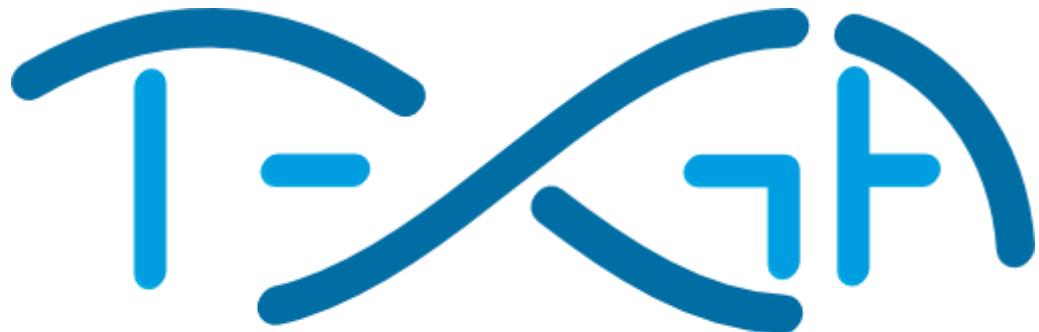


# **TEGA: Tools for Evolutionary and Genetic Analysis**

Versión 1.0.0

Dario E. Elias & Eva C. Rueda

2018



TOOLS for EVOLUTIONARY and GENETIC ANALYSIS

Versión del documento: 1.0.1

Copyright (C) 2018 Dario E. Elias & Eva C. Rueda.

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".

# Índice

<b>Introducción</b>	<b>4</b>
Sitio WEB de TEGA	5
Licencia	5
Cómo citar a TEGA	5
<b>Instalación</b>	<b>6</b>
Imagen de VirtualBox	6
Requerimientos	6
Instalación y configuración	7
Recomendaciones	10
<b>Uso de la plataforma</b>	<b>11</b>
Opciones de menú	11
Padrones	12
Pantalla de consulta	12
Pantalla de visualización	14
Pantalla de edición	15
Mapa de Muestras	17
Atributos dinámicos	18
Importación de Datos	21
Importación de Muestras	21
Importación de valores de atributos dinámicos de las muestras	23
Importación de Loci y Alelos	24
Ejemplos	26
Análisis de genotipos	27
Conjuntos de Muestras	27
Análisis de genotipos	28
Procedimientos de Análisis	29
Padrón de Procedimientos	29
Parámetros y Categorías de Procedimientos	31
Pasos de la ejecución de Procedimientos	32
Procedimientos disponibles	33
Structure	33
Genepop	34
DAPC	34
Índices y Validaciones	36
Ejemplos	36
Structure: Rueda et al 2013	37
DAPC: Kamvar et al 2014	38

Usuarios	40
Publicación de datos	41
Parámetros de TEGA	42
Copias de Seguridad	43
Generación	43
Restauración	44
<b>Desarrollo</b>	<b>45</b>
Aspectos generales	45
Instalación de fuentes	45
Diagrama de entidad relación	46
<b>Bibliografía</b>	<b>47</b>
<b>Historial de cambios</b>	<b>49</b>
<b>Anexo: GNU Free Documentation License</b>	<b>50</b>

# Introducción

La genética de poblaciones, como parte de la biología evolutiva, se ocupa del estudio de las diferencias genéticas entre e intra poblaciones [1]. Los investigadores utilizan diariamente datos genéticos y ambientales de muestras, almacenándolos en archivos de texto u hojas de cálculo, lo que hace imposible mantener su integridad, trazabilidad y acceso.

TEGA es una plataforma *standalone* basada en WEB, desarrollada para facilitar el análisis y la gestión de datos de genética de poblaciones. Fue diseñada para ser instalada por investigadores en sus laboratorios. Los objetivos de TEGA son:

## 1) Facilitar la gestión de datos

TEGA tiene una estructura basada en entidades, cada una tiene vistas con funciones básicas para: crear, leer, actualizar y eliminar (CLAE). También es posible importar masivamente los datos de muestras y genotipos (loci y alelos). En algunas entidades, como Muestras, Proyectos y Análisis de Genotipos, es posible adjuntar archivos (por ejemplo, imágenes y documentos). Además, dada la gran cantidad de datos que se pueden vincular a las muestras, TEGA le permite al usuario crear atributos dinámicos y vincularlos a diferentes entidades. Además, se puede visualizar la posición geográfica de las muestras con OpenStreetMap (<https://www.openstreetmap.org>).

Por otro lado, TEGA posee un módulo para la gestión y ejecución de procedimientos de análisis de datos. Un usuario con el rol Investigador puede crear procedimientos y adjuntar los archivos de ejecución y configuración; indicar los datos de entrada (por ejemplo, datos de muestras y alelos) y los parámetros para la ejecución. Luego, estos procedimientos pueden ser ejecutados desde la vista de Análisis de Genotipos.

## 2) Proveer una vía de ejecución de los procedimientos de análisis

TEGA tiene una entidad llamada Análisis de Genotipos para la gestión de datos relacionados con la ejecución de procedimientos de análisis genéticos. Inicialmente, el usuario debe crear Conjuntos de Muestras que agruparán las muestras según un criterio específico (por ejemplo, sitios de muestreo o fecha de recolección). Luego, el usuario debe crear un nuevo Análisis de Genotipo, seleccionando los conjuntos de muestras, loci y el proyecto vinculado al análisis. Finalmente, es posible ejecutar los procedimientos de análisis desde la interfaz de la plataforma. Cuando se ejecuta un procedimiento, el análisis de genotipo no puede ser editado ni eliminado, y cuando finaliza su ejecución, el usuario puede acceder a los archivos resultantes desde la vista de edición del análisis. De esta forma, TEGA vincula los resultados con los datos de entrada, el procedimiento y los parámetros utilizados, facilitando la trazabilidad del análisis.

Si bien TEGA está diseñada para que los usuarios (miembros del equipo de investigación) puedan crear sus propios procedimientos, en la primera versión implementamos métodos comunes en los estudios de genética de poblaciones, como STRUCTURE [2], y proporcionamos datos empíricos para probarlos [3,4].

### **3) Brindar un medio para publicar datos, procedimientos y resultados**

TEGA tiene diferentes roles de usuario para permitir el uso privado de los datos hasta el día de su publicación. Los roles Anónimo e Invitado están destinados a personas ajenas al equipo de investigación, que tienen acceso de solo lectura a datos públicos. Los roles de Administrador e Investigador están destinados al equipo de investigación, que tiene acceso a datos públicos y privados, y puede llevar a cabo operaciones de CLAE y ejecutar procedimientos de análisis.

Cuando se publican los resultados de un proyecto, los usuarios de la plataforma (con el rol de Investigador o Administrador) pueden modificar el estado del proyecto de privado a público. Esta acción cambiará el estado de las muestras, alelos, loci y análisis de genotipos relacionados con el proyecto, permitiendo a los usuarios con el rol Anónimo o Invitado visualizarlos. De esta forma, TEGA simplifica la publicación de datos y resultados. Además, también es posible cambiar el estado de los procedimientos de análisis.

Creemos que esta característica de TEGA facilitará el acceso a datos y procedimientos, permitiendo la reproducción de los trabajos. Por ello consideramos que TEGA está alineada con las necesidades actuales de la comunidad científica reflejadas en los movimientos Open Data y Open Science [5,6].

## Sitio WEB de TEGA

El sitio WEB del proyecto TEGA es <https://github.com/darioelias/TEGA>, allí se encuentran los fuentes, documentos y versiones.

## Licencia

TEGA se encuentra licenciada con GNU AGPL-3.0 (<https://www.gnu.org/licenses/agpl-3.0.html>).

El logo de TEGA fue diseñado por Ernesto Goddio y se encuentra licenciado bajo CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0>).

## Cómo citar a TEGA

En el sitio WEB de TEGA se menciona el artículo que debe citarse.

# Instalación

En esta sección se mencionan los requerimientos para la ejecución de TEGA y los pasos de para su instalación.

## Imagen de VirtualBox

Es posible descargar una imagen de VirtualBox (<https://www.virtualbox.org/>) con la plataforma instalada:

- Español:

<https://drive.google.com/file/d/1zSNGEbSHgyvf6bWHqeDRZO9G-RLiAH4R/view?usp=sharing>

- Inglés:

<https://drive.google.com/file/d/1Z9viT5ebMIRX6sIdlq96qSk-esHD3pNR/view?usp=sharing>

Sistema Operativo:

- Debian 9
- Usuarios del Sistema Operativo (usuario:contraseña):
  - root:root
  - user:user
  - tegar (usuario del sistema)
- Usuarios de PostgreSQL
  - tegar:tegar

Plataforma:

- Directorio instalación: `/home/user/TEGA/`
- Ejecución de la plataforma:
  - a. Iniciar sesión del Sistema Operativo con el usuario *user*
  - b. Abrir una terminal y ejecutar:
    - i. `cd TEGA`
    - ii. `java -jar *.war`
  - c. Una vez iniciada, abrir el navegador e ingresar en `http://localhost:8080`
- Se encuentran creados los usuarios indicados en la Tabla 1. Para crear nuevos usuarios es necesario indicar un mail válido en la plataforma (paso 5 de la sección *Instalación y configuración*) y reiniciarla.
- Están importados los datos de ejemplo (ver sección *Importación de Datos / Ejemplos*):
  - Rueda et al 2013
  - Kamvar et al 2014
  - Completo

## Requerimientos

- Software

- Sistema operativo: Linux (TEGA fue desarrollada y testeada en Debian 8 y 9).
  - JAVA 8. Guía de instalación de OpenJDK 8: <http://openjdk.java.net/install> (solo es necesario el Java Runtime Environment)
  - Python v2.7.9: Guía de instalación: <https://www.python.org/downloads>
  - PostgreSQL v9.4.10. Guía de instalación: <https://www.postgresql.org/download>
  - Paquetes de Linux (Debian):
    - zip
    - parallel [7]
    - libssl-dev
    - libcurl4-openssl-dev
    - libxml2-dev
    - libopenblas-base
    - rsync
    - gawk
    - sudo (solo si se utiliza un usuario específico para la ejecución de los procedimientos de análisis)
  - R v3.4.3: Guía de instalación: <https://cran.r-project.org>
  - Paquetes de R (instalar con sus dependencias):
    - popr v2.5.0 [8]
    - hierfstat v0.04-22 [9]
- Hardware (mínimo)
    - Memoria RAM: 1GB .
    - CPU: 2.00GHz.
    - Disco rígido: 1GB.

## Instalación y configuración

1. Descargar la última versión de TEGA: <https://github.com/darioelias/TEGA/releases>
2. Descomprimir el archivo:

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

En la carpeta donde se lo descomprime se encontrará:

- Ejecutable de la plataforma: *TEGA-[VERSION].war*
- Carpetas de TEGA:
  - i. config: contiene los archivos de configuración (\*.yml) utilizados durante el inicio de la plataforma. Y el archivo de configuración del procedimiento de backup.
  - ii. scripts: contiene los scripts utilizados durante la instalación
  - iii. proc: contiene los scripts para la generación de los backups y para la ejecución de los procedimientos de análisis.
  - iv. archivos: carpeta donde se almacenarán los archivos de las ocurrencias (los archivos subidos por los usuarios a la plataforma y los archivos resultante de los procedimientos de análisis).
  - v. ejemplos: archivos de ejemplo para la importación de muestras y alelos (ver sección *Importación de Datos / Ejemplos*), y ejemplos de

- los archivos exportados por la plataforma al ejecutar un procedimiento (ver sección *Padrón de Procedimientos*).
- vi. manual: carpeta con los instructivos de TEGA.
3. Crear un usuario en PostgreSQL para la plataforma (guía para crear un usuario <https://www.postgresql.org/docs/9.4/static/app-createuser.html>).
  4. Crear una base de datos en PostgreSQL (guía para crear base de datos <https://www.postgresql.org/docs/9.4/static/sql-createdatabase.html>). Verificar que el usuario de la plataforma tenga todos los permisos (guía para asignar permisos <https://www.postgresql.org/docs/9.4/static/sql-grant.html>)
  5. Configuración de la plataforma: se realiza a través del archivo config/application-prod.yml, el cual puede ser editado con un editor de texto, la documentación de cada opción puede ser consultada en:  
<http://docs.spring.io/spring-boot/docs/current/reference/html/common-application-properties.html>
    - Configuraciones de la base de datos:
      - i. `spring.datasource.url`: url de conexión a la base de datos. En caso de utilizar un servidor local se puede indicar:  
`jdbc:postgresql://localhost:5432/NOMBRE_BD`  
(reemplazar NOMBRE\_BD por el nombre de la base de datos creada en el paso 4).
      - ii. `spring.datasource.username`: nombre del usuario creado en el paso 3.
      - iii. `spring.datasource.password`: contraseña del usuario creado en el paso 3.
    - Configuraciones del mail: propiedades `spring.mail.*`, por defecto se encuentra configurado un mail de ejemplo. También es necesario configurar la propiedad `jhipster.mail.from`.
    - Configuraciones de SSL (opcional): propiedades `server.ssl.*`, por defecto se encuentra desactivado.
    - Indicar un valor aleatorio (alfanumérico de 40 caracteres) en la propiedad `jhipster.security.rememberMe.key`.
  6. Configuración de procedimiento de backups: se realiza a través del archivo config/backup.config.sh. Es necesario configurar los parámetros:
    - Relacionados a la base de datos:
      - i. host: Host donde se encuentra el motor de bases de datos
      - ii. port: Puerto para acceder al motor de bases de datos
      - iii. db: Nombre de la base de datos (NOMBRE\_BD)
      - iv. user: Nombre del usuario creado en el paso 3.
      - v. passDB: Contraseña del usuario creado en el paso 3.
    - Relacionados a la encriptación del backup:
      - i. passEncrip: Clave utilizada para encriptar la copia de seguridad.
    - Relacionados a los directorios de TEGA: (utilizar direcciones absolutas)
      - i. dirTEGA: dirección a la carpeta de TEGA.
      - ii. dirBackups: dirección de la carpeta donde se almacenarán las copias de seguridad.
      - iii. dirTmp: dirección a la carpeta donde se almacenarán los archivos temporales durante la generación de la copia de seguridad.

- iv. dirLogs: dirección a la carpeta donde se almacenarán los logs de ejecución.
  - cantBackupsMax: Cantidad máxima de copias de seguridad a mantener en la carpeta de copias de seguridad.
7. Ejecutar la plataforma con el comando:
- ```
java -jar TEGA-[VERSION].war
```
- (considerando que el archivo TEGA.war se encuentra en la carpeta actual. Las carpetas *config*, *proc* y *archivos* deben encontrarse en el mismo directorio que el ejecutable).
8. Ejecutar el script *implementacion.[lenguaje].sql* (ubicado en la carpeta scripts de TEGA):
- ```
psql -h HOST -d NOMBRE_BD -U USUARIO -p PUERTO -a -f implementacion.[lenguaje].sql
```
- Reemplazar HOST, NOMBRE\_BD, USUARIO y PUERTO por los valores correspondientes. Es posible importar las descripciones de los parámetros, procedimientos y usuarios en Español o en Inglés, según se prefiera se debe utilizar el archivo *implementacion.es.sql* o *implementacion.en.sql* respectivamente (solo ejecutar uno).
9. En esta instancia ya es posible ingresar a la plataforma con los usuarios predefinidos (Tabla 1). Si se utilizó la configuración por defecto es posible acceder a la plataforma, desde la computadora donde se la instaló, con la dirección:
- <http://localhost:8080>

Login	Clave	Rol
admin	admin	Implementador
administrador	administrador	Administrador
investigador	investigador	Investigador
invitado	invitado	Invitado
anonimo	anonimo	Anónimo

**Tabla 1.** Usuarios disponibles por defecto.

10. Configuración de parámetros de TEGA: estos parámetros pueden ser configurados desde PostgreSQL o desde el padrón de parámetros de la plataforma (ver sección de *Parámetros*). Los principales parámetros a configurar son:
- NOMBRE\_PLATAFORMA: Nombre de la plataforma
  - HTML\_HOME\_ES: HTML de la pantalla de inicio (Español)
  - HTML\_HOME\_EN: HTML de la pantalla de inicio (Inglés)
  - HTML\_FOOTER\_ES: HTML del pie de las páginas (Español)
  - HTML\_FOOTER\_EN: HTML del pie de las páginas (Inglés)
  - URL\_LOGO\_ICO: URL para el icono de la plataforma
  - URL\_LOGO\_MENU: URL para el logo del menú
  - MAPA\_MUESTRAS\_LATITUD: Latitud inicial del Mapa de Muestras (decimal)

- MAPA\_MUESTRAS\_LONGITUD: Longitud inicial del Mapa de Muestras (decimal)
- ROL\_USUARIO\_DEFECTO: Rol de usuario por defecto. Por defecto se encuentra configurado el rol Invitado.

Al configurar los parámetros *HTML\_HOME\_\** y *HTML\_FOOTER\_\**, recordar mencionar el sitio web de TEGA y su licenciamiento. Además mencionar la licencia de los datos públicos de esta implementación.

Una vez configurados los parámetros, si se tenía abierta la plataforma en el navegador, es necesario volver a cargarla (shift + F5 o shift + click en el botón de actualizar del navegador).

11. Configuración de la ejecución de los procedimientos de análisis: es posible configurar el usuario que se utilizará para ejecutar los procedimientos. En caso de utilizar el mismo usuario con el que se ejecuta la plataforma no realizar lo siguientes pasos. Para utilizar un usuario distinto se debe indicar el nombre del usuario en el parámetro de TEGA: *USUARIO\_PROCEDIMIENTOS*.

Además el usuario con el que se ejecuta la plataforma debe encontrarse en */etc/sudoers.tmp*. Se recomienda no solicitar la contraseña para el uso de sudo en este usuario, para ello en */etc/sudoers.tmp* se debe indicar los privilegios:

*ALL=(ALL:ALL) NOPASSWD:ALL*

En caso de que se desee solicitar la contraseña para el uso de sudo, esta debe ser indicada en el parámetro de TEGA: *CLAVE\_USUARIO\_PROCEDIMIENTOS*.

## Recomendaciones

Recomendaciones para el uso en producción de TEGA:

- Cambiar la clave del usuario *admin* y eliminar los usuarios de ejemplo.
- Utilizar un programador de tareas (como cron) para ejecutar la generación de copias de seguridad al menos una vez por mes.
- Utilizar un usuario específico para la ejecución de procedimientos de análisis (ver sección *Pasos de la ejecución de Procedimientos*). Y que dicho usuario sea un usuario del sistema con los permisos mínimos (ej. que no se encuentre en */etc/sudoers.tmp*).
- Quitar el permiso de lectura (a nivel de grupo y otros) a los archivos de la carpeta *config*.
- Configurar el uso del protocolo SSL para las conexiones.
- Además de instalar TEGA en un ambiente de producción crear un ambiente de testing para probar las configuraciones y procedimientos antes de enviarlas a producción.

# Uso de la plataforma

En esta sección se explican las funcionalidades básicas de TEGA.

## Opciones de menú

- Inicio: opción para acceder a la pantalla de inicio de la plataforma.
- Análisis:
  - Conjuntos de Muestras: padrón de conjuntos de muestras, entidad que permite agrupar muestras en un conjunto para ser utilizadas en los análisis.
  - Análisis de genotipos: padrón de análisis de genotipos, entidad relacionada a la ejecución de los procedimientos de análisis.
  - Procedimientos: padrón de los procedimientos de análisis, entidad vinculada a los procedimientos creados por el usuario para ser ejecutados desde los análisis de genotipos.
  - Parámetros de Procedimientos: padrón de parámetros de los procedimientos.
  - Categorías de Parámetros: padrón de las categorías de los parámetros de procedimientos, esta entidad permite agrupar a los parámetro de cada procedimiento.
- Entidades Esp. (específicas):
  - Muestras: padrón de muestras, entidad que contiene los datos ambientales y de gestión de las muestras.
  - Mapa de Muestras: pantalla de visualización geográfica de las muestras.
  - Loci: padrón de loci, entidad relacionada a los marcadores moleculares.
  - Alelos: padrón de alelos, entidad relacionada a los valores de los alelos de cada muestra y locus.
- Entidades Gral. (generales):
  - Países: padrón de países.
  - Provincias: padrón de provincias.
  - Localidades: padrón de localidades.
  - Regiones: padrón de regiones, entidad vinculada a las muestras para referenciar una región específica, como un río o una montaña.
  - Especies: padrón de especies.
  - Instituciones: padrón de instituciones, entidad vinculada a la institución que realizó la recolección de las muestras.
  - Profesionales: padrón de profesionales, entidad vinculada a la persona que realizó la recolección de las muestras (o el contacto con el instituto).
  - Tejidos: padrón de tejidos, entidad vinculada al tejido del que se extrajo la muestra.
  - Modos de recolección: padrón de modos de recolección, entidad relacionada a la técnica de recolección utilizada.
  - Proyectos: padrón de proyectos, entidad relacionada a un proyecto de investigación.

- Atributos: padrón de los atributos dinámicos, los cuales pueden ser creados por el usuario y vincularlos a las Muestras, Proyectos, Análisis de genotipos, Conjuntos de Muestras y Loci.
- Cuenta:
  - Ajustes: pantalla para modificar los datos del usuario actual.
  - Contraseña: pantalla para modificar la contraseña del usuario actual.
  - Sesiones: pantalla para visualizar las sesiones abiertas del usuario actual.
  - Cerrar sesión: opción para cerrar la sesión actual.
  - Iniciar sesión: pantalla para iniciar la sesión, indicando usuario y contraseña.
  - Crear una cuenta: pantalla para crear una nueva cuenta de usuario.
- Administración:
  - Gestión de usuarios: padrón de usuarios de la plataforma.
  - Métricas: pantalla para visualizar las métricas de la máquina virtual de JAVA (JVM) y peticiones de HTTP.
  - Salud: pantalla para visualizar el espacio libre del disco rígido.
  - Configuración: pantalla para visualizar las configuraciones de la plataforma (vinculado al archivo config/application-prod.yml).
  - Auditorías: pantalla para visualizar los eventos de autenticación de los usuarios.
  - Logs: pantalla para visualizar los logs de la plataforma.
  - API: pantalla de swagger para visualizar los servicios WEB de la plataforma.
  - Parámetros de TEGA: padrón de parámetros de la plataforma, entidad relacionada a parámetros internos de la plataforma.
  - Copia de seguridad: opción que permite crear una copia de seguridad de la plataforma completa y descargarla.
- Idioma: menú con los idiomas disponibles de la plataforma.

## Padrones

En esta sección se explican las características y funcionalidades básicas de los padrones de TEGA.

TEGA posee por cada entidad un padrón con las funcionalidades básicas: Alta, Baja, Modificación, Consulta y Exportación (ABMCE).

Todas las entidades poseen un campo denominado ID que es el identificador de cada ocurrencia, el ID es asignado automáticamente por la plataforma. Por otro lado, todos los padrones poseen un campo denominado “Código”, el cual debe ser único y debe ser asignado por el usuario. En el caso de las muestras existen dos códigos: Externo e Interno, el externo es el código de la muestra asignado por el instituto que recolectó la muestra, y el interno es el asignado por el equipo de investigación que utiliza TEGA.

### Pantalla de consulta

En esta pantalla es posible listar, crear, eliminar y exportar las ocurrencias del padrón.

Las pantallas de consultas tienen los siguientes componentes (Figura 1):

1. Botón para crear una nueva ocurrencia: al presionarlo se abre la ventana para crear una nueva ocurrencia.
2. Criterios de búsqueda:
  - a. Los criterios de tipo texto (como el Código) son case-insensitive y su evaluación es por inclusión.
  - b. Para eliminar un criterio de búsqueda basta con blanquear el casillero.
  - c. Para aplicar los criterios de búsqueda es necesario presionar el botón de actualizar (ver ítem 3.a).
3. Grilla con ocurrencias que cumplen los criterios de búsqueda:
  - a. Al hacer click en los encabezados de las columnas es posible cambiar el orden de las filas en la grilla.
  - b. En cada fila hay botones para visualizar, editar o eliminar la fila.
4. Botones para Actualizar, Importar, Exportar y Eliminar (AIEE):
  - a. Actualizar: actualiza la lista de ocurrencias a partir de los criterios de búsqueda.
  - b. Importar: abre la ventana de importación de datos. Por el momento solo se encuentra disponible en los padrones de Muestras y Alelos.
  - c. Exportar: exporta las ocurrencias que aparecen en la grilla (ítem 4) a un archivo CSV.
  - d. Exportar atributos: exporta los atributos dinámicos de las ocurrencias que aparecen en la grilla (ítem 4). Esta opción solo está disponible en las entidades que pueden ser asociadas a atributos dinámicos.
  - e. Eliminar: abre la ventana de eliminación masiva. Allí es posible seleccionar un grupo de ocurrencias en función de los criterios de búsqueda y eliminarlos.
5. Páginas disponibles
6. Cantidad de ocurrencias a mostrar por páginas

**Muestras**

**2) Criterios de búsqueda**

**1) Botón para crear**

ID	Código Interno	Conjunto de Muestras	Localidad
Pública	Código Externo	Región	Ubicación
Recolección Desde	Institución	Especie	
Recolección Hasta	Proyecto	Profesional	

**4) Botones AIEE**

**3) Grilla con ocurrencias**

ID	Código Interno	Código Externo	Fecha de Recolección	Región	Localidad	Especie	Ubicación	Vista	Editar	Eliminar
1	CONJUNTO01 EXT1 1	EXT1	1 ene. 2011	Region 1	Localidad 1	Especie 1	Estante 1	<span>Vista</span>	<span>Editar</span>	<span>Eliminar</span>
2	CONJUNTO01 EXT2 2	EXT2	1 ene. 2011	Region 1	Localidad 1	Especie 1	Estante 1	<span>Vista</span>	<span>Editar</span>	<span>Eliminar</span>
3	CONJUNTO01 EXT3 3	EXT3	1 ene. 2011	Region 1	Localidad 1	Especie 1	Estante 1	<span>Vista</span>	<span>Editar</span>	<span>Eliminar</span>
4	CONJUNTO01 EXT4 4	EXT4	1 ene. 2011	Region 1	Localidad 1	Especie 1	Estante 1	<span>Vista</span>	<span>Editar</span>	<span>Eliminar</span>
5	CONJUNTO01 EXT5 5	EXT5	1 ene. 2011	Region 1	Localidad 1	Especie 1	Estante 1	<span>Vista</span>	<span>Editar</span>	<span>Eliminar</span>
6	CONJUNTO01 EXT6 6	EXT6	1 ene. 2011	Region 1	Localidad 1	Especie 1	Estante 1	<span>Vista</span>	<span>Editar</span>	<span>Eliminar</span>
7	CONJUNTO01 EXT7 7	EXT7	1 ene. 2011	Region 1	Localidad 1	Especie 1	Estante 1	<span>Vista</span>	<span>Editar</span>	<span>Eliminar</span>

**3.a) Cambio del orden**

**3.b) Botones VEE**

**5) Páginas disponibles** →

**6) Filas por Pag.** →  Filas por Pag.

**Actualizar** **Importar** **Exportar** **Exp. Atributos** **Eliminar**

**Figura 1.** Pantalla de consulta de Muestras.

## Pantalla de visualización

Esta pantalla permite solo visualizar los atributos de una entidad, sin posibilidad de modificarlos (Figura 2).

**Crear o editar Proyecto**

ID	1
Código	PROYECTO 1
Fecha	2018-01-08 <input type="button" value=""/>
Detalle	PROYECTO 1
Público	<input checked="" type="checkbox"/>
Atributos	
Archivos	

**Figura 2.** Pantalla de visualización de Proyectos.

## Pantalla de edición

Esta pantalla se utiliza para el Alta y Edición de las ocurrencias (Figura 3.a). Para que se persistan los cambios es necesario presionar el botón de Guardar.

En las pantallas de edición es posible modificar los atributos de cada ocurrencia y vincularlas a ocurrencias de otras entidades. Formas de agregar y eliminar ocurrencias vinculadas (ejemplo con Conjuntos de Muestras):

1. Individualmente (Figura 3.b):
  - a. Agregar:
    - i. Primero se debe seleccionar el conjunto de muestras (la búsqueda se realiza por código y el detalle).
    - ii. Luego se debe presionar el botón de agregar.
  - b. Eliminar: se debe presionar el botón de eliminar de la fila correspondiente.
2. Masivamente (Figura 3.c):
  - a. Agregar:
    - i. Primero se debe presionar el botón de búsqueda:  , que abrirá la pantalla de búsqueda.
    - ii. Indicar los criterios de búsqueda, por ejemplo una región o un rango de fechas.
    - iii. Presionar el botón de actualizar, listará las ocurrencias que cumplan el filtro en la primer grilla.

iv. En la primer grilla es posible seleccionar individualmente las ocurrencias, tildando en la casilla de selección:  . También es posible tildar todas las casilla presionando el botón:  al pie de la grilla. Por otro lado el botón  deselecciona todas las filas de la grilla.

v. Presionar el botón  para agregar las ocurrencias.

- b. Eliminar: en la segunda grilla también es posible seleccionar las filas, y al presionar el botón  se las elimina.

Una vez que se termine de seleccionar las ocurrencias se debe presionar el botón de guardar para confirmar la selección.

Por otro lado, en el caso de los Alelos, es posible crearlos y eliminarlos desde la pantalla de edición de las muestras (Figura 3.d):

1. Para crear un nuevo alelo desde la pantalla de edición de la muestras hay que indicar los atributos y presionar el botón de aplicar.
2. Para editar un alelo hay que presionar el botón de editar de la fila, luego modificar los atributos y por último presionar el botón de aplicar.
3. Para descartar la creación o modificación de un alelo hay que presionar el botón de cancelar.
4. Para eliminar un alelo hay que presionar el botón de eliminación de la fila.

Para adjuntar un archivo a una ocurrencia es necesario que esta se encuentre creada (no es posible adjuntar archivos durante la creación, solo es posible durante la edición). Para adjuntar un archivo se debe (Figuras 3.e y 3.f):

1. Presionar el botón de enviar archivo en el panel de Archivos de la pantalla de edición. Al hacerlo se abrirá la pantalla de envío de archivos.
2. Seleccionar el archivo deseado, indicar un detalle y si es público.
3. Presionar el botón de “Enviar”, esto provocará el envío del archivo al servidor y se persistirá la vinculación del archivo con la ocurrencia.

Por otro lado, para descargar o eliminar un archivo solo es necesario presionar el botón de la fila correspondiente en la grilla de archivos.

**a)** Pantalla completa de edición de muestras.

**b)** Panel de selección de conjuntos de muestras:

- 1.a.i) Selector
- 1.a.ii) Botón para agregar
- 2.a.i) Botón de búsqueda
- 2.a.ii) Criterios de búsqueda
- 1.b) Botón de eliminación

**c)** Pantalla de selección masiva de conjuntos de muestras:

- 2.a.iv) Columna de selección
- 2.a.v) Botones de selección
- 2.a.iii) Botón para actualizar
- 2.b) Botones de selección y eliminación

**d)** Panel de alta, baja y modificación de alelos:

- 1/2) Atributos de alelos
- 1/2) Botón para aplicar
- 2) Botón de edición
- 3) Botón de cancelación
- 4) Botón de eliminación

**e)** Panel de alta y baja de archivos:

- 1) Botón para enviar archivos
- Botón de eliminación
- Botón para descargar el archivo

**f)** Pantalla de envío de archivos:

- 2) Archivo y atributos
- 3) Botón para confirmar el envío

**Figura 3.** Pantalla de edición de muestras: a) pantalla completa de edición, b) panel de selección de conjuntos de muestras, c) pantalla de selección masiva de conjuntos de muestras. d) panel de alta, baja y modificación de alelos, e) panel de alta y baja de archivos, f) pantalla de envío de archivos.

## Mapa de Muestras

El padrón de muestras además de las pantallas mencionadas también posee una pantalla para visualizar las coordenadas geográficas de cada muestra utilizando OpenStreetMap (<https://www.openstreetmap.org>).

Para visualizar las muestras se deben seguir los siguientes pasos (Figura 4):

1. Indicar los criterios de búsqueda (análogos a los de la pantalla de consulta)
2. Seleccionar un color: es posible cambiar el color por defecto haciendo click en el cuadrado coloreado.
3. Presionar el botón de aplicar.

Luego del paso 3 aparecerán en el mapa las muestras que cumplen los criterios y que posean las coordenadas geográficas, coloreadas con el color seleccionado.

Al presionar el botón de cancelar se eliminan del mapa todas las muestras.

Al hacer click en un ícono del mapa se despliega una ventana con el ID, el código Interno y el Externo.

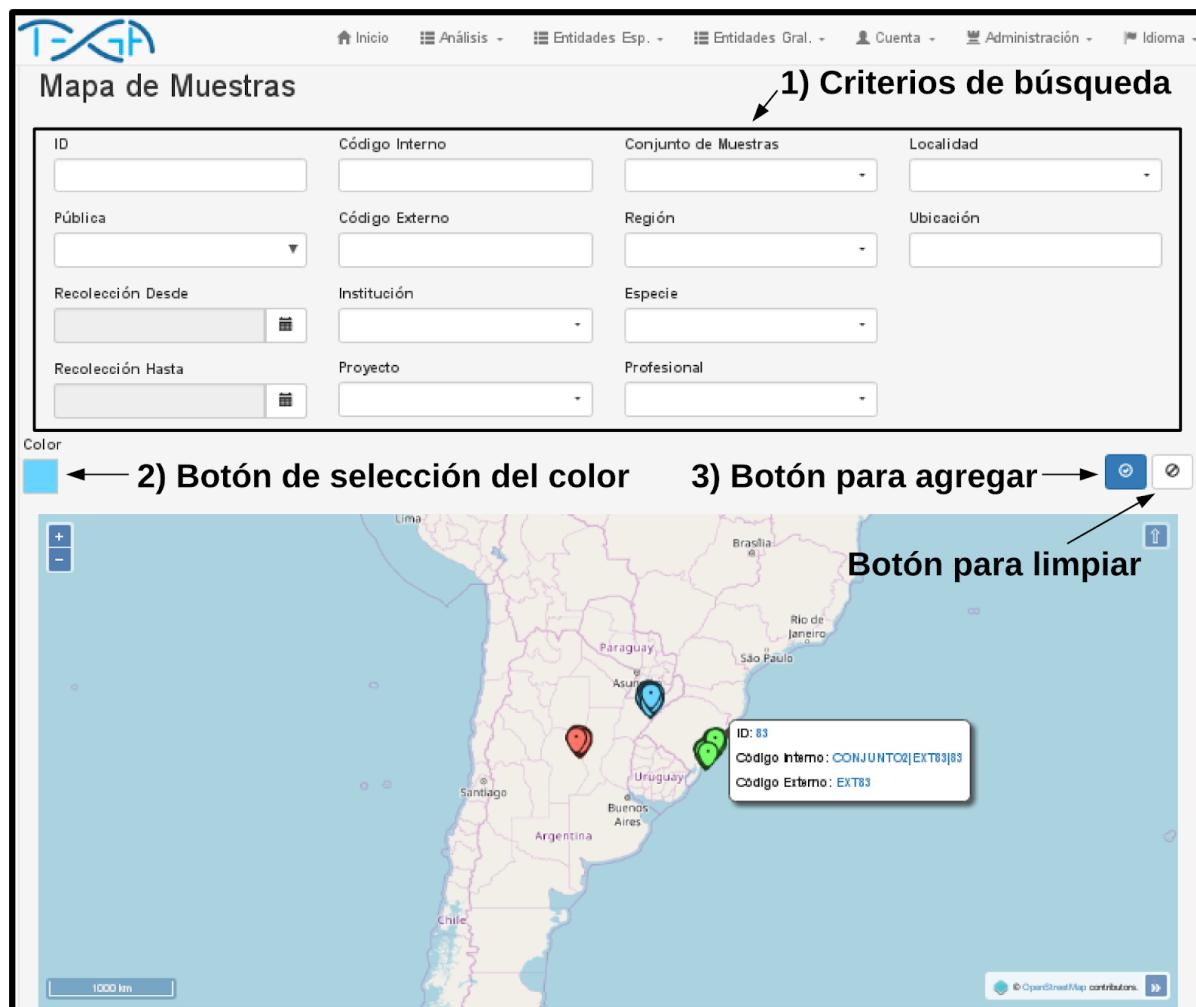


Figura 4. Pantalla del Mapa de Muestras.

## Atributos dinámicos

Dado que cada grupo de investigación puede estudiar distintas características del organismo o vincularlas a distintos factores ambientales, por ejemplo, la altura, la temperatura o la salinidad del agua, se desarrolló un módulo de atributos dinámicos para contemplar dicha variabilidad.

En la opción de menú *Entidades Gral./Atributos* es posible crear los atributos de cada entidad. Al crear un nuevo atributo se debe indicar (Figura 5.a):

- Código: debe ser único.
- Tipo: Indica el tipo del atributo (Carácter, Numérico, Entero, Lógico o Fecha). La plataforma validará el tipo de dato al importar o editar el atributo en las ocurrencias.
- Entidad: entidad a la que estará vinculado el atributo (Muestra, Proyecto, Análisis de Genotipos, Conjunto de Muestras o Locus).

- Detalle: descripción del atributo.
- Valor por defecto: el valor por defecto que tendrá el atributo en cada ocurrencia.

Para mantener la consistencia del tipo de los atributos en las ocurrencias, no es posible editar el Tipo, ni la Entidad de un atributo.

Una vez creado el atributo es posible modificar el valor de este en cada ocurrencia desde la pantalla de edición de la entidad correspondiente (Figura 5.b).

Por otro lado es posible exportar los valores de los atributos dinámicos desde la pantalla de consulta de cada entidad y es posible incluirlos en los procedimientos de análisis. La primer versión de TEGA no posee atributos dinámicos predefinidos.

Crear o editar Atributo a)

<b>Código</b>	longitud
<b>Tipo</b>	Numérico
<b>Entidad</b>	Muestra
<b>Detalle</b>	Longitud (cm)
<b>Valor por defecto</b>	0.00
<input type="button" value="Cancelar"/> <input type="button" value="Guardar"/>	

Crear o editar Muestra b)

ID 1	Latitud -30.5277364379831	Profesional Profesional 1	Especie Especie 1						
Código Interno CONJUNTO01 EXT1 1	Longitud -62.5196588969693	Modo Recolección Método 1	Institución Institution 1						
Código Externo EXT1	Región Region 1	Tejido Tejido 1	Ubicación Estante 1						
Fecha de Recolección 2011-01-01	Localidad Localidad 1	Sexo FEMENINO	Pública <input checked="" type="checkbox"/>						
Detalle Comentarios...									
<b>Alelos</b>									
<b>Conjunto de Muestras</b>									
<b>Proyectos</b>									
<b>Atributos</b> <table border="1"> <thead> <tr> <th>Código</th> <th>Detalle</th> <th>Valor</th> </tr> </thead> <tbody> <tr> <td>longitud</td> <td>Longitud (cm)</td> <td>10.42</td> </tr> </tbody> </table>				Código	Detalle	Valor	longitud	Longitud (cm)	10.42
Código	Detalle	Valor							
longitud	Longitud (cm)	10.42							
<b>Archivos</b>									
<input type="button" value="Cancelar"/> <input type="button" value="Guardar"/>									

**Figura 5.** Atributos dinámicos. a) Pantalla de creación y edición de atributos. b) Edición del valor de atributos dinámicos en una muestra.

# Importación de Datos

En esta sección se explica cómo importar masivamente los datos en TEGA. La importación de datos se divide en tres etapas: importación de Muestras, importación de los valores de atributos dinámicos de las Muestras e importación de Loci y Alelos.

## Importación de Muestras

Es posible importar masivamente los datos de las Muestras junto con los de sus entidades vinculadas (como Región, Tejido, Proyecto, etc.). Esta importación puede ser utilizada para el agregado de datos en la plataforma y para la modificación de datos existentes, por ejemplo modificar la Región de las Muestras indicadas. Es posible acceder a la pantalla de importación de muestras a través del botón de Importación ubicado en la pantalla de consulta de las muestras. En la pantalla de importación es necesario indicar el archivo a importar, el carácter separador de las columnas y en *Datos a importar* se debe indicar *Muestras*. Una vez procesado el archivo es posible visualizar la cantidad de entidades creadas o modificadas, y muestras no identificadas (Figura 6).

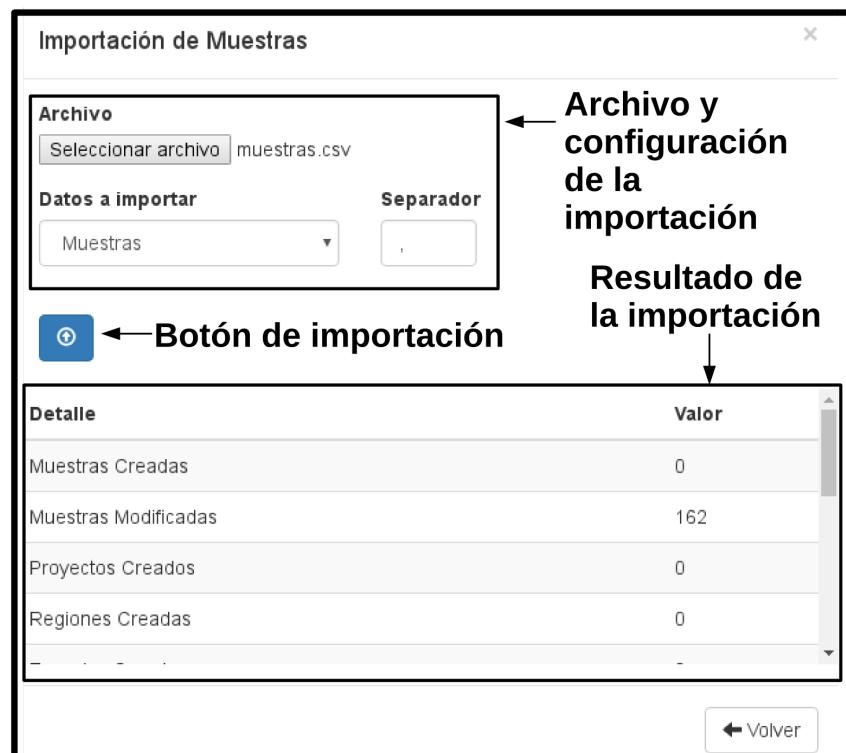
El formato del archivo a importar es texto plano, con columnas separadas por un mismo carácter, la primer fila indica el atributo que se importará en cada columna, no es necesario que se encuentren todos los atributos. Se debe respetar las mayúsculas y minúsculas en los nombres de los atributos (primera fila del archivo) y no deben incluir el acento gráfico (tilde). En los datos de las muestras (a partir de la segunda fila) no se distingue entre mayúsculas y minúsculas, y los datos pueden llevar el acento gráfico. Por ejemplo el código interno “código1” se considera igual a “Código1”.

Columnas disponibles en el archivo de importación:

- Columnas utilizada para identificar a las muestras: existen tres formas para identificar a las muestras:
  - a. Utilizando las columnas conjuntoMuestras y codigoExterno (Figura 7.a): El texto indicado en la columna conjuntoMuestras se utilizará para buscar un conjunto de muestras primero por el código y, si no se encuentra, por el detalle. Si no se encuentra se creará un nuevo conjunto de muestras. Luego la plataforma buscará las muestras del conjunto de muestras indicado, que posean el código externo indicado, si no existen muestras con esos filtros se creará una nueva muestra, su código interno se completará automáticamente concatenando el código del conjunto de muestras, el código externo y el ID.
  - b. Utilizando la columna de codigointerno (Figura 7.b): La plataforma buscará la muestra cuyo código interno sea igual al indicado, si no existe se creará una nueva muestra.
  - c. Utilizando la columna id (Figura 7.c): La plataforma buscará la muestra a partir del ID, si no la encuentra se creará una nueva muestra con el ID asignado por la plataforma.
- Columnas relacionadas a entidades: el texto que se indica en las columnas que se vinculan a otras entidades (como localidad y región), será utilizado para buscar la ocurrencia correspondiente primero a partir del código y luego a partir del nombre (o detalle). La búsqueda no distingue entre mayúsculas y minúsculas, además es

possible indicar acentos gráficos. En caso de que no se encuentre se creará la ocurrencia con el texto indicado como código y nombre (o detalle).

- conjuntoMuestras: relacionada a la entidad Conjuntos de Muestras.
  - localidad: relacionada a la entidad Localidades.
  - provincia: relacionada a la entidad Provincias.
  - pais: relacionada a la entidad Paises.
  - region: relacionada a la entidad Regiones.
  - profesional: relacionada a la entidad Profesionales.
  - institucion: relacionada a la entidad Instituciones.
  - especie: relacionada a la entidad Especies.
  - proyecto: relacionada a la entidad Proyectos.
  - tejido: relacionada a la entidad Tejidos.
  - modoRecolección: relacionada a la entidad Modos de Recolección
- Columnas relacionadas a atributos estáticos de las Muestras:
    - id: relacionada al ID de la muestra. Tipo: entero positivo, ej: 1027.
    - codigointerno: relacionada al Código Interno. Tipo: cadena de caracteres (max. 255), ej: "COD\_INT|1".
    - codigoExterno: relacionada al Código Externo. Tipo: cadena de caracteres (max. 255), ej: "COD\_EXT|1"
    - fechaRecolección: relacionada a la Fecha de Recolección. Tipo: fecha. Formatos disponibles: yyyy-MM-dd, yyyy/MM/dd, yyyyMMdd (yyyy: año, MM: mes, dd: día). Ejemplo: 2017-12-26, 2017/03/01, 20170121.
    - ubicacion: relacionada a la Ubicación. Tipo: cadena de caracteres, ej: "estante 3".
    - publico: relacionada al atributo Público. Tipo: lógico, ej: true.
    - latitud: relacionada a la Latitud. Tipo: numérico, ej: -38.0517.
    - longitud: relacionada a la Longitud. Tipo: numérico, ej: -57.5340.
    - sexo: relacionada al Sexo. Tipo: enumerado, valores posibles:
      - Femenino: F, FEMENINO o FEMALE
      - Masculino: M, MASCULINO o MALE
      - Indefinido: I, U, INDEFINIDO o UNDEFINED
    - detalle: relacionada a Detalle: Tipo: cadena de caracteres (max. 255), ej: "Comentario...".



**Figura 6.** Pantalla de importación de muestras.

```

1 conjuntoMuestras,codigoExterno,latitud,longitud,localidad,provincia,pais,region,fechaRecolección a)
2 CONJUNTO1,EXT1,-30.5277364379831,-62.5196588969693,Localidad 1,Provincia 1,Pais 1,Region 1,20110101
3 CONJUNTO1,EXT2,-30.7074601325742,-62.5334568025105,Localidad 1,Provincia 1,Pais 1,Region 1,20110101

```

```

1 codigoInterno,latitud,longitud,localidad,provincia,pais,region,fechaRecolección b)
2 CODIGO1,-30.5277364379831,-62.5196588969693,Localidad 1,Provincia 1,Pais 1,Region 1,20110101
3 CODIGO2,-30.7074601325742,-62.5334568025105,Localidad 1,Provincia 1,Pais 1,Region 1,20110101

```

```

1 id,latitud,longitud,localidad,provincia,pais,region,fechaRecolección c)
2 1,-30.5277364379831,-62.5196588969693,Localidad 1,Provincia 1,Pais 1,Region 1,20110101
3 2,-30.7074601325742,-62.5334568025105,Localidad 1,Provincia 1,Pais 1,Region 1,20110101

```

**Figura 7.** Archivos de ejemplo para la importación de Muestras. Las columnas se encuentran delimitadas por “,”. La primera línea es el encabezado, que indica el nombre de los atributos. Las segunda y tercera fila son datos de ejemplo de dos Muestras. Columnas utilizadas para identificar a las muestras: a) conjuntoMuestras y codigoExterno, b) codigoInterno y c) id.

## Importación de valores de atributos dinámicos de las muestras

La importación de los valores de atributos dinámicos de las muestras es análoga a la importación de las muestras. La pantalla de importación es la misma que la de las muestras pero en *Datos a importar* se debe indicar *Atributos extras de Muestra*. En el archivo de importación es necesario indicar las columnas identificadoras de las muestras y una columna con el código de cada atributo dinámico a importar (Figura 8). Tanto las Muestras como los Atributos deben haber sido creados antes de la importación. Durante la importación TEGA validará los valores de cada atributo, en caso de haber alguno inválido no será importado y se reportará la fila y el código del atributo relacionado. El formato de cada valor depende del tipo del atributo:

- Carácter: texto plano, las comillas no son necesarias. Ejemplo: Tools for Evolutionary and Genetic Analysis.
- Entero: valores enteros (cero, positivos o negativos). Ejemplos: 0, 10, -20.
- Numérico: valores numéricos, es posible utilizar notación científica, se debe utilizar el punto decimal. Ejemplos: 0.5, -1.4, 100, 1e-3.
- Lógico: valores posibles: 1 (verdadero) o 0 (falso).
- Fecha: formatos disponibles: yyyy-MM-dd, yyyy/MM/dd, yyyyMMdd (yyyy: año, MM: mes, dd: día). Ejemplo: 2017-12-26, 2017/03/01, 20170121.

```

1 id,peso,temperatura,fechaRecepcion
2 1,11.33,9.67,20110202
3 2,15.42,10.07,20110203
4 3,13.22,8.36,20110304

```

**Figure 8.** Archivo de ejemplo para la importación de valores de atributos dinámicos. La primera fila es el encabezado, indica los códigos de los atributos dinámicos a importar, el resto de las filas corresponden a tres muestras. En este ejemplo se utiliza el identificador interno (id) para identificar a cada muestra. Las columnas *peso*, *temperatura* y *fechaRecepcion*, corresponden a los atributos a importar en cada muestra.

## Importación de Loci y Alelos

Es posible importar masivamente los datos de los Loci y Alelos de cada muestra que ya se encuentre cargada en la plataforma.

Es posible acceder a la pantalla de importación de alelos a través del botón de Importación ubicado en la pantalla de consulta de los alelos. En la pantalla de importación es necesario indicar el archivo a importar, el tipo de importación, el carácter separador de las columnas y , si el tipo de importación es “Matriz”, el carácter utilizado para indicar los alelos nulos y el estado de los alelos (público o privado). Una vez procesado el archivo es posible visualizar la cantidad de alelos y loci creados, muestras no identificadas, loci no identificados, entre otros datos (Figura 9).

Al igual que en el archivo de las Muestras, el formato es texto plano, con columnas separadas por un carácter, y la primer fila indica el nombre de los atributos o el nombre de los Loci.

Existen dos tipos de archivo para esta importación:

- Matriz (Figuras 10.a, 10.b y 10.c): Las primeras columnas del archivo deben ser la(s) columna identificatoria de las muestras (análogas a la del archivo de importación de muestras). El resto de las columnas corresponden al valor de cada alelo en cada Loci. Por cada Locus que se desee importar se debe indicar el índice al final del nombre del Locus separándolo por un guión bajo. En caso de que el locus no exista se creará y se indicará como ploidía el valor del mayor índice. El valor de cada alelo es del tipo cadena de caracteres (max. 255). Los alelos nulos deben ser identificados con un carácter específico (ej. "?") el cual también se indicará en la pantalla de importación. El atributo Público de los alelos será completado con el valor del tilde titulado Público de la pantalla de importación.
- Lista (Figura 10.d): las columnas disponibles son:
  - id: ID del Alelo. Tipo: entero positivo, ej: 123.

- idMuestra: ID de la muestra. Tipo: entero positivo, ej: 123.
- codigointernoMuestra: código interno de la muestra. Tipo: cadena de caracteres (max. 255), ej: “COD\_INT|1”.
- idLocus: ID del locus. Tipo: entero positivo, ej: 123.
- codigoLocus: código del locus. Tipo: cadena de caracteres (max. 255), ej: “COD\_LOCUS”.
- indice: relacionado al Índice del Alelo. Tipo: entero positivo, ej: 123.
- valor: relacionado al Valor del Alelo. Tipo: cadena de caracteres (max. 255), ej: “1010”.
- publico: relacionado al atributo Público del Alelo. Tipo: lógico, ej: false.

Se debe indicar la columna idMuestra o codigointernoMuestra, no es necesario indicar ambas. De la misma forma con el Locus (idLocus o codigoLocus).

En caso de indicar la columna id la plataforma buscará el alelo en la base de datos a partir del id. Sino buscará el alelo que esté vinculado a la misma muestra y locus y posea el mismo índice, sino hay un alelo con esos datos se creará uno nuevo. En caso de que el locus no exista se creará.

**Importación de Alelos**

**Archivo**  
Seleccionar archivo alelos.csv

**Tipo de Importación**  
Matriz

**Separador**  
,

**Valor Nulo**  
?

**Público**

**Botón de importación**

**Resultado de la importación**

Detalle	Valor
Alelos Creados	0
Alelos Modificados	2198
Loci Creados	0
Loci Modificados	0
...	...

← Archivo y configuración de la importación

← Botón de importación

Resultado de la importación

← Volver

**Figura 9.** Pantalla de importación de loci y alelos.

```
1 conjuntoMuestras,codigoExterno,LOCUS1_1,LOCUS1_2,LOCUS2_1,LOCUS2_2  
2 CONJUNT01,EXT1,256,270,268,?  
3 CONJUNT01,EXT2,242,242,194,194
```

a)

```
1 codigoInterno,LOCUS1_1,LOCUS1_2,LOCUS2_1,LOCUS2_2  
2 CODIG01,256,270,268,?  
3 CODIG02,242,242,194,194
```

b)

```
1 id,LOCUS1_1,LOCUS1_2,LOCUS2_1,LOCUS2_2  
2 1,256,270,268,?  
3 2,242,242,194,194
```

c)

```
1 codigoInternoMuestra,codigoLocus,indice,valor,publico  
2 CONJUNT01|EXT1|1,LOCUS9,1,100,false  
3 CONJUNT01|EXT1|1,LOCUS9,2,200,true
```

d)

**Figura 10.** Archivo de ejemplo para la importación de alelos y loci. Las columnas se encuentran delimitadas por “,”. En a), b) y c) el tipo de archivo es Matriz, la primera línea es el encabezado, que indica el nombre de los loci a crear o actualizar. La segunda y tercer fila son datos de ejemplo de los alelos de dos muestras. Columnas utilizadas para identificar a las muestras: a) conjuntoMuestras y codigoExterno, b) codigointerno y c) id. d) Tipo de archivo Lista, la primera línea es el encabezado, que indica el nombre de los atributos a importar. La segunda y tercera fila son datos de ejemplo de los alelos de dos muestras.

## Ejemplos

En la carpeta *ejemplos* de la plataforma existen archivos de importación de ejemplo, el carácter separador es la coma (,). Dentro de la carpeta *ejemplos* hay una carpeta por idioma (*en* y *es*), y dentro de ellas las siguientes carpetas:

- completo: ejemplo con todos los campos (descriptivos) de las muestras
  - samples.csv: archivo de importación de muestras (162 muestras).
  - alleles.csv: archivo de importación de alelos, con el formato de matriz (2198 alelos no nulos y 394 nulos).
- Rueda\_et\_al\_2013: ejemplo con datos del trabajo de Rueda et al, 2013.
  - samples.csv: archivo de importación de muestras.
  - alleles.csv: archivo de importación de alelos, con el formato de matriz.
- Kamvar\_et\_al\_2014: ejemplo con datos del trabajo de Kamvar et al, 2014.
  - samples.csv: archivo de importación de muestras.
  - alleles.csv: archivo de importación de alelos, con el formato de matriz.
- instructivo: archivos de ejemplo mencionados en el instructivo.
  - muestras: ejemplos de la Figura 7 y 8.
    - muestras\_a.csv: ejemplo con columnas identificadorias: conjuntoMuestras y codigoExterno.
    - muestras\_b.csv: ejemplo con columna identificatoria: codigointerno.
    - muestras\_c.csv: ejemplo con columna identificatoria: id.
    - muestras\_attributes.csv: ejemplo de importación de atributos dinámicos. Columna identificatoria: id.
  - alleles: ejemplos de la Figura 10.
    - alelos\_a.csv: formato de Matriz, con columnas identificadorias: conjuntoMuestras y codigoExterno.

- alelos\_b.csv: formato de Matriz, con columna identificatoria: codigointerno.
- alelos\_c.csv: formato de Matriz, con columna identificatoria: id.
- alelos\_d.csv: formato de Lista.

## Análisis de genotipos

En esta sección se explica cómo crear un Análisis de Genotipos y ejecutar los procedimientos de análisis. TEGA posee un padrón denominado Análisis de Genotipos, destinado a la gestión de los datos vinculados a la ejecución de procedimientos de análisis de genotipos.

### Conjuntos de Muestras

Para crear un Análisis de Genotipos primero es necesario armar uno o más Conjuntos de Muestras que agruparán las Muestras a estudiar según el criterio necesario, por ejemplo muestras de distintas regiones. Es posible crear un nuevo conjunto de muestras desde el padrón *Entidades esp. / Conjuntos de Muestras*, es necesario indicar un código único y las muestras que lo integrarán. Una vez completado el código y seleccionadas las muestras se debe presionar el botón de guardar (Figura 11).

ID	Código	Detalle
1	CONJUNTO01	CONJUNTO01

Muestras		
Muestra	Acción	
CONJUNTO01 EXT1 1	<a href="#">Comentarios...</a>	
CONJUNTO01 EXT2 2	<a href="#">Comentarios...</a>	
CONJUNTO01 EXT3 3	<a href="#">Comentarios...</a>	
CONJUNTO01 EXT4 4	<a href="#">Comentarios...</a>	

Figura 11. Pantalla de edición de los Conjuntos de Muestras.

## Análisis de genotipos

Luego de crear los conjuntos de muestras se debe crear un nuevo Análisis de Genotipos (Figuras 12.a y 12.b), donde se indicará el Proyecto, los Conjuntos de Muestras y Loci a estudiar. También es posible seleccionar individualmente o masivamente los Loci y los Conjuntos de Muestras. Una vez creado el análisis es posible visualizar los alelos de cada Muestra y Loci de forma matricial presionando el botón titulado “Alelos” en la pantalla de consulta (Figura 12.c).

Por otro lado es posible ejecutar los procedimientos de análisis desde la pantalla de consulta (Figura 12.d), presionando el botón  Proc. Al hacerlo aparecerá la pantalla de ejecución de procedimientos, que contiene una pestaña por cada procedimiento y en cada una los parámetros correspondientes. Al presionar el botón Enviar se iniciará la ejecución del procedimiento de la pestaña seleccionada, con los parámetros indicados. Durante la ejecución del procedimiento el análisis no podrá ser modificado, ni eliminado. Además será posible ver el log de ejecución desde el botón  Log en la pantalla de consulta.

Al finalizar la ejecución se podrán visualizar los archivos resultantes y los parámetros utilizados en la sección Resultados de Procedimientos en la ventana de edición o de visualización del Análisis (Figura 12.e). En dicha sección aparecerá un ítem desplegable por cada ejecución realizada en el Análisis.

**a)** Pantalla de consulta de Análisis de Genotipos.

**b)** Pantalla para crear o editar un Análisis de Genotipos.

**c)** Pantalla para editar los alelos del análisis.

**d)** Pantalla para ejecutar procedimientos de Análisis de Genotipos.

**e)** Sección de resultados de procedimientos de la pantalla de edición del análisis de genotipos.

**Figura 12.** Pantallas de Análisis de Genotipos: a) pantalla de consulta, b) pantalla de edición, c) pantalla de edición de alelos del análisis, d) pantalla de ejecución de procedimientos de análisis, e) sección de resultados de procedimientos de la pantalla de edición del análisis de genotipos, allí se visualizan los archivos resultantes y parámetros de la ejecución de los procedimientos.

## Procedimientos de Análisis

TEGA posee un módulo para la gestión de los procedimientos de análisis, es posible crear, editar y eliminar los procedimientos.

### Padrón de Procedimientos

El padrón de procedimientos se encuentra en *Análisis / Procedimientos*. Al crear un procedimiento es necesario indicar (Figura 13):

- Código: debe ser único y será utilizado para nombrar las carpetas de ejecución.
  - Nombre: utilizado en la pantalla de ejecución de procedimientos en los Análisis de Genotipos.
  - Detalle: descripción del procedimiento.
  - Comando de ejecución: comando que se utilizará para ejecutar el procedimiento.
- Ejemplo: ./script.sh

- Comando para mostrar el log de ejecución: este comando será utilizado cuando el usuario presione el botón  en la consulta de Análisis de Genotipos. Ejemplo: `./script_log.sh`
- Público: indica si el procedimiento puede ser visto por usuarios con el rol Invitado.
- Exportaciones: en esta sección se indica los archivos que generará TEGA al momento de ejecutar el procedimiento:
  - Genotipos (TEGA): Archivo de genotipos con el formato interno de TEGA. El nombre del archivo generado será: genotipos.tsv. Las columnas se encuentran separadas por tabs. Contiene las columnas:
    - conjunto\_id: ID del Conjunto de Muestras.
    - conjunto\_cod: Código del Conjunto de Muestras.
    - conjunto\_det: Detalle del Conjunto de Muestras.
    - muestra\_id: ID de la Muestra
    - muestra\_cod\_int: Código Interno de la Muestra
    - Luego posee una columna por cada Locus. En cada columna se indican los valores de los alelos, separandolos con el valor del parámetro de TEGA EXP\_PROC\_SEP\_ALELOS (por defecto "/"). El orden de los alelos es el indicado en el campo Índice. El valor de los alelos nulos es el indicado en el parámetro de TEGA EXP\_PROC\_VALOR\_NULO (por defecto "-9").
  - Genotipos (Structure): Archivo de genotipos con el formato básico de Structure. El nombre del archivo generado será: genotipos.str.
  - Genotipos (Genepop): Archivo de genotipos con el formato básico de Genepop [10]. El nombre del archivo generado será: genotipos.gen.
  - Muestras: Archivo con los campos básicos de las muestras. El nombre del archivo será: muestras.tsv.
  - Atrib. extras de Muestras: Archivo con los atributos extras de las muestras. El nombre del archivo será: muestras\_atributos.tsv. La columna ID corresponde al ID de la muestra.
  - Conjuntos de Muestras: Archivo con los datos de los Conjuntos de Muestras. El nombre del archivo será: conjuntos\_muestras.tsv.
  - Atrib. extras de Conjuntos de Muestras: Archivo con los atributos extras de los conjuntos de muestras. El nombre del archivo será: conjuntos\_muestras\_atributos.tsv. La columna ID corresponde al ID del conjunto de muestras.
  - Loci: Archivo con los campos básicos de los Loci. El nombre del archivo será: loci.tsv.
  - Atrib. extras de Loci: Archivo con los atributos extras de los loci. El nombre del archivo será: loci\_atributos.tsv. La columna ID corresponde al ID del locus.
  - Cantidades: archivo con la cantidad de ocurrencias de las entidades exportadas (Muestras, Loci, Alelos y Conjuntos de Muestras).

En la carpeta `ejemplos/[lenguaje]/proc` hay archivos de ejemplo de cada uno de estos archivos. En el caso de los archivos de atributos dinámicos, cada uno tiene cinco atributos, uno por cada tipo (carácter, numérico, entero, lógico y fecha).

- Archivos: los archivos necesarios para la ejecución del procedimiento. Los archivos deben ser cargados luego de crear el procedimiento.

**Crear o editar Procedimiento**

ID  
1

Código  
structure

Nombre  
Structure

Detalle

Comando de ejecución  
.structure.sh

Comando para mostrar el log de ejecución  
.log.sh

Público

**Exportaciones**

Genotipos (TEGA) Genotipos (Structure) Genotipos (GenePop)  
 Muestras Atrib. extras de Muestras Conjuntos de Muestras  
 Atrib. extras de Conjuntos de Muestras Loci Atrib. extras de Loci  
 Cantidades

**Archivos**

ID	Nombre	Detalle	Público
1	structure.sh	structure.sh	<input checked="" type="checkbox"/>
2	structure	structure	<input checked="" type="checkbox"/>
3	CLUMPP	CLUMPP	<input checked="" type="checkbox"/>

**Figura 13.** Pantalla de edición de los Procedimientos de análisis.

## Parámetros y Categorías de Procedimientos

Los Parámetros de Procedimientos le permiten al usuario configurar la ejecución del procedimiento, por ejemplo indicar la cantidad de repeticiones, el rango de grupos a detectar, el método estadístico a utilizar, etc. El padrón de estos parámetros se encuentra en *Análisis / Parámetros de Procedimientos*.

Al crear un Parámetro hay que indicar (Figura 14):

- Código: el código debe ser único.
- Detalle: descripción del parámetro.
- Tipo: tipo de dato del parámetro, los tipos disponibles son: Entero, Numérico, Carácter, Lógico y Fecha.
- Procedimiento: procedimiento al que estará vinculado el parámetro.
- Categoría: Categoría del parámetro. Las categorías permiten agrupar los parámetros de un procedimiento en secciones para facilitar su búsqueda al momento de ejecutar el procedimiento. El padrón de categorias se encuentra en: *Análisis / Categorías de Parámetros*.
- No Editable: Un parámetro no editable no puede ser modificado al momento de la ejecución del procedimiento.
- Valor: Valor por defecto del parámetro.

Crear o editar Parámetro de Procedimiento

<b>ID</b>	2
<b>Código</b>	STRUCTURE_BURNIN
<b>Detalle</b>	Length of burnin period
<b>Tipo</b>	Entero
<b>Procedimiento</b>	structure
<b>Categoría</b>	STRUCTURE_MAIN
<b>No editable</b>	<input checked="" type="checkbox"/>
<b>Valor</b>	10000
<input type="button" value="Cancelar"/> <input type="button" value="Guardar"/>	

**Figura 14.** Pantalla de edición de los Parámetros de Procedimientos.

## Pasos de la ejecución de Procedimientos

En esta sección se mencionan los pasos internos realizados por TEGA para la ejecución de los procedimientos (la ejecución se realiza desde la consulta de Análisis de Genotipos):

1. Se cambia el estado del Análisis de Genotipos de Disponible a Ejecutando.
2. Se crea una nueva ocurrencia de la entidad Ejecución, la cual será vinculada a los parámetros de procedimientos utilizados y a los archivos resultantes.
3. Se crea la carpeta destino donde se almacenarán los archivos de entrada y los archivos resultantes. La dirección de esta carpeta se almacenará en la Ejecución y será:  
[ARCHIVOS]/analisis\_genotipos/[ID\_ANALISIS]/[CODIGO\_PROC]\_[ID\_EJECUCION]  
Donde:
  - ARCHIVOS: es la carpeta indicada en el parámetro de TEGA DIRECTORIO\_ENTIDADES.
  - ID\_ANALISIS: el ID del Análisis de Genotipos seleccionado.
  - CODIGO\_PROC: código del procedimiento a ejecutar.
  - ID\_EJECUCION: ID de la Ejecución.
4. Se generan los archivos de entrada indicados en la sección Exportaciones del procedimiento a ejecutar.
5. Se genera el archivo con los parámetros del procedimiento indicados por el usuario. Este archivo contiene tanto los parámetros editables como los no editables. El nombre del archivo generado es: parametros.tsv. Posee las columnas:
  - categoria: Categoría del Parámetro

- parametro: Código del Parámetro.
  - tipo: Tipo del Parámetro. Valores posibles: CARACTER, ENTERO, NUMERICO, LOGICO y FECHA.
  - valor: Valor del Parámetro
6. Se ejecuta el comando para la ejecución de los procedimientos, indicado en el parámetro de TEGA COMANDO\_PROCEDIMIENTOS (por defecto ./proc/ejecutar\_procedimiento.sh). Este script recibe por parámetro:
- Comando a ejecutar del procedimiento (Comando de Ejecución)
  - Usuario del sistema operativo a utilizar para la ejecución (parámetro de TEGA USUARIO\_PROCEDIMIENTOS). En caso de que esté vacío se utilizará el usuario actual.
  - Dirección de la carpeta donde están almacenados los archivos del procedimiento.
  - Dirección destino donde se encuentran los archivos de entrada.
- El script realizará los siguientes pasos:
- a. Crear una carpeta temporal y copiar los archivos del procedimiento a esa carpeta.
  - b. En caso de indicar un usuario de ejecución, se asignará permiso de lectura y escritura al usuario en la carpeta temporal y la destino. También se le asignará el permiso de ejecución sobre los archivos de la carpeta temporal creada en a.
  - c. Se cambia el directorio actual al creado en a.
  - d. Se ejecuta el comando indicado. Se pasan los siguientes parámetros al comando de ejecución:
    - i. ID del Análisis de Genotipos
    - ii. Dirección absoluta a la carpeta temporal creada en a.
    - iii. Dirección absoluta a la carpeta destino.
  - e. Se elimina la carpeta temporal creada en a.
6. Al finalizar la ejecución del procedimiento se realizarán los siguientes pasos dentro de una misma transacción:
- Vinculación de los archivos de la carpeta destino a la Ejecución.
  - Cambio de estado del Análisis de Genotipos, de Ejecutando a Disponible.

## Procedimientos disponibles

La primer versión de TEGA posee cuatro implementaciones modestas de los procedimientos de análisis habituales en los estudios de genética de poblaciones.

### Structure

Este procedimiento busca determinar la estructura de la población utilizando el método de agrupamiento bayesiano implementado en Structure. El procedimiento permite indicar los parámetros de cada uno de los programas utilizados. Además es posible indicar el rango de K (cantidad de clusters) a estudiar y réplicas por K. El procedimiento tiene los siguientes pasos:

1. Ejecución de Structure para el rango de K y réplicas indicadas. Se utiliza la librería parallel para paralelizar la ejecución.
2. Ejecución de STRUCTURE HARVESTER [11] para determinar el K más adecuado. Se selecciona el de mayor Delta K.
3. Ejecución de CLUMPP [12] para integrar las réplicas del K seleccionado.
4. Ejecución de Distruct [13] para graficar la asignación de los clusters, tanto a los individuos como a las poblaciones.

Los archivos resultantes del procedimiento son:

- a. archivos\_exp.zip: contiene los archivos de entrada: cantidades.tsv, conjuntos\_muestras.tsv y parametros.tsv.
- b. structure.zip: contiene los archivos de entrada y logs de la ejecución de Structure.
- c. structure\_log.zip: contiene los archivos de logs de cada ejecución de Structure.
- d. structure\_out.zip: contiene los archivos de salida de Structure. Este archivo puede ser utilizado en otras plataformas como CLUMPAK [14].
- e. structureHarvester.zip: contiene archivos de salida de STRUCTURE HARVESTER.
- f. clumpp.zip: contiene los archivos de entrada y salida de CLUMPP.
- g. distruct.zip: contiene los archivos de entrada y salida de DISTSTRUCT.
- h. evanno.txt: posee la tabla con los valores de Delta K determinados con STRUCTURE HARVESTER.
- i. evanno.mk.txt: posee el K seleccionado por la plataforma (mayor Delta K).
- j. distruct.ind.ps: archivo en formato PostScript generado con Distruct donde se visualiza la asignación de cada grupo a cada individuo (generado a partir del K seleccionado y de la integración de las réplicas del K realizada con CLUMPP).
- k. distruct.pop.ps: archivo en formato PostScript generado con Distruct donde se visualiza la asignación de cada grupo a cada conjunto de muestras (generado a partir del K seleccionado y de la integración de las réplicas del K realizada con CLUMPP).
- l. ejecucion.log: log de la ejecución del procedimiento.

## Genepop

Este procedimiento realiza la ejecución de Genepop. Es posible indicar las opciones a ejecutar y los parámetros a utilizar. Los archivos resultantes son:

- genotipos.gen: archivo de genotipos en el formato de Genepop.
- settings.txt: archivo de entrada de Genepop que indica los parámetros y opciones de menú a ejecutar.
- parametros.tsv: parámetros del procedimiento.
- genepop.log: log de ejecución del procedimiento.
- genotipos.gen.\*: archivos de salida de Genepop correspondientes a las opciones de menú indicadas.

## DAPC

Este procedimiento busca determinar la estructura de la población utilizando el método multivariado DAPC (Discriminant analysis of principal components) [15]. El procedimiento tiene los siguientes pasos:

1. Determinación de K (cantidad de clusters): se utiliza la función *find.clusters* de adegenet [16]. Es posible indicar la cantidad de componentes principales (PC) a utilizar o el porcentaje de la varianza; el método estadístico a utilizar para determinar el mejor K y el criterio de selección de K, entre otros parámetros. Es posible desactivar este paso y realizar los siguientes pasos con los grupos determinados por los conjuntos de muestras.
2. Determinación de PCs: se determina la cantidad de PCs a utilizar en DAPC, esto se realiza con una validación cruzada, utilizando la función *xvalDapc* de adegenet. Es posible indicar la cantidad de repeticiones, el porcentaje del conjunto de entrenamiento, entre otros parámetros. El procedimiento determina la cantidad de PCs en dos pasos, primero se ejecuta *xvalDapc* con el rango de PCs por defecto, luego se ejecuta *xvalDapc* para un rango de PCs más acotado con centro en la cantidad de PCs seleccionada en el primer paso. La cantidad de PCs seleccionada es la que reporta el menor Error Medio Cuadrado.
3. Gráficos de dispersión y de contribución de variables: se realizan los gráficos de dispersión y de contribución de variables, usando las funciones *scatter* y *loadingplot* de adegenet. Además se grafican la probabilidad posterior de pertenencia a nivel de muestras y de conjunto de muestras.

Archivos resultantes:

- **archivos\_exp\_logs.zip**: contiene los archivos de entrada y logs del procedimiento (*dapc.log*, *genotipos.tsv*, *parametros.tsv*).
- **find\_clusters.zip**: contiene los archivos de salida de determinación de K.
  - *pc\_var.\**: Imagen y tabla de la cantidad de PCs retenidos por porcentaje de la varianza utilizada.
  - *k\_stat.\**: Imagen y tabla del valor del método utilizado para determinar el K, por K.
- **cv\_dapc.zip**: archivos de salida de la determinación de PCs (N indica si es la primera o segunda validación cruzada):
  - *cvN.cvr.csv*: tabla con el resultado de la validación cruzada.
  - *cvN.rmsenpp.csv*: Root Mean Squared Error por número de PC de PCA.
  - *cvN.png*: gráfico de proporción de aciertos por número de ejes de PCA.
- **dapc.zip**: archivos de salida de la ejecución de DAPC:
  - *pca\_scatter.png*: Gráfico de dispersión.
  - *pmp\_indplot.png*: Gráfico de probabilidad posterior de pertenencia a nivel de muestra.
  - *pmp\_boxplot.png*: Gráfico de probabilidad posterior de pertenencia a nivel de conjunto de muestras.
  - *loadingplot\_\*.png*: Gráficos de contribución de variables. Contribución de cada alelo a los componentes principales. El número en el nombre del archivo indica el componente.
  - *posterior.csv*: tabla con la probabilidad posterior de pertenencia de cada muestra a cada clusters.
  - *pmp.csv*: tabla transformada con la probabilidad posterior de pertenencia de cada muestra a cada clusters.
  - *var.contr.csv*: tabla con la contribución de cada variable a cada componente principal.

- k\_stat.png: Imagen del valor del método utilizado para determinar el K, por K.
- pca\_scatter.png: Gráfico de dispersión.
- pmp\_indplot.png: Gráfico de probabilidad posterior de pertenencia a nivel de muestra.
- pmp\_boxplot.png: Gráfico de probabilidad posterior de pertenencia a nivel de conjunto de muestras.
- resultados.csv: tabla con resultados de los métodos (K seleccionado en el paso a, cantidad de PCs seleccionados en el paso b, etc.)

## Índices y Validaciones

Este procedimiento genera varios índices descriptivos y gráficos. Existen parámetros para indicar las tablas y gráficos que se generarán. Los archivos generados son:

- archivos\_exp\_logs.zip: archivos de entrada y logs del procedimiento.
- alelos.zip: contiene tablas y gráficos de frecuencia y promedio de alelos por conjunto de muestras. Se utilizan funciones de poppr para los cálculos.
- allelic\_richness.csv: tabla con riqueza alélica por Loci y Conjuntos de Muestras. Solo disponible para loci diploides. Se utiliza la función *allelic.richness* de hierfstat
- basic\_stat.csv: tabla con índices básicos (Fst, Fis, Ho, Hs, etc.). Se utiliza la función *basic.stats* de hierfstat.
- missing.\*: tabla y gráfico de alelos perdidos. Se utiliza la función *info\_table* de poppr.
- mlg.\*: tabla y gráfico de la cantidad de genotipos multilocus (MLG). Se utiliza la función *mlg.table* de poppr.
- mlg\_rarefaction.png: curva de rarefacción calculada con MLG. Se utiliza la función *rarecurve* de vegan [17].
- ploidy.\*: tabla y gráfico de la ploidía por locus y muestra. Se utiliza la función *info\_table* de poppr.
- genotype\_curve.\*: tabla y gráfico de la curva de acumulación de genotipos. Se utiliza la función *genotype\_curve* de poppr.
- genotypic\_diversity.\*: tabla y gráfico con índices de diversidad. Se utiliza la función *poppr* de poppr.
- HW\_\*: tablas con test de equilibrio de Hardy-Weinberg. Se utiliza la función *hw.test* de pegas [18].
- locus\_table\_\*: tablas con estadísticas a nivel de loci. Se utiliza la función *locus\_table* de poppr.
- dist\_pop\_method\_\*: distancia entre conjuntos de muestras calcula con el método indicado por el usuario (ver parámetros del procedimiento). Se utiliza la función *dist.genpop* de adegenet.

## Ejemplos

TEGA posee archivos de muestras y de genotipos de ejemplo para probar los procedimientos. Estos archivos corresponden a los trabajos de Rueda et al 2013 y Kamvar et al 2014, si bien los ejemplos se centran en los procesos *Structure* y *DAPC*, también pueden ser utilizados para probar *Genepop* e *Índices y Validaciones*.

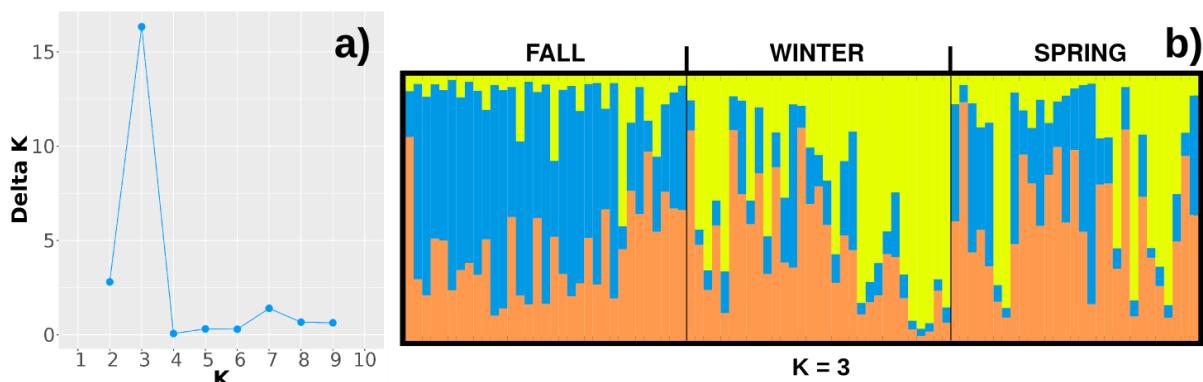
Structure: Rueda et al 2013

En el trabajo de Rueda et al 2013 se analizó la variación estacional en la estructura de la población de sábalos (*Prochilodus lineatus*) capturados en el río Uruguay, cerca de la ciudad Gualeguaychú (Entre ríos, Argentina). En este trabajo se utilizó Structure para analizar la estructura poblacional. En la carpeta *ejemplos/es/rueda\_et\_al\_2013* se encuentra el archivo de las muestras y el de alelos.

Los pasos para replicar el análisis con TEGA son:

1. Importar el archivo *muestras.csv*: la importación se realiza desde la consulta de muestras (ver la sección *Importación de Muestras*), se debe indicar la coma (,) como separador, y *Muestras* en Datos a Importar. Al importar las muestras se crearán automáticamente los Conjuntos de Muestras.
2. Importar el archivo *genotipos.csv*: la importación se realiza desde la consulta de Alelos (ver sección *Importación de Loci y Alelos*), se debe indicar el tipo de importación Matriz, la coma (,) como separador y “?” como valor nulo.
3. Crear un nuevo Análisis de Genotipos (ver sección *Análisis de Genotipos*): en la creación del análisis seleccionar los Loci: PL3, PL14, PL25, PL34, PL35, PL64, PL119, PL139 y PL216; y los Conjuntos de Muestras: RUEDA13|WINTER, RUEDA13|SPRING y RUEDA13|FALL.
4. Ejecutar el procedimiento Structure (ver sección Análisis de Genotipos): se deben indicar los parámetros (el resto por defecto):
  - STRUCTURE\_K\_HASTA: 10
  - STRUCTURE\_REPLICAS: 10
  - STRUCTURE\_BURNIN: 150000
  - STRUCTURE\_NUMREPS: 300000

Como resultado se aprecia que la cantidad de clusters óptima ( $K = 3$ ) y la asignación de cada uno a las muestras se corresponde con el trabajo de Rueda et al 2013 (Figura 15).



**Figura 15.** Resultados del procedimiento Structure con los datos de Rueda et al 2013. a) Gráfico de Delta K (archivo: *evanno.deltaK.png*). b) Asignación de la probabilidad de pertenencia de cada individuo a cada uno de los tres clusters (archivo: *distruct.ind.ps*).

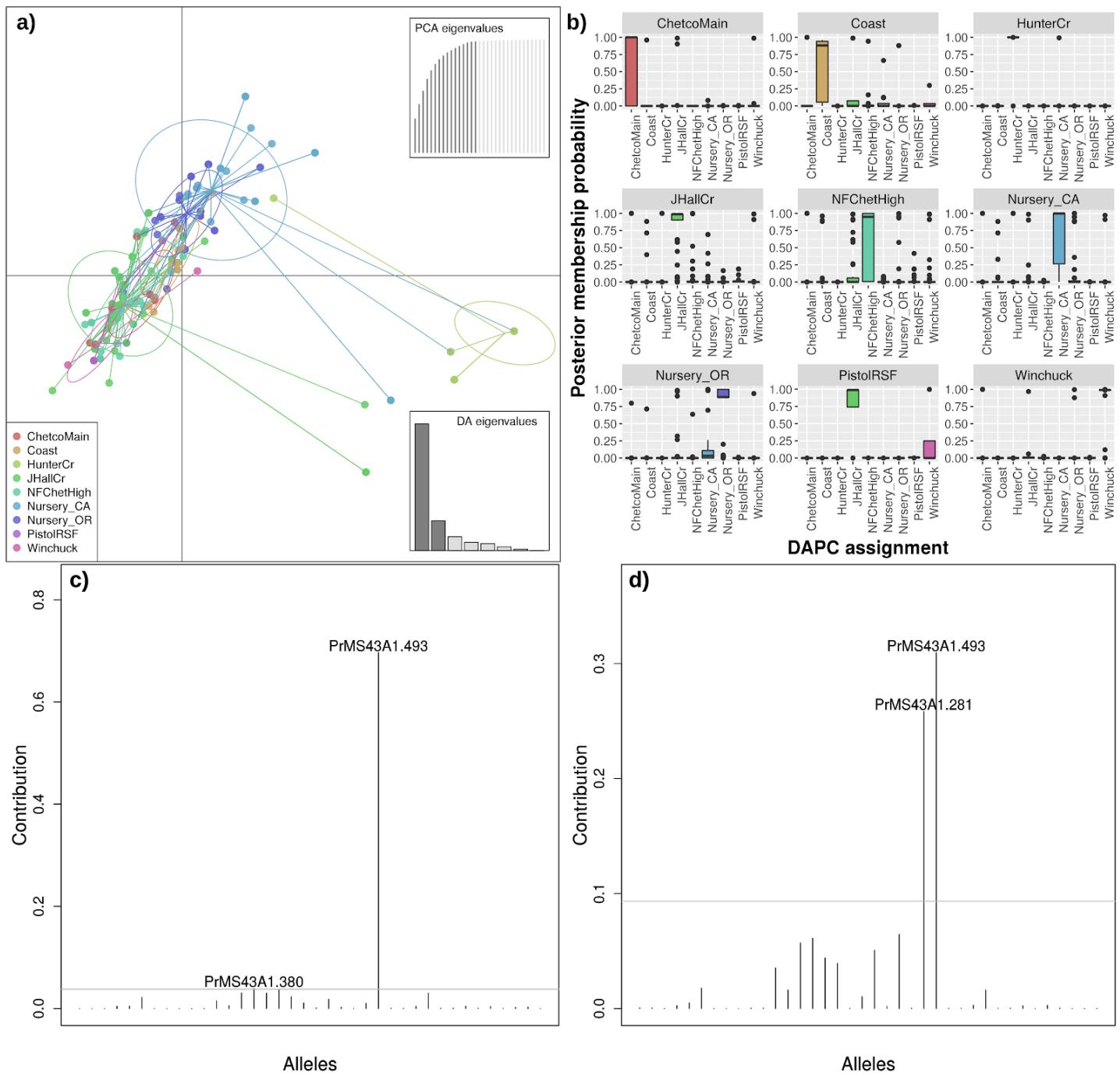
## DAPC: Kamvar et al 2014

Los datos de Kamvar et al 2014 fueron utilizados en el trabajo Kamvar et al 2015 [19], donde analizaron la estructura poblacional de *Phytophthora ramorum* en los bosques de Oregon y California, utilizando DAPC. En la carpeta `ejemplos/es/Kamvar_et_al_2014` se encuentra el archivo de muestras y el de alelos.

Los pasos para replicar el análisis con TEGA son:

1. Importar el archivo `muestras.csv`: la importación se realiza desde la consulta de muestras (ver la sección *Importación de Muestras*), se debe indicar la coma (,) como separador, y *Muestras* en Datos a Importar. Al importar las muestras se crearán automáticamente los Conjuntos de Muestras.
2. Importar el archivo `genotipos.csv`: la importación se realiza desde la consulta de Alelos (ver sección *Importación de Loci y Alelos*), se debe indicar el tipo de importación Matriz, la coma (,) como separador y "?" como valor nulo.
3. Crear un nuevo Análisis de Genotipos (ver sección *Análisis de Genotipos*): en la creación del análisis seleccionar los Loci: *PrMS6A1*, *Pr9C3A1*, *PrMS39A1*, *PrMS45A1* y *PrMS43A1*; y los Conjuntos de Muestras: *JHallCr*, *NFChetHigh*, *Coast*, *HunterCr*, *Winchuck*, *ChetcoMain*, *PistolRSF*, *Nursery\_CA* y *Nursery\_OR*.
4. Ejecutar el procedimiento DAPC (ver sección Análisis de Genotipos): se deben indicar los parámetros (el resto por defecto):
  - 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

Como resultado se aprecia que *HunterCr* se diferencia del resto de los Conjuntos de Muestras (Figura 16.a) y que la asignación de DAPC a cada muestra se correspondería a los Conjuntos de Muestras (sin considerar *PistolRSF*) (Figura 16.b). Además el alelo 493 del locus *PrMS43A1* sería el que más contribuye a esa la diferencia entre los conjuntos (Figura 16.c y 16.d). Estos resultados serían similares a los de Kamvar et al 2015.



**Figura 16.** Resultados del procedimiento DAPC con los datos de Kamvar et al. 2014. a) Gráfico de Dispersión (archivo *pca\_scatter.png*). b) Diagrama de cajas realizado a partir de la probabilidad de pertenencia de cada muestra a cada Conjunto de Muestras (archivo *pmp\_boxplot\_assig.png*). Contribución de los alelos al primer (c) y segundo (d) componente principal del análisis (archivos *loadingplot\_1.png* y *loadingplot\_2.png*). La mayor contribución la realiza el alelo 493 del locus PrMS43A1.

## Usuarios

En esta sección se explica cómo se realiza la gestión de usuarios en TEGA.

Como TEGA está enfocada a utilizarse durante el periodo de investigación, en el que se gestionan y analizan los datos, es necesario asegurar la privacidad de los datos hasta el momento de su publicación. Por este motivo existen distintos roles de usuarios:

- Anónimo: Tiene acceso a los datos públicos de las Muestras. Solo tiene permiso para visualizar los datos, no puede modificarlos ni eliminarlos.
- Invitado: Tiene acceso a los datos públicos de las Muestras, Alelos, Locus, Procedimientos y Análisis de Genotipos. Solo tiene permiso para visualizar los datos, no puede modificarlos ni eliminarlos.
- Investigador: Tiene acceso a los datos públicos y privados de la plataforma, puede realizar acciones ABMCE y ejecutar análisis de genotipos.
- Administrador: Tiene los mismos privilegios que el rol Investigador y también tiene acceso a la gestión de usuarios y configuraciones de la plataforma.
- Implementador: Rol destinado para los implementadores de la plataforma en la institución.

Los roles Anónimo e Invitado están destinados a las personas ajenas al grupo de investigación. Cuando una persona se registra en la plataforma se le asigna el rol indicado en el parámetro de TEGA *ROL\_USUARIO\_DEFECTO*, el cual puede ser Anónimo (ANONIMO) o Invitado (INVITADO, valor por defecto).

Los roles de Administrador e Investigador serían utilizados solo por el grupo de investigación.

Los usuarios con el rol de Administrador pueden acceder al padrón de Usuarios (opción de menú *Administración / Gestión de Usuarios*) (Figura 17). Allí pueden:

- Consultar los usuarios registrados en la plataforma, de forma análoga a otros padrones.
- Crear, Editar y Eliminar usuarios. Si bien un administrador puede crear un nuevo usuario, el flujo recomendado de creación de usuarios es:
  - a. La persona interesada en acceder a la plataforma debe crear su usuario en la pantalla de inicio (link “Crear una cuenta”, o la opción de menú *Cuenta / Crear una cuenta*).
  - b. La plataforma le enviará un mail a la dirección de mail indicada y la persona debe confirmar la registración (haciendo click en el link enviado en el mail). En esta instancia el usuario tendrá el rol de Anónimo.
  - c. El administrador debe asignar el rol deseado al usuario creado. Esto lo puede realizar desde la pantalla de edición del usuario, en el padrón de usuarios.
  - d. Una vez modificado el rol, si la persona tenía una sesión iniciada en la plataforma, debe cerrar la sesión y volver a iniciarla para visualizar los cambios.
- Tanto desde la pantalla de consulta, como desde la de edición, es posible cambiar el estado del usuario de Activado a Desactivado y viceversa. Un usuario desactivado no puede iniciar sesión en la plataforma.

Crear o editar un usuario

ID  
5

Login  
investigador

Nombre

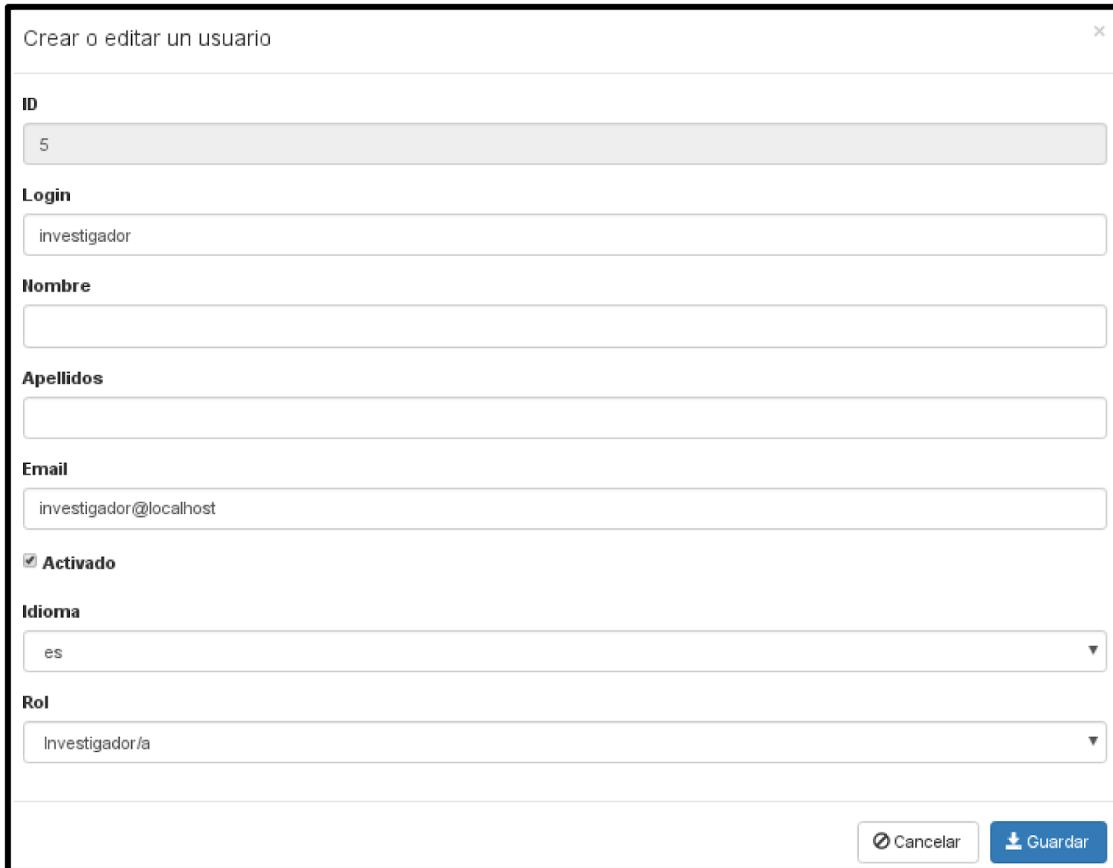
Apellidos

Email  
investigador@localhost

Activado

Idioma

Rol



**Figura 17.** Pantalla de edición de usuarios.

## Publicación de datos

En esta sección se explica cómo cambiar el estado (de privado a público y viceversa) de los datos masivamente.

Luego de publicar los resultados de una investigación los usuarios de la plataforma (con el rol de Investigador o Administrador) pueden cambiar el estado del proyecto vinculado a la publicación, pasándolo a público (Figura 18). Esto provocará que las Muestras, Alelos, Loci y Análisis de Genotipos vinculados al proyecto puedan ser accedidos por usuarios ajenos a la institución (usuarios con los roles Anónimo o Invitado).

**Crear o editar Proyecto**

ID	1
Código	PROYECTO 1
Fecha	2018-01-08 <input type="button" value=""/>
Detalle	PROYECTO 1
Público	<input checked="" type="checkbox"/>
Atributos	
Archivos	

**Figura 18.** Pantalla de edición de proyectos.

## Parámetros de TEGA

En esta sección se explica cómo se realiza la gestión de los parámetros de TEGA. Estos parámetros permiten modificar ciertas configuraciones de la plataforma, por ejemplo, el contenido de la pantalla de inicio, el logo, el carácter utilizado como separador en la exportaciones, etc. Los parámetros pueden ser modificados desde el padrón de parámetros (opción de menú *Administración / Parámetros de TEGA*), al cual solo pueden acceder los usuarios con el rol Administrador o Implementador (Figura 19).

La modificación o eliminación de un parámetro puede provocar errores en la plataforma, por lo que recomendamos modificarlos solo cuando se comprenda bien su impacto. A su vez, también recomendamos tener un ambiente de testing para probar distintas configuraciones.

Los parámetros poseen los siguientes campos:

- Código: el código del parámetro, debe ser único.
- Detalle: detalle del uso del parámetro en la plataforma.
- Tipo: tipo del valor del parámetro
  - Numérico (ej. 10.13)
  - Entero (ej. 1014)
  - Carácter (max. 10485760) (ej. “ejemplo”).
  - Lógico. La plataforma utiliza una casilla para los valores lógicos (en la base de datos se almacena como 1 y 0, verdadero y falso respectivamente).
  - Fecha (ej. 2017-01-26).
- Público: indica si un usuario anónimo puede acceder al parámetro. Utilizado principalmente para los parámetros relacionados a la pantalla de inicio.

- Valor: valor del parámetro, solo puede quedar vacío si el tipo de parámetro es Carácter.

Luego de modificar un parámetro, para visualizar el cambio es necesario recargar TEGA en el navegador (shift + F5).

Crear o editar Parámetro de TEGA

**ID**  
5

**Código**  
NOMBRE\_PLATAFORMA

**Detalle**  
Nombre de la plataforma

**Tipo**  
Caracter

**Público**

**Valor**  
NOMBRE\_PLATAFORMA

**Figura 19.** Pantalla de edición de parámetros.

## Copias de Seguridad

TEGA posee un procedimiento básico para la generación de copias de seguridad (o backups). La configuración del procedimiento se menciona en la sección *Instalación y Configuración de la plataforma*.

### Generación

Es posible generar una copia de seguridad desde la opción de menú *Administración / Copia de Seguridad*, solo los usuarios con el rol de Administrador o Implementador pueden acceder a esta opción. Por otro lado, también es posible ejecutar el procedimiento desde un programador de tareas, como cron. Para ello es necesario ejecutar el procedimiento de la siguiente forma:

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

Donde:

- DIR\_TEGA es la dirección a la carpeta de TEGA
- CRON: es el parámetro que le indica al procedimiento que la ejecución es automática.

La copia de seguridad incluye:

- Todos los archivos ubicados en la carpeta de TEGA (se excluyen las copias de seguridad).
- Copia de la base de datos.

Los archivos a incluir en la copia de seguridad son comprimidos y luego encriptado con *pgp*.

## Restauración

La restauración de una copia de seguridad es posible realizarla ejecutando el script:

*[DIR\_TEGA]/scripts/restore.sh [COPIA\_DE\_SEGURIDAD] [CLAVE\_ENCRYPTACIÓN]  
[HOST\_DB] [USUARIO\_BD] [CLAVE\_BD] [BD] [ENCODING] [LC\_COLLATE] [LC\_CTYPE]*

Donde:

- DIR\_TEGA es la dirección a la carpeta de TEGA
- COPIA\_DE\_SEGURIDAD: dirección y nombre del archivo de la copia de seguridad.
- CLAVE\_ENCRYPTACIÓN: clave de encriptación del archivo (indicada en el archivo de configuración del backup).
- HOST\_DB: host del motor de la base de datos (localhost).
- USUARIO\_BD: usuario con el que se creará la BD y se importarán los datos.
- CLAVE\_BD: clave del usuario de la BD.
- BD: nombre de la base de datos a crear.
- ENCODING, LC\_COLLATE y LC\_CTYPE: parámetros del comando CREATE DATABASE de PostgreSQL.

Este procedimiento realiza los siguientes pasos:

1. Desencripta la copia de seguridad.
2. Descomprime la copia de seguridad en el directorio actual.
3. Si los parámetros de la base de datos son indicados (HOST\_DB, USUARIO\_DB, etc.):
  - a. Elimina, si existe, la base de datos indicada.
  - b. Crea una nueva base de datos con el nombre indicado.
  - c. Importa los datos en la base de datos.

En caso de utilizar un usuario o base de datos distintas a la configuración de TEGA es necesario actualizar manualmente la configuración de TEGA (archivos config/\*.yml). Luego es posible ejecutar TEGA como se indica en la sección *Instalación y Configuración de la plataforma*. Debe asegurarse de que se encuentren instaladas las librerías y herramientas mencionadas en la sección *Requerimientos*.

# Desarrollo

En esta sección se mencionan los aspectos generales del desarrollo de TEGA.

## Aspectos generales

Para el desarrollo de TEGA se utilizó el generador de proyectos JHipster (<http://www.jhipster.tech>). En el Back-End se utilizó el lenguaje JAVA con el framework Spring Boot (<https://projects.spring.io/spring-boot>), y el motor de bases de datos PostgreSQL (<https://www.postgresql.org>). Mientras que en el Front-End se utilizó el lenguaje JavaScript con AngularJS (<https://angularjs.org>) y Bootstrap (<http://getbootstrap.com>). La totalidad de las librerías utilizadas puede ser consultada en el repositorio de TEGA (<https://github.com/darioelias/TEGA>).

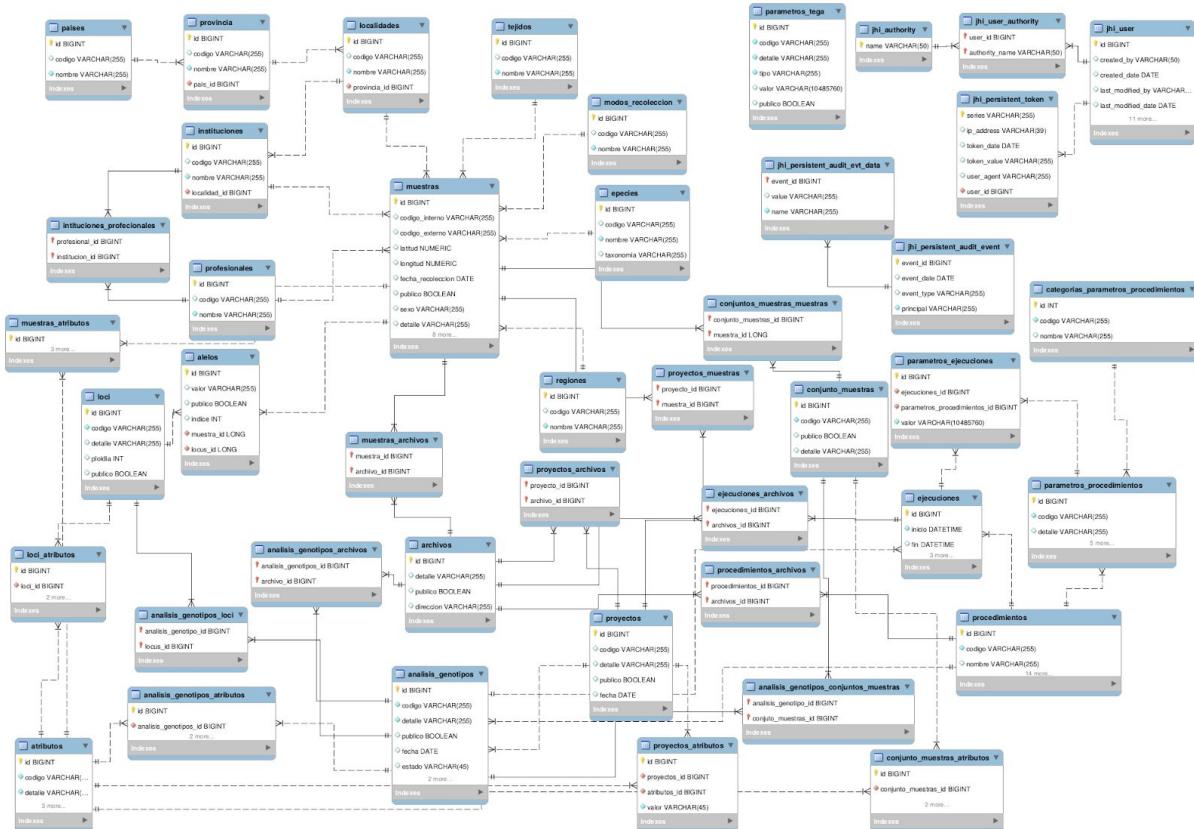
## Instalación de fuentes

Pasos para instalar los fuentes de TEGA:

1. Instalar los requerimientos de la sección *Requerimientos*
2. Instalar JAVA 8 Development Kit (JDK). Guía de instalación de OpenJDK 8: <http://openjdk.java.net/install/>
3. Instalar git v2.1.4. Guía de instalación de git: <https://git-scm.com/download/linux>
4. Instalar MAVEN (<https://maven.apache.org>)
5. Instalar Node.js v4.7.0 (<https://nodejs.org/en>)
6. Actualizar NPM. Comando: `npm install -g npm`
7. Instalar bower y gulp-cli. Comando: `npm install -g bower gulp-cli`
8. Clonar el repositorio de TEGA de github. Comando:  
`git clone git://github.com/darioelias/TEGA.git`
9. Copiar la carpeta `config_ejemplo` y renombrarla por `config`
10. Crear una base de datos y configurar TEGA. Pasos 3 a 10 de la sección *Instalación y configuración de la plataforma*. Además de configurar el archivo `application-prod.yml`, configurar el archivo `application-dev.yml`.
11. Ejecución en desarrollo: el script `dev_comp.sh` permite ejecutar maven con la configuración de desarrollo.
12. Ejecución en producción: el script `prod_comp.sh` permite ejecutar maven con la configuración para construir y empaquetar la plataforma (generando el archivo WAR en la carpeta `target`). Para más información consultar la documentación de JHipster o el archivo `jhipster.readme.txt`.

# Diagrama de entidad relación

El diagrama de entidad relación solo es orientativo (Figura 20) y puede ser visualizado con MySQL Workbench (<https://www.mysql.com/products/community>). El archivo del modelo se encuentra en la carpeta “modelo” del repositorio de git de TEGA.



**Figura 20.** Diagrama de entidad relación de TEGA visualizado con MySQL Workbench. El diagrama es solo orientativo.

# Bibliografía

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, Donnelly 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 Szöecs 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.

## **Historial de cambios**

14/05/2018: Creación del manual. Versión 1.0.0

11/01/2019: Correcciones menores del texto. Agregado de la sección Imagen de VirtualBox.

Versión 1.0.1

# Anexo: GNU Free Documentation License

GNU Free Documentation License

Version 1.3, 3 November 2008

Copyright (C) 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc.

<<https://fsf.org/>>

Everyone is permitted to copy and distribute verbatim copies  
of this license document, but changing it is not allowed.

## 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. Secondly, 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.