

UPOV/INF/17/2 Draft 6**Original:** Inglés**Fecha:** 10 de junio de 2021

para examen por correspondencia

**PROYECTO
(REVISIÓN)**

DIRECTRICES PARA LOS PERFILES DE ADN: SELECCIÓN DE MARCADORES MOLECULARES Y CREACIÓN DE UNA BASE DE DATOS (“DIRECTRICES BMT”)*Documento preparado por la Oficina de la Unión**para su examen por**el Comité Técnico, el Comité Administrativo y Jurídico y el Consejo en 2021**Descargo de responsabilidad: el presente documento no constituye un documento de política u orientación de la UPOV*

Nota sobre el presente proyecto

Se indica ~~tachado~~ (sombreado en gris) el texto que se propone suprimir del documento [UPOV/INF/17/1](#).

Se indica subrayado (sombreado en gris) el texto que se propone insertar en documento [UPOV/INF/17/1](#).

ÍNDICE

A.	INTRODUCCIÓN	3
B.	PRINCIPIOS GENERALES	3
1.	Selección de una metodología de marcadores moleculares	3
2.	Selección de marcadores moleculares	4
2.1	<i>Criterios generales</i>	4
2.2	<i>Criterios para tipos específicos de marcadores moleculares</i>	4
3.	Acceso a la tecnología	5
4.	Material a analizar	6
4.1	<i>Fuente del material vegetal</i>	6
4.2	<i>Tipo de material vegetal</i>	6
4.3	<i>Tamaño de la muestra</i>	6
4.4	<i>Muestra de referencia de ADN</i>	6
5.	Estandarización de los protocolos analíticos	7
5.1	<i>Introducción</i>	7
5.2	<i>Criterios de calidad</i>	7
5.3	<i>Fase de evaluación</i>	7
5.4	<i>Evaluación de los datos moleculares</i>	8
6.	Bases de datos	8
6.1	<i>Tipo de base de datos</i>	8
6.2	<i>Modelo de base de datos</i>	9
6.3	<i>Diccionario de datos</i>	9
6.4	<i>Relación de los cuadros</i>	10
6.5	<i>Transferencia de datos a la base de datos</i>	11
6.6	<i>Acceso a los datos y titularidad</i>	12
6.7	<i>Análisis de los datos</i>	12
6.8	<i>Validación de la base de datos</i>	12
7.	Resumen	12
	GLOSARIO	13
	Microsatélites o Secuencias Simples Repetidas (SSR)	13
	Polimorfismos de Nucleótido Único (SNP)	13
	Secuencias Polimórficas Amplificadas y Cortadas (CAPS)	13
	Regiones Amplificadas Caracterizadas por Secuencia (SCARs)	14
	Pig-tail	14
	Alelo nulo	14
	Bandas Stutter	14
A.	INTRODUCCIÓN	3
B.	PRINCIPIOS GENERALES	3
1.	Selección de marcadores moleculares	4
1.1	<i>Conjuntos de variedades para el proceso de selección</i>	4
1.2	<i>Marcadores moleculares: consideraciones respecto del rendimiento</i>	5
2.	Selección del método de detección	6
2.1	<i>Métodos de determinación del perfil de ADN: consideraciones generales</i>	6
2.2	<i>Acceso a la tecnología</i>	7
3.	Validación y armonización de un conjunto de marcadores y el método de detección	7
3.1	<i>Validación y armonización: consideraciones generales</i>	7
3.2	<i>Consideraciones respecto del rendimiento: validación de los marcadores y los métodos</i>	7
3.3	<i>Consideraciones respecto de la coherencia</i>	7

4.	Creación de una base de datos específica para una especie.....	8
4.1	Recomendaciones para diseñar bases de datos.....	8
4.2	Requisitos del material vegetal.....	9
4.3	Tratamiento de los datos de secuencias.....	11
4.4	Tipo de base de datos.....	11
4.5	Modelo de base de datos.....	11
4.6	Diccionario de datos.....	12
4.7	Acceso a los datos y titularidad.....	14
5.	Análisis de Intercambio de los datos.....	14
5.1	Situaciones de intercambio de datos.....	14
5.2	Métodos de intercambio de datos.....	14
6.	Resumen.....	14
C.	LISTA DE SIGLAS.....	17
	<u>ANEXO SITUACIONES DE INTERCAMBIO DE DATOS Y MÉTODOS DE TRANSMISIÓN</u>	

A. INTRODUCCIÓN

El presente documento (Directrices BMT) tiene por finalidad ofrecer orientaciones para la elaboración de metodologías armonizadas destinadas a sobre los principios armonizados para el uso de marcadores moleculares con objeto de generar datos moleculares de gran calidad para diversas aplicaciones. En el presente documento solo se contemplan los marcadores moleculares de ADN.

Las Directrices BMT también tienen el propósito de abordar la creación de bases de datos de perfiles moleculares de variedades vegetales, obtenidos posiblemente en laboratorios diferentes y mediante el empleo de tecnologías también diferentes. Además, el objetivo del presente documento es lograr un alto grado de calidad de ~~los~~ marcadores y estimular el deseo de obtener datos reproducibles con la ayuda de esos marcadores aun cuando se den situaciones en las que los equipos y los reactivos químicos sean diferentes. Es preciso tomar precauciones específicas para asegurar la calidad de las entradas que se incluyen en la base de datos.

B. PRINCIPIOS GENERALES

Para determinar el perfil de ADN de una variedad vegetal, se requiere un conjunto de marcadores moleculares y un método para detectarlos. Dos conjuntos distintos de marcadores moleculares detectados mediante el mismo método darán lugar a dos perfiles diferentes de ADN para una misma variedad. Por el contrario, es previsible que empleando dos métodos distintos para detectar los alelos específicos de un determinado conjunto de marcadores moleculares se obtengan perfiles de ADN idénticos. No es necesario estandarizar el método de detección ni la tecnología, siempre que el rendimiento cumpla los criterios de calidad y los perfiles de ADN obtenidos sean coherentes. Sea cual sea la tecnología utilizada para detectar conjuntos definidos de marcadores, el genotipo de una variedad determinada no debe verse afectado.

Los conjuntos de marcadores moleculares, los métodos de detección de marcadores y el proceso subsiguiente de desarrollo de la base de datos pueden subdividirse en cinco fases:

1. Selección de los marcadores moleculares
2. Selección del método de detección
3. Validación y armonización del método de detección
4. Creación de la base de datos
5. Intercambio de datos

En el presente documento se describen estas fases con mayor detalle. Las fases se consideran independientes del grado de desarrollo de las tecnologías de genotipado y de futuras mejoras en la secuenciación de alto rendimiento.

1. Selección de una metodología de marcadores moleculares

1.1 Los siguientes son criterios importantes para seleccionar una metodología:

- a) reproducibilidad de la producción de datos entre distintos laboratorios y plataformas de detección (tipos de equipo diferentes);
- b) repetibilidad en el tiempo;
- c) poder de discriminación;
- d) posibilidades para la creación de bases de datos; y
- e) accesibilidad de la metodología.

1.2 Dado el constante avance de la tecnología y la disponibilidad de equipos nuevos, para garantizar la viabilidad de las bases de datos es importante que la interpretación de los datos producidos sea independiente del equipo usado para su obtención. Esto es lo que sucede, por ejemplo, en el caso de los datos de secuenciación de ADN. Al principio, estos datos se obtenían empleando cebadores marcados radiactivamente y geles de secuencia, mientras que ahora pueden obtenerse mediante el uso de colorantes fluorescentes seguido de una separación en sistemas de electroforesis capilar en gel de alto rendimiento y sumamente automatizados.

1.3 A pesar de estas diferencias, los datos obtenidos por medio de las distintas técnicas son coherentes entre sí e independientes de la técnica empleada para su obtención. Esto también puede aplicarse a los datos obtenidos mediante, por ejemplo, el uso de microsátélites (secuencias simples repetidas, SSR) de ADN o de polimorfismos de nucleótido único (SNP). Esta capacidad de repetibilidad y reproducibilidad es un factor importante en la creación, el funcionamiento y la perdurabilidad de las bases de datos, y es de gran importancia en la creación de una base de datos con un mantenimiento centralizado y con información verificada procedente de fuentes diversas.

1.4 Las técnicas moleculares que pueden utilizarse para definir los perfiles de las variedades están limitadas por el requisito de que los datos han de ser repetibles, reproducibles y coherentes. Así, si bien en el ámbito de la investigación se han utilizado con éxito diversas técnicas de perfiles de ADN multilocus, en muchas de ellas la codominancia no es fácil de registrar y la reproducibilidad de los patrones de bandas complejos entre laboratorios que usan equipos diferentes puede ser un problema.

1.5 Estos factores plantean dificultades en el contexto de los perfiles de variedades. Por consiguiente, el presente documento se centra en las consideraciones y recomendaciones en torno a los usos de los SSR (microsátélites) que han sido bien definidos e investigados y, por cuanto respecta al futuro, en la información sobre secuenciación (es decir, los SNP o polimorfismos de nucleótido único). Otras técnicas basadas en la información sobre la secuencia del ADN, como las secuencias polimórficas amplificadas y cortadas (CAPS) y las regiones amplificadas caracterizadas por secuencia (SCAR) posiblemente también puedan satisfacer los criterios antes mencionados, pero su empleo en la definición de los perfiles de ADN de las variedades vegetales no se ha explorado todavía.

1. Selección de marcadores moleculares

1.1 Conjuntos de variedades para el proceso de selección

Para determinar el perfil de ADN de las variedades vegetales y crear la base de datos, los marcadores moleculares han de seleccionarse en función del objetivo. Para iniciar el proceso de selección de marcadores, se necesita un número conveniente de variedades (conjunto de desarrollo) que refleje en la mayor medida posible la diversidad observada en el grupo, el cultivo, la especie o el tipo para el que los marcadores deben ser discriminatorios. La selección se refina determinando el perfil de otras variedades (conjunto de validación) a fin de medir el rendimiento de los marcadores. Para elegir el conjunto de validación pueden aplicarse los criterios siguientes:

- a) variedades o líneas de gran semejanza genética, NIL, RIL;
- b) líneas parentales y su descendencia;
- c) variedades genéticamente cercanas pero morfológicamente distintas (por ejemplo, mutantes);
- d) algunas variedades morfológicamente cercanas pero con genealogía diferente;
- e) distintos lotes de la misma variedad;
- f) orígenes diferentes de la misma variedad.

1.2 Marcadores moleculares: consideraciones respecto del rendimiento

A continuación se indican los criterios generales para seleccionar un marcador o un conjunto de marcadores específicos; estos criterios son aplicables independientemente del uso al cual se destinen los marcadores:

a) Repetibilidad, solidez y reproducibilidad en distintos laboratorios a efectos de evaluación de los datos;

b) Posibles fuentes de marcadores moleculares

- Marcadores moleculares procedentes de recursos públicos
- Marcadores moleculares procedentes de recursos privados, evaluación y selección de chips y matrices comerciales específicos para determinadas especies
- Marcadores moleculares seleccionados a partir de datos de secuencias de reciente obtención;

c) Evitar, en la medida de lo posible, marcadores con alelos "nulos" (alelo cuyo efecto es la ausencia de un producto PCR a nivel molecular) – nuevamente, no es esencial pero sí aconsejable.

d) Ha de permitir una evaluación fácil, objetiva e irrefutable de los perfiles de los marcadores. Los marcadores con buen rendimiento son preferibles a los perfiles complejos de marcadores que son susceptibles de interpretación. Las respuestas dicotómicas claras también facilitan la armonización;

e) En general, los marcadores codominantes son preferibles a los marcadores dominantes porque su poder de discriminación es mayor;

f) Marcadores ubicados en regiones codificantes o no codificantes; y

g) El uso de marcadores moleculares es específico para una especie determinada, por lo que deben tenerse en cuenta las características de reproducción o multiplicación de esta.

Como es sabido, determinados usos pueden imponer otras consideraciones adicionales, entre ellas:

a) El número de marcadores debe estar en equilibrio con la exactitud del genotipo que se requiera en función del objetivo. El número de marcadores para obtener la resolución o el poder de discriminación necesarios depende del tipo de marcador (dominante o codominante, bialélico o multialélico), de la especie y de la calidad del rendimiento de los marcadores;

b) La información sobre el genoma y el desequilibrio de ligamiento debe ser reflejo de los objetivos. Conocer la posición física o genética de los marcadores seleccionados en el genoma no es indispensable, pero ayuda a seleccionar marcadores adecuados.

2. Selección de marcadores moleculares

2.1 Criterios generales

A continuación se indican los criterios generales para seleccionar un marcador o un conjunto de marcadores específicos; estos criterios son aplicables a todos los marcadores moleculares independientemente del uso al cual se destinen, si bien, como es sabido, determinados usos pueden imponer otros criterios adicionales:

a) nivel útil de polimorfismo;

b) repetibilidad y reproducibilidad en distintos laboratorios a efectos de evaluación de los datos;

c) distribución conocida de los marcadores a través del genoma (es decir, posición en el mapa) – aun no siendo indispensable, esta información es útil y ayuda a evitar la selección de marcadores que puedan estar ligados entre sí; y

d) evitar, en la medida de lo posible, marcadores con alelos "nulos" (alelo cuyo efecto es la ausencia de un producto PCR a nivel molecular) – nuevamente, no es esencial pero sí aconsejable.

2.2 Criterios para tipos específicos de marcadores moleculares

2.2.1 Marcadores microsatélite

2.2.1.1 El análisis de secuencias simples repetidas (SSR o microsatélites: véase el Glosario) utilizando la reacción en cadena de la polimerasa (PCR) está actualmente muy extendido y ofrece varias ventajas:

2.2.1.2 Los marcadores SSR se expresan de modo codominante y por lo general son fáciles de evaluar (registrar) y pueden mapearse sin dificultad. Dichos marcadores han sido empleados y analizados en distintos laboratorios, y en condiciones experimentales específicas pueden ser sólidos y repetibles. Además, pueden analizarse por medio de secuenciadores automáticos de ADN no radiactivos de alto rendimiento, basados ya sea en electroforesis en gel o en electroforesis capilar, siendo posible analizar varios de ellos al mismo tiempo (multiplexing).

2.2.1.3 La selección de marcadores de gran calidad es esencial para un análisis eficaz con microsatélites. Ello exige tomar en consideración, entre otros:

- a) El grado de *stutter* (la producción de una serie de una o más bandas cuyo tamaño difiere en 1 unidad de repetición).
- b) (n+1) picos; la Taq polimerasa suele agregar 1 pb al extremo de un fragmento. Esto puede evitarse mediante la utilización de cebadores "con pigtail" (véase el Glosario).
- c) El tamaño del producto amplificado.
- d) La separación efectiva entre los diversos alelos en sistemas de detección adecuados.
- e) La evaluación fiable y reproducible de los alelos en diferentes sistemas de detección.
- f) El nivel de polimorfismo entre las variedades (obsérvese que ello requiere el análisis de una cantidad significativa de variedades).
- g) Evitar posibles ligamientos.

2.2.1.4 Para poder evaluar los SSR en laboratorios distintos y por medio de equipos de detección diferentes, es de vital importancia que los alelos de referencia (es decir, conjuntos de variedades) se definan y se incluyan en todos los análisis. Estos alelos de referencia son necesarios porque los marcadores de peso molecular se comportan de modo diferente entre los distintos sistemas de detección generalmente disponibles y, por tanto, no son indicados para la identificación de los alelos.

2.2.1.5 Los cebadores utilizados en cada laboratorio deben ser sintetizados por un proveedor garantizado, para reducir la posibilidad de obtener perfiles de ADN diferentes debido a la utilización de cebadores sintetizados por distintas fuentes.

2.2.2 Polimorfismos de nucleótido único (SNP)

Los polimorfismos de nucleótido único (SNP: véase el Glosario) pueden detectarse por secuenciación del ADN, una técnica habitual que por lo general muestra elevados niveles de repetibilidad en el tiempo y reproducibilidad entre laboratorios. Sin embargo, la detección de los SNP específicos puede efectuarse por diversas técnicas, muchas de las cuales no son todavía rutinarias. Por su propia naturaleza, en las plantas diploides, los SNP sólo tienen dos estados alélicos, si bien en las poliploides esto puede variar por efecto de la dosificación. La estructura simple de los SNP hace que su evaluación sea relativamente sencilla y fiable. También significa que, para definir el perfil de un genotipo determinado de forma eficaz y eficiente, será preciso analizar una gran cantidad de marcadores, ya sea individualmente o por combinación de varios marcadores.

2. Selección del método de detección

2.1 Métodos de determinación del perfil de ADN: consideraciones generales

2.1.1 A continuación se señalan consideraciones importantes a la hora de seleccionar métodos de determinación del perfil de ADN:

- a) reproducibilidad de la producción de datos con y entre distintos laboratorios y plataformas de detección (tipos de equipo diferentes);
- b) repetibilidad en el tiempo;
- c) poder de discriminación;
- d) tiempo y mano de obra;
- e) consistencia del rendimiento en cuanto a tiempo y condiciones (sensibilidad a leves variaciones en el protocolo o las condiciones);
- f) flexibilidad del método, posibilidad de emplear distinto número de muestras o de marcadores;

- g) la interpretación de los datos obtenidos no depende del equipo utilizado;
- h) viabilidad de las bases de datos;
- i) accesibilidad de la metodología;
- j) no depender de un aparato, unos reactivos químicos, un proveedor, unos socios o unos productos determinados;
- k) susceptible de automatización;
- l) apto para la multiplexación; y
- m) rentable (existe un equilibrio entre los costos, el número de muestras y el número de marcadores).

3.2.2 Acceso a la tecnología

Algunos marcadores moleculares y materiales están a libre disposición del público. Pero debido a la gran inversión que probablemente deberá hacerse para obtener, ~~por ejemplo,~~ marcadores SSR de gran calidad, los marcadores y otros métodos o materiales pueden estar amparados por derechos de propiedad intelectual. La UPOV ha elaborado directrices, que han de seguirse, para la utilización de los productos y las metodologías que son objeto de derechos de propiedad intelectual; ~~a efectos del presente documento han de seguirse~~. Es recomendable abordar todos los aspectos concernientes a los derechos de propiedad intelectual al principio de cualquier trabajo de desarrollo.

3. Validación y armonización de un conjunto de marcadores y el método de detección

3.1 Validación y armonización: consideraciones generales

Los marcadores moleculares y los métodos de detección han de ser sólidos y dar lugar a perfiles de ADN coherentes. El rendimiento de los marcadores moleculares y los métodos de genotipado se evalúa mediante un proceso de validación. Cuando se trata de bases de datos compartidas, el proceso de armonización incluye la evaluación de la coherencia de los perfiles de ADN en distintos laboratorios utilizando equipos y reactivos químicos diferentes. El uso de marcadores y métodos validados permitirá obtener resultados armonizados.

3.2 Consideraciones respecto del rendimiento: validación de los marcadores y los métodos

El conjunto de marcadores seleccionado debe ser apto para cumplir su función. Ha de medirse la exactitud. Para determinar la idoneidad de un método y un conjunto de marcadores de ADN, deben tenerse en cuenta varios aspectos:

- a) la capacidad de discriminación o informativa;
- b) la repetibilidad: un mismo operador obtiene resultados idénticos en un ensayo con el mismo método e idénticas muestras de ensayo, en el mismo laboratorio y con el mismo equipo, en intervalos breves de tiempo;
- c) la reproducibilidad: distintos operadores obtienen resultados idénticos en un ensayo con el mismo método e idénticas muestras de ensayo, en el mismo o en distintos laboratorios y con equipos diferentes;
- d) la solidez: es una medida de su capacidad para no resultar afectado por desviaciones pequeñas pero deliberadas respecto de las condiciones experimentales descritas en los parámetros del procedimiento y proporciona una indicación de su fiabilidad durante el uso normal; y
- e) la tasa de error.

Las definiciones de las características de rendimiento están basadas en la norma ISO 16 577:2016

3.3 Consideraciones respecto de la coherencia

Para obtener resultados coherentes, el proceso de armonización de los marcadores y los métodos en distintos laboratorios, cuando se trata de una base de datos compartida (ensayo interlaboratorios (*ring test*)) ha de contemplar lo siguiente:

- a) La utilización, en todos los laboratorios, de una colección definida de variedades que represente una gran diversidad de alelos, como referencia para evaluar la coherencia entre ellos;
- b) La inclusión de duplicados, submuestras, plantas de una variedad para verificar la coherencia de los perfiles de ADN y calcular la tasa de error en distintos laboratorios;

- c) Acuerdos sobre la evaluación de los datos moleculares. La necesidad de elaborar un protocolo para valorar alelos o bandas en distintos laboratorios dependerá del tipo de marcador utilizado (por ejemplo, resulta indispensable si se utilizan marcadores SSR). Dicho protocolo deberá contemplar el modo de valorar los elementos siguientes:
- i. alelos raros (alelos en un locus determinado que aparecen con una frecuencia que está por debajo de un umbral acordado (comúnmente, entre el 5 y el 10%) para una población);
 - ii. alelos nulos (alelos cuyo efecto es la ausencia de un producto PCR a nivel molecular);
 - iii. Bandas “tenues” (bandas con intensidad inferior a un umbral de detección acordado, establecido ya sea empírica o automáticamente, y cuya valoración puede ser cuestionada);
 - iv. datos ausentes (cualquier locus para el cual, por alguna razón, no hay datos registrados en una o más variedades); y
 - v. bandas monomórficas o alelos no informativos (alelos o bandas que aparecen en todas las variedades analizadas, es decir, que no son polimórficos en una determinada colección de variedades).

4. Creación de una base de datos específica para una especie

Los datos que se almacenan en una base de datos y el modo de almacenarlos deben ser un reflejo de su proceso de obtención. Por lo tanto, al crear la base de datos deben considerarse distintos grados de tratamiento de estos (es decir, datos brutos, datos de secuencias...). En la base de datos han de almacenarse los resultados finales, por ejemplo, el perfil de ADN y su proceso de obtención, indicando la descripción del método de laboratorio y los pasos computacionales.

4.1 Recomendaciones para diseñar bases de datos

A la hora de diseñar bases de datos han de tenerse en cuenta los aspectos siguientes:

- a) La arquitectura de la base de datos ha de ser flexible; por ejemplo, se deben poder almacenar archivos planos y archivos comprimidos.
- b) Se requieren cuadros y entradas independientes para el trabajo experimental de laboratorio, el tratamiento de los datos y la valoración de los alelos.
- c) Debe almacenarse información a distintos niveles, por ejemplo, la valoración de los alelos y las reglas de interpretación que sustentan la decisión, así como los enlaces a los datos brutos (archivos TIFF, archivos BAM) que se generaron.
- d) Para los datos de secuenciación, se empleará la versión 4.2 o una superior del formato estándar VCF o BCF. En las entradas de las cabeceras debe figurar el nombre y la versión de los distintos *scripts* utilizados para mapear las lecturas de secuencias, filtrar las lecturas e identificar y filtrar las variantes, de manera que un bioinformático pueda repetir el análisis.
- e) Si se trata de réplicas cuyos perfiles de ADN no coinciden, el registro ha de señalarse o descartarse, si procede. Las reglas aplicadas en estos casos han de quedar documentadas en un depósito de códigos de dominio público, al que se hará referencia en el archivo VCF. En el caso de las variedades heterogéneas, también pueden utilizarse frecuencias.
- f) Validación de los datos de los archivos VCF o BCF conforme a las especificaciones pertinentes.
- g) Datos de fácil intercambio (por ejemplo, API).

4.2.2 Material a analizar Requisitos del material vegetal

Deben tenerse en cuenta ~~Las cuestiones principales en relación con el material~~ la fuente, el tipo de material y cuántas muestras ~~se precisa~~ se han de ~~analizar~~ almacenar y compartir en la base de datos.

4.2.1 Fuente del material vegetal

El material vegetal a analizar deberá ser una muestra auténtica y representativa de la variedad y, a ser posible, habrá sido obtenida a partir de la muestra de la variedad examinada a los efectos de los Derechos de Obtentor o del registro oficial. Para usar ~~estas muestras de material presentado para el examen a efectos de los Derechos de Obtentor o del registro oficial~~ se requiere el permiso de la autoridad pertinente, bien sea el obtentor o el conservador, según corresponda. Se debería poder identificar el origen del material vegetal del que se tomaron las muestras por si posteriormente se demostrara que algunas de ellas no eran representativas de la variedad.

4.2.2 Tipo de material vegetal

El tipo de material vegetal a muestrear para la extracción del ADN y el procedimiento empleado para el muestreo dependerán en gran medida del cultivo o la especie vegetal de que se trate. Por ejemplo, en el caso de las variedades de propagación por semilla pueden utilizarse éstas como fuente del ADN, mientras que en el caso de las variedades de multiplicación vegetativa el ADN puede extraerse del material foliar. Sea cual fuere la fuente del material, el método empleado para el muestreo y la extracción del ADN debería estar ~~normalizado y~~ documentado. Además, es menester verificar mediante análisis de ADN la coherencia de los resultados obtenidos por los métodos de muestreo y extracción.

4.2.3 Tamaño y tipo de la muestra (muestras en bloque o individuales)

Es indispensable que las muestras tomadas para el análisis sean representativas de la variedad. ~~A este respecto d~~ Deberán tomarse en consideración las particularidades de su reproducción sexuada o su multiplicación vegetativa (véase la Introducción General). ~~El tamaño de la muestra deberá fijarse teniendo en cuenta los procedimientos estadísticos adecuados.~~

4.2.4 Muestra de referencia de ADN

~~Es aconsejable~~ Se puede crear una colección de ~~muestras~~ de referencia de ADN con el material vegetal muestreado ~~conforme a las secciones 4.1, 4.2 y 4.3. Poder almacenar las muestras de referencia de ADN y suministrarlas a otros laboratorios representa una ventaja. El método de muestreo debe ajustarse a los procedimientos recomendados y se deben establecer criterios de calidad para la extracción de ADN. Es necesario documentar tanto el muestreo como la extracción.~~

Las muestras de ADN deberían almacenarse de forma tal que se evite su degradación (por ejemplo, a $-80\text{ }^{\circ}\text{C}$). La transferencia de las muestras de referencia de ADN se describe en la sección 1 del documento TGP/5.

5. Estandarización de los protocolos analíticos

5.1 Introducción

~~El presente documento no pretende suministrar protocolos técnicos pormenorizados para la producción de perfiles de ADN de variedades. En principio se puede utilizar cualquier metodología analítica apropiada, pero es importante validar la metodología por el conducto adecuado. La validación puede realizarse a través de algún método reconocido a nivel internacional o bien elaborando un enfoque orientado a los resultados. Cualquiera sea el caso, el presente documento ofrece algunas consideraciones útiles de carácter general.~~

~~El método a utilizar para la composición de genotipos y la creación de bases de datos ha de ser técnicamente sencillo de aplicar, fiable y sólido, y ha de permitir la evaluación fácil e irrefutable de los perfiles de los marcadores en laboratorios diferentes. Ello requiere un cierto grado de estandarización, por ejemplo, en relación con la selección de los marcadores, los alelos de referencia, y la valoración de alelos (alelos calling/scoring).~~

5.2 Criterios de calidad

5.2.1 Es importante considerar criterios de calidad en relación, por ejemplo, con:

- a) la calidad del ADN;
- b) los métodos de extracción de ADN;
- c) las secuencias de cebadores utilizadas;
- d) la polimerasa a emplear en las metodologías basadas en PCR;
- e) para las metodologías basadas en PCR, la cantidad o concentración de cada componente PCR y de los demás componentes; y
- f) las condiciones del ciclo de PCR.

5.2.2 La metodología debe enunciarse de forma detallada en un protocolo.

5.3 Fase de evaluación

5.3.1 Introducción

Para seleccionar marcadores adecuados y producir protocolos de laboratorio aceptables para una especie dada, es aconsejable una fase de evaluación preliminar en la que participe más de un laboratorio (es decir, un método de validación reconocido internacionalmente; por ejemplo, un *ring test* de conformidad con las normas acordadas a nivel internacional. Esta fase tendrá como principal objetivo seleccionar un conjunto de marcadores y normalmente incluirá la evaluación de los marcadores existentes, ya sea publicados o bien disponibles por otros medios. El número de marcadores a evaluar es variable y dependerá de las posibilidades que ofrezcan las distintas especies. Los marcadores deben proceder de fuentes fiables (por ejemplo, de publicaciones revisadas por homólogos) y obtenerse a través de proveedores garantizados. La decisión final acerca del número a evaluar será una ponderación entre los costos y la necesidad de producir al final del proceso un conjunto satisfactorio de marcadores acordados. El objetivo es producir un conjunto satisfactorio de marcadores que pueda ser analizado con seguridad y que pueda ser analizado, evaluado y registrado de forma reproducible en laboratorios diferentes, utilizando potencialmente distintos tipos de equipos y distintas fuentes de reactivos químicos, etc.

5.3.2 Elección de variedades

Como base de la fase de evaluación se elegirá un número conveniente de variedades, en función de la variabilidad genética dentro de la especie y el tipo de variedad en cuestión. La selección de variedades debería reflejar un nivel apropiado de diversidad y, si fuera posible, deberían incluir algunas variedades estrechamente relacionadas y algunas otras morfológicamente similares, para que sea factible evaluar el grado de discriminación en tales casos.

5.3.3 Interpretación de los resultados

En la siguiente etapa de evaluación se incluirá, a ser posible, un método de validación reconocido internacionalmente para valorar la metodología de forma global y objetiva. Cualquier marcador que en esta fase de evaluación ocasione dificultades en cualquiera de los laboratorios participantes deberá ser rechazado para su uso ulterior. Si tenemos en cuenta que la mayoría de los errores en los análisis de grandes colecciones de variedades proceden de errores de valoración, las bases de datos deberían ser creadas a partir de muestras duplicadas (p. ej., distintas submuestras de semilla de la misma variedad), analizadas por más de un laboratorio. El hecho de poder intercambiar las submuestras (o el ADN extraído de ellas) en la eventualidad de que se produzca alguna discrepancia, hace que este planteamiento sea muy eficaz para poner de relieve los errores de muestreo o aquellos debidos a la heterogeneidad dentro de las propias muestras, y para eliminar los posibles artefactos de laboratorio.

5.4 Evaluación de los datos moleculares

Paralelamente a la fase de evaluación habrá de prepararse un protocolo de evaluación de alelos/bandas. Dicho protocolo deberá contemplar el modo de evaluar los elementos siguientes:

- a) Alelos raros (alelos en un locus determinado que aparecen con una frecuencia que está por debajo de un umbral acordado (comúnmente, entre el 5 y el 10%) para una población);
- b) Alelos nulos (alelos cuyo efecto es la ausencia de un producto PCR a nivel molecular);
- c) Bandas "tenues" (bandas con intensidad inferior a un umbral de detección acordado, establecido ya sea empírica o automáticamente, y cuya valoración puede ser cuestionada);
- d) Datos ausentes (cualquier locus para el cual por cualquier razón no hay datos registrados en una o más variedades);
- e) Bandas monomórficas (alelos/bandas que aparecen en todas las variedades analizadas, es decir, que no son polimórficas en una determinada colección de variedades).

4.3 Tratamiento de los datos de secuencias

Un registro detallado del proceso de tratamiento de los datos puede incluir la siguiente información:

- a) tipo de herramientas y versiones utilizadas
- b) línea de comandos utilizada para la herramienta y umbrales;
- c) recuentos de reproducibilidad;
- d) posibilidad de intercambiar datos y proceso de intercambio;
- e) si es posible, deben almacenarse los datos brutos de las alineaciones (archivos BAM o CRAM);
- f) los archivos VCF para varias muestras no son adecuados; debe figurar un archivo VCF por variedad;
- g) si se almacenan los archivos VCF, se deben registrar todas las posiciones (de variantes y no variantes) y su profundidad;
- h) para los métodos de detección se han de considerar y comparar enfoques tanto heurísticos como probabilísticos;
- i) las bases de datos deben facilitar la entrada y salida de los datos de identificación de variantes en un formato estándar (VCF o BCF);
- j) el proceso de tratamiento de los datos debe quedar reflejado en un detallado archivo de registro que se almacenará junto con los datos de identificación de variantes;
- k) si es posible, deben almacenarse los datos brutos para que el tratamiento de los datos pueda repetirse con nuevas herramientas o actualizaciones; y
- l) debe registrarse un valor p o la incertidumbre para un alelo determinado.

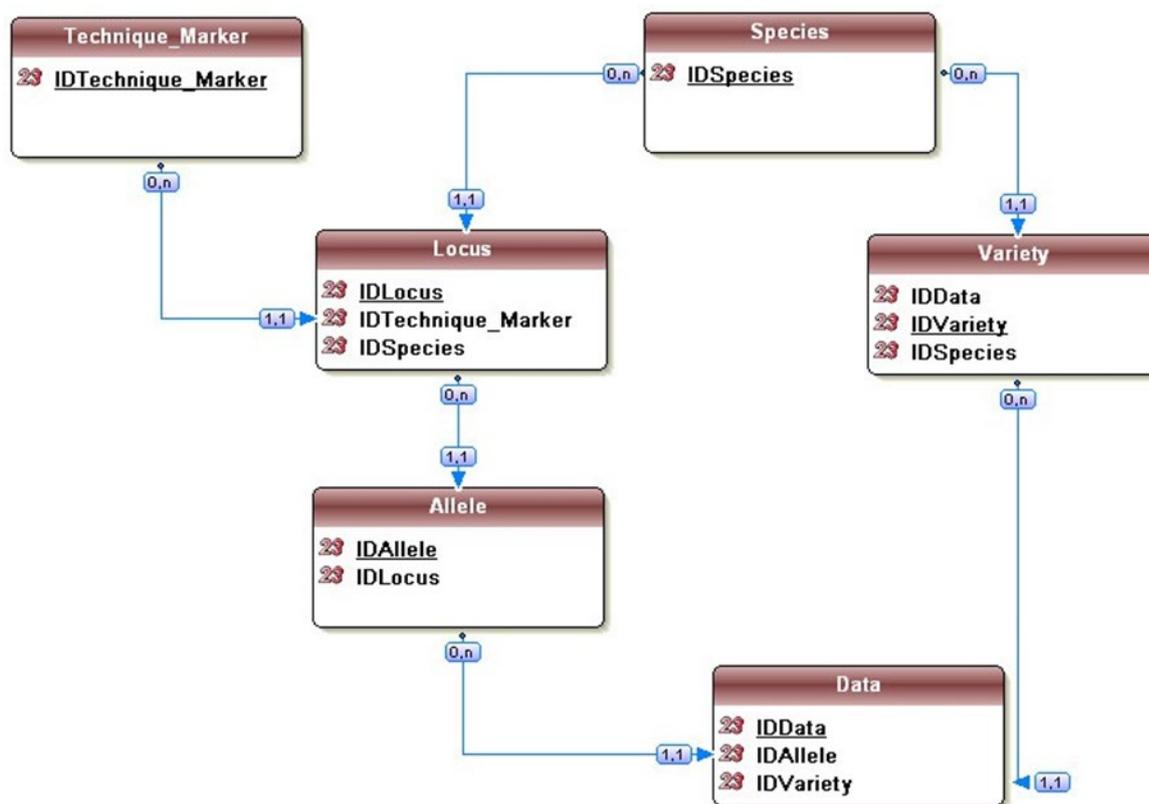
6. Bases de datos

6.14.4 Tipo de base de datos

Los datos moleculares pueden almacenarse de muchas maneras diferentes y, en consecuencia, es importante que la estructura desarrollada para la base de datos sea compatible con todos los usos a los que pretendidamente se destinarán los datos.

6.24.5 Modelo de base de datos

La definición del modelo de base de datos estará a cargo de expertos en bases de datos de T.I., quienes contarán para ello con la colaboración de los usuarios de la base de datos. El modelo de base de datos habrá de contener, como mínimo, seis objetos básicos: *Species* (especie); *Variety* (variedad); *Technique* (técnica); y *Allele* (alelo). *Marker detection method* (método de detección del marcador); *Marker* (marcador); *Locus* (locus); y *Allele* (alelo). Para las variantes obtenidas a partir de datos de secuenciación, los archivos VCF pueden almacenarse en una base de datos relacional o no SQL. En este caso, cada registro de una variante en la base de datos tiene una versión genómica, un cromosoma, una posición y un alelo de referencia definidos.



6-34.6 Diccionario de datos

4.6.3.1 En una base de datos, cada uno de los objetos configura un cuadro en el que se definen los campos. Por ejemplo:

- a) **Técnica/Código:** Tipo de marcador: designa el código o el nombre de la técnica o el tipo de marcador utilizado, por ejemplo, SSR, SNP, etc.
- b) **Posición en el genoma de referencia o código del locus:** designa—De ser posible, se debe proporcionar una versión del ensamblaje del genoma, el cromosoma y la posición si se dispone de un genoma de referencia de la especie de que se trate, p. ej., SL2.50ch05:63309763 en el caso del tomate (*Solanum lycopersicum*), la versión 2.50 del ensamblaje del genoma, el cromosoma 5 y la posición 63309763. Si no se dispone de un genoma de referencia o se desconoce la posición, puede utilizarse el nombre o el código del locus de la especie en cuestión, por ejemplo gwm 149, A2, etc.
- c) **Código alelo:** designa—Genotipo: En el caso de los perfiles SNP, debe proporcionarse la composición alélica del SNP o el MNP, por ejemplo, A/T o A/A. Si se trata de otras técnicas, el genotipo designa el nombre o el código del alelo de un locus determinado de la especie de que se trate, por ejemplo, 1, 123, etc.
- d) **Valor de los datos:** Profundidades alélicas o valor de los datos: En el caso de los SNP obtenidos de datos de secuenciación de nueva generación, estos deben indicar la profundidad de cobertura de los alelos, por ejemplo, 10/20 para un alelo A/T, en el que A se ha cubierto mediante 10 lecturas y T mediante 20 lecturas. De lo contrario, se designa el valor de los datos de una muestra determinada respecto de un alelo presente en un locus determinado, por ejemplo, 0 (ausencia), 1 (presencia), 0,25 (frecuencia), etc.
- e) **Variedad:** Denominación de la variedad o referencia del obtentor: la variedad es el objeto para el que se han obtenido los datos.
- f) **Tipo de variedad:** por ejemplo, línea endógama o híbrido.

f)g) especie: la especie se designa mediante el nombre botánico o el nombre nacional común que, a veces, se refiere también al tipo de variedad (por ejemplo, variedades del tipo de verano o de primavera, etc.). Se recomienda la utilización del código de la UPOV ~~evitaría para evitar~~ los problemas que plantean los sinónimos ~~por lo que resultaría ventajoso en términos de coordinación~~.

6.4.3.2 En el “diccionario de datos” debe especificarse, en cada cuadro, el número de campos, su nombre y definición, y los posibles valores y normas que han de aplicarse.

6.4 Relación de los cuadros

6.4.1 La relación entre los cuadros es un aspecto importante en el diseño de la base de datos. A continuación se presenta la relación que se establece entre ellos:

Cuadro	Enlace	Cuadro	Descripción
Mujer	0 — o 1 a n (0, n)	Hijo	0: es posible que una mujer no tenga hijos 1 a n: es posible que una mujer tenga de 1 a n hijos (en cuyo caso es madre)
Hijo	1 a 1 (1,1)	Mujer	Un hijo determinado solo tiene una madre biológica

6.4.2 En el siguiente cuadro figura la relación entre los seis objetos básicos que debe contener como mínimo la base de datos, como se propone en el modelo de base de datos de la Sección 6.2:

Cuadro	Enlace	Cuadro	Descripción
Técnica/marcador	0 — o 1 a n	Locus	0: Una técnica/un marcador puede figurar en Técnica/marcador, aunque en la base de datos todavía no se esté utilizando ningún locus/alelo 1 a n: un tipo de marcador determinado puede proporcionar de 1 a n loci útiles
Locus	1 a 1	Técnica/marcador	Un locus determinado se define dentro del ámbito de una técnica/un marcador determinados
Locus	1 a n	Alelo	Por cada locus, puede describirse 1 o más de 1 alelo
Alelo	1 a 1	Locus	Un alelo determinado se define dentro del ámbito de un Locus determinado
Alelo	0 — o 1 a n	Datos	0: puede definirse un alelo determinado, pero sin datos 1 a n: un alelo determinado puede estar presente en 1 a n datos
Datos	1 a 1	Alelo	los datos corresponden a un alelo determinado
Variedad	0 — o 1 a n	Datos	0: no hay datos para la variedad 1 a n: hay datos para la variedad
Datos	1 a 1	Variedad	los datos corresponden a una variedad determinada
Datos	1 a 1	Especie	los datos se obtienen para una variedad determinada y, luego, para la especie de la variedad
Especie	0 — o 1 a n	Datos	0: es posible que no haya datos para una especie. 1 a n: es posible que haya de 1 a n datos para una especie.

6.5 Transferencia de datos a la base de datos

Al transferir y transcribir los datos es posible introducir errores; para reducir esta posibilidad, es aconsejable automatizar en la mayor medida posible la transferencia de los datos a la base de datos.

6.6.4.7. Acceso a los datos y titularidad

Se recomienda abordar todas las cuestiones concernientes a la titularidad de los datos y el acceso a los datos de la base de datos en el momento de iniciar cualquier trabajo.

6.75. Análisis de Intercambio de los datos

~~El método de análisis vendrá determinado por la finalidad a la que hayan de destinarse los datos; por tanto, las presentes directrices no incluyen ninguna recomendación específica al respecto.~~

~~6.8 Validación de la base de datos~~

~~Al concluir la primera fase de la base de datos es recomendable llevar a cabo una “prueba a ciegas”, es decir, distribuir un número de muestras a distintos laboratorios y solicitarles que las identifiquen utilizando el protocolo acordado y la base de datos.~~

5.1 Situaciones de intercambio de datos

A efectos de la cooperación, el modelo de datos debe admitir distintas situaciones, como el intercambio de datos generados a partir de un conjunto estándar de marcadores para un cultivo determinado (situación n.º 1) y la búsqueda y consulta de datos de determinadas variedades generados a partir del mismo conjunto estándar de marcadores (situación n.º 2). En el Anexo “Situaciones de intercambio de datos y métodos de transmisión de datos” se ofrecen detalles técnicos de ambas situaciones.

5.2 Métodos de intercambio de datos

5.2.1 Los datos a transmitir pueden contener información diversa, como loci, muestras, ADN y datos y perfiles de huellas genéticas. El método de transmisión de datos se determinará con arreglo al contenido que se ha de transmitir y debe tenerse en cuenta lo siguiente:

- a) la cantidad de datos;
- b) la complejidad de los datos;
- c) los requisitos para las funciones de búsqueda.

En el Anexo “Situaciones de intercambio de datos y métodos de transmisión de datos” se ofrecen detalles técnicos de ambas situaciones.

5.2.2 Los formatos de datos empleados habitualmente son zip, csv, json y xml. A continuación se indican las características de cada uno:

1) El formato zip admite diversos archivos de información sobre los datos en formato original y, gracias a su elevado índice de compresión de datos y facilidad de transmisión, resulta adecuado para datos complejos y de gran volumen.

2) El formato csv es más adecuado para información sobre datos en formato sencillo, con la ventaja de que contiene menos datos no válidos y la velocidad de procesamiento es mayor.

3) Los formatos json y xml pueden contener información sobre datos más complejos e información más redundante, pero ambos muestran una buena legibilidad.

7.6. Resumen

A continuación se ofrece un resumen del enfoque recomendado para la determinación de alta calidad del perfil de ADN de las variedades, incluida la selección y utilización de marcadores moleculares, para la creación de bases de datos centralizadas así como la creación de bases de datos moleculares compartidas y sostenibles de perfiles de ADN de variedades (es decir, bases de datos que en el futuro puedan llenarse con datos procedentes de una diversidad de fuentes, independientemente de la tecnología empleada).

- a) considerar el enfoque cultivo a cultivo;
- b) acordar un tipo de marcador aceptable y la fuente del mismo;
- c) acordar las plataformas y los equipos de detección aceptables;
- d) acordar los laboratorios que participarán en la prueba;
- e) acordar las cuestiones relativas a la calidad (véase la sección 5.2);
- f) verificar la fuente del material vegetal empleado (véase la sección 4);

- g) acordar los marcadores a utilizar en una fase de evaluación preliminar en la que se contará con la colaboración de más de un laboratorio y se usarán equipos de detección diferentes (~~véase la sección 2~~);
- h) realizar una evaluación (~~véase la sección 5.3~~);
- i) elaborar y acordar un protocolo para la evaluación de los datos moleculares (~~véase la sección 5.4~~);
- j) acordar el material vegetal y el conjunto de referencias a analizar y la fuente de los mismos;
- k) analizar la colección de variedades acordada, en distintos laboratorios y con distintos equipos de detección, utilizando muestras duplicadas y, si surgiesen problemas, intercambiando muestras o extractos de ADN;
- l) usar en todos los análisis variedades, muestras de ADN y alelos de referencia; utilizar referencias (variedades, muestras de ADN o alelos, según corresponda) en todos los análisis;
- m) verificar todas las etapas (inclusive la entrada de datos); lograr la mayor automatización posible;
- n) efectuar una “prueba a ciegas” de la base de datos en laboratorios diferentes;
- o) adoptar ~~los~~ procedimientos para la adición de nuevos datos.

GLOSARIO

Microsatélites o Secuencias Simples Repetidas (SSR)

Los microsatélites, o secuencias simples repetidas (SSR), son secuencias de ADN que se repiten en tándem, por lo general con una unidad de repetición de 2 a 4 pares de bases (por ejemplo, GA, CTT y GATA). En muchas especies, se ha demostrado la existencia de múltiples alelos para algunos microsatélites, debido a las variaciones en el número de copias de esta unidad de repetición. Los microsatélites pueden analizarse por la técnica de la PCR utilizando cebadores específicos, procedimiento conocido como aproximación por microsatélites de sitio de secuencia marcada (STMS). Los alelos (productos de la PCR) pueden separarse por electroforesis en gel de agarosa o en gel de poliacrilamida. Para desarrollar microsatélites de sitio de secuencia marcada se necesita información sobre la secuencia de ADN que flanquea al microsatélite. Si bien esta información puede obtenerse en ocasiones a partir de las bases de datos de secuencias de ADN existentes, otras veces ha de conseguirse de forma empírica.

Polimorfismos de Nucleótido Único (SNP)

Los polimorfismos de nucleótido único (SNP) son variaciones que se producen en la secuencia de ADN cuando se altera un nucleótido único (A, T, C o G) de la secuencia del genoma. Por ejemplo, un SNP podría cambiar una secuencia de ADN de AAGGCTAA a ATGGCTAA. Por lo general, para que una variación pueda considerarse un SNP debe ocurrir al menos en el 1% de la población. El número potencial de marcadores SNP es muy elevado, lo cual significa que sería factible encontrarlos en cualquier parte del genoma. Los SNP pueden producirse tanto en las zonas codificadoras (genes) del genoma como en las zonas no codificadoras. El descubrimiento de los SNP trajo aparejada la secuenciación comparativa de varios individuos de una población. La identificación de los SNP potenciales se realiza comúnmente mediante la comparación de secuencias alineadas tomadas de las bases de datos de secuencias disponibles. Aunque los SNP pueden detectarse con relativa facilidad mediante PCR + electroforesis en gel, se están desarrollando procedimientos de análisis a gran escala y micromatrices que permitirán la evaluación automática de centenares de loci con SNP de una sola vez.

Secuencias Polimórficas Amplificadas y Cortadas (CAPS)

Las secuencias polimórficas amplificadas y cortadas (CAPS) son fragmentos de ADN que primero se amplifican mediante PCR utilizando cebadores específicos de entre 20 y 25 pb y a continuación se digieren con una endonucleasa de restricción. Posteriormente, la electroforesis en gel de los productos digeridos permite identificar los polimorfismos de longitud producidos por la variación en la aparición de los sitios de restricción. En comparación con otros marcadores, como los RFLP, los polimorfismos son más difíciles de identificar debido al reducido tamaño de los fragmentos amplificados (300-1800 pb). Sin embargo, el análisis CAPS no requiere hibridación Southern y detección radioactiva. Hasta la fecha, en general las CAPS se han venido aplicando principalmente en los estudios de cartografía génica.

Regiones Amplificadas Caracterizadas por Secuencia (SCARs)

Las regiones amplificadas caracterizadas por secuencia (SCARs) son fragmentos de ADN amplificados mediante PCR utilizando cebadores específicos de entre 15 y 30 pb, diseñados a partir de secuencias polimórficas identificadas con anterioridad. La utilización de cebadores PCR más largos permite a las SCAR evitar el problema de baja reproducibilidad. También suelen ser marcadores codominantes. Las SCAR son específicas de cada locus y se han venido aplicando en los estudios de cartografía génica y en la selección asistida por marcador.

Pig-tail

En el análisis SSR, se denomina pig tail la adición de una secuencia de oligonucleótidos a los cebadores empleados en la PCR, como una forma de mejorar la claridad de los productos de la amplificación y reducir los artefactos.

Alelo nulo

En el análisis SSR, un “alelo nulo” es un alelo en un locus determinado cuyo efecto se aprecia como la ausencia de algún producto de la PCR.

Bandas Stutter

En el análisis SSR, las “bandas *stutter*” hacen referencia a una serie de una o más bandas que aparecen tras la PCR y cuyo tamaño difiere en 1 unidad de repetición.

C. LISTA DE SIGLAS

API	Interfaz de programación de aplicaciones (<i>Application Programming Interface</i>)
BAM	Mapa de alineación binaria (<i>Binary Alignment Map</i>)
BCF	Formato binario de colaboración (<i>Binary Call Format</i>)
CRAM	Mapa comprimido de alineación de secuencias basada referencias (<i>Compressed Reference-oriented Alignment Map</i>)
MNP	Polimorfismo de varios nucleótidos (<i>Multiple Nucleotide Polymorphism</i>)
NGS	Secuenciación de nueva generación (<i>Next Generation Sequencing</i>)
NIL	Línea casi isogénica (<i>Near Isogenic Line</i>)
RIL	Línea recombinante endógama (<i>Recombinant Inbred Line</i>)
SAM	Mapa de alineación de secuencias (<i>Sequence Alignment Map</i>)
SNP	Polimorfismo de nucleótido único (<i>Single Nucleotide Polymorphism</i>)
SQL	Lenguaje de consulta estructurado (<i>Structured Query Language</i>)
SSR	Secuencias simples repetidas (<i>Simple Sequence Repeats</i>)
TIFF	Formato de archivo de imagen con etiquetas (<i>Tagged Image File Format</i>)
VCF	Formato de llamada de variantes (<i>Variant Call Format</i>)

[Sigue el Anexo]

ANEXO

SITUACIONES DE INTERCAMBIO DE DATOS Y MÉTODOS DE TRANSMISIÓN**A: Situaciones de intercambio de datos**

Situación n.º 1: intercambio de datos generados a partir de un conjunto estándar de marcadores para un cultivo determinado

Para intercambiar datos sobre el conjunto de marcadores empleado para un cultivo determinado, se puede recurrir al siguiente servicio en línea:

https://office.org/locus?upov_code={upovcode}&type={marker type}&method={observation method}

Por ejemplo, para obtener información sobre conjuntos de marcadores SSR para el maíz con el método CE, se ha de acceder a la siguiente dirección:

https://office.org/locus?upov_code=ZEAAA_MAY&type=SSR&method=CE

El resultado sería el siguiente:

<code>{"techniqueid":</code>	<code>l.</code>	<code>l.</code>
<code>"CN SSR ZEAA MAY CE V</code>	<code>["alleleid": "248/271",</code>	<code>["alleleid": "256/256",</code>
<code>1",</code>	<code>"examplevariety":</code>	<code>"examplevariety":</code>
<code>"description": "Laboratory</code>	<code>l.</code>	<code>l.</code>
<code>method description"</code>	<code>["alleleid": "248/290",</code>	<code>["alleleid": "256/264",</code>
<code>["locusid": "M01",</code>	<code>"examplevariety":</code>	<code>"examplevariety":</code>
<code>"alleles":</code>	<code>l.</code>	<code>l.</code>
<code>["alleleid": "238/256",</code>	<code>["alleleid": "250/250",</code>	<code>["alleleid": "256/266",</code>
<code>"examplevariety":</code>	<code>"examplevariety":</code>	<code>"examplevariety":</code>
<code>l.</code>	<code>l.</code>	<code>l.</code>
<code>["alleleid": "238/271",</code>	<code>["alleleid": "250/252",</code>	<code>["alleleid": "256/271",</code>
<code>"examplevariety":</code>	<code>"examplevariety":</code>	<code>"examplevariety":</code>
<code>l.</code>	<code>l.</code>	<code>l.</code>
<code>["alleleid": "246/246",</code>	<code>["alleleid": "250/256",</code>	<code>["alleleid": "256/284",</code>
<code>"examplevariety":</code>	<code>"examplevariety":</code>	<code>"examplevariety":</code>
<code>l.</code>	<code>l.</code>	<code>l.</code>
<code>["alleleid": "246/248",</code>	<code>["alleleid": "250/275",</code>	<code>["alleleid": "256/286",</code>
<code>"examplevariety":</code>	<code>"examplevariety":</code>	<code>"examplevariety":</code>
<code>l.</code>	<code>l.</code>	<code>l.</code>
<code>["alleleid": "246/250",</code>	<code>["alleleid": "252/256",</code>	<code>["alleleid": "258/258",</code>
<code>"examplevariety":</code>	<code>"examplevariety":</code>	<code>"examplevariety":</code>
<code>l.</code>	<code>l.</code>	<code>l.</code>
<code>["alleleid": "246/254",</code>	<code>["alleleid": "252/260",</code>	<code>["alleleid": "264/284",</code>
<code>"examplevariety":</code>	<code>"examplevariety":</code>	<code>"examplevariety":</code>
<code>l.</code>	<code>l.</code>	<code>l.</code>
<code>["alleleid": "246/256",</code>	<code>["alleleid": "252/271",</code>	<code>["alleleid": "271/292",</code>
<code>"examplevariety":</code>	<code>"examplevariety":</code>	<code>"examplevariety":</code>
<code>l.</code>	<code>l.</code>	<code>l.</code>
<code>["alleleid": "246/260",</code>	<code>["alleleid": "252/273",</code>	<code>l.</code>
<code>"examplevariety":</code>	<code>"examplevariety":</code>	<code>l.</code>
<code>l.</code>	<code>l.</code>	<code>["locusid"]="M02".</code>
<code>["alleleid": "246/277",</code>	<code>["alleleid": "252/282",</code>	<code>"alleles": [...]</code>
<code>"examplevariety":</code>	<code>"examplevariety":</code>	<code>]} vi</code>
<code>l.</code>	<code>l.</code>	
<code>["alleleid": "246/284",</code>	<code>["alleleid": "254/254",</code>	
<code>"examplevariety":</code>	<code>"examplevariety":</code>	
<code>l.</code>	<code>l.</code>	
<code>["alleleid": "246/288",</code>	<code>["alleleid": "254/271",</code>	
<code>"examplevariety":</code>	<code>"examplevariety":</code>	
<code>l.</code>	<code>l.</code>	
<code>["alleleid": "248/250",</code>	<code>["alleleid": "254/284",</code>	
<code>"examplevariety":</code>	<code>"examplevariety":</code>	
<code>l.</code>	<code>l.</code>	
<code>["alleleid": "248/256",</code>	<code>["alleleid": "254/286",</code>	
<code>"examplevariety":</code>	<code>"examplevariety":</code>	

Situación n.º 2: búsqueda y consulta de datos de determinadas variedades generados a partir del mismo conjunto estándar de marcadores

Para buscar y consultar datos moleculares de una variedad, se puede recurrir al siguiente servicio en línea:
https://office.org/variety?id={irn}&techniqueid={technique_code} vi

Por ejemplo,

[https://office.org/variety?id=XU_30201800000140 &techniqueid= CN SSR ZEAA MAY CE V 1](https://office.org/variety?id=XU_30201800000140&techniqueid=CN_SSR_ZEAA_MAY_CE_V_1) vi

El resultado sería el siguiente:

```
{ "techniqueid": "CN SSR ZEAA MAY PAGE ",  
  "varietyid": " XU 30201800000140 ",  
  "computationalsteps": "xxxxxxxxxxxxx"  
  "data":  
  [  
    { "id": "M01",  
      "value": "254/254"  
    },  
    { "id": "M02",  
      "value": "347/347"  
    },  
    { "id": "M03",  
      "value": "292/292"  
    },  
    { "id": "M04",  
      "value": "361/361"  
    },  
    ...  
  ]  
} vi
```

B: Métodos de transmisión de datos

A continuación se ofrece un ejemplo de creación de un paquete de datos de huella genética en formato zip para su transmisión. En primer lugar, han de asignarse identificadores independientes a las muestras, el ADN, los datos de huella genética y el atlas de la huella genética. De este modo, el archivo de datos en formato json contendrá toda la información sobre los loci, las muestras y el ADN. Cada conjunto de datos de huella genética se almacena por separado en un archivo json. El identificador de la huella genética estará vinculado al correspondiente locus de los datos de huella genética, y todos los archivos de datos de huella genética y del espectro de la huella genética se almacenarán por separado en el directorio correspondiente. El paquete de datos de huella genética tendrá la siguiente estructura:

```
zip/markers.json  
zip/samples.json  
zip/dnas.json  
zip/genes/gene_id_1.json  
zip/genes/gene_id_2.json  
.....  
zip/genes/gene_id_n.json  
zip/maps/map_id_1.png  
zip/maps/map_id_2.png  
.....  
zip/maps/map_id_m.png
```

El paquete de datos de huella genética en formato zip puede ampliarse para introducir más información. El archivo de datos de huella genética constituye el núcleo del paquete y de la correlación, por lo que puede analizarse correctamente la correlación entre las partes y los datos pueden transmitirse entre diferentes sistemas.