



P  
633.18  
GAR

P 633.18  
GAR

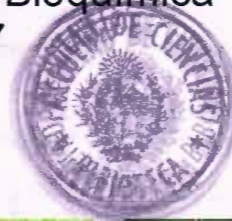


Facultad de Ciencias  
Universidad de la República

UNIVERSIDAD  
DE LA REPUBLICA  
URUGUAY

# CLASIFICACIÓN ASISTIDA POR MARCADORES MOLECULARES PARA DIFERENCIACIÓN DE BIOTIPOS CON CARACTERÍSTICAS DE MALEZA Y CULTIVARES COMERCIALES DE ARROZ

Trabajo especial II  
Licenciatura en Bioquímica  
2007



12500



17/04/07: AUTOS. - 12/11/07

**Silvia Garaycochea**

**Tutores: Ing. Agr. Fabián Capdevielle, Msc.  
Ing. Agr. Bioq. Victoria Bonnacarrère, Msc.**

P  
633.18 GAR  
Clasificación asistida po...  
Garaycochea Solsana, Silvia Raquel



FC/12500

## AGRADECIMIENTOS

Quiero agradecer al Instituto Nacional de Investigación Agropecuaria (INIA), por permitir la realización de mi trabajo de tesis en sus instalaciones.

Especialmente dar las gracias a mis tutores, Fabián y Victoria, por la paciencia, el aguante y por todo lo que me han enseñado que me ha ayudado a crecer a nivel académico y personal.

Muchas, muchas gracias!!!!, por lo que me dieron y me siguen dando.

Agradecer a todo el personal de la Unidad de Biotecnología, los que me apoyaron de una forma u otra para llevar a cabo mi trabajo.

A todos mis compañeros del laboratorio, con los que hemos compartido muchos momentos, especialmente a Paola y Diego (siempre dando "pa' delante" y aguantando la cabeza).

Un gracias enorme a toda mi familia, por apoyarme desde el principio al final (llegó el final !!). Mamá, Papá, Nacho, Andre y Guada muchas gracias. También a mis amigos por todo el apoyo. Especialmente muchas gracias a Martín, por el apoyo, confianza y aguante.

GRACIAS a todos !!!!!

# CLASIFICACIÓN AISISTIDA POR MARCADORES MOLECULARES PARA DIFERENCIACIÓN DE BIOTIPOS CON CARACTERISTICAS DE MALEZA Y CULTIVARES COMERCIALES DE ARROZ

<b>RESUMEN</b>	<b>2</b>
<b>INTRODUCCIÓN</b>	<b>3</b>
Importancia del arroz rojo como maleza	3
Identificación temprana de arroz como método de prevención	8
<b>OBJETIVO</b>	<b>10</b>
<b>MATERIALES Y MÉTODOS</b>	<b>11</b>
Material verde	11
Determinación de viabilidad, germinación y dormancia de las semillas	12
Aislamiento del ADN	13
Marcadores microsatélites	14
Análisis de datos	15
<b>RESULTADOS Y DISCUSIÓN</b>	<b>17</b>
Determinación de la viabilidad y germinación	18
Extracción de ADN y obtención de SSR	19
<b>CONCLUSIONES</b>	<b>36</b>
<b>BIBLIOGRAFÍA</b>	<b>37</b>
<b>ANEXO</b>	

# CLASIFICACIÓN ASISTIDA POR MARCADORES MOLECULARES PARA DIFERENCIACIÓN DE BIOTIPOS CON CARACTERÍSTICAS DE MALEZA Y CULTIVARES COMERCIALES DE ARROZ

## Resumen

El objetivo de este trabajo es ajustar un sistema de clasificación asistida por marcadores moleculares para diferenciación de biotipos de arroz rojo e identificación de sus híbridos respecto a cultivares de arroz. Los diferentes biotipos de arroz rojo constituyen poblaciones anuales de *Oryza sativa* L. con características típicas de malezas y cuyas semillas usualmente presentan pericarpio con coloración rojiza. Dichos biotipos se encuentran frecuentemente bajo condiciones de siembra directa en las mayores áreas de cultivo de arroz; además de presentar similitud con cultivares de los subgrupos Indica o Japónica de arroz, existen evidencias de flujo genético con arroz cultivado que dificultaría la diferenciación morfo-fisiológica respecto a cultivares de arroz.

Los microsatélites mapeados en el genoma de arroz que fueron utilizados en este trabajo (RM215, RM251, RM253 y RM234) se seleccionaron en base a estudios previos (Gealy et al, 2002) que indicaban una posible diferenciación maleza-cultivar para poblaciones de arroz cultivadas en EEUU. Los alelos existentes para estos 4 microsatélites fueron determinados en una muestra de cultivares uruguayos y accesiones de arroz rojo colectadas en la región utilizando un secuenciador automático ABI 310. Esta información se utilizó para validar un procedimiento de clasificación asistida por marcadores moleculares basado en el uso del algoritmo k-NN, evaluado previamente para marcadores AFLP (Capdevielle et al, 2003). Los porcentajes de clasificación correcta para muestras de tipo maleza y del tipo cultivar estuvieron entre 95% y 99% (validación cruzada utilizando SAS proc discrim), permitiendo la identificación de cruzamientos arroz rojo x cultivar realizados en forma experimental. Estos resultados indican que la diferenciación por técnicas moleculares de biotipos de arroz rojo respecto a cultivares utilizados a nivel agroindustrial sería un procedimiento aplicable para apoyar procesos de control de calidad en lotes de semillas básicas, en Uruguay y en otros países de la región .

## INTRODUCCIÓN

### *Importancia del arroz rojo como maleza*

Bajo el nombre genérico de “arroz rojo” se incluyen una serie de biotipos de arroz con características de tipo “maleza” pertenecientes a la misma especie (*Oryza sativa* L.) que el arroz cultivado en Uruguay y otras regiones del mundo, y cuya principal característica distintiva es la de poseer pericarpio de color rojizo, negro o marrón. Las características del arroz rojo asociadas con su comportamiento como maleza incluyen un mayor desarrollo vegetativo de las plantas, alta capacidad de dispersión y persistencia de las semillas y otras adaptaciones fisiológicas que favorecen su desarrollo como especie competidora respecto a plantas de la especie cultivada; estas características resultan particularmente problemáticas en el contexto de sistemas de cultivo mecanizados que utilizan variedades de arroz de tipo semi-enano para maximizar la producción por unidad de superficie. Desde un punto de vista fenológico, los diferentes biotipos de arroz rojo (tipo maleza) que se encuentran bajo condiciones de siembra directa en las mayores áreas de cultivo de arroz -Latinoamérica y Norte América, Caribe, África, y algunas regiones del Sur y Sudeste de Asia- constituirían poblaciones anuales de *O. sativa* L [Cao, *et al.*, 2006]. Algunos estudios sobre la diversidad genética dentro de estas poblaciones de arroz han reportado que tanto los biotipos cultivados como los biotipos con características de maleza se diferenciarían en clases de germoplasma denominadas *Indica* y *Japónica* en base en características morfológicas y fisiológicas, isoenzimas, RFLP (Restriction fragment length polymorphism), RAPD (Random amplified polymorphic DNA) [Suh *et al.*, 1997] y microsatélites (SSR, simple sequences repeat) [Vaughan *et al.*, 2001].

La adaptación de las plantas silvestres a los cambios ambientales causados por los procesos de domesticación de las especies cultivadas, la aparición de híbridos derivados de plantas cultivadas y especies silvestres reproductivamente compatibles [Baker 1974, Wet y Harlan, 1975] y la segregación de híbridos intervarietales, podrían estar jugando un rol importante en el origen de las poblaciones con características de malezas [Burger *et al.*, 2006]; cada uno de estos procesos estarían involucrados en los cambios adaptativos que favorecen la competitividad de las malezas respecto a las formas cultivadas, dependiendo de la localización geográfica de las plantas, así como de las prácticas agrícolas pasadas y actuales [Mortimer, 1994].

Existen diferentes hipótesis sobre el origen del arroz tipo maleza que consideran las posibles contribuciones de procesos tales como pre – adaptación e introgresión [Mortimer, 1994]. Los ecotipos de arroz maleza se encontrarían pre-adaptados a las condiciones de siembra directa en las que se realiza el cultivo por compartir gran parte de las características funcionales con las variedades modernas seleccionadas por programas de fitomejoramiento [Oka, 1988], existiendo asimismo evidencias de introgresión genética entre variedades cultivadas y ecotipos de arroz rojo.

Esto determinaría que los híbridos entre diferentes formas de *O. sativa* L y sus poblaciones derivadas presenten diferentes combinaciones de caracteres funcionales asociados con el proceso de domesticación (transición entre biotipos no cultivados y biotipos cultivados ancestrales), que representarían posibles clases fenotípicas (con extremos definidos como “cultivar” y “maleza”) con diferente grado de adaptación a los sistemas agrícolas actuales. Entre los factores que condicionarían la existencia de dichas clases fenotípicas se plantean los siguientes: 1) inmigración de biotipos de arroz rojo desde otros ambientes (dispersión de semillas asociadas con variedades cultivadas), 2) segregación de características de tipo maleza a partir de cultivares tradicionales (más similares a los tipos ancestrales), y 3) hibridación e introgresión genética entre biotipos de arroz rojo y variedades cultivadas (“mimetización” de los biotipos de arroz rojo).

En Uruguay se encuentran comúnmente dos tipos de arroz rojo que se diferencian por la coloración de sus glumas (figura 1). El “arroz negro” el que presenta una arista larga (4-5 cm) y resistente y el “arroz rojo” propiamente dicho, de glumas ocres y de arista corta; ambos son de grano corto, grueso y pubescente [Zorrilla de San Martín, 1992].



Sus efectos negativos desde el punto de vista agronómico, al igual que en otras malezas comunes, se deben a su competencia por diferentes factores productivos

(nutrientes, agua, luz, etc.) con el arroz cultivado, con la particularidad de que el arroz rojo se ve favorecido por casi todas las labores que se realizan para el cultivo. Además existe el agravante de que no existen controles químicos adecuados por tratarse del mismo género y especie [Hoagland y Paul, 1978]; sus perjuicios van más allá del campo, afectando no solamente la productividad del arroz cultivado sino también la calidad de los lotes cosechados, esto último debido a que la coloración y textura del pericarpio [Hoagland y Paul, 1978] afectan los procesos industriales de descascarado y pulido. Fenotípicamente el arroz tipo maleza está caracterizado por el desgrane y dormancia de sus semillas [Cao, *et al.*, 2006], lo que favorece su dispersión y persistencia en el campo; la dormancia es una característica particularmente negativa desde el punto de vista del control de esta maleza, ya que se ha observado que las semillas de arroz rojo pueden estar dormantes sin perder la viabilidad en el suelo entre 7 a 15 años [Naredo *et al.*, 1998].

Las plantas de “arroz rojo” son plantas más altas con respecto a las variedades cultivadas (figura 2), poseen hojas ásperas de color verde claro, forman un mayor número de macollos y son de hábito decumbente. Además, presentan alta plasticidad fenotípica y capacidad para la dispersión de sus semillas [Gealy y Black, 1998]. Previo a la floración es muy difícil identificar con seguridad los biotipos de arroz rojo y diferenciarlos de cualquiera de las variedades cultivadas, ya que sólo se dispone de información sobre algunas características distintivas, como la altura de la planta, el color de hojas y la pubescencia, las cuales no son exclusivas de un biotipo. Además, se debe tener en cuenta que puede existir un alto grado de fertilidad en los cruzamientos entre variedades cultivadas y biotipos de arroz rojo, con el consiguiente flujo genético entre poblaciones que dificulta la discriminación de plantas de *O. sativa* L. con características de tipo maleza.



**Figura 2:** Comparación de las características morfológicas entre arroz tipo maleza y arroz cultivado

Los cruzamientos entre plantas de diferentes biotipos se producen porque, aún cuando el arroz cultivado es principalmente autógamo (polinización y fertilización ocurren con la flor cerrada), puede presentar porcentajes de polinización cruzada desde un 0,5 % [Gealy *et al.*, 2002] a más del 5 %, lo que implica que algunas plantas de las variedades cultivadas se crucen entre ellas e incluso con otras especies del genoma AA, grupo al que pertenecen tanto el arroz cultivado como el arroz rojo (ambos biotipos pertenecientes a la especie *Oryza sativa* L.) [Singh y Khush, 2000]. Es por lo tanto frecuente en zonas con alta contaminación de arroz rojo y acumulación de muchos años de cultivo, encontrar diversidad dentro de los biotipos de arroz rojo, algunos de los cuales son muy parecidos a las variedades utilizadas en esas zonas, debido a un proceso de asimilación que se conoce como "mimetización"; se ha propuesto que esta capacidad le permite al arroz rojo adaptarse rápidamente a cambios en el sistema de cultivo [Zorrilla de San Martín, 1992].

Luego de la floración, las diferencias fenotípicas permiten su identificación en el campo y posterior raleo, pero de esta forma ya se está en presencia de contaminación. Tomando en cuenta la característica de dormancia, la capacidad de dispersión debido a la estructura de la panoja, la competencia por luz, nutriente y espacio de la maleza, el raleo durante el ciclo de cultivo no es un método eficaz.



La creación de variedades con características tales como resistencia a herbicidas, tanto por transgénesis como por mutaciones inducidas, se han llevado a cabo como mediada para combatir este tipo de maleza. Dada la alta tasa de fertilidad entre la maleza y el cultivo, no es una buena estrategia a largo plazo [Oard *et al*, 2000], ya que los genes que codifican para la nueva característica tiene la capacidad de introgresión en las especies de maleza a través de la hibridación (polinización cruzada) con las variedades transgénicas o mutantes [Chen *et al.*, 2004], provocando nuevas dificultades en el manejo. Por lo tanto, es imprescindible el desarrollo de métodos que permitan la diferenciación entre arroz rojo y arroz cultivar en etapas más tempranas del ciclo.

El arroz rojo no ha sido un problema grave en la producción arrocera uruguaya en los últimos veinte años, debido fundamentalmente a una exitosa campaña de control –con erradicación de plantas de tipo maleza- durante las etapas de producción de semillas para las diferentes variedades utilizadas por los productores. Sin embargo, existe creciente preocupación respecto a mantener estas condiciones privilegiadas de producción. El aumento sostenido del área de cultivo que ha llevado a una intensificación de la rotación (mayor número de años bajo cultivo de arroz), el aumento de los volúmenes de semilla a producir (con mayor requerimiento de áreas libres de malezas), la preocupante situación de expansión del área donde existe arroz rojo en las zonas arroceras vecinas, especialmente en Río Grande del Sur, hacen necesario plantearse estrategias complementarias para la prevención [Zorrilla de San Martín, 1992].

La principal medida preventiva sería la utilización de semillas para la siembra directa con calidad genética y pureza física certificada, para mantener los campos libres de arroz rojo a largo plazo; en este caso se considera imprescindible poder apoyar la detección convencional de lotes contaminados con arroz rojo (identificación de plantas y de semillas de la maleza en base a caracteres morfo-fisiológicos diferenciales) con técnicas de alta sensibilidad que permitan identificar aquellos casos donde puede existir contaminación genética (plantas y semillas pertenecientes a híbridos entre arroz rojo y variedades cultivadas) dentro de los procesos de producción de semillas de nuevas variedades desarrolladas en Uruguay [Capdevielle *et al*, 2004]. El uso de clasificación asistida por marcadores moleculares puede ser una herramienta práctica para la identificación de una potencial contaminación de lotes de semilla por arroz rojo, contribuyendo de esta forma al control de calidad para la producción de semillas de nuevas líneas

## Identificación temprana de arroz rojo como método de prevención

Para la implementación de estrategias de prevención que impidan la contaminación de los campos con arroz rojo, se hace necesario el uso de tecnologías que permitan conocer con mayor profundidad el objeto del problema. Con este fin varios estudios sobre la diversidad genética del arroz rojo se han llevado a cabo, con la utilización de diversas metodologías moleculares [Estorninos *et al.*, 2000; Vaughan *et al.*, 2001; Gealy *et al.*, 2002; etc.]. El mayor conocimiento acerca de la diversidad genética y comportamiento del arroz rojo, así como la elucidación de su origen y los procesos evolutivos de este, son de ayuda para establecer estrategias efectivas para su control [Pysek y Prach, 2003], permitiendo además establecer herramientas que mejoren el análisis de semillas para su certificación. En la actualidad la certificación de lotes de semillas, la que incluye determinación de presencia de arroz rojo en los lotes, utiliza métodos basados principalmente en identificación fenotípica, haciendo de este un método de control con alto contenido subjetivo.

Con el fin de establecer tecnologías que permitan detectar contaminación de arroz rojo antes de llegar al campo, acortar los tiempos necesarios para este análisis y sobre todo disponer de métodos objetivos, se comenzó con el desarrollo de métodos moleculares. Basados en estudios realizados en poblaciones de arroz rojo de EEUU [Vaughan *et al.*, 2001], se iniciaron estudios preliminares a efectos de explorar las relaciones genéticas entre biotipos de arroz rojo y cultivares uruguayos mediante marcadores AFLP (Amplified fragment length polymorphism; Federicci *et al.*, 2003) con el fin de conocer la capacidad adaptativa del complejo de arroz maleza. Asimismo se evaluó un procedimiento de clasificación asistida por dichos marcadores como herramienta para la discriminación del arroz maleza del cultivar, de manera de ayudar en la prevención de contaminación genética [Capdevielle *et al.*, 2003a]. Para ello, los datos obtenidos fueron analizados con el uso de la tecnología de k-nearest neighbor algorithm (k-NN), implementado en el sistema SAS. Dentro de este análisis, los autores seleccionaron aquellos polimorfismos que permitieran discriminar entre arroz rojo y arroz cultivar, logrando un acierto del 92% con la utilización de solamente 4 loci AFLP de los 256 fragmentos obtenidos; asimismo se obtuvieron porcentajes de clasificación correcta superiores a 90 % para asignación de diferentes tipos de arroz rojo a subgrupos caracterizados por presencia/ausencia de aristas y color de cáscara. En este caso el análisis discriminante permitió clasificar biotipos de arroz rojo con un acierto del 95-100% [Capdevielle *et al.*, 2003a] en base a los loci AFLP seleccionados.

Una de las hipótesis que pueden formularse a partir de las conclusiones de este trabajo, dado el alto porcentaje de acierto en la discriminación entre maleza y cultivar lograda con solamente 4 loci de tipo AFLP, es que sería posible generar una base de información que utilice un número limitado de loci específicos -como los correspondientes a marcadores de tipo microsatélite- para alcanzar un alto grado de precisión en la clasificación de muestras de *O. sativa* L. pertenecientes a diferentes biotipos (arroz rojo versus variedades cultivadas). Como caso de aplicación de este tipo de procedimiento para un sistema de producción de semillas certificadas, estos marcadores podrían permitir la discriminación de lotes conteniendo trazas de arroz rojo (mezclas de semillas), así como detectar la presencia de híbridos entre arroz rojo y variedades cultivadas que no sean diferenciables por otros métodos convencionales.

Los marcadores microsatélites o simples secuencias repetidas (SSR), se encuentran ampliamente distribuidos en el genoma de arroz, presentan un extenso polimorfismo constituyendo loci genéticos, altamente variable, multialélicos y son de gran contenido informativo [McCouch *et al.*, 1997]. Además de las ventajas de esta técnica, se le suma toda la información disponible en bases de datos públicas para el arroz (considerado como un cultivo modelo para estudios genómicos en cereales), donde existe una importante infraestructura de bases de datos (Gramene, Oryzabase, INE, TIGR) que facilita este tipo de estudios, posibilitando la extrapolación de datos desde mapas genéticos ya generados para estudios comparativos, reduciendo así tiempo y costos.

Gealy *et al.* (2002) evaluaron un conjunto de marcadores de tipo microsatélite mapeados en el genoma de arroz que se encuentran en diferentes frecuencias dentro de poblaciones de arroz rojo (AR), cultivares (CV) e híbridos (ARxCV; CVxAR) identificados en el sur de EEUU. Para ello fueron utilizados 18 SSR que en conjunto proporcionaban una cobertura mínima de los 12 cromosomas del genoma de arroz. Los datos obtenidos fueron analizados a través del software SAS IML para estimar las distancias genéticas entre diferentes accesiones, logrando la formación de tres grupos con el uso de PROC CLUSTER (SAS, 1999). En estos grupos no solo quedan discriminadas las accesiones de AR de los CV, sino también quedan en grupos diferentes las accesiones de AR con cáscara negra sin arista o corta y los de cáscara roja con arista. La mayoría de los híbridos incluidos en el estudio se encontraron en el grupo de CV, donde se observan dos subgrupos: en uno se encontraban los híbridos conocidos y los posibles, y en el segundo subgrupo los cultivares y dos híbridos. De esta forma, confirmaron su hipótesis inicial de que era posible identificar diferentes grupos y detectar los híbridos entre poblaciones de diferentes grupos con el uso de esta metodología. En el estudio de las frecuencias alélicas de los tres grupos, se encontró la existencia de al menos 10 alelos exclusivos de cada grupo, los que eran

generados por cuatro microsatélites. Basados en esta observación, los autores utilizaron 4 de los 18 SSR iniciales, elegidos aquellos que al menos presentaban un alelo exclusivo del grupo de AR o de CV, con el fin de disminuir el conjunto de microsatélites necesario para discriminar entre AR, CV e híbridos. La información obtenida a partir de los alelos generados por estos cuatro SSR elegidos no sólo fueron útiles para la diferenciación de híbridos, AC y AR, sino también pueden ser utilizados para apoyar otros estudios sobre la biología del arroz rojo y para determinar el grado de hibridación existente entre AR y AC [Gealy *et al.*, 2002].

## **OBJETIVO**

El objetivo de este trabajo es evaluar un método de clasificación asistida por marcadores moleculares en cuanto a su capacidad para discriminar biotipos de arroz rojo (AR) y cultivares (CV) de arroz desarrollados en Uruguay, así como híbridos CV x AR producidos en forma experimental. Para ello se plantea utilizar información sobre genotipado para un conjunto de marcadores microsatélites (basados en Gealy *et al.*, 2002) - obtenida a partir de accesiones de arroz rojo recolectadas en campos uruguayos y de la región -, para validar un algoritmo de análisis discriminante (k-NN) que pueda ser utilizado en la diferenciación de biotipos de arroz rojo y cultivares uruguayos y en la identificación de híbridos (basado en Capdevielle *et al.*, 2003a).

## **MATERIALES Y MÉTODOS**

### **Material vegetal**

La muestra de la población de arroz rojo sobre la que fue realizado el estudio estuvo compuesta por 43 accesiones uruguayas, 6 accesiones de Río Grande del Sur (Brasil) y 7 italianos. En arroz cultivado se trabajó con 7 de los cultivares más usados en la región y un grupo de 7 híbridos obtenidos mecánicamente en el país. El sistema utilizado para la nomenclatura para cada accesión de arroz se incluye en la Tabla N° 1, en anexo se incluye tabla con información agronómica disponible para algunas accesiones utilizadas. Las semillas de cada accesión fueron germinadas en medio MS + AG3, con temperatura constante de 27° C durante 7 días, para la obtención del material verde necesario para la extracción de ADN.

**Tabla N° 1**

Muestras	Procedencia (s. de colecta)	Tipo	Muestras	Procedencia (s. de colecta)	Tipo
2	Vichadero -1	Rojo	47	Itapebí - 4	Rojo
4	Vichadero - 3	Rojo	48	Itapebí_5	Rojo
5	Tacuarembó - 1	Rojo	49	Itapebí_6	Rojo
9	Tacuarembó - 5	Rojo	51	Itapebí - 8	Rojo
10	Tacuarembó - 6	Rojo	64	Bella Unión_1	Rojo
11	Zapata - 1	Rojo	68	F1(ARP/INIA Zapata) -1	Híbrido
12	Zapata - 2	Rojo	69	F1(ARP/INIA Caraguata)- 2	Híbrido
13	Zapata - 3	Rojo	70	F1(ARP/Bluebelle)-3	Híbrido
14	Zapata - 4	Rojo	71	F1(ARP/INIA Tacuari)-4	Híbrido
15	Zapata - 5	Rojo	72	F1(ARP/El Paso 144)-5	Híbrido
16	Zapata - 6	Rojo	73	F1(ARN/INIA Zapata)-6	Híbrido
17	Río Branco - 1	Rojo	74	F1 (ARN/INIA Caraguatá)- 7	Híbrido
18	Río Branco_2	Rojo	75	F1(ARN/Bluebelle)-8	Híbrido
19	Río Branco - 3	Rojo	76	F1( ARN/INIA Tacuarí)-9	Híbrido
20	Río Branco - 4	Rojo	79	Italia_2 (b177)	Rojo
22	La charqueada_1	Rojo	84	Italia_7 (b197)	Rojo
23	La charqueada - 2	Rojo	86	Italia_9 (b199)	Rojo
24	Rincón de Ramírez	Rojo	87	Italia_10 (b201)	Rojo
26	Desconocida - 1	Rojo	90	Italia_13 b(b208)	Rojo
27	Desconocida - 2	Rojo	94	Italia_17 (b232)	Rojo
28	Desconocida - 3	Rojo	96	Italia_19 (b235)	Rojo
31	INIA Caraguatá	Cultivar	101	Balilla	Rojo
32	INIA Zapata	Cultivar	102	IRGA 417	Rojo
33	INIA Cuaró	Cultivar	105	Río Grande del sur -1	Rojo
34	Bluebelle	Cultivar	106	Río Grande del sur -2	Rojo
35	EEA 404	Cultivar	129	Río Grande del sur -3	Rojo
36	El Paso 144	Cultivar	132	Río Grande del sur -4	Rojo
37	INIA Olimar	Cultivar	135	Río Grande del sur -5	Rojo
43	Itapebí_2	Rojo	144	Río Grande del sur -6	Rojo
46	Itapebí - 3	Rojo			

**Determinación de viabilidad, germinación y dormancia de las semillas**

La viabilidad de todas las accesiones de arroz rojo se estimó utilizando el test de TETRAZOLIUM [Chalam *et al*, 1967], antes de ser sometidas a tratamientos de ruptura de dormancia. Para ello, se embebieron 10 semillas (sin cáscara), en agua destilada durante 3 horas a 28°C. Luego se cortaron longitudinalmente por el medio del embrión. Estas fueron incubadas en una solución de Tetrazolium 1% por 1 hora a 40°C en oscuridad. A continuación se lavaron tres veces en agua destilada para remover el exceso de solución. Las semillas que presentaban el

embrión completamente teñido (coloración rojiza), fueron las consideradas viables.

Aquellas semillas que demostraron ser viables pero que no germinaron en las condiciones descritas fueron sometidas a diferentes tratamientos para la ruptura de la dormancia (Naredo *et al.*1998; Com. Pers. Gealy). Cada uno de los tratamientos se detalla en la tabla N° 2. Para cada uno de ellos fueron utilizadas 10 semillas intactas y estériles de tres biotipos: uno de buena germinación (control), otro de germinación regular y un tercero que nunca había germinado.

**Tabla N° 2:** Tratamientos para ruptura de dormancia

Tratamiento	Condiciones de la semilla	Condiciones de germinación		
1 - 50 °C 7 días	Sin Cáscara *1	30 °C	Luz (12/12)	en gasa
2 - 50 °C 14 días	Sin cáscara	30 °C	Luz (12/12)	en gasa
3 - 1000 ppm GA3 48hs *2	Con cáscara	30 °C	Luz (12/12)	en gasa
4 - 0.1 M HNO <sub>3</sub> 48 hs *2	Con cáscara	30 °C	Luz (12/12)	en gasa
5 - 1,4 mM GA3 + 4 mM KNO <sub>3</sub> 48 hs	Con cáscara	30 °C	Luz (12/12)	en gasa

\*1: Se saca la cáscara luego del tratamiento con calor,

\*2: agregado antes de la germinación. GA3: Ácido Giberélico.

Luego de seleccionado el tratamiento, todas las semillas de las accesiones que presentaban porcentajes de viabilidad entre 60 – 100 % fueron tratadas. Las semillas que germinaron luego del tratamiento de ruptura de dormancia fueron pasadas a medio (MS + GA3) en tubo Eppendorf de 2,5 ml a 25° C durante 24 horas, para un mejor desarrollo del brote. Luego de las 24 horas, se pasaron a medio líquido MS + BAP a 25° C, para la obtención de la cantidad suficiente de material verde (plántulas de 5 cm.). Para la ruptura de la dormancia y el estímulo de su germinación se colocaron 10 semillas intactas y estériles de cada accesión con porcentajes de viabilidad de entre 60 a 100%, en placas de petri sobre papel de filtro embebido en una mezcla de 1,4 mM ácido giberélico (GA3) y 4 mM nitrato de potasio (KNO<sub>3</sub>). Se incubaron en cámaras de germinación a temperatura constante de 30°C, durante 48 horas.

## **Aislamiento del ADN**

Para el aislamiento del ADN genómico a partir del material verde obtenido bajo las condiciones antes descritas, se utilizó el protocolo de extracción indicado en el Manual "Mutant germplasm characterization using molecular markers" (FAO/IAEA), con modificaciones menores. Se molieron de 2 a 4 g de tejido vegetal en mortero con nitrógeno líquido, colocando el material finamente molido en tubos con 5 ml de buffer de extracción precalentado a 65°C (2% CTAB, 1,4 M NaCl, 20 mM ácido etilendiamintetracético [EDTA] pH 8, 100 mM Tris pH 8, 2% PVP – 40, 0.125% de  $\beta$  – mercaptoetanol agregado en el momento). Luego de incubar las muestras a 65°C durante 20 minutos con agitación suave y periódica, se le agregó 1 volumen de Cloroformo-isoamílico (24:1) y se mezcló cuidadosamente. Posteriormente se centrifugaron a 12000 rpm durante 20 minutos a 4°C. El sobrenadante fue transferido a otro tubo, donde se precipitó el ADN con 0.7 volúmenes de isopropanol. Las muestras fueron centrifugadas a 5000 rpm durante 20 minutos a temperatura ambiente (TA). Los precipitados fueron lavados con etanol 70% y centrifugados a 5000 rpm durante 5 minutos.

En el caso de arroz rojo, 30 accesiones no germinaron, por lo que se determinó su viabilidad y sus condiciones de dormancia para determinar el material de partida para la extracción de ADN de acuerdo al método descrito en la siguiente sección. Para el aislamiento del ADN genómico de semillas no viables, se utilizó el mismo protocolo de extracción pero partiendo de 5 a 10 granos de semillas molidas, previa remoción de cáscara. Los precipitados de ADN obtenidos en ambos procedimientos fueron resuspendidos en 100  $\mu$ l de agua milliQ estéril, los que fueron cuantificados por electroforesis en gel de Agarosa 0,9 % y revelado con tinción de bromuro de etidio. Se utilizó 5  $\mu$ l de MassRuler (103ng/ $\mu$ l, Fermentas) como marcador de peso molecular. La concentración de ADN se estimó en gel por comparación con el marcador de peso molecular y se corroboró por medición espectrofotométrica. La concentración de todas las muestras fue ajustada a 100 ng/ $\mu$ l

## **Marcadores microsatélites**

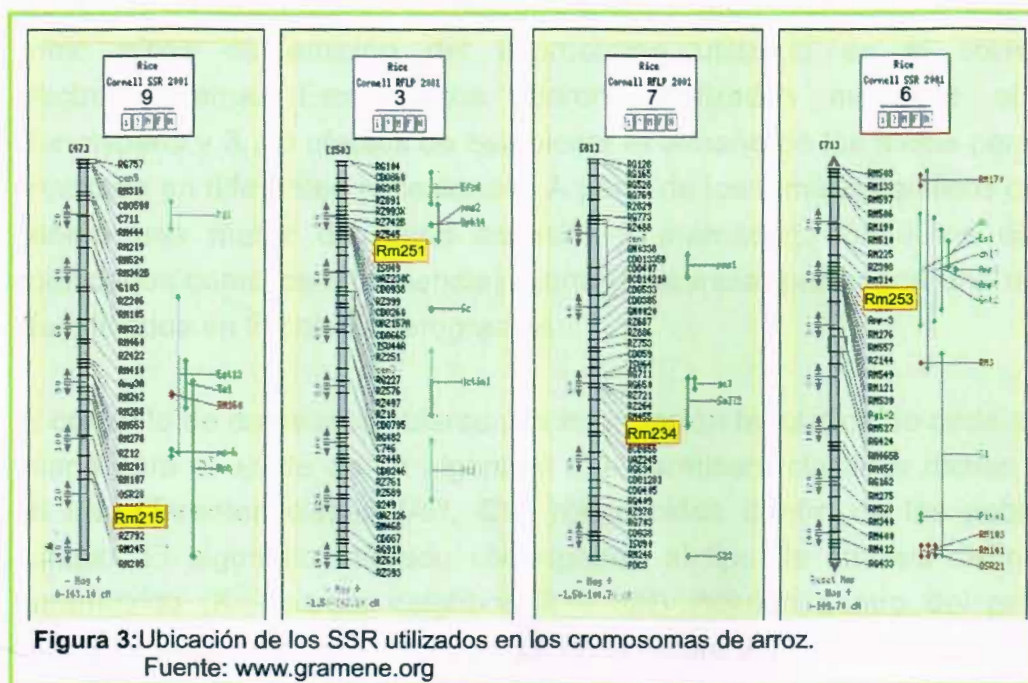
Los microsatélites utilizados fueron RM215, RM234, RM251 y RM253, los que se encuentran mapeados en el genoma de arroz (figura 3, Tabla N° 3). La selección de estos cuatro microsatélites, fue basada en estudios previos [Gealy *et al*, 2002] que indicaron que estos presentaban diferentes frecuencias entre biotipos de arroz rojo y cultivares de arroz utilizados en EEUU. Los cebadores (oligonucleótidos) para estos marcadores fueron obtenidos comercialmente. La reacción en cadena de la polimerasa (PCR) para las reacciones de amplificación

fue ajustada previamente. En la tabla N° 3 se detallan los datos reportados para cada uno de los SSR, en la figura 3 se muestra el mapeo en los cromosomas de arroz de cada uno de los microsatélites.

**Tabla N° 3:** Microsatélites utilizados para el genotipado de la población de estudio.

Fuente: [www.gramene.org](http://www.gramene.org)

Marcadores microsatelites de arroz	N° Cromosoma	N° de alelos Reportados	Tipo de repetición y largo	Rango de tamaño (pb)
RM 215	9	4	(CT)16	147 - 153
RM 234	7	7	(CT)25	133 - 163
RM 251	3	7	(CT)29	114 - 150
RM 253	5	7	(GA)25	125 - 143



**Figura 3:** Ubicación de los SSR utilizados en los cromosomas de arroz.

Fuente: [www.gramene.org](http://www.gramene.org)

Para la reacción se PCR se utilizaron 20 µl de reacción conteniendo 20 mM de cebadores, 5 mM desoxirribonucleótidos(dNTPs), 5 µM desoxirribonucleótidos marcados con fluoro cromos (relación 1000:1 con respecto a dNTPs) (Applied Biosystem), 1,5 mM MgCl<sub>2</sub> (Applied Biosystem), buffer 10X (Applied Biosystem), 200 ng ADN molde, 0,5 unidades de Taq polimerasa (5U/ul) (Applied Biosystem) sujeto al siguiente ciclo de PCR: 1 ciclo a 95°C por 7 minutos, 35 ciclos de 95°C por 1 minuto, 55°C por 1 minuto, 72°C por 1 minuto, 72°C por 5 minutos. Para el control de la reacción se utilizaron geles de agarosa 2% revelado con tinción de bromuro de etidio, se utilizaron 3 µl de generuler 100 pb (0,5µg/µl, Fermentas)



como marcador de peso molecular. Los productos de amplificación fueron separados mediante electroforesis en capilar usando el módulo FAM-JOE-TAMRA. 1  $\mu$ l de las amplificaciones fueron mezcladas con 11.7  $\mu$ l de Hi-Di™formamide (PE Applied Biosystem, Foster City, CA) y 0.3  $\mu$ l de marcador de peso molecular de ADN marcado con ROX (GeneScan®-500 [ROX]™)(PE Applied Biosystem, Foster City, CA). Las muestras se desnaturalizaron a 95°C por 5 min, se mantuvieron en hielo durante 5 min y se cargaron en un secuenciador automático ABIPRISM 310 Genetic Analyzer (PE Applied Biosystem, Foster City, CA). Los parámetros de corrida fueron los siguientes: tiempo de inyección 10 a 15 s, voltaje de la electroforesis 15 kv, tiempo de colección de datos 20min y temperatura de corrida 60°C.

### **Análisis de datos**

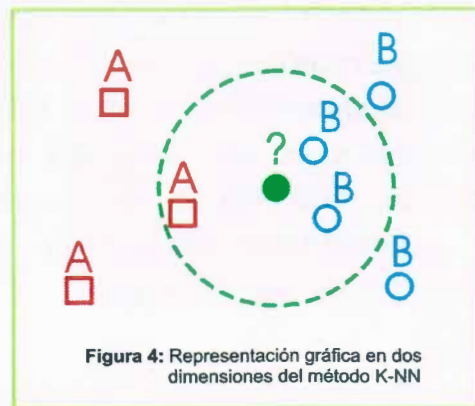
El análisis primario de los datos obtenidos del secuenciador ABI 310 se realizó con el programa GeneScan® (PE Applied Biosystem, Foster City, CA), donde los productos de amplificación correspondientes a cada microsatélite son observados como picos de emisión del fluorocromo utilizado en el correspondiente electroferograma. Estos picos fueron analizados mediante el programa Genotyper® v 3.7 a efectos de establecer el tamaño de los alelos para cada SSR analizado en diferentes accesiones. A partir de los tamaños alélicos obtenidos se generó una matriz de datos accesión x marcador, donde los datos fueron codificados como cero (ausencia) ó uno (presencia) para cada uno de los alelos identificados en los electroferogramas.

El conjunto de datos moleculares y la información fenotípica de cada accesión fue usada para el ajuste de un algoritmo que permitiera clasificar dichas accesiones en las diferentes clases (AR, CV) reconocidas dentro de las poblaciones en estudio. El algoritmo utilizado corresponde al tipo de análisis discriminante no paramétrico (K -nearest neighbor, k - NN) incluido dentro del procedimiento DISCRIM implementado en SAS/STAT® software 9.1) [SAS Institute, 2002-2003], utilizando una combinación de variables como predictores de la clase a la que pertenece cada observación (en este caso las variables predictivas son los marcadores moleculares analizados).

El algoritmo k-nearest neighbor clasifica cada observación del conjunto en estudio (sobre las que se quiere tener una predicción), en base al conjunto de referencia definido (aquellas observaciones para las que se conoce su clase) de acuerdo a los siguientes criterios: 1) encuentra los k vecinos más cercanos (K=1 en este caso) dentro de las observaciones del conjunto de referencia, y 2) predice la clase según la del vecino más cercano (si K>1, utiliza la clase más frecuente), por ejemplo, elige la clase que es más común entre los k - vecinos más cercanos

[Capdevielle, 2001]. Para realizar las predicciones en clases se definió el vecindario (neighborhood) por la distancia desde  $x$  a los  $k$ th puntos más cercanos del conjunto de referencia; luego se estima la probabilidad de que pertenezcan a la clase  $j$ , por la proporción de los puntos de referencia que pertenecen a dicha clase entre los  $k$  más cercanos a  $x$ .

**Figura 4:** Representación gráfica en dos dimensiones del método K-NN. Los círculos y cuadrados representan las observaciones integrantes del conjunto de referencia. El círculo punteado representa el vecindario definido para ese conjunto de observaciones.

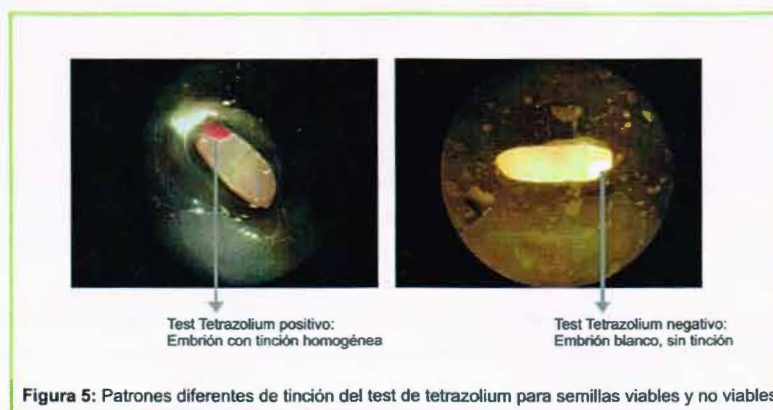


Utilizando la proporción de los puntos de referencia que pertenecen a una clase en particular (por ejemplo, arroz rojo versus cultivares) dentro de los en el “conjunto de entrenamiento” como la estimación de  $f(j | x)$ , la clasificación de una línea estaría dada por la “etiqueta” (AR ó CV) que corresponde al grupo de  $k$  vecinos más cercanos con la mayor estimación de  $f(j | x)$ . Cada observación en el conjunto de referencia es tratada por turnos (método “leave-one-out” de validación cruzada) como si fuera parte de un conjunto de observaciones a ser clasificadas (conjunto de estudio). Las distancias de todas las otras observaciones del conjunto de referencia son computadas respecto a cada observación incluida en el conjunto de estudio, y clasificadas utilizando el algoritmo k.NN. La clasificación para cada una de las observaciones del conjunto en estudio son comparadas con las “etiquetas” previamente disponibles para los miembros de cada grupo definido a priori (AR ó CV), lo que permite obtener por validación cruzada un rango de errores (y sus correspondientes aciertos) para el procedimiento de clasificación asistida por marcadores [Capdevielle et al., 2003b]

## RESULTADOS Y DISCUSIÓN

### Determinación de la viabilidad y germinación

La determinación de la viabilidad de las semillas se realizó como fue descrito en materiales y métodos con el Test de TETRAZOLIUM (test de viabilidad) [Chalam et al., 1967]. Éste tiene la capacidad de diferenciar entre células vivas y muertas, basado en la actividad de la enzima deshidrogenasa (enzima de la respiración) ubicada en los embriones, como resultado de la liberación de iones hidrogeno, los que reducen a la solución de tetrazolium (2,3,5 trifenil tetrazolium clorado), formando el compuesto llamado FORMAZAN, el que tiene una coloración rojiza [www.seedlab.oscs.orst.edu/Page\_Technical\_Brochures/TZ, revisado 22/07/2007]. La viabilidad de las semillas es interpretada de acuerdo a los patrones de tinción de los embriones y la intensidad de la coloración. En la figura 5 se muestran ejemplos de los patrones de coloración obtenidos para las semillas testeadas (embriones rojos – viables, embriones blancos – no viables).



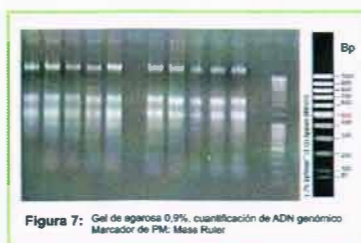
Los porcentajes de viabilidad fueron calculados para cada una de las accesiones testeadas. Las que presentaron porcentajes de viabilidad entre 60 - 100 %, fueron consideradas como semillas viables en estado de dormancia, dado que la dormancia – suspensión temporal visible del crecimiento del meristema – es una de las características más relevantes de esta maleza. Para la germinación de las accesiones dormantes fue necesario realizar tratamientos utilizando temperaturas y productos químicos que induzcan la ruptura de dormancia. Para ello se evaluaron cinco tratamientos en tres muestras que presentaban diferentes valores estimados de germinación: una de buena germinación, otra de germinación regular y una tercera que nunca germinó.

Para los tratamientos basados en la exposición de las semillas a altas temperaturas (tratamientos 1 y 2), no se observó ruptura de dormancia, mientras que para los tratamientos basados en compuestos químicos: Ác. Giberélico (hormona que estimula el crecimiento de los meristemas y activador del crecimiento de los embriones somáticos in Vitro) y  $KNO_3$  (estimulador de la germinación), sí se observó ruptura de dormancia (hubo germinación). En los tratamientos 3 (1000 ppm GA3 48hs) y 4 (0.1 M  $HNO_3$  48 hs) el brote observado fue mucho más pequeño y débil que el observado en el tratamiento 5 (1.4 mM GA3 + 4 mM  $KNO_3$ ). Este último fue seleccionado para el tratamiento de todas las muestras que presentaron entre 60 – 100 % de viabilidad, en base una mejor y más eficiente ruptura de la dormancia.

Luego del tratamiento de ruptura de la dormancia, las semillas con brote, fueron pasadas a tubos Eppendorff con medio MS + GA3, para favorecer un mejor desarrollo de la plántula. Cuando las plántulas alcanzaron los 5 cm., fueron subcultivadas en tubos con medio MS liq + BAP para un mejor desarrollo y mayor obtención de material verde. La secuencia seguida se muestra en figura 6.



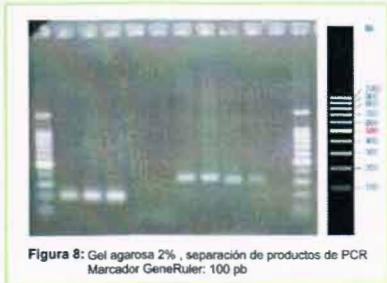
### Extracción de ADN y obtención de SSR



Mediante la utilización del protocolo de extracción descrito se obtuvo el banco de ADN integrado por muestras de cada accesión (AR, CV e híbridos) incluida en la población en estudio. La calidad del ADN se calculó en base al coeficiente entre la medida de la absorbancia a  $\lambda = 280$  nm (longitud de onda donde absorben los ácidos nucleicos) y a  $\lambda = 260$  nm (longitud de onda donde absorben las proteínas) y se consideraron aceptables aquellos que tuvieron concentraciones mayores a 500 ng/ul y un coeficiente de calidad entre 1,7 – 1,9. El rango de aceptación de la calidad del ADN fue definido en base a alta especificidad y sensibilidad de los marcadores microsatélites. Como criterio adicional para evaluar

la calidad de las muestras de ADN se corrieron geles de agarosa para corroborar las lecturas espectrofotométricas obtenidas (figura 7).

La obtención de la información molecular para cada muestra se llevó a cabo a través de la amplificación de los SSRs seleccionados mediante la reacción de PCR, siguiendo el procedimiento descrito antes. La verificación de estas reacciones se realizó a través de electroforesis en geles de agarosa 2% (figura 8), como etapa previa al análisis de las muestras mediante el secuenciador automático ABI 310. Paralelamente con las muestras se corrió un marcador de peso molecular (100 pb)



validando la corrida electroforética y permitió la estimación de los tamaños de los productos de amplificación.

A partir de los resultados anteriores, y considerando la información extraída de la base de datos de Gramene ([www.gramene.org](http://www.gramene.org)) acerca de los tamaños esperados para los amplicones de los microsatélites en uso, se establecieron los parámetros de corrida para detectar los fragmentos mediante electroforesis capilar en el secuenciador; en este trabajo se utilizó una resina de genotipado (POP 4: performance optimized polymer, para corridas rápidas) para el llenado del capilar. La utilización de métodos automatizados de mayor procesividad (utilizando detección por fluorescencia) permite el análisis de una gran cantidad de genotipos en menor tiempo.

Las muestras son separadas dentro del capilar electroforéticamente, y la detección de los fragmentos se realiza cuando los mismos atraviesan una “ventana” existente en el capilar, donde se combinan un emisor de luz láser y un detector de fluorescencia. Este sistema permite detectar los fragmentos amplificados de ADN que fueron marcados por incorporación de dNTPs unidos a fluorocromos que fueron incluidos en la reacción de PCR. El monitor de fluorescencia de tipo CCD detecta los espectros de emisión correspondientes a cada fluorocromo utilizado, interpretando las señales en función del tiempo de corrida como electroferogramas [User Manual ABI 310 Genetic Analyzer]. Los resultados obtenidos del secuenciador automático fueron analizados primariamente con el software ABI Genescan®, donde se determinaron los tamaños en pares de bases de cada uno de los picos obtenidos en la corrida. Seguidamente se utilizó el software ABI Genotyper®, con el que se determinaron los tamaños alélicos utilizados para los siguientes análisis. En caso de existir diferentes perfiles en los picos identificados en un electroferograma se realizaron correcciones manuales en los valores que el

software utiliza por defecto en comparación con estándares de peso molecular, a efectos de eliminar posibles errores en la determinación automática de los tamaños. En la figura 9 se muestra como ejemplo uno de los electroferogramas obtenidos.

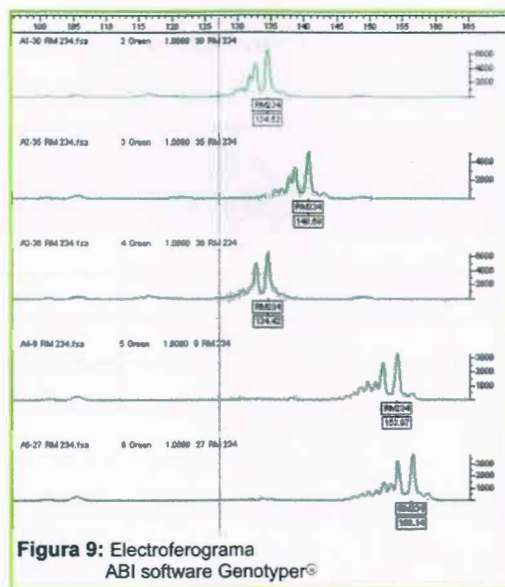


Figura 9: Electroferograma  
ABI software Genotyper®

Los errores en la determinación de los tamaños alélicos pueden deberse al deslizamiento (“slippage”) de la Taq polimerasa y/o al agregado de bases extras en el fin de cada fragmento durante la reacción de PCR, observándose picos no definidos o los “llamados picos sombra” [Ginot *et al.*, 1996]. Estos tipos de errores son detectables al observar la morfología de los picos (figura 10), pero en el análisis automático no serían detectados como errores, por lo que este tipo de artefactos experimentales podrían ser tomados como un tamaño válido de alelo en un análisis no supervisado.

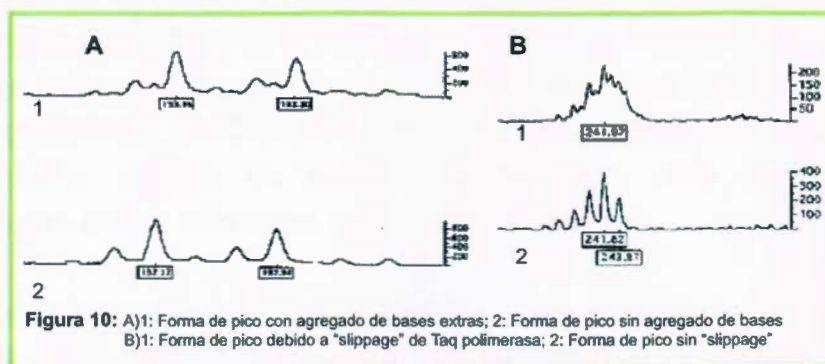


Figura 10: A)1: Forma de pico con agregado de bases extras; 2: Forma de pico sin agregado de bases  
B)1: Forma de pico debido a “slippage” de Taq polimerasa; 2: Forma de pico sin “slippage”

Se observaron diferencias en la cantidad de alelos para los cuatro microsatélites, con respecto a lo reportado en Gramene para cultivares de arroz. Para los loci RM215, RM234, y RM251 se encontró mayor cantidad de alelos por locus, mientras

que para RM253 fue menor. En la tabla N° 4 se detallan la cantidad de alelos hallados en este trabajo y los reportados previamente para cada uno de los microsatélites.

**Tabla N°4:** Número de alelos encontrados y reportados para cada locus de tipo SSR analizado.

Microsatélites	N° de alelos observados	N° de alelos Reportados
RM 215	7	4
RM 234	8	7
RM 251	10	7
RM 253	6	7

Los datos moleculares obtenidos junto con la información agronómica de cada una de las accesiones se integraron en una matriz utilizada como referencia para la clasificación de accesiones en las diferentes clases pre-definidas. La clasificación de datos es el proceso que encuentra propiedades en común de un grupo de objetos en una base de datos y usa esas propiedades para construir un modelo para la clasificación de éstos en diferentes clases [Johnson y Wichern, 1998]. Para la construcción del modelo de clasificación, una base de datos de muestras es tratada como conjunto de entrenamiento. Cada observación incluida en el conjunto de entrenamiento tiene información sobre el mismo conjunto de variables y está asociada a una clase (identidad) [Capdevielle *et al.*,2003b] . En este estudio el conjunto de entrenamiento se compuso con los datos moleculares (variables) obtenidos para cada uno de los microsatélites y la información agronómica (clase fenotípica, definida como AR ó CV) para cada una de las accesiones que integraban la población en estudio. El método K – NN (K nearest neighbor), implementado dentro del procedimiento DISCRIM de SAS/ STAT® 9.1 [SAS Institute 2002-2003], fue utilizado como método de análisis discriminante para llevar a cabo la clasificación dentro de las diferentes clases fenotípicas.

Los datos obtenidos fueron utilizados para establecer el conjunto de entrenamiento que permitió el ajuste del procedimiento antes descrito (k-NN), donde la totalidad de las observaciones y variables fueron incluidas para este análisis. En el conjunto de entrenamiento estuvieron incluidos también 9 híbridos (muestras sin asignación de clase previa), los cuales actuaron como muestras problema que representan casos de aplicación del procedimiento desarrollado. Los porcentajes de clasificación correcta estimados por validación cruzada se muestran en la tabla N° 5:

**Tabla N° 5:** Porcentajes de clasificación obtenidos para los datos utilizados para ajustar el procedimiento k-NN (resumen de validación cruzada usando el método "leave-one-out" implementado en SAS)

Número de observaciones y porcentaje clasificado en tipo			
De TIPO	Cultivar	Rojo	Total
	1	8	9
	11,11	88,89	100,00
Cultivar	6	1	7
	85,71	14,29	100,00
Rojo	0	43	43
	0	100,00	100,00

Las accesiones de arroz rojo en este caso fueron clasificadas dentro de la clase correspondiente con porcentaje de acierto cercano al 100% (considerando tamaño de muestra utilizado), mientras que uno de los cultivares fue clasificado como arroz rojo, lo cual sería considerado como un falso positivo. Del total de los 9 híbridos incluidos en el análisis son clasificados como rojos (resultado esperado) 8 de 9 (88.89%). En la tabla N° 2 del anexo se muestra el detalle de clasificación para cada uno de las accesiones.

El cálculo de estimación del error se basó en la comparación del error que generaría clasificar por azar las muestras en cada una de las clases (AR ó CV) versus la clasificación utilizando el modelo generado por el algoritmo k-NN (cada observación se clasifica respecto al conjunto de referencia definido). La comparación de los errores estimados mediante validación cruzada permite evaluar la capacidad predictiva del algoritmo utilizado para discriminar entre las clases AR y CV. En el esquema siguiente se muestran las tasas de errores para una clasificación basada en el azar (considerando una probabilidad "anterior" de 0.5, es decir que cualquier accesión pudiera ser considerada como AR ó CV) respecto a la utilización del procedimiento k-NN.

	CULTIVAR	ROJO	Total
Tasa error	0.1429	0.0000	0.0714
Anteriores	0.5000	0.5000	

Como se observa en estos resultados, la utilización del modelo generado por el algoritmo k-NN para la clasificación permitió disminuir considerablemente el error respecto a lo esperado por azar. El porcentaje de clasificación para las accesiones de arroz rojo fue cercano a 100% cuando se utilizó clasificación asistida por marcadores mientras que si fuera por azar sería de 50%. Para la clasificación en



cultivares la tasa de error también disminuye, pasando de 50 % de clasificación correcta por azar a aproximadamente 86% de clasificación correcta utilizando k-NN. Se estimó luego para el mismo modelo la tasa de error pero esta vez tomando en cuenta las proporciones, es decir considerando la cantidad accesiones pertenecientes a cada una de las clases (AR ó CV) en el conjunto de muestras utilizado para ajustar el modelo de clasificación para K-NN. Los valores estimados de error en este caso fueron los siguientes:

	CULTIVAR	ROJO	Total
Tasa	0.1429	0.0000	0.0200
Anteriores	0.1400	0.8600	

En este caso las diferencias de error entre la asignación al azar de las muestras a cada clase respecto a la clasificación utilizando el modelo k-NN son menores, debido a que la población en estudio está compuesta en un 90% de muestras de arroz rojo. En este caso sin embargo no se logra la clasificación correcta de la totalidad de los híbridos AR x CV considerados como muestras problema; por lo tanto la utilización de un procedimiento clasificatorio incluyendo la totalidad de los alelos identificados no cumpliría con el requisito de clasificar las muestras correspondientes a híbridos en la clase AR, teniendo como resultado un falso negativo.

La importancia de ajustar el algoritmo de clasificación para poder identificar los híbridos como parte de la clase AR se debe a la necesidad de detectar contaminación de genoma rojo existente debido a posibles flujos genéticos entre muestras de tipo AR y cultivares. Aunque el arroz es una especie autógama presenta porcentajes de alogamia (0.5%) [Gealy *et al.* 2003], lo que permitiría el cruce entre estos, dado que por pertenecer a la misma especie son sexualmente compatibles. A través del flujo génico entre arroz rojo y cultivares, los tipo maléza podrían adquirir características de los cultivados, dificultando su identificación a través de los métodos tradicionales. En este caso si bien el uso de todas las variables como clasificadores en el modelo utilizado fue capaz de describir las diferencias entre los cultivares y las accesiones de arroz rojo, no fue posible detectar casos que corresponderían a posibles contaminaciones genéticas (AR x CV). El origen del error en la clasificación puede tener diferentes explicaciones dadas por causas biológicas o atribuibles a una falta de ajuste del procedimiento discriminante utilizado.

Si el error hubiera tenido causas biológicas se centraría principalmente en el modo de cómo son controladas y heredadas las características que determinan las mayores diferencias fenotípicas entre arroz rojo y cultivares. Éstas son características multigénicas, sin conocimiento previo para las malezas de cuanto

aporta cada uno de los genes implicados para determinar la característica [Dekkers y Hospital, 2002]. Podría entonces el híbrido, producto del cruzamiento de un AR (dador) y un CV (receptor), no haber incorporado en su genoma las regiones aportadas por el dador donde se ubican las variantes alélicas que determinan las diferencias principales entre los mismos. Otra causa biológica posible podría ser que la utilización de un número limitado de loci no permitiera explorar un conjunto de regiones del genoma donde se pudieran encontrar diferencias que permitieran discriminar entre muestras de arroz rojo y cultivares uruguayos. Pero de haber sido así no se hubieran hallados los porcentajes de clasificación correcta de todas las accesiones de arroz rojo, mostradas en la tabla N° 5.

Otra de las causas posibles de error se refiere a una falta de ajuste del algoritmo utilizado en cuanto a su capacidad predictiva al ser aplicado a nuevas accesiones (AR x CV) no incluidas en el conjunto de entrenamiento (incluyendo AR y CV exclusivamente). Este tipo de error es conocido como sobre-ajuste ("overfitting") y afecta a los algoritmos utilizados en clasificación cuando el número de variables predictivas utilizadas es demasiado grande con respecto al tamaño de la población en estudio. Para evitar este problema se deben utilizar técnicas adicionales para seleccionar un número menor de variables de alto valor discriminante, a efectos de mejorar la capacidad del modelo ajustado para ser generalizable a futuros casos de uso [Tetko et al., 1995].

A efectos de optimizar la clasificación desde el punto de vista de la identificación de los híbridos AR x CV se seleccionó un subconjunto de variables (tabla N° 3 anexo). La elección de las variables se realizó mediante el procedimiento SAS PROC STEPDISC®, el cual categoriza las variables según el grado de contribución a la clasificación. Este asigna valores para el estadístico F según el grado de información aportado por la variable en cuanto a discriminar entre las clases pre-definidas (AR ó CV), es decir que selecciona aquellas variables que maximizan las diferencias entre las observaciones presentes en la población en estudio. El método usado dentro del sistema SAS (proc stepdisc) para calcular el valor F para las n variables seleccionadas se denomina "forward" (paso hacia delante), donde la variable de mayor valor F es seleccionada en cada iteración. Una vez incorporada una nueva variable al modelo, el valor F es recalculado dentro del resto de las variables -esta vez sin la variable seleccionada anteriormente- quedando seleccionada de igual forma la segunda, repitiendo el procedimiento en sucesivas iteraciones hasta incorporar un total de n variables que contribuyan mayormente a la clasificación; de esta forma se genera un subgrupo de variables que maximizan la discriminación entre clases a la población.

El algoritmo K-NN fue corrido nuevamente, utilizando esta vez al subgrupo de variables seleccionado por SAS PROC STEPDISC, obteniendo los nuevos

porcentajes de clasificación para cada una de las observaciones, basados en el conjunto de variables (marcadores) seleccionados. Estos porcentajes se detallan en la tabla N° 6 y a continuación las estimaciones de los errores, calculados de igual manera que en el caso anterior.

**Tabla N° 6:** Porcentajes de clasificación obtenidos utilizando variables seleccionadas (resumen de validación cruzada usando vecino más cercano).

Número de observaciones y porcentajes clasificado por tipo			
DE TIPO	Cultivar	Rojo	Total
	0	9	9
	0	100	100
<b>Cultivar</b>	6	1	7
	85,71	14,29	100
<b>Rojo</b>	0	43	43
	0	100	100

Las estimaciones de errores para este caso fueron:

1 – sin tener en cuenta las proporciones de cada clase

	CULTIVAR	ROJO	Total
Tasa	0.1429	0.0000	0.0714
Anteriores	0.5000	0.5000	

2- Teniendo en cuenta las proporciones de cada clase

	CULTIVAR	ROJO	Total
Tasa	0.1429	0.0000	0.0200
Anteriores	0.1400	0.8600	

Si bien los valores globales de clasificación correcta son similares a los obtenidos anteriormente, la diferencia principal con la primera corrida del algoritmo k-NN es que en este caso se logra la clasificación correcta del 100% de los híbridos incluidos en el análisis (detalle de clasificación en tabla N° 4 anexo), solucionando de esta forma el error que habría ocurrido al utilizar todas las variables como clasificadores, con el consiguiente sobre-ajuste del modelo de clasificación. En este caso, considerando el objetivo planteado, el ajuste de un procedimiento discriminante basado en información de un conjunto reducido de marcadores permite no solamente discriminar entre muestras correspondientes a biotipos con



características de maleza y cultivares, sino también detectar la presencia de posibles fuentes de contaminación genética (híbridos AR x CV).

Además de la información de tipo (arroz rojo, cultivar) también se contó con descripciones morfológicas de las semillas y plantas, como presencia o largo de aristas, ápice, color de cáscara, hábito de crecimiento (Tabla 1 anexo). Estas características son utilizadas como descriptores morfológicos a la hora de clasificar las accesiones de arroz tipo maleza y entre las mismas se encuentran algunas de las principales diferencias fenotípicas respecto a los cultivares de arroz. A efectos de estimar si el procedimiento era capaz además de describir las diferencias entre los tipos maleza y cultivares, sino también dentro de la población de rojos estudiada, el conjunto de entrenamiento fue redefinido para incluir accesiones que contaban con información morfológica, para las que se dispone de información sobre todas las variantes moleculares (alelos) determinadas. Los resultados de la clasificación se muestran en la tabla N° 7.

**Tabla N° 7:** Porcentajes de clasificación obtenidos para los datos calibrados (resumen de validación cruzada usando vecino más cercano).

Número de observaciones y porcentajes clasificado por tipo			
DE TIPO	Cultivar	Rojo	Total
	1	1	2
	50	50	100
<b>Cultivar</b>	6	0	6
	100	0	100
<b>Rojo</b>	0	26	26
	0	100	100

Cuando el procediendo es corrido para clasificar por tipo, nuevamente observamos resultados similares a la primer corrida, donde las muestras correspondientes a arroz rojo son clasificados correctamente al igual que los cultivares, y los dos híbridos incluidos quedan clasificados uno como rojo y otro como cultivar, por lo que se estaría nuevamente en presencia de un falso negativo. Los resultados primarios del PROC DISCRIM se muestran en la Tabla N° 5 de anexo, donde se observa la clasificación de cada muestra individual.

Siguiendo el enfoque propuesto anteriormente para solucionar el error generado (overfitting) por el exceso de variables (marcadores) incluido en comparación con el número de observaciones, nuevamente se seleccionó un subgrupo de marcadores a través del procedimiento SAS PROC STEPDISC, método forward. (tabla N° 6 anexo)

Una vez obtenido el subgrupo de variables fue corrido el procedimiento SAS PROC DISCRIM para este conjunto de datos. Los porcentajes de clasificación obtenidos se muestran en la tabla N° 8

**Tabla N° 8:** Porcentajes de clasificación obtenidos para los datos calibrados  
( resumen de validación cruzada usando vecino más cercano)

Número de observaciones y porcentajes clasificado por tipo			
DE TIPO	Cultivar	Rojo	Total
	0	2	2
	0	100	100
Cultivar	5	1	6
	83,33	16,67	100
Rojo	0	26	26
	0	100	100

Como se observa en la tabla de resultados el porcentaje de clasificación para las accesiones de arroz rojo se mantiene como en el caso anterior de 100%, mientras que en el caso de los cultivares es menor, aproximadamente de 84%. Sin embargo los dos híbridos AR x CV incluidos son clasificados correctamente. Para el objetivo trazado este subgrupo de variables sería entonces de mayor valor predictivo en comparación con el total de variables. En la tabla N° 7 del anexo se muestra el detalle de la clasificación para cada una de las accesiones.

Si se observa la tabla N° 7, donde se detalla la clasificación para cada una de las observaciones, se nota que el cultivar etiquetado como similar a los biotipos de arroz rojo (falso positivo) es EEA 404. Este cultivar es considerado de tipo tradicional (menor adaptación a sistemas de cosecha mecanizada) con características morfo-fisiológicas diferentes a la de la mayoría de los cultivares y se encuentra fuera de los sistemas de selección hace mucho tiempo. Esto podría explicar porque dicho cultivar podría presentar mayor similitud a los biotipos con características de maleza que respecto a las variedades modernas.

A los cambios que permiten la obtención de mejores variedades se los conoce como síndrome de domesticación [Li *et al.*, 2006]. En general, los cambios que más se destacan en la diferenciación entre una variedad sometida a mejoramiento (presión de selección artificial) y las que no los están (selección natural), son con respecto a ciertas características adaptativas que faciliten el manejo agronómico de las formas cultivadas, buscando por ejemplo la reducción del desgrane y dormancia de las semillas, la sincronización de la maduración, el incremento del número de inflorescencias y del rendimiento de granos [Hancock, 2004].

Fue posible suponer a priori que entre las variedades modernas y los biotipos con características de maleza debían existir diferencias a nivel de genoma, dadas las

diferencias fenotípicas que existen entre ellos, y que por lo tanto podrían encontrarse evidencias de asociación entre los marcadores discriminantes y características vinculadas con el síndrome de domesticación. Es de conocimiento que el grupo maleza no es uniforme, también existen diferencias fenotípicas entre ellos, alguna de ellas se enumeran en la tabla N° 1 del anexo. En consecuencia se discute la posibilidad de que el grupo de microsatélites utilizados para la discriminación entre accesiones de arroz rojo y cultivares pudieran también describir las diferencias fenotípicas presentes en la población de arroz maleza estudiada.

Para ello se utilizó el conjunto de muestras con características morfológicas, la misma cantidad de variables, pero esta vez se seleccionó cada una de las características: color de la cáscara, ápice, arista, hábitos de crecimiento como clase, utilizando el mismo procedimiento de clasificación (k-NN) que en los casos anteriores. Debido al pequeño tamaño de muestra disponible para estos casos el procedimiento de validación cruzada fue reemplazado por una re-sustitución de las observaciones dentro del conjunto de entrenamiento (lo que determina menor precisión en las estimaciones de error). La asignación a las clases fenotípicas definidas se puede ver a través de los resultados de la resustitución mostrada en la tabla N° 9, donde se encuentran para cada una de las clases y la asignación a cada una de ellas de las observaciones.

**Tabla N° 9:** Resumen de resustitución usando algoritmo k-NN

Número de observaciones y porcentajes clasificado por tipo				
color cascara	Negro	Ocre	Paja	Total
Negro	8	0	0	8
	100	0	0	100
Ocre	0	1	0	1
	0	100	0	100
Paja	0	0	13	13
	0	0	100	100

Número de observaciones y porcentajes clasificado por tipo				
Arista	Ausente	Larga	Corta	Total
Ausente	13	0	0	13
	100	0	0	100
Larga	0	7	0	7
	0	100	0	100
Corta	0	0	2	2
	0	0	100	100

Número de observaciones y porcentajes clasificado por tipo			
Apice	Paja	Purpura	Total
Paja	13	0	13
	100	0	100
Purpura	0	8	8
	0	100	100

Los resultados obtenidos por resustitución muestran que la totalidad de las observaciones con clase anteriormente asignada, fueron clasificadas correctamente. Esta capacidad podría ser explicada por la ubicación en el genoma de los SSR utilizados, indicando que los marcadores microsatélites utilizados tienen el potencial de detectar variaciones en regiones cromosómicas donde existen determinantes genéticos de las características fenotípicas que presentan los biotipos con características de maleza estudiados.

El análisis realizado para la discriminación entre clases con diferentes características morfológicas sugeriría entonces que los SSR utilizados podrían estar localizados en regiones del genoma donde se localizan genes o QTLs involucrados en la determinación de los fenotipos característicos de accesiones de arroz rojo. Por lo tanto se buscaron referencias sobre la base genética de algunas de estas características involucradas en el síndrome de domesticación (transición a formas cultivadas de *O. sativa* L.) en diversas bases de datos integradas en Gramene ([www.gramene.org](http://www.gramene.org)).

El color del pericarpio de la semilla, como se planteó anteriormente, es una de las características que definen a los biotipos con características de maleza (perjudicando la calidad industrial del arroz cultivado), siendo por ello motivo de diversos estudios buscando entender su base genética y fisiológica. Mediante estudios de genética clásica y comparativa han sido identificado QTLs directamente relacionados con la coloración del grano de arroz. Recientemente fueron identificados dos loci involucrados en la coloración del pericarpio. Uno de ellos denominado Rc se encuentra en el cromosoma 7, mientras que otro loci denominado Rd se encuentra en el cromosoma 1. Cuando ambos loci están presentes, producen semillas de color rojo. Mientras que cuando se presenta solo el Rc produce coloración de semilla marrón y la presencia sola de Rd no presenta fenotipo. [Sweeney *et al.*, 2006].

El gen Rc es requerido para la coloración rojiza del pericarpio en arroz (*O. sativa* L.), y se encuentra ubicado en una región de 18.5 kb del cromosoma 7, asociado a genes que estarían involucrados en la determinación de otras características típicas de malezas, el desgrane y la dormancia. Este gen presenta 3 alelos que dan coloraciones diferentes (Rc que produce semillas con líneas marrones sobre fondo rojo; Rc-s que produce rojo más tenue; rc alelo nulo). El alelo que determina la coloración rojiza (Rc y Rc-s), fue considerado como mutante, debido a que su fenotipo difiere de los cultivares modernos. Sin embargo la acción de Rc es dominante sobre el pericarpio blanco (rc), sugiriendo que son los cultivares modernos los que presentan el alelos mutante (no funcional), como una versión de su ancestro (*O. ruffipogon* L., semillas rojas).

Este gen *Rc* codifica una proteína bHLH (proteína básica hélice-loop-hélice) de 668 aminoácidos., que presenta dominios de bHLH, comunes a factores de transcripción que participan en la regulación de la síntesis de pigmentos (en este caso de proantocianidinas, pigmento responsable de la coloración roja). Mediante estudios de comparación de secuencias entre una variedad de arroz cultivado (*O.sativa*), una especie de arroz ancestral (*O. rufipogon*) y dos líneas mutantes para *Rc*, se observó la existencia de una delección de 14 pb en el gen *Rc* de los cultivares (pericarpio blanco), concluyendo que la misma sería responsable de la pérdida de color. Esta conclusión ha sido respaldada por estudios de expresión donde se observaron niveles de expresión del gen *Rc* similares para semillas blancas (cultivar), como para rojas (*O. rufipogon*), con la particularidad de que el transcrito del cultivar era más corto (14 bases menos), concordando con la información de secuencia. [Sweeney *et al.*, 2006], dando como resultado una proteína no funcional.

La ubicación del QTL responsable de la coloración del pericarpio fue determinada mediante mapeo genético, con el uso de marcadores microsatélites flanqueantes. Se encontró que uno de los SSR utilizados en este estudio está mapeado en una región cercana a dicho QTL. El microsatélite RM234 ubicado en el cromosoma 7,

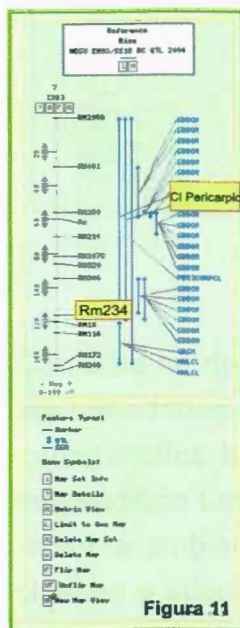


Figura 11

está co-localizado con el QTL involucrado en la coloración del pericarpio, dentro del intervalo de confianza definido para el mismo (líneas azules), (figura N° 11). La distancia entre el gen *Rc* y el marcador RM234, es aproximadamente de 50-60 cM, distancia que puede ser considerada demasiado grande para inferir la co-segregación. Sin embargo, existen evidencias acerca de la existencia de grandes bloques conservados que no han sufrido recombinación (haplotipos) a través de generaciones (especialmente para especies autóгамas altamente seleccionadas a partir de un número limitado de formas ancestrales, Garris *et al.* 2003), lo que determinaría altos niveles

de desequilibrio genético (LD) [Daly *et al.*, 2001; Gabriel *et al.*, 2002] en este tipo de poblaciones. Esto permite sugerir que el SSR RM234 no solo estaría ligado al gen *Rc* sino que muy probablemente presente desequilibrio genético respecto a la región cromosómica donde se localiza este gen.

Otra de las características importante que diferencian un cultivar de una maleza, el desgrane de las semillas, característica presente en malezas y no en cultivares, debido a que esta causa grandes pérdidas de rendimiento, ha sido uno de los objetivos centrales de los procesos de selección realizados por diferentes programas de mejoramiento.



Se han reportados cuatro loci dentro del genoma de arroz que estarían involucrados en el desgrane. Estos son *sh1* en el cromosoma 11 [Nagao y Takahashi, 1963], *sh2* en el cromosoma 1 [Oba *et al.*, 1990], *Sh3* en cromosoma 4 [Eiguchi y Sano, 1990; Nagai *et al.*, 2002], y *Sh4* en el cromosoma 3 [Flukuta *et al.*, 1994; Fukuta y Yagi, 1998]. Además de otros QTLs ubicados en los cromosomas 1,



3, 4, 7, 8,11 [Xiong *et al.*, 1999; Cai y Morishimia, 2000; Bres-Party *et al.*, 2001; Thomson *et al.*, 2003], también involucrados con dicha característica. En la figura N° 12 se muestra el mapa genético del cromosoma 3, donde se destacan la ubicación de otro de los SSRs utilizados en nuestro estudio (RM251) y uno de los QTL identificados para esta característica. El intervalo de confianza para la ubicación del QTL se encuentra entre 130 – 200 cM, mientras que el marcador se ubica aproximadamente a 60 cM de este. Como en el caso anterior; basados en la existencia de grandes bloques de conservación (LD) y conociendo a través de los resultados obtenidos en este trabajo que éste marcador es uno de los que presenta alelos que maximizan la discriminación entre cultivares y rojos (ver tabla N° 6 anexo), se considera la posible asociación del marcador vía desequilibrio genético como una posible explicación del valor predictivo de este marcador para las clases fenotípicas analizadas.

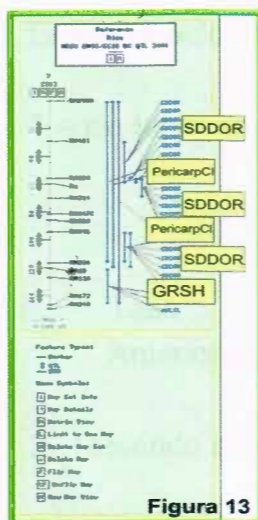
Como se ha discutido anteriormente otra característica importante que diferencia a las variedades cultivadas de las malezas es la dormancia de las semillas. Esta característica ha sido definida como genéticamente compleja y cuyo efecto está influenciado tanto por la base genética determinada por diferentes QTLs como por factores ambientales [Anderson *et al.* 1993]. Éstos se encuentran organizados en clusters e interactúan unos con otros con altos grados de epistasia, sugiriendo que para el control de la dormancia, actuaría una compleja red [Gu *et al.* 2004].

El arroz presenta grandes diferencias para esta característica, donde los genotipos que presentan mayor grado de dormancia se han encontrado en accesiones no domesticadas de especies salvajes (*O ruffipogon* L.) y biotipos con características de maleza (*O. sativa* L.) [Oka, 1988; Suh *et al.*, 1997]. Estos genotipos no domesticados podrían incluir un mayor número de variantes alélicas para la dormancia de las semillas, los que podrían haber sido eliminados por la domesticación en los cultivados. [Gu *et al.*, 2004]. Se ha reportado la existencia de seis QTL para la dormancia en arroz, los cuales están organizados en dos clusters [Gu *et al.*, 2005]. Cuatro de los seis QTLs se encuentran agrupados en uno de los clusters, y están muy relacionados con otras características de las malezas como el color de pericarpio, o el desgrane [Gu *et al.*, 2005], como se ha discutido

anteriormente. Estos presentan una ubicación no aleatoria, la que se asemeja a la estructura de los QTLs del síndrome de domesticación [Doebley y Stec, 1991]. El segundo grupo estaría formado por los dos QTLs restantes, que no presentarían asociación con loci de otras características [Gu *et al.*, 2005].

La ubicación de los QTLs del primer grupo fue identificada por Gu *et al.* (2004), mediante la construcción de mapas genéticos localizándolos en los cromosomas 4, 6, 7, 8, 12, delimitados por marcadores microsatélites. El segundo grupo de QTLs, no han sido mapeados en detalle, pero estudios recientes [Gu *et al.*, 2005] sugieren su ubicación en regiones de los cromosomas 1, 2, 3, 5, y 11. Este último grupo tiene efectos menores sobre la determinación de la dormancia que los pertenecientes al primer cluster. En este caso no se dispone de suficientes referencias para plantear que el SSR RM253, mapeado en el cromosoma 5, podrían estar asociado con alguno de los QTLs del segundo cluster; para confirmarlo serían necesarios estudios más detallados de la región, a través de mapeos finos del cromosoma y los QTLs involucrados.

Existen evidencias de la asociación de entre el color del pericarpio y la dormancia de las semillas en *Arabidopsis* y trigo [Debeaujon *et al.*, 2000; Flintman *et al.*, 2002], así como también para biotipos de arroz rojo, posiblemente generada por la existencia de altos niveles de desequilibrio genético en las poblaciones iniciales de las formas cultivadas [Gu *et al.*, 2003].



Específicamente se ha encontrado asociación entre el gen Rc (responsables de la coloración), ubicado en el cromosoma 7, ligado a uno de los QTLs de dormancia co-localizado en el mismo cromosoma, siendo éste aparentemente, uno de las regiones cromosómicas más importantes para la determinación de las características de dormancia en cereales [Gu *et al.*, 2004]. Este "bloque de genes adaptativos" se habría ido perdiendo a través del proceso de domesticación [Ji *et al.*, 2006]; es interesante destacar, que tanto para la característica dormancia de semilla, así como para desgrane y color del pericarpio hay una región cromosómica en común, donde existen QTLs para las tres características en el cromosoma 7 (Figura 13).

El origen del arroz rojo (tipo maleza) no está claro aún. La existencia de biotipos con estas características en Uruguay podría deberse a la introducción de lotes de semilla de arroz contaminados con formas de tipo maleza desde Europa, lugar desde donde fue introducido el cultivo en etapas tempranas del desarrollo agrícola previas a la amplia mecanización que favorece el uso de formas semi-enanas de

alto potencial. Dado que en el conjunto de muestras estudiadas se contó con accesiones de arroz rojo procedentes de regiones de cultivo localizadas en Brasil (Río grande del Sur) así como accesiones de origen italiano introducidas como referencia de formas ancestrales, se ajustó un procedimiento discriminante para la clasificación de estas en base a su origen, basado en la información de cada uno de los microsatélites analizados; para ello se corrió nuevamente el procedimiento PROC DISCRIM SAS 9.1, pero esta vez utilizando como clase el origen de cada una de las accesiones.

Los resultados obtenidos para esta clasificación se muestran en la tabla N° 10, el detalle de clasificación se muestra en tabla N° 7 de anexo.

**Tabla N° 10:** Porcentajes de clasificación obtenidos para los datos calibrados:  
(resumen de validación cruzada)

Número de observaciones y porcentaje clasificado en tipo				
De País	Brasil	Italia	Uruguay	Total
Brasil	6 100	0 0	0 0	6 100
Italia	0 0	7 100	0 0	7 100
Uruguay	0 0	0 0	39 100	39 100

Las estimaciones de error para este caso fueron:

1 – sin tener en cuenta las proporciones de cada clase:

	Brasil	Italia	Uruguay	Total
Tasa	0.0000	0.0000	0.0000	0.0000
Anteriores	0.3333	0.3333	0.3333	

2- Teniendo en cuenta las proporciones de cada clase:

	Brasil	Italia	Uruguay	Total
Tasa	0.0000	0.0000	0.0000	0.0000
Anteriores	0.1154	0.1346	0.7500	

A partir de los resultados obtenidos en la clasificación por país de origen de las accesiones de arroz rojo, se puede inferir que los microsatélites utilizados permitirían discriminar entre accesiones procedentes de diferentes regiones geográficas con valores cercanos a 100% de acierto (dentro de los límites del

tamaño de muestra utilizado. La posibilidad de diferenciar entre biotipos de arroz rojo de diferente procedencia estaría indicando que las poblaciones locales han desarrollado adaptaciones a partir de las formas de *O. sativa* L. originalmente introducidas con características de maleza, por lo que a pesar de sus diferencias son claramente discriminables como grupo respecto a los cultivares modernos.

El arroz cultivado presenta una estructura genética determinada en gran parte por sus ancestros. Es conocida la existencia de dos sub-especies dentro del arroz cultivado (*O. sativa* L.), indica y japónica, que habrían surgido por eventos separados de domesticación de su ancestro común *O. ruffipogon* L., en diferentes regiones y a diferentes tiempos [Garris *et al.*, 2005]. Dado que el arroz rojo pertenece a la misma especie que los cultivares es de esperar que los biotipos con características de maleza existentes en las diferentes regiones se adapten a las condiciones de cultivo para sobrevivir. La existencia de cruzamientos naturales entre biotipos con características de maleza y cultivares determina la existencia de un complejo dinámico en constante evolución [Vaughan *et al.*, 2001]; por lo tanto se considera necesario profundizar en el conocimiento de la estructura genética y diversidad de las poblaciones locales, considerando especialmente el posible impacto del flujo genético entre arroz rojo y cultivares respecto a la incorporación de genes para resistencia a herbicidas dentro de los sistemas productivos de la región.

## CONCLUSIONES

1) La utilización de un número reducido de microsatélites para genotipado de muestras de arroz rojo y cultivares utilizados en Uruguay permitió ajustar un modelo de clasificación asistida por marcadores basado en el algoritmo k-NN

2) El conjunto de entrenamiento definido por muestras de arroz rojo y cultivares a los que se aplicó el modelo de clasificación permitiría aplicar el alto valor predictivo para las clases AR y CV a la identificación de híbridos entre biotipos con características de maleza y cultivares

2) La diferenciación genética de biotipos de arroz rojo respecto a cultivares de arroz puede ser una valiosa herramienta para apoyar procesos de control de calidad en lotes de semillas básicas (madre y fundación) para variedades utilizadas por la industria arrocera de Uruguay.

## BIBLIOGRAFÍA

- **Anderson, J.**, Sorrells, M., Tanksley, S. (1993). 'RFLP análisis of genomic regions associated with resistente to preharvest spouting in wheat' *Crop Sci.* 33:453-459.
- **Baker, J.B. and Sonnier E.A.** 1983. Red rice and its control. Proceeding Conference on Weed Control in Rice, Philippines. IRRI pag: 327-333.
- **Bres – Patry, C.**, Lorieux, M., Clement, G., Bangrantz, M., Ghesquiere, A. (2001). 'Heredity and genetic mapping of domestication-related traits in a temperate japonica weedy rice. *TAG* 102:118-126.
- **Burger, J.C.**, Lee, S., Ellstrand, N.C. (2006). 'Origin and genetic structure of feral rye in the western United State. *Molecular Ecology.* 15:2527-2539.
- **Cai, H. and Morishima, H.** (2000). 'Genomic regions affecting seed shattering and seed dormancy in rice. *TAG* 100:840-846.
- **Cao, Q.**, Lu, B.R., Xia, H., Rong, J., Sala, F., Spada, A., Grassi, F. (2006) 'Genetic Diversity and origin of weedy rice (*Oryza sativa f. spontanea*) population found in north – eastern China revealed by simple sequences repeat (SSR) markers. *Annals of Botany* 98:1241-1252.
- **Capdevielle, F.** (2001) Evaluation of a discriminant analysis procedure combining agronomic and molecular marker information for germplasm improvement in rice. MSc Thesis Series, Louisiana State University.
- **Capdevielle, F.**, Federicci, M.T., Saldain, N., Vaughan, D. (2003a). 'Differentiation of Uruguayan weedy rice, and cultivars using markers-assisted clasification'. En: Conferencia Internacional de Arroz de clima tempaldo, 3º, Punta del Este, Uruguay.
- **Capdevielle, F.**, Pinson, S., Oard, J. (2003b). 'Marker – assisted classification of Lemont x Teqing RILs into disease response groups: comparison of discriminate analysis and neural network algorithms, In Proceedings 3<sup>rd</sup> International Temperate Rice Conference, Uruguay.
- **Capdevielle, F.**, Bonnacarrère, V., Alonso, M., Garaycochea, S., Valenzuela, S. (2004). 'Clasificación asistida por marcadores moleculares aplicada al programa de

mejoramiento genético de arroz'. Instituto Nacional Agropecuaria (INIA). Uruguay. Actividades de difusión 373, cap.5 pag. 33-38.

- **Chalam, G.V.,** Singh, A., Douglas, J.E. (1967). Seed testing manual. Indian Council of Agricultural Research and United State Agency for International Development. New Delhi.
- **Chen, L.J.,** Lee, D.S., Song, Z.P., Suh, H.S., Lu, B.R. 2004. 'Gene flow from cultivated rice (*Oryza sativa*) to its weedy and wild relatives'. *Annals of Botany* 93: 67-73.
- **Daly M.J,** Rioux J.D., Schaffner S.F., Hudson T.J., Lander E.S. (2001) 'High-resolution haplotype structure in the human genome'. *Nature Genet.* 29:229-232.
- **Debeaujon, I.,** Leon-Kloosterziel, M., Koornneef, M. (2000). 'Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*'. *Plant Physiol.* 122:403-413.
- **Dekkers J.C.M. and Hospital F.** (2002). 'The use of molecular genetics in the improvement of agricultural populations'. *Nature* 3: 22-32.
- **Doebly, J. and Stec, A.** (1991). 'Genetics analysis of the morphological differences between maize and teosinte'. *Genetics* 129:285-295.
- **Eiguchi, M., and Sano, Y.** (1990). A gene complex responsible for seed shattering and panicle spreading found in common wild rices'. *Rice Genet. Newsl.* 7:105-107.
- **Estorninos, L.E.,** Burgos, N.R., Gealy, D.R., Stewart, J.M., Talbert, R.E. (2000). 'Genetic characterization of red rice populations using molecular markers. Pages 76-84 in R.J. Norman and C.A. Beyrouthy eds. B.R. Wells Rice Research Studies – 1999. Series 476 Arkansas: Arkansas Agricultural Experiment Station.
- **Federicci, M.T.,** Vaughan, D., Tomooka, N., Kaga, A., Wang Wang, X., Doi, K., Francis, M., Zorrilla, G., Saldain, N. (2003). Analysis of Uruguayan weedy rice genetic diversity using AFLP molecular markers. *EJB*, vol 4, N°3.
- **Flintham, J.,** Adlam, R., Bassoi, M., Holdsworth, M., Gale, M.(2002). 'Mapping genes for resistance to sprouting damage in wheat'. *Euphytica* 126:39-45.
- **Fukuta, Y and Yagi T.** (1998). 'Mapping of a shattering resistance gene in a mutant line SR-5 induced from an indica rice variety, Nan-jing1'. *Breed. Sci.* 48:345-348.

- **Fukuta, Y.**, Yoshida, K., Fukui, K., Kobayashi, A. (1994). 'Analysis of shattering degrees and abscission layer development in shattering resistant mutant lines induced from and indica rice (*O. sativa* L.) variety , nan-jing 11'. *J. Breed.* 44:195-200.
- **Gabriel SB**, Schaffner S.F., Nguyen H., Moore J.M., Roy J., Blumenstiel B., Higgins J., DeFelice M., Lochner A., Faggart M., Liu-Cordero S.N., Rotimi C., Adeyerno A., Cooper R., Ward R., Lander E.S., Daly M.J., Altshuler D. (2002) 'The structure of haplotype blocks in the human genome'. *Science* 296:2225-2229.
- **Garris A**, McCouch S, Kresovich S (2003) 'Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.)'. *Genetics* 165: 759-769
- **Gealy, D.R. and Black, H.L.** (1998). 'Activity of glufosinate (Liberty) against red rice biotypes in glufosinate – resistant Glumont rice. Page: 41-48. In R.J. Norman and T.H. Johnson, eds. B.R. Wells Rice Research Studies – 1997. Series 460. Arkansas: Arkansas Agricultural Experiment Station.
- **Gealy, D.R.**, Tai, T. H., Sneller, C.H. (2002). 'Identification of red rice, rice and hybrid populations using microsatellite markers'. *Weed Sci.* 50:333-339.
- **Gealy, D.R.**, Mitten, D.H., Rutger, J.N. (2003). 'Gene flow between red rice (*Oryza sativa*) and herbicide-resistant rice (*O. sativa*): implications for weed management'. *Weed Technology* 17: 627–645.
- **Ginot, F.**, Bordelais, I., Nguyen, S., Gyapay, G. (1996). 'Correction of some genotyping errors in automated fluorescent microstellite analysis by enzymatic removal of one base overhangs.' *Nucleic Acid Research* 24: 540-541.
- **Gu, X.**, Chen, Z., Foley, M. (2003). 'Inheritance of seed dormancy in weedy rice. *Crop Sci.* 43:835-843.
- **Gu, X.**, Shahryar, F., Kianian, F., Foley, M. (2004). 'Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (*O. sativa*)'. *Genetics* 166:1503-1516.
- **Gu, X.**, Shahryar, F., Kianian, F., Hareland, G., Hoffer, B., Foley, M. (2005). 'Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*O. sativa*)'. *TAG* 110:1108-1118.
- **Hancock JF.** 2004. 'Plant Evolution and the Origin of Crop Species', 2nd ed. Cambridge, MA, USA: CABI Publishing.



- **Hoagland, R.E., Paul, R.V.** (1978). 'A comparative SEM study of red rice and several commercial (*Oryza sativa*) varieties. *Weed Sci.* 26:619-625.
- **Ji, H., Chu, S., Jiang, W., Cho, Y., Hahn, J., Eun, M., McCouch, S., Koh, H.** (2006). 'Characterization and mapping of shattering mutant in rice that corresponds to block of domestication genes'. *Genetics* 173: 995 – 1005.
- **Johnson R. A. and Wichern D. W.** (1998) 'Applied multivariate statistical analysis'. Prentice Hall Inc., New Jersey.
- **Li, C., Zhou A., Sang T.** 'Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*'. *New Phytologist* (2006) 170: 185–194
- **Manual 3rd FAO/IAEA Interregional Training Course** on "Mutant Germplasm Characterisation Using Molecular Markers". International Atomic Energy Agency (IAEA), Viena, Austria. (2002)
- **McCouch, S.R., Chen, O., Panaud, S., Temnykh, Y., Xu, Y., Cho, G., Huang, N., Ishii, T, Blair, M.** (1997). 'Microsatellite marker development, mapping and applications in rice genetics and breeding'. *Plant Mol. Biol.* 35:89-99.
- **Mortimer, A.M.** (1994). 'The classification and ecology of weeds' In 'Weed management for developing countries' eds. R. Lebrada, J. Caseley and C. Parker, FAO, Rome pag 7-20.
- **Nagai, Y., Sobrizal, P., Sanchez, T., Kurakazu, K., Doi, K.** (2002). 'Sh3, a gene for seed shattering, commonly found in wild rices'. *Rice Genet. Newsl.* 19: 74–75.
- **Nagao, S., and Takahashi.** (1963). 'Triallelic construction of twelve linkage group in Japanese rice. *J. Fac. Agric. Hokkaido Univ.* 53:72-130.
- **Naredo, M.E.B., Juliano, A.B., Lu, B.R., Guzman, F., Jackson, M.T.** (1998). 'Responses to seed dormancy-breaking treatments in rice species (*Oryza L.*)'. *Seed Sci. & Technology.* 26:675-689 Appendix 3-4.
- **Oard, J., Cohn, M., Linscombe, S., Gealy, D., Gravois, K.** (2000). 'Fields evaluation of seed production, shattering, and dormancy in hybrid populations of transgenic rice (*Oryza sativa*) and weed , red rice (*Irysa sativa*). *Plant Sci.* 157: 13-22.

- **Oba, S.**, Kicuchi, F., Maruyama, K. (1990). 'Genetic análisis of semidwarfness and grain shattering of Chinese rice (*O. sativa* L) variety "Ai-Jio-Nan-Te"'. J. Breed. 40: 13-20.
- **Oka, H. I.** (1988). 'The origin of cultivated rice'. Elsevier Ámsterdam pag. 254.
- **Pysek, P. and Prach, K.** (2003). 'Research into plant invasions in a crossroads region: history and focus'. Biological Invasions 5: 337-348.
- **Singh, R.J. and Khush, G.S.** (2000). 'Cytogenetics of rice. Page: 287-311 in J.S. Nanda, ed. Rice Breeding and Genetics – Research Priorities and Cahllenges. Enfield, NH: Science Publishers.
- **Suh, H. S.**, Sato, Y., Monshima, H. (1997). 'Genetic characterization of weedy rice (*Oryza sativa* L.) based on morpho-physiology, isozymes and RAPD markers. Theor. Appl. Genet. 94:316-321.
- **Sweeney, M.**, Thomson, M., Pfeil, B., McCouch, S. (2006). 'Caught Red-Handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice'. The Plant Cell, 18:283-294.
- **Tetko, I.** Livingstone, D., Luik, A. (1995). 'Comparison of overfitting and overtraining'. Journal of Chemical Information and Computer Sciences. 35(5): 826-833.
- **Thomson, M.J.**, Tai, T.H., McClung, M., Lai, X., Hinga, M.E. (2003). 'Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcrosses population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson'. TAG 107: 479-493.
- **Vaughan, L. K.**, Ottis, B. V., Prazak-Havey, A. M., Bormans, C.A., Sneller, C., Chandler, J.M., Park, W.D. (2001). Is red rice found in commercial rice really *Oryza sativa*?. Weed Sci. 49:468-476.
- **Wet, J.M.J., Harlan J.R.** (1975). 'Weeds and domesticates; evolution in man-made habitats. Economic Botany 29:99-107.
- **Xiong, L.**, Liu, D., Dai, K., Xu, C.G., Zhang. (1999). 'identification of genetics factor controlling domestication-related traits of rice using an F2 population of cross between *Oryza sativa* and *O. ruffipogon*. Theor. Appl. Genet. 98: 243 – 251.

- **Zorrilla de San Martín, G.** (1992). 'Arroz rojo conózcalo y combátalo'. Boletín de divulgación N° 20. Instituto Nacional de Investigación Agropecuaria (INIA). Uruguay.

# ANEXO

**TABLA N° 1:** Listado de muestras con información agronómica

Muestras	Procedencia (sitio de colecta)	Cáscara	Ápice	Tipo de arista	Hábito crecimiento	Grupo AFLP	Tipo
2	Vichadero -1	Negro	Rojo	Largo	sv	A	Rojo
4	Vichadero - 3	Paja	Paja	Ausente	sp	C	Rojo
5	Tacuarembó - 1	Paja	Paja	Ausente	sp	C	Rojo
9	Tacuarembó - 5	Paja	Paja	Ausente	m	C	Rojo
10	Tacuarembó - 6	Paja	Paja	Ausente	v	B	Rojo
11	Zapata - 1	Paja	Rojo	Corto	sv	B	Rojo
12	Zapata - 2	Negro	Rojo	Largo	sv	A	Rojo
13	Zapata - 3	Negro	Rojo	Corto	sv	A	Rojo
14	Zapata - 4	Paja	Paja	Ausente	sv	C	Rojo
15	Zapata - 5	Negro	Negro	Largo	sp	A	Rojo
16	Zapata - 6	Negro	Rojo	Largo	p	A	Rojo
17	Río Branco - 1	Paja	Paja	Ausente	sp	C	Rojo
19	Río Branco - 3	Paja	Paja	Ausente	sp	A	Rojo
20	Río Branco - 4	Paja	Paja	Ausente	v	C	Rojo
23	La charqueada - 2	Negro	Rojo	Largo	sv	A	Rojo
24	Rincón de Ramírez	Negro	Rojo	Largo	sp	A	Rojo
26	Desconocida - 1	Paja	Paja	Ausente	m	C	Rojo
27	Desconocida - 2	Paja	Paja	Ausente	sp	B	Rojo
28	Desconocida - 3	Negro	Rojo	Largo	sv	B	Rojo
31	INIA Caraguatá	Paja	Paja	Ausente	v	B	Cultivar
34	Bluebelle	Ocre	Paja	Ausente	v	B	Cultivar
35	EEA 404	Paja	Paja	Ausente	v	B	Cultivar

Hábitos de crecimiento referencias:

m – medio

p – postrado

sp – semi postrado

v - vertical

sv – semi vertical

Tabla Nº 2: Detalle de clasificación por tipo

Sistema SAS

09:50 Tuesday, August 14, 2007

Procedimiento DISCRIM

Resultados de clasificación para los datos calibrados: WORK.ARROZ\_ALELOS  
Resultados de validación cruzada usando Vecino más cercano

Función de la distancia cuadrada

$$D^2(X,Y) = (X-Y)' \text{COV}^{-1} (X-Y)$$

Probabilidad posterior de miembro en cada tipo

$m_k(X)$  = Proporción de obs en grupo k en el vecino más cercano de X

$$\text{Pr}(j|X) = \frac{m_j(X) \text{PRIOR}_j}{\sum_k (m_k(X) \text{PRIOR}_k)}$$

Probabilidad posterior de pertenencia a Tipo

Procedencia	De Tipo	Clasificado en Tipo	CULTIVAR	ROJO
Vichader1	ROJO	ROJO	0.0000	1.0000
Vichader3	ROJO	ROJO	0.0000	1.0000
Tacuarem1	ROJO	ROJO	0.0000	1.0000
Tacuarem5	ROJO	ROJO	0.0000	1.0000
Tacuarem6	ROJO	ROJO	0.0000	1.0000
Zapata_1	ROJO	ROJO	0.0000	1.0000
Zapata_2	ROJO	ROJO	0.0000	1.0000
Zapata_3	ROJO	ROJO	0.0000	1.0000
Zapata_4	ROJO	ROJO	0.0000	1.0000
Zapata_5	ROJO	ROJO	0.0000	1.0000
Zapata_6	ROJO	ROJO	0.0000	1.0000
Rio_Bran1	ROJO	ROJO	0.0000	1.0000
Rio_Bran2	ROJO	ROJO	0.0000	1.0000
Rio_Bran3	ROJO	ROJO	0.0000	1.0000
Rio_Bran4	ROJO	ROJO	0.0000	1.0000
La_charq1	ROJO	ROJO	0.0000	1.0000
La_charq2	ROJO	ROJO	0.0000	1.0000
Rincon_d	ROJO	ROJO	0.0000	1.0000
Desconoc1	ROJO	ROJO	0.0000	1.0000
Desconoc2	ROJO	ROJO	0.0000	1.0000
Desconoc3	ROJO	ROJO	0.0000	1.0000
INIA_Car	CULTIVAR	CULTIVAR	1.0000	0.0000
INIA_Zap	CULTIVAR	CULTIVAR	1.0000	0.0000
INIA_Cua	CULTIVAR	ROJO *	0.0000	1.0000
Bluebell1e	CULTIVAR	CULTIVAR	1.0000	0.0000
EEA_404	CULTIVAR	CULTIVAR	1.0000	0.0000
E1_Paso_	CULTIVAR	CULTIVAR	1.0000	0.0000
INIA_O1i	CULTIVAR	CULTIVAR	1.0000	0.0000
Itapebi2	ROJO	ROJO	0.0000	1.0000
Itapebi3	ROJO	ROJO	0.0000	1.0000
Itapebi4	ROJO	ROJO	0.0000	1.0000
Itapebi5	ROJO	ROJO	0.0000	1.0000
Itapebi6	ROJO	ROJO	0.0000	1.0000
Itapebi8	ROJO	ROJO	0.0000	1.0000
BellaUni1	ROJO	ROJO	0.0000	1.0000
F1_ARP_I1		ROJO *	0.0000	1.0000
F1_ARP_I2		ROJO *	0.0000	1.0000
F1_ARP_B3		ROJO *	0.0000	1.0000
F1_ARP_I4		ROJO *	0.0000	1.0000
F1_ARP_E5		ROJO *	0.0000	1.0000
F1_ARN_I6		ROJO *	0.0000	1.0000
F1_ARN_I7		CULTIVAR *	1.0000	0.0000
F1_ARN_B8		ROJO *	0.0000	1.0000
F1_ARN_I9		ROJO *	0.0000	1.0000
Italia_2	ROJO	ROJO	0.0000	1.0000
Italia_7	ROJO	ROJO	0.0000	1.0000
Italia_9	ROJO	ROJO	0.0000	1.0000
Italia_10	ROJO	ROJO	0.0000	1.0000
Italia_13	ROJO	ROJO	0.0000	1.0000
Italia_17	ROJO	ROJO	0.0000	1.0000
Italia_19	ROJO	ROJO	0.0000	1.0000
Balilla	ROJO	ROJO	0.0000	1.0000
IRGA_417	ROJO	ROJO	0.0000	1.0000
Rio_gran1	ROJO	ROJO	0.0000	1.0000
Rio_gran2	ROJO	ROJO	0.0000	1.0000
Rio_gran3	ROJO	ROJO	0.0000	1.0000
Rio_gran4	ROJO	ROJO	0.0000	1.0000
Rio_gran5	ROJO	ROJO	0.0000	1.0000
Rio_gran6	ROJO	ROJO	0.0000	1.0000

\* Observación mal clasificada

**Tabla N° 3:** Variables seleccionadas por el procedimiento SAS STEP DISC

Sistema SAS

09:50 Tuesday, August 14, 2007

Procedimiento STEPDISC

Resumen de la selección hacia delante

Pr > Paso media	Número en ASCC	Introducido	R-cuadrado parcial	F-Valor	Pr > F	Lambda de Wilks	Pr < Lambda	Correlación canónica cuadrada de la
0.0060	1	RM215_149	0.1468	8.26	0.0060	0.85321655	0.0060	0.14678345
0.0003	2	RM251_119	0.1646	9.26	0.0038	0.71274048	0.0003	0.28725952
<.0001	3	RM251_118	0.2065	11.97	0.0012	0.56557507	<.0001	0.43442493
<.0001	4	RM251_138	0.2729	16.89	0.0002	0.41123086	<.0001	0.58876914
<.0001	5	RM234_155	0.1836	9.89	0.0030	0.33573534	<.0001	0.66426466
<.0001	6	RM253_124	0.0781	3.64	0.0630	0.30950643	<.0001	0.69049357
<.0001	7	RM251_148	0.1486	7.33	0.0098	0.26352488	<.0001	0.73647512
<.0001	8	RM251_117	0.1366	6.49	0.0147	0.22751658	<.0001	0.77248342
<.0001	9	RM253_132	0.0816	3.55	0.0667	0.20895795	<.0001	0.79104205
<.0001	10	RM253_130	0.1656	7.74	0.0083	0.17435108	<.0001	0.82564892
<.0001	11	RM253_151	0.0619	2.51	0.1215	0.16355229	<.0001	0.83644771
<.0001	12	RM253_138	0.0668	2.65	0.1122	0.15262928	<.0001	0.84737072
<.0001	13	RM215_145	0.0809	3.17	0.0836	0.14028810	<.0001	0.85971190
<.0001	14	RM234_135	0.1169	4.63	0.0383	0.12388562	<.0001	0.87611438
<.0001	15	RM215_148	0.1471	5.86	0.0209	0.10566272	<.0001	0.89433728
<.0001	16	RM234_139	0.0619	2.18	0.1494	0.09911838	<.0001	0.90088162
<.0001	17	RM253_140	0.0672	2.31	0.1387	0.09245684	<.0001	0.90754316
<.0001	18	RM253_120	0.0355	1.14	0.2940	0.08917828	<.0001	0.91082172
<.0001	19	RM234_146	0.0570	1.81	0.1883	0.08409766	<.0001	0.91590234
<.0001	20	RM251_123	0.0541	1.66	0.2080	0.07954834	<.0001	0.92045166

**Tabla N° 4:** Detalle de clasificación por tipo con variables seleccionadas por STEPDISC

Procedimiento DISCRIM  
 Resultados de clasificación para los datos calibrados: WORK.ARROZ\_ALELOS  
 Resultados de validación cruzada usando Vecino más cercano

Función de la distancia cuadrada

$$D(X,Y) = (X-Y)' \text{COV}^{-1} (X-Y)$$

Probabilidad posterior de miembro en cada Tipo

$m_k(X)$  = Proporción de obs en grupo k en el vecino más cercano de X

$$\text{Pr}(j|X) = \frac{m_j(X) \text{PRIOR}_j}{\sum_k (m_k(X) \text{PRIOR}_k)}$$

Probabilidad posterior de pertenencia a Tipo

Procedencia	De Tipo	Clasificado en Tipo	CULTIVAR	ROJO
Vichader	ROJO	ROJO	0.3000	0.7000
Vichader	ROJO	ROJO	0.0000	1.0000
Tacuarem	ROJO	ROJO	0.0000	1.0000
Tacuarem	ROJO	ROJO	0.0000	1.0000
Tacuarem	ROJO	ROJO	0.0000	1.0000
Zapata_1	ROJO	ROJO	0.2857	0.7143
Zapata_2	ROJO	ROJO	0.3000	0.7000
Zapata_3	ROJO	ROJO	0.0000	1.0000
Zapata_4	ROJO	ROJO	0.0000	1.0000
Zapata_5	ROJO	ROJO	0.3000	0.7000
Zapata_6	ROJO	ROJO	0.3000	0.7000
Rio_Bran	ROJO	ROJO	0.0000	1.0000
Rio_Bran	ROJO	ROJO	0.3000	0.7000
Rio_Bran	ROJO	ROJO	0.3000	0.7000
Rio_Bran	ROJO	ROJO	0.0000	1.0000
La_charq	ROJO	ROJO	0.3000	0.7000
La_charq	ROJO	ROJO	0.3000	0.7000
Rincon_d	ROJO	ROJO	0.3000	0.7000
Desconoc	ROJO	ROJO	0.2857	0.7143
Desconoc	ROJO	ROJO	0.3000	0.7000
Desconoc	ROJO	ROJO	0.0000	1.0000
INIA_Car	CULTIVAR	ROJO	*	0.0000
INIA_Zap	CULTIVAR	ROJO	*	0.4886
INIA_Cua	CULTIVAR	ROJO	*	0.0000
Bluebell	CULTIVAR	ROJO	*	0.4886
EEA_404	CULTIVAR	ROJO	*	0.4886
El_Paso	CULTIVAR	ROJO	*	0.4886
INIA_Car	CULTIVAR	ROJO	*	0.4886
Itapebi_	ROJO	ROJO	0.0000	1.0000
Itapebi_	ROJO	ROJO	0.3000	0.7000
Itapebi_	ROJO	ROJO	0.0000	1.0000
Itapebi_	ROJO	ROJO	0.0000	1.0000
Itapebi_	ROJO	ROJO	0.0000	1.0000
Itapebi_	ROJO	ROJO	0.0000	1.0000
BellaUni	ROJO	ROJO	0.3000	0.7000
F1_ARP_I		ROJO	*	0.2905
F1_ARP_I		ROJO	*	0.2905
F1_ARP_B		ROJO	*	0.2905
F1_ARP_I		ROJO	*	0.2905
F1_ARP_E		ROJO	*	0.2905
F1_ARN_I		ROJO	*	0.0000
F1_ARN_I		ROJO	*	0.0000
F1_ARN_B		ROJO	*	0.0000
F1_ARN_I		ROJO	*	0.2905
Italia_2	ROJO	ROJO	0.0000	1.0000
Italia_7	ROJO	ROJO	0.0000	1.0000
Italia_9	ROJO	ROJO	0.0000	1.0000
Italia_1	ROJO	ROJO	0.0000	1.0000
Italia_1	ROJO	ROJO	0.3000	0.7000
Italia_1	ROJO	ROJO	0.2857	0.7143
Italia_1	ROJO	ROJO	0.0000	1.0000
Balilla	ROJO	ROJO	0.0000	1.0000
IRGA_417	ROJO	ROJO	0.3000	0.7000
Rio_gran	ROJO	ROJO	0.3000	0.7000
Rio_gran	ROJO	ROJO	0.0000	1.0000
Rio_gran	ROJO	ROJO	0.0000	1.0000
Rio_gran	ROJO	ROJO	0.0000	1.0000
Rio_gran	ROJO	ROJO	0.0000	1.0000

\* observación mal clasificada

**Tabla N° 5:** Detalle de clasificación por tipo para grupo de muestras con características morfológicas

Procedimiento DISCRIM  
 Resultados de clasificación para los datos calibrados: WORK.ARROZ\_ALELOS  
 Resultados de validación cruzada usando Vecino más cercano

Función de la distancia cuadrada

$$D^2(X,Y) = (X-Y)' \text{COV}^{-1}(X-Y)$$

Probabilidad posterior de miembro en cada tipo

$m_k(X)$  = Proporción de obs en grupo k en el vecino más cercano de x

$$Pr(j|X) = \frac{m_j(X) \text{PRIOR}_j}{\sum_k (m_k(X) \text{PRIOR}_k)}$$

Probabilidad posterior de pertenencia a tipo

Procedencia	De tipo	Clasificado en tipo	cultivar	rojo
Vichader1	rojo	rojo	0.0000	1.0000
Vichader3	rojo	rojo	0.0000	1.0000
Tacuarem1	rojo	rojo	0.0000	1.0000
Tacuarem5	rojo	rojo	0.0000	1.0000
Tacuarem6	rojo	rojo	0.0000	1.0000
Zapata_1	rojo	rojo	0.0000	1.0000
Zapata_2	rojo	rojo	0.0000	1.0000
Zapata_3	rojo	rojo	0.0000	1.0000
Zapata_4	rojo	rojo	0.0000	1.0000
Zapata_5	rojo	rojo	0.0000	1.0000
Zapata_6	rojo	rojo	0.0000	1.0000
Río_Bran1	rojo	rojo	0.0000	1.0000
Río_Bran3	rojo	rojo	0.0000	1.0000
Río_Bran4	rojo	rojo	0.0000	1.0000
La_Charg	rojo	rojo	0.0000	1.0000
Rincón_d	rojo	rojo	0.0000	1.0000
Desconoc1	rojo	rojo	0.0000	1.0000
Desconoc2	rojo	rojo	0.0000	1.0000
Desconoc3	rojo	rojo	0.0000	1.0000
INIA_Car	cultivar	cultivar	1.0000	0.0000
INIA_Zap	cultivar	cultivar	1.0000	0.0000
Bluebell	cultivar	cultivar	1.0000	0.0000
EEA_404	cultivar	cultivar	1.0000	0.0000
El_Paso_	cultivar	cultivar	1.0000	0.0000
INIA_Oli	cultivar	cultivar	1.0000	0.0000
Itapebí_	rojo	rojo	0.0000	1.0000
Itapebí_	rojo	rojo	0.0000	1.0000
Itapebí_	rojo	rojo	0.0000	1.0000
F1_ARN_I		cultivar *	1.0000	0.0000
F1_ARN_I		rojo *	0.0000	1.0000
Río_Gran	rojo	rojo	0.0000	1.0000



**Tabla N° 6:** Variables seleccionadas por el procedimiento SAS STEPDISC para la clasificación de accesiones con características morfológicas.

Sistema SAS

14:46 Friday, August 31, 2007 17

Procedimiento STEPDISC

Resumen de la selección hacia delante

Pr > Paso ASCC	Número en	Introducido	R-cuadrado parcial	F-Valor	Pr > F	Lambda de wilks	Pr < Lambda	Correlación canónica cuadrada de la media
0.0297	1	RM234_135	0.1479	5.21	0.0297	0.85207101	0.0297	0.14792899
0.0042	2	RM251_118	0.1959	7.07	0.0126	0.68512821	0.0042	0.31487179
0.0002	3	RM251_138	0.2640	10.04	0.0037	0.50427350	0.0002	0.49572650
<.0001	4	RM215_148	0.3051	11.85	0.0019	0.35042735	<.0001	0.64957265
<.0001	5	RM251_116	0.2683	9.53	0.0048	0.25641026	<.0001	0.74358974
<.0001	6	RM234_141	0.4000	16.67	0.0004	0.15384615	<.0001	0.84615385
<.0001	7	RM234_144	0.2245	6.95	0.0145	0.11930926	<.0001	0.88069074
<.0001	8	RM215_143	0.2065	5.98	0.0225	0.09467456	<.0001	0.90532544
<.0001	9	RM215_146	1.0000	Infin	<.0001	0.00000000	<.0001	1.00000000

**Tabla N° 7:** Detalle de clasificación por tipo para grupo de muestras con características morfológicas con variables seleccionadas por STEPDISC

Procedimiento DISCRIM  
 Resultados de clasificación para los datos calibrados: WORK.ARRDZ\_ALELOS  
 Resultados de validación cruzada usando Vecino más cercano

Función de la distancia cuadrada

$$D^2(X,Y) = (X-Y)' \text{COV}^{-1} (X-Y)$$

Probabilidad posterior de miembro en cada tipo

$m_k(X)$  = Proporción de obs en grupo k en el vecino más cercano de X

$$\text{Pr}(j|X) = \frac{m_j(X) \text{PRIOR}_j}{\sum_k (m_k(X) \text{PRIOR}_k)}$$

Probabilidad posterior de pertenencia a tipo

Procedencia	De tipo	Clasificado en tipo	cultivar	rojo
Vichader1	rojo	rojo	0.2066	0.7934
Vichader3	rojo	rojo	0.2066	0.7934
Tacuarem1	rojo	rojo	0.2066	0.7934
Tacuarem5	rojo	rojo	0.2066	0.7934
Tacuarem6	rojo	rojo	0.0000	1.0000
Zapata_1	rojo	rojo	0.1969	0.8031
Zapata_2	rojo	rojo	0.2066	0.7934
Zapata_3	rojo	rojo	0.0000	1.0000
Zapata_4	rojo	rojo	0.2066	0.7934
Zapata_5	rojo	rojo	0.2066	0.7934
Zapata_6	rojo	rojo	0.2066	0.7934
Río_Bran1	rojo	rojo	0.2066	0.7934
Río_Bran3	rojo	rojo	0.2066	0.7934
Río_Bran4	rojo	rojo	0.2066	0.7934
La_Charg	rojo	rojo	0.2066	0.7934
Rincón_d	rojo	rojo	0.2066	0.7934
Desconoc1	rojo	rojo	0.2066	0.7934
Desconoc2	rojo	rojo	0.2066	0.7934
Desconoc3	rojo	rojo	0.0000	1.0000
INIA_Car	cultivar	cultivar	1.0000	0.0000
INIA_Zap	cultivar	cultivar	1.0000	0.0000
Bluebell	cultivar	cultivar	1.0000	0.0000
EEA_404	cultivar	rojo *	0.0000	1.0000
El_Paso_	cultivar	cultivar	1.0000	0.0000
INIA_Oñ	cultivar	cultivar	1.0000	0.0000
Itapebí_	rojo	rojo	0.2066	0.7934
Itapebí_	rojo	rojo	0.1969	0.8031
Itapebí_	rojo	rojo	0.0000	1.0000
F1_ARN_I		rojo *	0.0000	1.0000
F1_ARN_I		rojo *	0.2031	0.7969
Río_Gran	rojo	rojo	0.0000	1.0000

\*observación mal clasificada

Tabla N° 8: Detalle de clasificación por país como clase

Sistema SAS

09:50 Tuesday, August 14, 2007

Procedimiento DISCRIM  
 Resultados de clasificación para los datos calibrados: WORK.ARROZ\_ALELOS  
 Resultados de validación cruzada usando Vecino más cercano

Procedencia	Probabilidad posterior de pertenencia a País				
	De País	Clasificado en País	B	I	U
Vichader1	U	U	0.0000	0.0000	1.0000
Vichader2	U	U	0.0000	0.0000	1.0000
Tacuarem1	U	U	0.0000	0.0000	1.0000
Tacuarem5	U	U	0.0000	0.0000	1.0000
Tacuarem6	U	U	0.0000	0.0000	1.0000
Zapata_1	U	U	0.0000	0.0000	1.0000
Zapata_2	U	U	0.0000	0.0000	1.0000
Zapata_3	U	U	0.0000	0.0000	1.0000
Zapata_4	U	U	0.0000	0.0000	1.0000
Zapata_5	U	U	0.0000	0.0000	1.0000
Zapata_6	U	U	0.0000	0.0000	1.0000
Rio_Bran1	U	U	0.0000	0.0000	1.0000
Rio_Bran2	U	U	0.0000	0.0000	1.0000
Rio_Bran3	U	U	0.0000	0.0000	1.0000
Rio_Bran4	U	U	0.0000	0.0000	1.0000
La_charq1	U	U	0.0000	0.0000	1.0000
La_charq2	U	U	0.0000	0.0000	1.0000
Rincon_d	U	U	0.0000	0.0000	1.0000
Desconoc1	U	U	0.0000	0.0000	1.0000
Desconoc2	U	U	0.0000	0.0000	1.0000
Desconoc3	U	U	0.0000	0.0000	1.0000
INIA_Car		U	*	0.0000	1.0000
INIA_Zap		B	*	1.0000	0.0000
INIA_Cua		U	*	0.0000	1.0000
Bluebell		U	*	0.0000	1.0000
EEA_404		U	*	0.0000	1.0000
El_Paso		I	*	0.0000	1.0000
INIA_Oli		U	*	0.0000	1.0000
Itapebi2	U	U	0.0000	0.0000	1.0000
Itapebi3	U	U	0.0000	0.0000	1.0000
Itapebi4	U	U	0.0000	0.0000	1.0000
Itapebi5	U	U	0.0000	0.0000	1.0000
Itapebi6	U	U	0.0000	0.0000	1.0000
Itapebi8	U	U	0.0000	0.0000	1.0000
BellaUni	U	U	0.0000	0.0000	1.0000
Fl_ARP_I	U	U	0.0000	0.0000	1.0000
Fl_ARP_I	U	U	0.0000	0.0000	1.0000
Fl_ARP_B	U	U	0.0000	0.0000	1.0000
Fl_ARP_I	U	U	0.0000	0.0000	1.0000
Fl_ARP_E	U	U	0.0000	0.0000	1.0000
Fl_ARN_I	U	U	0.0000	0.0000	1.0000
Fl_ARN_I	U	U	0.0000	0.0000	1.0000
Fl_ARN_B	U	U	0.0000	0.0000	1.0000
Fl_ARN_I	U	U	0.0000	0.0000	1.0000
Italia_2	I	I	0.0000	1.0000	0.0000
Italia_7	I	I	0.0000	1.0000	0.0000
Italia_9	I	I	0.0000	1.0000	0.0000
Italia_1	I	I	0.0000	1.0000	0.0000
Italia_1	I	I	0.0000	1.0000	0.0000
Italia_1	I	I	0.0000	1.0000	0.0000
Balilla	U	U	0.0000	0.0000	1.0000
IRGA_417	U	U	0.0000	0.0000	1.0000
Rio_gran1	B	B	1.0000	0.0000	0.0000
Rio_gran2	B	B	1.0000	0.0000	0.0000
Rio_gran3	B	B	1.0000	0.0000	0.0000
Rio_gran4	B	B	1.0000	0.0000	0.0000
Rio_gran5	B	B	1.0000	0.0000	0.0000
Rio_gran6	B	B	1.0000	0.0000	0.0000

\* Observación mal clasificada

12500