

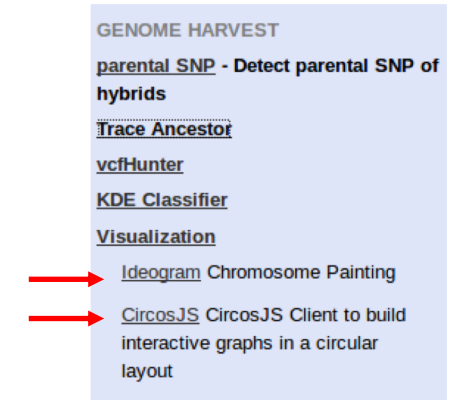
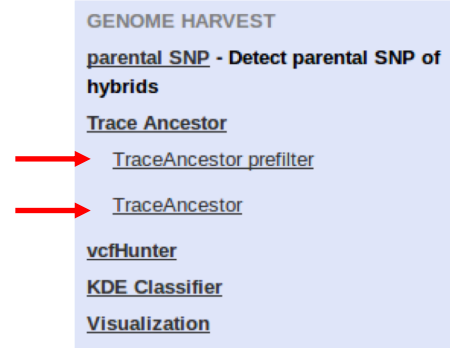
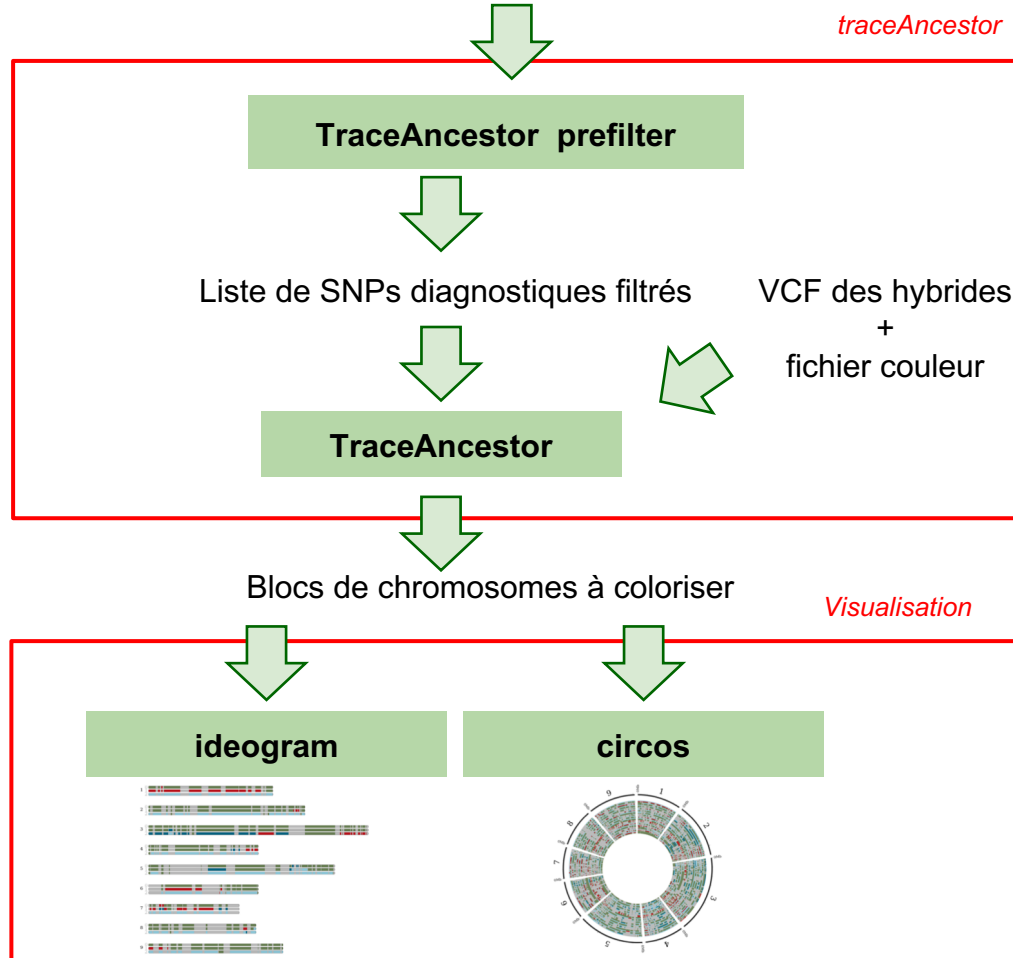


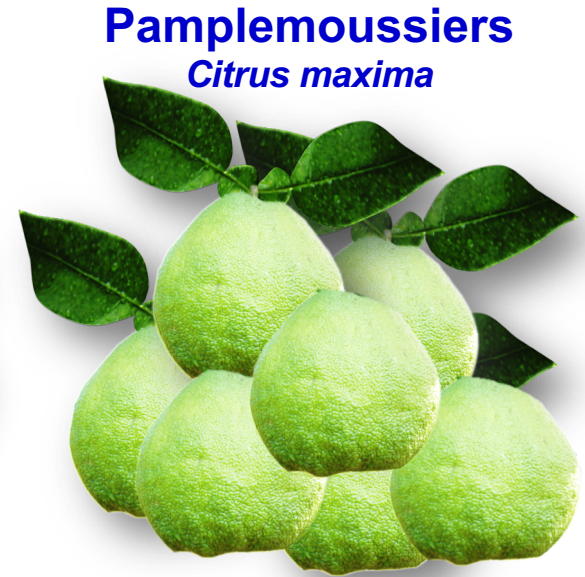
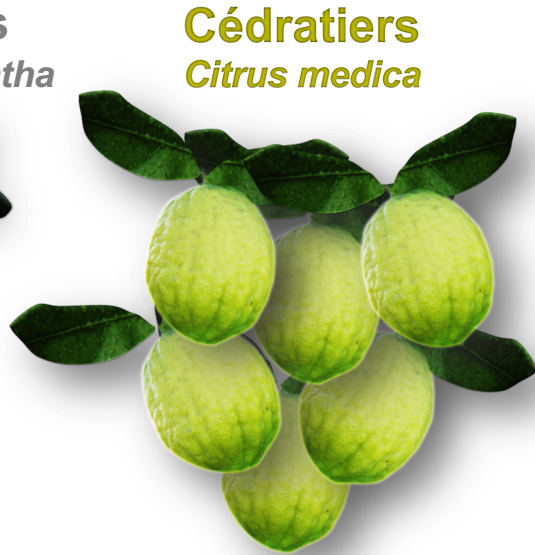
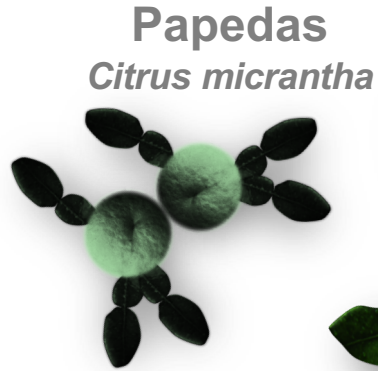
- Les parents/ancêtres sont déjà identifiés (pamplemousse, mandarinier, cédrat, micrantha)
- **Méthode** : Pour un individu donné, mesurer fenêtre par fenêtre sur un chromosome la fréquence de présence de SNPs ancestraux.

2 outils dans galaxy :

- **TraceAncestor prefilter**
- **TraceAncestor**

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





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:

$$G_{st} = \frac{H_{eTot} - H_s}{H_{eTot}} = \frac{H_{eTot} - \sum h_e / n}{H_{eTot}}$$

Données
manquantesIndices 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

Données
manquantesIndices 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 $F > 0.8 \rightarrow \text{ALT}$
 - Si $F < 0.2 \rightarrow \text{REF}$



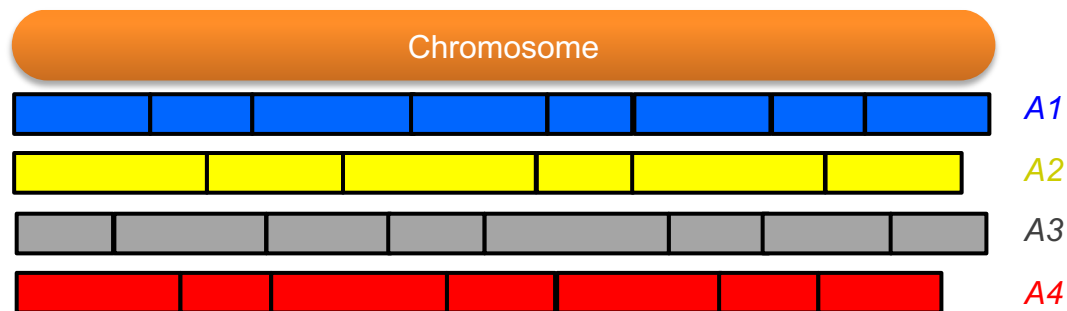
#CHROM	POS	REF	ALT	%Nref	GSTA1	GSTA2	GSTA3	GSTA4	FA1	FA2	FA3	FA4
1	85524	A	G	0.3103448276	0.2	0.2	0.2	1	0	0	0	1
1	108710	A	T	0.6206896552	0.2	1	0.2	0.2	0	1	0	0
1	108741	T	A	0.2413793103	0.2	0.2	1	0.2	0	0	1	0
1	109226	A	T	0	0.2	0.2	0.2	1	0	0	0	1
1	109661	A	G	0.3448275862	0.2	0.2	1	0.2	0	0	1	0
1	110915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

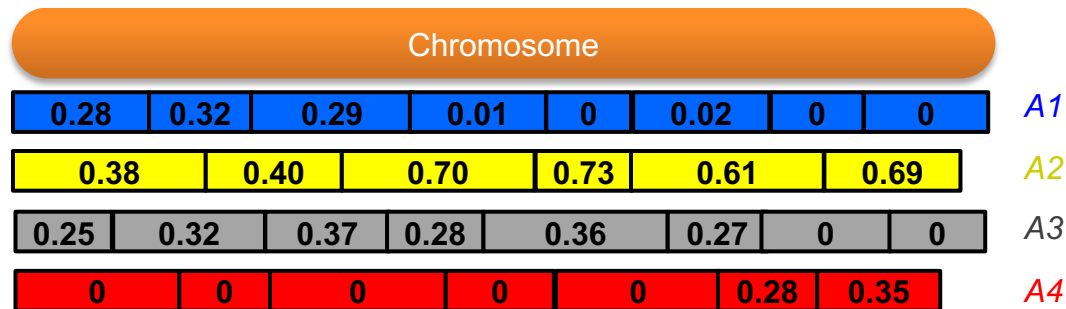
Découpage du chromosome en fenêtres
non chevauchantes de 10 SNPs



TraceAncestor

Etape 2

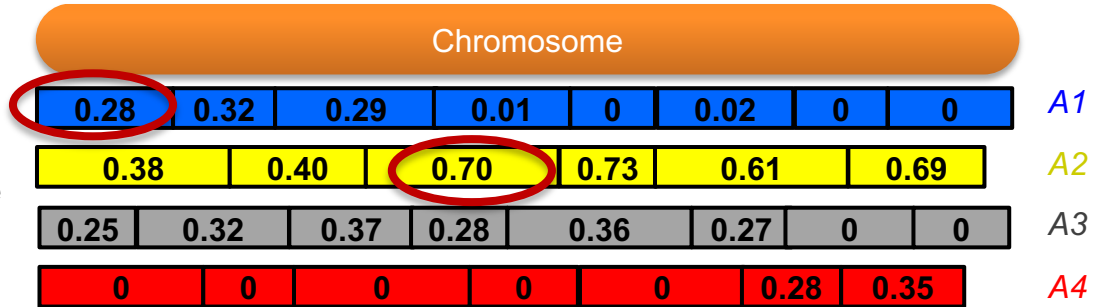
Calcul de la fréquence des reads ancestraux par ancêtre et par fenêtre de 10 SNPs.



TraceAncestor

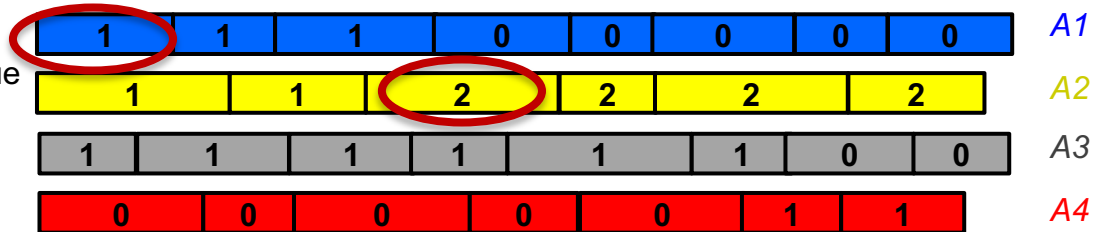
Etape 2

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 $(-3 < \text{LOD} < 3) \Rightarrow$ indétermination

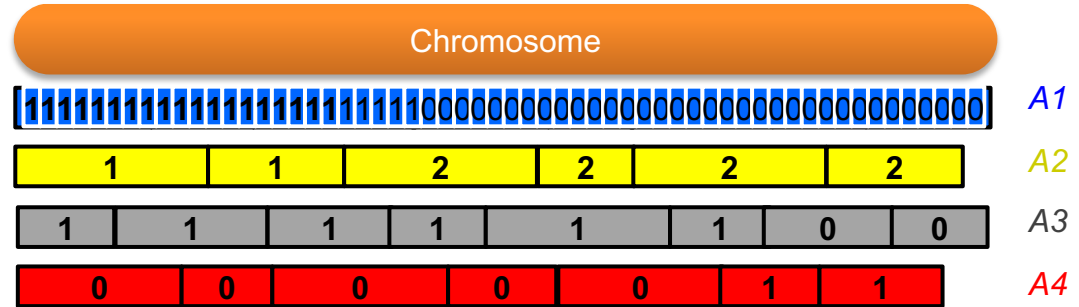
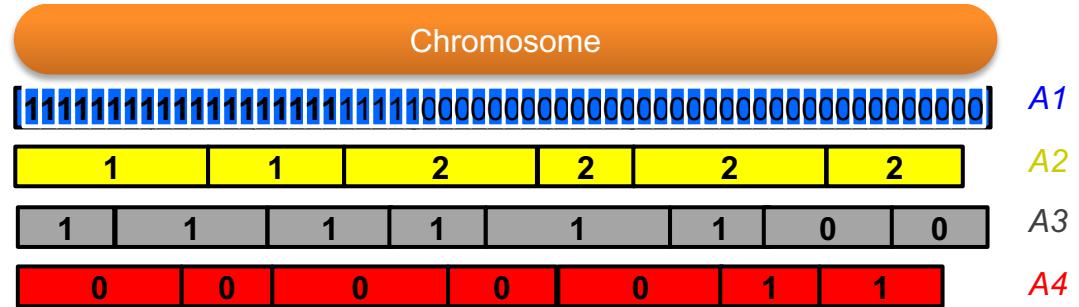
TraceAncestor

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



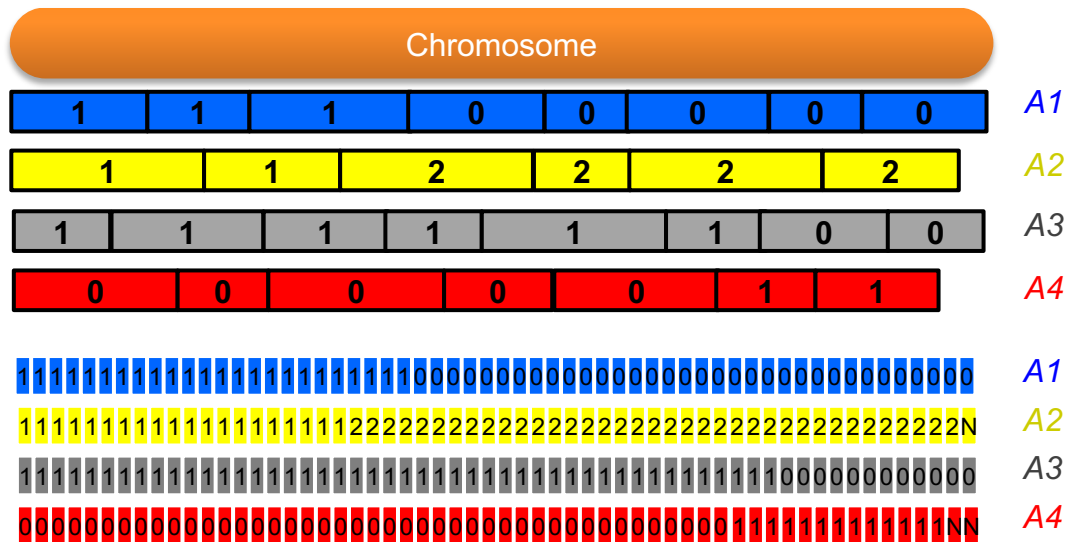
TraceAncestor

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

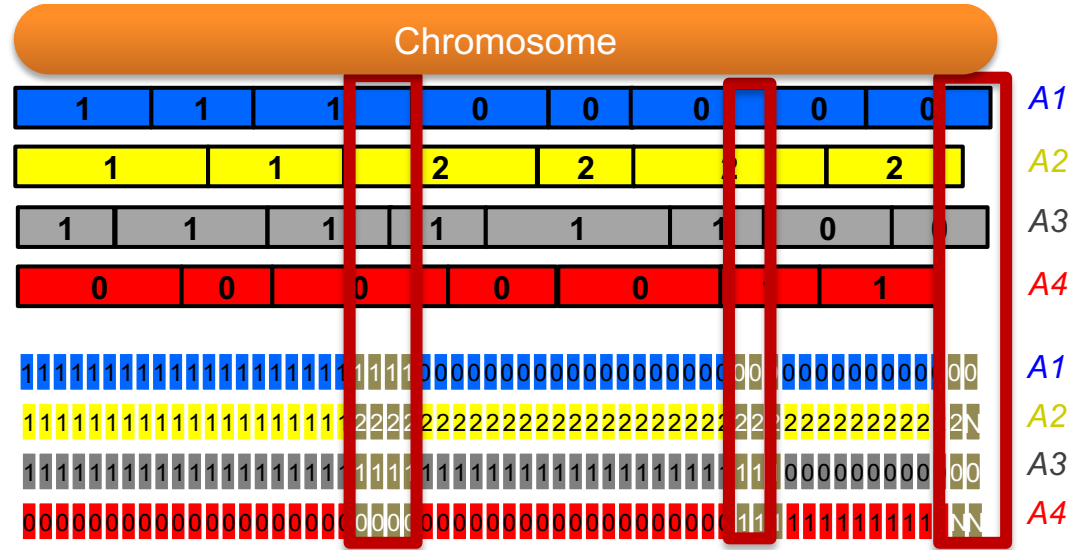
Estimation du dosage allélique de chaque ancêtre par fenêtre de 10SNP

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



UTILISATION DE TRACE ANCESTOR SOUS GALAXY

<http://galaxy.southgreen.fr/galaxy/>

ETAPE 1 : se connecter à galaxy

The screenshot displays the Galaxy web interface. At the top, a dark navigation bar contains the 'Galaxy' logo and several menu items: 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The 'User' menu is open, showing 'Login' and 'Register' options. On the right of the navigation bar, it indicates 'Using 0 bytes'. The main content area is divided into three sections. On the left is a 'Tools' sidebar with a search bar and a list of tool categories including 'Get Data', 'Send Data', 'BASIC TOOLS', 'Text Manipulation', 'Filter and Sort', 'mpEff tools', and 'Join, Subtract and Group'. The central section is titled 'Login' and contains a form with fields for 'Username / Email Address' (filled with 'formation1@cirad.fr') and 'Password' (masked with dots). Below the password field are links for 'Forgot password?' and 'Reset here', and a 'Login' button. On the right is a 'History' sidebar with a search bar and a message stating 'Unnamed history' and '0 b'. A blue information box in the history sidebar explains that the history is empty and provides links to 'load your own data' or 'get data from an external source'.

Galaxy Analyze Data Workflow Shared Data Visualization Help User Using 0 bytes

Tools search tools

[Get Data](#)
[Send Data](#)
BASIC TOOLS
[Text Manipulation](#)
[Filter and Sort](#)
[mpEff tools](#)
[Join, Subtract and Group](#)

Login

Username / Email Address:
formation1@cirad.fr

Password:
••••••••

[Forgot password?](#) [Reset here](#)

Login

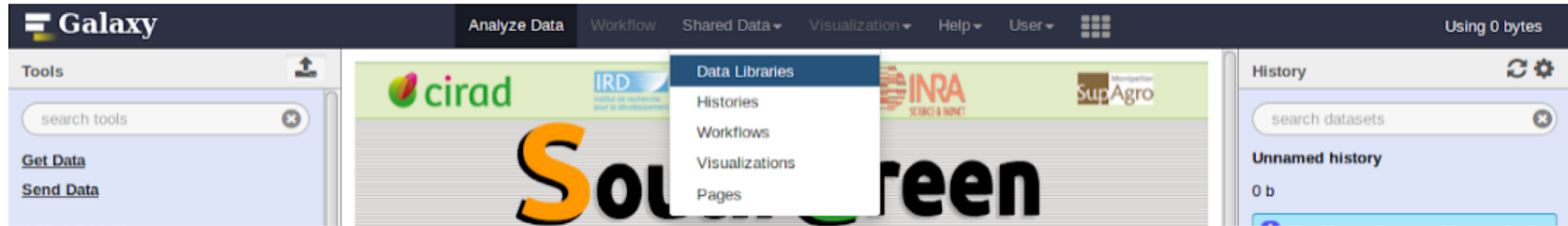
History search datasets

Unnamed history
0 b

i This history is empty. You can [load your own data](#) or [get data from an external source](#)

UTILISATION DE TRACE ANCESTOR SOUS GALAXY

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

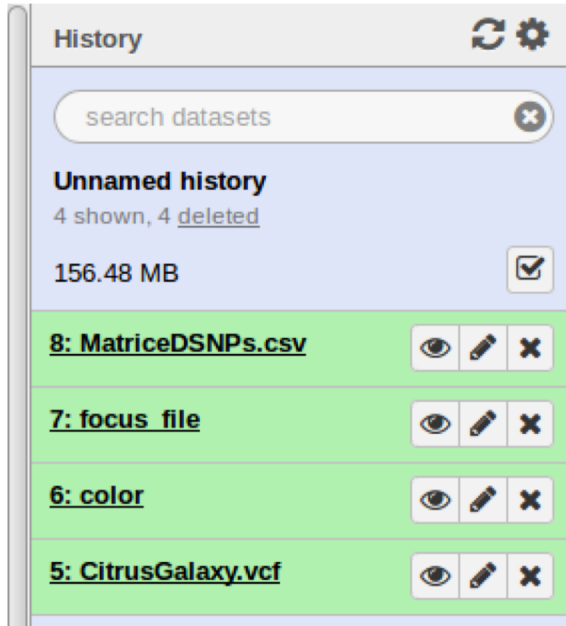


DataLibrary → GenomeHarvest → trainings_painting → TraceAncestor

The screenshot shows the 'Data Libraries' page in Galaxy. The breadcrumb navigation is 'Libraries / GenomeHarvest / trainings_painting / TraceAncestor'. A red arrow points to the 'to History' button. Below the navigation, a table lists the data items in the 'TraceAncestor' library.

<input type="checkbox"/>	name	description	data type	size	time updated (UTC)		
<input type="checkbox"/>	CircusGalaxy.vcf		vcf	155.4 MB	2018-06-29 04:19 AM		
<input type="checkbox"/>	color		txt	44 bytes	2018-06-29 03:20 AM		
<input type="checkbox"/>	focus_file		txt	270 bytes	2018-06-29 03:20 AM		
<input type="checkbox"/>	MatriceDSNPs.csv		pileup	1.1 MB	2018-06-29 03:20 AM		

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

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

TraceAncestor prefilter (Galaxy Version 0.1.0) Options


matrix table



 8: MatriceDSNPs.csv

Missing data threshold

GST threshold

 **Execute**

→ 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

TraceAncestor (Galaxy Version 0.1.0) Options

matrix table -t
9: Ancestral markers matrix

vcf of hybrid population -v
5: CitrusGalaxy.vcf

ploidy -p
3

color file -c
6: color

number of markers by windows -w
10

LOD value -l
3

threshold for the calcul of LOD score -s
0.99

Windows size (in k-bases) -k
100

Which hybrids do you want to work on?
Several hybrids

focus file (several hybrids) -f
7: focus_file

✓ Execute

← 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

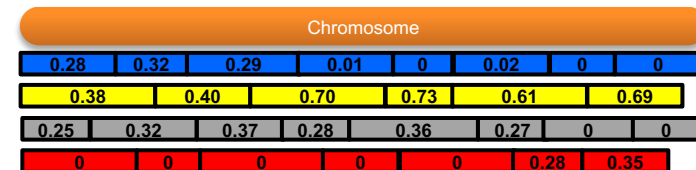
OUTPUTS



→ 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
Sample6 1	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

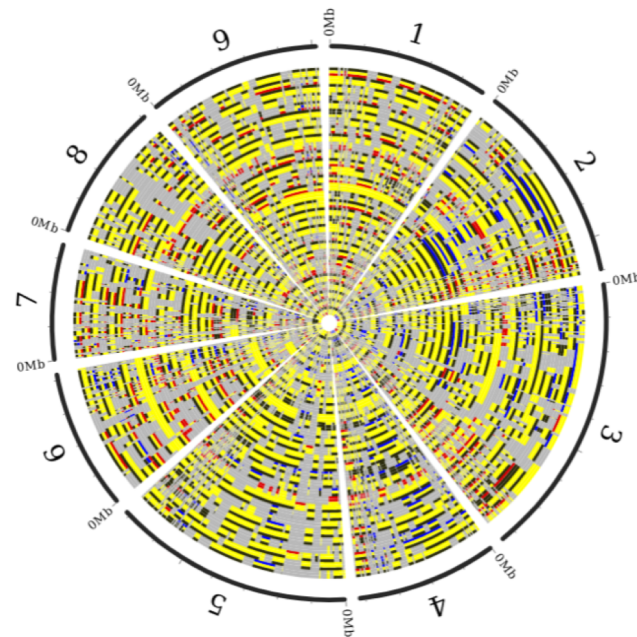
Track data

   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

GENERAL ▾

TRACKS ▾

CHROMOSOME ▾

STACK TRACKS ▾

RESET

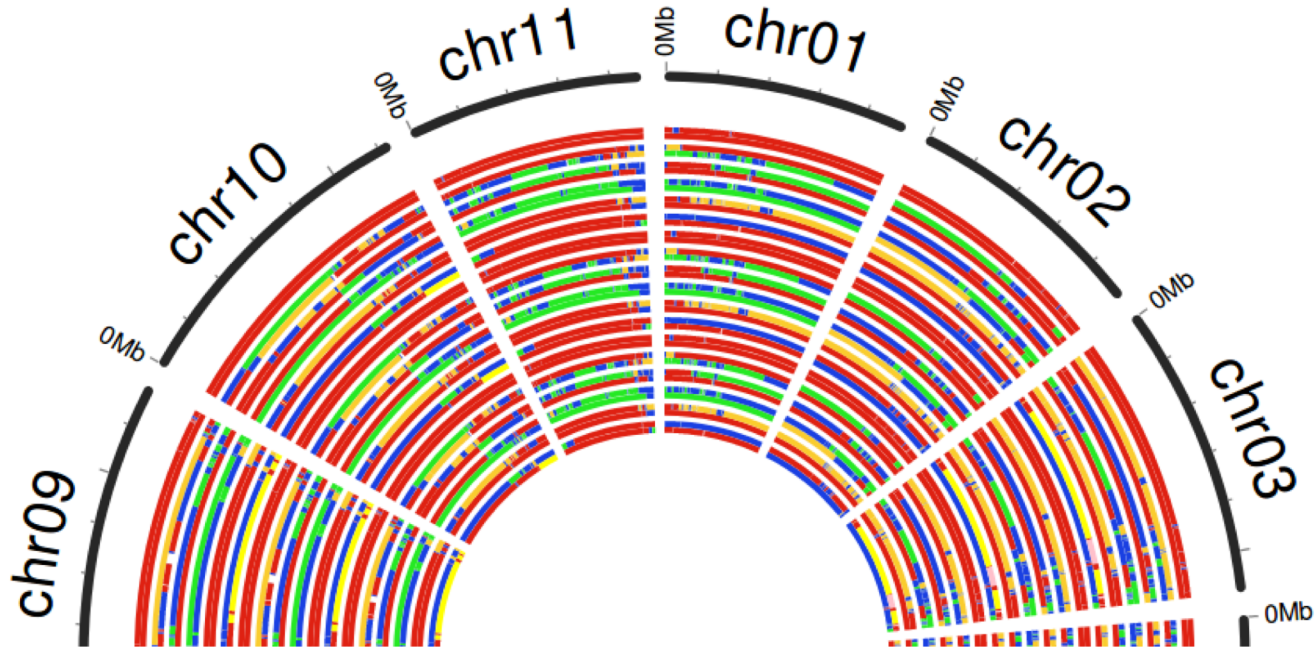
LOAD AN EXAMPLE

LOAD MOSAIC EXAMPLE

(RE)LOAD CIRCOS

DOWNLOAD

RESET ZOOM



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

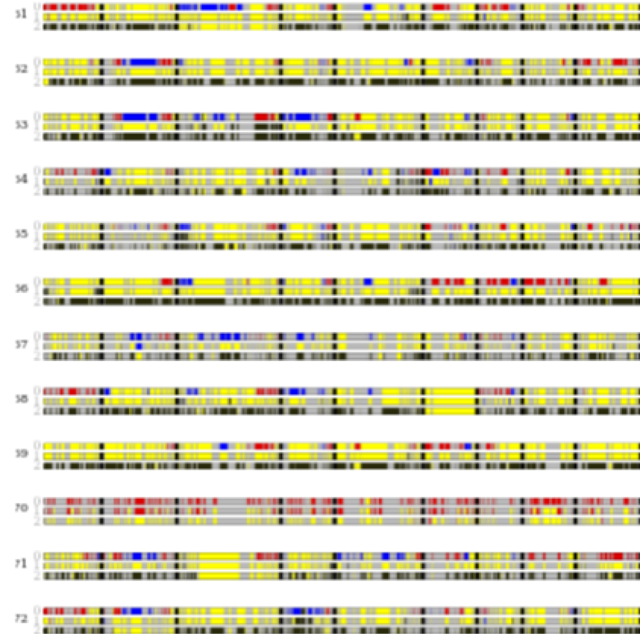
chr haplotype start end #color

Ploidy

3

ploidy

✓ Execute



UTILISATION DE TRACE ANCESTOR SOUS GALAXY

Longueur des chromosomes

Chro - haplo - début - fin - couleur

Télécharger en png

