PGR management in the genomics era
WP3

Lorenzo Barchi, Ezio Portis, Giuseppe Aprea, Jan Bartos, Jaroslav Dolezel, Giovanni Giuliano

UNIVERSITÀ DI TORINO

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Omics technologies represent a leap forward for the conservation, management and characterization of PGR

Current PGR management does not involve the routine use of genomic tools to trace accessions during seed regeneration or vegetative propagation

Taxonomical classification on morphological traits alone can be error-prone

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Genomics-based identification should become the golden standard for the management of PGR.

Genebank accessions should have their DNAs (accessions) barcoded or sequenced

PGR *in situ* managing can benefit from the use of genomics

Many smaller genebanks, *in situ* collection holders and final users of the PGR as well as scientists working on PGR simply don’t have the facilities and skills to use these technologies.

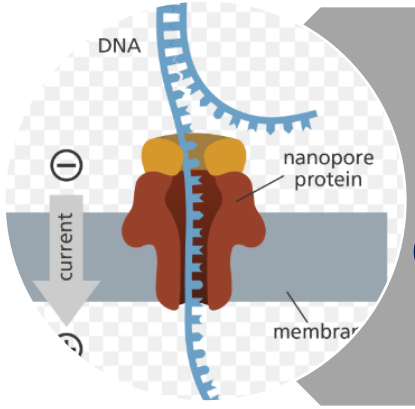
PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Refine and suggest protocols and methods for DNA barcoding (D3.2)

DNA (accessions) barcoding and genetic classification based on:

Reduced representation sequencing (RRS).

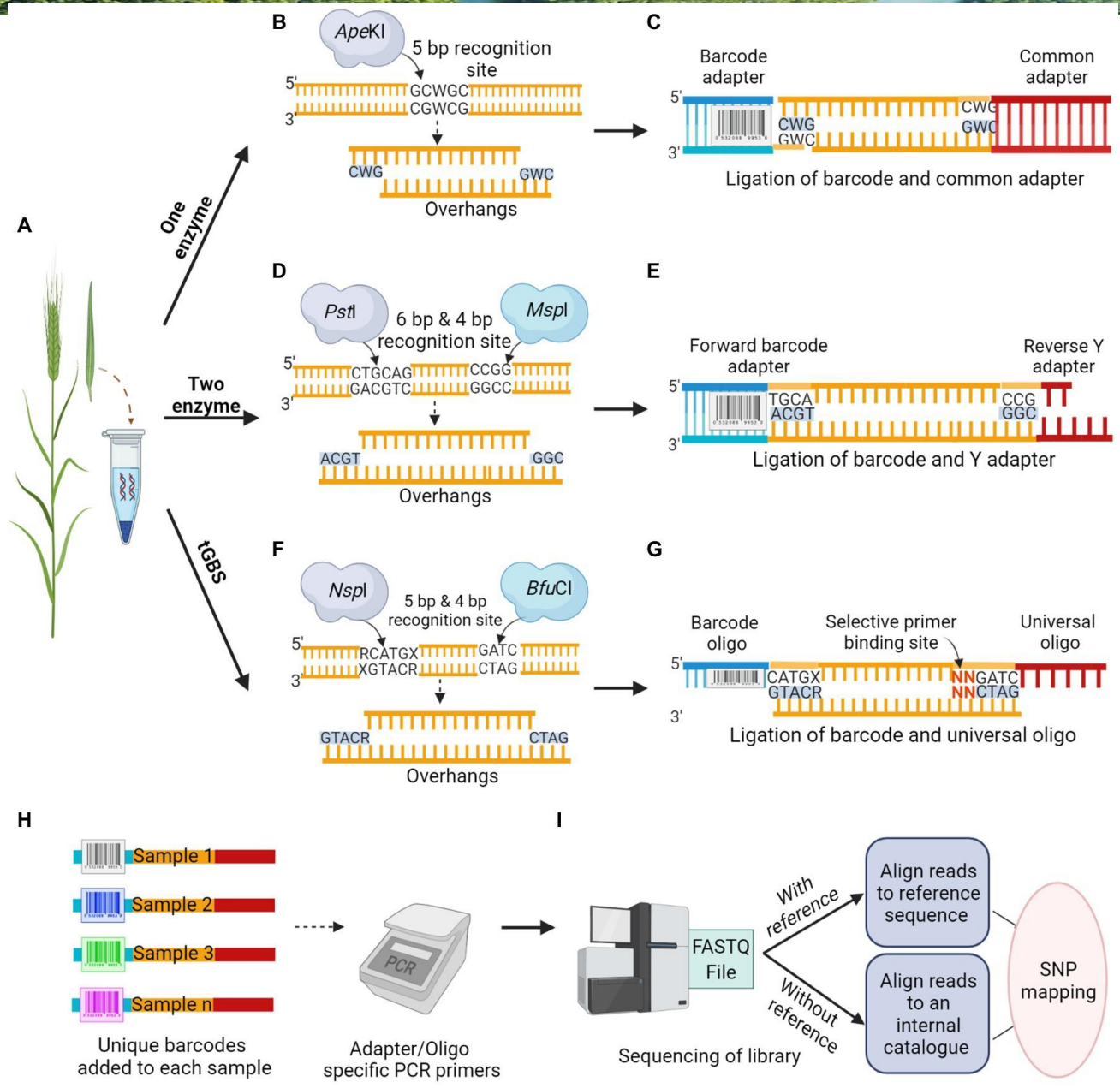
Using a set of <u>common</u> SNPs, it is now possible to identify thousands of polymorphism at low cost for accession identification



Whole genome resequencing (short reads) and chromosome scale sequencing (based on long reads)

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

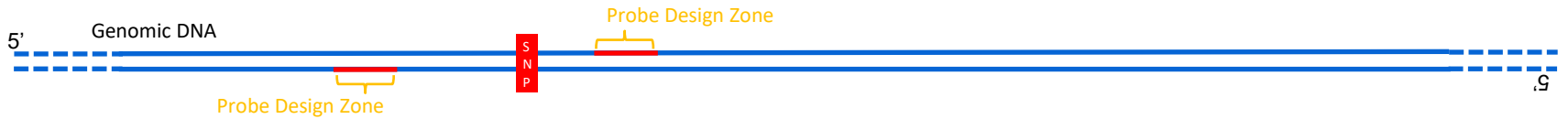
GBS: genotyping by sequencing

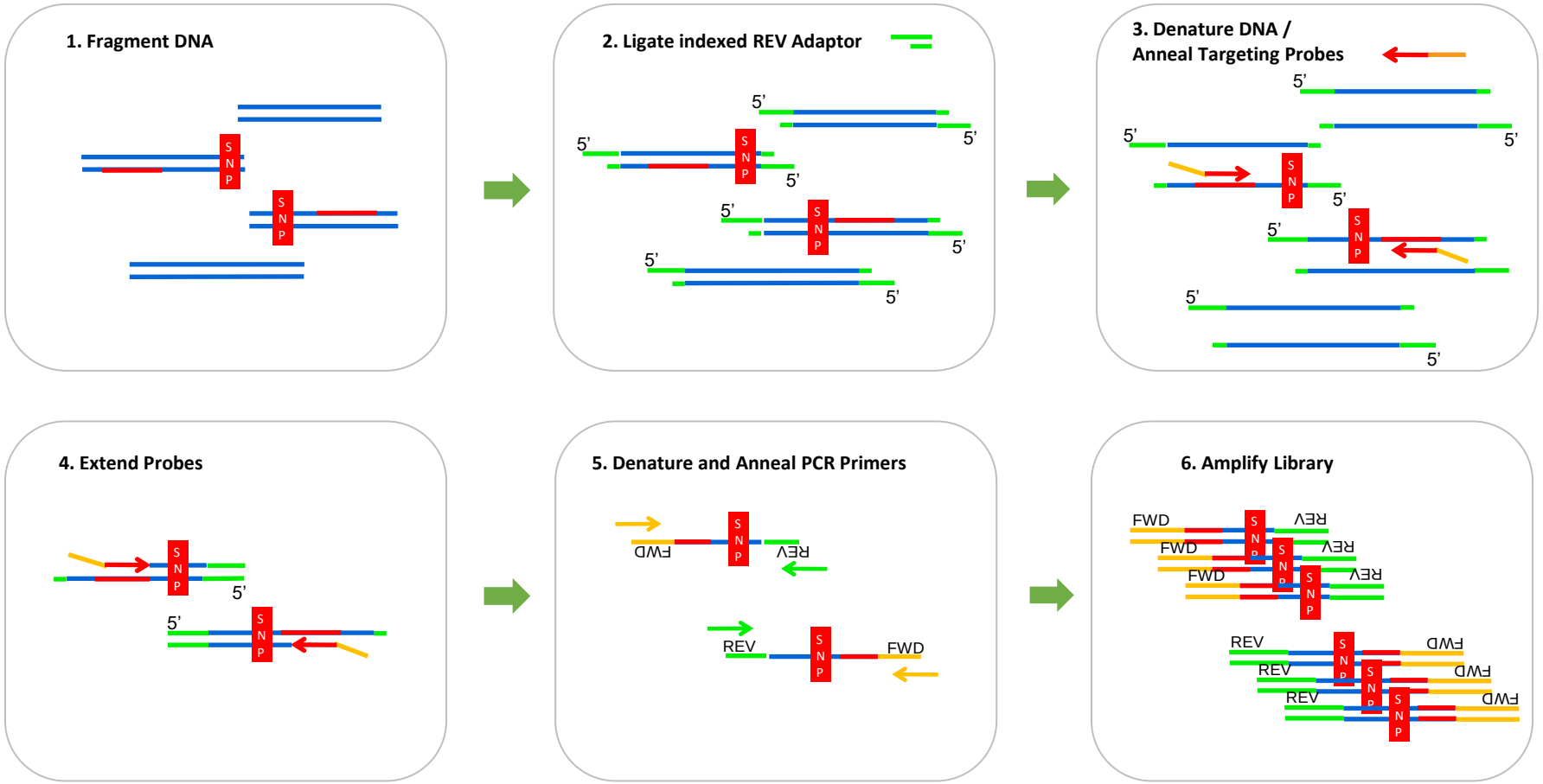


PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Reduced representation sequencing

Single Primer Enrichment Technology (SPET)

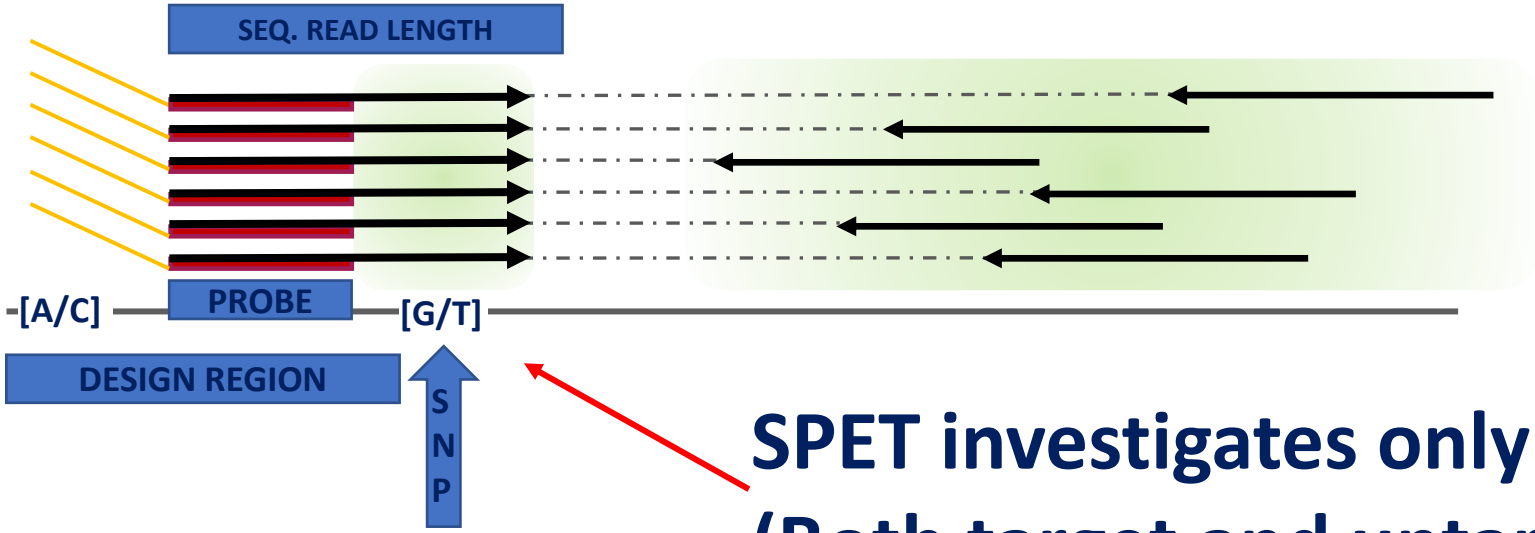




PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Reduced representation sequencing

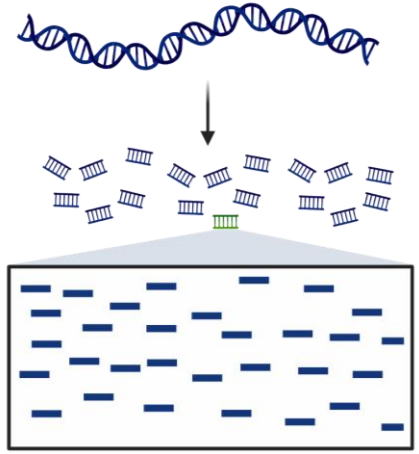
Target Enrichment Application: Genotyping By Sequencing

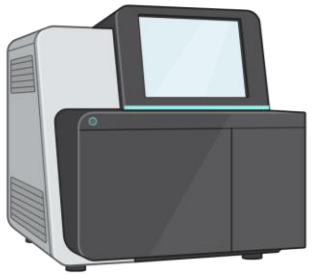


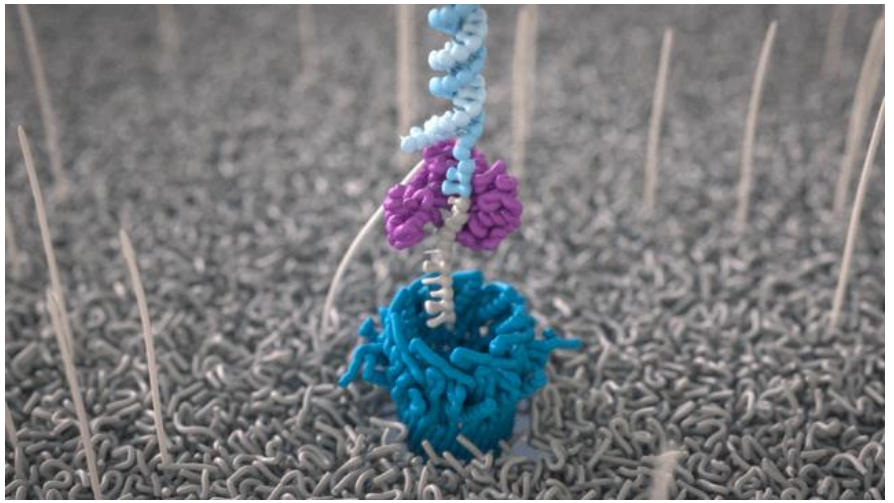
SPET investigates only TARGET loci (Both target and untargeted SNPs)

- Highly-multiplex genotyping (up to 3,092)
- Enables detection of a large number of known SNPs
- Every sequencing read is informative

WGS

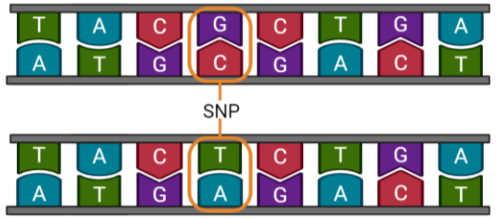




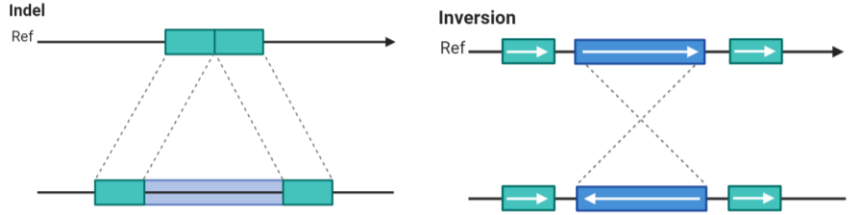


Sequencing data as well as Pangenome construction

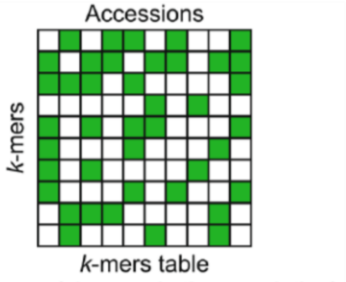
SNPs



Structural variations (SVs)



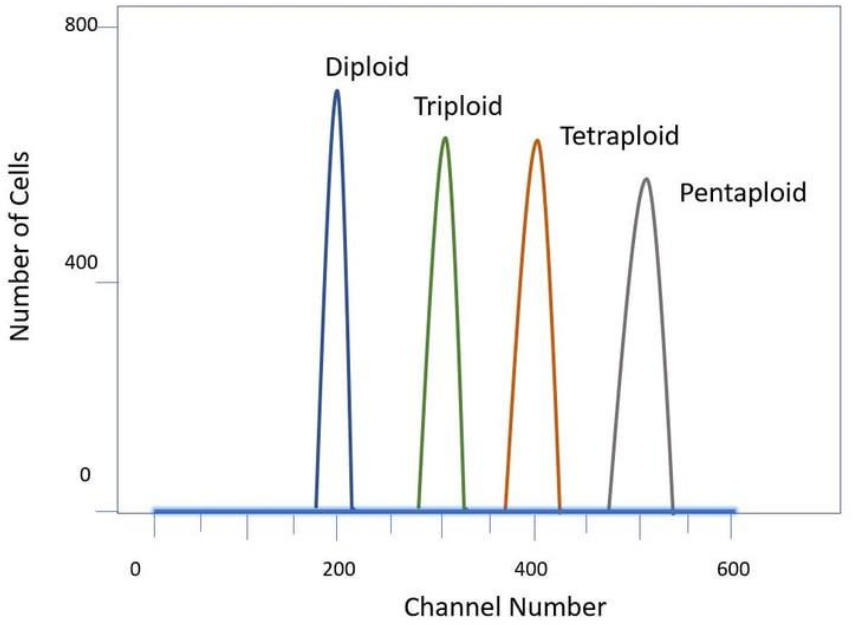
K-mers



D3.2: Cytogenomics

Cytogenomics protocols for:

Ploidy/aneuploidy/genome size determination (especially for clonally propagated species)



Providing precious information on chromosome number, karyotype and genomic constitution of CWR (Crop Wild Relatives)

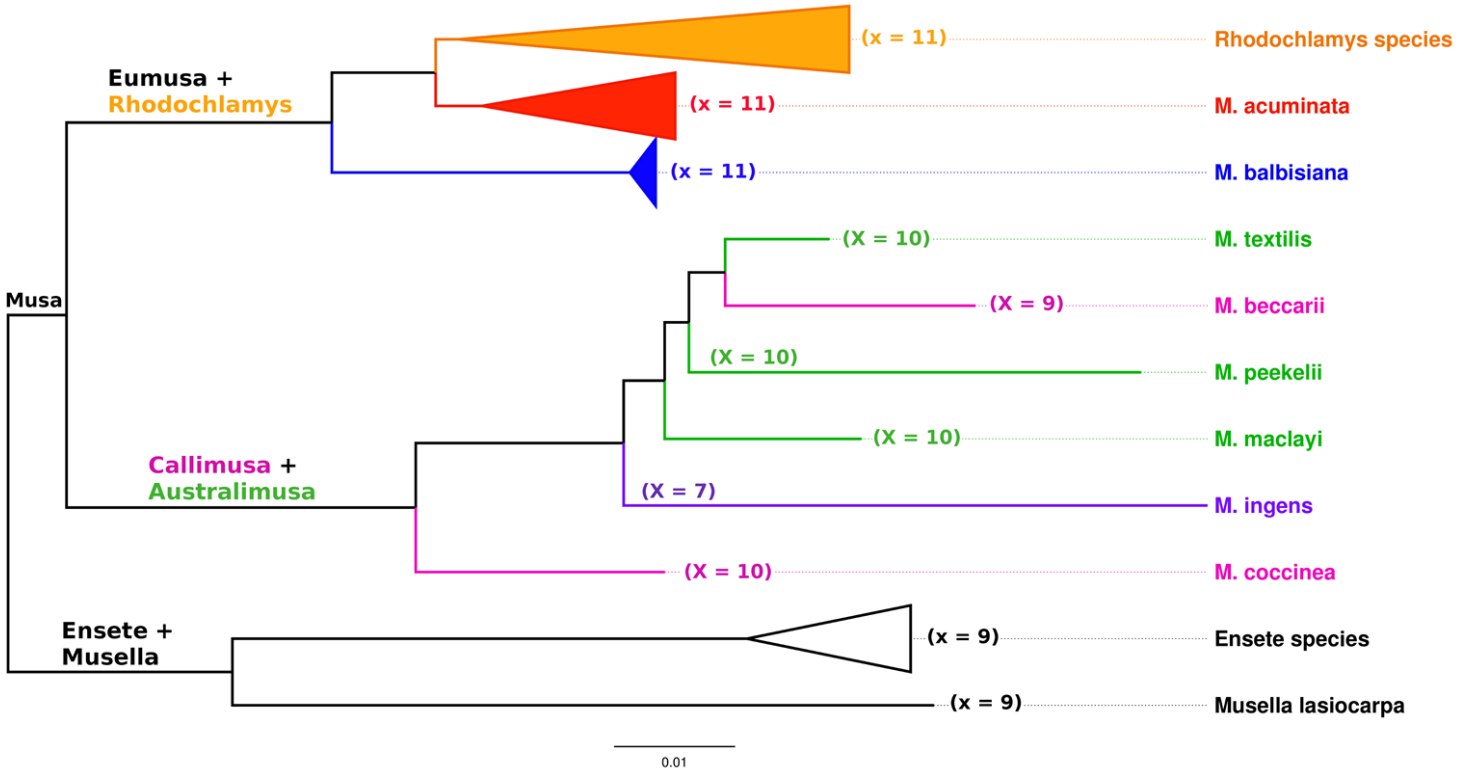
Useful for introgression by chromosome-mediated gene transfer.

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Examples from the genus *Musa*

- Basic chromosome number, ploidy and genome size vary between closely related species

- Differences between karyotypes of species with the same ploidy and occurrence of structural chromosome heterozygosity

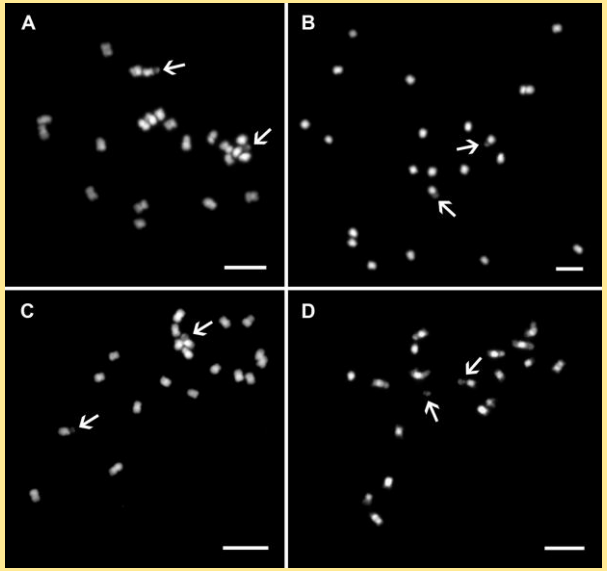


PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Cytogenomics methods

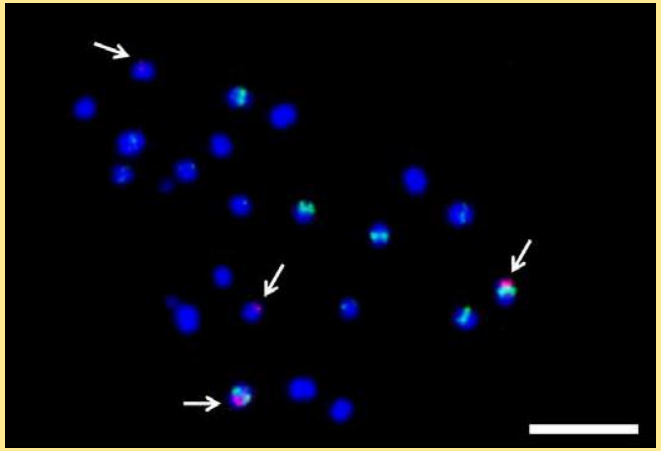
- Chromosome staining for determination of ploidy and chromosome number
- Molecular cytogenetics (fluorescence *in situ* hybridization – FISH) for detailed karyotype analysis

Chromosome counting

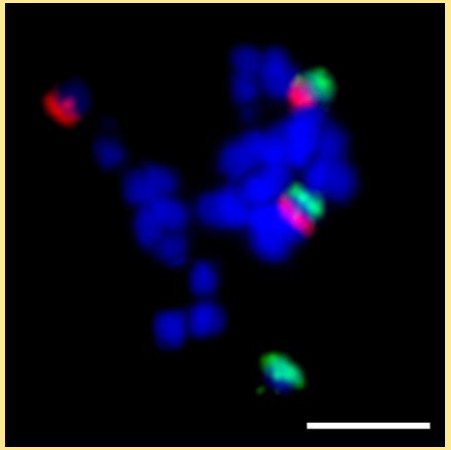


A) *Musa acuminata* (2n=2x=22); B) *Musa laterita* (2n=2x=22); C) *Musa maclayi* (2n=2x=20); D) *Musa beccarii* (2n=2x=18) Arows indicate NOR regions. Bar = 5 µm

Molecular cytogenetics



Fluorescence in situ hybridization (FISH) with probes to 5S rDNA (green) and satellite CL18 (red) in *Musa schizocarpa* (2n=2x=22). Chromosomes were counterstained with DAPI (blue). Bar = 5 µm



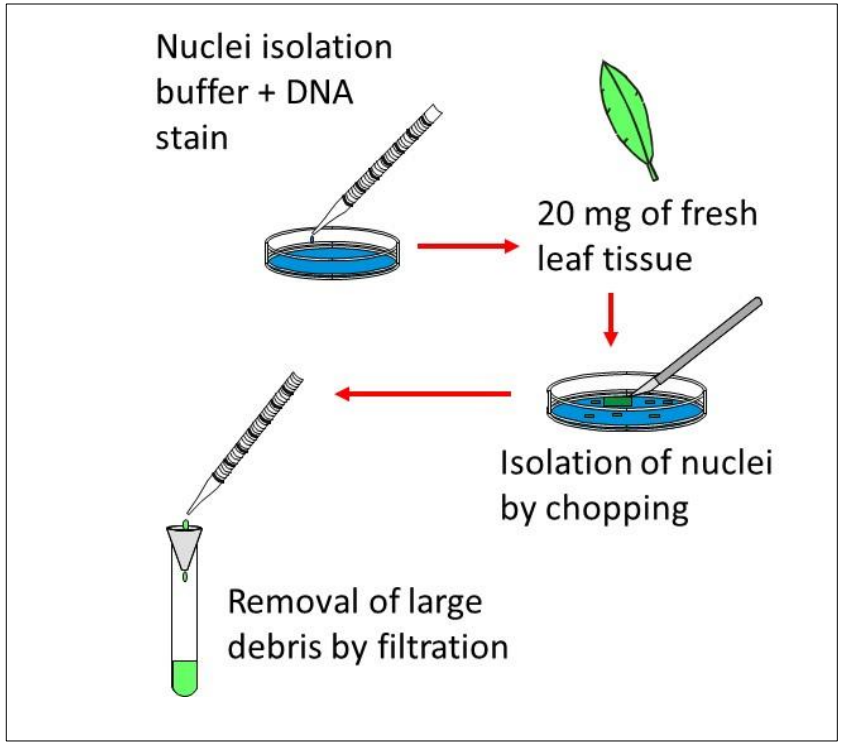
Chromosome-arm specific oligo painting FISH in aneuploid clone 'Imbogo' (2n = 3x – 1 = 32). Long arm of chromosome 3 (green); short arm of chromosome 8 (red). Chromosomes were counterstained with DAPI (blue). Bar = 5 µm

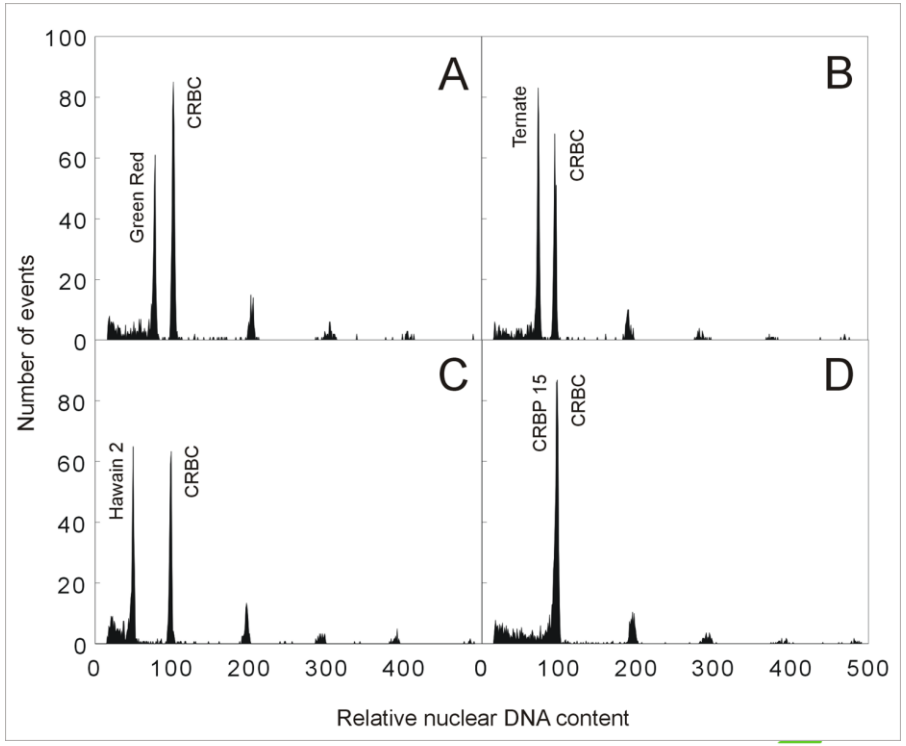
Šimoníková et al., Plants 2022

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Ploidy/aneuploidy/genome size determination

- Flow cytometry of nuclear DNA content as a rapid, accurate, easy and generally applicable method
- Characterization of >1500 accessions in *Musa* genebank (ITC collection at KU Leuven, Belgium)





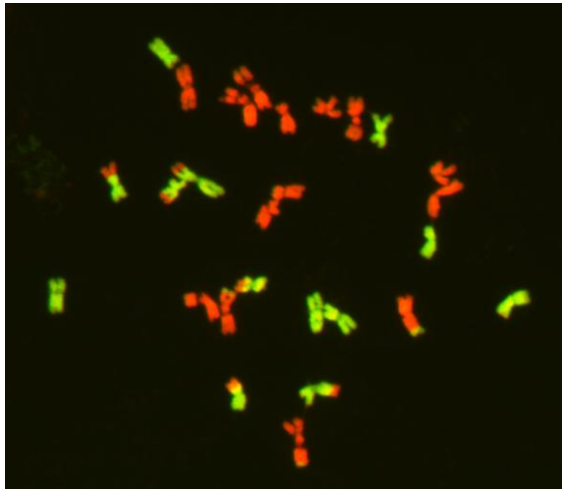
Šimoníková et al., Plants 2022

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Determination and facilitation of interspecific crossing/ introgression programs

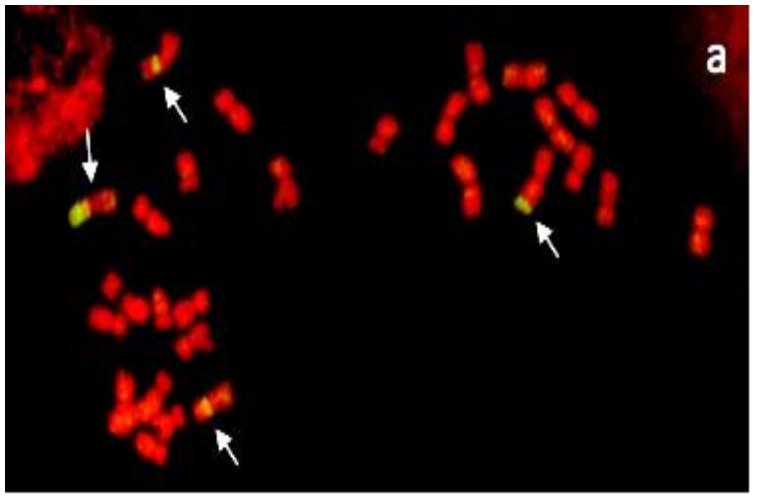
- Genome composition of Festuca x Lolium hybrids and introgression lines
- Molecular cytogenetics (genomic *in situ* hybridization – GISH) for identification of parental chromosomes in interspecific hybrids and introgressions

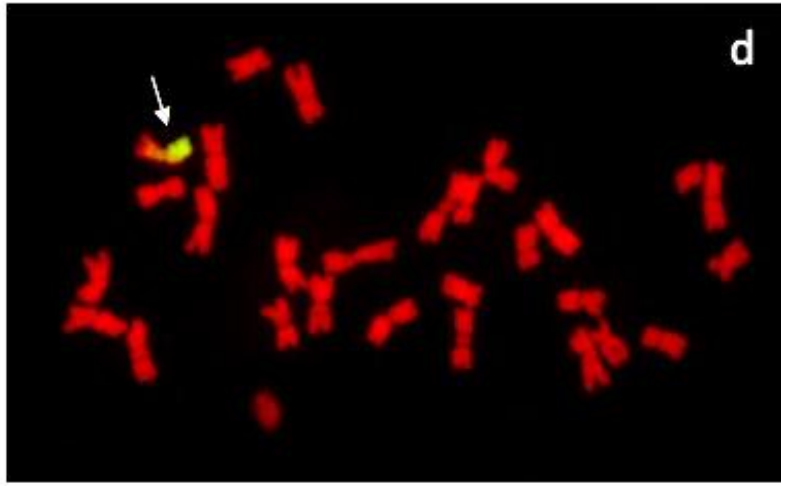
Festuca x Lolium hybrid



Fescue Ryegrass

Introgression lines



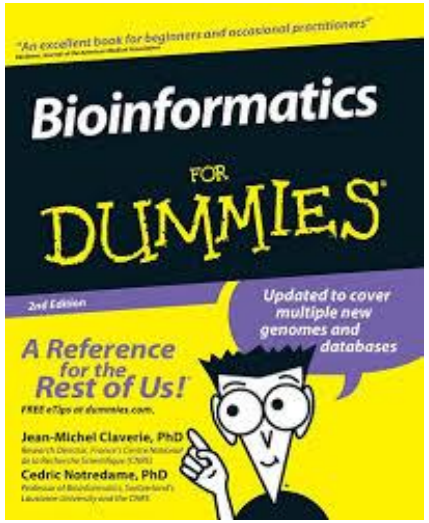


Kopecký et al., 2017, 2019, Majka et al. 2023; Mahelka et al., 2023

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Bioinformatics

An interdisciplinary field of science that develops methods and software tools for understanding biological data.



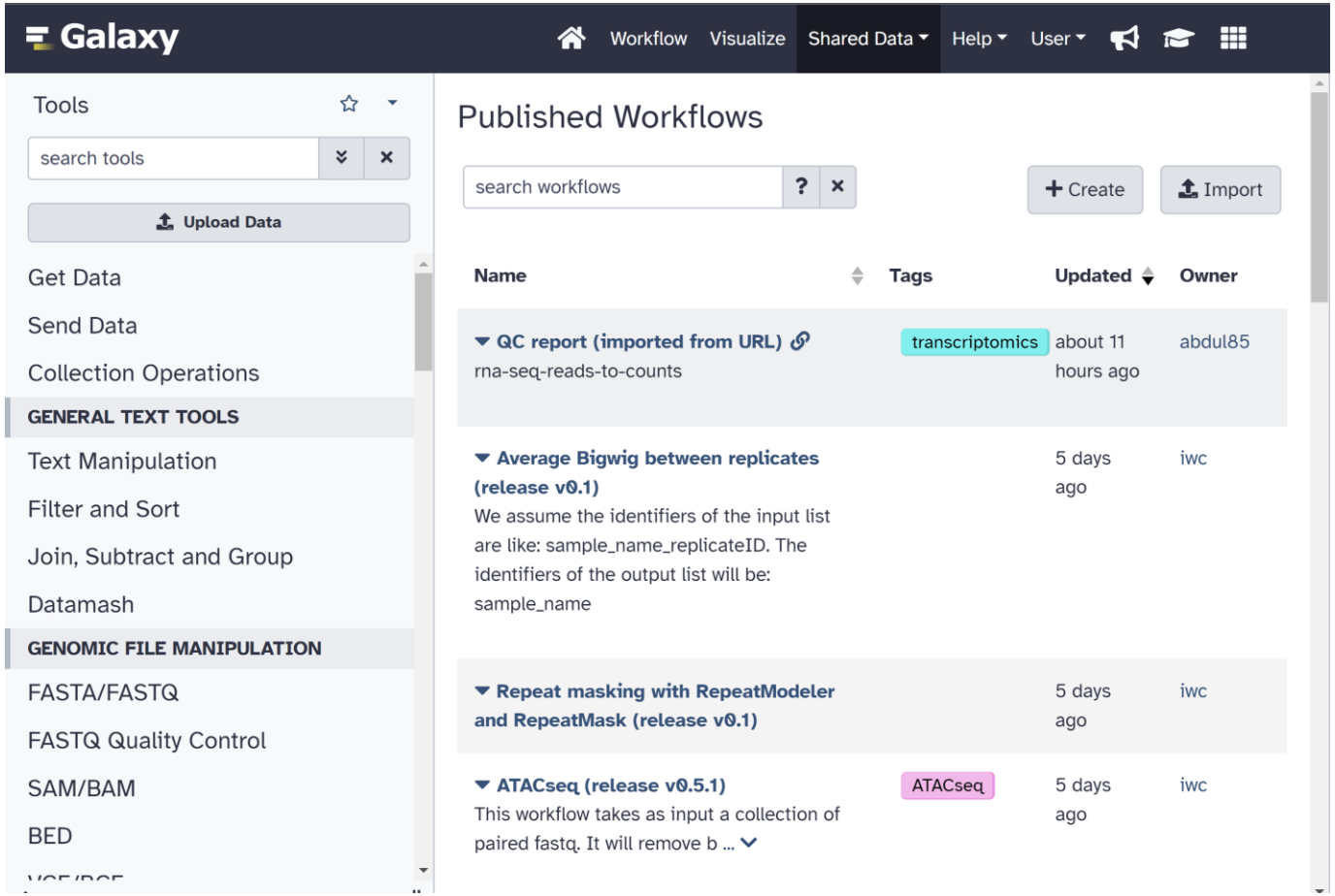
It’s not effortless!

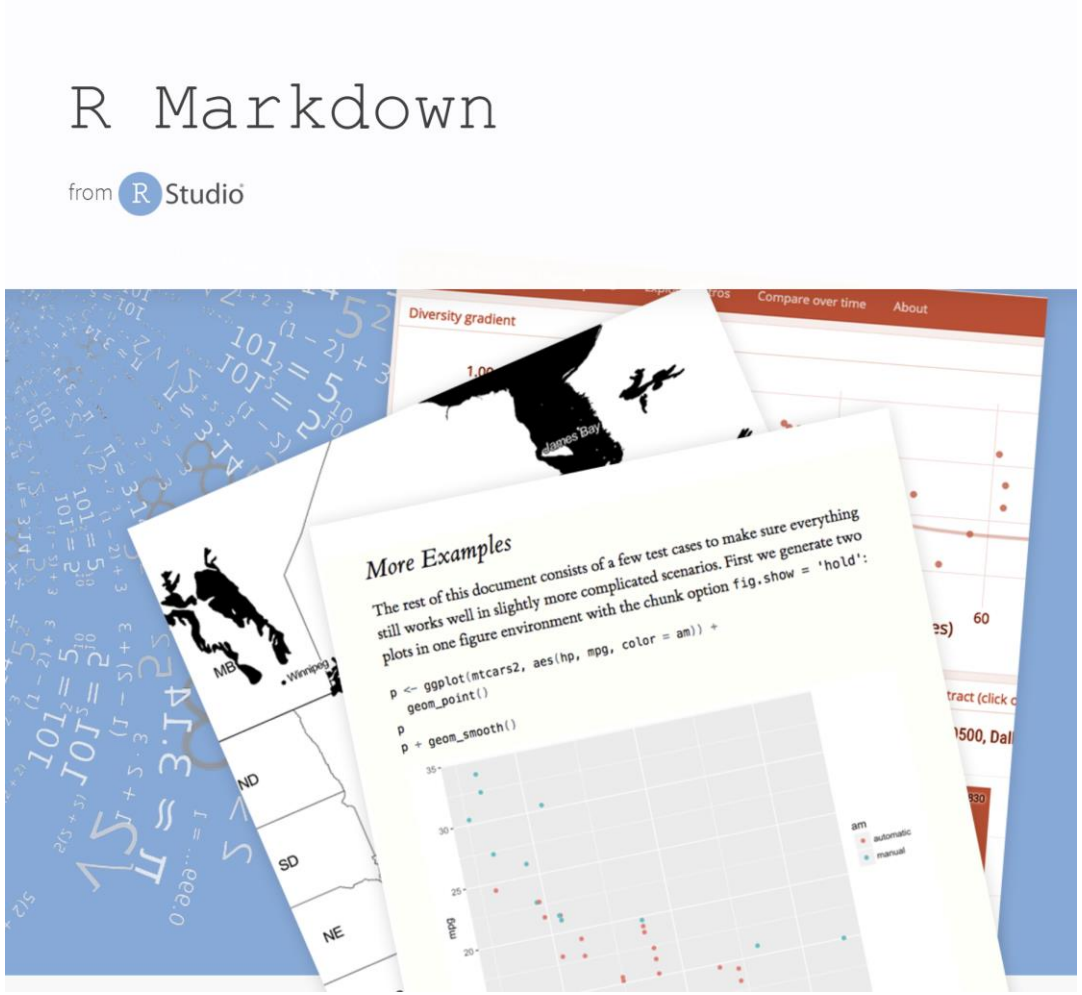
PRO-GRACE aims to provide and refine bioinformatic methods for different purposes (D3.5)

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Bioinformatics

PRO-GRACE aims to provide and refine bioinformatic methods for different purposes (D3.5)





PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Duplicates and misclassified

Based on SNPs identified using reduced representation/whole genome and chromosome scale (re)sequencing

Protocol will be developed/exploited to correct misidentifications, spot duplicates within and between collections

Exploit the barcode as well as a DOI associated to each of them, to minimize misidentifications and manage excessive levels of duplication

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Duplicates and misclassified

the plant journal SEB
SOCIETY FOR EXPERIMENTAL BIOLOGY

Original Article | Open Access

Analysis of >3400 worldwide eggplant accessions reveals two independent domestication events and multiple migration-diversification routes

Lorenzo Barchi, Giuseppe Aprea, M. Timothy Rabanus-Wallace, Laura Toppino, David Alonso, Ezio Portis, Sergio Lanteri, Luciana Gaccione, Emmanuel Omondi, Maarten van Zonneveld, Roland Schafleitner, Paola Ferrante, Andreas Börner, Nils Stein, Maria José Díez, Veronique Lefebvre, Jérémy Salinier, Hatice Filiz Boyaci, Richard Finkers, Matthijs Brouwer, Arnaud G. Bovy, Giuseppe Leonardo Rotino, Jaime Prohens, Giovanni Giuliano ... See fewer authors

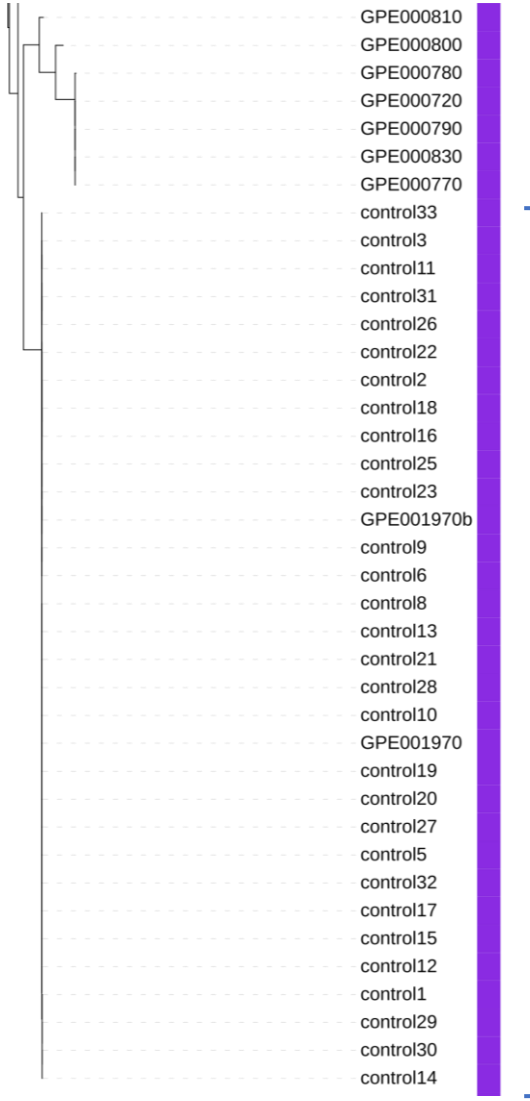
First published: 08 September 2023 | https://doi.org/10.1111/tpj.16455

Our approach: calculate identity by state (IBS) in pair-wise comparisons

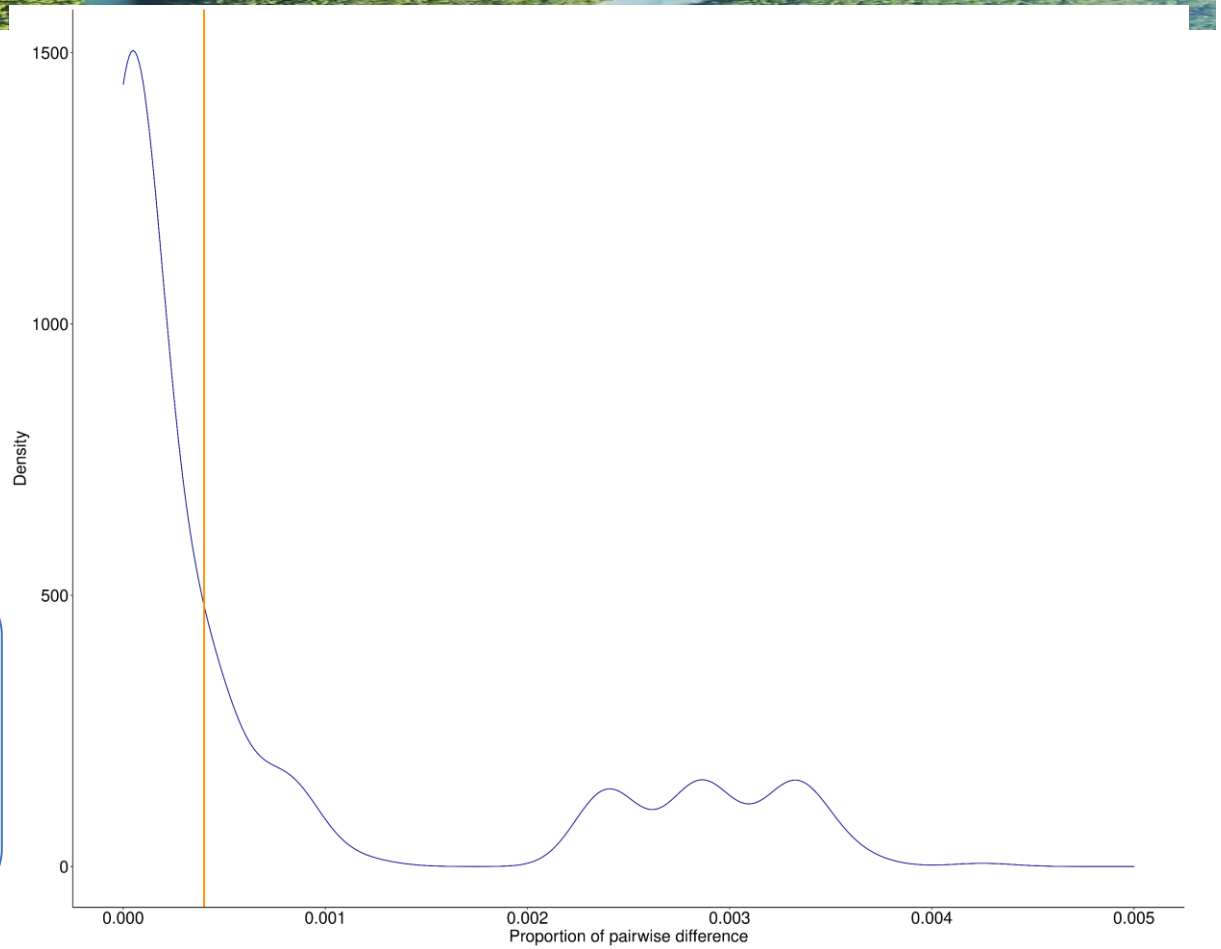
Which threshold to define accessions as duplicates?

Used replicates, i.e. additional samples from a well-known accession

Duplicates and misclassified



ML tree clearly shows identity among replicates



The distribution on controls IBSs allowed to impose a cut-off value of dissimilarity <= 0.0005

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Duplicates and misclassified

- For misclassification detection, for each accession, we determined the set of the 10 nearest neighbours and verified the species they belong to.
- A potential misclassification was called when the species match rate was less than 35%.
- The putative mislabeled accessions were eventually re-assigned to another species according to:

(i) assigned to a group having at least six representatives of another species and

(ii) after manual curation based on genebank passport data.

Gaps

CWRs are source of genes for resistance to biotic and abiotic stresses exploitable for domesticated species breeding

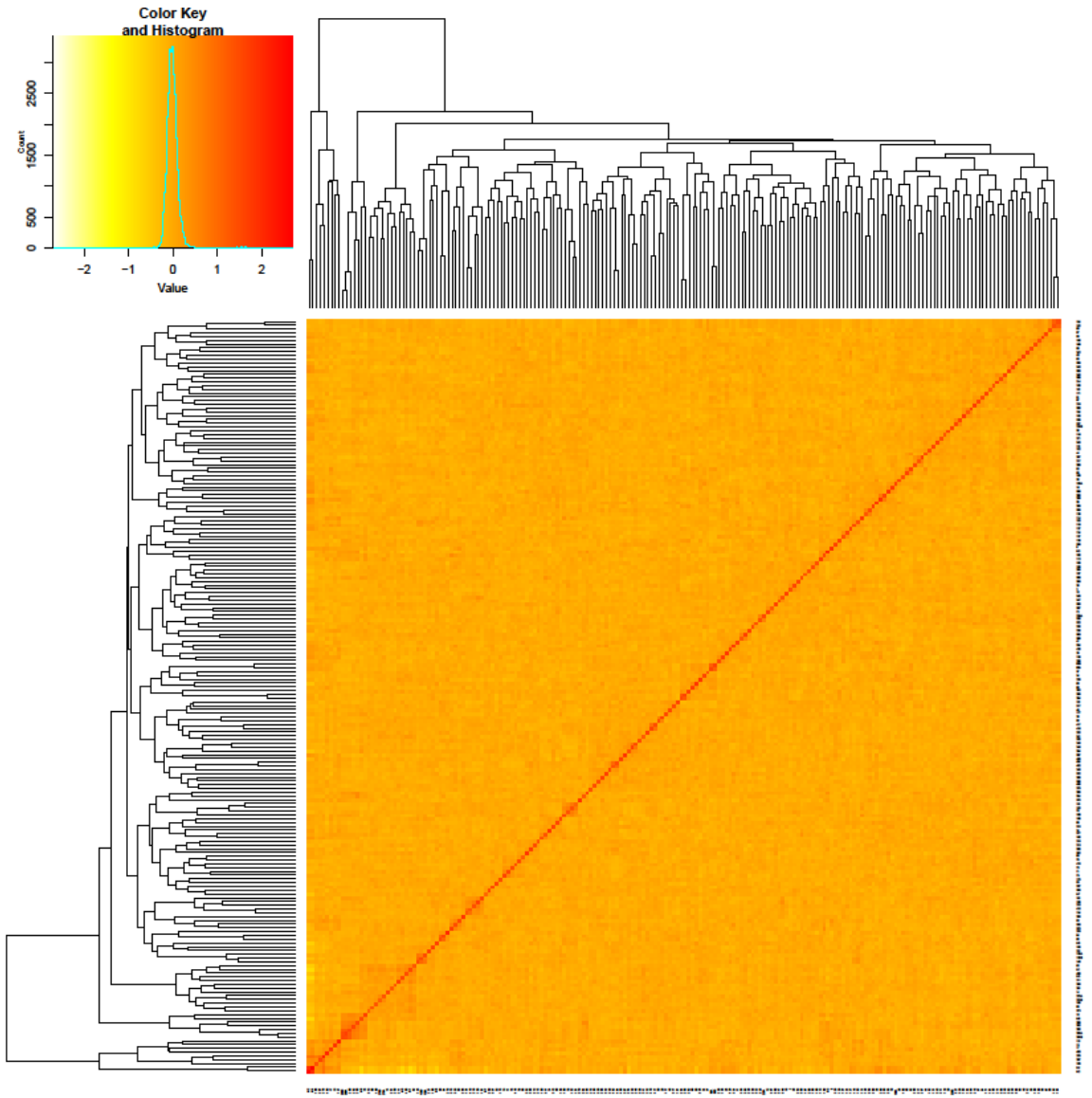
Critical gaps in ex situ and in situ collections with respect to natural genetic variability of a species have been identified.
CWR are also poorly represented in *ex situ* and *in situ* collections

Procedure for population structure and kinship estimates will be provided.
Together with barcoding, this will help to assist gap analysis and to reconstruct missing relations

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Kinship

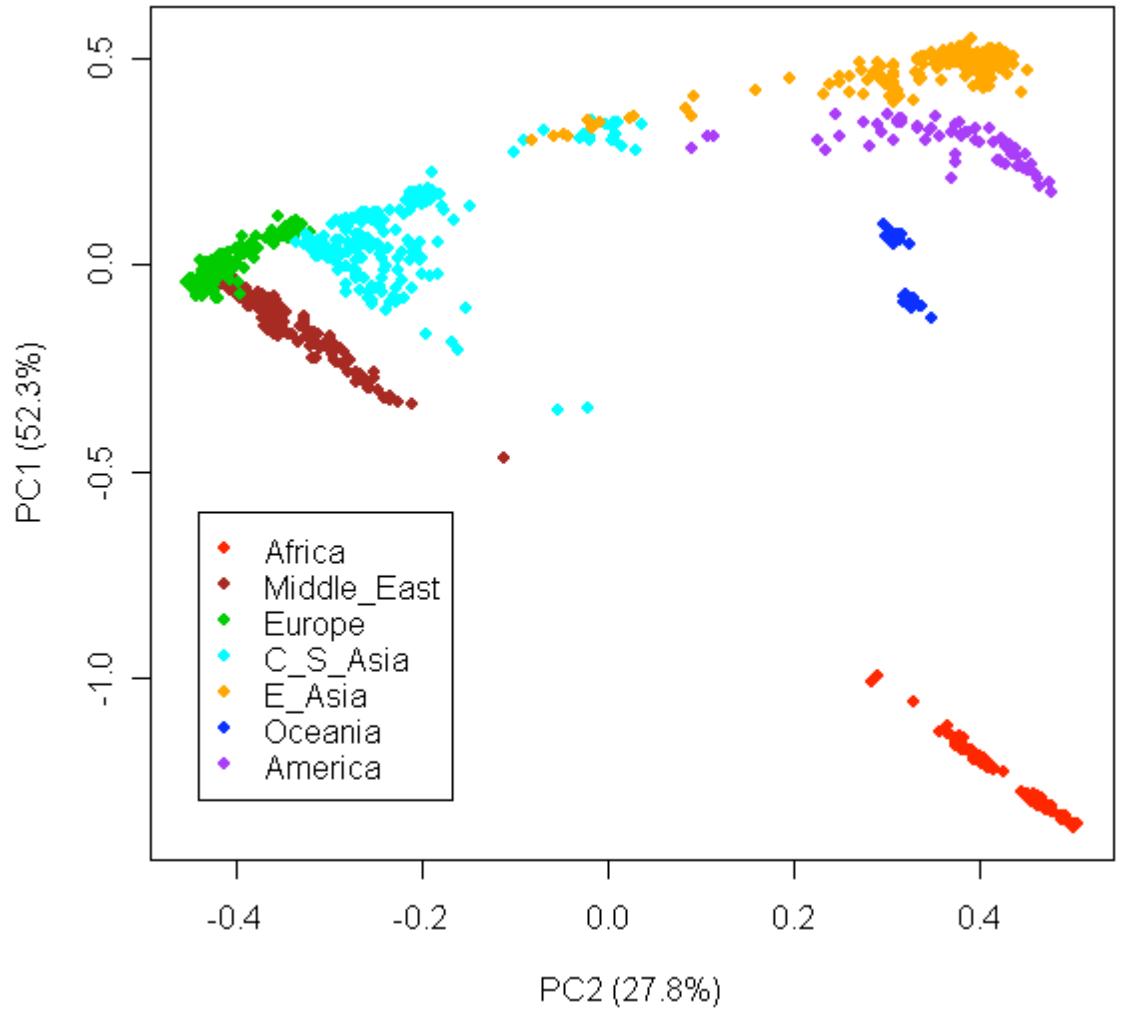
Genetic resemblance between individuals



PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Population stratification

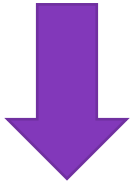
Population stratification is the presence of a systematic difference in allele frequencies between subpopulations in a population due to different ancestry



PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Core collection

Exploiting diversity/population structure/pedigree information

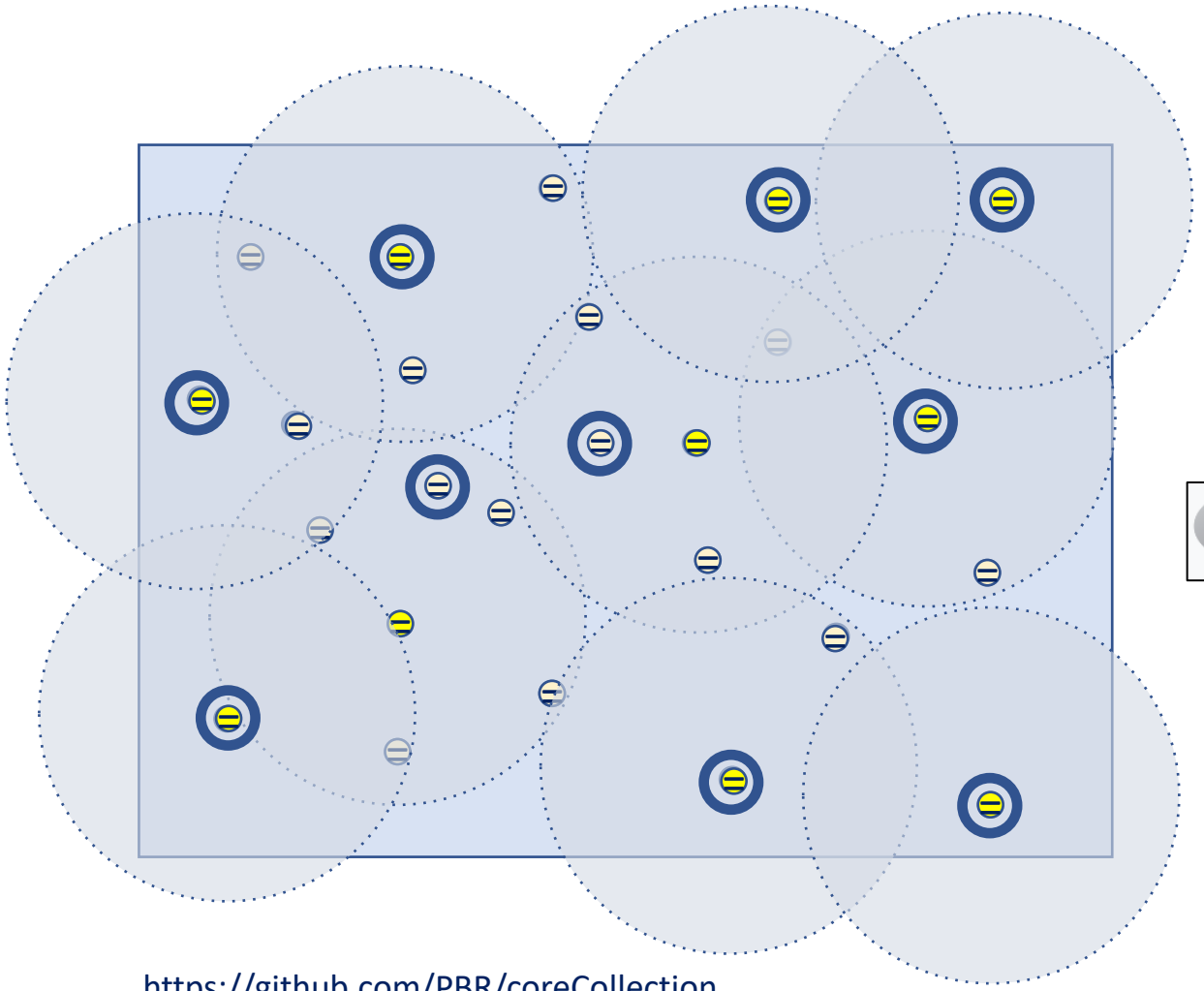


Construct core collections representative of the genetic variation of a much larger genepool

Set up pipelines for QTL and GWA studies which are now routinely used in the agronomic evaluation of large collections of germplasm

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Build up the eggplant core collection



https://github.com/PBR/coreCollection

SNPs → **VCF**

↓

Distance Matrix

↓

WAGENINGEN UNIVERSITY & RESEARCH

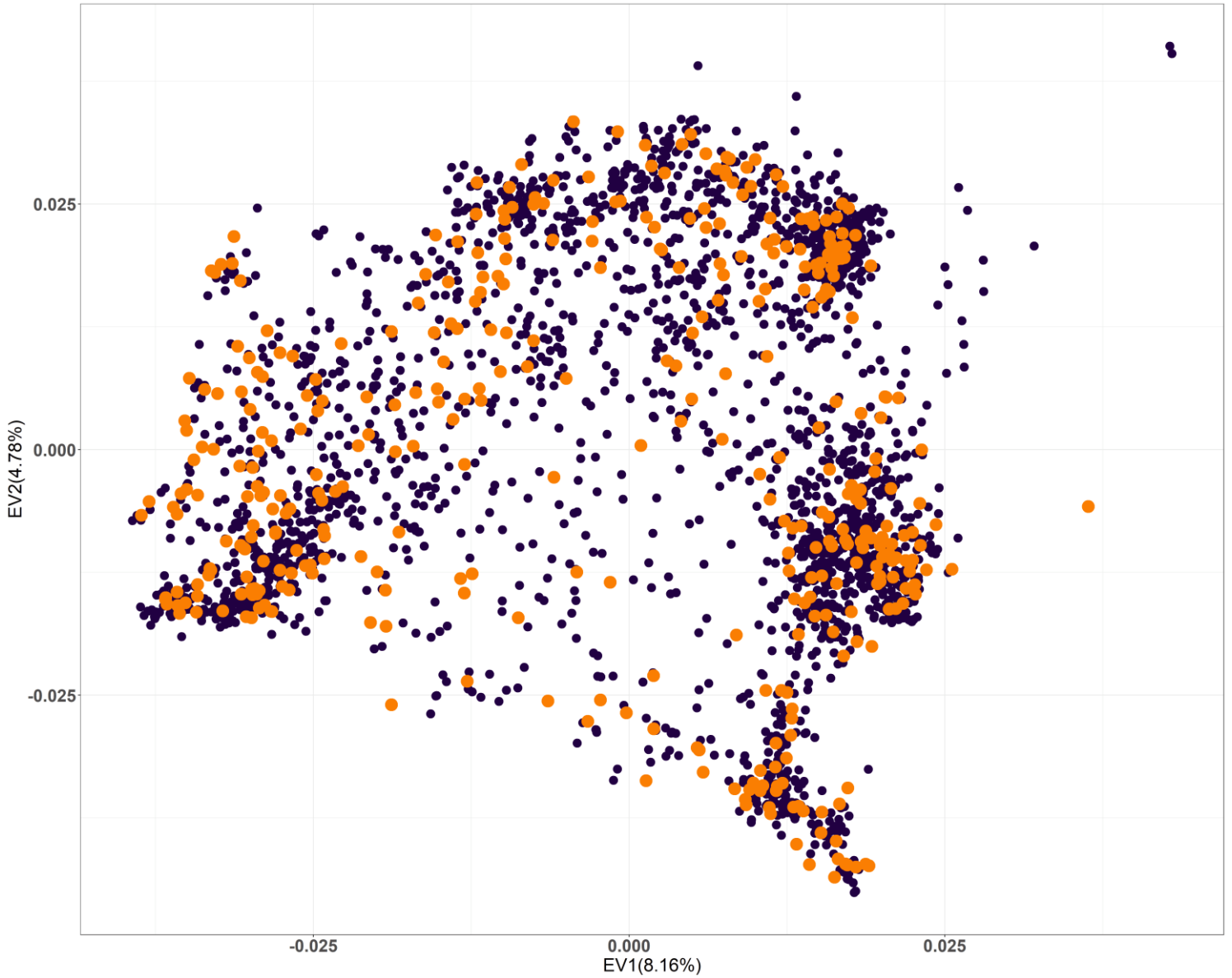
R

CoreCollection algorithm ← **Preselected individuals**

↓

Core Collection

Build up the eggplant core collection



Core collection

Whole collection

QTL and GWA

Phenotype

Genotype

SNPs, k-mers and SVs

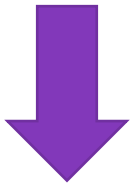
GAPIT

Genomic Association and Prediction Integrated Tool

(Version 3)

Pangenomics

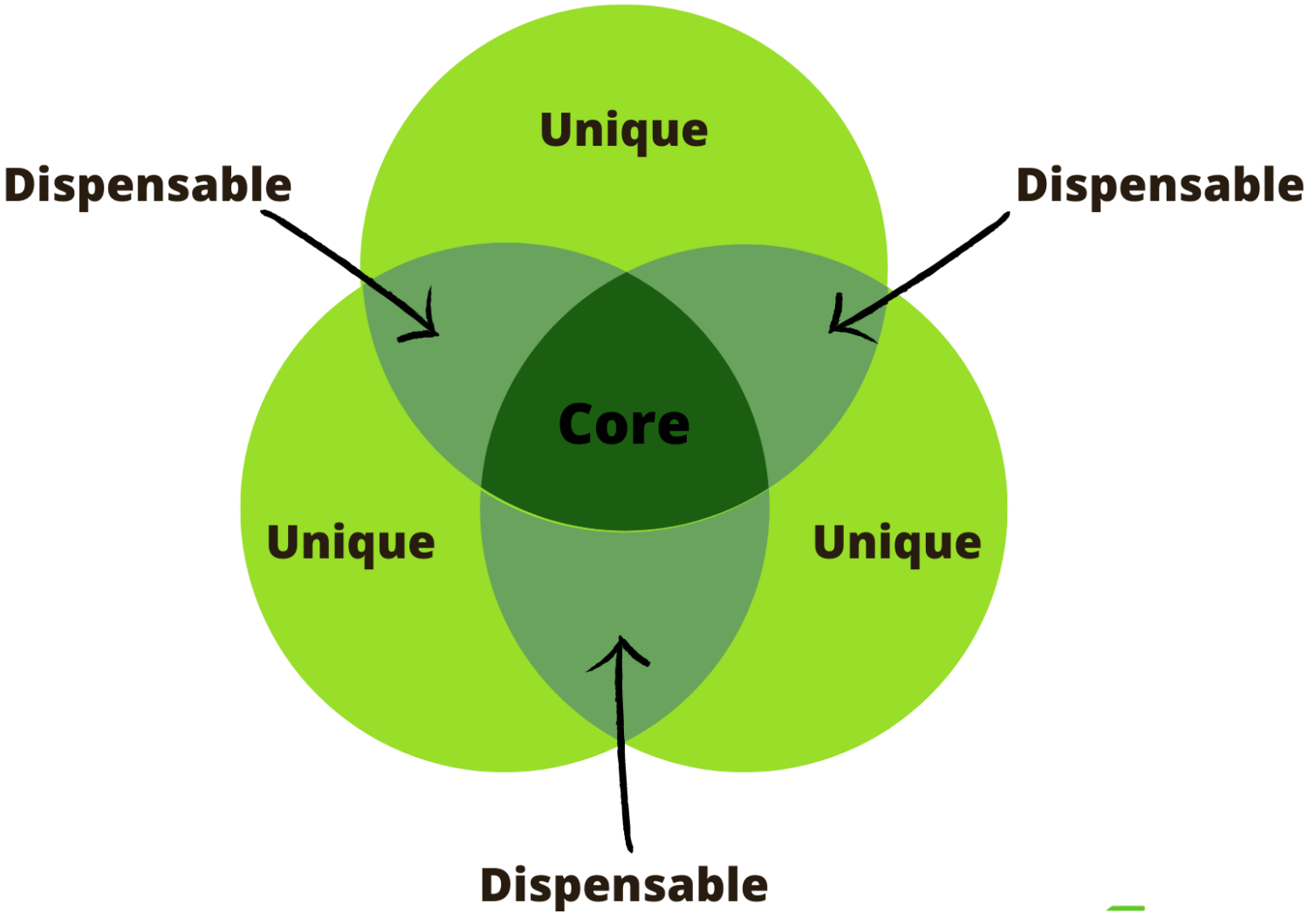
Pangenome construction can provide a comprehensive picture of the allelic variation within a single or among multiple species



Provide methodologies for pangenome construction, graph-pangenome, SNPs and indels identification as well as presence-absence variants (PAVs) and Selective Sweeps identification linked to key traits

PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Pan-genome



PRO-GRACE
PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

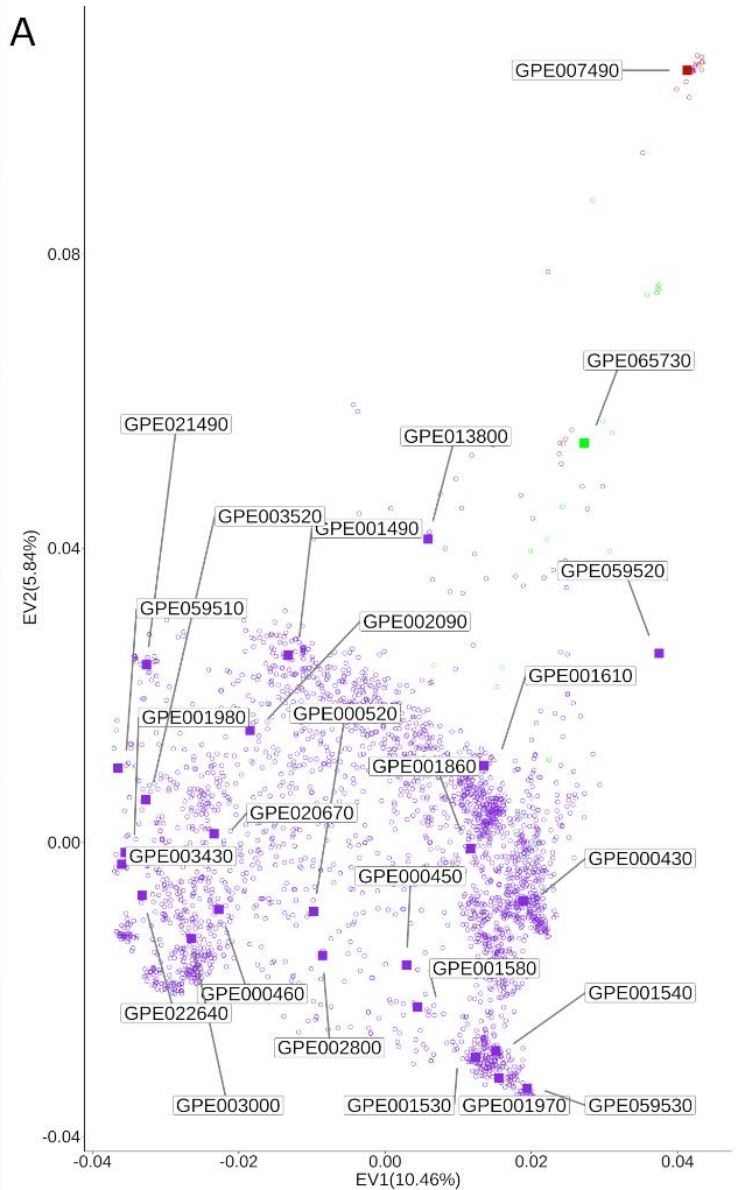
Eggplant reseq Illumina 20X

- **Reference eggplant inbred line ‘67/3’**

- **23** *Solanum melongena*

representative of the worldwide phenotypic and genetic diversity of the species

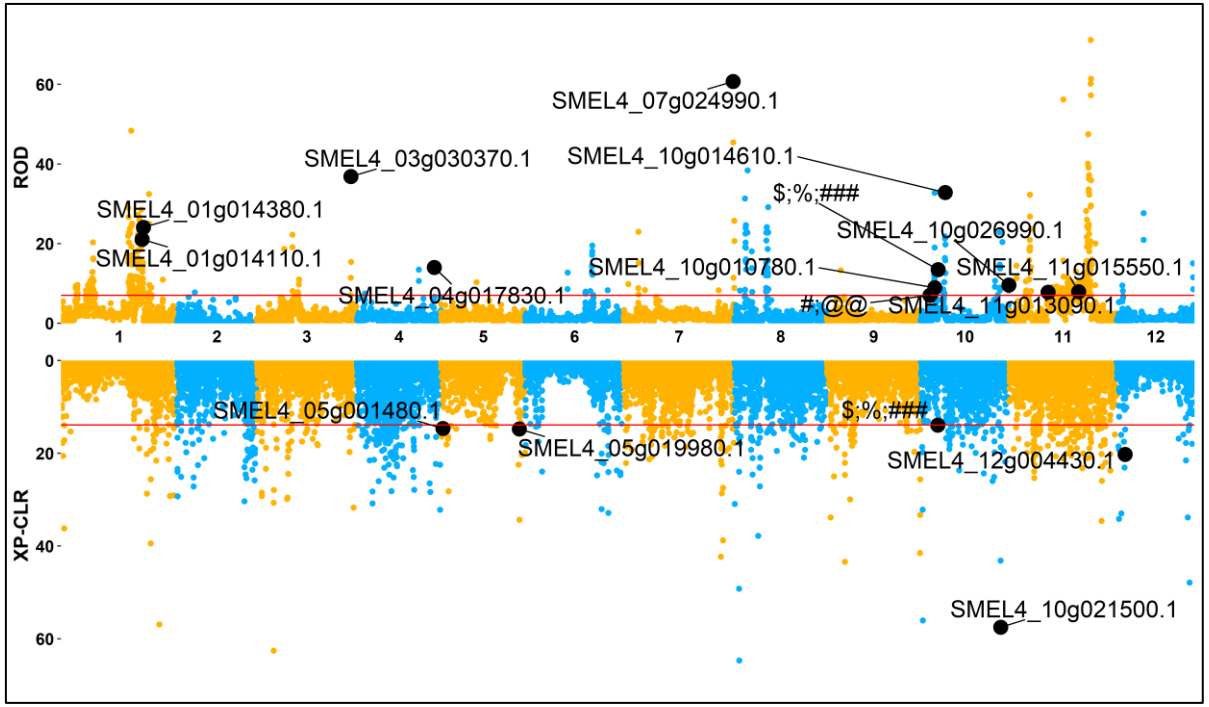
- **1** *Solanum insanum*
- **1** *Solanum incanum*



Selective sweeps and domestication

53 selective sweeps (spanning 35.3 Mb) identified

Non-anthocyanic vs anthocyanic fruits



Candidate genes identified in the SS regions.

Homologs of :

- Anthocyanin biosynthetic genes
  FLAVONOL SYNTHASE (*FLS*)
  CHALCONE ISOMERASE (CHI)
  ANTHOCYANIDIN REDUCTASE(ANR)
  DIHYDROFLAVONOL 4-REDUCTASE (FRR)

- MATE and ABC transporters mediating anthocyanin sequestration in other species

- *Prx31* peroxidase involved in anthocyanin degradation

- Transcription factors involved anthocyanin / proanthocyanidin regulation such as MYB14 and SmelMYBL1

Positions of previously discovered QTLs and QTNs

$ Barchi *et al.*, 2012;
Cericola *et al.*, 2014;
& Portis *et al.*, 2014;
# Frary *et al.*, 2014;
% Portis *et al.*, 2015;
@@ Toppino *et al.*, 2016;
Mao *et al.*, 2020;
@ Wei *et al.*, 2020a;
PI locus from Miyatake *et al.*, 2020

Several **QTLs** and **QTNs** controlling fruit anthocyanin pigmentation **co-localized with the SS on chr. 10**

PRO-GRACE

PROMOTING A PLANT GENETIC RESOURCE COMMUNITY FOR EUROPE

Thank you for your attention