**Title: Accurate sex identification of ancient elephant and other animal remains using low coverage DNA shotgun sequencing data**

Alida de Flamingh\*, Ashley Coutu†, ‡,§, Alfred L. Roca\*,\*\*,††

& Ripan S. Malhi \*,††,§§

***Author affiliation:***

\*Program in Ecology, Evolution and Conservation Biology, University of Illinois at Urbana-Champaign (UIUC), Urbana IL, 61801, USA

†Department of Archaeology, University of Cape Town, Cape Town, 7700, RSA

‡BioArCh, University of York, York YO10 5NG, UK

§Pitt Rivers Museum, University of Oxford, Oxford OX1 3PP, UK

\*\*Department of Animal Sciences, UIUC, Urbana IL, 61801, USA

††Carl R. Woese Institute for Genomic Biology, UIUC, Urbana IL, 61801, USA

§§Department of Anthropology, UIUC, Urbana IL, 61801, USA

Supplementary Appendix 1:

Appendix 1. Steps for adapting the Rx\_identifier.r script from Mittnik *et al.* (2016) to any target species that has a chromosome level reference genome and XY sex determination.

1. Align the target species sequence reads to the chromosome level reference genome
   1. Convert resulting SAM alignment files to BAM files and filter out unmapped and low-quality reads
   2. Remove PCR duplicates, sort and index SAM files using SAMtools (or similar software)
   3. Generate index statistic files for each BAM file using the “idxstats” function in SAMtools (or similar software). Add the suffix “.idxstats” to each resulting file.
   4. the "idxstats" files are used as input for the Rx\_identifier R script. Each row in an “idxstats” file contains the reference sequence name, sequence length, # mapped reads and # unmapped reads.
   5. Verify that idxstats files contain the number of reads mapped to each chromosome and determine which row number represents the X chromosome.
2. Adapt Rx\_identifier.r script for the target species of interest
   1. Download the Rx\_identifier.r script from Mittnik *et al.* (2016).
   2. Edit line 15 in the script so that nrows reflect the number of chromosomes (autosomal + sex chromosomes) of your target species reference genome (e.g. nrows = 24 for humans, nrows = 28 for African savanna elephants – see Appendix 2, Rx\_identifier.r script that has been adapted for elephants).
   3. Rows 24 to 47 calculates for each chromosome the ratio of the alignments to each chromosome to the total number of alignments to autosomes and sex chromosomes
      1. Adjust the script to reflect the number of chromosomes in the target species (i.e. add or remove lines per chromosome so that the final number of lines of script are equal to the number of chromosomes on the target species reference genome).
      2. In the Mittnik *et al.* (2016) script line 23 in the idxstats files represent the X chromosome, therefore Rt23 in the Rx\_identifier.r script (line 46 in the script) represents the ratio of alignments to the X chromosome to the total number of alignments to all chromosomes
      3. For the script adapted for elephants (Supplementary Appendix 2), the “idxstats” files contain X chromosome information on the 28th line of each file, therefore Rt28 represents the X chromosome.
   4. Edit line 49 in the script so that the numerator represents the ratio associated with the X chromosome. Add or remove fractions so that all chromosomes in the reference sequence are represented in the equation. However, since one is comparing the X chromosome to autosomal chromosomes, do not include the Y chromosome ratio (e.g. Rt24 from line 47 in the Mittnik *et al.* (2016) script) in the calculation.
   5. Edit line 52 so that the SE measures the amount of variability in the Rx mean compared with the number of autosomes in the target species (22 in humans, 27 in elephants).
   6. Save and add the modified script to the same folder as the “idxstat” files and run the script in R (R version 3.6.1. R Core Team, 2019).
   7. For each “idxstats” file the script produces a result file with the suffix “.Rx” which contains information on the sex determination statistics and whether sex identification is possible for the individual.

References:

Mittnik, A., C. C. Wang, J. Svoboda, and J. Krause, 2016 A molecular approach to the sexing of the triple burial at the upper paleolithic site of Dolní Vêstonice. PLoS One 11: 1–9.