

HOS 6236 Molecular Marker Assisted Plant Breeding

Fall 2017

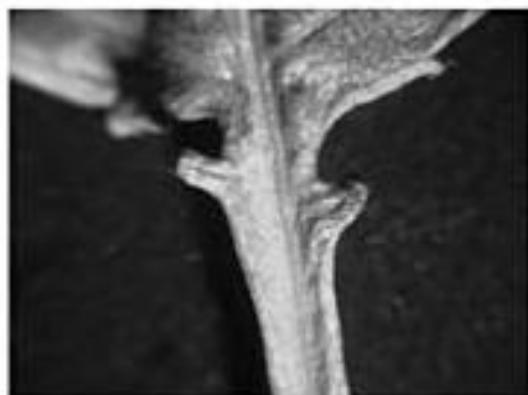
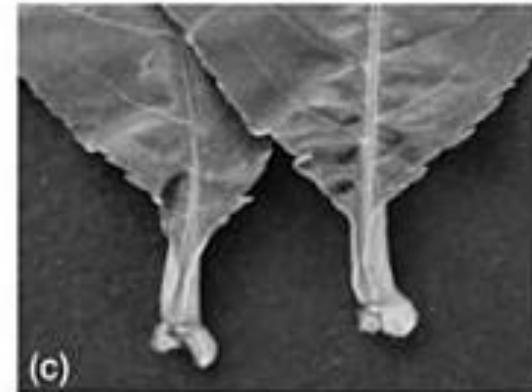
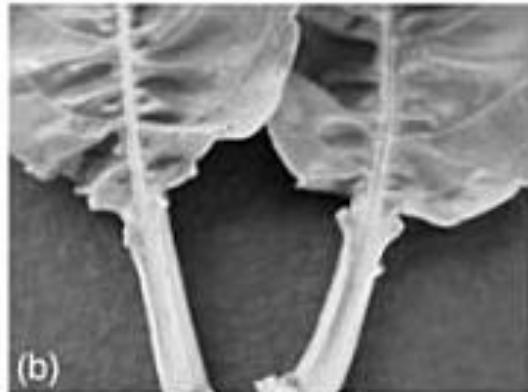
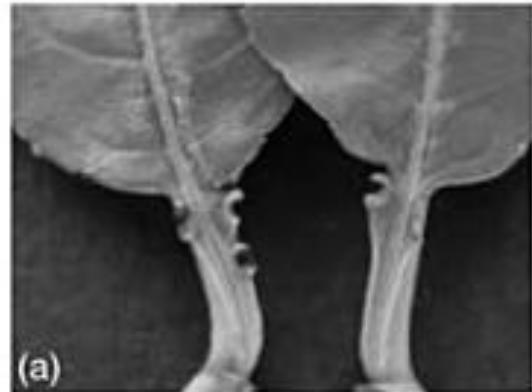
Last Class:

Recurrent selection methods and BLUP

Todays Class:

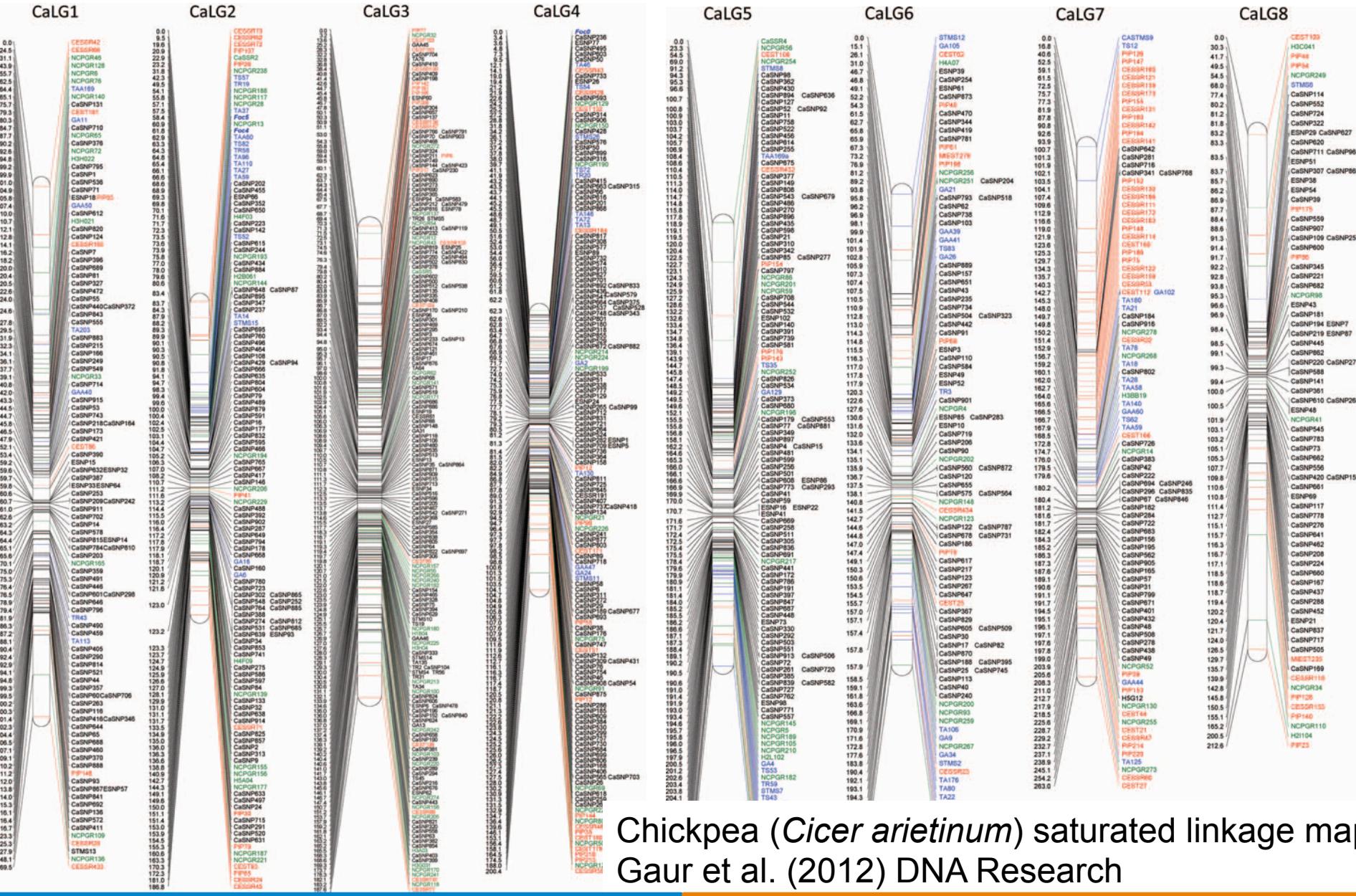
DNA Molecular Markers

Morphological Markers



Segregation for peach (*Prunus persica*) leaf glands
Bassi and Monet (2008)

DNA Markers



Chickpea (*Cicer arietinum*) saturated linkage map Gaur et al. (2012) DNA Research

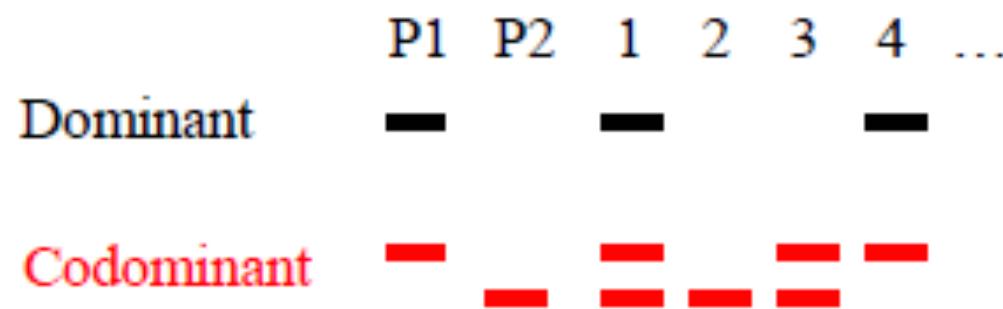
Genome variation

- Single nucleotide polymorphism
 - Coding region vs. non-coding region
 - Synonymous vs. non-synonymous
 - Missense vs nonsense
- Complex polymorphisms (multi-SNP)
- Structural variations
 - Insertions and deletions (indels)
 - Copy number variations (CNV)
 - Presence/Absence variations (PAV)
 - DNA rearrangements



What Makes a Good DNA Marker?

- High level of polymorphism – multi allelic
- Exhibit codominance (heterozygote can be distinguished)



What Makes a Good DNA Marker?

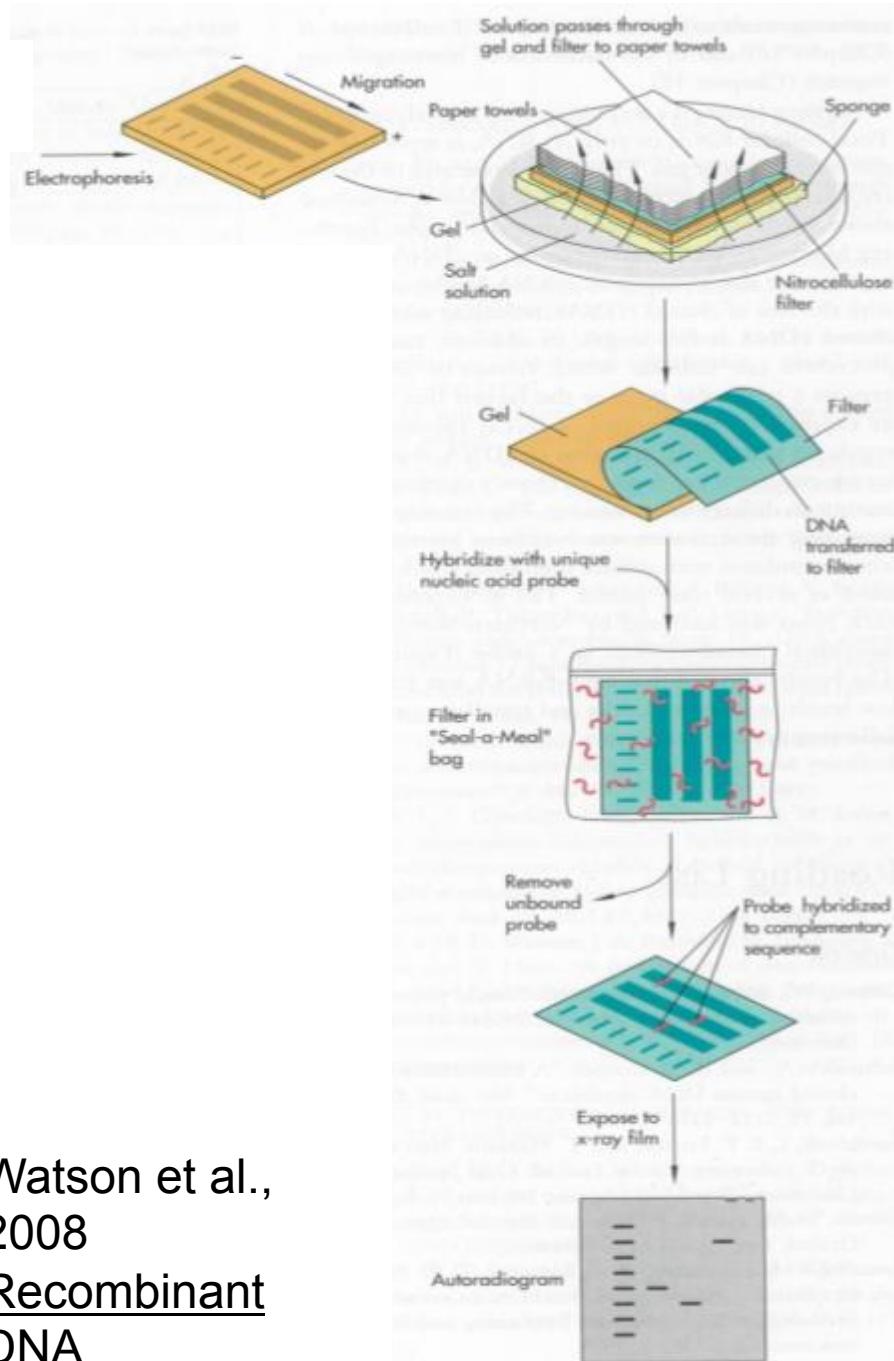
- Locus specific
- Even and frequent distribution over the genome
- Easy detection/visualization methods
- High repeatability
- Scalable for high throughput genotyping
- Low development and operational costs



Hybridization-Based Markers

- Require prior sequence knowledge for probe development
- PCR not necessary





RFLP (Restriction Fragment Length Polymorphism)

Botstein et al. (1980)

Construction of a genetic linkage map in man using restriction fragment length polymorphisms.

Amer. J. Hum. Genet.
32:314-331.

Watson et al.,
2008
Recombinant
DNA

PCR-Based Markers

1. Require prior sequence knowledge for primer development

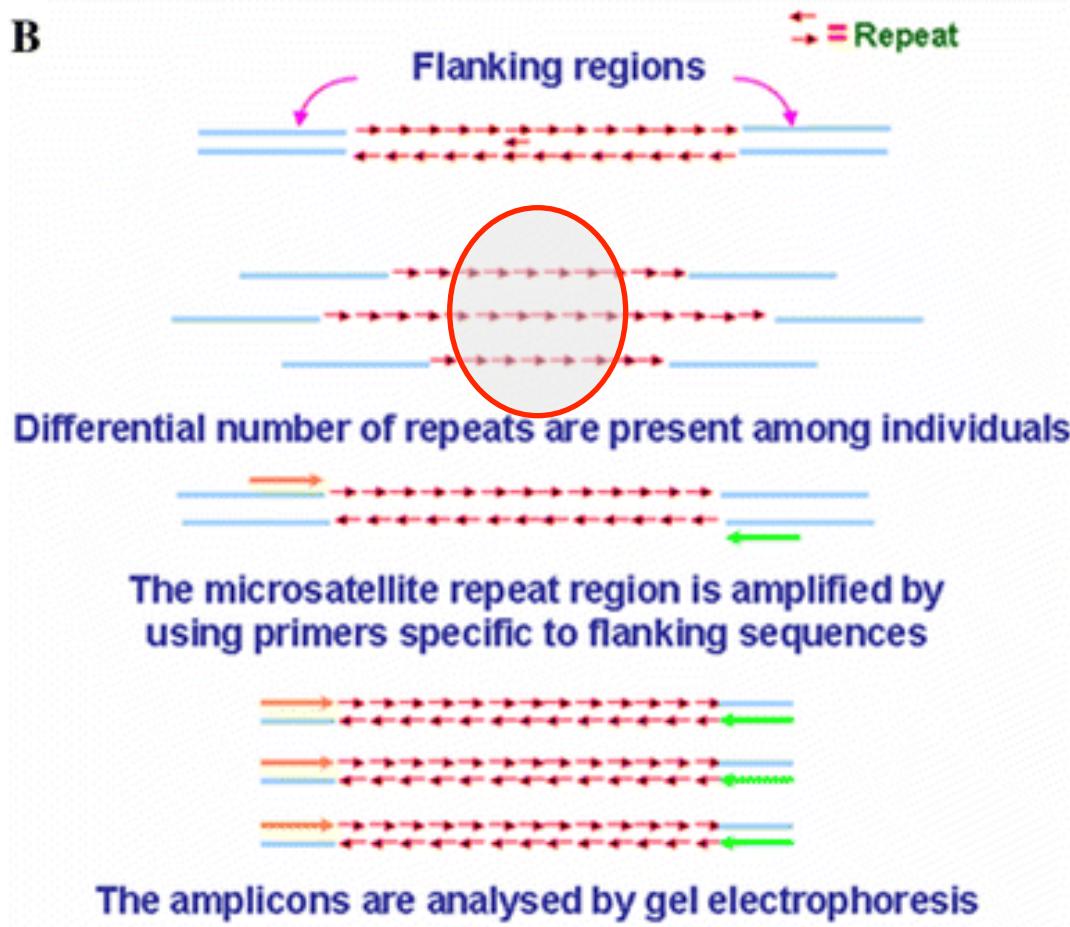
OR

2. Use arbitrary sequences to prime multiple loci in the genome



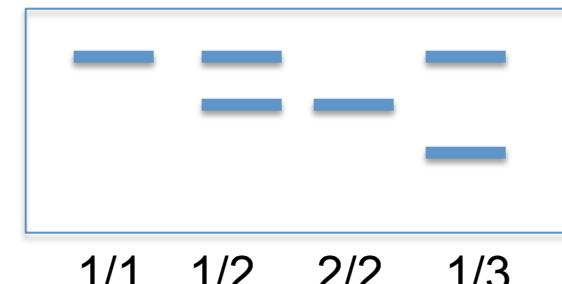
SSR (Simple Sequence Repeat)

B



Also known as:

- Microsatellites
- Short Tandem Repeats



Agarwal et al., 2008

SSR

Repeated units of nucleotide motifs < 10bp

- Examples: (CA) n , (AAT) n , (GATA) n

Widely distributed in eukaryotic genomes

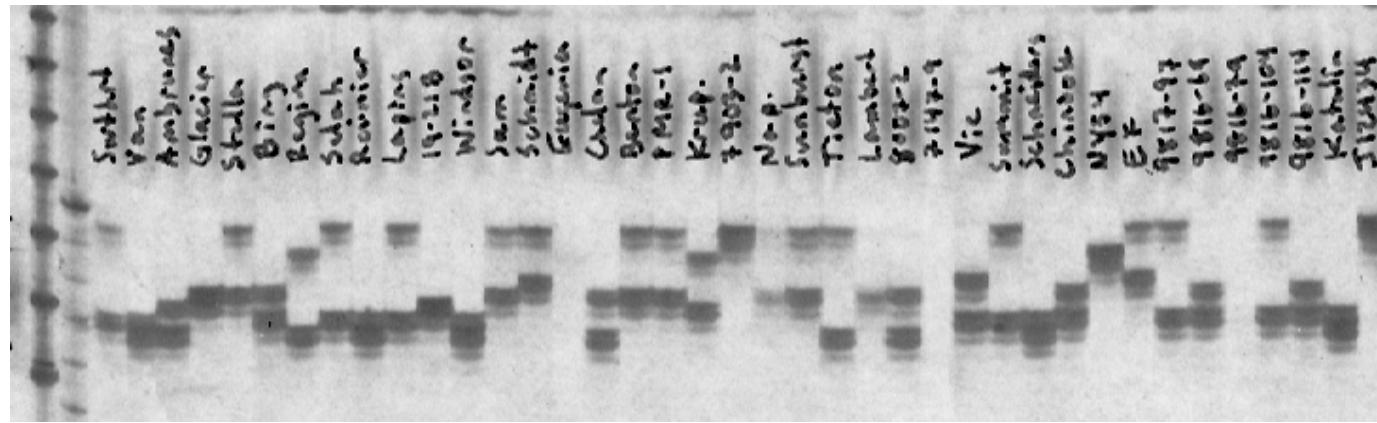
High level of allelic variation – Hypervariable

- DNA strand mis-pairing during replication

Primers designed from unique sequence in the repeat- flanking regions

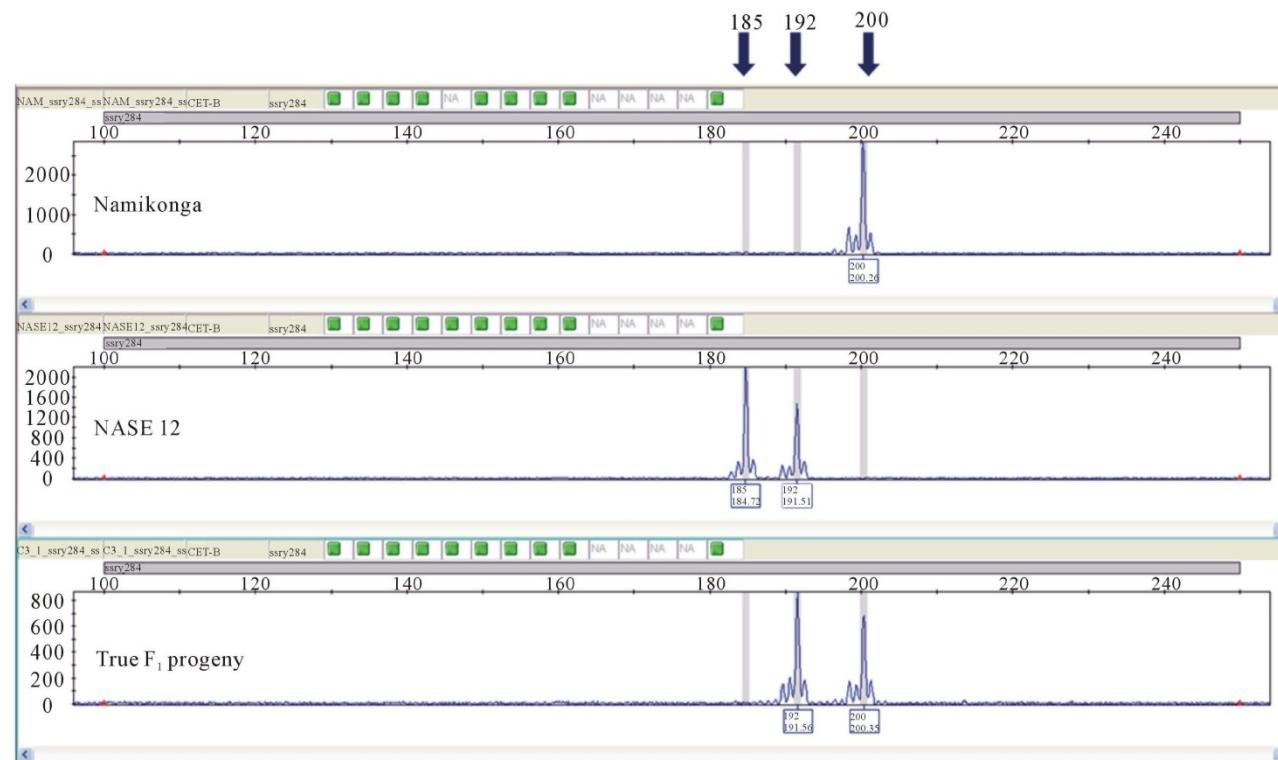


SSR



Silver stained
Poly-
acrylamide gel

Parental survey
for allele
differences in
cherry



Capillary
electrophoresis
of fluorescently
labeled SSR
markers in
cassava (Vincent
et al. American J
Plant Sci 5: 119)

PCR-Based Methods

1. Require prior sequence knowledge for primer development
2. Use short, arbitrary sequences to prime multiple loci in the genome
 - Randomly amplified polymorphic DNA (RAPD)
 - Amplified fragment length polymorphisms (AFLP)

RAPD (Random Amplified Polymorphic DNA)

- Utilizes small primers of arbitrary sequence
 - 10 base pair (10-mer) is common
- Primers vary in GC content
- Primers should bind to many different sites in the genome
- When a single primer binds twice within ~3000 bp in the correct orientation for PCR, amplification will occur



AFLP (Amplified Fragment Length Polymorphism)

- Proprietary technology (Keygene)
- AFLP combines restriction digestion with arbitrary primer amplification
- No prior sequencing needed
- High level of polymorphism for each selective primer combination
- Also usually dominant markers



SNP Genotyping Examples

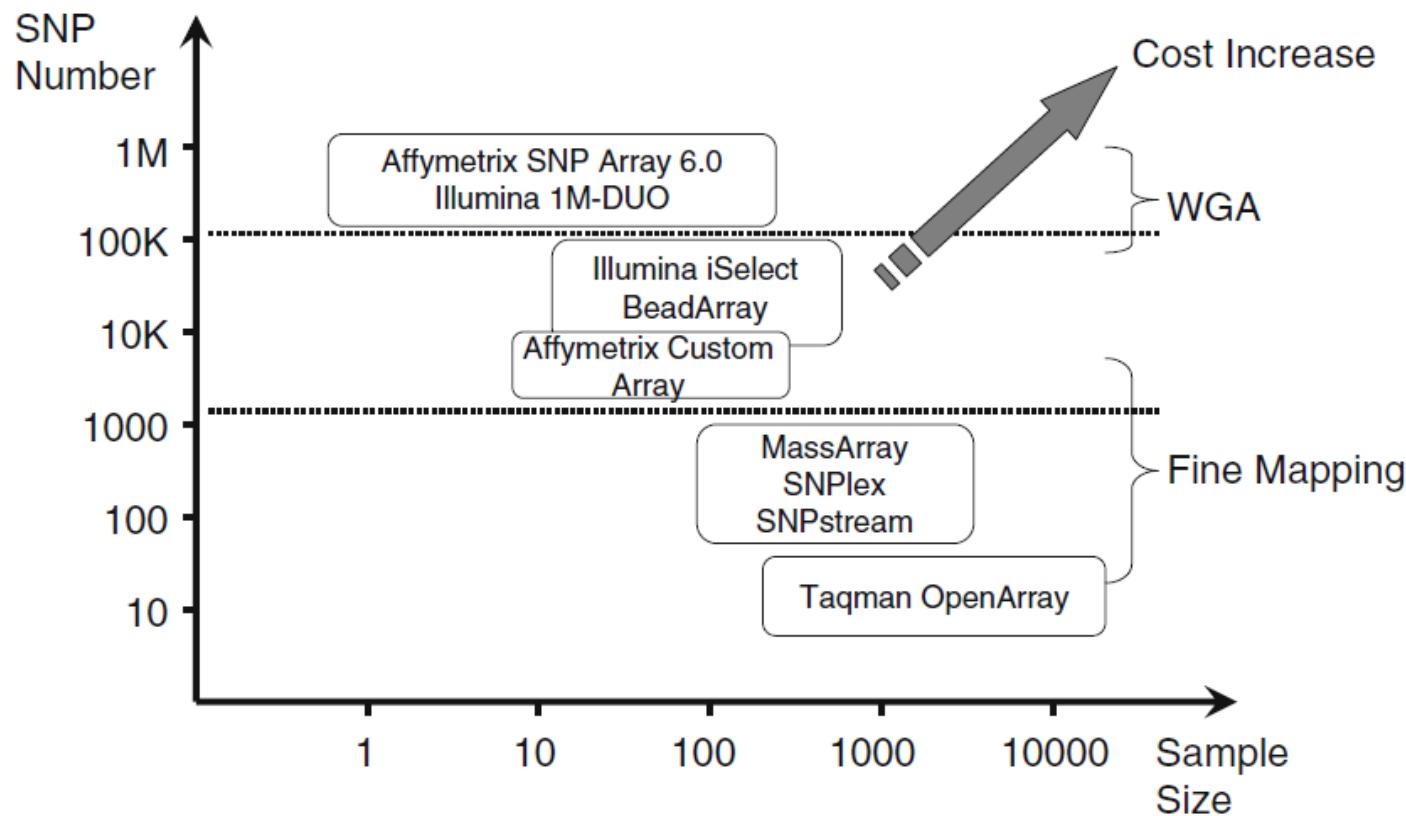
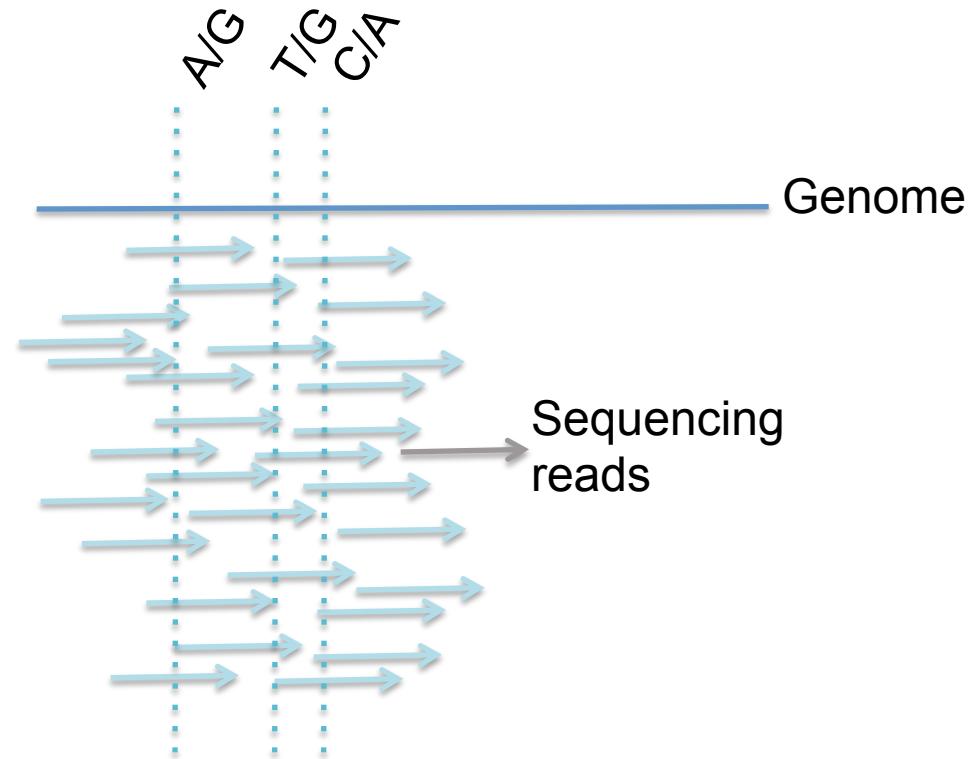
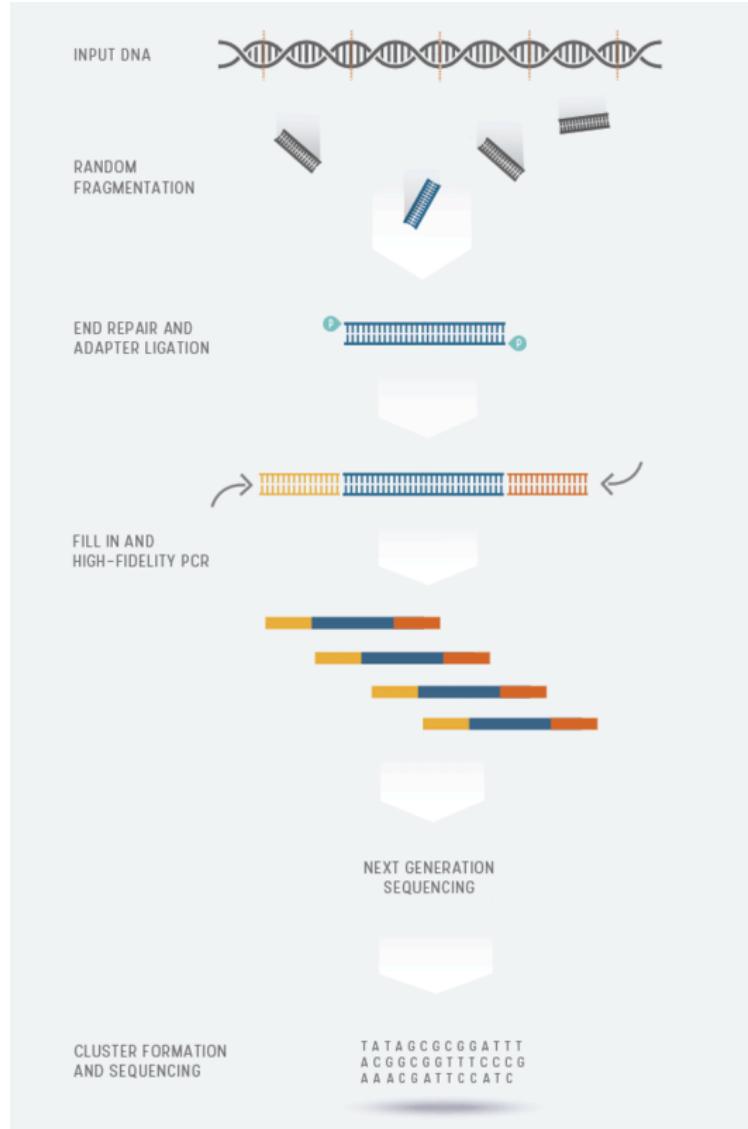


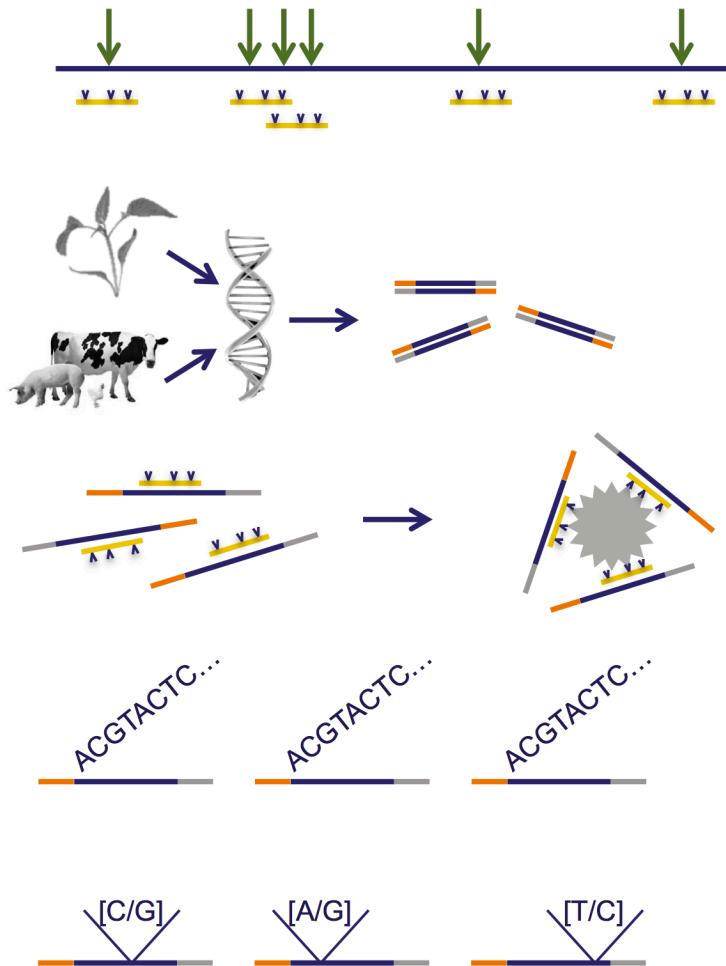
Fig. 16.1. An overview of platforms with regard to throughput of single nucleotide polymorphisms and sample size. Platforms are selected on the basis of reasonable running costs.

Ding and Jin, 2009

Whole Genome Sequencing



Sequence Capture



Custom probe design and synthesis

- Probe sets readily available for several species
- Custom probe design for maximum performance

Library preparation

- Robust, high-throughput and automated for scalability

Capture enrichment

- Automated and scalable process for any project size

Sequencing

- Various sequencing methods available to satisfy on project's demand

Data analysis

- Identification of SNPs and custom bioinformatics needs

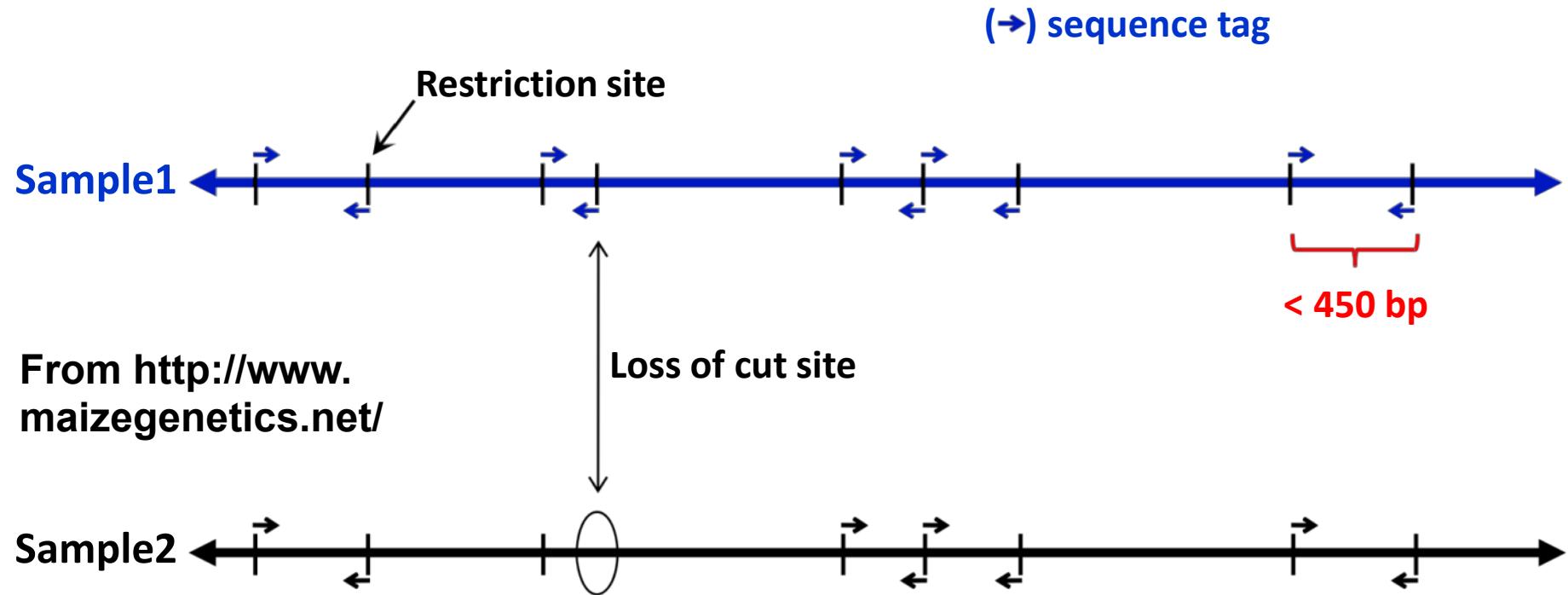


Regions of interest

Genome

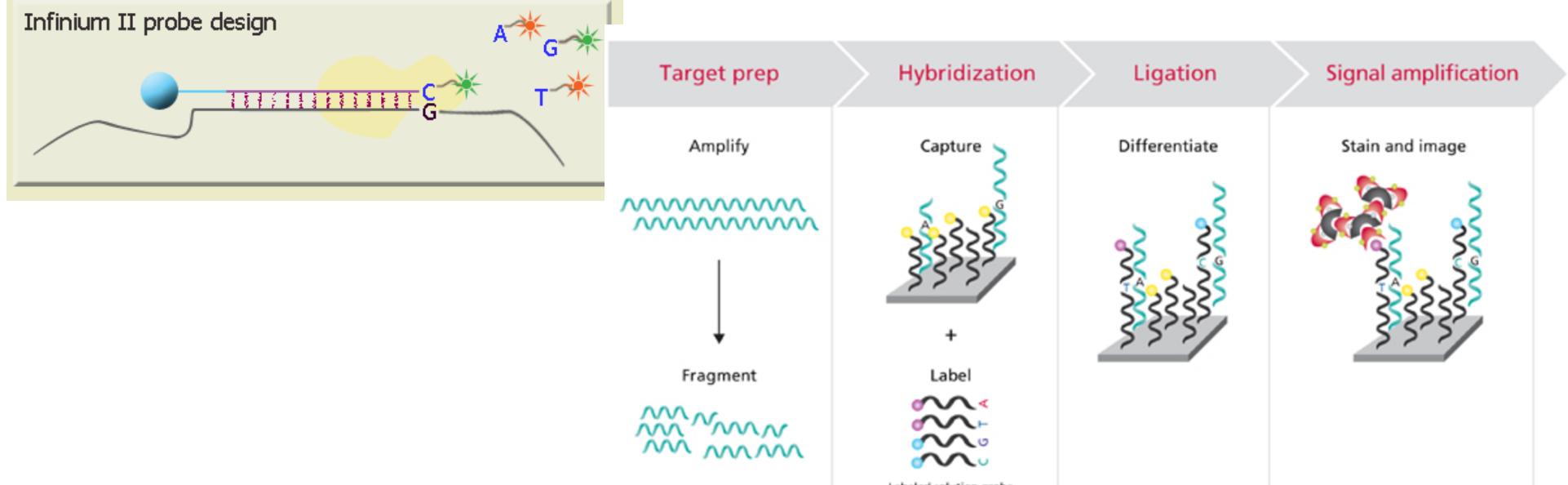
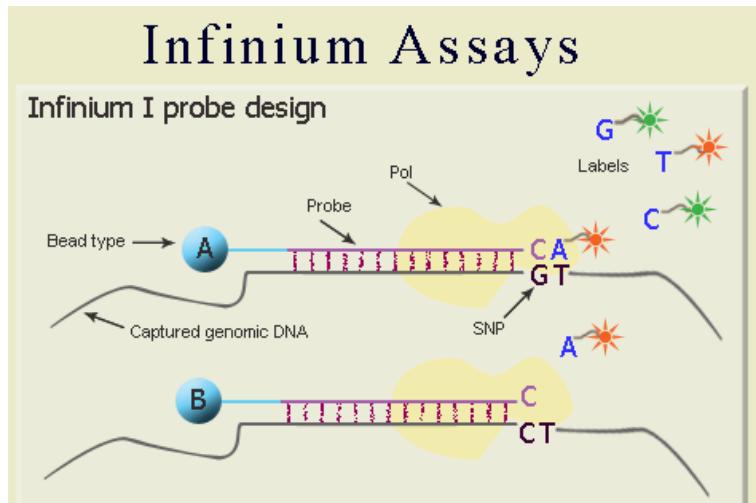
Probes

Genotype by sequencing (GBS)

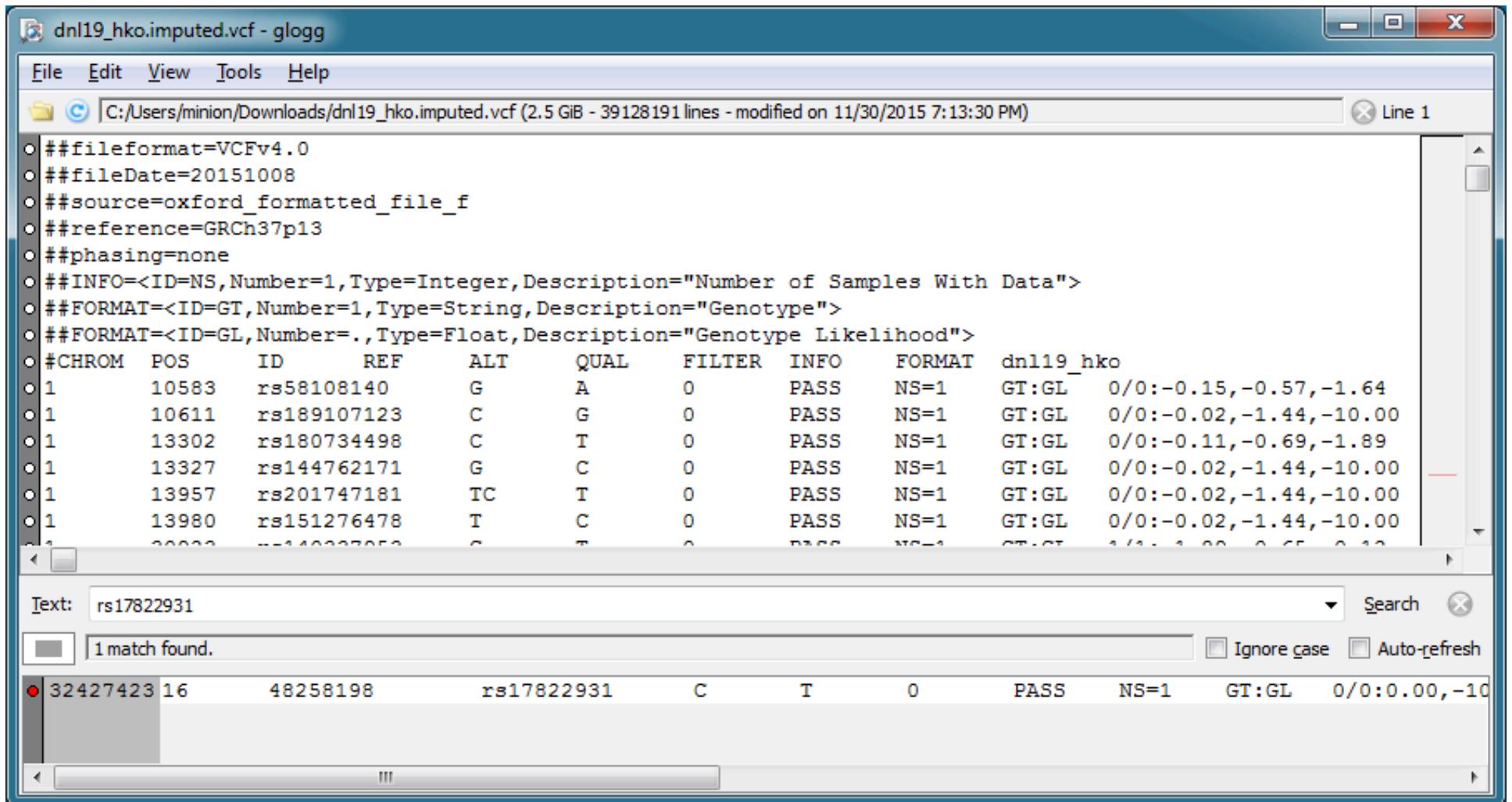


- Focuses NextGen sequencing power to ends of restriction fragments
- Scores both SNPs and presence/absence markers

SNP chips



Data format (VCF)



The screenshot shows a Windows application window titled "dn19_hko.imputed.vcf - glogg". The menu bar includes File, Edit, View, Tools, and Help. The toolbar has icons for Open, Save, and Close. The status bar shows "Line 1". The main area displays a VCF file with the following header and data:

```
##fileformat=VCFv4.0
##fileDate=20151008
##source=oxford_formatted_file_f
##reference=GRCh37p13
##phasing=none
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GL,Number=.,Type=Float,Description="Genotype Likelihood">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT dn19_hko
1 10583 rs58108140 G A 0 PASS NS=1 GT:GL 0/0:-0.15,-0.57,-1.64
1 10611 rs189107123 C G 0 PASS NS=1 GT:GL 0/0:-0.02,-1.44,-10.00
1 13302 rs180734498 C T 0 PASS NS=1 GT:GL 0/0:-0.11,-0.69,-1.89
1 13327 rs144762171 G C 0 PASS NS=1 GT:GL 0/0:-0.02,-1.44,-10.00
1 13957 rs201747181 TC T 0 PASS NS=1 GT:GL 0/0:-0.02,-1.44,-10.00
1 13980 rs151276478 T C 0 PASS NS=1 GT:GL 0/0:-0.02,-1.44,-10.00
1 20000 --14000000000 C T ^ PASS NS=1 GT:GL 1/1: 1.00 0.65 0.10
```

A search bar at the bottom left contains "Text: rs17822931" with a "Search" button. Below it, a message says "1 match found." with checkboxes for "Ignore case" and "Auto-refresh". The search results are shown in a table:

CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	dn19_hko
1	32427423	16	48258198	rs17822931	C	T	0	PASS	NS=1 GT:GL 0/0:0.00,-10

<https://dna.land/static/img/vcf-info/win/13-glogg-snp-found.png>

Summary

- Many types of molecular markers available
- Type(s) chosen for use will depend on many factors
- As sequencing costs become lower, SNPs and sequence tags are supplanting length polymorphisms
- Marker polymorphism is the basis of linkage mapping, genomic selection and molecular breeding

