

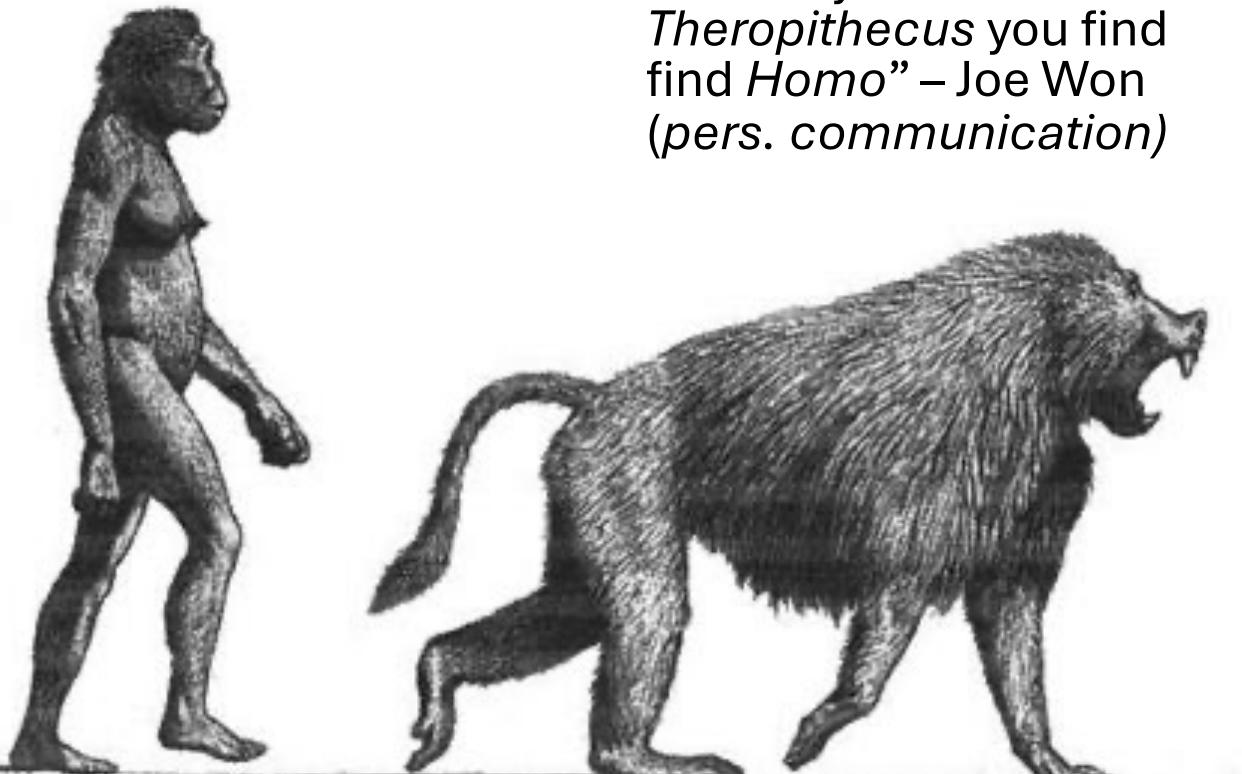
Gelada genomes highlight events of gene flow, hybridisation and local adaptation that track past climatic changes



Matteo Caldon¹  | Giacomo Mutti^{1,2,3} | Alessandro Mondanaro⁴ | Hiroo Imai⁵  |
Takayoshi Shotake⁶ | Gonzalo Oteo Garcia^{1,7,8}  | Gurja Belay⁹ | Jordi Morata¹⁰ |
Jean-Rémi Trotta¹⁰ | Francesco Montinaro^{11,12} | Spartaco Gippoliti^{13,14} |
Cristian Capelli^{1,15} 

Bioinformatic Workflow and Results Joseph Won

Theropithecus gelada



- Geladas are the last surviving species of an extremely successful Pleistocene taxon *Theropithecus*
- The only fully gramnivorous (grass eaters) primates
- “Where you find *Theropithecus* you find *Homo*” – Joe Won (pers. communication)

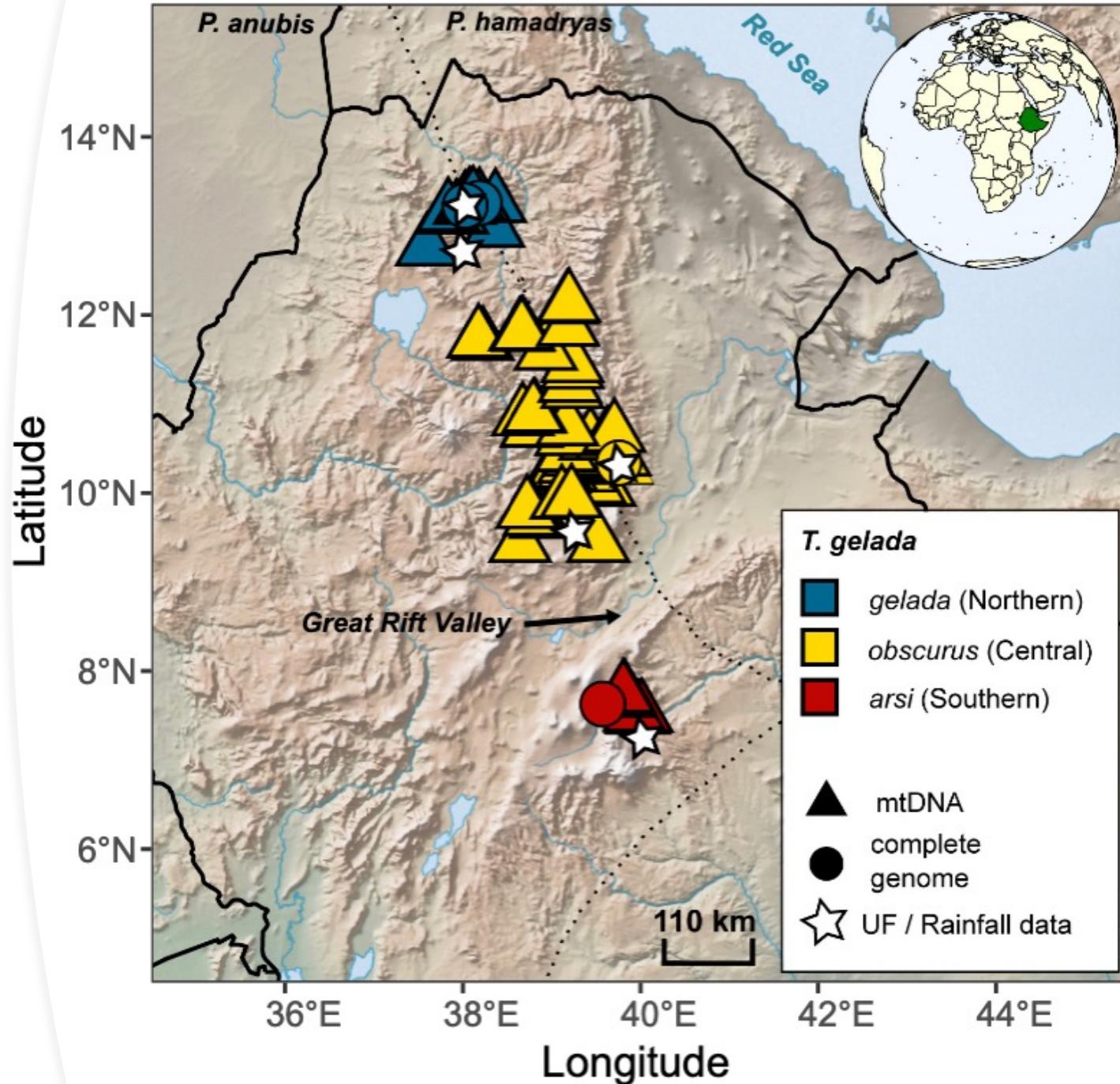


Extant populations and Research Questions

There are three populations of *Theropithecus gelada* in the Ethiopian highlands

Research Questions

1. How are these populations related? (mtDNA, Y-Chromosomes, autosomes)
2. How has environmental instability affected Gelada populations?
3. Has introgression occurred between *Theropithecus* and other Papionin species? (e.g. *Papio*, *Lophocebus*)

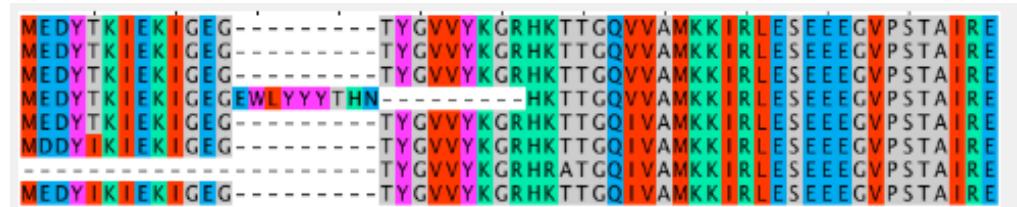


Materials and Quality Control

- *Theropithecus gelada*
 - Previous studies (Shotake et al., 2016)
 - Southern populations (30x coverage)
 - Zoo (33.65 coverage)
- Other papionins
 - *Lophocebus aterrimus*
 - *Papio hamadryas* and *Papio anubis*
 - *Mandrillus leucophaeus*
- **fastp v.0.23.2** – preprocessing tool
 - Adapter trimming
 - Quality filtering
- **BWA-MEM2 v.2.2.1** – mapping to reference genome
 - Tgel_1.0 (Chiou et al., 2022) – *T. gelada*
 - As they were comparing with other genera, they mapped to those also
- **Picard v.2.8.1** – mark duplicate reads
 - Avoid counting duplicates multiple times during variant calling

mtDNA Workflow

- **GetOrganelle v1.7.5**
 - newly sequenced samples were assembled using this tool
- **MAFFT v7.490**
 - Alignment tool for homologous sequences
- **IQ-TREE v.2.2.0:**
 - maximum likelihood methods
- **BEAST v.2.7.1:**
 - Bayesian analysis tool used for estimating evolutionary trees and divergence times.
 - Molecular clock:
 - You can enforce monophyly (2 Macaca mtDNA samples)
 - You can also fossil calibrate (papiro-theropithecus = 5mya)



.fa

```
mammal_data_analysis_gtr --zsh - 80x24
Reading alignment file /users/josephwon/COMPGEN/mammal_data_analysis_gtr/m11337_subsampled_gb.fas ... Fasta format detected
Alignment most likely contains DNA/RNA sequences
Alignment has 9 sequences with 1666 columns, 235 distinct patterns
157 parsimony-informative, 223 singleton sites, 1286 constant sites
Gap/Ambiguity Composition p-value
1 monodelphis_domestica 0.00% passed 56.62%
2 ornithorhynchus_anatinus 24.67% passed 32.66%
3 pan_troglodytes 0.00% passed 98.76%
4 gorilla_gorilla 0.00% passed 98.76%
5 sarcophilus_harrisii 0.24% passed 66.94%
6 notamacropus_eugenii 0.48% passed 55.16%
7 capra_hircus 0.00% passed 55.63%
8 sus_scrofa 0.12% passed 57.52%
9 heterocephalus_glaber_female 62.06% failed 3.67%
WARNING: 1 sequences contain more than 50% gaps/ambiguity
*** TOTAL 9.73% 1 sequences failed composition chi2 test (p-value<5%; df=3)
Create initial parsimony tree by phylogenetic likelihood library (PLL)... 0.000 seconds
NOTE: 0 MB RAM (0 GB) is required!
Estimate model parameters (epsilon = 0.100)
```

IQ-TREE

Or

BEAST

Y-Chromosome Workflow

- **VCFTools v.0.1.16**
 - called the variants on the Y chromosome
 - Ploidy = 1 (b/c Y- Chromosome)
- **BCFtools v.1.15.1**
 - filtered constant sites, indels and variants in repetitive regions
 - “minimum read depth (DP) lower than 250, maximum DP greater than 600, minimum mapping quality less than 55 and absent in less than 10 samples were removed. Finally, the resulting variants were thinned out by keeping only variants at least 100 bp apart”
- **BEAST v.2.7.1:**
- **poppr R package:**
 - Construct a phylogenetic tree using the Bruvo distance

VCFTools v.0.1.16

VCFTools

A set of tools written in Perl and C++ for working with [VCF files](#), such as those generated by the [1000 Genomes Project](#).

Project website: <https://vcftools.github.io/>

License

The program package is released under the GNU Lesser General Public License version 3.0 (LGPLv3). See the [LICENSE](#) file for the complete LGPL license text.

BCFtools v.1.15.1

build passing build passing downloads 655k

This is the official development repository for BCFtools. It contains all the vcf* commands which previously lived in the htllib repository (such as vcfcheck, vcfmerge, vcfisec, etc.) and the samtools BCF calling from bcftools subdirectory of samtools.

For a full documentation, see [bcftools GitHub page](#).

Other useful links:

File format specifications live on [HTS-spec GitHub page htllib samtools tabix](#)

Citing

Please cite this paper when using BCFtools for your publications.
<http://samtools.github.io/bcftools/howtos/publications.html>



BEAST v.2.7.1:

.gVCF

poppr R package:

Autosomal Workflow

- **GATK**
 - called the variants on the Y chromosome
 - Ploidy = 1 (b/c Y- Chromosome)
- **EIGENSOFT**
 - Perform PCA on molecular data
- **PLINK v.1.9**
 - Calculate heterozygosity
 - Total length
 - Runs of homozygosity (RoH)
- **qpDstat (D-statistics)**
 - Test if there is gene flow between populations
- **Admixtools2 R Package**
 - Explore gene flow EVENTS

```
[HEADER LINES]
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA12878
20 10001019 . T G 364.77 . AC=1;AF=0.500;AN=2;BaseQRankSum=0.699;ClippingRankSum=0.000;DP=27;ExcessHet=3.0103;FS=0.000;MLEAC=0.500;MLEAF=0.500;N=2;NS=1;PValue=0.000;QD=1.333;RD=0.000;SOR=0.000
20 10001298 . T A 884.77 . AC=2;AF=1.00;AN=2;DP=30;ExcessHet=3.0103;FS=0.000;MLEAC=1.000;MLEAF=1.000;N=2;NS=1;PValue=0.000;QD=1.333;RD=0.000;SOR=0.000
20 10001436 . A AAGGCT 1222.73 . AC=2;AF=1.00;AN=2;DP=29;ExcessHet=3.0103;FS=0.000;MLEAC=1.000;MLEAF=1.000;N=2;NS=1;PValue=0.000;QD=1.333;RD=0.000;SOR=0.000
20 10001474 . C T 843.77 . AC=2;AF=1.00;AN=2;DP=27;ExcessHet=3.0103;FS=0.000;MLEAC=1.000;MLEAF=1.000;N=2;NS=1;PValue=0.000;QD=1.333;RD=0.000;SOR=0.000
20 10001617 . C A 493.77 . AC=1;AF=0.500;AN=2;BaseQRankSum=1.63;ClippingRankSum=0.000;DP=27;ExcessHet=3.0103;FS=0.000;MLEAC=0.500;MLEAF=0.500;N=2;NS=1;PValue=0.000;QD=1.333;RD=0.000;SOR=0.000
```

GATK – variant calling .gVCF

EIGENSOFT

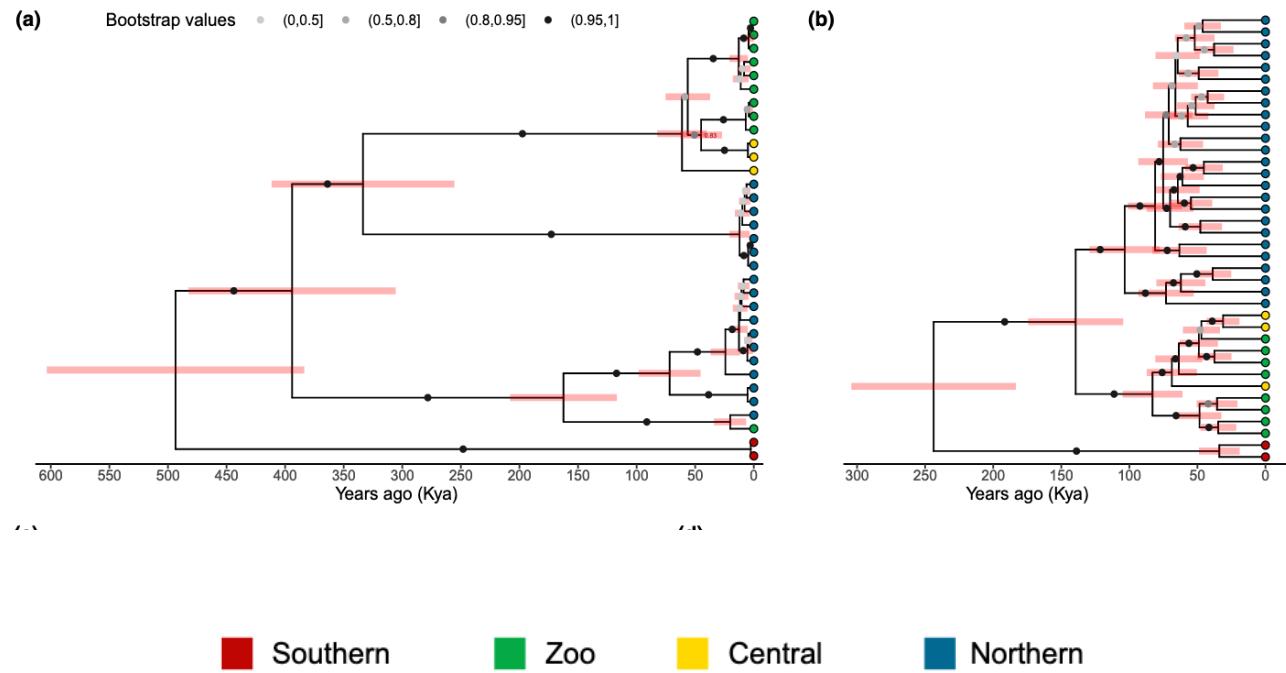
Admixtools2 R Package

PLINK v.1.9

qpDstat (D-statistics)



Uniparental Results



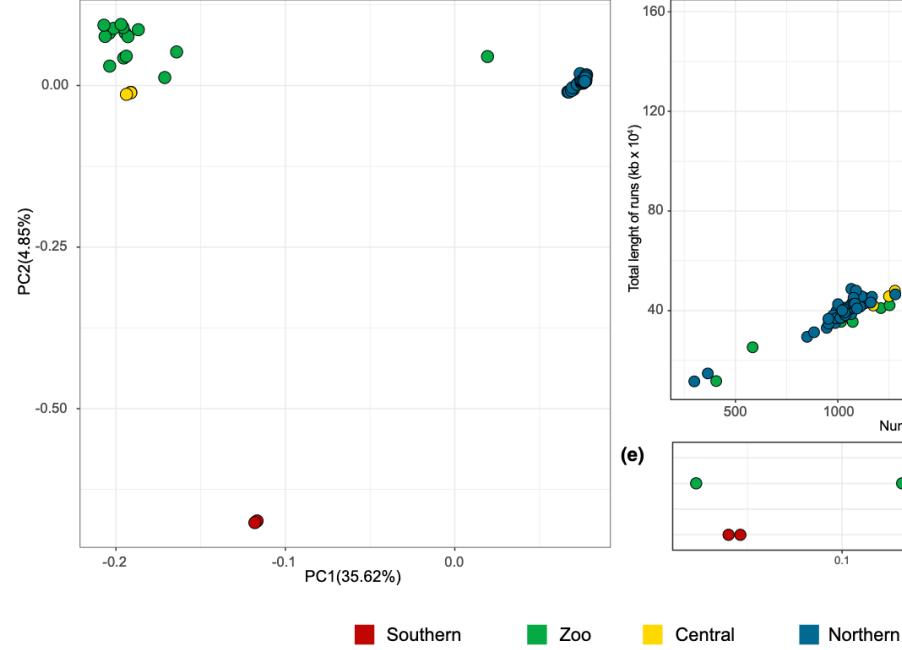
- mtDNA

- Southern geladas are deeply divergent and form a sister group to Northern and Central populations
- Central geladas are monophyletic but nested within Northern

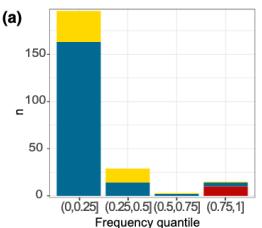
- Y-Chromosome

- Also clusters the populations but Northern and Central form sister clades
- TMRCA is .24 MYA for Y Chromosome but .49 MYA for mtDNA

(c)



(a)



(b)

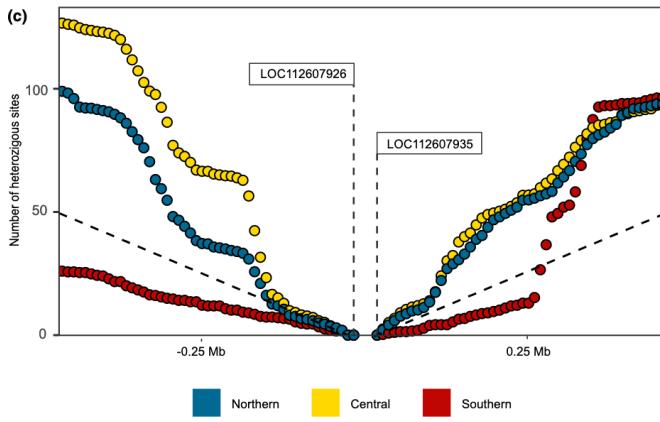
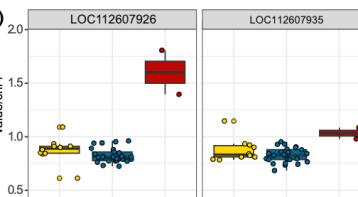


FIGURE 5 CNVs and amylase gene copy number. (a) Population-specific CNVs divided according to their frequency in each population. (b) Normalised genomic coverage over the two loci annotated as pancreatic amylase in geladas. (c) Cumulative number of heterozygous positions in 10-kb windows, starting from the beginning/end of the two geladas pancreatic amylase gene average and extended for 0.5 Mb. Reported is the average across individuals for the indicated population, each individual estimate normalised for the average number of heterozygous positions estimated on chromosome 1 in each individual (see Section 2); dashed lines refer to the expected cumulative curve for 10-kb windows each displaying the average number of heterozygous positions along chromosome 1.

Autosomal Results

- Variation
 - PCA and Heterozygosity measurements indicate southern populations are distinct and less diverse
- Demographic History
 - All experienced a decrease in population size and connectivity in last 500,000 years
- Hybridization
 - D statistics revealed *Papio* shares more alleles with Geladas than *Lophocebus*
 - 11% contribution from *Theropithecus* to *P. hamadryas*
- Copy Number Variants
 - A duplication in the pancreatic amylase gene was found in southern populations
 - Positive selection

Importance of Population Genomics



Plio-Pleistocene
Migration

Environmental
Shifts and Conservation



Papionin love/Hybridization

