

## IV. Conservation genomics

# Conservation biology

## **Conservation biology**

aims to maintenance of biological diversity, protection of species from extinction, preservation of their habitats, and protection of ecosystems against destruction

## **Conservation genetics**

application of genetic methods in conservation biology

## **Conservation genomics**

part of conservation genetics dealing with genomic data

# IUCN RedList

## IUCN

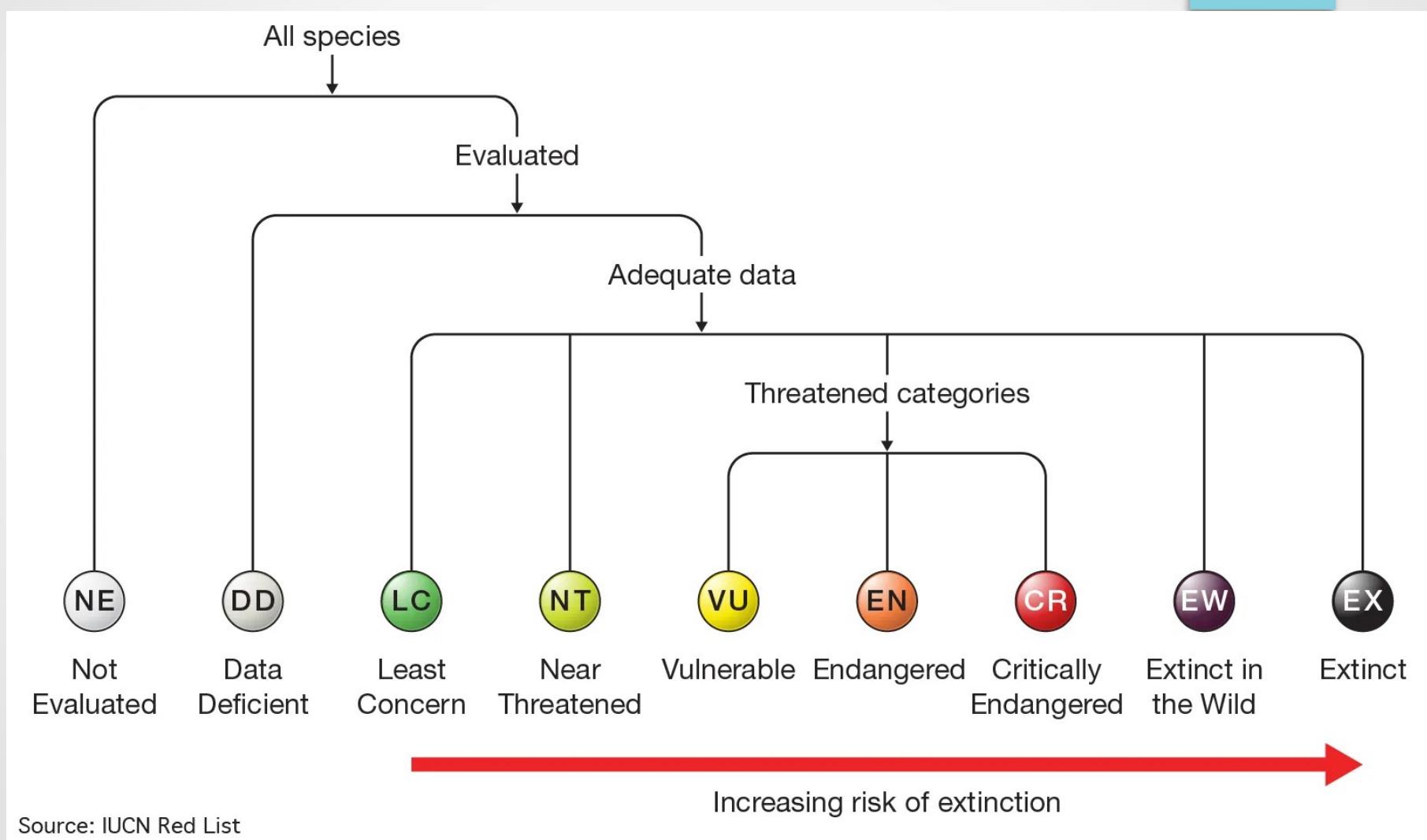
- International Union for Conservation of Nature
- Established in 1948

## IUCN Red List

- Established in 1964
- <https://www.iucnredlist.org>
- Contains information for animal, fungi and plant species on:
  - global conservation status and treats
  - range and habitat
  - ecology
  - 150,388 species assessed up to date

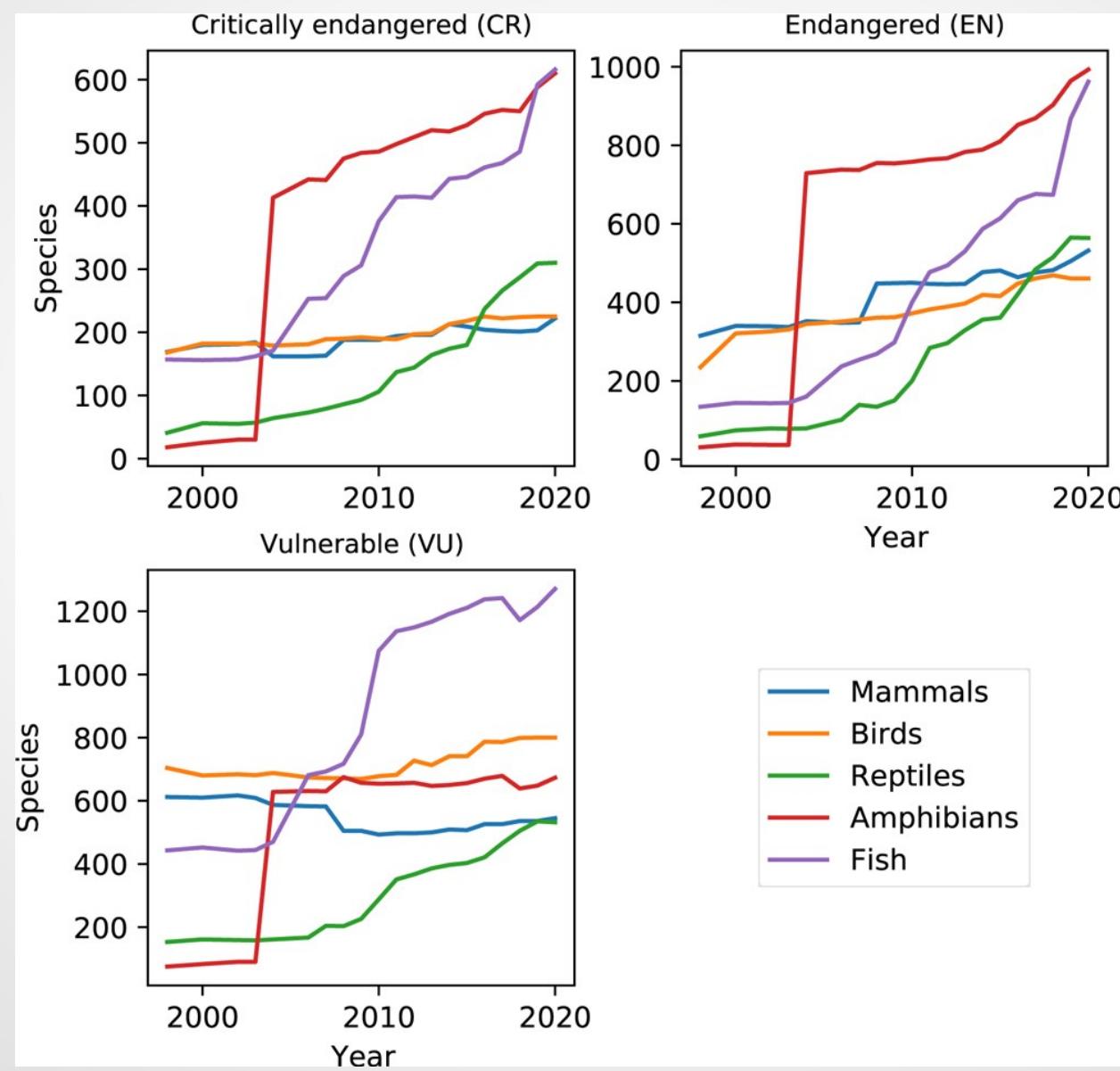


# IUCN RedList categories

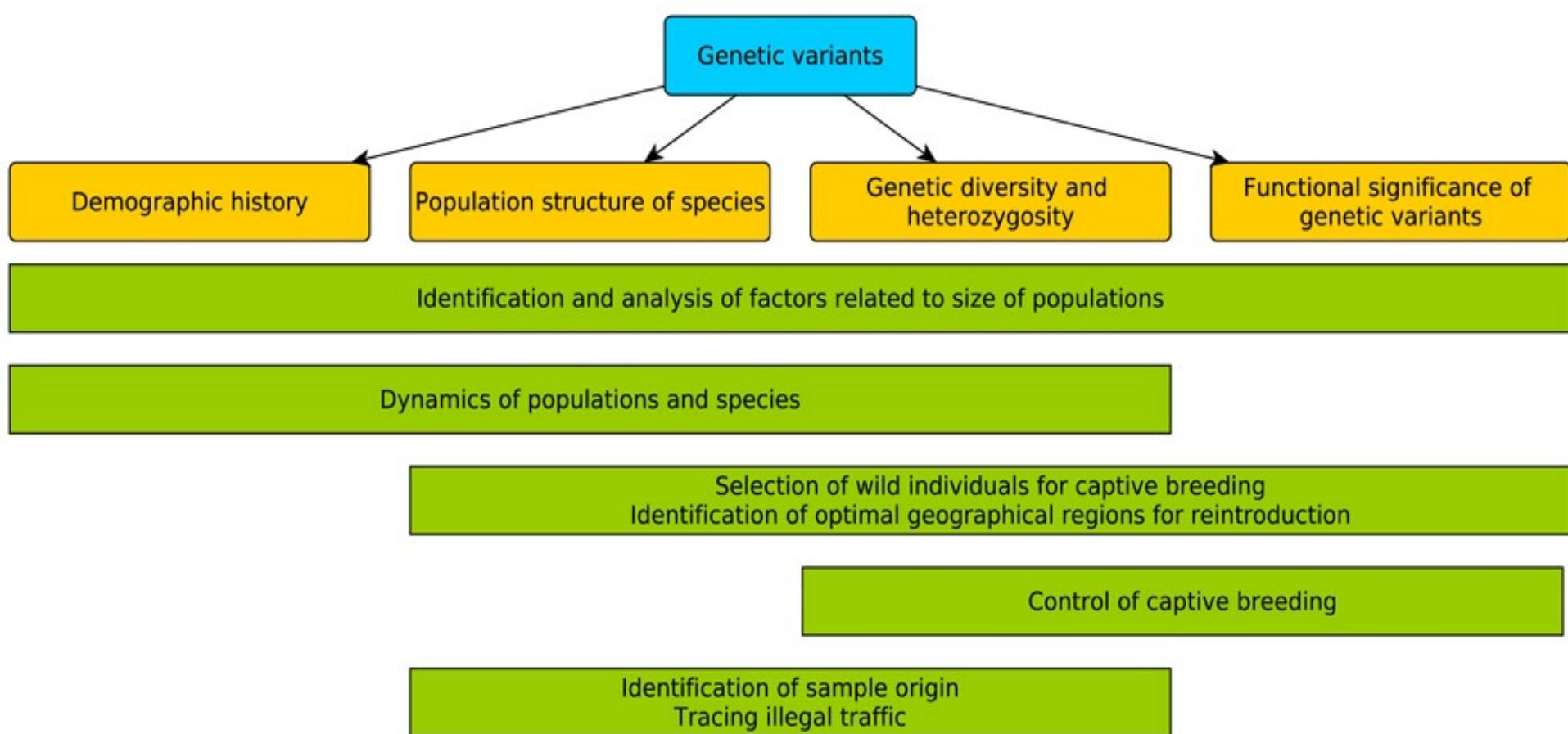


Source: IUCN Red List

# 6<sup>th</sup> mass extinction



# Significance of genetics for conservation



# Birth of conservation genomics



## First publication

Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii*

(Tasmanian devil)

Miller et al, 2011

<https://animals.sandiegozoo.org/animals/tasmanian-devil>

## Main features of first publication

- *de novo* assembly of reference genome for nonmodel species
- genetic diversity analysis on whole genome level
- development of test panel using whole genome data to genotype multiple samples and analyze population structure
- search for candidate deleterious mutations associated with diseases

# Main problems of conservation genomics

- I. Lack of samples
- II. Low budgets
- III. Sometimes big and complex genomes

# Approaches for big datasets

## Reduced-representation methods

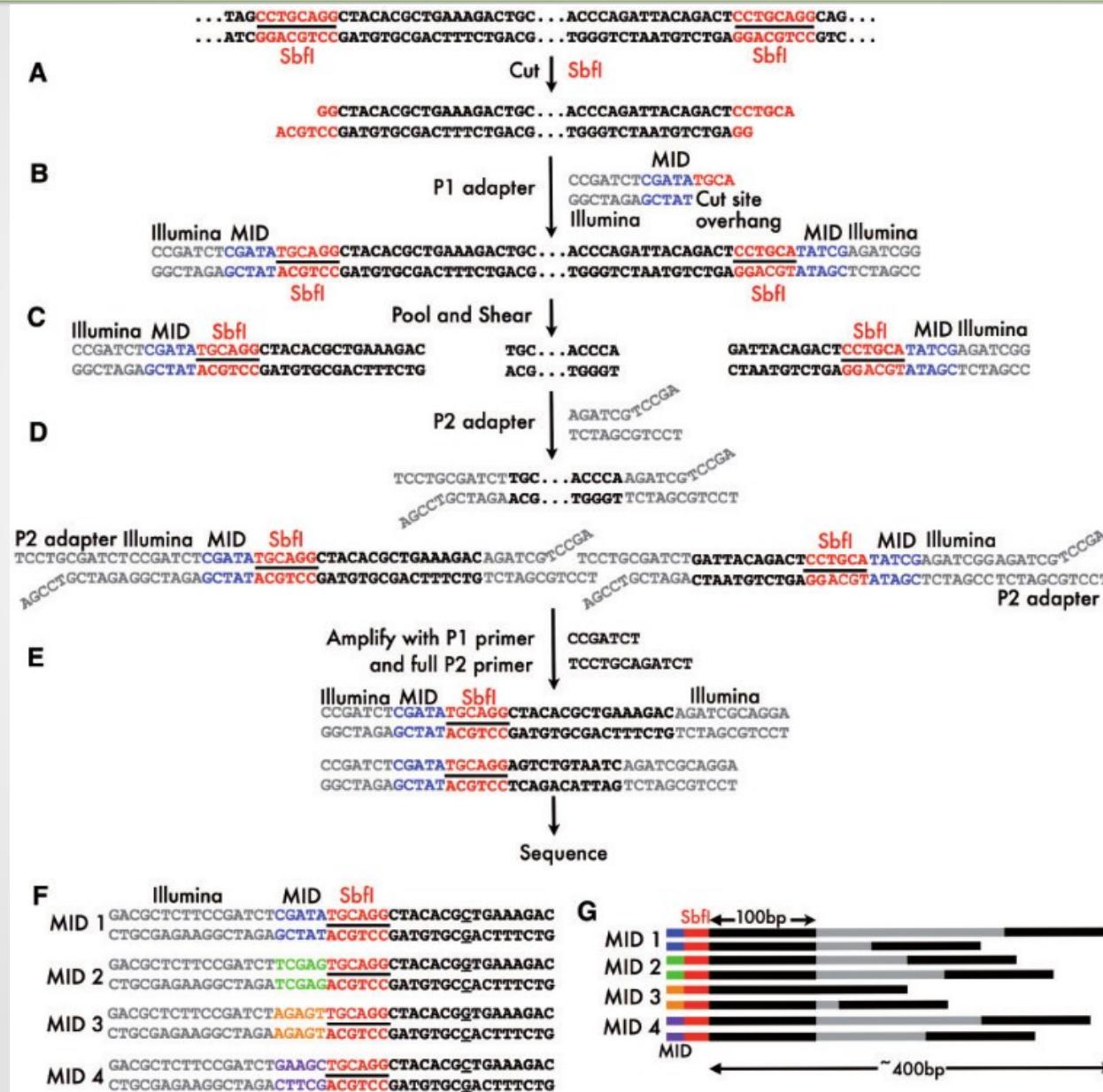
- Hybrid capture arrays
- RADseq
- Genotyping-in-thousands
- Microarrays

## Full representation methods

- WGS (resequencing)
- Pool-seq
- Ultra low coverage WGS + variant imputation

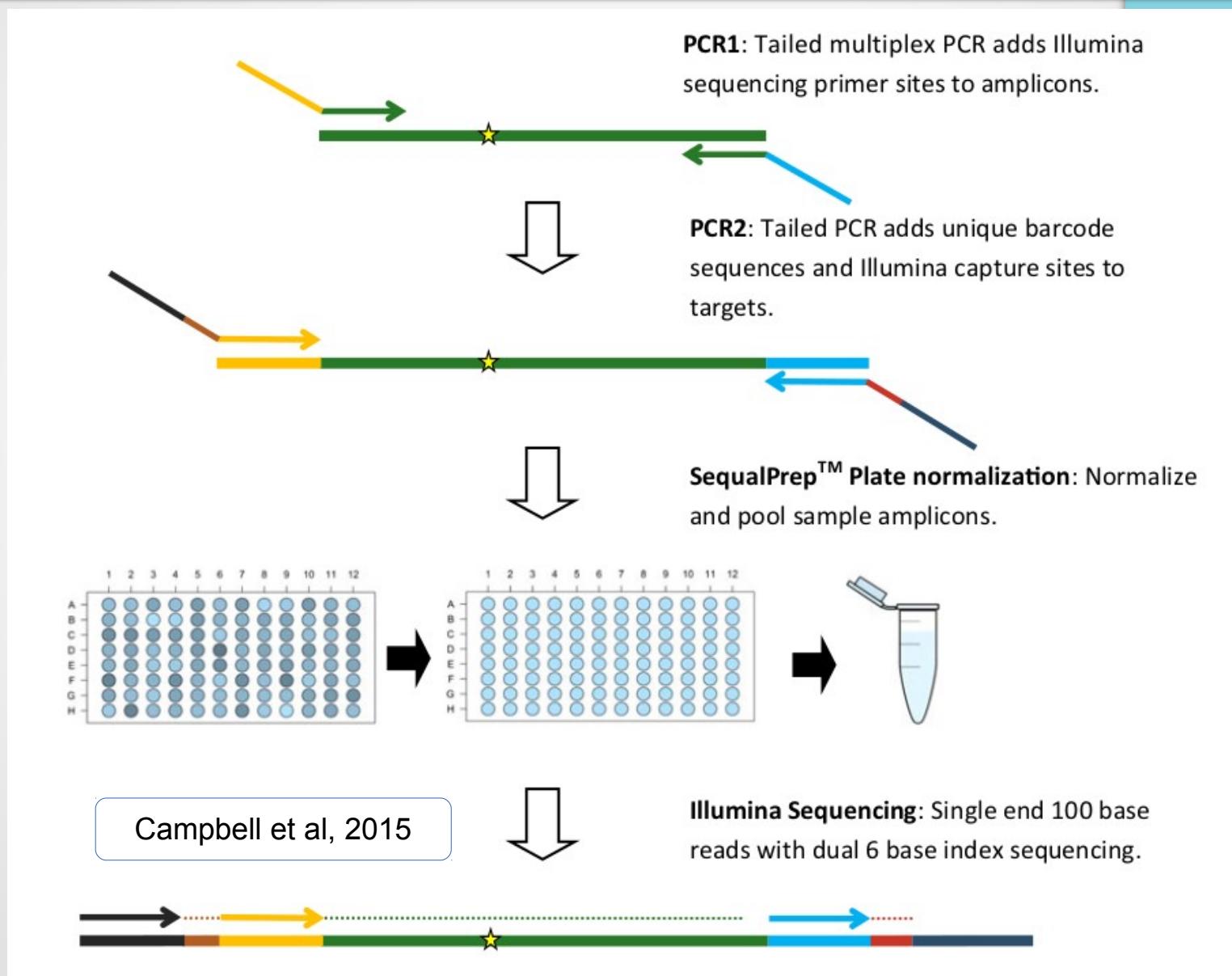
# RADseq

## Restriction site-associated DNA sequencing

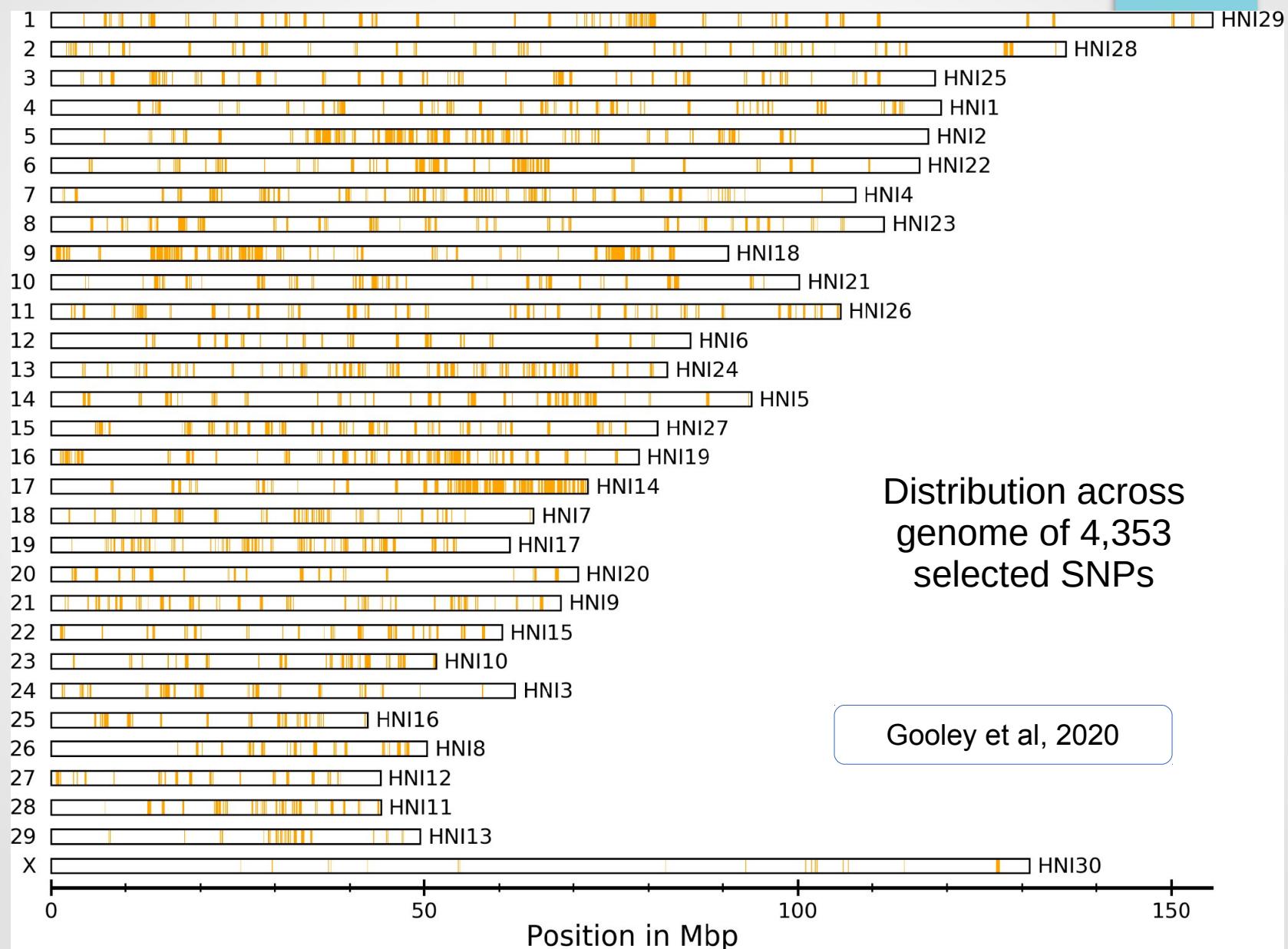


Davey and Blaxter, 2010

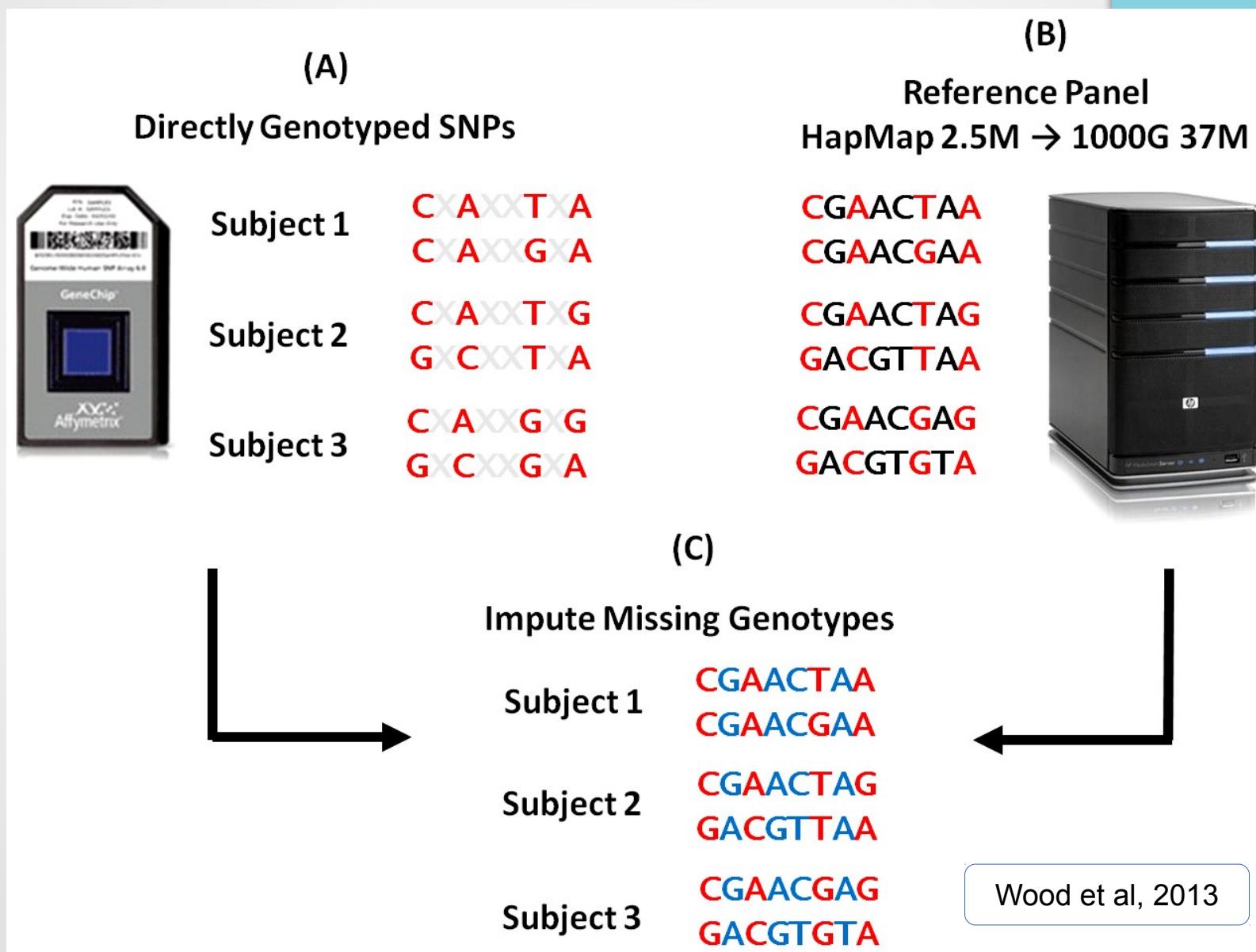
# Genotyping-in-thousands



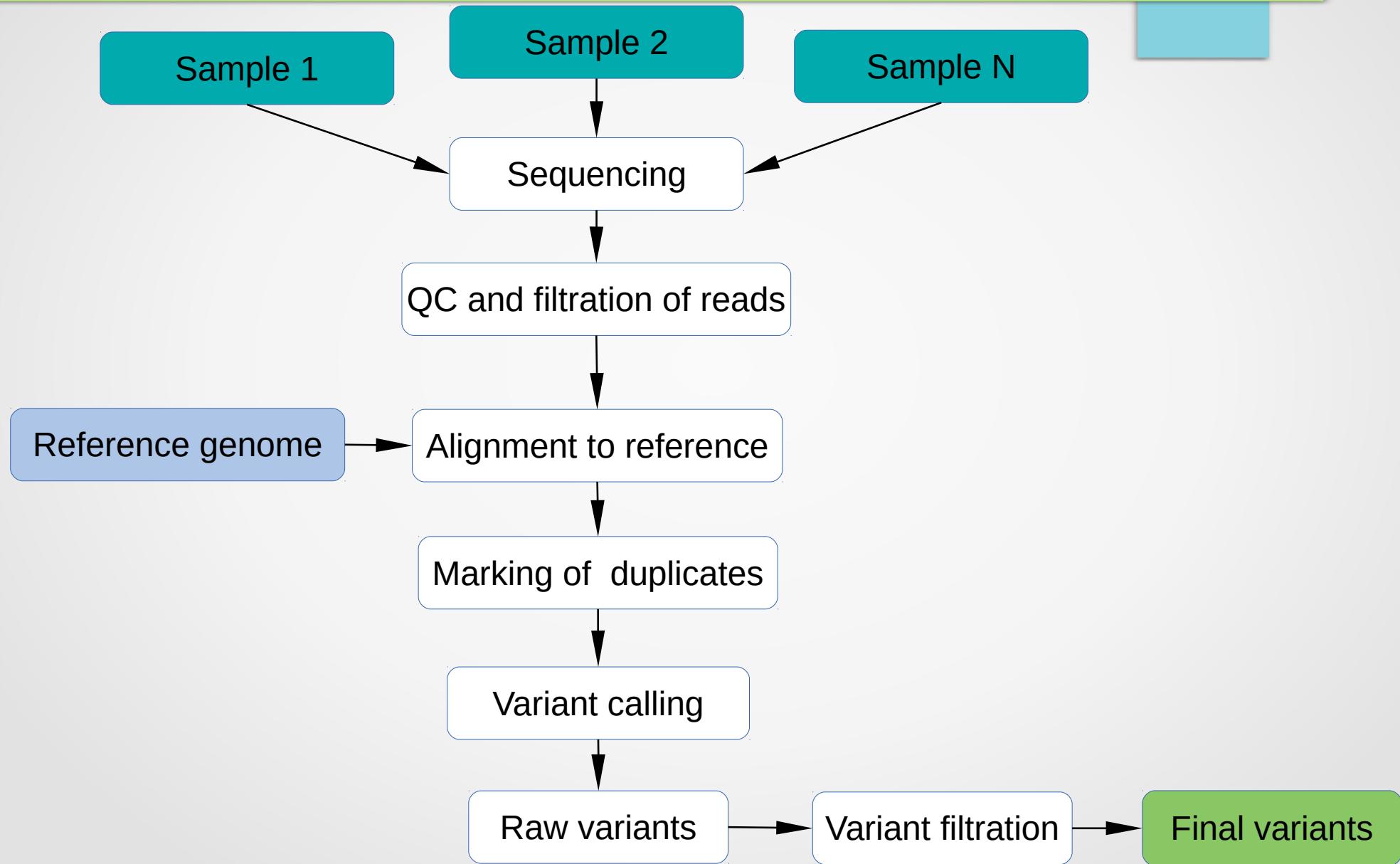
# Hybrid-capture array (sable antelope case)



# SNP imputation



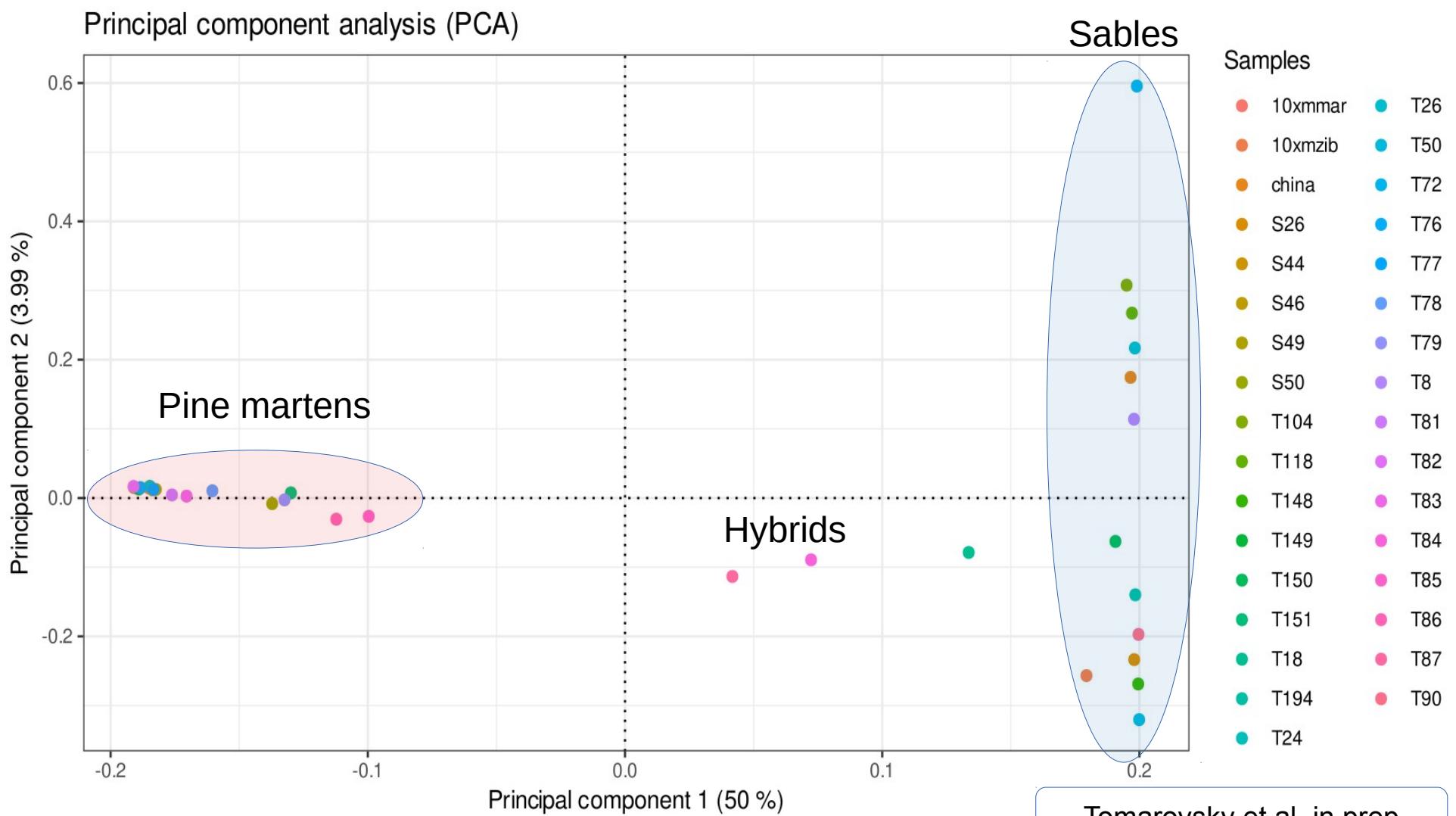
# Pipeline for detecting variants from sequencing data



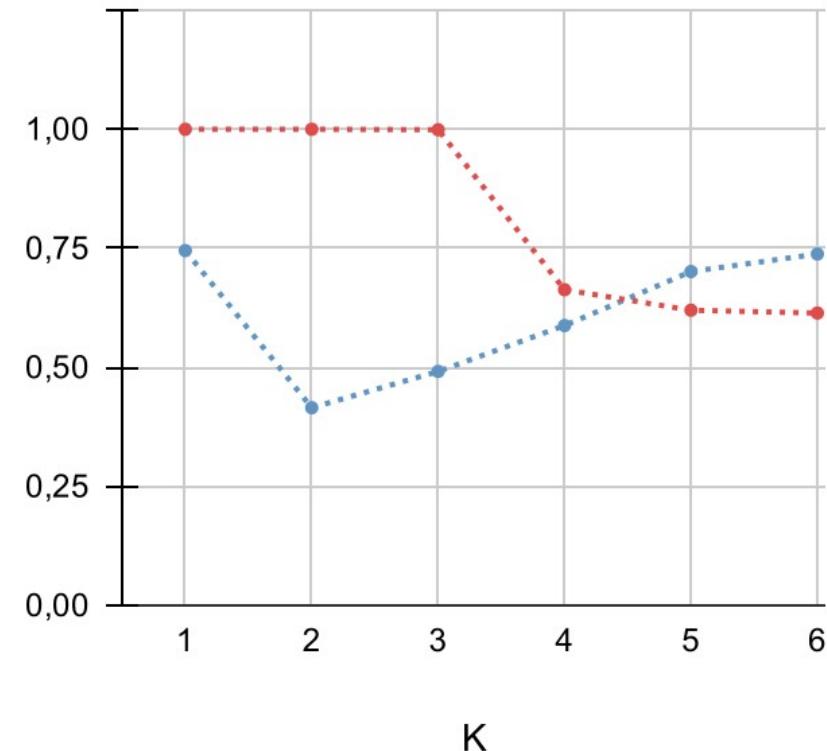
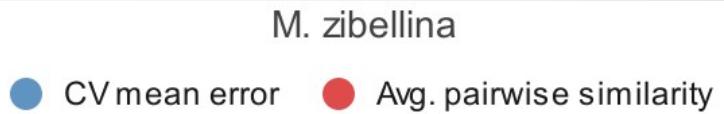
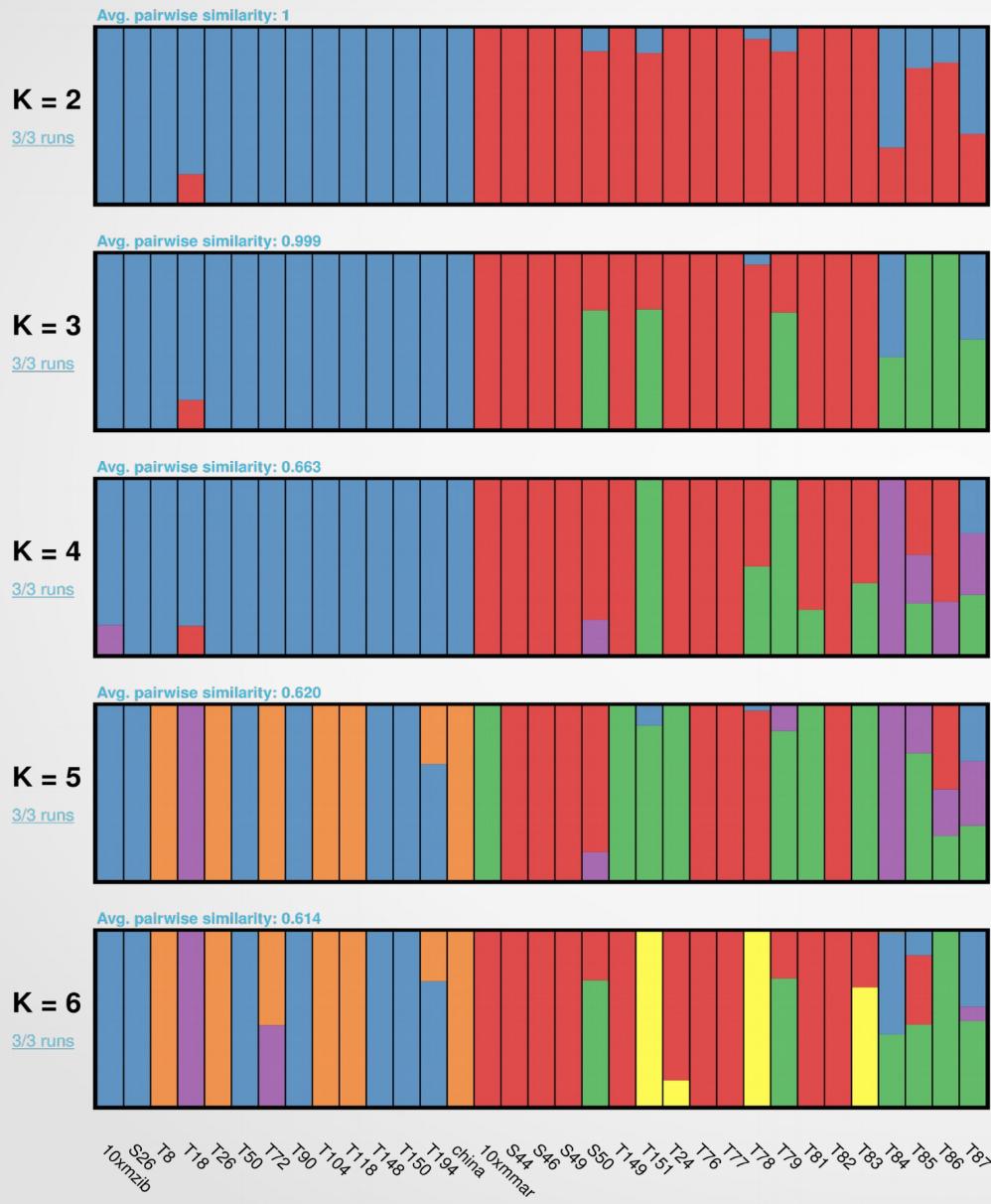
# Analysis of genetic variants

- **Principal component analysis (PCA)**  
from millions SNPs to few principal components
- **Admixture analysis**  
identification of ancestry for individual samples
- **Demographic history analysis**  
reconstruction of trajectory for effective population size in the past
- **Genetic diversity analysis**
- **Identification of deleterious variants**

# Principal component analysis (PCA)



# Admixture

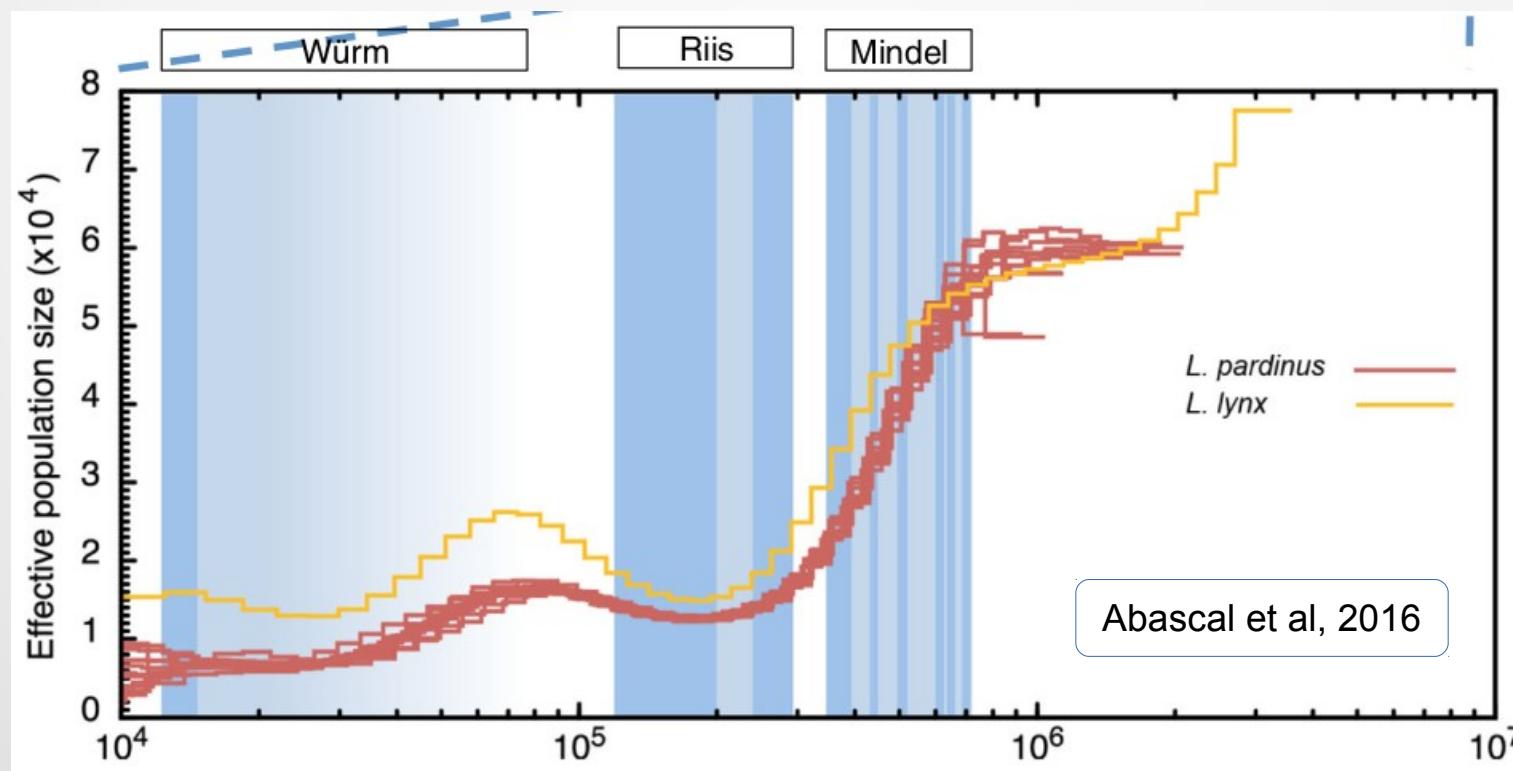


Tomarovsky et al, in prep

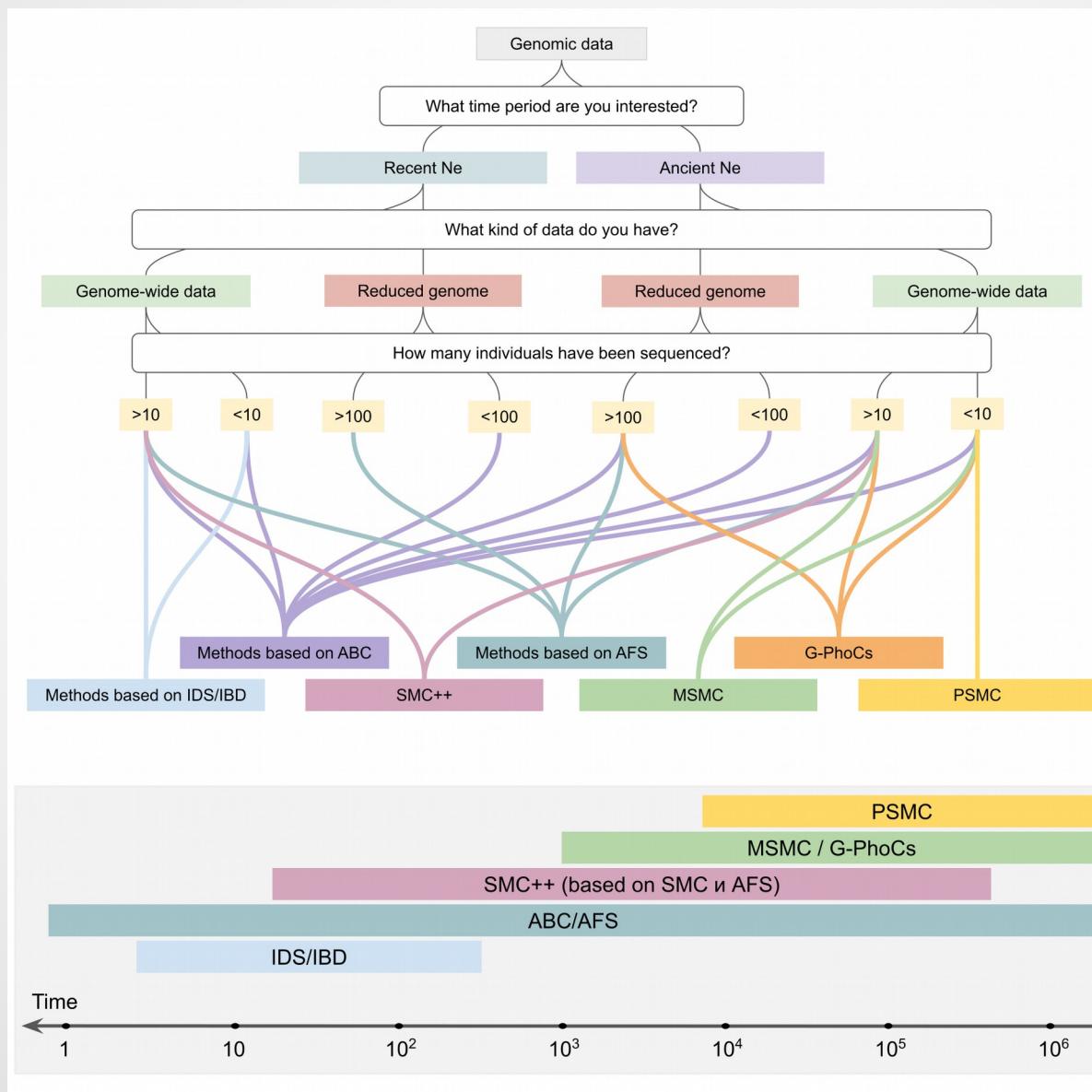
# Demography

## Demographic history

- reconstruction of effective population size changes in the past
- some methods allow to do it even from a single genome



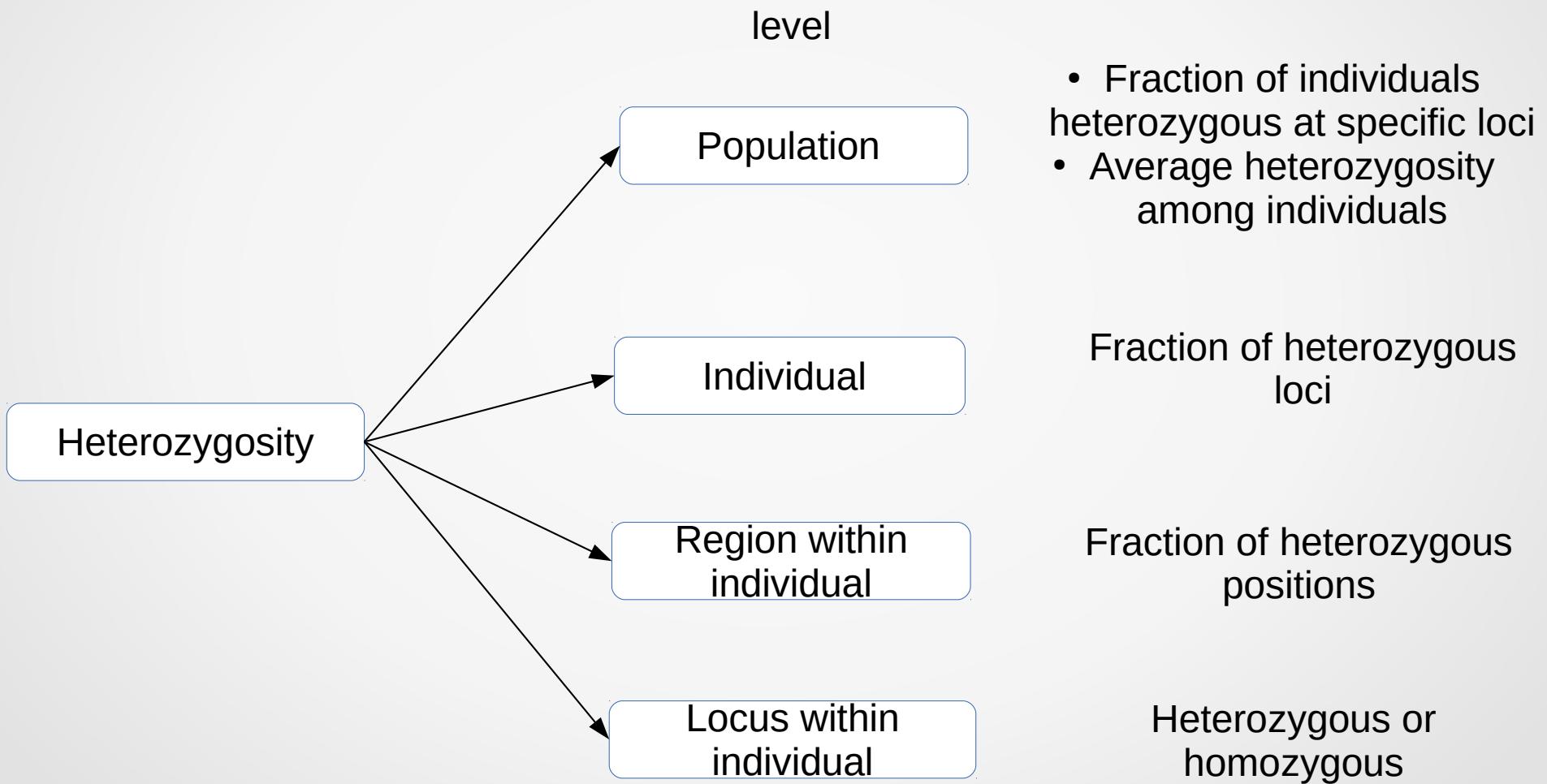
# Methods for demography reconstruction



## Indicators used to assess genetic diversity

- Wright's  $F$ -statistics
- haplotype diversity ( $H$ )
- nucleotide diversity ( $\pi$ )
- heterozygosity
- etc

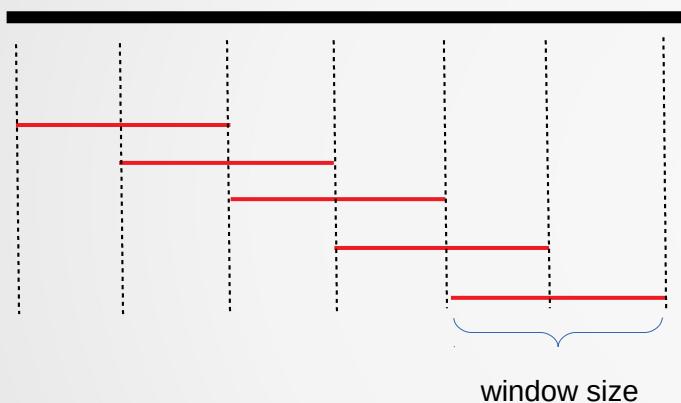
# Heterozygosity



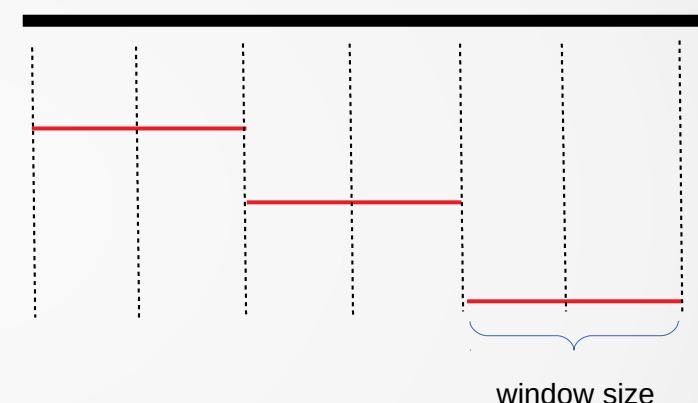
# Sliding windows approach

**Way to switch from single value of average heterozygosity to distribution**

overlapping windows



non-overlapping windows

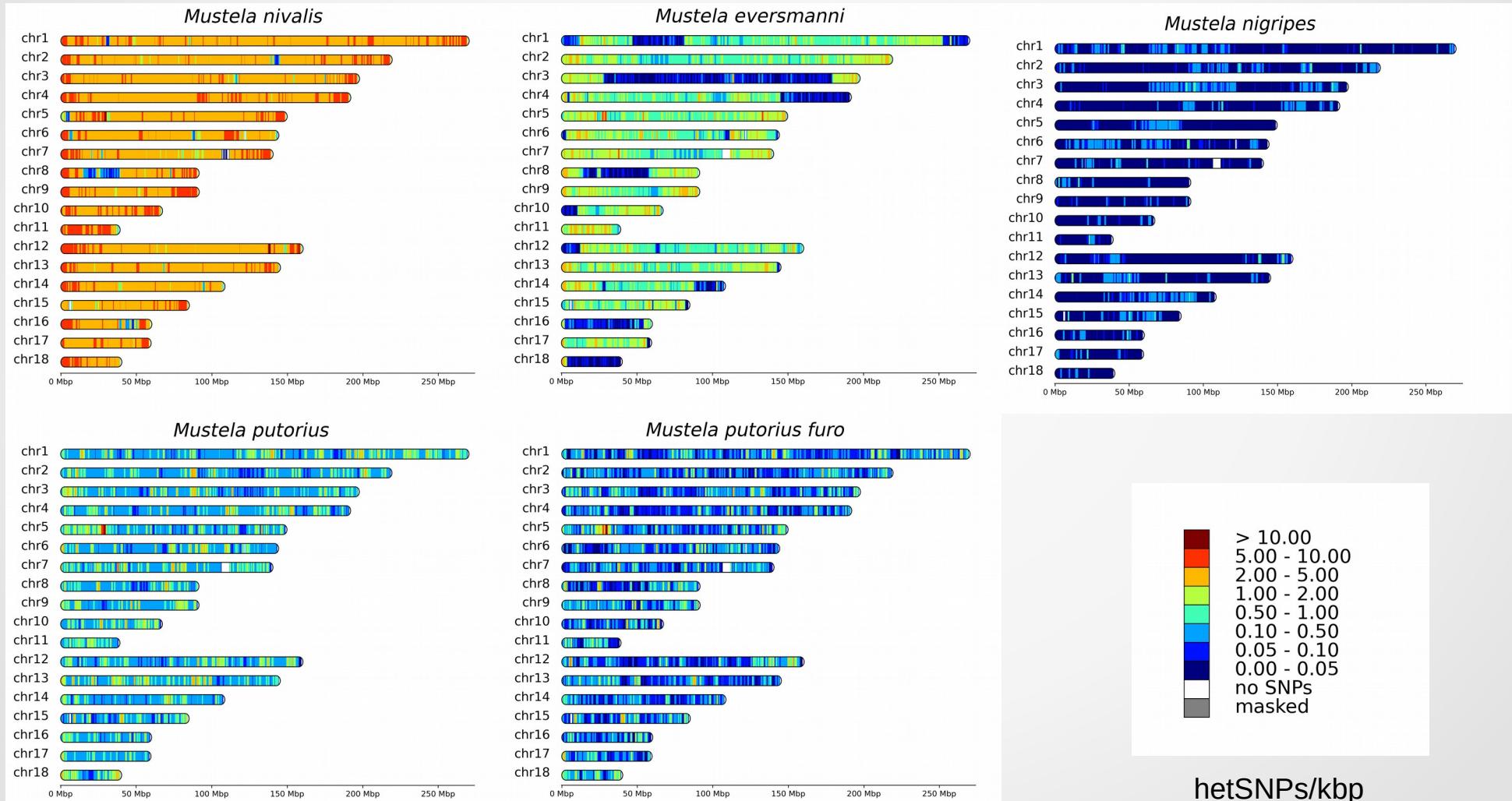


Cut genome into windows and count heterozygous genetic variants in each window.

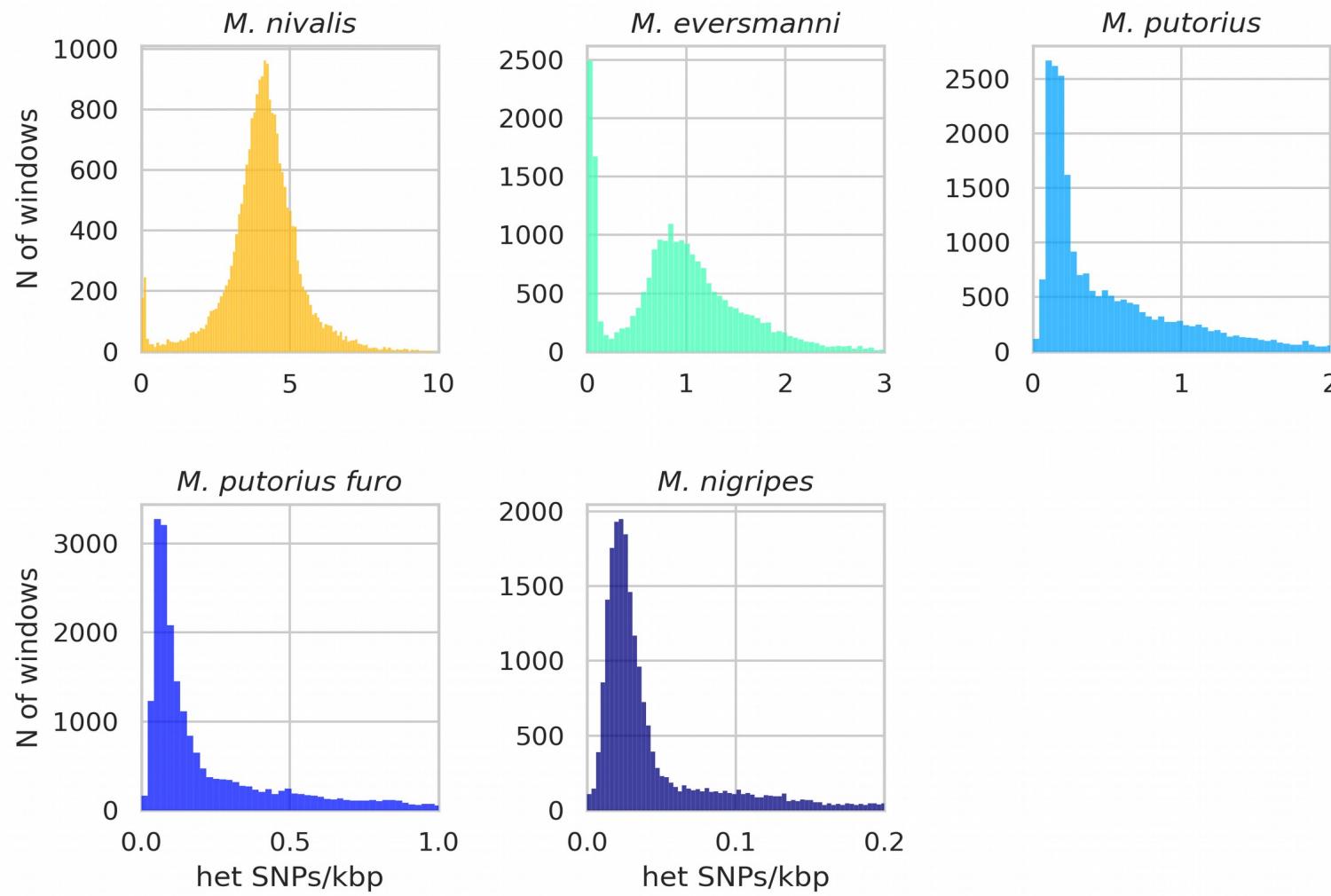
Then deal with distribution of counts.

Usually window size is between 50 - 1000 kbp.

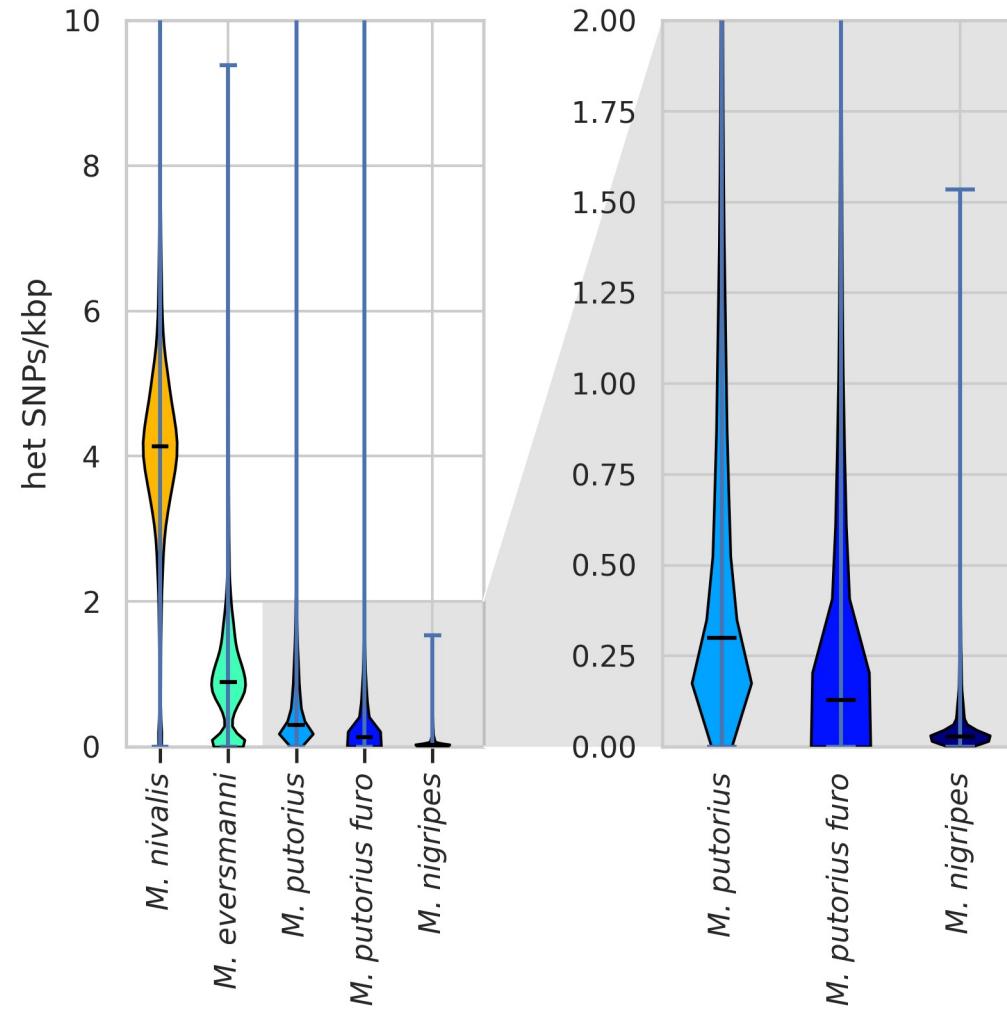
# Visualization of heterozygosity distribution: heatmap



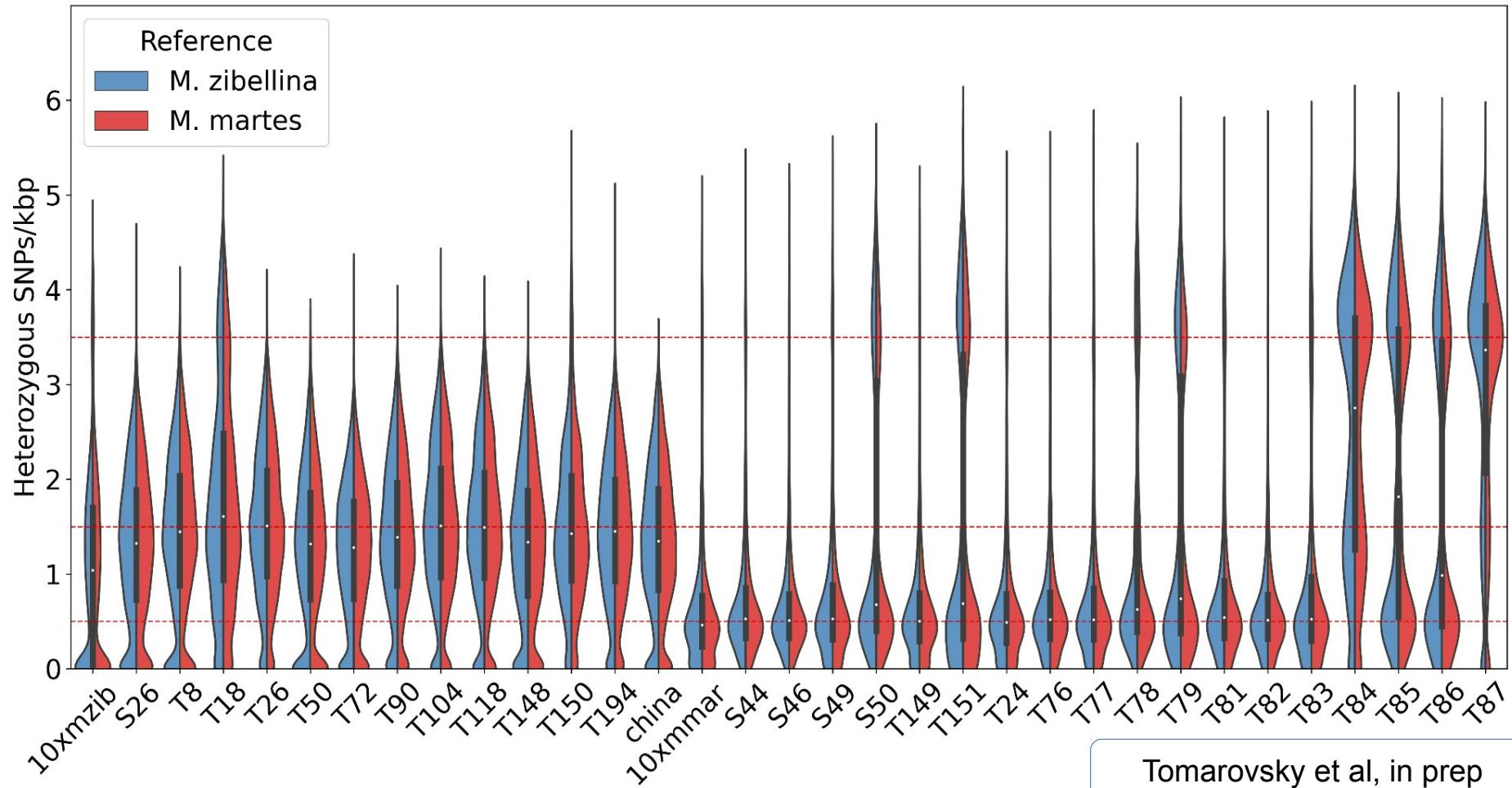
# Visualization of heterozygosity distribution: histogram



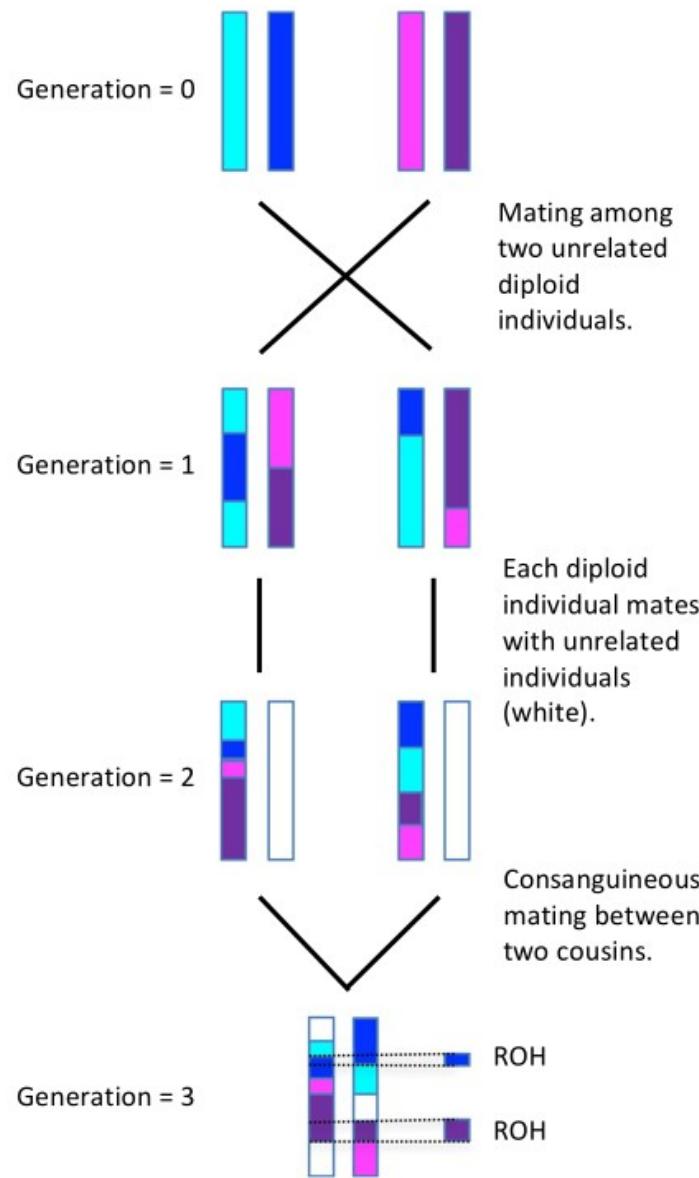
# Visualization of heterozygosity distribution: violin plot



# Visualization of heterozygosity distribution: violin plot



# Runs of homozygosity

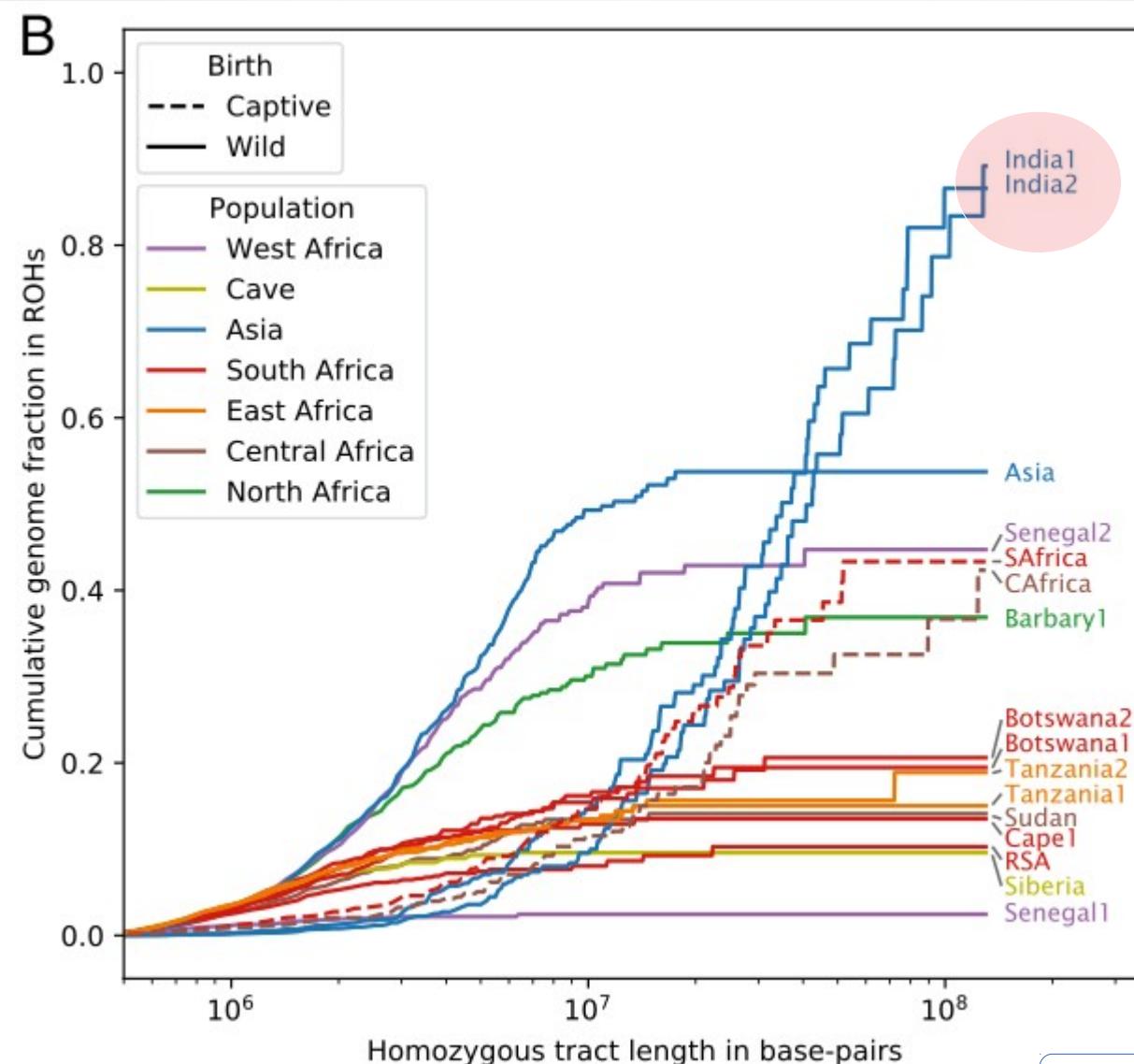


**RoH = run of homozygosity**

- long homozygous region in the genome
- idealization due to sequencing errors and repetitive sequences

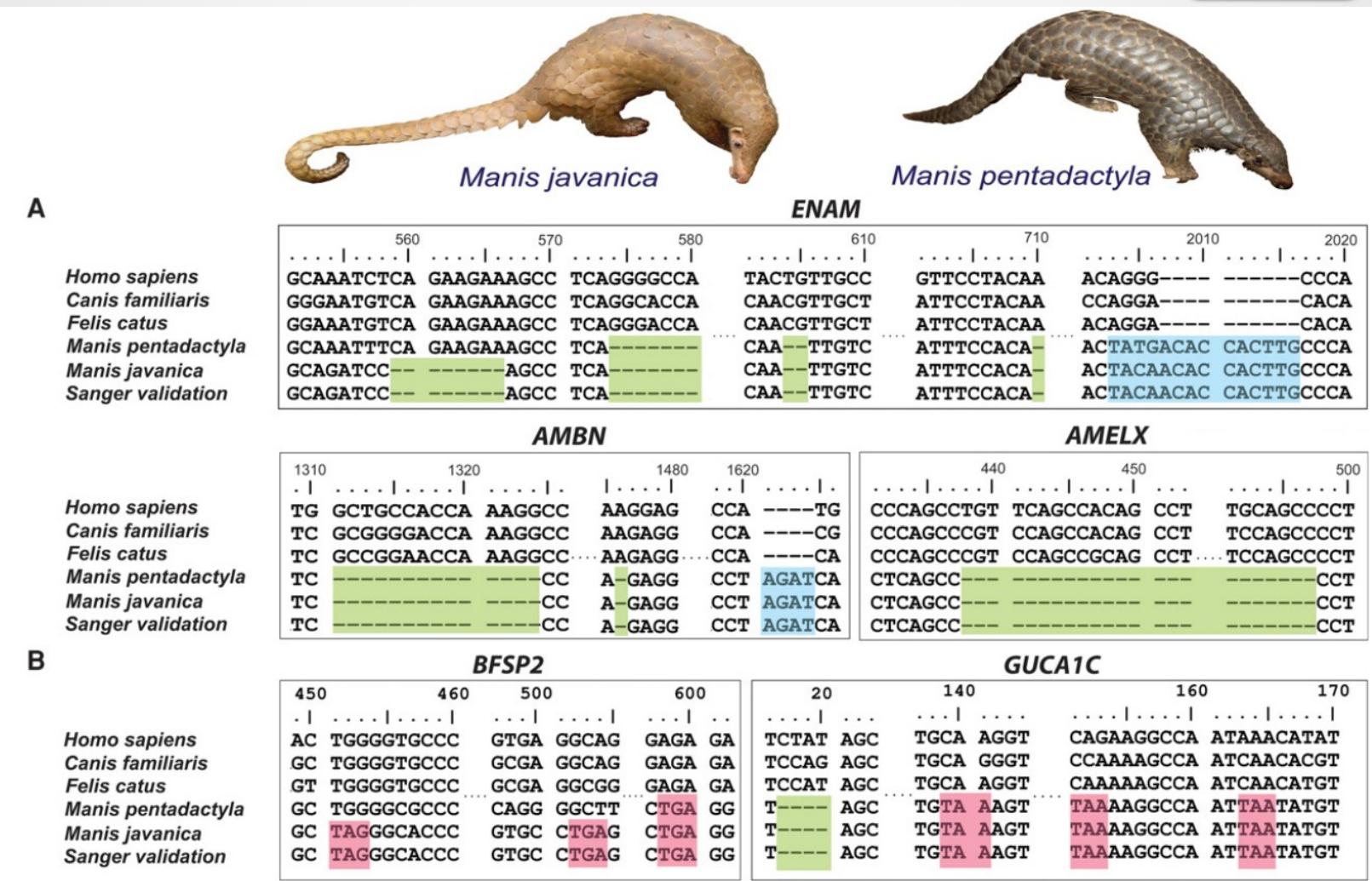
Brunische-Olsen, 2018

# ROHs in lions



# Pseudogenized genes in pangolin species

Vision related genes



Choo et al, 2016

Teeth-development genes

# Indels in interferon ε in pangolins

>40 species without insertion

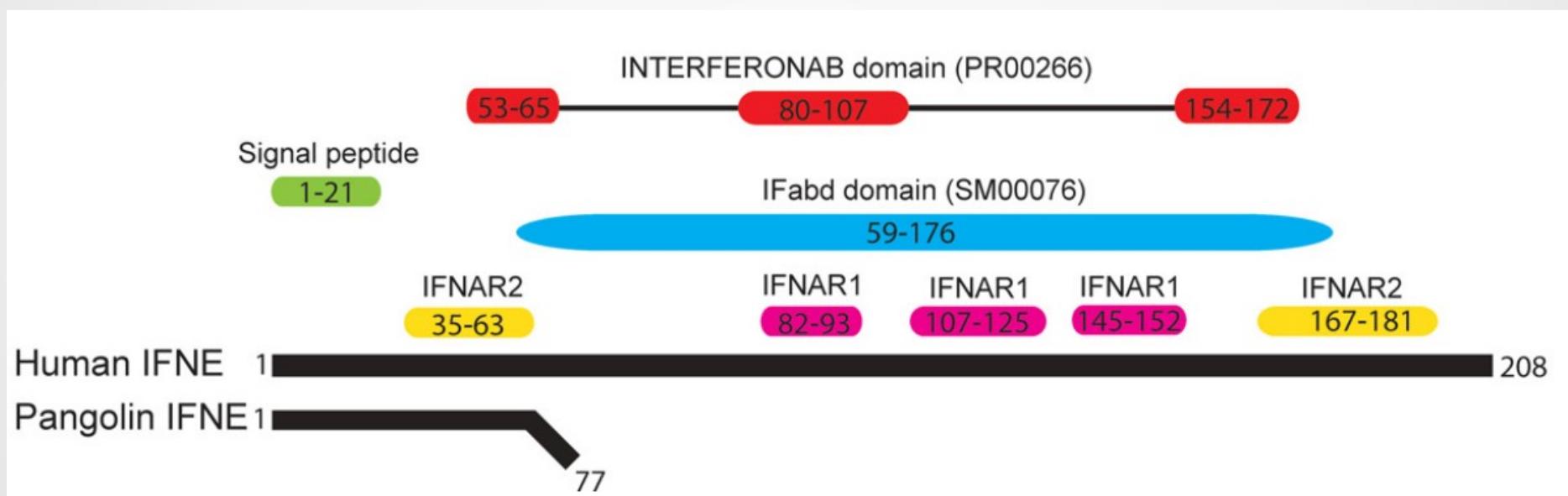
A

**IFNE**

*Homo sapiens*  
*Pan paniscus*  
*Pan troglodytes*  
*Gorilla gorilla*  
*Pongo abelii*  
*Nomascus leucogenys*  
*Rhinopithecus roxellana*  
*Macaca fascicularis*  
*Colobus angolensis palliatus*  
*Macaca mulatta*  
*Mandrillus leucophaeus*  
*Cercocebus atys*  
*Macaca nemestrina*  
*Papio anubis*  
*Chlorocebus sabaeus*  
*Aotus nancymaae*  
*Callicebus jacchus*  
*Galeopterus variegatus*  
*Pteropus vampyrus*

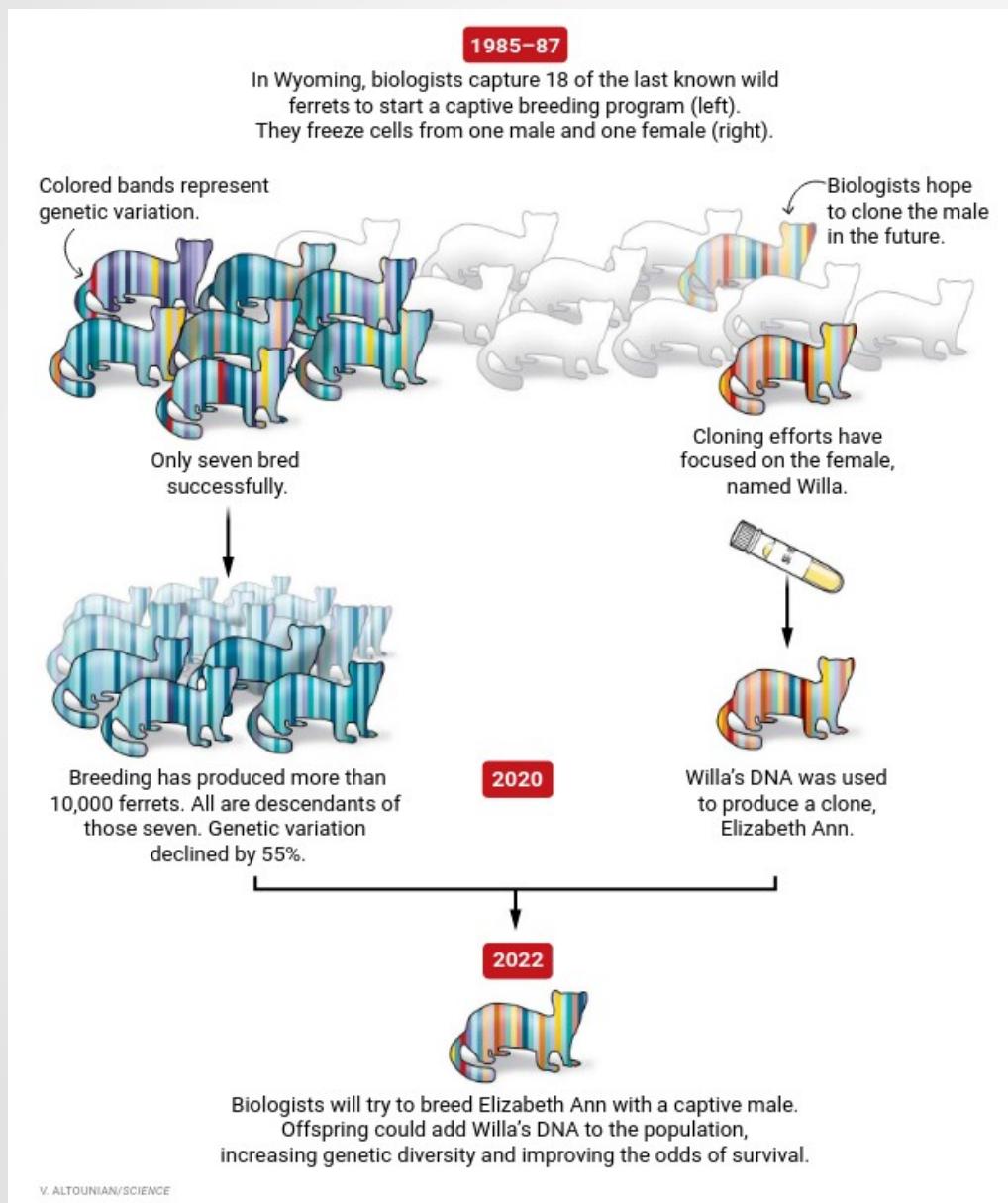
	Nucleotide sequences						Protein sequences	
	200	220	240	265	540	545	65	75
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAGCC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCT	TCAGCAGTAC	TCT	AA ATCAGCC		FLLPQKS-	-LSLQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAGCC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTTTGAGTCC	TCAGCAGTAC	TCT	AA ATCAACC		FLLPQKS-	-LSPQQY
TCAGAAG-	-	T CTGTGAGTTC	TCAACAGCAC	TCT	AA ATCAACT		FMLPQKS-	-VSSQQH
TCAGAAG-	-	T CTATGAGTTC	TCAACAGCAC	TCT	AA ATCAACC		FMLPQKS-	-MSSQQH
CCAGAAG-	-	T CTGCGAATCC	TCAGCAATAAC	TTT	AA ATCAACA		FLLPQKS-	-ANPQQY
CCAGAAG-	-	T CTCTGAATCC	TCACCTGTAC	TCT	AA ATCAACC		FLLPQKS-	-LNPHLY
<hr/>								
<i>Nannospalax galili</i>	TCAGAAG-	-	T CTGTGGCTCC	TTACCAAGTAT	TAT	AA ATCAACA	FLLPQKS-	-VAPYQQ
<i>Chrysochloris asiatica</i>	CCAGAAG-	-	T CTATGAATCC	TCACCAAGTAC	TCT	AA ATTATAAC	FMLPQKS-	-MNPHQY
<i>Erinaceus europaeus</i>	CCTGAAG-	-	T CTCCAAGTCC	TCACTGGTAT	TCT	AA ATCAACC	FLLPLKS-	-PSPHWY
<i>Ictidomys tridecemlineatus</i>	TCAGAAT-	-	T CTGTGAATTTC	TCACCAAGTAC	TCT	AA ATCAACC	FRLPQNS-	-VNFHQY
<i>Sorex araneus</i>	CCTGAGG-	-	T CTATGAATCC	TGACCAAAAC	TCT	AA ATTCAACC	FLLPLRS-	-MNPDQN
<i>Dipodomys ordii</i>	TTGGAAA-	-	T CCATGAGTCC	TCACCAAGTAC	TAT	AA GTTAACC	FLFPWKS-	-MSPHQY
<i>Ochotona princeps</i>	CCAGATG-	-	T CTGTGAATCT	TTATCAGTAC	TCT	AA ATCAACC	FGLPQMS-	-VNLYQY
<i>Echinops telfairi</i>	TCAGAAG-	-	G CCATAAAACCC	TCCTCGATGC	TCT	AA ATTAATC	FLLPQKA-	-INPPRC
<i>Elephantulus edwardii</i>	CCACAAG-	-	T CCATGAATTC	TCACCAAGGAC	TCT	AA GTCATTC	FALPHKS-	-MNSHQD
<i>Cricetulus griseus</i>	TTGGCAG-	-	T CTGTGAGTCG	TCACCAAGTAC	TCT	AA ATCAACC	FQLPWQS-	-VSRHQY
<i>Mesocricetus auratus</i>	TCTGCAG-	-	T CTGTGAGTCG	TCACCAAGTAC	TCT	AA ATCAACC	FLLPLQS-	-VSRHQY
<i>Mus musculus</i>	TCAGCAG-	-	C CTGTGAGTCC	TCACCAAGTAC	TGT	AA ATCCACC	FLLPQQP-	-VSPHQY
<i>Manis pentadactyla</i>	CCACCAATA	CCAGGATCCT	CCTGTATCTT	TCCTGAATCC	TCACTAGTAC	-CT A -----C	FLLRHQYQDP	PVSFLNPHY
<i>Manis javanica</i>	CCACCAATA	CCAGGATCCT	CCTGTATCTT	TCCTGAATCC	TCACTAGTAC	-CT A -----C	FLLPHQYQDP	PVSFLNPHY
Sanger validation	CCACCAATA	CCAGGATCCT	CCTGTATCTT	TCCTGAATCC	TCACTAGTAC	-CT A -----C	FLLPHQYQDP	PVSFLNPHY

# Shorten interferon ε protein in pangolins



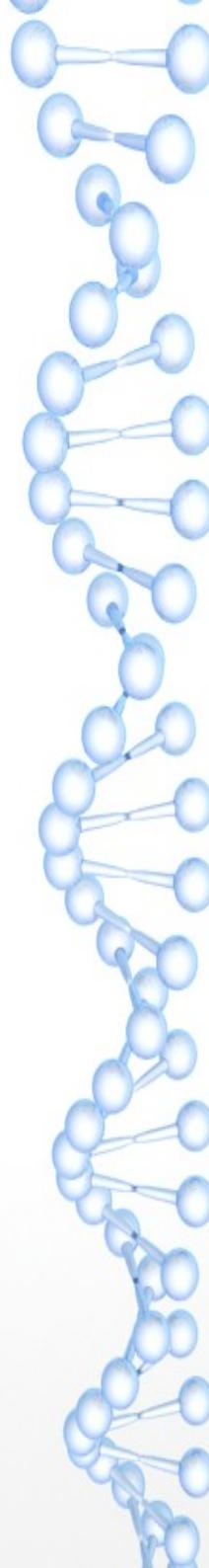
Choo et al, 2016

# Cloning from biobanks as a way to restore diversity



Elizabeth Ann,  
the cloned black-footed ferret

Rachel Fritts, 2018



**End of module IV**