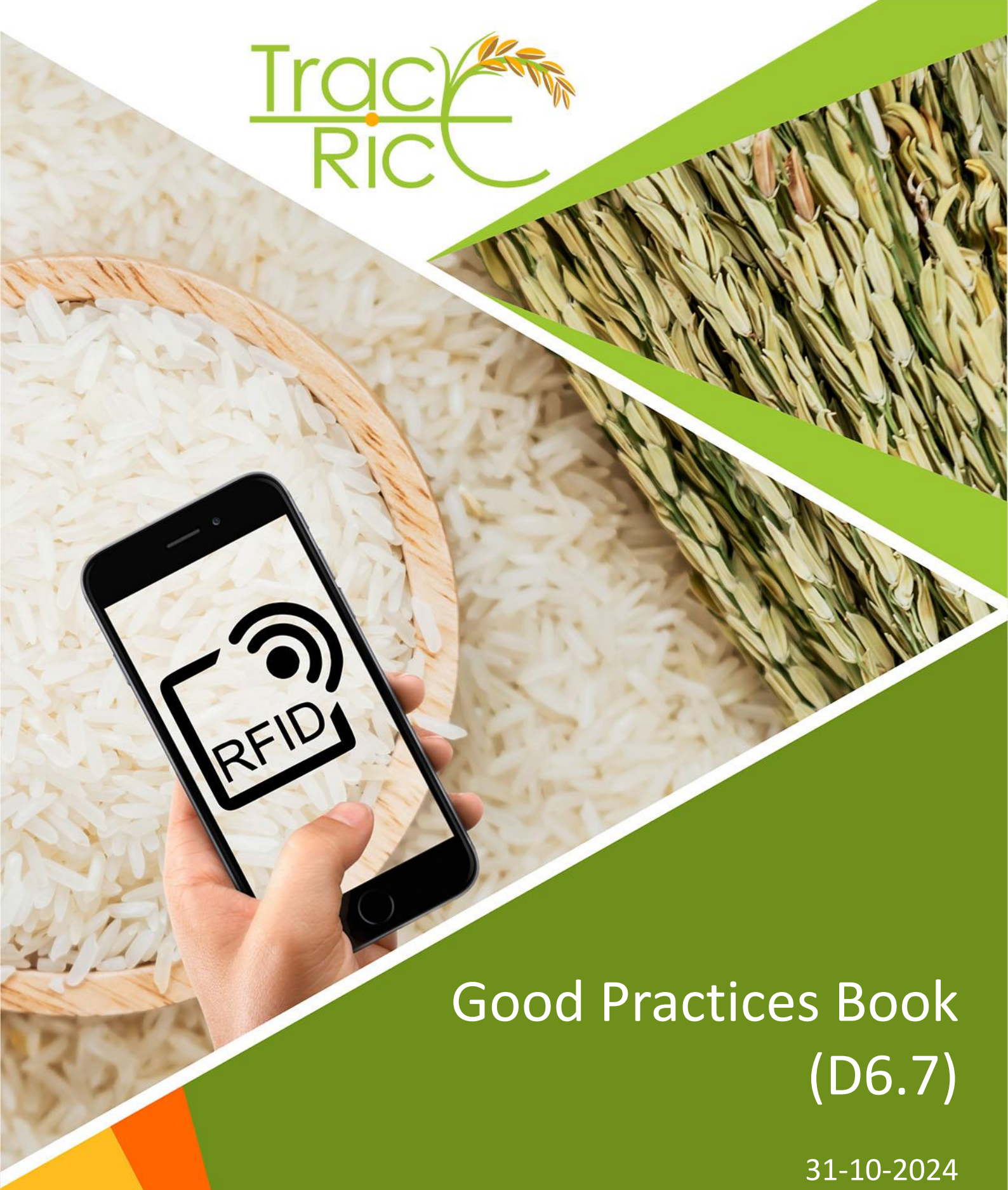


Good Practices Book (D6.7)

31-10-2024

TRACE-RICE with Grant nº 1934, is part of the PRIMA Programme supported under Horizon 2020, the European Union's Framework Programme for Research and Innovation

PRIMA
PARTNERSHIP FOR RESEARCH AND INNOVATION IN THE MEDITERRANEAN AREA

GOOD PRACTICES for rice authentication

Publication date: 31 October 2024

Authors: M. Beatriz Vieira & M. Margarida Oliveira

INDEX

ABOUT

Tracing rice and valorizing side streams along Mediterranean blockchain

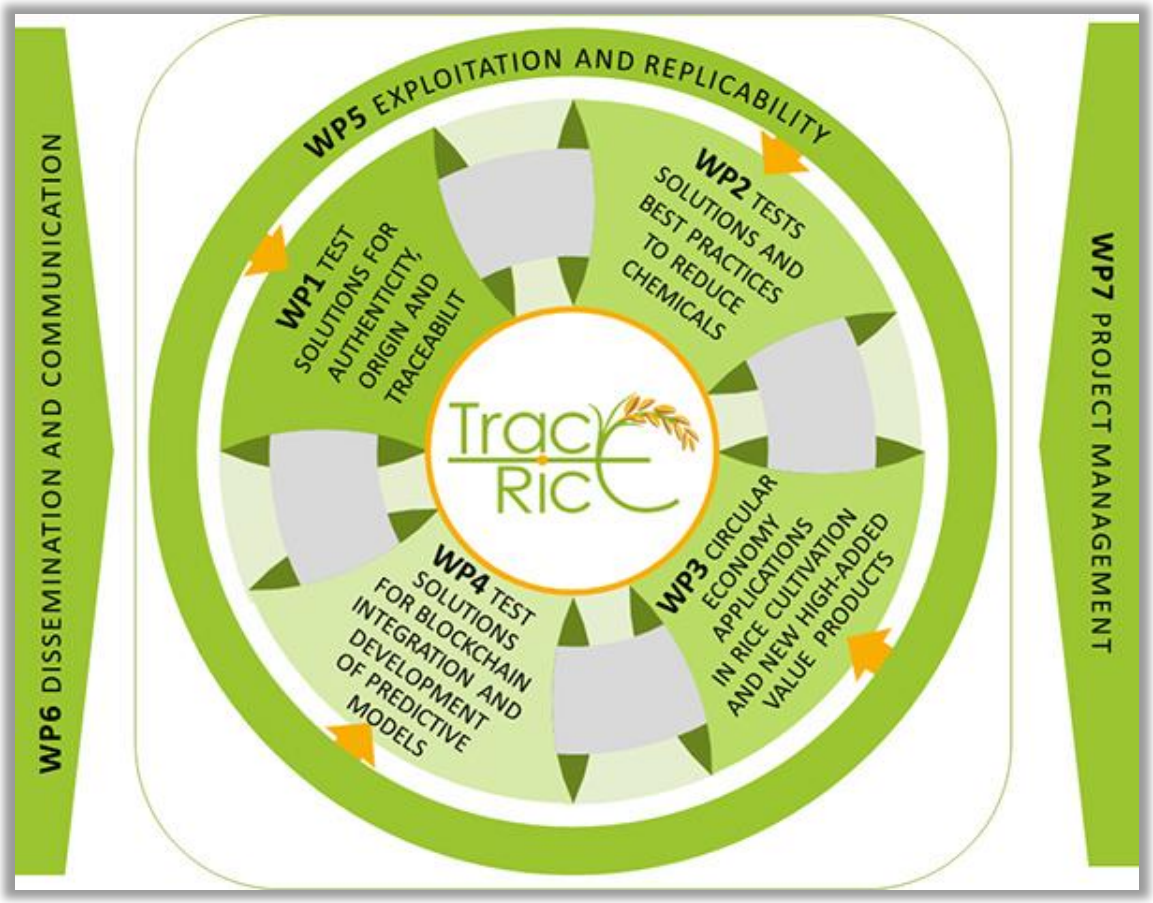
*TRACE-RICE with Grant nº 1934, (call 2019, section 1 Agrofood) is part of the **PRIMA Programme** supported under Horizon 2020*

TARGETS the adoption of cost-efficient and environmentally safe tools for traceability, authenticity, contaminant mitigation and conversion of by-products to innovative rice base food produced in the Mediterranean.

JOINS public and private stakeholders across the Mediterranean to deliver innovative solutions for fraud and safety challenges in the rice sector with an integrated full-chain approach ('farm to fork') in 48 months.

The TRACE-RICE project is structured in 7 Work Packages (WP).

All the WP has a strong relation among them to obtain the expected results, as is reflected in the following figure:



This Booklet of Good Practices was developed as a deliverable (6.7) of WP6 and contains a case-study developed in WP1.

PARTNERS

Instituto Nacional de Investigação Agrária e Veterinária, I.P.

iBET Instituto de Biologia Experimental e Tecnológica

ITQB NOVA UNIVERSIDADE NOVA DE LISBOA

CASA DO ARROZ

ernesto morgado desde1920

MATER DYNAMICS

BGI www.bqi.pt

Grupo Desarrollo

iata Instituto de Agroquímica y Tecnología de Alimentos

ALEXANDRIA UNIVERSITY

What is a GOOD PRACTICE? As a TRACE-RICE project concept

As per some of FAO criteria:

Effective

An effective way of achieving success.
A positive impact in the community.

Sustainable

Environmentally, economically, and socially sustainable.

Feasible

Technically feasible.
Easy to learn and understand.

Replicable

Potential for replicability and adaptability for similar situations

CONTEXT

Rice is the primary staple food for about half of the world's population, providing 20% of the calories consumed worldwide.

The relevance of rice in the European diet has been increasing due to its fundamental role in modern and healthy diets.

In Europe, rice is primarily produced in the Mediterranean region, where the climate and conditions support the cultivation of diverse rice varieties.

Rice has a rich genetic diversity covering many species and origins, some more valued than others.

THE PROBLEM…

Rice-based foods are prone to **fraudulent varietal claims**, where high-quality accessions are mixed/replaced by cheaper, lower-quality ones.

TRACE-RICE offers an innovative solution to fraud and safety challenges focusing on natural, healthy and tasteful rice-based foods by applying new technologies for product traceability. It provides an integrated full chain approach (from farm to fork), for raw rice and ready-to-eat rice, which will enhance the competitiveness of SMEs operating in the rice sector. In WP1, a method has been developed to specifically target fraudulent claims using a DNA-based method.

OBJECTIVE OF THIS BOOKLET

To propose an

Effective

Sustainable

Feasible

Replicable

DNA-based method to identify **rice varietal adulteration**.
Using the particular case of the EU market.

1. CHOOSING THE METHOD

foods

MDPI

Review

DNA-Based Tools to Certify Authenticity of Rice Varieties—An Overview

Maria Beatriz Vieira †, Maria V. Faustino †, Tiago F. Lourenço and M. Margarida Oliveira *

Instituto de Tecnologia Química e Biológica António Xavier (ITQB NOVA), Universidade Nova de Lisboa Av. da República, 2780-157 Oeiras, Portugal; marybtv@itqb.unl.pt (M.B.V.); mariafaustino@itqb.unl.pt (M.V.F.); tsantos@itqb.unl.pt (T.F.L.)
* Correspondence: mmolive@itqb.unl.pt
† These authors contributed equally to this work.

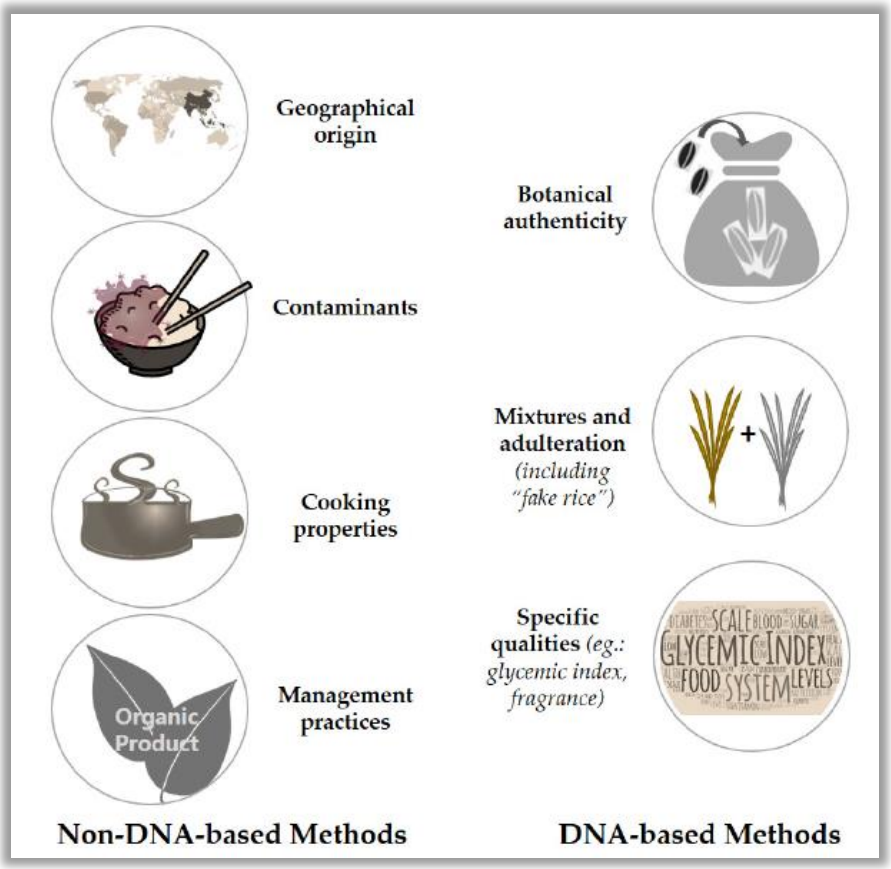
Abstract: Rice (*Oryza sativa* L.) is one of the most cultivated and consumed crops worldwide. It is mainly produced in Asia but, due to its large genetic pool, it has expanded to several ecosystems, latitudes and climatic conditions. Europe is a rice producing region, especially in the Mediterranean countries, that grow mostly typical *japonica* varieties. The European consumer interest in rice has increased over the last decades towards more exotic types, often more expensive (e.g., aromatic rice) and Europe is a net importer of this commodity. This has increased food fraud opportunities in the rice supply chain, which may deliver mixtures with lower quality rice, a problem that is now global. The development of tools to clearly identify undesirable mixtures thus became urgent. Among the various tools available, DNA-based markers are considered particularly reliable and stable for discrimination of rice varieties. This review covers aspects ranging from rice diversity and fraud issues to the DNA-based methods used to distinguish varieties and detect unwanted mixtures. Although not exhaustive, the review covers the diversity of strategies and ongoing improvements already tested, highlighting important advantages and disadvantages in terms of costs, reliability, labor-effort and potential scalability for routine fraud detection.

check for updates

Citation: Vieira, M.B.; Faustino, M.V.; Lourenço, T.F.; Oliveira, M.M.

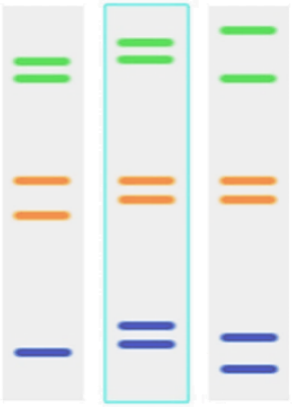
In 2022, **TRACE-RICE** published a review on diverse methods being used for authenticating rice varieties

Considering the specific goals, different methods may be employed. For varietal identification, we consider **DNA-based** methods, a **Good Practice**



Vieira & Faustino *et al.* 2022. https://doi.org/10.3390/foods11030258

From the DNA-based methods, we chose polymerase chain reaction (**PCR**) to amplify specific molecular markers or genomic variants. This strategy has the additional benefit of being relatively inexpensive and applicable to any lab using simple biomolecular equipment.



SETTING UP YOUR STUDY

In order to develop this method, multiple choices must be considered:

- **Which varieties** to use?
- How to **extract the DNA?**
- **Which markers** to employ?
- How to **visualize the markers' pattern?**
- What are the **constraints** associated with each step?

2. SELECTING THE VARIETIES for experimental design

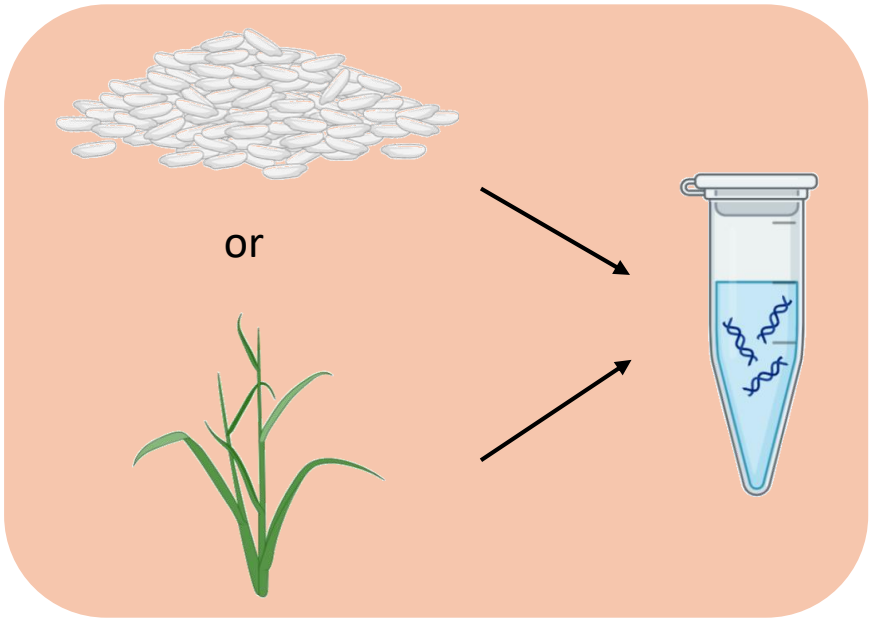
The varieties to use in the study always depend on the specific goal. Usually, it is desirable to ensure diversity of rice varieties. Also, the incorporation of basmati or other aromatic varieties can be important, given the fraud cases already described.

TRACE-RICE focused on accessions circulating (or soon to be released) in the Mediterranean market.

Considering the particular objectives of the TRACE-RICE project, this selection considered 22 varieties, 20 of which were sequenced through the project.
This table comprises the list of targeted accessions:

	Variety	Relevant in	Origin
1	Ariete	Portugal	Italy
2	Teti	Portugal	Italy
3	Ronaldo	Portugal	Italy
4	CL-28	Portugal	Italy
5	Maçarico	Portugal	Portugal
6	Caravela	Registered in PT 2021	Portugal
7	Arborio	Italy	Italy
8	Carnaroli	Italy	Italy
9	Elettra	Italy	Italy
10	Ulisse	Italy	Italy
11	J. Sendra	Spain	Spain
12	Bomba	Spain	Spain
13	Puntal	Spain	Spain
14	Gageron	France	France
15	Manobi	France	France
16	Arelate	France	France
17	Giza 177	Egypt	Egypt
18	Giza 181	Egypt	Egypt
19	Super Basmati	Europe	Pakistan
20	Basmati Type III	Europe	India
21	Albatros	Portugal	Italy
22	Lusitano	Portugal	Italy

3. DNA EXTRACTION



High-quality DNA is crucial for high-throughput techniques. However, for PCR amplification, several simpler methods can be employed, including the use of extraction kits.

Part of TRACE-RICE's project involved obtaining the whole-genome sequence of 20 varieties and therefore, we used and found the best results by following a protocol of extraction using the reagent Cetyltrimethylammonium bromide (CTAB).

An optimized CTAB extraction method was used, either for leaves (1), or seed flour (2), when the former was not possible.

The quality of the DNA was inferred by the ratio of absorbance at 260 and 280 nm, and at 260 and 230 nm (3). A260/A280 ratio of 1.7–2.0 and A260/A230 ratio of 2.0-2.2 is ideal

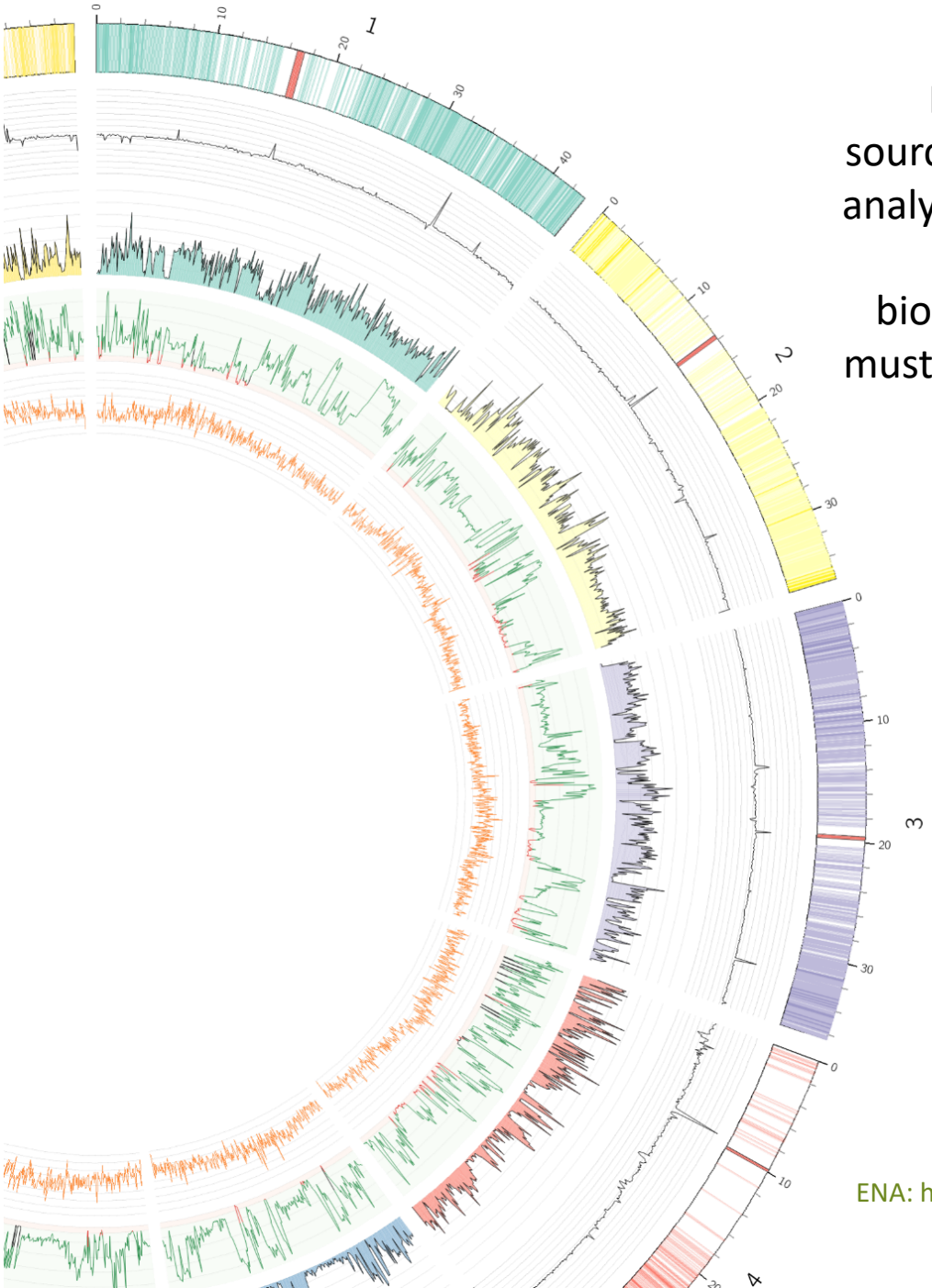
When developing the method, the biological material must have no mixtures, otherwise we cannot identify them or adulteration in a test.

(1) Doyle and Doyle, 1990, DOI: 10.1007/978-3-642-83962-7_18
(2) Datukishvili *et al,.* 2010, DOI: 10.1111/j.1365-2621.2010.02261.x
(3) Measured in a NanoDrop, Thermo Fisher Scientific equipment

4. GENOMIC DATA ANALYSIS

To apply/develop this method, the knowledge of the genome of each accession is imperative.

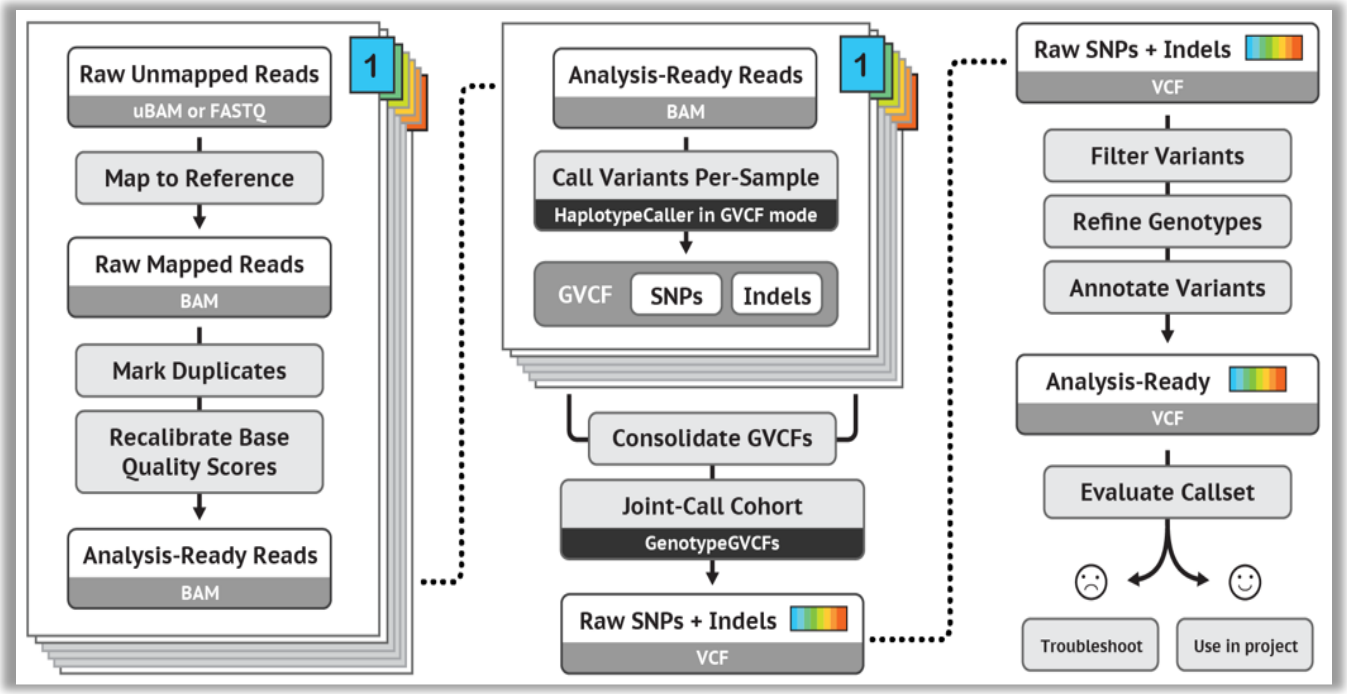
Files with the sequencing results of numerous rice genotypes are already available at public databases, such as the European Nucleotide Archive (ENA). Two of the genomes reported in this booklet were obtained from the ENA resource.



Regardless of the source of the data, an analysis of the **quality** and other bioinformatics steps must be conducted to obtain genomic variants.

ENA: https://www.ebi.ac.uk/

We suggest following the "**Best practices for** single nucleotide polymorphisms (**SNP**) **and** insertions and deletions (**InDel**) **discovery in germline DNA**" using the **GATK** set of bioinformatics tools.



From this workflow, a file containing the genomic variants (specifically, SNPs and InDels) of all the inputted samples, is obtained.

The next step is to select **which molecular markers** to use in the method. As described in Vieira & Faustino *et al.* (2022), the most used markers for adulteration identification are microsatellites and SNPs.

In this case-study, considering that TRACE-RICE aimed for the use of PCR and detection by electrophoresis gel, **InDels longer than 5 bps were employed**.

https://gatk.broadinstitute.org/

Vieira & Faustino *et al.*2022. https://doi.org/10.3390/foods11030258

5. SELECTING THE DISCRIMINATING INDELS

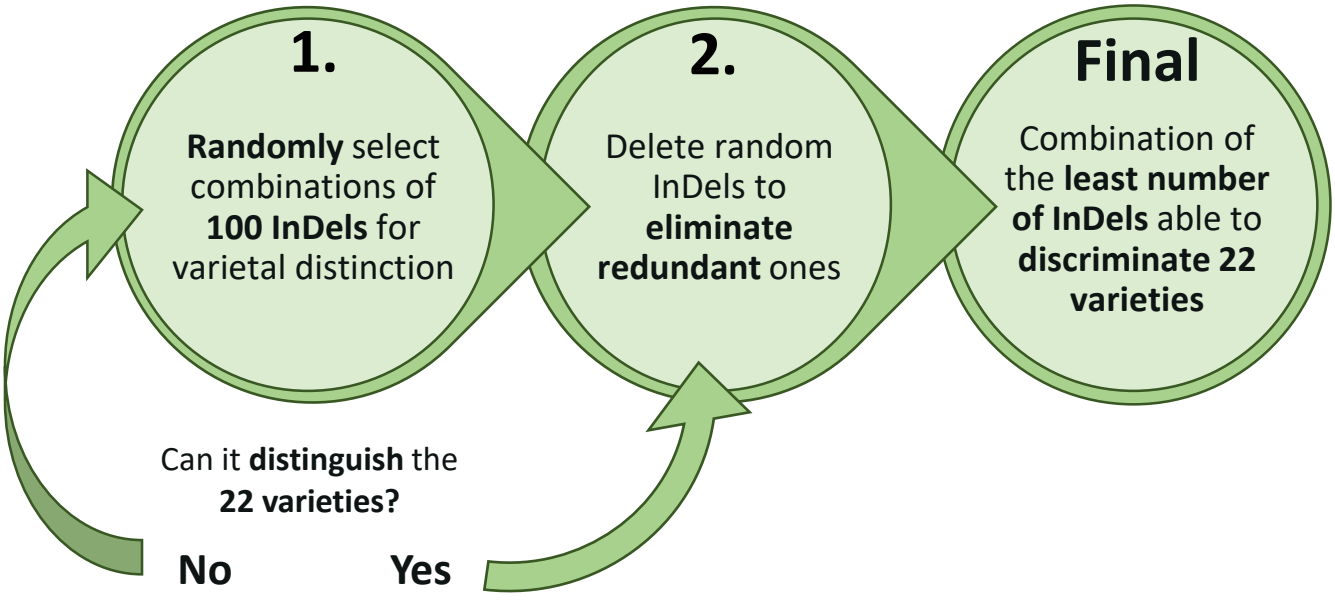
After defining the type of markers to use, it is necessary to identify which InDels have sufficient discriminating for varietal identification.

We suggest the application of the **Conditional Random Selection (CRS) method**. This method was developed by Yuan *et al.* (2022) to select SNPs for **rice varietal identification,**, although it was not experimentally validated.

The CRS method has two main stages:

1. Preliminary screening of the variants;
2. Deletion of redundant polymorphisms.

A bioinformatics R script was adapted for **TRACE-RICE**'s case-study comprising two key steps:

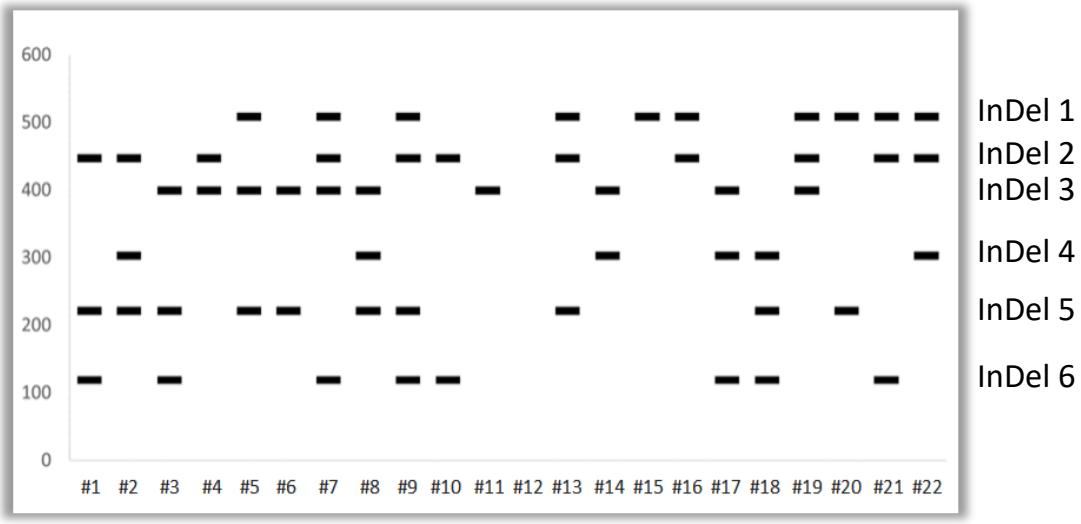


From a final list of **97,576 InDels**, the result of the CRS method application was a specific combination of **six InDels** with enough discriminating potential.

Considering the objectives of the TRACE-RICE project and the low number of InDels needed, efforts were directed toward developing a multiplex PCR-based method.

In this way, only one reaction is needed to detect all the markers, reducing time and overall cost.

Considering a pattern of presence *vs.* absence of each InDel in a certain variety, an *in silico* prediction of a multiplex gel can be represented as in the image:



Varieties

(numbers represent each accession as described in the table in page 14)

6. PRIMER DESIGN

The first step to any PCR optimization is Primer Design.
In our case-study, InDels are the target of the amplification, therefore, one of the primers (in each pair) flanks the target region (of either an insertion or deletion).
The following aspects must be taken into consideration when designing primers for multiplex PCR:

Length

- Primer length (18-25 bps)
- Multiplex PCR product sizes should be selected so that each one is visible in an electrophoresis gel

In our case-study, the six pairs of primers were individually designed using Primer3 and Blast from NCBI.
Primer size varied between 20 and 29 bps and product size between 119 and 508 bps.

Melting temperature

- Primers to use in the same reaction should have a similar Tm when possible. Variation between 3-5 ºC is acceptable

Primer melting temperature (Tm) influences the annealing temperature to use in the PCR. Maintaining the Tm around 60ºC is desirable but not always possible, as in this case-study, considering that InDels are present in repetitive regions of the genome and often with high GC content.

Primer3: https://www.ncbi.nlm.nih.gov/tools/primer-blast/

Specificity

- Primers should be highly specific to the target sequence since in a single pool there is competition between primers

Specificity is a key characteristic of the primers in multiplex reactions. Depending on the target sequence, this can be a difficult part of the designing process. The use of InDels proved to complicate the design, given the high prevalence of repetitive regions with similarities to the target of the primer.

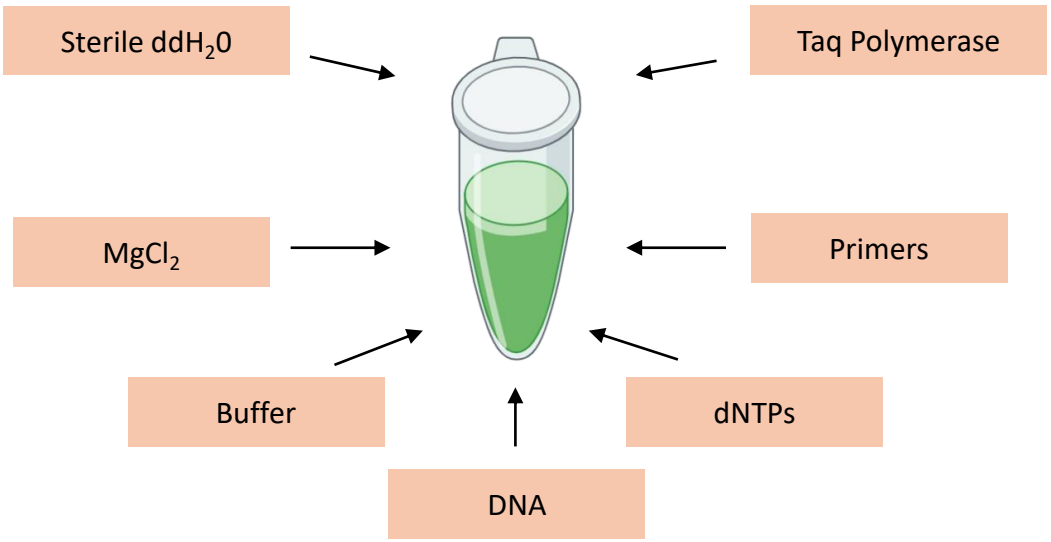
Recommendations for primer design

- Tm calculation: 2°C x (A+T) + 4°C x (G+C) (normally output from the primer designing online tools)
- Avoid complementarity in the 2–3 bases at the 3' end of the primer pairs
- Avoid mismatches between the 3' end of the primer and the template
- Avoid runs of over 3 GCs at the 3' end of the primer
- Avoid complementarity within primers and between the primer pairs in a multiplex reaction
- Avoid a T as ultimate base at the 3' end
- Ensure primer sequence is unique for the template sequence
- Use a concentration of 0.1–1.0 µM of each primer. For many applications, a primer concentration of 0.2 µM will be sufficient

7. PCR OPTIMIZATION

Once all the primers are designed, they must be tested/validated experimentally. Ideally, in the end, each PCR product's amplification pattern should correspond to the one predicted *in silico*. Multiple reasons can explain some cases when it doesn't, and we have highlighted some of them in the "Troubleshooting" section.

How to prepare your **reaction mix** (per tube):

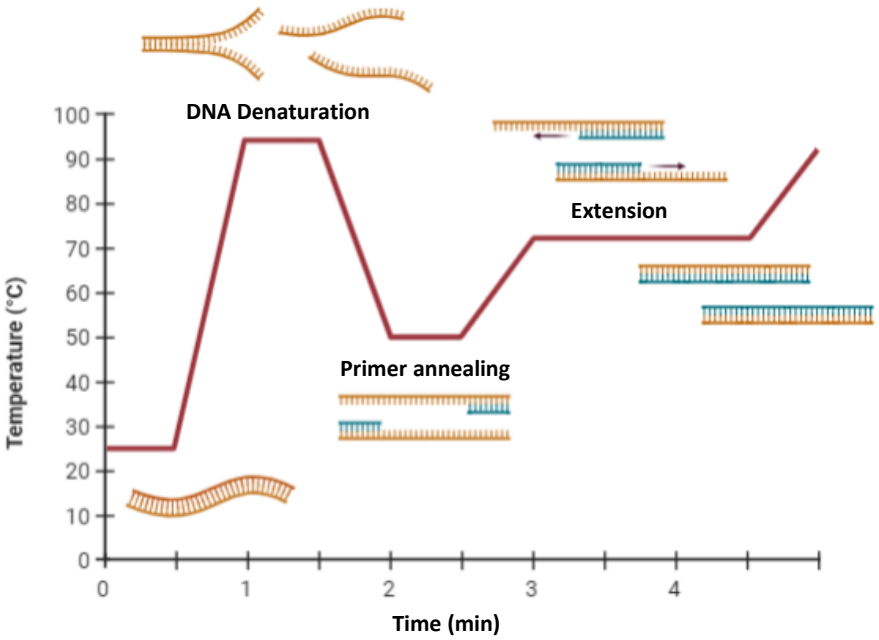


Component	Quantity for a 25µL reaction	Considerations
ddH_2O	to 25 µL	Double distilled water should be sterile
$MgCl_2$	1-5 mM	Should be optimized. Usually, higher concentration provides more amplification, but less specificity
Taq Buffer	1 X	Is specific for the Taq polymerase being used
dNTPs	0.2-0.4 mM	Higher concentration increases yield
Primers	0.2-0.5 µM	Crucial part of the optimization
Taq Polymerase	1 U	Follow manufacturer's instructions
DNA	5-500 ng	Too much DNA can inhibit amplification

Besides the reaction components, the PCR steps may also be optimized. A standard reaction contains 30-35 cycles of 3 main steps:

PCR Step	Temperature	Duration
Denature template	95 °C	1 min
Anneal primers	55-65 °C	30 s - 2 min
Extension	72 °C	30 s - 2 min

The following image represents the main steps of one PCR cycle, showing what happens at molecular level.



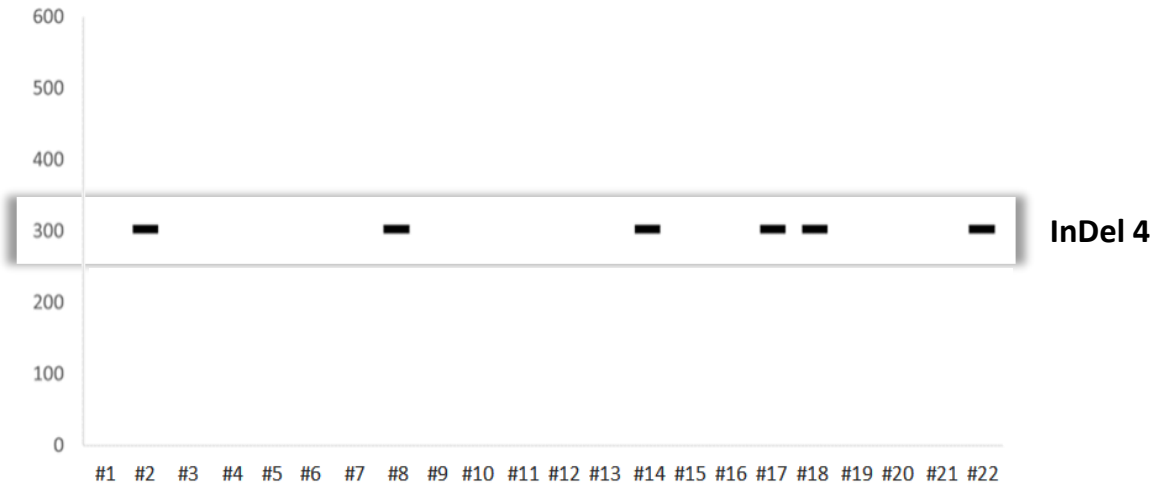
These steps may differ according to the selected Taq polymerase, and the annealing temperature should be optimized according to the melting temperatures of the primer pairs. The extension duration depends on the size of the PCR products.

Always consider that, if the goal is to implement multiplex PCR, the annealing temperatures used for each primer pair should be as close as possible, even though multiplex Taq polymerases sometimes tolerate a larger interval of temperatures.

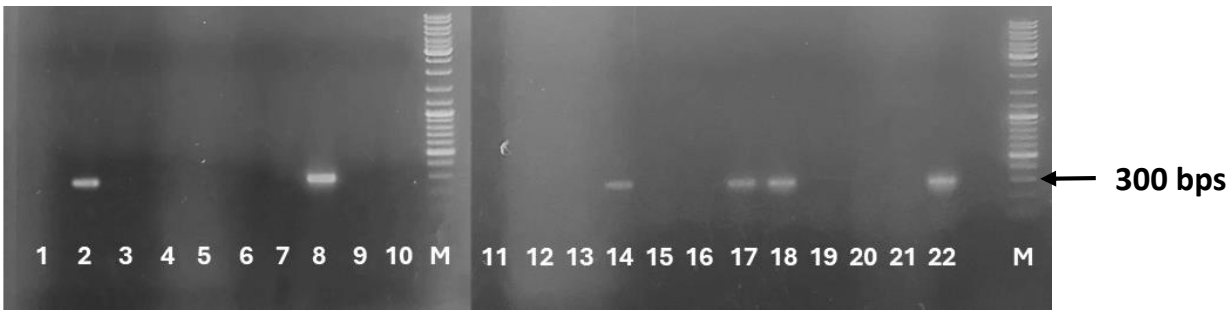
In our case-study, the 'simplex' PCR optimization was first performed for each InDel. The annealing temperatures for the six primer pairs ranged between 58 and 64 ºC, which is an amplitude higher than optimal for a multiplex reaction. The final extension time was 30 seconds considering the expected size of the bands.

The PCR products were revealed in an electrophoresis agarose (1%) gel stained with Ethidium Bromide (EtBr). Other stainings may also be used.

The expected *in silico* results of the detection of InDel 4 is shown in the following scheme.



Each number from 1 to 22 corresponds to one variety and, in the following gel image the prediction was confirmed. A molecular size marker was used to assess the size of the amplified band (in this case, 303 bps).



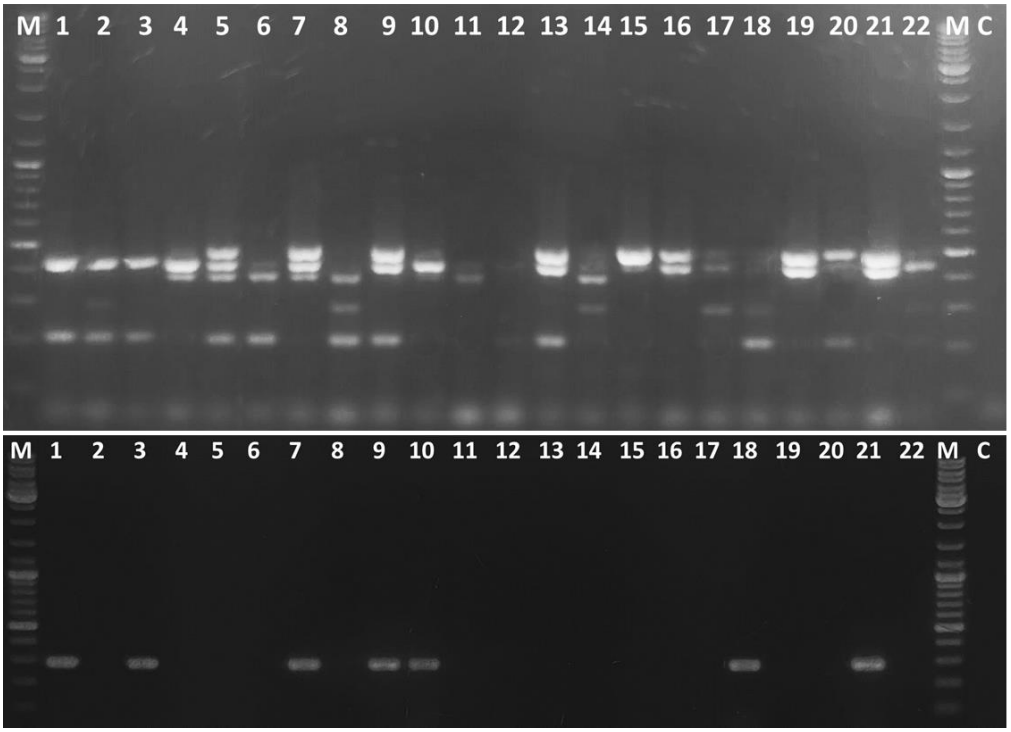
8. MULTIPLEX PCR

As previously mentioned, the final goal of this method was to detect all of the patterns of 6 InDels through multiplex PCR. This implied further steps of optimization and the use of specific buffers.

After several tests, a multiplex-optimized buffer was chosen, the Multiplex PCR 5X Master Mix (NEB, USA). The instructions for the PCR program were followed and the only changes needed were regarding the primers concentration and selected annealing temperature.

One issue we encountered involved a primer pair (detecting InDel 6) that, contrary to expectations, amplified in all varieties. Adjustments to the PCR program and modifications to the reaction mixture composition did not successfully eliminate the false positives. To address this, InDel 6 was removed from the multiplex and analyzed independently.

The following image shows the final multiplex PCR results (discussed in the following page), along with the single PCR for InDel 6.



The optimized annealing temperature was 64 °C and the final concentration of primers varied between 0.2 and 0.48 µM. The PCR program corresponded to: 1 min of denaturation at 95 °C, followed by 35 cycles of 20 secs at 95 °C, 1 min at 64 °C, 20 secs at 68 °C, and a final extension period of 5 minutes at 68 °C, and pausing at 10 °C.

Most of the expected results were experimentally validated however, there were exceptions. Still, it was possible to discriminate 16 varieties from all others using the multiplex reaction plus a single additional PCR.
The other six accessions were grouped into 3 pairs each of them with the same fingerprint (not discriminated). For these cases, different grain biometry or biochemical differences (eg., starch composition) may solve the problem.

The table below shows the final patterns of absence vs. presence ("0" vs. "1") of bands comparing with the expected results (in green when corresponding and in yellow when unexpected). Furthermore, it shows the additional strategies applicable for the non discriminated pairs (**a**- biochemical data (starch composition and glycaemic index); **b**- biometric data (length and chalkiness percentage of the grain); **c**- biochemical and biometric data (grain features and cooking parameters).

	a		**a**											**b**		**c**	**b**		**c**			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	Albatros	Arborio	Arelate	Ariete	Basmati Type III	Bomba	Caravela	Carnaroli	CL-28	Elettra	Gageron	Giza 177	Giza 181	J Sendra	Lusitano	Maçarico	Manobi	Puntal	Ronaldo	Super Basmati	Teti	Ulisse
InDel 1	0	0	0	0	1	0	1	0	1	0	0	0	1	0	1	1	0	0	1	1	1	0
InDel 2	1	1	1	1	1	0	1	0	1	1	0	0	1	0	0	1	0	0	1	0	1	1
InDel 3	0	0	0	1	1	1	1	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0
InDel 4	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	1
InDel 5	1	1	1	0	1	1	0	1	1	0	0	0	1	0	0	0	0	1	0	1	0	0
InDel 6	1	0	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	1	0

9. TROUBLESHOOTING

Problem detected	Possible explanation and Solution
Amplification of InDels in every variety (including negative ones)	Primers with low specificity. Design new longer primers, or keep optimizing PCR conditions – Increase annealing temperature, reduce primer or template DNA concentration. Reduce final $MgCl_2$ concentration. In the case of unspecific bands, the duration of elongation can also be altered.
Multiple non-specific bands	
Band not amplifying in any variety	PCR conditions too stringent. Lower the annealing temperature, add $MgCl_2$ and DMSO (Dimethyl Sulfoxide), that not only increases the yield, it also increases the specificity.
Internal/negative controls with amplification	Stock/working solution of one component is contaminate with DNA. Prepare new working solution of primers and sterile water. If the problem still maintains, gradually replace each component with freshly prepared ones until the problem is solved.
InDels amplifying a pattern different from the predicted one	Sanger sequence the PCR product to ensure that the band amplifying is the expected one. Considering that InDels are in a highly repetitive region, the primers may just have more affinity to flanking regions in certain varieties than the others. Different DNA polymerases may also give better results.

FINAL COMMENTS

This booklet outlines what the TRACE-RICE project considers to be best practices for rice authentication.

While developed within the European market context, this work has potential applications in broader contexts. One limitation is that marker selection is specific and dependent on the chosen varieties, which requires access to whole-genome sequences and repeated use of the CRS bioinformatics tool.

Initially designed to distinguish seeds of different accessions, this approach can also be applied to rice-based foods to identify unwanted mixtures or adapted for varietal identification in other crop species.

Additionally, this booklet includes troubleshooting strategies for issues encountered during the project, providing a foundation for future experiments.

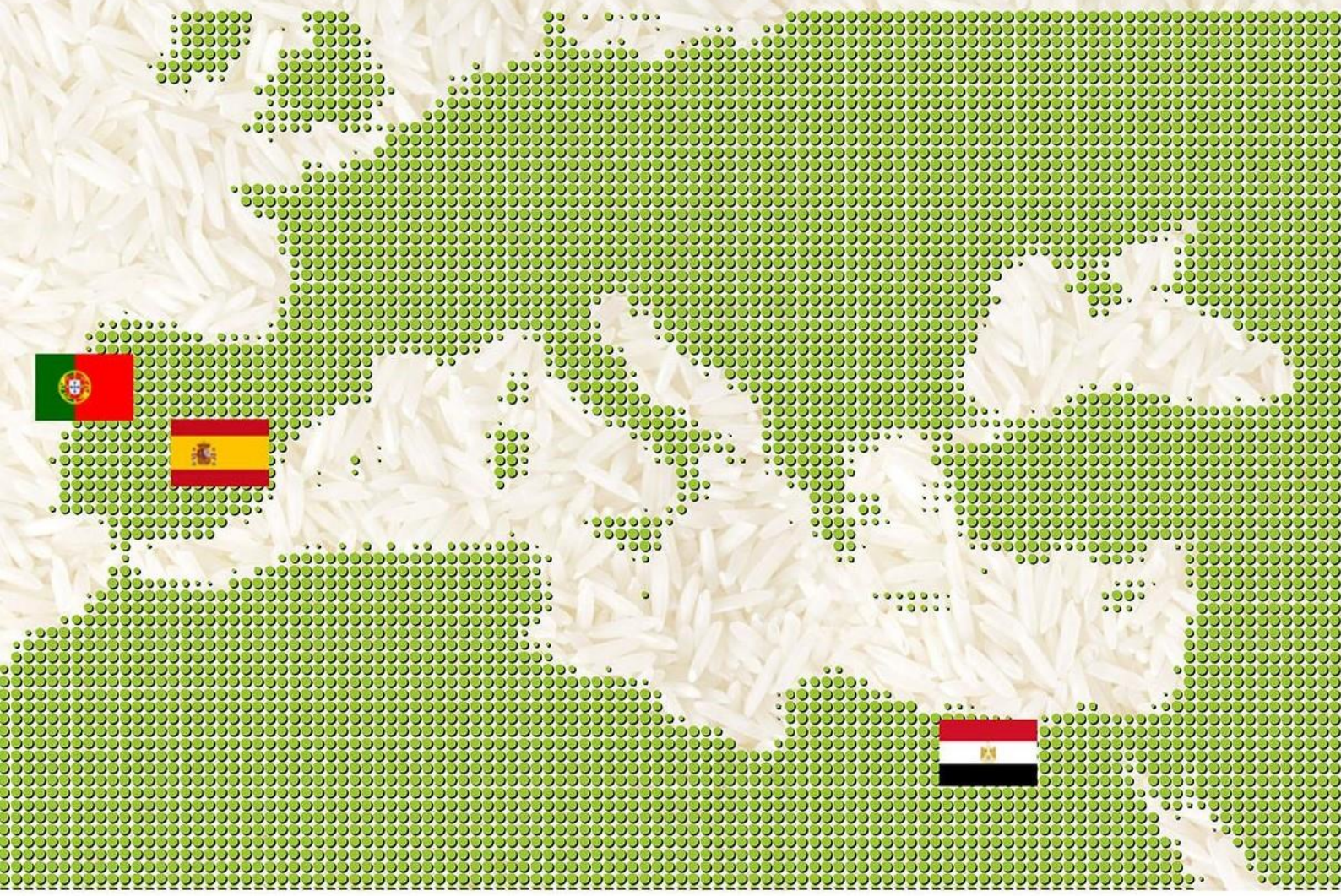
In conclusion, this method in TRACE-RICE could offer a cost-effective alternative to more complex methods for detecting fraudulent variety claims, aligning with our primary goal.

WEBSITE, CONTACTS AND ADDITIONAL INFORMATION

http://trace-rice.eu/

Email: postmaster@trace-rice.eu

TRACE-RICE



TRACE-RICE Consortium

Instituto Nacional de Investigação Agrária e Veterinária, I.P.

iBET Instituto de Biologia Experimental e Tecnológica

ITQB Universidade Nova de Lisboa

Casa do Arroz

ernesto morgad desde 1920

Mater Dynamics

BGI www.bgi.pt

Grupo Desarrollo

iata Instituto de Agroquímica y Tecnología de Alimentos

Alexandria University