

Capítulo 3. Diversidad y estructura genética del germoplasma de arroz (*Oryza sativa* L.) en Chile

Viviana Becerra V., Mario Paredes C., Alexis Salas B., Gabriel Donoso Ñ.

El Programa de Recursos Genéticos de INIA posee una colección activa de germoplasma de 1.200 genotipos, provenientes, principalmente, de países de clima templado. Debido a que las condiciones climáticas que presenta el país son días largos y bajas temperaturas, y a que la producción en Chile es principalmente de un grano comercial largo-ancho, el Programa de Mejoramiento Genético de Arroz (PMGA) debe desarrollar sus propias variedades. Esta situación deja al programa de mejoramiento genético en una situación compleja para obtener germoplasma, ya que no existen otros países donde se presente esta combinación de requerimientos climáticos y de mercado juntos. Germoplasma adaptado a las condiciones climáticas chilenas se podría encontrar en el norte de China, Corea y Japón. Sin embargo, el grano comercial de esos países es el típico arroz *japonica*, más pequeño y redondeado que el requerido en nuestro país. De ahí que una de las principales limitaciones de los programas de mejoramiento genético de arroz sea la baja diversidad genética disponible, lo que se ha traducido en el desarrollo de variedades con una alta vulnerabilidad genética (Kaneda, 1985; Didlay, 1990; Cuevas-Pérez et al., 1992; Montalban et al., 1998; Guimarães, 2002; Leung et al., 2015; Becerra et al., 2017).

Tradicionalmente, el germoplasma de arroz se ha evaluado por características morfológicas, agronómicas, de tolerancia a frío, rendimiento y calidad de grano. Sin embargo, las principales limitantes de esta caracterización son la influencia ambiental, el tiempo requerido para coleccionar los datos y el desconocido número de genes involucrados en estas características (Becerra y Paredes, 2000).

Actualmente, los marcadores moleculares permiten detectar cambios en el genotipo del individuo con una mayor precisión, lo que permite complementar la caracterización fenotípica convencional (Gonzaga et al., 2015; Liu et al., 2015; Reig-Valiente et al., 2016; Ndjiondjop et al., 2018; Wang et al., 2018; Park et al., 2019; Kishor et al., 2020). En este sentido, la diversidad genética de la colección del germoplasma utilizado por el PMGA no ha sido evaluada. Basado en lo anterior, en este capítulo se describen los análisis realizados sobre una muestra representativa del material genético utilizado por el PMGA con microsatélites (SSR) y polimorfismos de nucleótidos únicos (SNP) para poder estimar la diversidad genética de este material.

Situación de la diversidad y estructura genética del germoplasma de arroz en Chile

Parámetros de diversidad del germoplasma de arroz en Chile

Para determinar el nivel de diversidad, se analizaron 248 accesiones representativas del germoplasma de arroz del PMGA de INIA mediante microsatélites (SSR) ya que representó cerca de un 20% de la colección de trabajo del PMGA. Esta muestra estuvo constituida por accesiones de arroz tipo *japonica* templado y cuatro accesiones aromáticas (Basmati y Sugandh) como germoplasma contrastante. El germoplasma *japonica* estuvo formado por variedades comerciales chilenas antiguas y en uso, variedades comerciales extranjeras y líneas experimentales del PMGA de INIA, y de otros programas de mejoramiento como CIAT, IRRI y de Europa.

Las 248 accesiones fueron evaluadas con 30 microsatélites polimórficos, con la metodología descrita en Becerra et al. (2017). Estas accesiones presentaron un total de 183 alelos, con un promedio de 6,1 alelos por locus, y con un rango entre 2 y 14 (Cuadro 1). Los loci (SSR) que detectaron mayor polimorfismo fueron RM8068 (14 alelos), RM286 y RM44 (11 alelos), y RM547 (10 alelos). En contraste, los loci menos polimórficos fueron RM17 (2 alelos) y RM502, RM482, RM561, RM509 y RM555 (3 alelos). Los parámetros de diversidad genética obtenidos en este estudio (Cuadro 1) muestran que la frecuencia promedio del alelo más común fue 0,61 y el marcador RM555 mostró la mayor frecuencia alélica (0,87). La frecuencia promedio del alelo más escaso fue 0,028, mientras que los marcadores RM1261 y RM243 detectaron la menor frecuencia alélica.

Por otro lado, el promedio de genotipos diferenciados fue de 7,03 y RM8068 detectó 17 genotipos, siendo el marcador más eficiente, seguido de RM44 que diferenció 15 genotipos. La diversidad genética (H_e) promedio del germoplasma fue de 0,52, mientras que la mayor diversidad (0,73) fue detectada por los marcadores RM44 y RM241. La heterocigosidad promedio fue muy baja (0,01) y el mayor valor (0,05) detectado fue con el RM560. Cabe mencionar que varios loci (53 %) detectaron homocigosidad completa. El contenido de información polimórfica (PIC) promedio del material estudiado fue de 0,47.

En comparación con la diversidad genética de germoplasma de otras regiones, Lu et al. (2005) analizaron 145 accesiones de arroz de U.S.A. con 169 SSR, detectando 870 alelos con un promedio de alelos por locus de 5,2 (fluctuando entre 2 -21 alelos) y con un PIC promedio de 0,46, similar a la diversidad genética promedio del germoplasma chileno. La heterocigosidad en este estudio fue de 3,1 %, muy alta comparada con el valor chileno.

Estudios realizados en 150 variedades en China detectaron 1063 alelos (Zhang et al., 2011), con un promedio de 3,9 alelos por locus y un rango entre 2 y 12 alelos por locus, valores inferiores a los valores del material genético usado en Chile. El valor PIC promedio fue de 0,48. El promedio de alelos por locus para *indica* y *japonica* fue 3,71 y 3,26, respectivamente. La diversidad genética promedio para el arroz tipo *indica* y *japonica* fue de 0,48 y 0,45, respectivamente. En un segundo estudio realizado en China en 416 accesiones analizadas con 100 SSR, se detectaron 390 alelos, con un promedio de 3,9 alelos por locus, una diversidad genética de 0,47 y un valor del PIC 0,42 (Jin et al., 2010).

Otras colecciones mundiales del arroz también han sido sujetas a estudios de su diversidad genética. Es así como Yan et al. (2010), estudiando una colección mundial del USDA de 1.794 accesiones provenientes de 112 países y que conforman 14 regiones geográficas, determinaron un promedio de alelos por locus de 7,8, una frecuencia del alelo más común de 0,52, una diversidad genética de 0,61 y un valor de PIC de 0,57. Por su parte, Xu et al. (2004) determinaron en una colección de 236 accesiones *indica* y *japonica*, analizadas con 60 SSR, un total de 714 alelos con un promedio de 11,9 por locus, con una fluctuación de 2 a 34 alelos, un valor PIC promedio de 0,74 para la colección mundial (rango 0,17 - 0,92), y un valor PIC promedio de 0,50.

Un estudio genético realizado por Garris et al. (2005) mostró que los grupos *indica* y *japonica* tropical contienen un mayor porcentaje de polimorfismo (99 %) y que tienen un mayor número de alelos por locus, 7,26 y 6,09 respectivamente. En el mismo estudio, el grupo de *japonica* templado presentó la menor diversidad genética, comparado con los otros grupos de arroz, con un 91 % de loci polimórficos y 4,9 alelos por locus. El valor promedio general de PIC entre los cultivares fue 0,67, y el valor PIC de *japonica* templado fue de sólo 0,37, PIC inferior al detectado en el germoplasma usado en el germoplasma chileno (0,47).

Cuadro 1. Microsatélites, número de alelos, frecuencia mayor y menor de alelos, número de genotipos, diversidad génica, heterocigosidad y contenido de la información polimórfica (PIC).

Marcador SSR	Número de alelos	Mayor frecuencia alélica	Menor frecuencia alélica	Número de genotipos	Diversidad Génica (He)	Heterocigosidad	PIC
RM1164	5	0,62	0,004	5	0,51	0,00	0,43
RM8068	14	0,61	0,004	17	0,61	0,02	0,60
RM413	4	0,60	0,004	5	0,51	0,02	0,41
RM286	11	0,58	0,012	13	0,64	0,01	0,62
RM10	6	0,46	0,008	6	0,62	0,00	0,55
RM502	3	0,79	0,004	4	0,34	0,02	0,28
RM21	6	0,77	0,008	7	0,39	0,01	0,37
OSR28	8	0,48	0,004	12	0,67	0,03	0,62
RM44	11	0,44	0,004	15	0,73	0,03	0,70
RM276	5	0,77	0,012	5	0,39	0,00	0,36
RM259	7	0,52	0,004	7	0,63	0,00	0,57
RM1230	5	0,90	0,004	5	0,19	0,00	0,19
RM482	3	0,58	0,034	3	0,51	0,00	0,41
RM537	4	0,76	0,008	4	0,37	0,00	0,32
RM547	10	0,62	0,004	11	0,59	0,01	0,56
RM561	3	0,46	0,190	5	0,63	0,04	0,55
RM560	4	0,47	0,020	5	0,61	0,05	0,53
RM1261	7	0,55	0,002	8	0,60	0,02	0,54
RM241	10	0,42	0,016	10	0,73	0,00	0,69
RM426	8	0,67	0,004	8	0,53	0,00	0,51
RM17	2	0,64	0,357	2	0,46	0,00	0,35
RM243	7	0,47	0,002	8	0,64	0,01	0,57
RM406	6	0,81	0,008	6	0,33	0,00	0,31
RM447	4	0,85	0,008	4	0,26	0,00	0,24
RM509	3	0,77	0,036	3	0,36	0,00	0,32
RM510	5	0,52	0,008	5	0,52	0,00	0,41
RM525	7	0,42	0,012	10	0,71	0,04	0,66
RM555	3	0,87	0,028	3	0,24	0,00	0,22
RM583	7	0,49	0,004	8	0,69	0,01	0,65
RM5463	5	0,47	0,021	7	0,62	0,03	0,54
Promedio	6,1	0,61	0,028	7,03	0,52	0,01	0,47

Fuente: Becerra et al., 2017

En relación a la diversidad genética en material silvestre, un estudio de 10 especies silvestres determinó un mayor PIC (0,78) en comparación a las variedades de U.S.A. que tuvieron el PIC de 0,39, y una mayor diversidad genética promedio de 0,43, fluctuando entre 0,03 a 0,86 (Agrama y Eizenga, 2008).

En resumen, los valores de diversidad en el germoplasma chileno son, en general, inferiores a los reportados anteriormente, al considerar toda la estructura genética del arroz, es decir, *indica*, *aus*, *aromatica*, *japonica* templada y tropical (Garris et al., 2005; Lipitan et al., 2007; Caicedo et al., 2007; Liakat et al., 2011; Jamil et al., 2013), pero es muy similar comparado con los valores reportados en germoplasma *japonica* templado.

Las diferencias observadas en la diversidad entre diferentes poblaciones de arroz, se deben a su historia demográfica. La información generada por la secuenciación del arroz sugiere que *indica* y *japonica* templada divergieron hace 440.000 años atrás. Por otro lado, se ha determinado que *japonica* templado tiene una menor diversidad genética y una estrecha relación genética con *japonica* tropical (Ni et al., 2002), ya que posiblemente *japonica* templado se derivó del grupo *japonica* tropical (Garris et al., 2005) y en cada una de estas separaciones se han producido cuellos de botella que ha significado una reducción de la diversidad genética de la nueva población emergente (Zhu et al., 2007). Ante este escenario, las accesiones de *japonica* templado poseen una menor diversidad genética y se necesita ampliar esta base genética, principalmente a través de la incorporación de germoplasma de zona de origen (Zeng et al., 2007) o de grupos cercanos genéticamente como el *japonica* tropical.

Estructura del germoplasma

Los microsatélites son marcadores codominantes que tienen un amplio rango de aplicaciones prácticas en el campo del mejoramiento genético. Este trabajo es el primer análisis genotípico integral de un número representativo (248 individuos) del germoplasma *japonico templado* adaptado a las condiciones chilenas, combinado con un gran número (183) de loci.

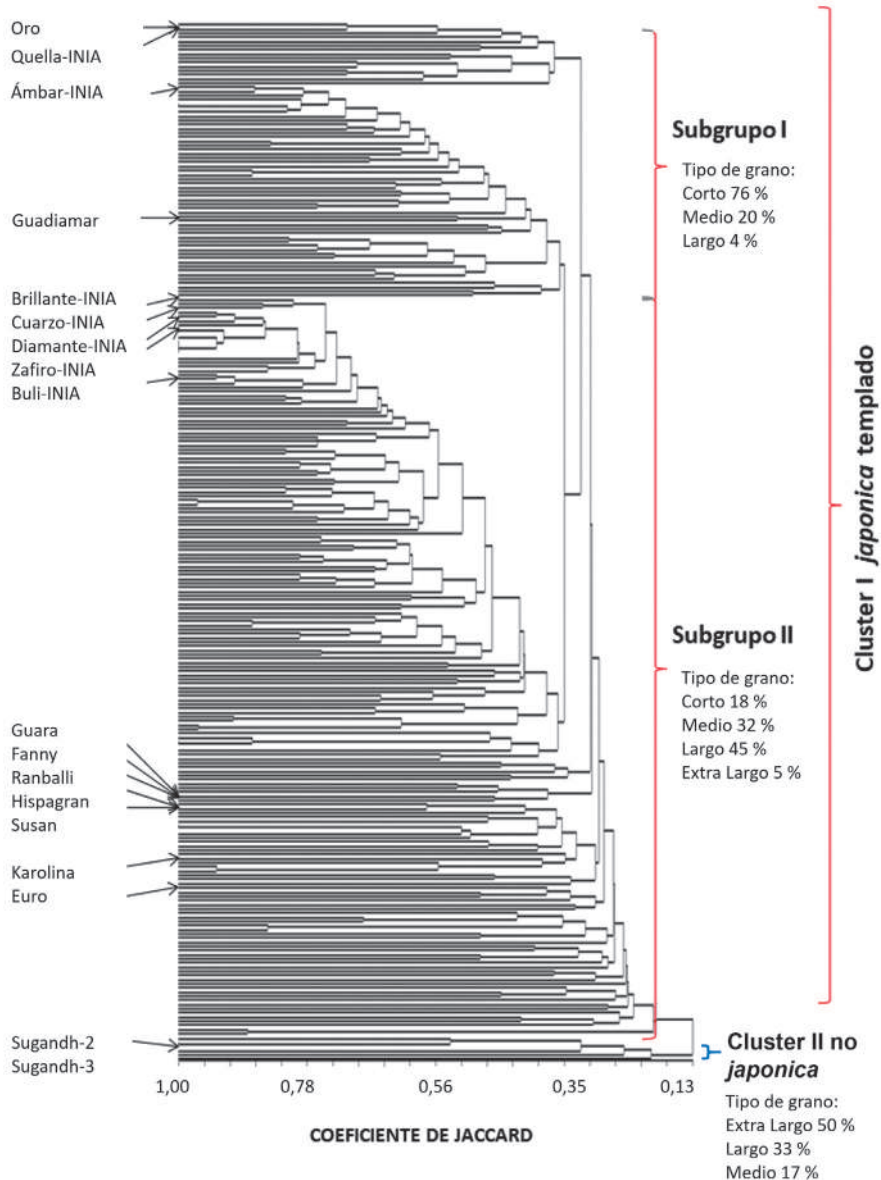
Para determinar la estructura genética del germoplasma de arroz usado en Chile, se generó un dendrograma que agrupó las accesiones en dos grupos principales (Figura 1). El grupo I incluyó 242 accesiones de *japonico* templado, mientras que el grupo II incluyó las 6 accesiones no *japonica* usadas en este estudio. Cabe señalar que la mayoría de los 248 genotipos analizados fueron genéticamente discriminados por los 30 SSR, con la excepción de los siguientes grupos de genotipos: INIA27-Quila256101; INIAG152-INIAG144; RQuila636-INIAG99-INIA169-INIAG165; Quila230602-Quila230603; Quila194603-Quila194602; y 'Sungandh-2'-'Sugandh-3'. El coeficiente de correlación cofenética entre la matriz cofenética y la matriz original de los datos SSR fue de 0,80, lo que significa que la matriz original está bien representada en el dendrograma.

El grupo I, representado por los 243 genotipos de *japonica* templado, se subdividió en dos subgrupos principales, los cuales tendieron a agruparse por tipo de grano. El primer subgrupo incluyó 66 genotipos, en su mayoría con grano corto (76 %) y medio (20 %), con una relación largo ancho de 2,0. Este subgrupo incluyó a las variedades chilenas de grano corto, 'Oro', 'Quella-INIA' y 'Ámbar-INIA'.

Por otro lado, el subgrupo II incluyó 176 genotipos, representado por accesiones de grano largo (45%) y medio (32%) con una relación largo ancho de 2,5 a 3,0. Este subgrupo incluyó las variedades chilenas de grano largo-ancho 'Brillante-INIA', 'Cuarzo-INIA', 'Diamante-INIA', 'Zafiro-INIA' y 'Buli-INIA'. Además, estas variedades presentaron una diversidad genética muy baja entre ellas, lo que señala el alto nivel de parentesco existente (Figura 1). Es así como 'Diamante-INIA' y 'Zafiro-INIA' presentaron un alto índice de similitud genética (96 %), a pesar de que fueron desarrolladas mediante diferentes métodos de selección, pedigrí y selección recurrente, lo que señala la escasa diversidad genética de los padres utilizados en su generación.

Por otro lado, las variedades introducidas desde Europa también tendieron a agruparse entre ellas ('Guara', 'Fanny', 'Hispagran', 'Susan', 'Karolina', 'Euro'), aunque son más lejanas genéticamente a las variedades chilenas.

El grupo II, que incluyó los seis genotipos no *japonica*, estuvo conformado por genotipos de grano extra largo (50 %) y largo (33 %), con una relación largo ancho sobre 3,0, entre ellos Basmati (*aromatica*), 'Sugandh-2', 'Sugandh-3' (arrozces aromáticos), 'Chu Xiang' (China), PRA557 (Madagascar) y Quila 261601. El material chileno Quila 261601 se agrupó junto a ellos, debido a que uno de sus progenitores es 'Sugandh-2'.



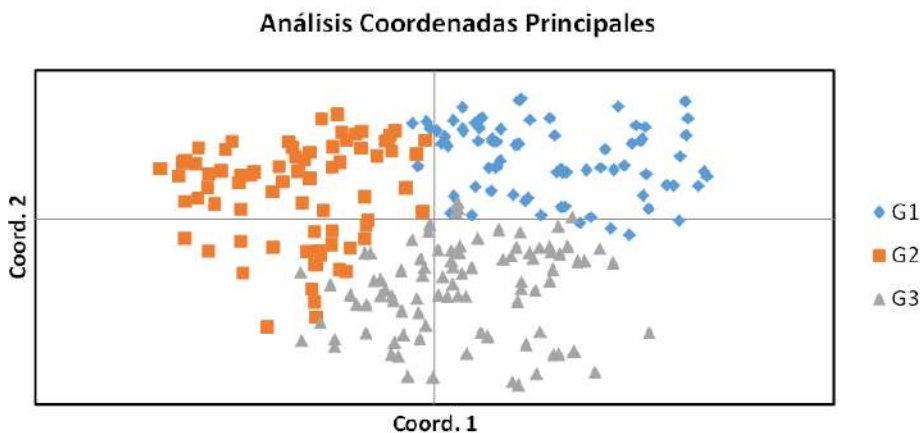
Fuente: Becerra et al., 2017.

Figura 1. Dendrograma de diversidad genética de 249 accesiones de arroz *japonica* templada chilenas, evaluadas con 30 microsatélites (Coeficiente de Jaccard).

La agrupación de las accesiones por tipo de grano ha sido reportada anteriormente en variedades de la India y Taiwán (Kumari et al., 2011; Chuang et al., 2011). En este caso, se postula que algunos SSR podrían estar asociados con regiones del genoma que determinan el tamaño del grano (Huang et al., 2013). La presencia de este tipo de marcadores sería de gran utilidad para la selección temprana, considerando que el tamaño de grano es uno de los principales criterios para determinar la estructura del germoplasma de arroz y la calidad comercial del grano (Courtois et al., 2012).

Cabe destacar que el tamaño de grano no tuvo una absoluta relación en la separación de los dos sub grupos observados en el dendrograma, pues en ambos subgrupos se observó una mezcla de los tres tipos de granos. Históricamente, las primeras variedades chilenas fueron del tipo corto-ancho, pero desde principios de 1980 hubo un gran cambio en la preferencia del consumidor y se cambió al consumo de las variedades de grano largo-ancho. El tipo de grano largo-ancho son generalmente producidos en África, América del Sur y Central, Norte América y el Caribe, y la mayoría de estos genotipos son clasificados como *japonica* tropical. Por otro lado, las accesiones del norte de China, Japón y Corea y parte de Europa, Asia Central y Rusia son clasificadas como *japonica* templada, pero poseen un grano pequeño a mediano. Por lo tanto, el gran desafío de Chile es desarrollar un germoplasma *japonica* templado con grano largo-ancho, precoz, de calidad y tolerante al frío en las etapas vegetativas y reproductivas.

El análisis de Coordenadas Principales (C) muestra la presencia de tres grupos (G1, G2 y G3), similar a lo mostrado por el dendrograma. Estas agrupaciones se distribuyen a lo largo de C1 y C2 en la Figura 2, donde el grupo 1 se concentra principalmente en el segundo cuadrante, 2/3 del Grupo 2 se concentra en el cuadrante 1, y 2/3 del grupo 3 se concentra en el cuadrante 4. Los grupos 2 y 3 entremezclan algunos genotipos en el cuadrante 3, lo mismo que los grupos 1 y 2 en el cuadrante 2. En general, cerca de un 70 % de la variación total se explica por las tres primeras coordenadas, lo que indica que los tres grupos de arroz *japonica* templada y el germoplasma no *japonica* son diferentes entre ellos.



Fuente: Becerra et al., 2017.

Figura 2. El análisis de Coordenadas Principales muestra la distribución especial de 249 accesiones de arroz *japonica templada* evaluadas con 30 microsatélites.

Determinación de la estructura de poblaciones mediante SSR

La estructura de la población de las 248 accesiones se analizó utilizando el método Bayesiano en el que cada grupo o población se basa en la probabilidad para cada número de grupos (K). Este

enfoque permite determinar el número de grupos existentes en la población en base a una mayor probabilidad (Lu et al., 2005).

Selección de deme (K) para SSR

En este estudio, los valores K fueron barridos desde 2 a 20, con 100 repeticiones de determinación con el programa 'Structure', para cada K (Figura 3). El primer panel muestra el valor promedio de máxima verosimilitud y el error para valor de K. Los datos del primer gráfico son derivados a la primera y segunda, para posteriormente calcular el delta K para cada posición. Este análisis permitió determinar un valor de K=6, como el óptimo para los datos de poblaciones ensayados.

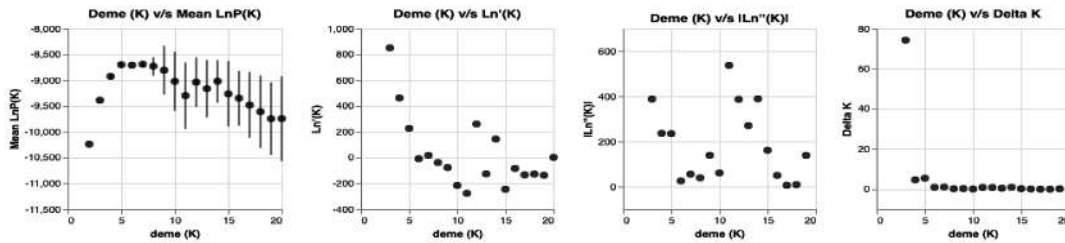


Figura 3. Exploración de valores K y resultados de máxima verosimilitud para el análisis de SSR.

Análisis de estructura de la población mediante SSR

Para el análisis de los SSR, se construyó una matriz con 248 individuos y 30 marcadores microsatélites como loci, con período de iteración de 10,000 y 20,000 repeticiones de MCMC, con un modelo Bayesiano y de máxima verosimilitud ejecutadas con el programa 'Structure' (Pritchard et al., 2000). Estos cálculos fueron paralelizados utilizando 'Structure threader' (Pina-Martins et al., 2017), realizando un barrido de valores de 'deme' desde K=2 a K=20 con 100 repeticiones de corrida de 'Structure' por K. Para obtener el mejor K se analizaron múltiples corridas de 'Structure' con el método de Evanno (Evanno et al., 2005) y se determinó el K analizando visualmente el gráfico de delta K v/s K. Al converger el valor de delta a un mínimo, se puede seleccionar uno de estos K. Se eligió aquél con menor error desde el gráfico. Este procedimiento permitió identificar a 6 grupos de genotipos.

Se seleccionó la primera repetición correspondiente al K seleccionado y se ordenó por el valor máximo de cada fracción y al grupo correspondiente. Este procedimiento permitió identificar grupos análogos a la función de clasificación de 'Structure'. Posteriormente, la composición de cada individuo fue normalizada y reducida dimensionalmente con T-SNE (Li et al., 2017), para luego ser analizada por los algoritmos de varios tipos de agrupamiento: jerárquico con el método de Ward (grupos definidos visualmente), aglomerativo con el método K-means (grupos asignados arbitrariamente), por agrupamiento jerárquico basado en densidades, HDBSCAN (McInnes et al., 2017) (no se define el número de grupos, pero sí el número mínimo de individuos).

En un gráfico de barras acumulado se grafica la proporción de cada individuo para el K=6. Los individuos son etiquetados y ordenados por el máximo valor de cada individuo y el K de pertenencia (Figura 4). En la parte superior del gráfico de barras se colorean los grupos identificados por el sorteo de 'Structure' (Structure - 6 grupos), los de agrupamiento por K-means (K-means - 6 grupos), y clustering por HDBSCAN (HDBSCAN - 5 grupos).

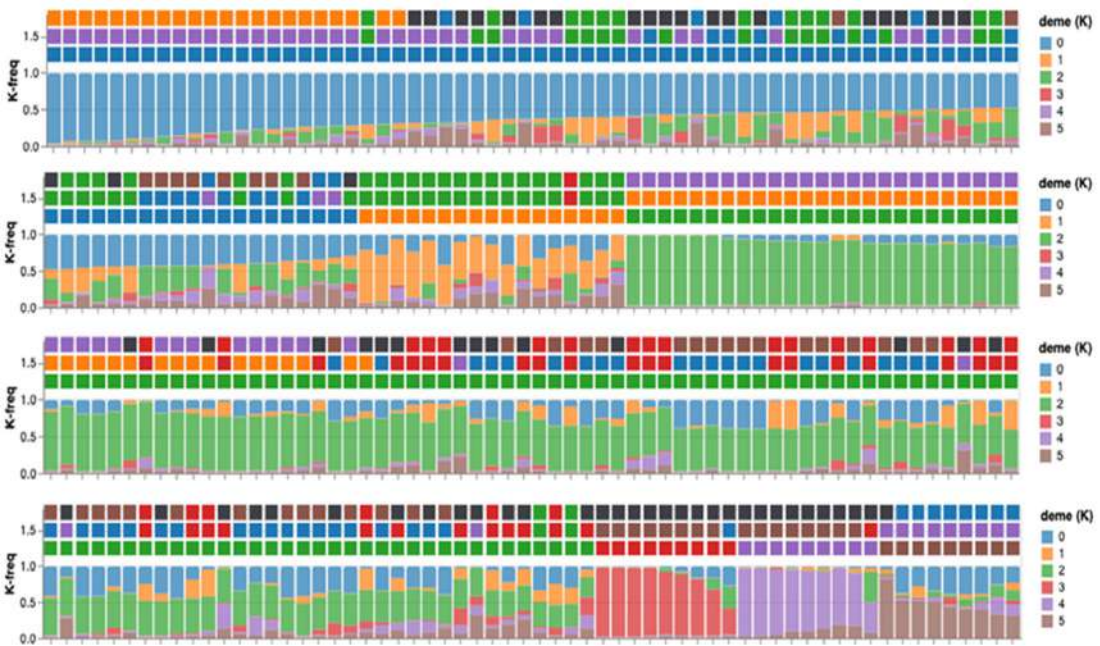


Figura 4. Germoplasma de *Oryza sativa* ordenado por fracciones genómicas obtenidas mediante el análisis de SSR.

Agrupamiento jerárquico de distancia genética mediante SSR

Se observa un dendrograma con el análisis de la matriz de distancia y el método de Ward truncado a 100 hojas, donde los valores del eje X son el número de hojas colapsado o la etiqueta del individuo. Para determinar los grupos se realizó un corte arbitrario a 60 de distancia, donde se identificaron 5 grupos (en colores verde, rojo, celeste, moradas y amarillo) (Figura 5).

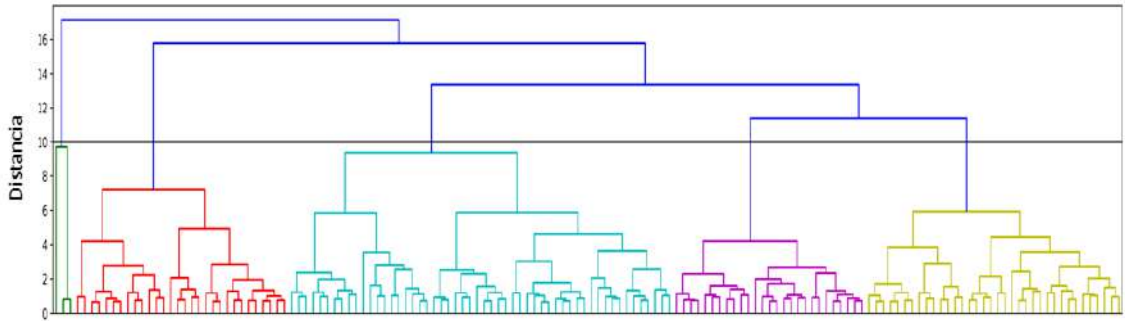


Figura 5. Dendrograma Jerárquico de Agrupamiento (Truncado), basado en distancia genética de *Oryza sativa* analizado mediante SSR.

Determinación de la estructura de poblaciones mediante SNP

Selección de 'deme' (K) para SNP

Los valores K fueron explorados mediante barridos desde 2 a 20, con 100 repeticiones de determinación con 'Structure' para cada K (Figura 6). El primer panel muestra el valor promedio de máxima verosimilitud y el error para valor de K. Los datos del primer gráfico son derivados a la primera y segunda, para posteriormente calcular el delta de K para cada posición. Este procedimiento permitió determinar el valor de K=6, como el óptimo para los datos poblaciones ensayados.

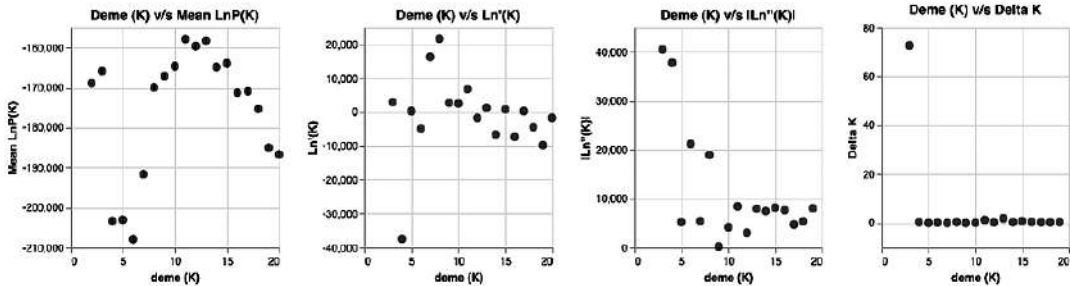


Figura 6. Exploración de valores K y resultados de máxima verosimilitud para el análisis de SNP.

Análisis de estructura de la población mediante SNP

Para el análisis de los SNP, se construyó una matriz con los mismos 248 genotipos de *O. sativa* y 3196 loci desde los SNP. 'Structure' (Pritchard et al., 2000) fue ejecutado con parámetros por defecto de quemado de 10,000 y 20,000 repeticiones de MCMC. La ejecución de 'Structure' fue paralelizada por 'Structure threader' (Pina-Martins et al., 2017), en un barrido de 2-20 K en 100 repeticiones con un total de 2,000 simulaciones. Para obtener el mejor K se analizaron múltiples corridas de 'Structure' con el método de Evanno y se determinó el K, analizando visualmente el gráfico de delta K v/s K. Al converger el valor de delta a un mínimo, se puede seleccionar uno de estos K, para decidir aquel con menor error desde el gráfico.

Al igual que en el análisis de los SSR, para los SNP se seleccionó la primera repetición correspondiente al K seleccionado, y se ordenó por el valor máximo de cada fracción y a qué 'cluster' corresponde, lo cual permitió identificar grupos análogamente a la función 'Clasificación' de 'Structure'. Luego, la composición de cada individuo fue normalizada y reducida dimensionalmente con T-SNE (Li et al., 2017), para después ser analizada por los algoritmos de agrupamiento: jerárquico con método de Ward (grupos definidos visualmente), aglomerativo con el método K-means (grupos asignados arbitrariamente), por 'clustering' jerárquico basado en densidades, HDBSCAN (McInnes et al., 2017) (no se define el número de grupos, pero sí el número mínimo de individuos).

En un gráfico de barras acumulado, se grafica la proporción de cada individuo para el K=6. Los individuos son etiquetados y ordenados por máximo el valor de cada individuo y el K de pertenencia.

En la parte superior del gráfico de barras se colorean los grupos identificados por el sort de 'Structure' (Structure - 6 grupos), los de agrupamiento por K-means (K-means - 6 grupos), y 'clustering' por HDBSCAN (HDBSCAN - 3 grupos) (Figura 7).

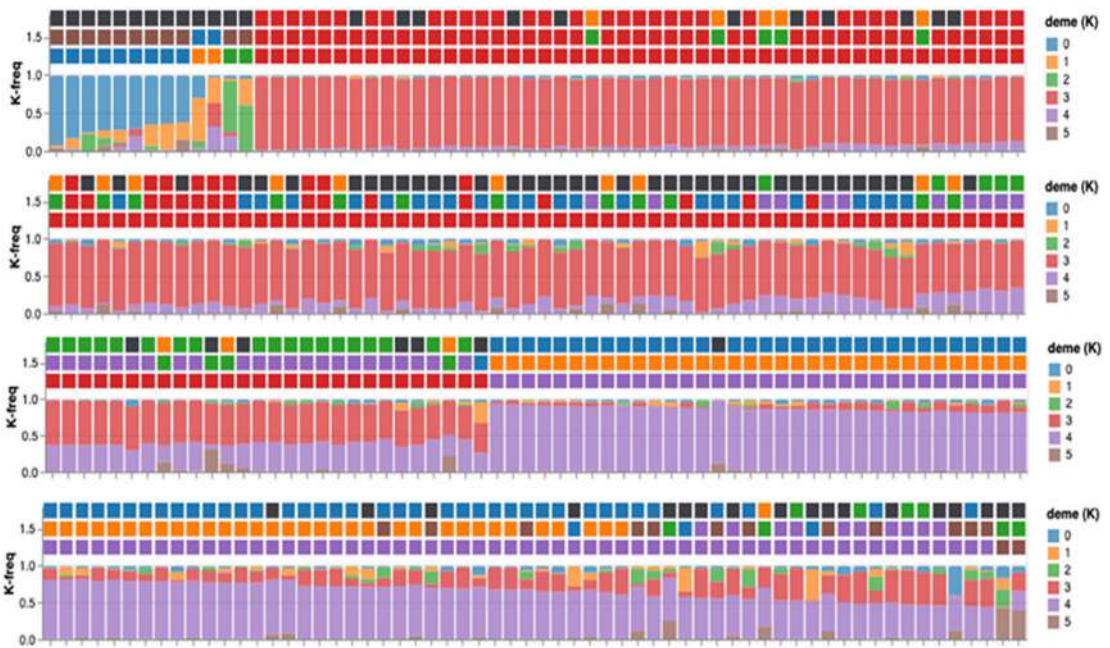


Figura 7. Germoplasma de *Oryza sativa* ordenado por fracciones genómicas obtenidas mediante el análisis de SNP.

Agrupamiento jerárquico por matrices de distancia mediante SNP

Se observa un dendrograma con el análisis de la matriz de distancia y el método de Ward truncado a 150 hojas, donde los valores del eje X son el número de hojas colapsado o la etiqueta del individuo. Para determinar los grupos se realizó un corte arbitrario a 60 de distancia. Se identificaron 5 grupos (en colores verde, rojo, celeste, morado y café) (Figura 8).

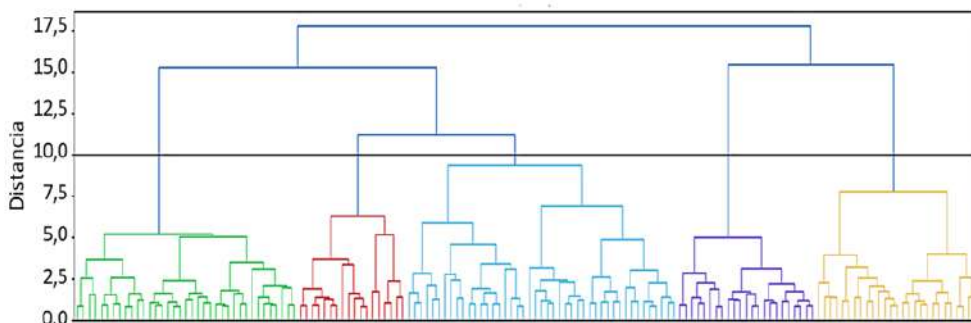


Figura 8. Dendrograma Jerárquico de Agrupamiento (Truncado) de *Oryza sativa* mediante SNP.

El número de grupos obtenidos fue el mismo para SSR y SNP; sin embargo, las agrupaciones de las accesiones son distintas con estos marcadores. Con SNP no hubo una agrupación relacionada con el tipo de grano. El alto número de grupos identificados indica que el material genético, pese a ser *japonica* templado, tiene un alto grado de introgresiones provenientes de otro tipo de germoplasma, producto de las numerosas introducciones y cruza reiteradas en el PMGA. A pesar de las variaciones detectadas por los SSRs y los SNPs en el germoplasma, éstas no han influido en gran medida en el nivel de diversidad genética observada a nivel molecular como fenotípico en las variedades desarrolladas.

Situación de la diversidad y estructura genética de variedades comerciales en Chile

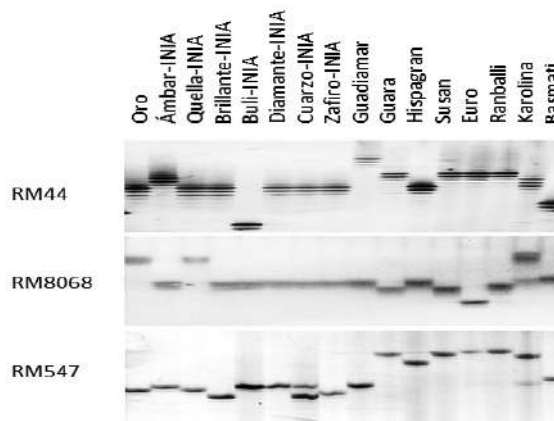
Los estudios de variedades de arroz mediante herramientas moleculares son escasos en nuestro país, siendo una actividad de crucial importancia para un programa de mejoramiento genético, pues permite identificar y certificar variedades, como también determinar la relación genética entre ellas. En Chile existen variedades comerciales de grano largo ancho y corto principalmente, las cuales han sido generadas por el programa de mejoramiento genético de arroz del INIA (Paredes et al., 2015).

El 'fingerprinting' o huella genética como tecnología, para elaborar perfiles genéticos únicos mediante el uso de microsatélites, marcadores posibles de usar a nivel operativo, permite distinguir variedades de arroz (McCouch et al., 2002; Garris et al., 2005; Lu et al., 2005; Chuang et al., 2011; Islam et al., 2011; Kanawapee et al., 2011), incluso cuando las variedades tienen un fondo genético muy estrecho (muy emparentadas) (Zhu et al., 2012). También se ha logrado identificar variedades de alto rendimiento (Xin et al., 2005) y adulteraciones de arroz élite de grano fino en lotes de semillas (Kumari et al., 2011). Incluso en Taiwán, los SSR se utilizaron para identificar el arroz pulido, comercializado en el mercado interno, para prevenir la actividad comercial fraudulenta de variedades protegidas de arroz (Chuang et al., 2011).

El objetivo de este estudio fue determinar el grado de parentesco entre las variedades comerciales chilenas y su relación con variedades de otras regiones geográficas. Las variedades chilenas analizadas fueron: 'Oro', 'Ámbar-INIA', 'Quella-INIA', 'Brillante-INIA', 'Buli-INIA', 'Diamante-INIA', 'Cuarzo-INIA', 'Zafiro-INIA', y las extranjeras 'Guadamar', 'Guara', 'Hispagran', 'Susan', 'Euro', 'Ranballi' y 'Karolina', usando Basmati del grupo de arroces aromáticos como genotipo diferenciador, y utilizando microsatélites.

Parámetros de diversidad genética de las variedades comerciales

Cuando se analizaron 16 variedades comerciales chilenas con 54 SSR, se detectó un total de 191 alelos, con una media de 3,5 alelos por locus (Becerra et al., 2015). Un 65 % de los SSR analizados detectaron sólo 2 o 3 alelos por locus, comparados con los 15 alelos por locus, observados en variedades brasileñas (Brondani et al., 2006). Los loci más polimórficos fueron RM547 (8 alelos), RM44 (7 alelos) y RM8068 (6 alelos) (Figura 9). La diversidad genética promedio fue de 0,50 y el promedio de heterocigosidad fue 0,002, valor esperado en una especie de autopolinización.



Fuente: Becerra et al., 2015.

Figura 9. Patrón de bandas detectadas en 16 variedades comerciales de arroz con los SSR RM44, RM8068, y RM547.

El valor promedio PIC fue de 0,44 con un rango de fluctuación de 0,12 a 0,83. Setenta y dos alelos únicos fueron detectados en ciertas variedades (frecuencia $\leq 0,07$). Los marcadores microsatélites polimórficos que muestran diferentes bandas podrían ser utilizados como identificadores para una variedad específica.

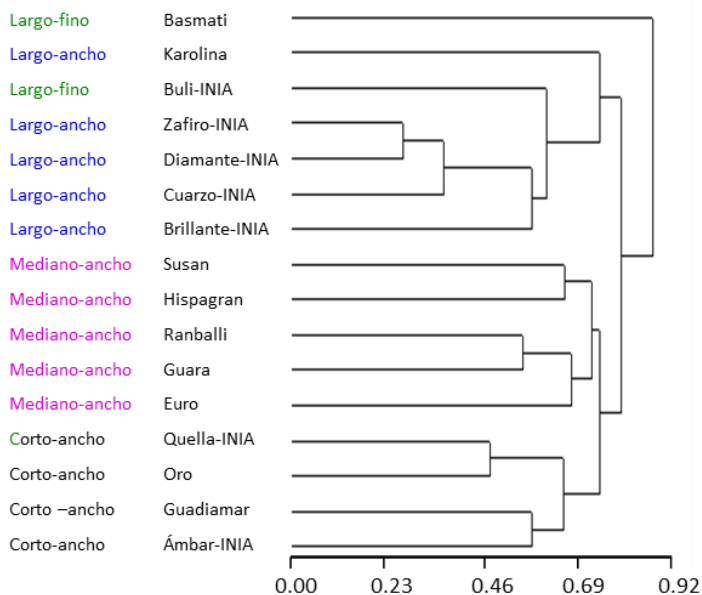
Dado el alto grado de parentesco de las variedades generadas por el programa de arroz del INIA, fue necesario usar un mayor número de SSR (54 SSR) para realizar este estudio. El promedio de alelos fue de 3,5, inferior a los 4,4 obtenidos por Bonow et al. (2009) y a los 11,7 reportados por Rahman et al., (2010) en variedades indias locales.

Este resultado confirma que la diversidad genética es baja entre las variedades comerciales analizadas. Los marcadores RM547 y RM1164 fueron muy importantes en la diferenciación genética de las variedades, lo que corrobora la utilidad de los microsatélites en la identificación de variedades que exhiben una base genética estrecha (Zhu et al., 2012).

Es sabido que el material genético de los programas de mejoramiento es más uniforme debido a la alta presión de selección y al uso de ciertos métodos de selección. Es más, el uso repetido de los padres agronómica e industrialmente aceptados, como progenitores en el programa de cruza, ha contribuido aún más a disminuir la diversidad del material genético. Ante esta situación, es importante la introducción constante de germoplasma desde otras regiones, para aumentar la base genética de los programas de mejoramiento genético.

Agrupamiento de variedades comerciales basado en distancia genética mediante SSR

Las 16 variedades analizadas se dividieron en dos grupos principales. El primero contiene las 15 variedades *japonica* agrupadas separadamente del genotipo aromático (Basmati) con una media general de 0,76. El análisis de clúster por el método UPGMA condujo a un dendrograma con un índice de correlación cofenético de 0,93, lo que indica que las distancias reflejan la situación real de la muestra estudiada (Figura 10).



Fuente: Becerra et al., 2015.

Figura 10. Dendrograma basado en la distancia genética de 16 variedades comerciales basadas en un análisis de 54 microsatélites (Coeficiente de Rogers).

El dendrograma muestra dos subgrupos principales dentro de las variedades *japonica*, con una tendencia a agruparse por tipo de grano (Figura 10). La rama inferior contiene las variedades de grano-corto 'Quella-INIA', 'Oro', 'Guadiamar', 'Ámbar-INIA' y las de grano mediano-ancho 'Susan', 'Hispagran', 'Ranballi', 'Euro' y 'Guara', con valores de relación largo-ancho entre 1,8 y 2,2.

La variedad 'Oro' tuvo una amplia aceptación en el mercado nacional desde 1960 hasta 1980, momento en que las variedades de grano largo ancho empezaron a aumentar su participación en el mercado. Esta variedad ha sido utilizada como progenitor en el PMGA de INIA, de hecho, fue uno de los padres de la variedad de grano corto 'Quella-INIA', creada en 1979, lo que puede explicar la cercanía genética de estas dos variedades (Figura 10).

Por el contrario, el subgrupo de la rama superior incluye la mayoría de las variedades comerciales INIA de semilla de largo-ancho, como 'Diamante-INIA', 'Zafiro-INIA', 'Cuarzo-INIA' y 'Brillante-INIA', así como la variedad húngara 'Karolina'. Los resultados obtenidos señalan la cercanía genética de las variedades 'Diamante-INIA' y 'Zafiro-INIA' y su relación con 'Cuarzo-INIA' y 'Brillante-INIA' (Figura 11). Por otro lado, la variedad 'Buli-INIA' que también se encuentra en este grupo, se diferencia de las anteriores por poseer un grano largo-fino. Cabe destacar que las variedades comerciales INIA de grano largo-ancho en uso, contienen a 'Diamante-INIA' en su fondo genético, pues esta variedad es y ha sido usada por largo tiempo como progenitora en las cruces que generan nuevas variedades. Ello explica su cercanía genética ya que fue la primera variedad de grano largo-ancho sembrada en el país y creada por INIA. Separadamente, se ubica Basmati (grano largo-fino) en la rama superior del dendrograma.

Estos resultados coinciden con los encontrados por Kumari et al. (2011), donde los microsatélites fueron capaces de diferenciar las variedades de India de grano medio-fino y variedades de arroz nacionales y extranjeras en Taiwán (Chuang et al., 2011). Previamente, se ha informado que algunos SSR están asociados con regiones de ADN genómico que determinan el tipo de grano (Huang et al., 2013).

La necesidad de utilizar 54 SSR para discriminar entre 16 variedades, principalmente de tipo *japonica*, indica el alto grado de parentesco, especialmente entre las variedades chilenas. Esto a diferencia de otros estudios, en los que se han usado un menor número de SSR para discriminar híbridos (He et al., 2012) y variedades (Brondani et al., 2006; Rahman et al., 2010).

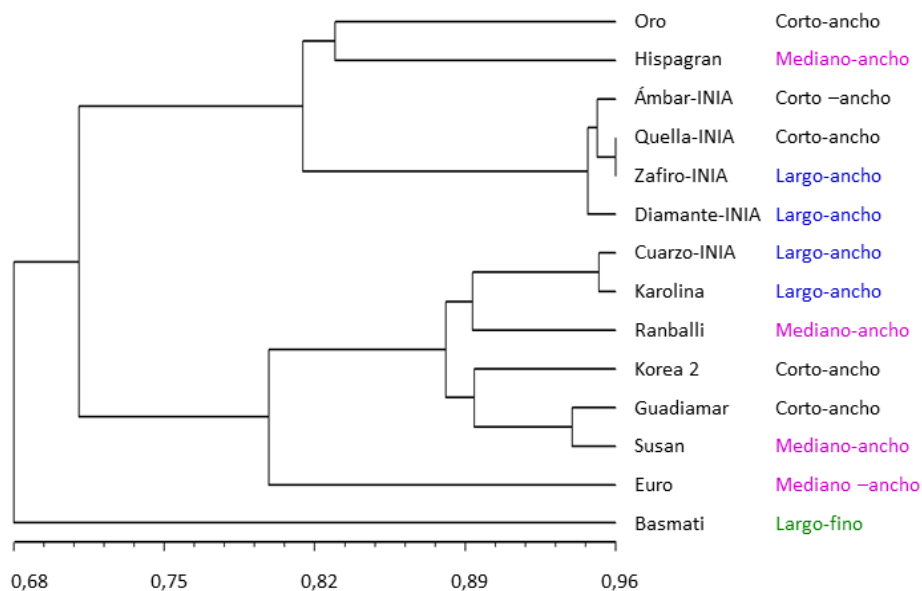
Agrupamiento de variedades comerciales basado en distancia genética mediante SNP

Los SNP corresponden a un cambio único de un par de bases en la secuencia de ADN y que ocurre con una frecuencia alta en el genoma de las especies. Estos cambios pueden ocurrir en regiones codificantes y no codificantes del ADN nuclear y de los plastidios, pudiendo afectar características específicas, o bien representan variaciones neutrales (McCouch et al., 2010). Los SNPs tienen dos alelos, lo que significa que en una población existe la posibilidad de dos pares de bases por cada sitio del SNP y su frecuencia a lo largo de cada cromosoma ha mostrado una distribución desuniforme del polimorfismo (Nasu et al., 2002).

Hoy, los SNP están reemplazando los microsatélites (SSR) como marcadores de ADN, debido a su mayor abundancia en el genoma, mayor cobertura y posición conocida en el genoma (Thomson, 2014), lo que aumenta la exactitud en mapeo y la resolución. Además, son automatizados y tienen un costo-efectividad creciente (McCouch et al. 2010, Tam et al., 2019). Una gran cantidad de marcadores SNP han sido desarrollados para el estudio del arroz (Feltus et al., 2004; Yamamoto et al., 2010; Arai-Kichise, 2011; Mvuyekure et al., 2018).

Catorce variedades comerciales de arroz fueron analizadas por 3050 SNP (Geneseek). El análisis de similitud de los datos obtenidos muestra, en el dendrograma, dos grupos principales (Figura

11). El primero contiene 6 variedades, 'Oro' asociada con 'Hispagran' y 'Ámbar-INIA', 'Quella-INIA', 'Zafiro-INIA' y 'Diamante-INIA'. Este grupo se mezclan variedades de grano corto, mediano y largo-ancho. El segundo grupo que incluye 'Cuarzo-INIA', 'Karolina', 'Ranballi', 'Korea 2', 'Guadamar', 'Susan' y 'Euro' corresponden a variedades de grano largo-ancho y corto. En forma separada se encuentra el genotipo aromático (Basmati), con una media general de 0,68. Este resultado confirma que los SNP fueron eficientes en diferenciar genotipos de distinto origen.



Fuente: Becerra et al., 2015.

Figure 11. Dendrograma de 14 variedades comerciales basadas en un análisis de SNP (Coeficiente de Jaccard).

Sin embargo, comparando los resultados obtenidos con los SSR y SNPs en el análisis genético de las variedades comerciales, aquellos obtenidos con los SSR presentan una mayor relación con características agronómicas, pues separa las variedades por tipo de grano con mayor exactitud.

Los marcadores SNP son una poderosa herramienta genética, no solo para investigación genómica, ya que, en mejora genética, pueden ser usados para mapeo de QTLs, clonación posicional, haplotipos y análisis de pedigrí, prueba de pureza de semillas e identificación varietal (Nasu et al., 2002; Mc Nally et al., 2009; Mvuyekure et al. 2018).

Consideraciones finales

Este es el primer estudio de diversidad genética del germoplasma del PMGA de INIA mediante SSR y SNP, lo cual representa cerca de un 20% de la colección de trabajo del programa. Los valores de diversidad en el germoplasma chileno son, en general, inferiores a los reportados cuando se considera toda la estructura genética del arroz, pero es similar comparado con valores reportados en germoplasma *japonica* templado.

Dentro del germoplasma se observa una alta cercanía genética entre las variedades chilenas, lo que fue confirmado por el alto número de SSR utilizado. Como se mencionó anteriormente, esta situación se puede deber a los métodos de mejoramiento utilizados y a la baja diversidad genética de los progenitores utilizados. Ante este escenario, este germoplasma necesita ampliar su base genética, principalmente a través de la incorporación de germoplasma de zona de origen

o de grupos cercanos genéticamente, como el *japonica* tropical. No se debe olvidar que el gran desafío de Chile es desarrollar variedades *japonica* templado de grano largo-ancho, precoces, de alta calidad de grano y tolerantes al frío en las etapas vegetativa y reproductiva.

El fondo genético del material analizado con SSR y SNP, generó el mismo número de grupos. Sin embargo, las agrupaciones de las accesiones son distintas con estos marcadores. La agrupación de las accesiones del material chileno mediante SSR, tendió a ser por tipo de grano, lo cual ha sido reportado con anterioridad en otros germoplasmas de arroz. Ello se podría deber a que algunos SSR podrían estar asociados con regiones del genoma que determinan el tamaño del grano. Estos marcadores SSR podrían ser de gran utilidad para la selección temprana, considerando que el tamaño del grano es uno de los principales criterios para determinar la estructura del germoplasma de arroz y la calidad comercial del mismo, lo que también se observó cuando se analizaron sólo las variedades comerciales. En cambio, con SNP no hubo una agrupación relacionada con el tipo de grano. Ello indica que la selección de uno u otro marcador molecular debe estar relacionado con el objetivo del estudio.

El alto número de grupos identificados por ambos marcadores indica que el material genético del PMGA, pese a ser clasificado como *japonica* templado, tiene un alto grado de introgresiones provenientes de otro tipo de germoplasma, posiblemente producto de las numerosas introducciones y cruces reiteradas en el PMGA.

Referencias

- Agrama, H.A., Eizenga, G.C. 2008. Molecular diversity and genome-wide linkage disequilibrium patterns in a worldwide collection of *Oryza sativa* and its relatives. *Euphytica* 160:339-355.
- Arai-Kichise, Y. 2011. Discovery of genome-wide DNA polymorphisms in a landrace cultivar of *japonica* rice by whole-genome sequencing. *Plant Cell Physiol.* 52:274-282.
- Becerra, V., Paredes, M. 2000. Uso de marcadores bioquímicos y moleculares en estudios de diversidad genética. *Agric. Técnica (Chile)* 60:270-281.
- Becerra, V., Paredes, M., Ferreira, M., et al. 2017. Assessment of the genetic diversity and population structure in temperate *japonica* rice germplasm used in breeding in Chile, with SSR markers. *Chil. J. Agric. Res.* 77:15-26.
- Becerra, V., Paredes, M., Gutiérrez, E., et al. 2015. Genetic diversity, identification, and certification of Chilean rice varieties using molecular markers. *Chil. J. Agric. Res.* 75:267-274.
- Bonow, S., Von Pinho, V.R., Vieira, G.C., et al. 2009. Microsatellite markers in and around rice genes: Applications in variety identification and DUS testing. *Crop Sci.* 48:880-886.
- Brondani, C., Oliveira, T., Hideo, P., et al. 2006. Determination of genetic variability of traditional varieties of Brazilian rice using microsatellite markers. *Genet. Mol. Biol.* 29:676-684.
- Caicedo, A.L., Williamson, S.H., Hernández, R.D., et al. 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet.* 3:1745-1756.
- Chuang, H., Lur, H.S., Hwu, K.K., et al. 2011. Authentication of domestic Taiwan rice varieties based on fingerprinting analysis of microsatellite DNA markers. *Bot. Stud.* 52:393-405.
- Courtois, B., Frouin, J., Greco, R., et al. 2012. Genetic diversity and population structure in a European collection of rice. *Crop Sci.* 52:1663-1675.
- Cuevas-Pérez, F., Guimaraes, E., Berrio, L., et al. 1992. Genetic base of irrigated rice in Latin America and the Caribbean, 1971 to 1989. *Crop Sci.* 32:1054-1059.
- Didlay, R. 1990. Contribution of ancestral lines in the development of new cultivars of rice. *Crop Sci.* 30:905-911.

- Evanno, G., Regnaut, S., Goudet, J. 2005. Detecting the number of clusters of individuals using software structure: a simulation study. *Mol. Ecology* 14:2611-2620.
- Feltus, F.A., Wan, J., Schulze, S.R., et al. 2004. An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. *Genome Res.* 14:1812-1819.
- Garris, A.J., Tai, T.H., Coburn, J., et al. 2005. Genetic structure and diversity in rice *Oryza sativa* L. *Genetics* 169:1631-1638.
- Gonzaga, Z., Aslam, K., Septiningsih, E., et al. 2015. Evaluation of SSR and SNP markers for molecular breeding in rice. *Plant Breed. Biotech.* 3:139-152.
- Guimarães, E. 2002. Genetic diversity of rice production in Brazil. p. 11-35. In Nguyen, V.N. (ed.) *Genetic diversity in rice. Case studies from Brazil, India and Nigeria.* FAO, Rome, Italy.
- He, Z., Xie, F., Chen, L., et al. 2012. Genetic diversity of tropical hybrid rice germplasm measured by molecular markers. *Rice Sci.* 19:193-201.
- Huang, R., Jian, L., Zheng, J., et al. 2013. Genetic bases of rice grain shape: so many genes, so little known. *Trends Plant Sci.* 18:218-226.
- Islam, M.M., Hoque, M.E., Rabbi, S.M.H.A., et al. 2011. DNA Fingerprinting and diversity analysis of BRRI hybrid varieties and their corresponding parents. *Plant Tissue Cult. Biotechnol.* 21:189-198.
- Jamil, M., Rana, I.A., Ali, Z., et al. 2013. Estimation of genetic diversity in rice (*Oryza sativa* L.) genotypes using Simple Sequence Repeats. *Mol. Plant Breed.* 36:285-291.
- Jin, L., Lu, Y., Xiao, P., et al. 2010. Genetic diversity and population structure in a diverse set of rice germplasm for association mapping. *Theor. Appl. Genet.* 121:475-487.
- Kanawapee, N., Sanitchon, J., Srihaban, P., et al. 2011. Genetic diversity analysis of rice varieties (*Oryza sativa* L.) differing in salinity tolerance based on RAPD and SSR markers. *Electron. J. Biotechnol.* 14:1-14.
- Kaneda, C. 1985. Development of very high yielding rice varieties. *Farm. Japan* 19:25-29.
- Kishor, D., Seo, J., Chin, J., et al. 2020. Evaluation of whole-genome sequence, genetic diversity, and agronomic traits of Basmati rice (*Oryza sativa* L.). *Front. Genet.* 11:86.
- Kumari, A.K., Dugarani, V.C., Vanisree, S., et al. 2011. Molecular fingerprinting of the elite, fine-grain type rice variety, Samba Mahsuri (BPT 5204) and assessment of genetic purity in seed-lots of the variety using SSR markers. *Int. J. Sci. Res.* 1:1-16.
- Leung, H., Raghavan, C., Zhou, B., et al. 2015. Allele mining and enhanced genetic recombination for rice breeding. *Rice* 8:34.
- Li, W., Cerise, J., Yang, Y., et al. 2017. Application of T-SNE to human genetic data. *J. Bioinform. Comput. Biol.* 15:1750017.
- Liakat-Ali, M., McClung, A.M., Jia, M.H., et al. 2011. A rice diversity panel evaluated for genetic and agro-morphological diversity between subpopulations and its geographic distribution. *Crop Sci.* 51:2021-2035.
- Lipitan, V., Brar, D.S., Abe, T., et al. 2007. Assessment of genetic diversity of Philippine rice varieties carrying good quality traits using SSR markers. *Breed. Sci.* 57:263-270.
- Liu, D., Wang, J., Wang, X. et al. 2015. Genetic diversity and elite gene introgression reveal the *japonica* rice breeding in northern China. *J. Integr. Agric.* 14:811-822.

- Lu, H., Redus, M.A., Coburn, J.R., et al. 2005. Population structure and breeding patterns of 145 US rice cultivars based on SSR markers analysis. *Crop Sci.* 45:66-76.
- McCouch, S.R., Teytelman, L., Xu, Y., et al. 2002. Development and mapping of 2240 new SSR markers for rice. *DNA Res.* 9:199-207.
- McCouch, S.R., Zhao, K., Wright, M., et al. 2010. Development of genome-wide SNP assay for rice. *Breed. Sci.* 60:524-535.
- McInnes, L., Healy, J., Astels, S. 2017. HdbSCAN: Hierarchical density based clustering. *J. Open Source Softw.* 2:205.
- McNally, K.L., Childs, K.L., Bohnert, R., et al. 2009. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. USA* 106:961-967.
- Montalban, R., Destro, D., Silva, E., et al. 1998. Genetic base of Brazilian upland rice cultivars. *J. Genet. Breed.* 52:203-209.
- Mvuyekure, S., Sibiya, J., Derera, J., et al. 2018. Assessment of genetic diversity of rice based on SNP markers for selection of parents for sheat rot (*Sarocladium oryzae*) resistance breeding. *S. Afr. J. Plant Soil* 35:51-59.
- Nasu, S., Suzuki, J., Ohta, R., et al. 2002. Search and for analysis of single nucleotide polymorphisms (SNP) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Res.* 9:163-171.
- Ndjiondjop, M., Semagn, S., Sow, M., et al. 2018. Assessment of genetic variation and population structure of diverse rice genotypes adapted to lowland and upland ecologies in Africa using SNPs. *Front. Plant Sci.* 9:446.
- Ni, J., Colowit, P.M., Mackill, D.J. 2002. Evaluation of genetic diversity in rice subspecies using microsatellite markers. *Crop Sci.* 42:601-607.
- Paredes, M., Becerra, V., Vega, A. 2015. Variedades tradicionales desarrolladas y recomendadas. p. 37-39. En Paredes, M., Becerra, V. (eds.) *Manual de Producción de Arroz: Buenas Prácticas Agrícolas (BPA)*. Instituto de Investigaciones Agropecuarias, INIA. 100p.
- Park, J., Yang, W., Kwon, Y. et al. 2019. Assessment of the genetic diversity of rice germplasm characterized by black-purple and red pericarp color using Simple Sequence Repeat markers. *Plants* 8:471.
- Pina-Martins, F., Silva, D., Fino, J., et al. 2017. Structure_threader: An improved method for automation and parallelization of programs Structure, fastStructure and MaverickK on multicore CPU Systems. *Mol. Ecol. Res.* 17(6):e268-e274.
- Pritchard, J. K., Stephens, M., Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Rahman, M.S., Sohag, M.K.H., Rahman, L. 2010. Microsatellite based DNA fingerprinting of 28 local rice (*Oryza sativa* L.) varieties of Bangladesh. *J. Bangladesh Agric. Univ.* 8:7-17.
- Reig-Valiente, J., Viruel, J., Sales, E., et al. 2016. Genetic diversity and population structure of rice varieties cultivated in temperate regions. *Rice* 9:58. Doi:10.1186/s12284-016-0130-5.
- Tam, N.T., Dwiyanti, M.S., Koide, Y., et al. 2019. Profiling SNP and nucleotide diversity to characterize Mekong Delta rice landraces in Southeast Asian populations. *Plant Genome* 12:190042.

- Thomson, M.J. 2014. High-throughput SNP genotyping to accelerate crop improvement. *Plant Breed. Biotechnol.* 2:195-212.
- Wang, W., Maulen, R., Hu, Z., et al. 2018. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557:43.
- Xin, Y., Zhang, Z., Xiong, Y., et al. 2005. Identification and purity test of super hybrid rice with SSR molecular markers. *Rice Sci.* 12:7-12.
- Xu, Y., Beachel, H., McCouch, S. 2004. A marker-based approach to broadening the genetic base of rice in the USA. *Crop Sci.* 44:1947-1959.
- Yamamoto, T., Nagasaki, H., Yonemaru, J., et al. 2010. Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. *BMC Genomics* 11:267.
- Yan, W.G., Agrama, H., Jia, M., et al. 2010. Geographic description of genetic diversity and relationships in the USDA rice world collection. *Crop Sci.* 50:2406-2417.
- Zeng, Y., Zhang, H., Li, Z., et al. 2007. Evaluation of genetic diversity of rice landraces (*Oryza sativa* L.) in Yunnan, China. *Breed. Sci.* 57:91-99.
- Zhang, H.D., Zhang, M., Wang, M., et al. 2011. A core collection and mini core collection of *Oryza sativa* L. in China. *Theor. Appl. Genet.* 122:49-61.
- Zhu, Y., Qin, G., Hu, J., et al. 2012. Fingerprinting and variety identification of rice (*Oryza sativa* L.) based on simple sequence repeat markers. Seed Science Center, College of Agriculture and Biotechnology, Zhejiang University, China. *Plant Omics J.* 5:421-426.
- Zhu, Q., Zheng, X., Luo, J., et al. 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol. Biol. Evol.* 24:875-888.

