



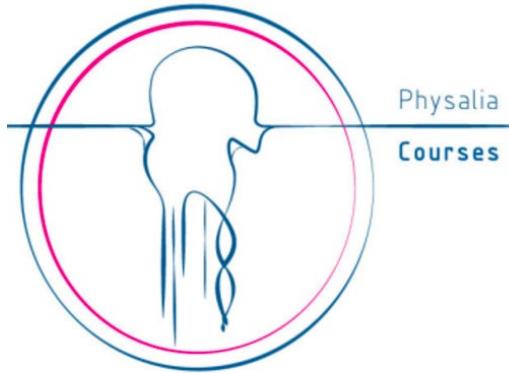
# How are you feeling today?

22 responses



eager to learn  
motivated  
**excited**  
**tired**   **tired too**  
curious  
energised  
need coffee  
optimistic  
enthusiastic  
pumped  
sleepy





# Adaptation Genomics Course

Mafalda Ferreria, PhD & Angela Fuentes Pardo, PhD

June 24 - 28, 2024

# Mafalda Ferreira, Ph.D

[mafalda.ferreira@zoologi.su.se](mailto:mafalda.ferreira@zoologi.su.se) | @MafaldaSFe

- Assistant Professor
- SciLifeLab Fellow



Stockholm  
University  SciLifeLab



- Genetic basis of local adaptation
- Evolution of adaptive traits across species
- Role of hybridization for adaptation
- Genetic architecture of adaptation (chromosomal rearrangements)
- Conservation genomics

# Angela Fuentes Pardo, PhD

[apfpgen@gmail.com](mailto:apfpgen@gmail.com) | @apfuentes7 | [fuentespardo.weebly.com](http://fuentespardo.weebly.com)

- Data steward / Data scientist
- Researcher



Atlantic herring  
(*Clupea harengus*)



Atlantic horse Mackerel  
(*Trachurus trachurus*)



Bean beetle  
(*Acanthocelides obtectus*)



- Population genomics
- Genetic basis of adaptation
- Conservation and fisheries genomics



UPPSALA  
UNIVERSITET

Postdoc



DALHOUSIE  
UNIVERSITY

PhD



BSc

# Course objectives

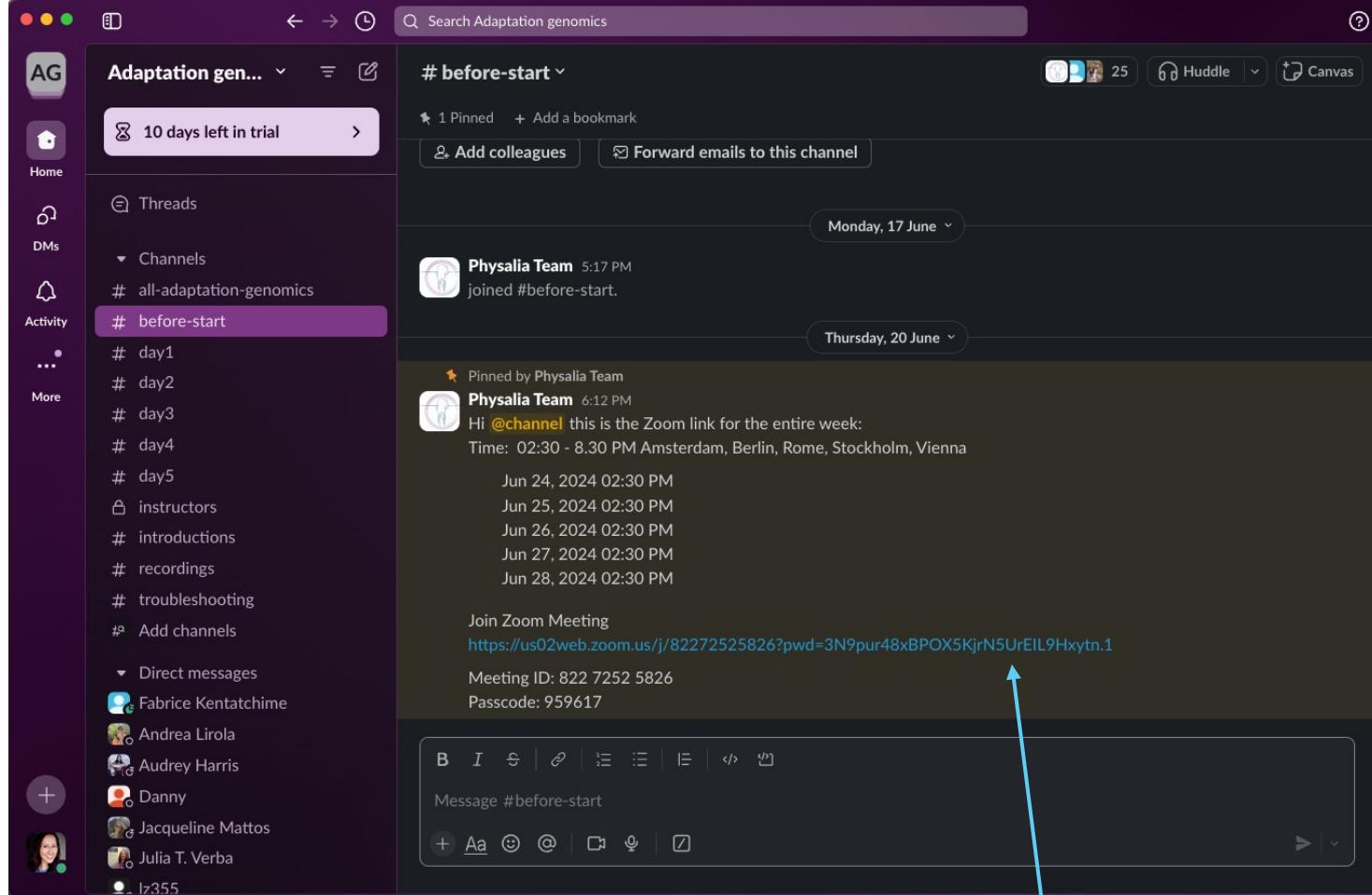
- Introduce **fundamental concepts** (theory, bioinformatics, sequencing and analytical methods)
- Provide **hands-on computational experience**
- Understand the **strengths and limitations** of various approaches
- Foster **critical thinking** to be applied on your research

# Schedule

| CET<br>(Central Europe) | PST<br>(West America) | Monday<br>(Day 1)   | Tuesday<br>(Day 2) | Wednesday<br>(Day 3) | Thursday<br>(Day 4) | Friday<br>(Day 5)  |
|-------------------------|-----------------------|---------------------|--------------------|----------------------|---------------------|--------------------|
| <b>14:30 - 15:30</b>    | <b>5:30 - 6:30</b>    | <b>Introduction</b> | <b>Lecture 2-1</b> | <b>Lecture 3-1</b>   | <b>Lecture 4-1</b>  | <b>Lecture 5-1</b> |
| 15:30 - 16:30           | 6:30 - 7:30           | Tutorial 1          | Tutorial 2         | Tutorial 3           | Tutorial 4          | Tutorial 5         |
| 16:30 - 17:00           | 7:30 - 8:00           | <i>Break</i>        | <i>Break</i>       | <i>Break</i>         | <i>Break</i>        | <i>Break</i>       |
| <b>17:00 - 18:00</b>    | <b>8:00 - 9:00</b>    | <b>Lecture 1-2</b>  | <b>Lecture 2-2</b> | <b>Lecture 3-2</b>   | Tutorial 4          | <b>Wrap-up</b>     |
| 18:00 - 20:30           | 9:00 - 11:30          | Tutorial 2          | Tutorial 2         | Tutorial 3           | <b>Open lab</b>     | <b>Q&amp;A</b>     |

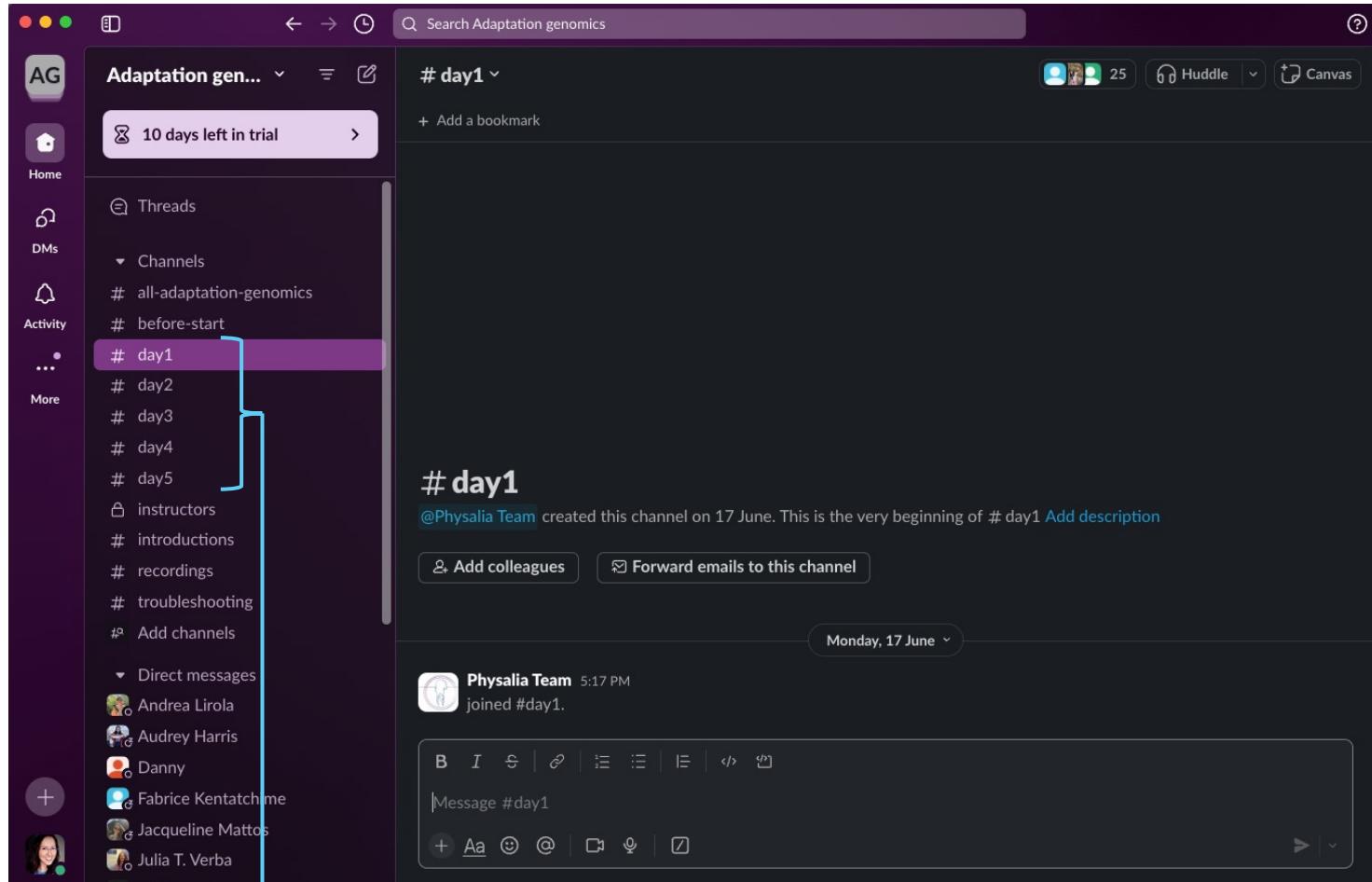
- 5 days. Each day, two lectures followed by self-guided practical exercises
- Let's make them interactive, don't hesitate to ask questions! 😊  
(live questions during lectures or written in the dedicated Slack channel)
- The instructors will be available for questions and support over the duration of the course

# Slack



Zoom link for the course

# Slack



Post your question in the  
dedicated Slack channel

# Schedule

| CET<br>(Central Europe) | PST<br>(West America) | Monday<br>(Day 1)   | Tuesday<br>(Day 2) | Wednesday<br>(Day 3) | Thursday<br>(Day 4) | Friday<br>(Day 5)  |
|-------------------------|-----------------------|---------------------|--------------------|----------------------|---------------------|--------------------|
| <b>14:30 - 15:30</b>    | <b>5:30 - 6:30</b>    | <b>Introduction</b> | <b>Lecture 2-1</b> | <b>Lecture 3-1</b>   | <b>Lecture 4-1</b>  | <b>Lecture 5-1</b> |
| 15:30 - 16:30           | 6:30 - 7:30           | Tutorial 1          | Tutorial 2         | Tutorial 3           | Tutorial 4          | Tutorial 5         |
| 16:30 - 17:00           | 7:30 - 8:00           | <i>Break</i>        | <i>Break</i>       | <i>Break</i>         | <i>Break</i>        | <i>Break</i>       |
| <b>17:00 - 18:00</b>    | <b>8:00 - 9:00</b>    | <b>Lecture 1-2</b>  | <b>Lecture 2-2</b> | <b>Lecture 3-2</b>   | Tutorial 4          | <b>Wrap-up</b>     |
| 18:00 - 20:30           | 9:00 - 11:30          | Tutorial 2          | Tutorial 2         | Tutorial 3           | <b>Open lab</b>     | <b>Q&amp;A</b>     |

- Work on the tutorials at your own pace, and when it is most convenient for you
- Don't worry if you can't finish some exercises
  - We provide the input files for each tutorial, so they can be run independently

# Schedule

| CET<br>(Central Europe) | PST<br>(West America) | Monday<br>(Day 1)   | Tuesday<br>(Day 2) | Wednesday<br>(Day 3) | Thursday<br>(Day 4) | Friday<br>(Day 5)  |
|-------------------------|-----------------------|---------------------|--------------------|----------------------|---------------------|--------------------|
| <b>14:30 - 15:30</b>    | <b>5:30 - 6:30</b>    | <b>Introduction</b> | <b>Lecture 2-1</b> | <b>Lecture 3-1</b>   | <b>Lecture 4-1</b>  | <b>Lecture 5-1</b> |
| 15:30 - 16:30           | 6:30 - 7:30           | Tutorial 1          | Tutorial 2         | Tutorial 3           | Tutorial 4          | Tutorial 5         |
| 16:30 - 17:00           | 7:30 - 8:00           | <i>Break</i>        | <i>Break</i>       | <i>Break</i>         | <i>Break</i>        | <i>Break</i>       |
| <b>17:00 - 18:00</b>    | <b>8:00 - 9:00</b>    | <b>Lecture 1-2</b>  | <b>Lecture 2-2</b> | <b>Lecture 3-2</b>   | Tutorial 4          | <b>Wrap-up</b>     |
| 18:00 - 20:30           | 9:00 - 11:30          | Tutorial 2          | Tutorial 2         | Tutorial 3           | <b>Open lab</b>     | <b>Q&amp;A</b>     |

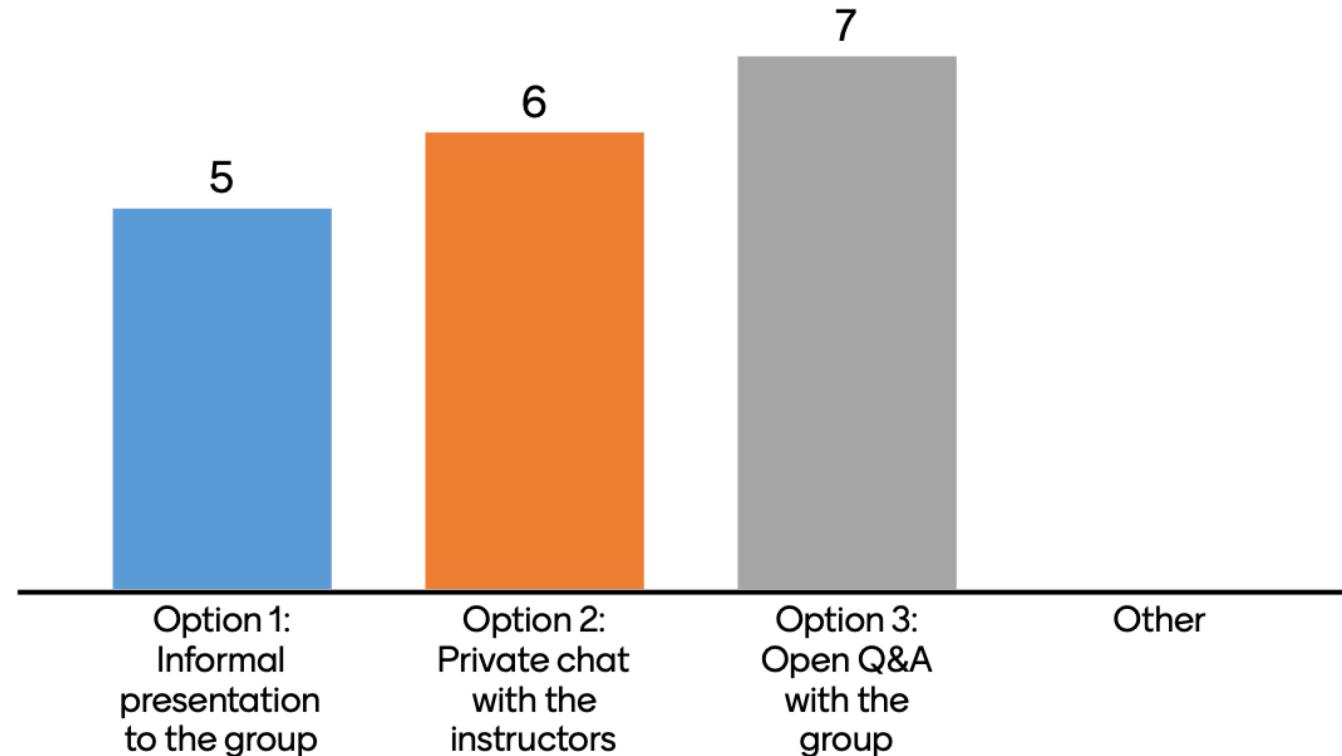
- Work on the tutorials at your own pace, and when it is most convenient for you
- Don't worry if you can't finish some exercises
  - We provide the input files for each tutorial, so they can be run independently
- Open lab, time to complete exercises or try optional tutorials
- You will have access to the AWS server from 12:30 to 22:30 (CET)
  - If you need a time extension, please contact Carlo

# Schedule

| CET<br>(Central Europe) | PST<br>(West America) | Monday<br>(Day 1)   | Tuesday<br>(Day 2) | Wednesday<br>(Day 3) | Thursday<br>(Day 4) | Friday<br>(Day 5)  |
|-------------------------|-----------------------|---------------------|--------------------|----------------------|---------------------|--------------------|
| <b>14:30 - 15:30</b>    | <b>5:30 - 6:30</b>    | <b>Introduction</b> | <b>Lecture 2-1</b> | <b>Lecture 3-1</b>   | <b>Lecture 4-1</b>  | <b>Lecture 5-1</b> |
| 15:30 - 16:30           | 6:30 - 7:30           | Tutorial 1          | Tutorial 2         | Tutorial 3           | Tutorial 4          | Tutorial 5         |
| 16:30 - 17:00           | 7:30 - 8:00           | <i>Break</i>        | <i>Break</i>       | <i>Break</i>         | <i>Break</i>        | <i>Break</i>       |
| <b>17:00 - 18:00</b>    | <b>8:00 - 9:00</b>    | <b>Lecture 1-2</b>  | <b>Lecture 2-2</b> | <b>Lecture 3-2</b>   | Tutorial 4          | <b>Wrap-up</b>     |
| 18:00 - 20:30           | 9:00 - 11:30          | Tutorial 2          | Tutorial 2         | Tutorial 3           | <b>Open lab</b>     | <b>Q&amp;A</b>     |

- Wrap-up of tutorials
- Q&A
  - Option 1: Students' informal presentations:
    - Max. 3 minutes (intro to the study, faced **obstacles/questions** and **how you are trying to solve them**)
    - 4 min feedback form the group
  - Option 2: Private discussion with instructors (7 min each)
  - Option 3: Open questions and answers with the group

# What do you like to do during the Q&A session?



# Outline of the course

## Day 1

Introduction to adaptation genomics

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

## Practical

Outlier loci 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)

Open Lab

# Outline of the course

## Day 5

Diversity of methods to study the genomics of adaptation  
Validation of candidate loci

## Practical

Functional annotation of candidate loci for adaptation  
Wrap-up of tutorials  
Q & A

# GitHub page of the course

[https://github.com/MafaldaSFerreira/physalia\\_adaptation\\_course-2024](https://github.com/MafaldaSFerreira/physalia_adaptation_course-2024)

The screenshot shows the GitHub repository page for 'physalia\_adaptation\_course-2024'. The repository is public and owned by 'MafaldaSFerreira'. The main interface includes a navigation bar with links for Product, Solutions, Open Source, Enterprise, Pricing, a search bar, and sign-in options. Below the header, there are tabs for Code, Issues, Pull requests, Actions, Projects, Security, and Insights. The 'Code' tab is selected, showing the 'main' branch with 26 commits. A detailed commit history is listed, showing changes made to various files and folders over the past week. To the right of the code area, sections for About, Releases, Packages, and Contributors are displayed. The 'About' section notes 'No description, website, or topics provided.' The 'Contributors' section lists two individuals: 'MafaldaSFerreira' (Mafalda S. Ferreira) and 'apfuentes' (Angela P. Fuentes-Pardo).

MafaldaSFerreira / physalia\_adaptation\_course-2024 Public

Code Issues Pull requests Actions Projects Security Insights

main 2 Branches 0 Tags Go to file

MafaldaSFerreira changes 28a068c · 2 hours ago 26 Commits

| File/Folder                                | Change               | Time        |
|--|----------------------|-------------|
| 00_before_the_course                       | update main          | 2 days ago  |
| 00_documents                               | update popfiles day1 | last week   |
| 01_day1                                    | changes              | 2 hours ago |
| 02_day2                                    | changes              | 2 hours ago |
| 03_day3                                    | corrections          | 5 days ago  |
| 04_day4                                    | updates main         | 2 days ago  |
| 05_day5                                    | changes              | 2 days ago  |
| images                                     | changes              | 2 hours ago |
| lectures                                   | clean up             | last week   |
| .DS_Store                                  | modified day04       | 2 days ago  |
| .gitignore                                 | fixes 01_day1        | last week   |
| Connection_to_the_Amazon_EC2_service_20... | update               | last week   |
| README.md                                  | changes              | 2 hours ago |

About

No description, website, or topics provided.

Readme

Activity

0 stars

2 watching

0 forks

Report repository

Releases

No releases published

Packages

No packages published

Contributors 2

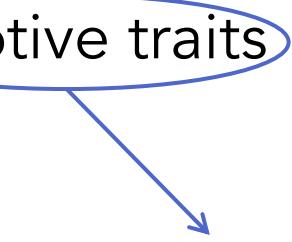
MafaldaSFerreira Mafalda S. Ferreira

apfuentes Angela P. Fuentes-Pardo

# Adaptation genomics

# Adaptation genomics

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



a characteristic of an organism  
that **enhances** its survival or  
**reproductive success** in a  
particular **environment**

# Adaptation genomics

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

- **Genetic basis of traits** = loci that control the 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 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 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 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 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 is to understand  
the **genomic basis** and  
**architecture** of adaptive traits

**Ecology**

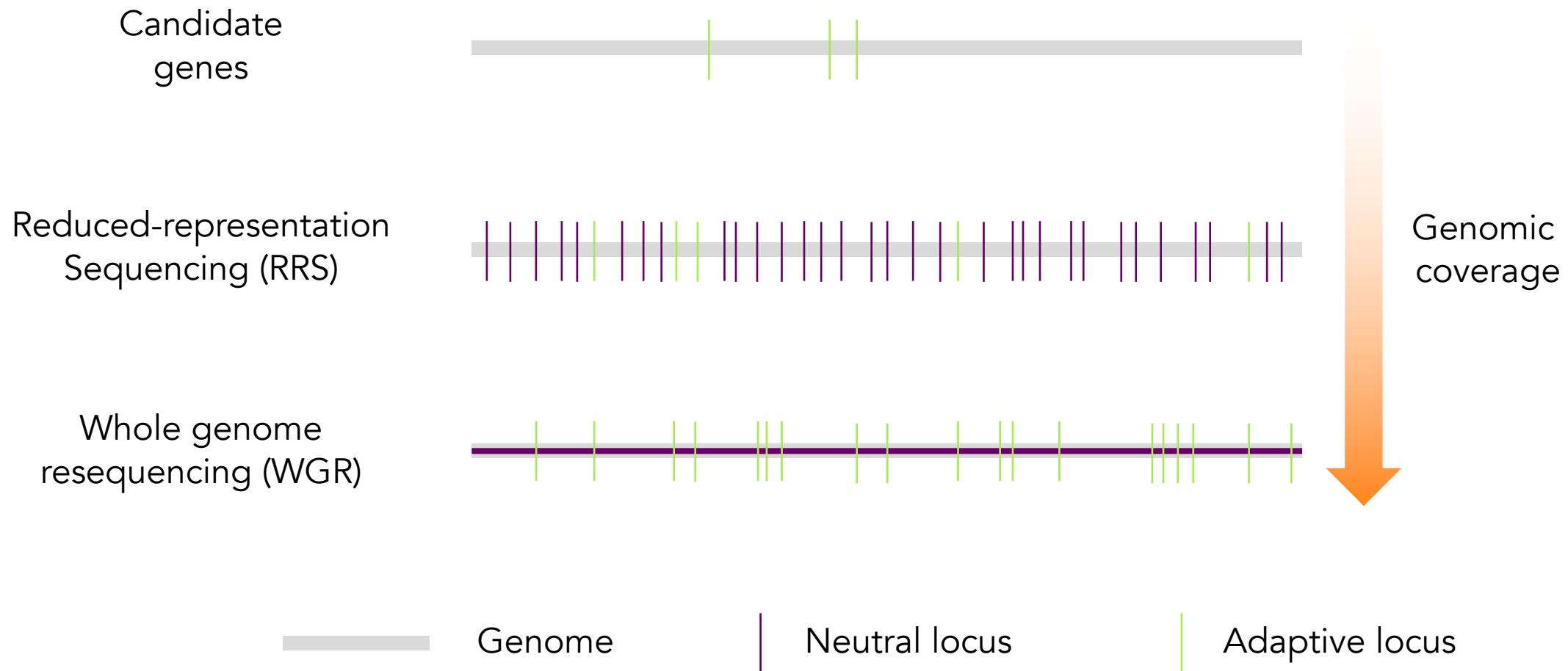
**Evolution**

**Conservation**

**Management**

Physiology, molecular evolution,  
biodiversity, speciation...

# Sequencing approaches



Genetics

Fraction of genome

Sanger  
sequencing

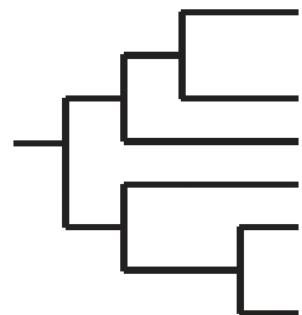
Genomics

Whole genome  
re-sequencing

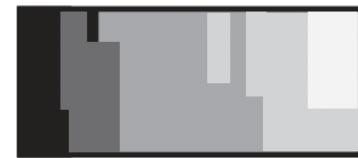
ddRAD

RADtag (Baird 2008)

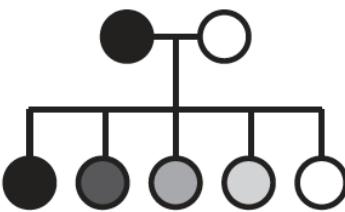
Phylogeny



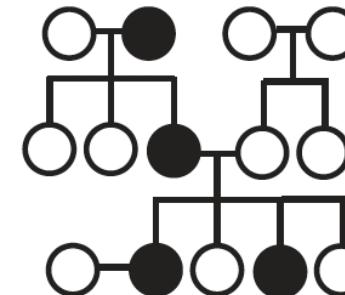
Population Structure



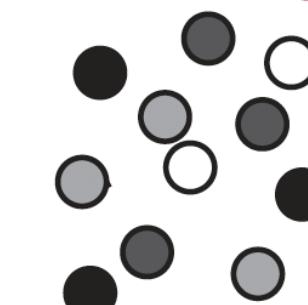
QTL  
Mapping



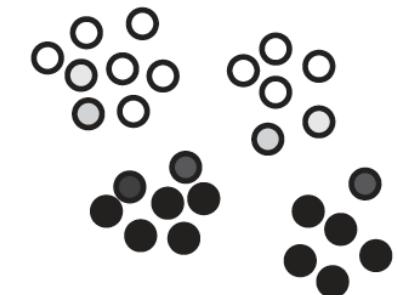
Pedigree  
Mapping



Association  
Mapping



Population  
Genomic Scans



Divergence limited

Recombination limited

Linkage Diseq. limited

# Reduced-representation approaches

- RADseq/GBS
  - Exome/exon capture
  - SNP chip
- 
- ]
- ]

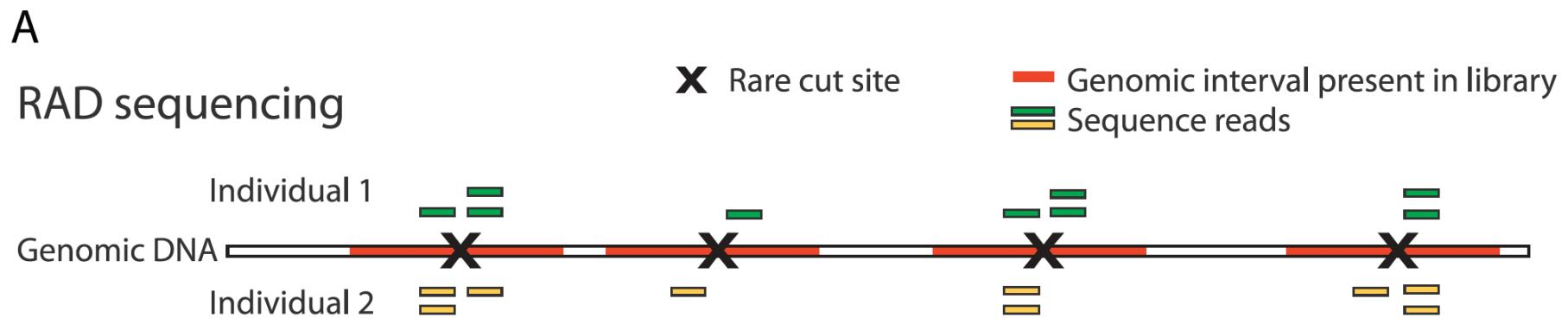
Random sampling of the genome

Targeted capture of loci of interest

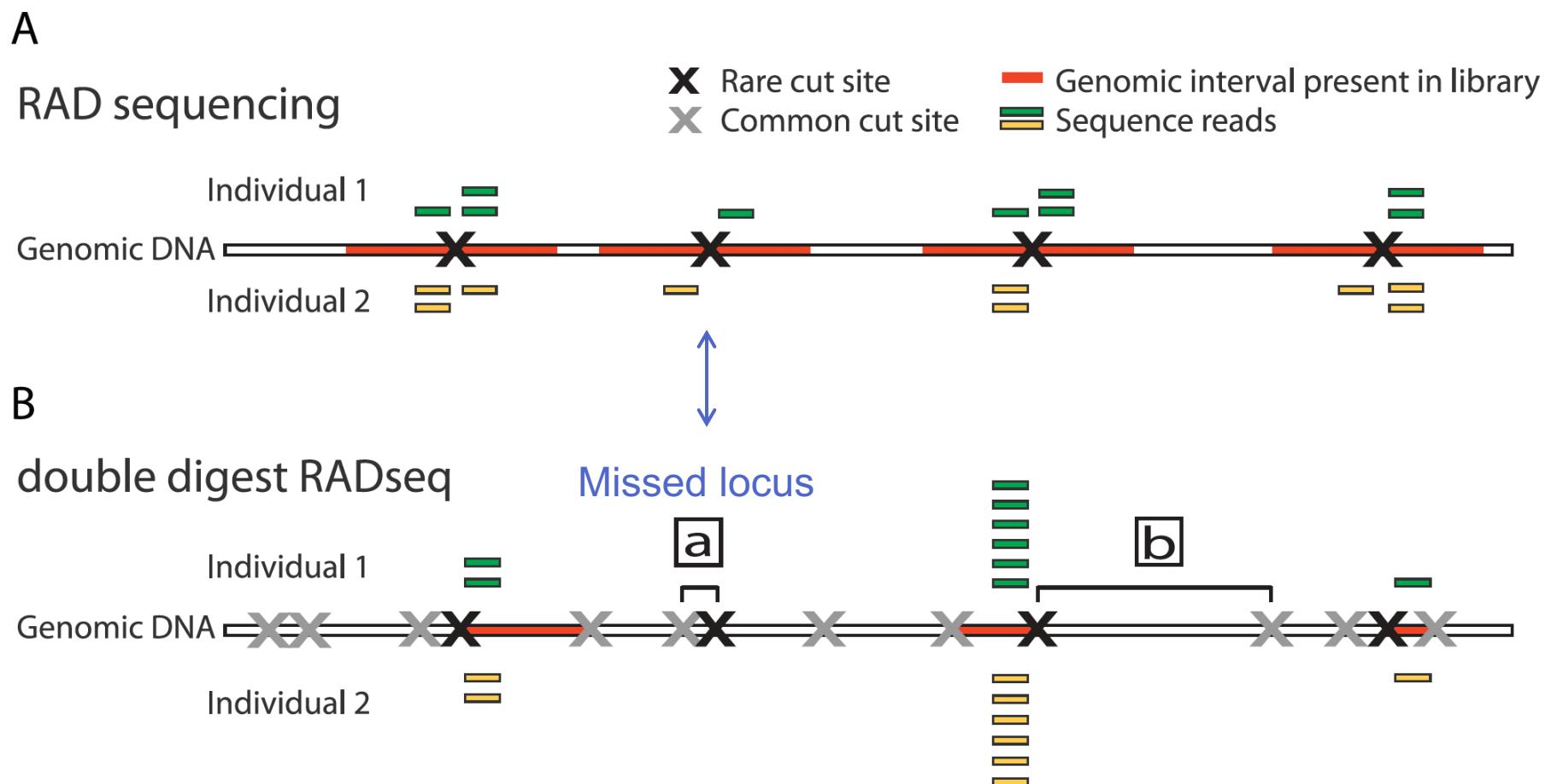
RADseq : Restriction-site associated DNA sequencing

SNP : Single nucleotide polymorphism

# Random approaches : RADseq



# Random approaches : RADseq



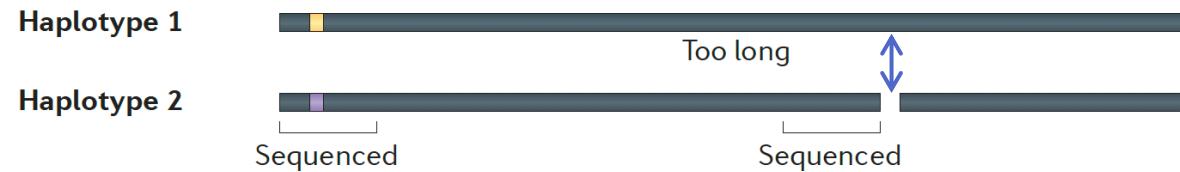
**a**

Cut site A

Cut site B



RADseq

**Single vs. double digestion****Allele dropout**

Restriction site 1

Restriction site 2

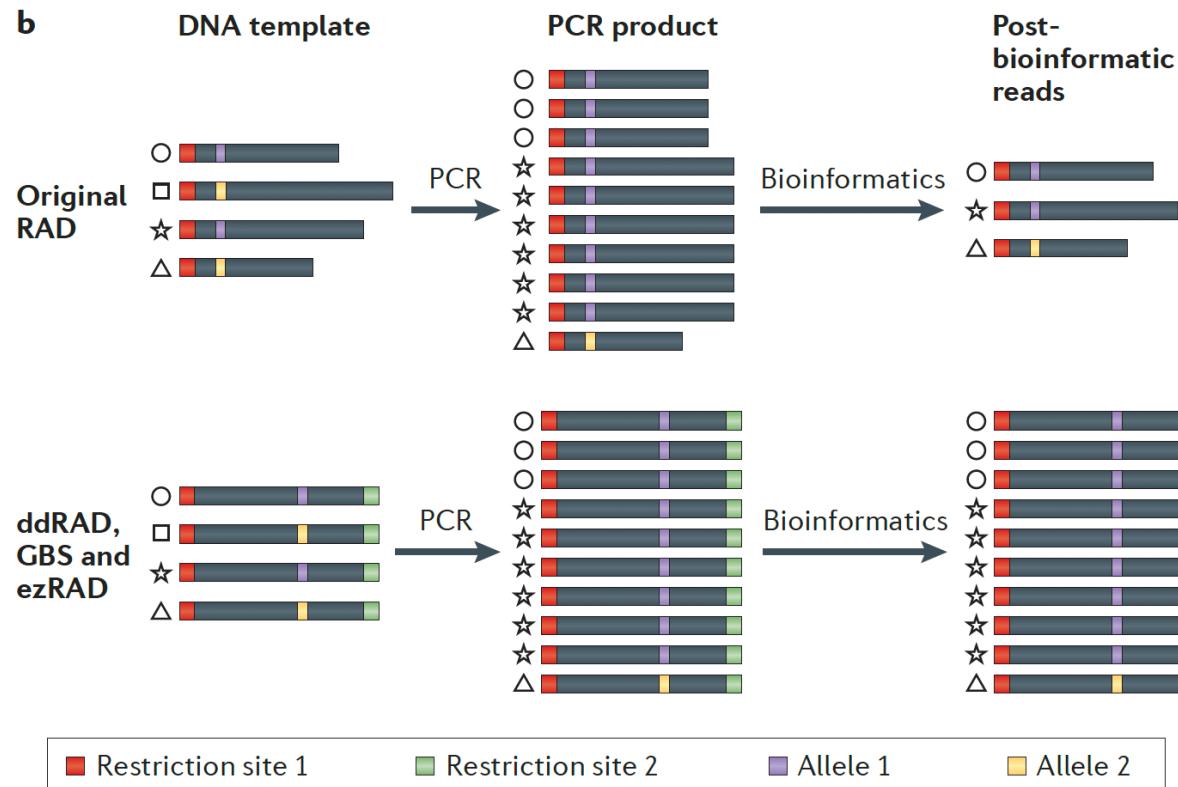
Allele 1

Allele 2

Andrews et al. 2016, Nature Gen. Rev.

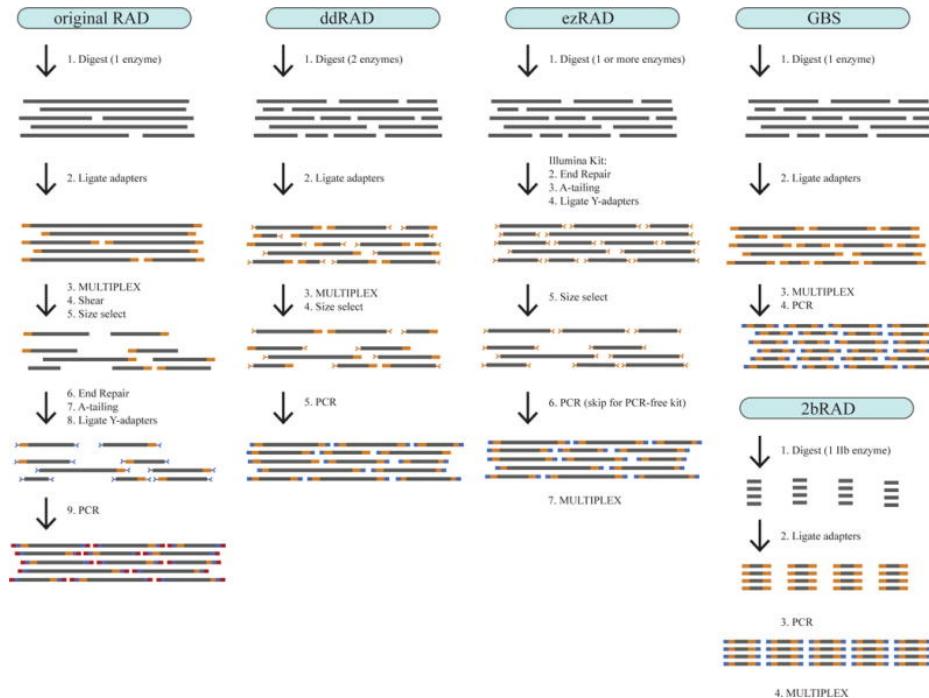


# Single vs. double digestion



- Target coverage may vary significantly between samples and methods

# Many different RADseq protocols ("flavors")



|   | 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.

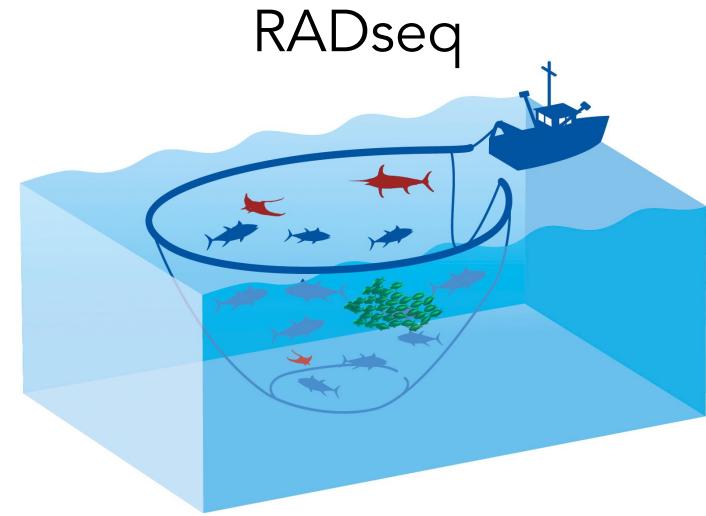
<sup>¶</sup>2bRAD can detect PCR errors by mismatch among forward and reverse reads on individual strands.

\*\*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 sequencing approaches

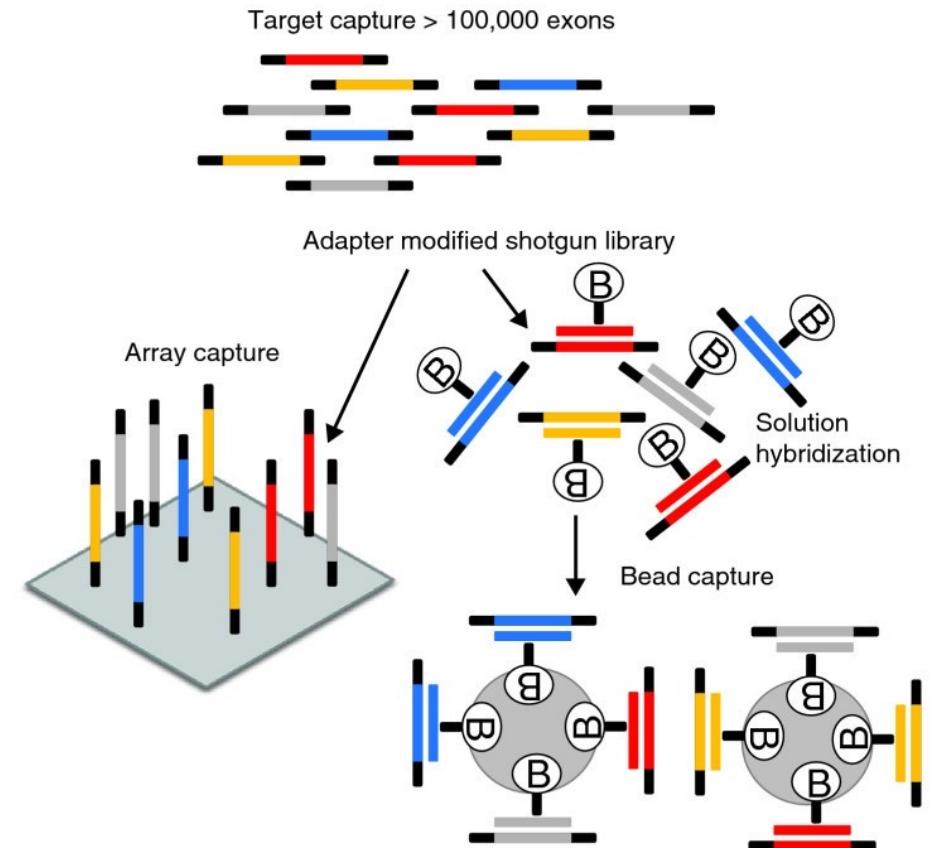


Targeted sequencing - RAPTURE

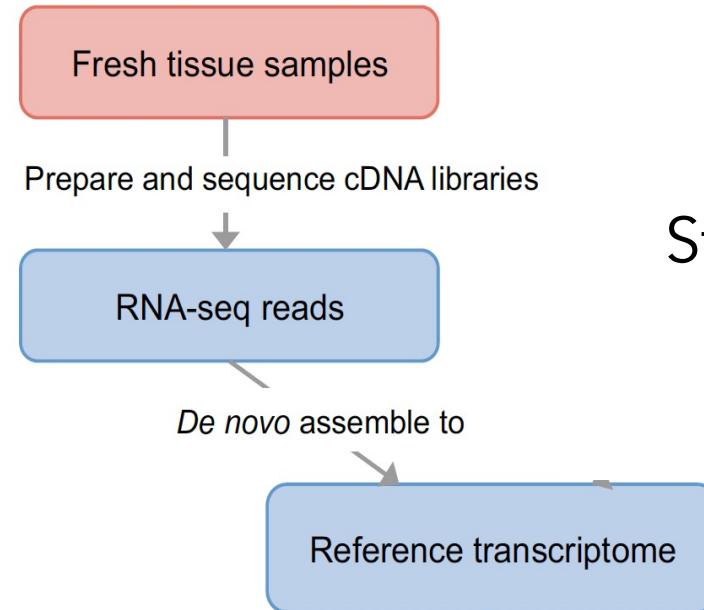


# Targeted approaches : Exome/exon capture

- Targets protein coding genes (or other known loci as well)
- While probes are generally available for some model species (human, mouse), they have to be designed for other species

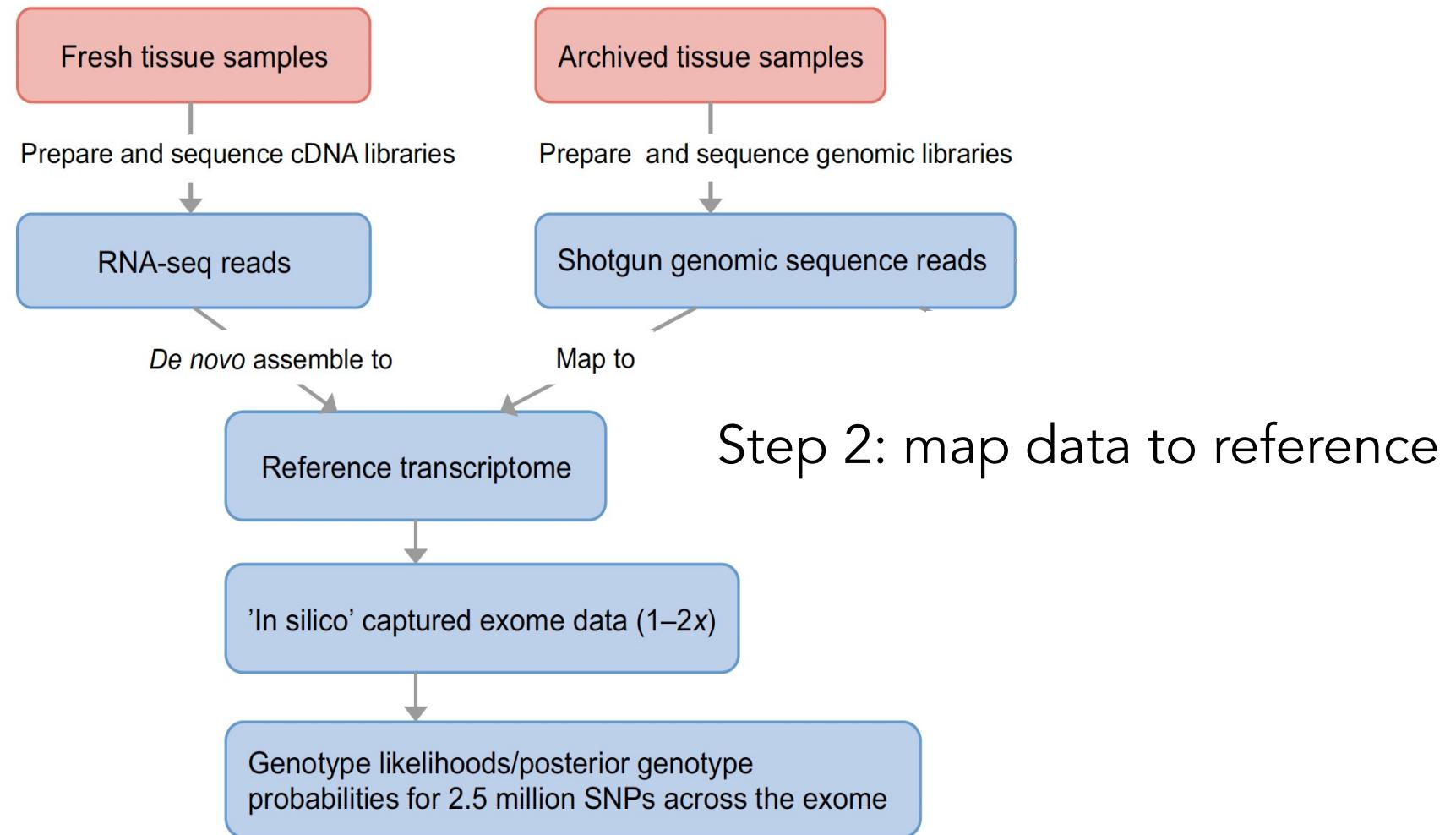


# Targeted approaches : Exome capture (*in silico*)

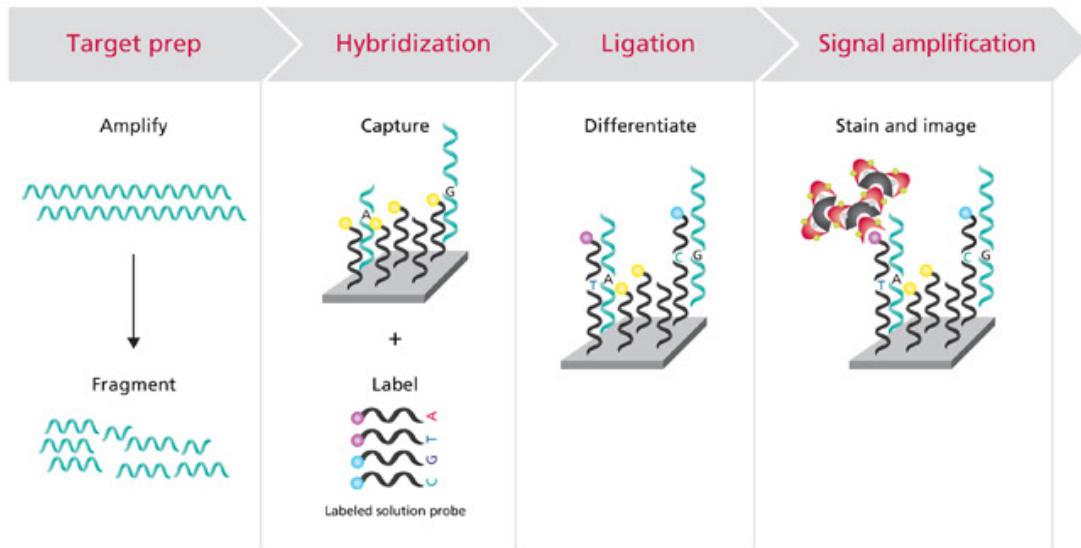


Step 1: build a reference transcriptome

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

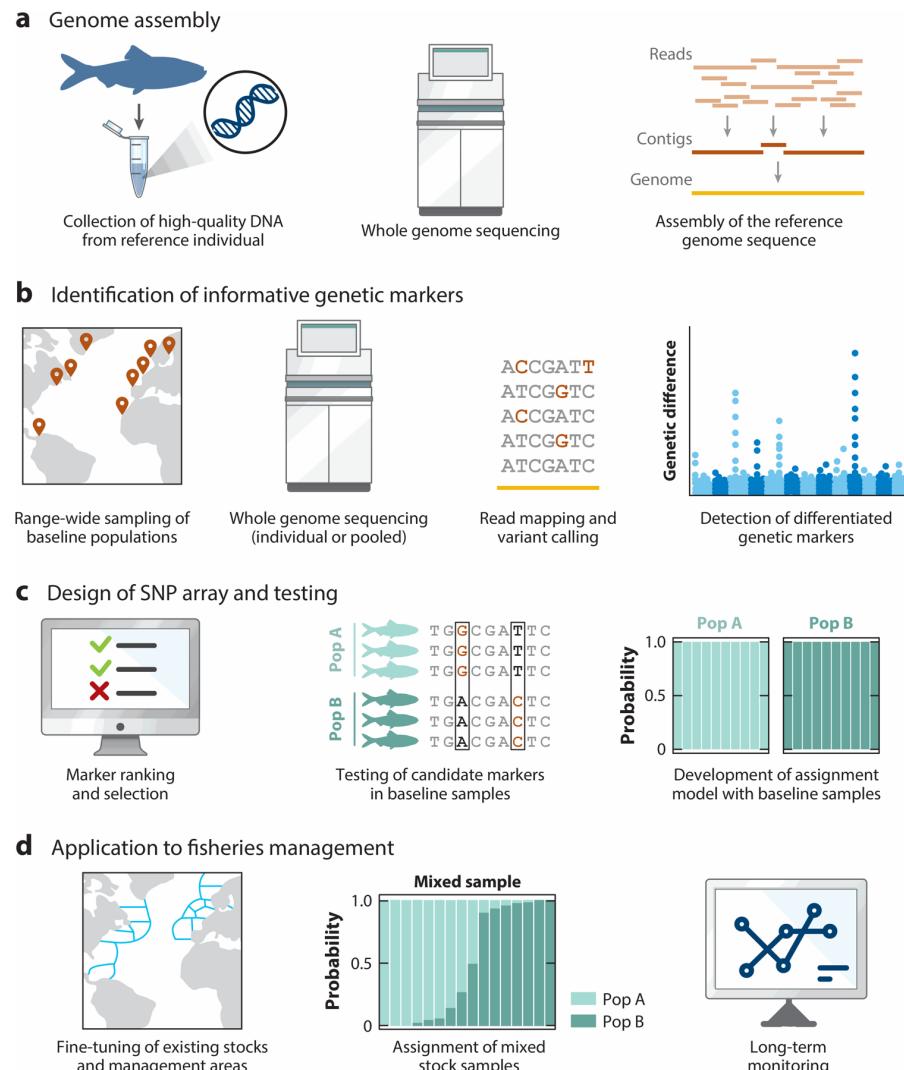


# Targeted approaches : SNP chip



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

# Targeted approaches - SNP chip



Atlantic herring  
(*Clupea harengus*)

- 5k Affymetrix Axiom genotyping array
- SNPs identified from whole genome resequencing data of 70+ populations
  - Neutral and outlier loci

→ For population genetics and fish stock monitoring

Andersson et al. 2024. Annual Review of Animal Biosciences

# Whole genome resequencing

Short-read sequencing :



Long-read sequencing :



Linked-reads technology :  
(synthetic long reads)



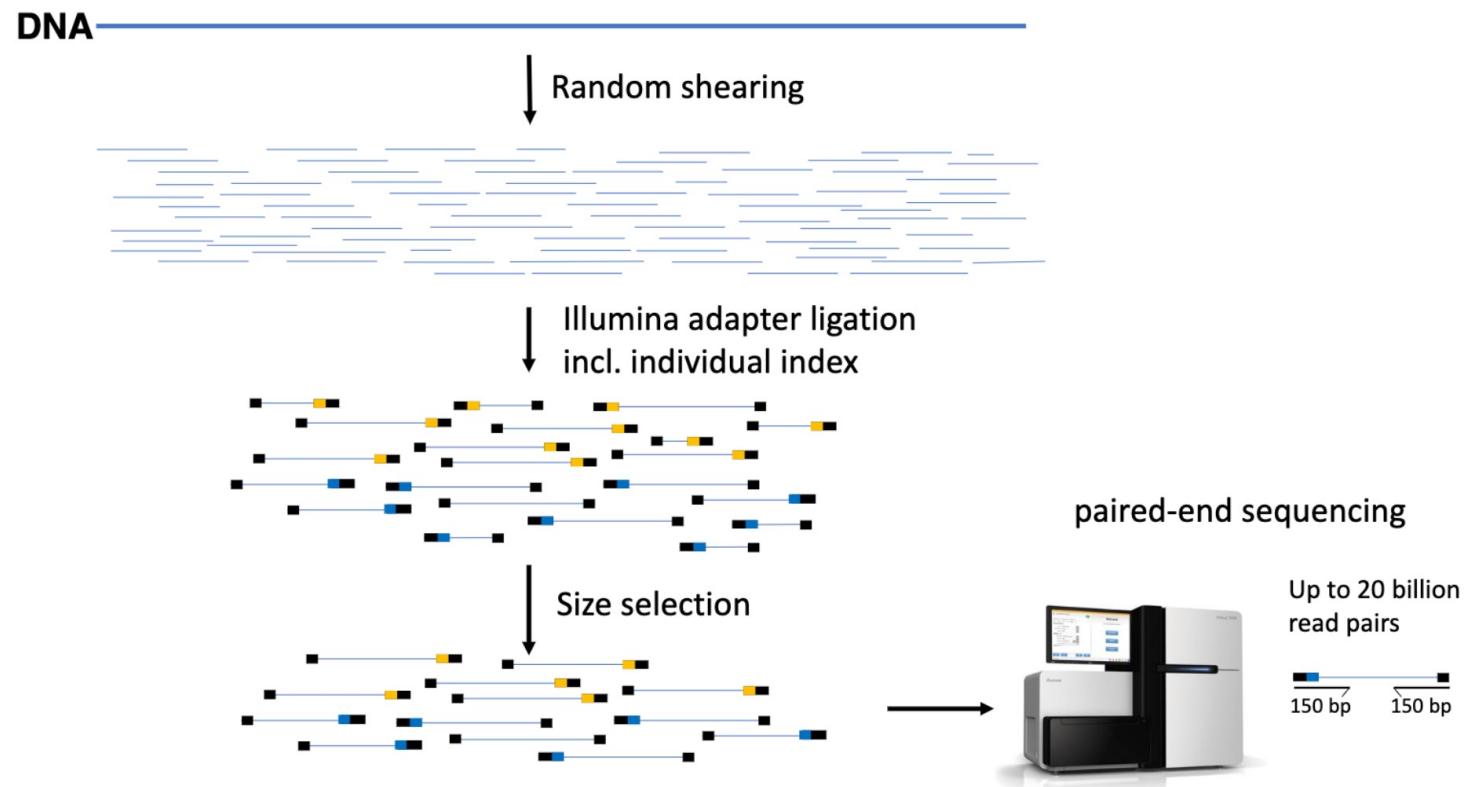
???

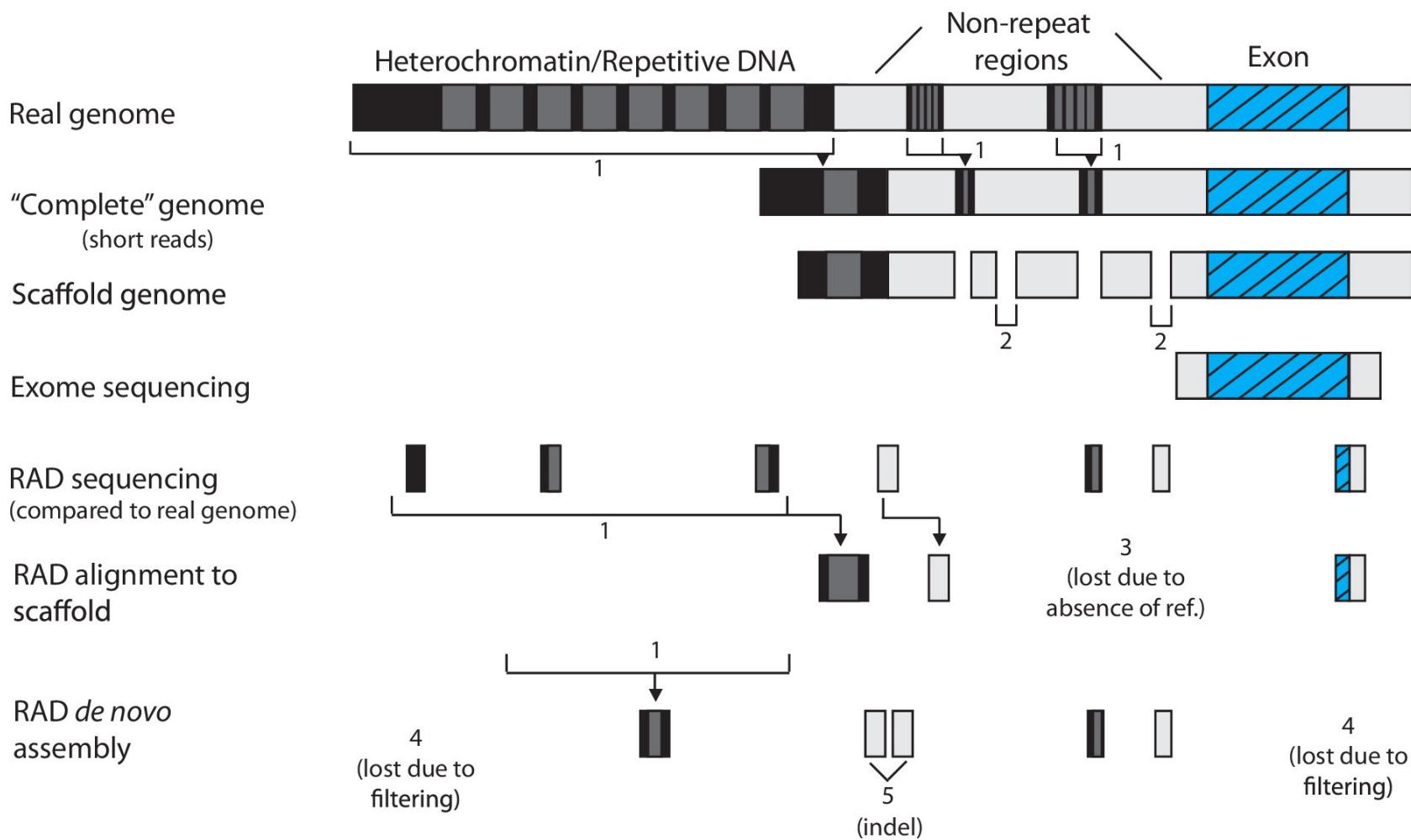
Tell-seq  
Haplotyping  
Stlfr...

= short-reads + barcodes by molecules

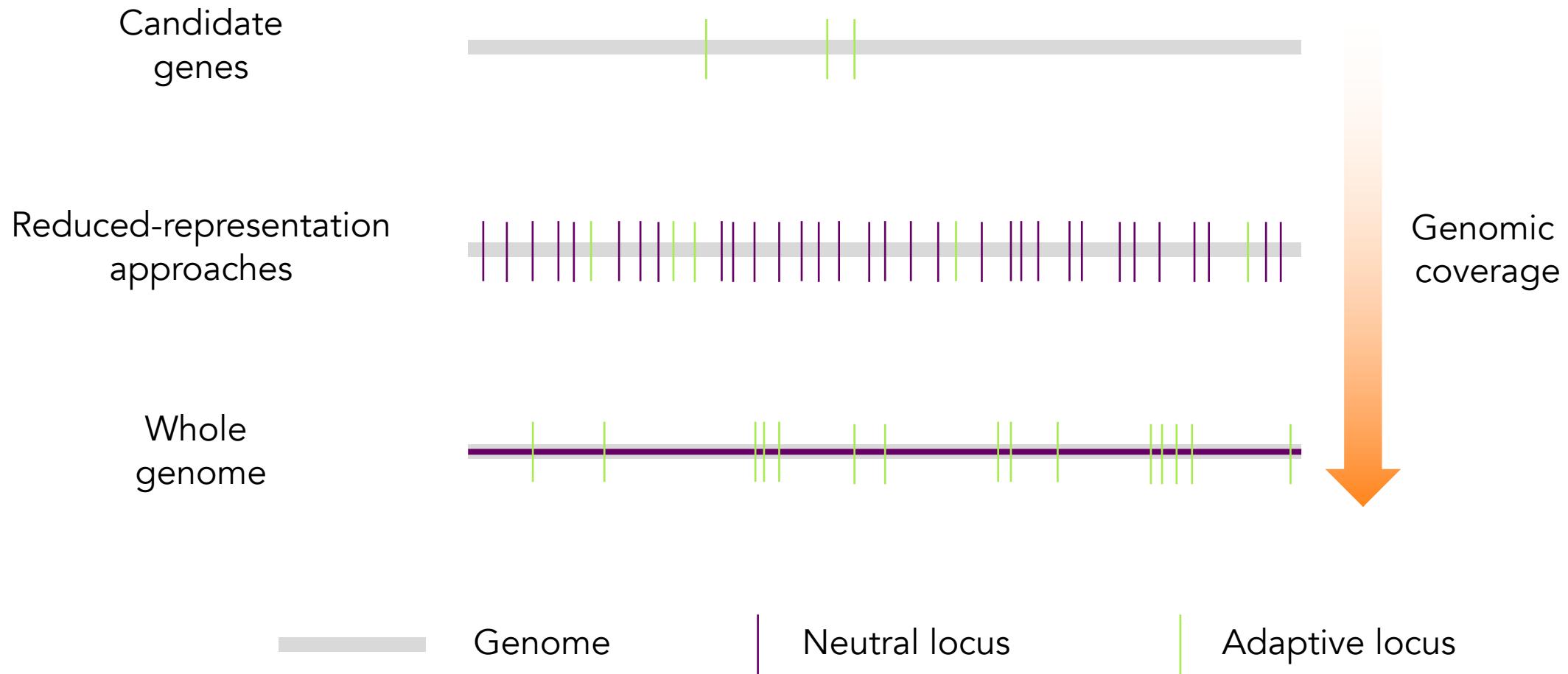
# Whole genome resequencing

Short-read sequencing :

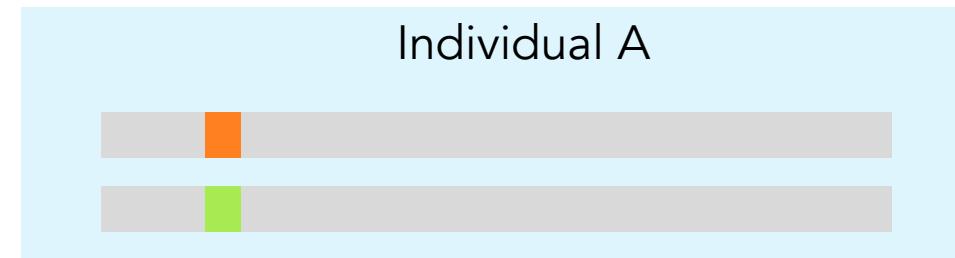




# Sequencing approaches



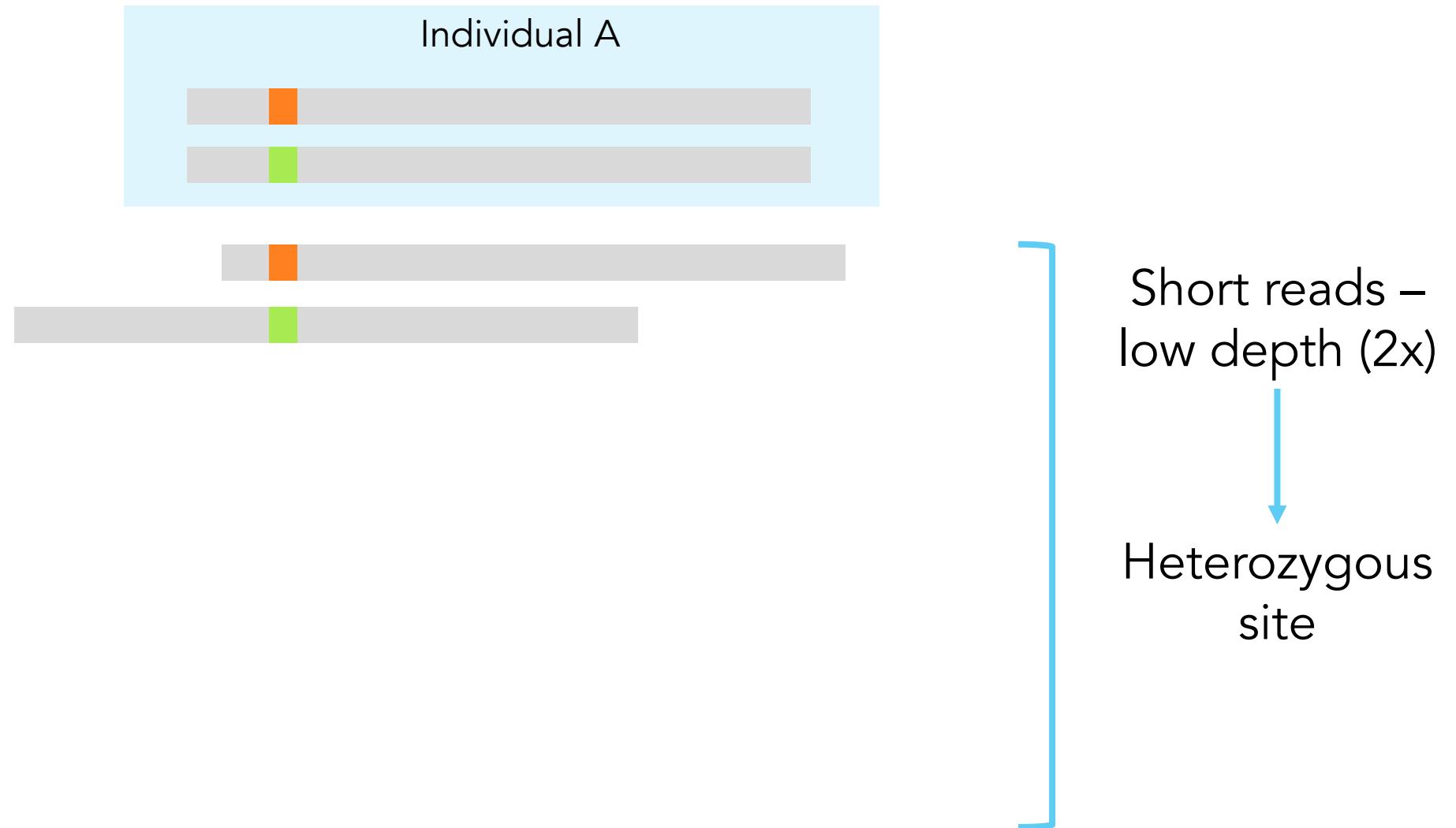
Sequencing depth = the number of reads covering a given site



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

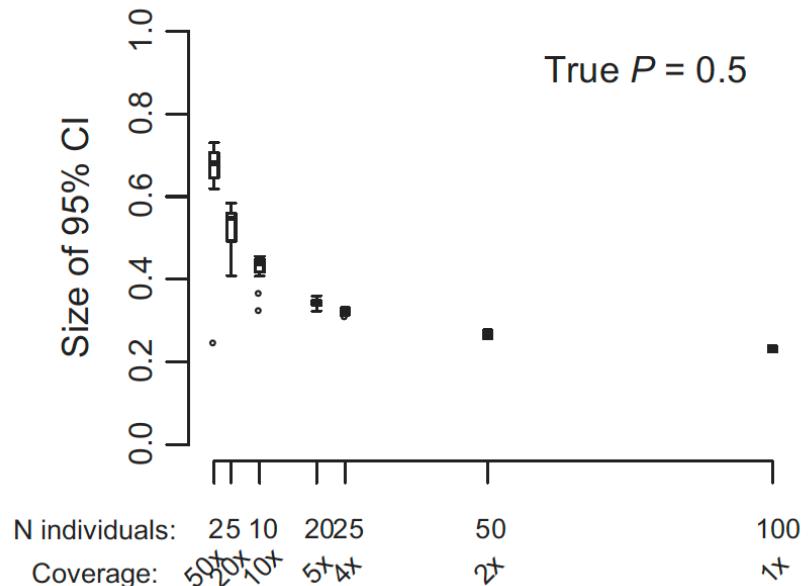
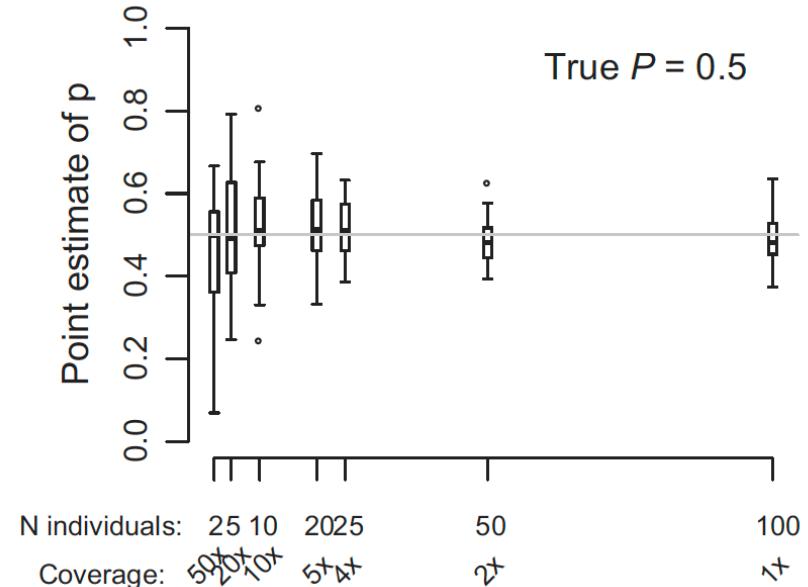


# Low-coverage sequencing

*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

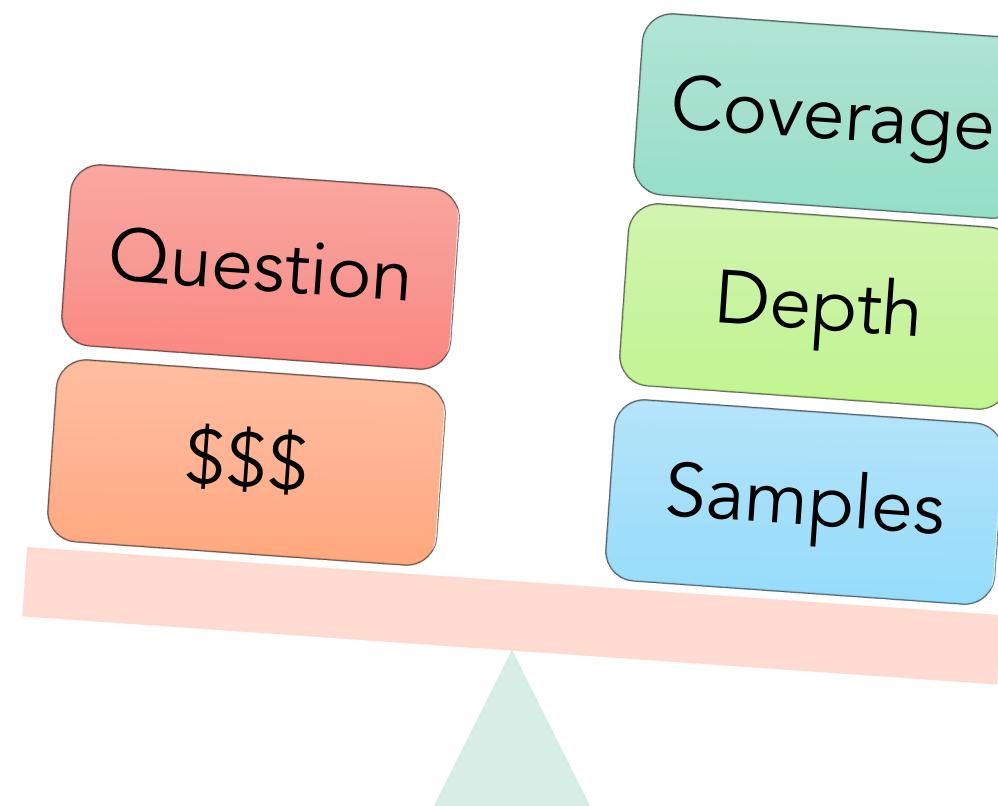
Question

\$\$\$

Coverage

Depth

Samples



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.