

Population Genomics and Demographic History of South Asian Megafauna

Masters in Bioinformatics - Spring 2026

Laurids Svenningsen
202006255

Msc. Student Bioinformatics,
Aarhus University

Juraj Bergman
Supervisor

Tenure-track Assistant Professor,
Department of Molecular Biology and
Genetics - BiRC,
Aarhus University

11.06.2026

Abstract

Understanding the demographic histories of megafauna is essential for conservation biology and for resolving the relative roles of climate change and human impact in driving population decline. This thesis assembles one of the most comprehensive population-level whole-genome sequencing datasets for living megafauna to date, encompassing 534 individuals across eleven species from South Asia and Africa. Using a largely automated bioinformatics pipeline created for this study, long-term effective population size (N_e) trajectories were inferred using SMC++, alongside recent trajectories estimated using GONE2, providing complementary reconstructions spanning deep evolutionary history to the present. SMC++ trajectories revealed a broadly consistent pattern across species, with elevated N_e during the mid to late Pleistocene followed by progressive declines toward the present. In several species, this peak broadly overlaps the Last Glacial Maximum ($\sim 20,000$ years ago). Contrary to expectations, no clear distinction was observed between the demographic histories of South Asian and African megafauna, suggesting that shared global drivers — particularly climate cycles and human expansion — have shaped megafauna demography more strongly than region-specific factors. GONE2-derived recent N_e trajectories showed predominantly negative correlations with human impact variables across most species, consistent with associations between increasing human populations, land use expansion, and megafauna decline over recent centuries. Mixed-effects modelling of SMC++-derived trajectories against palaeoclimate and human arrival pressure indicated that the human-only model provided the best fit to the data. Both human pressure and mean annual temperature showed negative associations with N_e , although neither effect reached statistical significance, likely reflecting limited statistical power given the number of species analysed. Together, these results support a model in which megafauna demographic histories have been shaped by a combination of long-term climatic processes and more recent anthropogenic pressures, with human impact emerging as a key driver of recent population decline across species.

Contents

Abstract	2
1 Introduction	6
1.1 Ecological Background	6
1.1.1 South Asia	6
1.1.2 Africa as a Comparison	7
1.2 Effective population size	8
1.3 Project Aims	9
2 Materials and Methods	11
2.1 Data Acquisition	11
2.1.1 Sample Acquisition and Filtering	11
2.1.2 Reference Genome Selection	11
2.2 Sequence Processing and Alignment	12
2.2.1 Reference Genome Indexing and Preparation	12
2.2.2 Sample Download and Mapping Preparation	12
2.2.3 Read Mapping	12
2.2.4 Variant Calling and Processing	13
2.2.5 Genotyping	13
2.3 Quality Control and Filtering	13
2.4 Population Structure and Relatedness Analysis	14
2.4.1 Population Structure Inference Using ADMIXTURE	14
2.4.2 Principal Component Analysis	15
2.4.3 Relatedness Estimation	15
2.5 Effective Population Size Inference	16
2.5.1 Preparation and Data Conditioning	16
2.5.2 Masking	16
2.5.2.1 Mappability	16
2.5.2.2 Coverage	17
2.5.2.3 Repetitive Regions	17
2.5.2.4 Runs of Homozygosity	17
2.5.3 Long-term N_e Inference (SMC++)	18
2.5.4 Recent N_e Inference (GONE2)	19
2.6 Mixed-Effects Modelling of N_e Trajectories	19
2.6.1 Data Compilation	19

2.6.2	Mixed-effects modelling	20
2.7	Correlation Analysis of Recent Demography and Human Impact	21
2.7.1	Data Alignment and Interpolation	21
2.7.2	Correlation Analysis	21
2.8	Code Availability	22
3	Results	23
3.1	Coverage and Quality Assessment	23
3.2	Population Structure	23
3.2.1	ADMIXTURE	23
3.2.2	Principal Component Analysis	24
3.2.3	Heterozygosity	25
3.3	Long-term N_e Trajectories (SMC++)	27
3.4	Recent N_e Estimates (GONE2)	29
3.5	Mixed-Effects Modelling Results	30
3.5.1	Model comparison	30
3.5.2	Fixed effects	31
3.5.3	Predicted vs observed N_e	32
3.6	Human Impact Correlation Results	32
4	Discussion	34
4.1	Dataset and Quality Considerations	34
4.2	Reliability of Demographic Inference	36
4.2.1	Interpreting the IICR	36
4.2.2	SMC++-Specific Limitations	36
4.2.3	GONE2-Specific Limitations	38
4.3	Comparative Demographic Histories Across Species	39
4.3.1	Recent Drivers: Correlations with Human Impact	40
4.4	Long-term Drivers: Mixed-Effects Modelling	42
4.5	Conclusion	44
4.6	Outlook and Future Directions	45
	References	47
	Appendix A Tables	i
A.1	Projects Accessions and Sample Counts	i
A.2	References Genomes, Generation Times and Mutation Rates	ii
	Appendix B Figures	iii
B.1	ADMIXTURE Clustering Results	iii

B.2 Coverage-Based Quality Control Results	iv
B.3 Heatmap of Masking Statistics	vi
B.4 Correlations Between Recent Ne and HYDE Data	ix
B.5 Combined SMC++ Trajectories Used for Modelling	xiv
B.6 Temperature Variable Used in Mixed-Effects Modelling	xiv
B.7 Human Pressure Used in Mixed-Effects Modelling	xv