



HAL
open science

Phasing haplotypes in rabbit using long reads technology

Julie Demars

► **To cite this version:**

Julie Demars. Phasing haplotypes in rabbit using long reads technology. Journées Plateforme Get-Plage, Nov 2017, Toulouse, France. hal-02784960

HAL Id: hal-02784960

<https://hal.inrae.fr/hal-02784960>

Submitted on 4 Jun 2020

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.



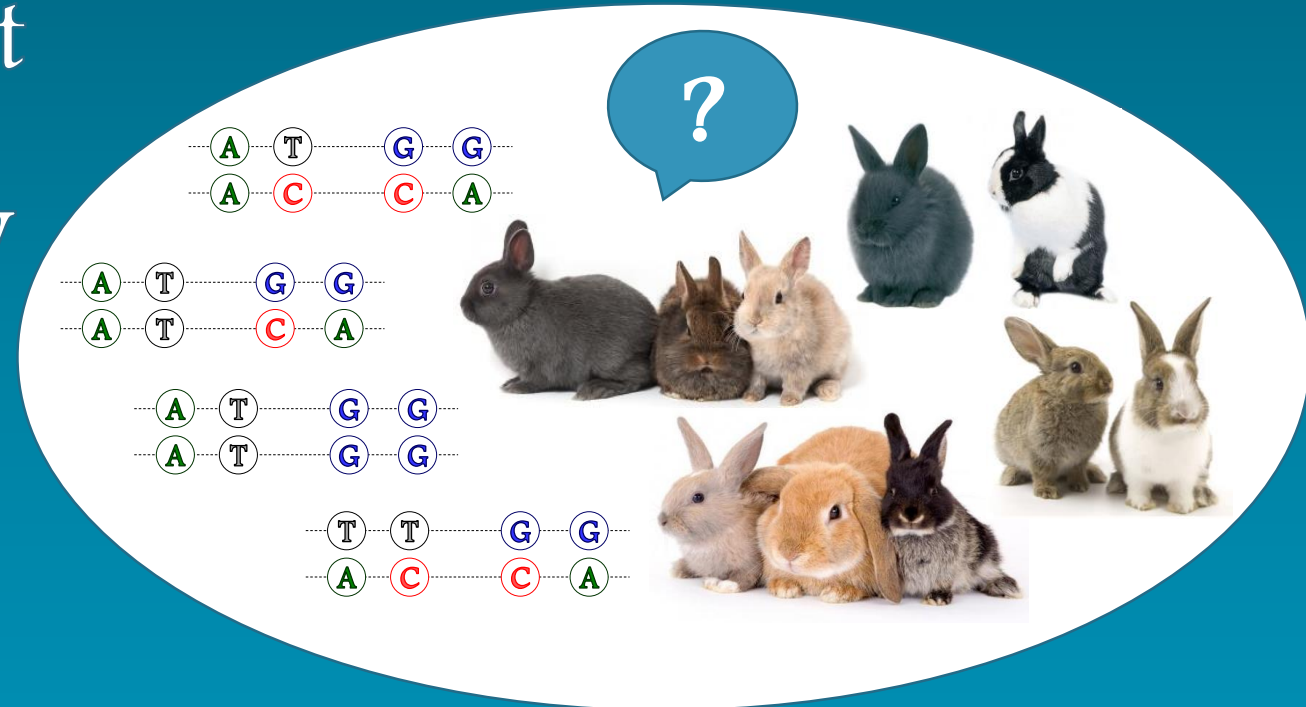
Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

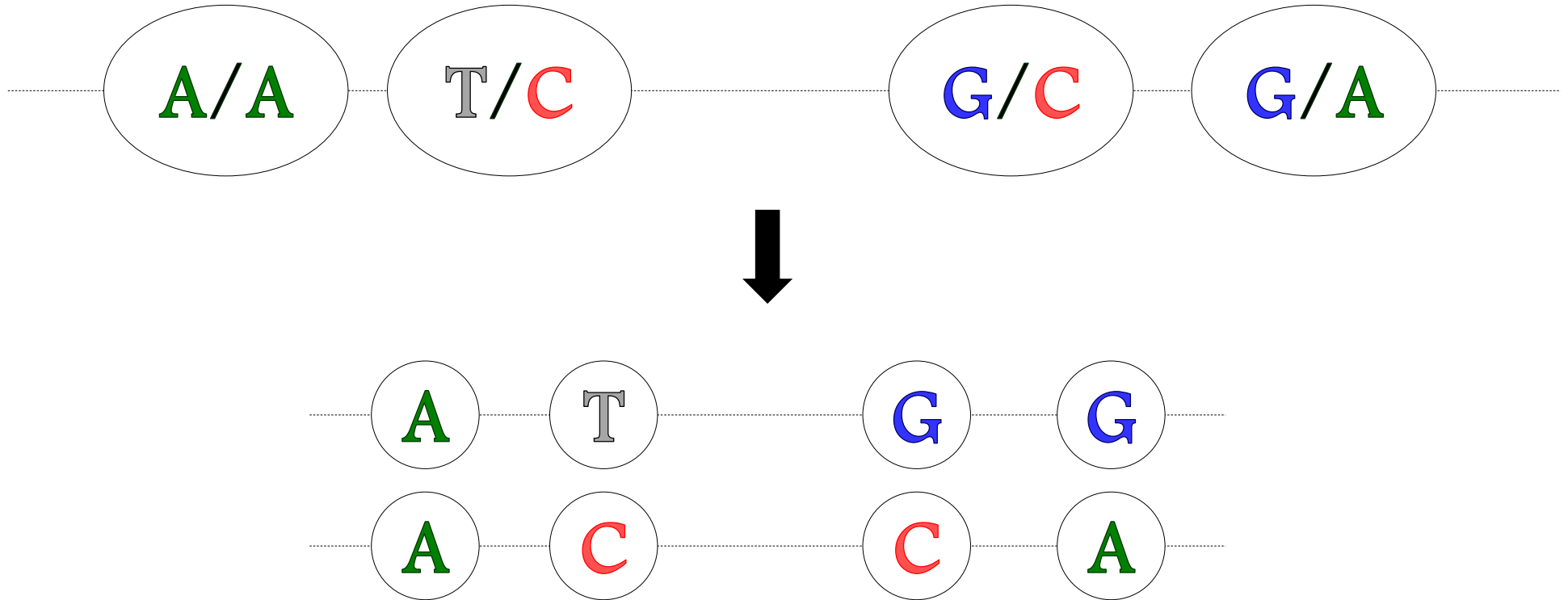
Phasing haplotypes in rabbit using long reads technology

Julie Demars

GenPhySE~Genetics, Physiology and Livestock Systems

Long reads : Dream or Reality ?
28th of November 2017





Fundamental aspect of genetics that is relevant in many applied problems

- **Allele-specific expression or methylation**

Genomic imprinting

- **Highly heterozygous regions**

HLA genes

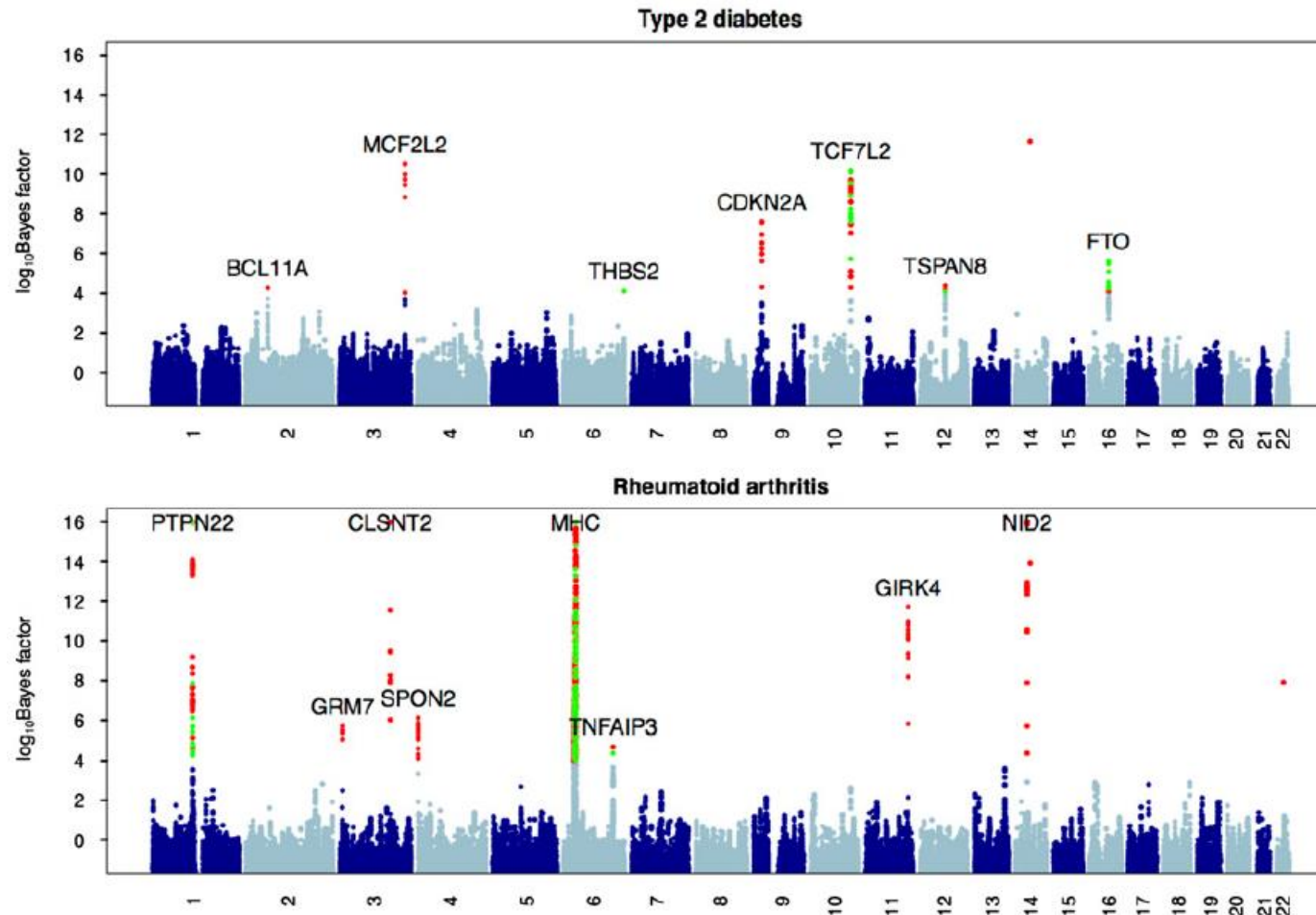
- **Population genetics and genome-wide association studies**

Polygenic traits, Allelic heterogeneity

Genetics. 2014 Jul;197(3):823-38. doi: 10.1534/genetics.114.164814. Epub 2014 May 8.

Detecting local haplotype sharing and haplotype association.

Xu H¹, Guan Y².

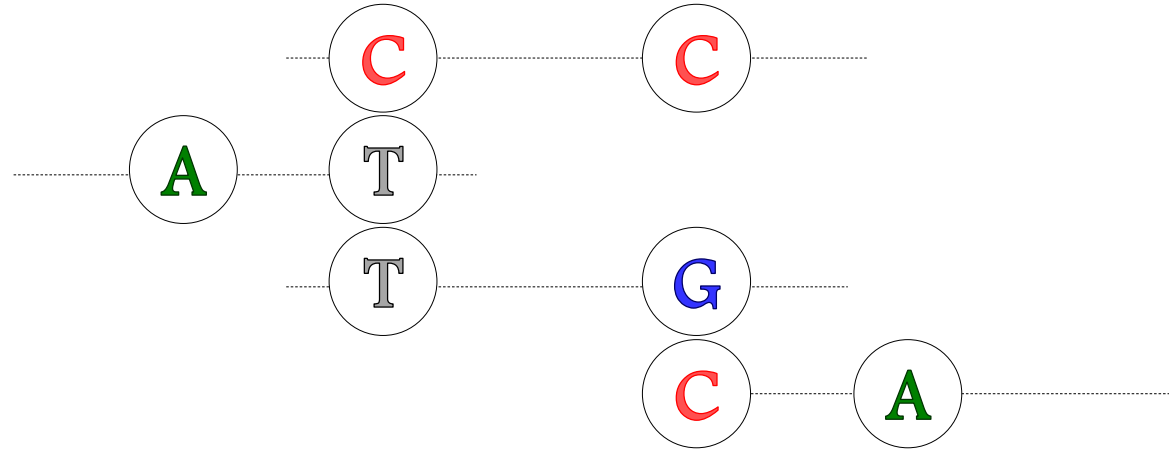


General principle of genome phasing

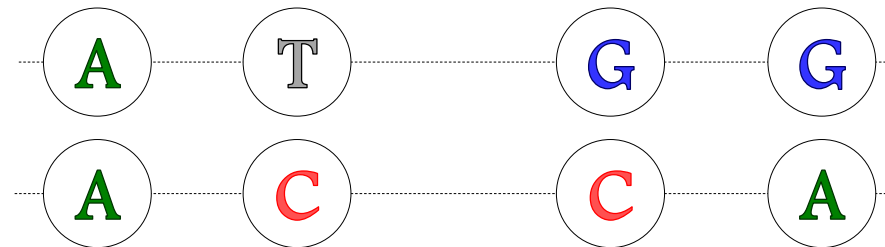
Unphased genome



Sequence reads



Phased results



- **Scale of datasets**

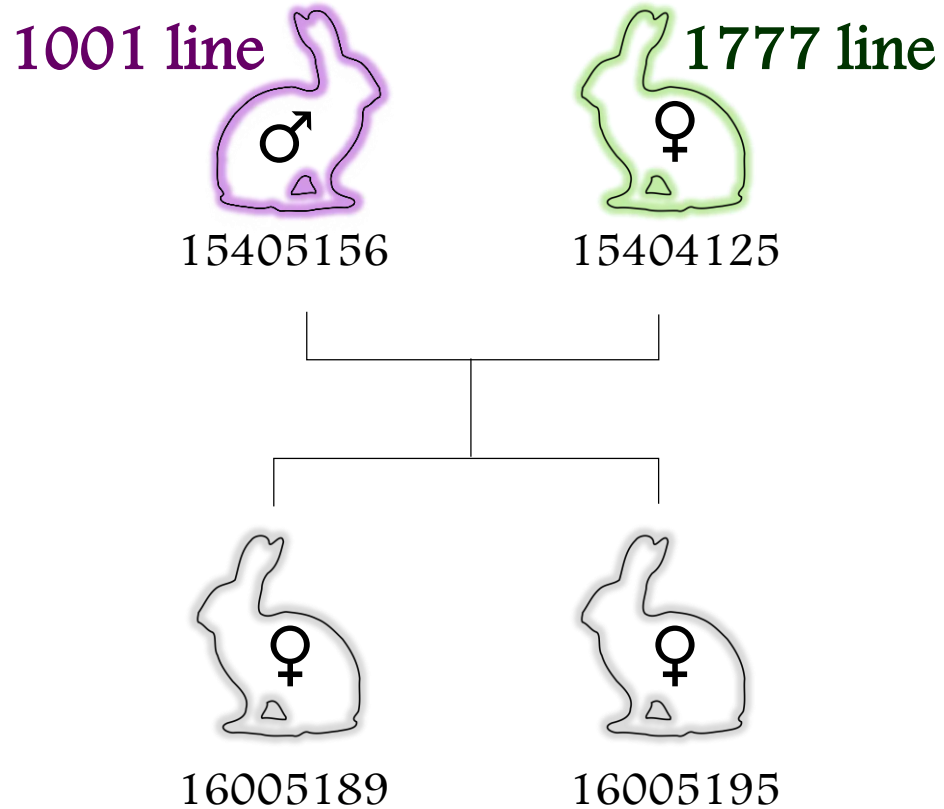
From SNP beadchips to whole genome

- **Statistical models and computational approaches**

Time consuming : balance quality of phasing/computational cost

- **Laboratory-based experimental methods**

Long reads produce virtual multi-kilobases reads on regular sequencers



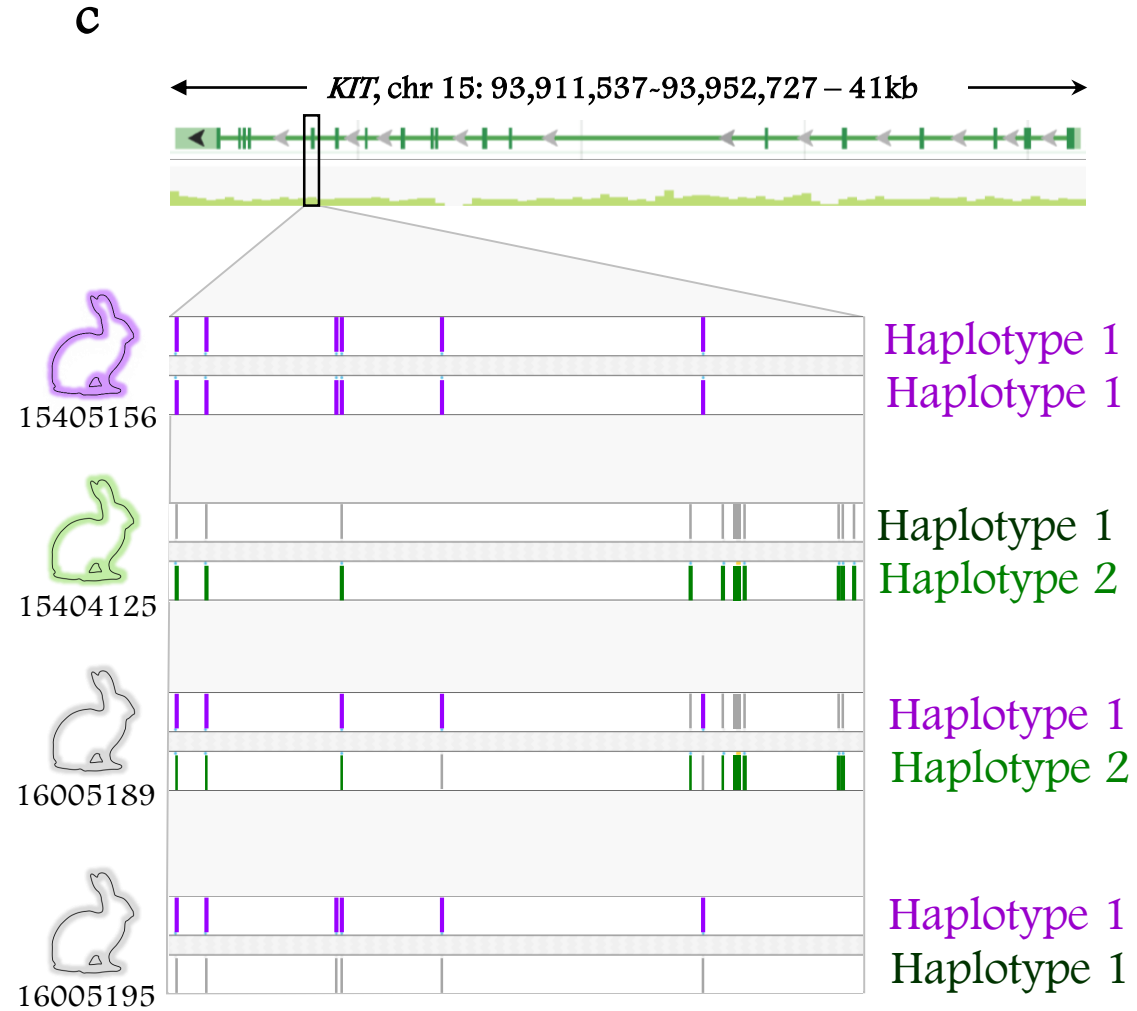
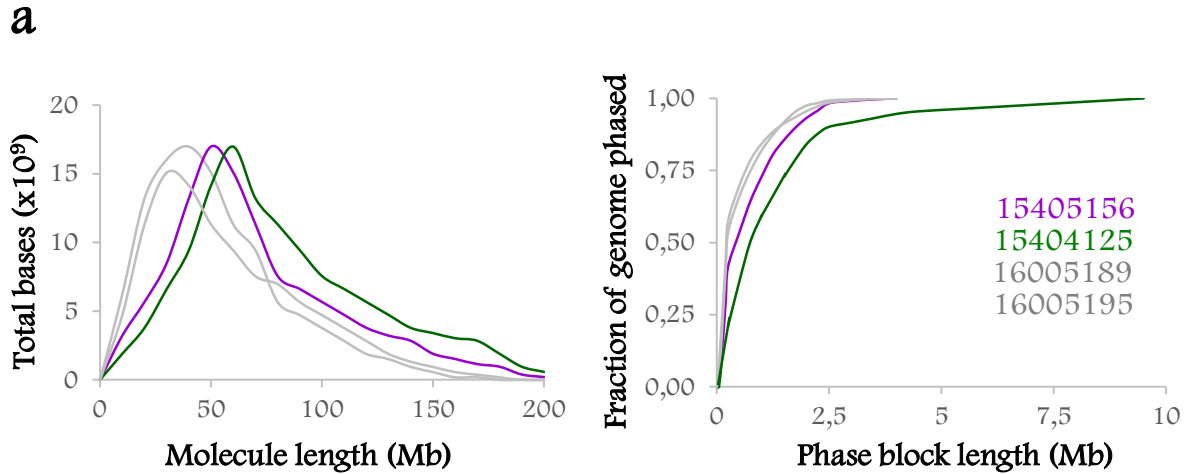
Rabbit Reference Genome (*Oryctolagus cuniculus*)

Version	OryCun2 (GCA_000003625.1), 1st assembly
Reference rabbit	New Zealand
Sequencing Depth	7X
Chromosomes	21
Size (Mb)	2247,75
Genes	20999
Scaffolds	3318
Size (Mb)	489,69
Genes	8099

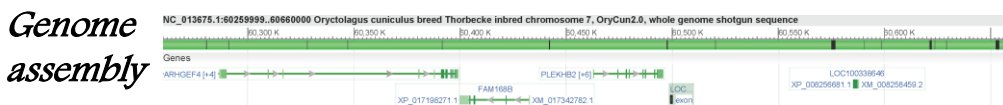
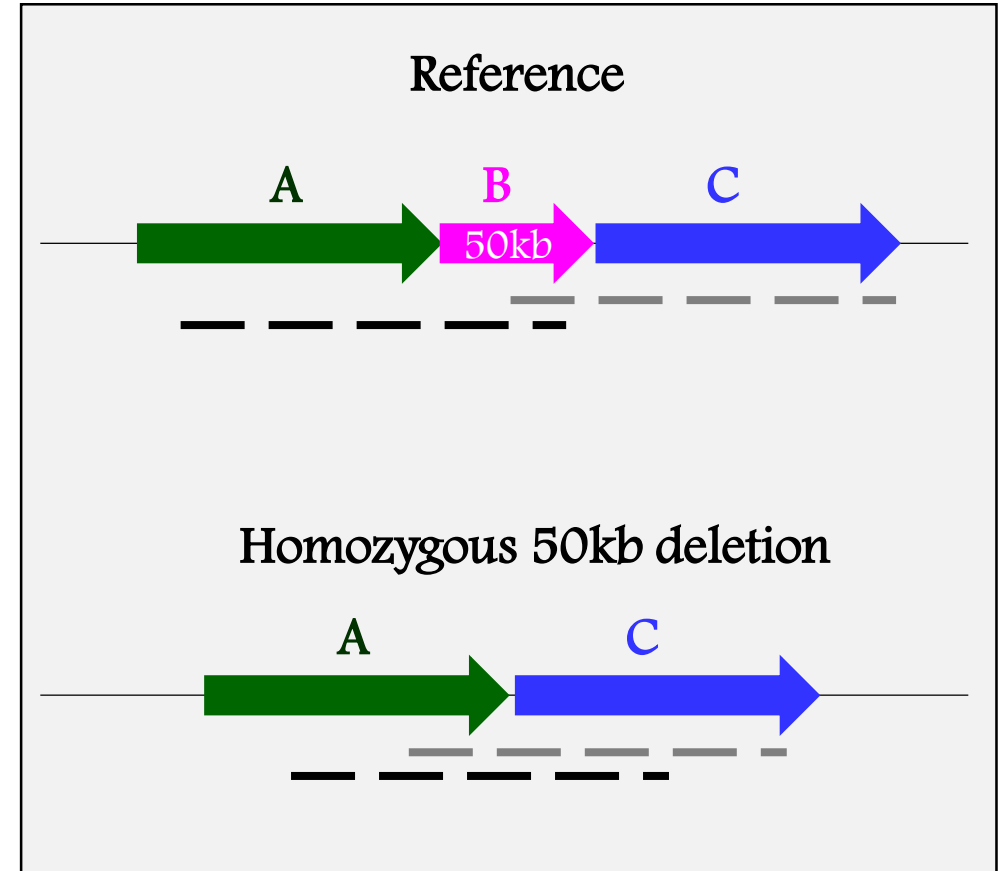
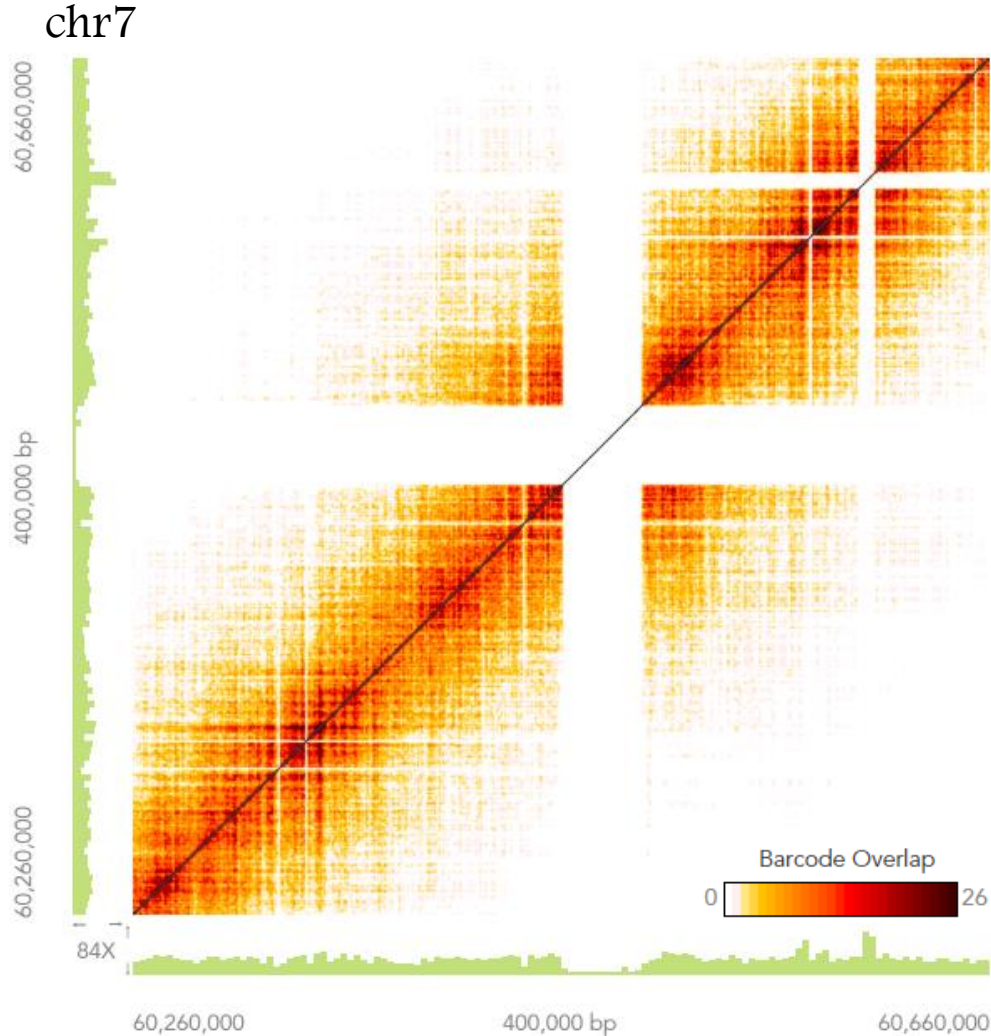
Summary of phasing results

Phased WGS	15405156 Male Father	15404125 Female Mother	16005189 Female Offspring	16005195 Female Offspring
Fold coverage	32	28	26	28
% aligned	81	80	80	77
% duplication	4,42	6,09	6,30	6,66
Relative genomic equivalents per partition	0,362	0,321	0,342	0,508
Number of molecules (millions)	1,59	1,59	1,40	1,34
Length-weighted mean molecule length (kb)	63,17(±41,95)	73,90(±48,02)	45,73(±30,30)	54,06(±43,98)
% SNPs phased	97	97	98	97
N50 phase block (kb)	875,49	1086,73	609,92	838,05
Longest phase block (kb)	4 417,55	9 930,61	3 725,93	4 599,55
Long structural variants (number)	162	183	197	155
Short structural variants (number)	49 846	45 901	46 265	47 723

Phase the « full » spectrum of called variants

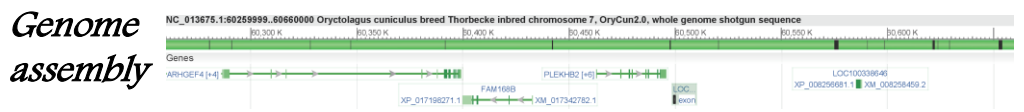
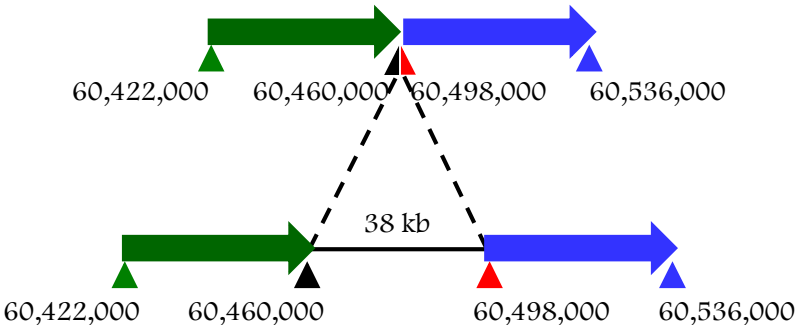
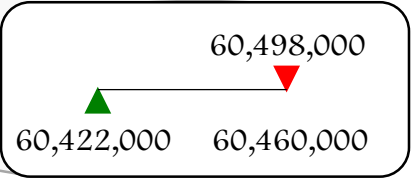
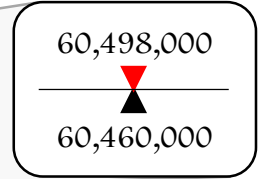
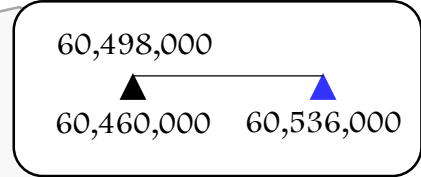
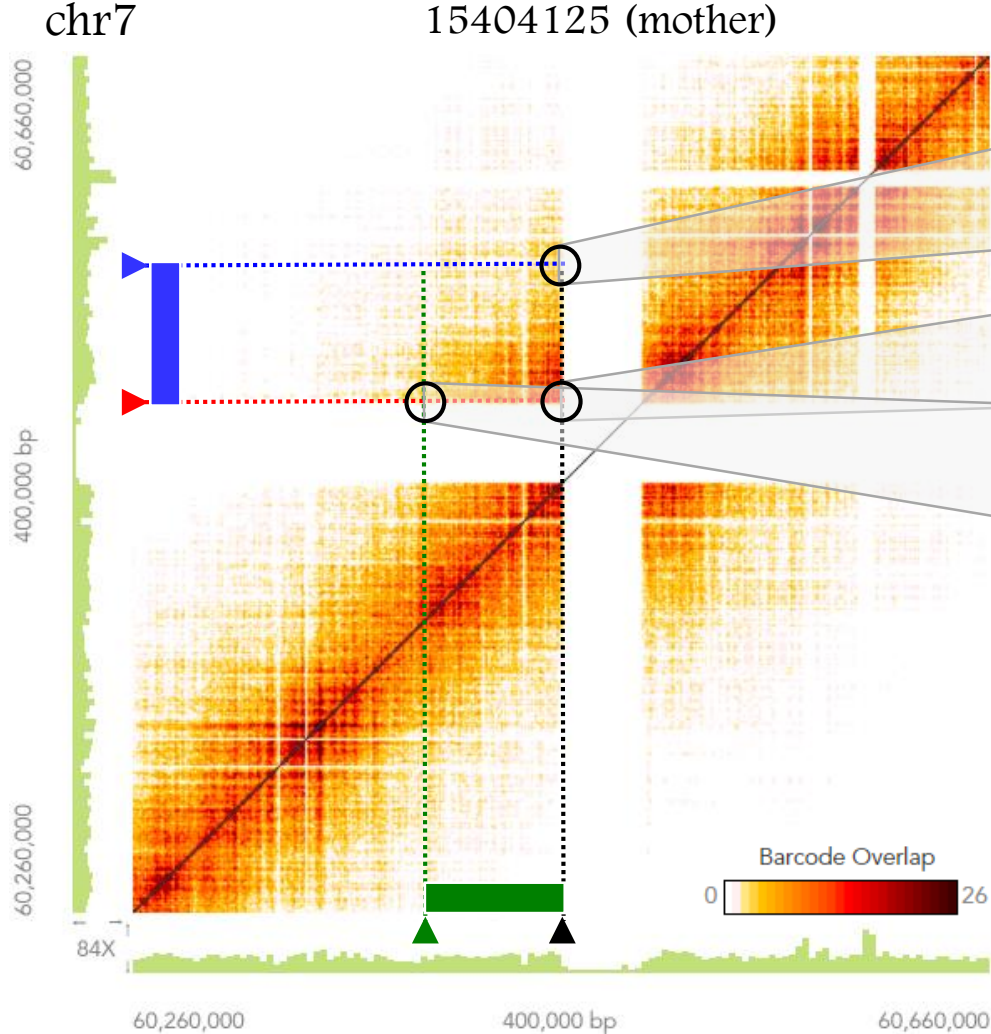


Homozygous deletion



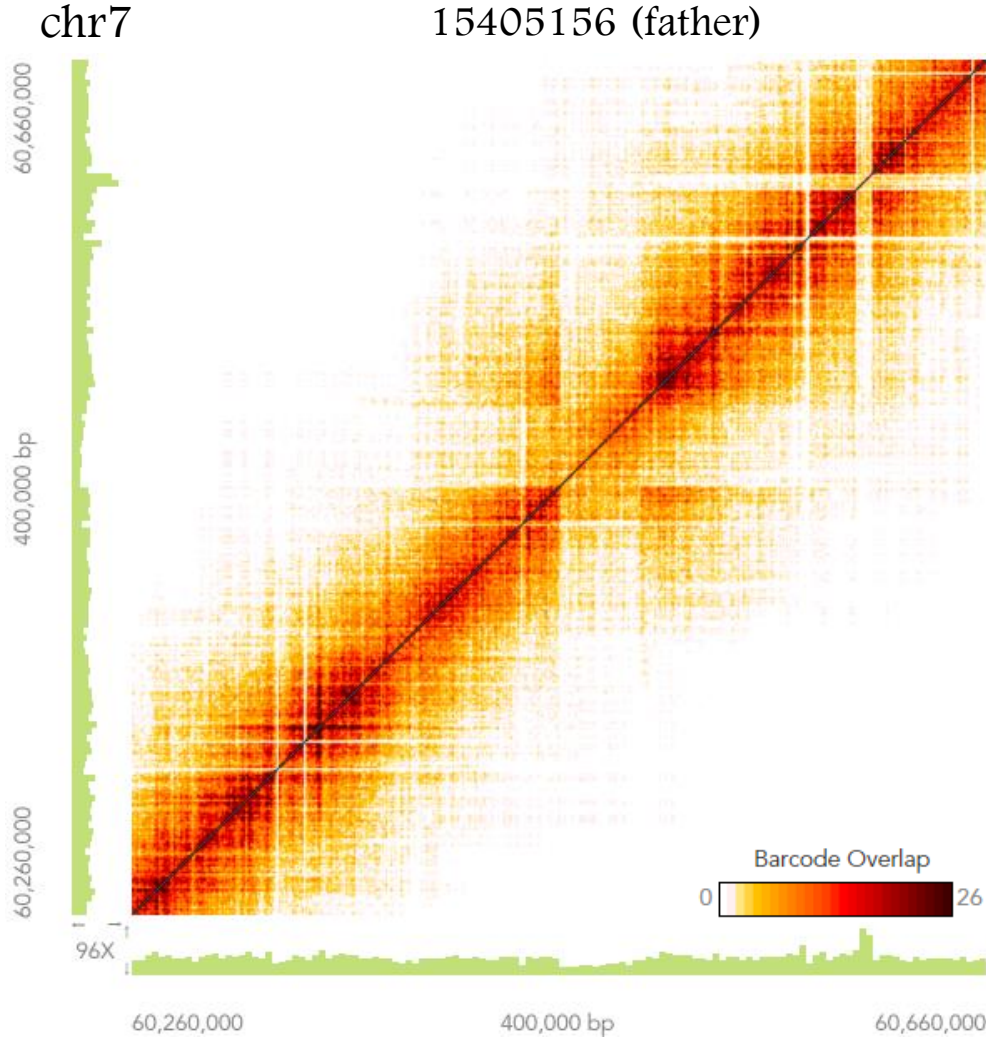
Homozygous deletion

15404125 (mother)

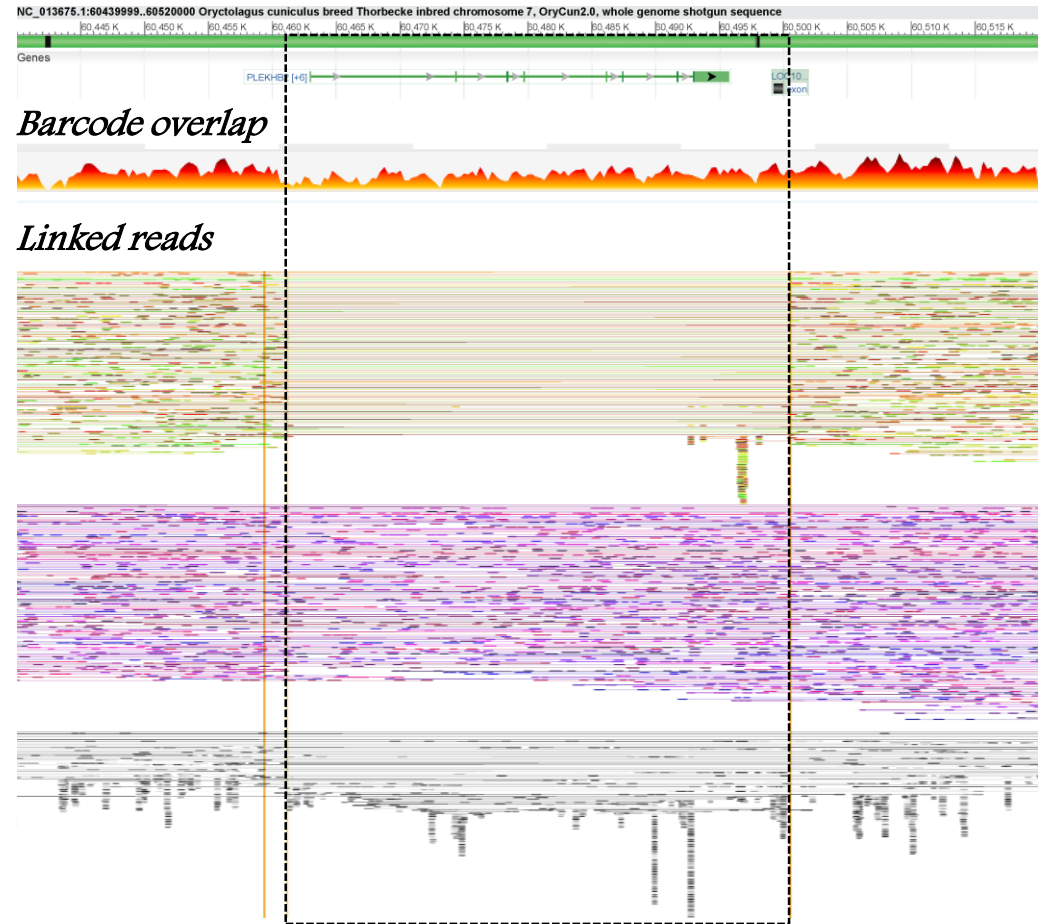


Heterozygous deletion

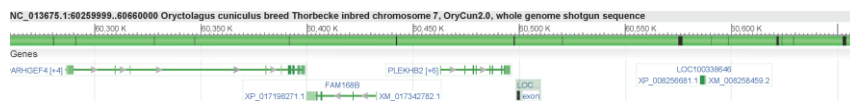
15405156 (father)



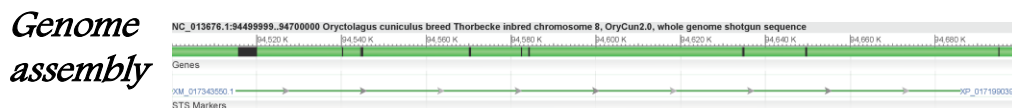
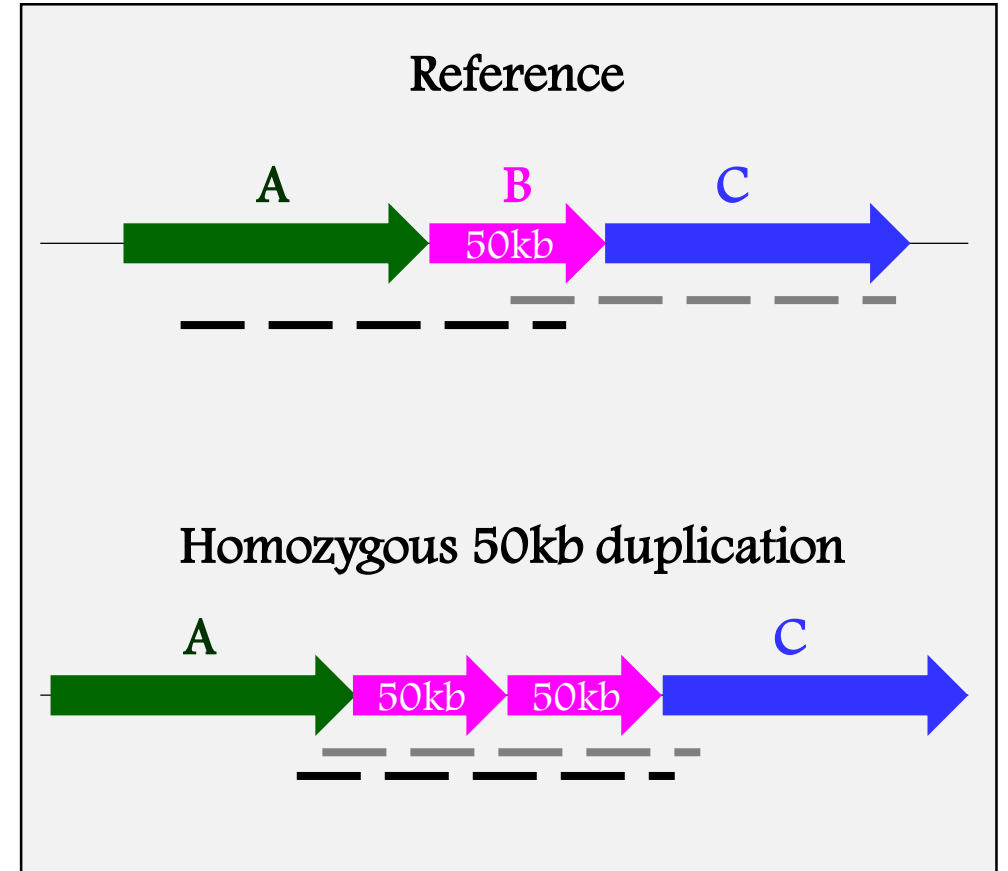
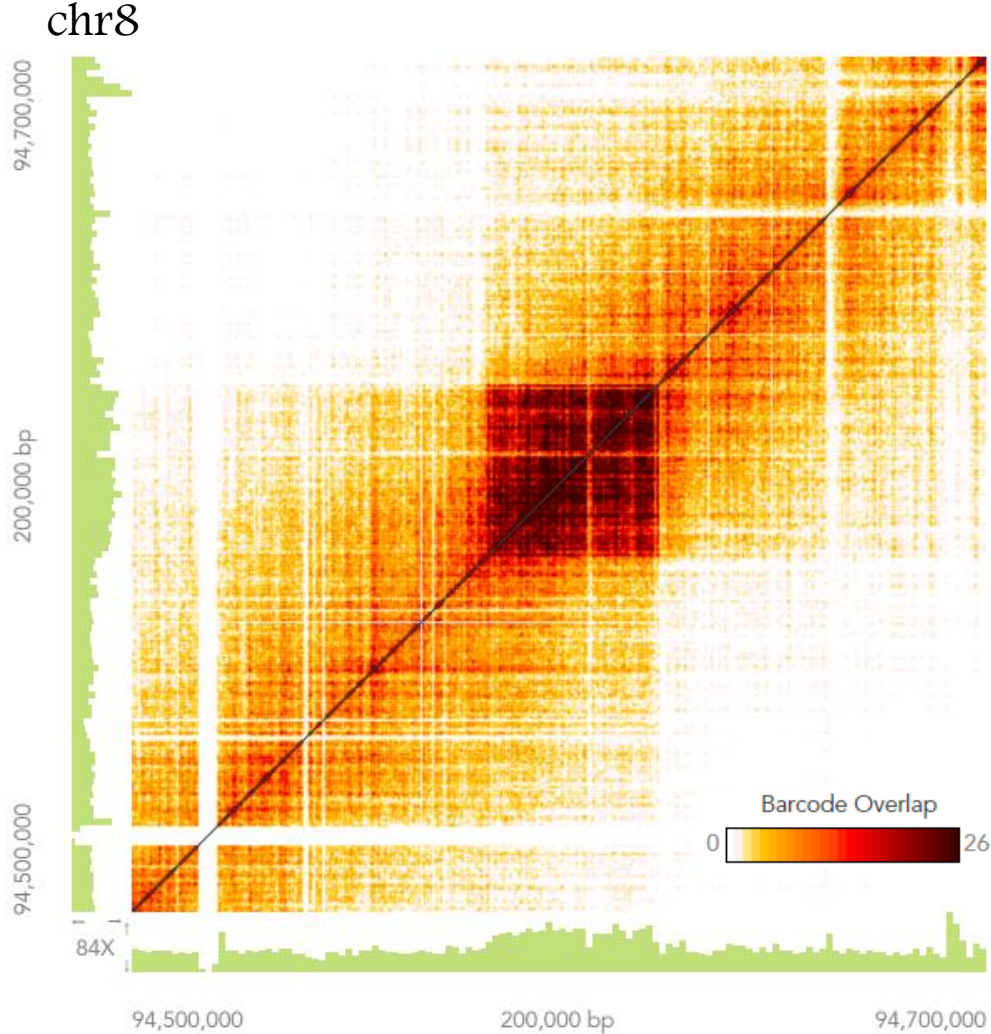
Genome assembly



Genome
assembly



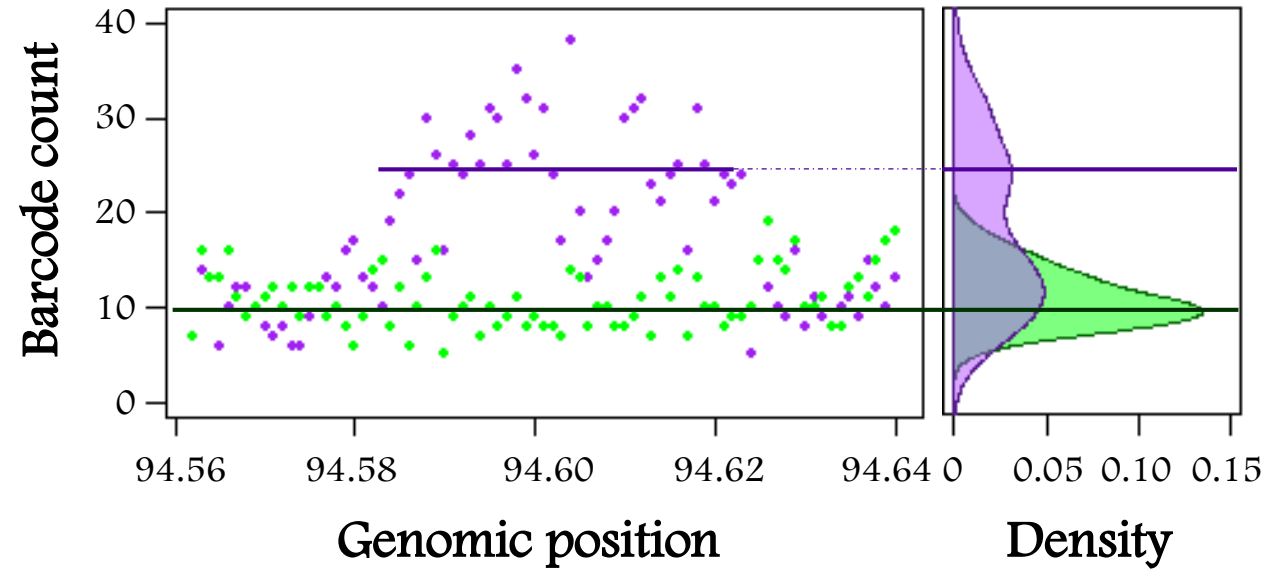
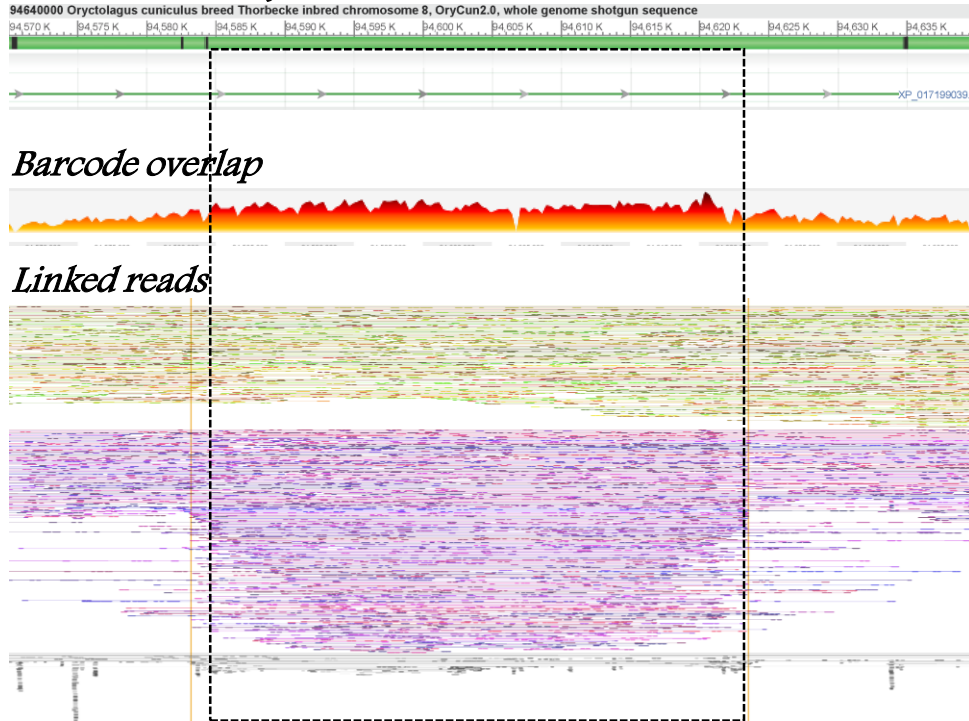
Duplication



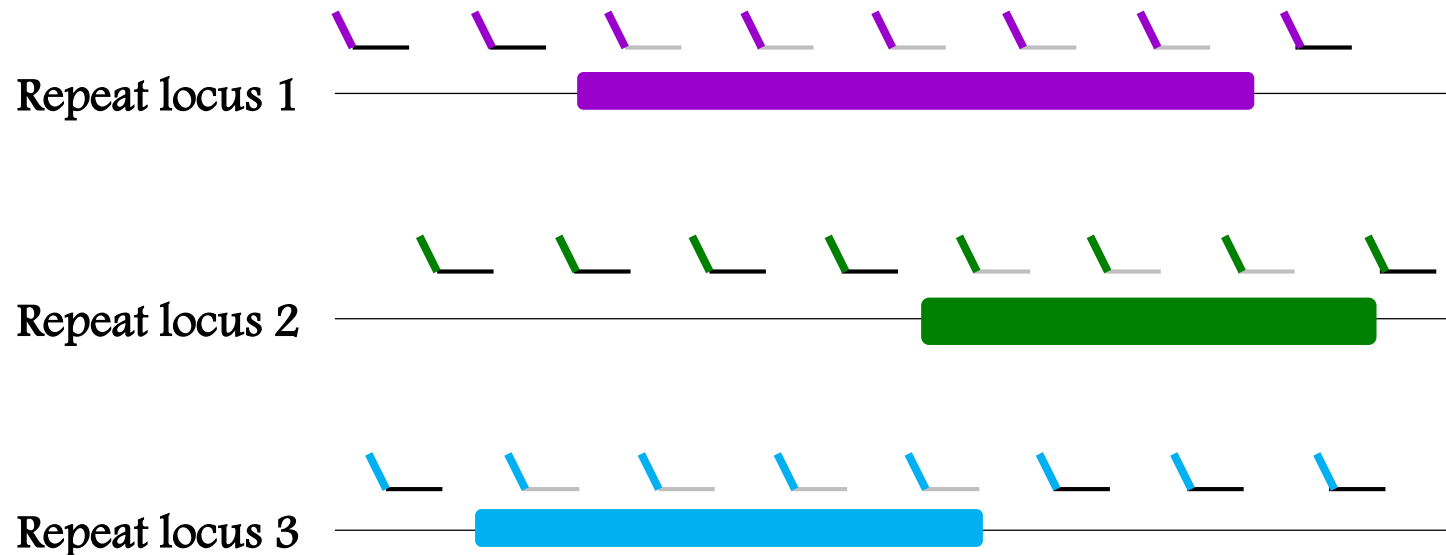
Heterozygous duplication

15404125 (mother)

Genome assembly



- Problem to assign reads in repetitive elements
- Information from 10X barcodes rescue unmapped reads



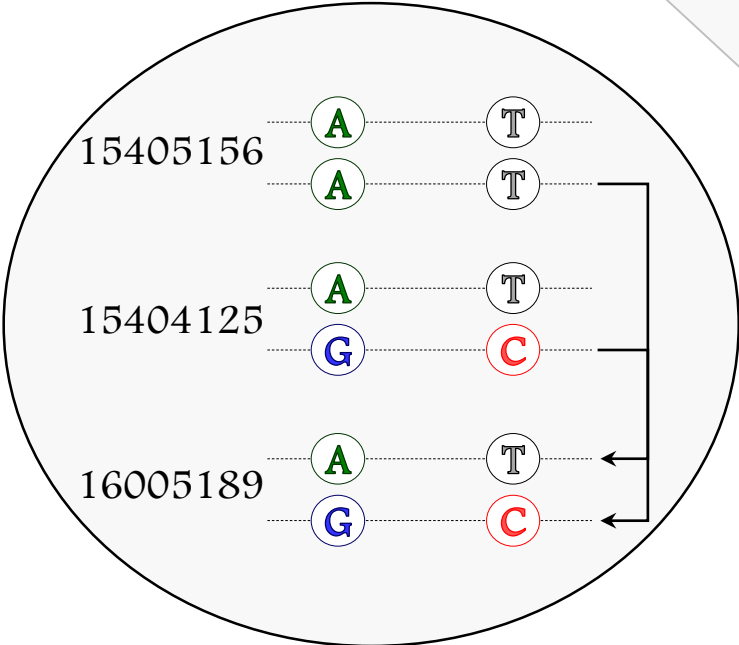
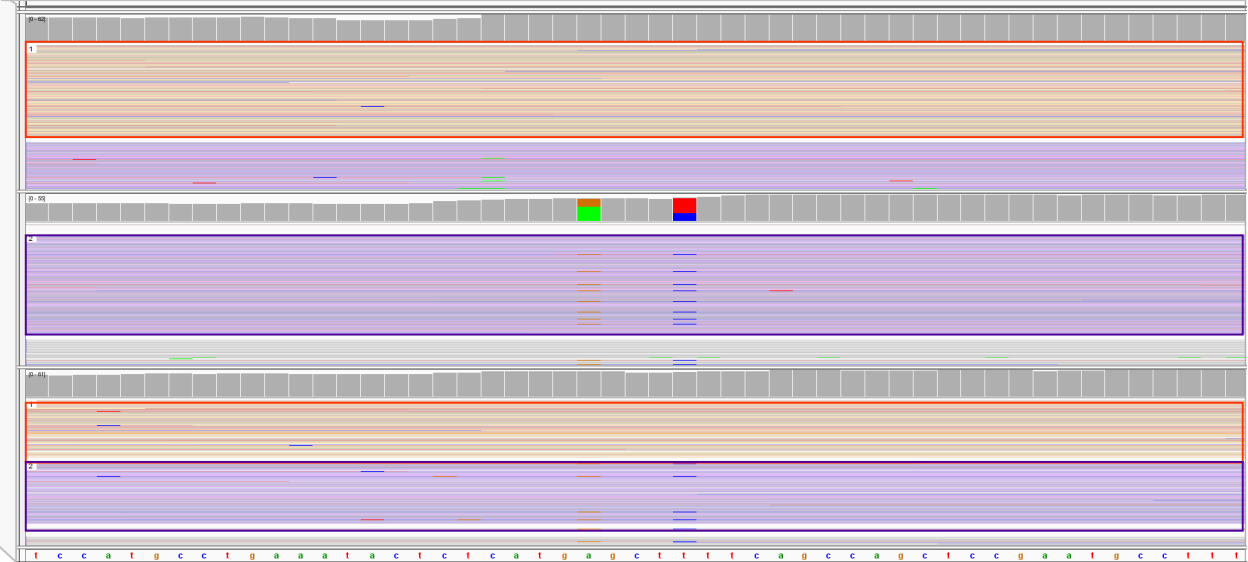
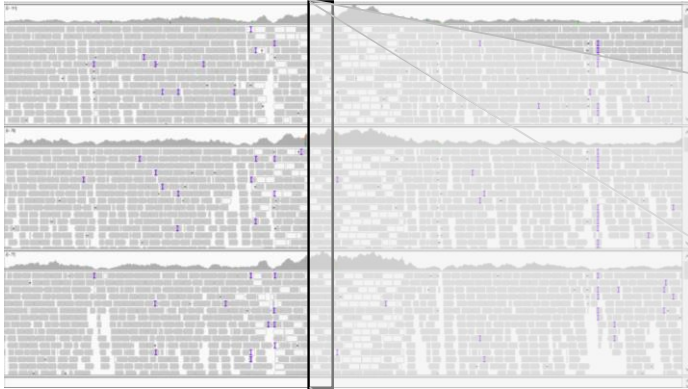
Recover variants in repetitive regions



15405156

15404125

16005189



- **Resolve the genome into long (megabases) phase blocks**

Phase the “full” spectrum of called variants

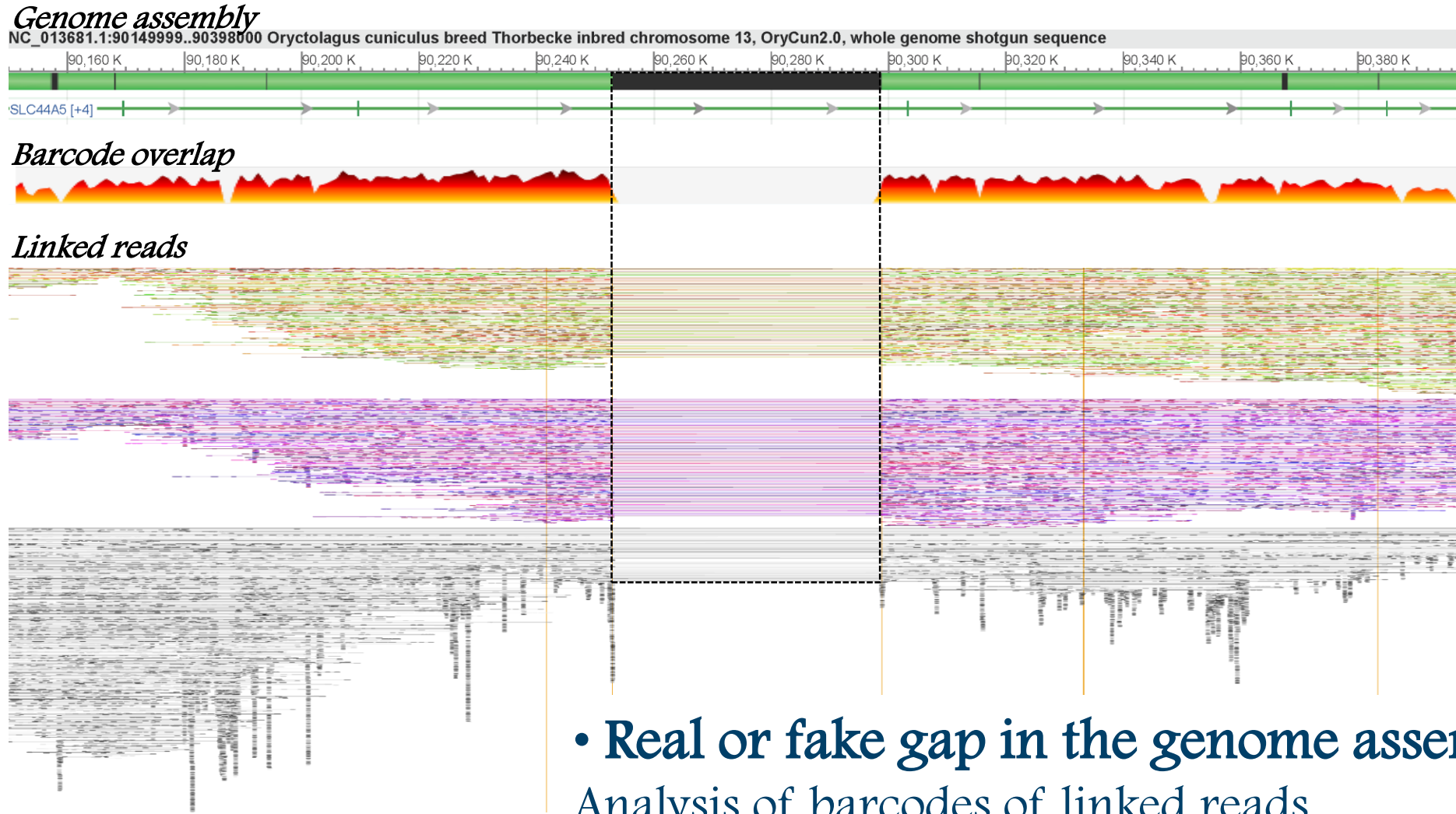
- **Identify structural variants and breakpoints**

Insertions, deletions, duplications, translocations...

- **Rescue variants in inaccessible parts of the genome**

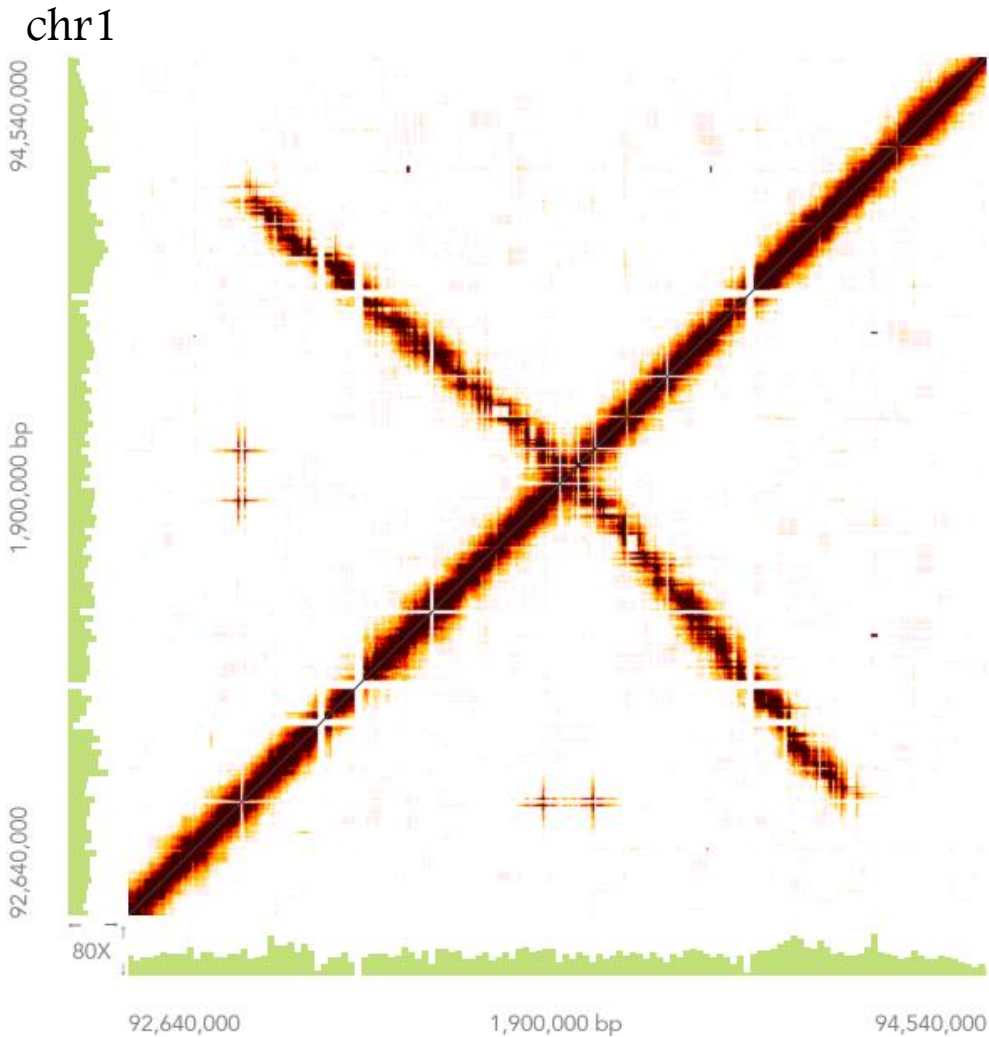
Confidently map reads in repetitive regions

- **Improve genome assembly**

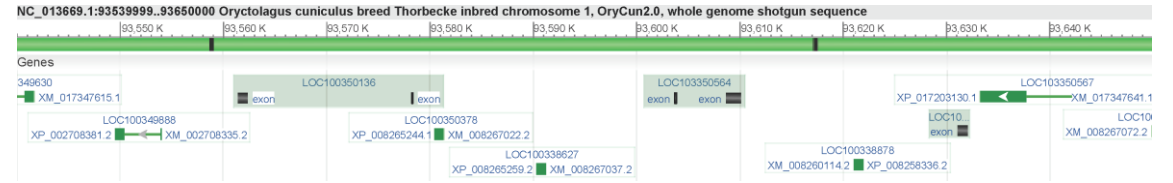




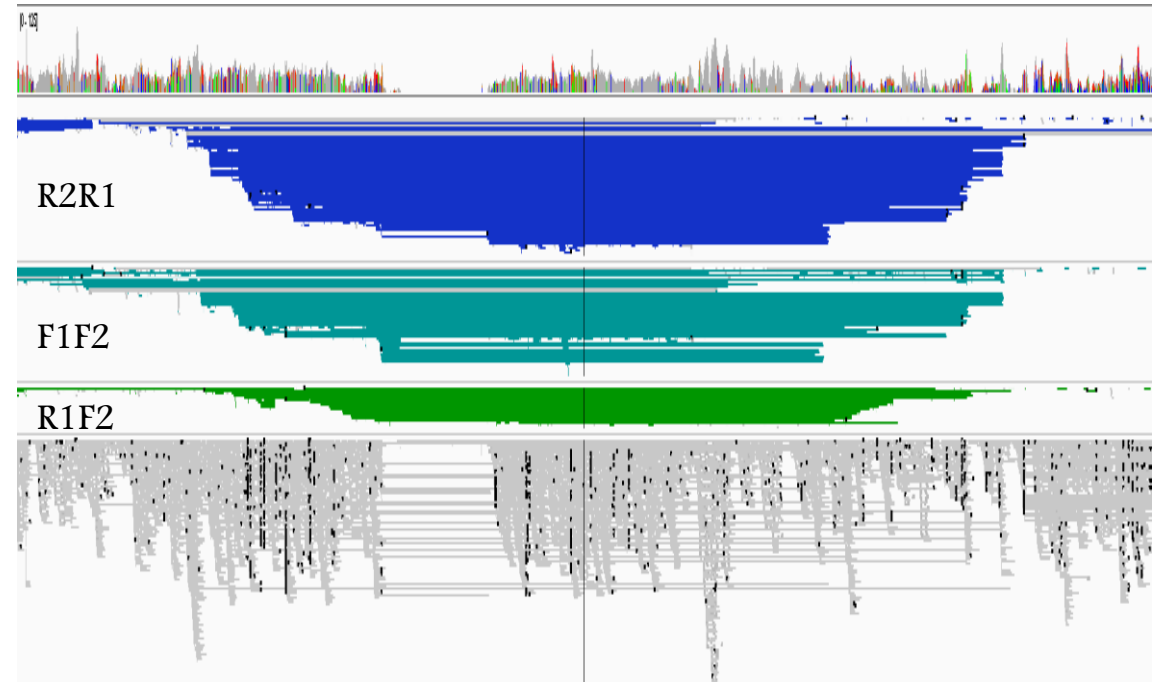
Complex rearrangement including inversion



Genome assembly



IGV window (pair orientation)



- Genotoul GeT~PlaGe



Olivier Bouchez
Cécile Donnadiou
Pauline Heuillard
Claire Kuchly
Sophie Valière

- GenPhySE



Isabelle Hochu
Sophie Leroux
Juliette Riquet
Julien Ruesche

- PECTOUL



Patrick Aymard
Elodie Balmisse
Virginie Helies