Nanopore Adaptive sampling: A powerful approach to decipher the genetic diversity of the NLRome/resistome in melon

Javier Belinchon-Moreno

To cite this version:

Javier Belinchon-Moreno. Nanopore Adaptive sampling: A powerful approach to decipher the genetic diversity of the NLRome/resistome in melon. Conference, talk, Oxford Nanopore Technologies, Mar 2023, Paris institut du cerveau (ICM), France. hal-04037782

HAL Id: hal-04037782
https://hal.inrae.fr/hal-04037782

Submitted on 20 Mar 2023

Nanopore Adaptive sampling: A powerful approach to decipher the genetic diversity of the NLRome/resistome in melon

Javier BELINCHON-MORENO

Nanopore Day, Paris, 21st March 2023

RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité

INRAE

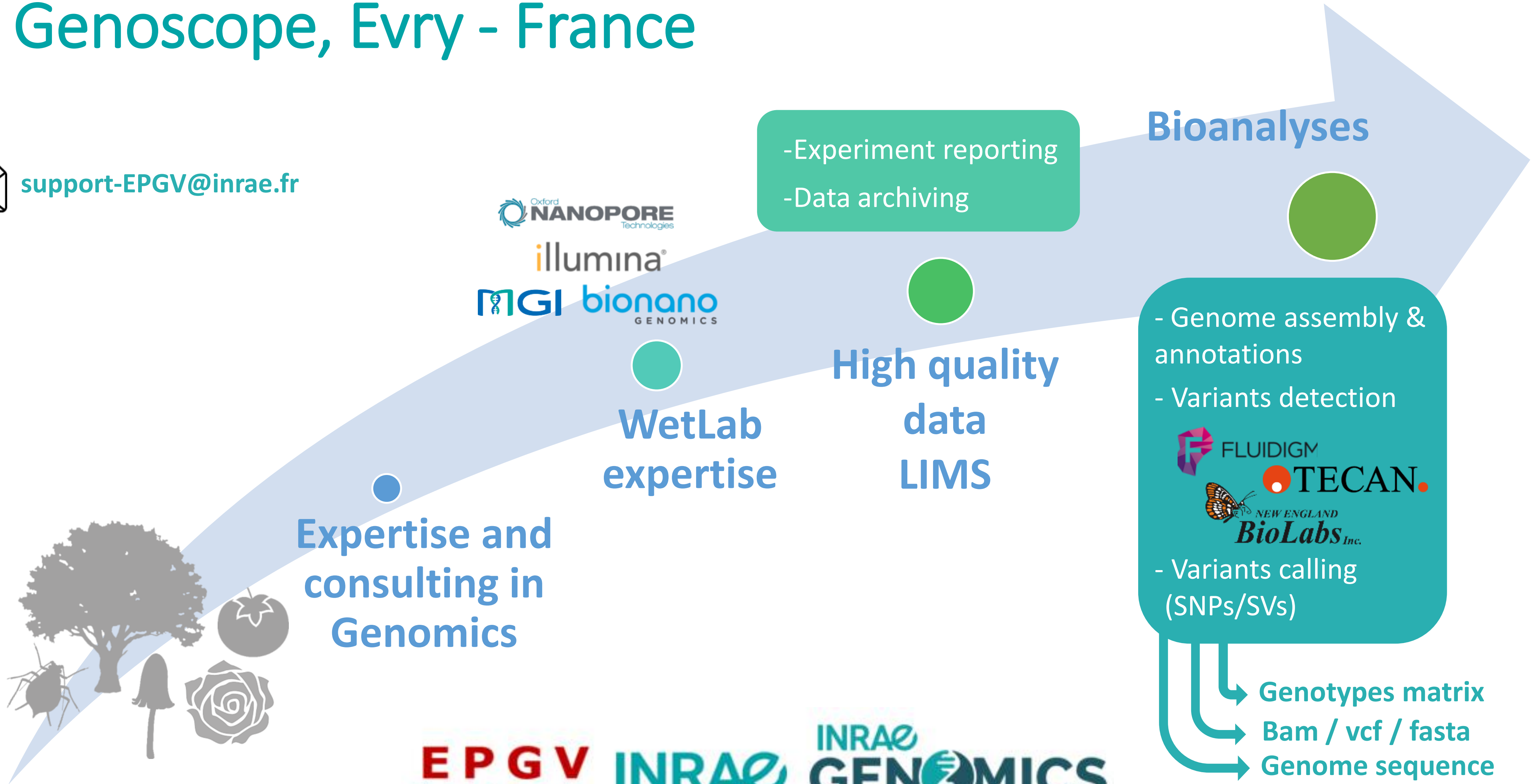
EPGV
Etude du Polymorphisme
des Génomes Végétaux

université
PARIS-SACLAY

GAFL
INRAE AVIGNON

EPGV – INRAE: plant genomic platform located at Genoscope, Evry - France

support-EPGV@inrae.fr



Expertise and consulting in Genomics

WetLab expertise

- Experiment reporting
- Data archiving

High quality data LIMS

Bioanalyses

- Genome assembly & annotations
- Variants detection
- Variants calling (SNPs/SVs)

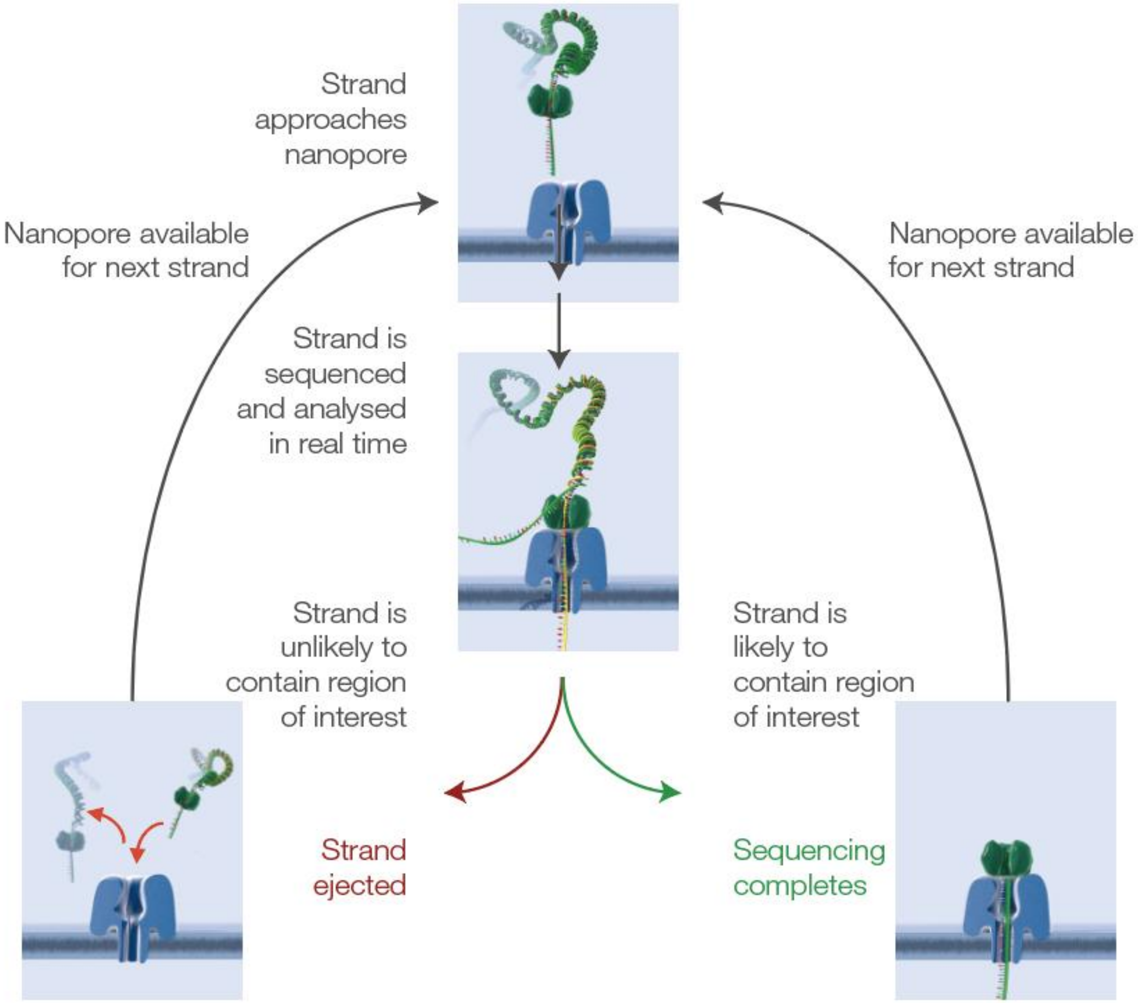
Genotypes matrix

Bam / vcf / fasta

Genome sequence

What is Nanopore adaptive sampling?

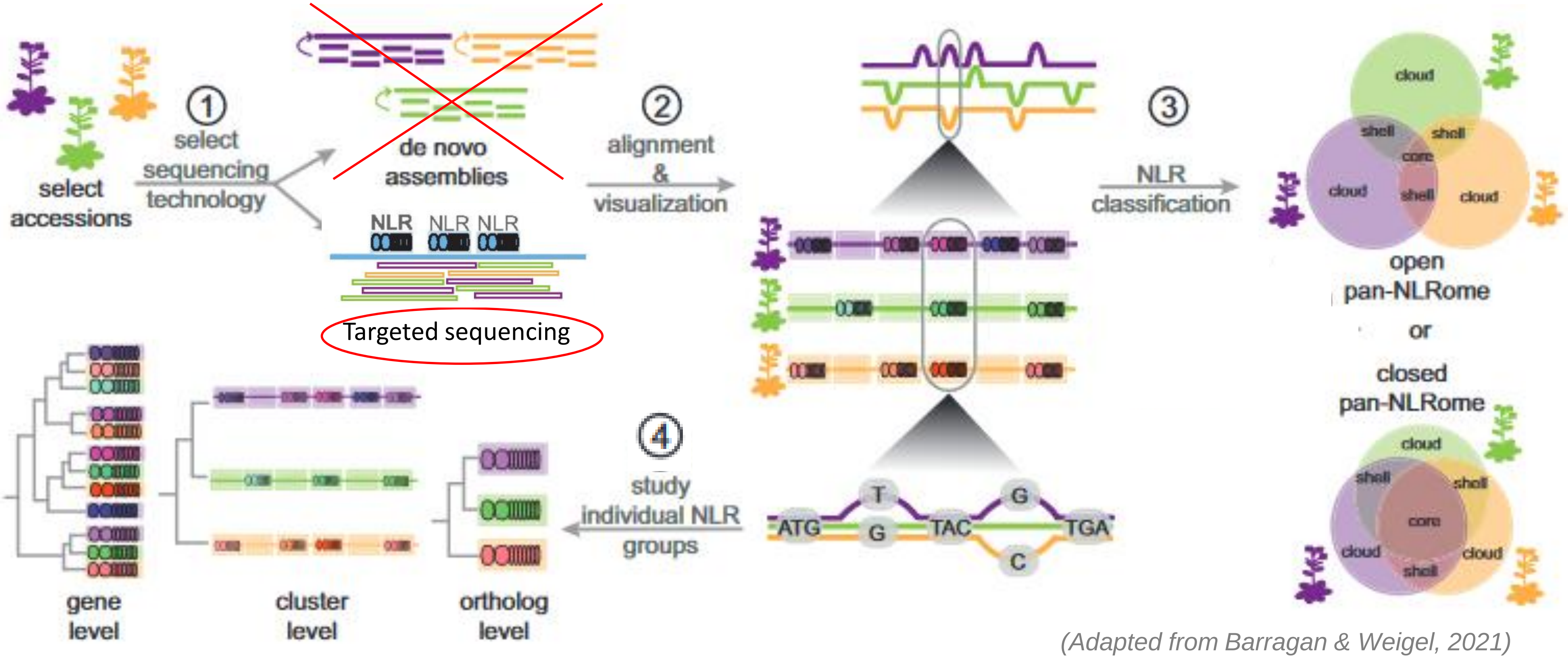
- Accept/reject molecules based on **small initial part** of sequence (~500 bp)
- **Enrich/deplete** specific regions of interest.
- Increase on-target data, reduce time-to-answer.
- **Real time** target selection: **easy** library preparation.



Our objective: Nanopore adaptive sampling to decipher the genetic diversity of the NLRome in melon

Reference genome not enough → Construction of **pan-NLRome**

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



(Adapted from Barragan & Weigel, 2021)

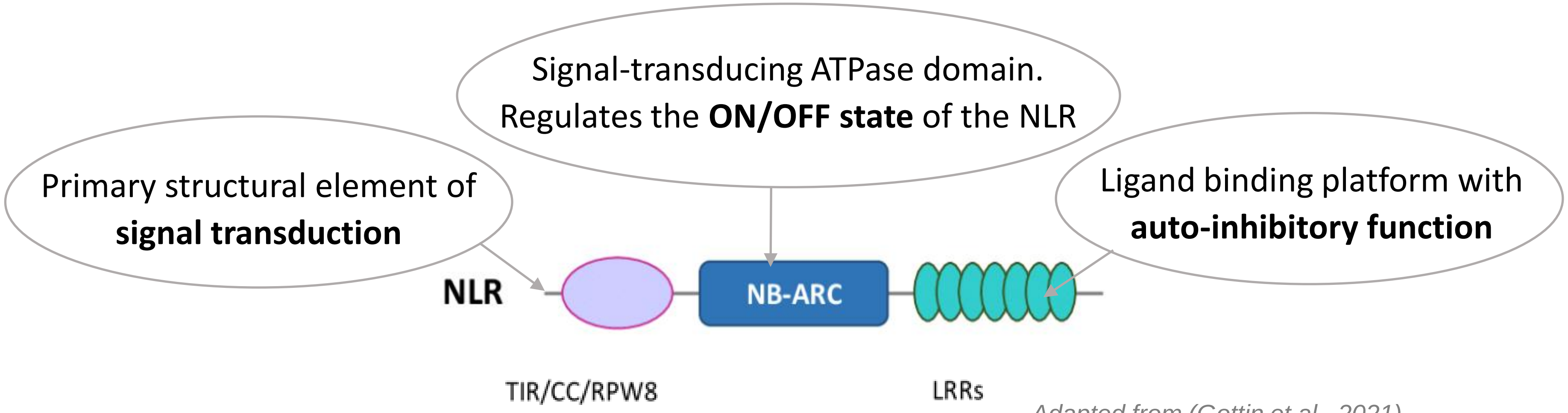
NLR resistance genes: a conserved structure

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

Intracellular immune receptor proteins

Major family of plant-resistance (R) genes

Broad range of **resistance**



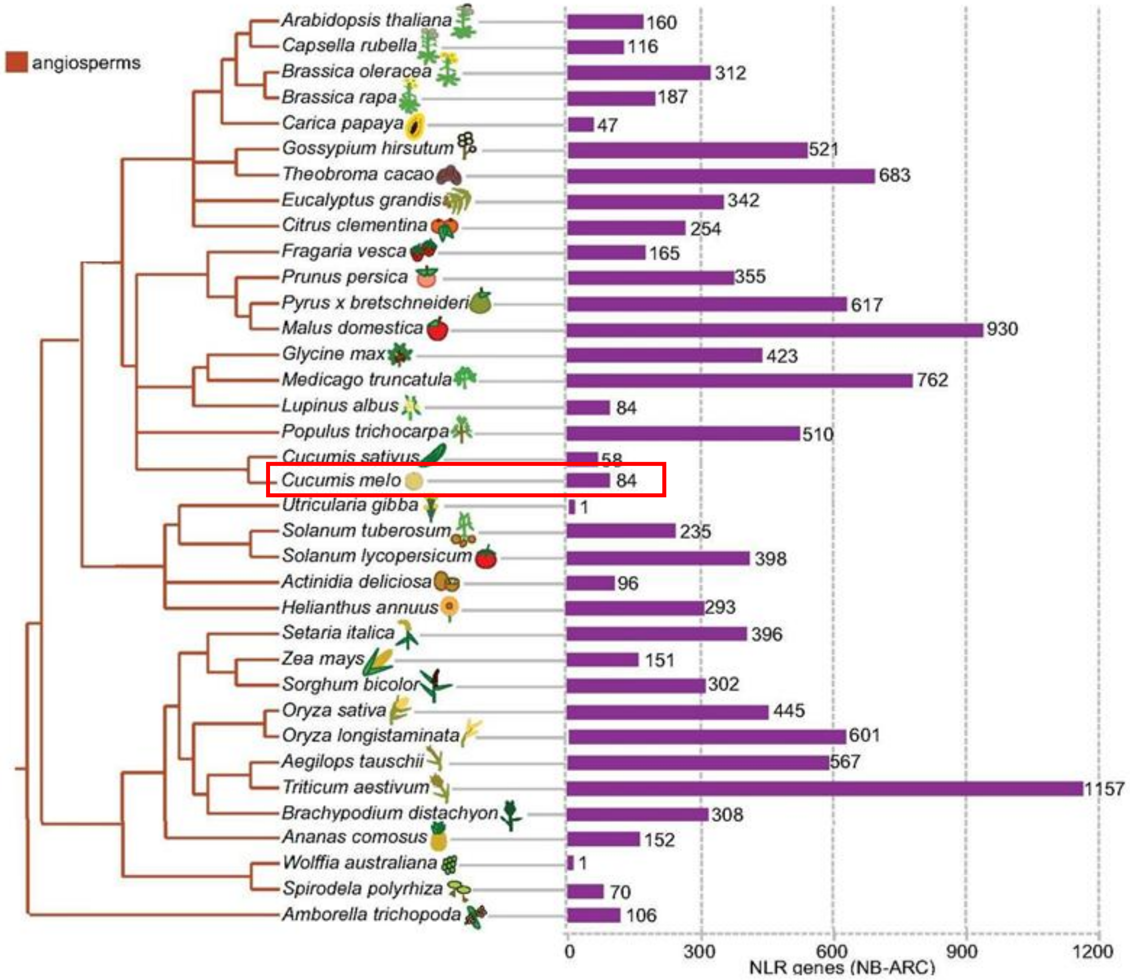
Signal-transducing ATPase domain.
Regulates the **ON/OFF state** of the NLR

Primary structural element of
signal transduction

Ligand binding platform with
auto-inhibitory function

Adapted from (Gottin et al., 2021)

Variable number of NLR genes between plant species



Adapted fom (Barragan & Weigel, 2021)

Usually grouped into **clusters**

High level of presence/absence polymorphisms (PAV)

Low frequency in the Cucurbitaceae family

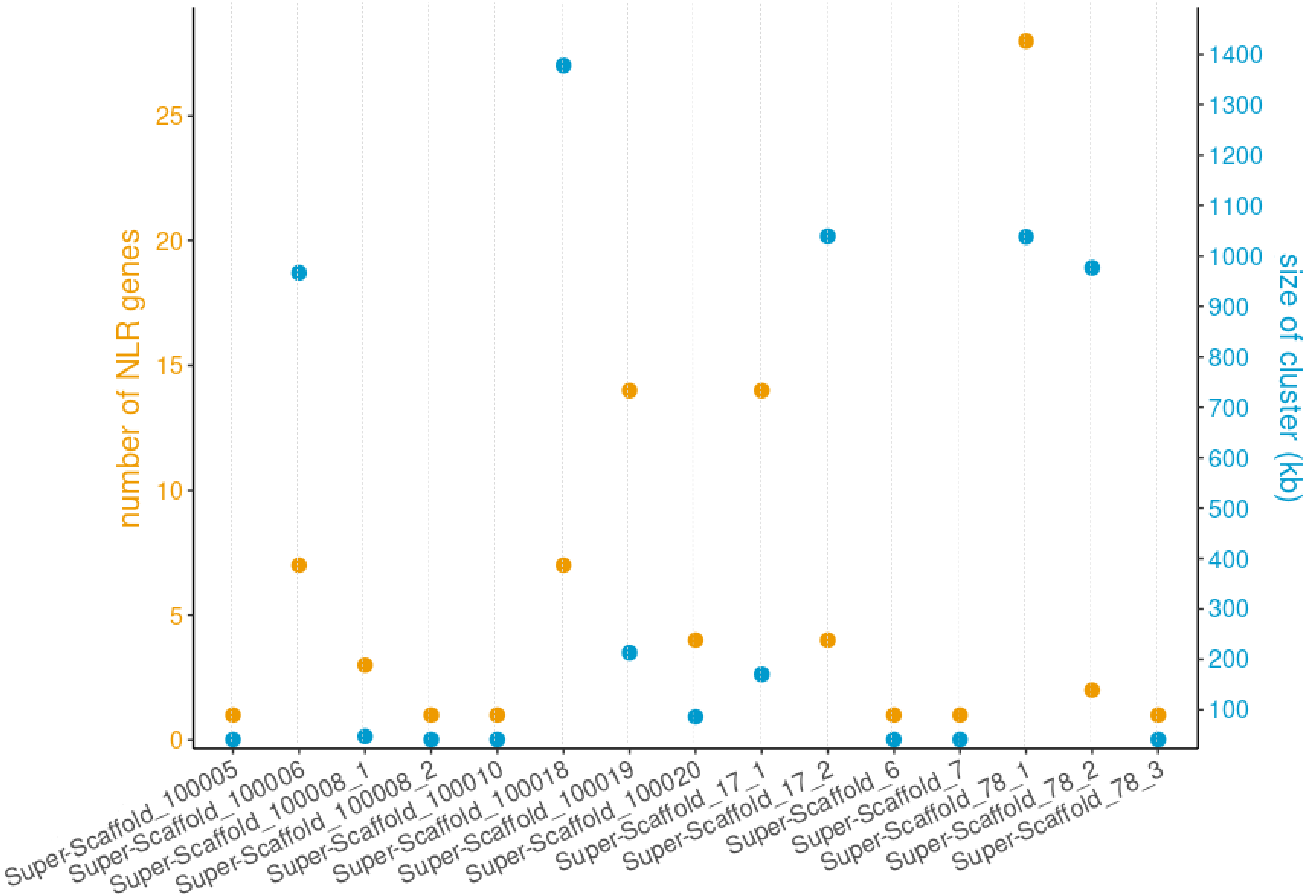
C.melo variety Anso77

84 NLR genes in 15 clusters

NLRGenomeSweeper *(Toda et al., 2020)*

Our target regions: 15 NLR-gene clusters

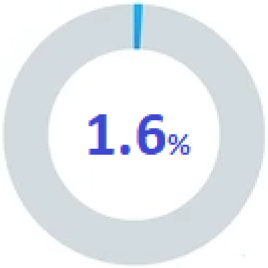
Reference genome → Variety **<u>Anso77</u>** (draft genome produced at INRAE-GAFL/EPGV)



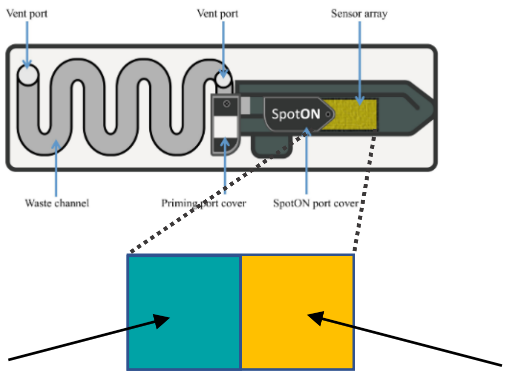
Reference genome (.fasta)
+
Target regions (.bed)

<u>Clusters</u>:
NLR genes + 20kb-extra flanking
Variable in size
Variable in number of genes

Genome length ~380Mb
Target regions length ~6.16Mb



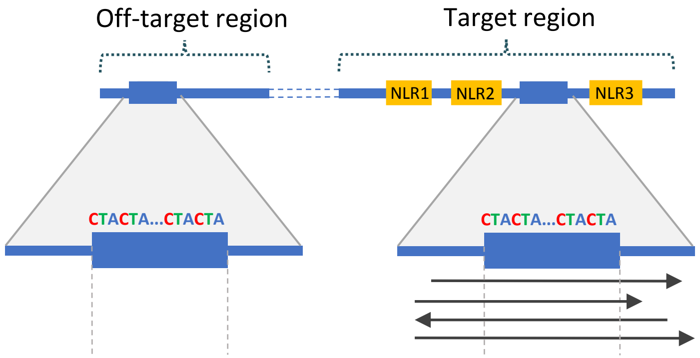
Our first experience: Variety Anso-77 (same as provided reference)



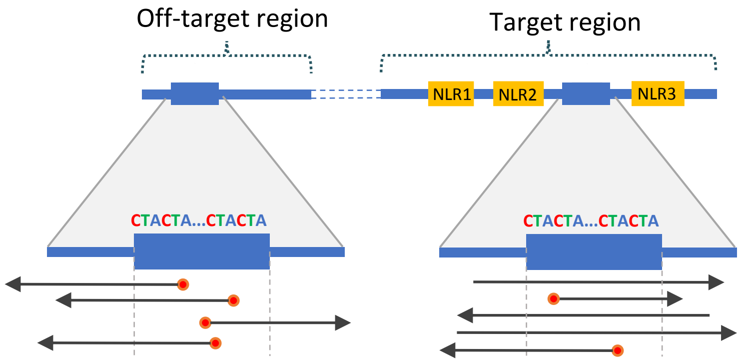
NAS

Standard sequencing

Masking repetitive elements >200kb



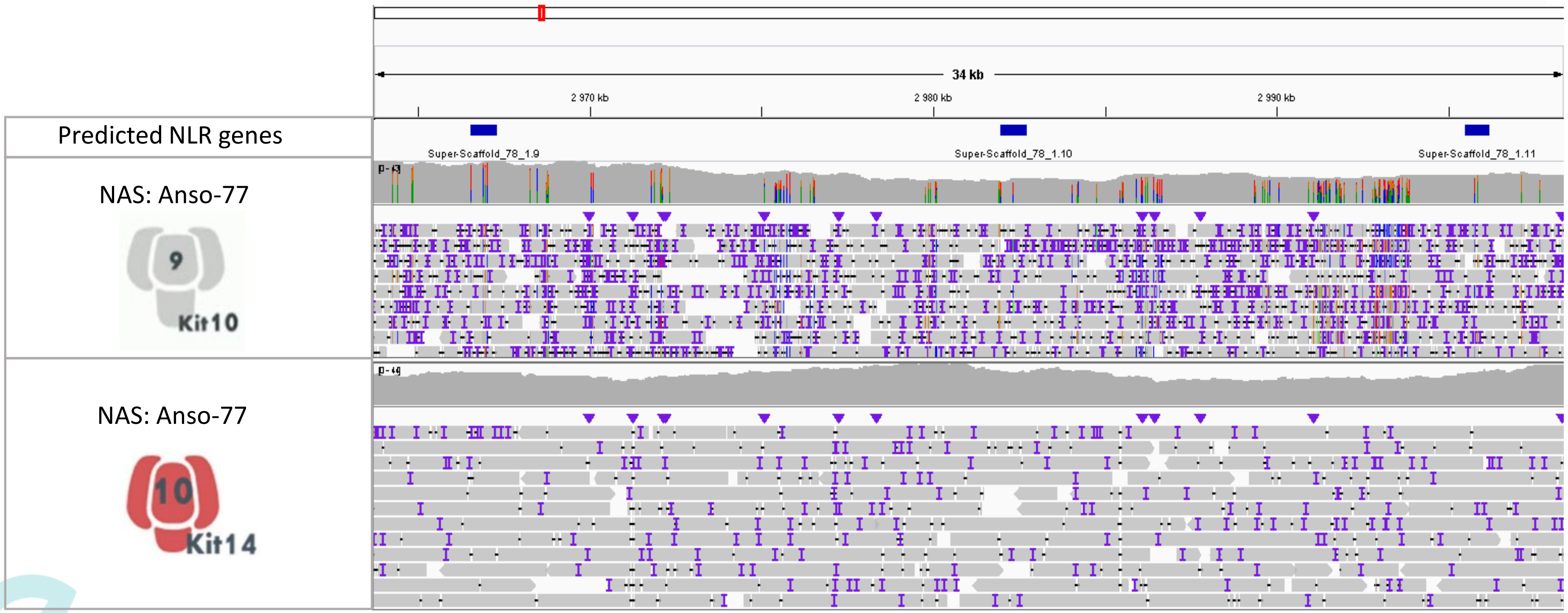
Not masking repetitive elements

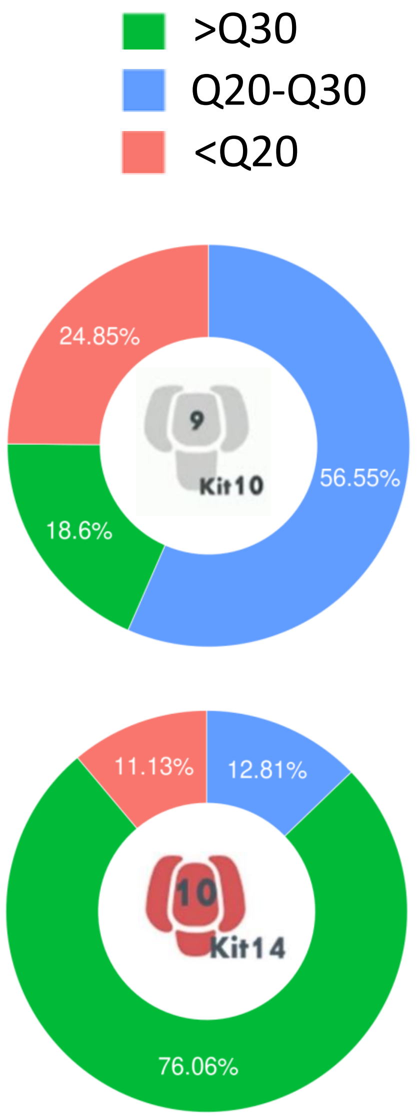


Nanopore kit evolution: From kit10 to kit14

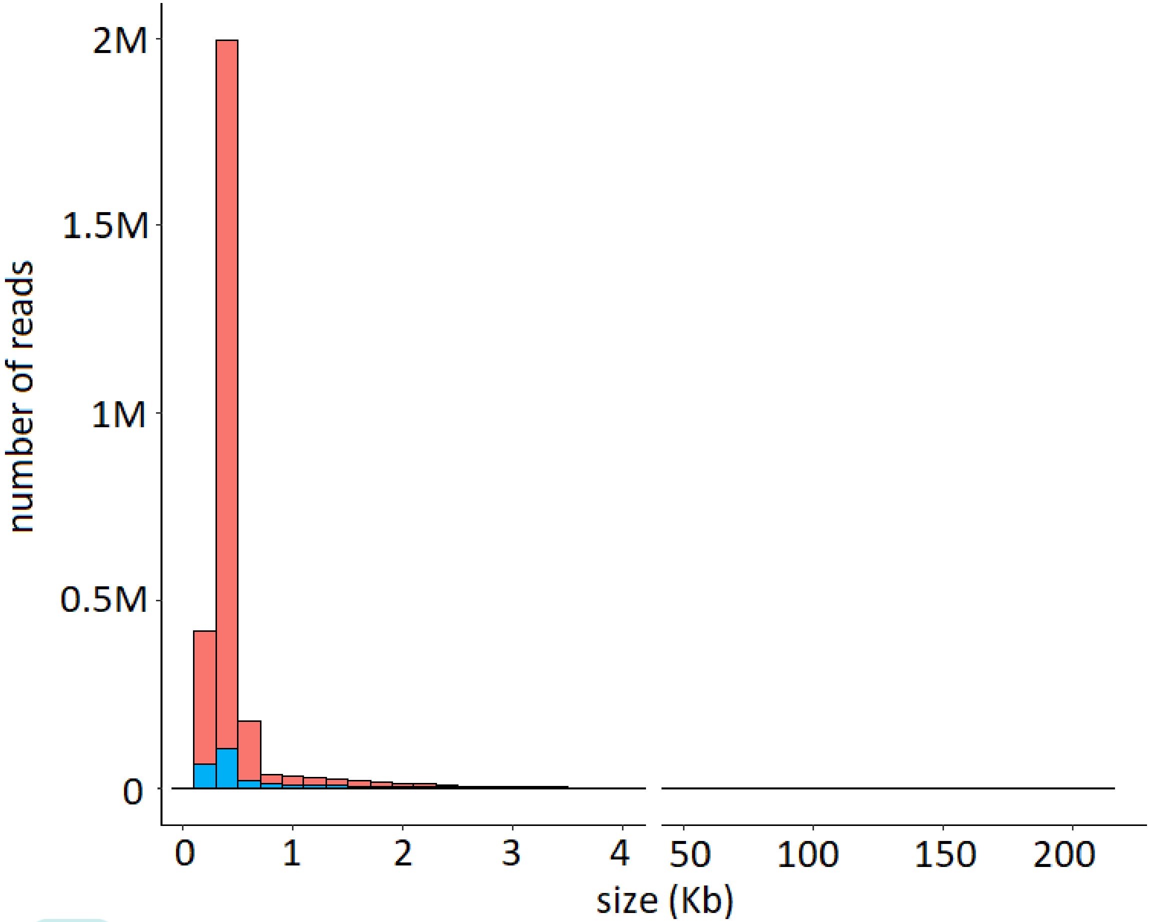
Kit14
R10.4.1
+
E8.2

- **Increased accuracy**, improved sequencing **speed**
- **Reactive consumption reduced**, adaptors capture improved
- Duplex sequencing, **better throughput**

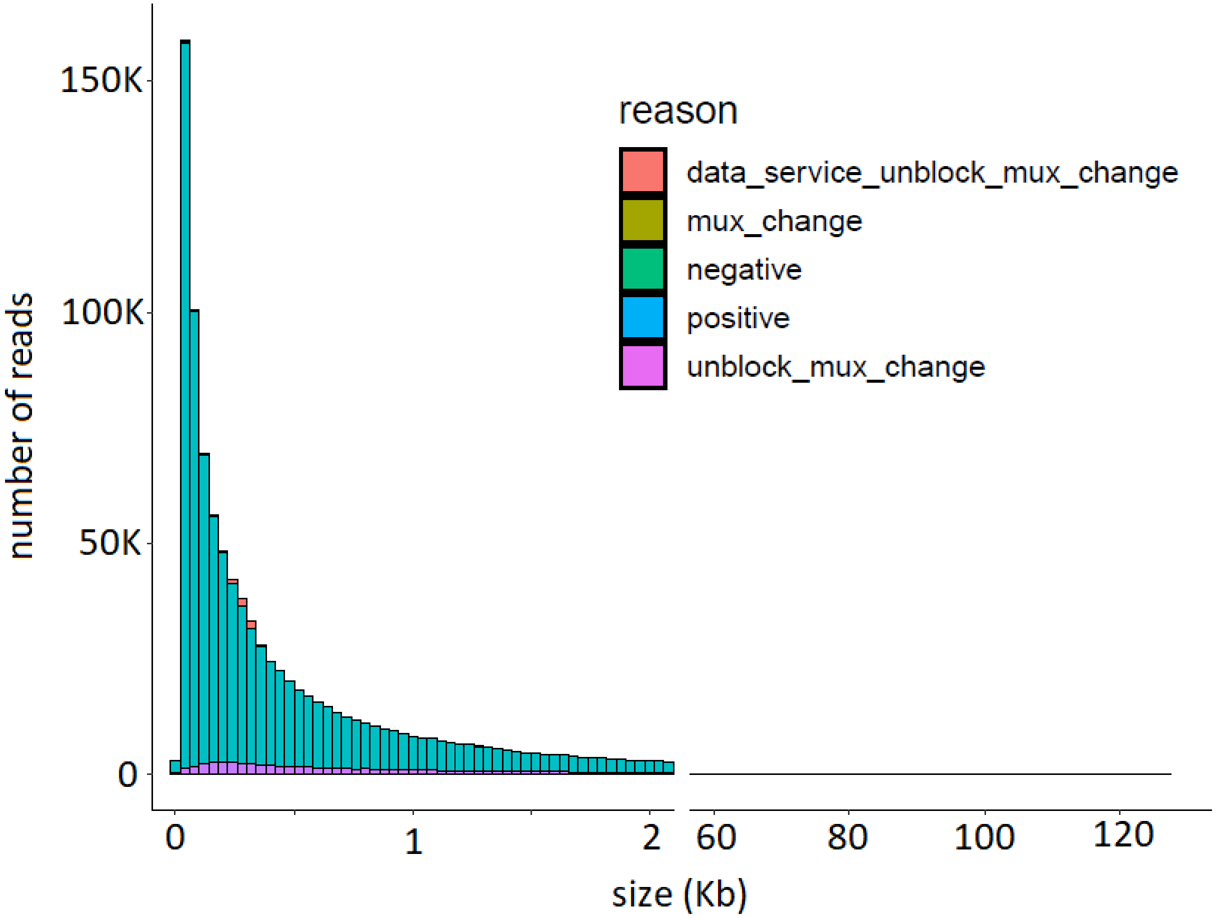




NAS: A characteristic distribution of read lengths

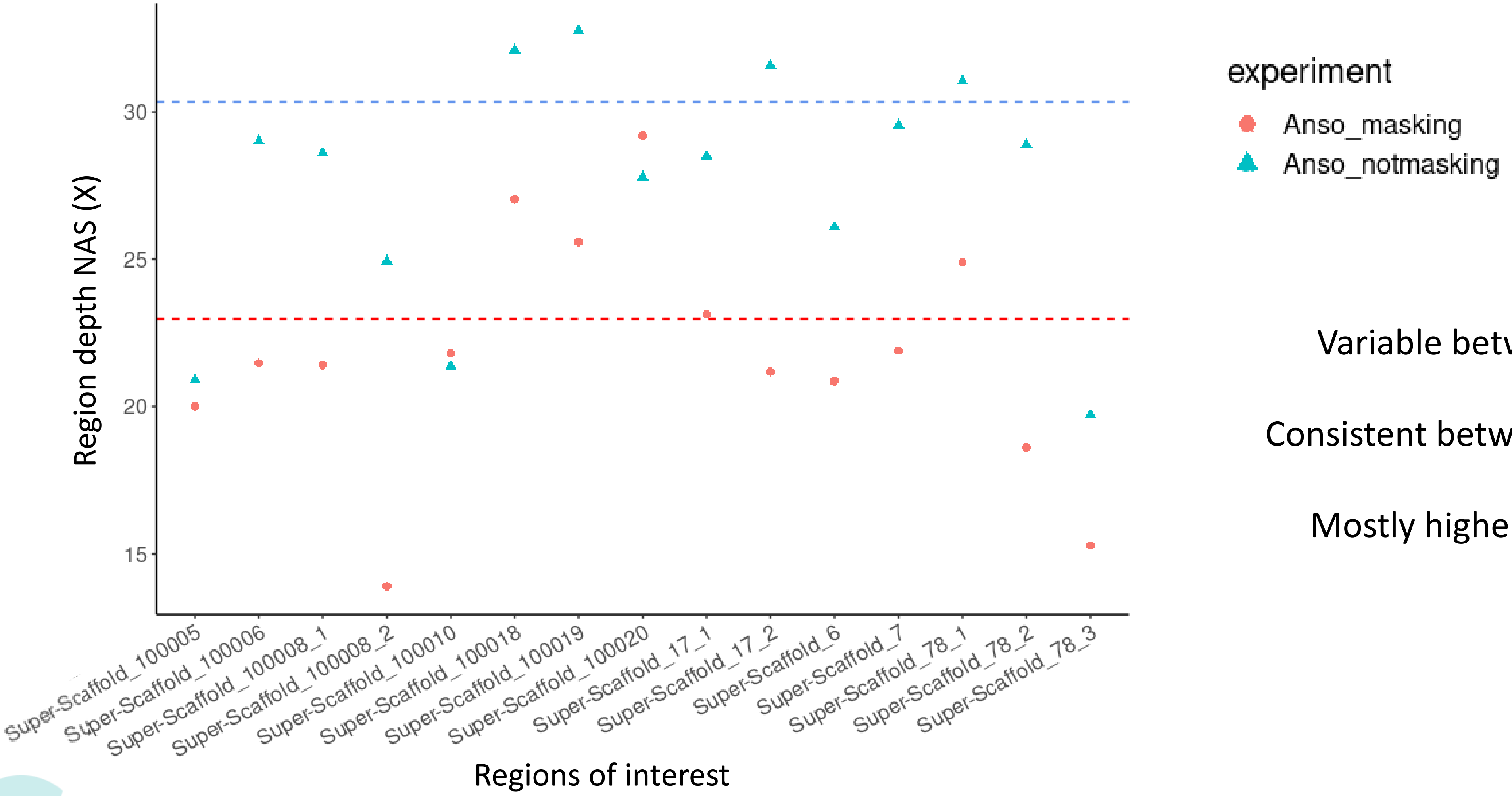


Nanopore adaptive sampling



Standard sequencing

Statistics: regions depth

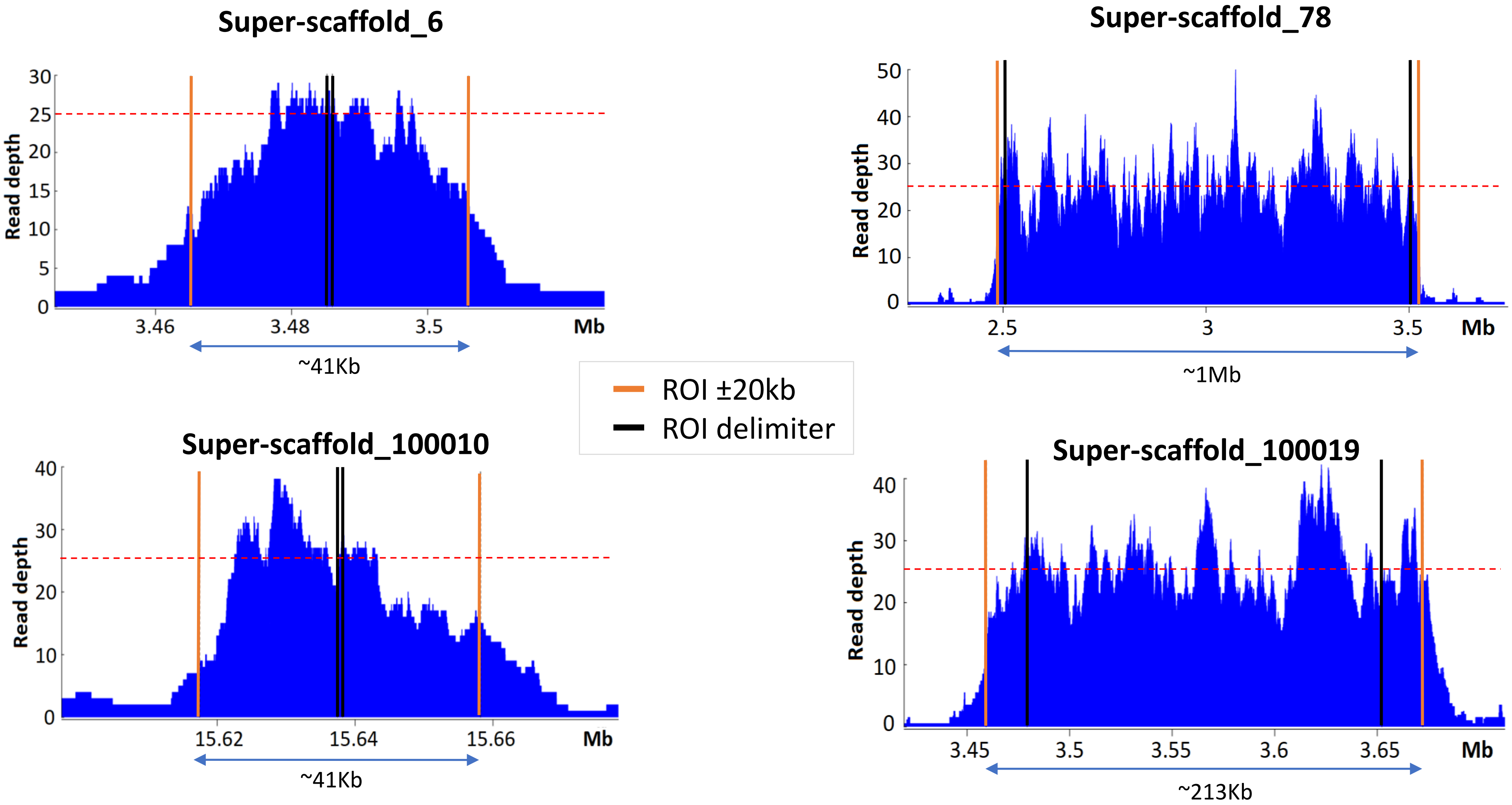


Variable between clusters

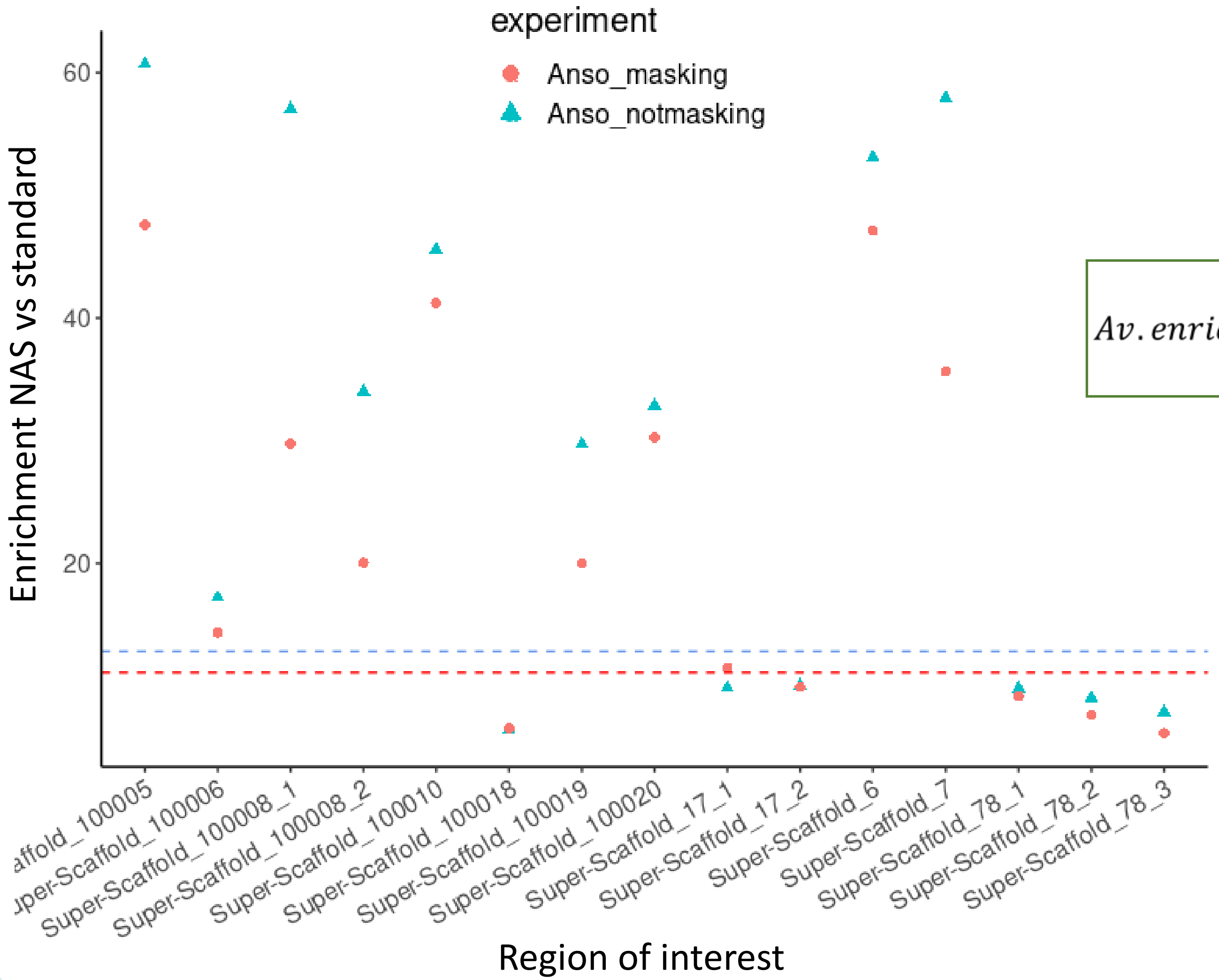
Consistent between experiences

Mostly higher not masking

Visual enrichment of our target regions



Good enrichment of our target regions



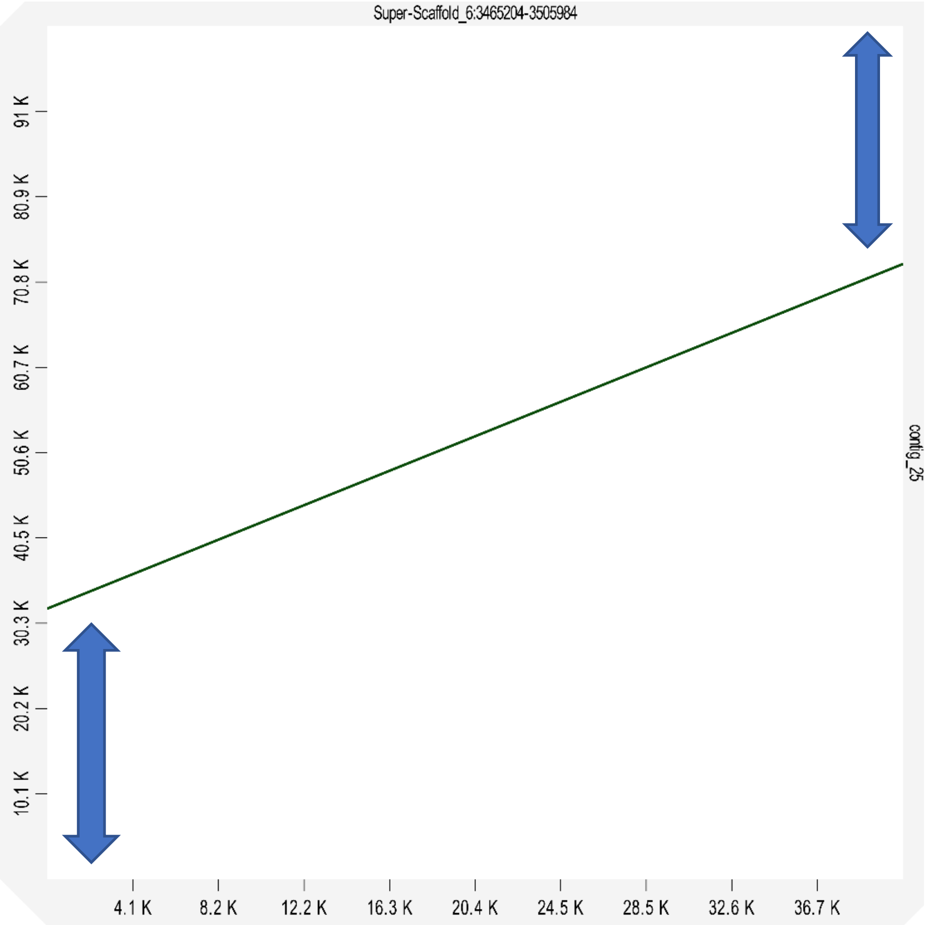
$$Enrichment_{region} = \frac{\frac{region\_cov\_NAS}{chrom\_cov\_NAS}}{\frac{region\_cov\_standard}{chrom\_cov\_standard}}$$

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

- Variable between clusters
- Variable between experiences
- But consistent between experiences for clusters under 20X
- Average ~equal between experiences

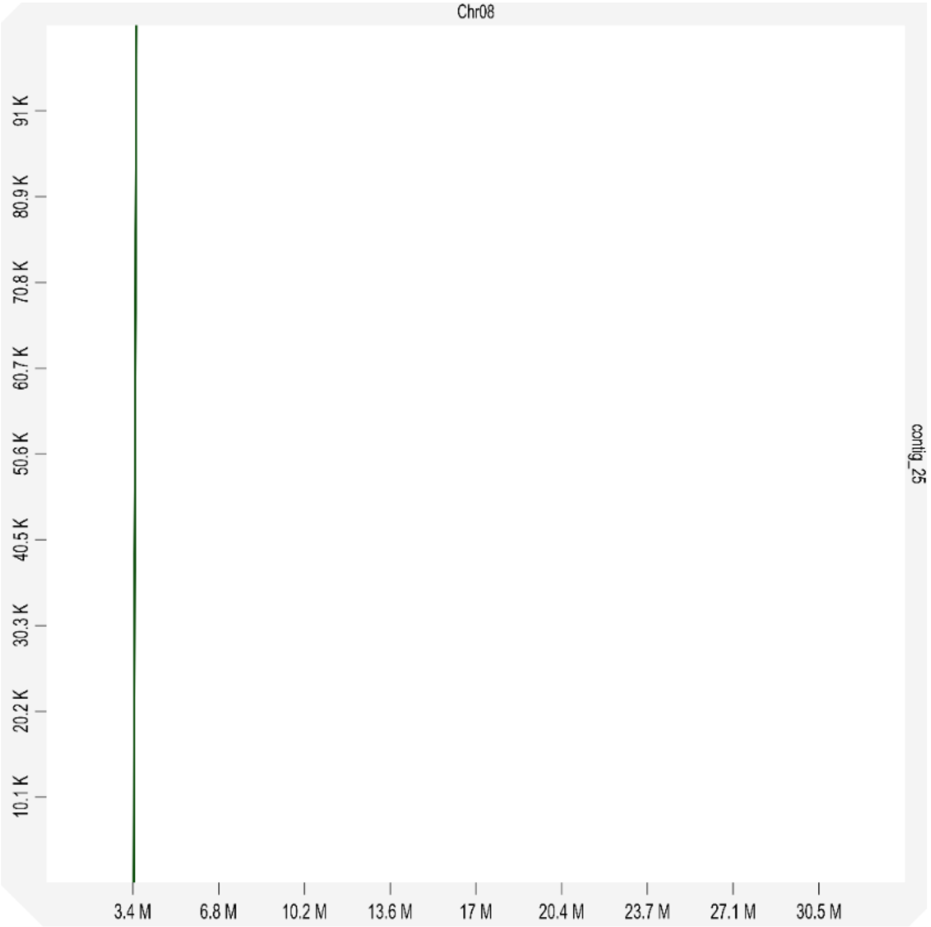
Genome assemblies of target regions: statistics

Anso-77 reference: one gene cluster



Anso-77 assembly

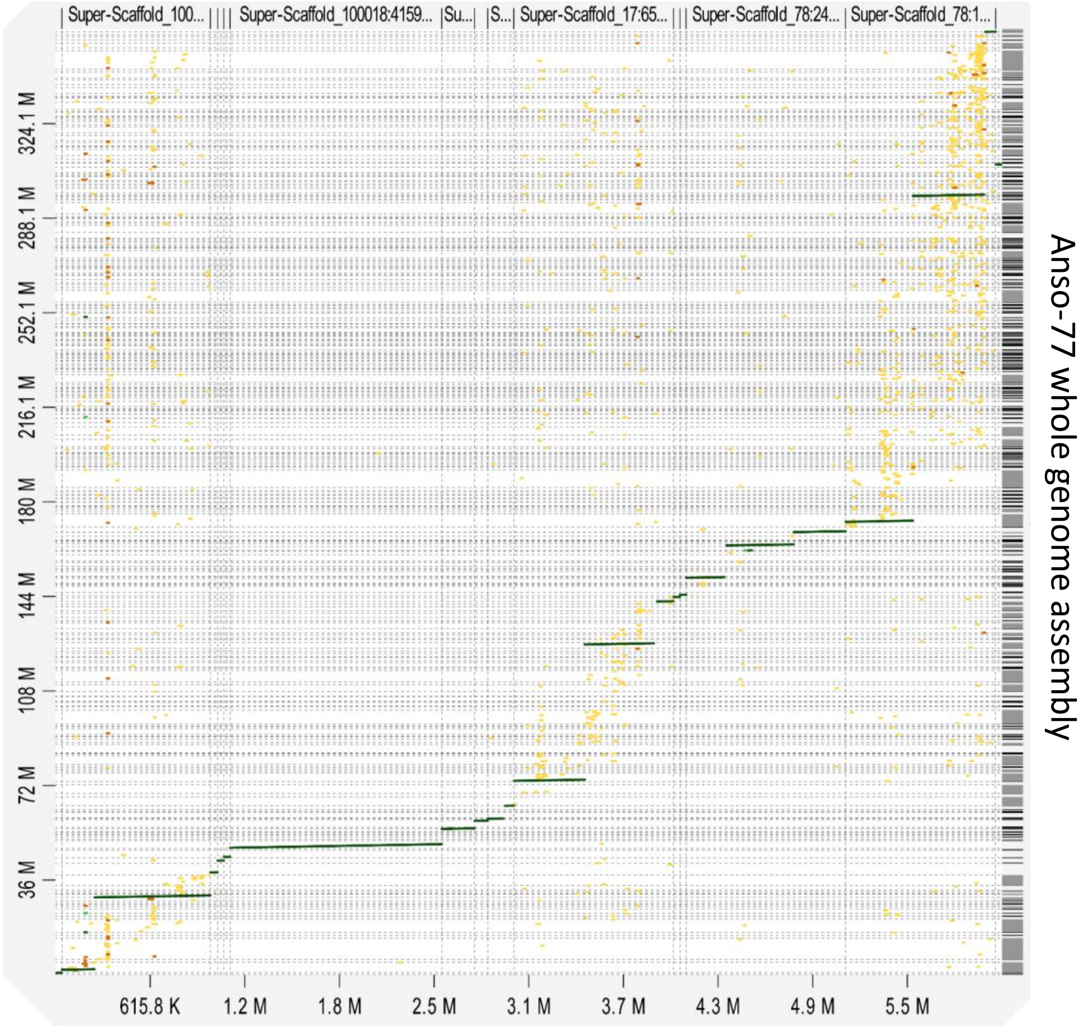
Anso-77 reference: complete chromosome



Anso-77 assembly

Genome assemblies: Good assembly of target regions

Anso-77 reference: 15 gene clusters

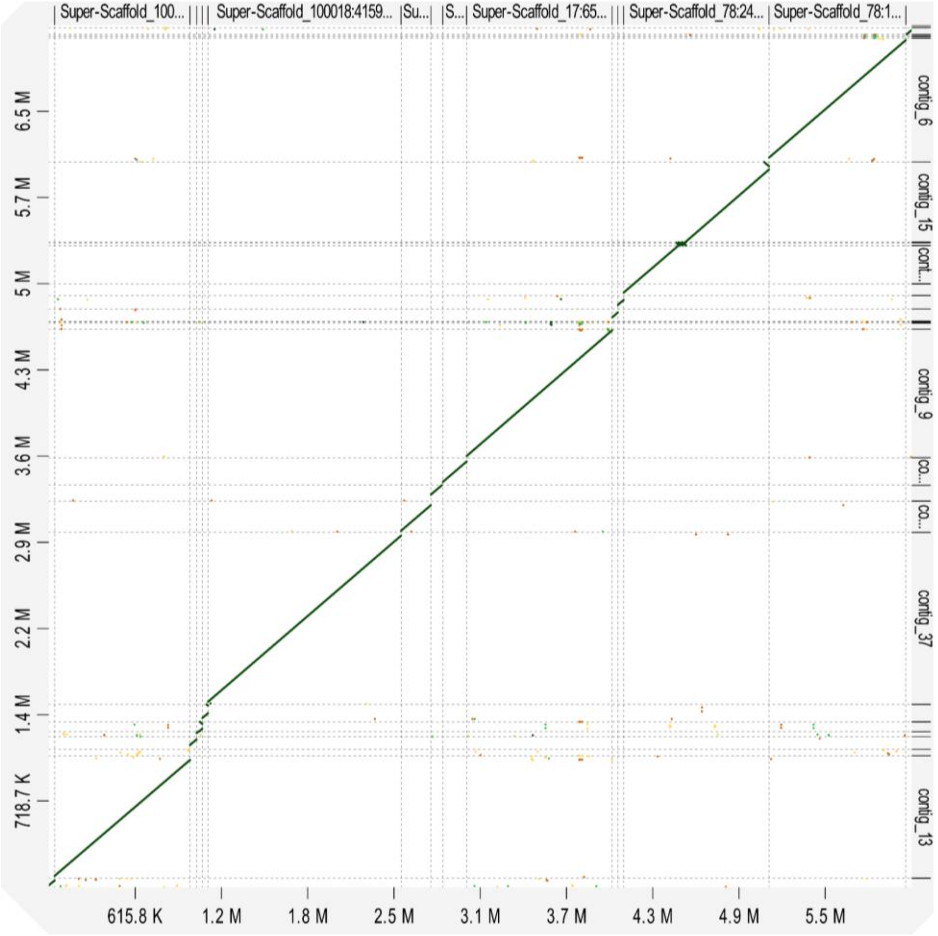


Assembly using WGS data:

- Target regions **fragmented** into multiple contigs
- Target regions **poorly covered** compared to adaptive sampling

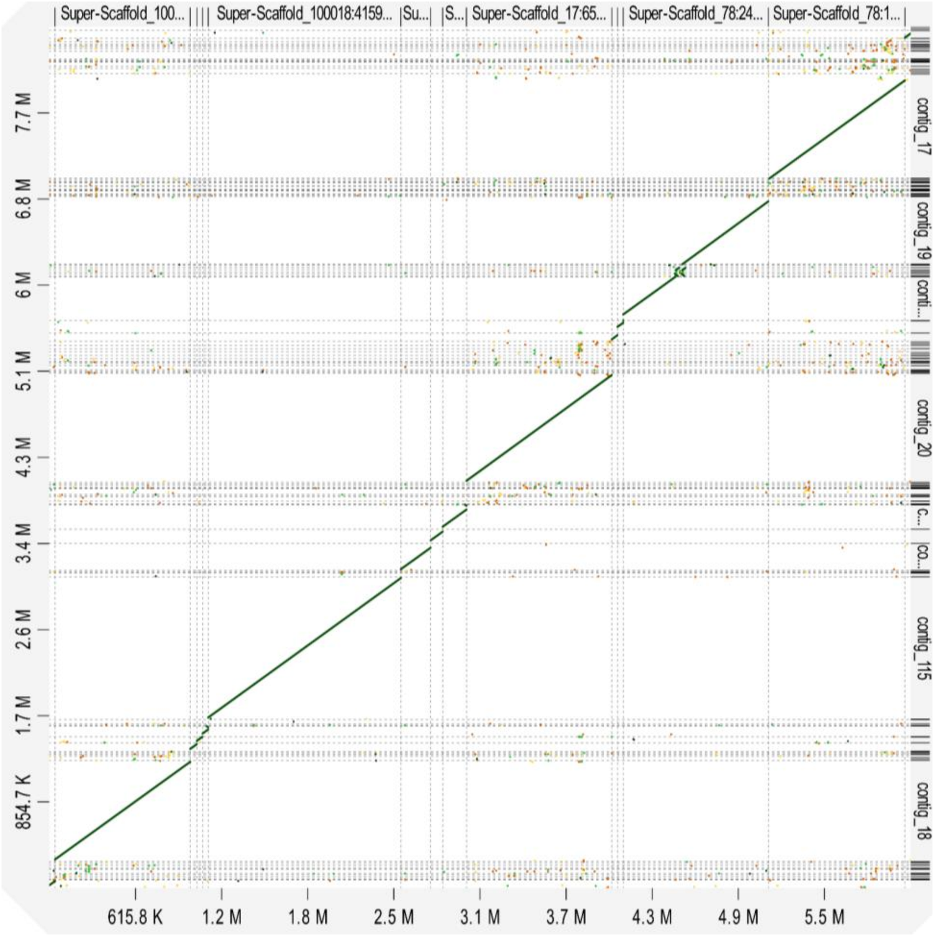
Genome assemblies: Good assembly of target regions

Anso-77 reference: 15 gene clusters



Anso-77 assembly **not** masking

Anso-77 reference: 15 gene clusters



Anso-77 assembly masking

Target regions fully assembled with a good depth

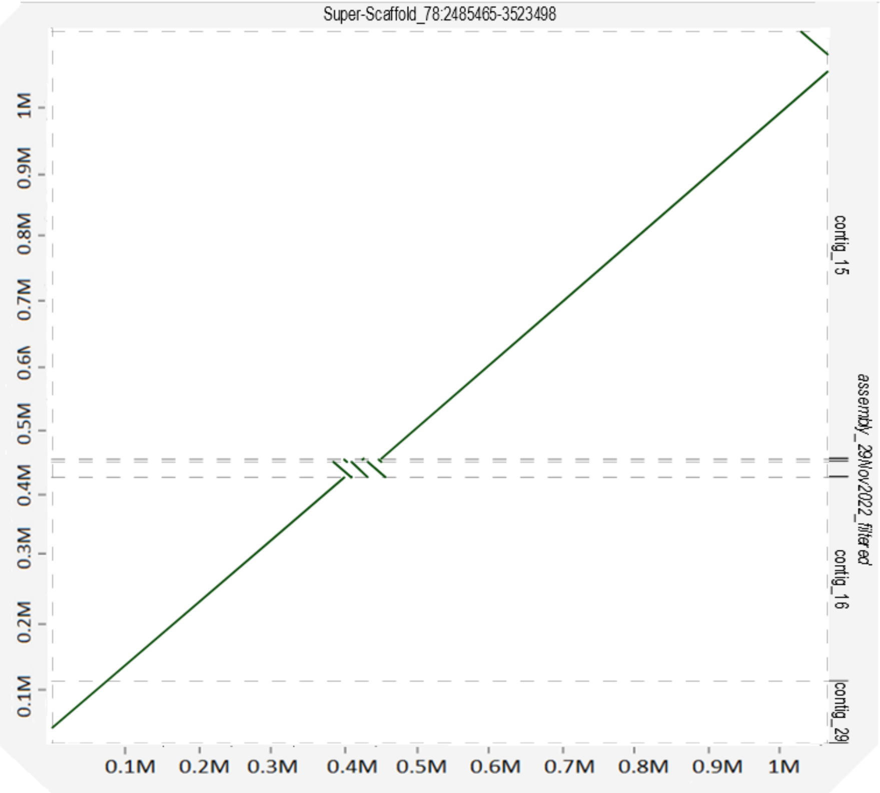
Genome assemblies: cluster Vat

Cluster Vat: Most complex cluster of NLR genes in melon

- 28 NLR genes
- 1Mb length

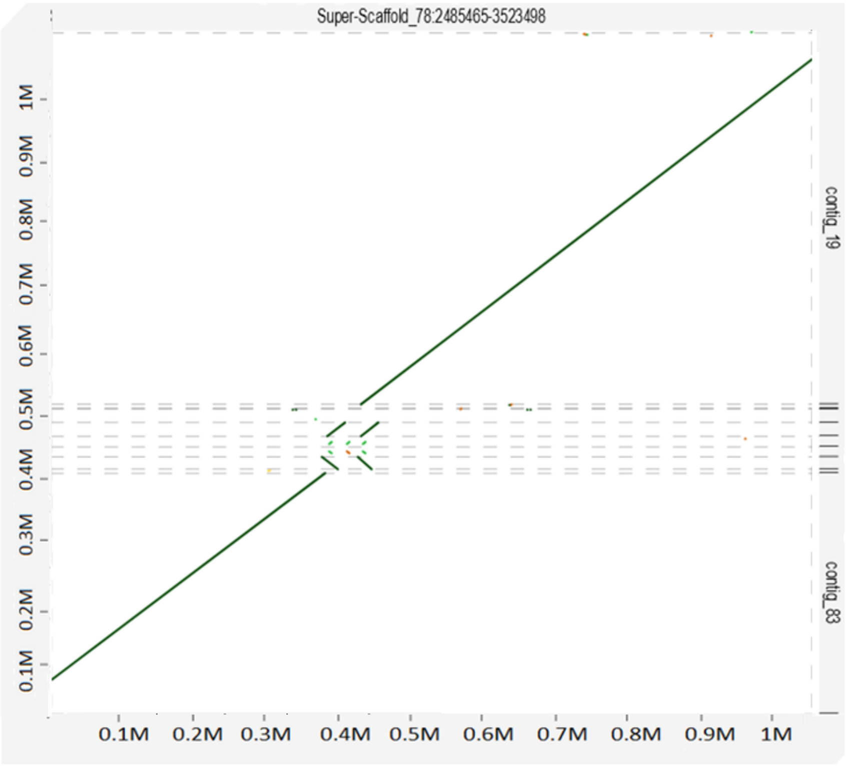
But also the most studied one (INRAE-GAFL)

Anso-77 reference: cluster Vat



Anso-77 assembly cluster Vat **not** masking

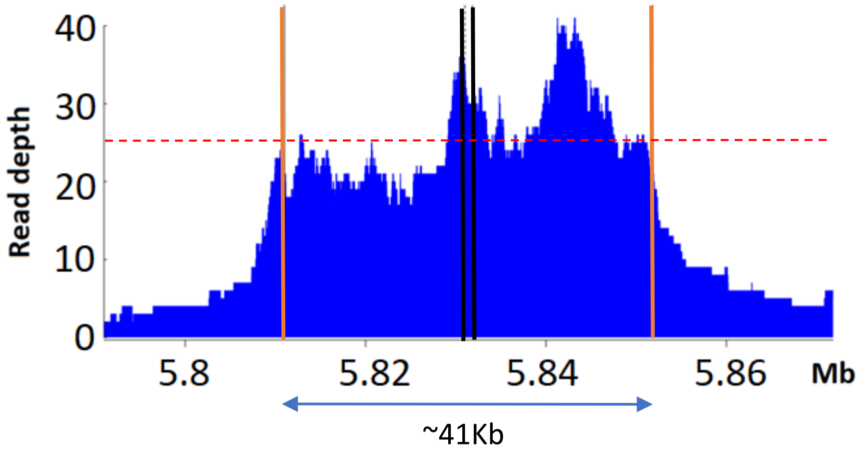
Anso-77 reference: cluster Vat



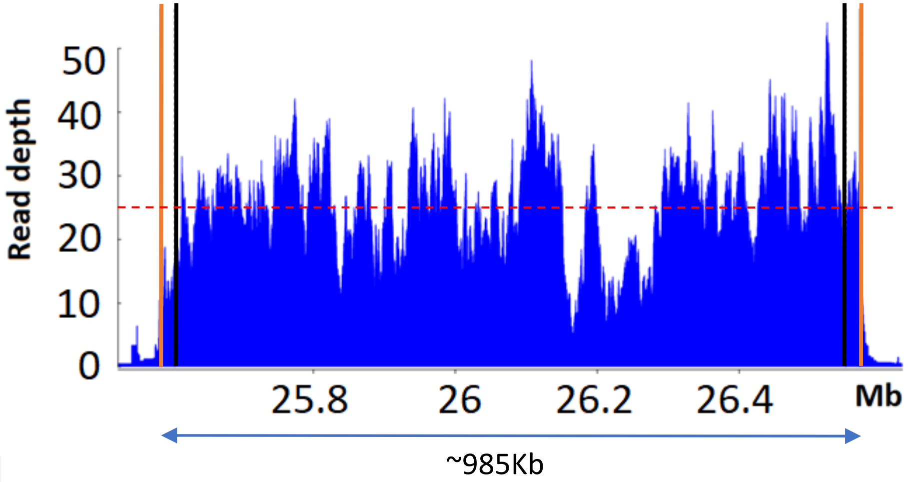
Anso-77 assembly cluster Vat masking

Preliminary results using a different variety: Doublon

Chr06

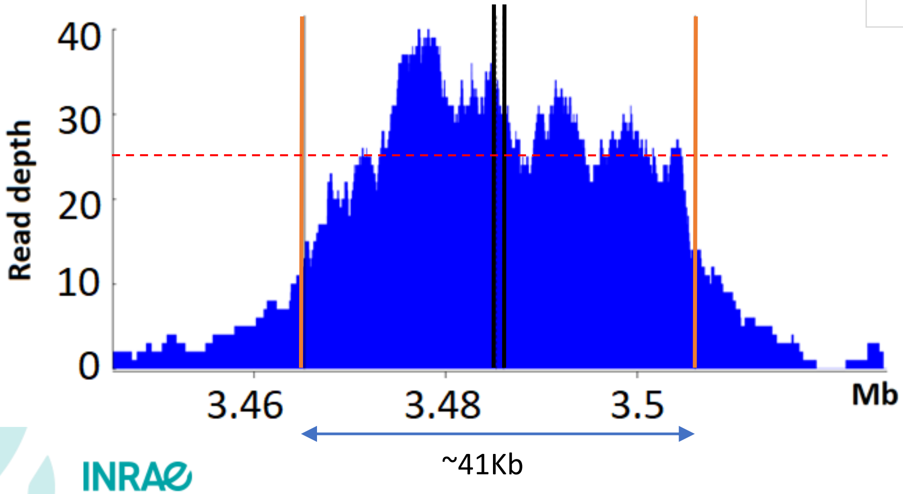


Chr05

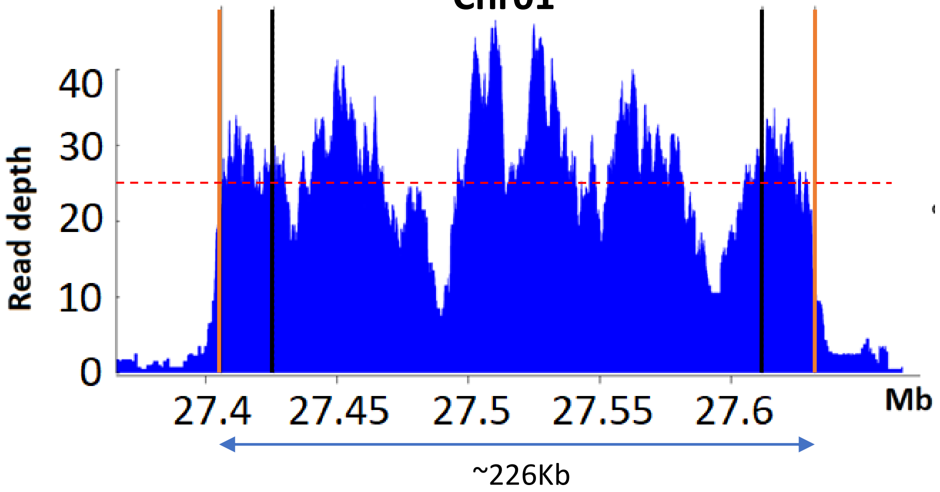


- 🟧 ROI ±20kb
- ⬛ ROI delimiter

Chr08



Chr01



Preliminary results using a different variety: Doublon

Vat genes

Doublon mapped on **Anso** genome (Vat cluster)

(Adapted from Chovelon et al., 2021)

Perspectives

Short-term perspectives:

1. Run on PromethION using a variety **genetically distant** from the provided reference: ‘Canton’
2. **Multiplexing** on PromethION using **two barcodes**: Varieties ‘Anso’ + ‘Doublon’

Middle-term perspectives:

1. **Increase the 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 diversity** of melon **NLR genes** using a **large amount of varieties**.

Acknowledgements

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

INRAE-GAFL
Nathalie Boissot
Veronique Chovelon
Jacques Lagnel
Vincent Rittener-Ruff

Genoscope
Stefan Engelen
Corinne Cruaud

Oxford Nanopore Technologies
François Berdou
Helen Neal
Barbara Ottolini

Disclaimer:
Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.