



FACULTAD DE
AGRONOMIA
UNIVERSIDAD DE LA REPUBLICA



UNIVERSIDAD
DE LA REPÚBLICA

EJEMPLOS DEL USO DE LOS MODELOS MIXTOS EN MEJORAMIENTO GENÉTICO VEGETAL

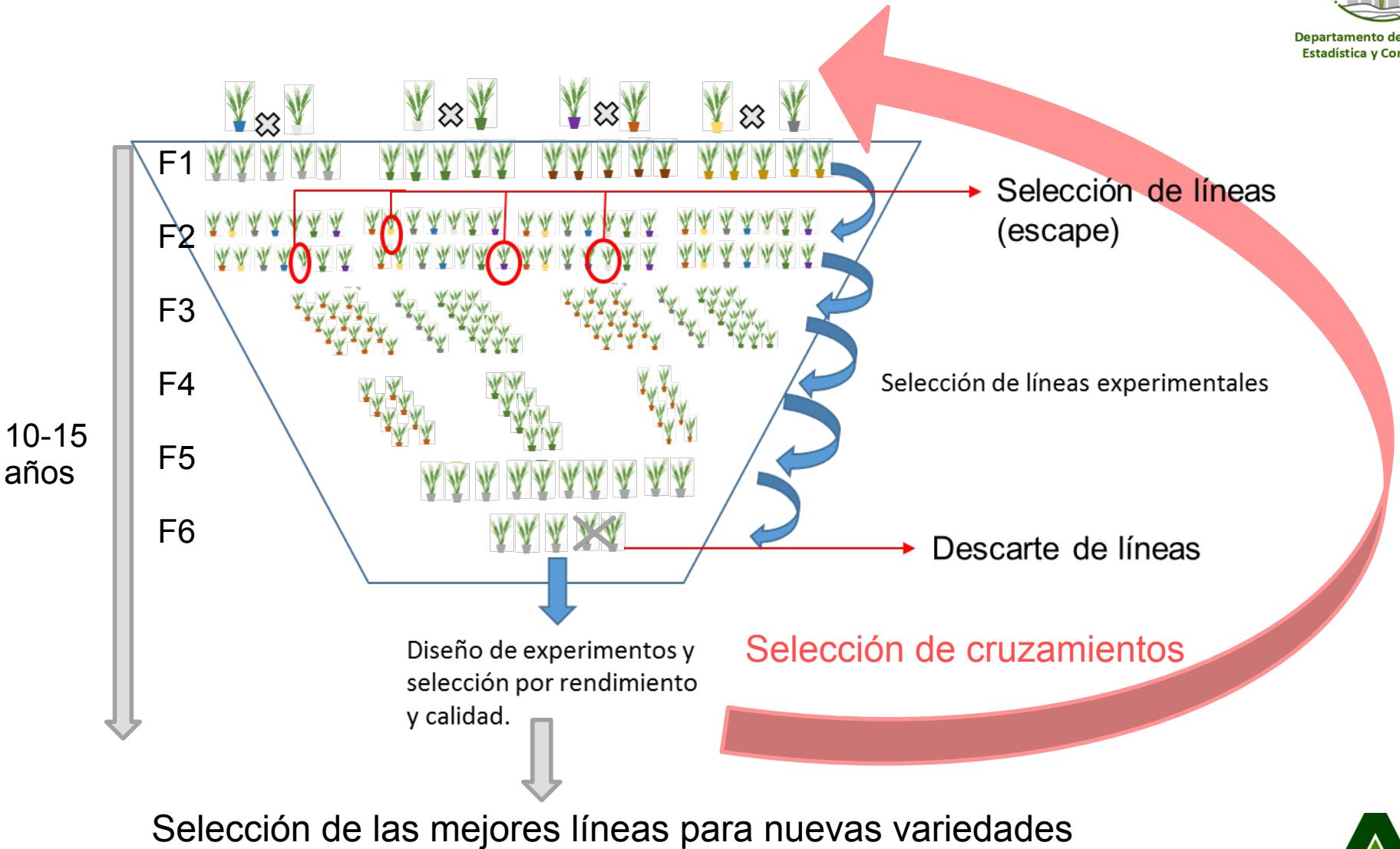
Bettina Lado, Martín Quincke y Lucía Gutiérrez

Departamento de Biometría, Estadística y Computación, Facultad de Agronomía, Universidad de la República



Departamento de Biometría,
Estadística y Computación

Programa de mejoramiento vegetal



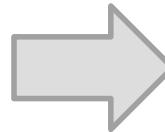
Predicciones Genómicas - del Fenotipo



Departamento de Biometría,
Estadística y Computación



MODELO
ESTADÍSTICO



Predicciones de
variables
fenotípicas

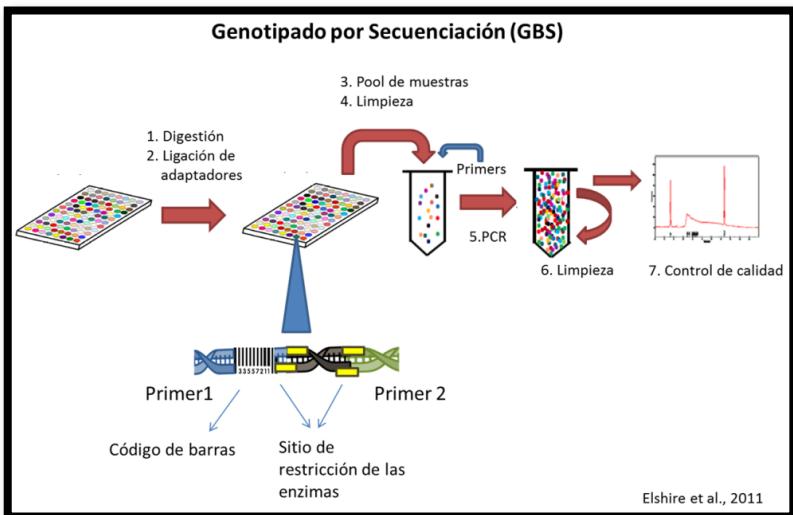
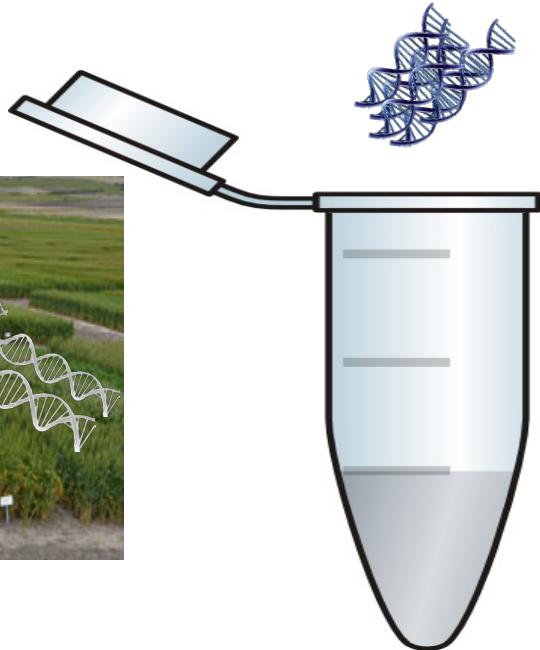


FACULTAD DE
AGRONOMIA
UNIVERSIDAD DE LA REPÚBLICA



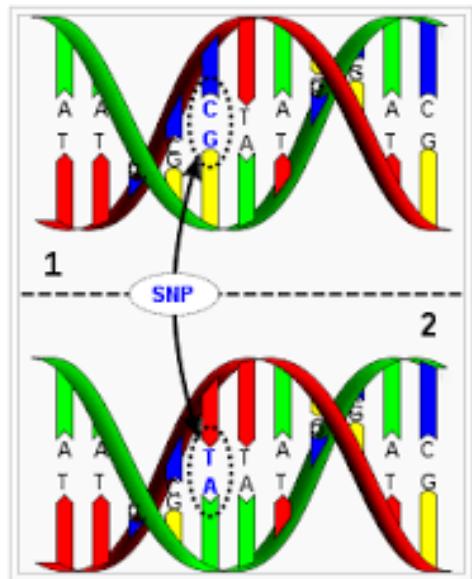
Población de entrenamiento

Rendimiento
Calidad
Enfermedades





Problema de $p>>n$: más marcadores que individuos



MATRIZ

	SNP1	SNP2	...					
1	A	G	NA	T	C	C	NA	...
2	A	C	NA	T	C	C	A	...
3	NA	NA	T	T	C	C	T	...
4	A	C	T	T	NA	C	A	...
5	A	C	T	NA	C	C	A	...
6	A	NA	T	NA	C	NA	NA	...
7	A	G	T	T	C	C	T	...
...

MATRIZ $M_{n \times p}$

	-1	1	NA	-1	-1	-1	NA	...
1	-1	-1	NA	-1	-1	-1	-1	...
2	-1	-1	NA	-1	-1	-1	-1	...
3	NA	NA	-1	-1	-1	-1	1	...
4	-1	-1	-1	-1	NA	-1	-1	...
5	-1	-1	-1	NA	-1	-1	-1	...
6	-1	NA	-1	NA	-1	NA	NA	...
7	-1	1	-1	-1	-1	-1	1	...
...



Modelo Ridge Regression (modelos de penalización):

$$y = X\beta_{RR} + \varepsilon_{RR}$$

Matriz de incidencia con
información de marcadores
moleculares

$$E(e) = 0$$

$$\text{Var}(e) = R = I\sigma_e^2$$

$$\hat{\beta}_{argmin} = \left\{ \sum_i \left(y_i - \sum_j x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^p \|\beta_j\|^2 \right\}$$

Se penalizan las estimaciones de los coeficientes de regresión mediante la inclusión de un parámetro λ que produce una contracción de los coeficientes entorno a cero (Hoerl y Kennard 1970). Así, se reduce la co-linealidad (redundancia) entre los predictores (marcadores)



Modelos Mixtos en el contexto de mejoramiento vegetal

Modelo Ridge Regression (modelos de penalización):

$$E(e) = 0$$

$$\text{Var}(e) = R = I\sigma_e^2$$

M(nxp):Matriz
de marcadores
moleculares

$$y = X\beta_{blup} + Z\mathbf{M}\boldsymbol{\mu}_{blup} + \boldsymbol{\varepsilon}_{blup}$$

$$E(y) = X\beta$$

$$\text{Var}(y) = V$$

$$E(u) = 0$$

$$\text{Var}(u) = I\sigma_u^2$$

La flexibilidad de los modelos mixtos de modelar las relaciones dentro de grupos de datos, utilizando matrices de varianza-covarianzas, es fundamental en su utilización en el mejoramiento vegetal.



Estimaciones de los efectos del modelo

$$\hat{u}_{blup} = GM'Z'V^{-1} \left(y - X\hat{\beta}_{blup} \right)$$

$$\hat{\beta}_{blup} = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

$$V = ZMGM'Z' + R$$

Ecuaciones de Henderson

$$X\beta_{blup} + ZMu_{blup} + \varepsilon_{blup}$$

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}ZM \\ M'ZR^{-1}X & M'Z'R^{-1}ZM + G^{-1} \end{pmatrix} \begin{pmatrix} \beta_{blup} \\ u_{blup} \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ M'Z'R^{-1}y \end{pmatrix}$$

$$\begin{pmatrix} \beta_{blup} \\ u_{blup} \end{pmatrix} = \begin{pmatrix} X'X & X'ZM \\ M'Z'X & M'Z'ZM + \lambda I \end{pmatrix}^{-1} * \begin{pmatrix} X'y \\ M'Z'y \end{pmatrix}$$



$\lambda = \sigma_e^2 \sigma_u^{-2}$ parámetro de penalización



Modelos Mixtos en el contexto de mejoramiento vegetal

Modelo GBLUP:

$$E(e) = 0$$

$$Mu_{blup}$$

$$Var(e) = R = I\sigma_e^2$$

$$y = X\beta_{blup} + Z\overline{g_{blup}} + \varepsilon_{blup}$$

$$E(y) = X\beta$$

$$Var(y) = V$$

$$E(g) = 0$$

$$Var(g) = G = K\sigma_g^2 = MM'\sigma_u^2$$



Modelos Mixtos en el contexto de mejoramiento vegetal



Departamento de Biometría,
Estadística y Computación

Modelo GBLUP:

$$E(e) = 0$$

$$Mu_{blup}$$

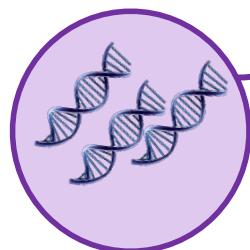
$$Var(e) = R = I\sigma_e^2$$

$$y = X\beta_{blup} + Zg_{blup} + \varepsilon_{blup}$$



$$E(u) = 0$$

$$Var(g) = G = K\sigma_g^2$$

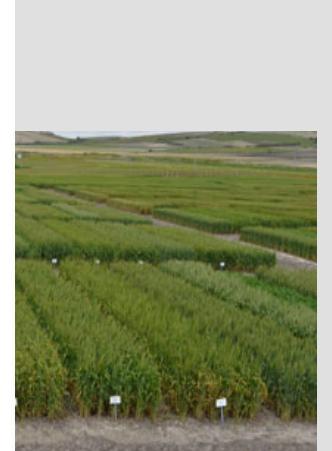
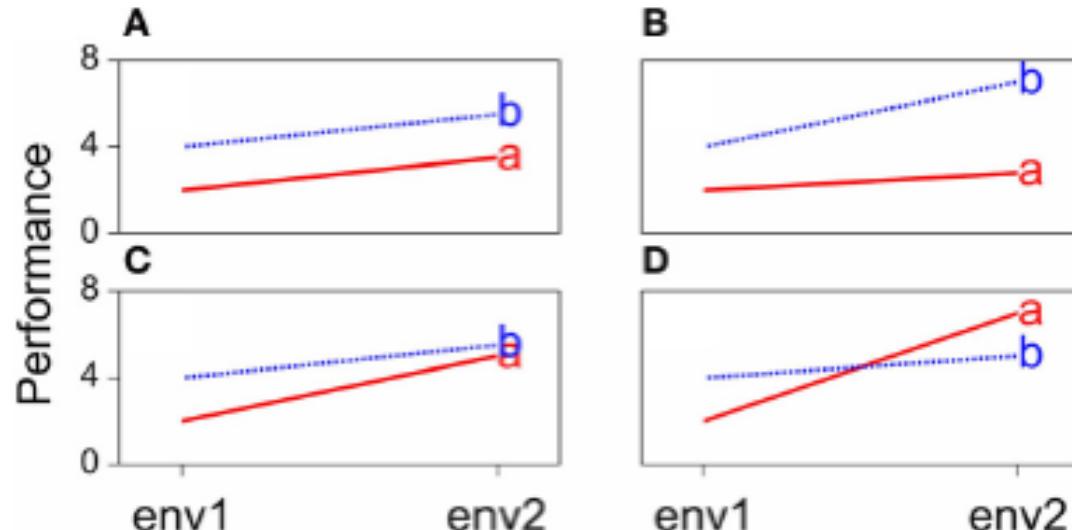


1	2	3	4	5	6	7	...	
1	1	0.85	0.55	0.90	0.45	0.35	0.66	...
2	0.85	1	0.50	0.82	0.36	0.55	0.25	...
3	0.55	0.50	1	0.69	0.75	0.78	0.35	...
4	0.90	0.82	0.69	1	0.58	0.59	0.25	...
5	0.45	0.36	0.75	0.58	1	0.65	0.15	...
6	0.35	0.55	0.78	0.59	0.65	1	0.40	...
7	0.66	0.25	0.35	0.25	0.15	0.40	1	...
...



FACULTAD DE
AGRONOMIA
UNIVERSIDAD DE LA REPUBLICA

Población de entrenamiento



Modelo multi-ambiente (GBLUP_{gxe})

$$y = \mu + Zg + \epsilon$$

g: vector de valor genético del individuos

$g \sim N(0, \rho \otimes K\sigma_{ge}^2)$ matriz de correlaciones entre ambientes

Múltiples caracteres



Departamento de Biometría,
Estadística y Computación

Población de entrenamiento



Datos fenotípicos de múltiples
caracteres

Modelo multi-caracter (GBLUPm)

$$y = \mu + Zg + \epsilon$$

g: vector de valor genético del individuos

$g \sim N(0, \Sigma \otimes K\sigma_{ge}^2)$ matriz de correlaciones entre caracteres



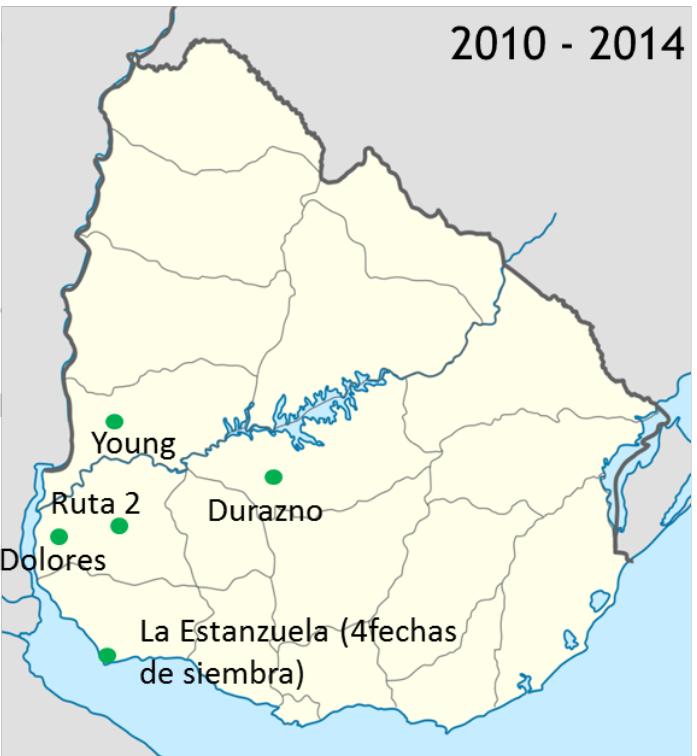
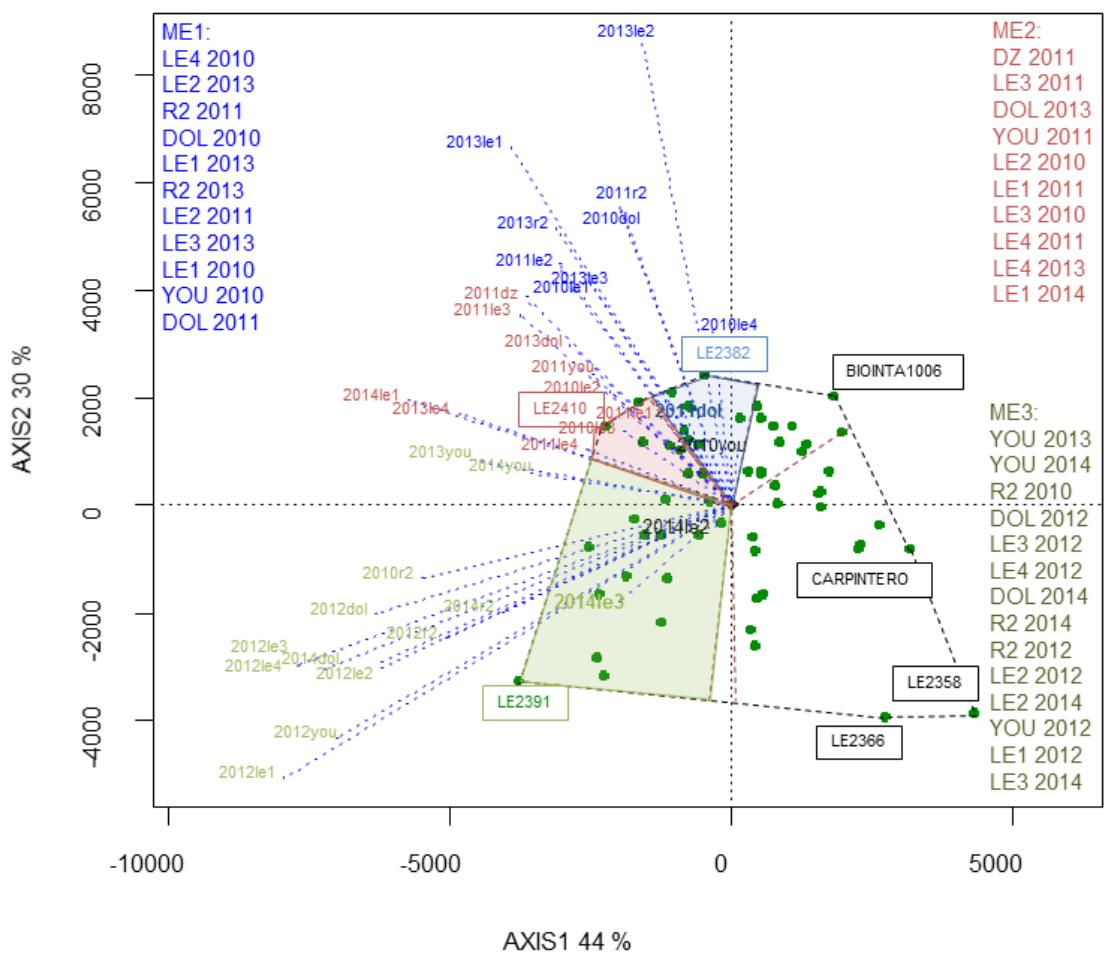
FACULTAD DE
AGRONOMIA
UNIVERSIDAD DE LA REPUBLICA

Objetivo General: Optimizar herramientas para la aplicación de selección genómica en distintos estadios de un programa de mejoramiento genético de trigo para la predicción de caracteres múltiples, el manejo de la interacción genotipo por ambiente y la selección de los mejores cruzamientos

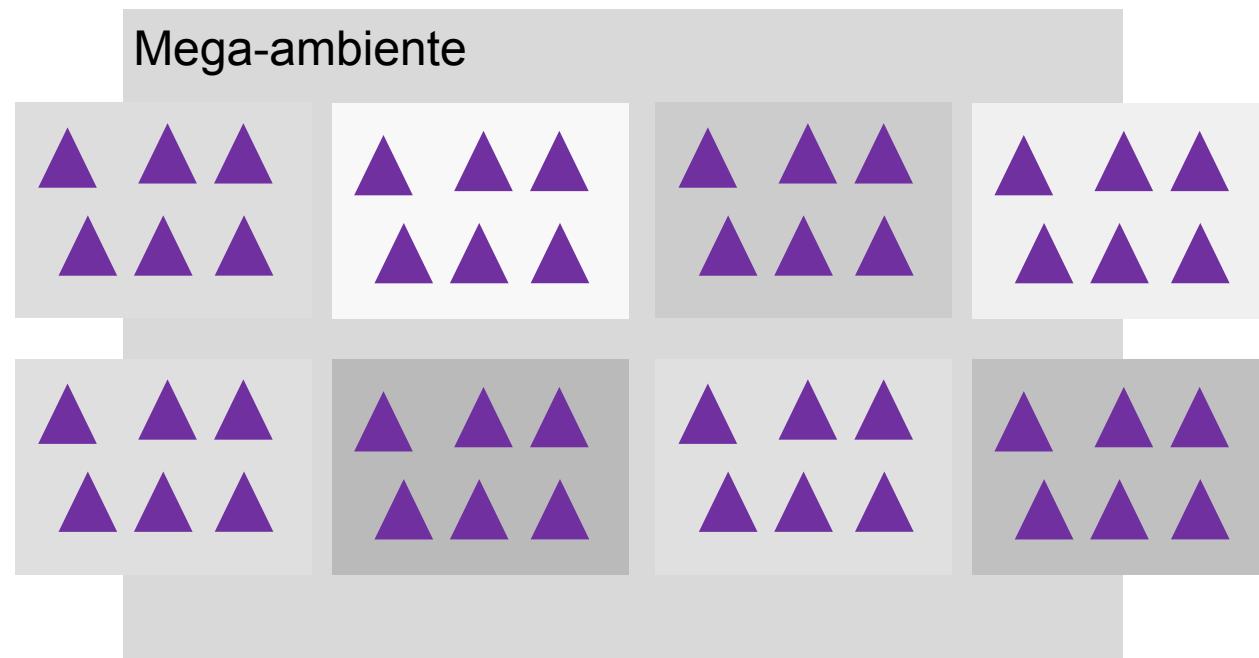


OBJETIVOS

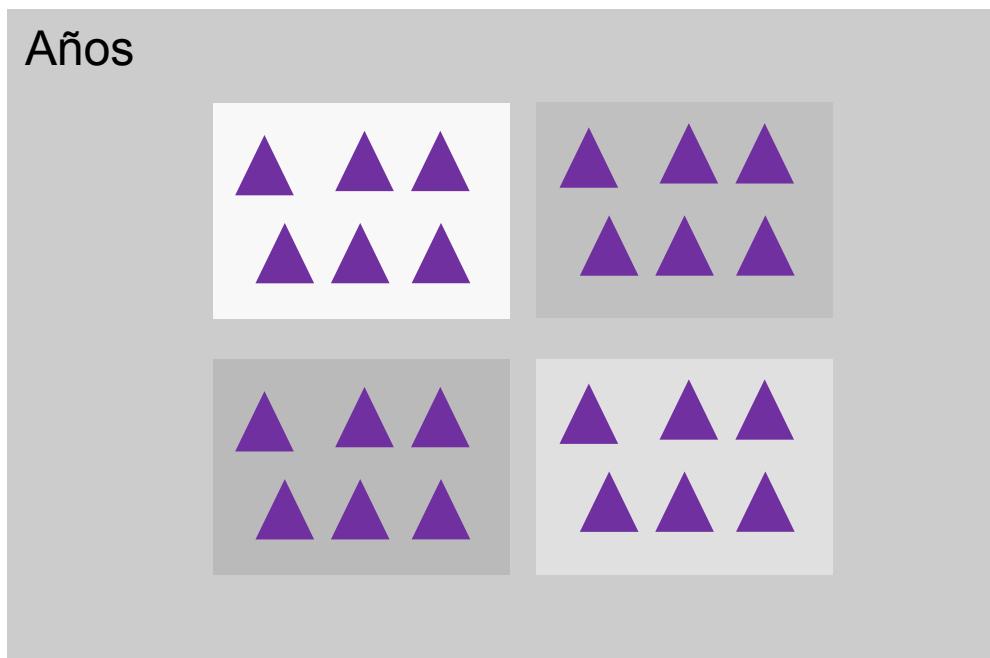
1) Mejorar la capacidad predictiva del valor genético de los individuos modelando la IGA en la predicción de nuevos individuos



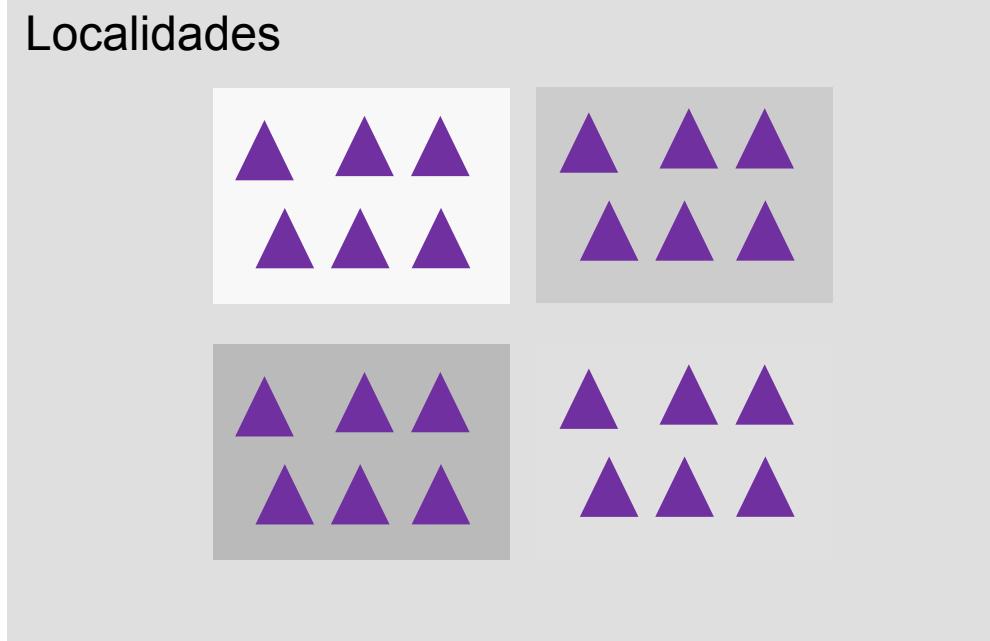
¿Es posible diseñar grupos de ambiente en los cuales la predicción de medias sea mejor que la predicción específica por ambiente?



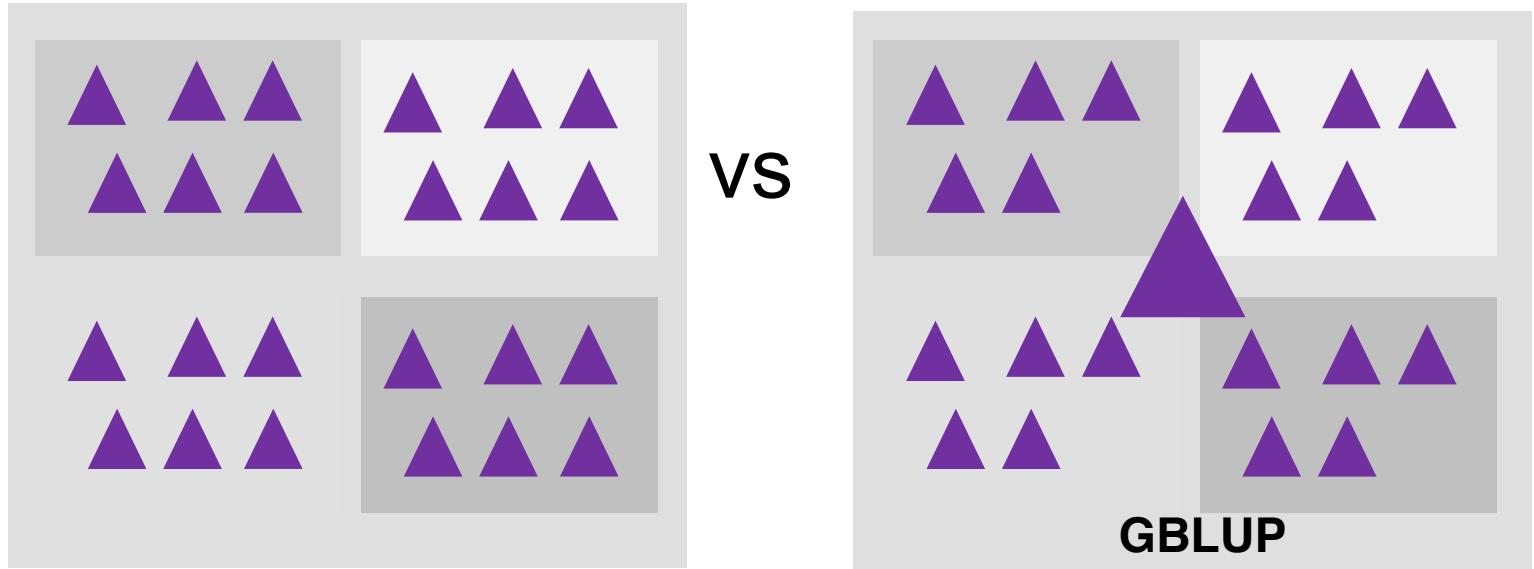
¿Es posible diseñar grupos de ambiente en los cuales la predicción de medias se mejor que la predicción específica por ambiente?



¿Es posible diseñar grupos de ambiente en los cuales la predicción de medias se mejor que la predicción específica por ambiente?



¿Es posible diseñar grupos de ambiente en los cuales la predicción de medias se mejor que la predicción específica por ambiente?



GBLUP_{gxe}

$$y_{ge} = Zg_{ge} + \varepsilon_{ge}$$

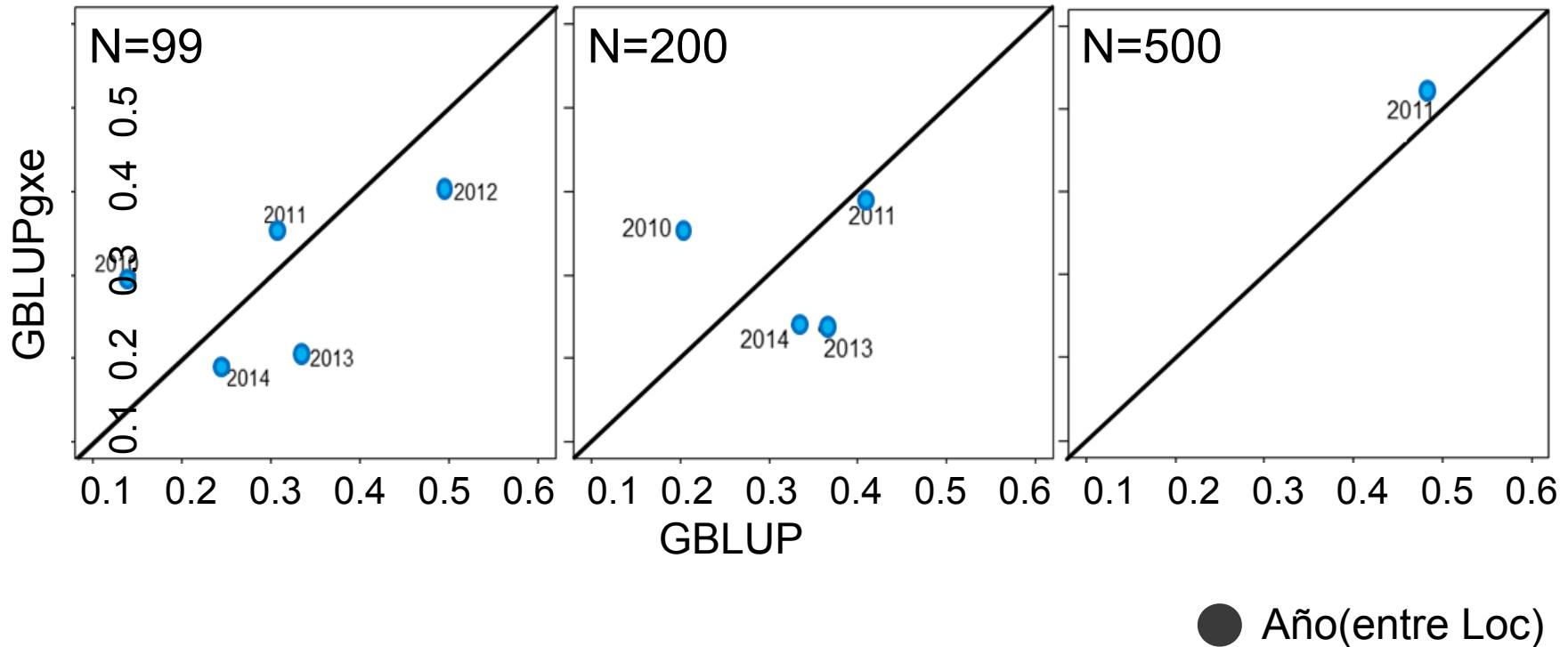
$$g_{ge} \sim N(0, (\rho \otimes K)\sigma_{ge}^2)$$

$$y = Zg + \varepsilon$$

$$g \sim N(0, K\sigma_g^2)$$

Comparación de modelos usando validación cruzada 7-fold

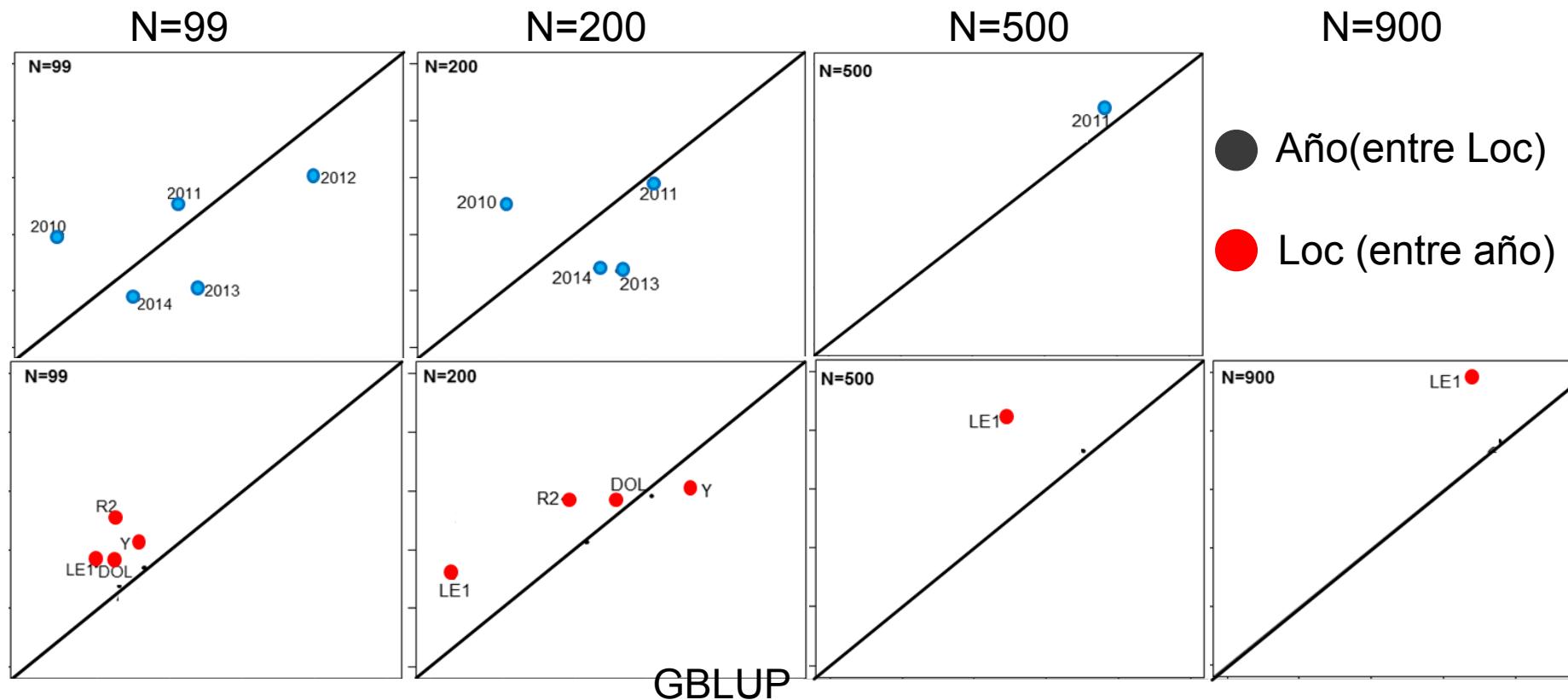
Comparación de modelos GBLUP vs GBLUPgxe



La IGA entre localidad en un mismo año es baja

Las predicciones dentro la mayoría de los años es mas precisa con el modelo GBLUP

Comparación de modelos multi-ambiente en SG



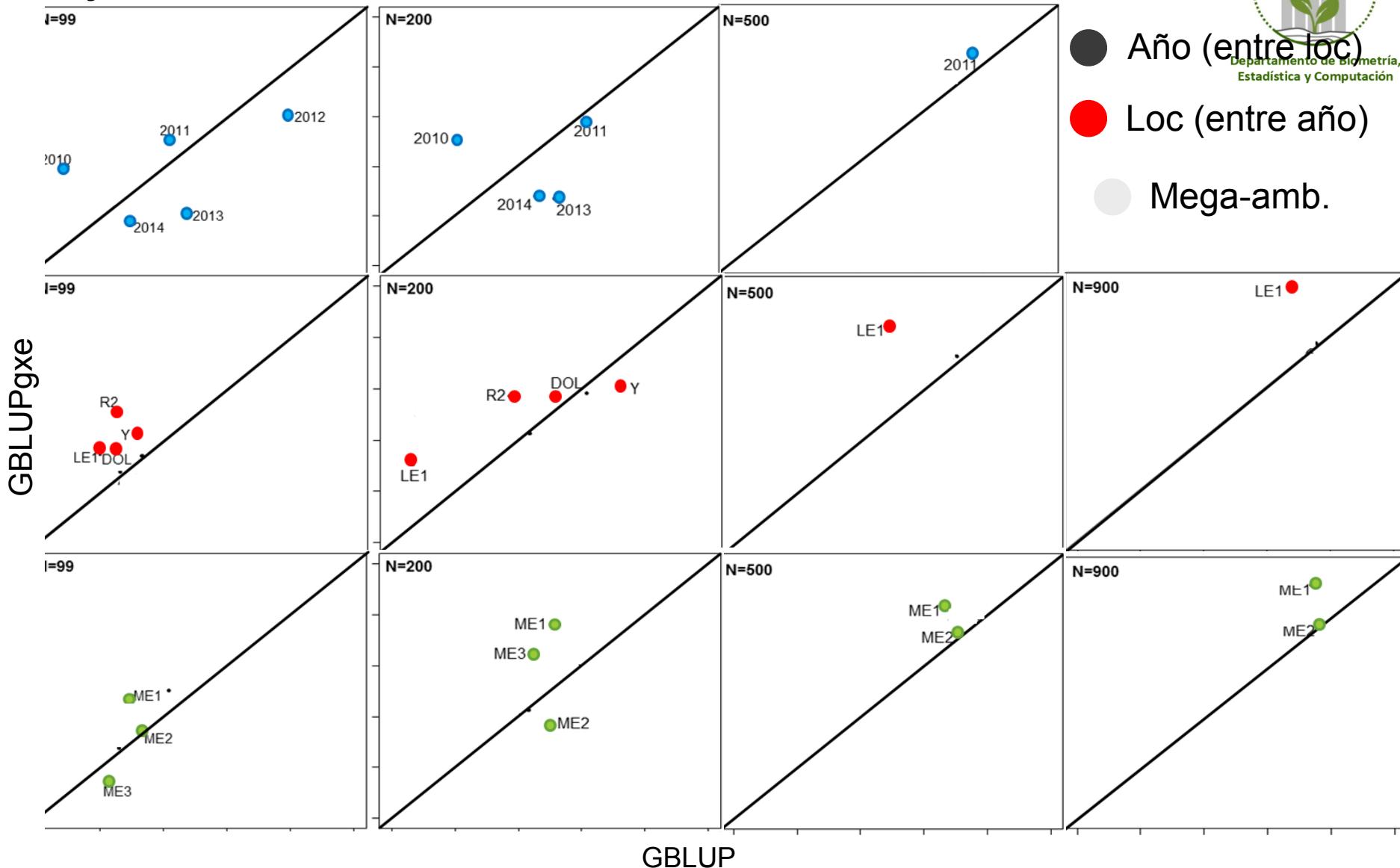
La IGA entre años dentro de las localidades es más alta

Predicciones en la mayoría de las localidades es mas precisa con el modelo
GBLUPgxe



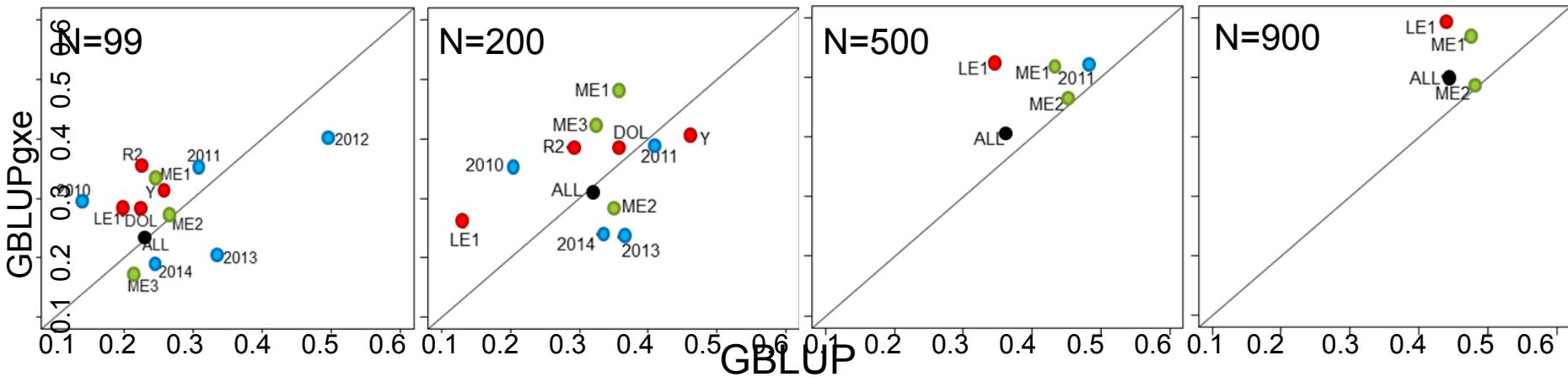
Departamento de Biometría,
Estadística y Computación

Objetivo 1: Resultados



Las predicciones dentro de mega-ambientes son más precisas utilizando el modelo GBLUPgxe

Comparación de modelos multi-ambiente en SG



El modelo GBLUPgxe tiende a ser más preciso que el modelo GBLUP al aumentar el tamaño de la población

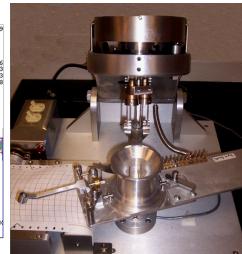
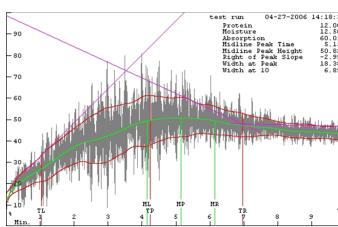
En presencia de mayores niveles de IGA se observa un mejor desempeño de los modelos multi-ambiente



OBJETIVOS

Objetivos Específicos:

- 1) Mejorar la capacidad predictiva del valor genético de los individuos modelando la IGA en la predicción de nuevos individuos y de individuos parcialmente fenotipados
- 2) Comparar la habilidad predictiva de modelos que incluyen caracteres múltiples correlacionados buscando optimizar la asignación de los recursos de fenotipado.
- 3) Predecir la varianza genética de la progenie considerando la población en equilibrio o desequilibrio de ligamiento y comparar la selección de los mejores cruzamientos incluyendo o no la varianza en dicha selección.



RESULTADOS



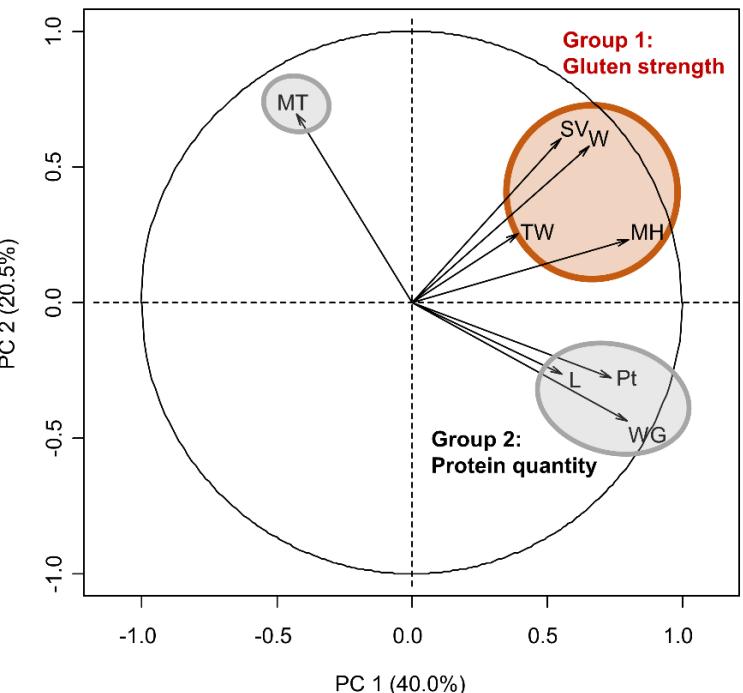
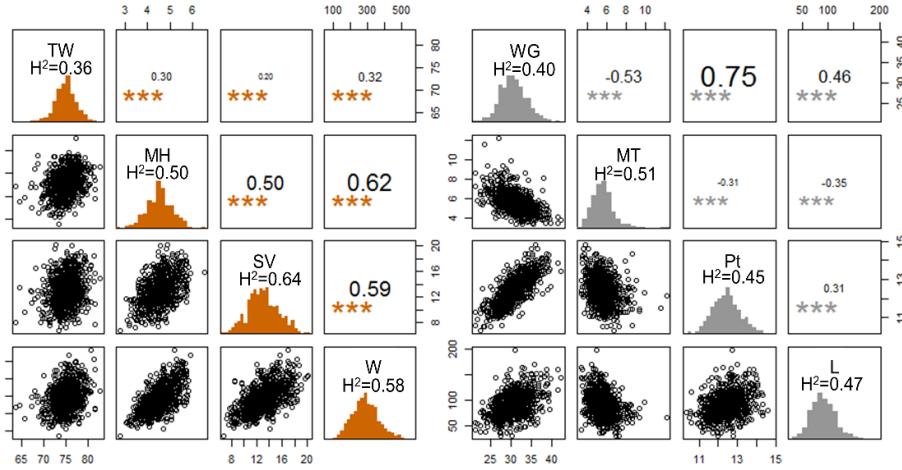
Departamento de Biometría,
Estadística y Computación



- Altura máxima del mixograma (HMAX)
- peso hectolítrico (PH)
- índice de sedimentación (SDS)
- parámetro W del alveograma

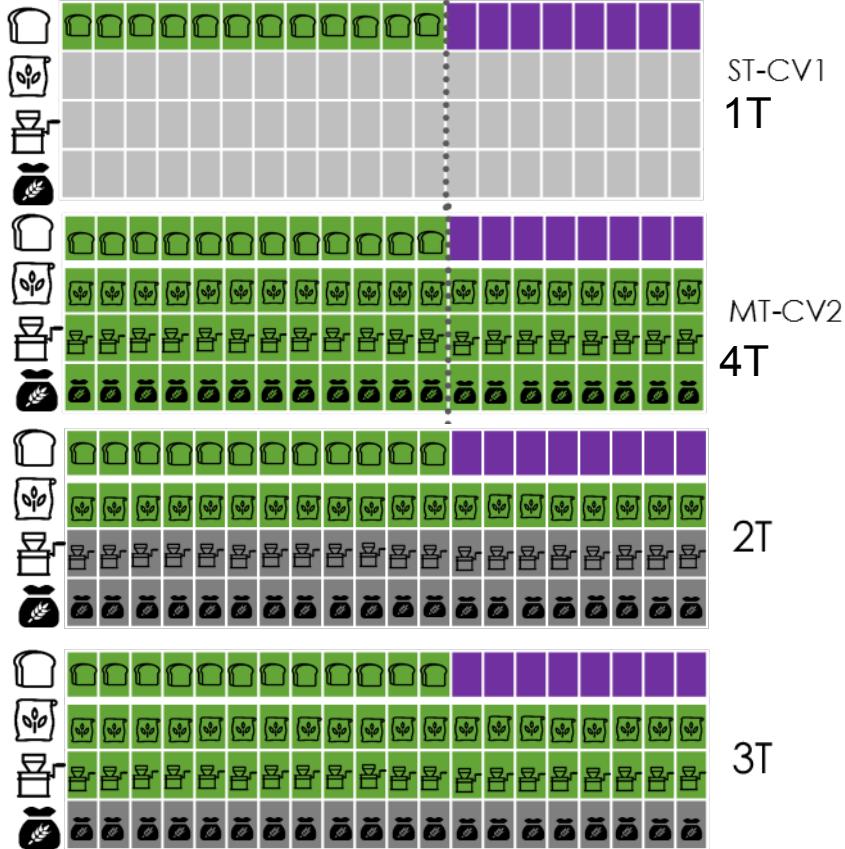
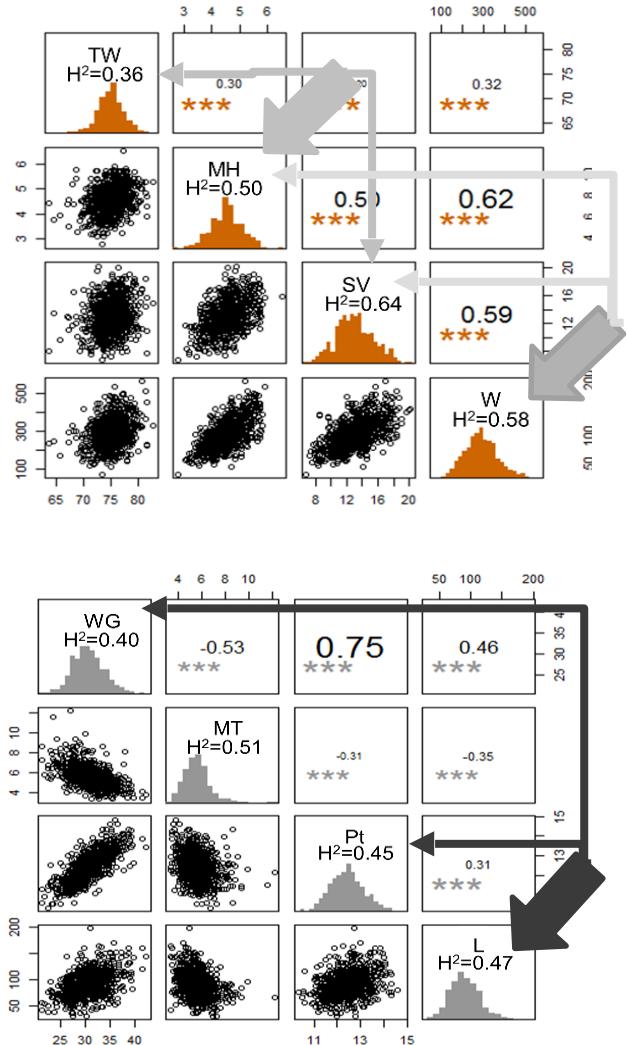


- Proteína total (Pt)
- Gluten húmedo (GH)
- parámetro L del alveograma
- Tiempo de mezclado del mixograma



ESTADÍSTICA
DE LA REPÚBLICA
NACIONAL

Comparación estrategias de predicción



Entrenamiento
Predicción

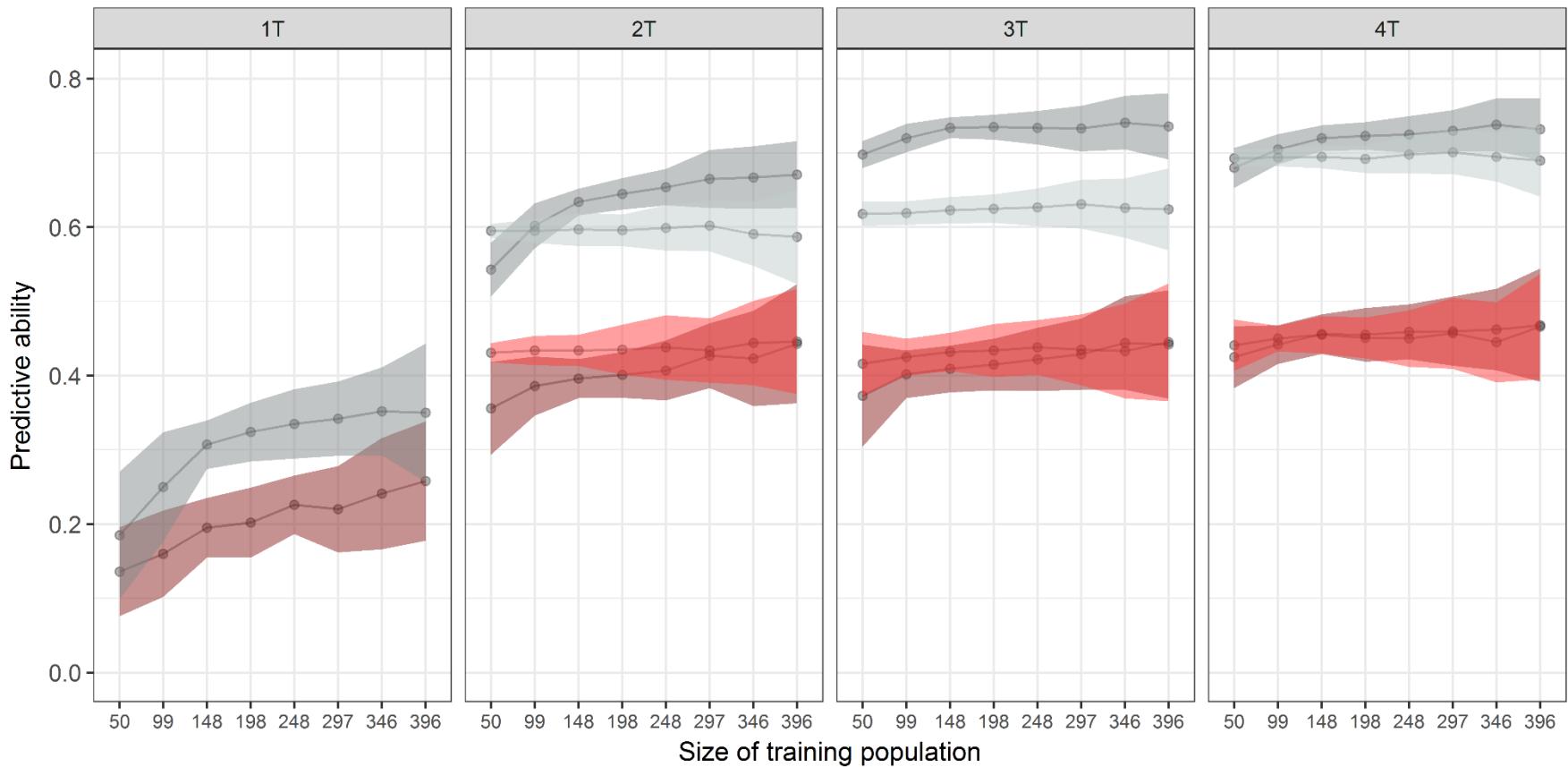


RESULTADOS



Departamento de Biometría,
Estadística y Computación

● L_G ● L_{NG} ● W_G ● W_{NG}



La inclusión de dos más caracteres incrementa la precisión de las predicciones si estos están correlacionados con el carácter a predecir



FACULTAD DE
AGRONOMÍA
UNIVERSIDAD DE LA REPÚBLICA



Selección de cruzas

- 1) Mejorar la capacidad predictiva del valor genético de los individuos modelando la IGA en la predicción de nuevos individuos y de individuos parcialmente fenotipados
- 2) Comparar la habilidad predictiva de modelos que incluyen caracteres múltiples correlacionados buscando optimizar la asignación de los recursos de fenotipado.
- 3) Predecir la varianza genética de la progenie considerando la población en equilibrio o desequilibrio de ligamiento y comparar la selección de los mejores cruzamientos incluyendo o no la varianza en dicha selección.

Predicción de la varianza asumiendo equilibrio de ligamiento

$$\hat{V}_{LE_{ij}} = \sum_{k=1}^{k=M} \left[\left(1 - (p_{k+}(ij) - p_{k-}(ij))^2 \right) * \hat{u}_k^2 \right]$$

k : k-ésimo SNP

i : i-ésimo parente 1

j : j-ésimo parente 2

Término que considera las diferencias entre los padres

$p_{k+}(ij)$: frecuencia del bialélo +1

$p_{k-}(ij)$: frecuencia del bialélo -1

Efecto de los marcadores

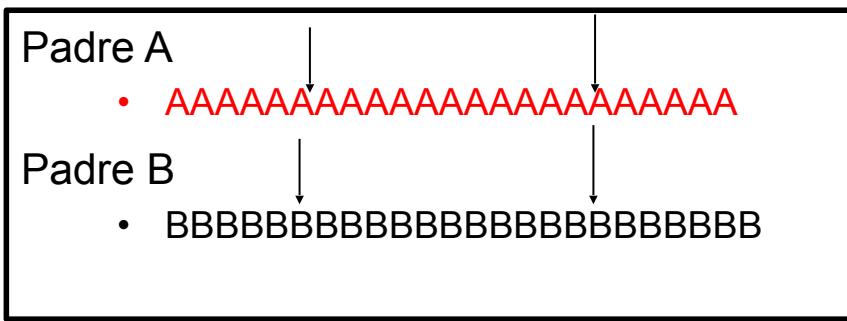
1000 valores genéticos de la progenie se simularon asumiendo $\sim N(MP, V_{LE})$

No se requiere de marcadores con posición en el genoma



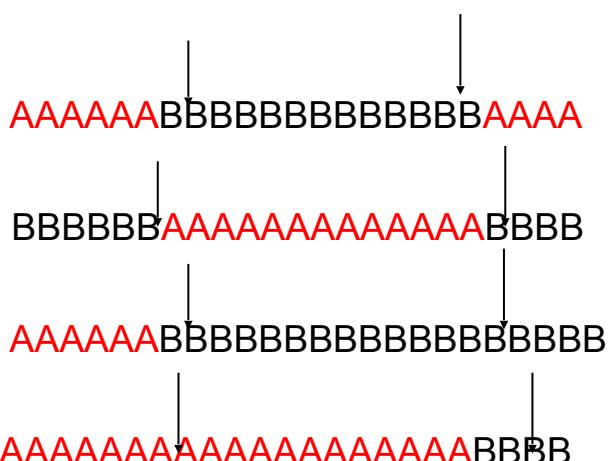
Predicción de la varianza considerando desequilibrio de ligamiento

Puntos aleatorios de recombinación



Requiere marcadores con posiciones para simular puntos de recombinación

Progenie genérica



Se *simularon* 1000 progenies (RILs) por cada cruce

$$\hat{V}_{LD_{ij}} = var(\hat{g}_{1000 \text{ RILs}})$$

'PopVar' R package. Mohammadi et al. 2015



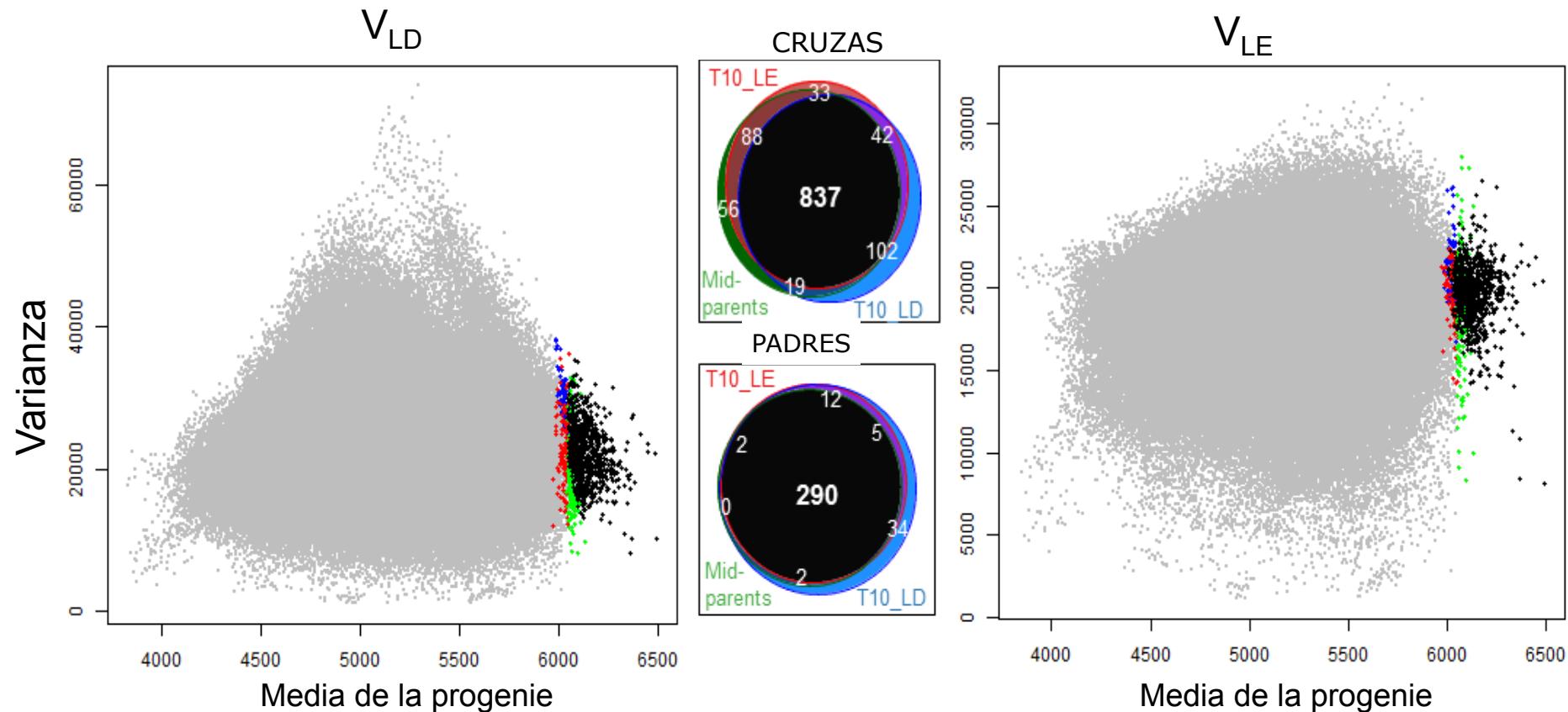
RESULTADOS



Departamento de Biometría,
Estadística y Computación

Selección por la media del 10% superior de la progenie
Varianza vs. media del valor genético de los padres de toda la
usando V_{LD} y V_{LE} , y por la media de los padres.

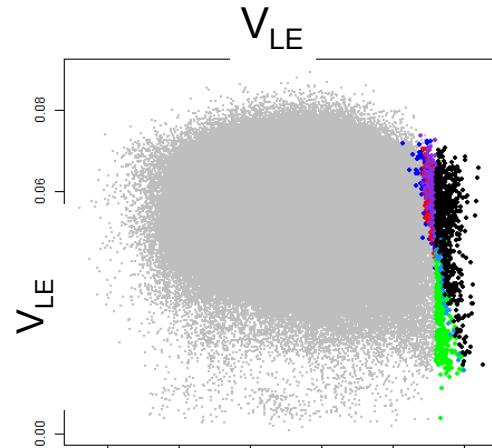
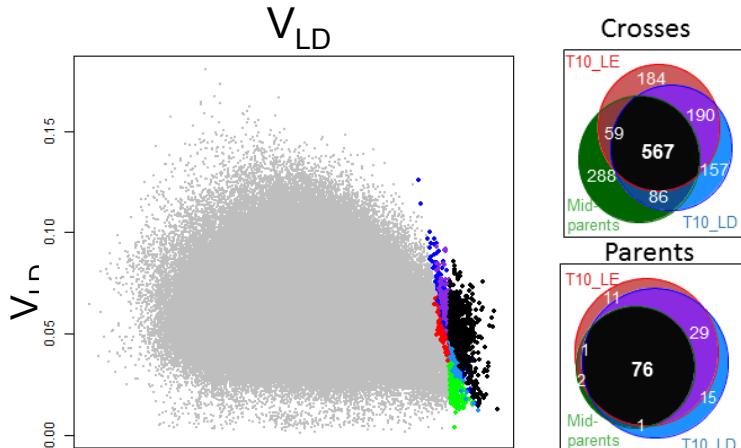
La varianza se calculó asumiendo la población en equilibrio de
ligamiento y considerando el desequilibrio de ligamiento



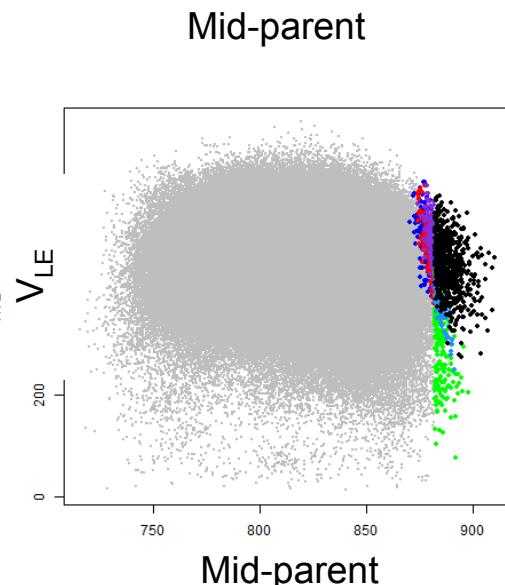
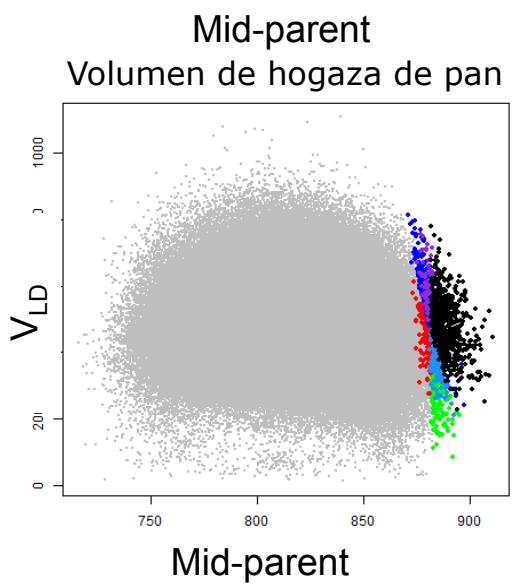
La selección de las mejores cruzas para rendimiento estuvo
determinada por el valor de la media de los padres.

Calidad panadera en trigo

Contenido de proteína en grano



Para calidad panadera hubieron menos cruzas en común que para rendimiento



De todos modos aún un alto número de cruzas fueron comunes a los 3 métodos.



Instituciones que financiaron la investigación



UNIVERSIDAD
DE LA REPÚBLICA



Instituto Nacional de Investigación Agropecuaria
U R U G U A Y



AGENCIA NACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN

GRACIAS