

 Open access • Book Chapter • DOI:10.1007/978-1-4939-6658-5\_5

## Mining Plant Genomic and Genetic Data Using the GnpIS Information System.

— [Source link](#) 

Anne-Françoise Adam-Blondon, Michael Alaux, Sophie Durand, Thomas Letellier ...+17 more authors

**Institutions:** Université Paris-Saclay, Agro ParisTech

**Published on:** 01 Jan 2017 - Methods of Molecular Biology (Humana Press, New York, NY)

**Topics:** Genetic architecture

Related papers:

- [New Technologies InterStoreDB: A Generic Integration Resource for Genetic and Genomic Data](#)
- [SNPTrackTM : an integrated bioinformatics system for genetic association studies](#)
- [BASC: an integrated bioinformatics system for Brassica research](#)
- [AraPheno and the AraGWAS Catalog 2020: a major database update including RNA-Seq and knockout mutation data for Arabidopsis thaliana.](#)
- [Analysis and visualization of Arabidopsis thaliana GWAS using web 2.0 technologies.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/mining-plant-genomic-and-genetic-data-using-the-gnpis-1kp54ofuxc>



**HAL**  
open science

## Mining plant genomic and genetic data using the GnpIS information system

Anne-Francoise Adam-Blondon, Michael Alaux, Sophie Durand, Thomas Letellier, Guillaume Merceron, Nacer Mohellibi, Cyril Pommier, Delphine Steinbach, Françoise Alfama-Depauw, Joëlle Amselem, et al.

### ► To cite this version:

Anne-Francoise Adam-Blondon, Michael Alaux, Sophie Durand, Thomas Letellier, Guillaume Merceron, et al.. Mining plant genomic and genetic data using the GnpIS information system. Aalt D.J van Dijk. Plant Genomics Databases. Methods and Protocols, 1533, Springer, 2017, Methods in Molecular Biology, 978-1-4939-6658-5. 10.1007/978-1-4939-6658-5\_5 . hal-03206731

**HAL Id: hal-03206731**

**<https://hal.archives-ouvertes.fr/hal-03206731>**

Submitted on 23 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Mining plant genomic and genetic data using the GnpIS information system

**Running Head :** Mining GnpIS

## **Authors**

A-F Adam-Blondon<sup>1</sup>, M. Alaux<sup>1\*</sup>, S. Durand<sup>1\*</sup>, T. Letellier<sup>1\*</sup>, G. Merceron<sup>1\*</sup>, N. Mohellibi<sup>1\*</sup>, C. Pommier<sup>1\*</sup>, D. Steinbach<sup>1,2\*</sup>, F. Alfama<sup>1</sup>, J. Amselem<sup>1</sup>, D. Charruaud<sup>1,3</sup>, N. Choisne<sup>1</sup>, R. Flores<sup>1</sup>, C. Guerche<sup>1</sup>, V. Jamilloux<sup>1</sup>, E. Kimmel<sup>1</sup>, N. Lapalu<sup>1,4</sup>, M. Loaec<sup>1</sup>, C. Michotey<sup>1</sup>, H. Quesneville<sup>1</sup>

\* The authors have contributed equally to the work

<sup>1</sup> Research Unit in Genomics-Info UR1164, INRA, Université Paris-Saclay, Versailles, Route de Saint-Cyr, 78026 Versailles Cedex, France.

<sup>2</sup> Research Unit GQE-Le Moulon UMR 320, INRA, Université Paris-Sud, Université Paris-Saclay, CNRS, AgroParisTech, Ferme du Moulon, 91190 Gif-sur-Yvette, France

<sup>3</sup> ADRINORD Espace Recherche Innovation, 2 rue des Canoniers, 59800 Lille, France

<sup>4</sup> UMR BIOGER, UMR1290, INRA, AgroParisTech, 1 avenue Lucien Bretignières, 78850 Thiverval-Grignon, France

## **Corresponding author : A-F Adam-Blondon**

Research Unit in Genomics-Info UR1164, INRA, Université Paris-Saclay, Versailles, Route de Saint-Cyr, 78026 Versailles Cedex, France.

Email : [afadam@versailles.inra.fr](mailto:afadam@versailles.inra.fr)

## **Summary**

GnpIS is an information system designed to help scientists working on plants and fungi to decipher the molecular and genetic architecture of trait variations by facilitating the navigation through genetic, genomic and phenotypic information. The purpose of the present chapter is to illustrate how users can (i) explore data sets from phenotyping experiments in order to build new datasets for studying genotype x environment interactions in traits, (ii) browse into the results of other genetic analysis data such as GWAS to generate or check working hypothesis about candidate genes, or to identify important alleles and germplasms for breeding programs and (iii) explore the polymorphism in specific area of the genome using InterMine, JBrowse tools embedded in the GnpIS information system.

## **Key Words**

GnpIS, GWAS, polymorphism, phenotypes, markers, genetic resources, GMOD genome browser, InterMine

## **1. Introduction**

In the last decades, many research programs in biology have been profoundly affected by a major technological evolution: the very fast development of high-throughput technologies that produce massive amounts of data (e.g. new generation sequencing methods, high-density genotyping arrays, improvements in proteomics and metabolomics data acquisition, image treatments for phenotype observations). One of the consequences of the emergence of these technologies is the critical need for managing high volumes of data in efficient computer systems. In addition, strong initiatives from international policy makers stressed the commitment for academics research to provide open and interoperable data resources, which should enhance the reuse of datasets, hence reducing the cost of research projects (see for instance [1,2]).

GnplS is an original information system built to act as a central repository for plant science data. It implements an efficient and reliable information system over a computer infrastructure able to meet the challenge of huge data management. It was designed to integrate heterogeneous data types, allowing to bridge genomics, genetics and phenomics data [3]. As a reference data warehouse, it promotes data exchange and dissemination to the international scientific community. For this purpose, GnplS uses internationally recognized formats, controlled vocabularies or ontologies, or other standards for data exchange [4].

The main focus of GnplS is to help plant and fungi scientists to decipher the molecular and genetic architecture of trait variations by facilitating the navigation through genetic, genomic and phenotypic information. The purpose of the present chapter is to illustrate how users can (i) explore data sets from phenotyping experiments used for genetic analysis to build datasets for studying genotype x environment interactions in traits, (ii) browse into the data results of genetic analysis such as genome-wide association studies (GWAS) to generate new working hypothesis about candidate genes, or to identify important alleles and germplasms for breeding programs, and (iii) explore the polymorphism in specific area of the genome.

## **2. Materials**

GnpIS relies on cutting edge data warehouse technologies using (i) a relational database to integrate and link data in a coherent and consistent framework, and (ii) data marts and (iii) NoSQL technologies to speed up the searches. It can be accessed through web graphical interfaces at the URL: <https://urgi.versailles.inra.fr/gnpis>. It guides the users to dedicated interfaces: <https://urgi.versailles.inra.fr/Data/Genome/Genome-data-access> for accessing genomes, <https://urgi.versailles.inra.fr/GnpMap> for genetic maps, <https://urgi.versailles.inra.fr/GnpSNP> for polymorphisms, <https://urgi.versailles.inra.fr/epheisis/> for phenotyping experiments, <https://urgi.versailles.inra.fr/association/> for phenotypes to genotypes associations, and <https://urgi.versailles.inra.fr/siregal/> for the genetic resources.

It gives access to data about genomes via the GBrowse and JBrowse, two GMOD genome browsers [5]. Genetic resources data follows the MultiCrop Passport Data format (MCPD, [6]). Genetic maps, QTLs (Quantitative Trait Loci), phenotypes to genotypes associations, polymorphism data such as SSR (Simple Sequence Repeats) and SNP (Single-Nucleotide Polymorphism) use relevant ontologies or controlled vocabularies when they exist.

Moreover, the GnpIS portal allows intuitive google-like queries using a free text field, or complex queries on dedicated data subsets of the information system (grapevine data and wheat chromosome 3B data) using dedicated InterMines ([7]; <http://urgi.versailles.inra.fr/GrapeMine/begin.do>; <http://urgi.versailles.inra.fr/Wheat3BMine/begin.do>).

The datasets used to illustrate the present chapter were produced by different groups of researchers working on traits evaluation in wheat [8], on association genetics in tomato [9], on grapevine genomics [10, 11]. Many others are available in GnpIS, derived from projects in which the French Institute for Agronomical Research (INRA) is partner.

### 3. Methods

#### 3.1. Building a dataset for the meta-analysis of phenotyping experiments

Studying genetic determinism of traits requires the statistically sound phenotypic evaluation of genotypes panels. It is also true for the meta-analysis of experimentations obtained in multiple environments, that may allow a better understanding of the influence of fluctuating environments (year/climate/location, management techniques, etc.). The building of such datasets for meta-analysis relies on a data integration process. This often implies to unambiguously identify in heterogeneous data sources the pivot scientific objects, such as plant materials or traits. Data integration is a long process that can be facilitated by integrative information systems such as GnpIS (Note 1). It also relies on the use of ontologies and controlled vocabularies for traits, and precise identifiers for germplasms, to ensure the consistency of the newly built datasets.

GnpIS gives access to multi-local and multi-annual phenotyping trials for several species and to the ontologies used. To select relevant data, three sets of filters are available: “Genotype” for selecting germplasm/plant material, “Variables” for trait and environment variables selection, and “Trial” for trials descriptors and metadata. The protocol below describes the way to build a dataset from a panel of evaluated genotypes, for a set of traits in a maximum number of experimental conditions, to better isolate the genetic component of the traits variation.

1. Go to <https://urgi.versailles.inra.fr/gnpis> and click on “Phenotypes” and on the left menu on “experimental data” or go directly to <https://urgi.versailles.inra.fr/ephegis/ephegis/viewer.do>
2. Plant material selection: the “Genotype” tab from the form is selected by default. In the example shown in fig. 1, the focus is on bread wheat by selecting the genus “*Triticum* L.” and the species “*Triticum aestivum aestivum*”.
3. Variables selection: clicking on the “Variable” tab shows all the variables consistent with the current search parameters, i.e. “*Triticum aestivum aestivum*”. To focus on yield as a trait of bread wheat (for instance), it is possible to find in the Small grain cereal

Network Ontology the three following traits: “yield”, “thousand grain weight” and “test weight”. Clicking on these three variables and then on the “Results” button allows retrieving results from 635 trials (at the date this chapter was written, as in the rest of this document).

4. Refine the selection: go back to the selection form using the “Back to form” button and select the “Trial” tab. It is possible, for instance, to refine by data set: selecting “INRA Wheat Network BRC accession (A series)” in the “Add trial by data set” box allows retrieving 230 trials after clicking on the “Results” button (fig. 2). It is also possible to filter by years and/or locations, which is a way to focus on some macro environment variables. Leaving empty a filter box means not filtering on the corresponding criteria.
5. Data export: the result page of GnpIS gives a summary of the filters applied and a preview of the results of the query. Two tables are provided: the “Trial table” with the average results from several repetitions and the “Rep table” that gives access to the raw data for each repetition. It is possible to download all the results either in a simple tabular format or in an ISA Tab exchange data format (fig. 2). The ISA Tab exchange data format ensures the traceability of the metadata associated to the data [4]. The downloaded data includes the year, the genotype (LotNumber), information on the position of the lot in the experiment (X and Y) and information on the treatments used in the experiment (here in the “itk” column). It allows seeking statistical correlation between those different factors (fig. 2). It is possible for users having a private access (Note 2) to the database to get climatic data associated to the experiments selected (Note 3).

### **3.2. GWAS data exploration**

With the increased number of plant genome sequencing projects and the development of high-throughput genotyping methods, it is now possible for a growing number of crops to genotype enough SNP to conduct association mapping approaches, either at the whole-genome scale



(GWAS) or in regions in which QTL have been detected by genetic linkage approaches. Thanks to the combination of approaches based on association genetics and QTL mapping, it is now possible i) to fine map QTL ii) to identify new markers useful for selection, iii) to find interesting alleles in genetic resources collections and to use them in new material dedicated to selection. Nowadays, GWAS is becoming a mandatory approach in the strategies developed by scientists and breeders to understand the genetic architecture of traits, and the relationships between genotype and phenotype. GnpIS was recently extended to store GWAS experiments results including data and metadata.

GnpIS provides as metadata (i) the plant material used in the GWAS experiment with unambiguous identification of accessions, (ii) the name of the panel in which are gathered these genotypes and some information on its genetic structure: kinship between individuals and linkage disequilibrium (LD) between polymorphisms, and (iii) the statistical model used to detect associations between markers and phenotypes. The data provided consist in (i) the genotypic and the phenotypic values used as entry in the model and (ii) the  $p$ -values of the association between each marker and the studied traits. The interface allows exploring GWAS data to discover the best markers and alleles associated to traits of interest with Q-Q Plots (quantile-quantile plots) and Manhattan Plots ( $-\log_{10}(P)$  genome-wide association plots). In turn, such marker identification allows finding candidate genes underlying the phenotypic variation through their integration with other data types present in the GnpIS data warehouse, such as genomic annotation data, expression data, genetic mapping data or other QTL data.

1. Go to <https://urgi.versailles.inra.fr/gnpis> and click on “Association” or go directly to <https://urgi.versailles.inra.fr/association>.
2. In the left menu, in “Queries”, select “Associations” and then select the genus of the species of interest (e.g. *Solanum* L.)
3. It is then possible to filter the data by traits, markers or panel. For instance, selecting Ascorbate content and Fructose and filtering on the “solcap” panel with the Panel tab, retrieved at the date the chapter was written 5524 unique markers with a result of association to one or both traits. All the results are displayed in two tables and

downloadable into a csv format (fig. 3). It is possible to visualize a boxplot of the phenotypic value of each allele at the marker (fig. 3). Note that at the bottom of the boxplot, a table sorts the accessions of the panel according to their genotypes at the marker and displays of each of them, their phenotypic value (fig 3). The name of the accessions is clickable to display more information.

4. The detail of the phenotyping and genotyping experiments used as inputs can be displayed by clicking on the panel name in the table. It displays the panel card and all the experiments linked to it. The phenotyping and genotyping data are displayed (Note 4) and downloadable in GnpIS (see paragraph 3.1).
5. Several tools allows going deeper into the results:
  - a. Graphical and dynamic display of the associations with a Q-QPlot or a Manhattan Plot (fig. 3).
  - b. New filters (by value, by chromosome, by trait...) facilitate selection and display.
  - c. Links to markers positions on the reference genome of the selected genus (fig. 4). A direct link to the browser of the reference genome (when available) is clickable in the "Association results" table, column "Genome Browser" (fig. 4). In the Manhattan plot, a pop-up window giving basic information on a marker (name, chromosome, position,  $Pval$ ,  $-\log Pval$ ), a link to the marker card and again to the browser can be obtained by pointing the marker (zoom until it works; fig. 4). Depending on genomes, the browser gives access to gene annotations, expression data, epigenetic marks, polymorphisms, etc. The marker card gives access to other types of information linked to the marker: genetic and physical maps, associated QTLs, sequences.

### **3.3. Exploring sequence diversity in a genomic region**

GnpIS allows bridging regions identified in QTL analysis, or association mapping, with genomic annotations and other associated genetic information. For instance, the GSVIVG01013466001

grapevine gene (other ID: Vv18s0122g00190) is located in a genome region under study in relation with the Fleshless berry locus [9].

1. One way to retrieve information on the region of interest is to use the GrapeMine (<http://urgi.versailles.inra.fr/GrapeMine/>) by clicking on InterMine, then on Grape and typing Vv18s0122g00190 in the search box. The tool gives access to a page gathering all the information on the gene (structure, sequence, annotation etc., fig. 5) and allows discovering other data stored in GnpIS via the links available in the right panel (i.e. the Vitis Gbrowse, datamart or else using the quick search, see below). A specific tutorial on GrapeMine (and on Wheat3BMmine) is available via a “documentation” link at the top right side of the window. One of the main interests of the InterMine tool is that it allows retrieving information linked to a genome interval or to lists of features and not only to a single feature as shown in figure 5.
2. The second way to find useful data is to go directly to <https://urgi.versailles.inra.fr/gnpis> and enter the gene ID (GSVIVG01013466001) in the text-based quick search box. Clicking on an exact match sends on a zoom of the GSVIVG01013466001 gene in the Vitis 12X Genome browser (GBrowse) (fig. 6).
3. Go to the menu “Select Tracks” at the top of the search GBrowse box (Note 5). A list of tracks is provided that can be added or hidden from the GBrowse view. Add the “Genoscope annotation” and the “V1\_NR” annotation tracks. Click “All on” in the polymorphism check box to show all the tracks of this category or select the three tracks from the “berry size project” (example in fig. 6). Click on the “Back to the browser” link to get the genomic view updated with the selected tracks.
4. The gene transcripts predicted by two different automatic annotations corresponding to the same gene (the gene ID is GSVIVG01013466001 in the Genoscope annotation and Vv18s0122g00190 in the CRIBI V1 annotation) are shown together with all the SNPs and insertions/deletions that were detected in the different parts of the gene.
5. It is possible to select any of the SNP markers in the area and to click on it to open a pop-up that gives a minimum information on the SNP (SNP ID, the type of

polymorphism, the position of the genome) and additional links like for instance towards Ensembl Plants. In the example shown in fig. 6, the SNP chosen is Vv\_1274223. Clicking on “Link to SNP card in GnpIS” in the pop-up allows to open the “Polymorphic locus card” of the GnpIS database in which additional information can be found ([https://urgi.versailles.inra.fr/GnpSNP/snp/card/snp.do?name=Vv\\_1274223](https://urgi.versailles.inra.fr/GnpSNP/snp/card/snp.do?name=Vv_1274223); fig. 7)

- a. The flanking sequences of this SNP that can be downloaded in a FASTA format. These sequences are based on the reference sequence that was used for SNP calling, which is also downloadable.
  - b. All the genotypes observed at the same genome position for a re-sequencing experiment are listed (“Linked with variations” block) and can be viewed in more details (click on “view list”) in a clickable table. The genotype table (<https://urgi.versailles.inra.fr/GnpSNP/snp/genotypeAction.do?action=BATCH&batchId=1608>) can be obtained by clicking on “view list of variations”, then in the table appearing, on the name of the experiment (in the example shown in fig. 7, “VVC2972A-batch1”) and finally on “view the genotype table” (fig.8). This matrix of the genotypes for all the markers of the region and all the plant material involved in this resequencing analysis can be exported in a csv format (export button at the bottom of the page).
  - c. All the corresponding sequenced accessions (“Linked with lines” block in fig. 7) and their taxon name are listed in a table (click on “view list”). Clicking on the accession name gives access to all the experiments of SNP detection for the given accession.
6. To know more about the accessions, go back to the text-based quick search box in GnpIS portal (<https://urgi.versailles.inra.fr/gnpis>) and enter the accession ID (e.g. 324Mtp43, or “Charger”). The system returns a table with several options “phenotypes, genetic resources, polymorphisms...”, depending on the data linked to the term found in the database. Genetic resources links to all the information associated to the accession: accession name, taxon, holding institute, origin, evaluation data, etc. (see

fig. 9). Clicking on the basket at the bottom right allows ordering the accession to a Biological Resource Center if the accession is maintained at INRA.

7. Another way is to click on the “Genetic Resources” module of GnpIS and on “Taxons” in the left menu. Typing the cultivar name “Cabernet franc” in the “Scientific name” box returns all the accessions available for this cultivar together with the linked data (fig 10).

#### **4. Notes**

<sup>1</sup>GnpIS is under continuous improvement: problems can be indicated by an Email to [urgi\\_contact@versailles.inra.fr](mailto:urgi_contact@versailles.inra.fr)

<sup>2</sup>The management and access to private data is made possible in the frame of project in collaboration with INRA.

<sup>3</sup>Access to climatic data must be asked only for a short list of experiments as the query retrieves a large number of data.

<sup>4</sup>Some link may display an error message. This is often due to the fact that the linked data are private.

<sup>5</sup>Some unwanted old configurations may be stored in your cache for some applications or pages of GnpIS interface. You may have to clean the cache of your web explorer to solve the problem or type Ctrl+F5.

#### **5. Acknowledgements**

We gratefully thank Aminah-Olivia Keliet, Btissam Aissaoui, Laura Burlot, Loic Couderc, Guillaume Cornut, Mathieu Labernardière, Mathilde Lainé, Aristide Lebreton, Florian Philippe, Sandrine Nsigue-Meilo and Daphnée Verdelet for help in some developments, data insertion, trainings on GnpIS in the last four years. We also warmly thank our past and present projects partners that are providing the data inserted in the database and who are at the origin of many improvements. For the most important recent improvements, we specially thank Stephane Nicolas, Mathilde Causse, Christopher Sauvage, Jacques Le Gouis, Alain

Charcosset, Patrice This, Gilles Charmet, François-Xavier Oury, Arnaud Gauffreteau, Etienne Paux and Frédéric Choulet. GnpIS has been developed in the last 4 years with the support of INRA, the ANR projects of the “Investment for the Future” call Phenome (<https://www.phenome-fppn.fr/>), Aker (<http://www.aker-betterave.fr/en/>), Amaizing (ANR-10-BTBR-03), Breedwheat (ANR-10-BTBR-02) and Peamust (ANR-11-BTBR-02), the ANR project GnpAsso (ANR-10-GENM-0006), the TransPLANT FP7 European infrastructure project (project n°283496).

## 6. References

- 1- [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)
- 2- Kaiser J (2015) U.S. agencies fall in line on public access. *Science* 349 :167
- 3- Steinbach, D; Alaux, M; Amselem, J; Choisine, N; Durand, S; Flores, R; Keliet, AO; Kimmel, E; Lapalu, N; Luyten, I; Michotey, C; Mohellibi, N; Pommier, C; Reboux, S; Valdenaire, D; Verdelet, D; Quesneville, H (2013) GnpIS: an information system to integrate genetic and genomic data from plants and fungi. *Database* (Oxford), Bat058. doi: 10.1093/database/bat058
- 4- Krajewski P, Chen D, Ćwiek H, van Dijk ADJ, Fiorani F, Kersey P, Klukas C, Lange M, Markiewicz A, Nap JP, van Oeveren J, Pommier C, Scholz U, van Schriek M, Usadel B, Weise S (2015) Towards recommendations for metadata and data handling in plant phenotypic experiments. *J Exp Bot*, 66(18):5417-27. doi:10.1093/jxb/erv271
- 5- Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH (2009) JBrowse: a next-generation genome browser. *Genome Res*, 19(9):1630-8. doi: 10.1101/gr.094607.109.
- 6- <http://www.biodiversityinternational.org/e-library/publications/detail/faobiodiversity-multi-crop-passport-descriptors-v2-mcpd-v2/>
- 7- Smith RN, Aleksic J, Butano D, Carr A, Contrino S, Hu F, Lyne M, Lyne R, Kalderimis A, Rutherford K, Stepan R, Sullivan J, Wakeling M, Watkins X, Micklem G (2012) InterMine:

a flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics*, 28(23):3163-5. doi: 10.1093/bioinformatics/bts577.

- 8- Oury FX, Godin C, Mailliard A, Chassin A, Gardet O, Giraud A, Heumez E, Morlais JY, Rolland B, Rousset M, Trottet M, Charmet, G (2012) A study of genetic progress due to selection reveals a negative effect of climate change on bread wheat yield in France. *Europ J Agronomy*, 40 , 28-38 . doi: 10.1016/j.eja.2012.02.007
- 9- Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, Fernie AR, Causse M (2014) Genome-Wide Association in Tomato Reveals 44 Candidate Loci for Fruit Metabolic Traits. *Plant Physiol*, 165, 1120-1132. DOI: 10.1104/pp.114.241521
- 10- Houel C, Bounon R, Chaïb J, Guichard C, Péros JP, Bacilieri R, Dereeper A, Canaguier A, Lacombe T, N'Diaye A, Le Paslier MC, Vernerey MS, Coriton O, Brunel D, This P, Torregrosa L, Adam-Blondon AF (2010) Patterns of sequence polymorphism in the fleshless berry locus in cultivated and wild *Vitis vinifera* accessions. *BMC Plant Biol*, 10:284. doi:10.1186/1471-2229-10-284
- 11- Adam-Blondon A-F, Jaillon O, Vezzulli S, Zharkikh A, Troggio M, Velasco R (2011) Genome Sequence Initiatives. in: A-F Adam-Blondon, JM Martinez-Zapater, Chittaranjan Kole (eds) *Genetics, Genomics and Breeding of Grapes*. Science Publishers and CRC Press. pp 211-234. ISBN 9781578087174

## Figures Captions

### Phenotypes

Genotype Variable Trial Reset Form Results

Reset Tab

Add accession by Genus

====> Choose a species

Add accession by Panel

Add accession by Collection


Add accession by Name

Paste your Accession Name list

#### Accession

1-10 of 1,729

Accession Number	Accession Name	Taxon Name
Alberic	Alberic	Triticum aestivum aestivum
29843	ALTIGO	Triticum aestivum aestivum
37771	AO00001	Triticum aestivum aestivum
AO01001	AO01001	Triticum aestivum aestivum
AO03001	AO03001	Triticum aestivum aestivum
AO03002	AO03002	Triticum aestivum aestivum
AO03003	AO03003	Triticum aestivum aestivum
AO03004	AO03004	Triticum aestivum aestivum
AO04001	AO04001	Triticum aestivum aestivum
AO05001	AO05001	Triticum aestivum aestivum

 Plant Material Export

**figure 1.** Query page of the GnpIS module for phenotypes (<https://urji.versailles.inra.fr/ephep/ephep/viewer.do>). A filter on experiments involving bread wheat genotypes has been applied. It is possible to export the list of the selected accessions (plant material) and their identifiers in a tabular format.




## Phenotypes

[Back to Form](#)

Search parameter(s):

Genus selected:   
Number of Taxon specie(s) selected: 1  
Number of Variable(s) selected: 3

 [Get ClimatK Data](#)



**DATA SUMMARY**  
Trials: 230  
Trial : [BTH\\_Champagne-céréales\\_2004\\_SetA1](#)  
Site : Champagne-céréales  
Data Available  
-----  
Trial : [BTH\\_Champagne-céréales\\_2005\\_SetA1](#)  
Site : Champagne-céréales  
Data Available  
-----  
Trial : [BTH\\_Champagne-céréales\\_2006\\_SetA1](#)

Available Phenotyping Campaign(s)

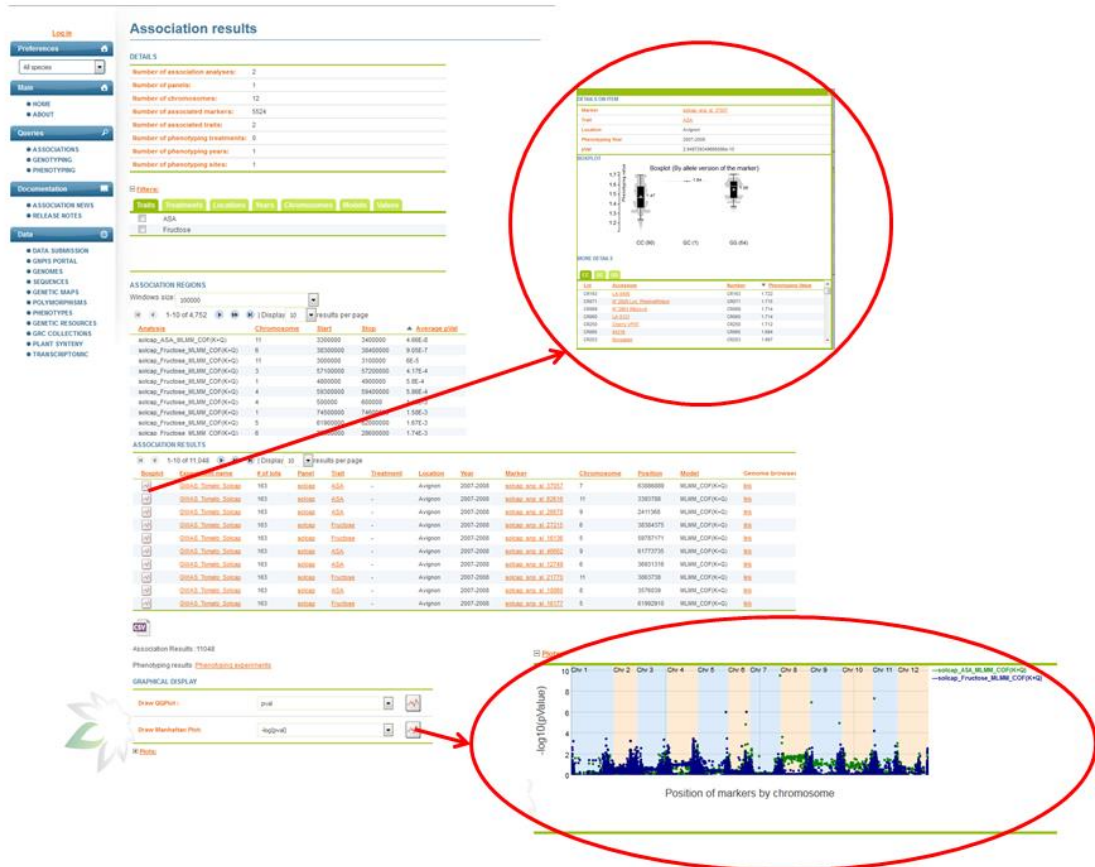
Level: Trial

1-10 of 1,190 | Display 10 results per page

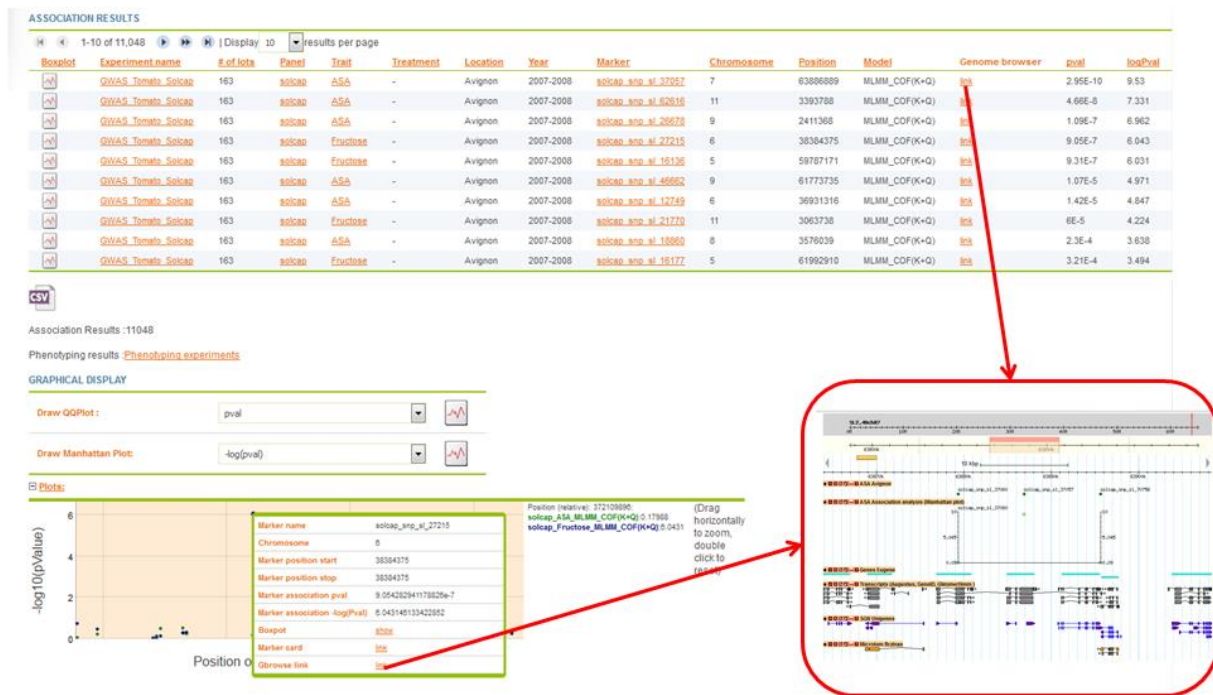
lotNumber	itk	trialName	trialSite	year	X	Y	yield(rdt)	thousand grain we
Barok	t treated	BTH_Champagne-céréales_2007_SetA1	Champagne-céréales	2007			103,6	
Hendrix	t treated	BTH_Champagne-céréales_2007_SetA1	Champagne-céréales	2007			83,8	
CF04076	t treated	BTH_Champagne-céréales_2007_SetA1	Champagne-céréales	2007			87,4	
Toisonдор	t treated	BTH_Champagne-céréales_2007_SetA1	Champagne-céréales	2007			93,5	
AO06313	t treated	BTH_Champagne-céréales_2007_SetA1	Champagne-céréales	2007			91,2	
AO05004	t treated	BTH_Champagne-céréales_2007_SetA1	Champagne-céréales	2007			91,6	
AO05001	t treated	BTH_Champagne-céréales_2007_SetA1	Champagne-céréales	2007			93,1	
DI05014	t treated	BTH_Champagne-céréales_2007_SetA1	Champagne-céréales	2007			87,6	
Koreli	t treated	BTH_Champagne-céréales_2007_SetA1	Champagne-céréales	2007			90,3	
DI04018	t treated	BTH_Champagne-céréales_2007_SetA1	Champagne-céréales	2007			92	

 [Ephesis data export](#)  [Ephesis IsaTab export](#)

**figure 2.** Result table of the GnpIS module for phenotypes showing a summary of the Search parameters and a preview of the selected trials.



**figure 3.** Display of the results of association genetic analysis in GnpIS: summary of the experiment, a set of filters allowing refining the display, and the table of the *pValues* for each marker x trait combination. It is also possible to display the boxplot of the trait values observed for each allelic combination, a Q-Q Plot of the *pValues* or a Manhattan plot of the  $-\log(pValues)$  of the markers x traits associations.



**figure 4.** GnpIS allows easy navigation between the association results presented in the table or plots and the annotation of the corresponding genomic regions displayed in genome browsers.

**GrapeMine** v1.0 An integrated database for grapevine data

Home Templates Lists QueryBuilder Regions Data Sources API Help

Contact Us | Log In

Search:  GO

**Gene : Vv18s0122g00190 *V. vinifera***

DB identifier: Vv18s0122g00190

---

**Genome Feature**

Region: gene Length: 4383 (kb) [link]

Location: chr18:228349-232731 reverse strand

---

**Overlapping Features**

Genome features that overlap coordinates of this Gene

Exons: 14, Genes: 1, Transcripts: 2, CDSs: 2, Polymorphisms: 2, SNPs: 18

Show all in a table >

---

**1 Organism**

Name: Vitis vinifera Taxon id: 29790

---

**1 Data Sets**

Name: Vitis vinifera L. URL: [link]

---

**7 Exons**

DB identifier	Length	Chromosome Location	Organism Name
Vv18s0122g00190.e1	497 (kb) [link]	chr18: 232325-232731	Vitis vinifera
Vv18s0122g00190.e2	43 (kb) [link]	chr18: 231841-231991	Vitis vinifera
Vv18s0122g00190.e3	95 (kb) [link]	chr18: 231651-231745	Vitis vinifera
Vv18s0122g00190.e4	24 (kb) [link]	chr18: 231531-231564	Vitis vinifera
Vv18s0122g00190.e5	93 (kb) [link]	chr18: 229914-229968	Vitis vinifera
Vv18s0122g00190.e6	43 (kb) [link]	chr18: 229769-229810	Vitis vinifera
Vv18s0122g00190.e7	274 (kb) [link]	chr18: 228349-228622	Vitis vinifera




---

**1 Transcripts**


DB identifier	Length	Chromosome Location
Vv18s0122g00190.t1	4383 (kb) [link]	chr18: 228349-232731

---

Questions? Comments? Click here!

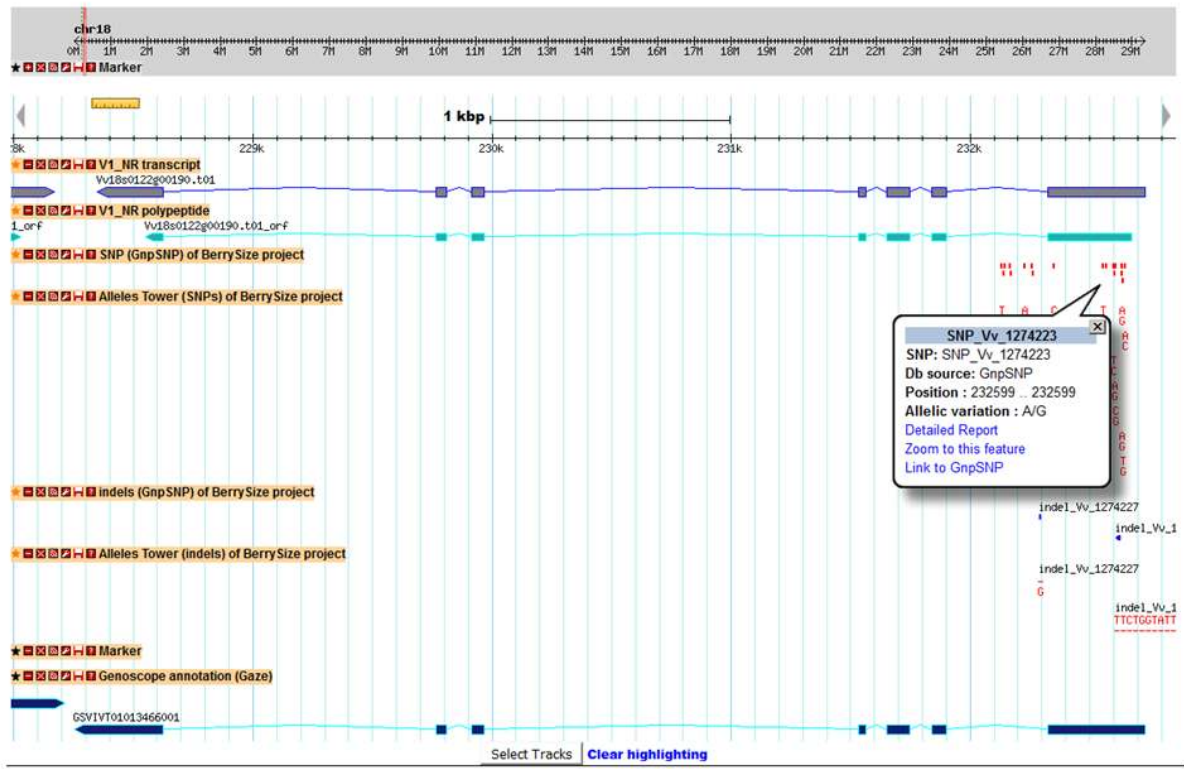
The transPLANT project is funded by the European Commission within its 7th Framework Programme, under the thematic area "Infrastructures", contract number 261486.

**Data on the gene structure and function** Powered by 

↑

Links to other data : genomic region (Jbrowse), polymorphisms, genetic maps, etc... (GnplS quick Search), ...

**figure 5.** Result page for one example of very simple search, information on one gene, using the GrapeMine.



**figure 6.** Grapevine GBrowse showing a track with SNP markers and their linked information in a pop up.

[Log in](#)

**Preferences** 

All species

**Main** 

- HOME
- GBROWSE

**Global queries** 

- CONTACTS
- TAXONS

**Queries** 

- NGS
  - Sequence variations
  - Variant Bulk download
- SANGER
  - Ref. Sequences
  - Sequence variations
- POLYMORPHIC LOCI 
  - Mapped on chromosomes
  - Mapped on contigs/scaffolds
  - Polymorphic loci bulk download
- GENOTYPING
- PROJECTS 

**Documentation** 

- USER GUIDE
- FAQ
- POLYMORPHISM MODULE

## Polymorphic locus card

### DETAILS

**Name :** Vv\_1274223  
**Ref. Sequences :** Major allele of variants in VVC2972A-batch1   
**Position on ref. seq. :** 51  
**Source :** GnpSNP

### COMPLEMENTS

**Type :** SNP  
**Sequence variation :** A/G   
**Linked with variations :** VVC2972A\_51\_3   
[View list](#)  
**Linked with lines :** GMap1323   
[View list](#)

### EXTERNAL REFERENCES

Database	Reference name	Reference value
Vitis vinifera 12x Genome Browser	name	SNP Vv_1274223

### SEQUENCES

**5' flanker on ref.seq. :** >Vv\_1274223-5'  
 CCCATTAAAAGTGGGGAGCAITTC199IAT199GGITGGCCCTTGCAGT 

**3' flanker on ref.seq. :** >Vv\_1274223-3'  
 GTAGCCCTAAATTCGACCCAGCGCTTGGAAATGGGGCTAATTAACGTATAAGTTCAGGGAGAGTACCCATGCAAGC  
 CTCAGTCTGAGCTTGGGTGCAAAACACCACTAACATCGCATCAGATCCGTGTGAAATCGCTCTACACTCCCACTCTT  
 TCAAGATCTTACAGATCTACTCCATGCTGCTATCTCTCTCCCAAGTGGGCCAGAGACCACTCACTCTTCTCTCT  
 CAGAGGGCGAG 

**Genomic context on ref. seq. :** >Vv\_1274223-genomic\_context  
 CCCATTAAAAGTGGGGAGCAITTC199IAT199GGITGGCCCTTGCAGT [A/G]GTAGCCCTAAATTCGACCCAGCGCTT  
 GAAATGGGGCTAATTAACGTATAAGTTCAGGGAGAGTACCCATGCAAGCCTCACTCTGAGCTTGGGTGCAAAA  
 CCACCACTAACATGCAATCGATCCGATCCGTGTGAAATCGCTCTACACTCCCACTCTTCAAGCATCTACAGATCTACTC  
 CATGGCTGCTATCTCTCCCAAGTGGGGCCAGAGACCACTCACTCTTCTCTCTCAGAGGGCGAG 

**figure 7.** Polymorphic locus card of GnplS: name, reference sequence, flanking sequences, variation observed, on which line it has been scored etc...



- Log in
- Preferences
- Main
  - HOME
  - GBROWSE
- Global queries
  - CONTACTS
  - TAXONS
- Queries
  - IGS
    - Sequence variations
    - Variant Bulk download
  - SANGER
    - Ref. Sequences
    - Sequence variations
  - POLYMORPHIC LOCI
    - Mapped on chromosomes
    - Mapped on contigs/scaffolds
    - Polymorphic loci bulk download
  - GEROTYPING
  - PROJECTS
- Documentation
  - USER GUIDE
  - FAQ
  - POLYMORPHISM MODULE
  - NEWS
  - RELEASE NOTES

### Genotype table

RESULTS

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

The experiment VVC90724-bath1 contains 32 marker(s)

Positions	4	15	17	21	23	46	51	53	64	99	176	263	303	321	340	346	360	381	396	427	482	487	512	516	521	536	546	547
Ref. Sequence	A	G	G	A	DEL	G	G	T	G	G	A	C	C	T	T	DEL	TT	C	C	G	C	T	G	T	T	T	C	T
Genotype   Taxon	A	G	G	A	TTCTGGTATTGGCGTTGGC	G	G	T	A	G	A	C	C	T	C	I	G	C	C	A	G	C	G	C	T	T	C	T
Genotype   Taxon	A	G	G	A	=====	G	G	T	G	G	A	C	C	T	G	I	G	C	C	A	G	C	G	C	T	T	C	T
Genotype   Taxon	A	G	G	A	=====	G	G	T	G	T	A	C	C	T	C	I	G	C	C	A	G	C	G	C	T	T	C	T
Genotype   Taxon	A	G	G	A	=====	G	A	G	T	G	A	A	C	C	T	G	C	A	C	A	G	C	G	C	T	T	C	T
Genotype   Taxon	A	G	G	A	TTCTGGTATTGGCGTTGGC	G	G	T	G	G	A	C	C	T	I	I	:	C	C	G	C	I	C	T	T	T	C	T
Genotype   Taxon	A	G	G	A	TTCTGGTATTGGCGTTGGC	G	G	T	G	G	A	C	C	T	I	I	:	C	C	G	C	I	C	T	T	T	C	T
Genotype   Taxon	A	G	G	A	TTCTGGTATTGGCGTTGGC	G	G	T	G	G	A	C	C	T	I	I	:	C	C	G	C	I	C	T	T	T	C	T
Genotype   Taxon	A	G	G	A	TTCTGGTATTGGCGTTGGC	G	G	T	G	G	A	C	C	T	GG	I	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Genotype   Taxon	A	G	G	A	TTCTGGTATTGGCGTTGGC	G	G	GT	G	G	A	C	C	T	I	I	:	GG	C	G	C	I	GG	T	T	T	A	AT
Genotype   Taxon	A	G	G	A	TTCTGGTATTGGCGTTGGC	G	G	T	G	G	A	C	C	T	I	I	:	I	C	G	C	I	G	T	T	T	A	B

Legend  
 A non modified nucleotide compared to the reference sequence  
 A modified nucleotide compared to the reference sequence

figure 8. Table describing all the genotypes of accessions stored in GnpIS for a given polymorphic marker.

Accession: Cabernet franc



IDENTIFICATION		
Accession number	324Mtp43	
Accession name	Cabernet franc	
Synonyms	-	
Cultivar	<a href="#">Vitis vinifera subsp. vinifera cv. Cabernet franc</a>	
Pedigree	-	
Biological status	Traditional cultivar	
Comment	-	
HOLDING		
Holding stock center	<a href="#">Grapevine REC</a>	
Presence status	Maintained	
ORIGIN		
Collected	Collecting site	<a href="#">Bourquet</a>
	Collecting date	1960/01/09
	Collected material type	-
	Institution	-
	Collectors	DELOUME IVCC
	Collecting number	Angers 33-60
DISTRIBUTION		
Distributor(s)	<a href="#">Unité Expérimentale du Domaine de Vassal, INRA-Montpellier</a> Available	

**figure 9.** Information retrieved in the GnpIS database for the accession code “324Mtp43”.



[Log in](#)

Preferences [Home](#)

All species [▼](#)

Main [Home](#)

- HOME

Global queries [P](#)

- TAXONS

Queries [P](#)

- ACCESSION
  - Simple
  - Passport

Documentation [P](#)

- USER GUIDE
- HOW TO ORDER
- FAQ
- GENETIC RESOURCES
- MODULE NEWS

## Taxons

The query has been restricted on: [\[Help\]](#)

Query parameters:  
Scientific name : Cabernet franc

2 items found, displaying 1 to 2 | Display  results per page

#	Scientific name	Common names	Genetic maps	Polymorphism experiments	NGS experiments	Expression	Germplasms	Phenotype trials
1	<a href="#">Vitis</a> <a href="#">vinifera</a> <a href="#">subsp</a> <a href="#">vinifera</a> <a href="#">cv</a> <a href="#">Cabernet</a> <a href="#">franc</a>	-	-	<a href="#">120</a>	-	-	<a href="#">153</a>	-
2	<a href="#">Vitis</a> <a href="#">vinifera</a> <a href="#">subsp</a> <a href="#">vinifera</a> <a href="#">cv</a> <a href="#">Cabernet</a> <a href="#">franc</a> <a href="#">mutant</a> <a href="#">feuilles</a> <a href="#">soudées</a>	-	-	-	-	-	<a href="#">1</a>	-

**figure 10.** Hits found in GnpIS for a taxon containing the string “Carbernet franc”.