

## AVANCES DE RESULTADOS DEL PROYECTO MAPEO ASOCIATIVO PARA LA IDENTIFICACIÓN DE MARCADORES ASOCIADOS A RENDIMIENTO, CALIDAD Y RESISTENCIA A ENFERMEDADES

V. Bonnacarrère<sup>1</sup>, S. Garaycochea<sup>2</sup>, L. Gutierrez<sup>3</sup>, J. Rosas<sup>4</sup>, N. Berberian<sup>5</sup>, S. Fernández<sup>6</sup>, S. Martínez<sup>7</sup>, F. Perez<sup>8</sup>, P. Blanco<sup>9</sup>

### 1. INTRODUCCIÓN

En el marco del nuevo plan estratégico de INIA, en el año 2011 se comenzó a ejecutar el proyecto “Mapeo asociativo para la identificación de marcadores asociados a rendimiento, calidad y resistencia a enfermedades en la población de mejoramiento de arroz de INIA” cuyo principal objetivo es identificar marcadores moleculares para contribuir a acelerar los procesos de obtención de variedades. En este proyecto se planteó la identificación de marcadores moleculares del tipo SNP (del inglés, Single Nucleotide Polymorphism) asociados a características del cultivo que son objetivo del programa de mejoramiento. En este proyecto se usó la estrategia de Mapeo Asociativo (MA) para la identificación de éstos marcadores. Esta estrategia requiere la identificación de una población de mapeo adecuada (líneas de arroz), el genotipado de la misma, la determinación de estructura de la población, el fenotipado de las líneas y la búsqueda de asociación entre genotipo y fenotipo.

La población de estudio de este proyecto está constituida por 665 líneas, 330 de la ssp. Japonica y 335 de Indica, en etapas finales de evaluación del programa de mejoramiento. Estas líneas fueron genotipadas usando una técnica de secuenciación de ADN conocida como GBS (del inglés Genotyping By Sequencing). (Elshire et al, 2011). Hasta el presente, este proyecto ha avanzado en el genotipado de las líneas seleccionadas mediante la identificación de SNPs y en el análisis estadístico de las mediciones fenotípicas determinadas por los mejoradores. La calidad de los datos fenotípicos es fundamental para el hallazgo de marcadores asociados, por lo que se ha puesto especial énfasis en el análisis estadístico de los mismos. En este reporte se presentan los resultados del análisis de datos fenotípicos sobre todas las líneas del programa de mejoramiento y no únicamente en los 665 genotipos analizados en el proyecto ya que el objetivo fue ajustar los mejores modelos de análisis a utilizar para cada una de las variables. A continuación se describen los materiales y métodos y los principales resultados obtenidos.

### 2. MATERIALES Y MÉTODOS

#### *Extracción de ADN, genotipado y análisis bioinformático*

A partir de plantas individuales de cada una de las líneas de arroz se extrajo ADN de alta calidad utilizando el kit de Qiagen (DNeasy Plant Extraction Kit). Las muestras de ADN fueron enviadas al Instituto para la Diversidad Genómica (Institute for Genomic Diversity) de la Universidad de Cornell en donde fueron genotipadas por GBS. La identificación de SNP fue implementada en Tassel. Los genomas de referencia utilizados fueron *O. sativa ssp japonica* cv Nipponbare y *O. sativa ssp indica* cv 9311

#### *Fenotipado de las líneas seleccionadas*

El germoplasma seleccionado fue en las zafra 2011 y 2012 para Rendimiento (Rend); Rendimiento sano, seco y limpio (SSL), % blanco total (BT\_porc); % de grano entero (Ent\_porc); % yesado total (Yes\_porcTot); %de yeso en el grano entero (Yes\_porcEnt); % de área yesada del grano (AY\_porc); % entero (EntOK); % mancha en grano (Mancha); Largo del grano (Largo); Ancho del grano (Ancho); Largo por ancho del grano (LxA); Blancura total (Blancura total) y blancura vitrea, (Blancura\_vitrea). Para el año 2012 se agregaron las siguientes variables: Ciclo vegetativo; altura (H); Alto de Hoja Bandera (HB); Peso (Peso); y parámetros enfermedades del tallo: Scl\_Max; Scl\_x; Rhiz\_max; y Rhiz\_x, P.

#### *Análisis estadístico de datos fenotípicos*

Para los genotipos Indica los datos fenotípicos del 2011 consistieron en 24 ensayos cada uno con un set de genotipos evaluados en un diseño de bloques completos al azar (BCA) con dos repeticiones. Se

<sup>1</sup> DSc INIA, Unidad de Biotecnología [vbonnacarrere@inia.org.uy](mailto:vbonnacarrere@inia.org.uy)

<sup>2</sup> MSc. INIA, Unidad de Biotecnología. [sgaraycochea@inia.org.uy](mailto:sgaraycochea@inia.org.uy)

<sup>3</sup> PhD Facultad de Agronomía, UdelaR, Departamento de Biometría, Estadística y Cómputos

<sup>4</sup> MSc INIA, Programa Arroz [jrosas@inia.org.uy](mailto:jrosas@inia.org.uy)

<sup>5</sup> Est. Facultad de Agronomía, UdelaR. Departamento de Biometría, Estadística y Cómputos

<sup>6</sup> Lic. INIA, Unidad de Tecnologías de la Información [sfernandez@inia.org.uy](mailto:sfernandez@inia.org.uy)

<sup>7</sup> Ing. Agr. INIA, Programa Arroz [smartinez@inia.org.uy](mailto:smartinez@inia.org.uy)

<sup>8</sup> PhD INIA, Programa Arroz [fperez@inia.org.uy](mailto:fperez@inia.org.uy)

<sup>9</sup> MSc. INIA, Programa de Arroz. [pblanco@inia.org.uy](mailto:pblanco@inia.org.uy)

usaron ocho testigos comunes y se evaluaron 1920 líneas. Los datos fenotípicos del 2012 consistieron en 12 ensayos cada uno con un set de genotipos evaluados en un diseño de BCA con tres repeticiones. Se realizó un análisis descriptivo de cada una de las variables a estudiar y se mapearon las diferentes variables en el campo para identificar patrones de heterogeneidad espacial.

Para esto se construyeron mapas de rendimiento para cada uno de los testigos así como se utilizaron herramientas geo-estadísticas para obtener la interpolación de valores de rendimiento a partir de los residuos de un modelo de análisis conjunto de los ensayos para los testigos repetidos. Cada una de las variables fue analizada mediante tres modelos que consideraron la heterogeneidad espacial de diferente forma. Indicadores de ajuste de modelos (i.e. AIC, BIC, LRT) fueron utilizados para seleccionar el mejor modelo para cada variable. El primer modelo (M1) analiza cada uno de los ensayos con su diseño experimental por separado: [1]  $y_{ij} = \mu + G_i + \beta_j + \epsilon_{ij}$  donde  $y_{ij}$  es la variable de respuesta (cada una de las variables estudiadas),  $\mu$  es la media general,  $G_i$  es el efecto del  $i$ -ésimo genotipo (línea),  $\beta_j$  es el efecto del  $j$ -ésimo bloque y  $\epsilon_{ij}$  es el residual del modelo asociado a cada unidad experimental. El segundo modelo (M2) considera todos los ensayos simultáneamente modelándolos como un efecto aleatorio: [2]  $y_{ijk} = \mu + G_i + E_j + \beta_{k(j)} + \epsilon_{ijk}$  donde  $y_{ijk}$ ,  $\mu$  y  $\epsilon_{ijk}$  son definidos de igual forma que en [1],  $E_j$  es una variable aleatoria asociada a la  $j$ -ésimo ensayo ( $E_j \sim N(0, \sigma_E^2)$ ) y  $\beta_{k(j)}$  es el efecto del  $j$ -ésimo bloque dentro del  $k$ -ésimo ensayo. El tercer modelo (M3) considera todos los ensayos simultáneamente a la vez que una corrección espacial por ubicación en el campo (i.e. columna e hilera): [3]  $y_{ijklm} = \mu + G_i + E_j + \beta_{k(j)} + C_{l(j)} + H_{m(j)} + \epsilon_{ijklm}$  donde  $y_{ijklm}$ ,  $\mu$ ,  $E_j$ ,  $\beta_{k(j)}$  y  $\epsilon_{ijklm}$  son definidos de igual forma que en [1] y [2],  $C_{l(j)}$  es una variable aleatoria asociada a la  $l$ -ésima columna dentro de ensayo ( $C_{l(j)} \sim N(0, \sigma_C^2)$ ),  $H_{m(j)}$  es una variable aleatoria asociada a la  $m$ -ésima hilera dentro de ensayo ( $H_{m(j)} \sim N(0, \sigma_H^2)$ ). Para Japonica se utilizaron los datos fenotípicos de los ensayos con líneas avanzadas evaluados durante los años 2010, 2011 y 2012. El diseño para evaluación de los genotipos fue de BCA, con dos o tres repeticiones, en parcelas de 1,2 x 3,5 m con seis hileras. Se evaluaron 1006 líneas en 2010, 1627 líneas en 2011 y 807 en 2012, utilizando cuatro testigos comunes a todos los ensayos. Se compararon los modelos M1, M2 y M3 descriptos respectivamente en las ecuaciones [1], [2] y [3]. Se estimó para cada variable y cada año la heredabilidad en el sentido amplio, o repetitividad, de acuerdo a la siguiente expresión: [4]  $H^2 = \frac{\sigma_G^2}{\sigma_P^2}$

donde  $H^2$  es la repetitividad,  $\sigma_G^2$  es la varianza genotípica y  $\sigma_P^2$  la varianza fenotípica

### 3. RESULTADOS

Del genotipado por GBS se identificaron un total de de 107.752 SNP con una frecuencia promedio de 0.28 SNP/Kb cuando se utilizó como genoma de referencia el de la subespecie japónica. El conjunto de SNPs identificados contra el genoma de Japonica fue el elegido para continuar con los estudios de asociación por la calidad del genoma de referencia.

En el análisis estadístico de los fenotipos de la población de mapeo de la ssp. Indica se detectó la presencia de heterogeneidad espacial mostrada por el rendimiento diferencial de los testigos en los diferentes ensayos (Figura 1.A) y por el mapa de rendimiento obtenido a partir de la interpolación de los residuales de rendimiento de un modelo de análisis conjunto de los ensayos con los testigos (Figura 1.B).

Debido a la heterogeneidad espacial presente en el campo y a las dimensiones del experimento, en el año 2011 la corrección espacial fue muy importante, siendo el modelo M3 el mejor para la mayoría de las variables (Cuadro 1). En el año 2012, el modelo con corrección espacial fue mejor únicamente en la mitad de las variables. Las variables que se beneficiaron por la corrección espacial no presentan características en común entre sí. Cuando se compara el ranking de las líneas al utilizar diferentes modelos estadístico para estimar las medias; para el caso de la variable rendimiento, de las mejores 15 líneas identificadas con el M3, 12 son recuperadas con M2, pero únicamente 9 son recuperadas como mejores líneas cuando se analizan con M1. Esto indica que la búsqueda de modelos apropiados es muy importante para identificar las mejores líneas y que el análisis por ensayo (M1) no es un buen modelo.

Al igual que para los genotipos Indica, para los Japonica el modelo con mejor ajuste para la mayoría de las variables fue el M3 especialmente para el año 2011, en el que éste modelo fue el mejor para la totalidad de las variables (Cuadro 2). Asimismo, en ese año se observan menores repetitividad para la mayoría de las variables (Cuadro 5), lo que refuerza la necesidad de corregir por variabilidad espacial para obtener buenos estimadores de las medias fenotípicas. Al igual que en los genotipos Indica,

existe un cambio en el ranking de las líneas al utilizar diferentes modelos estadístico para estimar las media. Para el caso de la variable rendimiento, de las mejores 15 líneas identificadas con M3, 11 son recuperadas con el M2 y 10 son recuperadas como mejores líneas cuando se usa M1.

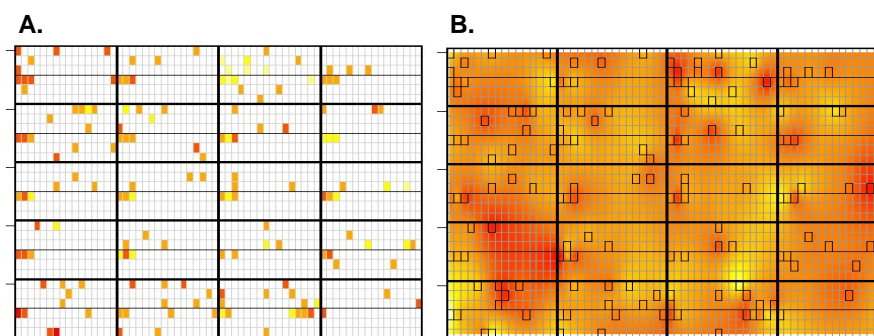


Figura 1. **A.** Rendimiento de los testigos repetidos en el año 2011 para 20 de los 24 ensayos establecidos en el campo. **B.** Mapa de rendimiento obtenido a partir de los ensayos para 20 de los 24 ensayos del año 2011

Cuadro 1. Listado del mejor modelo para cada una de las variables evaluadas en líneas tipo Indica en los años 2011 y 2012

Variable	2011	2012	Variable	2011	2012	Variable	2011	2012
Ancho	M2	M2	H	-	M2	rend	M3	M3
AY	M3	M3	HB	-	M3	Rhizmax	-	M2
BrancTot	M3	M2	Largo	M2	M2	Rhizprom	-	M2
BrancVit	M3	M3	LxA	M2	M2	Scimax	-	M2
BT	M3	M2	Mancha	M3	M2	Sciprom	-	M2
Ent	M3	M3	Peso	-	M3	SSL	M3	M3
EntOK	M3	M3	Pyri	-	M3	YesEnt	M3	M2
						YesTot	M3	M3

Cuadro 2. Mejores modelos para las variables comparadas y repetibilidad de las variables evaluadas en líneas tipo Japonica en los años 2010, 2011 y 2012

Variable	Comparación de modelos			Repetitividad			Variable	Comparación de modelos			Repetitividad		
	2010	2011	2012	2010	2011	2012		2010	2011	2012	2010	2011	2012
Rend	M3	M3	M3	0,52	0,18	0,39	Ent	M3	M3	M3	0,70	0,51	0,72
H	M2	M3	M3	0,60	0,50	0,49	Yeso	M2	M3	M2	0,21	0,34	0,23
Flor	M3	M3	M3	0,73	0,55	0,11	YRC	M3	M3	M3	0,26	0,49	0,70
SCL	M3	M3	M3	0,40	0,24	0,32	Manch	M2	M3	M3	0,06	0,00	0,35
ROS	M3	M3	M2	0,31	0,08	0,03	MRC	M2	M3	M3	0,07	0,00	0,36
BT	M2	M3	M2	0,54	0,51	0,60							

### 3. CONCLUSIONES

El avance del proyecto ha sido muy exitoso ya que se ha logrado obtener información genotípica de alta calidad y cantidad para todos los genotipos evaluados y se han ajustado modelos estadísticos para el análisis de los fenotipos que serán utilizados en el mapeo por asociación. Por otro lado es de gran relevancia el hecho de haber establecido un grupo multidisciplinario e interinstitucional ya que la ejecución de este tipo de proyectos requiere masa crítica con experiencia en diversas áreas del conocimiento.

### 4. BIBLIOGRAFÍA

ELSHIRE, R.; GLAUBITZ, J.; SUN, Q.; POLAND, J.; KAWAMOTO, K.; BUCKLER, E.; MITCHELL, S. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. Plos One v. 6, no.5, e 19379.