

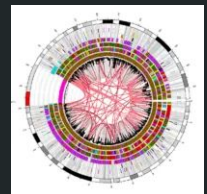
Population Genomics

Stories of bears, penguins and maybe beans



Emiliano Trucchi

Workshop on Genomics 2020
Cesky Krumlov



Emiliano Trucchi

Research on: Inferences of demographic and adaptive processes in natural populations of (mostly) non-model species.

Keywords: population genetics, genomics, molecular ecology, **conservation genetics**, phylogeography, phylogenetics, bioinformatics.

My academic record: University of Roma3 (Master), Tor Vergata (PhD), Oslo (postdoc 1 - Marie Curie), Vienna (postdoc 2), Ferrara (researcher), Marche Polytechnic (assistant professor).

emiliano.trucchi@univpm.it



CEES
Centre for Ecological and Evolutionary Synthesis



What is population genetics? (back to basics)

“The study of the **genetic composition** of natural populations and its evolutionary causes and consequences” (Coop 2019)

Genetic drift

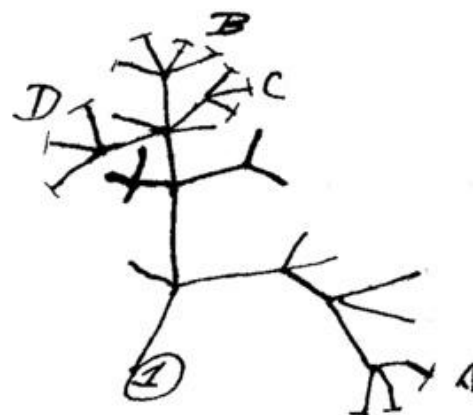
Migration

Recombination

Mutation

Selection

I think



What is population genetics? (back to basics)

A set of **theoretical models** to understand how these forces interact

Simple as any model simplifying reality

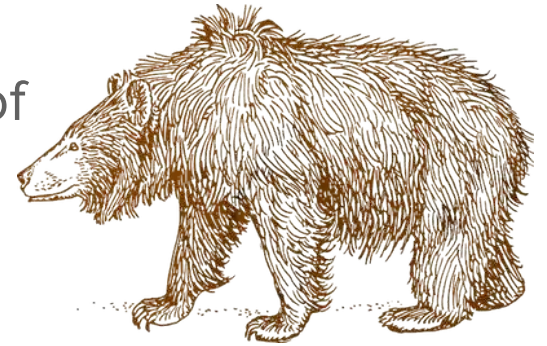
(... all models are wrong but some are useful, Box 1979)

Useful for understanding real patterns - give accurate predictions
(e.g., medical genetics, crop improvements, species conservation)

Powerful as the basic rules of genetic transmissions are simple and universal

Population genomics as a tool in conservation biology

SIDE A: The extraordinary genomic history of the endangered Apennine Brown Bear



SIDE B: Climate-driven range shifts in fragmented ecosystems

Bonus track: Ancient genomes reveal early farmers selected common bean while preserving diversity



SIDE A

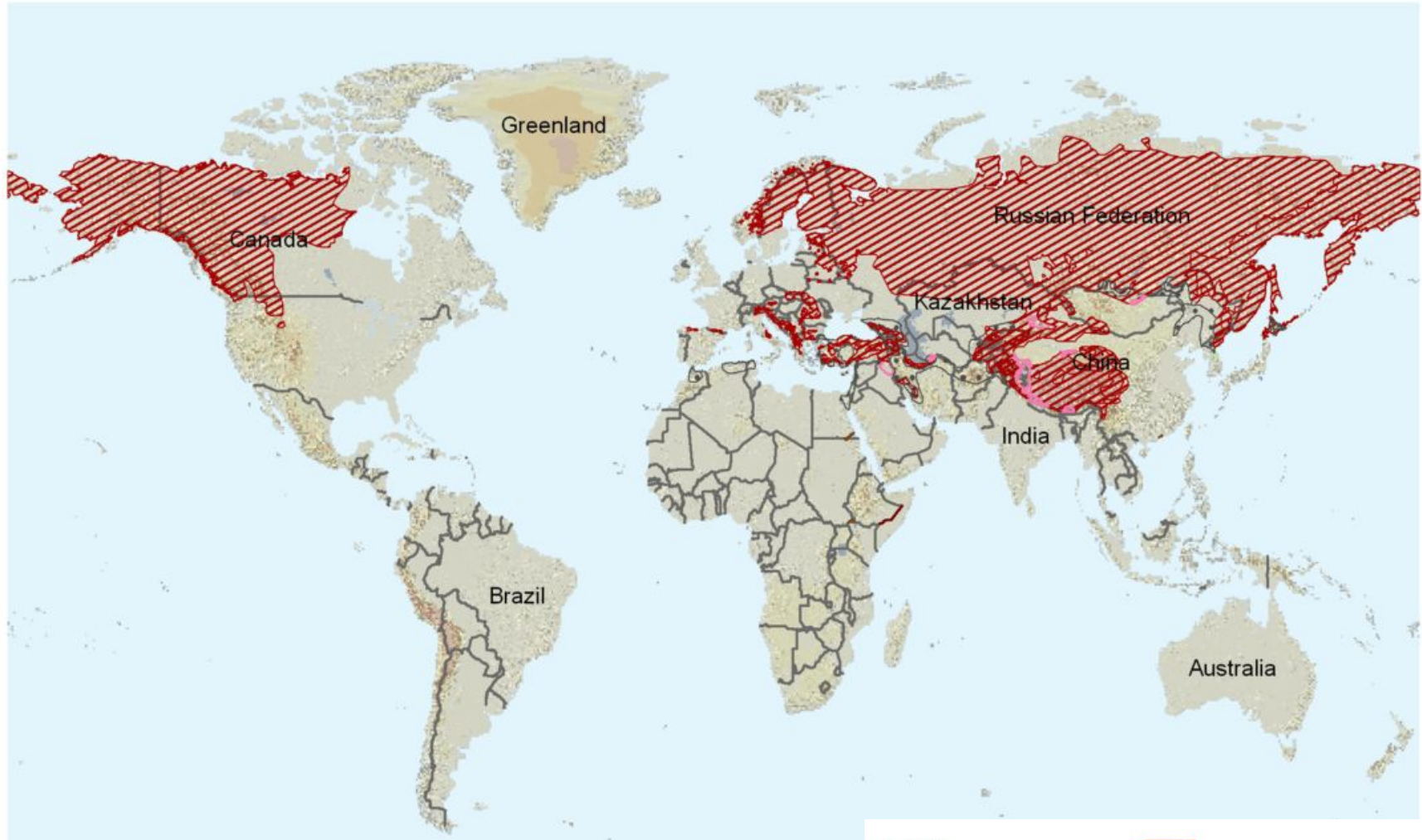


A brown bear is walking across a rocky, grassy hillside. The bear is positioned in the center-right of the frame, facing left. The terrain is covered with green grass and scattered grey rocks of various sizes. The background shows a continuation of the hillside with more rocks and vegetation.

The extraordinary genomic history of the endangered Apennine Brown Bear

Brown bear distribution

Generally holarctic with a more southern range in Asia



No global conservation issue



THE IUCN RED LIST
OF THREATENED SPECIES™



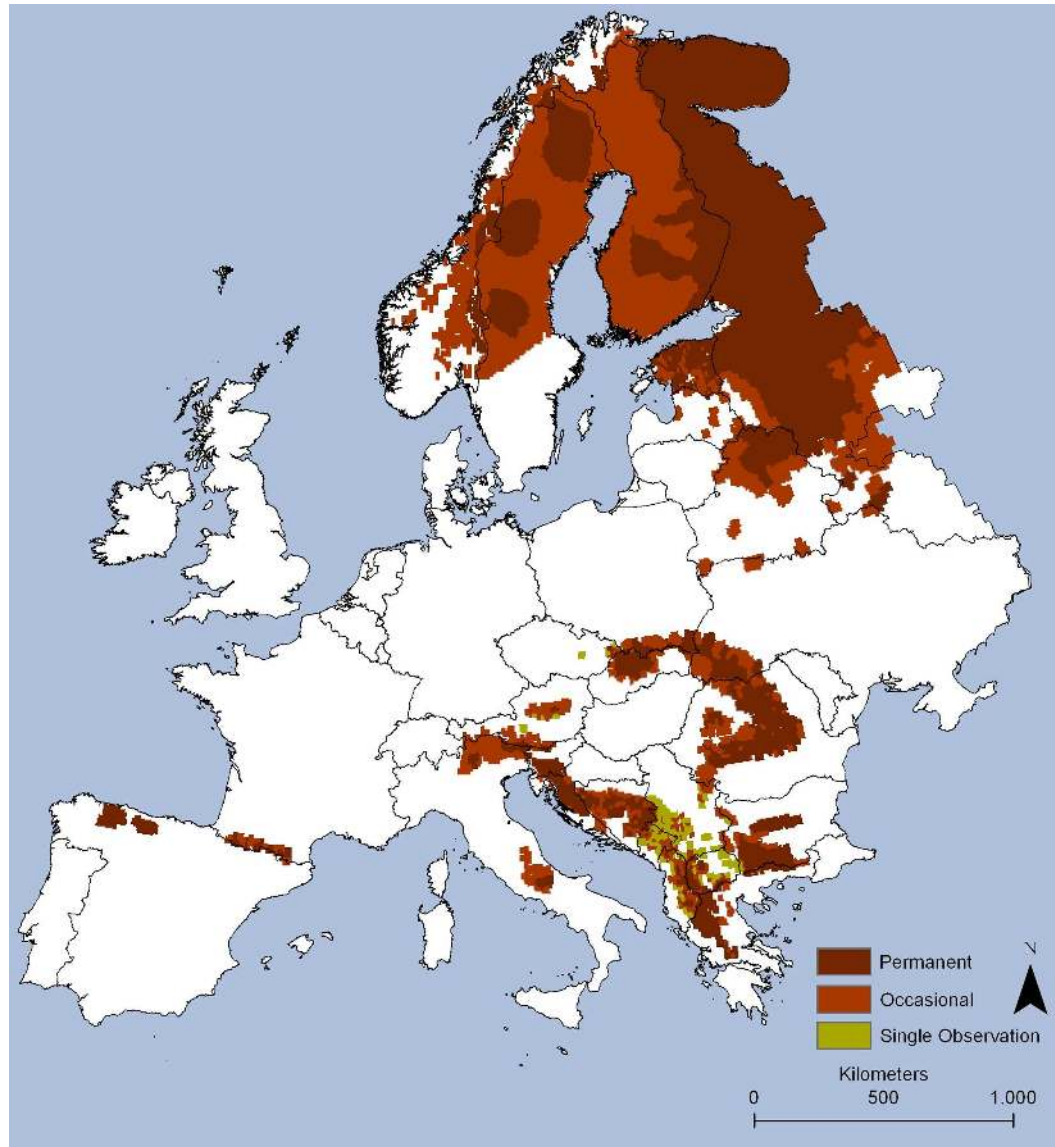
Brown bear distribution

Fragmented in South-Eastern Europe



More or less stable population,
ca. 17,000 individuals

(Status of Large Carnivore
Populations in Europe
2012-2016)



Brown bear distribution

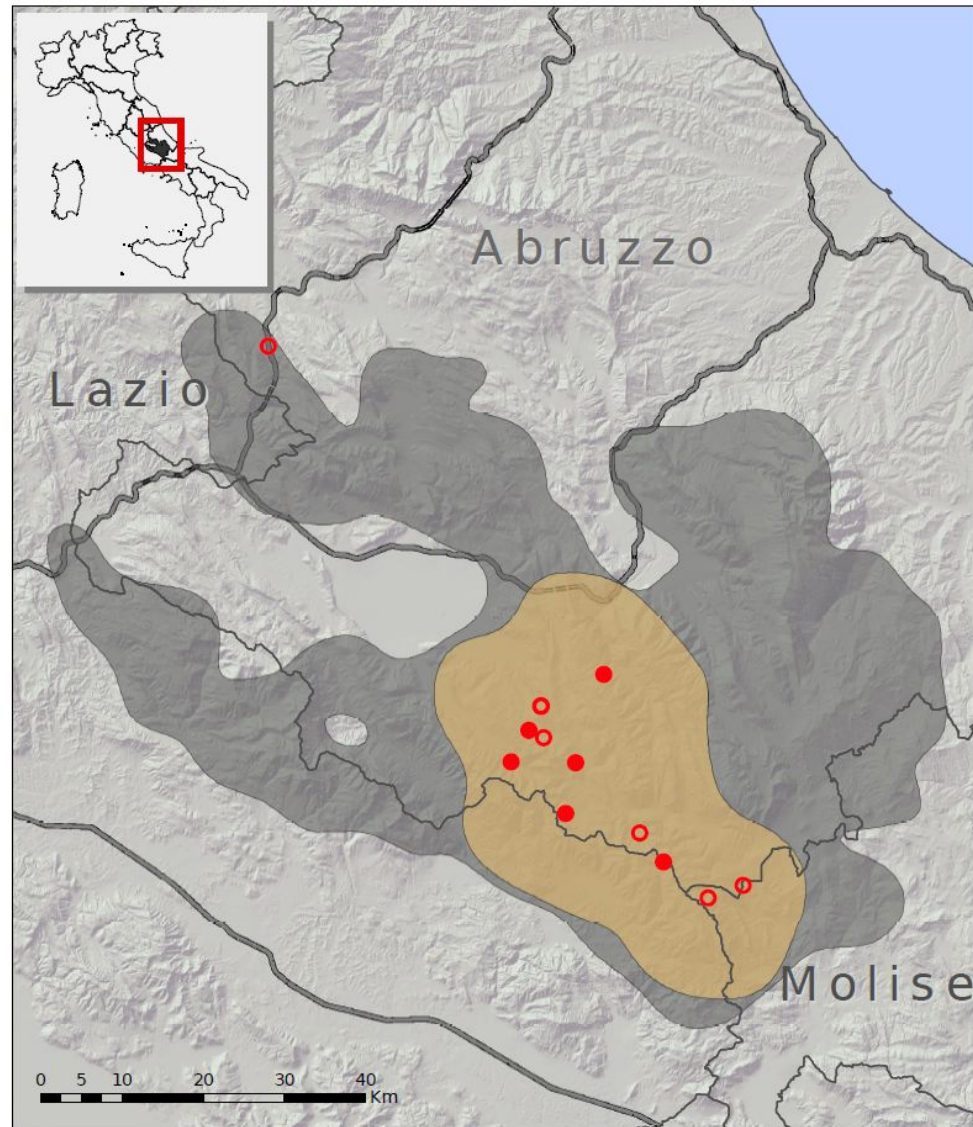
A small and isolated population in Central Italy



Less than 100 individuals since it the population has been checked (last century). Now ca. **50** individuals



National park since 1923



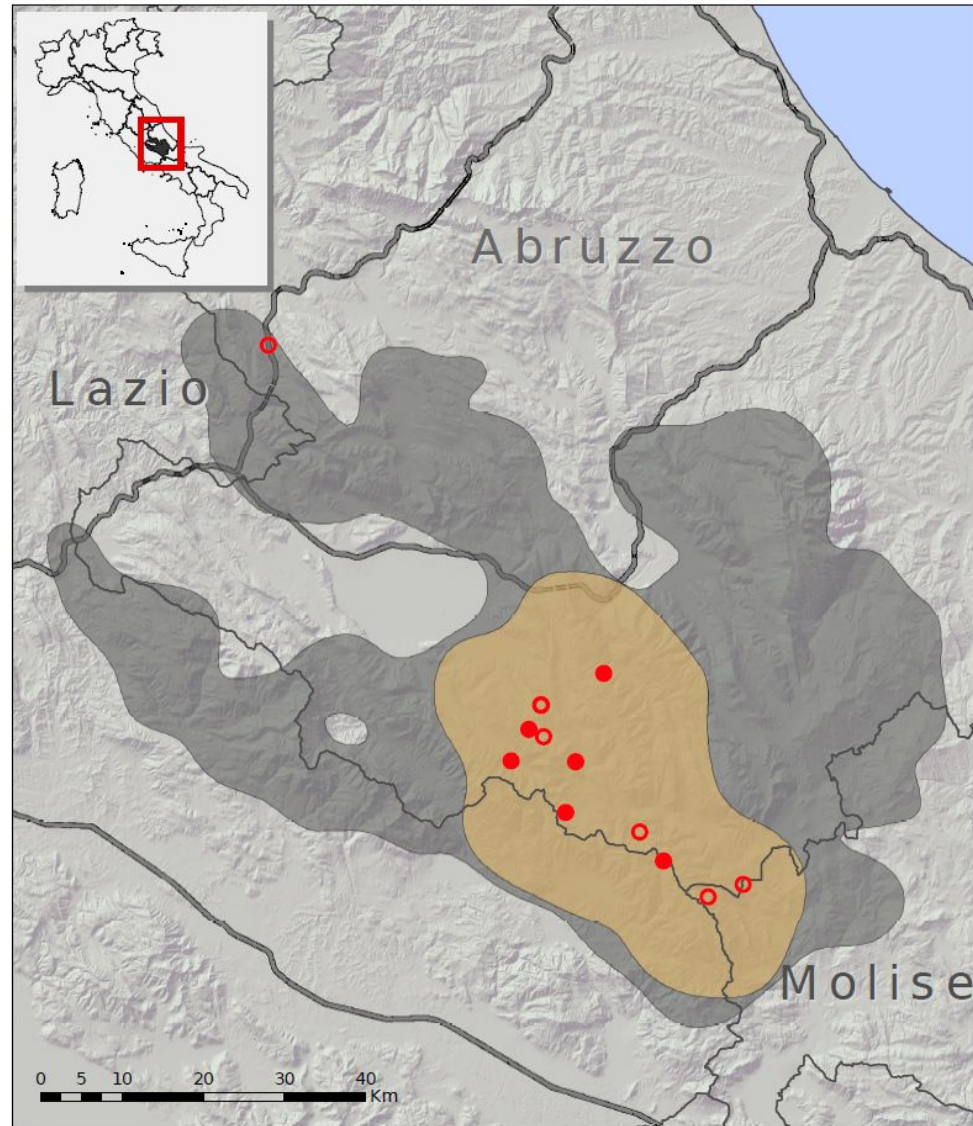
Brown bear distribution

A small and isolated population in Central Italy



Less than 100 individuals since it the population has been checked (last century). Now ca. **50** individuals

How long has this population been so small and isolated?



Brown bear distribution

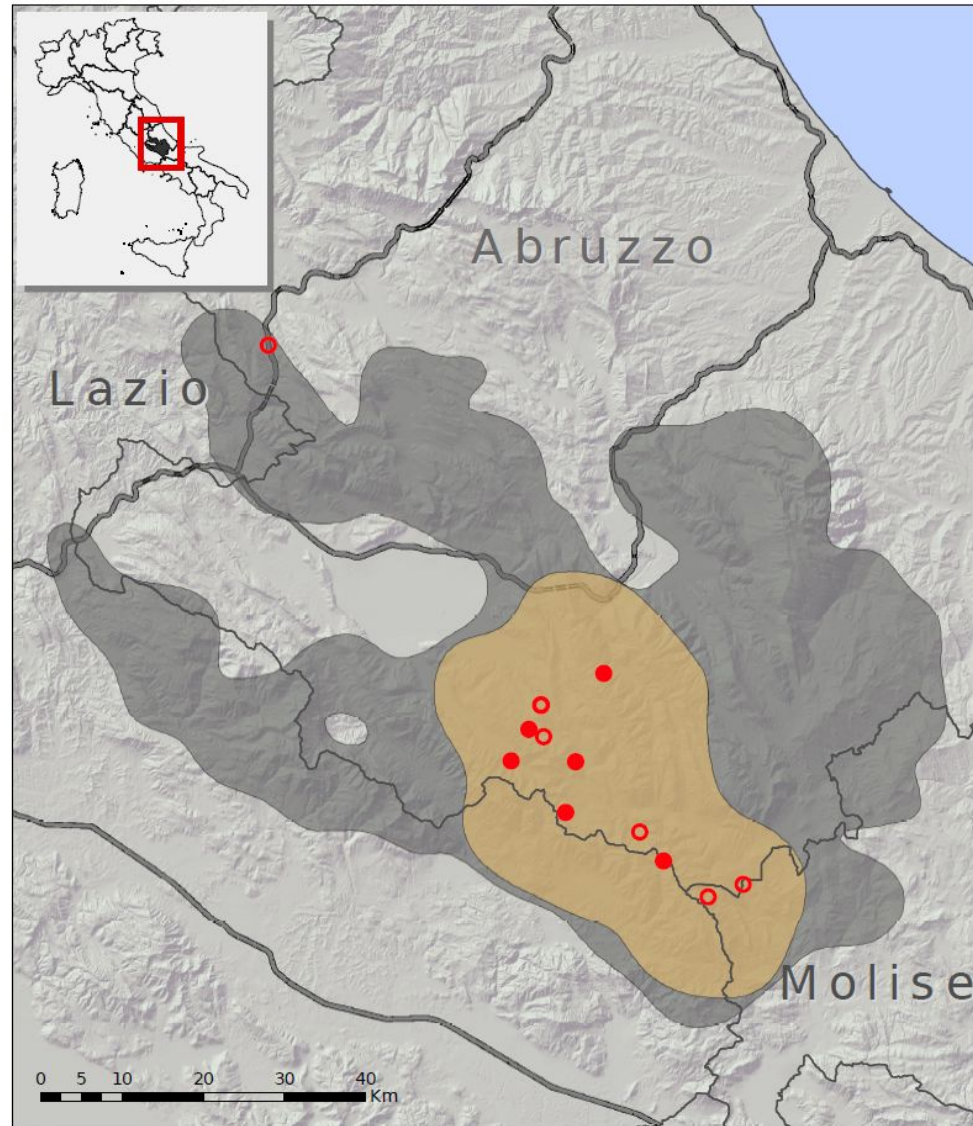
A small and isolated population in Central Italy



Less than 100 individuals since it the population has been checked (last century). Now ca. **50** individuals

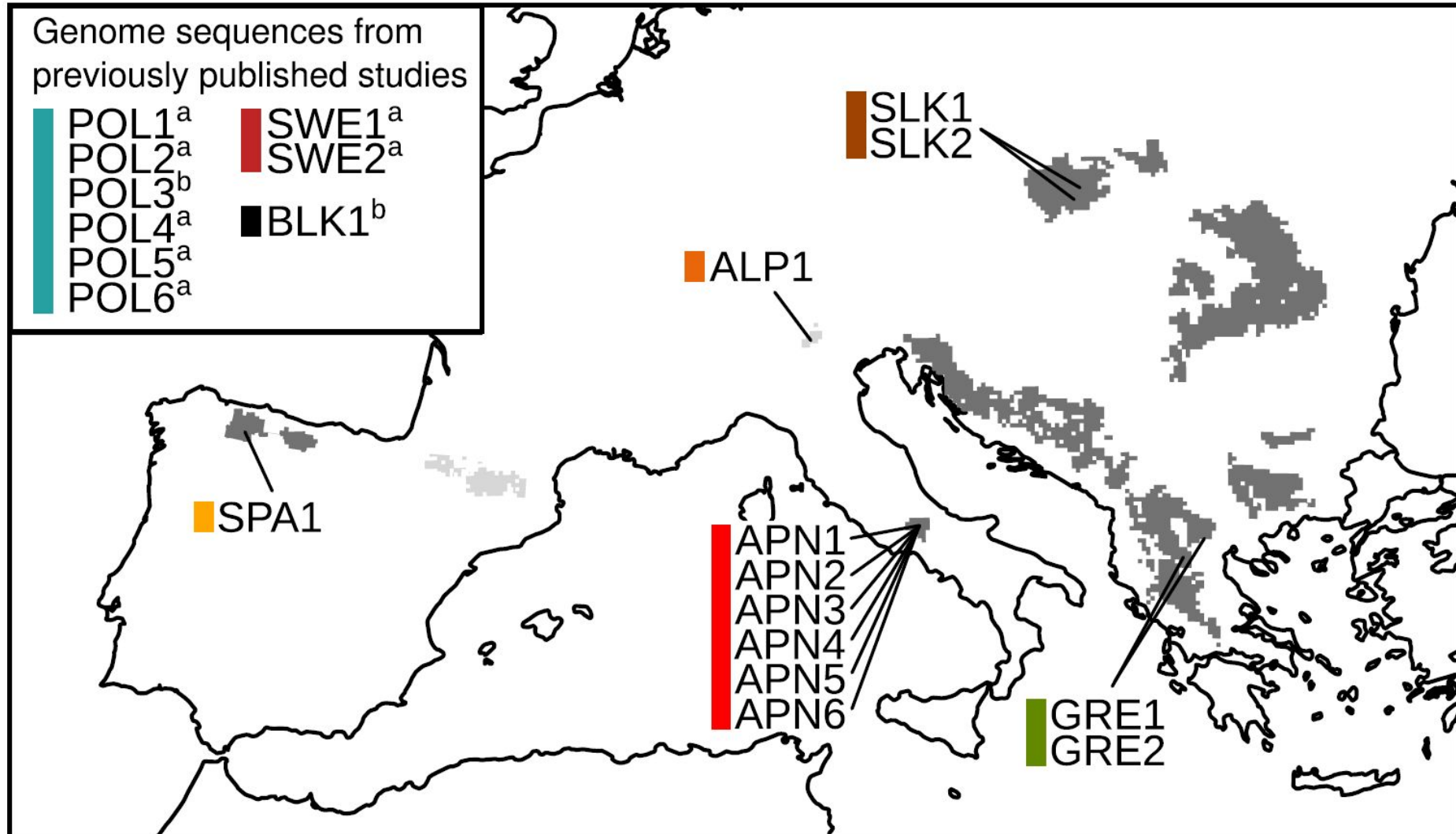
How long has this population been so small and isolated?

6 WGS (●),
12 whole mt genome
sequenced (○)

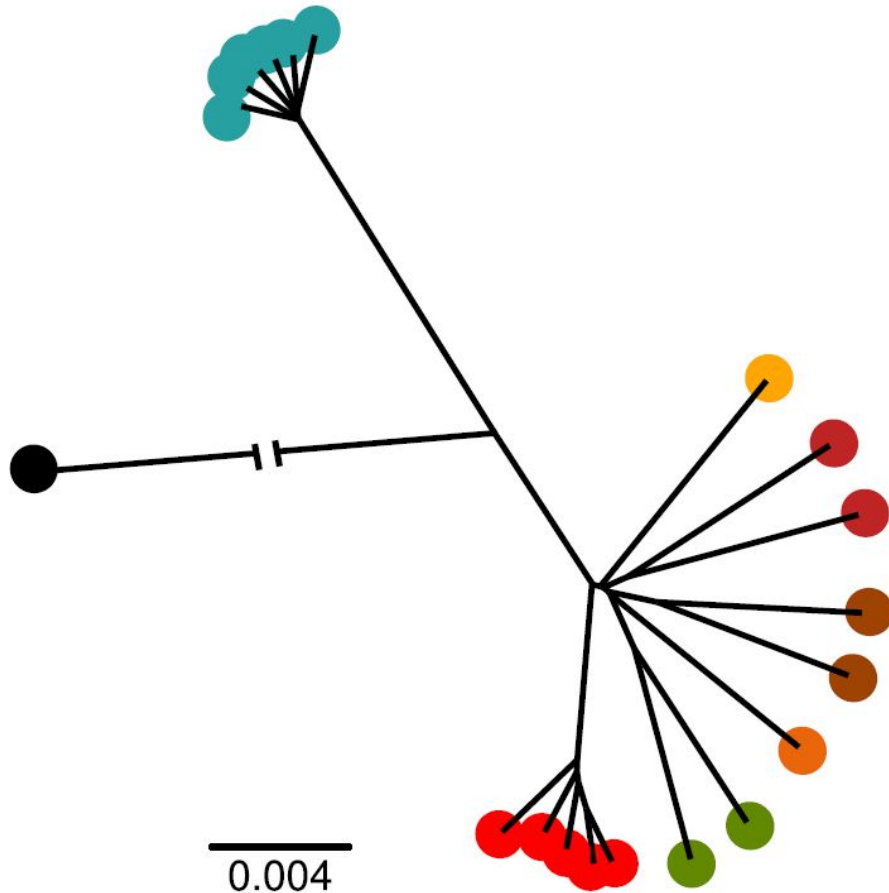


Whole-genome data

A few other samples from other European populations ...

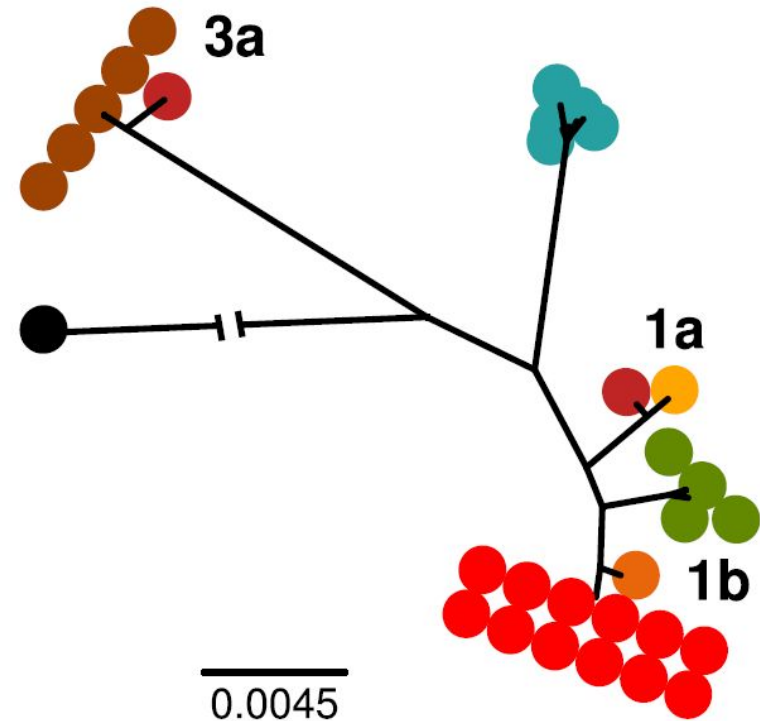


1,842,042,551 nuclear bp



ANGSD, ngsDist, nj from ape R package

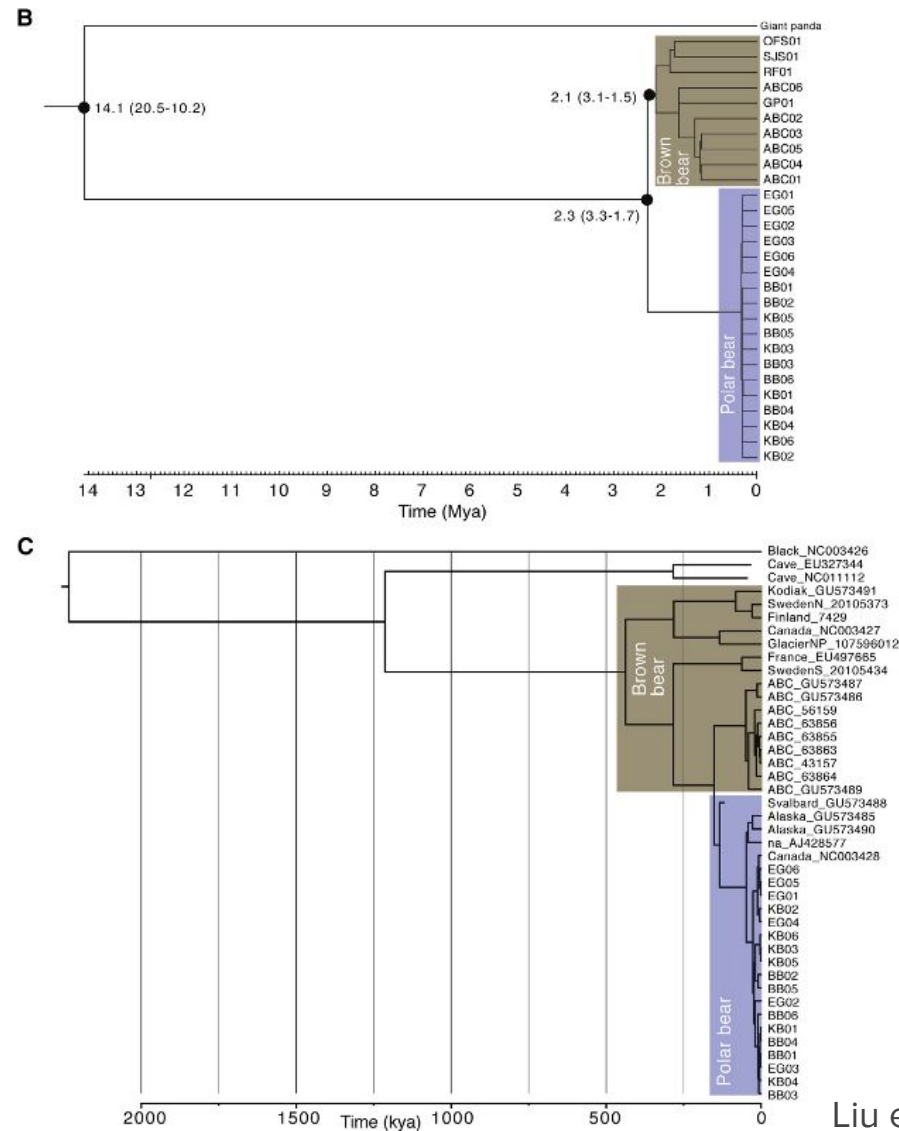
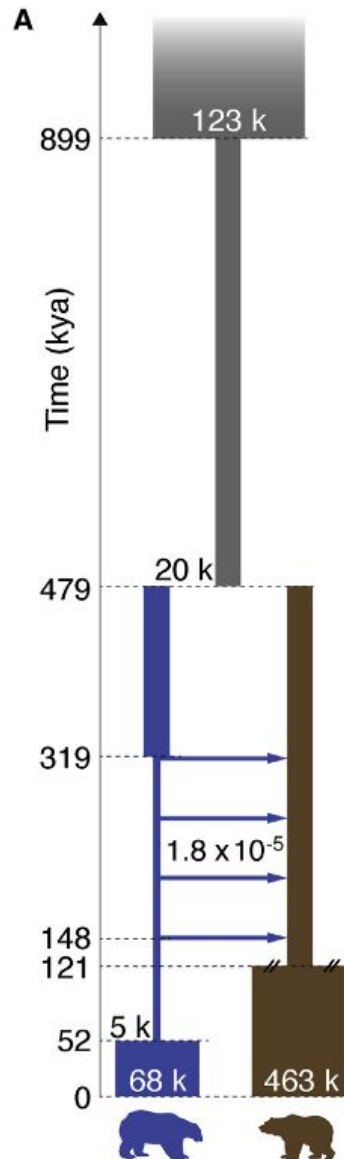
16,485 mitochondrial bp



Geneious, MAFFT, genetic distance (HKY), NJ

Structure of genomic diversity

Whole-genome and mitochondrion contrasting histories

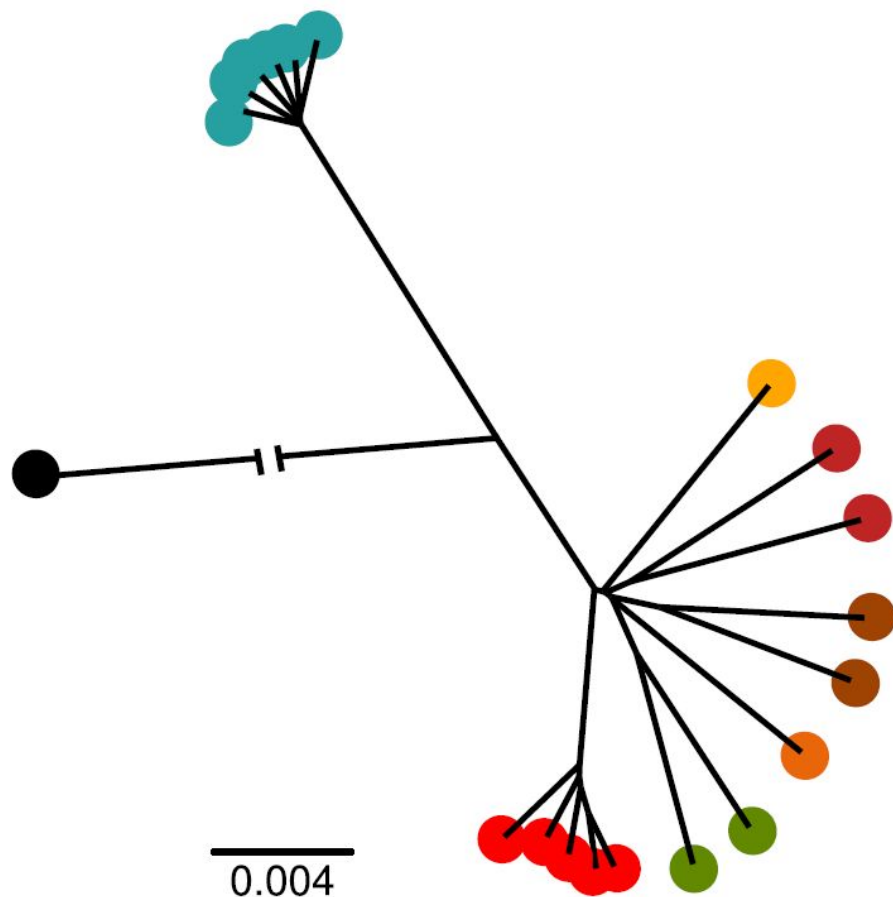


Structure of genomic diversity

Y-chr agrees with whole-genome: sex-biased dispersal?

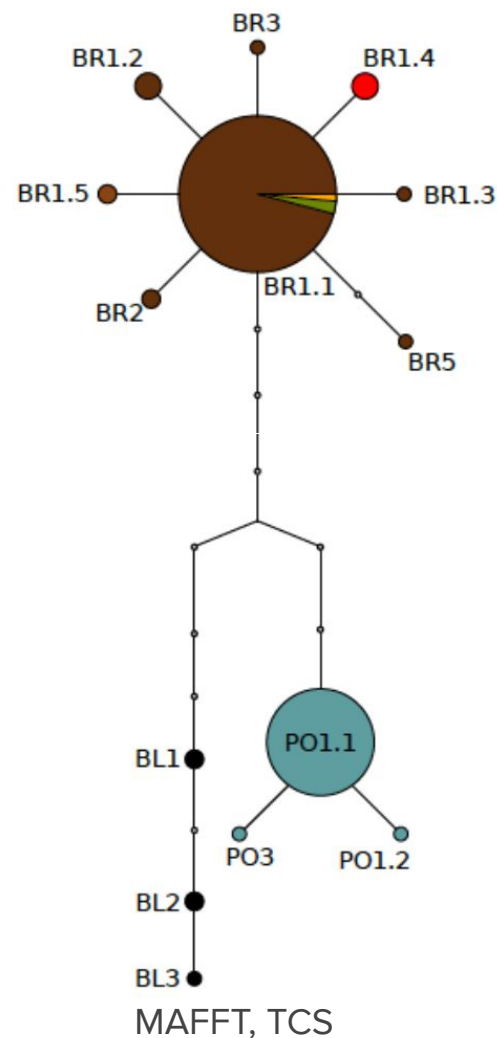


1,842,042,551 nuclear bp



ANGSD, ngsDist, nj from ape R package

5.3 Kb Y-chromosome

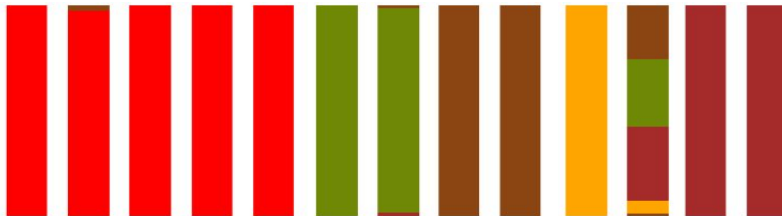


Structure of genomic diversity

Robust to decay of linkage disequilibrium



7,971 SNPs, 150bp thinning

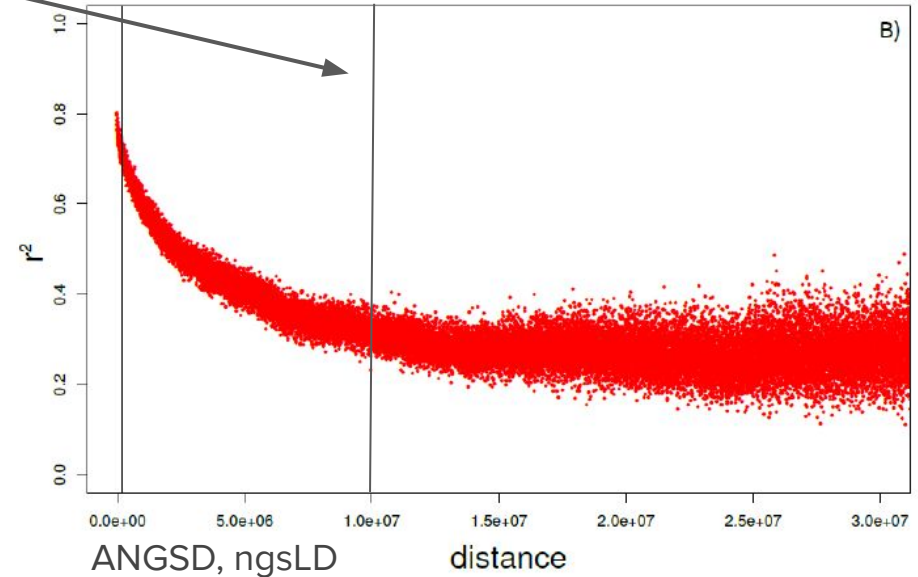
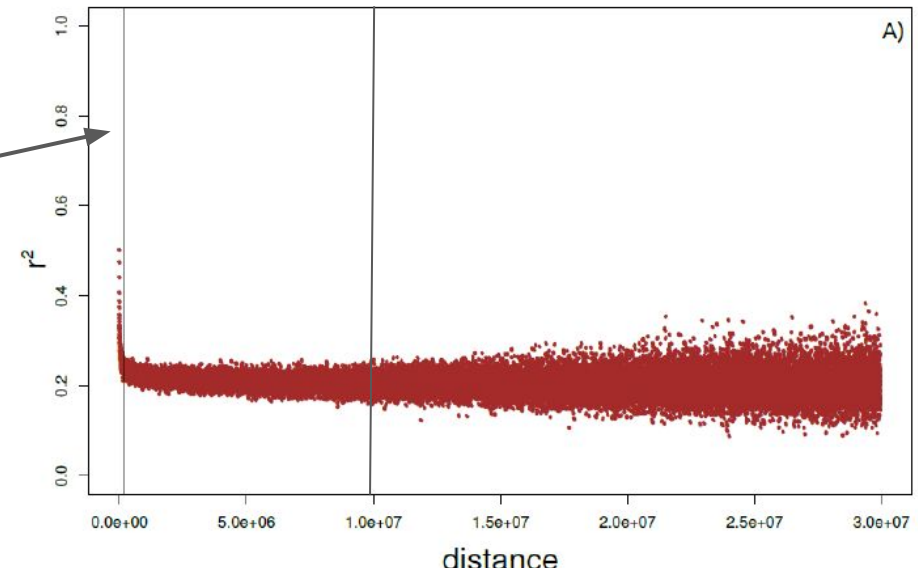


100 SNPs, 10Mb thinning



APN2
APN3
APN4
APN5
APN6
GRE1
GRE2
SLK1
SLK2
SPA1
ALP1
SWE1
SWE2

STRUCTURE, non-cds, non-rep



ANGSD, ngsLD

distance

Genomic diversity - average

Low but the polar bear is worse



High coverage:

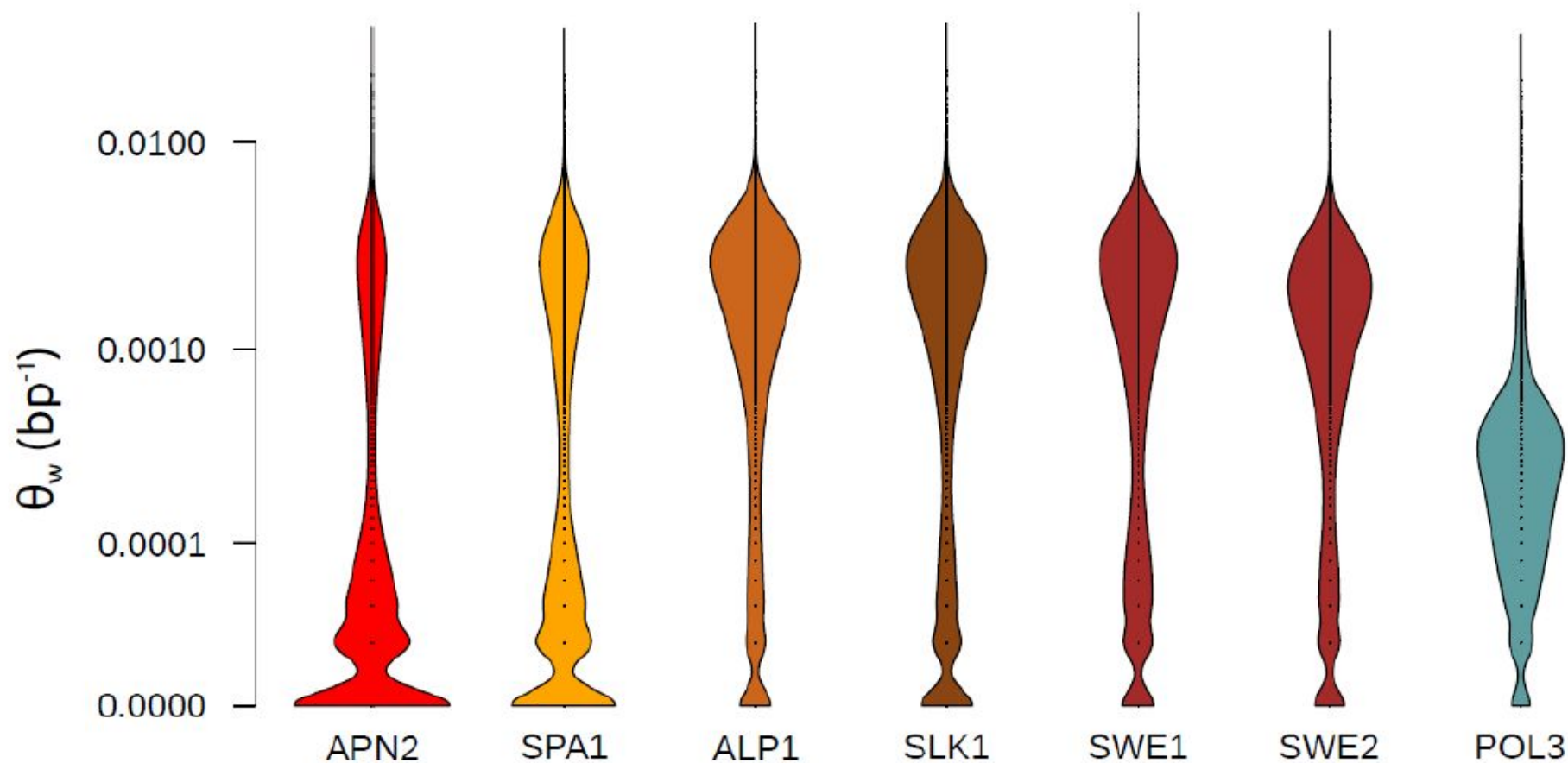
vcftools, SNP density, 50kb windows
(then downsampled and analyzed as
low coverage)

Low coverage:

ANGSD, realSFS, do_theta,
50kb windows

Genomic diversity - distribution

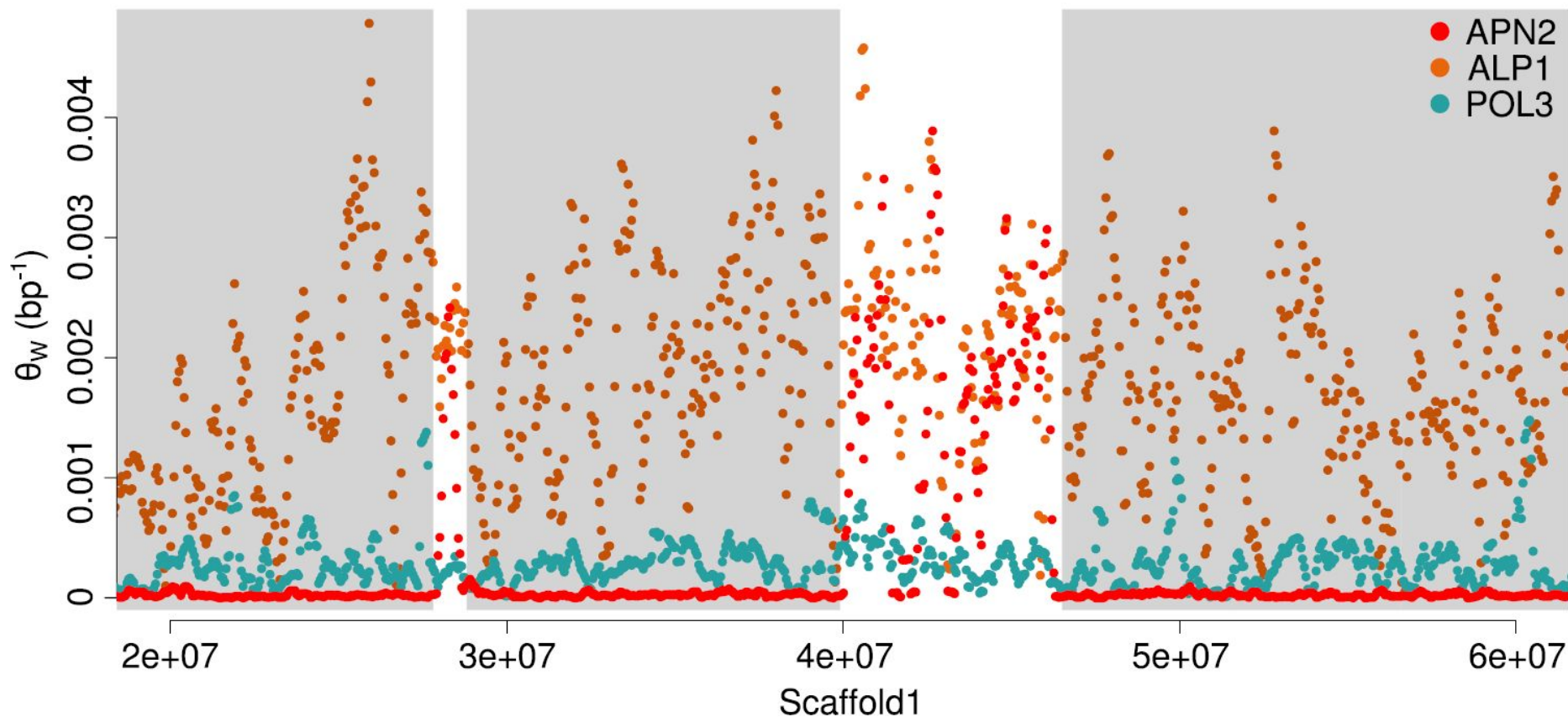
Low but the polar bear is worse...maybe not!



vcftools, SNP density in 50kb windows

Genomic diversity - distribution

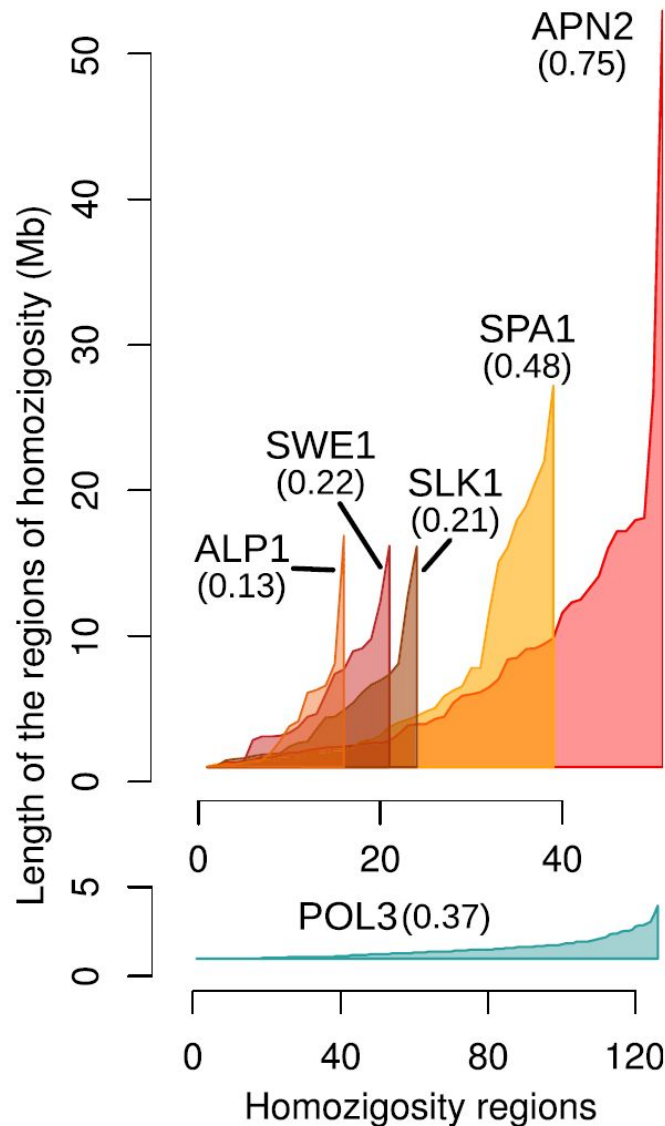
Long stretches with no diversity in the Apennine bear



vcftools, SNP density in 50kb windows

Inbreeding estimates

Much higher than any other European bear

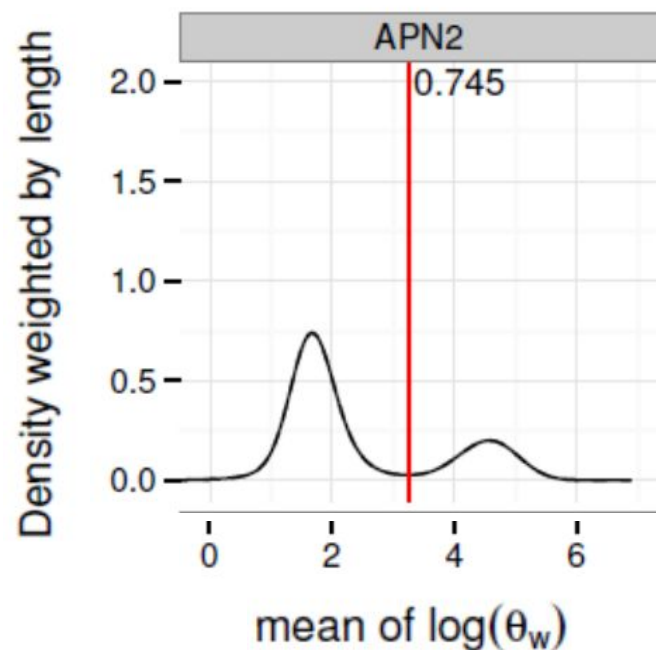
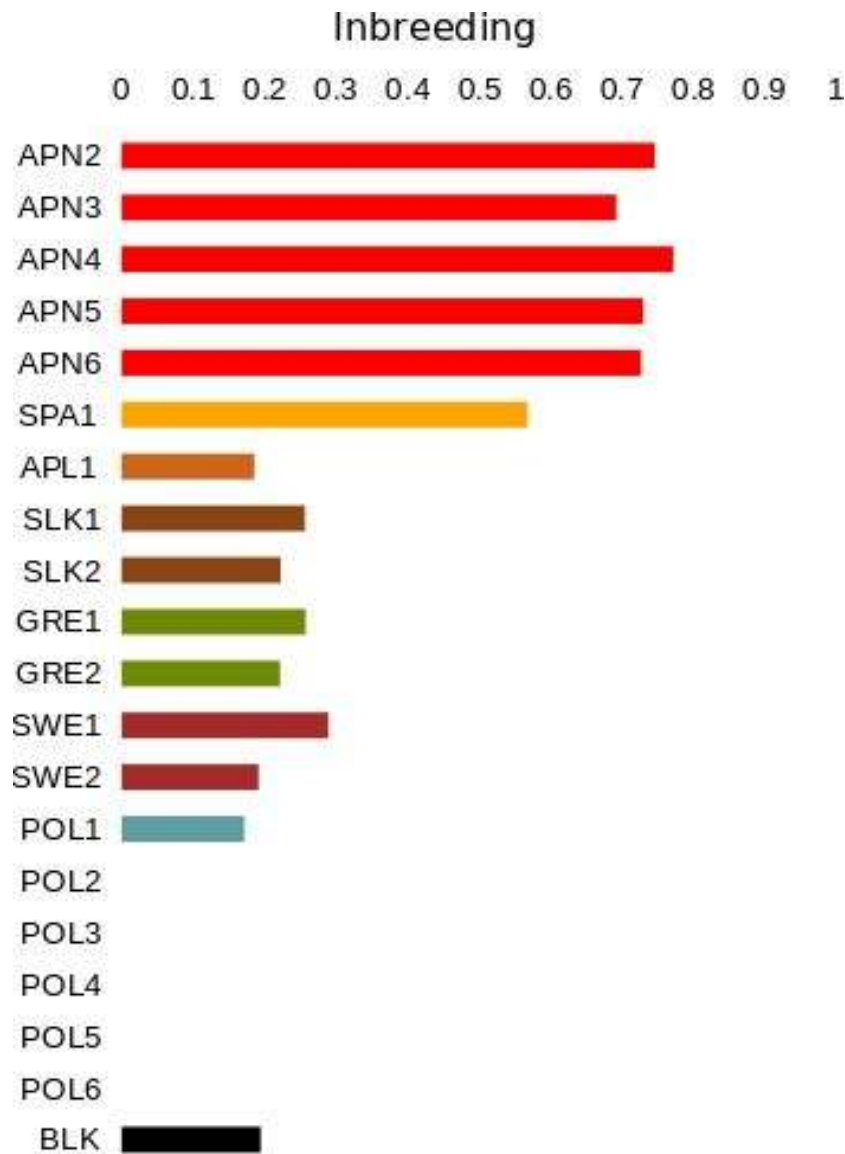


Regions longer than 1Mb with less than 25 segregating sites in the 13 longest scaffolds, ca. 0.5 Gb - sorted

(percent of the analyzed scaffolds)

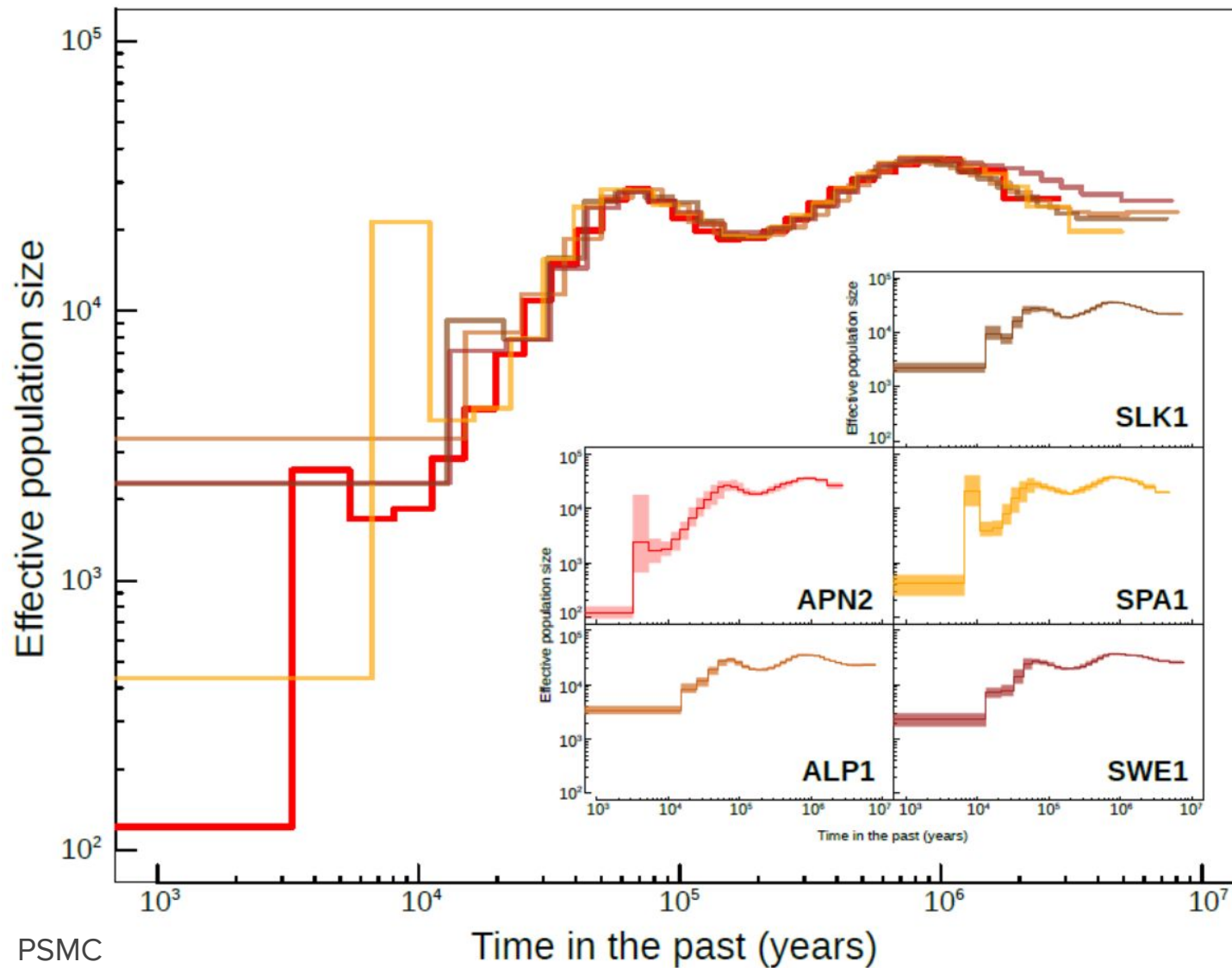
Inbreeding estimates

Much higher than any other European bear



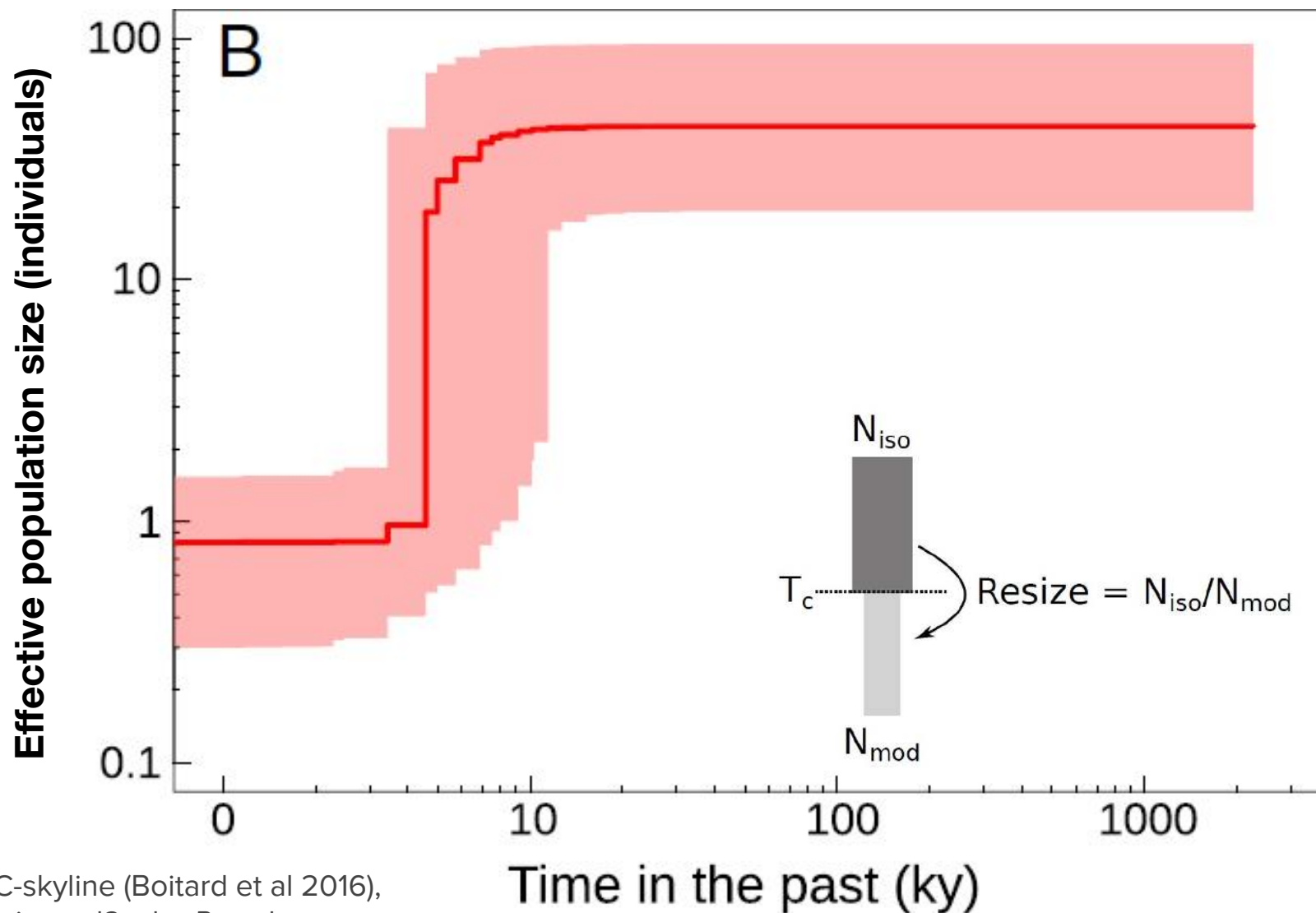
Based on the proportion of the genome segments that are mostly homozygous (Prüfer et al 2014)

Apennine population declined more than other Eu pops



Past demography

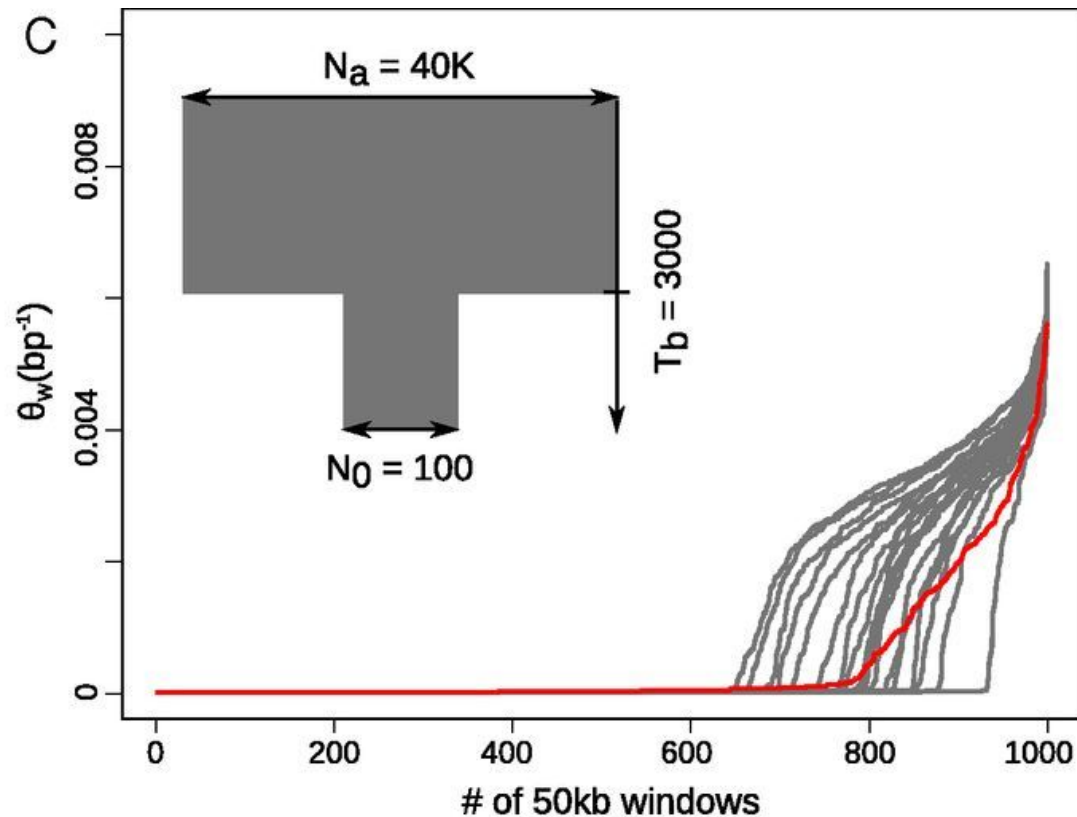
...and to very low N_e



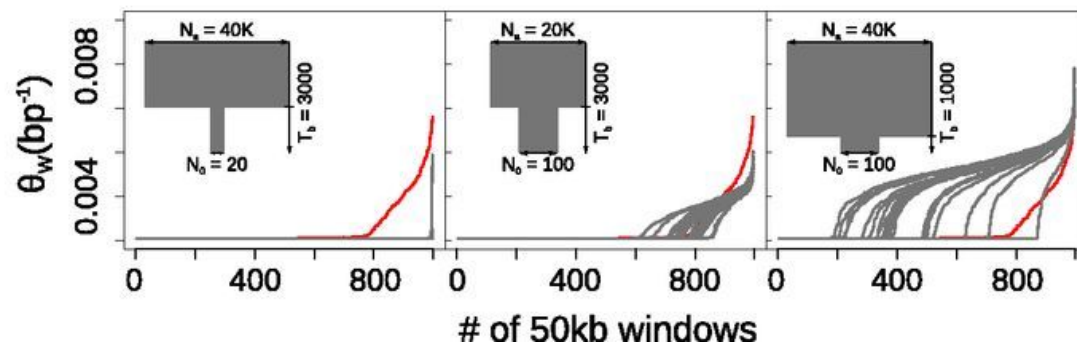
ABC-skyline (Boitard et al 2016),
fastsimcoal2, abc R package

Past demography

Decline 3,000 years ago to about 100 individuals

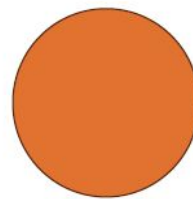
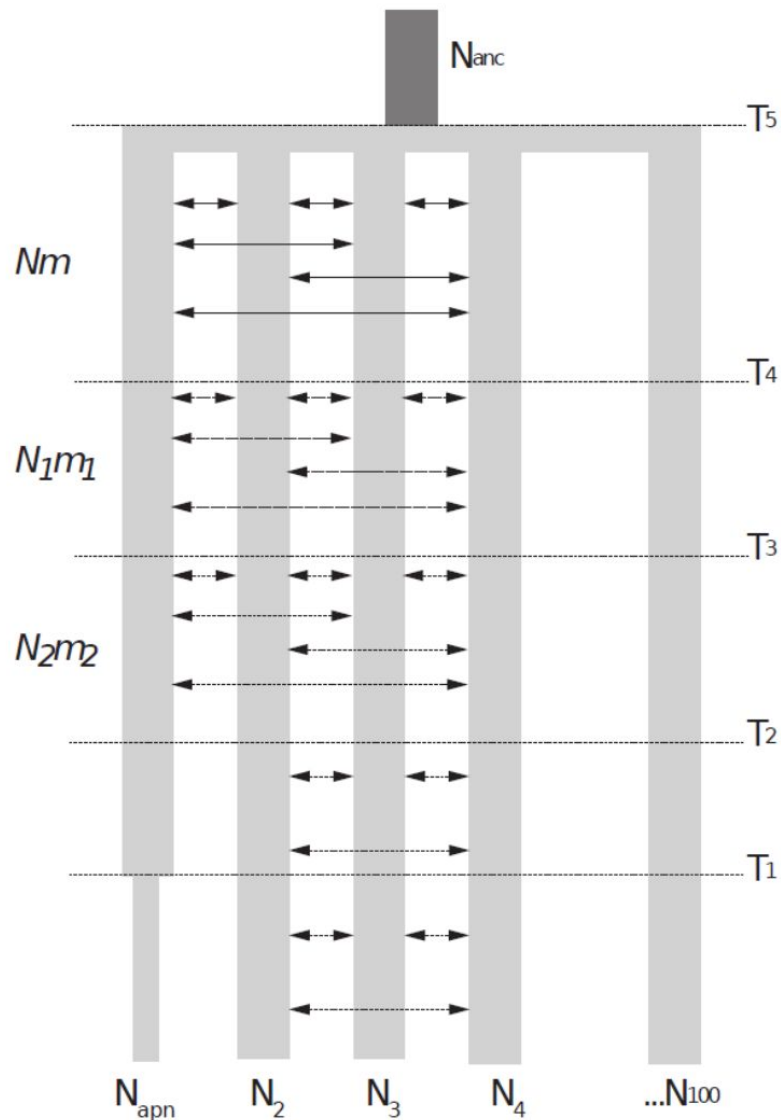


Simulation of expected pattern of homozygosity regions using *ms*



Past demography

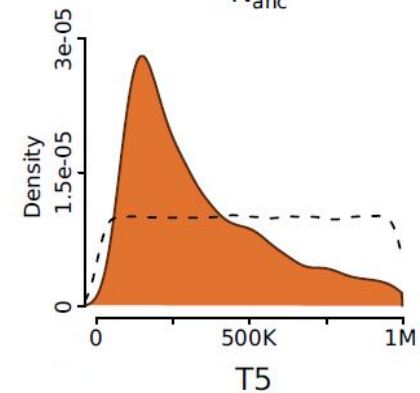
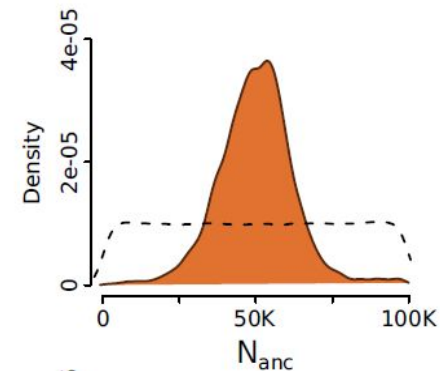
Fragmentation of a large European population



Ancestral population



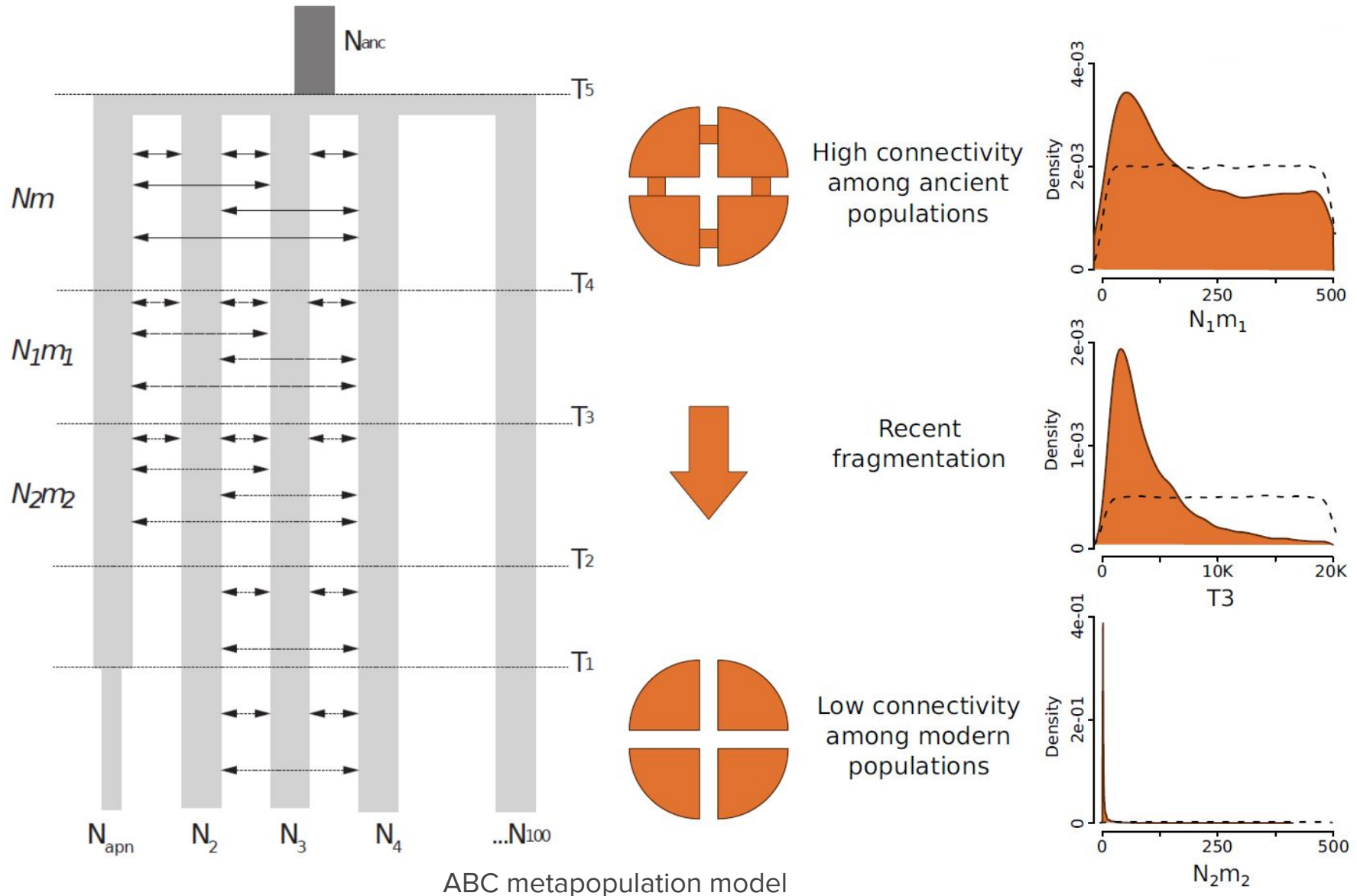
Ancient European spread



ABC metapopulation model

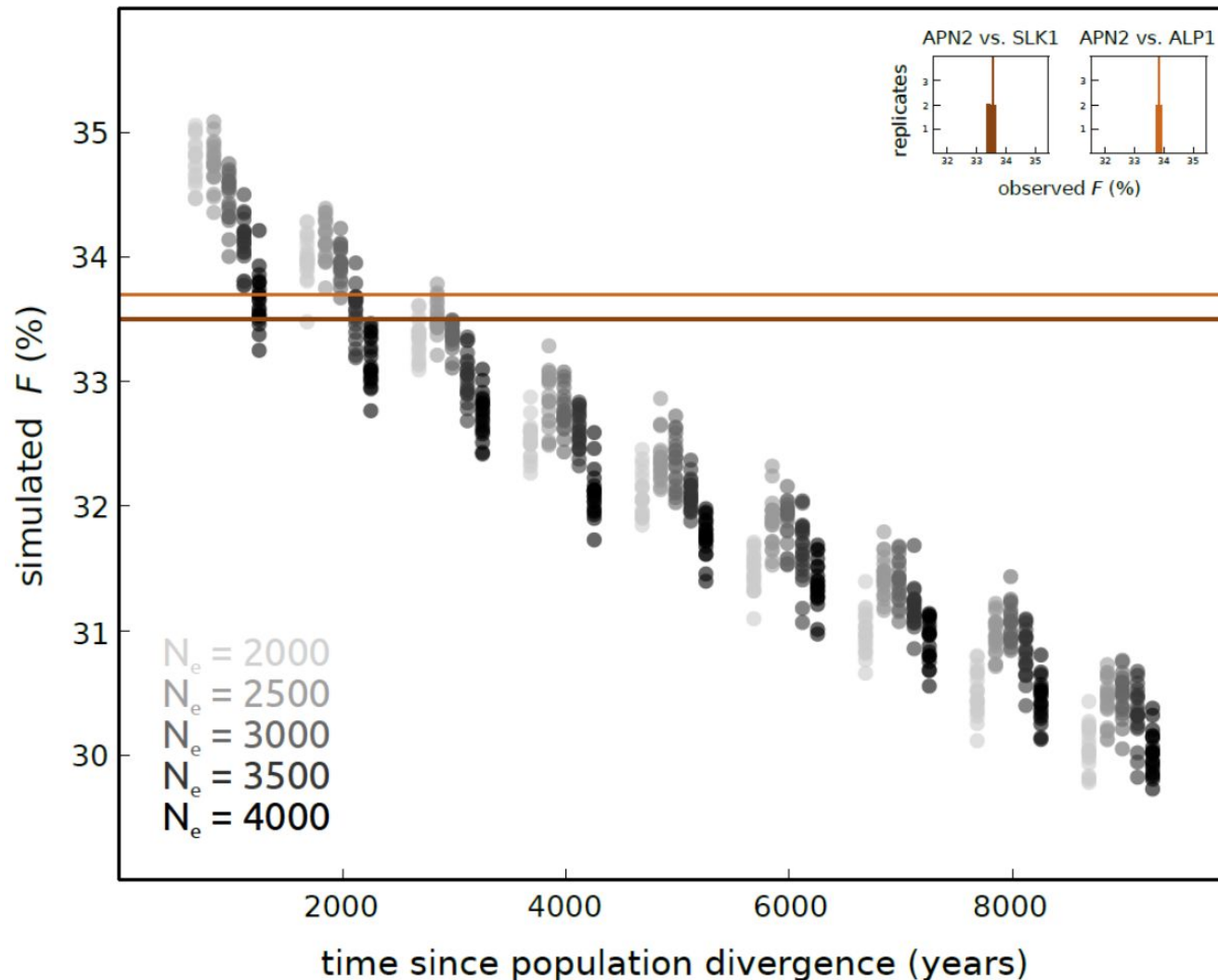
Past demography

Fragmentation of a large European population



Past demography

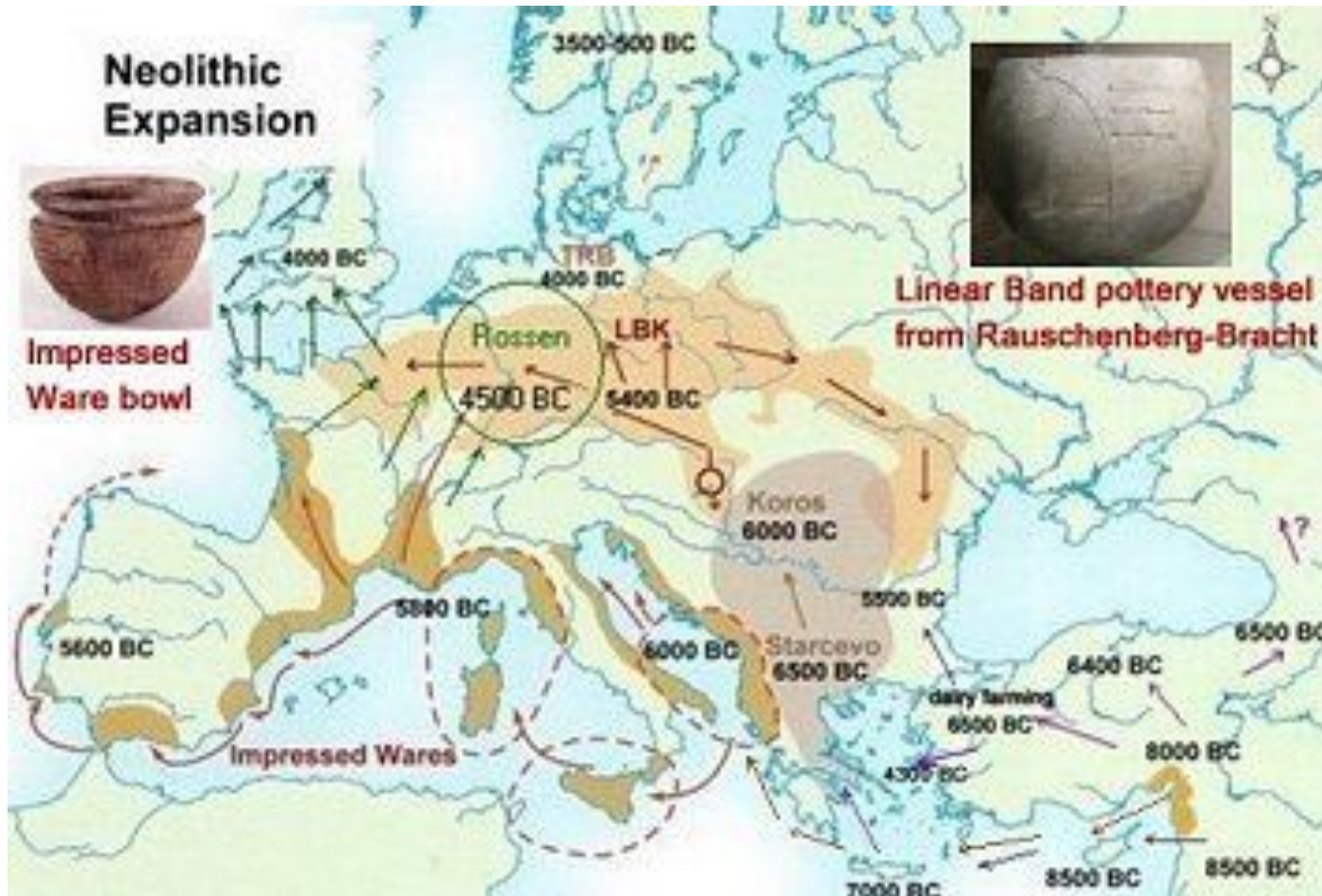
Apennine bears have been isolated since then



F statistic (Green et al 2010), simulation with ms

Past demography

Expansion of Neolithic farming in Europe burned forests



Why is this population still there?
Given such high extinction probability



Selective processes

Retention of non-synonymous polymorphisms



Reference genome	Two APN		SLK1+SLK2		GRE1+GRE2	
	P_{n-syn}	P_{syn}	P_{n-syn}	P_{syn}	P_{n-syn}	P_{syn}
ALP1	0.34	0.25	0.55	0.54	0.54	0.52
SLK1	0.36	0.26	-	-	0.54	0.51

McDonald-Kreitman like test

Fraction of heterozygous sites in a European bear that are still heterozygote in the Apennine bear, for synonymous and nonsynonymous sites

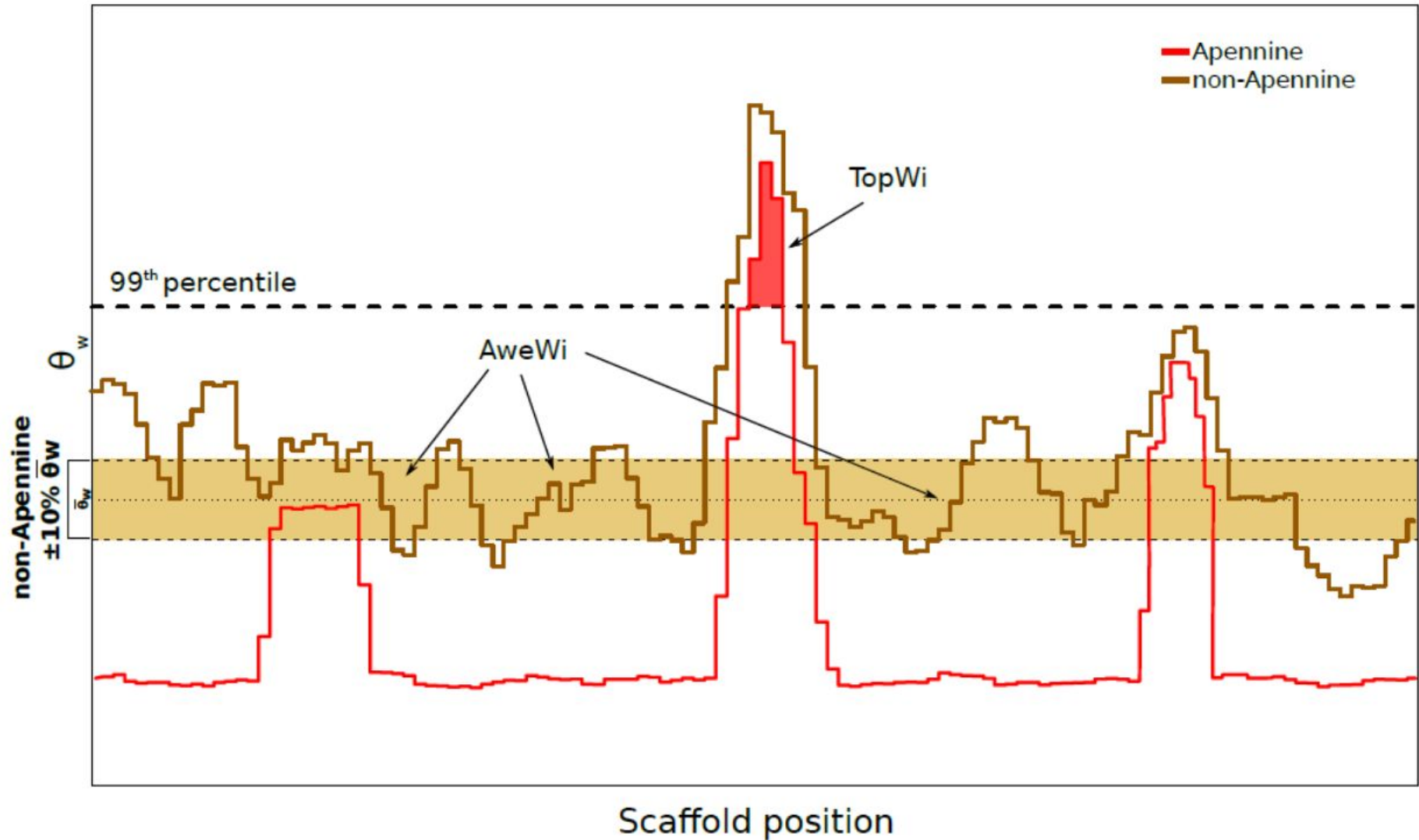
$P_{n-syn} = P_{syn}$: drift

$P_{n-syn} > P_{syn}$: **balancing selection**

$P_{n-syn} < P_{syn}$: directional selection

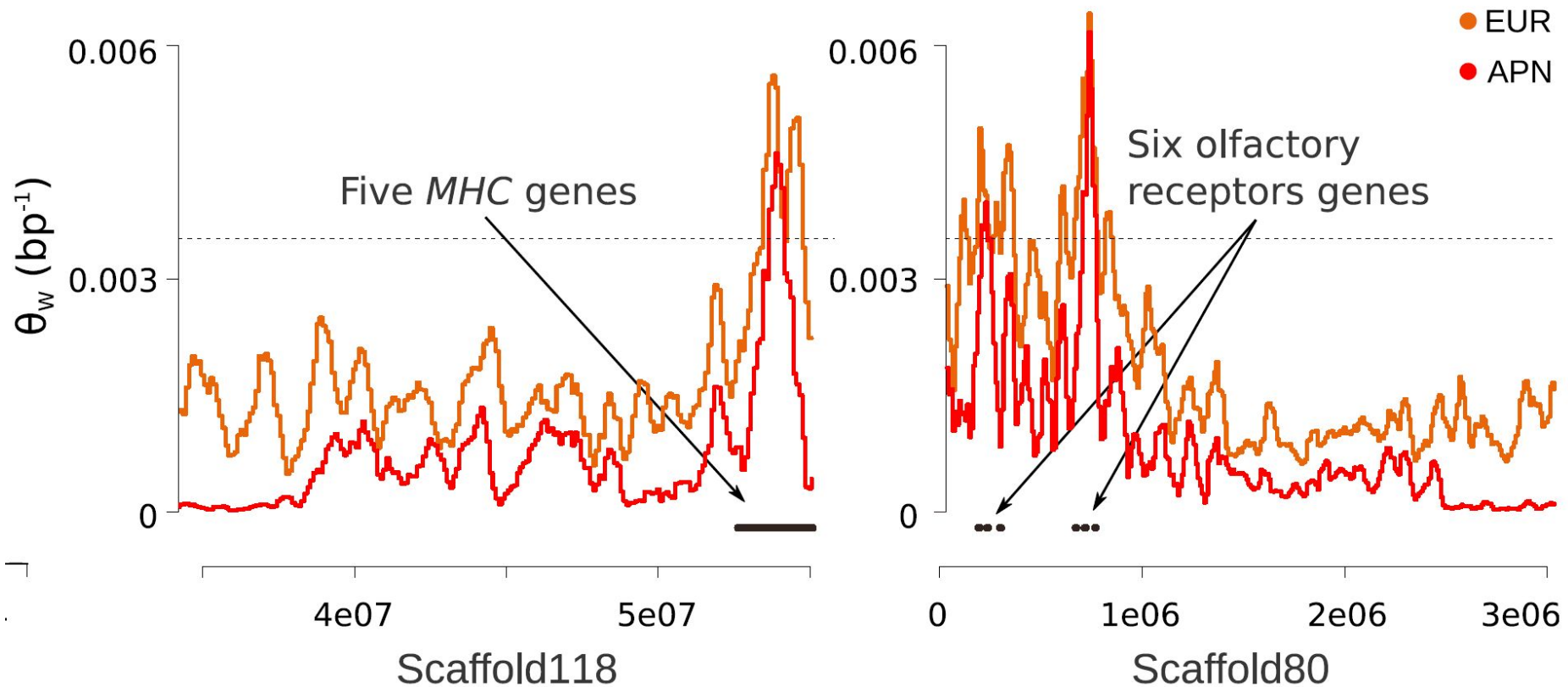
Selective processes: balancing

Retention of high diversity regions (HDR)



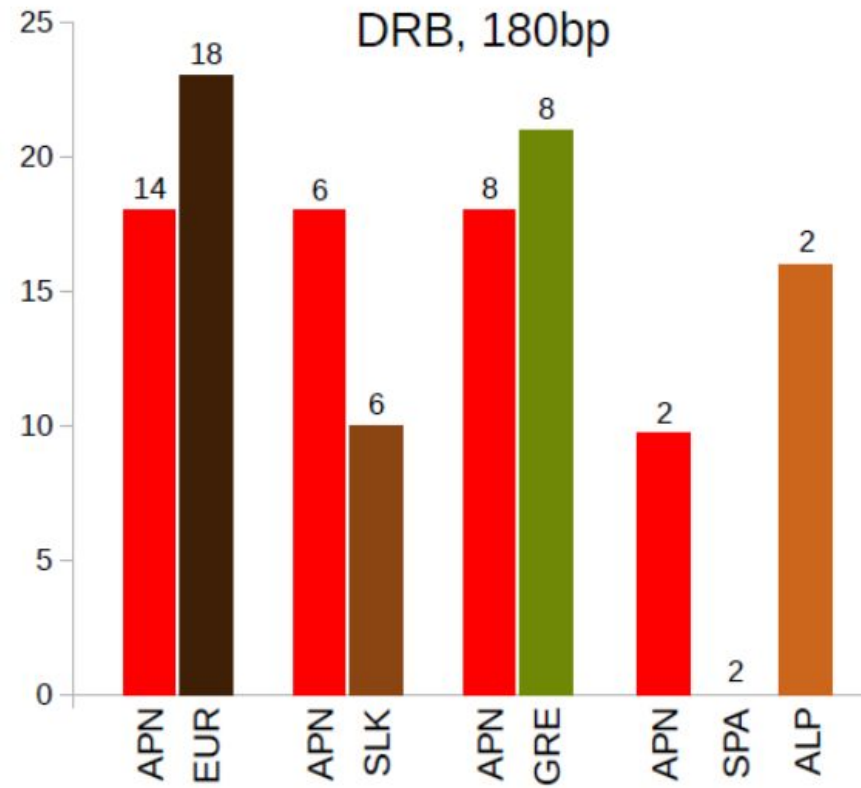
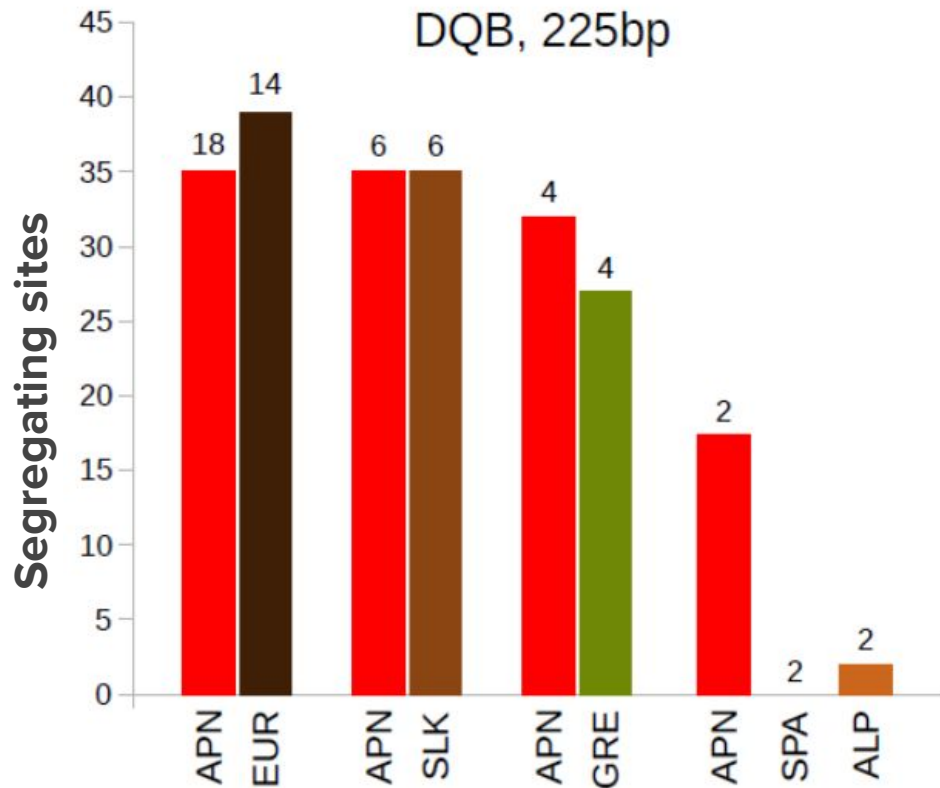
Selective processes: balancing

HDR are enriched in immune and olfactory genes



Selective processes: balancing

MHC genes are as diverse as in the rest of Europe



Sanger-sequencing of two *MHC* class II loci in additional individuals from Apennine and the rest of Europe

Genetic load

Accumulation of deleterious effects

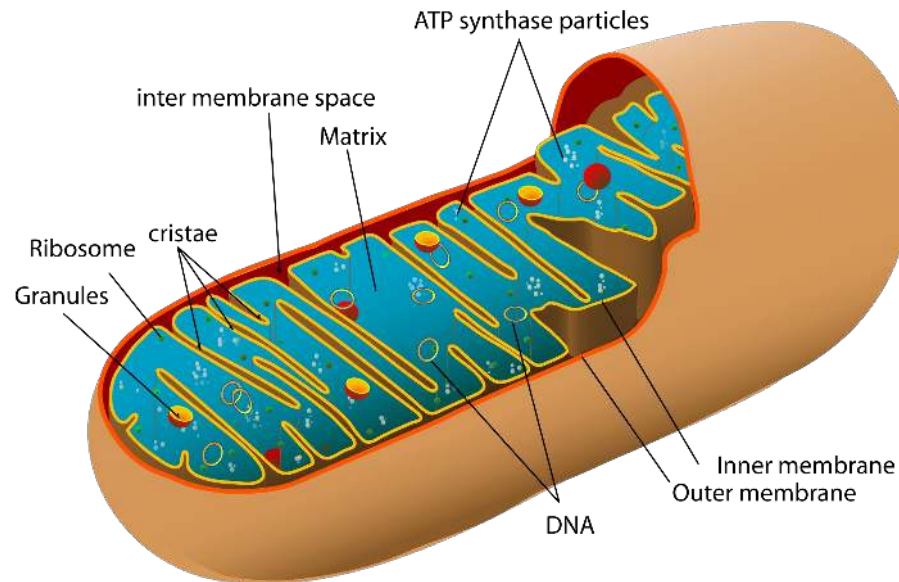


Predicted deleterious substitution fixed

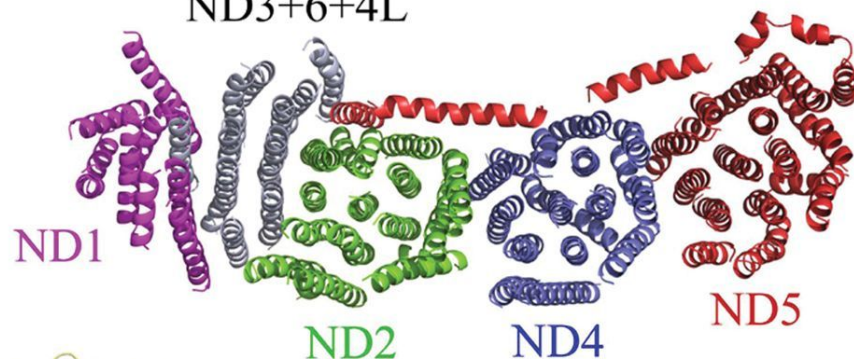
in the Apennine bears (by Panther + Polyphen): 40 + 4 stop codon

None in the other European bears!

5 in the **mt ND5** of which the most deleterious is not found elsewhere



ND3+6+4L



Gene	Position	#APNs (out of 6)	#NonAPNs (out of 45)	Score
ND5	G526E	6	0	-4.28
ND5	P447S	6	5	-3.59
ND5	T555A	6	3	-2.52

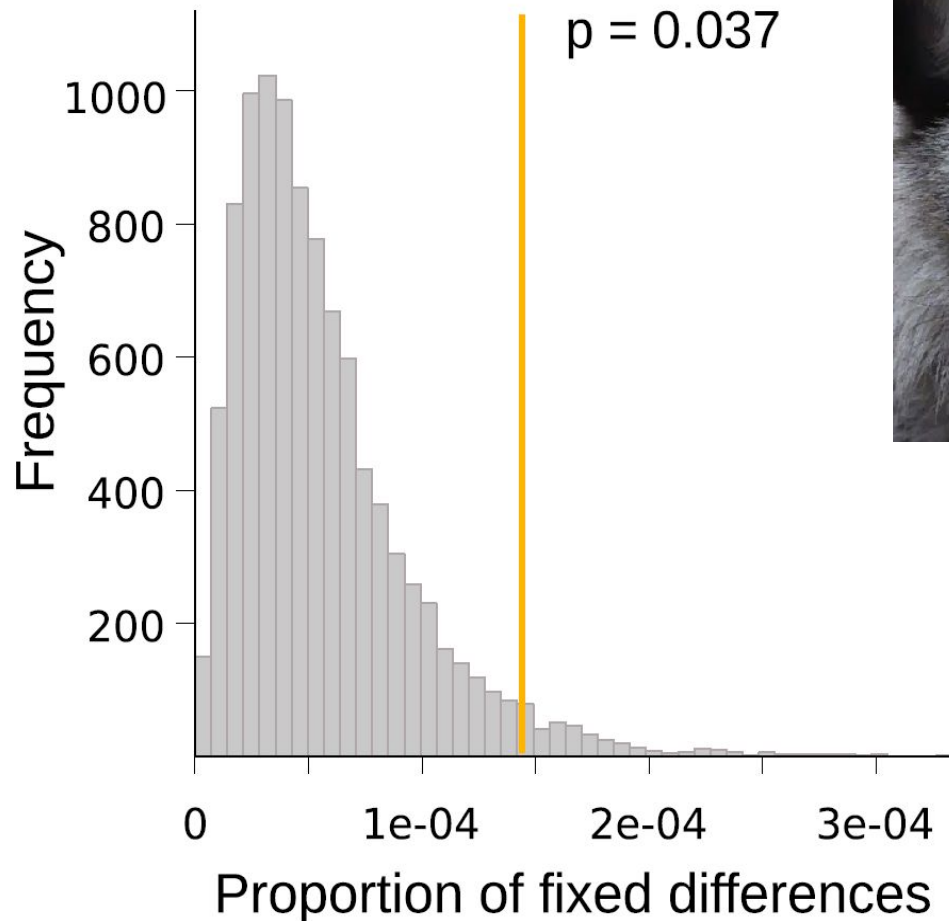
Genetic load in a specific group of genes

Are Apennine bears less aggressive?



Fixed differences in 22 “tameness” genes

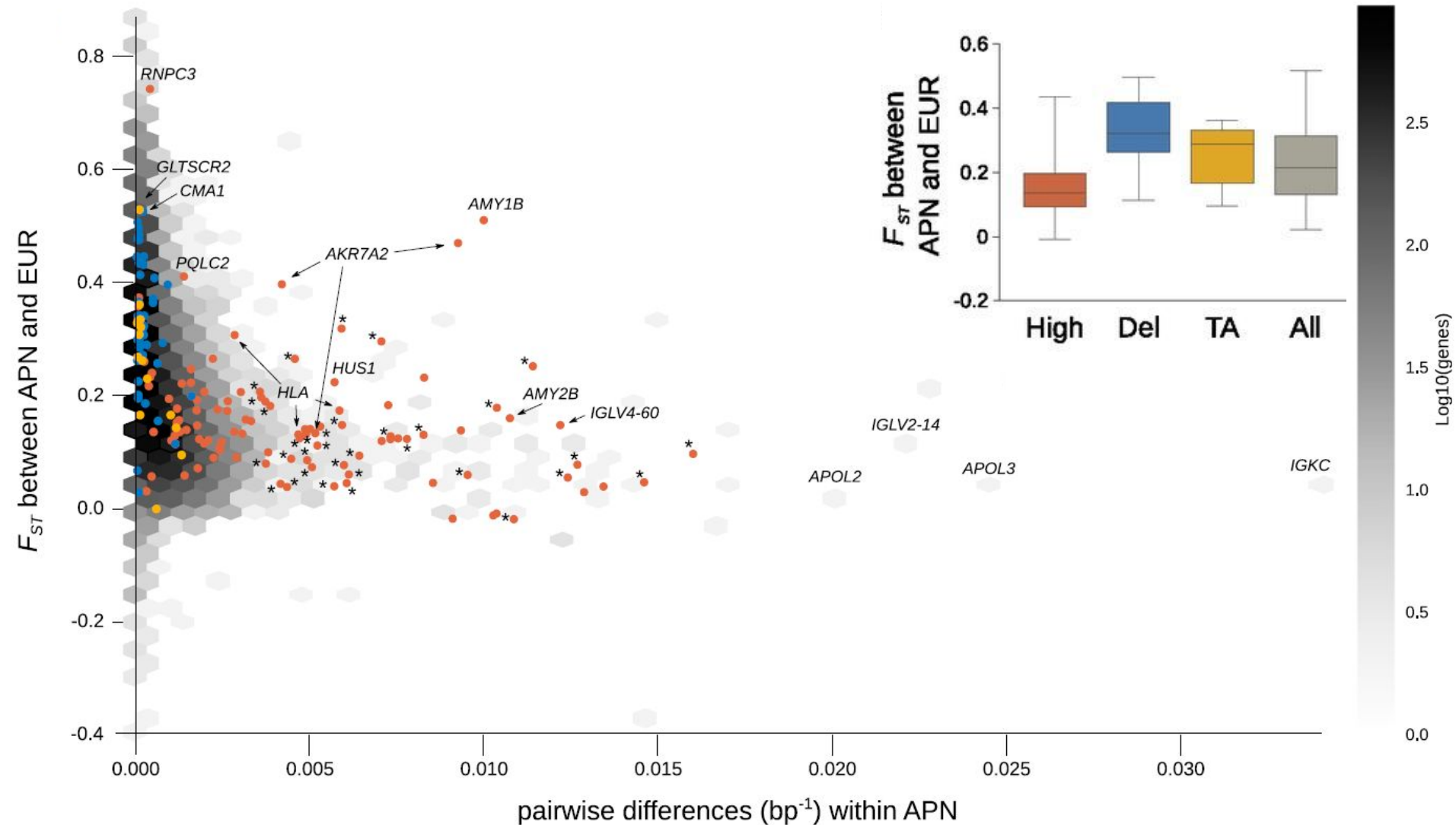
Not all deleterious fixations come to harm



Pattern confirmed with **additional individuals** in three of these genes:
PLAXNB1, DCC, DLL3 (not yet checked in the others)

Maintaining and losing diversity in genes

Balancing selection and deleterious alleles fixation by drift



Concluding



Surviving and diverging at very small population size

Apennine brown bear population is isolated since 300 generations (3,000 years) likely due to slash-and-burn agriculture during Neolithic expansion in Europe.

Population size has been rather small (less than 100-300) since then causing **high inbreeding and massive loss of diversity genome-wide**.

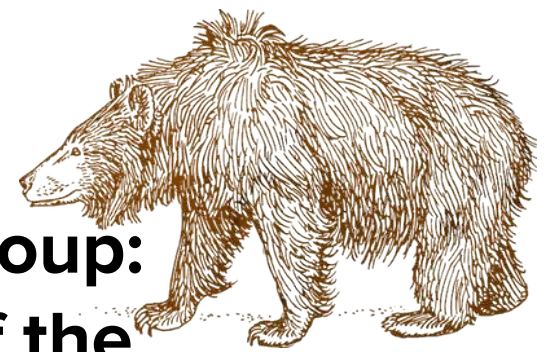
Genetic load of this population is high due to **fixation of deleterious substitutions** by genetic drift.

Some regions retaining **high genetic diversity due to functional or structural reasons** (duplicated loci, ectopic recombination) could have favoured survival.

Some alleles fixed by drift (or selected by hunting?) could have **changed behavior-related traits** decreasing human persecution of this population and, again, favouring survival.

(This work was done with just a handful of whole-genomes but about three years of work of quite some people!!)

Survival and divergence in a small group: The extraordinary genomic history of the endangered Apennine brown bear stragglers



Andrea Benazzo*, Emiliano Trucchi*, James A. Cahill, Pierpaolo Maisano Delser, Stefano Mona, Matteo Fumagalli, Lynsey Bunnefeld, Luca Cornetti, Silvia Ghirotto, Matteo Girardi, Lino Ometto, Alex Panziera, Omar Rota-Stabelli, Enrico Zanetti, Alexandros Karamanlidis, Claudio Groff, Ladislav Paule, Leonardo Gentile, Carles Vilà, Saverio Vicario, Luigi Boitani, Ludovic Orlando, Silvia Fuselli, Cristiano Vernesi, Beth Shapiro, Paolo Ciucci, and Giorgio Bertorelle (2017) **PNAS**, 114, E9589-E9597





SIDE B

Climate-driven range shifts in fragmented ecosystems

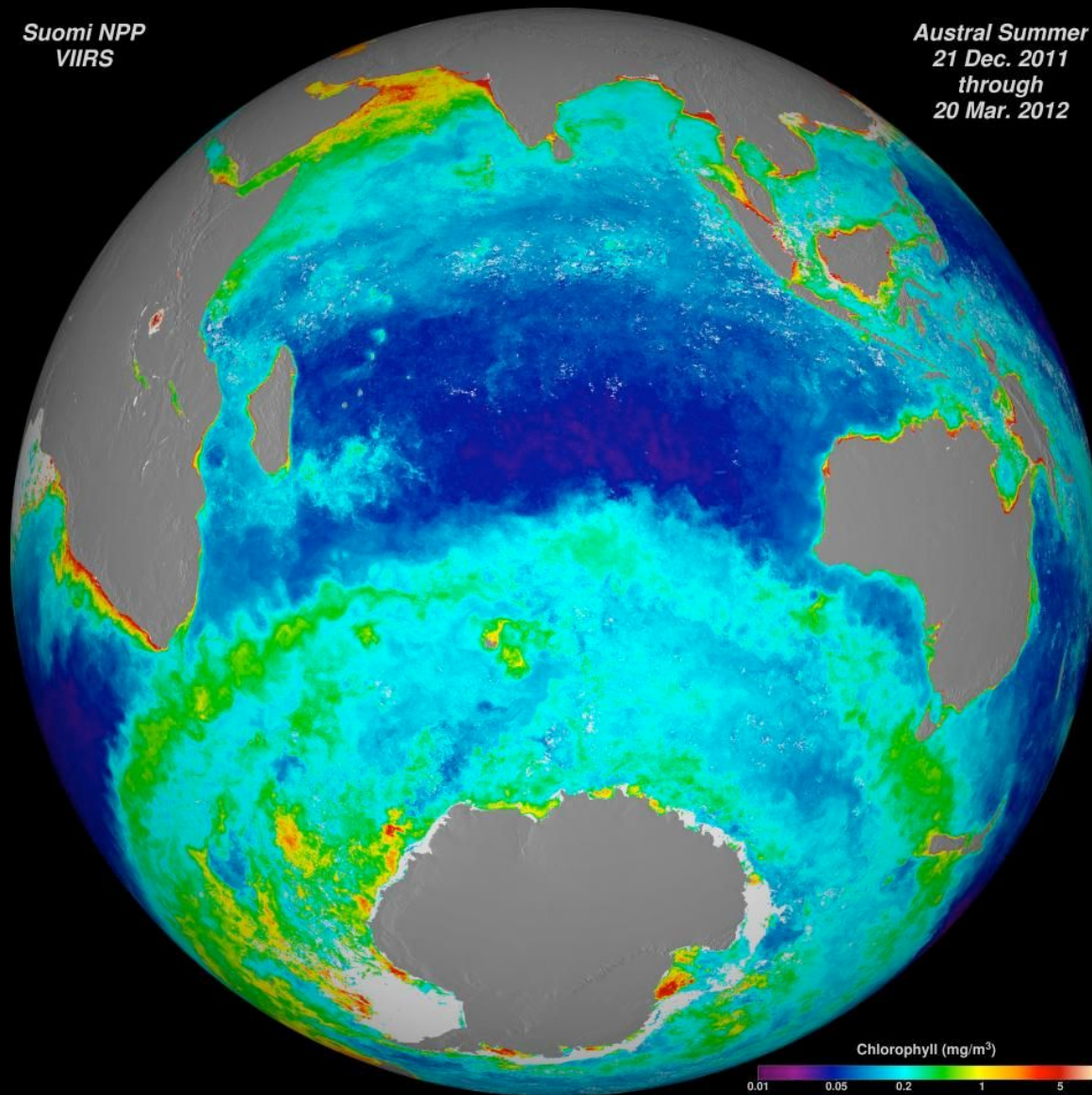
Understanding the impact of climate change on (sub-)Antarctic upper-level predators



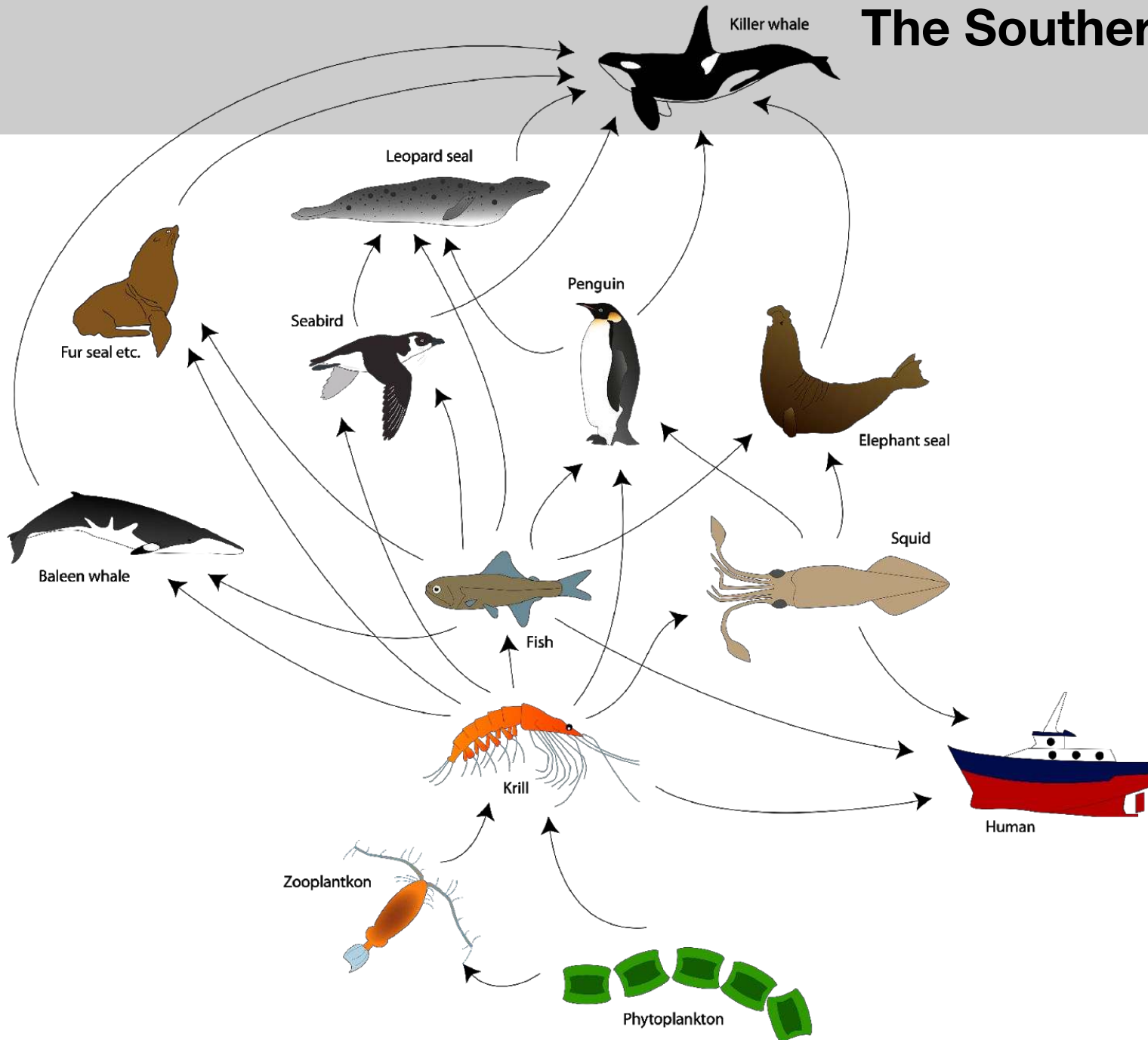
The Southern Ocean: The Antarctic Polar Front



The Southern Ocean: Chlorophyll and marine productivity



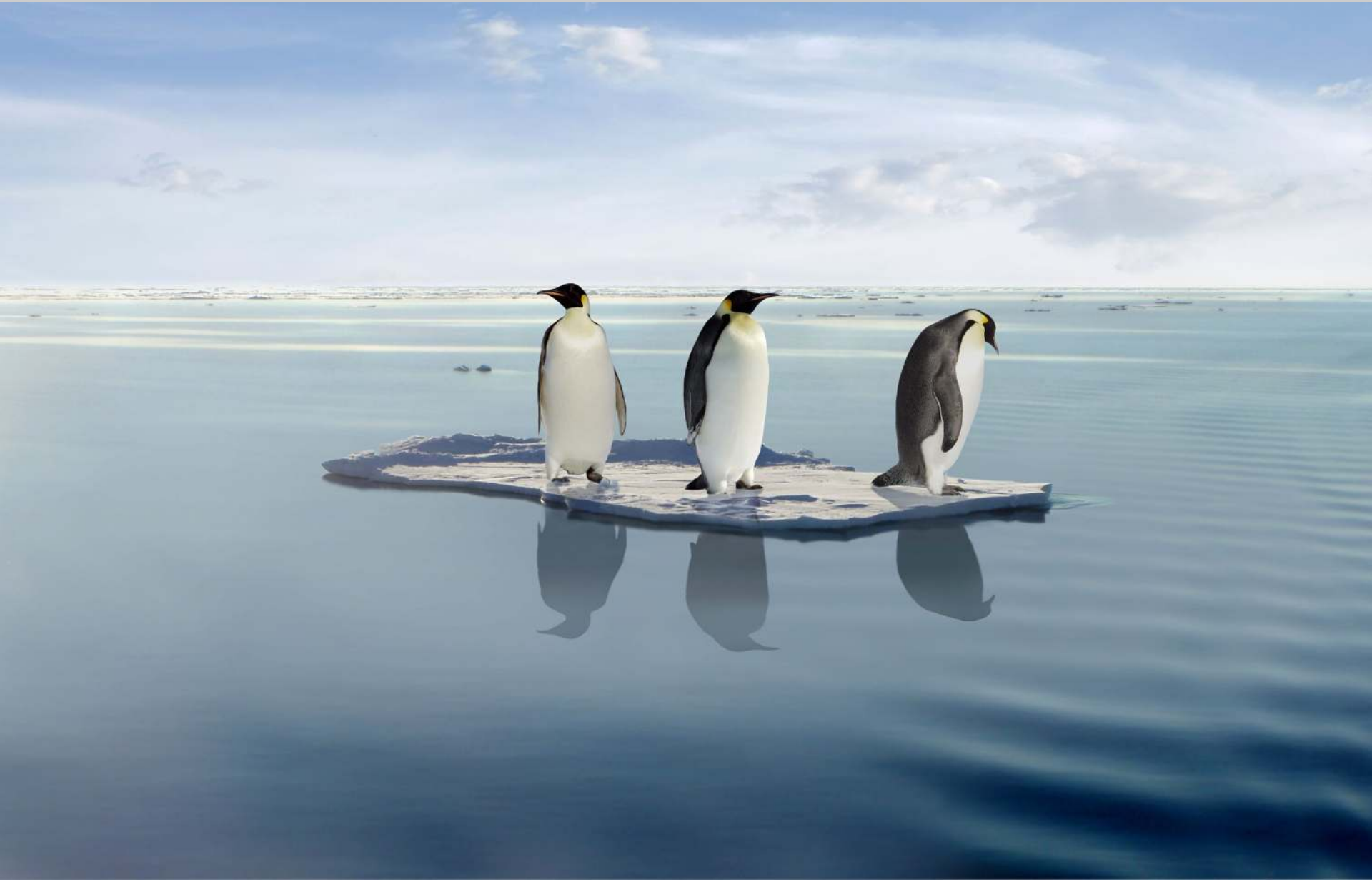
The Southern Ocean: Food web



Global warming: A fact



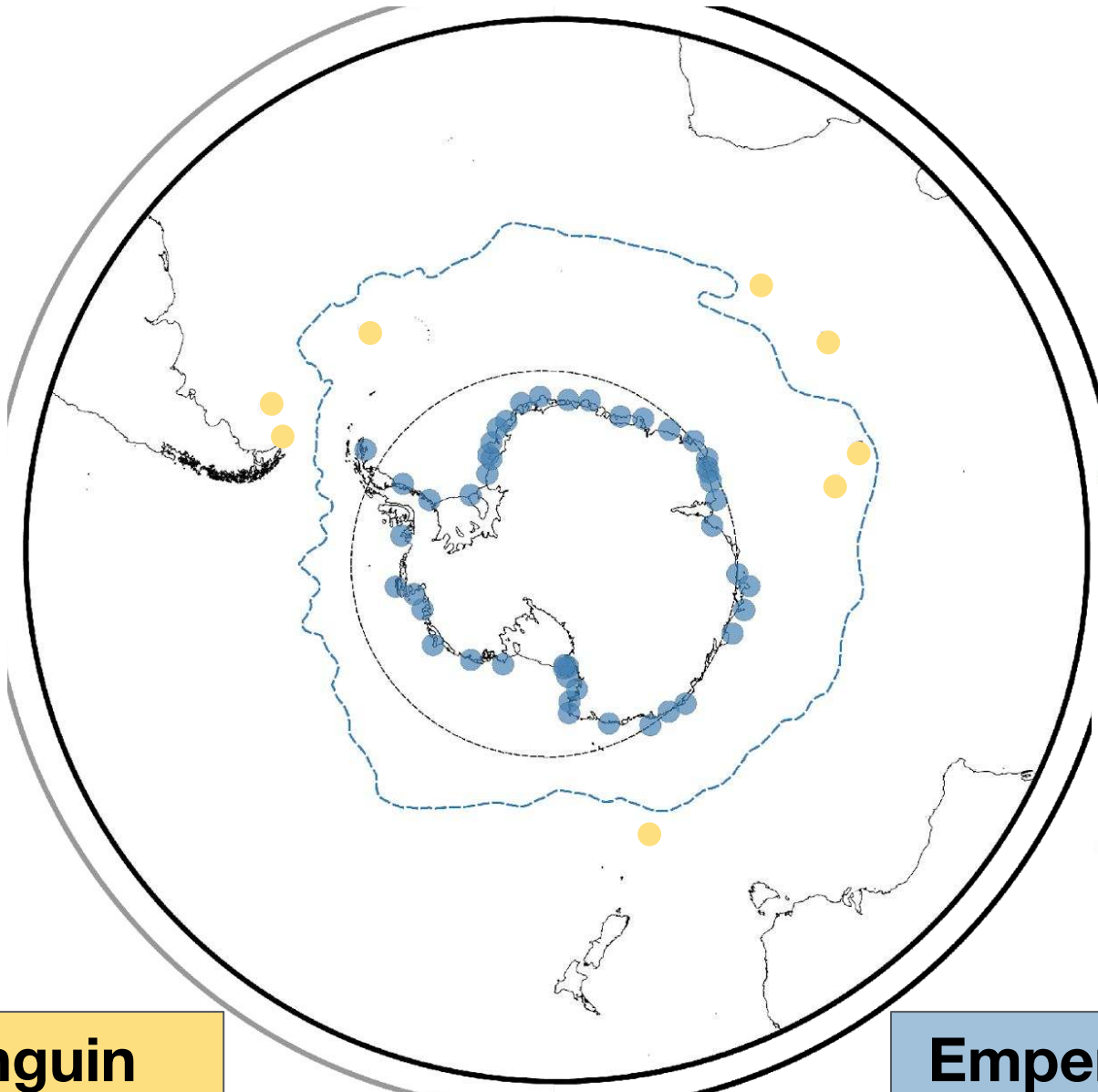
Global warming: Which is the effect in Antarctic ecosystems?



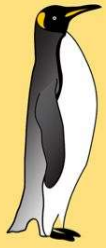
The Southern Ocean: Penguins' cradle



King penguin

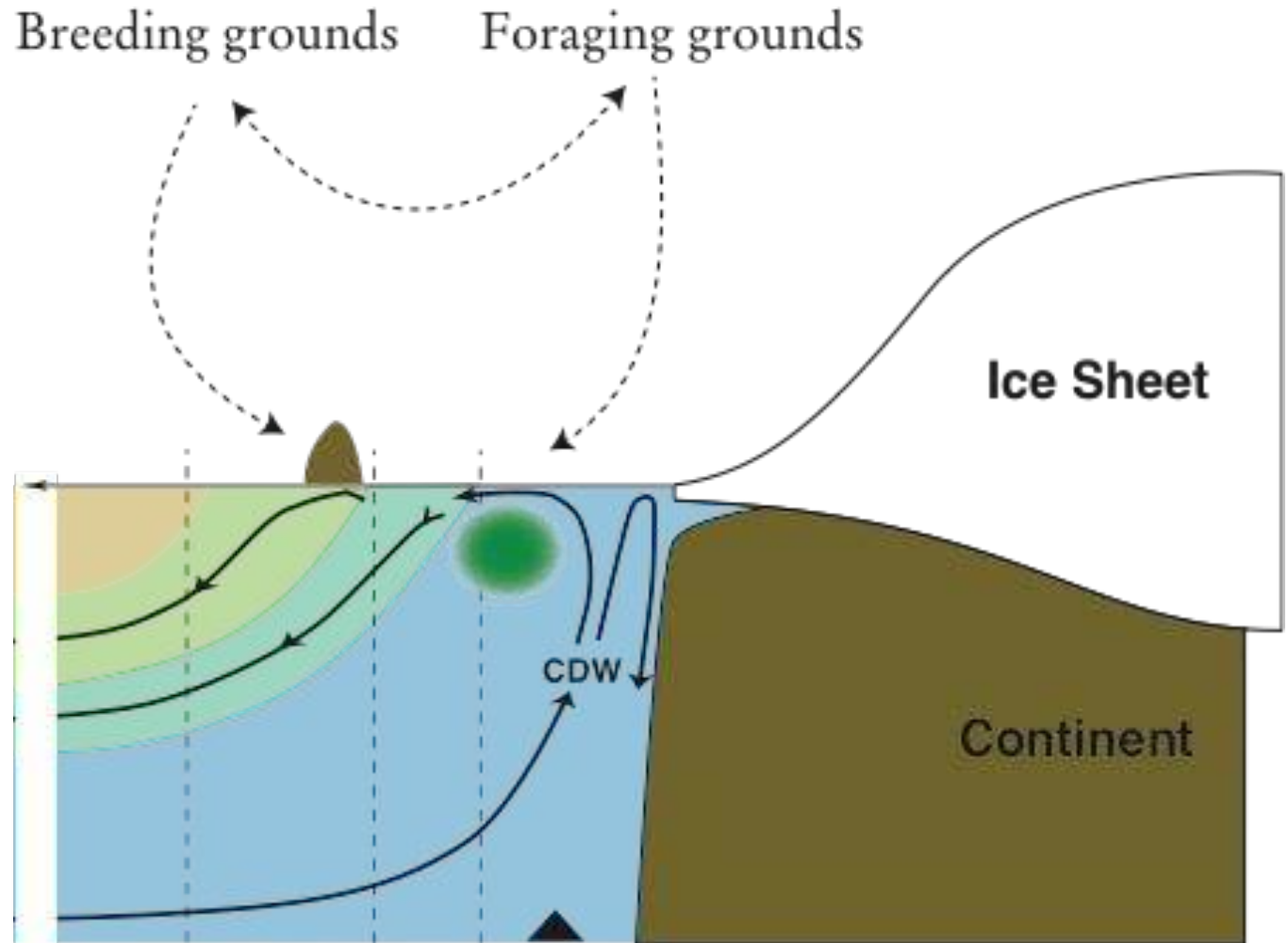


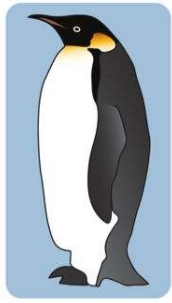
Emperor penguin



The King penguin's foraging strategy:

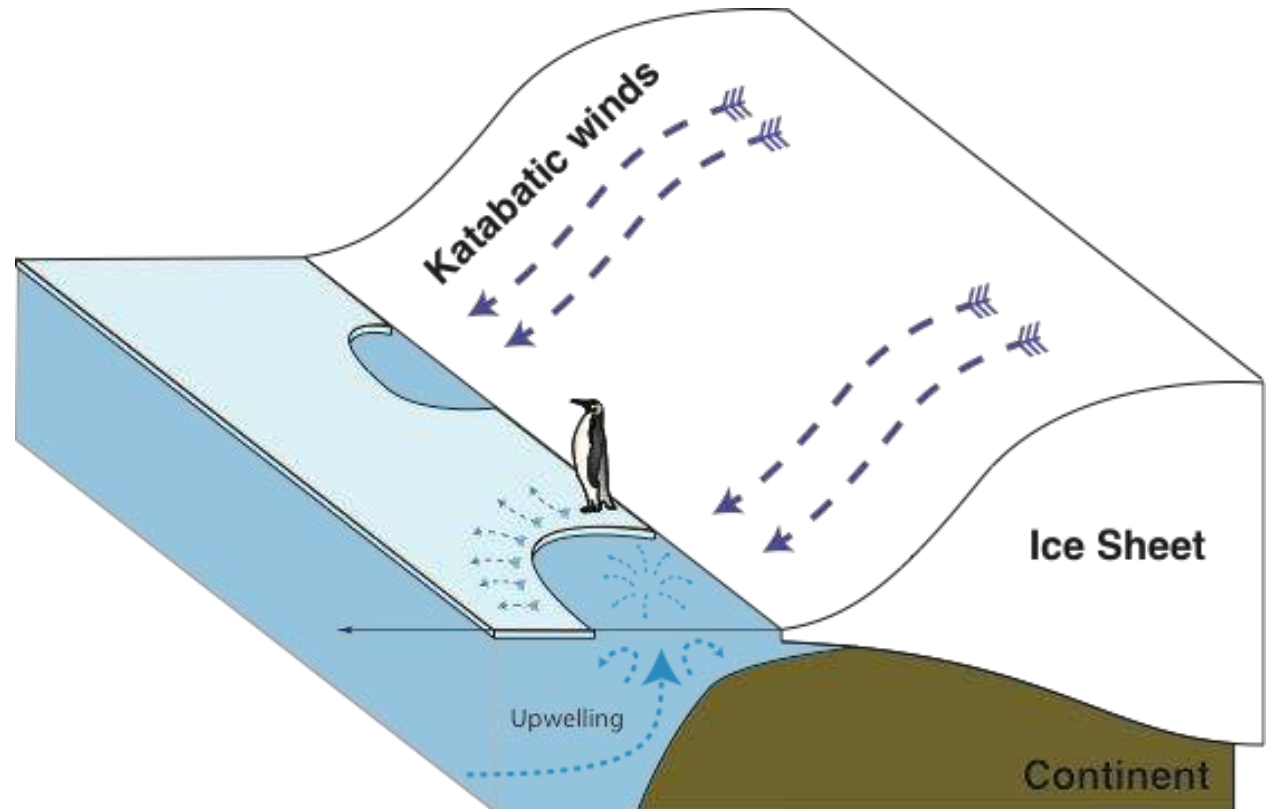
The polar front area

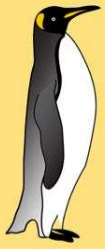




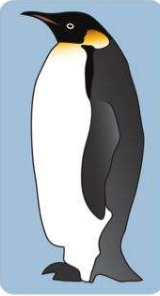
The Emperor penguin's foraging strategy:

The coastal polynyas



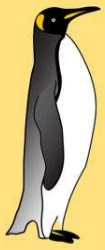


Question

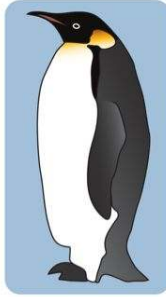


Which will be the **effects of climate change** on **demography and distribution** of the two species?





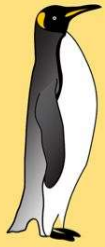
Methods



Species distribution modeling...

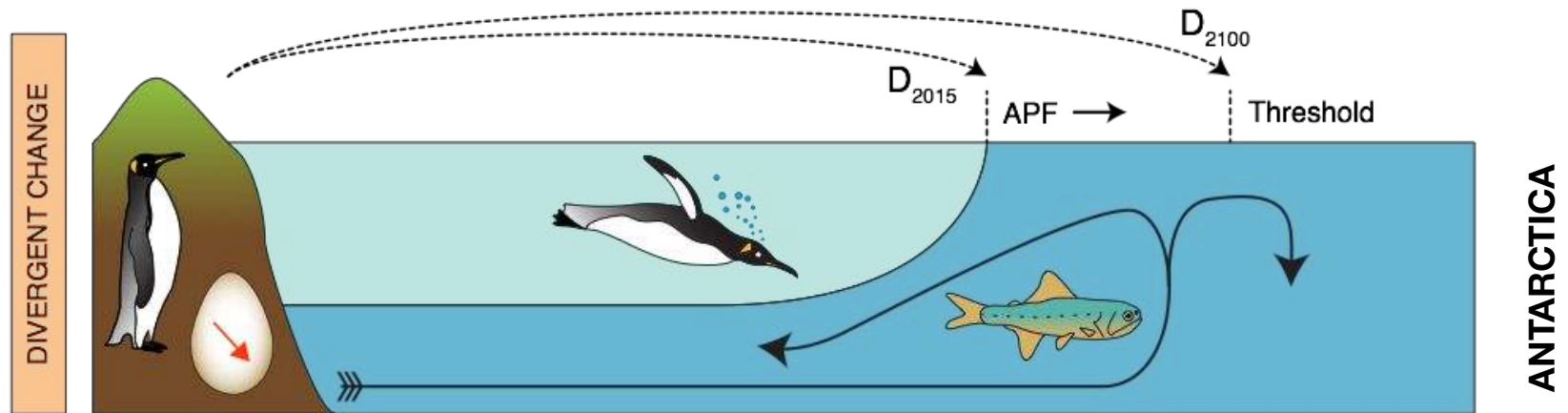
...coupled with **genomics** (RADseq + WGS) to validate model assumptions and hindcasting results

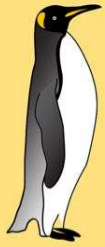




Our model of habitat suitability

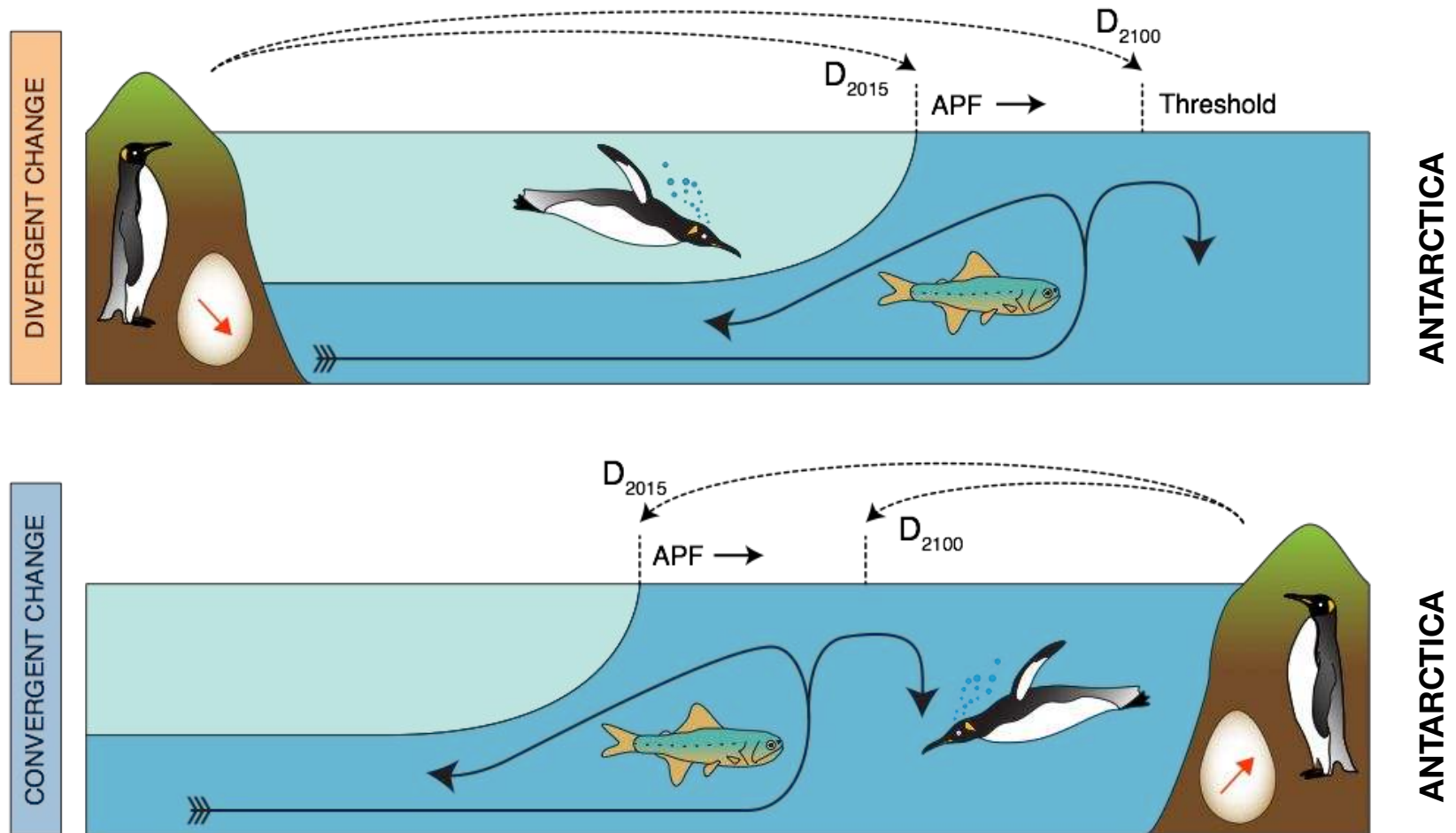
Increasing temperature and shifting Polar Front

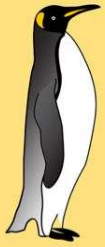




Our model of habitat suitability

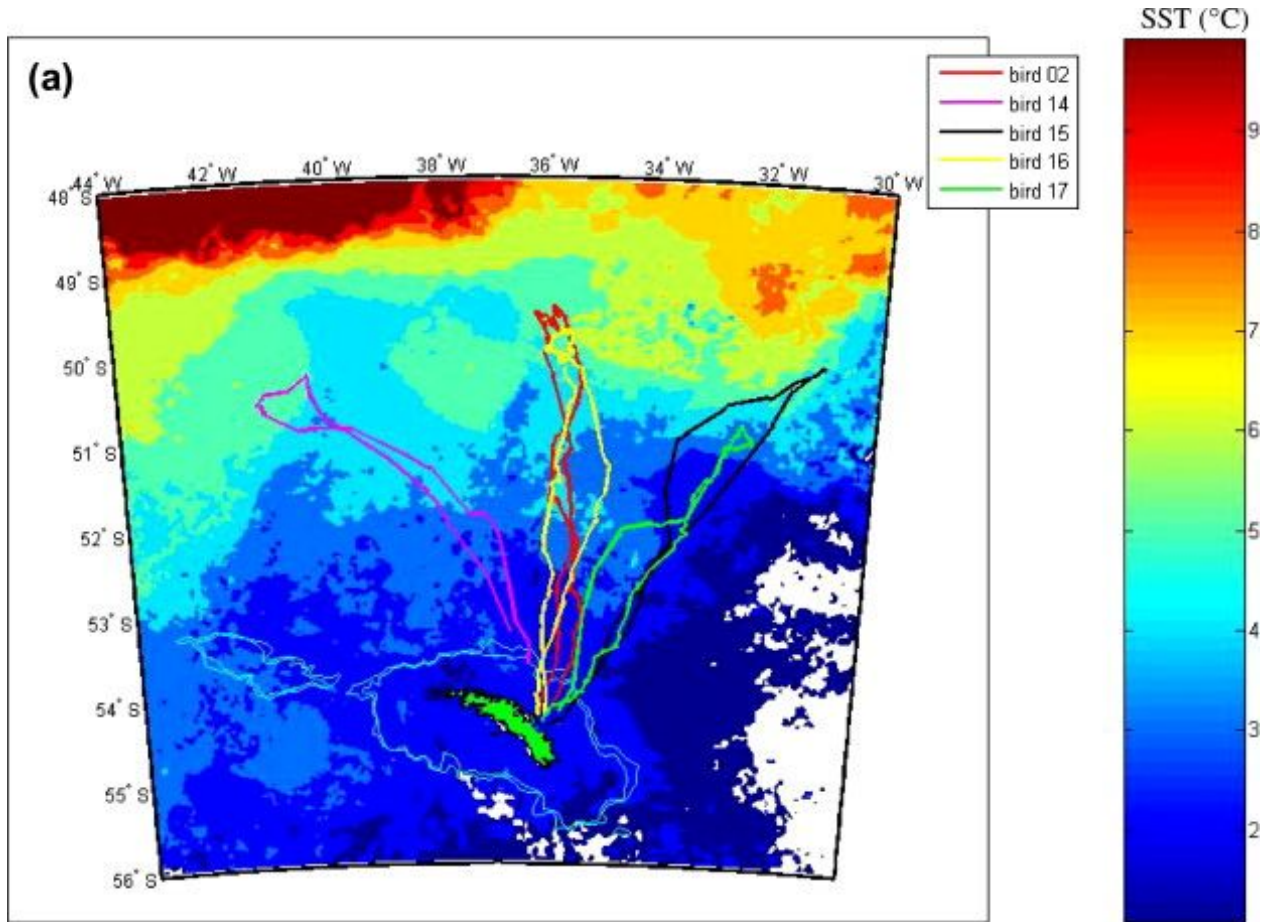
Increasing temperature and shifting Polar Front



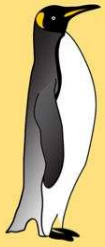


Bio-physical ecological niche model

Major constraints

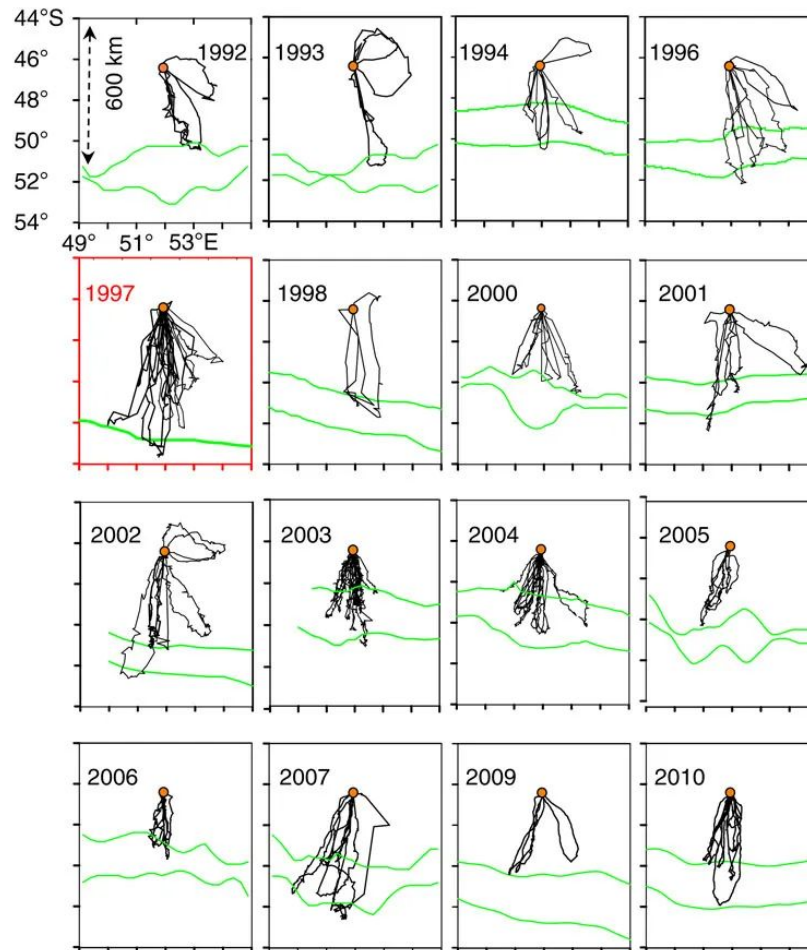


Sea Surface Temperature (SST)
and efficient
foraging

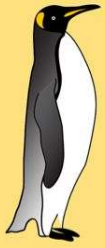


Bio-physical ecological niche model

Major constraints

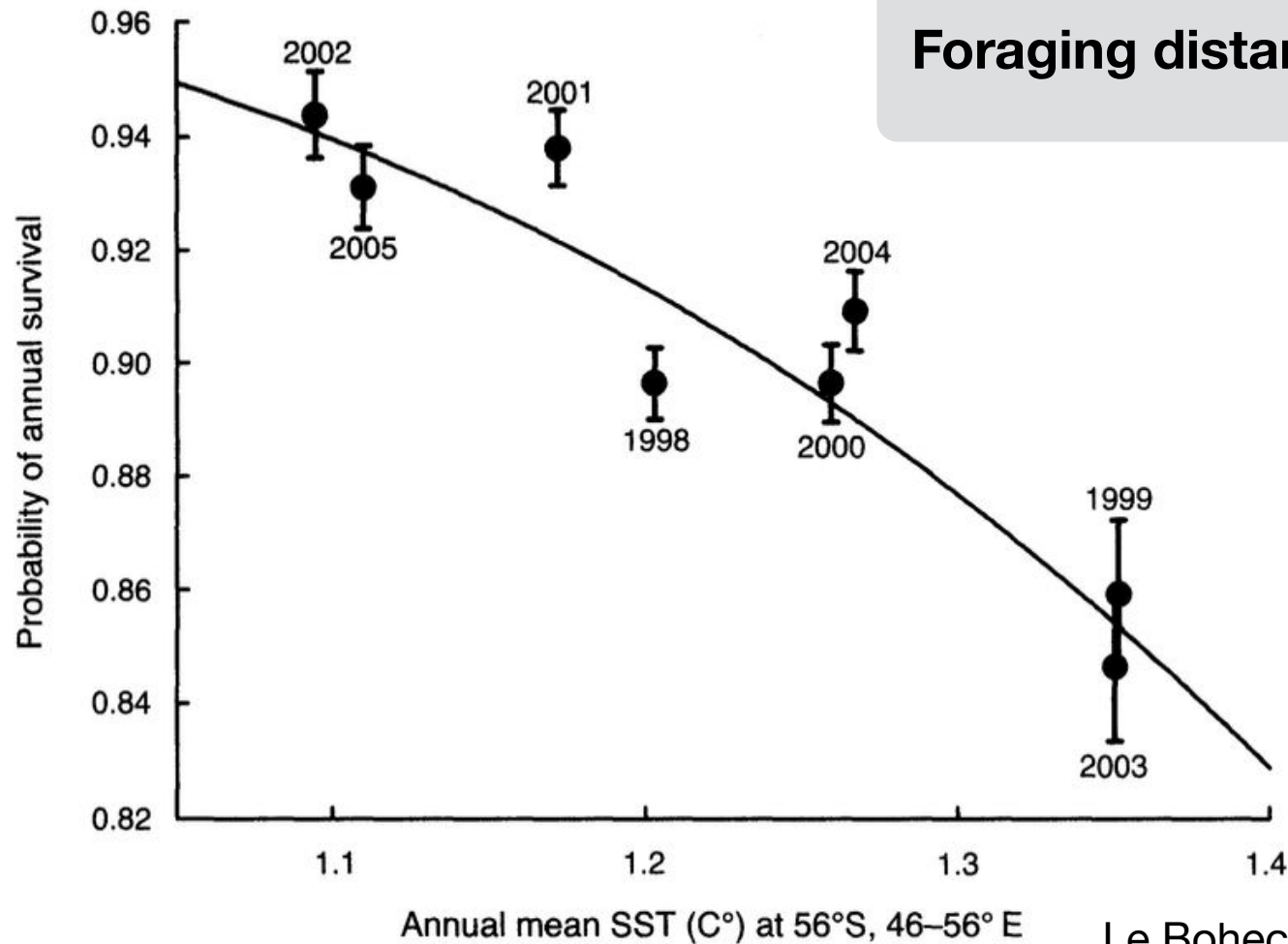


Sea Surface Temperature (SST)
and efficient
foraging

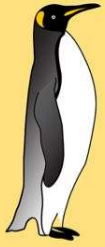


Bio-physical ecological niche model

Major constraints

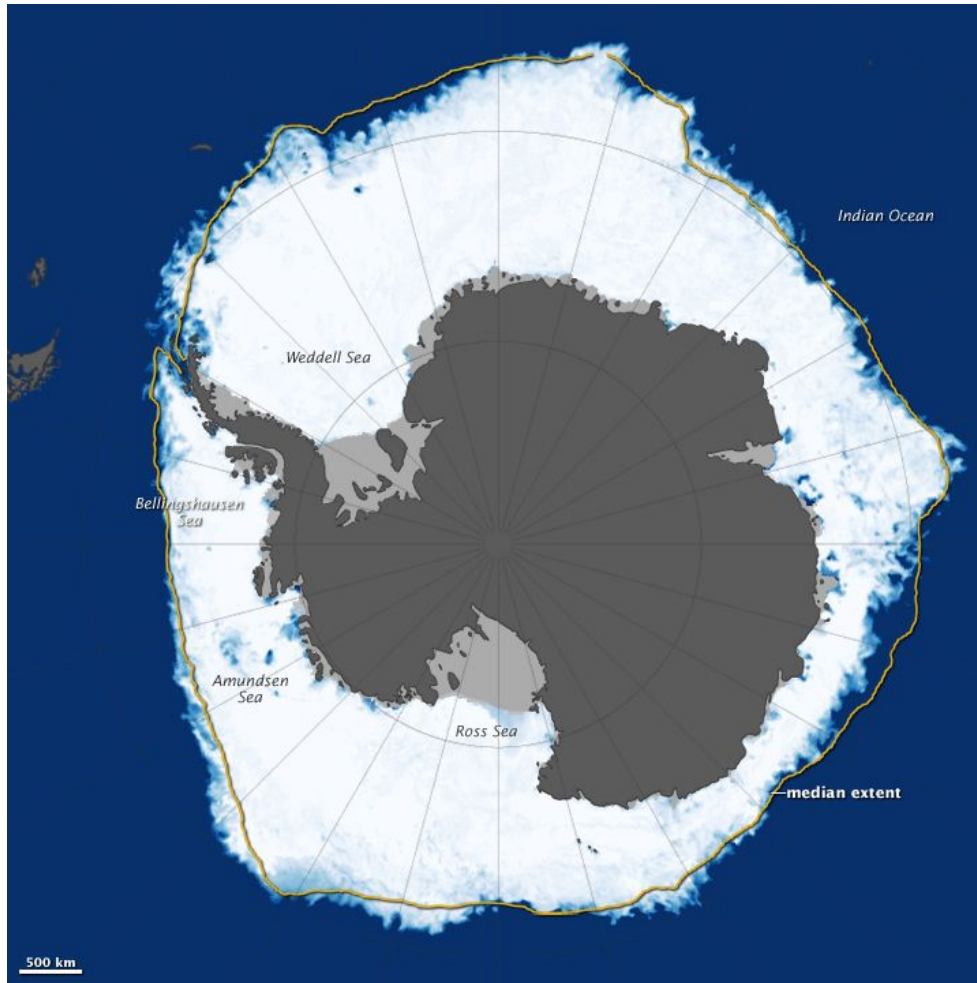


Le Bohec *et al.* 2008

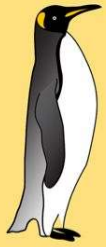


Bio-physical ecological niche model

Major constraints



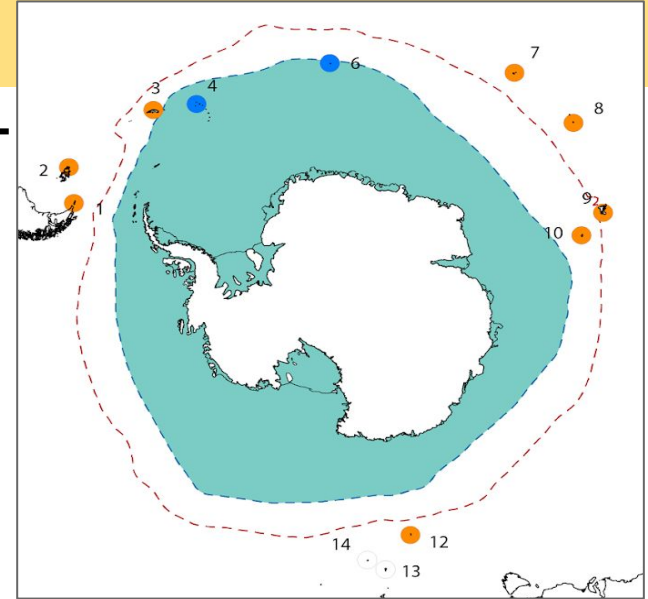
Temperature -> year-round
ice-free breeding grounds

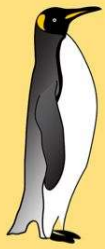


Our model finds all suitable islands

Accurate description of current distribution

PRESENT

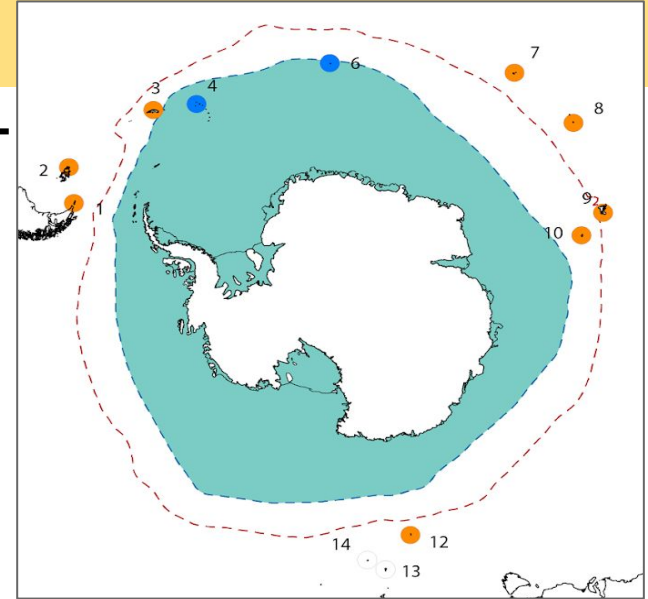




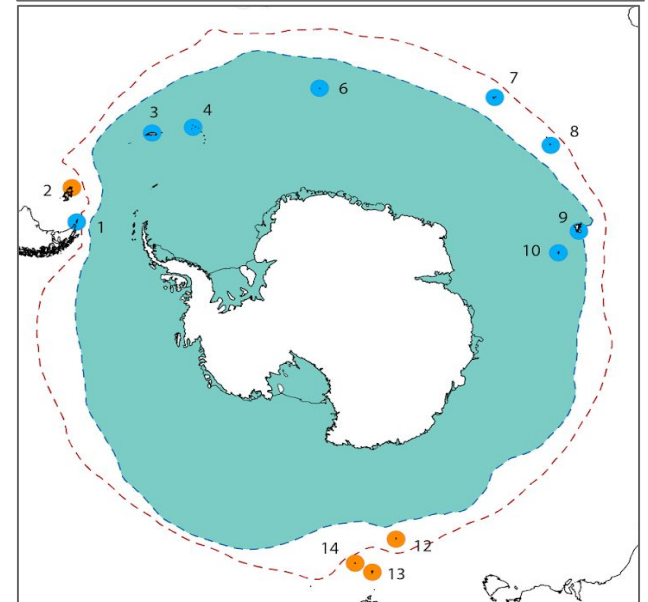
Our model suggests cold ice age refugia

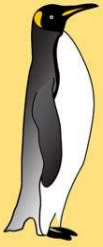
Hindcasting very few suitable islands at the LGM

PRESENT



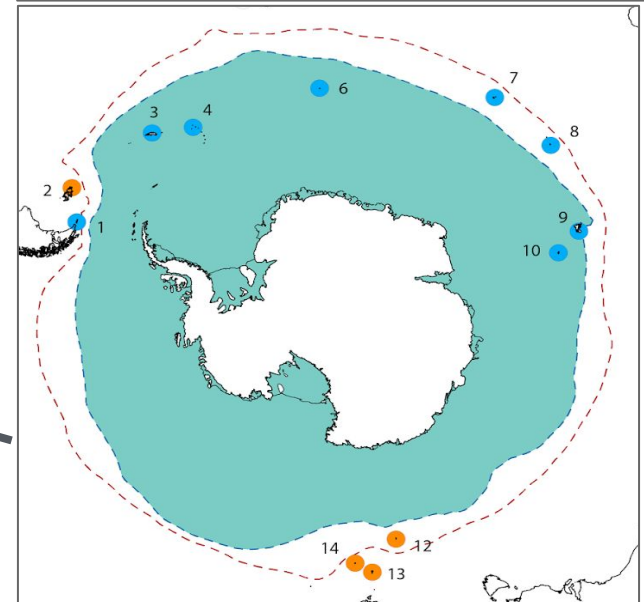
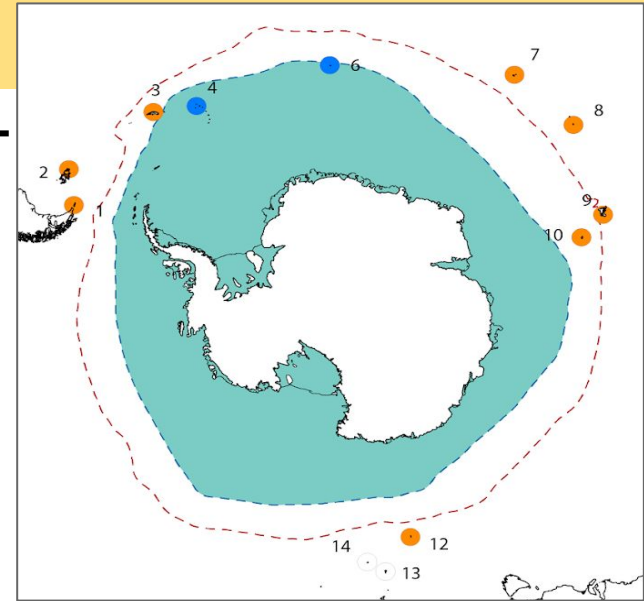
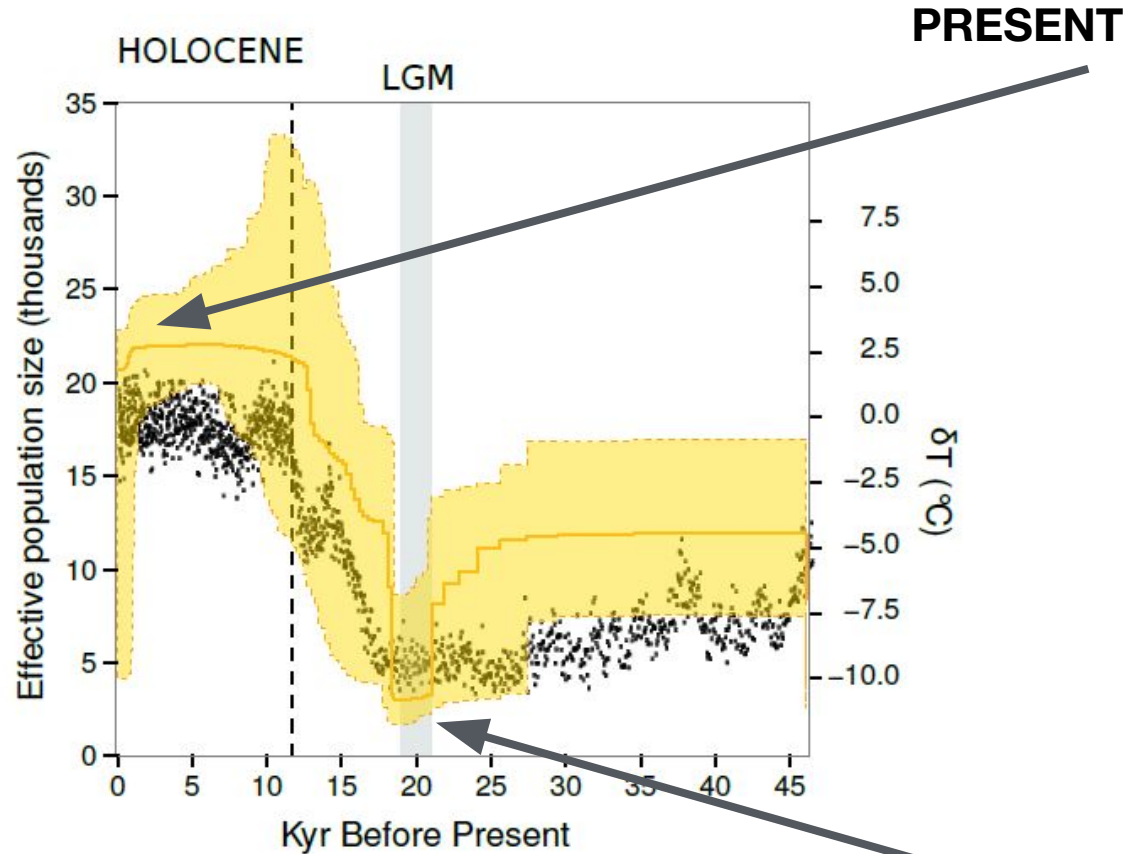
PAST





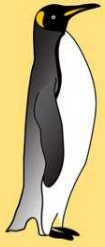
Genomic inference of past demography

In agreement with the species distribution model



Stairway plot using unfolded site frequency spectrum estimated by ANGSD

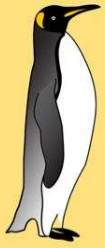
PAST



What is our prediction for the future?

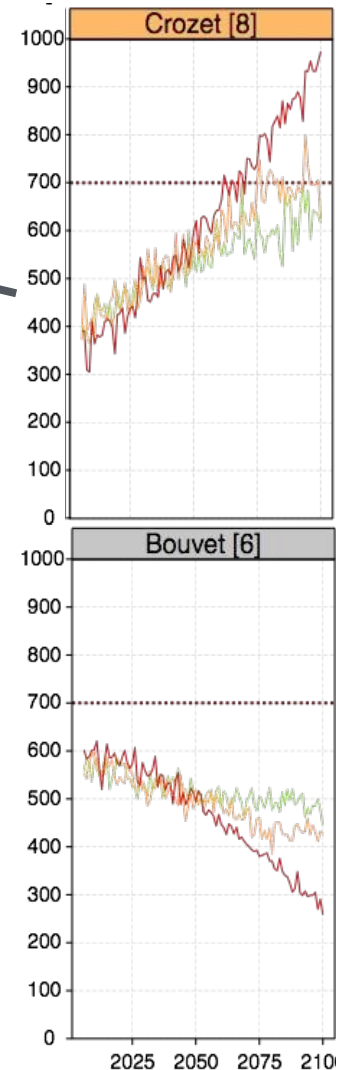
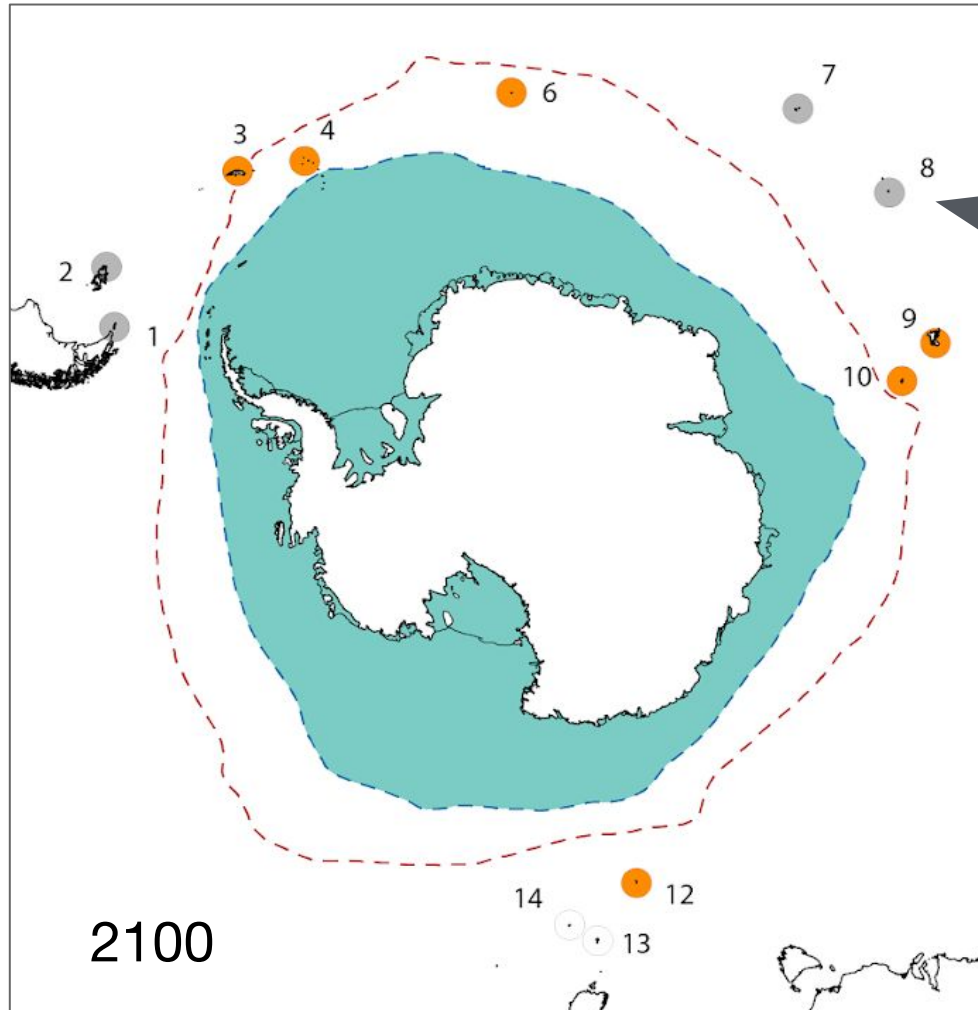
Under ongoing climate change scenario (up to 2100)

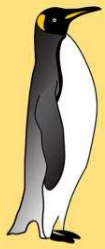




Our model predicts largest colonies will be lost

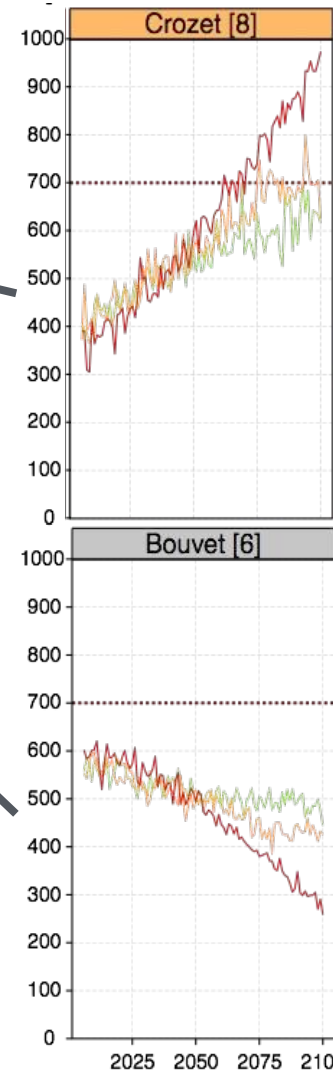
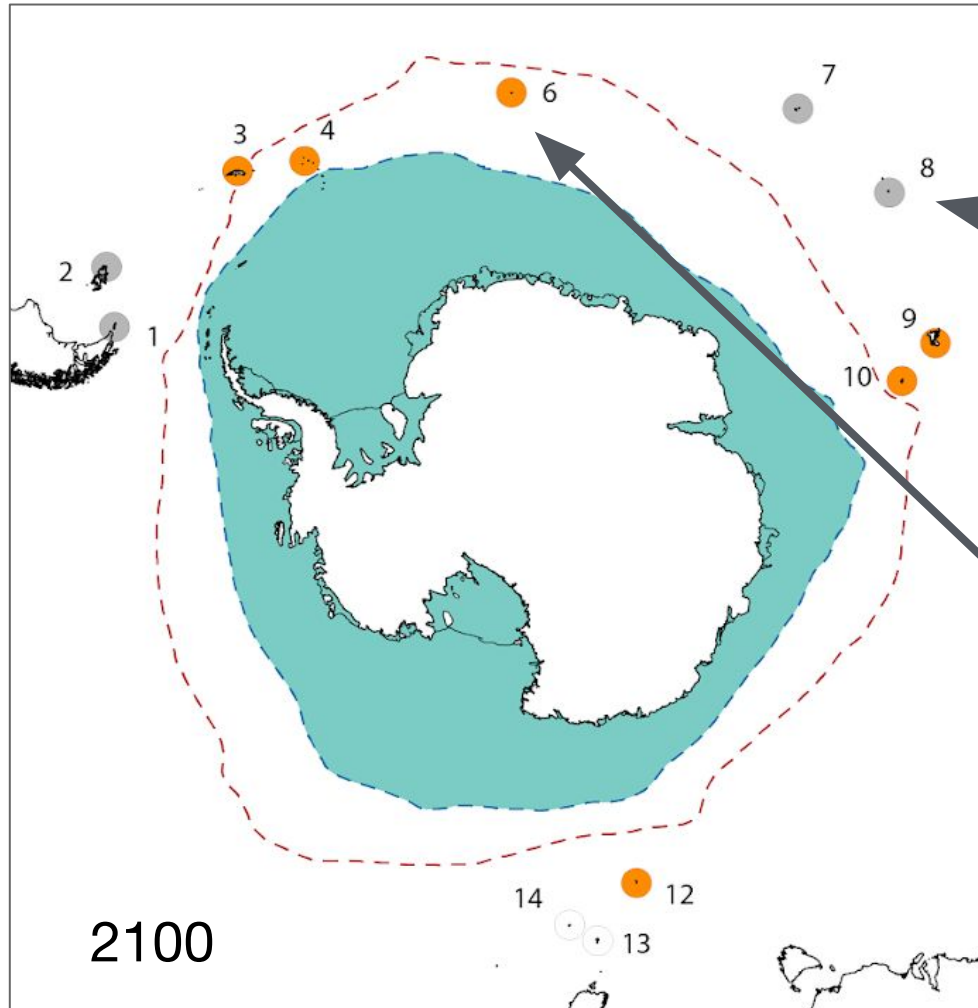
Few islands will become suitable



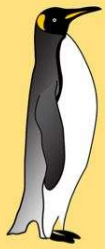


Our model predicts largest colonies will be lost

Few islands will become suitable

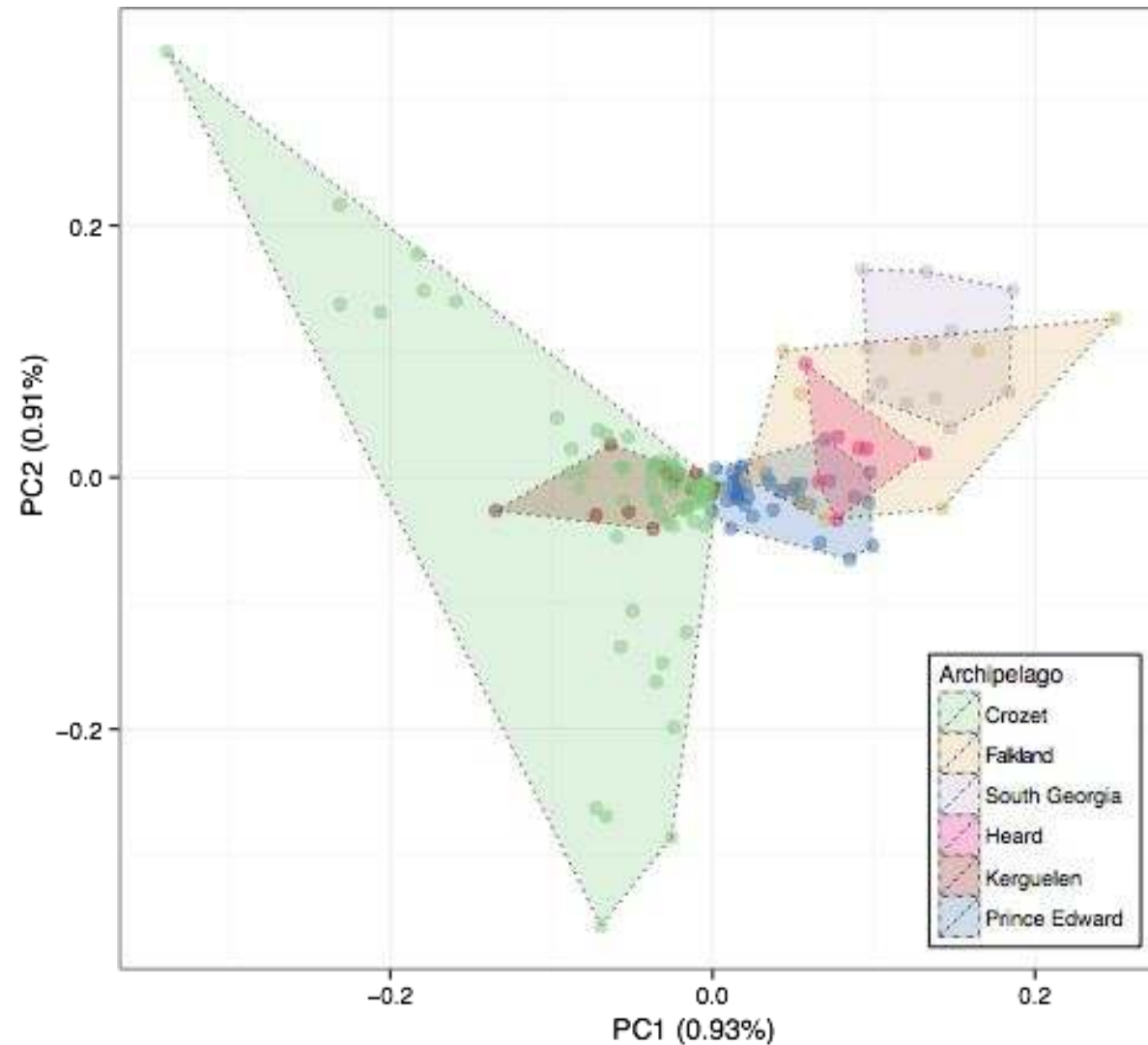


PROJECTED FORAGING DISTANCE



Genomic inference of high gene flow

World-wide connectivity in the King penguin

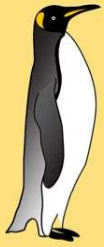


One single **highly connected** metapopulation

Long-range dispersal is not an issue

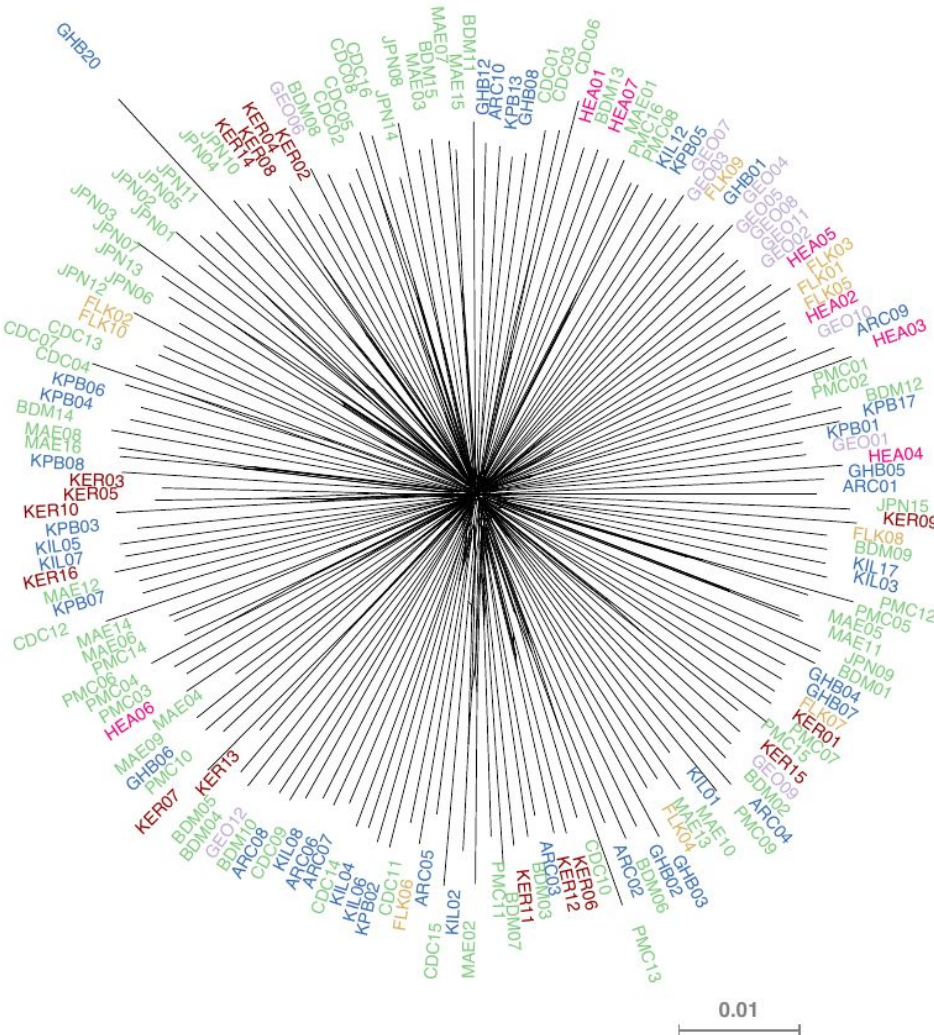
About **50,000 loci** genotyped in ca. **160 individuals** from **13 colonies** in **6 archipelagos**.

PCA



Genomic inference of high gene flow

World-wide connectivity in the King penguin

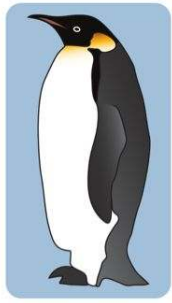


One single **highly connected** metapopulation

Long-range dispersal is not an issue

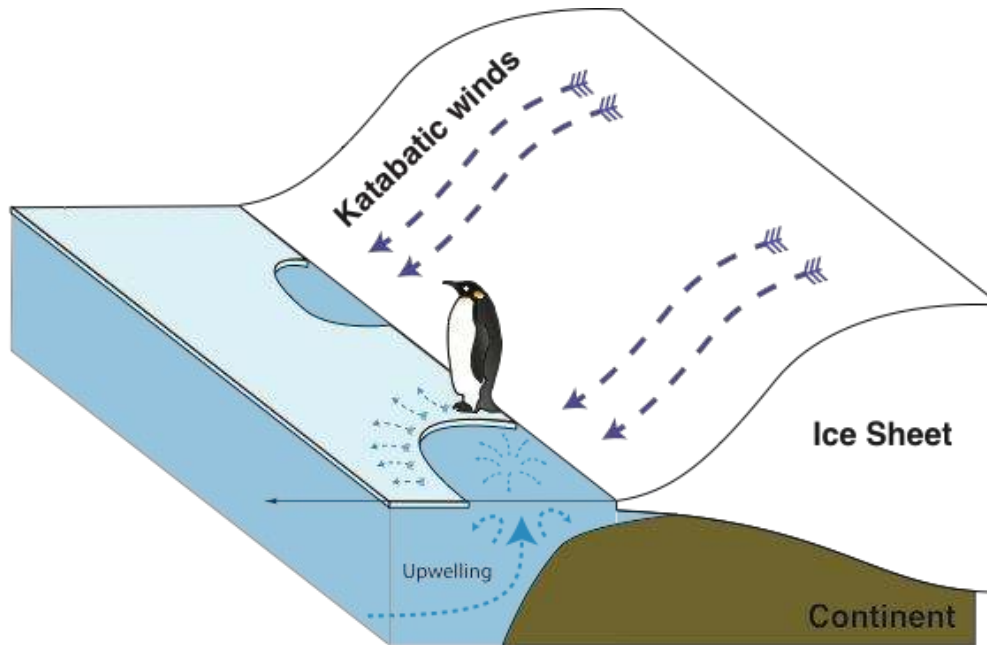
About **50,000 loci** genotyped in ca. **160 individuals** from **13 colonies** in **6 archipelagos**.

NJ tree

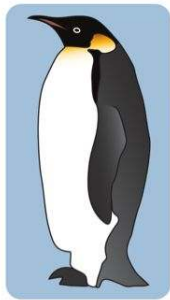


The Emperor penguin's foraging strategy:

Coastal polynyas are maintained by wind stress

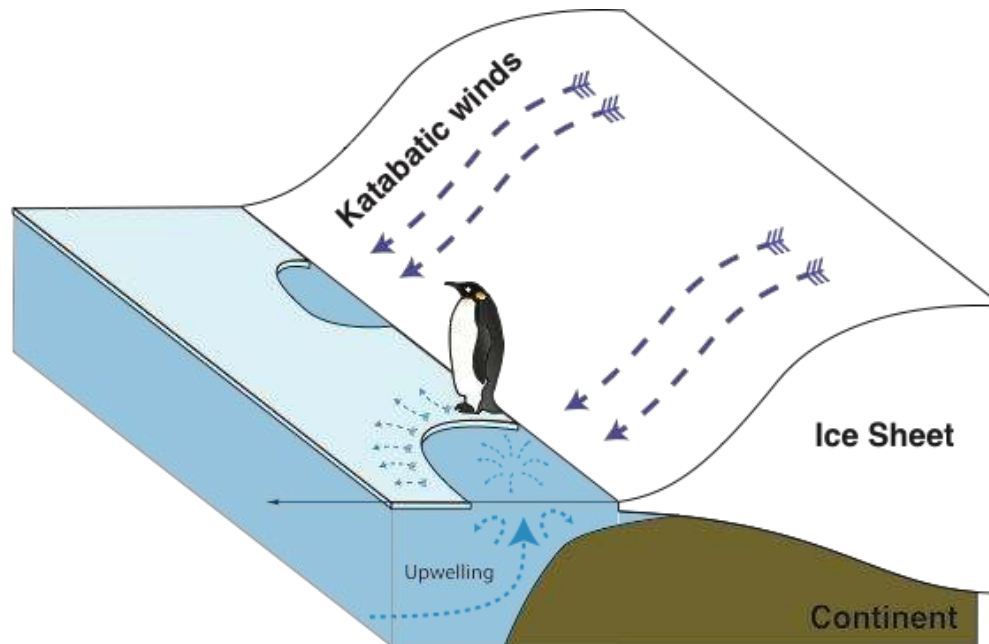


Modelling wind stress on
coastal ice area



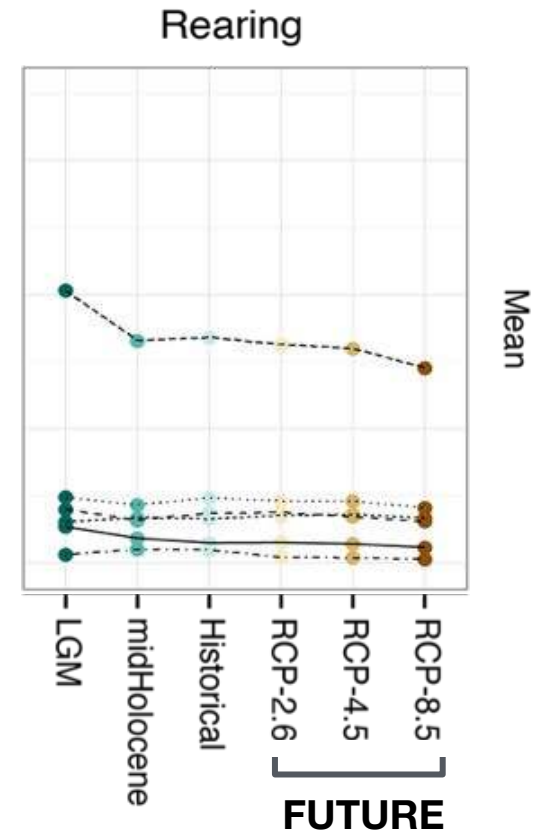
The Emperor penguin's foraging strategy:

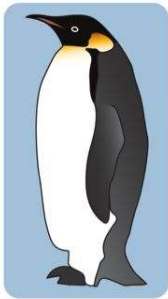
Coastal polynyas are maintained by wind stress



6 models assessed for **4 variables**,
under **6 « forcing periods »**

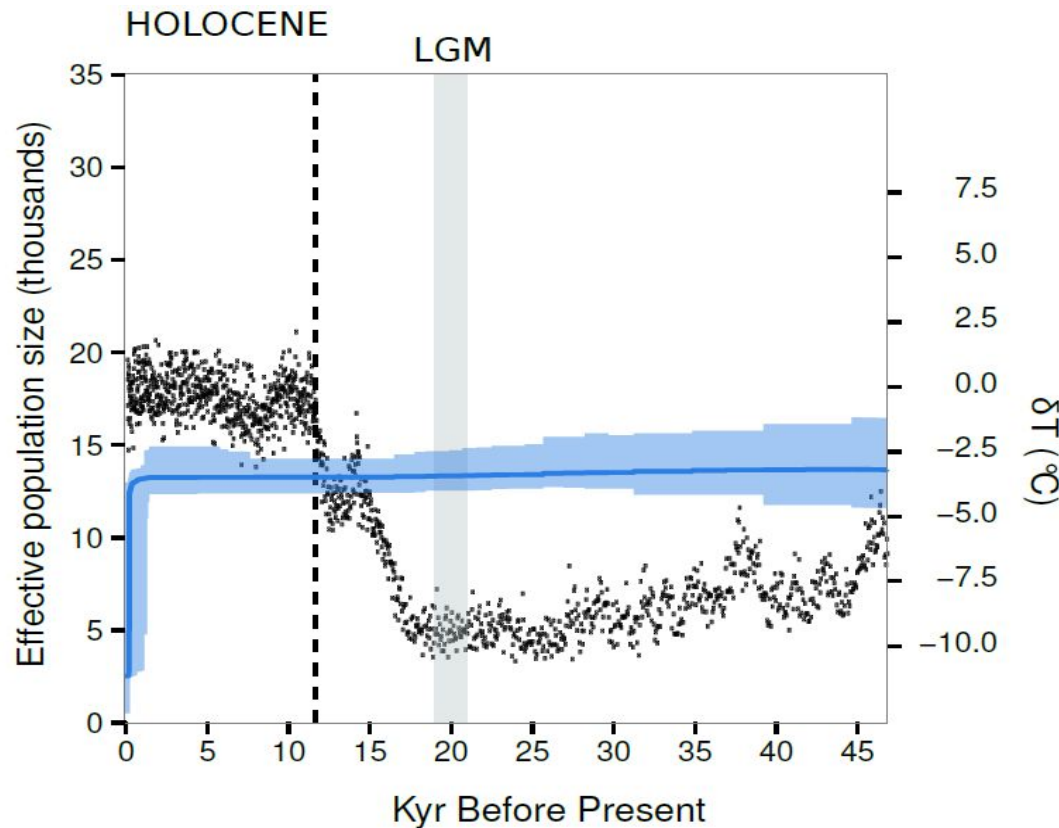
Modelling wind stress on
coastal ice area





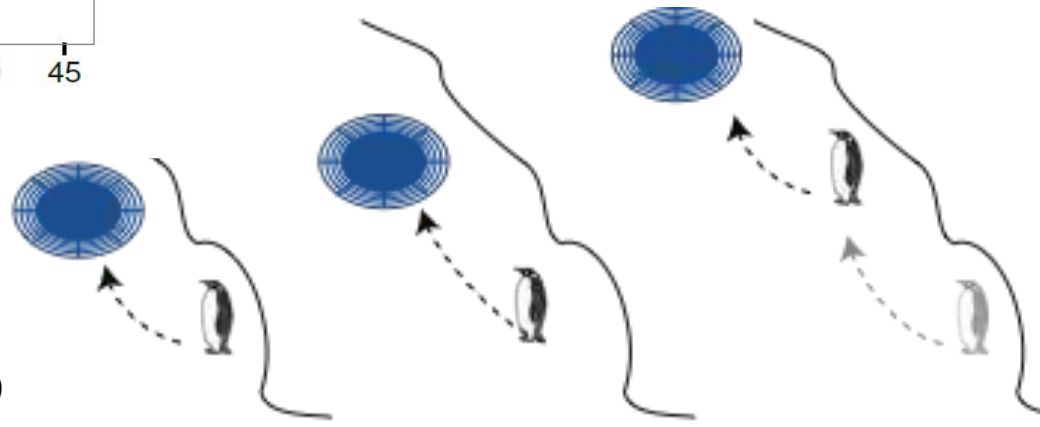
Genomic inference shows a constant population

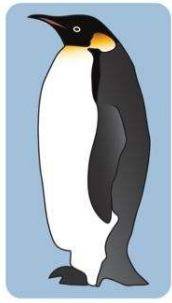
No detectable effect of past climatic shifts



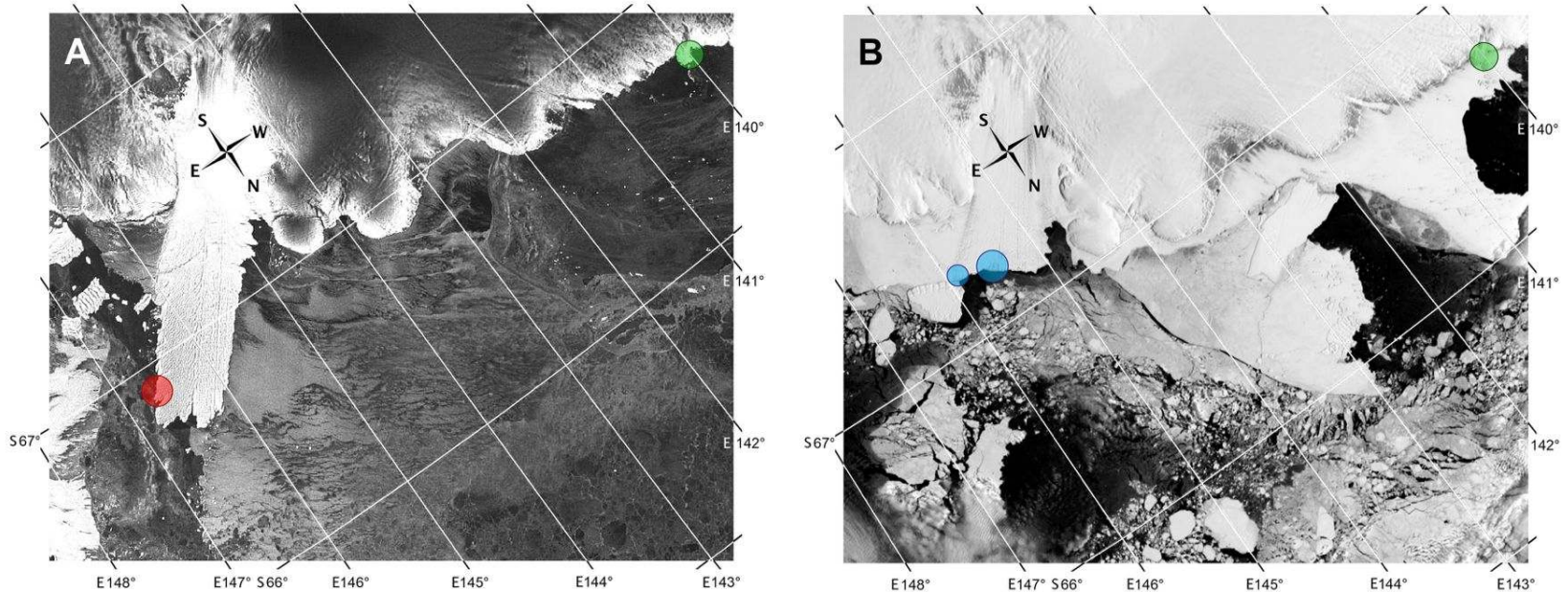
Polynyas move along the coast but were (and likely will be) always present

Stairway plot using unfolded site frequency spectrum estimated by ANGSD



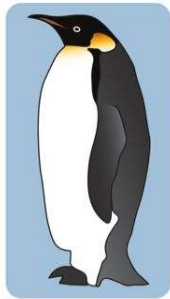


« Catastrophic pulse dispersal »?



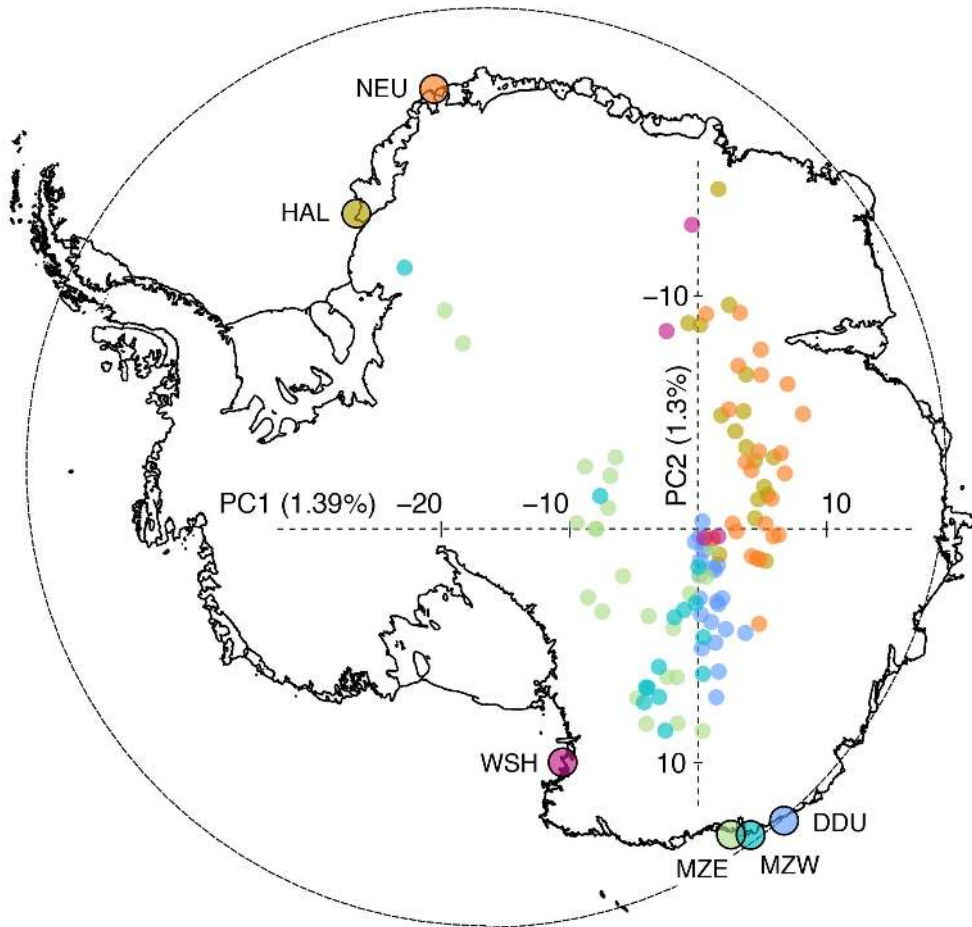
Mertz glacier, Adélie Land, before & after 2010

« Mass dispersal » events: Emperors need to stay flexible



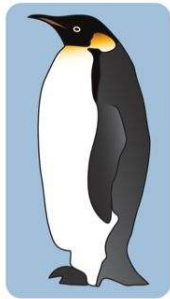
Genomic-based inference of high gene flow

Full circumpolar migration in the Emperor penguin



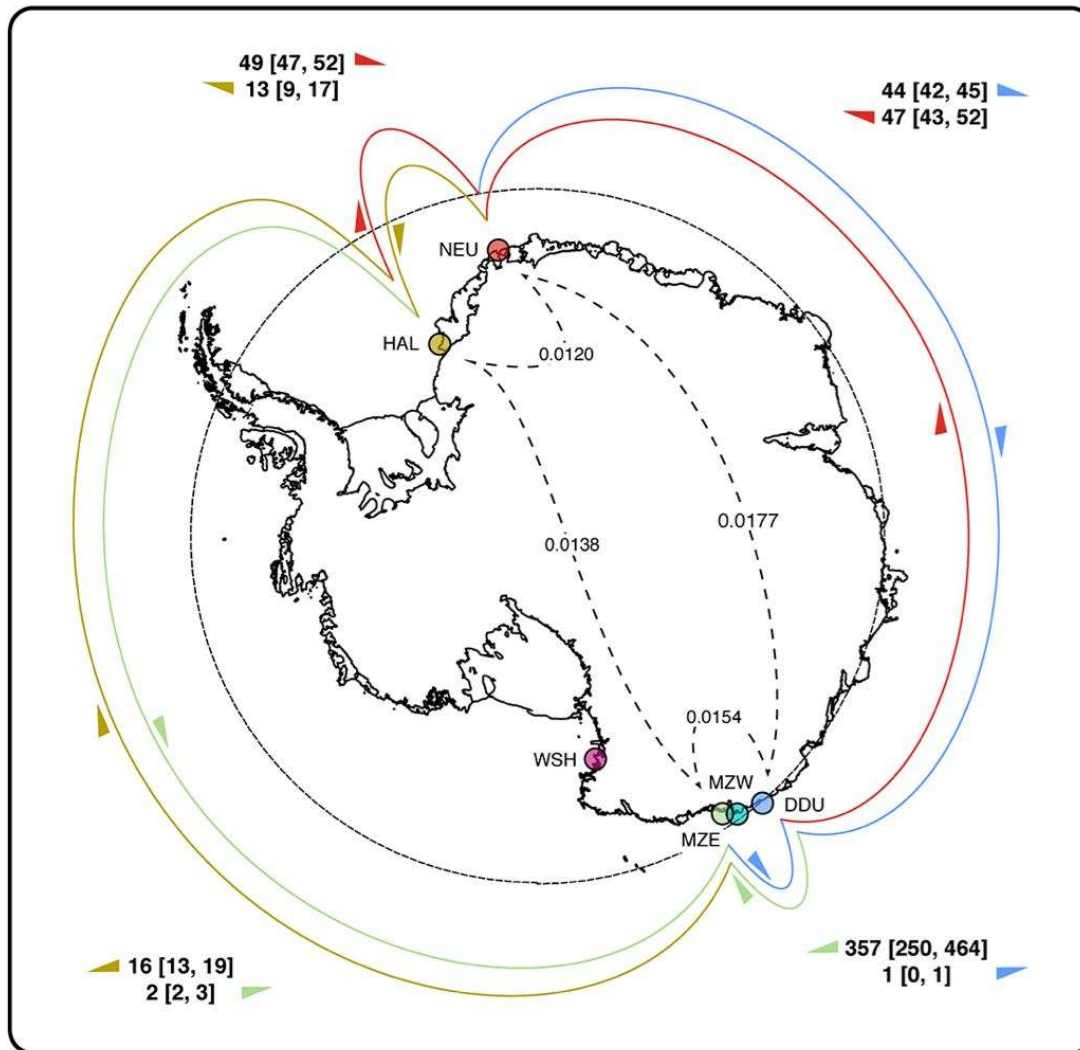
Very **weak** signal of **genetic differentiation** between colonies from the opposite site of Antarctica.

About **35,000 loci** genotyped in ca. **110 individuals** from **6 colonies**.



Genomic-based inference of high gene flow

Full circumpolar migration in the Emperor penguin



One single **highly connected** metapopulation

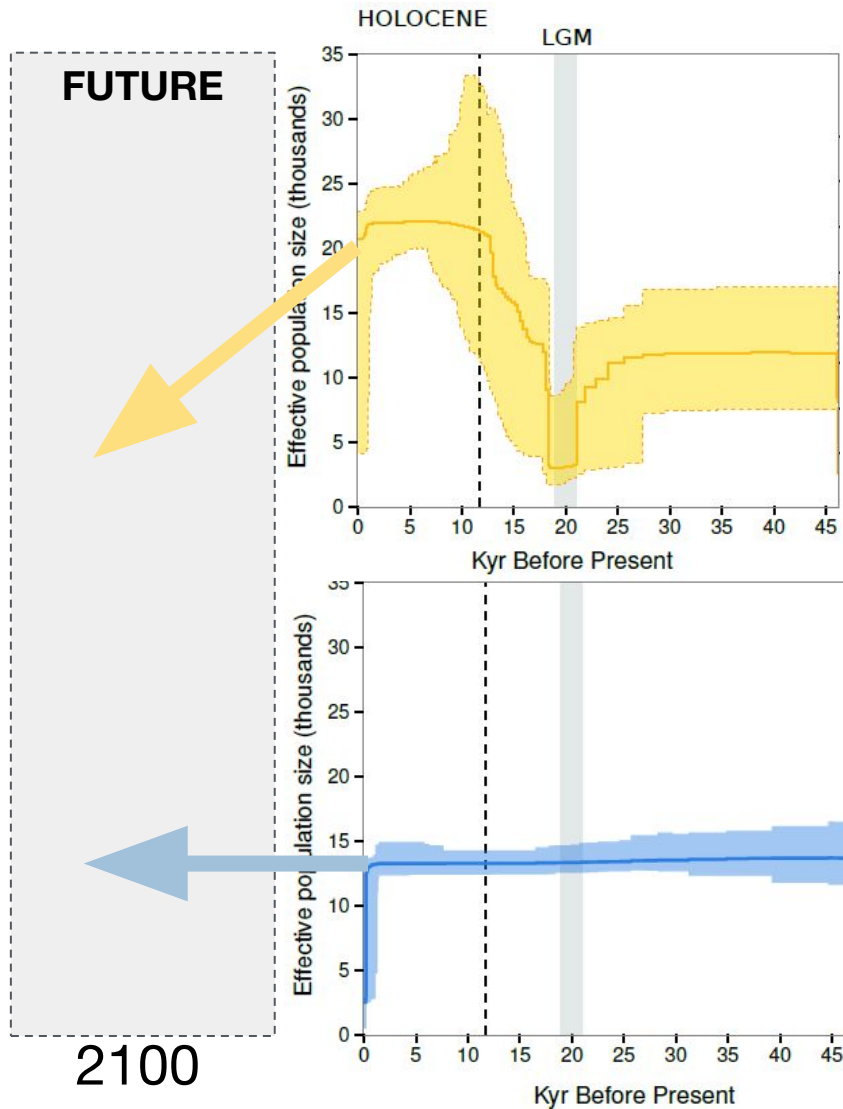
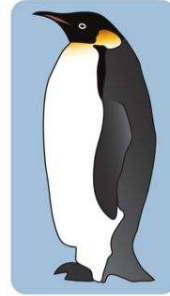
Long-range dispersal is not an issue

About **35,000 loci** genotyped in ca. **110 individuals** from **6 colonies**.

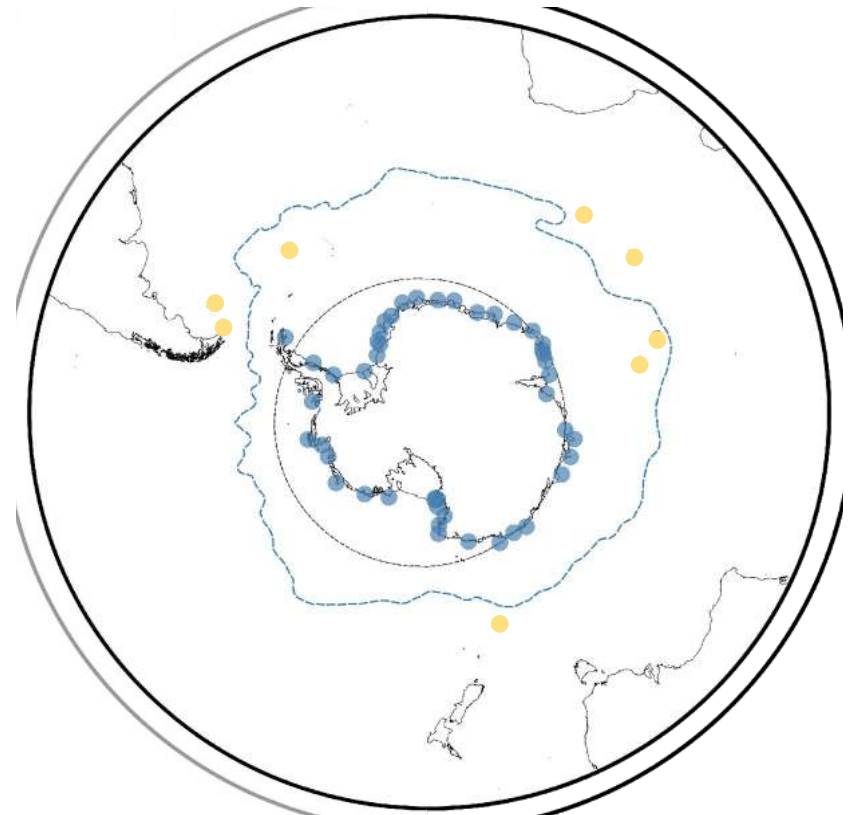
Migration rate (around) and F_{st} (inside) estimates

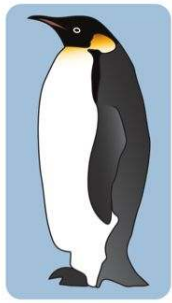


Conclusions



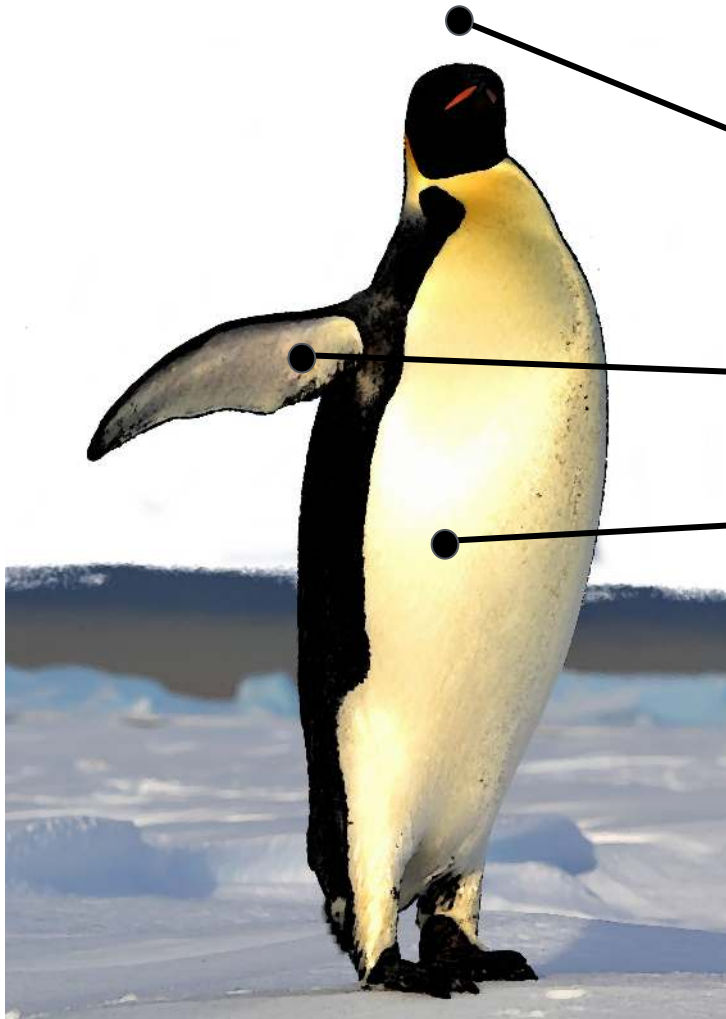
Habitat fragmentation more than extreme-cold adaptation drives response to **global warming**





Research in progress

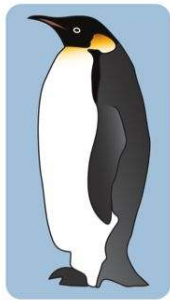
Understanding the genetic basis of the adaptations to the coldest environment on Earth



Behaviour

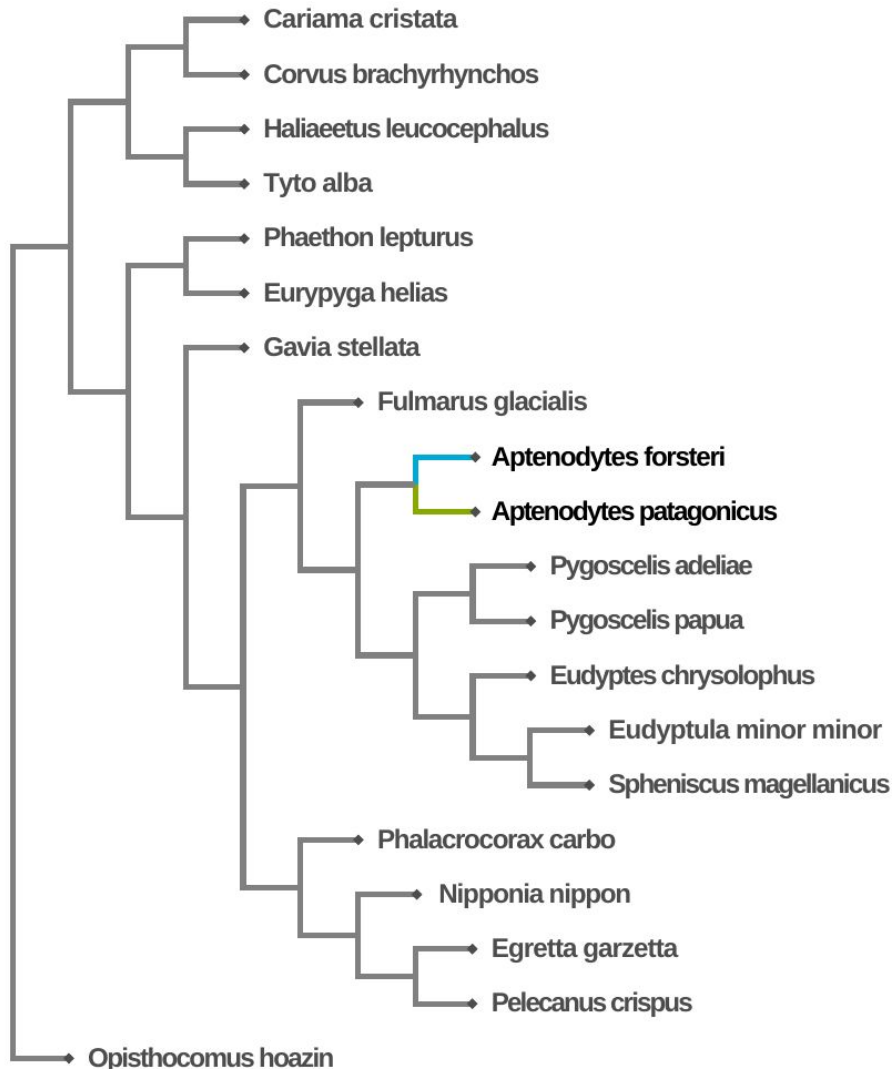
Insulation

Metabolism



Research in progress

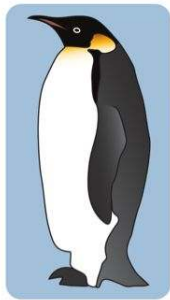
Understanding the genetic basis of the adaptations to the coldest environment on Earth



Long term selection

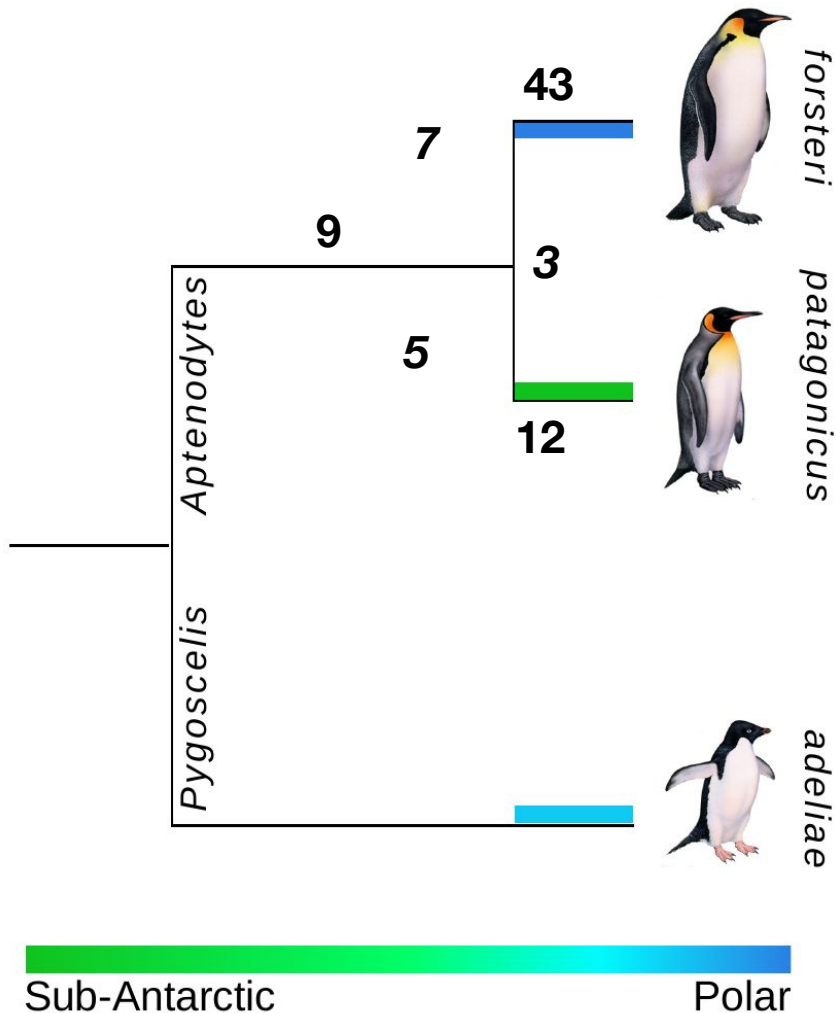
genes with significantly different dN/dS along the target branches

Using published reference genome CDS data within a custom pipeline to **find orthologs and run PAML**



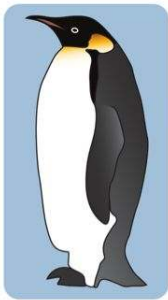
Research in progress

Understanding the genetic basis of the adaptations to the coldest environment on Earth



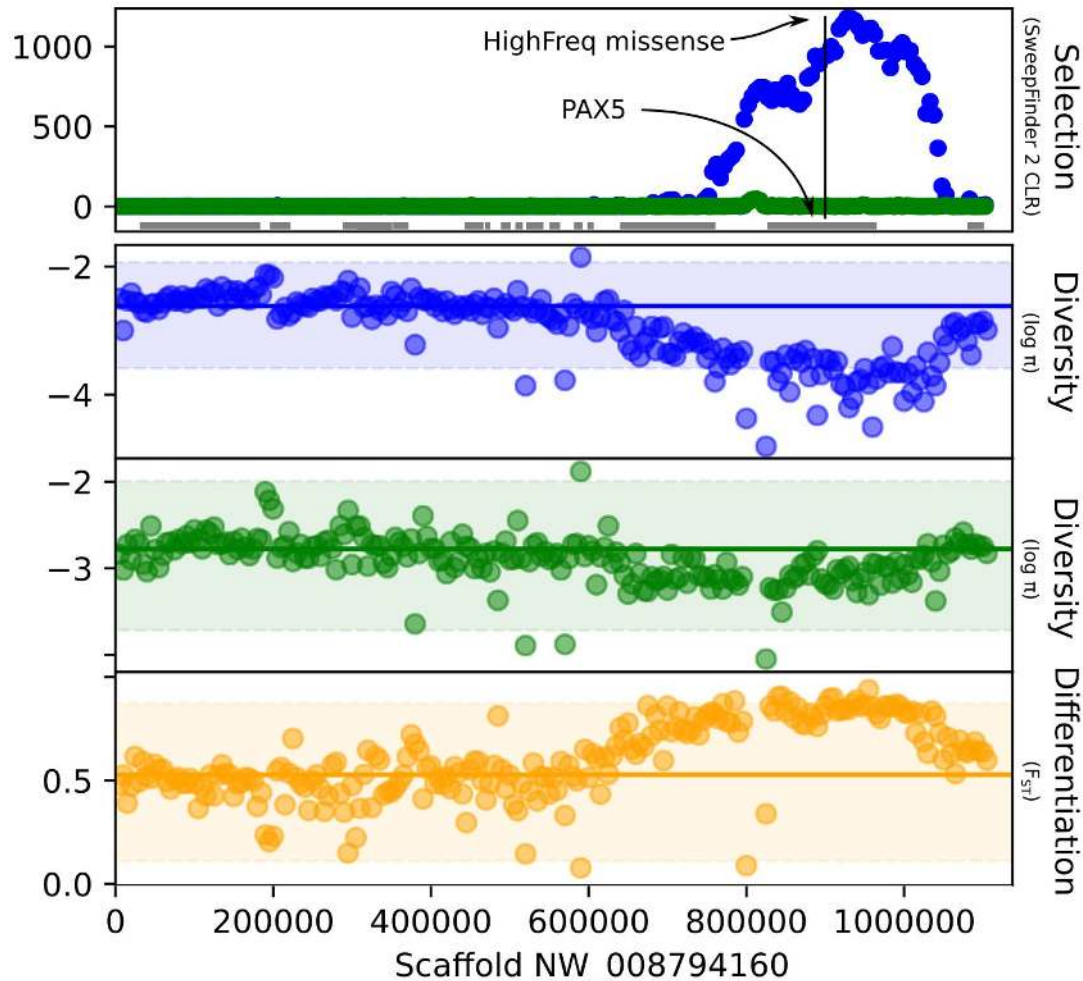
Long term selection

Testing ca. 7,500 orthologs, 3.5X more genes show a signal in the Emperor branch



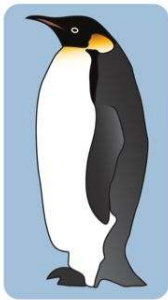
Research in progress

Understanding the genetic basis of the adaptations to the coldest environment on Earth



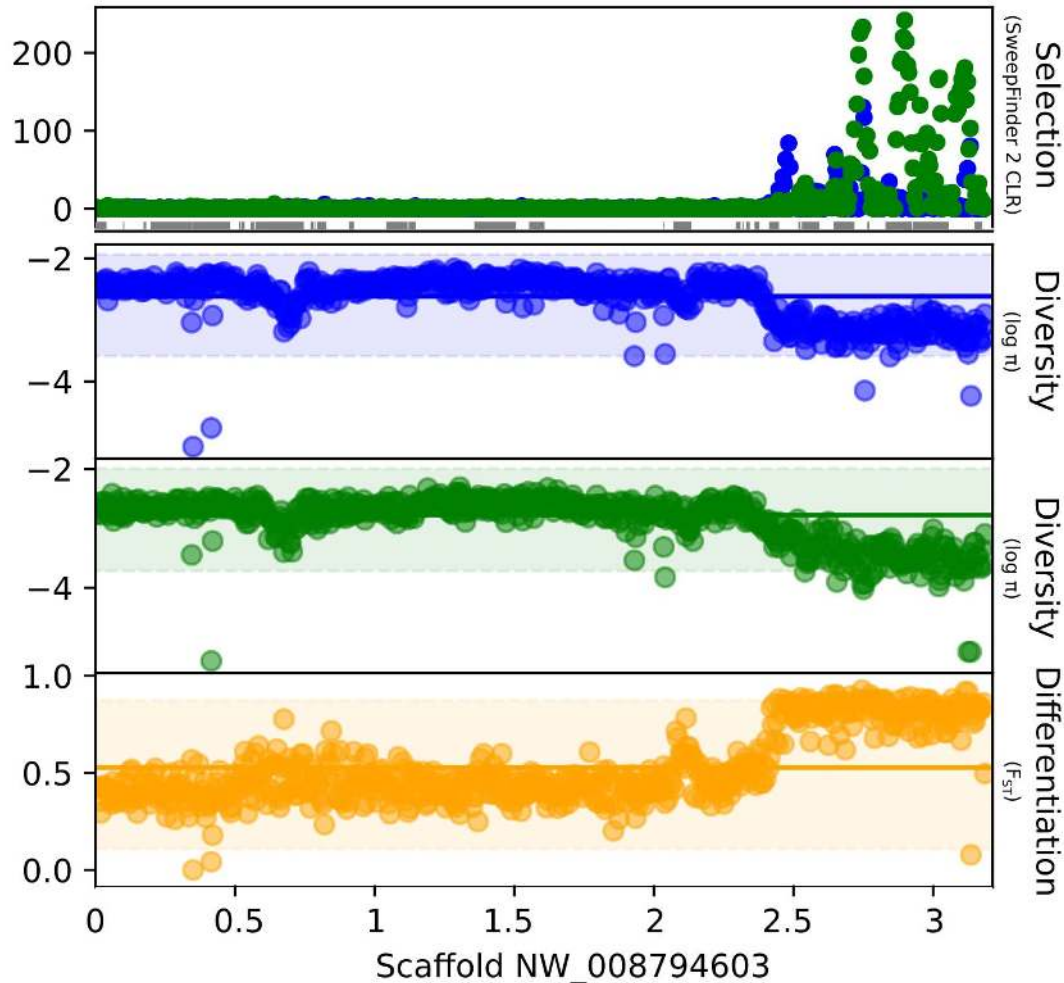
Clear differential selection signature in the Emperor (blue) vs the King (green) penguin

Whole-genome sequencing of 24 Emperor, 24 King, 3 Adelie and 3 Gentoo for **selection scan** with different approaches



Research in progress

Understanding the genetic basis of the adaptations to the coldest environment on Earth



Overlapping selection signature in the Emperor (blue) vs the King (green) penguin

How **recombination rate** landscape constraints adaptation?

Acknowledgements

Robin Cristofari



Céline Le Bohec



Andrea Benazzo, Giorgio Bertorelle, Francesco Bonadonna, Yves Cherel, Yvon Le Maho, Xiaoming Liu, Virginie Raybaud, Pierre Pistorius, Paul J. Ponganis, Nils Chr. Stenseth, Phil N. Trathan, Jason D. Whittington, Enrico Zanetti, Daniel P. Zitterbart

Cristofari et al 2016 Full circumpolar migration ensures evolutionary unity in the Emperor penguin. *Nature communications* 7
Cristofari et al 2018 Climate-driven range shifts of the king penguin in fragmented ecosystems. *Nature Climate Change* 8, 245-251
Trucchi et al 2019 Reply to: 'The role of ocean dynamics in king penguin range estimation'. *Nature Climate Change* 9, 122





BONUS TRACK

Ancient genomes reveal early farmers selected common beans while preserving diversity

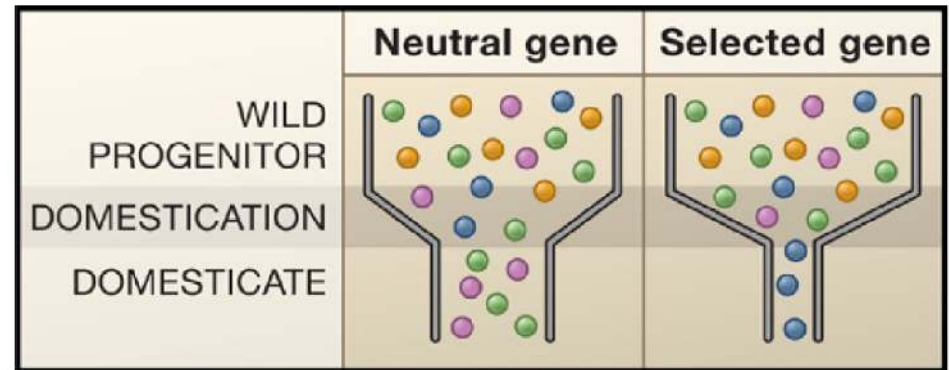


Understanding tempo and mode of domestication



Selection on traits of interests

Genetic **drift** due to
the domestication bottleneck



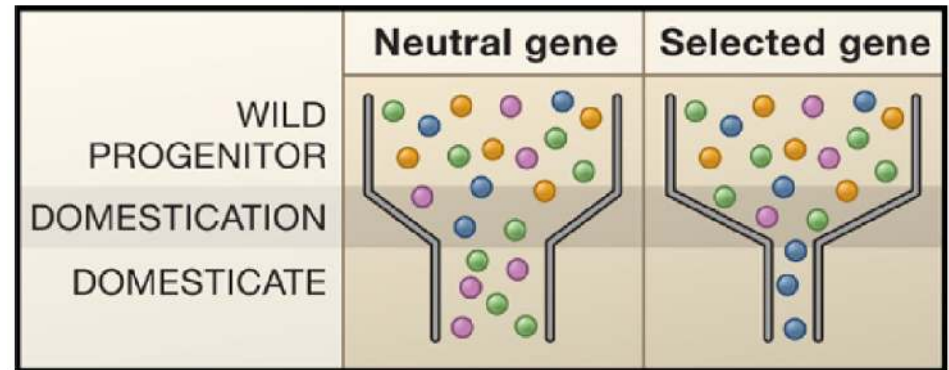
Doebley and Gaut 2006

Understanding tempo and mode of domestication



Selection on traits of interests

Genetic **drift** due to
the domestication bottleneck

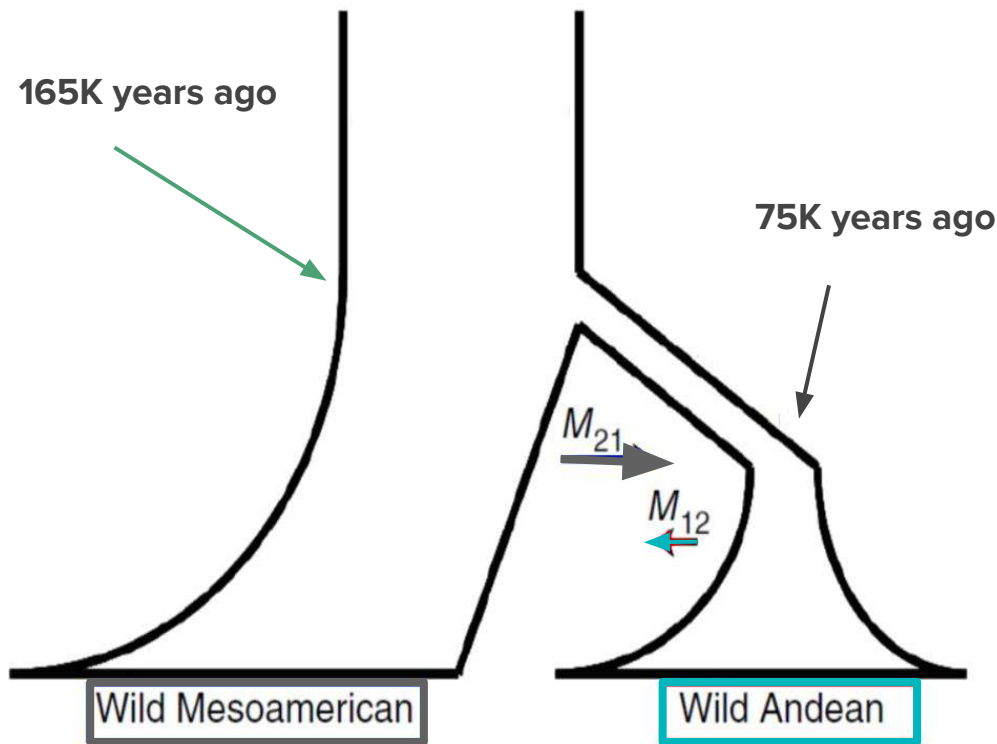


Doebley and Gaut 2006

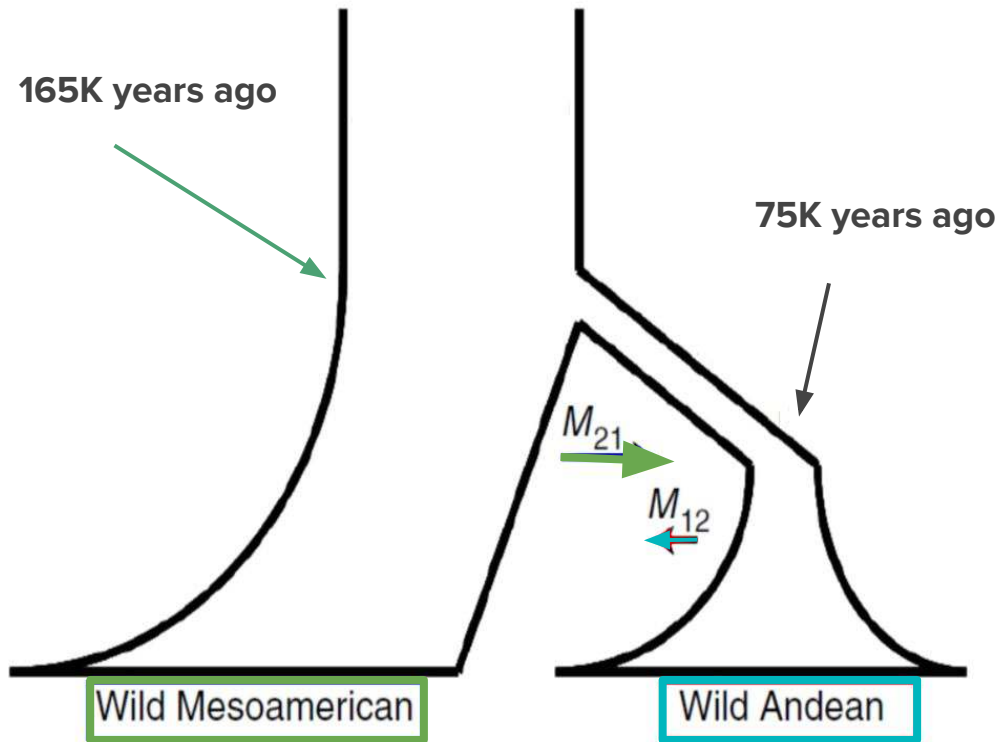
Which is the underlying genetics of selected phenotypic traits?

What is the temporal dynamics of both selection and loss of genetic diversity (genomic erosion)?

Common bean double independent domestication in meso and south America



Common bean double independent domestication in meso and south America



Schmutz et al 2014



Bitocchi et al 2012; Gaut 2014

Ancient genomics focus on south America domestication



5 museum collections

7 geographic area

9 archaeological sites
in **North-West Argentina**

150 bean seeds collected
(and 16 pod fragments)



Ancient genomics focus on south America domestication

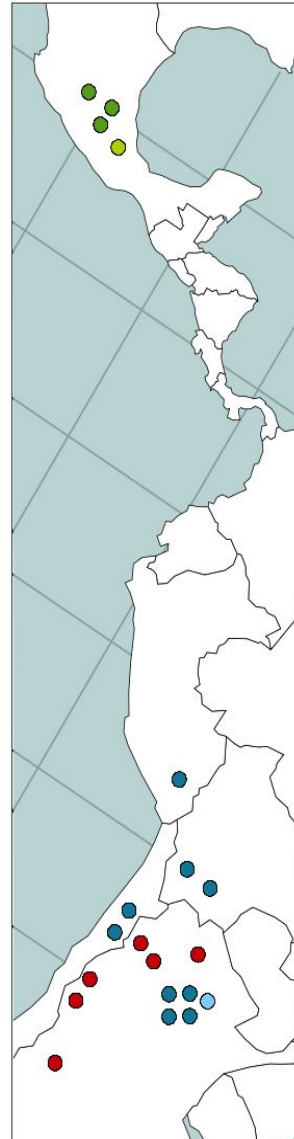


Preliminary low coverage sequencing of 30 ancient seeds

15 ancient seeds whole-genome sequenced at **4-18X** coverage

14 modern seeds, wild and domesticated

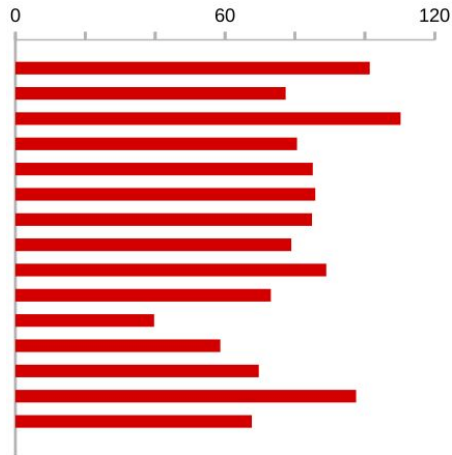
Domesticated
Ancient
Modern andean
Modern mesoamerican
Wild
Andean
Mesoamerican



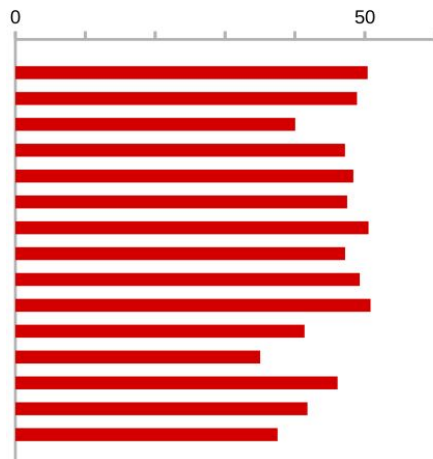
High quantity and quality of endogenous DNA in ancient common bean seeds



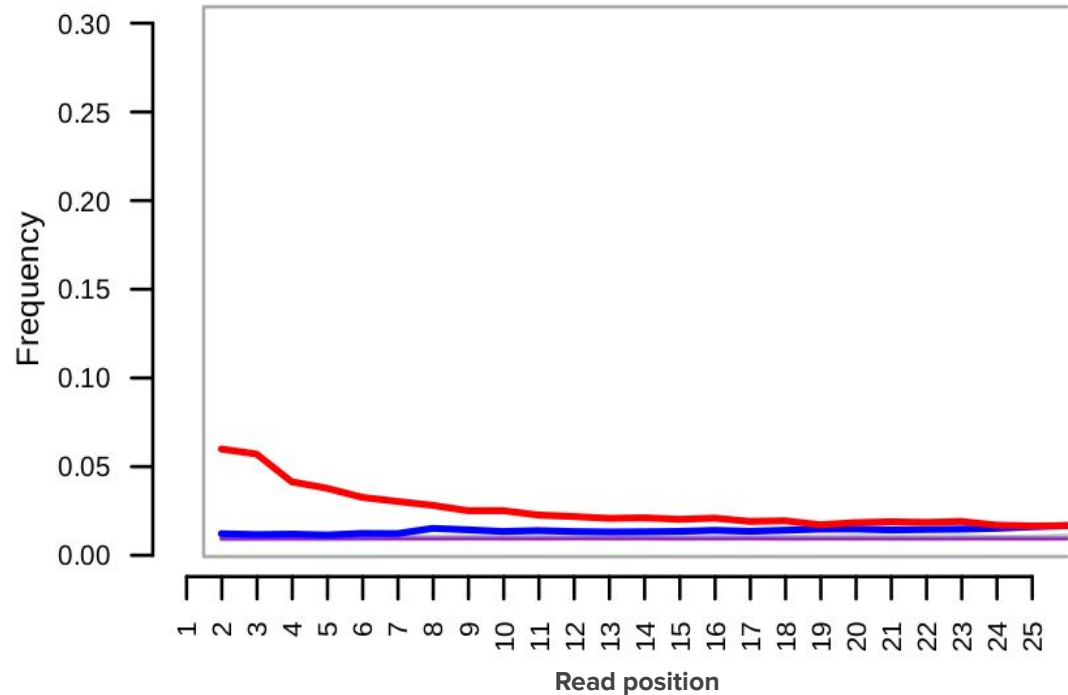
Average read length (bp)

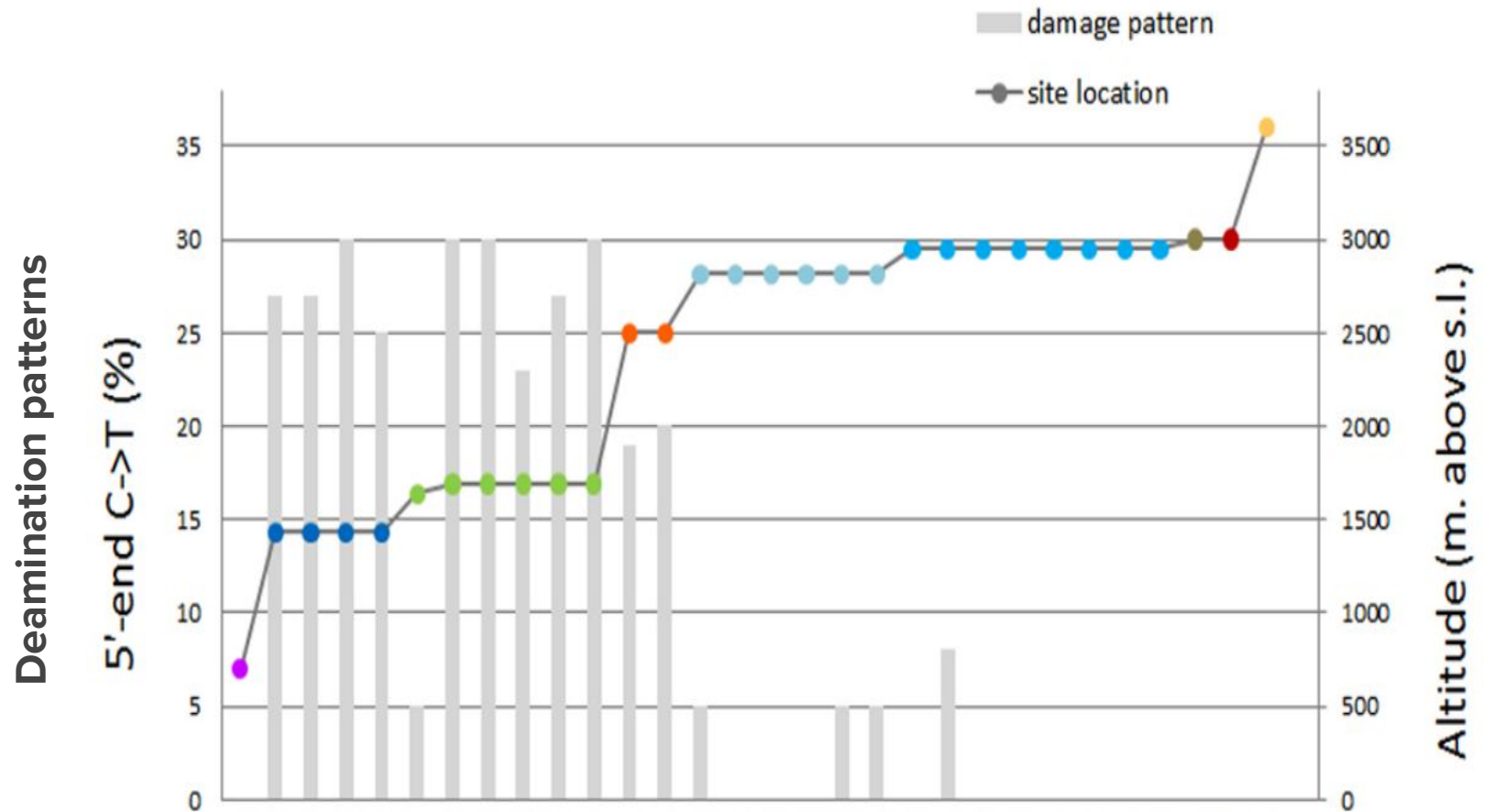


Endogenous DNA (%)

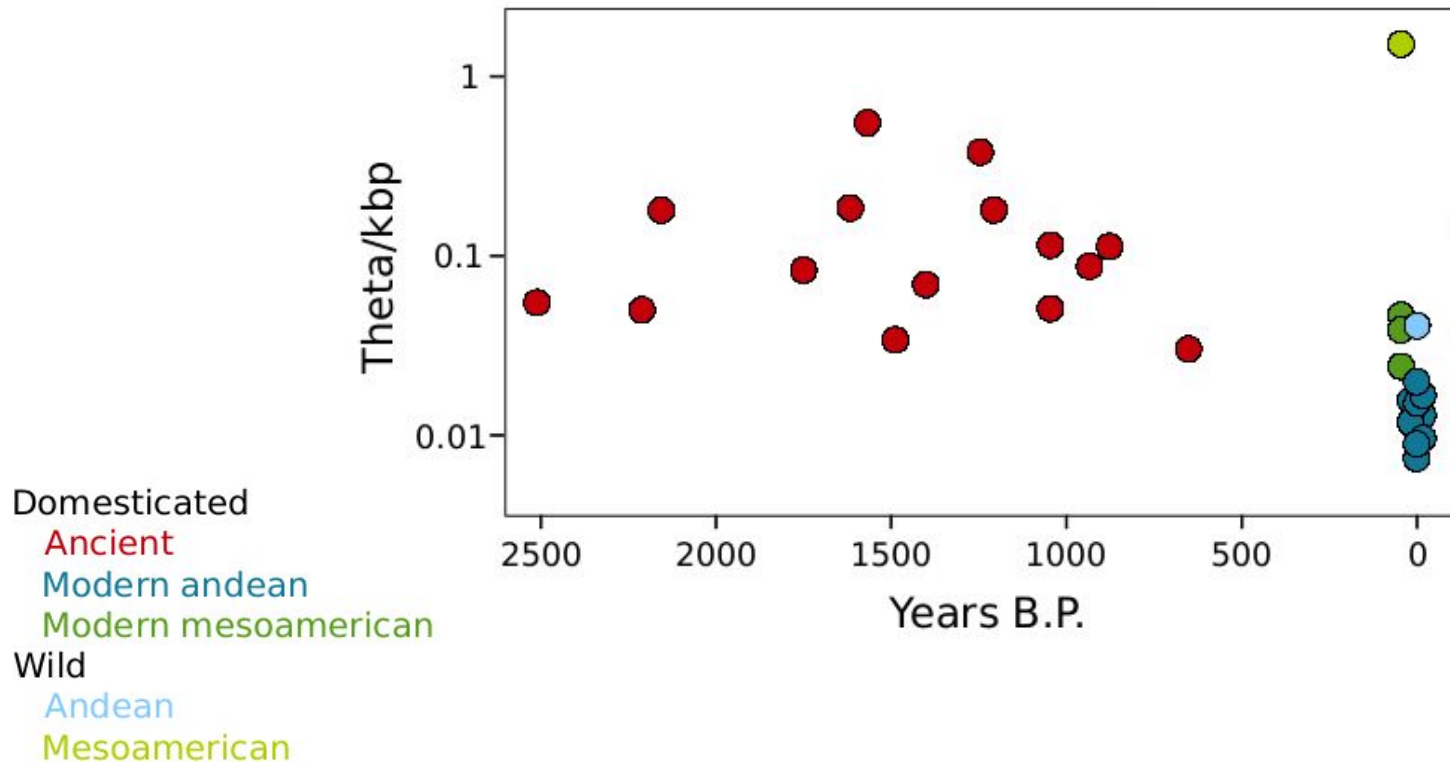


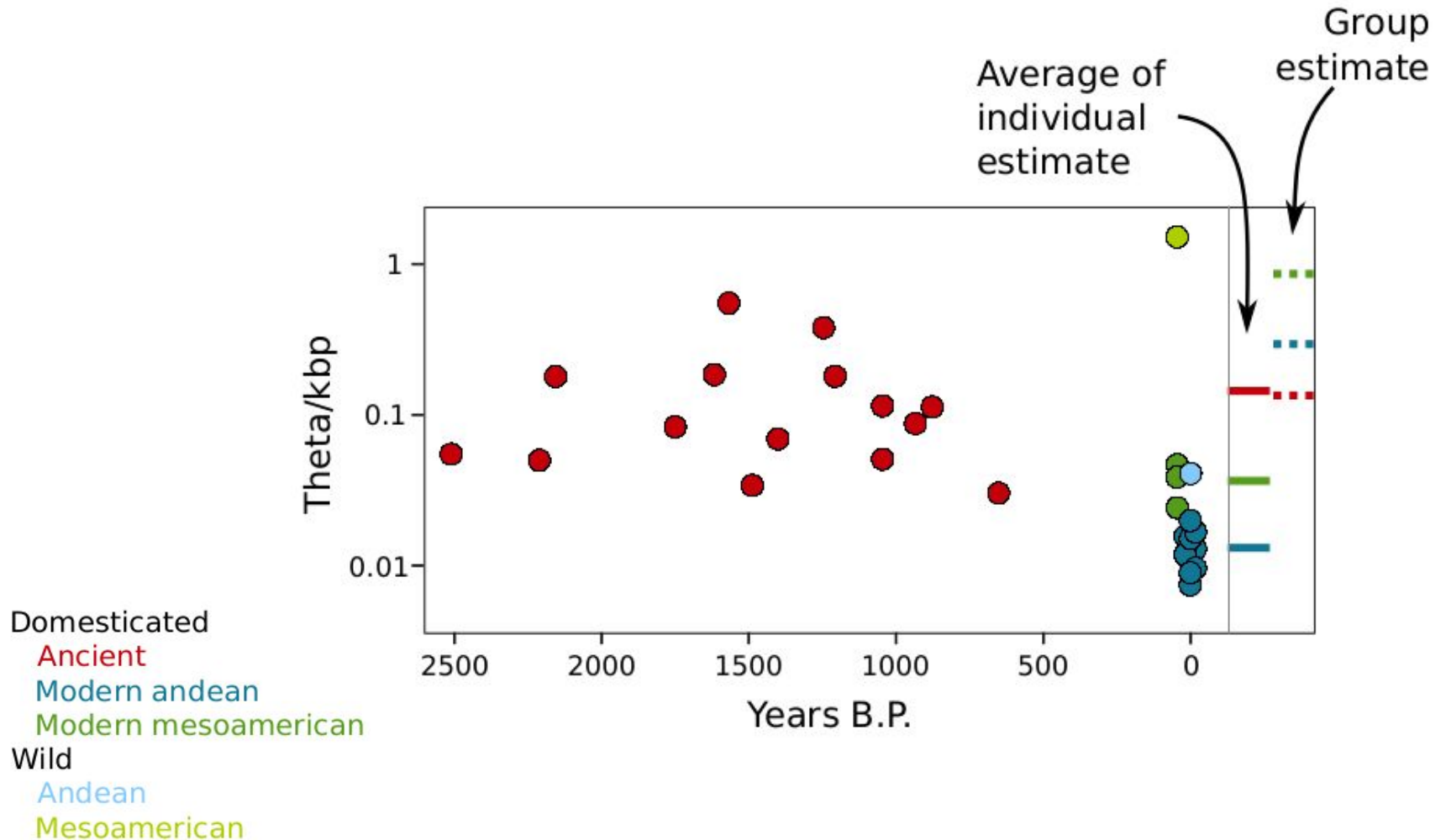
Deamination patterns typical of ancient DNA



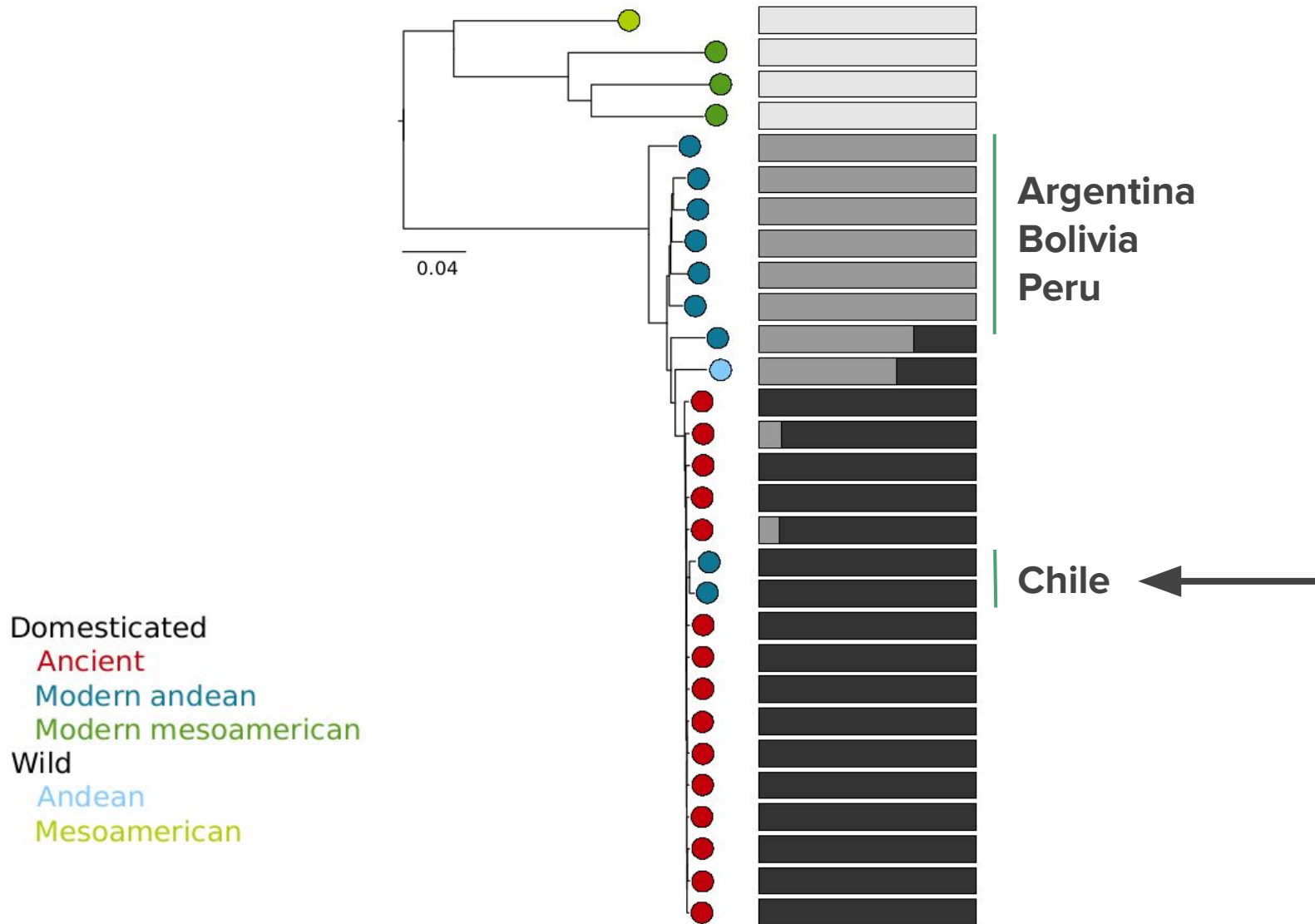


Loss of genetic diversity is recent in common bean domestication





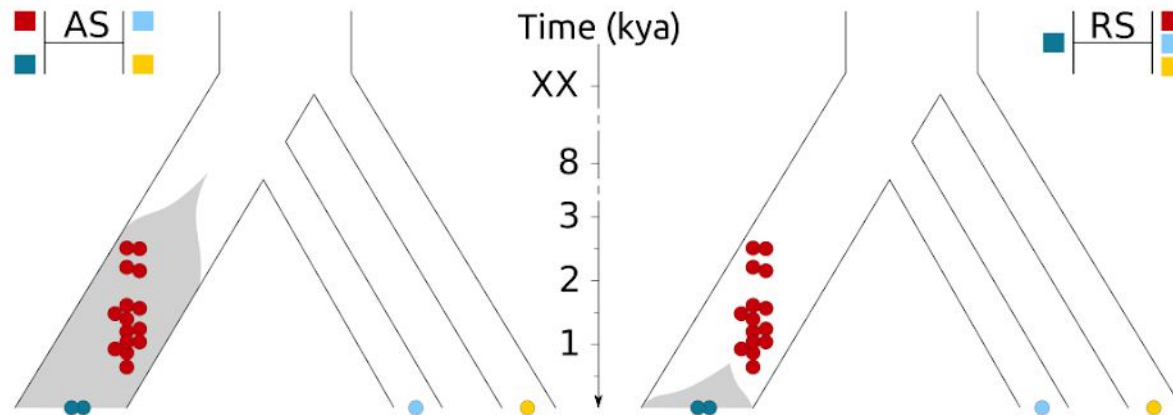
All ancient seeds belong to the same genomic clade



Gene-by-gene scan of enrichment in fixed alternative alleles



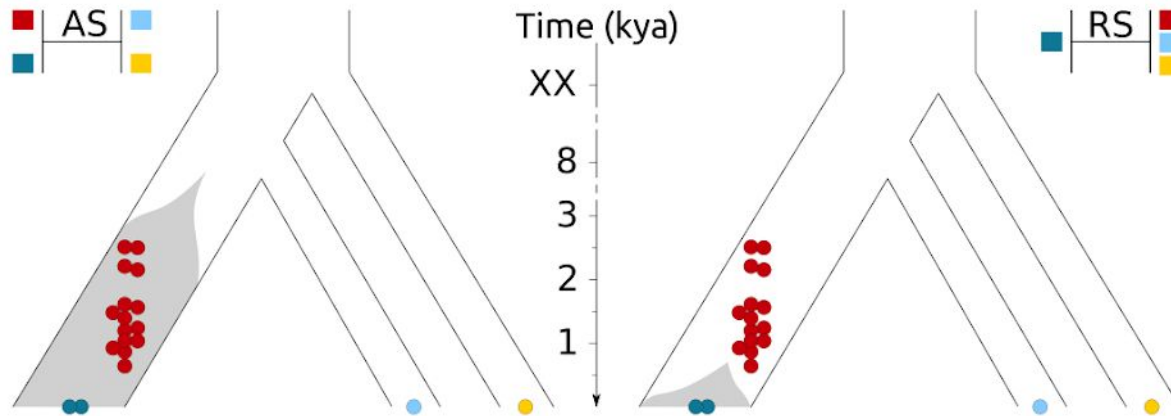
Timing of selection



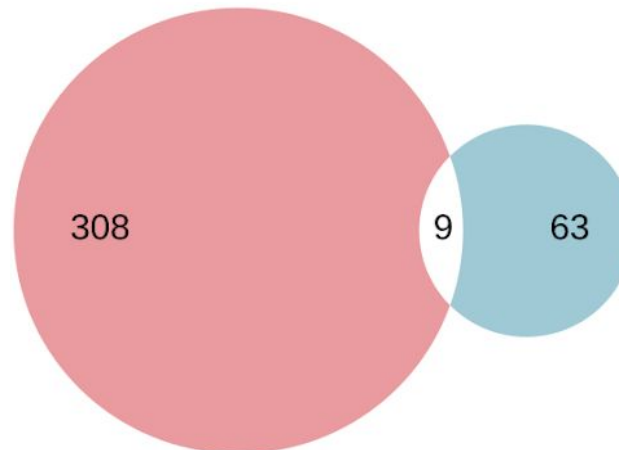
Selection affected 4.5X more genes in ancient than in recent times



Timing of selection



Selected genes



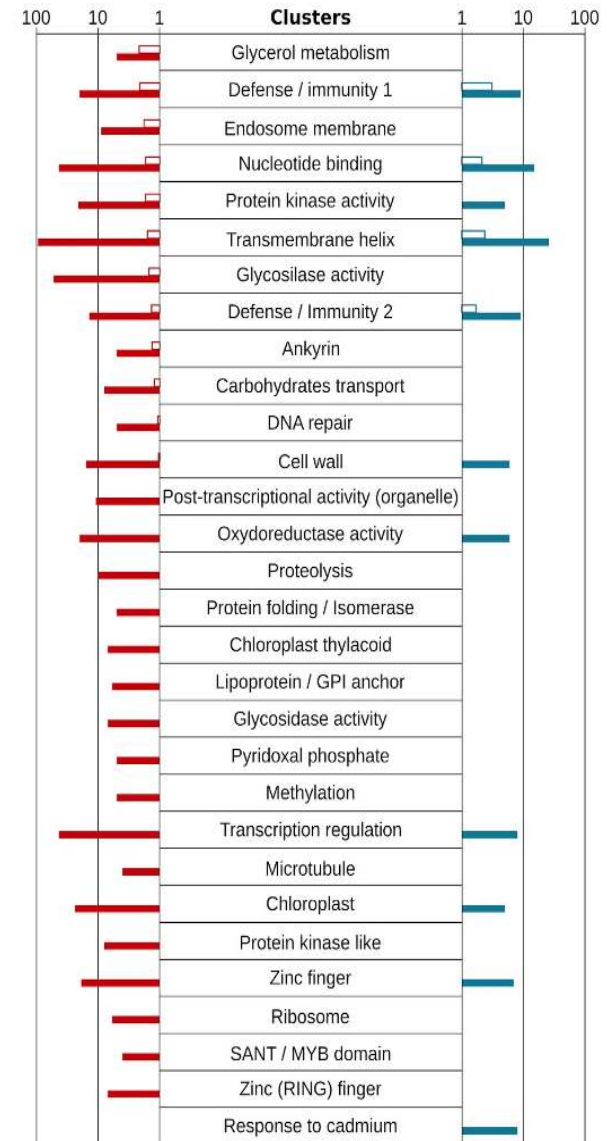
Different functional gene groups were selected by ancient vs. recent selection



Ancient selection targets: glycerol metabolism, carbohydrate and sugar transport and metabolism, intracellular transport, regulatory elements, modification of proteins, glycosylation

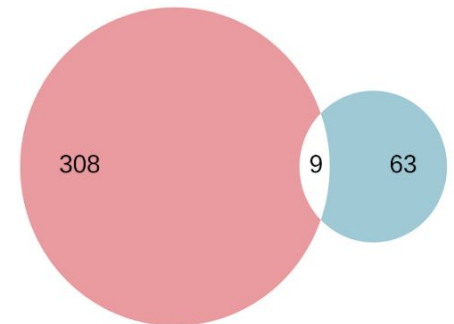
Recent selection targets: immunity and defense, regulatory elements and transmembrane transport

A few immune genes show signature of both ancient and recent selection



Seeds or cultivars **exchange** was common

More sustainable than modern breeding programs
(since Green revolution)



Concluding, ancient selection but recent gene erosion characterize bean domestication

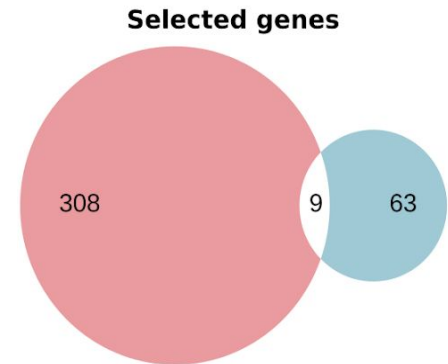
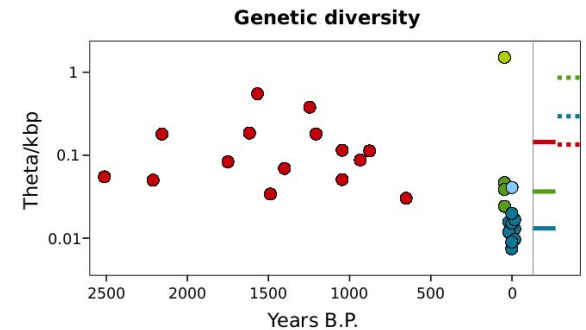


Early selection was probably based on **larger number** of seeds

Seeds or cultivars **exchange** was common

More sustainable than modern breeding programs (since Green revolution)

Modern landraces from Chile are the most similar to the ancient Andean cultivars





Museo di La Plata, La Plata

Museo di Storia Naturale di San Rafael, San Rafael

Museo Prof. M. Gambier, San Juan

Museo di Storia Naturale M. Lillo, San Miguel de Tucuman

Museo Archeologico Pio Pablo Diaz, Chaci



Alice Iob

Andrea Benazzo

Giorgio Bertorelle



Martina Lari

Stefania Vai

David Caramelli



Bastiaan Star

Hugo de Boer

Sanne Boessenkool

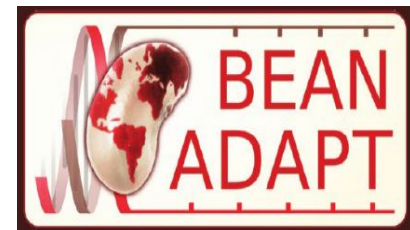


Elena Bitocchi

Elisa Bellucci

Laura Nanni

Roberto Papa



Trucchi et al (2019) Ancient genomes reveal early farmers selected common beans while preserving diversity. **BioRxiv**, <https://doi.org/10.1101/791806>

Take home message?

Any species can be a ***model species***

Analyses are more important than data

Get many ***collaborators***



Enjoy Genomics!



Emiliano Trucchi

Workshop on Genomics 2020
Cesky Krumlov

