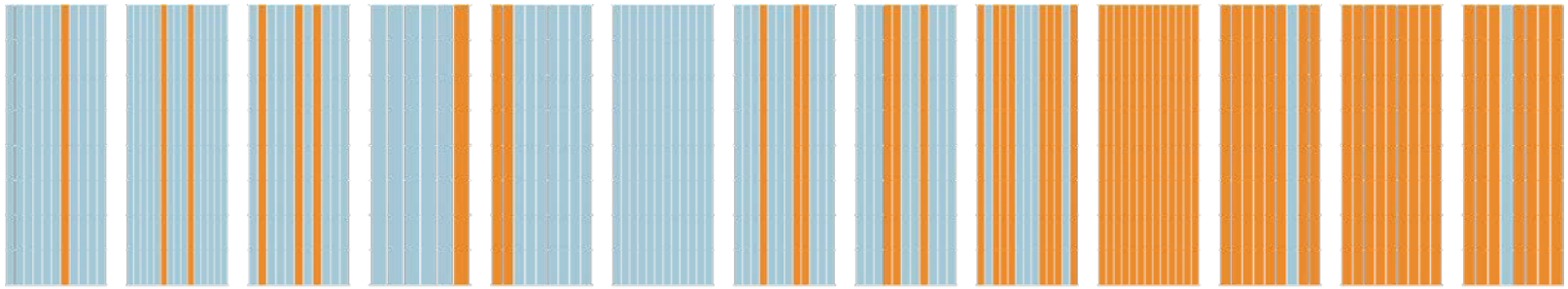
Genetics in support of fisheries and aquaculture management

17-19 September
Faro, Portugal

ICES CIEM | UAlg | CCMAR Centro de Ciências do Mar



Genetic and genomic tools - from proteins to genomes and back

17 September
Faro, Portugal

ICES CIEM | UAlg | CCMAR Centro de Ciências do Mar

What is a genetic marker?

"A genetic marker is an easily identifiable piece of genetic material, usually DNA, that can be used in the laboratory to tell apart cells, individuals, populations, or species."(The United States Department of Agriculture (USDA), 2006)

"Ideally, a genetic marker is an inherited variant that can be easily scored, varies within and between populations, and has a **negligible effect on phenotype and, in particular, on fitness**" (Baron et al. (2007) *Evolution*, New York, USA)

"A genetic marker is a gene or DNA sequence with a **known location on a chromosome** that can be used to identify cells, individuals or species. It can be described as a variation (which may arise due to mutation or alteration in the genomic loci) that can be observed." (www.wikipedia.org)

What can we use genetic markers for?

....all sorts of things, really

- Ecological inferences (e.g. track individuals)
- Population genetic inferences (e.g. estimate connectivity/levels of gene flow)
- Evolutionary inferences, i.e. the effects of evolution on the genomes of individuals, populations and species

Evolutionary forces? (4)

"Why Dag! I haven't seen you in years! You haven't evolved a bit!

Genetic markers

Phenotypes, e.g. enzymes

Used as a proxy for genetic variation

Organelles, e.g. mitochondria, chloroplasts

mtDNA is maternally inherited, haploid and without recombinatio

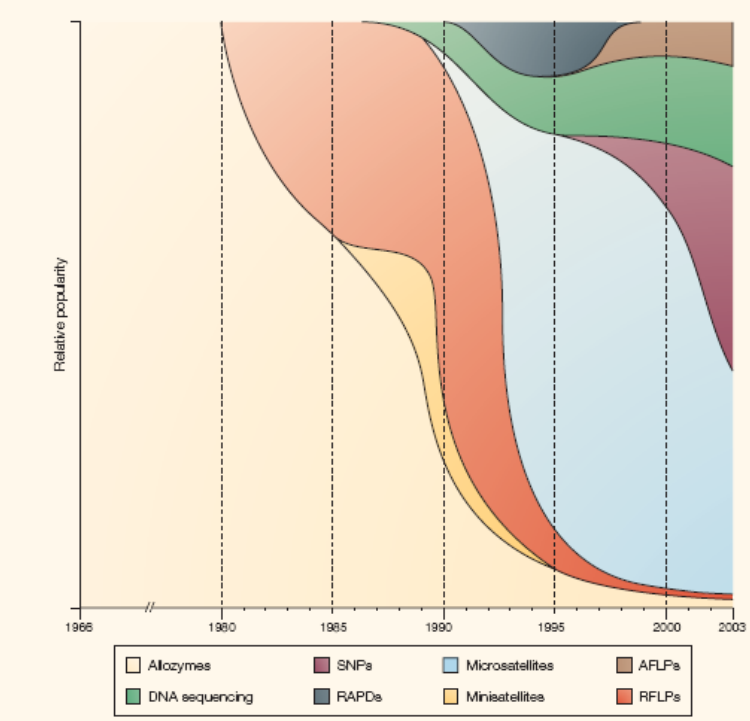
- ¼ Ne of nuclear DNA
- Only tracks female gene flow
- Behaves as a single locus

mtDNA has been widely used in phylogeographic studies
One particular gene, COI, used for species separation (barcoding)

Nuclear DNA

Diploid, inherited from both parents, codes most the genes, harbours lots of variation

Genetic markers



- Biochemical/Protein level

  Allozymes (Variants of an enzyme coded by one particular locus)
  Haemoglobin variants

- Molecular/DNA level

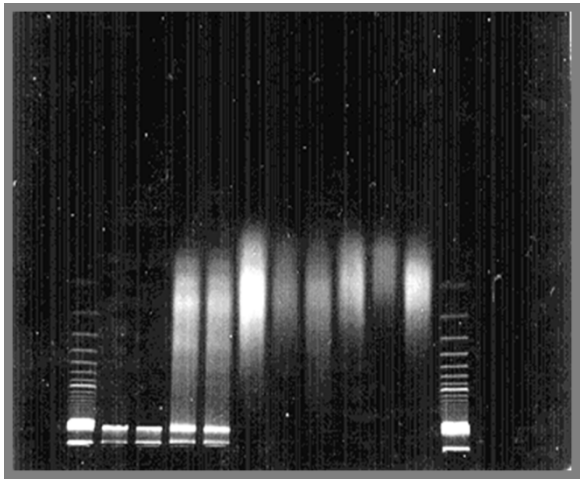
  Repeat variation (e.g. microsatellites)
  Single nucleotide polymorphisms (SNPs)

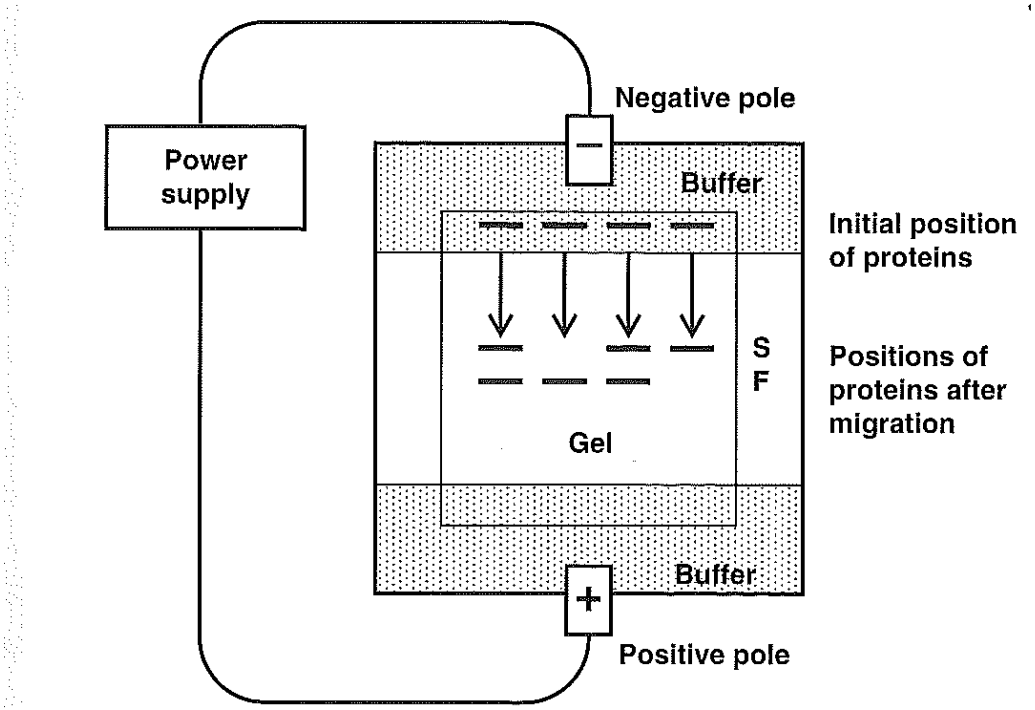
Schlötterer 2004

Electrophoresis

Use electrical current to separate proteins or DNA or based on charge and/or size

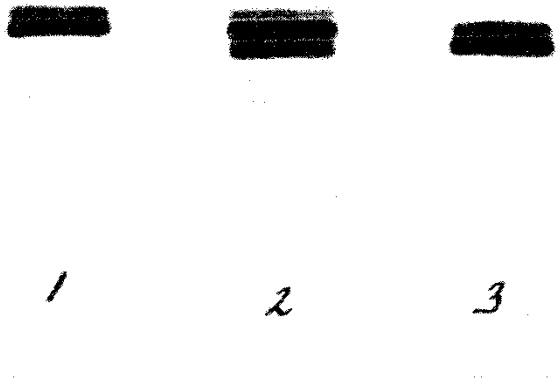
Enzymes: Enzyme specific dyes
DNA: Ethidium bromide + UV

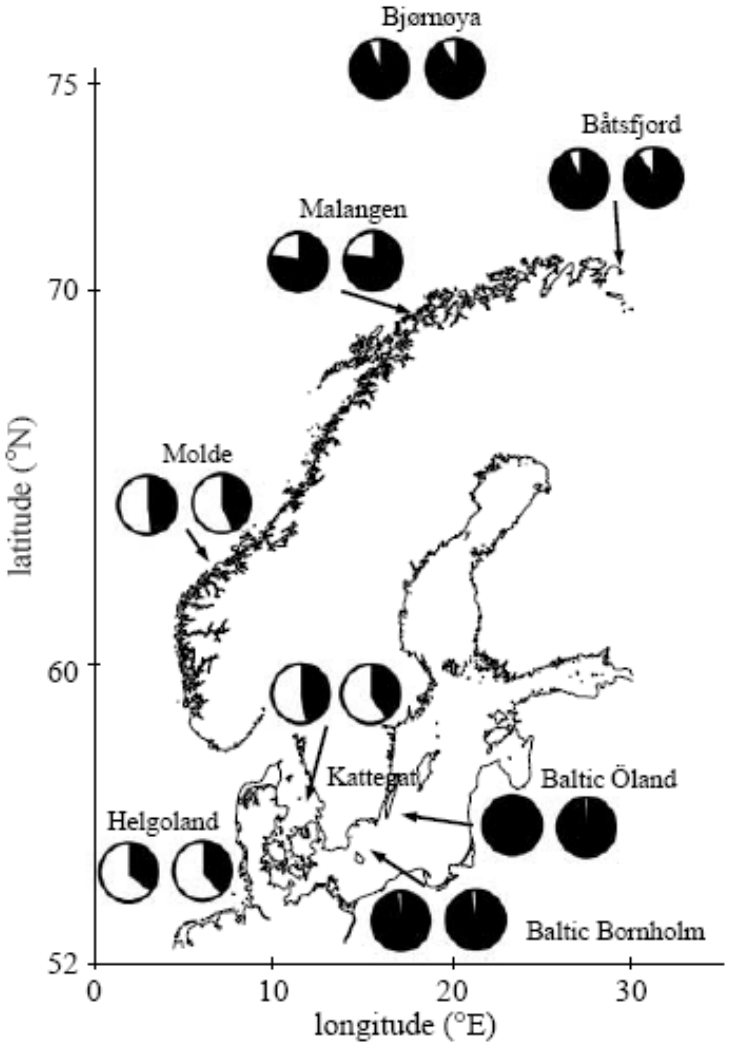




Genetic markers - Biochemical

Haemoglobin as an example





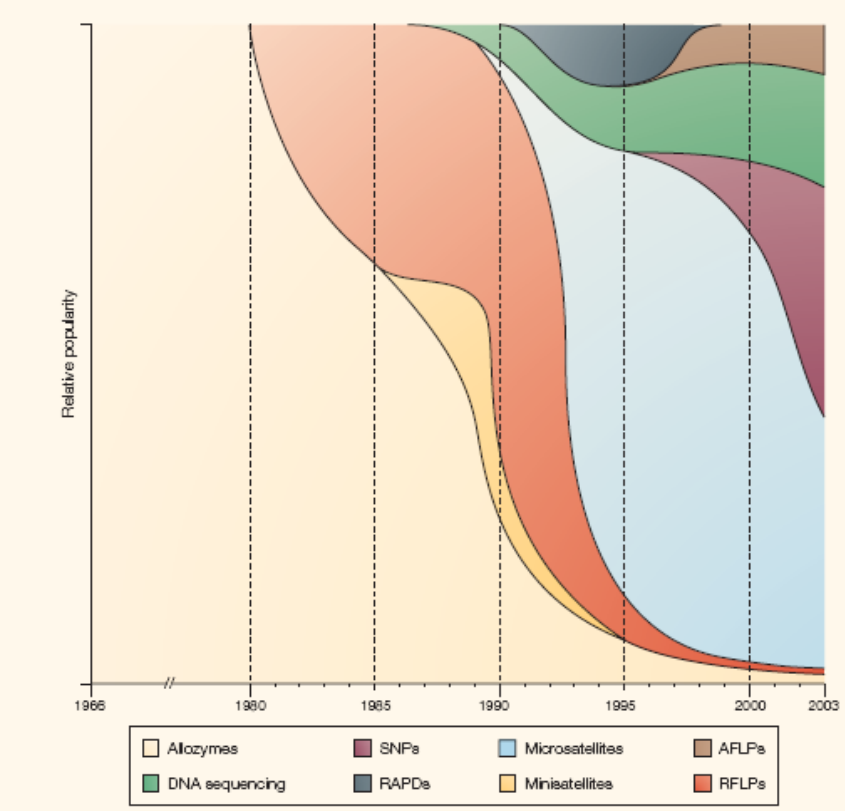
Sick 1961; Andersen et al. 2009

Genetic markers

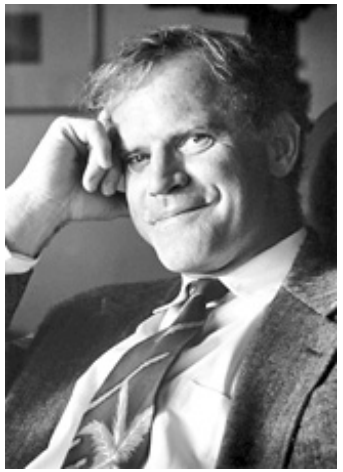
Biochemical/Protein level

- Low number of alleles/low level of genetic variation
- Only detects a proportion of present variation
- May be under selection
- Easy to apply (but difficult to master!)

Virtually abandoned today…..



Schlötterer 2004



PCR

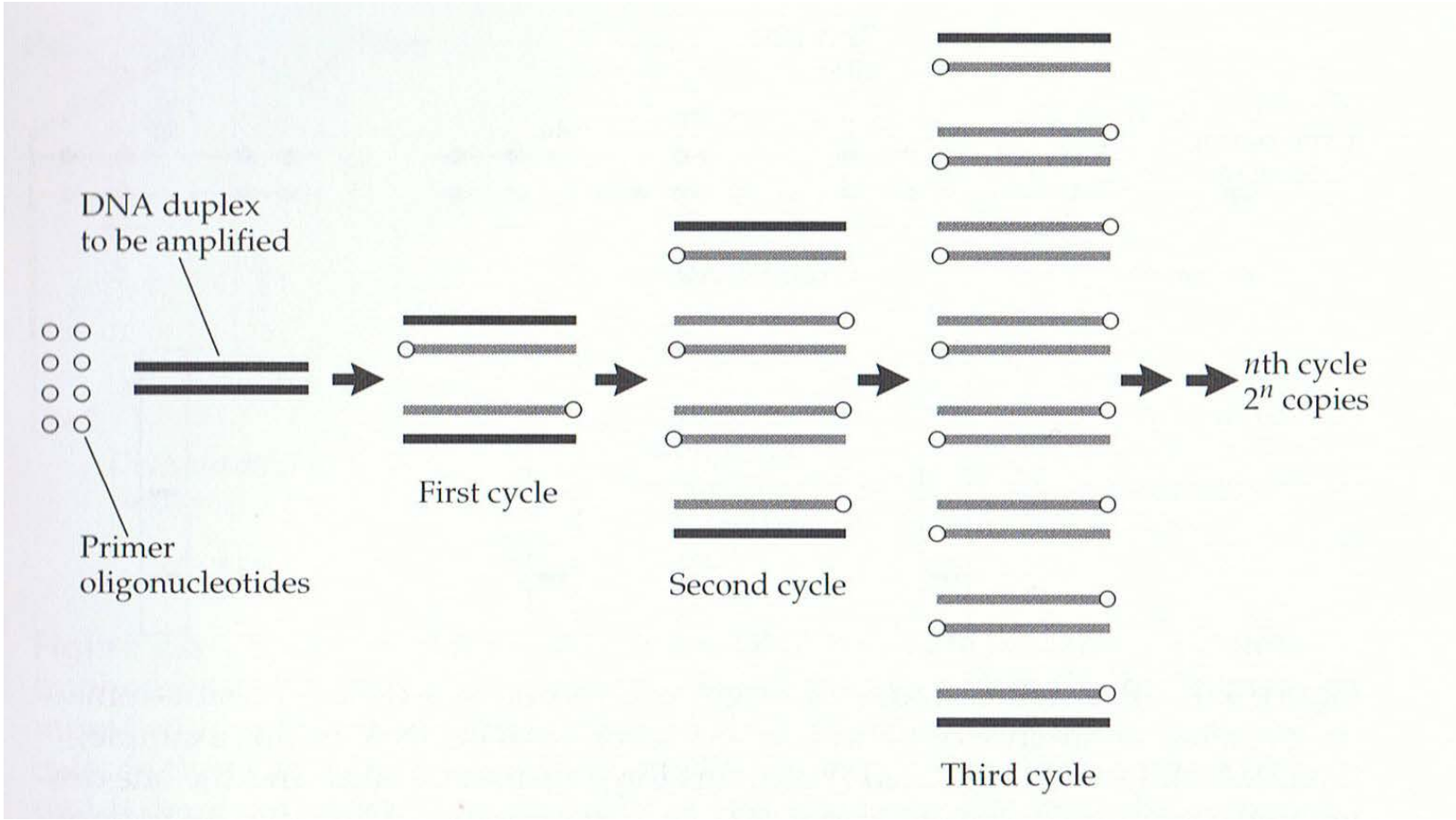
- Developed by Kerry Mullis in 1983
- Nobel prize in 1993
- "What if I had not taken LSD ever; would I have still invented PCR?I don't know. I doubt it. I seriously doubt it.“

- Enzymatic amplification of DNA
- Essentials:
  - –Heating and cooling
  - –A thermo-stabile DNA polymerase

Polymerase chain reaction (PCR)

Flanking sequence, known | Target sequence | Flanking sequence, known

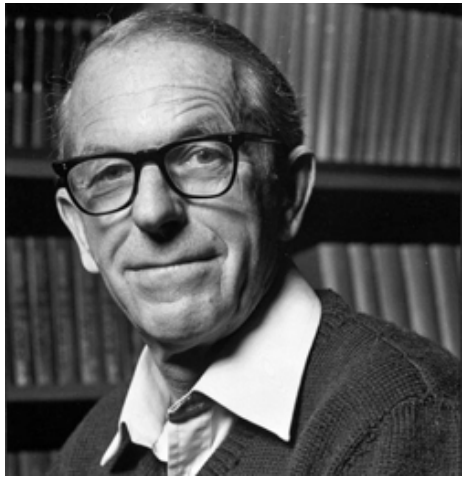
AGCCGTGCTCAGCAATXXXXXXXXXXXXXXXXXXXXXXAGGCATAAACGTTGGAC...



Sanger sequencing

- Sequencing method described by Sanger and Coulson in 1977

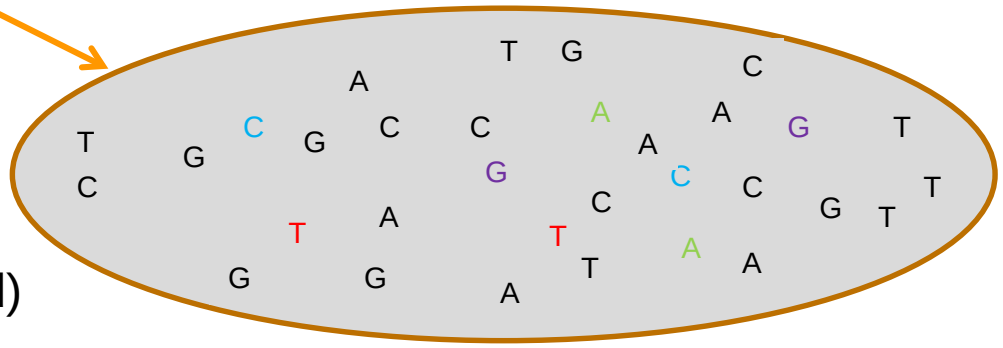
- For the next 25 years, ‘Sanger sequencing’ was practically the only DNA sequencing method used

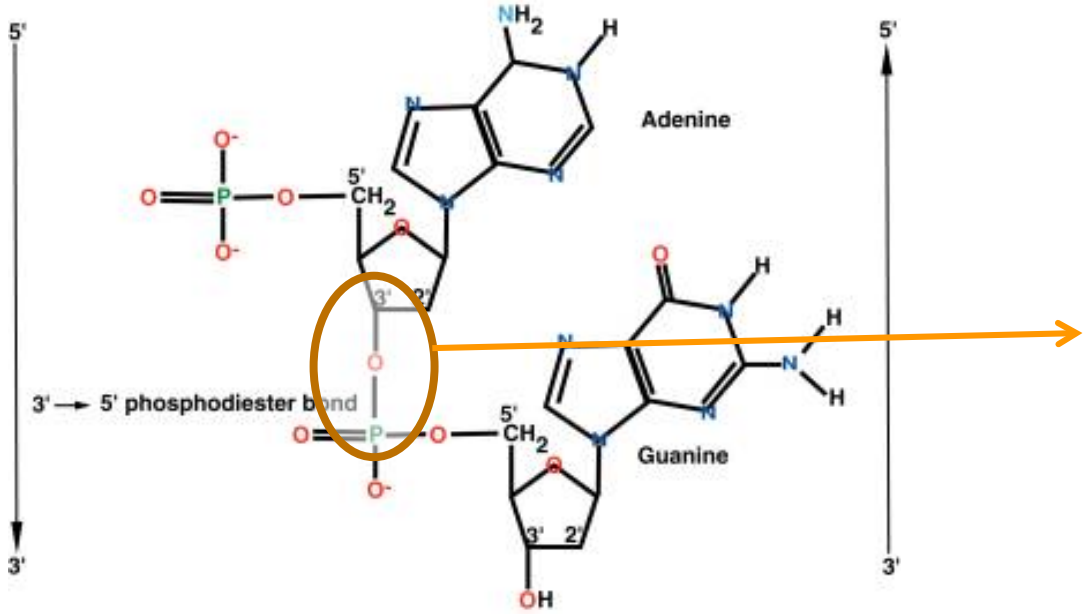


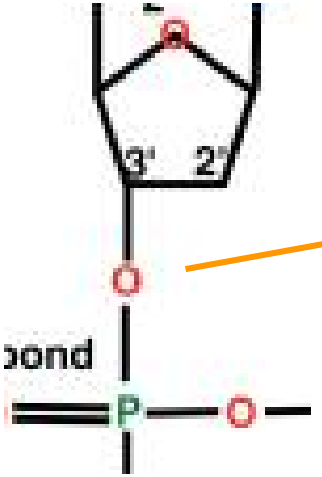
Sanger sequencing

PCR reaction mix

Incl.
dNTP's (black)
&
ddNTP's (coloured)







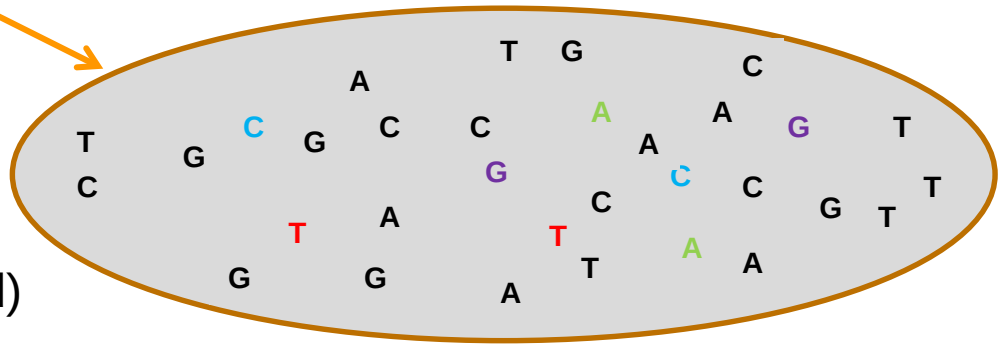
dNTP = -OH

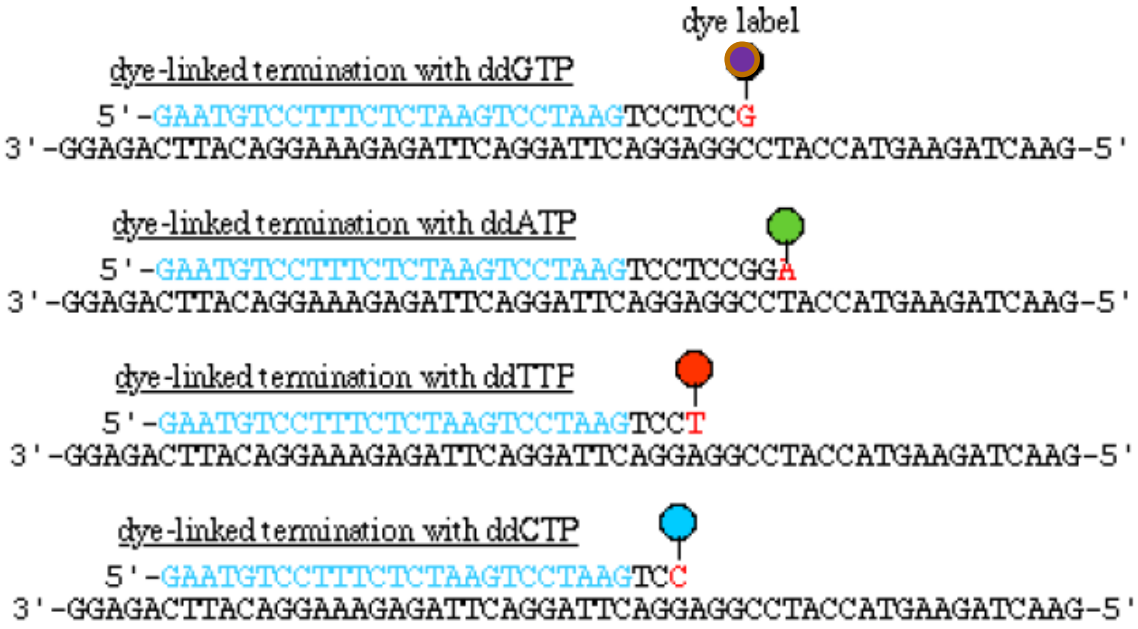
ddNTP = -H

Sanger sequencing

PCR reaction mix

Incl.
dNTP's (black)
&
ddNTP's (coloured)

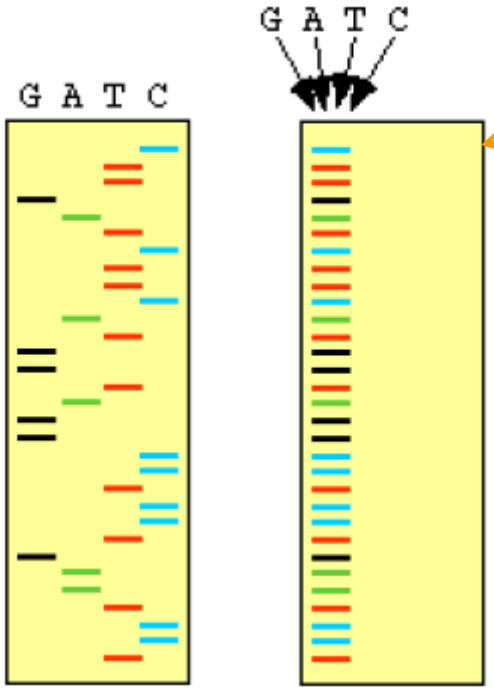




Sanger sequencing

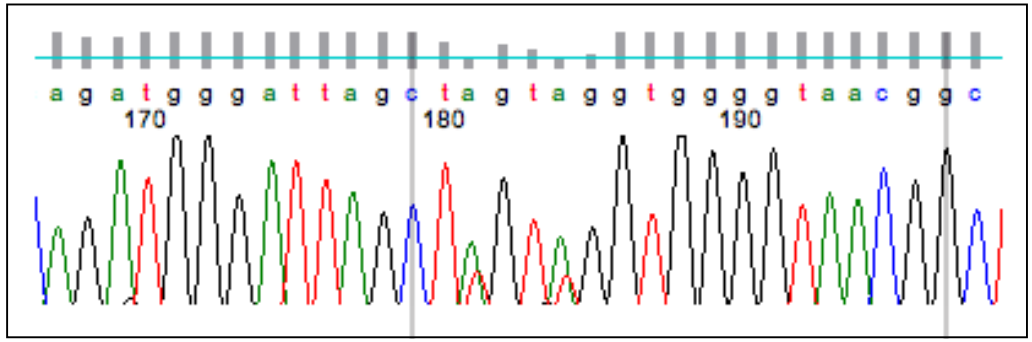
Sanger sequencing

PCR sequence products are analysed by electrophoresis.
Distinguishes DNA fragment lengths by 1 bp precision and records the fluorescent dye at the terminal ddNTP.



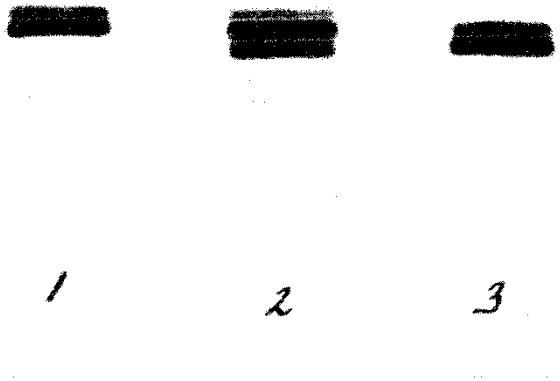
Output read on a gel

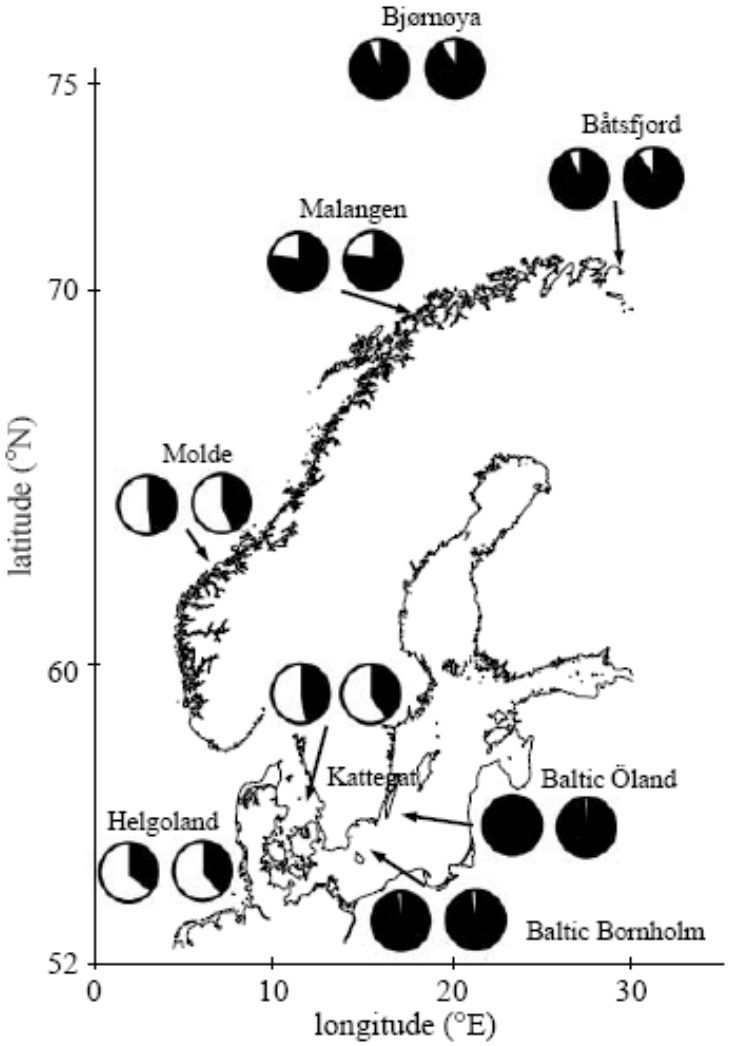
Output read as a chromatogram



Genetic markers - genes

Haemoglobin as an example

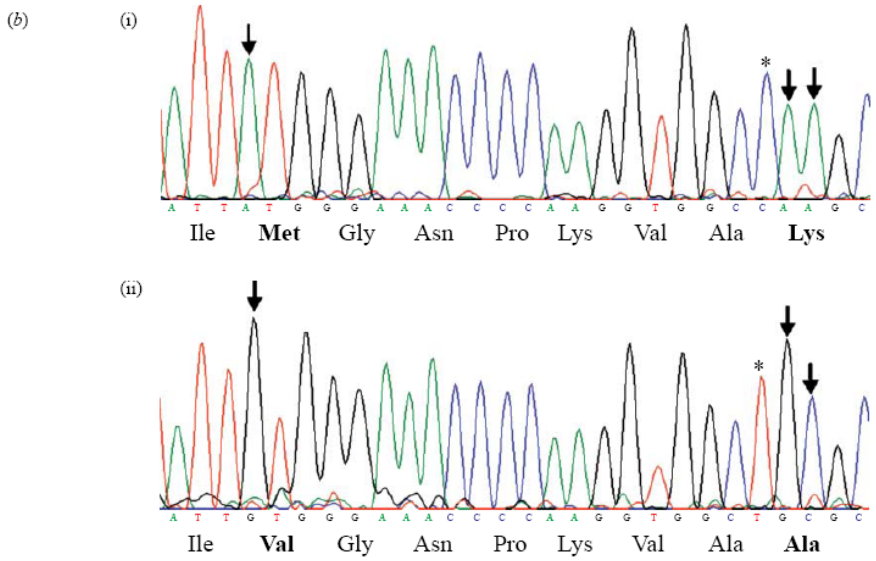


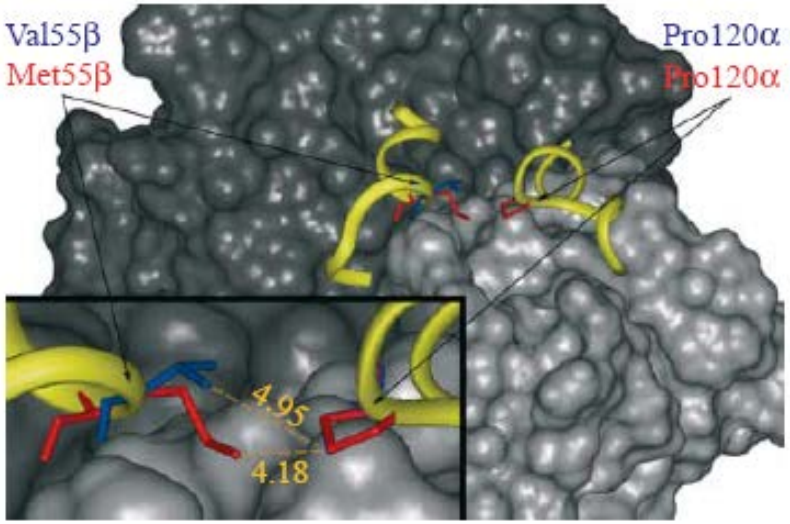


Sick 1961; Andersen et al. 2009

Genetic markers - genes

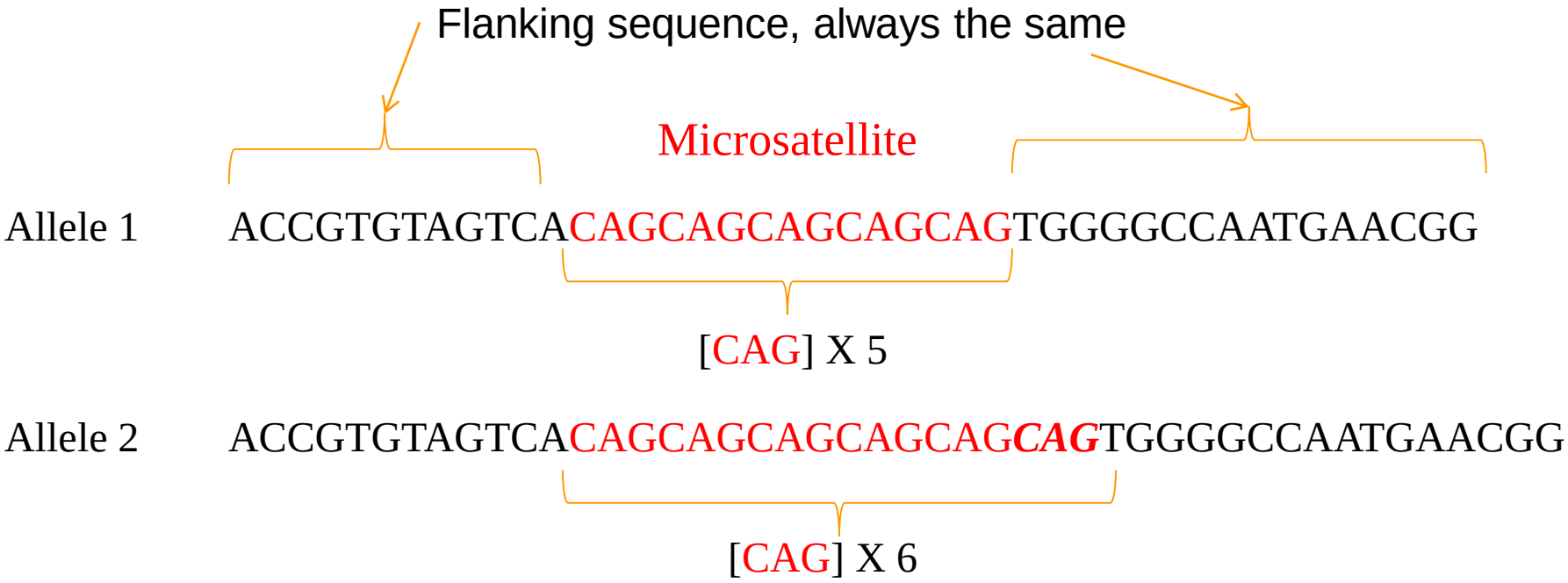
The DNA version



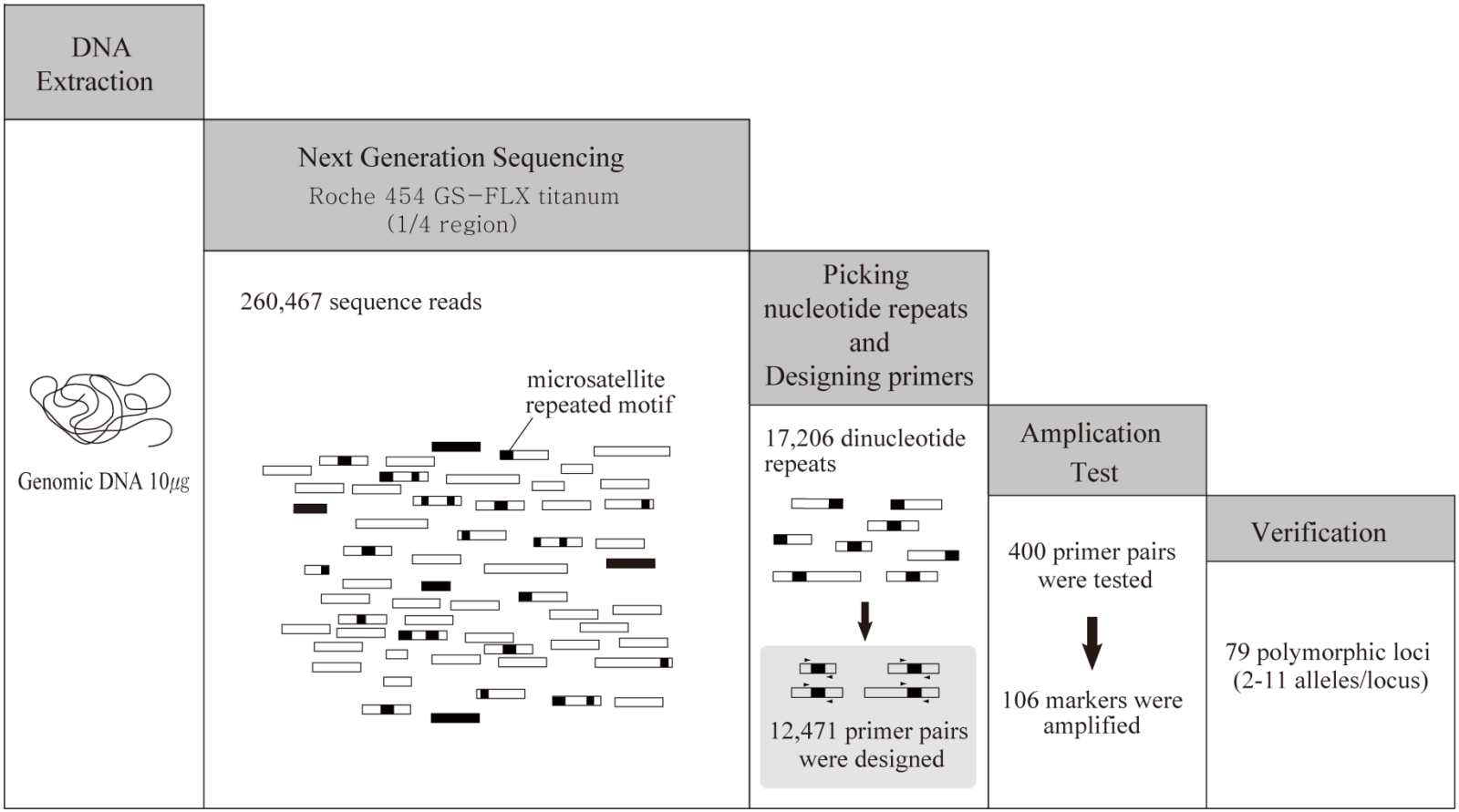


Andersen et al. 2009

Genetic markers – repeat variation



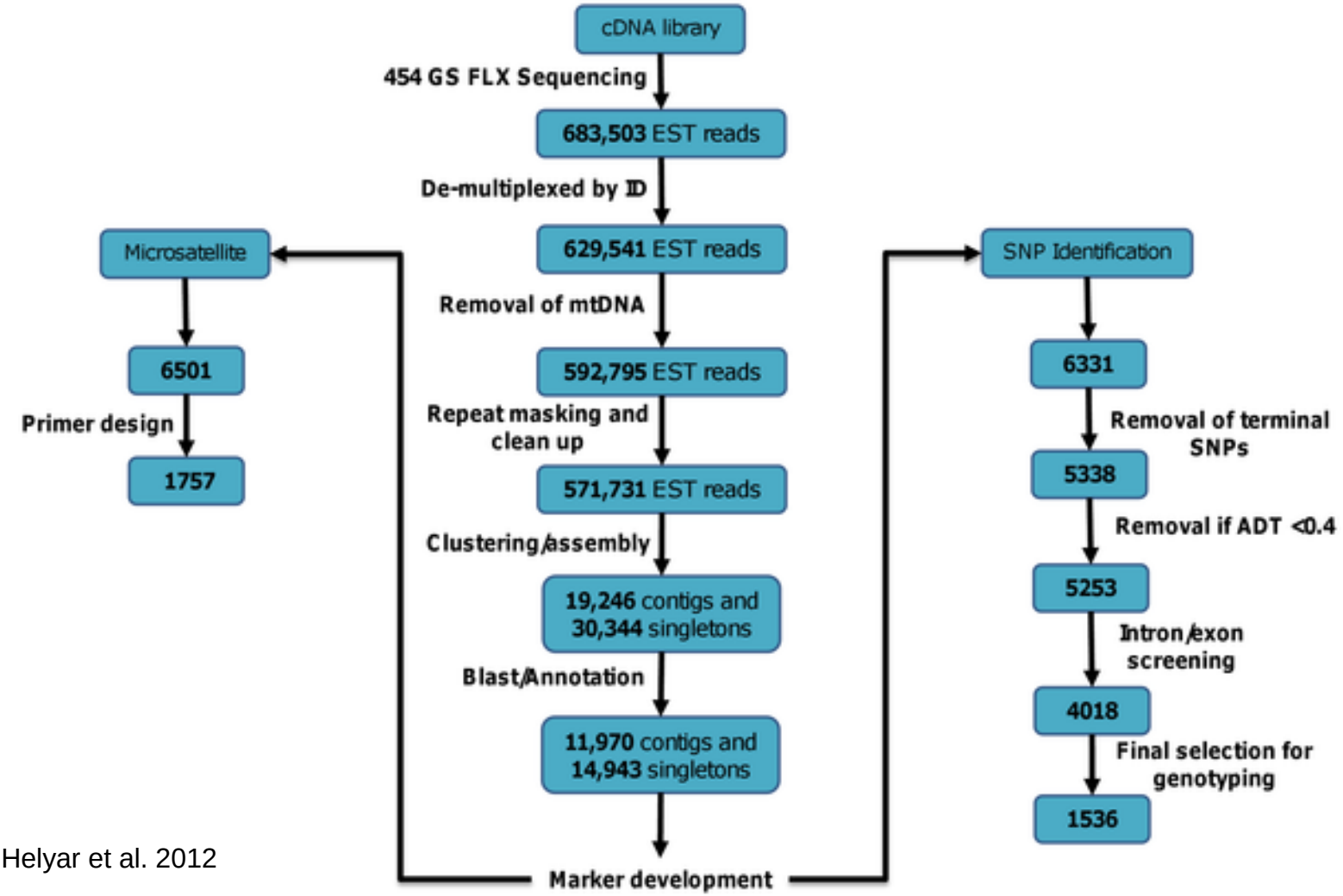
- High number of alleles/high level of genetic variation
- Often polymorphic in closely related species
- Need to know flanking sequence in focal or related species
- Takes time to develop
- Length variants may be difficult to separate in the lab and calibrate across labs


DNA Extraction
Genomic DNA 10μg
Next Generation Sequencing
Roche 454 GS-FLX titanum
(1/4 region)
260,467 sequence reads
microsatellite repeated motif
Picking nucleotide repeats and Designing primers
17,206 dinucleotide repeats
12,471 primer pairs were designed
Amplication Test
400 primer pairs were tested
106 markers were amplified
Verification
79 polymorphic loci (2-11 alleles/locus)

Genetic markers – Single nucleotide polymorphisms (SNPs)

- Only two alleles (mostly)
- Often species specific
- Need to know flanking sequence and SNP in focal species
- Abundant in the genome (in genes and outside genes)
- Variants easy to identify in the lab
- Good for high throughput genotyping
- Easy to calibrate

SNP development in herring

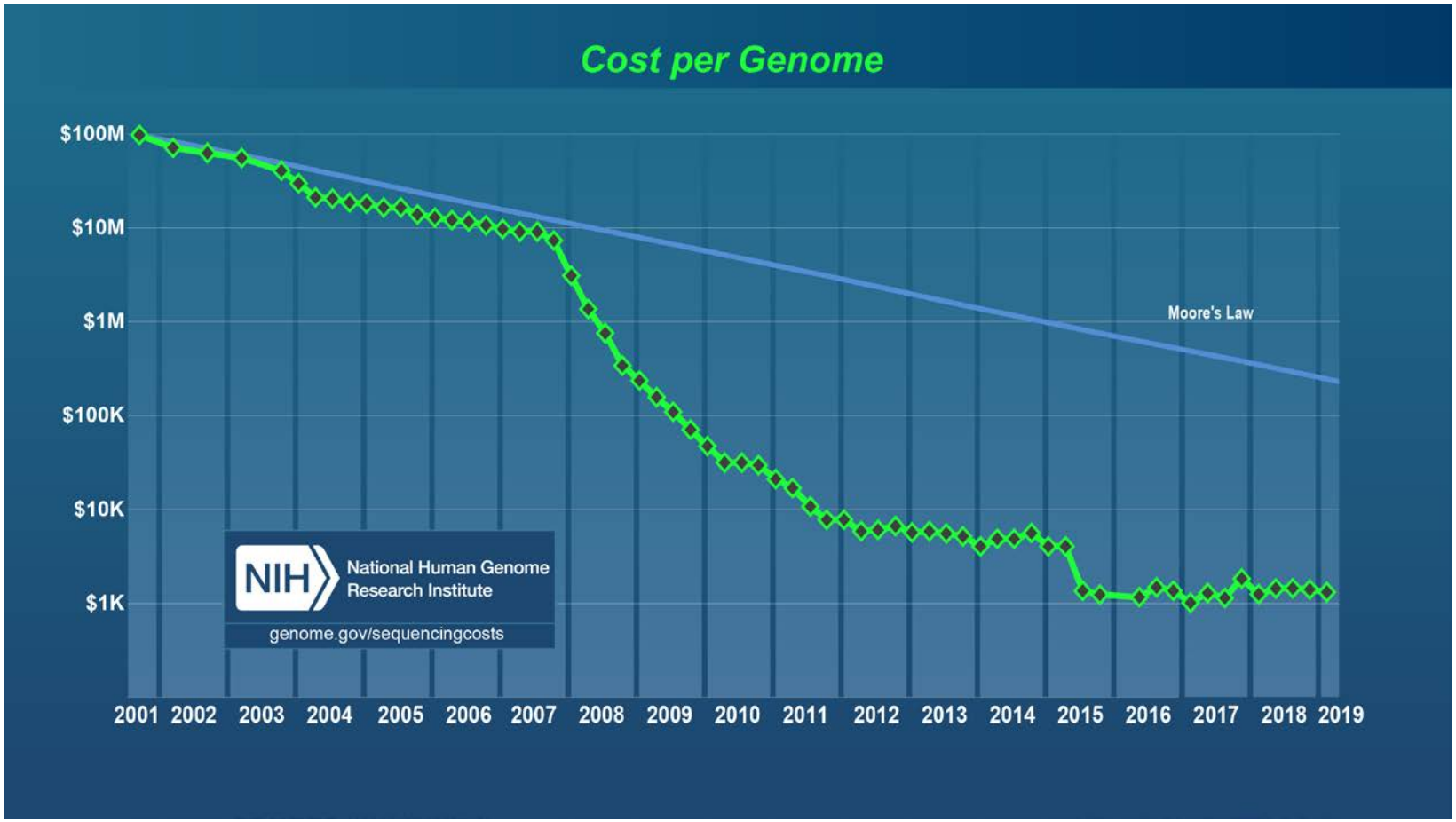


Helyar et al. 2012

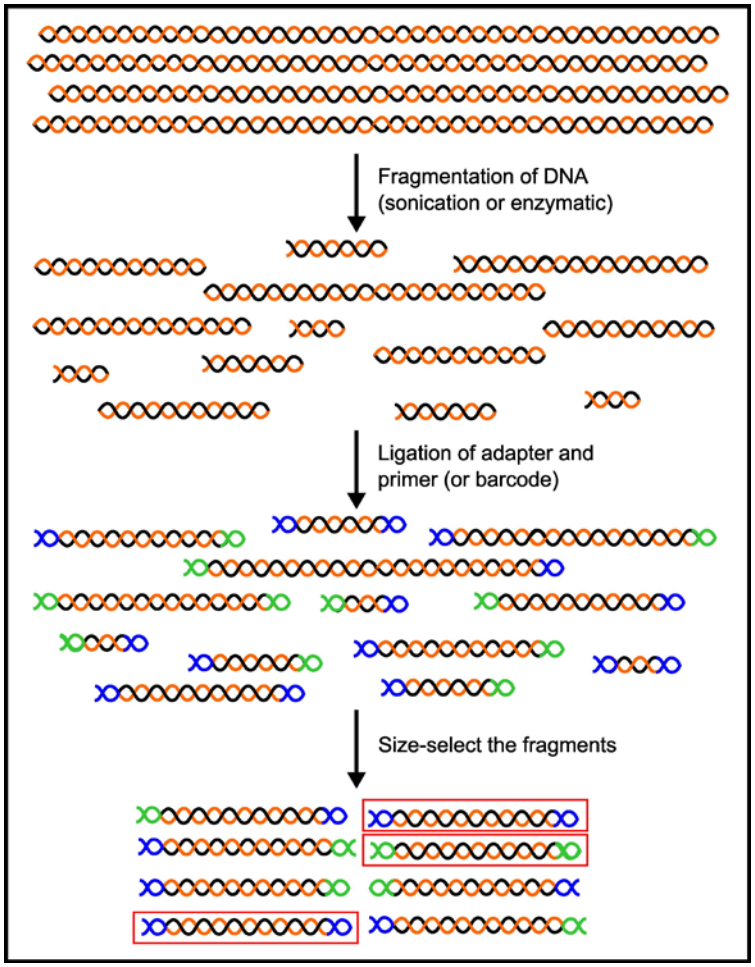
Next generation sequencing

- Generates thousands of randomly amplified sequences in one run
- The methodologies have been available since 2005
- Human genome was sequenced with the Sanger method over 10 years for a total cost of ~ $3 billion
- A new version was sequenced in 2007 with the 454 technology (next generation technology) in 2 months for ~ $2 million
- In 2019 the prize is approximately $1500 using Illumina technology

Cost per genome

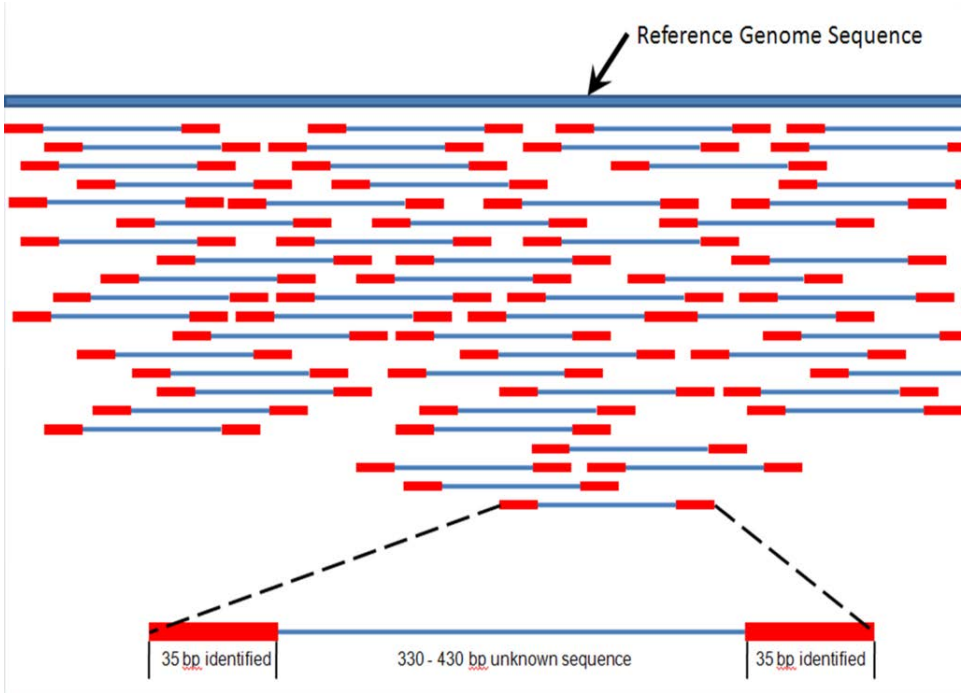


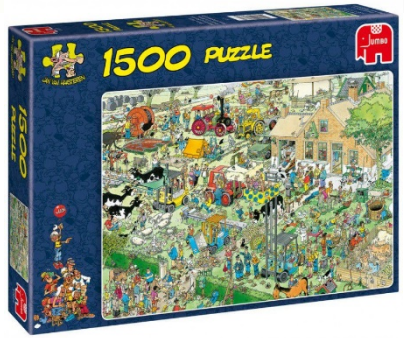
”NGS” sequencing 2019



Illumina Novaseq 6000. 20 billion reads of 150 bp

Bioinformatics





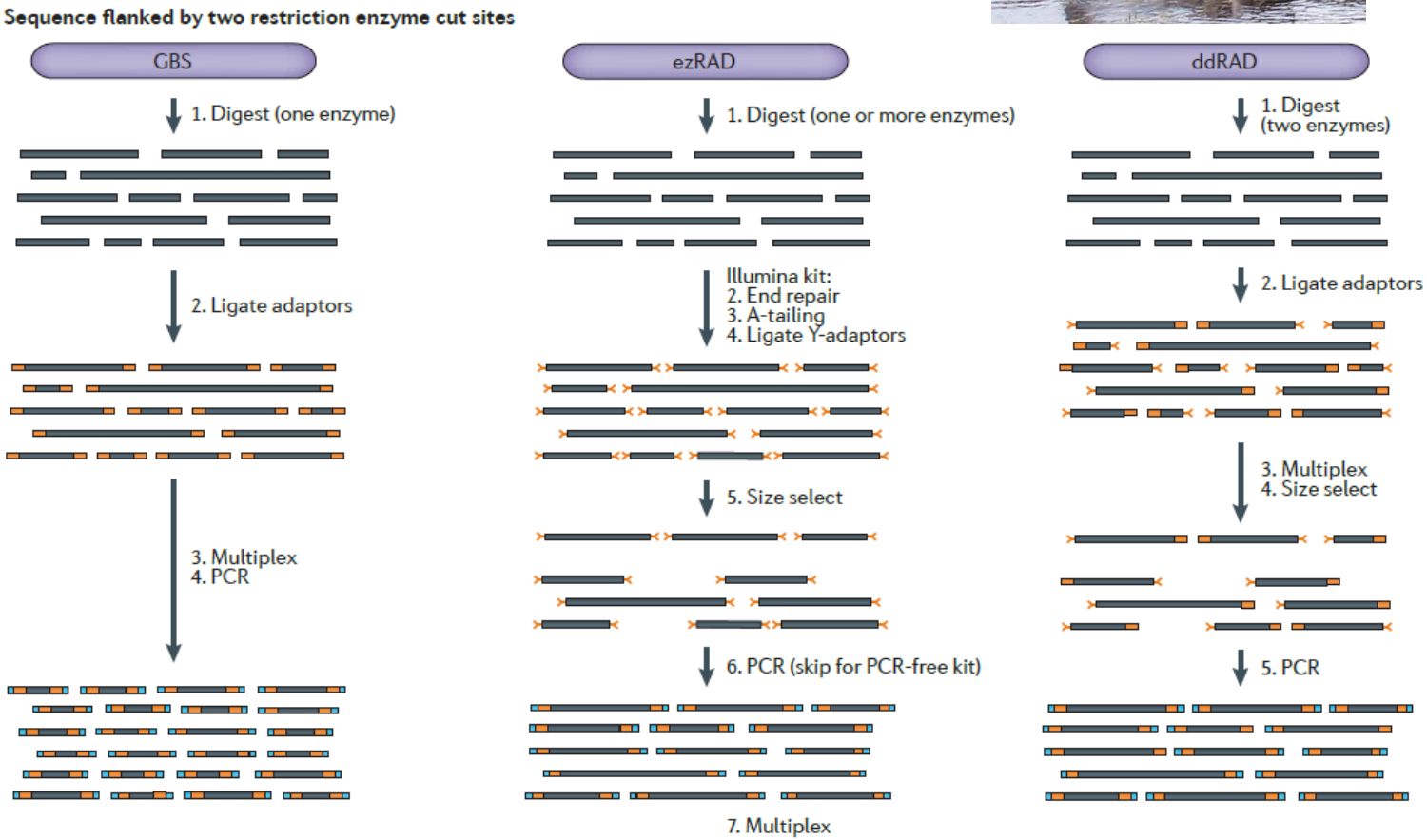
13,3 jigsaw puzzles with
1500 pieces

Next generation sequencing

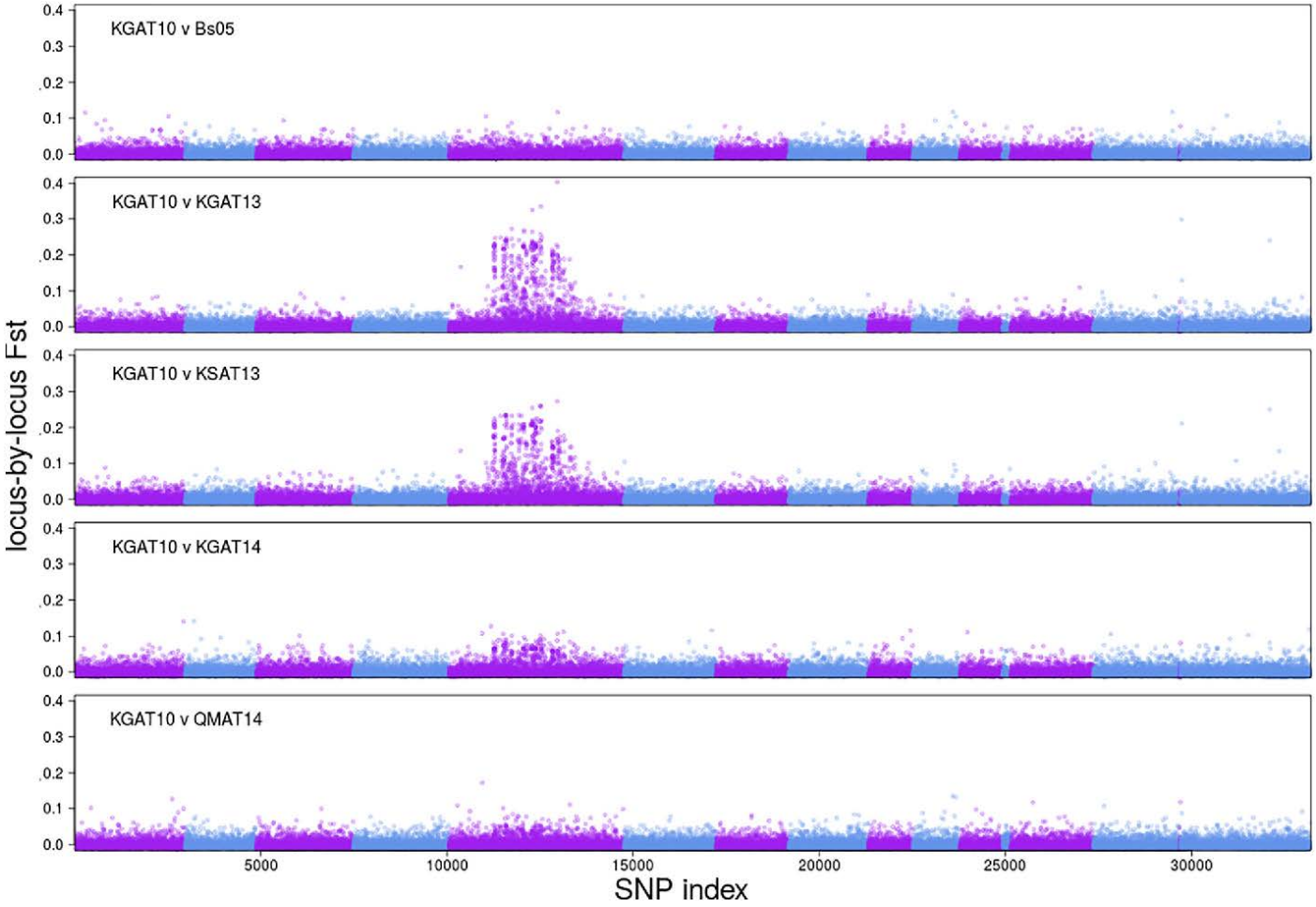
Applications

- SNP (or other marker) development (example)
- Next generation genotyping (RAD-sequencing)
- Sequence entire genomes (new genomes and re-sequencing)

Reduced representation sequencing

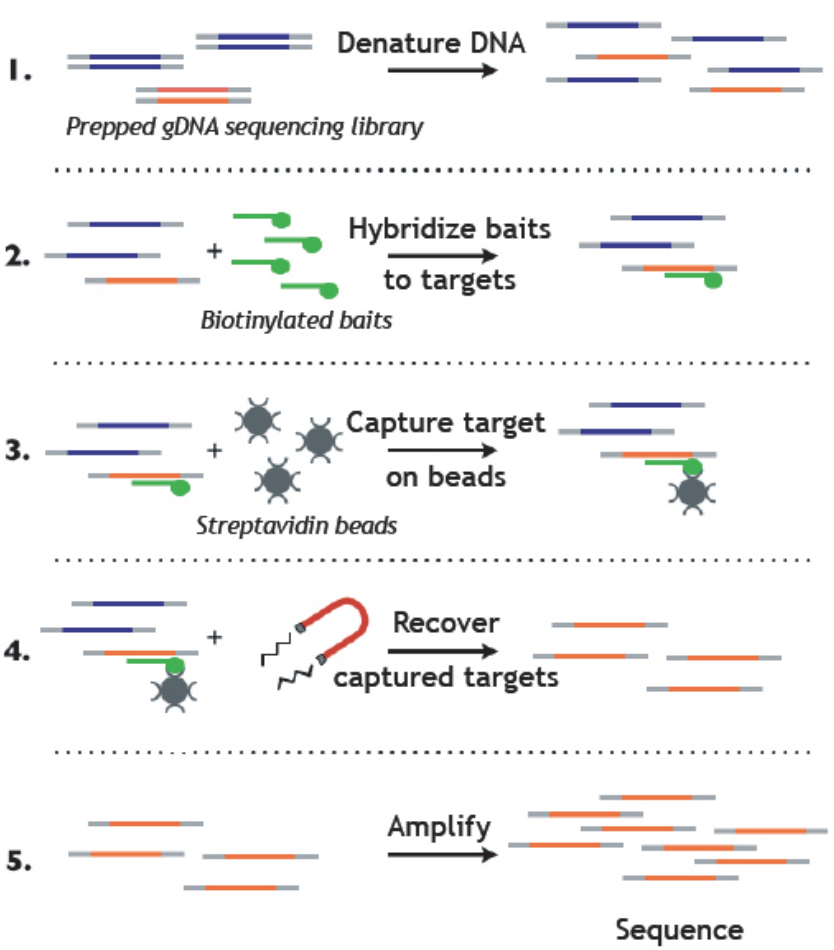


200 K SNP Chip for salmon louse



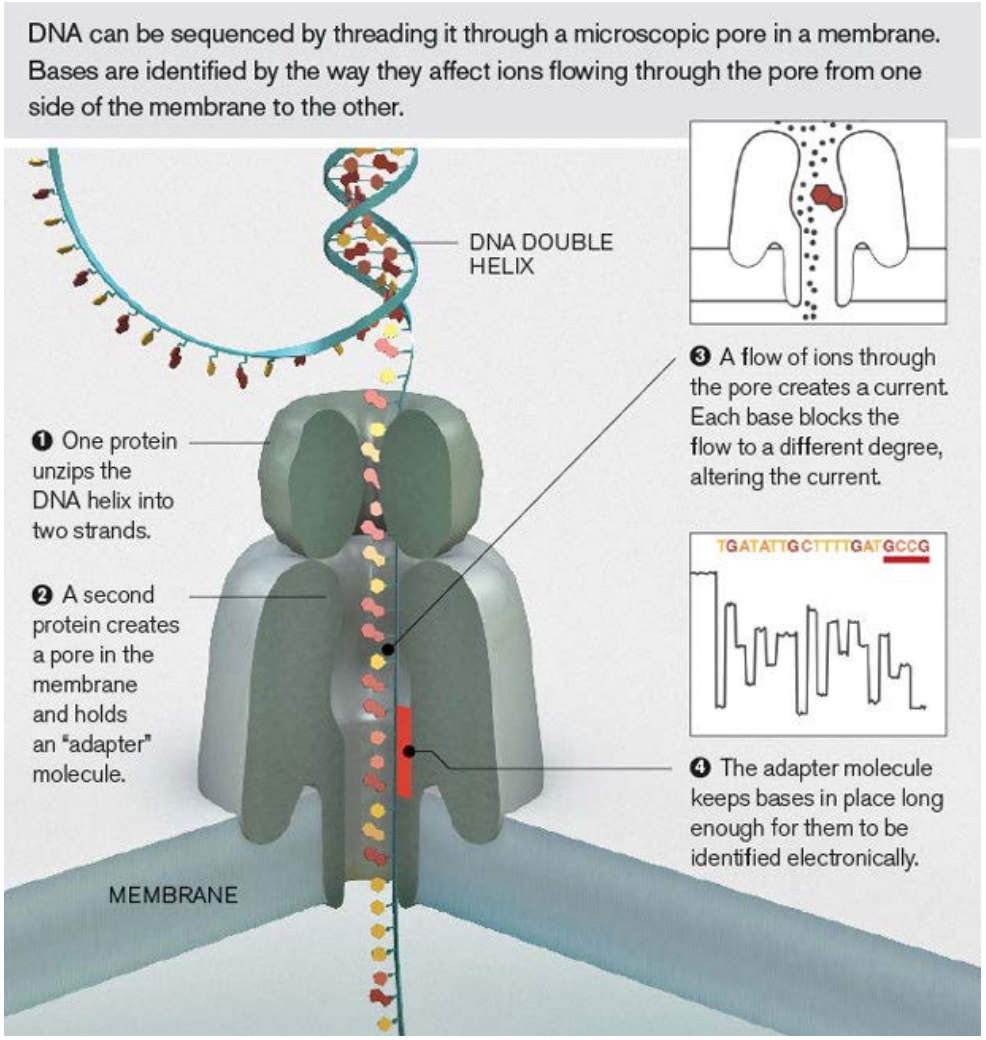
Messmer et al. 2018

DNA capture (historical material)



Next generation sequencing

Single molecule sequencing (3rd generation sequencing)



http://www.technologyreview.com

New technology – Oxford Nanopore

- Portable 3rd generation sequencing unit SmidgION

Summary

History	1960s	2000s	2019
Markers	Gene products (e.g. allozymes)	DNA, non-coding (e.g. microsatellites) + a few genes	DNA, coding and non-coding, candidate genes (SNPs)
# markers	~10-20	~10-20 (highly variable)	1000+ Genomes
What did we learn?	Genetic differences! Population structure (and adaptation)	Population structure at fine geographical scales History, demography and exchange between populations	What does structure really mean? Functional biodiversity Adaptation
Evolutionary forces	Neutral (and non-neutral)	Neutral	Neutral and non-neutral

Genomic differentiation

Le Moan et al.


WHERE?
WHEN?
WHY?
WHO?
WHAT?
HOW?
WHEN?
WHO?
WHAT?
WHERE?
HOW?
HOW?
WHEN?
WHERE?
WHERE? WHICH? WHOSE? WHEN? WHY?
WHOSE?
WHICH?
WHOSE?
WHEN? WHERE?
HOW?
WHAT? HOW? WHY?
WHO? WHERE? WHAT? HOW?
WHY? HOW?
WHAT? HOW? WHY?
WHO? WHERE? WHAT? HOW?
WHERE? WHICH? WHOSE? WHEN? WHY?
WHY? HOW? WHERE?
WHO? WHOSE?
WHERE? WHAT? HOW?
WHERE? WHICH? WHOSE? WHEN? WHY?
WHAT?
WHY?
HOW?
WHERE?
WHO?
WHERE? WHAT? HOW?
WHERE? WHICH? WHOSE? WHEN? WHY?
HOW? WHERE? WHY?
WHERE?
HOW?
HOW? WHO?
WHO? WHERE? WHAT? HOW?
WHAT? HOW? WHY?
WHERE?
WHAT?
HOW?
HOW?
WHAT?
HOW?
WHAT? WHERE? HOW?