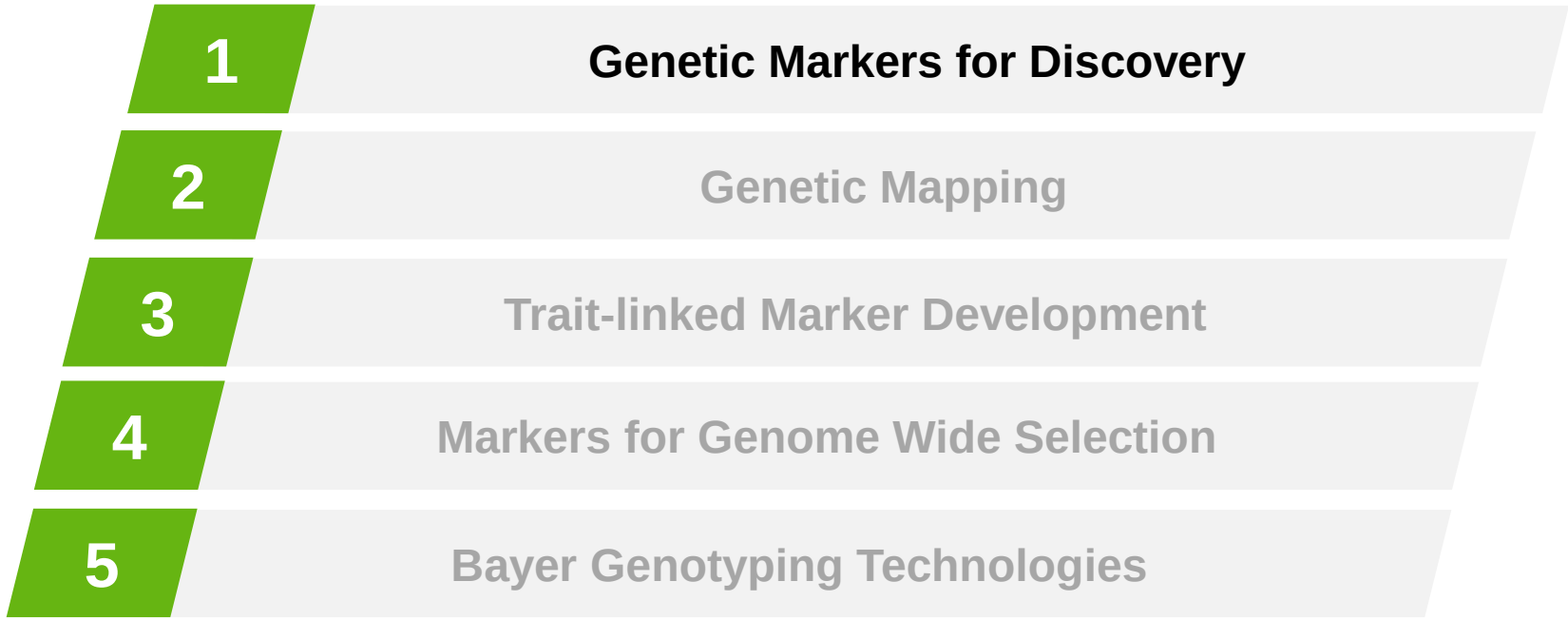
Discovery & Development of Genetic Markers for use in High Throughput Labs

Bayer Russia Molecular Marker Training

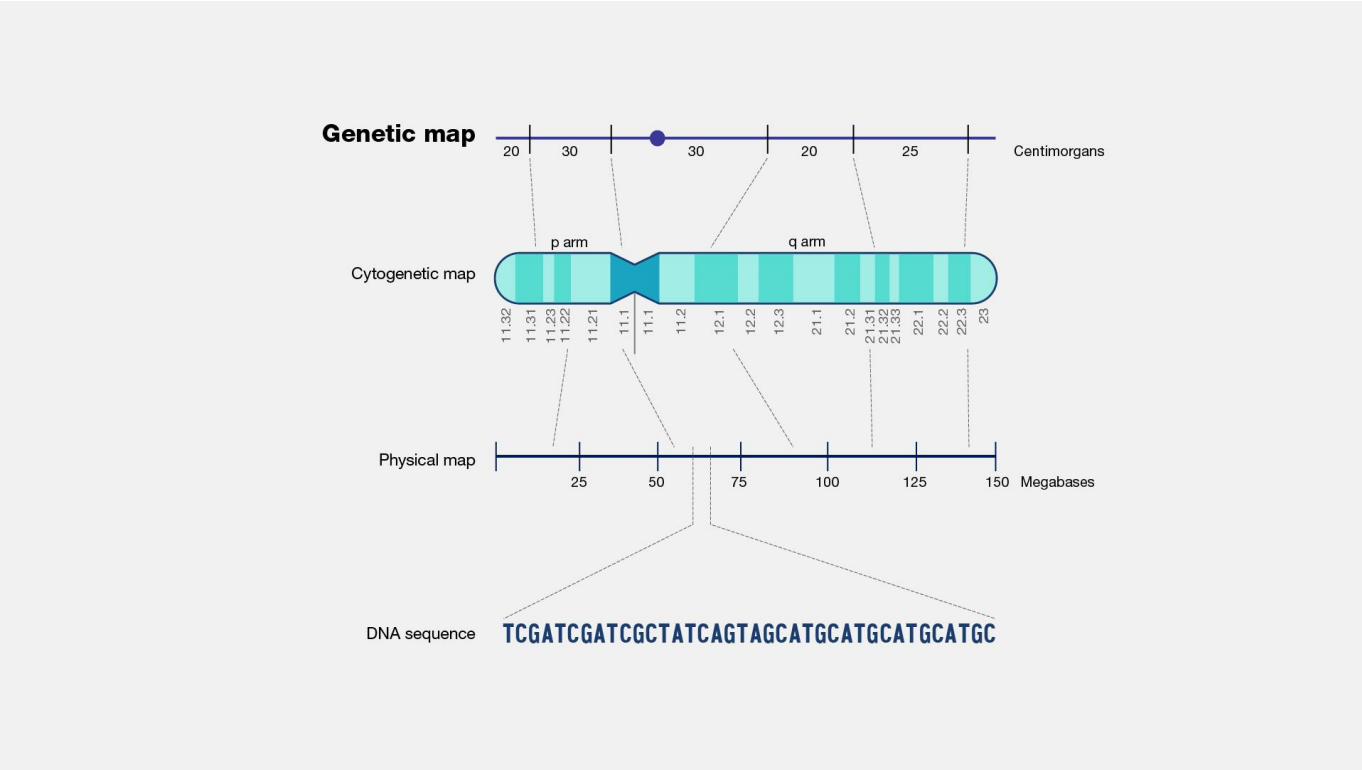
25 July 2023

Bayer Russia Molecular Marker Training :
Genetic Markers and Technologies



What is genetic marker?

A genetic marker is a DNA sequence with a known physical location on a chromosome.



https://www.genome.gov/genetics-glossary/Genetic-Map

Genetic markers available

Genetic marker can be described as a variation (which may arise due to mutation or alteration in the genomic loci) that can be observed.

What is an ideal genetic marker?

- co-dominant
- evenly distributed throughout genome
- highly reproducible
- easy to detect higher level of polymorphism

Most common genetic markers

- single nucleotide polymorphisms (SNPs)
- simple sequence length polymorphisms (SSLPs)
- insertions/deletions

SNP

TCGATCCAAATAATTAAATTA
TCGATCCAAACAATTAAATTA

Insertions/deletions

TAAGATCCTTCAGGAAAATTCCAAAAGGGATGCAACTG
TAAGATCCTTCAGGAAAA----------AAGGGATGCAACTG

SSLP

TACAATAATTCTATATATAAGTCACGAACAGAAAATG
TACAATAATTCTATATATATATATATATATATAAGTCACGAACAGAAAATG

Genetic markers available

Genetic Markers

Classical markers

Morphological
- Phentotype (visible)

Biochemical
- Isozyme
- Protein banding

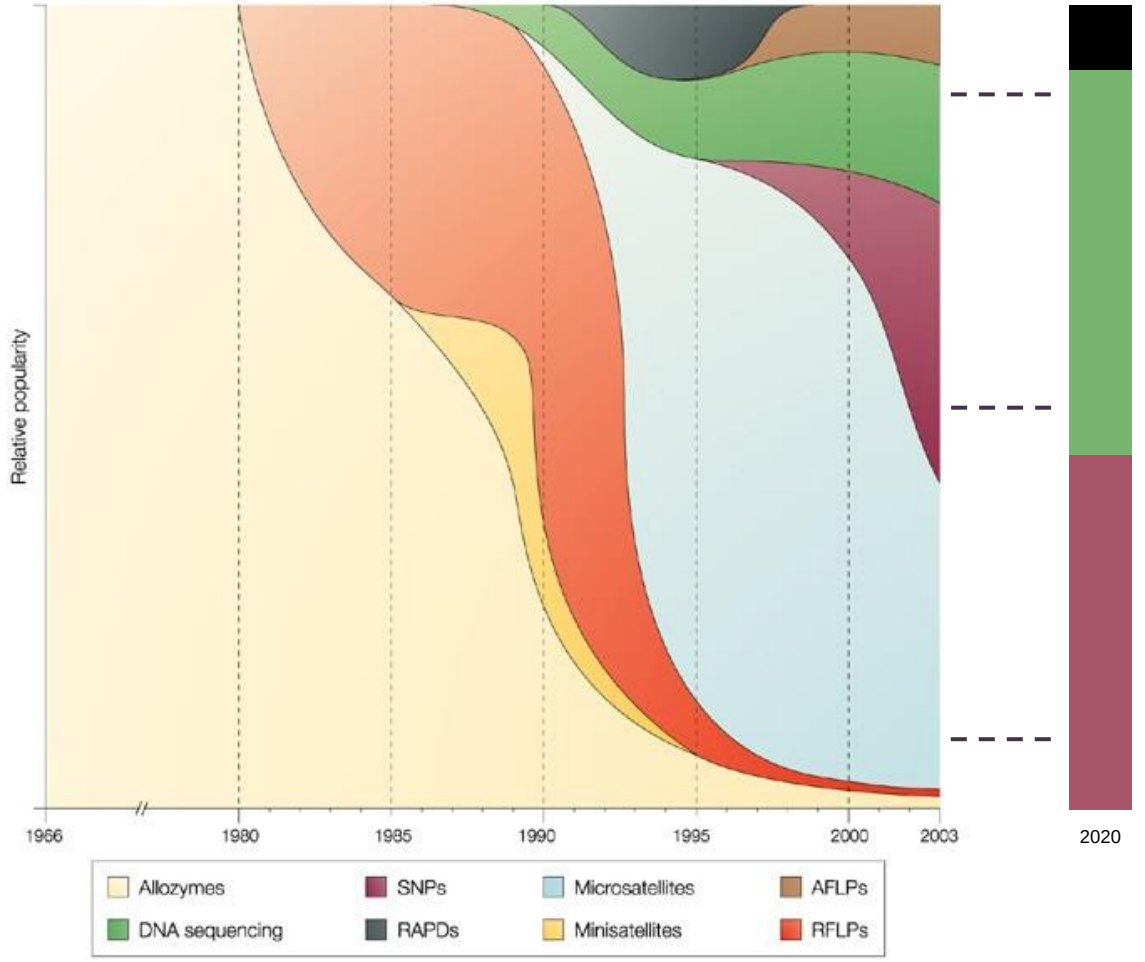
Cytological
- C-banding
- G-banding

DNA markers

PCR-based
- RAPD
- AFLP
- CAPS
- SSR
- STS
- KASP -SNP

Hybridization-based
- RFLP
- DArT
- SilicoDArT
- Microarray (SNP chip)

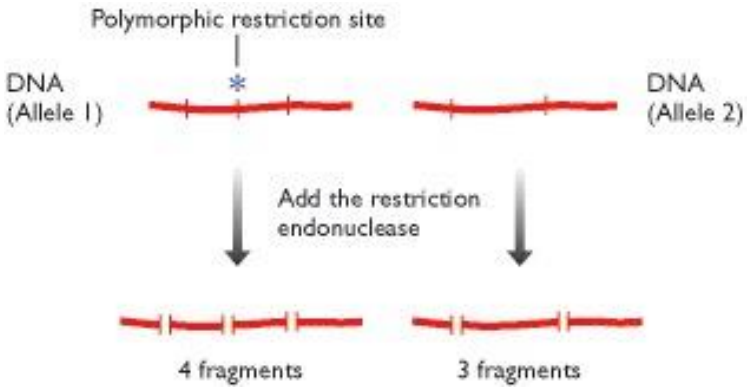
Sequencing-based
- GBS-SNP
- DArTSeq-SNP

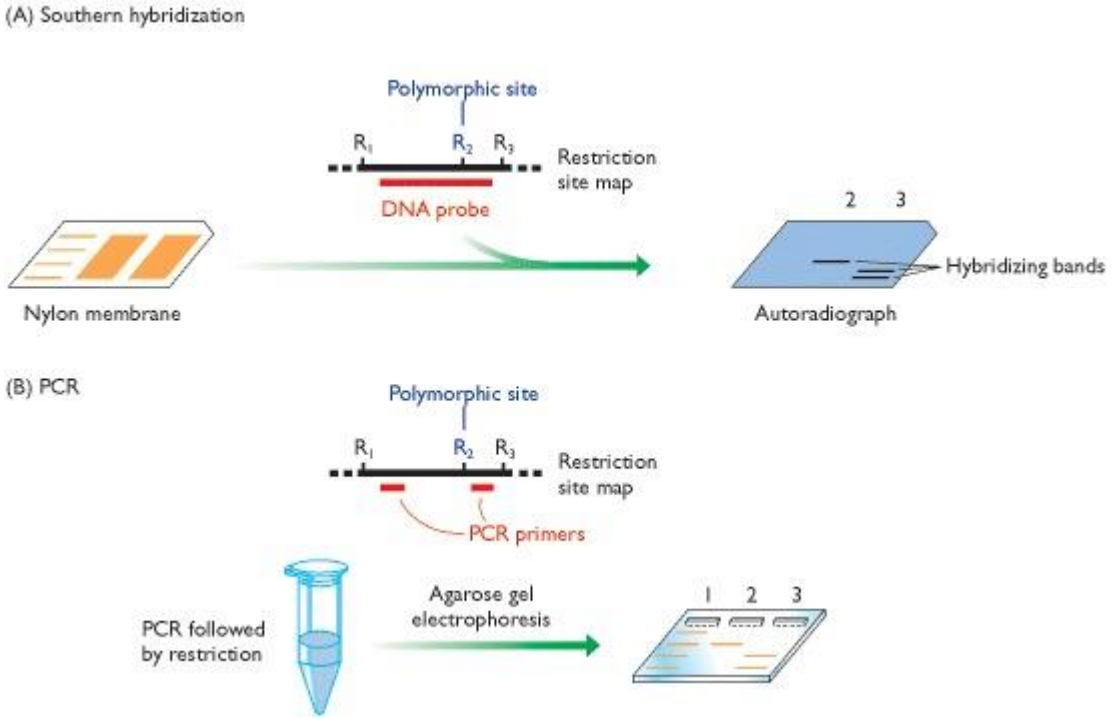


https://doi.org/10.1038/nrg1249

Nature Reviews | Genetics

Restriction Fragment Length Polymorphism (RFLP)

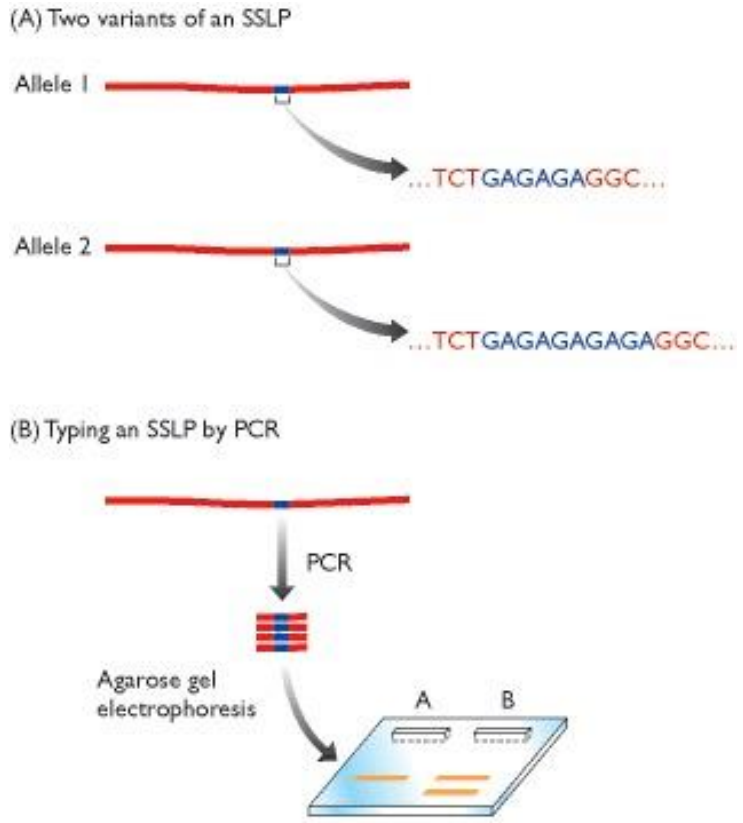




https://www.ncbi.nlm.nih.gov/books/NBK21116/

- co-dominant
- highly locus-specific
- time consuming and labor intensive

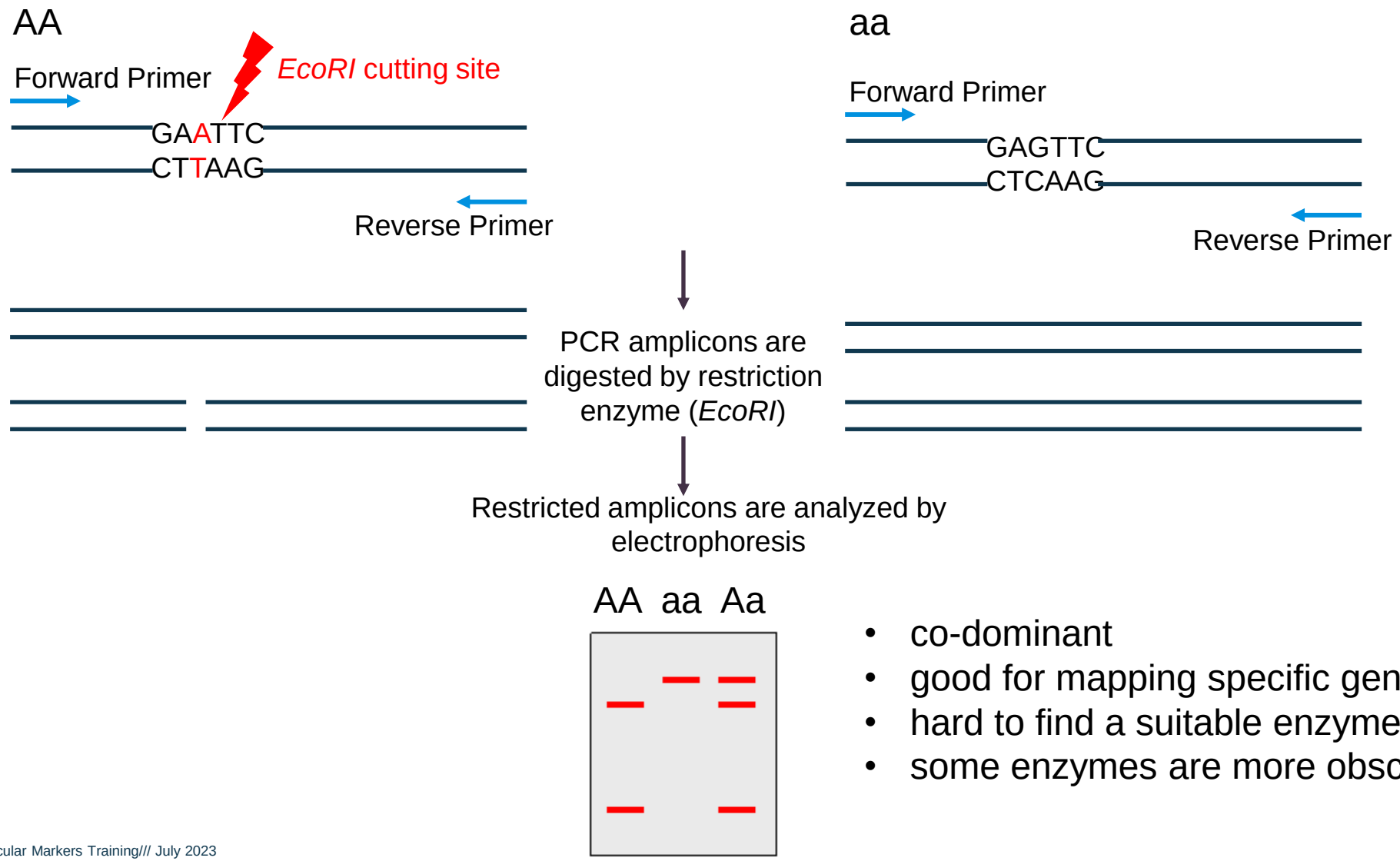
Simple Sequence Repeat (SSR) assay



https://www.ncbi.nlm.nih.gov/books/NBK21116/

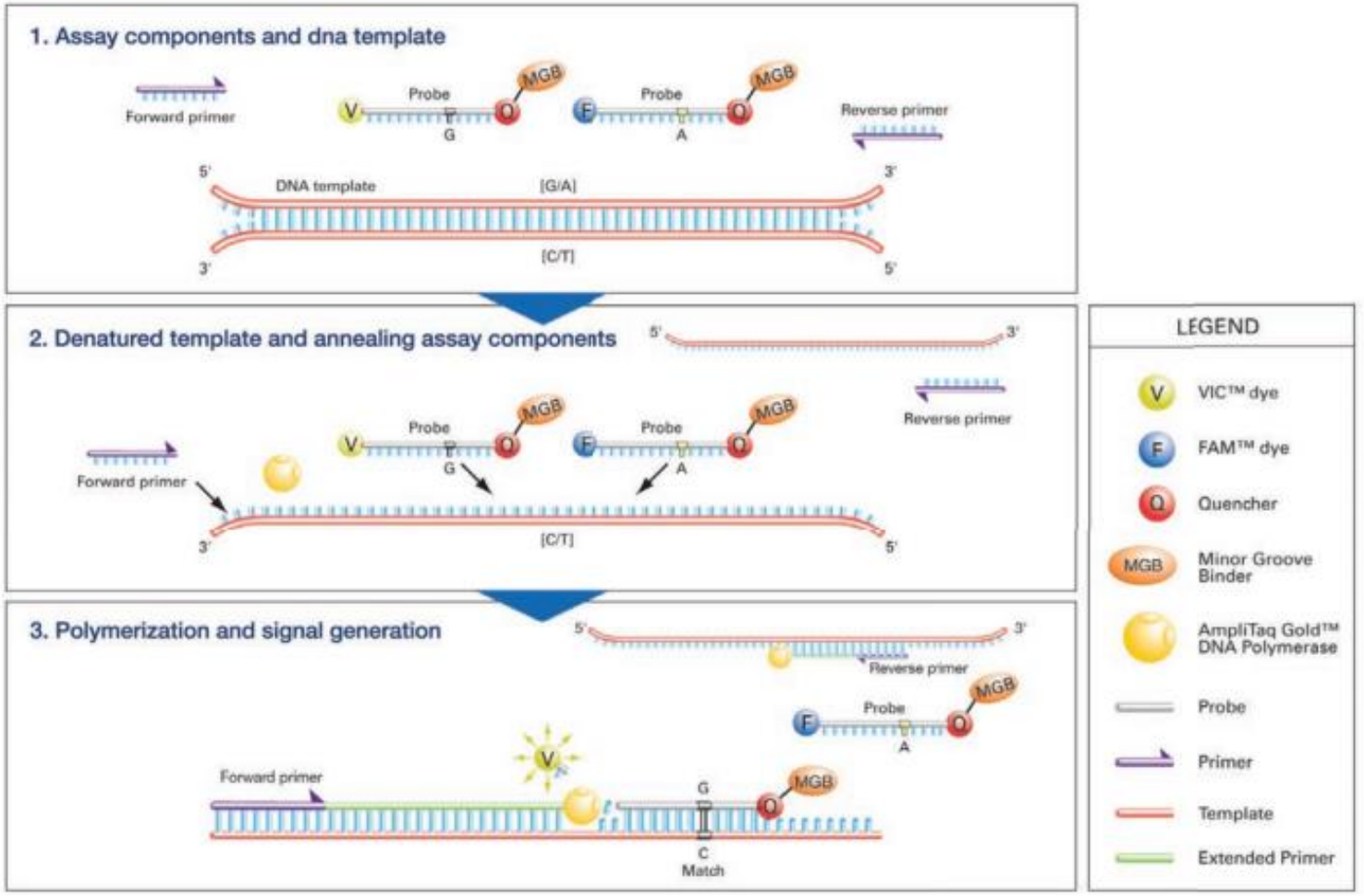
- co-dominant
- multiple alleles
- discovery is time consuming

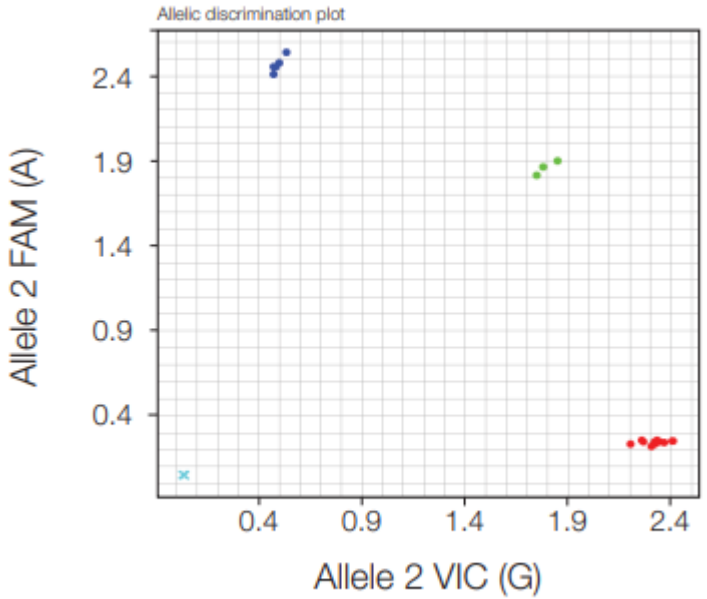
Cleaved Amplified Polymorphic Sequence (CAPS) assay



- co-dominant
- good for mapping specific genes
- hard to find a suitable enzyme
- some enzymes are more obscure/expensive

TaqMan SNP genotyping assay

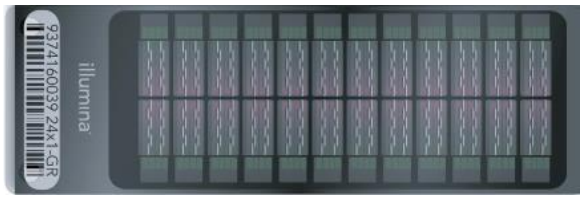


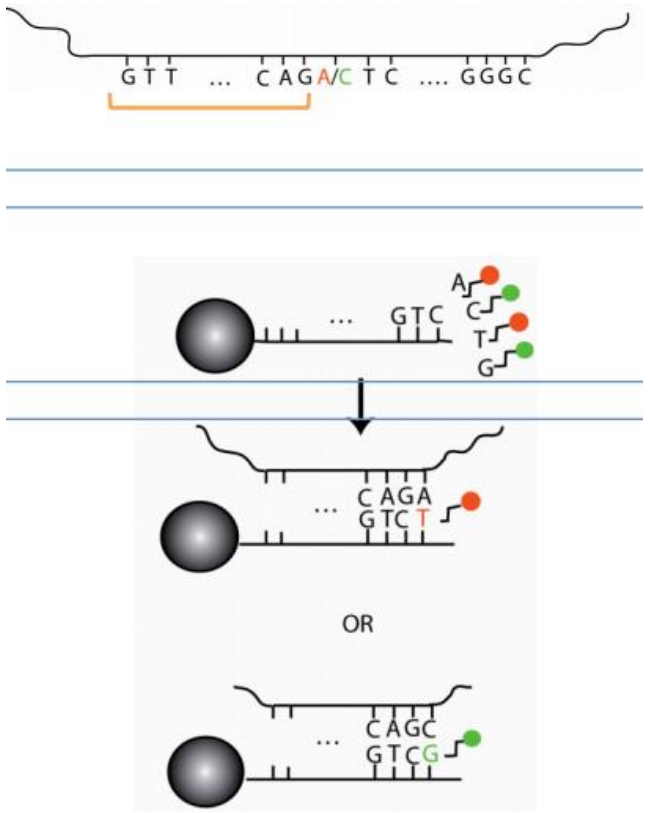


- co-dominant
- good for mapping specific genes
- probe is expensive

https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_040597.pdf

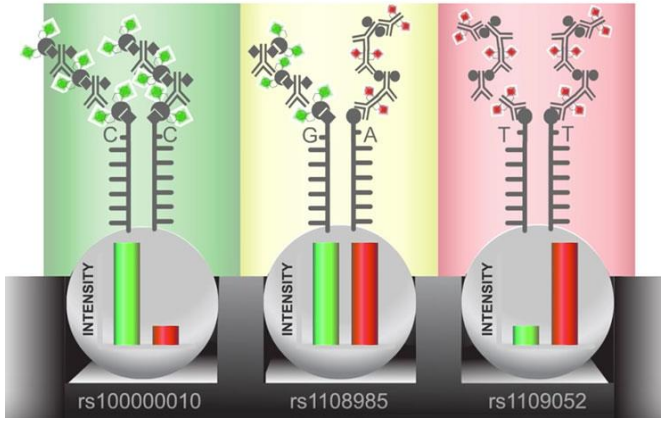
Single nucleotide polymorphism (SNP) arrays

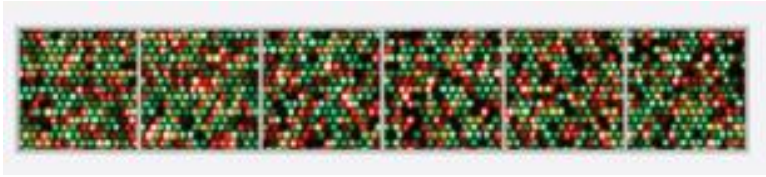




- An Illumina BeadChip contains up to millions of bead types.
- One bead type corresponds to each allele per SNP locus.
- Hundreds of thousands to millions of genotypes for a single individual can be assayed at once.

- Attached to each Illumina bead is a 50-mer sequence complementary to the sequence adjacent to the SNP site.
- The single-base extension (**T** or **G**) that is complementary to the allele carried by the DNA (**A** or **C**, respectively) then binds and results in the appropriately-colored signal (red or green, respectively).





https://www.illumina.com/science/technology/microarray.html

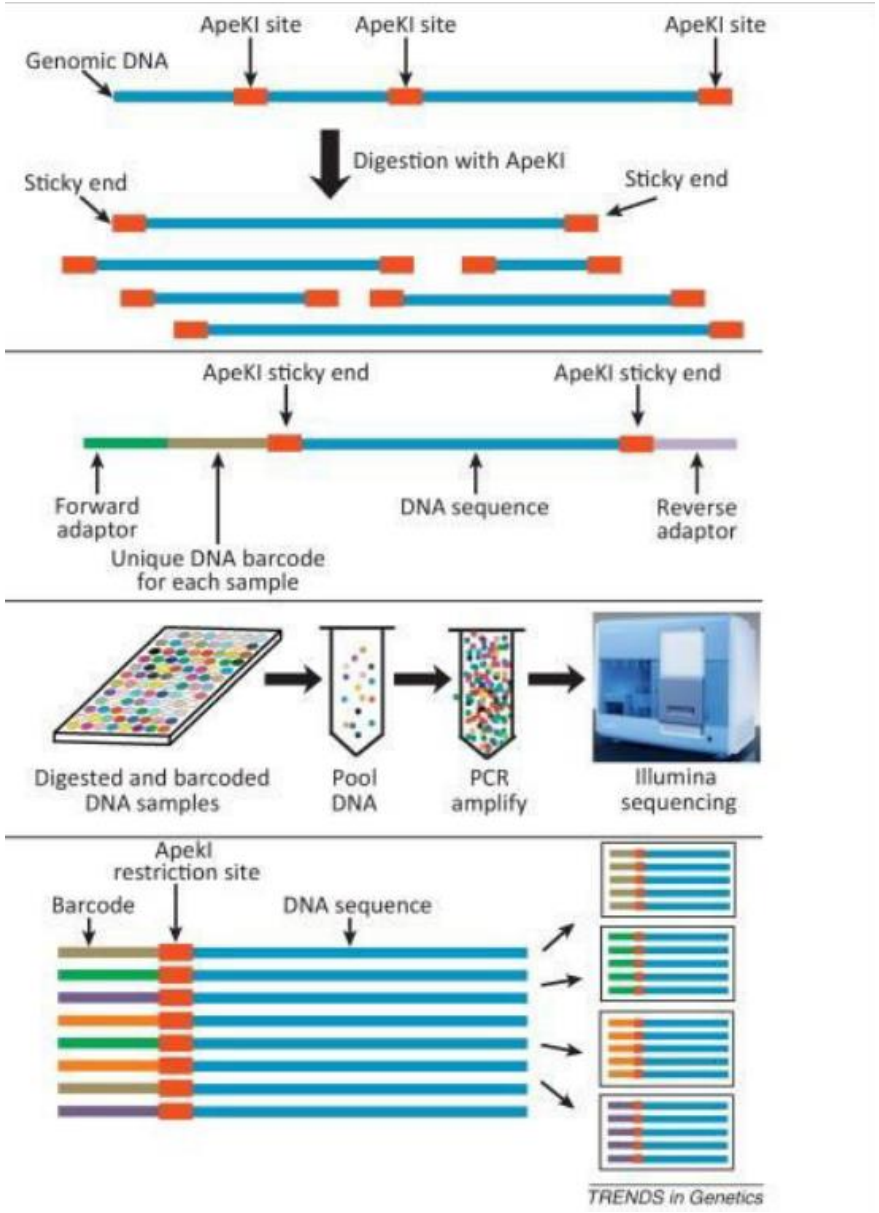
Genotyping by sequencing (GBS)

Step 1
construct reduced representation libraries by digesting each DNA sample with a restriction enzyme (ApeKI).

Step 2
Ligate custom “barcoded adaptors to sticky ends of restriction site. Each sample has its own unique barcode sequence.

Step 3
Pool digested and barcoded DNA into a single tube. Perform PCR amplification, library preparation, and sequencing on illumina platform.

Step 4
Use barcodes to assign sequences to samples. Provide a file of DNA sequence data for each sample.



Genotyping by sequencing (GBS)

--- a multiplex targeted sequencing method

Step 1 Select a set of SNPs obtained through GBS and design a panel of multiplexed PCR primers based on flanking sequences of SNPs.

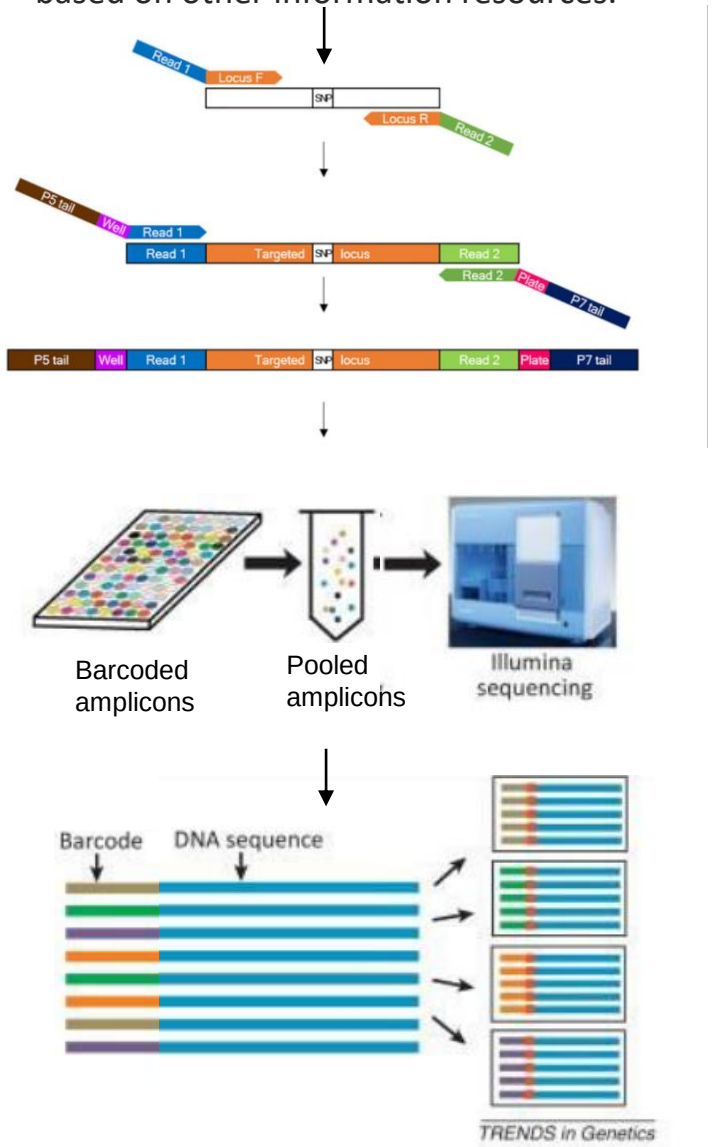
Step 2 Multiple target loci by amplified from a panel of multiplexed primers during first PCR step.

Step 3 Ligate index barcoded adaptors to amplicons through second PCR step. Each sample has its own unique index.

Step 4 Quantify and pool each library for sequencing.

Step 5 Use barcodes to assign sequences to samples. Provide a file of DNA sequence data for each sample.

A set of selected SNP markers obtained through GBS or based on other information resources.



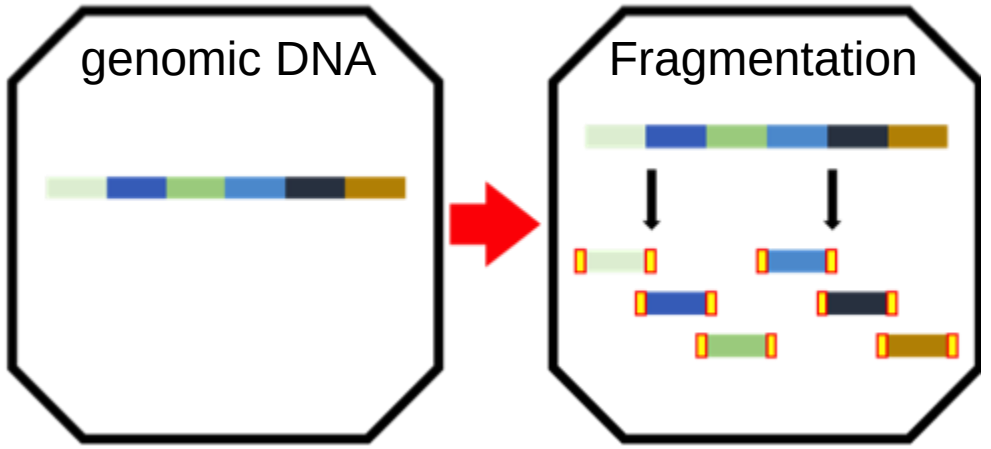
Low-coverage sequencing

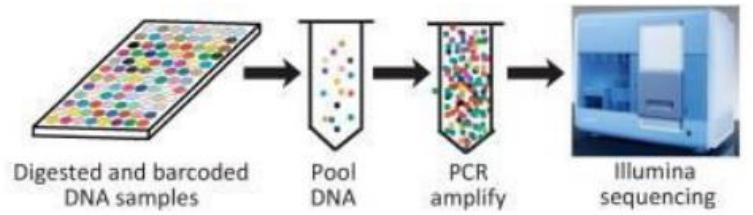
Step 1
Shear long genomic DNA into small fragments.

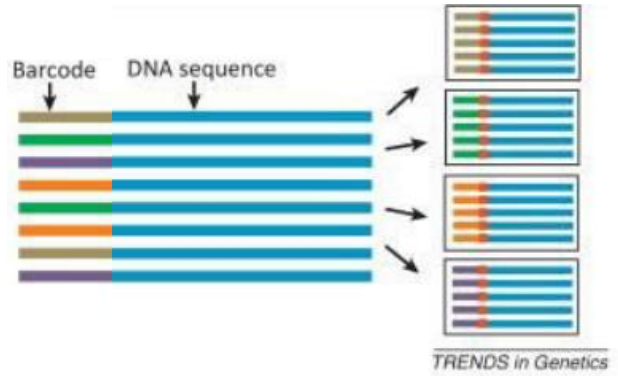
Step 2
Ligate custom “barcoded adaptors to fragmented DNAs. Each sample has its own unique barcode sequence.

Step 3
Pool barcoded DNA into a single tube. Perform PCR amplification, library preparation, and sequencing on illumina platform.

Step 4
Use barcodes to assign sequences to samples. Provide a file of DNA sequence data for each sample.





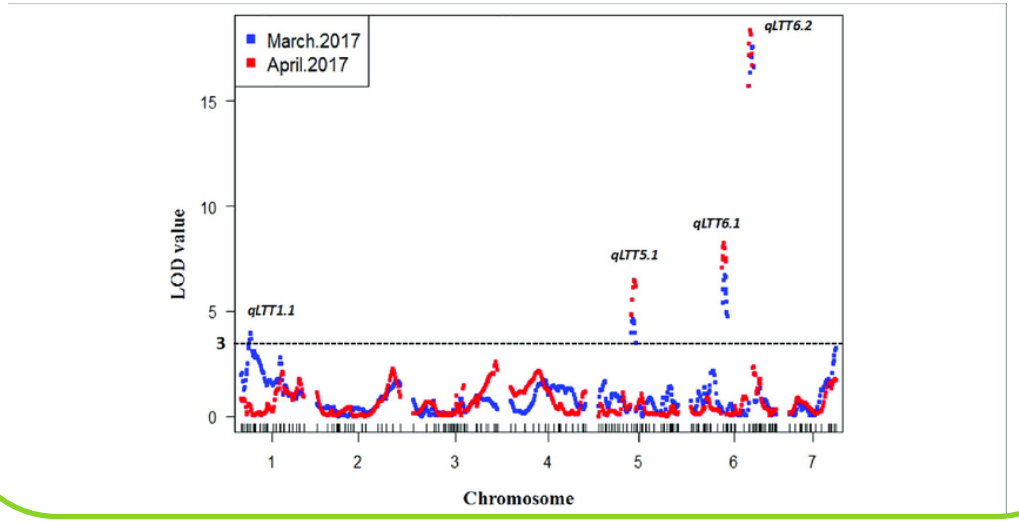


Uses of genetic markers---Genetic mapping

GWAS

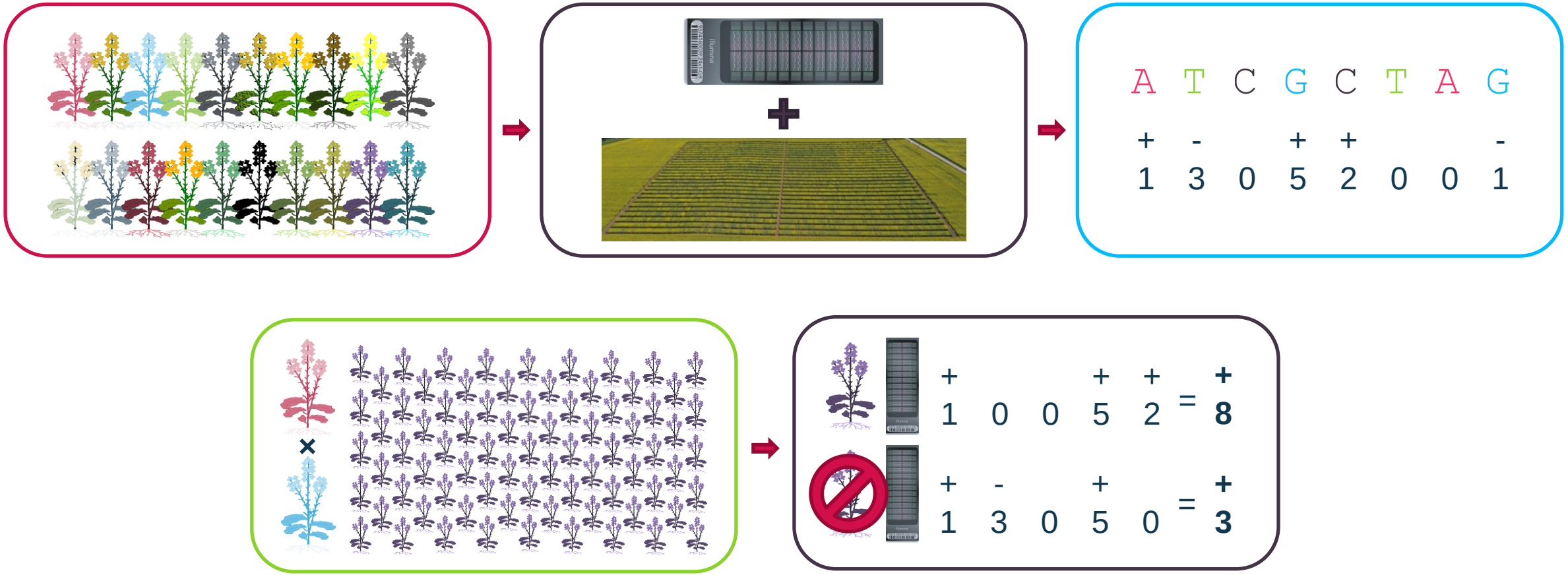
https://doi.org/10.3390/plants10050895

QTL analysis

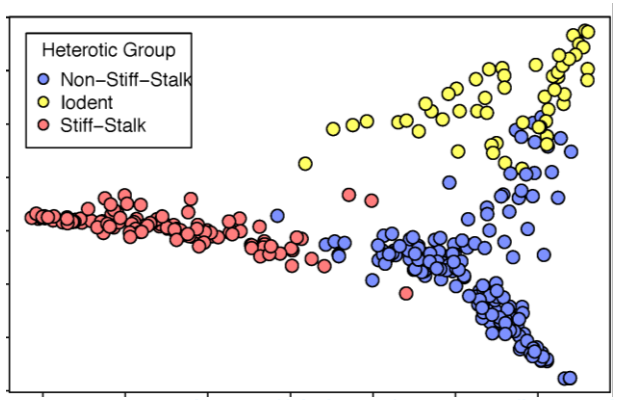


https://www.frontiersin.org/articles/10.3389/fpls.2019.01620/full

Uses of genetic markers
--- Marker-Assisted Selection using genome-wide markers

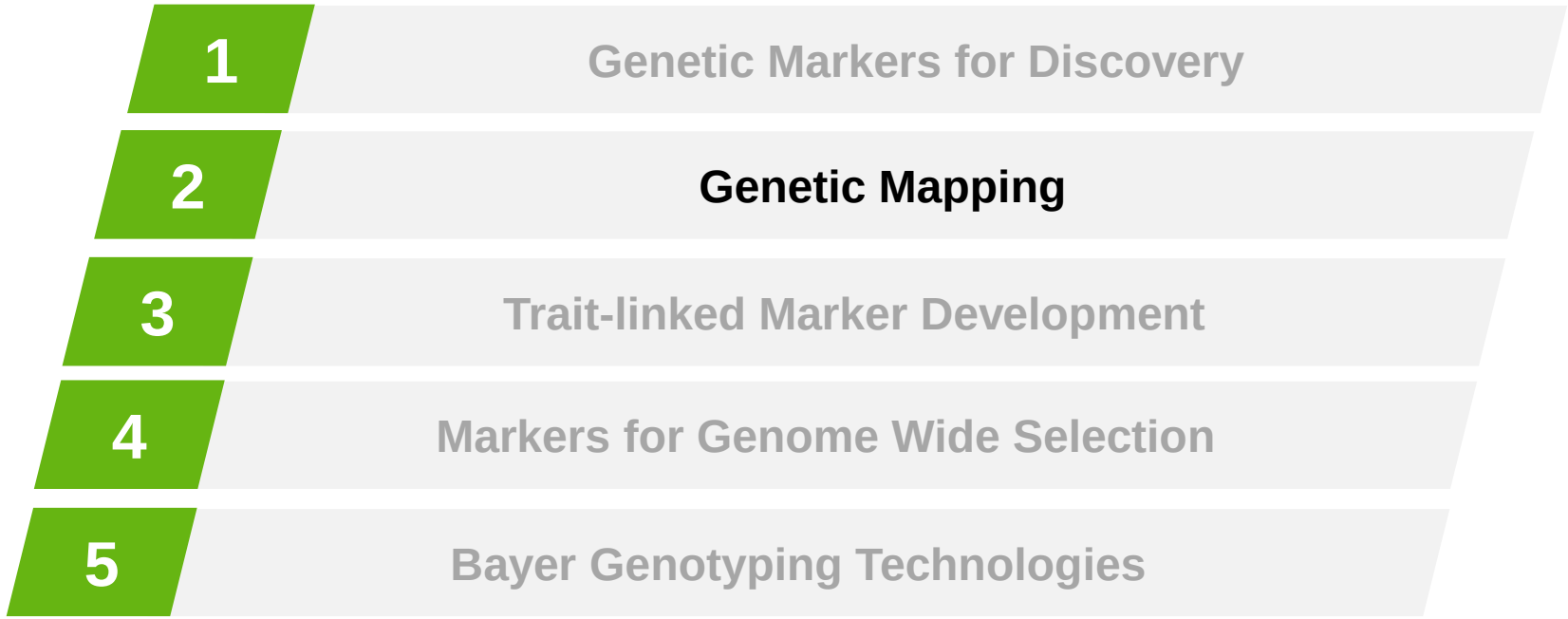


Uses of genetic markers—Genetic diversity analysis



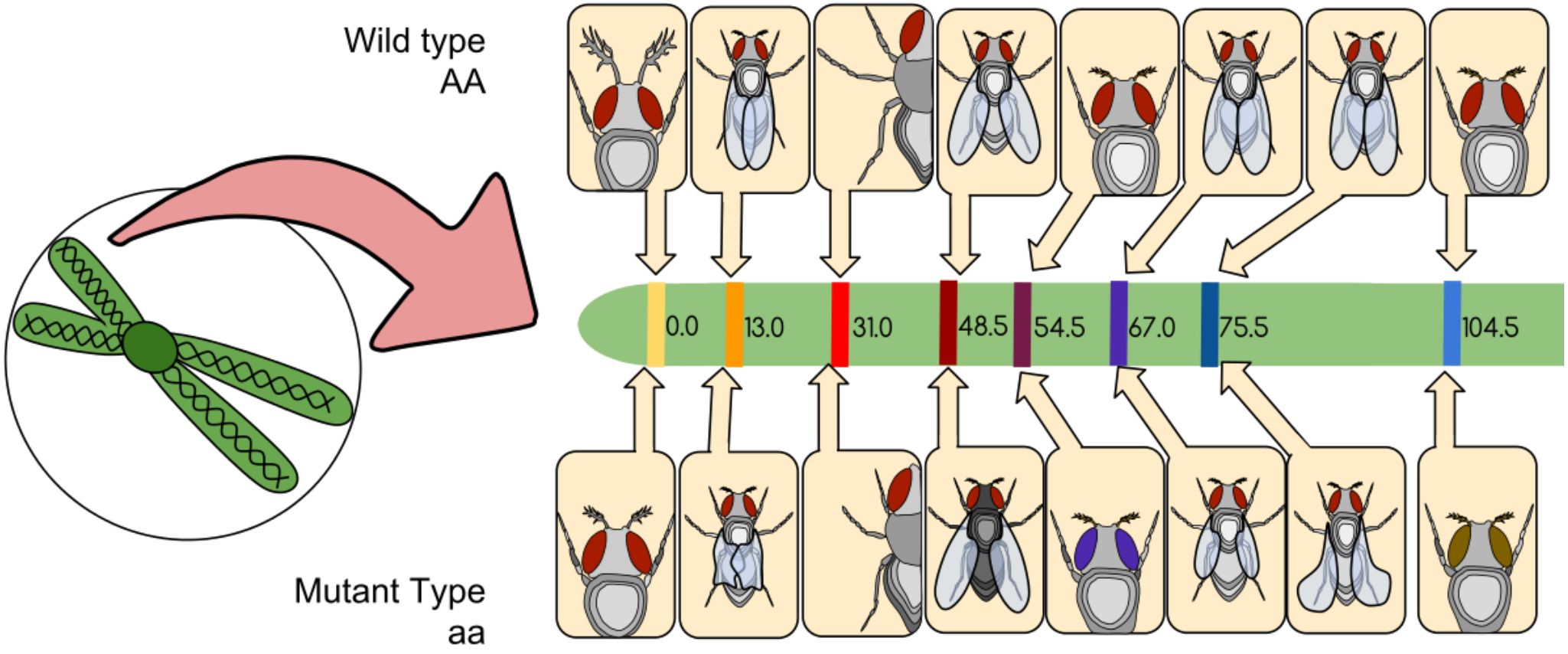
https://doi.org/10.1371/journal.pone.0189277

Bayer Russia Molecular Marker Training :
Genetic Markers and Technologies



What is genetic mapping?

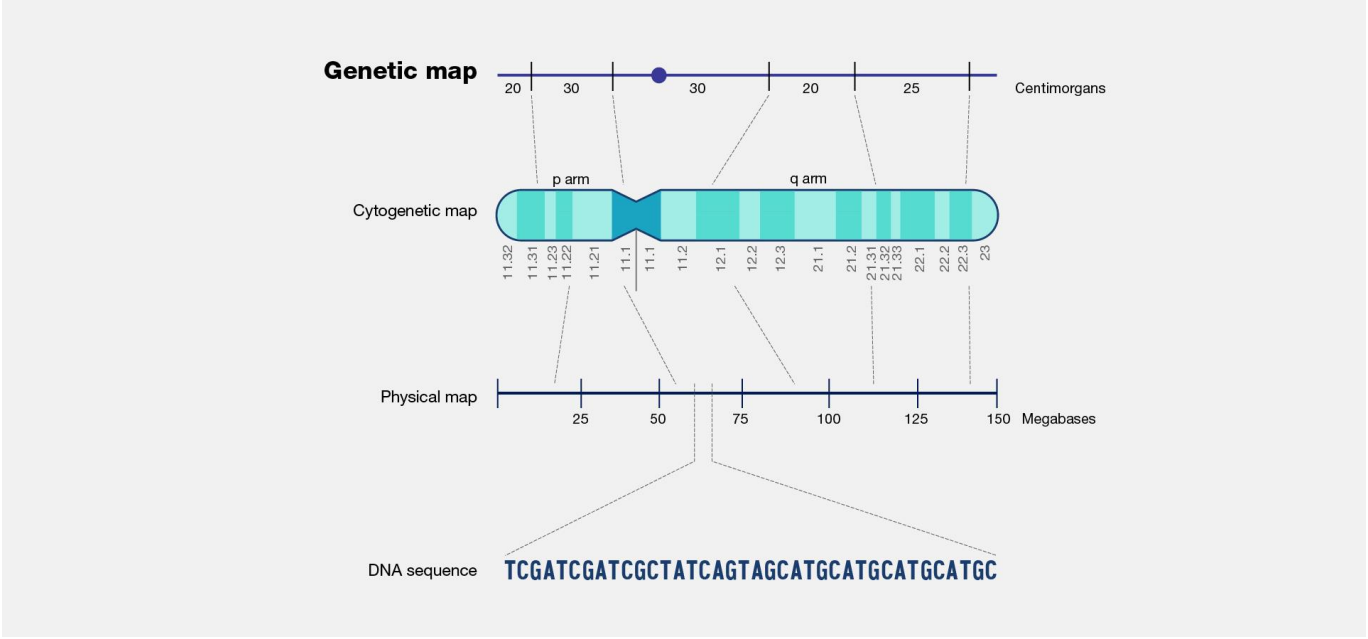
Gene mapping or genome mapping describes the methods used to identify the location of a gene on a chromosome and the distances between genes.



Mapping approaches

1. **genetic maps** (also known as linkage maps): distances are based on the genetic linkage information.
- offering insights into the nature of different regions of the chromosome
- recombination rates, and such rate is often indicative of gene-rich vs usually gene-poor regions of the genome

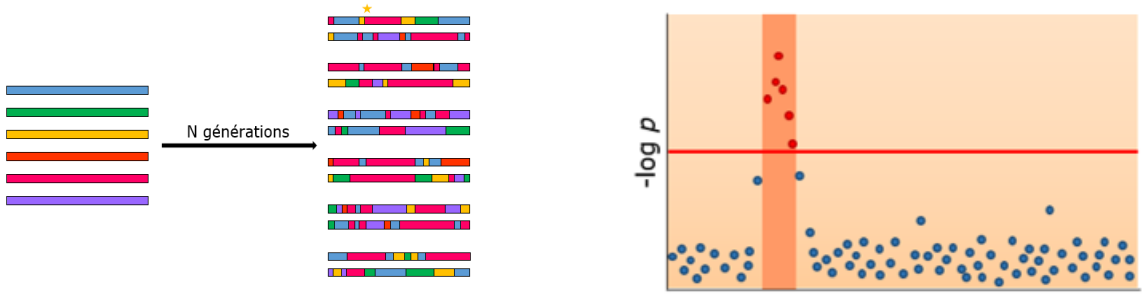
2. **physical maps**: use actual physical distances usually measured in number of base pairs.
- a more accurate representation of the genome



https://www.genome.gov/genetics-glossary/Genetic-Map

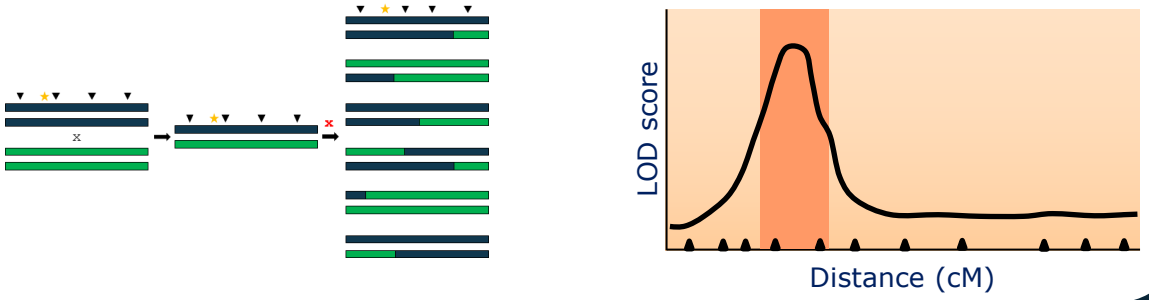
Native Trait discovery toolbox

GWAS – Genome Wide Association Study



- Uses historical data
- Scans large germplasm pools
- Requires data QC

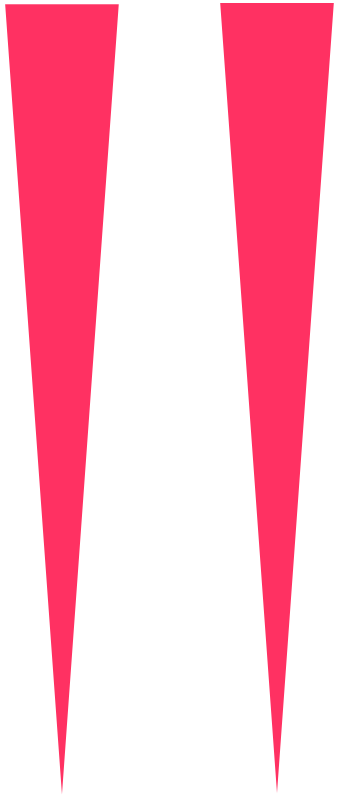
Bi-parental mapping



- Requires population development with contrasting parents for the trait of interest
- Targets specific germplasm pool

Amount of data

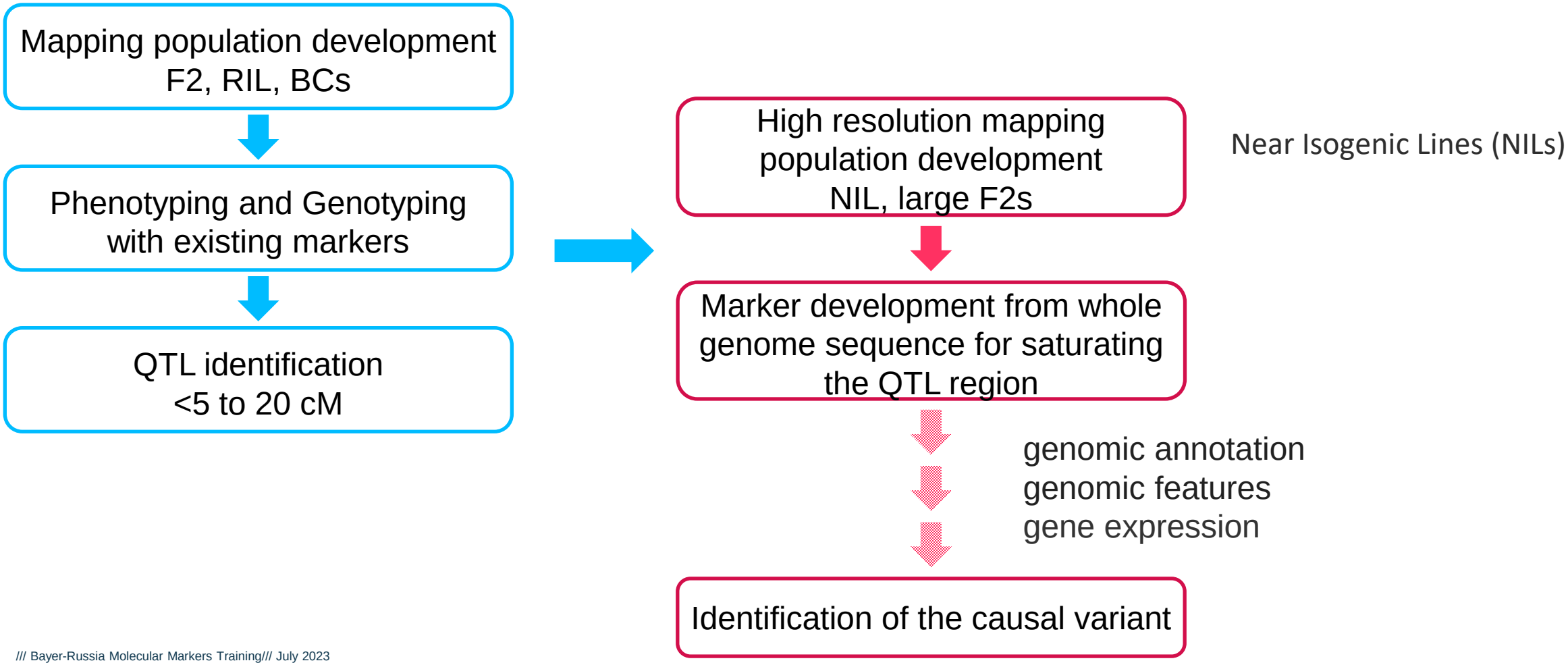
Allele Frequency

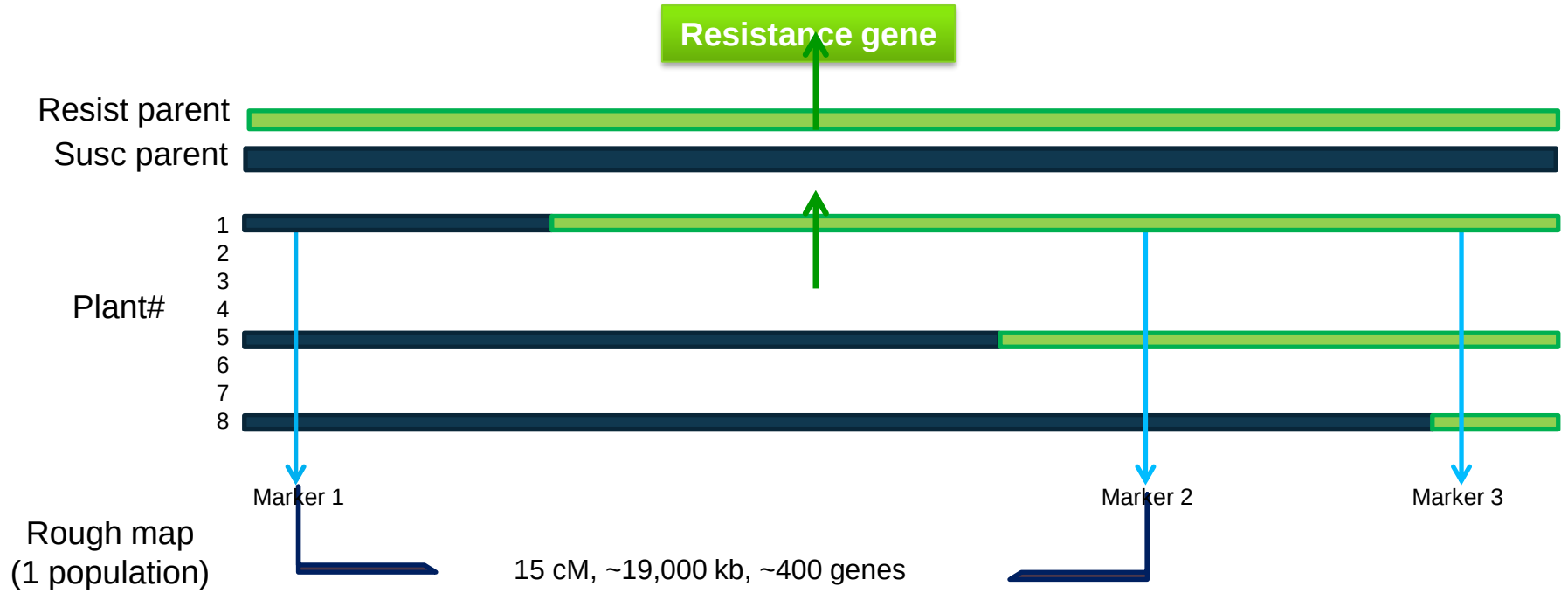


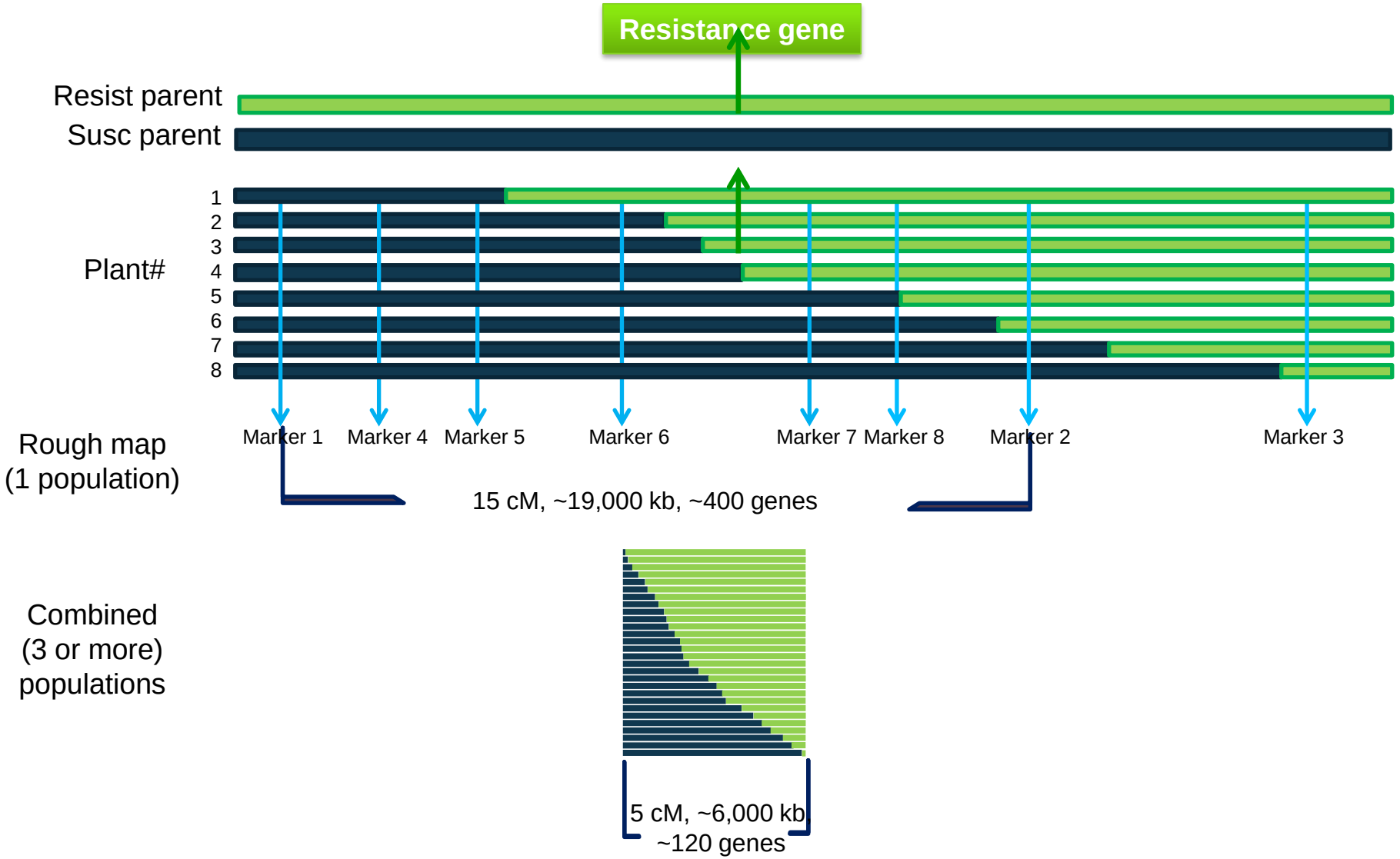
Adapted from Mackay *et al.*, 2009

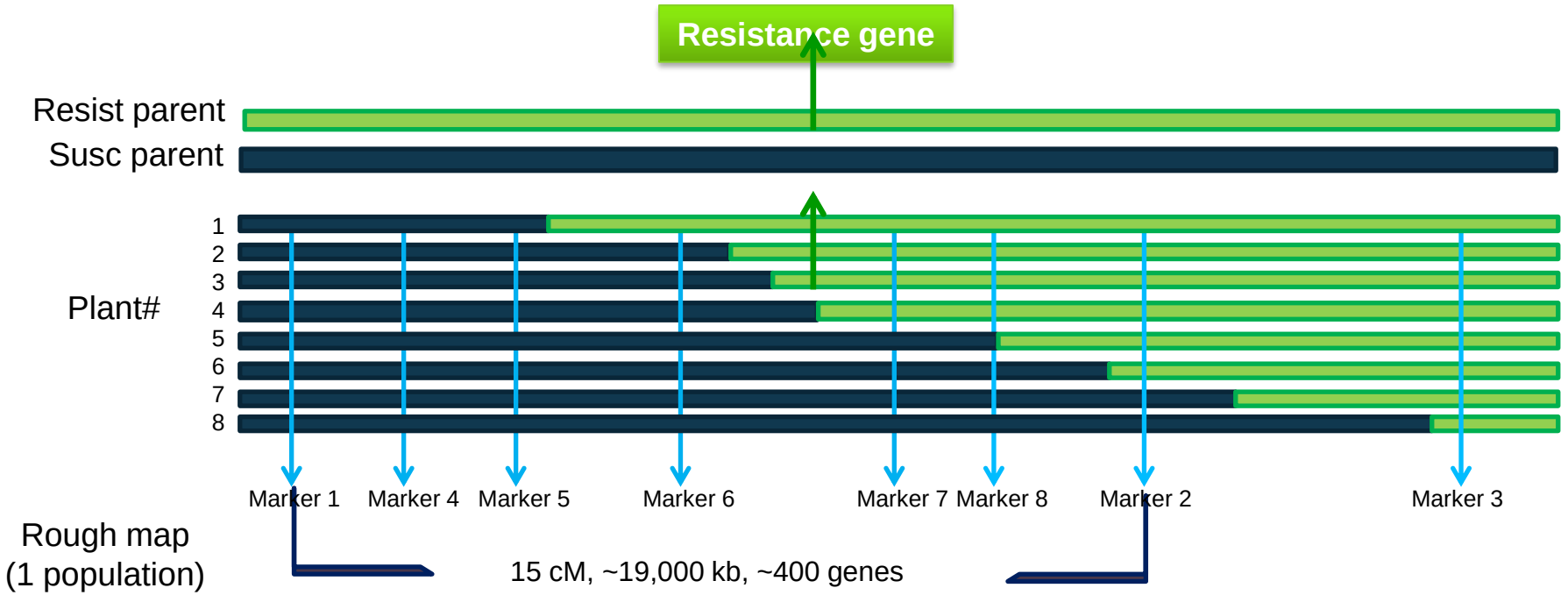
Fine-mapping: what are the causal SNPs?

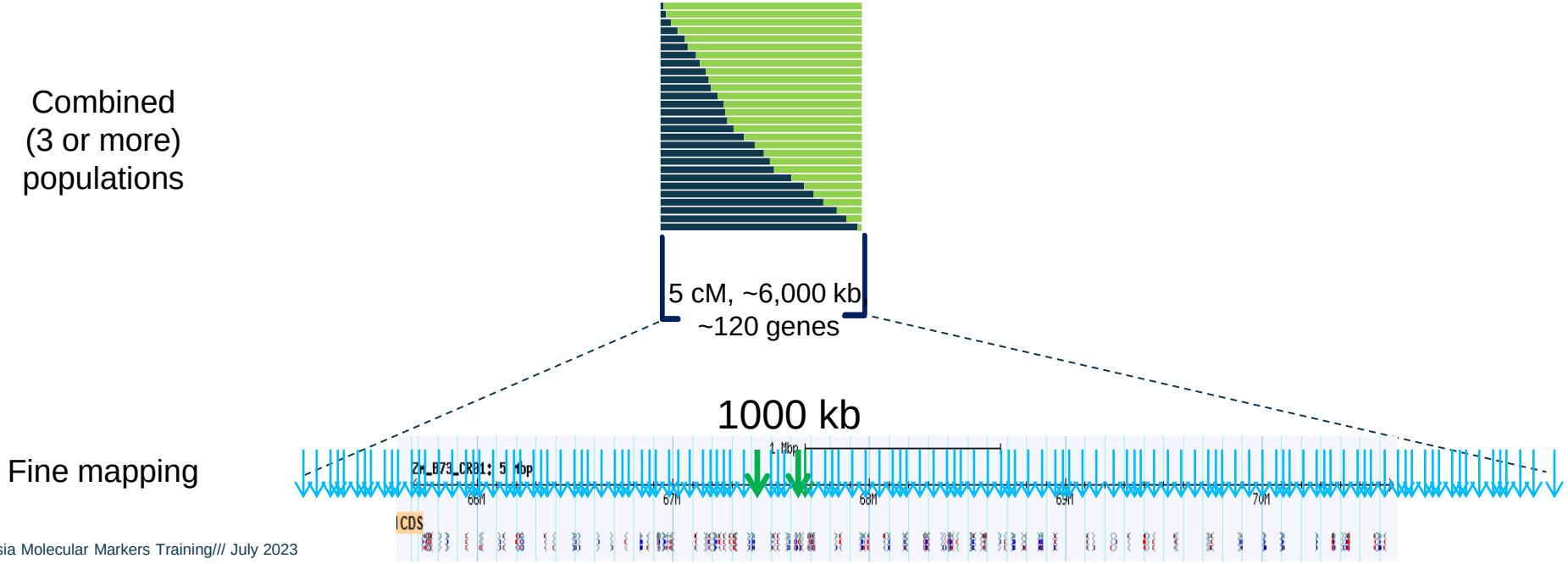
Fine mapping is the process by which the location of a QTL is reduced from an initial interval of 20 cM or more to an interval of a few cM or less.



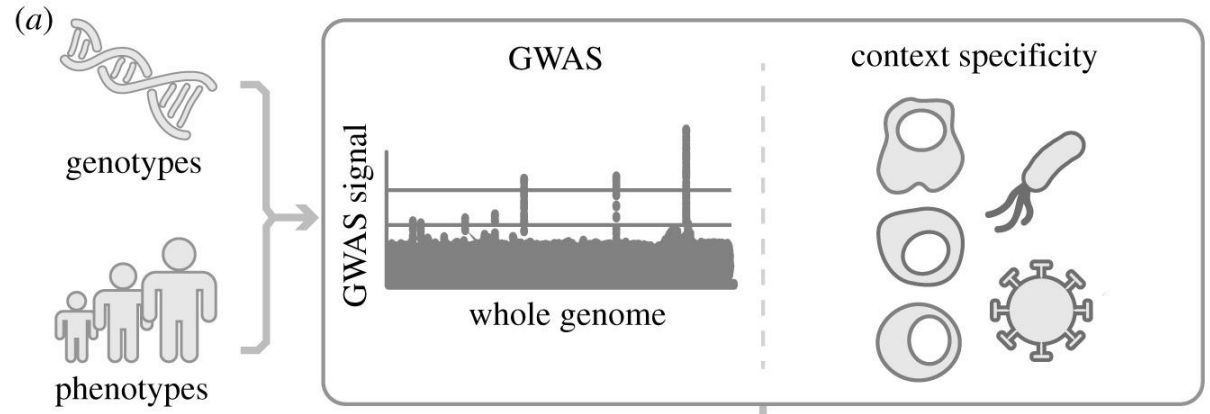

Resistance gene
Resist parent
Susc parent
Plant#
1
2
3
4
5
6
7
8
Marker 1
Marker 2
Marker 3
Rough map
(1 population)
15 cM, ~19,000 kb, ~400 genes


Resistance gene
Resist parent
Susc parent
Plant#
1
2
3
4
5
6
7
8
Marker 1
Marker 4
Marker 5
Marker 6
Marker 7
Marker 8
Marker 2
Marker 3
Rough map
(1 population)
15 cM, ~19,000 kb, ~400 genes
Combined
(3 or more)
populations
5 cM, ~6,000 kb
~120 genes

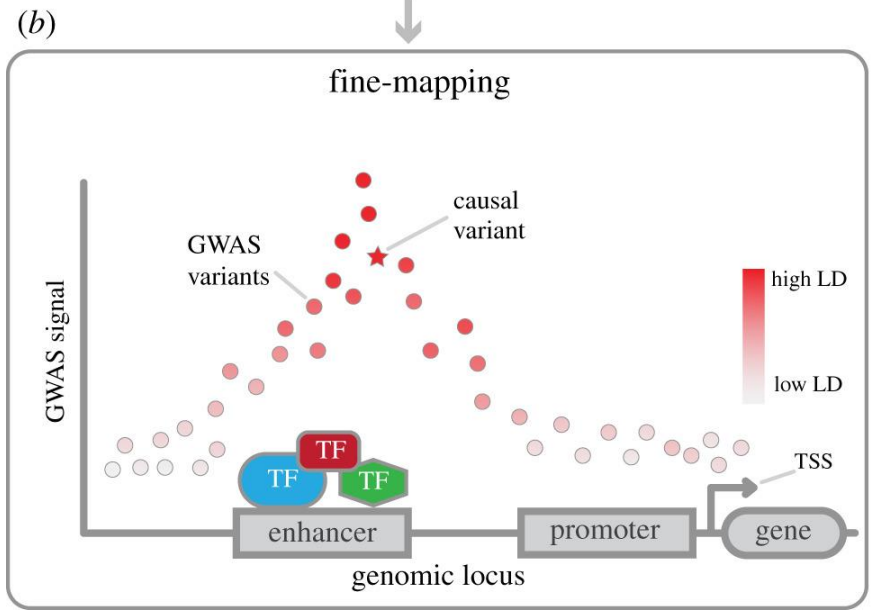

Resistance gene
Resist parent
Susc parent
Plant#
1
2
3
4
5
6
7
8
Marker 1
Marker 4
Marker 5
Marker 6
Marker 7
Marker 8
Marker 2
Marker 3
Rough map
(1 population)
15 cM, ~19,000 kb, ~400 genes

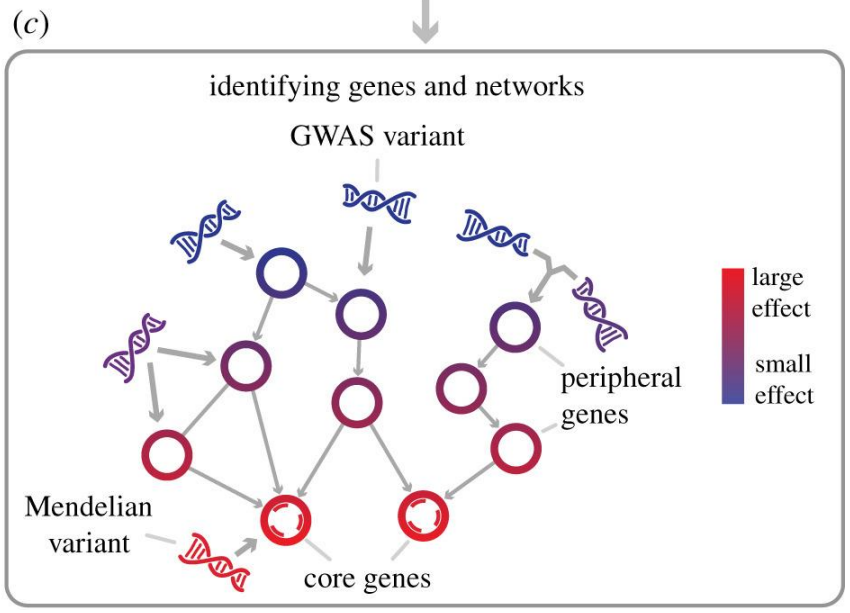

Combined
(3 or more)
populations
5 cM, ~6,000 kb
~120 genes
1000 kb
Fine mapping
Zm_B73_CR01: 5 Mbp
66M
67M
68M
69M
70M
CDS

Fine-mapping and gene prioritization approaches



functional /regulatory elements





https://royalsocietypublishing.org/doi/10.1098/rsob.190221

Why fine-mapping?

---Enable rapid, routine high-resolution mapping of traits for Breeding and Disease program.

1. Reduction of a QTL linkage interval to a **smaller region**

// fewer, higher predictive markers

// more precise selection models

// improve stacking

// minimize the interpretation of pleiotropy and close linkage

// Allele/region IP

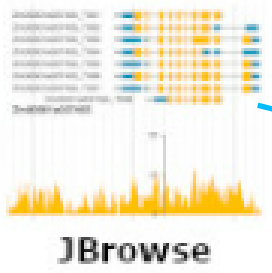
2. Identification of causal **genes**

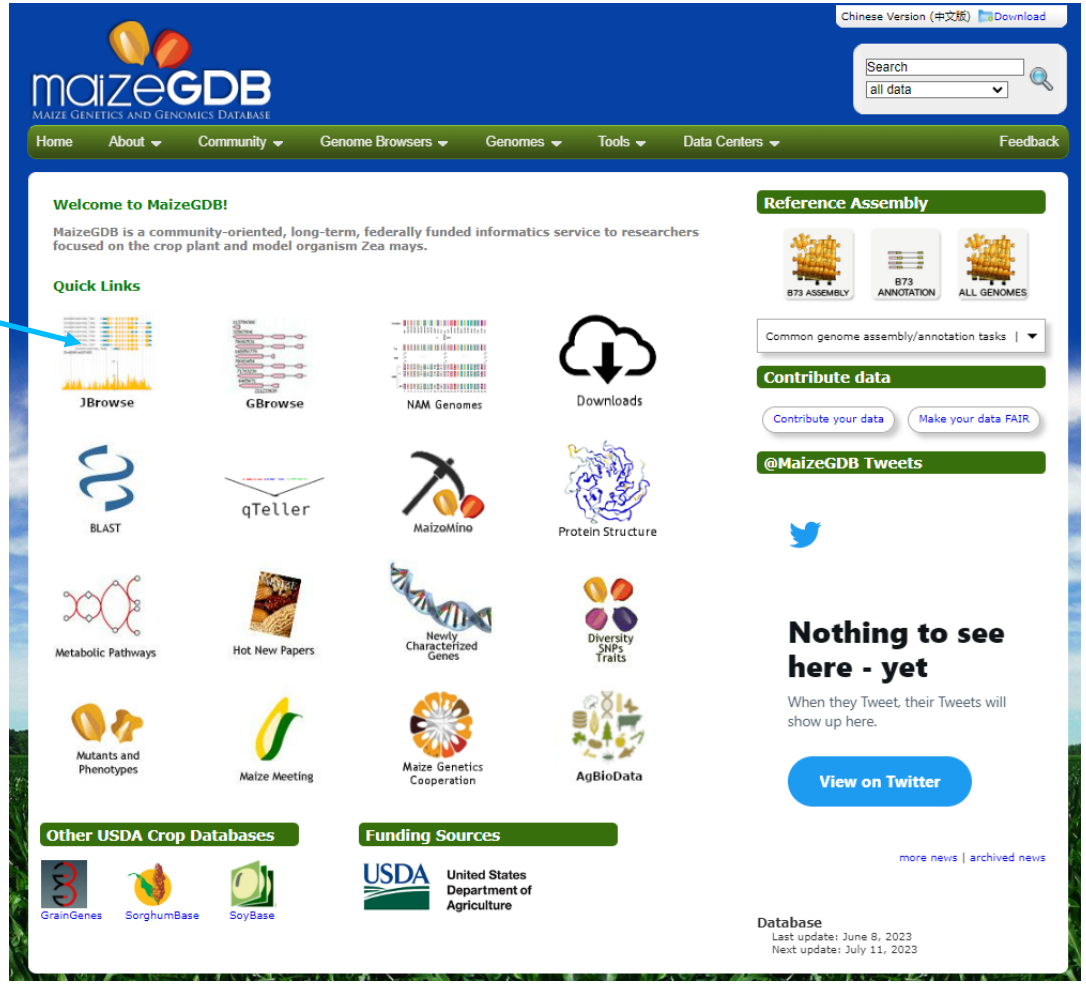
molecular characterization and functional studies of the relationship between gene and trait, and the wider genetic network

3. Pinpoint of the **causative genomic lesion**

Suggest experimental perturbation - editing

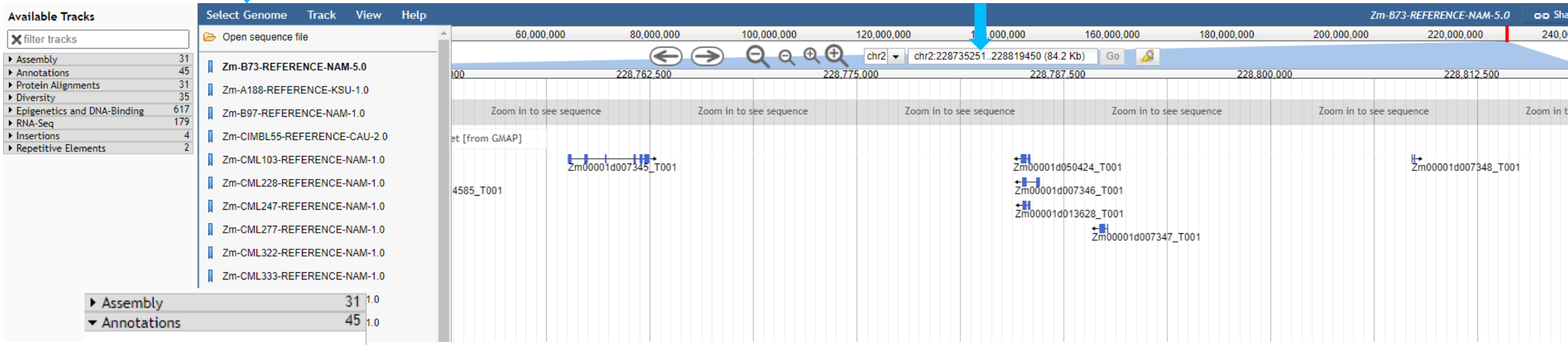
Bioinformatics tools—JBrowse

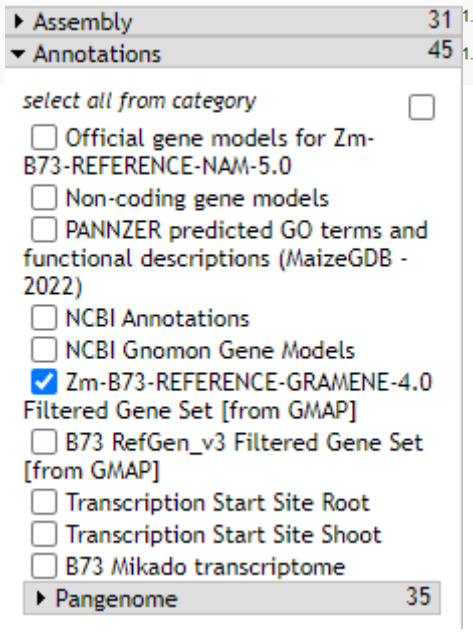




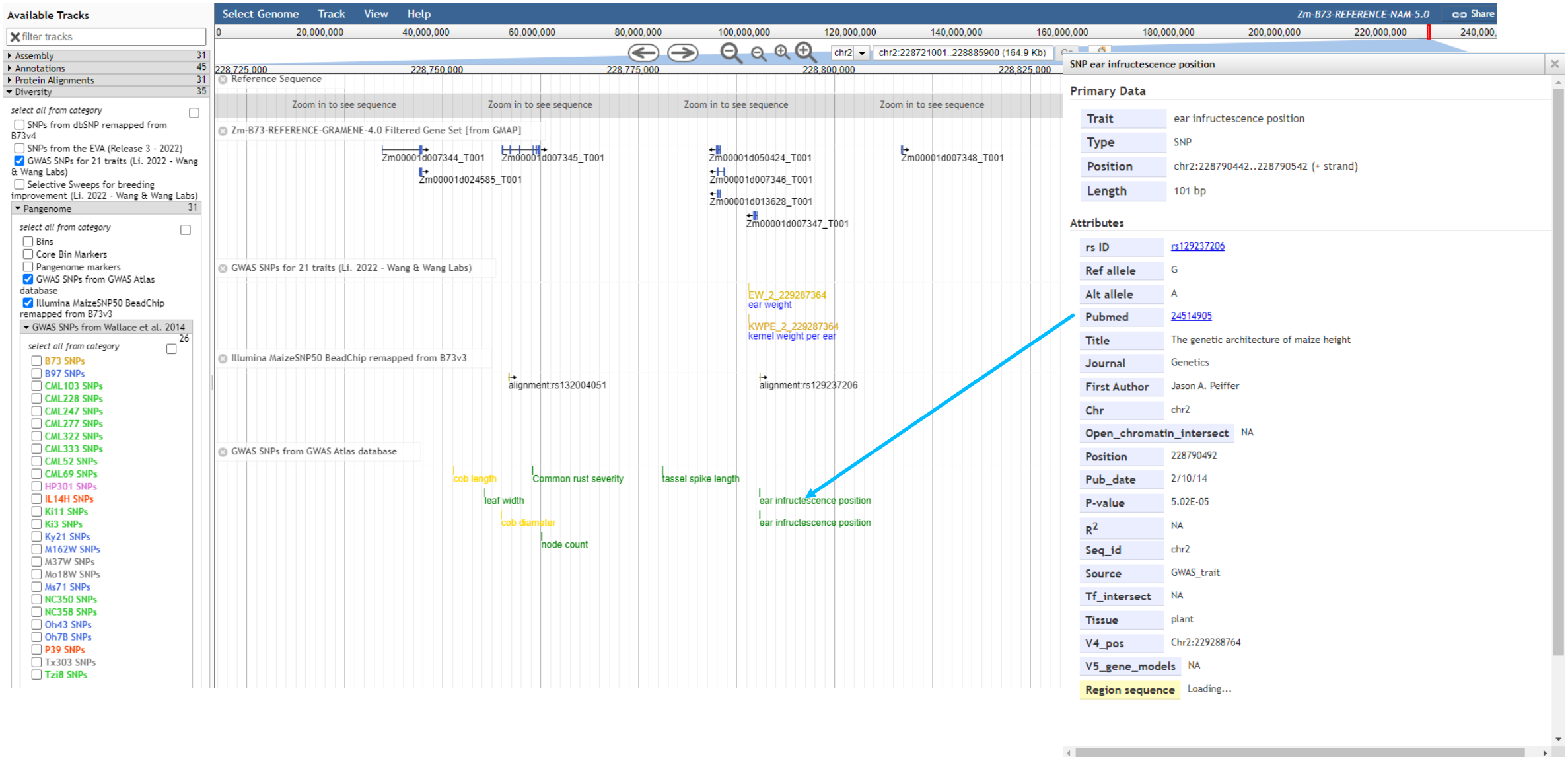
https://www.maizegdb.org/

Bioinformatics tools—JBrowse

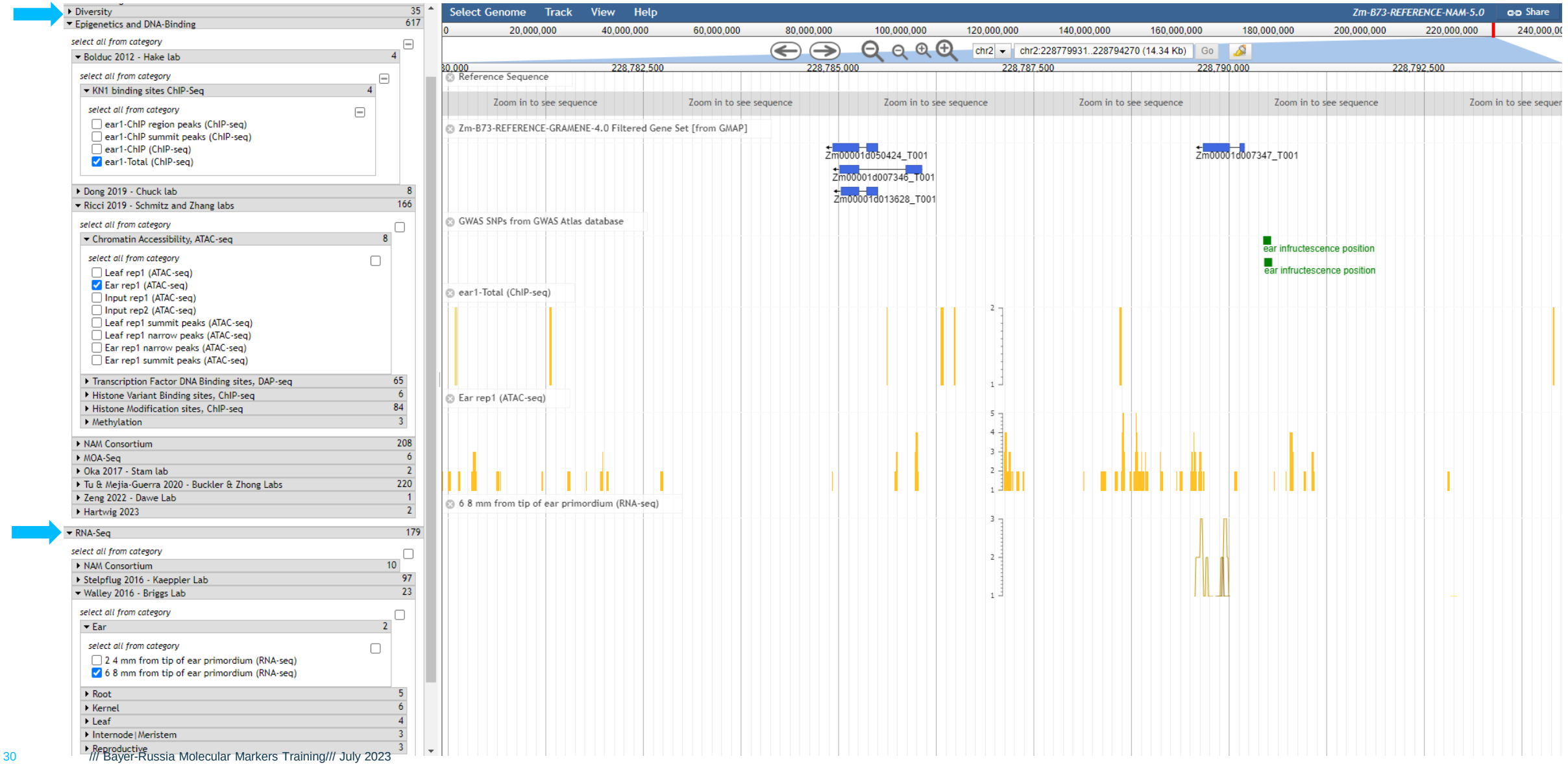




Bioinformatics tools—JBrowse



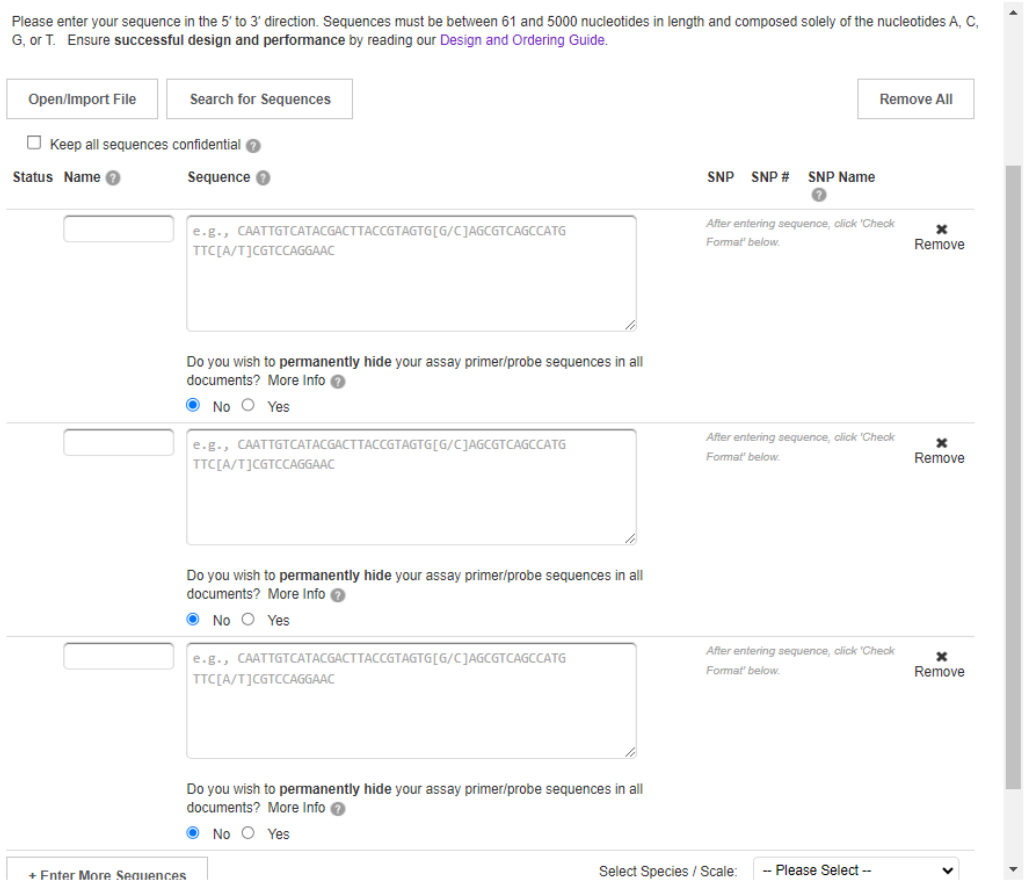
Bioinformatics tools—JBrowse



Bioinformatics tools—TaqMan Marker design

The Custom TaqMan® Assays Design Pipeline

https://www.thermofisher.com/order/custom-genomic-products/tools/genotyping/



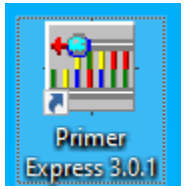
- At least one target site
- Target sites that are more than 40 bases away from the 5´ and 3´ ends.
- Target sites that are more than two bases away from any Ns.

AGTGAACGCGATA[G/A]GCANCTCCTGCCC

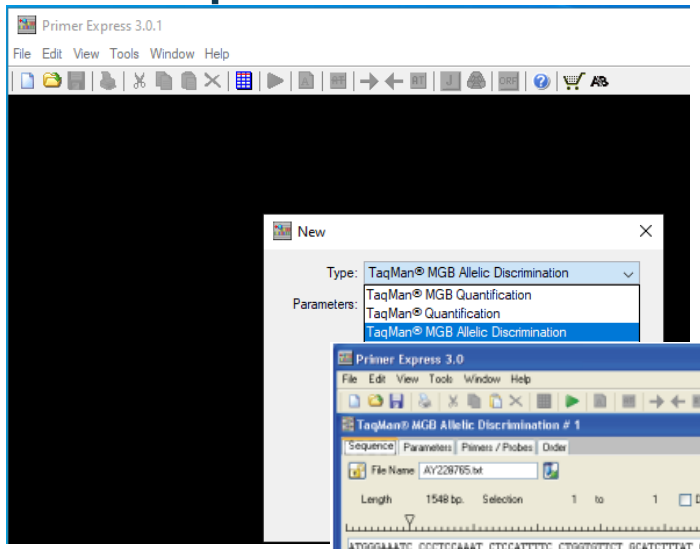
If this is your target site, verify that no Ns are within two bases.

- Target sites that are more than two bases away from any other SNPs.
- At least one specifc SNP for assay design (and mask all the remaining nontarget SNPs with Ns).

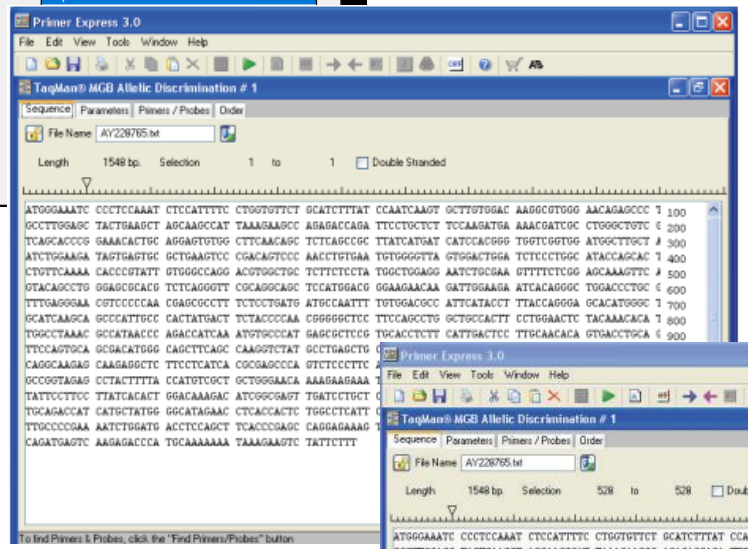
Bioinformatics tools—TaqMan Marker design



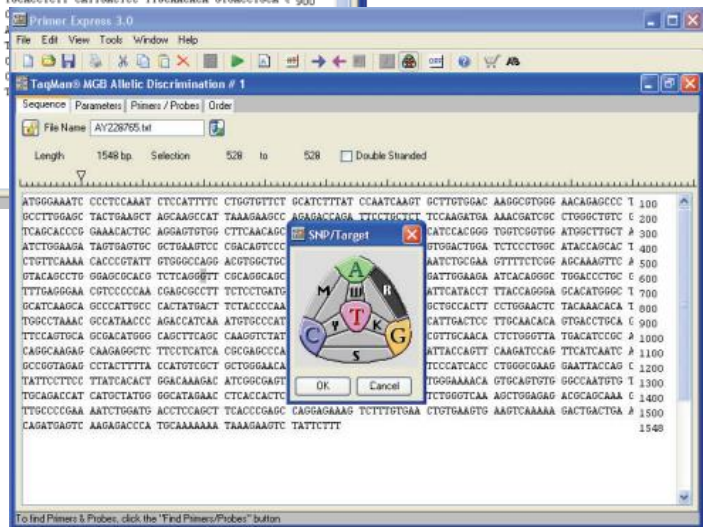
1. Select **File** > **New** to open the New dialog box



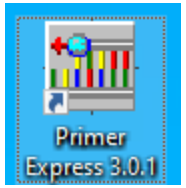
2. Select **Tools** > **Add DNA File**.
Note you can also copy and paste or type your sequence file in the Sequence tab



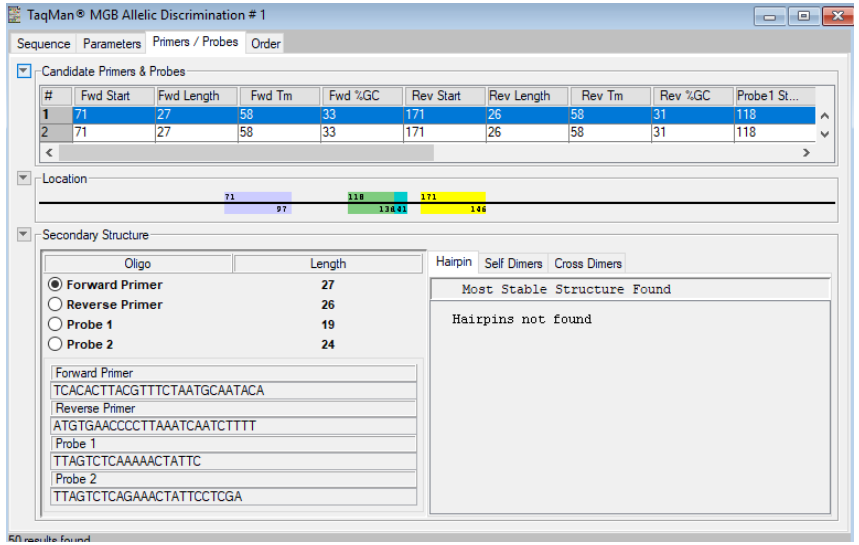
3. Highlight the SNP target site.
Edit > **Annotate** > **SNP Target**

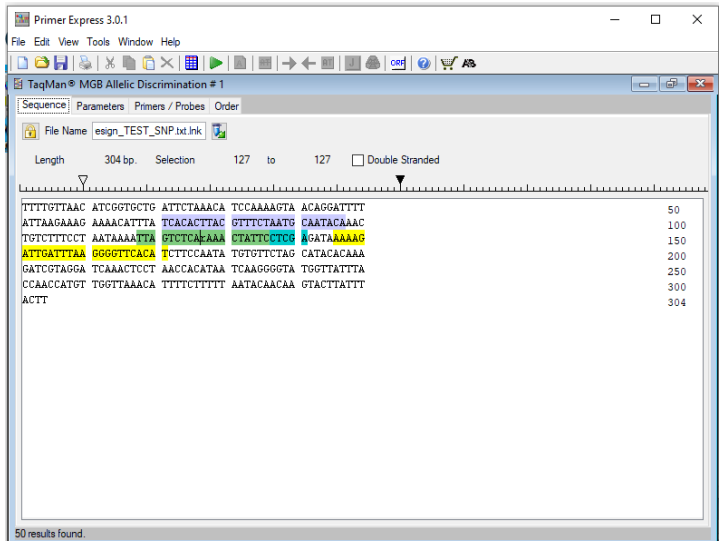


Bioinformatics tools—TaqMan Marker design

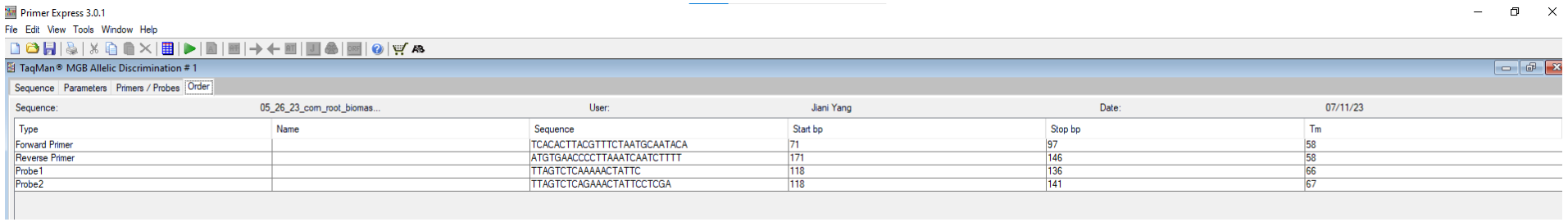


4. Select **Tools** > **Find Primers/Probes**





5. In the **Primer/Probe tab**, select the primer and probe set you want to order.
Click on the **Order tab** and Click **AB store** on the toolbar to go the online store



Bioinformatics tools—R script

R is an open-source programming language for statistical analysis and graphing

Many resources are available ranging from fully developed code for certain applications to online communities that help troubleshoot code.

JOURNAL ARTICLE

AlphaSimR: an R package for breeding program simulations

R Chris Gaynor ✉, Gregor Gorjanc, John M Hickey

G3 Genes|Genomes|Genetics, Volume 11, Issue 2, February 2021, jkaa017,
https://doi.org/10.1093/g3journal/jkaa017
Published: 07 December 2020 **Article history** ▾

Zhang *et al. BMC Bioinformatics* (2023) 24:199
https://doi.org/10.1186/s12859-023-05318-9

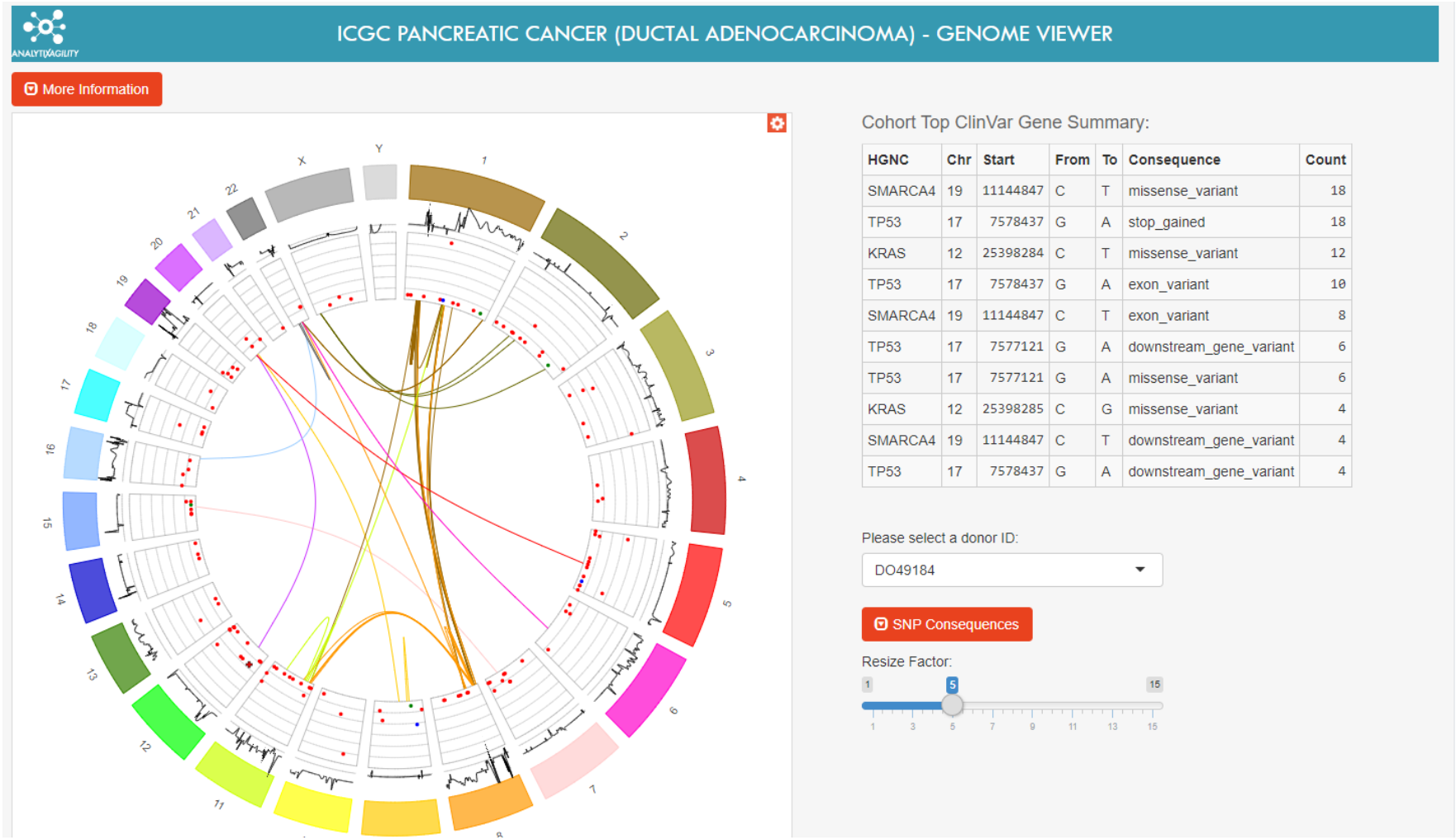
BMC Bioinformatics

SOFTWARE Open Access

geneHapR: an R package for gene haplotypic statistics and visualization

Renliang Zhang[1], Guanqing Jia[1*] and Xianmin Diao[1*]

Bioinformatics tools—Shiny for R: Genome Browser



HGNC	Chr	Start	From	To	Consequence	Count
SMARCA4	19	11144847	C	T	missense_variant	18
TP53	17	7578437	G	A	stop_gained	18
KRAS	12	25398284	C	T	missense_variant	12
TP53	17	7578437	G	A	exon_variant	10
SMARCA4	19	11144847	C	T	exon_variant	8
TP53	17	7577121	G	A	downstream_gene_variant	6
TP53	17	7577121	G	A	missense_variant	6
KRAS	12	25398285	C	G	missense_variant	4
SMARCA4	19	11144847	C	T	downstream_gene_variant	4
TP53	17	7578437	G	A	downstream_gene_variant	4

Shiny is an open source R package for building web applications using R

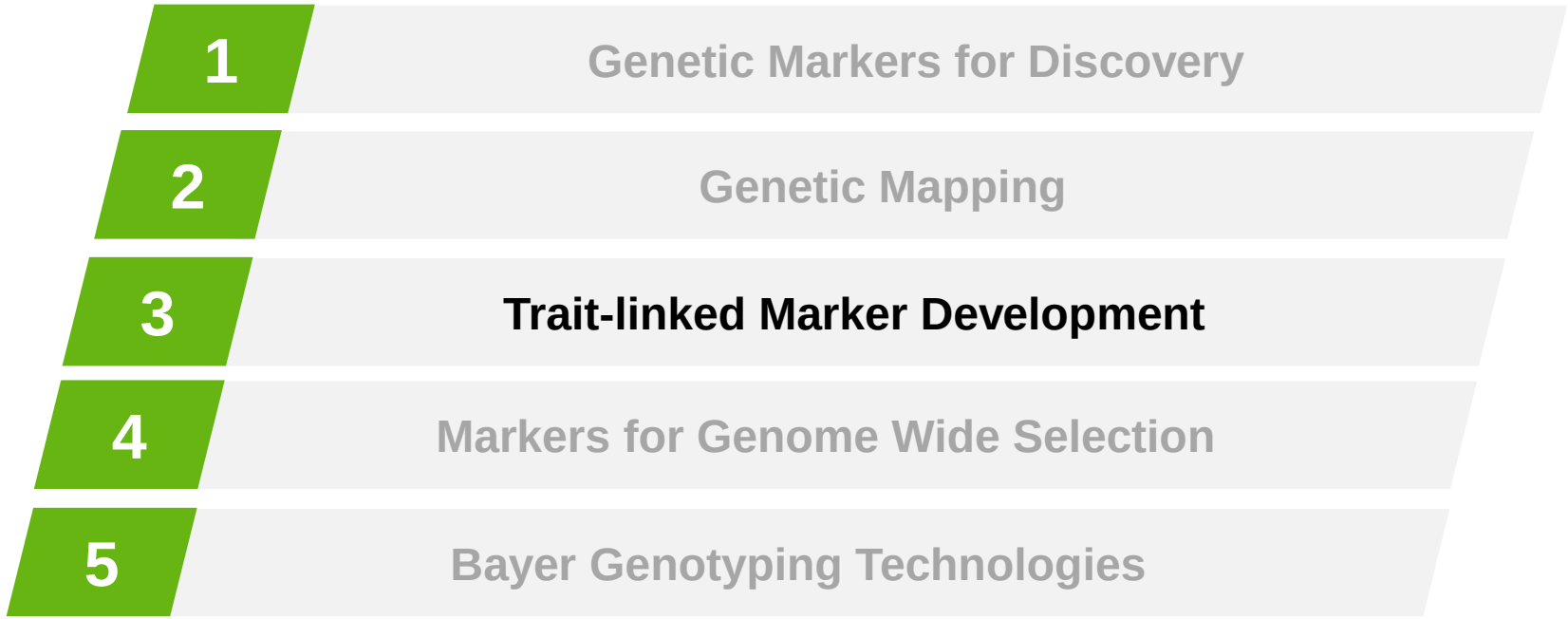
For example shown here, visualization is based on Circos, a way of visualizing whole genomes.

.

https://shiny.posit.co/r/gallery/life-sciences/genome-browser/

Q&A
Discussion

Bayer Russia Molecular Marker Training :
Genetic Markers and Technologies



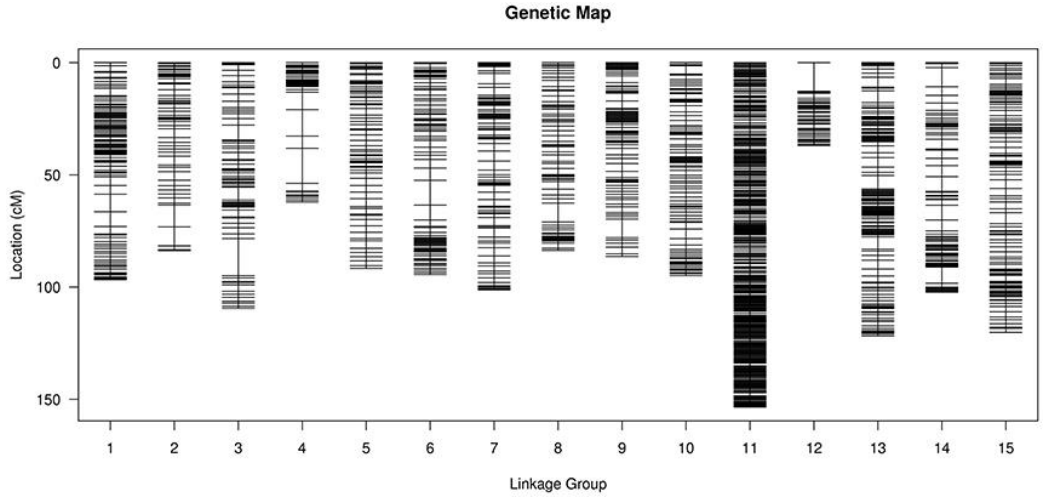
Two types of genetic markers

Regular markers

// Spanned all over the genome

// Polymorphism with no associated trait

// Can be used to build genetic maps



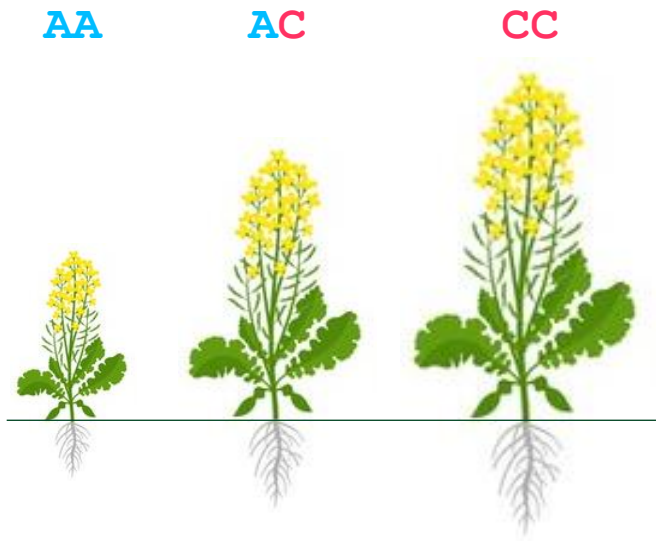
https://doi.org/10.3389/fpls.2018.00885

Trait markers

// Located in specific genomic areas

// Linked to the variation of a given trait

// Can be used to follow said trait (MAS)



Technology we’re leveraging to develop desired traits

Discovery Phase:
Genome Editing Approach to Short Stature Corn

Breeding | Edited | Wild Type

Plant Breeding

Desired gene

Plant with A Trait + Plant with B Trait

Many genes transferred

Plant C (with traits A + B)

Biotechnology

Desired gene

Source Genome + Target Plant Genome

Only selected gene transferred

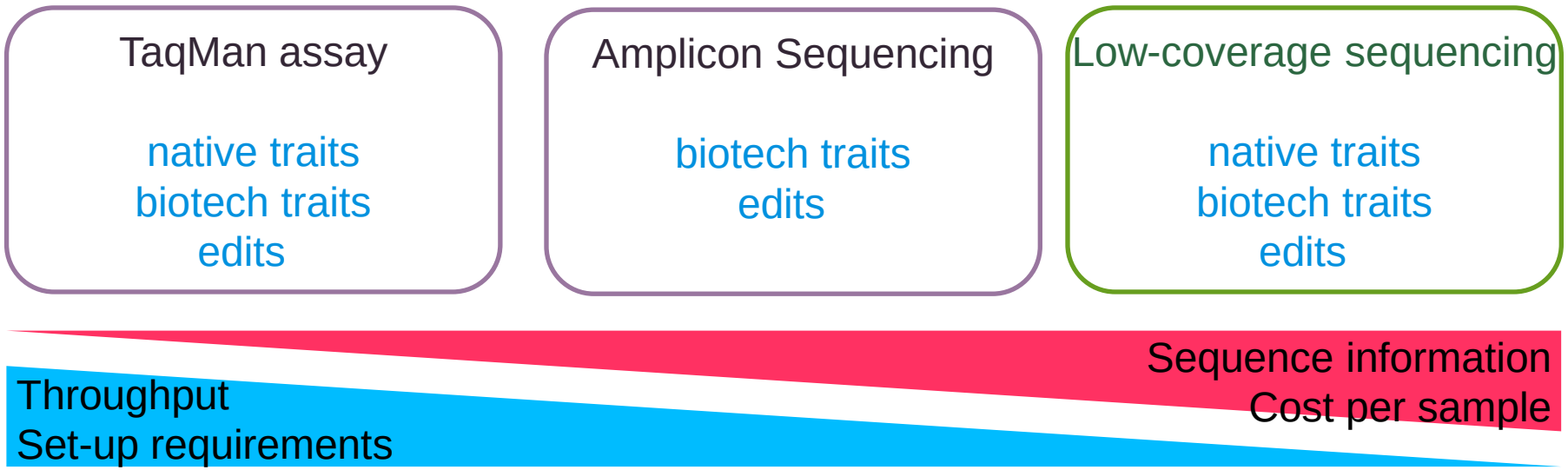
Modified Plant Genome

Gene Editing

Deactivate an unfavorable characteristic

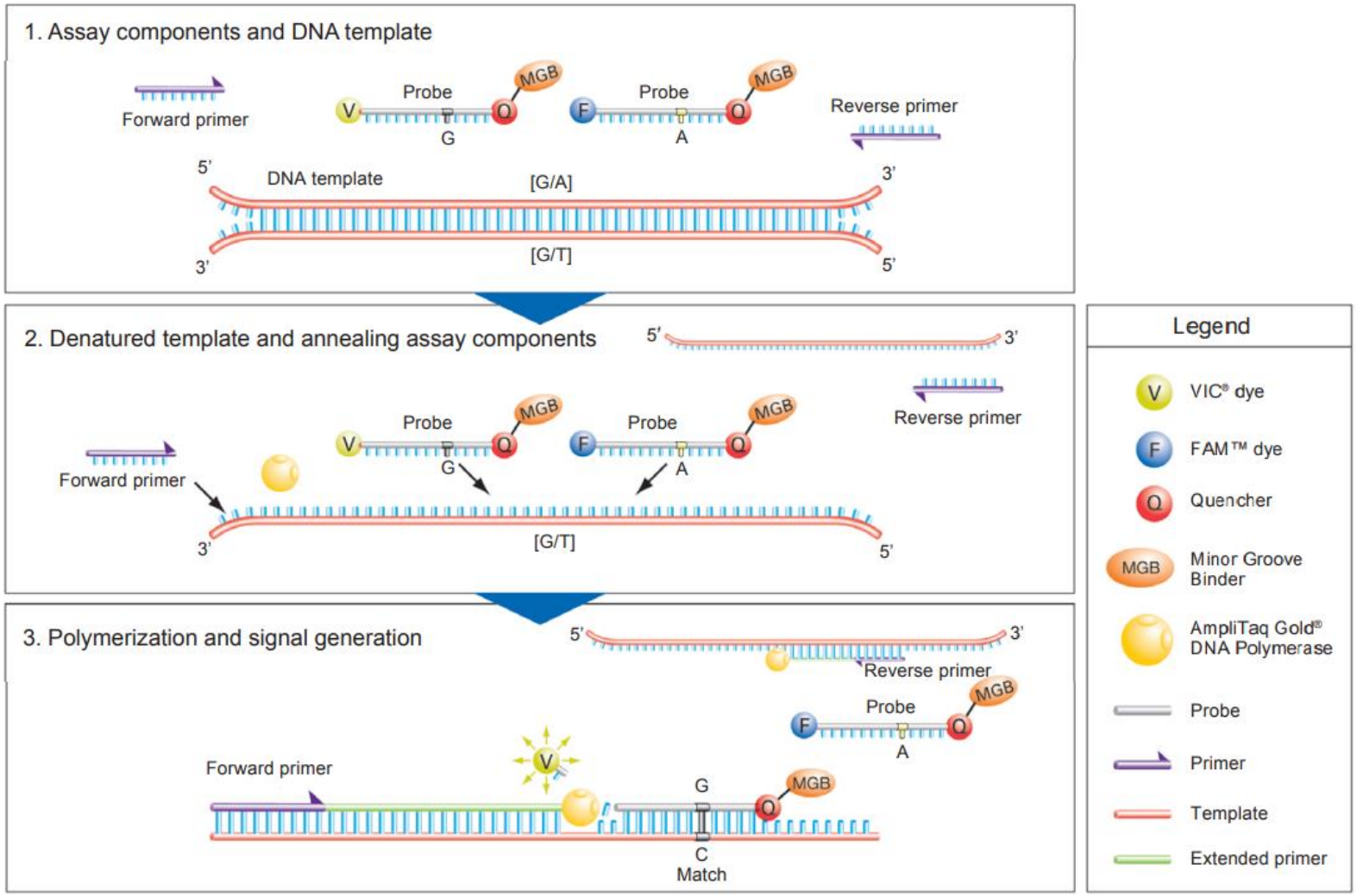
Enable a beneficial characteristic

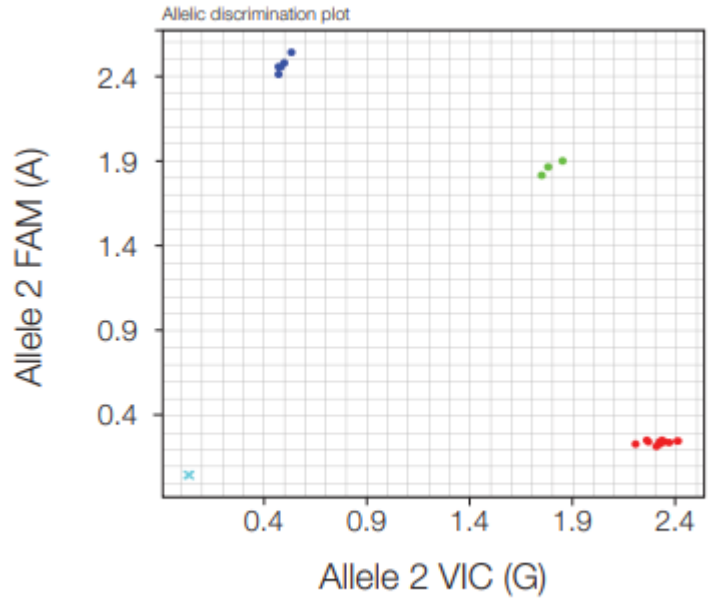
DNA assays are essential to crop science pipelines



Causal variant discovery for native traits

---TaqMan SNP genotyping assay





- co-dominant
- good for mapping specific genes
- Probe is expensive

https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_040597.pdf

Causal variant discovery for native traits

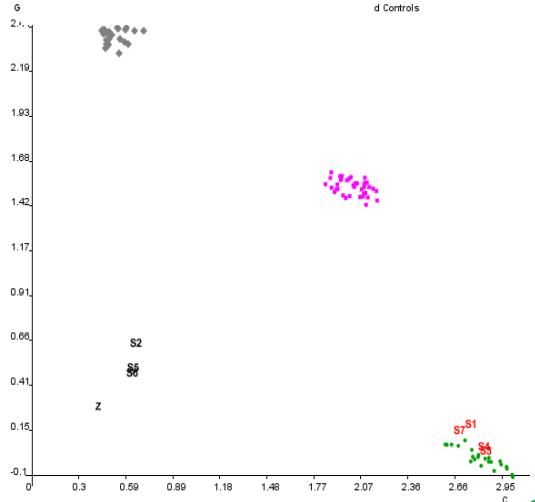
TaqMan Assay Validation: Score Plot Analysis

1. Assay validation using historical data
2. Assay validation using new population:

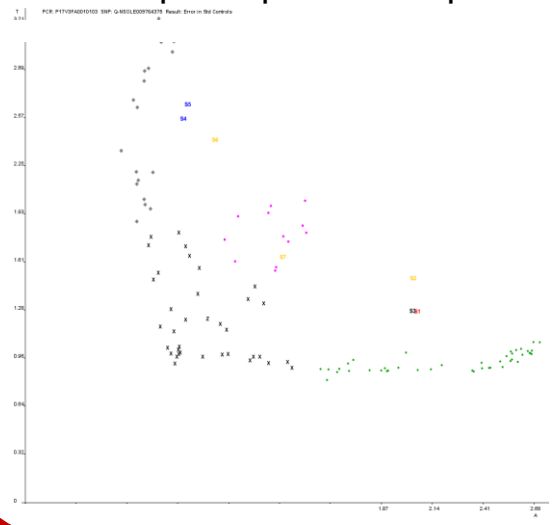
    F2 or any segregating population

    Or, it can be a mixture of the parental genotypes (two homozygous classes) and an F1 (heterozygote) in a 1:2:1 ratio

Example of good score plot



Example of poor score plot



Causal variant discovery for native traits

Genetic Accuracy

1. QTL effect requirements

- QTL identified via **linkage analysis**
  - Effect must be confirmed in two other populations (three pops total)
  - Minimum requirement for percent variation explained: R^2>15%
  - QTL window < 5cM
- QTL identified via **association analysis**
  - Confirm QTL location and efficacy in at least one bi-parental population (F2 or BCX) or through haplotype mining
    - Minimum requirement for percent variation explained in bi-parental population: R^2>15%
    - Haplotype mining: Pedigree and genotypic analysis via Trait Analysis Team of ancestral donor and recoveries combined with phenotypic data
- QTL or gene identified through **public literature**
  - Confirm in bi-parental or haplotype mining/Association analysis within breeding germplasm
  - Minimum requirement for percent variation explained
  - QTL window < 5cM

Causal variant discovery for native traits

Genetic Accuracy

2. Genetic accuracy panel requirements

Genetic Accuracy panel should constitute of the germplasm pool into which the trait will be deployed.

The germplasm or list of lines that constitute the deployment pool should be determined based on

- Pipeline materials (List of lines/inbreds that constitute the breeding pipeline for the crop)
- Heterotic pool (Male or Female)
- Geography
- Maturity
- Other

*Always remember to include the **trait donors** in the deployment pool.

Causal variant discovery for native traits

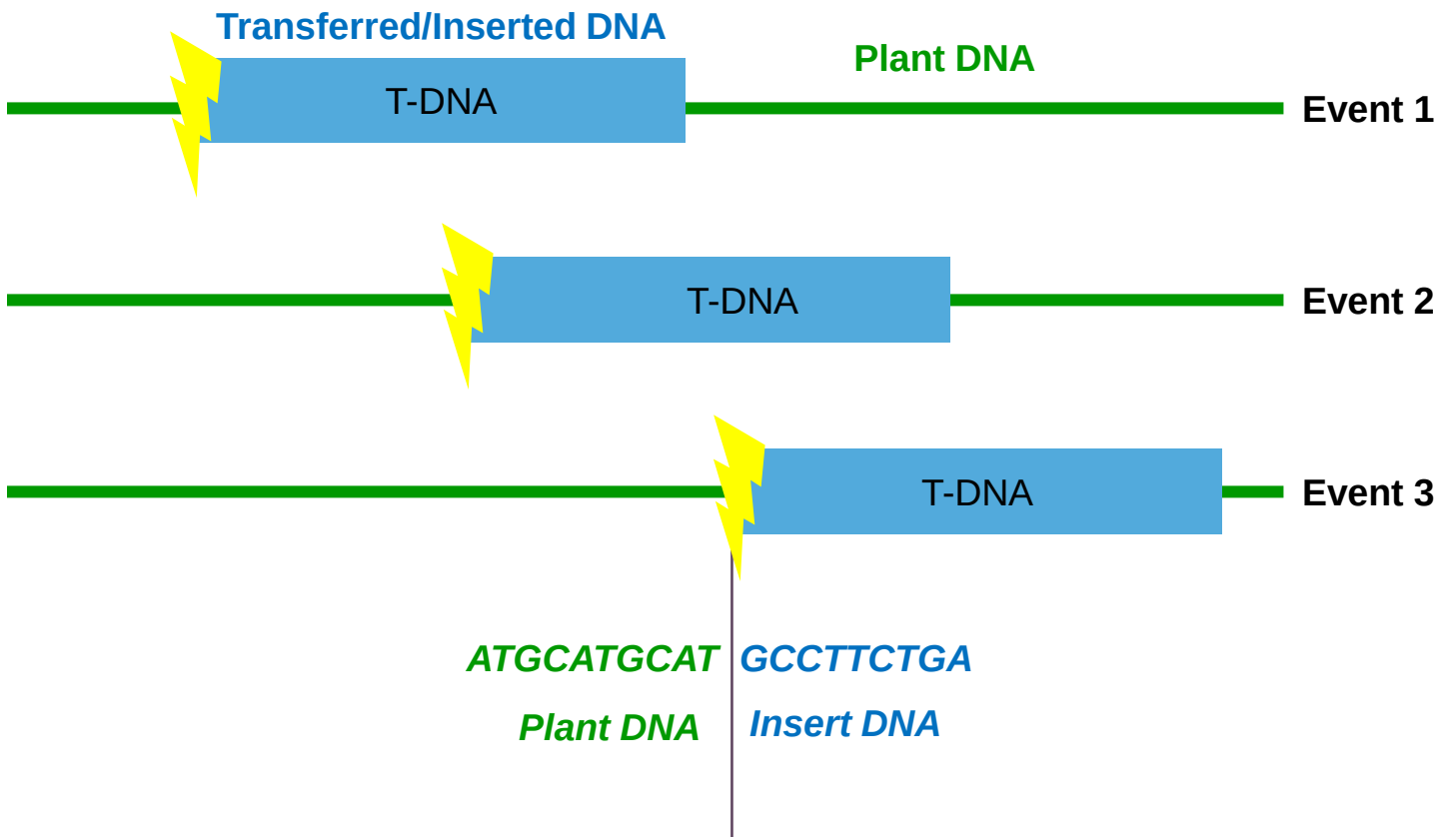
Genetic Accuracy

3. SNP accuracy requirements

- SNP for trait association must be within 5cM of QTL in mapping analysis.
- False positive rate <5 %: Allele of SNP associated with favorable phenotype should be found in less than 5 % of unfavorable lines
- False negative rate <10%: Allele of SNP associated with unfavorable phenotype found in less than 10 % of favorable phenotype lines

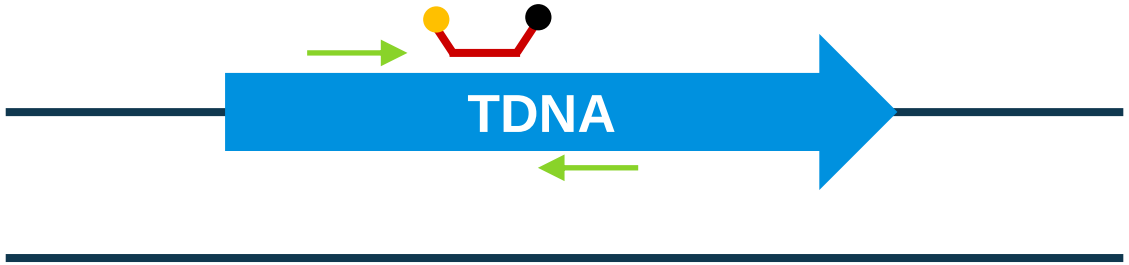
Genomic characterization for biotechnology traits

A Transformation Event is Defined by T-DNA/Plant DNA Junction



Genetic element-based assays and event junction assays can both be used to assay for transgenic events

Genetic element-based assay

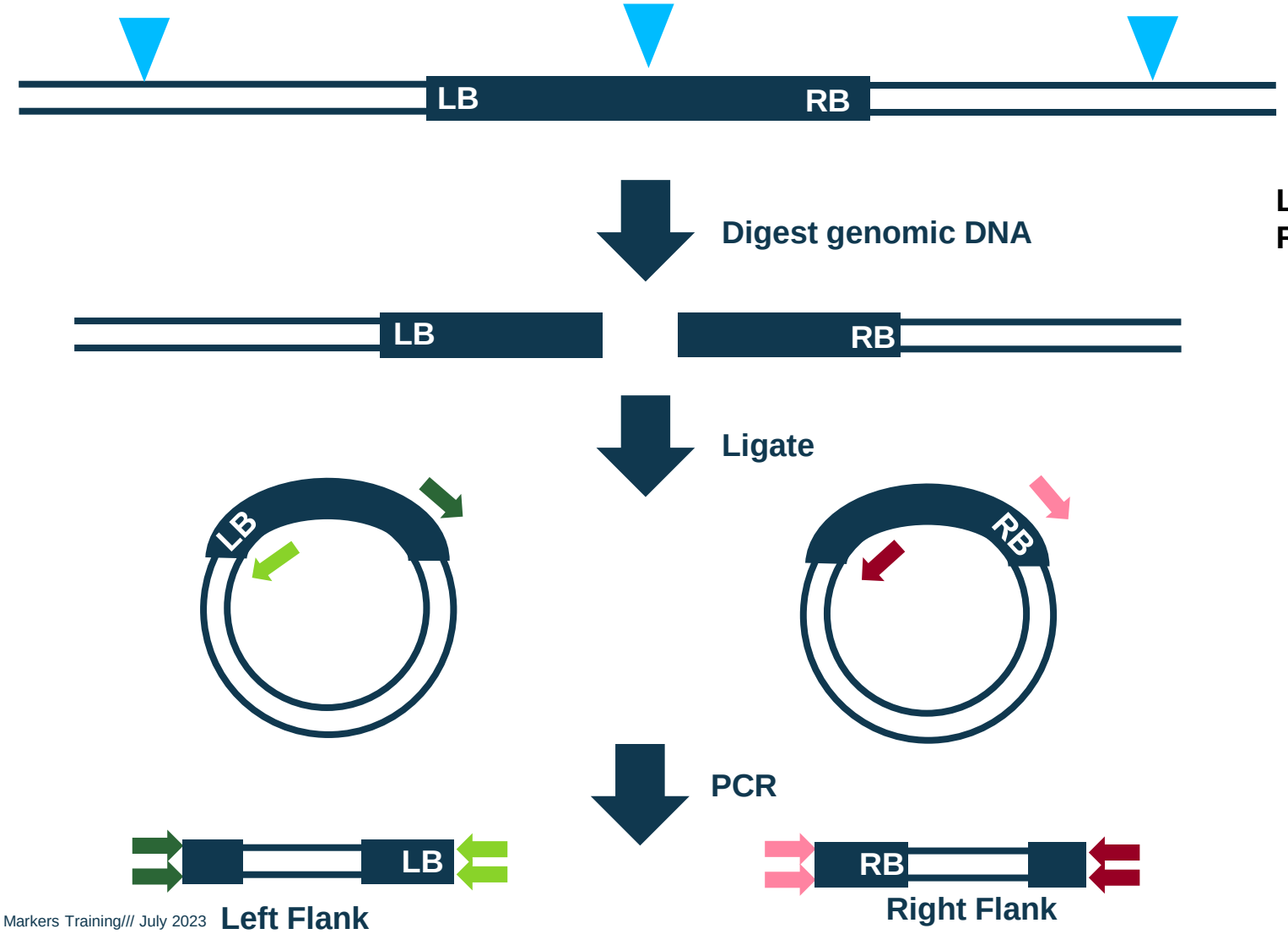


- Detects all events from construct with the element sequence
- Can detect events from different constructs if constructs contains the same element sequence
- Used to determine copy number of TDNA
- Does not require genomic flank DNA sequence

Event-specific junction assay

- Typically designed to detect one TDNA insertion
- Often used when the same event will be used or grown many times (as in commercial traits)
- Will not detect multiple events or events from a different construct
- Can be designed as RT-PCR assay or endpoint TaqMan® assay
- Requires isolation of genomic flank DNA for assay development

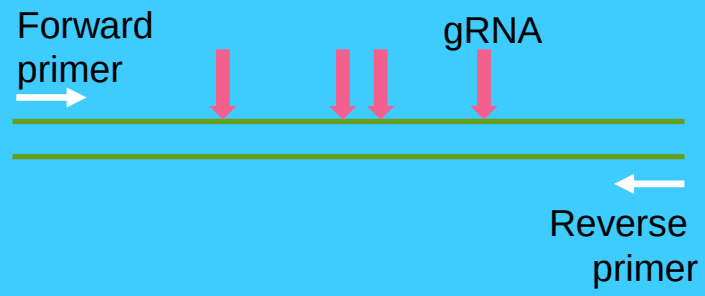
Inverse PCR can be used to obtain DNA sequence of genomic flank of transgene inserts to develop event-specific assays



LB = Left border TDNA element
RB = Right border TDNA element

Causal variant discovery for gene editing

PCR based method



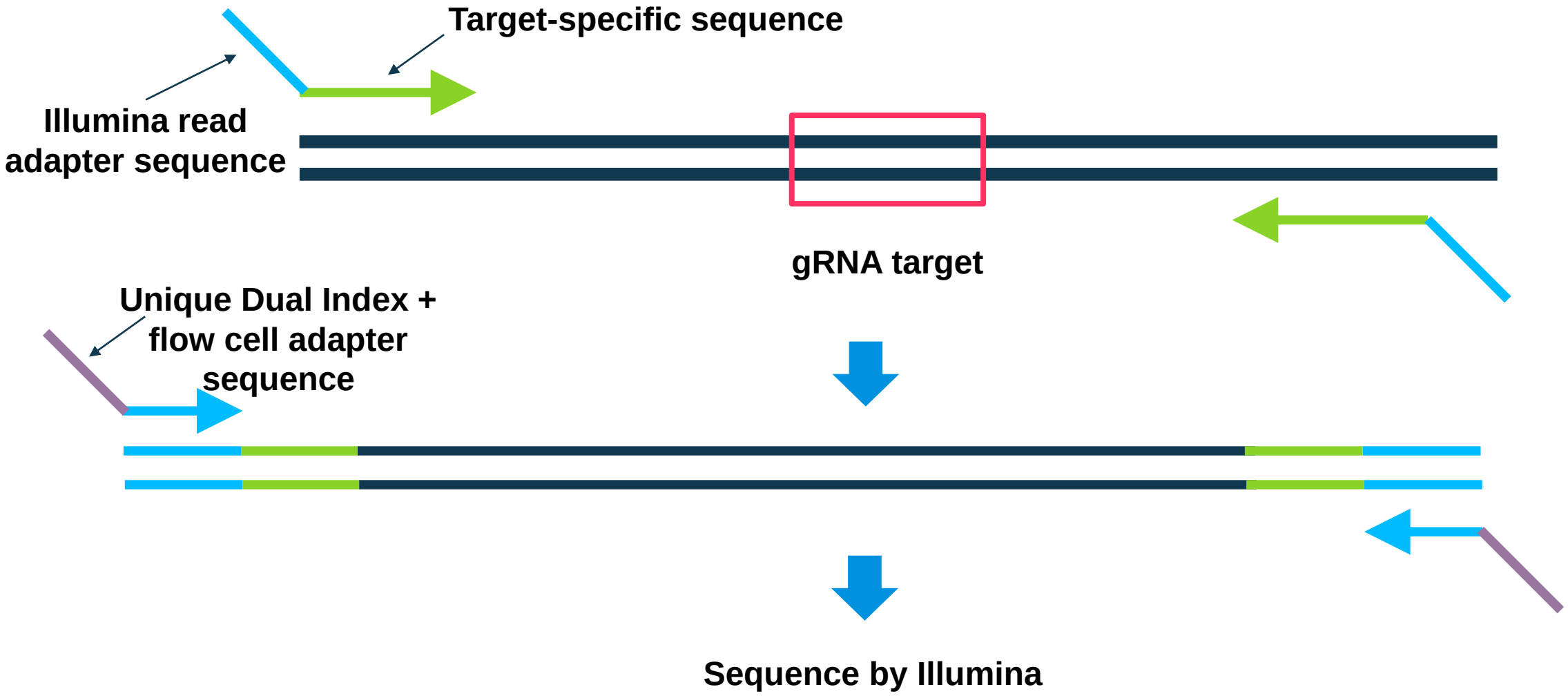
Presence of gene edits
Sequence confirmation of edits

Keep amplicon size within ideal PCR requirements (~300-600bp)
Have gRNA positioned <150bp from the gene specific primers, forward or reverse primers.


```
CGACCTCGGCATCGCGCCGCTCATCCGCGCCGACGAGGCGGGCACCGCGCGCGCCTCCGCCGAC WT
CGACCTCGGCATCGCGCCGCTCATCCGCGCC- ACGAGGCGGGCACCGCGCGCGCCTCCGCCGAC gn1a-1
CGACCTCGGCATCGCGCCGCTCATCCGCGC-- ACGAGGCGGGCACCGCGCGCGCCTCCGCCGAC gn1a -2
CGACCTCGGCATCGCGCCGCTCATCCGCGC-- - CGAGGCGGGCACCGCGCGCGCCTCCGCCGAC gn1a -3
CGACCTCGGCATCGCGCCGCTCA---------- CGAGGCGGGCACCGCGCGCGCCTCCGCCGAC gn1a-10
---------- ----------AC gn1a-112
CGACCTCGGCATCGCGCCGCTCATCCGCGCCG---------- gn1a-115
```

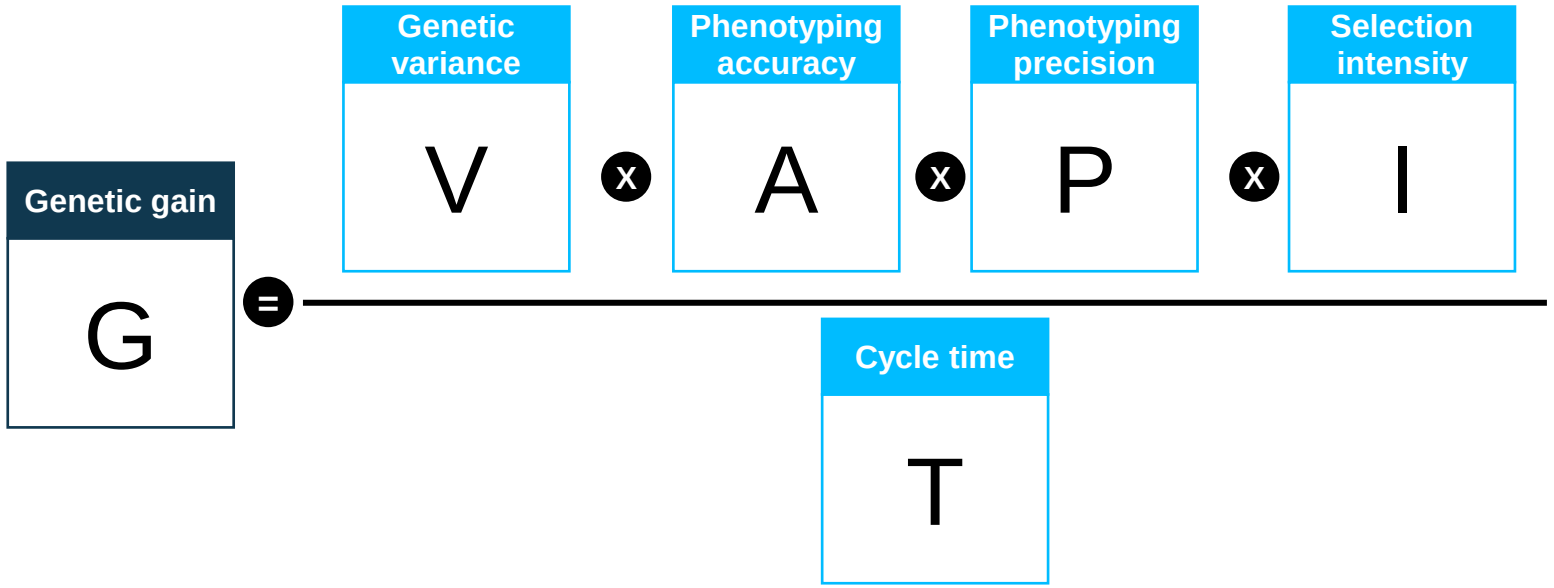
Li et al., 2016

Amplicon sequencing for gene edits involves a two-step PCR method



The breeder’s equation:

Molecular breeding has positive impact to V and T



Breeders have been successful whenever they had access to useful **genetic variation** and selection has focused on the **right traits** measured with the **right protocol** in the **right environment**

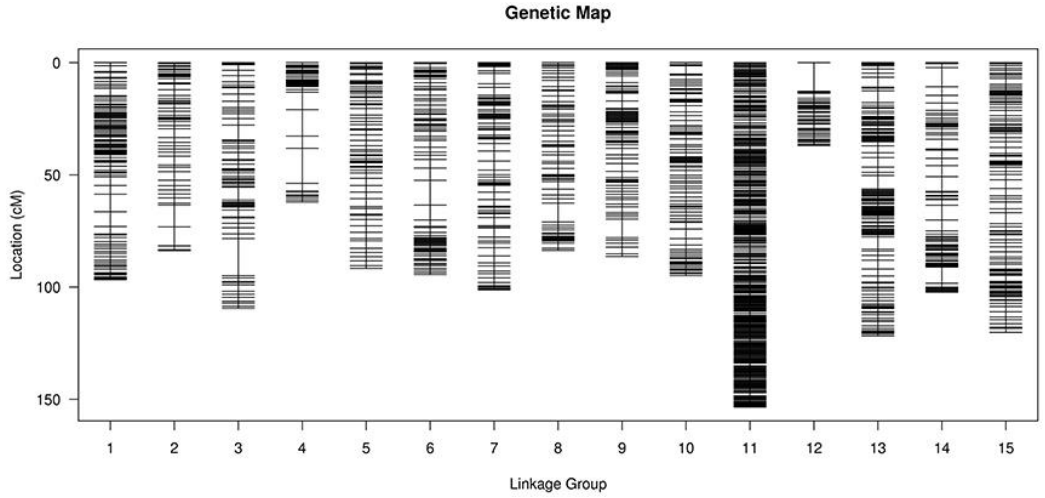
Two types of molecular markers

Genome-wide markers

// Span all over the genome

// Polymorphic with no associated trait

// Can be used to build genetic maps



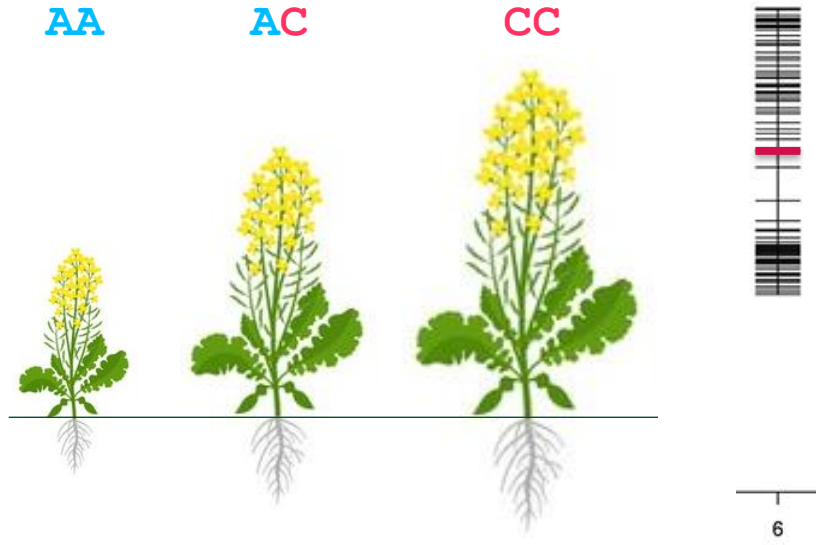
https://doi.org/10.3389/fpls.2018.00885

Trait markers

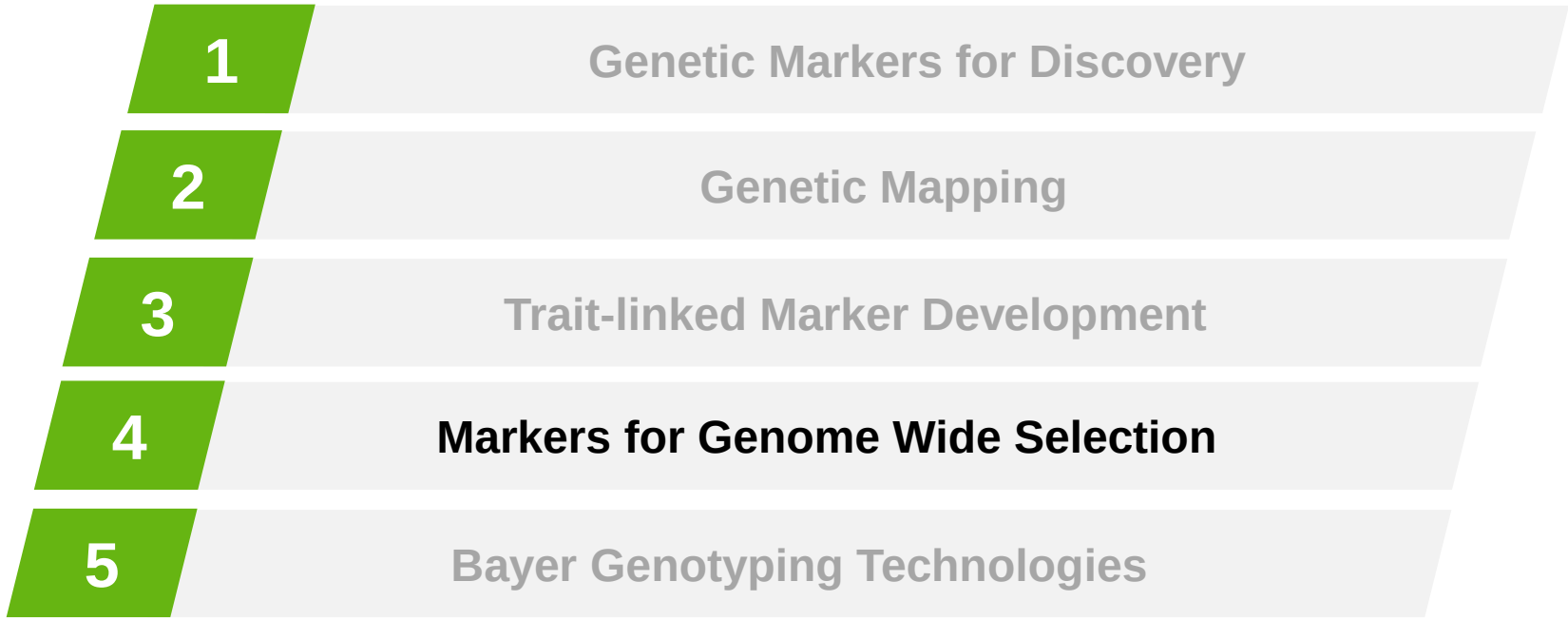
// Located in specific genomic areas

// Linked to the variation of a given trait

// Can be used to follow said trait (MAS)



Bayer Russia Molecular Marker Training :
Genetic Markers and Technologies



Agenda

- Overview
- Selection of candidate markers
- Designability screening
- Marker Selection
- Marker/Sample Validation
- Deployment, Versioning and Maintenance

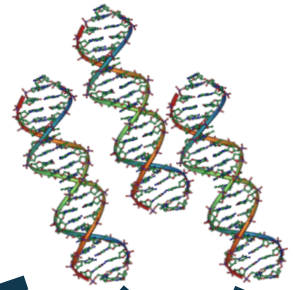
Marker data is generated by DNA sequencing and probe - based methods

Seed
(Single or Bulk)

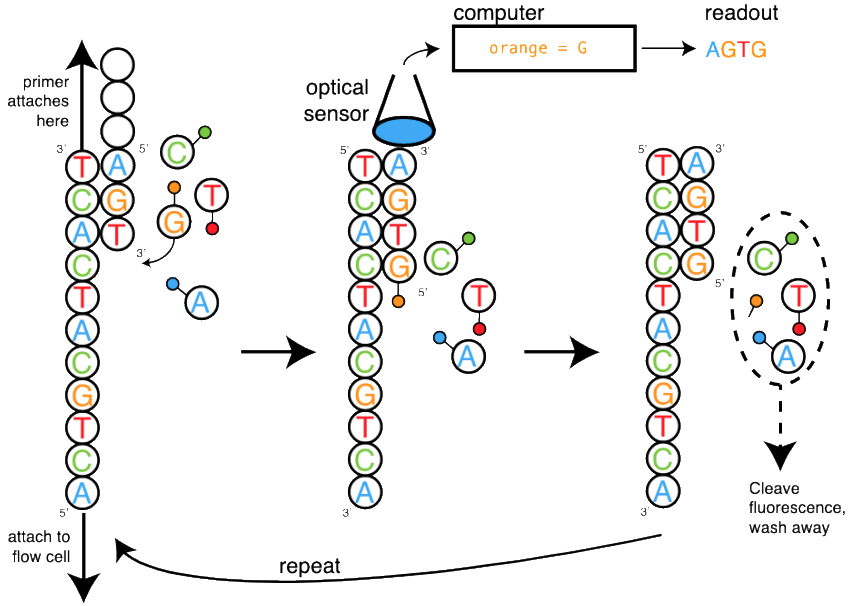
OR

Leaf Tissue

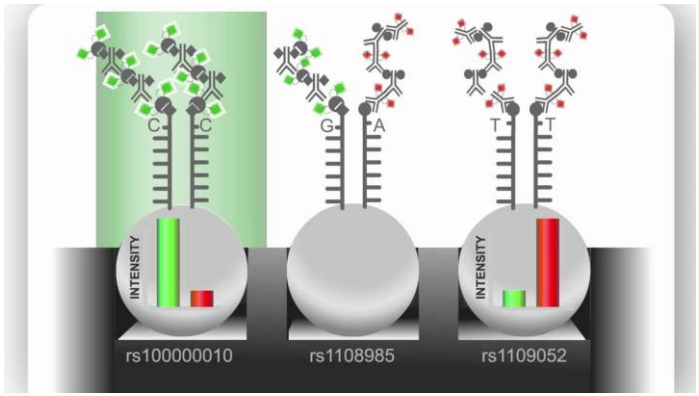
DNA
Extraction



DNA Sequencing (WGS,GBS,etc)

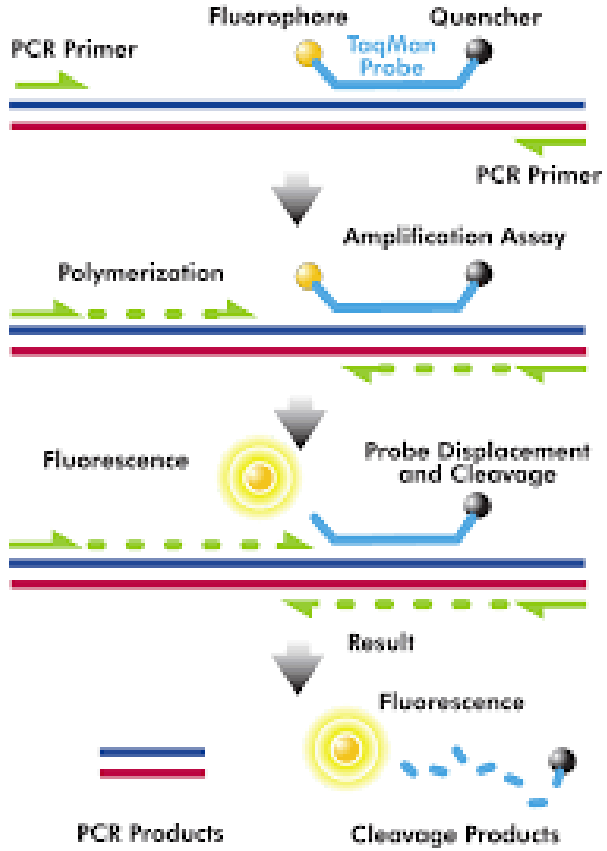


Infinium Assay (Fingerprinting)

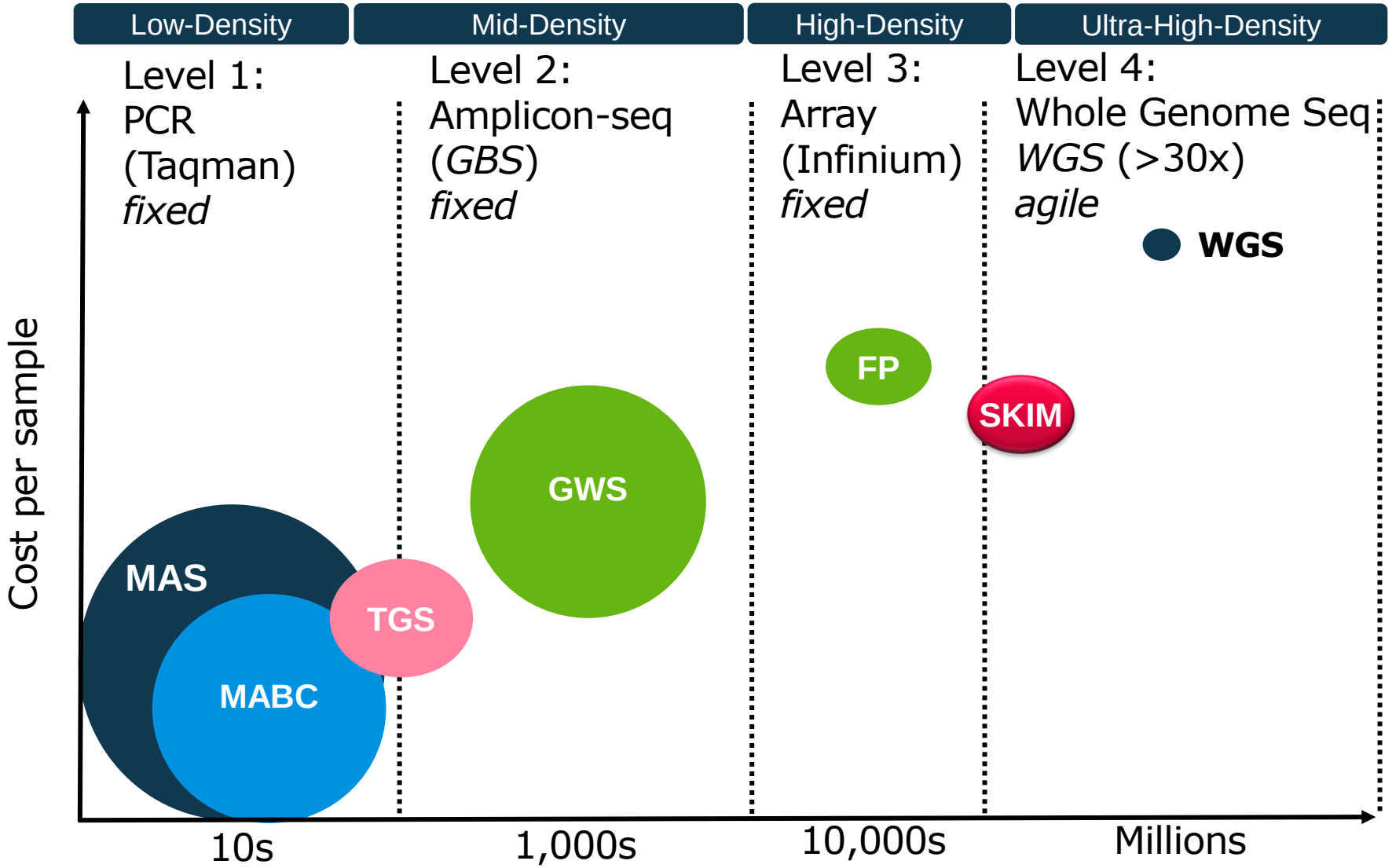


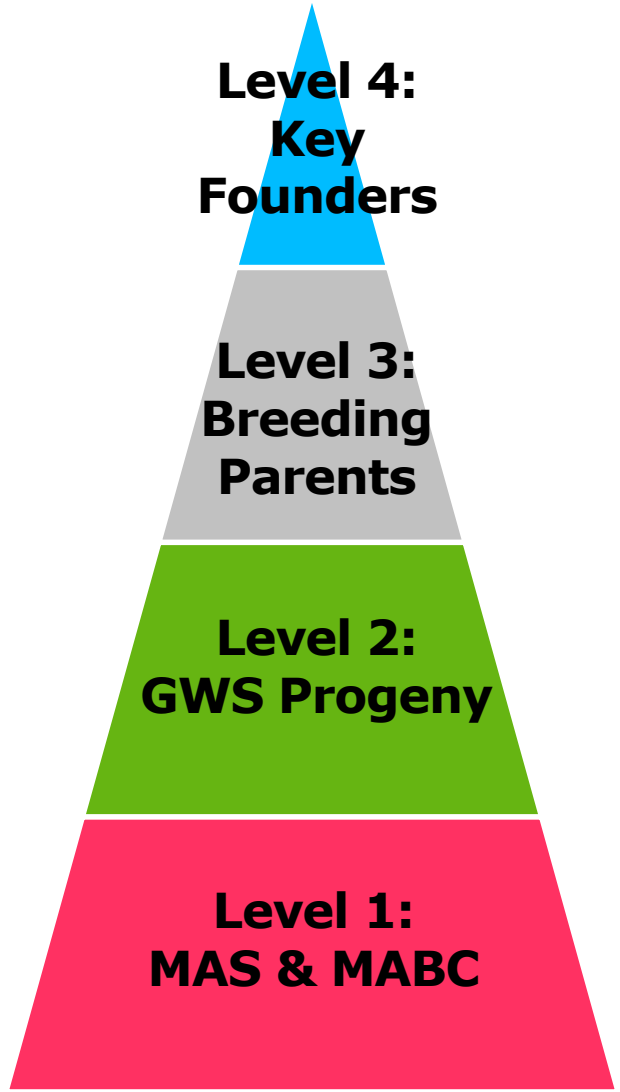
AA AT TT

TaqMan Assay

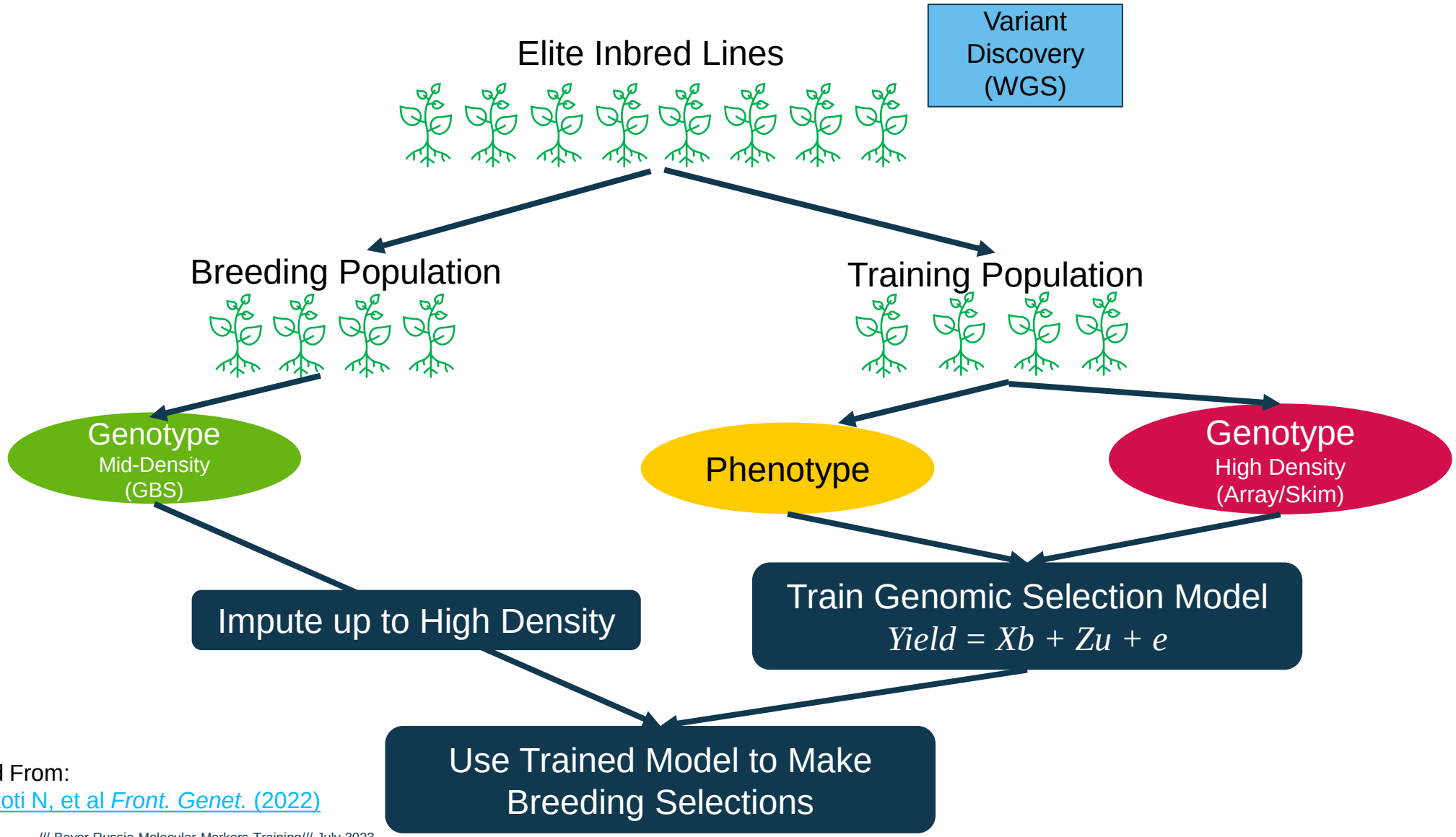


Production genotyping technologies at Bayer Crop Science - Precision Genomics





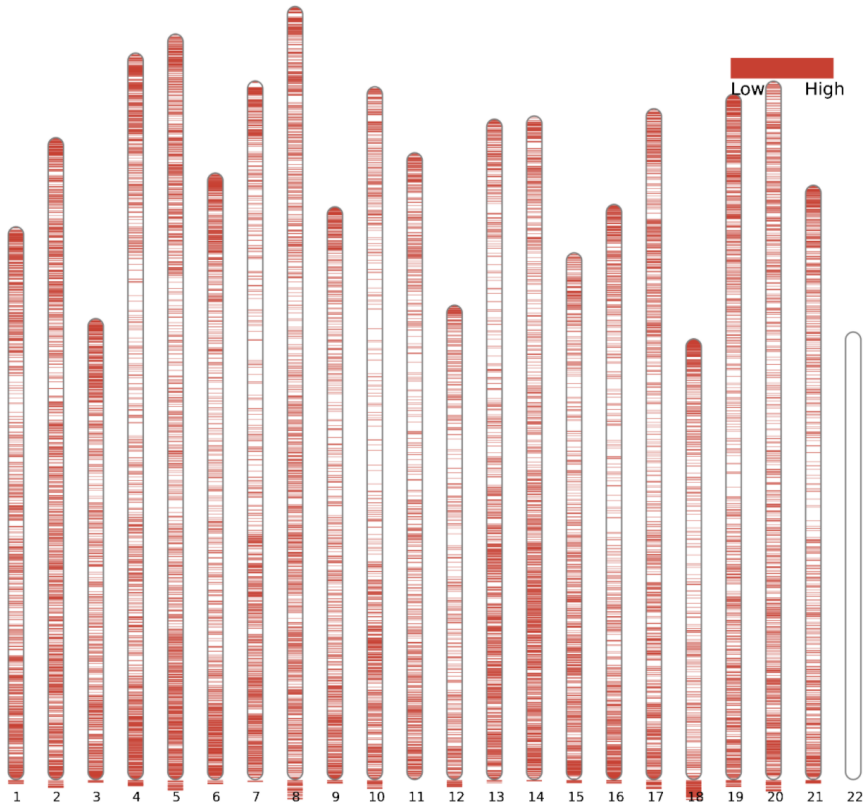
Genomic Selection



Adapted From:
Buhdlakoti N, et al *Front. Genet.* (2022)

Objective: Develop reduced-representation marker set(s) that covers as many haplotypes as possible, constrained by per-sample assay cost

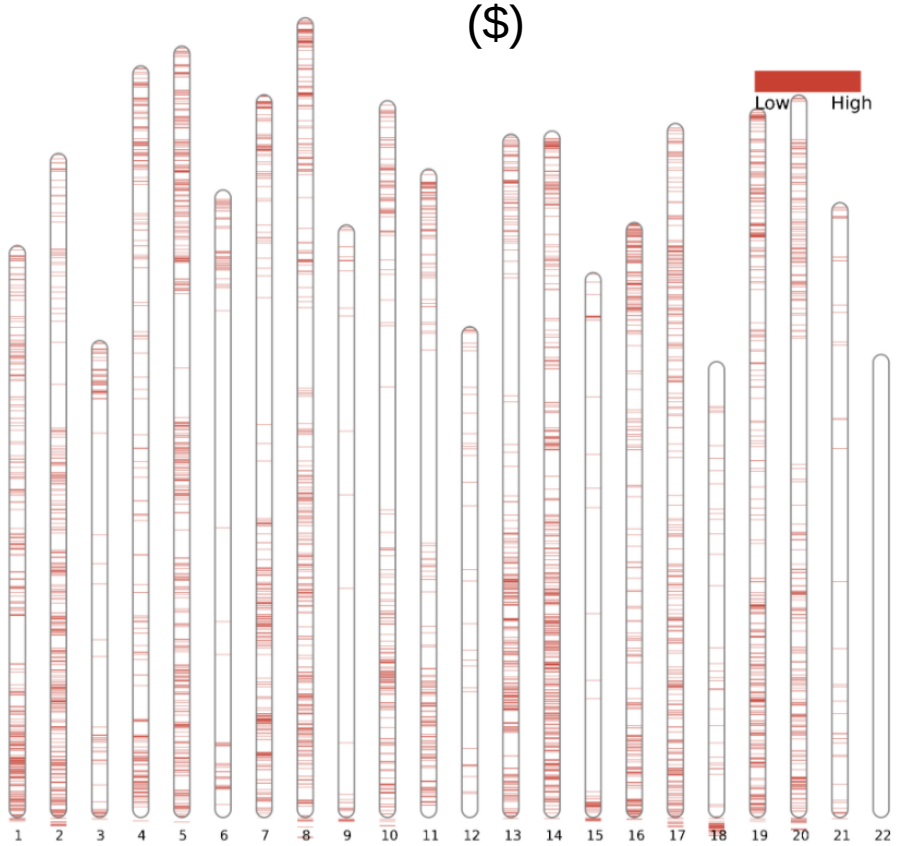
Full Set of Markers
($$$)



Select Markers by

- Parental Haplotypes
- Assay performance
- Required/Must-in Status

Reduced-Representation Set
($)



How to Build a Marker Set

Form Candidate Pool of Markers

- Collect all marker sequences and metadata
- Define germplasm to test
- Correct inconsistencies in alleles
- Correct genetic map

Predict Marker "Designability"

- Select markers to screen
- Determine genotyping platform
- Send markers to vendor

Select Markers

- Select markers to validate
- Manufacture assays/probes

Validate Selected Markers

- Measure success metrics compared to legacy data
- Drop low-performing markers
- Backfill must-in markers

Deployment, Versioning, Maintenance

- Put marker set into production
- Update production pipelines
  - Lab pipeline
  - Analysis pipeline

Genotyping Platforms

Designing Markers for Genomic Selection

Form Candidate Pool of Markers

Discovery: Whole Genome Sequencing

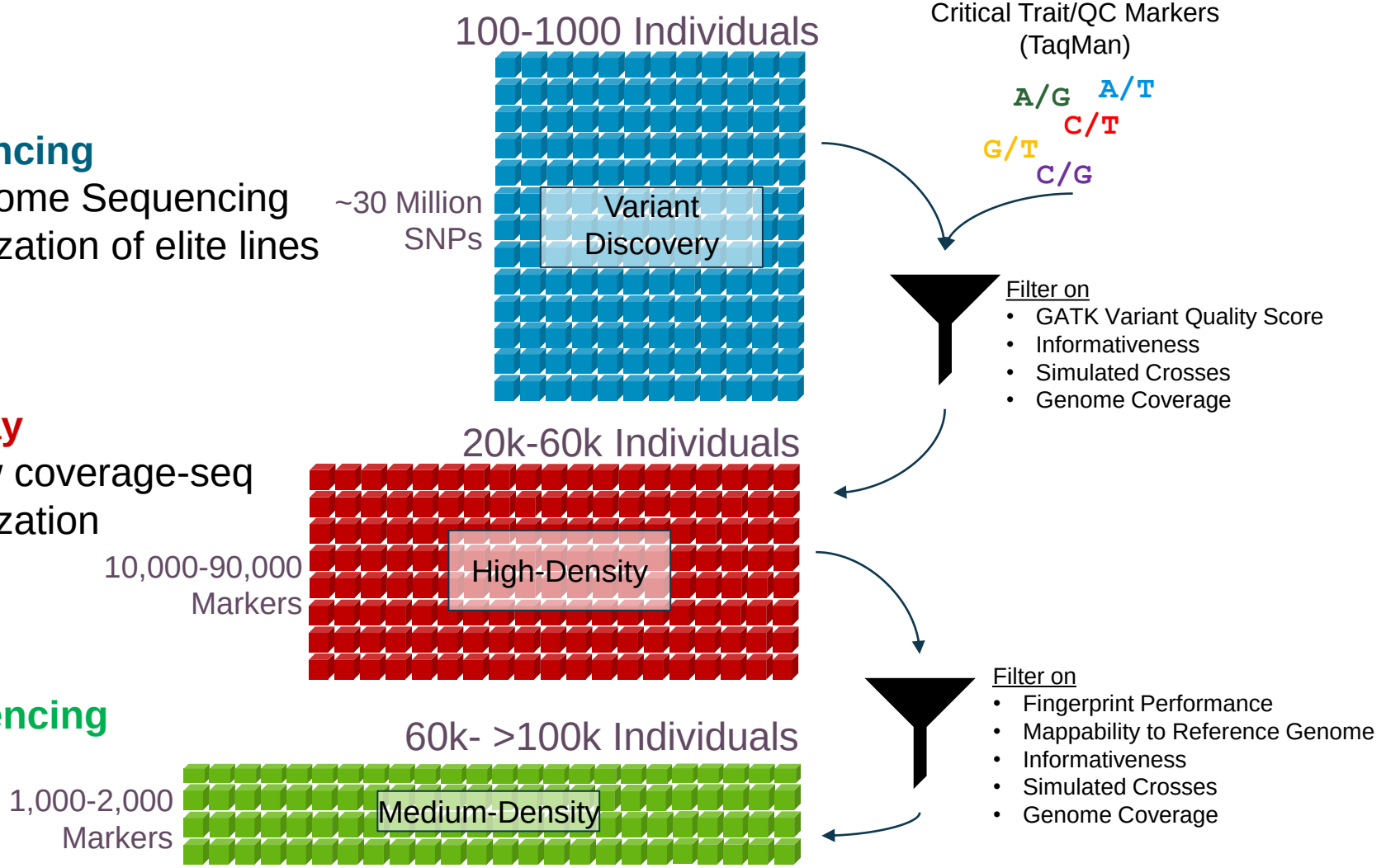
- Discovered with >30x Whole Genome Sequencing
- Purpose: Deep genetic characterization of elite lines
- Cost: $$$$

Production: Genotyping Microarray

- Genotyped on a microarray or low coverage-seq
- Purpose: Deep genetic characterization *with high-quality markers*
- Cost: $$

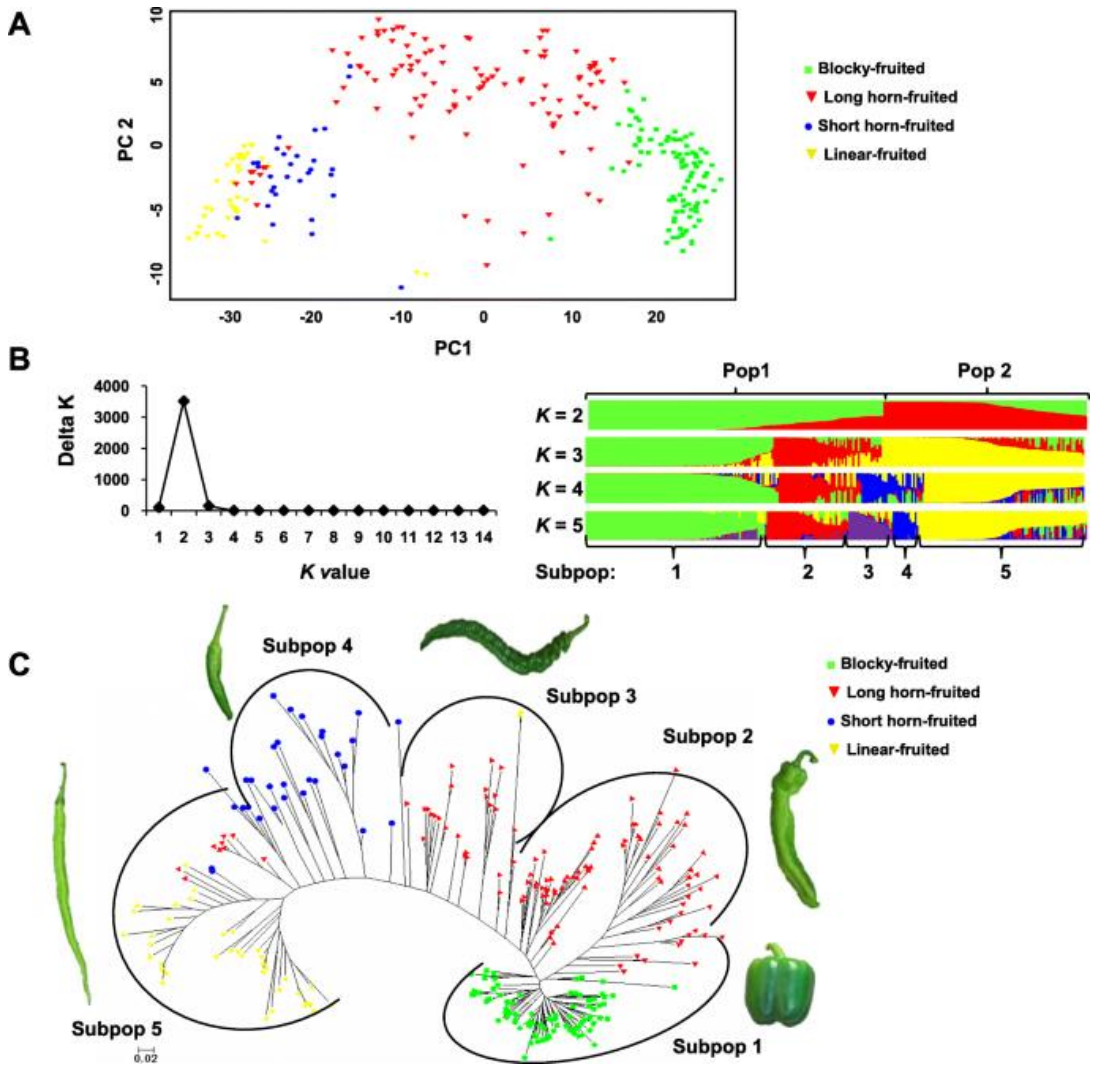
Production: Genotyping by Sequencing

- Multiplexed amplicon sequencing
- Purpose: Genomic Selection
- Cost: $



Developing a Diversity Panel – Pepper Example

Form Candidate Pool of Markers

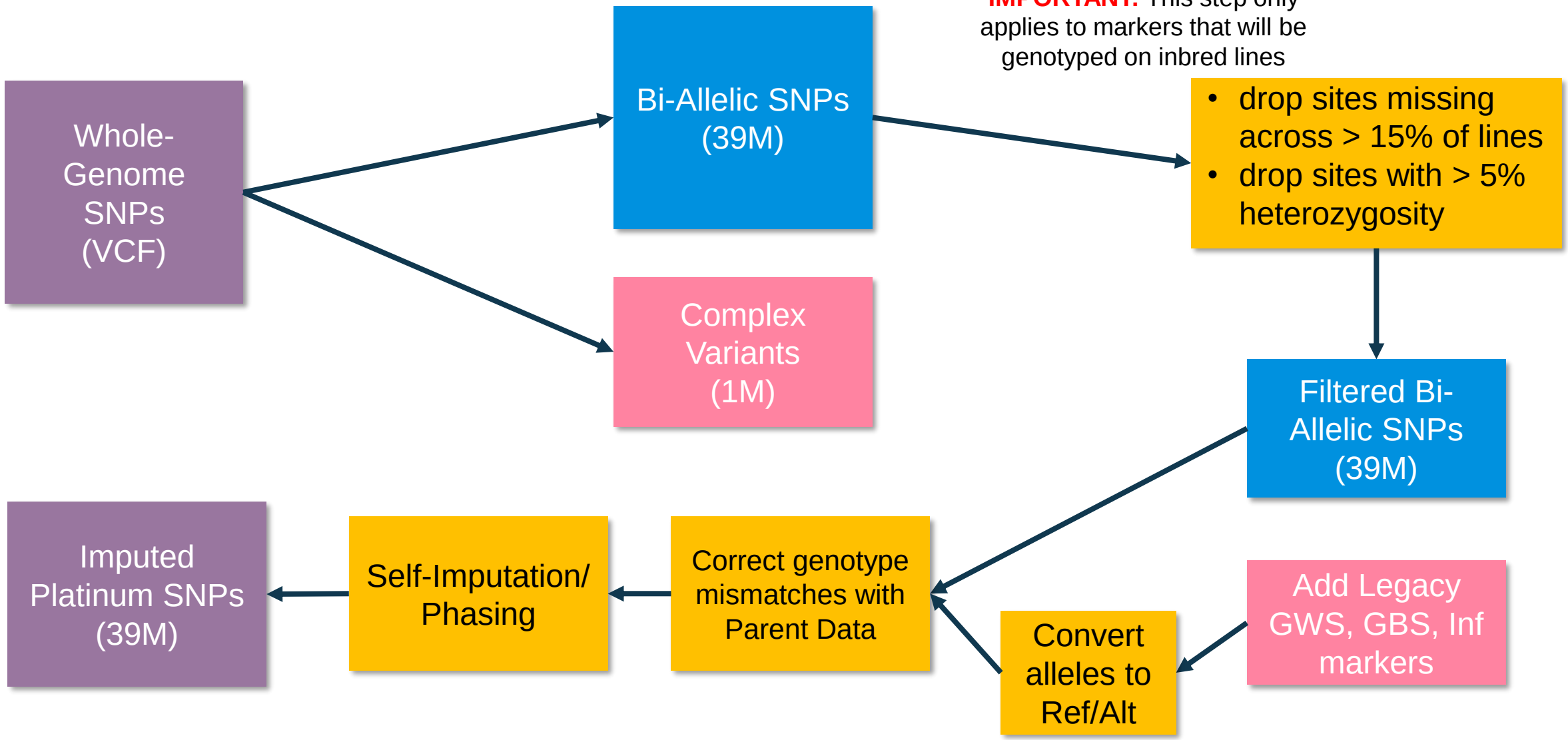


Du, H., Yang, J., Chen, B. *et BMC Plant Biol* **19**, 578 (2019).

- When selecting a marker validation diversity panel select markers that can distinguish between:
  - Major sub-populations
  - Major phylogenetic groups
  - Major phenotypic groups

Selecting Reliable Markers from Whole-Genome SNPs

Form Candidate Pool of Markers

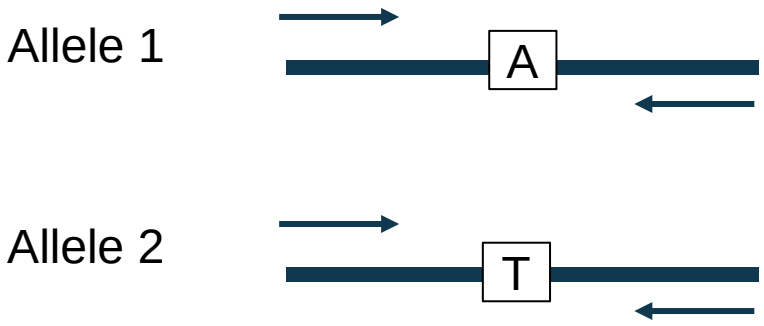


Form Candidate Pool of Markers

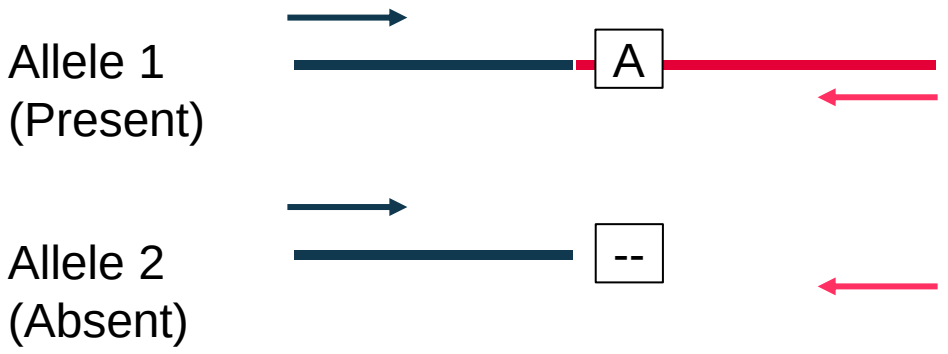
Required/Must-in Markers

- These are markers that are specifically requested by the breeders
  - Also contains supply chain QA/QC markers (10-30 markers)
- Most of these markers are designed to track markers associated with a trait, and some of the designs can be somewhat complex

SNP Marker



Large Indel Marker

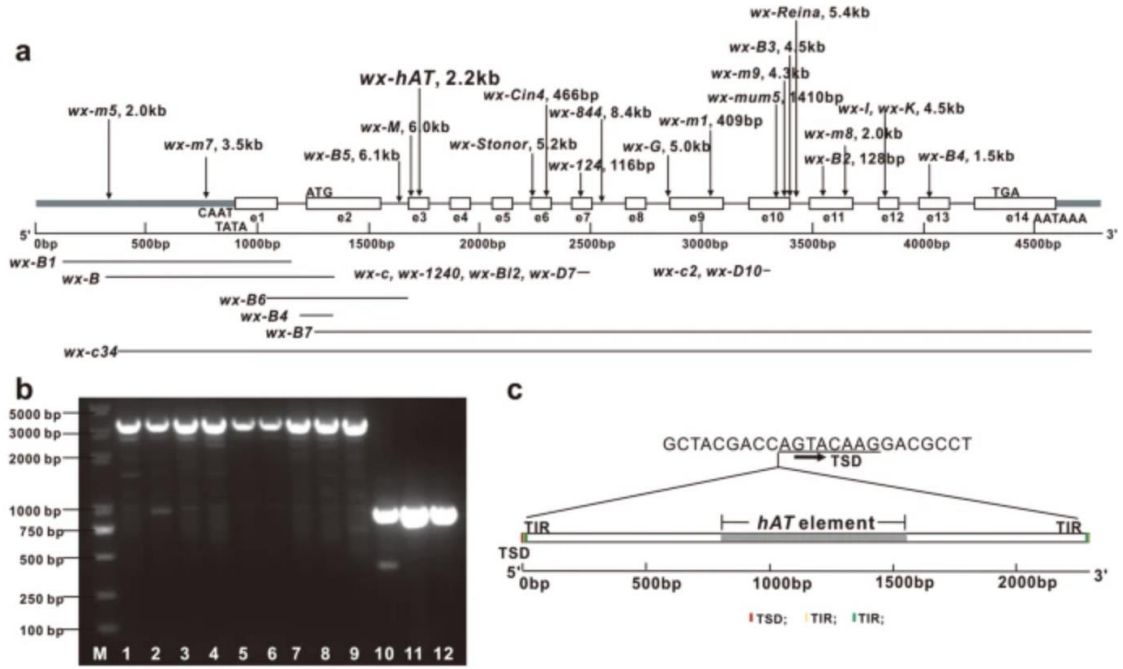


Form Candidate Pool of Markers

Trait markers can have complex/indel designs

Examples

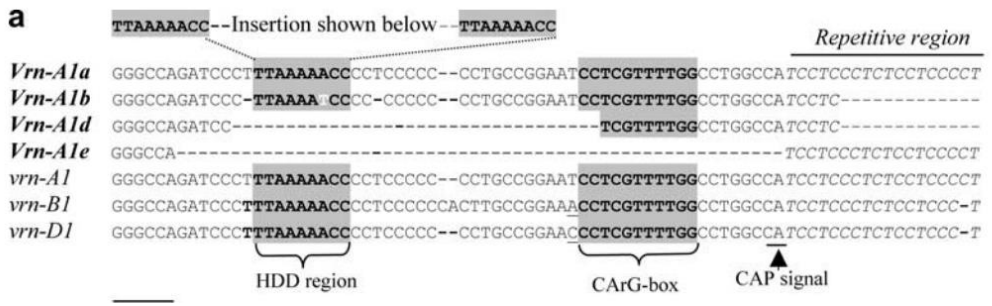
Waxy Locus in Maize: 2kb *hAT* Element insertion



M. Luo, *Nature Sci Reports* (2020)

7 Vernalization Genes in Wheat

(the fact that this is in a known repetitive region adds to the ~~pain~~ fun)



L. Yan, *Theor Appl Genet* (2004)

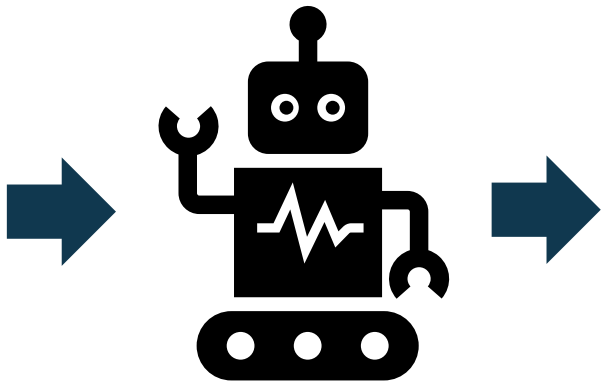
Predict Marker "Designability"

Predicting Designability based on Sequence

Marker Flanking Sequence

Marker	Sequence
M_1	ATACAGA[C/G]CTGAACA
M_2	GTCCCGT[A/T]CAGGGGT
…	…
M_N	GTGGGAT[A/C]CAAGGCC

Vendor Software



Scores and Designs

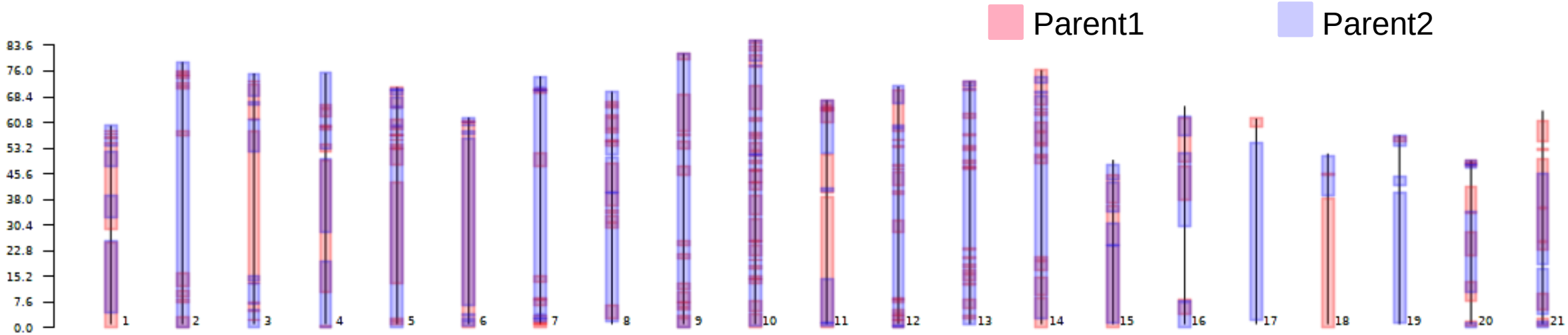
Marker	Score	Primers/Probes
M_1	100%	ATACAGA
M_2	50%	CAGGGGT
…	…	…
M_N	90%	CAAGGCC

- Marker Flanking Sequence can come from:
  - A reference genome
  - Targeted sequencing
  - Previous design from a different platform

- Marker Flanking Sequences are scored by:
  - Repetitiveness
  - Probability of cross-hybridization
  - Additional sequence features
- Include markers with favorable designability scores
- Remove markers with unfavorable scores

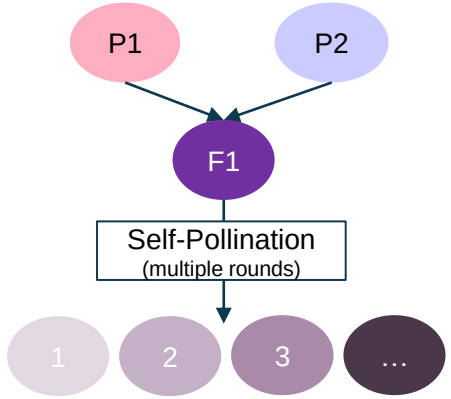
Select Markers

Selecting markers using parental information



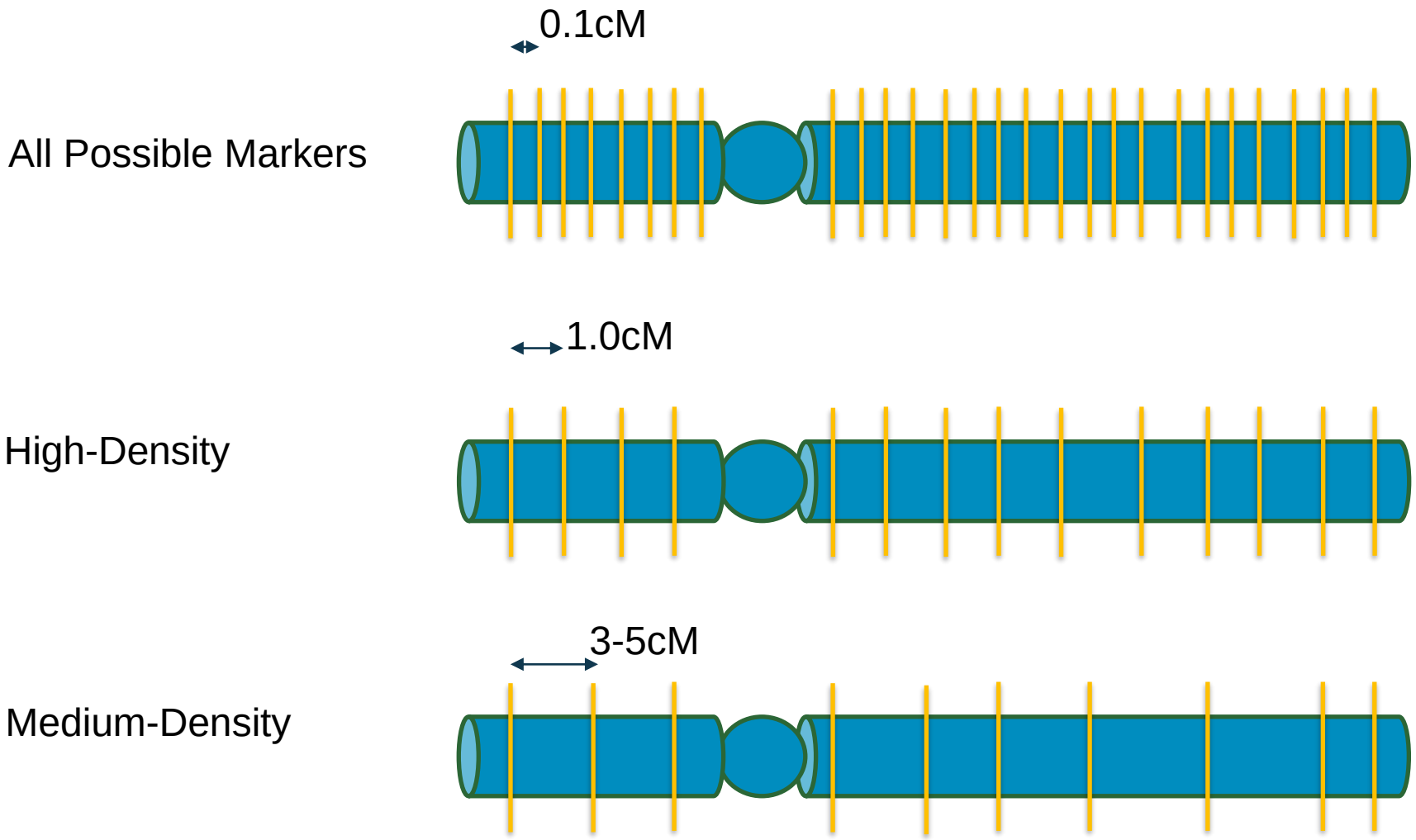
1. Using genotype data on the parents, assign marker alleles to each parent

2. Determine the minimum number of markers that will cover the maximum number of haplotypes across all pedigrees



Selecting the minimum number of markers to maximize genetic map coverage

Select Markers

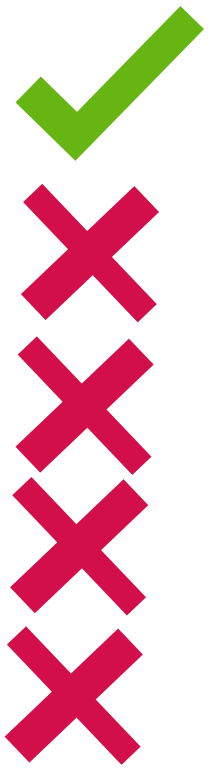


- The exact amount of spacing is decided by breeding stakeholders and optimized by genome size and recombination rate
- If gaps appear, they can be filled by gap-filling algorithms

Validate Selected Markers

Selecting markers using assay performance metrics

	Plant1	Plant2	Plant3	Plant4	Plant5	Plant6
Marker1	AA	AA	AA	BB	BB	BB
Marker2	AA	BB	BB	BB	BB	BB
Marker3	BB	Missing	BB	Missing	Missing	AA
Marker4	BB	BB	BB	BB	BB	BB
Marker5	AB	AB	AB	AB	AB	AB



- Low het
- High MAF
- No missing

- Low MAF

- High Missing

- All homozygous BB

- High heterozygosity

Determining Genotyping Accuracy with Controls

Validate Selected Markers

Genotypes From New Marker Set

	Plant1	Plant2
Marker1	AA	AA
Marker2	AA	AA
Marker3	BB	AA
Marker4	BB	BB
Marker5	AB	AB

Control Genotypes (WGS, Array, or GBS)

	Plant1	Plant2
Marker1	AA	AA
Marker2	AA	BB
Marker3	BB	BB
Marker4	BB	BB
Marker5	AB	AB

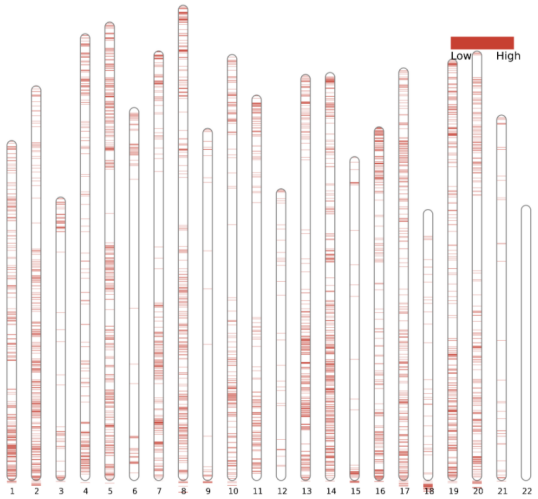
Concordance Matrix

	Plant1 Match?	Plant2 Match?	Marker Concordance
Marker1	1	1	100%
Marker2	1	0	50%
Marker3	1	0	50%
Marker4	1	1	100%
Marker5	1	1	100%
Sample Concordance	100%	60%	

Deployment, Versioning, Maintenance

Marker set versioning and deployment

New, Validated Marker Set



Database

- Tracks minor marker set/ genetic map updates

Markers Database

- Tracks all markers across all platforms

Genetic Maps Database

- Tracks **major** genetic map updates across all platforms

Data Warehouse

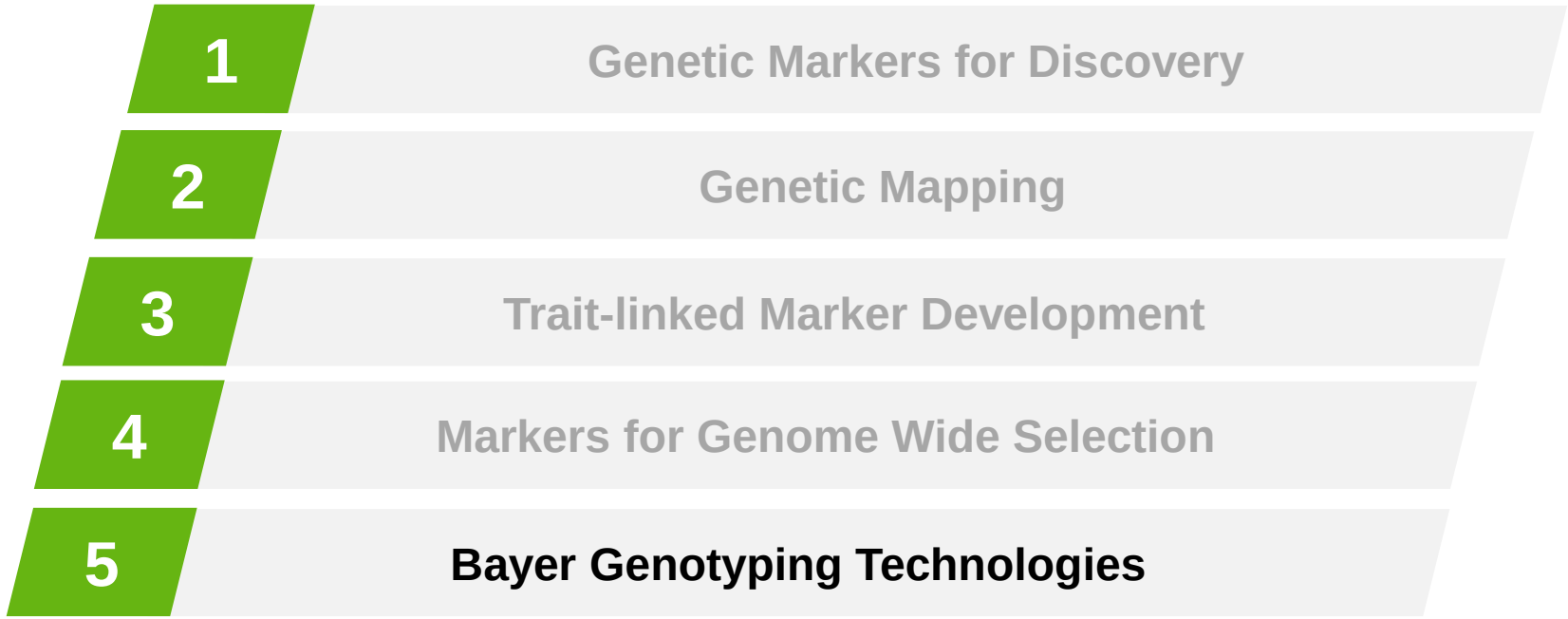
- Exposes marker data and genetic maps as query-able tables to end-users

// Both marker sets and genetic maps are given unique Identifiers, which are updated with every new version

// New markers require an additional registration step in database, where they are assigned new IDs

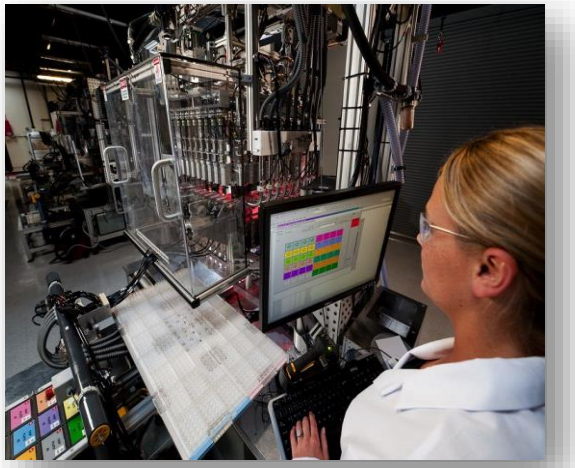
Q&A
Discussion

Bayer Russia Molecular Marker Training :
Genetic Markers and Technologies



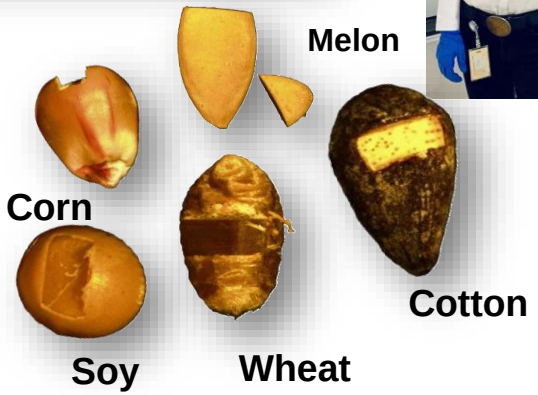
Pipelines need to be scalable to find the rare, high performers.

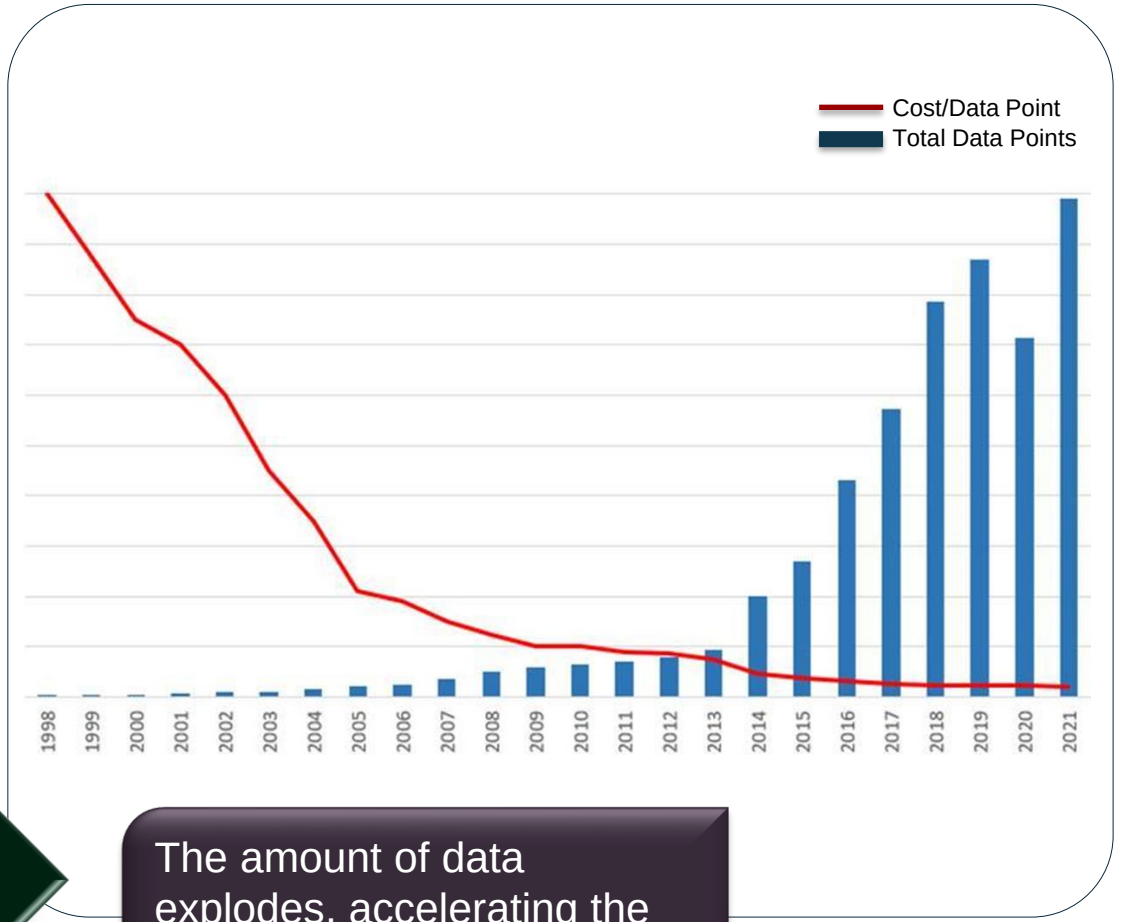
Scaling is enabled by increased capacity and decreased cost



Robotics Automation

Seed Chipping





The amount of data explodes, accelerating the ability to realize step-change improvements

Genotyping Technology Comparison

~10k to ~800k markers

High density markers

HD-Array by Affymetrics or Infinum. Selected markers which are best fit for the breeding pool as Fingerprinting (FP) platform. Used on parental / elite germplasm in later breeding pipeline.

Medium density markers

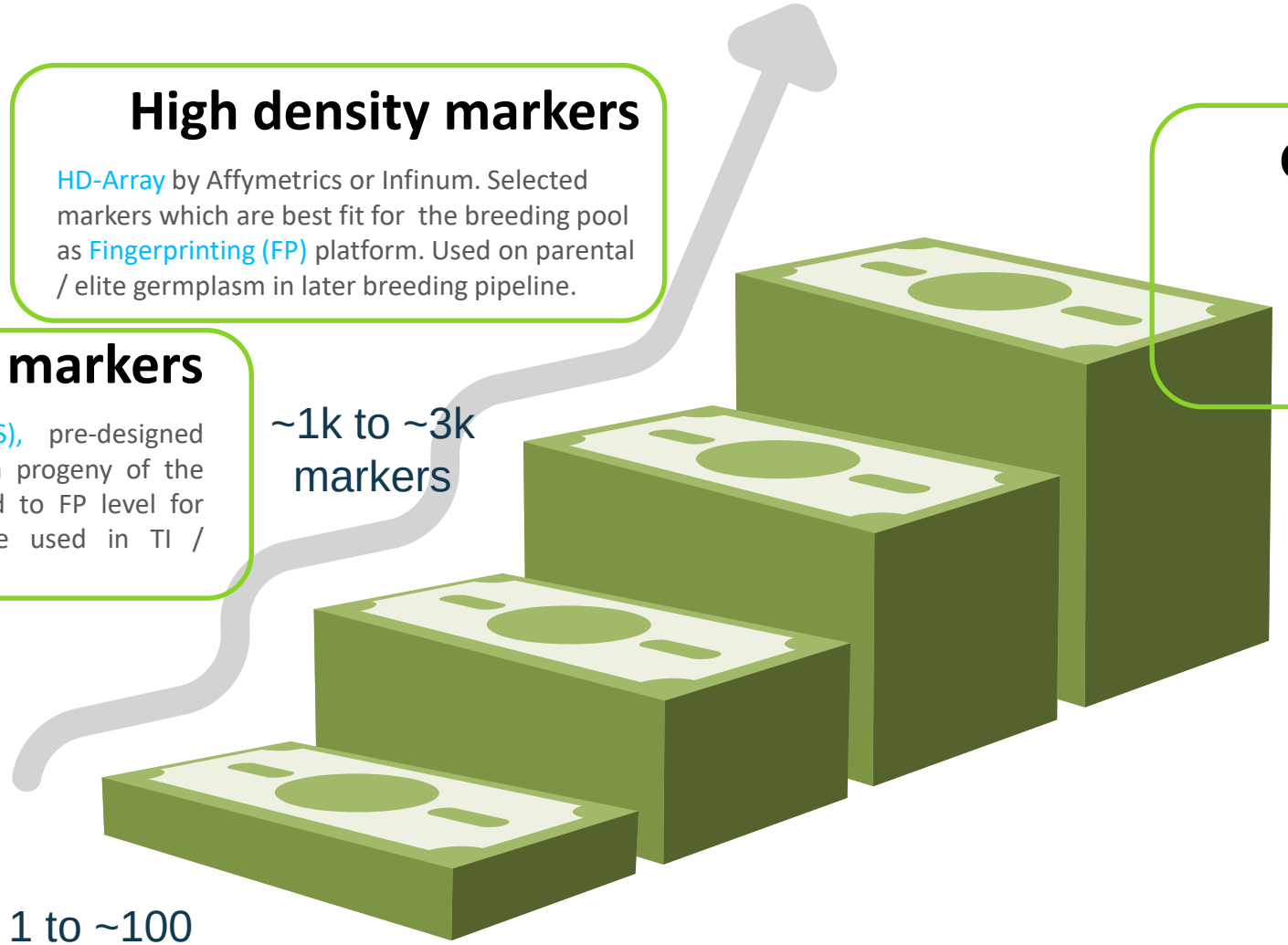
Genotyping by sequencing (GBS), pre-designed marker set based on FP. Used on progeny of the breeding populations and imputed to FP level for genomic prediction. Can also be used in TI / Discovery.

~1k to ~3k markers

Low density markers

Taqman / KASP markers. Used to track traits controlled by limited number of genes.

1 to ~100 markers



Genomic sequence

Elite / core / founder germplasm to identify necessary polymorphisms to strengthen the genomics platform and additional coverage to support traits discovery.

Low density genotyping

Equipment

Thermocycler or Hydrocycler

Plate Reader

OR

qPCR instruments

Talent

Assay Design

Experimental Design

Bench Scientists

Assay Scoring

Scaling

Liquid handling instrumentation
(Tecan, Biomek, Firefly)

Robotics for plate movement

Consider shift work for quicker turn-around times

Medium density genotyping

Equipment

Thermocycler or Hydrocycler

Short-read sequencer

Talent

Marker Selection

Experimental Design

Bench Scientists

Computational Biology

Scaling

Liquid handling instrumentation (Tecan, Biomek, Firefly, etc)

Robotics for plate movement

Consider shift work for quicker turn-around times

High density genotyping and genomic/transcriptomic sequencing

Equipment

Thermocycler or Hydrocycler

Appropriate array preparation and reader instrumentation

Short-read sequencer
(for low coverage seq)

Talent

Marker Selection

Experimental Design

Bench Scientists

Computational Biology

Scaling

Scalable high quality DNA preparation

Liquid handling instrumentation
(Tecan, Biomek, Firefly)

Genotyping does not have to be developed in-house as several global companies offer genotyping services

Examples for genotyping/sequencing companies

- Eurofins
- LGC
- SGS Institut Fresenius
- Gencove
- Paragon Genomics

*All company names are used for illustrative purposes only. Inclusion does not imply Bayer Crop Science endorsement.

Considerations when leveraging a 3rd party

- When scoping the project, important to have key information available
  - Expected number of samples/ batch
  - Single project or recurring submissions
  - Expected turn around time
  - What development or customization will be needed? Will you work together or do you need 3rd party to develop?
- Data flow from 3rd party back to breeding program
  - Is analysis required by 3rd party or will raw data be supplied?
  - Will standard data report be sufficient?
  - How will information be exchanged?

Q&A
Discussion

DISCLAIMER

THE INFORMATION CONTAINED HEREIN IS EXPERIMENTAL IN NATURE AND IS PROVIDED "AS IS". BAYER MAKES NO REPRESENTATIONS AND EXTENDS NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO THE MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, OR THAT THE USE OF THE INFORMATION WILL NOT INFRINGE ANY THIRD-PARTY PATENT, COPYRIGHT, TRADEMARK, OR OTHER PROPRIETARY RIGHTS.