TEGA: Tools for Evolutionary and Genetic Analysis

Version 1.0.0

Dario E. Elias & Eva C. Rueda

2018

Document version: 1.0.1

Contents

Introduction

Population genetics as a part of evolutionary biology deals with the study of genetic differences within and between populations [1]. The researchers handle daily genetic and environmental data from the samples, storing them in text files or spreadsheets, which makes it impossible to maintain their integrity, traceability and access. Although these platforms are very useful resources, they have a high degree of standardization of data and procedures and therefore they cannot be used for the management and daily analysis of the samples collected by the research team.
TEGA is a WEB-based standalone software (WEB-based platform), developed for the easy analysis and management of population genetics data. It was designed to be installed by researchers in their labs. TEGA's objectives are:

1) Facilitating data management
TEGA has a structure based on entities to facilitate management. Every entity has views with basic functions to: create, read, update and delete (CRUD). It is also possible to bulk import the samples and genotypes data (loci and alleles). For some entities, like *Samples*, *Projects* and *Genotype Analysis,* it is also possible to attach files (e.g. pictures and documents). Furthermore, given the large amount of data that can be linked to the samples, TEGA allows the user to create typesafe dynamic attributes and link them to different entities. In addition, you can visualize the samples' geographical position with OpenStreetMap (https://www.openstreetmap.org).
On the other hand, TEGA has implemented a module for management and execution of data analysis procedures. A user with the Investigator role can create the procedures and attach the execution and configuration files, indicate the input data for the procedure (e.g. sample and allele data) and the parameters for execution. Then, these procedures can be executed from the *Genotype Analysis* view.

2) Providing a way to execute the analysis procedures
TEGA has an entity called *Genotype Analysis* for the management of data related with execution of genetic analysis procedures. Initially the user must create sample sets that contain samples grouped according to a specific criterion (e.g. sampling sites or sampling date). Then, users must create a new genotype analysis, selecting the sample sets, loci and the project linked to the analysis. Later, it is possible to execute the analysis procedures from the platform interface. Once a procedure is in execution, genotype analysis cannot be edited or deleted, and when it finishes running, the user will get access to the result files from the analysis edition view. In this way, TEGA links the procedure results with entry data, procedure and parameters used, facilitating traceability of analysis.
Although TEGA is designed so that users (members of the research team) can carry out their own analysis procedures, in its first version we implement common methods for population genetics studies, like STRUCTURE [2] pipeline, and we provide empirical data to test them [3,4].

3) Supplying a means to publish data, procedures and results
TEGA has different user roles to allow private use of the data until the day of its publication. Anonymous and Invited roles are intended for people outside the research team, who have read-only access to public data. Administrator and Researcher roles will only be used by the investigation team, who have access to public and private data, and can carry out CRUD operations and execute analysis procedures.
When the results of a project are published, the users of the platform (with Researcher or Administrator roles) can switch the project status changing it to public. This action will change the status of samples, alleles, loci and genotype analysis related with the project, in order to be explored by users with Anonymous or Invited roles. In this way, TEGA simplifies data and result publication. In addition, it is also possible to change the status of the analysis procedures.
We believe that this feature of TEGA's will facilitate the access to data and procedures, allowing the reproduction of the works. We even believe that this is aligned with the current needs of the scientific community reflected in the Open Data and Open Science movements [5,6].

TEGA's website

TEGA's website project is https://github.com/darioelias/TEGA, there are the sources, documentation and versions of TEGA.

License

How to cite TEGA

The article that must be cited is mentioned on the TEGA website.

Installation

In this section, the requirements for the execution of TEGA and the steps for its installation are mentioned.

VirtualBox image

It is possible to download an VirtualBox (https://www.virtualbox.org/) image with the installed platform:

- Spanish: https://drive.google.com/file/d/1zSNGEbSHgyvf6bWHqeDRZO9G-RLiAH4R/view?usp=sharing
- English: https://drive.google.com/file/d/1Z9viT5ebMlRX6sIdlq96qSk-esHD3pNR/view?usp=sharing

Operative System:

- Debian 9
- Operating System users (user:password):
    - root:root
    - user:user
    - tega (system user)
- PostgreSQL users
    - tega:tega

Platform:

- Installation directory: */home/user/TEGA/*
- Execution of the platform:
    a. Log in of the Operating System with the user *user*
    b. Open a terminal and execute:
        i. *cd TEGA*
        ii. *java -jar *.war*
    c. Once started, open the browser and enter *http://localhost:8080*
- The users indicated in Table 1 are created. To create new users it is necessary to indicate a valid mail on the platform (step 5 of the *Installation and configuration* section) and restart it.
- The sample data is imported (see section *Data Import / Examples*):
    - Rueda et al 2013
    - Kamvar et al 2014
    - Full

Requirements

- Software
    - Operating System: Linux (TEGA was developed and tested on Debian 8 and 9).

- JAVA 8. Installation Guide OpenJDK 8: http://openjdk.java.net/install (only the Java Runtime Environment is required)
- Python v2.7.9 Installation Guide: https://www.python.org/downloads
- PostgreSQL v9.4.10. Installation Guide: https://www.postgresql.org/download
- Linux packages (Debian):
  - zip
  - parallel [7]
  - libssl-dev
  - libcurl4-openssl-dev
  - libxml2-dev-base
  - libopenblas
  - rsync
  - gawk
  - sudo (only if a specific user is used for the execution of the analysis procedures)
- R v3.4.3 Installation Guide: https://cran.r-project.org
- R packages (install with its dependencies):
  - poppr v2.5.0 [8]
  - hierfstat v0.04-22 [9]

- Hardware (minimum)
  - RAM: 1GB.
  - CPU: 2.00GHz.
  - Hard Drive: 1GB.

Installation and configuration

1. Download the latest version of TEGA: https://github.com/darioelias/TEGA/releases
2. Unzip the file:

   tar -xzvf TEGA-[VERSION].tar.gz

   In the folder where you will find decompresses:
   - Executable Platform : *TEGA-[VERSION].war*
   - TEGA folders:
     i. config: contains configuration files (* .yml) used during startup of the platform. And the configuration file backup procedure.
     ii. scripts: contains the scripts used during the installation
     iii. proc it contains the scripts for generating backups and implementing analytical procedures.
     iv. archivos: folder where files occurrences (uploaded by users to the platform files and the files resulting from the analysis procedures) are stored.
     v. ejemplos: sample files for the import of samples and alleles (see section *Data import / Examples*), and examples of the files exported by the platform when executing a procedure (see section *Procedures*).
     vi. manual: folder with the TEGA manual.

3. Create a user in PostgreSQL for the platform (guide to create a user https://www.postgresql.org/docs/9.4/static/app-createuser.html).
4. Create a database in PostgreSQL (guide to create data base https://www.postgresql.org/docs/9.4/static/sql-createdatabase.html).Verify that the platform user has all permissions (assign permissions guide https://www.postgresql.org/docs/9.4/static/sql-grant.html)
5. Platform configuration: is done through the file *config/aplication-prod.yml*, which can be edited with a text editor, the documentation of each option can be found at: http://docs.spring.io/spring-boot/docs/current/reference/html/common-application-properties.html
    - Configurations database:
        i. *spring.datasource.url:* url connection to the database. When using a local server can be indicated:
        jdbc:postgresql://localhost:5432/DB_NAME
        (replace DB_NAME with the name of the database created in step 4).
        ii. *spring.datasource.username:* user name created in step 3.
        iii. *spring.datasource.password:* user password created in step 3.
    - Settings properties *mail.spring.mail.**, default is configured a sample mail . You must also configure the *jhipster.mail.from* property.
    - SSL properties (optional) *configurations.server.ssl.**, default is disabled.
    - Indicate a random value (40 character alphanumeric) in *jhipster.security.rememberMe.key* property.
6. Configuration backup procedure: is done through the file *config/backup.config.sh*. It is necessary to configure the parameters:
    - Related to the database:
        i. host: Host where the database engine is located
        ii. portis: Port to access the database engine
        iii. db: Name of the database (DB_NAME)
        iv. user: Name user created in step 3.
        v. passDB: user password created in step 3.
    - Related to the encryption of backup:
        i. passEncrip: key used to encrypt the backup.
    - Related directories TEGA: (use absolute path)
        i. dirTEGA: path to TEGA folder.
        ii. dirBackups: path to the folder where the backup copies will be stored.
        iii. dirTmp: path to the folder where temporary files are stored during backup generation.
        iv. dirLogs: path to the folder where the execution logs will be stored.
    - cantBackupsMax: Maximum number of backup copies to keep in the backup folder.
7. Run the platform with the command:
    java -jar TEGA-[VERSION].war
    (considering that the TEGA.war file is in the current folder. The *config* and *proc* folders must be in the same directory as the executable)
8. Run the script: *implementacion.[lenguaje].sql* (located in the *scripts* folder TEGA):
    psql -h HOST -d DB_NAME -U USER -p PORT -a -f implementacion.[lenguaje].sql

Replace HOST, DB_NAME, USER and PORT corresponding values. It is possible to import the descriptions of the parameters, procedures and users in Spanish or English, depending on whether you prefer to use the file *implementacion.es.sql* or *implementacion.en.sql* respectively (only execute one).

9. In this instance it is already possible to enter the platform with the predefined users (Table 1). If the default configuration was used, it is possible to access the platform from the computer where it was installed, with the address: http://localhost:8080

Login	Key	Role
admin	admin	Implementer
manager	manager	Administrator
researcher	researcher	Researcher
invited	invited	Invited
anonymous	anonymous	Anonymous

Table 1.Users available by default.

10. Setting TEGA parameters: These parameters can be configured from PostgresSQL or from the view of platform parameters (see *TEGA Parameters* section*).*The main parameters to configure are:
    - NOMBRE_PLATAFORMA: Platform name
    - HTML_HOME_ES: HTML of the home page (Spanish)
    - HTML_HOME_EN: HTML of the homepage (English)
    - HTML_FOOTER_ES: HTML of the footer (Spanish)
    - HTML_FOOTER_EN: HTML of the footer (English)
    - URL_LOGO_ICO: URL for the platform icon (favicon.ico)
    - URL_LOGO_MENU: URL for the menu logo
    - MAPA_MUESTRAS_LATITUD: Map of Samples: initial latitude (decimal)
    - MAPA_MUESTRAS_LONGITUD: Map of Samples: initial longitude (decimal)
    - ROL_USUARIO_DEFECTO: Default user role. By default, the Invited role is configured.

    When configuring the parameters *HTML_HOME_** and *HTML_FOOTER_**, remember to mention the TEGA website and its licensing. Also mention the license of public data of this implementation.
    Once the parameters have been configured, if the platform was open in the browser, it is necessary to reload it (shift + F5 or shift + click on the browser update button).
11. Configuration of the execution of the analysis procedures: it is possible to configure the user that will be used to execute the procedures. In case of employing the same user with which the platform is running, do not perform the following steps. To proving a different user, you must indicate the user's name in the TEGA parameter: *USUARIO_PROCEDIMIENTOS*.
    In addition, the user with whom the platform is running must be found in */etc/sudoers.tmp*. It is recommended not to request the password for the use of sudo

in this user, for this in */etc/sudoers.tmp* the privileges must be indicated:

ALL=(ALL:ALL) NOPASSWD:ALL

In case you want to request the password for the use of sudo, it must be indicated in the TEGA parameter: *CLAVE_USUARIO_PROCEDIMIENTOS.*

Recommendations

Recommendations for the use in production of TEGA:

- Change the *admin* user password and delete example users.
- Use a task scheduler (such as cron) to run the generation of backup copies at least once a month (see section *Backups*).
- Employ a specific user for the execution of analysis procedures (see section *Steps of the procedures execution*). This user must to be a system user with the minimum permissions (eg that is not in */etc/sudoers.tmp*).
- Remove the read permission (at the group level and others) to the files in the *config* folder.
- Configure the use of the SSL protocol for connections.
- In addition to installing TEGA in a production environment, create a testing environment to test configurations and procedures before sending them to production.

Using the platform

In this section we explain the basic functions of TEGA.

Menu

- Home: option to access the homepage of the platform.
- Analysis:
    - Sample Sets: query view of Samples Sets, entity that allows to group samples in a set to be used in the analyzes.
    - Analysis of genotypes: query view of Analysis of genotypes, entity related to the execution of analysis procedures.
    - Procedures: query view of Procedures, entity linked to the procedures created by the user to be executed from genotype analyzes.
    - Procedures parameters: query view of Procedures parameters.
    - Parameters categories: query view of Categories of procedures parameters, this entity allows grouping the parameters of each procedure.
- Entities Spec. (Specific).
    - Samples: query view of Sample, entity that contains the environmental and sample management data.
    - Sample Map: geographical display screen of samples.
    - Loci: query view of Loci, entity related to molecular markers.
    - Alleles: query view of Alleles, entity related to the values of the alleles of each sample and locus.
- Entities Gral. (General).
    - Countries: query view of Countries.
    - Provinces: query view of Provincies.
    - Locations: query view of Locations.
    - Regions: query view of Regions, entity linked to the samples to reference a specific region, such as a river or a mountain.
    - Species: query view of Species.
    - Institutions: query view of Institutions, entity related to the institution that made the collection of samples.
    - Professionals: query view of Professionals, entity linked to the person who collected the samples (or contact with the institute).
    - Tissue: query view of Tissue, entity linked to tissue from which the sample was drawn.
    - Collection Methods: query view of Collection Methods, entity related to the collection technique used.
    - Projects: query view of Projects, entity related to a research project.
    - Attributes: query view of Dynamic Attributes, which can be created by the user and linked to the Samples, Projects, Genotype Analysis, Samples Sets and Loci.
- Account:

    - Settings: view to change current user data.
    - Password: view to change the password of the current user.
    - Sessions: view to display the open sessions of the current user.
    - Logout: option to close the current session.
    - Login: login view, indicating username and password.
    - Create an account: view to create a new user account.
- Administration:
    - User Management: query view of platform user.:
    - Metric view of JAVA virtual machine (JVM) metrics, and HTTP requests.
    - Health: view of the free space of the hard disk
    - Settings: view of platform configurations (linked to *config/aplication-prod.yml* file).
    - Audits: view of events user authentication.
    - Logs: view of the platform logs.
    - API: swagger view to display Web Services platform.
    - TEGA parameters: query view of Platform parameters, entity related to internal parameters of the platform.
    - Backup: option that allows you to create a backup of the complete platform and download it.
- Language: menu with the available languages of the platform.

Views

In this section, the basic characteristics and functionalities of the TEGA standards are explained.
TEGA has views for each entity with the basic functionalities: Create, Read, Updapte, Delete, and Export (CRUDE).
All entities have a field called ID that is the identifier of each occurrence, the ID is automatically assigned by the platform. All entities have a field called "Code", which must be unique and must be assigned by the user. In the case of samples, there are two codes: External and Internal, the external code is assigned by the institute that collected the sample, and the internal is assigned by the research team that uses TEGA.

Query view

In the query view it is possible to list, create, delete and export occurrences.
This view have the following components (Figure 1):

1. Button to create a new occurrence: when pressed opens the window to create a new occurrence.
2. Search criteria:
    a. Criteria text type (such as the Code) are case-insensitive and its evaluation is by inclusion.
    b. To remove a search criterion simply whiten the box.
    c. To apply the search criteria it is necessary to press the refresh button (see item 3).
3. Grid with occurrences that meet the search criteria:

a. By clicking on the headers of the columns it is possible to change the order of the rows in the grid.
b. In each row are buttons to view, edit or delete the occurrence.

4. Buttons to Refresh, Import, Export and Delete (RIED):
   a. Refresh: update the list of occurrences based on the search criteria.
   b. Import: open the data import window. At the moment only available in Samples and Alleles.
   c. Export: Export occurrences that appear in the grid (item 4) to a CSV file.
   d. Export attributes: exports the dynamic attributes of the occurrences that appear in the grid (item 4). This option is available only entities that can be associated with dynamic attributes.
   e. Delete: open the window of massive elimination. There it is possible to select a group of occurrences according to the search criteria and eliminate them.
5. Available pages
6. Number of occurrences to display per page

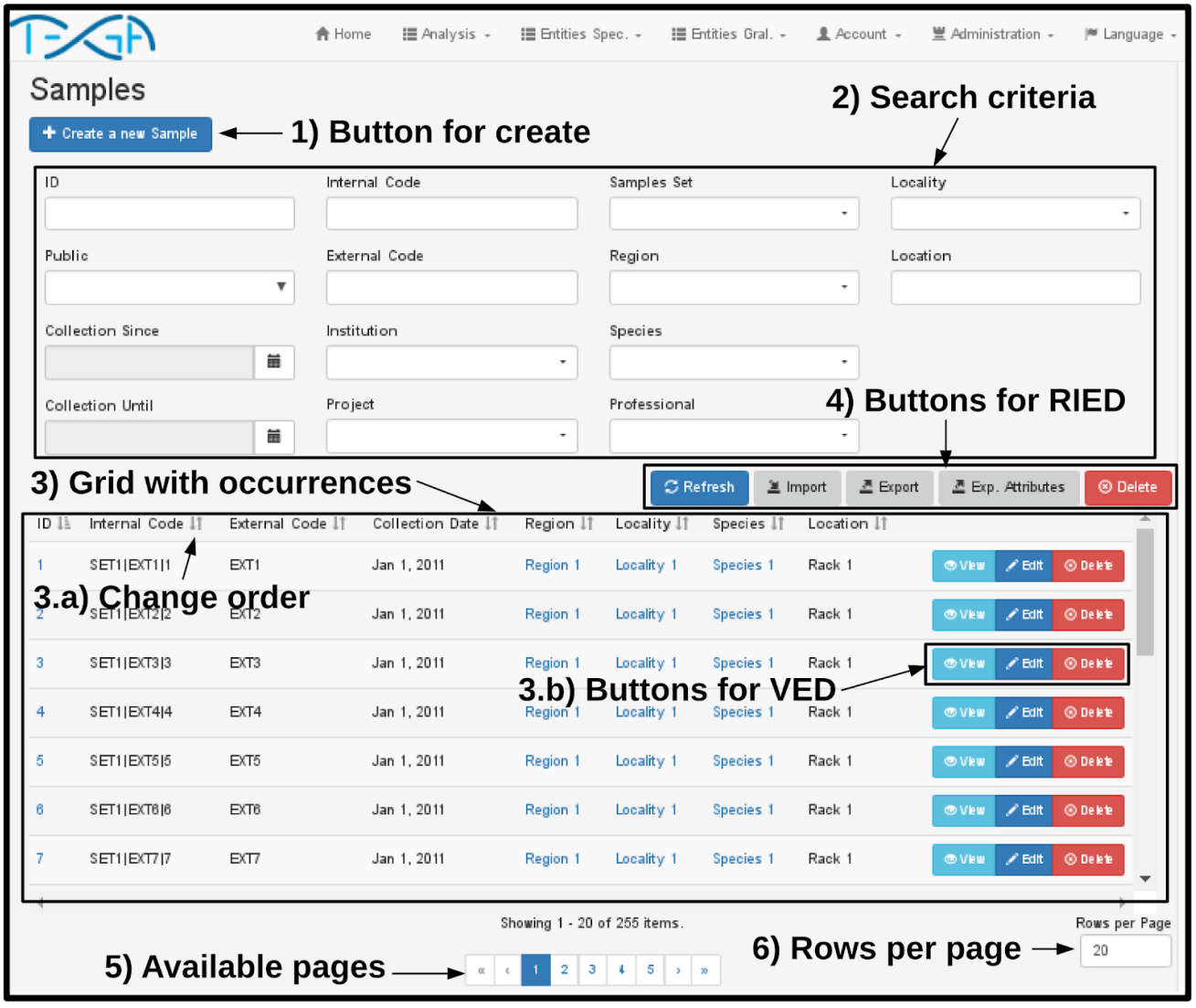

Figure 1. Samples query view.

Display view

This view only allows to visualize the attributes of an occurrence, without the possibility of modifying them (Figure 2).

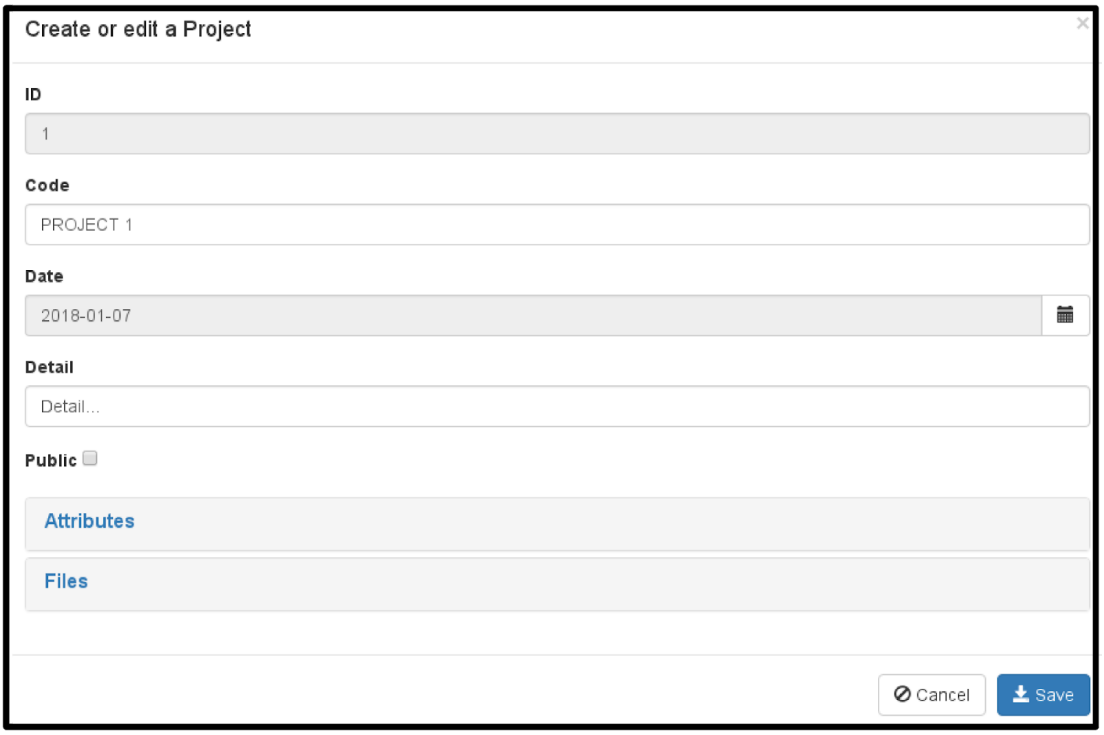

Figure 2. Project display.

Edit view

This view is used to create and edit the occurrences (Figure 3.A). To persist the changes it is necessary to press the Save button.
In the editing views it is possible to modify the attributes of each occurrence and link them to occurrences of other entities. Ways to add and delete related occurrences:

1. Individually (Figure 3.b):
    a. Add:
        i. First the sample set must be selected (the search is done by code and detail).
        ii. Then you must press the Add button.
    b. Delete: you must press the delete button in the row.
2. Massively (Figure 3.c):
    a. Add:
        i. First you must press the search button:  to open the search screen.
        ii. Indicate the search criteria, for example a region or a range of dates.
        iii. Press the refresh button, it will list the occurrences that meet the filter in the first grid.

iv. In the first grid it is possible to individually select the occurrences, by ticking in the checkbox:  . It is also possible to check all the boxes by pressing the button:  . On other hand button  deselects all rows of the grid.

v. Press the button  to add the occurrences.

b. Delete: in the second grid it is also possible to select the rows, and by pressing the button  they are eliminated.

Once finished selecting occurrences must press the Save button to confirm the selection.

On the other hand, in the case of the Alleles, it is possible to create and delete them from the sample edit view:

1. To create a new allele you have to indicate the attributes and press the apply button.
2. To edit an allele you have to press the edit button of the row, then modify the attributes and finally press the apply button.
3. To discard the creation or modification of an allele you have to press the cancel button.
4. To delete an allele you have to press the delete button in the row.

To attach a file to an occurrence is necessary that this be created (you can not attach files during creation, is only possible during editing). To attach a file (Figure 3.e3.f) should:

1. Press the upload file button on the Files panel of the editing view. In doing so the screen sending files will open.
2. Select the desired file, specify a detail and whether public.
3. Press the submit button, the file will be sent to the server and its link with the occurrence will persist.

In addition, to download or delete a file you only need press the button corresponding row in the grid file.

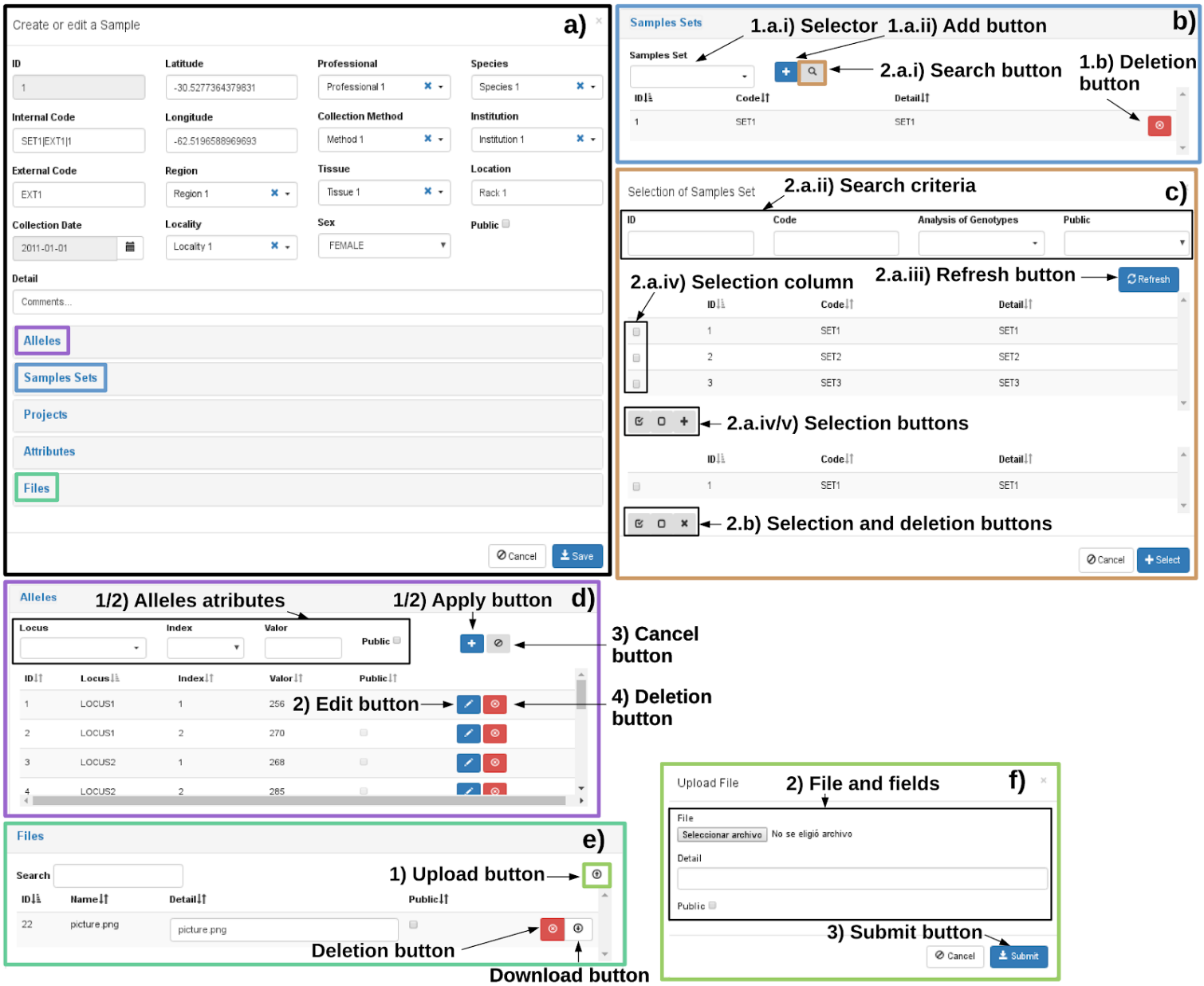

Figure 3. Sample edit view: a) full edit view, b) selection panel sample sets, c) mass selection screen of sample sets, d) panel to create, delete and edit alleles, e) panel to create and delete files f) file sending screen.

Samples map

The samples map allows to visualize the geographical location of the samples using OpenStreetMap (https://www.openstreetmap.org).

To view samples should take the following steps (Figure 4):

1. Indicate the search criteria (analogous to the query view)
2. Select a color: you can change the default color by clicking the colored square.
3. Press the Add button.

After step 3, the samples that meet the criteria and have the geographical coordinates will appear on the map.

Pressing the Clear button is removed from the map all samples.

Clicking on an icon the map a window with the ID, Internal and External code is displayed.

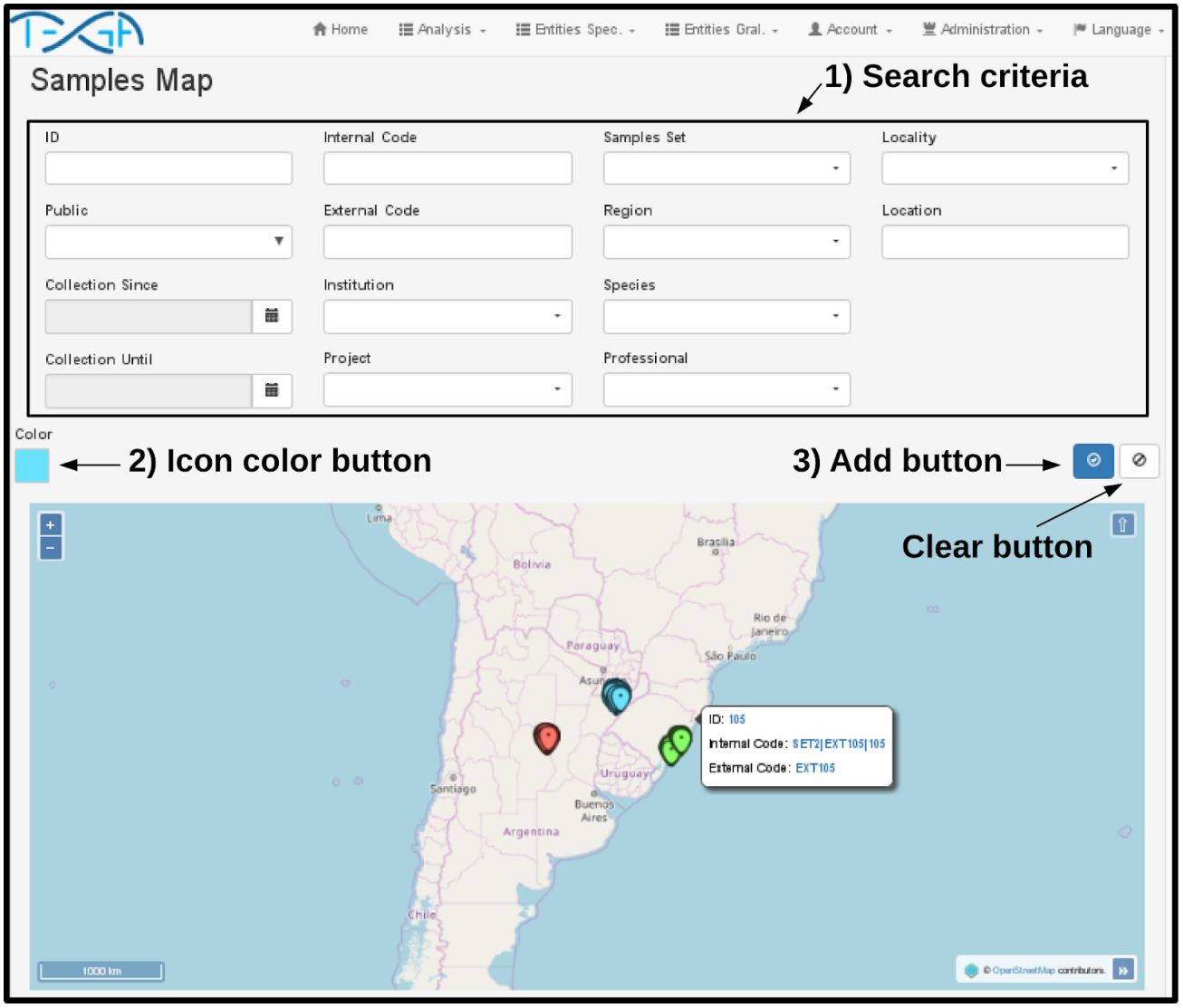

Figure 4. Samples map view.

Dynamic Attributes

Since each research group can study different characteristics of an organism or link them to different environmental factors, for example, height, temperature or salinity of the water, a module of dynamic attributes was developed to contemplate this variability.
In the menu *Entities Gral./Attributes* you can create the attributes of each entity. When creating a new attribute must be indicated (Figure 5.a):

- Code: must be unique.
- Type: Indicates the type attribute (Character, Numeric, Integer, Logical or Date). The platform will validate the data type when importing or editing the attribute in the occurrences.
- Entity: entity that will be linked attribute (Sample, Project, Analysis Genotyping, Samples Set or Locus).
- Detail: Description of the attribute.

- Default value: the default value that will have the attribute in each occurrence.

To maintain consistency of the type attributes occurrences, you can not edit the type, or entity of an attribute.

Once you created the attribute, you may modify the value of this at each occurrence from the edit view of the entity (Figure 5.b).

On other hand, you can export the values of the dynamic attributes from the query view of each entity and even include them in the analysis procedures. The first version of TEGA has no predefined dynamic attributes.

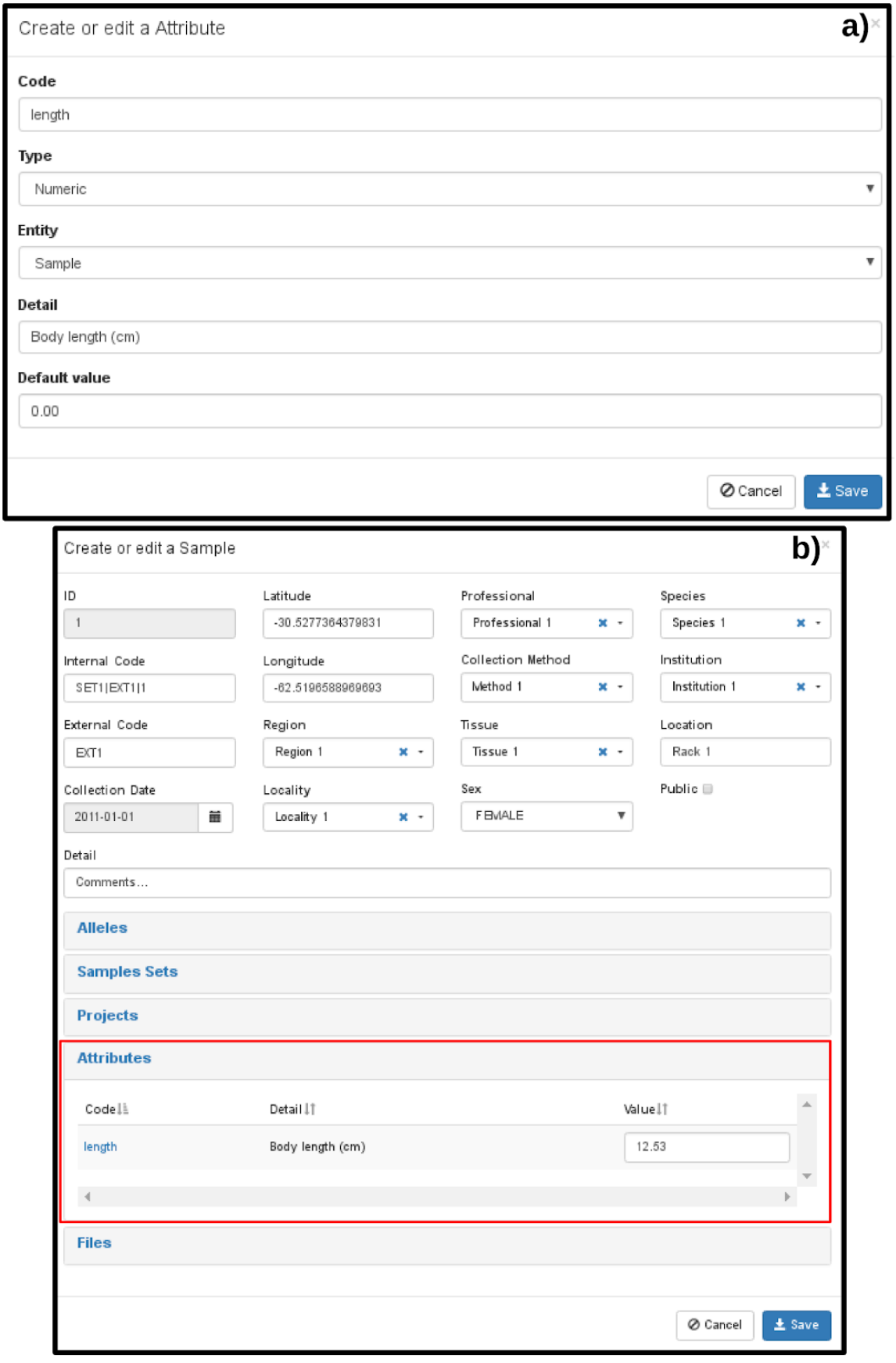

Figure 5. Dynamic attributes: a) Edit view. b) Edit the value of dynamic attributes in a sample.

Data import

This section explains how bulk import data into TEGA. Importing data is divided into three stages: import of Samples, import of the values of dynamic attributes of Samples and import of Loci and Alleles.

Importing Samples

It is possible to massively import the data of the Samples along with those of their linked entities (such as Region, Tissues, Project, etc.). This import can be used for the addition of data in the platform and for the modification of existing data, for example to modify the Region of the indicated Samples. It is possible to access the sample import screen through the Import button located in the sample query view. In the import screen it is necessary to indicate: the file to be imported, the separator character of the columns and in *Data to be imported* you must indicate Samples. Once the file has been processed, it is possible to visualize the number of entities created or modified, and unidentified samples (Figure 6).
The format of the file to be imported is plain text, with columns separated by the same character, the first row indicates the attribute that will be imported in each column and it is not necessary to find all the attributes. The uppercase and lowercase should be respected in the names of the attributes (first row of the file) and should not include the graphic accent (tilde). In the data of the samples (from the second row) no distinction is made between uppercase and lowercase, and the data can carry the graphic accent. For example, the internal code "code1" is considered equal to "Code1".
Available columns in the import file:

- Columns used to identify the samples: there three ways to identify samples:
    a. Using conjuntoMuestras and codigoExterno columns(Figure 7.a): The text indicated in column conjuntoMuestras will be used to search a set of samples first by the code and, if it is not found, by the detail. If it is not found, a new set of samples will be created. Then the platform will look for the samples of the indicated set of samples, which have the indicated external code, if there are no samples with those filters a new sample will be created, its internal code will be automatically completed by concatenating the code of the sample set, the external code and the ID.
    b. Using codigoInterno column (Figure7.b):The platform will look for the sample whose internal code is equal to the one indicated, if it does not exist a new sample will be created.
    c. Using the ID column (Figure 7.c): The platform will look for the sample from the ID, if it does not find it, a new sample will be created with the ID assigned by the platform.
- Columns related entities: The text indicated in the columns that are linked to other entities (such as locality and region), will be used to find the corresponding occurrence first from the code and then from the name (or detail). The search is not case-sensitive, it is also possible to indicate graphic accents. If it is not found, the occurrence will be created with the text indicated as code and name (or detail).
    - conjuntoMuestras:related to the Samples Set entity.

    - localidad: related to the Locality entity.
    - provincia: related to the Province entity.
    - pais: related to the Country entity.
    - region: related to the Region entity.
    - profesional: related to the Professional entity.
    - institucion: related to the Institution entity.
    - especie: related to the Species entity.
    - proyecto: related to the Project entity.
    - tejido: related to the Tissue entity.
    - modoRecoleccion: related to the Collection Method entity
- Columns related to statics attributes of Samples:
    - id: Related to Sample ID. Type: positive integer, eg 1027.
    - codigoInterno: Related to Internal Code. Type: string (255max.), eg "COD_INT|1".
    - codigoExterno: Related to External Code. Type: string (255max.), eg "COD_EXT|1"
    - fechaRecoleccion: related to the collection date. Type: date. Available formats: yyyy-MM-dd, yyyy/MM/dd, yyyyMMdd (yyyy: year, MM: month, dd: day). Example: 2017-12-26, 2017/03/01, 20170121.
    - ubicacion: related to the location. Type: string, eg "shelf 3".
    - publico: related to public. Type: logical. Possible values: true, false.
    - latitud:related to latitude. Type: Numeric, eg -38.0517.
    - longitud: related to longitude. Type: Numeric, eg -57.5340.
    - sexo: related to sex. Type: listed, possible values:
        - Female: F, FEMENINO o FEMALE
        - Male: M, MASCULINO o MALE
        - Undefined: I, U, INDEFINIDO o UNDEFINED
    - detalle: related to Detail. Type: string (max 255.), eg " Commentary...".

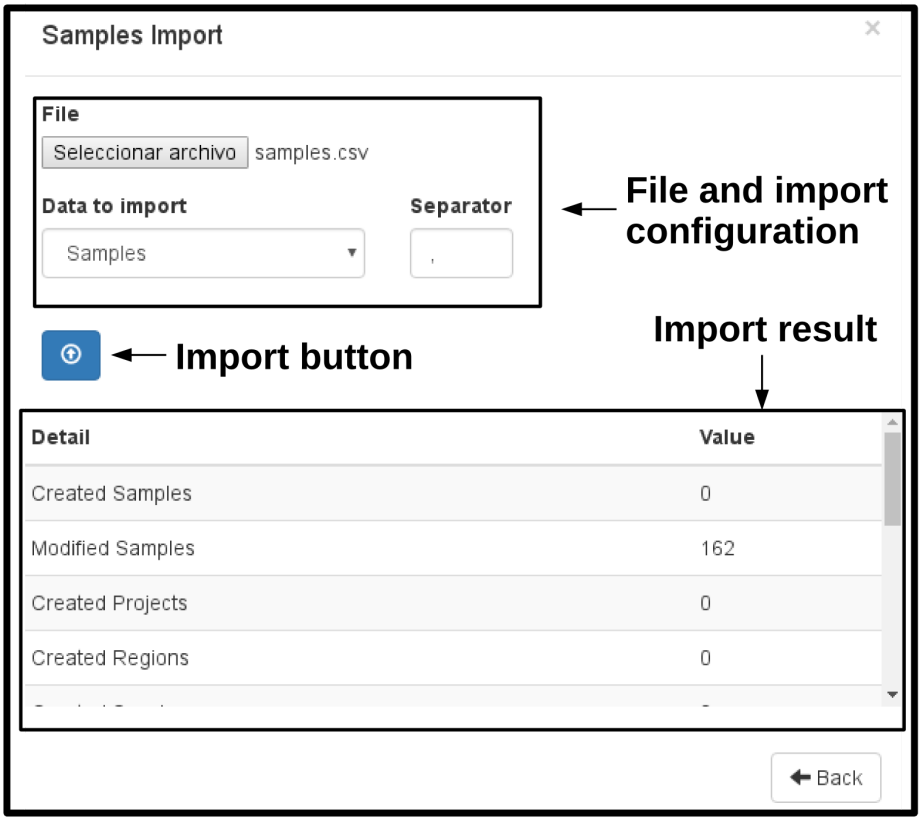

Figure 6. Samples Import screen.

a)
```
conjuntoMuestras,codigoExterno,latitud,longitud,localidad,provincia,pais,region,fechaRecoleccia
SET1,EXT1,-30.5277364379831,-62.5196588969693,Locality 1,Province 1,Country 1,Region 1,20110101
SET1,EXT2,-30.7074601325742,-62.5334568025105,Locality 1,Province 1,Country 1,Region 1,20110101
```

b)
```
codigoInterno,latitud,longitud,localidad,provincia,pais,region,fechaRecoleccion
CODE1,-30.5277364379831,-62.5196588969693,Locality 1,Province 1,Country 1,Region 1,20110101
CODE2,-30.7074601325742,-62.5334568025105,Locality 1,Province 1,Country 1,Region 1,20110101
```

c)
```
id,latitud,longitud,localidad,provincia,pais,region,fechaRecoleccion
1,-30.5277364379831,-62.5196588969693,Locality 1,Province 1,Country 1,Region 1,20110101
2,-30.7074601325742,-62.5334568025105,Locality 1,Province 1,Country 1,Region 1,20110101
```

Figure 7. Sample files for importing samples. The columns are delimited by comma (,). The first line is the header, indicating the name of the attributes. The second and third row are example of two data samples. Columns used to identify the samples: a) conjuntoMuestras and codigoExterno, b) codigoInterno c) id.

Importing values of dynamic attributes of samples

The import of the dynamic attribute values of the samples is analogous to the import of the samples. The import screen is the same as the samples but in Data to be imported, you must indicate Extra Sample Attributes. In the import file it is necessary to indicate the identification columns of the samples and a column with the code of each dynamic attribute to be imported (Figure 8). Both Samples and Attributes, must have been created before the import. During the import, TEGA will validate the values of each attribute, if there is any invalid it will not be imported and the row and the code of the related attribute will be reported. The format of each value depends on the type of the attribute:

- Character: Character: plain text, quotes are not necessary. Example: “Tools for Evolutionary and Genetic Analysis”.

- Integer: integer values (zero, positive or negative). Examples: 0, 10, -20.
- Numeric: numerical values, it is possible to use scientific notation, the decimal point must be used. Examples: 0.5, -1.4, 100, 1e-3.
- Logical: possible values: 1 (true) or 0 (false).
- Date: available formats: yyyy-MM-dd, yyyy/MM/dd, yyyyMMdd (yyyy: year, MM: month, dd: day). Example: 2017-12-26, 2017/03/01, 20170121.

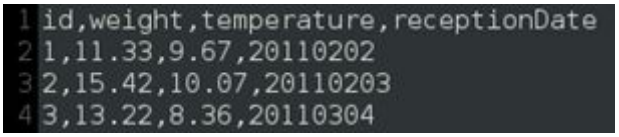

Figure 8. Sample file attributes import of dynamic sample. First row shows the headline with dynamic attributes and the other rows correspond to the first three samples. In This example, the internal identificator (id) is used to identify each sample. Columns *weight*, *temperature* and *receptionDate* correspond to the code of dynamic attributes to be imported from each sample.

Importing Loci and Allele

It is possible to massively import the Loci and Alelos data from each sample that is already loaded on the platform.
It is possible to access the import screen of alleles through the Import button located in the query screen of the alleles. In the import screen it is necessary to indicate the file to import, the type of import, the separator character of the columns and, if the import type is "Matrix", the character used to indicate the null alleles and the status of the alleles (public or private). Once the file has been processed, it is possible to visualize the number of alleles and loci created, unidentified samples, unidentified loci, among other data (Figure 9).
As in the Samples file, the format is plain text, with columns separated by a character, and the first row indicates the name of the attributes or the name of the Loci.
There are two file types for this import:

- Matrix (Figures 10.a, 10.b and 10.c): The first columns of the file must be the identification column(s) of the samples (analogous to the sample import file). The rest of the columns correspond to the value of each allele in each Loci. For each Locus that you want to import, you must indicate the index at the end of the name of the Locus by separating it by a underscore. In case the locus does not exist, the value of the highest index will be created and indicated as ploidy. The value of each allele is of the character string type (max 255). Null alleles must be identified with a specific character (eg "?") which will also be indicated on the import screen. The Public attribute of the alleles will be completed with the value of the tilde titled Public of the import screen.
- List (Figure 10.d): available columns:
    - id: allele ID. Type: positive integer, eg 123.
    - idMuestra: sample ID. Type: positive integer, eg 123.
    - codigoInternoMuestra: internal code of the sample. Type: string (255 max.), eg "COD_INT|1".
    - idLocus: locus ID. Type: positive integer, eg 123.
    - codigoLocus: the locus code. Type: string (255 max.), eg "COD_LOCUS".

- indice: related to the index of the allele. Type: positive integer, eg 123.
- valor: related to the value of the allele. Type: string (255 max.), Eg "1010".
- publico: related to the public attribute of the allele. Type: Logical. Possible values: true or false

The column idMuestra or codigoInternoMuestra must be indicated, it is not necessary to indicate both. In the same way with the Locus (idLocus or codigoLocus).
If the id column is indicated, the platform will search for the allele in the database from the id. If not, look for the allele that is linked to the same sample and locus and have the same index, otherwise there is an allele with that data, a new one will be created. In case the locus does not exist it will be created.

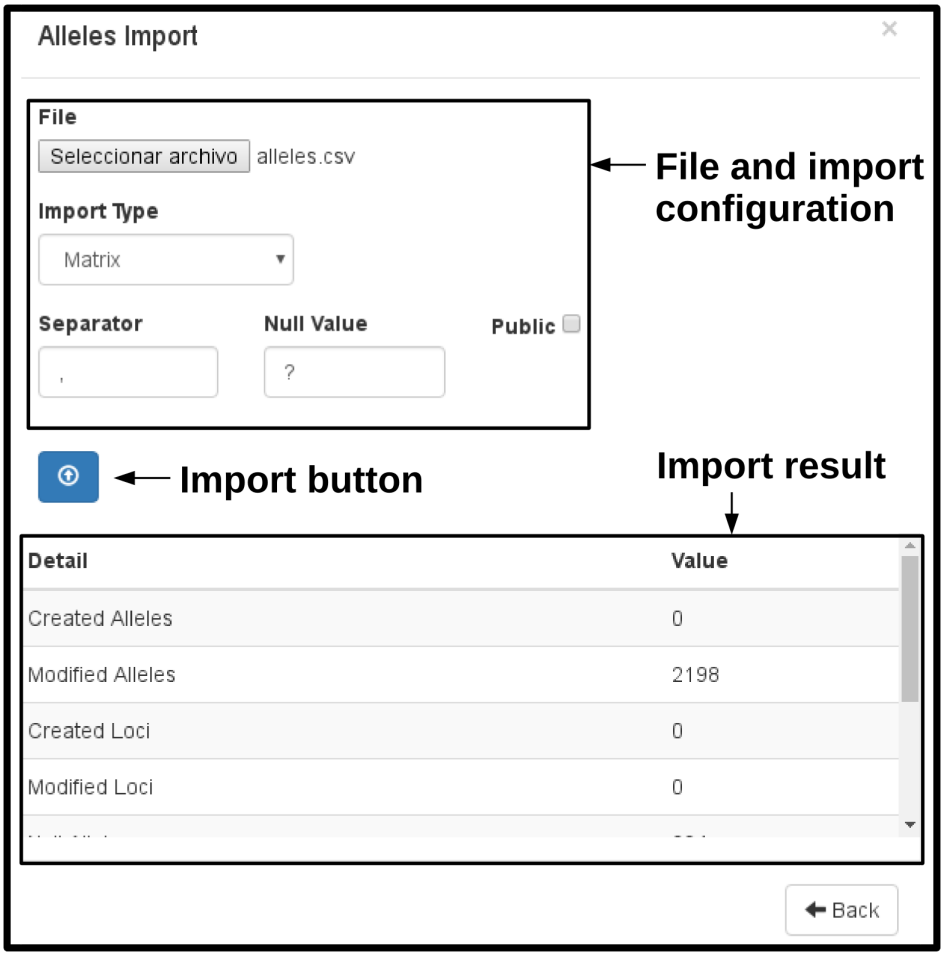

Figure 9. Import screen of loci and alleles.

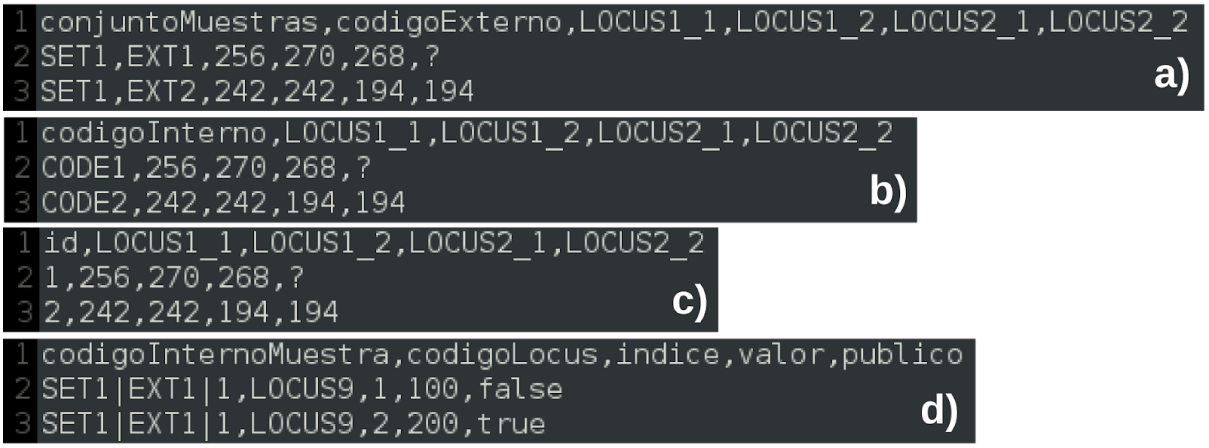

Figure 10. Sample file for the import of alleles and loci. The columns are delimited by comma (,). In a), b) and c) the file type is Matrix, the first line is the header, which indicates the name of the loci to be created or updated. The second and third rows are sample data of the alleles of two samples. Columns used to identify the samples: a) conjuntoMuestras y codigoExterno, b) codigoInterno y c) id. d) File type List, the first line is the header, which indicates the name of the attributes to be imported. The second and third rows are sample data of the alleles of two samples.

Examples

In the *examples* folder of the platform there are sample import files, the separator character is the comma (,). Within the examples folder there is a folder by language (*en* and *es*), and within them the following folders:

- full: example with all fields (descriptive) of the samples
    - samples.csv: sample import file (162 samples).
    - alleles.csv: alleles import file, with matrix format (2198 alleles not null and 394 null).
- Rueda_et_al_2013: example with data from the work of Rueda et al., 2013.
    - samples.csv: sample import file.
    - alleles.csv: alleles import file, with matrix format.
- Kamvar_et_al_2014: example with data from the work of Kamvar et al., 2014.
    - samples.csv: sample import file.
    - alleles.csv: alleles import file, with matrix format.
- instructive: example files mentioned in the manual.
    - samples: examples of Figure 7 and 8.
        - samples_a.csv: example with identification columns: conjuntoMuestras y codigoExterno.
        - samples_b.csv: example with identification column: codigoInterno.
        - samples_c.csv: example with identification column: id.
        - samples_attributes.csv: example of importing dynamic attributes. Identification column: id.
    - alleles: examples of Figure 10.
        - alleles_a.csv: matrix format, with identification columns: conjuntoMuestras y codigoExterno.
        - alleles_b.csv: matrix format, with identification column: codigoInterno.

- alleles_c.csv: matrix format, with identification column: id.
- alleles_d.csv: list format.

Genotype Analysis

This section explains how to create a Genotype Analysis and execute the analysis procedures. TEGA has an entity called Genotype Analysis, aimed at the management of data linked to the execution of genotype analysis procedures.

Samples Sets

To create a Genotype Analysis, it is first necessary to assemble one or more Samples Sets that will group the Samples to be studied according to the necessary criteria, for example samples from different regions. It is possible to create a new set of samples from the query view (*Analysis / Sample Sets*) and it is necessary to indicate a unique code and the samples that will integrate it. Once the code has been completed and the samples selected, the Save button must be pressed (Figure 11).

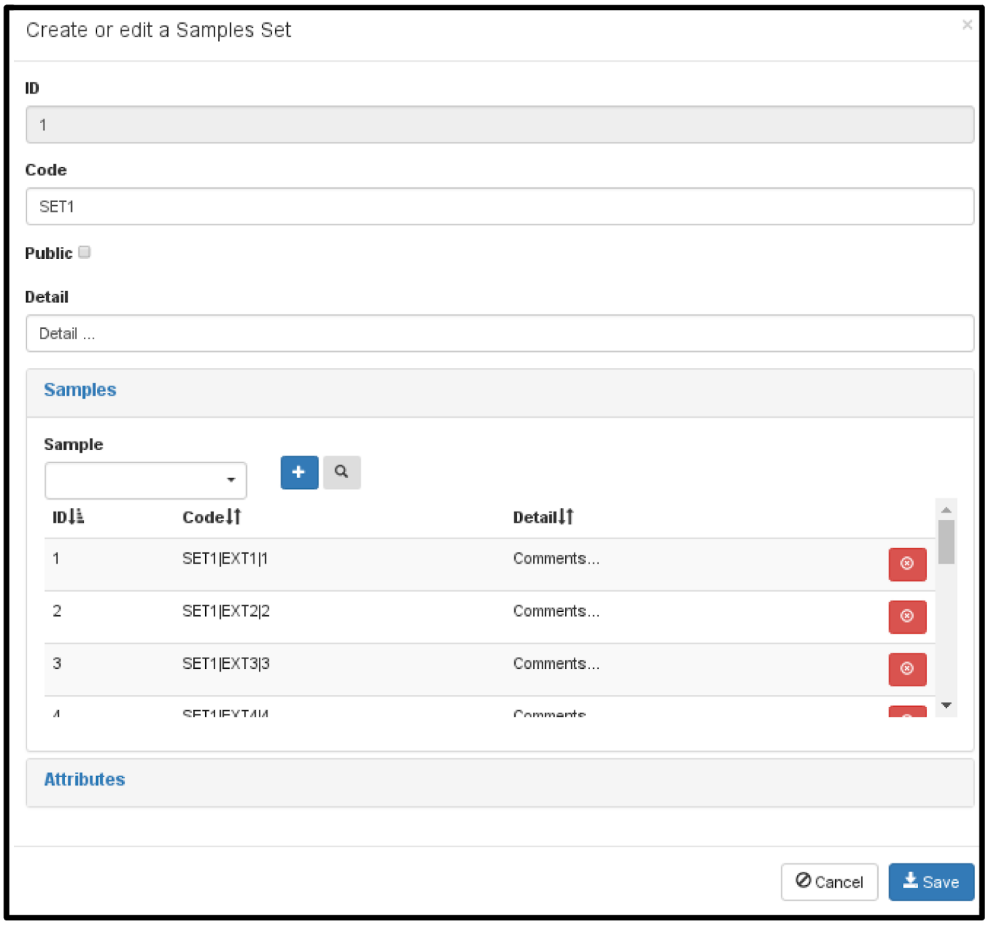

Figure 11. Sample Sets edit view.

Genotype analysis

After creating the sample sets, a new Genotype Analysis must be created (Figures 12.a and 12.b), where the Project, the Samples Sets and Loci to be studied will be recorded. It is also possible to individually or massively select the Loci and the Samples Sets. Once the analysis is created, it is possible to visualize the alleles of each Sample and Loci in a matrix form by pressing the button titled "Alleles" in the query view (Figure 12.c).
On the other hand it is possible to execute the analysis procedures from the query view (Figure 12.d), by pressing the button 

. When doing so, the procedure execution screen will appear, which contains a tab for each procedure and in each one the corresponding parameters. Pressing the Submit button will start the execution of the procedure of the selected tab, with the indicated parameters. During the execution of the procedure, the analysis can not be modified or eliminated. It will also be possible to see the execution log from the button 

in the query view.
At the end of the execution, the resulting files and the parameters used can be viewed in the Procedures Results section at the edit or display views of the genotypes analysis (Figure 12.e). In this section a drop-down item will appear for each execution made in the analysis.

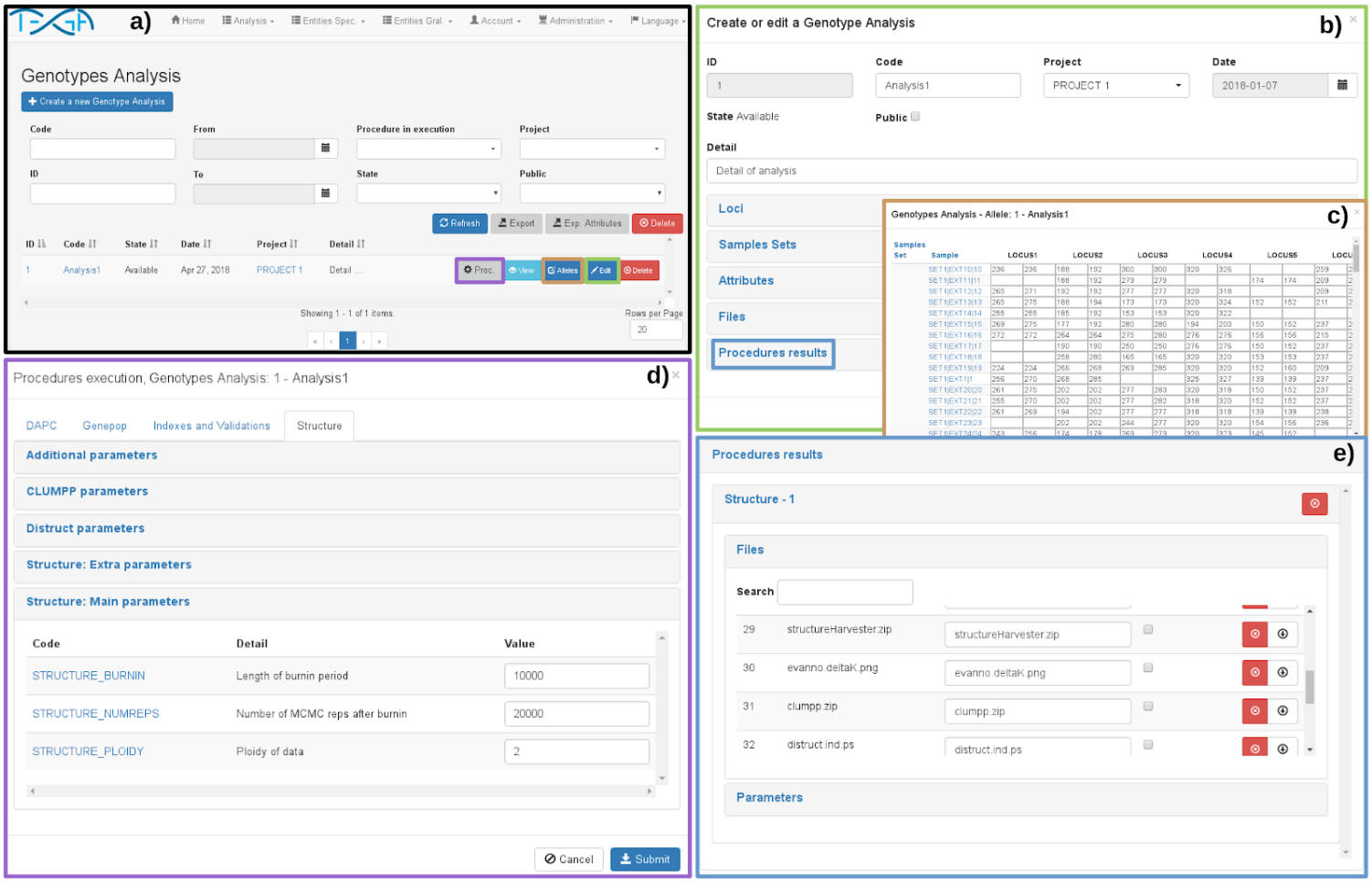

Figure 12. Genotype analysis views: a) query view, b) edit view, c) analysis allele editing view, d) analysis procedures execution view, e) procedures results section of the edit view of genotypes analysis, there are visualized the resulting files and parameters of the execution of the procedures.

Analysis Procedures

TEGA has a module for the management of the analysis procedures where is possible to create, edit and eliminate the procedures.

Procedures

The procedure query view is found in *Analysis / Procedures*. When creating a procedure it is necessary to indicate (Figure 13):

- Código: Code: must be unique and will be used to name the execution folders.
- Nombre: Name: used in the execution screen of procedures in the Genotype Analysis.
- Detail: description of the procedure.
- Execution command: command that will be used to execute the procedure. Example: ./script.sh

- Command to show the execution log: this command will be used when the user presses the button  in the Genotype Analysis query. Example: ./script_log.sh
- Public: indicates whether the procedure can be seen by users with the Invited role.
- Exports: this section indicates the files that TEGA will generate when executing the procedure:
    - Genotypes (TEGA): File of genotypes with the internal format of TEGA. The name of the generated file will be: genotipos.tsv. The columns are separated by tabs. Contains the columns:
        - conjunto_id: samples set ID.
        - conjunto_cod: samples set code.
        - conjunto_det: samples set detail.
        - muestra_id: sample ID.
        - muestra_cod_int: sample internal code.
        - Then it has a column for each Locus. In each column, the values of the alleles are indicated, separating them with the value of the TEGA parameter EXP_PROC_SEP_ALELOS (default "/"). The order of the alleles is indicated in the Index field. The value of the null alleles is the one indicated in the TEGA parameter EXP_PROC_VALOR_NULO (default "-9").
    - Genotypes (Structure): Genotypes file with the Structure format. The name of the generated file will be: genotipos.str.
    - Genotypes (Genepop): Genotypes file with the Genepop [10] format. The name of the generated file will be: genotipos.gen.
    - Samples: File with the static fields of the samples. The name of the file will be: muestras.tsv.
    - Attrib. extras of Samples: File with the extra attributes of the samples. The name of the file will be: muestras_atributos.tsv. The ID column corresponds to the sample ID.
    - Samples Sets: File with the static fields of the Samples Sets. The name of the file will be: conjuntos_muestras.tsv.
    - Attrib. extras of Samples Sets: File with the extra attributes of the sample sets. The name of the file will be: conjuntos_muestras_atributos.tsv. The ID column corresponds to the ID of the sample set.
    - Loci: File with the static fields of Loci. The name of the file will be: loci.tsv.
    - Attrib. Loci extras: File with the extra attributes of the loci. The name of the file will be: loci_atributos.tsv. The ID column corresponds to the locus ID.
    - Quantities: file with the number of occurrences of the exported entities (Samples, Loci, Alleles and Samples Sets).

  In the *ejemplos/[language]/proc* folder there are sample files of each of these files. In the case of dynamic attribute files, each has five attributes, one for each type (character, numeric, integer, logical, and date).
- Files: the files necessary for the execution of the procedure. The files must be loaded after creating the procedure.

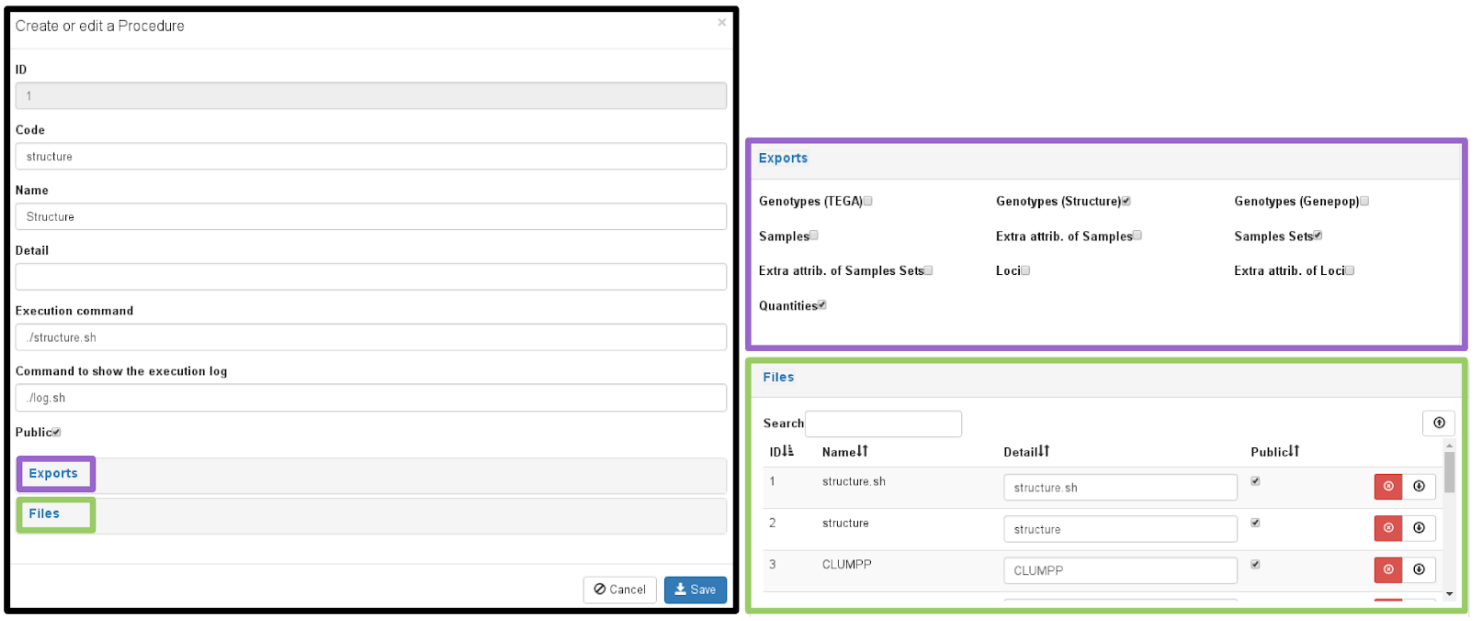

Figure 13. Edit view of the analysis procedures.

Parameters and Procedures Categories

The Procedure Parameters allow the user to configure the execution of the procedure, for example indicate the number of repetitions, the range of clusters to be detected, the statistical method to be used, etc. The query view of these parameters is found in *Analysis / Procedure Parameters.*

When creating a Parameter it is necessary to indicate (Figure 14):

- Code: the code must be unique.
- Detail: description of the parameter.
- Type: Data type of the parameter, the available types are: Integer, Numeric, Character, Logical and Date.
- Procedure: procedure to which the parameter will be linked.
- Category: Parameter category. The categories allow grouping the parameters of a procedure into sections to facilitate its search when executing the procedure. The query view of categories is found in *Analysis / Categories of Parameters*.
- Not Editable: A non-editable parameter cannot be modified at the time of execution of the procedure.
- Value: Default value of the parameter.

Figure 14. Edit view of the Procedure Parameters.

Steps of the procedures execution

In this section the internal steps carried out by TEGA for the execution of the procedures are mentioned (the execution is done from the Genotype Analysis query):

1. The status of the Genotype Analysis is changed from Available to Executing.
2. A new occurrence of the Execution entity is created, which will be linked to the parameters of the procedures used and the resulting files.
3. The destination folder is created in which the input files and the resulting files will be stored. The address of this folder will be stored in the Execution: [FILES]/analisis_genotipos/[ID_ANALYSIS]/[CODE_PROC]_[ID_EXECUTION]
   Where:
   - FILES: is the folder indicated in the TEGA parameter DIRECTORIO_ENTIDADES.
   - ID_ANALYSIS: the ID of the selected Genotype Analysis.
   - CODE_PROC: code of the procedure to execute.
   - ID_EXECUTION: Execution ID.
4. The input files indicated in the Export section of the procedure to be executed are generated.
5. The file is generated with the procedure parameters selected by the user. This file contains both editable and non-editable parameters. The name of the generated file is: parametros.tsv. It has the columns:

- categoria: parameter category.
- parametro: parameter code..
- tipo: parameter type. Possible values: CARACTER (character), ENTERO (integer), NUMERICO (numeric), LOGICO (logic) y FECHA (date).
- valor: parameter value.

6. The command for the execution of the procedures is executed, indicated in the TEGA COMANDO_PROCEDIMIENTOS parameter (by default ./proc/ejecutar_procedimiento.sh). This script receives by parameter:
   - Command to execute the procedure (Execution Command).
   - User of the operating system to be employed for the execution (parameter of TEGA USUARIO_PROCEDIMIENTOS). In case it is empty, the current user will be used.
   - Path of the folder where the procedure files are stored.
   - Destination path where the input files are located.

   The script will perform the following steps:
   a. Create a temporary folder and copy the procedure files to that folder.
   b. In case of indicating an execution user, read and write permission will be assigned to the user in the temporary and destination folders. The execution permission will also be assigned to the files of the temporary folder created in a.
   c. The current directory is changed to the one created in a.
   d. The indicated command is executed. The following parameters are passed to the execution command:
      i. Genotype analysis ID.
      ii. Absolute path to the temporary folder created in a.
      iii. Absolute path to the destination folder.
   e. The temporary folder created in a is deleted.
6. At the end of the execution of the procedure, the following steps will be carried out within the same transaction:
   - Linking the files in the destination folder to the Execution.
   - Change of status of the Genotype analysis, from Executing to Available.

Procedures available

The first version of TEGA has four modest implementations of the usual analysis procedures in population genetics studies.

Structure

This procedure seeks to determine the structure of the population using the Bayesian grouping method implemented in Structure. The procedure allows to indicate the parameters of each programs used. It is also possible to indicate the range of K (number of clusters) to be studied and repetitions by K. The procedure has the following steps:

1. Execution of Structure for the range of K and repetitions indicated. The parallel library is used to parallelize the execution

2. Execution of STRUCTURE HARVESTER [11] to determine the most suitable K. The one with the highest Delta K is selected.
3. Execution of CLUMPP [12] to integrate the repetitions of the selected K.
4. Execution of Distruct [13] to graph the assignation of clusters, both to individuals and populations.

The files resulting from the procedure are:

a. archivos_exp.zip: contains the input files: cantidades.tsv, conjuntos_muestras.tsv y parametros.tsv).
b. structure.zip: contains the input files and logs of the execution of Structure.
c. structure_log.zip: contains the log files of each execution of Structure.
d. structure_out.zip: contains the output files of Structure. This file can be used in other platforms such as CLUMPAK [14].
e. structureHarvester.zip: contains output files of STRUCTURE HARVESTER.
f. clumpp.zip: contains the CLUMPP input and output files.
g. distruct.zip: contains the input and output files of DISTRUCT.
h. evanno.txt: includes the table with the values of Delta K determined with STRUCTURE HARVESTER.
i. evanno.mk.txt: has the K selected by the platform (largest Delta K).
j. distruct.ind.ps: file in PostScript format generated with Distruct where the assignment of each group to each individual is visualized (generated from the selected K and the integration of the K repetitions made with CLUMPP).
k. distruct.pop.ps: file in PostScript format generated with Distruct where the assignment of each group to each samples set is visualized (generated from the selected K and the integration of the K repetitions made with CLUMPP).
l. ejecucion.log: log of the execution of the procedure.

Genepop

This procedure performs the execution of Genepop. It is possible to indicate the options to be executed and the parameters to be used. The resulting files are:

- genotipos.gen: genotype file in the Genepop format.
- settings.txt: Genepop input file that indicates the parameters and menu options to execute.
- parametros.tsv: procedure parameters.
- genepop.log: log of procedure execution.
- genotipos.gen.*: genepop output files corresponding to the menu options indicated.

DAPC

This procedure seeks to determine the structure of the population using the multivariate method DAPC (Discriminant analysis of principal components) [15]. The procedure has the following steps:

1. Determination of K (number of clusters): the function *find.clusters* of adegenet [16] package is used. It is possible to indicate the quantity of principal components (PC) to be used or the percentage of the variance; the statistical method to be used to determine the best K and the selection criterion of K, among other parameters. It is

possible to deactivate this step and perform the following steps with the groups determined by the samples sets.

2. Determination of PCs: the number of PCs to be used in DAPC is determined, this is done with a cross validation, using the *xvalDapc* function of adegenet package. It is possible to indicate the number of repetitions, the percentage of the training set, among other parameters. The procedure determines the number of PCs in two steps, first it executes *xvalDapc* with the range of PCs by default, then it executes *xvalDapc* for a range of PCs more bounded with center in the number of PCs selected in the first step. The number of PCs selected is the one that reports the least Mean Square Error.
3. Scatter and variable contribution graphs: scatter and variable contribution graphs are made, using the *scatter* and *loadingplot* function of adegenet package. In addition, the posterior membership probability at the sample level and the samples set is plotted.

Resulting files:

- archivos_exp_logs.zip: contains the input files and logs of the procedure (dapc.log, genotipos.tsv, parametros.tsv).
- find_clusters.zip: contains the output files of K determination.
    - pc_var.*: image and table of the number of PCs retained by percentage of the variance used.
    - k_stat.*: image and table of the value of the method used to determine K, by K.
- cv_dapc.zip: output files of PC determination (N indicates whether it is the first or second cross-validation):
    - cvN.cvr.csv: table with the result of the cross validation.
    - cvN.rmsenpp.csv: root mean squared error by number of PCs of PCA.
    - cvN.png: graph of ratio of hits by number of axes of PCA.
- dapc.zip: output files of the DAPC execution:
    - pca_scatter.png: scatter plot.
    - pmp_indplot.png: graph of posterior membership probability at the sample level.
    - pmp_boxplot.png: graph of posterior membership probability at the samples set level.
    - loadingplot_*.png: graphs of contribution of variables. Contribution of each allele to the main components. The number in the file name indicates the component.
    - posterior.csv: table with the posterior membership probability of each sample to each clusters.
    - pmp.csv: transformed table with the posterior membership probability of each sample to each clusters.
    - var.contr.csv: table with the contribution of each variable to each principal component.
- k_stat.png: image and table of the value of the method used to determine K, by K.
- pmp_indplot.png: graph of posterior membership probability at the sample level.
- pmp_boxplot.png: graph of posterior membership probability at the samples set level.

- resultados.csv: table with results of the methods (K selected in step a, number of PCs selected in step b, etc.).

Indexes and Validations

This procedure generates several descriptive and graphic indices. There are parameters to indicate the tables and graphs that will be generated. The generated files are:

- archivos_exp_logs.zip: input files and procedure logs.
- alelos.zip: contains tables and graphs of frequency and average of alleles per set of samples. Poppr functions are used for calculations.
- allelic_richness.csv: table with allelic wealth by Loci and Samples Sets. Only available for diploid loci. The *allelic.richness* function of hierfstat is used.
- basic_stat.csv: table with basic indexes (Fst, Fis, Ho, Hs, etc.). The *basic.stats* function of hierfstat is used.
- missing.*: table and graph of missing alleles. The *info_table* function of poppr is used.
- mlg.*: table and graph of the quantity of multilocus genotypes (MLG). The *mlg.table* function of poppr is used.
- mlg_rarefaction.png: rarefaction curve calculated with MLG. It uses the *rarecurve* function of vegan [17].
- ploidy.*: table and graph of ploidy by locus and sample. The *info_table* function of poppr is used.
- genotype_curve.*: table and graph of the curve of accumulation of genotypes. The *genotype_curve* function of poppr is used.
- genotypic_diversity.*: table and chart with diversity indexes. The *poppr* function of poppr is used.
- HW_*: tables with Hardy-Weinberg equilibrium test. The *hw.test* function of pegas [18] is used.
- locus_table_*: tables with statistics at the loci level. The *locus_table* function of poppr is used.
- dist_pop_method_*: distance between samples sets calculated with the method indicated by the user (see procedure parameters). The *dist.genpop* function of adegenet is used.

Examples

TEGA has sample files to test the procedures. These files correspond to the publications of Rueda et al. 2013 and Kamvar et al. 2014, although the examples focus on the *Structure* and *DAPC* processes, they can also be used to test *Genepop* and *Indexes and Validations*.

Structure: Rueda et al. 2013

In the work of Rueda et al. 2013, the seasonal variation in the population structure of the sábalos (*Prochilodus lineatus*) captured in the Uruguay River, near the city of Gualeguaychú (Entre ríos, Argentina), was analyzed. In this work Structure was used to analyze the population structure. In the folder *ejemplos/en/Rueda_et_al_2013* are the samples and alleles files.
The steps to replicate the analysis with TEGA are:

1. Import the samples.csv file: the import is done from the sample query (see the section Importing Samples), the comma (,) must be indicated as a separator, and Samples in *Data to be Imported*. When the samples are imported, the Sample Sets will be created automatically.
2. Import the file alleles.csv: the import is made from the Alleles query view (see section Importing of Loci and Alleles), you must indicate the type of import Matrix, the comma (,) as a separator and "?" as null value.
3. Create a new Genotype Analysis (see Genotyping Analysis section): in the creation of the analysis select the Loci: *PL3, PL14, PL25, PL34, PL35, PL64, PL119, PL139* and *PL216*; and the Sample Sets: *RUEDA13|WINTER, RUEDA13|SPRING* and *RUEDA13|FALL*.
4. Execute the Structure procedure (see Genotype Analysis section): the parameters must be indicated (the rest by default):
    - STRUCTURE_K_HASTA: 10
    - STRUCTURE_REPLICAS: 10
    - STRUCTURE_BURNIN: 150000
    - STRUCTURE_NUMREPS: 300000

As a result, it can be seen that the optimal number of clusters (K = 3) and the assignment of each to the samples corresponds to the work of Rueda et al. 2013 (Figure 15).

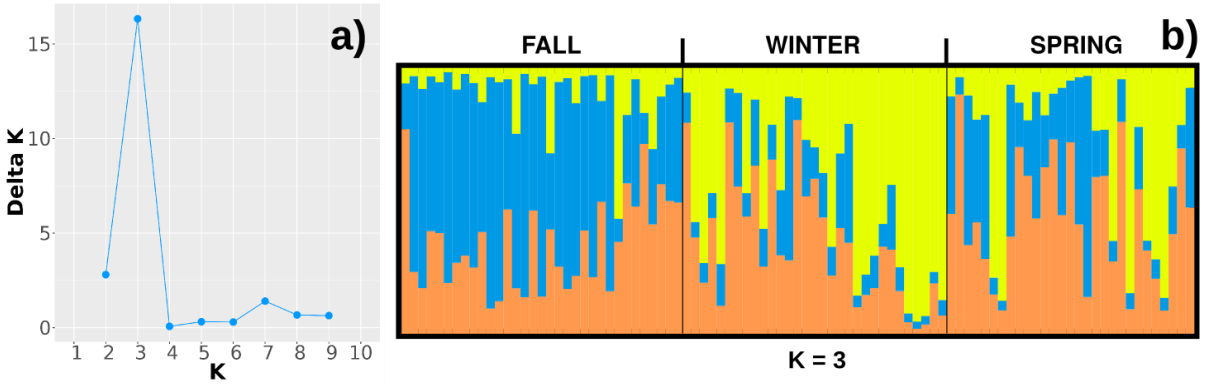

Figure 15. Example of the Structure procedure with data from Rueda et al. 2013. a) Plots of mean Delta K. b) Assignment probabilities for each individual (vertical bars) into one of three clusters.

DAPC: Kamvar et al. 2014

The Kamvar et al. 2014 data were used in the Kamvar et al.. 2015 [19] work, they analyzed the population structure of Phytophthora ramorum in the Oregon and California forests, using DAPC. In the folder *ejemplos/en/Kamvar_et_al_2014* are the samples and alleles files. The steps to replicate the analysis with TEGA are:

1. Import the samples.csv file: the import is done from the sample query (see the section Importing Samples), the comma (,) must be indicated as a separator, and Samples in *Data to be Imported*. When the samples are imported, the Sample Sets will be created automatically.
2. Import the file alleles.csv: the import is made from the Alleles query view (see section Importing of Loci and Alleles), you must indicate the type of import Matrix, the comma (,) as a separator and "?" as null value.

3. Create a new Genotype Analysis (see Genotyping Analysis section): in the creation of the analysis select the Loci: *PrMS6A1, Pr9C3A1, PrMS39A1, PrMS45A1* and *PrMS43A1*; and the Sample Sets: *JHallCr, NFChetHigh, Coast, HunterCr, Winchuck, ChetcoMain, PistolRSF, Nursery_CA* and *Nursery_OR*.
4. Execute the Structure procedure (see Genotype Analysis section): the parameters must be indicated (the rest by default):
    - DAPC_FC_EJECUTAR: NO.
    - DAPC_CV_REP: 250
    - DAPC_SP_P_BOX_NCOL: 3
    - DAPC_SP_P_IND_NCOL: 3
    - DAPC_SP_POSI_PCA: topright
    - DAPC_SP_P_BOX_XLEN: 11
    - DAPC_SP_P_BOX_YLEN: 11

As a result, it can be seen that HunterCr differs from the rest of the Sample Sets (Figure 16.a) and that the assignment of DAPC to each sample would correspond to the Samples Sets (without considering PistolRSF) (Figure 16.b). In addition, the 493 allele at PrMS43A1 locus would contribute the most to that difference between sets (Figure 16.c and 16.d). These results would be similar to those of Kamvar et al. 2015.

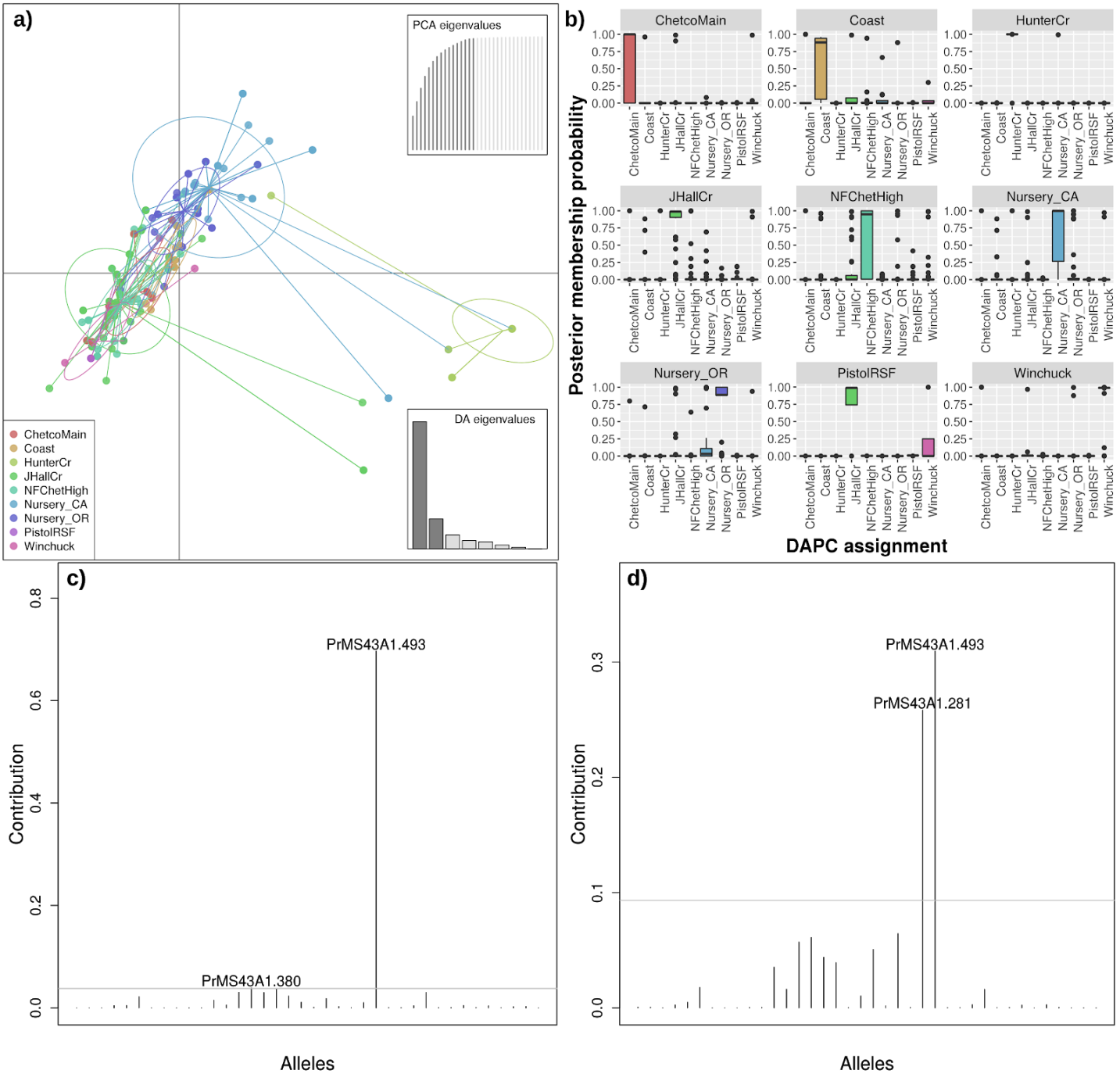

Figure 16. Results of the DAPC procedure with the data of Kamvar et al. 2014. a) Scatterplot (file pca_scatter.png). b) Boxplot made from the posterior membership probability of each sample to each Samples Set (file pmp_boxplot_assig.png). Contribution of the alleles to the first (c) and second (d) principal components of the analysis (files loadingplot_1.png and loadingplot_2.png). The greatest contribution is made by the 493 allele at PrMS43A1 locus.

Users

This section explains how user management is performed in TEGA.
As TEGA is focused to be used during the research period, in which the data is managed and analyzed, it is necessary to ensure the privacy of the data until the moment of its publication. For this reason there are different user roles:

- Anonymous: Has access to the public data of the Samples. Has permission to view and export the data, can not modify or delete them.
- Invited: Has access to the public data of the Samples, Alleles, Locus, Procedures and Genotype Analysis. Has permission to view and export the data, can not modify or delete them.
- Researcher: Has access to public and private data of the platform, can perform CRUDE actions and execute genotype analysis.
- Administrator: Has the same privileges as the Investigator role and also has access to the management of users and configurations of the platform.
- Implementer: Role for the implementers of the platform in the institution.

The Anonymous and Invited roles are intended for people outside the research group. When a person registers in the platform, he is assigned the role indicated in the TEGA parameter *ROL_USUARIO_DEFECTO*, which can be Anonymous (ANONIMO) or Invited (INVITADO, default value).
The roles of Administrator and Researcher would be used only by the research group.
Users with the Administrator role can access to the User query view (*Administration / User Management*) (Figure 17). There they can:

- Consult registered users in the platform, analogously to other views.
- Create, Edit and Delete users. While an administrator can create a new user, the recommended flow of user creation is:
    a. The person interested in accessing to the platform must create their user on the home screen (link *Register a new account*, or the menu option *Account / Register*).
    b. The platform will send an email to the indicated email address and the person must confirm the registration (by clicking on the link sent in the email). In this instance, the user will have the role of Anonymous.
    c. The administrator must assign the desired role to the created user. This can be done from the user edit view.
    d. Once the role has been modified, if the person had a session started on the platform, he must close the session and start it again to see the changes.
- Both from the query view and from the edit view, it is possible to change the user's status from On to Off and vice versa. A disabled user cannot log into the platform.

Figure 17. User edit view..

Data accessibility

This section explains how to change the status (from private to public and vice versa) of the data massively.
After publishing the results of an investigation, the users of the platform (with the role of Researcher or Administrator) can change the status of the project linked to the publication, passing it to the public (Figure 18). This will cause the Samples, Alleles, Loci and Genotypes Analysis linked to the project to be accessed by others users outside the institution (users with the Anonymous or Invited roles).

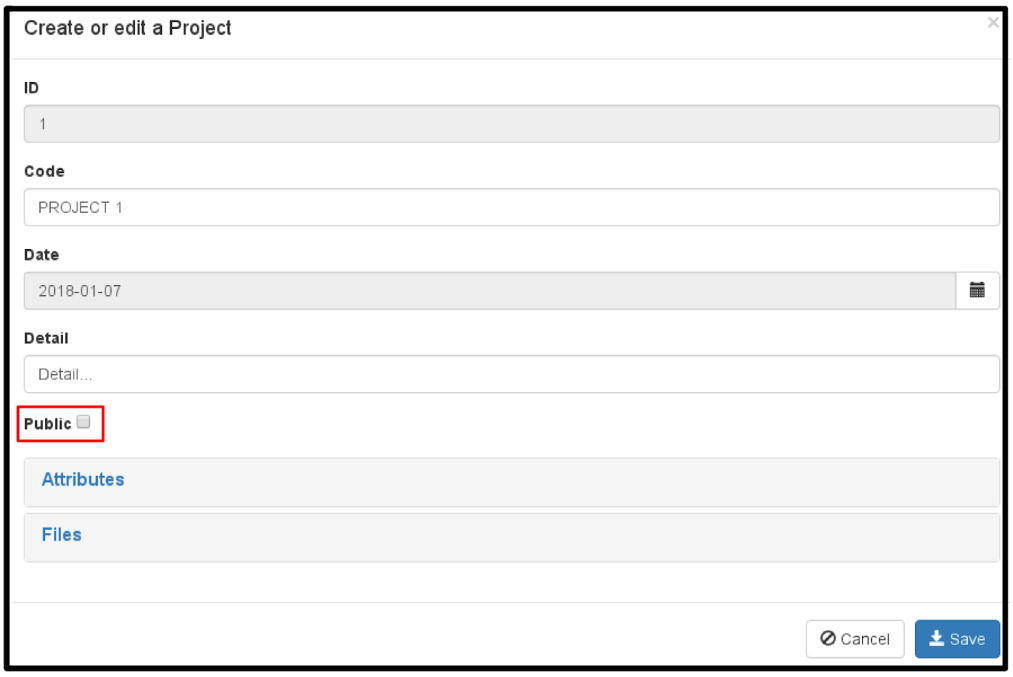

Figure 18. Project edit view.

TEGA parameters

This section explains how to manage the TEGA parameters. These parameters allow modifying certain configurations of the platform, for example, the content of the home screen, the logo, the character used as a separator in exports, etc. The parameters can be modified from the parameter query view (*Administration / TEGA Parameters*), which can only be accessed by users with the Administrator or Implementer role (Figure 19).
Modifying or deleting a parameter could cause fail in the platform, so we recommend modifying them only when the effect is well understood. In turn, we also recommend having a testing environment to test different configurations.
The parameters have the following fields:

- Code: the code of the parameter, must be unique.
- Detail: detail of the use of the parameter in the platform.
- Type: data type of the parameter, the available types are: Integer, Numeric, Character, Logical and Date.
- Public: indicates if an anonymous user can access the parameter. Mainly used for parameters related to the home page.
- Value: value of the parameter, it can only be empty if the parameter type is Character.

After modifying a parameter, to visualize the change it is necessary to reload TEGA in the browser (shift + F5).

Figure 19. TEGA parameter edit view.

Backups

TEGA has a basic procedure for backups generation. The configuration of the procedure is mentioned in the section *Installation and Configuration*.

Generation

It is possible to generate a backup from the menu option *Administration / Backup*, but only users with the role of Administrator or Implementer can access this option. On the other hand, it is also possible to execute the procedure from a task scheduler, such as cron. For this it is necessary to execute the procedure in the following way:

[DIR_TEGA]/proc/backups.sh CRON [DIR_TEGA]/config/backup.config.sh

Where:

- DIR_TEGA: path to TEGA folder
- CRON: parameter to tell the procedure that the execution is automatic.

The backup includes:

- All files located in the TEGA folder (backup copies are excluded).
- Copy of the database.

The files to be included in the backup are compressed and then encrypted with *pgp*.

Restore

The restoration of a backup is possible by executing the script:
[DIR_TEGA]/scripts/restore.sh [BACKUP] [ENCRYPTION_KEY] [HOST_DB] [USER_DB] [PASS_DB] [DB] [ENCODING] [LC_COLLATE] [LC_CTYPE]
Where:

- DIR_TEGA: path to TEGA folder
- BACKUP: path and file name of the backup file.
- ENCRYPTION_KEY: encryption key of the file (indicated in the backup configuration file).
- HOST_DB: host of the database engine.
- USER_DB: user with whom the database will be created and the data will be imported.
- PASS_DB: password of the DB user.
- DB: name of the database to create.
- ENCODING, LC_COLLATE y LC_CTYPE: parameters of the CREATE DATABASE command of PostgreSQL.

This procedure performs the following steps:

1. Decrypt the backup.
2. Unzip the backup in the current directory.
3. If database parameters are indicated (HOST_DB, USER_DB, etc.):
    a. Remove, if it exists, the indicated database.
    b. Create a new database with the indicated name.
    c. Imports the data in the database.

In case of using a user or database other than the TEGA configuration, it is necessary to manually update the TEGA configuration (*config/*.yml* files). It is then possible to run TEGA as indicated in the *Installation and Configuration* section. You must ensure that the libraries and tools mentioned in the *Requirements* section are installed.

Development

In this section, the general aspects of the development of TEGA are mentioned.

General aspects

For the development of TEGA the project generator JHipster (http://www.jhipster.tech) was used. In the Back-End the JAVA language was used with the Spring Boot framework (https://projects.spring.io/spring-boot), and the PostgreSQL database engine (https://www.postgresql.org) . While in the Front-End JavaScript was used with AngularJS (https://angularjs.org) and Bootstrap (http://getbootstrap.com).
All the libraries used can be consulted in the TEGA repository (https://github.com/darioelias/TEGA).

Sources installation

Steps to install the TEGA sources:

1. Install the requirements of the *Requirements* section
2. Install JAVA 8 Development Kit (JDK). OpenJDK 8 installation guide: http://openjdk.java.net/install/
3. Install git v2.1.4. Git installation guide: https://git-scm.com/download/linux
4. Install MAVEN (https://maven.apache.org)
5. Install Node.js v4.7.0 (https://nodejs.org/en)
6. Update NPM. Command: *npm install -g npm*
7. Install bower y gulp-cli. Command: *npm install -g bower gulp-cli*
8. Clone the github TEGA repository. Command: *git clone git://github.com/darioelias/TEGA.git*
9. Copy the *config_ejemplo* folder and rename it by *config*
10. Create a database and configure TEGA. Steps 3 to 10 of the section *Installation and Configuration*. In addition to configuring the *application-prod.yml* file, you must configure the *application-dev.yml* file.
11. Running in development: the *dev_comp.sh* script allows you to run maven with the development configuration.
12. Execution in production: the *prod_comp.sh* script allows you to run maven with the configuration to build and package the platform (generating the WAR file in the *target* folder). For more information consult the JHipster documentation or the *jhipster.readme.txt* file.

Entity-relationship diagram

The entity-relationship diagram is only illustrative (Figure 20) and can be viewed with the MySQL Workbench (https://www.mysql.com/products/community). The model file is located in the *model* folder of the git repository of TEGA.

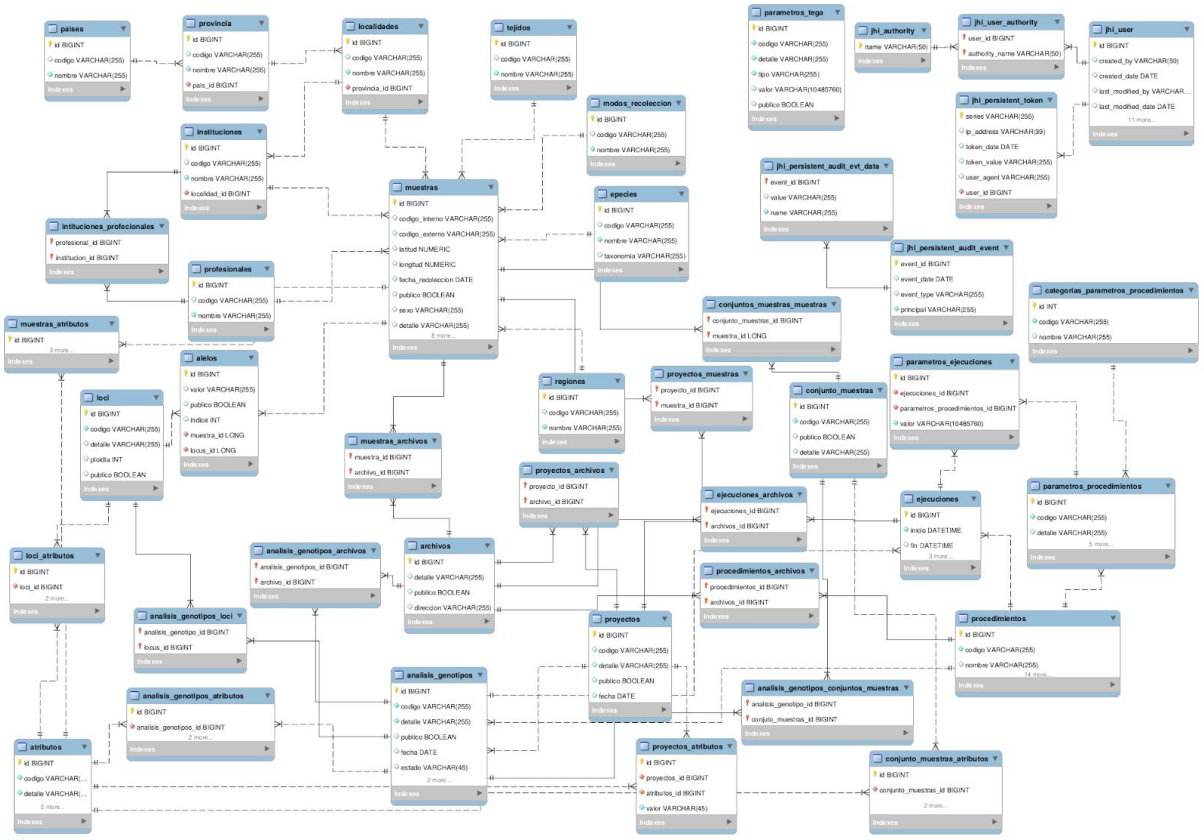

Figure 20. Diagram of entity-relationship of TEGA displayed with MySQL Workbench. The diagram is only illustrative.

Bibliography

1. Charlesworth B, Charlesworth D. Population genetics from 1966 to 2016. Heredity. 2017;118: 2–9. doi:10.1038/hdy.2016.55
2. Pritchard JK, Stephens M, Donelly P. Inference of population structure using multilocus genotype data. Genetics. 1998;155: 945–959.
3. Rueda EC, Carriquiriborde P, Monzón AM, Somoza GM, Ortí G. Seasonal variation in genetic population structure of sábalo (Prochilodus lineatus) in the Lower Uruguay River. Genetica. 2013;141: 401–407. doi:10.1007/s10709-013-9739-0
4. Kamvar ZN, Larsen MM, Kanaskie AM, Hansen EM, Grünwald NJ. Sudden_Oak_Death_in_Oregon_Forests: Spatial and temporal population dynamics of the sudden oak death epidemic in Oregon forests. Zenodo. 2014.
5. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. SCIENTIFIC STANDARDS. Promoting an open research culture. Sci. 2015;348: 1422–1425. doi:10.1126/science.aab2374
6. Fecher B, Friesike S. Open Science: One Term, Five Schools of Thought. Opening Science. Springer International Publishing; 2012. pp. 17–47. doi:10.1007/978-3-319-00026-8_2
7. Tange, O. (2011). Gnu parallel-the command-line power tool. The USENIX Magazine, 36(1), 42-47.
8. Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ, 2, e281.
9. Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Resources, 5(1), 184-186.
10. Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular ecology resources, 8(1), 103-106.
11. Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources, 4(2), 359-361.
12. Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23(14), 1801-1806.
13. Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Resources, 4(1), 137-138.
14. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular ecology resources, 15(5), 1179-1191.
15. Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC genetics, 11(1), 94.
16. Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24(11), 1403-1405.
17. Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2017). vegan:

Community Ecology Package. R package version 2.4-2. https://CRAN.R-project.org/package=vegan
18. Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics, 26(3), 419-420.
19. Kamvar, Z. N., Larsen, M. M., Kanaskie, A. M., Hansen, E. M., & Grünwald, N. J. (2015). Spatial and temporal analysis of populations of the sudden oak death pathogen in Oregon forests. Phytopathology, 105(7), 982-989.

Change history

14/05/2018: Creation of the manual. Version 1.0.0
11/01/2019: Minor text corrections. Added the VirtualBox Image section. Version 1.0.1

Annex: GNU Free Documentation License

GNU Free Documentation License
Version 1.3, 3 November 2008

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a

licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain

ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "publisher" means any person or entity that distributes copies of the Document to the public.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
D. Preserve all the copyright notices of the Document.
E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
H. Include an unaltered copy of this License.
I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document,
unaltered in their text and in their titles. Section numbers
or the equivalent are not considered part of the section titles.
M. Delete any section Entitled "Endorsements". Such a section
may not be included in the Modified Version.
N. Do not retitle any existing section to be Entitled "Endorsements"
or to conflict in title with any Invariant Section.
O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or
appendices that qualify as Secondary Sections and contain no material
copied from the Document, you may at your option designate some or all
of these sections as invariant. To do this, add their titles to the
list of Invariant Sections in the Modified Version's license notice.
These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains
nothing but endorsements of your Modified Version by various
parties--for example, statements of peer review or that the text has
been approved by an organization as the authoritative definition of a
standard.

You may add a passage of up to five words as a Front-Cover Text, and a
passage of up to 25 words as a Back-Cover Text, to the end of the list
of Cover Texts in the Modified Version. Only one passage of
Front-Cover Text and one of Back-Cover Text may be added by (or
through arrangements made by) any one entity. If the Document already
includes a cover text for the same cover, previously added by you or
by arrangement made by the same entity you are acting on behalf of,
you may not add another; but you may replace the old one, on explicit
permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License
give permission to use their names for publicity for or to assert or
imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this
License, under the terms defined in section 4 above for modified
versions, provided that you include in the combination all of the
Invariant Sections of all of the original documents, unmodified, and
list them all as Invariant Sections of your combined work in its
license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on

covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after

your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See https://www.gnu.org/licenses/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in

part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright (c) YEAR YOUR NAME.
> Permission is granted to copy, distribute and/or modify this document
> under the terms of the GNU Free Documentation License, Version 1.3
> or any later version published by the Free Software Foundation;
> with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
> A copy of the license is included in the section entitled "GNU
> Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the
> Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.