Characterization of the genetic and functional diversity of the NLRome/resistome in melon (Cucumis melo L.) using Nanopore Adaptive sampling (NAS)

Javier Belinchon-Moreno

HAL Id: hal-04071831
https://hal.inrae.fr/hal-04071831v1

Submitted on 17 Apr 2023

Exploring the genetic and functional diversity of the NLRome/resistome in melon (*Cucumis melo* L.) using Nanopore Adaptive sampling (NAS)

Javier BELINCHON-MORENO

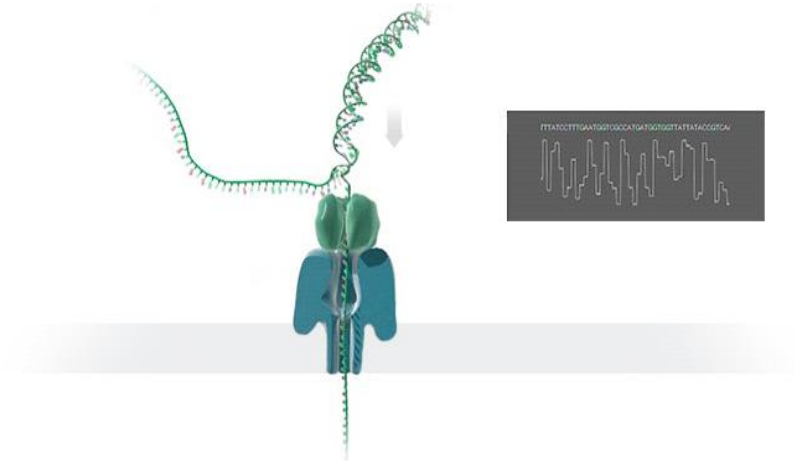Journées Jeunes Chercheurs, Clermont-Ferrand, 3rd April 2023

Oxford Nanopore Sequencing Technology (ONT)

Nanopore embedded in an electrically resistant polymer membrane

Read from *cis* to *trans*

Constant voltage applied → Read change of voltage

Flongle	MinION	MinION Mk1c	GridION	PromethION[24]	PromethION[48]
Theoretical Max: 2.8 Gb	Theoretical Max: 50 Gb	Theoretical Max: 50 Gb	Theoretical Max: 50 Gb x 5 = 250 Gb	Theoretical Max: 290 Gb x 24 = 7 Tb	Theoretical Max: 290 Gb x 48 = 14 Tb

(Nanopore London Callings, 2022)

What is Nanopore adaptive sampling?

- Accept/reject molecules based on **small initial part** of sequence (~500 bp)

- **Enrich/deplete** specific regions of interest.

- Increase on-target data, reduce time-to-answer.

- **Real time** target selection: **easy** library preparation.
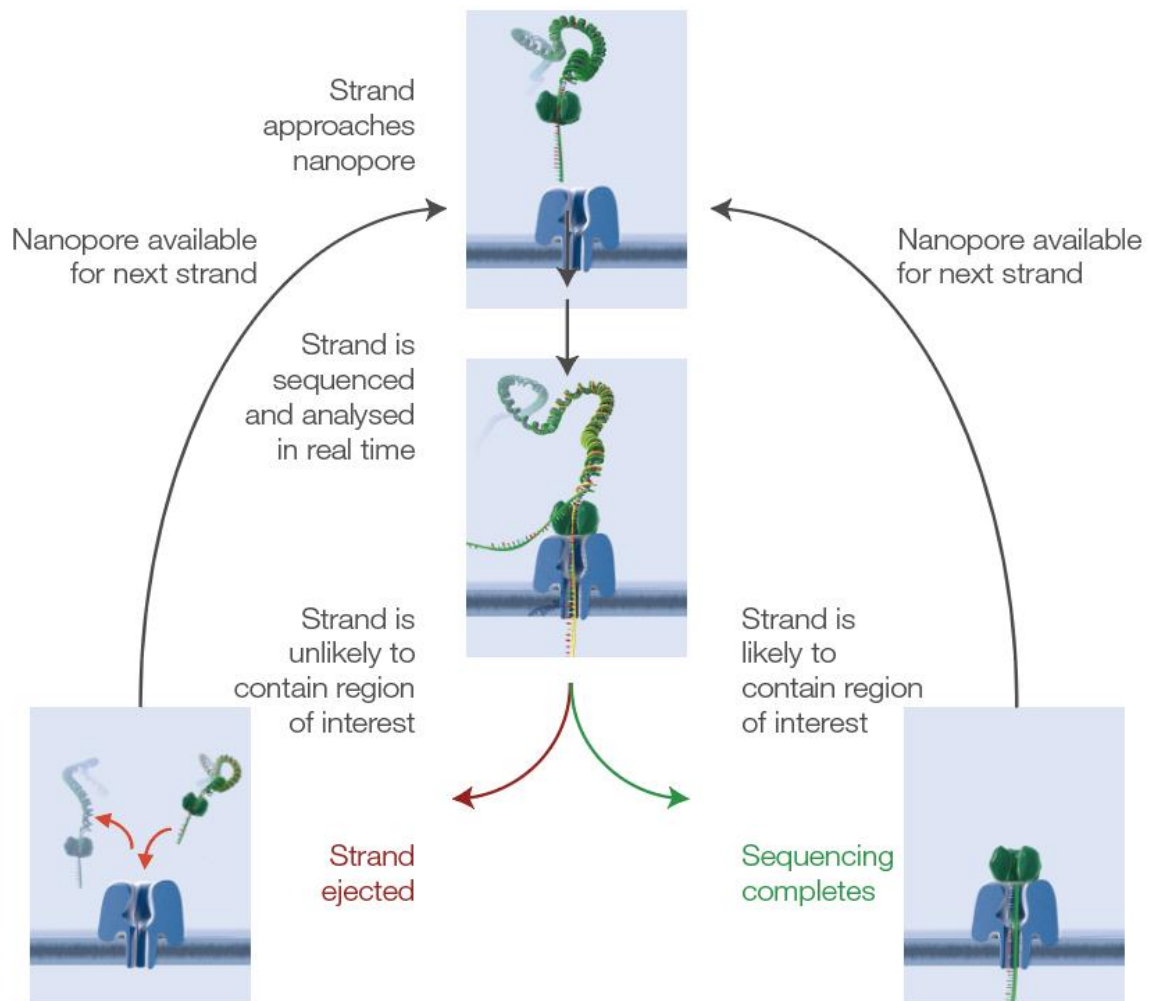
Strand approaches nanopore

Nanopore available for next strand

Strand is sequenced and analysed in real time

Nanopore available for next strand

Strand is unlikely to contain region of interest

Strand is likely to contain region of interest

Strand ejected

Sequencing completes

(Nanopore London Callings, 2022)

1 Reference genome not enough ➡ Construction of **pan-NLRome**

High number of varieties (~150 varieties) ➡ Need for **selective sequencing**

(Adapted from Barragan & Weigel, 2021)
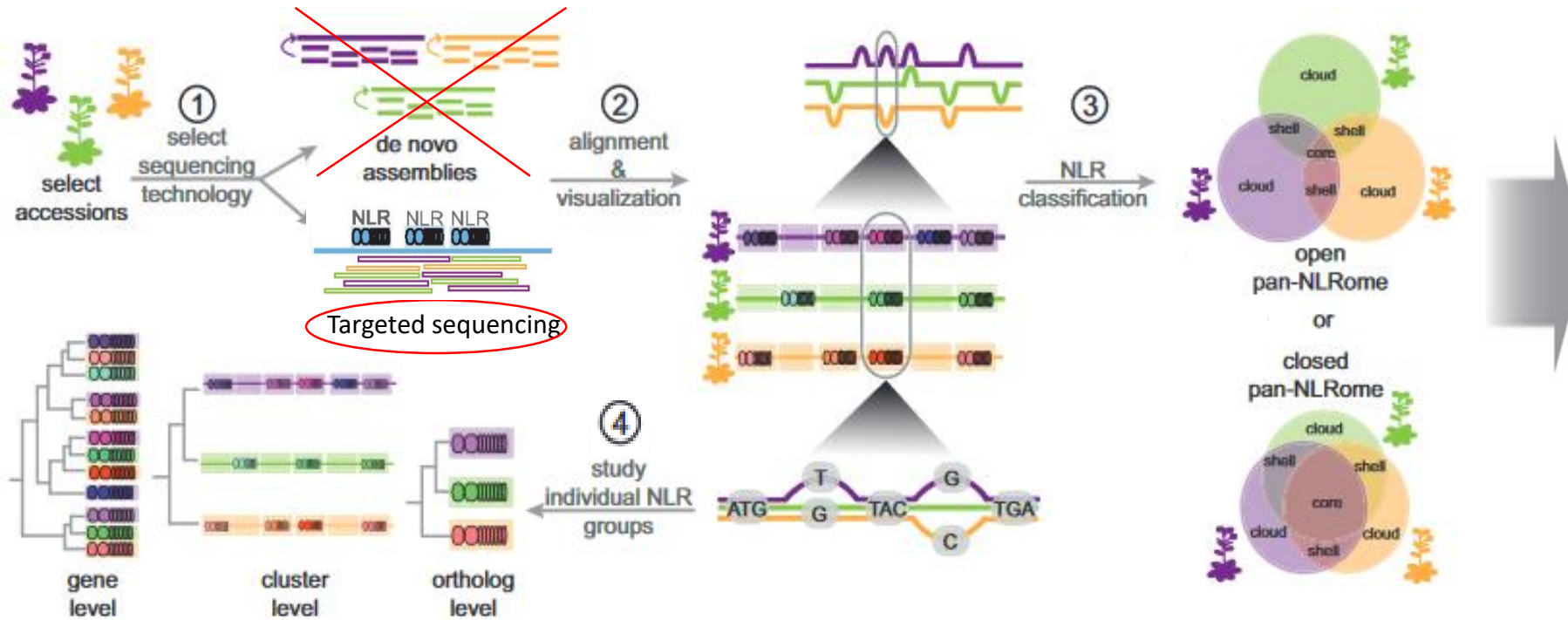
2

Understand its role in the expression of immunity against a wide range of pests and pathogens (GWAS)

NLR resistance genes: a conserved structure

Nucleotide binding site (NB) -leucine rich repeat (LRR)

Intracellular immune receptor proteins

Major family of plant-resistance (R) genes

Broad range of **resistance**

Adapted from (Gottin et al., 2021)

NLR — TIR/CC/RPW8 — NB-ARC — LRRs

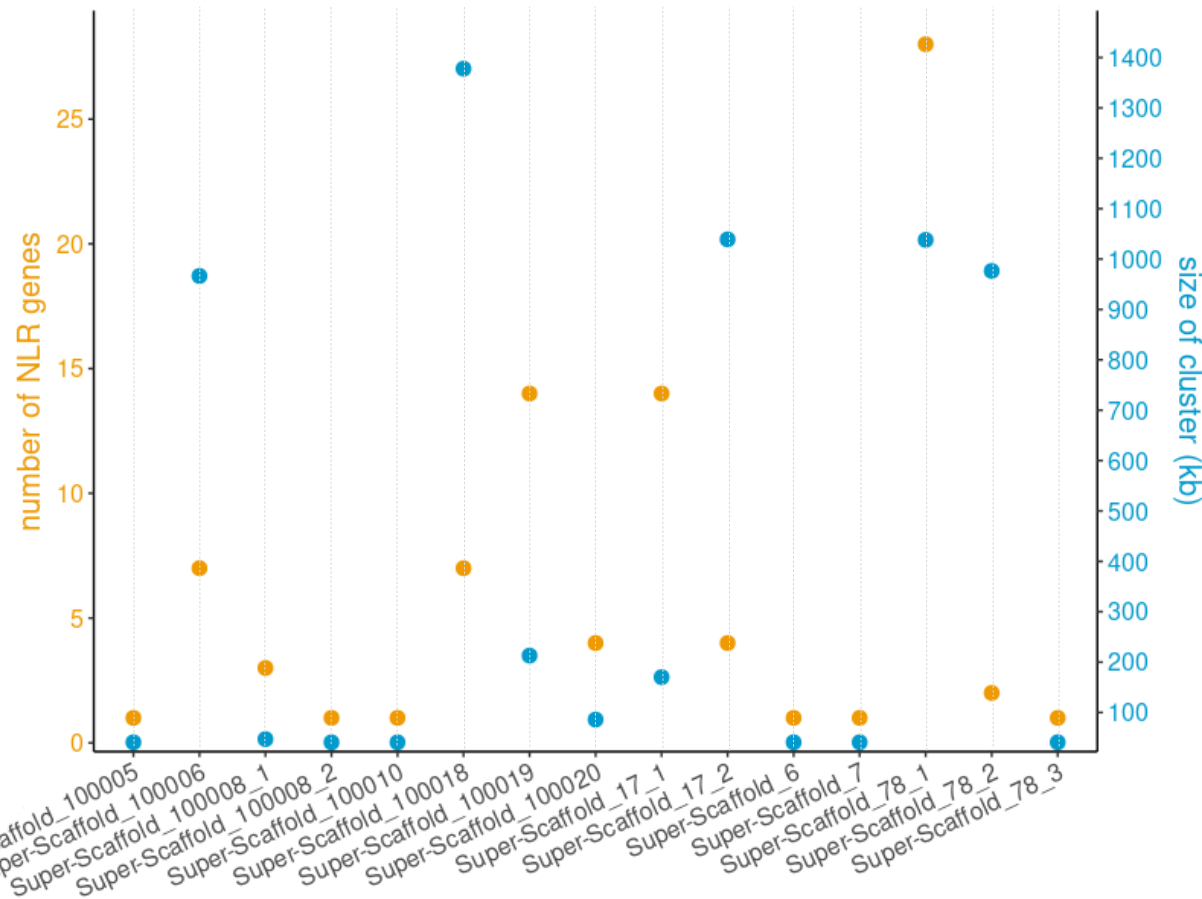*F. oxysporum*   *A. gossypii*   Oidium   *P. infestans*

Usually grouped into **clusters**

High level of presence/absence polymorphisms (PAV)

Low frequency in the Cucurbitaceae family
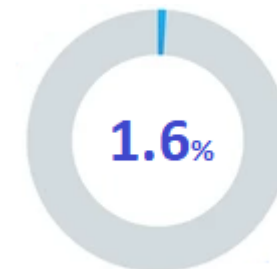
Our target regions: 15 NLR-gene clusters

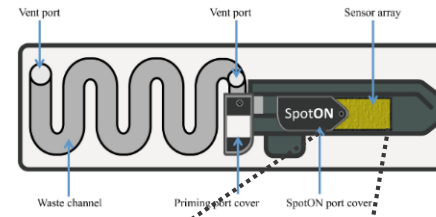Reference genome ➡ Variety **Anso77** (draft genome produced at INRAE-GAFL/EPGV)

Reference genome (.fasta)
+
Target regions (.bed)

Clusters:
NLR genes + 20kb-extra flanking
Variable in size
Variable in number of genes

Genome length ~380Mb
Target regions length ~6.16Mb

1.6%

INRAE

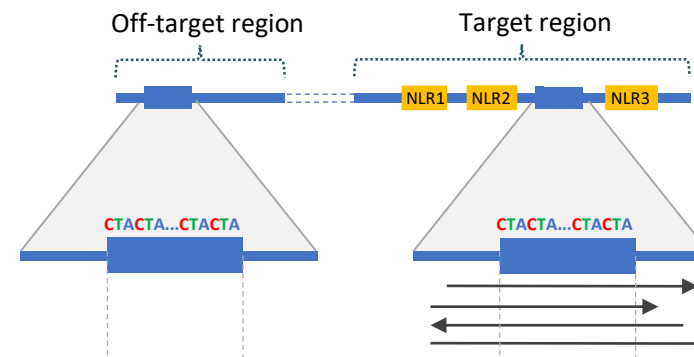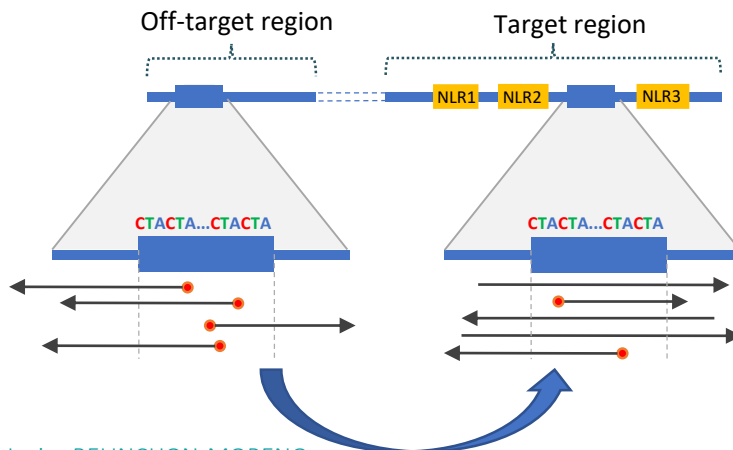# Our first experience: Variety Anso-77 (same as provided reference)
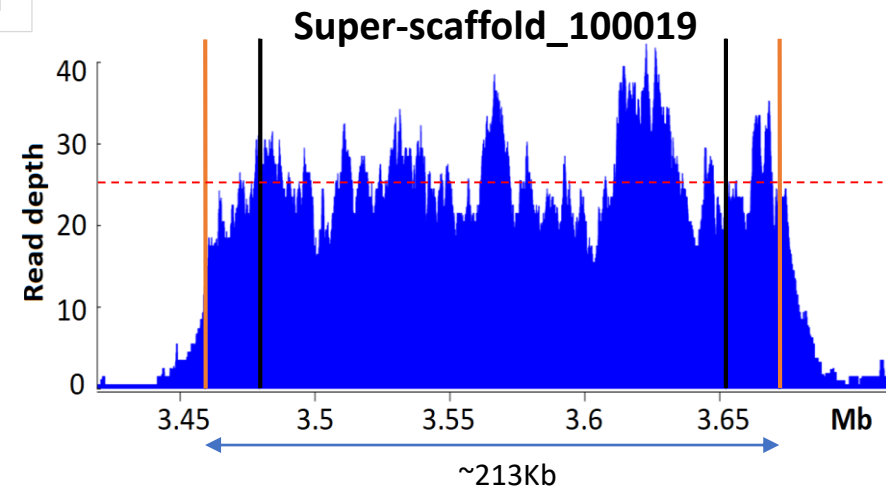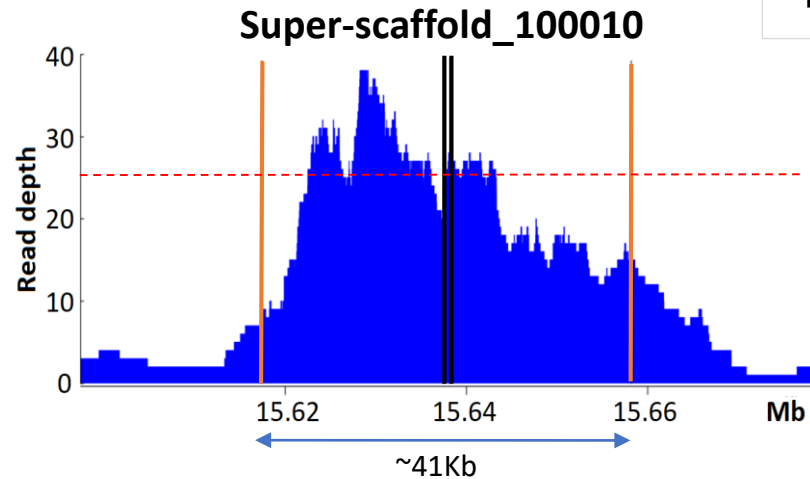

NAS

Standard sequencing
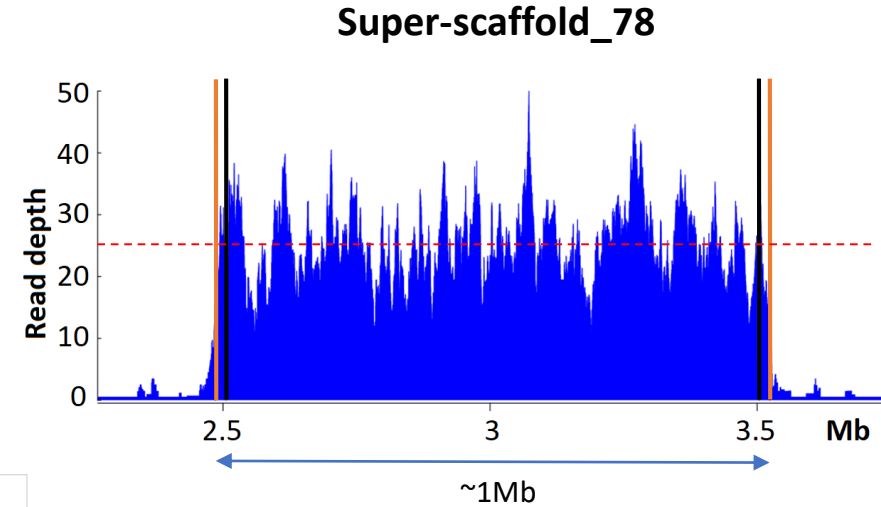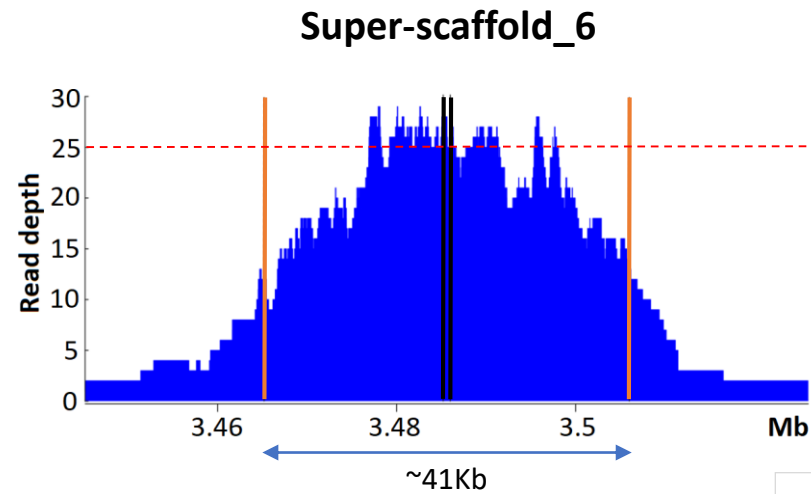
Not masking repetitive elements

Masking repetitive elements >200pb

INRAe

Visual enrichment of our target regions

Good enrichment of our target regions

Enrichment NAS vs standard

Regions of interest
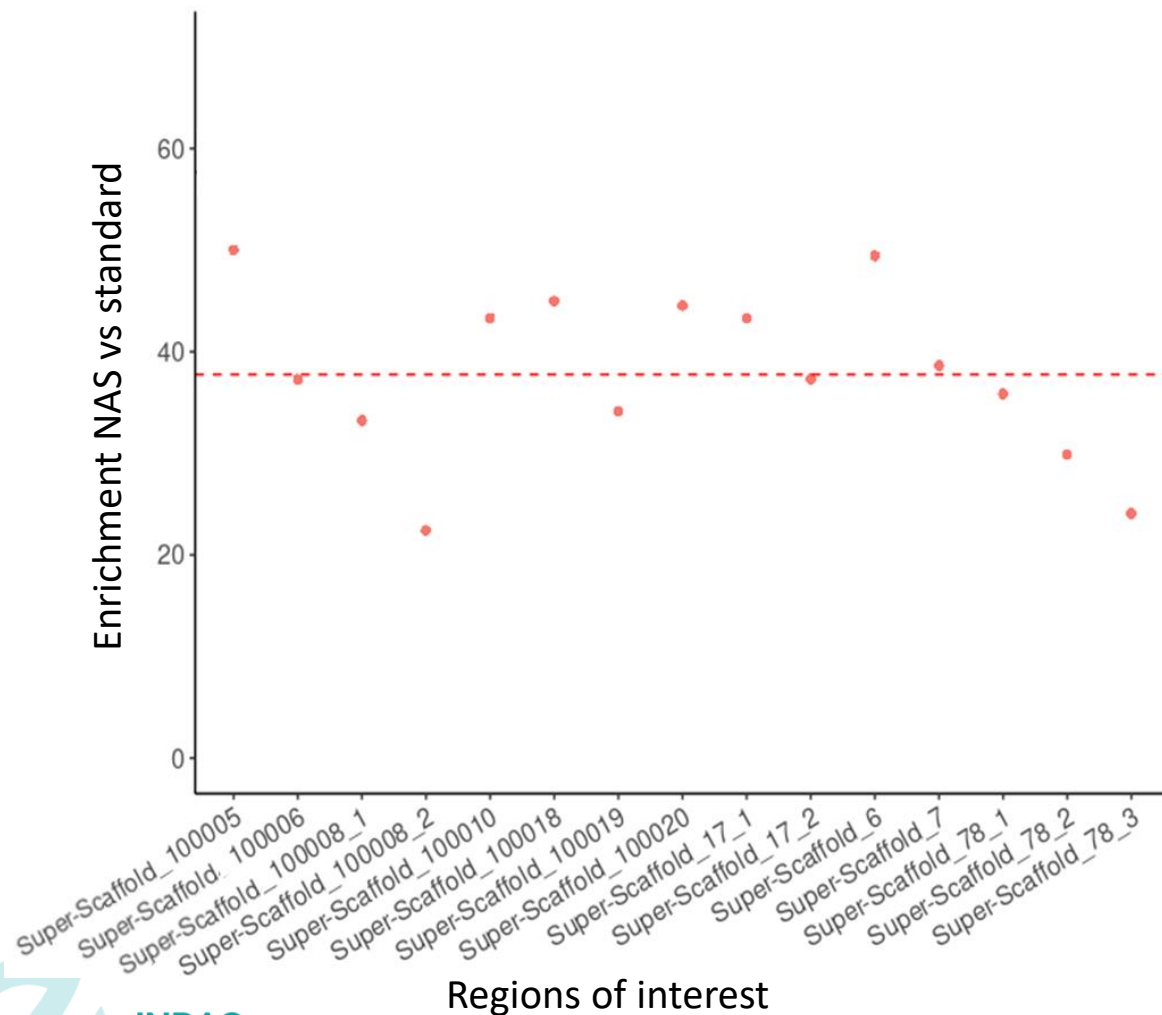
$$Enrichment_{region} = \frac{\dfrac{region\_cov\_NAS}{chrom\_cov\_NAS}}{\dfrac{region\_cov\_standard}{chrom\_cov\_standard}}$$

$$Av.\,enrichment = \sum_{region=1}^{N} \frac{enrichment_{region}(X) * region\_size(bp)}{total\_size\_regions(bp)}$$

- Variable between clusters

- Very high in average

$$\frac{region\_cov\_NAS}{region\_cov\_standard} \implies \sim 3\text{-}4$$

Genome assemblies: Good assembly of target regions compared to a standard sequencing

Anso-77 reference: 15 gene clusters

Anso-77 whole genome assembly
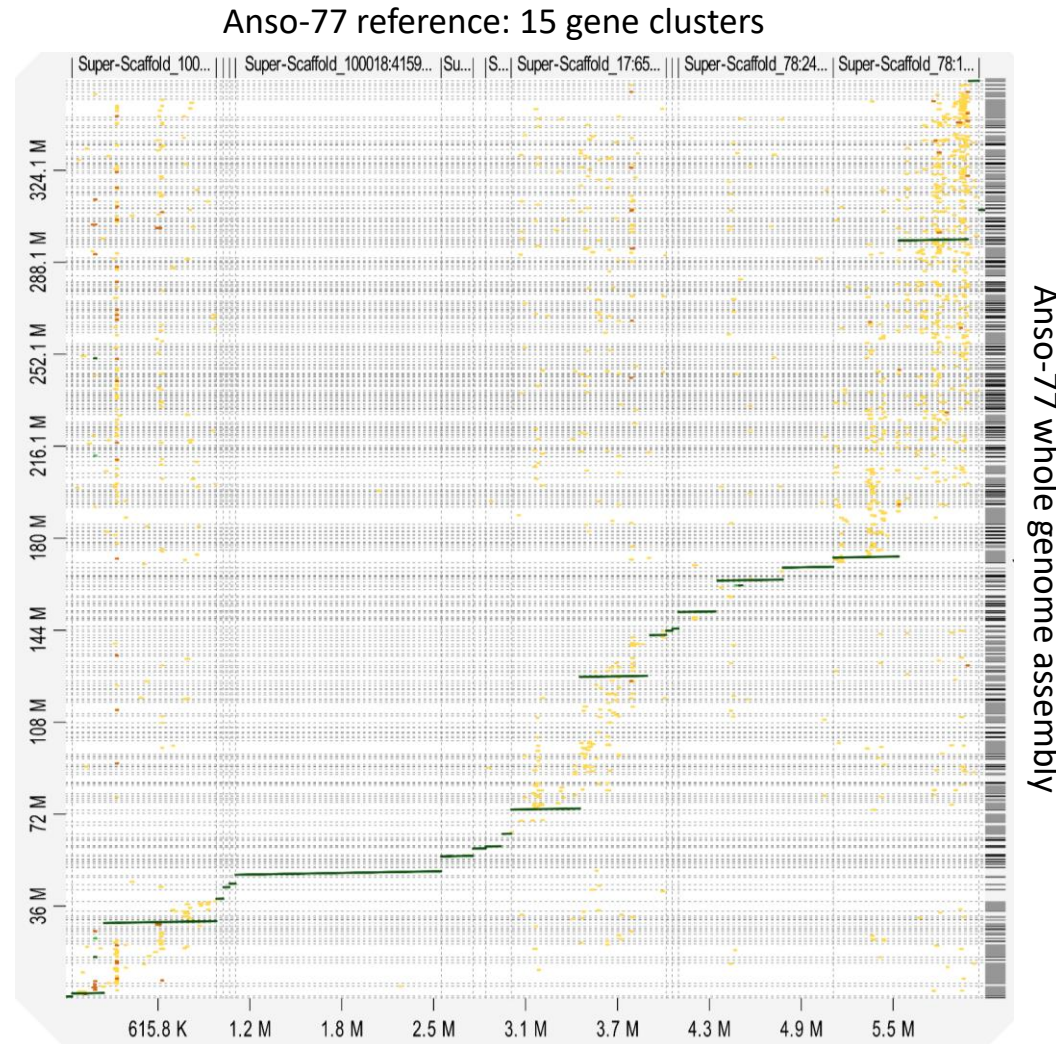
Assembly using WGS data:

- Target regions **fragmented** into multiple contigs

- Target regions **poorly covered** compared to adaptive sampling

Genome assemblies: Good assembly of target regions

Anso-77 reference: 15 gene clusters
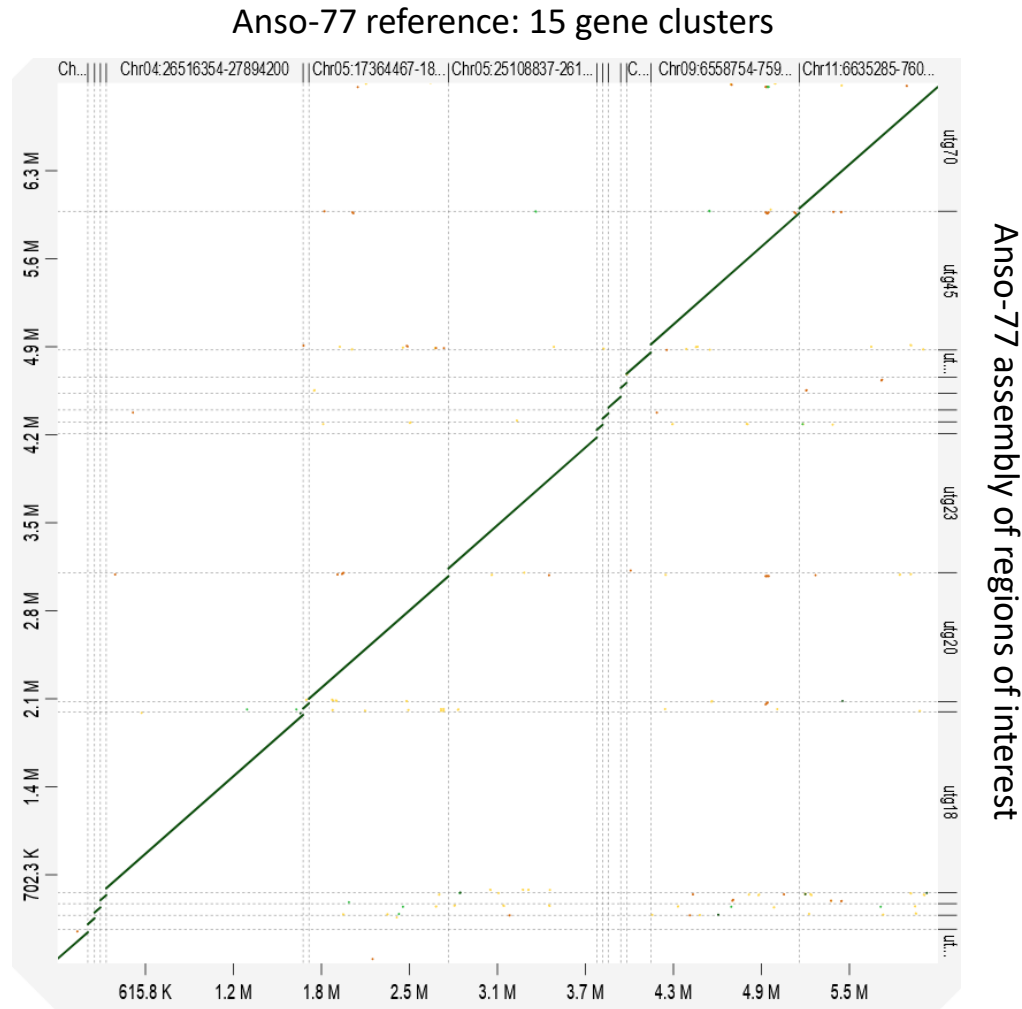
ANSO-77 – inodorus (ssp. melo)

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

NAS

Standard sequencing
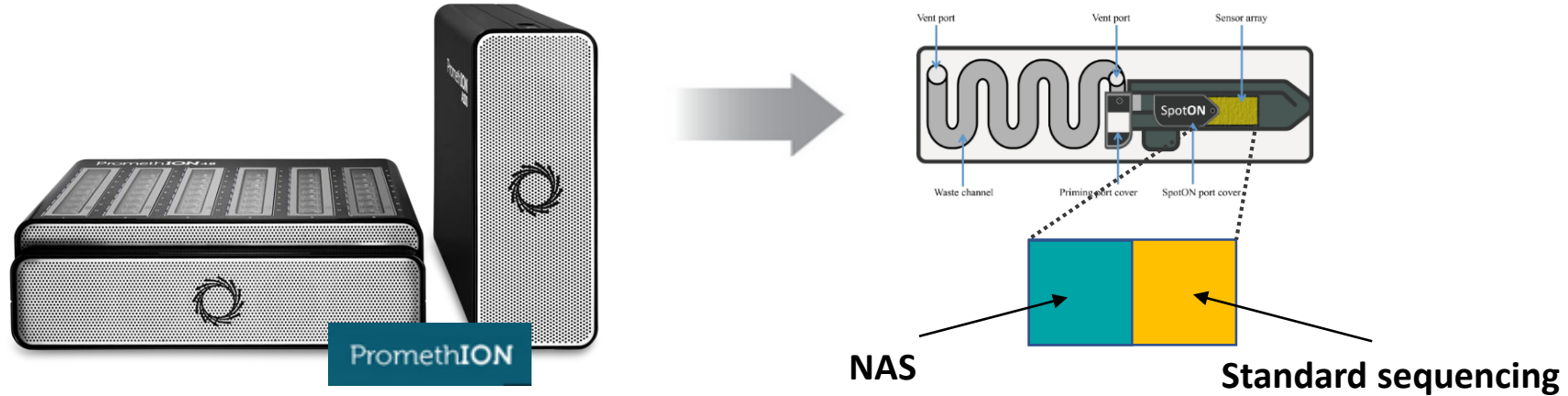
Single variety on promethION

2 barcodes on promethION

4 barcodes on promethION

8 barcodes on promethION

Preliminary results using different varieties: Doublon

Doublon: predicted 15 gene clusters
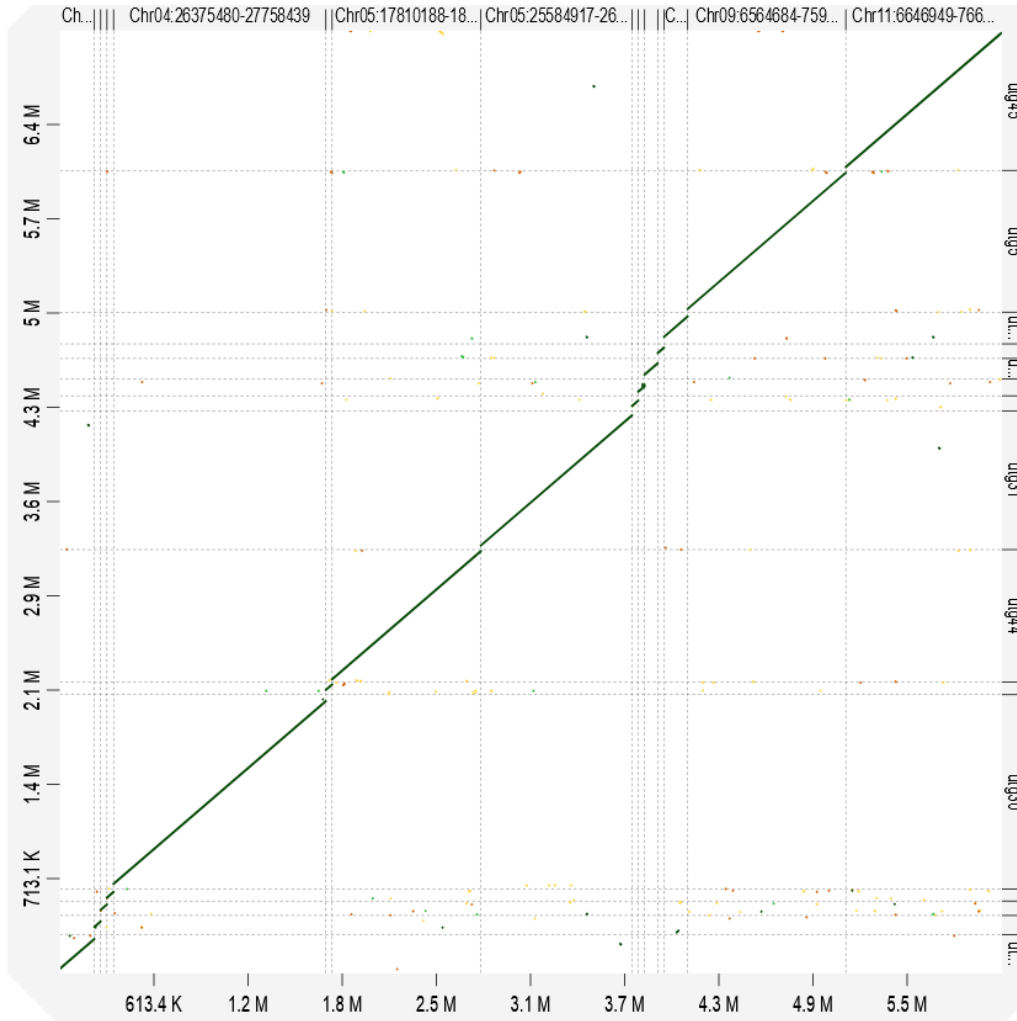
DOUBLON – cantalupensis (ssp. melo)

Target regions fully assembled

All predicted NLR genes (76) found in the assembly

number_contigs	15
total_length	7.131 Mb
largest_contig	1.470 Mb
GC(%)	33.31
N50	1.060 Mb
N75	1.003 Mb
L50	3
L75	5

Preliminary results using different varieties: Zhimali

Anso-77: predicted 15 gene clusters
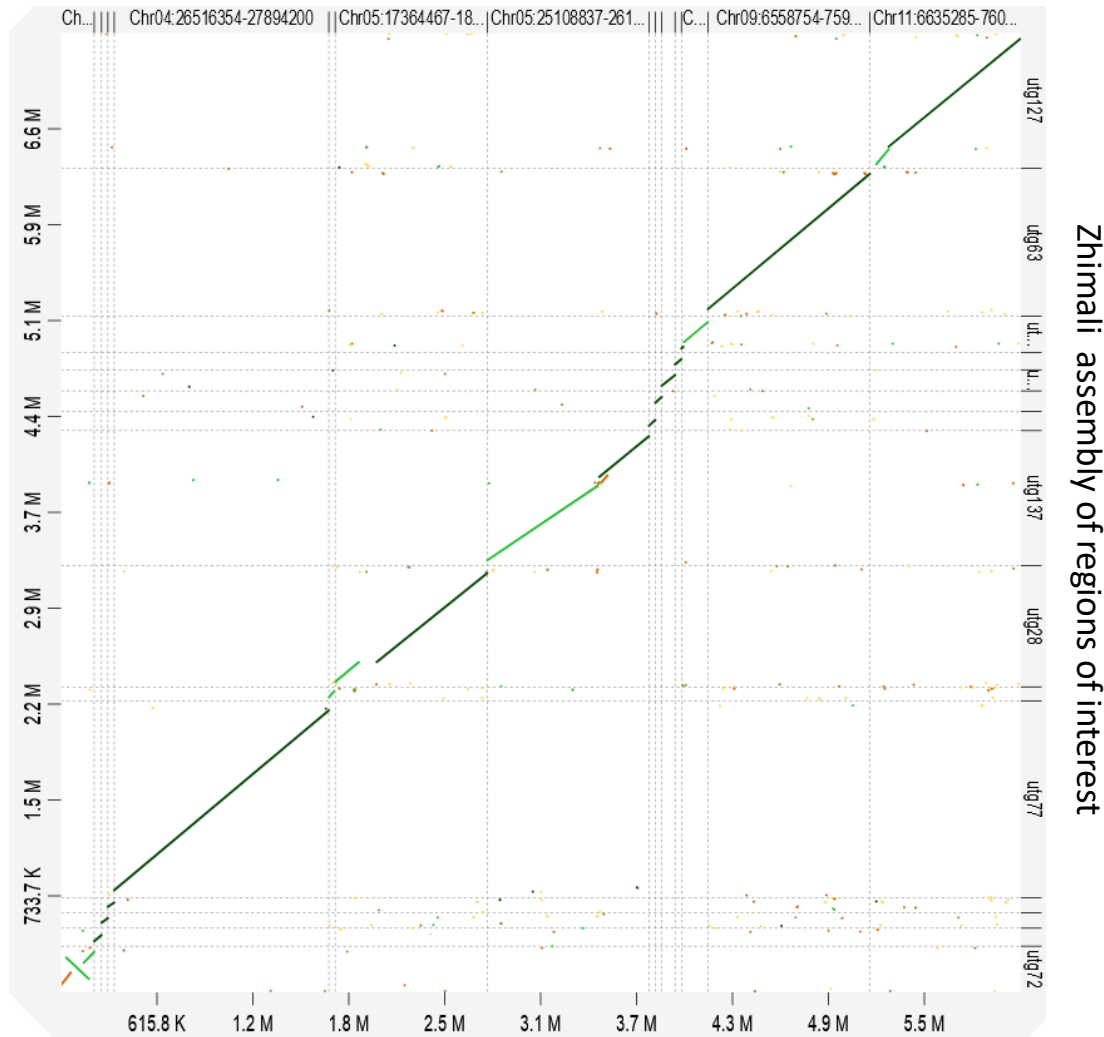
ZHIMALI – chinensis (ssp. agrestis)

Target regions fully assembled

78 NLR genes found in the assembly

number_contigs	15
total_length	7.635 Mb
largest_contig	1.476 Mb
GC(%)	33.23
N50	1.086 Mb
N75	1.035 Mb
L50	4
L75	5

INRAⒺ

3 April 2023 / Javier BELINCHON-MORENO

Perspectives

Short-term perspectives:

1. **Increase gradually the number of barcodes** on PromethION.

2. Start evaluating the available methods for pangenome construction.

Middle-term perspectives:

1. **Find the optimal number of barcodes** on PromethION: Selective sequencing of 150 varieties.

2. Evaluate the ability to detect **structural variants** and **copy number variations** between varieties.

Long-term perspectives:

1. Perform a similar experience using a **diploid heterozygote** species with a **higher number of NLR genes** and **more complex clusters**: *Vitis vinifera.*

In conclusion

What Nanopore adaptive sampling offers?

- Nanopore adaptive sampling is a **simple**, **reliable**, **efficient** and **cost-saving** approach.

  - ✓ Standard DNA extraction (no need for HMW DNA)
  - ✓ Standard library preparation
  - ✓ No need of probes design/synthesis
  - ✓ No need of short-read polishing with Kit14
  - ✓ Real-time sequencing
  - ✓ Reduced off-target volume of data compared to WGS

What can we do?

1. Adaptive sampling allows to efficiently retrieve (**mapping** and **assembly**) our **15 ROI in melon**.

2. It will allow to **explore the genetic and functional diversity** of melon **NLR genes** using a **large amount of varieties**.

Acknowledgements