TraceAncestor

PATRICK
OLLITRAULT

- Les parents/ancêtres sont déjà identifiés (pamplemousse, mandarinier, cédrat, micrantha)

- **Méthode** : Pour un individu donné, mesurer fenêtre par fenêtre sur un chromosome la fréquence de présence de SNPs ancestraux.

2 outils dans galaxy :

- **TraceAncestor prefilter**

- **TraceAncestor**

Matrice contenant les indices de différenciation par SNP (GST) par ancêtre

traceAncestor

TraceAncestor prefilter

Liste de SNPs diagnostiques filtrés

VCF des hybrides
+
fichier couleur

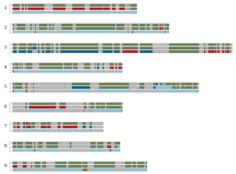
TraceAncestor

Blocs de chromosomes à coloriser

Visualisation

ideogram

circos



GENOME HARVEST

parental SNP - **Detect parental SNP of hybrids**

Trace Ancestor

- TraceAncestor prefilter
- TraceAncestor

vcfHunter

KDE Classifier

Visualization

GENOME HARVEST

parental SNP - **Detect parental SNP of hybrids**

Trace Ancestor

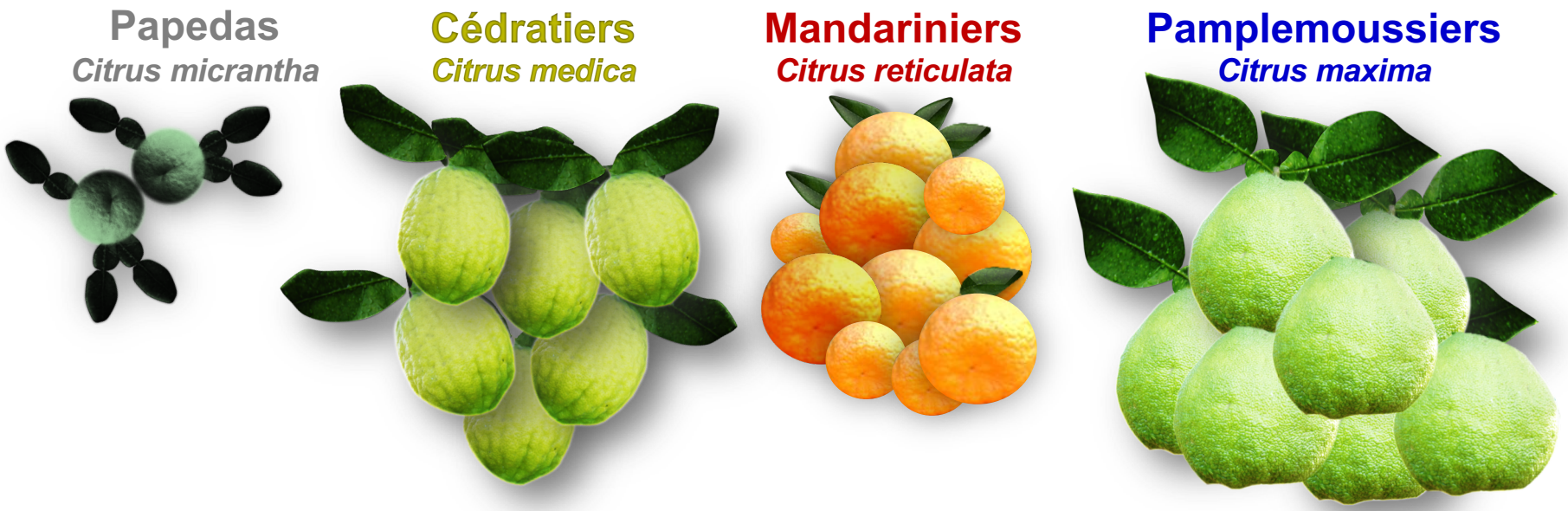
vcfHunter

KDE Classifier

Visualization

- Ideogram Chromosome Painting
- CircosJS CircosJS Client to build interactive graphs in a circular layout

Genotyping by Sequencing (GBS)



Diagnostic markers of the **four basic taxa** was estimated based on the G_{ST} parameter (coefficient of gene differentiation which measure the differentiation among subpopulations; Nei, 1973).

We considered **two sub-populations: 1**- the **taxa concerned** and **2- a theoretical population of the 3 other basic taxa**.
The **higher the value is**, the **more differentiated the taxa are**.
G_{st} is defined as the ratio between the inter population diversity and the total diversity:

$$Gst = \frac{He_{Tot} - Hs}{He_{Tot}} = \frac{He_{Tot} - \sum he/n}{He_{Tot}}$$

TraceAncestor prefilter

→ Sélection des marqueurs diagnostiques

Données manquantes

Indices de différenciation (GST) par ancêtre (chacun en comparaison aux 3 autres)

Fréquence de l'allele ALT

#CHROM	POS	REF	ALT	%Nref	GSTA1	GSTA2	GSTA3	GSTA4	FA1	FA2	FA3	FA4
1	85524	A	G	0.3103448276	0.2	0.2	0.2	1	0	0	0	1
1	108710	A	T	0.6206896552	0.2	1	0.2	0.2	0	1	0	0
1	108741	T	A	0.2413793103	0.2	0.2	1	0.2	0	0	1	0
1	109226	A	T	0	0.2	0.2	0.2	1	0	0	0	1
1	109661	A	G	0.3448275862	0.2	0.2	1	0.2	0	0	1	0
1	110915	A	C	0.3448275862	1	0.2	0.2	0.2	0	1	1	1

- Tri en fonction des données manquantes (< 0.3 par défaut)
- Tri en fonction des valeurs de GST (> 0.9 par défaut). Si GST fort pour un ancêtre→ la diversité allélique totale à cette position est majoritairement expliquée par cet ancêtre

TraceAncestor prefilter

→ Sélection des marqueurs diagnostiques

Données manquantes

Indices de différenciation (GST) par ancêtre (chacun en comparaison aux 3 autres)

Fréquence de l'allele ALT

#CHROM	POS	REF	ALT	%Nref	GSTA1	GSTA2	GSTA3	GSTA4	FA1	FA2	FA3	FA4
1	85524	A	G	0.3103448276	0.2	0.2	0.2	1	0	0	0	1
1	108710	A	T	0.6206896552	0.2	1	0.2	0.2	0	1	0	0
1	108741	T	A	0.2413793103	0.2	0.2	1	0.2	0	0	1	0
1	109226	A	T	0	0.2	0.2	0.2	1	0	0	0	1
1	109661	A	G	0.3448275862	0.2	0.2	1	0.2	0	0	1	0
1	110915	A	C	0.3448275862	1	0.2	0.2	0.2	0	1	1	1

- Définition de la valeur de l'allèle ancestral -> REF ou ALT?
  - Si F > 0.8 → ALT
  - Si F < 0.2 → REF

TraceAncestor prefilter

→ Sélection des marqueurs diagnostiques

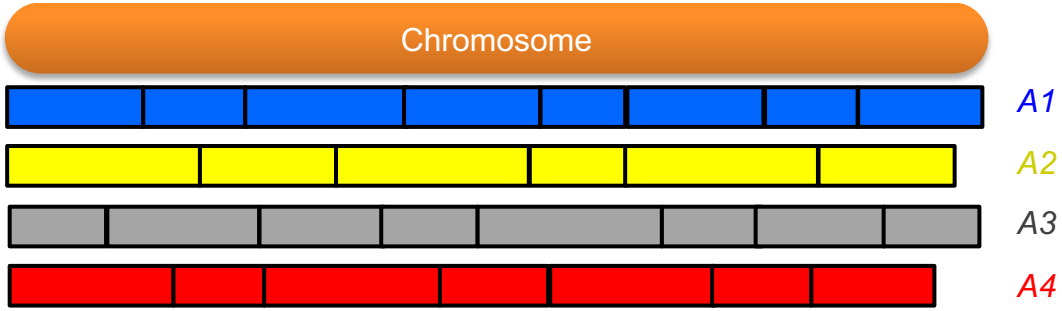
#CHROM	POS	REF	ALT	%Nref	GSTA1	GSTA2	GSTA3	GSTA4	FA1	FA2	FA3	FA4
1	85524	A	G	0.3103448276	0.2	0.2	0.2	1	0	0	0	1
1	108710	A	T	0.6206896552	0.2	1	0.2	0.2	0	1	0	0
1	108741	T	A	0.2413793103	0.2	0.2	1	0.2	0	0	1	0
1	109226	A	T	0	0.2	0.2	0.2	1	0	0	0	1
1	109661	A	G	0.3448275862	0.2	0.2	1	0.2	0	0	1	0
1	110915	A	C	0.3448275862	1	0.2	0.2	0.2	0	1	1	1

ancestor	chromosome	position	allele
A3	1	108741	A
A4	1	109226	T

TraceAncestor

Etape 1

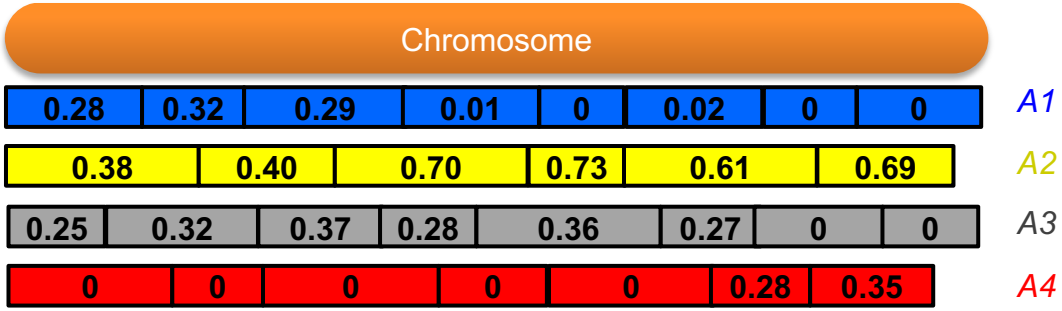
Découpage du chromosome en fenêtres non chevauchantes de 10 SNPs



TraceAncestor

Etape 2

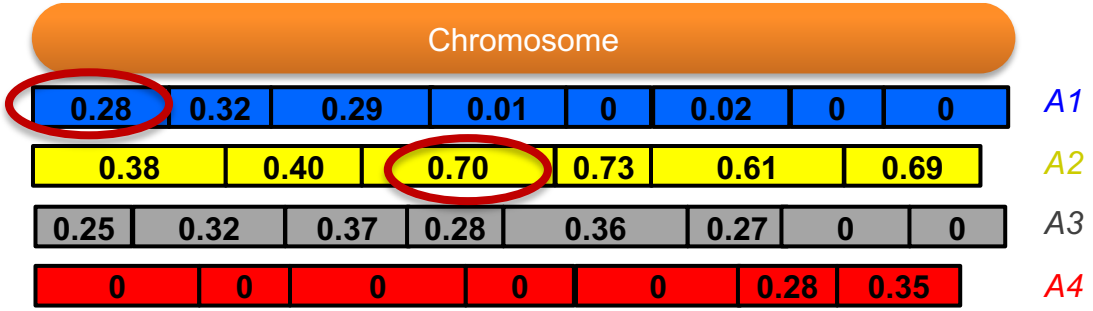
Calcul de la fréquence des reads ancestraux par ancêtre et par fenêtre de 10 SNPs.



TraceAncestor

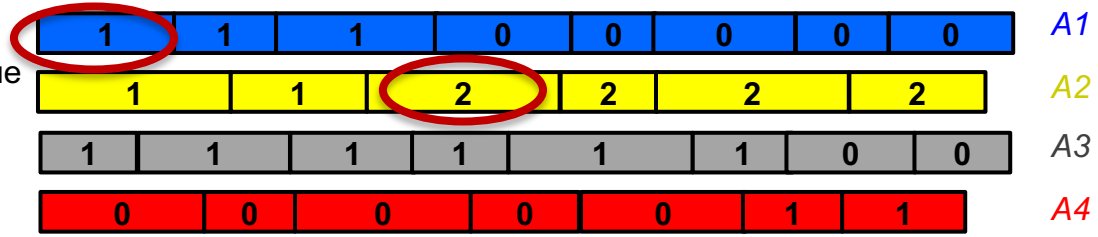
Etape 2

Calcul de la fréquence des reads ancestraux par ancêtre et par fenêtre de 10 SNPs.



Etape 3

Estimation du dosage allélique de chaque ancêtre par fenêtre de 10 SNPs



Test de vraisemblance (LOD) des différentes hypothèses 2 à 2, entre la fréquence observée et théorique pour triploïde

Diploid: 0.05 / 0.5 / 0.95

Triploid: 0.05 / 0.33 / 0.66 / 0.95

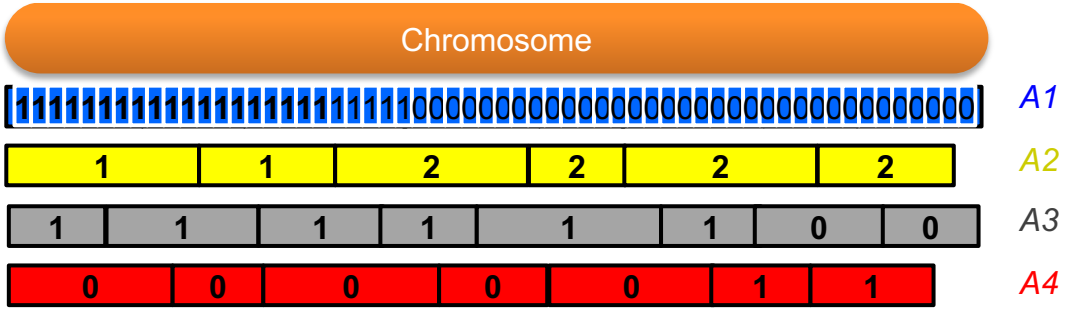
Tetraploid: 0.05 / 0.25 / 0.5 / 0.75 / 0.95

Si (-3 < LOD < 3) => indétermination

TraceAncestor

Etape 3

Estimation du dosage allélique de chaque ancêtre par fenêtre de 10SNP



Etape 4

Division du chromosome en sous-fenêtres non chevauchantes de 100kb. Le dosage allélique des fenêtres de 10 SNP est reporté dans les fenêtres de 100Kb

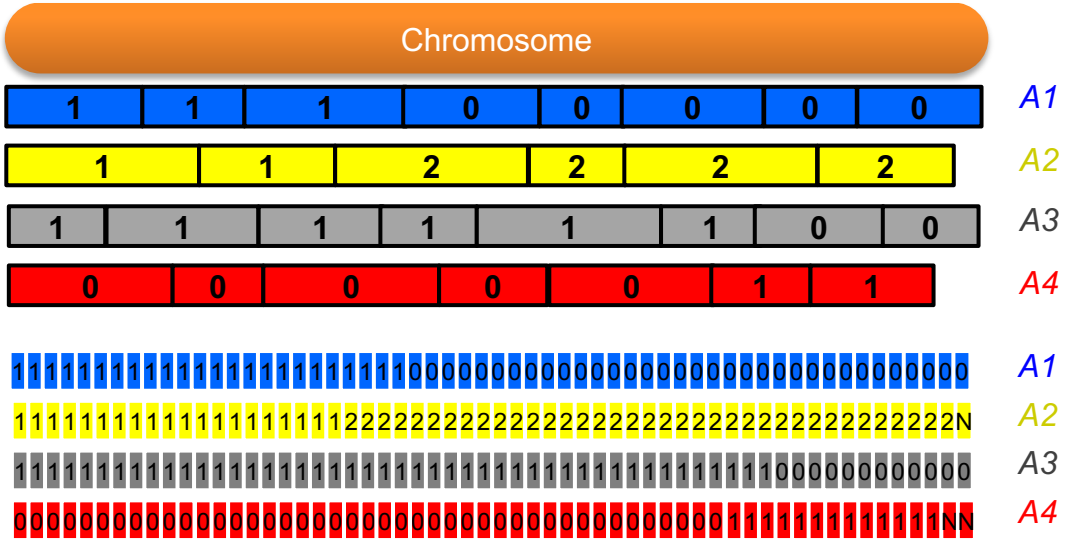
TraceAncestor

Etape 3

Estimation du dosage allélique de chaque ancêtre par fenêtre de 10SNP

Etape 4

Division du chromosome en sous-fenêtres non chevauchantes de 100kb. Le dosage allélique des fenêtres de 10SNP est reporté dans les fenêtres de 100Kb



TraceAncestor

Etape 3

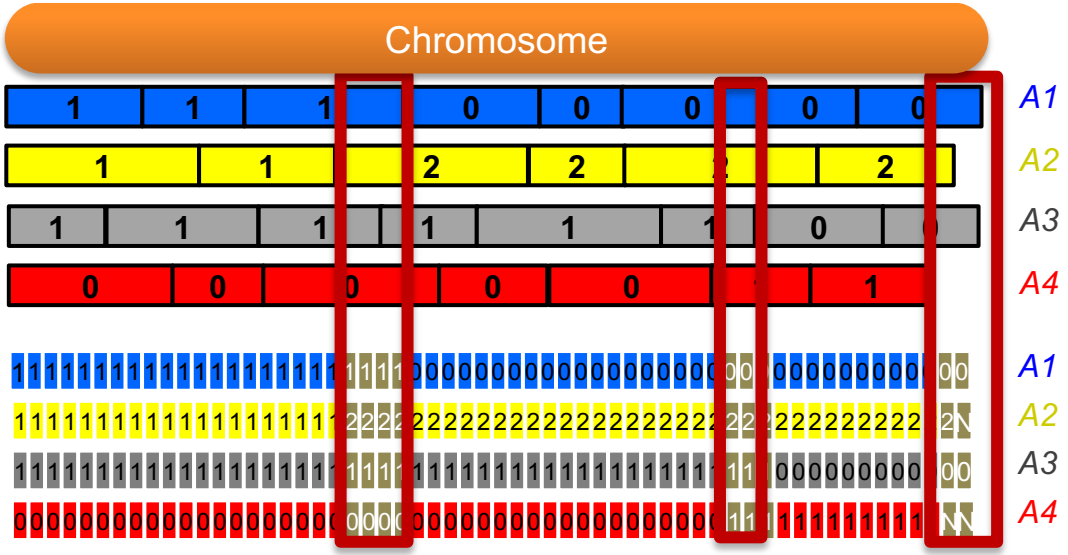
Estimation du dosage allélique de chaque ancêtre par fenêtre de 10SNP

Etape 4

Division du chromosome en sous-fenêtres non chevauchantes de 100kb. Le dosage allélique des fenêtres de 10SNP est reporté dans les fenêtres de 100Kb

Si la somme du dosage allélique de tous les ancêtres pour une fenêtre est différente de la ploïdie:

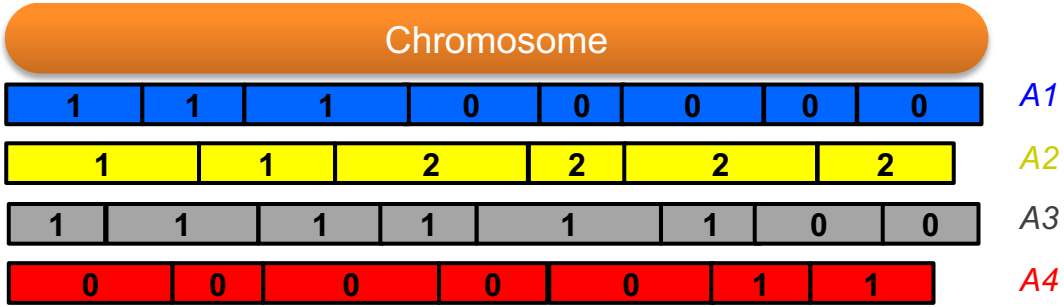
→ indetermination



TraceAncestor

Etape 3

Estimation du dosage allélique de chaque ancêtre par fenêtre de 10SNP

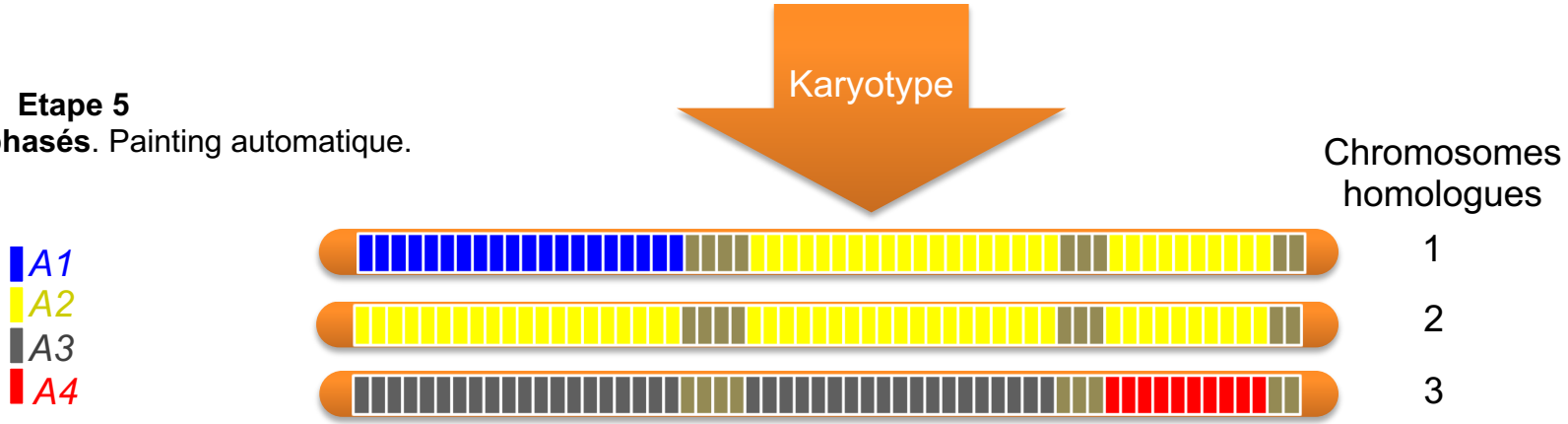


Etape 4

Division du chromosome en sous-fenêtres non chevauchantes de 100kb. Le dosage allélique des fenêtres de 10 SNPs est reporté dans les fenêtres de 100Kb

Etape 5

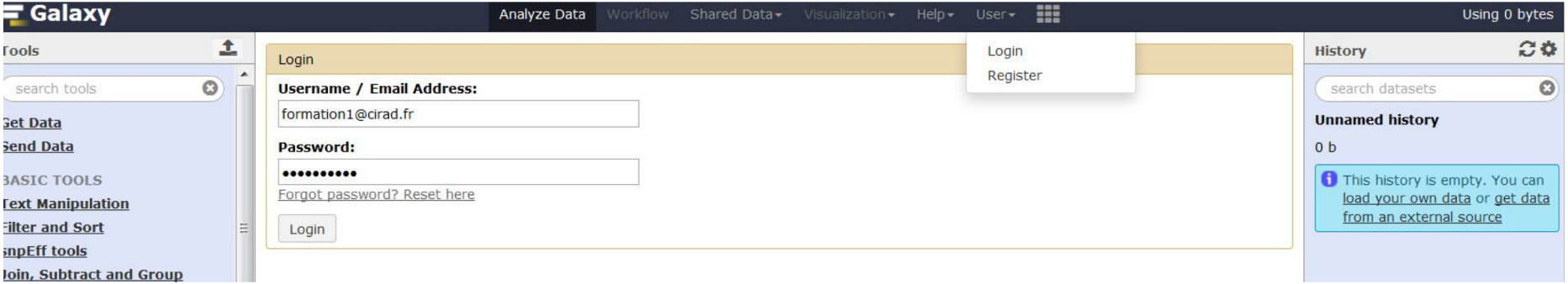
Caryotypes non-phasés. Painting automatique.



UTILISATION DE TRACE ANCESTOR SOUS GALAXY

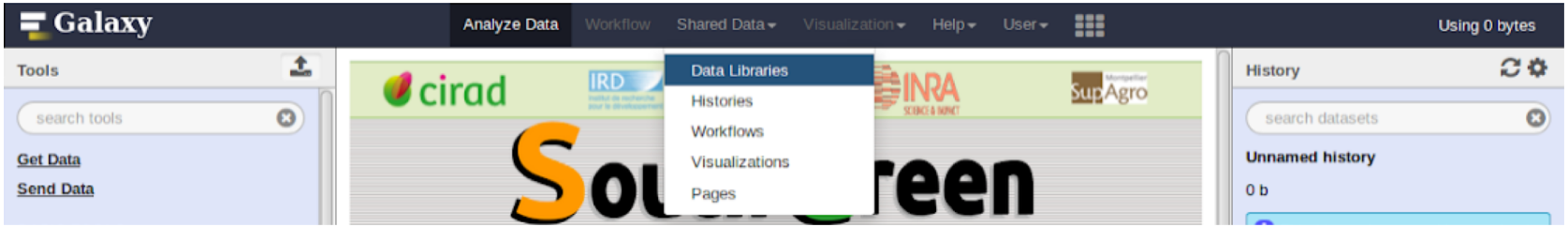
http://galaxy.southgreen.fr/galaxy/

ETAPE 1 : se connecter à galaxy

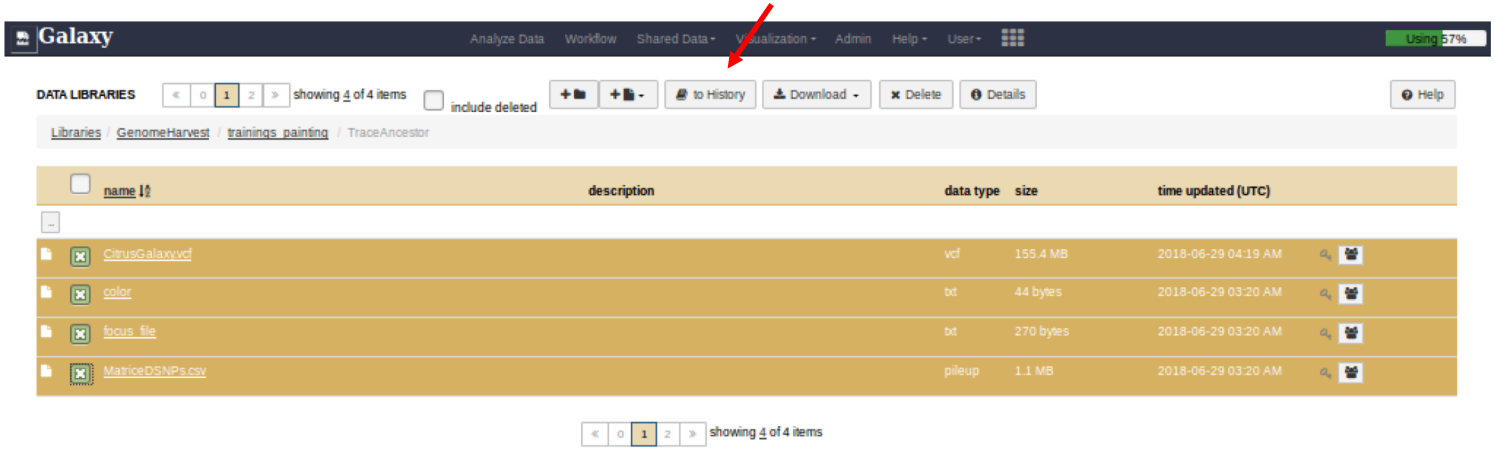


UTILISATION DE TRACE ANCESTOR SOUS GALAXY

ETAPE 2 : Charger les données tests de la librairie partagée “TraceAncestor” vers l’historique



DataLibrary → GenomeHarvest → trainings_painting → TraceAncestor



UTILISATION DE TRACE ANCESTOR SOUS GALAXY

Matrice contenant les données GST

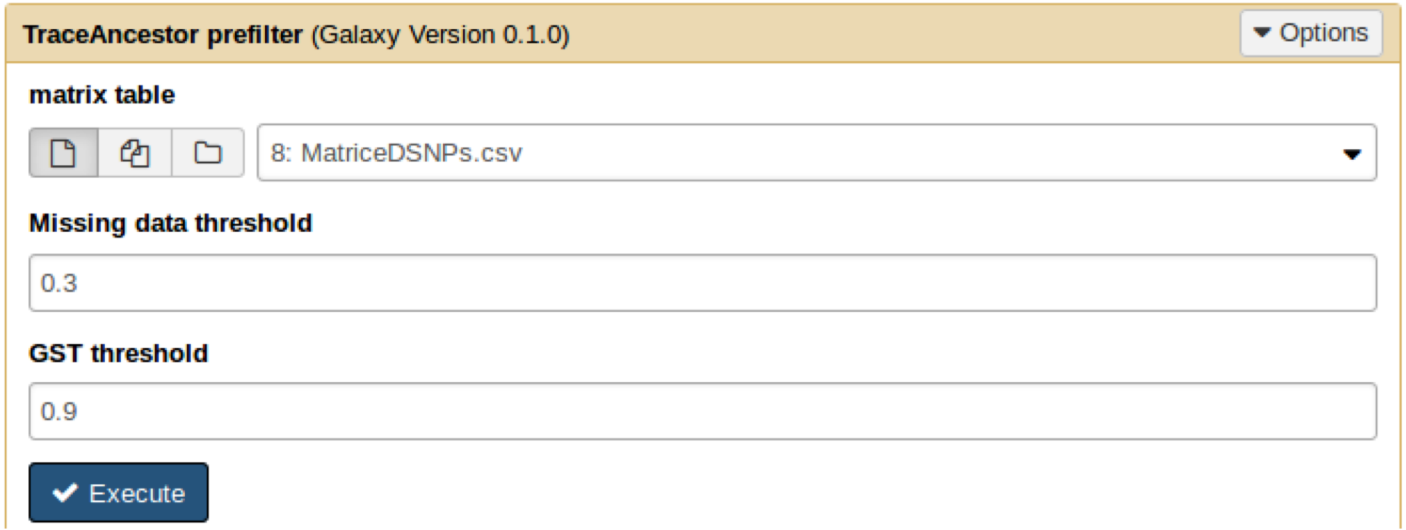
Fichier contenant des noms d'hybrides spécifiques sur lesquels on veut réaliser le painting

Association couleur - ancêtre

VCF des hybrides

UTILISATION DE TRACE ANCESTOR SOUS GALAXY

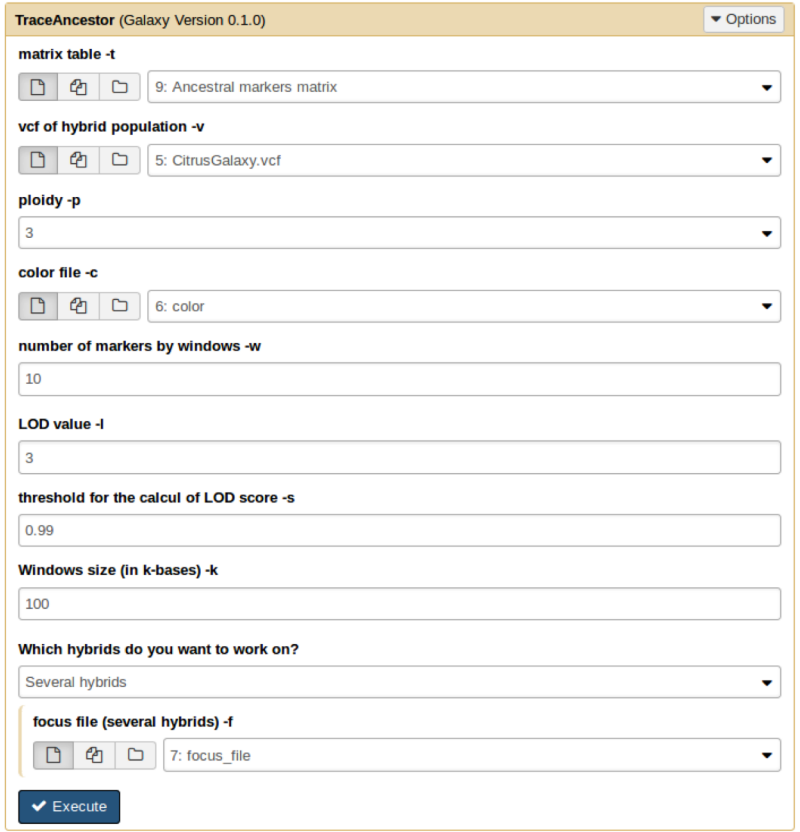
ETAPE 3 : Lancer TraceAncestor prefilter pour obtenir la matrice des marqueurs diagnostiques



→ Output : *Ancestral markers matrix* (matrice contenant les marqueurs diagnostiques filtrés)

UTILISATION DE TRACE ANCESTOR SOUS GALAXY

ETAPE 4 : Lancer TraceAncestor pour obtenir les fichiers de blocs de chromosomes à coloriser



- Nombre de marqueurs par fenêtres
- Valeur du LOD à partir de laquelle une hypothèse est acceptée
- Taux d'erreurs acceptée
- Taille des sous-fenêtres
- Choix du focus pour le painting:
  - Un individu
  - Plusieurs individus
  - Tous les individus

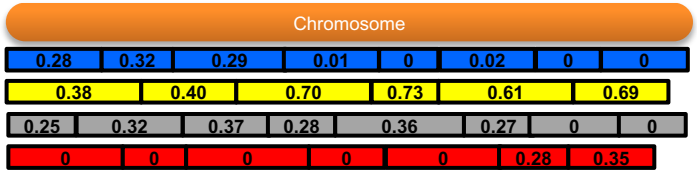
UTILISATION DE TRACE ANCESTOR SOUS GALAXY

OUTPUTS

→ Différents outputs si on choisit de faire le focus sur un seul individu

Ancestors frequency along the chromosomes

Hybrid	Ancestry	Chromosome	Position_Start	Position_End	Frequence
Sample61	A1	1	1	1029641	0.3038



Circos Painting

1	1	1700000	#DF0101	Sample61
1	1700001	2200000	#B9B9B9	Sample61

Chromosomes length for circos

1	28919326
2	36354460

Ideogram Painting

1	0	1	1700000	#DF0101
1	0	1700001	2200000	#B9B9B9

Chromosomes length for ideogram

Sample61	305908623	012
Sample62	305908623	012

UTILISATION DE TRACE ANCESTOR SOUS GALAXY

ETAPE 5 : Visualisation

App web circos : http://genomeharvest.southgreen.fr/visu/circosJS/demo/index.php

Circos

CircosJS CircosJS Client to build interactive graphs in a circular layout (Galaxy Version 0.0.1)

Values for Chromosome Length

169: Chromosomes length for circos

track

1: track

Track name

stack

Track type

Stack

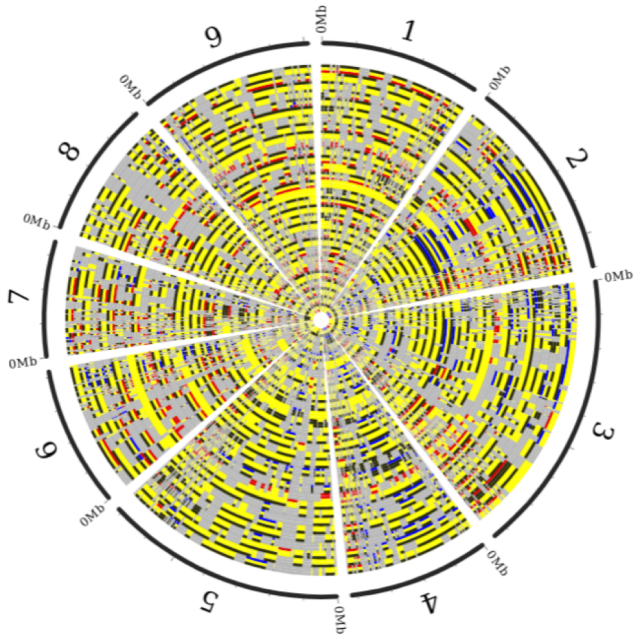
Track data

168: circos painting

Insert track

Track goes here

Execute

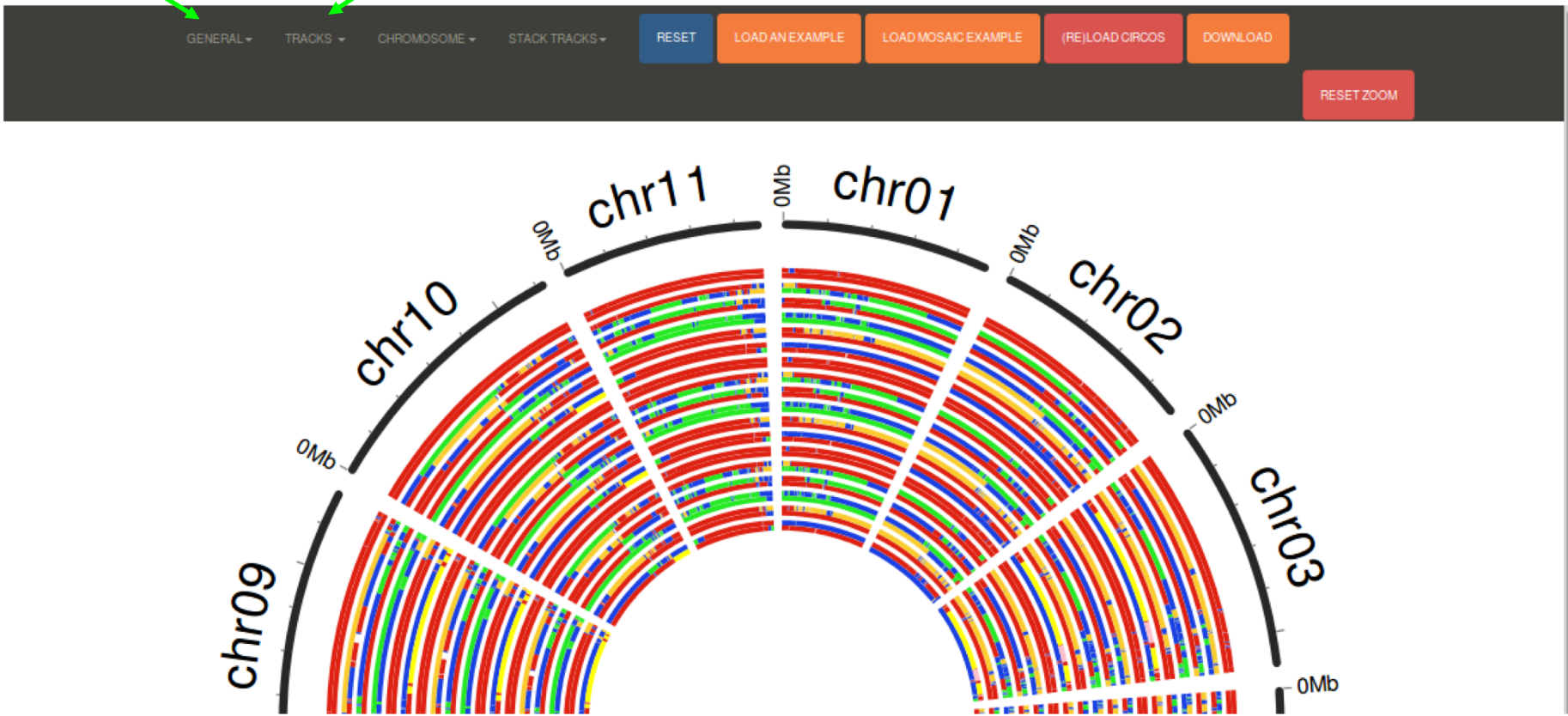


UTILISATION DE TRACE ANCESTOR SOUS GALAXY

Longueur des chromosomes

Chro - début - fin - couleur - individu

Télécharger en png



UTILISATION DE TRACE ANCESTOR SOUS GALAXY

ETAPE 5 : Visualisation

App web ideogram: http://genomeharvest.southgreen.fr/visu/ideogram/newindex.php

Ideogram

Ideogram Chromosome Painting (Galaxy Version 0.0.1)

Values for Chromosome Length

167: Chromosomes length for ideogram

chr length

Ancestral Blocks

166: Ideogram painting

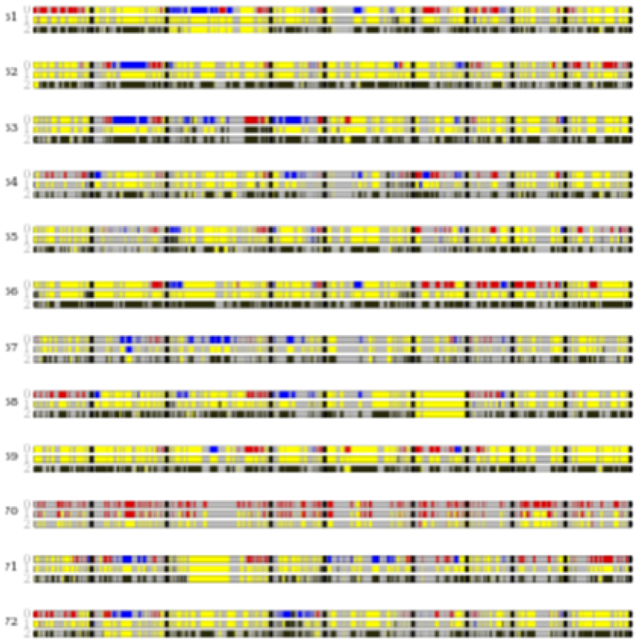
chr haplotype start end #color

Ploidy

3

ploidy

Execute



UTILISATION DE TRACE ANCESTOR SOUS GALAXY

Longueur des chromosomes

Chro - haplo - début - fin - couleur

Télécharger en png