

## Creating an history: All analysis will be found in this history

The screenshot shows the Galaxy web interface at <http://galaxy.southgreen.fr/galaxy/>. The interface has a dark blue header with the Galaxy logo and a search bar. A green banner at the top center says "0 datasets have been deleted permanently". On the left, there's a sidebar with a "Tools" section containing links like "Get Data", "Send Data", and various analysis categories such as "BASIC TOOLS", "SEQUENCE ANALYSIS", "NGS ANALYSIS", "SNP ANALYSIS", and "Rice Variant Analysis". On the right, a "History" panel is open, showing a list of actions: "HISTORY LISTS", "Saved Histories", "Histories Shared with Me", "HISTORY ACTIONS", and a dropdown menu where "Create New" is selected. Other options in the dropdown include "Copy History", "Share or Publish", "Show Structure", "Extract Workflow", "Delete", and "Delete Permanently". Below these are sections for "DATASET ACTIONS", "DOWNLOADS", and "OTHER ACTIONS".

# Renaming the History (easier to find analysis)

The screenshot shows the Galaxy web interface at [galaxy.southgreen.fr/galaxy/](http://galaxy.southgreen.fr/galaxy/). The left sidebar contains a tree view of available tools categorized by analysis type: BASIC TOOLS, SEQUENCE ANALYSIS, NGS ANALYSIS, SNP ANALYSIS, and others. The main workspace is currently empty. The right panel is titled "History" and displays a message: "This history is empty. You can load your own data or get data from an external source". A button labeled "Click to rename history" is present.

# Loading existing dataset: To test tools

The screenshot shows the Galaxy bioinformatics platform interface. At the top, there is a navigation bar with links for Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, User, and a search bar for 'Galaxy southgreen'. A red circle labeled '1' is positioned above the 'Shared Data' button. Below the navigation bar, the main content area features the South Green bioinformatics platform branding. A red circle labeled '2' is positioned above the 'Data Libraries' section in the dropdown menu. The 'Data Libraries' menu includes options for Histories, Workflows, Visualizations, and Pages. The central part of the screen displays a 'GWAS' workflow card. It includes a scatter plot visualization, a brief description of the 'SNIPlay3 GWAS workflow' (Tassel-based GWAS workflow (GLM model) including population structure and correction for structure (Dereeper et al, 2015)), an input requirement ('VCF + Phenotypic tabulated file'), and a 'Access workflow' button. To the left and right of the central card are other workflow cards for NGS analyses, Structural variations, and various genomic analysis tools. At the bottom, there is a section titled 'How to load big datasets?' with a diagram showing a flow from 'My file' through an HPC Cluster to 'My Galaxy History', and a note about tool availability.

1

2

Galaxy

analyze Data Workflow Shared Data Visualization Admin Help User

Galaxy southgreen

Tools

Get Data Send Data

BASIC TOOLS

Text Manipulation Filter and Sort Join, Subtract and Group Convert Formats

SEQUENCE ANALYSIS

Gene/Protein prediction

EMBOSS

Operate on Genomic Intervals

Fetch Sequences

Genomics

Fetch Alignments

Extract Features

NGS ANALYSIS

NGS: Quality Control

NGS : Mapping

NGS: GATK Tools

NGS: GATK2 Tools

NGS: SAM/BAM Manipulations

NGS: RNASeq

NGS: Assembly

NGS: Small RNAs

Bedtools

Picard Tools

SNP ANALYSIS

NGS: SNP Calling

VarScan

Population structure

GWAS

VCFtools

Tassel GBS (Version 4.0)

Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs))

cirad

IRD

INRA

Montpellier SupAgro

Welcome to GALAXY

Our pre-configured and validated workflows

NGS analyses

SNP calling

SNP analysis

GWAS

GWAS

SNIPlay3 GWAS workflow. Tassel-based GWAS workflow (GLM model) including population structure and correction for structure (Dereeper et al, 2015)

Input: VCF + Phenotypic tabulated file

Access workflow

Structural variations

Chrom. reconstruction

Metagenomics

Gene families

Mosaic Genomes

These workflows as part of the services provided by South Green

How to load big datasets?

1 My file → HPC Cluster → 2 My Galaxy History

Choose FTP file

In order to figure out which tools were made available by our team, please activate the "tool search" functionality from the Options drop-down and type "south green" in the lookup filter.

History

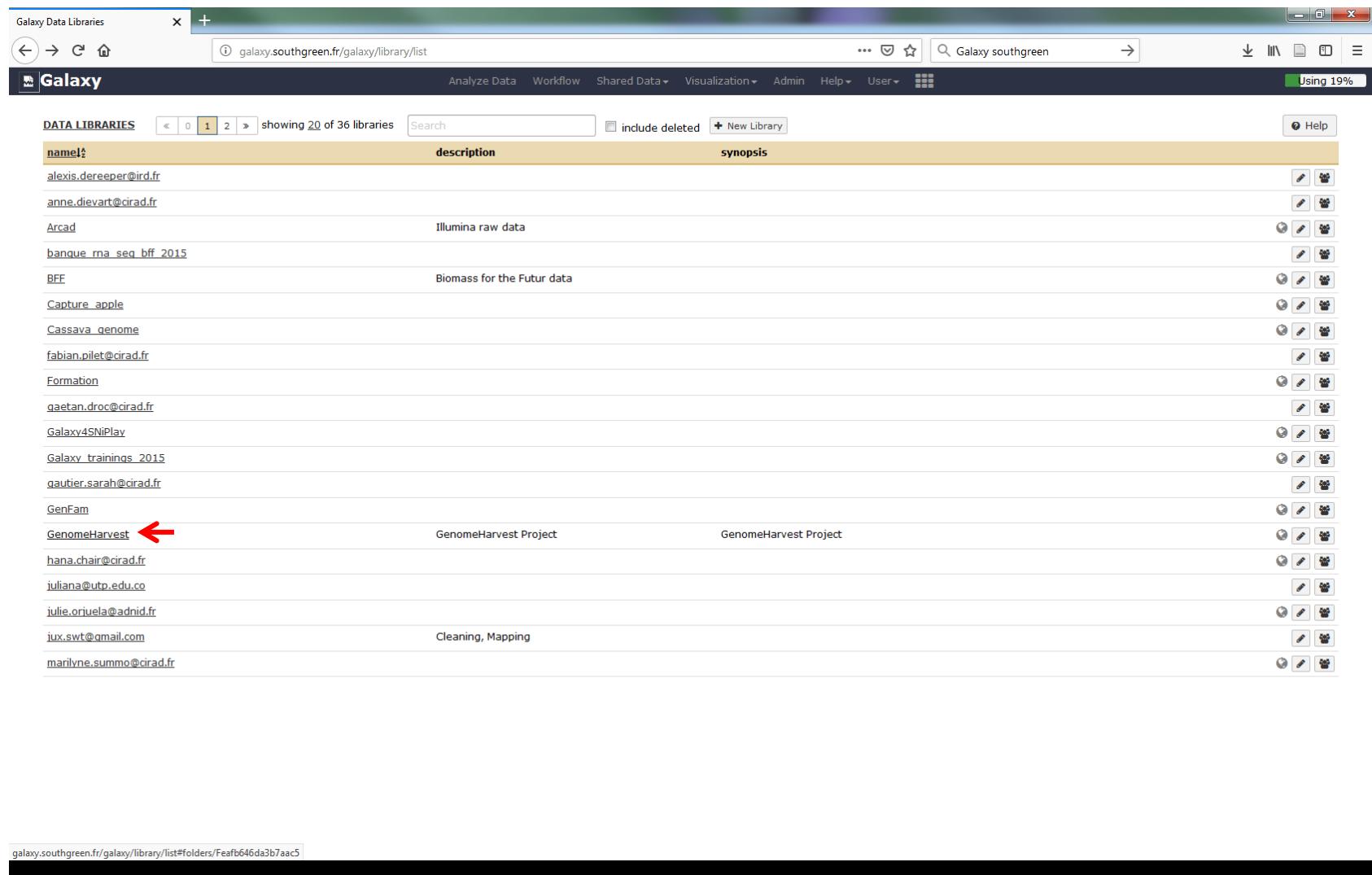
search datasets

VCFHunter

0 b

This history is empty. You can load your own data or get data from an external source

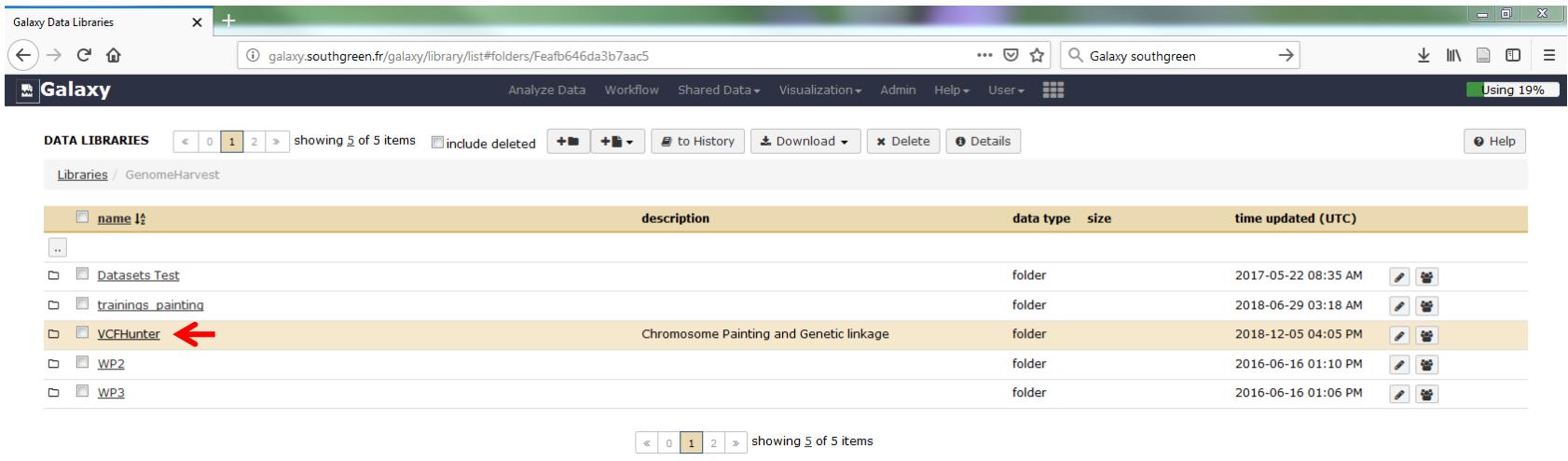
# Loading existing dataset: To test tools



The screenshot shows the Galaxy web interface with the URL `galaxy.southgreen.fr/galaxy/library/list` in the address bar. The page displays a list of 36 data libraries, with the 20th library, "GenomeHarvest", highlighted by a red arrow. The columns in the table are "name", "description", and "synopsis". The "GenomeHarvest" entry has a "description" of "GenomeHarvest Project" and a "synopsis" of "GenomeHarvest Project". The interface includes a navigation bar with links like Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, User, and a search bar.

name	description	synopsis
alexis.dereeper@ird.fr		
anne.dievert@cirad.fr		
Arcad	Illumina raw data	
banque_rna_seq_bff_2015		
BFF	Biomass for the Futur data	
Capture_apple		
Cassava_genome		
fabian.pilet@cirad.fr		
Formation		
gaetan.droc@cirad.fr		
Galaxy4SNPlay		
Galaxy_trainings_2015		
gautier.sarah@cirad.fr		
GenFam		
GenomeHarvest	GenomeHarvest Project	GenomeHarvest Project
hana.chair@cirad.fr		
juliana@utp.edu.co		
julie.oruuela@adnid.fr		
jux.swt@gmail.com	Cleaning, Mapping	
marilyne.summo@cirad.fr		

# Loading existing dataset: To test tools



The screenshot shows the Galaxy Data Libraries interface. The URL in the address bar is `galaxy.southgreen.fr/galaxy/library/list#folders/Feafb646da3b7aac5`. The search bar contains "Galaxy southgreen". The main content area displays a list of data libraries:

name	description	data type	size	time updated (UTC)	actions
Datasets Test		folder		2017-05-22 08:35 AM	
trainings_painting		folder		2018-06-29 03:18 AM	
VCFHunter	Chromosome Painting and Genetic linkage	folder		2018-12-05 04:05 PM	
WP2		folder		2016-06-16 01:10 PM	
WP3		folder		2016-06-16 01:06 PM	

A red arrow points to the "VCFHunter" folder in the list.

The bottom status bar shows the URL `galaxy.southgreen.fr/galaxy/library/list#folders/F52d6bdfafedbb5e5`.

# Loading existing dataset: To test tools

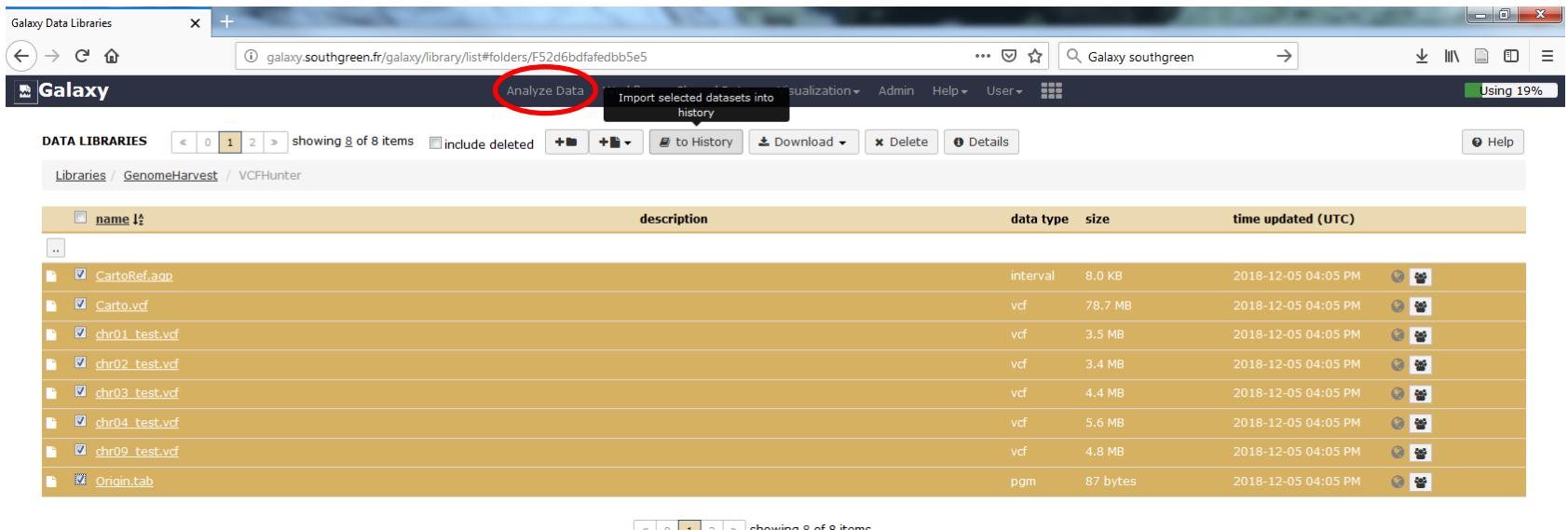
The screenshot shows the Galaxy web interface with the following details:

- Header:** Galaxy Data Libraries, galaxy.southgreen.fr/galaxy/library/list#folders/F52d6bdfafedbb5e5, Galaxy southgreen, Using 19%.
- Top Bar:** Analyze Data, Import selected datasets into history (highlighted), Visualization, Admin, Help, User, Grid View, Help.
- Section Headers:** DATA LIBRARIES, Libraries / GenomeHarvest / VCFHunter.
- Table Headers:** name, description, data type, size, time updated (UTC).
- Data:** A list of 8 items, all checked (indicated by a blue checkmark icon). The items are: CartoRef.app (interval, 8.0 KB, 2018-12-05 04:05 PM), Carto.vcf (vcf, 78.7 MB, 2018-12-05 04:05 PM), chr01\_test.vcf (vcf, 3.5 MB, 2018-12-05 04:05 PM), chr02\_test.vcf (vcf, 3.4 MB, 2018-12-05 04:05 PM), chr03\_test.vcf (vcf, 4.4 MB, 2018-12-05 04:05 PM), chr04\_test.vcf (vcf, 5.6 MB, 2018-12-05 04:05 PM), chr09\_test.vcf (vcf, 4.8 MB, 2018-12-05 04:05 PM), and Origin.tab (pgm, 87 bytes, 2018-12-05 04:05 PM).
- Pagination:** showing 8 of 8 items.

Red annotations:

- ① Points to the "To History" button at the top of the list.
- ② Points to the "Import selected datasets into history" button in the top navigation bar.

# Going back to history to analyze data



The screenshot shows the Galaxy Data Libraries interface. At the top, there is a navigation bar with a search bar containing "Galaxy southgreen". Below the navigation bar is a toolbar with various icons. A red circle highlights the "Analyze Data" button in the toolbar. The main area displays a table of datasets:

name	description	data type	size	time updated (UTC)
CartoRef.app		interval	8.0 KB	2018-12-05 04:05 PM
Carto.vcf		vcf	78.7 MB	2018-12-05 04:05 PM
chr01_test.vcf		vcf	3.5 MB	2018-12-05 04:05 PM
chr02_test.vcf		vcf	3.4 MB	2018-12-05 04:05 PM
chr03_test.vcf		vcf	4.4 MB	2018-12-05 04:05 PM
chr04_test.vcf		vcf	5.6 MB	2018-12-05 04:05 PM
chr09_test.vcf		vcf	4.8 MB	2018-12-05 04:05 PM
Origin.tab		pgm	87 bytes	2018-12-05 04:05 PM

At the bottom of the table, there is a page navigation bar showing "showing 8 of 8 items".

# Going back to history

The screenshot shows the Galaxy bioinformatics platform interface on a web browser. The URL is [galaxy.southgreen.fr/galaxy/](http://galaxy.southgreen.fr/galaxy/). The page title is "Galaxy". The top navigation bar includes "Analyze Data", "Workflow", "Shared Data", "Visualization", "Admin", "Help", "User", and a "History" section.

The main content area features the "SouthGreen bioinformatics platform" logo, which includes logos for cirad, IRD, Biodiversity, INRA, and SupAgro. Below the logo, it says "Welcome to GALAXY" and "Our pre-configured and validated workflows".

The left sidebar contains a list of tools categorized under various sections:

- NGS: Quality Control**
- NGS: Mapping**
- NGS: GATK Tools**
- NGS: GATK2 Tools**
- NGS: SAM/BAM Manipulations**
- NGS: RNASeq**
- NGS: Assembly**
- NGS: Small RNAs**
- Bedtools**
- Picard Tools**
- SNP ANALYSIS**
- NGS: SNP Calling**
- VarScan**
- Population structure**
- GWAS**
- VCFtools**
- Tassel GBS (Version 4.0)**
- Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs))**
- GENOME HARVEST**
- TransPo-RG Transfer of Position to Resequenced Genome**
- parental SNP - Detect parental SNP of hybrids**
- Visualization**
- TraceAncestor**
- vcfHunter**
- KDE\_classifier**
- METAGENOMICS**
- FROGS**
- EVOLUTION/PHYLOGENY**
- Comparative Genomics**
- NCBI BLAST+**
- Genfam**
- Protein analyses**
- STATISTICS/GRAFICS**
- Statistics**

The central area displays a heatmap visualization titled "Chromosome reconstruction". It includes a description of the "Scaffehunter" tool, which assembles scaffolds into pseudomolecules using markers genotyped in a population (Martin et al, 2016). The input is Fastq + FASTA. A button "Access workflow" is present.

The right sidebar shows the "History" section with a list of loaded datasets:

- VCFHunter 8 shown
- 100.33 MB
- 8: Origin.tab
- 7: chr09\_test.vcf
- 6: chr04\_test.vcf
- 5: chr03\_test.vcf
- 4: chr02\_test.vcf
- 3: chr01\_test.vcf
- 2: Carto.vcf
- 1: CartoRef.agp

A red box highlights the list of datasets, and the text "Loaded datasets" is overlaid in red at the bottom right of the box.

# Chromosome painting with vcfHunter tool

Developed to answer banana problematics



*Musa balbisiana*  
2n = 2x = 22  
1 to 1.2pg

B genome



*Musa acuminata*  
2n = 2x = 22  
1.1 to 1.3pg

A genome



Nearly half of the banana production worldwide rely on interspecific hybrids of various ploidy (**AB**, **AAB**, **ABB**, **AAAB**).

**What is the composition of A and B genomes along chromosomes of cultivated banana hybrids?**

# Chromosome painting with vcfHunter tool

What is the contribution of ancestral genomes along chromosomes of cultivated hybrids?

Several tools developed under vcfHunter toolbox for this purpose:

The screenshot shows the Galaxy web interface running on the SouthGreen bioinformatics platform. The left sidebar contains a list of tools, with 'vcfHunter' highlighted by a red circle labeled '1'. The main content area displays a 'SNP calling' workflow, showing a visualization of chromatograms and a 'How to load big datasets?' section. The right sidebar shows a history of 8 VCFHunter datasets.

**Left Sidebar (Tools):**

- NGS: SNP Calling
- Varscan
- Population structure
- GWAS
- VCFTools
- Tassel GBS (Version 4.0)
- Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs))
- GENOME HARVEST
- TransPoRG Transfer of Position to Resequenced Genome
- parental SNP - Detect parental SNP of hybrids
- Visualization
- TraceAncestor
- vcfHunter** (highlighted with red circle 1)
- VCF Filter
- vcf2allPropAndCov (highlighted with red box 2)
- vcf2allPropAndCovByChr
- vcf2popNew
- RecombCalculatorDDose
- Draw\_dot\_plot
- KDE\_classifier
- METAGENOMICS
- FROGS
- EVOLUTION/PHYLOGENY
- Comparative Genomics
- NCBI BLAST+
- Genfam
- Protein analyses
- STATISTICS/GRAFICS
- Statistics
- Graph/Display Data
- SOUTHGREEN PROJECTS
- SNiPlay3
- GNAnt Tools

**Main Content Area:**

Welcome to **GALAXY**

Our pre-configured and validated workflows

**SNP calling**

The SNP Calling is based on the **GATK toolkit**, using either UnifiedGenotyper or HaplotypeCaller module.

**Input:** BAM alignment files + FASTA for reference  
**Output:** VCF (Variant call Format) file

**Access workflow**

**SNP analysis** **GWAS** **Structural variations** **Chrom. reconstruction** **Metagenomics** **Gene families** **Mosaic Genomes**

These workflows as part of the services provided by **South Green**

**How to load big datasets?**

My file → 1 → HPC Cluster → 2 → My Galaxy History

**In order to figure out which tools were made available by our team, please activate the "tool search" functionality from the Options drop-down and type "south green" in the lookup filter.**

**Right Sidebar (History):**

- History
- search datasets
- VCFHunter 8 shown
- 100.33 MB
- 8: Origin.tab
- 7: chr09\_test.vcf
- 6: chr04\_test.vcf
- 5: chr03\_test.vcf
- 4: chr02\_test.vcf
- 3: chr01\_test.vcf
- 2: Carto.vcf
- 1: CartoRef.agp

# Chromosome painting with vcfHunter tool

Used data: A file containing ancestral (non admixed) accession origin

Ancestral origin

Accession name in the vcf

	P2	AA
T01	BB	
T02	BB	
T03	AA	
T04	AA	
T05	AA	
T06	AA	
T07	AA	
T08	BB	
T10	AA	
T11	AA	

History

search datasets

VCFHunter  
8 shown  
100.33 MB

1: CartoRef.agp

2: Carto.vcf

3: chr01\_test.vcf

4: chr02\_test.vcf

5: chr03\_test.vcf

6: chr04\_test.vcf

7: chr09\_test.vcf

8: Origin.tab

# Chromosome painting with vcfHunter tool

Used data: Several vcf files containing the genotypes of 15 accessions

The screenshot shows the Galaxy web interface with the following elements:

- Left sidebar (Tools):** Lists various bioinformatics tools including NGS, Population structure, GWAS, VCFtools, Tassel GBS, Rice Variant Analysis, GENOME HARVEST, TransPo-RG Transfer of Position to Resequenced Genome, parental SNP - Detect parental SNP of hybrids, Visualization, TraceAncestor, vcfHunter, KDE\_classifier, METAGENOMICS, FROGS, EVOLUTION/PHYLOGENY, Comparative Genomics, NCBI BLAST+, Genfam, Protein analyses, STATISTICS/GRAFICS, Statistics, Graph/Display Data, SOUTHGREEN PROJECTS, SNIPplay3, and GNPAnnot Tools.
- Middle panel:** Displays a table of VCF data. The columns are labeled: #CHROM, POS, ID, REF, ALT, and QUAL. The data shows variants for chromosome chr09 across multiple positions (e.g., 26844, 38559, 38565, 38574, 38610, 38651, 38707, 38713, 42909, 47454, 51113, 58026, 63859, 63864, 63876, 63889, 63901, 63924, 67990, 68003, 68024, 73617, 78839, 78853, 78866, 79031, 79081, 80569, 80631).
- Right panel (History):** Shows a list of datasets. A red box highlights the first five entries, which are VCF files: 8: Origin.tab, 7: chr09\_test.vcf, 6: chr04\_test.vcf, 5: chr03\_test.vcf, 4: chr02\_test.vcf, and 3: chr01\_test.vcf. An arrow labeled "2" points from the top of this list to the right. Another red box highlights the last two entries: 2: Carto.vcf and 1: CartoRef.app. An arrow labeled "1" points to the "Carto.vcf" entry. A large red callout bubble at the bottom right contains the text: "5 vcf files (1 for each chr) for 15 accessions (11 ancestral and 4 hybrids)".
- Bottom:** A black bar with a red arrow labeled "3" pointing to its center.

# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession

The screenshot shows the Galaxy bioinformatics platform interface. On the left, a sidebar lists various tools and projects, including 'vcfHunter' (circled with red number 2). In the center, the SouthGreen bioinformatics platform homepage is displayed, featuring the SouthGreen logo and pre-configured workflows. A large red arrow labeled '1' points from the sidebar towards the central workflow area. Another red arrow labeled '3' points from the sidebar towards the bottom right corner of the screen.

**Galaxy** Using 19%

Analyze Data Workflow Shared Data Visualization Admin Help User

**Galaxy**

Tools

- NGS: SNP Calling
- VarScan
- Population structure
- GWAS
- VCFTools
- Tassel GBS (Version 4.0)
- Rice Variant Analysis (Rice 3k, IRGIM, High Density Rice Array (HDRA, 700k SNPs))
- GENOME HARVEST
- TransPo-RG Transfer of Position to Resequenced Genome
- parental SNP - Detect parental SNP of hybrids
- Visualization
- TraceAncestor
- vcfHunter** ②
- VCF Filter
- vcf2allPropAndCov
- vcf2allPropAndCovByChr
- vcf2popNew
- RecombCalculatorDDose
- Draw\_dot\_plot
- KDE\_classifier
- METAGENOMICS
- FROGS
- EVOLUTION/PHYLOGENY
- Comparative Genomics
- NCBI BLAST+
- Genfam
- Protein analyses
- STATISTICS/GRAFICS
- Statistics
- Graph/Display Data
- SOUTHGREEN PROJECTS
- SNiPlay3
- GNPAnnot Tools

**cirad** **IRD** **Bioversity** **INRA** **SupAgro**

# SouthGreen bioinformatics platform

## Welcome to — GALAXY

### Our pre-configured and validated workflows

**SNP calling**

The SNP Calling is based on the **GATK toolkit**, using either UnifiedGenotyper or HaplotypeCaller module.

**Input:** BAM alignment files + FASTA for reference  
**Output:** VCF (Variant call Format) file

**Access workflow**

**SNP analysis** **GWAS** **Structural variations** **Chrom. reconstruction** **Metagenomics** **Gene families** **Mosaic Genomes**

These workflows as part of the services provided by [South Green](#)

**How to load big datasets?**

1. My file → HPC Cluster → My Galaxy History

Choose FTP file

In order to figure out which tools were made available by our team, please activate the "tool search" functionality from the Options drop-down and type "south green" in the lookup filter.

History

search datasets

VCFHunter 8 shown

- 100.33 MB
- 8: Origin.tab**
- 7: chr09\_test.vcf**
- 6: chr04\_test.vcf**
- 5: chr03\_test.vcf**
- 4: chr02\_test.vcf**
- 3: chr01\_test.vcf**
- 2: Carto.vcf**
- 1: CartoRef.agp**

# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 19%

History search datasets VCFHunter 8 shown 100.33 MB

vcf2allPropAndCov (Galaxy Version 0.1.0)

vcf collection No data dataset collection available.

origin --origin 8: Origin.tab A 2 column file containing accession name (col1), origin/group (Col2)

accession to work with

ploidy of the accession 2

ancestral accession can't have missing data no

all accessions should have the allele yes

Execute

Author Guillaume Martin  
Galaxy integration Aurore Comte  
Support For any questions about Galaxy integration, please send an e-mail to [aurore.comte@ird.fr](mailto:aurore.comte@ird.fr)

**vcf2allPropAndCov**

**Description**

This program perform two things based on a vcf.

1. It plots for an accession, the allele coverage alongs its chromosomes.
2. It identify, based on known ancestral accessions in the vcf, the alleles specific to each groups and plot the alleles proportion at a site in the accession along chromosomes.

**Inputs:**

VCF collection : one per chromosome or a single vcf for all chromosomes  
Origin : A 2 column file containing accession name (col1), origin/group (Col2)

**Origin.tab**

P2 AA  
T01 BB  
T02 BB

**Outputs:**

# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession

The screenshot shows the Galaxy web interface with the following details:

- Left Sidebar (Tools):** Includes sections like NGS: SNP Calling, VarScan, Population structure, GWAS, VCFtools, Tassel GBS (Version 4.0), Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs)), GENOME HARVEST, TransPo-RG Transfer of Position to Resequenced Genome, Parental SNP - Detect parental SNP of hybrids, Visualization, TraceAncestor, vcfHunter, Evolution/Phylogeny, Comparative Genomics, NCBI BLAST+, Genfam, Protein analyses, Statistics, Graph/Display Data, and Southgreen Projects.
- Main Content Area:** Displays the **vcf2allPropAndCov** tool configuration. The "vcf collection" dropdown is set to "No data dataset collection available." The "origin --origin" dropdown is set to "8: Origin.tab". The "accession to work with" field is empty. The "ploidy of the accession" dropdown is set to "2". The "ancestral accession can't have missing data" dropdown is set to "no". The "all accessions should have the allele" dropdown is set to "yes". A "Execute" button is present. Below the form, there are three authorship notes: Author Guillaume Martin, Galaxy integration Aurore Comte, and Support for any questions about Galaxy integration, please send an e-mail to [aurore.comte@ird.fr](mailto:aurore.comte@ird.fr).
- Description:** Describes the program's purpose: "This program perform two things based on a vcf.  
1. It plots for an accession, the allele coverage alongs its chromosomes.  
2. It identify, based on known ancestral accessions in the vcf, the alleles specific to each groups and plot the alleles proportion at a site in the accession along chromosomes."
- Inputs:** Lists "VCF collection : one per chromosome or a single vcf for all chromosomes" and "Origin : A 2 column file containing accession name (col1), origin/group (Col2)".
- Origin.tab:** Contains the following content:

```
P2 AA
T01 BB
T02 BB
```
- Outputs:** (Listed below the inputs)
- Right Panel (History):** Shows a list of datasets:
  - 8: Origin.tab
  - 7: chr09\_te
  - 6: chr04\_te
  - 5: chr03\_te
  - 4: chr02\_te
  - 3: chr01\_te
  - 2: Carto.vcf
  - 1: CartoRef.appA context menu is open over the first five items, with options: Hide datasets, Unhide datasets, Delete datasets, Undelete datasets, Permanently delete datasets, Build Dataset List (circled with red number 3), Build Dataset Pair, and Build List of Dataset Pairs. A red bracket labeled 1 points to the menu icon. Red numbers 2 and 3 are circled on the menu options.

# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession

The screenshot shows the Galaxy web interface with a 'Create a collection from a list of datasets' dialog box open. The dialog lists several VCF files: chr09\_test.vcf, chr04\_test.vcf, chr03\_test.vcf, chr02\_test.vcf, and chr01\_test.vcf. There are 'Discard' buttons next to each file name. Below the list is a 'Name:' input field containing 'VCF.conf'. At the bottom right of the dialog is a 'Create list' button. Red circles labeled '1' and '2' are overlaid on the 'Name:' field and the 'Create list' button respectively. The background shows the Galaxy tool panel on the left and a list of datasets on the right.

# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession

The screenshot shows the Galaxy web interface with the following details:

- Left Sidebar (Tools):** Includes sections like NGS: SNP Calling, VarScan, Population structure, GWAS, VCFtools, Tassel GBS (Version 4.0), Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs)), GENOME HARVEST, TransPo-RG Transfer of Position to Resequenced Genome, parental SNP - Detect parental SNP of hybrids, Visualization, TraceAncestor, vcfHunter, VCF Filter, vcf2allPropAndCov (highlighted with a red circle and arrow labeled 2), vcf2allPropAndCovByChr, vcf2popNew, RecombCalculatorDDose, Draw\_dot\_plot, KDE\_classifier, METAGENOMICS, FROGS, EVOLUTION/PHYLOGENY, Comparative Genomics, NCBI BLAST+, Genfam, Protein analyses, STATISTICS/GRAFICS, Statistics, Graph/Display Data, SOUTHGREEN PROJECTS, SNiPlay3, GNPAnnot Tools.
- Middle Panel (vcf2allPropAndCov Tool Configuration):**
  - vcf collection:** No data dataset collection available.
  - origin --origin:** 8: Origin.tab (A 2 column file containing accession name (col1), origin/group (Col2))
  - accession to work with:** (empty input field)
  - ploidy of the accession:** 2
  - ancestral accession can't have missing data:** no
  - all accessions should have the allele:** yes
  - Execute:** (checkbox checked)
- Bottom Panel (Tool Description):**
  - vcf2allPropAndCov**
  - Description:** This program perform two things based on a vcf.
    1. It plots for an accession, the allele coverage alongs its chromosomes.
    2. It identify, based on known ancestral accessions in the vcf, the alleles specific to each groups and plot the alleles proportion at a site in the accession along chromosomes.
  - Inputs:** VCF collection : one per chromosome or a single vcf for all chromosomes  
Origin : A 2 column file containing accession name (col1), origin/group (Col2)
  - Origin.tab:**

P2 AA  
T01 BB  
T02 BB
  - Outputs:** (List of outputs shown in the History panel)
- Right Panel (History):** Shows a list of datasets and tools used:
  - VCFHunter**: 9 shown, 100.33 MB
    - 7: chr09\_test.vcf (checked)
    - 6: chr04\_test.vcf (checked)
    - 5: chr03\_test.vcf (checked)
    - 4: chr02\_test.vcf (checked)
    - 3: chr01\_test.vcf (checked)
    - 2: Carto.vcf (unchecked)
    - 1: CartoRef.agp (unchecked)
  - Others:** VCF.conf, Origin.tab, VCF Hunter configuration files.

# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession

The screenshot shows the Galaxy web interface with the vcf2allPropAndCov tool selected. The tool configuration is as follows:

- 1** VCF.conf: Set to 9: VCF.conf
- 2** Origin.tab: Set to 8: Origin.tab
- 3** Kunnan: Set to Kunnan
- 4** 2 (diploid): Set to 2
- 5** no: Set to no
- 6** yes: Set to yes
- 7** Execute: The execute button.

**vfc2allPropAndCov**

**Description**

This program perform two things based on a vcf.

1. It plots for an accession, the allele coverage alongs its chromosomes.
2. It identify, based on known ancestral accessions in the vcf, the alleles specific to each groups and plot the alleles proportion at a site in the accession along chromosomes.

**Inputs:**

VCF collection : one per chromosome or a single vcf for all chromosomes  
Origin : A 2 column file containing accession name (col1), origin/group (Col2)

**Origin.tab**

P2 AA  
T01 BB  
T02 BB

**Outputs:**

The history panel on the right shows the following entries:

- VCFHunter 9 shown 100.33 MB
- 9: VCF.conf
- 8: Origin.tab
- 7: chr09\_test.vcf
- 6: chr04\_test.vcf
- 5: chr03\_test.vcf
- 4: chr02\_test.vcf
- 3: chr01\_test.vcf
- 2: Carto.vcf
- 1: CartoRef.agp

# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession

Galaxy x + [galaxy.southgreen.fr/galaxy/](http://galaxy.southgreen.fr/galaxy/) Galaxy southgreen Using 19%

**Tools**

- Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs))
- GENOME HARVEST
- TransPo-RG Transfer of Position to Resequenced Genome
- parental SNP - Detect parental SNP of hybrids
- Visualization
- TraceAncestor
- vcfHunter
  - VCF Filter
  - vcf2allPropAndCov
  - vcf2allPropAndCovByChr
  - vcf2popNew
  - RecombCalculatorDDose
  - Draw\_dot\_plot
  - KDE\_classifier
- METAGENOMICS
- FROGS
- EVOLUTION/PHYLOGENY
- Comparative Genomics
- NCBI BLAST+
- Genfam
- Protein analyses
- STATISTICS/GRAFICS
- Statistics
- Graph/Display Data
- SOUTHGREEN PROJECTS
- SNIPplay3
- GNPAnnot Tools
- GNPAnnot Converters
- ESTtik
- Expression data
- SAT
- CMAEE tools

**vcf2allPropAndCov (Galaxy Version 0.1.0)**

**vcf collection**  
9: VCF.conf

**origin --origin**  
17: Kunnan\_stats.tab  
A 2 column file containing accession name (col1), origin/group (Col2)

**accession to work with**

**ploidy of the accession**  
2

**ancestral accession can't have missing data**  
no

**all accessions should have the allele**  
yes

Execute

Author Guillaume Martin  
Galaxy integration Aurore Comte  
Support For any questions about Galaxy integration, please send an e-mail to [aurore.comte@ird.fr](mailto:aurore.comte@ird.fr)

**vcf2allPropAndCov**

**Description**  
This program perform two things based on a vcf.  
1. It plots for an accession, the allele coverage alongs its chromosomes.  
2. It identify, based on known ancestral accessions in the vcf, the alleles specific to each groups and plot the alleles proportion at a site in the accession along chromosomes.

**Inputs:**  
VCF collection : one per chromosome or a single vcf for all chromosomes  
Origin : A 2 column file containing accession name (col1), origin/group (Col2)

**Origin.tab**  
P2 AA  
T01 BB  
T02 BB

**Outputs:**

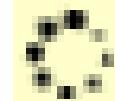
History search datasets 13 shown, 4 deleted 101.51 MB

12: Kunnan\_stats.tab  
16: Kunnan\_AlleleOriginAndRatio.tab  
15: Kunnan\_Ratio.png  
14: Kunnan\_Cov.png  
9: VCF.conf  
8: Origin.tab  
7: chr09\_test.vcf  
6: chr04\_test.vcf  
5: chr03\_test.vcf  
4: chr02\_test.vcf  
3: chr01\_test.vcf  
2: Carto.vcf  
1: CartoRef.agp

**4 files will be generated**



Send to cluster



Running on cluster



21: Finished

# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession

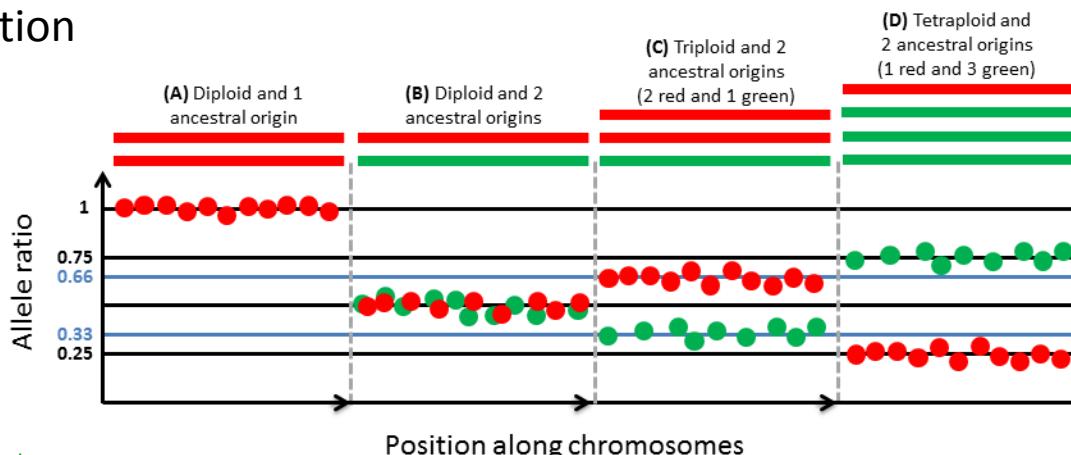
The screenshot shows the Galaxy web interface with the following details:

- Tool Selection:** vcf2allPropAndCov (Galaxy Version 0.1.0)
- Inputs:**
  - vcf collection: 9: VCF.conf
  - origin --origin: 17: Kunnan\_stats.tab
  - accession to work with: (empty)
  - ploidy of the accession: 2
  - ancestral accession can't have missing data: no
  - all accessions should have the allele: yes
- Description:** vcf2allPropAndCov
- Description:** This program performs two things based on a vcf.
  - It plots for an accession, the allele coverage along its chromosomes.
  - It identifies, based on known ancestral accessions in the vcf, the alleles specific to each group and plots the allele proportion at a site in the accession along chromosomes.
- Inputs:** VCF collection : one per chromosome or a single vcf for all chromosomes  
Origin : A 2 column file containing accession name (col1), origin/group (Col2)
- Outputs:** (Listed in the History panel)
- History Panel:** Shows the following items:
  - VCFHunter (13 shown, 4 deleted)
    - 101.51 MB
    - 17: Kunnan\_stats.tab
    - 16: Kunnan\_AlleleOriginAndRatio.tab
    - 15: Kunnan\_Ratio.png
    - 14: Kunnan\_Cov.png
  - 9: VCF.conf
  - 8: Origin.tab
  - 7: chr09\_test.vcf
  - 6: chr04\_test.vcf
  - 5: chr03\_test.vcf
  - 4: chr02\_test.vcf
  - 3: chr01\_test.vcf
  - 2: Carto.vcf
  - 1: CartoRef.agp

# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession : How the program work?

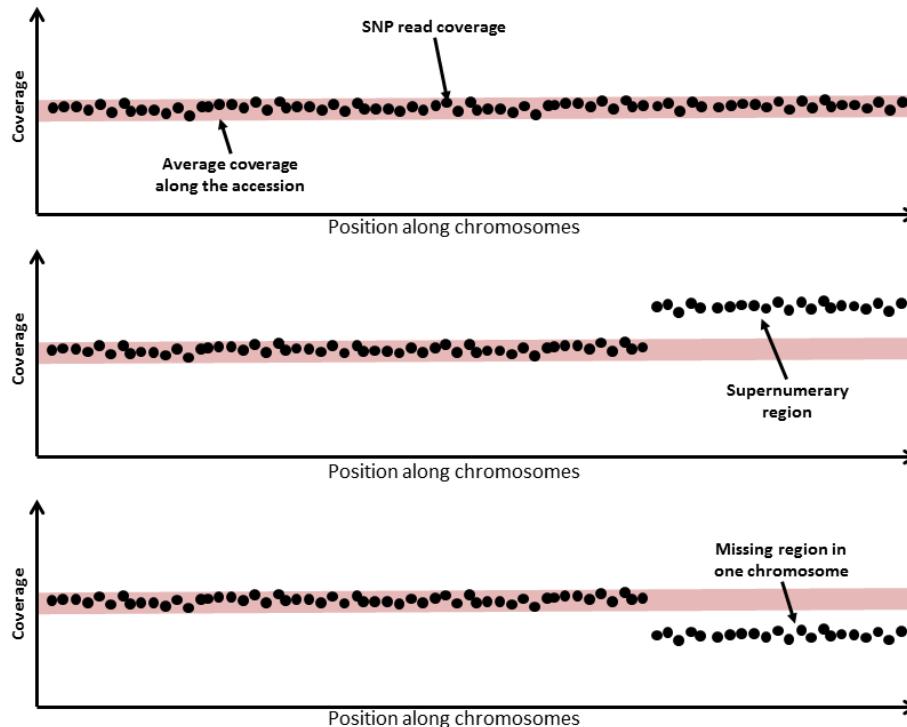
- Based on the origin.tab file, attribute an allele to an ancestral group according to the following rule:
  - ✓ If the allele is found **only** in member of an ancestral group (*i.e.* absent from all member of the other(s) group(s)), then the allele is attributed to this group.
- Possible variants to the rule:
  - ✓ Allele should be present in all member of the ancestral group (“yes” or “no”) 
  - ✓ Missing data are accepted in member of ancestral group (“yes” or “no”) 
- Calculate, in the studied accession, the proportion of read having this allele
- Plot this proportion



# Chromosome painting with vcfHunter tool

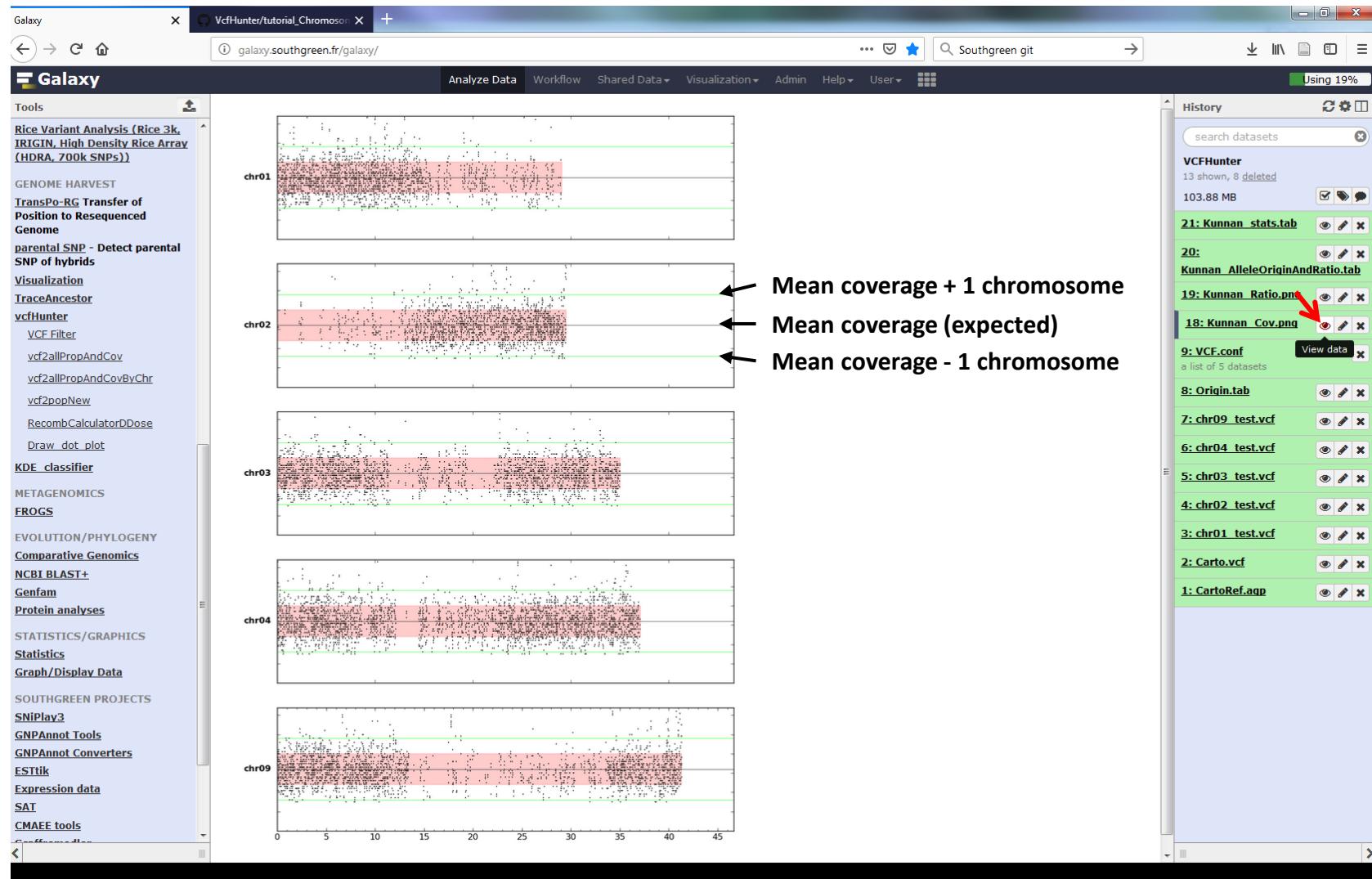
## 1- Chromosome painting of one accession : How the program work?

- Plot normalized variant site coverage



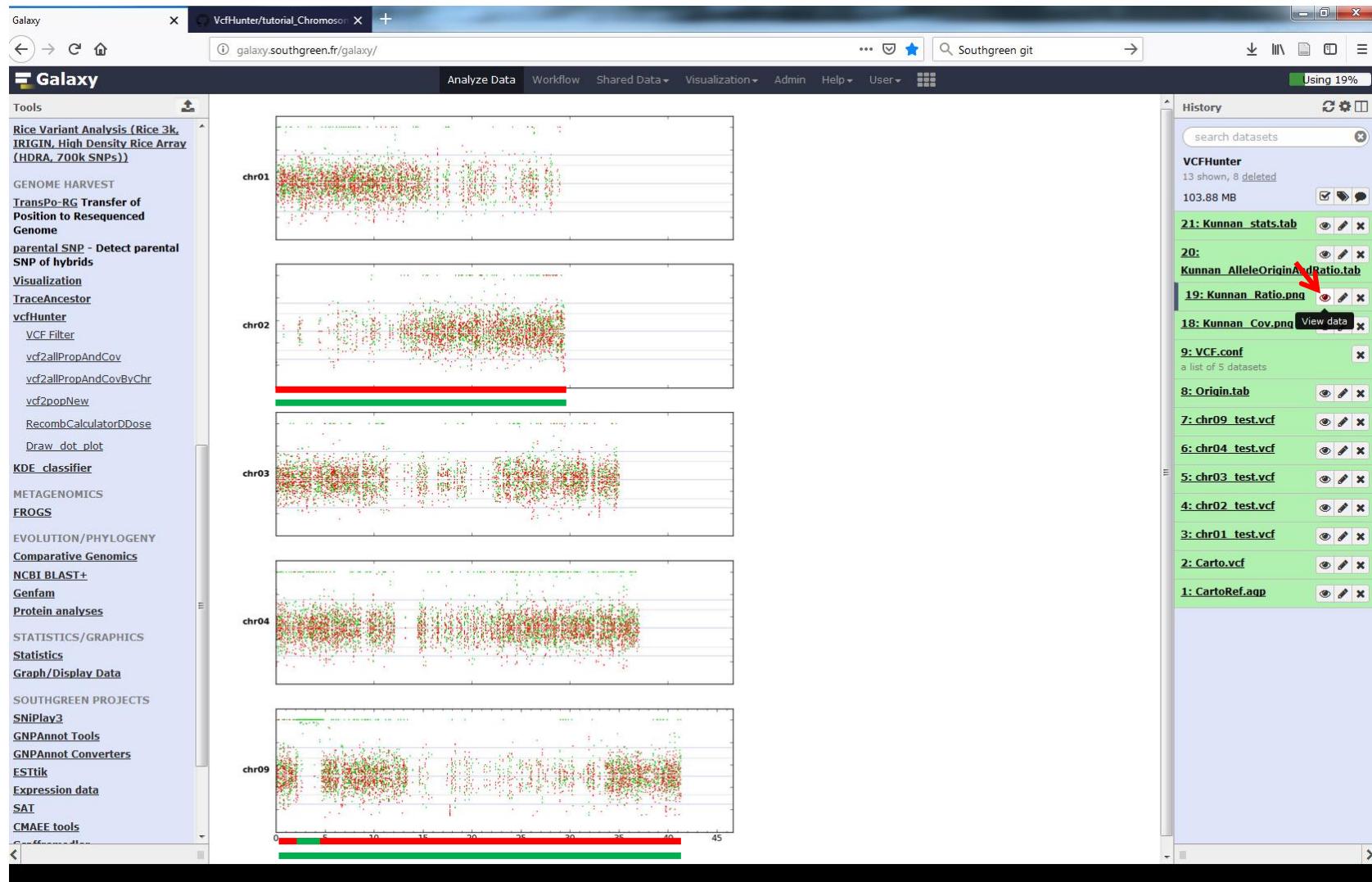
# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession : Outputs description



# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession : Outputs description



# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession : Outputs description

The file used to draw the allele ratio figure

Allele ratio in the accession

Allele ancestral group

Attributed allele

Position

Chromosome

The screenshot shows the Galaxy web interface with a VCF file open in a central panel and a history list on the right. The VCF file contains genetic data for chromosome 02, with columns for Position, Allele, Ref, Alt, and Depth. The history list shows various files generated by the vcfHunter tool, including 'Kunnan.stats.tab' (highlighted with a red arrow), 'Kunnan\_AlleleOriginAndRatio.tab', 'Kunnan\_Ratio.png', 'Kunnan\_Cov.png', 'VCF.conf', 'Origin.tab', and several test.vcf files for chromosomes 01 through 09.

Position	Allele	Ref	Alt	Depth
chr02 86818	G	BB	0.22580645161290322	
chr02 771450	G	BB	0.5263157894736842	
chr02 771451	A	AA	0.47368421052631576	
chr02 771458	A	BB	0.55	
chr02 771458	G	AA	0.45	
chr02 2273733	G	BB	0.5	
chr02 2273733	T	AA	0.5	
chr02 2273745	T	BB	0.5151515151615151	
chr02 2273745	A	AA	0.4848484848484486	
chr02 2287602	T	BB	0.44	
chr02 2287602	C	AA	0.56	
chr02 2343526	C	BB	0.5555555555555556	
chr02 2343526	T	AA	0.4444444444444444	
chr02 2427206	C	BB	0.535142857142857	
chr02 2427206	G	AA	0.4642857142857143	
chr02 2423280	A	BB	0.595238092380952	
chr02 2423280	G	AA	0.40476190476190477	
chr02 2423283	G	BB	0.5714285714285714	
chr02 2423283	A	AA	0.42857142857142853	
chr02 2423293	T	BB	0.5909090909090909	
chr02 2423293	C	AA	0.4090909090909091	
chr02 24232400	C	BB	0.5348837209302325	
chr02 24232400	A	AA	0.46511627906976744	
chr02 2447292	T	BB	0.4375	
chr02 2447292	A	AA	0.5625	
chr02 2447319	T	BB	0.42424242424242425	
chr02 2447319	A	AA	0.5757575757575758	
chr02 2447328	G	BB	0.375	
chr02 2447328	A	AA	0.625	
chr02 2480263	G	BB	0.48	
chr02 2480263	C	AA	0.52	
chr02 2480266	A	BB	0.4782608695652174	
chr02 2487990	G	BB	0.4722222222222222	
chr02 2487990	A	AA	0.5277777777777778	
chr02 2575326	T	BB	0.5714285714285714	
chr02 2575326	C	AA	0.42857142857142855	
chr02 2696749	T	BB	0.6	
chr02 2696749	C	AA	0.4	
chr02 4387494	C	BB	0.5	
chr02 4387494	T	AA	0.5	
chr02 4405457	C	BB	0.5675675675675675	
chr02 4405457	G	AA	0.43243243243243246	
chr02 4454139	T	BB	0.41379310344827586	

# Chromosome painting with vcfHunter tool

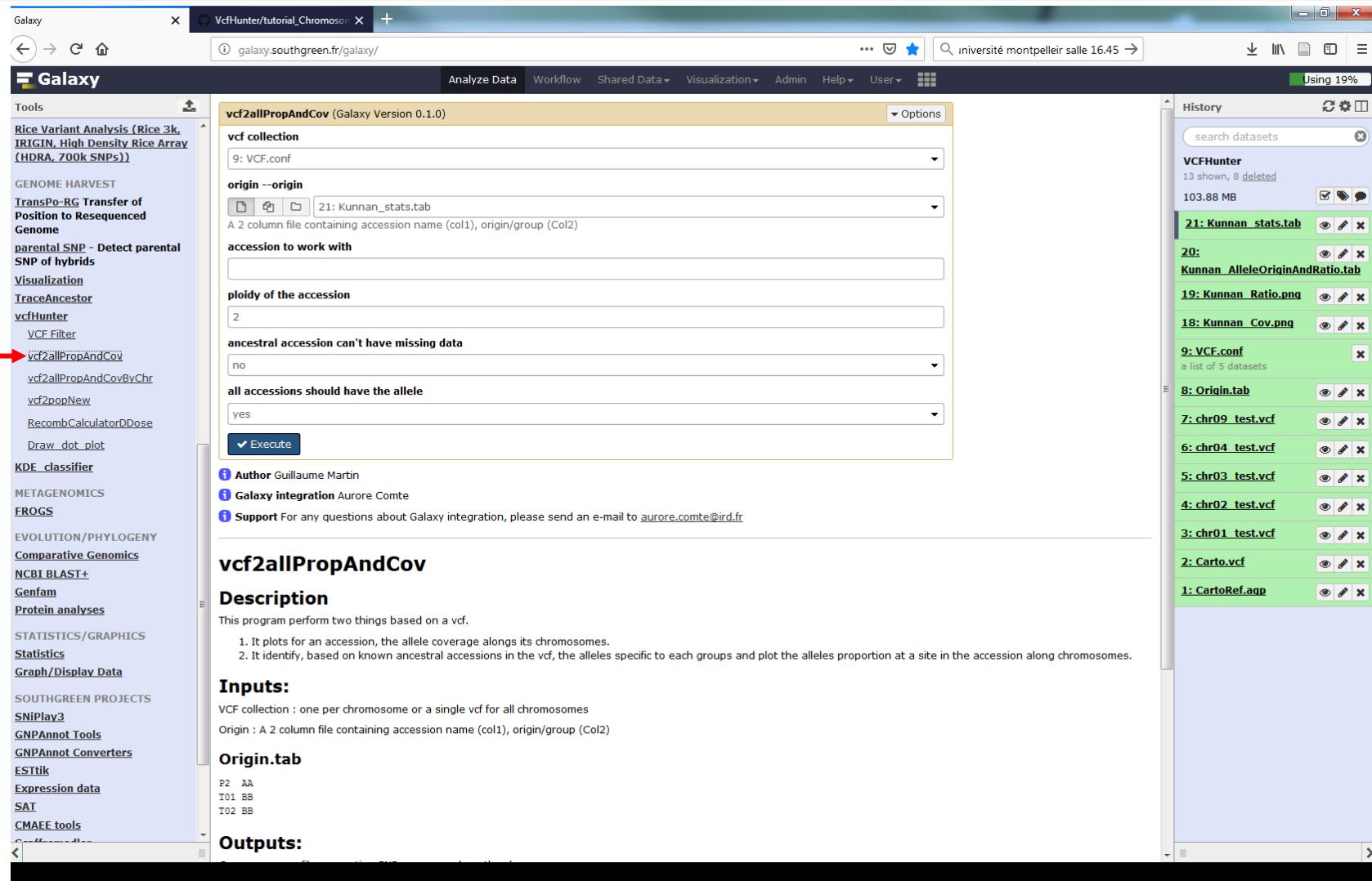
## 1- Chromosome painting of one accession : Outputs description

The screenshot shows the Galaxy web interface with the following details:

- Tools Panel:** Shows various bioinformatics tools categorized under different sections like Rice Variant Analysis, GENOME HARVEST, TransPoRG Transfer of Position to Resequenced Genome, parental SNP - Detect parental SNP of hybrids, Visualization, TraceAncestor, vcfHunter, KDE\_classifier, METAGENOMICS, FROGS, EVOLUTION/PHYLOGENY, Comparative Genomics, NCBI BLAST+, Genfam, Protein analyses, STATISTICS/GRAFICS, Statistics, Graph/Display Data, SOUTHGREEN PROJECTS, SNIPPlay3, GNPAnnot Tools, GNPAnnot Converters, ESTtik, Expression data, SAT, CMAEE tools.
- Workflow Tab:** Analyze Data, Workflow, Shared Data, Visualization, Admin, Help, User.
- History Tab:** search datasets, VCFHunter (13 shown, 8 deleted), 103.88 MB, 21: Kunnan\_stats.tab (highlighted with a red arrow), 20: Kunnan\_AlleleOriginAndRatio.tab, 19: Kunnan\_Ratio.png, 18: Kunnan\_Cov.png, 9: VCF.conf, 8: Origin.tab, 7: chr09\_test.vcf, 6: chr04\_test.vcf, 5: chr03\_test.vcf, 4: chr02\_test.vcf, 3: chr01\_test.vcf, 2: Carto.vcf, 1: CartoRef.agp.
- Content Area:** Allele global statistics

# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession : To go further



The screenshot shows the Galaxy web interface with the following details:

- Left Sidebar (Tools):** Includes sections like Tools, Rice Variant Analysis, GENOME HARVEST, vcfHunter, and many others.
- Main Content Area:** The **vcf2allPropAndCov** tool is selected. It has several input fields:
  - vcf collection:** Set to "9: VCF.conf".
  - origin --origin:** Set to "21: Kunnan\_stats.tab". A note says: "A 2 column file containing accession name (col1), origin/group (Col2)".
  - accession to work with:** An empty text input field.
  - ploidy of the accession:** Set to "2".
  - ancestral accession can't have missing data:** Set to "no".
  - all accessions should have the allele:** Set to "yes".
- Bottom Description:** Provides a brief description of the tool's function and its inputs/outputs.
- Right Sidebar (History):** Shows a list of datasets and analyses, including "VCFHunter", "21: Kunnan\_stats.tab", and various "Kunnan" files.

Running the analysis with other options, other accessions of the vcf.

Other hybrids names in the vcf GP1 (triploid), P1 (tetraploid) and P025 (triploid)

# Chromosome painting with vcfHunter tool

## 1- Chromosome painting of one accession : Additional information

The screenshot shows the Galaxy web interface with the following details:

- Tool Details:** The main panel displays the "vcf2allPropAndCov" tool from the "VcfHunter" category. It includes sections for "Description", "Inputs", "Outputs", and "For more informations". A large button labeled "Small description" is overlaid on the right side of the tool details.
- Parameter Configuration:** Above the tool details, there are two dropdown menus:
  - "ancestral accession can't have missing data": set to "no"
  - "all accessions should have the allele": set to "yes"A "Execute" button is located below these dropdowns.
- Tool History:** On the right, the "History" panel shows a list of 13 datasets, all named "Kunnan" with various file extensions (tab, png, vcf). A red circle labeled "1" points to the "vcf2allPropAndCov" entry in the history, and a red circle labeled "2" points to the "Kunnan" entries.
- Left Sidebar:** The sidebar lists various Galaxy tools and categories, including "Rice Variant Analysis", "TransPo-RG Transfer of Position to Resequenced Genome", "parental SNP - Detect parental SNP of hybrids", "Visualization", "TraceAncestor", "vcfHunter", "VCF Filter", "vcf2allPropAndCov", "vcf2allPropAndCovByChr", "vcf2popNew", "RecombCalculatorDDose", "Draw\_dot\_plot", "KDE\_classifier", "METAGENOMICS", "FROGS", "EVOLUTION/PHYLOGENY", "Comparative Genomics", "NCBI BLAST+", "Genfam", "Protein analyses", "STATISTICS/GRAFICS", "Statistics", "Graph/Display Data", "SOUTHGREEN PROJECTS", "SNIPplay3", "GNPAnnot Tools", "GNPAnnot Converters", "ESTtik", "Expression data", "SAT", "CMAEE tools", and "CartoRef".

# Chromosome painting with vcfHunter tool

## 2- Comparison of several accessions

The screenshot shows the Galaxy web interface with the following details:

- Left Sidebar (Tools):**
  - Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs))
  - GENOME HARVEST
  - TransPo-RG Transfer of Position to Resequenced Genome
  - parental SNP - Detect parental SNP of hybrids
  - Visualization
  - TraceAncestor
  - vcfHunter
    - VCF Filter
    - vcf2allPropAndCov
    - vcf2allPropAndCovByChr
    - vcf2popNew
    - RecombCalculatorDDose
    - Draw\_dot\_plot
  - KDE\_classifier
  - METAGENOMICS
  - FROGS
  - EVOLUTION/PHYLOGENY
  - Comparative Genomics
  - NCBI BLAST+
  - Genfam
  - Protein analyses
  - STATISTICS/GRAFICS
  - Statistics
  - Graph/Display Data
  - SOUTHGREEN PROJECTS
  - SNPlay3
  - GNPAnnot Tools
  - GNPAnnot Converters
  - ESTtik
  - Expression data
  - SAT
  - CMAEE tools
- Main Panel (vcf2allPropAndCovByChr Configuration):**
  - 1** VCF.conf (dropdown menu set to 9: VCF.conf)
  - 2** Origin.tab (dropdown menu set to 8: Origin.tab)
  - 3** T04,T02,Kunnan,GP1,P025,P1 (text input field)
  - 4** 3 (triploid -not very important-) (text input field)
  - 5** no (dropdown menu set to no)
  - 6** yes (dropdown menu set to yes)
  - 7** Execute button
- Right Panel (History):**
  - 9: VCF.conf
  - 8: Origin.tab
  - 7: chr09\_test.vcf
  - 6: chr04\_test.vcf
  - 5: chr03\_test.vcf
  - 4: chr02\_test.vcf
  - 3: chr01\_test.vcf
  - 2: Carto.vcf
  - 1: CartoRef.agp

# Chromosome painting with vcfHunter tool

## 2- Comparison of several accessions: output description

The screenshot shows the Galaxy web interface with the following details:

- Tools Panel:** Shows various bioinformatics tools categorized under different sections like GENOME HARVEST, METAGENOMICS, and SOUTHGREEN PROJECTS.
- Job Queue Status:** A green message box indicates "1 job has been successfully added to the queue - resulting in the following datasets".
- History Panel:** Displays a list of datasets generated by the VcfHunter tool, including:
  - 1: 1.22 MB
  - 102: output (highlighted with a red arrow)
  - 21: Kunnan\_stats.tab
  - 20: Kunnan\_AlleleOriginAndRatio.tab
  - 19: Kunnan\_Ratio.png
  - 18: Kunnan\_Cov.png
  - 9: VCF.conf
  - 8: Origin.tab
  - 7: chr09\_test.vcf
  - 6: chr04\_test.vcf
  - 5: chr03\_test.vcf
  - 4: chr02\_test.vcf
  - 3: chr01\_test.vcf
  - 2: Carto.vcf
  - 1: CartoRef.agp
- Red Callout Box:** A red rounded rectangle contains the text "A collection of several files", pointing towards the History panel.

# Chromosome painting with vcfHunter tool

## 2- Comparison of several accessions: output description

Galaxy    VcfHunter/tutorial\_Chromoson    Stockholm format - Wikipedia

galaxy.southgreen.fr/galaxy/ Analyze Data Workflow Shared Data Visualization Admin Help User Using 19%

Tools

Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs))

GENOME HARVEST

TransPo-RG Transfer of Position to Resequenced Genome

parental SNP - Detect parental SNP of hybrids

Visualization

TraceAncestor

vcfHunter

VCF Filter

vcf2allPropAndCov

vcf2allPropAndCovByChr

vcf2popNew

RecombCalculatorDose

Draw\_dot\_plot

KDE\_classifier

METAGENOMICS

FROGS

EVOLUTION/PHYLOGENY

Comparative Genomics

NCBI BLAST+

Genfam

Protein analyses

STATISTICS/GRAFICS

Statistics

Graph/Display Data

SOUTHGREEN PROJECTS

SNiPlay3

GNPAnnot Tools

GNPAnnot Converters

ESTtik

Expression data

SAT

CMAFE tools

1 job has been successfully added to the queue - resulting in the following datasets:  
You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

output  
a list of datasets

RatioAndCov\_chr01\_1\_Cov.png

RatioAndCov\_chr01\_1\_Ratio.png

RatioAndCov\_chr02\_1\_Cov.png

RatioAndCov\_chr02\_1\_Ratio.png

RatioAndCov\_chr03\_1\_Cov.png

RatioAndCov\_chr03\_1\_Ratio.png

RatioAndCov\_chr04\_1\_Cov.png

RatioAndCov\_chr04\_1\_Ratio.png

RatioAndCov\_chr09\_1\_Cov.png

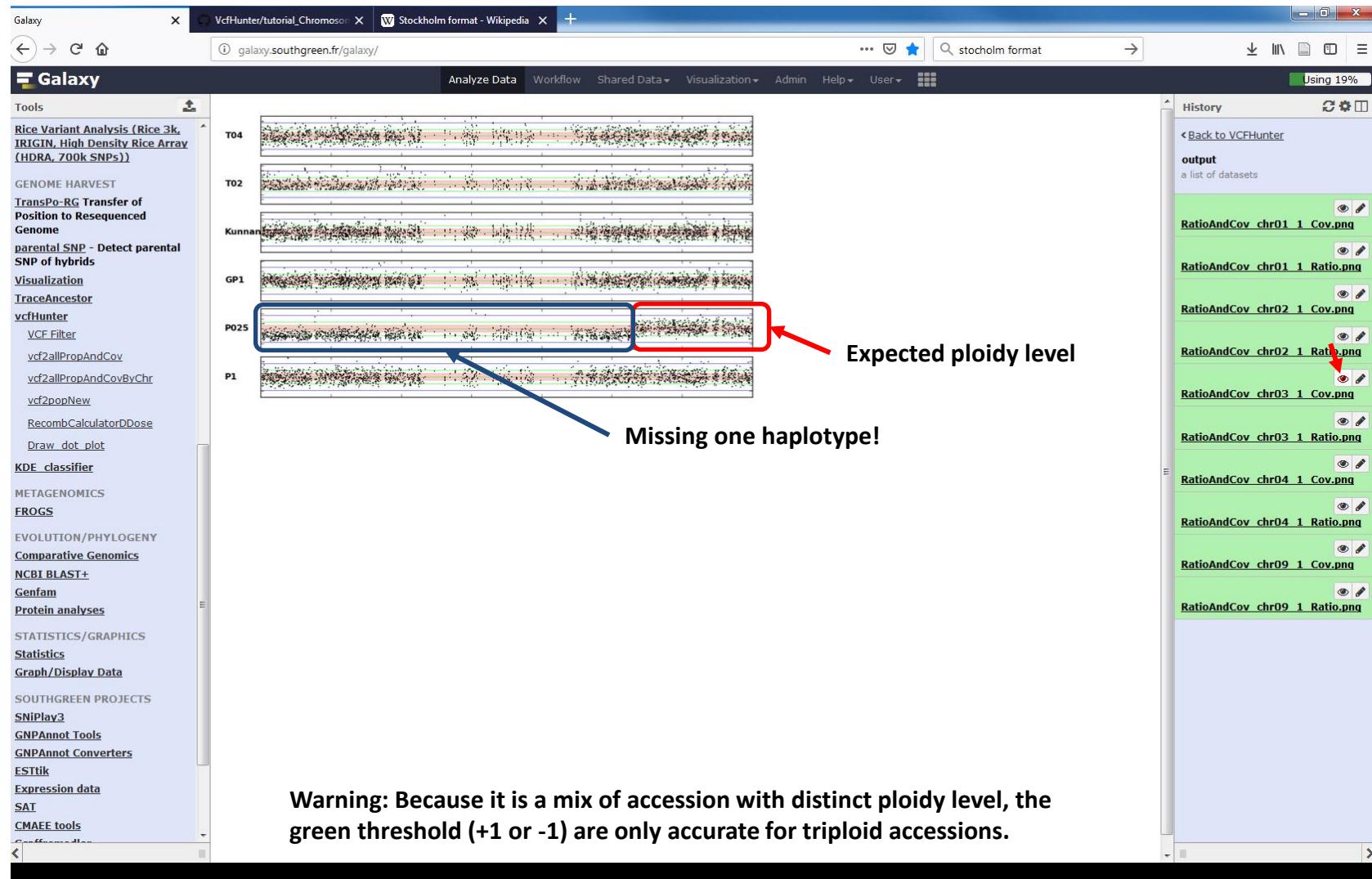
RatioAndCov\_chr09\_1\_Ratio.png

A collection of several files  
For each chromosomes:  
• A coverage file  
• A ratio file

The number of files per chromosome depends on the of the number of accessions treated (there can be at most 15 accessions per picture)

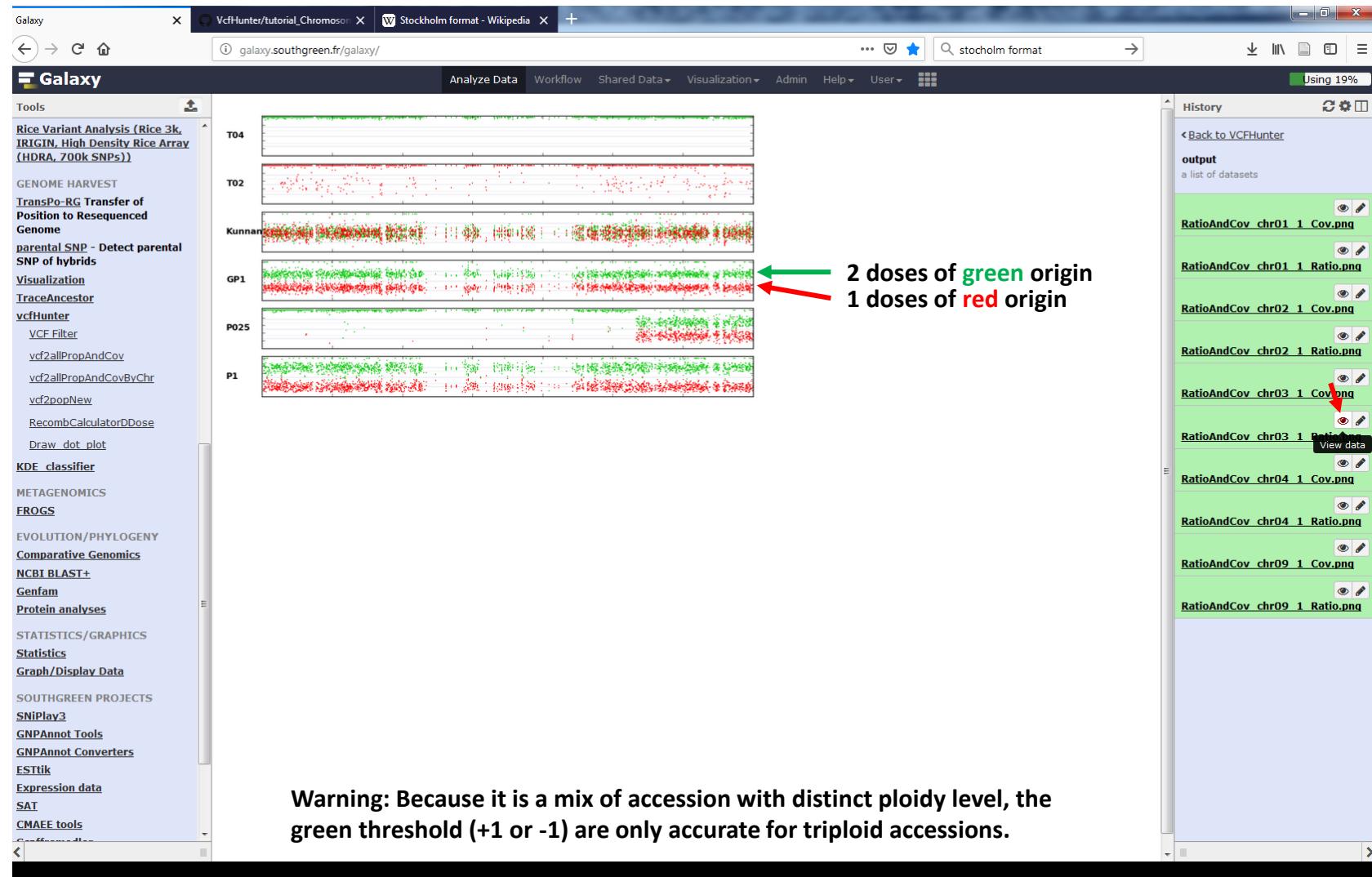
# Chromosome painting with vcfHunter tool

## 2- Comparison of several accessions: output description



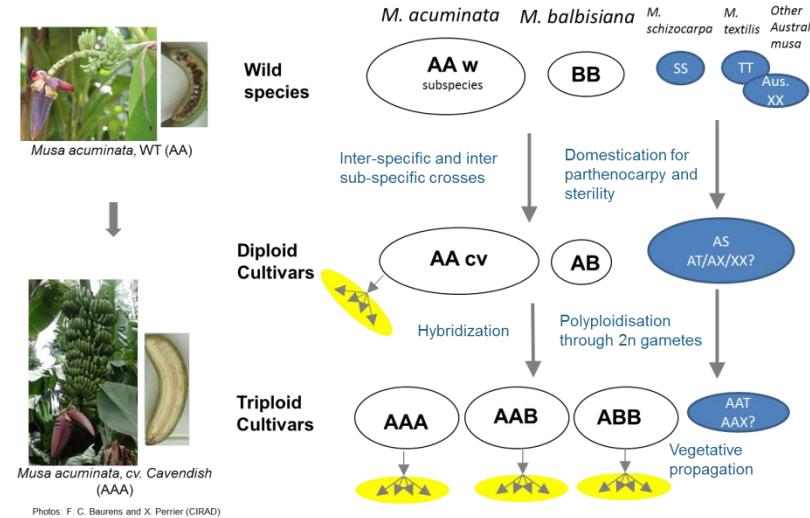
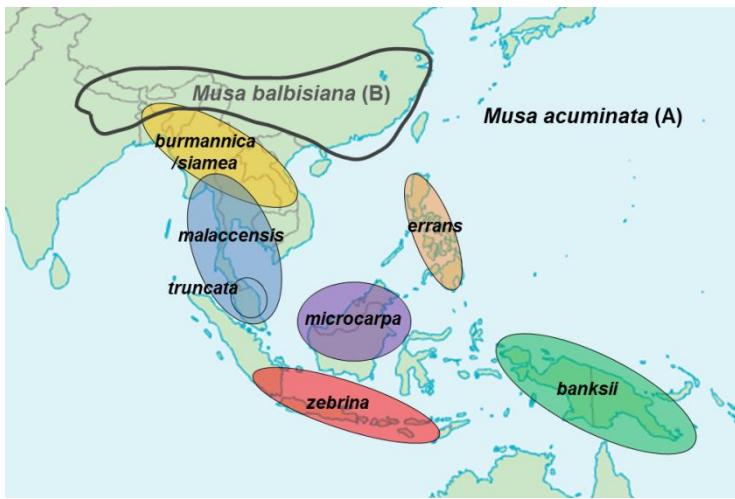
# Chromosome painting with vcfHunter tool

## 2- Comparison of several accessions: output description



# Genetic mapping analysis with vcfHunter tool

Developed to answer banana problematics



- Cultivated banana are hybrids between distinct species and subspecies showing chromosomal structural rearrangements

**What are the chromosomal structures of species and subspecies implicated in cultivated bananas?**

**What are the consequences of these structural variations on chromosomal pairing, recombination and segregation?**

# Genetic mapping analysis with vcfHunter tool

What are the chromosomal structures of plants?

What are the consequences of these structural variations on chromosomal pairing, recombination and segregation

Several tools developed under vcfHunter toolbox for this purpose:

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 19%

**Galaxy**

Tools

- NGS: SNP Calling
- Varscan
- Population structure
- GWAS
- VCFTools
- Tassel GBS (Version 4.0)
- Rice Variant Analysis (Rice 3k, IRIGEN, High Density Rice Array (HDRA, 700k SNPs))
- GENOME HARVEST
- TransPoRG Transfer of Position to Resequenced Genome
- parental SNP - Detect parental SNP of hybrids
- Visualization
- TraceAncestor
- vcfHunter** ②
- VCF Filter
- vcf2allPropAndCov
- vcf2allPropAndCovByChr
- vcf2popNew ①
- RecombCalculatorDDose
- Draw dot plot
- KDE classifier
- METAGENOMICS
- FROGS
- EVOLUTION/PHYLOGENY
- Comparative Genomics
- NCBI BLAST+
- Genfam
- Protein analyses
- STATISTICS/GRAFICS
- Statistics
- Graph/Display Data
- SOUTHGREEN PROJECTS
- SNPplay3
- GNPAnnot Tools

cirad IRD INRA Montpellier SupAgro

**SouthGreen** bioinformatics platform

Welcome to GALAXY

Our pre-configured and validated workflows

**SNP calling**

The SNP Calling is based on the GATK toolkit, using either UnifiedGenotyper or HaplotypeCaller module.

Input: BAM alignment files + FASTA for reference  
Output: VCF (Variant call Format) file

Access workflow

SNP analysis GWAS Structural variations Chrom. reconstruction Metagenomics Gene families Mosaic Genomes

These workflows as part of the services provided by South Green

How to load big datasets?

1 My file → HPC Cluster → 2 My Galaxy History

Choose FTP file

In order to figure out which tools were made available by our team, please activate the "tool search" functionality from the Options drop-down and type "south green" in the lookup filter.

History

search datasets

VCFHunter 8 shown 100.33 MB

- 8: Origin.tab
- 7: chr09\_test.vcf
- 6: chr04\_test.vcf
- 5: chr03\_test.vcf
- 4: chr02\_test.vcf
- 3: chr01\_test.vcf
- 2: Carto.vcf
- 1: CartoRef.agp

# Genetic mapping analysis with vcfHunter tool

Used data: A vcf file containing genotypes for each individuals of a mapping population

The screenshot shows the Galaxy web interface with a VCF file loaded for analysis. The main pane displays the first megabyte of a large VCF file, which is over 11 GB in size. The file contains genomic data for chromosomes 1 through 22, with columns for CHROM, POS, ID, REF, ALT, QUAL, FILTER, INFO, FORMAT, and various sample IDs (P191, P134, P135, P136, P137, P039, P038, P132, P133, P035, P034, P03'). The INFO column includes depth information (e.g., P0/0:58, P0/1:53). The FILTER column shows mostly PASS. The FORMAT column indicates that the data is in VCFv4.2 format with AD, DP, and GC fields. The INFO column also includes GATKCommandLine VariantFiltration details like Date=2018-01-08 14:37:19.913025 and additional filters like CoverageFiltration and MissingDataFiltration.

On the right side, the History panel shows a workflow named "VCFHunter" that has run 14 times. It lists several output datasets, including "102: output" (a list of 10 datasets), "21: Kunnan\_stats.tab", "20: Kunnan\_AlleleOriginAndRatio.tab", "19: Kunnan\_Ratio.png", "18: Kunnan\_Cov.png", "9: VCF.conf", "8: Origin.tab", "7: chr09\_test.vcf", "6: chr04\_test.vcf", "5: chr03\_test.vcf", "4: chr02\_test.vcf", "3: chr01\_test.vcf", "2: Carto.vcf", and "1: Carto.Ref.agp". The dataset "4: chr02\_test.vcf" is highlighted with a red circle, and "2: Carto.vcf" is highlighted with a red arrow pointing towards it.

# Genetic mapping analysis with vcfHunter tool

Used data: An agp file locating scaffolds in the reference assembly

The screenshot shows the Galaxy web interface with the following components:

- Left Sidebar (Tools):** A tree view of available tools categorized by function: Tools, Get Data, Send Data, BASIC TOOLS, Text Manipulation, Filter and Sort, Join, Subtract and Group, Convert Formats, SEQUENCE ANALYSIS, Gene/Protein prediction, EMBOS, Operate on Genomic Intervals, Fetch Sequences, Genomics, Fetch Alignments, Extract Features, NGS ANALYSIS, NGS: Quality Control, NGS : Mapping, NGS: GATK Tools, NGS: GATK2 Tools, NGS: SAM/BAM Manipulations, NGS: RNASeq, NGS: Assembly, NGS: Small RNAs, Bedtools, Picard Tools, SNP ANALYSIS, NGS: SNP Calling, Varscan, Population structure, GWAS, VCFtools, Tassel GBS (Version 4.0), Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs)).
- Central Table:** A table titled "##agp-version 2.0" showing genomic data. The columns are labeled 1 through 9. The data includes chromosomes (chr01), positions, strand (W or N), scaffold names, and fragment information.
- Top Bar:** Shows the URL "galaxy.southgreen.fr/galaxy/", search bar with "google", and various navigation icons.
- Right Panel (History):** A list of datasets from the "VCFHunter" workflow:
  - 102: output (list of 10 datasets)
  - 21: Kunnan\_stats.tab
  - 20: Kunnan\_AlleleOriginAndRatio.tab
  - 19: Kunnan\_Ratio.png
  - 18: Kunnan\_Cov.png
  - 9: VCF.conf (list of 5 datasets)
  - 8: Origin.tab
  - 7: chr09\_test.vcf
  - 6: chr04\_test.vcf
  - 5: chr03\_test.vcf
  - 4: chr02\_test.vcf
  - 3: chr01\_test.vcf (highlighted with a red circle)
  - 2: Carto.vcf
  - 1: CartoRef.agp (highlighted with a red arrow)

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: vcf2PopNew

The screenshot shows the Galaxy web interface with the following details:

- Left Sidebar (Tools):**
  - Tassel GBS (Version 4.0)
  - Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs))
  - GENOME HARVEST
  - TransPo-RG Transfer of Position to Resequenced Genome
  - parental SNP - Detect parental SNP of hybrids
  - Visualization
  - TraceAncestor
  - vcfHunter**
    - VCF Filter
    - vcf2allPropAndCov
    - vcf2allPropAndCovByChr
    - vcf2popNew**
    - RecombCalculatorDDose
    - Draw\_dot\_plot
  - KDE\_classifier
  - METAGENOMICS
  - FROGS
  - EVOLUTION/PHYLOGENY
  - Comparative Genomics
  - NCBI BLAST+
  - Genfam
  - Protein analyses
  - STATISTICS/GRAFICS
  - Statistics
  - Graph/Display Data
  - SOUTHGREEN PROJECTS
  - SNPlay3
  - GNPAnnot Tools
  - GNPAnnot Converters
  - ESTtik
  - Expression data
  - SAT
- Central Panel (vcf2popNew Configuration):**
  - The vcf file --vcf**: 119: chr09\_test.vcf
  - Segregation tested --seg**: (empty input field)
  - Minimal read coverage for a marker in an accession --MinCov**: 10
  - Maximal read coverage for a marker in an accession --MaxCov**: 1000
  - Window for minority allele coverage frequency to be insufficient to call a heterozygous but too high to call an homozygous --WinFreq**: 0.01:0.1
  - Minimal read number of minor allele to call variant heterozygous --MinAlCov**: 1
  - Maximal missing data proportion in the progeny (Excluding parents) --miss**: 0.1
  - Add allele coverage information to genotype file --addcov**: no
  - accessions to exclude from the filtration --NoUsed**: Nothing selected
  - accessions to exclude from the analysis --exclude**: Nothing selected
  - The reference fasta file. --ref**: Nothing selected
- Right Panel (History):**
  - Back to VCFHunter
  - output**: a list of datasets
    - RatioAndCov\_chr01\_1\_Cov.png
    - RatioAndCov\_chr01\_1\_Ratio.png
    - RatioAndCov\_chr02\_1\_Cov.png
    - RatioAndCov\_chr02\_1\_Ratio.png
    - RatioAndCov\_chr03\_1\_Cov.png
    - RatioAndCov\_chr03\_1\_Ratio.png
    - RatioAndCov\_chr04\_1\_Cov.png
    - RatioAndCov\_chr04\_1\_Ratio.png
    - RatioAndCov\_chr09\_1\_Cov.png
    - RatioAndCov\_chr09\_1\_Ratio.png

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: vcf2PopNew

The vcf file --vcf  
119: chr09\_test.vcf

Segregation test  
Several segregating markers found in the vcf file.

Minimal read coverage  
10

If a lower value is found data point is converted to missing.

Maximal read coverage for a marker in an accession --MaxCov  
1000

If a greater value is found data point is converted to missing.

Window for minority allele coverage frequency to be insufficient to call a heterozygous but too high to call an homozygous --WinFreq  
0.01:0.1

(example: "0.05:0.1"). With the example if minority allele is in ]0.05:0.1] calling will become missing for this data point.

Minimal read number of minor allele to call variant heterozygous --MinAlCov  
1

Maximal missing data proportion in the progeny (Excluding parents) --miss  
0.1

greater missing proportion will result in removing the marker.

Add allele coverage information to genotype file --addcov  
no

If this option is passed, in addition to genotypes, alleles coverage information is also filled.

accessions to exclude from the filtration --NoUsed  
Nothing selected

A tabulated file containing in one column, names of accessions to exclude from the filtration (based on missing data and p-value) but which will be kept in final files.

accessions to exclude from the analysis --exclude  
Nothing selected

A tabulated file containing in one column, names of accessions to exclude from the analysis and the files.

The reference fasta file. --ref  
Nothing selected

If passed, a tag associated to the marker will be outputted in a fasta file. This tag will contain 125 bases before the marker and 125 bases after.

remove

History

output

a list of datasets

RatioAndCov\_chr01\_1\_Cov.png

RatioAndCov\_chr01\_1\_Ratio.png

RatioAndCov\_chr02\_1\_Cov.png

RatioAndCov\_chr02\_1\_Ratio.png

RatioAndCov\_chr03\_1\_Cov.png

RatioAndCov\_chr03\_1\_Ratio.png

RatioAndCov\_chr04\_1\_Cov.png

RatioAndCov\_chr04\_1\_Ratio.png

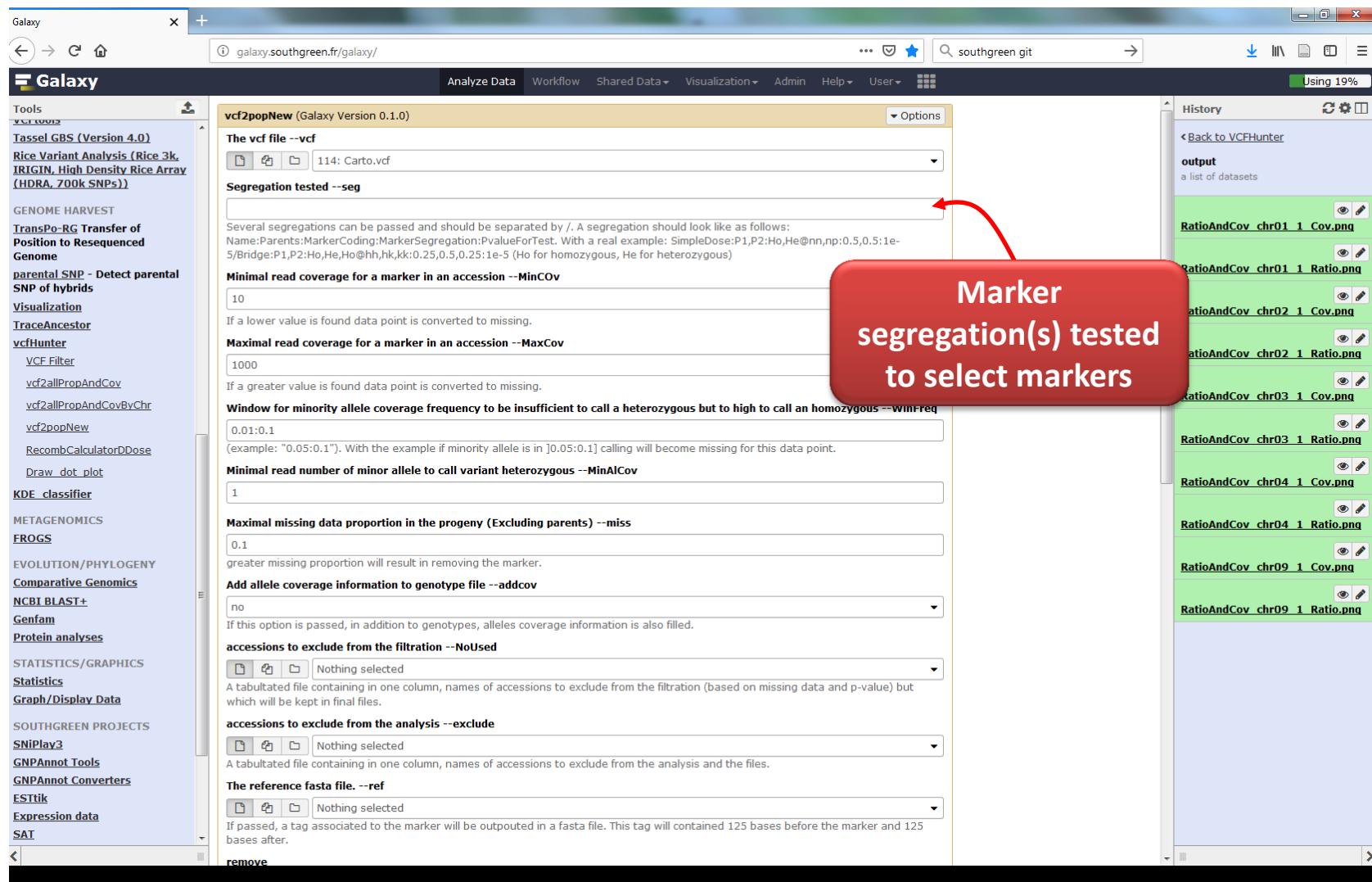
RatioAndCov\_chr09\_1\_Cov.png

RatioAndCov\_chr09\_1\_Ratio.png

Carto.vcf

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: vcf2PopNew



The screenshot shows the Galaxy web interface with the vcf2PopNew tool selected. The 'Segregation tested --seg' input field is highlighted with a red arrow and a callout box containing the text: "Marker segregation(s) tested to select markers".

**vcf2popNew (Galaxy Version 0.1.0)**

**The vcf file --vcf**  
114: Carto.vcf

**Segregation tested --seg**

Several segregations can be passed and should be separated by /. A segregation should look like as follows:  
Name:Parents:MarkerCoding:MarkerSegregation:PvalueForTest. With a real example: SimpleDose:P1,P2:Ho,He@nn,np:0.5,0.5:1e-5:Bridge:P1,P2:Ho,He,He@hh,hk,kk:0.25,0.5,0.25:1e-5 (Ho for homozygous, He for heterozygous)

**Minimal read coverage for a marker in an accession --MinCov**  
10

If a lower value is found data point is converted to missing.

**Maximal read coverage for a marker in an accession --MaxCov**  
1000

If a greater value is found data point is converted to missing.

**Window for minority allele coverage frequency to be insufficient to call a heterozygous but too high to call an homozygous --WinFreq**  
0.01:0.1

(example: "0.05:0.1"). With the example if minority allele is in ]0.05:0.1] calling will become missing for this data point.

**Minimal read number of minor allele to call variant heterozygous --MinAlCov**  
1

**Maximal missing data proportion in the progeny (Excluding parents) --miss**  
0.1

greater missing proportion will result in removing the marker.

**Add allele coverage information to genotype file --addcov**  
no

If this option is passed, in addition to genotypes, alleles coverage information is also filled.

**accessions to exclude from the filtration --NoUsed**  
Nothing selected

A tabulated file containing in one column, names of accessions to exclude from the filtration (based on missing data and p-value) but which will be kept in final files.

**accessions to exclude from the analysis --exclude**  
Nothing selected

A tabulated file containing in one column, names of accessions to exclude from the analysis and the files.

**The reference fasta file. --ref**  
Nothing selected

If passed, a tag associated to the marker will be outputted in a fasta file. This tag will contain 125 bases before the marker and 125 bases after.

**remove**

**History**

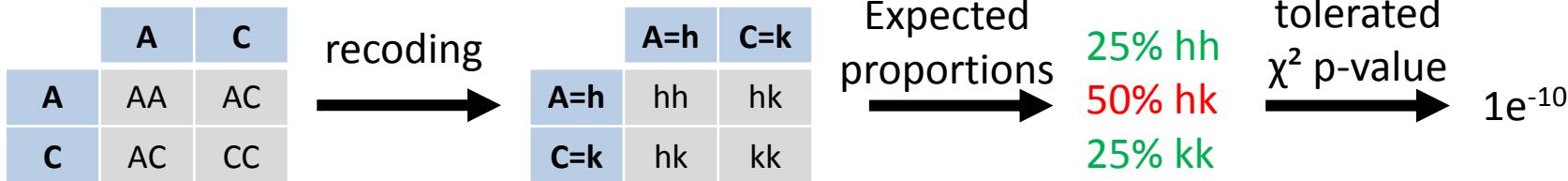
- RatioAndCov\_chr01\_1\_Cov.png
- RatioAndCov\_chr01\_1\_Ratio.png
- RatioAndCov\_chr02\_1\_Cov.png
- RatioAndCov\_chr02\_1\_Ratio.png
- RatioAndCov\_chr03\_1\_Cov.png
- RatioAndCov\_chr03\_1\_Ratio.png
- RatioAndCov\_chr04\_1\_Cov.png
- RatioAndCov\_chr04\_1\_Ratio.png
- RatioAndCov\_chr09\_1\_Cov.png
- RatioAndCov\_chr09\_1\_Ratio.png

# Genetic mapping analysis with vcfHunter tool

Selecting segregating markers in vcf files: vcf2PopNew

The expected marker segregation depends of the cross studied:

- ✓ Selfing of an heterozygous diploid accession:



Homozygous = Ho  
Heterozygous = He

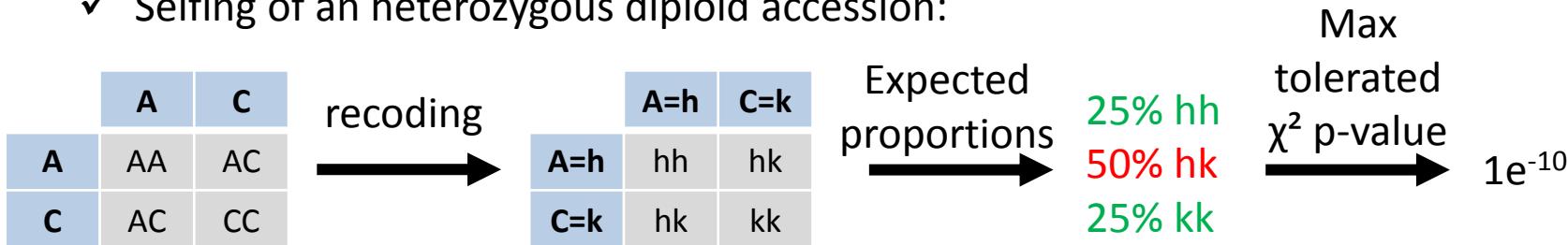
# Genetic mapping analysis with vcfHunter tool

Selecting segregating markers in vcf files: vcf2PopNew

The expected marker segregation depends of the cross studied:

- ✓ Selfing of an heterozygous diploid accession:

Homozygous = Ho  
Heterozygous = He



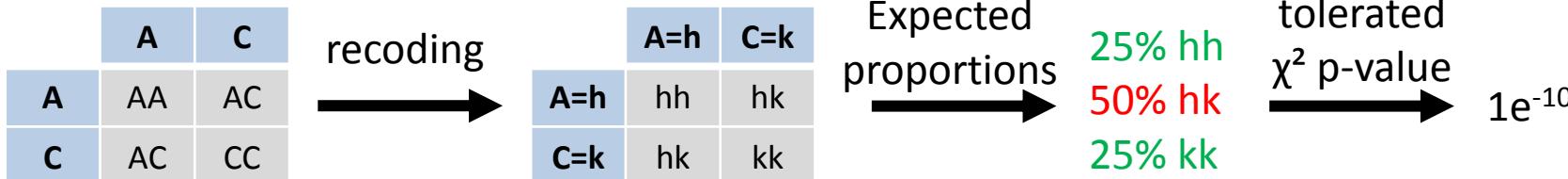
Bridge:P1,P2:Ho,He,Ho@hh,hk,kk:0.25,0.5,0.25:1e-10

# Genetic mapping analysis with vcfHunter tool

Selecting segregating markers in vcf files: vcf2PopNew

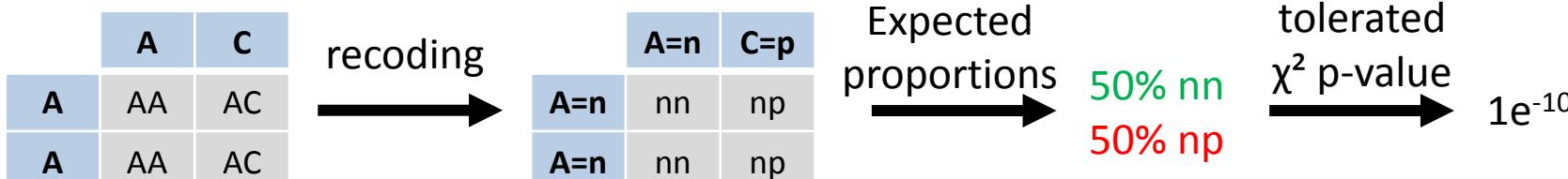
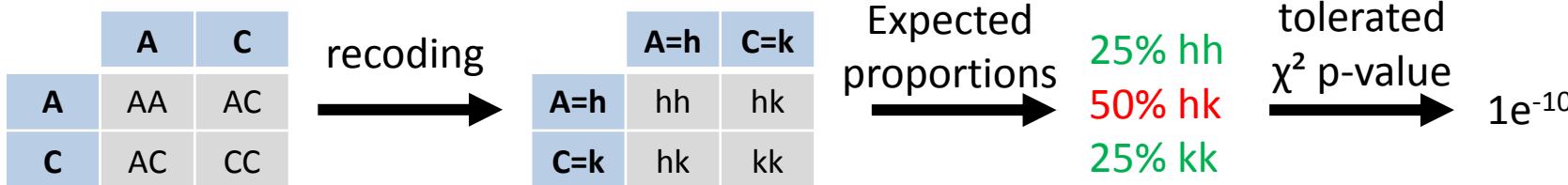
The expected marker segregation depends of the cross studied:

- ✓ Selfing of an heterozygous diploid accession:



**Bridge:P1,P2:Ho,He,Ho@hh,hk,kk:0.25,0.5,0.25:1e-10**

- ✓ Cross between two diploid heterozygous accessions:



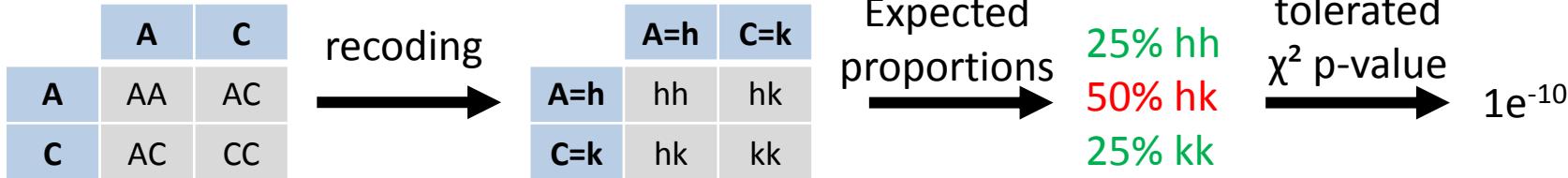
Homozygous = Ho  
Heterozygous = He

# Genetic mapping analysis with vcfHunter tool

Selecting segregating markers in vcf files: vcf2PopNew

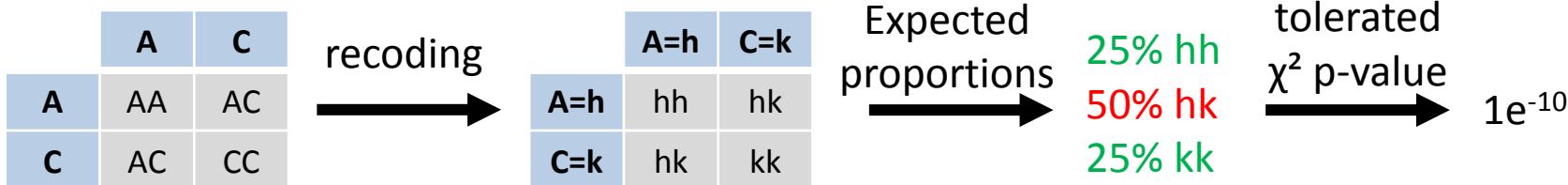
The expected marker segregation depends of the cross studied:

- ✓ Selfing of an heterozygous diploid accession:

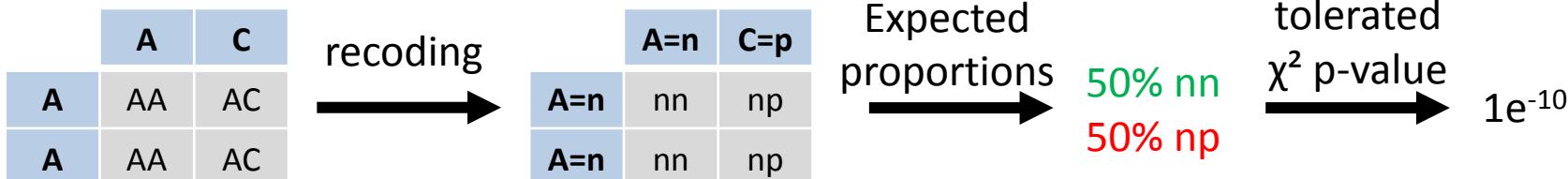


**Bridge:P1,P2:Ho,He,Ho@hh,hk,kk:0.25,0.5,0.25:1e-10**

- ✓ Cross between two diploid heterozygous accessions:



**Bridge:P1,P2:Ho,He,Ho@hh,hk,kk:0.25,0.5,0.25:1e-10**



**SimpleDose:P1,P2:Ho,He@nn,np:0.5,0.5:1e-10**

Homozygous = Ho  
Heterozygous = He

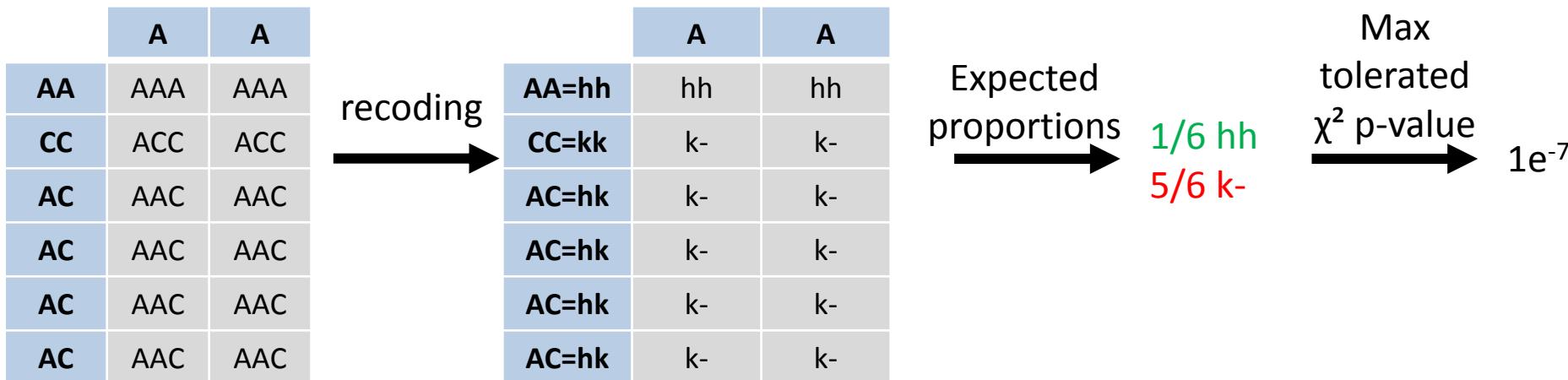
# Genetic mapping analysis with vcfHunter tool

Selecting segregating markers in vcf files: vcf2PopNew

The expected marker segregation depends of the cross studied:

Homozygous = Ho  
Heterozygous = He

- ✓ Selfing of an heterozygous diploid accession with an heterozygous tetraploid accession:  
An additional segregation tested: double dose markers (e.g: P2 genotype = A/A/C/C, P1 genotype = C/C)



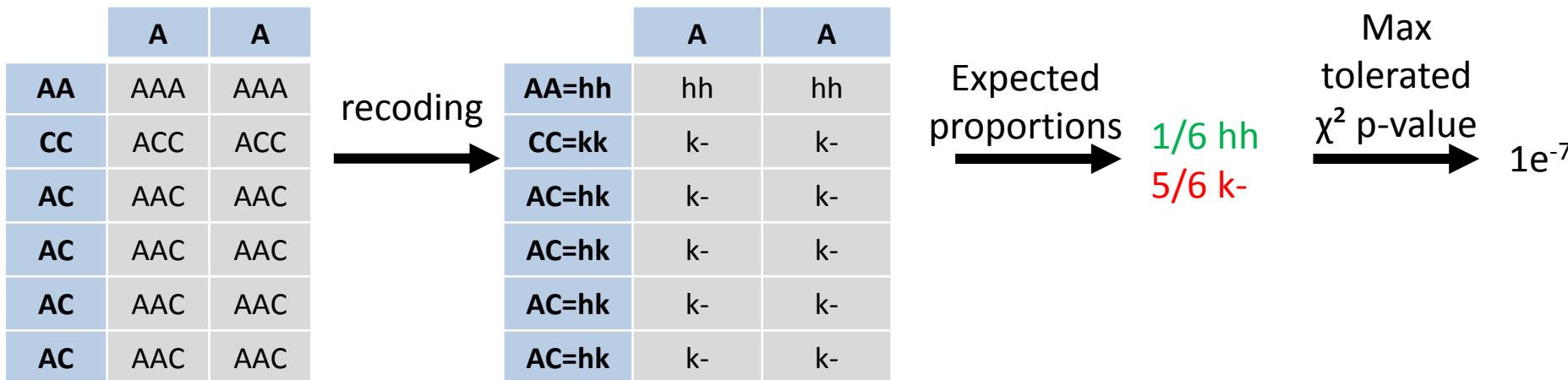
# Genetic mapping analysis with vcfHunter tool

Selecting segregating markers in vcf files: vcf2PopNew

The expected marker segregation depends of the cross studied:

Homozygous = Ho  
Heterozygous = He

- ✓ Selfing of an heterozygous diploid accession with an heterozygous tetraploid accession:  
An additional segregation tested: double dose markers (e.g: P2 genotype = A/A/C/C, P1 genotype = C/C)



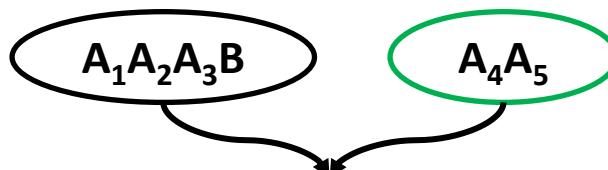
**DoubleDose:P1,P2:Ho,He@hh,k-:0.1667,0.8333:1e-7**

## Genetic mapping analysis with vcfHunter tool

Selecting segregating markers in vcf files: vcf2PopNew

The expected marker segregation depends of the cross studied:

- ✓ Selfing of an heterozygous diploid accession with an heterozygous tetraploid accession:  
An additional segregation tested: double dose markers
- ✓ This is the type of cross we have in our example!



With some regions which  
are AABB or A<sub>1</sub>A<sub>1</sub>A<sub>2</sub>B for  
example

**Genotyped progeny (180 individuals)**

- ✓ We will select markers for several segregations

**DoubleDose:P1,P2:Ho,He@hh,k-:0.1667,0.8333:1e-7**

**Bridge:P1,P2:Ho,He,Ho@hh,hk,kk:0.25,0.5,0.25:1e-10**

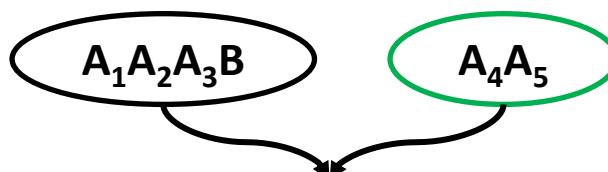
**SimpleDose:P1,P2:Ho,He@nn,np:0.5,0.5:1e-10**

## Genetic mapping analysis with vcfHunter tool

Selecting segregating markers in vcf files: vcf2PopNew

The expected marker segregation depends of the cross studied:

- ✓ Selfing of an heterozygous diploid accession with an heterozygous tetraploid accession:  
An additional segregation tested: double dose markers
- ✓ This is the type of cross we have in our example!



With some regions which  
are AABB or A<sub>1</sub>A<sub>1</sub>A<sub>2</sub>B for  
example

**Genotyped progeny (180 individuals)**

- ✓ We will select markers for several segregations

**DoubleDose:P1,P2:Ho,He@hh,k-:0.1667,0.8333:1e-7**

**Bridge:P1,P2:Ho,He,Ho@hh,hk,kk:0.25,0.5,0.25:1e-10**

**SimpleDose:P1,P2:Ho,He@nn,np:0.5,0.5:1e-10**

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: vcf2PopNew

The vcf file --vcf  
114: Carto.vcf

Segregation tested --seg

Several segregations can be passed and should be separated by /. A segregation should look like as follows:  
Name:Parents:MarkerCoding:MarkerSegregation:PvalueForTest. With a real example: SimpleDose:P1,P2:Ho,He@nn,np:0.5,0.5:1e-5  
Bridge:P1,P2:Ho,He,Ho@hh,hk,kk:0.25,0.5,0.25:1e-5 (Ho for homozygous, He for heterozygous)

Minimal read coverage for a marker in

10

If a lower value is found data point is considered missing.

Maximal read coverage for a marker in

1000

If a greater value is found data point is considered missing.

Window for minority allele coverage fraction

0.01:0.1

(example: "0.05:0.1"). With the example if minority allele is in ]0.05:0.1] calling will become missing for this data point.

Minimal read number of minor allele to call variant heterozygous --MinAlCov

1

Maximal missing data proportion in the progeny (Excluding parents) --miss

0.1

greater missing proportion will result in removing the marker.

Add allele coverage information to genotype file --addcov

no

If this option is passed, in addition to genotypes, alleles coverage information is also filled.

accessions to exclude from the filtration --NoUsed

Nothing selected

A tabulated file containing in one column, names of accessions to exclude from the filtration (based on missing data and p-value) but which will be kept in final files.

accessions to exclude from the analysis --exclude

Nothing selected

A tabulated file containing in one column, names of accessions to exclude from the analysis and the files.

The reference fasta file. --ref

Nothing selected

If passed, a tag associated to the marker will be outputted in a fasta file. This tag will contain 125 bases before the marker and 125 bases after.

remove

History

output

a list of datasets

RatioAndCov\_chr01\_1\_Cov.png

RatioAndCov\_chr01\_1\_Ratio.png

RatioAndCov\_chr02\_1\_Cov.png

RatioAndCov\_chr02\_1\_Ratio.png

RatioAndCov\_chr03\_1\_Cov.png

RatioAndCov\_chr03\_1\_Ratio.png

RatioAndCov\_chr04\_1\_Cov.png

RatioAndCov\_chr04\_1\_Ratio.png

RatioAndCov\_chr09\_1\_Cov.png

RatioAndCov\_chr09\_1\_Ratio.png

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: vcf2PopNew

The vcf file --vcf  
114: Carto.vcf

Segregation tested --seg  
:Ho,He@nn,np:0.5,0.5:1e-10/Bridge:P1,P2:Ho,He,Ho@hh,hk,kk:0.25,0.5,0.25:1e-10/DoubleDose:P1,P2:Ho,He@hh,k:-0.1667,0.8333:1e-7

Minimal read coverage for a marker in an accession --MinCov  
15

Maximal read coverage for a marker in an accession --MaxCov  
300

Window for minority allele coverage frequency to be insufficient to call a heterozygous but too high to call an homozygous --WinFreq  
0.01:0.1

Minimal read number of minor allele to call variant heterozygous --MinAlCov  
3

Maximal missing data proportion in the progeny (Excluding parents) --miss  
0.05

Add allele coverage information to genotype file --addcov  
no

accessions to exclude from the filtration --NoUsed  
Nothing selected

accessions to exclude from the analysis --exclude  
Nothing selected

The reference fasta file. --ref  
Nothing selected

remove

History

VCFHunter  
14 shown, 112 deleted, 10 hidden  
102.91 MB

126: output  
a list of datasets

125: Kunnan\_stats.tab

124: Kunnan\_AlleleOriginAndRatio.tab

123: Kunnan\_Ratio.png

122: Kunnan\_Cov.png

121: vcf.conf

120: Origin.tab

119: chr09\_test.vcf

118: chr04\_test.vcf

117: chr03\_test.vcf

116: chr02\_test.vcf

115: chr01\_test.vcf

114: Carto.vcf

113: CartoRef.agp

① 15  
② 300  
③ 0.01:0.1  
④ 3  
⑤ 0.05  
⑥ no

# Genetic mapping analysis with vcfHunter tool

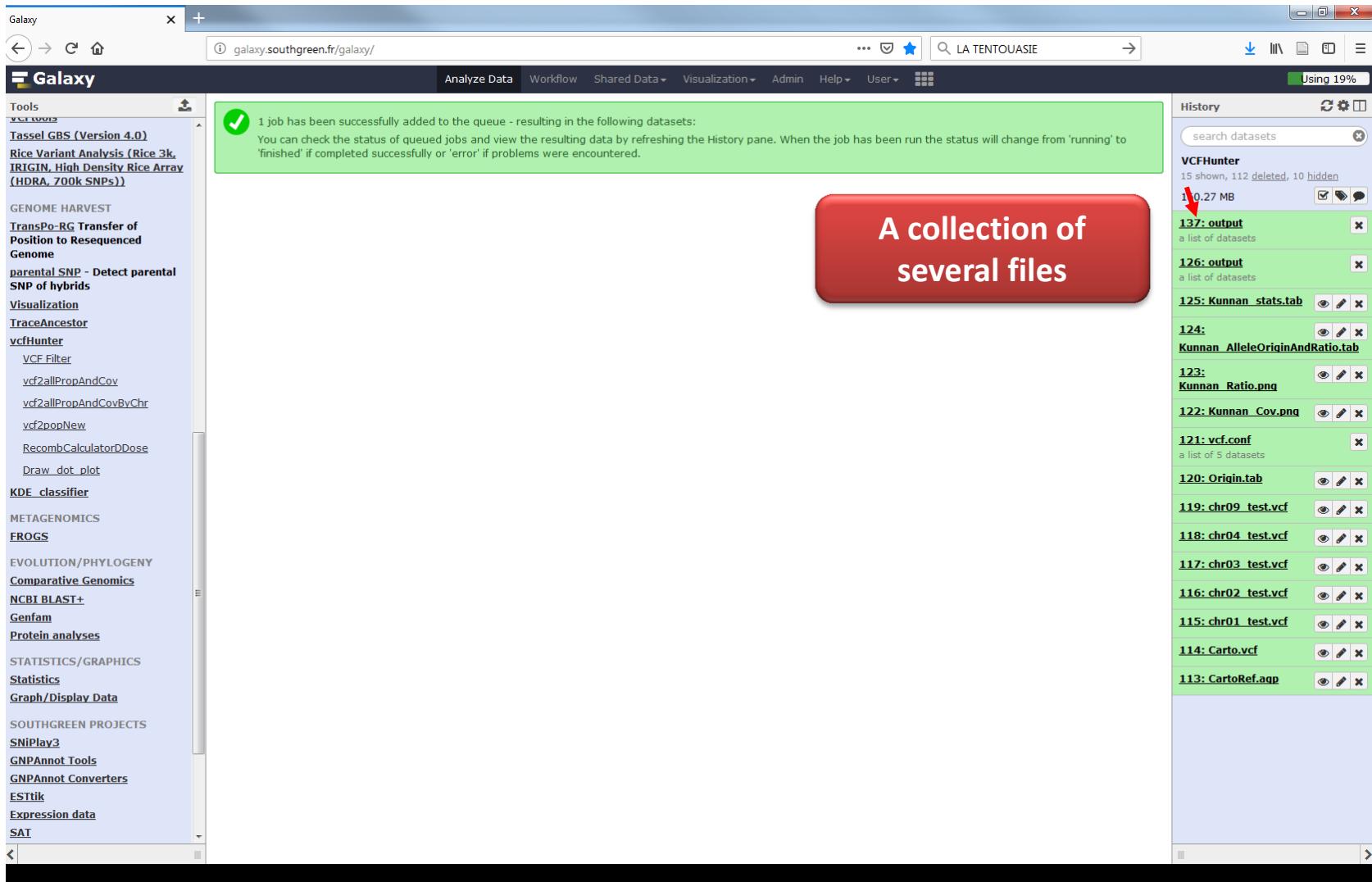
## Selecting segregating markers in vcf files: vcf2PopNew

The screenshot shows the Galaxy web interface with the following details:

- Left Sidebar (Tools):** Tassel GBS (Version 4.0), Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs)), GENOME HARVEST, TransPo-RG Transfer of Position to Resequenced Genome, parental SNP - Detect parental SNP of hybrids, Visualization, TraceAncestor, vcfHunter (selected), VCF Filter, vcf2allPropAndCov, vcf2allPropAndCovByChr, vcf2popNew, RecombCalculatorDDose, Draw\_dot\_plot, KDE\_classifier, METAGENOMICS, FROGS, EVOLUTION/PHYLOGENY, Comparative Genomics, NCBI BLAST+, Genfam, Protein analyses, STATISTICS/GRAFICS, Statistics, Graph/Display Data, SOUTHGREEN PROJECTS, SNIPlay3, GNPAAnnot Tools, GNPAAnnot Converters, ESTtik, Expression data, SAT.
- Center Panel (vcfHunter):**
  - Description:** If passed, a tag associated to the marker will be outpouted in a fasta file. This tag will contain 125 bases before the marker and 125 bases after.
  - remove:** For some programs, marker name length is limited. This option helps you to reduce marker names. By default marker name is chromosome name+M+site position. A string can be passed that will be searched and removed from all marker name. This is not necessary if your chromosome name is not too long.
  - Execute:** A button with a red circle containing the number 2.
  - Contributors:** Author Guillaume MARTIN (guillaume.martin@cirad.fr), Franc-Christophe BAURENS (franc-christophe.baurens@cirad.fr) and Olivier GARSMEUR (olivier.garsmeur@cirad.fr).
  - Support:** For any questions about Galaxy integration, please send an e-mail to auore.comte@ird.fr.
- Right Panel (History):** History search datasets. VCFHunter 14 shown, 112 deleted, 10 hidden. 102.91 MB. List of outputs:
  - 126: output (a list of datasets)
  - 125: Kunnan\_stats.tab
  - 124: Kunnan\_AlleleOriginAndRatio.tab
  - 123: Kunnan\_Ratio.png
  - 122: Kunnan\_Cov.png
  - 121: vcf.conf (a list of 5 datasets)
  - 120: Origin.tab
  - 119: chr09\_test.vcf
  - 118: chr04\_test.vcf
  - 117: chr03\_test.vcf
  - 116: chr02\_test.vcf
  - 115: chr01\_test.vcf
  - 114: Carto.vcf
  - 113: CartoRef.agp

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: outputs



The screenshot shows the Galaxy web interface with the URL [galaxy.southgreen.fr/galaxy/](http://galaxy.southgreen.fr/galaxy/). The main content area displays a green success message: "1 job has been successfully added to the queue - resulting in the following datasets: You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered." To the right, the "History" pane lists several datasets generated by the VCFHunter tool. A red callout box with the text "A collection of several files" is overlaid on the history list, pointing to the first dataset, "137: output".

A collection of several files

Dataset ID	Dataset Name	Description
137	output	a list of datasets
126	output	a list of datasets
125	Kunnan_stats.tab	
124	Kunnan_AlleleOriginAndRatio.tab	
123	Kunnan_Ratio.png	
122	Kunnan_Cov.png	
121	vcf.conf	a list of 5 datasets
120	Origin.tab	
119	chr09_test.vcf	
118	chr04_test.vcf	
117	chr03_test.vcf	
116	chr02_test.vcf	
115	chr01_test.vcf	
114	Carto.vcf	
113	CartoRef.agp	

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: outputs

This dataset is large and only the first megabyte is shown below.  
Show all | Save

② chromosome position P191 P134 P135 P136 P137 P039 P038 P132 P133 P035 P034 P037 P036 P031 P030 P03:

chromosome	position	P191	P134	P135	P136	P137	P039	P038	P132	P133	P035	P034	P037	P036	P031	P030	P03:
chr01M228	17228	A/A	A/T	A/T	A/G	A/T	./.	A/T	A/T	A/G	G/G	G/G	G/G	G/G	A/A	A/T	A/T
chr01M20916	20912	G/A	G/G	G/G	G/G	G/G	./.	C/T	T/C								
chr01M20916	20916	T/T	T/C	T/C	T/C	T/C	./.	C/T	T/C								
chr01M29287	29287	C/A	C/C	C/C	C/C	C/C	C/C	C/A									
chr01M37116	37116	T/T	T/C														
chr01M37131	37131	C/T	C/T	C/T	C/T	C/T	./.	C/T									
chr01M37155	37155	C/T	C/T	C/T	C/T	C/T	./.	C/T									
chr01M50285	50285	T/T	T/C	T/C	T/C	T/C	./.	T/T									
chr01M50313	50313	T/T	T/C	T/C	T/C	T/C	./.	T/T									
chr01M50330	50330	G/A															
chr01M53409	53409	C/C	C/G	C/G	C/G	C/G	./.	C/C	C/G	C/G	C/G	C/G	C/G	C/G	C/C	C/C	C/G
chr01M53451	53451	C/G	C/C	C/C	C/C	C/C	C/C	C/G									
chr01M59316	59316	G/A	G/G	G/A	G/A	G/A	G/A										
chr01M59428	59428	T/C	T/C	T/C	T/C	T/C	./.	T/C									
chr01M67527	67527	T/T	T/G	T/G	T/G	T/G	./.	T/G									
chr01M89917	89917	C/C	C/T	C/T	C/T	C/T	./.	C/T									
chr01M89923	89923	G/G	G/T	G/T	G/T	G/T	./.	G/G									
chr01M10518	10518	C/T	C/C	C/T	C/T	C/T	C/T										
chr01M127702	127702	T/G	T/T	T/G	T/T	T/G	./.	T/T	T/T	T/G	T/T	T/G	T/T	T/G	T/G	T/G	T/G
chr01M144740	144740	144740	144740	144740	144740	144740	144740	144740	144740	144740	144740	144740	144740	144740	144740	144740	144740
chr01M152926	152926	152926	152926	152926	152926	152926	152926	152926	152926	152926	152926	152926	152926	152926	152926	152926	152926
chr01M157399	157399	157399	157399	157399	157399	157399	157399	157399	157399	157399	157399	157399	157399	157399	157399	157399	157399
chr01M165977	165977	165977	165977	165977	165977	165977	165977	165977	165977	165977	165977	165977	165977	165977	165977	165977	165977
chr01M165990	165990	165990	165990	165990	165990	165990	165990	165990	165990	165990	165990	165990	165990	165990	165990	165990	165990
chr01M182847	182847	182847	182847	182847	182847	182847	182847	182847	182847	182847	182847	182847	182847	182847	182847	182847	182847
chr01M186534	186534	186534	186534	186534	186534	186534	186534	186534	186534	186534	186534	186534	186534	186534	186534	186534	186534
chr01M186547	186547	186547	186547	186547	186547	186547	186547	186547	186547	186547	186547	186547	186547	186547	186547	186547	186547
chr01M186553	186553	186553	186553	186553	186553	186553	186553	186553	186553	186553	186553	186553	186553	186553	186553	186553	186553
chr01M192113	192113	192113	192113	192113	192113	192113	192113	192113	192113	192113	192113	192113	192113	192113	192113	192113	192113
chr01M192120	192120	192120	192120	192120	192120	192120	192120	192120	192120	192120	192120	192120	192120	192120	192120	192120	192120
chr01M226490	226490	226490	226490	226490	226490	226490	226490	226490	226490	226490	226490	226490	226490	226490	226490	226490	226490
chr01M235828	235828	A/A	A/G	A/G	A/G	A/A	A/G	A/A	A/G	A/G	A/G	A/G	./.	A/A	A/G	A/A	A/G
chr01M246911	246911	T/C															
chr01M251887	251887	C/C	C/G	C/G	C/G	C/C	C/G	./.	C/C	C/G	C/G	C/C	C/C	C/C	C/C	C/C	C/C
chr01M26767	26767	G/G	G/T	G/T	G/T	G/G	G/T	G/G	G/T	G/T	G/T	G/T	G/T	G/G	G/G	G/G	G/G
chr01M340656	340656	C/C	C/G	C/G	C/G	C/C	C/G	C/C	C/G	C/G	C/G	C/G	C/G	C/C	C/C	C/C	C/C
chr01M381756	381756	T/T	T/A	T/A	T/A	T/T	T/A	T/T	T/A	T/A	T/A	T/A	T/T	T/T	T/T	T/T	T/T
chr01M395536	389536	C/C	C/T	C/T	C/T	C/T	./.	C/T	C/C	C/C	C/T	C/T	C/C	C/C	C/C	C/C	C/C
chr01M408718	408718	G/G	G/A	G/A	G/G	G/G	G/G	G/G	G/A	G/A	G/A	G/A	G/G	G/G	G/G	G/G	G/G
chr01M415461	415461	C/C	C/T	C/T	C/C	C/T	C/C	C/T	C/C	C/T	C/T	C/T	C/C	C/C	C/C	C/C	C/C
chr01M4466101	4466101	C/C	C/T	C/T	C/C	C/T	C/C	C/T	C/C	C/T	C/T	C/T	C/C	C/C	C/C	C/C	C/C
chr01M471076	471076	T/T	T/G	T/G	T/G	T/T	T/G	T/T	T/G	T/G	T/G	T/G	T/T	T/T	T/T	T/T	T/T
chr01M471092	471092	G/G	G/T	G/T	G/G												
chr01M480227	480227	G/G	G/T	G/T	G/G												

③ →

①

②

List of variant line passing missing data filter  
(not used after in this tutorial but can be of researcher interest)

LA TENTOUASIE

History

Back to VCFHunter

output  
a list of datasets

Pop.tab  
Pop\_report.tab  
Pop\_sub.vcf  
Pop\_tab\_Bridge.tab  
Pop\_tab\_DoubleDose\_P1.tab  
Pop\_tab\_DoubleDose\_P2.tab  
Pop\_tab\_SimpleDose\_P1.tab  
Pop\_tab\_SimpleDose\_P2.tab  
Pop\_tab\_unknown.tab

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: outputs

Galaxy

analyze Data Workflow Shared Data Visualization Admin Help User

LA TENTOUSAIE

Using 19%

History < Back to VCFHunter

output a list of datasets

Pop.tab Pop\_report.tab Pop\_sub.vcf Pop\_tab\_Bridge.tab Pop\_tab\_DoubleDose\_P1.tab Pop\_tab\_DoubleDose\_P2.tab Pop\_tab\_SimpleDose\_P1.tab Pop\_tab\_SimpleDose\_P2.tab Pop\_tab\_unknown.tab

X<sup>2</sup> p-value X<sup>2</sup> value

P148	P022	P023	P020	P021	P026	P027	P024	P025	P028	P029	P118	P-value	ChiSquare
A/T	A/A	A/T	A/T	A/T	A/T	A/T	A/A	A/A	A/T	A/A	0.3574985362366353	2.0572480225988703	
G/A	G/A	G/A	G/G	G/G	G/A	G/G	G/A	G/G	G/G	G/G	0.642410153454242	0.756480446927373	
T/C	T/T	T/C	T/C	T/C	T/C	T/C	T/T	T/T	T/C	T/C	0.6181543848150135	0.620340782122905	
C/A	C/A	C/C	0.2224172220150283	3.0064005649717513									
T/C	T/T	T/C	0.3574985362366353	0.0572480225988703									
C/T	C/C	C/T	0.6188652157518436	0.957355503159057									
C/T	C/C	C/T	0.7884679191557895	0.957355503159057									
T/C	T/T	./.	T/C	0.7884679191557895	0.4753271186440678								
G/A	./.	G/A	0.7884679191557895	0.4753271186440678									
C/G	C/C	C/G	0.4381323782633934	1.650468361581921									
C/G	C/G	C/C	0.2261394055217152	2.9732072625698325									
G/A	G/A	G/G	0.1762997977067675	3.471138674033149									
T/C	T/T	T/C	0.5170608008052784	1.319189616566847									
T/G	T/T	T/G	0.571516527594204	1.11892359550561									
C/T	C/C	C/T	0.4381323782633934	1.650468361581921									
G/T	G/G	G/T	0.4381323782633934	1.650468361581921									
C/T	C/T	C/C	0.2261394055217152	2.9732072625698325									
T/G	T/G	T/T	0.381156081165	3.7324857142857146									
A/C	0.89286592771468	0.517606519701618											
G/A	G/A	G/G	G/G	G/G	G/A	G/G	G/A	G/G	G/A	G/G	0.73075628253	4.42229438202247	
T/C	T/C	T/T	T/T	./.	T/C	T/G	T/G	T/G	T/G	T/G	0.3535405258	3.5095201117318435	
T/G	T/T	T/G	0.2316463274	0.470413966480447									
C/T	C/C	C/T	0.2316463274	0.470413966480447									
G/R	./.	G/A	0.46488715e-05	22.68171515416752									
A/G	A/A	A/G	0.381156081165	3.7324857142857146									
T/C	T/T	T/C	0.8112117096960386	0.4184524224395379									
A/T	A/A	A/T	0.4184524224395379	0.4184524224395379									
C/G	C/C	C/G	0.794167029372387	0.4609229508196721									
T/G	T/T	T/G	0.794167029372387	0.4609229508196721									
G/A	G/G	G/A	0.7528981835562427	0.5676505494505495									
G/T	./.	G/T	0.5249547448776666	1.2888644067796									
A/G	A/A	A/G	0.5249547448776666	1.2888644067796									
T/T	T/C	0.9366257998724724	1.303942876073232										
C/G	C/C	C/G	0.5249547448776666	1.2888644067796									
C/A	C/C	C/A	0.5249547448776666	1.2888644067796									
G/A	G/G	G/A	1.8989378722328712e-10	44.76911242937853									
A/G	A/A	A/G	0.6181543848150135	0.9620304782122905									
A/C	A/A	A/C	0.66431423222169	0.8180000000000001									
G/T	G/G	G/T	7068485671129155	0.693877653631285									
C/G	C/C	C/G	C/G	C/G	C/G	C/G	./.	C/G	C/G	C/G	0.7506785277537805	0.5735555555555556	
T/A	T/T	T/A	T/A	T/A	T/A	T/A	T/T	T/A	T/T	T/A	0.7068485671129155	0.693877653631285	
C/T	C/C	C/T	C/T	C/T	C/T	C/T	./.	C/C	C/C	C/T	0.8208553484317711	0.3773505617977528	
G/A	G/G	G/A	G/A	G/A	G/A	G/A	./.	G/G	G/A	G/G	0.6149273612478361	0.9725022598870057	
C/T	C/C	C/T	C/T	C/T	C/T	C/T	C/T	C/C	C/T	C/C	0.6614013865683321	0.8267887640449438	
C/T	C/C	C/T	C/T	C/T	C/T	C/T	C/T	C/C	C/T	C/C	0.7042246211332313	0.7013158192090396	
T/G	T/T	T/G	T/T	T/G	0.7068485671129155	0.693877653631285							
G/T	G/G	G/T	G/T	G/T	G/T	G/T	G/T	G/G	G/G	G/G	0.6614013865683321	0.8267887640449438	
G/A	G/G	G/A	G/A	G/A	G/A	G/A	G/A	G/G	G/A	G/G	0.6149273612478361	0.9725022598870057	
A/G	A/A	A/G	A/G	A/G	A/G	A/G	A/G	A/A	A/G	A/G	0.6614013865683321	0.8267887640449438	
T/C	T/T	T/C	T/T	T/C	0.7068485671129155	0.693877653631285							
T/A	T/T	T/A	T/T	T/A	0.7506785277537805	0.5735555555555556							
A/G	A/G	A/G	A/A	A/A	A/G	A/G	A/G	A/G	A/G	A/G	0.9860436804357866	0.028109242129956746	
G/A	G/G	G/A	G/G	G/G	0.3302183756773579	2.21600219780219							
C/T	C/C	C/C	C/T	C/T	C/T	C/T	C/T	C/T	C/C	C/T	0.482873229976734	1.456002247191012	
A/C	A/A	A/C	A/C	A/C	A/C	A/C	A/A	A/C	A/C	./.	0.5715165275941204	1.118923595505618	
A/G	A/A	A/G	A/G	A/G	A/G	A/G	A/A	A/G	A/G	A/A	0.66431423222169	0.8180000000000001	
C/T	./.	C/T	C/T	C/T	C/T	C/T	C/C	C/C	C/T	C/C	0.5249547448776666	1.28886440677966	
G/A	./.	G/A	G/A	G/A	G/A	G/A	./.	G/G	G/G	G/G	0.5249547448776666	1.28886440677966	
SAT	G/A	G/G	G/A	G/A	G/A	G/A	G/G	G/G	G/G	G/G	0.6181543848150135	0.9620340782122905	

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: outputs

Galaxy x +

galaxy.southgreen.fr/galaxy Analyze Data Workflow Shared Data Visualization Admin Help User

PASSED OPTIONS  
--vcf: /work/GALAXY/galaxy/database/files/091/dataset\_91984.dat  
--MinCov: 15  
--MaxCov: 300  
--WinFreq: 0.01:0.1  
--MinAlCov: 3  
--miss: 0.05  
--prefix: Pop

REPORT ON DATA POINTS  
Removed line due to duplicated position in vcf file: 0  
Converted to missing due to no reads: 27526 (0.9519824032316079%)  
Converted to missing due to insufficient reads coverage: 14073 (0.4867125031126359%)  
Converted to missing due to overcoverage: 80 (0.0027667874830534264%)  
Converted to missing due to too many variant: 1621 (0.05606203137537006%)  
Converted to missing due to ambiguous minority allele frequency (--WinFreq argument): 162131 (5.607275267686689%)  
Converted to missing due to ambiguous minority allele coverage (--MinAlCov argument): 0 (0.0%)

REPORT MARKER POINT  
Total marker treated: 15380  
Complete line converted to missing due to unexpected vcf format (no AD or GT or DP tags): 0 (0.0%)  
Marker with missing data on all individuals: 0 (0.0%)  
Monomorphic markers (converted to missing for convenience): 0 (0.0%)  
Di-allelic markers: 15331 (99.6814044213264%)  
More than di-allelic markers (converted to missing for convenience): 49 (0.3185955786736021%)  
Marker removed based on missing data cutoff (--miss argument): 4646 (30.208062418725618%)  
Marker removed based on ChiSquare cutoff (--pValue argument): 1836 (11.937581274382316%)  
Selected marker: 8898 (57.854356306892065%)  
Marker parsed in P1 file(s): 8178  
Marker parsed in unknown file(s): 163  
Marker parsed in Bridge file(s): 0  
Marker parsed in P2 file(s): 557

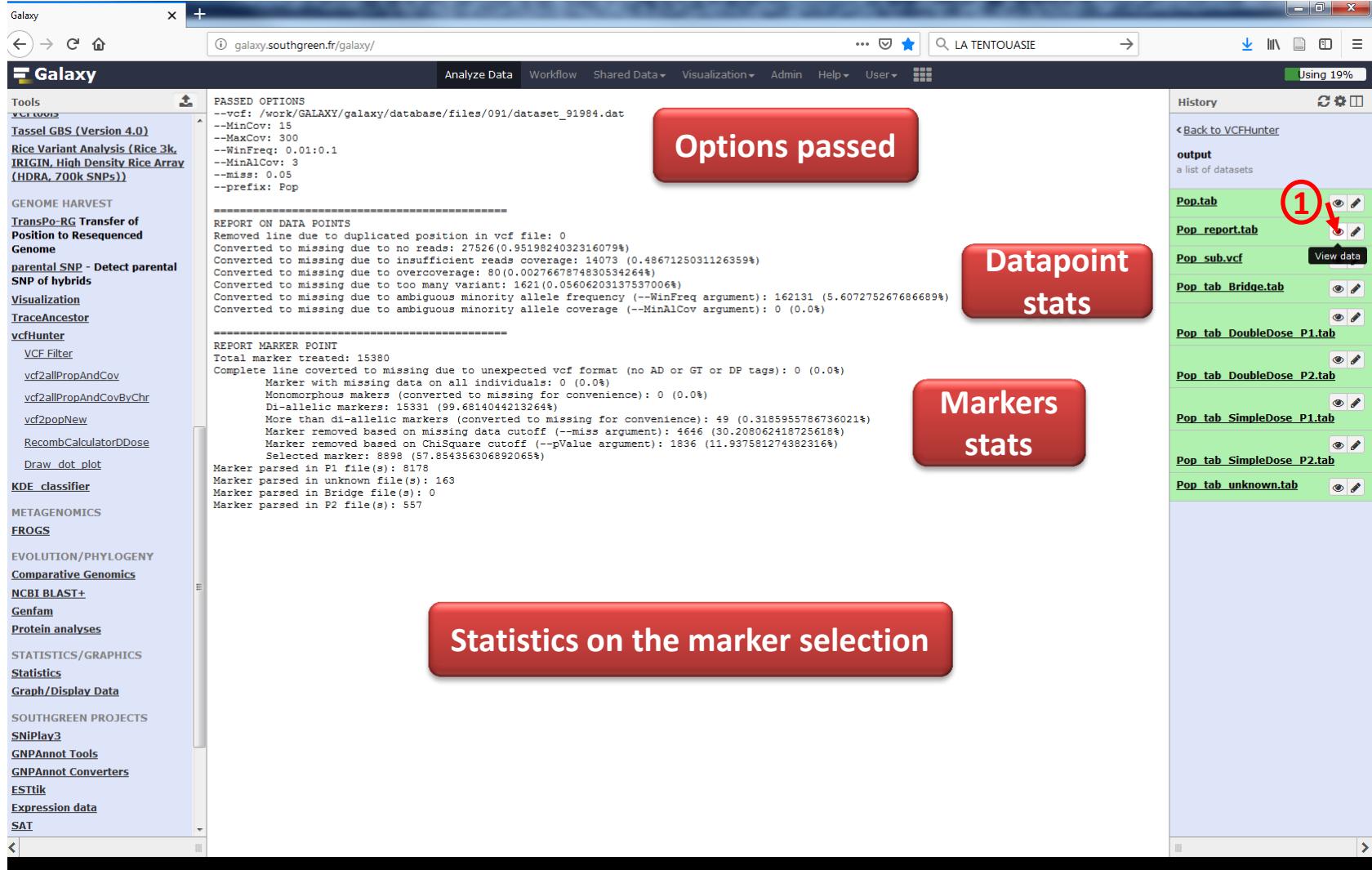
Options passed

Datapoint stats

Markers stats

Statistics on the marker selection

History  
Back to VCFHunter  
output  
a list of datasets  
**Pop.tab** (1)  
**Pop\_report.tab**  
**Pop\_sub.vcf** View data  
**Pop\_tab\_Bridge.tab**  
**Pop\_tab\_DoubleDose\_P1.tab**  
**Pop\_tab\_DoubleDose\_P2.tab**  
**Pop\_tab\_SimpleDose\_P1.tab**  
**Pop\_tab\_SimpleDose\_P2.tab**  
**Pop\_tab\_unknown.tab**



# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: outputs

This dataset is large and only the first megabyte is shown below.  
[Show all](#) | [Save](#)

```
##fileformat=VCFv4.2
##FILTER=<ID=SpCluster,Description="SNPs found in clusters">
##FORMAT=<ID=AD,Number=.,Type=Integer,Description="Allelic depths for the ref and alt alleles in the order listed">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##GATKCommandLine.VariantFiltration=<ID=VariantFiltration,CommandLineOptions="analysis_type=VariantFiltration input_file=[] showFullBamList=false read_buffer_size=1000000000 additional.filter=<ID=TAGRemoval,Description="VariantSnpCluster are removed",Date="2018-01-08 14:37:19.913025"> additional.filter=<ID=CoverageFiltration,Description="Genotype having less than 15 x coverage and less than 3 x coverage for each allele are converted to missing"> additional.filter=<ID=MissingDataFiltration,Description="SNP with more than 11 genotype missing are removed",Date="2018-11-14 13:46:44.848676"> contig=<ID=chr01,length=29070452> contig=<ID=chr02,length=29511734> contig=<ID=chr03,length=35020413>
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT P191 P134 P135 P136 P137 P039 P038 P132 P133 P035 P034 P03
chr01 17228 . A T . PASS . GT:AD:DP:GC 0/0:0/47,0/47:5.2612933261e+14 0/0:1/59,21:80:7.5557837259e+22 0/0:1/57,23
chr01 20912 . G A . PASS . GT:AD:DP:GC 0/0:1/27,19:46:65536 0/0:0/52,0/52:2.80333245379e+16 0/0:0/59,0/59:7.1027738003
chr01 20916 . T C . PASS . GT:AD:DP:GC 0/0:0/46,0/46:2.43891551199e+14 0/0:1/36,16:52:1.1651126778e+12 0/0:1/44,15
chr01 29287 . C A . PASS . GT:AD:DP:GC 0/1:1/19,22:41:64:0 0/0:0/47,1:48:98747997863.5 0/0:1/47,12:59:1.1805916207
chr01 37116 . T C . PASS . GT:AD:DP:GC 0/0:0/45,0/45:1.10765491825e+14 0/0:1/61,12:73:2.17253329946e+18 0/0:1/52,13
chr01 37131 . C T . PASS . GT:AD:DP:GC 0/0:1/33,13:46:1.09951162778e+12 0/1:1/32,41:73:2.62144.0 0/0:1/47,18:652.88
chr01 37155 . C . PASS . GT:AD:DP:GC 0/0:1/33,13:46:1.09951162778e+12 0/1:1/32,41:73:2.62144.0 0/0:1/47,18:624.50
chr01 50285 . T . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 50313 . T . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 50330 . G . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 53409 . C . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 53451 . C . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 59316 . G . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 59428 . T . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 67527 . T . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 89917 . C . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 89923 . G . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 105158 . C . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 127702 . T . PASS . GT:AD:DP:GC 0/0:1/30,12 0/0:1/31,12
chr01 144740 . A C . PASS . GT:AD:DP:GC 0/0:1/19,15:34:256.0 0/0:1/41,26:67:1073741824.0 0/0:1/49,7:56:21425805.5722
chr01 152926 . G A . PASS . GT:AD:DP:GC 0/0:1/25,15:40:1048576.0 0/0:0/60,1.59812410501e+19 0/0:1/44,12:561.84
chr01 157399 . T C . PASS . GT:AD:DP:GC 0/0:1/17,15:32:16.0 0/0:0/38,0:38:4.36927180114e+11 0/0:1/19,14:33:1024.0 0/0
chr01 165977 . T G . PASS . GT:AD:DP:GC 0/0:0/51,0:51:1.2700083428e+16 0/0:1/44,17:61:1.80143985095e+16 0/0:1/34,13
chr01 165990 . C T . PASS . GT:AD:DP:GC 0/0:0/51,0:51:1.2700083428e+16 0/0:1/42,17:59:1.12589990684e+15 0/0:1/34,13
chr01 182847 . G A . PASS . GT:AD:DP:GC 0/0:1/30,14:44:4294967296.0 0/1:1/18,50:68:1.84467440737e+19 0/1:1/27,36
chr01 186534 . A G . PASS . GT:AD:DP:GC 0/0:0/52,0:52:2.8833245379e+16 0/0:1/44,15:59:2.88230376152e+17 0/0:1/41,14
chr01 186547 . T C . PASS . GT:AD:DP:GC 0/0:1/38,14:52:2.81474976711e+14 0/0:1/30,29:59:4.0 0/0:1/35,19:54:4294
chr01 186553 . A T . PASS . GT:AD:DP:GC 0/0:1/38,14:52:2.81474976711e+14 0/0:1/30,29:59:4.0 0/0:1/34,19:53:1073
chr01 192113 . C G . PASS . GT:AD:DP:GC 0/0:0/36,0:36:1.936338422010.7 0/0:1/32,40:2.81474976711e+14 0/0:1/38,10:48:7.20
chr01 192120 . T G . PASS . GT:AD:DP:GC 0/0:0/36,0:36:1.936338422010.7 0/0:1/31,8:39:7.0368741777e+13 0/0:1/38,10:48:7.20
chr01 192136 . G A . PASS . GT:AD:DP:GC 0/0:0/36,0:36:1.936338422010.7 0/0:1/31,8:39:7.0368741777e+13 0/0:1/39,10:49:2.88
chr01 226490 . G T . PASS . GT:AD:DP:GC 0/0:0/42,0:42:1.04092203766e+13 0/0:1/52,21:73:4.16168601843e+18 0/0:1/48,17
chr01 235828 . A T,G . PASS . GT:AD:DP:GC 0/0:0/31,0:31:1875986614.41 0/0:2/52,0:52:1.0552395351.5 0/0:2/40,0:16:56:63
chr01 246911 . T C . PASS . GT:AD:DP:GC 0/0:0/31,0:31:1875986614.41 0/0:2/52,0:52:1.0552395351.5 0/0:2/40,0:16:56:63
chr01 251887 . C G . PASS . GT:AD:DP:GC 0/0:0/52,0:52:2.80333245379e+16 0/0:1/56,13:69:2.76454365653e+22 0/0:1/36,12
chr01 251913 . C N . PASS . GT:AD:DP:GC 0/0:0/52,0:52:2.80333245379e+16 0/0:1/56,13:69:2.76454365653e+22 0/0:1/36,12
```

Vcf containing variant line passing missing data filter (not used after in this tutorial but can be of researcher interest)

History < Back to VCFHunter  
output a list of datasets  
Pop.tab  
Pop\_report.tab  
Pop\_sub.vcf  
Pop\_tab\_Bridge.tab View data  
Pop\_tab\_DoubleDose\_P1.tab  
Pop\_tab\_DoubleDose\_P2.tab  
Pop\_tab\_SimpleDose\_P1.tab  
Pop\_tab\_SimpleDose\_P2.tab  
Pop\_tab\_unknown.tab

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: outputs

The screenshot shows the Galaxy web interface with the URL [galaxy.southgreen.fr/galaxy/](http://galaxy.southgreen.fr/galaxy/). The left sidebar lists various bioinformatics tools and projects. The main workspace displays a table of segregating markers with columns: Marker, coding, ratio, rephased, and P191, P134, P135, P136, P137, P039, P038, P132, P133, P035, P034, P037, P036, P038. To the right, a list of output datasets is shown under the heading "output: a list of datasets". The datasets include:

- Pop.tab
- Pop\_report.tab
- Pop\_sub.vcf
- Pop\_tab\_Bridge.tab (circled with a red number 1)
- Pop\_tab\_DoubleDose\_P1.tab
- Pop\_tab\_DoubleDose\_P2.tab
- Pop\_tab\_SimpleDose\_P1.tab
- Pop\_tab\_SimpleDose\_P2.tab
- Pop\_tab\_unknown.tab

A large red callout box highlights the "Pop\_tab\_Bridge.tab" dataset, stating: "This is the file which should contain bridge markers is empty because there are no bridge markers".

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: outputs

This file contained double dose markers which are heterozygous in parent P1

Phase information (1 = marker has been rephased)

Expected proportion

Segregation type

Marker name (Chr+"M"+Pos)

The screenshot shows a Galaxy workflow interface with a VCF file loaded. A prominent red callout box in the center contains the text: "This file contained double dose markers which are heterozygous in parent P1". Below this, several annotations are overlaid on the VCF data:

- Phase information (1 = marker has been rephased):** Points to a marker with a double dose value (0.1667) and a rephased status (1).
- Expected proportion:** Points to another marker with a double dose value (0.1667) and a rephased status (1).
- Segregation type:** Points to a marker with a double dose value (0.1667) and a rephased status (1).
- Marker name (Chr+"M"+Pos):** Points to a marker with a double dose value (0.1667) and a rephased status (1).

The VCF data table shows various markers across chromosomes, with some markers having double dose values (e.g., 0.1667, 0.8333) and others having single dose values (e.g., 0, 1). The rephased column indicates whether the marker has been rephased.

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: outputs

The screenshot shows the Galaxy web interface with a VCF file analysis output. The main panel displays a table of marker data across various samples (P191, P134, P135, P136, P137, P039, P038, P132, P133, P035, P034, P037, P036, P038). The table includes columns for Marker, coding, ratio, and rephased values. A red callout box highlights the entry for marker chr01M4599854, which has a ratio of 0.1667, 0.8333. The right panel shows a list of datasets, with one item circled in red and labeled with a '1'.

This file contained double dose markers  
which are heterozygous in parent P2.

They are very few and this is good because  
P2 is diploid and thus double dose markers  
should not exist!

Marker	coding	ratio	rephased	P191	P134	P135	P136	P137	P039	P038	P132	P133	P035	P034	P037	P036	P038
chr01M4599854	hh, k-	0.1667, 0.8333	0	k-	k-	k-	k-	k-	hh	k-							
chr02M24813339	hh, k-	0.1667, 0.8333	0	k-	k-	k-	k-	hh	k-								
chr03M27342905	hh, k-	0.1667, 0.8333	0	k-													
chr03M29789501	hh, k-	0.1667, 0.8333	0	hh	k-	hh	k-	k-	k-	k-							
chr03M31059415	hh, k-	0.1667, 0.8333	0	k-													

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: outputs

Galaxy x +

galaxy.southgreen.fr/galaxy/ Analyze Data Workflow Shared Data Visualization Admin Help User History < Back to VCFHunter output a list of datasets

This dataset is large and only the first megabyte is shown below.  
Show all | Save

Marker	coding	ratio	rephased	P191	P134	P135	P136	P137	P039	P038	P132	P133	P035	P034	P037	P036	P031	P030	P03:
chr01M17228	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	--	np	np	np	nn	nn	np	nn	nn
chr01M20912	nn,np	0.5,0.5	0	np	nn	nn	np	np	np	np	nn	np	nn	np	nn	nn	np	nn	nn
chr01M20916	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M29287	nn,np	0.5,0.5	0	np	--	np	nn												
chr01M37116	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M50285	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M50313	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	np	nn	np	np	np	nn	nn	np	nn	nn
chr01M53409	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	--	nn	np	np	np	--	nn	np	nn	nn
chr01M53451	nn,np	0.5,0.5	0	np	nn	np	nn	np	nn	np	nn	np	np	--	np	--	np	np	nn
chr01M59316	nn,np	0.5,0.5	0	np	nn	np	nn	np	nn	np	nn	np	np	nn	np	np	np	np	nn
chr01M67527	nn,np	0.5,0.5	0	nn	np	np	np	--	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M89917	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M89923	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	np	np	nn	nn
chr01M105158	nn,np																		
chr01M127702	nn,np																		
chr01M152926	nn,np																		
chr01M157399	nn,np																		
chr01M165977	nn,np																		
chr01M165990	nn,np																		
chr01M186534	nn,np																		
chr01M192113	nn,np																		
chr01M192120	nn,np																		
chr01M192136	nn,np																		
chr01M226490	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	np	nn	nn	nn
chr01M235828	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	--	nn	np	nn	nn
chr01M251887	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	--	nn	np	np	np	nn	nn	np	nn	nn
chr01M251913	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	--	nn	np	np	np	nn	nn	np	nn	nn
chr01M251923	nn,np	0.5,0.5	0	nn	nn	nn	--	nn	nn	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M310320	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M310368	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M326767	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M340656	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M381756	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M389536	nn,np	0.5,0.5	0	nn	np	np	np	--	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M408718	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M4415461	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M466101	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M471076	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M471092	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M488027	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M502137	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M502140	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M502151	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M511685	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	--	np	nn	nn
chr01M512201	nn,np	0.5,0.5	0	np	np	np	np	nn	np	np	nn	np	nn						
chr01M528716	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M534858	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M548511	nn,np	0.5,0.5	0	nn	np	np	np	nn	np	nn	nn	np	np	np	nn	nn	np	nn	nn
chr01M563242	nn,np	0.5,0.5	0	np															

This file contained simple dose markers which are heterozygous in parent P1

History < Back to VCFHunter  
output a list of datasets  
**Pop.tab**   
**Pop\_report.tab**   
**Pop\_sub.vcf**   
**Pop\_tab\_Bridge.tab**   
**Pop\_tab\_DoubleDose\_P1.tab** 1   
**Pop\_tab\_DoubleDose\_P12.tab**   
**Pop\_tab\_SimpleDose\_P1.tab**   
**Pop\_tab\_SimpleDose\_P2.tab**   
**Pop\_tab\_unknown.tab**

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: outputs

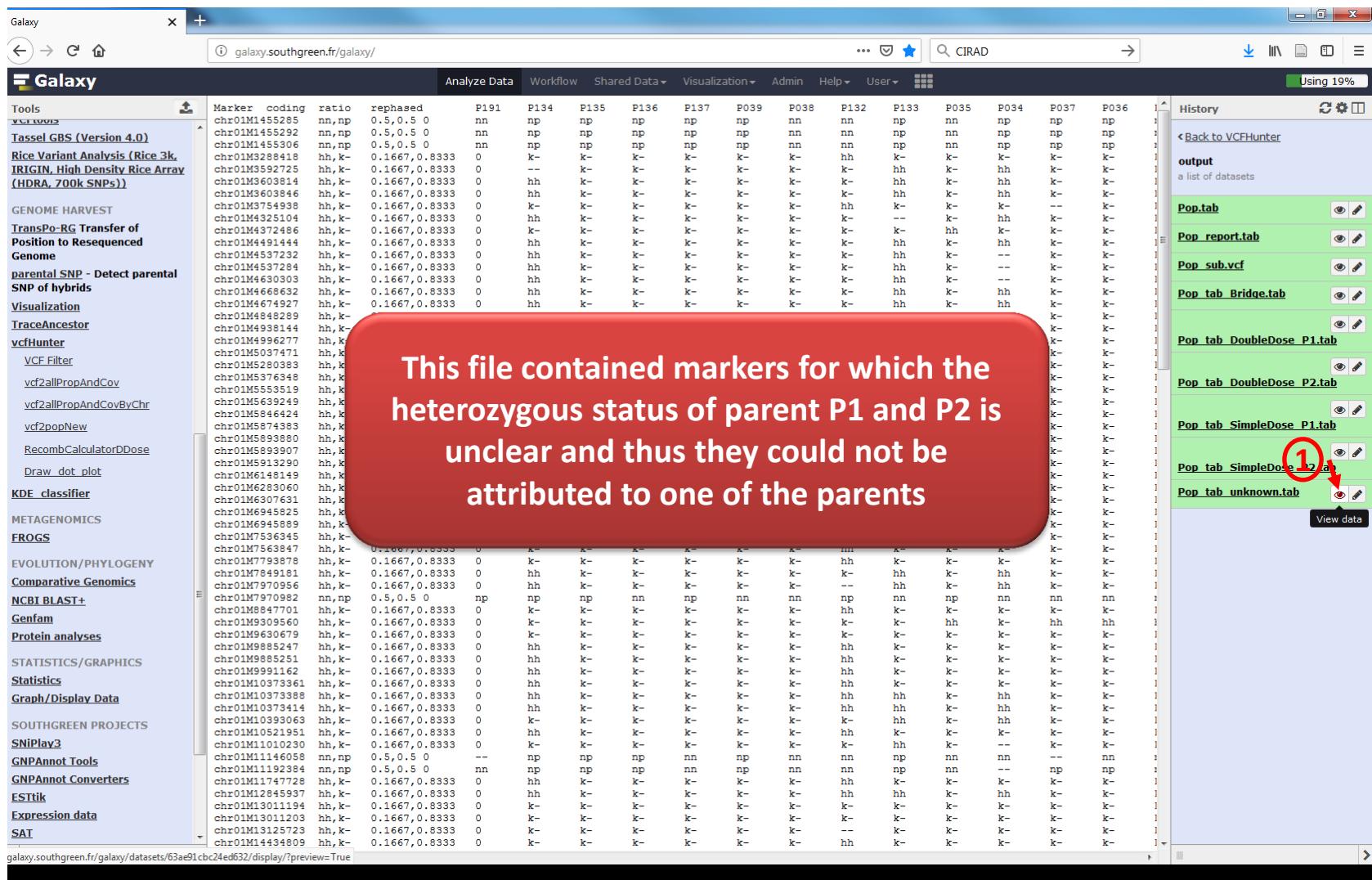
This file contained simple dose markers  
which are heterozygous in parent P2

Marker	coding	ratio	rephased	P191	P134	P135	P136	P137	P039	P038	P132	P133	P035	P034	P037	P036	
chr01M2909259	nn,np	0.5,0.5	0	np	np	nn	np	nn	nn	np	nn	np	nn	np	np	np	np
chr01M3084769	nn,np	0.5,0.5	0	np	np	nn	np	nn	nn	np	nn	np	nn	np	np	np	np
chr01M3230825	nn,np	0.5,0.5	0	np	np	nn	np	nn	nn	np	nn	np	nn	np	np	np	np
chr01M3285013	nn,np	0.5,0.5	0	nn	nn	nn	np	nn	nn	np	nn	np	nn	np	nn	nn	nn
chr01M3358254	nn,np	0.5,0.5	0	nn	nn	nn	np	nn	nn	np	nn	np	nn	np	nn	nn	nn
chr01M3395458	nn,np	0.5,0.5	0	np	np	nn	np	nn	nn	np	nn	np	nn	np	nn	nn	nn
chr01M3417999	nn,np	0.5,0.5	0	np	np	nn	np	nn	nn	np	nn	np	nn	np	nn	nn	nn
chr01M3493115	nn,np	0.5,0.5	0	np	np	nn	np	nn	nn	np	nn	np	nn	np	nn	np	np
chr01M3507847	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M3580672	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M3587744	nn,np	0.5,0.5	0	np	np	--	np	nn	np	np	nn	np	np	np	np	np	np
chr01M3775407	nn,np	0.5,0.5	0	np	np	nn	np	nn	nn	np	nn	np	nn	np	--	np	np
chr01M3775423	nn,np	0.5,0.5	0	--	nn	nn	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M3837583	nn,np	0.5,0.5	0	nn	nn	nn	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M3837669	nn,np	0.5,0.5	0	nn	nn	np	nn	np	np	np	nn	np	nn	np	nn	nn	nn
chr01M3847042	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	np	np	nn	np	np
chr01M3856981	nn,np	0.5,0.5	0	nn	nn	np	nn	np	np	np	nn	np	nn	np	nn	nn	nn
chr01M3881851	nn,np	0.5,0.5	0	nn	nn	np	nn	np	np	np	nn	np	nn	np	nn	nn	nn
chr01M3901090	nn,np	0.5,0.5	0	nn	nn	np	nn	np	np	np	nn	np	nn	np	nn	nn	nn
chr01M3901141	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M3913984	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M3926703	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M3958080	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4025771	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4025260	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4047381	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4047390	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4053672	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4127313	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4145915	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4150482	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4163170	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4167085	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4185588	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4222080	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4235526	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	nn	nn	nn
chr01M4242788	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	--	np	np
chr01M4242794	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	--	np	np
chr01M4269479	nn,np	0.5,0.5	0	--	nn	nn	np	nn	np	np	nn	np	nn	np	--	np	np
chr01M4288400	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4288430	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	--	np	np
chr01M4325073	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4334576	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4366480	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4426286	nn,np	0.5,0.5	0	nn	nn	np	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4426367	nn,np	0.5,0.5	0	--	nn	nn	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4426401	nn,np	0.5,0.5	0	--	nn	nn	np	np	np	np	nn	np	nn	np	--	nn	nn
chr01M4435206	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4435217	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4435339	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4526241	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4537287	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4557883	nn,np	0.5,0.5	0	--	nn	nn	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4577078	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	--	nn	nn
chr01M4598287	nn,np	0.5,0.5	0	nn	nn	np	nn	np	np	nn	nn	np	nn	np	--	nn	nn
chr01M4616823	nn,np	0.5,0.5	0	nn	nn	np	nn	np	np	nn	nn	np	nn	np	--	nn	nn
chr01M4630312	nn,np	0.5,0.5	0	nn	nn	np	nn	np	np	nn	nn	np	nn	np	--	nn	nn
chr01M4668683	nn,np	0.5,0.5	0	nn	nn	np	nn	np	np	nn	nn	np	nn	np	--	nn	nn
chr01M4692503	nn,np	0.5,0.5	0	np	np	nn	np	nn	np	np	nn	np	nn	np	--	nn	nn

# Genetic mapping analysis with vcfHunter tool

## Selecting segregating markers in vcf files: outputs

This file contained markers for which the heterozygous status of parent P1 and P2 is unclear and thus they could not be attributed to one of the parents



The screenshot shows the Galaxy web interface with a VCF file open. The left sidebar lists various tools and projects. A red callout box highlights a specific note about markers where parent status is unclear. On the right, a sidebar shows a list of output datasets, with 'Pop.tab' circled in red.

Tools

- Tassel GBS (Version 4.0)
- Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs))
- GENOME HARVEST
- TransPo-RG Transfer of Position to Resequenced Genome
- parental SNP - Detect parental SNP of hybrids
- Visualization
- TraceAncestor
- vcfHunter
  - VCF Filter
  - vcf2allPropAndCov
  - vcf2allPropAndCovByChr
  - vcf2popNew
  - RecombCalculatorDDose
  - Draw\_dot\_plot
- KDE\_classifier
- METAGENOMICS
- FROGS
- EVOLUTION/PHYLOGENY
- Comparative Genomics
- NCBI BLAST+
- Genfam
- Protein analyses
- STATISTICS/GRAFICS
- Statistics
- Graph/Display Data
- SOUTHGREEN PROJECTS
- SNIPlays
- GNPAnnot Tools
- GNPAnnot Converters
- ESTtik
- Expression data
- SAT

Analyze Data Workflow Shared Data Visualization Admin Help User

History

CIRAD

Using 19%

Pop.tab

Pop\_report.tab

Pop\_sub.vcf

Pop\_tab\_Bridge.tab

Pop\_tab\_DoubleDose\_P1.tab

Pop\_tab\_DoubleDose\_P2.tab

Pop\_tab\_SimpleDose\_P1.tab

Pop\_tab\_SimpleDose\_P2.tab

Pop\_tab\_unknown.tab

# Genetic mapping analysis with vcfHunter tool

## Calculating marker segregation distortion on parent P2 simple dose markers

The screenshot shows the Galaxy web interface with the following details:

- Tools Panel:** On the left, under the "vcfHunter" category, the "RecombCalculatorDDose" tool is highlighted with a red circle and arrow.
- Tool Configuration:** The "RecombCalculatorDDose (Galaxy Version 0.1.0)" tool is set up with:
  - The marker file matrix:** Input is "125: Kunman\_stats.tab" (output of vcf2popNew).
  - Are marker phased?**: Input is "n".
  - Analysis to perform:** Input is "R". Description: "R: Calculate recombination rate / S: Calculate segregation distortions".
- History Panel:** On the right, a list of datasets is shown, including "Pop.tab", "Pop\_report.tab", "Pop\_sub.vcf", "Pop\_tab\_Bridge.tab", "Pop\_tab\_DoubleDose\_P1.tab", "Pop\_tab\_DoubleDose\_P2.tab", "Pop\_tab\_SimpleDose\_P1.tab", "Pop\_tab\_SimpleDose\_P2.tab", and "Pop\_tab\_unknown.tab".

# Genetic mapping analysis with vcfHunter tool

Calculating marker segregation distortion on parent P2 simple dose markers

The screenshot shows the Galaxy web interface with the 'RecombCalculatorDDose' tool selected. The left sidebar lists various tools and projects. The main panel shows the 'RecombCalculatorDDose' tool configuration, including input and output fields. A red callout box highlights a file named 'Pop.tab' in the 'output' section of the history panel, stating: 'File which are in collections are not available through by galaxy workflow!!!'. Below this, another callout box states: 'The problem is that we want to access the "Pop\_tab\_SimpleDose\_P2.tab" file...'. To solve this, the instructions are: 1- get this file on our computer and 2- upload this file on galaxy but this time not as part of a collection.

File which are in collections are not available through by galaxy workflow!!!

The problem is that we want to access the "Pop\_tab\_SimpleDose\_P2.tab" file...

To solve this problem, we have to:

- 1- get this file on our computer
- 2- upload this file on galaxy but this time not as part of a collection

# Genetic mapping analysis with vcfHunter tool

## 1- Downloading the Pop\_tab\_SimpleDose\_P2.tab

The screenshot shows the Galaxy web interface with the following components:

- Left Sidebar:** Contains a list of tools and projects, including Tassel GBS (Version 4.0), RecombCalculatorDDose (Galaxy Version 0.1.0), and various genome harvest and VCF Hunter tools.
- Middle Panel:** Displays the RecombCalculatorDDose tool configuration. It includes fields for "The marker file matrix" (set to 125: Kunnan\_stats.tab) and "Analysis to perform" (set to R: Calculate recombination rate / S: Calculate segregation distortions). A red circle labeled **3** points to the "Execute" button.
- Center Panel:** A modal dialog titled "Ouverture de Galaxy145-[Pop\_tab\_SimpleDose\_P2.tab].null" is displayed. It shows the file path as "Galaxy145-[Pop\_tab\_SimpleDose\_P2.tab].null" and the type as "null File (321 Ko) à partir de: http://galaxy.southgreen.fr". It contains options to "Ouvrir avec" (Open with) or "Enregistrer le fichier" (Save file), with the latter being selected. A red circle labeled **4** points to the "OK" button.
- Right Panel:** The History panel shows a list of datasets:
  - Pop.tab
  - Pop\_report.tab
  - Pop\_sub.vcf
  - Pop\_tab\_Bridge.tab
  - Pop\_tab\_DoubleDose\_P1.tab
  - Pop\_tab\_DoubleDose\_P2.tab
  - Pop\_tab\_SimpleDose\_P1.tab (highlighted with a red circle labeled **1**)
  - Pop\_tab\_SimpleDose\_P2.tab (highlighted with a red circle labeled **2**)
  - Pop\_tab\_unknown.tab

# Genetic mapping analysis with vcfHunter tool

## 1- Uploading the Pop\_tab\_SimpleDose\_P2.tab onto Galaxy

The screenshot shows the Galaxy web interface with the following details:

- Tools Panel (Left):** Shows a sidebar with various bioinformatics tools categorized under sections like BASIC TOOLS, SEQUENCE ANALYSIS, NGS ANALYSIS, and SNP ANALYSIS.
- Tool Configuration (Center):** The "RecombCalculatorDDose (Galaxy Version 0.1.0)" tool is selected. The configuration includes:
  - Get Data:** A red arrow labeled "2" points to the "Upload File from your computer" button, and a red circle labeled "3" points to the file input field containing "125: Kunnan\_stats.tab".
  - The marker file matrix:** Set to "output of vcftoolsNew".
  - Are marker phased?**: Set to "n".
  - Analysis to perform:** Set to "R: Calculate recombination rate / S: Calculate segregation distortions".
  - Execute:** A blue "Execute" button.
- Help and Support (Bottom Left):** Includes links for the author, Galaxy integration, and support.
- Description and Details (Bottom Center):** Describes the tool's purpose and provides input and output specifications, along with citation information.
- History Panel (Right):** Shows a list of datasets uploaded to the history, including "Pop.tab", "Pop\_report.tab", "Pop\_sub.vcf", "Pop\_tab\_Bridge.tab", "Pop\_tab\_DoubleDose\_P1.tab", "Pop\_tab\_DoubleDose\_P2.tab", "Pop\_tab\_SimpleDose\_P1.tab", "Pop\_tab\_SimpleDose\_P2.tab", and "Pop\_tab\_unknown.tab".

# Genetic mapping analysis with vcfHunter tool

## 1- Uploading the Pop\_tab\_SimpleDose\_P2.tab onto Galaxy

The screenshot shows the Galaxy web interface with a floating "Download from web or upload from disk" dialog box. The dialog lists a single dataset: "Galaxy145-[Pop\_tab\_SimpleDose\_P2.tab].null" with a size of 321.5 KB. The dialog has tabs for "Regular" and "Composite", and buttons for "Choose local file", "Choose FTP file", "Paste/Fetch data", "Pause", "Reset", "Start", and "Close". A red arrow points from a circled '1' in the left sidebar to the "Start" button. Another red arrow points from a circled '2' to the "Close" button. A third red arrow points from a circled '3' to the "Start" button.

Drag and drop the downloaded file

1

2

3

Inputs:

The marker file matrix = Pop\_tab\_SegregationName\_Parent.tab from vcftoolsNew

Output:

REC.tab: A tabulated file of pairwise marker recombination rate (analysis to perform R).  
SegDist.tab: A tabulated file of marker segregation distortions (analysis to perform S).

Citations  Show BibTeX

Baurens, Franc-Christophe and Martin, Guillaume and Hervouet, Catherine and Salmon, Frédéric and Yohomé, David and Ricci, Sébastien and Rouard, Mathieu and Habas, Remy and Lemainque, Arnaud and Yahiaoui, Nabila and et al. (2018). Recombination and large structural variations shape interspecific edible banana genomes. In *Molecular Biology and Evolution*, [doi:10.1093/molbev/msy199][Link]

# Genetic mapping analysis with vcfHunter tool

## Calculating marker segregation distortion on parent P2 simple dose markers

The screenshot shows the Galaxy web interface with the following steps highlighted:

1. The marker file matrix dropdown is set to "147: Galaxy145-[Pop\_tab\_SimpleDose\_P2.tab].null".
2. The "Are marker phased?" dropdown is set to "n".
3. The "Analysis to perform" dropdown is set to "S".
4. The "Execute" button is highlighted.

A red dashed arrow points from step 4 to the "History" panel on the right, which shows a new entry: "147: Galaxy145-[Pop\_tab\_SimpleDose\_P2.tab].null". A callout box labeled "Downloaded file appears" is positioned over this entry.

**RecombCalculatorDDose**

**Description**

This program performs recombination frequencies between two pairs of markers. It can also calculate marker segregation distortion.

**Inputs:**

The marker file matrix = Pop\_tab\_SegregationName\_Parent.tab from vcf2popNew

**Output:**

REC.tab: A tabulated file of pairwise marker recombination rate (analysis to perform R).  
SegDist.tab: A tabulated file of marker segregation distortions (analysis to perform S).

**Citations**  Show BibTeX

Baurens, Franc-Christophe and Martin, Guillaume and Hervouet, Catherine and Salmon, Frédéric and Yohomé, David and Ricci, Sébastien and Rouard, Mathieu and Habas, Remy and Lemainque, Arnaud and Yahiaoui, Nabila and et al. (2018). Recombination and large structural variations shape interspecific edible banana genomes. In *Molecular Biology and Evolution*, [doi:10.1093/molbev/msy199][Link]

**History**

- VCFHunter
- 16 shown, 112 deleted, 19 hidden
- 160.58 MB
- 147: Galaxy145-[Pop\_tab\_SimpleDose\_P2.tab].null
- 137: output

**Kunnan\_AlleleOriginAndRatio.tab**

- 123: Kunnan\_Ratio.png
- 122: Kunnan\_Cov.png
- 121: vcf.conf
- 120: Origin.tab
- 119: chr09\_test.vcf
- 118: chr04\_test.vcf
- 117: chr03\_test.vcf
- 116: chr02\_test.vcf
- 115: chr01\_test.vcf
- 114: Carto.vcf
- 113: CartoRef.app

# Genetic mapping analysis with vcfHunter tool

## Calculating marker segregation distortion : output

Galaxy x +

galaxy.southgreen.fr/galaxy/ Analyze Data Workflow Shared Data Visualization Admin Help User

Using 19%

**Galaxy**

Tools

- Tassel GBS (Version 4.0)
- Rice Variant Analysis (Rice 3k, IRGIn, High Density Rice Array (HDRA, 700k SNPs))
- GENOME HARVEST
- TransPo-RG Transfer of Position to Resequenced Genome
- parental SNP - Detect parental SNP of hybrids
- Visualization
- TraceAncestor
- vcfHunter**
- VCF Filter
- vcf2allPropAndCov
- vcf2allPropAndCovByChr
- vcf2popNew
- RecombCalculatorDDose
- Draw\_dot\_plot
- KDE\_classifier
- METAGENOMICS
- FROGS
- EVOLUTION/PHYLOGENY
- Comparative Genomics
- NCBI BLAST+
- Genfam
- Protein analyses
- STATISTICS/GRAFICS
- Statistics
- Graph/Display Data
- SOUTHGREEN PROJECTS
- SNPlay3
- GNPAnnot Tools
- GNPAnnot Converters
- ESTtik
- Expression data
- SAT

1 2

1	2
chr01M2909259	1.0675463526216225
chr01M3084769	1.0587658909783821
chr01M3230825	1.123551035124867
chr01M3284968	0.4798874785587451
chr01M3285013	0.5276568002953714
chr01M3358254	0.8080719000803791
chr01M3395458	1.1328964223140199
chr01M3417999	1.0675463526216225
chr01M3493115	0.9336553965234122
chr01M3507847	0.5276568002953714
chr01M3580672	0.5726718245857745
chr01M3758744	0.9336553965234122
chr01M3775407	1.0675463526216225
chr01M3775423	0.34098131172920304
chr01M3837583	0.5276568002953714
chr01M3837669	0.43051680529937114
chr01M3847042	1.0587658909783821
chr01M3856981	0.4763798150217415
chr01M3881851	0.523767198186683
chr01M3901090	0.523767198186683
chr01M3901141	0.4763798150217415
chr01M3913984	0.8594678641097688
chr01M3926703	0.995462582229307
chr01M3958080	0.4763798150217415
chr01M4015771	0.8080719000803791
chr01M4032506	1.0587658909783821
chr01M4047381	1.1424366088102702
chr01M4047390	1.1424366088102702
chr01M4053672	1.1328964223140199
chr01M4127313	0.4798874785587451
chr01M4145915	0.5276568002953714
chr01M4150482	0.523767198186683
chr01M4163170	0.6277828140473716
chr01M4167085	1.1328964223140199
chr01M4185588	1.123551035124867
chr01M4222080	0.63257328864087
chr01M4235526	0.7454217266286189
chr01M4242788	0.995462582229307
chr01M4242794	1.0675463526216225
chr01M4269479	0.5276568002953714
chr01M4288400	0.5276568002953714
chr01M4288430	1.050163073656123
chr01M4325073	0.4798874785587451

A two column file with:  
1- marker name  
2- marker segregation distortion

History

search datasets

VCFHunter 17 shown, 112 deleted, 9 hidden 160.6 MB

148: SeqDist.tab

147: Galaxy145- [Pop tab SimpleDose P2.tab].nul

137: output

126: output

125: Kunnan\_stats.tab

124: Kunnan\_AlleleOriginAndRatio.tab

123: Kunnan\_Ratio.png

122: Kunnan\_Cov.png

121: vcf.conf

120: Origin.tab

119: chr09\_test.vcf

118: chr04\_test.vcf

117: chr03\_test.vcf

116: chr02\_test.vcf

115: chr01\_test.vcf

114: Carto.vcf

113: CartoRef.app

①

# Genetic mapping analysis with vcfHunter tool

## Calculating marker segregation distortion : output

Galaxy x +

galaxy.southgreen.fr/galaxy/ Analyze Data Workflow Shared Data Visualization Admin Help User

Using 19%

**Galaxy**

Tools

- Tassel GBS (Version 4.0)
- Rice Variant Analysis (Rice 3k, IRGIn, High Density Rice Array (HDRA, 700k SNPs))
- GENOME HARVEST
- TransPo-RG Transfer of Position to Resequenced Genome
- parental SNP - Detect parental SNP of hybrids
- Visualization
- TraceAncestor
- vcfHunter**
- VCF Filter
- vcf2allPropAndCov
- vcf2allPropAndCovByChr
- vcf2popNew
- RecombCalculatorDDose
- Draw\_dot\_plot
- KDE\_classifier
- METAGENOMICS
- FROGS
- EVOLUTION/PHYLOGENY
- Comparative Genomics
- NCBI BLAST+
- Genfam
- Protein analyses
- STATISTICS/GRAFICS
- Statistics
- Graph/Display Data
- SOUTHGREEN PROJECTS
- SNPlay3
- GNPAnnot Tools
- GNPAnnot Converters
- ESTtik
- Expression data
- SAT

1 2

1	2
chr01M2909259	1.0675463526216225
chr01M3084769	1.0587658909783821
chr01M3230825	1.123551035124867
chr01M3284968	0.4798874785587451
chr01M3285013	0.5276568002953714
chr01M3358254	0.8080719000803791
chr01M3395458	1.1328964223140199
chr01M3417999	1.0675463526216225
chr01M3493115	0.9336553965234122
chr01M3507847	0.5276568002953714
chr01M3580672	0.5726718245857745
chr01M3758744	0.9336553965234122
chr01M3775407	1.0675463526216225
chr01M3775423	0.34098131172920304
chr01M3837583	0.5276568002953714
chr01M3837669	0.43051680529937114
chr01M3847042	1.0587658909783821
chr01M3856981	0.4763798150217415
chr01M3881851	0.523767198186683
chr01M3901090	0.523767198186683
chr01M3901141	0.4763798150217415
chr01M3913984	0.8594678641097688
chr01M3926703	0.995462582229307
chr01M3958080	0.4763798150217415
chr01M4015771	0.8080719000803791
chr01M4032506	1.0587658909783821
chr01M4047381	1.1424366088102702
chr01M4047390	1.1424366088102702
chr01M4053672	1.1328964223140199
chr01M4127313	0.4798874785587451
chr01M4145915	0.5276568002953714
chr01M4150482	0.523767198186683
chr01M4163170	0.6277828140473716
chr01M4167085	1.1328964223140199
chr01M4185588	1.123551035124867
chr01M4222080	0.63257328864087
chr01M4235526	0.7454217266286189
chr01M4242788	0.995462582229307
chr01M4242794	1.0675463526216225
chr01M4269479	0.5276568002953714
chr01M4288400	0.5276568002953714
chr01M4288430	1.050163073656123
chr01M4325073	0.4798874785587451

A two column file with:  
1- marker name  
2- marker segregation distortion

148: SeqDist.tab

147: Galaxy145-  
146: Pop tab SimpleDose P2.tab.nul

137: output

126: output

125: Kunnan\_stats.tab

124: Kunnan\_AlleleOriginAndRatio.tab

123: Kunnan\_Ratio.png

122: Kunnan\_Cov.png

121: vcf.conf

120: Origin.tab

119: chr09\_test.vcf

118: chr04\_test.vcf

117: chr03\_test.vcf

116: chr02\_test.vcf

115: chr01\_test.vcf

114: Carto.vcf

113: CartoRef.app

# Genetic mapping analysis with vcfHunter tool

## Calculating pairwise marker recombination rate on P2 simple dose

The screenshot shows the Galaxy web interface with the following details:

- Tool Panel:** On the left, under the "Tools" section, the "RecombCalculatorDDose" tool is highlighted with a red circle and arrow.
- Workflow Tab:** The "Analyze Data" tab is selected.
- Input Fields:** The "marker file matrix" dropdown is set to "125: Kunman\_stats.tab" (output of vcf2popNew). The "Are marker phased?" dropdown is set to "n". The "Analysis to perform" dropdown is set to "R: Calculate recombination rate / S: Calculate segregation distortions".
- Buttons:** A "Execute" button is visible.
- Help and Support:** Includes links to the author (Guillaume MARTIN), Galaxy integration (Aurore Comte), and support email (aurore.comte@ird.fr).
- Description:** The tool is described as calculating frequencies of recombination observed between two pairs of markers, also calculating marker segregation distortion.
- Inputs:** The marker file matrix is specified as Pop\_tab\_SegregationName\_Parent.tab from vcf2popNew.
- Output:** REC.tab: A tabulated file of pairwise marker recombination rate (analysis to perform R). SegDist.tab: A tabulated file of marker segregation distortions (analysis to perform S).
- Citations:** A section listing publications by Baurens et al. (2018) related to recombination and structural variations in banana genomes.
- History:** On the right, a history panel shows a list of datasets, including Pop.tab, Pop\_report.tab, Pop\_sub.vcf, Pop\_tab\_Bridge.tab, Pop\_tab\_DoubleDose\_P1.tab, Pop\_tab\_DoubleDose\_P2.tab, Pop\_tab\_SimpleDose\_P1.tab, Pop\_tab\_SimpleDose\_P2.tab, and Pop\_tab\_unknown.tab.

# Genetic mapping analysis with vcfHunter tool

## Calculating pairwise marker recombination rate on P2 simple dose

The screenshot shows the Galaxy web interface with the following details:

- Tool:** RecombCalculatorDDose (Galaxy Version 0.1.0)
- Input:** The marker file matrix is set to "147: Galaxy145-[Pop\_tab\_SimpleDose\_P2.tab].null". A red circle labeled **1** points to this dropdown.
- Marker Phasing:** Set to "n" (unphased). A red circle labeled **2** points to this input field.
- Analysis:** Set to "R: Calculate recombination rate / S: Calculate segregation distortions". A red circle labeled **3** points to this dropdown.
- Execute:** A red circle labeled **4** points to the "Execute" button.
- Help:** Includes links to the author (Guillaume MARTIN), Galaxy integration (Aurore Comte), and support email.
- Description:** The program is designed to calculate frequencies of recombination observed between two pairs of markers. It can also calculate marker segregation distortion.
- Inputs:** The marker file matrix = Pop\_tab\_SegregationName\_Parent.tab from vcf2popNew.
- Output:** REC.tab: A tabulated file of pairwise marker recombination rate (analysis to perform R). SegDist.tab: A tabulated file of marker segregation distortions (analysis to perform S).
- Citations:** A section listing publications related to the tool, with a "Show BibTeX" link. One citation is highlighted:

Baurens, Franc-Christophe and Martin, Guillaume and Hervouet, Catherine and Salmon, Frédéric and Yohomé, David and Ricci, Sébastien and Rouard, Mathieu and Habas, Remy and Lemainque, Arnaud and Yahiaoui, Nabila and et al. (2018). Recombination and large structural variations shape interspecific edible bananas genomes. In *Molecular Biology and Evolution*, [doi:10.1093/molbev/msy199][Link]
- History:** A list of recent datasets and analyses, including:
  - 147: Galaxy145-[Pop\_tab\_SimpleDose\_P2.tab].null
  - 137: output
  - 126: output
  - 125: Kunnan\_stats.tab
  - 124: Kunnan\_AlleleOriginAndRatio.tab
  - 123: Kunnan\_Ratio.png
  - 122: Kunnan\_Cov.png
  - 121: vcf.conf
  - 120: Origin.tab
  - 119: chr09\_test.vcf
  - 118: chr04\_test.vcf
  - 117: chr03\_test.vcf
  - 116: chr02\_test.vcf
  - 115: chr01\_test.vcf
  - 114: Carto.vcf
  - 113: CartoRef.app

# Genetic mapping analysis with vcfHunter tool

## Calculating pairwise marker recombination rate: output

This dataset is large and only the first megabyte is shown below.  
Show all | Save

A square matrix of marker pairwise recombination rate

The Galaxy interface shows the following details:

- Tools:** Tassel GBS (Version 4.0), Rice Variant Analysis (Rice 3k, IRIGIN, High Density Rice Array (HDRA, 700k SNPs)), GENOME HARVEST, TransPoRG Transfer of Position to Resequenced Genome, parental SNP - Detect parental SNP of hybrids, Visualization, TraceAncestor, vcfHunter (selected), VCF Filter, vcf2allPropAndCov, vcf2allPropAndCovByChr, vcf2popNew, RecombCalculatorDDose, Draw\_dot\_plot, KDE\_classifier, METAGENOMICS, FROGS, EVOLUTION/PHYLOGENY, Comparative Genomics, NCBI BLAST+, Genfam, Protein analyses, STATISTICS/GRAFICS, Statistics, Graph/Display Data, SOUTHGREEN PROJECTS, SNIPPlay, GNPAnnot Tools, GNPAnnot Converters, ESTtik, Expression data, SAT.
- Analyze Data:** Workflow, Shared Data, Visualization, Admin, Help, User.
- Workflow:** This dataset is large and only the first megabyte is shown below. Show all | Save
- Visualization:** A square matrix of marker pairwise recombination rate
- Admin:** Using 19%
- History:** 18 shown, 113 deleted, 9 hidden. 166.13 MB. 150: REC.tab, 148: SeqDist.tab, 147: Galaxy145-. Pop\_tab\_SimpleDose\_P2.tab.null, 137: output, 126: output, 125: Kunnan\_stats.tab, 124: Kunnan\_AlleleOriginAndRatio.tab, 123: Kunnan\_Ratio.png, 122: Kunnan\_Cov.png, 121: vcf.conf, 120: Origin.tab, 119: chr09\_test.vcf, 118: chr04\_test.vcf, 117: chr03\_test.vcf, 116: chr02\_test.vcf, 115: chr01\_test.vcf, 114: Carto.vcf, 113: CartoRef.app.

# Genetic mapping analysis with vcfHunter tool

## Plotting pairwise recombination

The screenshot shows the Galaxy web interface with the 'Draw\_dot\_plot' tool selected. The interface includes a left sidebar with various bioinformatics tools, a central configuration panel for the 'Draw\_dot\_plot' tool, and a right panel showing the 'History' of previous runs.

**Tool Configuration (Left Panel):**

- Step 1:** A red circle highlights the 'Draw\_dot\_plot' link in the 'vcfHunter' section of the sidebar.
- Step 2:** A red circle highlights the 'REC.tab' dataset in the 'pairwise matrix marker file' dropdown.
- Step 3:** A red circle highlights the 'no' option in the 'I already have a locus file' dropdown.
- Step 4:** A red circle highlights the output dataset '[Pop\_tab\_SimpleDose\_P2.tab]' in the history panel.
- Step 5:** A red circle highlights the value 'n' in the 'A value specifying if the marker position should be defined based on physical position or not' dropdown.
- Step 6:** A red circle highlights the 'Execute' button.

**Description (Bottom Left):**

This program draw a dotplot based on marker pairwise recombination file obtained from RecombCalculatorDDose.

**Inputs:**

The pairwise matrix marker file = REC.tab from RecombCalculatorDDose  
Stats = SegDist.tab from RecombCalculatorDDose

**Outputs:**

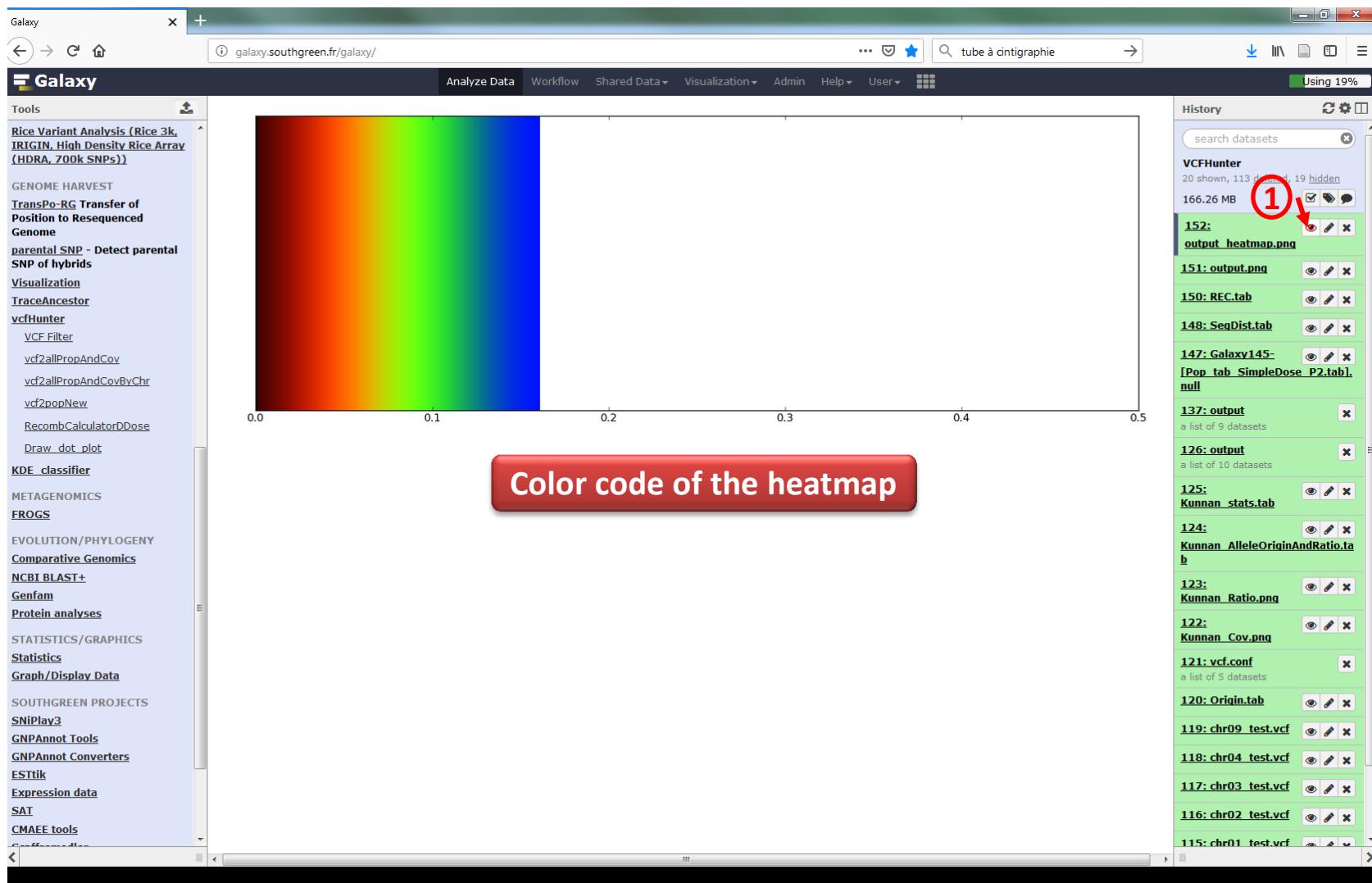
A dotplot file representing pairwise marker linkage

**History (Right Panel):**

- 150: REC.tab
- 148: SegDist.tab
- 137: output
- 126: output
- 125: Kunnan\_stats.tab
- 124: Kunnan\_AlleleOriginAndRatio.tab
- 123: Kunnan\_Ratio.png
- 122: Kunnan\_Cov.png
- 121: vcf.conf
- 120: Origin.tab
- 119: chr09\_test.vcf
- 118: chr04\_test.vcf
- 117: chr03\_test.vcf
- 116: chr02\_test.vcf
- 115: chr01\_test.vcf
- 114: Carto.vcf
- 113: CartoRef.app

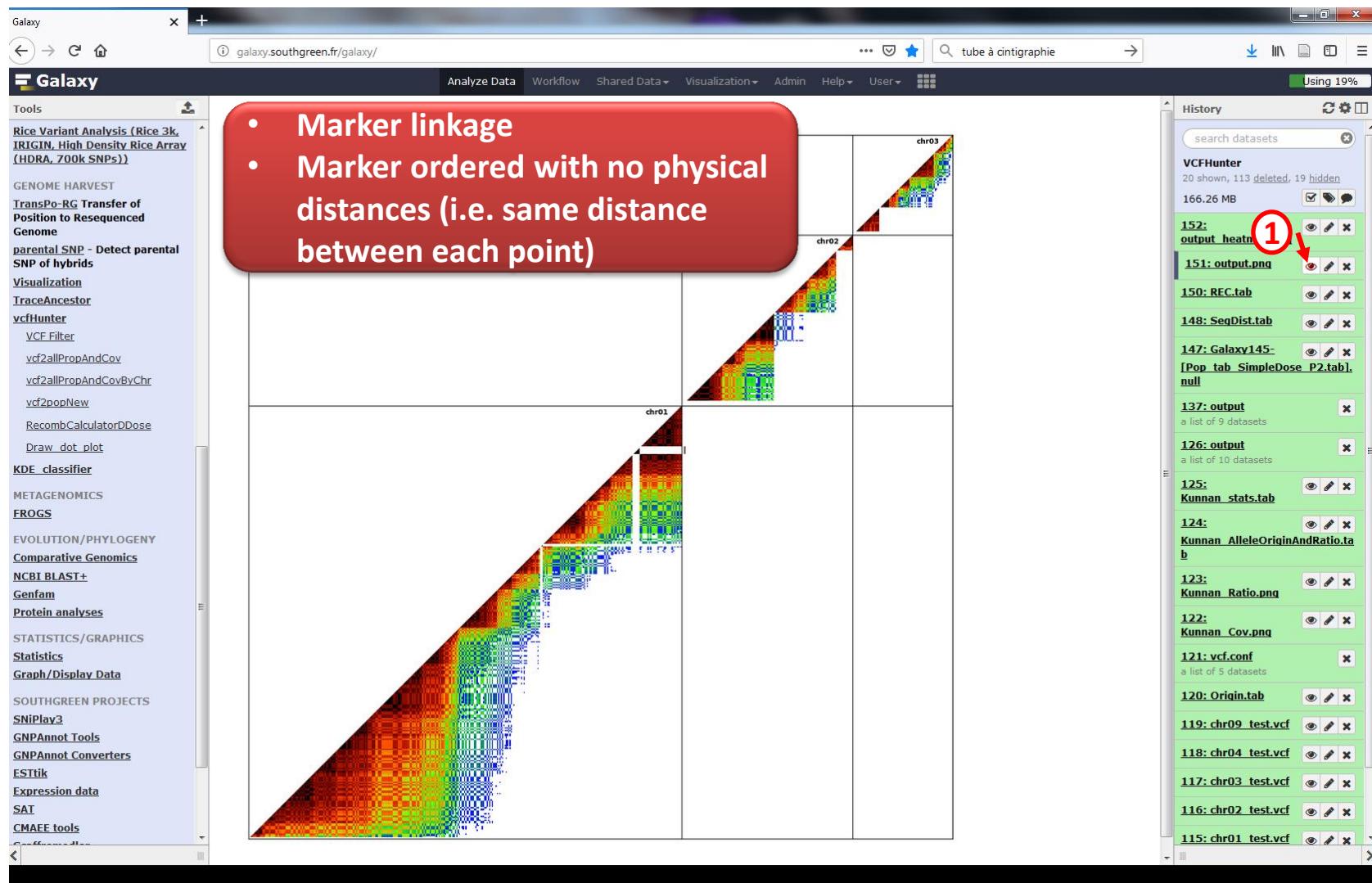
# Genetic mapping analysis with vcfHunter tool

## Plotting pairwise recombination: outputs



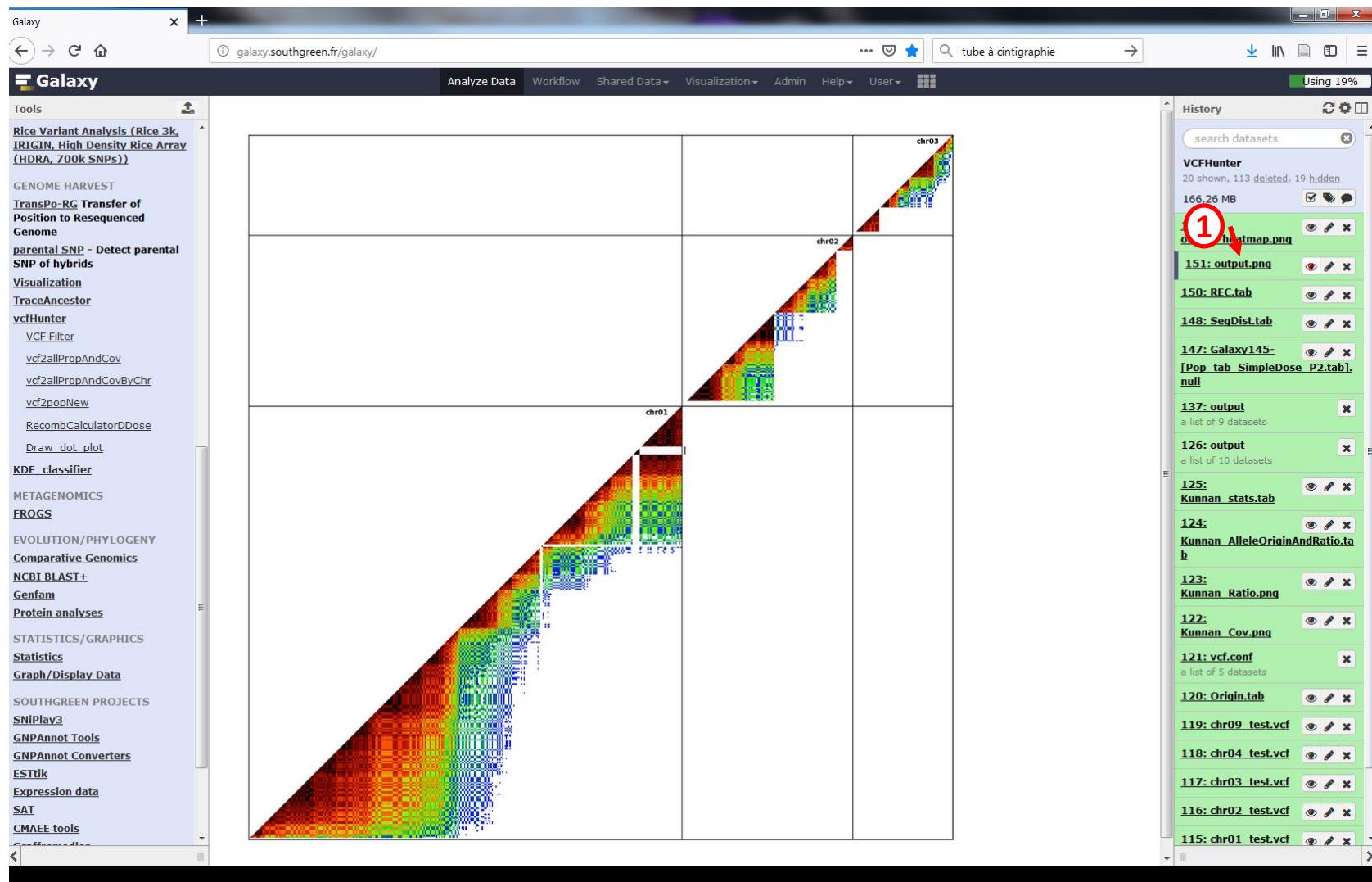
# Genetic mapping analysis with vcfHunter tool

## Plotting pairwise recombination: outputs



# Genetic mapping analysis with vcfHunter tool

## Plotting pairwise recombination with other options



# Genetic mapping analysis with vcfHunter tool

## Plotting pairwise recombination with other options

The screenshot shows the Galaxy web interface with the 'Draw\_dot\_plot' tool selected. The tool configuration includes:

- The pairwise matrix marker file:** 150: REC.tab (generated by RecombCalculatorDDose (REC.tab))
- I already have a locus file:** no (col1: marker name, col2: chromosome, col3: position)
- Pop\_tab\_SegregationName\_Parent.tab from vcf2popNew:** 147: Galaxy145-[Pop\_tab\_SimpleDose\_P2.tab].null (create your locus file)
- List of chromosomes to draw in this order:** (empty input field)
- Agr file locating scaffolds in the reference sequence:** 113: CartoRef.agp
- stats:** 148: SegDist.tab (A two column file with column 1: marker name, column2: statistics (SegDist.tab))
- A value specifying if the marker position should be defined based on physical position or not:** y

**History Panel:**

- 150: REC.tab
- 148: SegDist.tab
- 147: Galaxy145-[Pop\_tab\_SimpleDose\_P2.tab].null

**Tool Description:**

### Draw\_dot\_plot

#### Description

This program draw a dotplot based on marker pairwise recombination file obtained from RecombCalculatorDDose.

#### Inputs:

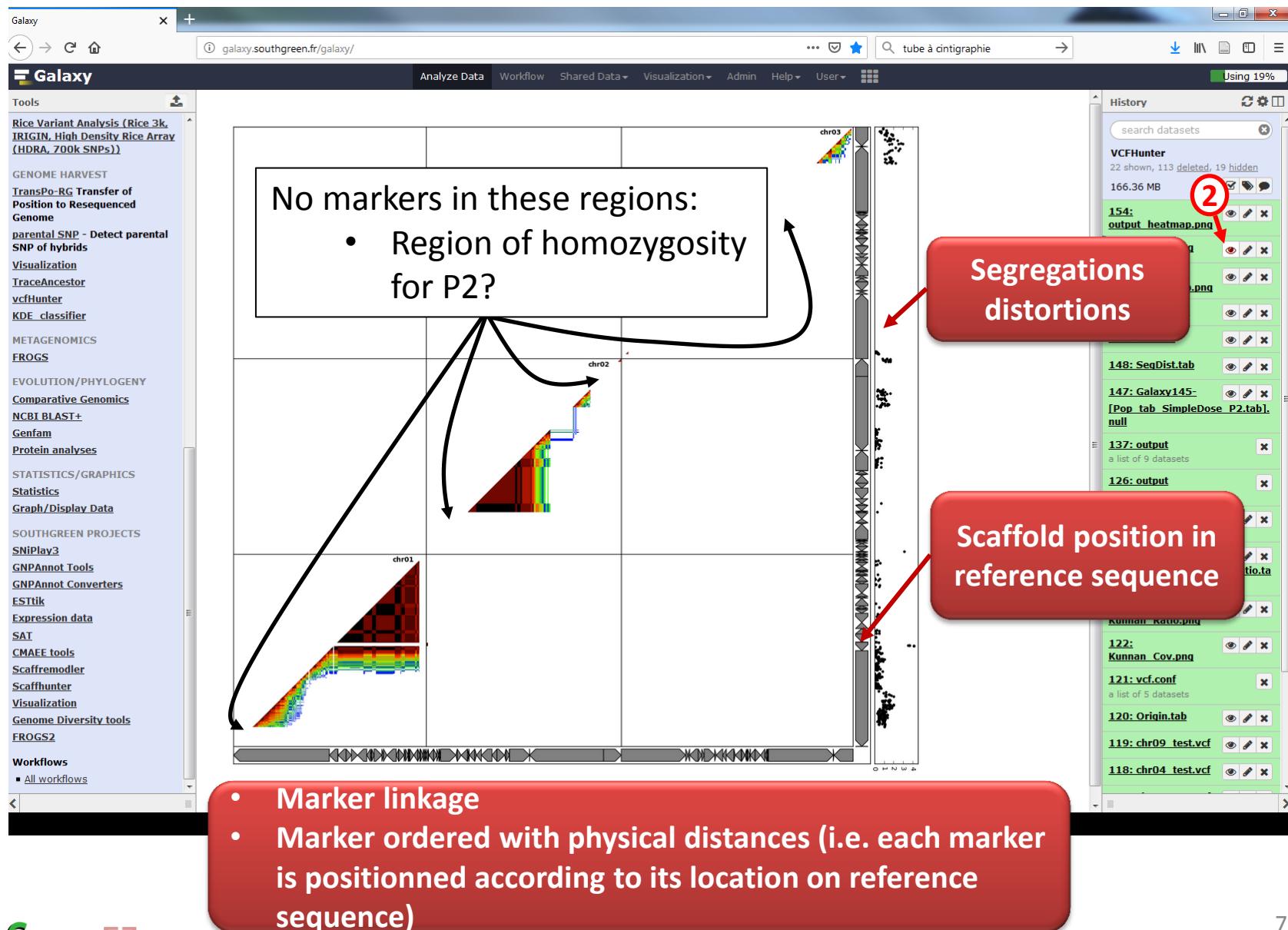
The pairwise matrix marker file = REC.tab from RecombCalculatorDDose  
Stats = SegDist.tab from RecombCalculatorDDose

#### Outputs:

A dotplot file representing pairwise marker linkage

# Genetic mapping analysis with vcfHunter tool

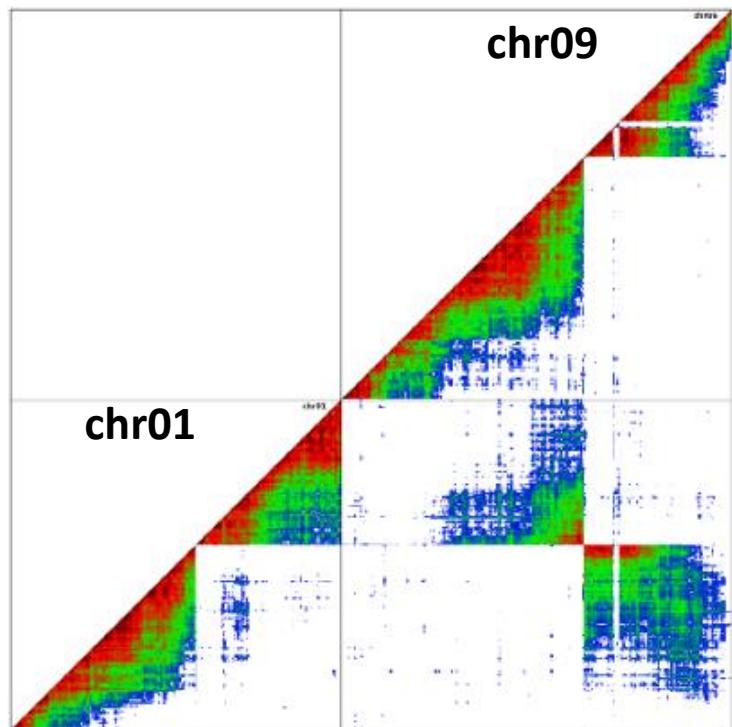
## Plotting pairwise recombination with other options



# Genetic mapping analysis with vcfHunter tool

Plotting pairwise recombination with other populations

Evidence of reciprocal translocations  
between chromosome 1 and 9



Evidence for large segregation  
distortions and inter-chromosomal  
linkage

