Workshop 1 - Application of Genetics to Small Pelagic Fish

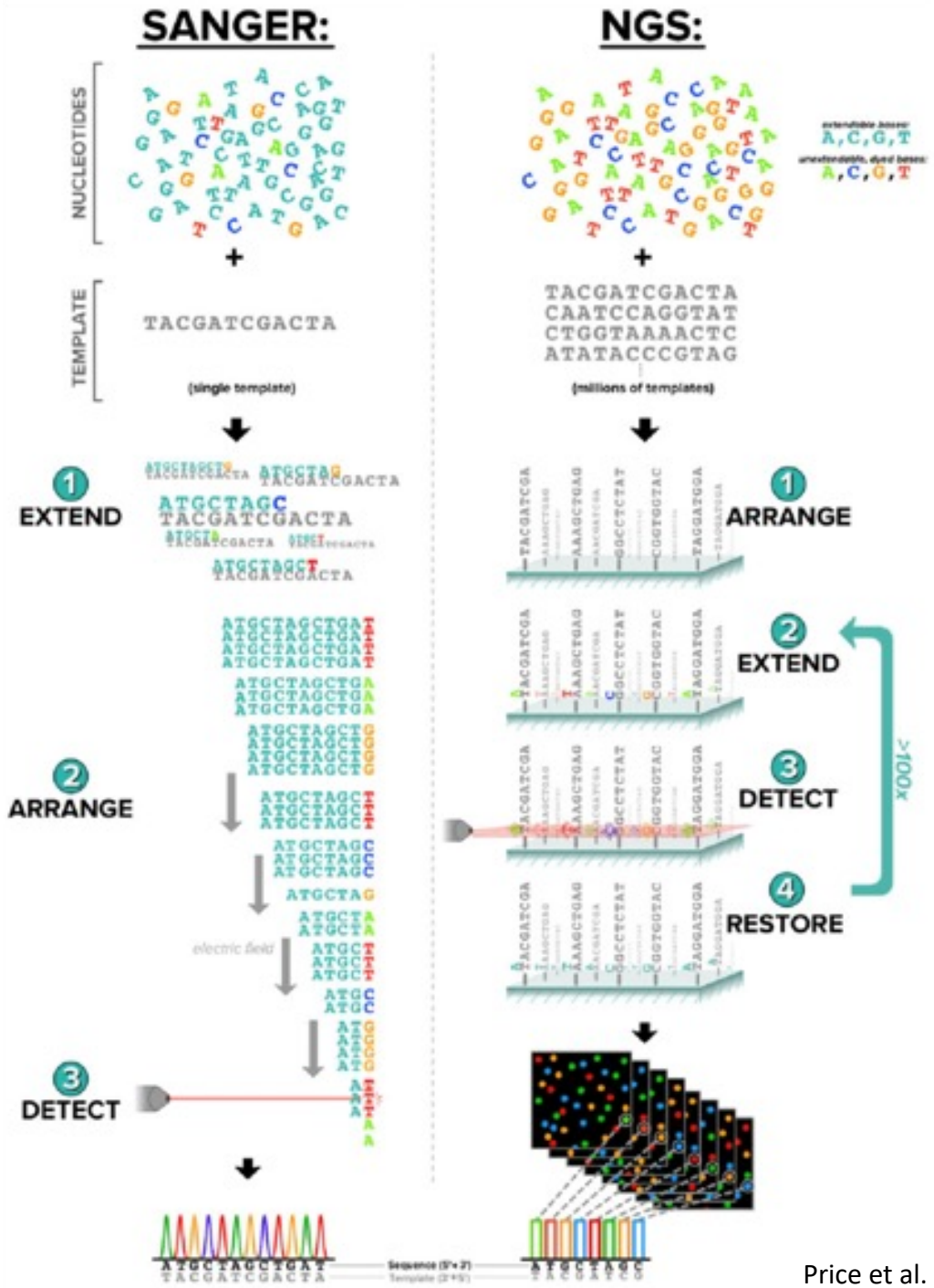
The potential of Next-Generation-Sequencing: from genes to genomes, and from single to multiple markers

Ana Veríssimo

cibio BIOPOLIS

Starting from the basics of NGS

- Moving from a single target to multiple targets.
- No need for pre-existing sequence information.
- Higher cost-efficiency in per bp sequencing costs.

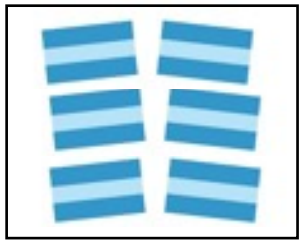


Price et al. 2018 *Biol Res Nursing*

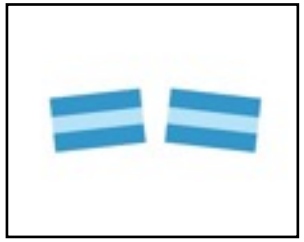
Starting from the basics of NGS

How do we do it?

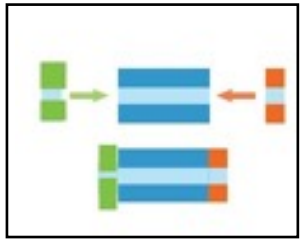
PCR amplification of a given target in many samples



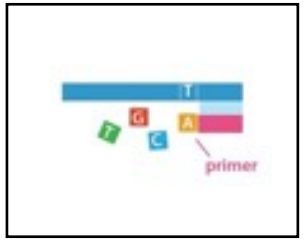
Extract DNA/RNA from **any tissue type and organism**



Break long chains/fragments into small pieces



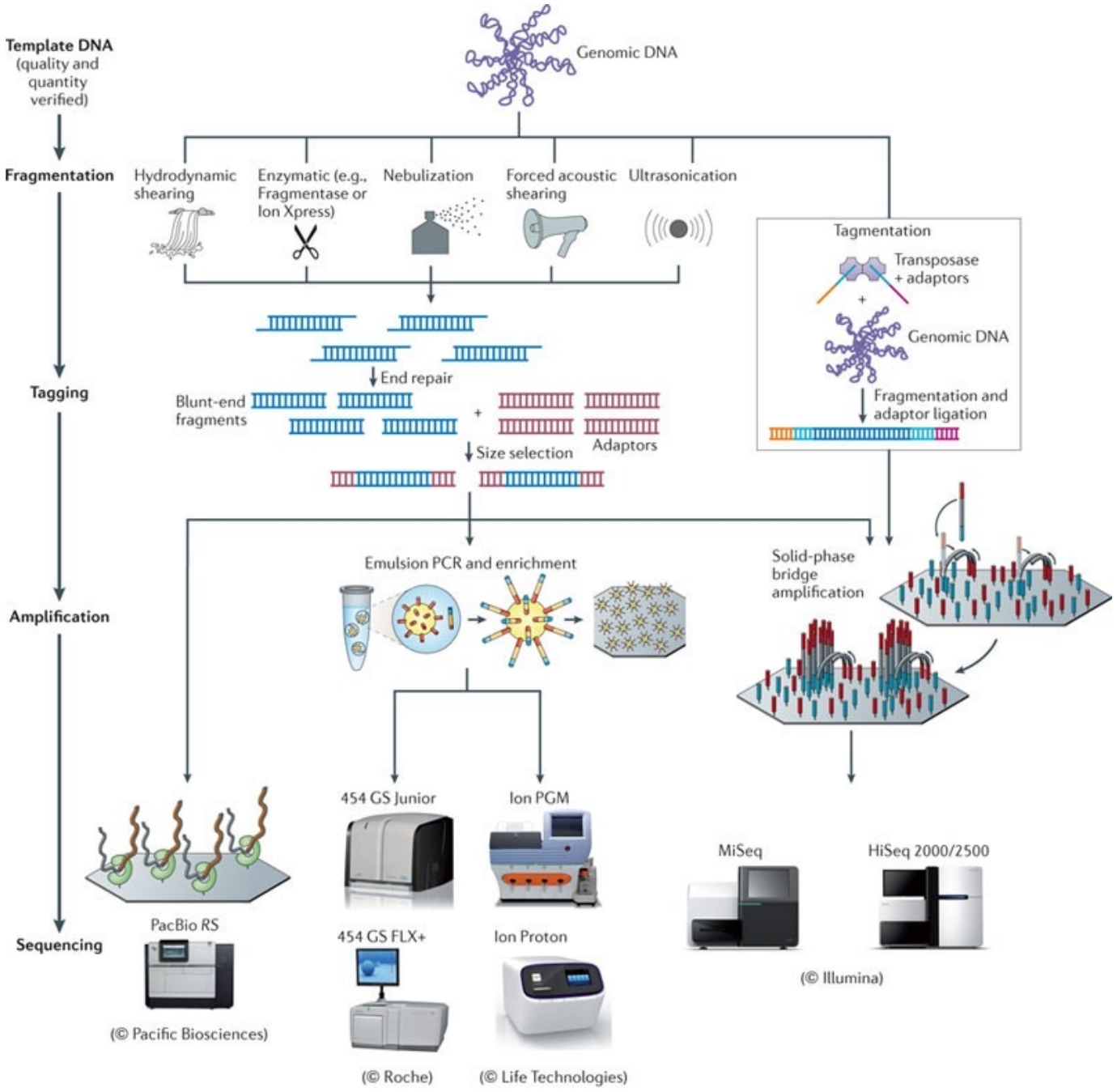
Add platform-specific adaptors for sequencing.



Massively parallel **sequencing of multiple fragments**.

Different platforms for NGS

- Different approaches to massively parallel sequencing



Loman et al. 2021 Nature Reviews | Microbiology

Different platforms for NGS

- Variable run outputs, error rates and sequence read lengths

Platform Name	Illumina HiSeq 2500	Ion Torrent-Proton II	PacBio RS II	OxFord Nanopore Minion
Instrument				
Cost (USD) **	690 k	224 k	695 k	1 k ***
Reagent cost Per run/per GB	4126/45.84	1000/20.41	100/1111.11	900/1000
Reads per run	300 millions	280 millions	0.03 millions	0.1 millions
Average Read length	2 × 150 bp	175 bp	14,000 bp	9,000 bp
Run time	10 h	5 h	2 h	6 h
Major errors	substitution	indel	indel	deletion
Error rate (%)	0.1	1	1	4
Amplification	bridgePCR	emPCR	none, SMS	none, SMS
Advantage	low cost per GB; high output	low cost	long reads; no amplification bias	long reads; no amplification bias
Disadvantage	high cost	homopolymer errors	low throughput; high cost	high error rate

* Sources: http://www.molecularecologist.com/next-gen-fieldguide-2014/ and websites of the companies;

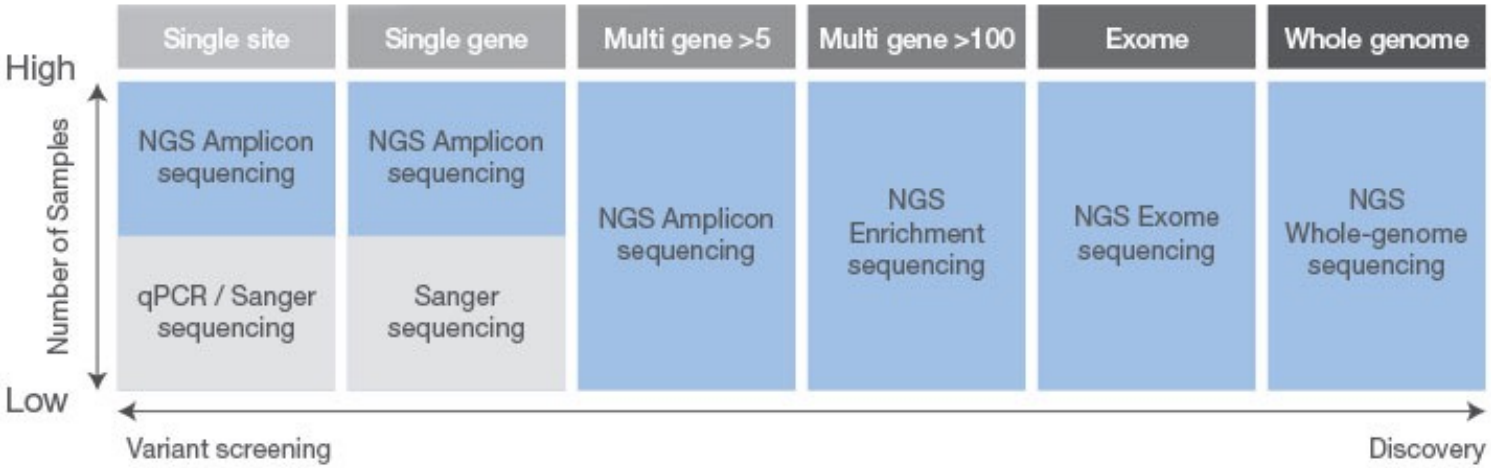
** Sources: http://www.molecularecologist.com/next-gen-table-3a-2014/;

*** Accessing fee. Sources: https://www.nanoporetech.com/products-services/minion-mki.

Ye et al. 2015 *Pharmaceutics*

What can you do with NGS?

Well, it depends…



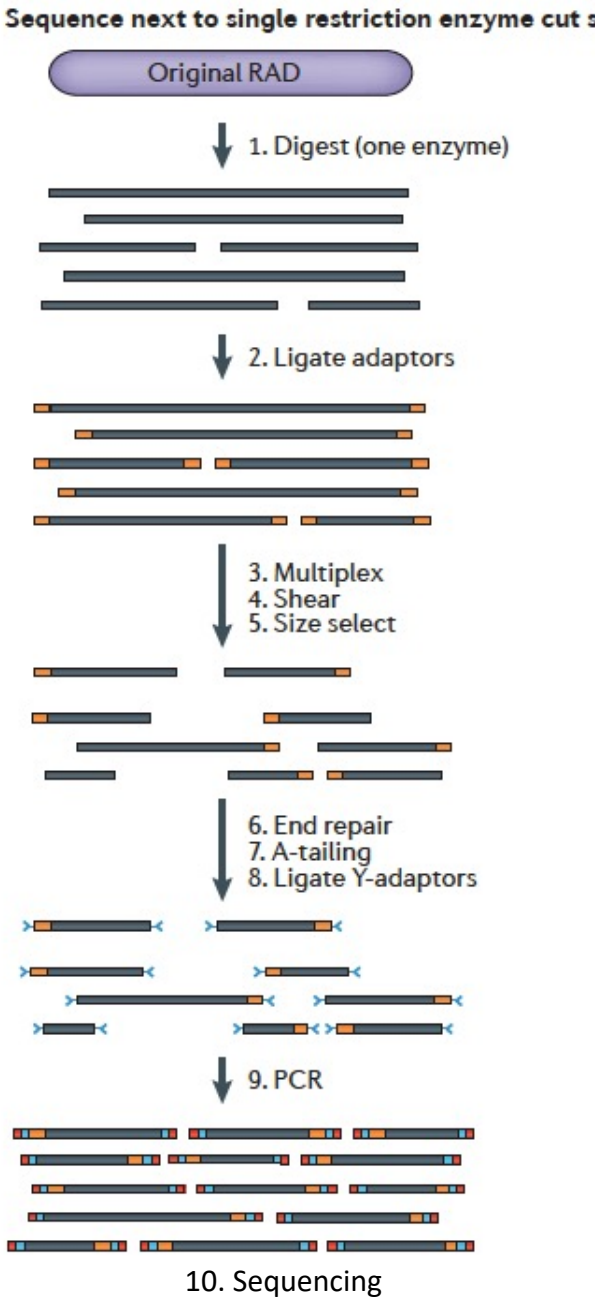
Illumina. com

Approaches & Case-studies

Genotyping by Sequencing

RADseq

(Restriction site Associated DNA Sequencing)

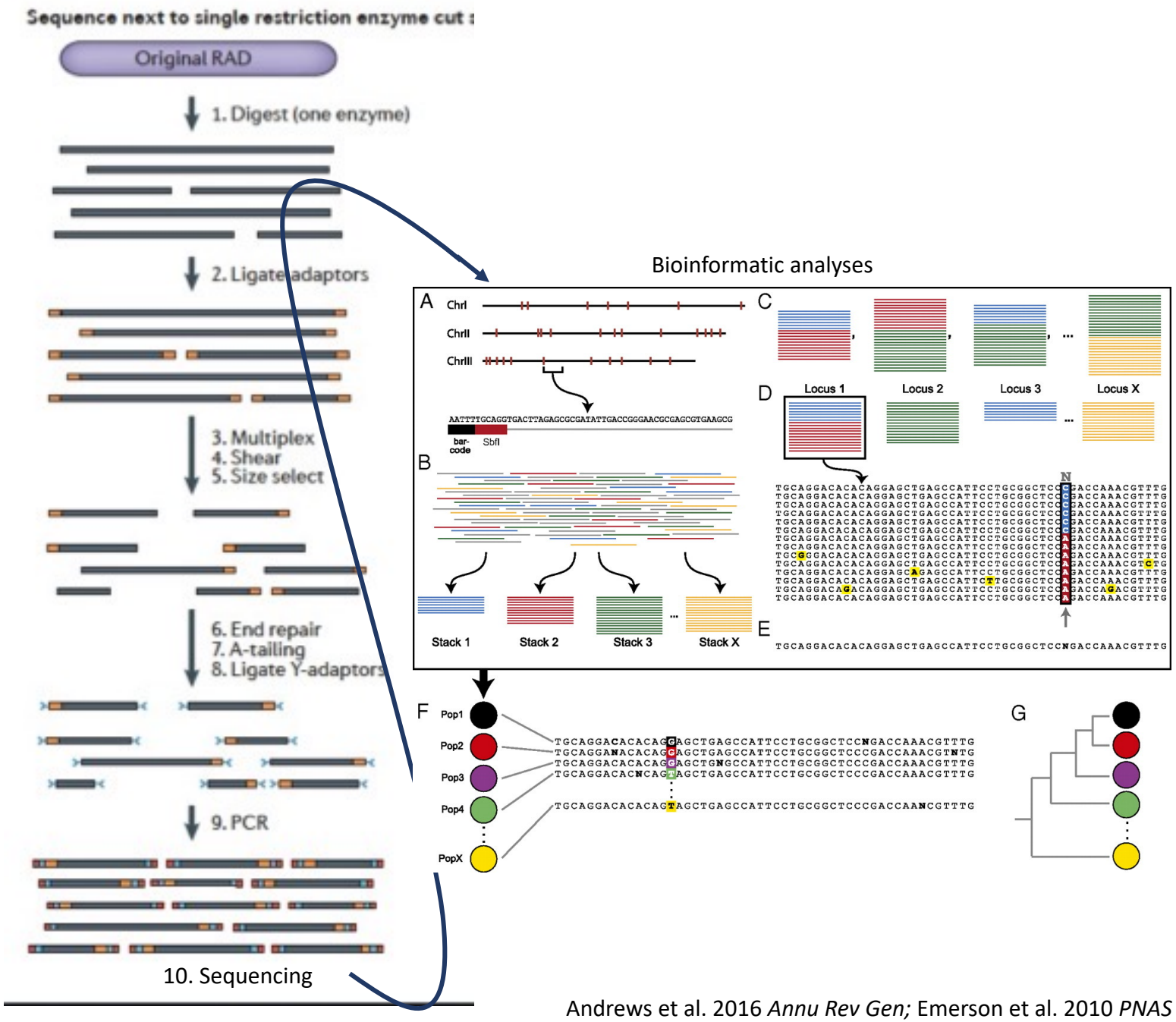


Andrews et al. 2016 *Annu Rev Gen*

Genotyping by Sequencing

RADseq

(Restriction site Associated DNA Sequencing)

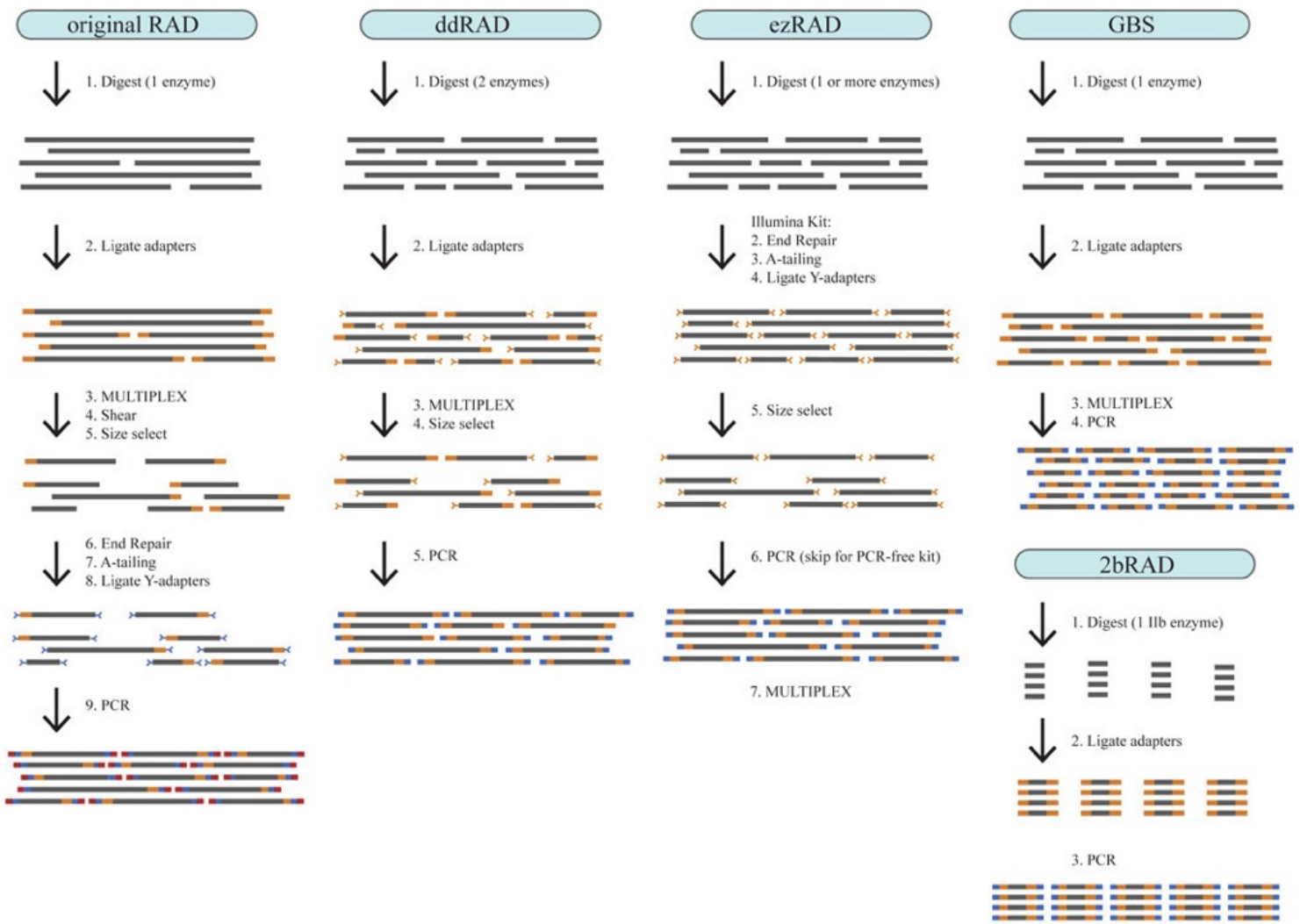


Andrews et al. 2016 *Annu Rev Gen;* Emerson et al. 2010 *PNAS*

Genotyping by sequencing

RADseq

Different approaches



Andrews et al. 2016 *Nat Rev Gen*

Genotyping by sequencing

RADseq

Different approaches have different trade-offs

Summary of trade-offs among five RADseq methods.

	Original RAD	2bRAD	GBS	ddRAD	ezRAD
Options for tailoring number of loci	Change restriction enzyme	Change restriction enzyme	Change restriction enzyme	Change restriction enzyme or size selection window	Change restriction enzyme or size selection window
Number of loci per 1 Mb of genome size *	30–500	50–1000	5–40	0.3–200	10–800
Length of single-end loci	≤1kb if building contigs; otherwise ≤300bp **	33–36bp	<300bp **	≤300bp **	≤300bp **
Cost per barcoded/indexed sample	Low	Low	Low	Low	High
Effort per barcoded/indexed sample	Medium	Low	Low	Low	High
Uses proprietary kit?	No	No	No	No	Yes
Can identify PCR duplicates?	with paired-end sequencing	No	with degenerate barcodes	with degenerate barcodes	No
Specialized equipment needed	Sonicator	None	None	Pippin Prep ***	Pippin Prep ***
Suitability for large or complex genomes ****	good	poor	moderate	good	good
Suitability for *de novo* locus identification (no reference genome) *****	good	poor	moderate	moderate	moderate
Available from commercial companies (in 2015)	Yes	No	Yes	Yes	No

* Estimated as follows: original RAD, assuming either a 6-cutter or 8-cutter; 2bRAD, assuming type IIB enzymes with recognition sites containing 5–7 specific nucleotides; GBS, values from Elshire *et al.*[66]; ddRAD, from Table 1 in Peterson *et al.*[14] and allowing for up to double the size range; ezRAD, values from Toonen *et al.*[16] for species with reference genomes.

** Based on current limits in sequencing technology

*** Can alternatively be used with standard gel equipment

**** Based on ability to reduce total number of loci and lengths of loci

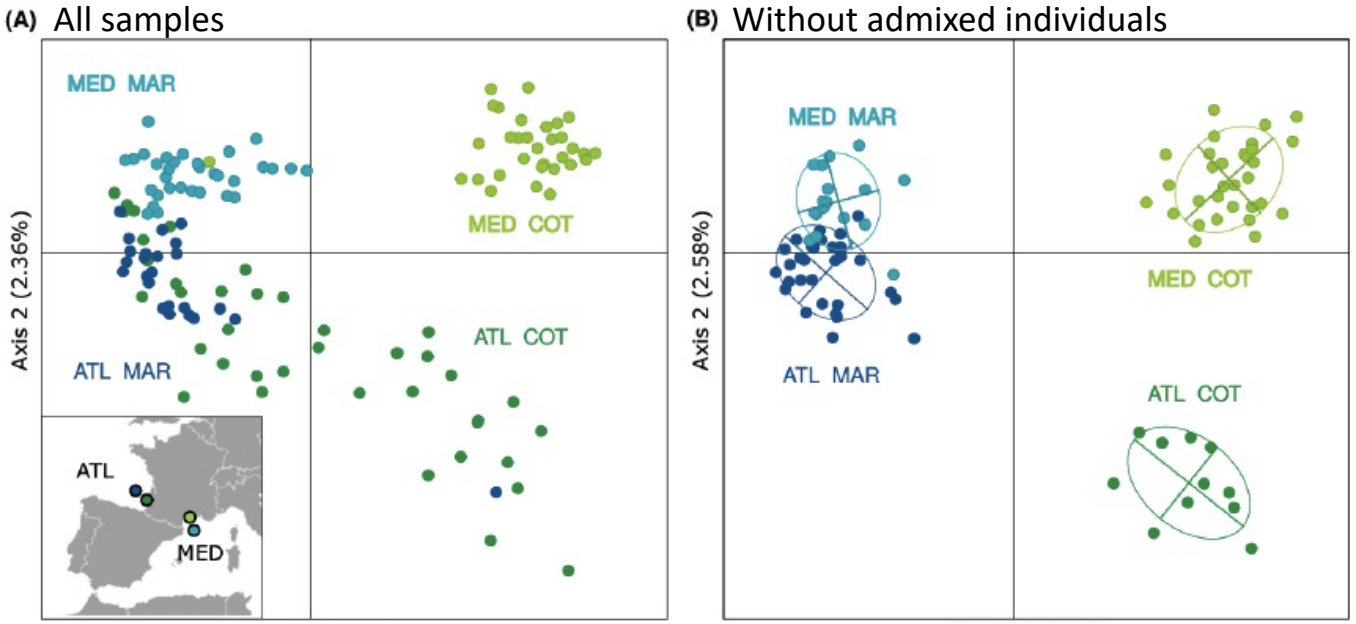
***** Based on lengths of loci to distinguish paralogs and duplicate sequence

Andrews et al. 2016 *Nat Rev Gen*

Genotyping by sequencing

RADseq

Genetic basis of coastal vs. marine ecotypes differentiation in the European anchovy *Engraulis encrasicolus*



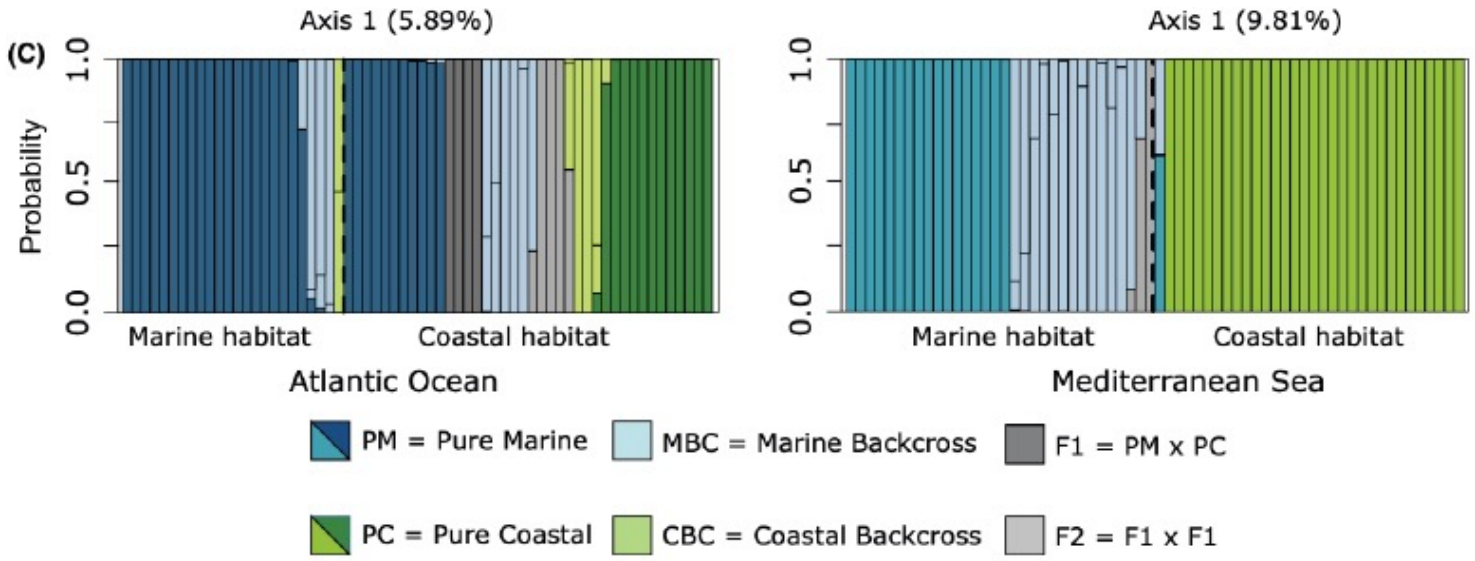
Habitat type is the major driver of genetic structure.

Le Moan et al. 2016 *Mol Ecol*

Genotyping by sequencing

RADseq

Genetic basis of coastal vs. marine ecotypes differentiation in the European anchovy *Engraulis encrasicolus*



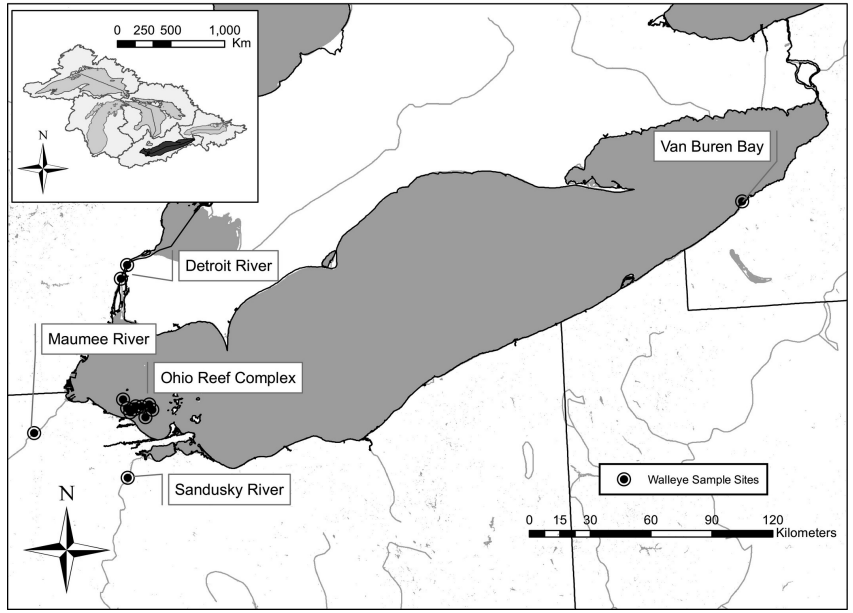
Despite the strong differentiation between ecotypes,
gene flow (mixing) is occurring between them in ATL and MED

Le Moan et al. 2016 *Mol Ecol*

Genotyping by sequencing

RADseq

Mixed-stock analyses of Lake Eerie walleye



Multiple spawning units with different productivity.

Stock mixing occurs in the eastern basin outside the spawning season.

Commercial fisheries in the eastern basin target a mix of fish from different stocks.

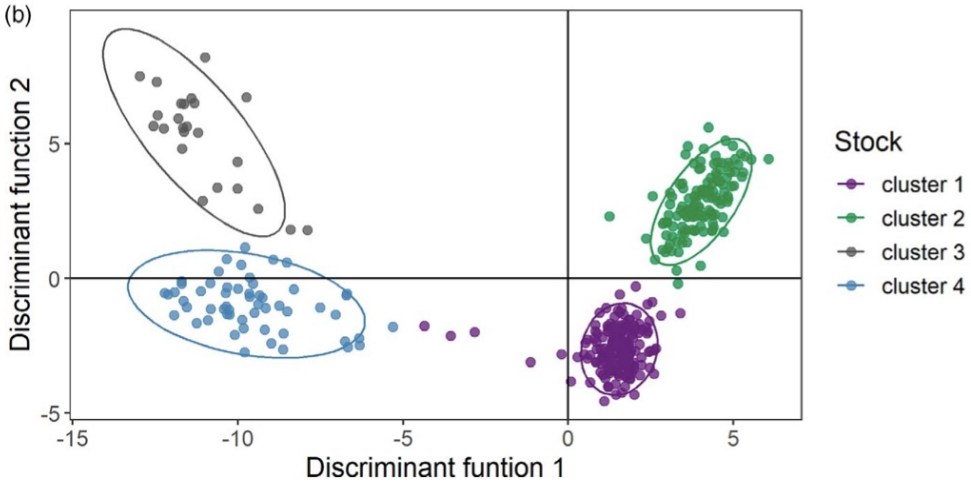
The problem: How much of the catch comes from the different stocks?

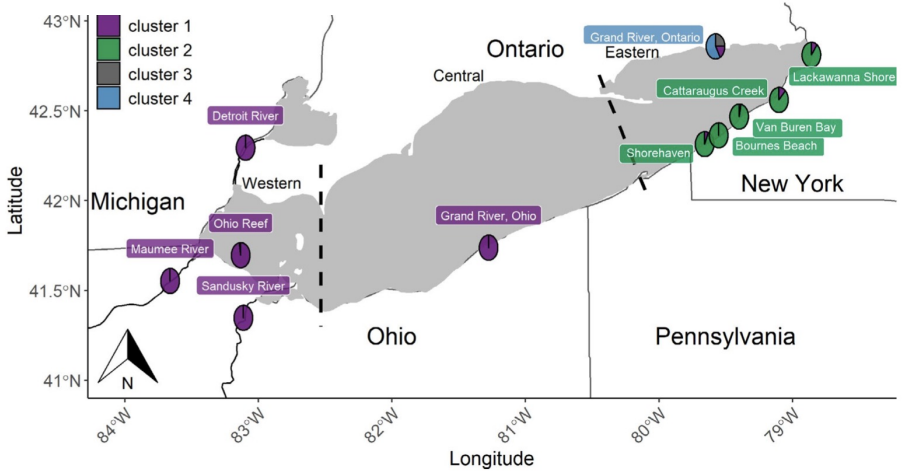
Euclide et al. 2020 *Ecol Appl*

Genotyping by sequencing

RADseq

Mixed-stock analyses of Lake Eerie walleye



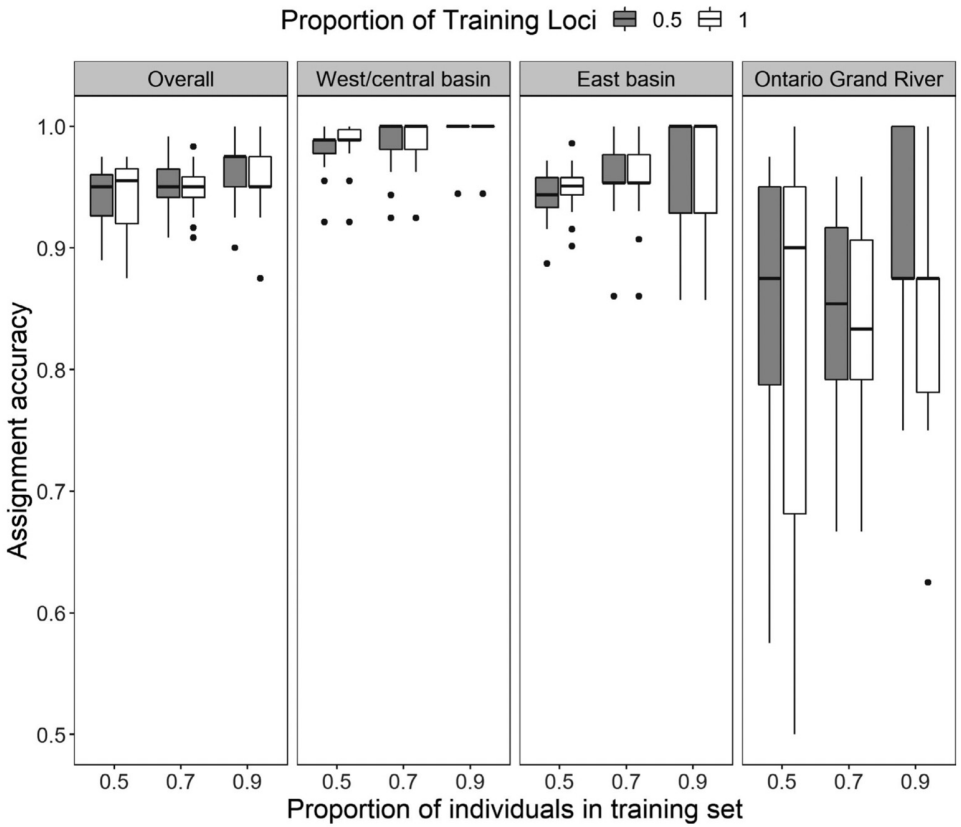


Genetic differentiation of four clusters encompassing the spawning stocks.

Euclide et al. 2020 *Ecol Appl*

SNP panels

Mixed-stock analyses of Lake Eerie walleye

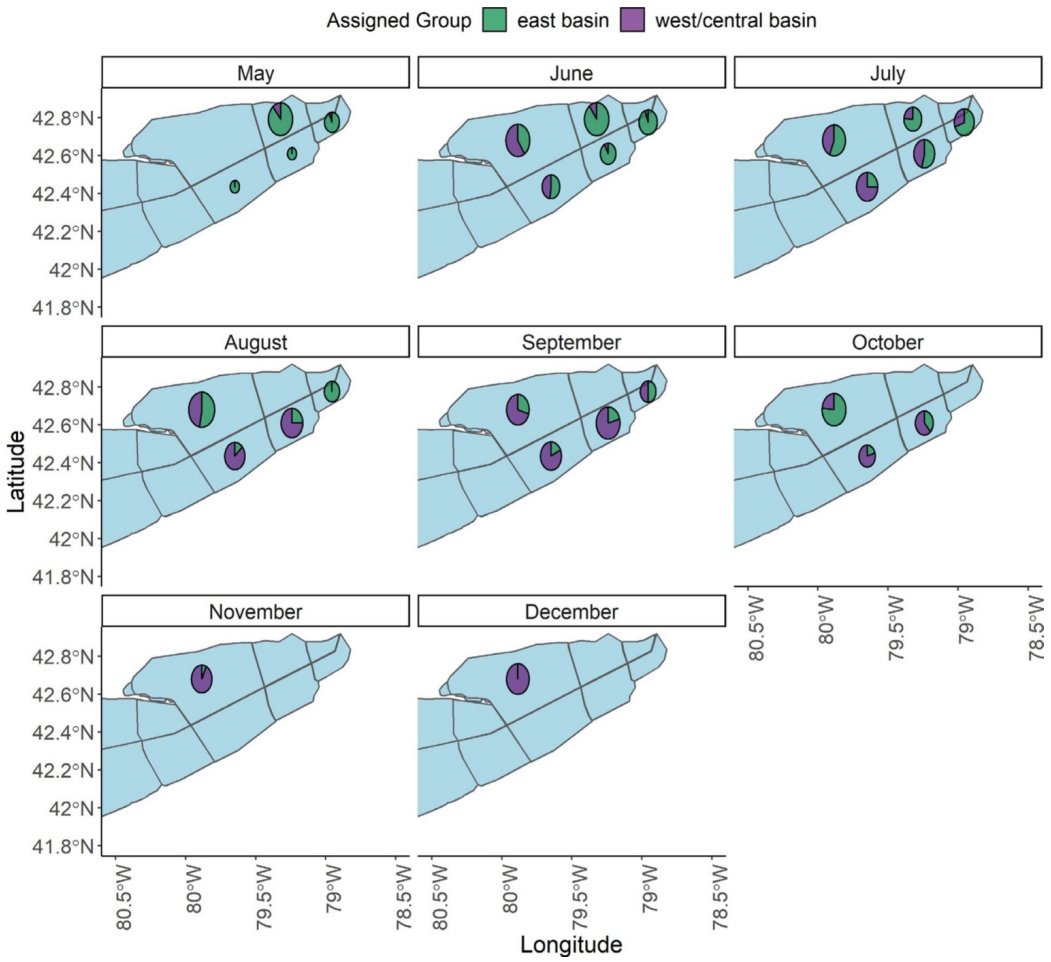


SNPs were able to discriminate individuals and re-assign them to their source populations.

Euclide et al. 2020 *Ecol Appl*

SNP panels

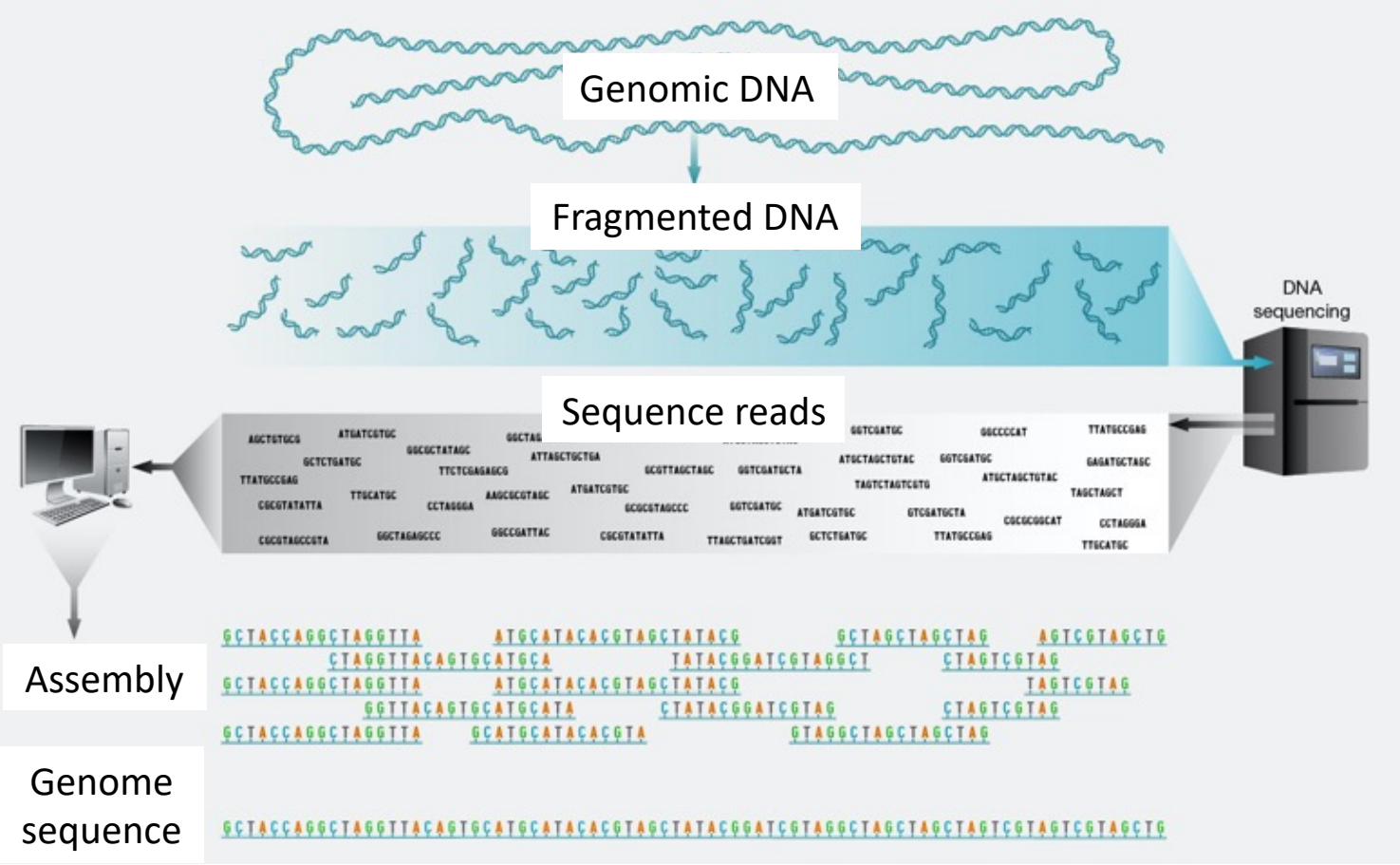
Mixed-stock analyses of Lake Eerie walleye



Assignment of harvested individuals from unknown origin was performed on eastern basin through time.

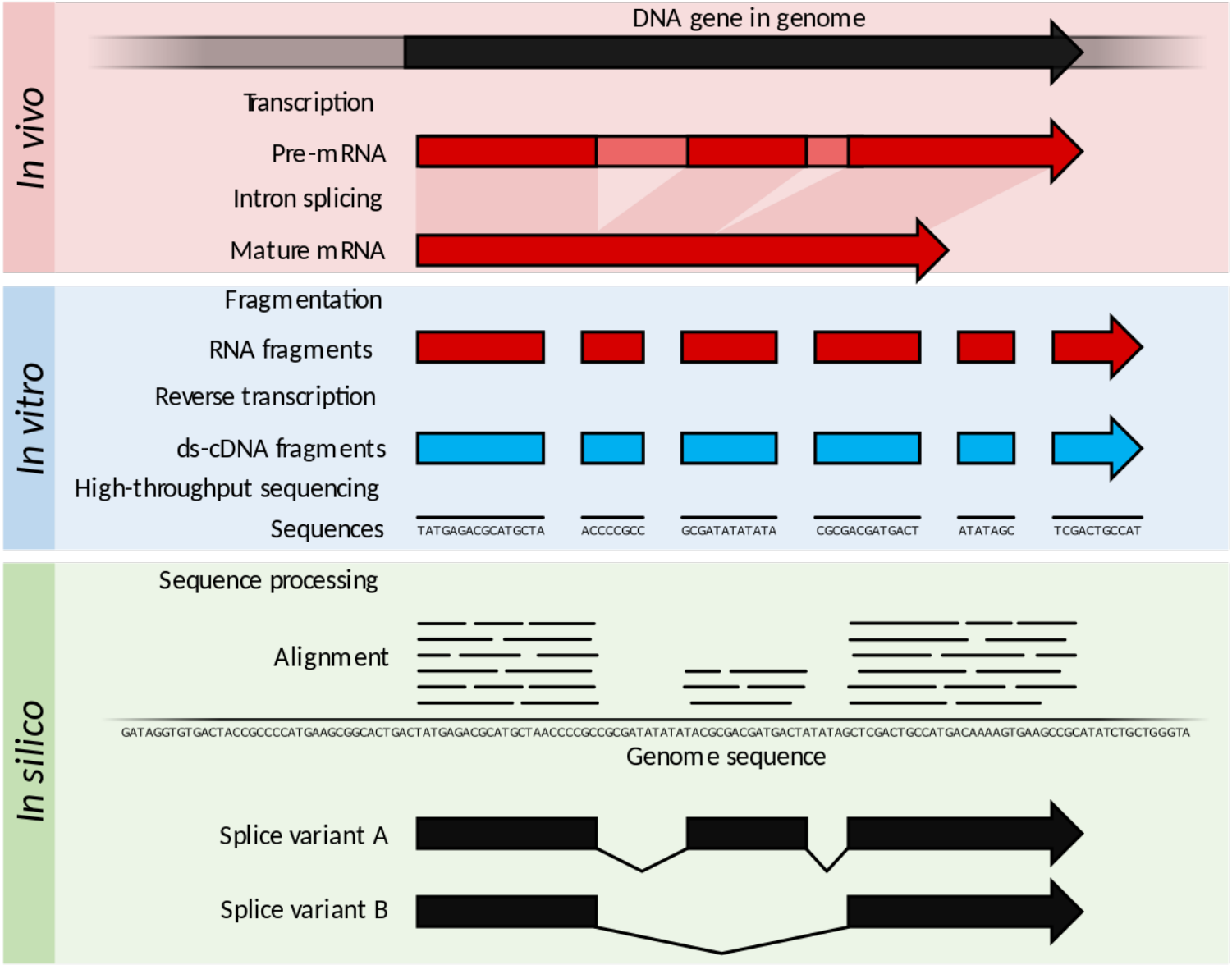
Euclide et al. 2020 *Ecol Appl*

Whole genome sequencing



NIH genome.gov

Whole transcriptome sequencing

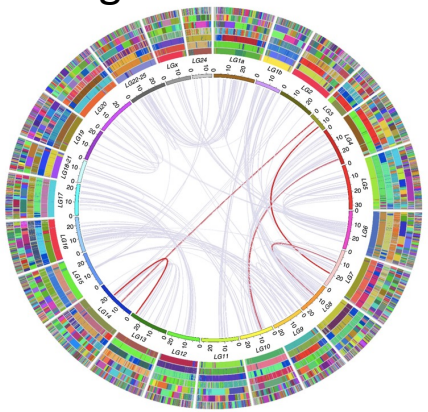


Lowe et al. 2017 *PLOS Comput Biol*

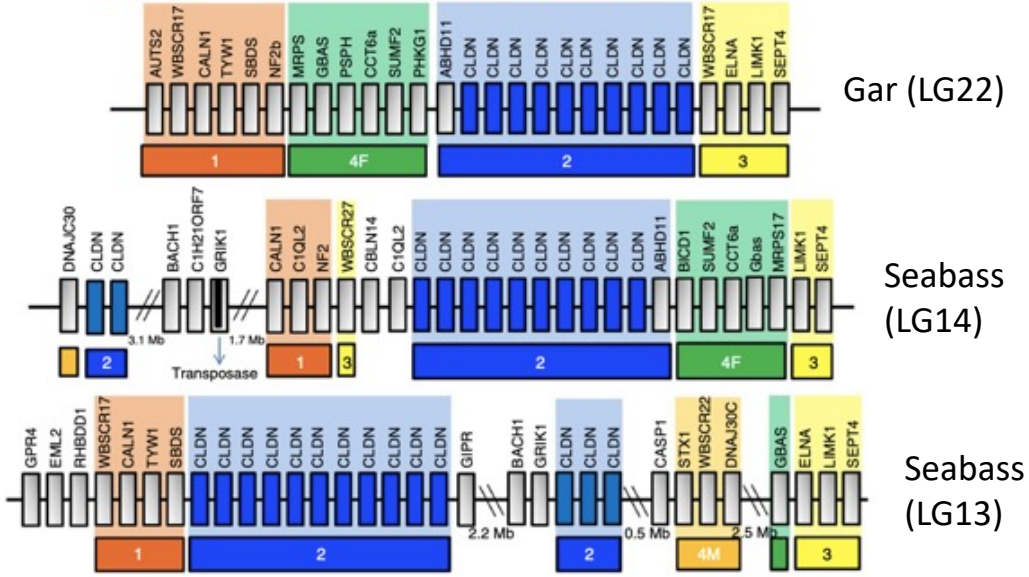
Whole genome/ transcriptome sequencing

Adaptation to euryhalinty in European seabass *Dicentrarchus labrax*

Whole genome assembly



Genome annotation using transcritome assembly



Gar (LG22)

Seabass (LG14)

Seabass (LG13)

Gene expansions in osmoregulatory gene families

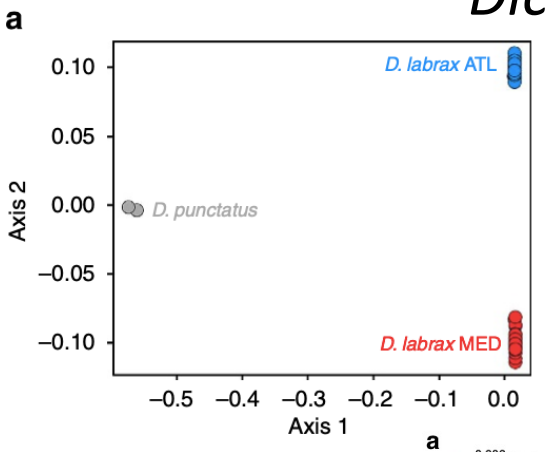
Tine et al. 2014 *Nat Comm*

Whole genome/ transcriptome sequencing

+

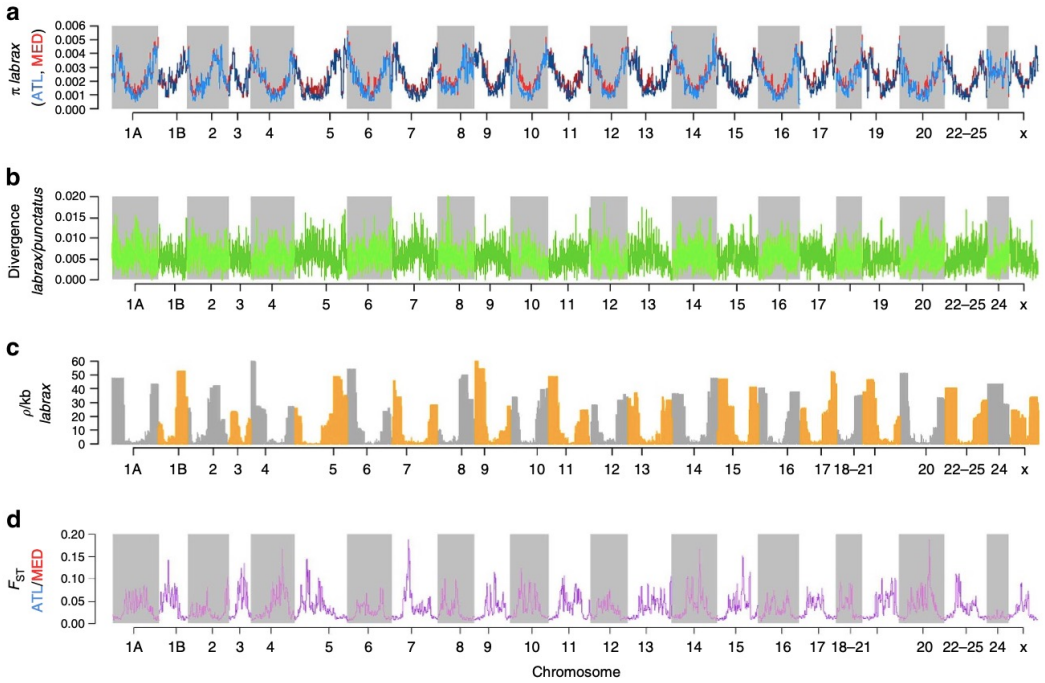
RADseq

Adaptation to euryhalinty in European seabass *Dicentrarchus labrax*



Distinct genetic lineages in ATL & MED.

But where are the differences?



Whole genome perspective of genetic variation.

Tine et al. 2014 *Nat Comm*

Metabarcoding

(Amplicon sequencing)

eDNA metabarcoding work flow

Single marker:
COI
MiFish
18S
Other…

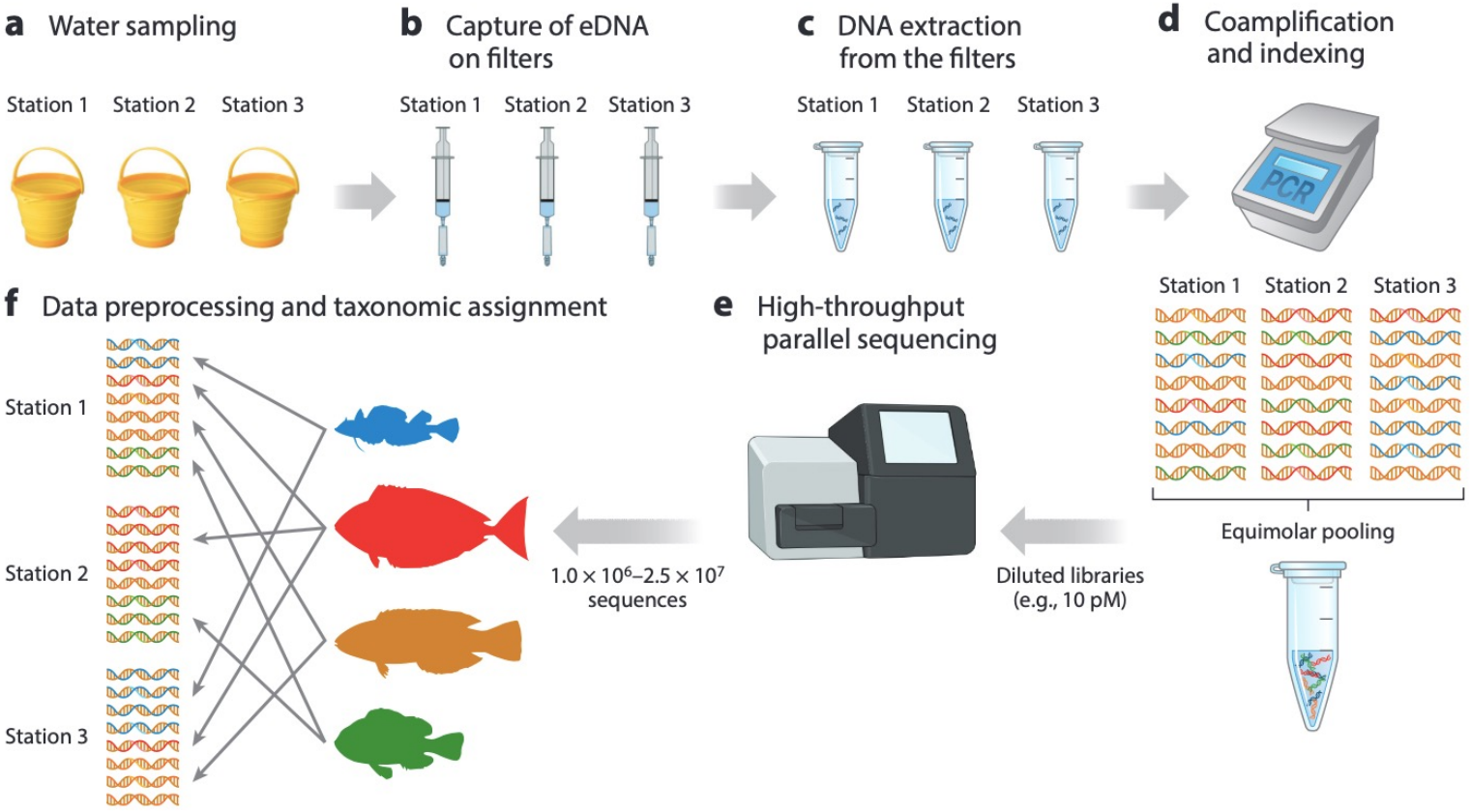


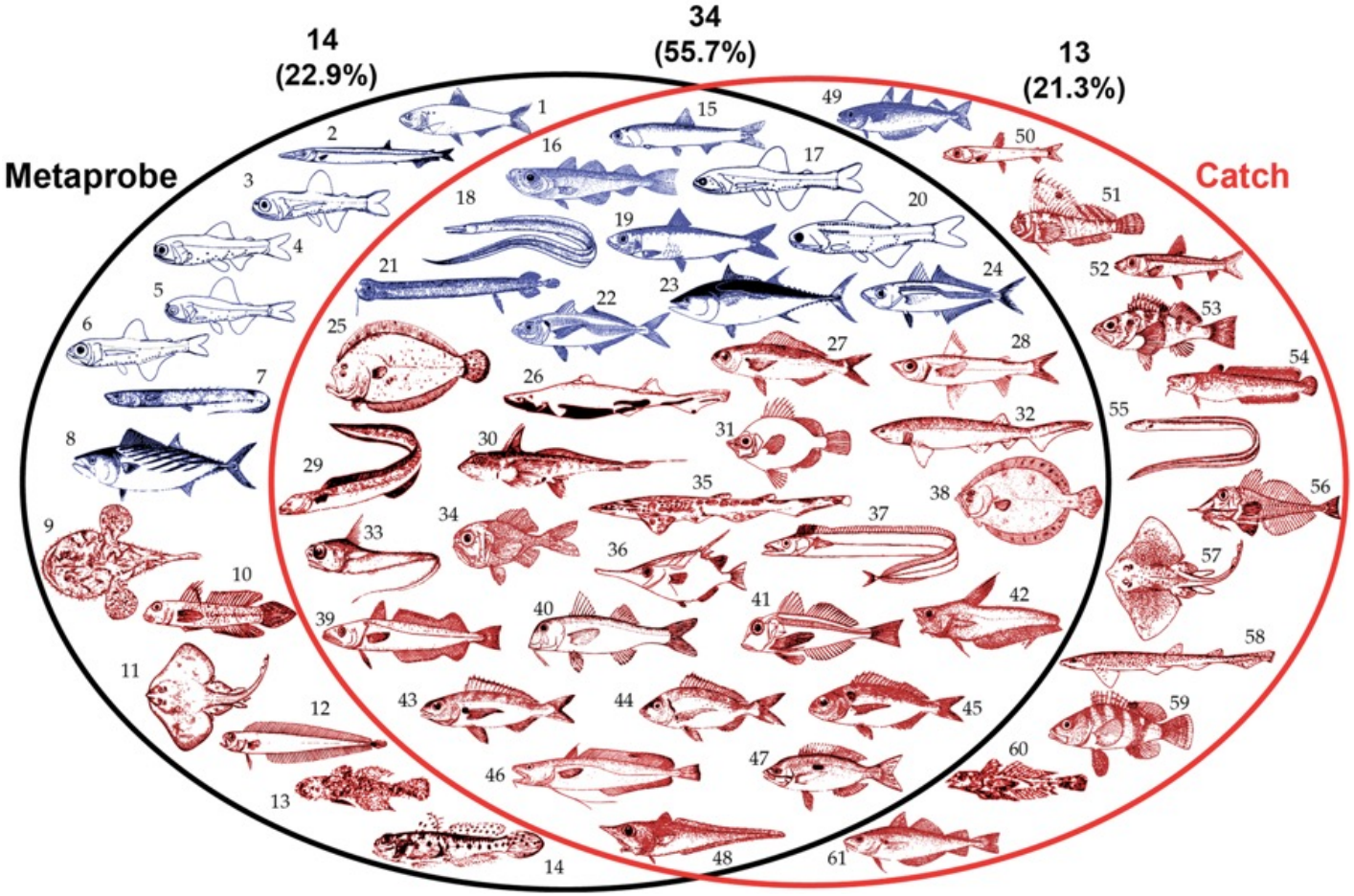
Figure 1

Miya 2022 *Annu Rev Mar Sci*

Metabarcoding

Biodiversity monitoring

Fishing vessels as ocean biodiversity samplers

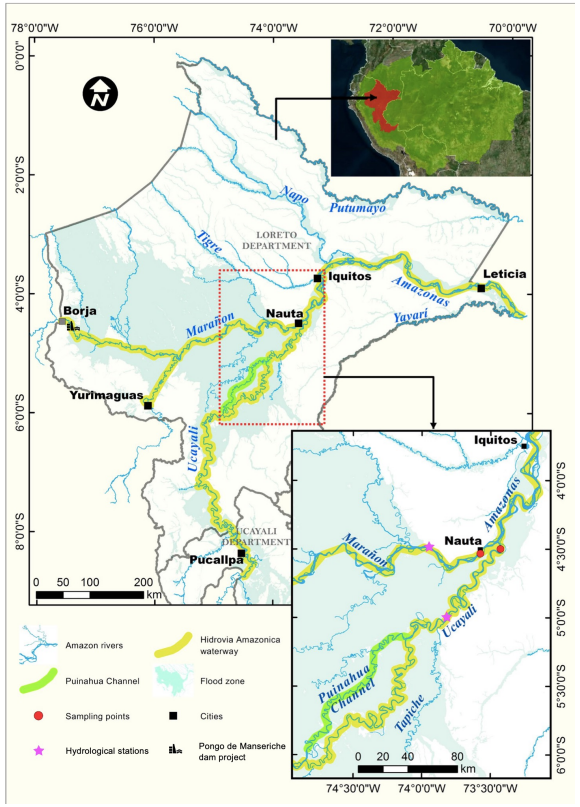


Maiello et al. 2022 *Fish Res*

Metabarcoding

Ichthyoplankton monitoring

Species-level ichthyoplankton diversity and dynamics



Goals

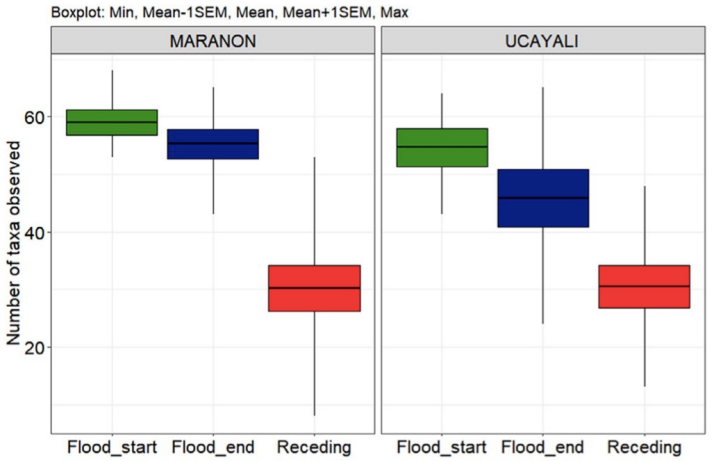
- no. of species spawning in the system.
- spawning periods & relative abundances with hydrological cycle
- contribution of a given river to larvae production of commercial species

Mariac et al. 2022 *Mol Ecol*

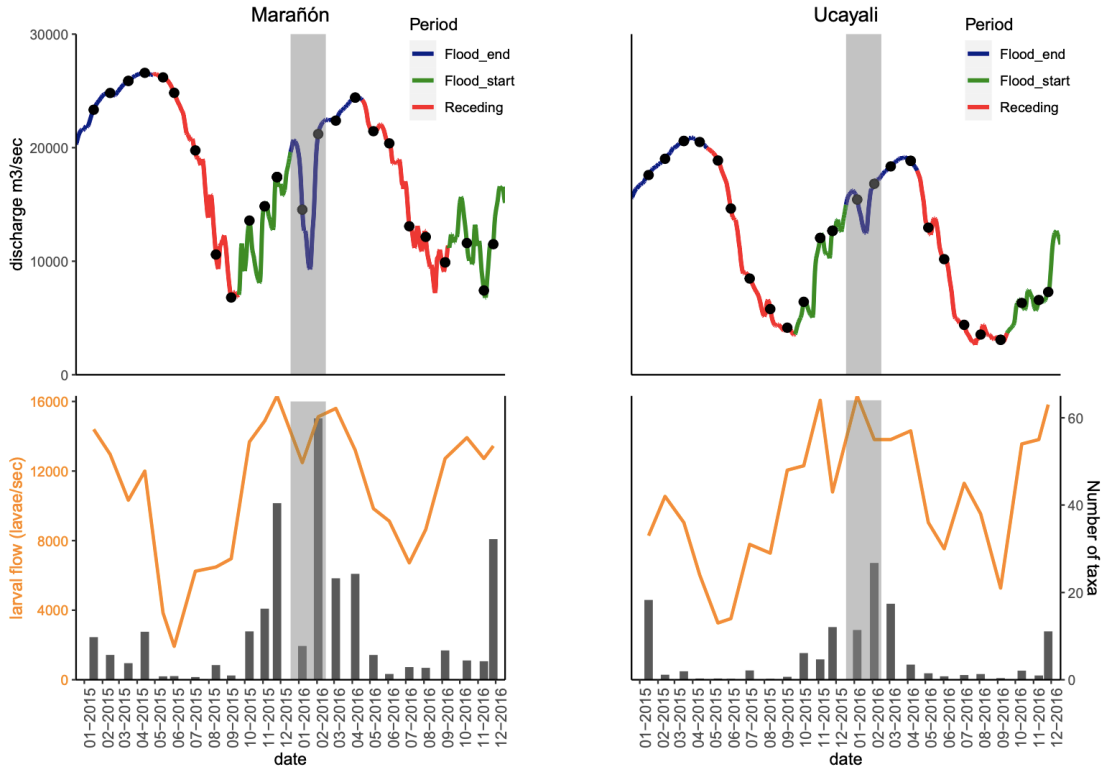
Metabarcoding

Ichthyoplankton monitoring

Species-level ichthyoplankton diversity and dynamics



Species diversity and larvae abundance varied along the hydrological cycle.



Mariac et al. 2022 *Mol Ecol*

Metabarcoding

Ichthyoplankton monitoring

Species-level ichthyoplankton dynamics

	Marañón	Flood start	Flood ends	Receding	Ucayali	Flood start	Flood ends	Receding
	C	Flood start	Flood end	Receding	C	Flood start	Flood end	Receding
Leporinus trimaculatus	2	0.2	0.7	0.0	1	0.9	0.1	0.0
Prochilodus nigricans	2	0.3	0.7	0.0	1	0.4	0.6	0.0
Prochilodus sp. aff. costatus	2	0.2	0.8	0.0	2	0.2	0.8	0.0
Psectrogaster amazonica	2	0.4	0.6	0.0	2	0.0	1.0	0.0
Rhaphiodon vulpinus	2	0.3	0.7	0.0	2	0.1	0.9	0.0
Thoracocharax stellatus †	2	0.2	0.8	0.0	2	0.0	1.0	0.0
Curimata cyprinoides	2	0.4	0.5	0.1	2	0.0	1.0	0.0
Curimatella meyeri	2	0.4	0.6	0.0	2	0.0	1.0	0.0
Hydrolycus scomberoides	2	0.4	0.6	0.0	2	0.3	0.7	0.0
Leporinus lacustris	2	0.4	0.5	0.1	1	0.8	0.1	0.1
Potamorhina altamazonica	2	0.4	0.6	0.0	2	0.0	1.0	0.0
Tetragonopterus argenteus †	2	0.4	0.6	0.1	1	0.9	0.1	0.0
Colossoma macropomum	3	0.3	0.0	0.7	1	0.9	0.0	0.1
Leporinus fasciatus	3	0.2	0.0	0.8	1	0.6	0.4	0.0
Semaprochilodus insignis	3	0.1	0.0	0.9				
Triportheus albus	3	0.0	0.0	0.9	3	0.1	0.0	0.9
Rhytiodus microlepis					1	0.0	1.0	0.0
Clupeiformes								
Pellona castelnaeana	3	0.4	0.0	0.6	1	0.9	0.1	0.0
Pellona flavipinnis	3	0.3	0.2	0.5	3	0.1	0.2	0.7
Perciformes								
Plagioscion auratus	3	0.3	0.0	0.7	3	0.0	0.0	1.0
Plagioscion squamosissimus	3	0.1	0.4	0.5	3	0.2	0.2	0.6
Siluriformes								
Calophysus macropterus	1	0.8	0.2	0.0	1	1.0	0.0	0.0
Hypophthalmus edentatus	1	0.7	0.1	0.2	1	0.9	0.1	0.0
Pimelodus sp. B CGD–2016 †	3	0.4	0.1	0.4	3	0.2	0.1	0.7
Pimelodus sp. C CGD–2016 †	1	0.7	0.1	0.3	3	0.4	0.0	0.6
Pseudoplatystoma tigrinum †	1	0.8	0.2	0.0	1	0.4	0.6	0.0
Amblydoras gonzalezi †	2	0.0	1.0	0.0	2	0.0	1.0	0.0

Relative read abundance of taxa in the samples was used to infer timing and duration of reproduction.

Mariac et al. 2022 Mol Ecol

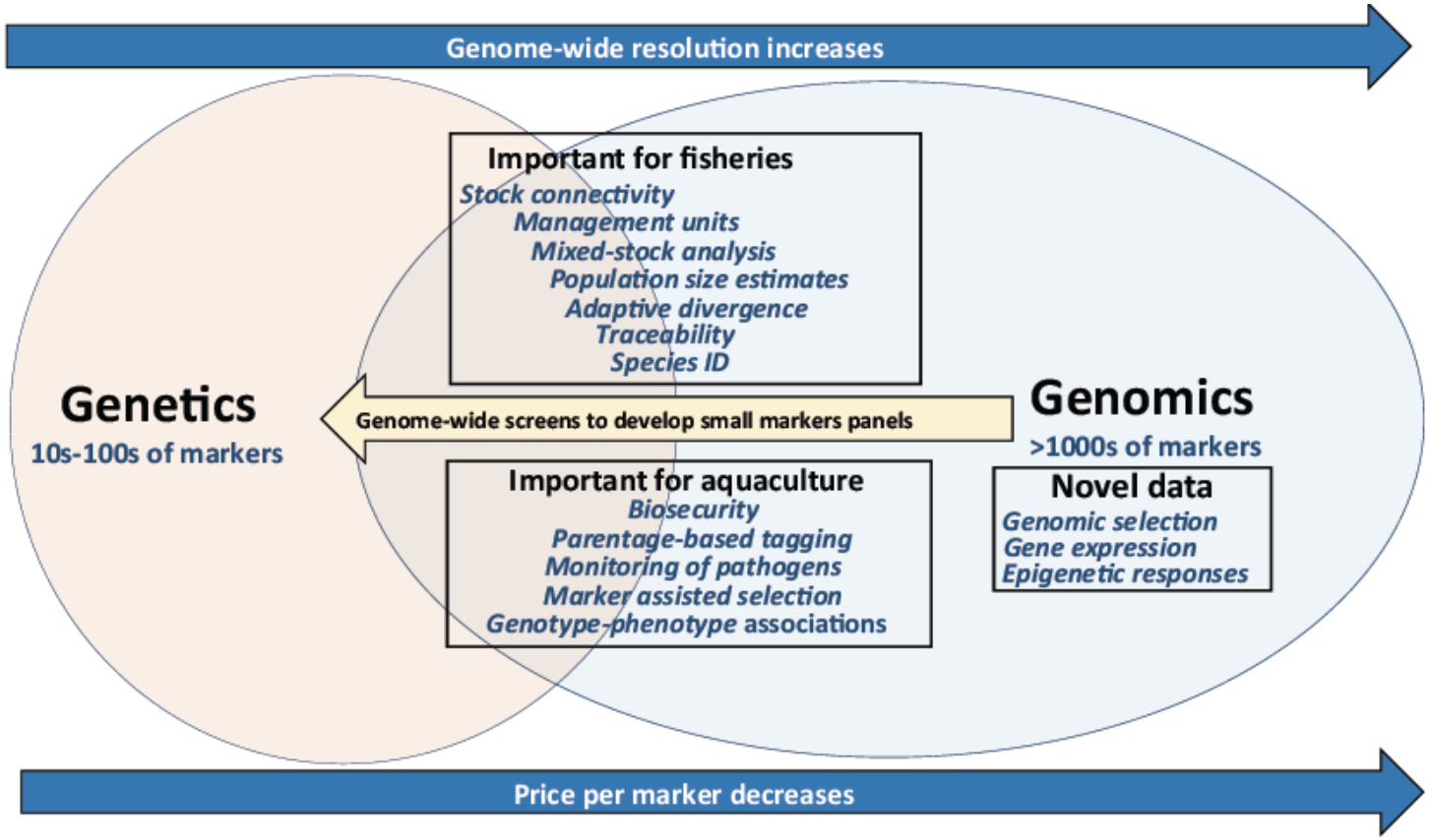
What can NGS do for you?

1. Choose your question

2. Decide on the samples needed

3. Decide on marker type and number

4. Check budget



Trends in Ecology & Evolution

Bernatchez et al. 2017 *TREE*

Questions?