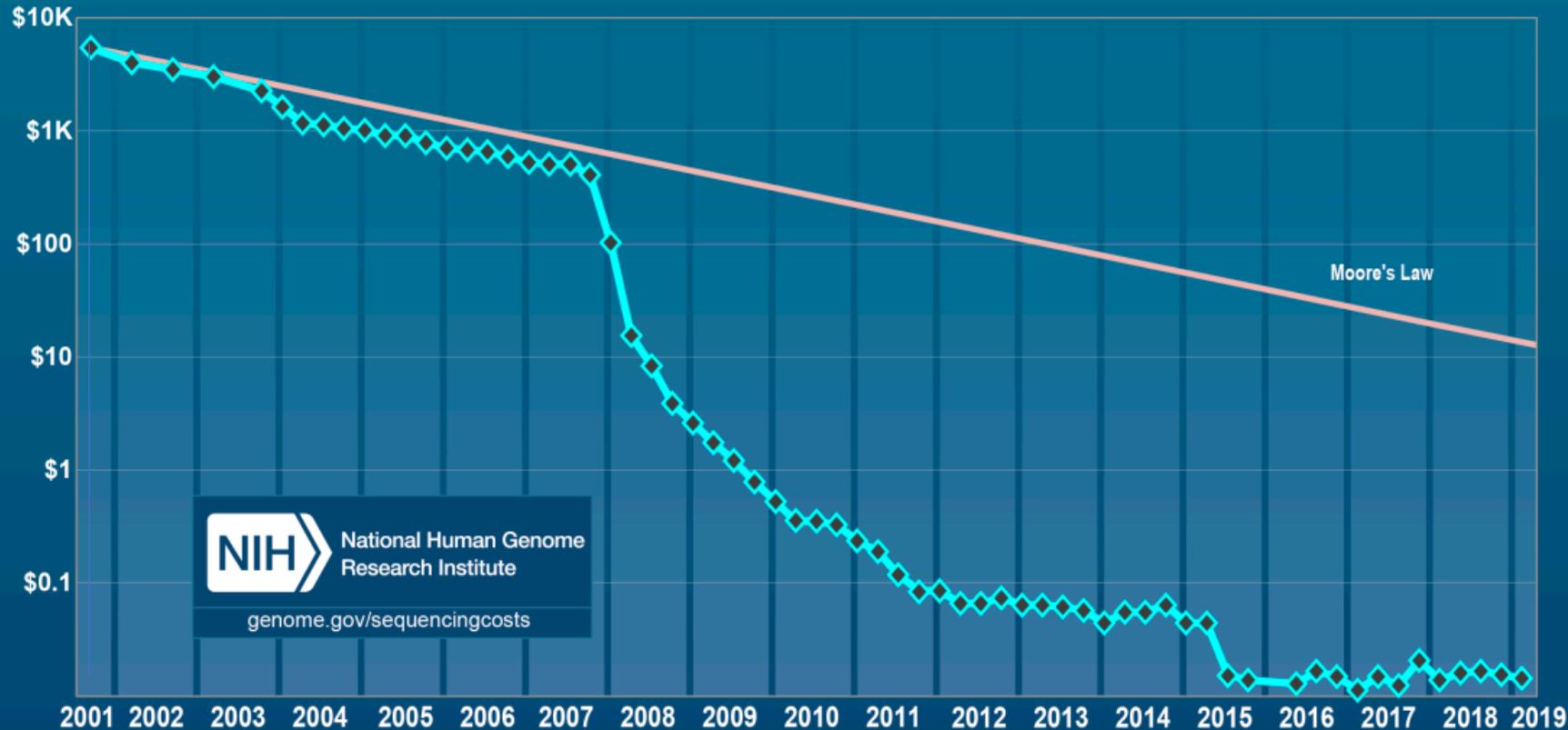
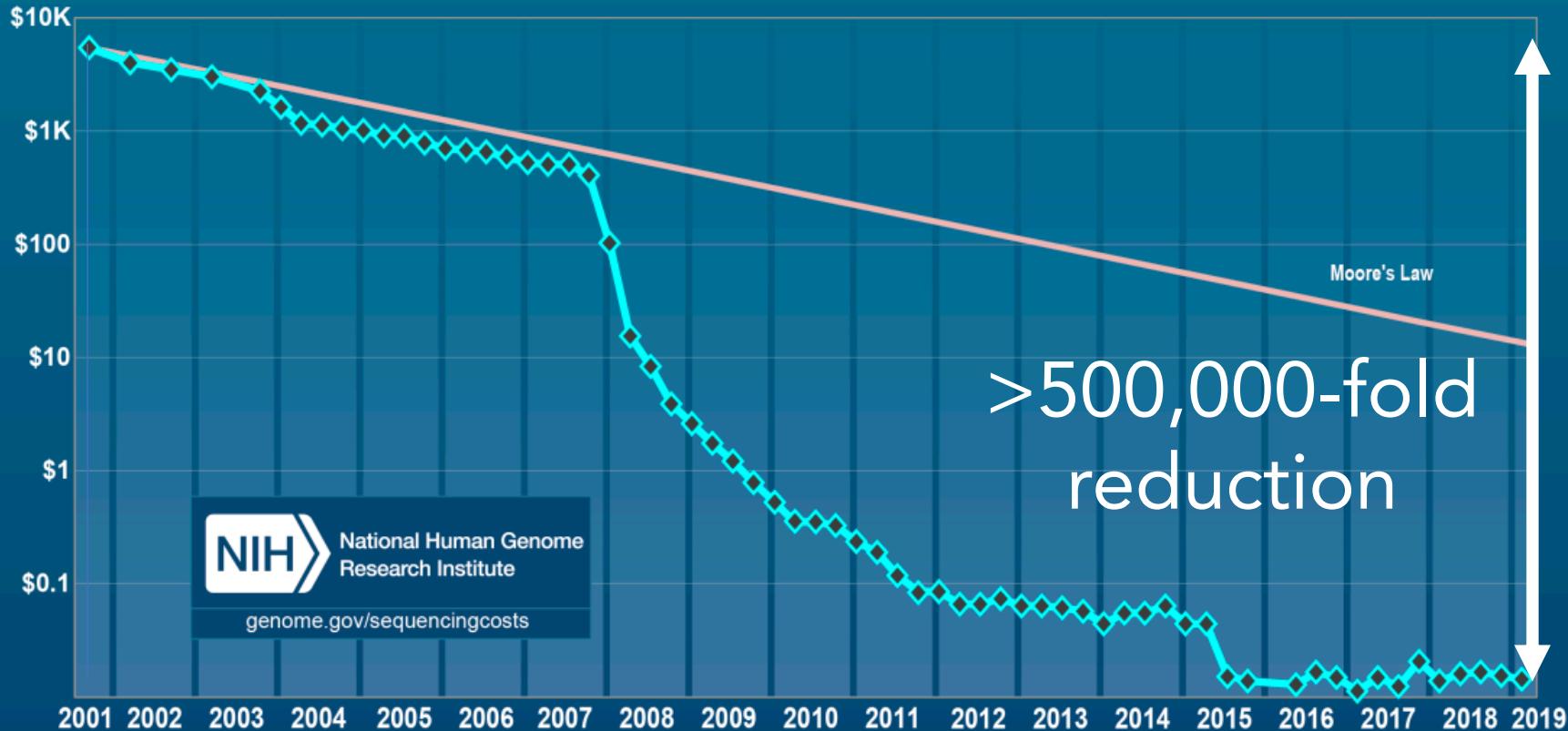


Introduction to low-coverage whole genome sequencing

Cost per Raw Megabase of DNA Sequence

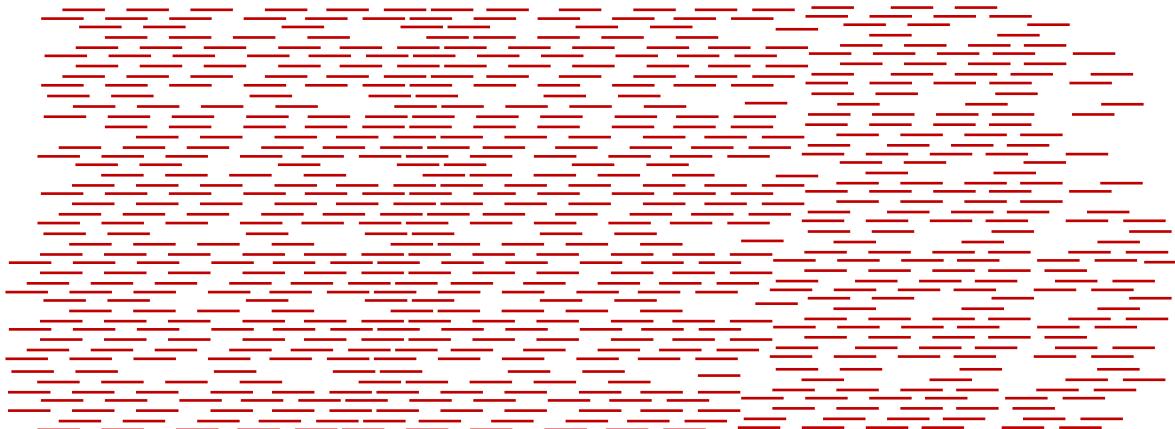


Cost per Raw Megabase of DNA Sequence

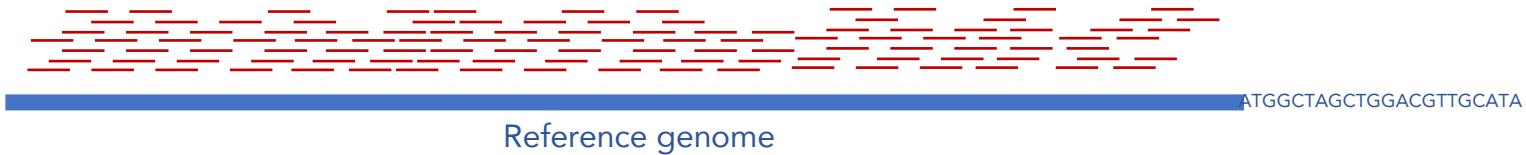


How to distribute sequencing effort?

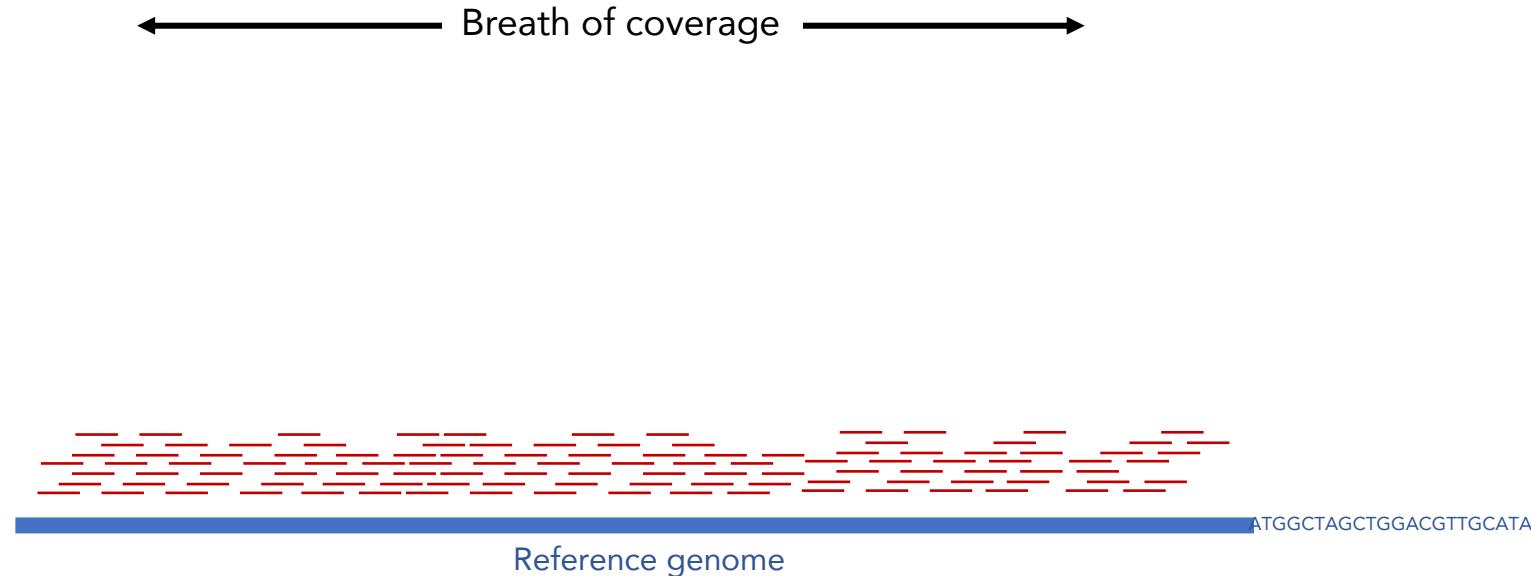
Raw sequence reads



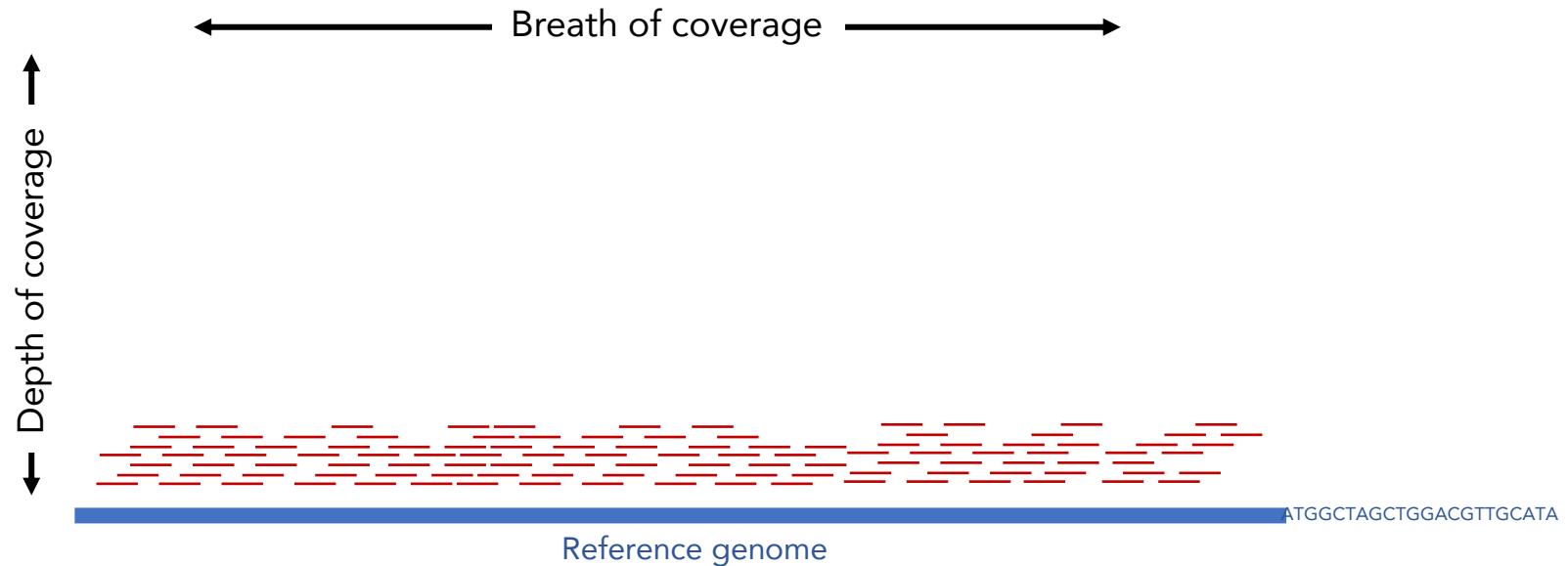
How to distribute sequencing effort?



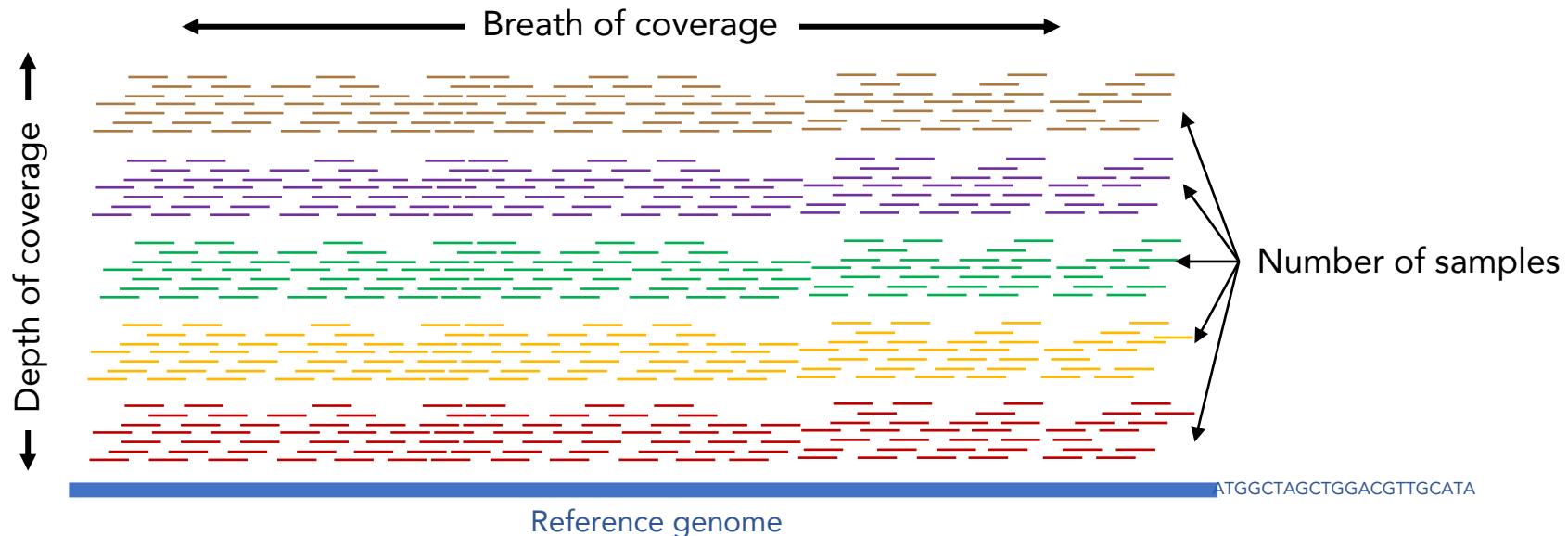
How to distribute sequencing effort?



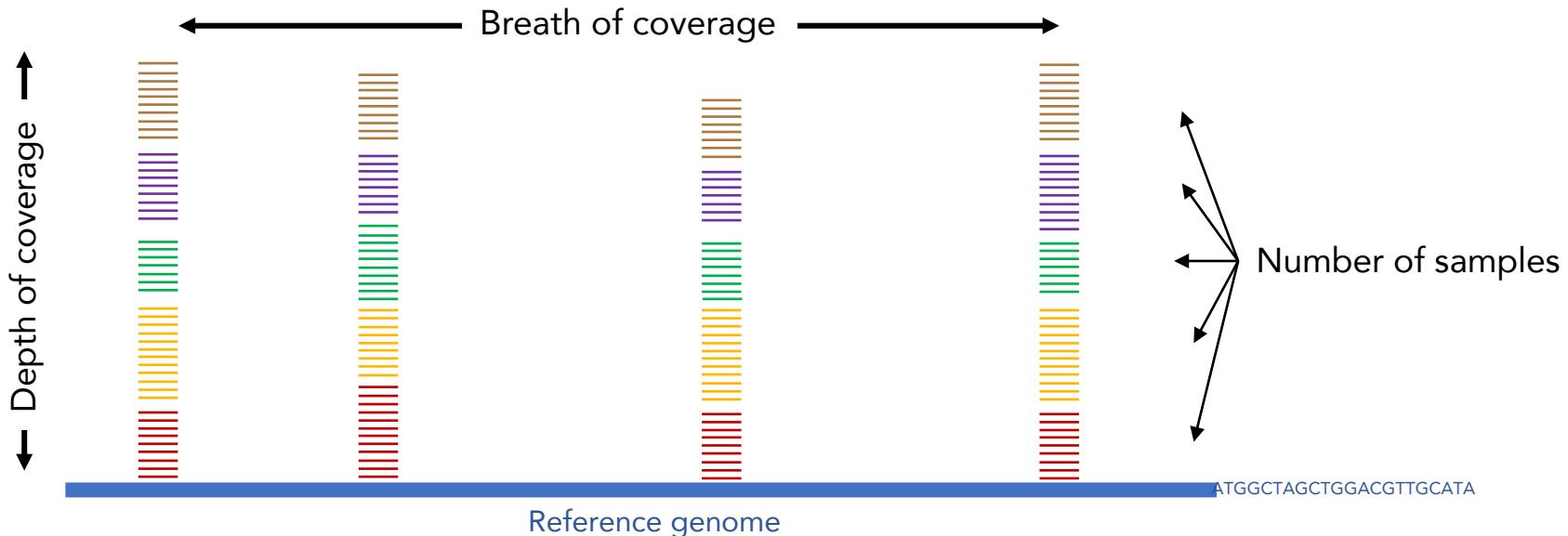
How to distribute sequencing effort?



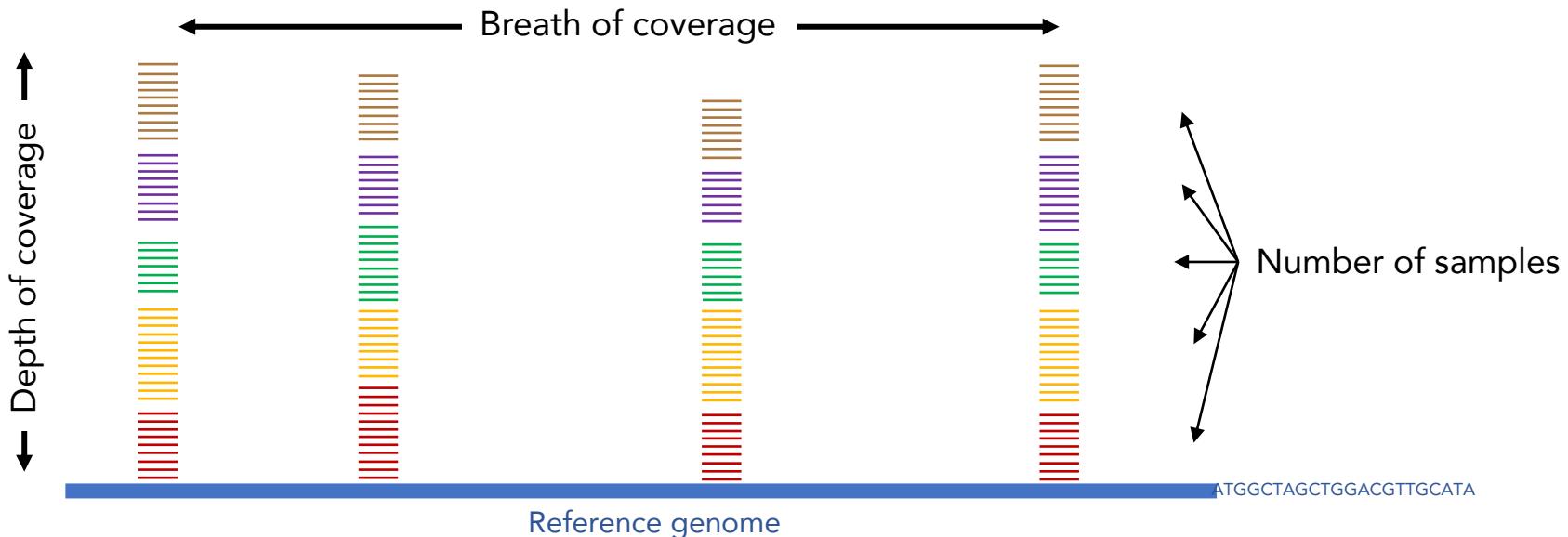
How to distribute sequencing effort?



RAD-seq approach

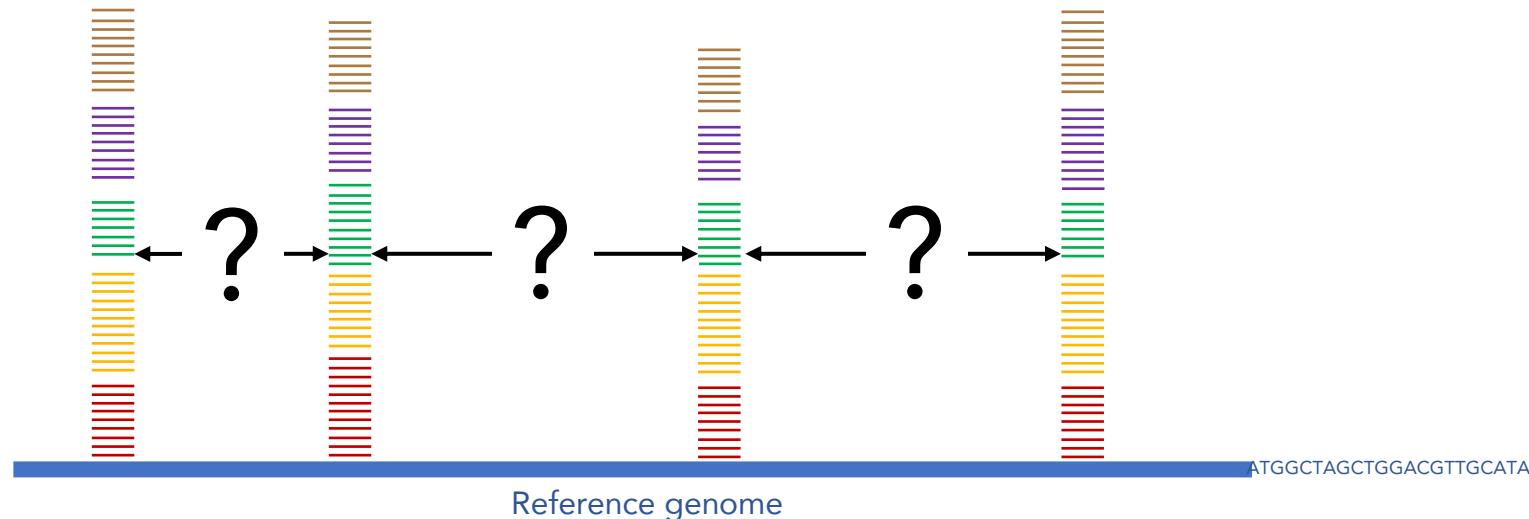


RAD-seq approach or SNP chip



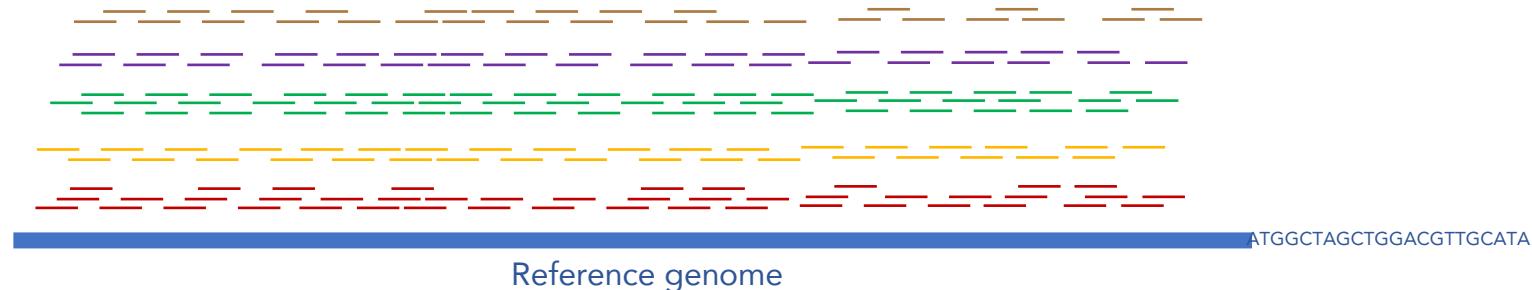
RAD-seq approach or SNP chip

What are we missing?



The value of trading depth for breath

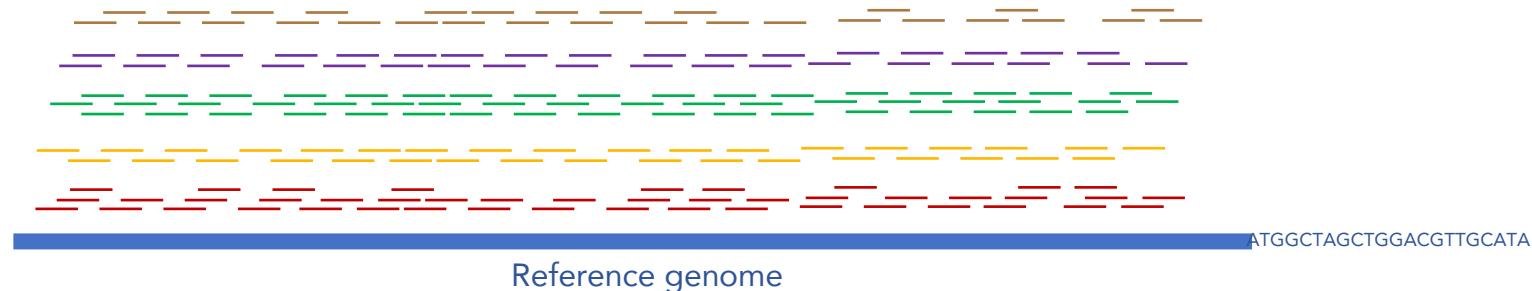
Maximizing the information content in our sequence data



The value of trading depth for breadth

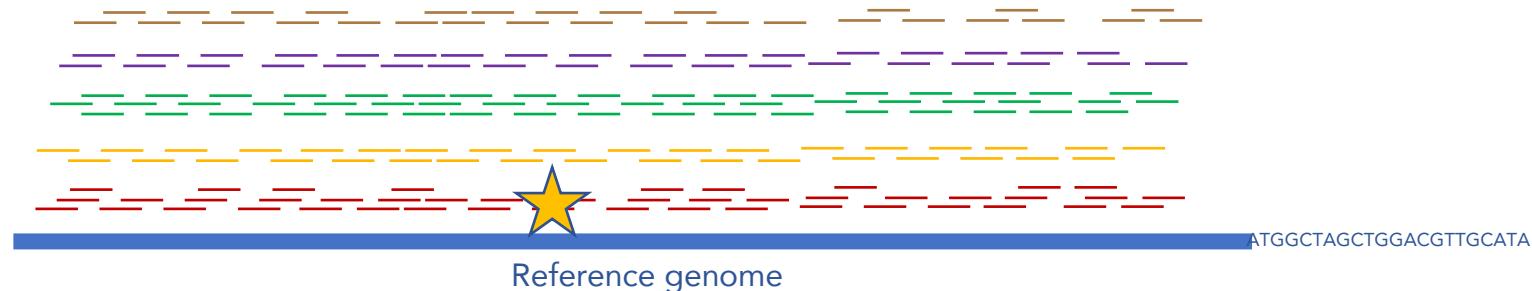
Maximizing the information content in our sequence data

But low-coverage precludes robust genotype calls



The value of trading depth for breath

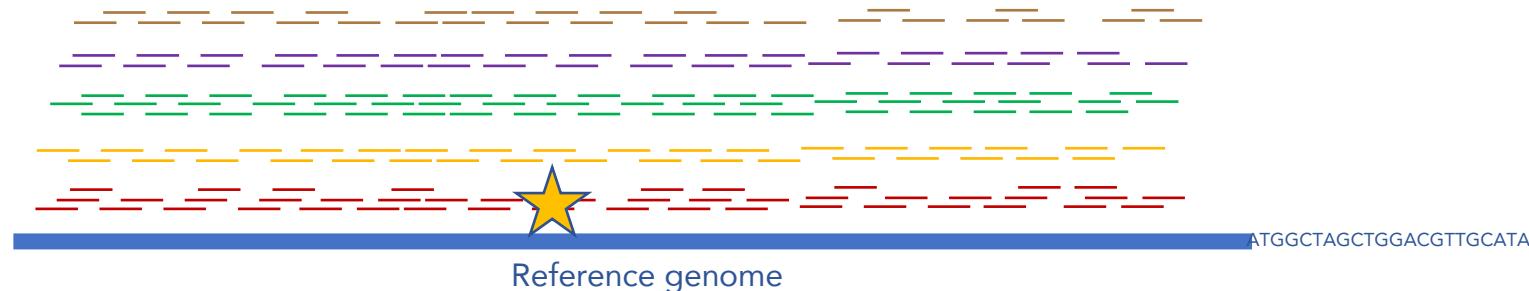
Do we care about the **genotype** at a particular **SNP** in a **particular individual**?



The value of trading depth for breath

Do we care about the **genotype** at a particular **SNP** in a **particular individual**?

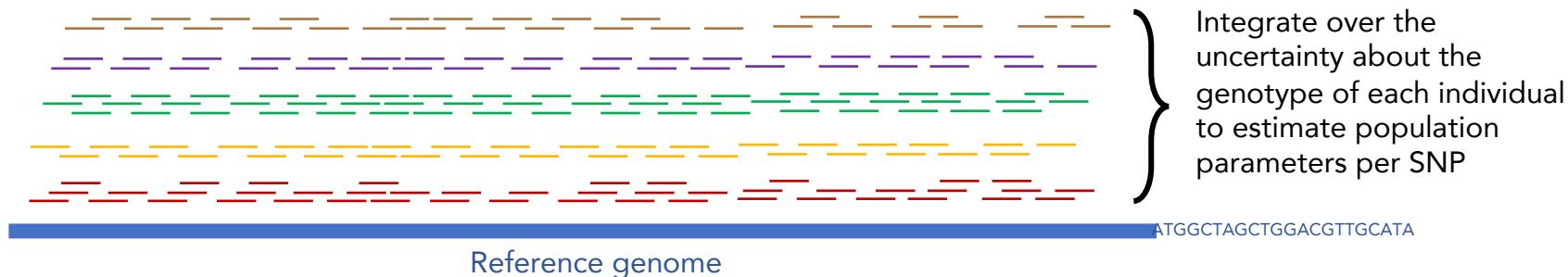
Often not



The value of trading depth for breadth

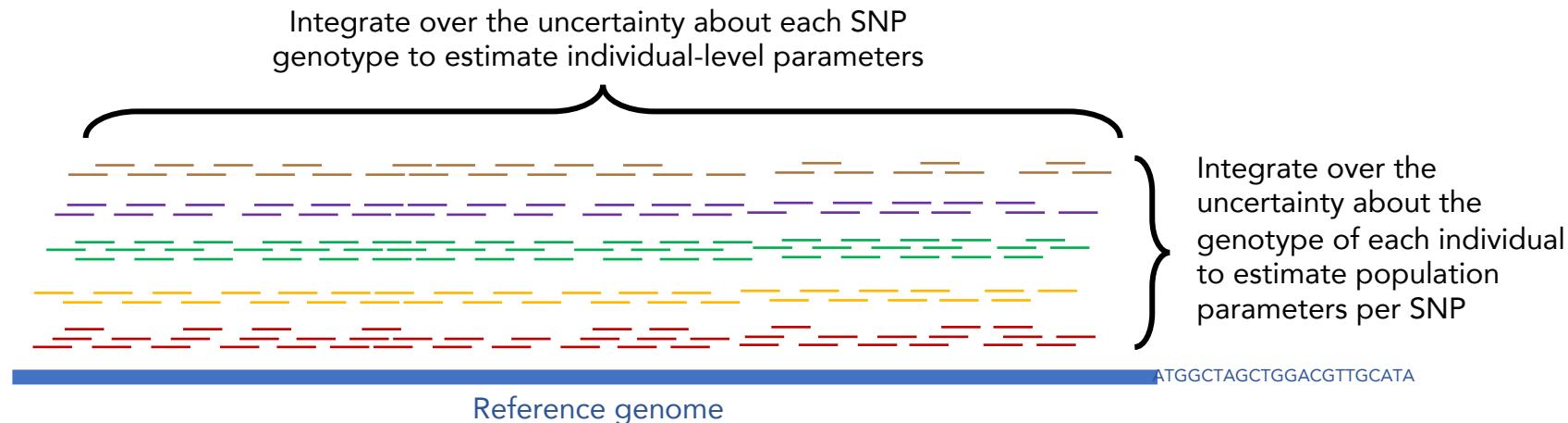
Do we care about the **genotype** at a particular **SNP** in a **particular individual**?

Often not



The value of trading depth for breadth

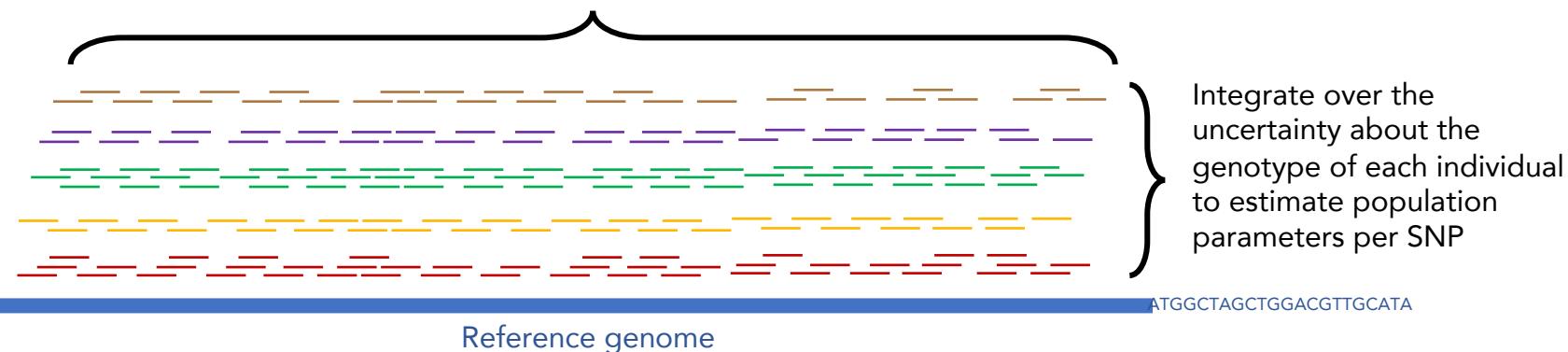
Do we care about the **genotype** at a particular SNP in a **particular individual**?



The value of trading depth for breadth

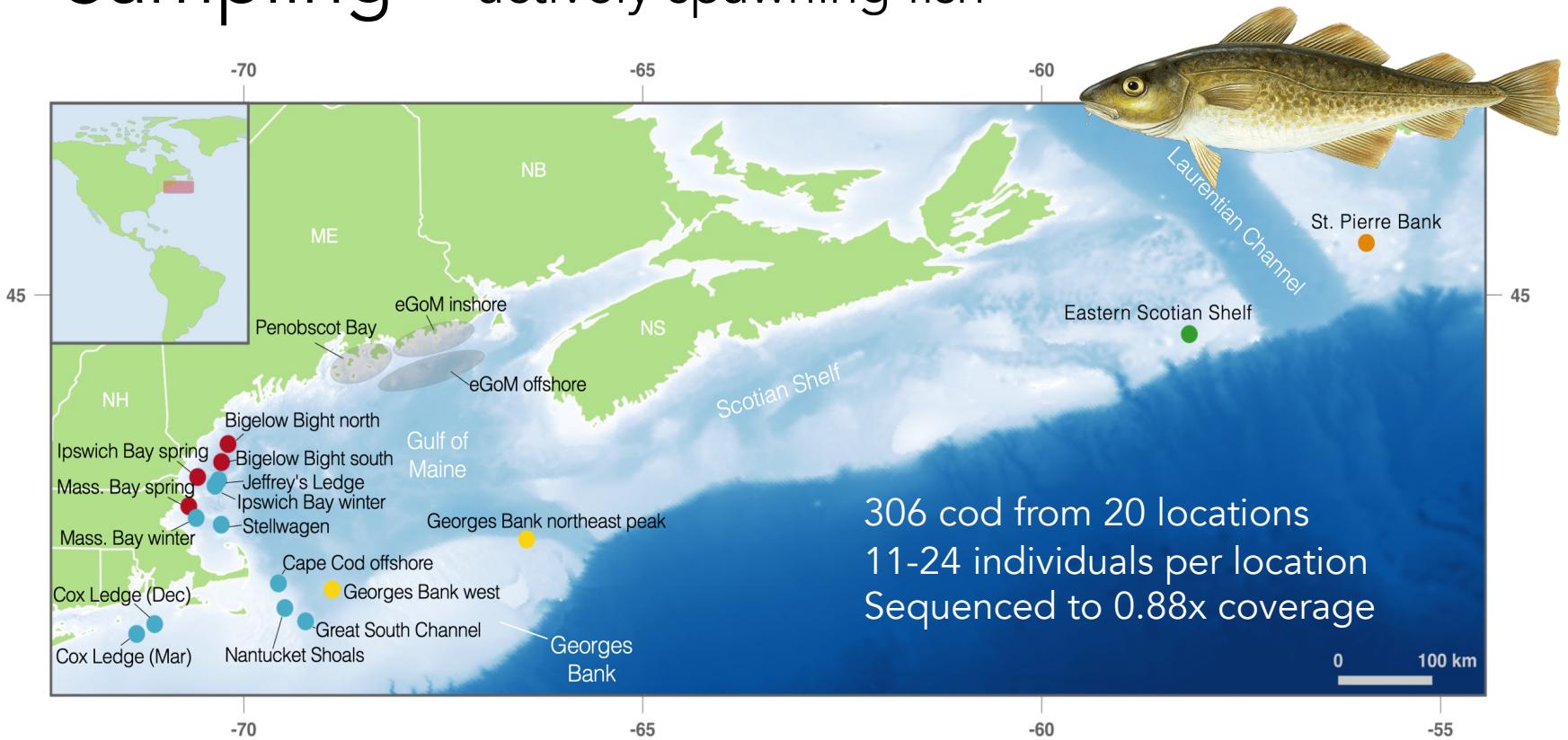
Robust inference when downstream analysis is conducted in a probabilistic framework that takes uncertainty into account

Integrate over the uncertainty about each SNP genotype to estimate individual-level parameters

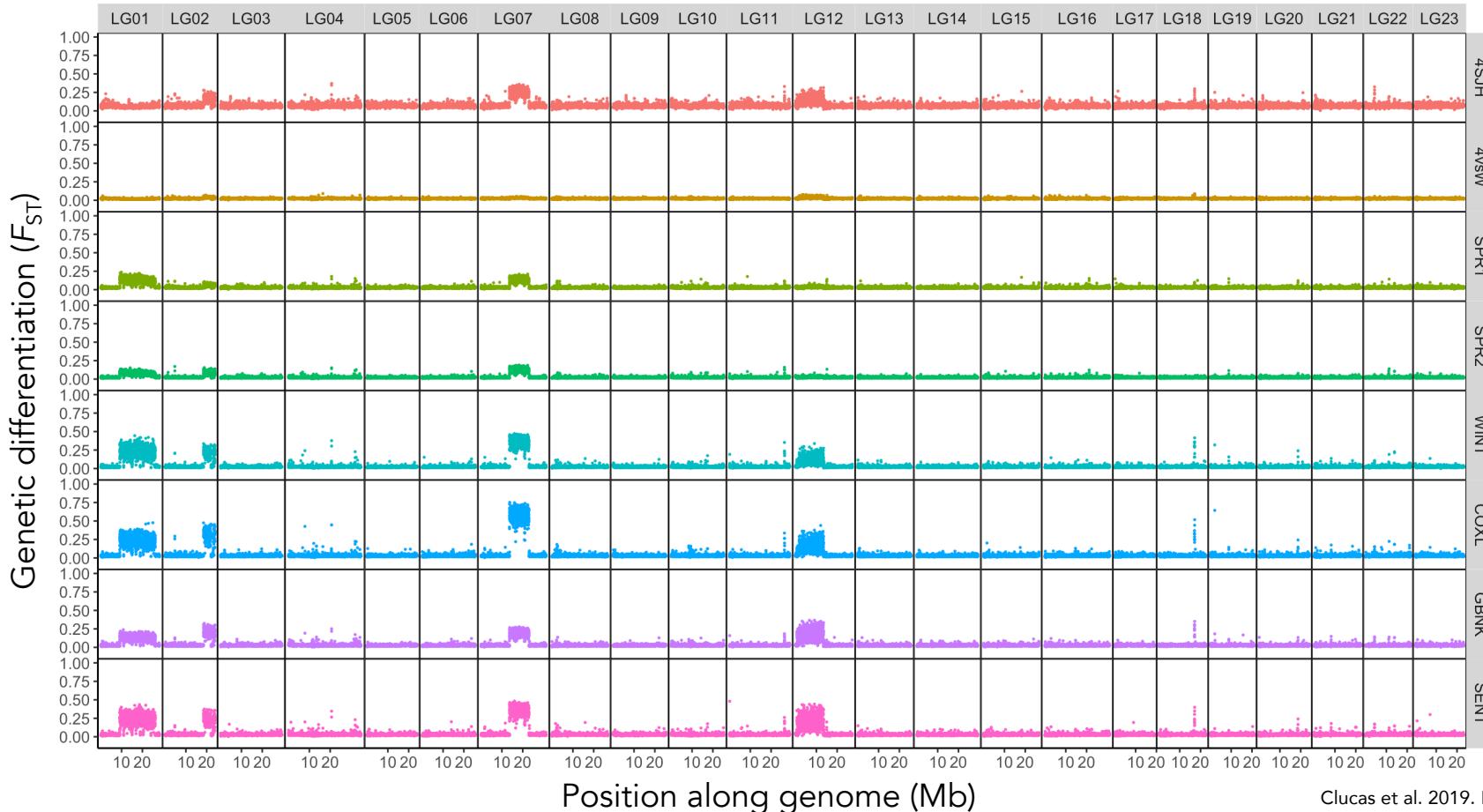


ATGGCTAGCTGGACGTTGCATA

Sampling – actively spawning fish



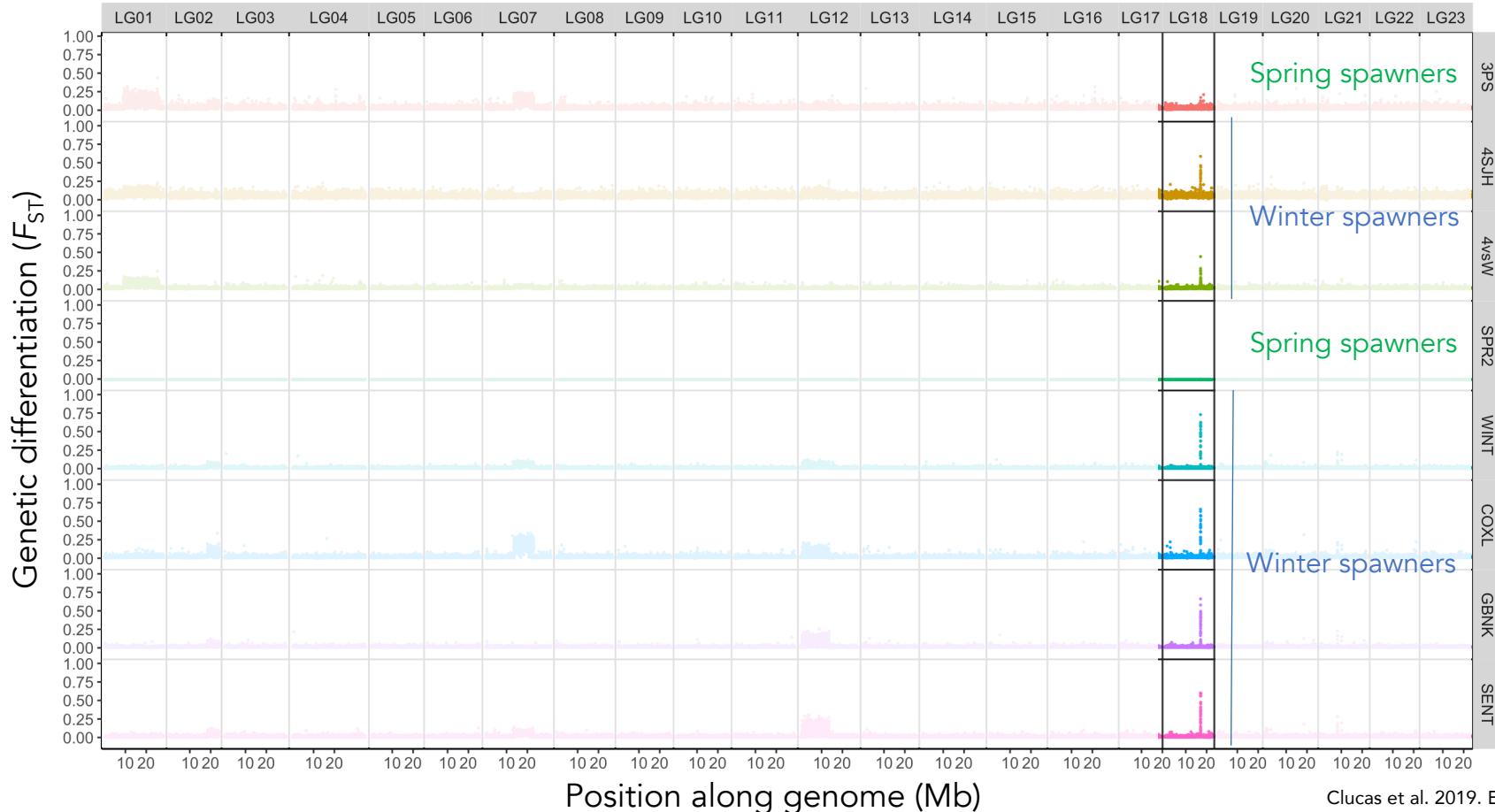
Pairwise comparisons to St. Pierre Bank, Canada (15 kb windows)



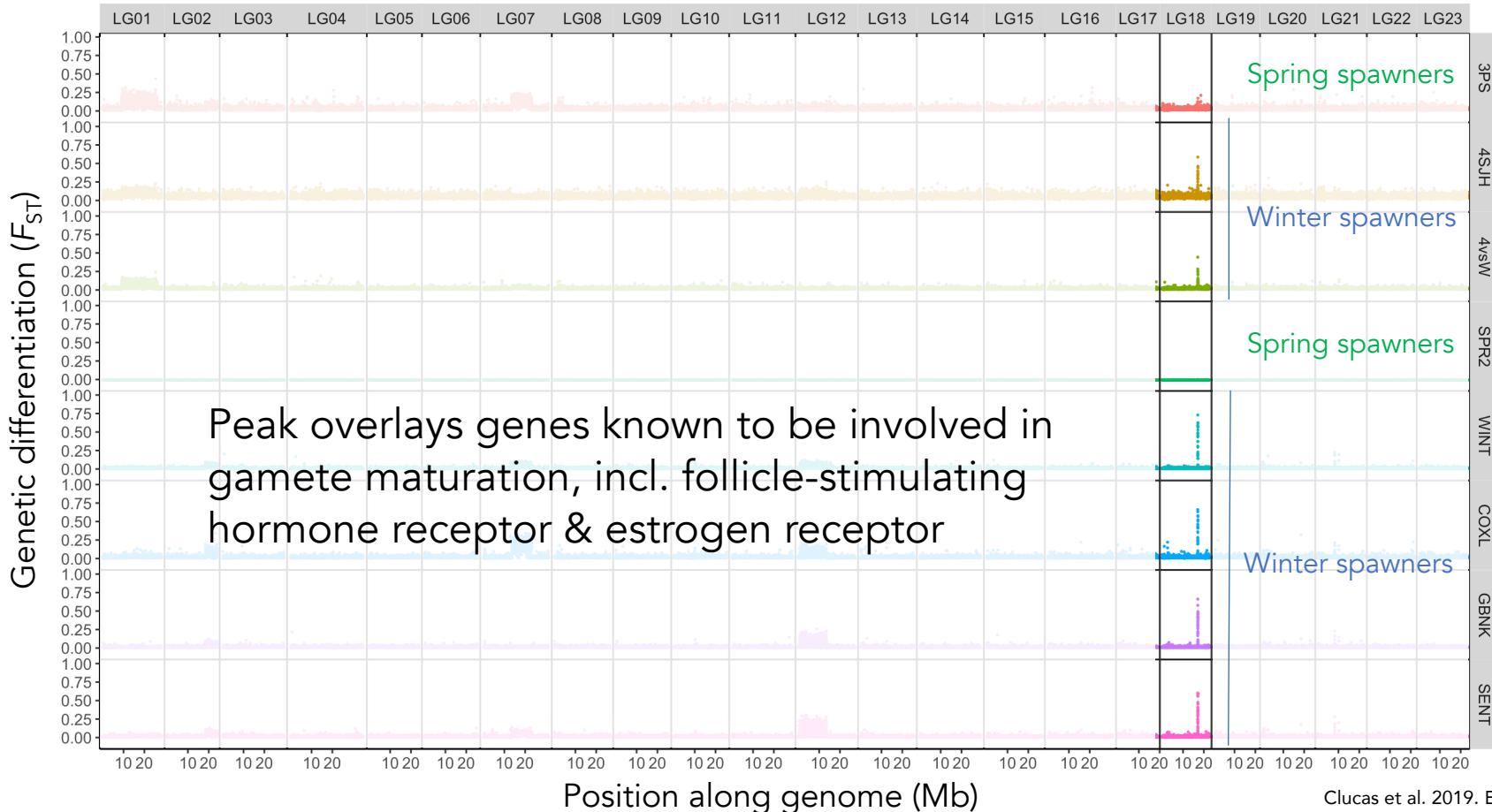
Position along genome (Mb)

Clucas et al. 2019. *Evol. Appl.*

Pairwise comparisons to the wGoM spring spawners (5 kb windows)

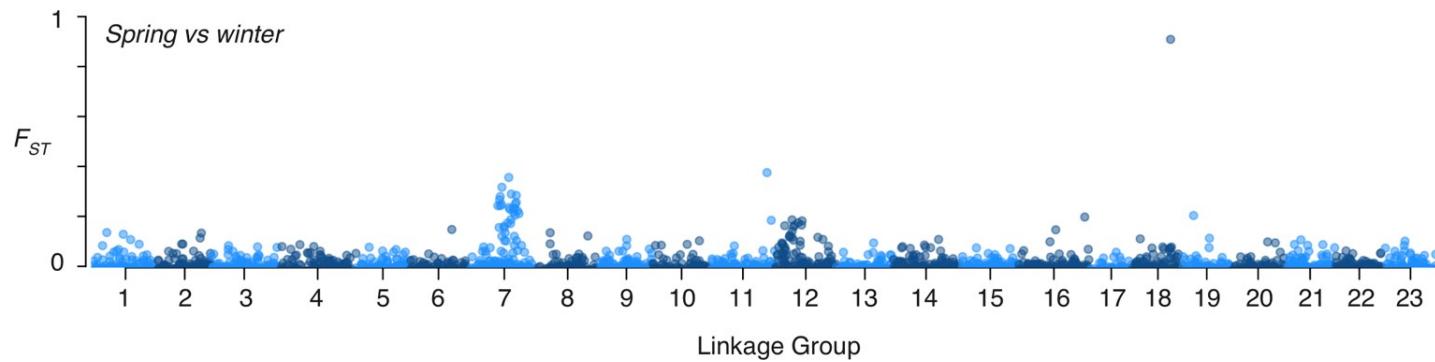


Pairwise comparisons to the wGoM spring spawners (5 kb windows)



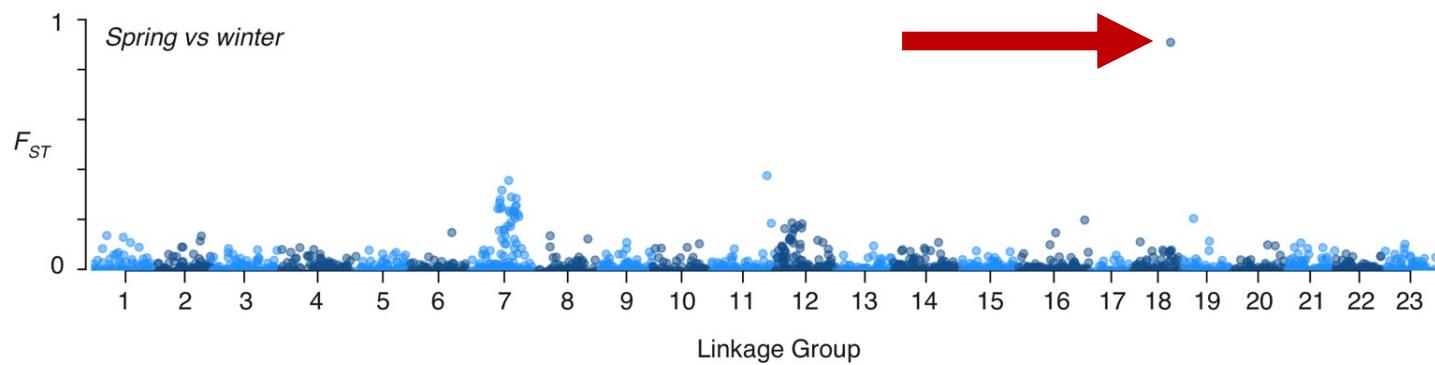
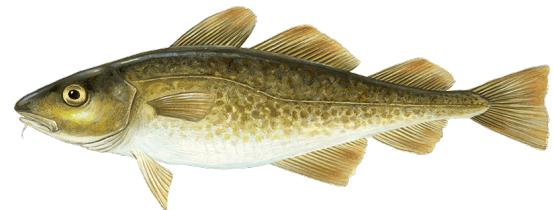
Compare to RADseq study of some of the same populations

3,128 SNPs

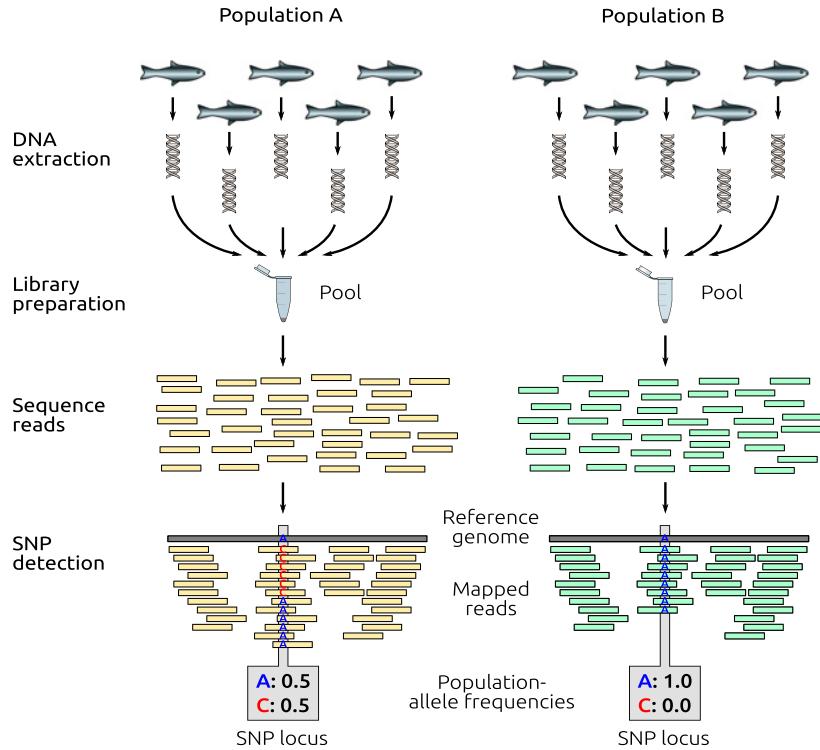


Compare to RADseq study of some of the same populations

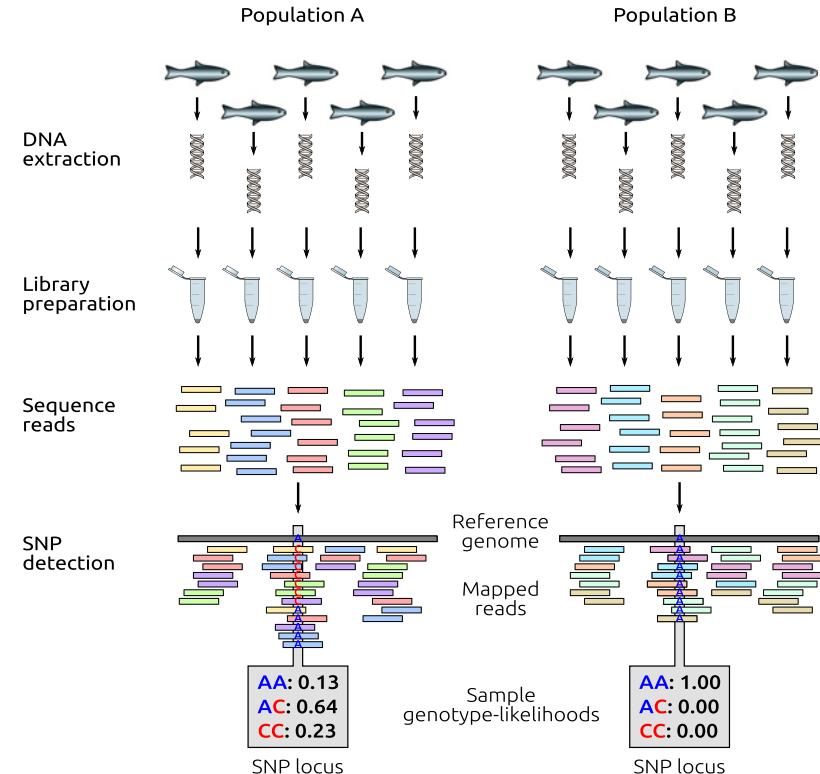
3,128 SNPs



Pool-seq

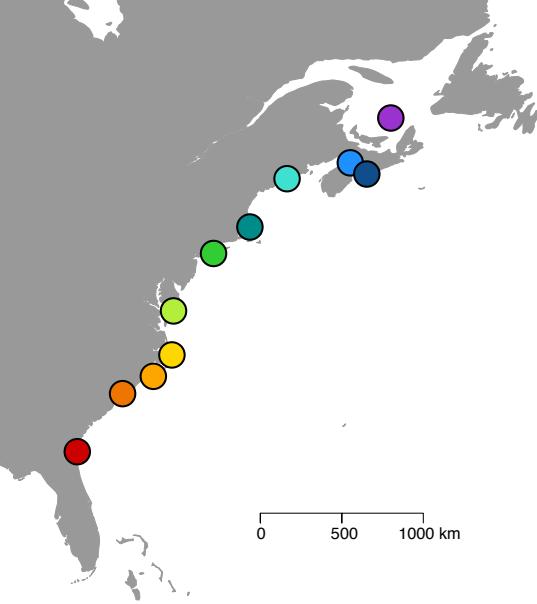


Low-coverage WGS



Advantages of lcWGS over Poolseq

- Can account for uneven contribution of samples in pool
- Retain individual information about e.g.
 - Cryptic population mixing
 - Relatedness
 - Estimates of individual heterozygosity or inbreeding
 - Linkage disequilibrium

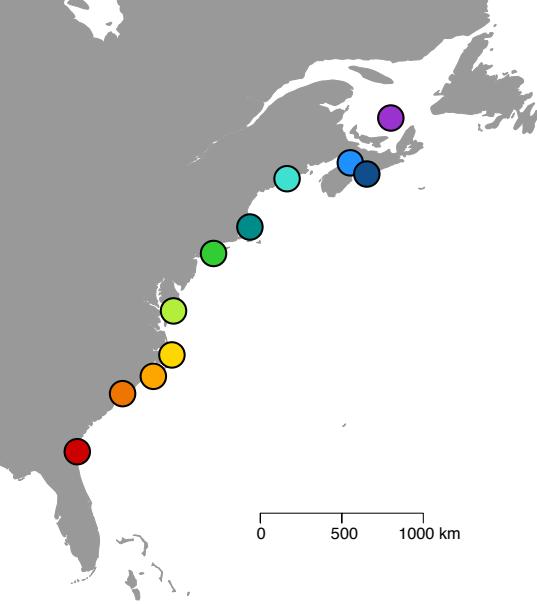


Individual-level analysis

Enables analysis of the genetic distance among individuals

n=50 silversides per site
(sampled during spring spawning)

Low-coverage 'in silico'
exome capture

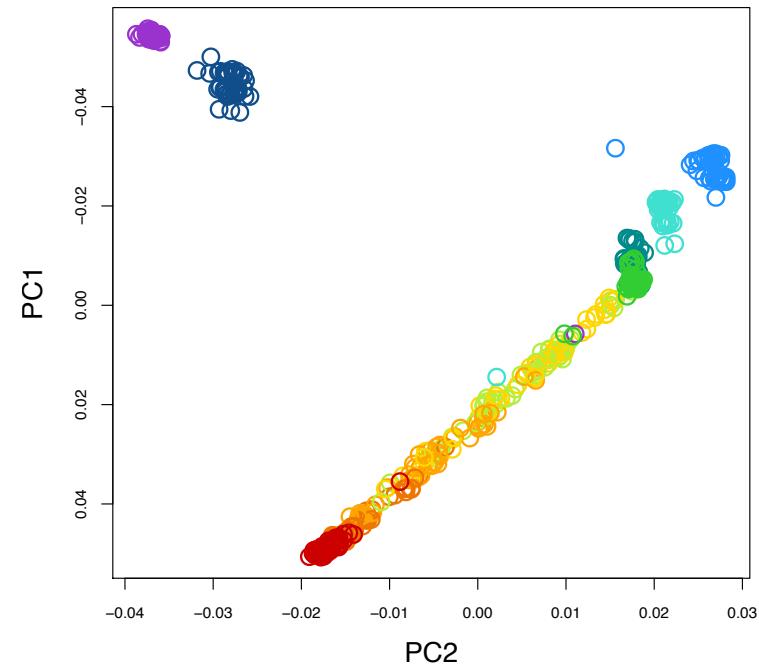


n=50 silversides per site
(sampled during spring spawning)

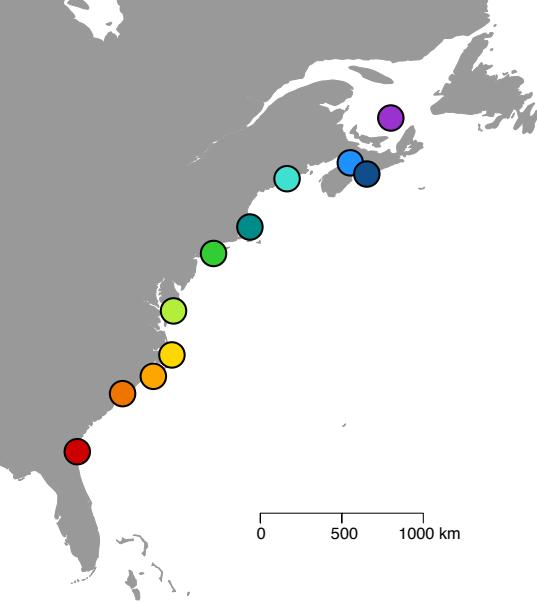
Low-coverage 'in silico'
exome capture (~1.3x)

Individual-level analysis

Enables analysis of the genetic distance among individuals



Multi-dimensional scaling of genetic distance (across 2.3 million SNPs)

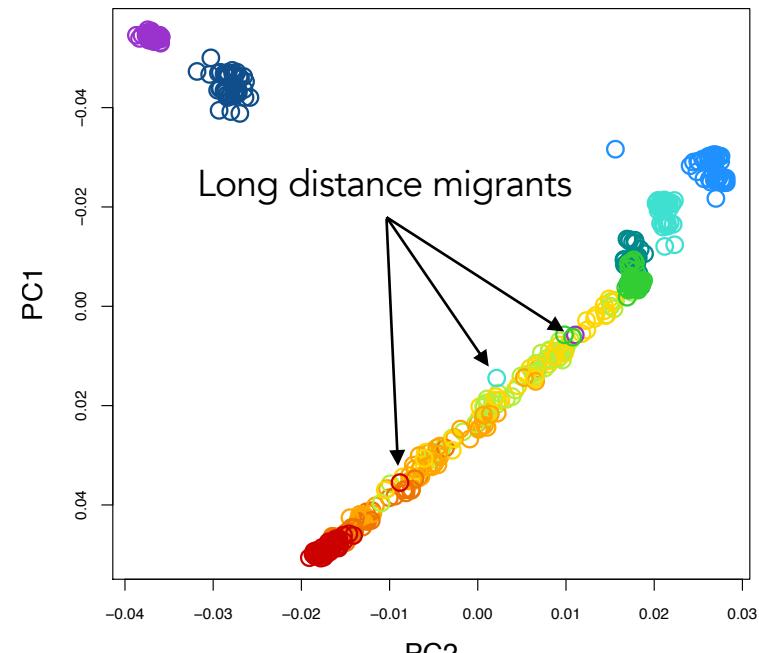


n=50 silversides per site
(sampled during spring spawning)

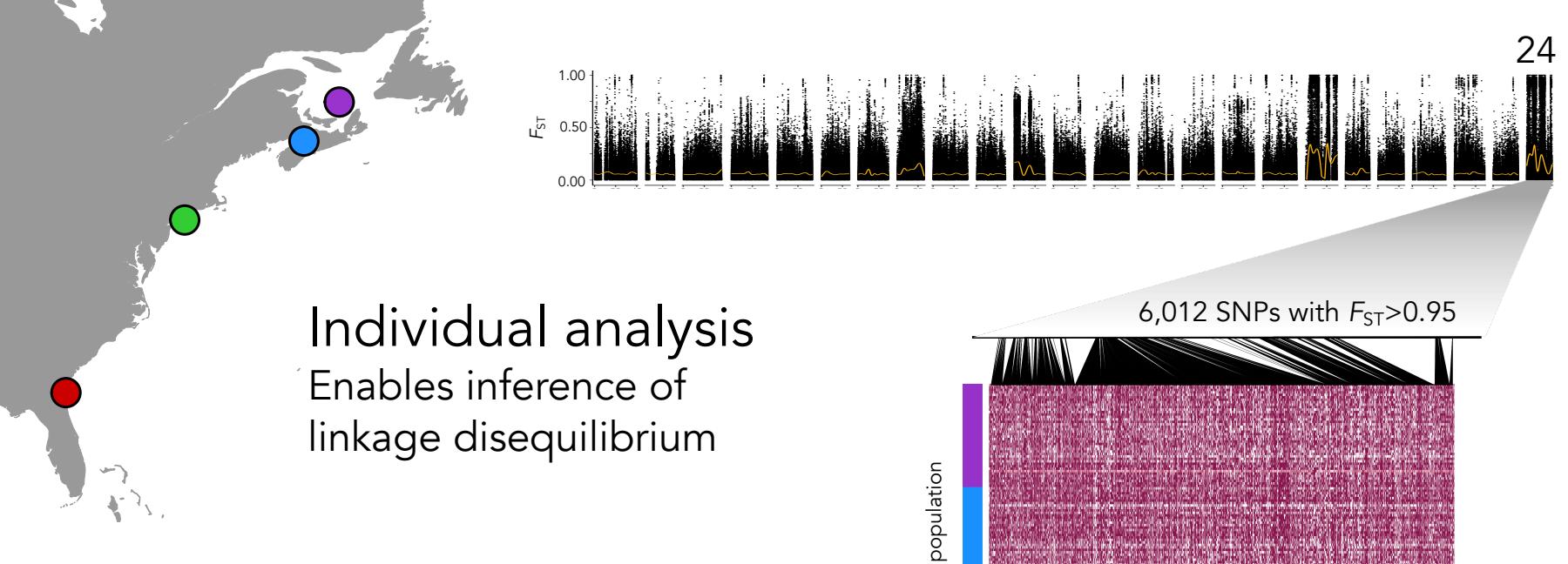
Low-coverage 'in silico' exome capture (~1.3x)

Individual-level analysis

Enables identification of divergent individuals and mixed samples



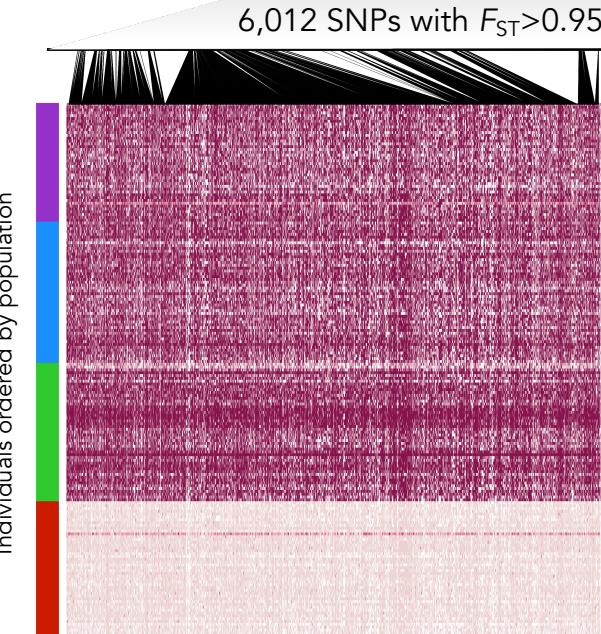
Multi-dimensional scaling of genetic distance (across 2.3 million SNPs)

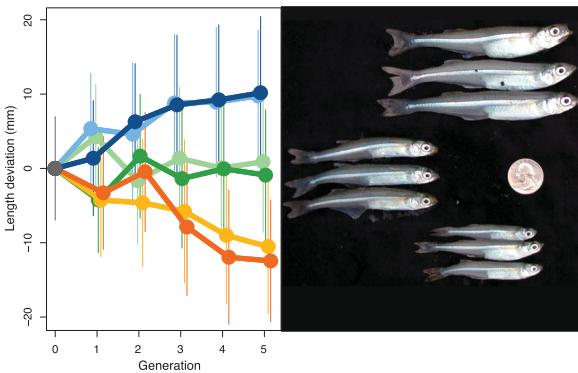


Individual analysis
Enables inference of
linkage disequilibrium

Most likely genotype for
each individual

- Northern homozygote
- Heterozygote
- Southern homozygote

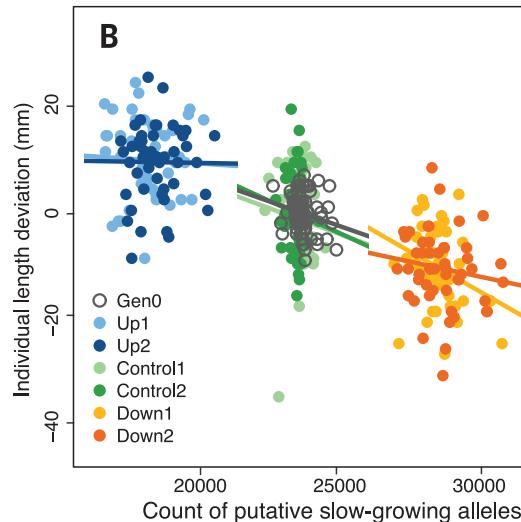




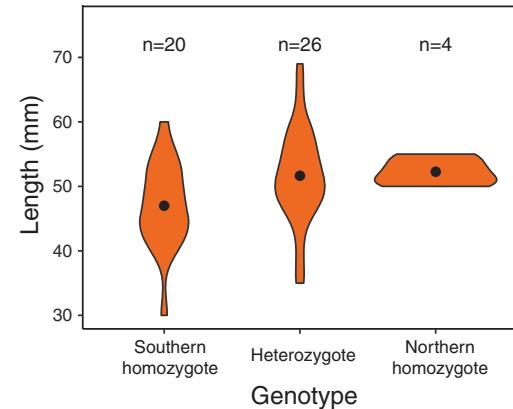
Individual-level analysis

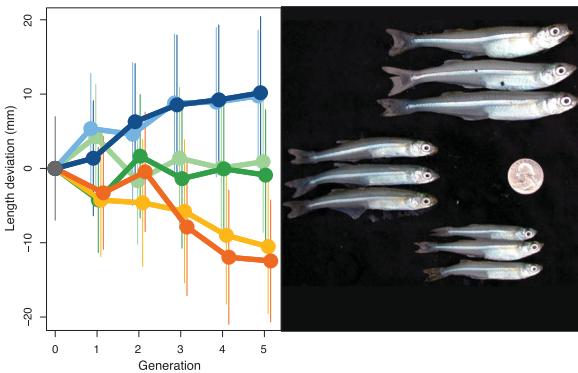
Enables individual genotype-phenotype association analysis

Polygenic score



Haplotyping inversions

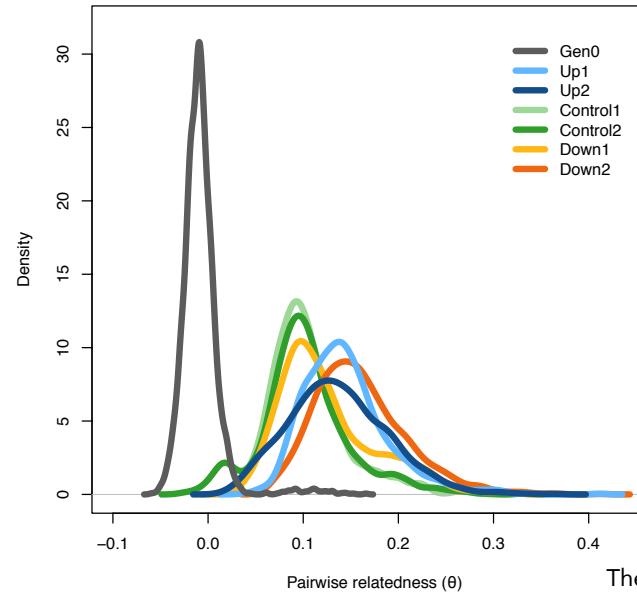




Individual-level analysis

Enables computation of relatedness and individual diversity statistics

Pairwise relatedness

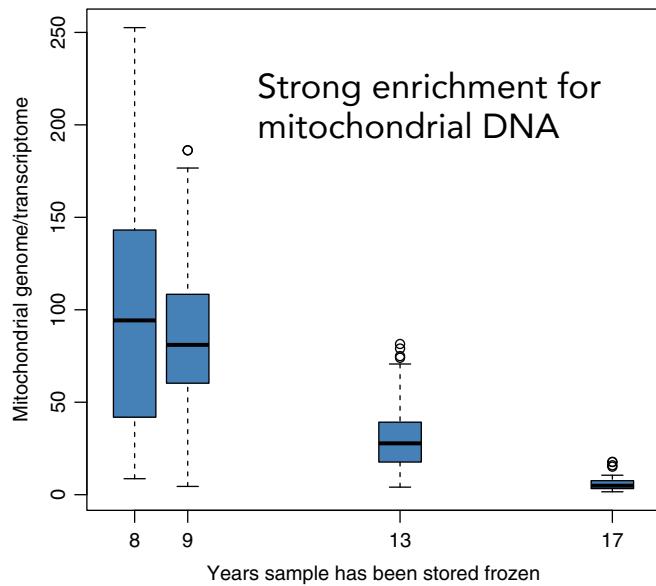


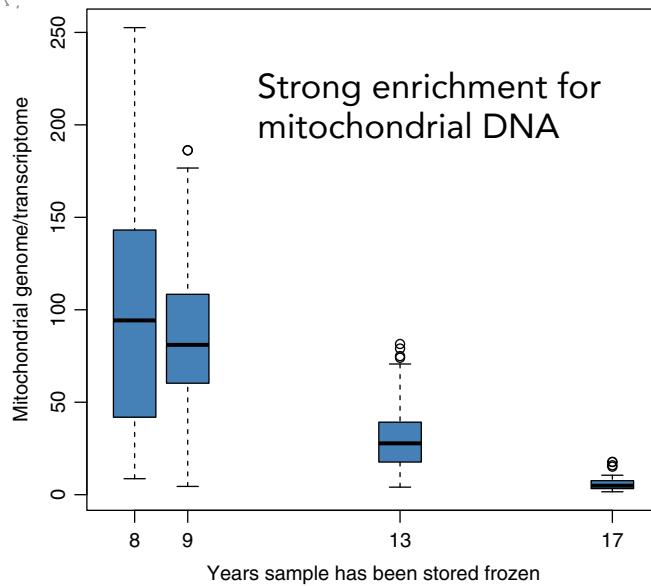
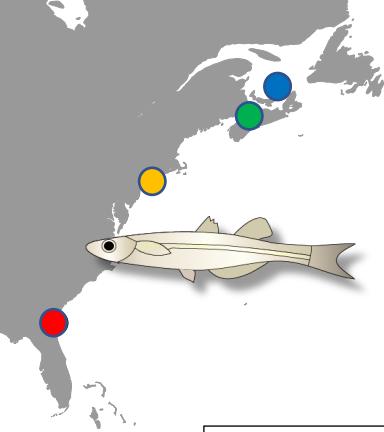
~50 silversides per population in generations 0 and 5

Low-coverage 'in silico' exome capture (~1.3x)

Individual-level analysis

Enables recovery of full mitochondrial genome sequences

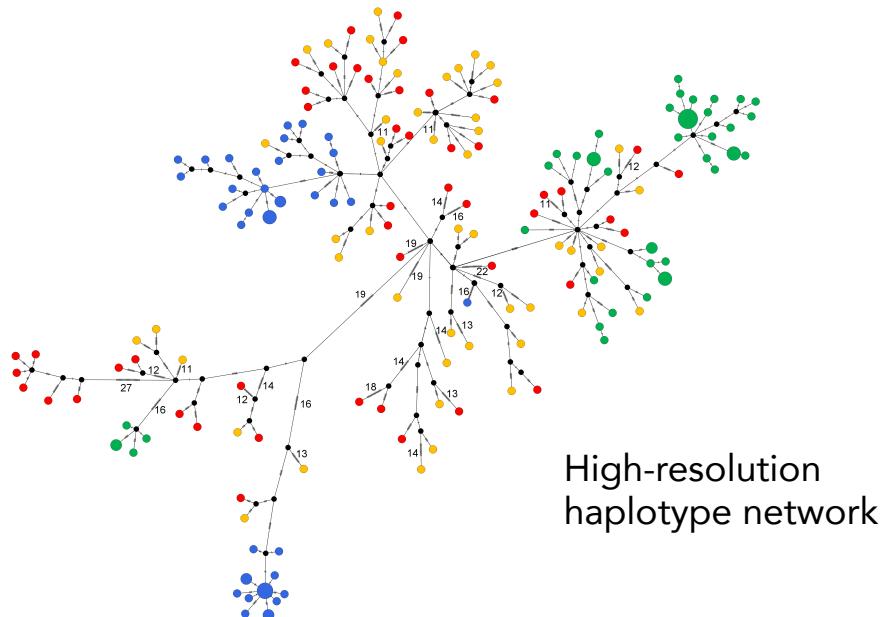




Therkildsen and Palumbi 2017. Mol. Ecol.

Individual-level analysis

Enables recovery of full mitochondrial genome sequences



Lou et al. 2018. Mar. Biol.

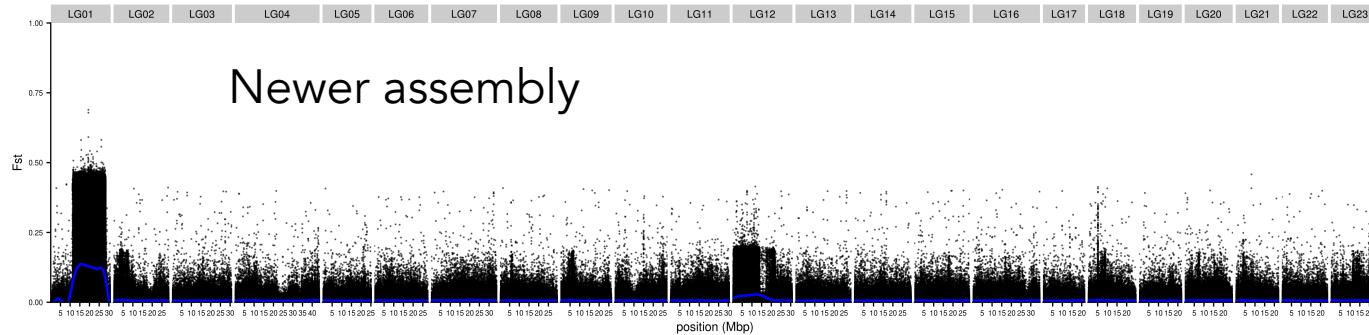
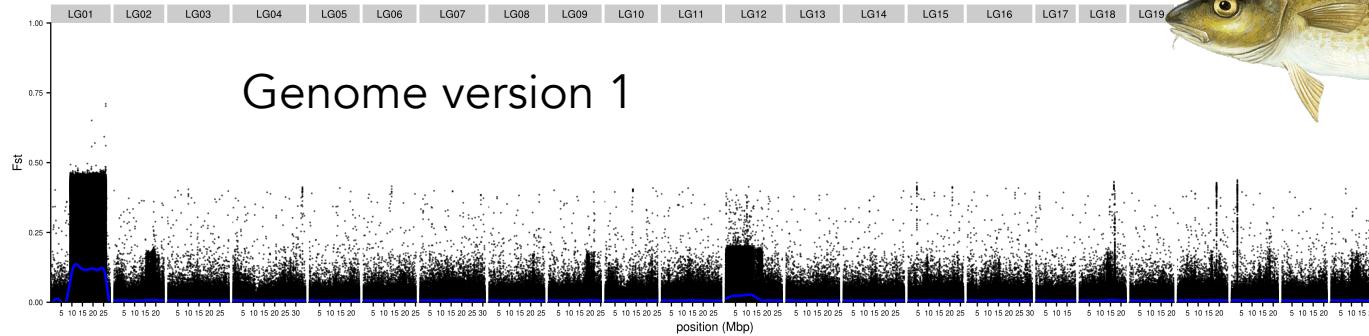
Cheap reads are short – We need a reference sequence for mapping

- *Do novo* genome assembly has become much more accessible even on modest budgets
- The quality of the new genome assemblies the literature is getting flooded with varies widely

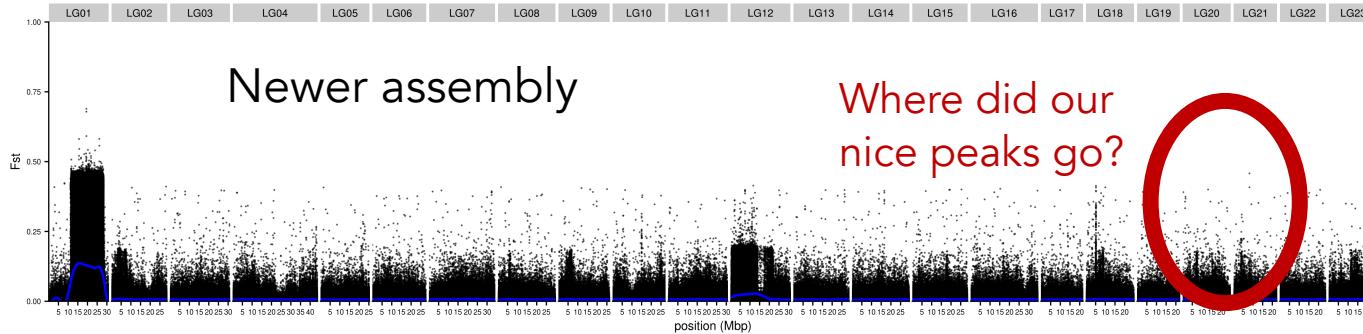
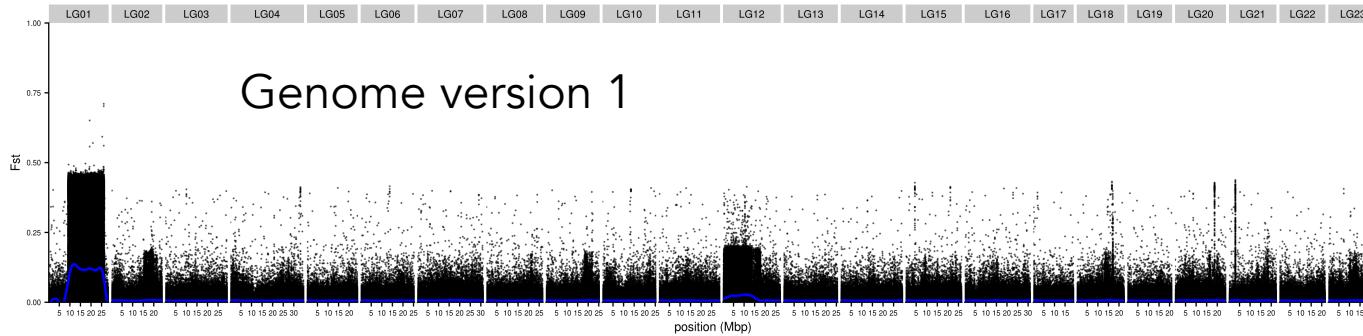
Discussion in breakout rooms

- How important is it to have a high-quality genome assembly?

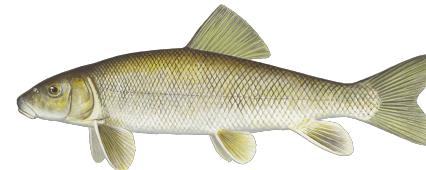
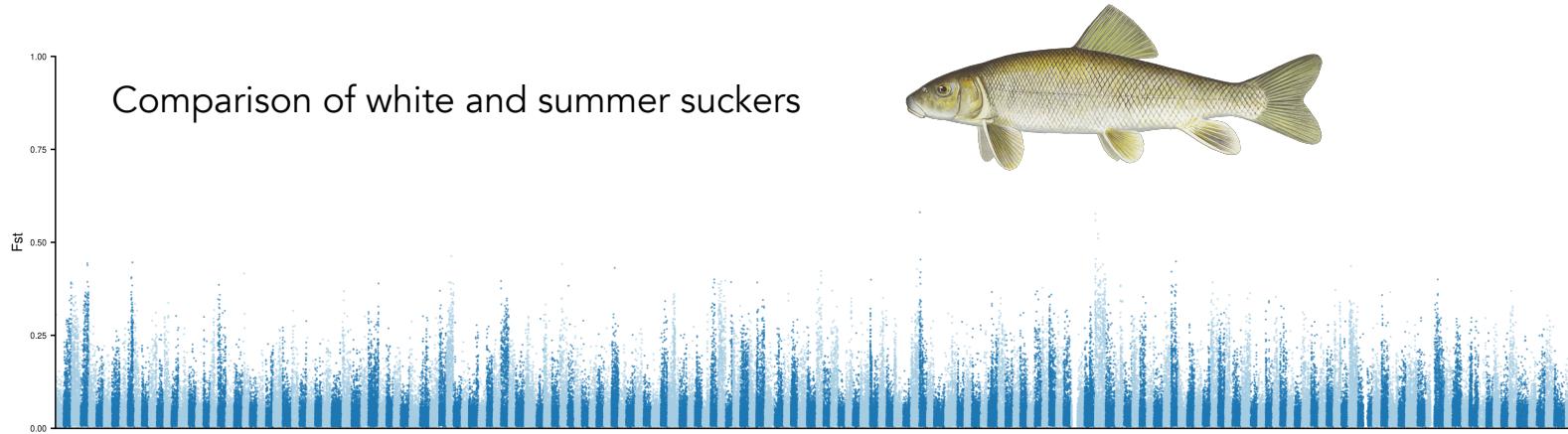
Examples from Atlantic cod



Examples from Atlantic cod

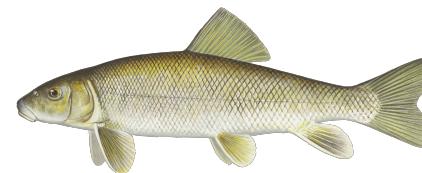


Highly fragmented genome assembly



Highly fragmented genome assembly

How do we even begin to make sense of this?



New technology for highly contiguous genome assembly

- Long-read sequencing
- Hi-C
- Optical mapping



VERTEBRATE
GENOMES
PROJECT

A PROJECT OF THE G10K CONSORTIUM

PHASE I TECHNOLOGY WHO WE ARE PUBLICATIONS NEWS EVENTS WAYS TO HELP CONTACT ERGA

DONATE

Welcome to the Vertebrate Genomes Project (VGP)

A mission to generate near error-free and complete genome assemblies of all living vertebrate species

LIST OF ~70,000 EXTANT VERTEBRATE SPECIES

GENOME ARK

All VGP data, raw reads, mito assemblies

VGP ENSEMBL

Annotations

Set of primary assemblies

UCSC VGP GB

Annotations

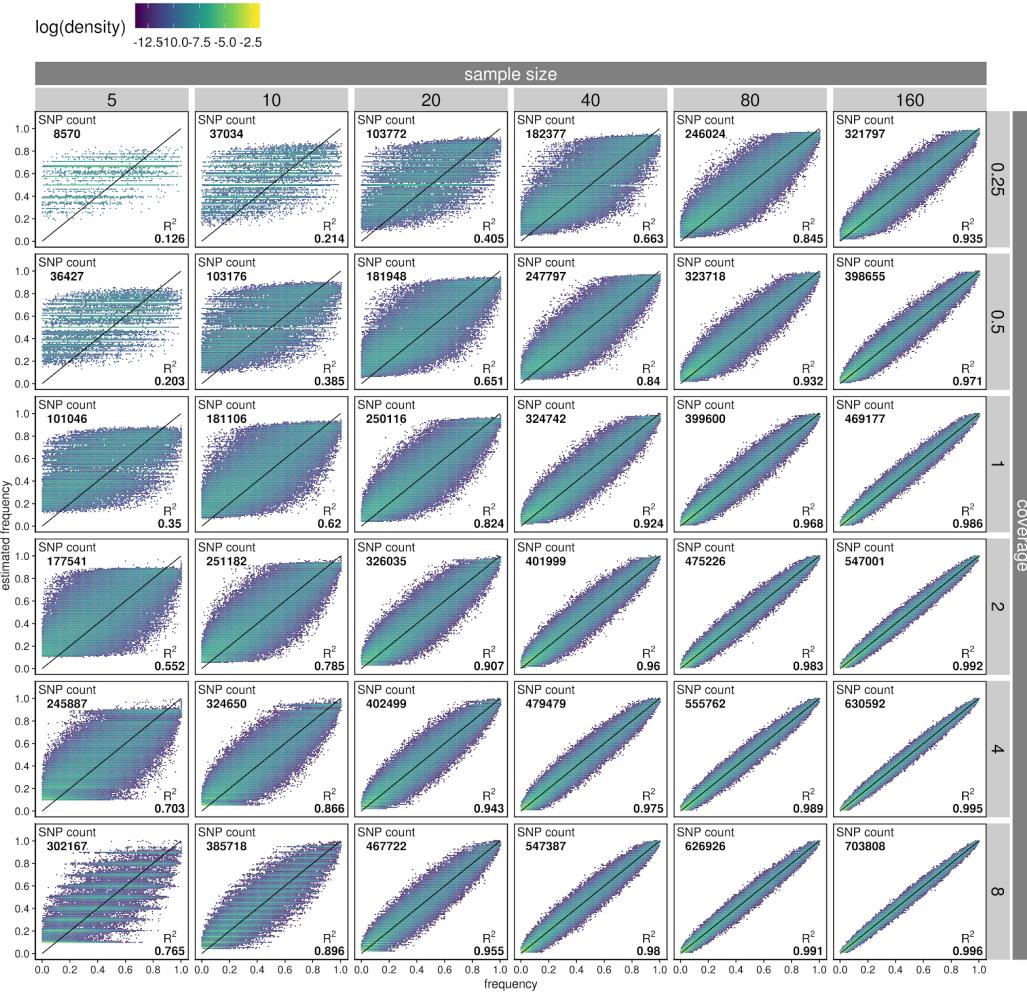
VGP GENBANK

Experimental design

- How low should we go?

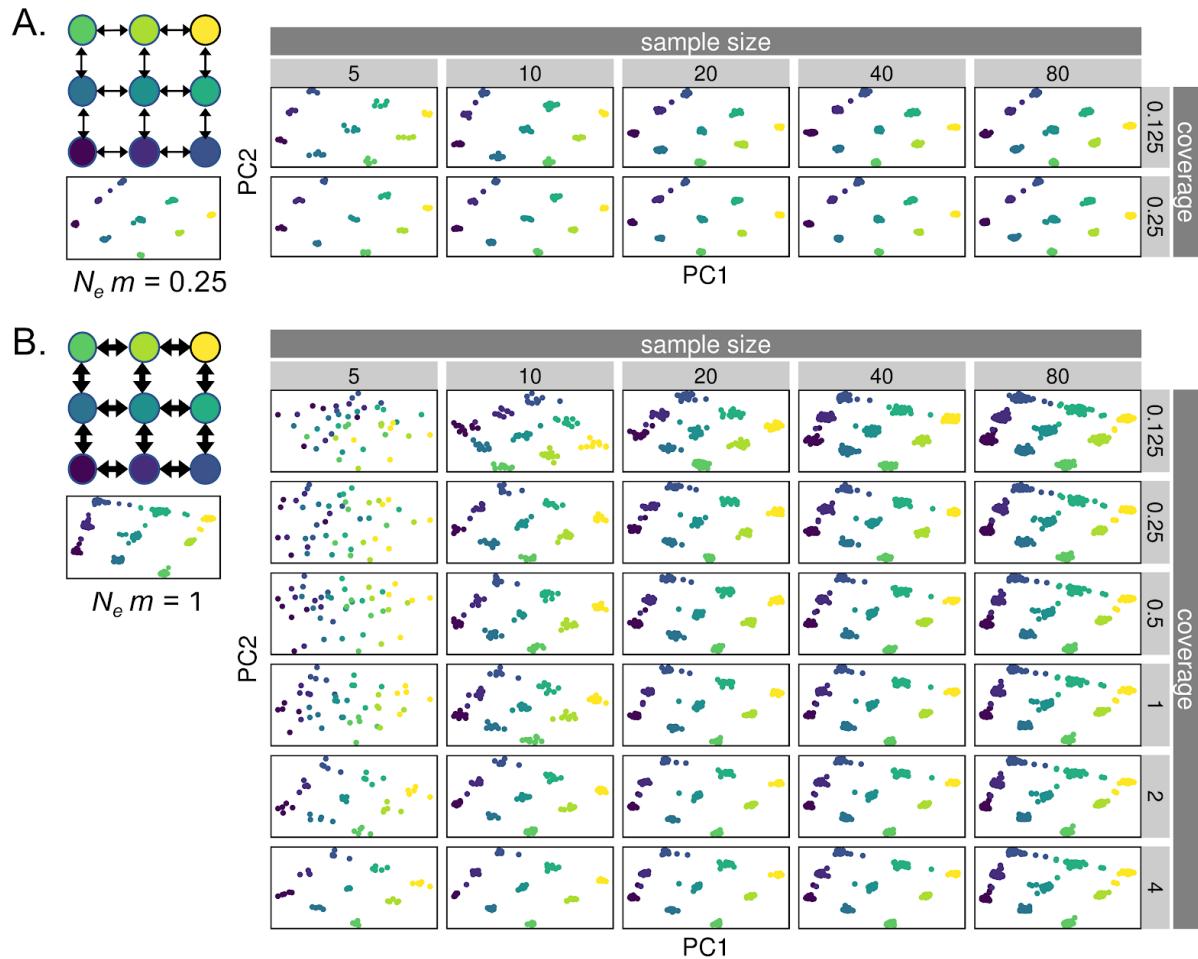
Allele frequency

Lower coverage of more individuals provides more accurate estimates



PCA

As long as the sample size is sufficient, population structure detected even at 0.125x coverage



Outlier detection

Peaks can be detected even at low coverage when the number of individuals is sufficient

