

## 22- VALIDACIÓN DE SNP ASOCIADOS A VARIABLES DE INTERÉS EN GERMOPLASMA JAPONICA TROPICAL DE INIA

J. E. Rosas<sup>1</sup>, B. Sprunck<sup>2</sup>, W. Iriarte<sup>3</sup>, I. Rebollo<sup>4</sup>, V. Bonnacarrère<sup>5</sup>, F. Molina<sup>6</sup>, P. Blanco<sup>7</sup>, F. Pérez de Vida<sup>8</sup>

**PALABRAS CLAVE:** Marcadores moleculares, Mapeo asociativo, Selección asistida.

### INTRODUCCIÓN

El programa de mejoramiento genético de arroz de INIA (PMGA) requiere la implementación de selección asistida por marcadores moleculares (MAS) en forma rutinaria para mejorar la eficiencia de la selección y acelerar la ganancia genética. Existen diversas estrategias para identificar marcadores asociados a rasgos de interés que permitan utilizar MAS. Una de ellas es el mapeo asociativo genómico (GWAS por sus siglas en inglés). El PMGA y la Unidad de Biotecnología de INIA ejecutaron el proyecto de mapeo asociativo en arroz (PMAA) estudiando la asociación entre más de 70.000 marcadores de tipo SNP y los principales rasgos cuantitativos de interés para el programa. Se identificaron varios loci asociados a estos rasgos (QTL). Las variables mapeadas fueron tiempo de floración (FT, Rosas *et al.*, 2017), altura de planta (PH, Rosas *et al.*, 2017), yesado (GC, Quero *et al.*, 2018), % de entero (PHR, Quero *et al.*, 2018), blanco total (YAM, Quero *et al.*, 2018) y resistencia a Brusone (LB, Rosas *et al.*, en preparación). Este trabajo busca aplicar esos descubrimientos en la implementación rutinaria de MAS en el PMGA, mediante la validación de los SNP identificados en una nueva población independiente a fin de corroborar los hallazgos del GWAS. El

objetivo de este trabajo fue validar los SNP identificados en estudios de GWAS previos determinando la asociación entre estos SNP y las variables de interés en una población independiente de la población de mapeo y representativa del germoplasma avanzado actual del PMGA.

### MATERIALES Y MÉTODOS

Selección de SNPs a validar. Se seleccionaron los SNPs más significativos (con mayor  $-\log_{10}(P)$  y menos datos perdidos) de los QTL con mayor proporción de varianza fenotípica explicada (PVE) encontrados en los análisis de GWAS realizados en el marco del PMAA mencionados en la introducción.

Población de validación. El PMGA seleccionó dos poblaciones (500 *índicas* y 500 *japónica tropical*) independientes de la población utilizada para el mapeo asociativo. Las líneas de tipo *índica* no presentaron la variabilidad genotípica necesaria para realizar la validación. Se trabajó por lo tanto con las 500 líneas de tipo *japónica tropical*.

Genotipado de SNPs. Se sintetizaron cebadores KASP (LGC Genomics) para cada SNP, utilizándolos para genotipar las líneas en un equipo QuantStudio (Applied Biosystems). El llamado de alelos se realizó con el software QuantStudio Design & Analysis (Applied Biosystems).

1 D.Sc., INIA. Programa Nacional de Investigación en Producción de Arroz, jrosas@inia.org.uy.

2 Br., Estudiante Lic. Cs. Biológicas, Fac. de Ciencias UdelaR

3 Br., INIA Unidad de Biotecnología

4 Ing.Agr., Estudiante Ms. Ciencias Agrarias, Fac. de Agronomía UdelaR / becaria INIA

5 D.Sc. INIA Unidad de Biotecnología,

6 Ph.D., INIA. Programa Nacional de Investigación en Producción de Arroz,

7 M.Sc., INIA. Programa Nacional de Investigación en Producción de Arroz, hasta junio 2018

8 Ph.D., Programa Nacional de Investigación en Producción de Arroz

Datos fenotípicos y estimación de medias fenotípicas. Se consolidaron para su análisis conjunto los datos de ensayos de campo de los años 2003, 2004, 2006 y de 2013 a 2017 de los materiales de tipo *japónica* tropical con datos para las mismas variables mapeadas (FT, PH, GC, PHR, YAM, LB), agregándose además contenido de amilosa (AC). Se predijeron los efectos genotípicos de las líneas (BLUPs) para cada variable fenotípica con el modelo de la ecuación 1:

Ecuación [1]

$$Y_{ijkl} = \mu + L_i + B_{j(k)} + E_k + Z_l + e_{ijkl}$$

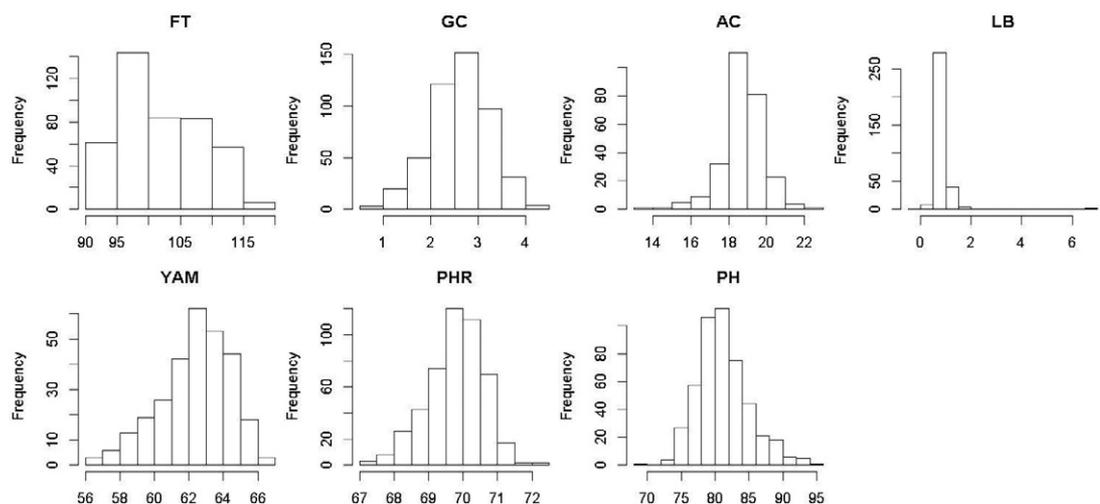
donde  $Y_{ijkl}$  es la observación de la variable de respuesta para el cultivar  $i$  en el bloque  $j$ , ensayo  $k$  y zafra  $l$ ;  $\mu$  es el efecto fijo de la media general;  $C_i$  es el efecto aleatorio del cultivar  $i$ ;  $B_{j(k)}$  es el efecto aleatorio del bloque  $j$  anidado en el ensayo  $k$ ;  $E_k$  es el efecto aleatorio del ensayo  $k$ ;  $Z_l$  es el efecto aleatorio de la zafra  $l$ , y  $e_{ijkl}$  es el efecto residual para el cultivar  $i$ , bloque  $j$ , ensayo  $k$ , y zafra  $l$ . Todos los efectos aleatorios se asumieron

independientes e idénticamente distribuidos. Los modelos se ajustaron con la función lmer del paquete lme4 en R.

Análisis de asociación SNP – fenotipo. Se ajustó un modelo de regresión lineal con los BLUPs de las líneas genotipadas como variable de respuesta y los valores genotípicos de cada SNP como variable regresora. Las regresiones se ajustaron con la función lm en R base. Se obtuvo el  $R^2$  de cada regresión como estimador de la proporción de la varianza fenotípica explicada (PVE) por cada SNP, así como el  $-\log_{10}(P)$  del coeficiente de regresión como estimador de la asociación SNP-fenotipo.

## RESULTADOS

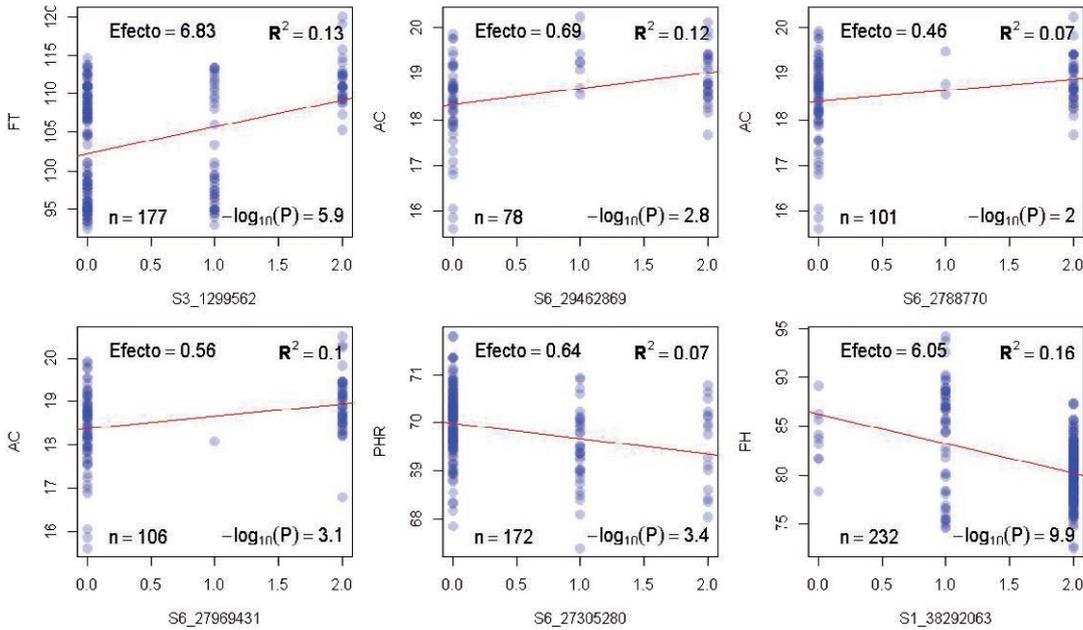
Las distribuciones de las variables fenotípicas para la población de validación se presentan en la figura 1. A excepción de resistencia a Brusone en hoja (LB), todas mostraron una buena adecuada dispersión para la validación.



**Figura 1.** Distribución de valores fenotípicos para las variables a validar. FT: tiempo de floración, GC: yesado, AC: amilosa, LB: brusone en hoja, YAM: entero, PHR: blanco total, PH: altura.

Las regresiones de SNPs seleccionados (sólo se muestran los que tuvieron un  $R^2$  mayor a 0,05) se muestran en la figura 2, detallando el efecto  $\beta$  estimado del SNP, el

$R^2$  de la regresión, el número de individuos con datos genotípicos y fenotípicos usado para ajustar el modelo, y el  $-\log_{10}(P)$  del  $\beta$  estimado.



**Figura 2.** Regresión de las variables fenotípicas de interés (FT= tiempo de floración, AC= contenido de amilosa, PHR= % de entero, PH= altura de planta) en los estados alélicos de los SNP (0= alelo con mayor frecuencia, 1= alelo con menor frecuencia) para las cuales se encontró un PVE mayor a 5%. Para cada SNP se muestra el efecto, el  $R^2$  de la regresión, el número de individuos con datos y el  $-\log_{10}$  del P-valor.

En general los SNPs tuvieron baja PVE. Sólo 4 SNP tuvieron PVE mayor a 10 ( $R^2$  mayor a 0,1), para tres variables (FT, AC, y PH). Los 10 SNPs con PVE mayor a 5% se señalan en el cuadro 1. Los efectos de estos SNPs son de buena magnitud, sobre todo considerando que fueron estimados a partir de

BLUPs, lo cual genera un encogimiento en la magnitud de los efectos estimados. Por ejemplo, el SNP S3\_1299562 tiene un efecto estimado de una semana en el tiempo de floración, y el S1\_38292063 de 6 cm en la altura de planta.

**Cuadro 1.** Parámetros de validación del SNP seleccionados. Se marcan en negrita los que superan el 5% de PVE, y subrayados los que superan el 10% y tienen un efecto de magnitud relevante desde el punto de vista práctico para el programa de mejoramiento.

Variable	SNP	$-\log_{10}(P)$	PVE	Efecto	n
<b>FT</b>	<b><u>S3_1299562</u></b>	<b><u>5,91</u></b>	<b><u>12,60</u></b>	<b><u>6,83</u></b>	<b><u>177</u></b>
<b>FT</b>	<b><u>S3_10096438</u></b>	<b><u>2,70</u></b>	<b><u>6,00</u></b>	<b><u>4,43</u></b>	<b><u>158</u></b>
GC	S6_2140954	0,24	0,20	0,15	189
GC	S6_2145195	0,06	0,00	0,06	180
AC	S6_2140954	0,02	0,00	0,02	100
AC	S6_2145195	1,23	3,50	1,14	102
AC	S6_26898904	0,26	0,40	0,13	102
<b>AC</b>	<b><u>S6_27365083</u></b>	<b><u>2,31</u></b>	<b><u>7,50</u></b>	<b><u>0,55</u></b>	<b><u>104</u></b>
<b>AC</b>	<b><u>S6_29462869</u></b>	<b><u>2,76</u></b>	<b><u>12,20</u></b>	<b><u>0,69</u></b>	<b><u>78</u></b>
<b>AC</b>	<b><u>S6_29480471</u></b>	<b><u>2,10</u></b>	<b><u>7,00</u></b>	<b><u>0,48</u></b>	<b><u>100</u></b>
<b>AC</b>	<b><u>S6_2788770</u></b>	<b><u>2,04</u></b>	<b><u>6,70</u></b>	<b><u>0,46</u></b>	<b><u>101</u></b>
<b>AC</b>	<b><u>S6_27969431</u></b>	<b><u>3,07</u></b>	<b><u>10,20</u></b>	<b><u>0,56</u></b>	<b><u>106</u></b>
LB	S9_9786203	0,11	0,10	0,02	75
YAM	S6_26898904	0,38	0,70	0,43	102
<b>YAM</b>	<b><u>S6_27365083</u></b>	<b><u>2,16</u></b>	<b><u>6,90</u></b>	<b><u>1,17</u></b>	<b><u>104</u></b>
YAM	S6_29462869	0,06	0,00	0,09	78
YAM	S6_29480471	0,01	0,00	0,02	100
<b>YAM</b>	<b><u>S6_2788770</u></b>	<b><u>2,22</u></b>	<b><u>7,40</u></b>	<b><u>1,20</u></b>	<b><u>101</u></b>
YAM	S6_27969431	1,56	4,60	0,81	106
PHR	S6_26894513	2,17	3,80	0,36	194
<b>PHR</b>	<b><u>S6_27305280</u></b>	<b><u>3,38</u></b>	<b><u>7,10</u></b>	<b><u>0,64</u></b>	<b><u>172</u></b>
<b>PH</b>	<b><u>S1_38292063</u></b>	<b><u>9,87</u></b>	<b><u>16,40</u></b>	<b><u>6,05</u></b>	<b><u>232</u></b>

## CONCLUSIONES

Se identificaron dos SNP (S3\_1299562 y S1\_38292063) con altos PVE y efectos en FT y PH que pueden ser utilizados en un set para selección asistida en germoplasma *japónica* tropical de INIA. Si bien FT y PH son variables con alta heredabilidad y de fácil selección fenotípica, estos SNPs tendrían utilidad en un esquema de avance rápido generacional combinado con selección asistida. Es necesario validar los SNP para resistencia a Brusone en una población con variabilidad fenotípica y genotípica adecuada, así como definir una población de tipo *índica* adecuada para la de validación de los SNPs identificados por GWAS y otros SNP reportados en la literatura. La selección de poblaciones de validación se vería facilitada

al disponerse de una base de datos consolidada con información fenotípica, genotípica y de pedigrí del PMGA.

## BIBLIOGRAFÍA

- Quero, G., Gutierrez, L., Monteverde.** 2018. Genome-wide association study using historical breeding populations discovers genomic regions involved in high-quality rice. *Plant Genome*, 11, no. 170076. doi: 10.3835/plantgenome2017.08.0076
- Rosas, J.E., Martínez, S., Blanco.** 2017. Resistance to multiple temperate and tropical stem and sheath diseases of rice. *Plant Genome*, 10(1): 1-13. doi:10.3835/plantgenome2017.03.0029