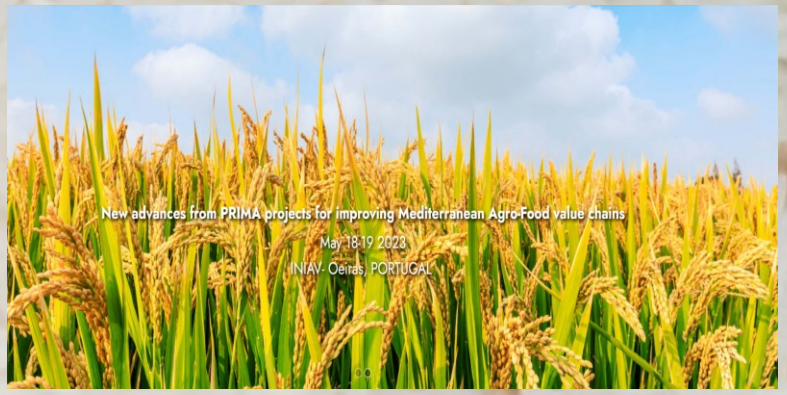


18-19 May 2023

Genome analysis and identification of molecular markers for rice (*Oryza sativa*) traceability

M. Beatriz Vieira, Pedro M. Barros, Tiago F. Lourenço & M. Margarida Oliveira

TRACE-RICE with Grant nº 1934, (call 2019, section 1 Agrofood) is part of the PRIMA Programme supported under Horizon 2020, the European Union's Framework Programme for Research and Innovation

The relevance of rice

Rice is the **primary staple food** (> 3.5 billion people)

20% of **total calories consumed** worldwide

Increase of **relevance in Europe** → Produced in the EU **Mediterranean** countries

Rice has rich genetic diversity → Great number of **varieties**

With ranges of quality and value

Rice adulteration and traceability

Rice-based foods are prone to **fraud and adulteration**

Fraudulent variety claims

Loss of value for the consumer

Jeopardize value of honest producers

Rice adulteration and traceability

Rice-based foods are prone to **fraud and adulteration**

Develop **tools** to **identify undesirable mixtures** and to **discriminate between varieties**

Molecular markers-based methods are efficient

PCR is relatively **inexpensive**

Whole-genome data is increasing

Purpose and methodology

Develop a **DNA-based method** to **discriminate varieties** that **circulate** on the **European market**

With **minimal associated costs**

Resequencing of **20 rice varieties**

Bioinformatic analysis of resequenced genomes

Identification of polymorphisms and **design of a functional detection method**

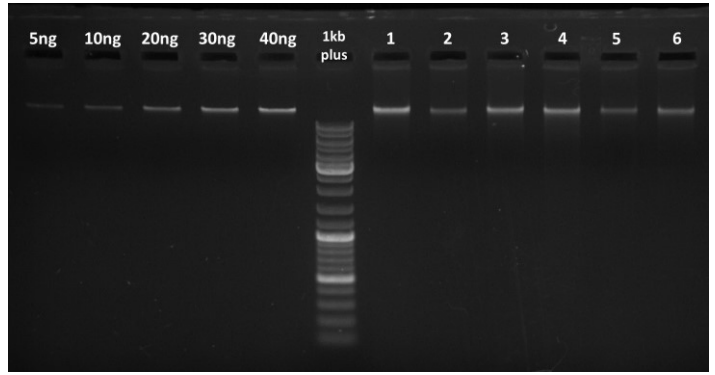
Whole-genome sequencing of the studied varieties

- Arborio
- Ulisse
- Ronaldo
- Ariete
- Giza 181
- Arelate
- Caravela
- Super Basmati
- Maçarico
- J. Sendra
- Gageron
- Manobi
- Carnaroli
- Elettra
- Teti
- CL-28
- Giza 177
- Albatros
- Lusitano
- Basmati type III

DNA extracted from **leaves**

Confirmed **DNA quality** and **concentration**

A260/A230 λDNA
A260/A280



Shipped for **sequencing**

Humanizing Genomics macrogen

Quality control

Sequencing (Illumina platform)

TruSeq DNA PCR-Free

Raw FASTQ files => **READS**

Did not germinate

DNA extracted from **seed flour**

Whole-genome sequencing of the studied varieties

Arborio
Ulisse
Ronaldo
Ariete
Giza 181
Arelate
Caravela
Super Basmati
Maçarico
J. Sendra
Gageron
Manobi
Carnaroli
Elettra
Teti
CL-28
Giza 177
Albatros
Lusitano
Basmati type III

+

Bomba
Puntal

Previously sequenced
(Reig-Valiente *et al.* (2016))

Bioinformatics analysis of whole-genome sequences

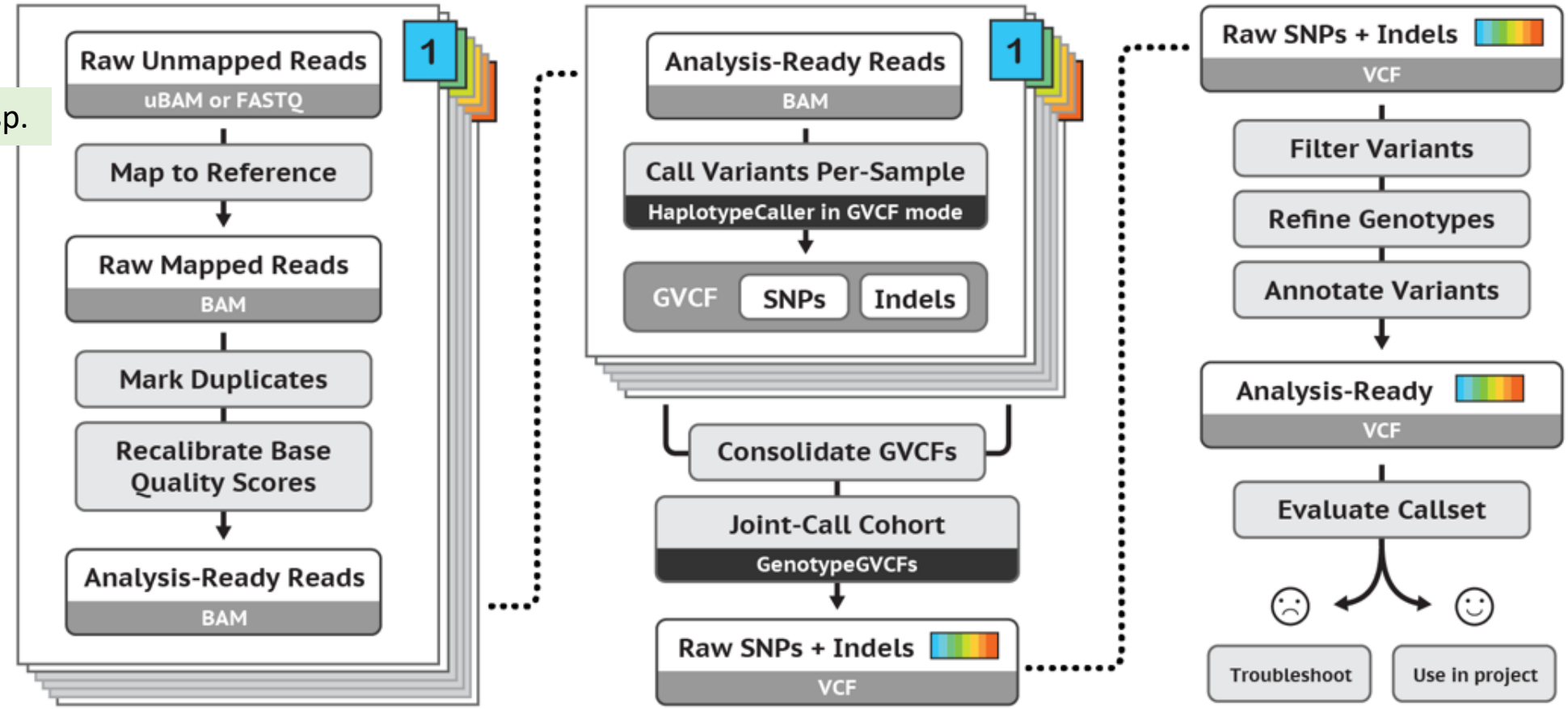
+

Short variants calling (identification of polymorphisms)

From reads to variants…

GATK workflow for variants calling

Nipponbare sp.



https://gatk.broadinstitute.org/

InDel markers were selected for varietal discrimination

Over 4M variants were extracted from the **22 genotypes**

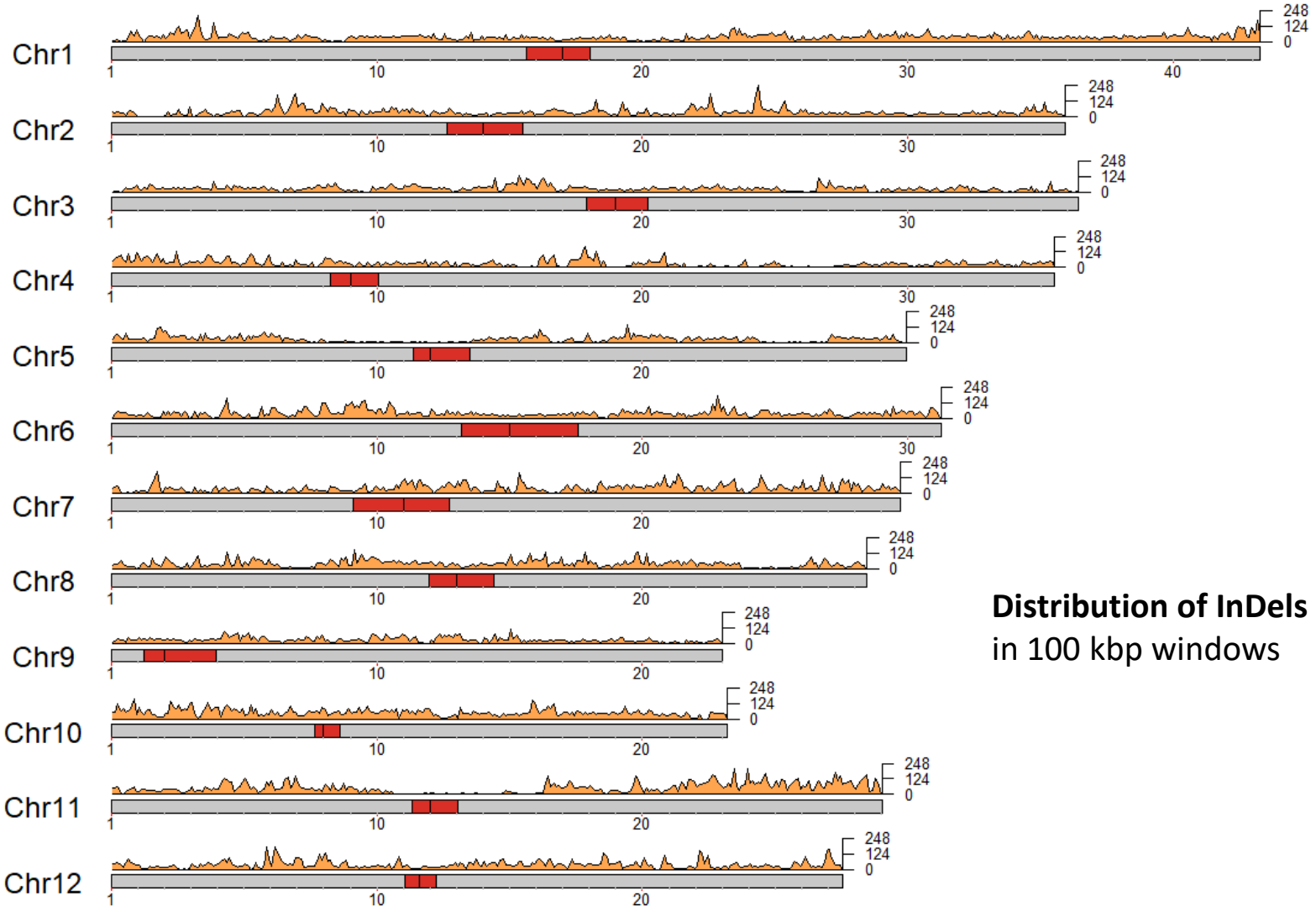
SNPs (3,612,131)

+

InDels (1,217,118 => **97,576**)

>5 bp
MAF > 0.1
Normalized

What are the **InDels** with enough **discriminating potential** to **identify** each and all of the **varieties?**



Distribution of InDels in 100 kbp windows

The Conditional-Random-Selecting (CRS) method

Yuan *et al.* (2022)

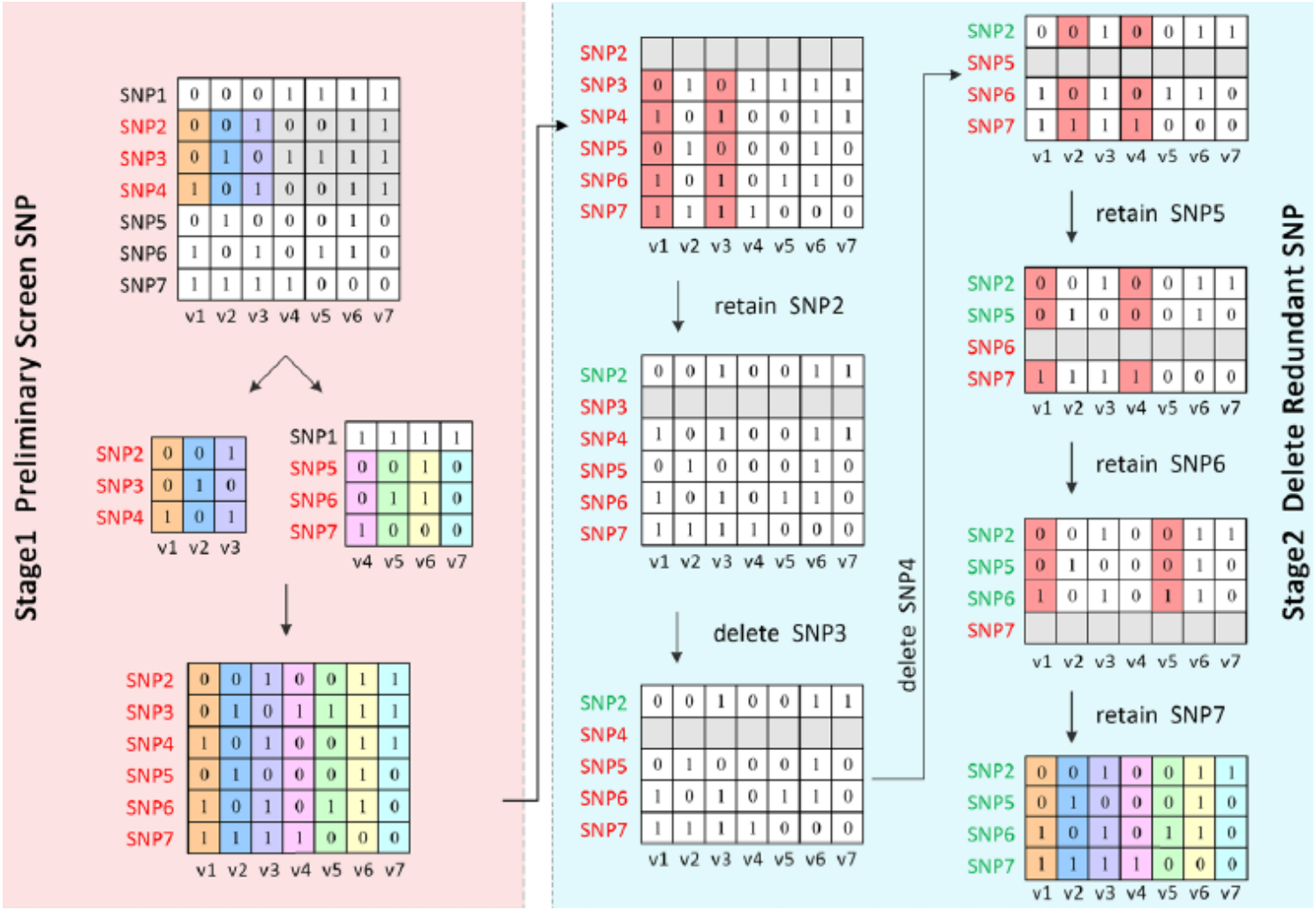
Yuan *et al. BMC Bioinformatics* (2022) 23:30
https://doi.org/10.1186/s12859-022-04562-9

BMC Bioinformatics

RESEARCH Open Access

Effective identification of varieties by nucleotide polymorphisms and its application for essentially derived variety identification in rice

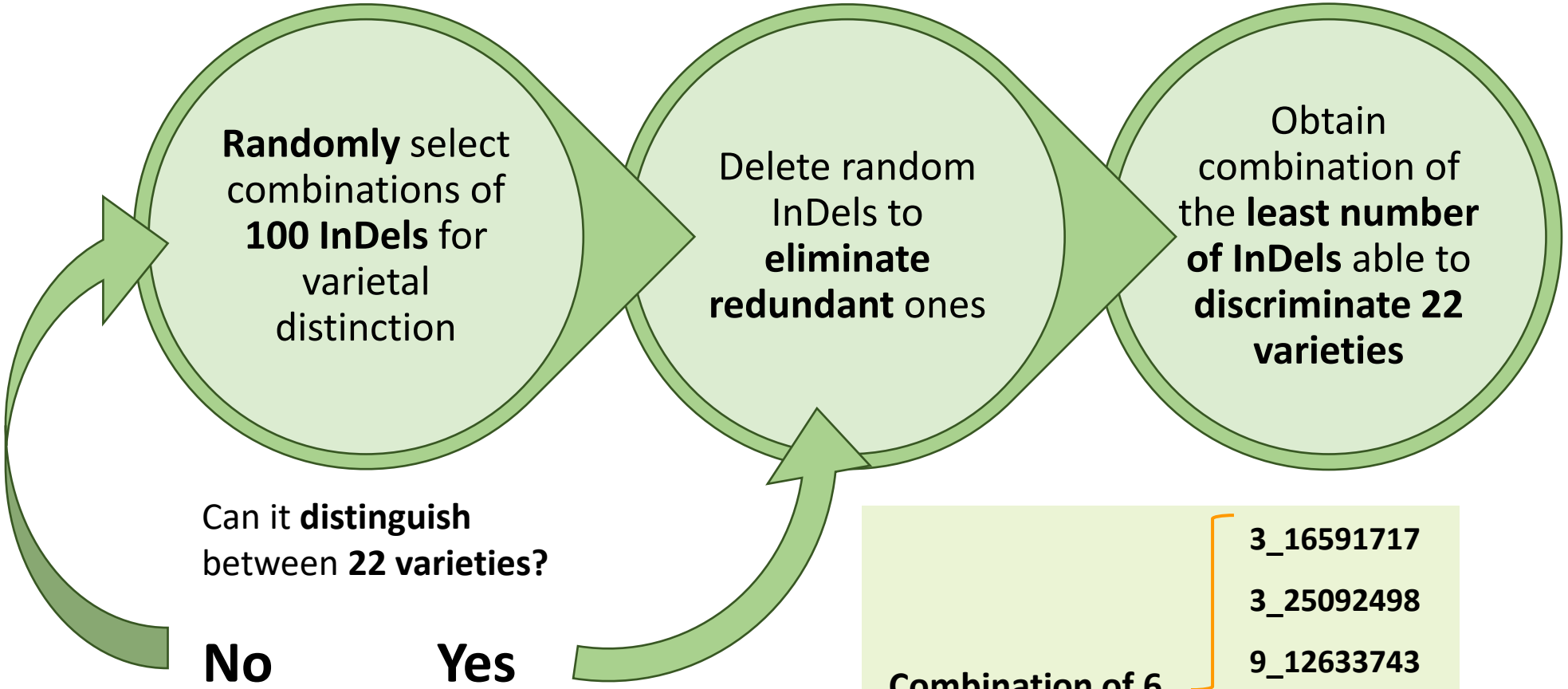
Xiong Yuan[1†], Zirong Li[1†], Liwen Xiong[1], Sufeng Song[2], Xingfei Zheng[3], Zhonghai Tang[4], Zheming Yuan[1*] and Lanzhi Li[1*]



Find the **lowest number** of **SNPs** that **discriminates** a number of **varieties**

Eliminates redundant SNPs

6 InDels were selected using the CRS method



Combination of 6
Chr_Position

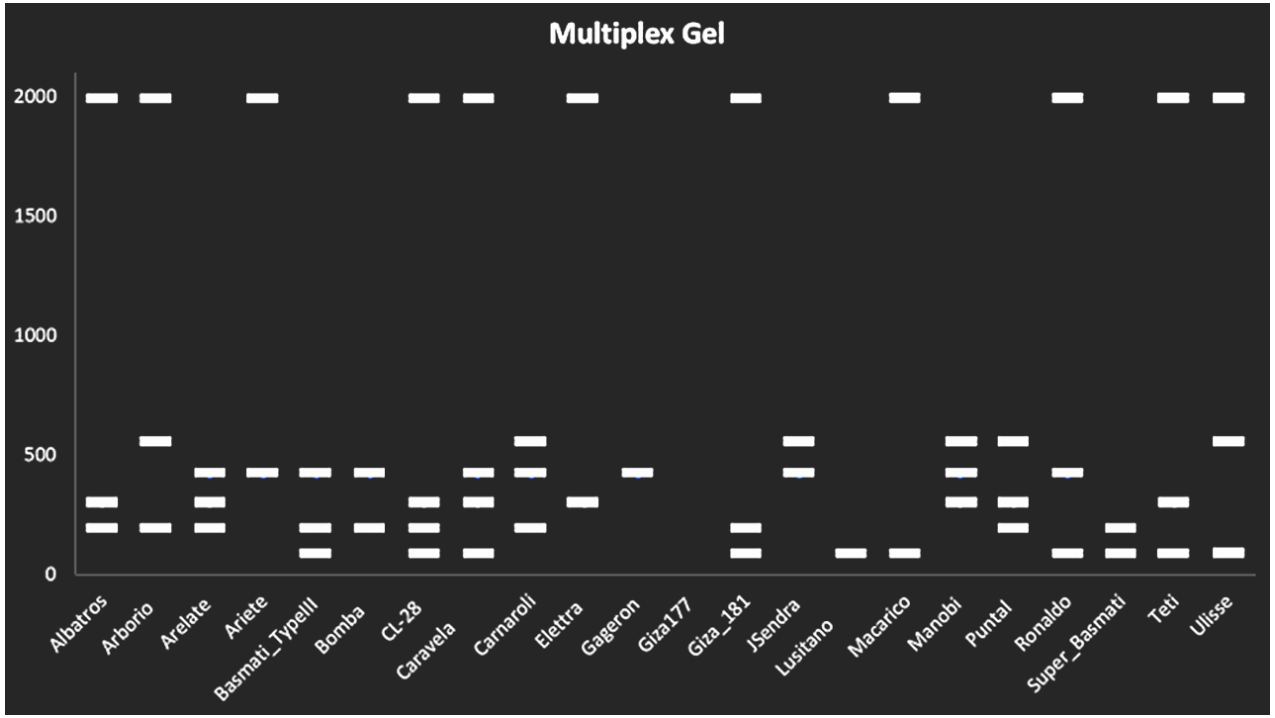
- **3_16591717**
- **3_25092498**
- **9_12633743**
- **11_7229053**
- **11_10274009**
- **11_26962480**

Predicted results confirm the discriminating potential

Presence vs **absence** detection

Primer design is **complex** (many sequence repeats)

Predicted results of an **electrophoresis** gel of the **multiplex-PCR products:**



<u>Ongoing</u>:

- "**Uniplex**"-PCR optimization
- **Single-seed DNA extraction**
- Development of **multiplex system** for **electrophoresis** detection

Take home message

Outcomes

Obtained and analysed the **whole-genome sequence** of **20 varieties**

Optimized the **CRS method** for **InDels** selection → **6 InDels**

Ongoing and future perspectives

Experimental validation of the method

Application of **multiplex-PCR** for the InDels detection

Validation of the method to **other sequenced varieties**

Acknowledgements

Prof. Margarida Oliveira

Pedro Barros

Tiago Lourenço

Plant Functional Genomics lab

Green-it
Bioresources 4 Sustainability

GPlantS
Genomics of Plant Stress

thank you!

M. Beatriz Vieira

marybtv@itqb.unl.pt