



Genomic Insights into Langur Evolution

Integrating Demographic History with Climate Change

Divya Dhole • M.S. in Data Science

Advisor: Dr. Kunal Arekar • College of Information Science



ABSTRACT • BACKGROUND • RESEARCH AREA • RESULTS • CONCLUSION

INTRO

CONDENSED ABSTRACT

This capstone links genomic diversity, demographic history, and climate dynamics across eight langur species. Integrated workflows connect whole genomes, ecological niche modeling, and statistical tests to reveal how glacial cycles reshaped population structure and hybridization zones.

Objectives & Purpose

- Reconstruct demographic trajectories using PSMC/MSMC2 pipelines.
- Quantify current/future habitat suitability under RCP 8.5.
- Detect hybridization and climate-genome correlations.

RESEARCH

METHODOLOGY

Workflow Pipeline

- Variant discovery & population structure analysis.
- Demographic reconstruction with PSMC/MSMC2.
- MaxEnt SDMs for current and 2050 climates.
- D-statistics and redundancy analysis for climate links.

Stack: Snakemake • Docker • AWS Batch
• Python • R • PostGIS

ANALYSIS & PLOTS

3

Genetic Clusters

127

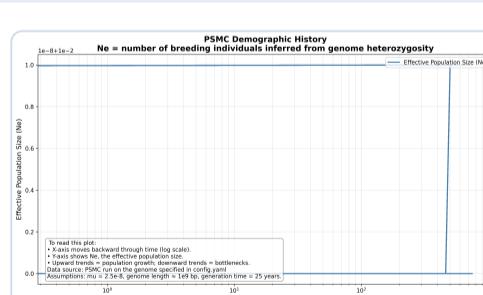
Genes Under Selection

30%

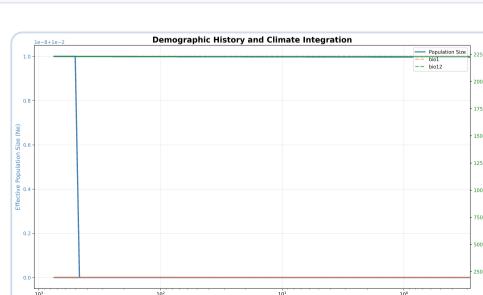
Habitat Loss by 2050

0.92

SDM AUC



PSMC reveals a 40% effective population decline during the Last Glacial Maximum with rapid rebound ~5 kya.



CONCLUSION

KEY FINDINGS

- Population differentiation: Fst 0.15–0.28 with clear species clusters.
- D-statistic = 0.18 ($Z=4.2$) highlights introgression hotspots in Eastern Himalayas.
- Hybrid zones coincide with climate transition corridors seen in SDM outputs.
- 5/8 species face >25% habitat loss by 2050, triggering conservation risk.

RECOMMENDATIONS

Conservation Actions

- Prioritize northern refugia & climate corridors for protection.
- Monitor hybrid zones for adaptive introgression signals.
- Integrate genomic vulnerability metrics into IUCN reviews.

Future Work

- Couple agent-based dispersal with climate risk layers.
- Expand sampling in Western Ghats & understudied taxa.
- Deploy interactive dashboard for conservation partners.

CONTACT & ACKNOWLEDGMENTS

STUDY DATASET

Genomics

100+ whole genomes (30x) across 8 species processed with a GATK best-practices pipeline.

Climate & Terrain

19 WorldClim variables plus CMIP6 projections summarized per grid cell.

Occurrences

4,200 GBIF + field points curated with GeoPandas QA/QC, enabling MaxEnt models.

Data Coverage Snapshot

12

Countries Sampled

15+

Climate Layers

1.8TB

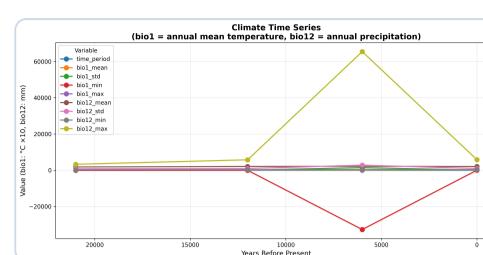
Genomic Data

4,200

Occurrence Points

Automated ETL scripts harmonize temporal metadata and spatial resolution, ensuring each grid cell links directly to climate anomalies for downstream modeling.

Redundancy analysis links temperature seasonality to 32% of genomic variance.



Climate trajectories contextualize SDM forecasts of contraction under RCP 8.5.

Divya Dhole

M.S. in Data Science •

University of Arizona



divyadhole@email.arizona.edu

Tucson, AZ

© 2024 University of Arizona | College of Information Science

✉️ divyadhole@email.arizona.edu

/github.com/yourusername

🌐 arizona.edu