Variations Structurales

Journée PEPI-IBIS

Jeudi/Vendredi 6/7 Juin 2019

Hélène Rimbert, UMR1095 GDEC INRA-UCA

Variation Structurales

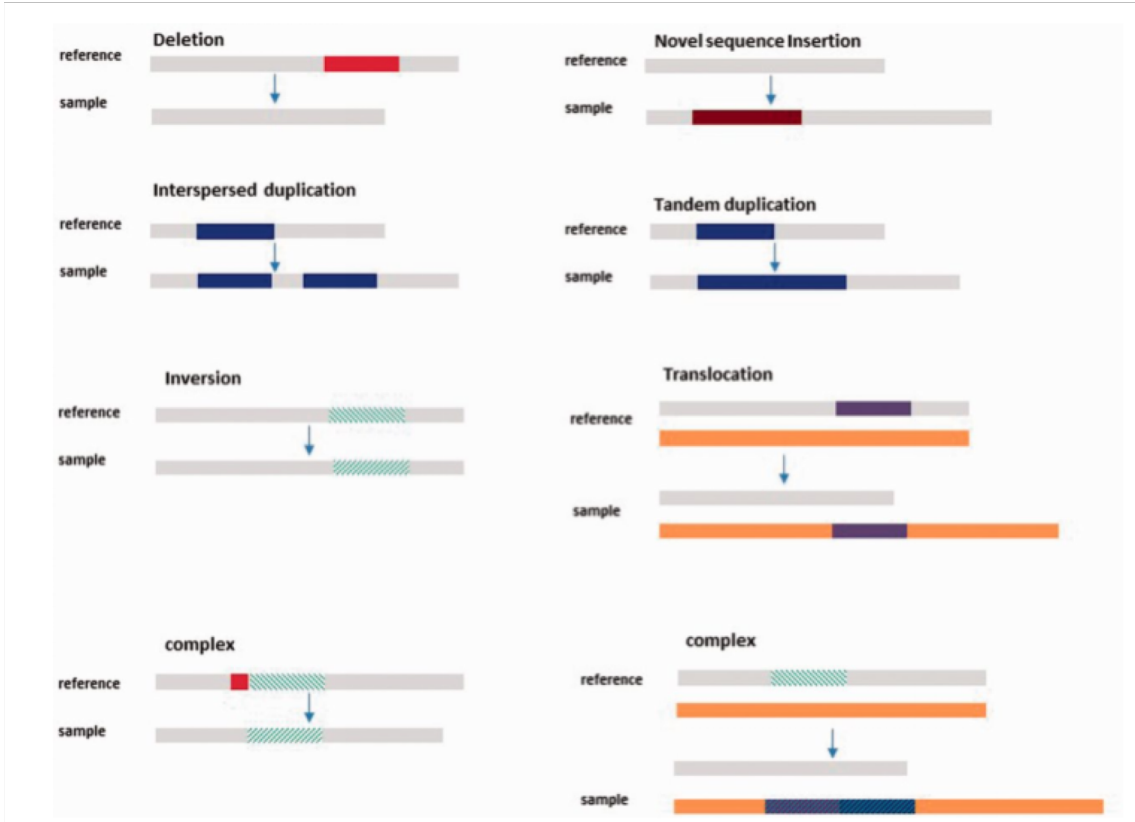
Translocations

Insertion/Délétions

Variations du nombre de copies

Quelle taille minimum?

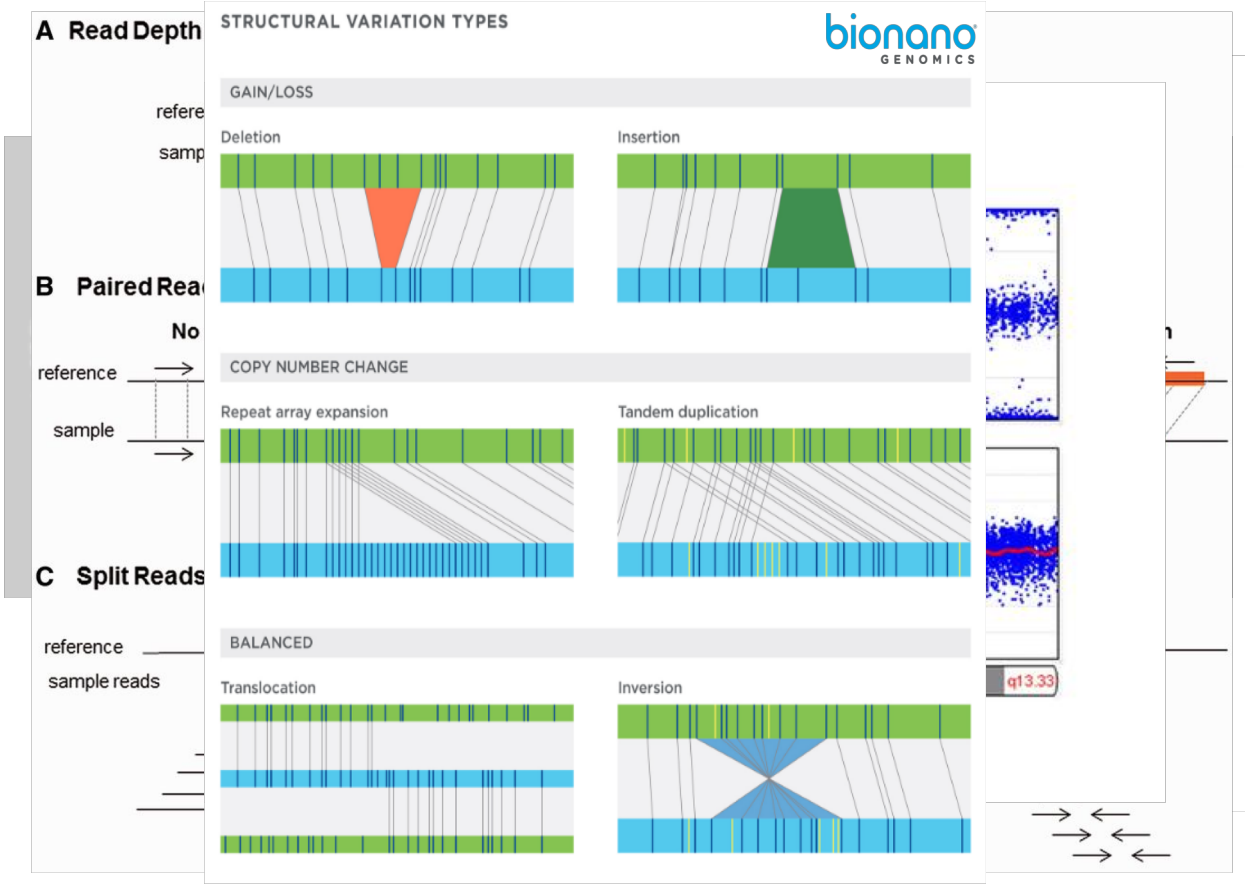
50bp ? 500 bp? 1kb ?



A decade of structural variants: description, history and methods to detect structural variation
Briefings in Functional Genomics, Volume 14, Issue 5, 1 September 2015

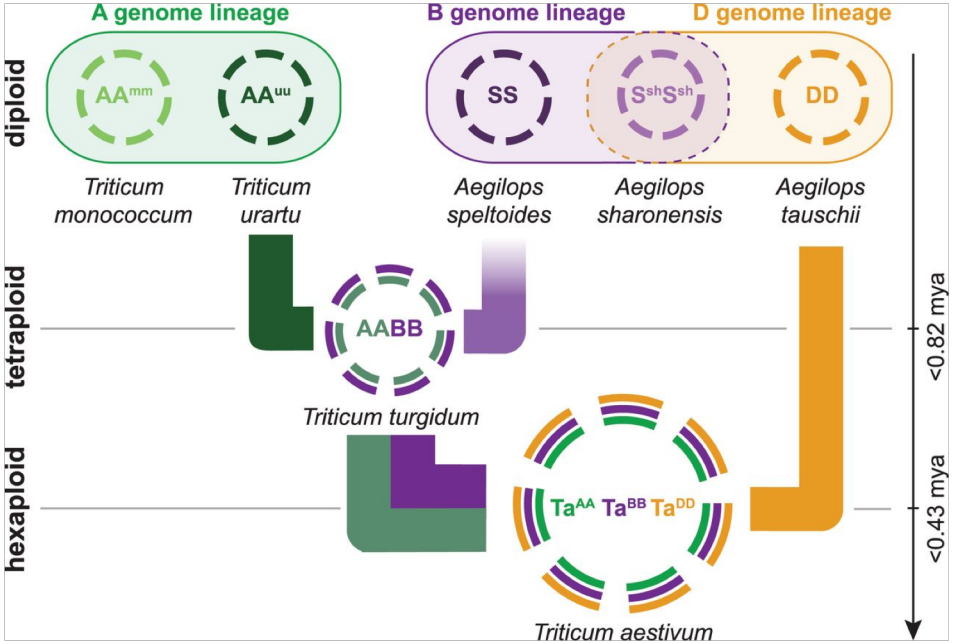
Variation Structurales - Méthodes

- Cytogénétique
- CGH array
- SNP genotyping (LRR/BAF)
- NGS
- Optical mapping



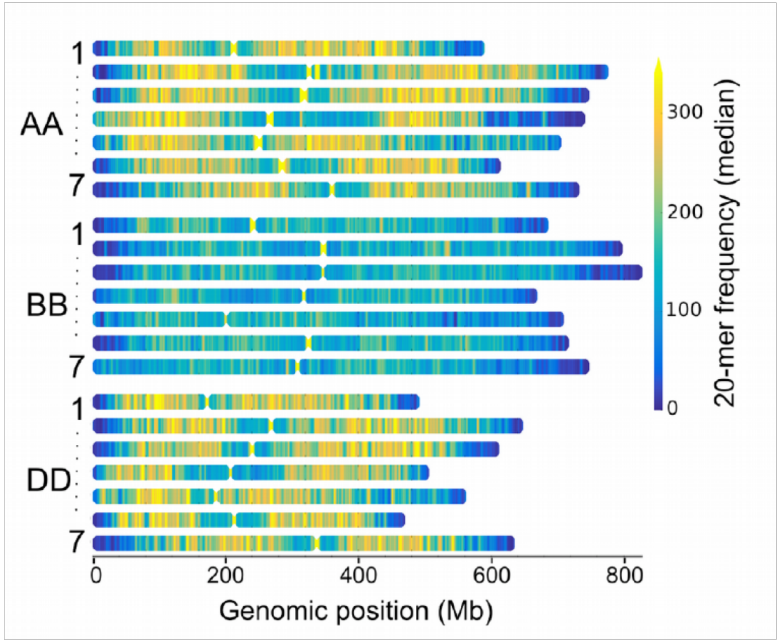
Triticum aestivum genome

T. aestivum history



Science. 2014 Jul 18;345(6194):1251788. doi: 10.1126/science.1251788.
A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome.
International Wheat Genome Sequencing Consortium (IWGSC).

Repeat Content: 85% TE



Science. 2018 Aug 17;361(6403). pii: eaar7191. doi: 10.1126/science.aar7191. Epub 2018 Aug 16.
Shifting the limits in wheat research and breeding using a fully annotated reference genome.

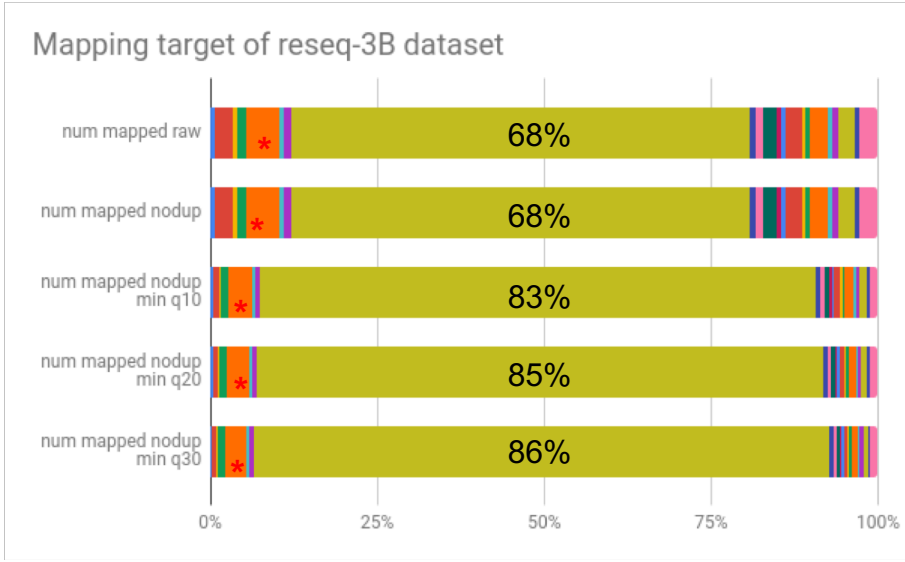
Variation Structurales - NGS sur 3B

Reséquençage chromosomes 3B de blé tendre

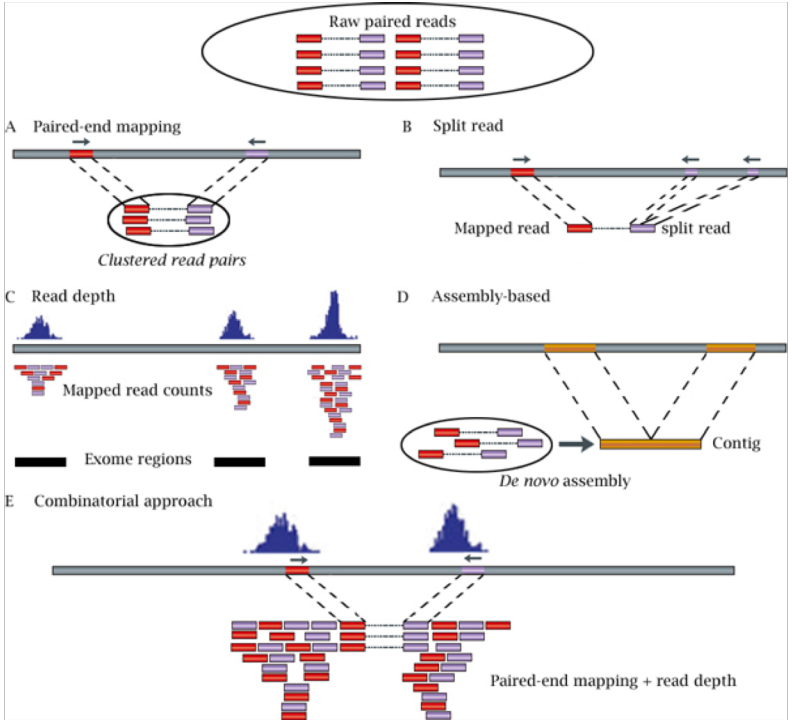
Isoler chromosome 3B: cytométrie de flux

Mapping whole genome

* contamination 2B: 3%



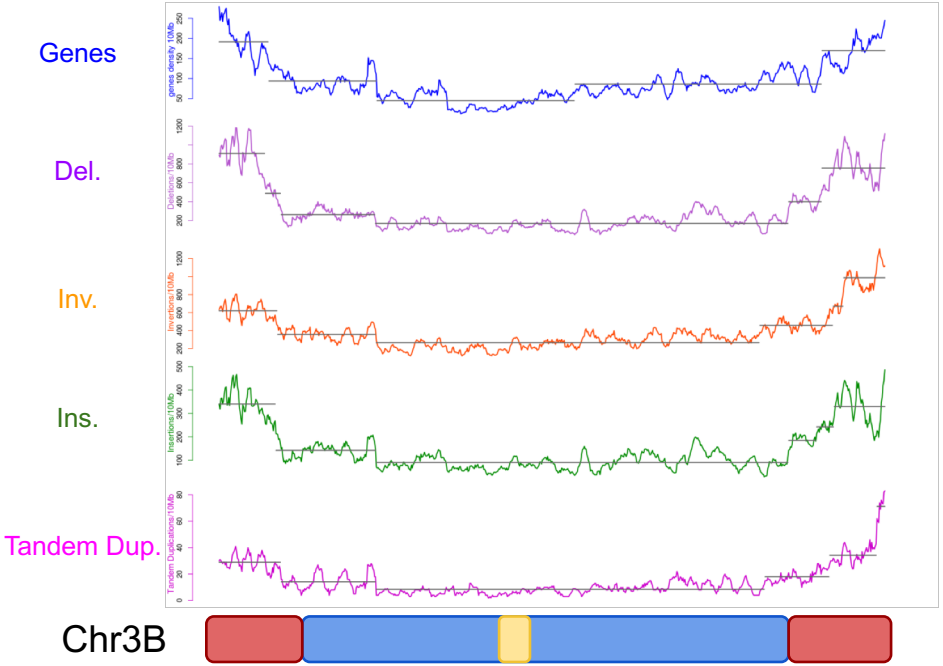
Variation Structurales - NGS sur 3B

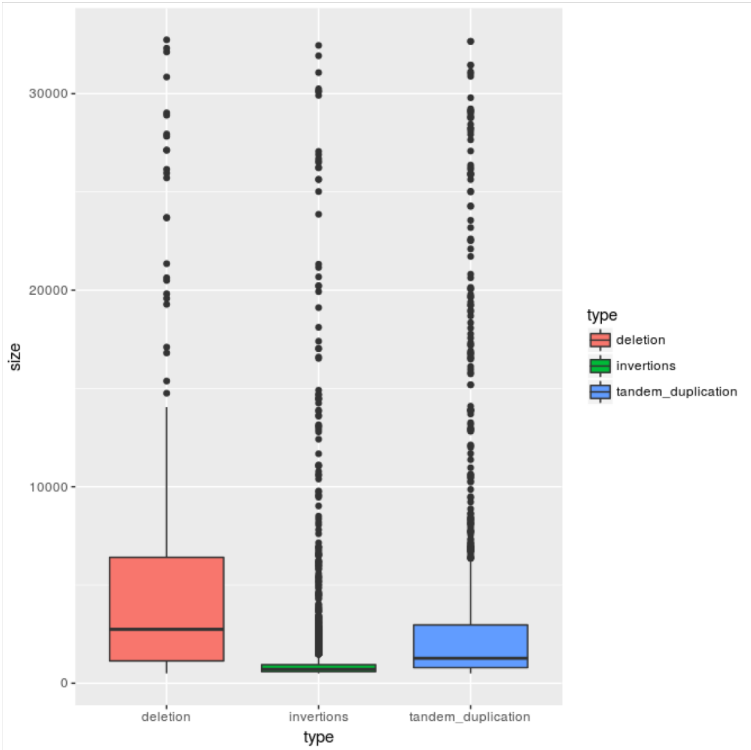


Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics, 2013 https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-S11-S1

Variation Structurales - NGS sur 3B

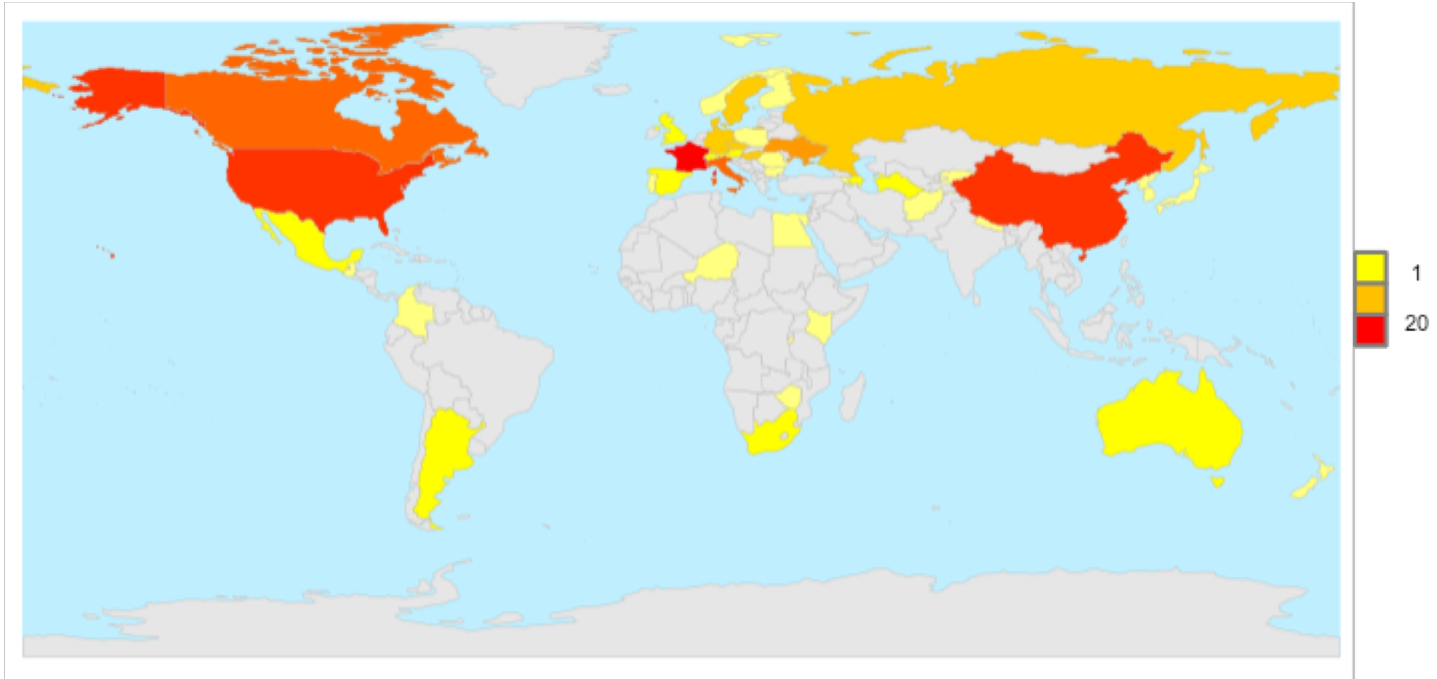
Pindel: Pair-end + Split-read si info disponible pour faire le calling des points de cassure.





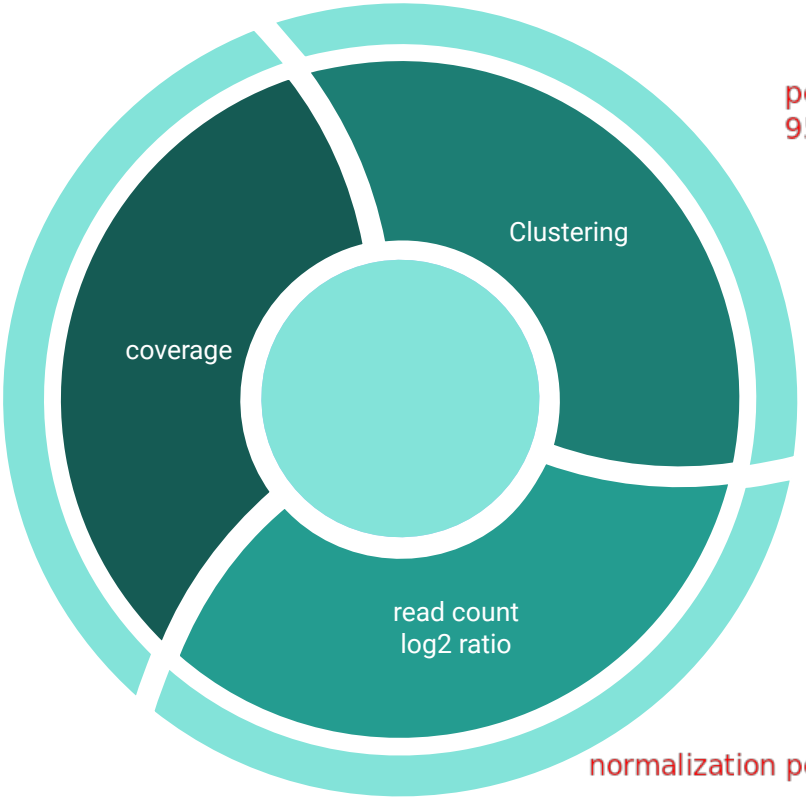
Variation Structurales - ExomeCap

Exome Capture sur 112 accessions sur 95k gènes (High Confidence)

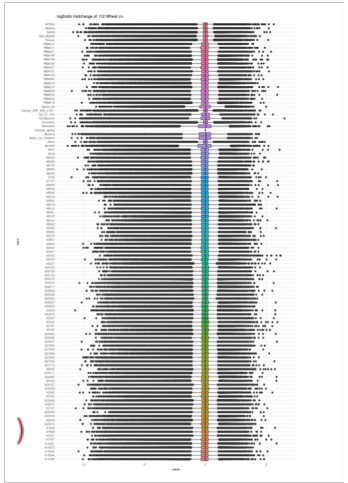


85% coverage moyen
min: 83% (Ding Xi 24)
max: 90% (Barbu du finistère)

Breed wheat



per genome
95k singletons / 110k genes



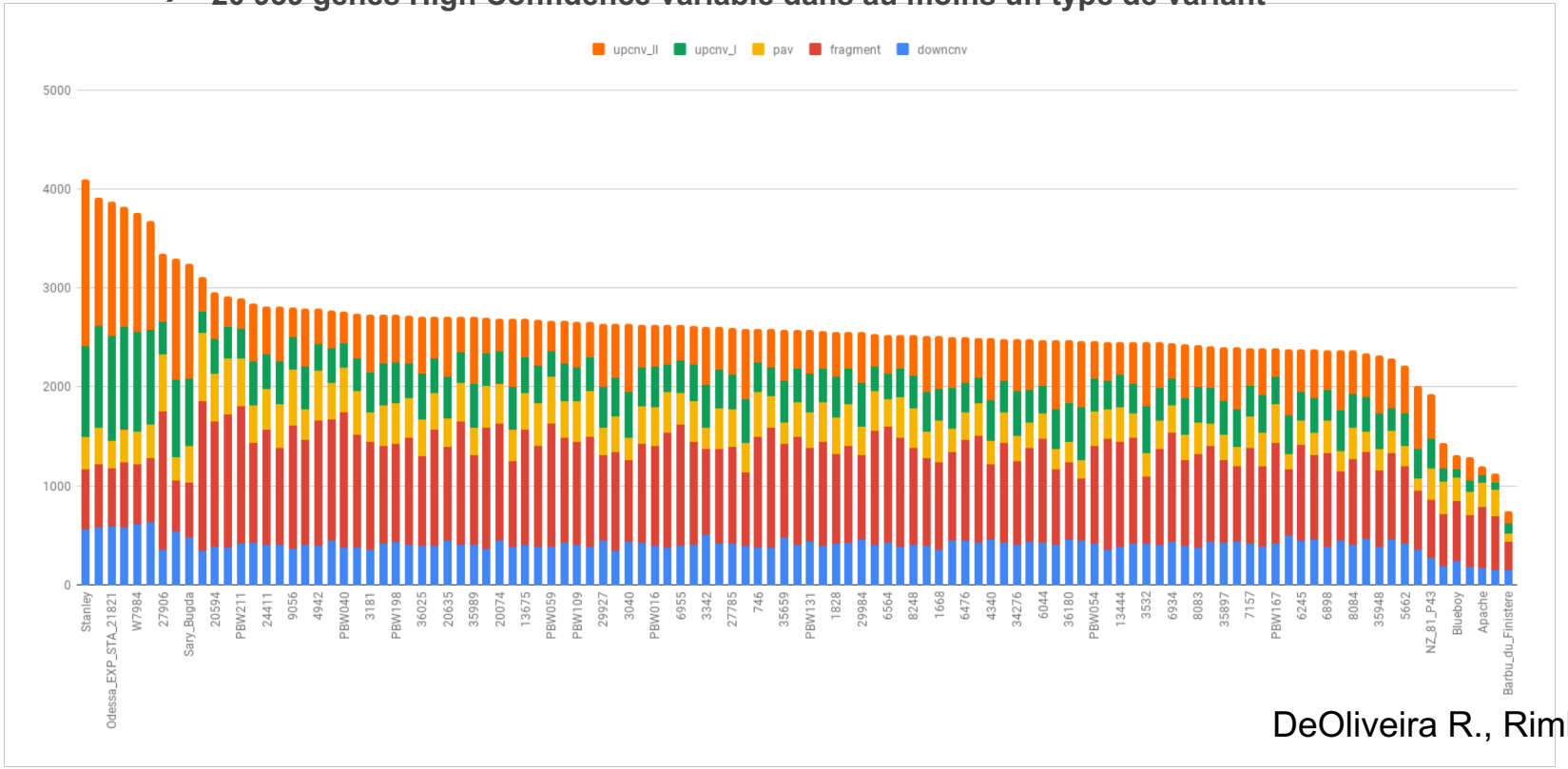
normalization per lib size

$$\log2\left(\frac{\text{Sample norm. read count}}{\text{CS norm. read count}}\right)$$

DeOliveira R. et al en révision

Variation Structurales - ExomeCap

➔ **20 939 gènes High Confidence variable dans au moins un type de variant**



DeOliveira R., Rimbert H.

Variation Structurales - ExomeCap

Visualisation & transfert de données: adapter le format VCF
Coordonnée = gene start

- ALT possibles:

##ALT=<ID=CNV:DEL,Description="Deletion">
##ALT=<ID=CNV:FRAGMENT,Description="Partial deletion">
##ALT=<ID=CNV,Description="Copy number variable region">
##ALT=<ID=CNV:UP,Description="CNV gain of copy compared to reference">

Variation Structurales - ExomeCap

Visualisation & transfert de données: adapter le format VCF
Coordonnée = gene start

- INFO
  - SVTYPE:CNV
  - END = gene stop coordinate
  - SVLENGTH = gene length
  - IMPRECISE
  - GENEID

Variation Structurales - ExomeCap

Visualisation & transfert de données: adapter le format VCF
Coordonnée = gene start

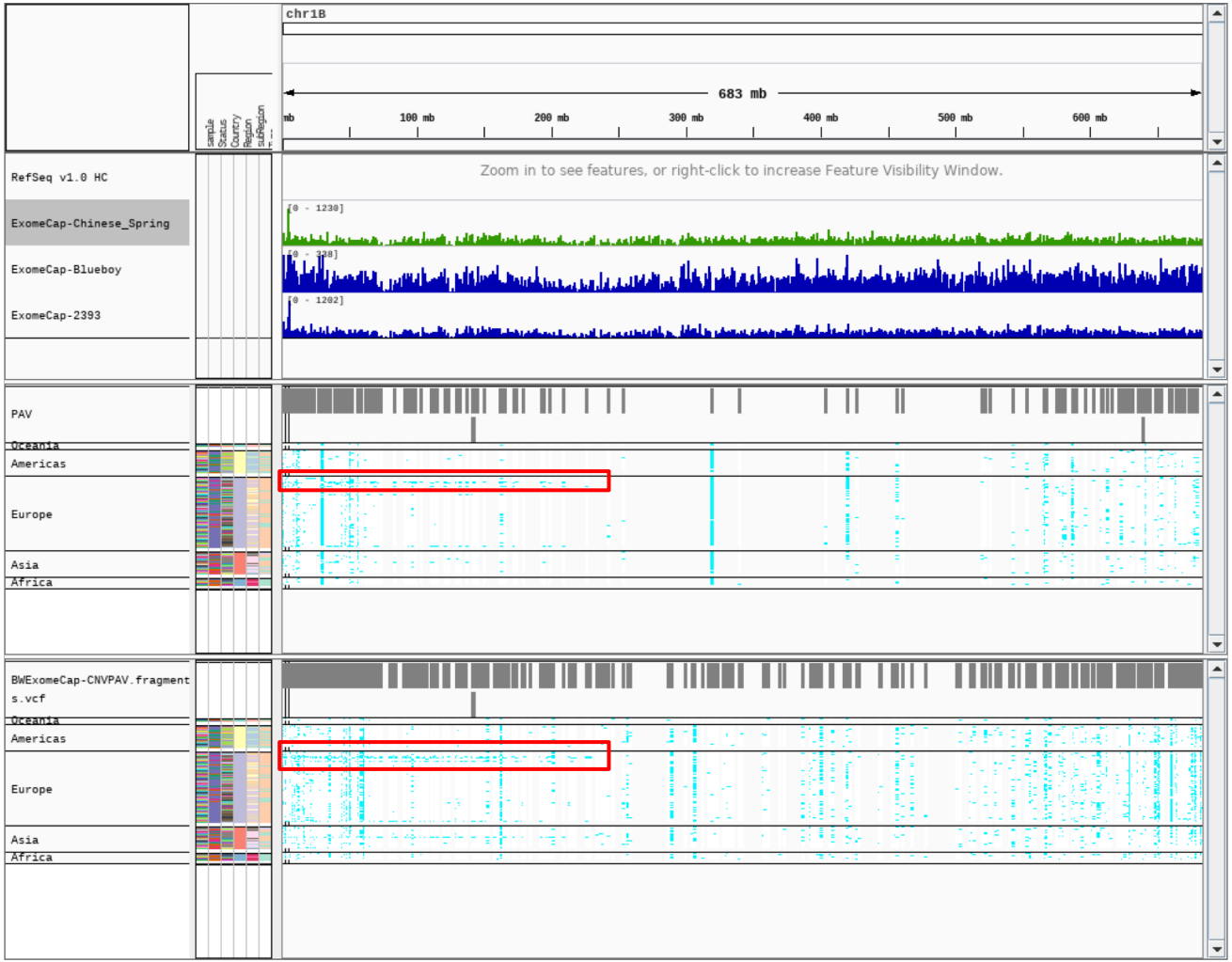
- FORMAT

##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=CV,Number=1,Type=Float,Description="Coverage">
##FORMAT=<ID=L2,Number=1,Type=Float,Description="Log2ratio of read count compared to reference">
##FORMAT=<ID=HZ,Number=1,Type=Integer,Description="SNV count">
##FORMAT=<ID=DESC,Number=1,Type=String,Description="SV type">

##fileformat=VCFv4.0
##fileDate=201811
##reference=IWGSC_RefSeqV1
##INFO=<ID=END,Number=1,Type=Integer,Description="End position of the variant described in this record">
##INFO=<ID=SVLEN,Number=1,Type=Integer,Description="Difference in length between REF and ALT alleles">
##INFO=<ID=IMPRECISE,Number=0,Type=Flag,Description="Imprecise structural variation">
##INFO=<ID=GENEID,Number=1,Type=String,Description="Gene Id">
##INFO=<ID=SVTYPE,Number=1,Type=String,Description="Type of structural variant">
##INFO=<ID=CLUSTER,Number=1,Type=String,Description="Clustering status">
##ALT=<ID=CNV:DEL,Description="Deletion">
##ALT=<ID=CNV:FRAGMENT,Description="Partial deletion">
##ALT=<ID=CNV,Description="Copy number variable region">
##ALT=<ID=CNV:UP1,Description="CNV gain of copy compared to reference with SNV evidence">
##ALT=<ID=CNV:UP2,Description="CNV gain of copy compared to reference without SNV evidence">
##ALT=<ID=CNV:DOWN,Description="CNV loss of copy compared to reference">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=CV,Number=1,Type=Float,Description="Coverage">
##FORMAT=<ID=L2,Number=1,Type=Float,Description="Log2ratio of read count compared to reference">
##FORMAT=<ID=HZ,Number=1,Type=Integer,Description="SNV count">
##FORMAT=<ID=DESC,Number=1,Type=String,Description="SV type">

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	13286
	13384	13444							
chr1A	1144332	BWSVGDEC0B2132C1	T	<CNV:UP>	99	PASS			
	SVTYPE=CNV;END=1145934;SVLEN=1603;GENEID=TraesCS1A01G001800;IMPRECISE;CLUSTER=singleton						GT:CV:L2:HZ:DESC	0/0:0.9261863:0.949	
739178896323:6:0	0/0:0.99560635:0.642489669063505:1:0		0/0:1:-0.0436596716404827:4:0						
chr1A	1159193	BWSVGDEC21151770	G	<CNV:UP>	99	PASS			
	SVTYPE=CNV;END=1162830;SVLEN=3638;GENEID=TraesCS1A01G001900;IMPRECISE;CLUSTER=singleton						GT:CV:L2:HZ:DESC	0/0:0.95174154411:0	
.343351028065941:8:0	0/0:0.97624471911:0.295465458199241:1:0		0/0:0.95871301789:-0.428160237959512:6:0						
chr1A	1209720	BWSVGDEC9FEAD8A2	A	<CNV:FRAGMENT>	99	PASS			
	SVTYPE=CNV;END=1212812;SVLEN=3093;GENEID=TraesCS1A01G002300;IMPRECISE;CLUSTER=singleton						GT:CV:L2:HZ:DESC	0/0:0.9579646:0.126	
258637758877:0:0	0/0:0.9496066833:-0.0155962336354707:0:0		0/0:0.8946173833:-0.78141369557814:0:0						
chr1A	1657114	BWSVGDEC4C1BFAEF	A	<CNV:FRAGMENT>,<CNV:DEL>,<CNV:UP>	99	PASS			
	SVTYPE=CNV;END=1666923;SVLEN=9810;GENEID=TraesCS1A01G002900;IMPRECISE;CLUSTER=singleton						GT:CV:L2:HZ:DESC	0/0:0.69946236:-0.115763967056228:0:0	
	0/0:0.71481482:0.175780884727282:0:0		1/1:0.2910048:-3.58752007143849:0:fragment						
chr1A	1668948	BWSVGDEC75335800	G	<CNV:FRAGMENT>,<CNV:DEL>	99	PASS			
	SVTYPE=CNV;END=1670358;SVLEN=1411;GENEID=TraesCS1A01G003000;IMPRECISE;CLUSTER=singleton						GT:CV:L2:HZ:DESC	0/0	
:1:0.395762330214666:0:0		0/0:0.8152174:0.387769739337288:0:0		0/0:0.7754764:0.229908896793696:0:0					
chr1A	1694198	BWSVGDECA2C987C7	C	<CNV:FRAGMENT>,<CNV:DEL>,<CNV:UP>	99	PASS			
	SVTYPE=CNV;END=1698603;SVLEN=4406;GENEID=TraesCS1A01G003100;IMPRECISE;CLUSTER=singleton						GT:CV:L2:HZ:DESC	0/0:1:0.0560910614844838:0:0	
	3/3:1:1.1998822928243:0:upcnv		0/0:0.99550705:0.865378577359578:0:0						
chr1A	2161813	BWSVGDEC6970D7FE	T	<CNV:FRAGMENT>,<CNV:DEL>,<CNV:UP>	99	PASS			
	SVTYPE=CNV;END=2163439;SVLEN=1627;GENEID=TraesCS1A01G003400;IMPRECISE;CLUSTER=singleton						GT:CV:L2:HZ:DESC	2/2:0:0:0:pav	2/2:0:0:0:pav
	0/0:0.797048:0.35331785609337:0:0								
chr1A	2377712	BWSVGDEC6414AAA8	T	<CNV:UP>	99	PASS			
	SVTYPE=CNV;END=2379143;SVLEN=1432;GENEID=TraesCS1A01G003800;IMPRECISE;CLUSTER=singleton						GT:CV:L2:HZ:DESC	0/0:0.76380	
15:-0.628541282371425:0:0		0/0:0.7638015:-0.72738326480568:1:0		0/0:0.6289308:-0.67050525855626:0:0					
chr1A	2540490	BWSVGDECE4744317	A	<CNV:FRAGMENT>	99	PASS			
	SVTYPE=CNV;END=2540958;SVLEN=469;GENEID=TraesCS1A01G004200;IMPRECISE;CLUSTER=singleton						GT:CV:L2:HZ:DESC	1/1:0.2621359:-3.77	

- IGV
- VCFTools
- BCFTools



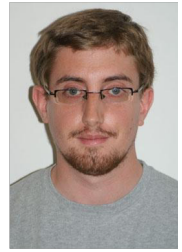
Variation Structurales - OTV

OTV: traces de PAV ?

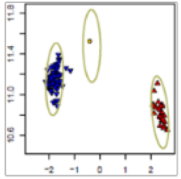
TaBW280K SNP array

X

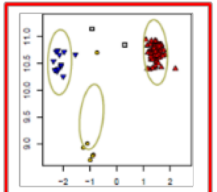
4,506 wheat lines



113,457 SNPs with less than 2% missing or heterozygous data



99,333 PolyHighRes



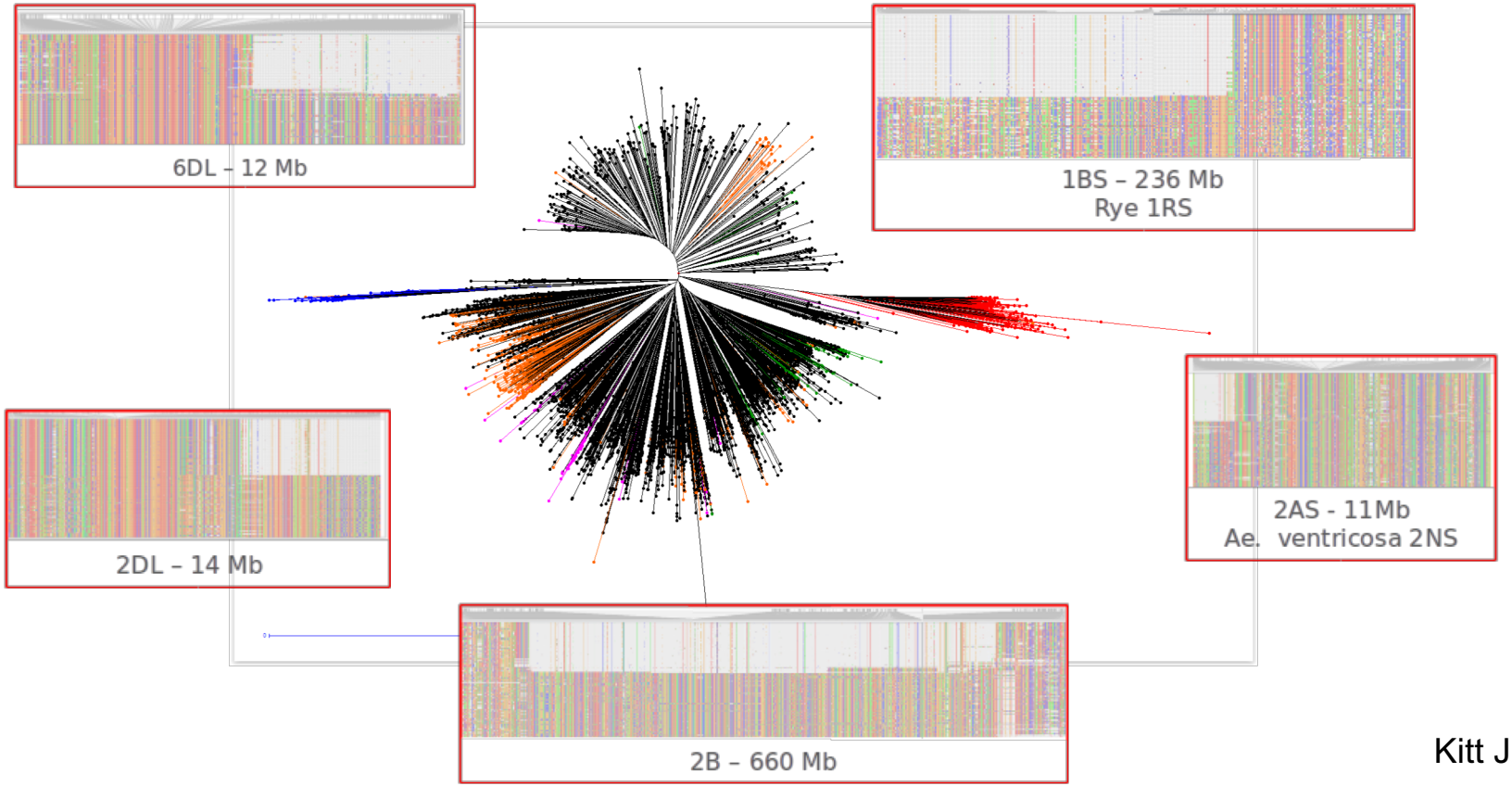
14,124 OTVs

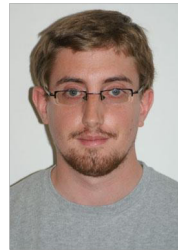
Presence-absence variations

8,741 haplotypes

Kitt J., Paux. E

Variation Structurales - OTV

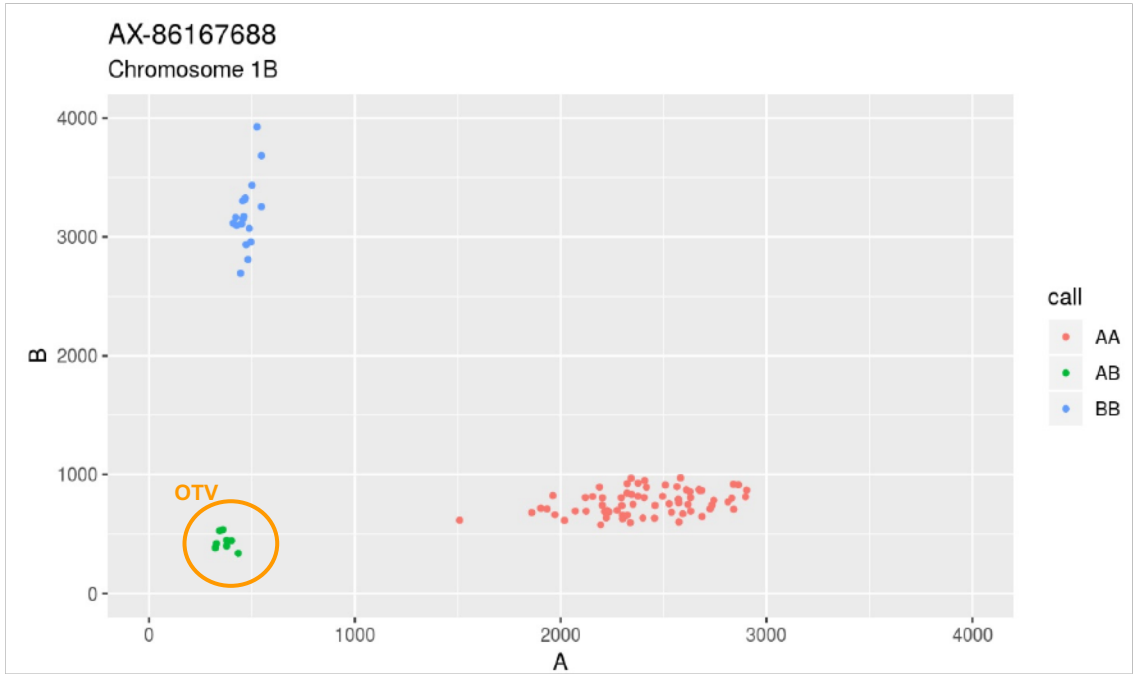


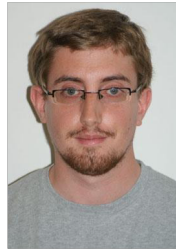


Kitt J., Paux. E

Variation Structurales - BAF & LRR

Variations d'intensité du signal = PAV ?





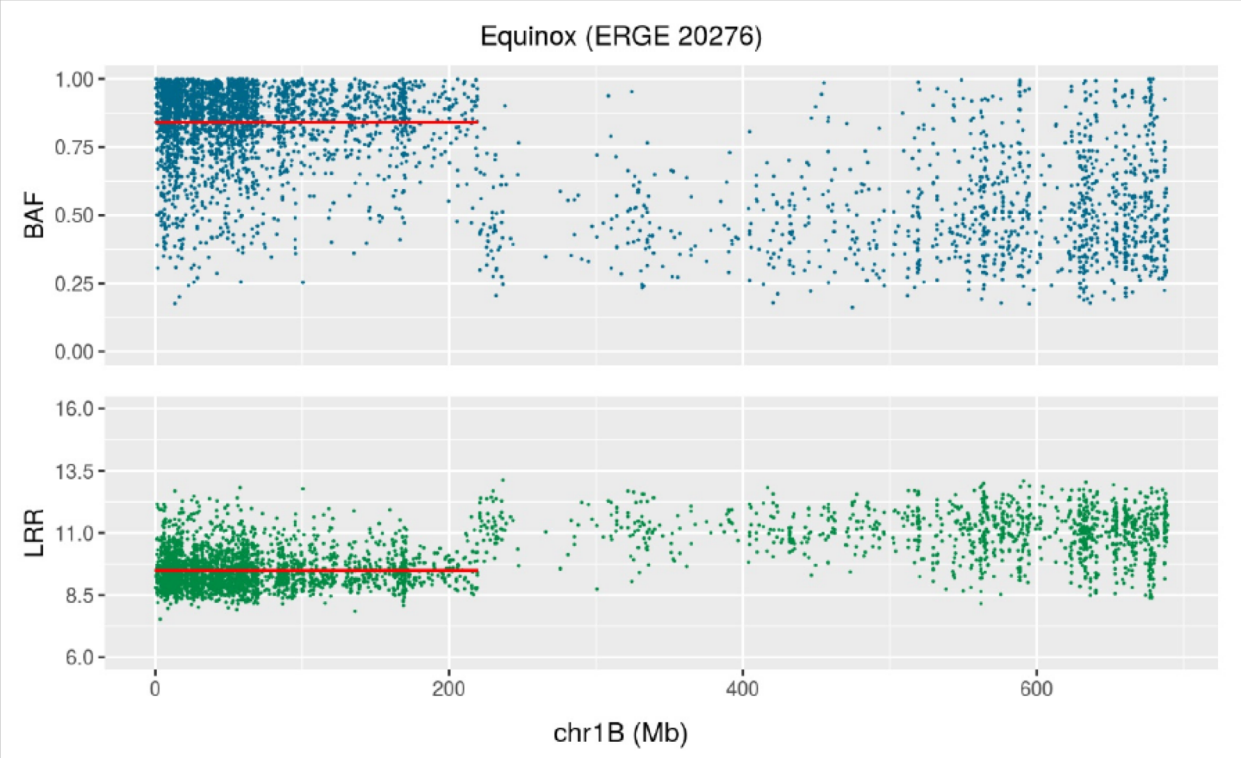
$$LRR = \sqrt{A^2 + B^2}$$

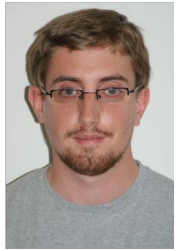
$$BAF = \frac{B}{A+B}$$

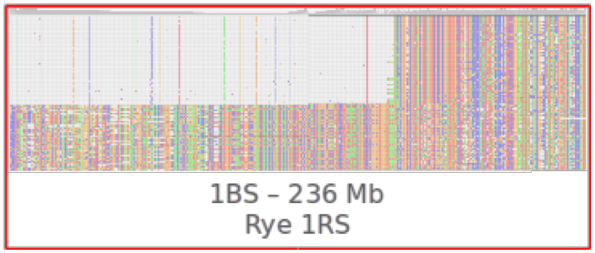
Kitt J., Paux. E

Variation Structurales - BAF & LRR

Exemple 1B (Rye 1RS, package R DNAcopy)





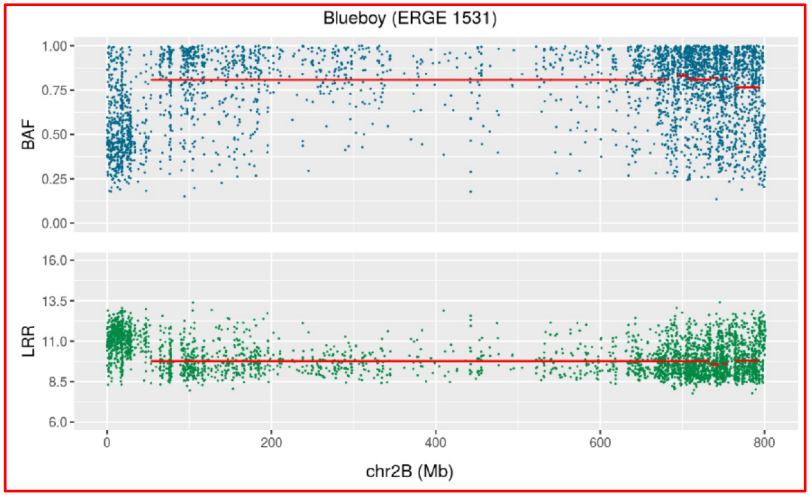


Kitt J., Paux. E

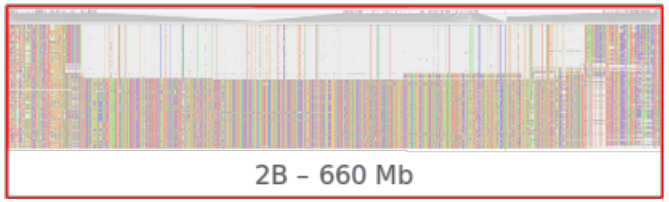
Variation Structurales

Exemple 2B, cv. Blueboy

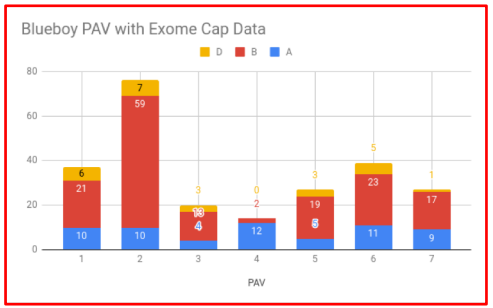
BAF/LRR



OTV



ExomeCap



Kitt J., Rimbert H., Paux E.

Merci de votre attention

SEVEN
De Oliveira Romain
Juery Caroline
Kitt Jonathan
* Paux Etienne
Ranoux Marion

Bioinfo
* Choulet Frédéric
Leroy Philippe
Monat Cécile
Papon Nathan
Rimbert Hélène

DORG
* Balfourier François

* Antoine Mahul
Nadia Goue
David Grimbichler