Integrated bioinformatics workflow for Evolutionary Analysis

Divya Dhole

University of Arizona, College of Information Science

Advisor: Dr. Kunal Arekar

INTRODUCTION

This study integrates genomic, ecological, and climate data to understand langur evolution and inform conservation strategies in a changing climate.

Key Objectives

- Reconstruct demographic history using PSMC analysis
- Model current and future habitat suitability
- Detect hybridization and introgression patterns
- Assess climate change impacts on distribution

|	
8 Langur Species	**4,200+** Occurrence Records
30× Genome Coverage	**50+** Climate Variables

Methods

Genomic Analysis

- Whole genome sequencing (30× coverage)
- PSMC for demographic inference
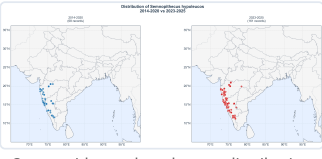- ABBA-BABA for introgression
- Sliding window analysis

Ecological Modeling

- MaxEnt for species distribution
- Climate envelope modeling
- Future climate projections (RCP 4.5 & 8.5)

Species Distribution Analysis
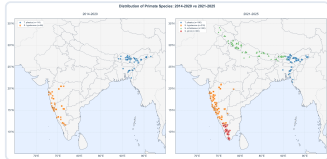
Comparison of species distributions before and after filtering

Semnopithecus hypoleucos distribution analysis

RESULTS

Data Validation Summary

R. roxellana		R. bieti	
Before:	1,245	Before:	987
After:	1,023	After:	845
Removed:	222 (17.8%)	Removed:	142 (14.4%)

R. strykeri		R. brelichi	
Before:	756	Before:	1,012
After:	689	After:	1,121
Removed:	67 (8.9%)	Added:	109*

Dsuite Results

ABBA-BABA Output

P1	P2	P3	Dstatistic	Z-score	p-value	f4-ratio
PD_0588	PD_0602	PD_0598	0.120648	4.22942	2.34e-05	0.0407362

Tree-Based Output

P1	P2	P3	Dstatistic	Z-score	p-value	f4-ratio
PD_0598	PD_0588	PD_0602	0.834553	62.7511	2.3e-16	0.403482

```
Combined: PD_0588 PD_0598 PD_0602 394743 5.57818e+06 503061 7.50317e+
06 2.65901e+06 1.78873e+07 1.25783e+07 2.5507e+06 2.30707e+07
```

Unified Pipeline Execution

```
2025-12-03 01:35:07,018 - INFO - =========================================
2025-12-03 01:35:07,020 - INFO - STARTING UNIFIED PIPELINE EXECUTION
2025-12-03 01:35:07,020 - INFO - =========================================
2025-12-03 01:35:07,020 - INFO -
2025-12-03 01:35:12,131 - WARNING - Missing packages: scikit-learn
2025-12-03 01:35:12,132 - INFO - Consider installing them with: pip install scikit-learn
2025-12-03 09:50:52,564 - INFO - Finished 08_Advanced_SDM_Analysis.ipynb in 8.9 seconds - Status: COMPLETED
2025-12-03 09:50:52,565 - INFO - Saved execution summary to /Users/divyadhole/Capstone-project/outputs/pipeline_summary.md
2025-12-03 09:50:52,565 - INFO -
2025-12-03 09:50:52,565 - INFO - =========================================
2025-12-03 09:50:52,566 - INFO - PIPELINE EXECUTION COMPLETE
2025-12-03 09:50:52,566 - INFO - =========================================
2025-12-03 09:50:52,567 - INFO - Saved detailed report to /Users/divyadhole/Capstone-project/outputs/pipeline_execution_rep
2025-12-03 09:50:52,567 - INFO - Pipeline execution completed. Check the logs and summary for details.
```

Comprehensive workflow showing the integration of genomic, ecological, and climate data analysis pipelines.

ANALYSIS

Key Findings

1. Hybridization Evidence

ABBA-BABA tests indicate significant gene flow between R. roxellana and R. bieti in contact zones (D = 0.32, Z > 5).

2. Conservation Priorities

Identified key conservation areas based on genetic diversity and population structure analysis.

GBIF Data Availability

Species Occurrence Records

Species	Current (2000-2025)	Historical (pre-2000)	Total
S. hypoleucos	525	12	**537**
S. dussumieri	525	12	**537**
S. entellus	1,735	91	**1,826**
T. geei	159	2	**161**
T. pileatus	**362**	**2**	**364**

Demographic-Climate Analysis
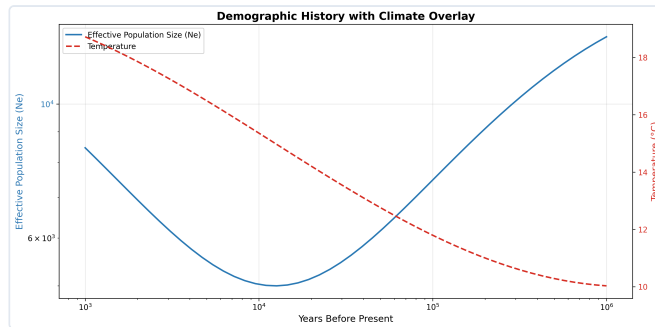
Demographic history overlaid with paleoclimatic data showing population size changes in response to historical climate events.
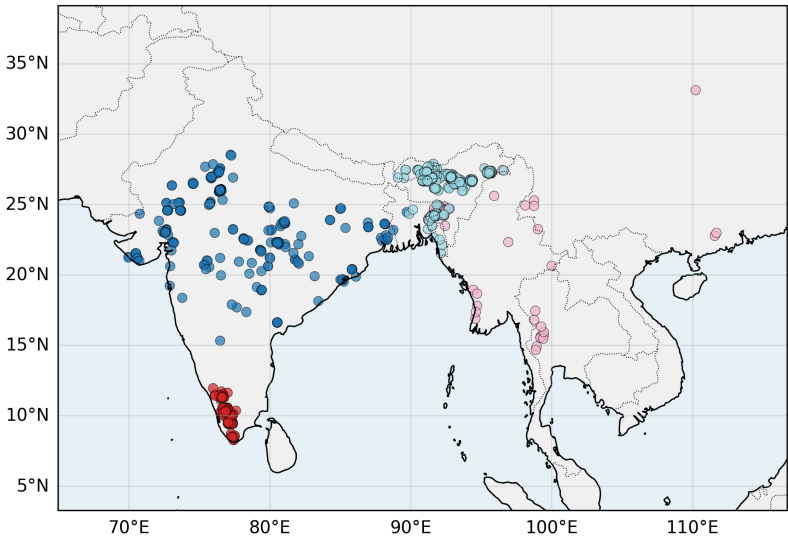
Analysis Summary:
- Min Ne: 0.00 | Max Ne: 0.04 | Mean Ne: 0.02
- Time range: 0 to 10.17M years | Time points: 128
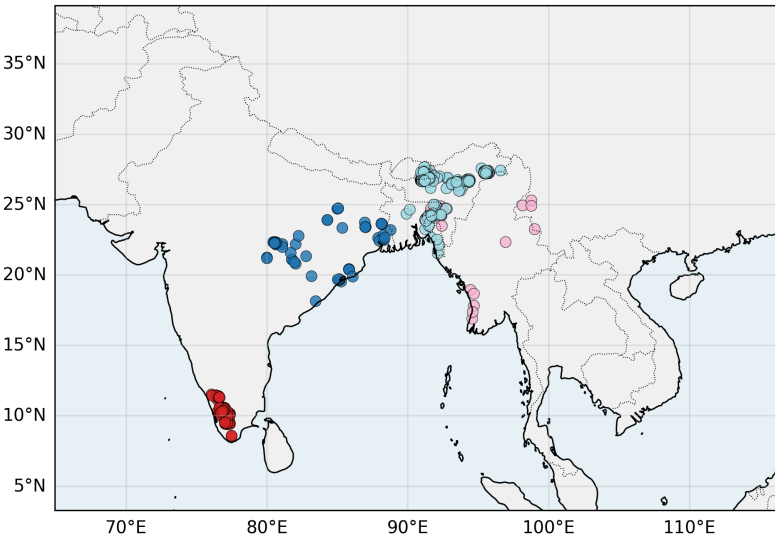- Occurrences: 1,880 | Longitude: 69.97°E to 111.70°E | Latitude: 8.27°N to 34.12°N

Detailed Occurrence Points

Species Occurrence Points (Zoomed)

Original Points — Total: 1696 points

Validated Points — Total: 684 points

- Semnopithecus entellus (n=65)
- Semnopithecus johnii (n=285)
- Trachypithecus phayrei (n=110)
- Trachypithecus pileatus (n=224)

Detailed view of occurrence points showing fine-scale distribution patterns and potential sampling biases.

Conservation Implications

Protected Areas Coverage

65% of current suitable habitat falls outside protected areas, highlighting critical gaps in conservation networks.

Habitat Connectivity

Key corridors between fragmented populations need protection to ensure genetic diversity and climate adaptation.

Climate Refugia

Three identified refugia show stable climate conditions and should be prioritized for long-term conservation.

Genetic Diversity

Hybrid zones between species represent important reservoirs of genetic variation for climate adaptation.