

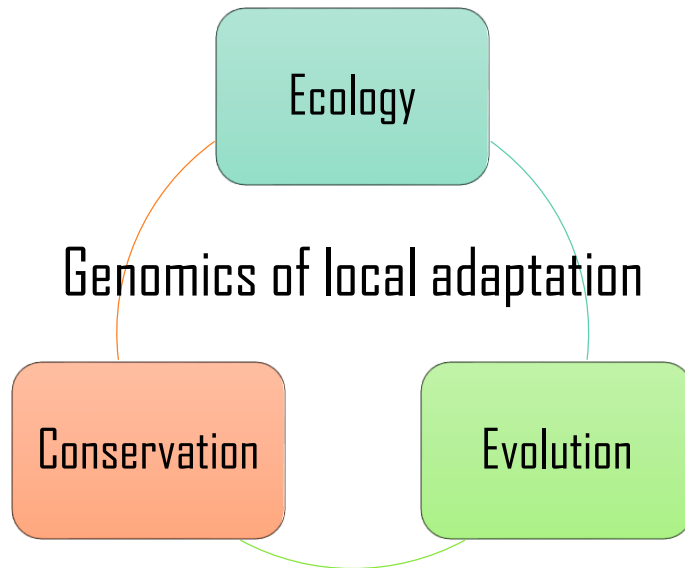
# Adaptation Genomics Course

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

May 30 - June 3, 2022



# Anna Tigano, Ph.D



Cornell University



[anna.tigano@ubc.ca](mailto:anna.tigano@ubc.ca) | [@t\\_annina](https://twitter.com/at_annina) | [annatigano.weebly.com](http://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

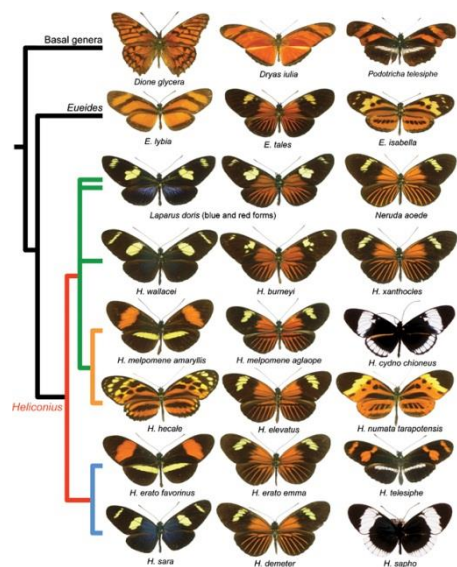


Claire Mérot



[claire.merot@gmail.com](mailto:claire.merot@gmail.com)  
[claire.merot@univ-rennes1.fr](mailto:claire.merot@univ-rennes1.fr)  
@ClaireMerot

[http://www.normalesup.org/~cmerot/index\\_en.html](http://www.normalesup.org/~cmerot/index_en.html)



Speciation in  
*Heliconius*  
butterflies

Evolution

Ecologie

*The evolution of  
biological diversity*

Génomique



Mimicry



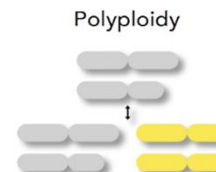
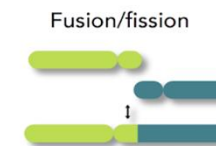
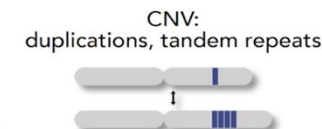
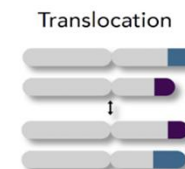
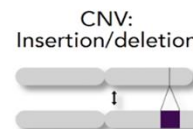
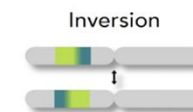
Environmental  
adaptation



Inversion polymorphism in  
*Coelopa frigida* seaweed flies



Structural  
Variants





# Yann Dorant, Ph.D



y.dorant@gmail.com | @YannDorant



- Population genetic structure
- Genome architecture and adaptation to environment
- Seascape genomics
- Transcriptomics and functional genomics
- Genomics for fisheries and aquaculture

**Claire**



**Yann**



**Anna**



# Schedule

CET (Europe)	PST (West America)	Monday	Tuesday	Wednesday	Thursday	Friday
14:30-17h	5h30-8h	Introduction/ Practical 1	Lecture 2-1/ Practical 2	Lecture 3-1/ Practical 3	Lecture 4-1/ Practical 4	Lecture 5-1/ Practical 5
		Break	Break	Break	Break	Break
18h-20h30	9h-11h30	Lecture 1-2/ Practical 2	Lecture 2-2/ Practical 2	Lecture 3-2/ Practical 3	Lecture 4-2/ Practical 4	Lecture 5-2/ Q&A

The instructors will take a break between the two periods but otherwise be available for questions and support (depending on time zone)  
You are welcome to work at your own pace and when it's most convenient to you

You have access to the AWS server from 12:30 to 22:30 (CET)

\*\*\* 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

Connect to the server  
From raw data to variant calling

# Outline of the course

## Day 2

Population structure as a confounding factor

Genomic signatures of selection

## Practical

Genetic diversity, population differentiation and structure



# Outline of the course

## Day 3

Outlier analyses and genotype-environment associations  
Confounding factors of signatures of selection

## Practical

Outlier analyses and genotype-environment associations

# Outline of the course

## Day 4

Detecting structural variation

Evolutionary significance of structural variation

## Practical

Analysis of haploblocks

Analysis of structural 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**

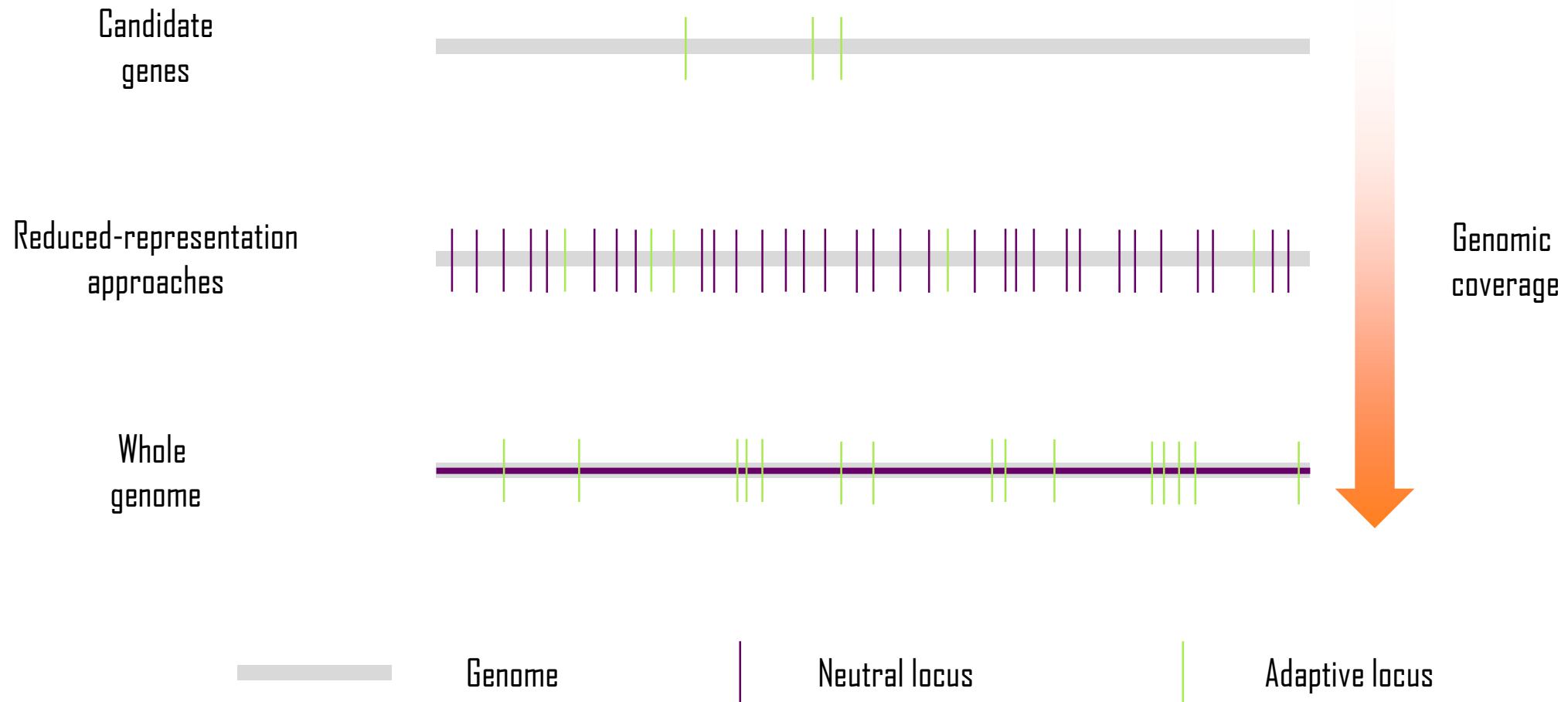
**Evolution**

**Conservation**

**Management**

Physiology, molecular evolution,  
biodiversity, speciation...

# Sequencing approaches



Genetics

Fraction of genome

Sanger sequencing

Genomics

Whole genome re-sequencing

RADtag (Baird 2008)

ddRAD

Phylogeny

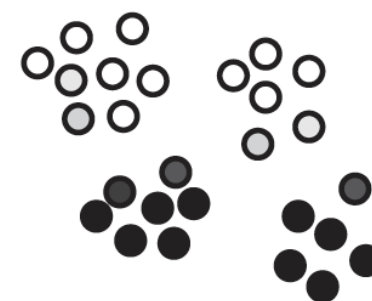
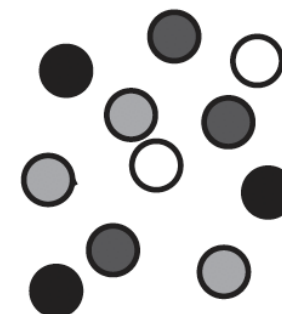
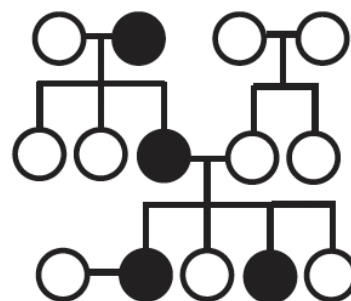
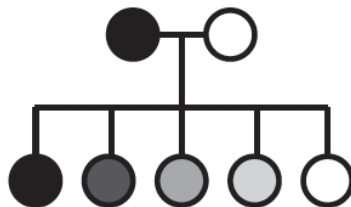
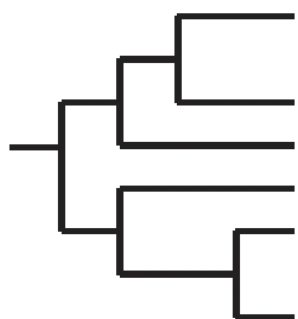
Population Structure

QTL Mapping

Pedigree Mapping

Association Mapping

Population Genomic Scans



Divergence limited

Recombination limited

Linkage Diseq. limited

# Reduced-representation approaches

- RADseq/GBS



Random sampling of the genome

- Exome/exon capture



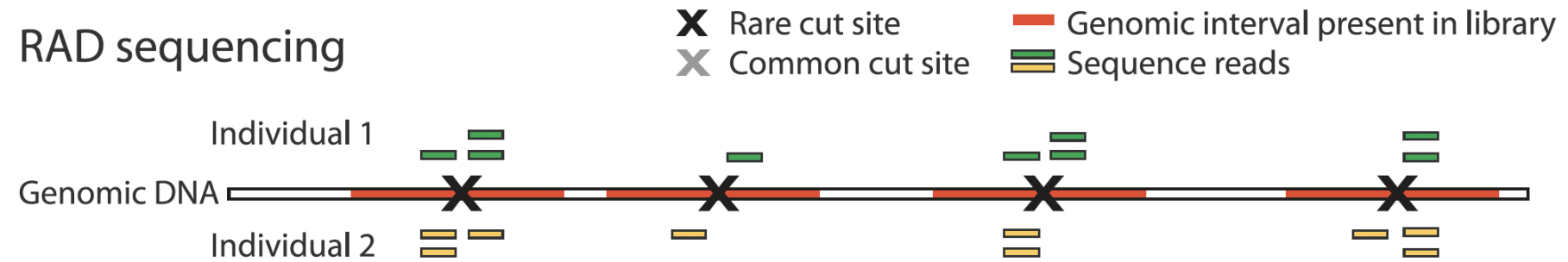
Targeted capture of loci of interest

- SNP chip

# RADseq

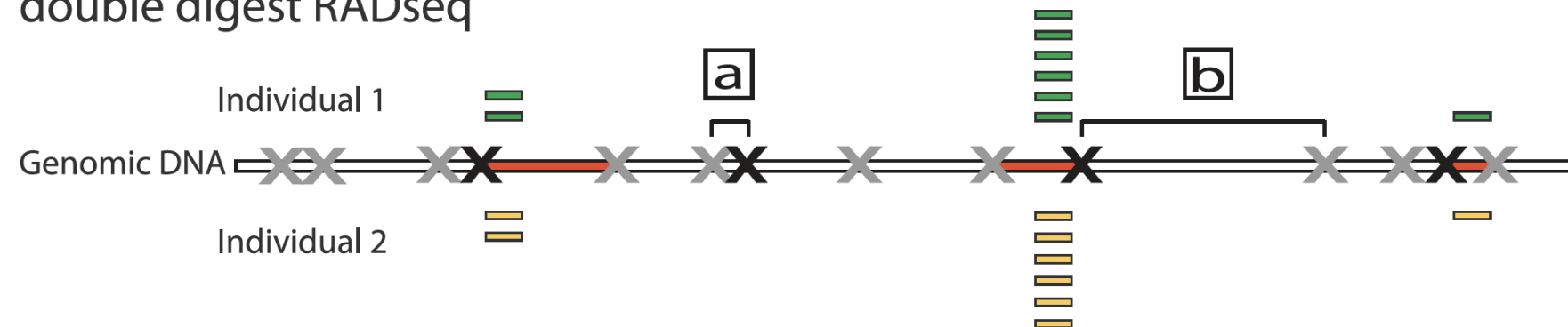
A

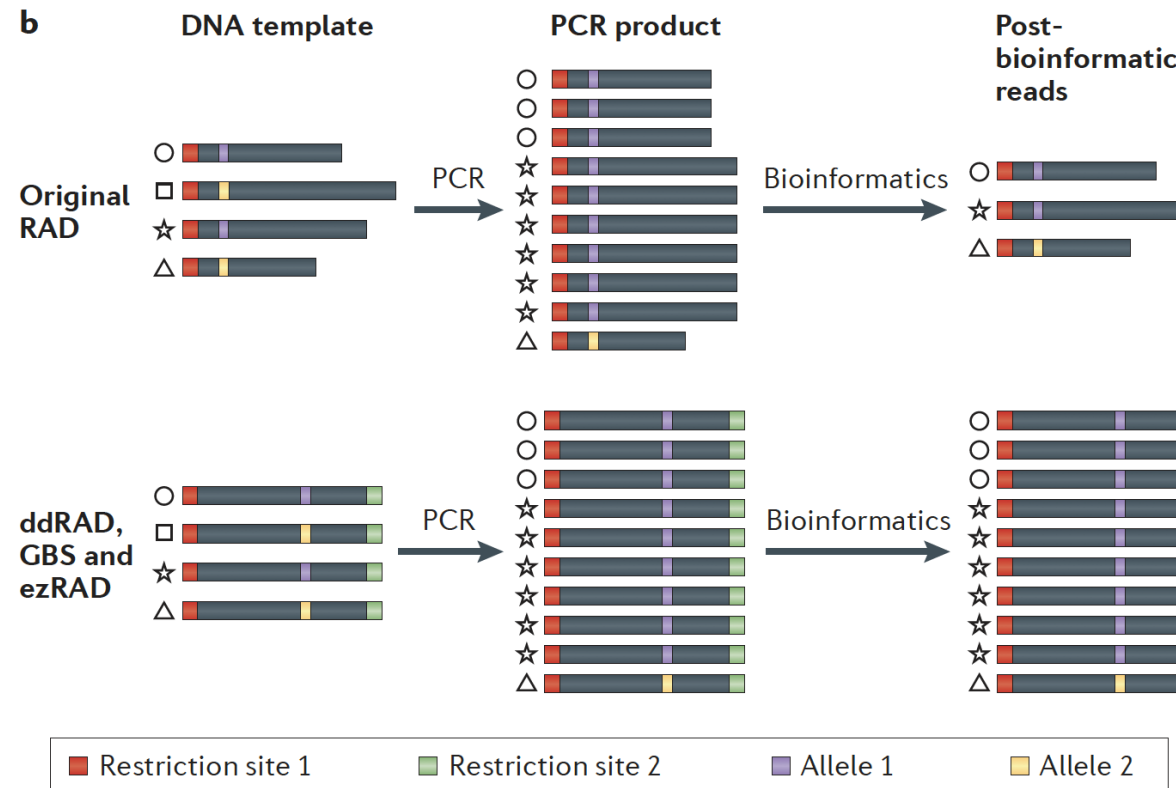
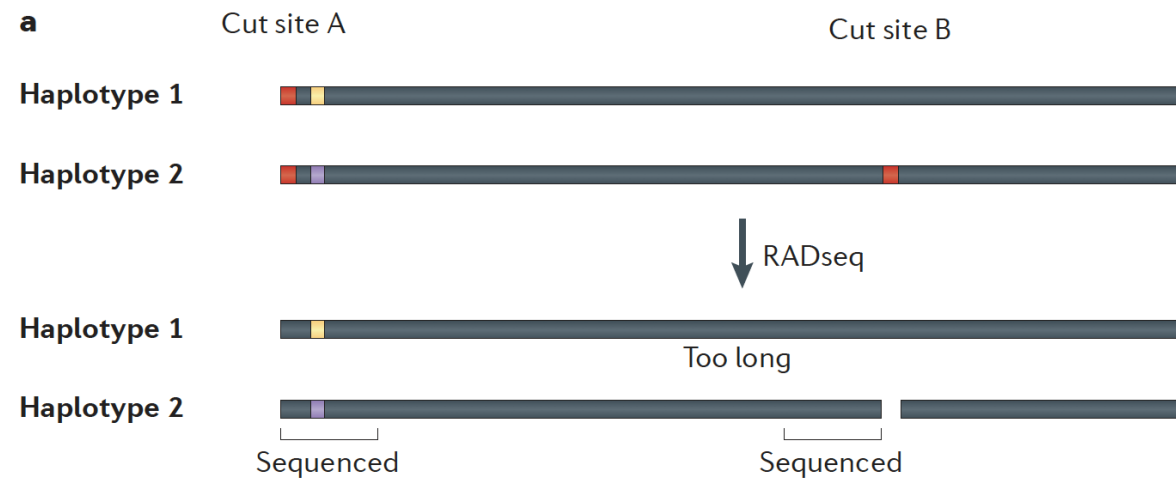
## RAD sequencing



B

## double digest RADseq





Single vs. double digestion



# Many different RADseq protocols

	mbRAD	ddRAD	ezRAD	2bRAD
Restriction cut sites per 10 kb*	~0.2–2.4	~3.7 × 10 <sup>-5</sup> –39	~39	~2.4
Postdigest fragment reduction	Size selection	Size selection	Size selection	Selective adapters
Contigs > 200 bp <sup>†</sup>	Yes	No	Some	No
Ability to blast/annotate <i>de novo</i> contigs	High	Mid	Mid	Low
Protocol complexity (# Steps) <sup>‡</sup>	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 <sup>§</sup>	Yes
ID of PCR duplicates	Yes	No	No <sup>§</sup>	No <sup>¶</sup>
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).

<sup>†</sup>When performing 100 bp reads such as on a HiSeq platform.

<sup>‡</sup>Not counting clean-up steps.

<sup>§</sup>ezRAD can be used with a PCR-free library preparation kit, thus removing the need to detect PCR duplicates.

<sup>¶</sup>2bRAD can detect PCR errors by mismatch among forward and reverse reads on individual strands.

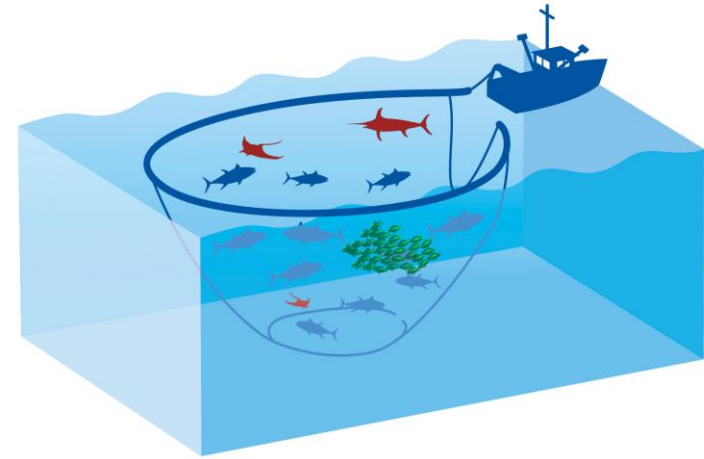
\*\*With some effort, the indexing for mbRAD can be modified to reduce the oligo counts to 22–37.

# 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



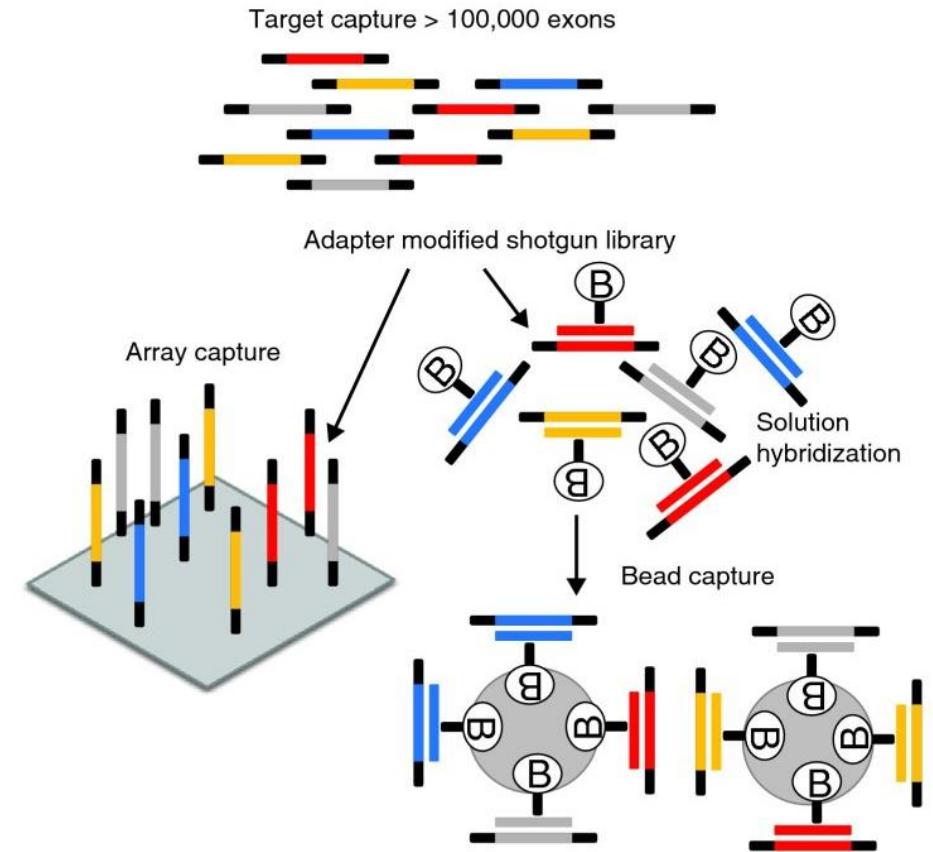
Targeted sequencing - RAPTURE



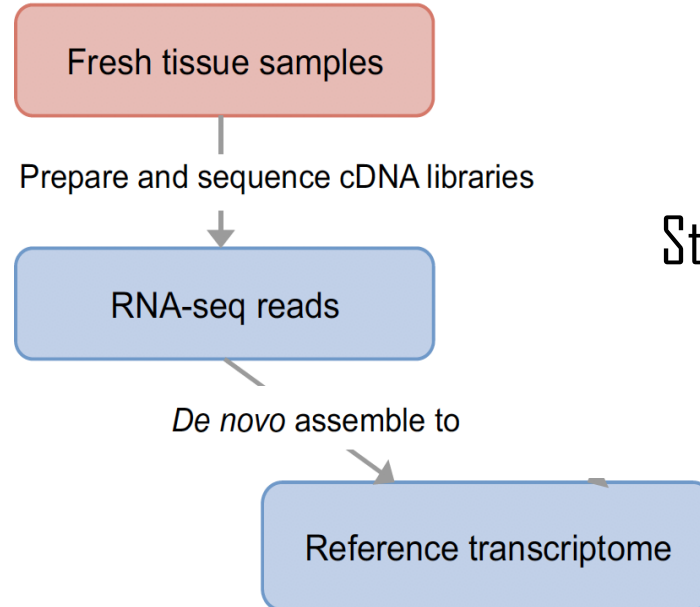
# Targeted approaches - Exome/exon capture

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.

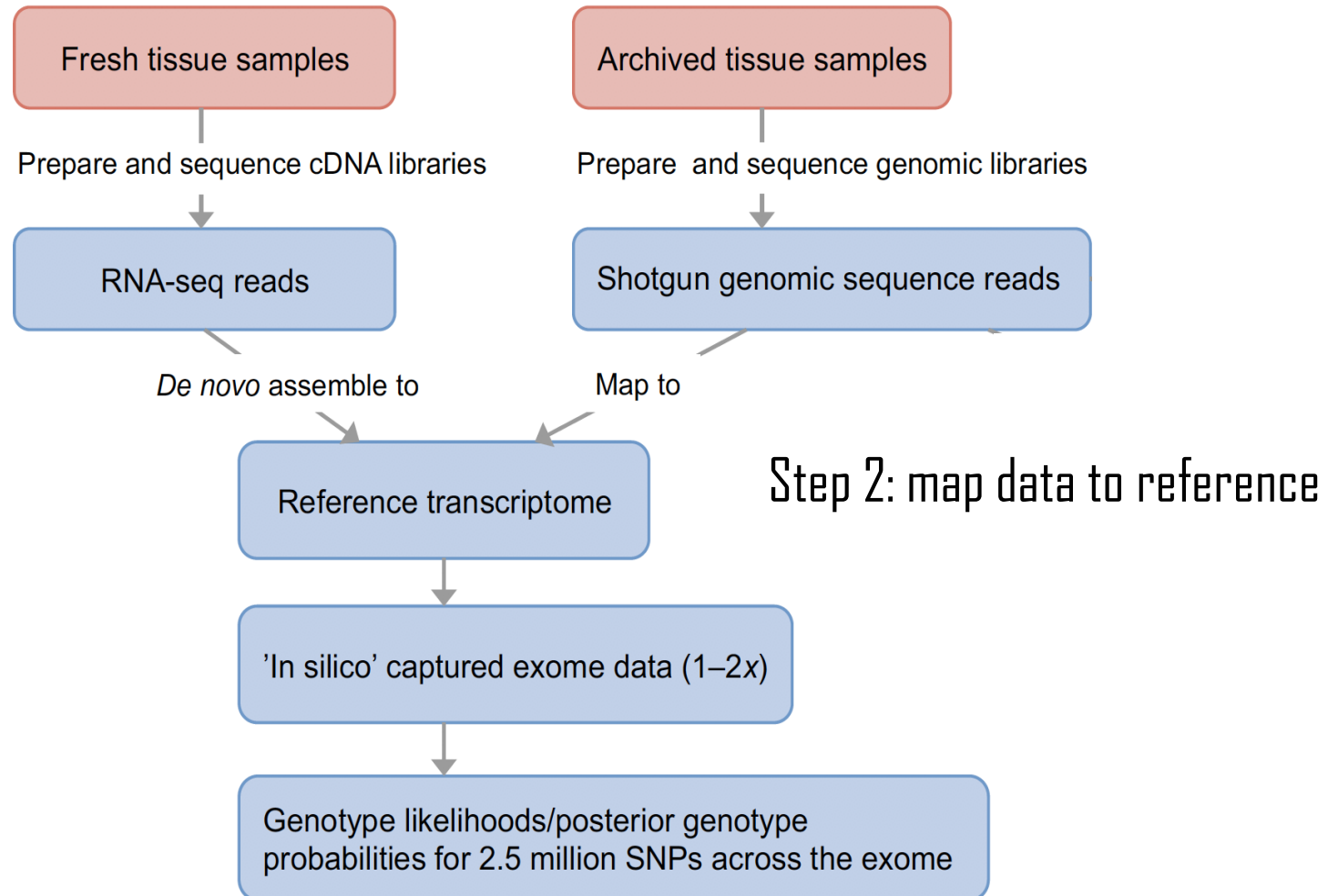


# Targeted approaches - Exome capture (*in silico*)

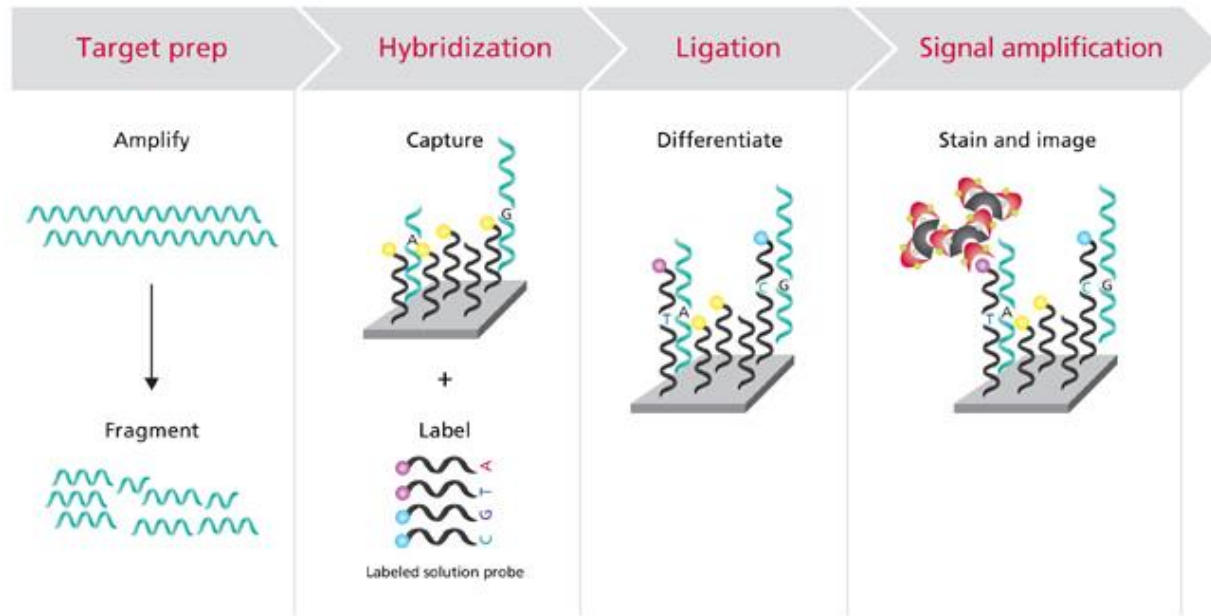


Step 1: build a reference transcriptome

# Targeted approaches - Exome capture (*in silico*)



# 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

# Whole genome resequencing

Short-read sequencing

illumina®

ion torrent  
⬇ \* △ ○ × □ + ≈

Long-read sequencing

pb PACBIO®

Oxford  
**NANOPORE**  
Technologies

Linked-reads technology  
(synthetic long reads)

10x GENOMICS  
???

Tell-seq  
Haplotagging  
Stlfr...

= short-reads + barcodes by molecules



# Whole genome resequencing

Short-read sequencing

illumina®

DNA

Random shearing

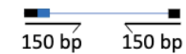
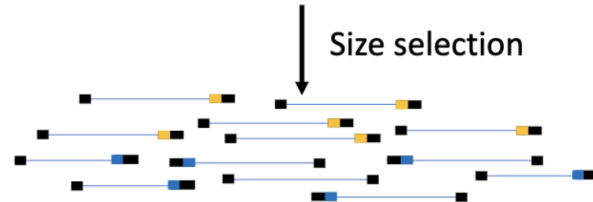
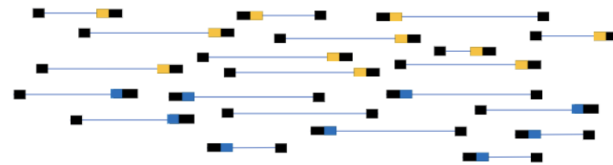
Illumina adapter ligation  
incl. individual index

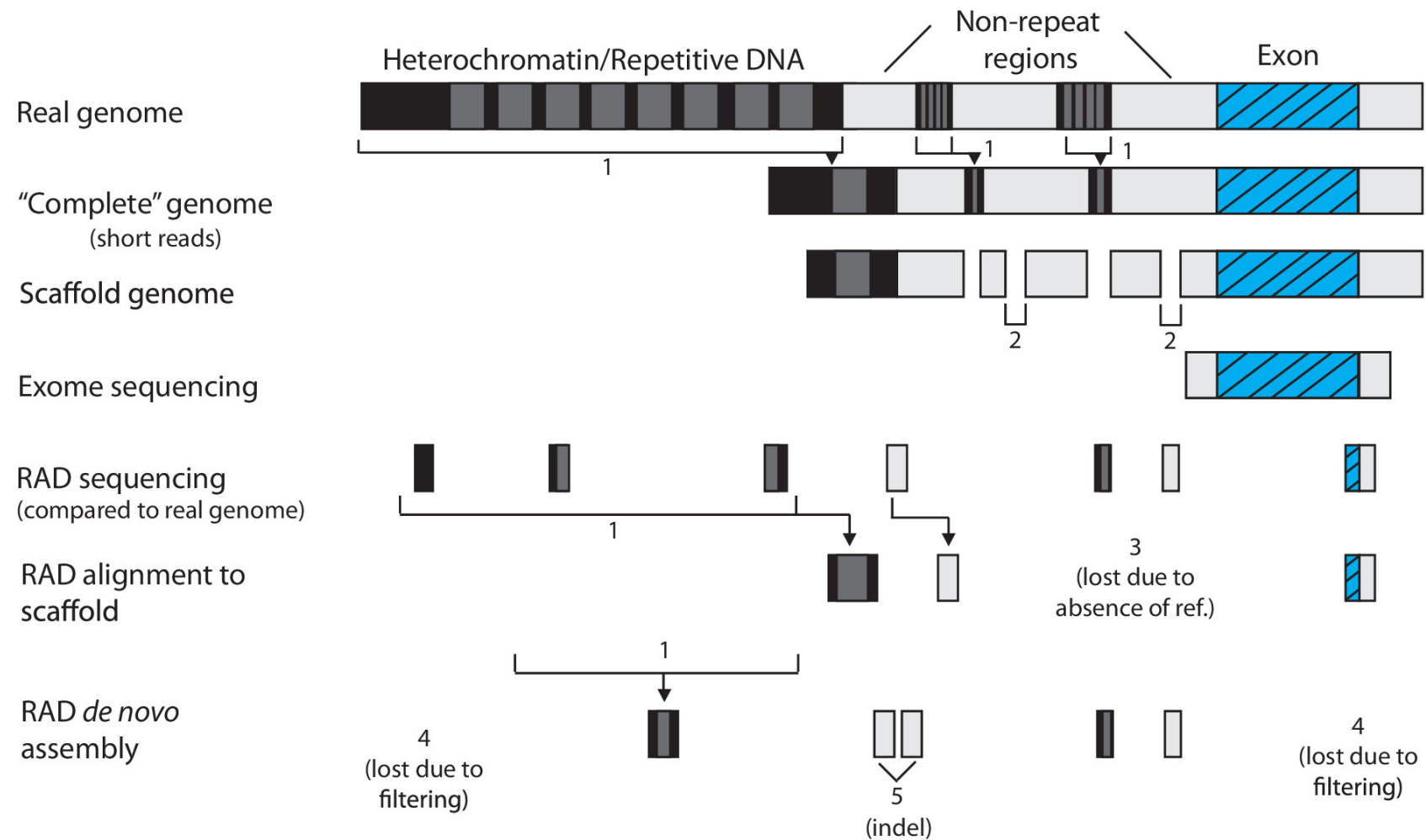
Size selection

paired-end sequencing

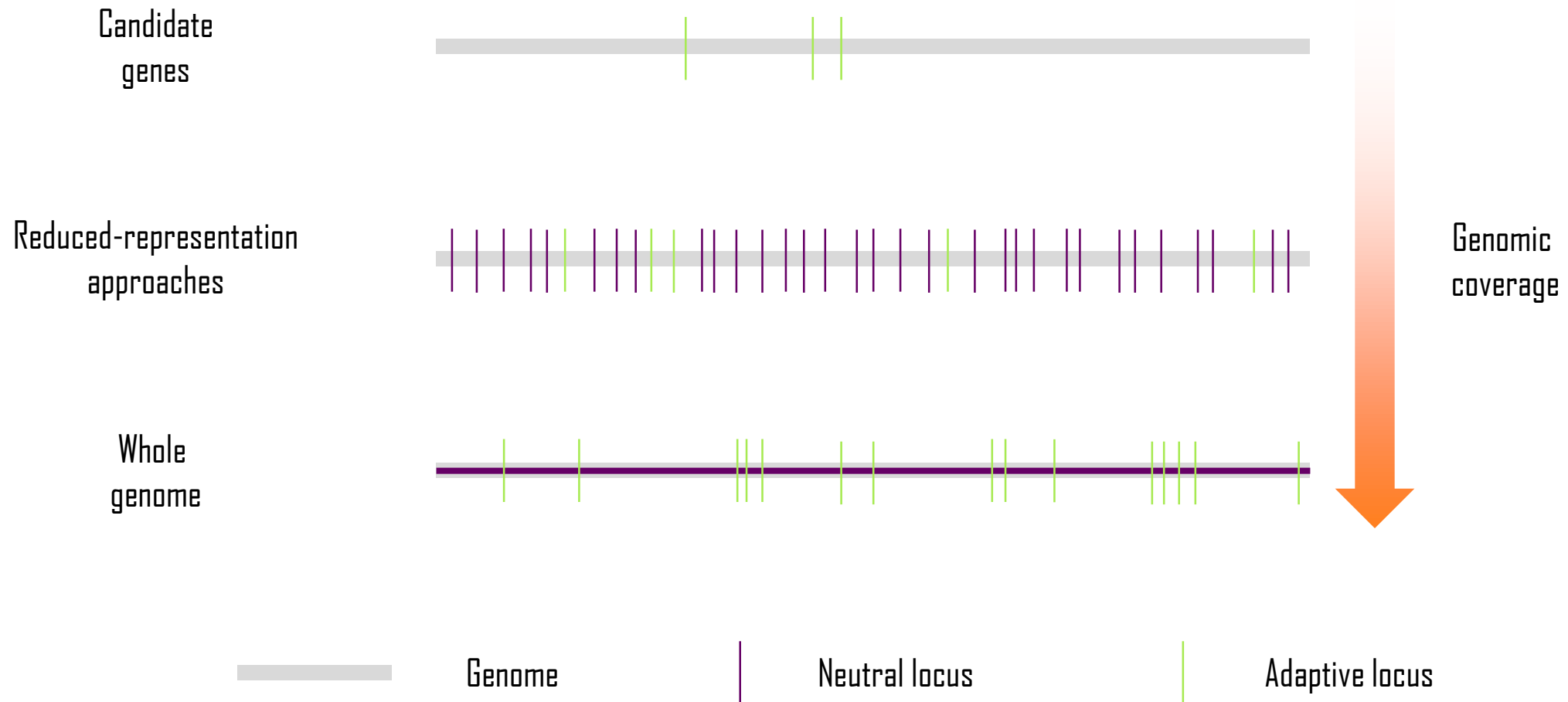
Up to 20 billion  
read pairs

150 bp 150 bp



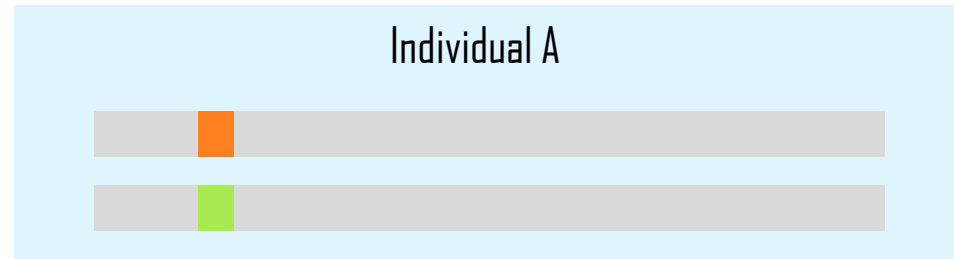


# Sequencing approaches



# Sequencing depth

= the number of reads covering a given site



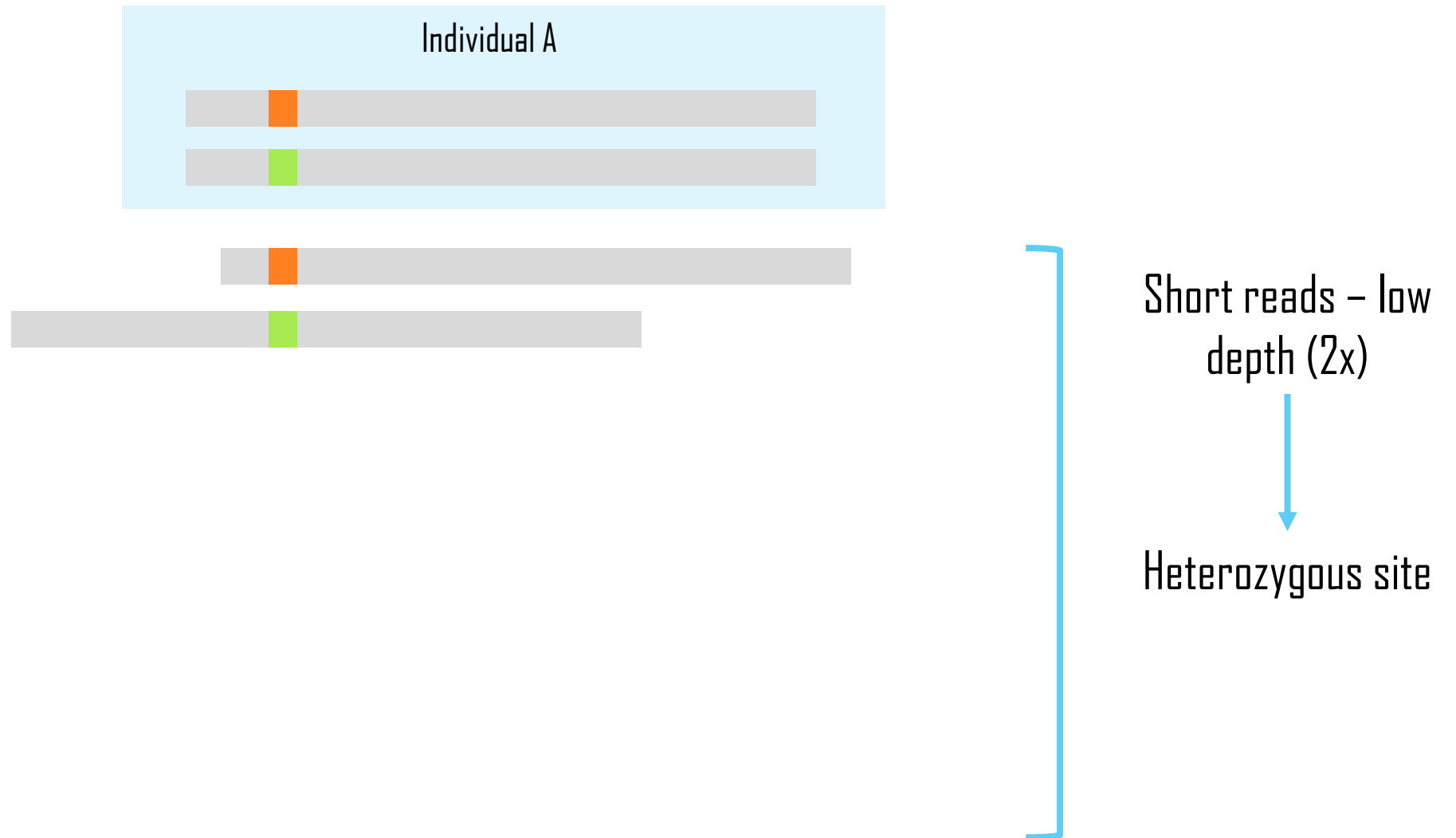
# Sequencing depth

= the number of reads covering a given site



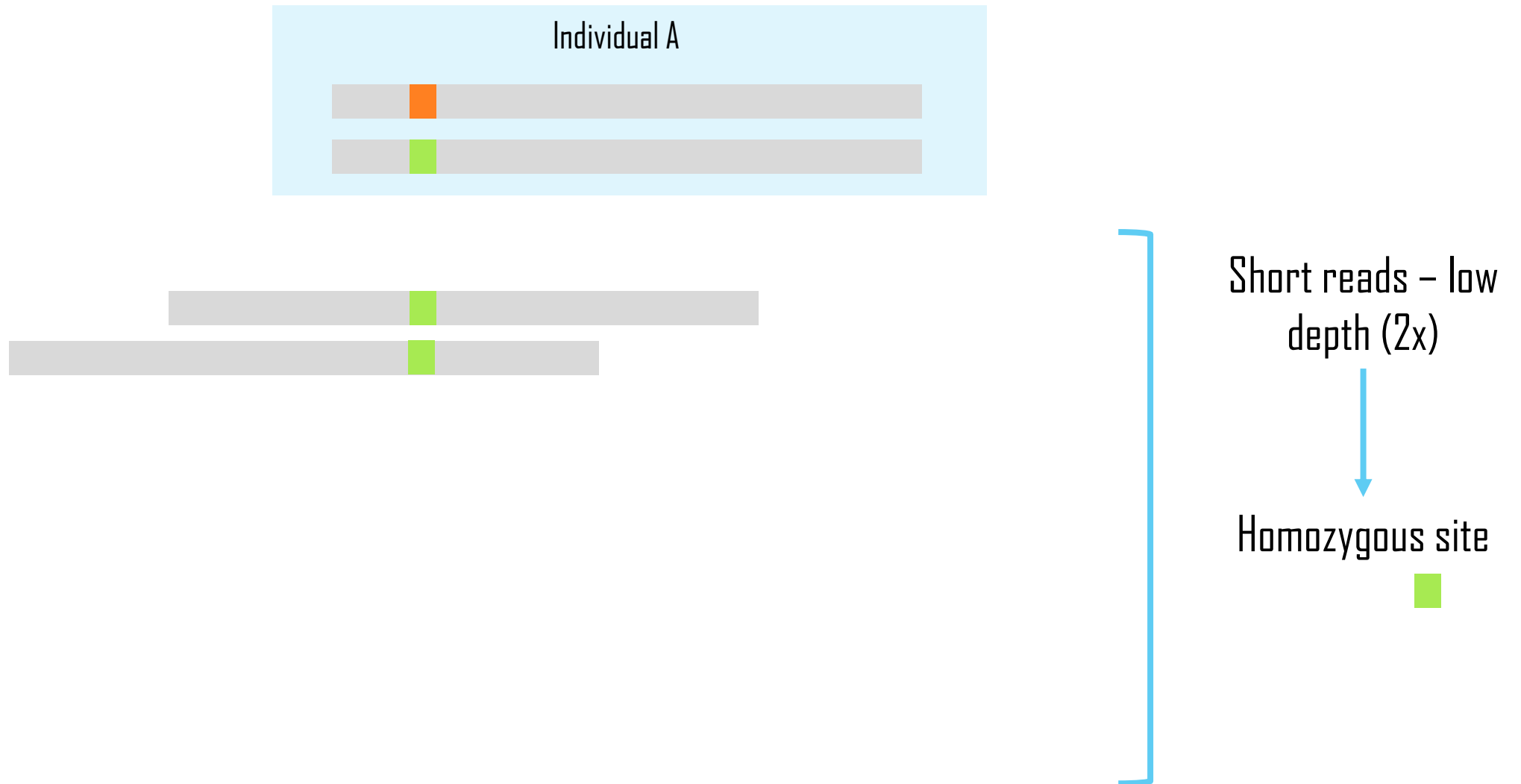
# Sequencing depth

= the number of reads covering a given site



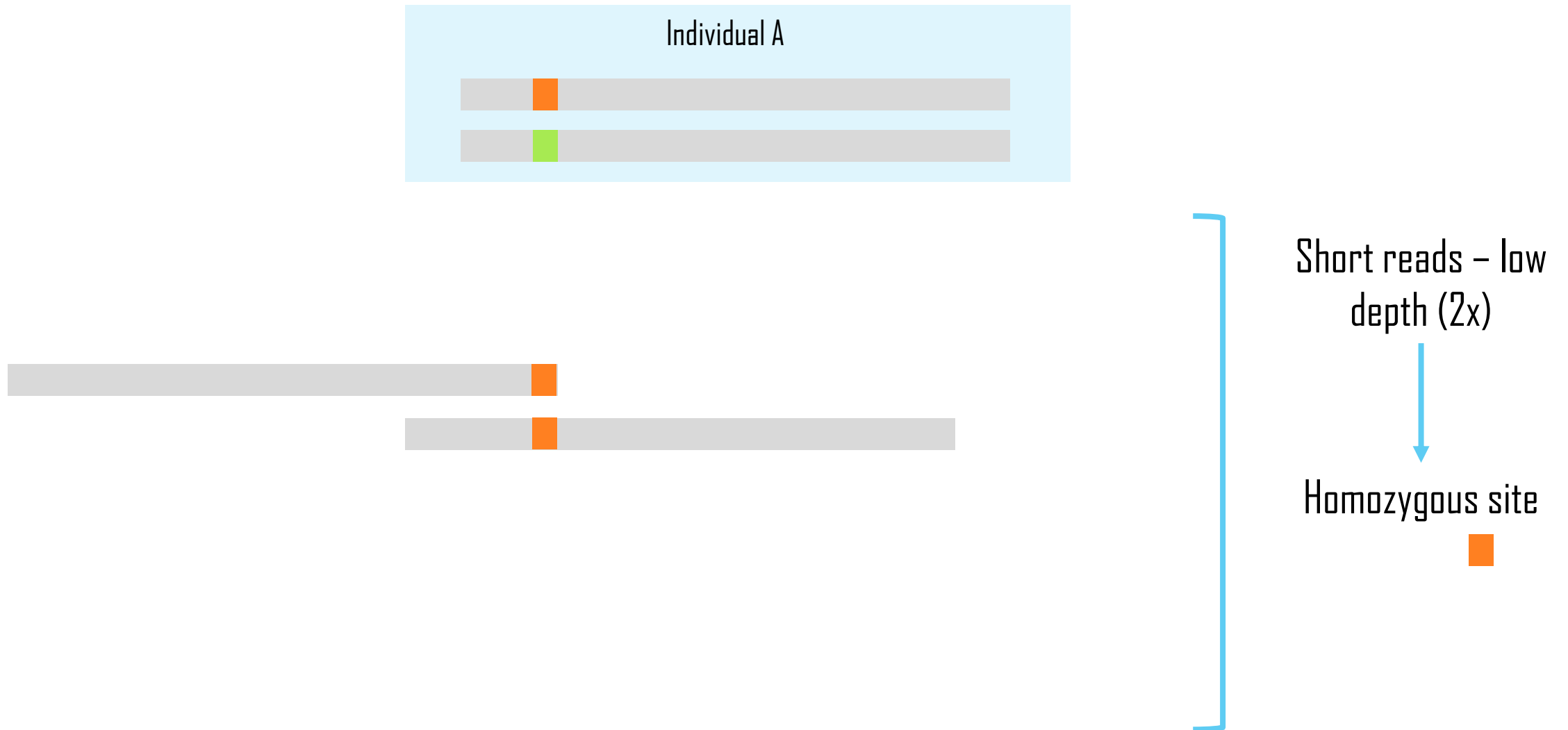
# Sequencing depth

= the number of reads covering a given site



# Sequencing depth

= the number of reads covering a given site



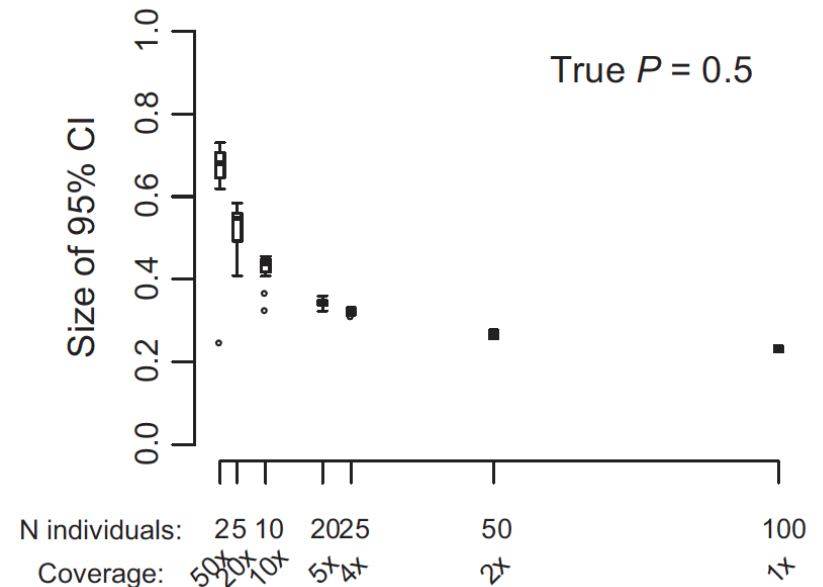
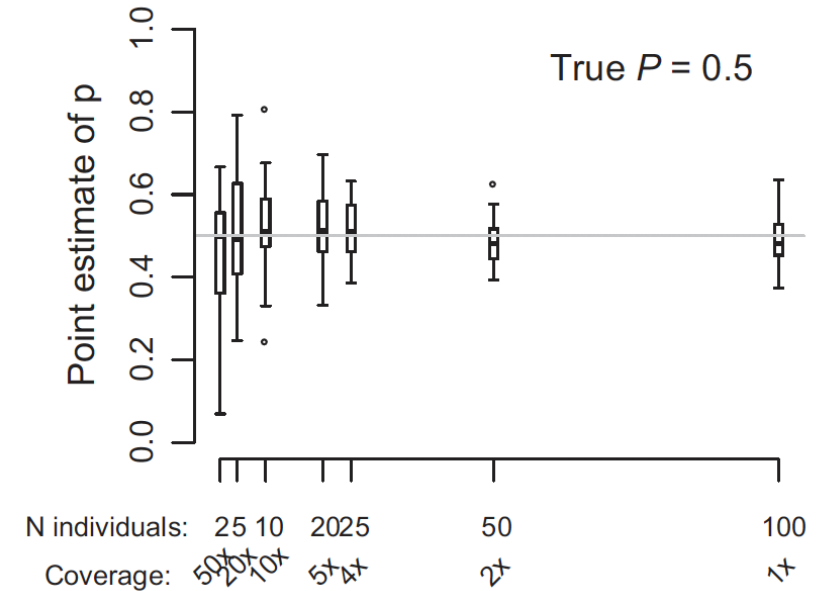


# Low-coverage sequencing

*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.



Application

Study design

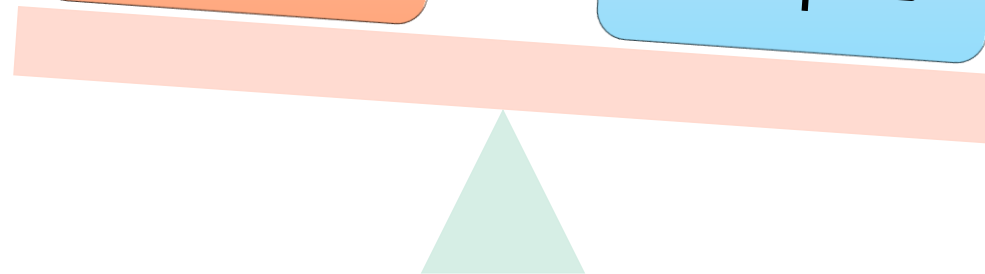
Question

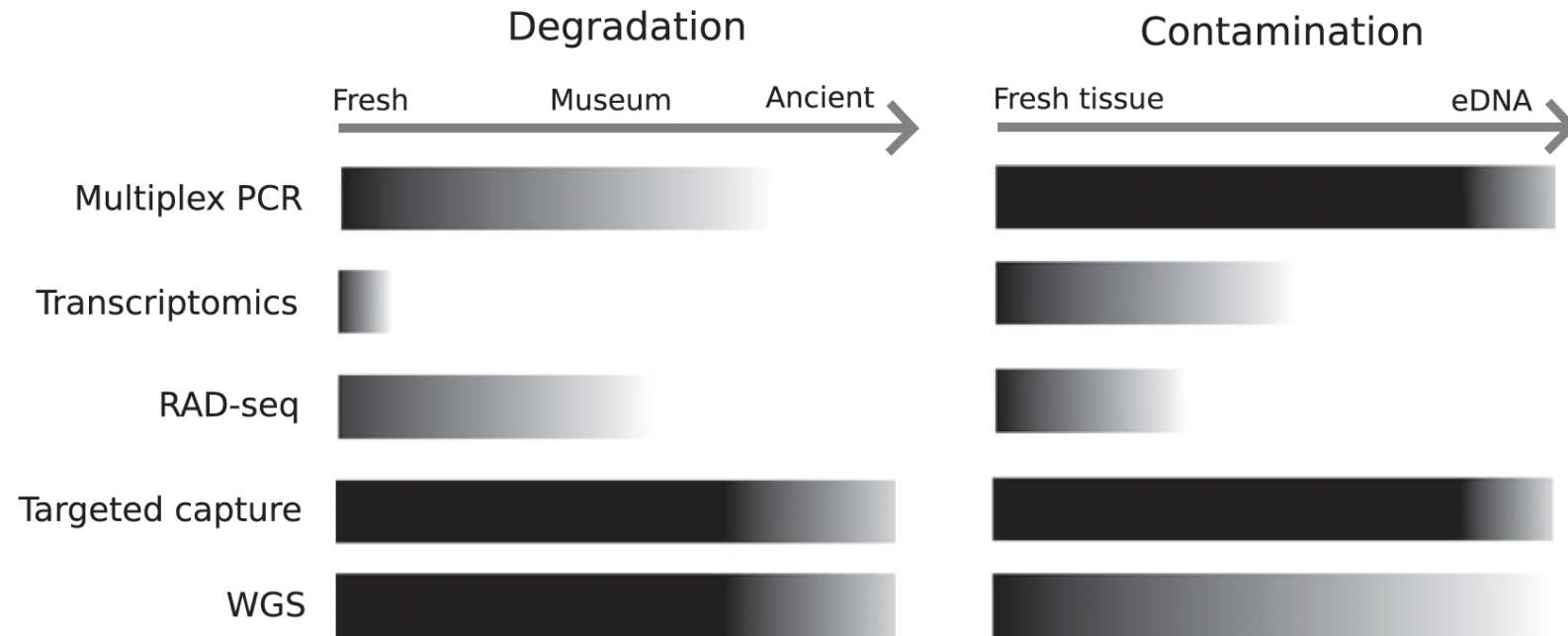
\$\$\$

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.