PALEOMIX ZONKEY V1.2.13 – DB REV. 20161101

A PIPELINE FOR DETECTION OF F1 HYBRIDS IN EQUIDS.

INTRODUCTION

The Zonkey Pipeline is a easy-to-use pipeline designed for the analyses of low-coverage, ancient DNA derived from historical equid samples, with the purpose of determining the species of the sample, as well as determining possible hybridization between horses, zebras, and asses. This is accomplished by comparing one or more samples aligned against the *Equus caballus* 2.0 reference sequence with a reference panel of modern equids, including wild and domesticated equids.

For more information, please refer to the the documentation for the Zonkey pipeline or the documentation for the PALEOMIX pipeline, on which the Zonkey pipeline is based.

ANALYSIS OVERVIEW

Zonkey run using database rev. 20161101. Data processed using pysam v0.14.1, SAMTools v1.7.0 [*Li et al. 2009*] and PLINK v1.7 [*Purcell et al. 2007*]; plotting was carried out using R v3.6.0. Additional tools listed below.
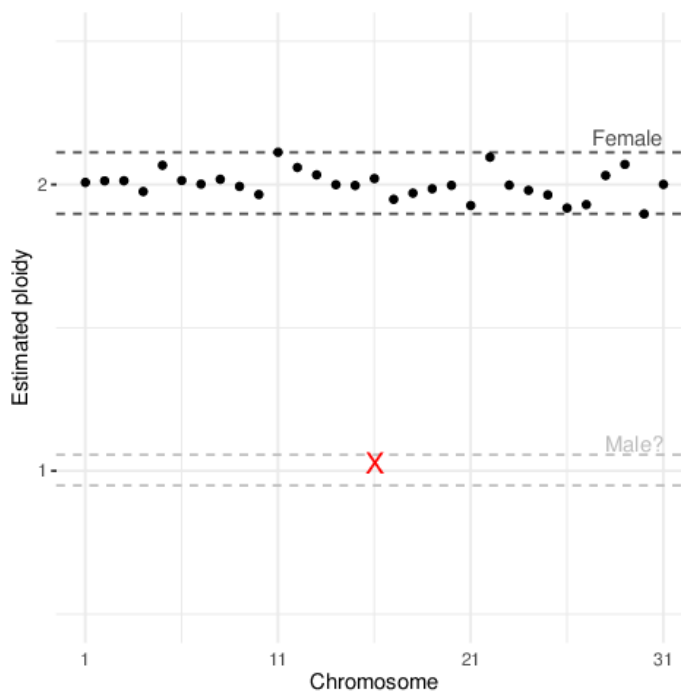
Nuclear report from '*/disk/franklin/data/ludovic/MINISEQ /220805_Gaetan_182_80PE/Mapping/EquCab2 /Mon2017x150_Ages_1_21912.Horse_nuc_wY.realigned.q25.bam*'

Number of reads processed:	281631
Number of reads overlapping SNPs:	167488
Number of SNPs used (incl. transitions):	167001
Number of SNPs used (excl. transitions):	51836

Mitochondrial report from '*/disk/franklin/data/ludovic/MINISEQ /220805_Gaetan_182_80PE/Mapping/EquCab2 /Mon2017x150_Ages_1_21912.Horse_mt.realigned.q25.bam*'

Reference sequence used:	5835107Eq_mito3
Reference sequence length:	16659
Number of sites covered:	11392
Percentage of sites covered:	68.4
Mean coverage per site:	1.2

Autosomes vs. sex-chromosomes:

REFERENCE PANEL

Group(2)	Group(3)	ID	Species	Sex	Sample Name	Publication
Caballine	Horse	HCab	*E. caballus*	Male	FM1798	doi:**10.1016/j.cub.2015.08.032**
		HPrz	*E. przewalskii*	Male	SB281	doi:**10.1016/j.cub.2015.08.032**
NonCaballine	Ass	AAsi	*E. a. asinus*	Male	Willy	doi:**10.1038/nature12323**
		AKia	*E. kiang*	Female	KIA	doi:**10.1073/pnas.1412627111**

	AOna	*E. h. onager*	Male	ONA	doi:**10.1073/pnas.1412627111**
	ASom	*E. a. somaliensis*	Female	SOM	doi:**10.1073/pnas.1412627111**
Zebra	ZBoe	*E. q. boehmi*	Female	BOE	doi:**10.1073/pnas.1412627111**
	ZGre	*E. grevyi*	Female	GRE	doi:**10.1073/pnas.1412627111**
	ZHar	*E. z. hartmannae*	Female	HAR	doi:**10.1073/pnas.1412627111**

ADMIXTURE ESTIMATES

Admixture proportions estimated using **ADMIXTURE** v1.3 *[Alexander et al. 2009]*, using default parameters.

2 ancestral groups
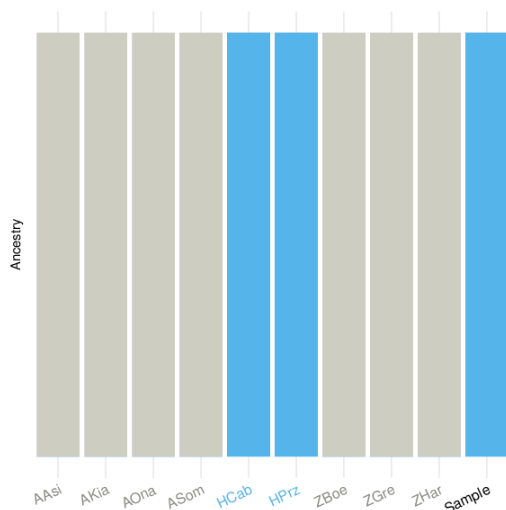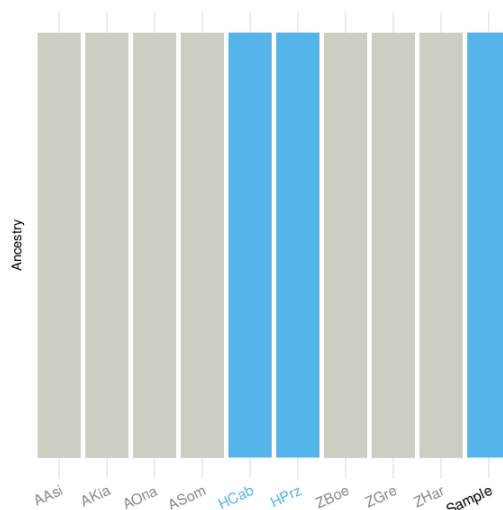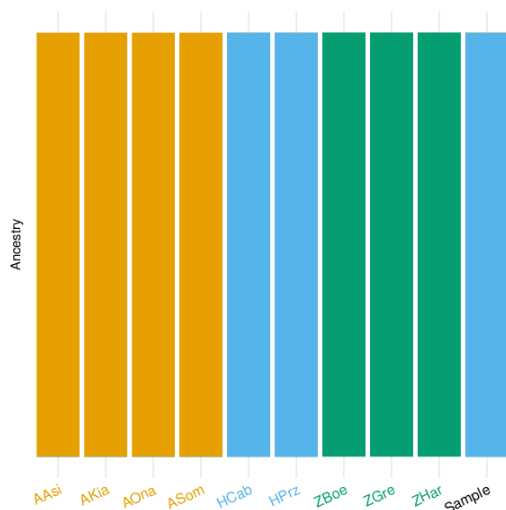
No admixture detected.

2 ancestral groups, excluding transitions

No admixture detected.

3 ancestral groups

No admixture detected.

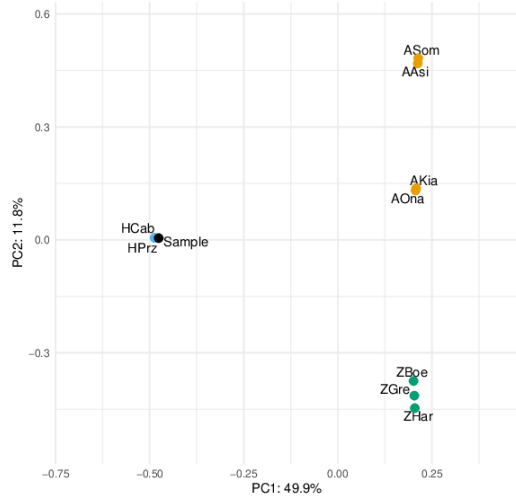**3 ancestral groups, excluding transitions**
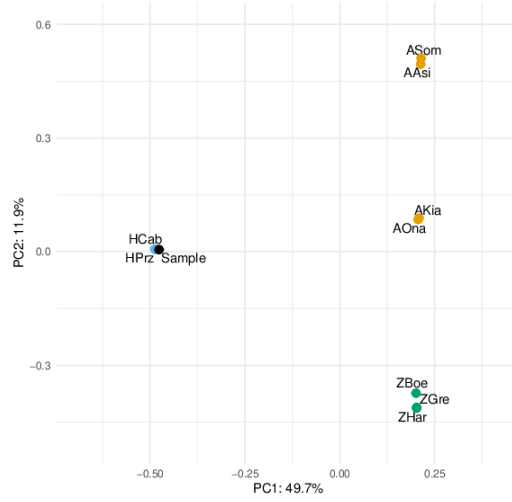

No admixture detected.

PCA PLOTS

Principal Component Analysis carried out using SmartPCA v16000, from the **EIGENSOFT** toolkit.

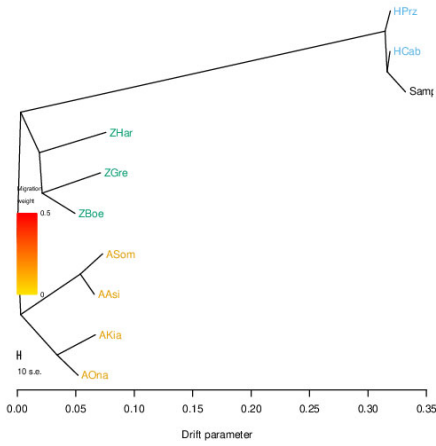**Including transitions**
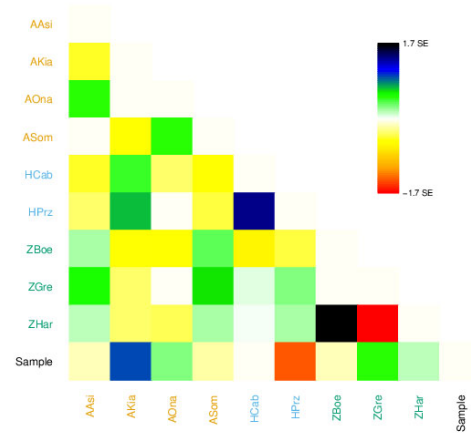
Excluding transitions

TREEMIX PLOTS

Detection of population mixture using **TreeMix** v1.13 *[Pickrell and Pritchard 2012]*; parameters were -k 0; -global; and supervised estimation using ancestral groups listed in the Reference Panel.
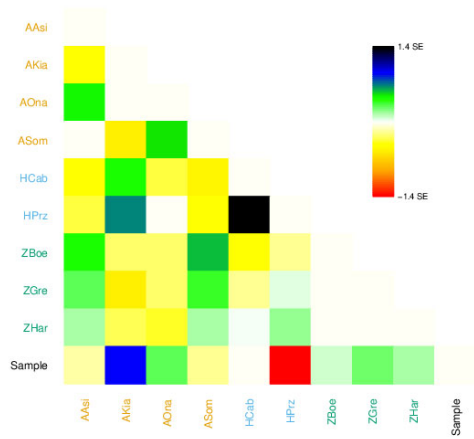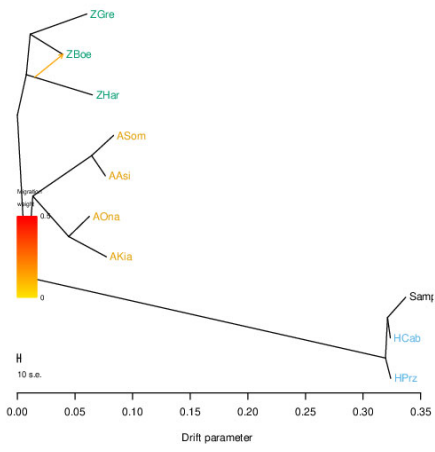
Including transitions

Edges = 0

Residuals

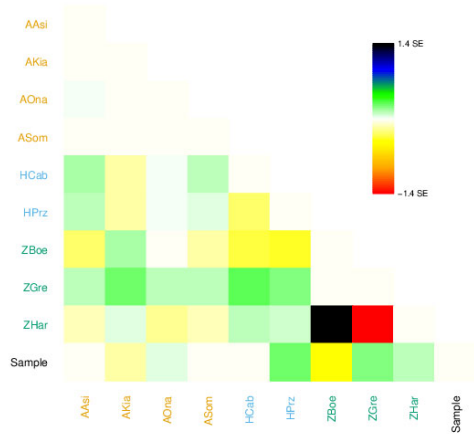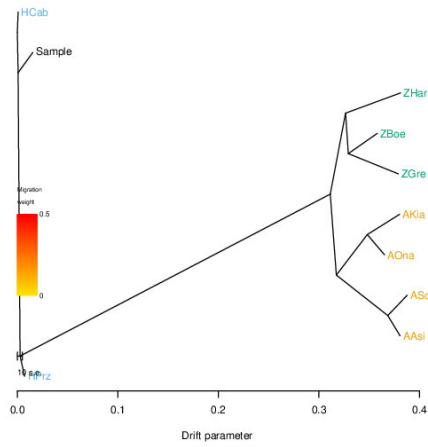Variance explained by model = 0.999994.

Edges = 1

Residuals

Variance explained by model = 0.999996.

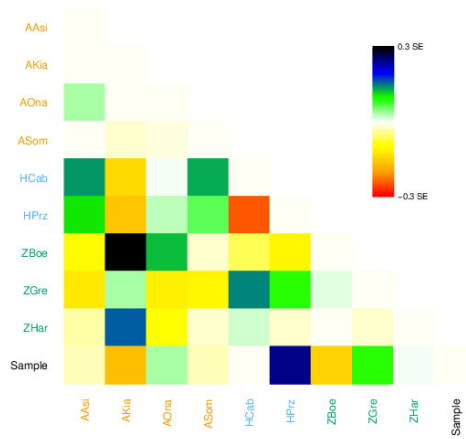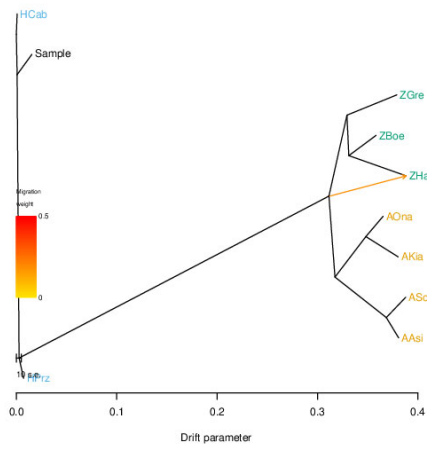# Excluding transitions

Edges = 0

Residuals

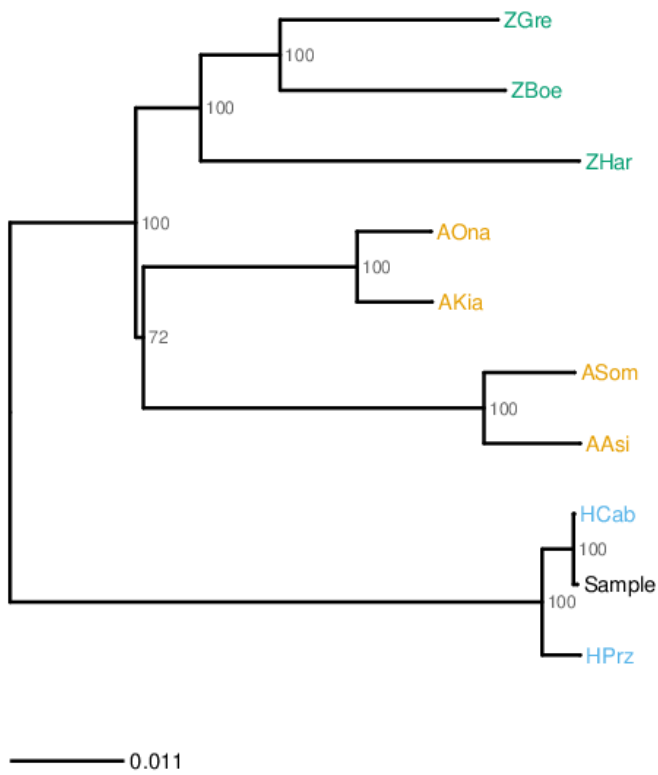Variance explained by model = 0.999994.

Edges = 1

Residuals

Variance explained by model = 0.999999.

MITOCHONDRIAL PHYLOGENY

Phylogenetic inference performed using RAxML v8.2.12 [*Stamatakis 2006*].

REFERENCES

- Alexander *et al*. "**Fast model-based estimation of ancestry in unrelated individuals**". *Genome Res*. 2009 Sep;19(9):1655-64. doi: **10.1101/gr.094052.109**. PMID: **19648217**.
- Li *et al*. "**The Sequence Alignment/Map format and SAMtools**. *Bioinformatics*. 2009 Aug 15;25(16):2078-9. doi: **10.1093/bioinformatics/btp352**. PMID: **19505943**.
- Pickrell and Pritchard. "**Inference of population splits and mixtures from genome-wide allele frequency data**". *PLoS Genet*. 2012;8(11):e1002967. doi: **10.1371/journal.pgen.1002967**. PMID: **23166502**.
- Purcell *et al*. "**PLINK: a tool set for whole- genome association and population-based linkage analyses**". *Am J Hum Genet*. 2007 Sep;81(3):559-75. PMID: **17701901**.
- Stamatakis. "**RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models**". *Bioinformatics*. 2006 Nov 1;22(21):2688-90. Epub 2006 Aug 23. doi: **10.1093/bioinformatics/btl446**. PMID: **16928733**.