

# Supplementary materials

## Contents

<b>1 Supplementary methods</b>	<b>2</b>
1.1 Bioinformatics workflow . . . . .	2
1.1.1 Schema . . . . .	2
1.1.2 Alignment, filtering and count: Pipeline single-meiosis.sh . . . . .	2
1.1.3 Cross with SNV: variantutils R functions [author: Charif D.] . . . . .	2
1.1.4 Individuals and Tetrad genotyping: HMM_model R functions [author: Robin S.] . . . . .	2
1.1.4.1 Individual level: . . . . .	3
1.1.4.2 Tetrade level . . . . .	3
1.1.4.3 HMM parameters . . . . .	3
1.1.4.4 Classification: . . . . .	4
1.1.4.5 Results: . . . . .	4
1.1.5 COs detection and classification: crossover programm . . . . .	4
1.1.6 NCOs detection: . . . . .	4
1.1.7 Tetrad analysis iteration . . . . .	4
1.1.8 Gold variants, CO and NCO annotation . . . . .	4
1.1.9 Track Length estimation . . . . .	4
1.1.10 Detection power . . . . .	5
<b>2 Supplementary Results</b>	<b>5</b>
2.1 Overview of sequencing data: . . . . .	5
2.2 Sensibility and specificity of the overall strategy . . . . .	5
2.2.1 The Score is highly correlated to the tetrads sequencing depth . . . . .	5
2.3 Gold SNPs descriptiv statistics: . . . . .	6
2.4 Tetrads Genotypes: . . . . .	7
2.5 Distribution of detected CO and NCO along the genome: . . . . .	12
2.6 Number of detected events per tetrads and genotypes . . . . .	14
2.7 Number of events and converted SNP per Tetrads . . . . .	15
2.8 Track Length analysis . . . . .	15
2.8.1 Descriptiv statistics: . . . . .	15
2.8.1.1 NCO . . . . .	15
2.8.1.2 CO . . . . .	15
2.8.1.3 Comparison of CO and NCO TL within Genotype . . . . .	15
2.8.2 Comparaison of the CO TL: between wt and each genotype . . . . .	16

2.8.3 Comparaison of the NCO TL: between wt and each genotype . . . . .	17
2.9 Distribution of CO and NCO in features . . . . .	17
2.10 nb SNP annot . . . . .	17
2.11 nb Event Annot . . . . .	17
2.12 Distribution of SNP in features . . . . .	17
2.13 Probability of conversion detection . . . . .	18
<b>3 Bibliographie</b>	<b>19</b>
<b>4 Session info</b>	<b>19</b>

# 1 Supplementary methods

## 1.1 Bioinformatics workflow

### 1.1.1 Schema

### 1.1.2 Alignment, filtering and count: Pipeline single-meiosis.sh

Tetrad's analysis was carried out by applying the overall strategy proposed by Anderson (2011) and implemented in the single-meiosis.sh pipeline avalaibale at .

- For each tetrads, sequences from the four individuals were mapped against a merged reference genome containing both the Columbia (TAIR10) and Landsberg (Zapata et al. (2016)) sequences. BWA (Li and Durbin (2009)) was used to align sequences with a set of parameters positioned to neither allow mismatches (or gap) nor multiple hits.
- Alignment were filtered out to remove non specific hit and PCR duplicates. Such parameters ensure only the maintenance of reads that are specific to a parental genome but also specific to a location within a genome. This parameters discards reads that span conversion but allowing mismatches introduced two much noises in the results. This settings was the best compromise between specificity and sensitivity.
- Each alignment file was then separated by parental genomes: Col-0 and Ler-0 to compute the number of reads covering each genome position thanks to the samtools mpileup programm.

### 1.1.3 Cross with SNV: variantutils R functions [author: Charif D.]

VCF files were crossed with the allelique SNVs list between Col-0 and Ler-0 provided by Zapata et al. (2016) (TabOfSNP.txt) and serve as input to the HMM.

### 1.1.4 Individuals and Tetrad genotyping: HMM\_model R functions [author: Robin S.]

An HMM model was implemented by S. Robin to obtain both the individuals and tetrad genotypes at each SNV position from the 4 count files (M1, M2, M3, M4) of a Tetrad.

Notations:

- $I = 4$  individuals forming a tetrade ( $i = 1 \dots I$ );
- $C = 5$  chromosomes per individuals ( $c = 1 \dots C$ );
- $T$  markers ( $T_c$  markers in chromosomes  $c$ :  $\sum_c n_c = n$ ,  $t = 1 \dots n_c$ );
- $R_{ict}$  = number of reads mapped onto marker  $t$  in the Col + Ler genome from chromosome  $c$  in individual  $i$ ;
- $Y_{ict}$  = number of reads mapped onto marker  $t$  in the Col genome from chromosome  $c$  in individual  $i$ .

**1.1.4.1 Individual level:** Because of the experimental setting, at every marker, the ratio  $Y_{it}/R_{it}$  is expected to be close to either:

- 1/2 (heterozygous genotype Col/Ler) or
- 1 (homozygous genotype Col)

This suggest a model with  $K = 2$  hidden states with respective emission distributions:

- $Y_{it} \sim \mathcal{B}(R_{it}, \gamma_1)$ , where  $\gamma_1 \simeq 1/2$ ;
- $Y_{it} \sim \mathcal{B}(R_{it}, \gamma_2)$ , where  $\gamma_2 \simeq 1$ .

The first goal is to recover for each individual at each marker the hidden state

$$Z_{it} \in \{1, 2\}$$

**1.1.4.2 Tetrade level** Tetrade genotype is determined by the states of the four individuals at a given marker. Therefore, we consider the joint hidden state  $S_t \in \{1, \dots, K^I\} = \{1, \dots, 16\}$  for the whole tetrade at position  $t$ , which is related to the four individual hidden states  $Z_{it}$  according to Table. 1. In the same table  $H_t$  is the number of heterozygous individuals at marker  $t$ . These states can be interpreted as follows:

- States  $S_t = 1$  and  $16$  ( $H_t = 4$  and  $0$ ) are expected to be absent;
- States  $S_t = 4, 6, 7, 10, 11$  and  $13$  ( $H_t = 2$ ) are expected to be predominant;
- States  $S_t = 2, 3, 5, 8, 9, 12, 14$  and  $15$  ( $H_t = 1$  or  $3$ ) correspond to NCO.

$S_t$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$Z_{1t}$	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
$Z_{2t}$	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
$Z_{3t}$	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
$Z_{4t}$	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
$H_t$	4	3	3	2	3	2	2	1	3	2	2	2	1	2	1	1

Table 1: Correspondence between the tetrade hidden state  $S_t$  and each of the four individual hidden states  $Z_{it}$ , as defined by the R instruction: `expand.grid((1:2),(1:2),(1:2),(1:2))`.

We used the hidden Markov model (HMM) framework where:

- the hidden states  $Z_{it}$  or  $S_t$  are supposed to form a Markov chain and where
- the observations  $Y_{it}$  are supposed to be independent conditionally on the hidden states with distribution controlled by the corresponding hidden state.

#### 1.1.4.3 HMM parameters

- The hidden states of each individuals have been considered jointly. Both the success probabilities  $(\gamma_1, \gamma_2)$  and the transition rates  $(\lambda, \mu)$  are common to all individuals, resulting in:

$$\hat{\gamma}_1, \hat{\gamma}_2, \hat{\lambda}, \hat{\mu} \rightarrow 4 \text{ parameters.}$$

- To take into account the non homogeneous distance separating the marker, we allowed heterogeneous transition probabilities (continuous time Markov chain). The frequencies of changes from one states to another are given by transition rates

$$\lambda_{k\ell} : k \rightarrow \ell.$$

In the binary case ( $Z \in \{1, 2\}$ ), denoting

$$\lambda : 1 \rightarrow 2, \quad \mu : 2 \rightarrow 1,$$

**1.1.4.4 Classification:** The inference on the hidden states is carried based on their conditional distribution given the observations  $Y = (Y_{it})$ , that is

$$p_{\hat{\theta}}(Z|Y).$$

The classification (that is the inference of the hidden state) had been made marker by marker. For each individual and each marker, we obtained the  $\{posterior\ probability\}$  provided by the EM alorithm.

$$\tau_{itk} = P_{\hat{\theta}}\{Z_{it} = k|Y\}.$$

A natural classification rule consists in using the  $\{maximum\ a\ posteriori\}$  (MAP) rule:

$$MAP_{it} = \arg \max_k \tau_{itk}.$$

#### 1.1.4.5 Results:

- Chromosom
- Position
- MAP
- $-\log_{10}(1 - MAP_{it})$
- MAP classification

#### 1.1.5 COs detection and classification: crossover programm

Thanks to the marker genotype obtained by the HMM model, we then used the crossover programm Anderson (2011) to detect and classify the CO event. Two COs could be detected if they are at least distant by 5kb.

#### 1.1.6 NCOs detection:

NCOs correspond to markers with the following states  $St = \{2, 3, 5, 8, 9, 12, 14, 15\}$ . Successiv markers with a genotype corresponding to one of this states and with a maximum distance of 2 Kb between each belong to the same event.

#### 1.1.7 Tetrads analysis iteration

1. A first analysis of all tetrads was done considering all of the SNV markers from Zapata et al. (2016) (All\_SNV=545481)
2. SNVs where then filtering according to the followig criterions:
  - Removing of SNVs associated to COs that are found in several tetrads (file Liste\_FAUX\_CO.txt)
  - Removing of markers with non 2:2 (fichier df\_Genotype\_SNP\_count.txt)
3. Re-analysis of all Tetrads with a reducing list of SNV (Gold\_Variants=522658).
4. Sensibility and specificity analysis (?)

#### 1.1.8 Gold variants, CO and NCO annotation

SNP have been annotated thanks to TxDb.Athaliana.BioMart.plantsmart28 database and the locateVariants function from the VariantAnnotation package: Coding, Intron, FiveUTR, ThreeUTR, Promoter up to 1000 bases.

Regarding, heterochromatine, centromere and transposable elements, we used a bed file from personal communication with Dr. Mezard.

#### 1.1.9 Track Length estimation

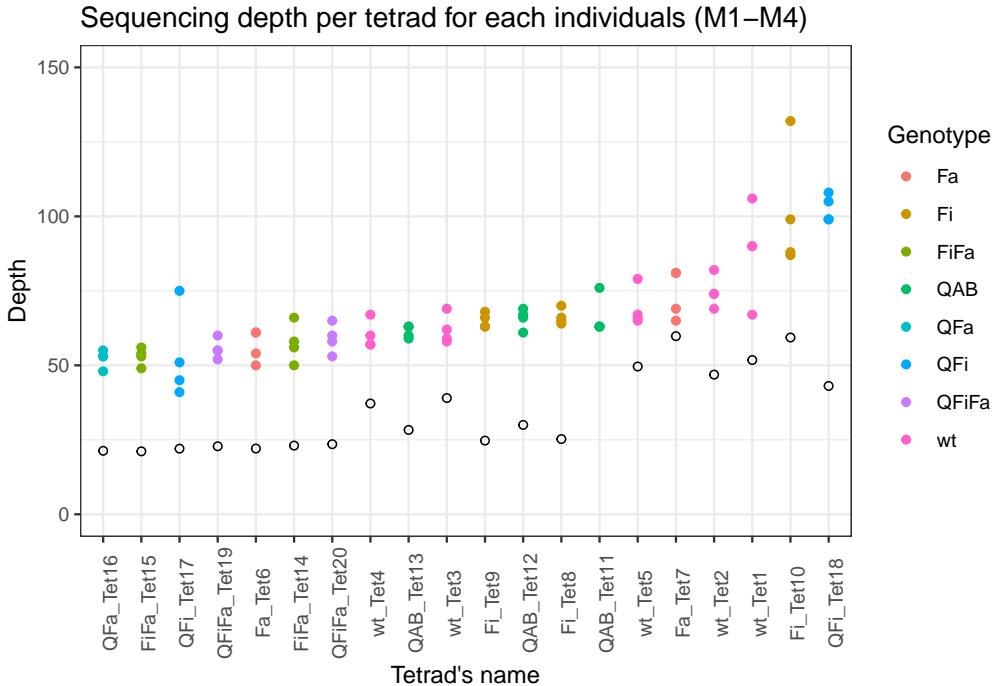
- median TL:
- min TL: between the last converted SNPs
- max TL: between the 2 flanking SNPs

### 1.1.10 Detection power

- For each window size (20, 50, 75, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000, 2500, 3000, 4000, 5000, 8000, 10000)
  - 10000 sequences of this length were sampled from the genome.
  - We computed the following statistics
    - \* mean converted SNP
    - \* number of windows with at least one converted SNP

## 2 Supplementary Results

### 2.1 Overview of sequencing data:



```
##  
##   Fa    Fi   FiFa    QAB    QFa    QFi  QFiFa     wt  
##   2     3     2     2     3     1     2     2     5
```

### 2.2 Sensibility and specificity of the overall strategy

#### 2.2.1 The Score is highly correlated to the tetrads sequencing depth

The score is positively correlated to the rate of true positive NCO

More the tetrad's depth is important, less the number of events to check is high and the highest is the rate of TP NCO.

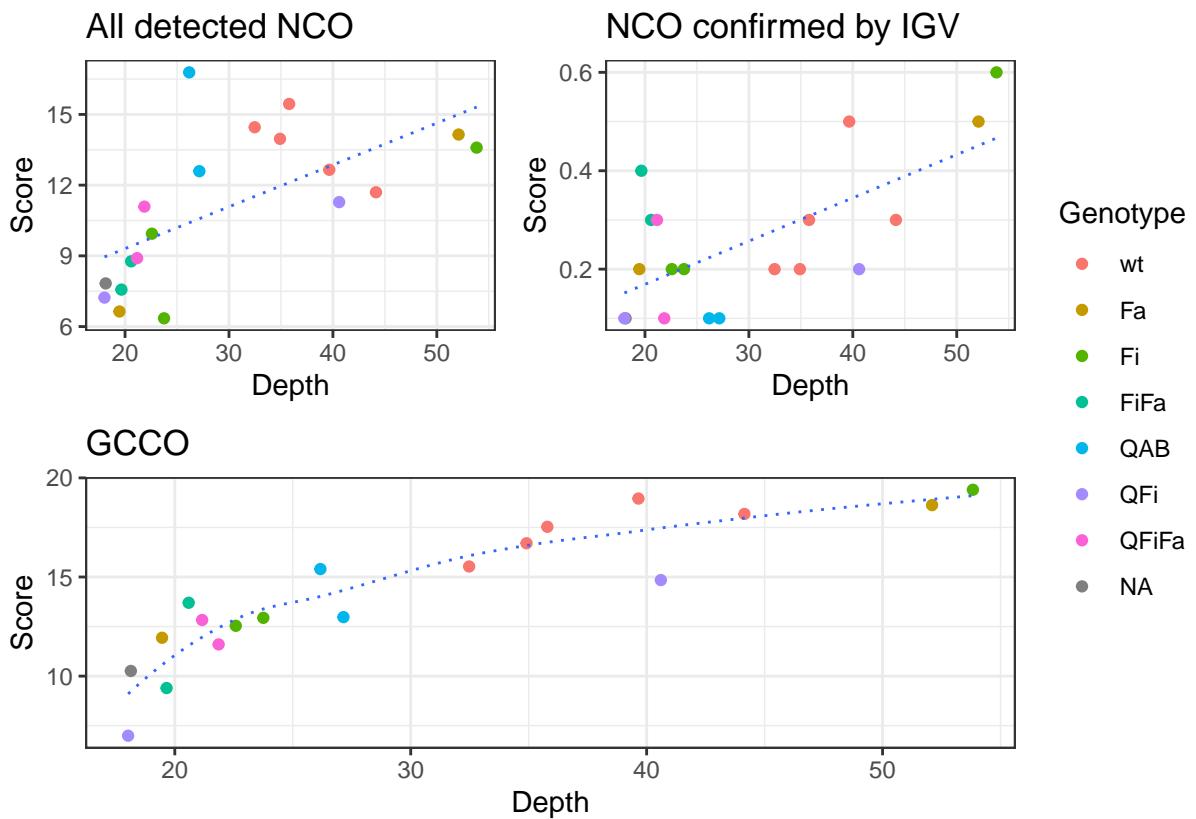
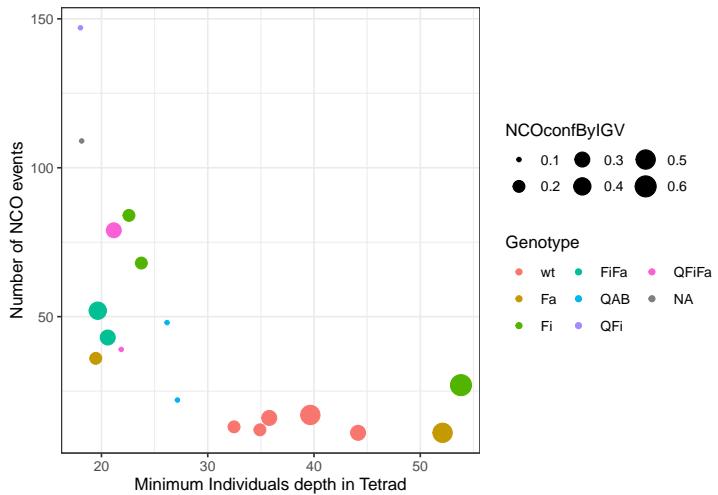


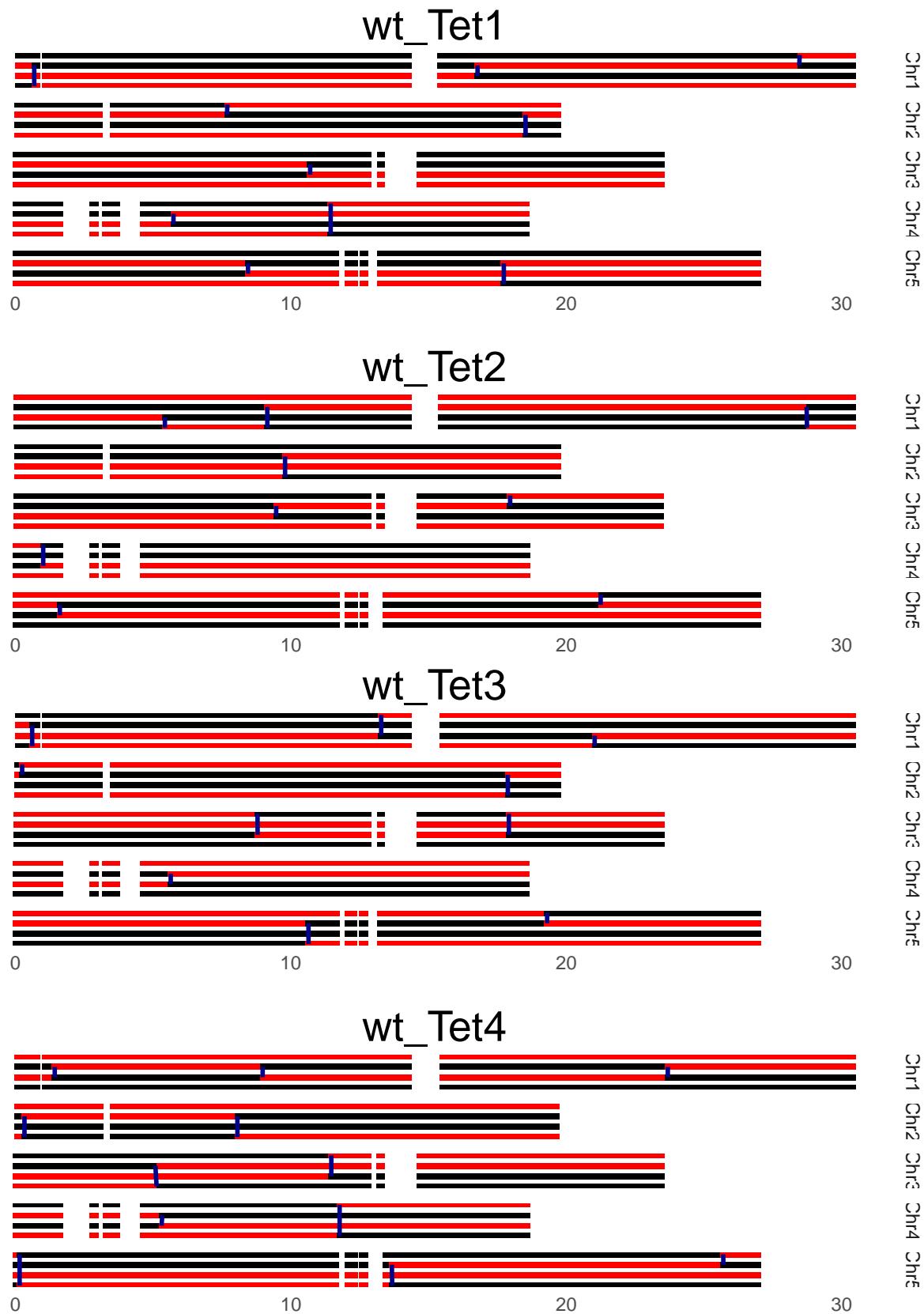
Figure 1: Supp Fig 2



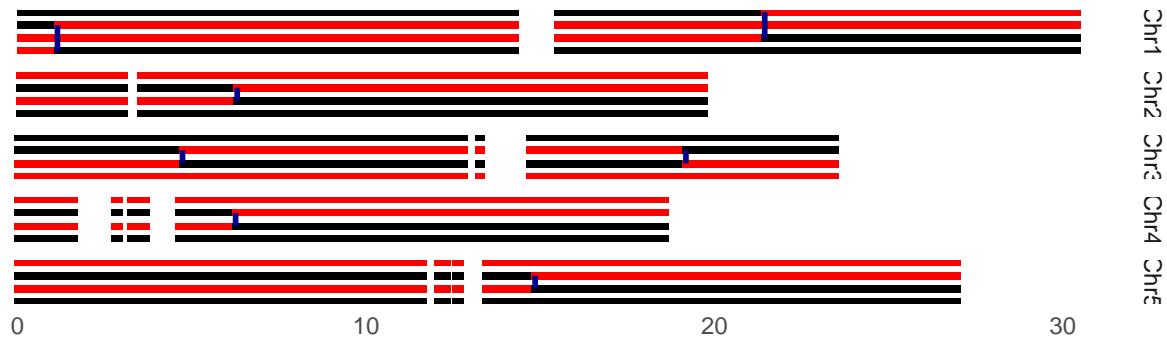
## 2.3 Gold SNPs descriptiv statistics:

```
## # A tibble: 5 x 4
##   Chr     max   mean median
##   <chr> <int> <dbl>  <dbl>
## 1 Chr1    996316 220.    43
## 2 Chr2    459751 208.    50
## 3 Chr3   1376607 233.    48
## 4 Chr4   1170202 231.    45
## 5 Chr5    519179 204.    43
```

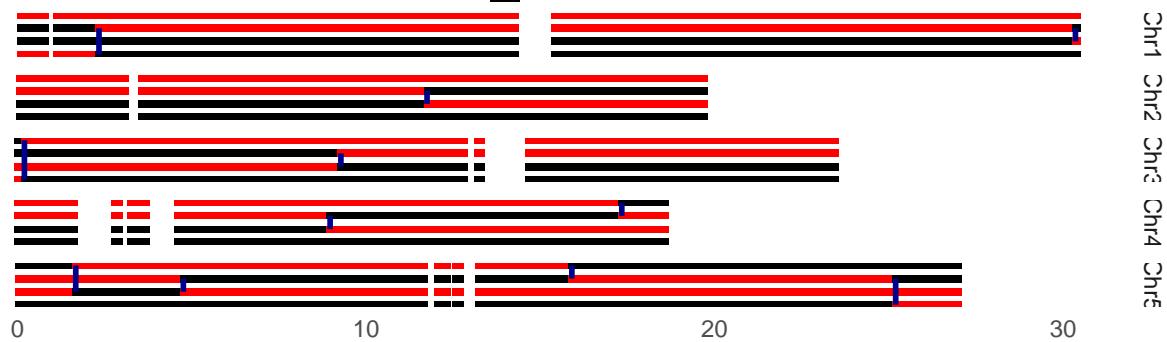
## 2.4 Tetrads Genotypes:



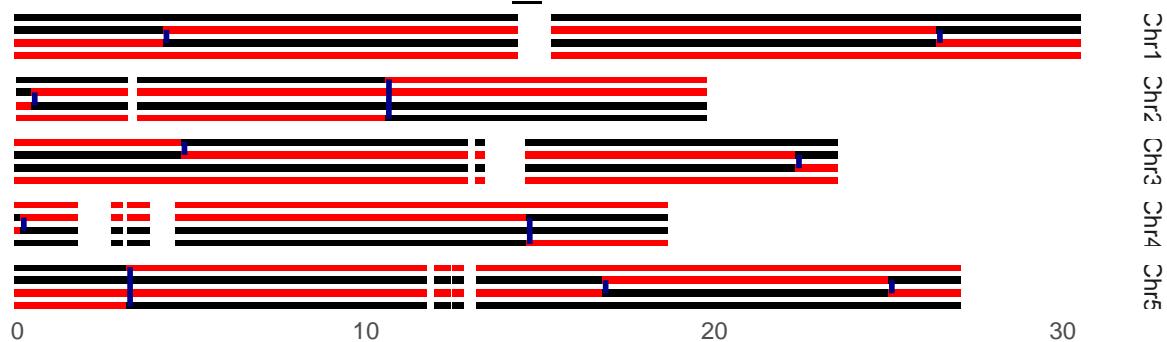
**wt\_Tet5**



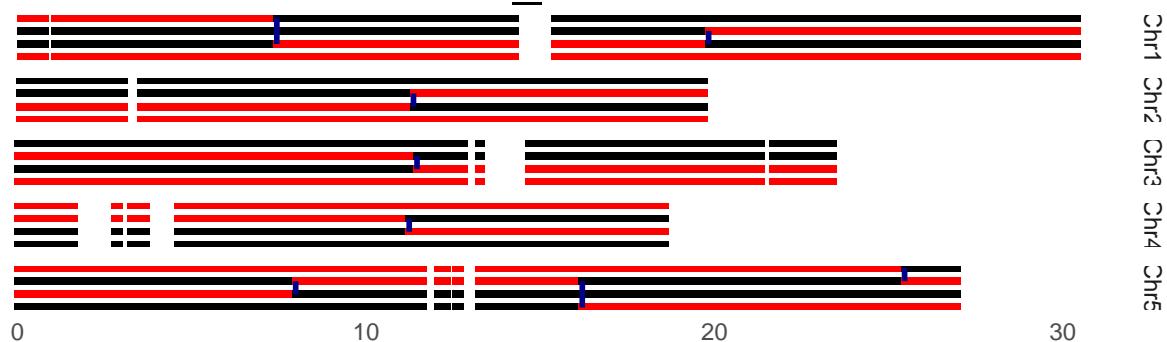
**Fi\_Tet10**



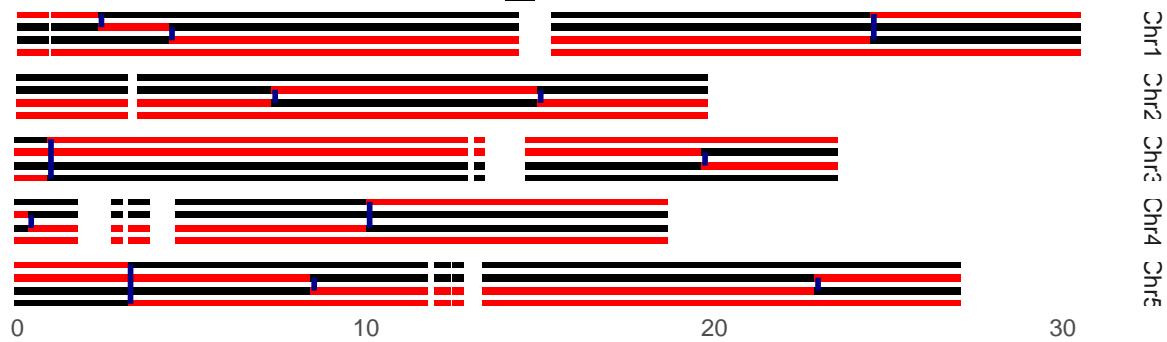
**Fa\_Tet6**



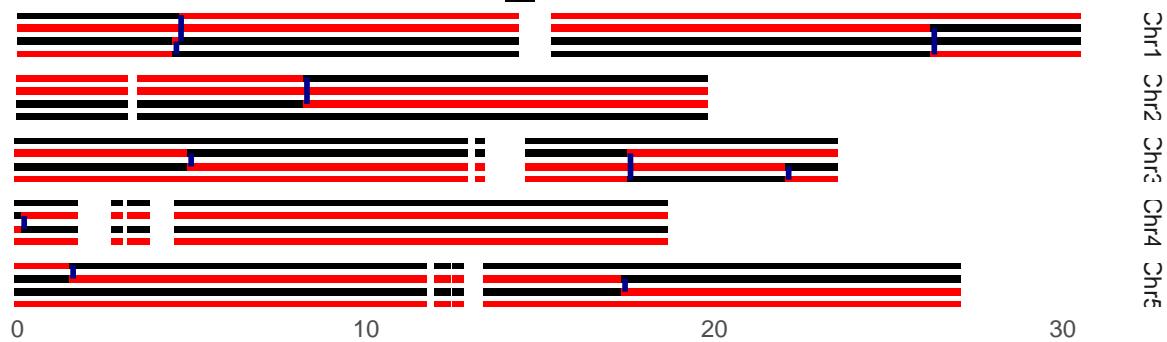
**Fa\_Tet7**



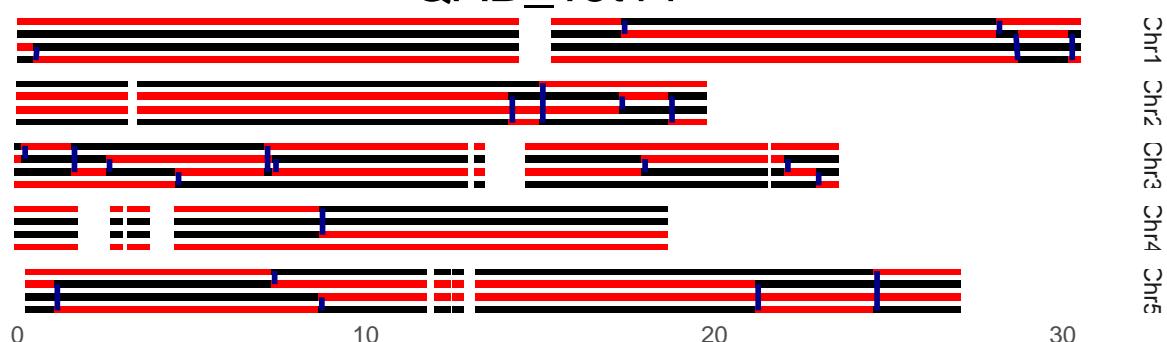
### Fi\_Tet8



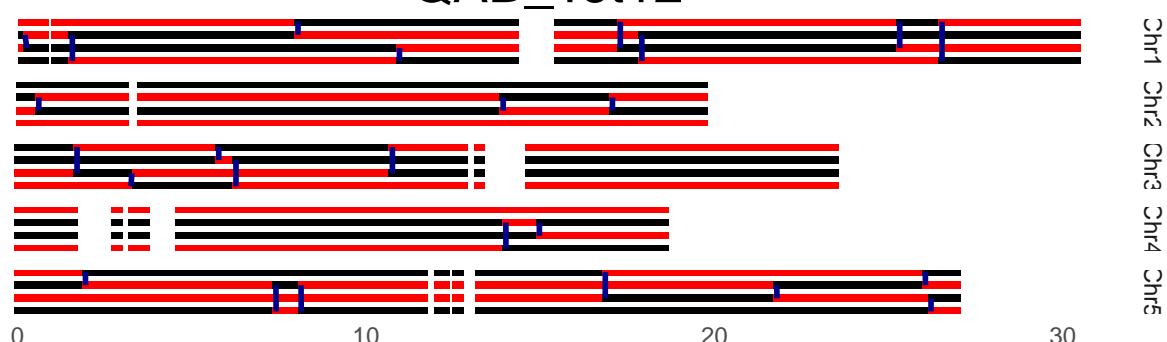
### Fi\_Tet9



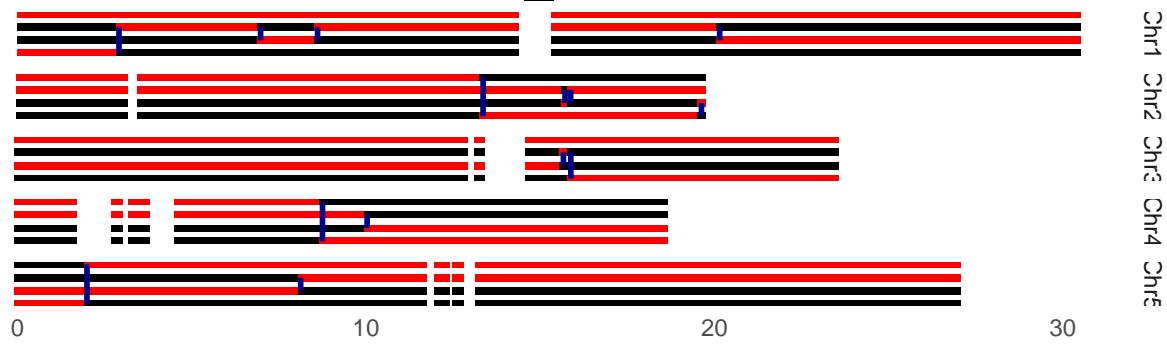
### QAB\_Tet11



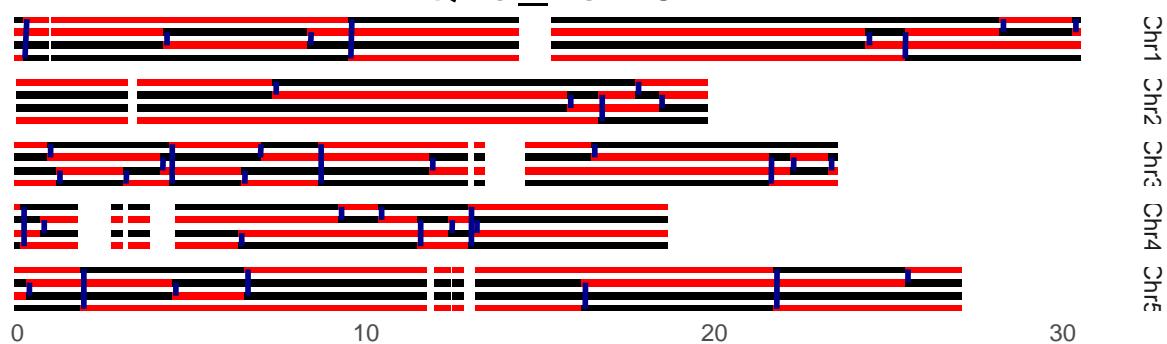
### QAB\_Tet12



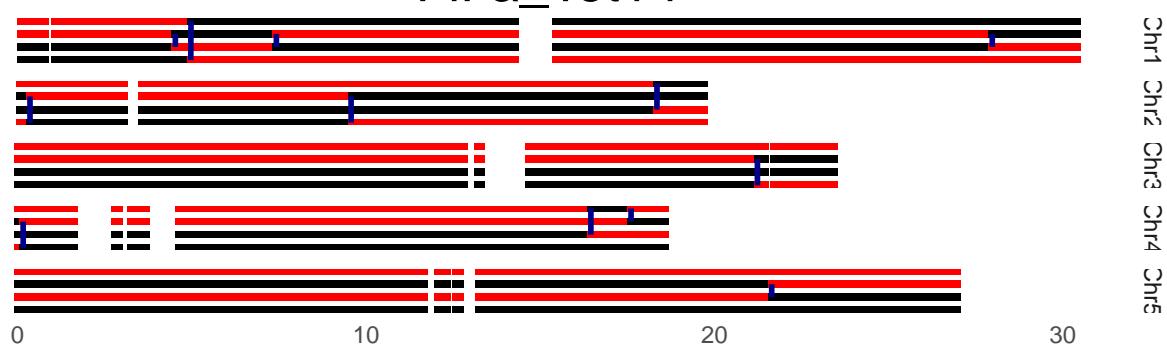
### QAB\_Tet13



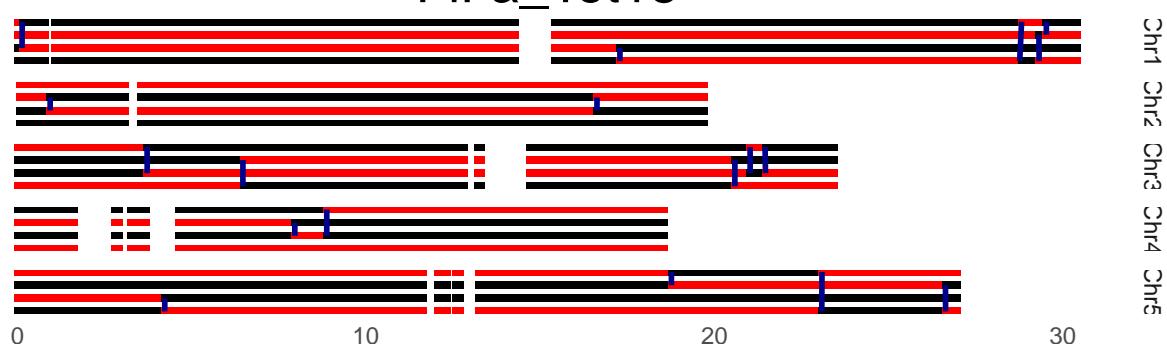
### QFa\_Tet16



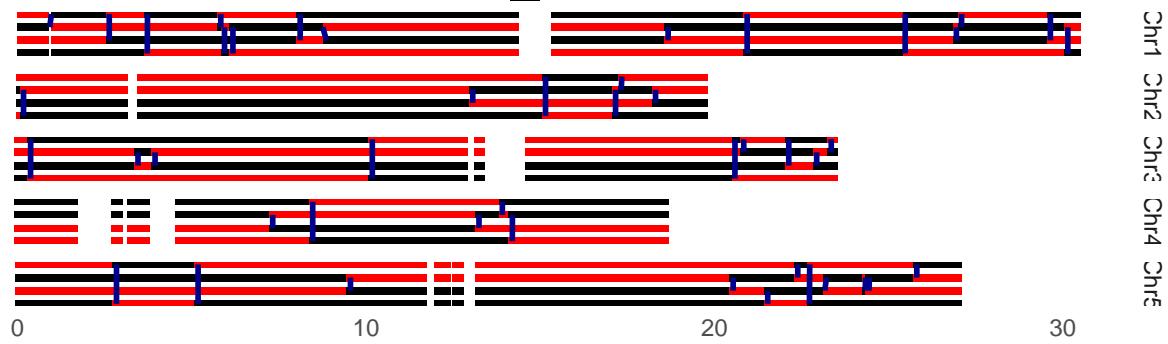
### FiFa\_Tet14



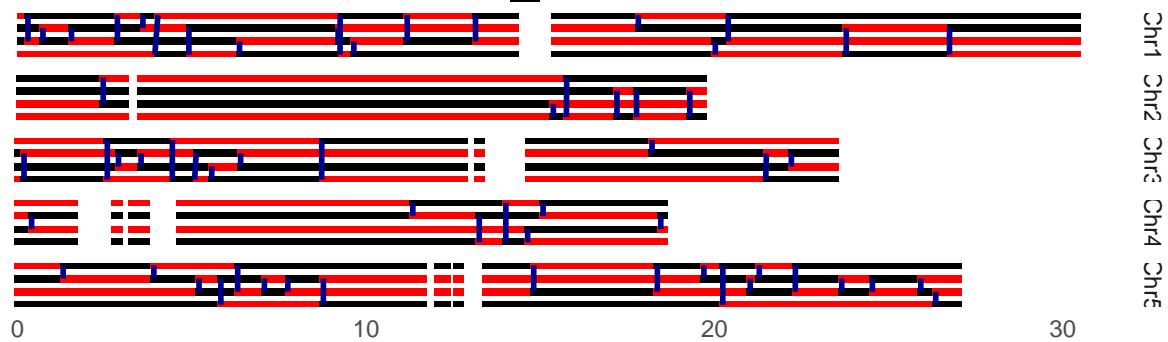
### FiFa\_Tet15



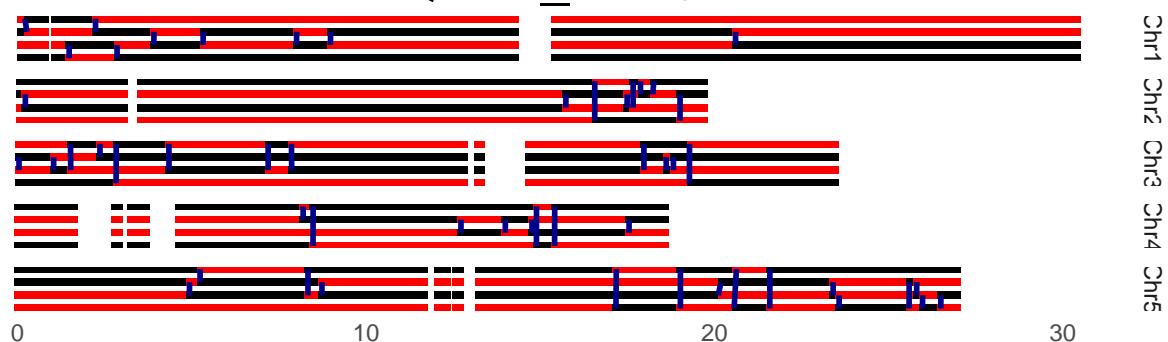
### QFi\_Tet17



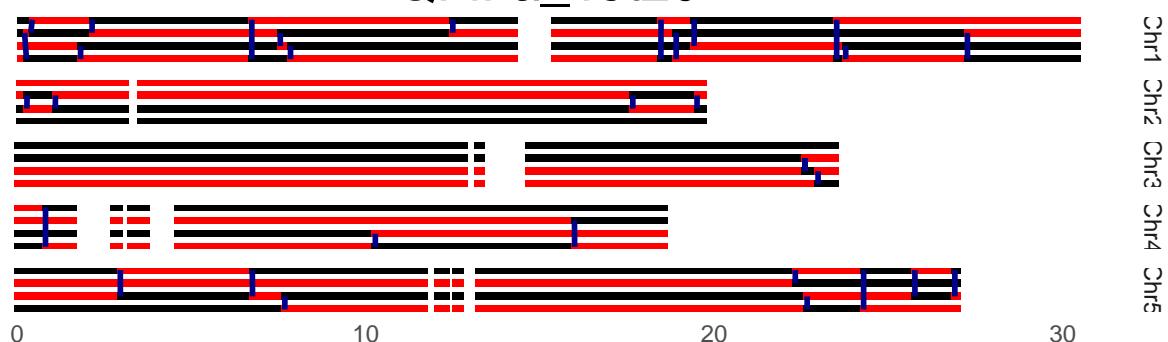
### QFi\_Tet18



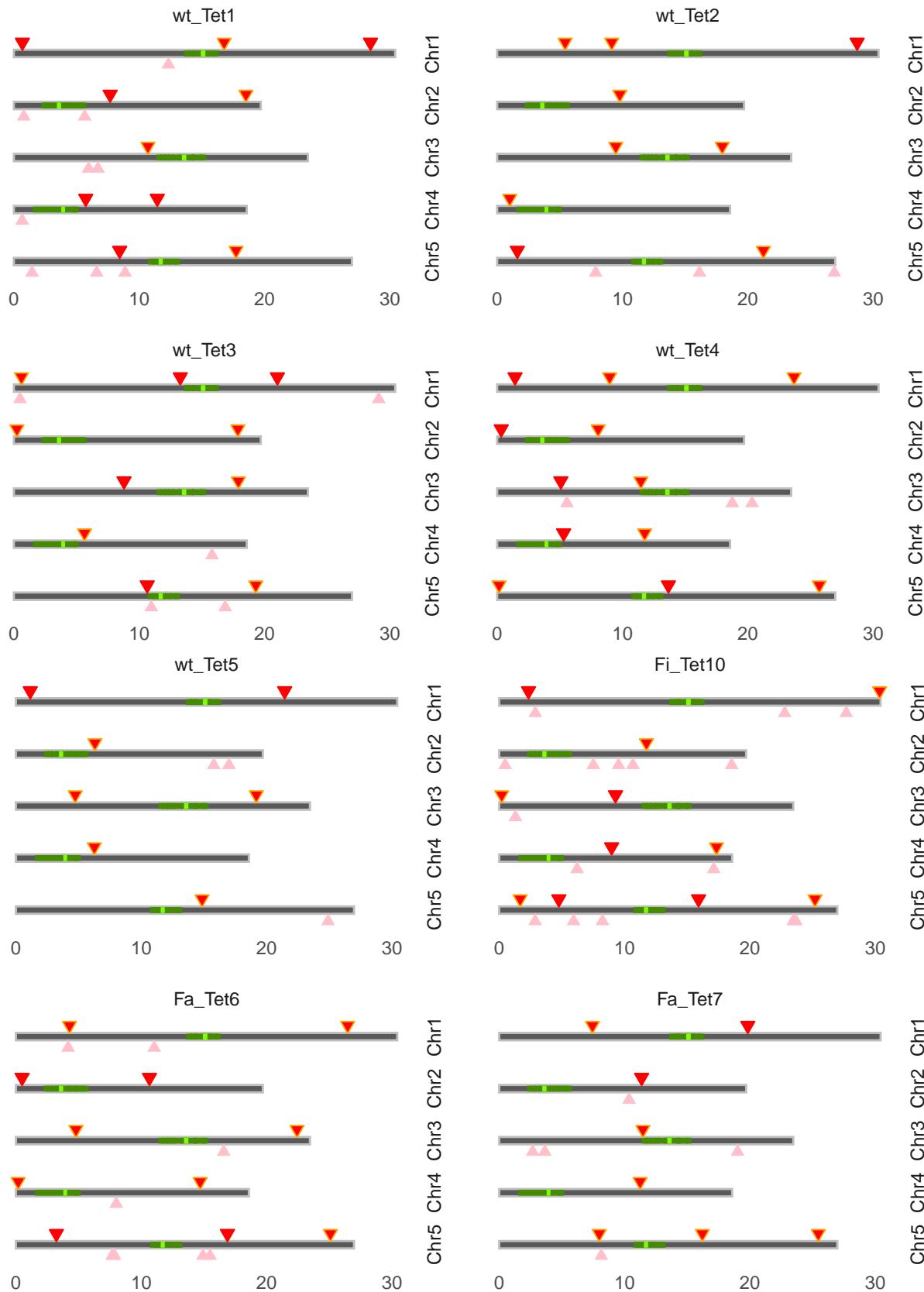
### QFiFa\_Tet19

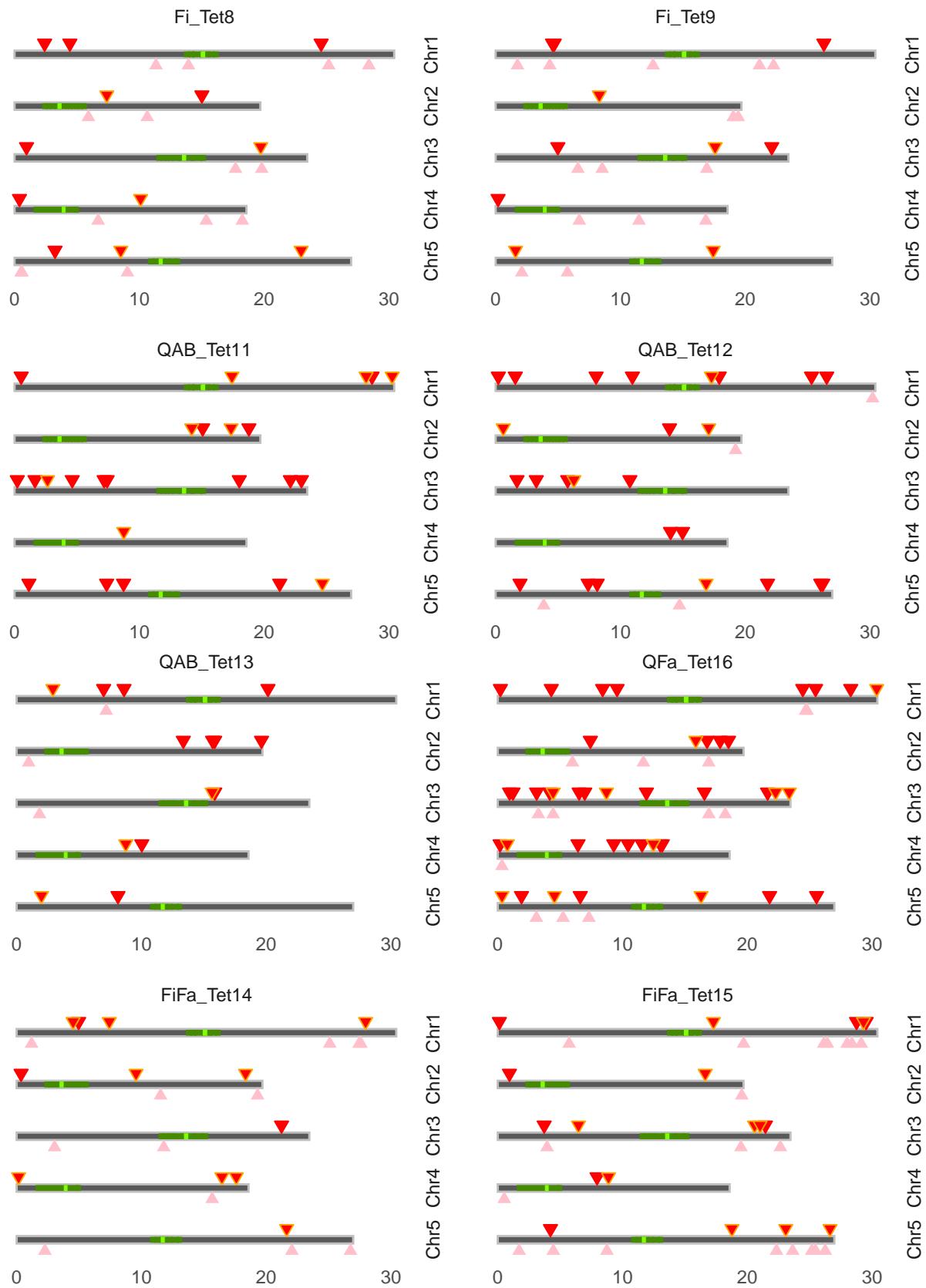


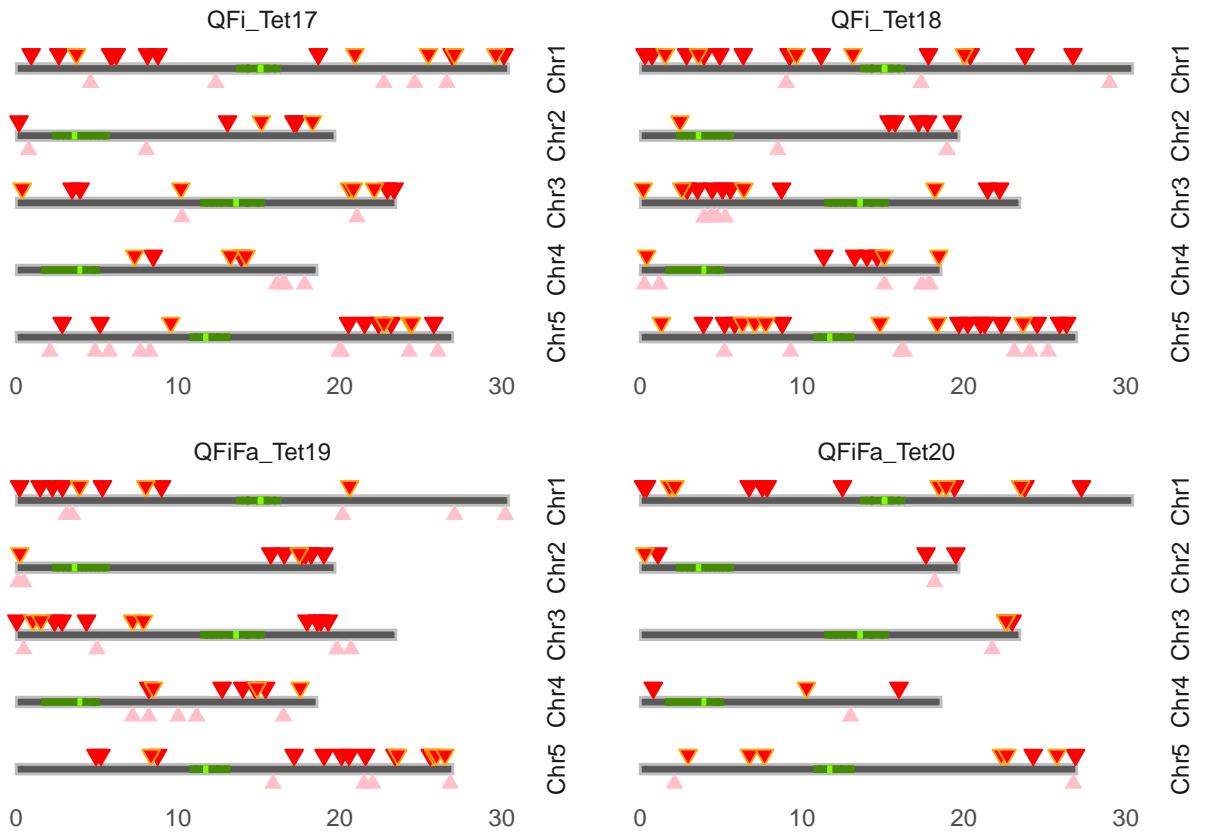
### QFiFa\_Tet20



## 2.5 Distribution of detected CO and NCO along the genome:







## 2.6 Number of detected events per tetrads and genotypes

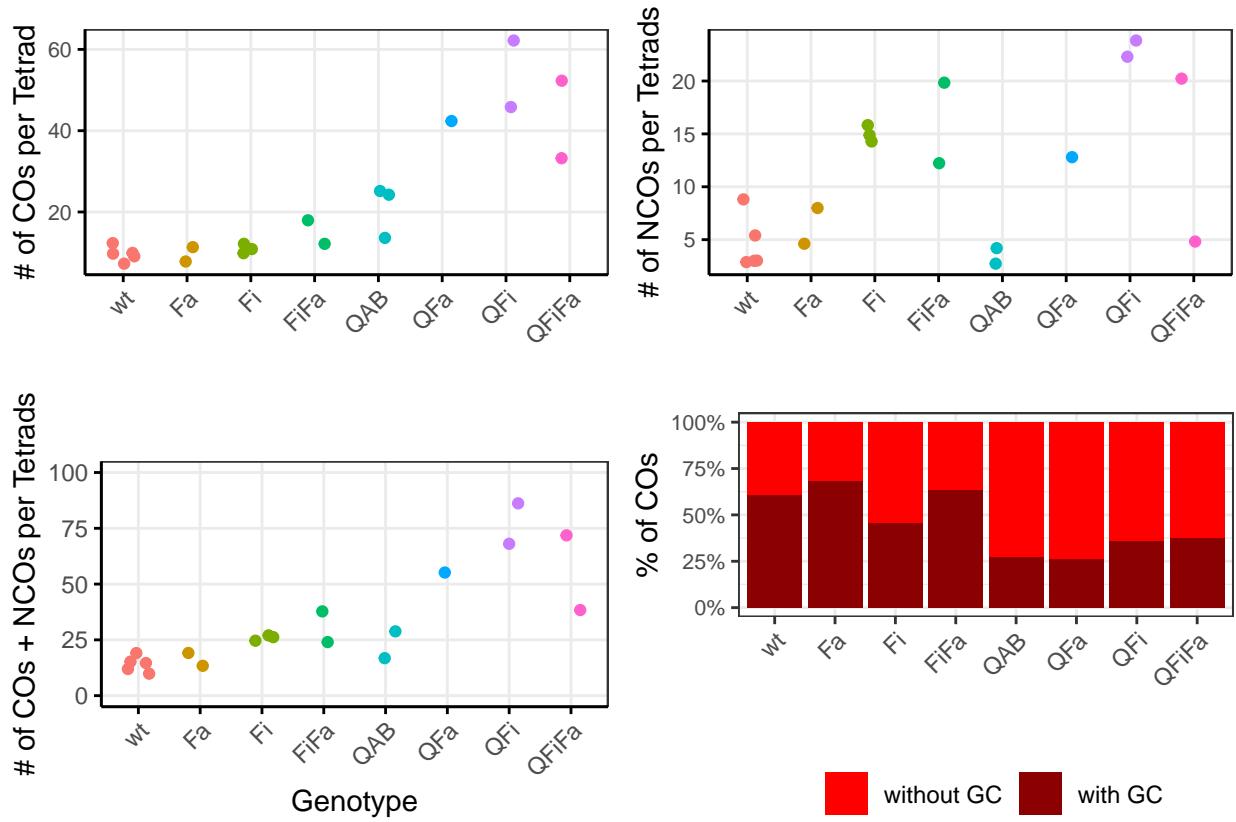


Figure 2: fig2

```

## TableGrob (2 x 2) "arrange": 4 grobs
##  z   cells   name      grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
## 3 3 (2-2,1-1) arrange gtable[layout]
## 4 4 (2-2,2-2) arrange gtable[layout]

```

## 2.7 Number of events and converted SNP per Tetrad

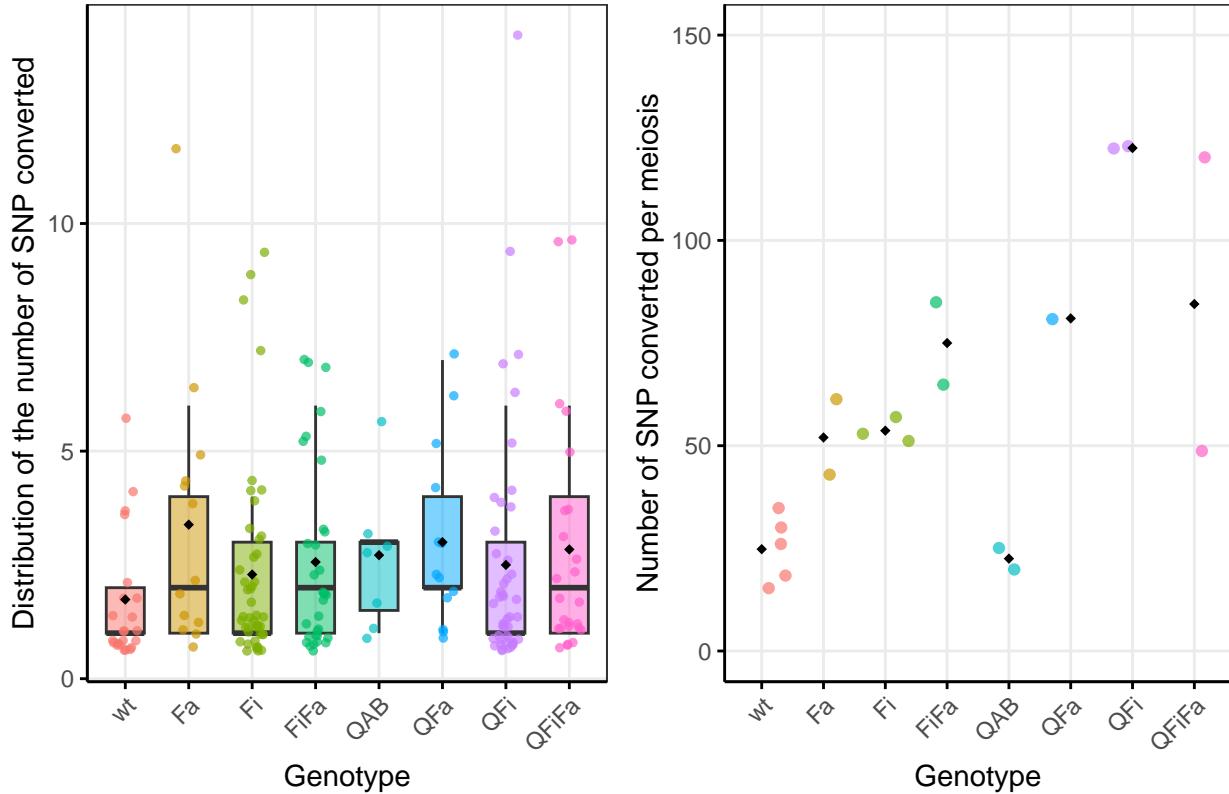


Figure 3: Supp Fig4

## 2.8 Track Length analysis

Dans l'analyse précédente, on avait pris cette décision:

When the two SNPs flanking the conversion events are at a cumulative distance greater than 1 kb (ie twice the distance separating on average 2 snps) then we consider that the estimate of the size of the conversion is not precise enough to be kept for the analysis.

### 2.8.1 Descriptive statistics:

2.8.1.1 NCO % latex table generated in R 4.4.0 by xtable 1.8-4 package % Fri Sep 20 17:18:30 2024

2.8.1.2 CO % latex table generated in R 4.4.0 by xtable 1.8-4 package % Fri Sep 20 17:18:30 2024

2.8.1.3 Comparison of CO and NCO TL within Genotype

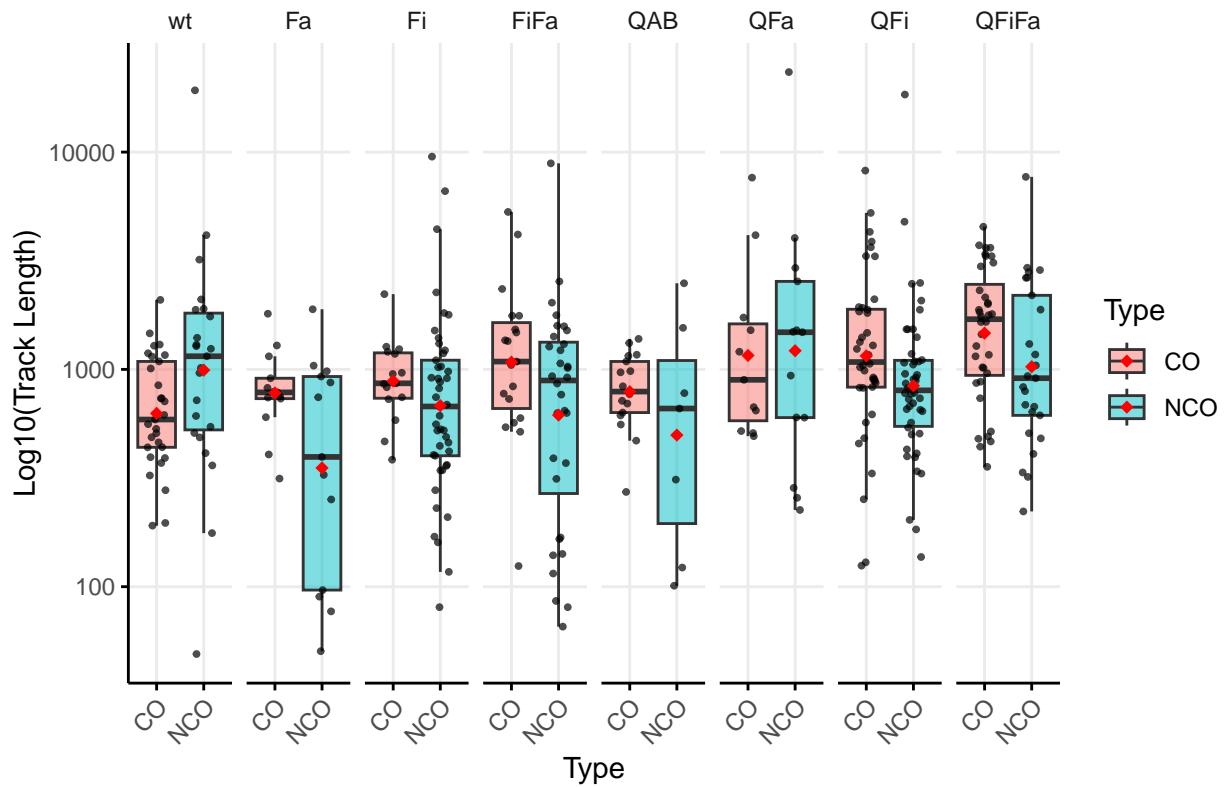
	Genotype	n	median.mean	minTL.mean	maxTL.mean
1	wt	23	2018.74	185.26	3853.22
2	Fa	13	595.65	26.00	1166.31
3	Fi	45	1149.77	83.44	2217.09
4	FiFa	32	1113.94	271.12	1957.75
5	QAB	7	859.50	89.43	1630.57
6	QFa	13	3094.85	492.38	5698.31
7	QFi	46	1370.43	215.96	2525.91
8	QFiFa	25	1501.34	532.12	2471.56

Table 2: table1

	Genotype	n	median.mean	minTL.mean	maxTL.mean
1	wt	29	742.78	255.90	1230.66
2	Fa	14	870.14	294.07	1447.21
3	Fi	15	1019.17	430.27	1609.07
4	FiFa	19	1659.61	896.68	2423.53
5	QAB	18	1406.83	738.11	2076.56
6	QFa	13	2746.00	1866.85	3626.15
7	QFi	49	1923.40	803.76	3044.04
8	QFiFa	40	2300.95	1160.45	3442.45

Table 3: table1

### Track Length by Genotype and Type



% latex table generated in R 4.4.0 by xtable 1.8-4 package % Fri Sep 20 17:18:31 2024

#### 2.8.2 Comparaison of the CO TL: between wt and each genotype

% latex table generated in R 4.4.0 by xtable 1.8-4 package % Fri Sep 20 17:18:31 2024

	Genotype	statistic	p.value	alternative	p.adjust	significant
1	Fa	110.00	0.20	two.sided	0.20	
2	Fi	416.00	0.19	two.sided	0.19	
3	FiFa	380.00	0.14	two.sided	0.14	
4	QAB	71.00	0.49	two.sided	0.49	
5	QFa	71.50	1.00	two.sided	1.00	
6	QFi	1178.00	0.03	two.sided	0.03	*
7	QFiFa	573.00	0.07	two.sided	0.07	
8	wt	222.00	0.04	two.sided	0.04	*

	Genotype	statistic	p.value	alternative	p.adjust	significant
1	Fa	143.00	0.22	two.sided	0.22	
2	Fi	143.00	0.07	two.sided	0.07	
3	QAB	185.00	0.17	two.sided	0.17	
4	FiFa	147.00	0.01	two.sided	0.01	*
5	QFa	93.00	0.04	two.sided	0.04	*
6	QFi	314.00	0.00	two.sided	0.00	*
7	QFiFa	198.00	0.00	two.sided	0.00	*

### 2.8.3 Comparaison of the NCO TL: between wt and each genotype

% latex table generated in R 4.4.0 by xtable 1.8-4 package % Fri Sep 20 17:18:31 2024

	Genotype	statistic	p.value	alternative	p.adjust	significant
1	Fa	222.00	0.02	two.sided	0.02	*
2	Fi	657.00	0.07	two.sided	0.07	
3	FiFa	446.00	0.19	two.sided	0.19	
4	QAB	104.00	0.27	two.sided	0.27	
5	QFa	138.00	0.72	two.sided	0.72	
6	QFi	613.00	0.29	two.sided	0.29	
7	QFiFa	291.00	0.95	two.sided	0.95	

Table 4: table1

## 2.9 Distribution of CO and NCO in features

### 2.10 nb SNP annot

```
##      UTR3 UTR5 Intron Coding TE Hetero
## GCCO    15   16    120     7  79      7
## NCO     2    5    29    24  50      7
```

### 2.11 nb Event Annot

```
##      UTR3 UTR5 Intron Coding TE Hetero
## GCCO    13   12     50    59  25      3
## CO      31   21     84   101  40      0
## NCO     1    4     14    10  14      4
```

### 2.12 Distribution of SNP in features

The datamart **TxDb.Athaliana.BioMart.plantsmart28** was used to annotate the SNP in the different features of the genome thanks to the locateVariants function from the VariantAnnotation package.

% latex table generated in R 4.4.0 by xtable 1.8-4 package % Fri Sep 20 17:18:41 2024

	Compartments	SumLength	NbSNP	feat.prop	snp.prop	GCCO	NCO	GCCO.event	CO.event	NCO.event
Intron	Intron	20610303	99430	16.57	24.25	120.00	29.00	50.00	84.00	14.00
Coding	Coding	51600369	97879	41.48	23.87	7.00	24.00	59.00	101.00	10.00
UTR5	UTR5	4123096	12522	3.31	3.05	16.00	5.00	12.00	21.00	4.00
UTR3	UTR3	6636022	20902	5.33	5.10	15.00	2.00	13.00	31.00	1.00
Hetero	Hetero	15711413	54382	12.63	13.26	7.00	7.00	3.00	0.00	4.00
TE	TE	24932539	124024	20.04	30.25	79.00	50.00	25.00	40.00	14.00

The number of converted SNP per features (summing up all tetrads results) is compared to the number of SNP awaited for each feature if the distribution follow the SNP distribution in features.

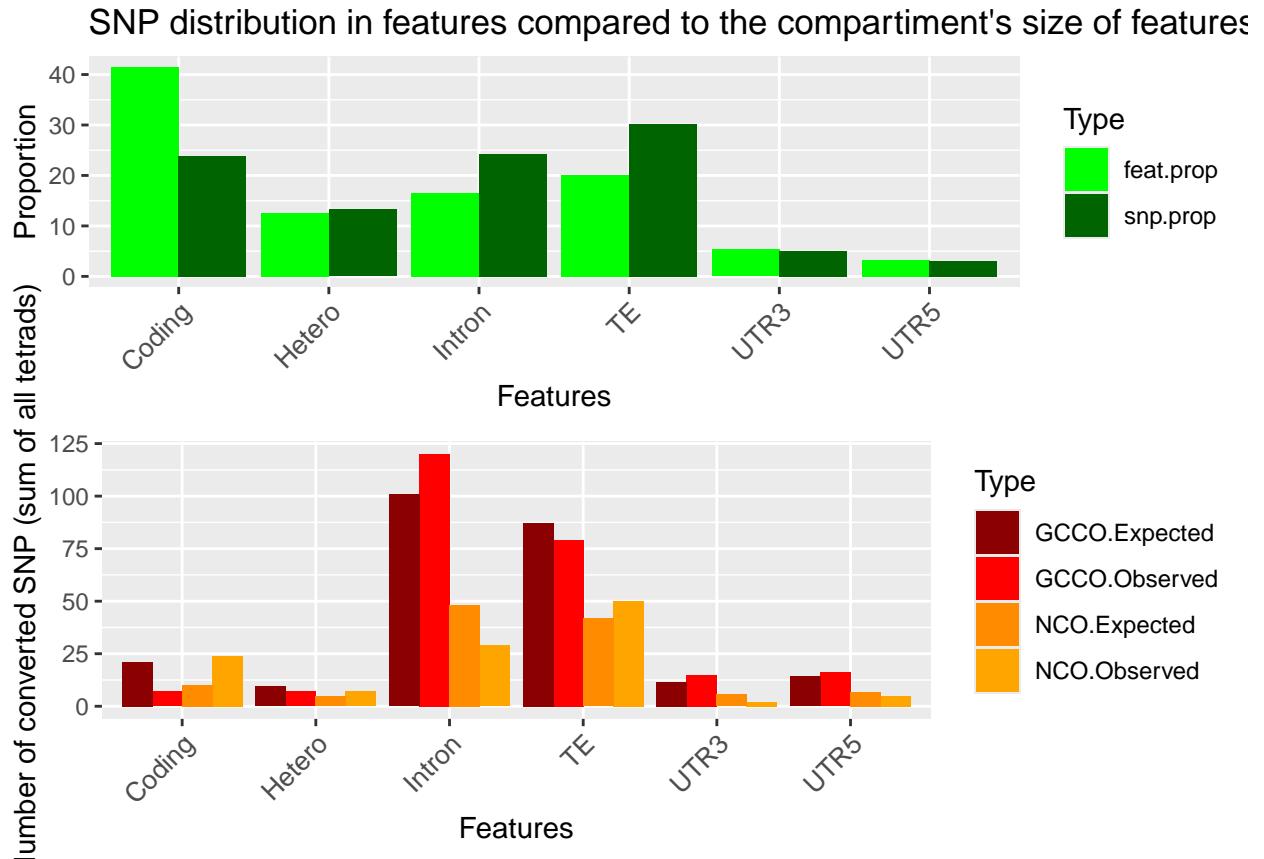
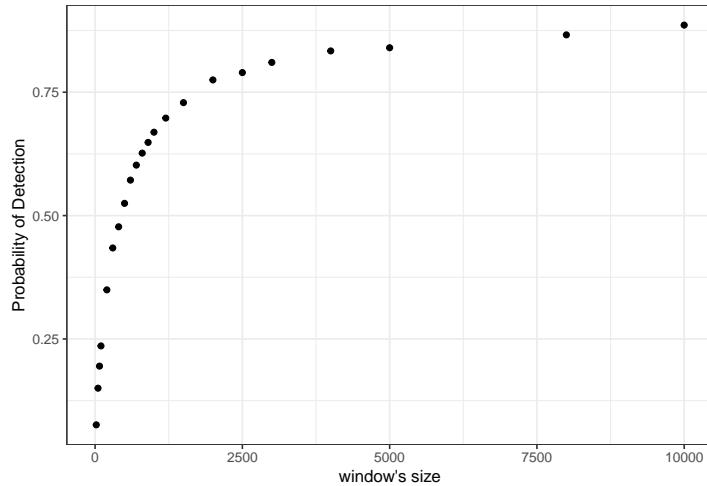


Figure 4: fig4

```
##
## Pearson's Chi-squared test
##
## data: rr[2:3, ]
## X-squared = 48.438, df = 5, p-value = 2.89e-09
```

## 2.13 Probability of conversion detection

Given the gold SNP distribution in the genome, we computed the probability of detecting a SNP in a window of a given size. We sampled 10000 windows of different sizes and computed the mean number of SNP detected and the number of windows with at least one SNP.



### 3 Bibliographie

### 4 Session info

```
sessionInfo()
```

```
## R version 4.4.0 (2024-04-24)
## Platform: aarch64-apple-darwin20
## Running under: macOS Ventura 13.1
##
## Matrix products: default
## BLAS:    /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
## LAPACK:  /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib; LAPACK version
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: Europe/Paris
## tzcode source: internal
##
## attached base packages:
## [1] stats4      stats       graphics    grDevices   utils       datasets    methods
## [8] base
##
## other attached packages:
## [1] TxDb.Athaliana.BioMart.plantsmart28_3.2.2
## [2] GenomicFeatures_1.56.0
## [3] AnnotationDbi_1.66.0
## [4] Biobase_2.64.0
## [5] GenomicRanges_1.56.0
## [6] GenomeInfoDb_1.40.1
## [7] IRanges_2.38.0
## [8] S4Vectors_0.42.0
## [9] BiocGenerics_0.50.0
## [10] patchwork_1.2.0
## [11] cowplot_1.1.3
## [12] broom_1.0.6
## [13] gridExtra_2.3
## [14] xtable_1.8-4
```

```

## [15] plyr_1.8.9
## [16] lubridate_1.9.3
## [17]forcats_1.0.0
## [18] stringr_1.5.1
## [19] dplyr_1.1.4
## [20] purrr_1.0.2
## [21] readr_2.1.5
## [22] tidyverse_2.0.0
## [23] tibble_3.2.1
## [24] ggplot2_3.5.1
## [25] tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] DBI_1.2.3                  bitops_1.0-7
## [3] rlang_1.1.4                 magrittr_2.0.3
## [5] matrixStats_1.3.0           compiler_4.4.0
## [7] RSQLite_2.3.7                mgcv_1.9-1
## [9] png_0.1-8                  systemfonts_1.1.0
## [11] vctrs_0.6.5                pkgconfig_2.0.3
## [13] crayon_1.5.2               fastmap_1.2.0
## [15] backports_1.5.0            XVector_0.44.0
## [17] labeling_0.4.3              utf8_1.2.4
## [19] Rsamtools_2.20.0            rmarkdown_2.27
## [21] tzdb_0.4.0                 UCSC.utils_1.0.0
## [23] ragg_1.3.2                 tinytex_0.52
## [25] bit_4.0.5                  xfun_0.44
## [27] zlibbioc_1.50.0             cachem_1.1.0
## [29] jsonlite_1.8.8              blob_1.2.4
## [31] highr_0.11                 DelayedArray_0.30.1
## [33] BiocParallel_1.38.0          parallel_4.4.0
## [35] R6_2.5.1                   stringi_1.8.4
## [37] rtracklayer_1.64.0           Rcpp_1.0.12
## [39] SummarizedExperiment_1.34.0 knitr_1.47
## [41] Matrix_1.7-0                splines_4.4.0
## [43] timechange_0.3.0            tidyselect_1.2.1
## [45] abind_1.4-5                rstudioapi_0.16.0
## [47] yaml_2.3.8                 codetools_0.2-20
## [49] curl_5.2.1                 lattice_0.22-6
## [51] withr_3.0.0                KEGGREST_1.44.0
## [53] evaluate_0.23               Biostrings_2.72.1
## [55] pillar_1.9.0                MatrixGenerics_1.16.0
## [57] generics_0.1.3              RCurl_1.98-1.14
## [59] hms_1.1.3                  munspell_0.5.1
## [61] scales_1.3.0                glue_1.7.0
## [63] tools_4.4.0                 BiocIO_1.14.0
## [65] GenomicAlignments_1.40.0   XML_3.99-0.16.1
## [67] grid_4.4.0                  colorspace_2.1-0
## [69] nlme_3.1-165               GenomeInfoDbData_1.2.12
## [71] restfulr_0.0.15             cli_3.6.2
## [73] textshaping_0.4.0            fansi_1.0.6
## [75] S4Arrays_1.4.1              gtable_0.3.5
## [77] digest_0.6.35               SparseArray_1.4.8
## [79] rjson_0.2.21                farver_2.1.2
## [81] memoise_2.0.1               htmltools_0.5.8.1
## [83] lifecycle_1.0.4              httr_1.4.7
## [85] bit64_4.0.5

```

*Analysis of Meiotic Recombination in Whole-Genome Datasets.” PLOS ONE* 6 (10): 1–17. <https://doi.org/10.1371/journal.pone.0025509>.

Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60.

Zapata, Luis, Jian Ding, Eva M. Willing, Björn Hartwig, Daniela Bezdan, Wenbin B. Jiao, Varun Patel, et al. 2016. “Chromosome-Level Assembly of *Arabidopsis Thaliana* Ler Reveals the Extent of Translocation and Inversion Polymorphisms.” *Proceedings of the National Academy of Sciences* 113 (28): E4052–60. <https://doi.org/10.1073/pnas.1607532113>.