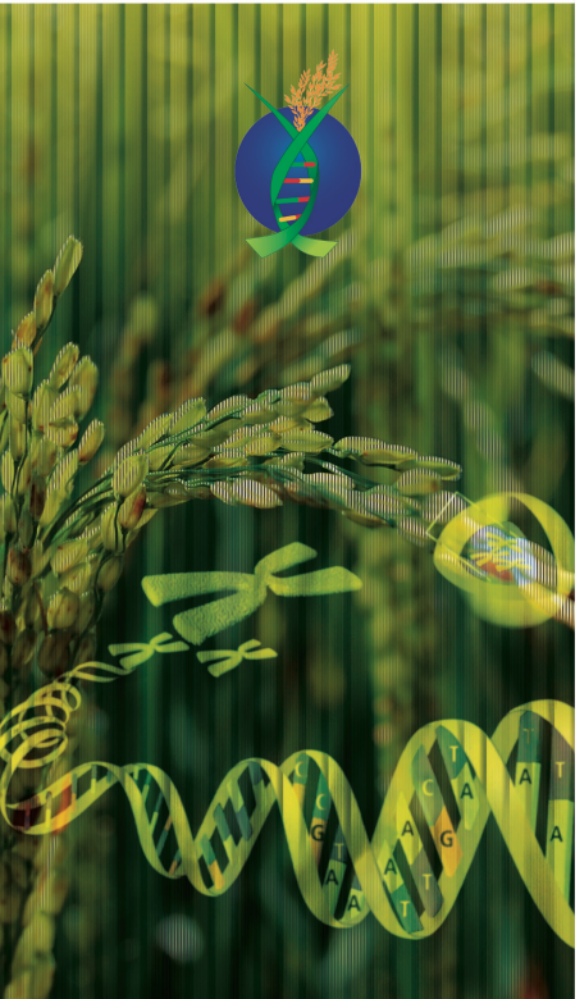
ISRFG 2021

HYBRID FORMAT

18th International Symposium on Rice Functional Genomics. Barcelona 3rd to 5th November, 2021



Certificate of Appreciation ISRFG 2021

This is presented to

Mathias Lorieux

for their contribution as a speaker at the

18th International Symposium on Rice Functional Genomics

held in Barcelona on 3rd - 5th November 2021 as an hybrid event

Barcelona, 3rd to 5th November, 2021

Blanca San Segundo
President of the Organising Committee
Centre for Research in Agricultural Genomics (CRAG), Barcelona

Meiotic recombination in rice is modified by structural genomic variation

Mathias Lorieux

18th International Symposium of Rice Functional Genomics
Barcelona, November 2021

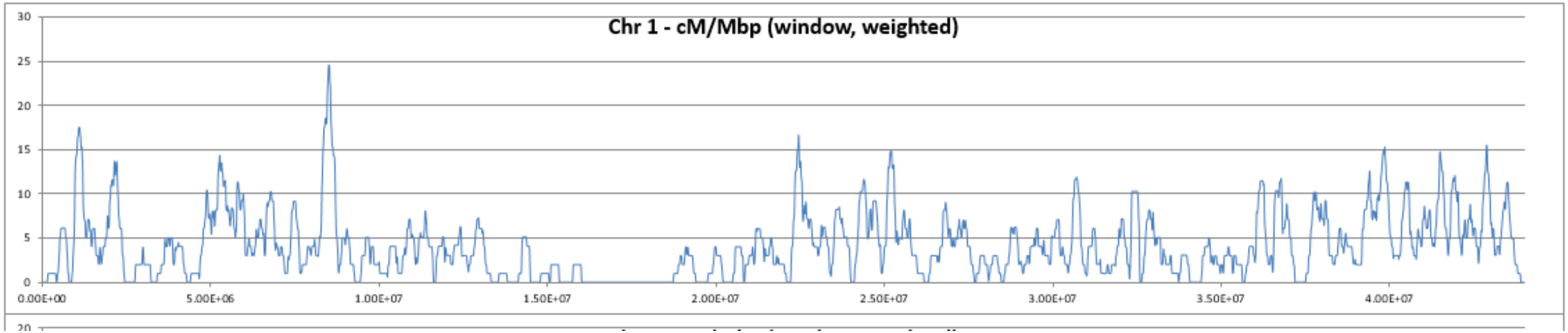
DIversity, **A**daptation and **DE**velopment *of plants*

Alliance

Meiotic recombination is not homogeneous

(Plant Genome I, 1992)

Local
recombination rate



telomere centromere telomere

Local recombination rate (in cM/Mbp) along rice chromosome 1

Why is recombination important?

- Academically interesting: how meiotic recombination works
- Genetic recombination is the basis of **plant breeding**

→ Phenotypes of progeny depend largely on recombination

→ More recombinants = higher $p(\text{interesting phenotype})$

Understanding the laws of recombination means ability to *predict* it...
...so we can select the best progenitors for breeding

Some known factors that affect recombination in plants

GC3 content

Serres-Giardi et al (2012). The Plant Cell, three grass species

Sequence peculiarities in and around ‘hot spots’

Kianian et al (2018). Nature Communications, maize

de Haas et al (2017). DNA research, tomato

Pericentromeric heterochromatin/DAPI-bright regions

Cheng et al (2001). Genome Research, Rice

de Haas et al (2017). DNA research, tomato

Certain genes families

Chen et al (2017) Plos One, Cotton

Genes and TEs abundance

Colomé-Tatché et al (2012). PNAS, Arabidopsis

DNA methylation

Mirouze et al (2012). PNAS, Arabidopsis

... and many more

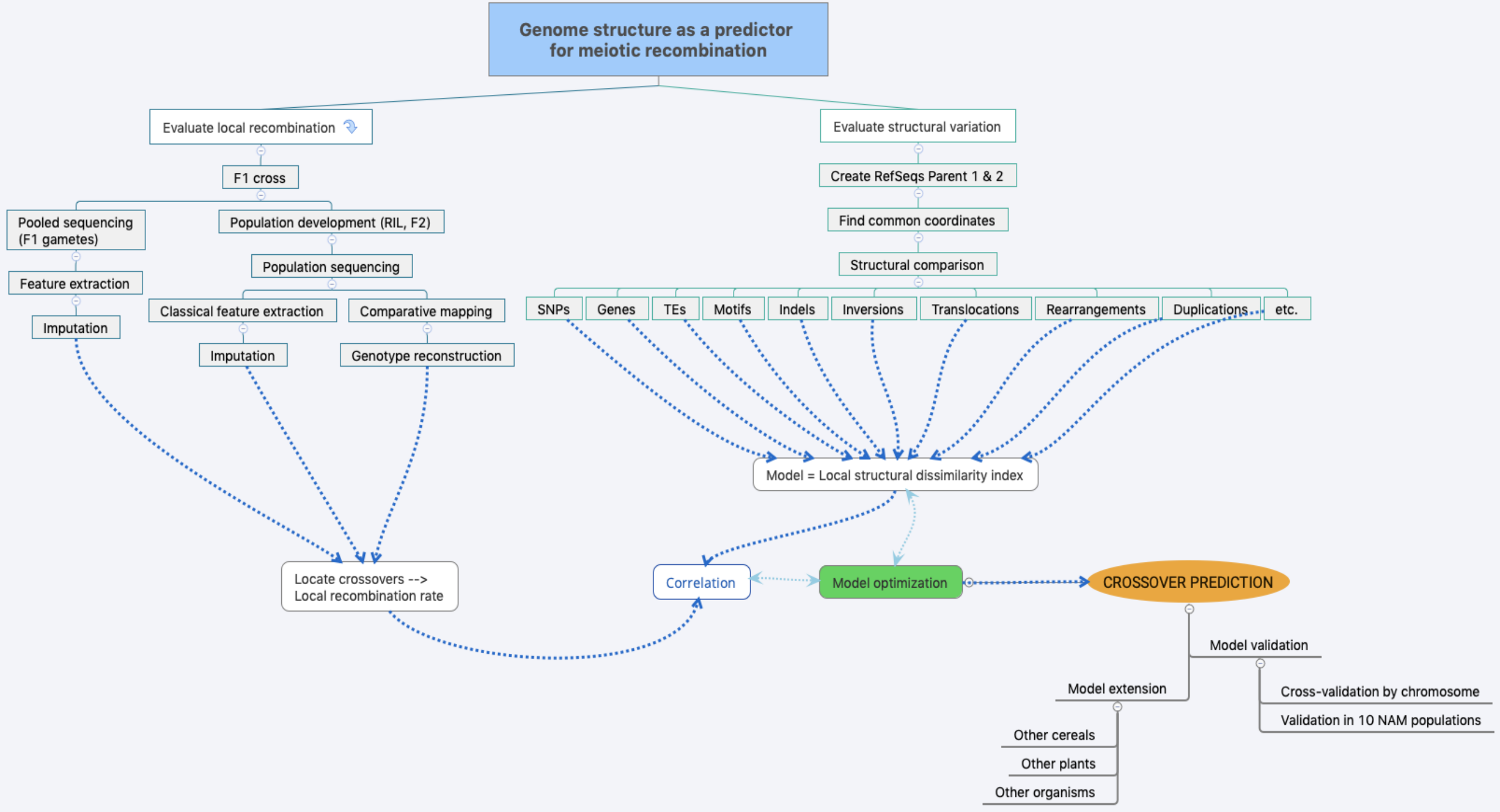
Is genomic structural variation also a factor?

Hypothesis: recombination rates in the F_1 hybrid are altered by structural differences between the parental genomes

Test the hypothesis: three steps

1. Estimate F_1 recombination as precisely as possible
2. Assess genomic variation between the parents
   - *not* Illumina re-sequencing
   - we want *de novo* sequences because we need the genome *structure*
3. Search for correlations between recombination and genomic sequence differences

If successful, derive a model for recombination based on parental sequence properties


Genome structure as a predictor for meiotic recombination
Evaluate local recombination
F1 cross
Pooled sequencing (F1 gametes)
Population development (RIL, F2)
Feature extraction
Population sequencing
Imputation
Classical feature extraction
Comparative mapping
Imputation
Genotype reconstruction
Evaluate structural variation
Create RefSeqs Parent 1 & 2
Find common coordinates
Structural comparison
SNPs
Genes
TEs
Motifs
Indels
Inversions
Translocations
Rearrangements
Duplications
etc.
Model = Local structural dissimilarity index
Locate crossovers --> Local recombination rate
Correlation
Model optimization
CROSSOVER PREDICTION
Model validation
Cross-validation by chromosome
Validation in 10 NAM populations
Model extension
Other cereals
Other plants
Other organisms

1. Estimating local recombination

$r = 0.59$

Recombination data [prototype]

- A population of 212 F_{12} *recombinant inbred lines* (RILs)
  [IR64 × Azucena] → F_1 → F_2 → ... → F_{12}

  inter-subspecific *indica* × *japonica* → reveals wide range of genomic differences (400-600k years divergence)

- A set of ~1.8 million SNP markers
  - Illumina ~1.5x paired-end whole-genome sequencing (WGS)

Problem 1: missing data

Reducing cost has a cost

- lots of missing data
- sequencing errors not corrected by other reads
- heterozygotes not seen

We need *imputation*

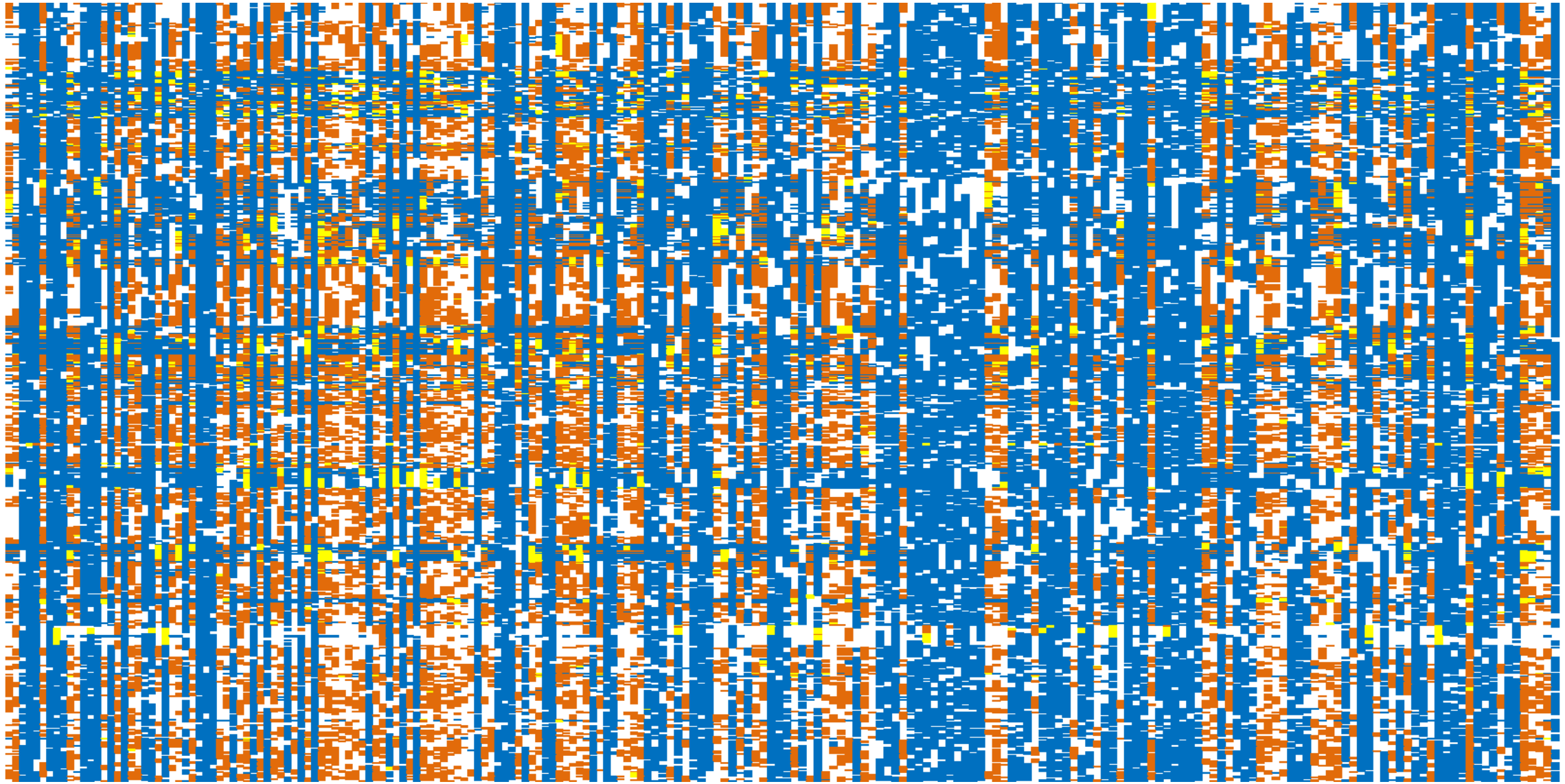
Simulation of a 1.5 X WGS coverage (Poisson distribution) on 100 discrete positions

Problem 2: noisy data

Individuals

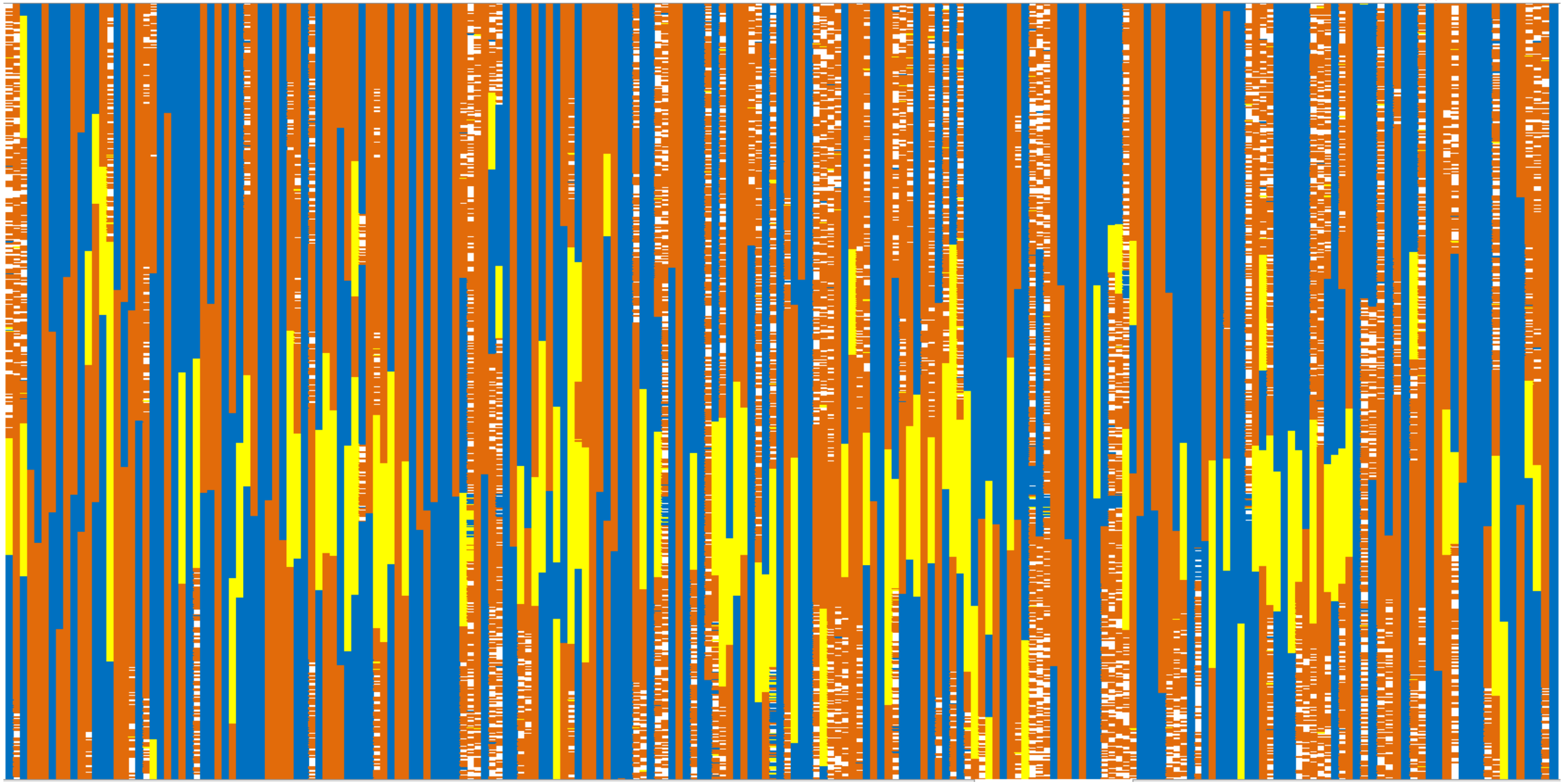
SNPs

[IR64 x Azucena] RILs (F_{12}), WGS@1.5X, raw data (filtered)



Imputation with Genotype-Corrector

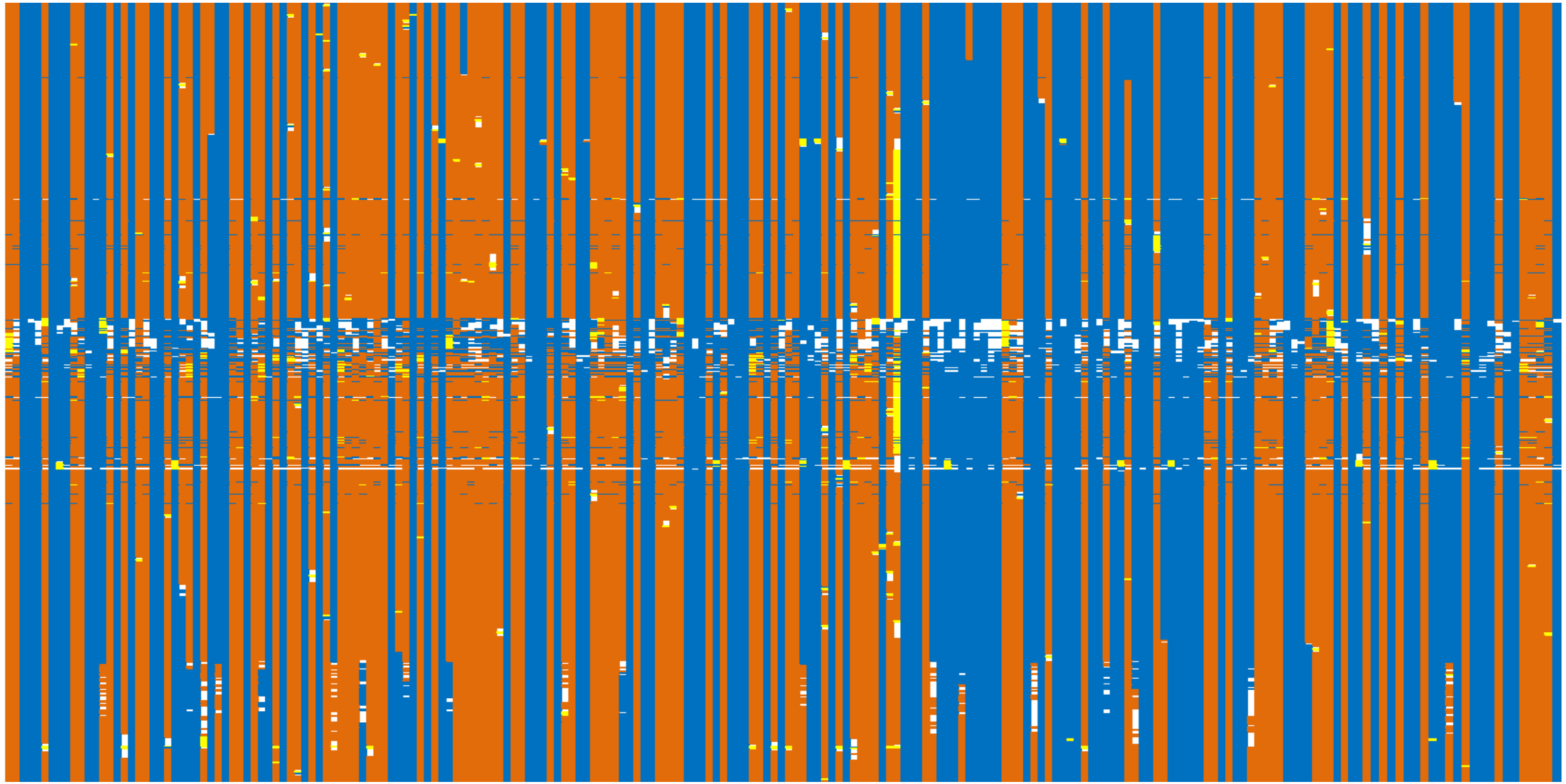
Genotype-Corrector: improved genotype calls for genetic mapping in F_2 and RIL populations
SCIENTIFIC REPORTS 2018



Imputation with Tassel-FSFHap

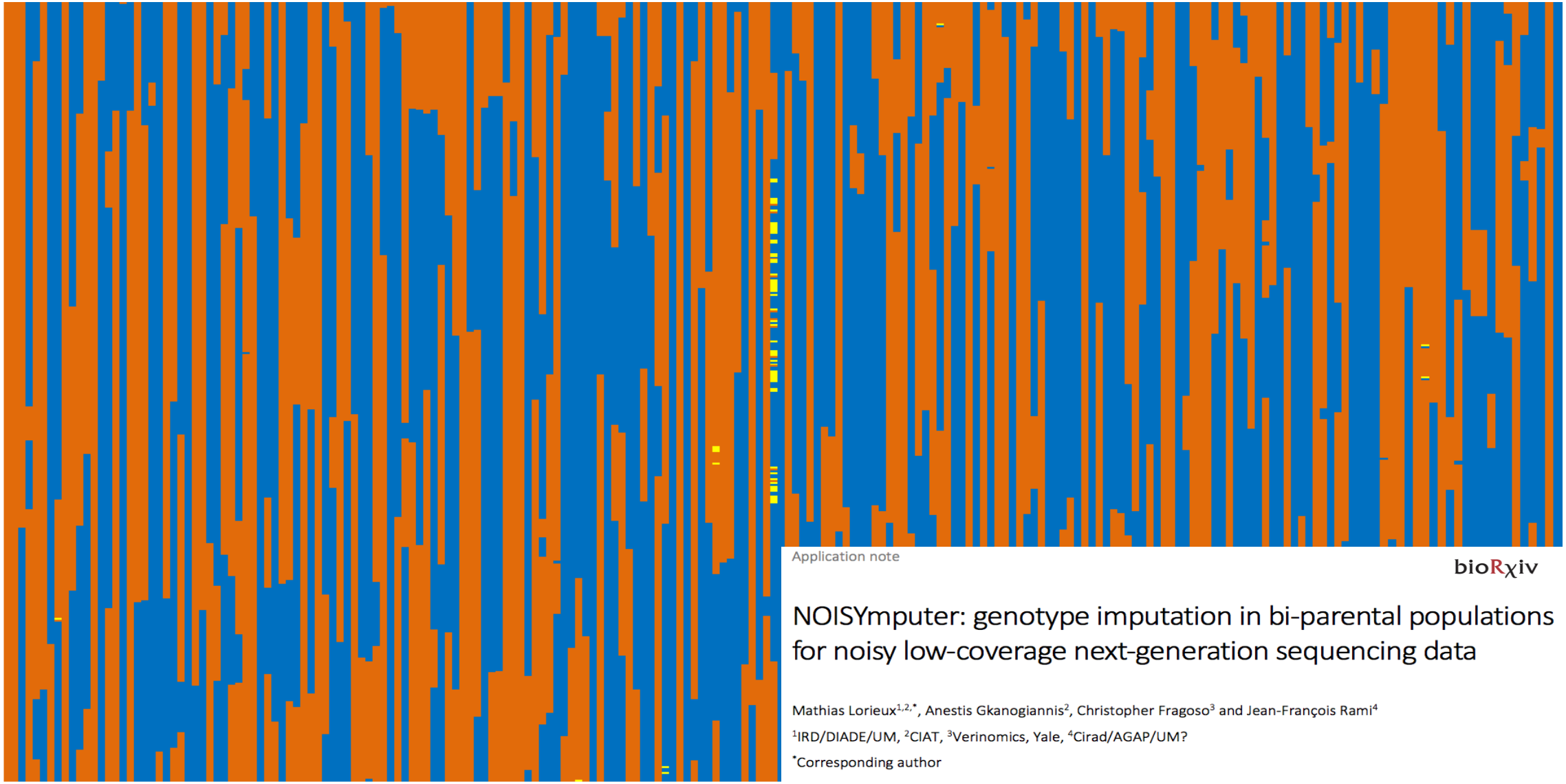
Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants

THE PLANT GENOME ■ NOVEMBER 2014 ■ VOL. 7, NO. 3



New imputation approach

v. 0.5: improved breakpoint impution in F_2s



Application note

bioRχiv

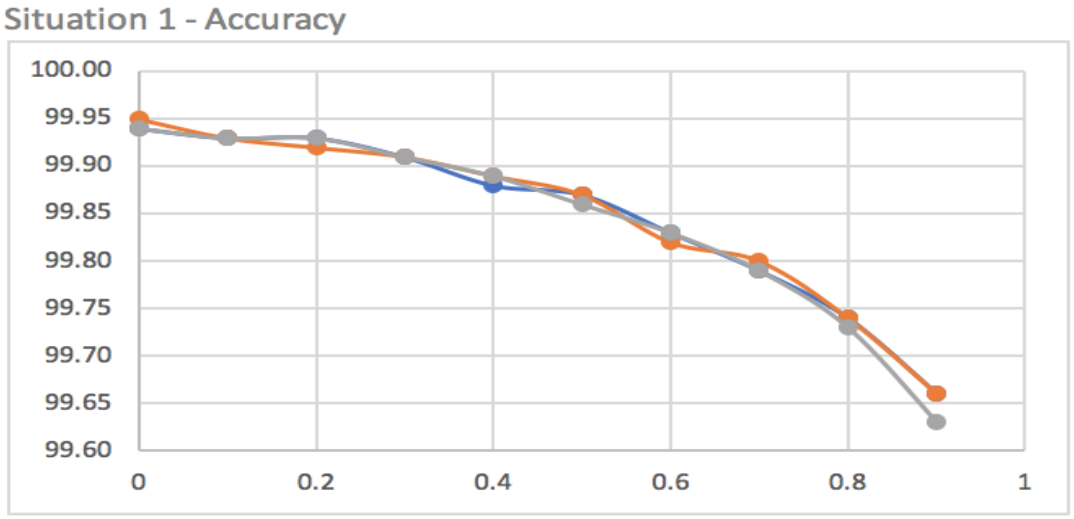
NOISYmputer: genotype imputation in bi-parental populations for noisy low-coverage next-generation sequencing data

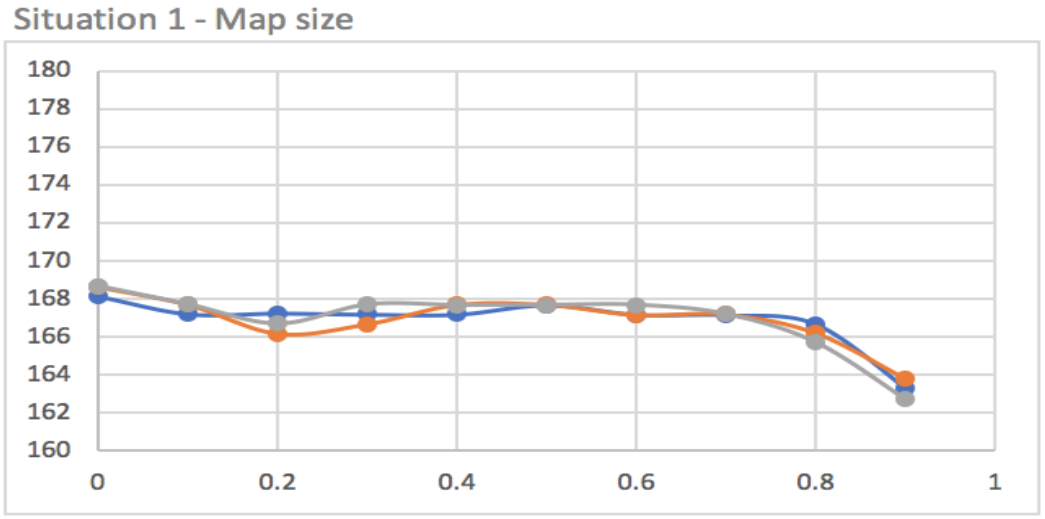
Mathias Lorieux[1,2,*], Anestis Gkanogiannis[2], Christopher Fragoso[3] and Jean-François Rami[4]

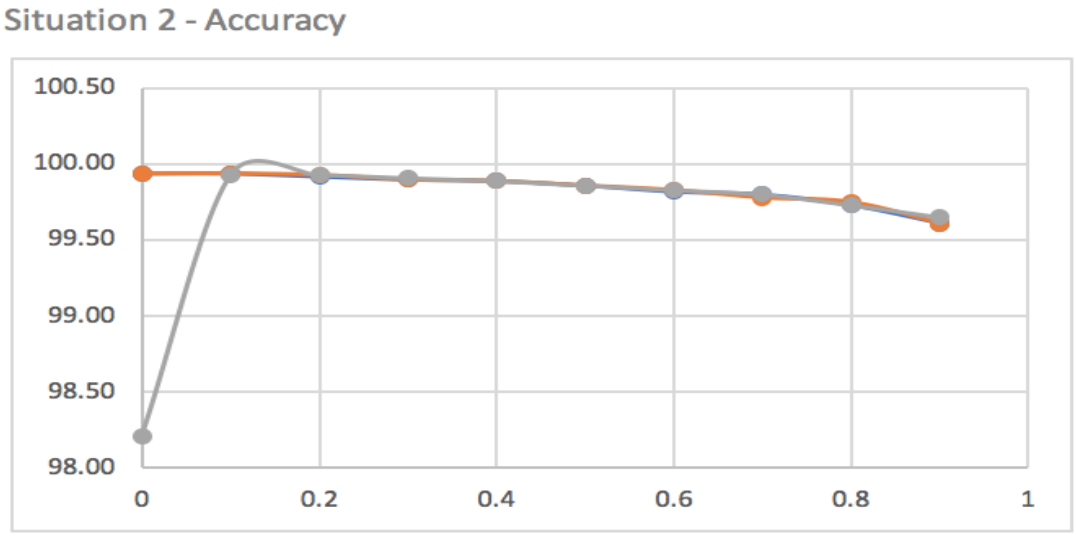
[1]IRD/DIADE/UM, [2]CIAT, [3]Verinomics, Yale, [4]Cirad/AGAP/UM?

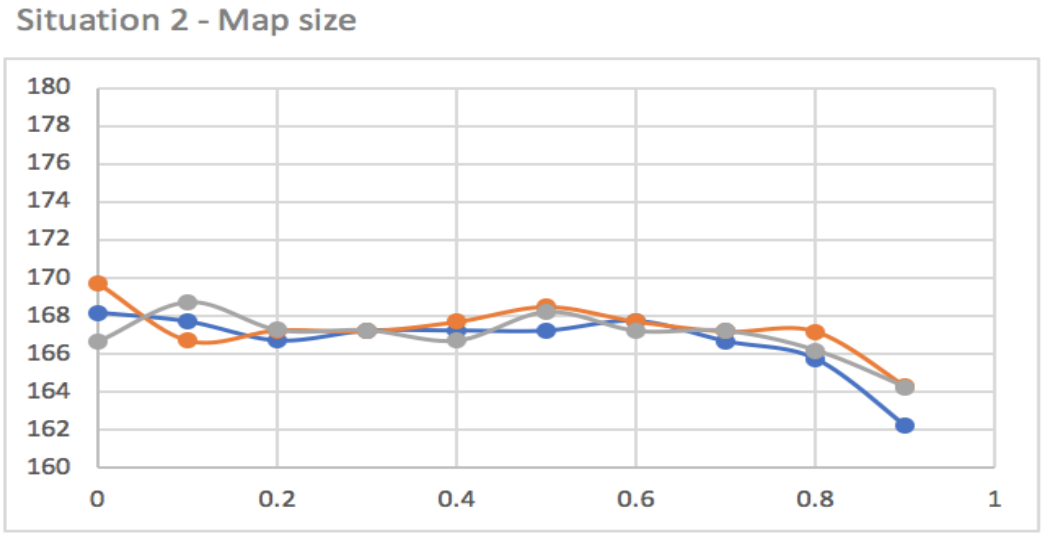
[*]Corresponding author

Imputation accuracy (simulated data)

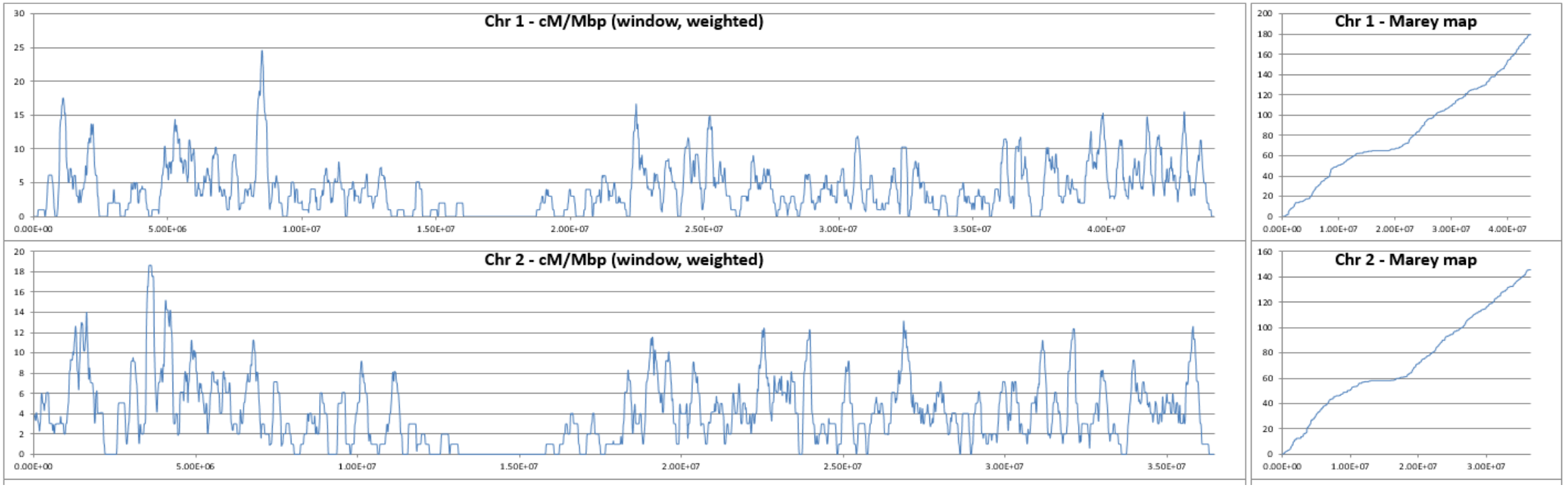








Now we can compute local recombination rates



Genetics and population analysis

Bioinformatics, 2017

Constructing linkage maps in the genomics era with MapDisto 2.0

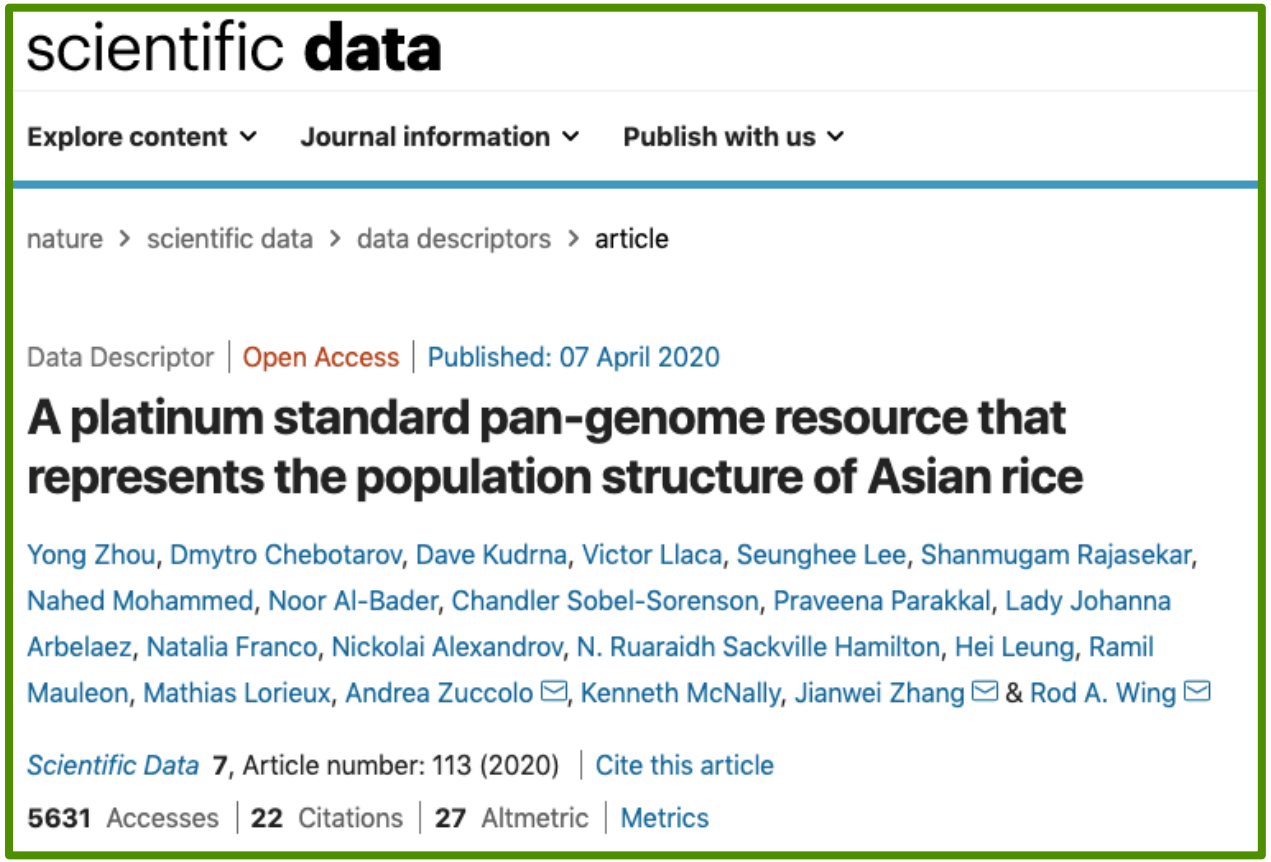
Christopher Heffelfinger[1], Christopher A. Fragoso[1] and Mathias Lorieux[2,3,*]

2. Estimating local genome variation

Genomic data

High-quality PacBio genomes of the two parents,
IR64 and Azucena – ~120x PacBio + ~40x Illumina

COLLEGE OF AGRICULTURE & LIFE SCIENCES
Arizona Genomics Institute



scientific data

Explore content | Journal information | Publish with us

nature > scientific data > data descriptors > article

Data Descriptor | Open Access | Published: 07 April 2020

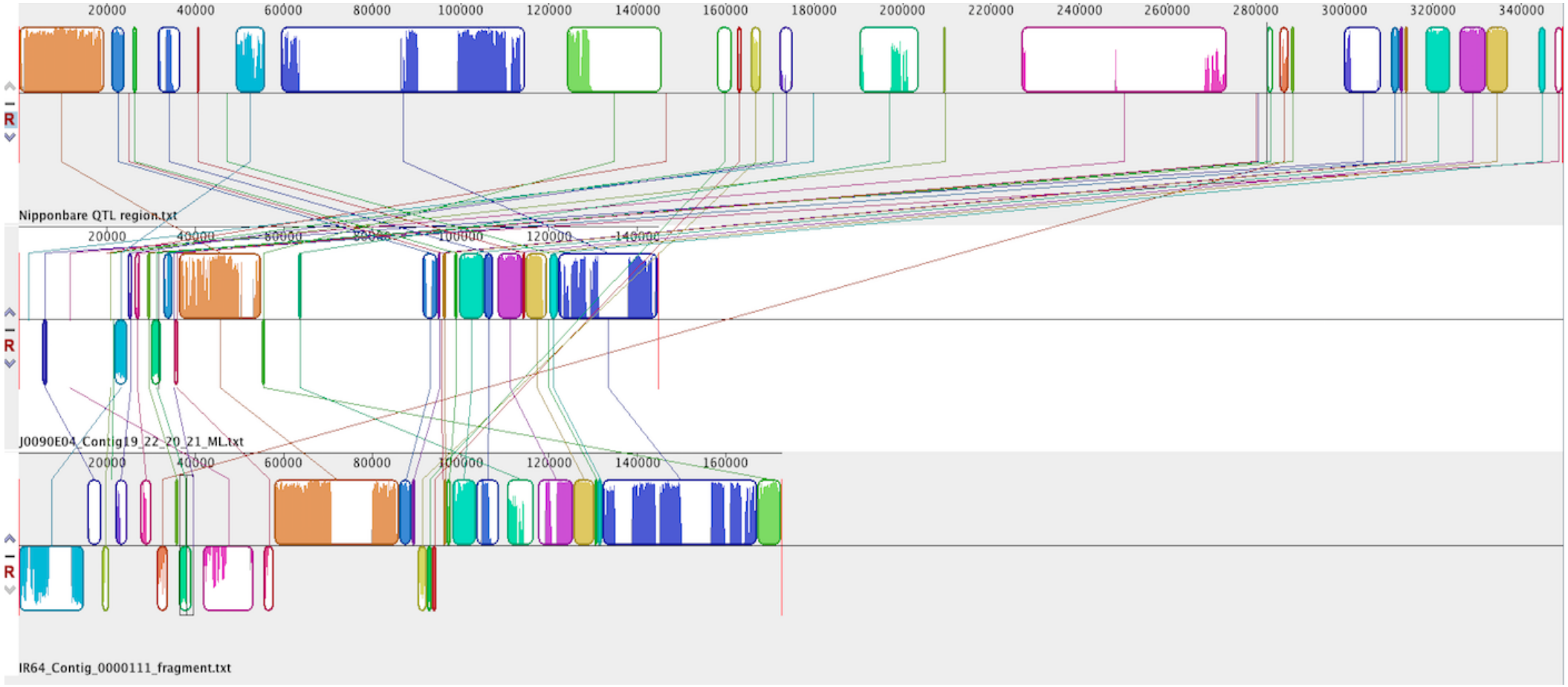
A platinum standard pan-genome resource that represents the population structure of Asian rice

Yong Zhou, Dmytro Chebotarov, Dave Kudrna, Victor Llaca, Seunghee Lee, Shanmugam Rajasekar, Nahed Mohammed, Noor Al-Bader, Chandler Sobel-Sorenson, Praveena Parakkal, Lady Johanna Arbelaez, Natalia Franco, Nickolai Alexandrov, N. Ruaraidh Sackville Hamilton, Hei Leung, Ramil Mauleon, Mathias Lorieux, Andrea Zuccolo, Kenneth McNally, Jianwei Zhang & Rod A. Wing

Scientific Data **7**, Article number: 113 (2020) | Cite this article

5631 Accesses | **22** Citations | **27** Altmetric | Metrics

IR64 and Azucena can vary substantially



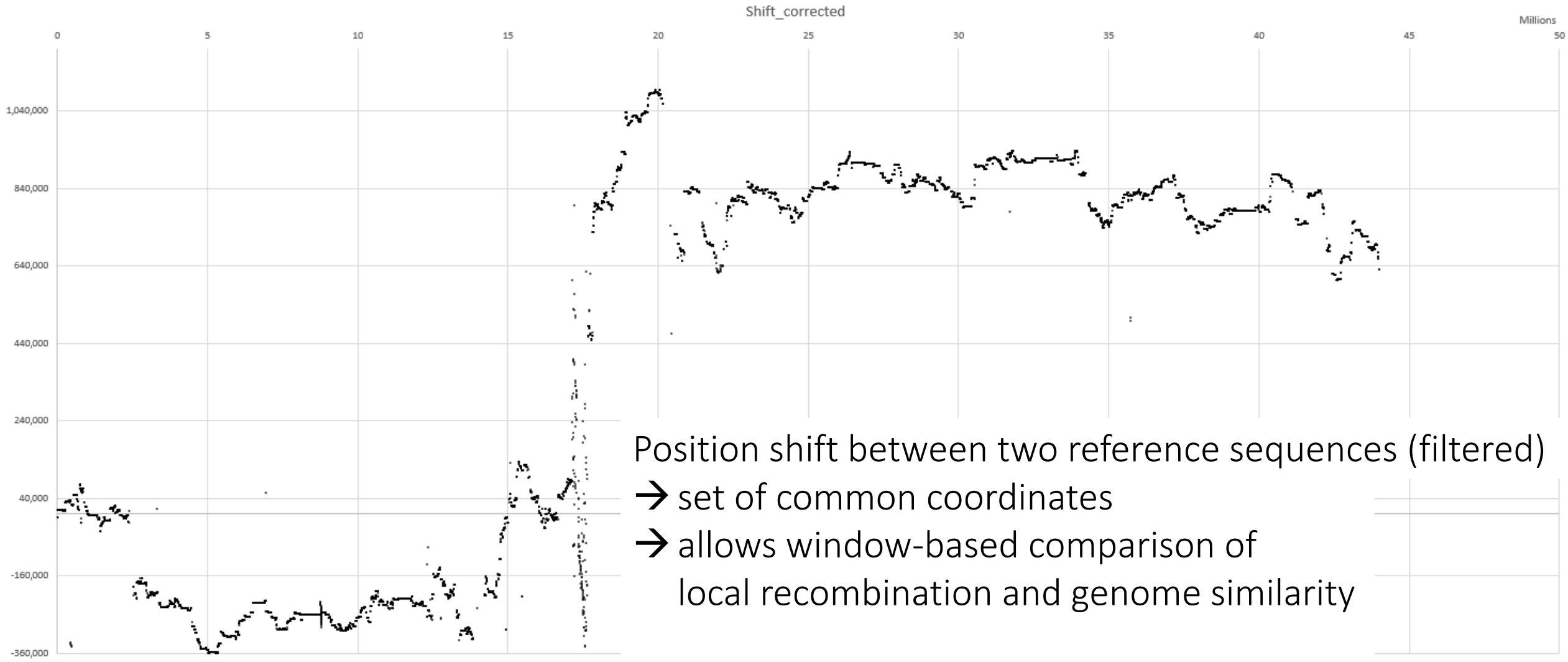
Azucena

(Nipponbare)

IR64

Graphics from Mauve analysis

Finding common coordinates



Position shift between two reference sequences (filtered)

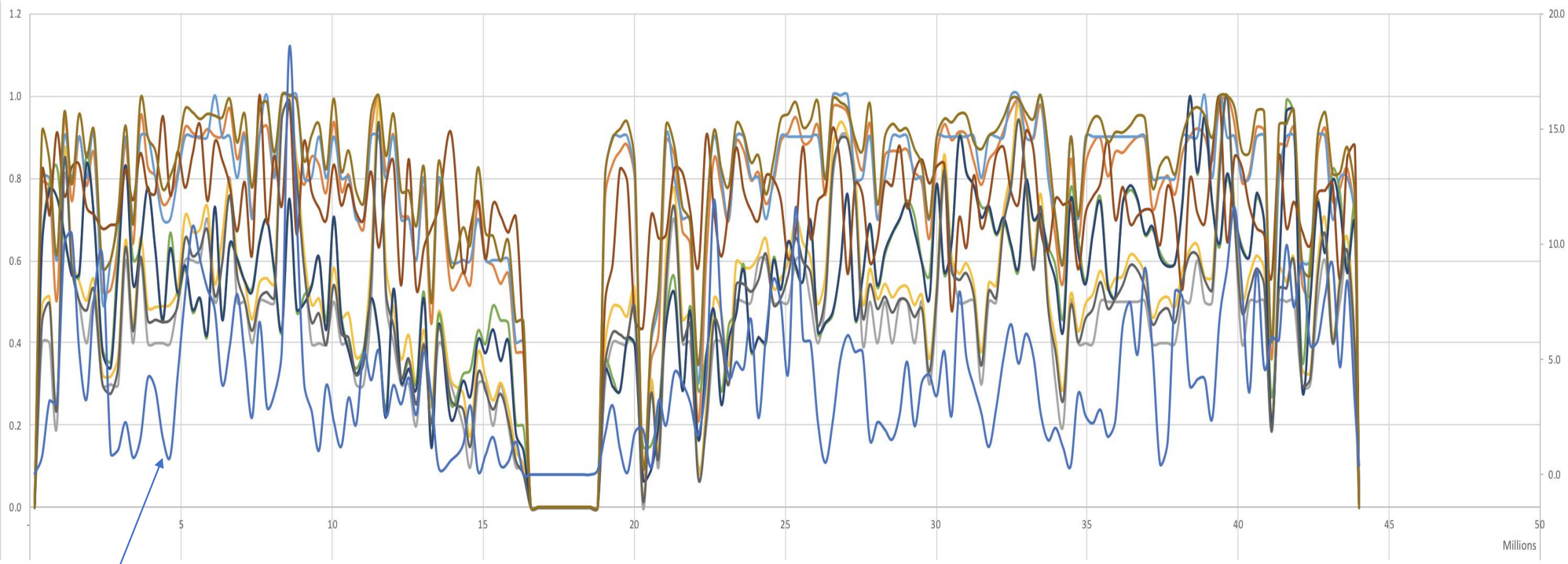
→ set of common coordinates

→ allows window-based comparison of local recombination and genome similarity

3. Comparing recombination with genome variation

r = 0.59

Comparing genomes: which indicator?

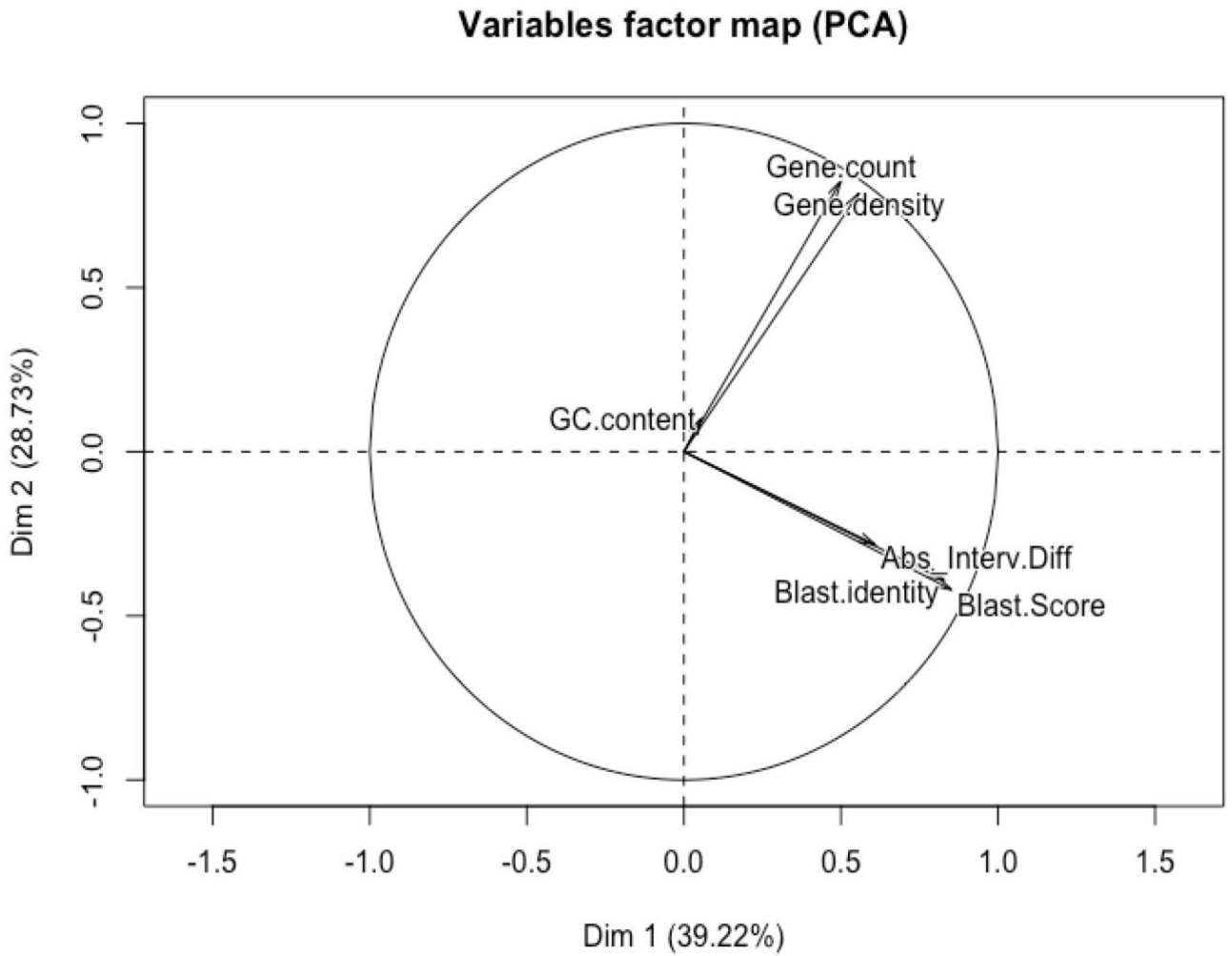


cM/Mbp

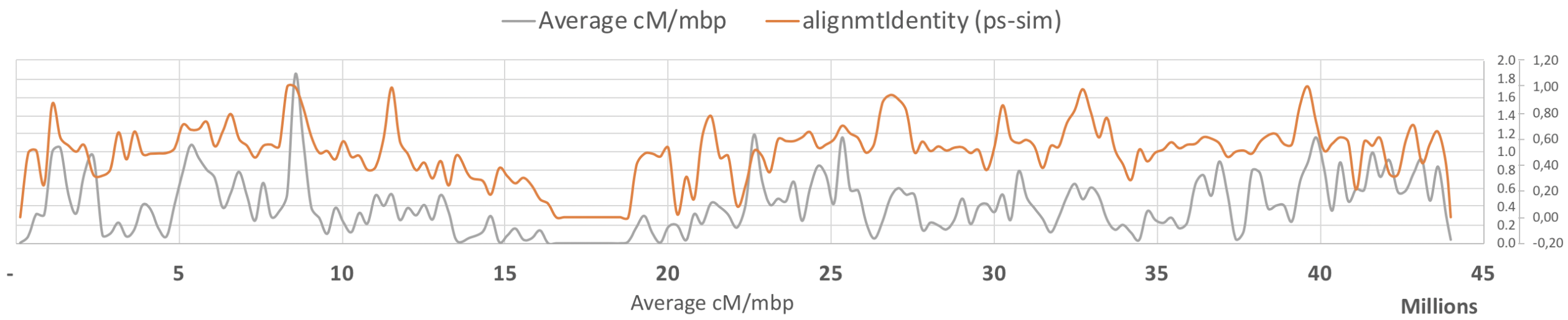
Other curves:

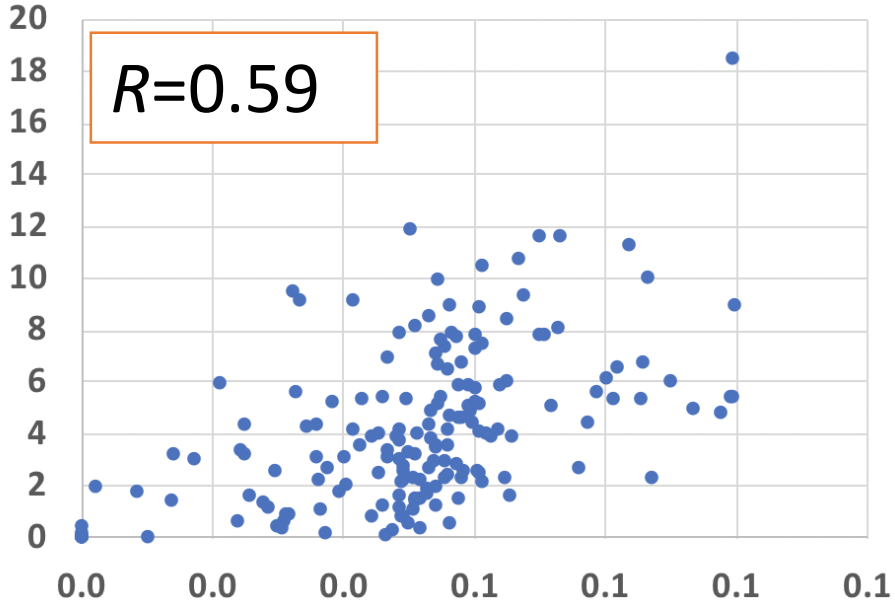
Chunk size difference; BLAST score; BLAST identity; GC %; Gene count; Gene density; TE density; Combined; SNP abundance

Finding uncorrelated features



Blast score + gene density *vs.* recombination





Let’s try something better

Mauricio Peñuela

Camila Riccio

Reference Sequence

Query Sequence

Pre-processing Protocol

- Alignment
- Variant calling
- Identity measure for each window

Based on nucmer outputs

Model

Step 1	Step 2	Step 3	Step 4
Case 1: penalty Case 2: award Case 3: correction	Zeroing negative values	CentO detection Centromere correction	Data smoothing

Prediction of recombination

$$Id(w) = window\_size - V(w) - I(w) - A(w)$$

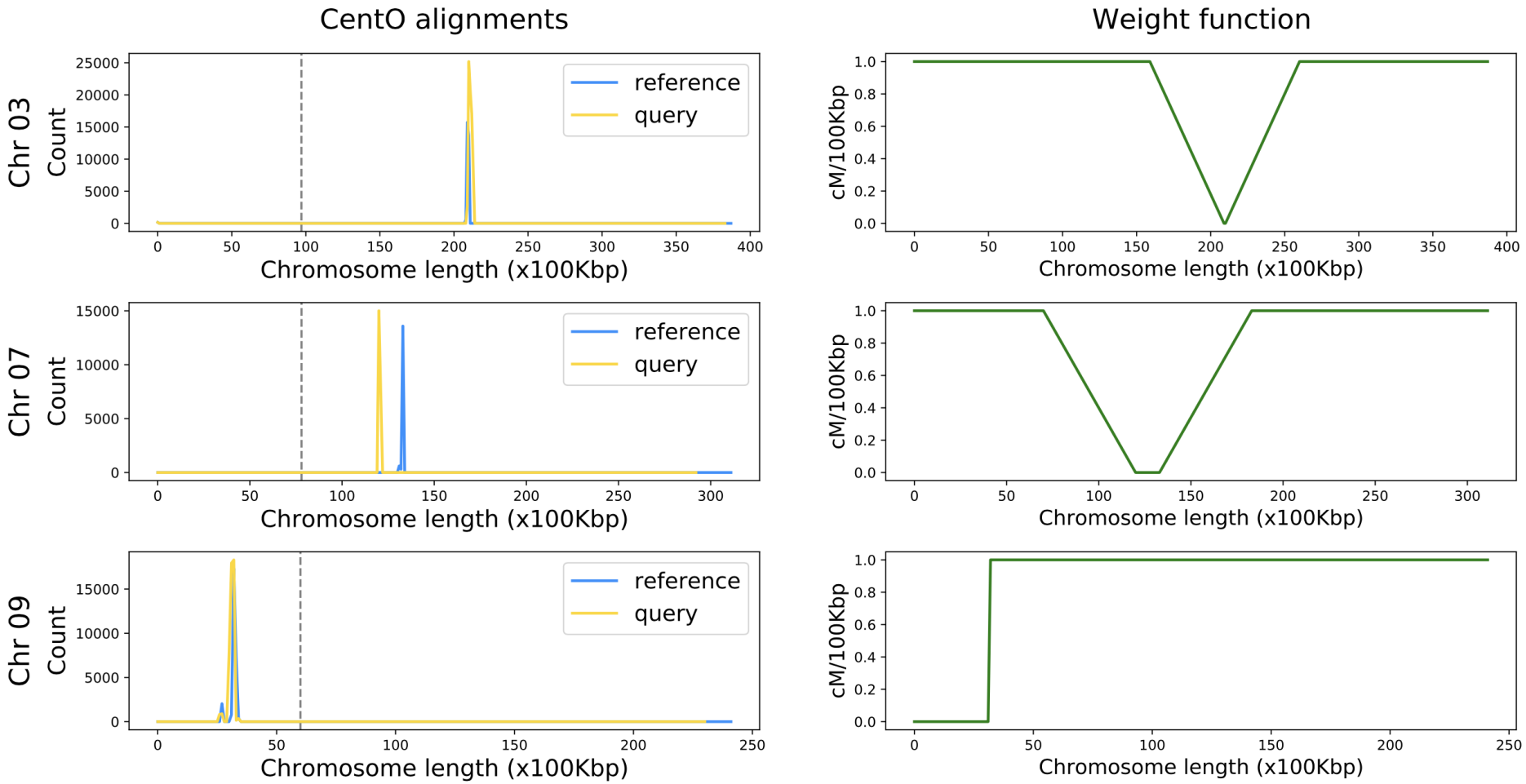
# of variants

 # bases in inversion

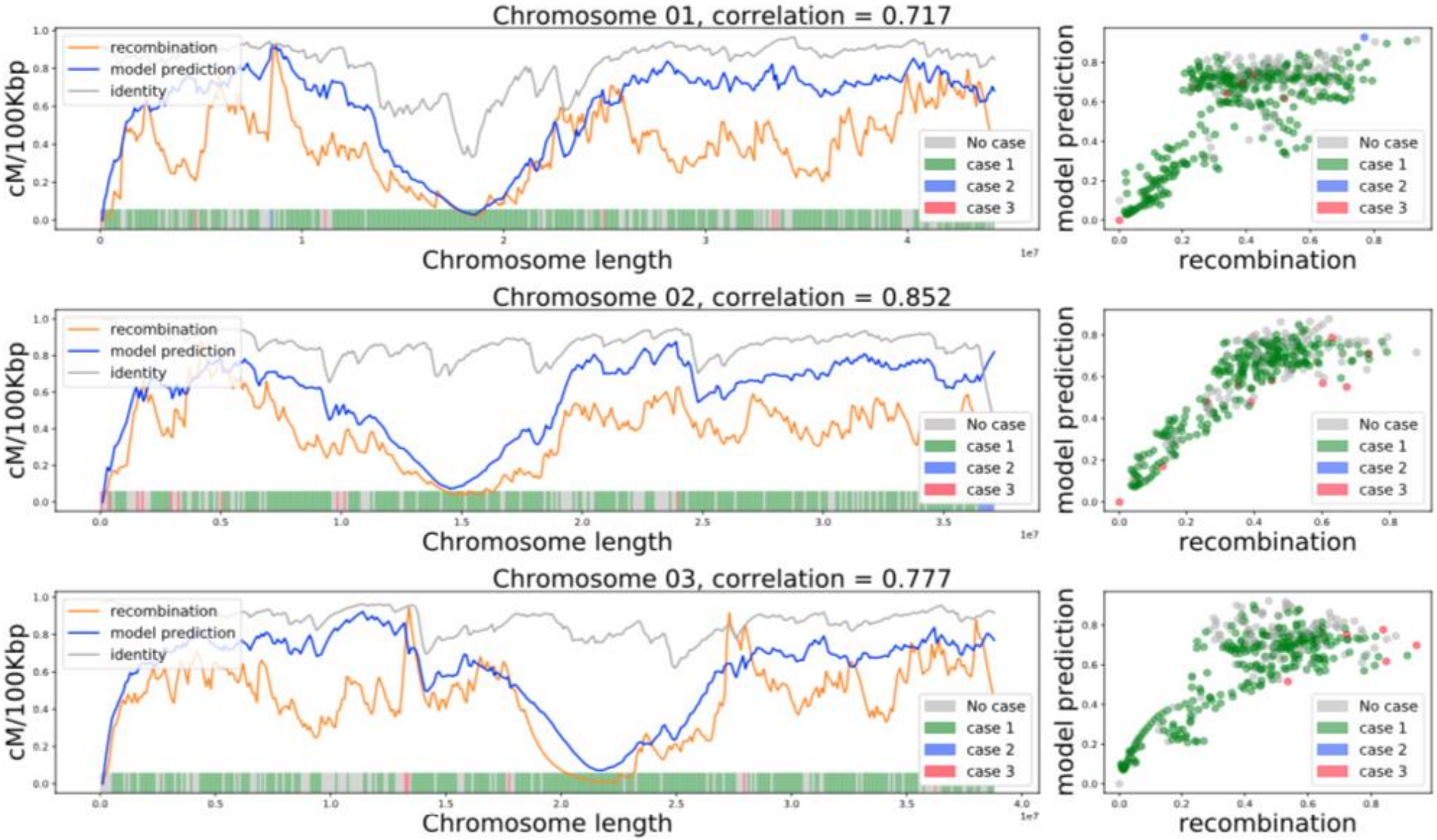
 # absent bases

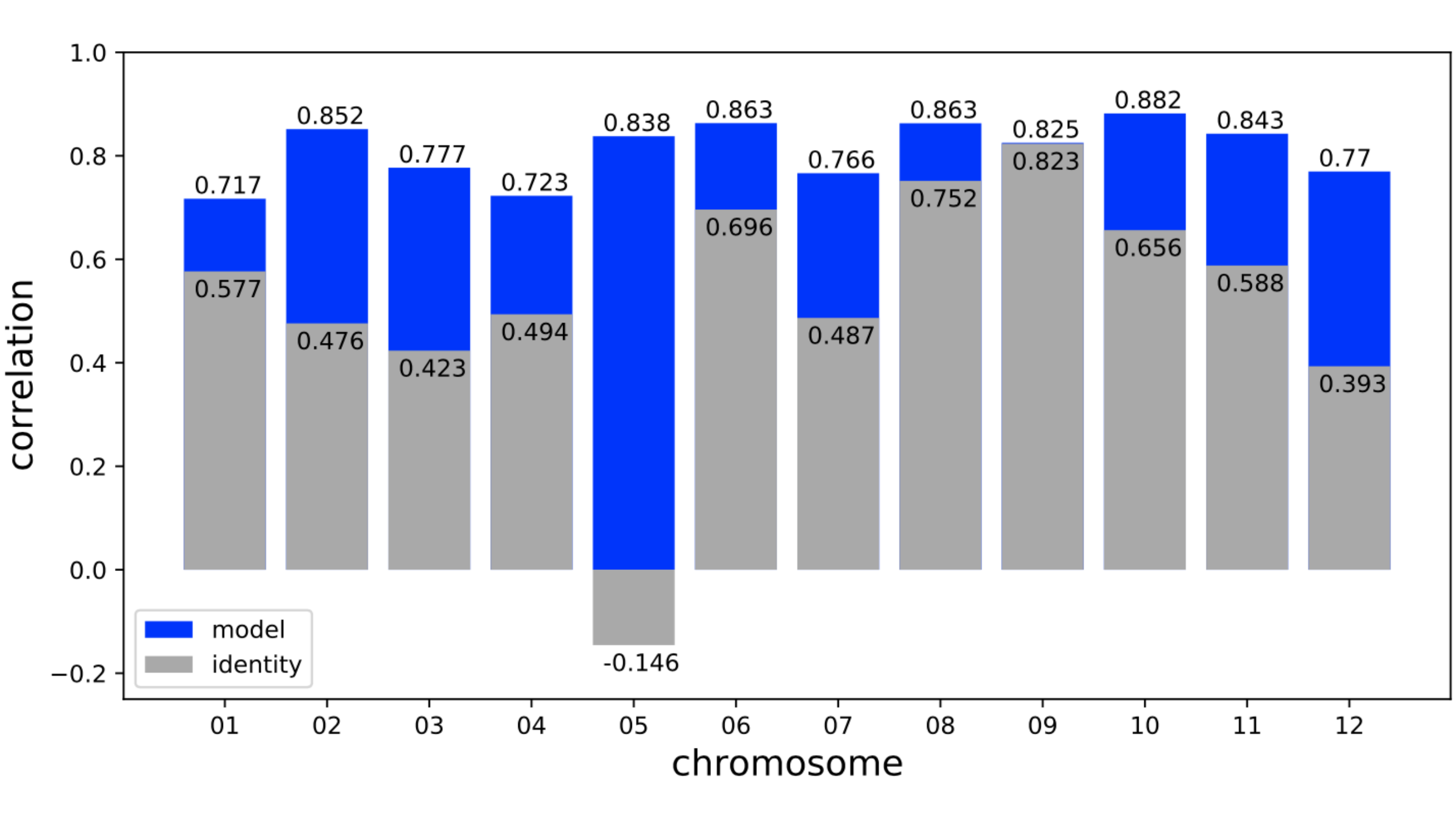
nucmer outputs parsing

Centromere detection using CentO sequences and centromere correction



Local recombination vs. model (chromosome 1)




1.0
0.8
0.6
0.4
0.2
0.0
−0.2
correlation
0.717
0.852
0.777
0.723
0.838
0.863
0.766
0.863
0.825
0.882
0.843
0.77
0.577
0.476
0.423
0.494
-0.146
0.696
0.487
0.752
0.823
0.656
0.588
0.393
model
identity
01
02
03
04
05
06
07
08
09
10
11
12
chromosome

Draft manuscript in prep

Sequence Identity Allows Prediction of Crossover Recombination

Mauricio Peñuela[1], Camila Riccio[1], Jorge Finke[1], Camilo Rocha[1], Anestis Gkanogiannis[2], Rod A. Wing[3], and Mathias Lorieux[2,4]

[1]Facultad de Ingeniería y Ciencias, Pontificia Universidad Javeriana. Cali, Colombia.
[2]AgroBiotechnology Unit, Alliance Bioversity-CIAT. Cali, Colombia.
[3]Arizona Genomics Institute, University of Arizona. Tucson, AZ, US
[4]DIADE, University of Montpellier, CIRAD, IRD. Montpellier, France.

Abstract

Meiotic recombination is a crucial cellular process, being one of the major drivers of evolution and adaptation of species. In plant breeding, crossing is used to introduce genetic variation among individuals and populations. A better characterization of the variation of the recombination rates along the chromosomes would enable breeding programmes to increase the chances of creating novel allele combinations, and more generally, to introduce new varieties with a collection of desirable traits. While different approaches to predict recombination rates for different species have been developed, they fail to estimate the outcome of a crossing between two specific accessions. This is missing in the panel of tools that breeders can use to reduce costs and execution times of crossing experiments. This papers builds on the hypothesis that chromosomal recombination correlates positively to a measure of sequence identity. In particular, we develop a model that uses sequence identity, combined with other features derived from genome alignment (including the number of variants, inversions, absent bases, and CentO sequences) to predict local chromosomal recombination in rice. Model performance is validated in an inter-subspecific *indica* x *japonica* cross, using 212 recombinant inbred lines. Across all 12 chromosomes, an average correlation of about 0.8 between experimental and prediction rates is achieved.

4. Now the life-size project

r = 0.59

LANDSREC: The high-resolution landscapes in rice meiotic recombination

E. Guiderdoni (PI) **M. Lorieux (co-PI)** **D. Zhou (co-PI)**

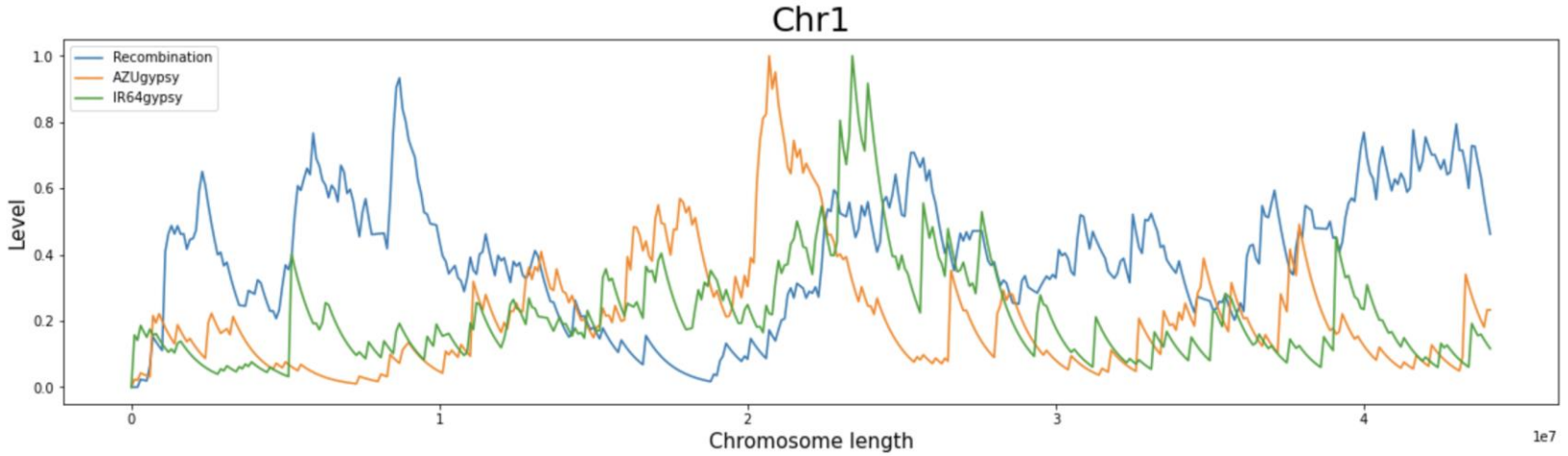
anr agence nationale de la recherche
AU SERVICE DE LA SCIENCE

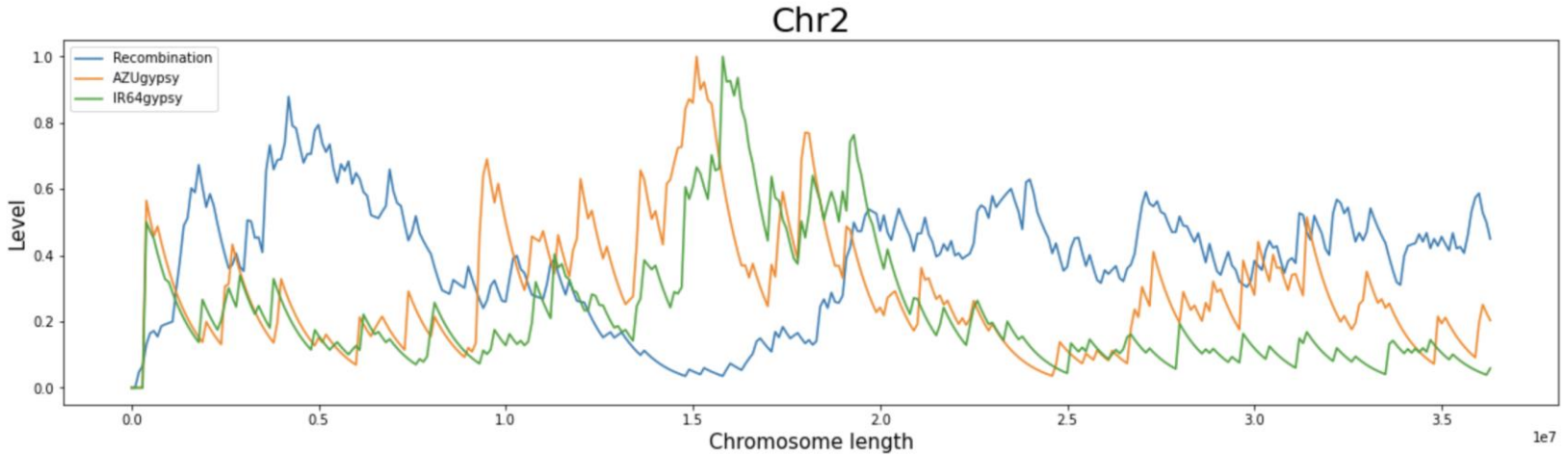
WP3

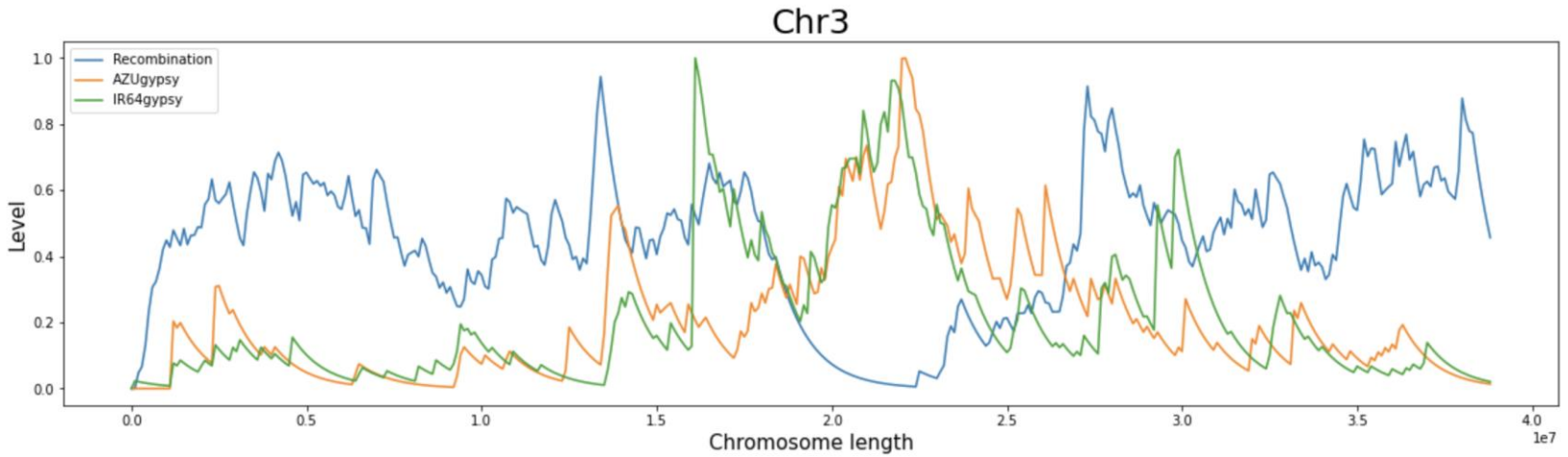
- A *new* population of **2,000 F_2** individuals from the same cross [IR64 × Azucena]
  - →10x more precise estimates of local recombination
  - →Direct estimation of the F_1 recombination rate
- 2x WGS for **each single** F_2 individual genome
- NOISYmputer v. 0.5 → Julia code jupyter{book}
- Look at motifs, GC3, etc 
- Methylation
- TEs, genes, duplications
- Better methods to calculate genome “similarity” (SyRI)
- Try AI for model construction
- Cross-validation
  - between chromosomes
  - between 10 Nested-Association Mapping populations

Yale

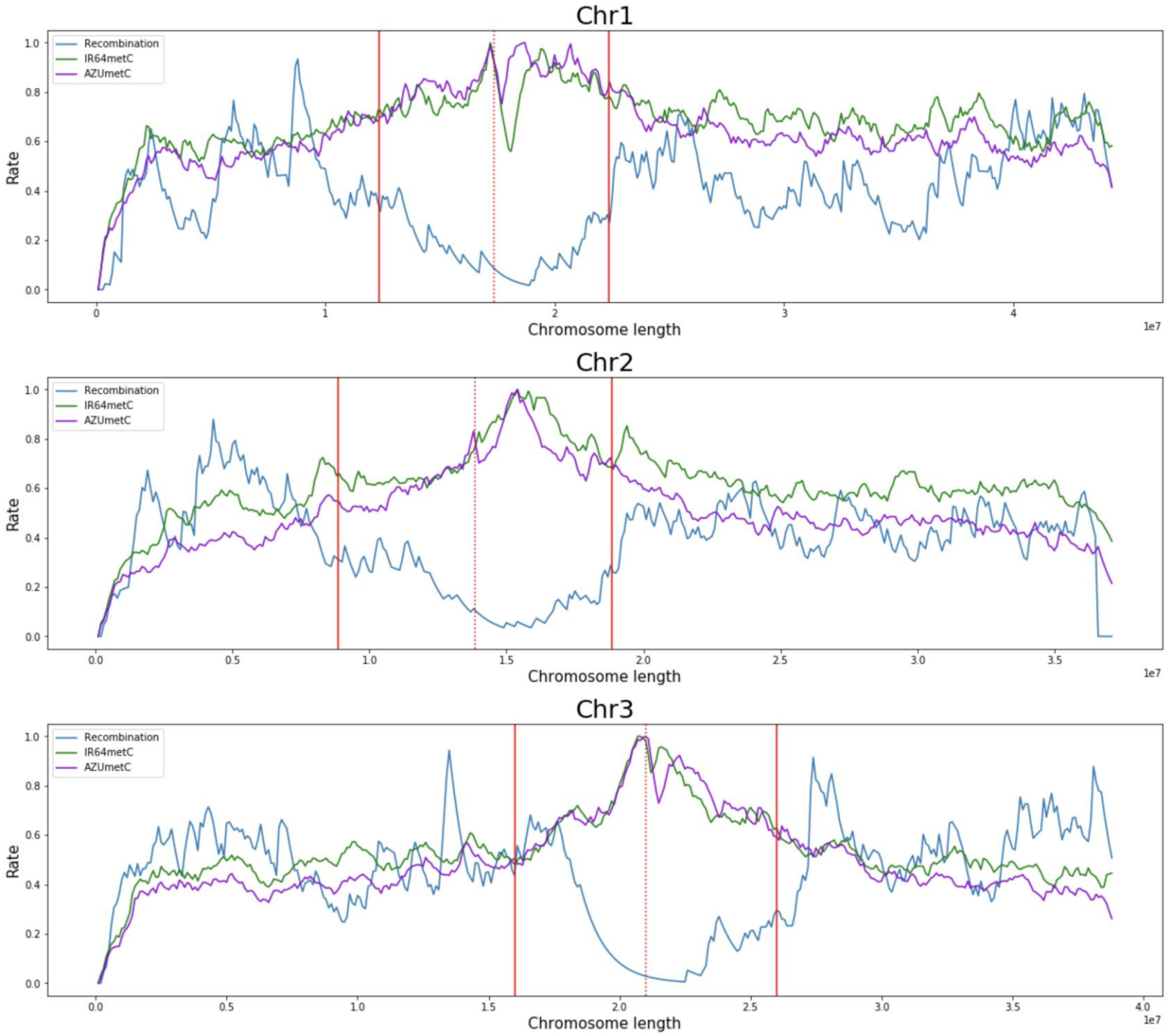
Gypsy elements vs. recombination







Methylation vs. recombination



A possible application

Predict recombination for all possible pairs of the 3,000 rice genomes

→ Breeders can choose the best predicted crosses (recombination in QTLs)

- Requires genome structure reconstruction from Illumina

LANDSREC: The high-resolution landscapes in rice meiotic recombination

*** Postdoc position at IRD, Montpellier ***

Mainly bioinformatics & AI

3 years, starts ~Feb 2022

mathias.lorieux@ird.fr

Many thanks to:

- Arizona Genomics Intitute
  - Rod Wing
  - Jianwei Zhang
  - Dave Kudrna
- Genoscope
  - Karine Labadie
- IRD
  - François Sabot
  - Christine Tranchant
  - Romain Guyot
- Yale University
  - Stephen Dellaporta
- The University of Sheffield
  - Julie Scholes

Funding:

- The RICE CRP
- France Génomique
- AGI
- BBSRC
- World bank - OMICAS
- USAID

*** Postdoc position at IRD, Montpellier ***

Mainly bioinformatics & AI

3 years, starts ~Feb 2022

mathias.lorieux@ird.fr