Rose GWAS browser

Survival guide

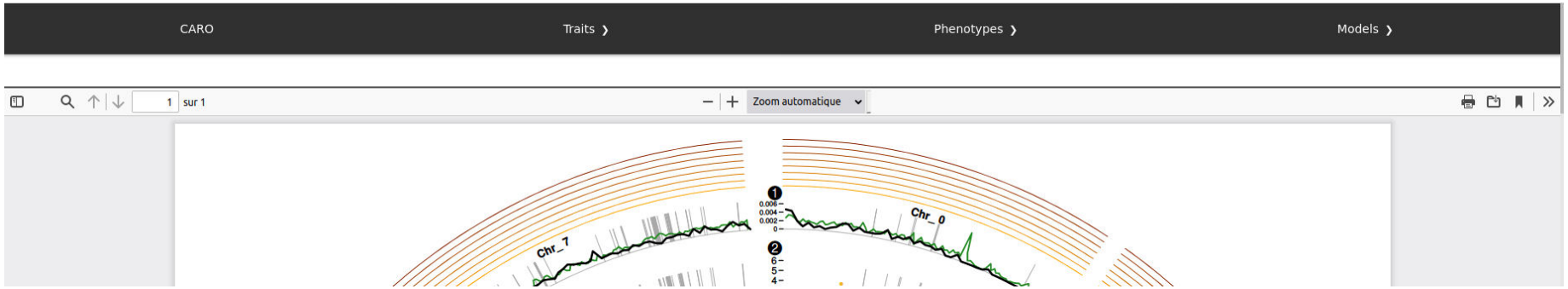
Leroy T., Albert E., Thouroude T., Baudino S., Caissard J-C., Chastellier A., Chameau J., Jeauffre J., Loubert T., Paramita S.N., Pernet A., Soufflet-Freslon A., Oghina-Pavie C., Foucher C., Hibrand-Saint Oyant L.*, Clotault J.* Dark side of the honeymoon: reconstructing the Asian x European rose breeding history through the lens of genomics, bioRxiv

Thibault Leroy - Last update: June 21th 2023

How to browse on this website?

1/ First **select a main trait** on the drop-down menu « Traits », e.g. « Black Spot »

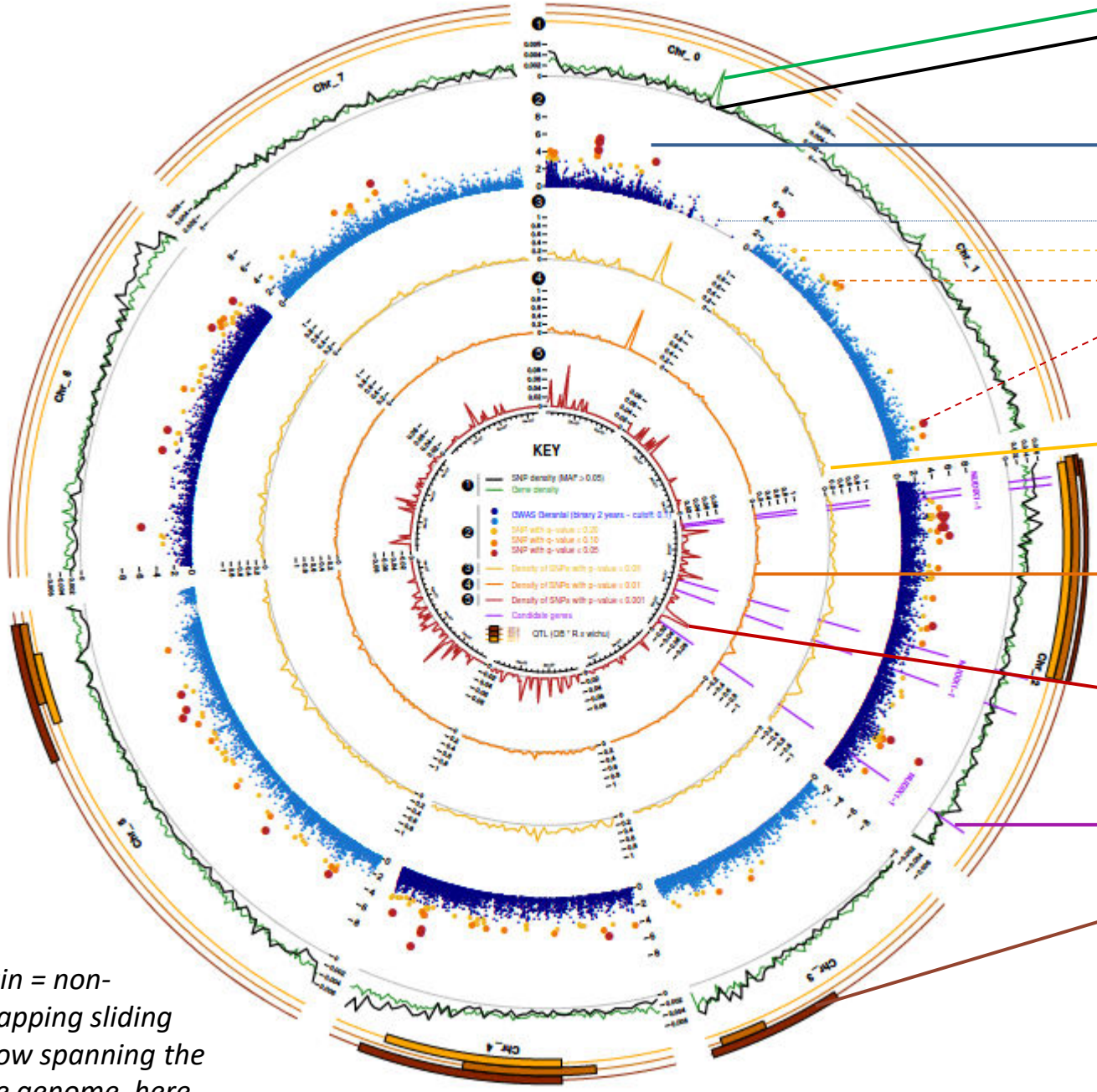
2/ Then you will **have access to a new drop-down menu « Phenotypes »**, « Parameters » or similar to more precisely select a phenotype of interest, *e.g.* « Blackspot - 3 years ». For some phenotypes, GWAS was also performed on (binary: 0/1 around the indicated threshold value)



3/ Then a pdf is expected to be open on the webpage. By default, the results shown are those from the general model of GWASpoly. For more information regarding the GWASpoly model, see slide 4 of this survival guide.

4/ To select any other model than the general model, a **new drop-down menu (« Models ») is then available**. Select the « additive » model for instance (see also slide 4).

How to read the results?



Gene density (500kb Slidwin)
SNP density (only those with MAF ≥ 0.05, 1Mb Slidwin)

'Manhattan plot' of the -log10(p-valeurs) for all SNPs with MAF ≥ 0.05

Blue dot= SNP with q-value > 0.2
Yellow= SNP with q-value ≤ 0.2
Orange = SNP with q-value ≤ 0.1
Red = SNP with q-value ≤ 0.05

The q-value is a FDR-corrected p-value

Density of associated SNPs (pval ≤ 0.05) per window (1Mb Slidwin)

Density of associated SNPs (pval ≤ 0.01) per window (1Mb Slidwin)

Density of associated SNPs (pval ≤ 0.001) per window (1Mb Slidwin)

Know candidate gene

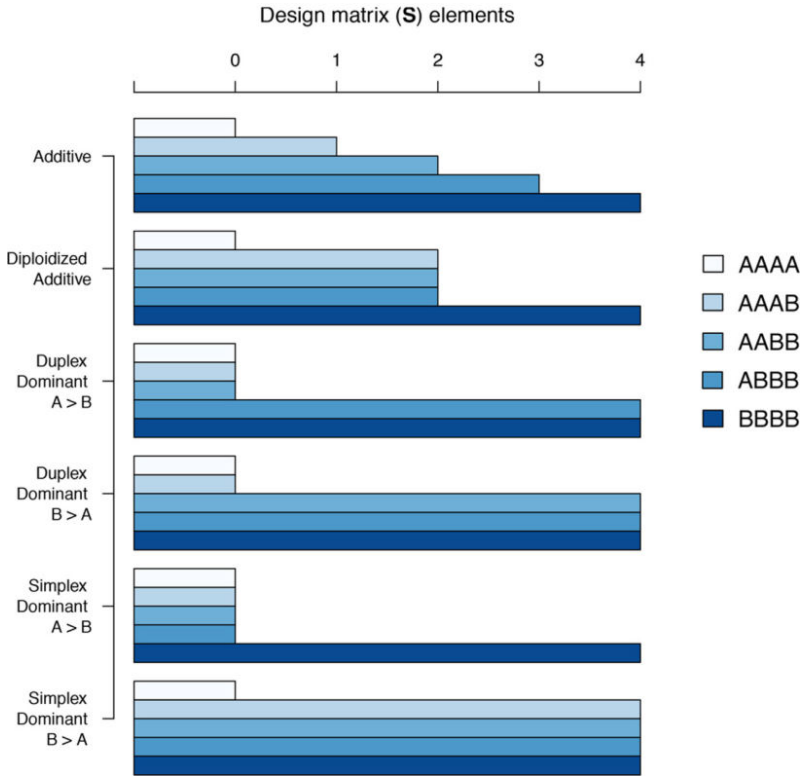
QTL OW, if available (one line per year)
If a line is split into two (e.g. here on chr2 for the 1st year, it means that a different QTL was observed on the two parental maps)

Slidwin = non-overlapping sliding window spanning the whole genome, here 1Mb slidwin used.

Why so many different GWAS models ?

It is important to note that performing GWAS on a polyploid species generates an additional degree of complexity regarding the genetic dominance of a given trait.

Note that even if all our 204 rose genotypes are not tetraploids, all were genotyped as tetraploid (most are still truly tetraploids, the remaining genotypes are mostly diploid). At each SNP, each individual has 5 possible classes of alleles B {0, 1, 2, 3, 4} depending on the observed dosage levels of allele B, *i.e.* AAAA, AAAB, AABB, ABBB or BBBB.



Rosyara *et al.* 2016 The Plant Genome (GWASpoly)

The general model is a simple model where the SNP effect can be on any of the classes. This model allows to rapidly explore the results, but sometimes highlights too many associated SNPs (and can provide a quite noisy signal)

The other models are more specific regarding the dominance (see opposite and the GWASpoly paper for details)

How to rapidly identify the most interesting genomic regions for each trait to focus on?

Main Trait (benzylalcohol)

Quantitative/qualitative phenotypic variation

GWAS model

compounds	traits	models	chr0	chr1	chr2	chr3	chr4	chr5	chr6	chr7
Benzylalcohol	quantitative	general	NA	NA	NA	NA+	NA+/-	NA	NA	NA
		diploidized general		+		++				++
		additive				+++	+			++
		diploidized additive		+		+++				++
		simplex dominant R				+++				+++
		simplex dominant A		++		++				++
		duplex dominant R	NA	NA	NA	NA+	NA+/-	NA	NA	NA
		duplex dominant A	NA	NA	NA	NA+	NA+/-	NA	NA	NA+/-
	qualitative 0.1	general	NA+/-	NA+/-	NA+/-	NA++	NA+/-	NA++	NA+	NA
		diploidized general				+++			+	
		additive			+	+			+	
	(pb en ligne)	diploidized additive				+++			++	
		simplex dominant R				+++			++	
		simplex dominant A				+++			++	+
		duplex dominant R				++		+	+	
		duplex dominant A				++			++	
	Qualitative 5	general	+/-		+/-	+++				
		diploidized general	+/-			+++				+
		additive				+++	+/-			++
		diploidized additive	+/-			+++				++
		simplex dominant R				+++				++
		simplex dominant A	+/-			+++				+/-
		duplex dominant R				+++				+/-
	(pb en ligne PD)	duplex dominant A				++				
Betacitronellol	quantitative	general			++			+	++	
		diploidized general						+/-	+/-	

Key

+++	Clear signal, main peak
++	Likely true signal, 2nd peak
+	weaker local signal
+/-	Even weaker local signal
NA	background noise too strong
NA+/-	background noise too strong, weak evidence
NA+	background noise too strong, but likely signal
NA++	background noise too strong, but clear signal
	none (or evidence considered to be too weak)

Whatever the analysis done (quali/quanti) and the model used, the main signal is on chr3

Secondary signal on chr6 if the variation is qualitative for genotypes with very low (<0.1) vs. at least moderate (>0.1) content in Benzylalcohol

Secondary signal on chr7 if the variation is quantitative or qualitative for genotypes with low (<5) vs. high (>5) content in Benzylalcohol

This spreadsheet is available for download here:

https://github.com/roseGWASbrowser/rosegwasbrowser.github.io/blob/master/Summary_GWAS_CARO_100322.xlsx