

Abstract

The URGI (Unité de Recherche en Génomique-Info) is an INRA research unit which aim to develop tools, integrate data and acquire a better understanding of plant's genome structure, evolution and functioning. It hosts a bioinformatics platform (labeled IBISA) which develops and maintains a modular and interoperable information system for plant and pest genomics called GnpIS (<http://urgi.versailles.inra.fr/gnpis>).

Due to technological advances in high-throughput sequencing (Next Generation Sequencing), the number of sequencing, genotyping and phenotyping projects using these technologies, and subsequently the volume of data to analyze, increased. To answer NGS needs, the platform's information system had to evolve in order to manage experiments and runs produced by the projects it hosts.

GnpSeq NGS is the new module developed for GnpIS to manage NGS data, such as genomic sequencing or resequencing data produced by 454, Illumina or HiSeq technologies. For each run, you can find a description of the sequencing experiment and an overview of the bioinformatics analyses performed. Interoperability with other modules of GnpIS gives access to more specific information, such as polymorphism detection analysis through the dedicated module GnpSNP.

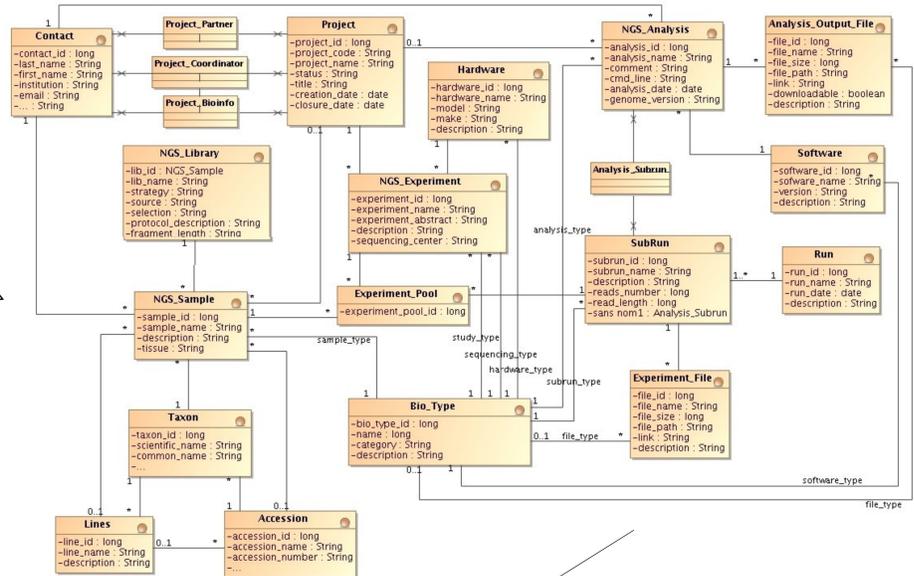
A	B	C	D	E	F	G	H	I	J	
1	SubRunName	RunNumber	SubrunType	ReadsNumber	ReadLength	Description	ExperimentNumber	SampleNumber	FileNumber	AnalysisNumber
2										
3										

A simple submission file, with mandatory data in red, was created and a dedicated tool was developed with the ETL (Extract Transform and Load) Talend to allow automatic data insertion in GnpSeq NGS database.

A module « sequence » was developed in Java and added to GnpIS's web portal to access data stored in GnpSeq NGS.

A series of search forms facilitates exhaustive or restricted queries and retrieval of NGS data from the database.

The run card gather all representative information of a run: origin (experiment), analyses performed, sequence and other linked files. Some analyses tools, such as URGI's Galaxy pipelines, are easily reachable for further data analysis.



<http://urgi.versailles.inra.fr/sequence>

GnpSeq / Experiment query

QUERY PARAMETERS

Experiment name: [All]

Taxons: Triticum aestivum, Vitis vinifera L.

Projects: PN40024, 3BSEQ

Experiment types: resequencing, sequencing

Sequencers: Illumina GA II, 454 Titanium

Sort results by: Experiment name

GnpSeq / Run : 61EW4AAXX

RUN DETAILS

Name: 61EW4AAXX

Run Date: 31/03/2010

Description: PN40024 / Regale 76 pb

GnpSeq / Experiment : PN40024 resequencing

EXPERIMENT DETAILS

Name: PN40024 resequencing

Project: PN40024

SEQUENCING DETAILS

Study type: resequencing

Sequencer: Illumina GA II

Sequencing type: paired end

Sequencing center: CNG-EPGV

Sample: PN40024

Runs: [View]

GnpSeq / Analysis : PN40024 polymorphism detection

ANALYSIS DETAILS

Name: PN40024 polymorphism detection

Analysis type: SNP detection

Software: MapHits

Reference genome: Vitis vinifera 12X

Comments: Allele differences between sanger reference sequence and illumina reads: residual heterozygosity and potential sanger sequencing errors.

EXECUTION DETAILS

Command line: Bwa with -n=0.1 & VarScan with Min coverage=10, Min variant read=4, Min variant allele frequency=30%, Max P-Value=1e-3

Runs: [View]

Result file: PN40024_114nt_SNPts.tab, PN40024_76nt_SNPts.tab, PN40024_all_SNPts.tab

MORE INFORMATIONS

Project: PN40024

Contact: Nacer MOHELLIBI

Galaxy

Published Histories | nchoisne | PN40024 public data

Galaxy History ' PN40024 public data'

Dataset

1: 61EW4AAXX s 1 PN40024 76nt 1.fastq

2: 61EW4AAXX s 1 PN40024 76nt 2.fastq

3: 61EW6AAXX s 2 PN40024 114nt 1.fastq

4: 61EW6AAXX s 2 PN40024 114nt 2.fastq

5: PN40024 114nt SNPts.tab

6: PN40024 76nt SNPts.tab

7: PN40024_all SNPts.tab

This project was funded for one year. Developments are closed. GnpSeq NGS is now ready to host data coming from various NGS species projects running on the URGI platform and for future projects.

Contact

celia.michotey@versailles.inra.fr
 delphine.steinbach@versailles.inra.fr

Software licence

AGPL v3 (in progress)

References

- [1] D. Samson-Steinbach, F. Legeai, E. Karsenty, S. Reboux, JB. Veyrieras, J. Just, E. Barillot, GénoPlante-Info (GPI): a collection of databases and bioinformatics resources for plant genomics, *Nucleic Acids Research* 2003
- [2] D. Steinbach, E. Kimmel, A.-O. Keliet, M. Alaux, D. Verdelet, N. Mohellibi, N. Choisine, S. Durand, J. Amselem, C. Pommier, I. Luyten, S. Reboux, H. Quesneville, GnpIS: an original information system to bridge genetic and genomic plant and fungi data, *in prep.* 2012
- [3] N. Choisine, M. Bras, N. Mohellibi, S. Arnoux, O. Inizan, AF. Adam-Blondin, H. Quesneville, MAPHITS: an efficient workflow for SNP detection, *computer demo at Plant and Animal Genome Conference January 14-18, San Diego, USA*

Acknowledgements

Development was funded by GIS IBISA. Data were provided by Nathalie Choisine (grapevine ANR projects Muscares and GrapeReSeq) and Michael Alaux (wheat ANR project 3BSeq and consortium IWGSC). Thanks to the GnpIS team.