Experimental Design

Joana Meier

Tree of Life Programme, Wellcome Sanger Institute

THE ROYAL SOCIETY

The Branco Weiss Fellowship Society in Science

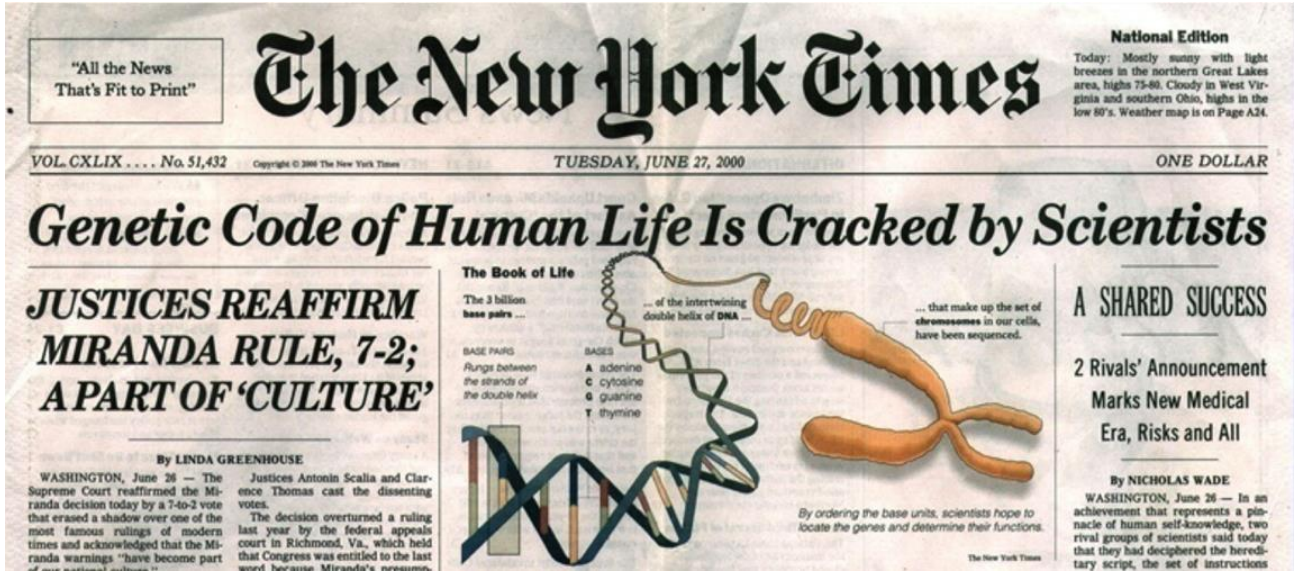
wellcome

UNIVERSITY OF CAMBRIDGE

Outline

- Sequencing technologies
- Sampling and sequencing strategy
- Case studies
- Ensuring you trust your data
- Breakout groups to discuss your experimental design
- Reference genomes

Human Genome Project



"All the News That's Fit to Print"

The New York Times

National Edition

Today: Mostly sunny with light breezes in the northern Great Lakes area, highs 75-80. Cloudy in West Virginia and southern Ohio, highs in the low 80's. Weather map is on Page A24.

VOL. CXLIX . . . . No. 51,432 TUESDAY, JUNE 27, 2000 ONE DOLLAR

Genetic Code of Human Life Is Cracked by Scientists

JUSTICES REAFFIRM MIRANDA RULE, 7-2; A PART OF 'CULTURE'

By LINDA GREENHOUSE

WASHINGTON, June 26 — The Supreme Court reaffirmed the Miranda decision today by a 7-to-2 vote that erased a shadow over one of the most famous rulings of modern times and acknowledged that the Miranda warnings "have become part

Justices Antonin Scalia and Clarence Thomas cast the dissenting votes.

The decision overturned a ruling last year by the federal appeals court in Richmond, Va., which held that Congress was entitled to the last

The Book of Life

The 3 billion base pairs ...

... of the intertwining double helix of DNA ...

... that make up the set of chromosomes in our cells, have been sequenced.

BASE PAIRS
Rungs between the strands of the double helix

BASES
A adenine
C cytosine
G guanine
T thymine

By ordering the base units, scientists hope to locate the genes and determine their functions.

The New York Times

A SHARED SUCCESS

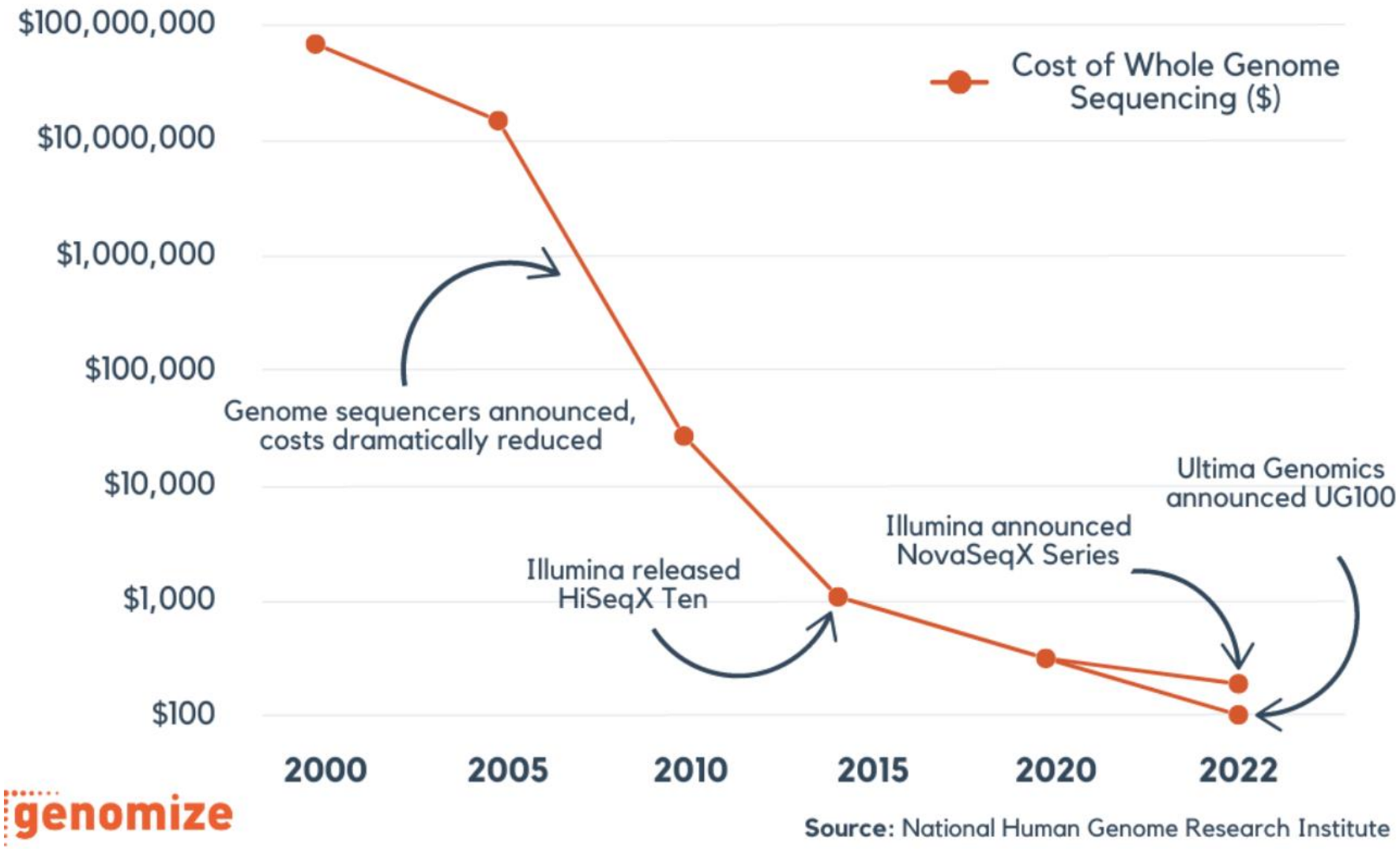
2 Rivals' Announcement Marks New Medical Era, Risks and All

By NICHOLAS WADE

WASHINGTON, June 26 — In an achievement that represents a pinnacle of human self-knowledge, two rival groups of scientists said today that they had deciphered the hereditary script, the set of instructions

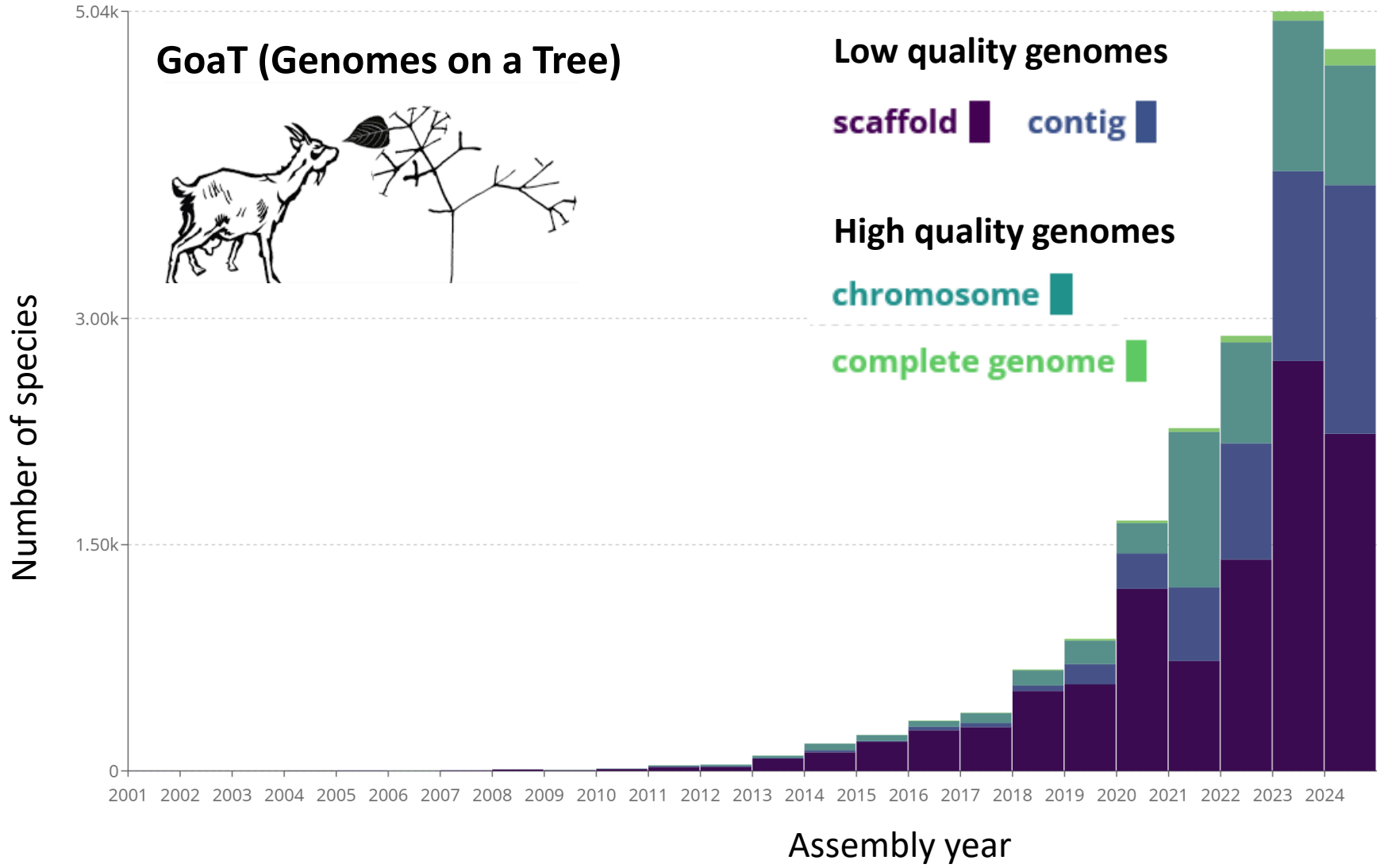
- Human genome project – started in 1990, completed in 2003
- Sequenced across ~20 institutions worldwide
- Cost an approximate $5 billion US dollars

Sequencing costs are decreasing rapidly



Since Oct 2023 PacBio Revio (66 Gbp per lane in 15 kb reads)

More and more species have reference genomes



Two main types of high-throughput sequencing

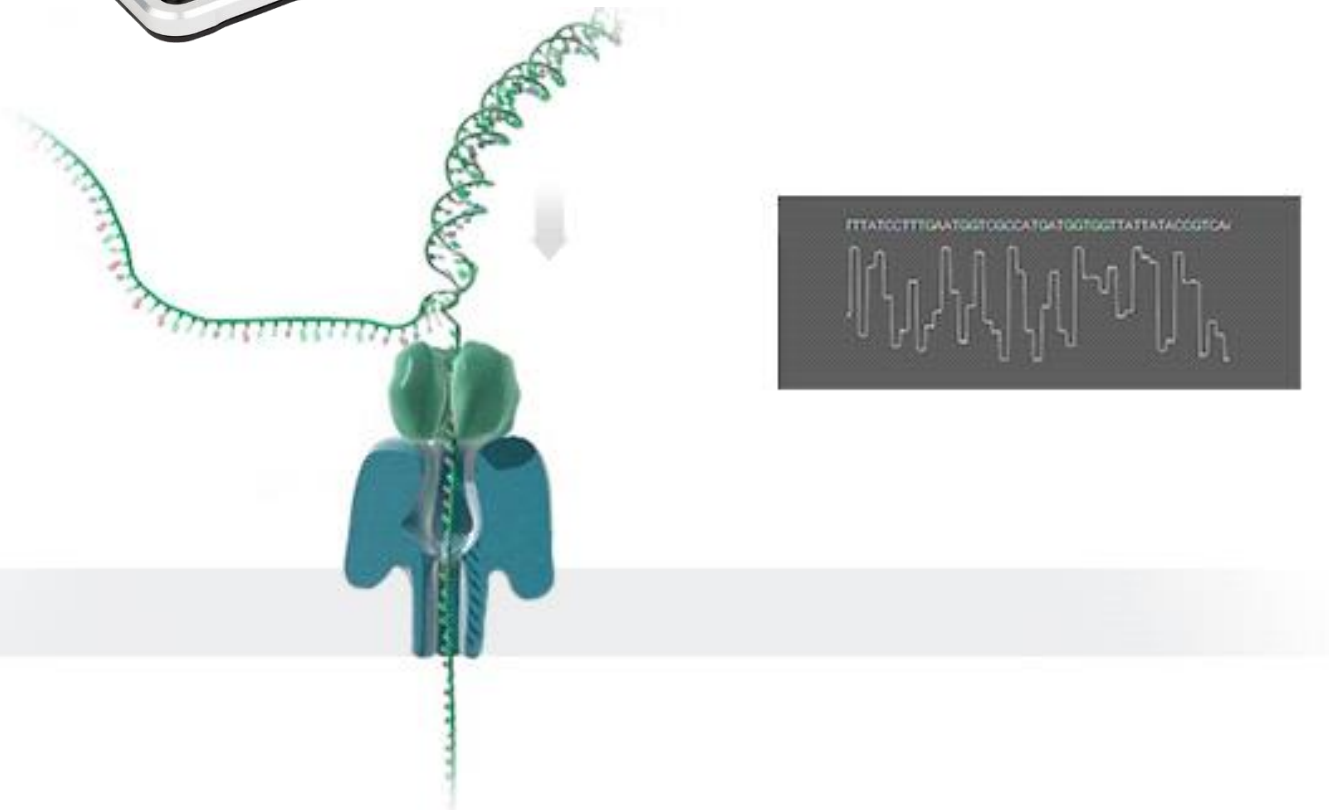
- **Short-read sequencing**
  - Reads are typically 150-300 bp long
  - Cheaper than long-read sequencing
  - E.g. Illumina, Ultima Genomics

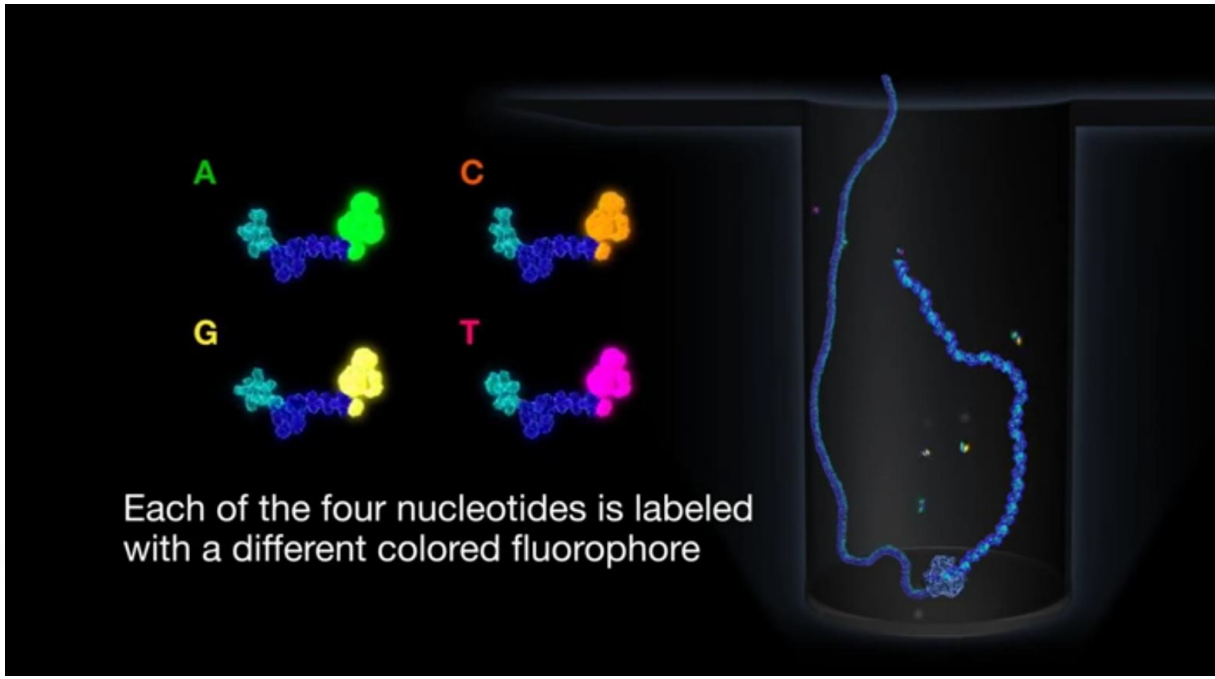
- **Long-read sequencing**
  - Reads are typically >10 kb long (PacBio: 10-20 kbp, ONT/Nanopore: >100 kbp)
  - More expensive than short-read technologies
  - Required for making a reference genome
  - E.g. PacBio or ONT/Nanopore

Long read sequencing technologies

ONT / Nanopore
(>100 kb reads)

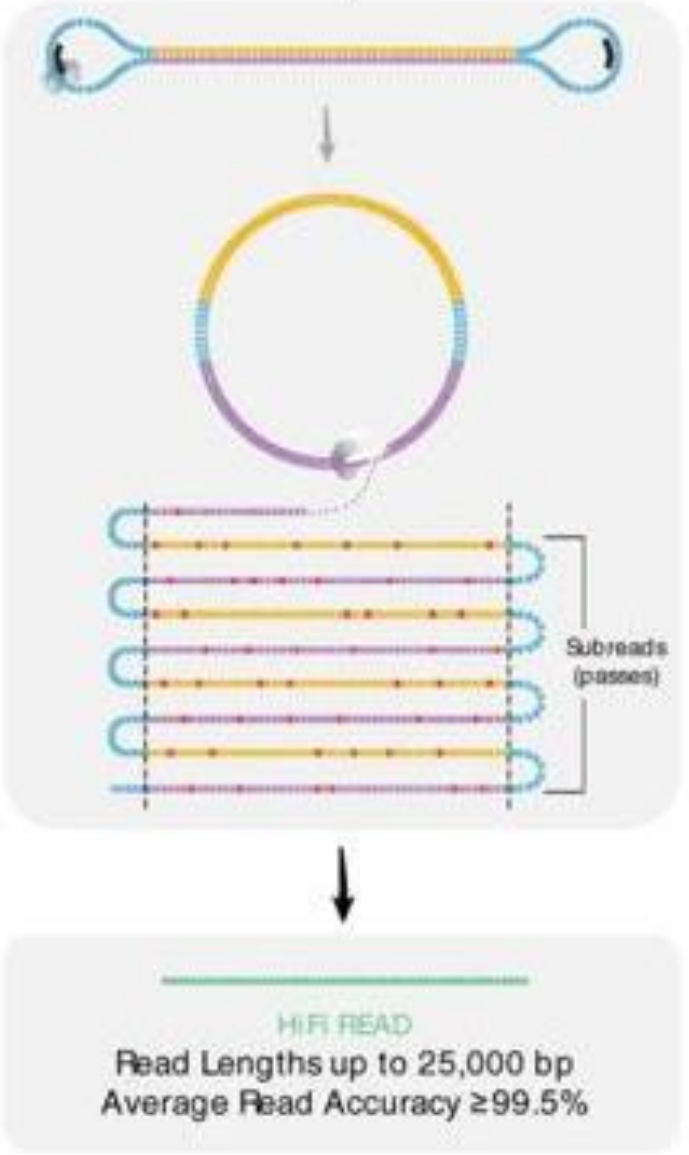
PacBio
~10-20 kb reads





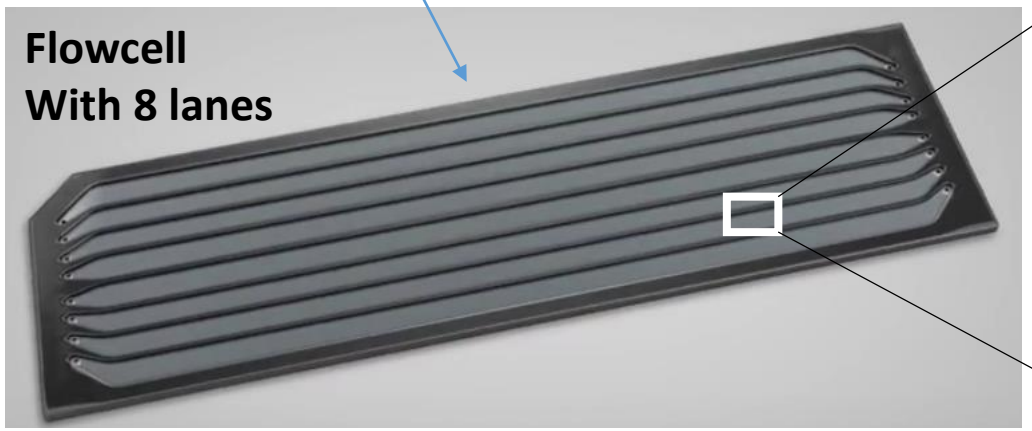
PacBio HiFi reads (99.95% accurary)

Each DNA-fragment is sequenced many times to get a high-quality consensus (=summary) read

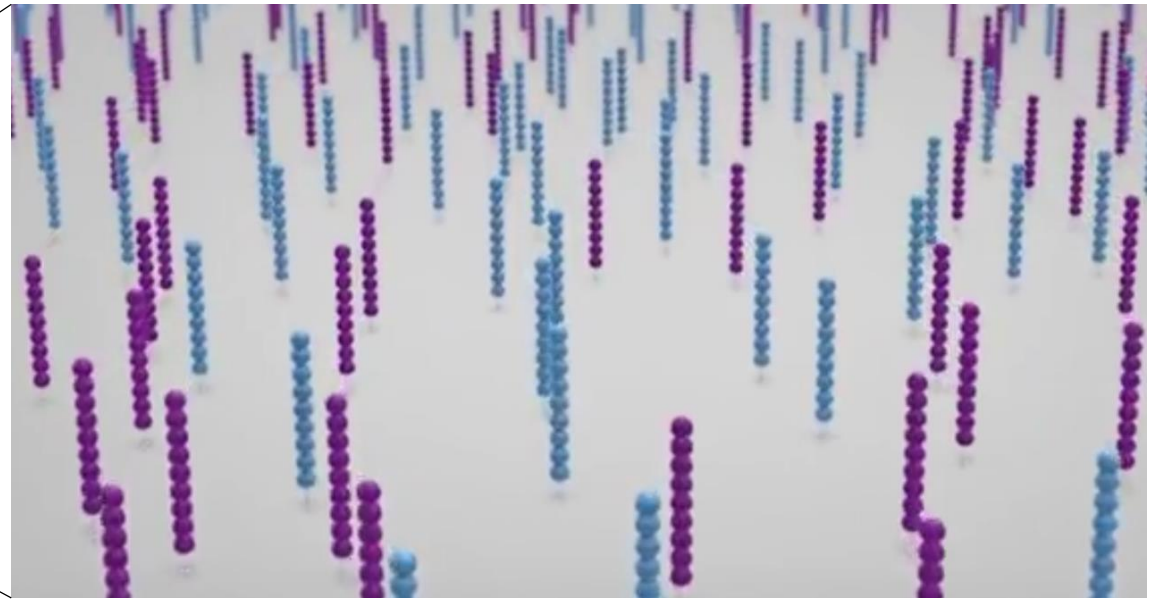


Illumina flowcell

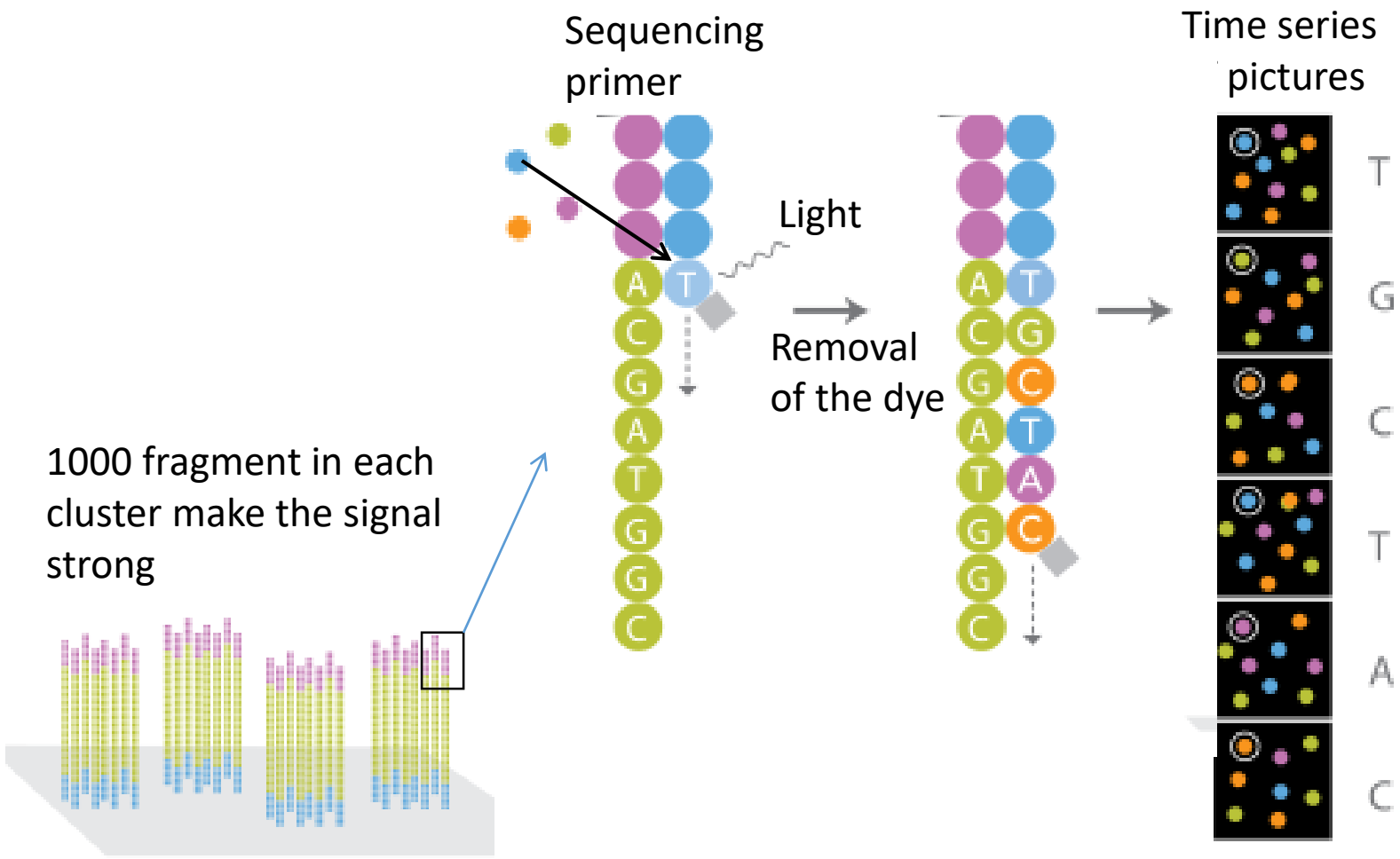
DNA fragments with Illumina adapters

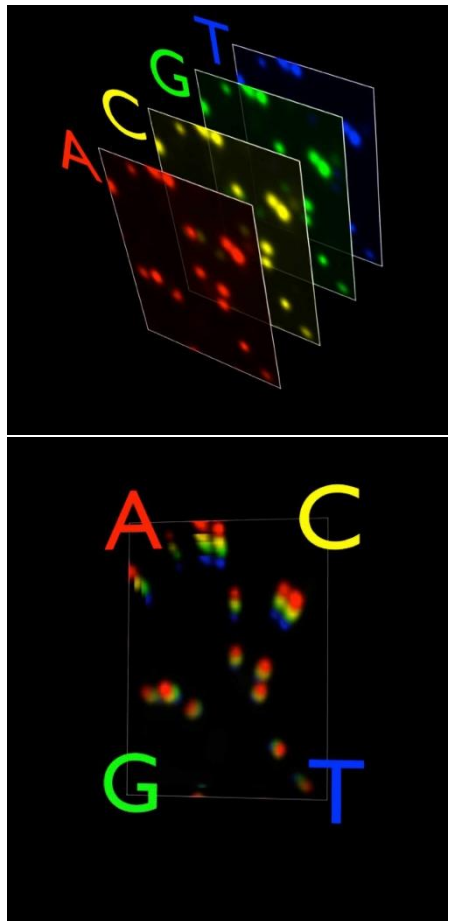


Each lane contains a dense lawn of Illumina primers



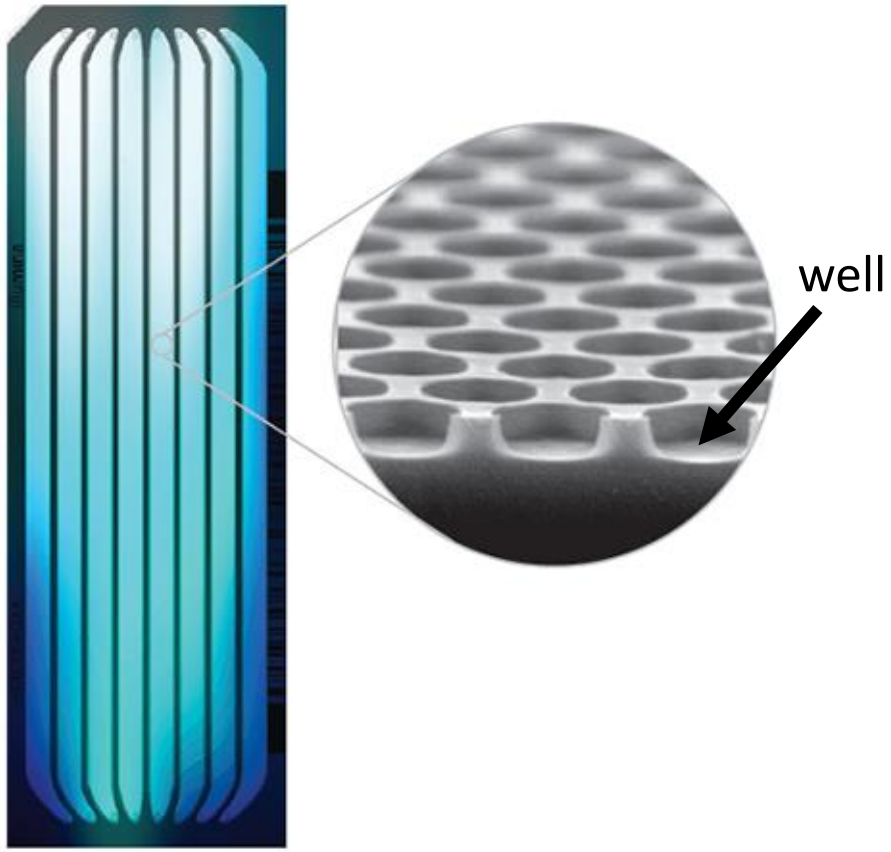
Short-read sequencing with Illumina

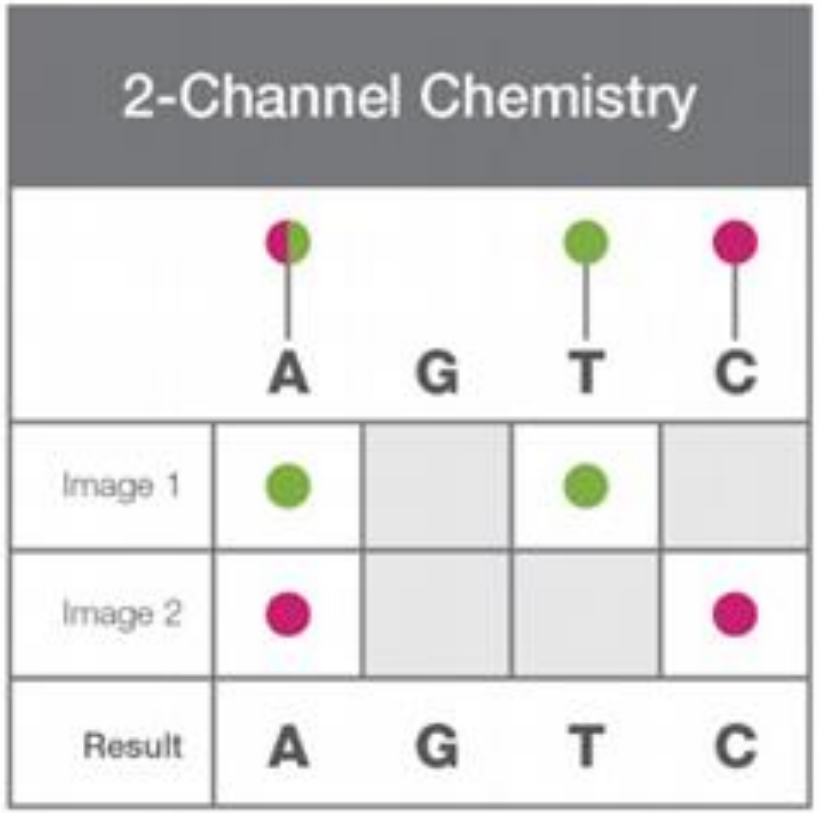




Newer Illumina machines use wells and only 2 colours

(e.g. Novaseq, Nextseq, MiniSeq. This makes it faster and cheaper)

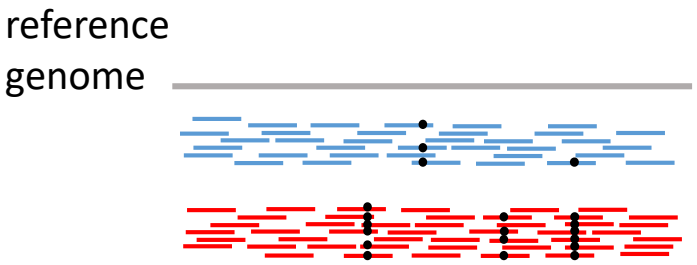




Sequencing approaches for speciation genomics

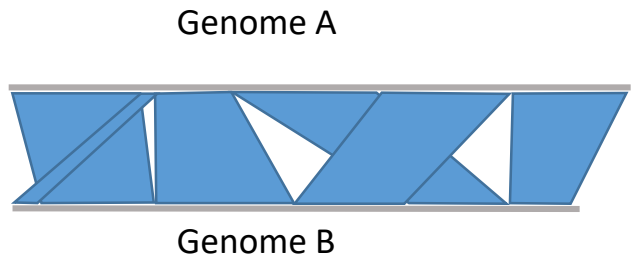
Whole-genome requencing (short-read data)

- Requires a reference genome
- individuals need to be from the same or closely related species
- Complete genome sequenced



Genome assembly comparisons (long-read data)

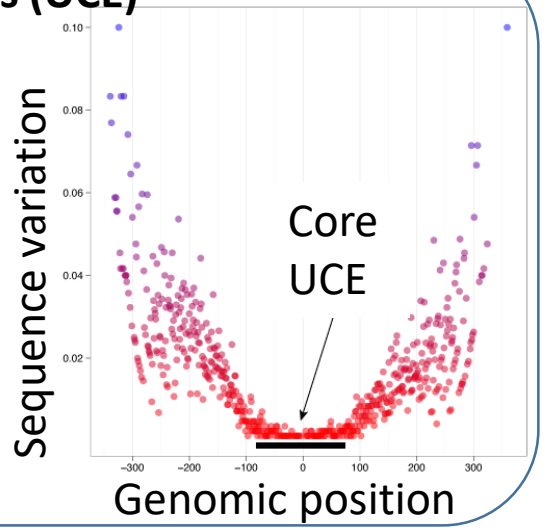
- Comparative genomics – studying structural variation between species, can be distantly related
- Gene expansions, transposable elements etc
- Phylogenomics across deeply divergent species
- Pangenomics – multi-genome assemblies to study within-species variation in structural variants



Reduced-representation techniques
(only parts of the genome sequenced)

Ultra-conserved elements (UCE)

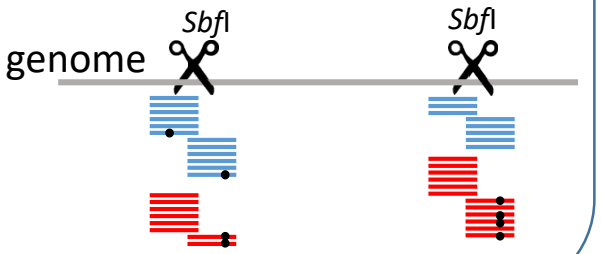
- Sequence capture with baits based on genomic regions that are conserved across many species
- Works with highly divergent species



Restriction Associated Sequencing (RAD)

(similar methods: GBS, ddRAD)

- does not require primers/baits or reference genome
- individuals need to be from the same or closely related species
- Information from thousands of loci distributed across the genome

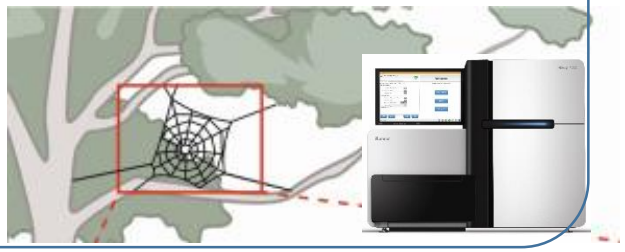


Targeted or amplicon sequencing, e.g. barcoding

- Sequencing one or few genes
- requires primers
- e.g. CO1 (mitochondrial barcoding region), advantage: large database (BOLD) available to compare to for species identification

Environmental DNA (eDNA)

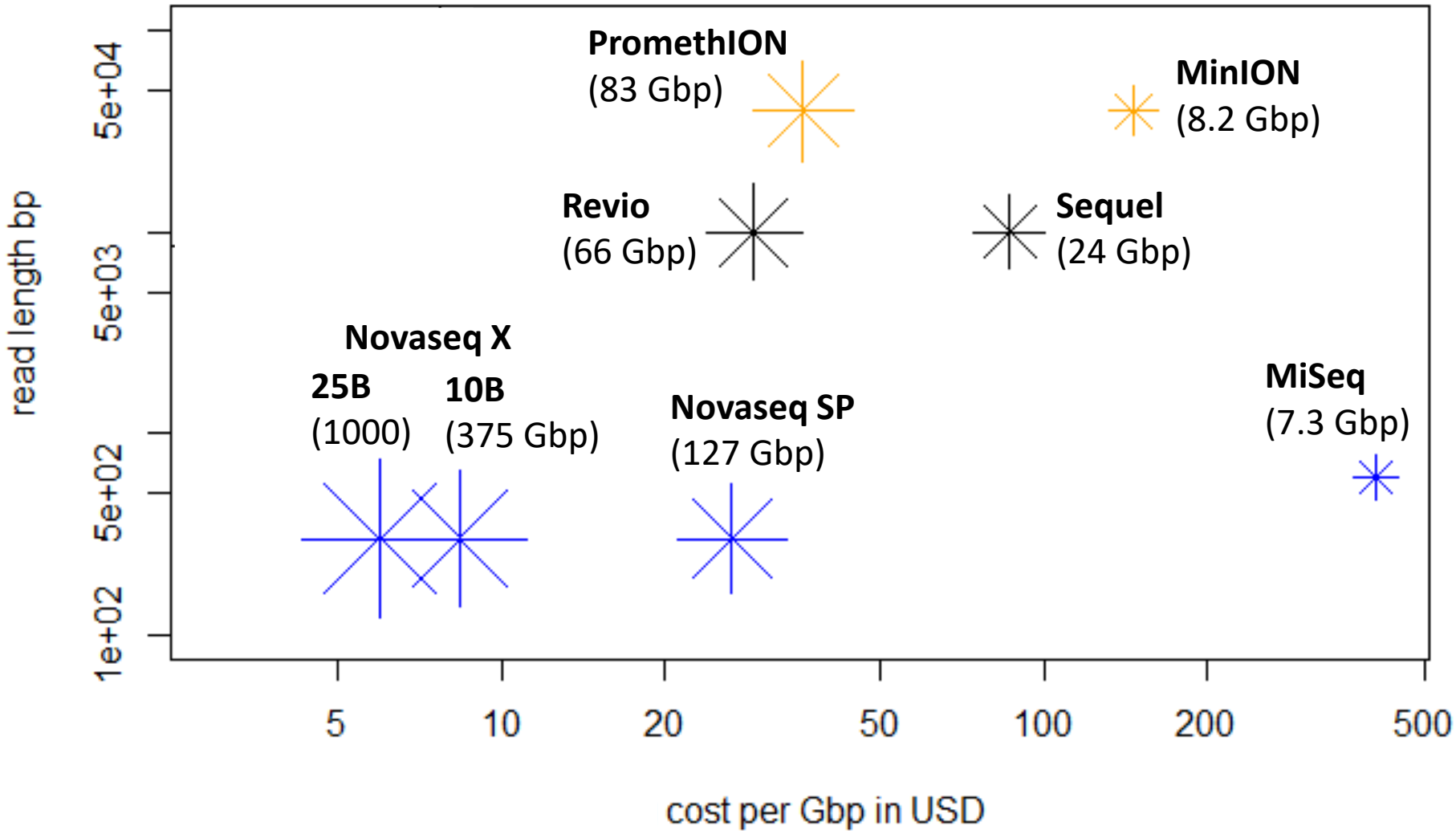
- Mostly CO1 sequencing from soil, water, air (spider webs)
- Studying local species richness



Read length versus per Gbp sequencing costs for different sequencing machines (note the axes are in logarithmic scale)

ONT
PacBio
Illumina

Star size shows the total throughput per lane, also given in parentheses ()



Sequencing strategy

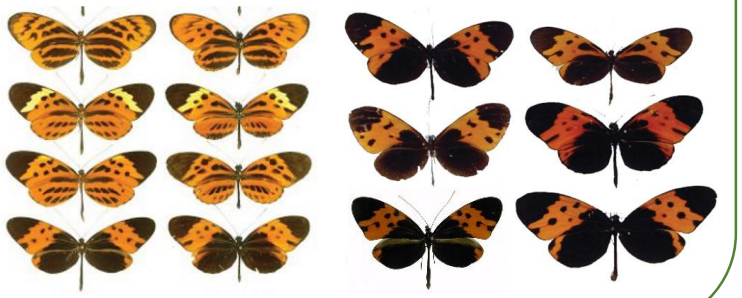
Total sequencing data gets divided by:

- Number of sites to sequence
  - Depends on genome size and sequencing stragegy, e.g. RAD versus whole-genome
- Number of specimens to sequence
  - and how many specimens per population, how many populations per species
- Sequencing depth (e.g. sequencing at 10x depth of coverage)

- Example: 1 NovaseqX 10B lane
  ~2.5 billion paired-end reads of 150 bp each -> 375 Gbp data
  - 100 whole-genomes of a species with 0.375 Gbp genome size at 10x coverage
  - 19 whole-genomes of a species with 1 Gbp genome size at 20x coverage
  - 375 individuals sequenced with a RAD sequencing approach resulting in 50 Mbp at a sequencing depth of 20x

Different questions require different experimental design

Resolving the taxonomy
- Placing potentially new species
- Species delineation

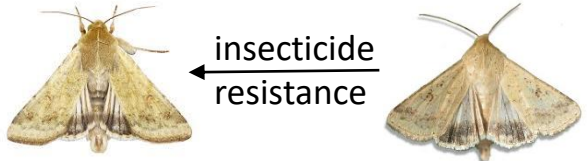


Adaptation to climate change
- Identifying genomic regions involved in adaptation

Pfenninger et al. 2021

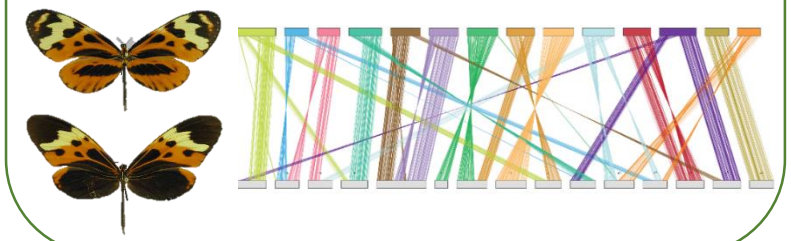
Are the species declining?
- Detecting past inbreeding
- Assessing genetic diversity

Detecting introgression
- Are populations/species hybridising now or in the past?
- Finding regions of adaptive introgression

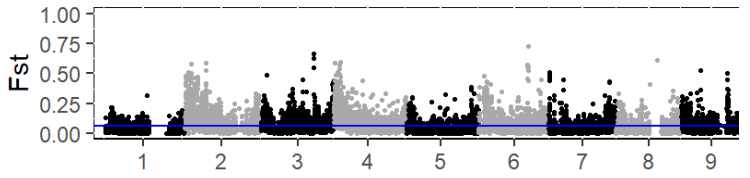


Helicoverpa zea *H. armigera*

Studying genome evolution
- Gene expansions, e.g. olfactory
- Chromosomal rearrangements
- Genome size evolution (TEs, etc)



What genes are under divergent selection
- Short-term (recent sweep)
- Long-term (fixed substitutions)



Sampling considerations

- Knowing the natural history and morphologically inferred taxonomy if your helps massively
- If you want to test for introgression, you need outgroups
- Aim to get even sampling of populations/species to compare
- Sampling sites matter:
  e.g. if you want to test for ongoing gene flow, you need specimens from sympatry or parapatry
- High number of specimens needed for:
  - Inferring recent demographic changes
  - Studying rare alleles
  - GWAS (inferring the genetic basis of traits, particularly if polygenic)

Case studies to discuss

Three examples with very different optimal sampling designs

- **Identifying barriers to gene flow and the genetic basis of relevant traits in a hybrid zone**

H. erato and *H. melpomene* in Ecuador

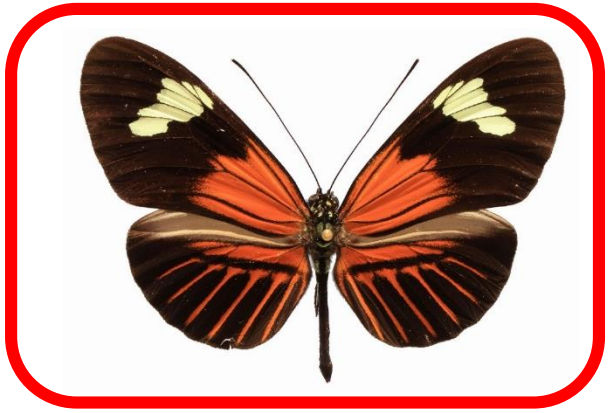
Heliconius erato

Highland

Lowland

Heliconius melpomene

Highland



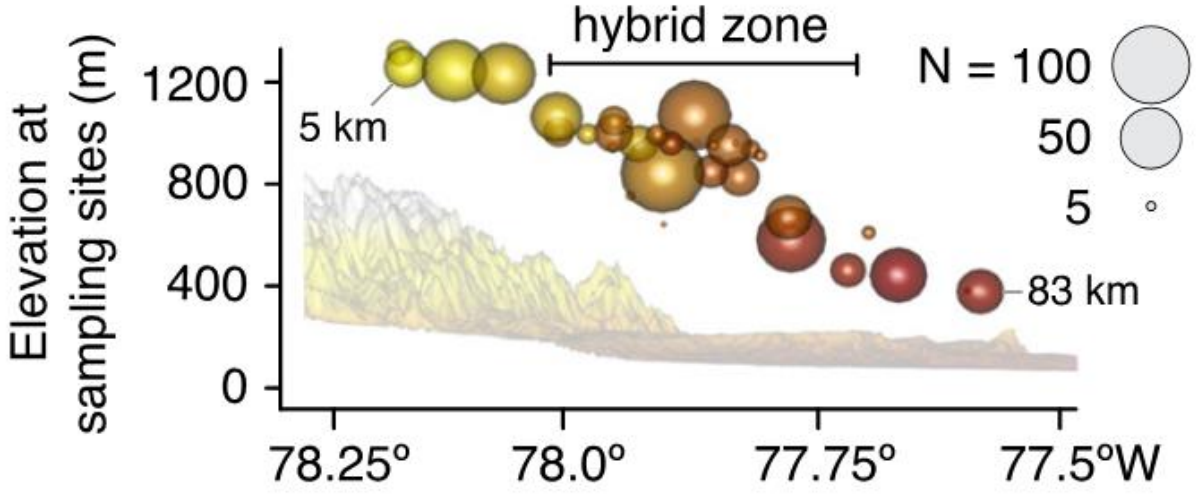
Lowland

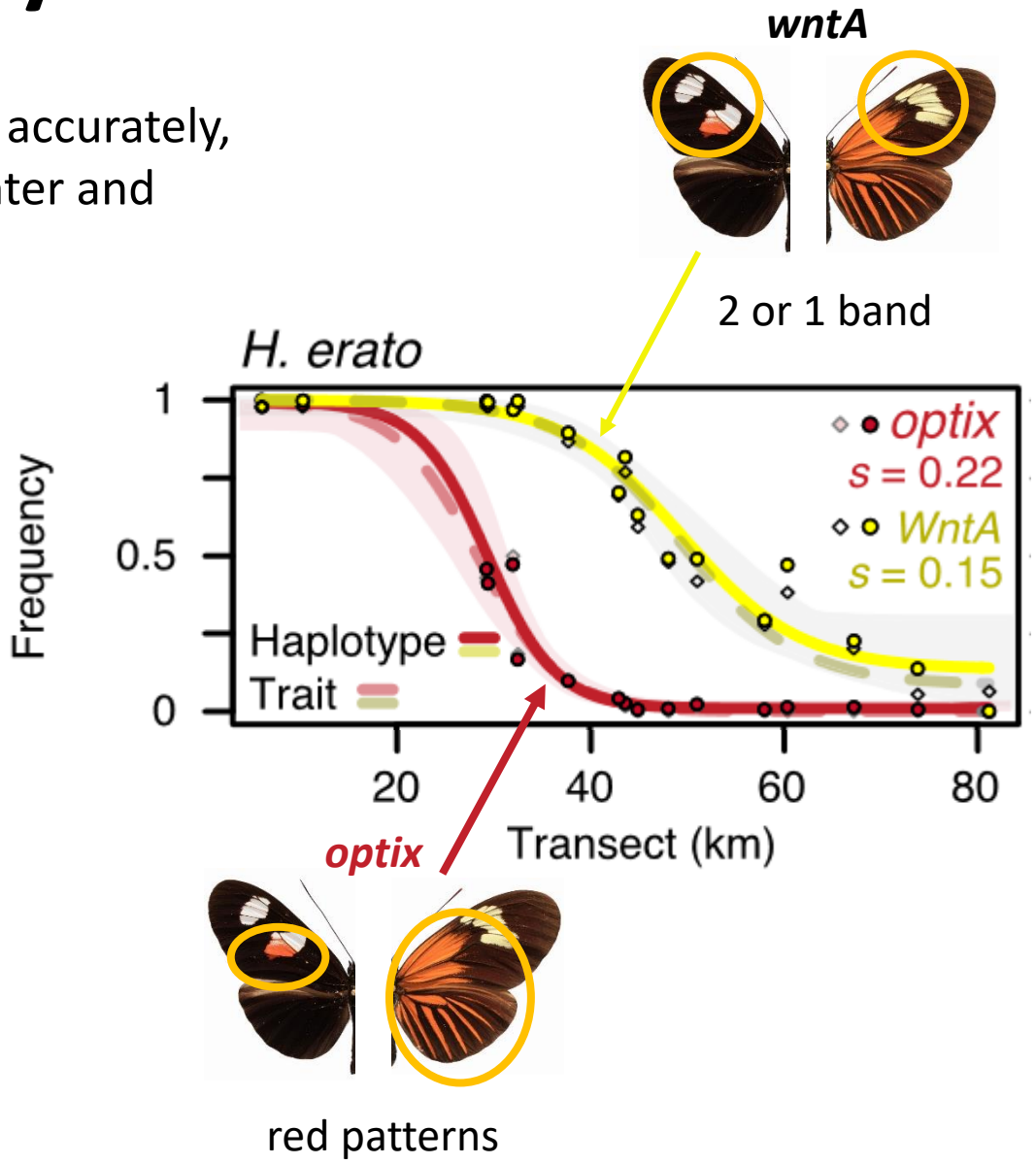
Patricio Salazar

Chris Jiggins

Heliconius erato hybrid zone

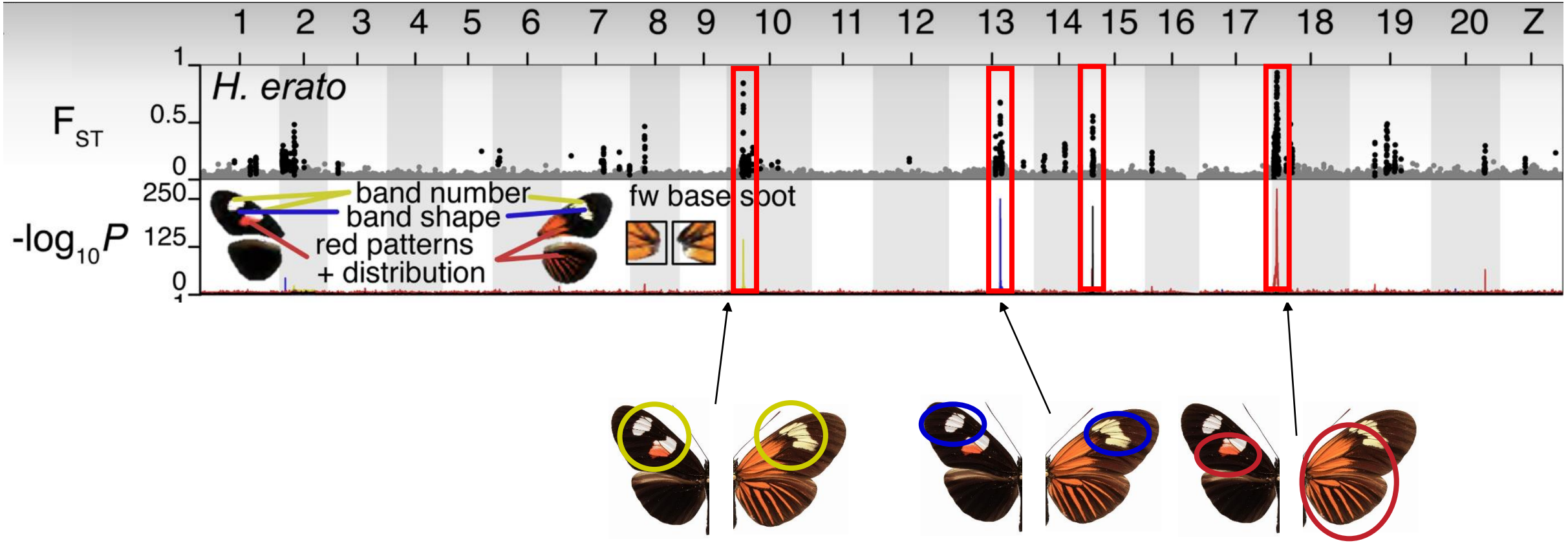
Many individuals per population to estimate allele frequency accurately, many populations along the hybrid zone to infer the cline center and shape accurately





Meier at al. 2021 PNAS

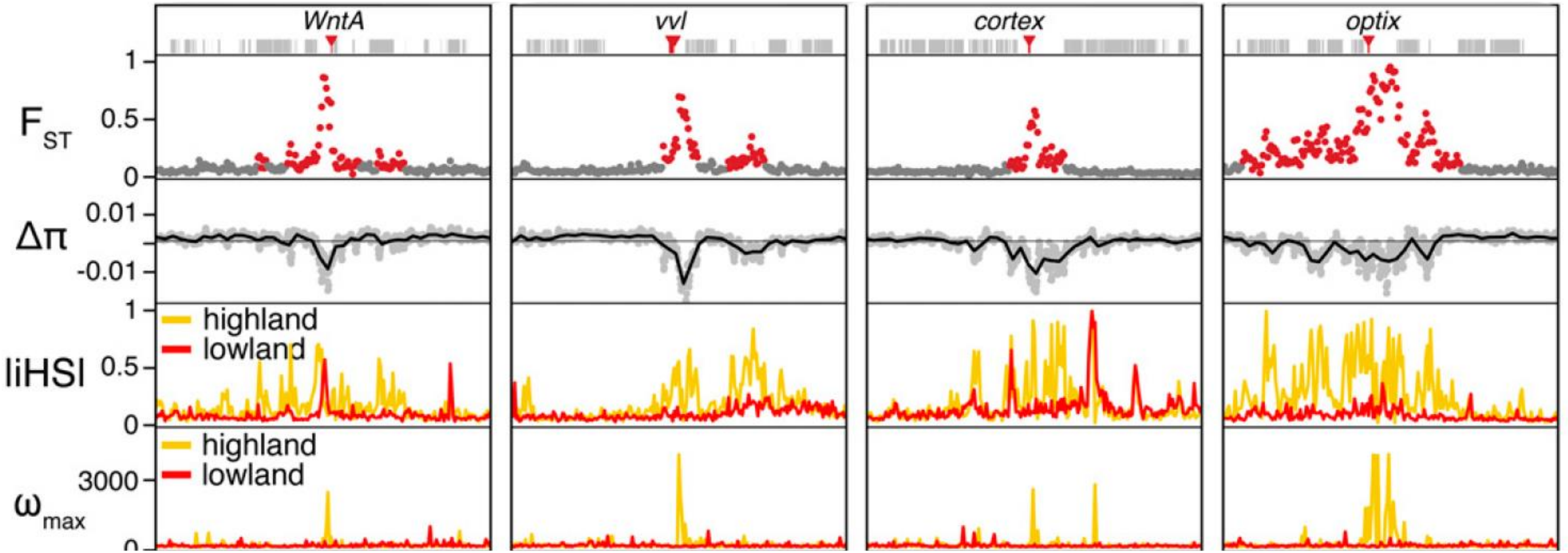
F_{ST} peaks coincide with GWAS peaks and show signatures of selective sweeps



2 or 1 band | band shape | red patterns

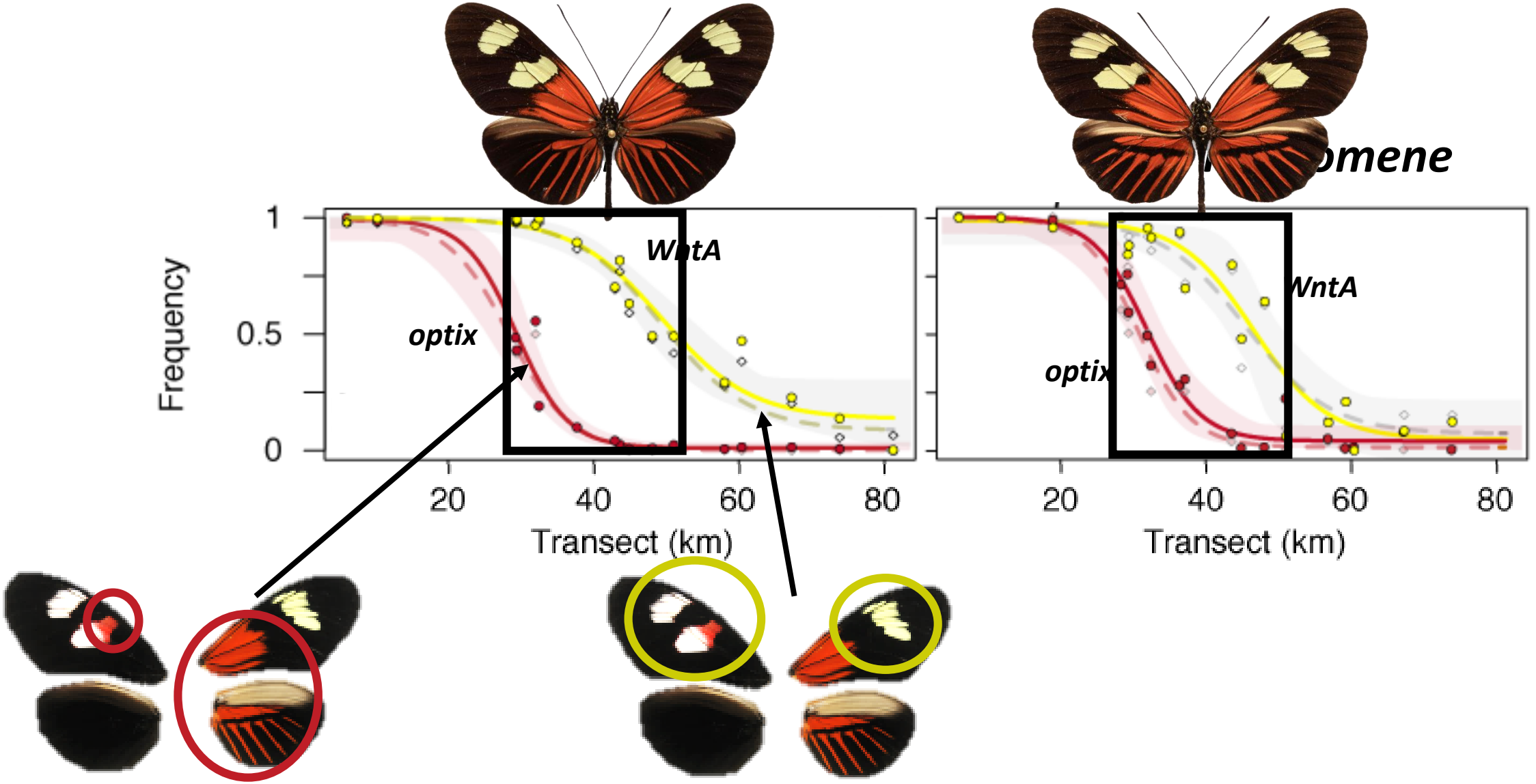
Meier at al. 2021 PNAS

F_{ST} peaks coincide with GWAS peaks and show signatures of selective sweeps



Meier at al. 2021 PNAS

In both *Heliconius erato* and *H. melpomene*, the clines at the major colour loci are shifted by 18 km

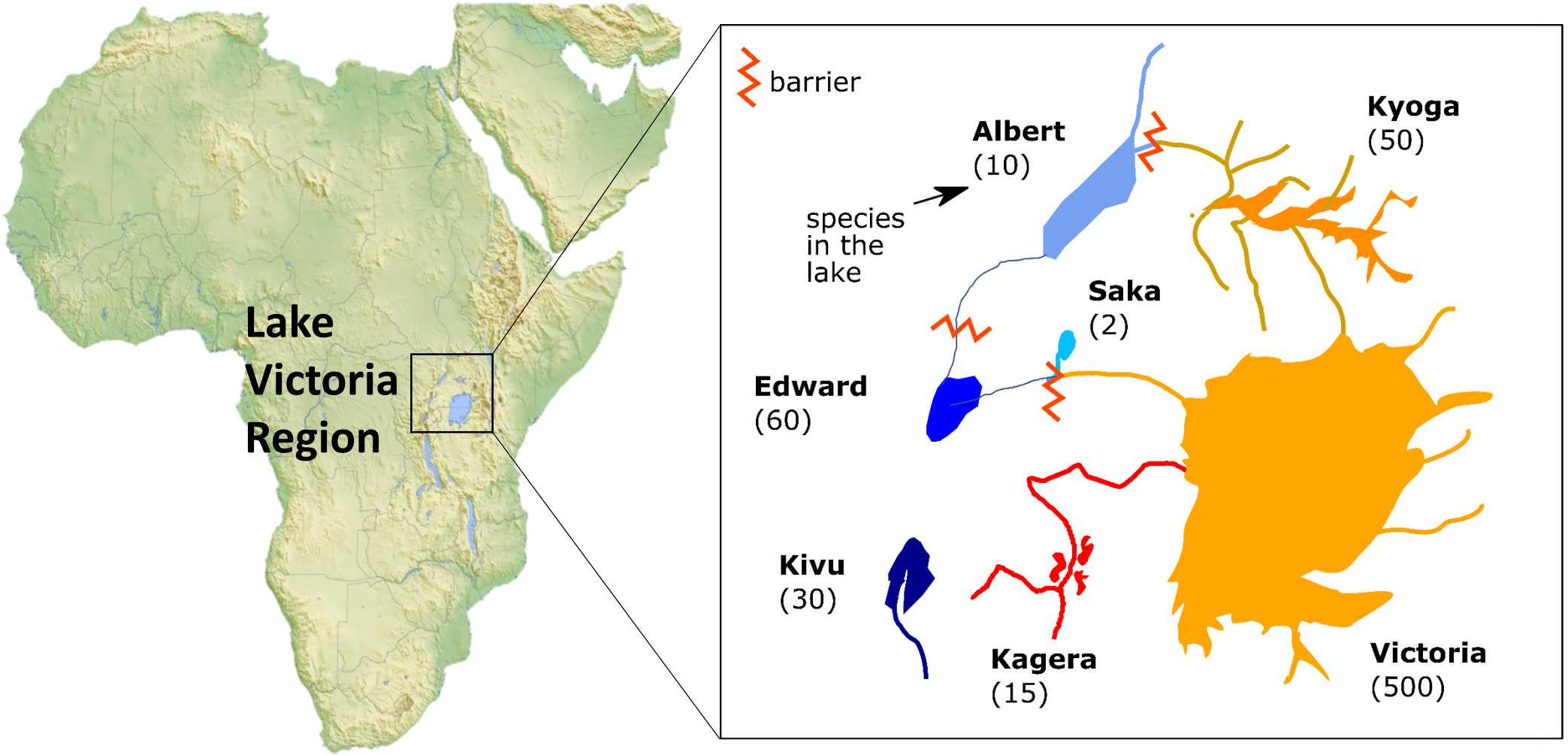


Case studies to discuss

Three examples with very different optimal sampling designs

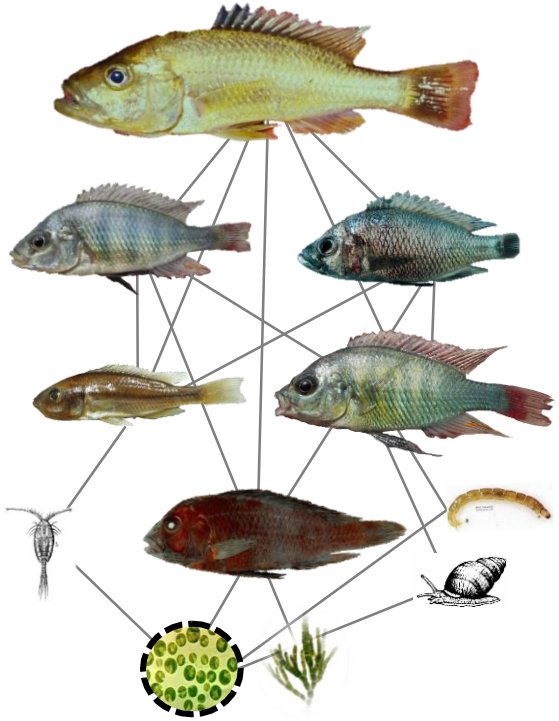
- Identifying barriers to gene flow and the genetic basis of relevant traits in a hybrid zone
- **Inferring if a fish species community evolved in a lake or if the different species independently colonised the lake.**

The same lineage of cichlid fishes diversified in all major lakes in the Lake Victoria Region in only 150,000 years

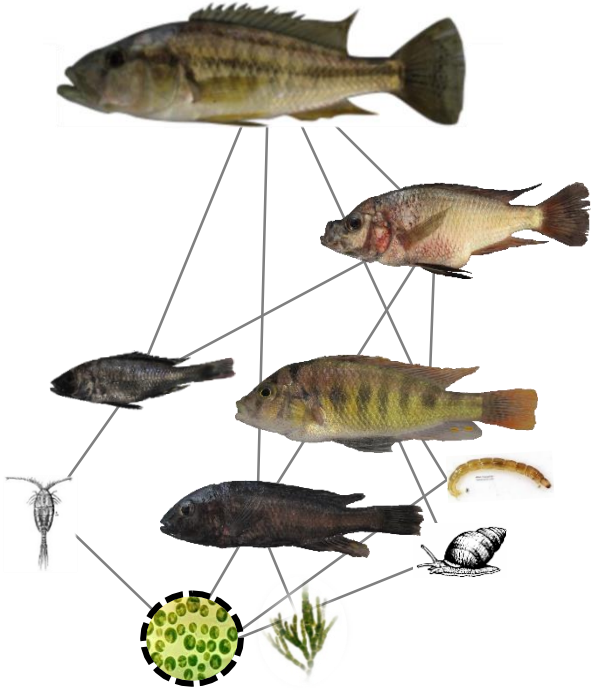


In each lake, cichlids occupy a wide range of ecological niches. But Lake Victoria is much younger! Only 16,000 years!

Lake Edward

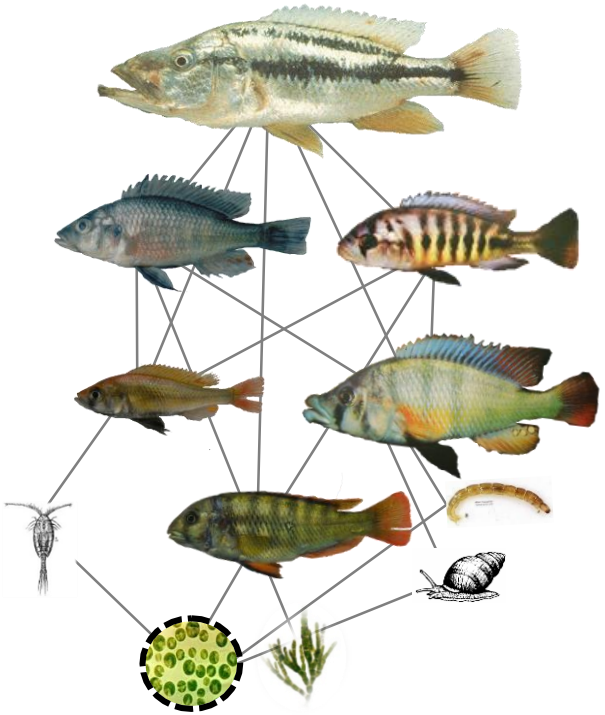


Lake Kivu

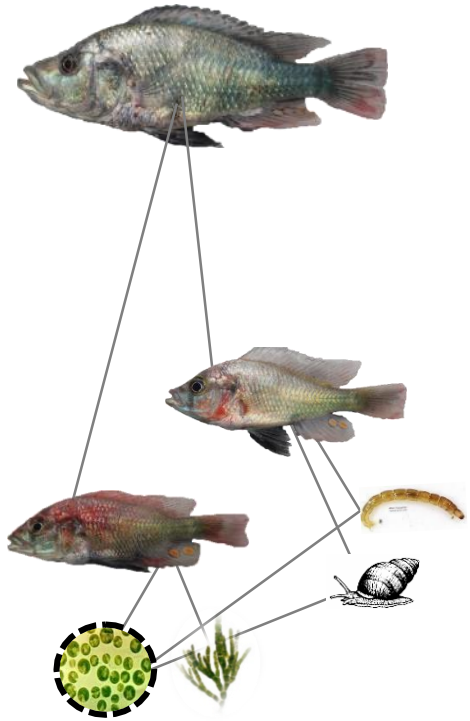


Less than 16,000 years old!

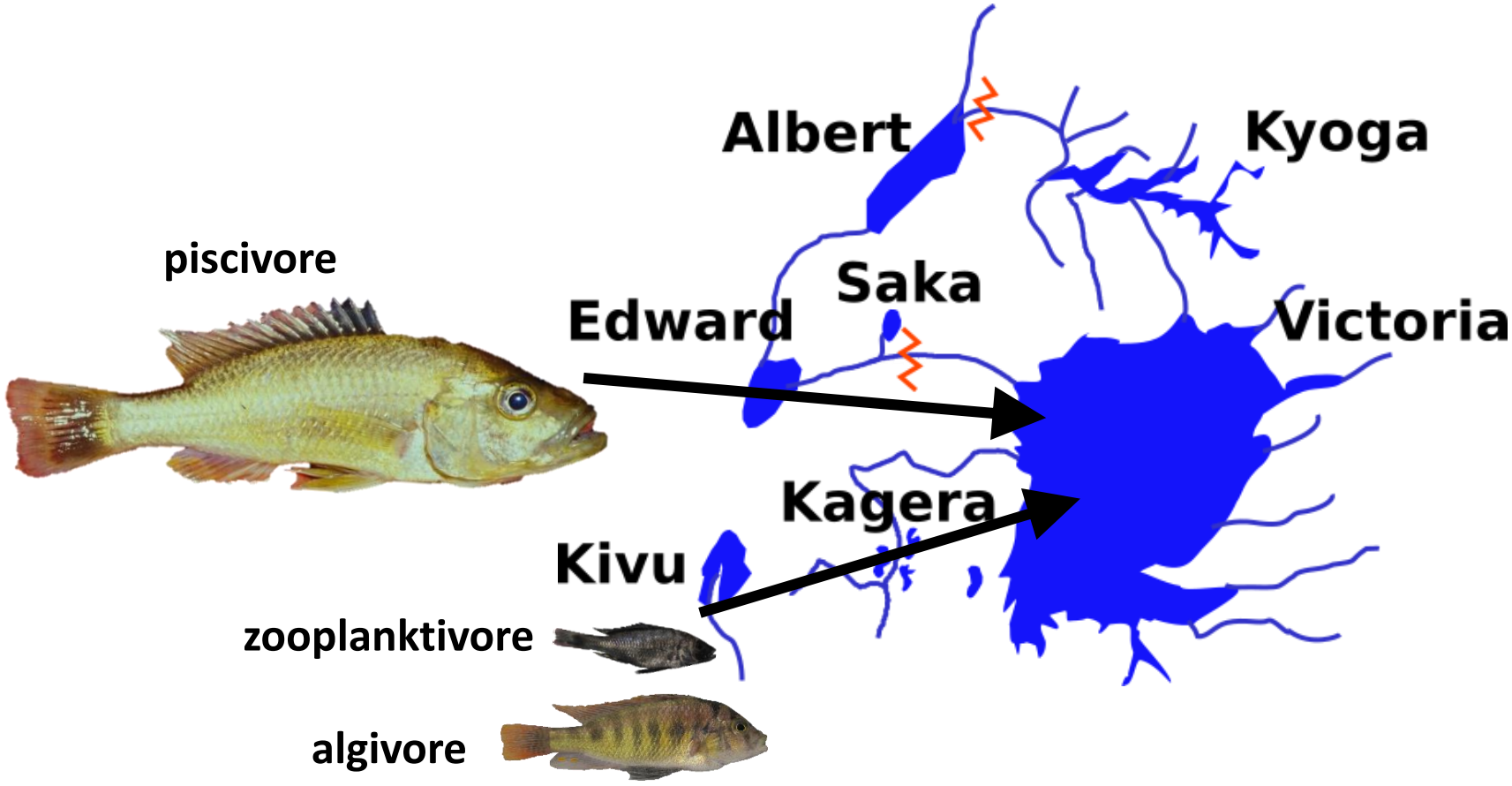
Lake Victoria (500 species)



Kagera Lakes

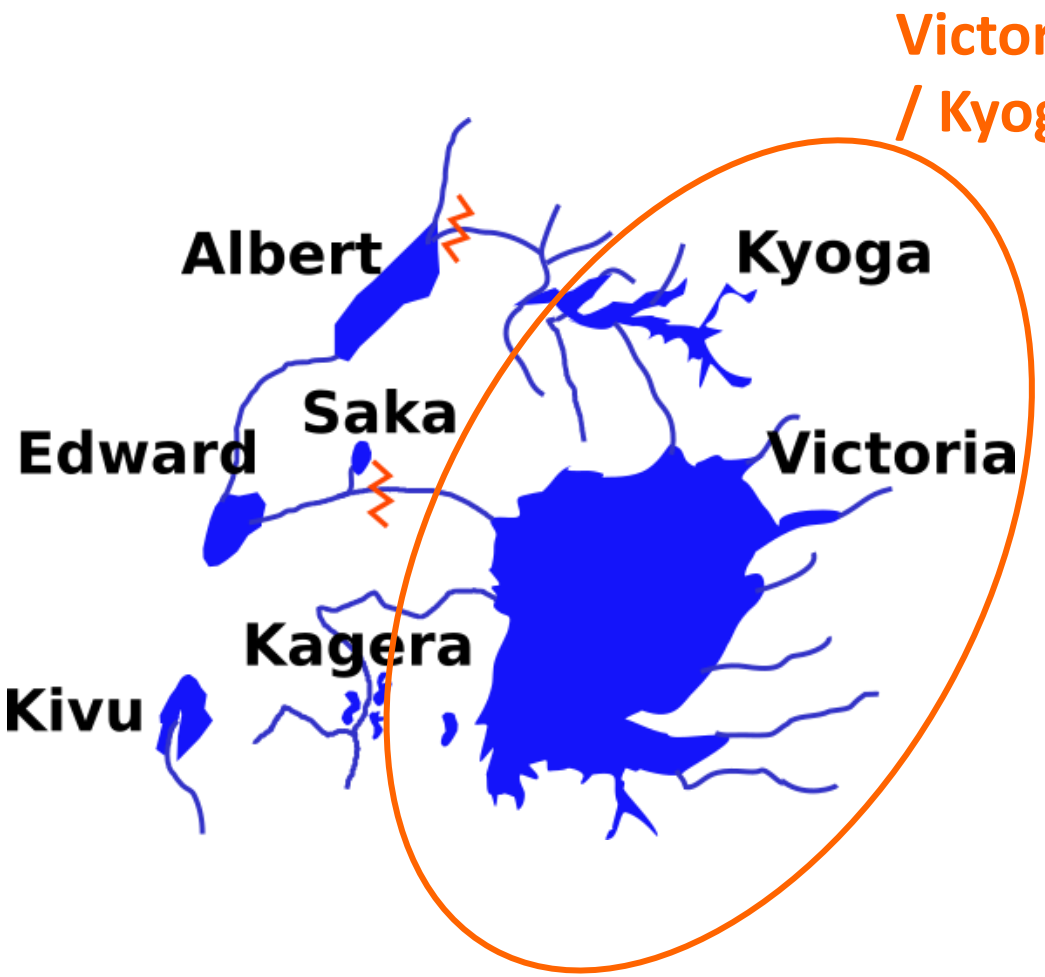


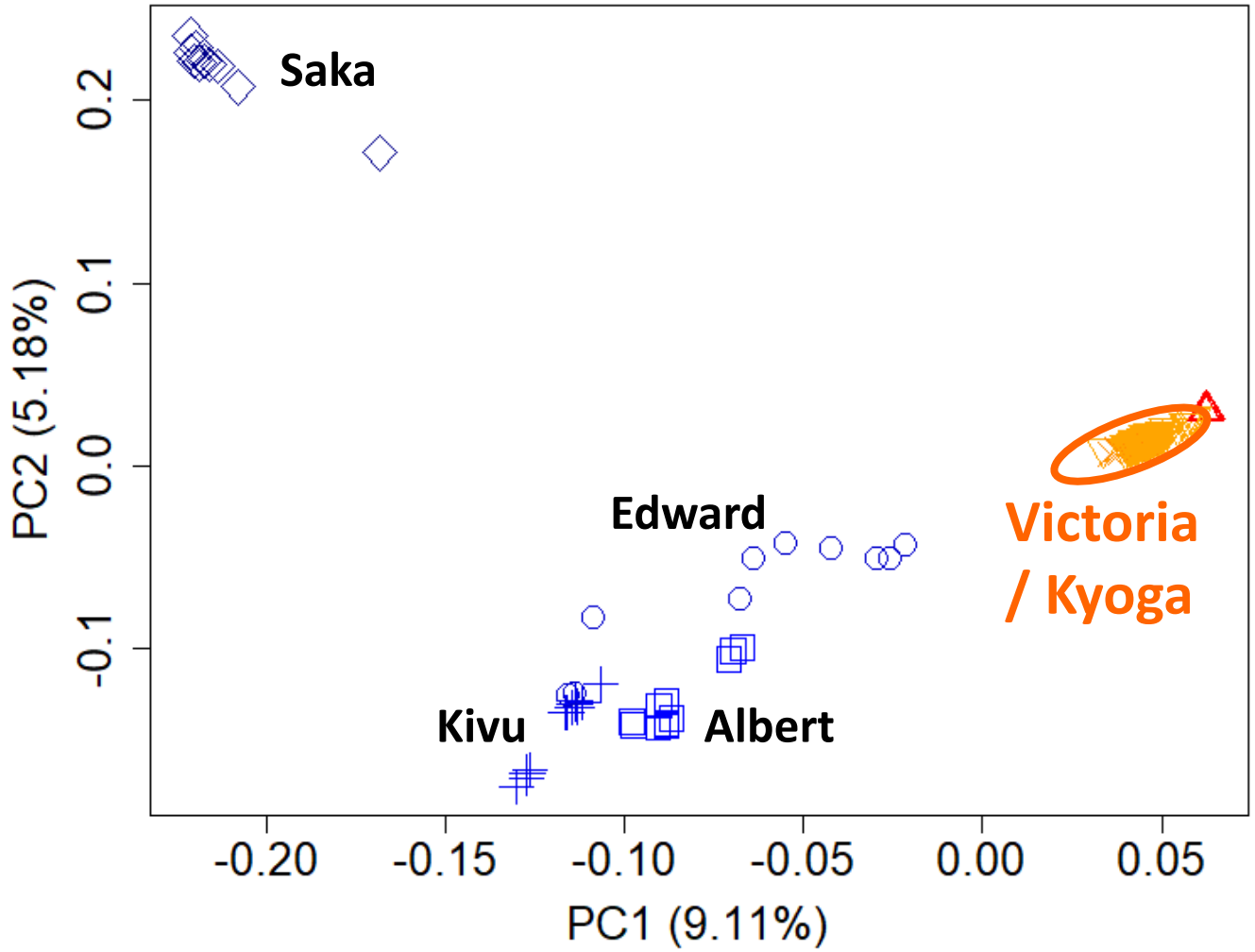
Did the different ecological groups colonise Lake Victoria independently?



PCA on whole-genomes separates Lake Victoria/Kyoga cichlids from others

(152 genomes, 1.6M LD-pruned SNPs)

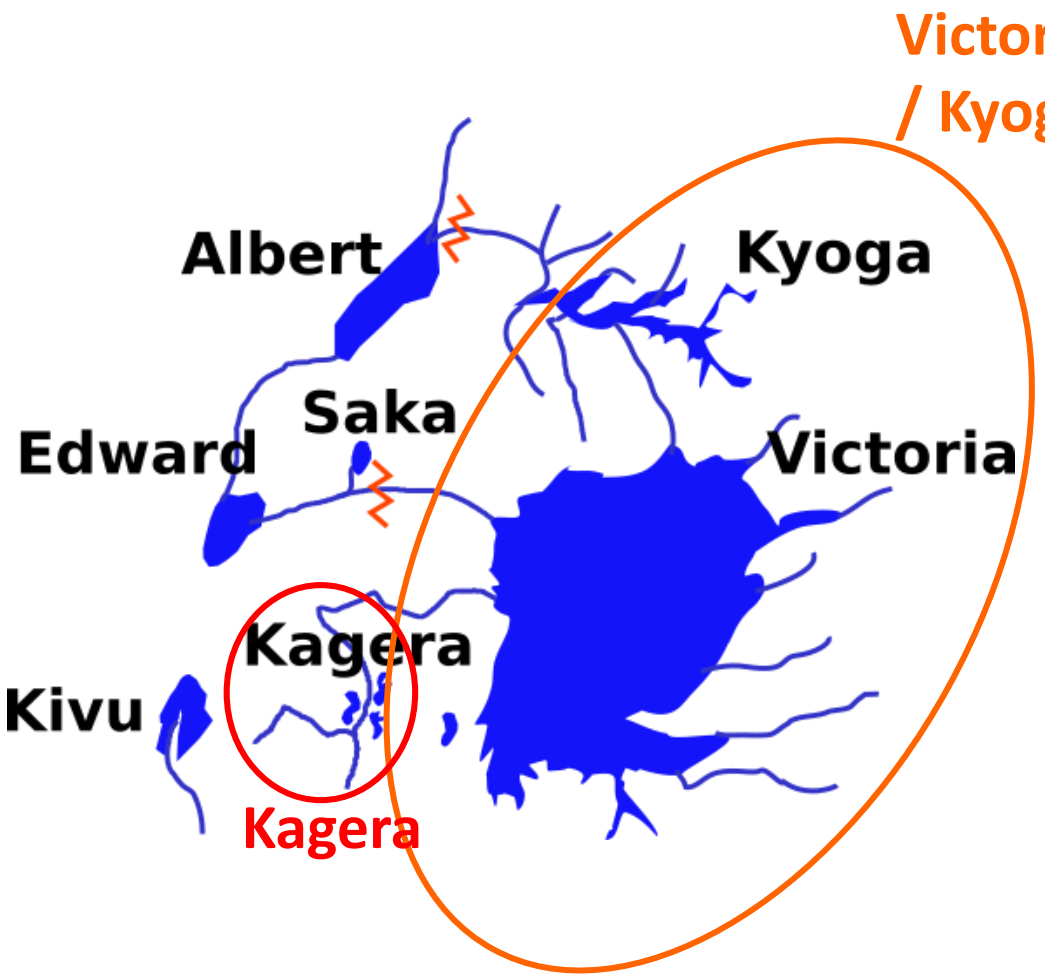


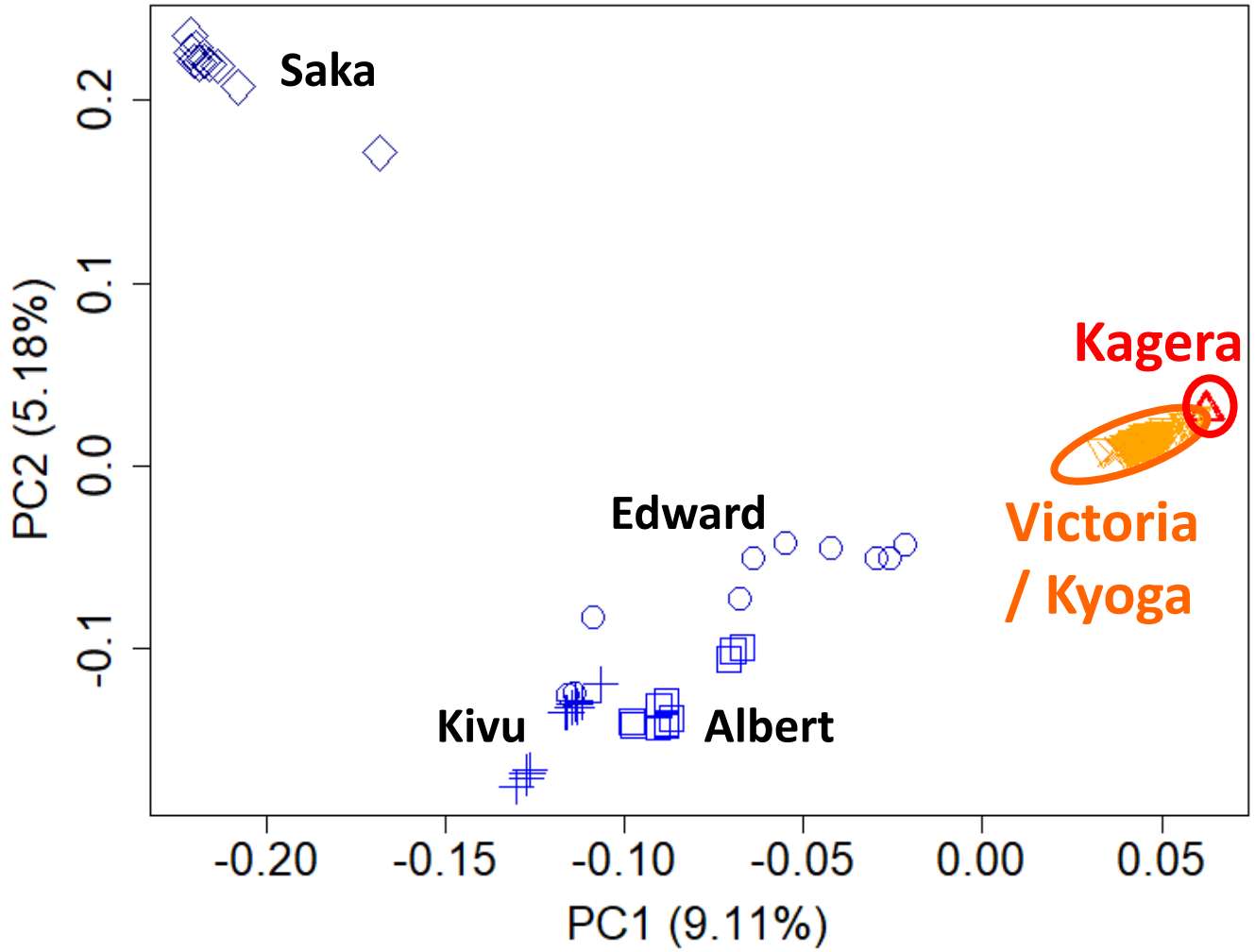


Meier *et al*. (2023) *Science*

PCA on whole-genomes separates Lake Victoria/Kyoga cichlids from others

(152 genomes, 1.6M LD-pruned SNPs)

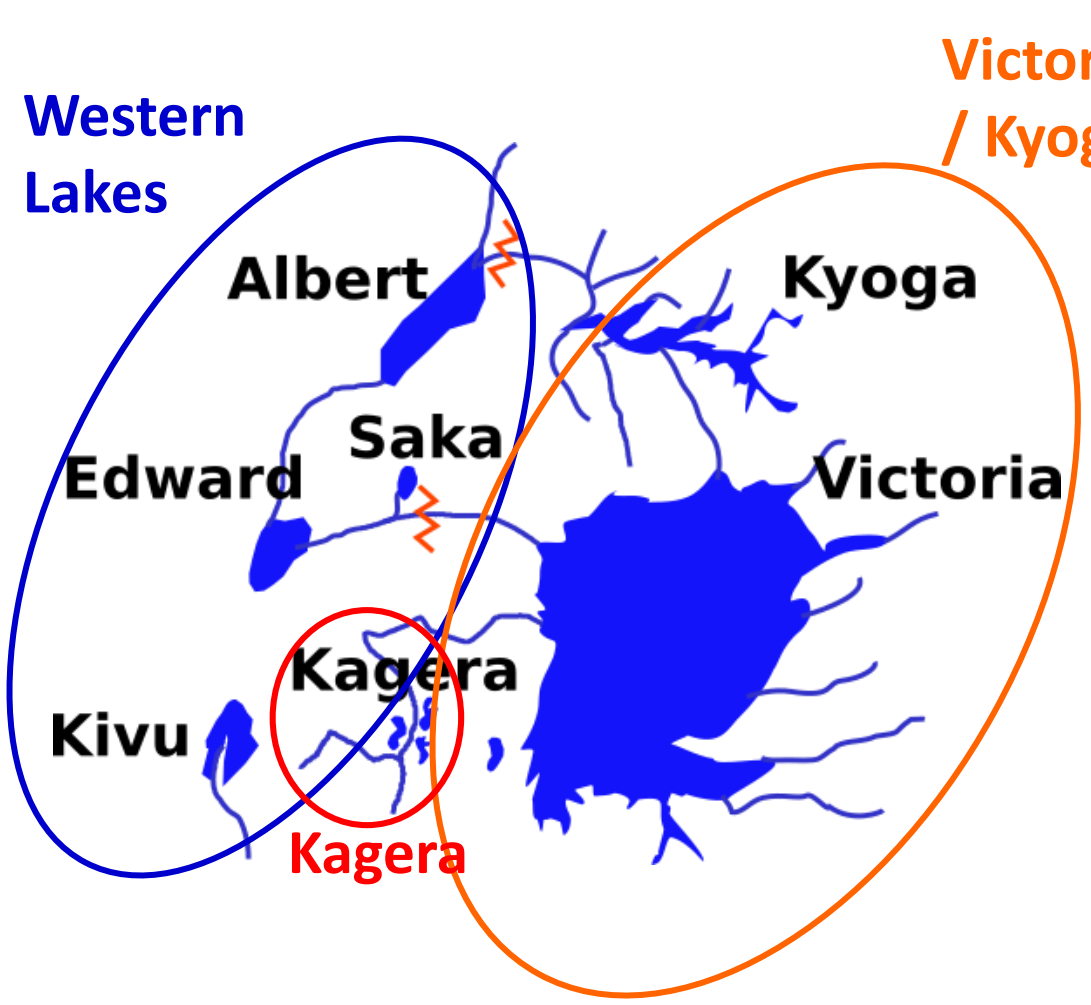


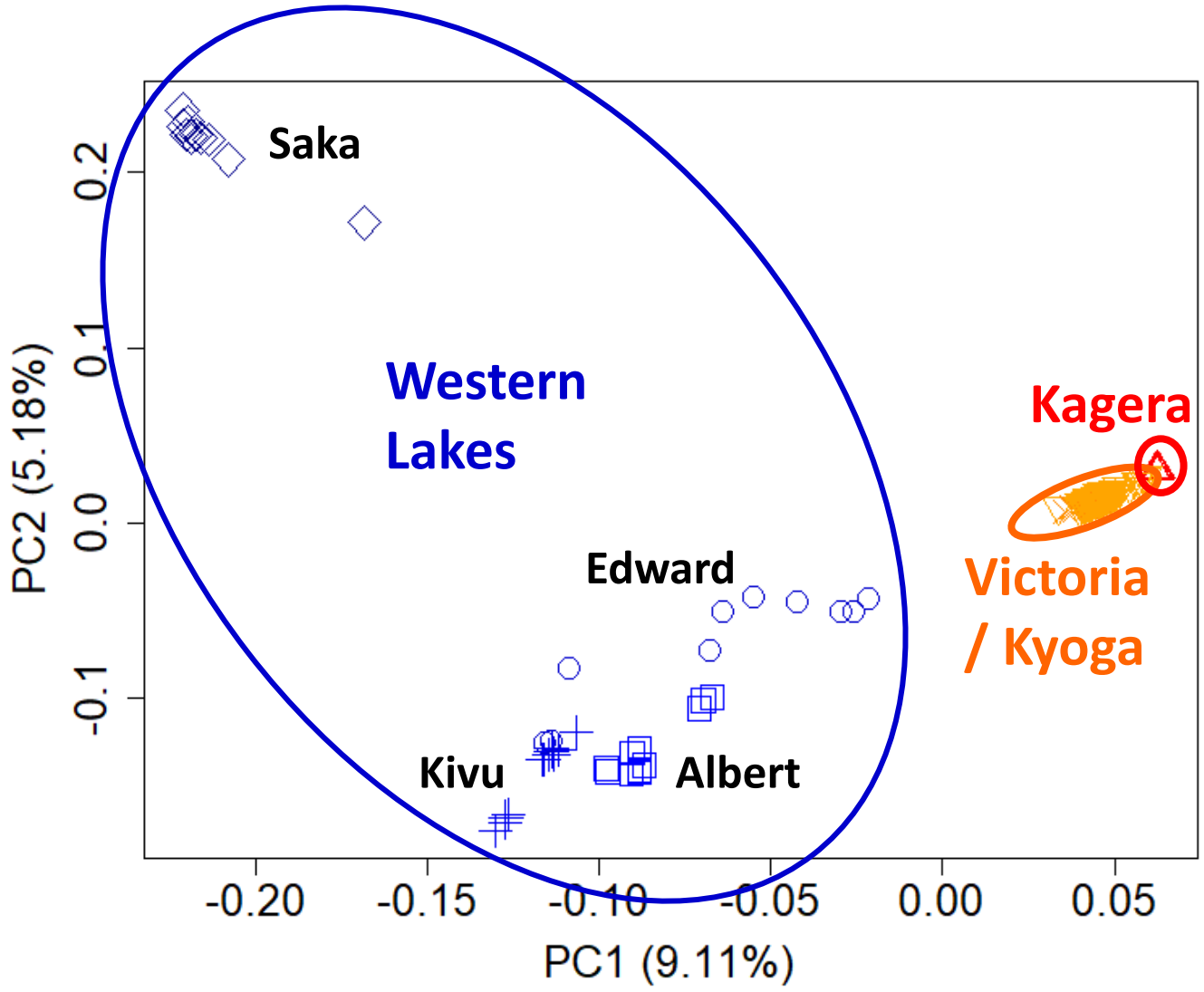


Meier *et al*. (2023) *Science*

PCA on whole-genomes separates Lake Victoria/Kyoga cichlids from others

(152 genomes, 1.6M LD-pruned SNPs)

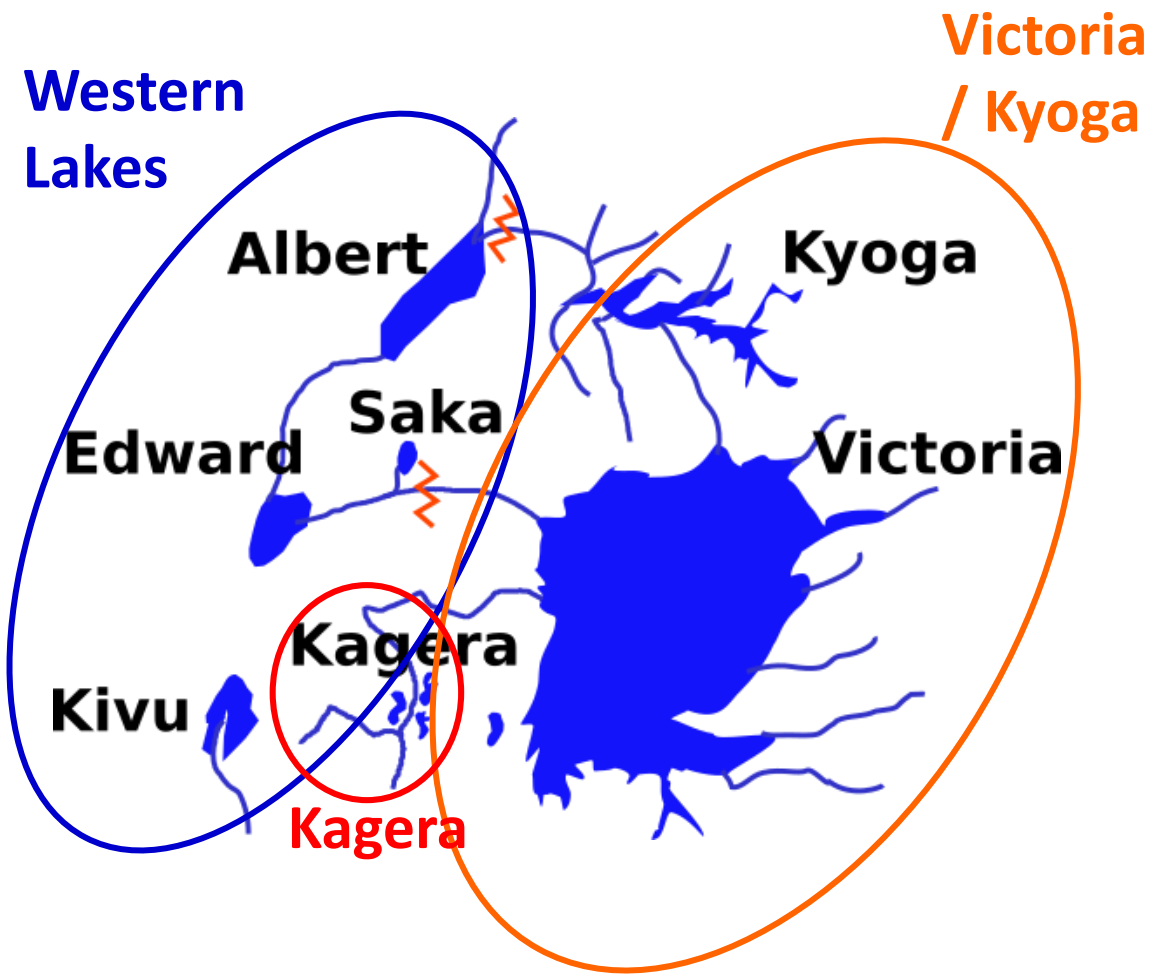




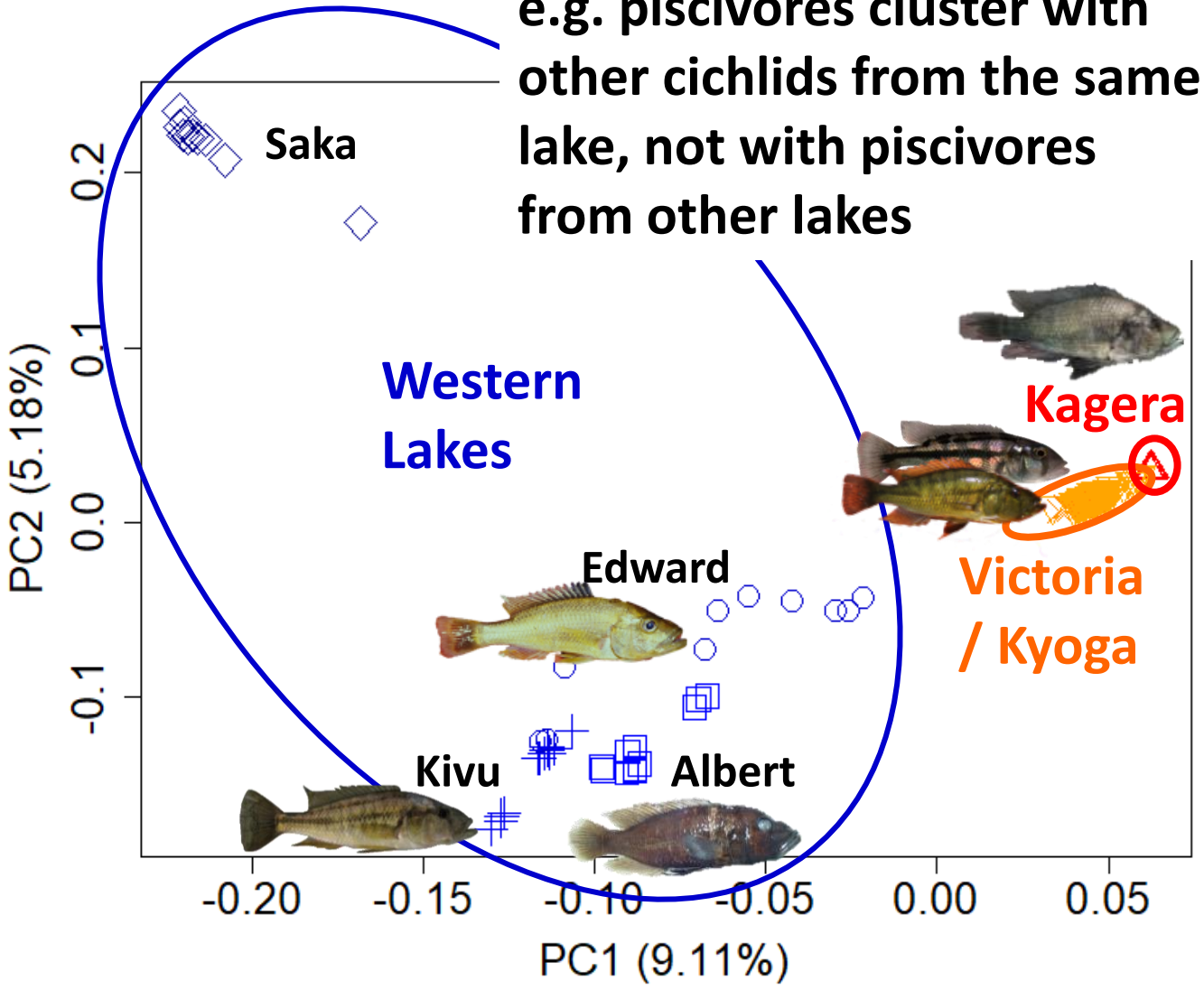
Meier *et al*. (2023) *Science*

PCA on whole-genomes separates Lake Victoria/Kyoga cichlids from others

(152 genomes, 1.6M LD-pruned SNPs)



e.g. piscivores cluster with other cichlids from the same lake, not with piscivores from other lakes

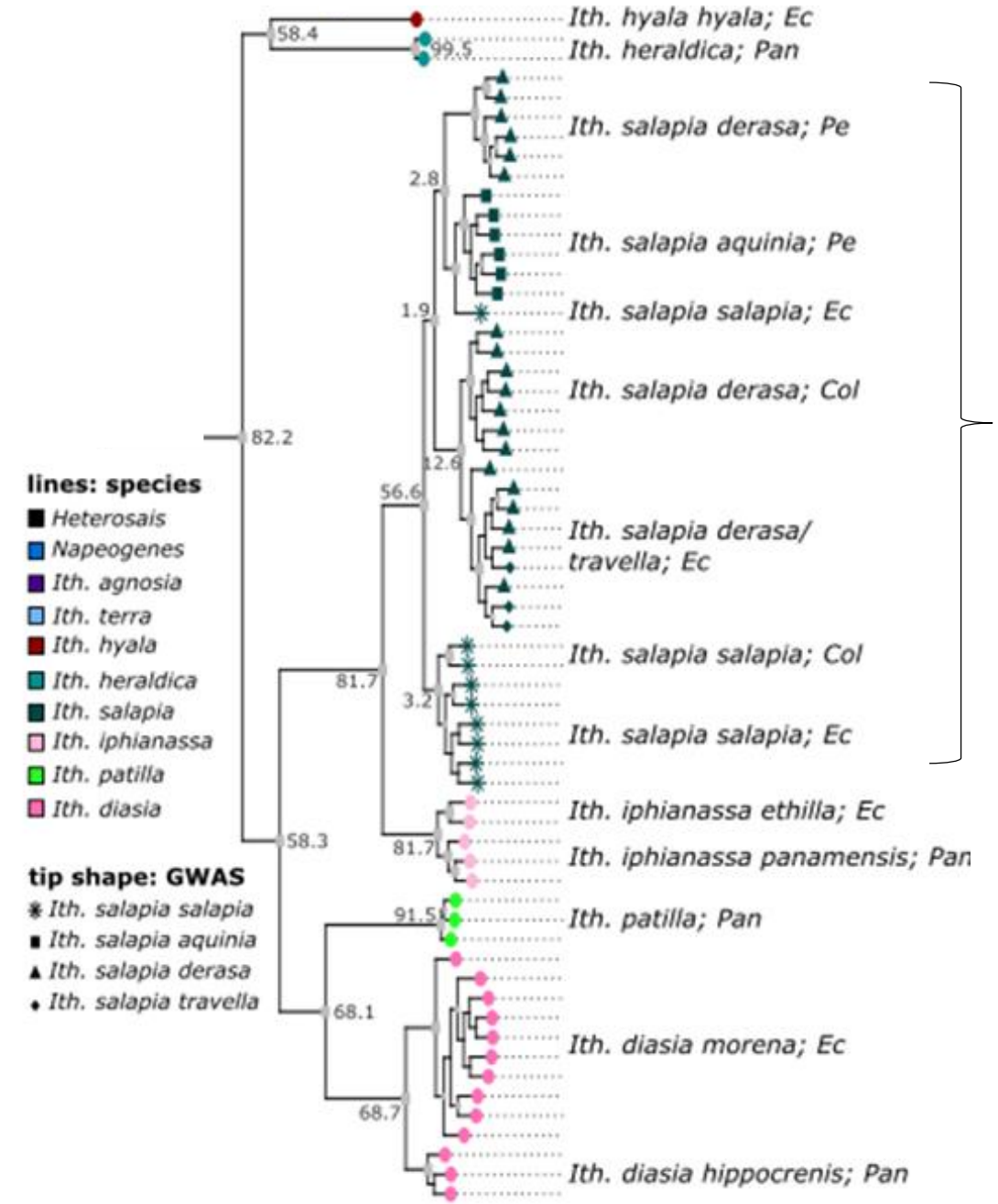


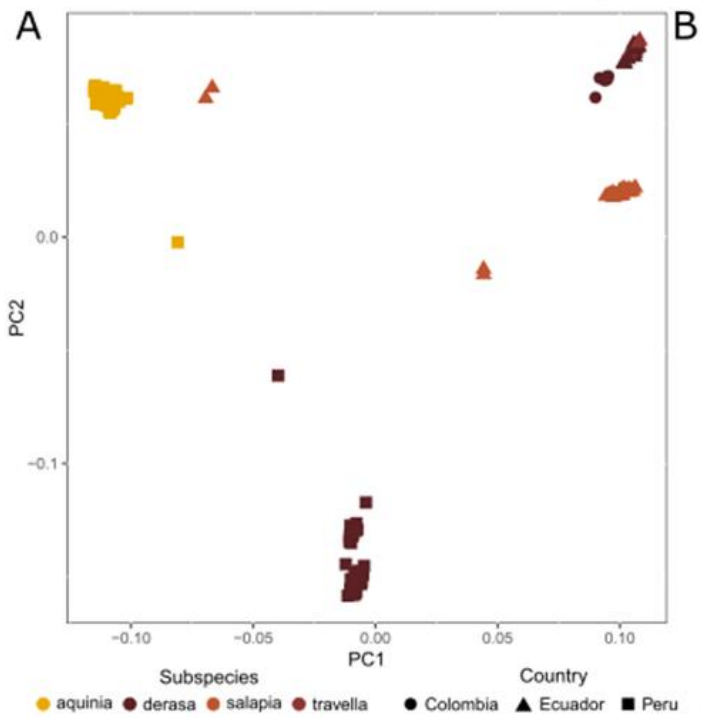
Meier *et al*. (2023) *Science*

Case studies to discuss

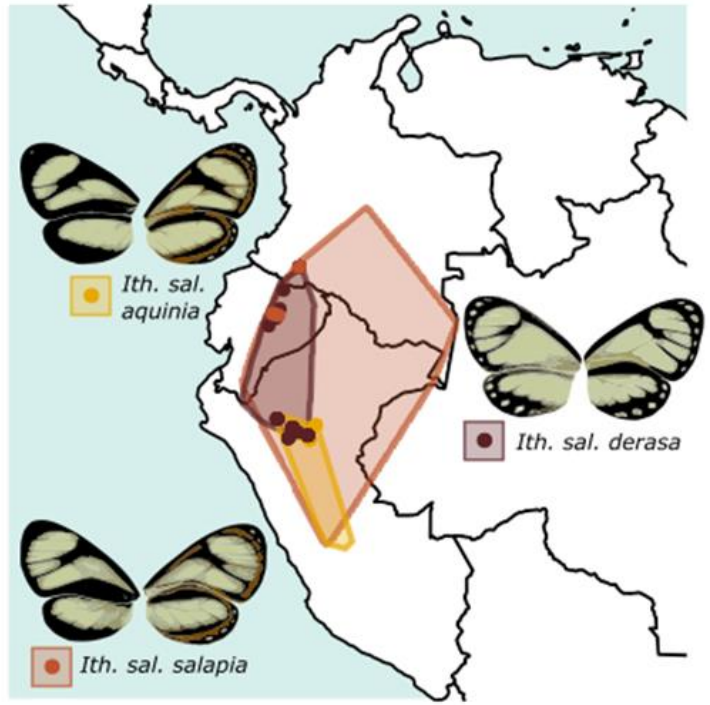
Three examples with very different optimal sampling designs

- Identifying barriers to gene flow and the genetic basis of relevant traits in a hybrid zone
- Inferring if a fish species community evolved in a lake or if the different species independently colonised the lake
- **Inferring if chromosomal rearrangements contribute to speciation in a butterfly species complex**

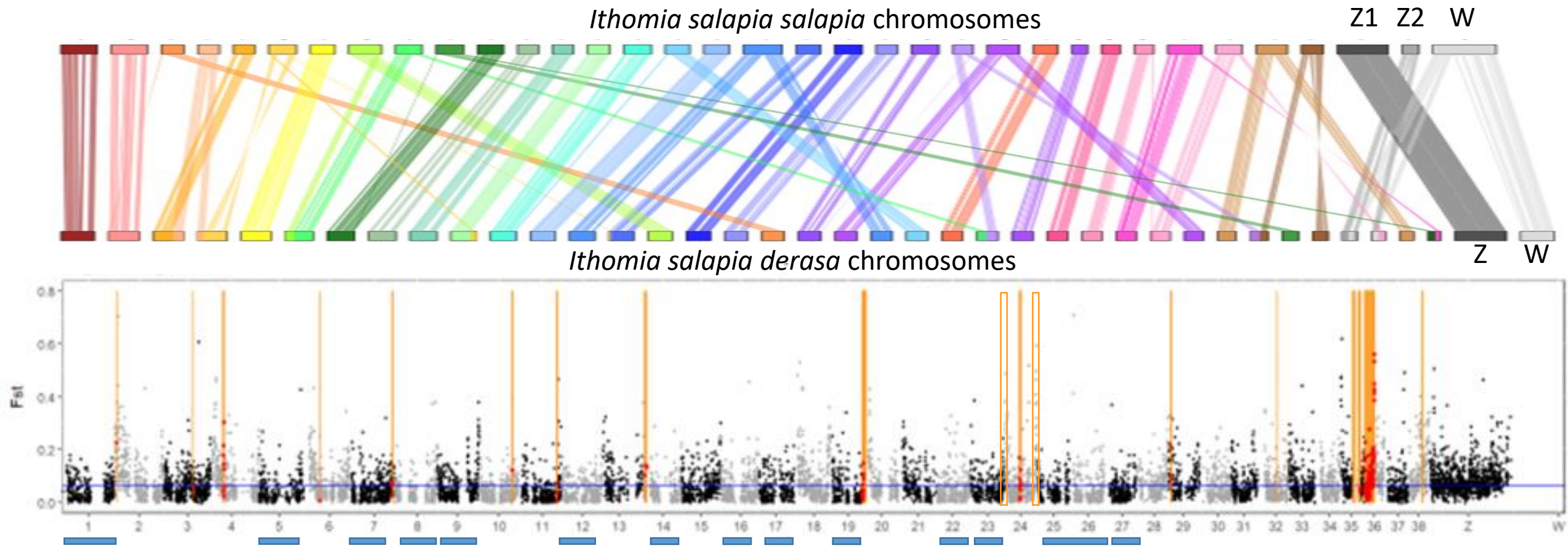




Ithomia salapia



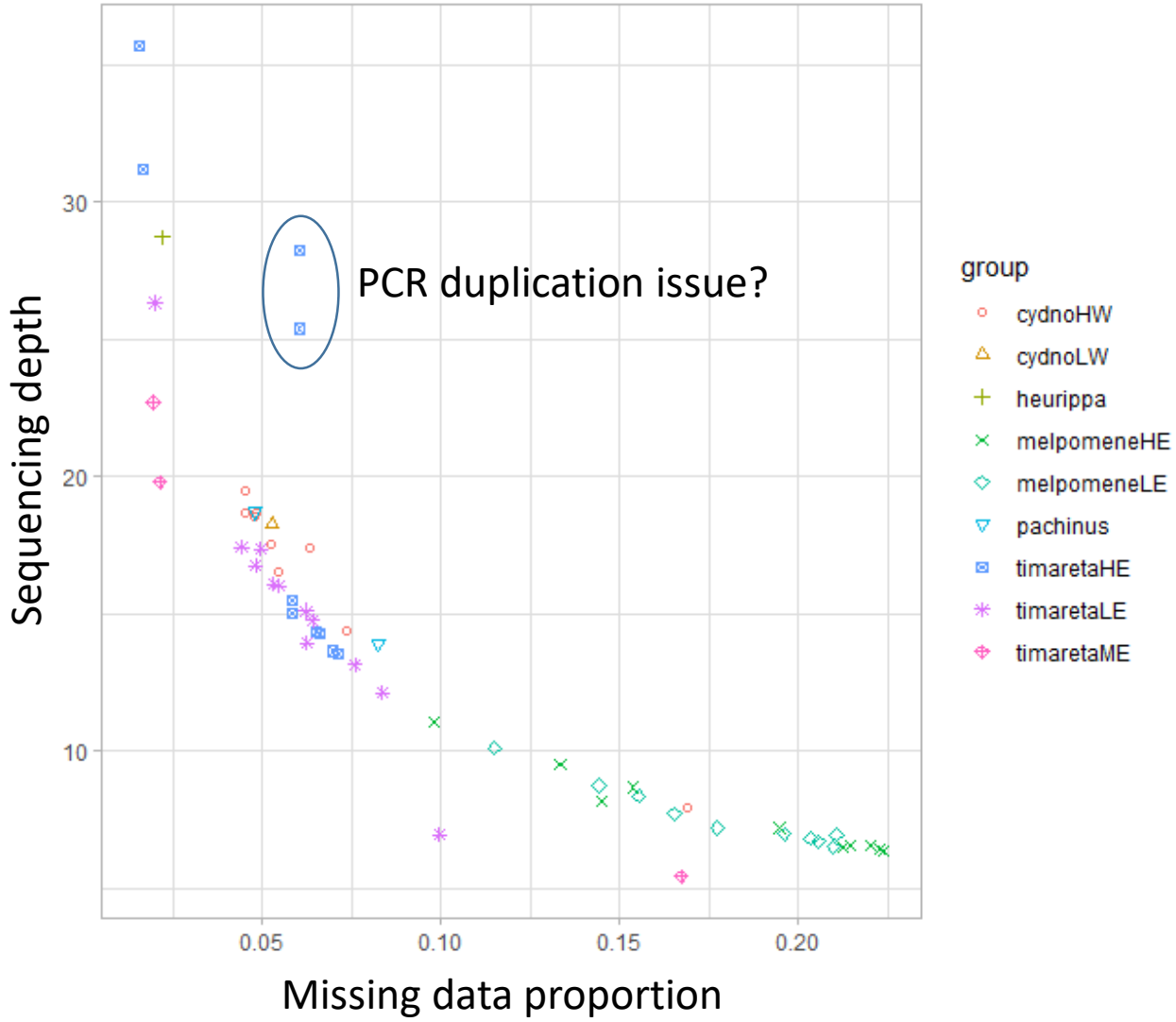
Ithomia salapia «subspecies» show massive chromosomal rearrangements, likely contributing to speciation



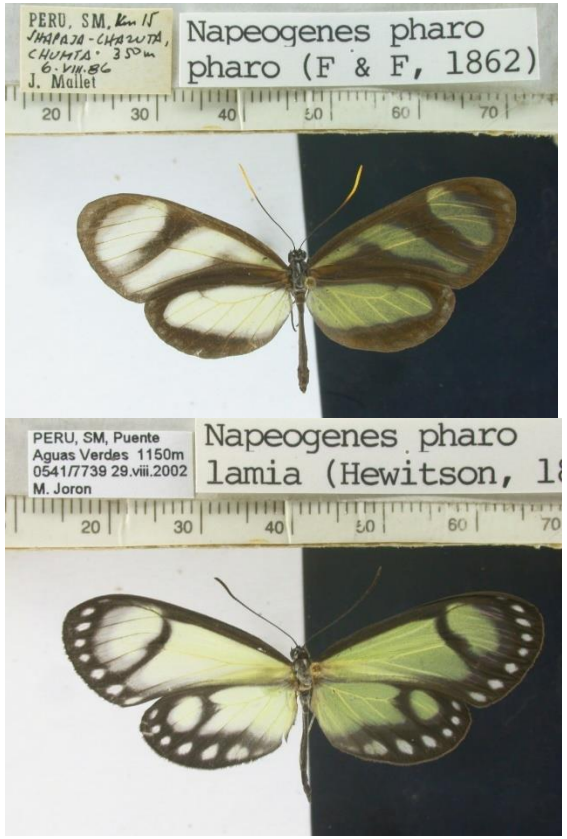
Once you get your data, make sure you trust it

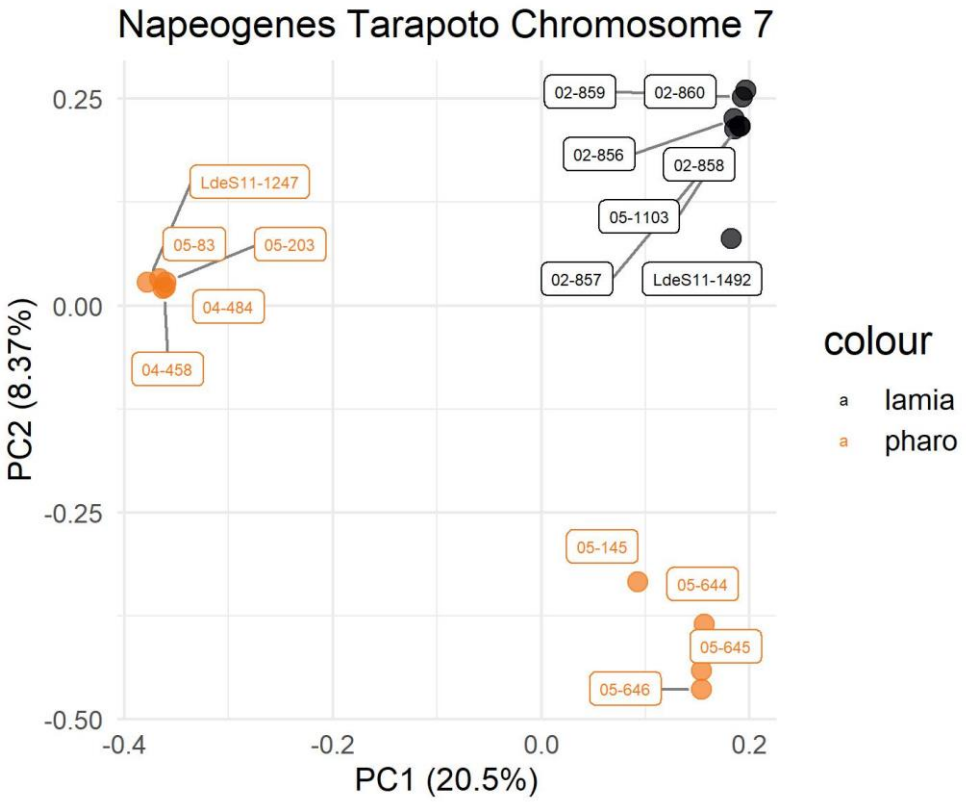
Get a feeling for your data, do quality checks at all stages and plot your data

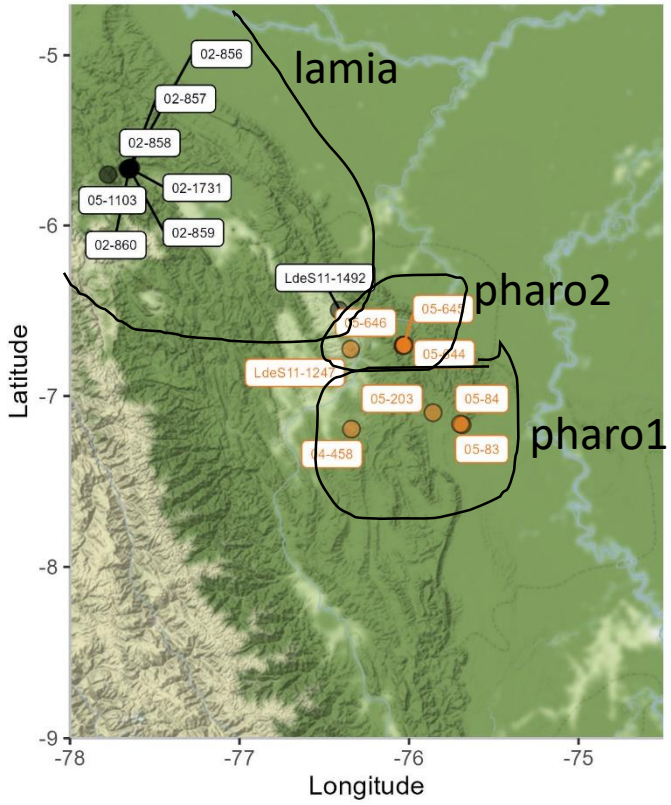
- Do any individuals show low mapping proportion?
- Do any individuals show high proportion of missing data or low sequencing depth?



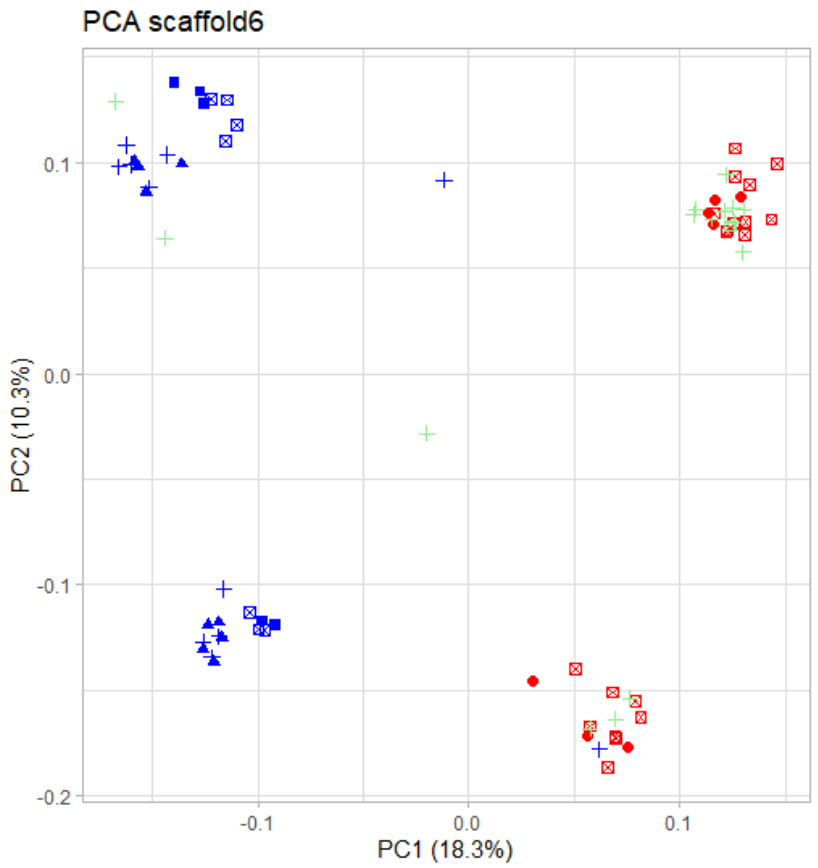
Once you get your data, make sure you trust it

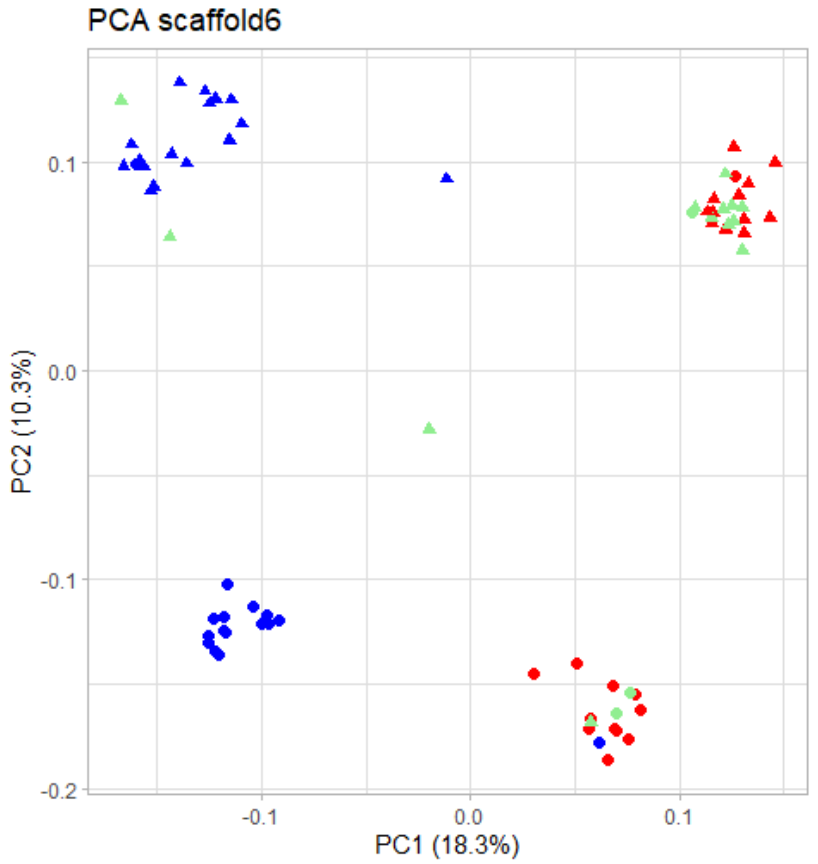






Once you get your data, make sure you trust it





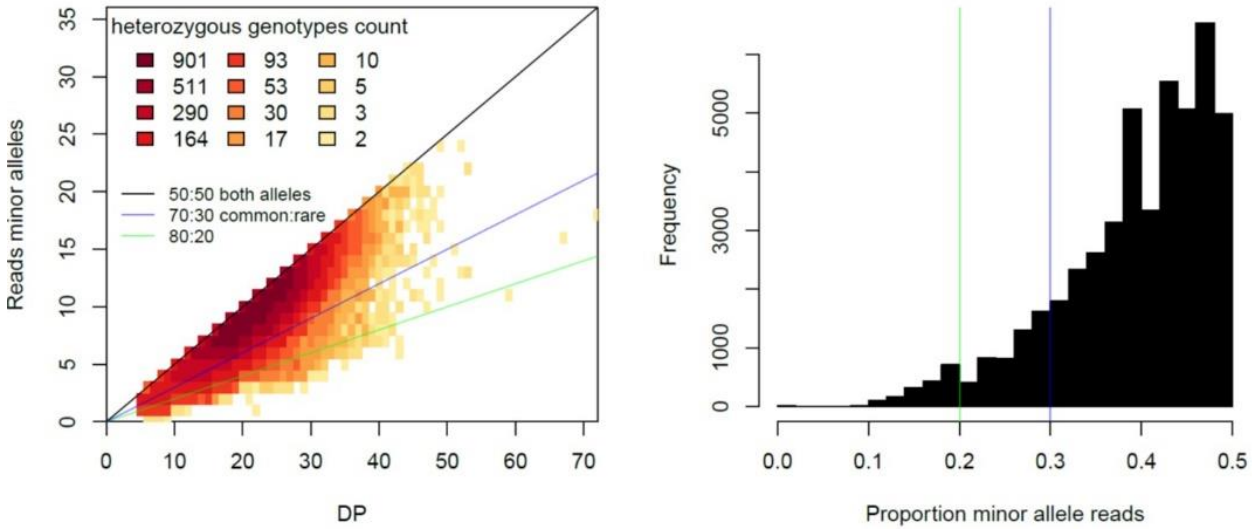
Ithomia salapia aquinia (Amazon)

Ithomia salapia derasa (Andes)

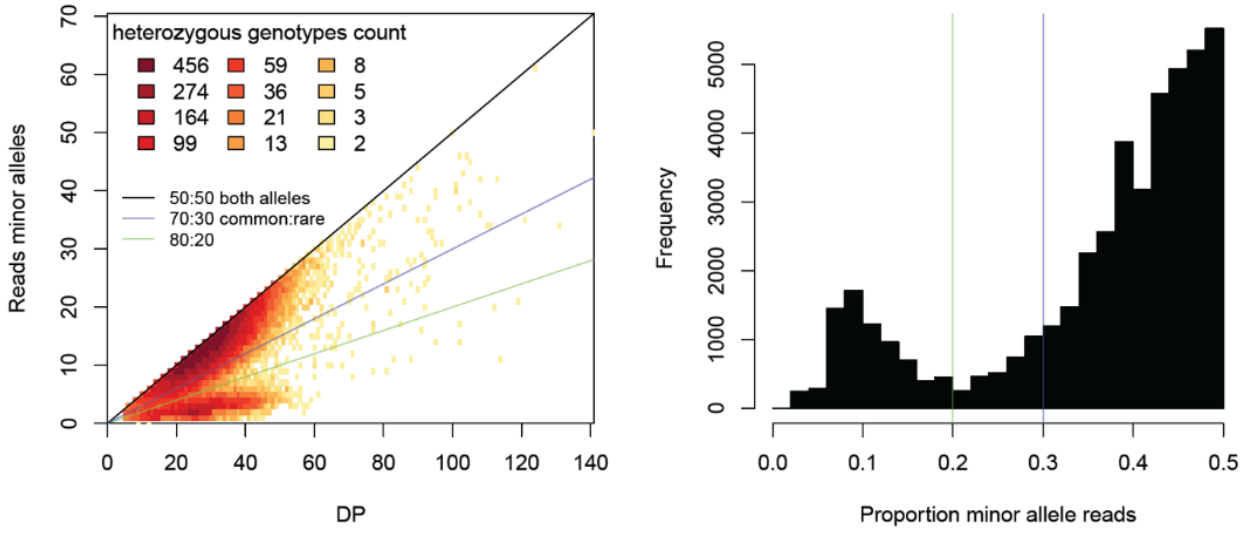
Once you get your data, make sure you trust it

Contamination can cause issues, including erroneous inference of introgression

well-sequenced individual



Individual with likely contamination

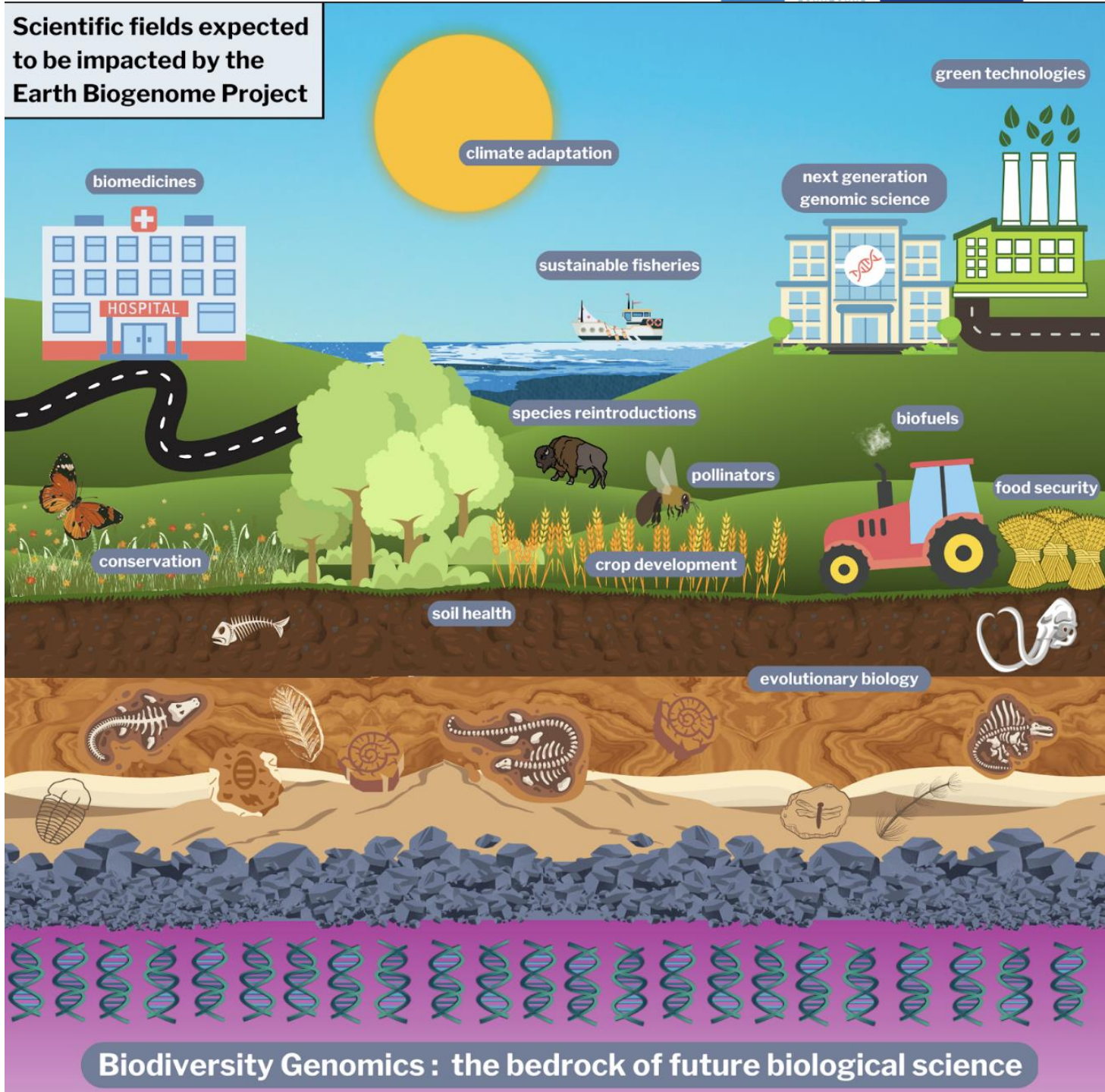


https://speciationgenomics.github.io/allelicBalance

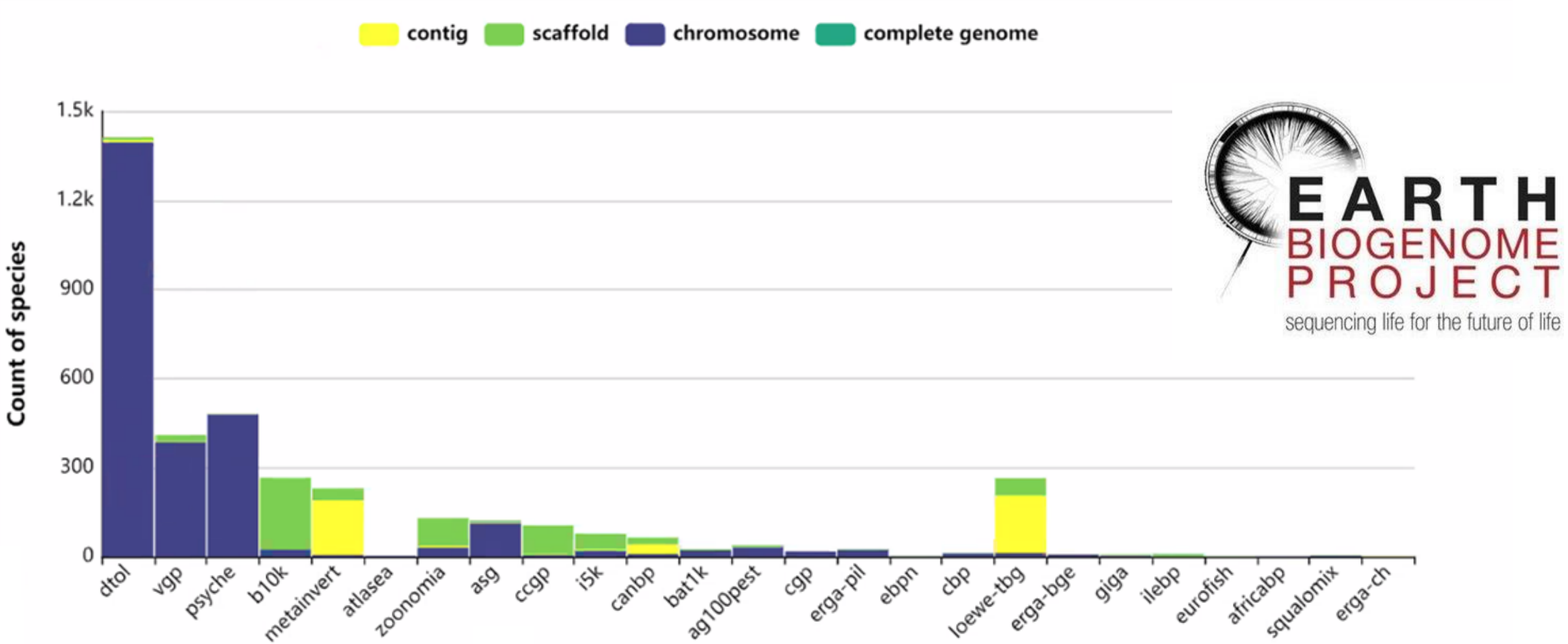

EARTH
BIOGENOME
PROJECT
Darwin
TREE
of
LIFE
Project Psyche
African BioGenome Project
Genomics for the future of biological diversity across Africa
ATLASea
Atlas des génomes marins
VGP
VERTEBRATE
GENOMES
PROJECT
A PROJECT OF THE G10K CONSORTIUM
BAT 1K
ERGA
CREATING A NEW FOUNDATION FOR BIOLOGY
Sequencing Life for the
Future of Life

Reference genomes have global impact in a wide range of areas important for planetary health:

- Biomedicines
- Climate adaptation
- Biomonitoring
- Green technologies
- Next generation genomic science
- Sustainable fisheries
- Species reintroductions
- Conservation
- Pollinators
- Biofuels
- Food security
- Crop development
- Soil health
- Evolutionary biology



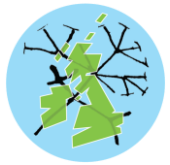
Large reference genome projects



EARTH BIOGENOME PROJECT

sequencing life for the future of life

Major Projects at the Tree of Life Programme at the Wellcome Sanger Institute, Cambridge, UK



- ★ **Darwin Tree of Life Project**
  - 70,000 species from Britain and Ireland [funded for ~8,000 species to 2026 currently]

- ★ **Aquatic Symbiosis Genomics**
  - 1,000 species (500 symbiotic systems) from marine and freshwater

- ★ **Project Psyche – Lepidoptera genomes for Europe**
  - Sequencing all 11,000 moths and butterflies of Europe

- ★ **Vertebrate Genomes Project**
  - VGP Phase 1 (ordinal - 260 species) and Phase 2 (family) goals

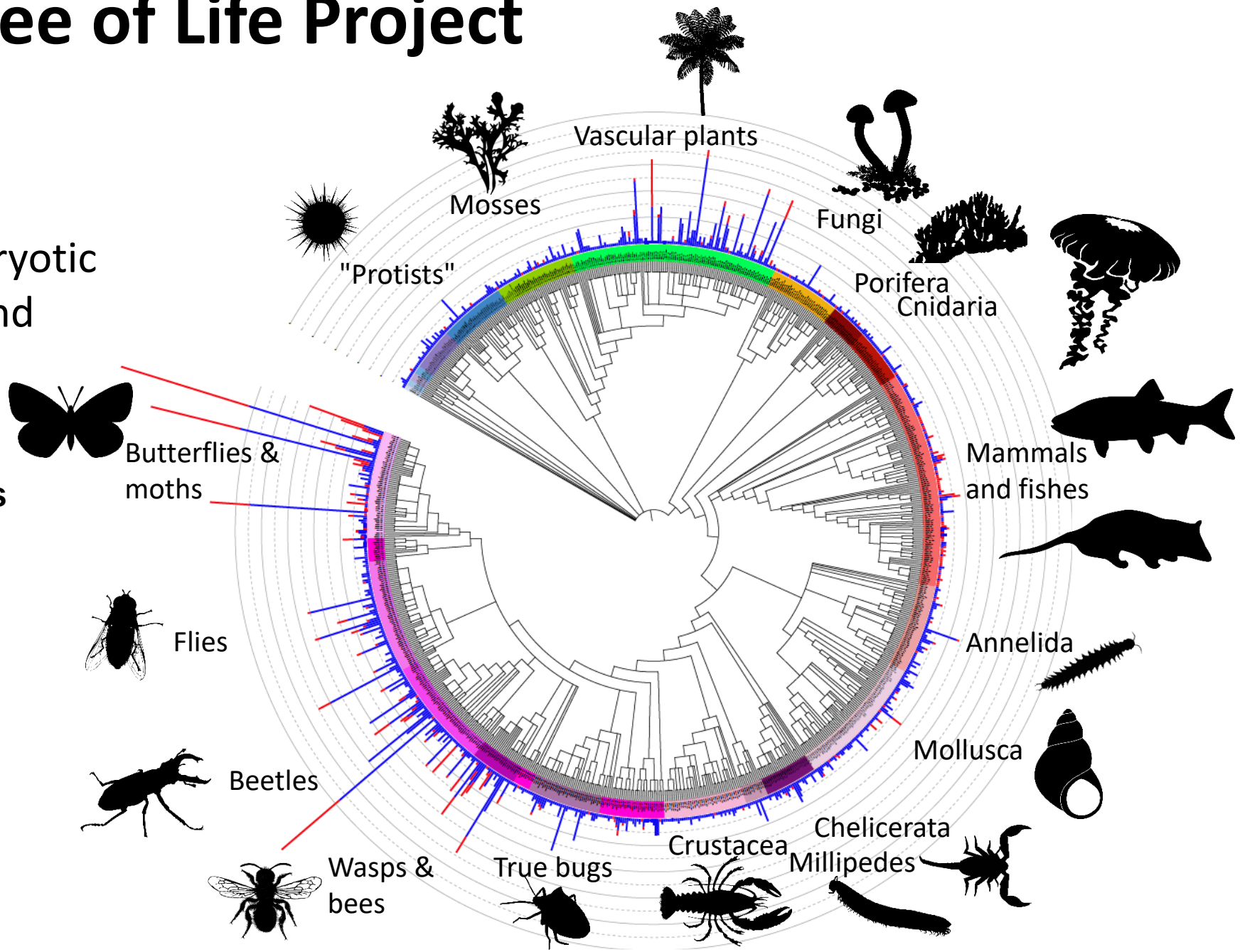
- ★ **European Reference Genome Atlas**
  - Sequencing the genomes of all species in the European continent (~500 species)

Darwin Tree of Life Project

sequence all ~70,000 eukaryotic species in Britain and Ireland

>5000 species in progress

>1500 genomes complete



wellcome

Project Psyche

Generate and explore reference genomes for all **~11,000 species** of **butterflies and moths (Lepidoptera)** of **Europe.**

Named after Psyche - Greek goddess of the soul, often depicted with butterfly wings

Credit: Sidon, Sarcophagus relief of Psyche. Livius.org

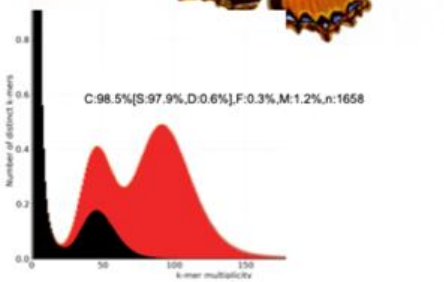
Tree of Life

wellcome sanger institute

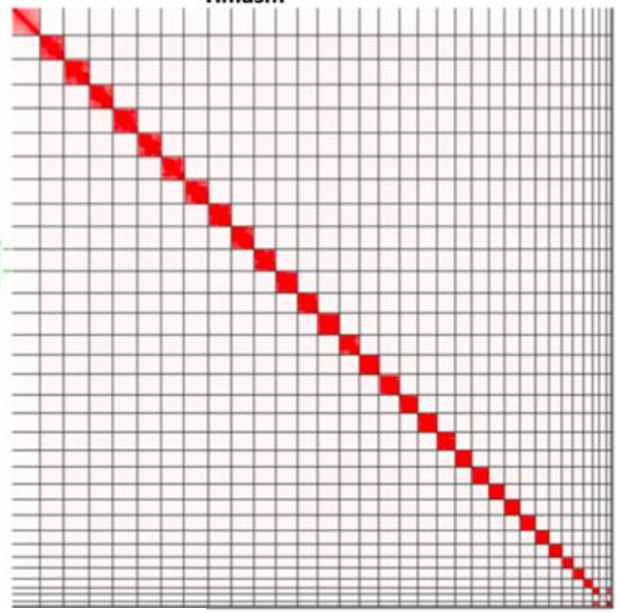
Why Lepidoptera?

Lepidoptera genomes tend to assemble very easily

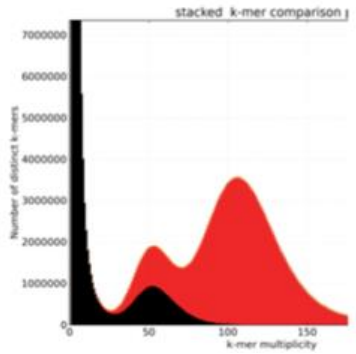
Nymphalis polychloros

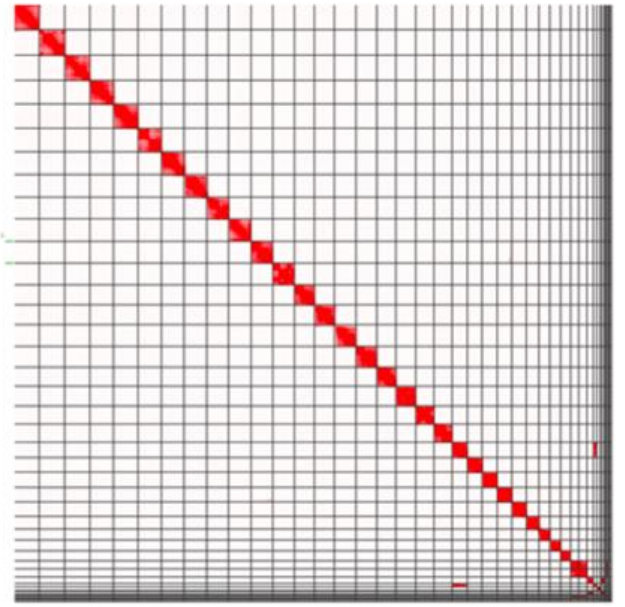


Hifiasm

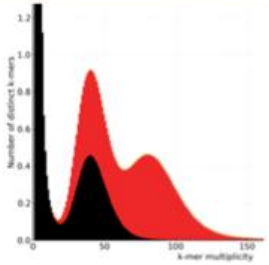


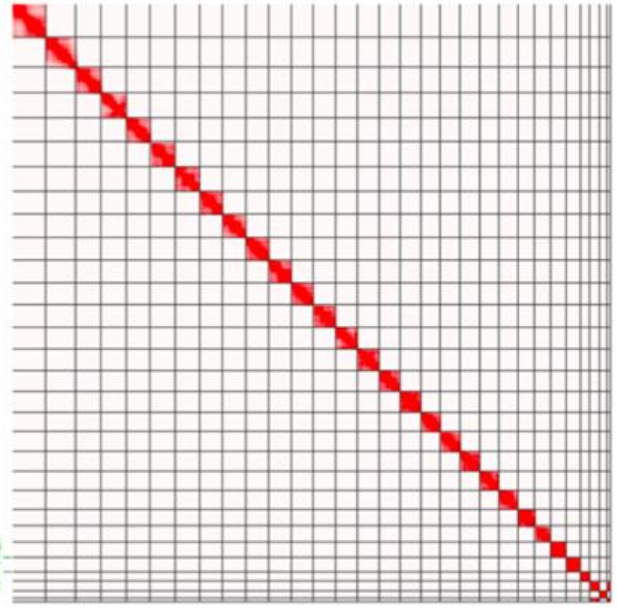
Pheosia tremula



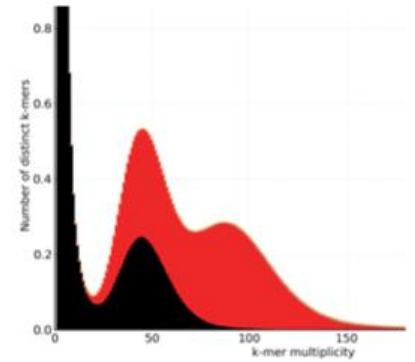


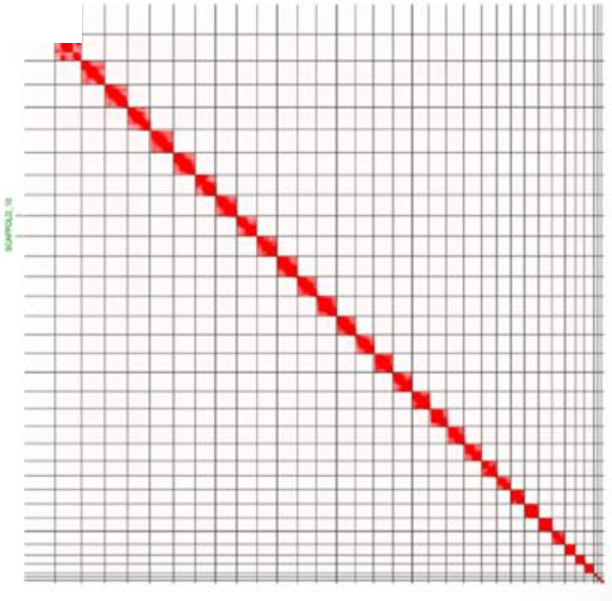
Deilephila porcellus





Colias croceus





Why sequence all species of Lepidoptera in Europe?

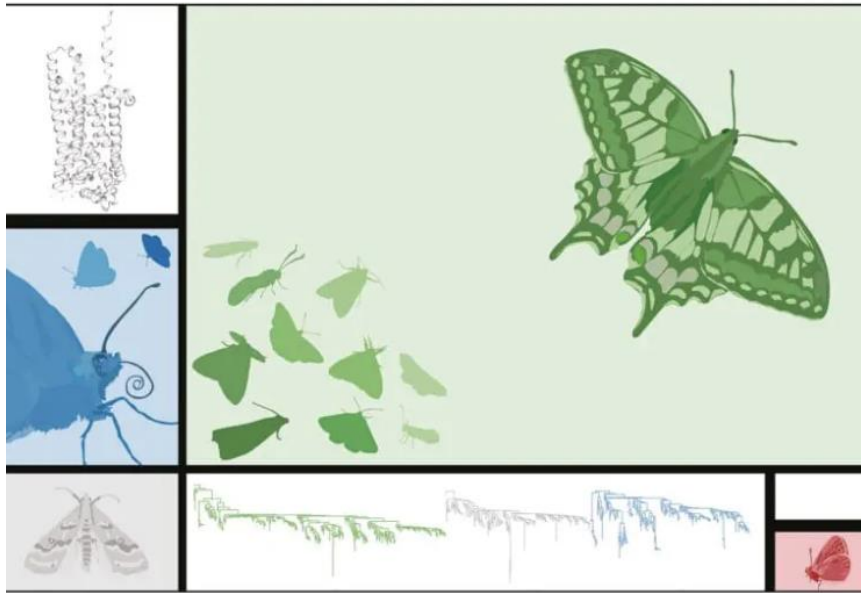
To answer wide-reaching questions:

- Chromosome evolution
- Evolution of opsin genes
- Genetic basis of migration
- Population genetics
- Genetic diversity predictors
- & many more…

Volume 40 • Number 12 • December 2023

MOLECULAR BIOLOGY AND EVOLUTION

academic.oup.com/mbe



Opsin Gene Duplication in Lepidoptera: Retrotransposition, Sex Linkage, and Gene Expression

Peter O Mulhair, Liam Crowley, Douglas H Boyes, Owen T Lewis, Peter W H Holland Author Notes

Molecular Biology and Evolution, Volume 40, Issue 11, November 2023,

nature ecology & evolution

Explore content About the journal Publish with us

nature > nature ecology & evolution > articles > article

Article | Open access | Published: 21 February 2024

Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera

Charlotte J. Wright, Lewis Stevens, Alexander Mackintosh, Mara Lawniczak & Mark Blaxter

Nature Ecology & Evolution **8**, 777–790 (2024) | Cite this article

16k Accesses | **13** Citations | **255** Altmetric | Metrics

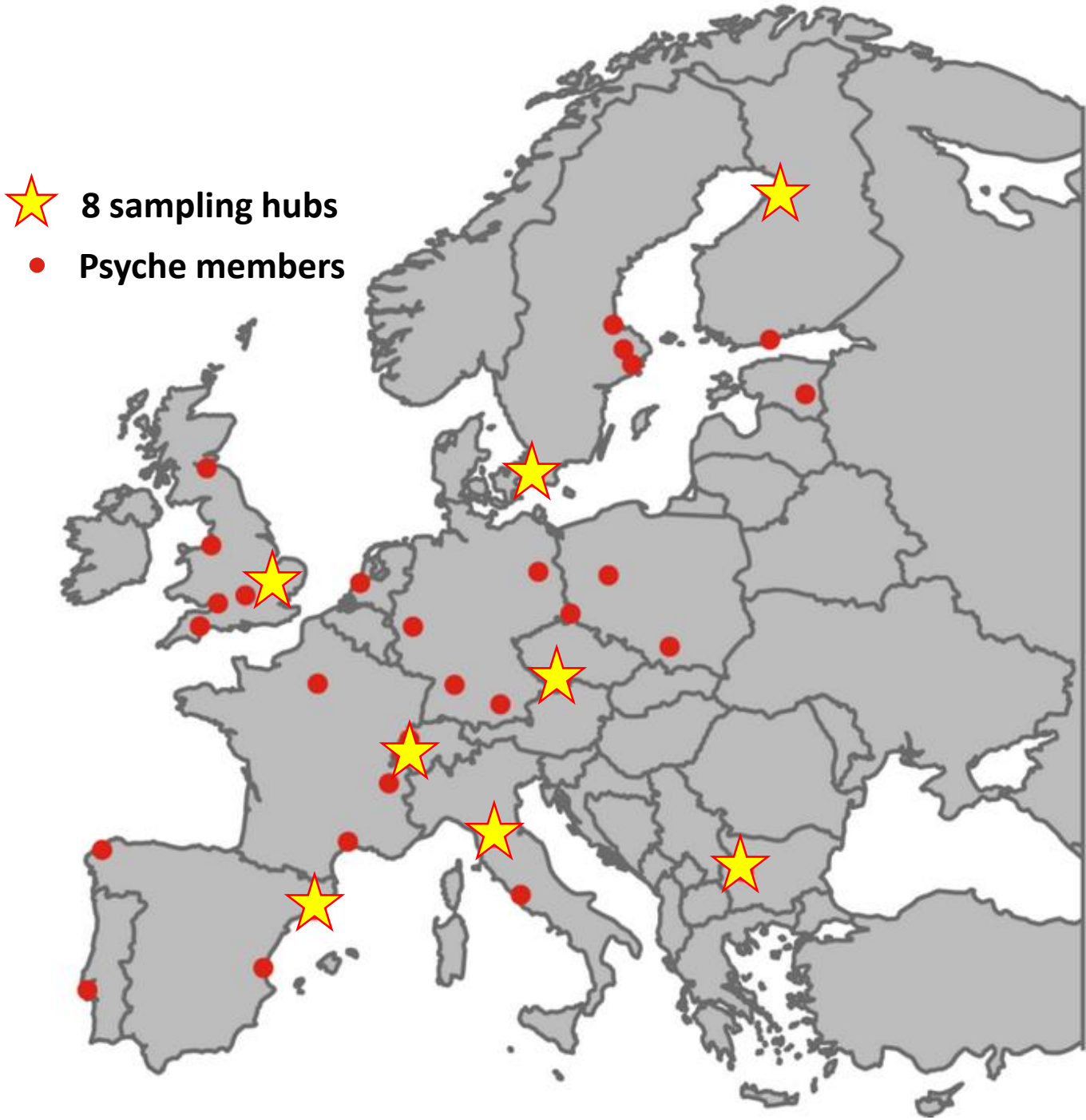
Psyche community

Charlotte Wright

Mark Blaxter

Joana Meier

Join us if you are interested in butterflies and moths



Phases of Project Psyche

Generate and explore reference genomes for all **~11,000 species** of **butterflies and moths (Lepidoptera)** of **Europe.**

Phase 1: Generate 2,000 genomes

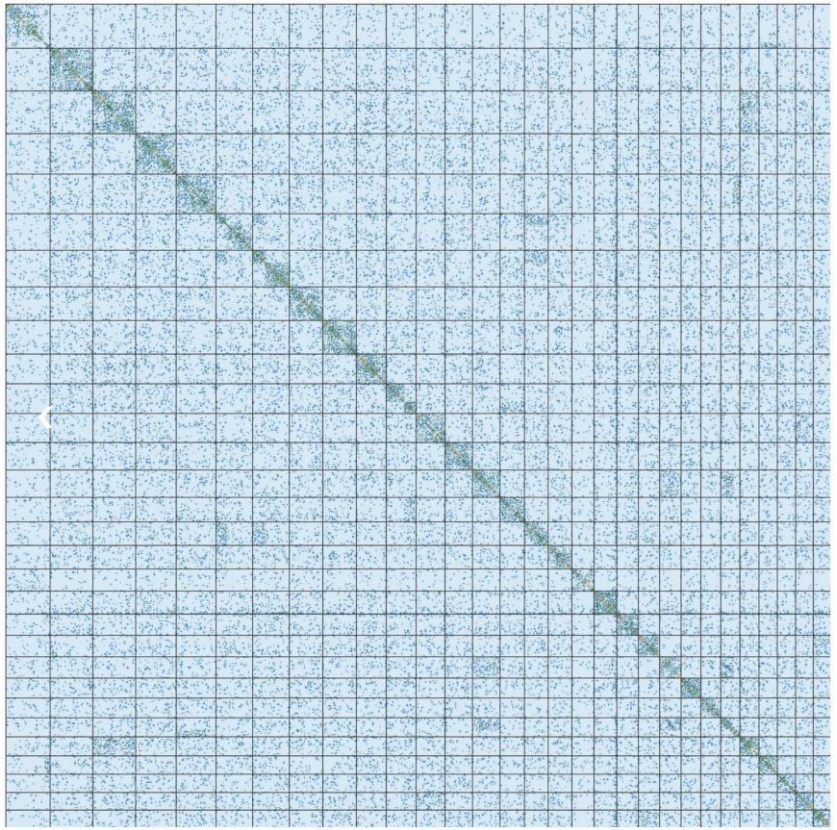
- Focus on taxonomic diversity
- Samples from sampling hubs
- Sequencing at the Wellcome Sanger Institute
- Writing a whitepaper to celebrate the first 1000 genomes
- Writing a set of 3-5 publications on the first 1000 genomes

Plans for later phases:

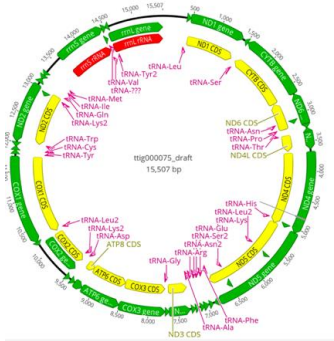
- Establishing more sampling hubs
- Establishing more sequencing hubs

Genome assembly of the moth and all its co-bionts

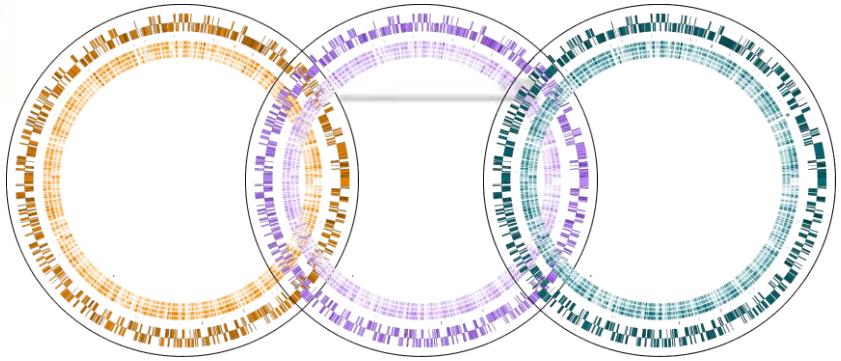
***Phalera bucephala*:**
one moth, five genomes



31 nuclear chromosomes



mitochondrial genome



Wolbachia A, B1, B2

Openly published genomes and Genome Notes: Collector(s) as first author(s)

https://www.darwintreeoflife.org/genomes/genome-notes

Wellcome Open Research

Wellcome Open Research 2023, 8:278 Last updated: 18 SEP 2023

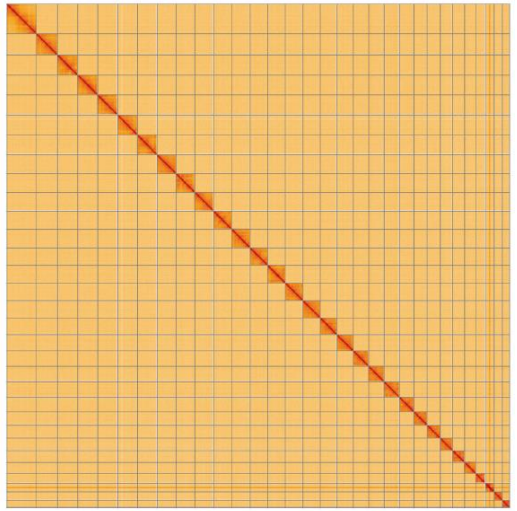
Check for updates

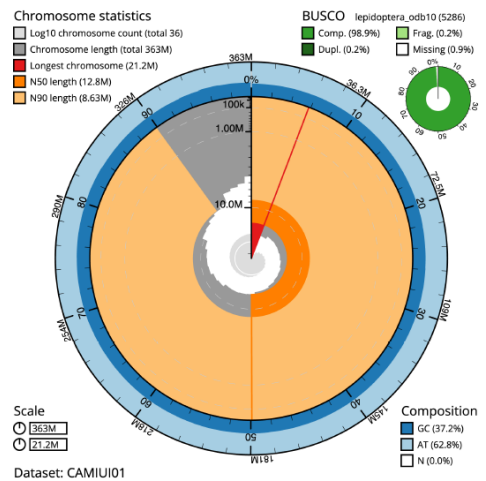
DATA NOTE

The genome sequence of the Dark Spectacle, *Abrostola triplasia* (Linnaeus, 1758) [version 1; peer review: 2 approved]

Douglas Boyes[1+], Owen T. Lewis[2],
University of Oxford and Wytham Woods Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]UK Centre for Ecology & Hydrology, Wallingford, England, UK
[2]University of Oxford, Oxford, England, UK





Psyche enables LepEU: population genomics across Lepidoptera

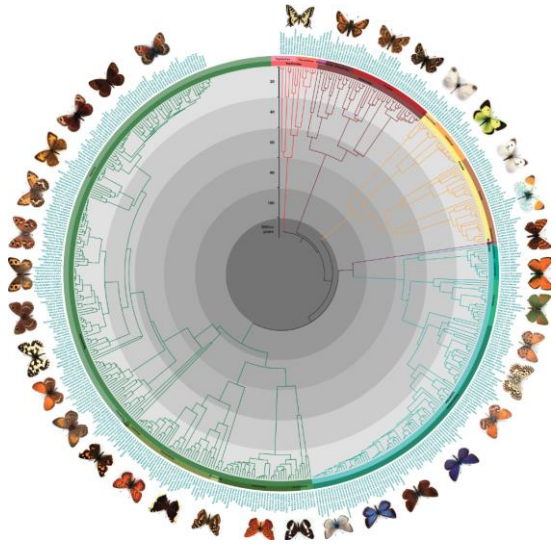
Project Psyche
Reference genomes of
all European Lepidoptera

LepEU: European Lepidopteran Population Genomics Consortia

https://lepeu.github.io/

Quantifying regional genetic diversity, inbreeding, and local adaptation of leps to inform biodiversity management in Europe

Patrícia Beldade
Lisbon University, Portugal



Chris Wheat
Stockholm University, Sweden

COST action:
Lep10KGenomes: Utilizing 10,000 genomes of European Lepidoptera

- Organise workshops and training courses
- Research exchanges for early career researchers
- Work with external stakeholders to address societal needs using genomic information
- Devise best practices
  - For reference genomes - **Psyche**
  - For population genomics - **LepEU**

Niklas Wahlberg
Lund University,
Sweden

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme
of the European Union

Find out more here:
https://tinyurl.com/Lep10k-COST

Join Project Psyche here:
https://tinyurl.com/projectpsyche

How to generate a reference genome?

Check if a reference genome exists

Genomes **o**n **a** **T**ree [(GoaT, https://goat.genomehubs.org/)](https://goat.genomehubs.org/)

GoaT - latest

Genomes on a Tree (GoaT)

Genomes on a Tree (GoaT): A versatile, scalable search engine for genomic and sequencing project metadata across the eukaryotic tree of life. Challis *et al.* 2023. Wellcome Open Res 2023, **8**:24
doi:10.12688/wellcomeopenres.18658.1

GoaT needs been built using GenomeHubs to help coordinate efforts across the Earth Biogenome Project (EBP) Network at all stages from planning through sequencing and assembly to publication. read more...

Search GoaT

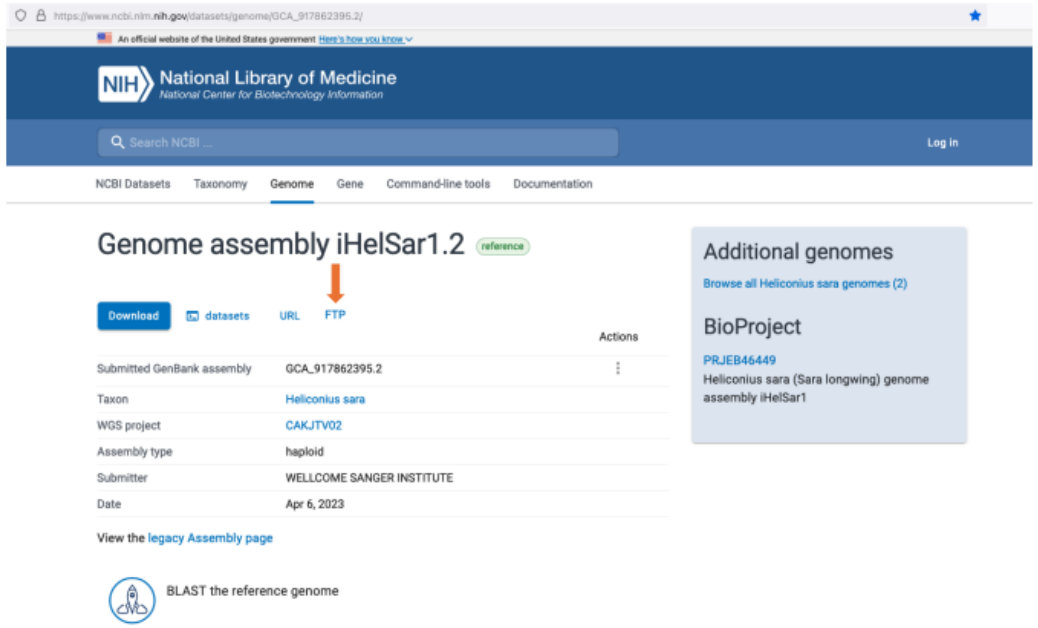
Type to search GoaT taxon index (e.g. Canidae)

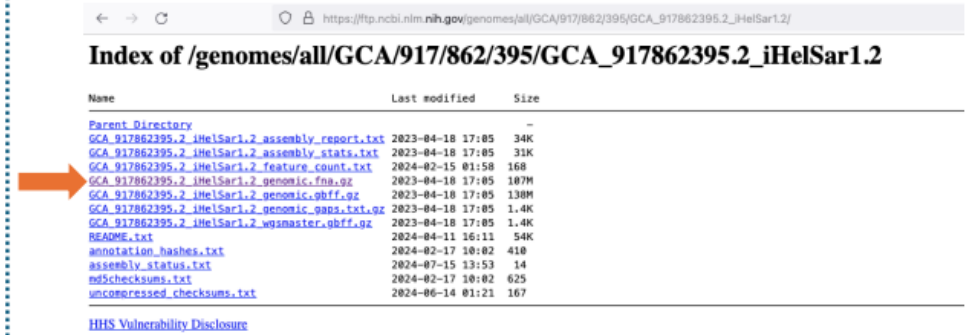
TAXON

include descendants: Off
include estimates: On
empty columns: Off
result columns
query builder
clear all

First, check if a reference genome already exists

- Genomes on a Tree [(GoaT, https://goat.genomehubs.org/)](https://goat.genomehubs.org/)
- If yes, get it e.g. from NCBI (www.ncbi.nlm.nih.gov/home/genomes)



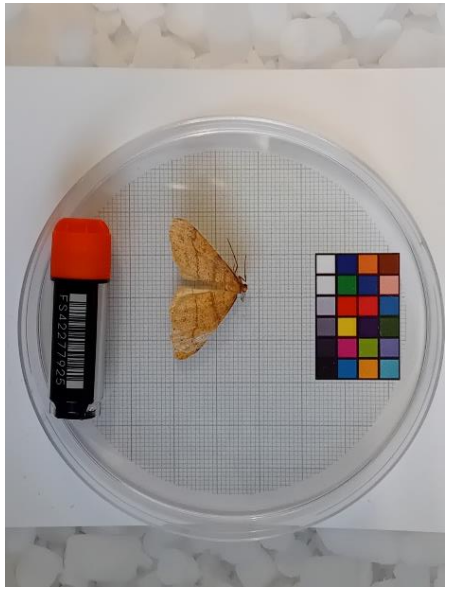


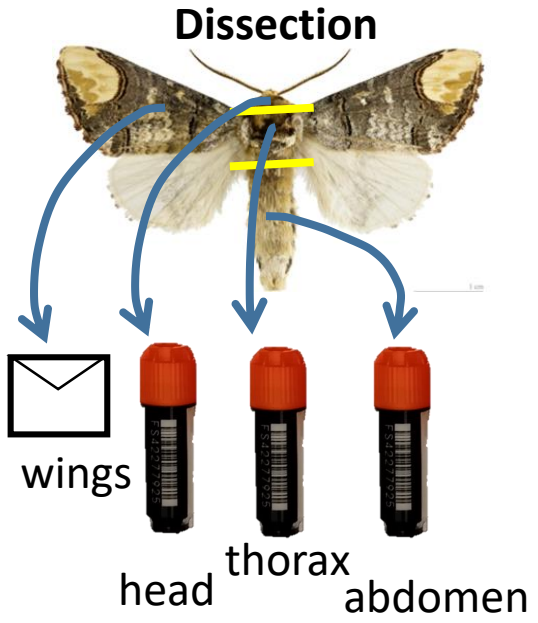

```
wget
https://ftp.ncbi.nlm.nih.gov/genomes/all/GC
A/917/862/395/GCA_917862395.2_iHelSar1.2/GC
A_917862395.2_iHelSar1.2_genomic.fna.gz
```

1. Getting specimens with well-preserved DNA

Dry shipper (-150°C)
(with liquid nitrogen in the walls)

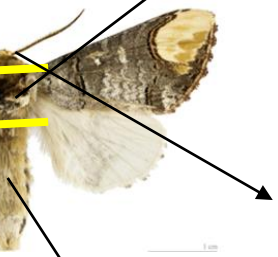
- Ideally:
  - Fresh tissue (or blood for birds or fishes)
  - Flash-frozen (in liquid nitrogen or cryogenic dry shipper) and kept at max -70°C
- ensure you have all metadata for the specimen (GPS coordinates, photo of the specimen, etc)





2. Sequencing the specimen

For Project Psyche
we use these tissues:



thorax

PacBio sequencing (long reads) to assemble contigs

ATGTGTCATGGGACATATGTGTCATGGGACATGAGAGAGAGA ← contig = consensus sequence of aligned reads

GAGAGAGAATGTGTCATGGGACAT
GGACATATGTGTCATGGGACATGA
TATGTGTCATGGGACATGAGAGAGAGA

head

HiC sequencing to scaffold contigs together

GTGTCATG ATGTGTCATG ATGAGAGAG GAGAGGGA
ATGTGTCATGGGACATATGTGTCATGGGACATGAGAGAGAGAGAGAGA GAGAGAGGGACATAAGAGTGTGTCATG

ATGTGTCATGGGACATATGTGTCATGGGACATGAGAGAGAGAGAGAGANNNNGAGAGAGGGACATAAGAGTGTGTCATG ← scaffold = sequence with gaps

abdomen

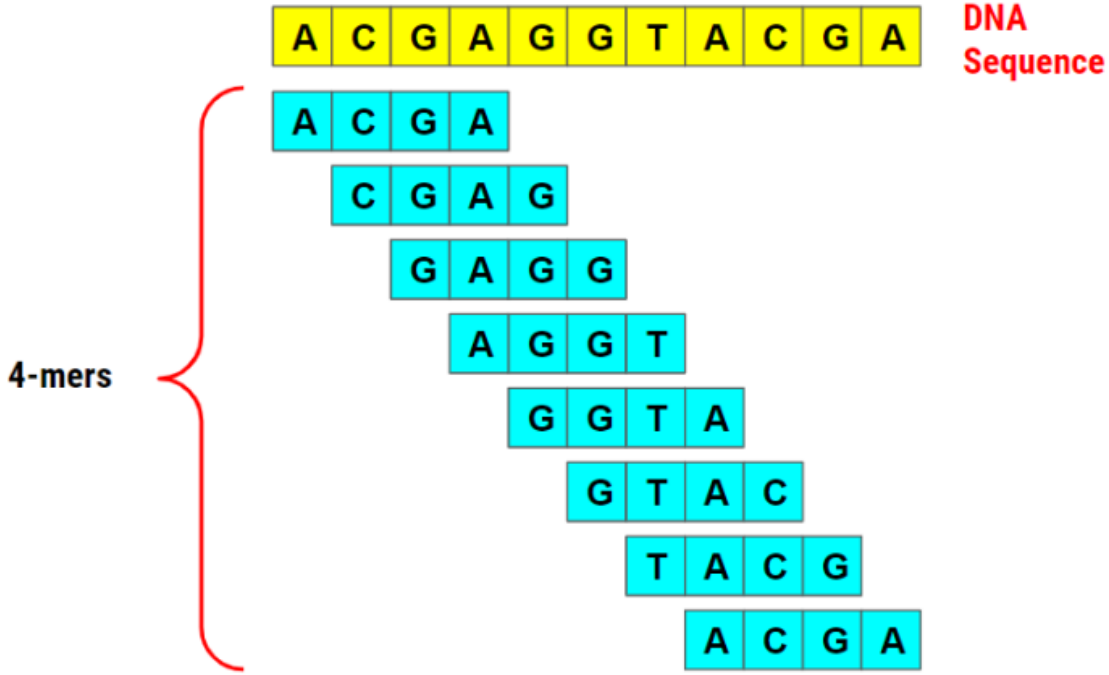
RNA sequencing for genome annotation

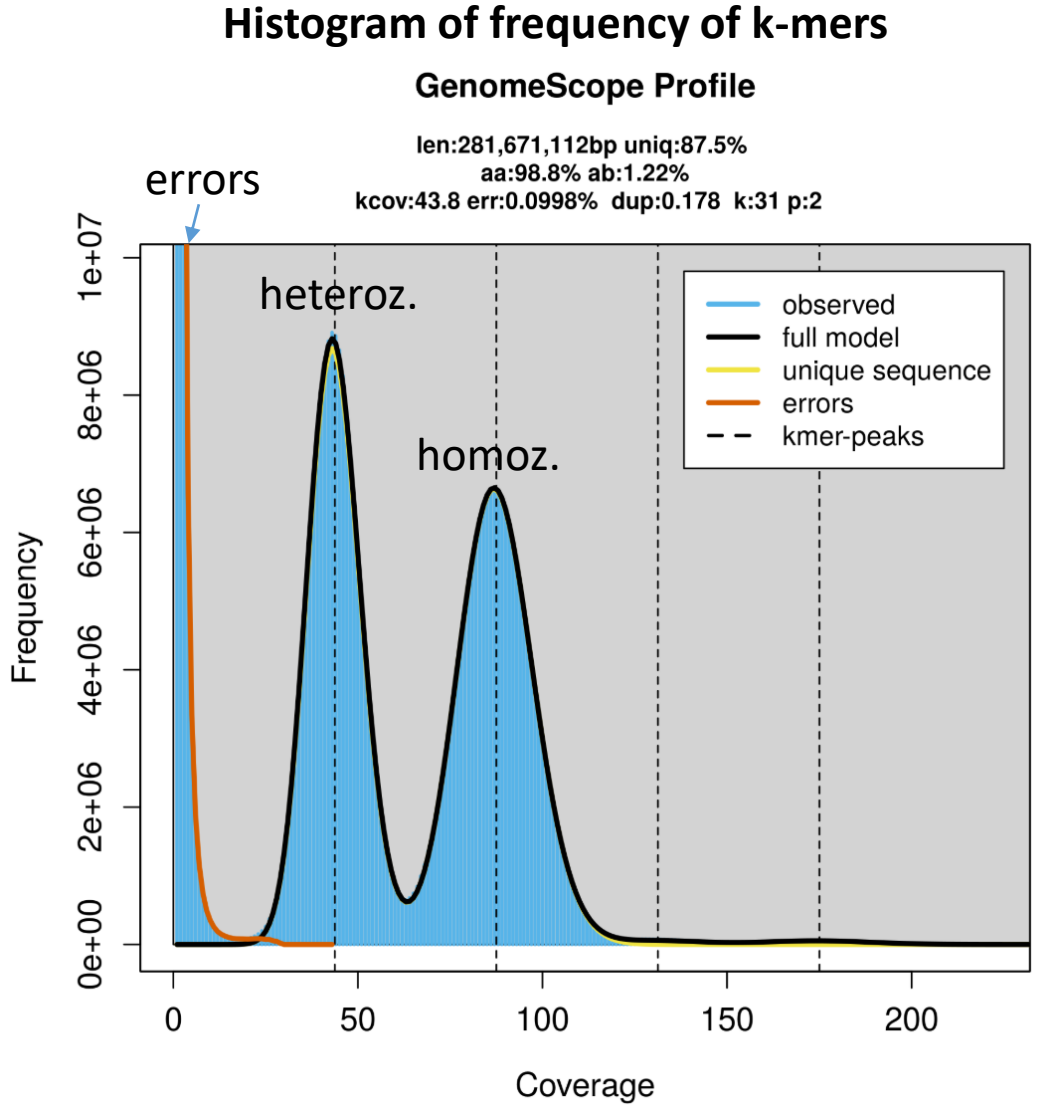
GTGAGTCATGGGAGAGGACATCAAAAAAAAAAAAAAA
AGAGGTCATGTGAGTACATCGATCGGAAAAAAAAAAAAAAA

For T2T (telomere-to-telomere) genomes, add ONT ultra-long reads (>150 kb)

3. First statistics with kmer analyses

Estimating the genome size, depth of coverage, ploidy, levels of heterozygosity & duplication, identifying contamination





Genomescope

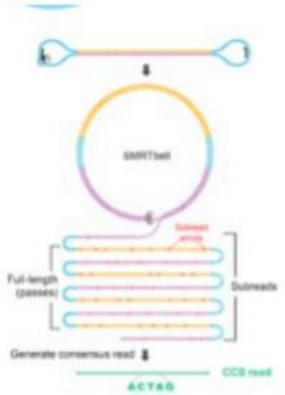
4. Assembling the genome

DToL Current Pipeline

- Sequencing technologies: PacBio HiFi + HiC (Arima or Qiagen)

MitoHiFi — *For mitochondria genome assembly*

Slide from Marcela Uliano-Silva



Assembly

Haplotype separation

Scaffolding

Hicanu or Hifiasm

Purge dups

Prim

Hap

Yahs scaffolding (Arima or Qiagen HiC)

Curated assembly

1- Kmer Jellyfish/ GenomeScope, asmstats, smudgeplot

2 - asmstats, BUSCO, merqury

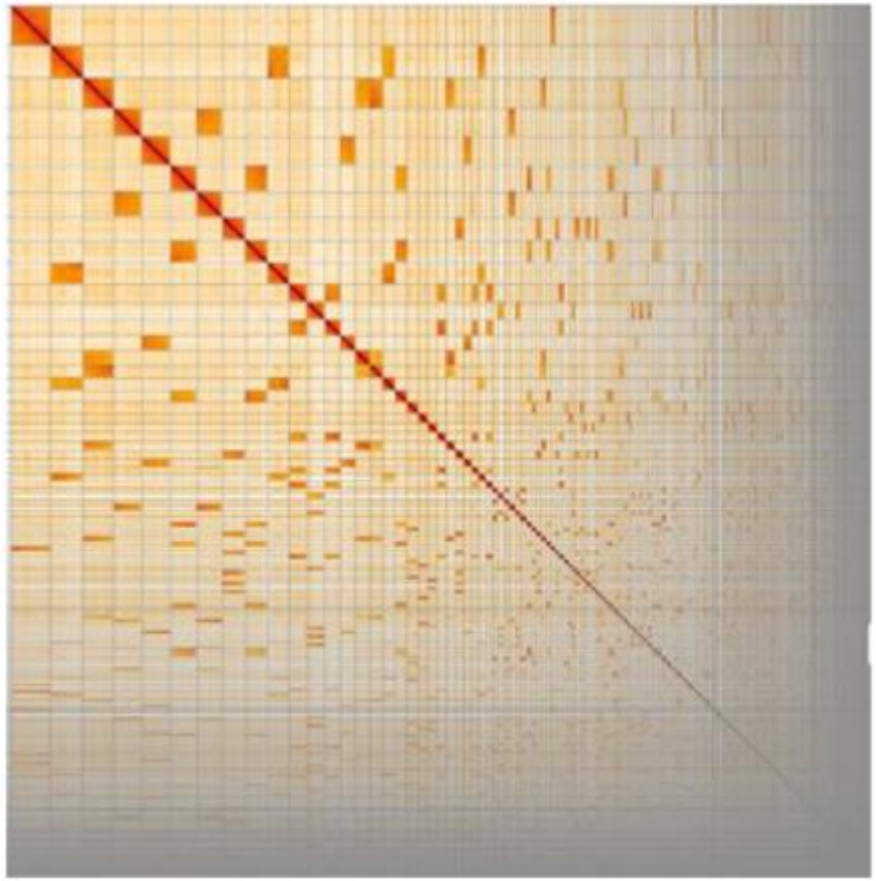
3 - asmstats, BUSCO, merqury

4 - asmstats, BUSCO, merqury, HiC heatmap

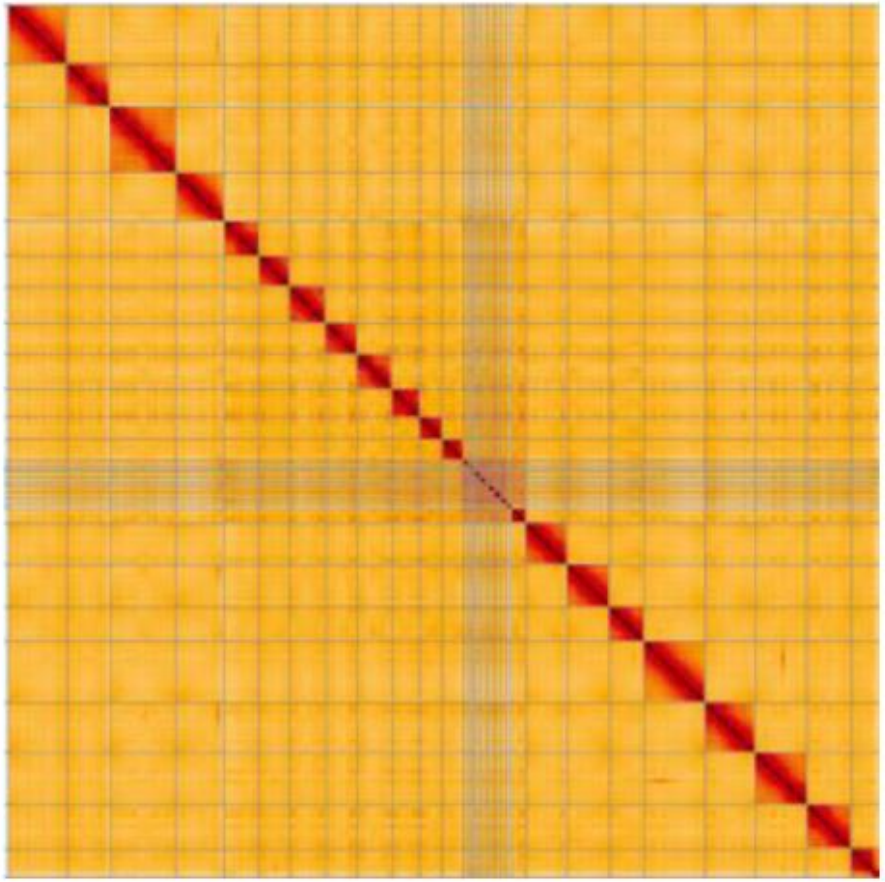
5 - asmstats, BUSCO, merqury, HiC heatmap

Manual curation to correct errors by assemblers

HiC map before manual curation



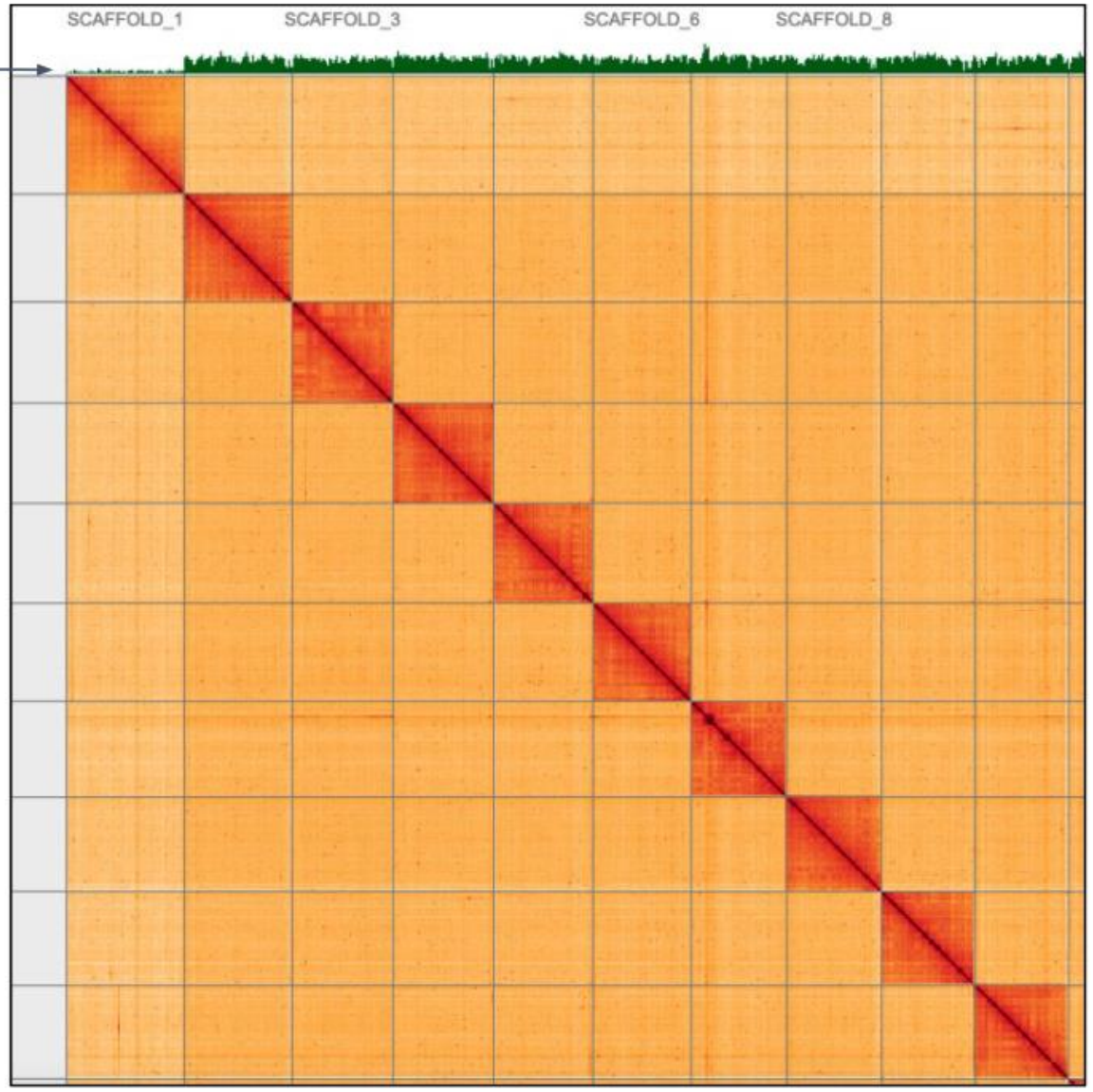
After manual curation with PretextView

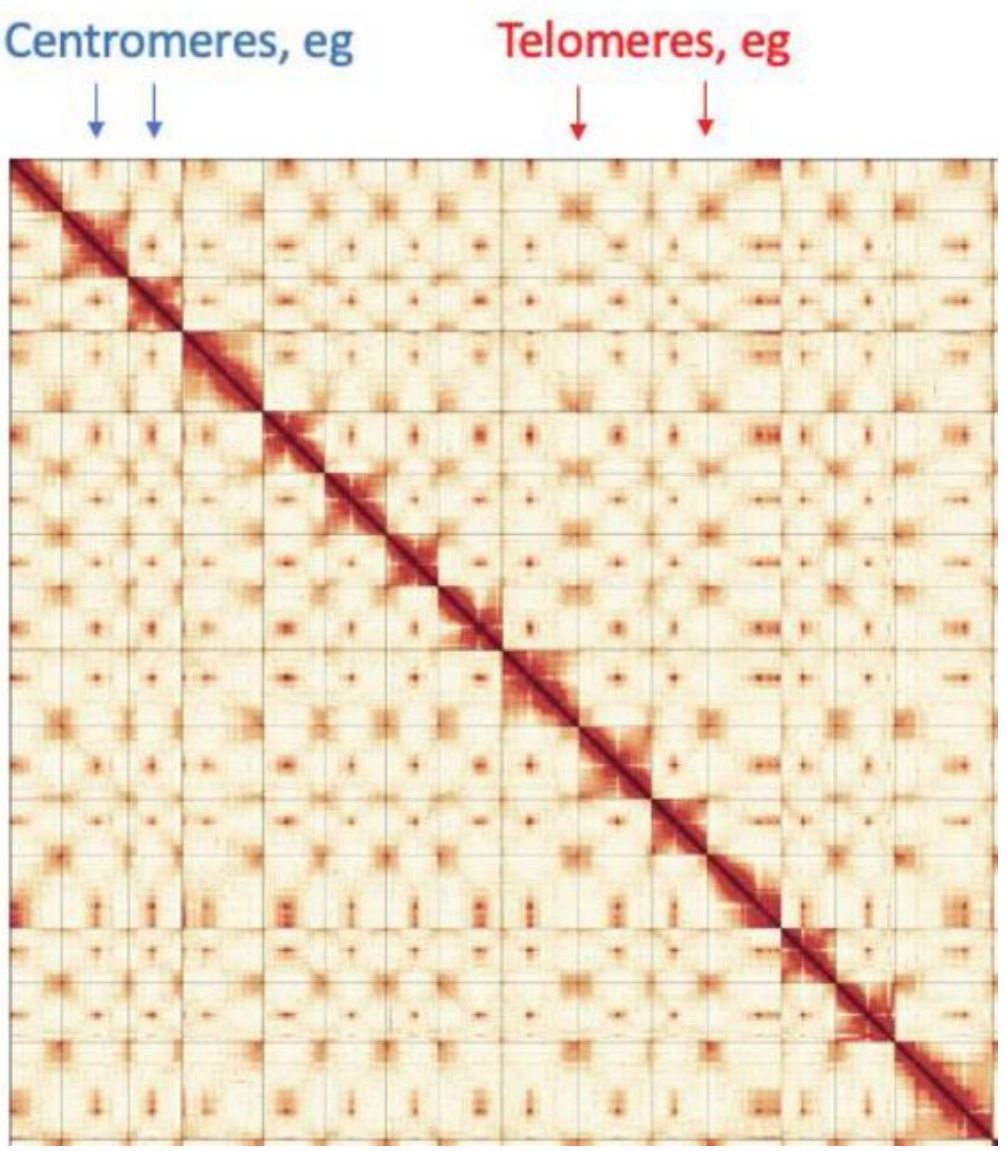


https://gitlab.com/wtsi-grit/rapid-curation/-/blob/main/Interpreting_HiC_Maps_guide.pdf

HiC maps will also show interesting features

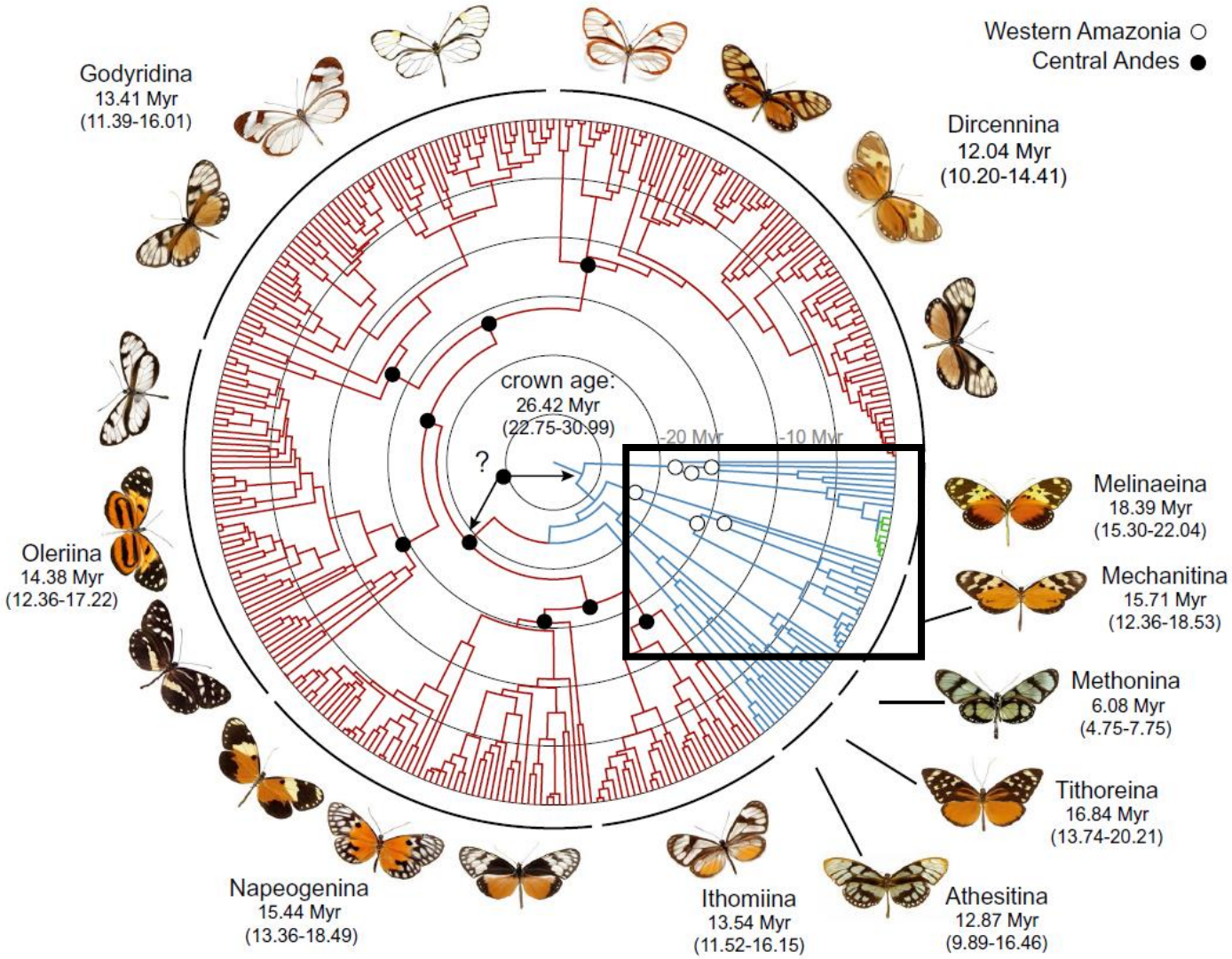
Identifying sex Chromosomes (half coverage)



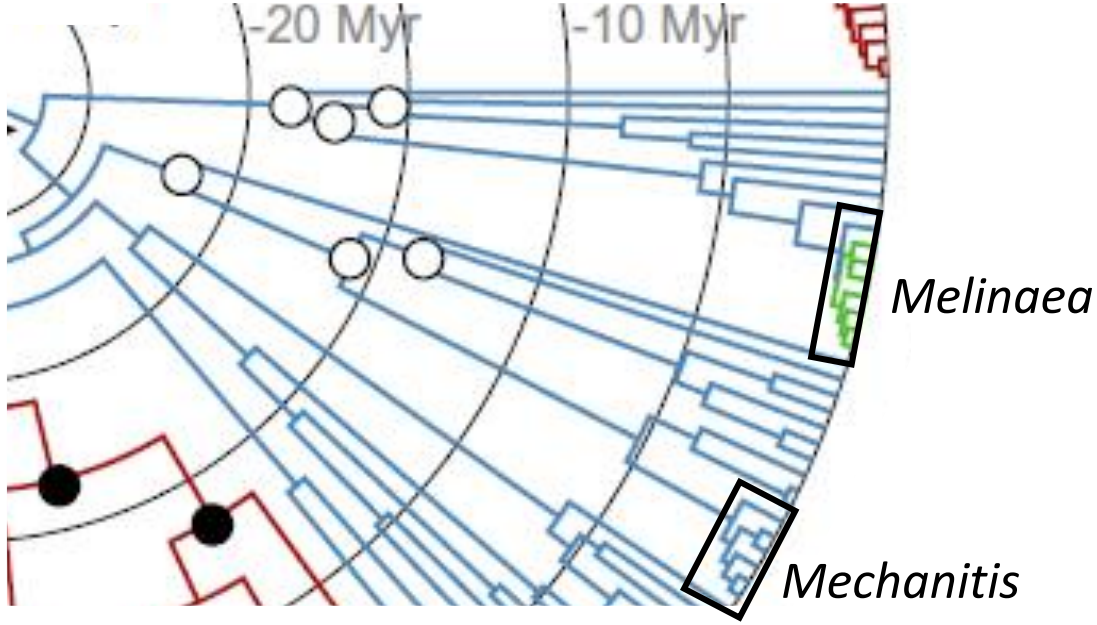


An example of using reference genomes from my own work

Comparing rapidly and slowly speciating Ithomiini genera



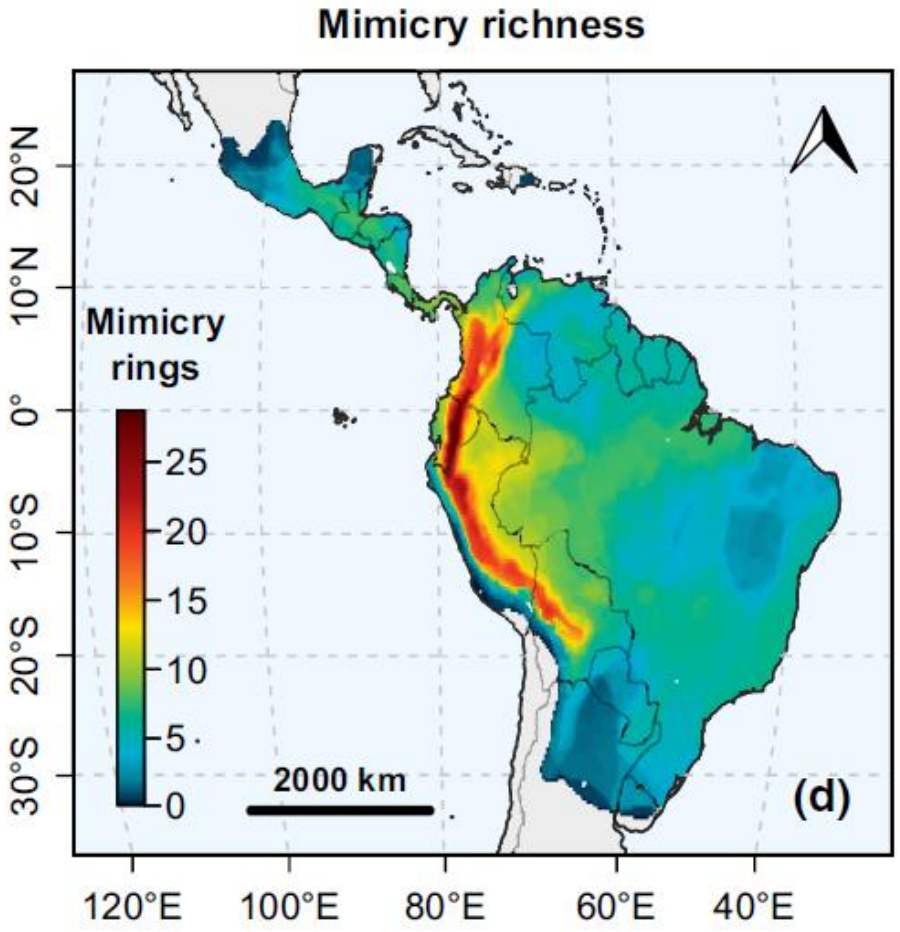
Time since last common ancestor



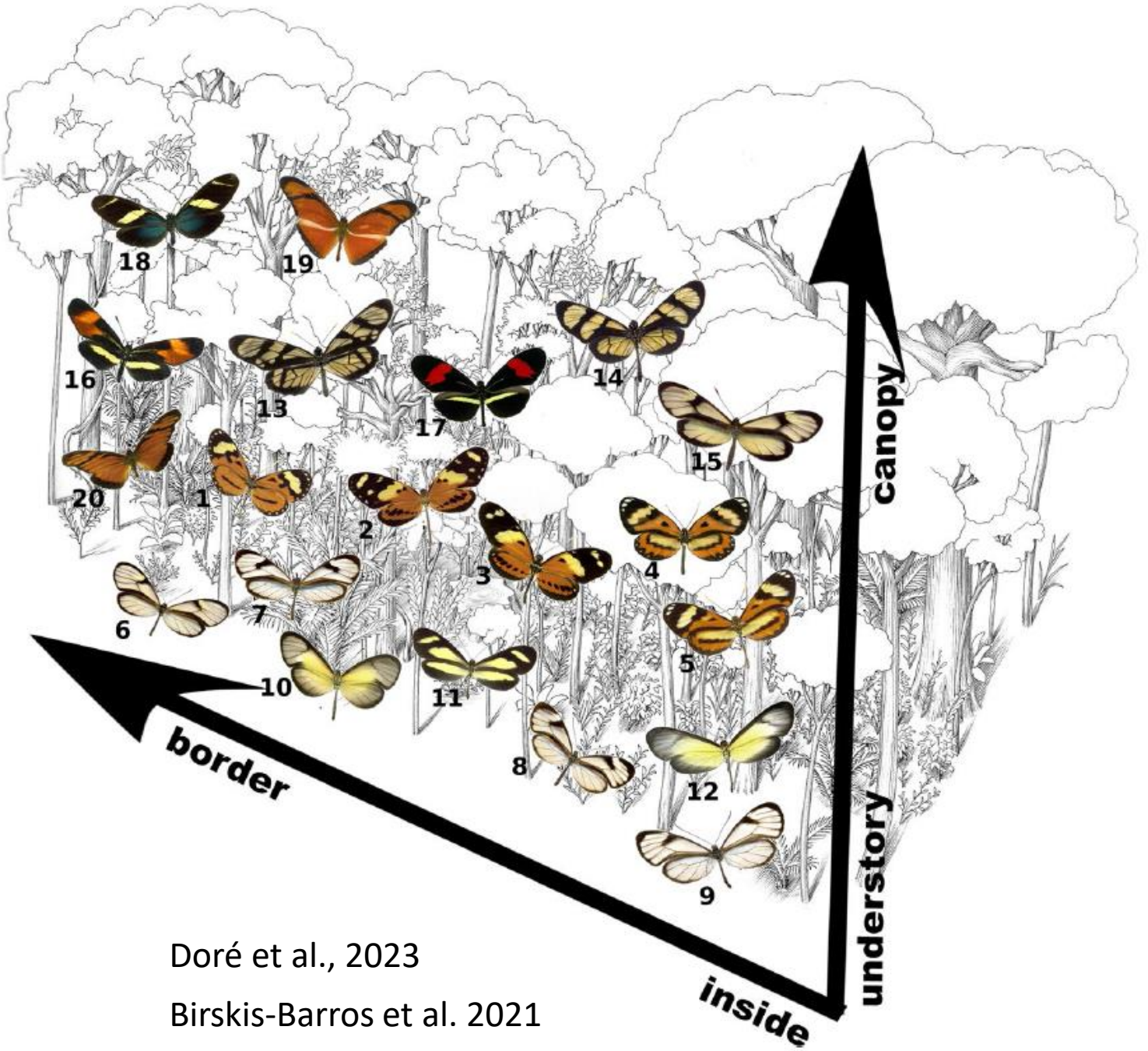
Chazot et al., 2019

Müllerian mimicry rings of ithomiini and others

Co-mimics converge in microhabitat use and associated traits



Doré et al., 2021

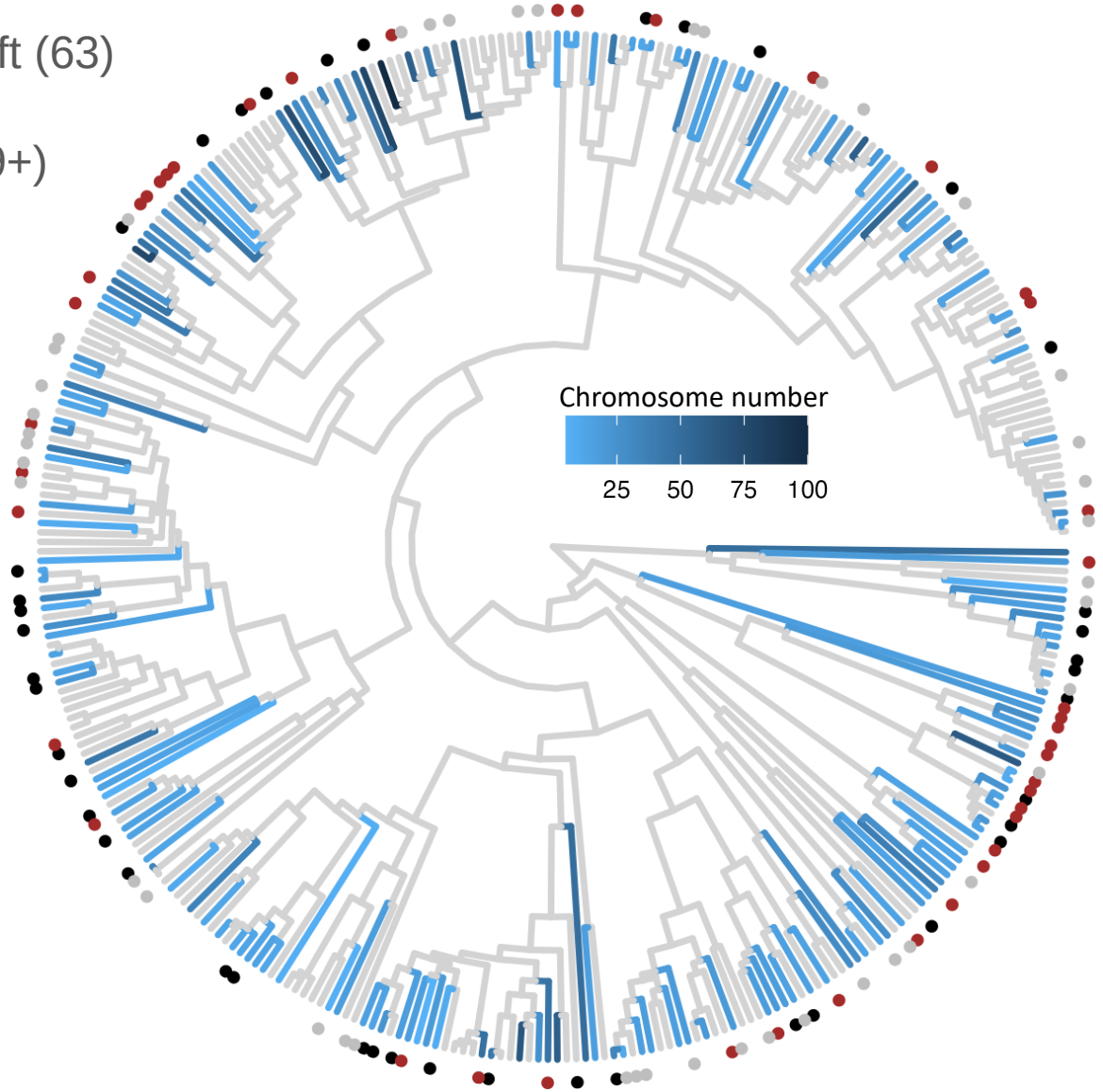


Doré et al., 2023

Birskis-Barros et al. 2021

Sequencing 200 ithomiini genomes

- Curated/Draft (63)
- Lab (47)
- Sampled (49+)



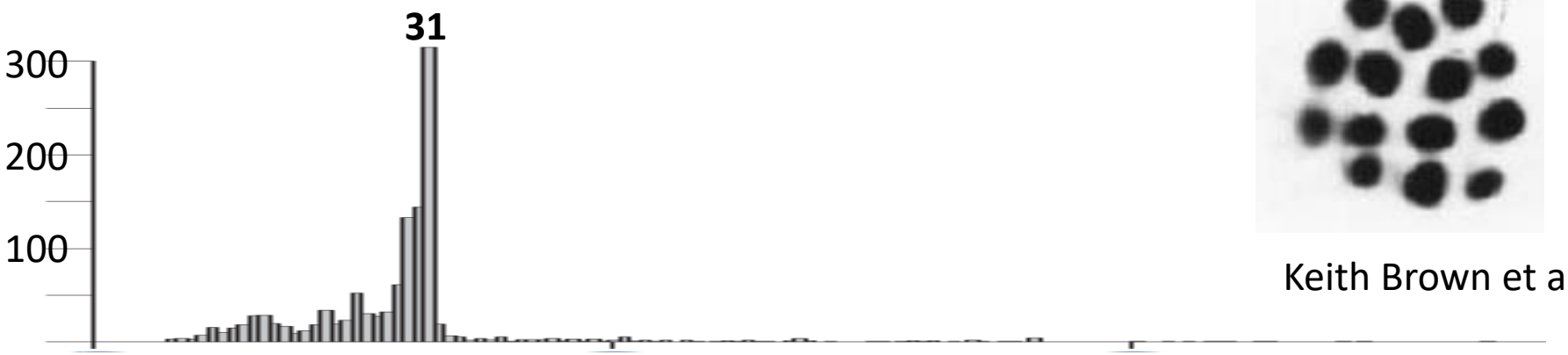
Patricio Salazar

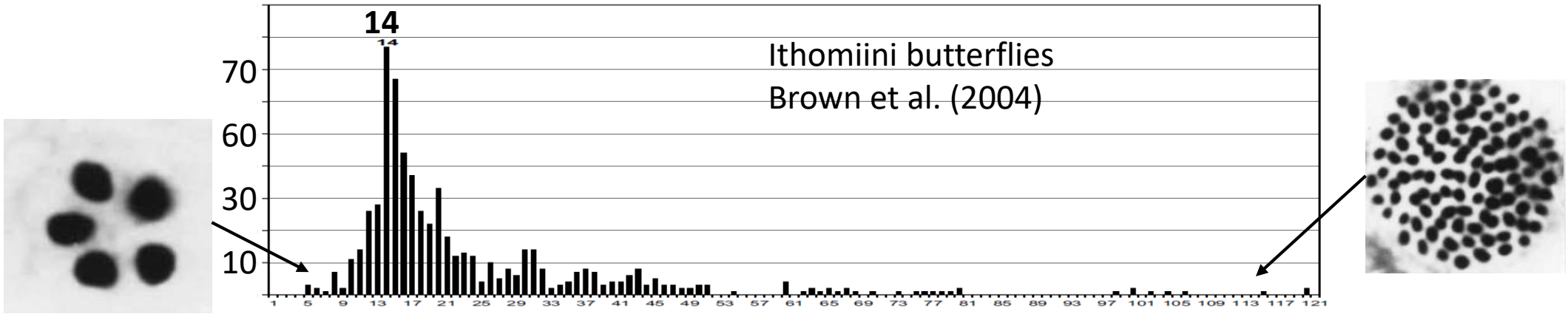
Karin Näsvall

Do chromosomal rearrangements contribute to rapid diversification?

most butterflies have 30-31 chromosomes

De Vos et al. (2020): 2400 Lepidoptera (including 171 ithomiini)

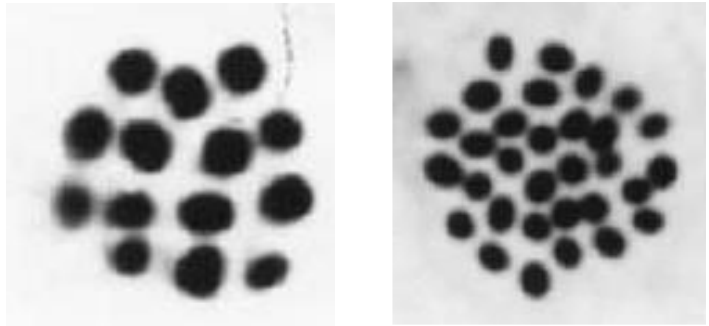




Ithomiini butterflies
Brown et al. (2004)

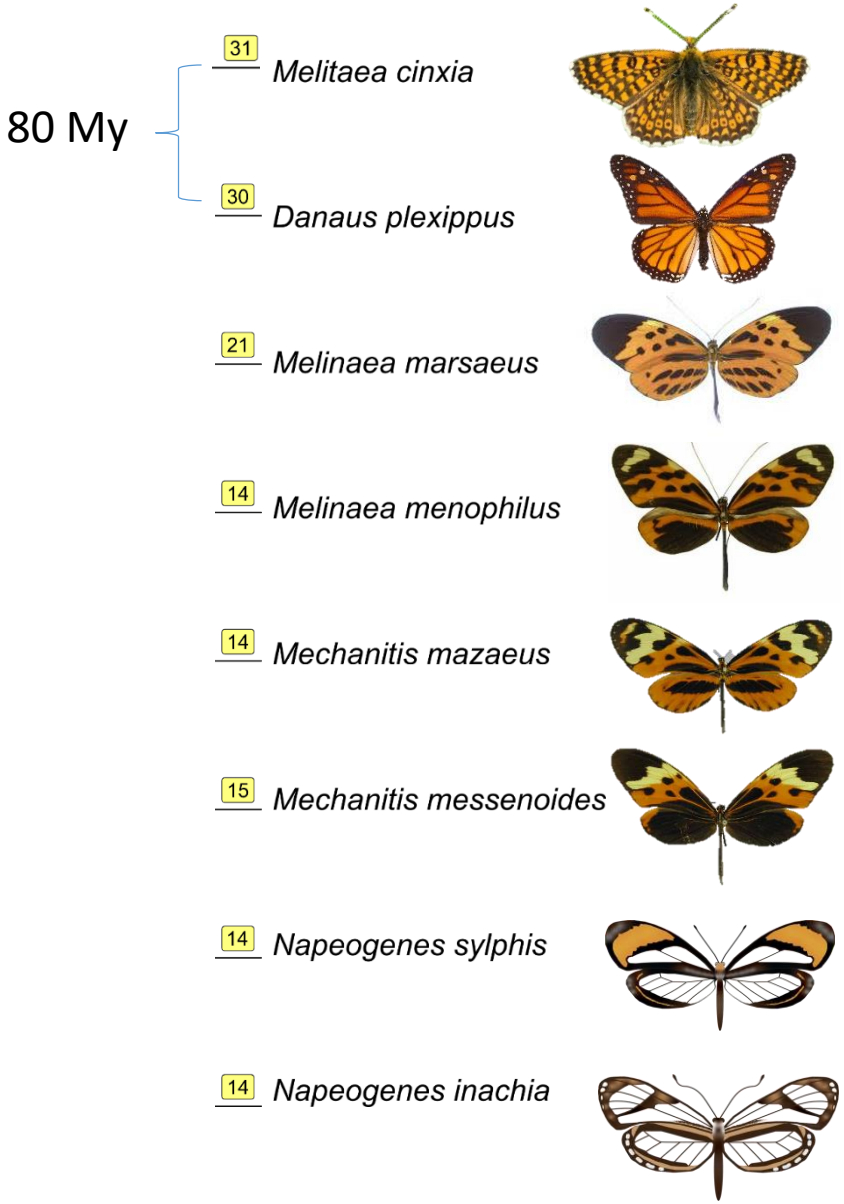
***Melinaea* species:**

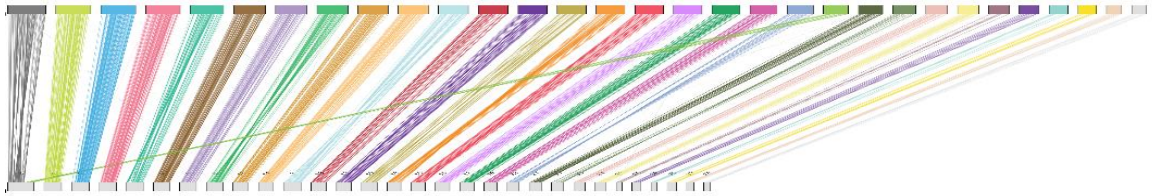
13-30 chromosomes



Keith Brown et al., 2004, Hereditas

Large-scale chromosomal rearrangements

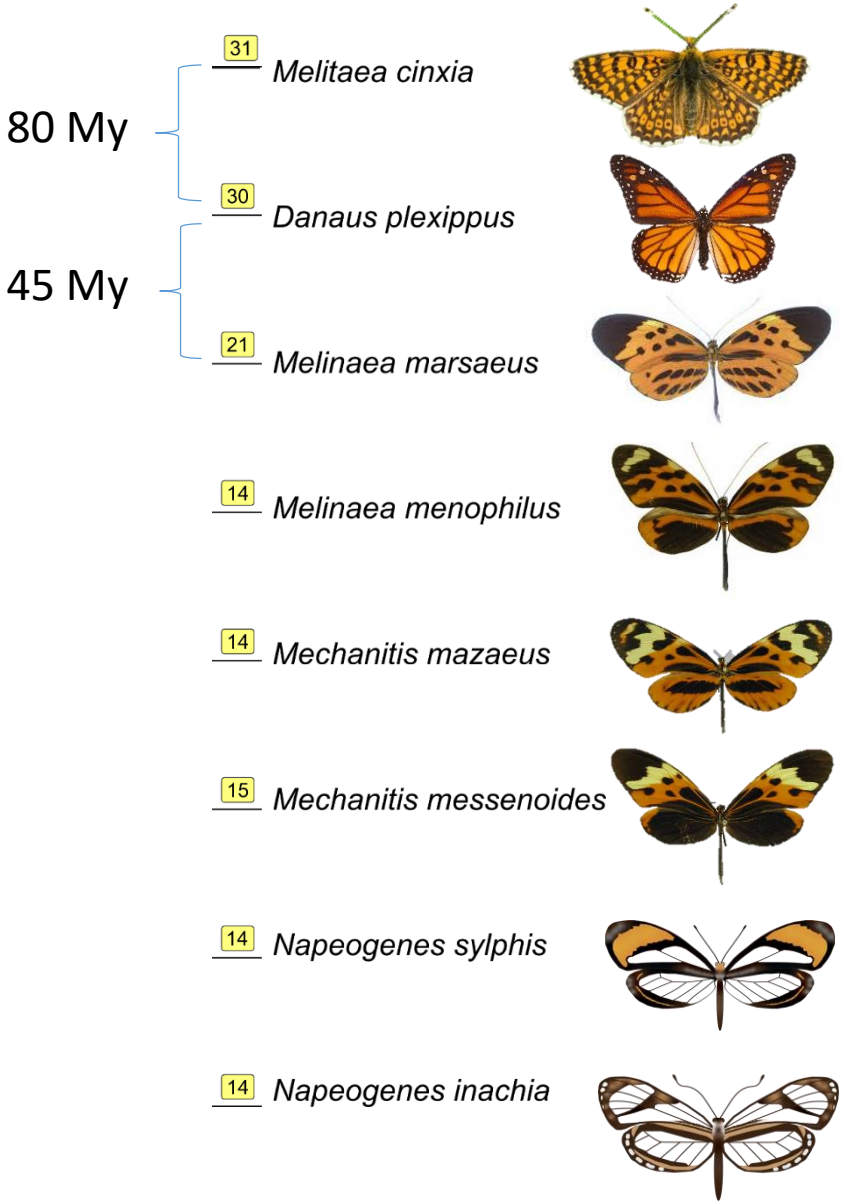


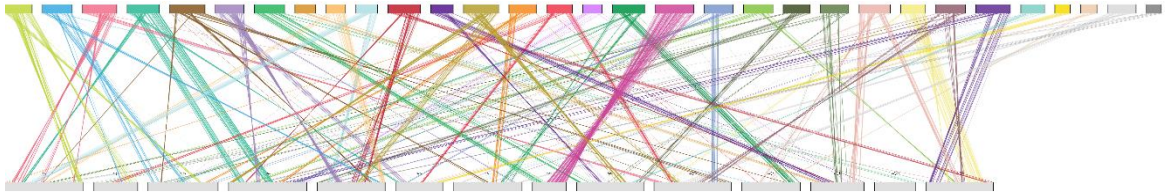


Karin Näsvall

Divergence times: Chazot et al. 2019

Large-scale chromosomal rearrangements

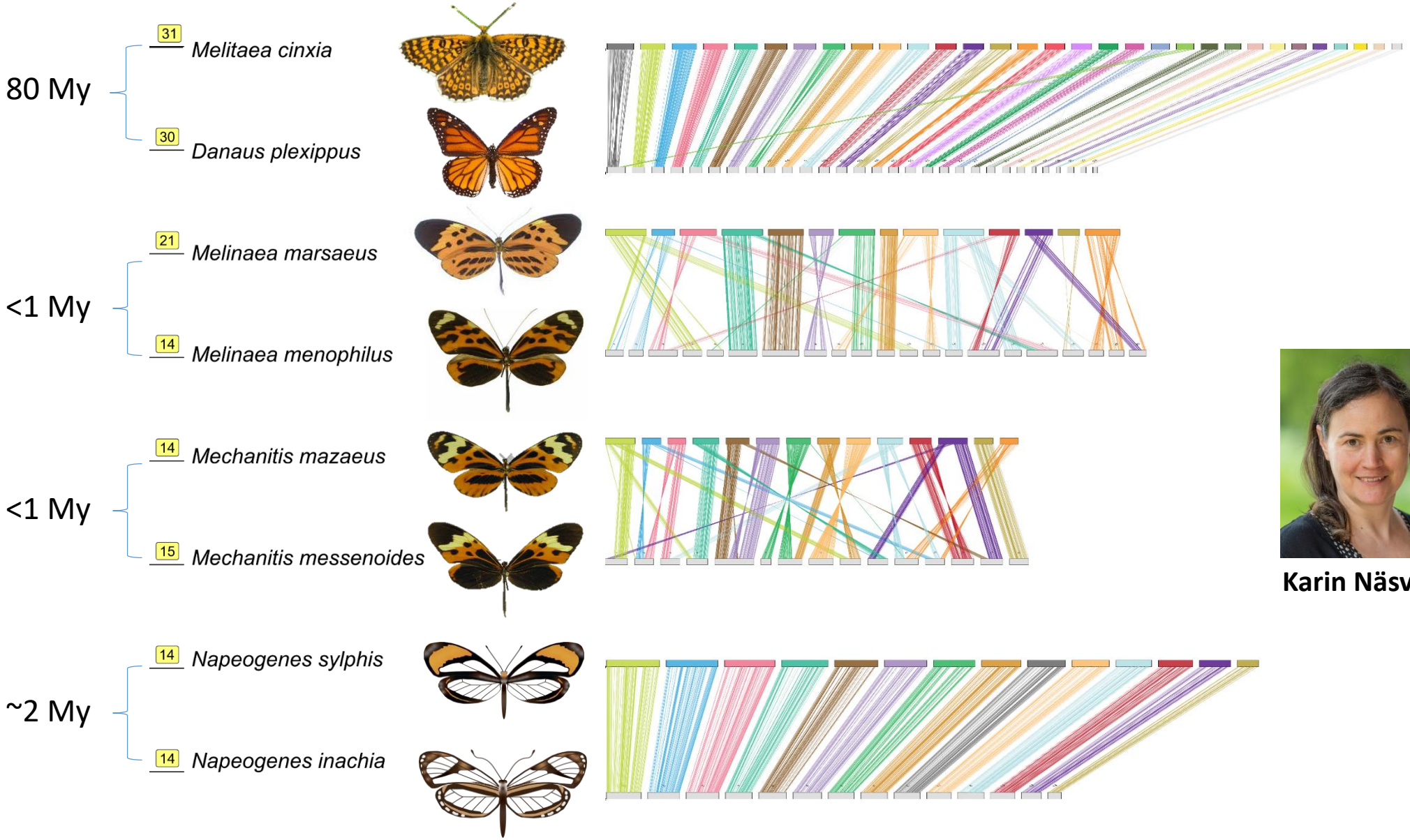




Karin Näsvall

Divergence times: Chazot et al. 2019

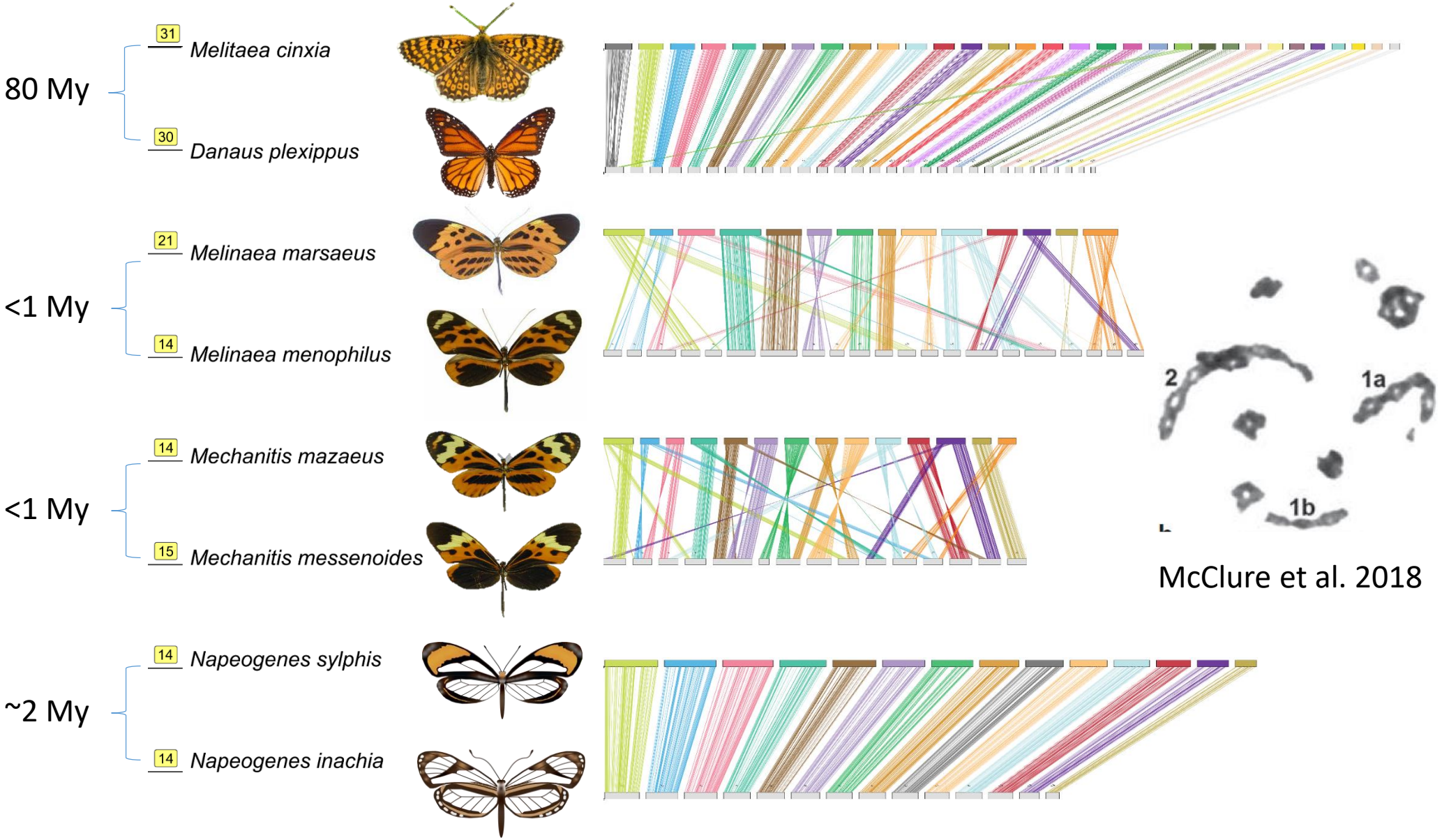
Large-scale chromosomal rearrangements



Karin Näsvall

Divergence times: Chazot et al. 2019

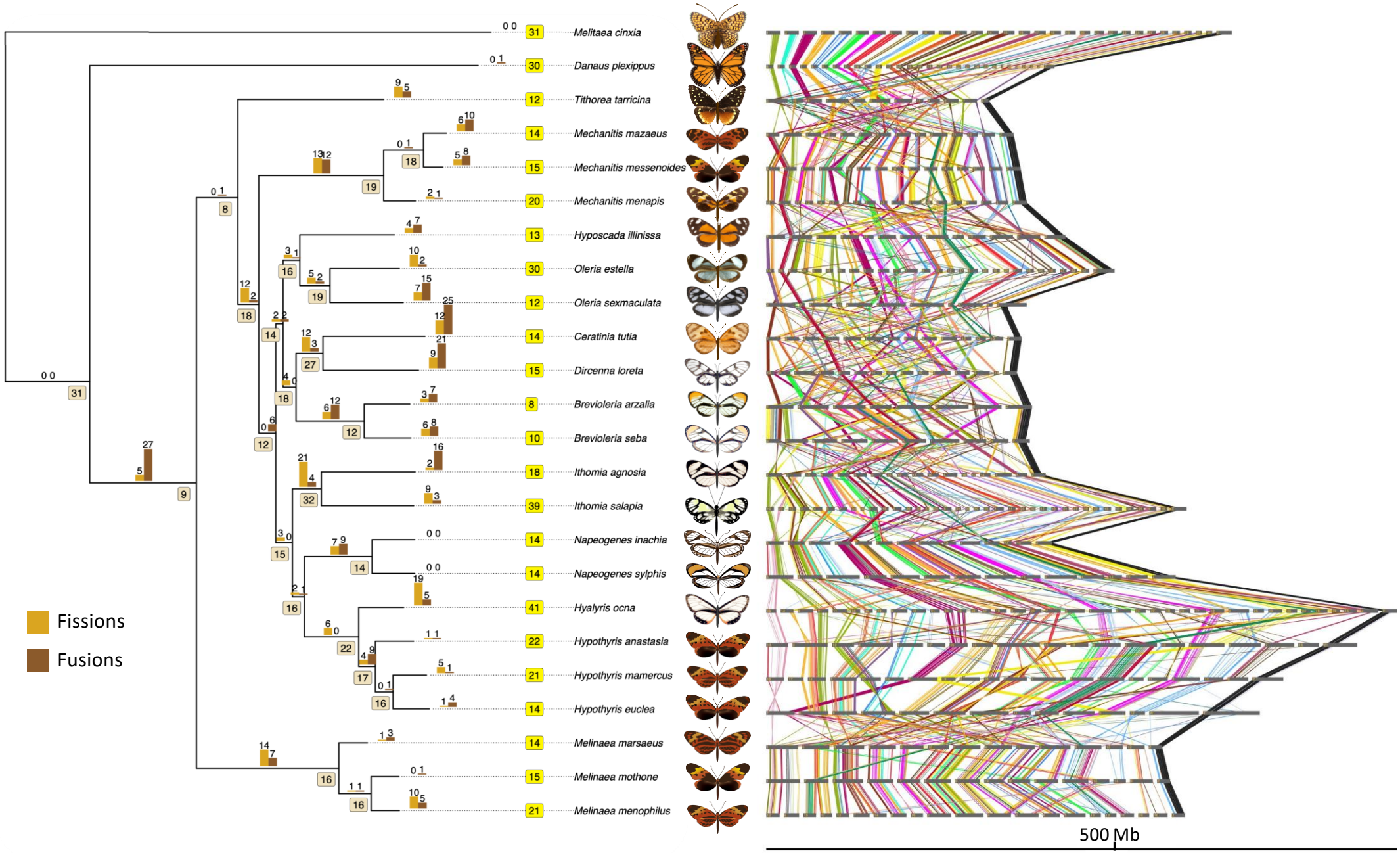
Large-scale chromosomal rearrangements



McClure et al. 2018

Divergence times: Chazot et al. 2019

Large-scale chromosomal rearrangements



Karin Näsvall

Species of recent radiations (<1 Mya) show highly rearranged chromosomes

M. macrinus (23)

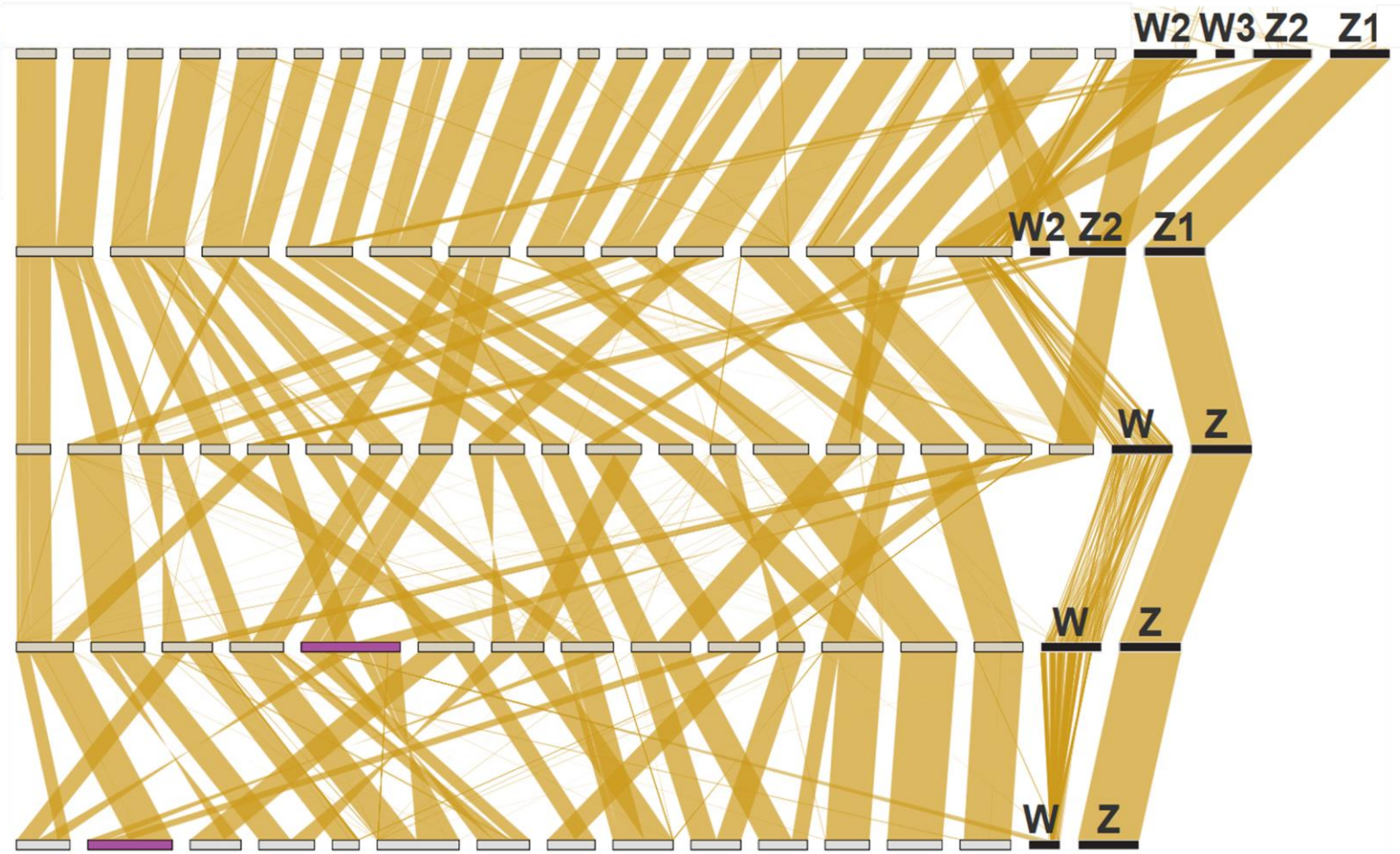
n chromosomes

M. mazaeus (14)

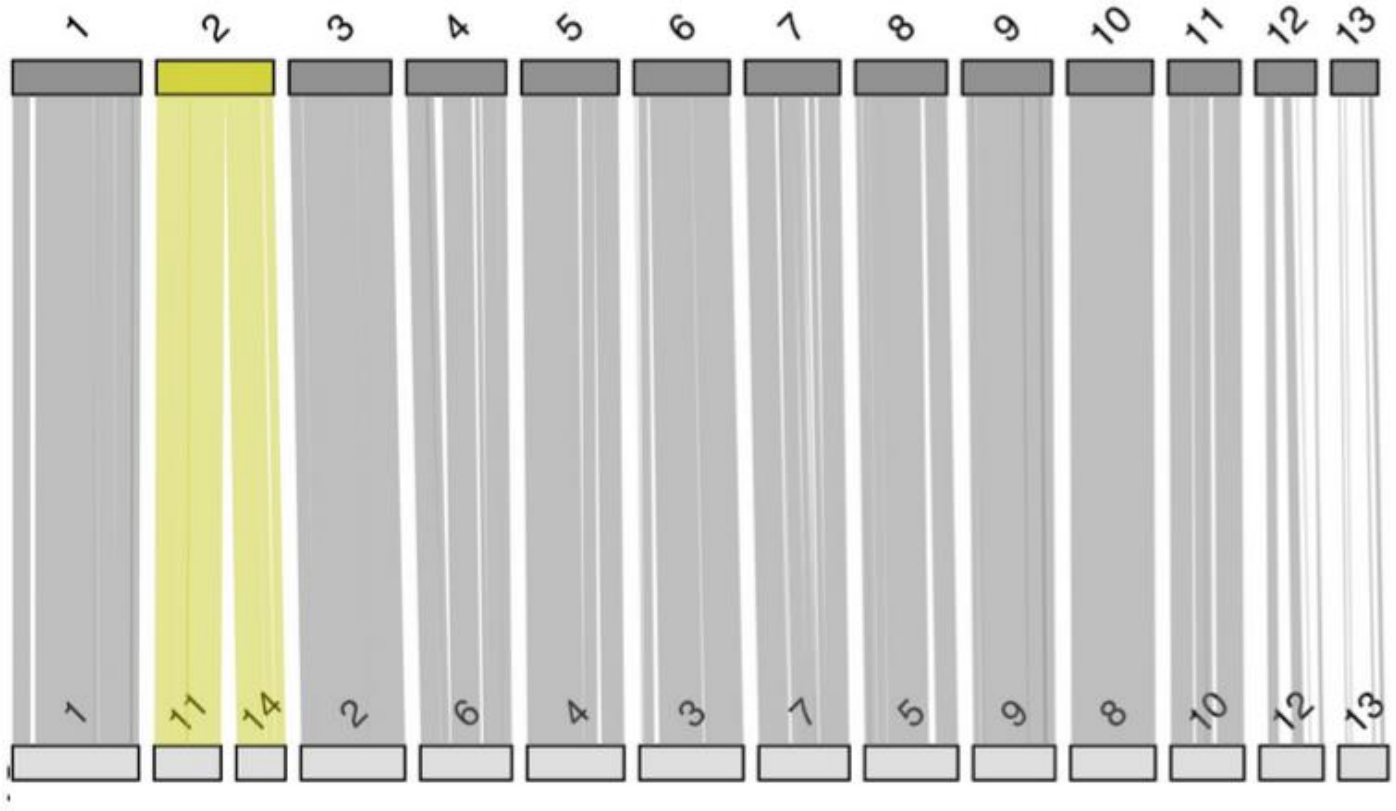
M. menapis (20)

M. messenoides (15

M. polymnia (15)

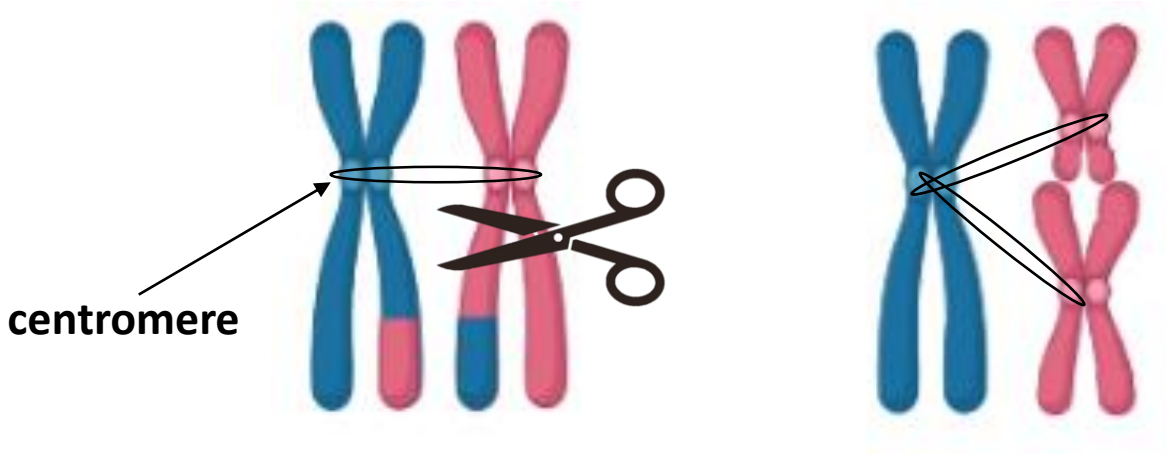


Some individuals are heterozygous for fission-fusion rearrangements

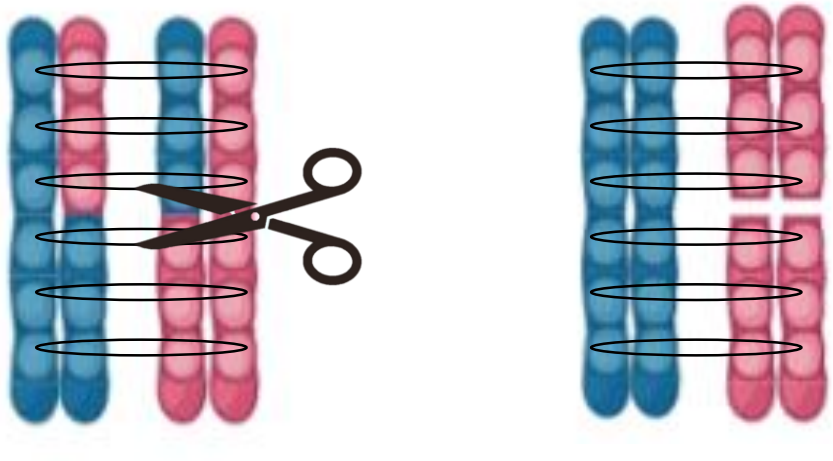


Butterflies have holocentric chromosomes

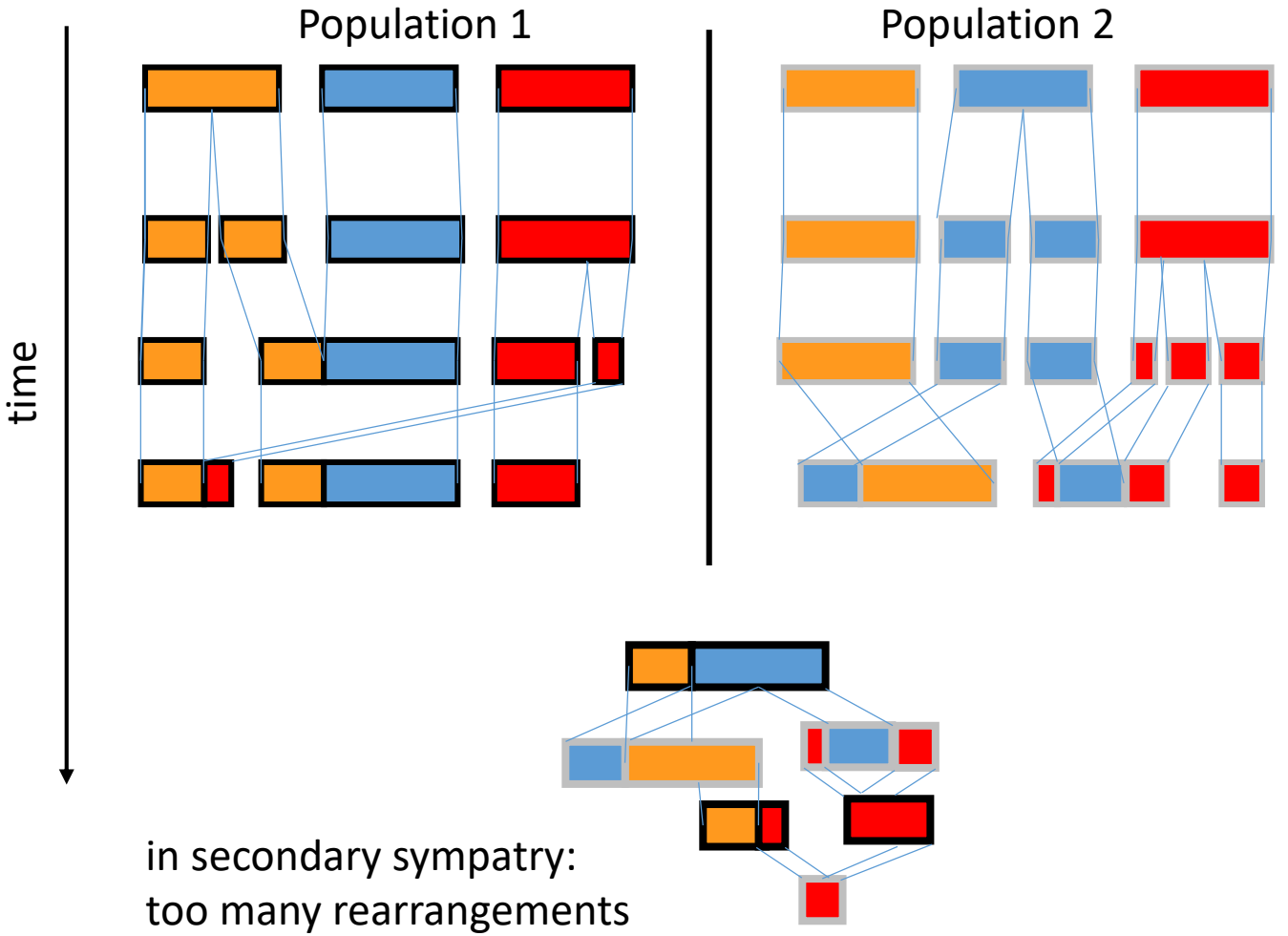
Monocentric
(like us)



Holocentric
(all butterflies)



Hypothesis: Chromosomal rearrangements during periods of geographic isolation drive speciation



in secondary sympatry:
too many rearrangements
lead to sterile hybrids

Van der Heijden et al., 2024

IKIAM
UNIVERSIDAD REGIONAL AMAZÓNICA

Many thanks to my wonderful team and key collaborator

@Sanger

Arif Maulana
PhD Student

Dr Karin Näsvall
Postdoctoral Fellow

Dr Nicol Rueda
Postdoctoral Fellow

Dr Patricio A. Salazar-Carrión
Postdoctoral Fellow

Jonah Walker
PhD Student

Eva van der Heijden
PhD student

@IKIAM en Ecuador

Franz Chandi

Kimberly Gavilanes

Alex Arias

María José Sánchez

Prof Caroline Bacquet