



**HAL**  
open science

# Exploitation de la puissance du séquençage pour l'amélioration des plantes

Damien Hinsinger

► **To cite this version:**

Damien Hinsinger. Exploitation de la puissance du séquençage pour l'amélioration des plantes. Mini congrès, CEA - Genoscope, Jun 2022, Evry, France. hal-03937956

**HAL Id: hal-03937956**

**<https://hal.inrae.fr/hal-03937956>**

Submitted on 13 Jan 2023

**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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License



# ➤ Exploitation de la puissance du séquençage pour l'amélioration des plantes

Damien Hinsinger

Unité INRAE EPGV – Etude du Polymorphisme des Génomes Végétaux

# ➤ Etude du polymorphisme des **génomomes végétaux**

## **3 génomes chez les plantes:**

Nucléaire  
Mitochondrial  
Chloroplastique

## **Génomomes de taille variable**

Arabidopsis: 120 Mb  
Tomate : 950 Mb  
Mais: 2 300 Mb  
Blé tendre: 17 000 Mb

## **Structure parfois polyploïde du génome**

Blé dur 4X ,Blé tendre 6X  
Rose de 2X à 10X

## **Diversité génétique large**

Espèces cultivées  
Espèces sauvages



# ➤ Etude du polymorphisme des génomes végétaux

Variété de polymorphismes allant

D'une paire de bases : SNP

De quelques paires de base : petits indels, PAV, SSR (microsatellites)

De nombreuses paires de bases/portion de chromosome : variations structurales (SV)



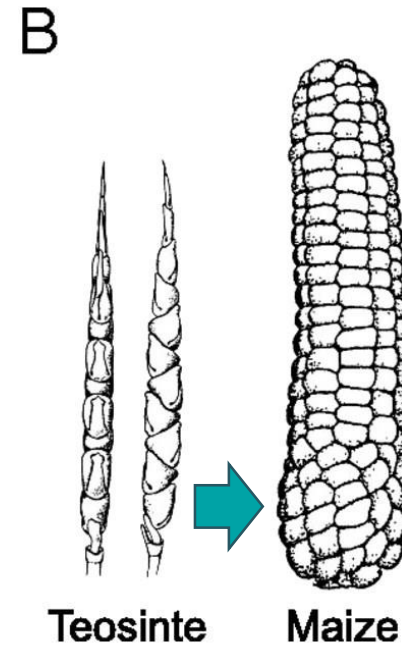
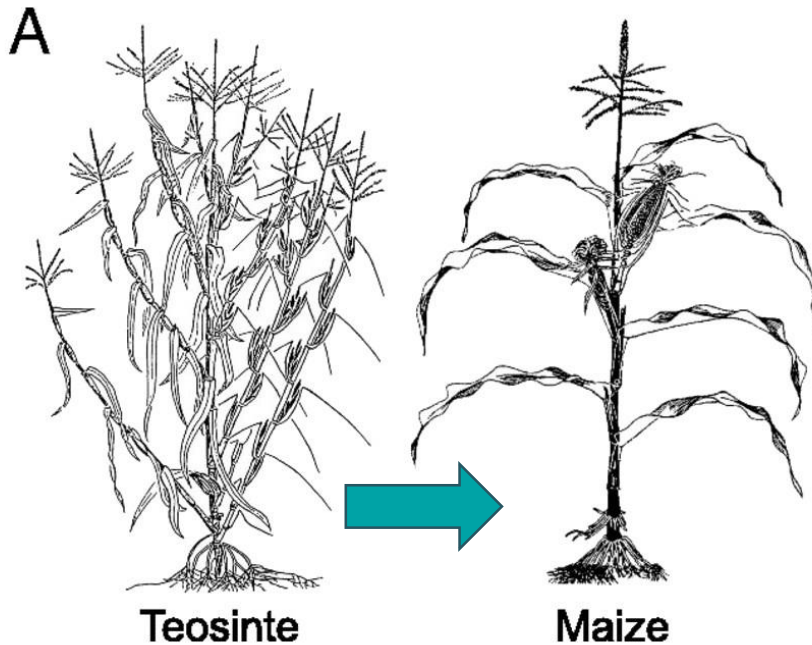
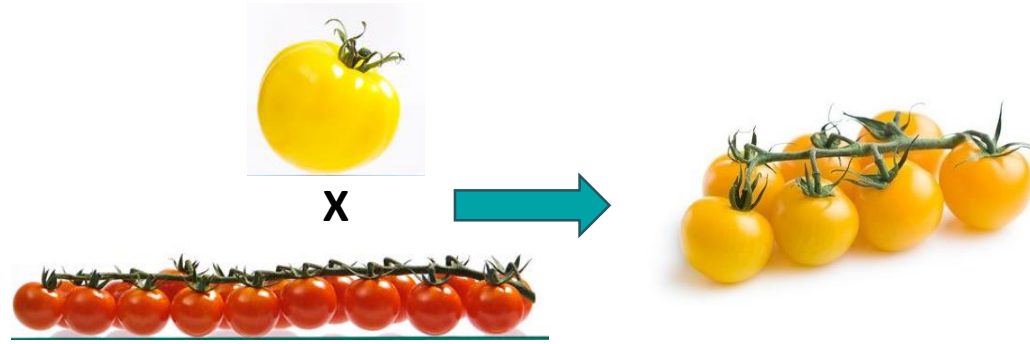
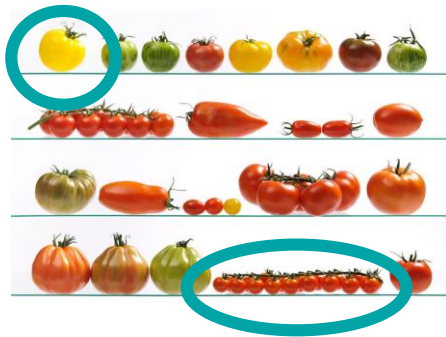
Génotypage d'individus pour 1 ou plusieurs

marqueurs génétiques (SNP ou SV) **SV: Structural Variant**  
(variation structurale).

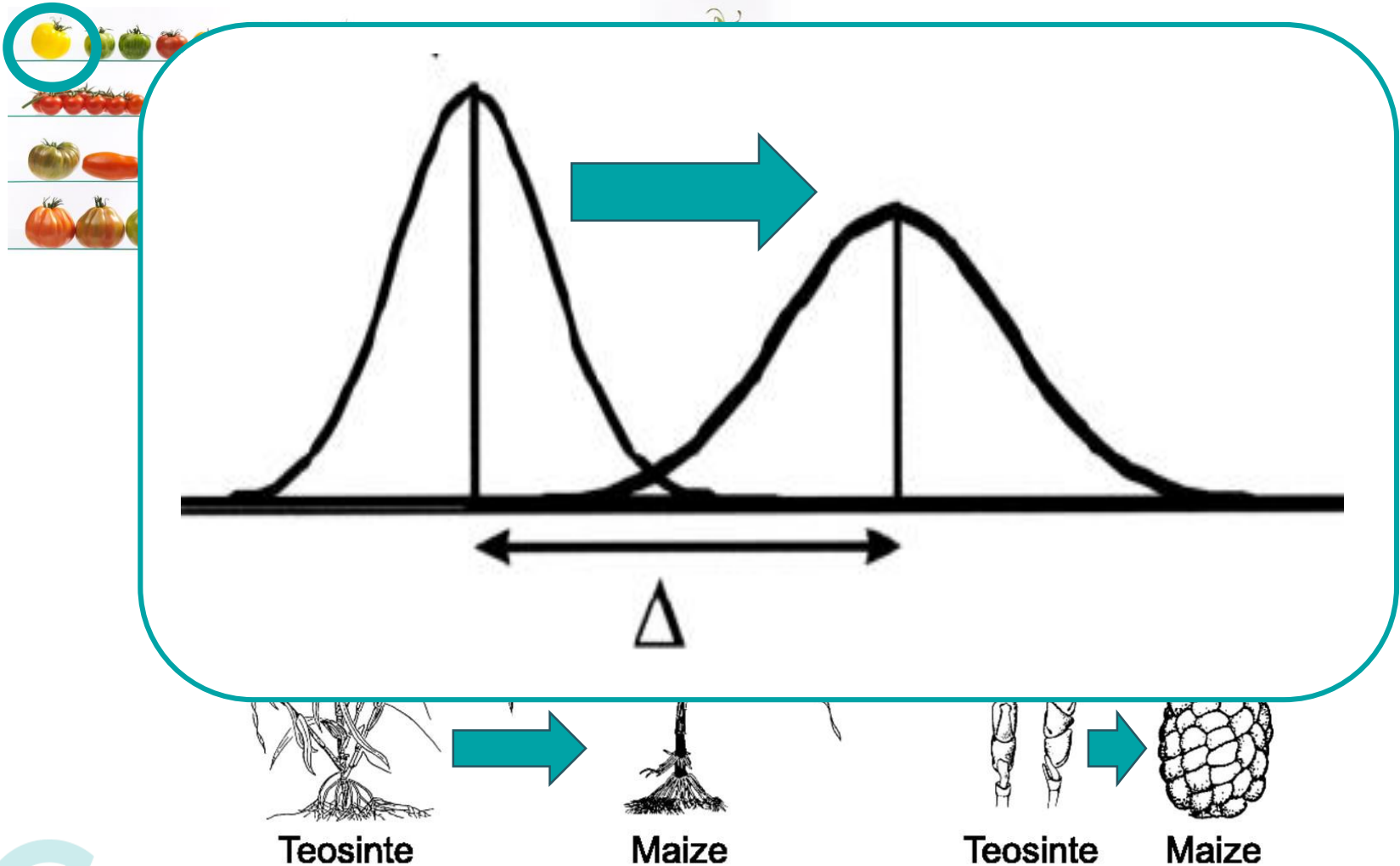
**SNP: Single Nucleotide Polymorphism**  
(polymorphisme au niveau d'un seul  
nucléotide).



# ➤ L'amélioration des plantes



# ➤ L'amélioration des plantes



# ➤ L'amélioration des plantes



## Besoins

Identifier les polymorphismes (SNPs / SVs) existants dans les populations « pré-breeding » pour les relier aux caractères d'intérêts

Suivre précisément les polymorphismes liés aux caractères d'intérêts dans les descendances issues de croisements



Teosinte



Maize



Teosinte



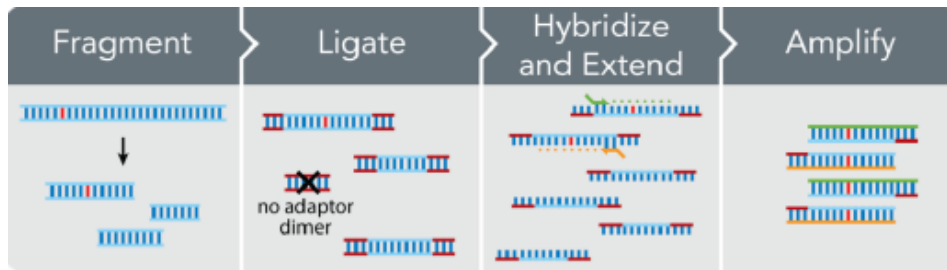
Maize



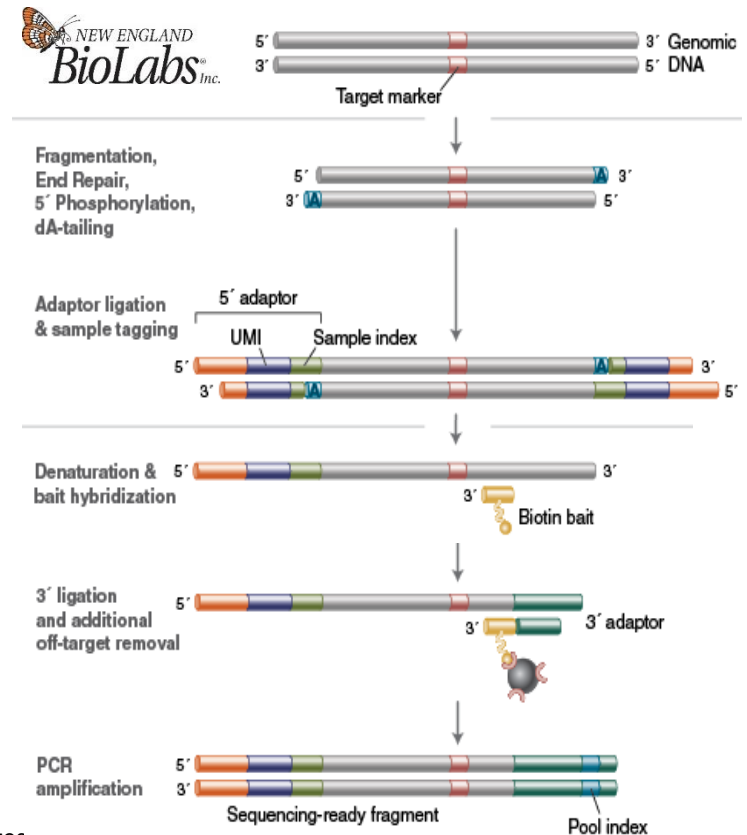
# ➤ Génotypage de SNPs ciblés par séquençage

Principe:

Hybridation et extension par sonde spécifique (Allegro) / capture du fragment ADN par sonde spécifique (NEB)  
 Amplification puis séquençage NGS partiel du brin synthétisé



Multiplexage jusqu'à 3072 (Allegro) / 768 (NEB) ind. per lane  
 500-80 000 SNP (Allegro) / 100-5 000 SNPs (NEB) en 1 seul design.



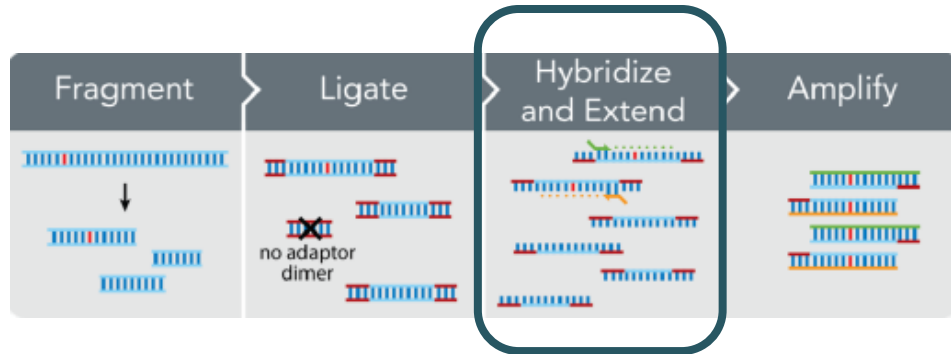
Exploitation de la puissance du séquençage pour l'amélioration des plantes

14 juin 2022 / Mini-congrès Genoscope / Damien Hinsinger

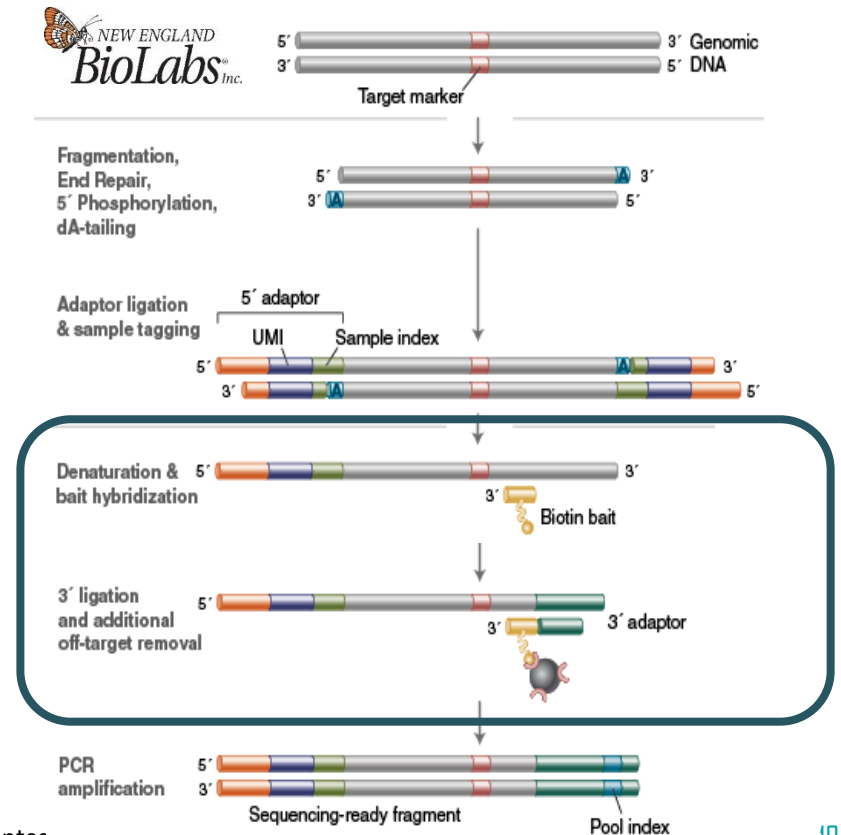
# ➤ Génotypage de SNPs ciblés par séquençage

Principe:

Hybridation et extension par sonde spécifique (Allegro) / capture du fragment ADN par sonde spécifique (NEB)  
 Amplification puis séquençage NGS partiel du brin synthétisé



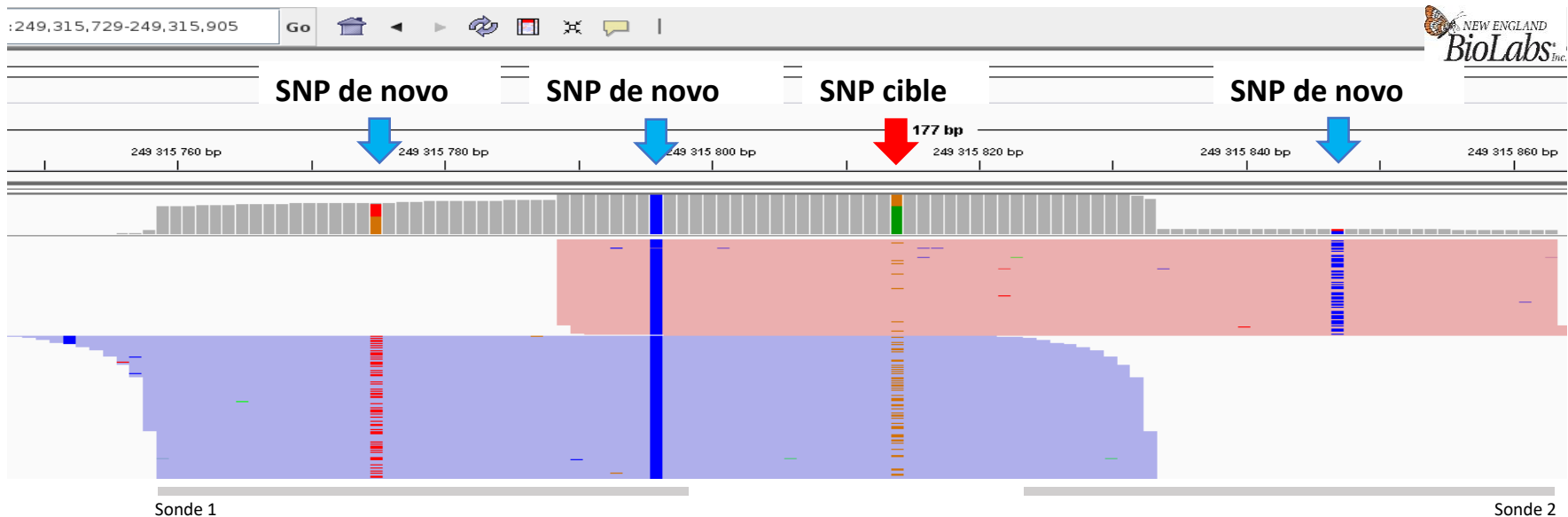
Multiplexage jusqu'à 3072 (Allegro) / 768 (NEB) ind. per lane  
 500-80 000 SNP (Allegro) / 100-5 000 SNPs (NEB) en 1 seul design.



Exploitation de la puissance du séquençage pour l'amélioration des plantes

14 juin 2022 / Mini-congrès Genoscope / Damien Hinsinger

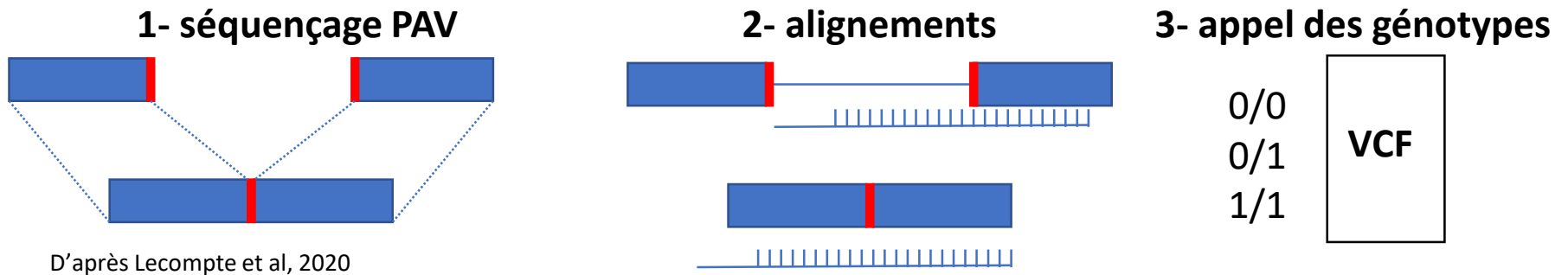
# ➤ SNPs *de novo* pour Diversité génétique



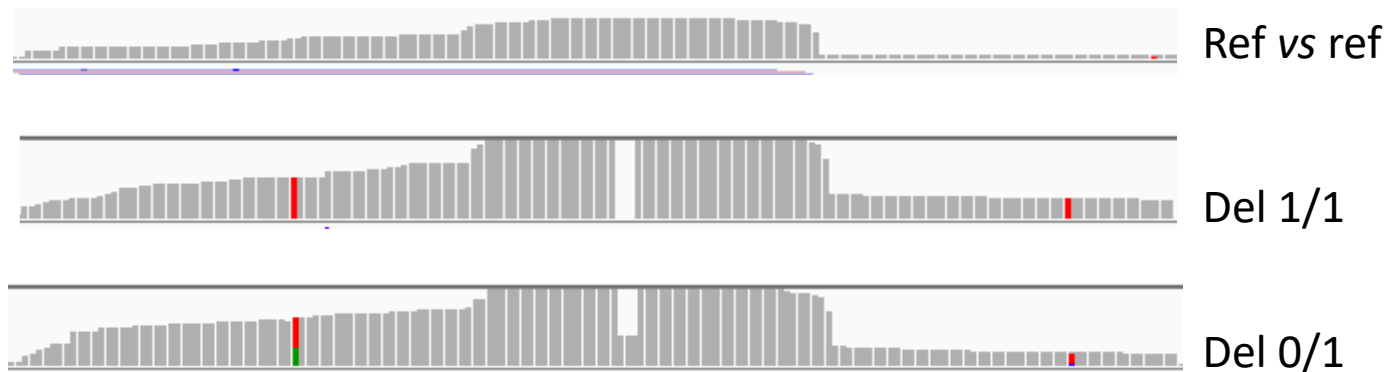
	NEB Next Direct Genotyping Solution		Allegro Genotyping	
	<i>Amaizing</i> (maïs) 110 SNPs		<i>Resilens</i> (lentille) 25 000 SNPs	<i>Amaizing</i> (maïs) 2 500 SNPs
<b>De novo SNPs + indels 1bp</b>				
Nb SNPs <i>de novo</i>	650 ~6x cibles		188 159 ~7,5x cibles	60 269 ~23x cibles
Distance au SNP ciblé	+- 100 nt		+- 100 nt (Ref partielle)	+-400 nt (Ref génome)
<b>Génotypes</b>				
Répétabilité	80%		84%	84,1 %
Données manquantes	22,3 % (12,4 % Ref)		17,5 % (12,1 % Ref)	14,1 % (Ref)

# ➤ Au-delà du SNP, quelles sont les perspectives?

Adapter les outils « génotypage ciblé par séquençage » au génotypage de variants structuraux (SVs)



✓ Exemple : Génotypage d'une délétion de 2 bases-alignement



# ➤ Variants Structuraux (poster Romane Guilbaud)

## Impliqués dans traits agronomiques / interactions biotiques

- Détection idéale en long reads (\$\$)
- Détection imprécise et fort taux de faux positifs (=FDR) en short reads
  - ➔ détection *outdatée*
  - ➔ génotypage de SVs connus possible, mais long reads pour la détection (\$\$)



Difficile à grande échelle

Besoin d'un outil pour filtrer les faux positifs dans pop. d'amélioration

# ➤ Variants Structuraux (poster Romane Guilbaud)

## Impliqués dans traits agronomiques / interactions biotiques

- Détection idéale en long reads (\$\$)
- Détection imprécise et fort taux de faux positifs (=FDR) en short reads
  - ➔ détection *outdatée*
  - ➔ génotypage de SVs connus possible, mais long reads pour la détection (\$\$)



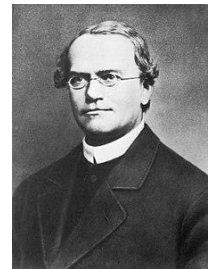
Difficile à grande échelle

## Besoin d'un outil pour filtrer les faux positifs dans pop. d'amélioration

Si trios parents/descendant disponibles (souvent – croisements)

➔ Ségrégation mendélienne

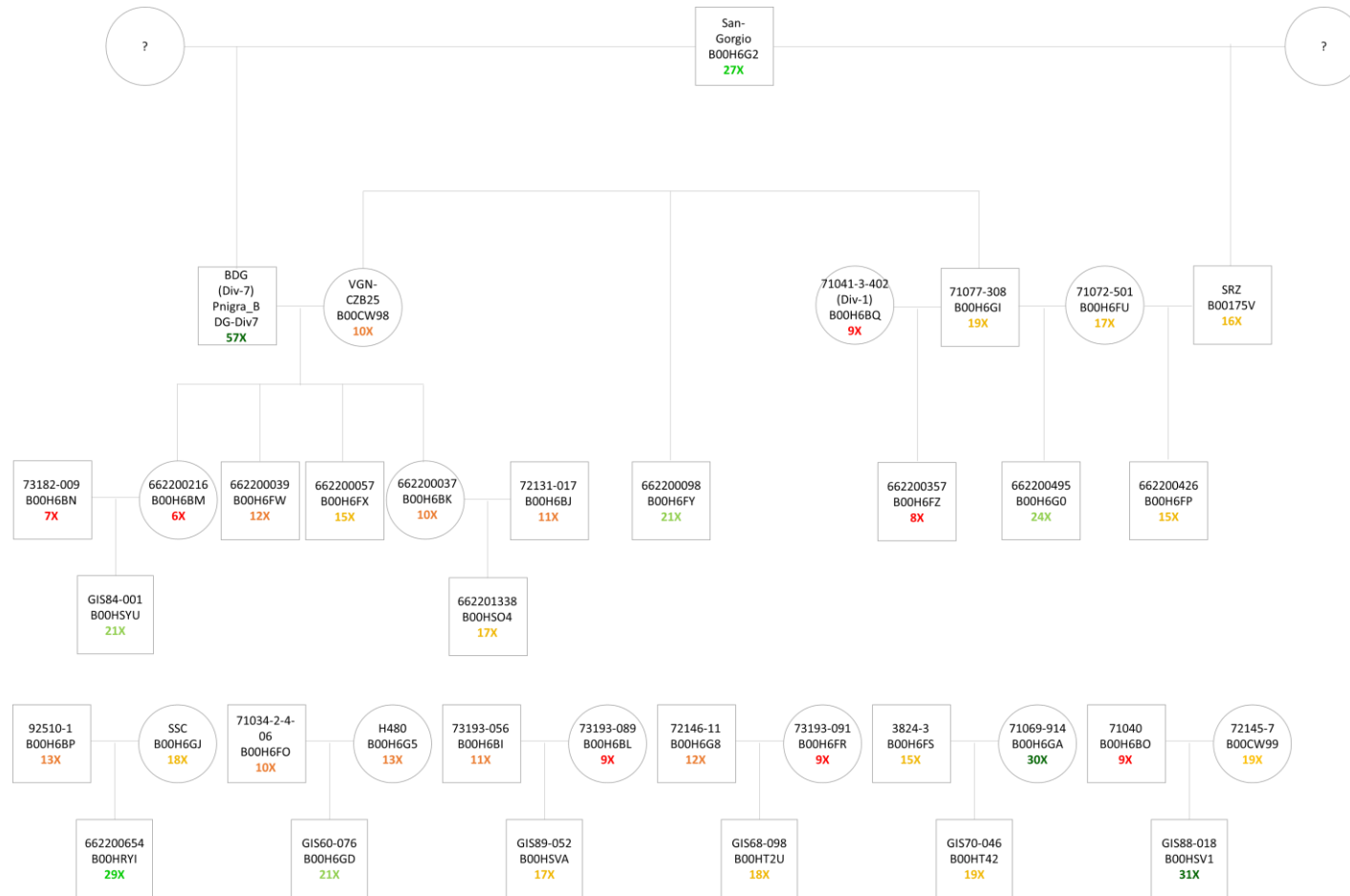
	x 0/0	0/1	1/1
0/0	0/0	0/0, 0/1	0/1
0/1	0/0, 0/1	0/0, 0/1, 1/1	0/1, 1/1
1/1	0/1	0/1, 1/1	1/1



G. Mendel  
(1822-1884)

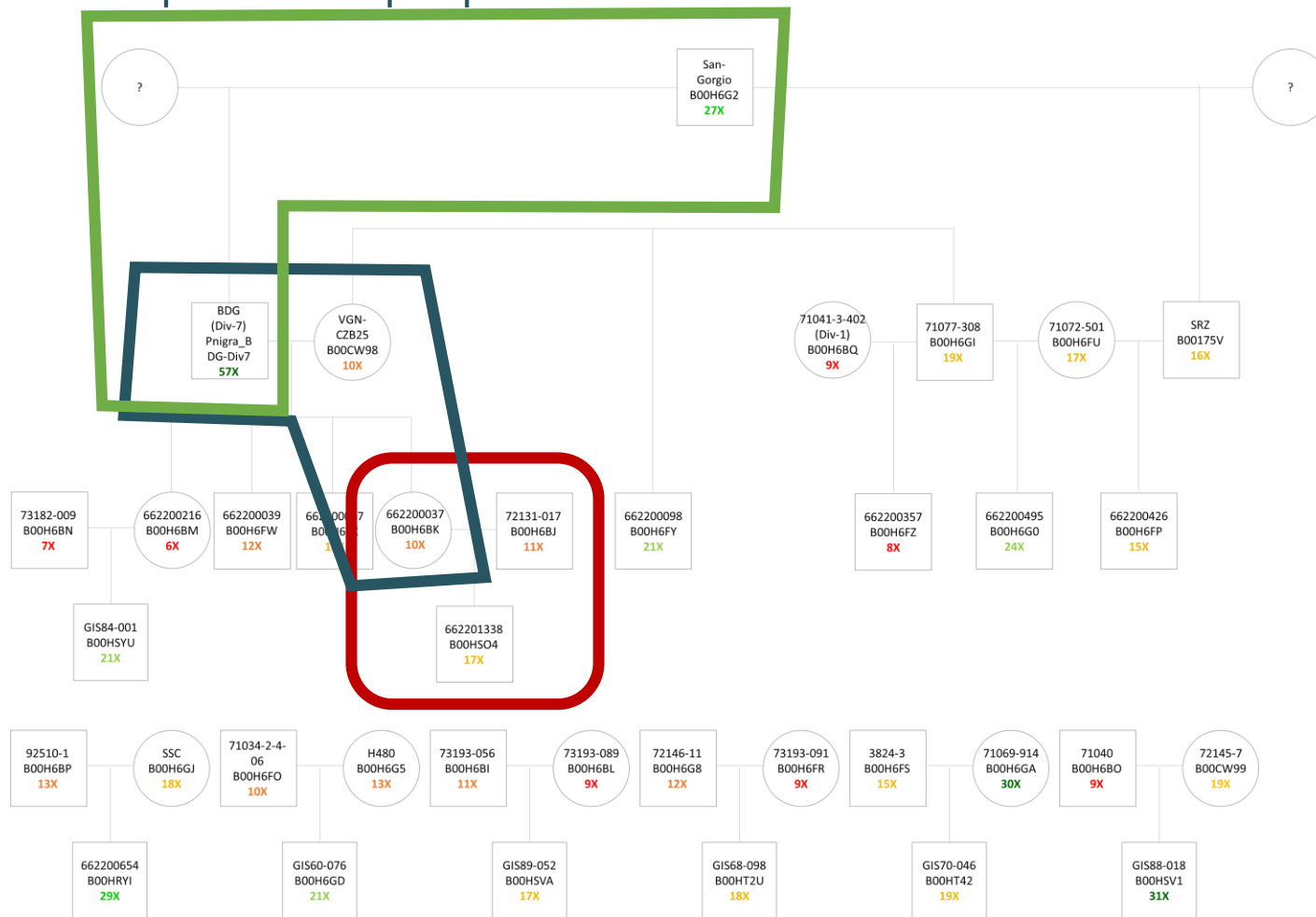
# ➤ Identification de nouveaux marqueurs – les SVs

Exemple chez le peuplier noir



# ➤ Identification de nouveaux marqueurs – les SVs

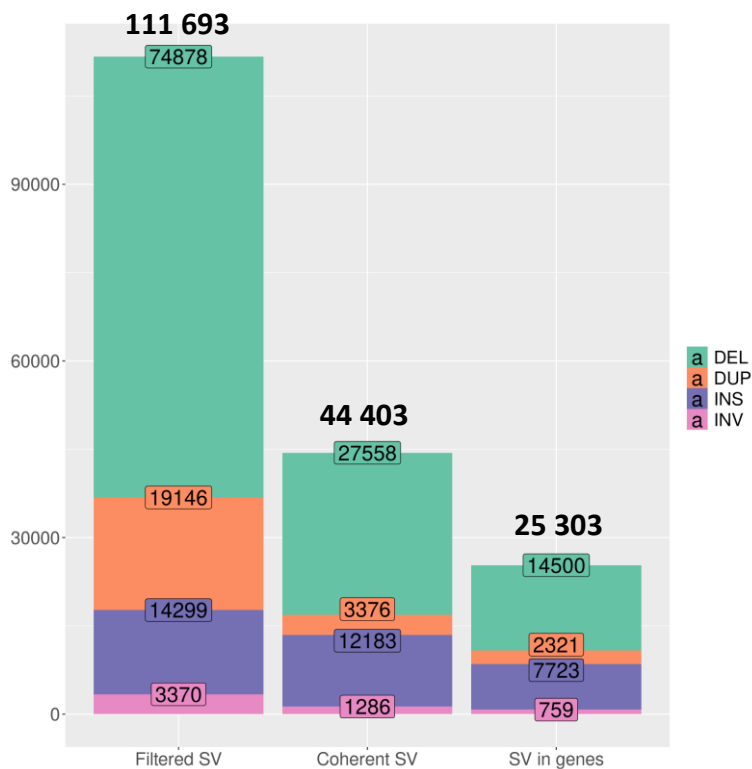
Exemple chez le peuplier noir



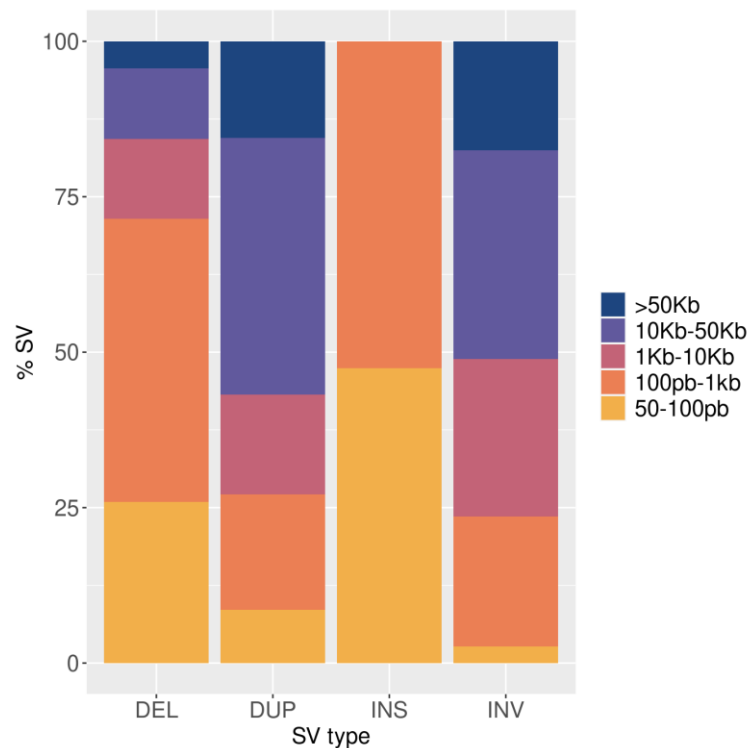


# ➤ Identification de nouveaux marqueurs – les SVs

Exemple chez le peuplier noir



Nombre et types de SVs



Pourcentage des SVs cohérents associés aux gènes en fonction de leur taille et de leur type



# ➤ Identification de nouveaux marqueurs – les SVs

Exemple chez le peuplier noir

- Intérêts :
  - Ségrégation : test très conservatif → - de SNPs, mais + robustes
  - Remet les short reads au goût du jour (SRA !!)  
sans nouveau séq. (\$\$ 😊)
- Limitations :
  - Profondeur (vieilles données ≈ faible prof.)
  - Pas tous SVs détectables (< 100 kb, types)



## Catalogue de SVs intra- et inter- spécifique

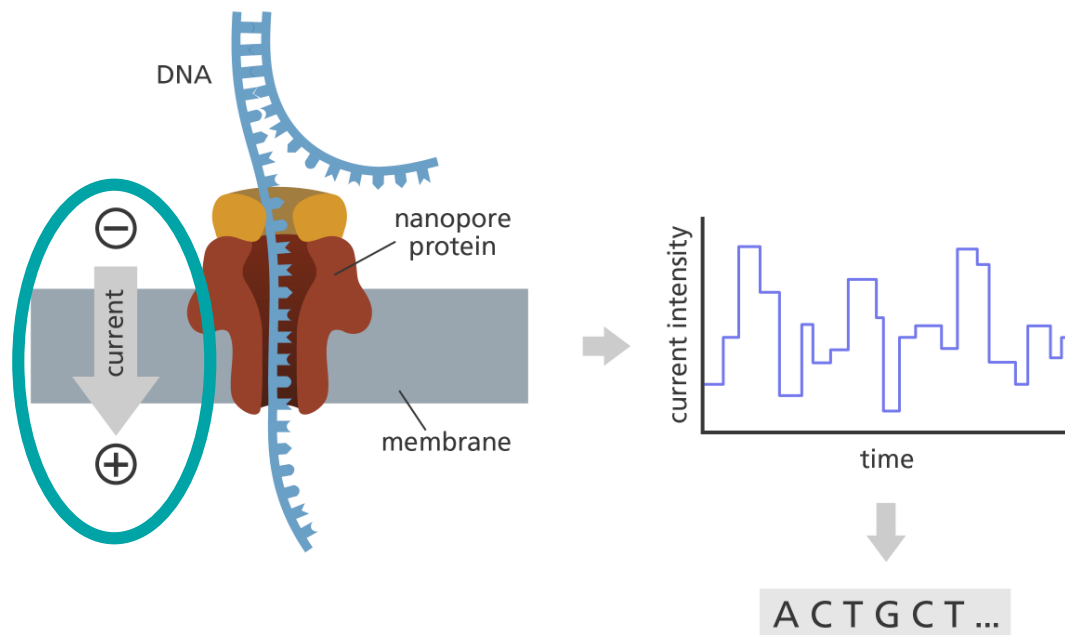
- Voir poster Romane Guilbaud 😊



# ➤ Adaptive Sampling – NAS

Nouveau développement – in progress

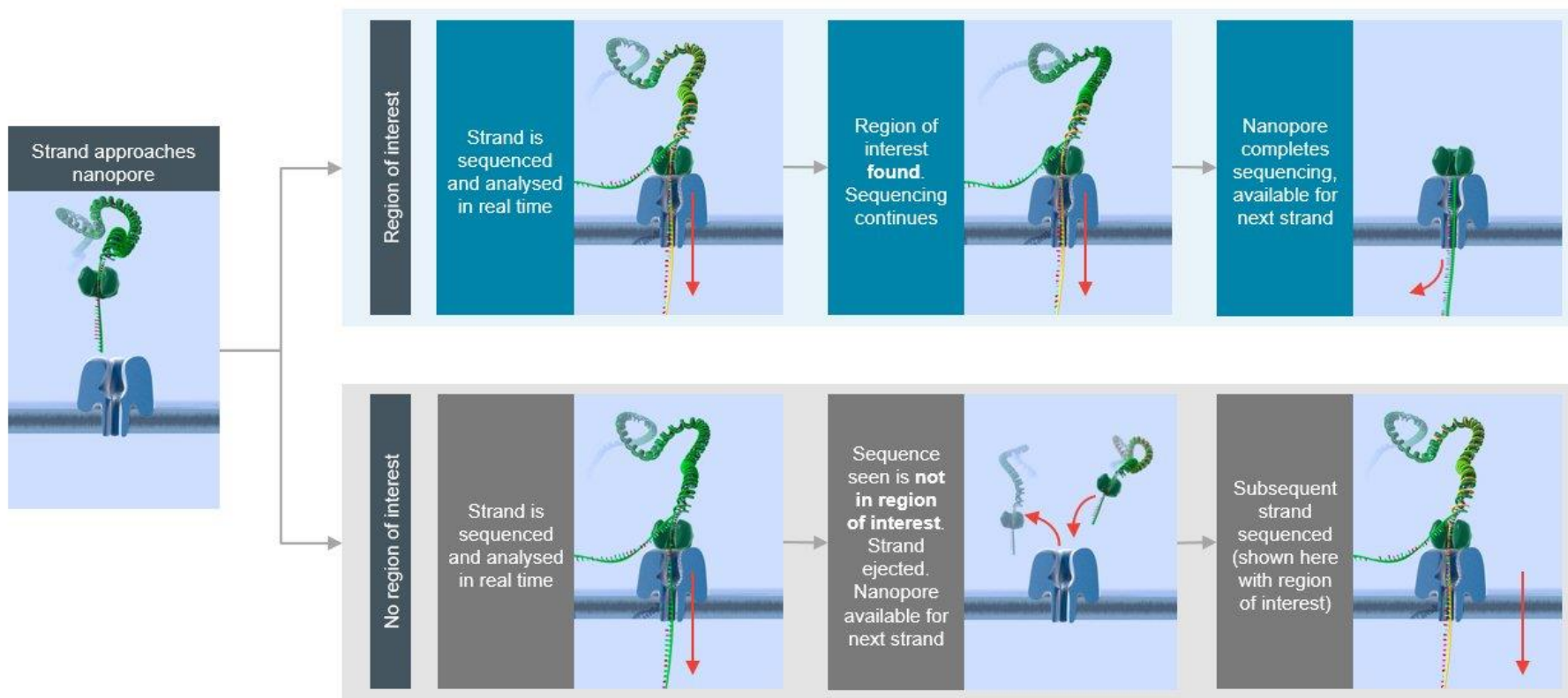
## Caractéristique unique et intrinsèque du séquençage Nanopore ONT



# ➤ Adaptive Sampling – NAS

Nouveau développement – in progress

## Caractéristique unique et intrinsèque du séquençage Nanopore ONT



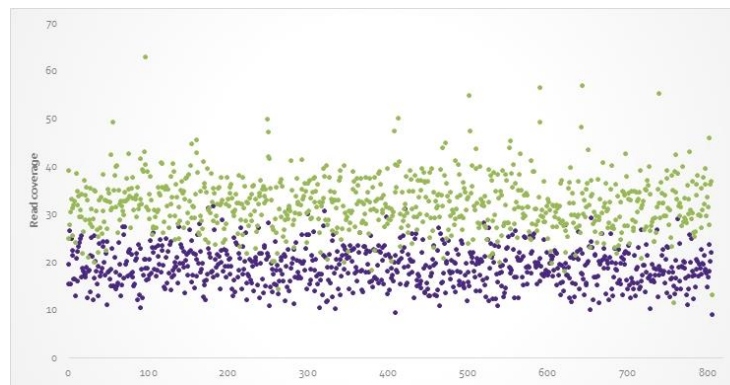
# ➤ Adaptive Sampling

Nouveau développement – in progress

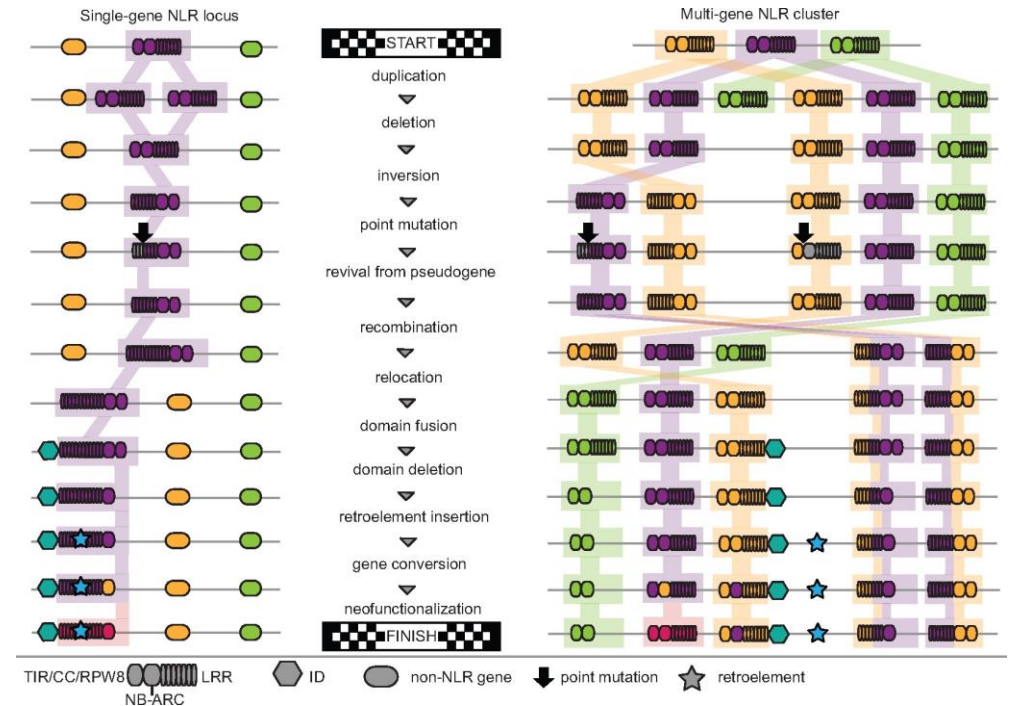
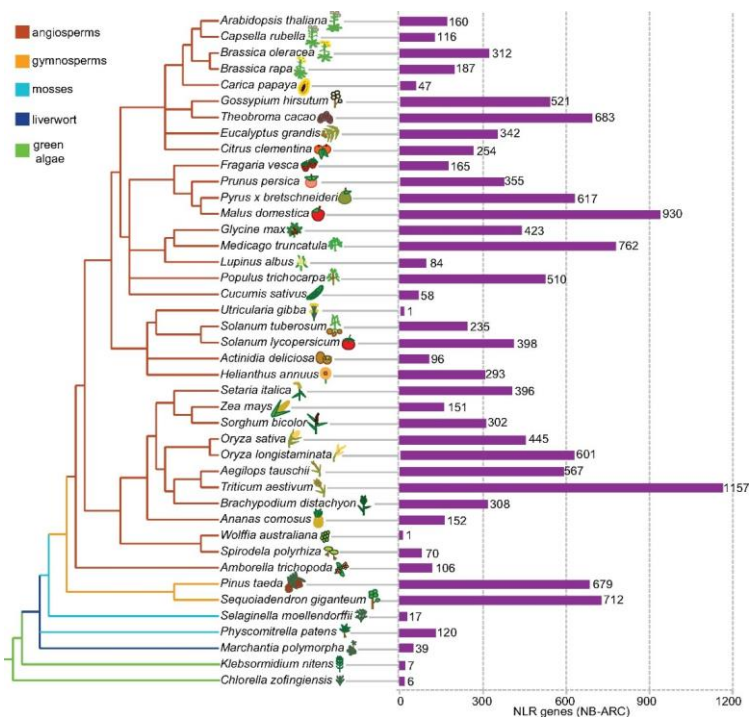
## Caractéristique unique et intrinsèque du séquençage Nanopore ONT

### Enrichissement / déplétion de « régions » cibles

- Enrichissement de taxons dans métagénomés ou avec contaminations (espèces rares, pathogènes dans tissus hôtes, etc)
- Enrichissement de gènes diagnostiques (maladies, traits agronomiques, etc)
- Déplétion de séquences assemblées (« genome finishing », contaminations)



# ➤ NLR-ome : « multiply and resist »



Barragan, A. C., & Weigel, D. (2021). Plant NLR diversity: the known unknowns of Pan-NLRomes. *The Plant Cell*, 33(4), 814-831.

# ➤ Clusters NLR chez *Cucumis melo*

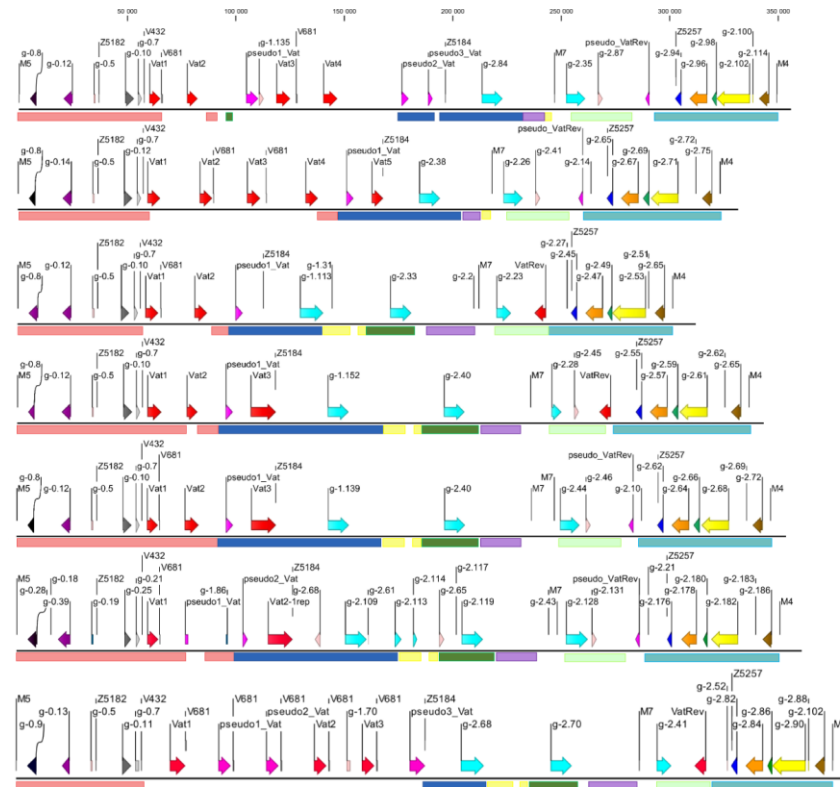


## Génome du melon

- ~450 Mb
- Lignées pures (ou quasiment)

## Clusters NLR

- >10 clusters (~80 gènes)
- VAT (~15 gènes) bien caractérisé dans plusieurs variétés (Illumina WGS)
- Nombre de gènes variable
- Différent gènes avec nombre de domaines variable
- 6 transposons insérés dans pseudo-gènes



Cucumis melo  
Cv DHL92

Cucumis melo  
Cv Anso77

Cucumis melo  
Cv PI 161375

Cucumis melo  
Cv Pyszawat

Cucumis melo  
Cv Doublet

Cucumis melo  
Cv Harukei-3

Cucumis melo  
Cv HS



INRAE

Exploitation de la puissance du séquençage pour l'amélioration des plantes

14 juin 2022 / Mini-congrès Genoscope / Damien Hinsinger

## ➤ Méthode

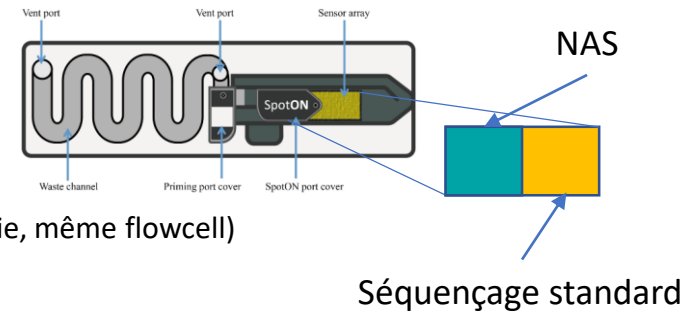
- **Design du panel de séquences à enrichir**

- Extraction des séquences des gènes NLR / clusters
- +10 kb de régions flanquantes en 5' et 3' pour chaque région → référence

➔ Target ~ 1% of the melon genome

- **Test sur demi-flowcell**

- 256 canaux « enrichis » / 256 canaux séquençant *as usual*
- Comparaison *in silico* de l'enrichissement relatif (même librairie, même flowcell)
- run de 90h



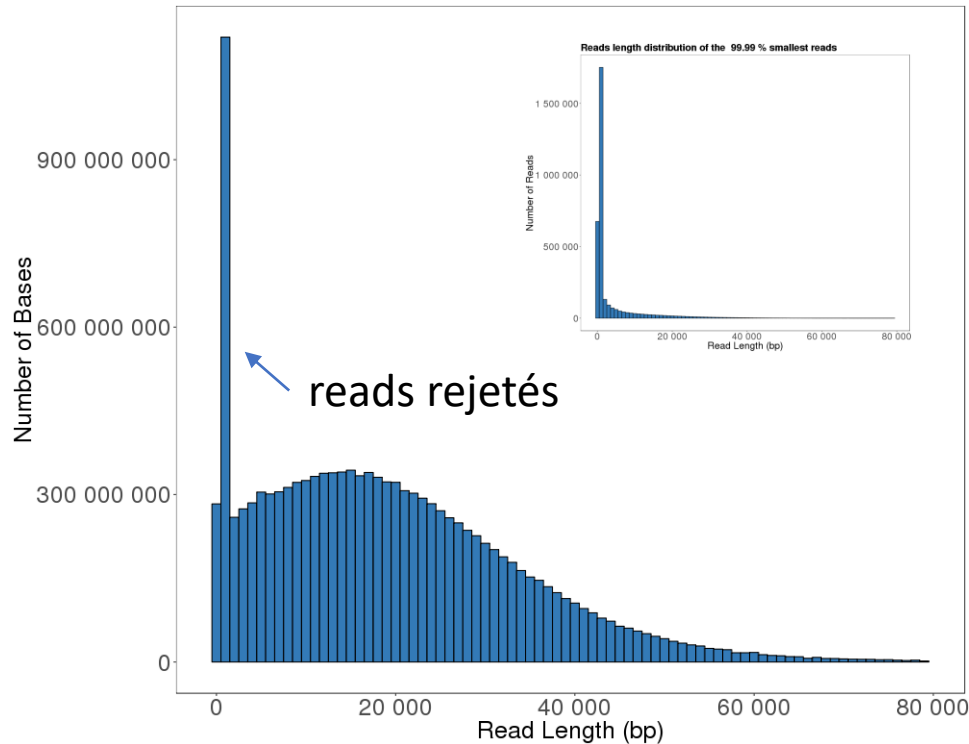


# ➤ Contrôle Qualité - séquençage

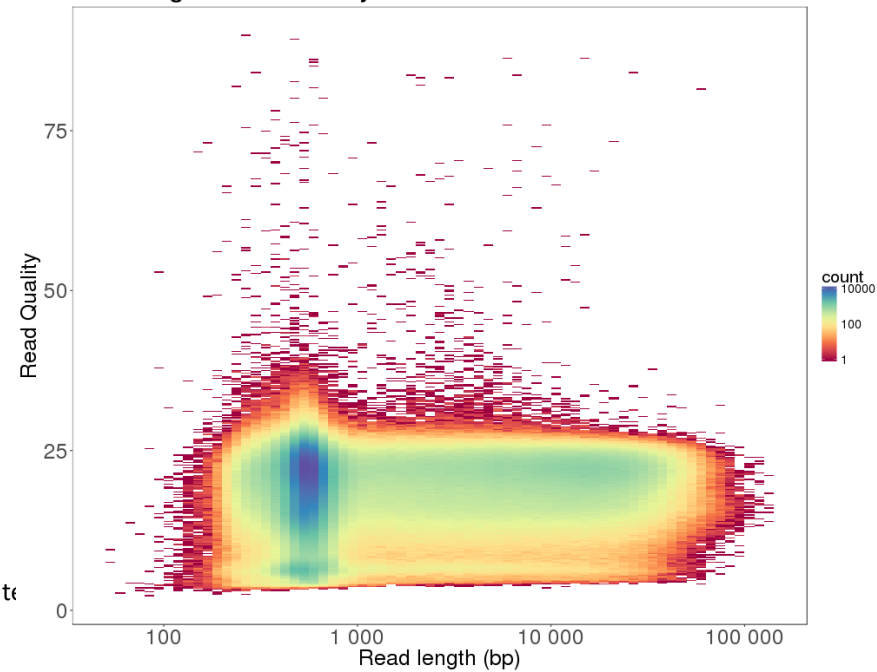
## QC séquençage Nanopore

rawYield	nbReads	avgSize	maxSize	N50	nbSup10000	nbSup5000	avgQ	%GC
12 640 451 249	3 381 069	3 739	136 830	17 098	411 272	626 344	19.27	36,90

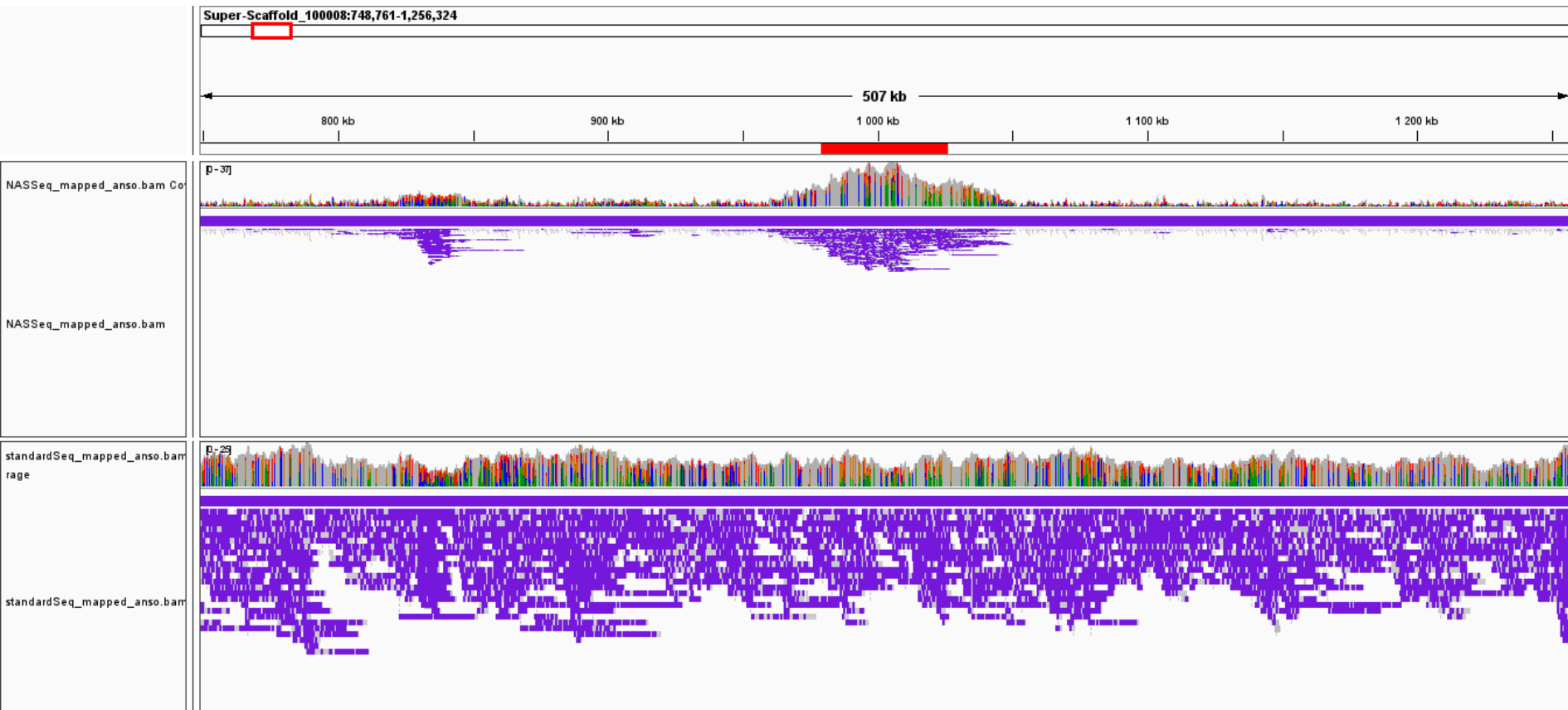
Cumulative bases distribution of the 99.99 % smallest reads



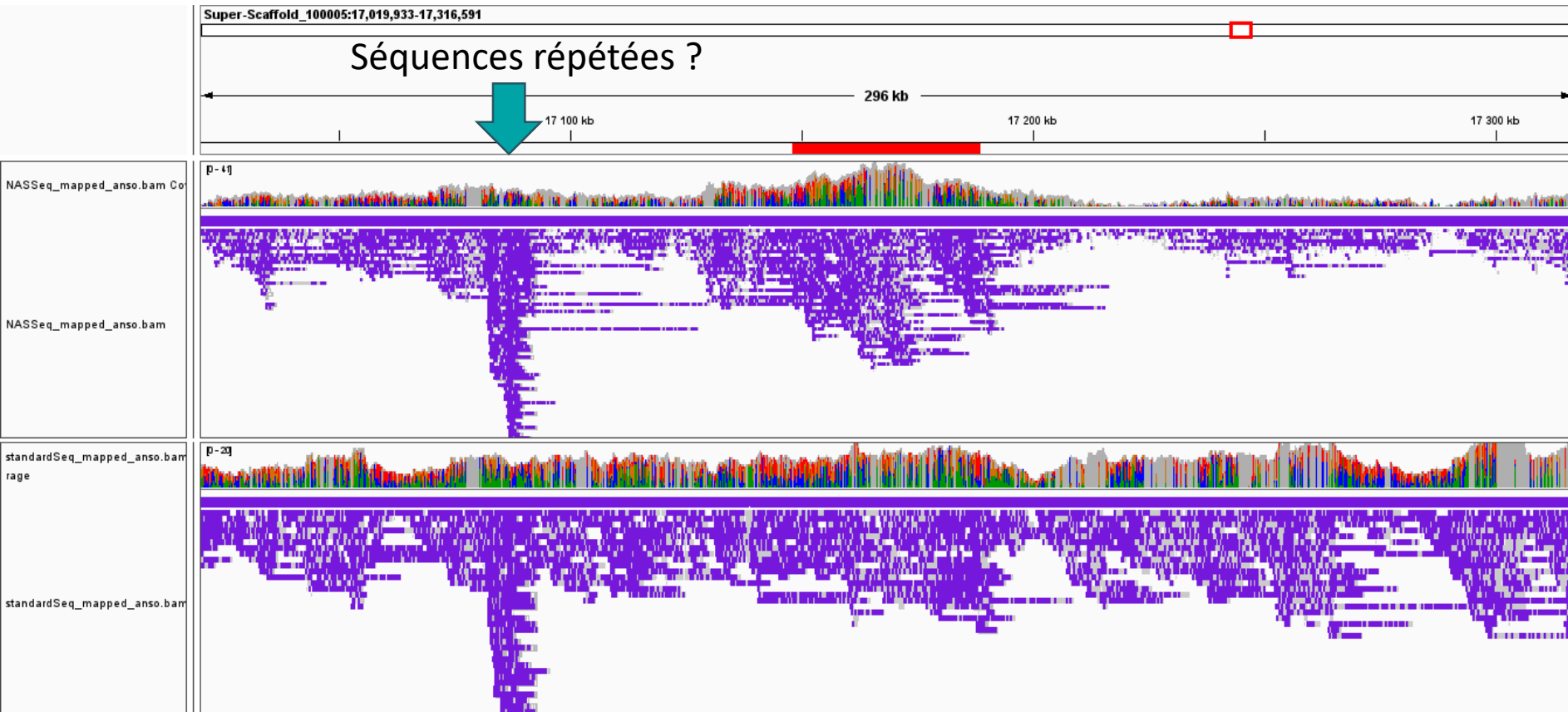
Read Length vs Read Quality



# ➤ Comparaison visuelle (mapping)



# ➤ Comparaison visuelle (mapping)

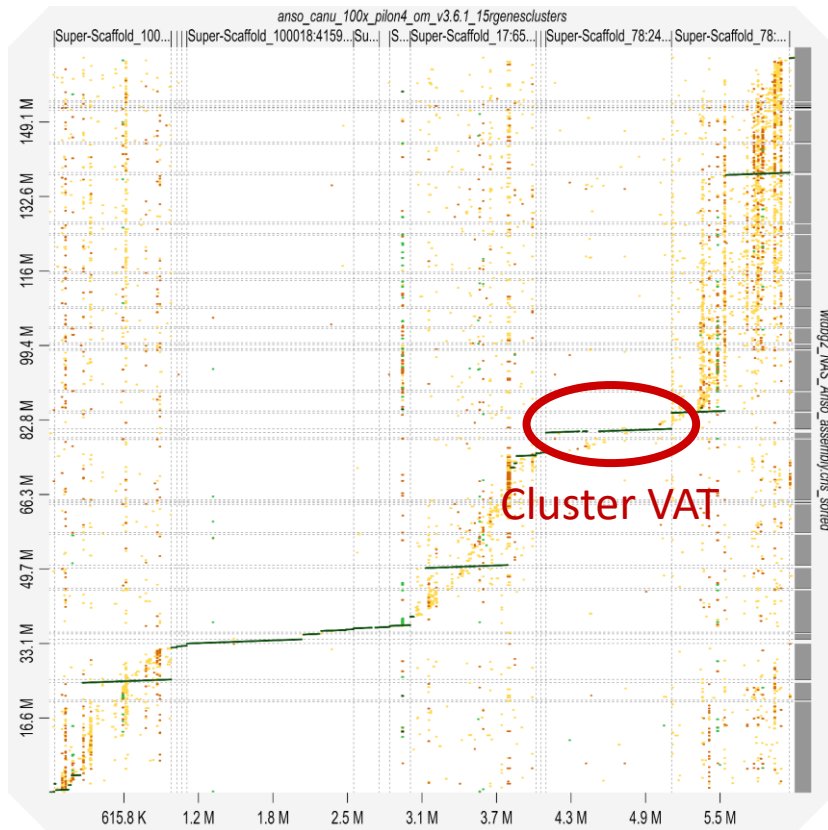


## ➤ Comparaison par régions (mapping)

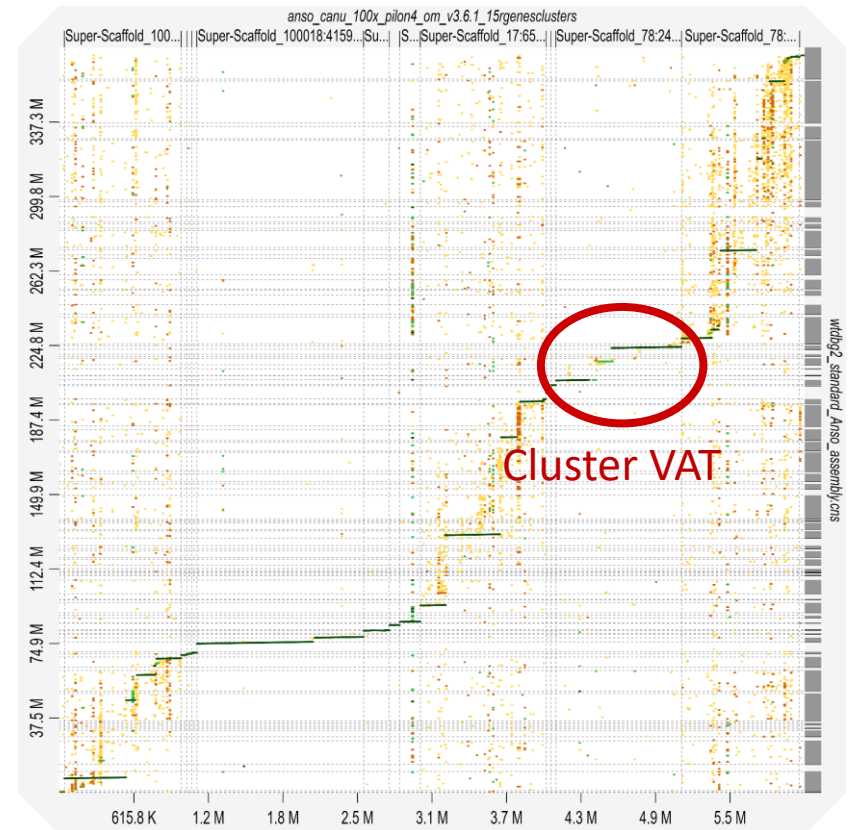
Chromosome	Chrom cov NAS	Region cov NAS	Chrom cov standard	Region cov standard	Enrichment NAS	Enrichment standard	enrichment
Super-Scaffold_6	6,9	31,22	14,58	16,6	4,52	1,14	3,97
Super-Scaffold_7	5,44	24,75	18,15	12,59	4,55	0,69	6,56
Super-Scaffold_17	9,02	35,16	16,37	15,56	3,90	0,95	4,10
Super-Scaffold_17	9,02	30,23	16,37	12,47	3,35	0,76	4,40
Super-Scaffold_78	10,13	34,86	15,17	15,21	3,44	1,00	3,43
Super-Scaffold_78	10,13	30,31	15,17	13,7	2,99	0,90	3,31 Cluster VAT
Super-Scaffold_78	10,13	19,15	15,17	11,48	1,89	0,76	2,50
Super-Scaffold_100005	7,17	26,29	15,33	14,11	3,67	0,92	3,98
Super-Scaffold_100006	8,15	30,79	14,25	13,4	3,78	0,94	4,02
Super-Scaffold_100008	6,34	24,09	15,58	13,38	3,80	0,86	4,42
Super-Scaffold_100008	6,34	26,34	15,58	14,46	4,15	0,93	4,48
Super-Scaffold_100010	9,26	24,02	20,84	13,26	2,59	0,64	4,08
Super-Scaffold_100018	9,06	34,58	18,95	15,87	3,82	0,84	4,56
Super-Scaffold_100019	6,23	34,96	17,57	13,38	5,61	0,76	7,37
Super-Scaffold_100020	4,9	23,83	15,95	13,52	4,86	0,85	5,74



# ➤ Comparaison de l'assemblage des régions cibles NAS vs standard



Nanopore Adaptive Sampling



Séquençage standard

# > Perspectives

## Génotypage SNPs

- « carrosser » le pipeline d'analyse pour prod. -> stage M1 Benjamin Loire
- Renforcer notre expertise -> IB21 (comparaison technos sur plusieurs espèces)
- Poursuivre les analyses sur SVs / reconstruction d'haplotypes (projet Amaizing)

## Génotypage SVs

- cf. poster Romane Guilbaud
- Comparer avec approches long reads (avec et sans trios)
- Valider certains SVs « prometteurs »

## Nanopore Adaptive Sampling

- Améliorer la référence (Vendelbo *et al.* 2022, GAFL, séquences répétées, etc)
- Test du kit 14 (output + Q20+)
- Tester autres usages potentiels (genome finishing, grandes régions / bras de chromosomes, etc) / sur d'autres espèces

## ➤ Remerciements

**EPGV**  
Etude du Polymorphisme  
des Génomes Végétaux

INRAE  
**GENOMICS**



**RésiLens**



INRAE **EPGV**  
Etude du Polymorphisme  
des Génomes Végétaux

Exploitation de la puissance du séquençage pour l'amélioration des plantes  
14 juin 2022 / Mini-congrès Genoscope / Damien Hinsinger