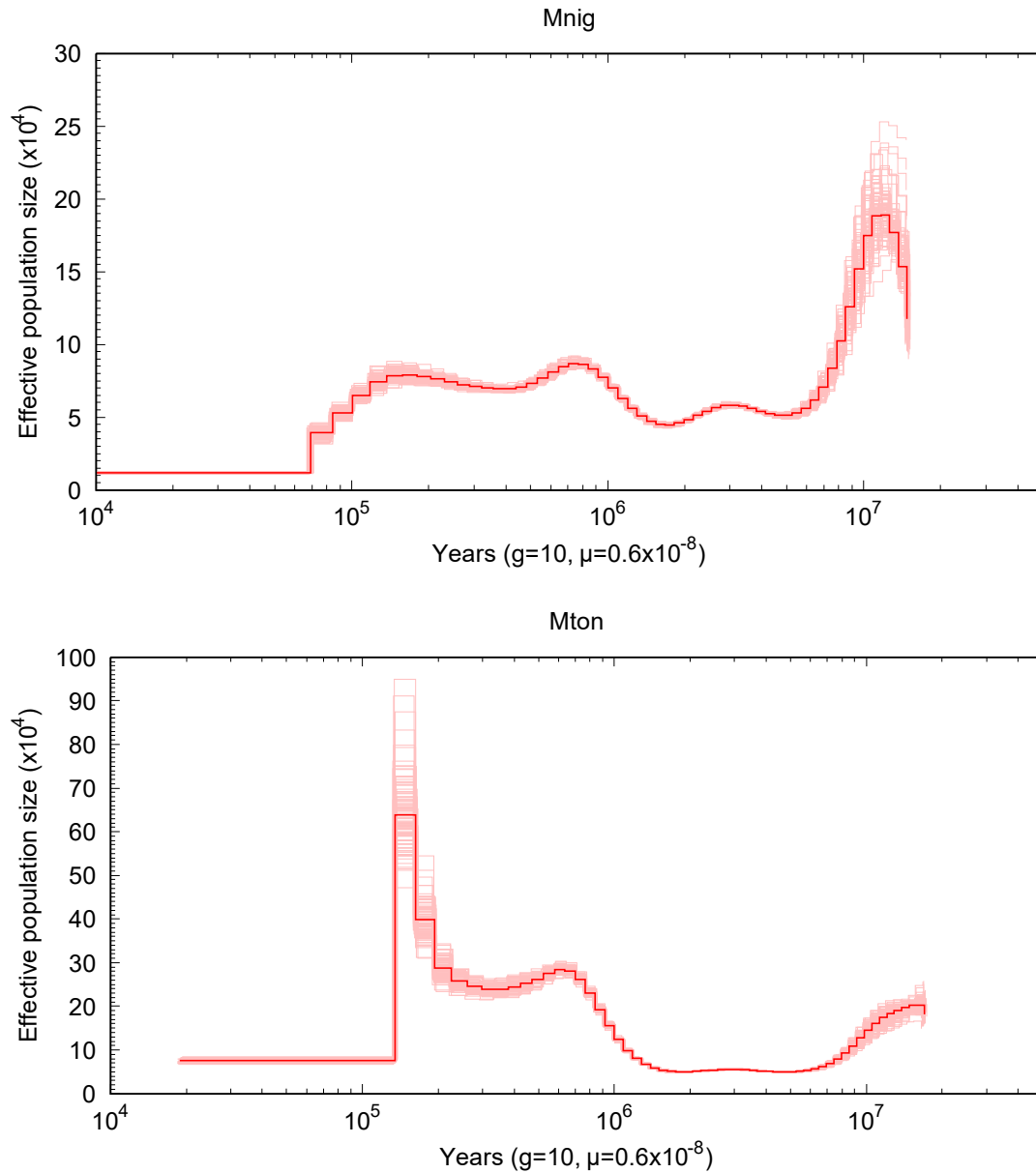
Supplementary Figure S7. Distribution of top 5% of windows with strongest introgression signals according to f_d across chromosomes. Species of species pairs belong to species groups *fascicularis* and *silenus*. Results show that top 5% of windows with strongest introgression signals for different species pairs had similar distributions across genome.











Supplementary Figure S8. Demographic history of each individual macaque genome with 100 bootstrap replicates. Each panel shows estimated ancestral effective population size calculated by PSMC. Bootstrap replicates are shown in light red. Sample names are given above each panel. g , generation time. μ , mutation rate.