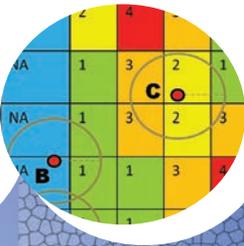
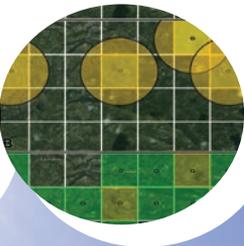


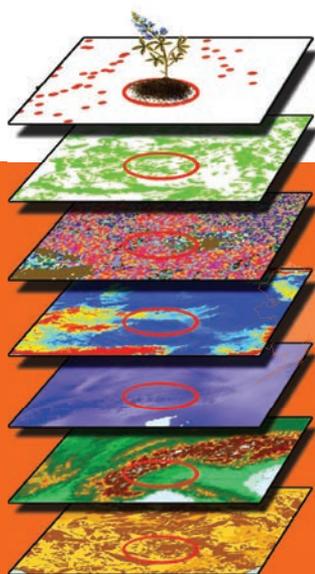
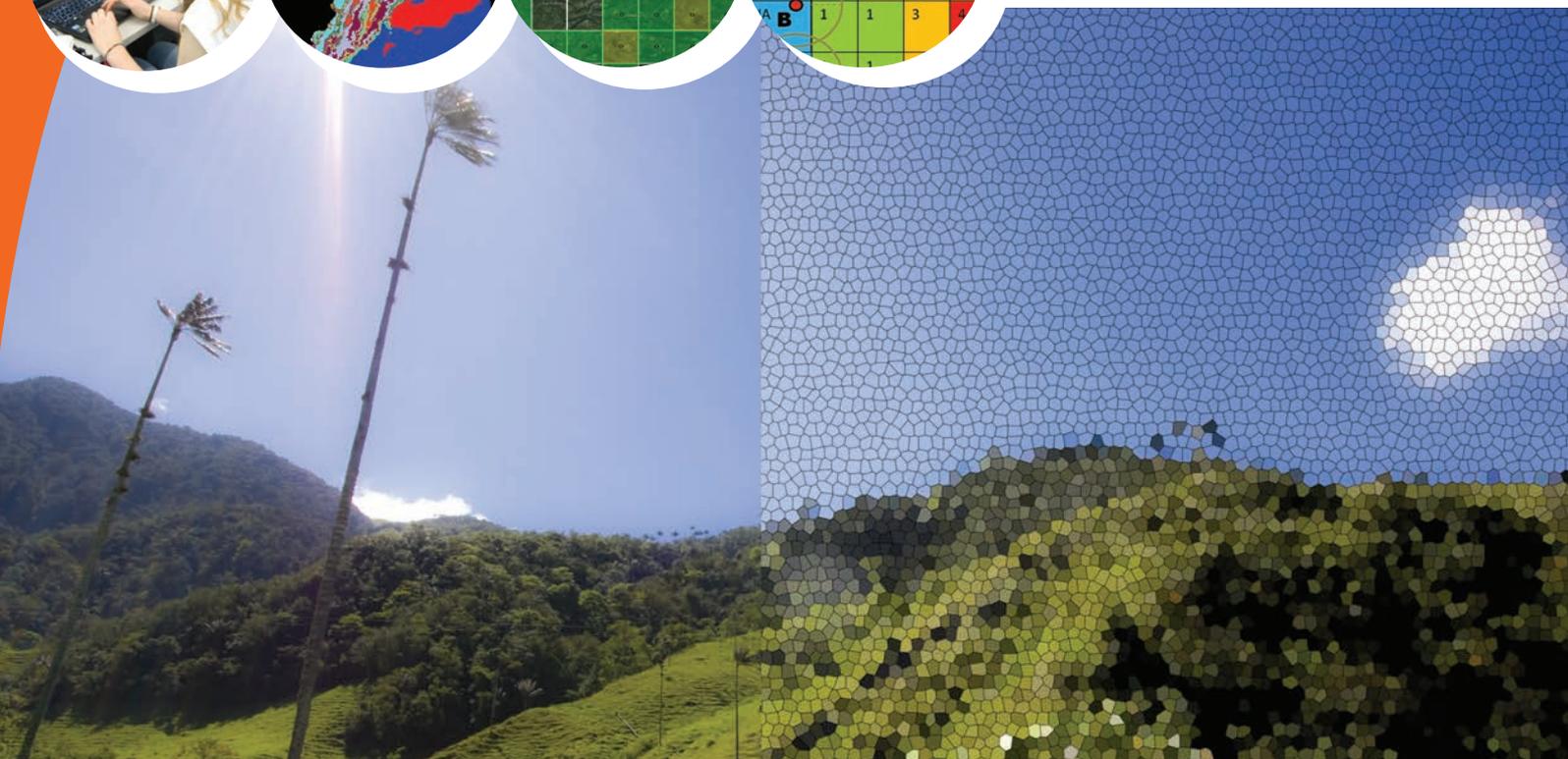


Tratado Internacional

SOBRE LOS RECURSOS FITOGENÉTICOS PARA LA ALIMENTACIÓN Y LA AGRICULTURA



*Herramientas*



# CAPFITOGEN

Programa para el Fortalecimiento  
de las Capacidades en Programas  
Nacionales de Recursos  
Fitogenéticos de América Latina

Versión 1.2



*Herramientas*

# CAPFITOGEN

**Programa para el Fortalecimiento  
de las Capacidades en Programas  
Nacionales de Recursos  
Fitogenéticos de América Latina**

**Versión 1.2**

**Autor de las herramientas:**

Mauricio Parra Quijano

*Consultor*

*Tratado Internacional sobre los Recursos Fitogenéticos  
para la Alimentación y la Agricultura, TIRFAA*

*FAO*

**Autores del manual acompañante:**

Mauricio Parra Quijano

Elena Torres Lamas, *Universidad Politécnica de Madrid (España)*

José María Iriondo Alegría, *Universidad Rey Juan Carlos (España)*

Francisco López, *Oficial de Apoyo del Tratado, TIRFAA, FAO*

Las denominaciones empleadas en este producto informativo y la forma en que aparecen presentados los datos que contiene no implican, por parte de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), juicio alguno sobre la condición jurídica o nivel de desarrollo de países, territorios, ciudades o zonas, o de sus autoridades, ni respecto de la delimitación de sus fronteras o límites. La mención de empresas o productos de fabricantes en particular, estén o no patentados, no implica que la FAO los apruebe o recomiende de preferencia a otros de naturaleza similar que no se mencionan.

Las opiniones expresadas en este producto informativo son las de su(s) autor(es), y no reflejan necesariamente los puntos de vista o políticas de la FAO.

ISBN 978-92-5-308255-1 (edición impresa)

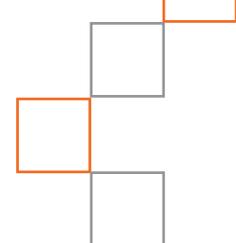
E-ISBN 978-92-5-308256-8 (PDF)

© FAO, 2014

La FAO fomenta el uso, la reproducción y la difusión del material contenido en este producto informativo. Salvo que se indique lo contrario, se podrá copiar, descargar e imprimir el material con fines de estudio privado, investigación y docencia, o para su uso en productos o servicios no comerciales, siempre que se reconozca de forma adecuada a la FAO como la fuente y titular de los derechos de autor y que ello no implique en modo alguno que la FAO aprueba los puntos de vista, productos o servicios de los usuarios.

Todas las solicitudes relativas a la traducción y los derechos de adaptación así como a la reventa y otros derechos de uso comercial deberán dirigirse a [www.fao.org/contact-us/licence-request](http://www.fao.org/contact-us/licence-request) o a [copyright@fao.org](mailto:copyright@fao.org).

Los productos de información de la FAO están disponibles en el sitio web de la Organización ([www.fao.org/publications](http://www.fao.org/publications)) y pueden adquirirse mediante solicitud por correo electrónico a [publications-sales@fao.org](mailto:publications-sales@fao.org).



## Índice

<b>1.</b>	Programa para el Fortalecimiento de las Capacidades en Programas Nacionales de Recursos Fitogenéticos en América Latina	5
<b>2.</b>	Herramientas CAPFITOGEN: características e instalación	9
<b>3.</b>	Herramienta GEOQUAL	23
<b>4.</b>	Herramienta ELC mapas	35
<b>5.</b>	Herramienta ECOGEO	49
<b>6.</b>	Herramienta Representa	59
<b>7.</b>	Herramienta DIV mapas	71
<b>8.</b>	Herramienta ColNucleo	95
<b>9.</b>	Herramienta FIGS_R	105
<b>10.</b>	Errores frecuentes	119
<b>11.</b>	Créditos	127
<b>12.</b>	Anexos	133



# 1. Programa para el Fortalecimiento de las Capacidades en Programas Nacionales de Recursos Fitogenéticos de América Latina

Bajo el auspicio del Tratado Internacional sobre los Recursos Fitogenéticos para la Alimentación y la Agricultura (TIRFAA) y la Agencia Española de Cooperación Internacional para el Desarrollo (AECID), entre los años 2008 y 2010 se desarrollaron dos talleres para la implementación del TIRFAA en el grupo de los países de América Latina y el Caribe (GRULAC) en Cartagena de Indias (Colombia, julio-agosto de 2008) y La Antigua (Guatemala, agosto de 2010). El éxito de esta actividad demostró la efectividad de este tipo de talleres en contribuir a la implementación de los objetivos del TIRFAA dentro de la comunidad de países GRULAC. La coordinación de los talleres entre la Secretaría del TIRFAA, la AECID, el Ministerio de Medio Ambiente y Medio Rural y Marino y el Centro Nacional de Recursos Fitogenéticos del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria CRF-INIA ha sido un factor decisivo en la consecución de los objetivos, dado el firme compromiso de España con el TIRFAA y los estrechos lazos existentes entre los Programas Nacionales de Conservación de Recursos Fitogenéticos e Institutos Nacionales de Investigación Agrícola (INIAs) de los países del GRULAC y España.

La buena experiencia cosechada en talleres previos y la importancia de impulsar algunos objetivos del TIRFAA en América Latina y el Caribe, en particular los contenidos en los artículos 5, 6, 7, 8 y 13.2 c, representó un incentivo clave para dar continuidad a este tipo de actividades. A la vez se hizo evidente la necesidad de dar mayor profundidad y contenido técnico a la continuidad a los talleres, estableciendo un programa de transferencia de tecnología, donde los talleres representan un elemento dentro de una estrategia de acción más amplia.

Tomando como base estos antecedentes y las necesidades de la región fue implementado el Programas para el Fortalecimiento de las Capacidades en Programas Nacionales de Recursos Fitogenéticos de América Latina - CAPFITOGEN. Este programa está enfocado en el desarrollo de tecnologías apropiadas para países con abundante agrobiodiversidad y limitados recursos económicos. Su función es la de desarrollar la tecnología, transferirla y proveer formación adecuada al personal técnico de los países de América Latina que son parte del Tratado.

La buena acogida de las herramientas y metodologías que se han desarrollado bajo el auspicio del Programa CAPFITOGEN en 2013, ha hecho que en algunos países objetivo del programa se hayan organizado talleres de índole nacional por iniciativa y búsqueda de financiación de los propios interesados. Al mismo tiempo se ha producido interés desde otros países y regiones que no eran objeto inicial



del Programa, quienes han venido demandando tanto las herramientas como actividades de transferencia y entrenamiento.

De esta forma el Programa CAPFITOGEN espera servir no sólo como generador y transferidor de tecnología apropiada, sino como un modelo de transferencia en sí mismo. Uno de sus aspectos más innovadores es el hecho de involucrar a quienes han desarrollado las metodologías científicas, a desarrollar las herramientas facilitadas basadas en sus metodologías y a realizar directamente las actividades de transferencia y entrenamiento técnico. Este modelo garantiza a los beneficiarios del Programa acceso directo a los científicos y desarrolladores para la resolución de sus dudas o exposición de sus casos. A la vez, los científicos se nutren directamente de las experiencias y problemáticas de los técnicos de los Programas Nacionales, lo cual incidirá en futuras investigaciones de mayor aplicación y más ajustadas a las necesidades reales.



## 2. Herramientas CAPFITOGEN: características e instalación

### 2.1. Origen

Las plantas cultivadas fueron en su día plantas silvestres cuya amplia batería genética les permitía resistir y adaptarse a los retos que un ambiente cambiante les planteaba, como una nueva plaga, una enfermedad, un herbívoro, una sequía, etc. El proceso de domesticación transformó estas plantas silvestres en los granos, legumbres, frutas y verduras que hoy conocemos; unos productos a la medida de las necesidades y gustos de los seres humanos. Pero para lograr estos productos hubo que acometer un intenso proceso de selección. En consecuencia, la base genética de las especies domesticadas es bastante reducida si la comparamos con el de sus ancestros silvestres.

La evolución de las especies cultivadas ha ido dejando productos entre los cuales también encontramos diferencias. Por ejemplo, las variedades e híbridos modernos, producidos a partir de la década de los años 60, son materiales con un potencial productivo muy alto pero de base genética demasiado estrecha si comparamos con las variedades de principio de siglo XX.

Así la domesticación, un proceso netamente selectivo, ha ido dejando en el camino muchas configuraciones genéticas en pos de obtener la "mejor" variedad. En ese proceso han desaparecido muchos genes valiosos que podrían aportar soluciones ante futuros problemas con nuestras productivas, homogéneas y vulnerables variedades modernas. Afortunadamente no todo se ha eliminado de los campos y agricultores de todo el mundo siguen conservando variedades heredadas de sus ancestros, probablemente no comerciales, pero con valores muy importantes para su cultura, hábitos alimenticios o incluso religiosos.

Conscientes de la pérdida progresiva de ese patrimonio genético, muchas naciones han venido rescatando y conservando esas variedades y plantas silvestres emparentadas, depositándolas en colecciones de germoplasma (semillas, tejidos, propágulos) desde los años 50.

#### 2.1.1 Conservar para conocer y utilizar

Las colecciones de germoplasma se diferencian esencialmente de una colección de museo en que lo conservado es para ser utilizado. El principal usuario son los mejoradores genéticos de cultivos, quienes buscan allí fuentes de caracteres de interés para ser transferidos a variedades modernas. Sin embargo, el uso eficiente y efectivo del germoplasma sólo es posible mediante el conocimiento que podamos generar del mismo. La obtención de ese conocimiento se lleva a cabo a través de

los procesos de caracterización y evaluación, los cuales requieren grandes esfuerzos en términos económicos y logísticos por parte de las entidades responsables de conservar la agrobiodiversidad. En el nivel nacional estas actividades se llevan a cabo a través de los Programas Nacionales y para el caso de Latinoamérica estos Programas están usualmente a cargo de los Institutos Nacionales de Investigación Agraria (INIAs).

### 2.1.2 Técnicas adecuadas para escenarios con escasos recursos

Los métodos para recolectar, conservar y caracterizar la agrobiodiversidad con estándares científicos, han venido usualmente desde regiones y centros donde los recursos económicos, la infraestructura o la capacitación del personal no representan mayores limitaciones. Esto ha generado que en países en desarrollo estas metodologías o no se puedan aplicar, o si se aplican, no se puedan extender a todo el germoplasma conservado. Una situación que contrasta con el hecho de que es en los países en desarrollo donde se concentra la mayor riqueza genética agrícola vegetal.

Este escenario ha conducido a que algunos grupos de investigación en el mundo hayan dirigido sus esfuerzos a explorar metodologías de menor coste y complejidad, que se adapten mucho mejor a las condiciones de Programas Nacionales de países en desarrollo. Las alternativas metodológicas incluyen el aprovechamiento de la información ambiental de los sitios de colecta (ecogeográfica), por ejemplo, para hacer estimaciones de la variabilidad genética que contendría el germoplasma o estimar con mayor éxito la probabilidad de encontrar genes de interés. Así mismo, se plantea el uso de los sistemas de información geográfica como medio de obtención y aprovechamiento de dichos datos ecogeográficos. Dado que la mayor parte de la información ecogeográfica y los programas informáticos para realizar los análisis son de acceso gratuito, la inversión se reduce a un ordenador con una configuración de tipo comercial y capacitación del personal. Se trata entonces de métodos compatibles con escenarios de recursos limitados, condición que ocurre a menudo en Programas Nacionales de países en desarrollo. Después de demostrar su efectividad en casos de estudio publicados en revistas científicas internacionales, esta nueva generación de metodologías fueron elegidas para ser adaptadas en forma de herramientas simplificadas y prácticas, y ser finalmente transferidas a Programas Nacionales de las regiones seleccionadas.

## 2.2 Características

Los procesos que engloban las aplicaciones de la ecogeografía y los sistemas de información geográfica (SIG) en la conservación y uso eficiente de los recursos fitogenéticos, tal y como han sido publicados en diversos medios de divulgación científica, implican la utilización de numerosos programas informáticos, algunos de ellos de no fácil acceso o no gratuitos, en secuencias metodológicas complejas. Por

tanto, el desarrollo de herramientas prácticas que permitan aplicar estos avances por parte de personal técnico sin mayor formación en manejo y programación de programas estadísticos y SIG, constituye el principal reto.

La solución reside en la utilización de un programa informático que reúna las siguientes características:

- a) Incluir todos los análisis estadísticos requeridos.
- b) Incluir todas las funciones de los SIG, tanto para el manejo de datos georreferenciados como para su análisis.
- c) Manejar bases de datos y los productos de los análisis estadísticos y de los SIG.
- d) Permitir guardar todos los resultados de los análisis en formatos informáticos convencionales.
- e) Ser de libre acceso y distribución.

En la actualidad, la comunidad científica dispone del lenguaje informático de programación llamado "R" (R. Cran, 2012) que cumple con las características necesarias para el desarrollo de las herramientas CAPFITOGEN. El lenguaje "R" es muy potente para el análisis pero cuenta con el principal inconveniente de que requiere conocimientos técnicos específicos para poder programar su funcionamiento. El Programa CAPFITOGEN ha superado este inconveniente involucrando a los equipos de investigación responsables de las metodologías originales en el desarrollo de las herramientas, en particular, en la programación de R. La forma en que se programa R busca integrar todas las funciones y los análisis requeridos bajo un mismo programa pero de manera que funcione con unos formatos de datos predefinidos, a través de los cuales se pueda introducir los datos de cualquier usuario en las rutinas de R y se logren resultados particulares.

El último desafío en la popularización del uso de estas herramientas era la simplificación de la forma en la que se introducen órdenes y datos y se ha afrontado gracias al desarrollo de una interfaz sencilla basada en programación java y html y con el uso de servidores virtuales Tomcat. Esta solución la ofrece la aplicación RWUI (<http://sysbio.mrc-bsu.cam.ac.uk/Rwui>), que al ser utilizada en el desarrollo de las herramientas CAPFITOGEN, se adapta con pequeñas modificaciones.

## 2.3 Instalación y ejecución

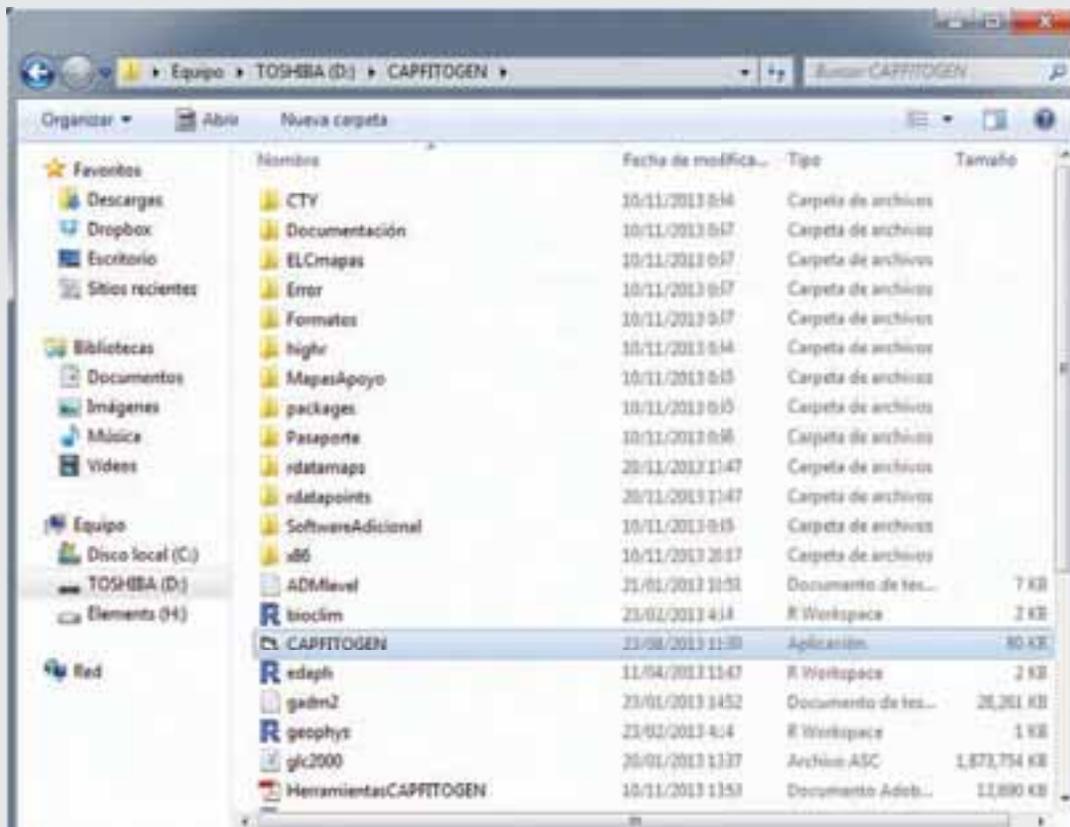
La versión 1.1 de las herramientas CAPFITOGEN requería un complejo proceso de instalación que incluía la instalación manual de R, el ambiente de ejecución de Java y del servidor virtual Tomcat, así como la modificación de variables de entorno de Windows. El antiguo procedimiento exigía al usuario hacer la instalación manual de los programas y los cambios de las variables de entorno, además de tener en cuenta el tipo de sistema operativo (32 o 64 bit) de la versión de Windows.

A partir de versión 1.2, el conjunto de herramientas CAPFITOGEN se entrega con un instalador que se encarga de realizar todos los pasos necesarios de instalación de programas y modificación de variables de entorno. Adicionalmente este instalador se encarga también de descomprimir e instalar todos los paquetes de R necesarios para que se lleven a cabo los análisis.

De esta manera se ha conseguido un paquete de herramientas listo para usar, los pasos para hacer la instalación y ejecución de las herramientas CAPFITOGEN son los siguientes:

- a) Si ha descargado las herramientas desde el sitio web dispuesto para ello (detalles en <http://www.agrobiodiversidad.org/blog/?p=1039>), tendrá un archivo con extensión .alz (comprimido) con el cuerpo principal de las herramientas y otros archivos .alz con la información ecogeográfica para cada país o región que se vaya a considerar como área de trabajo. Extraiga todos los archivos o carpetas que contienen esos archivos comprimidos, y luego arme el instalador de las herramientas poniendo la carpeta con el nombre del país o región en la carpeta "rdatamaps" del conjunto de carpetas y archivos del cuerpo principal de las herramientas, el cual se encuentra englobado en la carpeta "CAPFITOGEN".

**FIGURA 1.** Contenido típico de la memoria USB o archivo comprimido con las herramientas CAPFITOGEN, donde se resalta el archivo CAPFITOGEN.exe



Copie ésta carpeta preferiblemente en una unidad de memoria externa dedicada exclusivamente para las herramientas o directamente en el disco duro del PC, que frecuentemente corresponde a la unidad C:\. Los sistemas de descarga desde internet y armado de las herramientas pueden cambiar en versiones posteriores.

- b) Si ha accedido a las herramientas a través de una memoria USB suministrada por el Programa CAPFITOGEN, simplemente conéctela a uno de los puertos USB de su PC. Al ver el contenido de la memoria USB encontrará una carpeta llamada "CAPFITOGEN". No mueva el contenido de esta carpeta. Esta carpeta contiene la estructura de carpetas y archivos (Fig. 1) necesarias para el funcionamiento de las herramientas.
- c) En el grupo de archivos encontrará un solo archivo ejecutable (con extensión .exe), usualmente denominado "CAPFITOGEN.exe". Haga doble clic sobre este archivo e inmediatamente se desplegará una ventana como la que se muestra en la Fig. 2.
- d) Haga clic en el botón de "Instalar". El instalador le irá mostrando el avance en la instalación de los diferentes programas en un texto de color azul. Cuando termine, le aparecerá una ventana de color gris indicándole la finalización del proceso (Fig. 3). Haga clic en el botón "Aceptar" de esta ventana. No es necesario reiniciar la computadora.
- e) Inmediatamente después de terminar la instalación, en la ventana inicial, en donde se encontraba el botón "Instalar" aparecerá el botón "Ejecutar". Haga clic sobre este.
- f) En ese momento se abrirá una ventana de fondo negro en la que se cargará una serie de código en color blanco (ver Fig. 4). Este código corresponde a la programación para activar el servidor virtual tomcat. La carga de instrucciones finaliza cuando aparece la siguiente línea de texto indicando los milisegundos que ha tomado activar el servidor virtual:

INFO: Server startup in xxxx ms

**FIGURA 2.** Ventana de instalación de las herramientas CAPFITOGEN.



**FIGURA 3.** Ventana que indica que la instalación ha finalizado.



- g) Después de varios segundos de iniciar el proceso anterior, se abrirá automáticamente el programa de navegación de internet predeterminado (por ejemplo: Internet Explorer, Mozilla Firefox o Google Chrome) como muestra la Fig. 5. En la barra de direcciones del navegador aparecerá la siguiente instrucción:

<http://localhost:8080/start/>

**FIGURA 4.** Ventana de fondo negro donde se carga el servidor virtual. Esta ventana no se debe cerrar sino hasta la finalización de uso de las herramientas.

```

ene 17, 2014 9:02:10 AM org.apache.catalina.startup.HostConfig.deployDirectory
INFO: Despliegue del directorio D:\CAPFITOGEN\...
ene 17, 2014 9:02:10 AM org.apache.catalina.startup.HostConfig.deployDirectory
INFO: Despliegue del directorio D:\CAPFITOGEN\...
ene 17, 2014 9:02:10 AM org.apache.coyote.AbstractProtocol start
INFO: Starting ProtocolHandler ["http-bio-8080"]
ene 17, 2014 9:02:10 AM org.apache.coyote.AbstractProtocol start
INFO: Starting ProtocolHandler ["ajp-bio-8009"]
ene 17, 2014 9:02:10 AM org.apache.catalina.startup.Catalina start
INFO: Server startup in 2377 ms

```

En algunas ocasiones el proceso del punto "f", más arriba indicado, inicia antes de que el descrito en el punto "e" haya finalizado. Esto ocasiona que el servidor virtual no esté listo cuando es llamado desde el navegador, lo cual genera un aviso de error en el navegador (ver Fig. 6). El adelanto del proceso "f" sobre el "e" se da comúnmente la primera vez que se ejecutan las herramientas o cuando la configuración del equipo no está muy actualizada. Este problema se soluciona simplemente cerrando y abriendo el navegador y escribiendo en la barra de direcciones:

`http://localhost:8080/start/`

**FIGURA 5.** El navegador se abre automáticamente mostrando el panel de inicial.



**FIGURA 6.** Ejemplo de error que se produce cuando el navegador intenta abrir la dirección del servidor virtual sin que este se haya cargado por completo.



- h) Aparecerá en el navegador un panel de inicio (Fig. 5), que muestra en la parte izquierda la lista de herramientas disponibles y en la derecha los lenguajes disponibles para cada herramienta. Al hacer clic sobre los íconos de color verde se abrirá en el navegador el formulario de la herramienta y lenguaje seleccionado. En la Fig. 7 se muestra como ejemplo el formulario de la herramienta GEOQUAL. En este formulario aparece una pequeña barra de vínculos que se explica en la Fig. 8.

Para salir de la aplicación una vez haya terminado de utilizar las herramientas, basta con cerrar el navegador, cerrar la ventana de fondo negro y cerrar la ventana de la Fig. 2 haciendo clic en el botón "salir". Es posible que después de cerrar todas estas ventanas, Windows le pregunte si la aplicación fue correctamente instalada, por favor responda afirmativamente.

**FIGURA 7. Formulario de la herramienta GEOQUAL. 1. Vínculos sobre aspectos legales. 2. Espacio que indica la herramienta en uso. 3. Barra de vínculos. 4. Descripción breve de la herramienta. 5. Vínculo al manual de instrucciones. 6. Zona de introducción de parámetros. 7. Botón para iniciar el análisis.**

The screenshot shows the GEOQUAL web application interface. At the top, there is a header with the text "Tratado internacional" and "GEOQUAL" in a blue bar. Below the header, there is a description of the tool. The interface includes several input fields for parameters: "BUCARFITOGEN", "Superficie", "alt", "Diferencia", and "Diferencia". A "Calcular" button is located at the bottom. A small navigation bar with icons is visible above the input fields. The page is annotated with numbers 1 through 7 pointing to specific elements: 1 points to the top navigation bar, 2 points to the "GEOQUAL" header, 3 points to the navigation bar with icons, 4 points to the introductory text, 5 points to a link for the manual, 6 points to the input fields, and 7 points to the "Calcular" button.

En el momento en que se realiza la instalación, se añade un archivo de identificación en la unidad y ruta donde se encuentran los archivos y carpetas de las herramientas CAPFITOGEN. Ese archivo hace que el ordenador reconozca que la instalación ya ha sido realizada. De esta manera, cada vez que necesite usar de nuevo las herramientas, pulse sobre el archivo CAPFITOGEN.exe y se abrirá la ventana de instalación, pero esta vez directamente se mostrará el botón de "Ejecutar". A partir de este paso se repite todo el proceso desde el punto "d" indicado más arriba. Si por algún motivo la letra de la unidad o la ruta donde se encuentra el archivo ejecutable cambia, o se elimina el archivo de identificación accidentalmente, al hacer doble clic sobre CAPFITOGEN.exe aparecerá de nuevo el botón "Instalar". Se aconseja entonces proceder para reinstalar. Si se reinstalan nuevamente los programas R y Java no se producirán problemas en la ejecución de las herramientas.

**FIGURA 8.** Barra de vínculos (corresponde a la parte 3 de la Fig.6). Los botones cumplen las siguientes funciones: 1. Regresar al formulario de la herramienta seleccionada. 2. Ver manual de instrucciones de las herramientas CAPFITOGEN. 3. Información de contacto y apoyo. 4. Regreso a la página de selección de las herramientas (página de aterrizaje).



- i) Mientras el análisis se está ejecutando, Windows mostrará el puntero del ratón con el ícono de espera. Una vez el análisis ha concluido satisfactoriamente, la página que contiene el formulario volverá a situarse en la parte superior, mostrando el encabezado con el logo del TIRFAA. Al ir a la parte inferior de la página aparecerá un mensaje de "Proceso finalizado" (Fig.9), indicando que se deben buscar los resultados ya guardados. Si desea cambiar algún parámetro y volver a ejecutar el análisis, puede hacerlo desde el mismo formulario, cambiando el parámetro en cuestión y pulsando sobre el botón de "Analizar". Puede también borrar todo el contenido del formulario con el botón "Clear Page".

**FIGURA 9.** Vista de la parte inferior de la página del formulario una vez se ha concluido el análisis correctamente y los resultados están disponibles.



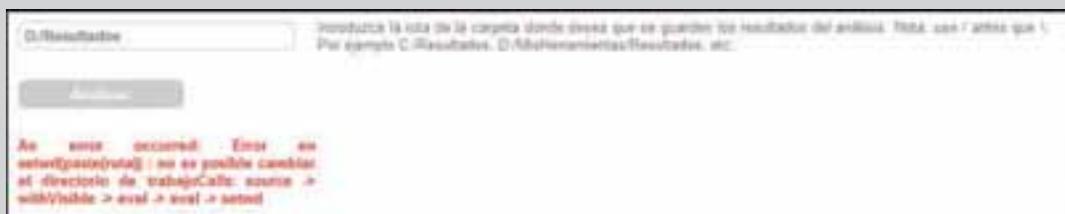
Los resultados de los análisis deben buscarse en la carpeta definida por el usuario en el parámetro "resultados". Los mapas, tanto en formato ráster (mallas de celdas con extensiones .grd y .gri), por ejemplo, los mapas producidos por la herramienta ELC mapas, como en formato vectorial (usualmente "shapefiles"), por ejemplo, los mapas de puntos, pueden abrirse con el programa DIVA GIS ya que son plenamente compatibles. Las tablas resultado se producen en formato de texto separado por tabulaciones, las cuales pueden abrirse desde programas como Microsoft Excel o las hojas de cálculo de Open Office.

- j) Es posible que cuando se sale de una herramienta hacia el panel de inicio (usando el botón  de la Fig. 8) y luego se regresa a la misma herramienta o simplemente se recarga de nuevo la dirección en el navegador, aparezca una ventana como la que muestra la Fig. 10 con instrucciones en inglés. En esta ventana se pregunta al usuario si desea volver a cargar una herramienta que previamente ha sido cargada. No ocurrirá ningún problema si se recarga la herramienta desde esta ventana, por lo que para continuar se hace clic sobre el botón "Run new copy of" (ejecutar nueva copia de) y el nombre de la herramienta.
- k) Cuando se produce un error en la ejecución de la herramienta, ya sea por una deficiencia en la instalación, por un parámetro introducido erróneamente o por defectos en las tablas que aportan información de pasaporte o caracterización, la página que contiene el formulario muestra la cabecera (de la misma manera que cuando el proceso finaliza exitosamente), pero el tiempo que tarda es considerablemente inferior a lo que tarda el proceso exitoso y en la parte inferior de la página, aparecerá un mensaje de error, como lo muestra la Fig. 10. Resulta posible detectar el origen del error a través del mensaje de error. Así, cuando aparezca este mensaje, generado directamente por el programa R, se debe comparar su contenido con la lista de mensajes del capítulo 10 "Errores frecuentes".

**FIGURA 10.** Ventana que indica que se está intentando recargar una herramienta previamente abierta. Para poder continuar, simplemente hacer clic sobre el botón "Run new copy of" y el nombre de la herramienta.



**FIGURA 11.** Vista de la parte inferior de la página del formulario cuando se ha producido algún tipo de error.



- l) Por las configuraciones de algunos computadores personales, es posible que durante la instalación no se carguen bien todos los paquetes que R requiere para llevar a cabo los análisis previstos. Esto producirá un error recurrente al ejecutar las herramientas, el cual aparecerá en color rojo como el de la Fig. 11 pero con el siguiente texto:

An error occurred: Error en library(nombre del paquete) : there is no package called 'nombre del paquete'Calls: source -> withVisible -> eval -> eval -> library

Este error, que aparece detallado en el capítulo 10 (Errores frecuentes), puede repetirse para todos los paquetes necesarios, haciendo el proceso de corrección propuesto en el capítulo 10 demasiado extenso. En este caso, se sugiere que el usuario elimine la carpeta "library" que encontrará en la ruta C:\rwin, luego copie la carpeta "library" que encontrará en la ruta CAPFITOGEN\Error y finalmente la pegue en la ruta C:\rwin. Esta acción eliminará por completo los problemas de ausencias de paquetes necesarios para R.

m) En algunas ocasiones, debido a un proceso deficiente de instalación o a una incorrecta ubicación de la carpeta "CAPFITOGEN", es posible que la interfaz no pueda activar R y enviarle la información necesaria para realizar los análisis. Este problema se hará evidente cuando después de hacer clic en el botón "analizar" y haber rellenado correctamente los parámetros, el tiempo de ejecución sea anormalmente corto y no aparezcan mensajes de error en color rojo sino el mensaje de "Proceso finalizado" y en la carpeta donde se deberían guardar los resultados no aparezca ningún archivo nuevo. En este caso asegúrese de dos aspectos:

1. Que los archivos y carpetas que componen las herramientas CAPFITOGEN están alojadas en una carpeta y no se encuentran directamente en la raíz del directorio. En otras palabras, la ruta para llegar a capfitogen.exe debería ser X:\CAPFITOGEN\capfitogen.exe y no X:\capfitogen.exe siendo X la letra de la unidad de disco. Si hiciera falta reinstale las herramientas asegurándose cumplir la condición anteriormente detallada.

2. Que en caso de que la carpeta CAPFITOGEN no se encuentre en la raíz del directorio (X:\CAPFITOGEN) sino en una ruta dentro de otras carpetas, algo de por sí no recomendable, esta ruta no debe contener espacios entre palabras. Por ejemplo, ubicar la carpeta en la ruta X:\Mis Documentos\CAPFITOGEN generará este error por el espacio que hay entre las palabras "Mis" y "Documentos".

Si a pesar de asegurarse que las dos condiciones anteriores se cumplen el problema se repite, póngase en contacto con la asistencia técnica del Programa.

n) Cuando el proceso que se ejecuta después de hacer clic en el botón "analizar" toma mucho tiempo (más de 15 o 20 minutos) es posible que la página que muestra el formulario de la herramienta cambie su apariencia y muestre un aviso de error de java, como el que aparece en la Fig. 12. Este cambio y este aviso de error no indican que el proceso haya finalizado abruptamente ni que no se produzcan resultados. Se trata de un error más de visualización de la interfaz que del proceso en R. Por ello, el paso siguiente es revisar si aparecen los archivos esperados como resultados en la carpeta designada para guardarlos.

**FIGURA 12.** Aviso de error de java que aparece cuando los procesos tardan demasiado tiempo.



```
Estado HTTP 500 - java.lang.IllegalStateException: setAttribute: La Sesión ya ha sido invalidada

Informe de Excepción
java.lang.IllegalStateException: setAttribute: La Sesión ya ha sido invalidada
El servidor encontró un error interno que fue tan grave que pudo o pudo ser inseguro.
StackTrace

java.servlet.ServletException: java.lang.IllegalStateException: setAttribute: La Sesión ya ha sido invalidada
org.apache.struts.action.RequestProcessor.processException(RequestProcessor.java:433)
org.apache.struts.action.RequestProcessor.processActionPerform(RequestProcessor.java:432)
org.apache.struts.action.RequestProcessor.process(RequestProcessor.java:236)
org.apache.struts.action.ActionServlet.doPost(ActionServlet.java:432)
javax.servlet.http.HttpServlet.service(HttpServlet.java:647)
javax.servlet.http.HttpServlet.service(HttpServlet.java:723)

StackTrace

java.lang.IllegalStateException: setAttribute: La Sesión ya ha sido invalidada
org.apache.catalina.session.StandardSession.setAttribute(StandardSession.java:1437)
org.apache.catalina.session.StandardSession.setAttribute(StandardSession.java:1402)
org.apache.catalina.session.StandardSessionFacade.setAttribute(StandardSessionFacade.java:156)
org.apache.struts.action.ActionServlet.doPost(ActionServlet.java:432)
org.apache.struts.action.RequestProcessor.processException(RequestProcessor.java:433)
org.apache.struts.action.RequestProcessor.process(RequestProcessor.java:236)
org.apache.struts.action.ActionServlet.doPost(ActionServlet.java:432)
javax.servlet.http.HttpServlet.service(HttpServlet.java:647)
javax.servlet.http.HttpServlet.service(HttpServlet.java:723)

La trama completa de la causa de este error se encuentra en los archivos de diario de Apache Tomcat/7.0.35.

Apache Tomcat/7.0.35
```



## 3. Herramienta GEOQUAL

### 3.1. ¿Qué es la Evaluación de la Calidad de la Georreferenciación en datos de Pasaporte?

Es una metodología que determina el grado de certidumbre contenido en algunos descriptores de pasaporte cuya función es definir inequívocamente el lugar donde el germoplasma fue recolectado. De esta manera, GEOQUAL hace una evaluación de la calidad de los datos de descripción de la localidad y de las coordenadas indicadas como sitio de recolección.

En términos amplios, el concepto de calidad aplicado a datos ha recibido diversas definiciones. En el ámbito geográfico, la definición de calidad como “aptitud de uso” o potencialidad de uso ha sido ampliamente aceptada (Chrisman, 1983). Esto relaciona directamente calidad y posibilidad de uso de los datos. La incertidumbre asociada a todo tipo de datos, es una propiedad de quien obtiene o usa los datos más que de los datos mismos. Por ello calidad e incertidumbre tiene un grado de subjetividad variable, la cual puede reducirse hasta cierto punto mediante el uso de metodologías que realicen evaluaciones sobre bases lo más objetivas posibles. En todo caso, calidad e incertidumbre se asumen como medidas de riesgo entendido y riesgo asumido (Chapman, 2005).

La necesidad de evaluar la calidad de la georreferenciación de la información disponible sobre la presencias o incluso ausencias de entidades biológicas, es un asunto palpable en diferentes áreas como la ecología, análisis espacial y patrones de distribución de especies. Son numerosos los estudios que apuntan a que dicha calidad es un asunto crítico en metodologías como la modelación de la distribución de especies. La certeza de la ocurrencia de una especie en un sitio determinado es determinante para cualquier método que use datos de presencia o ausencia como materia prima. Los siguientes estudios hacen referencia a este aspecto: Foley y colaboradores, 2009; Hill y colaboradores, 2009; Otegui y colaboradores, 2013.

Tener una estimación del grado de incertidumbre en la georreferenciación de sitios de presencia o ausencia de especies se convierte entonces en un aspecto clave antes de realizar cualquier análisis que utilice el aspecto espacial para el estudio de las distribuciones. Muchos análisis de este tipo desembocan en la toma de decisiones en aspectos prácticos de áreas como la conservación de la biodiversidad. Por tanto introducir información de base confiable alimentando análisis adecuados producirá resultados confiables y decisiones acertadas y oportunas.

### 3.2. Antecedentes de la herramienta GEOQUAL

La metodología detrás de GEOQUAL acumula cuatro años de desarrollo, desde que surgió la necesidad de un estimador de confianza (o de riesgo, según se vea) de la georreferenciación del sitio de recolección, que usualmente quedan reflejados en los datos de pasaporte. La necesidad en particular surgió a finales de 2009, al momento de preparar los datos de pasaporte del Inventario Nacional de Recursos Fitogenéticos Españoles para ser caracterizados ecogeográficamente. Tener una idea de la calidad de la georreferenciación en datos de pasaporte fue entonces una prioridad a la hora de crear el Sistema de Información Ecogeográfica de los Recursos Fitogenéticos Españoles (proyecto SIERFE, <http://www.sierfe.es>).

SIERFE es un sistema que permite seleccionar germoplasma a partir de la caracterización ambiental del sitio de recolección a través de un portal de internet. Al desarrollar GEOQUAL e incorporarlo en SIERFE, un estimador de la calidad permite a los usuarios de SIERFE (solicitantes de germoplasma, como mejoradores, científicos o agricultores) determinar su exigencia en términos de calidad de la georreferenciación a la hora de seleccionar germoplasma por variables ecogeográficas. Esto representa un avance único en el desarrollo de sistemas de información y selección de germoplasma. Más de 45.000 entradas del Inventario Español fueron caracterizadas ecogeográficamente y cada una recibió un valor de calidad en una escala de 0 a 100.

GEOQUAL fue entonces desarrollado específicamente para las características de los datos de pasaporte del Inventario Nacional de Recursos Fitogenéticos Españoles y fue posible utilizando diversos programas, la mayor parte de tipo comercial como ArcGIS de ESRI.

En 2011, en el marco del proyecto *PGR secure* del Séptimo Programa Marco de la Unión Europea (<http://www.pgrsecure.org>), fue necesaria la depuración de cuatro bases de datos que contenían información de ocurrencia de variedades y especies silvestres relacionadas con cuatro géneros de interés agrícola en Europa (*Avena*, *Beta*, *Brassica* y *Medicago*). Más de 33.000 registros recibieron un valor GEOQUAL, lo cual permitió no considerar o mejorar la calidad de alrededor de 4.000 registros.

A partir de entonces varios los investigadores europeos en temas de agrobiodiversidad se interesaron por GEOQUAL, demandando el desarrollo de una herramienta de fácil manejo que permitiera aplicar GEOQUAL sobre diferentes formatos de datos de presencia de especies.

Ya en 2012, cuando el programa CAPFITOGEN es aprobado y se deciden las herramientas a desarrollar, GEOQUAL se convirtió en una prioridad. Se trataba de abordar el reto de crear una herramienta de evaluación de la calidad de datos de georreferenciación que fuera sencilla, con toda la información necesaria precargada, que no requiriera mayores conocimientos en Sistemas de Información

Geográfica (SIG), para poder aplicarla, que viniera en una solución integrada (usando sólo un programa de SIG), que usara como base el formato de descriptores de pasaporte definidos por la FAO y Bioversity International en 2012 y que fuera transferible a técnicos de programas nacionales. La herramienta GEOQUAL que se presenta aquí, es por tanto, la evolución de una idea original transformada en tecnología de fácil adopción, reuniendo factores de adaptabilidad apropiados a las condiciones y necesidades de varios programas nacionales de conservación de recursos fitogenéticos.

### 3.3. Características de GEOQUAL

La herramienta GEOQUAL se compone de cuatro parámetros, tres de ellos aportan aproximaciones diferentes sobre la calidad de la georreferenciación (COORQUAL, SUITQUAL y LOCALQUAL) y un parámetro final (TOTALQUAL) que resume los tres primeros. Los parámetros base se calculan en rangos que van de cero a veinte, siendo cero calidad nula y 20 máxima calidad. En algunas ocasiones, dependiendo de los datos de pasaporte disponibles, se puede obviar el cálculo de LOCALQUAL, como se explica más adelante. Adicionalmente se genera un parámetro que transforma los valores iniciales de TOTALQUAL (0 a 40 ó 0 a 60) a un rango de evaluación de 0 a 100, para facilidad de uso e interpretación del valor de evaluación (TOTALQUAL100).

Es importante tener en cuenta que GEOQUAL opera sobre el formato de descriptores de pasaporte FAO-Bioversity publicado en el año 2012 con la adición de cuatro descriptores de localidad (ADM1, ADM2, ADM3 y ADM4) que tienen correspondencia con diferentes figuras administrativas según cada país (ver Anexo 9.1). Sin embargo, si los datos estuvieran en el formato FAO-IPGRI 2001, GEOQUAL podría también operar previo traslado de la información del formato 2001 al 2012 sin tener que añadir información para los nuevos campos que incluye la versión 2012, pero sí considerando la inclusión de los cuatro descriptores ADM.

La herramienta GEOQUAL incluye un modelo de tabla de descriptores de pasaporte basado en los descriptores multicultivo de FAO-Bioversity con la adición de los cuatro descriptores ADM en formato Excel (carpeta "DescriptoresPasaporteModelo", archivo "Tabla pasaporte modelo FAO\_Bioversity 2012 modificada.xls"), donde en color verde se señalan los descriptores indispensables para GEOQUAL y en color amarillo aquellos no indispensables pero si importantes. Los campos no señalados no son tenidos en cuenta por GEOQUAL pero si se debe mantener su posición dentro de la tabla (como en el caso de los señalados) para que GEOQUAL encuentre las variables que analiza en el sitio que espera encontrarlas. Como regla general, al llenar esta tabla, cuando no conozca la información que se le requiere, se debe escribir en el campo las letras NA, que normalmente indican "no aplica" pero que en el caso de GEOQUAL también indica que no hay información disponible.

### 3.3.1 Descripción de los parámetros base de GEOQUAL

#### 3.3.1.1 Parámetro COORQUAL

Es un parámetro que determina la calidad intrínseca de las coordenadas contenidas en los datos de pasaporte. Se determina a través de cuatro sub-parámetros inicialmente:

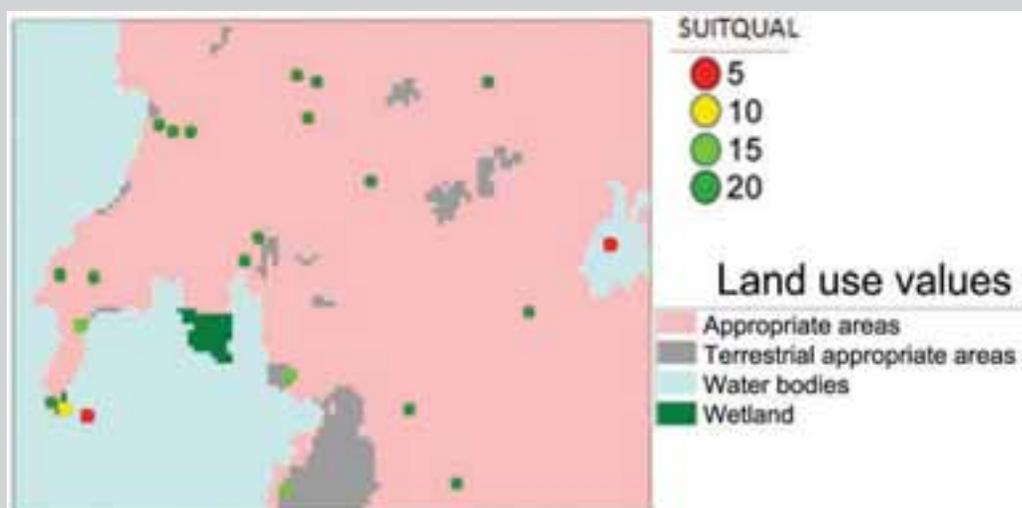
- a) **ERRORES:** Si las coordenadas en formato sexagesimal o decimal contienen valores fuera de marco posible dentro del sistema de coordenadas latlong WGS84. Usa los descriptores LATITUDE, LONGITUDE, DECLATITUDE y DECLONGITUDE.
- b) **PRECIS:** Aplica para coordenadas en formato sexagesimal que cumplen la codificación del listado de descriptores de pasaporte FAO-Bioversity 2012. Este subparámetro determina si las coordenadas fueron obtenidas con una precisión de segundos, minutos o grados. Usa los descriptores
- c) **GEORBLE:** Es una valoración de la posibilidad de obtener coordenadas del sitio de recolección a partir de los datos de descripción de localidad disponibles.
- d) **INTERTEMP:** Toma valores del descriptor COLLDATE y los interpreta en la medida de la posibilidad de uso de métodos de georreferenciación. Por ejemplo, para recolecciones ocurridas posteriormente al año 2000, se asume una alta probabilidad de uso de GPS, lo cual incrementaría la calidad de las coordenadas.
- e) **GEOREFMETH:** Valora el sistema por el cual se han asignado coordenadas al sitio de recolección. GEOREFMETH corresponde a un campo de la tabla de pasaporte FAO/Bioversity 2012. Solamente se considerará este subparámetro cuando se dispone de valores para todas las entradas de este campo.

Cada sub-parámetro se evalúa en un rango de cero a tres, donde cero corresponde a calidad mínima y tres a calidad máxima. Las combinaciones de los valores de cada sub-parámetro generan el parámetro COORQUAL en un rango de cero a veinte.

#### 3.3.1.2 Parámetro SUITQUAL

Este parámetro asigna un valor de calidad a las coordenadas de acuerdo a lo apropiado del sitio de recolección para el crecimiento de plantas, y diferenciando de la naturaleza de la entrada (silvestre o cultivada de acuerdo al descriptor SAMPSTAT). La información sobre las características del sitio de la recolección proviene de un mapa de uso de suelo (Global Land Cover 2000 o GLC2000). Se trata del mapa de cobertura global de libre acceso sobre uso de suelo más antiguo y que cuenta con una resolución apropiada (1 km). Las clases originales de este mapa se transforman en función de lo apropiado que resulta cada clase a la presencia de plantas cultivadas o silvestres, en una escala de 0 a 20.

**Figura 13.** Ejemplo de la obtención de valores *SUITQUAL* de acuerdo a valores de uso de suelo interpretados.

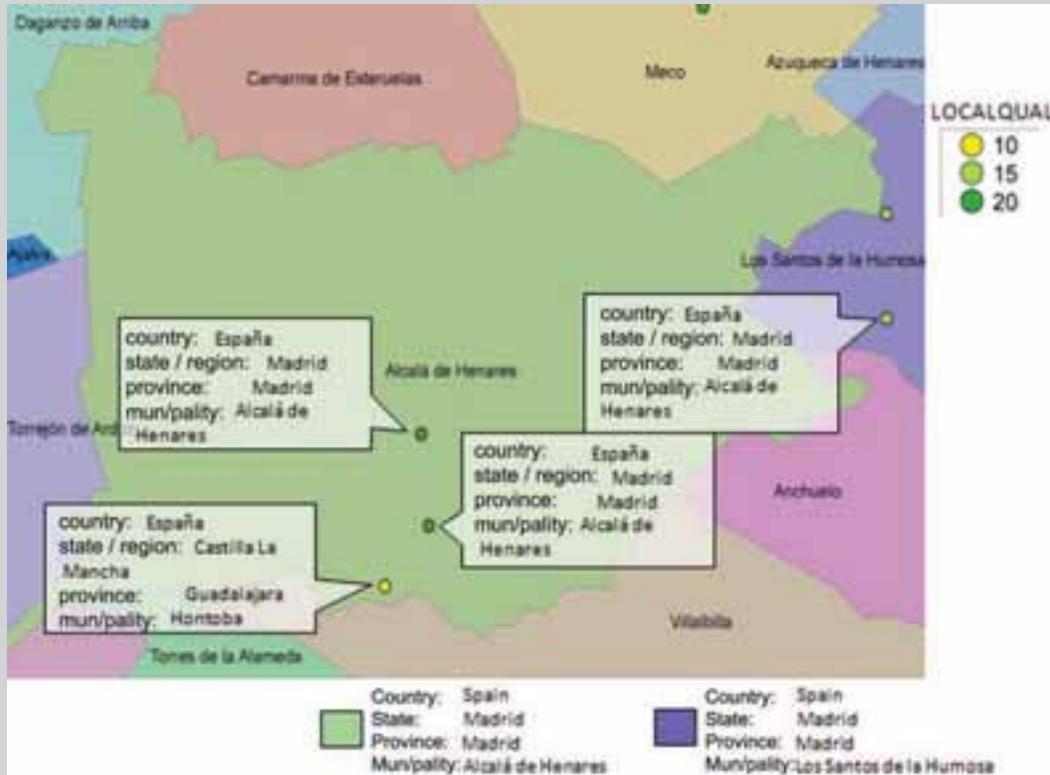


### 3.3.1.3 Parámetro LOCALQUAL

LOCALQUAL es el resultado de la comparación de la descripción de la localidad donde se recolectó el germoplasma proveniente de los campos ORIGCTY, ADM1, ADM2, ADM3, ADM4 y COLLSITE, con los campos ISO, NAME1, NAME2, NAME3 y NAME4 de la base de datos "Global Administrative Areas" (GADM) v2.0, extraídos por medio de las coordenadas provistas en DECLATITUDE y DECLONGITUDE (o a través de la transformación a formato decimal de LATITUDE y LONGITUDE). A diferencia del proceso que realiza "Check Coordinates" (comprobar coordenadas) incluido en DIVA-GIS donde la comparación es absoluta (los términos deben coincidir carácter por carácter para adjudicar coincidencia), GEOQUAL utiliza la distancia generalizada de Levenshtein a través de la función "agrep" del paquete base de R, que toma en cuenta el número de inserciones, eliminaciones o cambios de caracteres entre las dos cadenas que se comparan. Así, permitiendo un cierto número de este tipo de cambios, la función "agrep" es capaz de identificar concordancias aunque ocurran errores de tipografía o por el efecto de aquellos caracteres alfabéticos propios de ciertos idiomas que en ocasiones no se codifican correctamente (caso de la ñ y los acentos en el idioma castellano).

Para más seguridad, LOCALQUAL compara también con los campos VARNAME1, VARNAME2, VARNAME3 y VARNAME4 incluidos en GADM, que son variantes del nombre oficial de la unidad administrativa y que pueden ser usados por los curadores a la hora de registrar el germoplasma en sus bases de datos de pasaporte.

**FIGURA 14.** Ejemplo de la obtención de valores LOCALQUAL de acuerdo a la comparación de niveles administrativos de los datos aportados por el usuario y los extraídos por las coordenadas desde GADM.

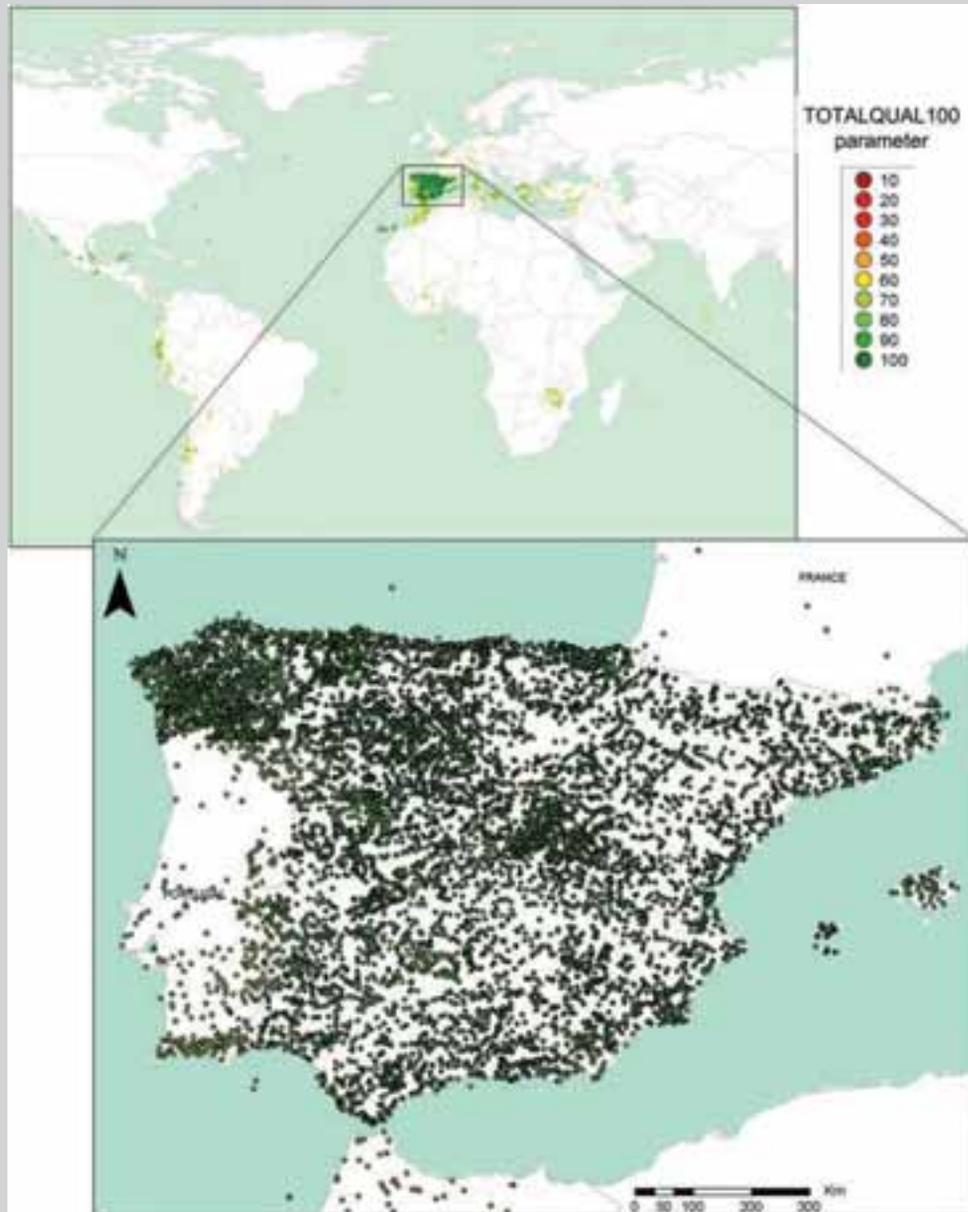


Finalmente, LOCALQUAL considera la serie de comparaciones positivas entre los diferentes emparejamientos (ORIGCTY con ISO, ADM1 con NAME1, etc.) para calcular un valor en un rango de cero a veinte.

### 3.3.2 Descripción de los parámetros TOTALQUAL y TOTALQUAL100

El parámetro final de resumen TOTALQUAL es simplemente la suma de los valores de COORDQUAL, SUITQUAL y LOCALQUAL. Por los rangos posibles de valores de estos tres parámetros, TOTALQUAL puede tomar valores entre 0 y 60. Sin embargo, para facilidad de interpretación y análisis de los resultados de GEOQUAL, se calcula también el parámetro TOTALQUAL100 que es una transformación de TOTALQUAL a un rango de valores de 0 a 100, donde 0 es calidad nula (incluyendo la falta de coordenadas) y 100 representa una calidad teórica óptima.

**FIGURA 15.** Resultados de la aplicación de GEOQUAL en el Inventario Nacional Español de Recursos Fitogenéticos. Se muestran los valores alcanzados de TOTALQUAL100.



### 3.3.3 Determinación de umbrales de calidad

GEOQUAL ha sido pensada desde su implementación inicial, como una metodología altamente objetiva, en la que el usuario interviene mínimamente en la obtención del valor definitivo. Sin embargo, toda determinación de calidad conlleva componentes subjetivos y GEOQUAL no es la excepción.



30

Por ejemplo, hay cierta subjetividad cuando se aplica a ciertas categorías de uso de suelo unos valores de aptitud para el crecimiento de plantas. También la definición de a partir de que valores se puede considerar alta o baja calidad es una cuestión subjetiva que tiene que ver con el observador más que con la técnica.

El umbral sobre el cual se considera que una entrada está correctamente georreferenciada usando valores GEOQUAL debe ser definida por el usuario de los datos en base a sus expectativas y necesidades. Se puede plantear el uso de diferentes umbrales dependiendo del uso que se le dará a los datos, del tipo de estudio al que serán sometidos, y de la precisión y exactitud de la información del resto de las fuentes en un determinado estudio. Es recomendable ver la distribución de valores TOTALQUAL100 del conjunto de entradas y así conocer de antemano que de aplicar un umbral muy exigente (cercano a 100) o poco exigente (por debajo de 50) se seleccionarían muchas o pocas entradas.

### 3.4. Utilización de la herramienta GEOQUAL

Una vez instalada las herramientas CAPFITOGEN y seleccionada la herramienta GEOQUAL, será necesario definir una serie de parámetros para que la programación R funcione correctamente.

Después de definir todos los parámetros y rutas que requiere GEOQUAL, al hacer clic en el botón "Analizar", se iniciará el proceso de análisis de la herramienta. Luego de un tiempo que puede variar por la introducción de algunos parámetros de resolución, tipo de análisis, tamaño de datos procesados o configuración de hardware del computador, GEOQUAL producirá resultados que guardará donde se le haya indicado (parámetro 3.4.1.6).

#### 3.4.1 Parámetros iniciales definidos por el usuario

##### 3.4.1.1 Parámetro: ruta

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN. Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

##### 3.4.1.2 Parámetro: pasaporte

Explicación: Escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama "tabla" debe escribir: "tabla.txt". Recuerde que este archivo debe ser guardado previamente en la carpeta "Pasaporte" que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN.

### 3.4.1.3 Parámetro: *precision*

Explicación: Elija utilizar mapas de alta o baja precisión para determinar si las coordenadas de un sitio de recolección caen en el mar y a cuanta distancia lo hace. Alta resolución puede ralentizar un poco el proceso en bases de datos muy grandes (por encima de 15.000 entradas con coordenadas).

### 3.4.1.4 Parámetro: *local*

Explicación: Indique si desea usar el parámetro LOCALQUAL para la evaluación de la calidad de la georreferenciación. LOCALQUAL es un parámetro de comparación entre localidad descrita y extraída por SIG. Si sus datos no contienen ninguna descripción de localidad o la descripción está contenida por completo en el campo COLLSITE, NO es conveniente usar esta opción.

### 3.4.1.5 Parámetro: *resultados*

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

## 3.5. Resultados de GEOQUAL

En la ruta y carpeta que haya determinado en “resultados” (parámetro 3.4.1.6) encontrará tres tablas y un mapa de puntos de tipo vectorial (shapefile).

### 3.5.1 Tablas

Las tablas producidas por GEOQUAL están en formato de texto delimitado por tabulaciones y pueden ser abiertas en programas como Excel, OpenOffice o R.

3.5.1.1 “PasaporteOriginalEvaluadoGEOQUAL.txt”: Es la tabla de pasaporte en el formato sugerido que originalmente se uso para el análisis, con la adición de cinco columnas con los valores obtenidos para los parámetros: SUITQUAL, LOCALQUAL, COORQUAL, TOTALQUAL y TOTALQUAL100.

3.5.1.2 “tabla\_de\_analisisGEOQUAL.txt”: Esta tabla contiene también todas las columnas de la tabla de pasaportes originalmente introducidas para su análisis, aunque en este caso sólo se incluyen las entradas con coordenadas. Sin embargo, lo más importante de esta tabla es que en ella se han añadido todas las columnas correspondientes a extracciones, interpretaciones o subparámetros que han sido necesarios para poder calcular los valores de los parámetros de GEOQUAL. La lista de variables adicionales incluidas en esta tabla y su explicación se encuentra en el Anexo 12.4.



### 3.5.2 Mapas

3.5.2.1 Mapa de puntos en formato vectorial tipo "shapefile". Este mapa viene acompañado de una tabla que incluye los valores de los parámetros de evaluación GEOQUAL, de tal manera que en DIVA-GIS se pueden mostrar los puntos con diferentes grados de color de acuerdo a su puntaje (calidad). Un "shapefile" se compone de hasta 6 archivos de igual nombre pero diferente extensión. En el caso de GEOQUAL, el shapefile sólo está compuesto por tres extensiones (.shp, .shx, y .dbf) y tiene el nombre de ShapefilePuntosGEOQUAL.

3.5.2.2 Mapa de puntos en formato Google Earth. Este mapa corresponde al archivo mapa\_puntos\_google.kml y si se tiene instalado el programa Google Earth, con sólo hacer doble clic sobre su nombre en el explorador de Windows, el mapa de puntos (en forma de tachuelas o chinchetas) se abrirá en dicho programa, ubicando los sitios de recolección sobre las imágenes de satélite. Al hacer clic sobre las chinchetas se abrirá una pequeña ventana mostrando el valor TOTALQUAL100 de cada entrada.

### 3.6. Referencias

Chapman, A.D. 2005. Principles of data quality, versión 1.0. Report of the Global Biodiversity Information Facility, Copenhagen.

Chrisman, N.R. 1983. The role of quality information in the long-term functioning of a GIS. Proceedings of AUTOCART06, 2: 303-321. Falls Church, VA: ASPRS.

FAO, IPGRI. 2001. Lista de descriptores de pasaporte para cultivos múltiples desarrollada por la FAO y el IPGRI .

FAO, BIOVERSITY. 2012. FAO/Bioversity multi-crop Passport descriptors V.2. Disponible en [http://www.bioversityinternational.org/index.php?id=19&user\\_bioversitypublications\\_pi1%5BshowUid%5D=6901](http://www.bioversityinternational.org/index.php?id=19&user_bioversitypublications_pi1%5BshowUid%5D=6901)

Foley, D.H., Wilkerson, R.C., Rueda, L.M. 2009. Importance of the “what,” “when,” and “where” of mosquito collection events. J Med Entomol. 2009 Jul;46(4):717-22.

Hill, A.W., Guralnick, R., Flemons, P., Beaman, R., Wieczorek, J., Ranipeta, A., Chavan, V., Remsen, D. 2009. Location, location, location: utilizing pipelines and services to more effectively georeference the world’s biodiversity data. BMC Bioinformatics. 2009 Nov 10;10 Suppl 14:S3. doi: 10.1186/1471-2105-10-S14-S3.

Otegui, J., Ariño, A.H., Encinas, M.A., Pando, F. 2013. Assessing the primary data hosted by the Spanish node of the Global Biodiversity Information Facility (GBIF). PLoS One. 2013;8(1):e55144. doi: 10.1371/journal.pone.0055144.

Soberón, J., Peterson, T. 2004. Biodiversity informatics: managing and applying primary biodiversity data. Phil. Trans. R. Soc. Lond. B. 359, 689-698.



## 4. Herramienta ELC mapas

### 4.1. ¿Qué es un mapa de caracterización ecogeográfica del terreno?

Se trata de un tipo de mapa donde podemos visualizar diferentes escenarios ambientales que pueden corresponderse con los diferentes procesos adaptativos de una especie vegetal a lo largo de un determinado territorio. Entre sus múltiples usos, los mapas de caracterización ecogeográfica del terreno (mapas ELC) son útiles para la conservación y uso razonable de la agrobiodiversidad.

La idea de expresar adaptación a través de mapas no es nueva. Desde mediados del siglo pasado vienen desarrollándose mapas de biomas, ecosistemas, regiones ecológicas, etc. Estos mapas representan unidades ambientales generalmente como regiones grandes y homogéneas. En principio los "climas" o "ambientes" (los términos se usaban indistintamente) que representan estos mapas eran usados en estudios de diversos tipos de organismos (plantas, animales, microorganismos). Algunos mapas afinaron más y representaban, por ejemplo, los climas afines a las formaciones vegetales descritas por Leslie Holdridge en 1947, aunque luego se generalizaron como "sistemas de clasificación de zonas de vida".

Estos mapas fueron de mucha utilidad para biólogos y naturalistas que intentaban comprender la distribución de organismos vivos en relación a la temperatura y la humedad. Aun hoy en día, el sistema de Holdridge tiene aplicación, por ejemplo, en estudios de cambio climático. Sin embargo, la mezcla a veces indistinguible de rasgos bióticos (vegetación) y abióticos (temperatura, precipitación) en estos mapas, la simplificación del componente abiótico en sólo dos factores y la forma de delinear las regiones (grandes, homogéneas y continuas) representaban un impedimento serio para su uso en estudios de adaptación en especies en particular.

La utilización de información sobre adaptación para diseñar una colecta, o tener criterios a la hora de conservar y utilizar de manera eficiente los recursos fitogenéticos tampoco es nada nuevo, aunque son escasos los trabajos publicados sobre la materia donde explícitamente se haga referencia a la adaptación. Existe, como referencia, un primer mapa ecogeográfico cuyo propósito era crear colecciones núcleo/nucleares (core collections) en 1997 (Tohme et al., 1995), donde al final del proceso de selección de entradas además de la dimensión ecogeográfica se consideraron otros criterios adicionales de diferente naturaleza.

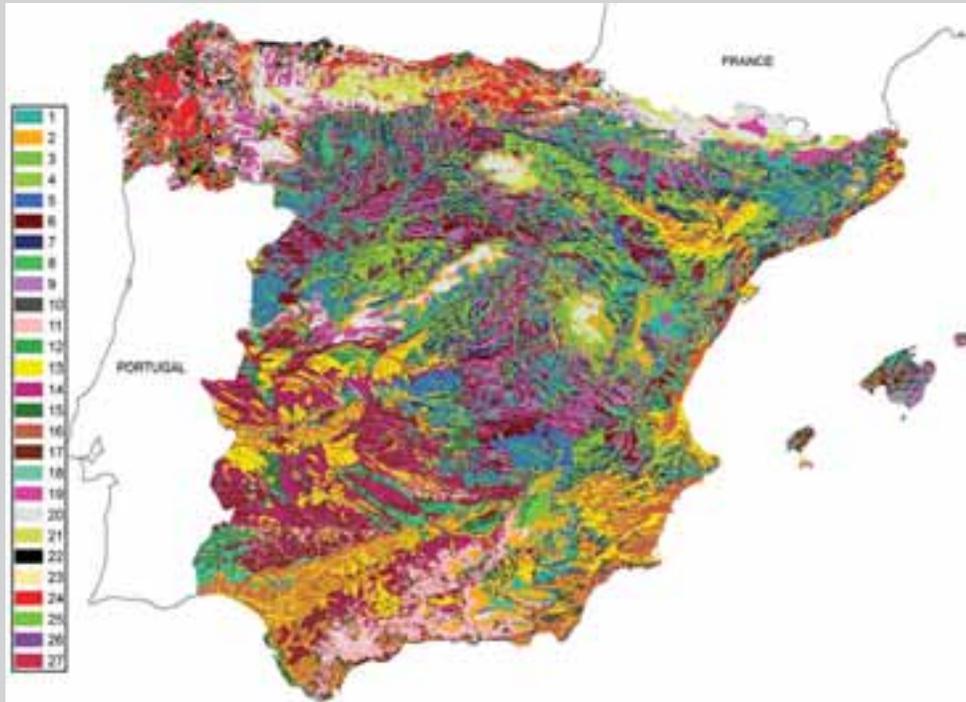
Desde entonces se han producido varios desarrollos: los programas SIG se han hecho gradualmente más flexibles y "amigables", e incluso algunos paquetes estadísticos actualmente incluyen utilidades y herramientas SIG; la información ecogeográfica disponible (en forma de capas SIG) es de mejor calidad y más accesible; los equipos informáticos de alta capacidad de análisis se venden a precios

reducidos y el acceso a internet se ha incrementado en países en desarrollo. Estos avances han repercutido en el desarrollo de mapas que representan diferentes escenarios de adaptación para plantas cultivadas y silvestres emparentadas. Es así como se obtuvo un primer mapa de caracterización ecogeográfica del territorio para España en 2005 (Parra Quijano et al., 2008) de tipo generalista (que pudiera ser aplicado a varias especies silvestres emparentadas con cultivadas) aunque sólo fue utilizado para algunas especies del género *Lupinus*. Este mapa, obtenido por técnicas de análisis multivariado y determinación del número de grupos por criterios bayesianos, representó las diferentes unidades ambientales de forma reticulada y como regiones homogéneas, pero de dimensiones reducidas y discontinuas. Estas características físicas ya contrastaban con los mapas bioclimáticos tradicionales. Otra diferencia fue la inclusión de variables de tipo geofísico y edáfico (además de las bioclimáticas), con la idea de representar los aspectos abióticos que pueden afectar el desarrollo de una planta desde un punto de vista agronómico.

A mediados de 2008 se desarrolló un nuevo mapa ecogeográfico para España Peninsular e Islas Baleares usando otras fuentes de información ecogeográfica. La metodología de creación de este nuevo mapa fue similar al anterior. En este punto lo que más interesaba a los investigadores era verificar que el mapa realmente reflejara escenarios adaptativos, es decir, realizar una validación. Para ello se evaluó el desempeño del nuevo mapa para ocho especies (cuatro leguminosas y cuatro gramíneas), dos de ellas silvestres emparentadas con cultivadas y las otras seis compuestas por variedades locales, usando su distribución y la variable "peso de la semilla" como variable fenotípica indicadora con valor adaptativo. Los resultados se compararon con dos mapas de referencia, uno con una estructura física similar al mapa ecogeográfico (alta reticulación, discontinuidad, unidades de tamaño reducido), pero construido sin tener en cuenta aspectos relacionados con la adaptación abiótica de plantas (mapa CORINE land cover, un mapa de uso de suelo, ver <http://www.eea.europa.eu/data-and-maps/explore-interactive-maps/corine-landcover-2006>) y otro con una estructura física diferente (más parecida a la de los mapas tradicionales) pero construido con un fin similar (mapa DMEER o mapa digital de regiones ecológicas europeas, ver <http://www.eea.europa.eu/data-and-maps/figures/dmeer-digital-map-of-european-ecological-regions>).

Los resultados del estudio de validación fueron variopintos. El mapa de caracterización ecogeográfica del territorio funcionó en general mejor para especies leguminosas que para gramíneas, aunque la excepción fue *Zea mays* para la cual el desempeño del mapa fue bastante aceptable. Como era de esperar, el mapa reflejó escenarios adaptativos para las dos especies silvestres, pero también produjo resultados satisfactorios en casos de especies compuestas por sólo variedades locales, como el caso de *Phaseolus vulgaris*. Como conclusión, los mapas de caracterización ecogeográfica del territorio son capaces de reflejar escenarios adaptativos y, por tanto, pueden ser utilizados en muchas actuaciones relacionadas con la colecta, conservación y utilización eficiente de los recursos fitogenéticos. Sin embargo, es recomendable crear mapas específicos para cada especie o grupo de especies relacionados filogenéticamente. Realizar mapas de caracterización ecogeográfica del terreno de tipo generalista puede resultar arriesgado a la hora

**FIGURA 16.** Mapa de caracterización ecogeográfica del territorio de España Peninsular y Baleares para *Lupinus*.



de sacar conclusiones para un grupo numeroso de especies, particularmente si no se realiza una validación del mapa. También es importante hacer una selección apropiada de las variables ecogeográficas involucradas en la creación del mapa, pero siempre representando los tres aspectos abióticos involucrados en el desarrollo de las plantas: bioclimático, geofísico y edáfico.

#### 4.2. Antecedentes de la herramienta ELC mapas

La herramienta ELC mapas que aborda este instructivo representa el desarrollo del concepto *Ecogeographical Land Characterization Maps* expuesto en la publicación de Parra Quijano y colaboradores (2012 A)

Este tipo de mapas han tenido usos diversos en colecta, conservación y uso de recursos fitogenéticos como en los trabajos de Parra Quijano y colaboradores (2011 A, 2011 B y 2012 B) o Thormann (2012).

El interés que esta metodología había despertado en varios equipos y proyectos de investigación en colecta, conservación y uso de recursos fitogenéticos contrastaba con una observación recurrente entre posibles usuarios: la metodología descrita en dicha publicación es un tanto compleja dado que mezcla técnicas de sistemas

de información geográfica (SIG) y análisis multivariado. Además, el desarrollo original implicaba el uso de un programa de análisis estadístico de pago. Estos aspectos limitaban de manera importante el desarrollo de mapas ecogeográficos de caracterización del terreno por parte de investigadores y técnicos.

### 4.3. Características de ELC mapas

La herramienta ELC mapas provee una nueva opción para desarrollar mapas de caracterización ecogeográfica del terreno mediante el uso de R, evitando las complicaciones anteriormente descritas. Este software, gratuito y con amplia capacidad de cálculo estadístico y potencia gráfica, permite integrar SIG y análisis multivariado. Así, la nueva herramienta ELC mapas es capaz de producir mapas de caracterización ecogeográfica del terreno sin necesidad de alternar entre diferentes programas, ni descargar y manipular información ecogeográfica. Sin embargo, es importante aclarar que los productos de la herramienta ELC mapas son mapas y tablas que pueden ser visualizadas en programas como DIVA-GIS, Google Earth o Microsoft Excel. Los mapas de caracterización ecogeográfica del terreno de la herramienta ELC mapas son así mismo un componente de otras herramientas como Representa.

ELC mapas usa dos procedimientos para determinar el número de grupos a usar dentro del método de agrupamiento. Estos procedimientos son:

- a) Un sistema simple que usa como algoritmo de agrupamiento K –means con determinación de punto de corte basado en la disminución de la suma de cuadrados intragrupo (Ketchen y Shook, 1996). El número de grupos óptimo es el alcanzado cuando la disminución de la sumas de cuadrados intragrupos entre una solución de  $n$  y una de  $n+1$  grupos es menor del 50%. Este método, también conocido como de “elbow” o codo, es el de más rápido cálculo y es capaz de actuar sobre grandes cantidades de datos sin grandes demoras, por lo cual se aconseja para países de gran tamaño.
- b) Método de agrupamiento de partición alrededor de los medoides (pam). Se utiliza el método de silhouette de interpretación y validación del número de grupos. Este sistema (en principio gráfico, posteriormente adaptado en R por el paquete fpc) permite determinar qué tan apropiadamente han sido los datos agrupados (Kaufman y Rousseeuw, 1987; Rousseeuw, 1987). Es un sistema que consume más recursos computacionales, por lo que aplicado a grandes conjuntos de datos, consumirá más tiempo.

Los métodos de determinación del número de grupos no son del todo objetivos, pues al menos el usuario debe determinar un máximo número de grupos que permitirá, y para el método del codo, el porcentaje de disminución es subjetivo, aunque basado en la observación de las gráficas de varianza intragrupos por número de grupos.

Dado que la información ecogeográfica a niveles de resolución de 1 km o incluso de 5 km es bastante voluminosa para un subcontinente completo como Latinoamérica,

la herramienta ELC mapas está inicialmente pensada para ser utilizada a nivel país, si bien la distribución de las especies objetivo o la distribución de las recolecciones de germoplasma pueda ir más allá de los límites de un país determinado. Se espera producir, en un corto plazo, una versión de ELC mapas con datos a nivel continental o subcontinental con menos resolución (10 km).

#### 4.4. La selección de las variables ecogeográficas

Este es un aspecto muy importante a tener en cuenta antes de usar la herramienta ELC mapas. El cambio, adición o eliminación de una sola variable de un sólo componente (bioclimático, geofísico o edáfico) cambiará sensiblemente la configuración final del mapa y su correlación con los escenarios adaptativos de la especie.

La técnica de mapas de caracterización ecogeográfica del terreno originalmente no contemplaba hacer mayor discriminación de variables dado que se pensaba en mapas de corte generalistas. Sin embargo, a medida que se reconoció que su capacidad de discriminar correctamente escenarios adaptativos se incrementaba al enfocarse en una especie en particular o un grupo de especies muy afines (en términos genéticos), surgió la idea de seleccionar las variables ecogeográficas de cada componente que más influyen en la adaptación abiótica de la especie o especies, y que por tanto determinan su distribución.

El proceso de selección es entonces un aspecto clave a la hora de obtener mapas más precisos en términos adaptativos. El listado de variables potencialmente seleccionables puede crearse a partir de:

- a) **Búsquedas bibliográficas:** No resulta difícil encontrar en literatura de tipo técnica y/o científica, como artículos, libros o documentos, algunas referencias sobre los factores ambientales que influyen, determinan o limitan la distribución de una especie. Frecuentemente se debe buscar la correspondencia entre las referencias sobre esos factores con las variables disponibles en forma de capas SIG para realizar el mapa.
- b) **Conocimiento experto:** La consulta de personas expertas en la especie o grupo de especies, sobre qué variables ecogeográficas pueden ser consideradas factores clave en su adaptación y distribución suele ofrecer información muy valiosa a la hora de seleccionar variables. Esta consulta introduce subjetividad en el proceso, algo a lo que no se debe temer en absoluto. En el proceso de creación de mapas de caracterización ecogeográfica del terreno, el conocimiento experto en estas fases previas puede ser la diferencia entre un mapa correctamente validado o un mapa con poco sentido en términos de adaptación de la especie objetivo. El aporte del conocimiento experto puede ser más decisivo en la medida que se consulte el mayor número de expertos posible y luego se determinen consensos. Un buen ejemplo de mapa de caracterización ecogeográfica del terreno en el que se ha utilizado conocimiento experto es el trabajo de Parra Quijano y colaboradores (2012 C).

En este estudio el mapa fue utilizado para la determinación de la localización idónea de reservas genéticas para varias especies del género *Beta* en Europa.

Posterior a la elaboración de este listado potencial de variables, se debe determinar por cada componente (bioclimático, geofísico y edáfico) qué variables pueden estar aportando información redundante. Para ello es importante realizar un análisis de correlaciones bivariadas o un análisis de colinealidad. En caso de encontrarse valores altos de correlación entre dos variables, que hagan parte de un mismo componente, se debería descartar una de ellas. Adicionalmente un análisis de componentes principales (en caso de que todas las variables sean cuantitativas) puede ayudar a determinar las relaciones de las variables y facilitar la selección final. No es recomendable utilizar más de cinco variables por componente pues la configuración de las zonas (celdas adyacentes con el mismo valor) del mapa resultante puede llegar a ser difícil de comprender. Así mismo, utilizar latitud y longitud (parámetros 4.5.1.7 y 4.5.1.8) ofrece mapas con zonas más grandes y mapas menos reticulados. El efecto contrario se obtiene al utilizar variables como "orientación" del componente geofísico.

Una vez determinada la lista final de variables, éstas se seleccionan en los parámetros bioclimv, geophysv y edaphv (parámetros 4.5.1.5, 4.5.1.6 y 4.1.5.9). La lista de variables completa que incluye la herramienta ELC mapas v1.2 se encuentra en la parte final de este documento (Anexos 12.1, 12.2 y 12.3).

## 4.5. Utilización de la herramienta ELC mapas

Una vez instaladas las herramientas CAPFITOGEN y seleccionada la herramienta ELC mapas, se deben especificar una serie de parámetros por parte del usuario.

### 4.5.1 Parámetros iniciales definidos por el usuario

#### 4.5.1.1 Parámetro: ruta

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN.  
Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo L:/CAPFITOGEN , D:/CAPFITOGEN, etc.

#### 4.5.1.2 Parámetro: primvez

Explicación: Si esta es la primera vez que utiliza la herramienta ELC mapas en este/a computador/a, haga clic en esta casilla.

#### 4.5.1.3 Parámetro: país

Explicación: Seleccione el país para el cual quiere construir el mapa ELC. Aparecerán múltiples países disponibles sólo si la herramienta incluye información para todos ellos, de lo contrario solo aparecerá un país determinado.

#### 4.5.1.4 Parámetro: *resol1*

Explicación: Seleccione el nivel de resolución que desea utilizar para la generación del mapa. Tenga en cuenta que 1x1 km ofrece mayor resolución, pero exige mucha más capacidad de cómputo y requiere mucho más tiempo que 5x5 km, especialmente en países de gran extensión. Ver el Anexo 12.5 sobre disponibilidad de resoluciones en relación al país o región seleccionados.

#### 4.5.1.5 Parámetro: *bioclimv*

Explicación: Seleccione las variables bioclimáticas (temperaturas, precipitaciones e índices) que desea incluir para generar el mapa ELC. Puede hacer selección múltiple de variables manteniendo pulsada la tecla Ctrl (control) y añadiendo más variables haciendo clic sobre ellas con el botón izquierdo del ratón.

#### 4.5.1.6 Parámetro: *geophysv*

Explicación: Seleccione las variables geofísicas (relativas al relieve y radiación solar) que desea incluir para generar el mapa ELC. Puede hacer selección múltiple de variables.

#### 4.5.1.7 Parámetro: *latitud*

Explicación: ¿Incluye la latitud en su mapa? Nota: incluyendo latitud y longitud y excluyendo orientación creará unidades ecogeográficas más contiguas y mapas menos reticulados.

#### 4.5.1.8 Parámetro: *longitud*

Explicación: ¿Incluye la longitud en su mapa? Nota: incluyendo latitud y longitud y excluyendo orientación creará unidades ecogeográficas más contiguas y mapas menos reticulados.

#### 4.5.1.9 Parámetro: *edaphv*

Explicación: Seleccione las variables edáficas (texturas, profundidad, pH, etc.) que desea incluir para generar el mapa ELC. Puede hacer selección múltiple de variables.

#### 4.5.1.10 Parámetro: *optim*

Explicación: Indique si requiere proceso de optimización. La optimización solo es recomendable para países de gran extensión (Argentina, Brasil, México) usando alta resolución (1x1 km). De lo contrario puede ralentizar el proceso antes que acelerarlo. Además, si incluye latitud o longitud como variables para crear el mapa ELC, no use "optim".

#### 4.5.1.11 Parámetro: maxg

Explicación: Indique el número máximo de agrupamientos por componente (bioclimático, geofísico y edáfico) que desea permitir (a mayor número, mapas con más categorías). Se recomienda que sean menos de cinco.

#### 4.5.1.12 Parámetro: metodo

Explicación: Seleccione uno de los métodos ofrecidos para generar los agrupamientos con corte objetivo. "elbow" o codo es el método más sencillo y rápido; "medoides" es más sofisticado y consume más recursos.

#### 4.5.1.13 Parámetro: resultados

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

Después de definir todos los parámetros y rutas que requiere ELC mapas, al hacer clic en "Analiza", iniciará el análisis la herramienta. Luego de un tiempo que puede variar por la introducción de algunos parámetros de resolución, tipo de análisis, tamaño de datos procesados o configuración de hardware del computador/a, ELC mapas producirá resultados que guardará donde se le haya indicado (parámetro 4.5.1.13).

## 4.6. Resultados de ELC mapas

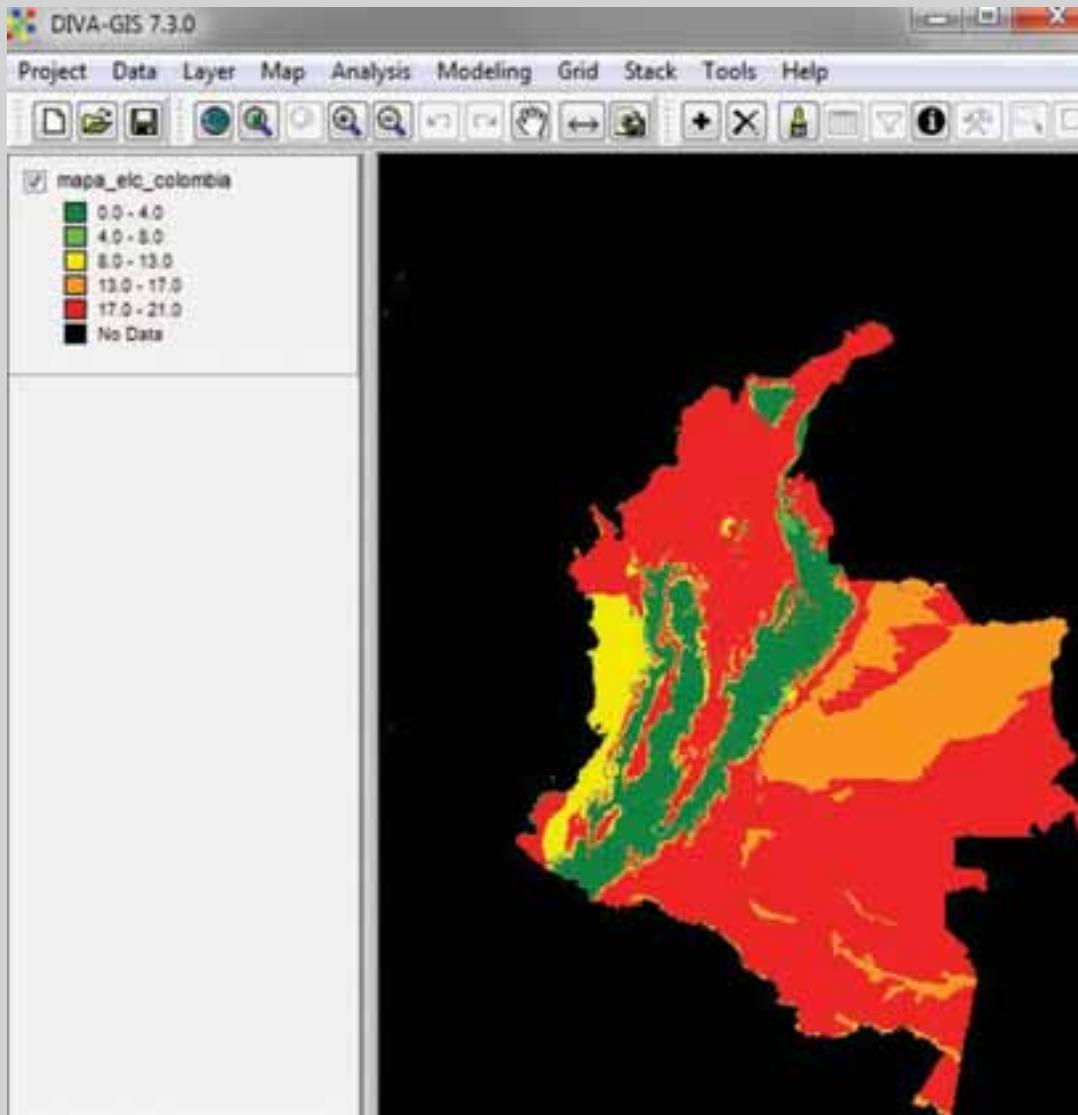
En la ruta y carpeta que haya determinado en "resultados" (parámetro 4.5.1.13) encontrará cinco mapas y tres tablas.

### 4.6.1 Los mapas

Corresponden al mapa ELC final del país o la región determinada en el parámetro 4.5.1.3 (mapa\_elc\_pais.grd, y mapa\_elc\_pais.gri, junto con la imagen mapa\_elc\_pais.png) y los mapas que representan las categorías resultantes de los componentes bioclimático, geofísico y edáfico (mapa\_bioclimatico\_pais.grd, mapa\_bioclimatico\_pais.gri, mapa\_geofisico\_pais.grd, mapa\_geofisico\_pais.gri, mapa\_edafico\_pais.grd y mapa\_edafico\_pais.gri). Todos estos mapas pueden ser abiertos en DIVA-GIS. Inicialmente DIVA-GIS abre los mapas como lo muestra la Fig. 17.

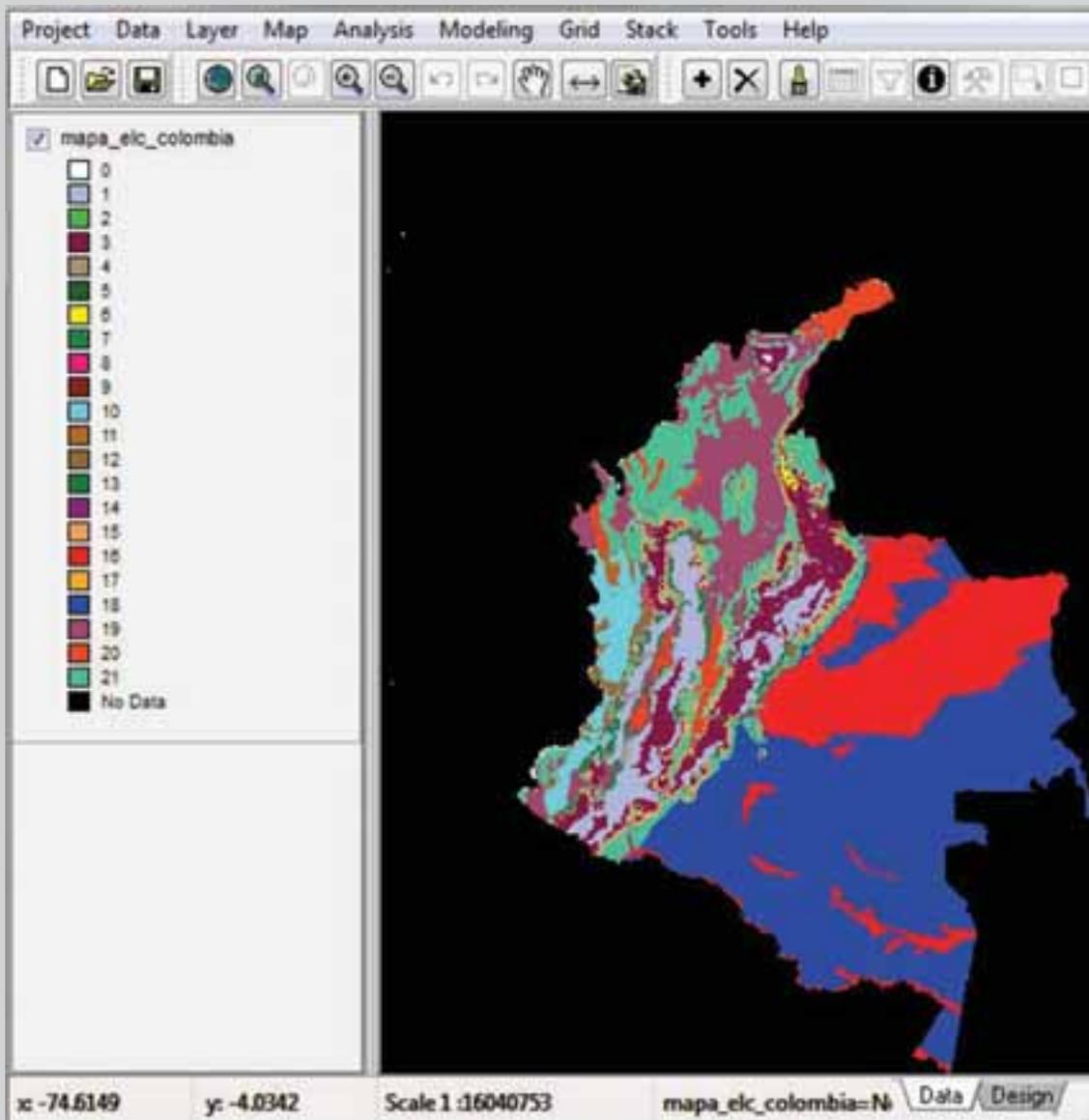
Sin embargo, su visualización se puede alterar al hacer doble clic sobre el cajón gris de la parte izquierda que representa esa capa. Agregando tantas filas hasta alcanzar el número de categorías que contiene el mapa, y luego aplicando una rampa de colores aleatorios, se puede obtener un mapa como el que muestra la Fig. 18. Es conveniente que los colores sean lo más diferentes posible para que se identifiquen perfectamente las categorías (escenarios ecogeográficos) presentes en el territorio.

**FIGURA 17.** Ejemplo de un mapa de Colombia producido por ELC mapas tal y como lo mostraría DIVA-GIS sin ningún tipo de edición. Las 20 categorías se muestran agrupadas en 5 rangos.



**NOTA:** Recuerde siempre que la categoría denominada como "0" (cero) no hace parte de las categorías ecogeográficas del mapa resultado, sino que es la denominación de todas las zonas para las cuales existe información de uno o dos componentes pero no de los tres. Por ejemplo, por obvias razones es normal que la información de suelos para zonas urbanas o cuerpos de agua sea nula, pero si puede existir información bioclimática e incluso geofísica para estas áreas. Esas áreas serán codificadas como "0".

**FIGURA 18.** Ejemplo de un mapa de Colombia producido por ELC mapas que muestra un color por cada una de sus categorías. Las propiedades del mapa tal y como lo abre DIVA-GIS fueron alteradas para mostrar cada categoría con un color diferente.



Adicionalmente a los mapas compatibles con DIVA-GIS se produce un mapa compatible con Google Earth, "mapa\_elc\_pais.kml". Siempre que se tenga instalado Google earth, este mapa se podrá abrir como una capa sobre las imágenes de Google earth al hacer doble clic sobre el archivo. Este mapa no puede ser manipulado (cambio de colores) y su calidad gráfica no es óptima.

## 4.6.2 Las tablas

Las tablas producidas por ELC mapas están en formato de texto delimitado por tabulaciones y pueden ser abiertas en programas como Excel, OpenOffice, o R. Normalmente al hacer clic con el botón derecho del ratón, en la opción "Abrir con" se ofrecen algunos de estos programas si están instalados.

4.6.2.1 "Tabla\_ELC\_celdas\_pais.txt". Esta tabla muestra los valores de las variables seleccionadas y los valores de las categorías ELC ("ELC\_CAT"), bioclimáticas, geofísicas y edáficas para cada centroide (fila) de cada celda que compone el territorio del país objeto de estudio. También incluye los valores de latitud y longitud de cada centroide.

4.6.2.2 "numero\_categorias\_pais.txt". Contiene un simple recuento de las categorías ecogeográficas producidas y representadas en el mapa ELC resultante (columna "N\_ELC\_CAT") y el número de categorías producidas por componente.

4.6.2.3 "Estadist\_ELC\_pais.txt", "Estadist\_BIOCLIM\_pais.txt", "Estadist\_EDAPH\_pais.txt" y "Estadist\_GEOPHYS\_pais.txt". En estas tablas se encuentran los estadísticos descriptivos (media, valor mínimo, valor máximo y desviación estándar) para cada una de las variables originales involucradas en la creación del mapa ELC y para los mapas de cada uno de los componentes (bioclimático, geofísico y edáfico) representados en el mapa ELC. Estas tablas se asemejan a la tabla suplementaria S2 presentada a manera de descripción de las categorías del mapa ELC en el artículo de Parra Quijano y colaboradores (2012 A).

## 4.7. Referencias

Kaufman, L. y Rousseeuw, P.J. 1987, Clustering by means of Medoids, in Statistical Data Analysis Based on the L1-Norm and Related Methods. Y. Dodge (eds), North-Holland, 405–416.

Ketchen, D. J. y Shook, C. L. 1996. The application of cluster analysis in Strategic Management Research: An analysis and critique. *Strategic Management Journal* 17(6): 441–458.

Parra-Quijano, M. Iriondo, J.M., De la Cruz, M., Torres, M.E. 2011 A. Strategies for the development of core collections based on ecogeographical data. *Crop Science* 51:656-666

Parra-Quijano, M. Iriondo, J.M., Torres, M.E., De la Rosa, L. 2011 B. Evaluation and validation of ecogeographical core collections using phenotypic data. *Crop Science* 51:694-703

Parra-Quijano, M.; Draper, D.; Torres, E. e Iriondo, J.M. 2008. Ecogeographical representativeness in crop wild relative *ex situ* collections. p. 249-273. *In* Maxted, N.; Ford-Lloyd, B.V.; Kell, S.P.; Iriondo, J.M.; Dulloo, M.E. y Turok, J. (ed.) *Crop wild relative conservation and use*. CAB International, Wallingford.

Parra-Quijano, M. Iriondo, J.M., Torres, M.E. 2012 A. Ecogeographical land characterization maps as a tool for assessing plant adaptation and their implications in agrobiodiversity studies. *Genetic Resources and Crop Evolution* 59(2):205-217 DOI 10.1007/s10722-011-9676-7

Parra-Quijano, M. Iriondo, J.M., Torres, M.E. 2012 B. Improving representativeness of genebank collections through species distribution models, gap analysis and ecogeographical maps. *Biodiversity and Conservation* 21:79-96 DOI 10.1007/s10531-011-0167-0

Parra-Quijano, M. Iriondo, J.M., Frese, L., Torres, M.E. 2012 C. Spatial and ecogeographic approaches for selecting genetic reserves in Europe. *En*: N. Maxted, M.E. Dulloo, B.V. Ford-Lloyd, L. Frese, J. Iriondo y M.A.A. Pinheiro de Carvalho (ed.) *Agrobiodiversity Conservation: securing the diversity of crop wild relatives and landraces*. CABI, Wallingford, UK.

Rousseeuw, P.J. 1987. "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". *Computational and Applied Mathematics* 20: 53–65. doi:10.1016/0377-0427(87)90125-7.

Thormann, I. 2012. Applying FIGS to crop wild relatives and landraces in Europe. *Crop Wild Relative* 8 14:16. <http://www.pgrsecure.org/publications>

Tohme, J., Jones, P., Beebe, S. y Iwanaga, M. 1995. The combined use of agroecological and characterisation data to establish the CIAT *Phaseolus vulgaris* core collection. p. 95-107. In Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L. y Morales, E.A.V. (eds.) Core collections of plant genetic resources. IPGRI, Rome.



## 5. Herramienta ECOGEO

### 5.1. Caracterización ecogeográfica del germoplasma

Se entiende caracterización ecogeográfica como el análisis de toda información ambiental del sitio donde crece un individuo o una población vegetal, directamente relacionada con el proceso de adaptación al entorno biótico o abiótico. En particular, las herramientas CAPFITOGEN permiten analizar sólo el componente abiótico, clasificado por tres aspectos principales, los cuales suelen considerarse en estudios de adaptación de cultivos (Ceballos-Silva y López-Blanco, 2003) o zonificación agrícola (Williams et al., 2008):

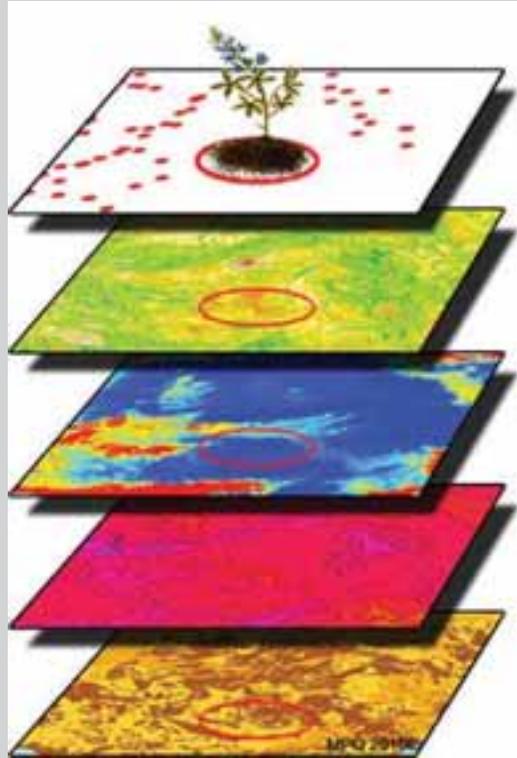
- a) Bioclimático: Hace referencia a factores relacionados con temperatura y precipitación. También incluye las relaciones entre temperatura y precipitación que se manejan a través de índices.
- b) Geofísico: Agrupa los factores topográficos y de relieve de mayor importancia, especialmente relacionados con la radiación solar.
- c) Edáfico: Factores relacionados con las condiciones físicas y/o químicas de la fracción del suelo de la que las plantas dependen.

Por tanto, caracterizar ecogeográficamente un conjunto de entradas implica asignar a cada una de ellas, información de tipo bioclimático, geofísico y edáfico del sitio de su recolección.

La información ecogeográfica del sitio de recolección por si sola muestra muchos rasgos adaptativos del germoplasma, y en conjunto con otro tipo de caracterizaciones, como la fenotípica o genotípica, puede resultar muy útil para explicar los patrones genéticos observados. En algunos casos, como los que se dan en ausencia de recursos económicos suficientes para otro tipo de estudios, la caracterización ecogeográfica puede considerarse una alternativa válida, sencilla y barata para facilitar la utilización del germoplasma por parte de mejoradores que buscan parentales con ciertos rasgos adaptativos en las colecciones.

El insumo más importante para realizar una caracterización ecogeográfica son las coordenadas o la descripción del sitio de recolección (de la que se pueda extraer una coordenada), que usualmente se registran en los descriptores de pasaporte en el momento de la recolección. Usando dichas coordenadas se asignará a cada entrada los datos que describen los rasgos ambientales más importantes del sitio donde fue recolectada. Esto implica que la calidad de esas coordenadas es un aspecto crucial para una asignación correcta de información ecogeográfica, razón por la que se sugiere el uso de la herramienta GEOQUAL antes de realizar una caracterización de este tipo.

**FIGURA 19.** Ilustración del proceso de extracción de información ecogeográfica que se realiza para un sitio de recolección a través de SIG.



Aparte de las coordenadas como materia prima, una caracterización ecogeográfica requiere de información ambiental que cubra toda el área de trabajo y de un programa especializado para la extracción de la información correspondiente a cada sitio de recolección. Los SIG son el tipo de programa que permiten realizar este proceso.

El producto de una caracterización ecogeográfica es similar al de cualquier otro tipo de caracterización, una matriz de datos, donde las filas suelen corresponder a las entradas y las columnas a los descriptores. A partir de esta matriz inicial es posible realizar análisis multivariados, como los que frecuentemente se realizan con otros tipos de caracterización, que permiten en este caso determinar relaciones de similitud ambiental entre los diferentes sitios de recolección. Un análisis factorial, como por ejemplo el de Componentes Principales (ACP), permitiría también conocer las relaciones entre las diferentes variables originalmente introducidas y crear variables sintéticas no correlacionadas que pudieran describir las afinidades ecogeográficas entre las entradas con un número reducido de componentes.

Es importante resaltar que la caracterización ecogeográfica arroja información sobre los sitios de recolección, no directamente de la naturaleza del germoplasma.

Por esto, los análisis de multivariados que funcionan sobre matrices de distancia o disimilitud, en este caso lo que reflejan es la afinidad ambiental e, indirectamente, adaptativa que hay entre los diferentes sitios de recolección. Entradas de una misma especie con patrones genotípicos o fenotípicos diferentes pueden ocurrir en escenarios ambientales muy parecidos o incluso indiferenciables.

## 5.2. Características de ECOGEO

La herramienta ECOGEO pone a disposición del usuario información ecogeográfica de más de cien variables (ver Anexos 12.1, 12.2 y 12.3), extraídas para una lista de entradas que el usuario introduzca en el análisis mediante el formato de datos de pasaporte FAO/Biodiversidad 2012, con modificaciones menores, formato que se utiliza en varias de las herramientas CAPFITOGEN.

La herramienta ECOGEO contiene toda la información necesaria, lo cual permite que el usuario no tenga que descargar ninguna información desde internet (no se requiere conexión, al igual que para las demás herramientas CAPFITOGEN). La información o capas de variables ecogeográficas están adaptadas y dispuestas para funcionar con las programaciones R de la herramienta.

El área de trabajo es un segundo aspecto que define el usuario. Frecuentemente el área de trabajo corresponde a los límites territoriales de un estado, tal y como viene definido en la base de datos global sobre áreas administrativas (<http://www.gadm.org>). Las variables o capas ecogeográficas están recortadas según sus límites, de tal manera que si se elige un país determinado y los datos de pasaporte incluyen coordenadas que indican sitios por fuera de este país, a las entradas correspondientes a dichas coordenadas no se les asigna ninguna información. Podrían existir opciones para hacer análisis más allá de las fronteras de un país determinado, que es lo que se considera una región o incluso un continente. Si hubiera disponibilidad de regiones o continentes dentro de la herramienta, (aparecerían en los listados del parámetro "país"), el usuario puede usar estas áreas de trabajo de mayor cubrimiento, sabiendo que seguramente el nivel de resolución de esta información será de menor alcance (tamaños de celda sobre los 10x10 km).

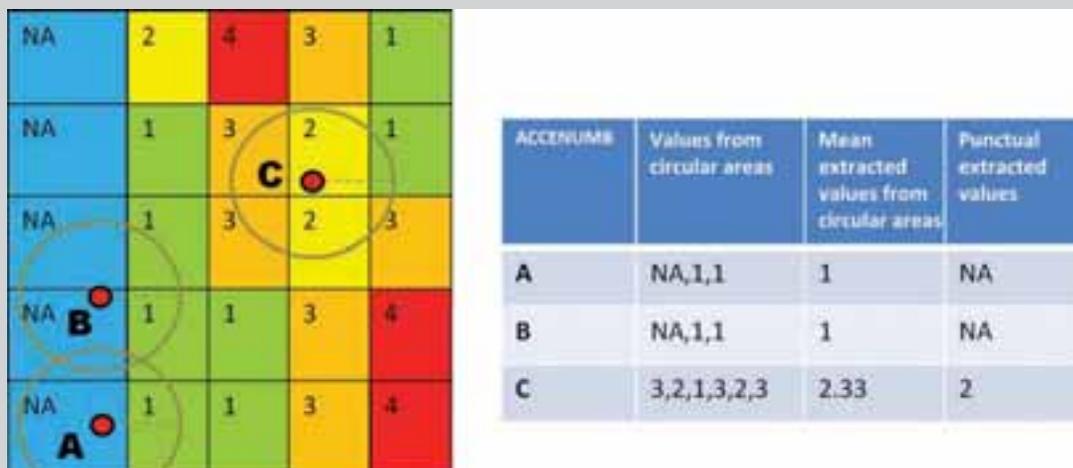
Hay un aspecto importante que introduce la herramienta ECOGEO. Se trata de la forma en la que se extrae la información ecogeográfica para un sitio de recolección. Usualmente las extracciones se realizan para el punto que señala la coordenada. Sin embargo, ocurren dos situaciones en las que la extracción "puntual" no refleje las condiciones abióticas más reales del sitio de recolección:

- A) Cuando se tiene poca información sobre las coordenadas o en general éstas no son de buena calidad, según lo puede señalar GEOQUAL u otras metodologías. Incluso en casos de especies de distribución costera, donde a pesar de una georreferenciación relativamente apropiada, una extracción puntual podría arrojar muchos valores "NA" (no información disponible) dado que los mapas/capas ráster de información ecogeográfica no se ajustan perfectamente al contorno de las líneas costeras.

- B) Cuando por diversas razones el sitio de recolección del germoplasma no corresponde exactamente al sitio donde la planta crece, pero éste se encuentra dentro de un perímetro relativamente conocido (por ejemplo cuando se recolecta germoplasma en mercados locales).

En estos casos el usuario puede utilizar la extracción de tipo "radial". En ella, el usuario indica a la herramienta el radio del área circular alrededor del punto que indica la coordenada, para la cual desea extraer información. De esta manera, ECOGEO extrae datos ecogeográficos de todas las celdas que se encuentran al interior del área circular, calcula su valor medio y asigna este valor a la entrada, repitiendo este proceso para todas las variables ecogeográficas por las que se pretende caracterizar el germoplasma (ver Fig. 20). Adicionalmente ECOGEO descarta valores "NA" para el cálculo de medias y la subsecuente asignación de valores. Para indicarle a ECOGEO que se desea realizar una extracción "radial", el usuario debe activar el parámetro buffy (apartado 5.3.1.6) y a continuación introducir el valor en metros del radio del área circular de extracción en el parámetro tamp (apartado 5.3.1.7).

**FIGURA 20.** Diferencias entre los valores asignados a partir de una extracción de tipo puntual y una radial. Las celdas en color azul y valores NA representan cuerpos de agua y los puntos de color rojo tres sitios de recolección (identificados con códigos ACCENUMB) ubicados a partir de sus coordenadas.



Una vez el usuario ha preparado su tabla de pasaporte de acuerdo al formato preestablecido, le ha indicado a la herramienta donde se encuentra, ha indicado el área de trabajo, el nivel de resolución y la forma de extracción que desea, sólo le resta definir las variables/capas de interés de cada aspecto (bioclimático, geofísico y edáfico) con las que quiere caracterizar los sitios de recolección del germoplasma.

Con la definición de estos parámetros, la herramienta ECOGEO buscará las variables/capas de información ecogeográfica de interés, las agrupará y extraerá la

información para cada coordenada desde el grupo de capas en un solo paso. Con la información extraída generará una tabla que se salvará en la dirección que el usuario defina en el parámetro "resultados".

Finalmente, si el usuario estuviera interesado en realizar un análisis de agrupamiento o un Análisis de Componentes Principales (ACP), puede indicar a la herramienta que ejecute dichos análisis y además puede señalar que tipo de agrupamiento desea o el número de componentes principales que desea retener, respectivamente. La herramienta ECOGEO producirá gráficos (dendrogramas o biplots) y tablas (valores y vectores principales y puntuaciones para los componentes retenidos) que se guardarán en la carpeta indicada en el parámetro "resultados".

### 5.3. Utilización de la herramienta ECOGEO

Una vez instalada las herramientas CAPFITOGEN y seleccionada la herramienta ECOGEO, será necesario definir una serie de parámetros para que la programación R funcione correctamente.

#### 5.3.1 Parámetros iniciales definidos por el usuario

##### 5.3.1.1 Parámetro: ruta

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN.  
Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

##### 5.3.1.2 Parámetro: pais

Explicación: Seleccione el país en el cual fueron recolectadas la mayoría o la totalidad de las entradas que desea analizar. Si las entradas han sido recolectadas en más de un país, considere seleccionar una región, subcontinente o continente (progresivamente se añadirán estas opciones).

##### 5.3.1.3 Parámetro: pasaporte

Explicación: Escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama 'tabla' debe escribir: 'tabla.txt'. Recuerde que este archivo debe ser guardado previamente en la carpeta 'Pasaporte' que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN.

##### 5.3.1.4 Parámetro: geoqual

Explicación: Seleccione esta opción si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50 columnas (y no las 45 del modelo de pasaporte que usan las herramientas CAPFITOGEN). Use por tanto la tabla

resultado de GEOQUAL v.2 denominada "PasaporteOriginalEvaluadoGEOQUAL.txt" como tabla de pasaporte en el punto anterior.

#### 5.3.1.5 Parámetro: *totalqual*

Explicación: Si su tabla de pasaporte proviene de GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

#### 5.3.1.6 Parámetro: *buffy*

Explicación: Marque esta opción si desea que la extracción de información ecogeográfica se haga sobre un área alrededor del sitio de recolección. Dejar sin marcar esta opción lleva a que la extracción se realice sólo para el punto que indican las coordenadas del sitio de recolección.

#### 5.3.1.7 Parámetro: *tamp*

Explicación: Especifique el radio (en metros) de un área circular alrededor del punto que indican las coordenadas del sitio de recolección, de la que se extraerá información ecogeográfica. Los valores extraídos del área circular serán promediados para obtener un único valor y no se tendrán en cuenta celdas sin valor. Se recomienda que este valor no sea inferior a la distancia de cada lado de celda del parámetro resol1.

#### 5.3.1.8 Parámetro: *resol1*

Explicación: Seleccione el nivel de resolución que desea utilizar para la extracción de información ecogeográfica. Note que 1x1 km ofrece mayor resolución pero exige más capacidad de cómputo en relación a 5x5 km, si bien no es un aspecto tan limitante como en la herramienta ELC mapas. Las resoluciones 10x10 y 20x20 km están restringidas para países de gran extensión, subcontinentes o continentes. Ver el Anexo 12.5 sobre disponibilidad de resoluciones en relación al país o región seleccionado.

#### 5.3.1.9 Parámetro: *bioclimsn*

Explicación: Seleccione esta opción si desea caracterizar por variables de tipo bioclimático (temperatura, precipitación e índices asociados).

#### 5.3.1.10 Parámetro: *bioclimv*

Explicación: Seleccione las variables bioclimáticas que desea incluir en la caracterización ecogeográfica. El total de variables seleccionables se detalla en el Anexo 12.1.

#### 5.3.1.11 *Parámetro: edaphsn*

Explicación: Seleccione esta opción si desea caracterizar por variables de tipo edáfico (texturas, profundidad, pH, etc.).

#### 5.3.1.12 *Parámetro: edaphv*

Explicación: Seleccione las variables edáficas que desea incluir en la caracterización ecogeográfica. El total de variables seleccionables se detalla en el Anexo 12.2.

#### 5.3.1.13 *Parámetro: geophysn*

Explicación: Seleccione esta opción si desea caracterizar por variables de tipo geofísico (relativas al relieve y radiación solar).

#### 5.3.1.14 *Parámetro: geophysv*

Explicación: Seleccione las variables geofísicas que desea incluir en la caracterización ecogeográfica. El total de variables seleccionables se detalla en el Anexo 12.3.

#### 5.3.1.15 *Parámetro: latitud*

Explicación: Incluye la latitud como variable de caracterización ecogeográfica?.

#### 5.3.1.16 *Parámetro: longitud*

Explicación: Incluye la longitud como variable de caracterización ecogeográfica?.

#### 5.3.1.17 *Parámetro: ecogeoclus*

Explicación: Seleccione esta opción si desea realizar un análisis de agrupamientos de las entradas caracterizadas ecogeográficamente.

#### 5.3.1.18 *Parámetro: ecogeoclustype*

Explicación: Elija el tipo de agrupamiento jerárquico que desea utilizar para los agrupamientos ecogeográficos: "single" = vecino más cercano, "complete" = vecindario más compacto, "ward" = método de mínima varianza de Ward, "mcquitty" = método de McQuitty, "average" = similaridad promedio (UPGMA), "median" = similaridad de la mediana, "centroid" = centroide geométrico, "flexible" = Beta flexible.

#### 5.3.1.19 *Parámetro: ecogeopca*

Explicación: Seleccione esta opción si desea realizar un análisis de componentes principales para las entradas caracterizadas ecogeográficamente.

#### 5.3.1.20 Parámetro: *ecogeopcaxe*

Explicación: Numero de componentes a retener (siempre menor al número de variables ecogeográficas) dentro del análisis PCA.

#### 5.3.1.21 Parámetro: *resultados*

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

### 5.4. Resultados de ECOGEO

En la ruta y carpeta que haya determinado en “resultados” (parámetro 5.3.1.21) encontrará dos figuras y cuatro tablas.

#### 5.4.1 Las figuras

Se trata de los archivos *dendrograma\_ecogeo.wmf* y *pca\_ecogeo.wmf* que corresponden a figuras vectoriales en formato metaarchivo de Windows. Estas figuras (un dendrograma y un biplot respectivamente) sólo se generan si se ha solicitado a la herramienta realizar un análisis de agrupamientos (parámetro 5.3.1.17) o un análisis de componentes principales (parámetro 5.3.1.19). Pueden abrirse e incluso modificarse en Microsoft Powerpoint o programas de edición de imágenes.

#### 5.4.2 Las tablas

Las cuatro tablas corresponden a dos tipos de información.

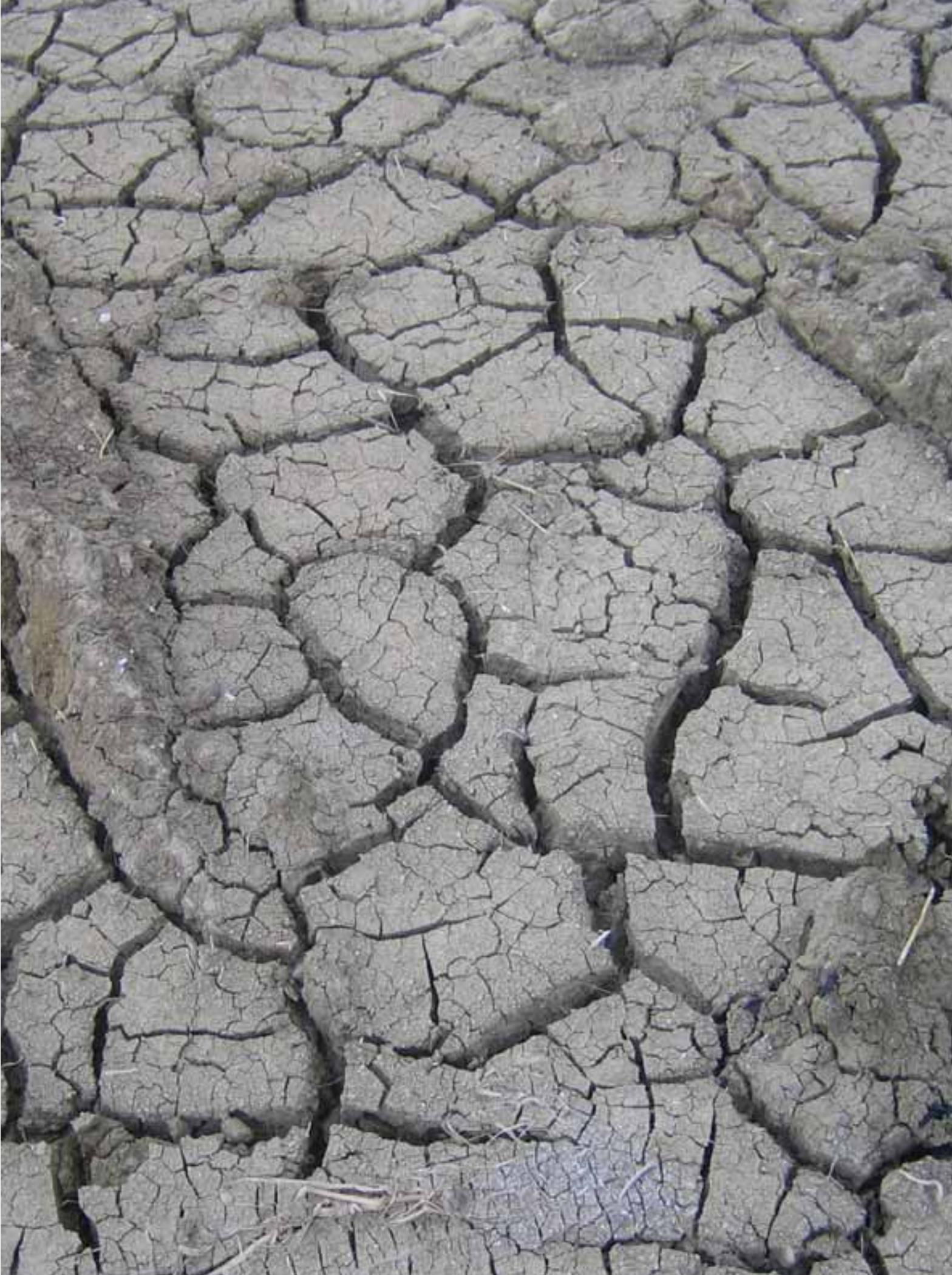
5.4.2.1 Tabla de caracterización ecogeográfica de las entradas: Se trata del archivo *TablaVarEcogeograficapais.txt*. Corresponde a la matriz inicial de caracterización, con tantas filas como entradas analizadas y tantas columnas como descriptores ecogeográficos.

5.4.2.2 Tablas resultado del análisis de componentes principales: Corresponden a los archivos *ecogeographic\_eigenvalues.txt* (tabla con los valores propios), *ecogeographic\_eigenvectors.txt* (tabla con los vectores propios) y *ecogeographic\_pcascotes.txt* (tabla con las puntuaciones de cada entrada para los componentes principales retenidos). Sólo se generarán si se ha solicitado a la herramienta realizar dicho análisis (parámetro 5.3.1.19).

## 5.5. Referencias

Ceballos-Silva, A. y López-Blanco, J. 2003. Evaluating biophysical variables to identify suitable areas for oat in Central Mexico: a multi-criteria and GIS approach. *Agriculture, Ecosystems and Environment* 95 (2003) 371–377.

Williams, C.L., Hargrove, W.W., Liebman, M. y James, D.E. 2008. Agro-ecoregionalization of Iowa using multivariate geographical clustering. *Agriculture, Ecosystems and Environment* 123 (2008) 161–174



## 6. Herramienta Representa

### 6.1. Concepto de representatividad en colecciones de germoplasma

Los aspectos sensibles que pueden poner en riesgo una conservación *ex situ* de recursos fitogenéticos exitosa se pueden dar en dos momentos particulares, o en el momento de la colecta o durante la conservación propiamente dicha. A través de la aplicación de técnicas apropiadas para el manejo del germoplasma, se puede reducir el riesgo de pérdidas de entradas durante el periodo de conservación. Sin embargo, el germoplasma que se lleva a conservación debe ser el reflejo de la diversidad genética más fiel posible de las poblaciones vegetales que ocurren en el campo. Durante la conservación en el mejor de los casos y sin que medien nuevas recolecciones, ese reflejo se mantendrá intacto. Esta situación pone en evidencia la importancia de realizar recolecciones de germoplasma que aseguren capturar la mayor diversidad genética posible. Qué tanto se parece la muestra de la diversidad conservada *ex situ* respecto a la diversidad genética total que ocurre en la naturaleza, es lo que se denomina **representatividad** de una colección de germoplasma.

La representatividad de una especie dentro de una colección de germoplasma puede determinarse a nivel intra e interpoblacional. Para el caso de una especie cultivada, el equivalente sería niveles intra e intervarietales. Son dos conceptos que al considerarse de forma global la representatividad de una colección son insolubles. A pesar de esto, y por cuestiones prácticas relacionadas con la forma de conservación del germoplasma, ambos conceptos se han trabajado por separado.

La representatividad intrapoblacional ha sido exhaustivamente estudiada como en los múltiples trabajos de Crossa y colaboradores (1994, 1997, 2011), lo cual ha desembocado en estrategias de recolección específicas de acuerdo a la biología reproductiva de la especie, a la distribución espacial de los individuos y al tamaño de la población. Básicamente lo que se busca es calcular, según el caso, el número mínimo de individuos a recolectar para asegurar la captura de la mayor parte de los alelos presentes en la población. En contraste, la forma como debería representarse interpoblacionalmente una especie en una colección ha sido menos abordada. Sin embargo, desde el desarrollo del concepto de colecciones núcleo o nucleares, la representatividad interpoblacional de una especie en una colección ha ganado importancia, dado que éstas subcolecciones sólo operan a este nivel (Brown, 1989; Yonezawa et al., 1995).

Una vez el concepto de representatividad de una colección de germoplasma se implanta en la comunidad de científicos y curadores en el ámbito de los recursos fitogenéticos, el siguiente paso fue determinar la forma más apropiada de calcularla. Si el objetivo de la conservación *ex situ* es capturar y mantener la mayor diversidad genética posible de una especie, la determinación ideal de la representatividad

sería en términos genéticos. En consecuencia, la fórmula para determinar la representatividad genética (RG) en porcentaje sería:

$$RG = (NAC * 100) / NAT$$

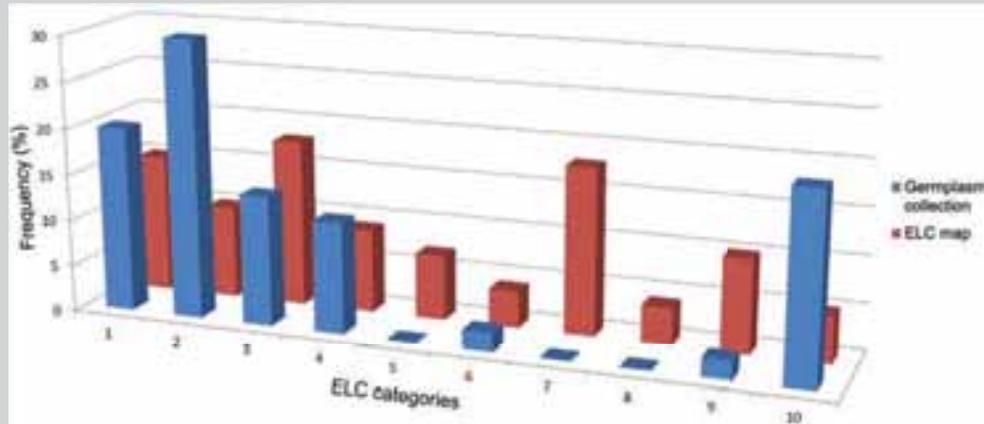
Donde NAT corresponde al número total de alelos de la suma de todos los loci estudiados que presenta la especie objetivo dentro del ámbito espacial (continente, país, región, etc.) de la colección a evaluar y NAC es el número de alelos de tales loci de esta especie capturados por dicha colección. Esta determinación ideal de la representatividad genética conlleva un impedimento de orden práctico. Conocer el número total de alelos que puede contener una especie en un territorio suficientemente grande como un país (espacio de trabajo usual de una colección de germoplasma de un Programa Nacional) o incluso a niveles muy inferiores a éste, resulta una tarea en la práctica inabordable para cualquier especie, excepto para aquellas en las que se sabe con certeza que están compuestas de muy pocas poblaciones. Si se considera el contexto de los recursos fitogenéticos para alimentación y agricultura, esta excepción es casi inexistente. Adicionalmente, el intentar calcular la RG lleva a que indirectamente se haya representado el 100% de los alelos, si el muestreo de todas las poblaciones implicase recolección de germoplasma. En otras palabras, si calcular la RG de una colección de germoplasma implica recolectar muestras y germoplasma de todas las poblaciones que componen la especie dentro de un territorio de trabajo, entonces, independientemente de lo difícil que pueda resultar esta tarea, la máxima representatividad ya sería alcanzada siempre que en la recolección se hayan seguido criterios adecuados de representatividad intrapoblacional.

Estas dificultades prácticas y logísticas han llevado a pensar en otras alternativas para determinar la representatividad de una colección. La representatividad ecogeográfica (RE) fue planteada por Parra-Quijano y colaboradores (2008) en el caso especial de colecciones *ex situ* de plantas silvestres emparentadas con cultivadas (CWR en sus siglas en inglés). En ella se plantea la posibilidad de usar mapas de caracterización ecogeográfica del territorio (como los producidos por ELC mapas) para conocer que tantas condiciones ambientales presentes en un marco espacial estarían representados en una colección de germoplasma.

Como ejemplo de dicha aplicación, la Fig. 21 muestra una distribución de las frecuencias para cada categoría ELC de una colección de germoplasma en contraste con la disponibilidad de dichas categorías en el total del marco espacial. Este ejemplo ficticio sirve para mostrar cómo la representatividad de una colección puede o no (como es el caso) estar sesgada respecto a la abundancia de unidades ambientales presentes en el espacio de trabajo. El contraste de valores para las categorías 2 y 7 dejan entrever que las dos distribuciones no guardan parecido y posiblemente una prueba de Chi cuadrado determinaría una asociación no significativa de las dos distribuciones.

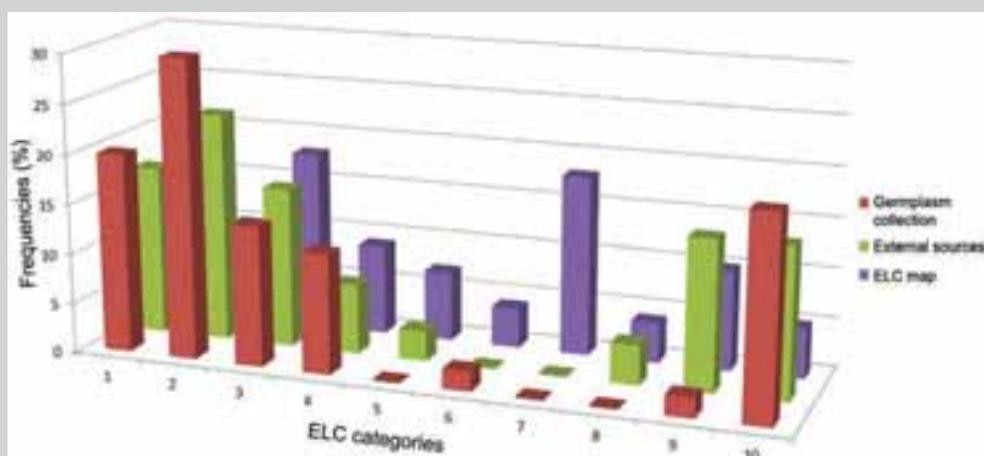
Sin embargo, la determinación más exacta de la RE se consigue a través de un análisis de faltantes o vacíos (más conocido por su denominación en inglés como "gap analysis"). Para realizarlo es necesario previamente compilar información de

**FIGURA 21.** Comparación de la representación de cada categoría ELC en la colección de germoplasma y la disponibilidad total de dichas categorías en el mapa ELC, medida en valores de frecuencias (en porcentaje).



otras fuentes externas a la colección, como otras colecciones de germoplasma o cualquier dato que indique la presencia de poblaciones de la especie objetivo (pliegos de herbario, bases de datos botánicas, referencias bibliográficas, etc.). Luego se compara la distribución de la frecuencia de las recolecciones de la colección a evaluar respecto a la de las fuentes externas y entonces se podrá visualizar claramente qué ambientes están subrepresentados en la colección.

**FIGURA 22.** Comparación de distribución de las frecuencias de sitios de recolección de la colección objetivo y de presencias de fuentes externas sobre diez categorías ELC. Se incluye también la distribución de la frecuencia de cada categoría en el total del mapa ELC.



La Fig. 22 ilustra el proceso de comparación previamente mencionado. Usando los mismos datos ficticios del ejemplo de la Fig. 21, en esta gráfica de barras se incluye (en color verde) la distribución de las frecuencias de las categorías ELC para datos de presencia provenientes de fuentes externas. En este caso es evidente el parecido entre las distribuciones colección objetivo y fuentes externas, aunque algunas diferencias son especialmente interesantes. Para las categorías 5 y 8, las fuentes externas indican presencia de la especie en esa unidad ambiental la cual no estaría representado en la colección, evidenciando la presencia de faltantes o vacíos ecogeográficos. Estos vacíos pueden ser útiles para la planificación de nuevas recolecciones de germoplasma, pues se puede priorizar la visita de estos ambientes pues se conoce donde ocurren estas poblaciones provenientes de fuentes externas.

Es importante hacer una aclaración respecto al tratamiento que se puede dar a los datos provenientes de fuentes externas. Tomando las presencias provenientes de otro banco de germoplasma como fuente externa, se puede conocer la representatividad de la colección objetivo globalmente, pero usar dichos datos para determinar sitios prioritarios para nuevas recolecciones puede llevar a recolectar duplicados intercolecciones.

### 6.3. Utilización de la herramienta Representa

Una vez instalada las herramientas CAPFITOGEN y seleccionada la herramienta Representa, será necesario definir una serie de parámetros para que la programación R funcione correctamente.

#### 6.3.1 Parámetros iniciales definidos por el usuario

##### 6.3.1.1 Parámetro: ruta

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN.  
Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

##### 6.3.1.2 Parámetro: internet

Explicación: Si desea descargar información sobre fuentes externas desde bases de datos de internet (GBIF) y tiene acceso a la red, por favor seleccione esta opción.

##### 6.3.1.3 Parámetro: pasaporte

Explicación: Escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama "tabla" debe escribir: "tabla.txt". Recuerde que este archivo debe ser guardado previamente en la carpeta "Pasaporte" que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN.

#### 6.3.1.4 Parámetro: *geoqual*

Explicación: Seleccione esta opción si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50 columnas (y no las 45 del modelo de pasaporte que usan las herramientas CAPFITOGEN). Use por tanto la tabla resultado de GEOQUAL v.2 denominada PasaporteOriginalEvaluadoGEOQUAL.txt como tabla de pasaporte en el punto anterior.

#### 6.3.1.5 Parámetro: *totalqual*

Explicación: Si su tabla de pasaporte proviene de GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

#### 6.3.1.6 Parámetro: *fext*

Explicación: Dispone de ocurrencias provenientes de fuentes externas (cualquier fuente de información diferente a la colección objeto del análisis de representatividad) en el formato requerido?.

#### 6.3.1.7 Parámetro: *fuentex*

Explicación: Indique el nombre del archivo que contiene las ocurrencias de fuentes externas en el formato requerido. Si el archivo se llama "FuentesExternas", en este campo debe aparecer "FuentesExternas.txt" (porque esta tabla debe estar en formato de texto delimitado por tabulaciones). No olvide que este archivo debe encontrarse en la carpeta llamada Pasaporte.

#### 6.3.1.8 Parámetro: *geoqualfe*

Explicación: La tabla de ocurrencias de fuentes externas, en el formato requerido, dispone en las columnas indicadas de información sobre evaluación de la calidad de la georreferenciación (aplicación previa de GEOQUAL)?.

#### 6.3.1.9 Parámetro: *totalqualfe*

Explicación: Si su tabla de ocurrencias de fuentes externas fue evaluada por GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

#### 6.3.1.10 Parámetro: *duplibg*

Explicación: Marque esta opción si considera como faltante las ocurrencias de fuentes externas que provienen de otros bancos o colecciones de germoplasma (Columna TYPESOURCE con valor 40). Tenga en cuenta que al marcar esta opción,

Ud. podría llegar a realizar recolecciones de poblaciones que otras colecciones ya tienen representadas, generando duplicación inter-colecciones. Atención: Marque esta opción si todas las ocurrencias de fuentes externas que introduce provienen de bancos o colecciones de germoplasma. De no hacerlo se generará un error.

#### 6.3.1.11 Parámetro: *gbifFE*

Explicación: Desea descargar datos de ocurrencia de fuentes externas desde el portal de Global Biodiversity Information Facility (GBIF)?. Esta opción requiere conexión a internet. Esta opción es incompatible con la inclusión de ocurrencias de fuentes externas aportadas por el usuario. En caso de marcar esta opción y aportar a la vez una tabla con ocurrencia de fuentes externas, solo se tendrá en cuenta esta última.

#### 6.3.1.12 Parámetro: *genero*

Explicación: Escriba el nombre del género de la especie objetivo del presente análisis. Este será el género para el cual se descargara información desde el portal de GBIF. No olvide primera letra en mayúscula. Si no activa el parámetro *gbifFE*, no es necesario que indique aquí ninguna información.

#### 6.3.1.13 Parámetro: *especie*

Explicación: Escriba el nombre de la especie (solo el epíteto) objetivo del presente análisis. Este nombre se unirá al del género para realizar la solicitud y descarga de información desde GBIF. Si desea descargar información para todo el género, solamente escriba aquí un \*. Todo el epíteto debe estar escrito en minúsculas. Si no activa el parámetro *gbifFE*, no es necesario que indique aquí ninguna información.

#### 6.3.1.14 Parámetro: *mapaelc*

Explicación: Indique el nombre del archivo que contiene el mapa ELC (producto de la aplicación de la herramienta ELCmapas), el cual debe encontrarse en la carpeta ELCmapas, que corresponde a una de las carpetas que componen el directorio CAPFITOGEN. Este mapa debe estar en formato DIVA-GIS, compuesto por los dos archivos con extensiones ".grd" y ".gri", tal y como lo produce ELCmapas. En esta caja de texto se debe escribir el nombre del archivo con la extensión ".grd". Así, si el nombre del mapa es "mapa\_elc\_spain", ud. debe escribir "mapa\_elc\_spain.grd".

#### 6.3.1.15 Parámetro: *statelc*

Explicación: Indique el nombre del archivo que contiene la tabla de estadísticas descriptivas del mapa ELC producido con la herramienta ELCmapas (esta herramienta suele nombrar a este archivo como "Estadist\_ELC\_" más el nombre del país, correspondiente al resultado 4.6.2.3). Al igual que el mapa ELC, este archivo debe estar situado en la carpeta ELCmapas. De igual forma el nombre debe escribirse junto con la extensión, en este caso al ser una tabla, la extensión es txt. Así si el archivo se llama "Estadist\_ELC\_spain" se debe escribir "Estadist\_ELC\_spain.txt".

#### 6.3.1.16 Parámetro: *distdup*

Explicación: Determine el valor de distancia (en km) por debajo del cual Ud. considerará que dos sitios de presencia o de recolección en realidad representan una misma población. El valor cero (por defecto) excluye del análisis de representatividad, entradas con idénticas coordenadas.

#### 6.3.1.17 Parámetro: *resultados*

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

## 6.4. Resultados de Representa

En la ruta y carpeta que haya determinado en “resultados” (parámetro 6.3.1.17) encontrará hasta cinco mapas y hasta cinco tablas.

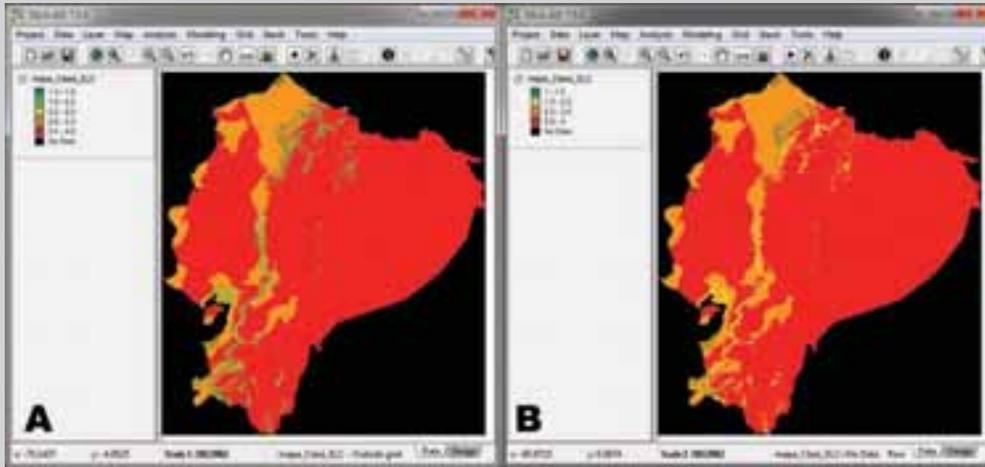
### 6.4.1 Los mapas

Se trata de dos mapas de puntos de tipo vectorial (shapefiles) y tres mapas ráster (formato .grid) que pueden visualizarse directamente en DIVA-GIS. Si no se introducen datos de fuentes externas, la lista de mapas se reduciría a tres (los descritos en los apartados 6.4.1.1, 6.4.1.2 y 6.4.1.4).

6.4.1.1 “mapa\_Class\_ELC.grd”. Corresponde a un mapa que agrupa en cuatro las categorías originales del mapa ELC. Estos grupos corresponden a la partición de las categorías según su frecuencia en la totalidad del territorio. La frecuencia se divide a partir de cuartiles. El grupo señalado como 1 corresponde a la frecuencia baja (por debajo del cuartil 0.25), el grupo 2 corresponde a la frecuencia media-baja (entre cuartiles 0.25 y 0.5 o mediana), el grupo 3 corresponde a la frecuencia media-alta (entre cuartiles 0.5 o mediana y 0.75) y grupo 4 que corresponde a frecuencia alta (por encima del cuartil 0.75). Cuando este mapa se abre en DIVA-GIS, por defecto se muestran cinco colores, por lo que es importante cambiar la visualización y que sólo se muestren cuatro colores, cada uno abarcando un rango que cubra cada valor de grupo (1 a 4). Este cambio se ilustra en la Fig. 23.

6.4.1.2 “mapa\_Class\_Sp.grd”. Corresponde a un mapa que agrupa en cuatro las categorías originales del mapa ELC. Estos grupos corresponden a la partición de las categorías ELC según la frecuencia en la ocurren las entradas. La frecuencia se divide a partir de cuartiles. El grupo señalado como 1 corresponde a la frecuencia baja (por debajo del cuartil 0.25), el grupo 2 corresponde a la frecuencia media-baja (entre cuartiles 0.25 y 0.5 o mediana), el grupo 3 corresponde a la frecuencia media-alta (entre cuartiles 0.5 o mediana y 0.75) y grupo 4 que corresponde a frecuencia alta (por encima del cuartil 0.75).

**FIGURA 23.** Configuración visual apropiada para mapas ráster de Representa. A) Visualización mapa\_Class\_ELC.grd tal y como se abre en DIVA-GIS, B) visualización ajustada a cuatro colores por cada grupo de frecuencia. Los escenarios adaptativos menos frecuentes (grupos baja y media baja frecuencia) aparecen en colores verde y amarillo.



6.4.1.3 "mapa\_Tipo\_faltante.grd". Este mapa es otra reclasificación de las categorías originales del mapa ELC. Este mapa sólo se produce cuando el usuario introduce datos de fuentes externas. Esta reclasificación corresponde a unos criterios expuestos en la tabla siguiente:

**TABLA 1.** Clasificación de categorías de mapa ELC de acuerdo a criterios de prioridad para futuras exploraciones

Clase	Diferencia Fuentes externas vs Banco de germoplasma (DIF) <sup>1</sup>	Clasificación frecuencia por ocurrencia de la especie <sup>2</sup>	Clasificación por frecuencia de categoría en mapa ELC <sup>3</sup>
0	No aplica	No aplica	No aplica
1	1	Baja o media baja	Baja o media baja
2	1	Baja o media baja	Media alta o alta
3	1	Media alta o alta	Baja o media baja
4	1	Media alta o alta	Media alta o alta
5	0.99-0.5	Baja o media baja	Baja o media baja
6	0.99-0.5	Baja o media baja	Media alta o alta
7	0.99-0.5	Media alta o alta	Baja o media baja
8	0.99-0.5	Media alta o alta	Media alta o alta

Clase	Diferencia Fuentes externas vs Banco de germoplasma (DIF) <sup>1</sup>	Clasificación frecuencia por ocurrencia de la especie <sup>2</sup>	Clasificación por frecuencia de categoría en mapa ELC <sup>3</sup>
9	0.01-0.499	Baja o media baja	Baja o media baja
10	0.01-0.499	Baja o media baja	Media alta o alta
11	0.01-0.499	Media alta o alta	Baja o media baja
12	0.01-0.499	Media alta o alta	Media alta o alta
13	0 y NA	No aplica	No aplica

<sup>1</sup> Este valor sale de la comparación entre ocurrencias de fuentes externas y bancos/colecciones de germoplasma en cada categoría usando la siguiente fórmula  $DIF = (FE/BG)/FE$  siendo FE el número de ocurrencias de fuentes externas y BG las correspondientes a banco de germoplasma.

<sup>2</sup> Esta clasificación es la misma que se muestra en el mapa 6.4.1.2

<sup>3</sup> Esta clasificación es la misma que se muestra en el mapa 6.4.1.1

Estas clases están relacionadas con la prioridad de visita o exploración que tendría cada categoría ecogeográfica en una futura recolección. La clase 1 agrupa las categorías que tendrían la máxima prioridad, mientras que la clase 2 tendría menos prioridad que la clase 1, y así consecutivamente hasta la clase 13.

El mapa, al abrirse en DIVA-GIS no muestra las 13 clases con un color individual para cada clase, sino que agrupa los 13 valores en cinco colores. Su visualización correcta se logra en DIVA-GIS cuando se añaden 8 colores más y se ajustan los rangos de valores de cada color (como en los mapas anteriores) al valor de una clase.

6.4.1.4 "Shapefile\_Puntos\_BG.shp". Mapa vectorial (shapefile) que representa los sitios de recolección de germoplasma del banco o la colección bajo evaluación de representatividad. La tabla que acompaña este mapa de puntos contiene todos los campos del formato de pasaporte FAO/Bioversity 2012.

6.4.1.5 "Shapefile\_FE\_class.shp". Mapa vectorial (shapefile) que representa las ocurrencias de fuentes externas. La tabla que acompaña este mapa de puntos presenta los siguientes campos adicionales al formato de datos de fuentes externas:

FE\_cat: Categoría del mapa ELC en la que ocurre la presencia

FE\_BG\_dif: Valor DIF (ver Tabla 1) para la categoría ELC en la que ocurre la presencia.

Class\_Sp: Indica el cuartil al que pertenece la categoría en la que ocurre la presencia de fuente externa, de acuerdo a la frecuencia de la especie.

Class\_ELC: Indica el cuartil al que pertenece la categoría en la que ocurre la presencia de fuente externa, de acuerdo a la frecuencia de la misma categoría en el mapa ELC.

Tipo\_falt: Indica a que clase pertenece la categoría en la que ocurre la presencia de fuente externa de acuerdo a la clasificación de la Tabla 1.

## 6.4.2 Las tablas

De igual manera que con los mapas, la lista de tablas puede reducirse de cinco a tres, dependiendo si el usuario introduce o no datos de fuentes externas.

6.4.2.1 “Tabla\_Fuentes\_Externas\_clasificadas.txt”. Corresponde a la misma tabla acompañante del shapefile del apartado 6.4.1.5, conteniendo las mismas variables.

6.4.2.2 “Tabla\_Resultados\_Representatividad.txt”. En esta tabla se presentan los resultados finales de evaluación de la representatividad, se hayan o no incluido datos de fuentes externas. Con esta tabla es posible crear en Excel las gráficas de barras como las de las Fig. 21 y 22. Finalmente en esta tabla se presenta toda la información requerida para el cálculo de los parámetros de la Tabla 1, incluyendo el valor de clase con el que se definen las prioridades.

6.4.2.3 “Tabla\_Resultados\_X2.txt”. Esta tabla muestra los resultados de la prueba de Chi cuadrado para determinar el grado de asociación entre dos distribuciones. Si se han introducido datos de fuentes externas, esta tabla contendrá los resultados de dos pruebas de Chi cuadrado: distribución banco/colección (o BG) vs. fuentes externas (FE) y banco/colección vs. distribución de frecuencias totales de las categorías en el mapa ELC.

6.4.2.4 “TablaClasificacionCuartilesEspecie.txt” y “TablaClasificacionCuartilesMapaELC.txt”. Estas dos tablas muestran los valores de los cuartiles 0.25, 0.5 (mediana) y 0.75 para la distribución de frecuencias de la especie y de las categorías del mapa ELC.

## 6.5. Referencias

Brown, A.H.D. 1989. The case for core collections. In : Brown, A.H.D., Frankel, O.H., Marshall, D.R. y Williams, J.T. (ed.) The use of plant genetic resources. Cambridge University Press, Cambridge, UK.

Crossa, J. y Vencovsky, R. 1994. Implications of the variance effective population size on the genetic conservation of monoecious species. *Theoretical and Applied Genetics* 89:936–942

Crossa, J. y Vencovsky, R. 1997. Variance effective population size for two-stage sampling of monoecious species. *Crop Science* 37:14–26

Crossa, J. y Vencovsky, R. 2011 Chapter 5: Basic sampling strategies: theory and practice. In: Guarino, L., Ramanatha Rao, V. y Goldberg, E. (ed.) *Collecting Plant Genetic Diversity: Technical Guidelines – 2011 Update*. Bioversity International. Available online (accessed 6 November 2013) [http://croptgenbank.sgrp.cgiar.org/index.php?option=com\\_content&view=article&id=671](http://croptgenbank.sgrp.cgiar.org/index.php?option=com_content&view=article&id=671)

Parra-Quijano, M.; Draper, D.; Torres, E. e Iriondo, J.M. 2008. Ecogeographical representativeness in crop wild relative *ex situ* collections. p. 249-273. In Maxted, N.; Ford-Lloyd, B.V.; Kell, S.P.; Iriondo, J.M.; Dulloo, M.E. y Turok, J. (ed.) *Crop wild*

relative conservation and use. CAB International, Wallingford.

Yonezawa, K.; Nomura, T. y Morishima, H. 1995. Sampling strategies for use in stratified germplasm collections. P. 35-53. *In* : Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L. y Morales, E.A.V. (ed.) Core collections of plant genetic resources. John Willey & sons, Chichester, UK.



## 7. Herramienta DIV mapas

### 7.1. Representación espacial de la diversidad local

En 2012 fue publicado un estudio sobre la presentación de patrones espaciales de diversidad genética a partir de marcadores neutrales tipo microsatélite para el caso de *Annona cherimola* (van Zonneveld et al., 2012). En este estudio se propone una forma diferente de mostrar cómo está distribuida la diversidad genotípica, basado en la estimación de parámetros propios de genética de poblaciones, pero que antes de ser aplicados a todas las muestras a la vez, en este caso la diversidad se estima a nivel local con la determinación de vecindades o áreas de influencia. El resultado de juntar los resultados de cada vecindad fue un mapa que muestra claramente donde se localizan los “puntos calientes” de diversidad. Las aplicaciones de esta metodología en la conservación *ex situ* e *in situ* de los recursos fitogenéticos (variedades locales y especies silvestres emparentadas con cultivadas) fueron evidentes.

Esta no ha sido la primera aproximación SIG o de geoestadística en el análisis de la diversidad genética, pues previamente se han realizado interpolaciones de datos genéticos (Hoffman et al., 2003). Sin embargo, la metodología de van Zonneveld y colaboradores representa una aplicación muy práctica y más sencilla en su análisis e interpretación.

Posteriormente Thomas y colaboradores (2012) aplicaron esa misma metodología para 993 individuos caracterizados por microsatélites de cacao (*Theobroma cacao*), junto con otros análisis, para identificar procesos evolutivos en esta planta cultivada.

A partir de la publicación de estos desarrollos fue posible entender paso a paso cómo opera el proceso de obtención de un mapa de este tipo. La metodología era evidentemente repetible y sólo varía el parámetro genético que se calcula entre las muestras que componen una vecindad. Así, si el parámetro expresa que tan diferente genéticamente son las muestras de una vecindad, el mapa podría denominarse como “mapa de diversidad”. A partir de esta base, fue posible desarrollar la herramienta “DIV mapas” y generalizar su aplicación más allá de datos de caracterización genotípica.

Es muy importante destacar que estos mapas muestran diversidad genotípica a nivel intraespecífico, un aspecto que los diferencia notablemente de los mapas de riqueza de especies o mapas de diversidad filogenética, los cuales trabajan a nivel interespecífico.

Mostrar la diversidad en forma de mapas tiene múltiples ventajas respecto a las formas en las que usualmente se presentan estos resultados. Los mapas de diversidad, basados en el desarrollo original de van Zonneveld y colaboradores (2012) permiten identificar de manera sencilla y rápida, las zonas o regiones donde se concentra una

alta variabilidad. Un mapa de estas características se convierte en una poderosa herramienta para la toma de decisiones en materia de conservación *ex situ* e *in situ*.

### 7.1.1 ¿Por qué un mapa de la diversidad ecogeográfica?

La diversidad ecogeográfica de un conjunto de accesiones es una medida de las diferencias que ocurren entre los escenarios adaptativos de donde provienen dichas accesiones, o en otras palabras, los sitios de recolección. Se recurre al término "escenario adaptativo" antes que "ambiente", porque para el cálculo de la diversidad ecogeográfica sólo se considerarían las características ambientales de tipo abiótico de mayor influencia en la distribución y ocurrencia de la especie objetivo, no todas las características ambientales disponibles.

La diversidad ecogeográfica, como otro tipo de diversidad, se determina a partir de datos de caracterización del germoplasma. La caracterización ecogeográfica se realiza mediante la extracción de información para cada coordenada mediante la utilización de un SIG, el cual ha sido previamente cargado con capas de información ambiental.

La visualización de la diversidad ecogeográfica en forma de mapa similar a los que ha desarrollado van Zonneveld y colaboradores (2012), facilita la comparación entre zonas o regiones en base a la diferencia de los escenarios adaptativos donde ocurren las entradas. Las zonas o regiones donde ocurren mayores diferencias se pueden traducir directamente en zonas donde cabe esperar germoplasma con adaptaciones más divergentes e indirectamente una posible ocurrencia de una mayor diversidad genotípica o fenotípica. Evidentemente la determinación de zonas con mayor diversidad genotípica o fenotípica es óptima cuando se realiza a través de datos de caracterización genotípica y fenotípica respectivamente. Sin embargo, a falta de éstos, un mapa de diversidad ecogeográfica puede servir de solución intermedia mientras se consiguen caracterizar genotípica y/o fenotípicamente las entradas. En cualquier caso, el escenario ideal para el análisis de la diversidad bajo esta nueva metodología, es cuando se pueden obtener mapas para los tres tipos de caracterización, dado que su contraste ofrece una visión biológica muy completa de la situación de los recursos fitogenéticos que ocurren dentro de un marco de trabajo.

## 7.2. Procedimiento de obtención de mapas de diversidad de DIV mapas

DIV mapas es una aplicación inspirada en la aplicación de van Zonneveld y colaboradores (2012) para chirimoya (*Annona cherimola*), pero incluye algunas diferencias respecto a la metodología original las cuales son más evidentes al contrastar los dos procesos. En este apartado se mostrará paso a paso como DIV mapas logra obtener mapas de diversidad.

DIV mapas determina medidas de diversidad local, es decir, entre las accesiones recolectadas en una zona con forma de cuadrícula con determinado tamaño y sus

vecindades (zona de influencia), usando como insumo información ecogeográfica, fenotípica o genotípica. Nótese que a partir de este punto nos referiremos a entradas y no a muestras, dado que la herramienta está pensada para ser utilizada en el entorno de recursos fitogenéticos, lo cual no implica que no pueda usarse en otros ámbitos biológicos. Como resultado, DIV mapas ofrece una ilustración gráfica que refleja los valores de las medidas de diversidad distribuidas en un espacio a manera de mapa, el cual facilita la visualización de zonas calientes de diversidad genética.

Para empezar es importante anotar que DIV mapas, así como otras herramientas incluidas en el presente manual y muchas otras aplicaciones de SIG y ecogeografía en recursos fitogenéticos, requiere que cada entrada este georreferenciada y que además lo esté de manera adecuada. El apartado 3 de este manual hace referencia a la herramienta GEOQUAL que permite conocer la calidad de la georreferenciación del sitio de recolección del germoplasma. Es aconsejable usar esta herramienta antes de usar DIV mapas, de manera que sólo entradas con suficiente calidad en su georreferenciación serán tenidas en cuenta en la obtención de mapas de diversidad. En cualquier caso, entradas sin coordenadas (campos DECLATITUDE ó LATITUDE y DECLONGITUDE ó LONGITUDE) no podrán ser incluidas en el análisis que realiza DIV mapas.

El segundo punto importante es que si se necesita obtener mapas de diversidad fenotípicos o genotípicos, los datos de caracterización de cada tipo deben ser arreglados según el formato que se suele proveer en la carpeta "Formatos" (archivos .xls de Excel). Si el tipo de mapa de diversidad requerido es de tipo ecogeográfico, DIV mapas incluye el mismo proceso de caracterización ecogeográfica de germoplasma que la herramienta ECOGEO (capítulo 5), por lo tanto para este aspecto no es necesario preparar tablas o matrices de datos de caracterización sino simplemente señalar las variables ecogeográficas por las cuales se quiere caracterizar las entradas.

DIV mapas aprovechará toda la información de caracterización disponible y válida, y en base a ella realizará los mapas de diversidad para cada aspecto individualmente. En ese sentido, las lista de entradas caracterizadas genotípica, fenotípica o ecogeográficamente pueden coincidir (lo cual favorece la interpretación de resultados) o no. Lo que es imprescindible es que los códigos de identificación de las entradas en las tablas de caracterización genotípica o fenotípica, deben estar incluidos en la tabla de pasaporte FAO/Bioversity 2012 que contiene la información de georreferenciación de los sitios de recolección.

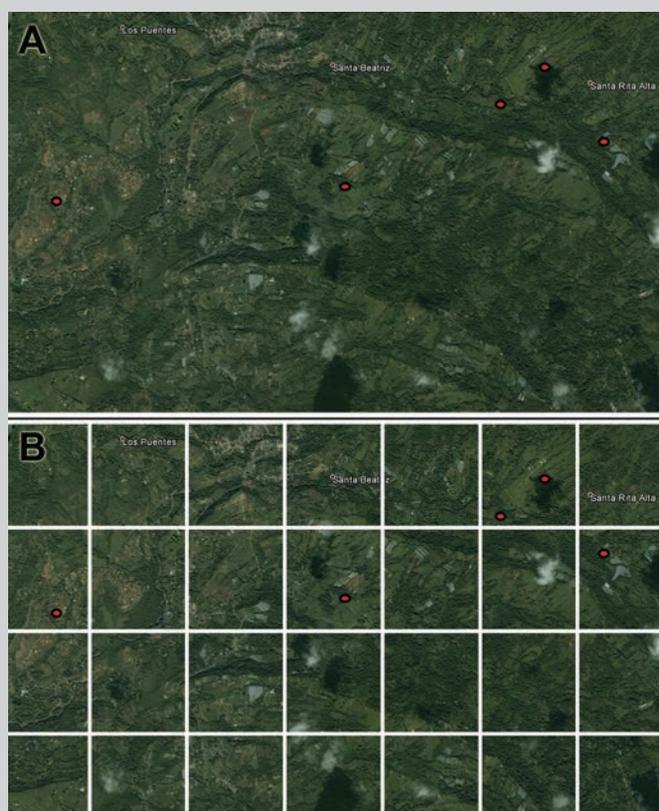
Una vez claras estas condiciones, los siguientes puntos muestran cómo DIV mapas obtiene los mapas de diversidad, independientemente de los datos de caracterización que use para ello.

### 7.2.1. Distribución de sitios de recolección y generación de retícula o malla

A partir de las coordenadas de cada sitio de recolección se genera un espacio de trabajo (x-min, y-min, x-max e y-max siendo x latitud e y longitud), sobre el cual se sobrepone una retícula o conjunto de celdas de forma cuadrada de un tamaño

definido por el usuario (ver Fig. 24). Adicionalmente se carga una capa con los centroides de cada celda de la retícula (ver Fig. 25 parte A). Cada centroide tiene un código de identificación.

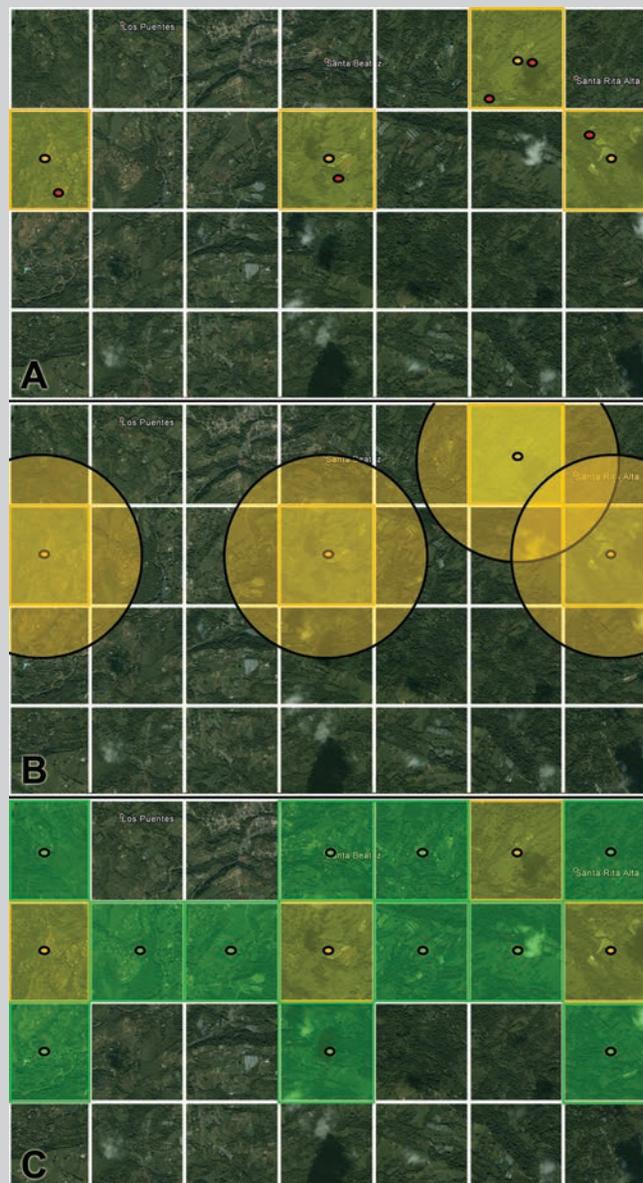
**FIGURA 24. Primer paso. A) distribución espacial de los sitios de recolección, B) superposición de retícula de dimensiones de celda (resolución) seleccionadas por el usuario.**



### 7.2.2. Selección de celdas con entradas y celdas vecindario

Del total de celdas de la retícula se seleccionan aquellas donde ocurren entradas. Adicionalmente el usuario determina un área de influencia, que señala como el radio de un área circular, la cual está relacionada con la biología reproductiva de la especie, el flujo genético y el manejo y la dispersión que puede generar el hombre, especialmente si se trata de una forma cultivada. Esta área de influencia se utilizará para determinar las celdas vecindario, que son las celdas donde no ocurren entradas pero que se encuentran suficientemente cerca a las inicialmente seleccionadas (celdas con entradas). Para ser celda vecindario, el centroide de ésta debe caer dentro de la proyección del área circular de influencia que se traza a partir del centroide de cada celda con entradas. El proceso de selección de celdas con entradas y de celdas vecindario se muestra en la Fig 25.

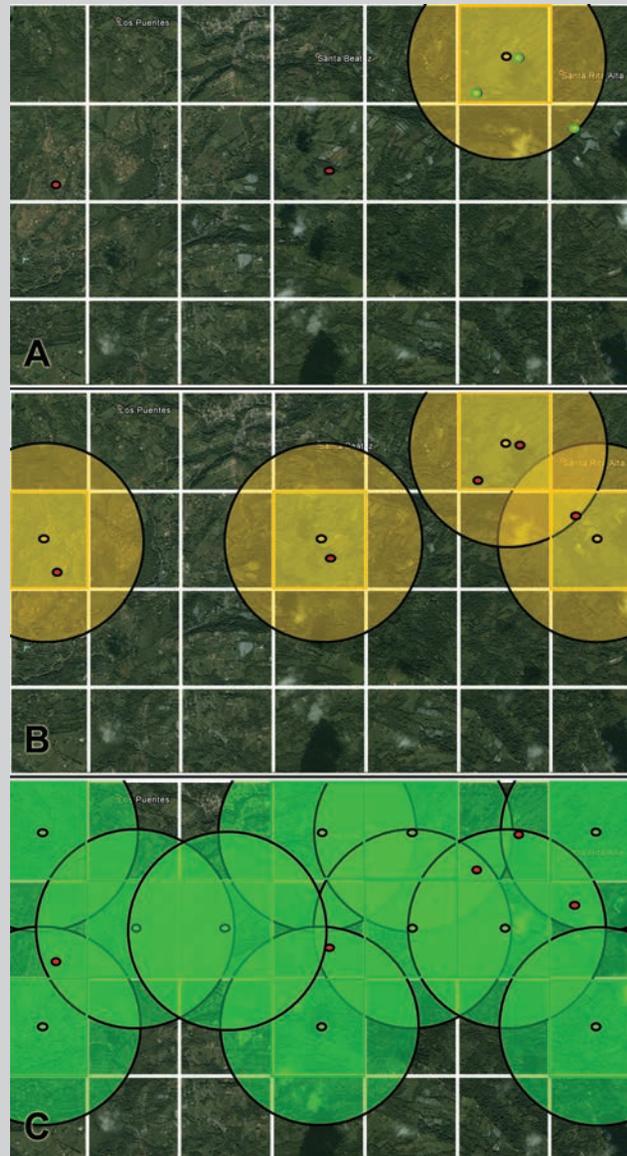
**FIGURA 25.** Segundo paso. A) Determinación de celdas con entradas y sus centroides, B) proyección de las áreas de influencia desde los centroides de celdas con entradas, C) determinación de celdas vecindario.



### 7.2.3. Determinación de entradas asociadas a celda con entradas y celdas vecindario

A partir de los centroides de las celdas con entradas y celdas vecindario se proyectan de nuevo las áreas circulares de influencia y a la lista resultante de entradas que caen en cada área se le asigna el código de identificación de su respectivo centroide (ver Fig. 26).

**FIGURA 26.** Tercer paso. A) Determinación de las entradas que ocurren dentro del área de influencia de una celda en singular, B) determinación de entradas que ocurren dentro de áreas de influencia de celdas con entradas, C) determinación de entradas que ocurren dentro de áreas de influencia de celdas vecindario.

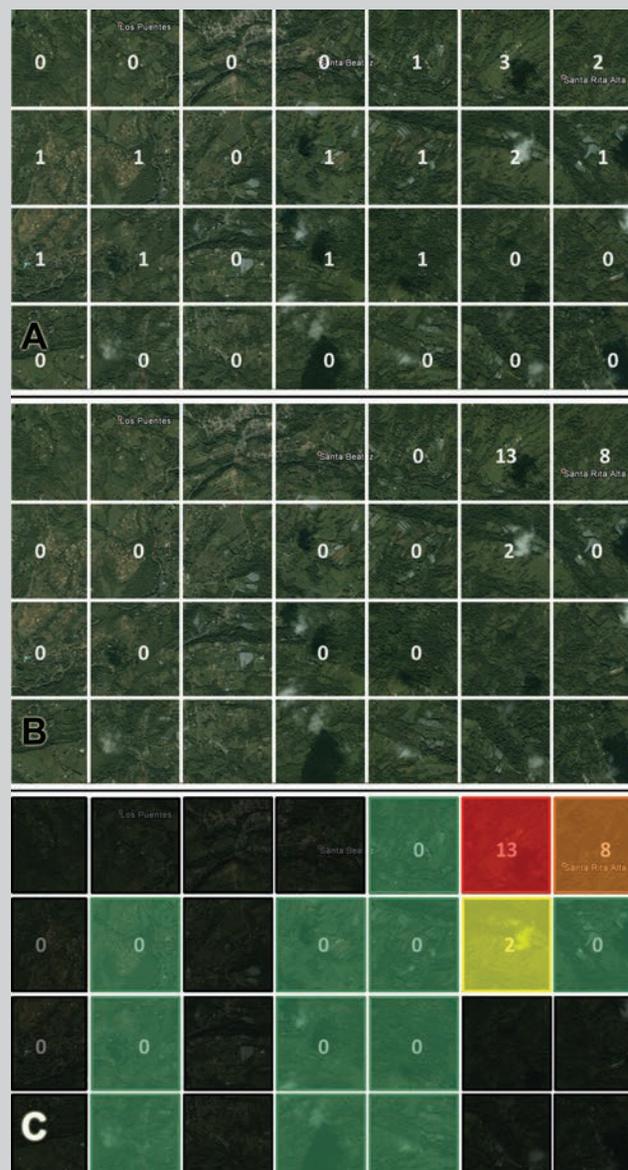


#### 7.2.4. Obtención de los mapas finales de diversidad

La lista de entradas por celda servirá para obtener matrices iniciales, con los datos de caracterización fenotípica, genotípica o ecogeográfica (según los datos que haya ingresado el usuario) como columnas y las entradas de cada centroide identificadas por su valor de ACCENUMB como filas. De esta forma se obtendrán tantas matrices iniciales como celdas con entradas y celdas vecindario han sido determinadas. A

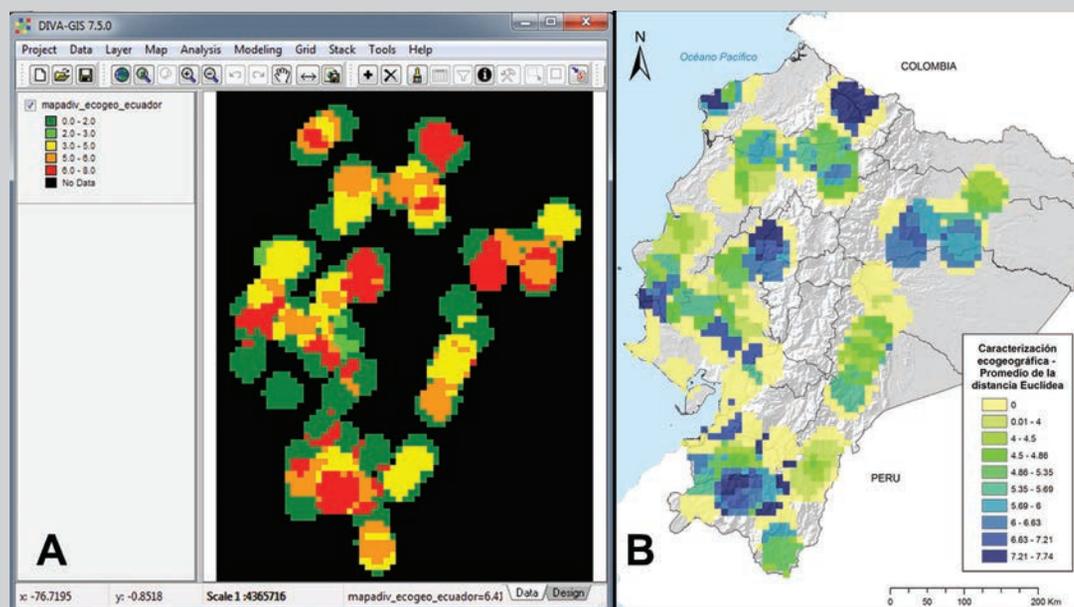
cada matriz inicial se aplica estandarizaciones de datos si se tratase de variables cuantitativas y posteriormente un coeficiente de distancia o similaridad/disimilaridad, el cual producirá a la vez, una matriz diagonal de distancia. A partir de esta matriz se calcula la distancia promedio de las entradas involucradas y ese valor promedio de distancia se adjudicará a cada código de centroide y a su respectiva celda. Esto permite a R producir mapas de celdas ráster (que reflejan los valores asignados (ver Fig. 27).

**FIGURA 27.** Cuarto paso. A) Número de entradas analizadas por celda, B) valores asignados a celdas de distancia genotípica, fenotípica o ecogeográfica promedio, C) Asignación de colores de una gradación acorde con los valores promedio de distancia.



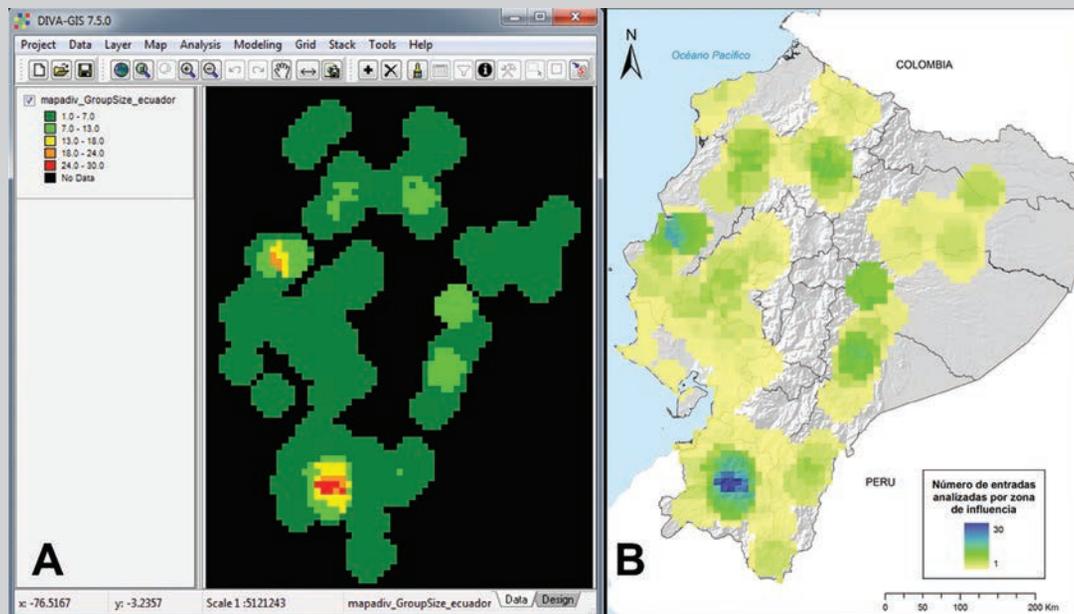
Para el caso de caracterizaciones genóticas, adicionalmente a la distancia o disimilaridad promedio, se puede calcular otros parámetros genéticos como el de la diversidad genética de Nei (1987) o la proporción de marcadores polimórficos para cada grupo de entradas dentro de cada área de influencia. Estos parámetros los calcula R desde las matrices iniciales de caracterización.

**FIGURA 28.** Visualización de mapas de diversidad. A) Mapa de diversidad ecogeográfica para la colección ecuatoriana de maní (*Arachis hypogaea*) producto de DIV mapas abierto en DIVA-GIS, B) mismo mapa en su versión final.



Finalmente, al visualizarse en DIVA-GIS el archivo ráster compuesto por las celdas a las que se le han asignado valores de los parámetros de diversidad, lo que hace el programa informático es asignar un color a cada celda de una gradación de colores, permitiendo observar rápidamente cuales son los lugares donde se presentan mayores niveles de diversidad, medida por valores promedio de distancia/disimilaridad u otros parámetros genéticos (ver Fig. 28). Si se han introducido datos de caracterización de diferente tipo, se producirán varios mapas, ocurriendo siempre la siguiente relación: uno para el aspecto ecogeográfico, uno para el aspecto fenotípico y uno o varios para el aspecto genotípico, dependiendo de si el usuario ha solicitado el cálculo de uno o varios parámetros. También se produce un mapa del número de entradas analizadas por celda como se muestra en la Fig. 29, que corresponde a lo que representa la Fig. 27 parte A. Este último mapa puede servir de apoyo para determinar la existencia de posibles sesgos en la recolección o en la interpretación de los patrones que se puedan encontrar en los mapas de diversidad.

**FIGURA 29.** Visualización de mapa de número de entradas analizadas por celda. A) Mapa abierto en DIVA-GIS, B) mismo mapa en su versión final.



### 7.2.5. Uso de remuestreo para eliminación de posibles sesgos de recolección

van Zonneveld y colaboradores (2012) sugieren usar un método llamado de rarefacción, el cual eliminaría el efecto de un posible sesgo en la recolección de las muestras analizadas. Thomas y colaboradores (2013) sugieren un método de remuestreo. Este último método está incorporado en DIV mapas. En él se usa como N (tamaño de la muestra) la mediana del número de entradas analizadas por celda. Celdas con un número de entradas inferior a la mediana son descartadas. Entonces el valor asignado a cada celda corresponde al promedio de los valores promedio de distancia obtenidos en cada proceso de remuestreo. Dependiendo del número de remuestreos que el usuario elija y del número de celdas que componga el mapa de diversidad, el proceso de obtención puede tomar desde minutos hasta horas.

### 7.2.6. Otros análisis

DIV mapas también permite realizar otros tipos de análisis, particularmente cuando se han introducido datos de caracterización de diferentes tipos. De esta manera, DIV mapas pregunta al usuario si desea realizar análisis de agrupamientos, análisis de ordenación de la misma manera que se realiza en la herramienta ECOGEO. Adicionalmente el usuario puede solicitar realizar comparaciones matriciales de Mantel (1967) entre matrices de distancia para el total de las entradas. De manera

automática, DIV mapas genera una matriz de distancias geográficas entre todos los sitios de recolección e introduce esta matriz en las correlaciones matriciales pareadas.

### 7.3. Formatos de datos de entrada en DIV mapas

Para el correcto funcionamiento de la herramienta DIV mapas, es necesario introducir información de diverso tipo en los formatos indicados. Usualmente estos formatos se encuentran en la carpeta "Formatos" dentro de la estructura de carpetas y archivos de las herramientas CAPFITOGEN. Dentro de esta carpeta encontraremos otra con el nombre "Formatos DIVmapas" y dentro de ella cinco archivos Excel.

#### 7.3.1. Modelo de datos de pasaporte

Como en otras herramientas CAPFITOGEN, los datos de pasaporte deben ingresar en formato FAO/Bioversity 2012 con modificaciones menores (archivo "ModeloDatosPasaporte\_FAO\_BIOVERSITY\_2012.xls"). Como DIV mapas permite trabajar con datos evaluados por su calidad de georreferenciación mediante la herramienta GEOQUAL, aparte del modelo de datos de pasaporte normal, se dispone de un modelo que incluye campos adicionales para los valores de evaluación GEOQUAL (archivo "ModeloDatosPasaporte\_FAO\_BIOVERSITY\_2012\_conDatosGEOQUAL.xlsx"). Sin embargo, la manera más fácil de usar datos de pasaporte evaluados por GEOQUAL es introducir directamente la tabla de pasaporte con datos de evaluación que produce GEOQUAL como resultado y que lleva el nombre de "PasaporteOriginalEvaluadoGEOQUAL.txt". Recuerde que esta tabla debe estar en formato de texto delimitado por tabulaciones, por lo que debe ser exportada desde Excel en este formato, y debe ser ubicada en la carpeta "Pasaporte" de la estructura de carpetas y archivos de las herramientas CAPFITOGEN.

#### 7.3.2. Modelo de datos fenotípicos

Al visualizar el contenido del formato de datos fenotípicos (archivo "ModeloDatosFenotipicos.xlsx") aparecerá una columna en verde (de relleno obligatorio) denominada "ACCENUMB" que corresponde al mismo código de identificación ACCENUMB de la tabla de datos de pasaporte. El orden en que se encuentren los códigos es indiferente. Dado que no siempre se cuenta con datos de caracterización fenotípica para todas las entradas presentes en los datos de pasaporte, el número de entradas presente en la tabla de datos fenotípicos puede ser menor al de datos de pasaporte. La situación que no debe ocurrir es que en la tabla de datos fenotípicos aparezcan entradas o códigos ACCENUMB que no aparezcan en la tabla de datos de pasaporte. Esto último generaría un error de procesamiento.

El resto de las columnas en el formato aparecen con los nombres "D1", "D2" y "D3". Estos nombres representan los nombres de los descriptores fenotípicos 1, 2 y 3. El formato sólo incluye tres columnas para los descriptores, pero teóricamente pueden ser tantos descriptores como el usuario disponga, alargando la secuencia

desde "D4" hasta donde sea necesario. Los nombres pueden ser cambiados (por ejemplo "D1" por "NGRANOS"), para facilidad del usuario. Si desea cambiar los nombres, tenga en cuenta dos recomendaciones. La primera es que no existan espacios dentro del nombre, que en lo posible el nombre este compuesto por menos de once caracteres y que especialmente no se repita ningún nombre. Esta última condición puede generar error.

La codificación de las variables fenotípicas tiene ciertas condiciones. Las variables, ya sean cuantitativas o categóricas deben estar expresadas numéricamente. Para las variables categóricas, los estados que durante la caracterización hayan sido descritos mediante códigos que incluyen caracteres alfabéticos o de otros tipos, deben ser transformados a códigos exclusivamente numéricos, sin líneas, puntos, comas o espacios. Si existen datos faltantes, éstos deben codificarse con los caracteres "NA".

Finalmente tenga en cuenta que la herramienta DIV mapas sólo reconoce la información en tablas cuando esta se encuentra en formato de texto separado por tabulaciones. En consecuencia, una vez la tabla de datos fenotípicos en Excel esté completa y acorde a los requerimientos previamente señalados, se debe exportar en formato de texto separado por tabulaciones y este archivo de texto debe ubicarse en la carpeta "Pasaporte" junto con el resto de tablas de datos de caracterización y la tabla de datos de pasaporte.

### 7.3.3. Modelo de tabla de naturaleza de las variables fenotípicas

En el caso de que disponga de datos de caracterización fenotípica y que desee utilizarla para generar un mapa de diversidad con DIV mapas, además de aportar la tabla de datos fenotípicos del apartado 7.3.2, es imprescindible rellenar la tabla que aparece con el nombre de "ModeloTablaNaturalezaVariables.xlsx", la cual indicará la naturaleza de cada variable o descriptor fenotípico incluido en la tabla de datos fenotípicos. Este archivo Excel contiene dos hojas de cálculo.

En la primera ("Natvariables") se encuentra la tabla de naturaleza de variables fenotípicas, que sólo contiene tres columnas. La primera es "ID" y en ella simplemente se asigna un número consecutivo a cada variable (1, 2, 3,...), de esta manera cada fila en esta tabla corresponde a una variable o descriptor fenotípico de la tabla de datos fenotípicos. La segunda es "NOMVAR" y corresponde exactamente a los nombres asignados a las variables o descriptores en la tabla de datos fenotípicos. La tercera y última columna es "NATVAR", en la que se debe indicar la naturaleza de la variable o descriptor correspondiente. Al situar el cursor sobre esta celda, se desplegará la lista de posibles valores para esta columna, a saber: binario simétrico, binario asimétrico, nominal, ordinal, cuantitativo.

Finalmente en la hoja de cálculo "Observaciones" se encuentran algunas indicaciones y ayudas para rellenar la tabla de la hoja de cálculo "Natvariables".

Al finalizar, la tabla de naturaleza de variables debe ser exportada en formato de texto delimitado por tabulaciones y debe situarse en la carpeta "Pasaporte", de igual manera que el resto de tablas de ingreso de información.

### 7.3.4. Modelo de datos genotípicos

Como se ha señalado previamente, DIV mapas ofrece la posibilidad de hacer mapas de diversidad a partir de caracterizaciones genotípicas de germoplasma, analizando la información proveniente de marcadores moleculares, como si estos fueran de tipo dominante. Esto implica que la tabla de datos genotípicos (archivo Excel "ModeloDatosGenotipicos0\_1.xlsx") contendrá variables de ausencia/presencia codificados como 0 y 1 respectivamente. Así, la estructura de esta tabla es muy similar a la de datos fenotípicos y se debe rellenar de manera similar, excepto porque todas las variables o descriptores en la tabla de datos genotípicos corresponden a variables binarias asimétricas y se codificarán con valores 0 y 1.

Al igual que con la información fenotípica, tenga en cuenta que la herramienta DIV mapas sólo reconoce la información en tablas cuando ésta se encuentra en formato de texto separado por tabulaciones. En consecuencia, una vez la tabla de datos genotípicos en Excel esté completa y acorde a los requerimientos señalados, se debe exportar en formato de texto separado por tabulaciones y este archivo de texto debe ubicarse en la carpeta "Pasaporte" junto con el resto de tablas de datos de caracterización y la tabla de datos de pasaporte.

## 7.4. Utilización de la herramienta DIV mapas

Una vez instalada las herramientas CAPFITOGEN y seleccionada la herramienta DIV mapas, será necesario definir una serie de parámetros para que la programación R funcione correctamente.

### 7.4.1 Parámetros iniciales definidos por el usuario

#### 7.4.1.1 Parámetro: ruta

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN.  
Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

#### 7.4.1.2 Parámetro: país

Explicación: Seleccione el país en el cual fueron recolectadas la mayoría o la totalidad de las entradas que desea analizar. Si las entradas han sido recolectadas en más de un país, considere seleccionar una región, subcontinente o continente (progresivamente se añadirán estas opciones).

#### 7.4.1.3 Parámetro: bootstrap

Explicación: Marque esta opción si desea calcular los valores para los mapas mediante bootstrapping (remuestreo). Tenga en cuenta que elegir esta opción reducirá áreas de análisis en los mapas allí donde la densidad de sitios de recolección es menor.

#### 7.4.1.4 Parámetro: *bootn*

Explicación: Especifique el número de remuestreos si ha elegido la opción de bootstrapping.

#### 7.4.1.5 Parámetro: *replac*

Explicación: Marque esta opción si desea que los remuestreos se realicen con reemplazamiento.

#### 7.4.1.6 Parámetro: *pasaporte*

Explicación: Escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama 'tabla' debe escribir: 'tabla.txt'. Recuerde que este archivo debe ser guardado previamente en la carpeta 'Pasaporte' que hace parte del conjunto de carpetas que componen la herramienta DIVmapas.

#### 7.4.1.7 Parámetro: *geoqual*

Explicación: Seleccione esta opción si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 50 columnas (y no las 45 del modelo de pasaporte que usan las herramientas CAPFITOGEN). Use por tanto la tabla resultado de GEOQUAL v.1.2 denominada 'PasaporteOriginalEvaluadoGEOQUAL.txt' como tabla de pasaporte en el punto anterior.

#### 7.4.1.8 Parámetro: *totalqual*

Explicación: Si su tabla de pasaporte proviene de GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

#### 7.4.1.9 Parámetro: *buffy*

Explicación: Marque esta opción si desea que la extracción de información ecogeográfica se haga sobre un área alrededor del sitio de recolección. Dejar sin marcar esta opción lleva a que la extracción se realice sólo para el punto que indican las coordenadas del sitio de recolección.

#### 7.4.1.10 Parámetro: *tamp*

Explicación: Especifique el radio (en metros) de un área circular alrededor del punto que indican las coordenadas del sitio de recolección, de la que se extraerá información ecogeográfica. Los valores extraídos del área circular serán promediados para obtener un único valor y no se tendrán en cuenta celdas sin valor. Se recomienda que este valor no sea inferior a la distancia de cada lado de celda del parámetro resol1.

#### 7.4.1.11 Parámetro: ecogeo

Explicación: Seleccione esta opción si desea obtener un mapa de diversidad ecogeográfica.

#### 7.4.1.12 Parámetro: resol1

Explicación: Seleccione el nivel de resolución que desea utilizar para la extracción de información ecogeográfica. Note que 1x1 km ofrece mayor resolución pero exige más capacidad de cómputo en relación a 5x5 km, si bien no es un aspecto tan limitante como en la herramienta ELC mapas. Las resoluciones 10x10 y 20x20 km están restringidas para países de gran extensión, subcontinentes o continentes. Ver el Anexo 12.5 sobre disponibilidad de resoluciones en relación al país o región seleccionada.

#### 7.4.1.13 Parámetro: bioclimsn

Explicación: Seleccione esta opción si desea usar variables de tipo bioclimático (temperatura, precipitación e índices asociados) para la generación de mapas de diversidad ecogeográfica.

#### 7.4.1.14 Parámetro: bioclimv

Explicación: Seleccione las variables bioclimáticas que desea incluir para generar el mapa de diversidad ecogeográfica. El total de variables seleccionables se detalla en el Anexo 12.1.

#### 7.4.1.15 Parámetro: edaphsn

Explicación: Seleccione esta opción si desea usar variables de tipo edáfico (texturas, profundidad, pH, etc.) para la generación de mapas de diversidad ecogeográfica.

#### 7.4.1.16 Parámetro: edaphv

Explicación: Seleccione las variables edáficas que desea incluir para generar el mapa de diversidad ecogeográfica. El total de variables seleccionables se detalla en el Anexo 12.2.

#### 7.4.1.17 Parámetro: geophysn

Explicación: Seleccione esta opción si desea usar variables de tipo geofísico (relativas al relieve y radiación solar) para la generación de mapas de diversidad ecogeográfica.

#### 7.4.1.18 Parámetro: geophysv

Explicación: Seleccione las variables geofísicas que desea incluir para generar el mapa de diversidad ecogeográfica. El total de variables seleccionables se detalla en el Anexo 12.3.

#### 7.4.1.19 *Parámetro: latitud*

Explicación: Incluye la latitud en su mapa de diversidad ecogeográfica? Nota: incluyendo latitud y longitud el aspecto de cercanía geográfica tendrá algún reflejo en el mapa final aunque solo a nivel local (dentro de cada zona de influencia).

#### 7.4.1.20 *Parámetro: longitud*

Explicación: Incluye la longitud en su mapa de diversidad ecogeográfica? Nota: incluyendo latitud y longitud el aspecto de cercanía geográfica tendrá algún reflejo en el mapa final aunque solo a nivel local (dentro de cada zona de influencia).

#### 7.4.1.21 *Parámetro: phenotip*

Explicación: Seleccione esta opción si desea obtener un mapa de diversidad fenotípica. Para ello deberá disponer de datos de caracterización o evaluación fenotípica (morfología, fenología, productividad, resistencia, etc.) siguiendo el formato indicado. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama 'fenotipos', en este espacio debe escribir 'fenotipos.txt'. Recuerde que esta tabla debe encontrarse en la carpeta Pasaporte, de la estructura de datos de las herramientas CAPFITOGEN.

#### 7.4.1.22 *Parámetro: phenot*

Explicación: Indique el nombre del archivo de texto que contiene los datos provenientes de la caracterización fenotípica en el formato indicado. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama 'fenotipos', en este espacio debe escribir 'fenotipos.txt'.

#### 7.4.1.23 *Parámetro: phenotv*

Explicación: Indique el nombre del archivo de texto que contiene la tabla que describe la naturaleza de cada variable fenotípica en el formato indicado. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama 'variablesfenotipo', en este espacio debe escribir 'variablesfenotipo.txt'. Esta tabla debe describir todas las variables incluidas en la tabla con los datos de la caracterización (punto anterior).

#### 7.4.1.24 *Parámetro: genotip*

Explicación: Seleccione esta opción si desea obtener un mapa de diversidad genotípica. Para ello deberá disponer de datos de caracterización o evaluación genotípica (como presencia o ausencia de marcador como ceros y uno) siguiendo el formato indicado. Recuerde que esta tabla debe encontrarse en la carpeta Pasaporte, de la estructura de datos de las herramientas CAPFITOGEN.

#### 7.4.1.25 Parámetro: *genot*

Explicación: Indique el nombre del archivo de texto que contiene los datos provenientes de la caracterización genotípica en el formato indicado. No olvide incluir el nombre de la extensión, así por ejemplo si la tabla se llama 'genotipos', en este espacio debe escribir 'genotipos.txt'.

#### 7.4.1.26 Parámetro: *neigd*

Explicación: Seleccione esta opción si desea obtener un mapa del índice promedio de diversidad genética de Nei (1987), un mapa de la proporción promedio de marcadores polimórficos y un mapa de número de entradas analizadas por celda.

#### 7.4.1.27 Parámetro: *csimilar*

Explicación: Indique el número correspondiente del coeficiente de similaridad que desea utilizar para generar el mapa de distancia genotípica promedio. 1 = Índice de Jaccard (1901), 2 = SMC de Sokal&Michaner 1958, 3 = Sokal y Sneath (1963) (S5 de Gower y Legendre), 4 = Rogers&Tanimoto (1960), 5 = Dice (1945), 6 = coeficiente de Hamann, 7 = Ochiai (1957), 8 = Sokal y Sneath (1963) (S13 de Gower y Legendre), 9 = Phi de Pearson, 10 = S2 de Gower y Legendre. La distancia (d) se obtiene como  $d=\sqrt{1-s}$  siendo s el coeficiente de similaridad.

#### 7.4.1.28 Parámetro: *rgrid*

Explicación: Elija el tamaño de celda (en km) que tendrá el/los mapas de diversidad que va a generar. Este parámetro está restringido a los siguientes valores: 1, 5, 10, 50 y 100 km (si escoge otro valor se producirá un error).

#### 7.4.1.29 Parámetro: *buffer*

Explicación: Elija el radio del área circular de influencia o vecindad (en km). Esta área se creará a partir de cada centroide de celda del mapa con sitios de recolección y genera grupos con las entradas cuyo sitio de recolección queda incluido. El valor de índices y distancias promedio de cada grupo será asignado a la celda desde cuyo centroide se trazó el área de influencia.

#### 7.4.1.30 Parámetro: *ecogeoclus*

Explicación: Seleccione esta opción si desea realizar un análisis de agrupamientos de todas las entradas para las cuales se ha extraído información ecogeográfica.

#### 7.4.1.31 Parámetro: *ecogeoclustype*

Explicación: Elija el tipo de agrupamiento jerárquico que desea utilizar para los agrupamientos ecogeográficos: 'single' = vecino más cercano, 'complete' = vecindario más compacto, 'ward' = método de mínima varianza de Ward, 'mcquitty' = método

de McQuitty, 'average' = similaridad promedio (UPGMA), 'median' = similaridad de la mediana, 'centroid' = centroide geométrico, 'flexible' = Beta flexible.

#### 7.4.1.32 Parámetro: *ecogeopca*

Explicación: Seleccione esta opción si desea realizar un análisis de componentes principales de todas las entradas para las cuales se ha extraído información ecogeográfica.

#### 7.4.1.33 Parámetro: *ecogeopcaxe*

Explicación: Número de componentes a retener (siempre menor al número de variables ecogeográficas) dentro del análisis PCA.

#### 7.4.1.34 Parámetro: *phenoclus*

Explicación: Seleccione esta opción si desea realizar un análisis de agrupamientos de todas las entradas para las cuales se aporta información fenotípica.

#### 7.4.1.35 Parámetro: *phenoclustype*

Explicación: Elija el tipo de agrupamiento jerárquico que desea utilizar para los agrupamientos fenotípicos: 'single' = vecino más cercano, 'complete' = vecindario más compacto, 'ward' = método de mínima varianza de Ward, 'mcquitty' = método de McQuitty, 'average' = similaridad promedio (UPGMA), 'median' = similaridad de la mediana, 'centroid' = centroide geométrico, 'flexible' = Beta flexible.

#### 7.4.1.36 Parámetro: *phenopca*

Explicación: Seleccione esta opción si desea realizar un análisis de componentes/ coordenadas principales de todas las entradas para las cuales se ha aportado información fenotípica.

#### 7.4.1.37 Parámetro: *phenopcaxe*

Explicación: Numero de componentes/coordenadas a retener (siempre menor al número de variables ecogeográficas) dentro del análisis PCA/PCoA.

#### 7.4.1.38 Parámetro: *phenovarq*

Explicación: Seleccione esta opción si la totalidad de las variables/descriptores fenotípicos corresponden a variables de tipo cuantitativo.

#### 7.4.1.39 Parámetro: *genoclus*

Explicación: Seleccione esta opción si desea realizar un análisis de agrupamientos de todas las entradas para las cuales se aporta información genotípica.

#### 7.4.1.40 Parámetro: *genoclustype*

Explicación: Elija el tipo de agrupamiento jerárquico que desea utilizar para los agrupamientos genotípicos: 'single' = vecino más cercano, 'complete' = vecindario más compacto, 'ward' = método de mínima varianza de Ward, 'mcquitty' = método de McQuitty, 'average' = similaridad promedio (UPGMA), 'median' = similaridad de la mediana, 'centroid' = centroide geométrico, 'flexible' = Beta flexible.

#### 7.4.1.41 Parámetro: *genopco*

Explicación: Seleccione esta opción si desea realizar un análisis de coordenadas principales de todas las entradas para las cuales se ha aportado información genotípica.

#### 7.4.1.42 Parámetro: *genopcoaxe*

Explicación: Numero de coordenadas a retener (siempre menor al número de variables ecogeográficas) dentro del análisis PCoA.

#### 7.4.1.43 Parámetro: *mantelt*

Explicación: Indique si desea analizar las correlaciones matriciales (Mantel, 1967) entre las posibles combinaciones de factores (ecogeográfico vs. fenotípico vs. genotípico). Se realizaran todas las posibles comparaciones de acuerdo a si se introdujeron datos fenotípicos o genotípicos o si se creó una matriz ecogeográfica a partir de sitios de recolección. Se generará e incluirá una matriz de distancias geográficas en las comparaciones matriciales pareadas.

#### 7.4.1.44 Parámetro: *mantelmeth*

Explicación: Seleccione el tipo de correlación a usar en la prueba de Mantel.

#### 7.4.1.45 Parámetro: *mantelper*

Explicación: Ingrese el número de permutaciones que desea para realizar la prueba de Mantel.

#### 7.4.1.46 Parámetro: *resultados*

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

## 7.5. Resultados de DIV mapas

En el caso de DIV mapas, la cantidad de gráficas, tablas y mapas puede ser muy variable, dependiendo de los datos que se introduzcan, las opciones que se indiquen y los análisis que desee realizar el usuario. Dado que la ejecución de DIV mapas

puede llegar a producir muchos resultados y que éstos pueden ser organizados de acuerdo al tipo de datos y/o análisis de donde provienen, DIV mapas crea varias carpetas dentro de la ruta indicada en el parámetro resultados (ver apartado 7.4.1.46). De esta manera, los resultados se guardan en las correspondientes carpetas temáticas que se explicarán en los siguientes apartados, y sólo se ubicará fuera de ellas el mapa de puntos correspondiente a los sitios de recolección en dos versiones (“ShapefilePuntosPasaporte.shp” y “mapa\_puntospas\_google.kml”) y una tabla (“Valor\_mediana\_para\_bootstrapping.txt”) que sólo aparecerá cuando se solicite procesos de remuestreo y que contiene el valor de la mediana utilizado como umbral en dicho proceso.

### 7.5.1 Carpeta “ClassicMultivariateResults\_pais”

Esta carpeta contiene gráficas (formato .wmf) y tablas (.txt) producto de análisis multivariados (análisis de agrupamientos y análisis de componentes principales), tal y como se detallan en los apartados 5.4.1 para las figuras y 5.4.2.2 para las tablas. Dependiendo de los datos introducidos, aparecerá en el nombre de los archivos los términos “ecogeographic” (proviene de la caracterización ecogeográfica), “genotypic” (proviene de la caracterización genotípica) y “phenotypic” (proviene de la caracterización fenotípica). Los resultados guardados en esta carpeta corresponden a los productos de análisis de todas las entradas a la vez, es decir, de la manera usual de análisis, tal y como los realiza la herramienta ECOGEO.

### 7.5.2 Carpeta “EcogeographicResults\_pais”

Esta carpeta aparecerá en el caso de que se haya solicitado la obtención de un mapa de diversidad ecogeográfica (ver apartado 7.4.1.11). Contiene el mapa de diversidad en tres formatos (imagen “mapadiv\_ecogeo\_pais.png”, Google earth “mapadiv\_ecogeo\_pais.kml”, y DIVA-GIS “mapadiv\_ecogeo\_pais.grd”). En estos mapas aparecen representados con diferentes colores el promedio de las distancias ecogeográficas de las entradas del área de influencia de cada celda. La distancia corresponde a la euclídea y tiene posibles valores desde 0 (para el caso en el que haya una sola entrada o que todas las entradas hayan sido recolectadas en ambientes idénticos) hasta infinito.

Respecto a las tablas, pueden encontrarse las siguientes:

7.5.2.1 “tabla\_estadisticas\_mapadiv\_ecogeo.txt”. Esta tabla muestra las estadísticas del mapa de diversidad ecogeográfica en términos de la distancia, es decir, la media, desviación estándar y valores máximos y mínimos de distancia, determinados en el conjunto de celdas que componen el mapa.

7.5.2.2 “TablaVarEcogeograficasecuador.txt”. Esta tabla contiene los datos de caracterización ecogeográfica de las entradas analizadas. Esta tabla se asemeja a la que produce la herramienta ECOGEO en el apartado 5.4.2.1.

7.5.2.3 "DistanciasMedias\_sin\_con\_bootstrap\_ECOGEO.txt". Esta tabla sólo estará disponible cuando se haya solicitado realizar análisis de remuestreo (apartado 7.4.1.3). En ella aparecen los valores de distancia promedio para cada celda (denominada aquí como "grupo") sin remuestreo (columna "WO\_bootstrap") y con remuestreo ("W\_bootstrap").

7.5.2.4 "TestBootstrapping.txt". Esta tabla sólo estará disponible cuando se haya solicitado realizar análisis de remuestreo (apartado 7.4.1.3). Esta tabla muestra dos test de asociación entre los valores de distancia promedio provenientes de procesos con remuestreo y sin remuestreo (tabla 7.5.2.3).

### 7.5.3 Carpeta "PhenotypicResults\_pais"

Esta carpeta aparecerá en el caso de que se haya solicitado la obtención de un mapa de diversidad fenotípica (ver apartado 7.4.1.11) y se haya aportado la tabla con los datos correspondientes. Contiene el mapa de diversidad en tres formatos (imagen "mapadiv\_phenot\_pais.png", Google earth "mapadiv\_phenot\_pais.kml", y DIVA-GIS "mapadiv\_phenot\_pais.grd"). En estos mapas aparecen representados con diferentes colores el promedio de las distancias fenotípicas de las entradas del área de influencia de cada celda. La distancia corresponde a 1-coeficiente general de similaridad de Gower (1971) y tiene posibles valores desde 0 (para el caso en el que haya una sola entrada o que todas las entradas presenten fenotipo idéntico) hasta 1 (máxima diferencia).

Respecto a las tablas, pueden encontrarse las siguientes:

7.5.3.1 "tabla\_estadisticas\_mapadiv\_phenot.txt". Esta tabla muestra las estadísticas del mapa de diversidad fenotípica en términos de la distancia, es decir, la media, desviación estándar y valores máximos y mínimos de distancia, determinados para todo el conjunto de celdas que componen el mapa.

7.5.3.2 "TestBootstrapping.txt". Esta tabla sólo estará disponible cuando se haya solicitado realizar análisis de remuestreo (apartado 7.4.1.3). Esta tabla muestra dos test de asociación entre los valores de distancia promedio provenientes de procesos con remuestreo y sin remuestreo.

### 7.5.4 Carpeta "GenotypicResults\_pais"

Esta carpeta aparecerá en el caso de que se haya solicitado la obtención de un mapa de diversidad genotípica (ver apartado 7.4.1.11) y se haya aportado la tabla con los datos correspondientes.

Al interior de la carpeta se encontrarán los siguientes mapas:

7.5.4.1 "mapadiv\_GenotDistance\_pais". Corresponde al mapa de diversidad genotípica medida en distancias promedio en tres formatos (imagen ".png",

Google earth “.kml”, y DIVA-GIS “.grd”). En estos mapas aparecen representados con diferentes colores el promedio de las distancias genotípicas de las entradas caracterizadas del área de influencia de cada celda. La distancia corresponde a 1-coeficiente de similaridad de Dice (1945) y tiene posibles valores desde 0 (para el caso en el que haya una sola entrada o que todas las entradas presenten fenotipo idéntico) hasta 1 (máxima diferencia).

7.5.4.2 “mapadiv\_GroupSize\_pais”. Corresponde al mapa de número de entradas analizadas por celda (imagen “.png”, Google earth.kml”, y DIVA-GIS “.grd”). En estos mapas aparecen representados con diferentes colores el número las entradas de cada área de influencia de cada celda.

7.5.4.3 “mapadiv\_NeiGeneDiversity\_pais”. Corresponde al mapa de diversidad genotípica medida por el índice de diversidad de Nei (1987) en tres formatos (imagen “.png”, Google earth “.kml”, y DIVA-GIS “.grd”). En estos mapas aparecen representados con diferentes colores el índice de diversidad antes mencionado, obtenido de las entradas caracterizadas del área de influencia de cada celda.

7.5.4.4 “mapadiv\_ProportionVariableMarkers\_pais”. Corresponde al mapa que muestra la proporción de marcadores polimórficos en tres formatos (imagen “.png”, Google earth “.kml”, y DIVA-GIS “.grd”). En estos mapas aparecen representados con diferentes colores la proporción de marcadores moleculares polimórficos obtenida de las entradas caracterizadas del área de influencia de cada celda.

Respecto a las tablas, pueden encontrarse las siguientes:

7.5.4.5 “tabla\_estadisticas\_mapa\_GenotDistance.txt”. Esta tabla muestra las estadísticas del mapa de diversidad genotípica (mapa 7.5.4.1) en términos de la distancia de Dice (1945), es decir, la media, desviación estándar y valores máximos y mínimos de dicha distancia, determinados para todo el conjunto de celdas que componen el mapa.

7.5.4.6 “tabla\_estadisticas\_mapa\_NeiGeneDiversity.txt”. Esta tabla muestra las estadísticas del mapa de diversidad genotípica (mapa 7.5.4.3) en términos del índice de diversidad genética de Nei (1987), es decir, la media, desviación estándar y valores máximos y mínimos para este índice, determinados para todo el conjunto de celdas que componen el mapa.

7.5.4.7 “NeiGeneDiversityMedias\_sin\_con\_bootstrap.txt”. Tabla con los índices de diversidad genética de Nei (1987) obtenidos para cada celda (aquí llamada “group”) sin remuestreo (columna “WO\_bootstrap”) y con remuestreo (“W\_bootstrap”). Esta tabla sólo estará disponible cuando se haya solicitado realizar análisis de remuestreo (apartado 7.4.1.3).

7.5.4.8 “ProportPolymorphMarkersMedias\_sin\_con\_bootstrap.txt”. Tabla con los valores de proporción de marcadores polimórficos obtenidos para cada celda (aquí

llamada "group") sin remuestreo (columna "WO\_bootstrap") y con remuestreo ("W\_bootstrap"). Esta tabla sólo estará disponible cuando se haya solicitado realizar análisis de remuestreo (apartado 7.4.1.3).

7.5.4.9 "DistanciasMedias\_sin\_con\_bootstrap\_DICE.txt". En esta tabla aparecen los valores de distancia promedio de Dice (1945) para cada celda (denominada aquí como "group") sin remuestreo (columna "WO\_bootstrap") y con remuestreo ("W\_bootstrap"). Esta tabla sólo estará disponible cuando se haya solicitado realizar análisis de remuestreo (apartado 7.4.1.3).

7.5.4.10 "TestBootstrappingDICE.txt". Esta tabla muestra dos test de asociación entre los valores de distancia promedio de Dice (1945) provenientes de procesos con remuestreo y sin remuestreo. Esta tabla sólo estará disponible cuando se haya solicitado realizar análisis de remuestreo (apartado 7.4.1.3).

7.5.4.11 "TestBootstrappingNei.txt". Esta tabla muestra dos test de asociación entre los valores del índice de diversidad de Nei (1987) provenientes de procesos con remuestreo y sin remuestreo. Esta tabla sólo estará disponible cuando se haya solicitado realizar análisis de remuestreo (apartado 7.4.1.3).

7.5.4.12 "TestBootstrappingPPM.txt". Esta tabla muestra dos test de asociación entre los valores de proporción de marcadores polimórficos provenientes de procesos con remuestreo y sin remuestreo. Esta tabla sólo estará disponible cuando se haya solicitado realizar análisis de remuestreo (apartado 7.4.1.3).

#### 7.5.4 Carpeta "MantelCorrelationResults\_pais"

En esta carpeta se guardarán todas las tablas con las matrices de distancias calculadas para todas las entradas a la vez ("Matriz\_distancia\_") y las tablas con los resultados de las correlaciones matriciales de Mantel (1967). El nombre de cada tabla indica la comparación de información realizada. Para las correlaciones donde intervienen datos genotípicos se usa la matriz de distancia de Dice. Por ejemplo, el archivo "Mantel\_genotypic\_Vs\_phenotypic.txt" contiene los resultados de la correlación matricial entre las distancias genotípicas (Dice) y las distancias fenotípicas (Gower). Es importante anotar que DIV mapas también calcula la matriz de distancias geográficas (calculadas en grados decimales) para poder hacer comparaciones matriciales respecto al componente de distancia geográfica.

## 7.6. Referencias

Damme, P., Garcia, W., Tapia, C., Romero, J., Manuel Sigueñas, M. y Hormaza, J.I. 2012. Mapping Genetic Diversity of Cherimoya (*Annona cherimola* Mill.): Application of Spatial Analysis for Conservation and Use of Plant Genetic Resources. PLoS ONE

7(1): e29845. doi:10.1371/journal.pone.0029845

Dice, L.R. 1945. Measures of the Amount of Ecologic Association Between Species. *Ecology* 26:297–302.

FAO, BIOVERSITY. 2012. FAO/Bioversity multi-crop Passport descriptors V.2. Disponible en

[http://www.bioversityinternational.org/index.php?id=19&user\\_bioversitypublications\\_pi1%5BshowUid%5D=6901](http://www.bioversityinternational.org/index.php?id=19&user_bioversitypublications_pi1%5BshowUid%5D=6901)

Gower, J.C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27: 857-74.

Hoffmann, M.H., Glaß, A.S., Tomiuk, J., Schmuths, H., Fritsch, R.M. y Bachmann, K. 2003. Analysis of molecular data of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS). *Molecular Ecology*, 12: 1007–1019

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach.

*Cancer Res.* 27: 209-220.

Thomas, E., van Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G., y van Etten, J. 2012. Present spatial diversity patterns of *Theobroma cacao* L. in the neotropics reflect genetic differentiation in pleistocene refugia followed by human-influenced dispersal. *PLoS ONE* 7(10): e47676. doi:10.1371/journal.pone.0047676

van Zonneveld M, Scheldeman X, Escribano P, Viruel MA, Van Damme P, et al. (2012) Mapping Genetic Diversity of Cherimoya (*Annona cherimola* Mill.): Application of Spatial Analysis for Conservation and Use of Plant Genetic Resources. *PLoS ONE* 7(1): e29845. doi:10.1371/journal.pone.0029845



## 8. Herramienta ColNucleo

### 8.1. Concepto de colección núcleo

Una colección núcleo o nuclear (en adelante núcleo) es una subcolección, o una fracción de una colección original, que se puede realizar por diversos motivos. El tamaño de la colección original es la mayor determinante a la hora de tomar la decisión de crear una colección núcleo. Las colecciones núcleo suelen considerarse una solución cuando el tamaño de las colecciones originales se convierte en un problema. Una colección de mayor tamaño puede representar un problema a la hora de multiplicar, caracterizar o evaluar germoplasma, en particular cuando los recursos económicos son limitados, seleccionar materiales para programas de mejoramiento, crear colecciones activas o de trabajo, etc. Dependiendo de las condiciones de cada sitio, una colección de mayor tamaño puede ser una conformada por 500, 1000, 2000 o más entradas.

Una colección núcleo suele tener un tamaño del 10% del total de entradas de la colección original, si bien hay estudios que sitúan el porcentaje óptimo por encima o por debajo de este valor (Parra Quijano et al., 2011a). A este porcentaje se le denomina "intensidad de muestreo".

Es importante anotar que de ninguna manera la determinación de una colección núcleo debe poner en riesgo la conservación de las entradas no seleccionadas, las que se conocen como "colección de reserva". Una colección núcleo puede facilitar la toma de decisiones en ciertas actividades de conservación y determinación de prioridades bajo condiciones de recursos insuficientes, lo cual no exime la responsabilidad de conservar la colección por completo. Por ejemplo cuando es necesario realizar una multiplicación de germoplasma, se cuenta con una colección núcleo y los recursos son limitados, se podría plantear multiplicar en primer término las entradas de la colección núcleo y con recursos adicionales en otro ciclo de multiplicación, el resto de la colección.

Independiente de los motivos que lleven a su creación, la principal característica de una colección núcleo respecto a otro tipo de subcolecciones es que ésta debe ser representativa de la diversidad genética contenida en la colección original. Esto implica que una colección núcleo debe contener entradas lo más disímiles entre sí, de tal manera que desde un punto de vista genético no se incluyan entradas duplicadas o altamente emparentadas (Brown, 1995). Por consiguiente para obtener un subconjunto de entradas disímiles genéticamente, es imprescindible contar con información sobre la composición genética de la colección, en otras palabras, disponer de datos de caracterización.

En este punto aparece uno de los primeros limitantes para obtener colecciones núcleo, en un escenario de recursos limitados, caracterizar una colección de más de

1000 o 2000 entradas puede resultar inabordable. Las caracterizaciones genotípicas y fenotípicas suelen demandar ingentes esfuerzos económicos y humanos que muchas instituciones no pueden asumir o lo hacen de manera parcial. Por lo tanto, para superar este tipo de impedimentos se puede utilizar otro tipo de datos de caracterización.

En ciertos casos, cuando se ha requerido la creación de colecciones núcleo y no se contaban con datos de caracterización para su obtención, la solución propuesta fue el uso de datos de pasaporte, particularmente aquellos datos administrativos que describen la localidad del sitio de la recolección (país, estado, provincia). Con ello, se pretendía asimilar diferentes unidades administrativas de recolección a diferentes ambientes y así lograr una colección núcleo representativa de todas las unidades administrativas y ambientes. De esta forma se crearon diversas colecciones núcleo administrativas en especies como cacahuete (Upadhyaya et al., 2003), guandul (Reddy et al., 2005), sésamo (Xiourong et al., 2000) o sorgo (Grenier et al., 2001). Este tipo de colecciones no asegurarían que la colección núcleo esté conformada por las entradas más disímiles en términos del ambiente donde fueron colectadas, dado que las diferentes unidades administrativas son divisiones antrópicas que no corresponden necesariamente a ambientes diferentes.

### 8.1.1 Estrategia de agrupación

El primer paso para la obtención de una colección núcleo es la formación de grupos en la colección original por afinidad. Como se ha comentado previamente, para obtener una colección núcleo es necesario contar con datos de caracterización ecogeográfica, genotípica, fenotípica o en el caso de colecciones núcleo administrativas, datos de pasaporte. Esta información se usa para crear grupos de entradas afines o similares. La creación de los grupos suele realizarse a través de métodos de clasificación multivariada a partir de los datos de caracterización del germoplasma.

En una variante para colecciones núcleo de tipo ecogeográfico, se propone no usar la caracterización del germoplasma sino del territorio donde este ocurre, a través del uso de mapas ecogeográficos de caracterización del territorio (mapas ELC). Así las entradas son agrupadas de acuerdo a la categoría del mapa en el que ocurren. Esto facilitaría la inclusión de nuevas entradas a la colección nuclear, dado que no sería necesario repetir el análisis de agrupamientos, simplemente conocer el grupo (categoría ecogeográfica del mapa) de pertenencia de la nueva entrada (Parra Quijano et al., 2011b).

### 8.1.2 Determinación de cuotas por estrategias de afijación

Posteriormente, se determina el número de entradas que serán seleccionadas de cada grupo de afinidad. Ese número o cuota la define la estrategia de afijación (*allocation strategy* en inglés) que el curador considere más adecuada. A medida que las colecciones núcleo se han popularizado, se han propuesto numerosas estrategias

de afijación. Gradualmente la complejidad y sofisticación de dichas estrategias se ha ido incrementando. Sin embargo algunos estudios comparativos muestran que las estrategias más complejas no necesariamente son las que producen colecciones nucleares más representativas (Parra Quijano et al., 2011b). Las estrategias más recurrentes, sencillas y testadas son las siguientes (Yonezawa et al., 1995):

- a) Aleatoria (R): Las entradas son seleccionadas de manera aleatoria del total de la colección. Los grupos creados por la estratificación son ignorados.
- b) Constante (C): El mismo número de entradas se selecciona de cada grupo, sin tener en cuenta el número de entradas que conforman el grupo.
- c) Proporcional (P): El número de entradas a seleccionar de cada grupo es proporcional al tamaño (número de entradas) de cada grupo.
- d) Logarítmico (L): El número de entradas a seleccionar de cada grupo es proporcional al logaritmo del tamaño (número de entradas) de cada grupo.
- e) Dependiente de la diversidad (G): El número de entradas a seleccionar de cada grupo es proporcional a la diversidad contenida en cada grupo. Esta estrategia requiere acceso a datos de caracterización, adicionalmente a los agrupamientos provenientes de la estratificación.

### 8.1.3 Información sobre disponibilidad de entradas

En buena parte de los estudios científicos sobre la creación de colecciones núcleo, se realizan simulaciones para determinar cuál es la estrategia de agrupamiento y de afijación que produce la colección núcleo más representativa para cada caso, utilizando para ellos el total de la colección. Sin embargo estas aproximaciones teóricas y simulaciones pueden producir colecciones núcleo que en la práctica podrían no constituirse debido a que las entradas seleccionadas no están disponibles. La disponibilidad de una entrada para ser parte de una colección núcleo, puede estar influenciada por varios factores, entre ellos el número de semillas disponible, si la entrada está sólo representada en la colección base o si la entrada tiene algún tipo de restricción en su uso y distribución. Por esta razón es importante que en el caso de la obtención de una colección núcleo de carácter práctico, se tenga en cuenta la información de disponibilidad que pueda aportar el curador de la colección.

## 8.2. Colecciones núcleo ecogeográficas

La caracterización ecogeográfica representa una alternativa para la obtención de colecciones núcleo. Teniendo en cuenta la relación entre fenotipo, genotipo y ambiente, una colección núcleo creada a partir de datos de caracterización ecogeográfica puede ser representativa no sólo de las condiciones ambientales de las poblaciones de procedencia de las entradas contenidas en la colección original sino también de sus fenotipos y genotipos, siempre que la representatividad sea evaluada mediante características fenotípicas o genotípicas relacionadas con la adaptación (Parra Quijano et al., 2011a).

El uso de datos de caracterización ecogeográfica del germoplasma en la obtención de colecciones núcleo se ha documentado desde 1995, en el caso del establecimiento de una colección núcleo de *Phaseolus vulgaris* en el Centro Internacional de Agricultura Tropical (CIAT) (Tohme et al., 1995). Dado que la amplia disponibilidad de SIG aplicados a recursos fitogenéticos y de capas de información ecogeográfica no fue posible hasta la década 2000-2010, las colecciones núcleo ecogeográficas no reaparecen en el contexto científico internacional hasta 2008, con el caso de *Trifolium spumosum* (Ghamkhar et al., 2008).

Posteriormente, en un par de estudios sobre diferentes tipos de colecciones núcleo ecogeográficas, se determinó que la combinación mapa de caracterización ecogeográfica del territorio (mapa ELC) como estrategia de agrupamiento y el tipo proporcional como estrategia de afijación, produjo colecciones núcleo de alta representatividad ecogeográfica y fenotípica para colecciones de *Lupinus spp.* y *Phaseolus vulgaris*, respectivamente (Parra Quijano et al., 2011a, 2011b). En dichos estudios, hasta 16 combinaciones diferentes de estrategias de agrupamiento y afijación, dieron resultados similares o inferiores en términos de representatividad ecogeográfica y fenotípica que la combinación mapa ELC junto con afijación proporcional.

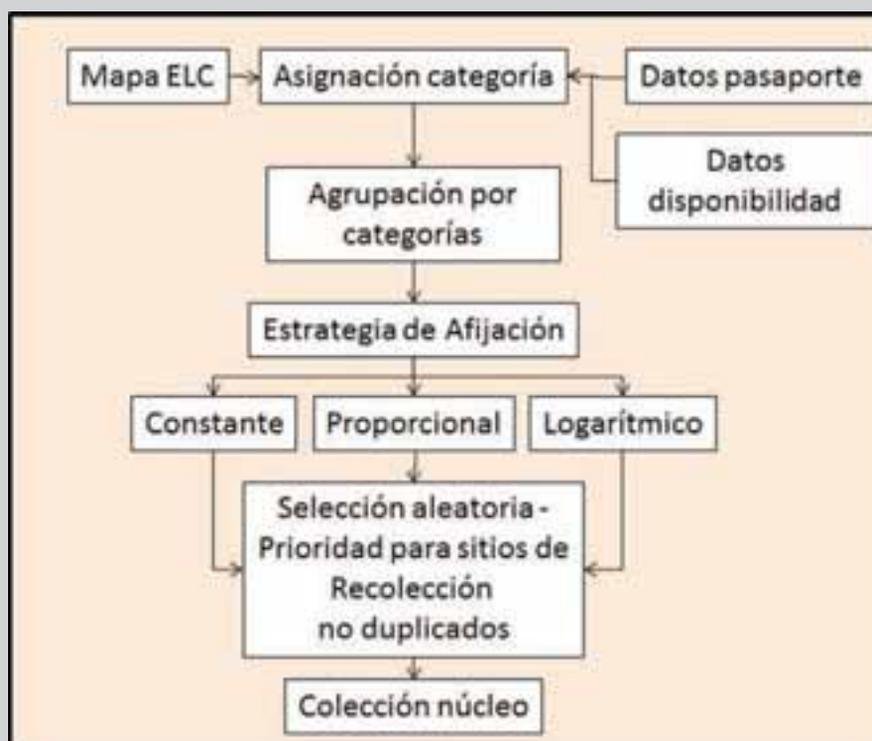
### 8.3. Obtención de colecciones núcleo ecogeográficas en ColNucleo

Siguiendo las recomendaciones de algunos estudios científicos sobre colecciones núcleo y representatividad, la herramienta ColNucleo permite obtener colecciones núcleo ecogeográficas mediante la combinación de agrupamiento por mapa ELC y tres métodos de afijación (C, P y L). El mapa ELC debe ser generado con la herramienta "ELC mapas" (ver Capítulo 4).

La Fig. 30 muestra que, en un primer paso, se extrae la categoría del mapa ELC correspondiente al sitio de recolección de cada entrada con coordenadas. Posteriormente las entradas se agrupan de acuerdo a la categoría del mapa ELC asignada. De acuerdo a la estrategia de afijación y la intensidad de muestreo que el usuario haya seleccionado, ColNucleo fija las cuotas o número de entradas de cada grupo que conformarán en conjunto la colección núcleo. Después, ColNucleo determina si la cuota se puede cumplir con entradas que no correspondan a duplicados geográficos (no necesariamente genético) y que el curador ha designado como "disponibles" en el caso de que el usuario haya señalado la opción de utilizar datos de disponibilidad. Las entradas no duplicado tendrán prioridad sobre entradas duplicado. Si la cuota es de menor tamaño que el número de entradas no duplicadas disponibles, se seleccionará aleatoriamente entre éstas. Si la cuota es de mayor tamaño se seleccionarán todas las entradas disponibles no duplicadas y el faltante se cubrirá con entradas duplicadas seleccionadas aleatoriamente. Finalmente las entradas seleccionadas serán señaladas con el número 1 (uno) en una nueva columna que se añadirá a la tabla de pasaporte de entrada. En el caso de haberse utilizado datos de disponibilidad, las colecciones núcleo obtenidas pueden ser incompletas al no disponerse de suficientes entradas para representar una o

varias categorías ELC. Por esta razón ColNucleo genera una tabla adicional en la que señala qué accesiones serían necesarias hacer disponibles para que la colección núcleo represente todas las categorías ELC de acuerdo a las cuotas fijadas.

**FIGURA 30.** Ilustración del proceso que sigue la herramienta ColNucleo para la obtención de colecciones núcleo ecogeográficas.



#### 8.4. Formato de tabla de pasaporte para ColNucleo

ColNucleo utiliza la tabla de pasaporte FAO/Bioersity 2012 con modificaciones que utiliza las herramientas GEOQUAL, Representa y ECOGEO, con la adición de un campo en la parte derecha, denominado "AVAILAB" que determina la disponibilidad de cada entrada. De esta manera, en la columna AVAILAB se codificará con el número 1 (uno) las entradas disponibles, con un 0 (cero) las no disponibles y con las letras NA las entradas para las cuales no se tiene información.

#### 8.5. Utilización de la herramienta ColNucleo

Una vez instaladas las herramientas CAPFITOGEN y seleccionada la herramienta ColNucleo, será necesario definir una serie de parámetros para que la programación R funcione correctamente.

## 8.5.1 Parámetros iniciales definidos por el usuario

### 8.5.1.1 Parámetro: ruta

Explicación: Ruta donde se ha copiado o se encuentran las herramientas CAPFITOGEN.  
Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo F:/CAPFITOGEN, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

### 8.5.1.2 Parámetro: pasaporte

Explicación: Escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama "tabla" debe escribir: "tabla.txt". Recuerde que este archivo debe ser guardado previamente en la carpeta "Pasaporte" que hace parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Ésta tabla contiene una columna adicional denominada "AVAILAB" respecto a la tabla de pasaporte usada como modelo para otras herramientas CAPFITOGEN. Ésta columna adicional indica la disponibilidad de cada entrada a ser seleccionada para una colección núcleo/nuclear.

### 8.5.1.3 Parámetro: geoqual

Explicación: Seleccione esta opción si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 51 columnas (y no las 46 del modelo de tabla exclusivo para CoINucleo sin haber sido sometida a GEOQUAL). Use por tanto la tabla resultado de GEOQUAL denominada PasaporteOriginalEvaluadoGEOQUAL.txt como tabla de pasaporte en el punto anterior.

### 8.5.1.4 Parámetro: totalqual

Explicación: Si su tabla de pasaporte proviene de GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores entre 0 (calidad nula) a 100 (máxima calidad).

### 8.5.1.5 Parámetro: mapaelc

Explicación: Indique el nombre del archivo que contiene el mapa ELC (producto de la aplicación de la herramienta ELCmapas), el cual debe encontrarse en la carpeta ELCmapas, la cual es una de las carpetas que componen el directorio CAPFITOGEN. Este mapa debe estar en formato DIVA-GIS (extensión .grd tal y como lo produce ELCmapas) y debe escribirse con la extensión. Así, si el nombre del mapa es "mapa\_elc\_spain", ud. debe escribir "mapa\_elc\_spain.grd".

### 8.5.1.6 Parámetro: statelc

Explicación: Indique el nombre del archivo que contiene la tabla de estadísticas descriptivas del mapa ELC producido con la herramienta ELCmapas (ésta herramienta suele nombrar a éste archivo como "Estadist\_ELC\_" más el nombre del país o región).

Al igual que el mapa ELC, este archivo debe estar situado en la carpeta ECLmapas. De igual forma el nombre debe escribirse junto con la extensión, en este caso al ser una tabla, la extensión es txt. Así, si el archivo se llama "Estadist\_ELC\_spain" se debe escribir "Estadist\_ELC\_spain.txt".

#### 8.5.1.7 Parámetro: *distdup*

Explicación: Determine el valor de distancia (en km) por debajo del cual Ud. consideraría que dos sitios de recolección en realidad representan una misma población (duplicado geográfico). El valor cero (valor mínimo y por defecto) da menos prioridad a la hora de la selección a entradas con idénticas coordenadas. En la medida que el valor de distancia aquí estipulado aumente, un mayor número de entradas se considerarían duplicados geográficos.

#### 8.5.1.8 Parámetro: *porcol*

Explicación: Corresponde a la intensidad de muestreo. Indique el tamaño que desea para la colección núcleo/nuclear, expresado en porcentaje del tamaño de la colección original (valores de 0 a 100). Por ejemplo, si la colección original contiene 2000 entradas y se desea una colección núcleo/nuclear de 200 entradas, el valor a indicar aquí será 10, o para una colección núcleo/nuclear de 300 entradas, el valor será 15.

#### 8.5.1.9 Parámetro: *estratcol*

Explicación: Seleccione la estrategia de afijación o fijación de cuotas de representación de cada categoría ecogeográfica del mapa ELC. Puede escoger entre las estrategias "C" constante (la misma cuota para todas las categorías), "P" proporcional (cuotas proporcionales a la cantidad de entradas de cada categoría) o "L" logarítmica (cuotas proporcionales al logaritmo de la cantidad de entradas de cada categoría).

#### 8.5.1.10 Parámetro: *availab*

Explicación: Seleccione esta opción si desea utilizar la columna sobre disponibilidad de entradas para ser seleccionadas con destino a una colección núcleo. Recuerde que para esta herramienta, la tabla de pasaporte incluye una columna denominada "AVAILAB", donde se indica la disponibilidad de las entradas de la colección original para integrar colecciones núcleo, a través de los códigos 0 (entrada no disponible), 1 (entrada disponible) y NA (sin información/no disponible). Si no desea utilizar información sobre disponibilidad, la selección de entradas se realizará sobre el total de las entradas. La disponibilidad es un criterio propio del curador y podría estar determinado por el número de semillas conservadas, su germinación u otros diversos factores.

#### 8.5.1.11 Parámetro: *resultados*

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / en lugar de \ en la indicación de la ruta de la carpeta. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

## 8.6. Resultados de ColNucleo

Una vez finalizado el análisis, ColNucleo producirá un total de cuatro tablas, si el usuario ha indicado utilizar los datos de disponibilidad (parámetro 8.5.1.10) o tres en el caso que haya decidido no utilizar dicha información.

8.6.1.1 "CoreCollection.txt". Esta tabla contiene la tabla de pasaporte de las entradas que ColNucleo ha seleccionado como colección núcleo ecogeográfica, con una columna adicional en el extremo derecho de la tabla denominada "BGcat", que indica el grupo o categoría del mapa ELC al que pertenece la entrada de acuerdo a la ubicación del sitio de recolección.

8.6.1.2 "CoreCollect\_Properties.txt". Esta tabla muestra varios de los parámetros introducidos por el usuario y que ColNucleo ha utilizado para obtener la colección núcleo. Los campos que incluye son "Allocation\_strategy" o estrategia de afijación, "Sample\_size" o porcentaje de intensidad de muestreo, "Use\_availability\_data" o si se usa o no datos de disponibilidad, "No\_access\_sampled", tamaño de la colección núcleo obtenida y "No\_access\_to\_be\_multiplied" (sólo para el caso de usar datos de disponibilidad) o número de entradas que no están disponibles o no se tiene información sobre su disponibilidad, y que deberían estarlo para lograr una colección núcleo completa.

8.6.1.3 "CoreCollect\_stats.txt". Esta tabla contiene estadísticas por cada categoría del mapa ELC (identificadas en la columna "ELC\_CAT"). Contiene las siguientes columnas a la derecha de "ELC\_CAT": "FREC\_W\_DUPL" que indica el número de entradas, incluyendo duplicados geográficos, cuyo sitio de recolección ocurre dentro de cada categoría, "FREC\_WO\_DUPL" indica lo mismo que la anterior columna pero excluyendo duplicados geográficos, "Porcent\_W\_DUPL" indica el porcentaje de entradas (contando duplicados) que ocurre en cada categoría, "FreqClass\_W\_DUPL" indica la clasificación en cuartiles de la frecuencia de ocurrencia en cada categoría de manera similar a como se indica en el apartado 6.4.1.1, "Duplicates" es el número de entradas duplicados geográficos por cada categoría, "N\_Availab" es el número de entradas disponibles en total (duplicados más no duplicados) para cada categoría, "N\_AvailabWO" es el número de entradas disponibles no duplicadas para cada categoría, "Q\_Even" ó "Q\_Prop" ó "Q\_Log" (aparecería el título de columna dependiendo del método de afijación seleccionado) es el número de entradas que deben representar *a priori* cada categoría, es decir, la cuota, y finalmente la columna CCfinal que corresponde al número de entradas que componen la colección núcleo ecogeográfica obtenida por ColNucleo en base a los parámetros introducidos y, en su caso, disponibilidad de entradas.

8.6.1.4 "EntriesToBeMultiplied.txt". Se trata de una tabla con la misma estructura de columnas que "CoreCollection.txt" sólo que en "EntriesToBeMultiplied.txt" aparecen las entradas que ColNucleo ha seleccionado como parte de la colección núcleo pero para las cuales no hay disponibilidad. En principio se asume, por el nombre del archivo, que pueden ser entradas que requerirían ser multiplicadas para ser disponibles y entonces hacer parte de la colección núcleo. Sin embargo, la razón por las cuales no están disponibles pueden ser variadas como se explica en el apartado 8.1.3.

## 8.7. Referencias

Brown, A.H.D. 1995. The core collection at the crossroads. p. 3–19. En Hodgkin, T., Brown, A.H.D., Hintum, T.J.L., Morales, E.A.V. (ed.) Core collections of plant genetic resources. John Wiley & Sons, New York, NY.

Ghamkhar, K., R. Snowball, B.J. Wintle, Brown, A.H.D. 2008. Strategies for developing a core collection of bladder clover (*Trifolium spumosum* L.) using ecological and agro-morphological data. Aust. J. Agric. Res. 59:1103–1112.

Grenier, C., Hamon, P., Bramel-Cox, P.J.. 2001. Core collection of sorghum: II. Comparison of three random sampling strategies. Crop Science. 41:241–246.

Parra Quijano, M., Iriondo, J.M., Torres, M.E., De la Rosa, L. 2011a. Evaluation and validation of ecogeographical core collections using phenotypic data. Crop Science 51:694-703.

Parra-Quijano, M., Iriondo, J.M., de la Cruz, M., Torres, M.E. 2011. Strategies for the development of core collections based on ecogeographical data. Crop Science 51:656-666.

Reddy, L.J., H.D. Upadhyaya, C.L.L. Gowda, S. Singh. 2005. Development of core collection in pigeonpea (*Cajanus cajan* (L.) Millspaugh) using geographic and qualitative morphological descriptors. Genetic Resources and Crop Evolution 52:1049–1056.

Tohme, J., P. Jones, S. Beebe, and M. Iwanaga. 1995. The combined use of agroecological and characterisation data to establish the CIAT Phaseolus vulgaris core collection. p. 95–107. In Hodgkin, T., Brown, A.H.D., Hintum, T.J.L., Morales, E.A.V. (ed.) Core collections of plant genetic resources. John Wiley & Sons, New York, NY.

Upadhyaya, H.D., Ortiz, R., Bramel, P.J., S. Singh, S. 2003. Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. Genet. Resour. Crop Evol. 50:139–148.

Xiurong, Z., Yingzhong, Z., Yong, C., Xiangyun, F., Qingyuan, G., Mingde, Z., Hodgkin, T. 2000. Establishment of sesame germplasm core collection in China. Genet. Resour. Crop Evol. 47:273–279.

Yonezawa, K., Nomura, T., Morishima, H. 1995. Sampling strategies for use in stratified germplasm collections. p. 35–53. In Hodgkin, T., Brown, A.H.D., Hintum, T.J.L., Morales, E.A.V. (ed.) Core collections of plant genetic resources. John Wiley & Sons, New York, NY.



## 9. Herramienta FIGS\_R

### 9.1. La Estrategia de Selección Focalizada de Germoplasma

La técnica de selección de germoplasma con fines de utilización denominada “Estrategia de Selección Focalizada de Germoplasma” o FIGS por sus siglas en inglés (Focused Identification of Germplasm Strategy), proviene de un concepto originalmente desarrollado por Mackay (1990).

Ésta técnica busca identificar entradas de una colección que tienen potencial de uso por parte de mejoradores. El potencial para uso en mejoramiento está basado en el uso de información ecogeográfica de los sitios de recolección y su asociación con caracteres de interés para el mejorador de cultivos (Mackay y Street, 2004).

Dado que FIGS usa variables ecogeográficas de tipo abiótico para la selección germoplasma, la asociación entre las variables ecogeográficas y el carácter de interés para el mejoramiento es directa si el carácter de interés es de tipo abiótico o indirecto para un carácter biótico. De esta manera, si un mejorador busca germoplasma con potencial y el carácter de interés es adaptación a condiciones de sequía, buscará directamente germoplasma cuyo sitio de recolección esté caracterizado por bajas precipitaciones. Si la característica de interés fuera de tipo biótica, como la resistencia a un patógeno, primero se debería establecer una relación entre una serie de variables ecogeográficas y la resistencia a dicho patógeno, para consecuentemente seleccionar germoplasma cuyo sitio de recolección reúna las condiciones ecogeográficas asociadas a la resistencia al patógeno.

Se pueden distinguir dos técnicas para la selección de germoplasma mediante FIGS. La primera es la de filtrado de entradas y la segunda es la técnica de calibración.

La técnica de filtrado selecciona entradas de una colección caracterizada ecogeográficamente, escogiendo aquellas que cumplen con ciertos valores o rangos para alguna de las variables por las cuales han sido caracterizadas. En ocasiones lo que se selecciona es una fracción de la distribución de una variable ecogeográfica en la colección caracterizada. Los valores, rangos o la fracción de la distribución y la variable ecogeográfica de selección los determina el investigador, curador o mejorador en base al conocimiento de la especie, de la variable ecogeográfica y del carácter de interés. Un ejemplo de la aplicación de éste método fue la selección FIGS indirecta para resistencia en trigo ante la plaga *Eurygaster integriceps* (El Bouhssini et al., 2009). Otro caso es la aplicación de FIGS de forma directa para la identificación de recursos genéticos de *Vicia faba* con adaptación a sequía (Khazaei et al., 2013).

La técnica de calibración requiere que la colección haya sido caracterizada ecogeográficamente en su totalidad o casi totalidad (entradas con coordenadas) y

que además se haya evaluado al menos parcialmente por el carácter de interés. La técnica de calibración se desarrolla en dos fases. En la primera fase, se utiliza análisis matemáticos y estadísticos para establecer la relación entre la presencia o ausencia del carácter de interés y una o varias variables ecogeográficas. Una vez probada esta relación, se realiza una predicción de presencia o ausencia del carácter de interés sobre la fracción de la colección no evaluada usando para ello la información ecogeográfica disponible para la totalidad de la colección. La predicción señalaría qué entradas serían potencialmente de interés para el mejoramiento de cultivos. La aplicación de la técnica de calibración se aprecia en los estudios de Endresen y colaboradores para cebada y trigo (Endresen, 2010; Endresen et al., 2012).

Por su propia naturaleza, la técnica de calibración se usa para FIGS de tipo indirecto, mientras la técnica de filtrado se ha usado para ambos tipos. La técnica de calibración es metodológicamente más compleja y sus resultados se asumen más precisos en la detección de entradas que poseen la característica de interés respecto a la técnica de filtrado. Sin embargo, la técnica de calibración tiene como desventaja el requisito de tener que contar con datos de evaluación parcial de la colección y que esos datos sean suficientemente fiables como para que la relación que se pueda establecer entre variable ecogeográfica y carácter de interés sea válida. Este hecho hace que su aplicación se vea restringida a un 22% de las colecciones, que es el porcentaje de colecciones nacionales que cuentan con algún tipo de evaluación por factores bióticos en 40 países, según el Segundo Reporte del Estado de los Recursos Fitogenéticos para la Alimentación y la Agricultura (FAO, 2010).

Independientemente de la forma en cómo se obtenga la subcolección FIGS, es altamente recomendable que se valide mediante ensayos de adaptación, tolerancia o resistencia, que efectivamente las entradas seleccionadas poseen el rasgo de interés por la cual fueron seleccionadas a través de las condiciones ecogeográficas de sus sitios de recolección.

## 9.2. Subcolecciones FIGS y colecciones núcleo

Una subcolección FIGS es el conjunto de entradas con potencial uso en el mejoramiento de una especie cultivada y que proviene de una selección FIGS.

Las subcolecciones FIGS, a diferencia de una colección núcleo no tiene necesariamente que ser representativas de la variabilidad de la colección original. Una subcolección FIGS convencional conlleva un sesgo (el interés del mejorador de cultivos) muy pronunciado en el momento de su selección, por lo que no cabe esperar encontrar alta representatividad .

Otra diferencia entre una colección núcleo y una subcolección FIGS es que para ésta última pueden establecerse tantas por especie como características de interés existan. En contraste, colecciones núcleo se suelen establecer una por especie.

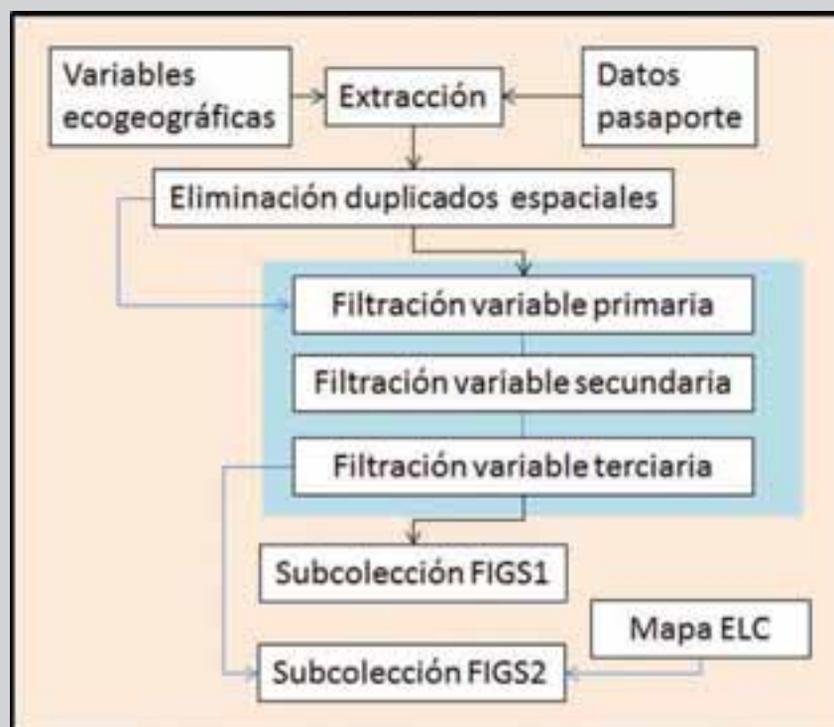
Sin embargo, como en el caso de las colecciones núcleo, establecer una o varias subcolecciones FIGS no debe poner en riesgo la conservación de las entradas no

seleccionadas. Si una colección núcleo sirve, por ejemplo, para priorizar algunas labores como la caracterización o evaluación de unas entradas en una colección cuando no se cuenta con medios para realizar la labor para la totalidad de colección, una subcolección FIGS lo que busca es potenciar la utilización del germoplasma contenido en una colección, facilitando al mejorador de cultivos, quien es su principal usuario, la búsqueda de material con potencial para ser integrado en sus programas de mejora.

### 9.3. Obtención de subcolecciones FIGS con la herramienta FIGS\_R

Con la herramienta FIGS\_R se puede obtener una subcolección FIGS mediante la técnica de filtrado. FIGS\_R permite utilizar hasta tres variables de selección de manera jerárquica, es decir, una variable primaria (de uso obligado) que ejerce el primer proceso de filtrado, una secundaria (opcional) que filtra sobre el subconjunto producto del primer filtrado y una variable terciaria (opcional y sólo funcional si se ha utilizado una variable secundaria) que filtra sobre el subconjunto producto del segundo filtrado. Como variable primaria, secundaria o terciaria, se puede escoger una de las 103 variables ecogeográficas (ver Anexos 12.1, 12.2 y 12.3) que se tienen a disposición en las herramientas CAPFITOGEN. La Fig. 31 muestra el proceso que sigue FIGS\_R para la obtención de subcolecciones FIGS.

**FIGURA 31.** Ilustración del proceso que sigue la herramienta FIGS\_R para la obtención de subcolecciones FIGS.



la forma de realizar la selección. La primera forma es la de la determinación por parte del usuario de un rango de valores que debe satisfacer una entrada para ser incluida en la subcolección FIGS. La segunda forma es la determinación de una fracción de la colección, expresada en porcentaje, cuyas entradas son las que poseen los valores más altos o más bajos en relación a la variable de selección.

FIGS\_R adopta algunos términos o definiciones del mejoramiento genético de cultivos, como intensidad de selección o diferencial de selección. El primero se usa para definir el porcentaje de la colección inicial que quedará incluido en la subcolección FIGS. El segundo hace referencia a la diferencia entre la media de la colección original y la subcolección FIGS para la o las variables de selección.

Adicionalmente, la herramienta FIGS\_R permite crear subcolecciones FIGS balanceadas ecogeográficamente. En otras palabras, si el usuario previamente ha generado un mapa ELC (a través de la herramienta ELC mapas, capítulo 4) y usa exclusivamente la segunda forma de selección (fracción de la colección), puede hacer que la herramienta: 1. Asigne categorías a cada entrada en base a la categoría del mapa ELC donde ocurre su sitio de recolección, y 2. seleccione de cada categoría la fracción de entradas con los valores más altos o bajos para la variable de selección. El balance mediante un mapa ELC tiende a producir subcolecciones FIGS con una mayor representatividad ecogeográfica aunque conservando en buena medida el potencial de utilización en programas de mejoramiento por la característica de interés.

Finalmente es importante tener en cuenta que la herramienta FIGS\_R puede tener en cuenta también información sobre disponibilidad de las entradas para ser seleccionadas. De esta manera, utiliza el mismo formato de entrada de datos (datos de pasaporte) que la herramienta ColNucleo, es decir, el formato usado para la herramienta GEOQUAL más la adición del campo "AVAILAB". Para rellenar el campo "AVAILAB" aplican los mismos criterios descritos en el apartado 8.4.

## 9.4. Utilización de la herramienta FIGS\_R

Una vez instalada las herramientas CAPFITOGEN y seleccionada la herramienta FIGS\_R, será necesario definir una serie de parámetros para que la programación R funcione correctamente.

### 9.4.1 Parámetros iniciales definidos por el usuario

#### 9.4.1.1 Parámetro: ruta

Explicación: Ruta donde se ha copiado o se encuentra la herramientas CAPFITOGEN.  
Nota: use / antes que \. Por ejemplo F:/, C:/CAPFITOGEN, D:/MisHerramientas/CAPFITOGEN, etc.

#### 9.4.1.2 *Parámetro: país*

Explicación: Seleccione el país en el cual fueron recolectadas la mayoría o la totalidad de las entradas que desea analizar. Si las entradas han sido recolectadas en más de un país, considere seleccionar una región, subcontinente o continente (progresivamente se añadirán estas opciones).

#### 9.4.1.3 *Parámetro: pasaporte*

Explicación: Escriba el nombre del archivo que contiene la tabla de pasaporte en formato de texto sin olvidar incluir la extensión (.txt). Por ejemplo, si el archivo se llama "tabla" debe escribir: "tabla.txt". Recuerde que este archivo debe ser guardado previamente en la carpeta "Pasaporte" que forma parte del conjunto de carpetas que componen el directorio CAPFITOGEN. Ésta tabla contiene una columna adicional denominada "AVAILAB" respecto a la tabla de pasaporte usada como modelo para otras herramientas CAPFITOGEN. Esta columna adicional indica la disponibilidad de cada entrada a ser seleccionada para una subcolección FIGS.

#### 9.4.1.4 *Parámetro: geoqual*

Explicación: Seleccione esta opción si los datos de pasaporte han sido analizados con la herramienta GEOQUAL y contienen por tanto 51 columnas (y no las 46 del modelo de tabla exclusivo para ColNucleo sin haber sido sometida a GEOQUAL). Use por tanto la tabla resultado de GEOQUAL denominada PasaporteOriginalEvaluadoGEOQUAL.txt como tabla de pasaporte en el punto anterior.

#### 9.4.1.5 *Parámetro: totalqual*

Explicación: Si su tabla de pasaporte proviene de GEOQUAL y desea considerar un mínimo de calidad para que sus datos sean incluidos en el análisis, determine el valor de TOTALQUAL100 a usar como umbral. Se admiten valores de 0 (calidad nula) a 100 (máxima calidad).

#### 9.4.1.6 *Parámetro: controlelc*

Explicación: Indique si desea usar un mapa ELC previamente creado, para determinar la distribución de las entradas que componen la subcolección FIGS sobre las diferentes categorías del mapa. Para usuarios avanzados, esta opción habilita la obtención de una subcolección FIGS adicional, en la que selección de las entradas se hace sobre cada categoría ELC. Se requiere para ello usar métodos de selección sobre fracciones de la distribución para todas las variables consideradas.

#### 9.4.1.7 Parámetro: *mapaelc*

Explicación: Indique el nombre del archivo que contiene el mapa ELC (producto de la aplicación de la herramienta ELCmapas), el cual debe encontrarse en la carpeta ELCmapas, la cual es una de las carpetas que componen el directorio CAPFITOGEN. Este mapa debe estar en formato DIVA-GIS (extensión .grd tal y como lo produce ELCmapas) y debe escribirse con la extensión. Así, si el nombre del mapa es "mapa\_elc\_spain", ud. debe escribir "mapa\_elc\_spain.grd".

#### 9.4.1.8 Parámetro: *statelc*

Explicación: Indique el nombre del archivo que contiene la tabla de estadísticas descriptivas del mapa ELC producido con la herramienta ELCmapas (ésta herramienta suele nombrar a éste archivo como "Estadist\_ELC\_" más el nombre del país o región). Al igual que el mapa ELC, este archivo debe estar situado en la carpeta ELCmapas. De igual forma el nombre debe escribirse junto con la extensión, en este caso al ser una tabla, la extensión es txt. Así, si el archivo se llama "Estadist\_ELC\_spain" se debe escribir "Estadist\_ELC\_spain.txt".

#### 9.4.1.9 Parámetro: *distdup*

Explicación: Determine el valor de distancia (en km) por debajo del cual Ud. consideraría que dos sitios de recolección en realidad representan una misma población (duplicado geográfico). El valor cero (valor mínimo y por defecto) da menos prioridad a la hora de la selección a entradas con idénticas coordenadas. En la medida que el valor de distancia aquí estipulado aumente, un mayor número de entradas se considerarían duplicados geográficos.

#### 9.4.1.10 Parámetro: *availab*

Explicación: Seleccione esta opción si desea utilizar la columna sobre disponibilidad de entradas para ser seleccionadas como integrantes de la subcolección FIGS. Esto implica dar prioridad a entradas disponibles pero no restringir la posible consideración de no disponibles. Recuerde que para esta herramienta, la tabla de pasaporte incluye una columna denominada "AVAILAB", donde se indica la disponibilidad de las entradas de la colección original para integrar colecciones núcleo, a través de los códigos 0 (entrada no disponible), 1 (entrada disponible) y NA (sin información/no disponible). La disponibilidad es un criterio propio del curador y podría estar determinado por el número de semillas conservadas, su germinación u otros diversos factores.

#### 9.4.1.11 Parámetro: *soloavailab*

Explicación: Seleccione esta opción si desea restringir la selección de entradas con destino a la subcolección FIGS exclusivamente a las entradas designadas como disponibles (valor 1 en el campo "AVAILAB").

#### 9.4.1.12 Parámetro: *resol1*

Explicación: Seleccione el nivel de resolución que desea utilizar para la extracción de información ecogeográfica. Note que 1x1 km ofrece mayor resolución pero exige más capacidad de cómputo en relación a 5x5 km, si bien no es un aspecto tan limitante como en la herramienta ELC mapas. Las resoluciones 10x10 y 20x20 km están restringidas para países de gran extensión, subcontinentes o continentes

#### 9.4.1.13 Parámetro: *buffy*

Explicación: Marque esta opción si desea que la extracción de información ecogeográfica se haga sobre un área alrededor del sitio de recolección. Dejar sin marcar esta opción lleva a que la extracción se realice sólo para el punto que indican las coordenadas del sitio de recolección.

#### 9.4.1.14 Parámetro: *tamp*

Explicación: Especifique el radio (en metros) de un área circular alrededor del punto que indican las coordenadas del sitio de recolección, de la que se extraerá información ecogeográfica. Los valores extraídos del área circular serán promediados para obtener un único valor.

#### 9.4.1.15 Parámetro: *variab1v*

Explicación: Seleccione una (1) variable ecogeográfica primaria por la cual desea seleccionar entradas con el fin de obtener una subcolección FIGS. Si elige seleccionar entradas por una o dos variables adicionales (variable secundaria y terciaria), la variable seleccionada en este punto será con la que se realice el primer filtrado.

#### 9.4.1.16 Parámetro: *variab1rang*

Explicación: Marque esta opción, si desea seleccionar entradas respecto a la variable primaria usando un intervalo de valores, es decir, indicando un valor máximo y mínimo que determine un rango que se usará para seleccionar las entradas con destino a la subcolección FIGS.

#### 9.4.1.17 Parámetro: *variab1min*

Explicación: Especifique el valor mínimo para la variable primaria con la que determinar el rango que se usará para seleccionar entradas con destino a la subcolección FIGS.

#### 9.4.1.18 Parámetro: *variab1max*

Explicación: Especifique el valor máximo para la variable primaria con la que determinar el rango que se usará para seleccionar entradas con destino a la subcolección FIGS.

#### 9.4.1.19 Parámetro: *variab1cola*

Explicación: Marque esta opción, si desea seleccionar entradas respecto a la variable primaria usando una fracción de la distribución, es decir, un porcentaje de la colección original cuyos valores son los más altos o bajos respecto a la variable primaria.

#### 9.4.1.20 Parámetro: *variab1vpor*

Explicación: Determine la fracción de la distribución (en porcentaje) que desea seleccionar para conformar la subcolección FIGS. Valores permitidos entre 0 y 100.

#### 9.4.1.21 Parámetro: *variab1vhl*

Explicación: Seleccione la fracción de la distribución que desea seleccionar en referencia a la variable primaria.

#### 9.4.1.22 Parámetro: *variab2*

Explicación: Marque esta opción si desea utilizar una variable secundaria para seleccionar entradas con destino a una subcolección FIGS. Con valores de esta variable se realizará una selección de entradas sobre el subconjunto previamente seleccionado con la variable primaria.

#### 9.4.1.23 Parámetro: *variab2v*

Explicación: Seleccione una (1) variable ecogeográfica secundaria por la cual desea seleccionar entradas con el fin de obtener una subcolección FIGS. Puede ser la misma variable primaria.

#### 9.4.1.24 Parámetro: *variab2rang*

Explicación: Marque esta opción, si desea seleccionar entradas respecto a la variable secundaria usando un intervalo de valores, es decir, indicando un valor máximo y mínimo que determine un rango que se usará para seleccionar las entradas con destino a la subcolección FIGS.

#### 9.4.1.25 Parámetro: *variab2min*

Explicación: Especifique el valor mínimo para la variable secundaria con la que determinar el rango que se usará para seleccionar entradas con destino a la subcolección FIGS.

#### 9.4.1.26 Parámetro: *variab2max*

Explicación: Especifique el valor máximo para la variable secundaria con la que determinar el rango que se usará para seleccionar entradas con destino a la subcolección FIGS.

#### 9.4.1.27 Parámetro: *variab2cola*

Explicación: Marque esta opción, si desea seleccionar entradas respecto a la variable secundaria usando una fracción de la distribución remanente, es decir, un porcentaje del subconjunto seleccionado por la variable primaria cuyos valores son los más altos o bajos respecto a la variable secundaria.

#### 9.4.1.28 Parámetro: *variab2vpor*

Explicación: Determine la fracción de la distribución remanente (en porcentaje) que desea seleccionar para conformar la subcolección FIGS, usando la variable secundaria. Valores permitidos entre 0 y 100.

#### 9.4.1.29 Parámetro: *variab2vhl*

Explicación: Seleccione la fracción de la distribución que desea seleccionar en referencia a la variable secundaria.

#### 9.4.1.30 Parámetro: *variab3*

Explicación: Marque esta opción si desea utilizar una variable terciaria para seleccionar entradas con destino a una subcolección FIGS. Con valores de esta variable se realizará una selección de entradas sobre el subconjunto previamente seleccionado con las variables primaria y secundaria. Si no se ha determinado previamente la utilización de una variable secundaria, seleccionar una terciaria no tendrá ningún efecto en la conformación de la subcolección FIGS.

#### 9.4.1.31 Parámetro: *variab3v*

Explicación: Seleccione una (1) variable ecogeográfica terciaria por la cual desea seleccionar entradas con el fin de obtener una subcolección FIGS. Puede ser la misma variable primaria o secundaria.

#### 9.4.1.32 Parámetro: *variab3rang*

Explicación: Marque esta opción, si desea seleccionar entradas respecto a la variable terciaria usando un intervalo de valores, es decir, indicando un valor máximo y mínimo que determine un rango que se usará para seleccionar las entradas con destino a la subcolección FIGS.

#### 9.4.1.33 Parámetro: *variab3min*

Explicación: Especifique el valor mínimo para la variable terciaria con la que determinar el rango que se usará para seleccionar entradas con destino a la subcolección FIGS.

#### 9.4.1.34 Parámetro: *variab3max*

Explicación: Especifique el valor máximo para la variable terciaria con la que determinar el rango que se usará para seleccionar entradas con destino a la subcolección FIGS.

#### 9.4.1.35 Parámetro: *variab3cola*

Explicación: Marque esta opción, si desea seleccionar entradas respecto a la variable terciaria usando una fracción de la distribución remanente, es decir, un porcentaje del subconjunto seleccionado por la variable primaria y secundaria cuyos valores son los más altos o bajos respecto a la variable terciaria.

#### 9.4.1.36 Parámetro: *variab3vpor*

Explicación: Determine la fracción de la distribución remanente (en porcentaje) que desea seleccionar para conformar la subcolección FIGS, usando la variable terciaria. Valores permitidos entre 0 y 100.

#### 9.4.1.37 Parámetro: *variab3vhl*

Explicación: Seleccione la fracción de la distribución que desea seleccionar en referencia a la variable terciaria.

#### 9.4.1.38 Parámetro: *resultados*

Explicación: Introduzca la ruta de la carpeta donde desea que se guarden los resultados del análisis. Nota: use / antes que \. Por ejemplo C:/Resultados, D:/MisHerramientas/Resultados, etc.

## 9.5. Resultados de FIGS\_R

Una vez finalizado el análisis y de acuerdo a la configuración de los parámetros anteriormente detallados, FIGS\_R producirá entre tres y cinco tablas.

Cuando no se incluye un mapa ELC en el análisis (parámetro 9.4.1.6, controlelc), sólo aparecerán las siguientes tres tablas:

9.5.1 "FIGS\_regular.txt". Ésta tabla contiene la identificación de las entradas seleccionadas con destino a la subcolección FIGS (campo "ACCENUMB"), las coordenadas del sitio de recolección ("DECLATITUDE" Y "DECLONGITUDE"), el campo de disponibilidad ("AVAILAB") y a continuación tantas columnas como variables de selección usadas.

9.5.2 "FIGS\_stat\_table.txt". Corresponde a la tabla que resume las características tanto de la colección original como de la subcolección FIGS a través de estadísticas como la intensidad de selección alcanzada, media, valor máximo y mínimo y diferencial de selección para cada una de las variables de selección utilizadas.

9.5.3 "Passport\_FIGS\_R.txt". Ésta tabla corresponde a la tabla de pasaporte que el usuario introduce al análisis más un campo adicional por cada variable de selección utilizada llamado "SEL\_VAR" seguido por el número 1, 2 o 3. Estos campos señalan con el número 1 las entradas que hacen parte de la subcolección FIGS y con "NA" las que no fueron seleccionadas. Así, en el proceso selectivo con la variable primaria (definida en el parámetro 9.4.1.15) se habrían seleccionado las entradas identificadas con un "1" en el campo "SEL\_VAR1", si se ha utilizado una variable de selección secundaria, las entradas seleccionadas en el segundo proceso de filtrado aparecerán identificadas con un "1" en el campo "SEL\_VAR2", y finalmente si se ha utilizado una variable de selección terciaria, las entradas seleccionadas en el tercer proceso de filtrado aparecerán identificadas con un "1" en el campo "SEL\_VAR3".

Cuando se incluye un mapa ELC para conocer las características ecogeográficas de la subcolección FIGS generada, aparecerá una nueva tabla:

9.5.4 "FIGS\_freq\_ELCmap.txt". Esta tabla muestra valores de frecuencia, número de duplicados y número de entradas disponibles para cada categoría ecogeográfica de manera similar a la tabla de la herramienta ColNucleo descrita en el apartado 8.6.1.3 ("CoreCollect\_stats.txt"). Adicionalmente aparecen en la parte izquierda de la tabla hasta tres nuevos campos identificados con el prefijo "FIGS\_var" y a continuación los números 1, 2 o 3. De esta manera, en el campo "FIGS\_var1" aparecerá el número de entradas seleccionadas por la variable primaria que fueron recolectadas en cada categoría ELC, en "FIGS\_var2" el número de entradas seleccionadas por la variable secundaria en el segundo proceso de filtrado de cada categoría ELC y en "FIGS\_var3" el número de entradas seleccionadas por la variable terciaria en el tercer proceso de filtrado de cada categoría ELC.

Finalmente, si para las variables de selección primaria, secundaria y terciaria se ha utilizado exclusivamente la segunda forma de selección (fracción de la colección), es decir, se ha marcado la opción "variab1cola", "variab2cola" y "variab3cola" (parámetros 9.4.1.19, 9.4.1.27 y 9.4.1.35 respectivamente), los resultados incluirán una quinta tabla:

9.5.5 "FIGS\_UnderELC.txt". Corresponde a una tabla con los mismos campos descritos anteriormente para la tabla "Passport\_FIGS\_R.txt" (apartado 9.5.3) pero en este caso sólo contiene las entradas que constituyen la subcolección FIGS balanceada mediante el mapa ELC, e incluyen los campos "SEL\_VAR1", "SEL\_VAR2" y "SEL\_VAR3" que indican con el valor 1 si esas entradas hubieran sido también seleccionadas en un esquema FIGS sin el uso de mapas ELC. Adicionalmente, al costado izquierdo de la tabla aparecen hasta 3 nuevos campos, identificados como "var\_eco1", "var\_eco2" y "var\_eco3" dependiendo de cuantas variables de selección se hayan utilizado. En cada uno de estos campos aparecerán los valores correspondientes a la extracción de cada variable de selección para cada sitio de recolección ("var\_eco1" valores de la variable primaria, "var\_eco2" valores de la variable secundaria y "var\_eco3" valores de la variable terciaria).

Adicionalmente, la tabla "FIGS\_freq\_ELCmap.txt" (apartado 9.5.4) incluirá hasta tres nuevos campos en el costado izquierdo, con las denominaciones "No\_by\_var1", "No\_by\_var2" y "No\_by\_var3". En estos campos se muestra el número de entradas seleccionadas para la subcolección FIGS balanceada por el mapa ELC en cada proceso de selección, "No\_by\_var1" para el primer proceso de filtrado por la variable primaria, "No\_by\_var2" para el segundo proceso de filtrado por la variable secundaria y "No\_by\_var3" para el tercer proceso de filtrado por la variable terciaria.

## 9.6. Referencias

El Bouhssini, M. E., Street, K., Joubi, A., Ibrahim, Z., Rihawi, F. 2009. Sources of wheat resistance to Sunn pest, *Eurygaster integriceps* Puton, in Syria. Genetic Resources and Crop Evolution 56: 1065–1069.

Endresen, D.T.F. 2010. Predictive association between trait data and ecogeographic data for Nordic barley landraces. Crop Science 50: 2418-2430.

Endresen, D.T.F., Street, K., Mackay, M., Bari, A., Amri, A., De Pauw, E., Nazari, K., Yahyaoui, A. 2012. Sources of resistance to stem rust (Ug99) in bread wheat and durum wheat identified using Focused Identification of Germplasm Strategy. Crop Science 52: 764-773.

FAO 2010. The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture. Rome

Khazaei, H., Street, K., Bari, A., Mackay, M., Stoddard, F.L. 2013. The FIGS (Focused Identification of Germplasm Strategy) approach identifies traits related to drought adaptation in *Vicia faba* genetic resources. PLoS ONE 8(5): e63107. doi:10.1371/journal.pone.0063107

Mackay, M.C. 1990. Strategic planning for effective evaluation of plant germplasm. p. 21-25 En: Srivastava, J.P., Damania, A.B. (eds). Wheat genetic resources: Meeting diverse needs. John Wiley & Sons, Chichester, UK.

Mackay, M. C., Street, K. 2004. Focused identification of germplasm strategy – FIGS. p. 138-141. En: Black, C.K., Panozzo, J.F., Rebetzke, G.J. (eds). Cereals 2004. Proceedings of the 54th Australian Cereal Chemistry Conference and the 11th Wheat Breeders' Assembly, 21-24 September 2004, Canberra, Australian Capital Territory (ACT). Cereal Chemistry Division, Royal Australian Chemical Institute, Melbourne, Australia.



## 10. Errores frecuentes

En la siguiente lista se encuentran muchos de los avisos de error (texto de color rojo en la parte inferior de la interfaz), o al menos los más frecuentes, que pueden aparecer durante la ejecución de alguna de las herramientas. Si durante su experiencia de trabajo con las herramientas CAPFITOGEN le aparecen avisos de error diferentes a los expuestos o tiene alguna duda sobre el funcionamiento de las mismas, puede visitar el foro de resolución de problemas en la siguiente dirección web: <http://www.agrobiodiversidad.org/foroCAPFITOGEN/>.

### 1. Sin aviso de error pero sin resultados en la carpeta designada para ello:

Herramienta: Cualquiera pues proviene de una ubicación inadecuada del conjunto de herramientas

Solución(es): Verifique que la estructura de datos de las herramientas (carpetas y archivos incluidos en la carpeta llamada CAPFITOGEN) se encuentran dentro de una ruta sin valores atípicos o sin espacios. Por ejemplo, si Ud. ha situado las herramientas en la carpeta "Mis documentos", la ruta podría ser así: C:\Mis documentos\CAPFITOGEN. Esta ruta contiene un espacio entre las palabras "Mis" y "documentos" lo cual puede generar este error, en el que aparentemente el análisis se ha ejecutado correctamente pero la carpeta designada para guardar los resultados está vacía. Se recomienda ubicar la carpeta CAPFITOGEN directamente en la raíz del directorio de la unidad de disco (en este caso así: C:\CAPFITOGEN).

### 2. Aviso de error:

An error occurred: Error en library(package name) : there is no package called 'package name'Calls: source -> withVisible -> eval -> eval -> library

Herramienta: Cualquiera pues proviene de la instalación

Solución(es): Este error indica que algún paquete de R que requiere la herramienta no fue correctamente instalado y por esa razón R no lo encuentra. Asegúrese que la estructura de carpetas y archivos de las herramientas no se encuentra localizada en la raíz del directorio (por ejemplo en K:). Si es así, cree una carpeta (usualmente con el nombre de CAPFITOGEN) en la raíz del directorio, corte toda la estructura de carpetas y archivos y péguela en la carpeta que acaba de crear. A continuación reinstale las herramientas. Si esta opción no funciona, pruebe a instalar el paquete manualmente. Para ello:

- a. En el código del error aparece el nombre del paquete no instalado, exactamente donde en el ejemplo dice 'package name'. Con este nombre de paquete, abra la carpeta "packages" dentro de la estructura de carpetas y archivos de CAPFITOGEN y allí encontrará una serie de archivos .zip con diversos nombres. Uno de ellos corresponderá con el nombre del paquete y estará acompañado de unos números relativos a la versión. Copie el nombre del archivo (completo, incluyendo la extensión .zip).
- b. Abra R, con el ejecutable que estará alojado en la ruta: C:\rwin\bin\i386\Rgui.exe y teclee la siguiente instrucción en la "R console":

```
Install.packages("X:/CAPFITOGEN/packages/nombreachivopaquete.zip")
```

X corresponde a la letra de la unidad donde se encuentran las herramientas CAPFITOGEN (cambie esta letra con la que sea acorde a su caso) y en la parte "nombreachivopaquete.zip" pegue el nombre del archivo que copió en el paso 1. A continuación teclee "enter".

- c. El programa empezará a instalar el paquete y cuando acabe aparecerá un aviso similar a este:

```
package 'sp' successfully unpacked and MD5 sums checked
```

- d. Pruebe que el paquete quedó correctamente instalado, para ello teclee a continuación:

```
library("package name")
```

"package name" es el nombre del paquete y como aparece en el aviso de error (sin **códigos de versiones ni extensión .zip**). Posteriormente pulse "enter". Aparecerá un aviso similar a este:

```
Mensajes de aviso perdidos
```

```
package 'cluster' was built under R version 2.15.3
```

- e. Pruebe de nuevo la herramienta. Es posible que aparezca el mismo error pero para un paquete diferente. Si es así, repita la operación hasta que cesen los avisos de error.

Este tipo de errores no son usuales desde que se mejoró el sistema de instalación, pero ocasionalmente se presentan, especialmente en la versión de Windows 8.

### 3. Aviso de error:

An error occurred: Error: no se puede ubicar un vector de tamaño X.X Gb

Herramienta(s): varias

Solución(es): Es un problema relacionado con el tamaño de las matrices que tiene

que manejar R. Suele solucionarse rebajando resoluciones de los mapas. Si se presenta en ELC mapas, se recomienda cambiar el método de determinación de número óptimo de grupos o incrementar el tamaño de celda en el parámetro resol1.

También es posible que este error aparezca en GEOQUAL por un error en el contenido de las tablas de pasaporte, específicamente cuando ocurren duplicados para el campo ACCENUMB. Este campo identifica inequívocamente cada entrada y un solo duplicado puede generar este aviso de error. La solución está en revisar en la tabla que no ocurran duplicados y si ocurren, asignar números o códigos únicos para cada entrada.

#### 4. Aviso de error:

An error occurred: Error en `sample.int(m, k)` : primer argumento inválido  
Calls: source ... withVisible -> eval -> eval -> kmeans -> sample.int

Herramienta(s): ELC mapas

Solución(es): Significa que alguna variable para esa región o país es constante y al estandarizarse produce una tabla de 0 filas que luego genera el error en Kmeans (método elbow). Se soluciona deseleccionando esa variable problemática. Para identificarla, es importante saber que suele corresponder a variables de suelo especialmente en países pequeños. Por ejemplo, la variable "profundidad" suele originar este problema. También en países secos, usar variables de mínimas precipitaciones suele producir este problema

#### 5. Aviso de error:

An error occurred: Error en `clara(sdata, k, ...)` : x is not a numeric dataframe or matrix.  
Calls: source -> withVisible -> eval -> eval -> pamk -> clara

Herramienta(s): ELC mapas

Solución(es): Significa que alguna variable para esa zona es constante y al estandarizarse produce una tabla de 0 filas que luego genera el error en medoides. Igual solución que error No. 2.

#### 6. Aviso de error:

An error occurred: Error en `kmeans(edaph[, -1], centers = i)` : more cluster centers than distinct data points.  
Calls: source -> withVisible -> eval -> eval -> kmeans

Herramienta(s): ELC mapas

Solución(es): Significa que se ha indicado un número máximo de grupos menor al

número objetivo que el método elbow determina como óptimo. Repetir operación con un número menor de grupos.

## 7. Aviso de error:

An error occurred: Error: objeto 'ecogeot' no encontrado

Herramienta(s): ECOGEO

Solución(es): Se debe seleccionar la opción geophyssv si se seleccionan variables geofísicas.

## 8. Aviso de error:

An error occurred: Error en validObject(.Object) : invalid class "SpatialPoints" object: bbox should never contain infinite valuesCalls: source ... SpatialPoints -> new -> initialize -> initialize -> validObject

Herramienta(s): Representa

Solución(es): Revise el archivo de texto "process\_info.txt" de la carpeta "Error" de la estructura de carpetas y archivos de las herramientas CAPFITOGEN. En la línea inferior de ese archivo de texto encontrará **que puede aparecer un aviso diciendo "ATENCIÓN!!, error al eliminarse todos los registros FE al considerar datos de otros bancos como no faltantes"**. Esto indica que la herramienta se ha quedado sin datos de fuentes externas pues todos los aportados provienen de "bancos de germoplasma" y al indicarse que éstos no deben tomarse como faltantes, se produce un error, al no quedar datos por analizar. Elimine la opción de utilizar fuentes externas o permita que Representa tome datos de otros bancos como faltantes (apartado 6.3.1.10).

## 9. Aviso de error:

An error occurred: Error en dist(x[ss[[i]], ], method = metric, ...) : no se permiten vectores de longitud negativaCalls: source ... withVisible -> eval -> eval -> pamk -> distcritmulti -> dist

Herramienta(s): ELC mapas

Solución(es): Aparece este error cuando el país o región es muy grande, la resolución es alta (un tamaño de celda menor) y se pide a la herramienta realizar la determinación del número óptimo de agrupamientos "medoides". La primera solución aplicable es ejecutar de nuevo el análisis pero usando el método "elbow". Si a pesar de esto, se genera otro error, la recomendación es usar una resolución menor (tamaño de celda mayor).

## 10. Aviso de error:

An error occurred: Error en `merge.data.frame(as.data.frame(x), as.data.frame(y), ...)` : no se permiten vectores de longitud negativa  
Calls: source ... merge -> merge.default -> merge -> merge.data.frame

Herramienta(s): ELC mapas

Solución(es): El error persiste pues las matrices que se generan son tan grandes que el método de determinación óptima de número de agrupamientos “elbow” tampoco es capaz de manejar ese tamaño de matrices. La solución ya pasa por usar una resolución menor (un tamaño de celda mayor).

## 11. Aviso de error:

An error occurred: Error en `.checkNumericCoerce2double(obj)` : cannot retrieve coordinates from non-numeric elements  
Calls: source ... coordinates -> .local -> do.call -> .checkNumericCoerce2double

Herramienta: GEOQUAL

Solución(es): Error en la codificación de las coordenadas o en la elaboración de la tabla de pasaporte. Para el primer caso, corrija manualmente en Excel y vuelva a salvar en texto delimitado por tabulaciones. Para el segundo caso, el orden de las variables no es el indicado y por eso las columnas correspondientes a las coordenadas se encuentran desplazadas. Ceñir estrictamente el orden de las variables al formato indicado y no añada columnas ni cambie su orden.

## 12. Aviso de error:

An error occurred: Error en `readChar(con, 5L, useBytes = TRUE)` : no se puede abrir la conexión  
Calls: source -> withVisible -> eval -> eval -> load -> readChar

Herramienta: Se puede presentar en cualquier herramienta

Solución(es): Este error suele corresponder a introducción errada de parámetros. Por ejemplo, en ELC mapas si a un país como Cuba se indica a la herramienta una resolución de celda de 10x10 km, puede aparecer este error. También cuando se indica una ruta equivocada donde se encuentra las herramientas o las tablas de pasaporte, etc. Para evitar este problema, revise parámetro por parámetro, confirmando que los valores son los correctos.

## 13. Aviso de error:

An error occurred: Error en `apply(x, 2, fun2)` : dim(X) must have a positive length  
Calls: source ... extract -> .xyValues -> .xyvBuf -> lapply -> FUN -> apply

Herramienta: Se puede presentar en herramientas con función de extracción radial

Solución(es): Este error puede darse cuando el usuario solicita una extracción radial utilizando un radio demasiado pequeño (parámetro `tamp`) respecto al tamaño de celda o resolución de las variables ecogeográficas (parámetro `resol1`). Por ejemplo, solicitar una extracción radial de 1000 m usando resoluciones de celda de 10x10 km aprox (5 arc-min). Una situación así produce valores nulos de extracción que luego generan este error. Pruebe con radios de tamaño mayor, especialmente mayores al tamaño de lado de cada celda y/o intente con una resolución mayor (por ejemplo, para una extracción radial de 1000 m pasar de "Celdas 5x5 km aprox (2.5 arc-min)" a "Celdas 1x1 km aprox (30 arc-seg)" podría solucionar el problema). Si aun así el problema no se soluciona, sería recomendable usar extracciones puntuales.

#### 14. Aviso de error:

An error occurred: Error en `colnames<-`(`\*tmp\*`, value = "ACCENUMB") : el atributo 'names' [1] debe tener la misma longitud que el vector [0]Calls: source -> withVisible -> eval -> eval -> colnames<-

Herramienta: Se puede presentar en herramientas en donde el usuario ingrese datos de pasaporte

Solución(es): Este aviso de error puede producirse cuando el usuario indica una tabla de pasaporte en el parámetro "pasaporte" que no tiene el número de columnas esperado, debido a se haya eliminado accidentalmente alguna columna, o que la herramienta espera columnas adicionales no incluidas (como es el caso de ColNucleo que espera la columna adicional "AVAILAB") o que se indica en el parámetro "geoqual" que la tabla contiene las cuatro columnas adicionales producto del análisis de la herramienta GEOQUAL y en realidad no las tiene. Cerciórese del contenido de la tabla pasaporte que ingresa y acorde a ello, usar el parámetro "geoqual".

#### 15. Aviso de error:

An error occurred: Error en if (any(puntosorig\$DECLATITUDE >= 90 | puntosorig\$DECLATITUDE <= : valor ausente donde TRUE/FALSE es necesarioCalls: source -> withVisible -> eval -> eval ó

An error occurred: Error en if (any(puntosorig\$DECLONGITUDE >= 180 | puntosorig\$DECLONGITUDE <= : valor ausente donde TRUE/FALSE es necesarioCalls: source -> withVisible -> eval -> eval

Herramienta: Se puede presentar en herramientas en donde el usuario ingrese datos de pasaporte

Solución(es): Existe un error en las coordenadas de al menos una entrada, y esto puede deberse a errores mismos de la codificación de las coordenadas o a que incluso no haya valor alguno (incluyendo NA) en los campos de las coordenadas. Para solucionar el problema en el primer caso, verifique la codificación de las coordenadas en sexagesimal, que concuerden con el formato FAO/Bioversity 2012 y para decimal, que los valores para DECLATITUDE estén entre -90 y 90 y para DECLONGITUDE entre -180 y 180. Si el error se debe al segundo caso (campos vacíos), esto puede deberse a la aparición de entradas "fantasma", las cuales se forman al crearse la tabla de pasaporte en excel y tener algunas filas de más, que lamentablemente no pueden ser fácilmente identificadas porque aparecen en blanco pero que al exportarse la tabla en formato de texto, en ese momento aparecen. El sistema entiende que son entradas, porque ocupan una fila, pero no tienen datos en absoluto, incluyendo coordenadas, por lo que se genera este error.



# 11. Créditos

## 11.1 Cómo citar CAPFITOGEN

La siguiente es la referencia para citar el uso de las herramientas CAPFITOGEN o del presente manual de usuario:

Parra-Quijano, M., Torres, E., Iriondo, J.M., López, F. 2014. Manual de usuario herramientas CAPFITOGEN versión 1.2. Tratado Internacional sobre los Recursos Fitogenéticos para la Alimentación y la Agricultura, FAO, Roma.

## 11.2 Programas informáticos usados en CAPFITOGEN

El desarrollo de las herramientas CAPFITOGEN ha sido posible gracias a la financiación del Ministerio de Asuntos Exteriores y de Cooperación de España y el Tratado Internacional sobre los Recursos Fitogenéticos para la Alimentación y la Agricultura.

Las herramientas CAPFITOGEN están soportadas por R versión 2.15.2 (<http://cran.r-project.org/>).

R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>

Para la interfaz facilitada para usuarios de GEOQUAL se utilizó el Rwui (<http://sysbio.mrc-bsu.cam.ac.uk/Rwui/>).

El desarrollo del instalador y la mejora de la interfaz de visualización fueron realizados por Lextrend S.L. (<http://www.lextrend.com/>).

## 11.3 Paquetes R

Las herramientas CAPFITOGEN usan los siguientes paquetes R:

sp (Edzer Pebesma, Roger Bivand, Barry Rowlingson, Virgilio Gomez-Rubio )

raster (Robert J. Hijmans y Jacob van Etten)

maptools (Roger Bivand, Nicholas Lewin-Koh)

rgdal (Roger Bivand, Keitt Tim, Rowlingson Barry)

rgeos (Roger Bivand, Colin Rundel, Edzer Pebesma, Karl Ove Hufthammer)

RJSONIO (Duncan Temple Lang)

googleVis (Markus Gesmann, Diego de Castillo)

cluster (Martin Maechler)

modeltools (Torsten Hothorn, Friedrich Leisch, Achim Zeileis)

fpc (Christian Hennig)

dismo (Robert J. Hijmans, Steven Phillips, John Leathwick y Jane Elith)

ade4 (Daniel Chessel, Anne-Beatrice Dufour y Stephane Dray)

labdsv (David W. Roberts)

vegan (Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L.Simpson, Peter Solymos, M. Henry H. Stevens y Helene Wagner)

Estos paquetes a su vez dependen de otros paquetes para su correcto funcionamiento. Agradecemos el aporte de todos los autores relacionados.

Algunos mapas de diversidad genotípica usan parámetros derivados de los scripts AFLPdat (<http://www.nhm.uio.no/english/research/ncb/aflpdat/>) de Dorothee Ehrich.

## 11.4 Datos

La fuente de información administrativa de alta precisión proviene de:

Global Administrative Areas GADM version 2 (<http://www.gadm.org>)

Los mapas de anillos de 1 y 10 km alrededor de áreas administrativas de alta precisión (GADM) se obtuvieron a través de la función Buffer (zona de influencia) en ArcGIS 10.

El mapa de anillos de 1, 10 y 20 km de baja precisión son una modificación del mapa ESRI world countries (2011) que se encuentra bajo licencia Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.

El mapa de uso de suelo utilizado para el cálculo del parámetro SUITQUAL es Global Land Cover (GLC) 2000 y su referencia es:

Bartholome E., Belward A.S., Achard F., Bartalev S., Carmona Moreno C., Eva H., Fritz S.,

Grégoire J.-M., Mayaux P. & Stibig H.-J. (2002). Global Land Cover mapping for the year 2000

– Project status November 2002, Office for Official Publications of the European Communities,

Luxembourg EUR 20524).

La fuente de información (variables) ecogeográfica proviene de:

Worldclim (<http://www.worldclim.org>)

Hijmans, R.J.; Cameron, S.E.; Parra, J.L.; Jones, P.G. and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25:1965-1978.

Harmonized world soil database

(<http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/>)

FAO/IIASA/ISRIC/ISSCAS/JRC. 2012. Harmonized World Soil Database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Modelos digitales de elevación (MDE) del Shuttle Radar Topography Mission (SRTM)

(<http://srtm.csi.cgiar.org/>)

Jarvis, A., H.I. Reuter, A. Nelson, E. Guevara, 2008, Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database:<http://srtm.csi.cgiar.org>.

## 11.5 Metodologías

La evaluación de la calidad de la georreferenciación en datos de pasaporte es una técnica originalmente desarrollada en el proyecto SIERFE (Sistema de Información Ecogeográfica para los Recursos Fitogenéticos Españoles). El desarrollo de GEOQUAL, los mapas de caracterización ecogeográfica del territorio y el concepto de Representatividad Ecogeográfica (RE) son un aporte original de Mauricio Parra Quijano, Elena Torres Lamas y José María Iriondo Alegría.



El concepto original considerado para el desarrollo de mapas de diversidad (herramientas DIV mapas) ha sido publicado por:

van Zonneveld M, Scheldeman X, Escribano P, Viruel MA, Van Damme P, et al. (2012) Mapping Genetic Diversity of Cherimoya (*Annona cherimola* Mill.): Application of Spatial Analysis for Conservation and Use of Plant Genetic Resources. PLoS ONE 7(1): e29845. doi:10.1371/journal.pone.0029845

La herramienta FIGS\_R incorpora ideas y desarrollos alcanzados en PGR Secure (<http://www.pgrsecure.org>), un proyecto colaborativo financiado bajo el Séptimo Programa Marco (THEME KBBE 2010.1.1-03, "Characterization of biodiversity resources for wild crop relatives to improve crops by breeding). Los conceptos y desarrollos introducidos en FIGS\_R provienen del trabajo del grupo de "Caracterización Predictiva" (Task 2.2) del WP2 "Informatics", liderado por Bioversity International (<http://www.bioversityinternational.org/>). Los investigadores involucrados en estos desarrollos son: Imke Thormann, Jacob van Etten y Sonia Dias (Bioversity), José Iriondo y Luisa Rubio (Universidad Rey Juan Carlos), Shelagh Kell (University of Birmingham), Dag Endresen (GBIF), Rosa García (CRF-INIA) y Mauricio Parra Quijano (TIRFAA).

## 11.6 Otros agradecimientos

Agradecimientos a Fernando Latorre (CRF-INIA, España) por su decidido apoyo al Programa CAPFITOGEN y el desarrollo de las herramientas. Agradecimientos también por su apoyo y comentarios a otros investigadores del CRF-INIA, en particular Lucía de la Rosa, Rosa García y Luis Ayerbe.

Agradecimientos a Robert J. Hijmans por su valiosa colaboración en la resolución de dudas con ciertos paquetes de R y su generosidad al permitir distribuir la información de worldclim y GADM dentro de las herramientas CAPFITOGEN.

También agradecemos a múltiples centros de investigación y conservación de recursos fitogenéticos y a muchos de sus investigadores y curadores por sus valiosos comentarios y aportes para la mejora de las herramientas CAPFITOGEN.



## 12. Anexos

### 12.1 Variables ecogeográficas disponibles. Variables bioclimáticas.

Código	Descripción de variables	Unidad	Fuente
prec_1	Precipitación media de enero	mm	Worldclim
prec_2	Precipitación media de febrero	mm	Worldclim
prec_3	Precipitación media de marzo	mm	Worldclim
prec_4	Precipitación media de abril	mm	Worldclim
prec_5	Precipitación media de mayo	mm	Worldclim
prec_6	Precipitación media de junio	mm	Worldclim
prec_7	Precipitación media de julio	mm	Worldclim
prec_8	Precipitación media de agosto	mm	Worldclim
prec_9	Precipitación media de septiembre	mm	Worldclim
prec_10	Precipitación media de octubre	mm	Worldclim
prec_11	Precipitación media de noviembre	mm	Worldclim
prec_12	Precipitación media de diciembre	mm	Worldclim
tmean_1	Temperatura media de enero	°C	Worldclim
tmean_2	Temperatura media de febrero	°C	Worldclim
tmean_3	Temperatura media de marzo	°C	Worldclim
tmean_4	Temperatura media de abril	°C	Worldclim
tmean_5	Temperatura media de mayo	°C	Worldclim
tmean_6	Temperatura media de junio	°C	Worldclim
tmean_7	Temperatura media de julio	°C	Worldclim
tmean_8	Temperatura media de agosto	°C	Worldclim
tmean_9	Temperatura media de septiembre	°C	Worldclim
tmean_10	Temperatura media de octubre	°C	Worldclim
tmean_11	Temperatura media de noviembre	°C	Worldclim
tmean_12	Temperatura media de diciembre	°C	Worldclim
tmin_1	Temperatura mínima de enero	°C	Worldclim
tmin_2	Temperatura mínima de febrero	°C	Worldclim
tmin_3	Temperatura mínima de marzo	°C	Worldclim
tmin_4	Temperatura mínima de abril	°C	Worldclim
tmin_5	Temperatura mínima de mayo	°C	Worldclim
tmin_6	Temperatura mínima de junio	°C	Worldclim
tmin_7	Temperatura mínima de julio	°C	Worldclim
tmin_8	Temperatura mínima de agosto	°C	Worldclim

## 12.1 Continuación

Código	Descripción de variables	Unidad	Fuente
tmin_9	Temperatura mínima de septiembre	°C	Worldclim
tmin_10	Temperatura mínima de octubre	°C	Worldclim
tmin_11	Temperatura mínima de noviembre	°C	Worldclim
tmin_12	Temperatura mínima de diciembre	°C	Worldclim
tmax_1	Temperatura máxima de enero	°C	Worldclim
tmax_2	Temperatura máxima de febrero	°C	Worldclim
tmax_3	Temperatura máxima de marzo	°C	Worldclim
tmax_4	Temperatura máxima de abril	°C	Worldclim
tmax_5	Temperatura máxima de mayo	°C	Worldclim
tmax_6	Temperatura máxima de junio	°C	Worldclim
tmax_7	Temperatura máxima de julio	°C	Worldclim
tmax_8	Temperatura máxima de agosto	°C	Worldclim
tmax_9	Temperatura máxima de septiembre	°C	Worldclim
tmax_10	Temperatura máxima de octubre	°C	Worldclim
tmax_11	Temperatura máxima de noviembre	°C	Worldclim
tmax_12	Temperatura máxima de diciembre	°C	Worldclim
bio_1	Temperatura media anual	°C	Worldclim
bio_2	Rango medio de temperaturas diurnas	°C	Worldclim
bio_3	Isotermalidad (BIOCLIM2/BIOCLIM7) (*100)		Worldclim
bio_4	Estacionalidad en la temperatura (desviación estándar*100)		Worldclim
bio_5	Máxima temperatura del mes más cálido	°C	Worldclim
bio_6	Mínima temperatura del mes más frío	°C	Worldclim
bio_7	Rango de temperatura anual (BIOCLIM5 - BIOCLIM6)	°C	Worldclim
bio_8	Temperatura media del cuarto más húmedo (3 meses más lluviosos)	°C	Worldclim
bio_9	Temperatura media del cuarto más seco (3 meses más secos)	°C	Worldclim
bio_10	Temperatura media del mes más cálido (3 meses más cálidos)	°C	Worldclim
bio_11	Temperatura media del cuarto más frío (3 meses más fríos)	°C	Worldclim
bio_12	Precipitación anual	mm	Worldclim
bio_13	Precipitación del mes más húmedo	mm	Worldclim
bio_14	Precipitación del mes más seco	mm	Worldclim
bio_15	Estacionalidad en la precipitación (coeficiente de variación)	mm	Worldclim

## 12.1 Continuación

Código	Descripción de variables	Unidad	Fuente
bio_16	Precipitación del cuarto más húmedo (3 meses más lluviosos)	mm	Worldclim
bio_17	Precipitación del cuarto más seco (3 meses más secos)	mm	Worldclim
bio_18	Precipitación del cuarto más cálido (3 más cálidos)	mm	Worldclim
bio_19	Precipitación del cuarto más frío (3 más fríos)	mm	Worldclim

## 12.2 Variables ecogeográficas disponibles. Variables edáficas.

Código	Descripción de variables	Unidad	Fuente
ref_depth	Profundidad referencia de la unidad de suelo	m	HWS Database
t_gravel	Contenido de grava en suelo superficial	%vol.	HWS Database
t_sand	Contenido de arena en suelo superficial	% peso	HWS Database
t_silt	Contenido de limo en suelo superficial	% peso	HWS Database
t_clay	Contenido de arcilla en suelo superficial	% peso	HWS Database
t_ref_bulk	Densidad aparente de referencia en suelo superficial	kg/dm <sup>3</sup>	HWS Database
t_oc	Contenido de carbón orgánico en suelo superficial	% peso	HWS Database
t_ph_h2o	pH en suelo superficial en solución suelo-agua	-log(H <sup>+</sup> )	HWS Database
t_cec_clay	Capacidad de intercambio catiónico de arcilla en suelo superficial	cmol/kg	HWS Database
t_cec_soil	Capacidad de intercambio catiónico en suelo superficial (general)	cmol/kg	HWS Database
t_bs	Saturación de bases en suelo superficial.	%	HWS Database
t_teb	Bases intercambiables totales en suelo superficial	cmol/kg	HWS Database
t_caco3	Carbonato de calcio en suelo superficial	% peso	HWS Database
t_caso4	Yesos en suelo superficial	% peso	HWS Database
t_esp	Sodicidad en suelo superficial	%	HWS Database
t_ece	Salinidad en suelo superficial	dS/m	HWS Database
s_gravel	Contenido de grava en subsuelo	%vol	HWS Database

## 12.2 Continuación

Código	Descripción de variables	Unidad	Fuente
s_sand	Contenido de arena en subsuelo	% peso	HWS Database
s_silt	Contenido de limo en subsuelo	% peso	HWS Database
s_clay	Contenido de arcilla en subsuelo	% peso	HWS Database
s_ref_bulk	Densidad aparente de referencia en subsuelo	kg/dm <sup>3</sup>	HWS Database
s_oc	Contenido de carbón orgánico en subsuelo	% peso	HWS Database
s_ph_h2o	pH en subsuelo en solución suelo-agua	-log(H <sup>+</sup> )	HWS Database
s_cec_clay	Capacidad de intercambio catiónico de arcilla en subsuelo	cmol/kg	HWS Database
s_cec_soil	Capacidad de intercambio catiónico en subsuelo (general)	cmol/kg	HWS Database
s_bs	Saturación de bases en subsuelo.	%	HWS Database
s_teb	Bases intercambiables totales en subsuelo	cmol/kg	HWS Database
s_caco3	Carbonato de calcio en subsuelo	% peso	HWS Database
s_caso4	Yesos en subsuelo	% peso	HWS Database
s_esp	Sodicidad en subsuelo	%	HWS Database
s_ece	Salinidad en subsuelo	dS/m	HWS Database

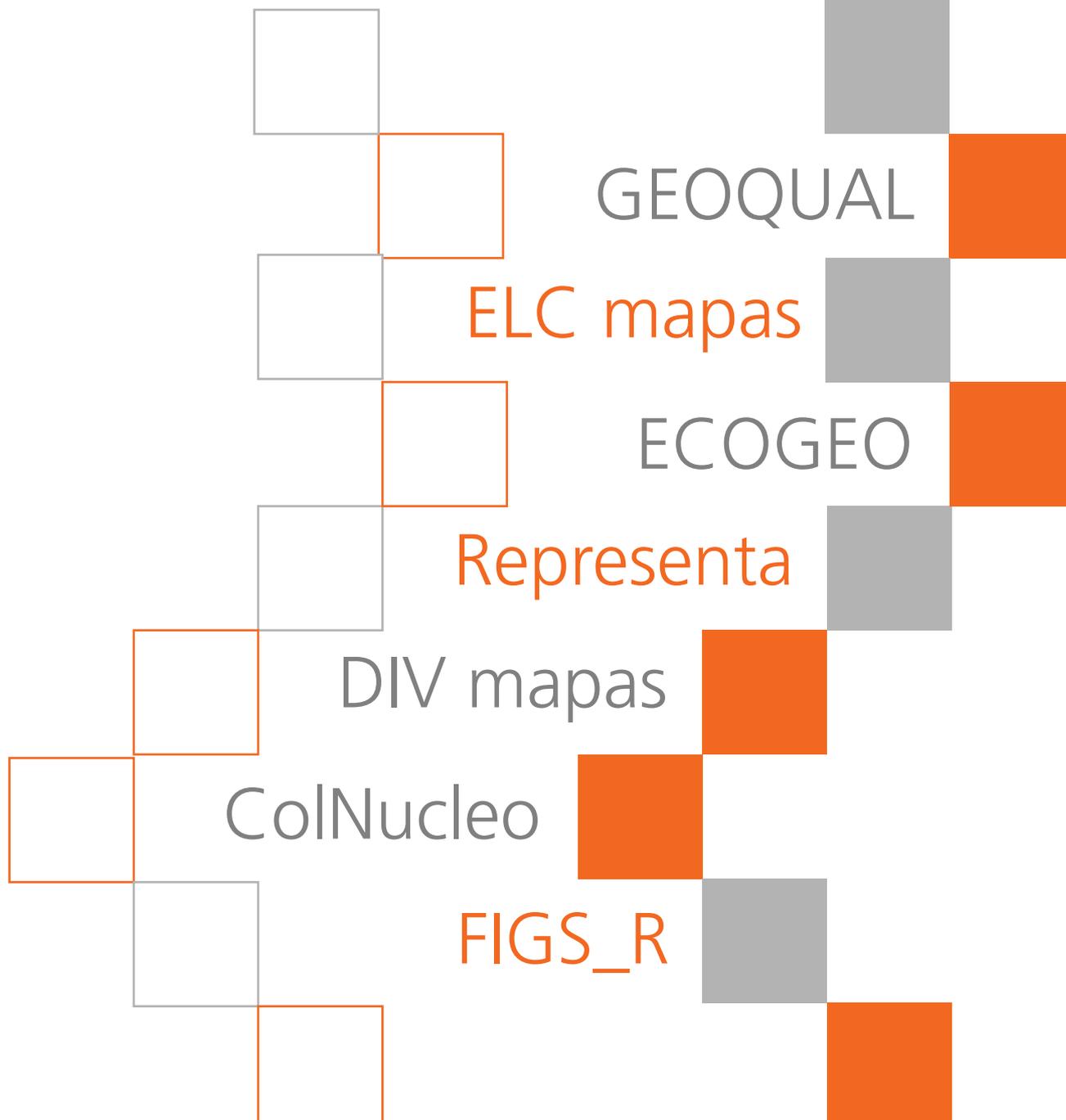
## 12.3 Variables ecogeográficas disponibles. Variables geofísicas.

Código	Descripción de variables	Unidad	Fuente
alt	Elevación. Metros sobre el nivel del mar	m	Worldclim
slope	Pendiente (en grados) de la superficie del terreno	°	SRTM MDE
aspect	Orientación (en grados) de la superficie del terreno	°	SRTM MDE
northness	Norticidad. 1 si la orientación tiende al norte, -1 al sur		SRTM MDE
eastness	Esticidad. 1 si la orientación tiende al este, -1 oeste		SRTM MDE
POINT_X	Longitud	°	
POINT_Y	Latitud	°	

Nota: Las direcciones web de las fuentes de información (worldclim, SRTM MDE y HWS Database) aparecen en el capítulo 11 (Créditos).

## 12.4 Explicación de las columnas adicionales en la tabla de resultados "tabla\_de\_analisisGEOQUAL.txt".

Variable	Explicación
globlandc	Valor extraído de GLC 2000(Global Land Cover 2000).
DISTOLAND	Anillo de distancia en el que caen las coordenadas. (0= tierra, 1=1km, 10=10km, etc.).
SUITQUAL	Parámetro SUITQUAL (valores de 0 a 20).
ID_0	Valor extraído de GADM que identifica el polígono de país.
ISO	Valor extraído de GADM que se compara con ORIGCTY.
NAME_0	Valor extraído de GADM que corresponde a la denominación completa de país.
ID_1	Valor extraído de GADM que identifica al polígono del nivel NAME_1.
NAME_1	Valor extraído de GADM que se compara con ADM1.
VARNAME_1	Valor extraído de GADM para nombres alternativos a NAME_1.
ENGTYPE_1	Valor extraído de GADM que define el tipo de administración que representa NAME_1 .
ID_2	Valor extraído de GADM que identifica al polígono del nivel NAME_2.
NAME_2	Valor extraído de GADM que se compara con ADM2.
VARNAME_2	Valor extraído de GADM para nombres alternativos a NAME_2.
ENGTYPE_2	Valor extraído de GADM que define el tipo de administración que representa NAME_2 .
ID_3	Valor extraído de GADM que identifica al polígono del nivel NAME_3.
NAME_3	Valor extraído de GADM que se compara con ADM3.
VARNAME_3	Valor extraído de GADM para nombres alternativos a NAME_3.
ENGTYPE_3	Valor extraído de GADM que define el tipo de administración que representa NAME_3.
ID_4	Valor extraído de GADM que identifica al polígono del nivel NAME_4.
NAME_4	Valor extraído de GADM que se compara con ADM4.
VARNAME_4	Valor extraído de GADM para nombres alternativos a NAME_4.



Con la financiación y la colaboración de:



Tratado Internacional sobre los Recursos Fitogenéticos  
para la Alimentación y la Agricultura

Teléfono: (+39) 06 5705 6343 • Fax (+39) 06 570 56347 • [capfitogen@fao.org](mailto:capfitogen@fao.org)

<http://www.planttreaty.org/capfitogen>

ISBN 978-92-5-308255-1



9 789253 082551

I3700S/1/03.14