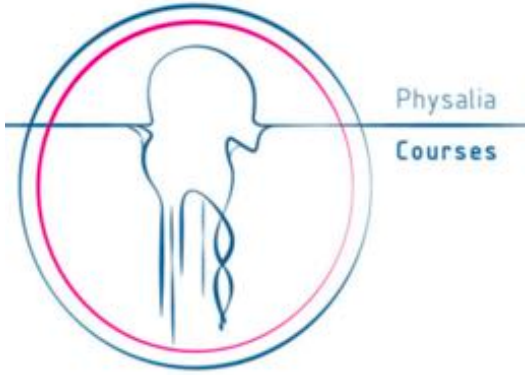


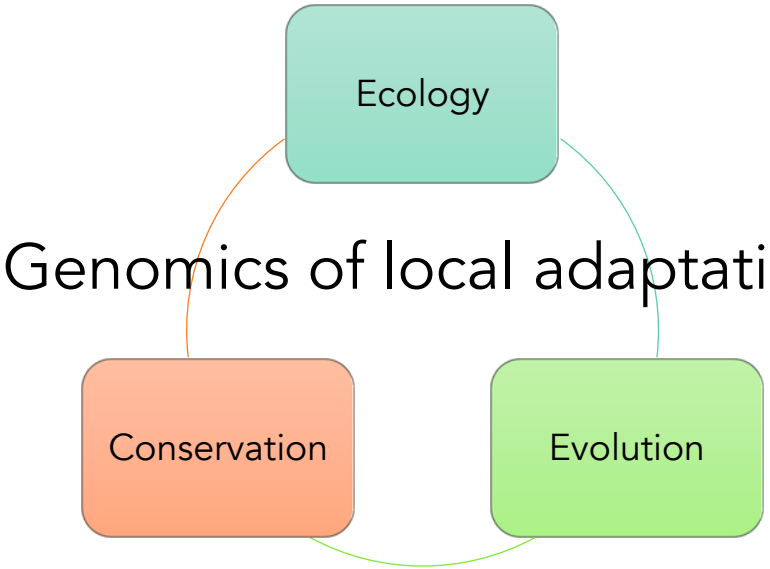
Adaptation Genomics Course

Anna Tigano, Ph.D. & Claire Mérot, Ph.D. & Yann Dorant

14-18 September 2020

Anna Tigano

anna.tigano@unh.edu | @t_annina | annatigano.weebly.com

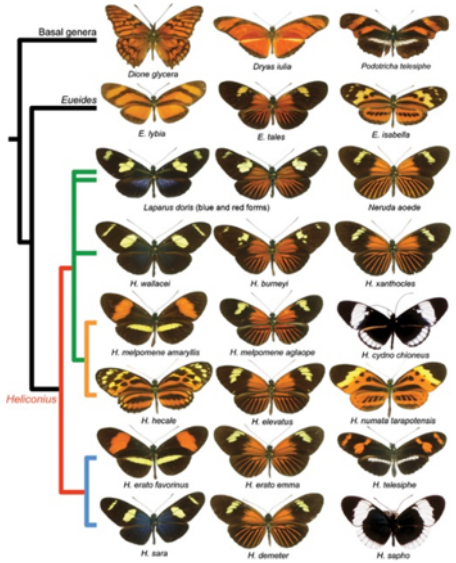


- Genomics of local adaptation with and without gene flow
- Genomic basis and architecture of adaptive traits
- Adaptation to extreme environments
- Structural variation and adaptation
- Conservation genomics

Claire Mérot

claire.merot@gmail.com
@ClaireMerot
http://www.normalesup.org/~cmerot/index_en.html

Faculté des sciences et de génie



Speciation in *Heliconius* butterflies

Mimicry

Environmental adaptation

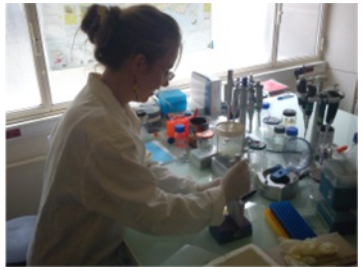
Evolution

Ecologie

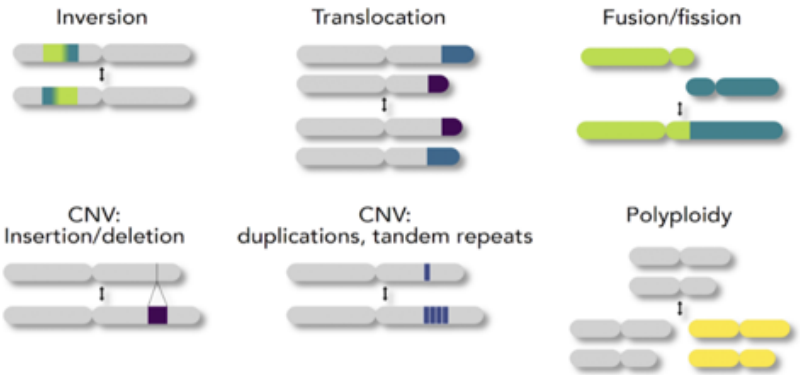
The evolution of biological diversity

Génomique

Inversion polymorphism in *Coelopa frigida* seaweed flies



Structural Variants



Yann Dorant

UNIVERSITÉ LAVAL

INSTITUT DE BIOLOGIE INTÉGRATIVE ET DES SYSTÈMES IBIS

- **Currently: PhD student in genomics (L. Bernatchez's Lab) Québec, Canada**
- Msc. Ecology of coastal areas and estuaries (France)
- Bsc. Biology (France)

Projects :

- **Population genomics of American lobster (H. americanus) in the Northwest Atlantic.**

- **Local adaptation spurred by structural variants in fish**

Bioinformatics

- **Code**
  - Bash
  - R
  - Python
- **Population structure**
  - Admixture
  - PCA
  - DAPC
  - IBD
- **Search for selection :**
  - OutFlank
  - BayPass
  - RDA
  - LFMM
- **Search for Structural variants:**
  - CNVs
  - Chromosomal inversions

Genomics

- RADseq/GBS
- Pool sequencing
- RAD-capture

Claire Yann Anna

Schedule

	Monday	Tuesday	Wednesday	Thursday	Friday
Morning 9-11:30	Lecture 1/ Practical 1	Lecture 1/ Lecture 2	Lecture 1/ Lecture 2	Lecture 1/ Practical 1	Lecture 1/ Lecture 2
Lunch 11:30-12:30	Break	Break	Break	Break	Break
Afternoon 12:30-15:00	Lecture 2/ Practical 2	Practical 1/ Practical 2	Practical 1/ Practical 2	Lecture 2/ Practical 2	Practical1/ Q&A

There will be a short 15 min break between lectures when both in the morning.
The instructors will take a break for lunch but otherwise be available for questions and support in the mornings and afternoons.
You are welcome to work at your own pace and when it's most convenient to you

You have access to the AWS server 9-15 EST
*** In case of particular needs ask Carlo for a time extension ***

Outline of the course

Day 1

Intro to adaptation genomics

Bioinformatics and sequencing approaches

Population genomics for adaptation

Practical

From raw data to variant calling

Outline of the course

Day 2

Genomic signatures of selection

Population structure as a confounding factor

Practical

Genetic diversity, population differentiation and structure

Outline of the course

Day 3

Confounding factors of signatures of selection

Outlier analyses and genotype-environment associations

Practical

Outlier analyses and genotype-environment associations

Outline of the course

Day 4

Structural variation

Large blocks of differentiation and structural variation

Practical

Analysis of haploblocks

Analysis of Copy Number Variants

Outline of the course

Day 5

Other methods to study the genomics of adaptation

Validation of candidate loci

Practical

Functional annotation of candidate loci for adaptation

Q & A

THE END!

Objectives

To get you familiar with bioinformatics, sequencing and analytical methods through the integration of *theory and empirical examples* to select the most appropriate approach to study the genomics of adaptation in your species of interest.

Adaptation genomics

The main goal of adaptation genomics is to understand the genomic basis and architecture of adaptive traits

Adaptation genomics

The main goal of adaptation genomics is to understand the genomic basis and architecture of adaptive traits

- Genetic basis of traits = loci that control the adaptive trait
- Genetic architecture = the interactions among alleles (dominance, epistasis, pleiotropy, polygeny)
- Genomic architecture = position of alleles and structural variants associated

Adaptation genomics

The main goal of adaptation genomics is to understand the genomic basis and architecture of adaptive traits

Ecology

Often local adaptations are not apparent, and we use a top-down approach to understand what species/populations are adapted to

Adaptation genomics

The main goal of adaptation genomics is to understand the genomic basis and architecture of adaptive traits

Ecology

Evolution

By identifying the genes underpinning local adaptation we can gain insights into the process of adaptation and the interplay among evolutionary forces

Adaptation genomics

The main goal of adaptation genomics is to understand the genomic basis and architecture of adaptive traits

Ecology

Evolution

Conservation

Understanding how organisms have adapted in the past can help us predict their potential to future changes in their environment

Adaptation genomics

The main goal of adaptation genomics is to understand the genomic basis and architecture of adaptive traits

Ecology

Evolution

Conservation

Management

Assessment of adaptive differentiation ensures appropriate management of population/species of socio-economic importance (e.g., fish stocks, game species)

Adaptation genomics

The main goal of adaptation genomics is to understand the genomic basis and architecture of adaptive traits

Ecology

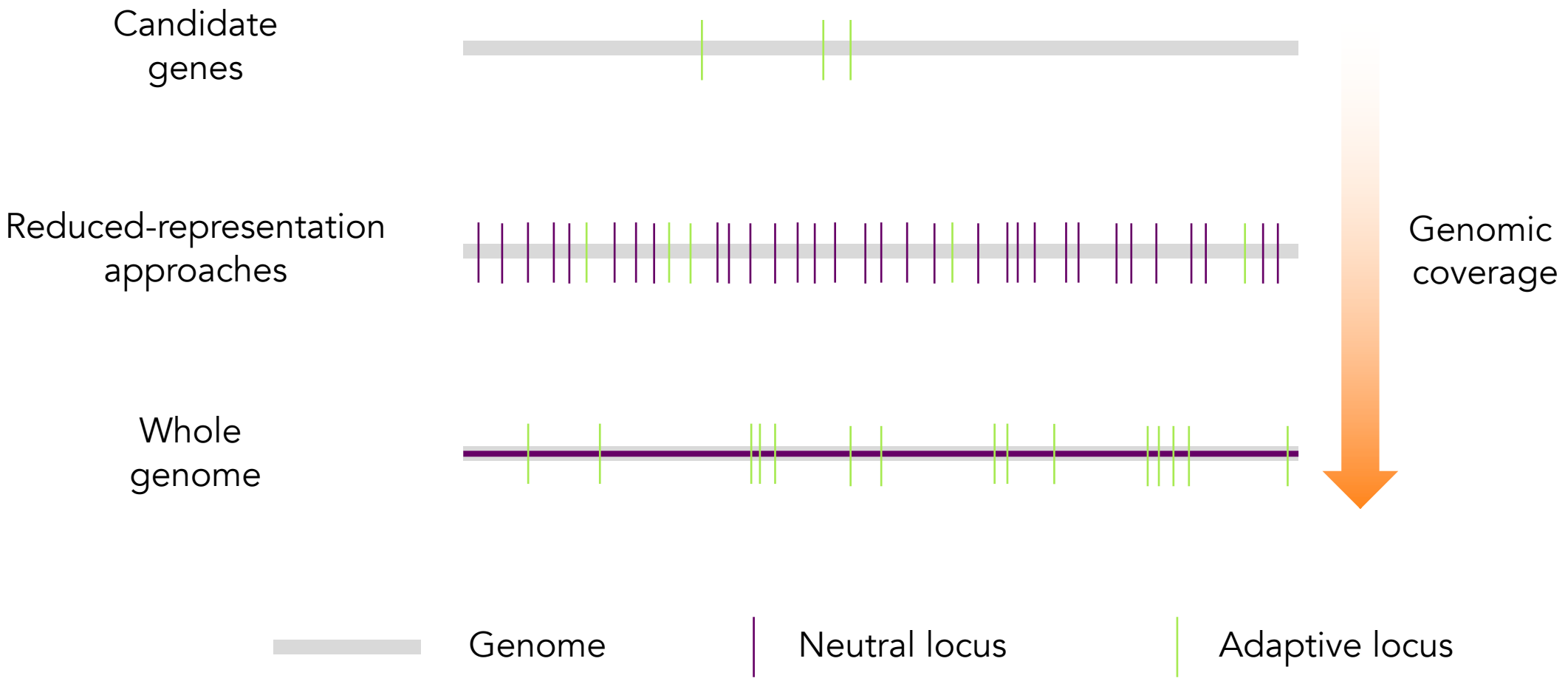
Evolution

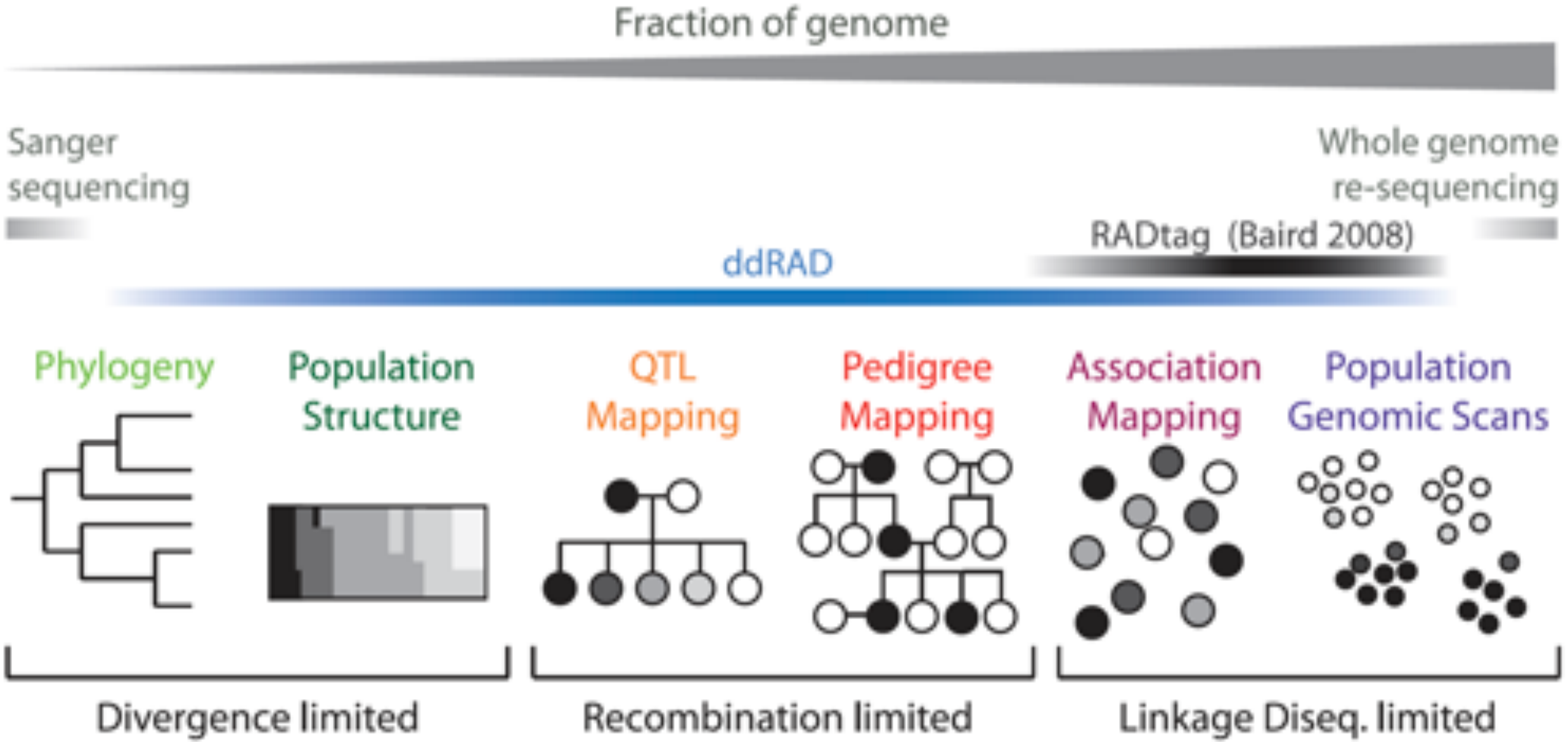
Conservation

Management

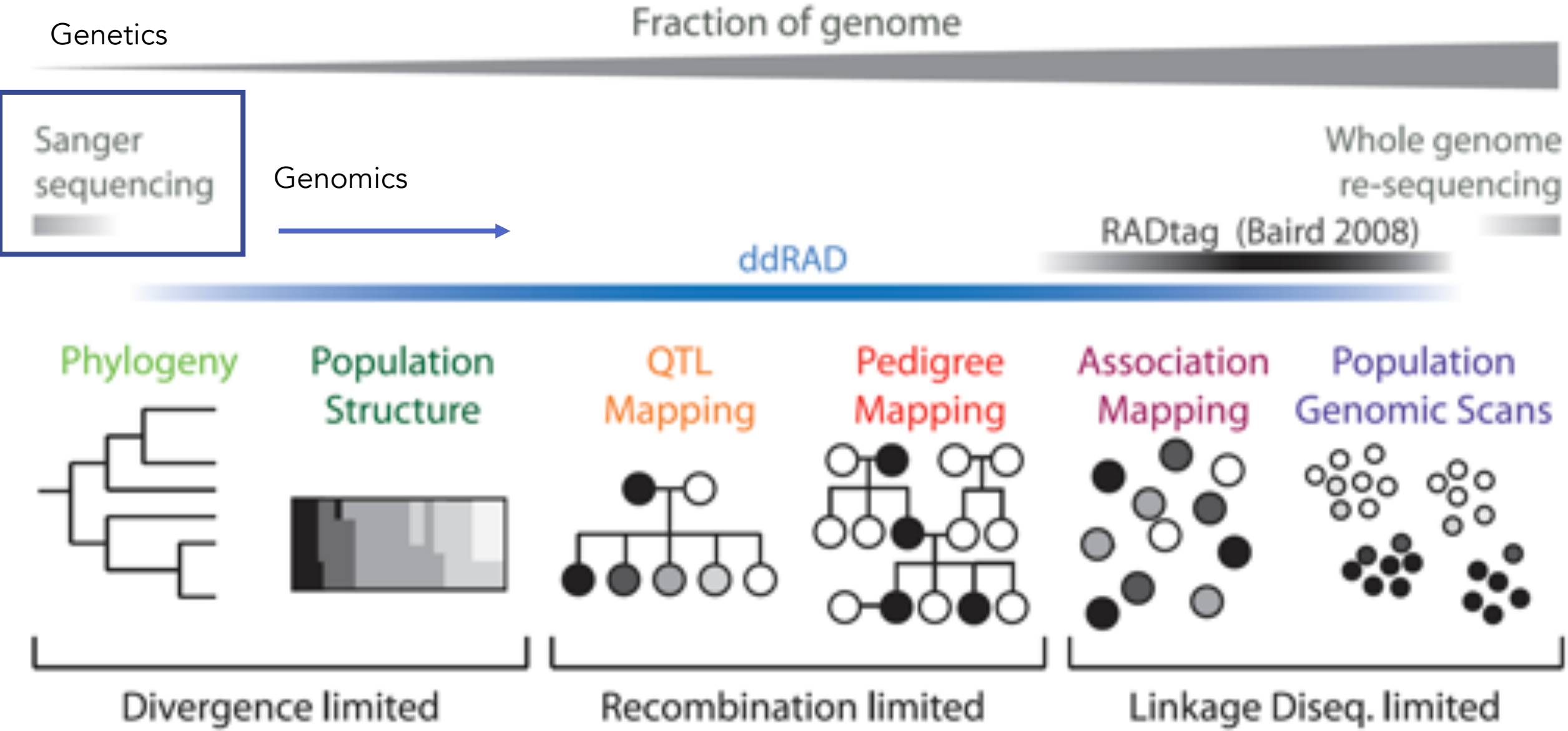
Physiology, molecular evolution, biodiversity, speciation…

Sequencing approaches





Peterson et al. 2012, PlosOne



Peterson et al. 2012, PlosOne

Reduced-representation approaches

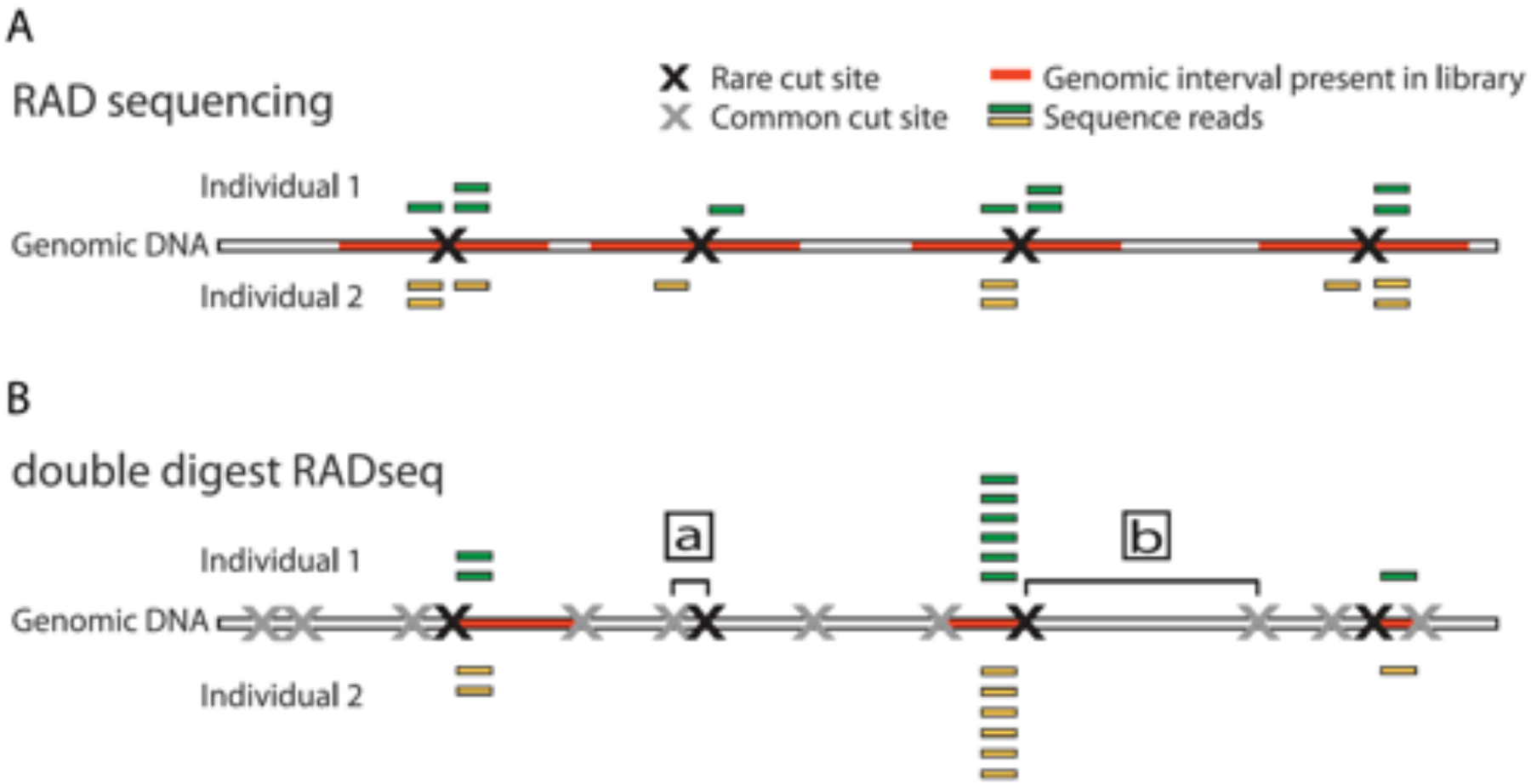
- RADseq/GBS

Random sampling of the genome

- Exome/exon capture
- SNP chip

Targeted capture of loci of interest

RADseq



Peterson et al. 2012, PlosOne

Single vs. double digestion

Andrews et al. 2016, Nature Gen. Rev.

Many different RADseq protocols

	mbRAD	ddRAD	ezRAD	2bRAD
Restriction cut sites per 10 kb*	~0.2–2.4	~3.7×10^{-5}–39	~39	~2.4
Postdigest fragment reduction	Size selection	Size selection	Size selection	Selective adapters
Contigs > 200 bp†	Yes	No	Some	No
Ability to blast/annotate *de novo* contigs	High	Mid	Mid	Low
Protocol complexity (# Steps)‡	6	4	4–6	3
Level of technical difficulty	High	Mid	Low	Low
Level of technical support	Low	Low	Mid-high	Low
Insert complexity (first × bases)	Low	Low	Very low	High
PCR AT/GC content, copy number Bias among loci	Yes	Yes	Yes, No§	Yes
ID of PCR duplicates	Yes	No	No§	No¶
Uniform locus length	No	No	No	Yes
Oligos required to uniquely identify and build 96 libraries	196**	31	20–22	37
Target insert size range	200–600 bp	Customizable	Customizable	33–36 bp

*These numbers represent only theoretical calculations for one enzyme (or enzyme combination). The number of fragments sampled will depend on size selection, genome composition, the number of enzymes used and the use of restrictive adapters (see 2bRAD).
†When performing 100 bp reads such as on a HiSeq platform.
‡Not counting clean-up steps.
§ezRAD can be used with a PCR-free library preparation kit, thus removing the need to detect PCR duplicates.
¶2bRAD can detect PCR errors by mismatch among forward and reverse reads on individual strands.
**With some effort, the indexing for mbRAD can be modified to reduce the oligo counts to 22–37.

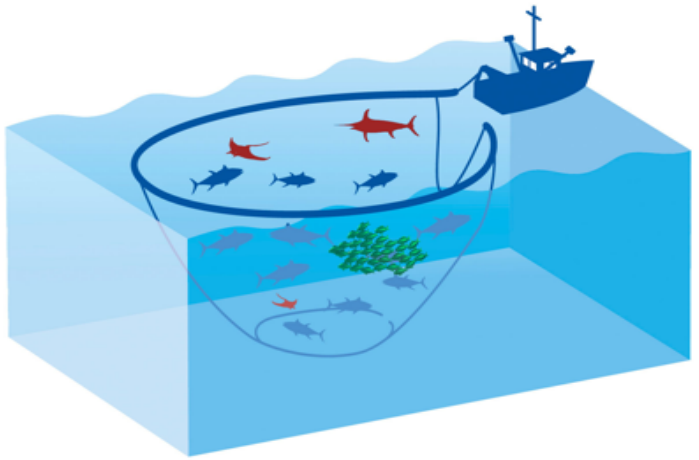
Puritz et al. 2014, Molecular Ecology

Targeted approaches

Advantages:

- Scalable and cost-effective
- Lower variance in target coverage
- More accurate SNP calling
- Higher reproducibility
- Can be combined with other reduced-representation approaches

RADseq

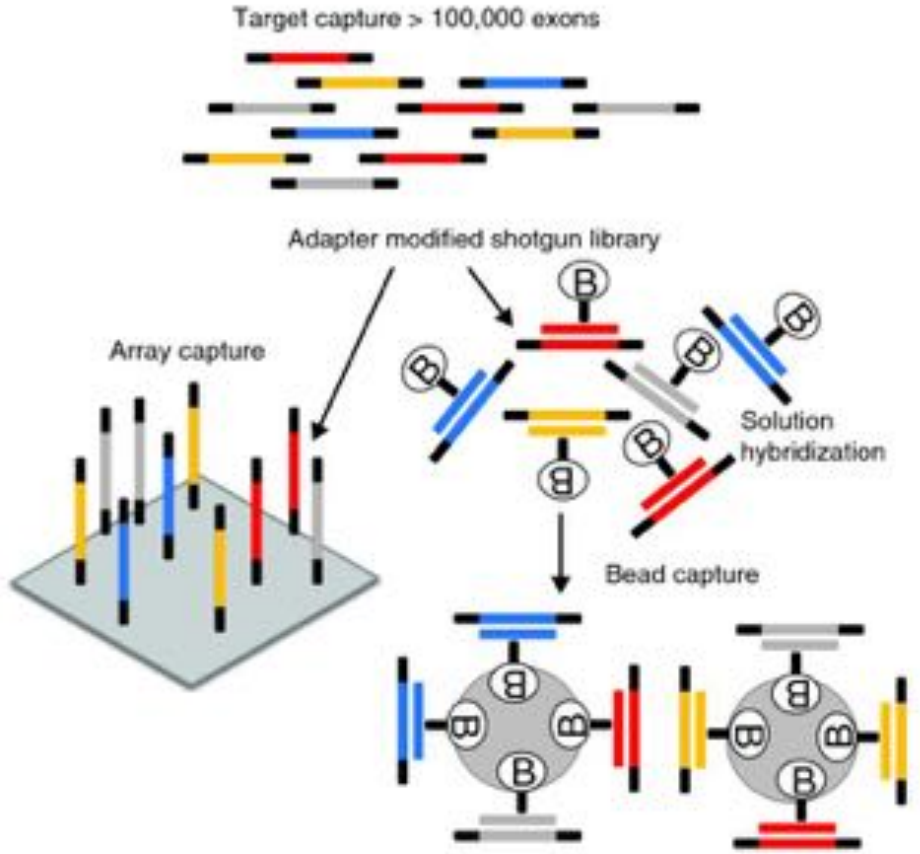


Targeted sequencing - RAPTURE

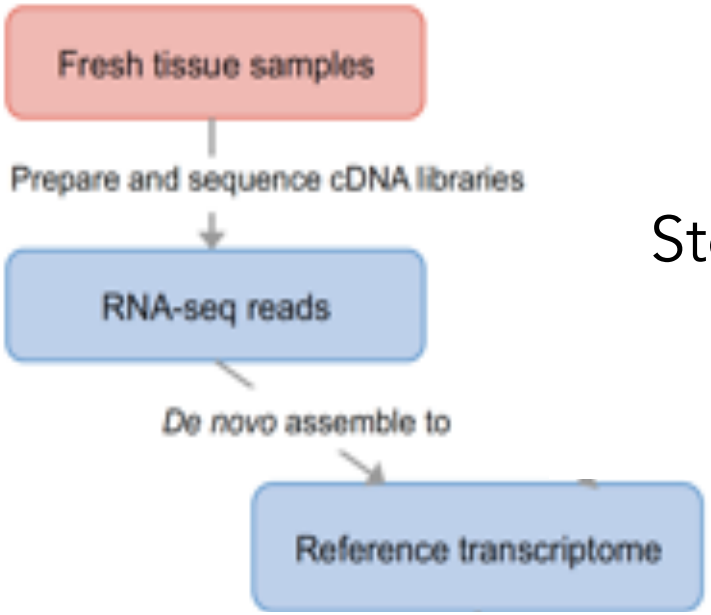
Targeted approaches - Exome/exon capture

Used to sequence protein coding genes (or other sequences as well).

While probes are generally available for some model species (human, mouse), they have to be designed for other species.



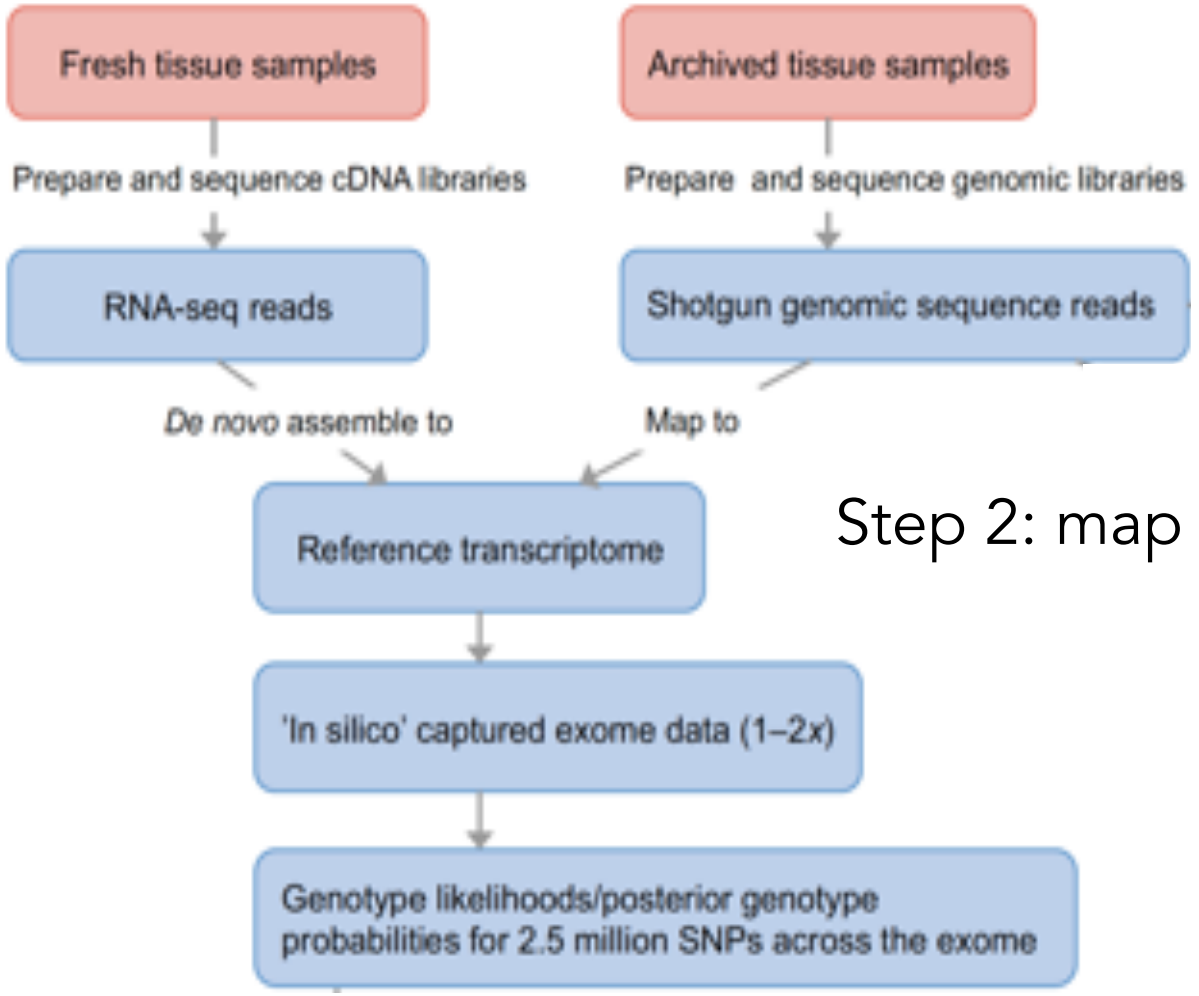
Targeted approaches - Exome capture (*in silico*)



Step 1: build a reference transcriptome

Therkildsen & Palumbi 2017, Mol. Ecol. Res.

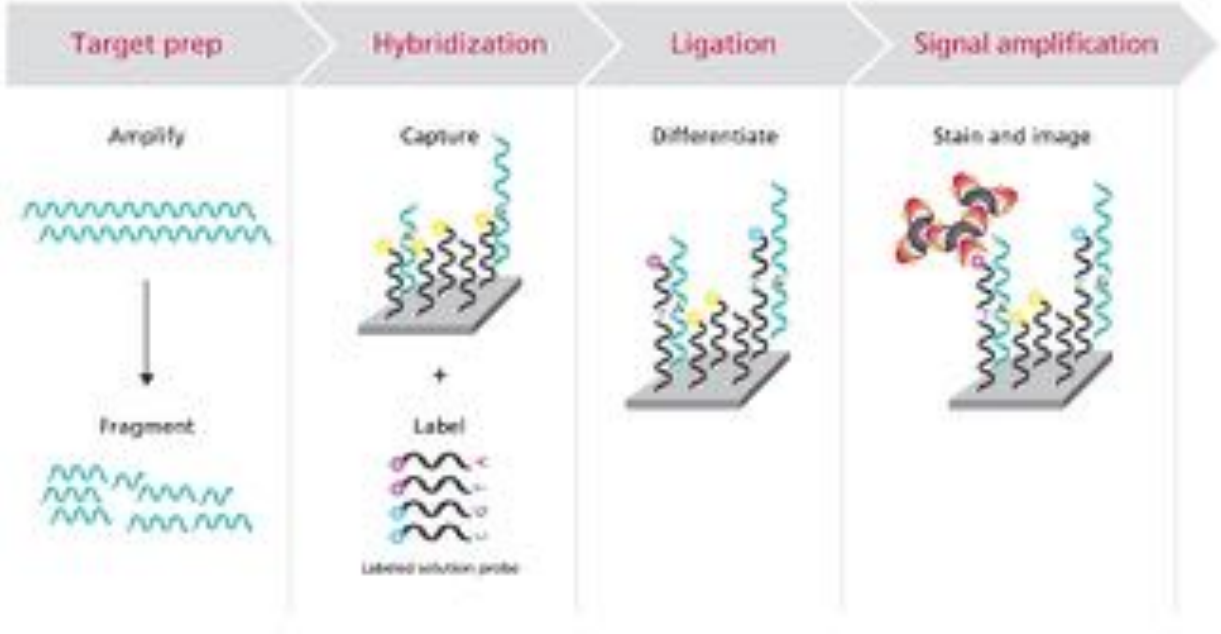
Targeted approaches - Exome capture (*in silico*)



Step 2: map data to reference

Therkildsen & Palumbi 2017, Mol. Ecol. Res.

Targeted approaches - SNP chip array



Cost-effective to genotype high number of SNPs in large number of samples.

Targeted approaches - SNP chip array

Antarctic fur seal

85k Affymetrix Axiom genotyping array includes SNPs from

- Previous RADseq markers
- transcriptome markers
- MHC loci

→ To identify loci of adaptive importance and monitor levels of standing genetic variation

Humble et al. 2020, G3

Whole genome resequencing

Short-read sequencing

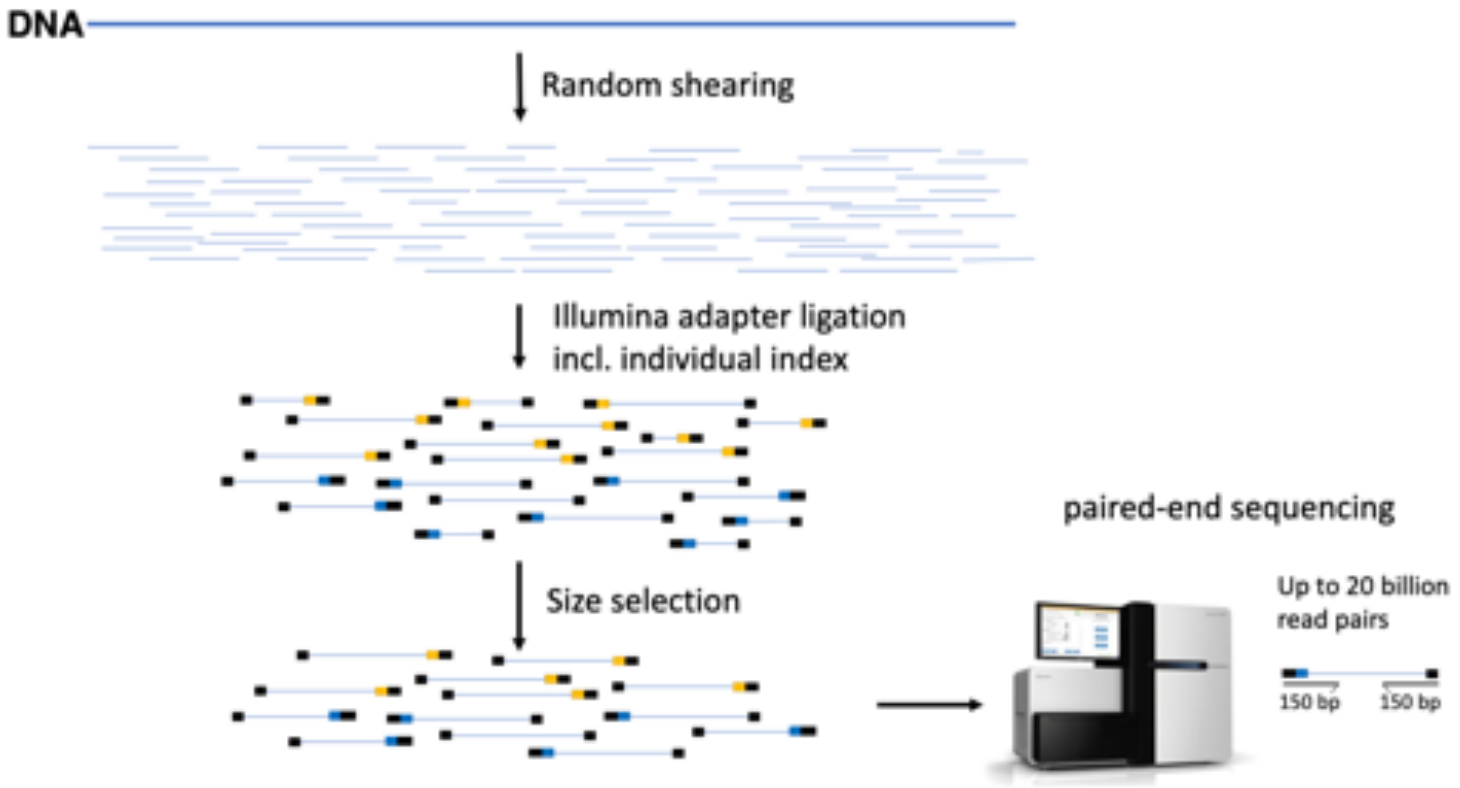
Long-read sequencing

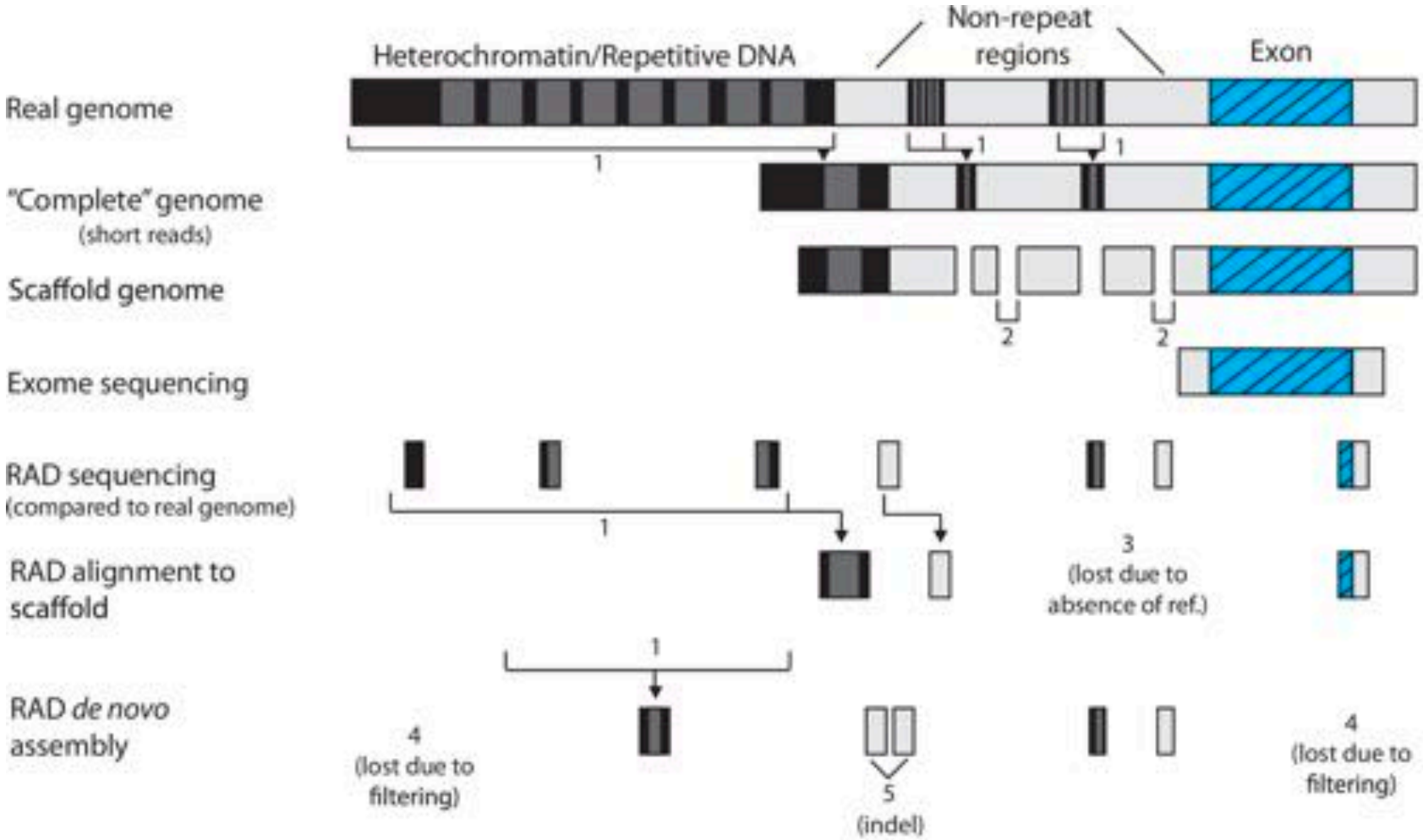
Linked-reads technology

???

Whole genome resequencing

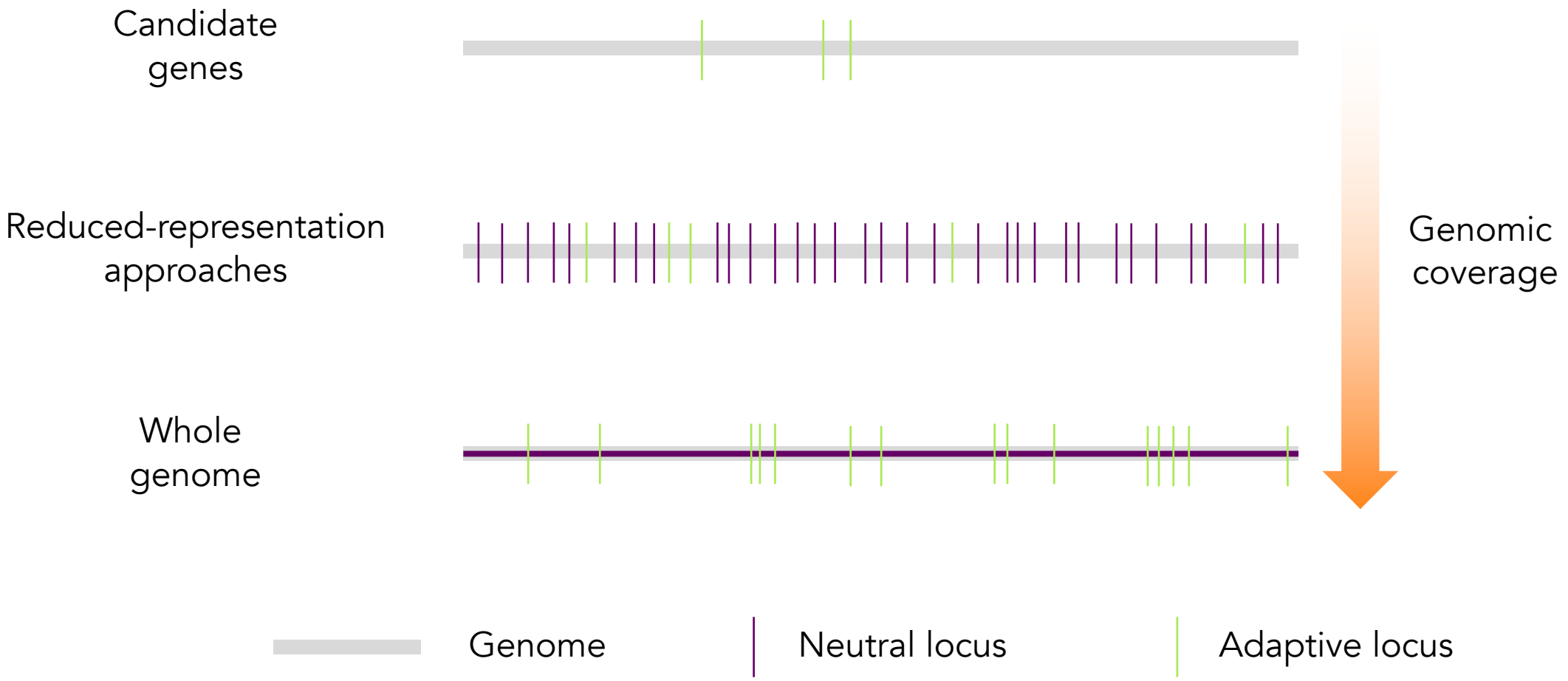
Short-read sequencing



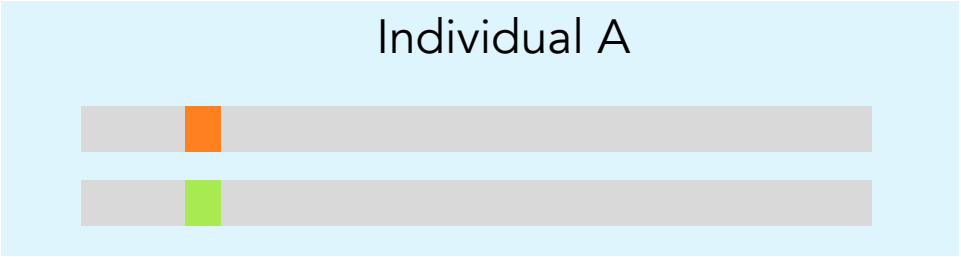


Hoban et al. 2016, American Naturalist

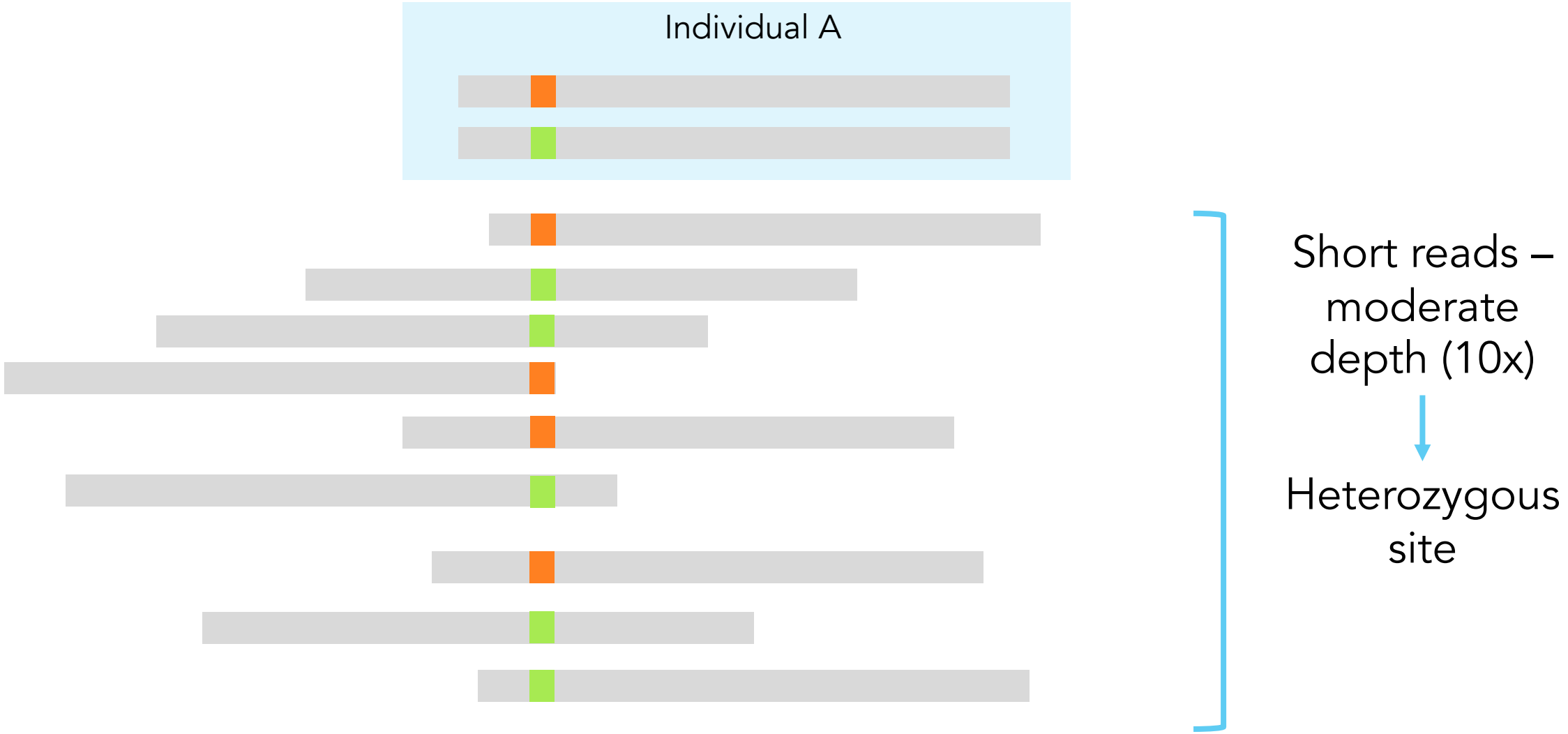
Sequencing approaches



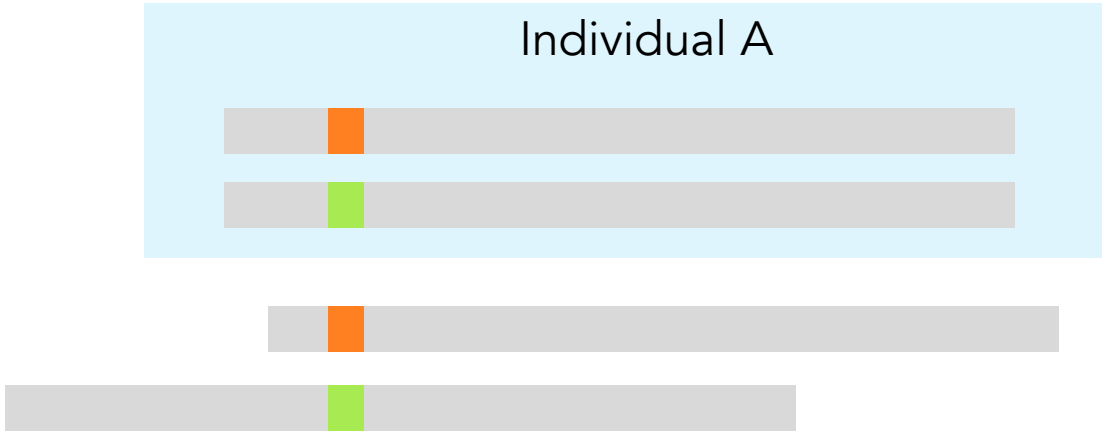
Sequencing depth = the number of reads covering a given site



Sequencing depth = the number of reads covering a given site



Sequencing depth = the number of reads covering a given site

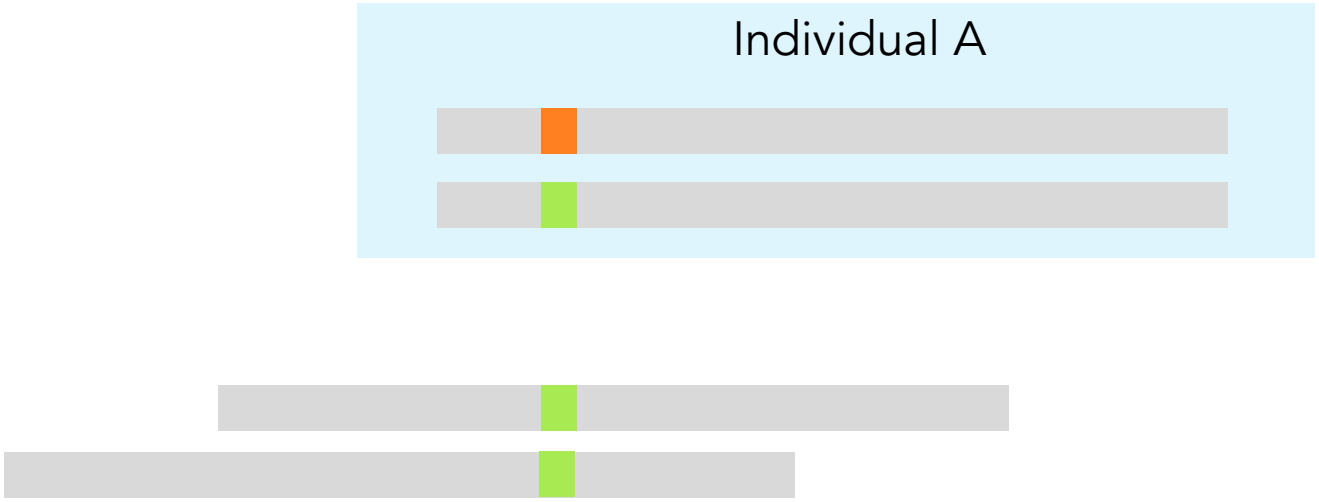


Short reads – low depth (2x)

↓

Heterozygous site

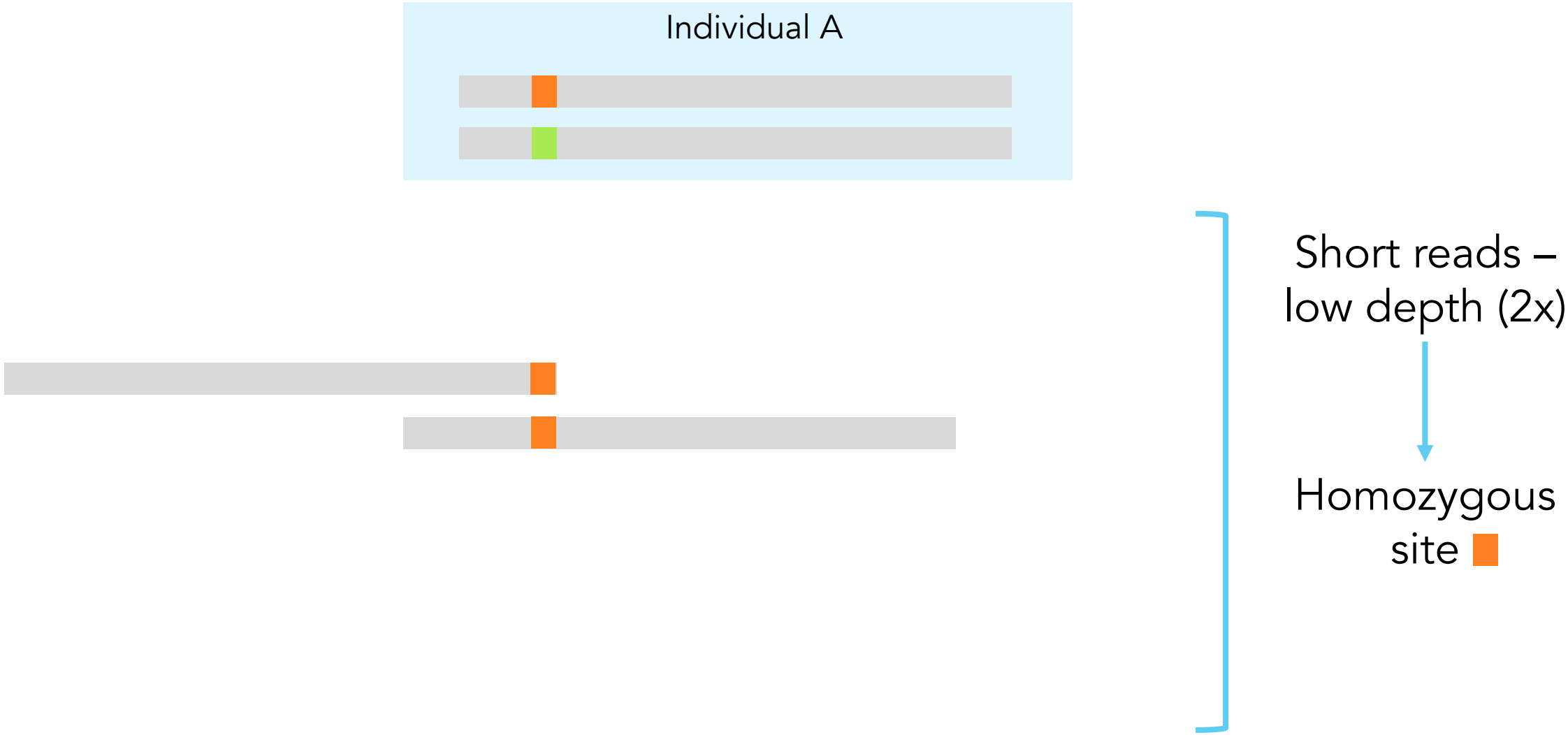
Sequencing depth = the number of reads covering a given site



Short reads – low depth (2x)

Homozygous site

Sequencing depth = the number of reads covering a given site

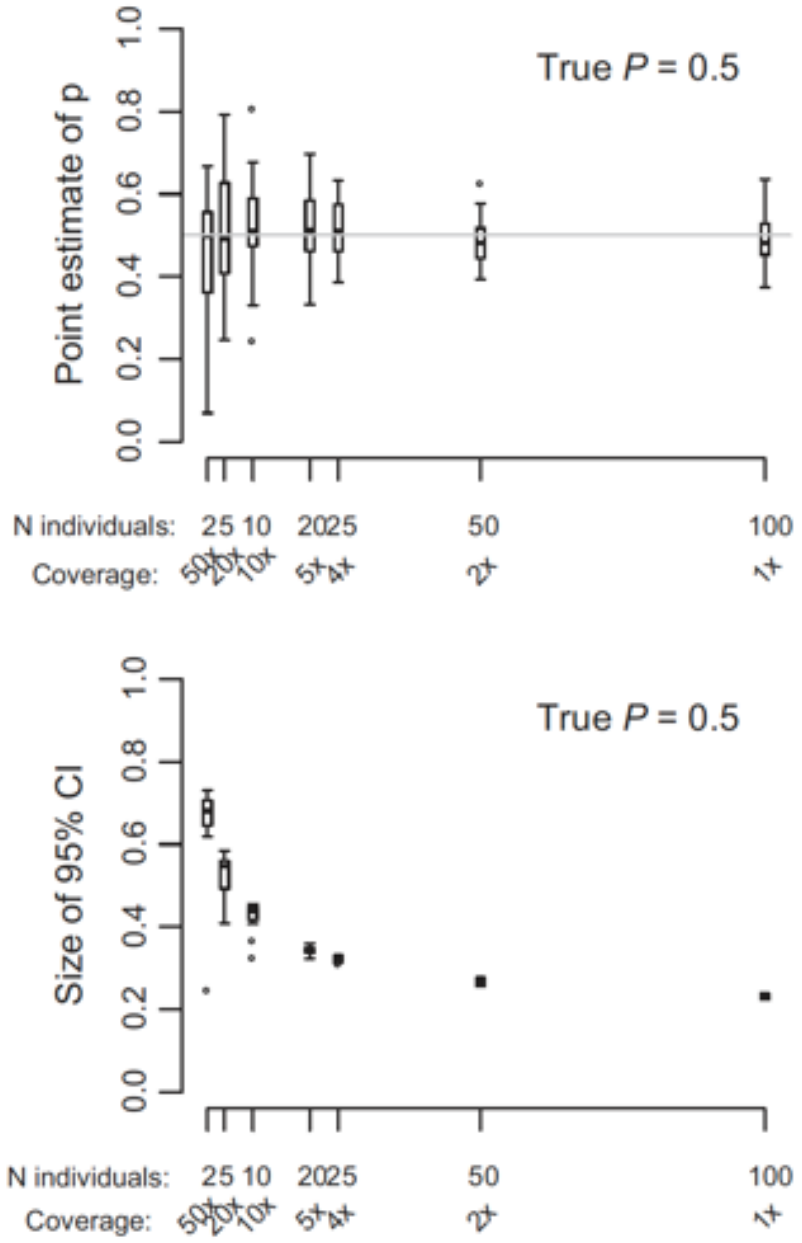


Low-coverage sequencing

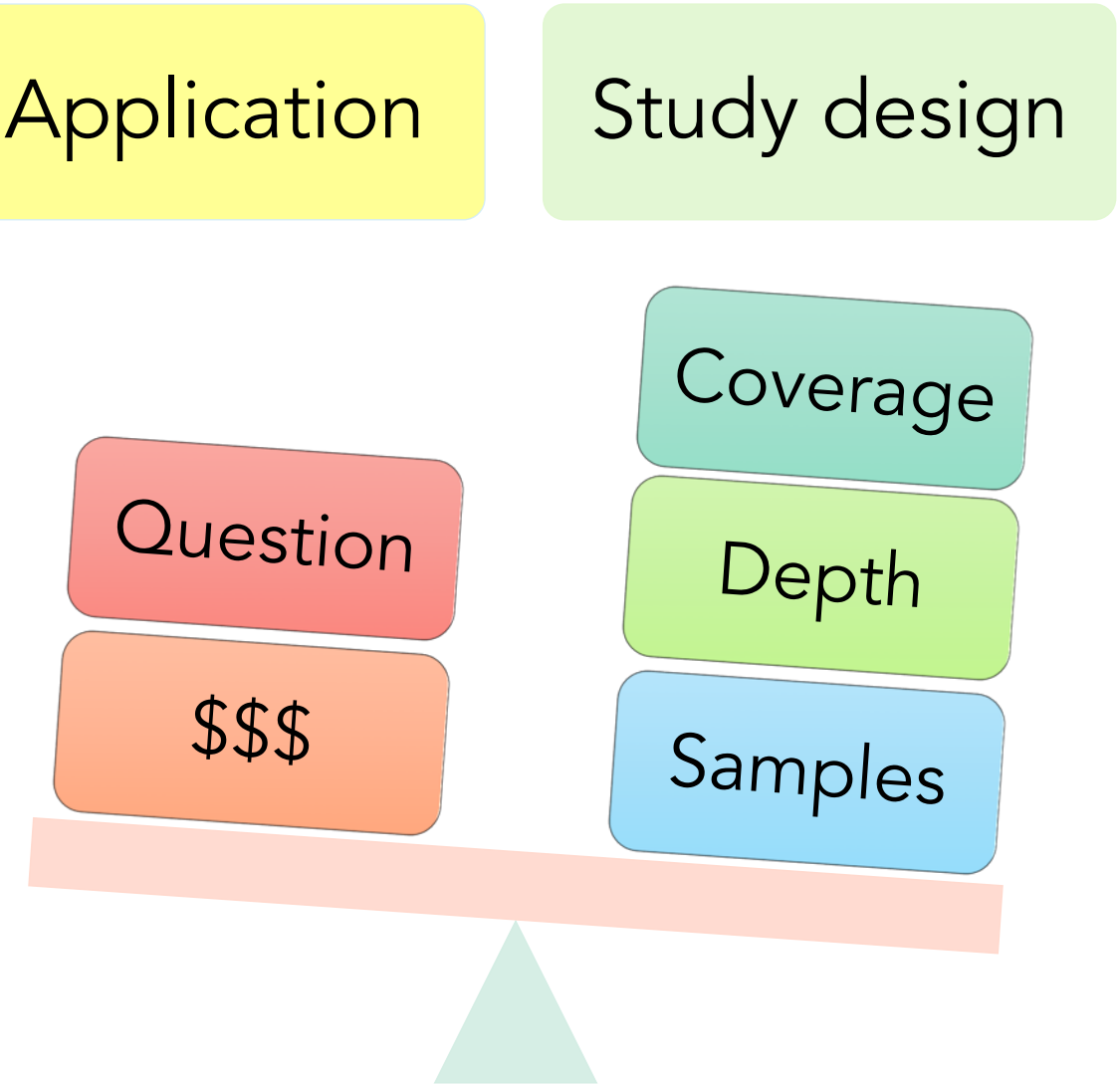
However, most population genomics analyses collapse genotypes to population allele frequencies.

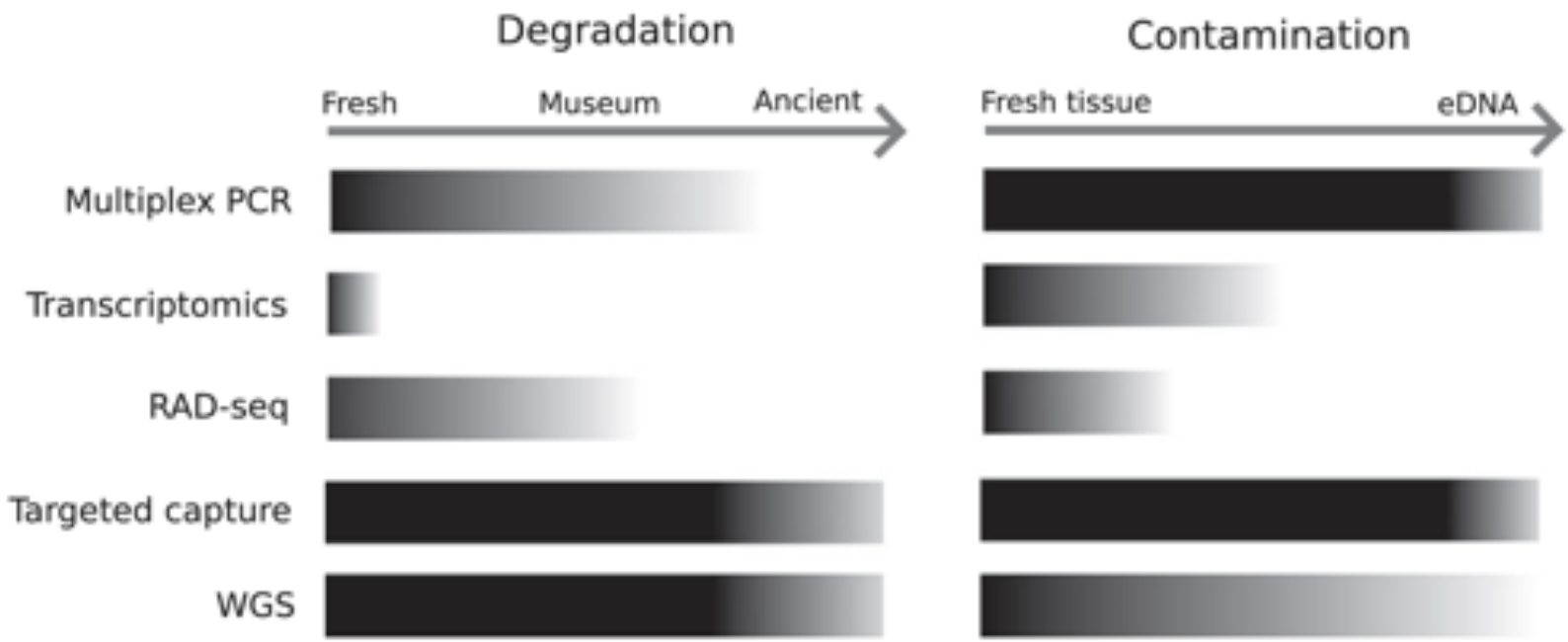
In those cases, high number of individuals at low depth provide more accurate estimates than a few individuals sequenced at higher depths.

When genotypes are necessary, they can be associated with genotype uncertainty in a probabilistic framework.



Buerkle and Gompert 2013, Molecular Ecology


Application
Study design
Question
$$$
Coverage
Depth
Samples



Jones & Good 2016, Molecular Ecology

In addition to all these technical aspects, there are many evolutionary and molecular factors to consider to choose the most appropriate sequencing approach for your study.

We will explore those throughout the rest of the week.

The end.