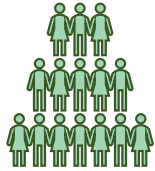


# From whole-genome data to the identification of molecular markers: steps to address rice fraudulent claims

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

Iberian Plant Biology 2023 | July 9-12



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



**Fraudulent variety claims**



Loss of value for the consumer



Jeopardize value of honest producers

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



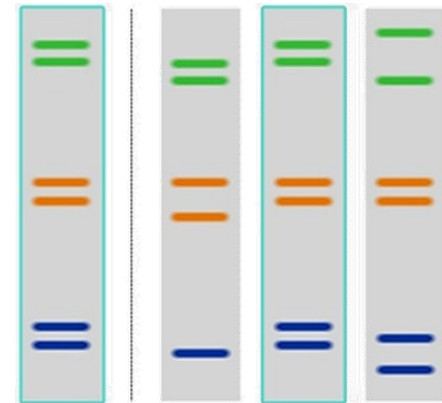
Develop **tools to identify undesirable mixtures**  
and  
to **discriminate** between **varieties**



**Molecular markers-based** methods are efficient

**PCR** is relatively **inexpensive**

**Whole-genome** data is increasing



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

With **minimal associated costs**

**Whole-genome sequencing**  
of **20 rice varieties**

**Bioinformatic analysis**  
of sequenced genomes



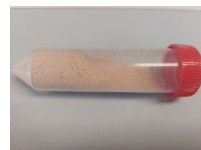
**Identification of polymorphisms**  
and **design of a functional**  
**detection method**

# 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

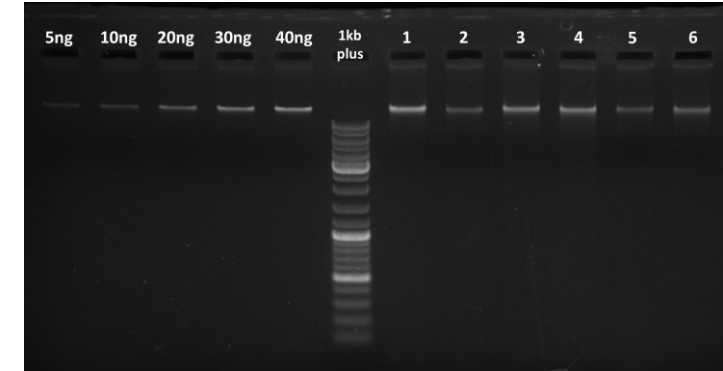


Did not germinate  
DNA extracted from seed flour



Confirmed DNA quality and concentration

A260/A230      λDNA  
A260/A280



Shipped for sequencing



Quality control

**Sequencing** (Illumina platform)

TruSeq DNA PCR-Free

Raw FASTQ files => **READS**

# The sequences of 2 additional varieties were added to the study

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))



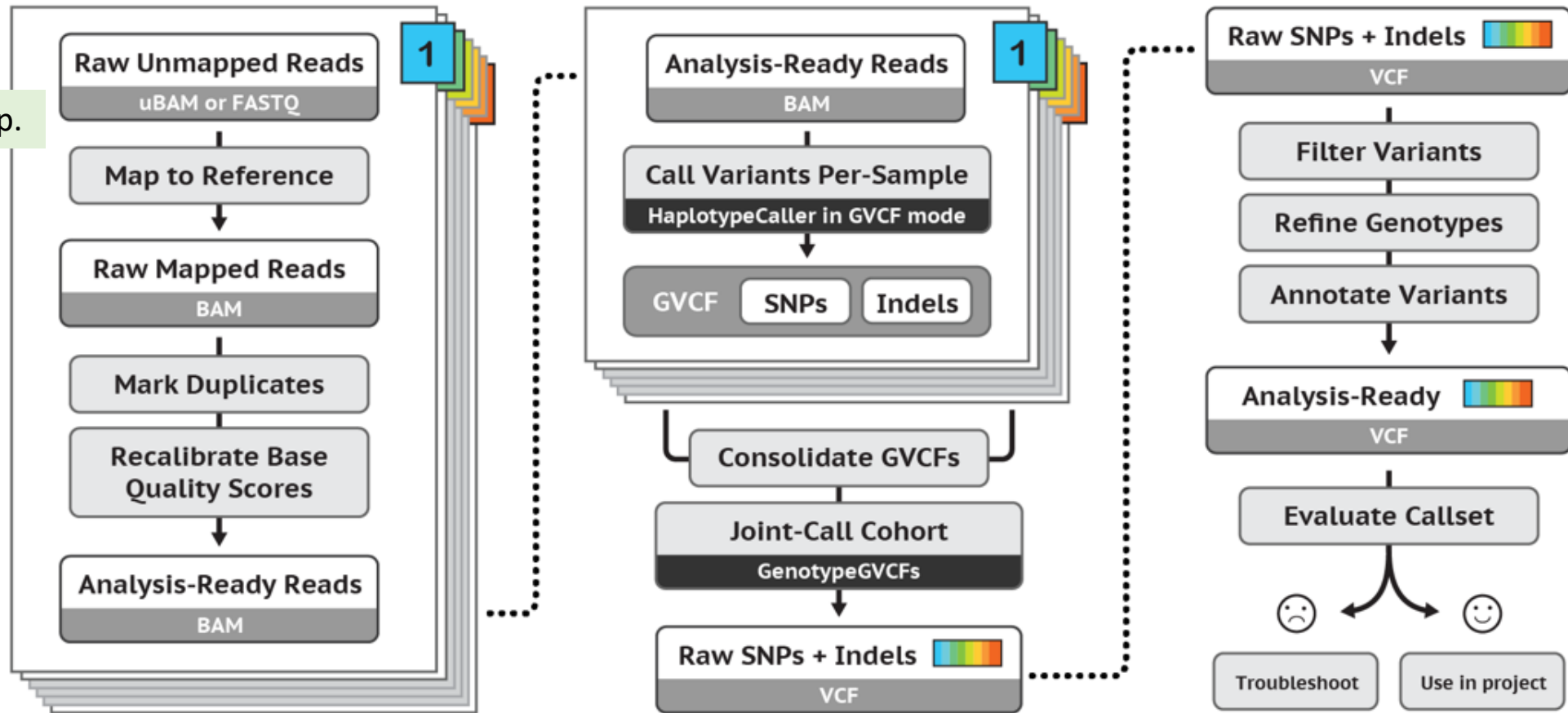
**Bioinformatics analysis of  
whole-genome sequences**



**Short variants calling**  
(identification of polymorphisms)

## GATK workflow for variants calling

Nipponbare sp.





# InDel markers were selected for varietal discrimination

Over 4M variants were extracted from the 22 genotypes

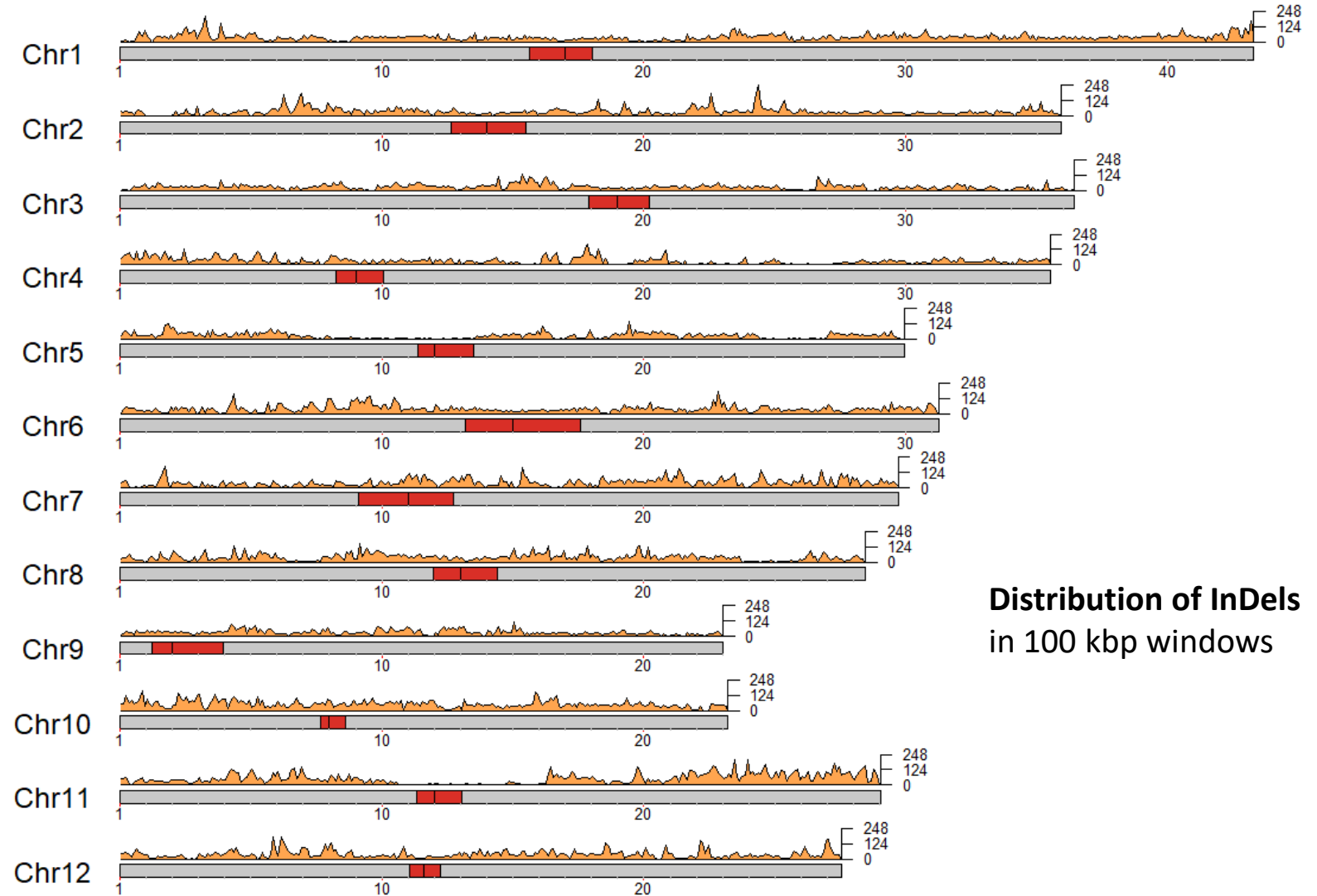
SNPs (3,612,131)



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

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

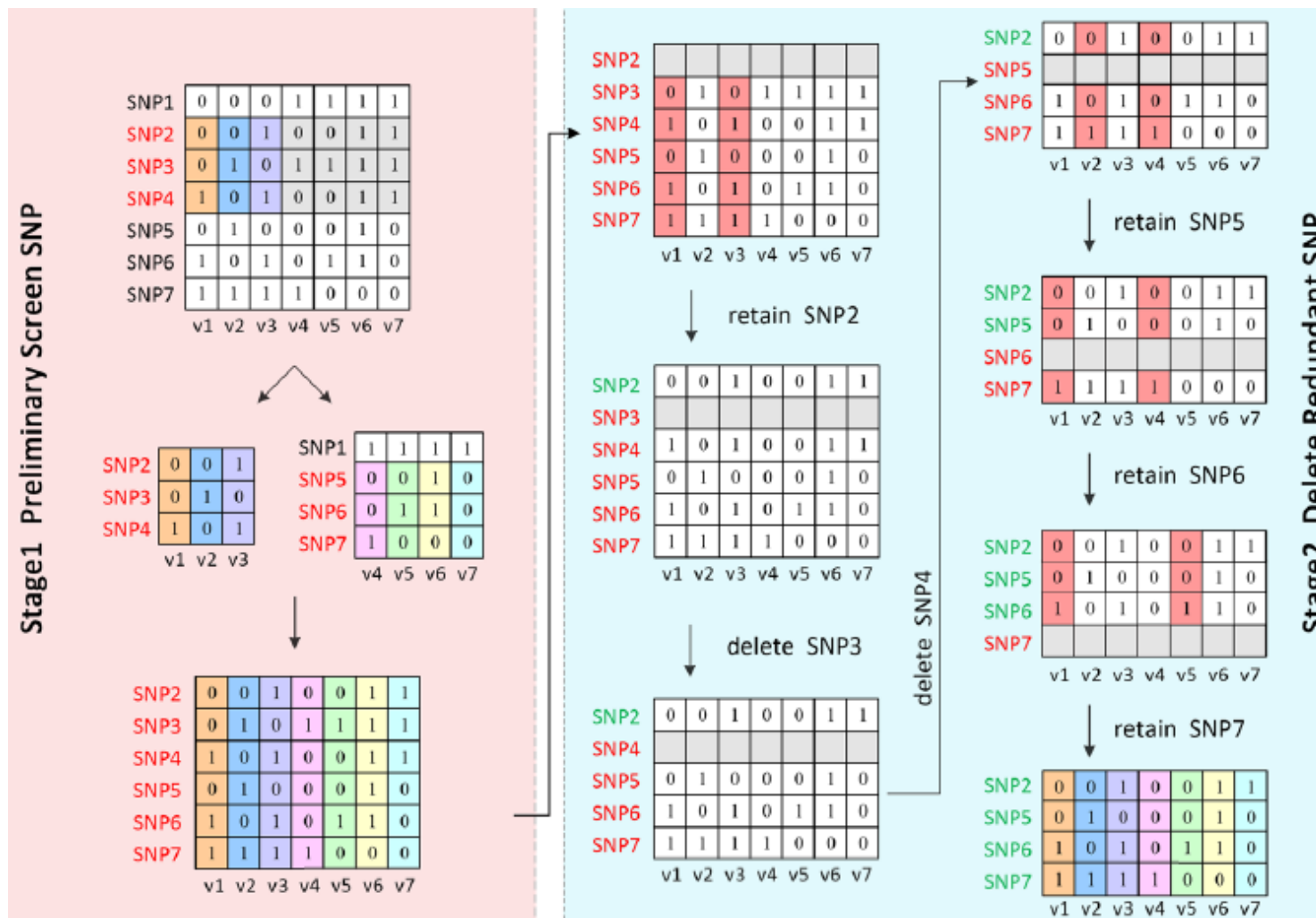


Xiong Yuan<sup>1†</sup>, Zirong Li<sup>1†</sup>, Liwen Xiong<sup>1</sup>, Sufeng Song<sup>2</sup>, Xingfei Zheng<sup>3</sup>, Zhonghai Tang<sup>4</sup>, Zheming Yuan<sup>1\*</sup> and Lanzhi Li<sup>1\*</sup>

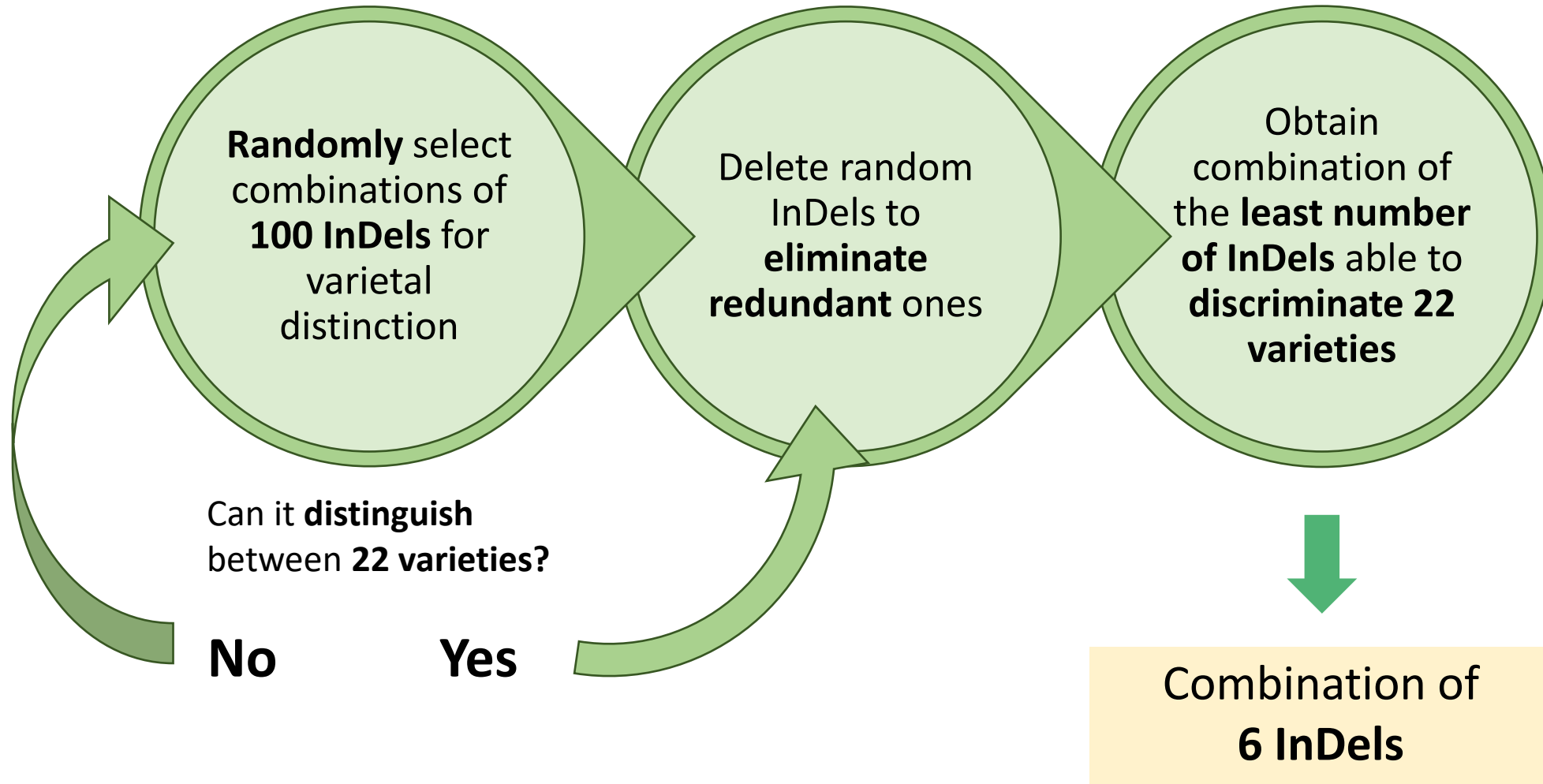
Find the **lowest number of SNPs** that **discriminates** a number of **varieties**

**Eliminates redundant SNPs**

Can it be applied to **InDels**?



# 6 InDels were selected using the CRS method



**Presence vs absence detection**

**Primer design is complex**

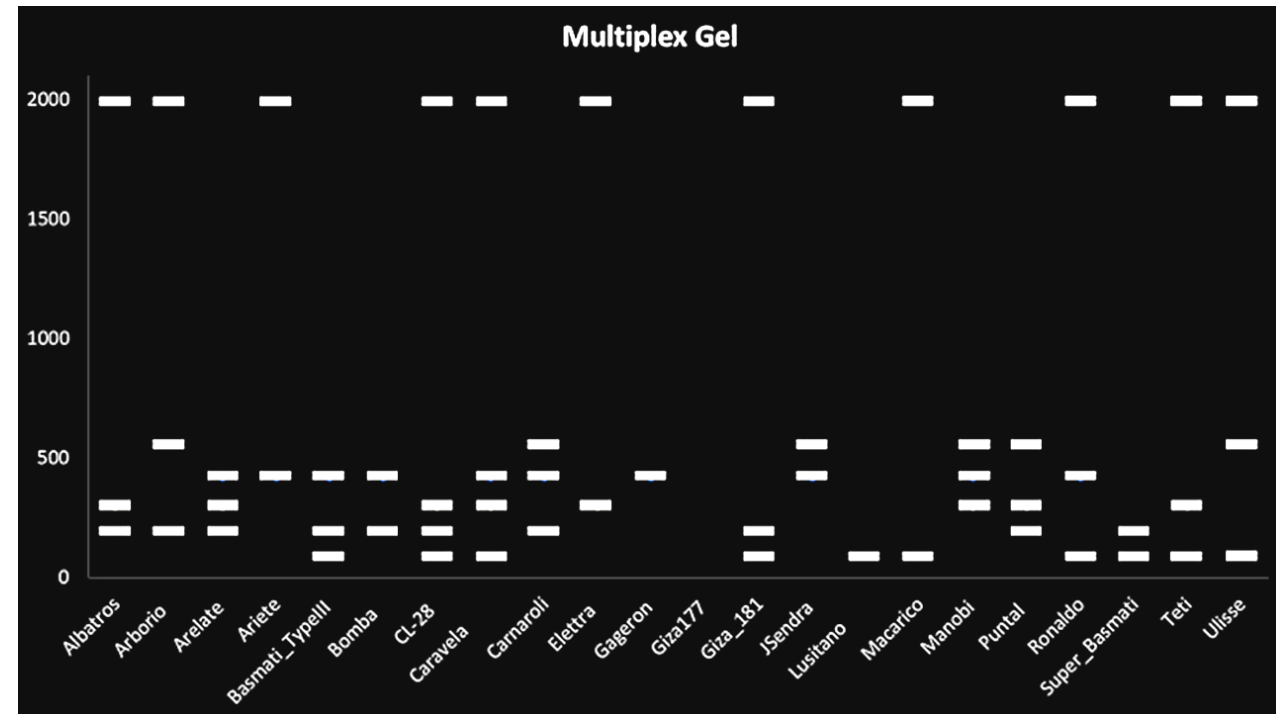
(many sequence repeats)



## Ongoing:

- Single PCR optimization
- Development of **multiplex system** for **electrophoresis detection**

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



## Outcomes

Whole-genome sequence of **20 varieties** => Obtained and analysed

CRS method => Optimized 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 for **other sequenced varieties**

# Acknowledgements



Prof. Margarida Oliveira

Pedro Barros

Tiago Lourenço

Plant Functional Genomics lab



**Green-it**  
Bioresources4Sustainability

**GPlants**  
Genomics of Plant Stress



TRACE-RICE with Grant n° 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



**itqb nova**