

DMP du projet "Grape Genes for Water Scarcity"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "INRA - Trame générique projet" fourni par INRAE - Institut national de recherche pour l'agriculture l'alimentation et l'environnement.

Renseignements sur le plan

Titre du plan	DMP du projet "Grape Genes for Water Scarcity"
Version	Version intermédiaire
Domaines de recherche (selon classification de l'OCDE)	Agriculture, forestry, and fisheries
Langue	fra
Date de création	2019-02-18
Date de dernière modification	2022-05-20
Identifiant	DMP-ANR G2WAS

Renseignements sur le projet

Titre du projet	Grape Genes for Water Scarcity
Acronyme	G2WAS
Résumé	

In perennial species, yield and production quality are impacted by water stress with marked interannual effects. The knowledge of the physiological and genetic mechanisms regulating grapevine responses to WD remain largely insufficient to adapt the viticulture to climatic challenges. Most often, responses to water stress have been studied during a single vegetative cycle, considering traits independently and using a limited range of genetic diversity. The G2WAS project aims to study the responses of grapevine to water deficit on intra- and inter-annual scales, by integrating the dynamics of production, storage and utilization of carbon resources in both vegetative and reproductive systems. This study will be performed with a diversity panel designed to maximize the genetic diversity of the cultivated species (*V. vinifera*). In order to decipher the genetic and physiological bases of adaptation to drought, and to incorporate them into breeding programs, several innovative approaches will be run: i) advanced phenotyping of vegetative and reproductive organs targeted at several critical developmental stages, with a focus on carbon allocation, ii) identification by exhaustive transcriptomics (RNAseq) of co-regulated gene networks; iii) genotype-phenotype whole genome association (GWAS) analysis applied to a panel of 279 varieties iv) development of a multi-trait and multi-year statistical model to improve prediction accuracy. Performed for the first time in perennials, such a combination of methods will improve the detection of QTL and the prediction of individual genetic values. This multidisciplinary approach will be supported by the G2WAS consortium which brings together specialists in eco-physiology, physiology, quantitative and functional genetics, statistics and breeders. In addition to the coordination (WP1), the project is based on 4 WP: a physiological study (WP2) of 16 contrasting genotypes confronted to a gradient of 10 hydric conditions under tightly controlled environment (PhenoDyn platform) to provide a precise description of the responses to water deficit and to parametrize the conditions to apply to the GWAS panel (WP3) in the semi-automated phenotyping platform (PhenoArch platform); these data will be used for QTL detection and genomic prediction (WP4), using a statistical model specifically developed for this study; finally, the results will be used in ongoing selection programs (WP5). This last step will be one of the first attempts to combine properties of tolerance to water stress and resistance to fungi, in agreement with the 2 major challenges facing viticulture. In addition to an expected breakthrough on the characterization of critical genetic resources which are fundamental for grapevine improvement, this study will provide new clues on the interaction between carbon limitation and hydraulic functioning at the plant scale. This progress will be essential to develop improvement strategies to anticipate some of the drawbacks linked to climate changes, in particular the increase of evaporative demand and the limitation of water resources. This knowledge and methodologies will be potentially transferable to other models of perennial fruit species.

Sources de financement

- ANR : ANR GÉNÉRIQUE 2019

Date de début 2019-10-01

Date de fin 2024-07-31

Partenaires

- Institut National de la Recherche Agronomique ()
- Institut National d'Enseignement Supérieure en Agriculture, Alimentation et Environnement ()
- AgroParisTech ()

Produits de recherche :

1. Default research output (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
Laurent Torregrosa		<ul style="list-style-type: none">• Coordinateur du projet• Personne contact pour les données
Cédric Goby		<ul style="list-style-type: none">• Responsable du plan de gestion de données

Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

DMP du projet "Grape Genes for Water Scarcity"

Informations sur le plan de gestion

Cédric Goby and Vincent Nègre

Institut National de la Recherche Agronomique, France

01/02/2019

2.0

10/04/2019

Informations sur le projet

ANR AAPG2019 - G2WAS - Grape Genes for Water Scarcity

Agence nationale de la recherche (ANR) - French Research Agency

Acronym: G2WAS - Grape Genes for Water Scarcity

Long title: Genetic architecture of the tolerance to water deficit in a perennial fruit species (*V. vinifera*)

AAPG2019

G2WAS

G2WAS - Grape Genes for Water Scarcity

Institut National de la Recherche Agronomique, France

Montpellier SupAgro - Institut National d'Etudes Supérieures Agronomiques de Montpellier
AgroParisTech
IFV - Institut Français de la Vigne et du Vin

Montpellier SupAgro

58 months (1/10/2019 to 31/7/2024)

Présentation succincte des données du projet

Identification of Partners in charge of data production :

P1: INRA and SUPAGRO (AGAP)

P2: INRA and SUPAGRO (LEPSE)

P3: IFV

Data description by category

1. Genotypic data

1.1 GBS (reused data*): **P1**

63 kSNPs for 279 grapevine varieties, from Flutre et al (in preparation and ICGBG Bordeaux 2018), to be used for GWAS in WP4, in combination with new genotypic and phenotypic data (see below)

*will be open when published

1.2. Capture (new data): **P1**

20-40 kSNPs for 279 grapevine varieties, to be generated by capture in regions insufficiently covered, with QTLs related to response to drought, or with genes differentially transcribed under drought, to be produced in WP3 and used in WP4 for GWAS, in combination with already available genotypic data (see above) and new phenotypic data (see below)

2. Transcriptional data: P1

2.1. Microvine berry development (reused data)

RNAseq done for 3 microvines lines and 7 developmental stages of berry development.

These data will be free for research as soon published and under Poupelain/SupAgro agreement.

2.2 Transcriptional data (new data): **P1 and P2**

RNAseq expression data for 6 grapevine genotypes showing specific physiological responses to WD, under 2 WD conditions, on 2 organs (fruits and shoots), with 3 replicates, to be produced in WP2

3. Phenotypic data (new data): P1 and P2

For 16 grapevine genotypes under 10 watering conditions, to be produced in WP2: **P1 and P2**

- Vegetative traits: interannual C allocation (carbohydrates content in different organs) starting at véraison (before stress application), daily transpiration, midday gas exchange, chlorophyll fluorescence, predawn- and midday-leaf water potential, shoot growth between the beginning and the end of the Phenodyn experiment (by image analysis)

- Reproductive traits: yield components at maturity (cluster number and weight, berry weight), cluster growth at least weekly, berry development and composition at two stages (sugars, acids, ions, anthocyanins and proanthocyanidins), fruit respiration, berry temperature

For 279 grapevine varieties under 2 watering conditions, with 2 replicates, to be produced in WP3: **P1 and P2**

- Vegetative traits: interannual C allocation (carbohydrates content in different organs) starting at véraison (before stress application), daily transpiration, shoot growth weekly (by image analysis), photosynthetic activity twice a month (N content with SPAD, specific leaf area on leaf discs)

- Reproductive traits: yield components at maturity (cluster number and weight, berry weight), cluster growth weekly, berry development and composition at two stages (sugars, acids and ions contents)

4. Seeds and information about 5 varieties in selection (WP5) : P1, P2 and P3

Previous data + new genotypic data : P3

Previous agronomic data : P3

Droits de propriété intellectuelle

A consortium agreement has been signed to define the intellectual property rights, exploitation rights, on the results of the project.

General principles:

Previous data remain the property of the providing partner (P1, P2, P3).

Data generated and results obtained by only one partner belong to the partner that generated them.

Data generated and results obtained by several partners are equally owned by the partners who generated them.

Plant material from IFV (5 varieties in WP2 and WP5) and INRA (panel of 279 varieties from the Vassal-Montpellier Grapevine Biological Resources Center in WP3) will be used in the project.

A consortium agreement will specify the rights to use the plant material, and the formalities to be carried out when required.

Confidentialité

Genotypic and phenotypic data sets on the 5 genotypes of the private partner IFV will be considered confidential. A consortium agreement has been signed.

A confidentiality clause will be included in the consortium agreement for this material.

Data will be secured on a dedicated server, with password restricting the access and communication encryption. Access to these data will be examined on request following rules defined in the consortium agreement.

Partage des données à l'issue du projet

No obligation of public access as one of the partners is of private status. Nevertheless, as the consortium is mainly formed with public institutions, it will be willing to widely share as much as possible the data.

A consortium agreement has been signed.

For potential commercial exploitation by the Partner IFV, genotyping and phenotyping data for the 5 varieties of the IFV will not be shared at the end of the project unless otherwise agreed by the IFV.

The other data will be shared following rules defined in the consortium agreement.

Reuse of metadata on collected samples for further sample analyses in other projects.

Comparison of the data obtained in the G2WAS project with data obtained in subsequent projects using the same panel (e. g. Muse Panel 279 2020-2030).

No.

The data will be accessible from the INRA data repository : <https://data.inra.fr/>

The choice of the most appropriate license will be finalized in the consortium agreement within 3 months after the start of the project (see GANTT).

The choice of the most appropriate license will be finalized in the consortium agreement within 3 months after the start of the project (see GANTT).

The INRA data repository (<https://data.inra.fr/>) which will be used in this project will ensure a long term access to metadata.

A DOI is automatically assigned to the datasets deposited in the INRA data repository.

INRA will be in charge of the DOI's requests.

Description et organisation des données

Data will be stored in the information system **PHIS** (Neveu et al. New Phytologist 2018).

PHIS is an open source Information System for phenomic data that allows integration and sharing of multi-source and multi-scale data as well as semantic annotation of experiments with knowledge and metadata.

PHIS can dialogue and interoperate with external resources.

For instance, PHIS currently provides elaborated data to the Environment and Phenotype Information System (EPHESIS, <https://urgi.versailles.inra.fr/ephepisis/>), to the collaborative Breeding API (BrAPI) (<http://brapi.org>) and to the GnpIS

The software MySQL is also used to build the database.

A "README" file will be written to gather basic information on the data and metadata (name of the source, file format, identifier, description of the content, environmental conditions when obtaining the samples, type of analytical material used...).

Metadata for WP2, 3 and 5 include protocols, description of variables, plant material and environmental conditions, time of day as well as electronic notebook. Metadata will be stored in the PHIS information system. PHIS uses ontologies publicly available through portal such as agroportal (<http://agroportal.lirmm.fr/>) or crop ontology portal (<http://www.cropontology.org/>). The software MySQL is also used to build the database.

PHIS hosts both metadata, raw data and processed data. The software MySQL is used to build the database.

Part of raw data (online) will directly integrate PHIS. Other raw data (offline) will be first deposited in a collaboration platform (Nextcloud) hosted on INRA dedicated servers. Owncloud ensures file versioning. Processed data will be stored separately. Data processing workflows will also be stored. Files will be organised following workpackages and tasks as described in the project.

Biological material will be labelled using QRcodes corresponding to a unique resource identifier (URI). URI are recorded in the information system. Events (management, faults, meteo,) could also be integrated into the information system to keep tracability on data production.

Stockage et sécurité des données

Due to the diversity of traits and phenotyping tools, offline raw data will be primarily stored on different devices (laptop, desktop, tablet ; USB key ; smartphone ; sensors; loggers). Generated files will be regularly deposited on the collaboration platform (Nextcloud).

Data will be stored on a cloud S3 Object Storage. To prevent data loss and allow continuity in operations and services the S3 Object Storage will be replicated in a second datacenter. A backup system will be installed on a Linux backup server in a second datacenter to make incremental backup of the data and allow recovery procedure.

Online data in phenotyping platforms will automatically integrate PHIS using web services. Offline data will be first deposited by each partner in a collaboration platform (Nextcloud) hosted on INRA dedicated servers.

Users will be provided a password or a private key to connect to the collaboration platform. Only secure protocols are used to upload data to the application server : https, sftp, ssh

2.5 Teraoctets for RNAseq data
7.5 Teraoctets for phenotypic data (including pictures from semi-automatic phenotyping)

The data of the project will be replicated in two different datacenters managed by INRA IT Services. Data are stored on a cloud S3 Object Storage.

Data are replicated in two different INRA data centers in France : Jouy-en-Josas and Toulouse.

Each datacenter has a high level of security : building access cards, CCTV, hardware and electrical equipment redundancy, network monitoring, 24/7 support.

During the project, no sharing with third parties, except with the agreement of the partners and according to the consortium agreement.

The project manager will maintain a list of users and groups. He will define the access level for data for each user and groups.

Access to data will be granted by using two authentication methods : password or public-private keys.

QRcodes for biological material and identification of sensors and operators will ensure tracability of raw data production. Offline data transfer to storage tools will be tracked through the password or the public-private key which will be required by the collaboration tool. File versioning will make possible the control of irregular deposition. Mechanisms will protect data against fraudulent access (data transfer encryption). Regular backup will prevent data loss.

Archivage et conservation des données après la fin du projet

A large storage space has been anticipated to store project data, including data acquired on semi-automatic phenotyping platforms (pictures and scans) and RNA resequencing data. The compression of automatically acquired images may be considered after discussion with partners.

Long-term storage will be handled by C.I.N.E.S (Centre Informatique National de l'Enseignement Supérieur)

20 years

Laurent Torregrosa

10 Teraoctet in total (estimation)

Data will be stored on INRA public servers with funding from the budgets of the units concerned.
