T5.4 Using genomics to track the evolution of heterogeneous (organic) material

Module 5 -Organic heterogeneous material (OHM) design and development

Training courses in Organic Breeding

07.03.2025

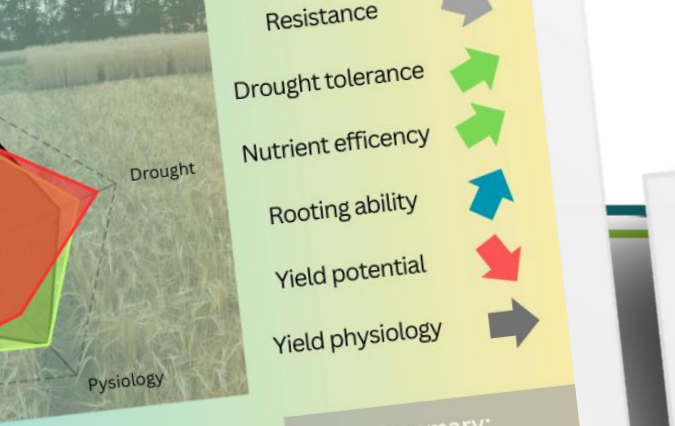
Michael Schneider, FiBL

LiveSeeding

Co-funded by the European Union

Funded by the European Union, the Swiss State Secretariat for Education, Research and Innovation (SERI) and UK Research and Innovation (UKRI). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or REA, nor SERI or UKRI.

UK Research and Innovation



Resistance
Drought tolerance
Nutrient efficency
Rooting ability
Yield potential
Yield physiology

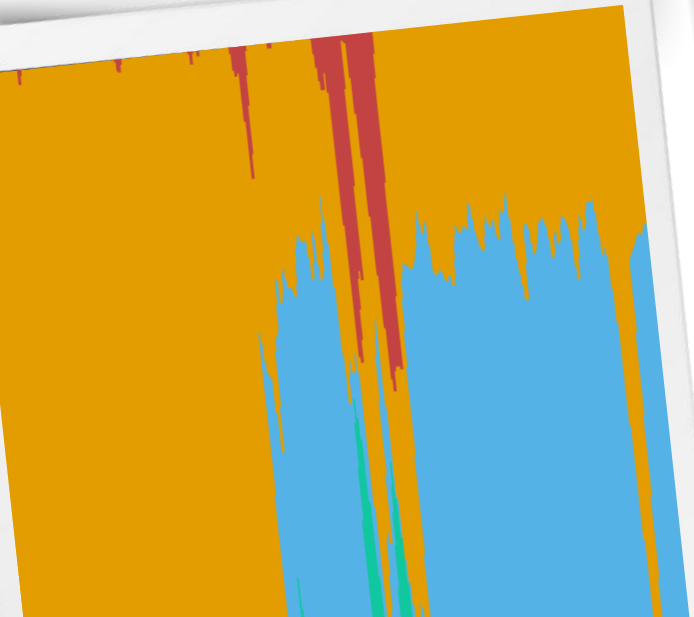
f the material

6th generation

Summary:

Strengths:
NUE &
Drought tolerance

Weakness:
Yield potential



Training in organic breeding

Module 4: Organic heterogeneous material (OHM) design and development

Unit 5.4: Using genomics to track the evolution of heterogeneous (organic) material

Michael Schneider, FiBL

Co-funded by the European Union

Funded by the European Union, the Swiss State Secretariat for Education, Research and Innovation (SERI) and UK Research and Innovation (UKRI).

UK Research and Innovation

Training in organic breeding organized in 5 Modules

1. Module 1 - Plant Genetic Resources (PGRs): collection, conservation and exchange to support the increase of agrobiodiversity in farming systems
2. Module 2 - Phenomics: approaches and tools for genetic resources and breeding material characterisation - FEBRUARY 3rd 2025, 9:00 to 17:30 CET
3. Module 3 - Breeding methods fundamentals - FEBRUARY 13th 2025, 9:00 to 18:00 CET
4. Module 4 - Development and application of molecular methods in organic breeding - MARCH 4th 2025, 9:00 to 18:00 CET
5. Module 5 - Organic heterogeneous material (OHM) design and development - MARCH 7th 2025, 9:00 to 18:00 CET

LiveSeeding

Planned for today

- **Give full walk through**
  - **Sampling populations**
  - **Isolation of genomic DNA/RNA**
  - **Genotyping approaches**
  - **Analysis pipelines to transform bare DNA into knowledge**
    - Simple Settings (e.g. Parents known, biparental cross)
    - Complex Settings (e.g. Multiple parents, no reference information)
  - **Results and assumptions to draw (by examples)**
- **Time for discussion & a small Quiz**

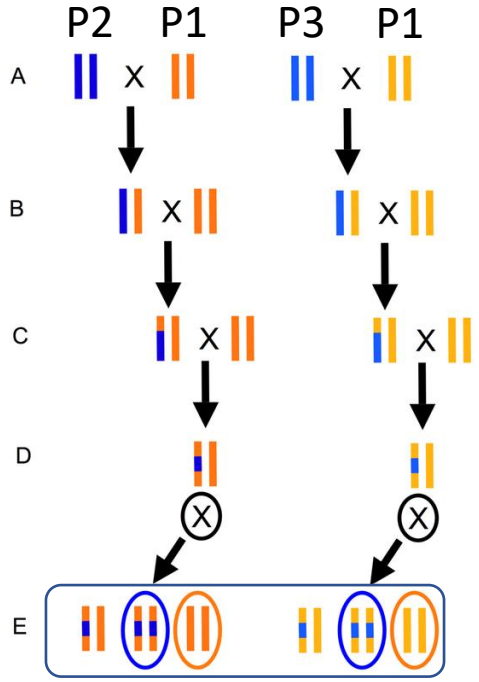
Questions are welcome anytime

LiveSeeding

Creating organic heterogenous material

Organic heterogenious material

3 parents crossed, 2 wild relatives

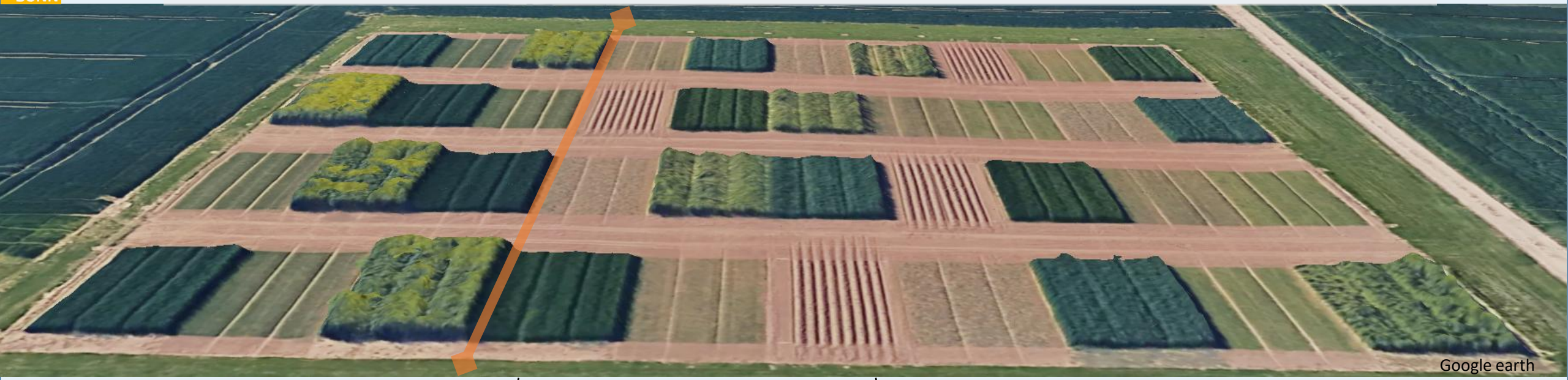


Bulk and send to field

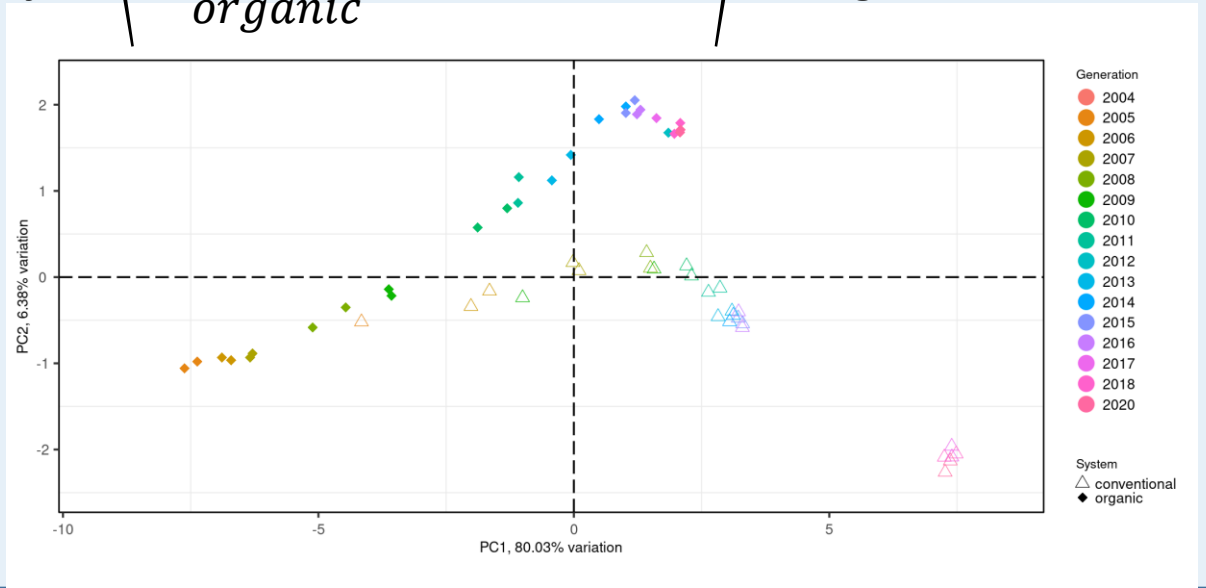
Modified figure from Lowry & Wills 2021; doi.org/10.1371/journal.pbio.1000500

LiveSeeding

Experimental evolution – natural adaptation to farming systems



Crop wild relative x elite variety $\rightarrow \left\langle \frac{conventional}{organic} Environment \right\rangle + 26\ generations\ - artificial\ selection = ???$



LiveSeeding

We want to know..

What is happening in these populations?

How are they adapting?

How does the environment impact the pop.?
(Climate, soil, farming practice,..)

LiveSeeding

Sampling populations / OHM

1. Step

get the DNA from the population

LiveSeeding

How to?

Money is a limited factor, so we are going to construct a pooled sample

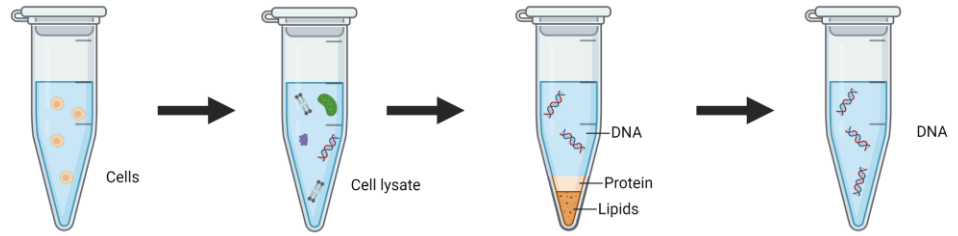
1. **Selecting x-hundred seeds as a representative sample for the entire population**
2. **Grinding the seeds to flour (medium fine)**
   - **No greenhouse required (+)**
   - **Low labour costs (++)**
   - **Liquid nitrogen required for oil rich seeds (-)**
   - **Grinding slow or requires special mill insets (-)**
   - **DNA extraction challenging (--)**

   or

2. **Collecting equally sized leaf discs for each genotype to test**
   - **Greenhouse required (-)**
   - **14 days delay until sample collection (-/+)**
   - **High labour demand due to leaf sampling (equal size) (--)**
   - **Liquid nitrogen requiered (-)**
   - **Collection in 5 ml Eppendorf tubes possible – parallelized grinding in a swing mill possible (+)**
   - **DNA extraction simple (++)**

3. **Extracting DNA (RNA) – for example with the Seed mini kit (Zymo)**
   - **Starting material**
     - Flour :
       - min. 7.5 ml falcon tubes and 1g Flour starting material
       - Additional purification steps required to remove secondary compounds (starch, oil)
       - Over night eluation in Water necessary to increase yield
     - Leaf discs
       - Grinding of disks to very fine powder – ensures equal contribution of each genotype
       - Standard DNA / RNA extraction procedure

https://www.retsch.de/de/produkte/zerkleinern/kugelmuehlen/schwingmuehle-mm-400/



Ref: https://www.aatbio.com/resources/assaywise/2022-11-1/overview-of-dna-extraction-methods

LiveSeeding

Keep in mind – the grinding step will effect your DNA fragment size and overall output

2. Step
performing the genotyping

LiveSeeding

Selecting the relevant genotyping approach

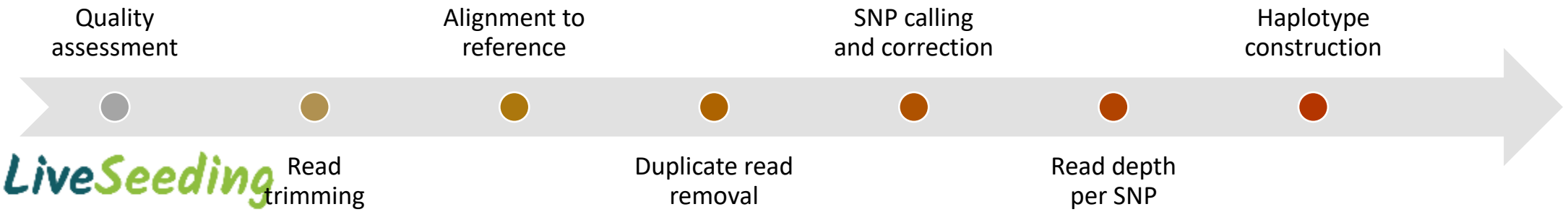
Illumina sequencing

Genome coverage (cost for 17 GB genome size)	Whole genome sequencing (WGS)	RNAseq[2]	Genotyping by sequencing (GBS)[3]	Oxford Nanopore / PacBio HiFi
Base precision	**99.999%**	**99.999%**	**99.999%**	99-99.99%
Biggest Advantage	Entire genome	Higher coverage	Higher coverage	**Haplotype calling**
Biggest Disadvantage	**Short fragments**	Expression bias	PCR duplication	Early stage tech
Costs	2	3	**1**	4
Usefulness	high	medium	low	**best**

[2] - not all genes are expressed, expression bias
[3] - few, targeted loci sequenced. Duplications not removable

The more related the parents are, less SNPs can be found => requires higher sequencing depth

Data processing



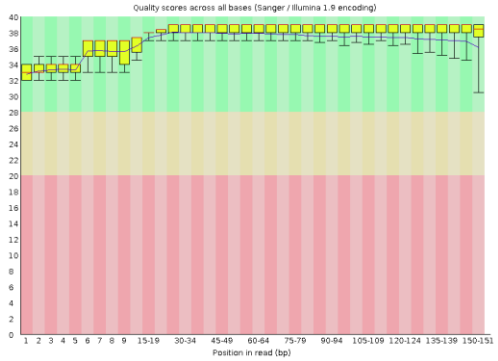
LiveSeeding

Technical requirements for poolseq

Knowledge of parental haplotypes

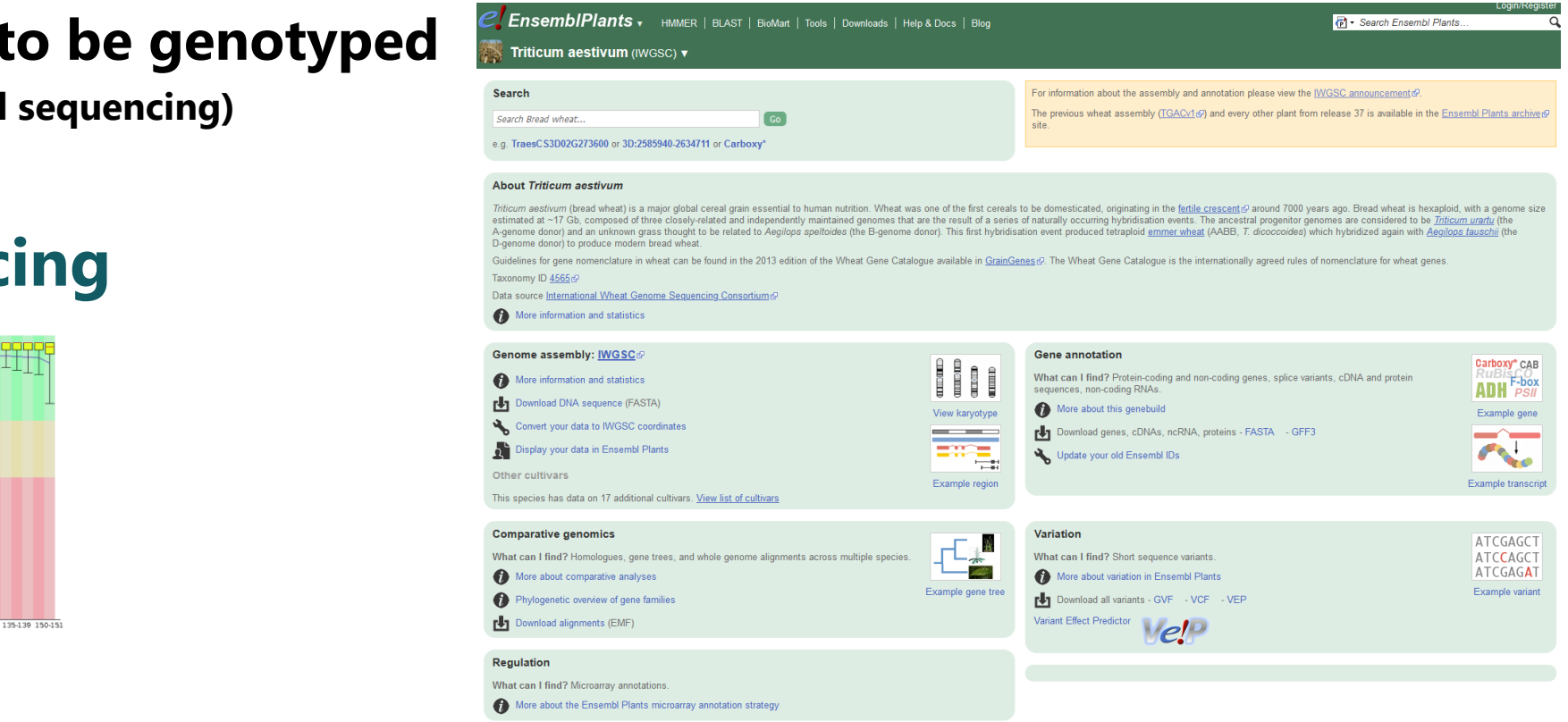
- **Parental lines need to be genotyped**
- **(Only necessary for short read sequencing)**

High quality Sequencing



Reference genome

Optional for improved results – SNP database



LiveSeeding

3. Step

Genotype cost efficient

LiveSeeding

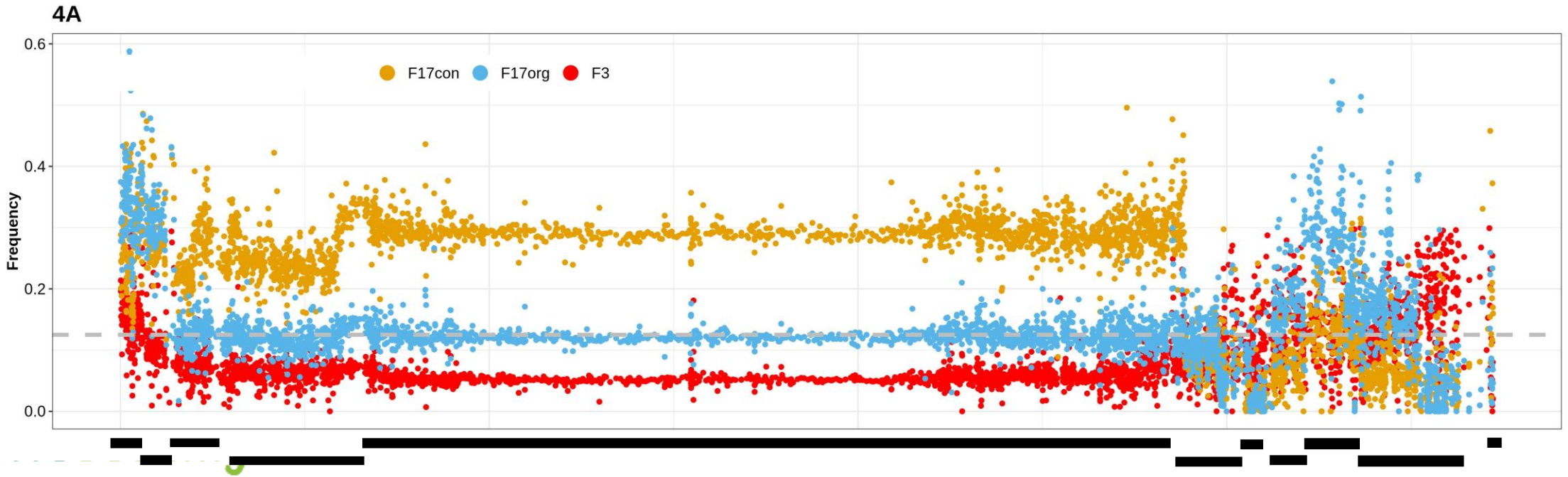
How can we estimate the accurate frequency (for low costs)?

The simple answer – from:

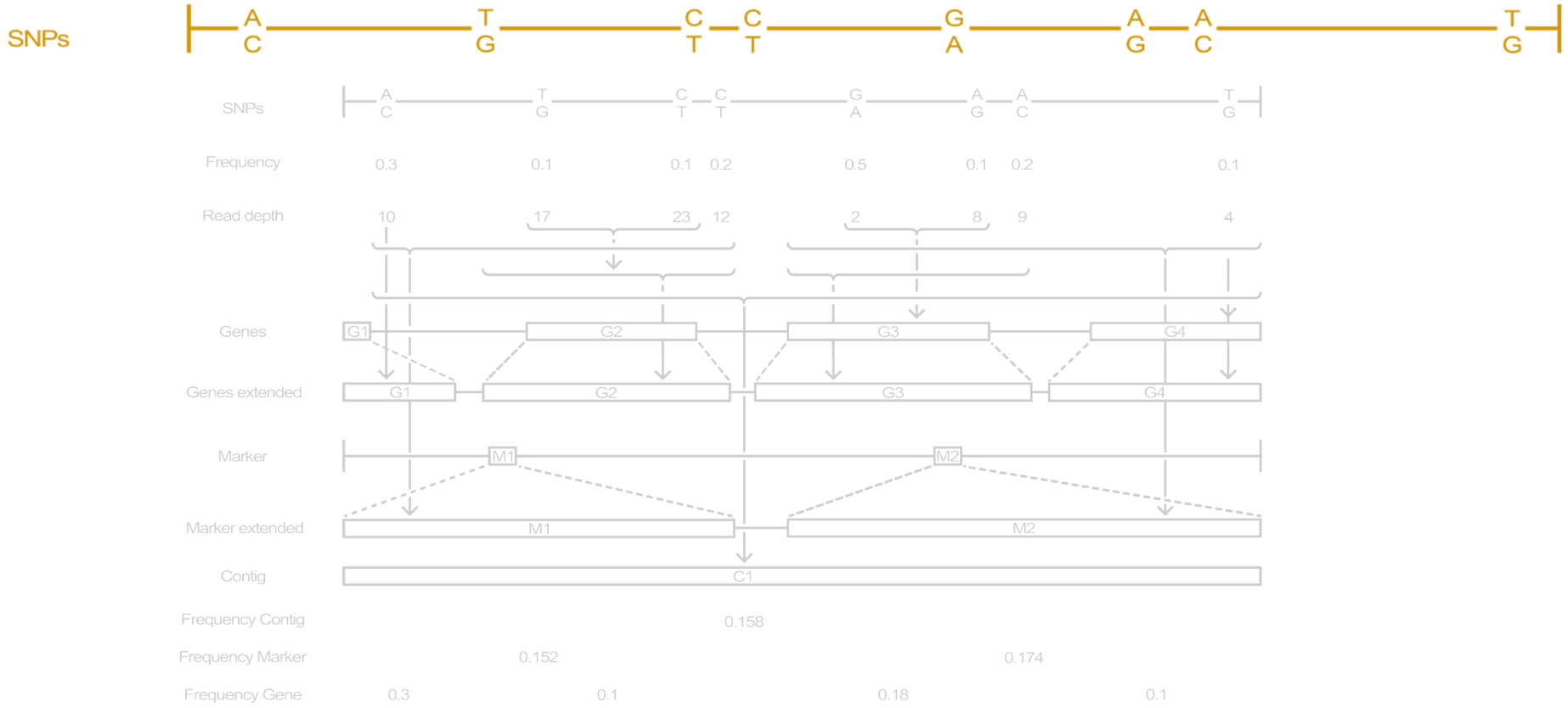
low coverage sequencing + bioinformatics + reference infromation

1. Knowledge about haplotypes – which SNPs allele are linked in the parental lines

2. Linkage disequillibrium – SNP variants are linked on the same genomic fragment



Strategies to increase power at low squencing depth



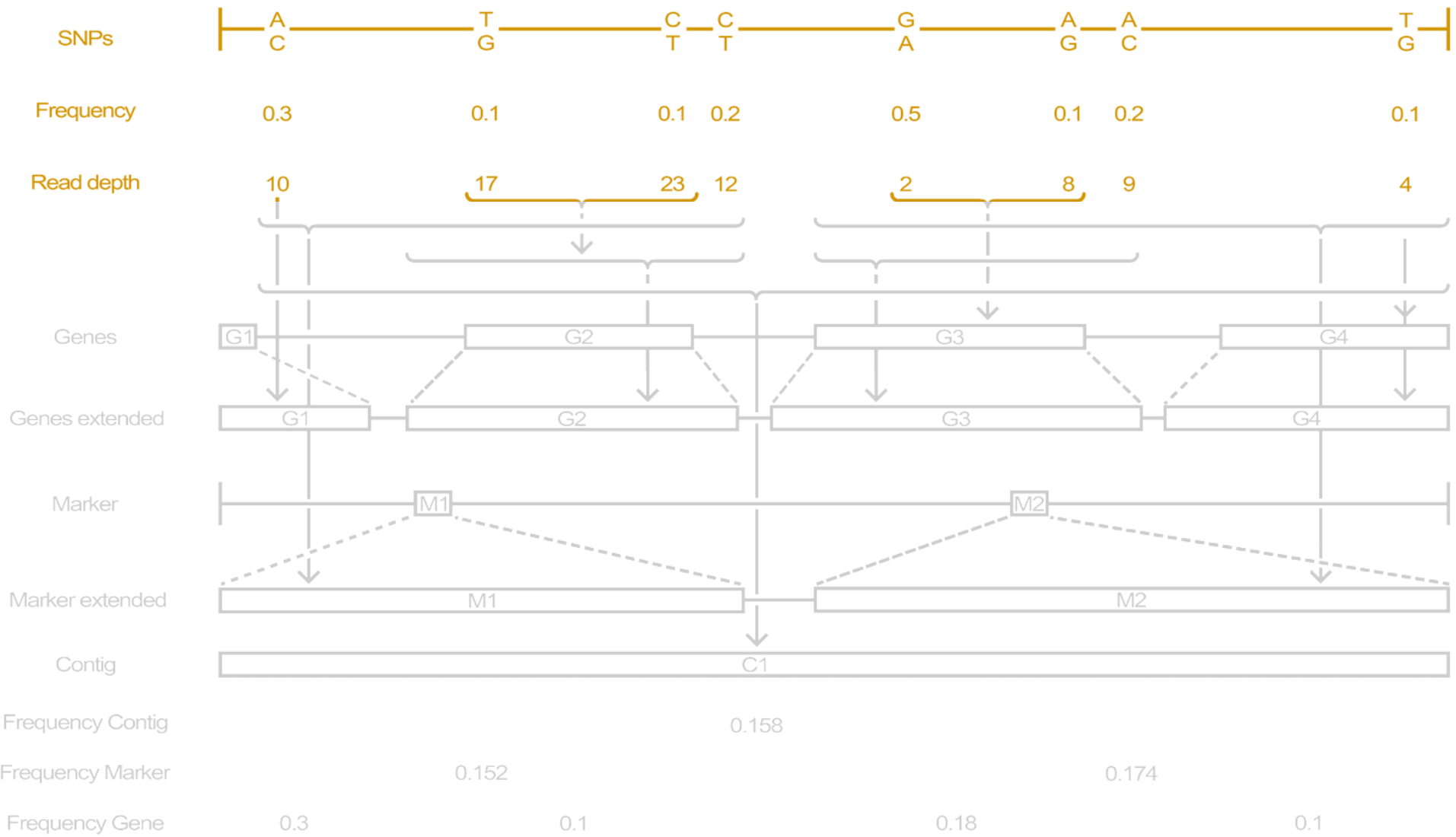
LiveSeeding

› Plant Methods. 2022 Mar 21;18(1):34. doi: 10.1186/s13007-022-00852-8.

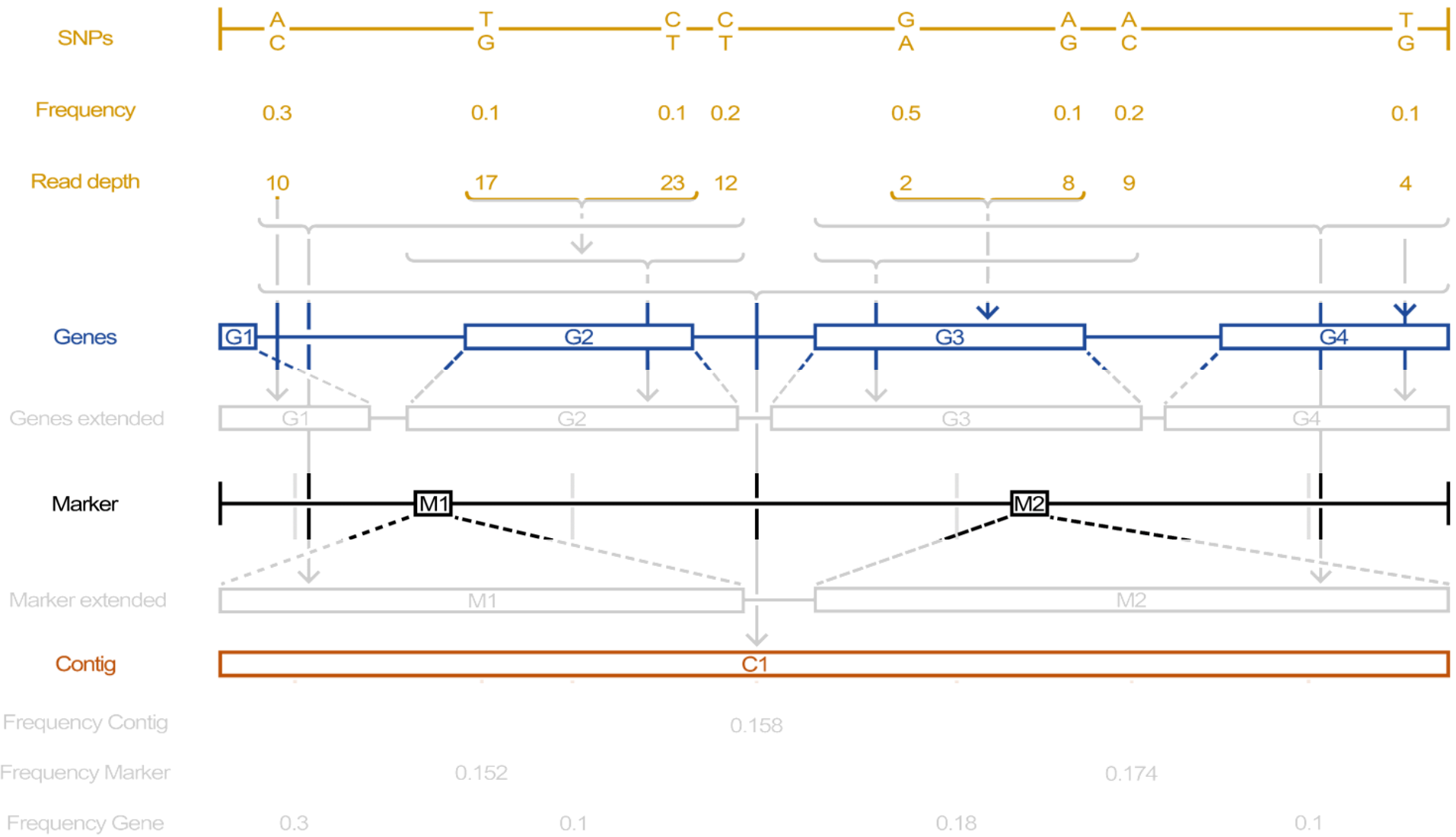
High-throughput estimation of allele frequencies using combined pooled-population sequencing and haplotype-based data processing

Michael Schneider [1 2], Asis Shrestha [1 2], Agim Ballvora [1], Jens Léon [3]

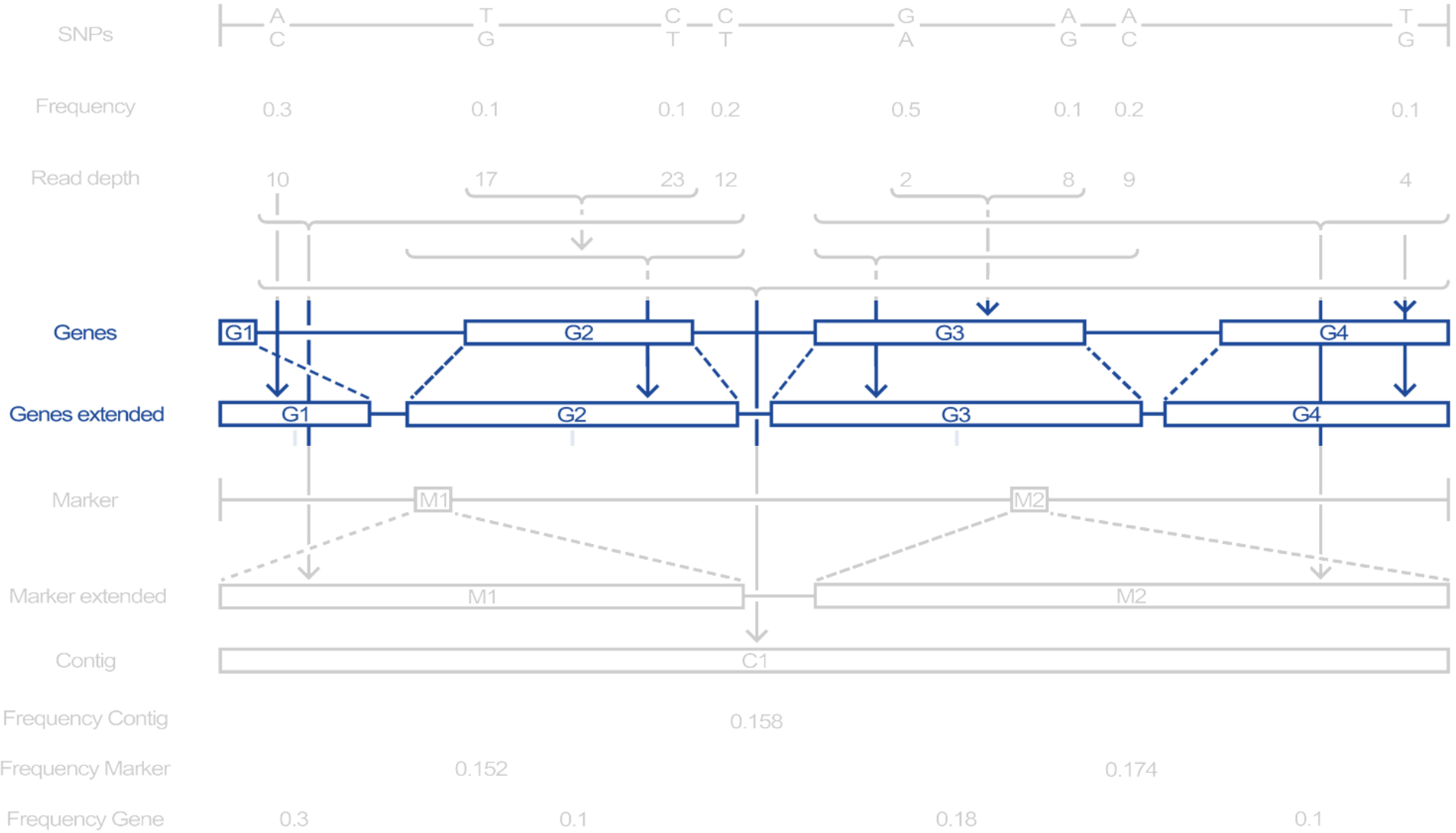
Strategies to increase power at low squencing depth



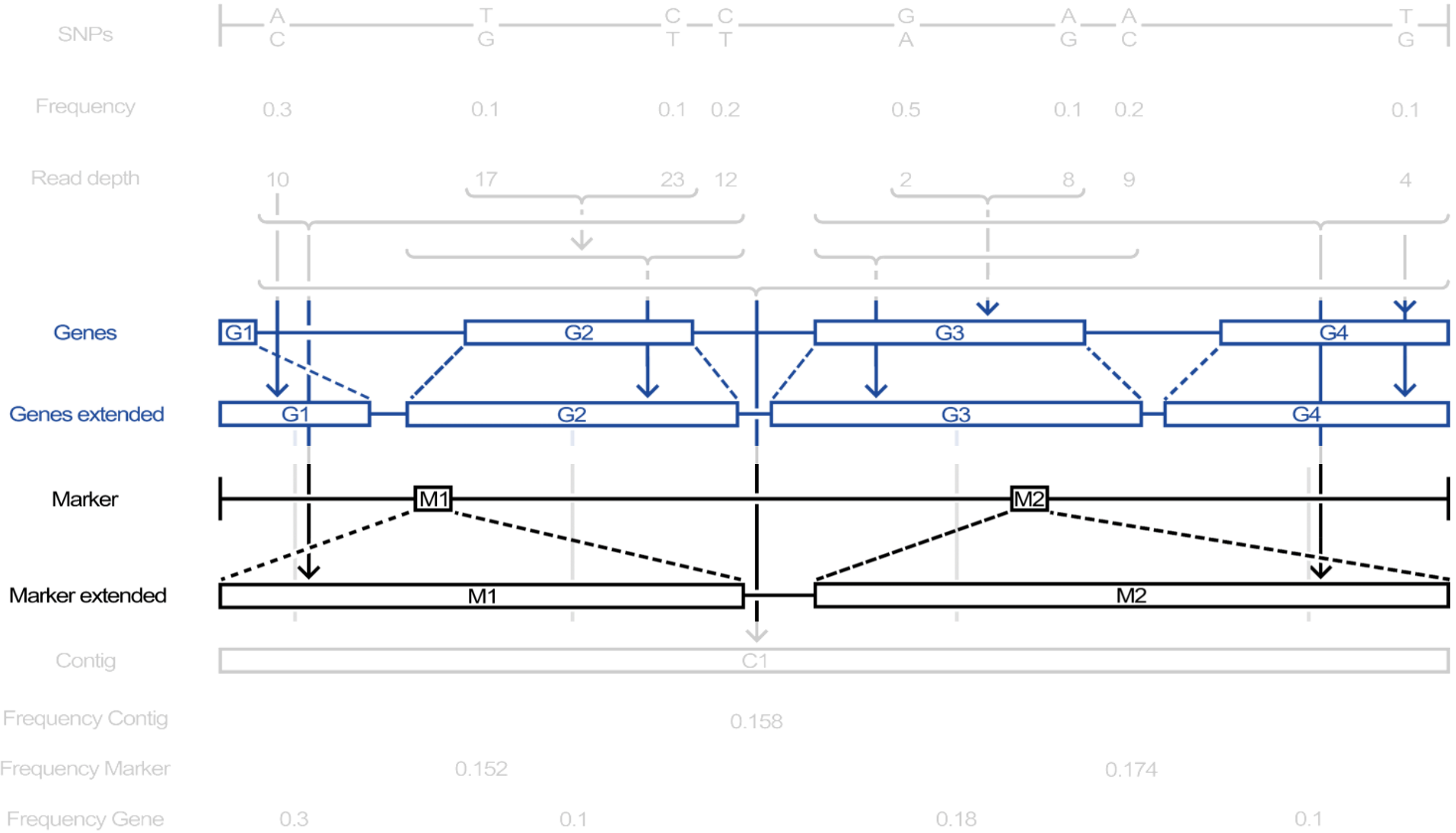
Strategies to increase power at low squencing depth



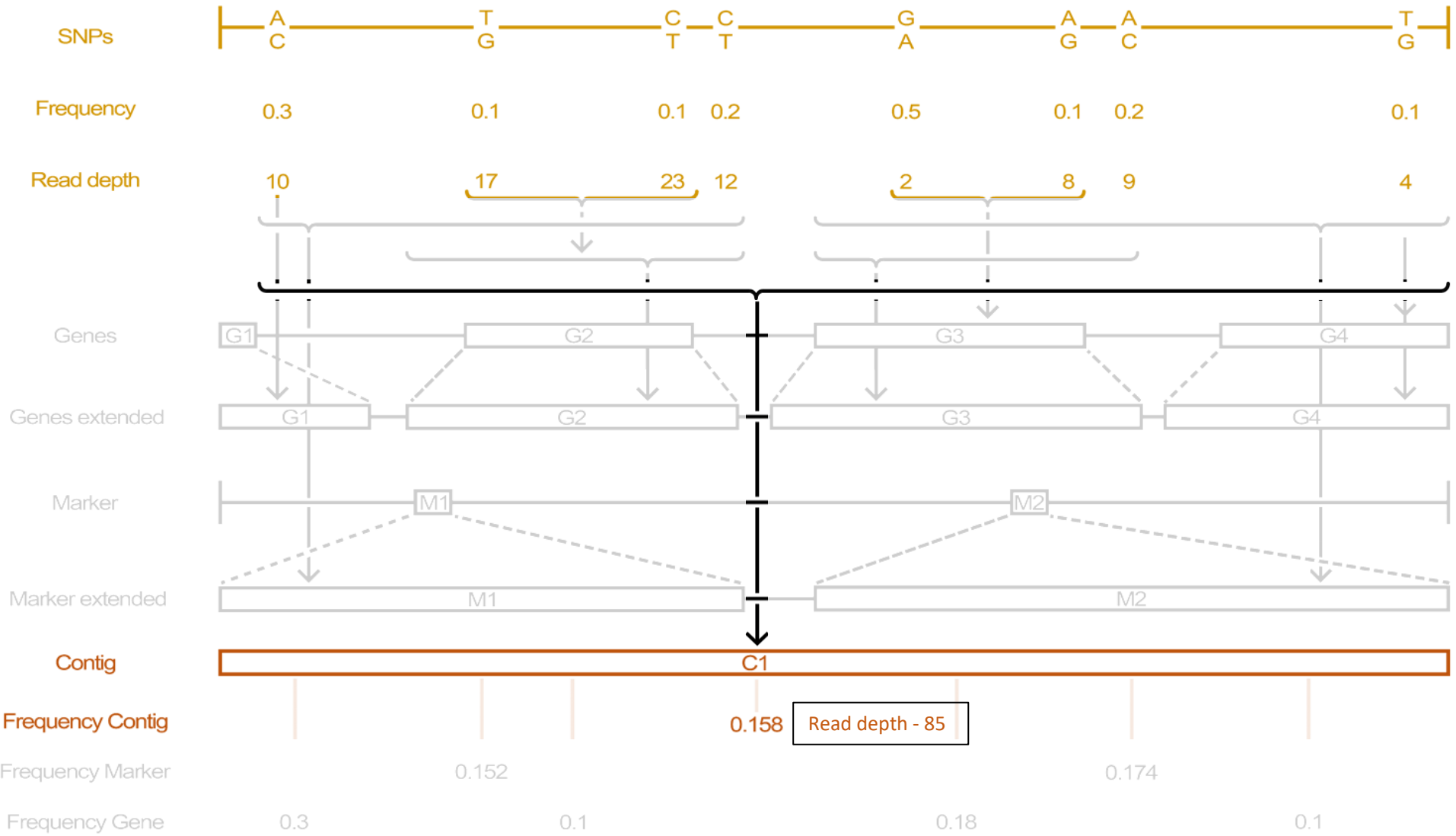
Strategies to increase power at low squencing depth



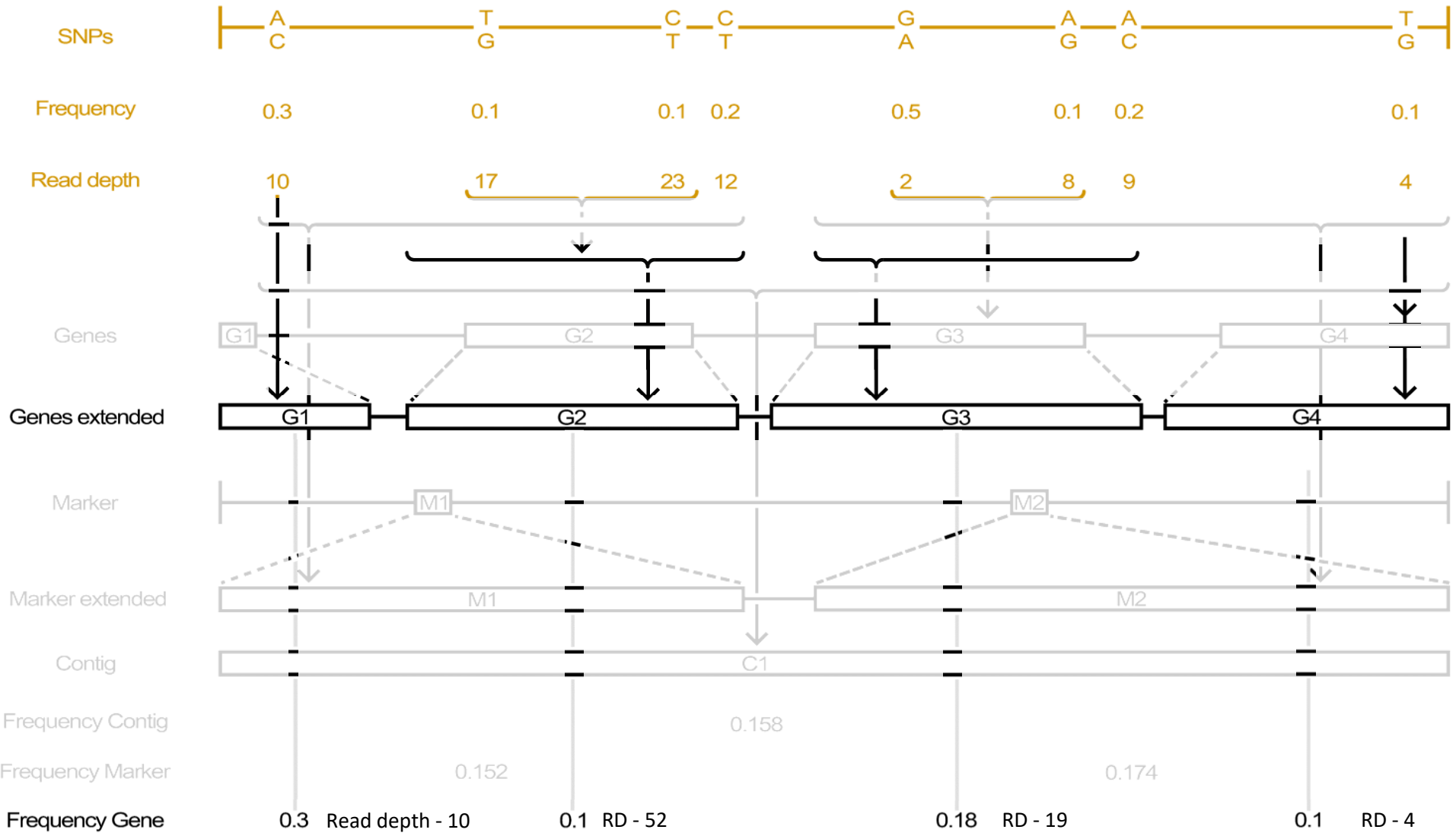
Strategies to increase power at low squencing depth



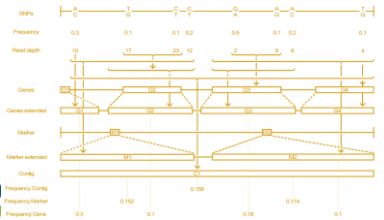
Strategies to increase power at low squencing depth



Strategies to increase power at low squencing depth



How does the theory convert into practice?



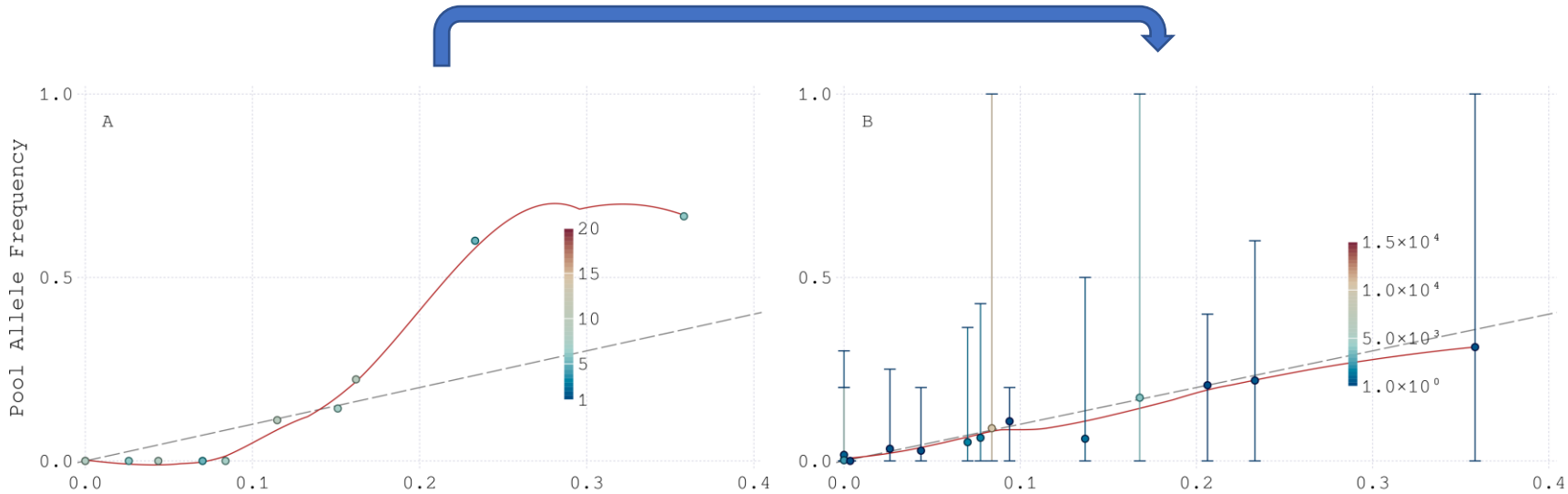
Comparing

Single SNP → Gene annotated haplotypes

288 genotypes, single and pool sequenced

KASP markers to pool sequencing

Using curated SNP databases to filter SNPs can improve the haplotype estimation accuracy



LiveSeeding

Summary

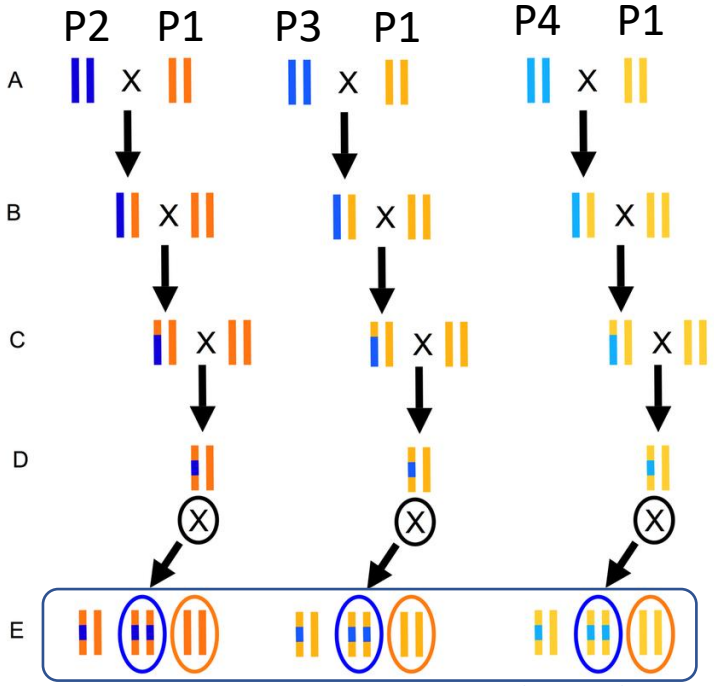
- Aggregating reads on the same recombination block can result in massive cost reductions compared to commonly nesseary sequencing depth (100-500x – cost reduction of 80-90%)
- The system works well if we have two parents which act as reference, so that the reads can be assigned to the respective parent

LiveSeeding

But how do we handle more complex systems?

Organic heterogenious material

4 parents crossed, 2 wild relatives

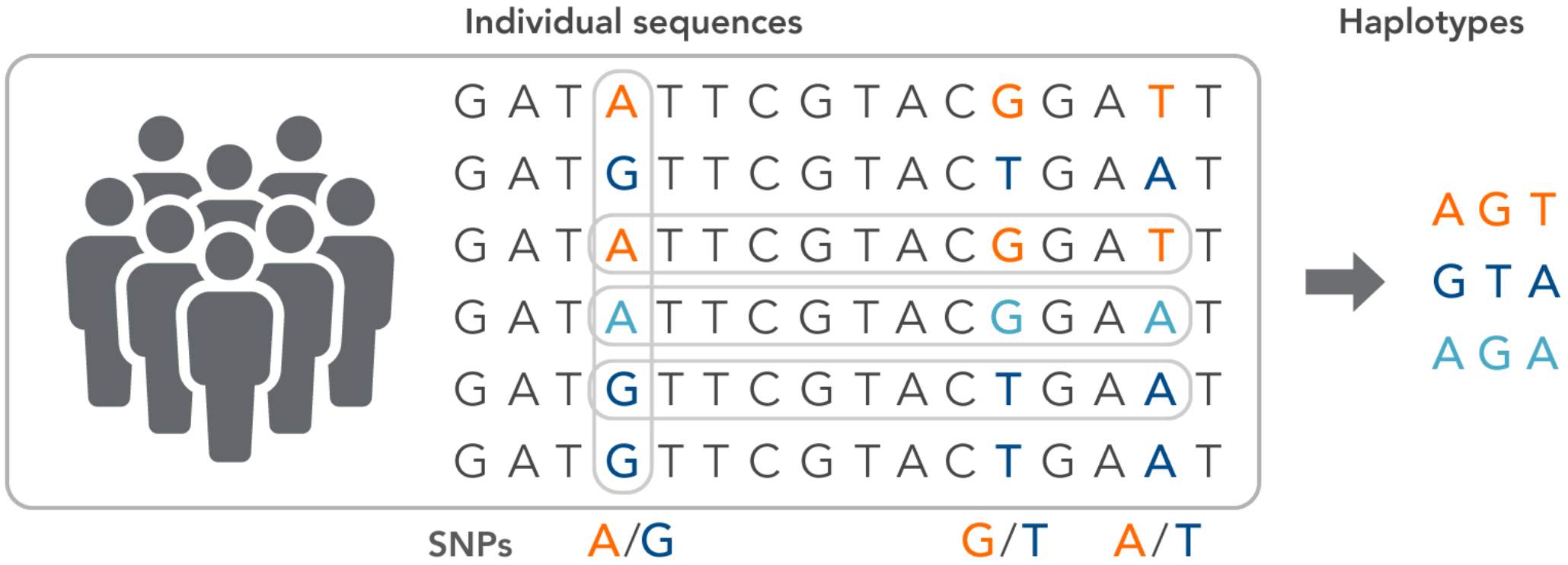


Modified figure from Lowry & Wills 2021; doi.org/10.1371/journal.pbio.1000500

Bulk and send to field

LiveSeeding

We got a problem..

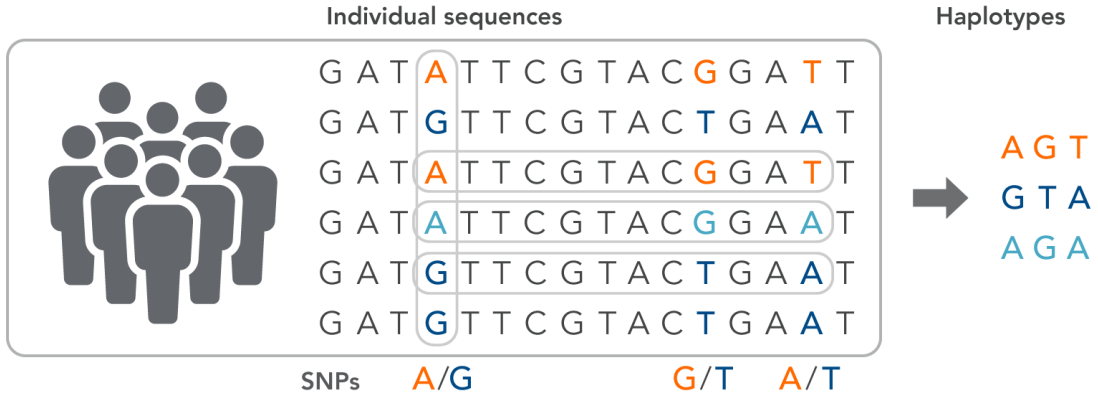


LiveSeeding

What can we do to overcome this dilemma?

- **Phasing of haplotypes**

  - **With a combination of SNP (haplotypes), we can differentiate the parents**
    - *ATA → P1, ATG → P2, CCG → P3*



LiveSeeding

We can achieve this with a long-molecule sequencing approach – PacBio or ONT

Advantages of long-read sequencing apporaches

1. **Phasing of haplotypes** 

   ➢ With a combination of SNP (haplotypes), we can differentiate the parents

2. **Methylation information** 

   ➢ short-term adaptation patterns related to the environment

3. **Microbe identification on long fragments** 

   ➢ Shotgun approach with long fragment might allow better classification and quantification (compared to 16S and ITS)

We cannot only get the genetic information, but also describe short-term adaptations (methylation) & the Mulitplication environment – describe the population as a result of the holobiont

LiveSeeding

How does the phasing to haplotypes work?

LiveSeeding

Phasing pipeline

01 Base calling

Dorado super accurate base calling

- min read length = 200
- quality cut off - phred = 12

02 Trimming reads

trimmomatic removal of

- first base pairs with low quality

03 Alignment

minimap2 alignment against the reference genome

- ont preset

04 Variant detection (~6.5 mio)

bcftools variant calling

- using a SNV data base to filter seq. artefacts
- *multi sample calling*
- extracting allelic depth
  - AD & filter QUAL > 40

05 Extraction of polimorphic loci (~4 mio)

identify relevant SNPs

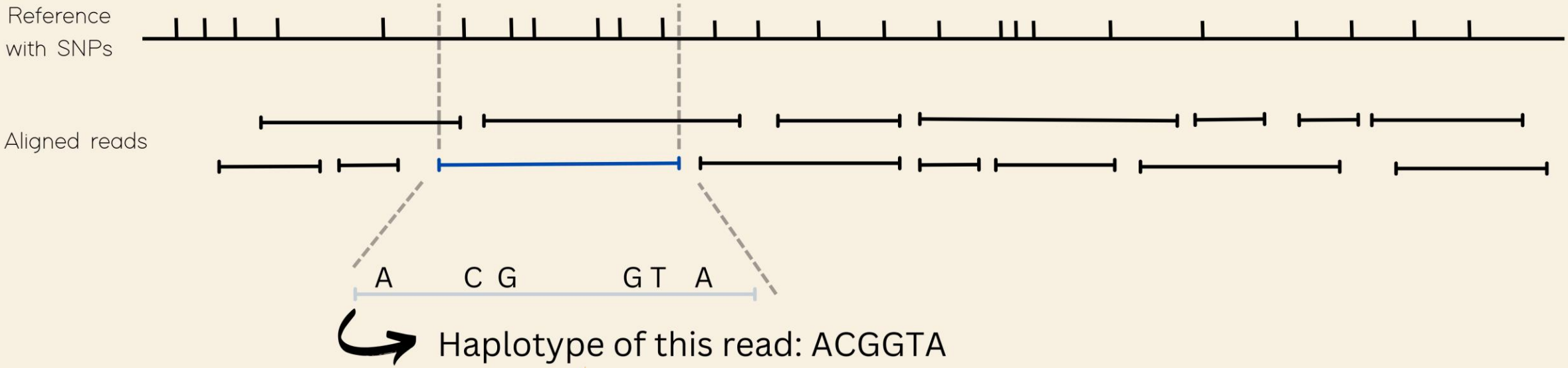
- dropping
  - homozygot loci
  - missing
  - more than 2 alleles

06 Phasing of single reads

- read SAM to work on every read
- using the CIGAR & seq to rebuild the alignment
- define the parental haplotypes
- Filter reads:
  - *63%* align. quality > 10 & no secondary align.
  - *61%* min length of 500 bp
  - *20%* drop reads with les than 3 SNPs
  - *20%* read quality > 15 (phred)

07 Haplotyping

- building contigs across the genome (500KB)
- count alignments overlapping contigs
  - count per parent in the population



Read String	FLAG Int64	Chr String	Start Int64	Qual Int64	CIGAR String	Seq String	Qid String	END Int64	SNPs Int64	alignRead String	alignQual String	Haplotype String
3444051c-0064-4193-a1bc-76f187c1…		chr1H	200008715	60	32S12M1D5M4D504M1D635M1D176M5D2M…	TATGTCCTCTACTCGTTCAGTTACGTATTGCT…	%%'(&''$$$$%'(+89@?::9;>@@AEFDD…	20001447	3	GAAGGTTGTTTG-AAGGA----TACAAAGGAA…	ECCJFSSKJIHB9EC65/9999-,,++,4BJP…	GGA
4481e85b-49d5-4164-918c-53721cfc…	1	chr1H	200012548	60	12S300M2D114M1I222M1I847M1D585M1…	GAATGTATTGCTTGTGCTTTAGCTTCCCCCTC…	%)++9;CB????JHFBCDCCISISKPJIGGLJ…	20002320	5	TGTGCTTTAGCTTCCCCCTCAGGGCTTCGTAC…	?HFBCDCCISISKPJIGGLJFKJMLGGJSSSH…	GAG
7924e052-c353-4b0f-b29b-b230d873…		chr1H	200012629	60	33S80M3D82M1I575M2D197M1D2M2I462…	GTTATGTTGGCTCCCTTCGTAGGTTGGTCTTG…	%&''''&$#######$$$$$$$%%%%%%$…	20001596	5	TGGATACCCAGAGGAGCAAGTAGGTGAAGCCG…	(((*35433339888A;;;<FLQRHSFSSIIS…	TGGGA
b620c372-ad83-4433-b5af-bd1ffa5f…	0	chr1H	200013001	60	33S663M1D461M3I218M1I2M1D297M1I2…	ATGTTGTGCCTGTTGCTCGTAGGTTGGTCTTG…	$$$&&&&&&&&'&%$%$#####$%''''''…	200015898	4	GGATTTAGTGCTCAATCTAGTTGGAGTTGTG…	9999@FECDHGGSNSMSNSJGHQHKHSSFHHH…	TGA
946b52d0-ced3-4da9-835b-3377c6be…	1	chr1H	200013700	60	55S312M1D681M2I11M1D243M1D94M1D61…	CAATGTATTGCTACTTTCACAGAGAGAGGATA…	*011CDKISJSKJGSLOJKJSJCIJJFSGDDJ…	200018360	4	ATTGCTACTTTCACAGAGAGAGGATAGGTCG…	DKISJSKJGSLOJKJSJCIJJFSGDDJGFHHI…	TGGA
e290a3e9-9463-45f9-8220-27f89644…		chr1H	200029155	60	40S285M1I3M2D18M1D52M2D552M1I364…	TTGTGTCTGCTTGTACTTCGTTCAGTTACGTA…	%'''(%%$$###$(**+**-11<><::9663…	20003541	4	AATCACAAATACAGGCCACATACTACAGGATG…	%%%%*,+-.-@@@AAFKJSNHHHB<96666=8…	CAA
b14bbf2f-956d-4029-b7f1-ca7b0708…		chr1H	200031226	60	39S27M1I78M1I17M1I129M1I15M1I176…	TTATGTTATGTGCCTGTACTTCGTTCAGTTAC…	$%&(())&&()&'''%%$%%('(+11221113…	20003584	5	TATGTATCTGAATTAGGGAAGGATATCTTTCC…	444++***12BA>>=;?@AKSSLPSPFD?4*;…	GAAC
d68109a3-d964-48db-b8b4-b1bd7297…	1	chr1H	200033046	60	226M1D506M1D258M1D6M1D208M2I3M1D…	TTGGATTGAGATGAAACTTGGAAATATGATCA…	(+34<;;;;;CDDEMCD213222136,((((-…	20004011	8	TTGGATTGAGATGAAACTTGGAAATATGATCA…	(+34<;;;;;CDDEMCD213222136,((((-…	CGTTCAAC
c1913875-04b8-4dc4-adea-5c680115…	0	chr1H	200033589	60	33S41M1I8M1D3M1D1M1D66M1I103M2D1…	TTGTTATGCTACTTGGTTACATTTAAGTATTG…	#$$$%%%%$###%%$$$%%%%%%#$&()*++…	20003756	8	TAAGGATGTTAGTTGGAGAATAGCAAGAATCA…	8::<9..-----0489ANIFLHIJGSEGE?>?…	AATCGTCG
0469dad8-f5be-4295-a988-874ec1dd…	16	chr1H	200033589	60	13S5M2I185M2D89M1I1M2I141M2D7M1I…	GGCCACCATTTGCTAAGGATATGTTAGTTGGA…	'))+00,)'&&*)),-+)**LGHSCHBBCBDI…	20003756	8	TAAGGATGTTAGTTGGAGAATAGCAAGAATCA…	),-+)LGHSCHBBCBDISHNGSSFMSKFPEHD…	GATCCTA
6bb0b2da-29d3-4dfd-a653-9b43941b…	0	chr1H	200051940	60	34S77M1I442M1I3M1D623M5D22M1I347…	ATGTTGCACATTTATTCGTTCAGTTACGTATT…	''',(%$####$%'''()2ACEFDD?BABACB…	20005646	5	GGAACAAACGCAAAAGAAACACAAAAAACAC…	CIDDSMKSSQGSIEHSSSDKHNS@@AHLSSHK…	AGAAA
ffedff3c-13c6-45c7-b457-8a131763…	0	chr1H	200052809	60	36S6M1D343M2D6M1D37M2I3M1D70M2D3…	ACTTTTATGTACTTACTCGTTCAGTTACGTAT…	#$%%&((''&&%$$$$%''--.99<>88420001…	20005694	5	ATCTC-TAAAGAGATTAGTAAGTGGATGAGG…	CCHJHD9B'''CGNHFQ>=====SHLKJLIIS…	AAGAA
5d55f6c0-122d-4808-a6f8-f3bf15aa…	1	chr1H	200066199	60	12S113M2D1M1D5M1D2M1D387M2I234M3…	CTATGTATTGCTTGAAGAATGGCTCTTCCTTG…	(+68BEJLSSHHJNMSGJJHSIISLMJGSFMI…	20007321	6	GAAGAATGGCTCTTCCTTGTCTTCATCTTTT…	JNMSGJJHSIISLMJGSFMIHSISLSIKSLJM…	TATTCA
a75f404b-f62e-52ea-aafd-1a660075…	0	chr1H	200069800	60	35S102M1D163M1I531M1I3M2I89M1I45…	TATGTGCATCTTACTTCGTTCAGTTACGTATT…	$%%''$$$%%%%''+0---..7766220002?…	20007630	13	GGCCCTTTTAAAAATCAGAACTAAAAGGGGC…	==>ASSDFA:4?ACKGIIEJJIJGEJEICEDA…	CTTAACTAGTTTAT
79d25f5b-e687-458b-b264-64a25811…	16	chr1H	200069824	60	31S44M1I226M1I17M1I5M1I98M1I235M…	ACATCACCATTGTGGCCTCGGAAACCAGATGA…	$%%%&''**))(),*'&&%%%%$$%(*+,*)0…	20007630	13	AAAGGGGCGATCAATGAATTGACCACGTGTAA…	0623:<;99999BA?>99<;<<<<GFSISSSI…	ATTCCATTTAT
a519e272-6857-4252-9a6a-30781b03…	0	chr1H	200070897	60	29S348M2D177M1I231M1D5M2D10M1D5M…	TTTGTCTGCTTGACTAGTTACGTATTGCTGAA…	$%%%%$$##%'&&'''&'(()3>>@IEGJGPS…	20007444	11	GAAACTAGTGTTCGATGGCCCAACAAATCCTA…	GPSGIHKGFIGH22EGKSLSSSHKSPPOJHJS…	GCTACTTATTA
bafeb1b6-d9bd-461d-829c-530a3be6…	0	chr1H	200072049	60	34S14M1D14M1D40M1D20M1D10M1D80M2…	ATGTGCCTGTTACTTGGTTCAGTTAGGTATTG…	$$$%%''++,-*))**,-..0/./2444335)…	20007503	11	TATACCTCAGCTTG-TCGATTGAGCCGAT-CC…	))))>ACA:9889:697,++--,---.//-9-*…	A-CGTTCA
19c51eec-ef59-4f2c-9706-a73708f4…	1	chr1H	200072362	60	17S18M4D65M1D12M1I10M1I5M4D4M1I1…	CATTGCCATGTATTGCTAGCAAATAAAATGAA…	'''&%%&**68DDFEHFII=@@A?@DCCCBEE…	20007543	11	AGCAAATAAAATGAAAAG----CTCTCTCGAA…	II=@@A?@DCCCBEECDE9999D;6--+*+.(…	GTCCACTAT
19c4c767-3d81-4623-bc57-6e3124af…		chr1H	200072562	60	40S94M1D113M1I86M1I199M1D1M1D261…	GTTGTGTAGCATCTGTCTTCAGACTCAGGTTG…	%'*+++*&%%%%%%&''')'&%%%%%&('(*…	20007981	19	ACAGTGGATGTGCAGGAGCAAGCGCTGATGCT…	2222269>B??;;655;;;;=>@@DJDIOSML…	TGCATTTTCAAACTATTT

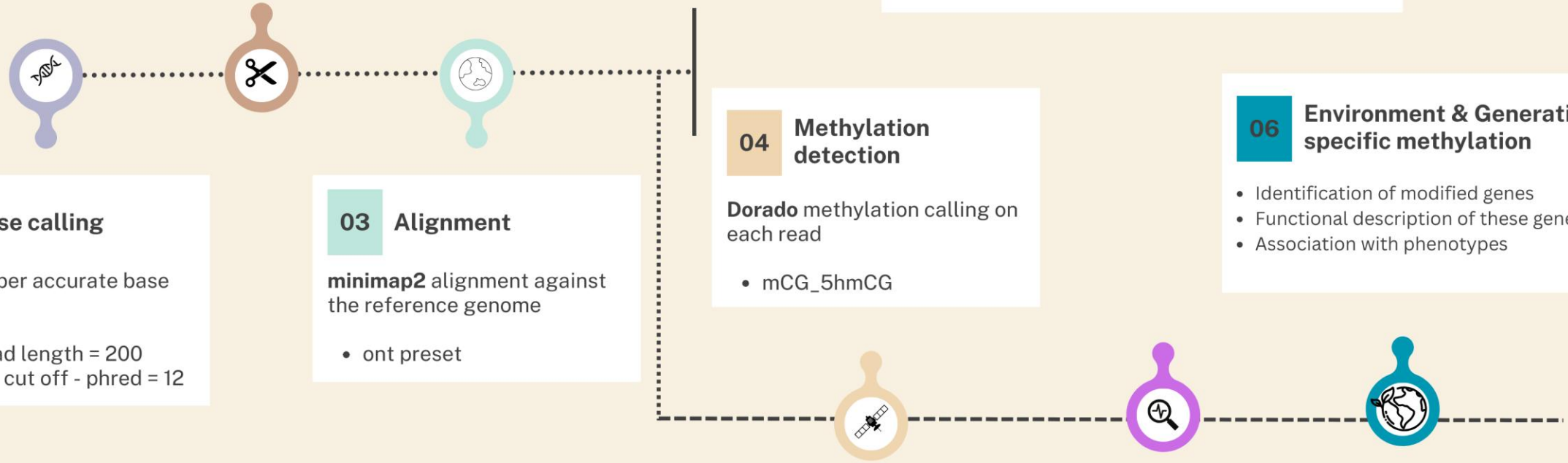
1. Align to reference genome

2. Identify SNPs in read

3. Extract SNPs

Methylation pipeline

(Epigenetic variations)



01 Base calling

Dorado super accurate base calling

- min read length = 200
- quality cut off - phred = 12

02 Trimming reads

trimmomatic removal of

- first base pairs with low quality

03 Alignment

minimap2 alignment against the reference genome

- ont preset

04 Methylation detection

Dorado methylation calling on each read

- mCG_5hmCG

05 Methylation pattern extraction and comparison

Modkit *pileup* and differntial methylation detection *dmr*

- *collection of methylated bases*
- *pairwise comparison (or multi)*
- *in annotated regions (5kb gene upstream)*

06 Environment & Generation specific methylation

- Identification of modified genes
- Functional description of these genes
- Association with phenotypes

Haplotyping Pipeline - parents unknown

01 Base calling

Dorado super accurate base calling

- min read length = 200
- quality cut off - phred = 12

02 Trimming reads

trimmomatic removal of

- first base pairs with low quality

03 Alignment

minimap2 alignment against the reference genome

- ont preset

04 Variant detection (~6.5 mio)

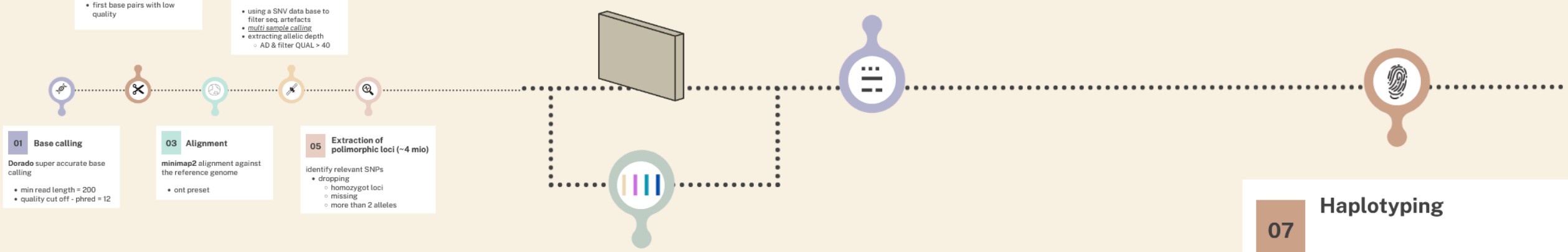
bcftools variant calling

- using a SNV data base to filter seq. artefacts
- *multi sample calling*
- extracting allelic depth
  - AD & filter QUAL > 40

05 Extraction of polimorphic loci (~4 mio)

identify relevant SNPs

- dropping
  - homozygot loci
  - missing
  - more than 2 alleles



06.1 Construct reference Haplotypes

Parents unknown - no natural reference

1. Concat ONT reads in long contigs
   - Using the SNP profile to differntiate phases
2. Aggregate contigs to chromosome-scale haplotypes
   - Match making based on:

   PCA + Kmeans clustering + Linkage disseq.
3. Reference haplotypes generated

06.2 Phasing of single reads

- read SAM to work on every read
- using the CIGAR & seq to rebuild the alignment
- track & define the denovo estimated haplotypes
- Filter reads

07 Haplotyping

- building contigs across the genome (500KB)
- count alignments overlapping contigs
  - count per parent in the population

So finally, what information can be obtained from the genotyping?

(1) Are the populations different? (Statistcally significant)

- **Ultra low (short-read) sequencing**

(2) What has changed?

- **Which regions are different?**
- **Low low sequencing**

(3) Which traits were subject of frequency changes?

- **Gene level differences (& Methylation patterns[2] & microbiome variations[2])**
- **Medium low sequencing**

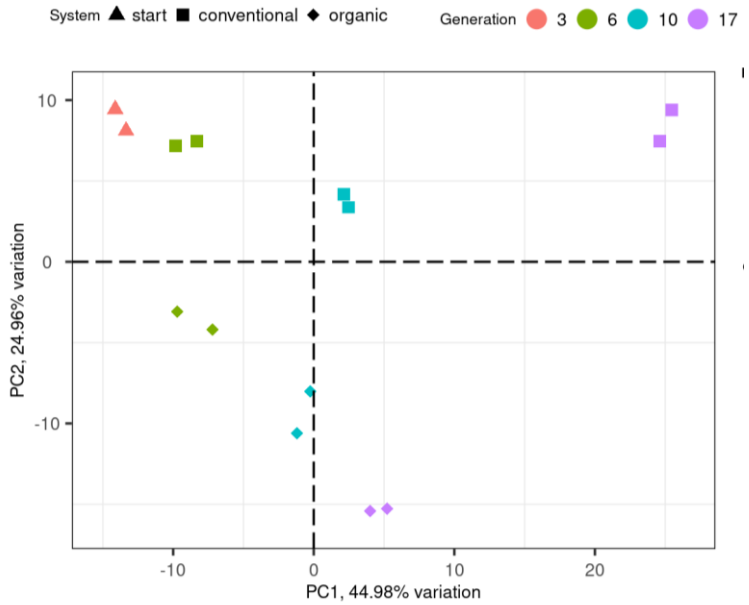
[2] - only available using the latest Oxford Nanopore Techique

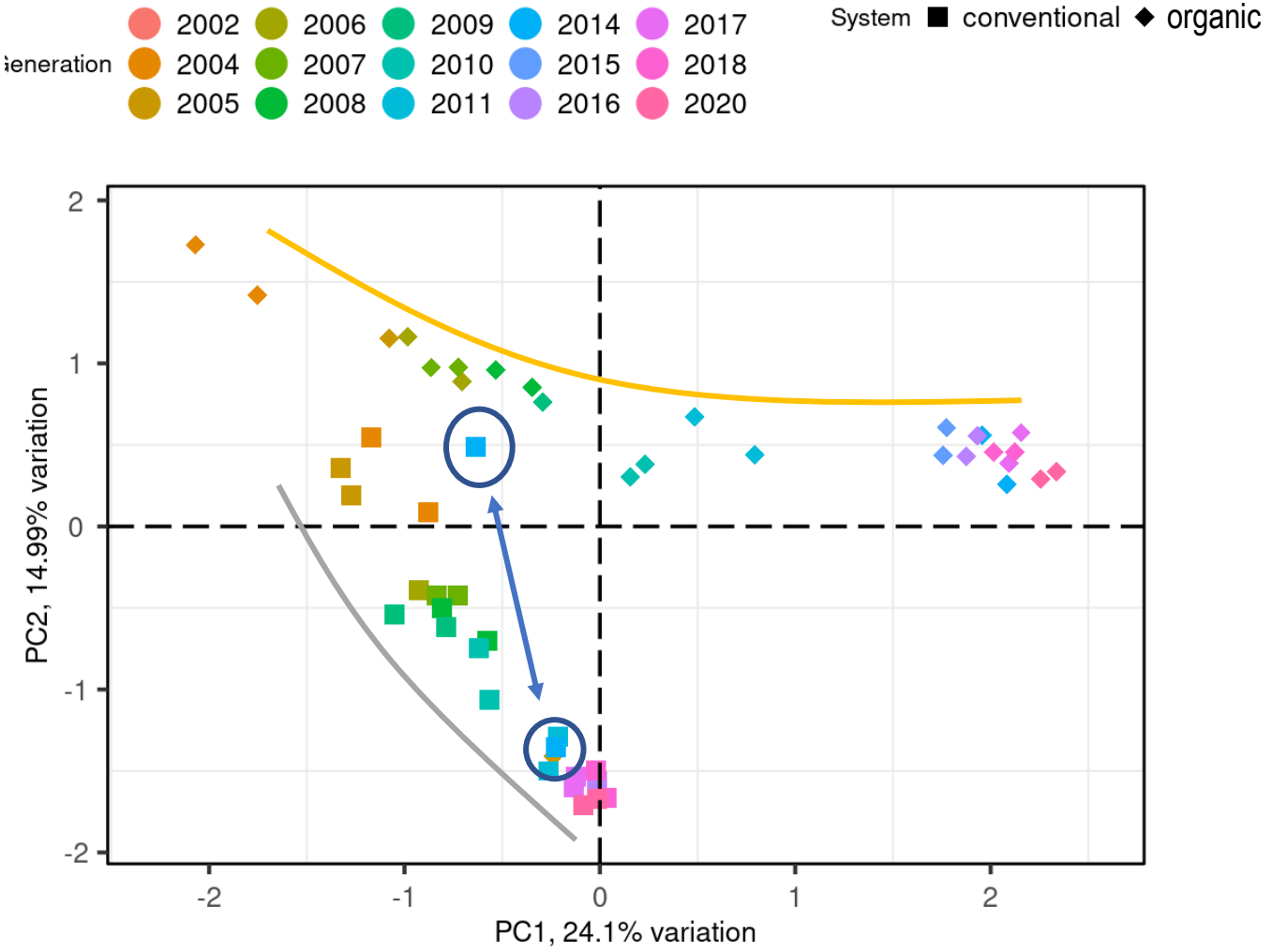
LiveSeeding

1 - Lowest level - are the populations different?

Low

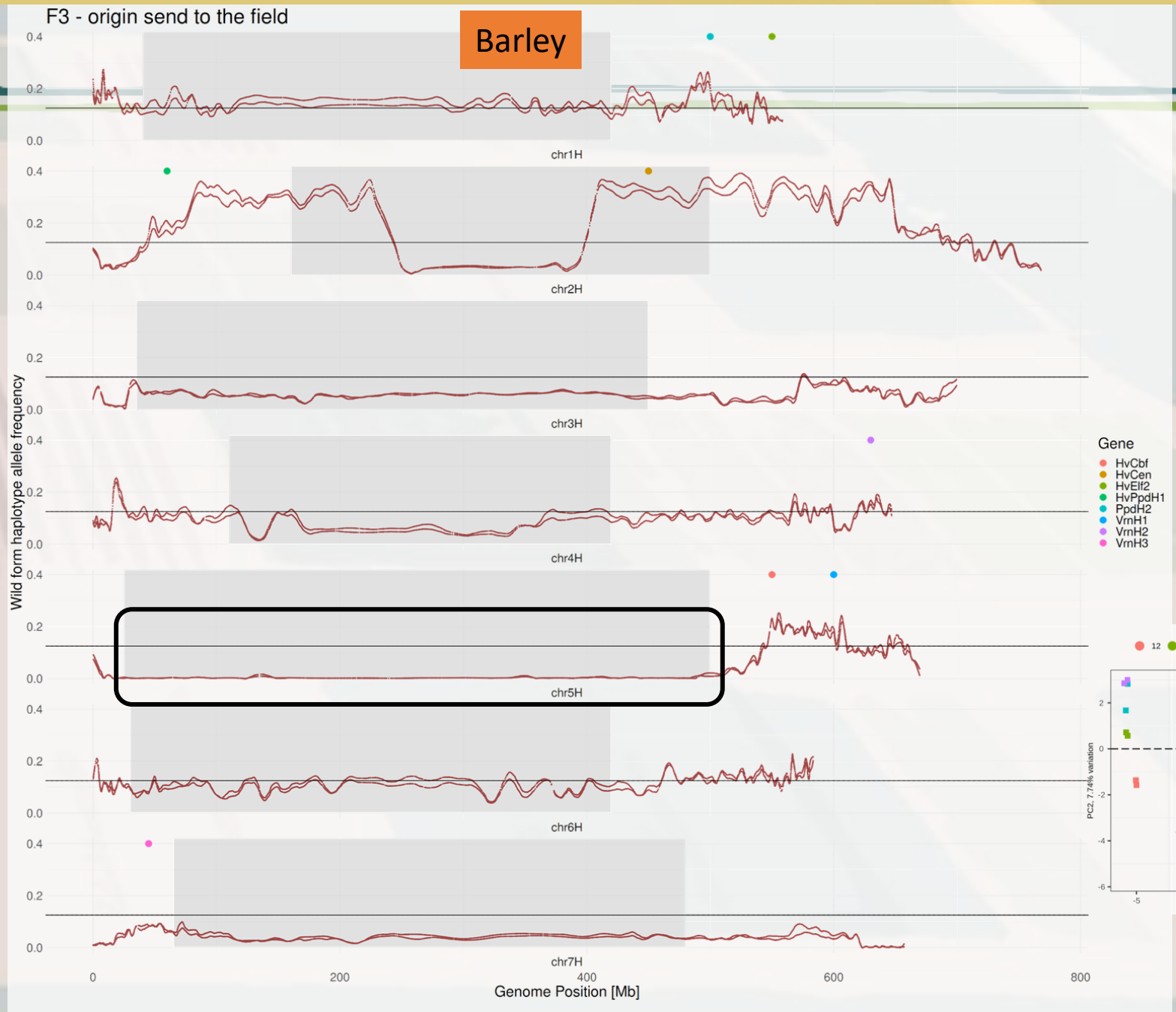
- **What is required?**
  - **Lowest sequencing depth**
  - **Any genotyping method**

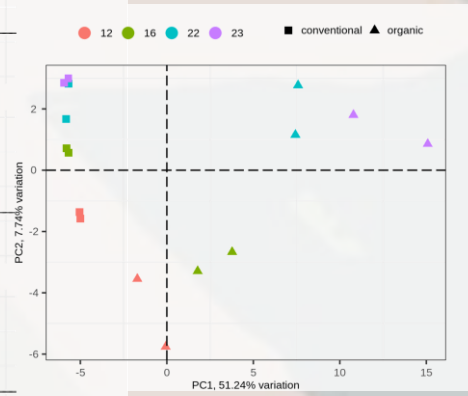




2- Intermediate level – approximate chromosomal regions under change

Mid





Mid

F12

Organic
Conventional
Origin (F3)

chr1H
chr2H
chr3H
chr4H
chr5H
chr6H
chr7H

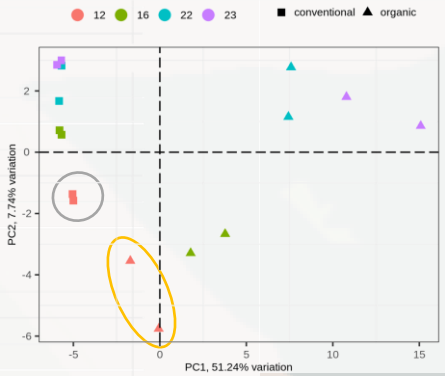
Wild form haplotype allele frequency

0.0
0.2
0.4

0
200
400
600
800

Genome Position [Mb]

Gene
HvCbf
HvCen
HvElf2
HvPpdH1
PpdH2
VrnH1
VrnH2
VrnH3



Mid

F16

Organic
Conventional
Origin (F3)

Wild form haplotype allele frequency

chr1H
chr2H
chr3H
chr4H
chr5H
chr6H
chr7H

Genome Position [Mb]

Gene
HvCbf
HvCen
HvElf2
HvPpdH1
PpdH2
VrnH1
VrnH2
VrnH3

12 16 22 23
conventional organic

PC2, 7.74% variation
PC1, 51.24% variation

Mid

F22

Organic
Conventional
Origin (F3)

chr1H
chr2H
chr3H
chr4H
chr5H
chr6H
chr7H

Wild form haplotype allele frequency

Genome Position [Mb]

Gene
HvCbf
HvCen
HvElf2
HvPpdH1
PpdH2
VrnH1
VrnH2
VrnH3

12 16 22 23
conventional organic

PC2, 7.74% variation
PC1, 51.24% variation

Mid

F23

Organic
Conventional
Origin (F3)

Wild form haplotype allele frequency

chr1H
chr2H
chr3H
chr4H
chr5H
chr6H
chr7H

Genome Position [Mb]

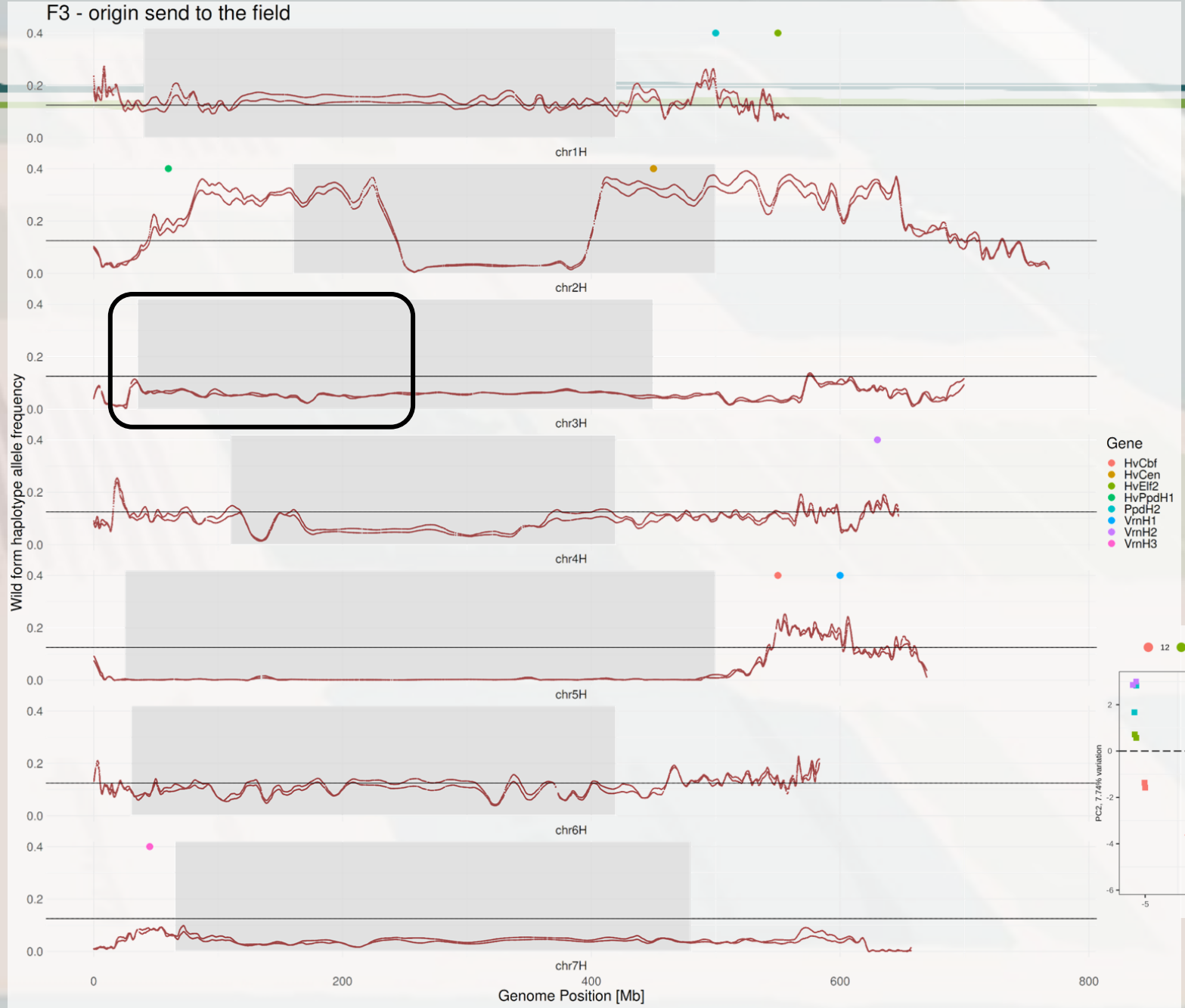
Gene
HvCbf
HvCen
HvElf2
HvPpdH1
PpdH2
VrnH1
VrnH2
VrnH3

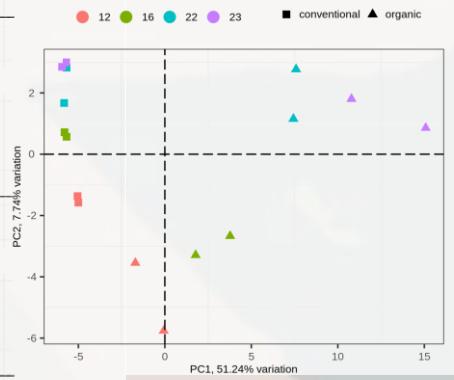
12 16 22 23
conventional organic

PC2, 7.74% variation
PC1, 51.24% variation

Mid

F3 - origin send to the field





Mid

F12

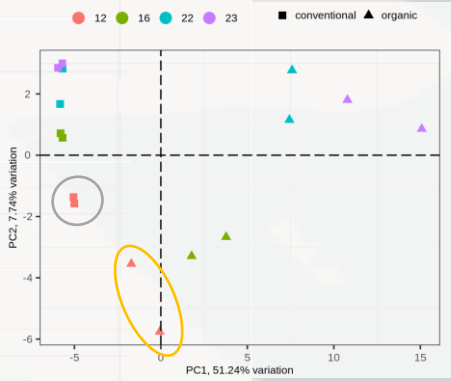
Organic
Conventional
Origin (F3)

Wild form haplotype allele frequency

chr1H
chr2H
chr3H
chr4H
chr5H
chr6H
chr7H

Genome Position [Mb]

Gene
HvCbf
HvCen
HvElf2
HvPpdH1
PpdH2
VrnH1
VrnH2
VrnH3



Mid

F16

Organic
Conventional
Origin (F3)

chr1H
chr2H
chr3H
chr4H
chr5H
chr6H
chr7H

Wild form haplotype allele frequency

0.0
0.2
0.4

0
200
400
600
800

Genome Position [Mb]

Gene
HvCbf
HvCen
HvElf2
HvPpdH1
PpdH2
VrnH1
VrnH2
VrnH3

12 16 22 23
conventional organic

PC2, 7.74% variation
PC1, 51.24% variation

Mid

F22

Organic
Conventional
Origin (F3)

chr1H
chr2H
chr3H
chr4H
chr5H
chr6H
chr7H

Wild form haplotype allele frequency

Genome Position [Mb]

Gene
HvCbf
HvCen
HvElf2
HvPpdH1
PpdH2
VrnH1
VrnH2
VrnH3

12 16 22 23 conventional organic

PC2, 7.74% variation
PC1, 51.24% variation

Mid

F23

Organic
Conventional
Origin (F3)

chr1H
chr2H
chr3H
chr4H
chr5H
chr6H
chr7H

Wild form haplotype allele frequency

Genome Position [Mb]

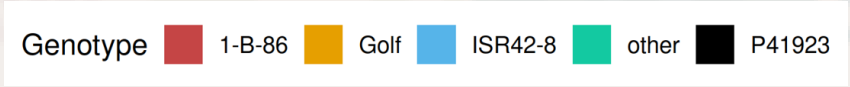
Gene
HvCbf
HvCen
HvElf2
HvPpdH1
PpdH2
VrnH1
VrnH2
VrnH3

12 16 22 23
conventional organic

PC2, 7.74% variation
PC1, 51.24% variation

How does it looks like in a more complex population?

Mid



BC2F3 (1998)

Natural adaptation for 25 years

BC2F28 (2023) [organic]

LiveSeeding

Home › Agronomy for Sustainable Development › Article

Deep genotyping reveals specific adaptation footprints of conventional and organic farming in barley populations—an evolutionary plant breeding approach

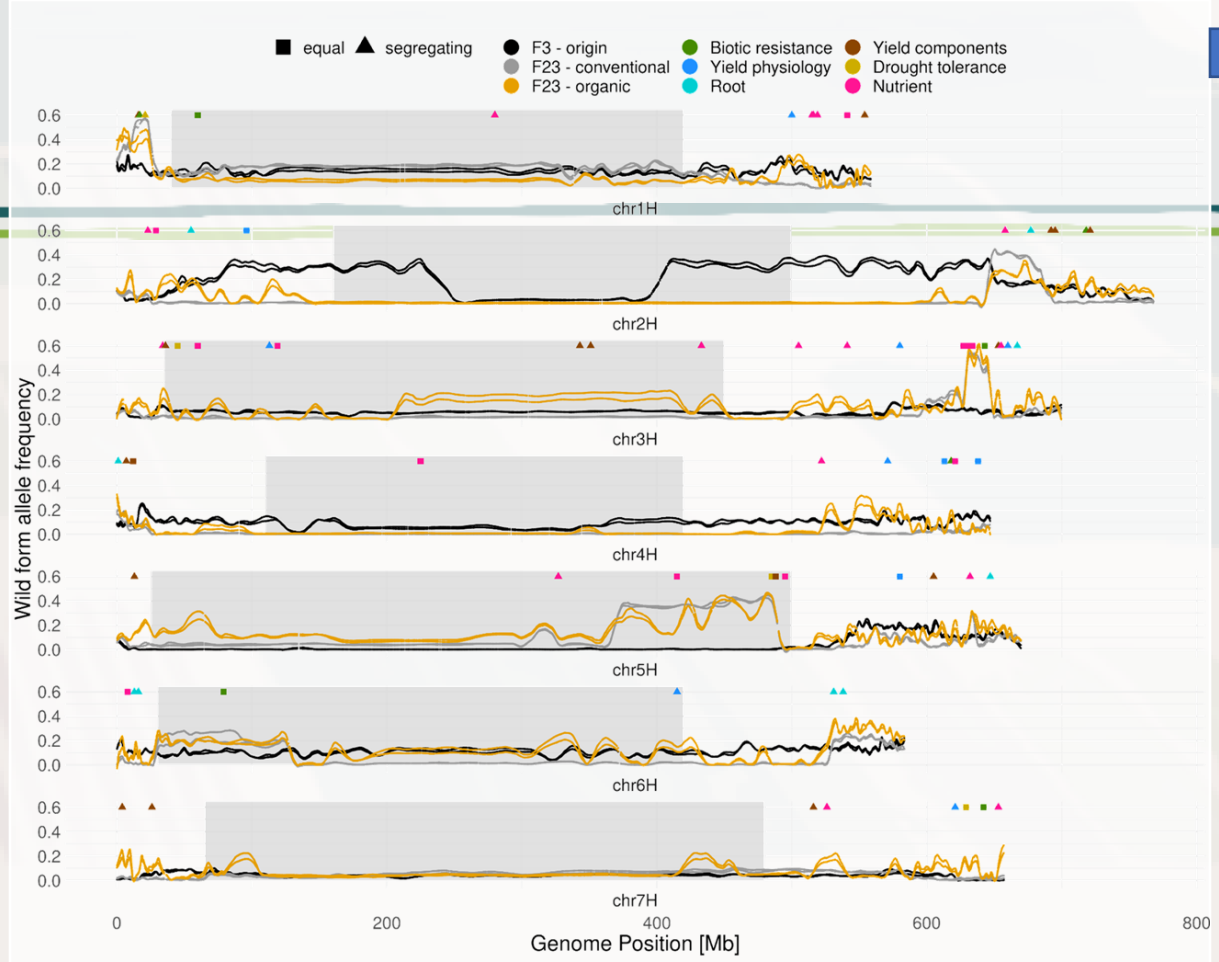
Research Article | Open access | Published: 08 May 2024
Volume 44, article number 33, (2024) Cite this article

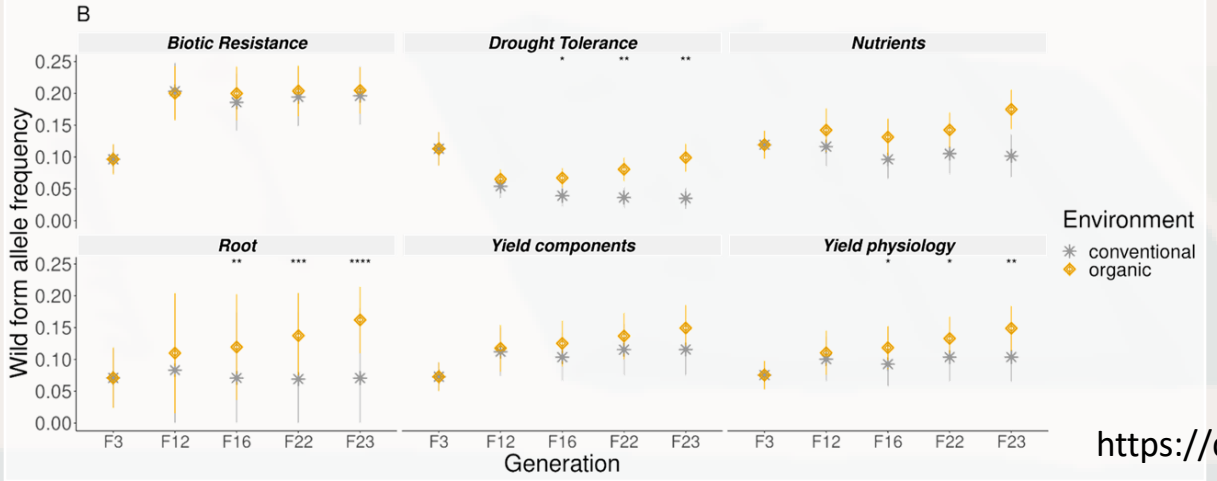
3- Highest level – identify candiate genes causing selection

High

https://doi.org/10.1007/s13593-024-00962-8

High



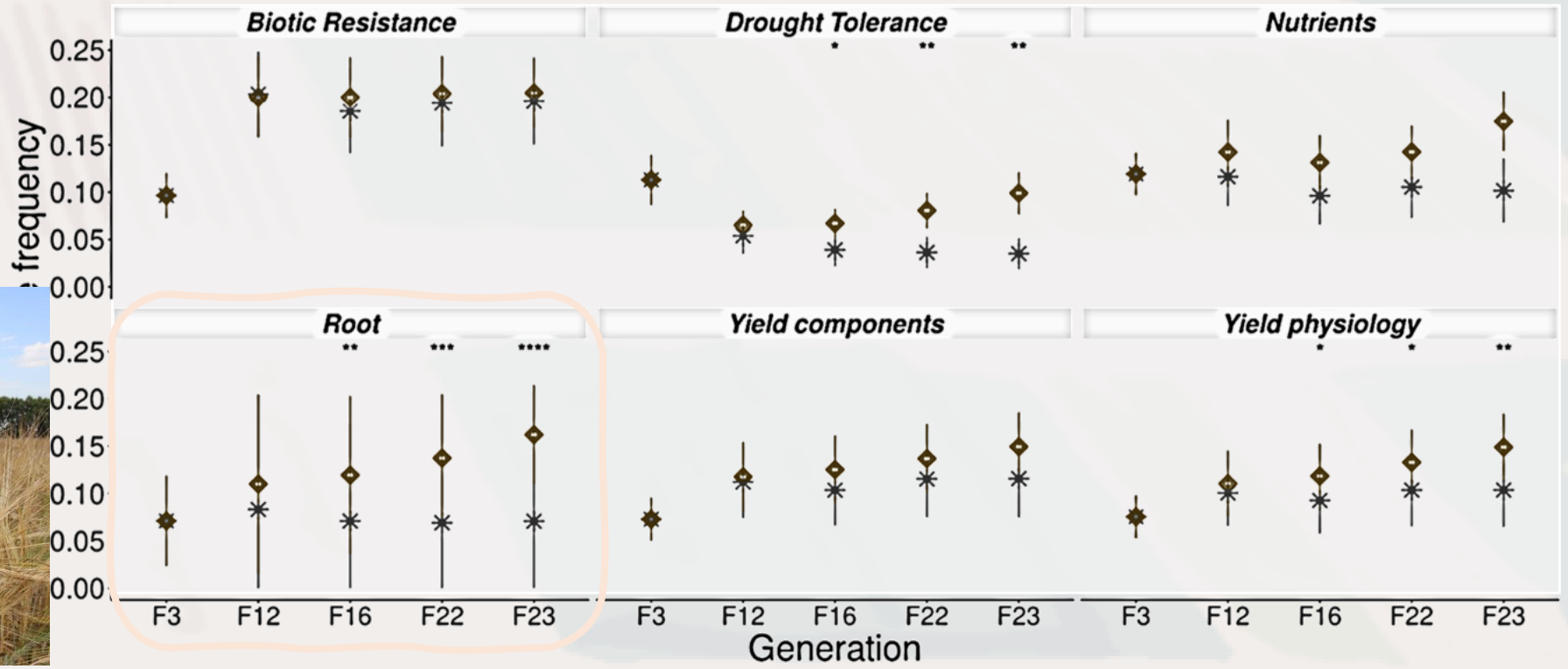


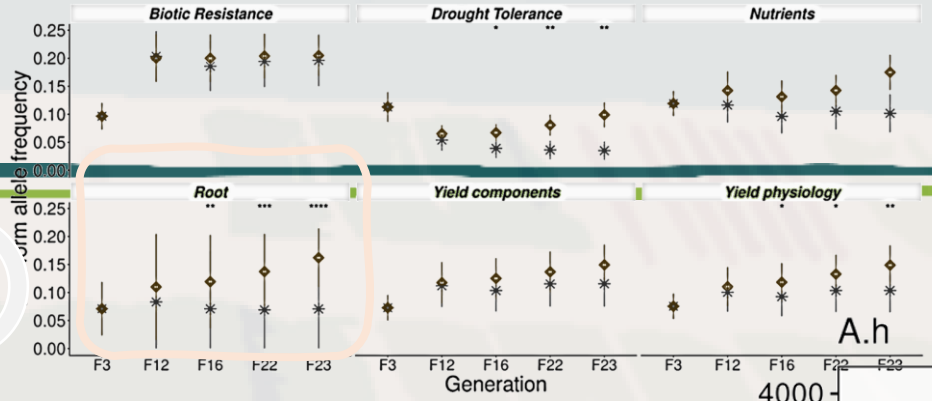
https://doi.org/10.1007/s13593-024-00962-8

3- Highest level – identify candiate genes causing selection

High

Converting the Frequencies into phenotypic characterisitcs



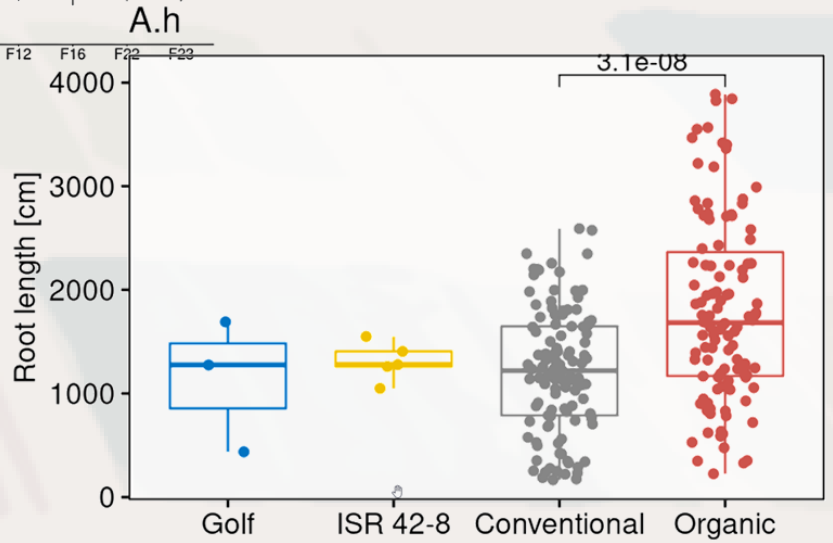


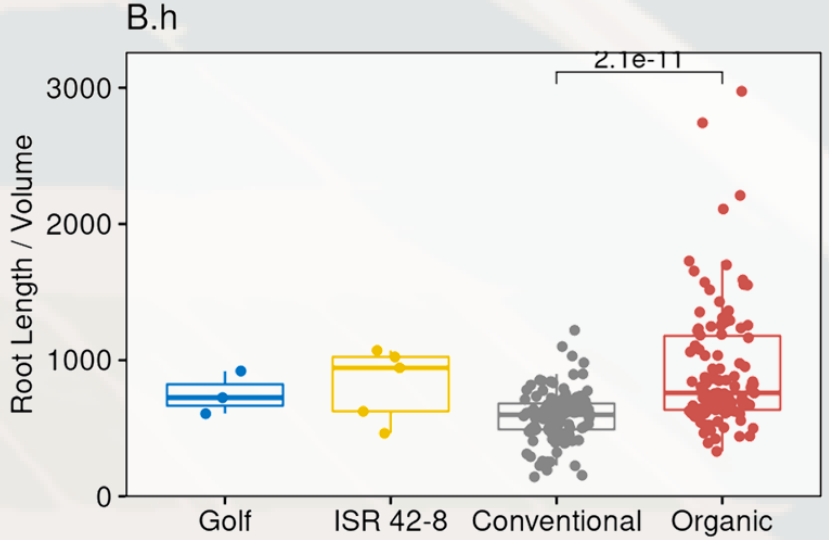
High

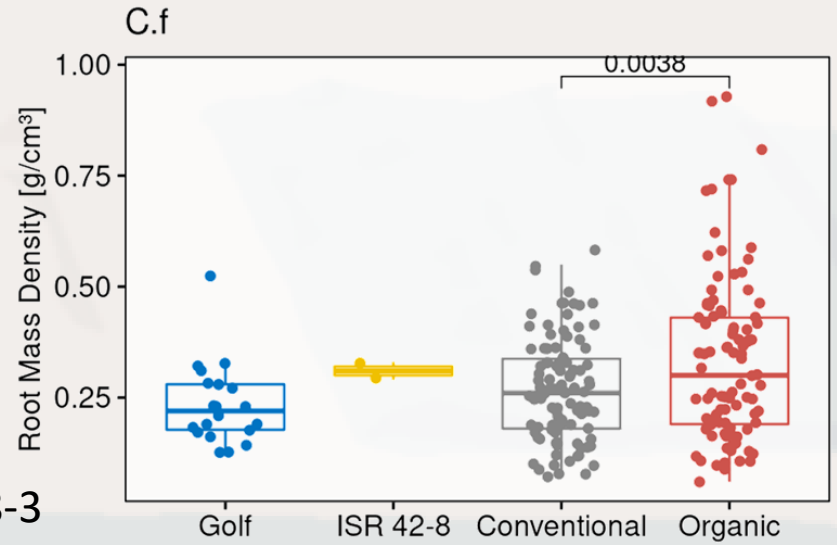
Parents

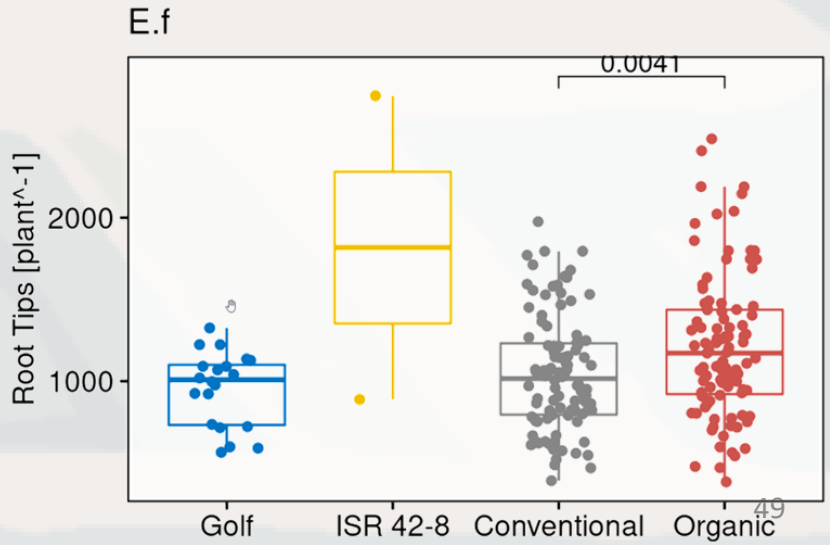
BC2F24

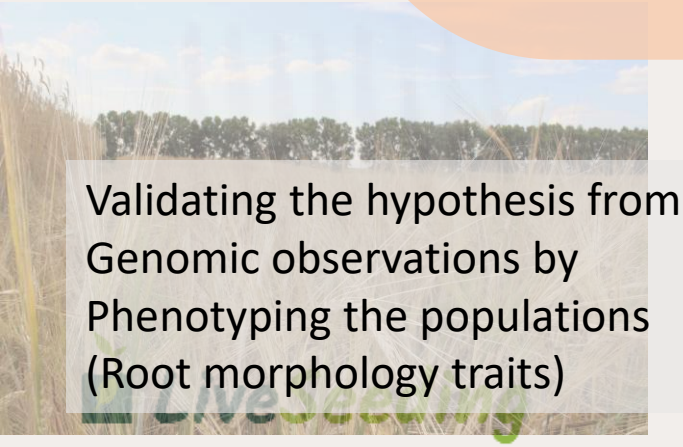
Golf ISR 42-8 Conventional Organic









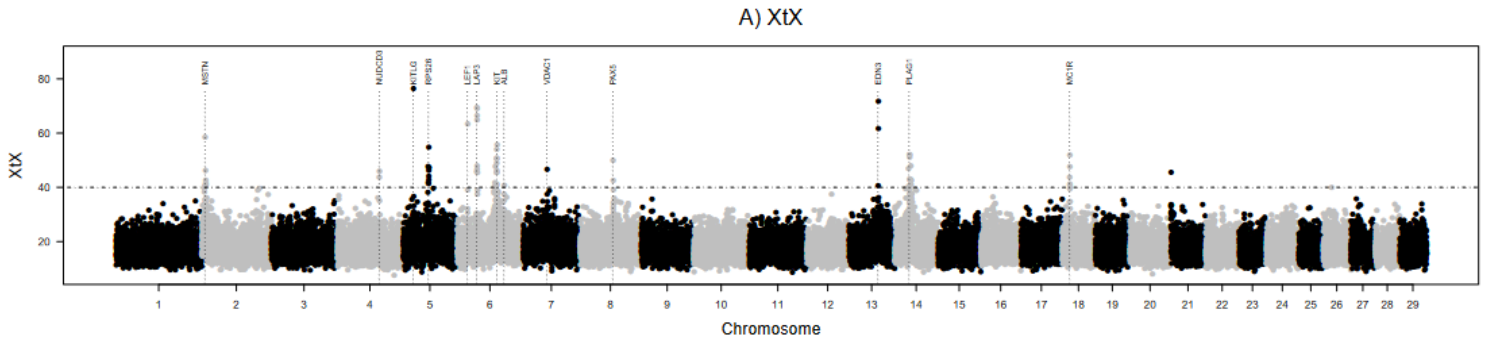


Validating the hypothesis from the Genomic observations by Phenotyping the populations (Root morphology traits)

https://doi.org/10.1038/s41598-022-23298-3

Interpreting allele frequencies is uncomfortable -

Can I have a p(like)-value?



GENETICS

Genetics. 2015 Dec; 201(4): 1555–1579.
Published online 2015 Oct 19. doi: 10.1534/genetics.115.181453
PMCID: PMC4676524
PMID: 26482796

Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates

Mathieu Gautier[1]

Associated Data

▸ Supplementary Materials

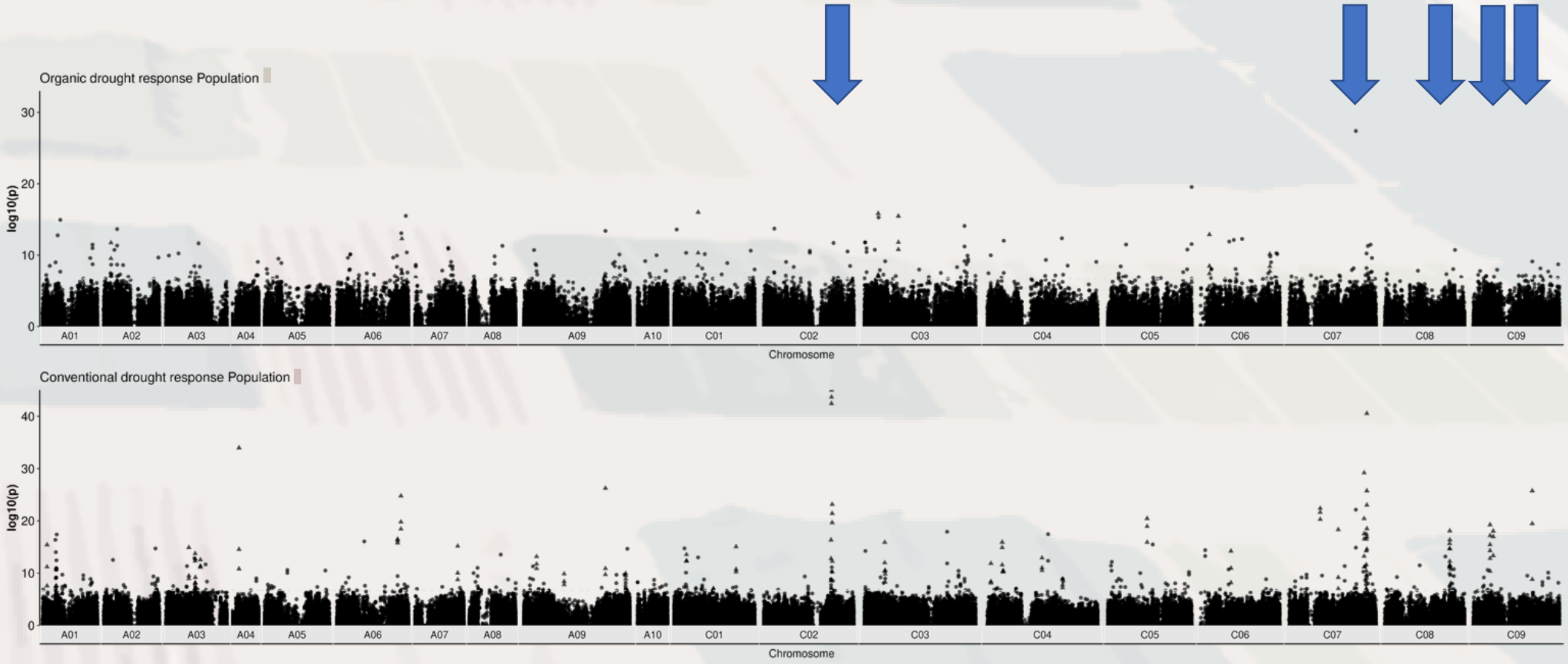
Abstract Go to: ▸

In population genomics studies, accounting for the neutral covariance structure across population allele frequencies is critical to improve the robustness of genome-wide scan approaches. Elaborating on the BayEnv model, this study investigates several modeling extensions (i) to improve the estimation accuracy of the population covariance matrix and all the related measures, (ii) to identify significantly overly differentiated SNPs based on a calibration procedure of the XtX statistics, and (iii) to consider alternative covariate models for analyses of association with population-specific covariables. In particular, the auxiliary variable model allows one to deal with multiple testing issues and, providing the relative marker positions are available, to capture some linkage disequilibrium information. A comprehensive simulation study was carried out to evaluate the performances of these different models. Also, when compared in terms of power, robustness, and computational efficiency to five other state-of-the-art genome-scan methods (BayEnv2, BayScEnv, BayScan, FLK, and LFMM), the proposed approaches proved highly effective. For illustration purposes, genotyping data on 18 French cattle breeds were analyzed, leading to the identification of 13 strong signatures of selection. Among these, four (surrounding the KITLG, KIT, EDN3, and ALB genes) contained SNPs strongly associated with the piebald coloration pattern while a fifth (surrounding PLAG1) could be associated to morphological differences across the populations. Finally, analysis of Pool-Seq data from 12 populations of *Littorina saxatilis* living in two different ecotypes illustrates how the proposed framework might help in addressing relevant ecological issues in nonmodel species. Overall, the proposed methods define a robust Bayesian framework to characterize adaptive genetic differentiation across populations. The BayPass program implementing the different models is available at http://www1.montpellier.inra.fr/CBGP/software/baypass/.

Keywords: genome scan, Bayesian statistics, association studies, linkage disequilibrium, Pool-Seq

LiveSeeding

Of course, here we go..



LiveSeeding

Summary – What have we seen so far?

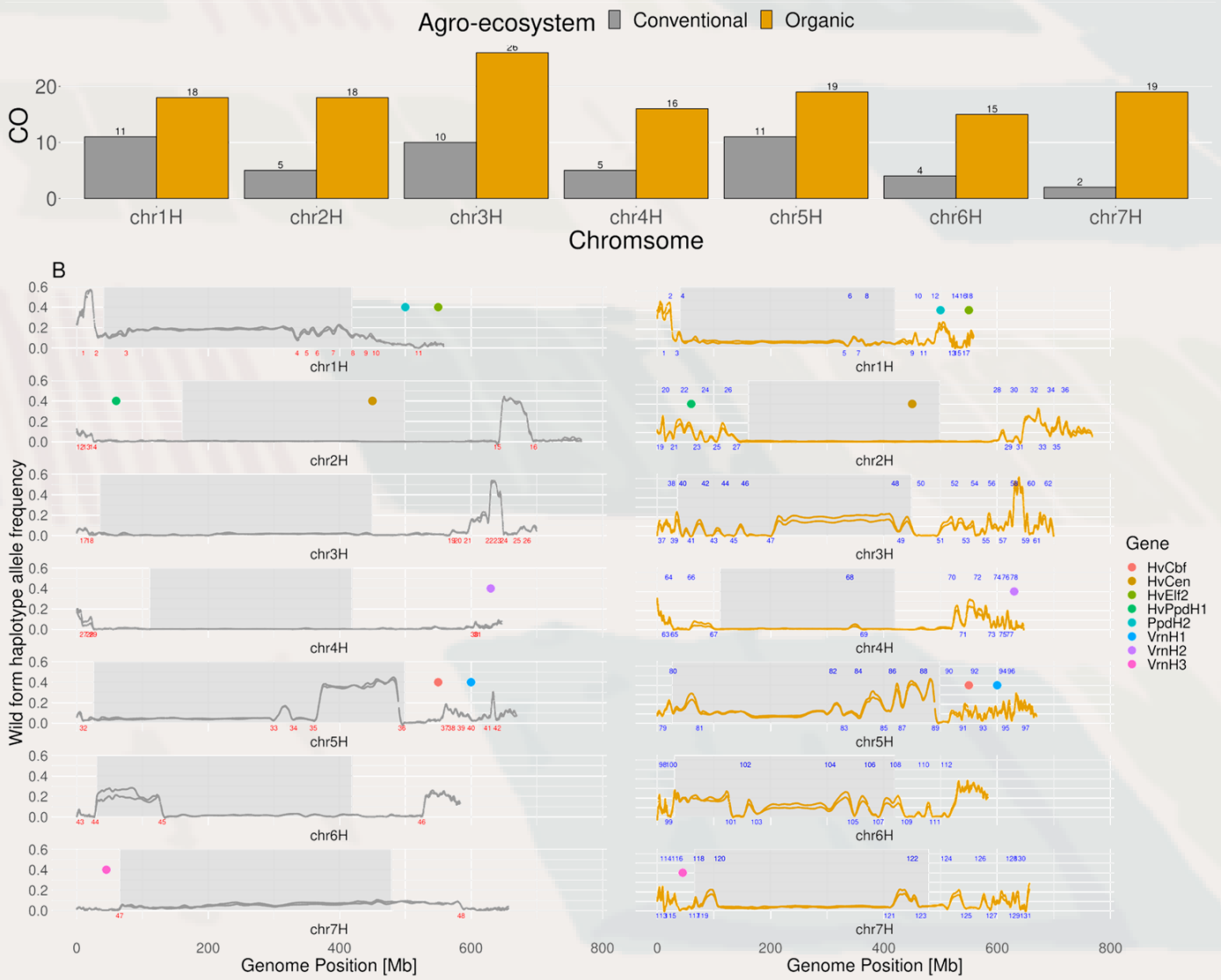
- Genotyping an entire population can provide a lot of information for **little costs**
- Changes across generations can be tracked with a **high precision**
- The changes can also serve to **predict** changes in the **phenotypic characterisitcs**
- Appling the latest advances in genotyping methods can also highlight ***Epigenetic & Microbial*** adaptations

So far? Is there more to it?

Yes, of course there is! We have not yet harvested all information from the genotyping

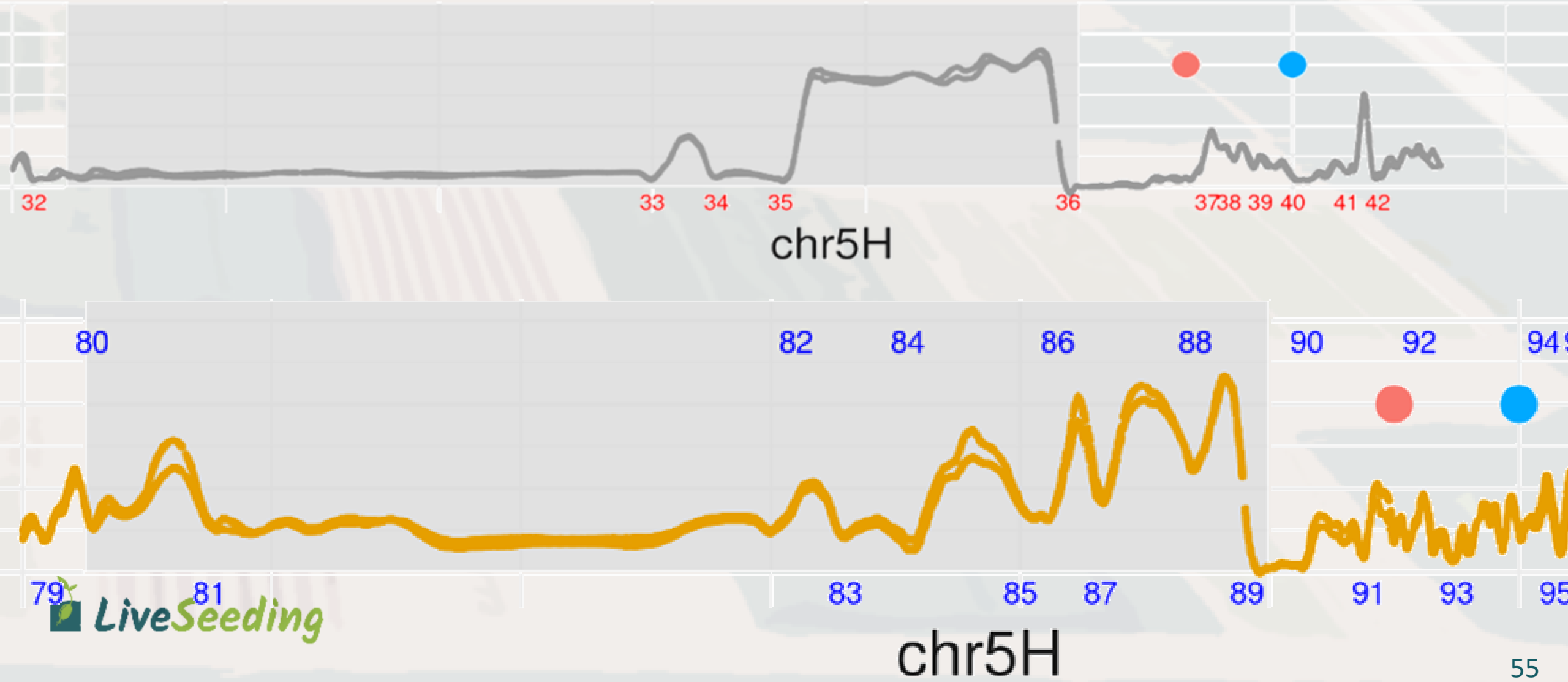
We should take everything we get, when we already spend money on the genotyping!

Recombination detection



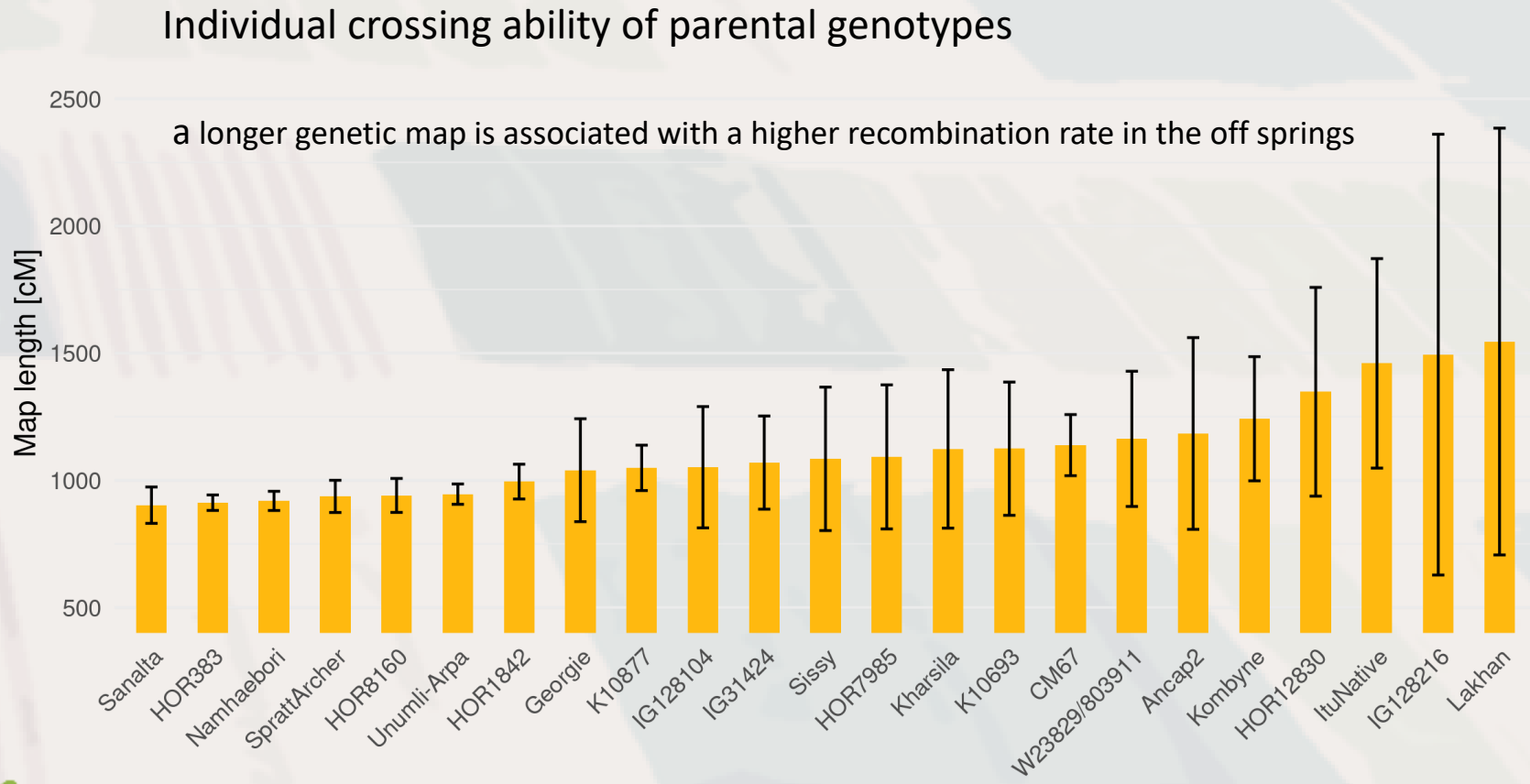
LiveSeeding

Recombination detection



LiveSeeding

The knowledge on the recombination rate provides also very useful information for the breeding of new variaties or populations



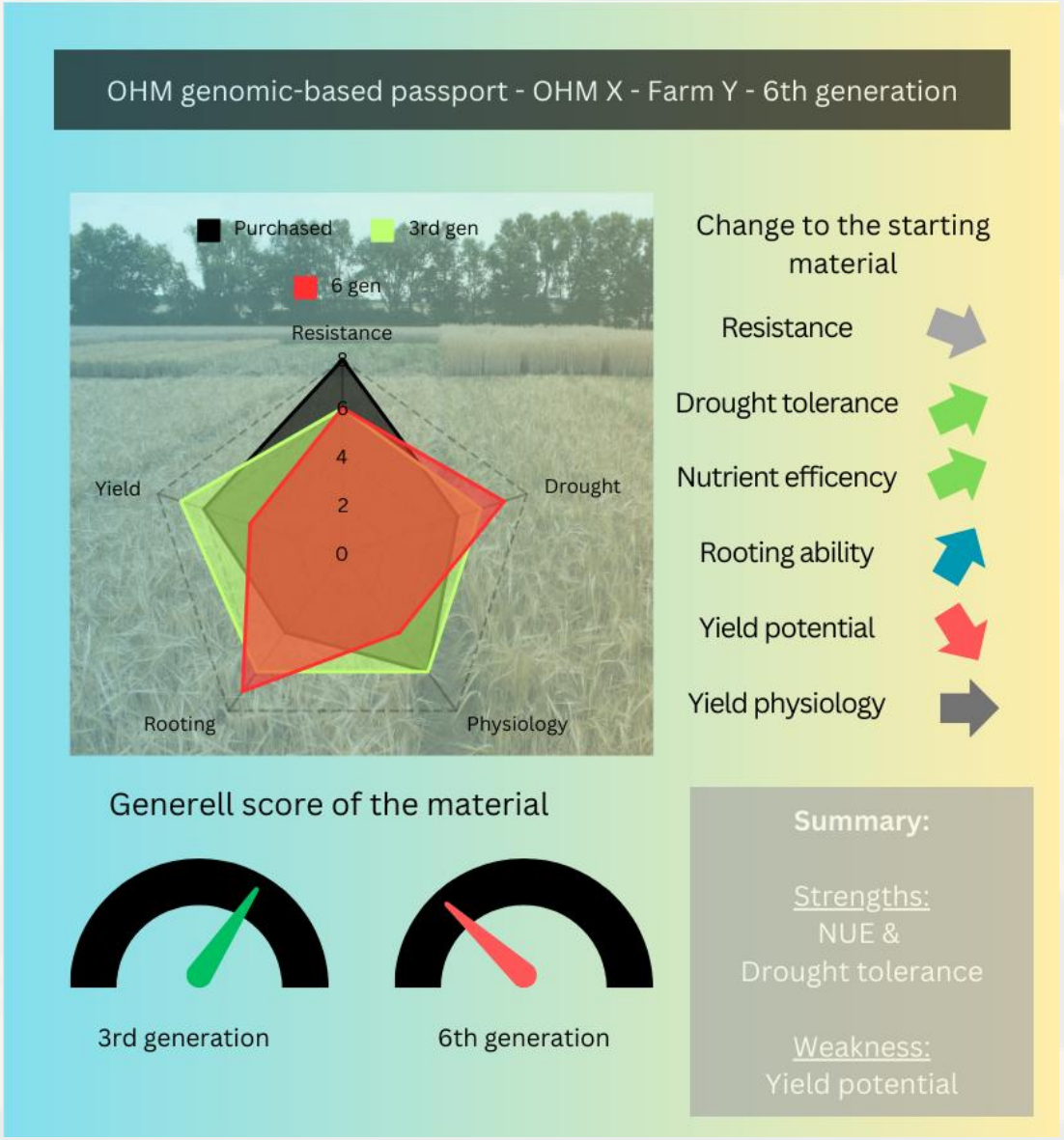
LiveSeeding

Final conclusions

- The presented genomic approach can provide deep information on populations, their change across years and support breeding of new variaties or (O)HM
- It helps to understand beneficial adaptations towards niche environments
- And it could serve as..

OHM Track alternative – how could the genomics help to track OHM and provide „open access“ to a constantly changing material?

How could it be applied in OHM tracking?



LiveSeeding

Thanks for your attentic

VI_EUCARPIA CONFERENCE

on breeding to meet environmental & societal challenges

26 TO 28 MAY 2025 PORTUGAL

Instituto Politécnico de Coimbra - ESAC

Polytechnic University of Coimbra · LiveSeeding · EUCARPIA Section Organic & Low Input Agriculture

LiveSeeding

www.LiveSeeding.eu
key results, newsletter, upcoming events, policy briefs, videos, training, practice abstract

LiveSeeding

Link to sister project on organic fruit breeding
www.InnOBreed.eu

LiveSeeding

Co-funded by the European Union

Funded by the European Union, the Swiss State Secretariat for Education, Research and Innovation (SERI) and UK Research and Innovation (UKRI). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or REA, nor SERI or UKRI.

UK Research and Innovation