

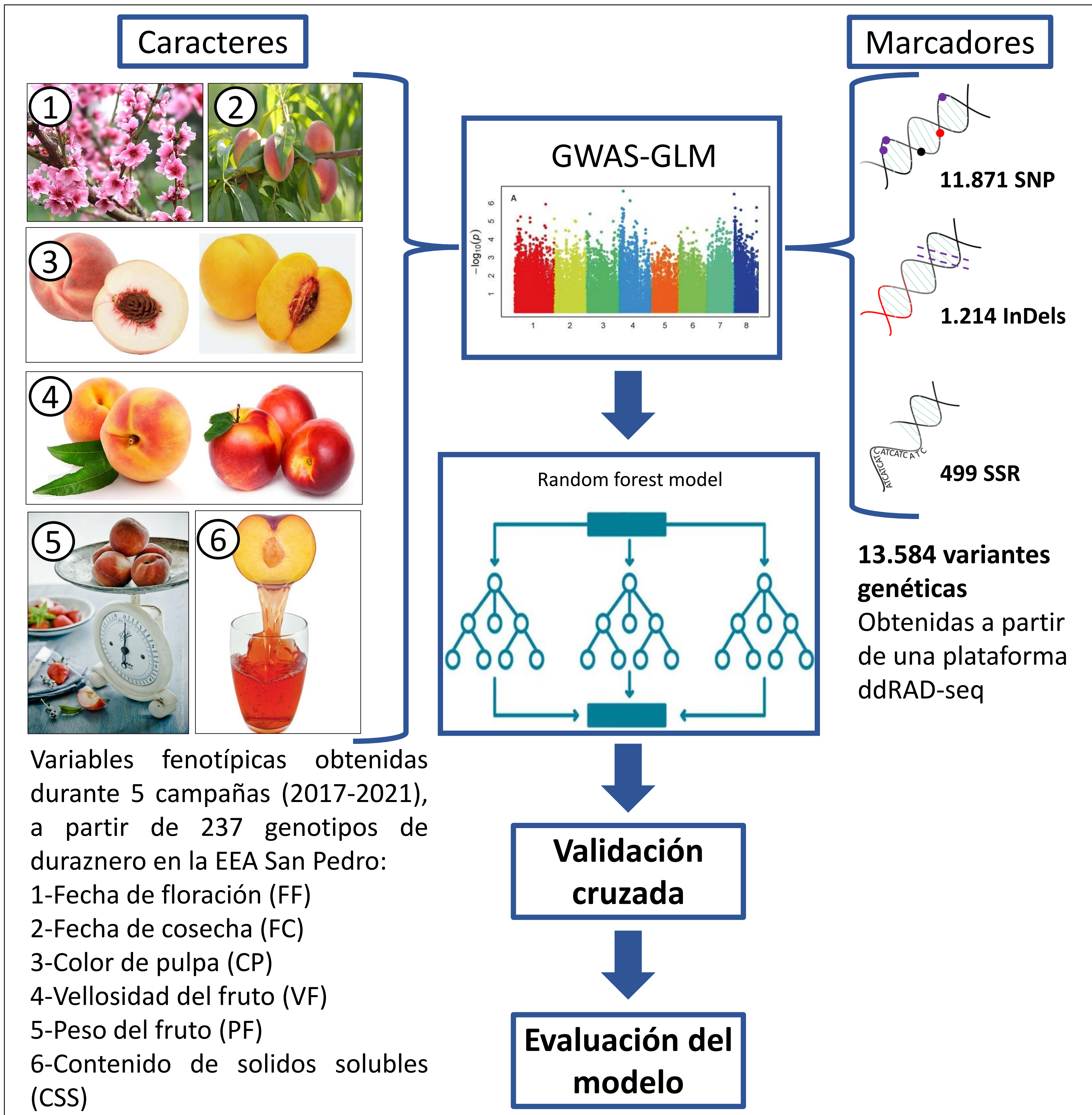


Aballay, Maximiliano M.<sup>1</sup>; Chirino, Julián S.<sup>2</sup>; Valentini, Gabriel H.<sup>2</sup>; Sánchez Gerardo<sup>1</sup>  
<sup>1</sup>Laboratorio de biotecnología, <sup>2</sup>Estación Experimental Agropecuaria (EEA) San Pedro, INTA, Ruta N° 9, Km 170, San Pedro (2930), Buenos Aires, Argentina.  
 E-mail: sanchez.gerardo@inta.gov.ar

## INTRODUCCIÓN

El mejoramiento genético de duraznero es una actividad a largo plazo que implica una gran inversión de tiempo en comparación con los cultivos anuales debido principalmente a dos desafíos: largos períodos juveniles y plantas de gran tamaño. El desarrollo en las tecnologías de secuenciación que se ha producido en los últimos años hace posible acceder a un gran caudal de información, ayudando a acelerar el proceso de mejoramiento. Este nivel de información nos permite aplicar técnicas de aprendizaje automático que pueden encontrar patrones, y realizar predicciones que ayuden al proceso de mejoramiento.

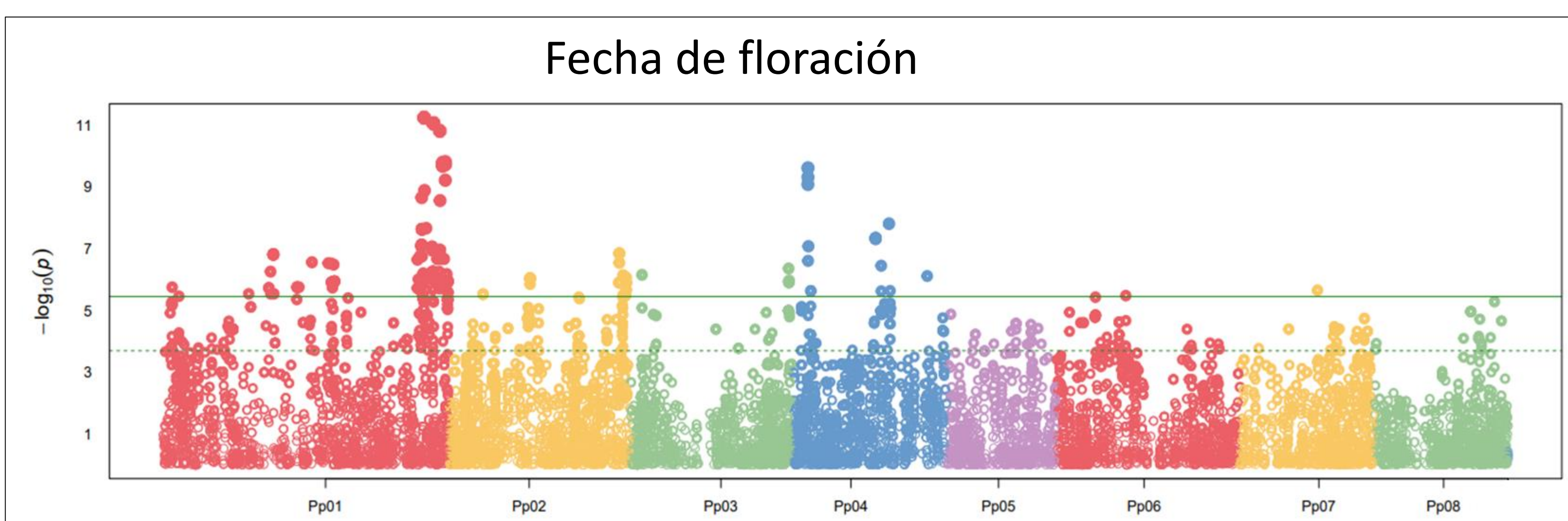
## MATERIALES Y MÉTODOS



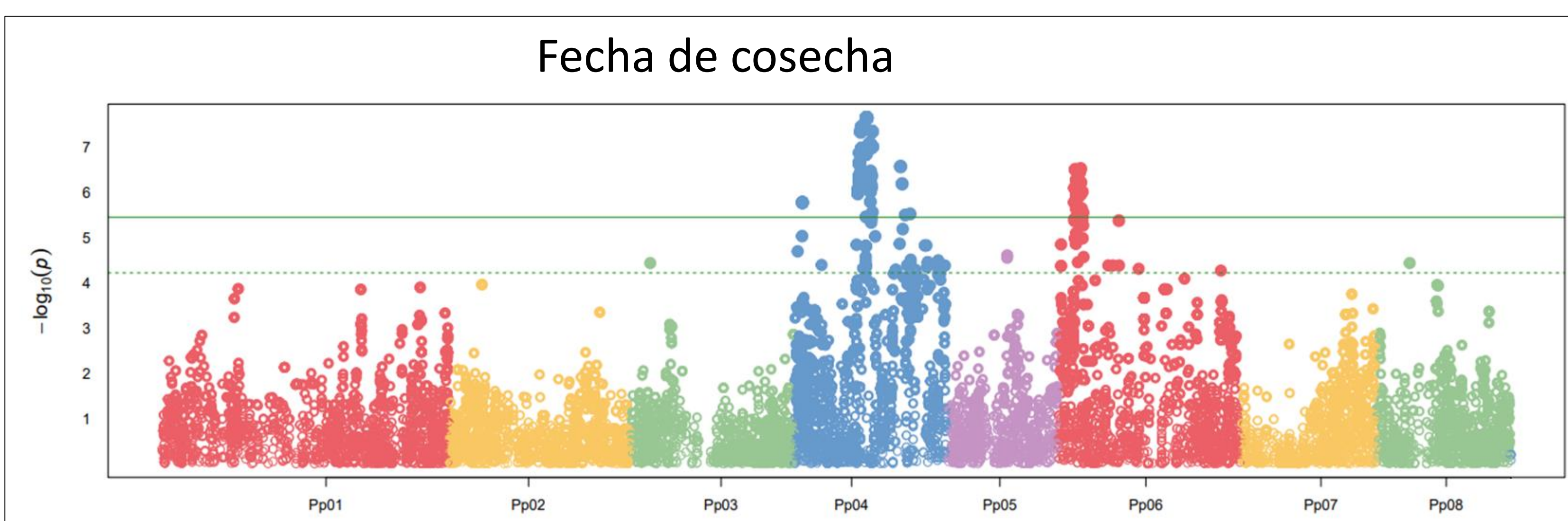
Variables fenotípicas obtenidas durante 5 campañas (2017-2021), a partir de 237 genotipos de duraznero en la EEA San Pedro:

- 1- Fecha de floración (FF)
- 2- Fecha de cosecha (FC)
- 3- Color de pulpa (CP)
- 4- Velloso del fruto (VF)
- 5- Peso del fruto (PF)
- 6- Contenido de sólidos solubles (CSS)

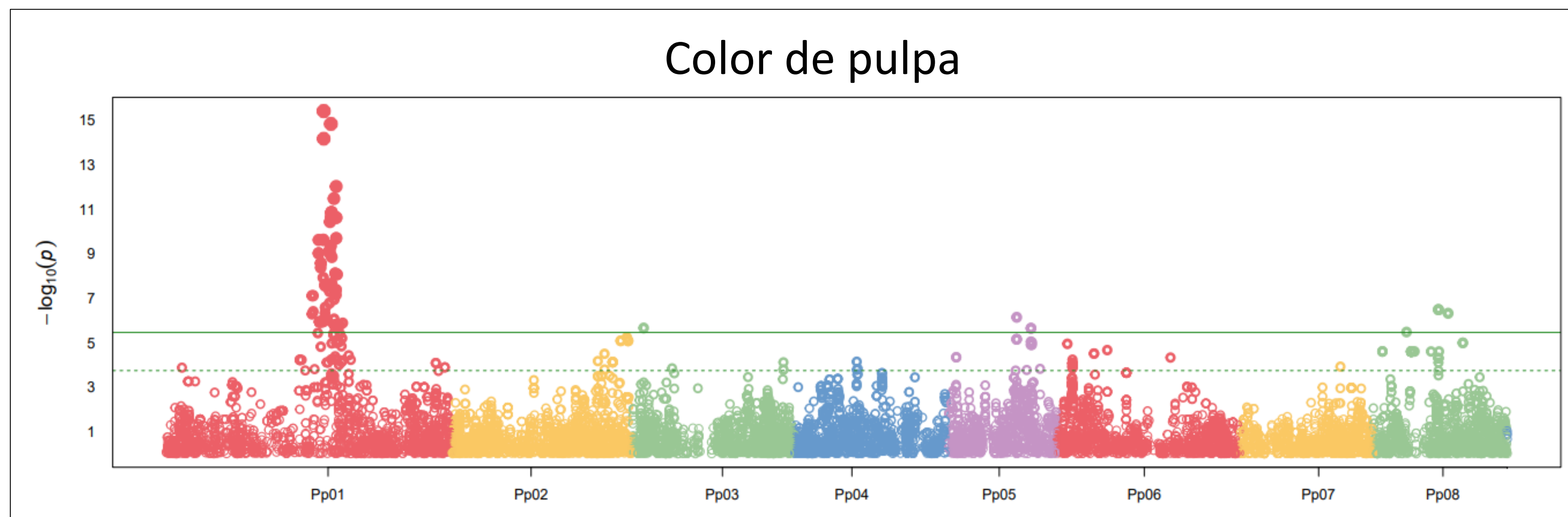
## RESULTADOS



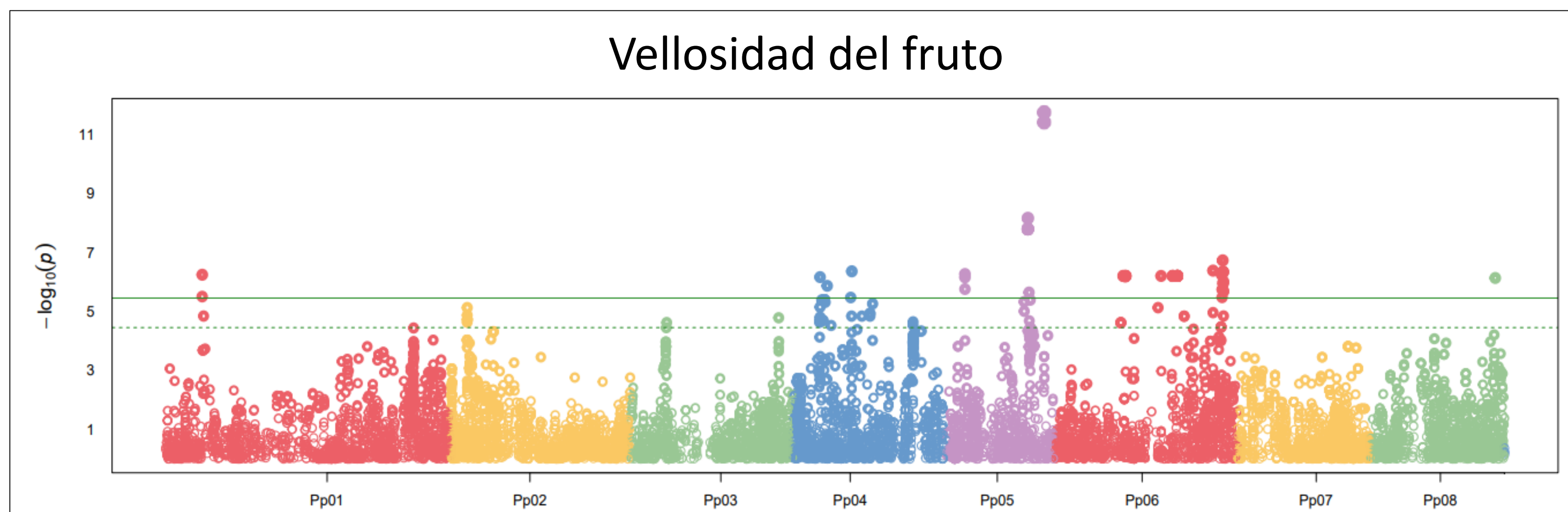
**Figura 1: Marcadores asociados a fecha de floración (campaña 2017).** En las 5 campañas se observó la presencia de un pico de asociación ubicado en el cromosoma 1, y dos picos de asociación ubicados en el cromosoma 4.



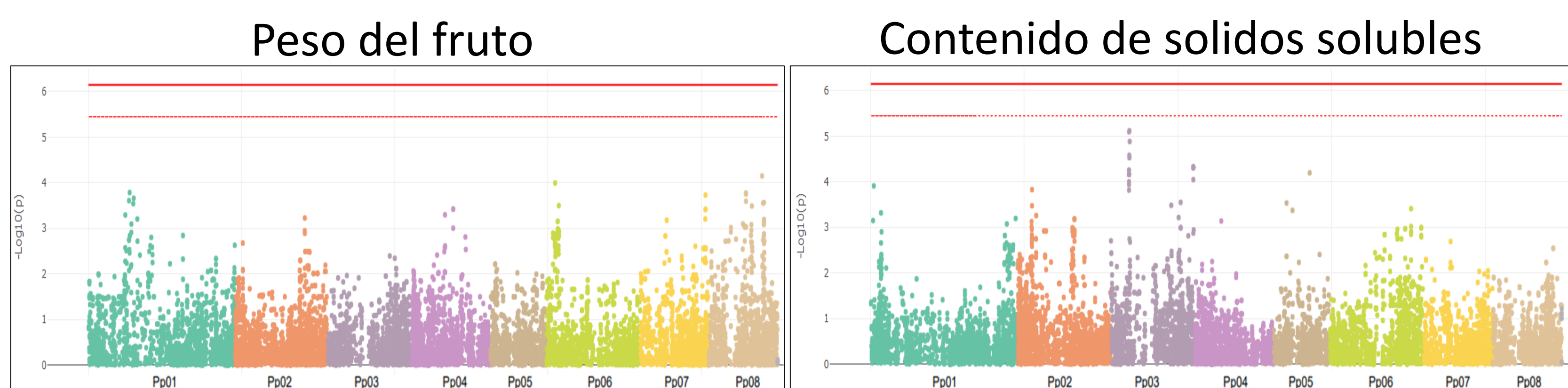
**Figura 2: Marcadores asociados a fecha de cosecha (campaña 2018).** En las 5 campañas se observó la presencia de un pico de asociación ubicado en el cromosoma 4, y en tres campañas se observó además la presencia de un pico de asociación en el cromosoma 6.



**Figura 3:** este carácter categórico presenta un alto valor de LOD con marcadores presentes en el cromosoma 1.



**Figura 4:** En los marcadores analizados se observan 2 SNP que presentan un alto de valor de LOD, los cuales se encuentran en una posición cercana al locus descrito como el responsable de este carácter.



**Figura 5:** Los caracteres PF y CSS no presentaron asociación en ninguna de las 5 campañas analizadas.

Carácter	Validación cruzada		Evaluación final	
	R <sup>2</sup> /Acc. del set de entrenamiento	R <sup>2</sup> /Acc. del set de validación	R <sup>2</sup> /Acc. del set de entrenamiento	R <sup>2</sup> /Acc. del set de prueba
FF	84,45%	72,65%	83,51%	78,34%
FC	89,66%	73,83%	88,95%	77,31%
CP	100,00%	86,40%	100,00%	93,90%
VF	100,00%	94,80%	100,00%	100,00%

**Figura 6: Resultados de la validación cruzada y evaluación final del modelo entrenado con los caracteres que presentaron control genético a través de GWAS (FF, FC, CP y VF).**

La validación cruzada y evaluación final del modelo son procesos que se realizaron para evaluar la capacidad que tienen los modelos para predecir correctamente el comportamiento de los caracteres analizados. Para esto se utilizó la métrica R<sup>2</sup> en las variables cuantitativas (FF y FC) y la métrica Accuracy en las variables categóricas (CP y VF). Las mismas se encuentran expresadas en valores porcentuales e indican el nivel de predicciones correctas realizadas por el modelo.

El proceso de validación cruzada presentó valores altos de Accuracy para los caracteres categóricos (CP y VF), en donde ambos presentaron un valor máximo (100%) para el set de entrenamiento, y valores de 86,40% (CP) y 94,80% (VF) para el set de validación. Estos caracteres tuvieron un desempeño similar en la evaluación final del modelo con valores de Accuracy máximos (100%) para el set de entrenamiento, y valores de 93,90% (CP) y 100,00% (VF) para el set de prueba.

Los caracteres cuantitativos (FF y FC) presentaron en la validación cruzada valores de R<sup>2</sup> superiores al 80% para el set de entrenamiento y valores superiores al 70% para el set de validación. En el caso de la evaluación final se observaron valores de R<sup>2</sup> superiores al 80% para el set de entrenamiento y valores superiores al 70% para el set de prueba.

## CONCLUSIONES

- Para el set de datos analizados, aquellos caracteres que mostraron asociación (GWAS), también presentaron un buen desempeño en las predicciones realizadas con valores mayores a 70% de Accuracy/R<sup>2</sup>.
- Las variables categóricas presentaron un mejor desempeño en las predicciones realizadas.
- Los resultados presentados en este trabajo demuestran la capacidad que tienen las herramientas de aprendizaje automático como apoyo para los programas de mejoramiento de duraznero.