

## A targeted sequencing approach to capture the resistance gene clusters diversity in melon

Javier Belinchon-Moreno

#### ▶ To cite this version:

Javier Belinchon-Moreno. A targeted sequencing approach to capture the resistance gene clusters diversity in melon. PAG XXX. Plant and Animal Genome Conference, Jan 2024, San Diego (California), United States. hal-04670453

## HAL Id: hal-04670453 https://hal.inrae.fr/hal-04670453v1

Submitted on 12 Aug 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# A targeted sequencing approach to capture the resistance gene clusters diversity in melon

Javier Belinchon-Moreno1,2, Aurelie Berard1, Aurelie Canaguier1, Isabelle Le-Clainche1, Véronique Chovelon2, Jacques Lagnel2, Vincent Rittener-Ruff2, Massyly Lebbat1, Damien Hinsinger1, Nathalie Boissot2, Patricia Faivre-Rampant1 Plant and Animal Genomics Conference, San Diego, CA, USA Cucurbi-genomics session 12<sup>th</sup> January 2023

## > NLR resistance genes



Intracellular immune receptor proteins

Major family of plant-resistance (R) genes

### Broad range of resistance



## Usually grouped into clusters

High level of presence/absence polymorphisms (PAV)

Low frequency in the Cucurbitaceae family

**81 NLR genes identified in DHL95** 

45% of them grouped into 9 clusters

## p. 3

(Chovelon et al., 2021)

# **Presence of TE inside the genes**

(pseudogenes)

**Highest PAV polymorphism density** 

(focus on 350Kb – M4M5)

# > Vat cluster in melon

Most studied NLR cluster in melon

Region of 1Mb on chromosome 5

Highest NLR gene density (23 genes by Gonzalez et al., 2014)



# > NLR clusters: a challenging characterization

- Specific role of each NLR remains largely unknown
- Characterize the NLR-type resistance in melon
- Contribution for sustainable agriculture
- Single reference genome cannot represent the full diversity of a species .....> Construction of NLRome

- Big challenge
- Complex gene structure (repetitions)
- Grouped into gene clusters, with plenty of repetitive elements (TE)
- Prone to duplication and transposition





- Sequencing using <u>short reads</u> are **ineffective**
- <u>Genotyping approaches</u> do **not provide reliable information**
- WGS using long reads may be so expensive (wetlab, info & bioinfo)

# > Nanopore adaptive sampling

- **Real time** target selection: **easy** library preparation.
- Accept/reject molecules based on small initial part of sequence (~800 bp)
- Enrich/deplete specific regions of interest.
- Increase on-target data, reduce time-to-answer.
- No need of DNA amplification
- No need of laborious or expensive experimental design



(Nanopore London Callings, 2022)

# > Our objective

INRA

12th January 2023 / Javier BELII







## > Nanopore adaptive sampling: our target regions



## > Our experiences

Testing on ANSO-77 (same as provided reference)



## > Our experiences

From ANSO-77 (same as provided reference) to more genetically distant accessions



# Sood enrichment of our target regions

Effectively reducing the off-target volume of data



# Assembly of 15 target regions of Anso77

Complete and continuous assembly



ANSO-77 – inodorus (ssp. melo)

Target regions fully assembled

## All predicted NLR genes (84) found in the assembly

number_contigs	15
total_length	7.023 Mb
largest_contig	1.443 Mb
GC(%)	33.23
N50	1.107 Mb
N75	1.026 Mb
L50	3
L75	5

## Results using different varieties Also complete and continuous assembly of all the NLR-clusters





DOUBLON – cantalupensis (ssp. melo)

#### Target regions fully assembled All predicted NLR genes (76) found in the assembly

ANANAS – ameri (ssp. melo)

Target regions fully assembled

85/85 NLR genes found in the assembly

VEDRANTAIS – cantalupensis (ssp. melo)

Target regions fully assembled 82/81 NLR genes found in the assembly

ZHIMALI – chinensis (ssp. agrestis)

Target regions fully assembled **79/79 NLR genes found in the assembly** 

# > Adaptive sampling limitations



# > A single reference is not enough: Necesity of a pan-NLRome

Accession differing from
 the reference by SV

Reference may contain no location to correctly map the reads

Mapping bias

- **True** even with the newer **long-read** sequencing approaches
- For this reason, SVs are much more poorly characterized than SNPs and short InDels (Sirén et al., 2021).
- The problem grows in highly complex regions with a large number of presence/absence polymorphisms and many **repetitive elements**





# > Pangenome construction tools

Using 7 varieties sequenced with NAS

	Minigraph (Li et al., 2020)	Minigraph-Cactus (Hickey et al., 2022)	<b>PGGB</b> (Garrison et al., 2023)
Sequence comparison	reference-based, progressive	reference-based, progressive	symmetric, all-vs-all
Resolution	SV only (variations >50 bp)	base-level (SNPs)	base-level (SNPs)
Scope	full assemblies	Non-centromeric	full assemblies
Full reconstruction possible	yes	yes	no
Short read mapping	untested	yes (fast)	untested
Long read mapping	yes (fastest)	yes	yes (slowest)

(Adapted from https://github.com/human-pangenomics/hpp\_pangenome\_resources)

## > Variant genotyping on the pangenome graph



# > Variant genotyping on the pangenome graph



## First attempts to build a pan-NLRome graph

Focus on the well studied M4-M5 region of Chr5 (350 Kb)



Graph saturation curve

100 accessions graph (minigraph-cactus)



- --minigraph
- --minigraph\_cactus

## First attempts to build a pan-NLRome graph

Focus on the well studied M4-M5 region of Chr5 (350 Kb)

	<b>2 melon acccessions</b> Anso77 and Doublon	419 557 bp 6 033 segments 8 189 edges
Variation present in graph	Mapping of Anso77 short re	ads Mapping of Doublon short reads
2157 variants found	1 variants PASS found	2149 variants PASS found
1665 SNPs	0 SNPs	1657 SNPs (99.5%)
462 InDels 1-50 bp	1 InDels 1-50 bp	462 InDels 1-50 bp (100%)
31 SVs > 50 bp	0 SVs > 50 bp	30 SVs > 50 bp (98%)
	7 melon acccessions Anso77 and Doublon	584 518 bp 24 820 segments 34 084 edges
Variation present in graph (only counting ANSO/DOUBLON) 1531 variants found 1199 SNPs 302 InDels 1-50 bp	Mapping of Anso77 short reads 7 variants PASS found 0 SNPs 6 InDels 1-50 bp 1 SVs > 50 bp	<ul> <li>Mapping of Doublon short reads         <ul> <li>1703 variants PASS found</li> <li>1305 SNPs (108.8%)</li> <li>372 InDels 1-50 bp (123%)</li> <li>26 SVs &gt; 50 bp (86 7%)</li> </ul> </li> </ul>
1 INR 4030 SVs > 50 bp		

## > First attempts to build a pan-NLRome graph

Focus on the well studied M4-M5 region of Chr5 (350 Kb)

**2 melon acccessions** Anso77 and Doublon



## > In conclusion

- Nanopore adaptive sampling is a simple, reliable, efficient and cost-saving target sequencing approach
- Adaptive sampling allows to efficiently retrieve (mapping and assembly) our 15 ROI in melon.
- A **reference genome is not enough** to characterise the diversity of the NLRome in melon
- Graph pangenomics still unstable and under developement
- The construction of a pangenome graph will allow the characterisation of the complex NLRome clusters using a reference-free approach that improves the common errors showing up in classical variant calling with a single reference





## Acknowledgements



## **INRAE-EPGV**

Patricia Faivre-Rampant Aurelie Berard Aurelie Canaguier Isabelle Le-Clainche Damien Hinsinger

## **INRAE-GAFL**

Nathalie Boissot Veronique Chovelon Jacques Lagnel Vincent Rittener-Ruff

## <u>Genoscope</u>

Corinne Cruaud Stefan Engelen

## PRIVATE PARTNERS



Gautier semences







@EPGV\_INRAE



support-EPGV@inrae.fr javier.belinchon-moreno@inrae.fr