

# Predicción genómica: avances y perspectivas para el mejoramiento del arroz

Título corriente: Predicción genómica para el mejoramiento del arroz

Jérôme Bartholomé<sup>1,2,3\*</sup>, Parthiban Thathapalli Prakash<sup>3</sup>, y Joshua N. Cobb<sup>4</sup>

<sup>1</sup> CIRAD, UMR Instituto AGAP, F-34398, Montpellier, Francia,

<sup>2</sup> Instituto AGAP, Univ Montpellier, CIRAD, INRAE, Montpellier SupAgro, Montpellier, Francia

<sup>3</sup> Plataforma de Mejoramiento del Arroz, Instituto Internacional de Investigación del Arroz, DAPO Box7777, Metro

Manila, Filipinas

<sup>4</sup> RiceTec Inc, PO Box 1305, Alvin, TX 77512 EE. UU.

\*Autor para correspondencia: [jerome.bartholome@cirad.fr](mailto:jerome.bartholome@cirad.fr)

## Resumen

La predicción genómica puede ser una herramienta poderosa para lograr mayores tasas de ganancia genética para rasgos cuantitativos si se integran completamente en una estrategia de reproducción. En el arroz como en otros cultivos, la

El interés en la predicción genómica es muy fuerte con una serie de estudios que abordan múltiples aspectos de su uso, que van desde los más conceptuales a los más prácticos. En este capítulo, nos

revisar la literatura sobre el arroz (*Oryza sativa*) y resumir las consideraciones importantes para el

integración de la predicción genómica en los programas de mejoramiento. El programa de cría en regadío de la

El Instituto Internacional de Investigación del Arroz se utiliza como un ejemplo concreto sobre el que proporcionamos datos.

y R scripts para reproducir el análisis, pero también para resaltar los desafíos prácticos con respecto a la

uso de predicciones. El dicho: *“Para alguien con un martillo, todo parece un clavo”*

describe una trampa psicológica común que a veces plagla la integración y

aplicación de las nuevas tecnologías a una disciplina. Hemos diseñado este capítulo para ayudar al arroz

los criadores evitan ese escollo y aprecian los beneficios y limitaciones de aplicar genómica

predicción, ya que no siempre es el mejor enfoque ni el primer paso para aumentar la tasa de

ganancia genética en todos los contextos.

**Palabras clave:** arroz, programa de mejoramiento, predicción genómica, selección genómica, *Oryza sativa*

# 1. Introducción

El objetivo de todo programa de fitomejoramiento es proporcionar variedades mejoradas que satisfagan las necesidades de las partes interesadas clave (participantes en la cadena de valor, desde agricultores hasta consumidores). Una limpieza comprensión de la biología y la genética de la especie combinada con un producto específico concepto son elementos clave para lograr este objetivo [1]. Sin embargo, el paisaje genético que un el criador necesita explorar para identificar productos superiores es muy grande y excede materialmente el capacidad de los programas de mejoramiento [2]. De hecho, el fitomejoramiento puede considerarse como un número juego donde los esquemas de mejoramiento están diseñados para aumentar la probabilidad de encontrar genotipos con combinaciones deseables de características utilizando una cantidad limitada de recursos [3].

El esquema de mejoramiento es el marco conceptual que captura todas las actividades que realiza un criador. durante un ciclo de cría. Un solo ciclo de reproducción se puede resumir en cuatro partes principales: creación, evaluación, selección y recombinación [4] y está diseñado para crear nuevas variaciones, evaluar con precisión el desempeño del germoplasma de mejoramiento y recombinar individuos para formar una cohorte mejorada. La evaluación es una parte central de un esquema de mejoramiento que implica múltiples pasos de fenotipado diseñados para estimar el valor genético heredable (o valor genético) de los candidatos a la selección [5]. En el caso del rendimiento, generalmente un conjunto de genotipos preseleccionados por rasgos altamente hereditarios se evalúan en ensayos multiambientales (MET) destinados para representar la población objetivo de entornos (TPE) en los que se espera que el producto realizar [6, 7]. Estos pasos finales del proceso de evaluación requieren importantes recursos y abarcan varios años en la mayoría de los programas de fitomejoramiento [3]. para superar esta limitación y aumento de la eficiencia de los programas de mejoramiento, varias metodologías y herramientas han surgido en las últimas tres décadas debido en gran parte a las mejoras en la caracterización de polimorfismos de ADN y poder de cómputo [8]. Entre ellos, los métodos que utilizar información molecular para inferir el rendimiento fenotípico (como la selección asistida por marcadores) [9, 10] y la selección genómica [11]) son herramientas importantes que permiten que los programas de mejoramiento modernos

maximizar el uso de sus recursos limitados. A diferencia de la clásica selección asistida por marcadores, la predicción genómica explica los loci de rasgos cuantitativos de efecto grande y pequeño, por lo tanto capturando una mayor proporción de la variación genética de un rasgo [12, 13].

El concepto de selección genómica fue propuesto por primera vez por Meuwissen et al. [11] para animales cría. En este estudio de simulación, los autores predijeron el valor genético basándose en marcadores de juveniles sin registros fenotípicos utilizando los animales de los dos anteriores generaciones para estimar los efectos del marcador. Obtuvieron altas precisiones para los valores predichos. valores genéticos (valores genéticos estimados genómicos - GEBV) y concluyó que este enfoque para aumentar la tasa de ganancia genética tiene potencial cuando se combina con técnicas para reducir intervalos de generación. La selección genómica comúnmente se refiere al proceso donde la selección los candidatos, que solo están genotipados, se seleccionan en función de su GEBV (predicciones genómicas).

Para lograr esto, primero se modela la relación marcador-fenotipo utilizando un conjunto de entrenamiento (un conjunto más pequeño conjunto representativo de individuos que refleja lo más fielmente posible la genética del individuos previstos para la predicción) en los que los datos de marcadores fenotípicos y de todo el genoma son a la vez generado [12, 14]. Para evaluar el desempeño de los modelos, la mayoría de las veces, la correlación entre los valores pronosticados y observados se calcula usando una población de validación cuyo la composición depende de la estrategia de validación [15]. Esta métrica generalmente se conoce como precisión o capacidad predictiva según con qué valores observados se comparan las predicciones: valores genéticos o rendimientos fenotípicos, respectivamente.

El desarrollo acelerado de la tecnología de genotipado de densidad media y alta durante la década de 2010 condujo al primer informe del uso práctico de la predicción genómica en el ganado lechero [16] seguido de contribuciones importantes de los mejoradores que trabajan en especies de plantas importantes para la agricultura [17, 18]. De hecho, la predicción genómica es ahora un intenso campo de investigación que busca optimizar su uso y integración en los programas de mejoramiento de plantas y animales a nivel mundial. Avances importantes se han hecho con respecto a nuestra comprensión de los principales factores que afectan la precisión de la

GEBVs incluyendo: el tamaño efectivo de la población del programa de mejoramiento, la heredabilidad y arquitectura genética de los rasgos objetivo, el tamaño y la composición de la población de entrenamiento, así como el número, distribución e informatividad de los marcadores [19]. predicción genómica

Los modelos y su implementación en herramientas de software también ha recibido especial atención con el fin de aprovechar de manera eficiente toda la información contenida no solo en conjuntos de datos genómicos y fenotípicos, sino también también en otras fuentes de datos “ómicos” [20]. Si bien los impulsores de la precisión de la predicción son cada vez mejor entendida, la cuestión de cómo la predicción genómica se integra mejor en un

La estrategia de fitomejoramiento existente sigue siendo un desafío ya que los programas de mejoramiento operan en una amplia variedad de contextos (características objetivo, especies, recursos, escala, etc.).

El arroz (*Oryza sativa*) es una especie modelo para la biología molecular [21] y un alimento básico para una gran parte de la humanidad. Se obtuvieron importantes ganancias en la productividad gracias a los esfuerzos de mejoramiento durante e inmediatamente después de la revolución verde [22, 23]. Estas mejoras fueron realizado principalmente a través de la selección fenotípica en grandes viveros de pedigrí segregantes [22, 24, 25]. El uso de marcadores moleculares también fue clave para la introgresión de alelos mayores que confieren resistencia al estrés biótico [26] o abiótico [27, 28]. El éxito de esta estrategia dependía en gran medida en la alta heredabilidad y arquitecturas genéticas simples de los rasgos bajo selección (planta altura, madurez, resistencia a enfermedades, tipo de grano) y las muy grandes y bien caracterizadas diversidad genética de *O. sativa* [29, 30] y especies estrechamente relacionadas como *O. glaberrima* (African arroz), *O. rufipogon* u *O. nivara* [31, 32]. Esto puede explicar por qué el interés por implementar la predicción genómica en la comunidad mundial de mejoramiento del arroz se ha retrasado en relación con los animales criar o criar cultivos tradicionalmente de polinización cruzada como el maíz. Durante este tiempo, lleva mencionando que algunos avances clave se realizaron a través de la mejora de la población a través de un estrategia de selección recurrente en América Latina [25, 33]. Sin embargo, más recientemente, la aceleración en la ganancia genética para el rendimiento en otras especies, los costos decrecientes del genotipado y la creciente

importancia de la sostenibilidad en la producción de arroz han contribuido a un mayor interés en

implementación de la predicción genómica en el mejoramiento del arroz.

En este capítulo, primero damos una visión general de la investigación sobre predicción genómica en arroz con un

centrarse en estudios que hagan uso de la estrategia en un programa de mejoramiento. Luego destacamos

Consideraciones importantes para la integración de la predicción genómica en un esquema de mejoramiento de arroz.

En esta segunda parte, aspectos como la identificación de los puntos de entrada para la selección genómica en un

esquema de crianza, el diseño efectivo de poblaciones de entrenamiento, estrategias para reducir el

se presenta el intervalo de generación y la importancia de los sistemas de gestión de datos. En el

tercera parte, tomamos el programa de mejoramiento del Instituto Internacional de Investigación del Arroz (IRRI) para cultivos de regadío

sistemas como un ejemplo para la integración de la predicción genómica en el desarrollo de un producto

programar y proporcionar los datos asociados y los scripts R para ejecutar e interpretar el análisis

(disponible en información complementaria). En la última parte, presentamos interesantes avances en

predicción genómica que puede ayudar aún más a los programas de mejoramiento de arroz a aumentar su eficiencia. Nuestro

El objetivo de este capítulo es proporcionar a los mejoradores de arroz una base sólida para comprender el

ventajas y limitaciones del uso de la predicción genómica en su estrategia de mejoramiento para maximizar

la tasa de ganancia genética para los rasgos relevantes. Debido a la fuerte presencia de arroz endogámico en Asia,

optó por centrar el alcance de este capítulo en el arroz asiático endogámico (*O. sativa*) aunque las especificidades

de aplicar la predicción genómica al arroz híbrido se abordan en menor medida. Por otro

punto de vista sobre la importancia de la predicción genómica para el mejoramiento del arroz, remitimos al lector a la

capítulos de libros de Spindel e Iwata [34] y Ahmadi *et al.* [35].

## 2. La predicción genómica funciona en arroz

La literatura sobre predicción genómica para especies de cultivo es muy rica. Con más de 50 estudios publicados

desde 2014 (Cuadro 1, [36–89]), la predicción genómica en arroz no es una excepción. Informamos sobre

la mayoría de los estudios publicados en arroz (ya sea exclusivamente o en conjunto con otras especies) para

para resaltar el volumen y la diversidad del trabajo realizado hasta la fecha y su relevancia para mejorar las estrategias de crianza. Para lograr esto último, enfatizamos intencionalmente los estudios enfocados en la integración con los programas de mejoramiento, que tienden a reportar los desafíos más prácticos de la implementación.

## 2.1 Descripción general

Los primeros estudios que informaron sobre el uso de la predicción genómica en arroz se publicaron en 2014 (Tabla 1). A pesar de la riqueza de recursos genómicos y de marcadores disponibles en el arroz, estos estudios llegaron, sorprendentemente, cinco años después de los primeros estudios sobre predicción genómica (usando datos reales) que publicado en maíz [90], trigo [91] o cebada [92]. La amplitud de los recursos genómicos disponibles al arroz y la profundidad de la diversidad genética que se ha caracterizado hasta ahora ha llevado a la descubrimiento de muchos QTL principales con tamaños de efecto razonables. Mientras que un único y valioso recurso para la comunidad de mejoramiento de arroz, el fuerte enfoque en el descubrimiento, caracterización y la introgresión de QTL de gran efecto de germoplasma exótico puede haber servido para retrasar la transición hacia la predicción genómica [93]. El tipo de poblaciones evaluadas en estos primeros estudios genómicos los estudios de predicción en arroz tienden a reforzar esa impresión (Figura 1A). De hecho, entre los primeros tres estudios publicados en 2014, dos se basaron en el mismo panel de diversidad [94] y uno en híbridos derivados de una población de mapeo (un F2 inmortalizado [95]). En general, los paneles de diversidad que fueron, en muchos casos, diseñados para estudios de asociación [94] representaron una gran proporción de los estudios publicados hasta el momento (Figura 1A). Para la mayoría de estos estudios, el objetivo era metodológico: comprensión de los impactos de la estructura de la población, integración de conocimiento sobre arquitectura genética de rasgos, optimización de conjuntos de entrenamiento, comparación de modelos o integración de modelos de cultivo sin implicación directa en un programa de mejoramiento. Dada la extensión de estructura de subpoblaciones ancestrales en arroz, el uso de paneles de diversidad para evaluar genómica Es probable que los modelos de predicción induzcan un sesgo en la estimación de la capacidad predictiva. De hecho si el

estructura de la población no se tiene en cuenta, la mayor parte de la capacidad de predicción puede surgir de la capacidad de predecir entre subpoblaciones y no dentro de subpoblaciones [36]. Aparte de estudios basados en paneles de diversidad, 16 estudios utilizaron líneas de reproducción, nueve estudios se centraron en híbridos, seis utilizaron una población de mapeo, cuatro estudios se basaron en poblaciones sintéticas y tres cultivares usados (Figura 1A).

Además de la amplia variedad de poblaciones encontradas en estos estudios, el tamaño de la población, el número de marcadores, el número de fenotipos o el número de ambientes utilizados para caracterizar las poblaciones fueron muy variables (Figura 1B). La población más grande El tamaño (2265) se logró utilizando datos disponibles públicamente del proyecto del genoma del arroz 3000 [85]. Dadas las limitaciones y dificultades que rodean la recopilación de datos de fenotipo de alta calidad, comprensiblemente, la mayoría de los estudios emplearon tamaños de población de alrededor de 300 (Tabla 1). En los casos en que Se utilizaron grandes poblaciones de 1.000 o 2.000 individuos, el fenotipado se hizo de forma muy número limitado de entornos (generalmente 1 o 2). De hecho, menos de la mitad de los estudios utilizaron más de tres ambientes para evaluaciones fenotípicas (Figura 1B). Entre los tres estudios teniendo información fenotípica en 10 o más ambientes (año, estación o lugar), dos son basado en germoplasma de programas de mejoramiento [49, 64] pero los conjuntos de datos estaban desequilibrados (no todos individuos fenotipados en todos los ambientes o genotipados). El tercer estudio de Jarquin *et al.* [86] usó la información de 51 entornos en combinación con la duración de los días para predecir los días a la partida de genotipos no probados. Entre la amplia variedad de características consideradas, el tiempo de floración (o madurez), la altura de la planta y el rendimiento de grano fueron los más comunes. El número de marcadores varió de 162 [48] a 4 millones [87], y la mayoría de los estudios utilizaron unos pocos miles marcadores (Tabla 1). Genotipado por secuenciación y SNP fijo (polimorfismo de un solo nucleótido) Los arrays fueron las tecnologías más utilizadas. En algunos casos, densidades de marcador muy altas se obtuvieron a través de la resecuenciación del genoma completo con una cobertura generalmente baja (1X o 2X) seguido de imputación [73, 85].



Los métodos estadísticos para la predicción genómica han sido un foco central de muchos estudios en todo el mundo.

especies donde se ha aplicado. A través de los 54 estudios de arroz, 33 métodos diferentes fueron

evaluado con el método genómico de mejor predicción lineal imparcial (GBLUP) siendo el más

utilizado (Figura 1C). Desde que se propuso este método [96], su flexibilidad y robustez han

le permitió convertirse rápidamente en un método de referencia para la cría de animales y plantas. Similar a

el pedigrí tradicional BLUP [97], GBLUP utiliza una matriz de relación aditiva que se basa en

marcadores en lugar de información de pedigrí. Varias extensiones o variaciones de este modelo aditivo

han sido propuestos para dar cuenta de la dominancia y/o la epistasis [37, 53] o para usar otras "ómicas"

datos (transcriptoma o metaboloma) para estimar la relación entre individuos [80, 89]. En

además de GBLUP, RKHS (espacio de Hilbert del núcleo reproductor), frecuentista y LASSO bayesiano

(menor contracción absoluta y operador de selección), RR-BLUP (regresión de cresta BLUP), RF

(bosque aleatorio), SVM (máquina de vectores de soporte), PLSR (regresión de mínimos cuadrados parciales),

BayesB y BayesC fueron los métodos más utilizados en estos estudios sobre arroz (Figura 1C). Otro

métodos de la gran familia de enfoques de aprendizaje automático, como el aumento de gradiente

máquina (GBM) o red neuronal artificial (ANN), también se evaluaron en el contexto de

predicción genómica con resultados mixtos [68, 85].

La composición del conjunto de validación, que puede desempeñar un papel importante en la determinación de la

la precisión de las predicciones dependía en gran medida de la estrategia de validación utilizada en cada estudio

(Tabla 1). Salam *et al.* [15] definió tres tipos principales de métodos de validación: validación cruzada

(validación de subconjuntos), validación entre conjuntos y validación de progenie dependiendo de la composición de

los conjuntos de entrenamiento y validación. La validación cruzada o validación de subconjuntos (K-fold, dejar uno fuera,

muestreo aleatorio o estratificado) fue, con mucho, la estrategia más utilizada entre todos los estudios que

hemos compilado (Fig. 1D). Este método de validación es muy conveniente porque solo necesita

divida sus datos en conjuntos de entrenamiento y validación para poder estimar precisiones sin un

conjunto de datos "independiente" (como se necesita para la validación entre conjuntos o progenie). Por su naturaleza, cruzada

la validación tiende a sobrestimar la precisión de la predicción en comparación con una validación más realista escenarios [57, 98]. La situación se vuelve aún más compleja cuando se utilizan modelos multivariados. usado [99]. Otro enfoque cercano a la validación cruzada, el método HAT [55, 100], se utilizó en cuatro estudios. Este método, basado en la matriz de sombrero de los efectos aleatorios, utiliza la predicción suma residual de cuadrados para estimar la precisión de la predicción y funciona en el contexto de GBLUP, Modelos RKHS y Bayesiano. Este método es considerablemente más rápido que la validación cruzada ya que no es necesario volver a entrenar al modelo [100]. Los métodos de validación entre conjuntos y progenie solo se utilizaron en tres estudios cada uno (Fig. 1D). Teniendo en cuenta el contexto de los programas de mejoramiento donde la integración de la predicción genómica está dirigida principalmente a reducir el tiempo de ciclo, la progenie la validación representa una evaluación más significativa del rendimiento de los modelos de predicción. De hecho, en el concepto inicial de selección genómica, Meuwissen *et al.* [11] progenie usada validación: los modelos se construyeron con datos de las generaciones 1001 y 1002 y las precisiones se calcularon usando los valores pronosticados y los verdaderos valores genéticos de la generación 1003. Además, el decaimiento del desequilibrio de ligamiento que ocurre entre los marcadores y QTL debido a la recombinación en las generaciones de progenie tiende a disminuir la precisión de las predicciones [101], pero los hace interpretables de manera más realista en términos de aplicaciones a la cría práctica. escenarios. Por ejemplo, Ben Hassen *et al.* [57] utilizó la validación de progenie de líneas endogámicas con un número limitado de individuos y encontró una menor capacidad predictiva en comparación con la validación cruzada para los mismos rasgos.

## 2.2 Hallazgos importantes y limitaciones actuales para la predicción genómica en arroz

### 2.2.1 Hallazgos importantes

La Tabla 1 proporciona un breve resumen del objetivo principal de cada estudio en esta revisión. El lector por lo tanto, puede orientarse hacia las publicaciones que son más relevantes para sus preguntas.

A continuación, resumimos resultados importantes centrándonos principalmente en aquellos más relacionados con la implementación en programas de mejoramiento.

un. *La predicción genómica funciona en diferentes contextos.* Los resultados más importantes que surgen de todos los estudios es que la predicción de los rendimientos basada en marcadores moleculares obras. De hecho, la precisión de los GEBV es relativamente alta incluso para rasgos como el rendimiento de grano. Muchos mejoradores de arroz están preocupados por la eficiencia de la predicción genómica, pero es evidente que no está justificado mirar la literatura sobre el arroz y más específicamente los estudios usando mejoramiento germoplasma [41, 45, 53, 54, 57, 62].

B. *Se puede aumentar la precisión de la predicción.* Los criadores pueden jugar con diferentes factores para aumentar la precisión de las predicciones o reducir el costo de implementación. De hecho, por optimizar la composición y evaluación del conjunto de entrenamiento, al enfocarse en información marcadores moleculares (polimórficos con una frecuencia de alelos menores de media a alta y a lo largo del genoma) o integrando datos adicionales (históricos, ambientales covariables, modelo de cultivo, ...) se pueden obtener mejores precisiones. El tamaño y el composición del conjunto de entrenamiento define la fuerza de la relación genética con el candidatos de selección, que es uno de los factores más importantes que impulsan la precisión. Por lo tanto, se han desarrollado algoritmos para seleccionar el conjunto de entrenamiento [39, 42, 46, 73, 78]. En cuanto a los marcadores moleculares, diferentes estudios muestran que la densidad de marcadores puede, en cierta medida, medida, puede reducirse sin afectar la precisión de la predicción. Por ejemplo, Arbeláez *et. Alabama*. [67] diseñó un ensayo SNP rentable con solo 1000 marcadores seleccionados para ser informativo en material de cría de élite y obtuvo buenas precisiones.

C. *Los modelos pueden predecir el rendimiento de la descendencia.* El concepto inicial de selección genómica fue basado en la predicción de los valores genéticos de la descendencia con el objetivo de disminuir la duración de los ciclos de cría [11]. Los muy pocos estudios sobre arroz que realizaron progenie validación [56, 57, 89] muestran resultados prometedores cuando la información de los padres se utiliza para

predecir el rendimiento de la progenie. Sin embargo, aún queda mucho por hacer en esa dirección, ya que la mayor parte del aumento en la ganancia genética relacionada con la integración de la predicción genómica es relacionados con la reducción del tiempo del ciclo de cría.

*D. La predicción genómica es eficiente en el contexto de los híbridos:* gran parte de las lecciones aprendidas

con respecto a las densidades de marcadores, la identificación del conjunto de entrenamiento y la selección del modelo se aplican igualmente a esquemas de reproducción híbridos y endogámicos. Los programas híbridos presentan desafíos únicos donde podrían aplicarse predicciones que no son aplicables a otros esquemas de mejoramiento. De nota es la predicción de cómo los machos y las hembras podrían combinarse para crear superior combinaciones híbridas. En el arroz híbrido hay alguna evidencia de que el desempeño híbrido es impulsado por una convergencia de genética aditiva de las líneas masculina y femenina.

La incorporación de parámetros no aditivos en la predicción no parece ayudar [37].

Si bien esto parece razonable, otros cultivos han mostrado un importante efecto no aditivo.

componente del rendimiento híbrido (p. ej., en maíz [102, 103]). Esta particular conclusión

probablemente estaba sesgado por una base genética muy estrecha y una precisión muy baja para inter-set

predicción del rendimiento de grano. También hay evidencia de que los modelos de rasgos múltiples pueden mejorar

Precisión de predicción para rasgos de baja heredabilidad en arroz híbrido [54,81]. Esto es de particular

importancia en el contexto híbrido ya que muchos rasgos (especialmente el costo de los rasgos de bienes como híbrido

rendimiento de producción de semillas) son particularmente difíciles de medir al principio de un programa de mejoramiento. A

conjunto particularmente único de fenotipos correlacionados asociados con programas híbridos es el

oportunidad de medir el rendimiento per se de los padres endogámicos, así como híbridos

rendimiento del mismo material. Uso de datos de fenotipo parental combinados con datos

sobre el rendimiento de los híbridos puede mejorar la precisión de la predicción del rendimiento del arroz híbrido en aproximadamente

13% [89].

*mi. Modelar GxE aumenta la precisión de la predicción.* Ya sea a través de entornos múltiples

modelos de predicción genómica [56, 62] o combinando modelos de crecimiento de cultivos y modelos genómicos

modelos de predicción [48, 86], varios estudios demostraron la mejor precisión de estos

enfoques para predecir actuaciones específicas del entorno. Una ventaja clave de la genómica selección sobre la selección fenotípica tradicional en el caso de modelos multiambientales, es la capacidad de los modelos para evaluar los efectos de los marcadores y los efectos de los marcadores por entorno interacciones y, en última instancia, aumentar la precisión de la predicción [18, 104]. con la integración de modelos de crecimiento de cultivos en el marco de predicción genómica, la respuesta del genotipo a las variaciones ambientales se modela lo que permite la predicción del rendimiento de candidatos de selección para entornos no probados [105]. Este enfoque es muy prometedor. para la mejora del arroz porque tiene más en cuenta GxE. Sin embargo, el uso rutinario de los modelos de crecimiento de cultivos en los programas de mejoramiento requieren una inversión sustancial en términos de datos adquisición y análisis y, por lo tanto, será interesante para sistemas específicos de arroz propensos a limitaciones ambientales.

*F. Las diferencias entre los modelos de predicción genómica son marginales.* La mayoría de los estudios

la comparación de modelos estadísticos para la predicción genómica encontró pequeñas o ninguna diferencia entre ellos en términos de precisión [20]. En general, ninguno de los modelos es consistentemente mejor para todos los rasgos o métodos de validación. GBLUP se suele utilizar como referencia debido a su sencillez, versatilidad para incluir diferentes tipos de información y robustez para arquitectura de rasgos diferentes. Modelos bayesianos (como B-LASSO, BayesB o BayesC) o RKHS puede funcionar mejor cuando se trata de rasgos influenciados por genes de efecto grande (como como tiempo de floración o resistencia al añublo). Los pocos estudios que utilizaron el aprendizaje automático métodos (como ANN o SVM) informaron resultados decepcionantes con muy variable incluso con una optimización de los parámetros [68, 85]. Trabajo adicional en este probablemente se necesita dirección para concluir sobre el interés de estos métodos para la rutina predicción genómica.

### 2.2.2 Limitaciones actuales

A pesar de la cantidad y diversidad de estudios, todavía hay algunos puntos que no están bien cubiertos en la literatura sobre el arroz. Dependiendo del contexto, pueden ser limitantes para aprovechar todo el potencial de la selección genómica.

- un. *La precisión por sí sola no es suficiente para evaluar la eficacia de la predicción genómica.* Casi todos los estudios basaron su evaluación de la selección genómica en la precisión de la predicciones Aunque la precisión es un factor importante para evaluar el modelo de predicción eficiencia, no informa sobre qué individuos son seleccionados *en fin* por los diferentes métodos. El diferencial de selección realizado probablemente sería una mejor métrica para comparar diferentes enfoques de predicción genómica a medida que los mejoradores consideran conjuntamente varios rasgos para avanzar en el material, lo que hace que la evaluación de los rasgos por separado sea menos relevante. Finalmente, como acertadamente señalan Bassi *et al.* [106], el fenotipo también es solo un predictor del verdadero valor genético y tiene una varianza de error como un GEBV.
- B. *La precisión de la predicción dentro de la familia no se tiene suficientemente en cuenta.* Ningún estudio sobre el arroz ha analizado en detalle la precisión de la predicción dentro de la familia utilizando múltiples familias o información de los padres como el conjunto de entrenamiento. En efecto, salvo el caso concreto de los estudios que utilizan una familia biparental, los informes sobre la precisión dentro de la familia son escasos. Esto es manifiesto también en la literatura híbrida donde la mayoría de los artículos se enfocan en predecir combinaciones híbridas y no intentar estimar la capacidad de combinación general entre una cohorte de nuevos machos o hembras. Sin embargo, este es un punto clave cuando se trata de la implementación de la predicción genómica ya que una mayor precisión dentro de la familia puede ayudar a aumentar la tasa de ganancia genética mientras se equilibra el nivel de consanguinidad en la población. Las diferencias entre cruces se predicen mejor tanto dentro como fuera entre las variaciones familiares son capturadas por el modelo [107, 108].

C. *Se descuidaron la calidad del grano o los rasgos de resistencia a enfermedades.* Ningún estudio relacionado con la nutrición

valor del grano pulido (contenido de zinc, índice glucémico, ...) se publicaron hasta la fecha.

Solo un estudio evaluó el potencial de la predicción genómica para ayudar a disminuir el nivel

de arsénico en el grano mediante mejoramiento genético [72]. En cuanto a la resistencia a enfermedades, el único estudio

de Huang *et al.* [75] informó precisiones que van desde 0,15 a 0,72 para la predicción de

resistencia a varios aislamientos de *Magnaporthe oryzae* (blast). Para la resistencia a las enfermedades, el arroz

Los genetistas se enfocan principalmente en los genes principales, pero también es importante enfocarse en la variación cuantitativa.

importante abordar preocupaciones como eludir las resistencias. Para el valor nutricional del grano,

las correlaciones negativas entre los rasgos se pueden abordar mejor utilizando la genómica de múltiples rasgos

predicción.

D. *La implementación en los programas de mejoramiento es secundaria.* Si bien es claro que la base

El objetivo de todos los estudios es mejorar nuestro conocimiento de la predicción genómica para optimizar

estrategias de mejoramiento, pocos ubican sus hallazgos en un caso concreto de mejoramiento

programa. Por ejemplo, Spindel *et al.* [45] propusieron integrar la predicción genómica en

una tubería de mejoramiento de arroz irrigado y discutió las ventajas y limitaciones de tal

un esquema. Sin embargo, para la mayoría de los estudios que trabajan en el mejoramiento de germoplasma (ver Tabla

1) este no es el caso. Por lo tanto, los resultados siguen siendo más teóricos que prácticos, ya que

tales análisis son importantes para justificar las inversiones en selección genómica y para

comprender las barreras potenciales para su implementación.

### 3. Integración de la predicción genómica en los programas de mejoramiento del arroz: aspectos clave

Los puntos de entrada para la selección genómica en un programa de mejoramiento de arroz variarán dependiendo del

objetivos del programa, la estrategia de mejoramiento en el lugar, la genética y/o ambiental

restricciones que el mejorador tiene que tener en cuenta, y el costo de genotipado y fenotipado del

rasgos bajo selección. Sin embargo, hay requisitos previos clave para evaluar antes de integrar un

preparación del programa de mejoramiento para implementar la predicción genómica. A falta de lo esencial componentes tales como (a) objetivos claros, (b) manejo meticuloso de datos, (c) operaciones, (d) fenotipado efectivo y (e) selección basada en BLUP, la aplicación de predicciones genómicas es extremadamente limitada [4]. Ejecución de predicciones genómicas utilizando datos de reproducción o conjuntos de entrenamiento especialmente diseñados son útiles para establecer la capacidad de referencia para hacer predicciones, pero integrar la tecnología en un programa de mejoramiento existente puede ser un desafío. Cría Los programas representan proyectos plurianuales que gestionan cohortes superpuestas de germoplasma, por lo que cambiar la estrategia a menudo se hace paso a paso para no interrumpir el desarrollo del producto proceso. El propósito de esta sección es proporcionar pautas sobre elementos importantes para considerar antes de implementar una estrategia de selección genómica en un programa de mejoramiento de arroz.

### 3.1 Mapear la estrategia de mejoramiento

El principal valor de la predicción genómica radica en su uso en la toma de decisiones para seleccionar eficientemente material de mejoramiento en una o varias etapas del esquema de mejoramiento. Por lo tanto, una clara La comprensión de la estrategia de mejoramiento y sus diferentes componentes es la base para una eficiente integración de la predicción genómica. A menudo, el esquema de crianza reside en la cabeza del criador, y traducir este conocimiento en un marco estructurado es un paso obligatorio para diseñar cuidadosamente esquemas alternativos [109]. La predicción genómica es una inversión a largo plazo para el programa de mejoramiento y la transición directa a una estrategia de selección genómica óptima no es siempre posible Por lo tanto, el equipo de mejoramiento debe elaborar un plan de transición y expertos con el fin de definir pasos claros para lograr los objetivos. Este aspecto no suele ser informado en la literatura sobre predicción genómica, ya que se trata de información más técnica en cuanto al esquema de crianza. En arroz, sólo un estudio situó los resultados en el marco de un programa de reproducción y detalló el uso de la predicción genómica y sus impactos potenciales [45]. Sin embargo, como se muestra en el trigo, este paso de caracterización del esquema de mejoramiento es esencial para la



integración o la optimización de la selección genómica en base a los conocimientos adquiridos durante la  
últimos años [106, 110].

Los esquemas de selección genómica óptimos no suelen ser simples evoluciones de la selección actual.

esquema. La mayoría de los esquemas de mejoramiento convencionales en arroz y cultivos autopolinizados en

En general, se basa en la cría de pedigrí [25], pero la selección genómica se adapta mejor a la reproducción recurrente .

esquemas de selección basados en cruces élite por élite para mejorar rasgos complejos. De hecho, un pozo

programa de mejoramiento estructurado donde el germoplasma élite ha sido claramente identificado y con un

tamaño de población efectivo pequeño ( $N_e$   $\approx$  40) es más probable que se beneficie del uso de genómica

predicción debido a un mayor desequilibrio de ligamiento entre marcadores y QTL, baja o ausencia de

estructura poblacional y mayor parentesco entre genotipos. Además, varios importantes

se necesitan cambios para aprovechar al máximo las predicciones genómicas: reducir el tiempo de ciclo, crear un entrenamiento

establecer, almacenar/utilizar datos fenotípicos y genotípicos, reasignar presupuesto y personal [106, 111].

Comprender las interconexiones entre estos cambios y cómo afectarán a la

La secuencia de operaciones en curso permite anticipar obstáculos potenciales.

#### *Recomendaciones clave:*

un. Definir claramente la estrategia de mejoramiento actual y sus objetivos.

B. Planificar la integración de la predicción genómica como una inversión a largo plazo con un claro  
mapa vial.

C. Utilice la selección recurrente en la población de élite para maximizar el potencial de la genómica  
predicción

## 3.2 Reducir el tiempo de ciclo

Un aspecto interesante de la selección genómica es que ha llevado a un mayor enfoque en la fundamentos del mejoramiento en la comunidad de fitomejoradores [112]. El concepto de respuesta a la selección capturada en la ecuación de Breeder es quizás el mejor ejemplo [4, 109]. Entre la parámetros de la ecuación, el intervalo de generación (o tiempo de ciclo) es el más fácil de entender y para jugar. Como destaca Meuwissen *et al.* [11] en su artículo seminal, el uso de predicciones genómicas pueden aumentar en gran medida la tasa de ganancia genética al reducir el tiempo del ciclo: *"Era llegó a la conclusión de que la selección de valores genéticos predichos a partir de marcadores podría aumentar sustancialmente la tasa de ganancia genética en animales y plantas, especialmente si se combina con reproducción técnicas para acortar el intervalo generacional"*. Esta conclusión fue confirmada quince años después. por el primer informe del impacto de la selección genómica en la tasa de ganancia genética en el ganado lechero [113]. Los autores encontraron una reducción drástica en el intervalo generacional relacionado con un fuerte aumento en la tasa de ganancia genética de los rasgos de rendimiento (50 - 100%). En el fitomejoramiento, los métodos para reducir el tiempo del ciclo (independientemente del uso de la selección genómica) se han estudiado para varias décadas ahora [114, 115]. El avance de generación rápida (RGA) o doble haploide son probablemente el más común en las especies de cultivo, incluso si se han utilizado enfoques más modernos propuesto últimamente [116, 117]. En arroz, RGA ha recuperado interés recientemente ya que es rentable forma de arreglar rápidamente el material (típicamente de F2 a F6 en un año) para su evaluación en réplicas ensayos [118]. Esto se puede realizar en invernaderos, casas de malla o en el campo dependiendo de la recursos disponibles. Para los criadores que trabajan en un esquema de crianza de pedigrí clásico, el uso de RGA podría ser un primer paso hacia la implementación de la selección genómica [119]. para la cría programas que ya implementan RGA o métodos similares para reducir el tiempo de ciclo, genómica la selección puede ayudar aún más a acortar el ciclo de reproducción. Sin embargo, esto requiere una genómica modelo de predicción que puede predecir de manera eficiente el valor genético de la próxima generación (descendencia). Por lo tanto, un conjunto de entrenamiento basado en material de uno o varios ciclos anteriores tiene que ser

constituida antes de implementar este tipo de régimen. Este es también el caso de los más agresivos. estrategias basadas en selección recurrente que apuntan a la recombinación de material no fijo (S0) seleccionado basándose únicamente en los valores predichos. En ese tipo de esquema, la parte de mejoramiento de la población es parcialmente desacoplado de la parte de desarrollo del producto que permite un año o incluso menos ciclo reproductivo [120, 121]. Por el momento, solo los estudios de simulación han reportado este tipo de ya que varios desafíos técnicos tienen que ser resueltos antes de la implementación. De hecho, un La reducción drástica del tiempo del ciclo de reproducción puede conducir a la superposición de actividades entre diferentes ciclos durante el período de transición que pueden interrumpir los ciclos en curso o aumentar sustancialmente la carga de trabajo

#### *Recomendaciones clave:*

- un. Use la predicción genómica junto con métodos robustos para producir líneas endogámicas (por ejemplo, rápido avance de generación) para reducir efectivamente el tiempo de ciclo.
- B. Tener en cuenta la restricción técnica asociada con la reducción del tiempo de ciclo en el hoja de ruta de predicción genómica.

### **3.3 Diseñar el conjunto de entrenamiento**

Una vez que se ha definido el punto de entrada de la predicción genómica en el esquema de mejoramiento, el El diseño del conjunto de entrenamiento es el primer paso hacia la implementación de la selección genómica. Tres Se deben tomar decisiones importantes con respecto al conjunto de entrenamiento: su composición y tamaño, su fenotipado y su genotipado. El criador debe encontrar un equilibrio entre estos tres aspectos. con el fin de optimizar el conjunto de entrenamiento según los recursos disponibles. Una forma sencilla para la mayoría programas de mejoramiento para comenzar es comenzar a genotipificar cada línea que ingresa a la prueba de rendimiento. A partir de ahí, esos conjuntos de datos se pueden optimizar empíricamente para aumentar la precisión de la predicción.

Es bien sabido que la precisión aumenta con el tamaño del conjunto de entrenamiento. teórico [122–124] y los estudios empíricos [65, 125, 126] sugieren que el tamaño del conjunto de entrenamiento debe maximizarse cuando se trata de rasgos complejos. Sin embargo, grandes conjuntos de entrenamiento no siempre son factibles principalmente debido a los costos de genotipado y fenotipado. Se desarrollaron varios métodos para optimizar la composición del conjunto de entrenamiento para lograr altas precisiones mientras se mantiene el tamaño a un número manejable [39, 42, 46, 69, 73, 78, 127–129]. Todos estos métodos utilizan el aditivo relaciones genéticas (generalmente basadas en datos de marcadores) para muestrear de manera óptima un conjunto de representantes genotipos. Un aspecto clave de la optimización del conjunto de entrenamiento es la definición de la predicción conjunto (candidatos de selección). De hecho, las estrechas relaciones genéticas entre el conjunto de entrenamiento y el los candidatos de selección son clave para maximizar la precisión de la predicción [130, 131]. Por lo tanto, la mayoría de los métodos de optimización están considerando conjuntamente los genotipos que compondrán el entrenamiento y los conjuntos pronosticados para calcular directamente los criterios basados en la relación (el promedio de los coeficientes de relación entre el conjunto de entrenamiento y el conjunto predicho [128, 132]) o para estimar criterios basados en la teoría de modelos mixtos (la varianza del error de predicción, el coeficiente de determinación o la precisión esperada [39, 78, 127]). En los casos en que la formación y la los conjuntos pronosticados provienen de la misma población (por ejemplo, candidatos de selección de la misma cohorte) o la información sobre los individuos predichos aún no está disponible (por ejemplo, descendencia), optimización Se han desarrollado métodos para minimizar las relaciones genéticas entre los individuos de la conjunto de entrenamiento [46, 73]. Dependiendo de la disponibilidad de datos y de los objetivos de predicción, la el criador puede elegir entre estos métodos de optimización para dar forma al conjunto de entrenamiento y actualizarlo cuando es necesario predecir candidatos de selección de un nuevo ciclo.

La optimización de la composición del conjunto de entrenamiento tiene que hacerse en conjunto con la estrategia de fenotipado. En la mayoría de los casos, los candidatos de selección que se utilizarán para actualizar la El modelo de predicción se evalúa para rasgos clave en MET para estimar G×E. Dado que el número total de las parcelas disponibles para la evaluación son casi fijas, el criador necesita equilibrar la población

tamaño con el nivel de replicación (dentro y entre entornos). Clásicamente, el nivel de la replicación aumenta durante el ciclo de reproducción para dedicar más recursos a un número menor de líneas más prometedoras en las etapas finales. En el contexto de la selección genómica donde el la unidad de evaluación son los alelos en lugar de los individuos, aumentando el tamaño del conjunto de entrenamiento mientras que la disminución del nivel de replicación tiende a aumentar la precisión de la predicción [133, 134].

El tamaño típico de una población de entrenamiento (150 - 300) para fenotipar en un clásico totalmente

Por lo tanto, el experimento replicado se puede multiplicar por 1,5 a 3 con pocas pruebas. Sin embargo lo es aconsejable tener un nivel suficiente de replicación dentro y entre entornos para: i) mantener

repetibilidad, especialmente para rasgos de baja heredabilidad, ii) evaluar el nivel de GxE y iii) evitar modelo problemas de convergencia con muy pocas réplicas. La limitación de la replicación usando pruebas escasas

Los enfoques también pueden ser una buena oportunidad cuando la disponibilidad de semillas es una limitación.

Finalmente, la tecnología utilizada para genotipar los conjuntos de entrenamiento y predichos necesita ser cuidadosamente considerado con el fin de capturar de manera eficiente distintos alelos QTL, así como la relación general en la población. Varios factores entran en juego al elegir o desarrollar el apropiado

tecnología de genotipado: costo, tipo de marcadores, densidad, informatividad en la población objetivo,

tasa de reproducibilidad, etc. En el caso de aplicar predicción genómica, una buena caracterización de

la diversidad genética manejada por el programa de mejoramiento es fundamental para determinar el marcador

densidad necesaria para lograr una precisión de predicción óptima. Se ha demostrado usando ambos

simulaciones deterministas [13, 135] y estocásticas [136] de que la densidad del marcador tiene que aumentar

cuando el tamaño efectivo de la población aumenta para mantener la precisión [135–137]. Sin embargo, la mayoría

estudios empíricos en arroz encontraron que la precisión alcanza una meseta cuando la densidad del marcador

va más allá de 2 a 5 marcadores por centiMorgan para programas de mejoramiento con una población efectiva

Talla inferior a 50.

*Recomendaciones clave:*

- un. Maximice la relación entre el entrenamiento y los conjuntos predichos cuando sea posible.
- B. Use pruebas escasas para el fenotipado con el fin de equilibrar el tamaño del conjunto de entrenamiento y el nivel de recursos disponibles.
- C. Evite usar un conjunto de entrenamiento de un canal de reproducción para predecir los candidatos de otra tubería de reproducción.

### 3.4 Generar e integrar datos de buena calidad

Como se destacó anteriormente, la adquisición y gestión de datos son componentes esenciales de un programa de cría. Todas las decisiones de avance se toman en base a los datos registrados de múltiples fuentes (campo, laboratorio, proveedor de servicios, etc.). Cuidadosa gestión de datos desde la semilla hasta el fenotipo y/o el genotipo deben estar en su lugar para garantizar la precisión. El uso de lo digital

Las herramientas de recopilación de datos son una forma clave de reducir tanto como sea posible los errores que pueden ser perpetuado durante el proceso de recopilación de datos. Al respecto, se ha demostrado con datos simulados que incluso un pequeño porcentaje de errores graves (0,1% o 1%) en fenotípica los registros pueden reducir severamente la respuesta a la selección [138]. Conclusiones similares también fueron encontrado cuando los errores están presentes en los registros genealógicos [139]. Además de datos precisos, sólidos y Se necesitan canalizaciones de análisis diseñadas apropiadamente para curar los datos y convertirlos en inteligencia interpretable. La predicción genómica agrega una capa adicional de complejidad en comparación a la selección tradicional asistida por marcadores en que puede requerir la integración de diferentes tipos de datos (fenotipos, genotipos, pedigrí y/o datos meteorológicos) recopilados durante varios años para ser útil. La consistencia del tipo y formato de datos y la estabilidad de las estructuras de datos a lo largo del tiempo es clave aspecto para aprovechar todo el poder de los datos históricos de reproducción para entrenar y actualizar continuamente modelos de predicción genómica [140].

Para ayudar a los criadores con la gestión de datos, las soluciones de software como Breeding

Sistema de Gestión (<https://bmspro.io>), Breeding4Results (B4R)

(<https://riceinfo.atlassian.net/wiki/spaces/ABOUT/pages/326172737/Breeding4Results+B4R>),

Breedbase (<https://breedbase.org>), o Gestión de datos genómicos de GOBii

(<https://gobiiproject.atlassian.net/wiki/spaces/GD/overview>) están disponibles y se utilizan en diferentes

organismos públicos. A pesar de los esfuerzos significativos para desarrollar canales de análisis (como el

RiceGalaxy, <https://galaxy.irri.org>, [141]) y el proyecto Breeding API (<https://brapi.org>)

diseñado para permitir la interoperabilidad entre bases de datos de fitomejoramiento, no eficiente de extremo a extremo

La solución está disponible públicamente para realizar la predicción genómica en el contexto de un mejoramiento aplicado.

programa. De hecho, varias limitaciones están presentes entre el software disponible para implementar

predicción genómica, incluida la falta de vínculos directos entre los datos genotípicos y fenotípicos,

capacidad analítica limitada de múltiples entornos o múltiples rasgos, sin posibilidad de integrar el dominio

o efectos epistáticos en un modelo de predicción, y ninguna integración significativa de los datos meteorológicos en

una canalización analítica. Por lo tanto, la mayoría de los programas públicos de mejoramiento extraen el fenotipo

y datos genotípicos de sus respectivos software de gestión de datos y utilizar análisis ad hoc

tuberías para ejecutar modelos de predicción genómica. Con suerte, proyectos como Breeding API o el

Sistema de mejoramiento empresarial (<https://ebs.excellenceinbreeding.org>) ofrecerá estas posibilidades en

el futuro cercano dentro de un marco coherente diseñado para permitir programas de mejoramiento aplicados.

#### *Recomendaciones clave:*

un. Utilice sistemas de recopilación de datos digitales cuando sea posible.

B. Trabaje con sistemas de gestión de datos y rutinas de análisis eficientes para genómica

predicción (GBLUP, RR-BLUP).

C. Use estructuras de datos genotípicos y fenotípicos consistentes a lo largo de los años para facilitar los datos

integración.

### 3.5 Tener en cuenta los costes

La integración de la selección genómica en un programa de mejoramiento es una inversión a largo plazo que debe traducirse en una mejor tasa de ganancia genética que valga la pena implementar. Incluso si las ventajas de utilizando la selección genómica son claros, el esquema de mejoramiento óptimo en relación con genética, operacional, y las restricciones de costos no son fáciles de identificar. Después de establecer una visión de lo que es óptimo, la necesidad de convertirse a esta nueva estrategia de una manera económica es probablemente la limitación más importante para el fortalecimiento de los modernos programas de mejoramiento. Sin embargo, hay varias palancas que pueden utilizarse para liberar recursos en un programa que tenga como objetivo desplegar completamente la selección genómica.

Las primeras palancas están relacionadas con el fenotipado. Gracias a la predicción genómica, algunos fenotipos los pasos pueden reducirse o incluso eliminarse ahorrando los costos relacionados *de facto*. De hecho, este es uno de las principales ventajas de la predicción genómica que, con las estructuras de datos adecuadas, permite tanto para una reducción del tiempo de ciclo como de los costos de fenotipado [111]. Los costos de fenotipado y el potencial para reemplazar una actividad de fenotipado con una predicción debe evaluarse cuidadosamente al planificar la integración de la predicción genómica, ya que a veces puede requerir una modificación del esquema de crianza. Un ejemplo clave de esto es el ahorro de costos incurrido al hacer la transición del programa tradicional de cría de pedigrí donde la selección que ocurre durante la fijación Los pasos (F2 a F5) se pueden retrasar hasta que se hayan extraído las líneas endogámicas sustituyendo un vivero de pedigrí basado en el campo con un método RGA basado en SSD mucho más económico y rápido. los los ahorros de costos realizados a este nivel pueden cubrir fácilmente el costo de la genotipificación, ya que el material avanzado a través de RGA es mucho menos costoso (alrededor de 1 dólar estadounidense por líneas F5/F6) [119]. Organizaciones Sin embargo, debe buscar estrategias presupuestarias multianuales para acomodar los costos fijos que pueden incurrir si las instalaciones de invernadero existentes no pueden aprovecharse para esta actividad. Capital inicial las inversiones a menudo se pueden pagar mediante costos operativos reducidos durante varios años.

Además, las organizaciones deben tener en cuenta la financiación adicional que podría generarse debido a



el aumento de la ganancia genética que acompañará a un acortamiento del ciclo de cría y una mejora en la precisión de la selección.

Otra forma directa de recuperar costos es usar la predicción genómica para reducir el volumen de un costoso ejercicio de fenotipado [73, 142]. Esto se puede hacer mediante el fenotipado selectivo de un subconjunto cuidadosamente elegido de un ensayo para rasgos costosos como bioquímica de granos u otra publicación recolectar rasgos y usar los ahorros de costos para pagar la toma de huellas dactilares de ADN. Adicionalmente, desarrollando un índice de rasgos correlacionados de alto rendimiento que puede ser menos costoso de medir u ofrecer un mayor rendimiento en comparación con el rasgo objetivo puede disminuir el costo de fenotipado y ofrecer precisión similar. En ese contexto, la predicción genómica de múltiples rasgos ofrece un marco ideal para integrar rasgos correlacionados para maximizar la precisión de la predicción [143].

Las segundas palancas están relacionadas con el genotipado. En un programa de fitomejoramiento, la elección del La tecnología de genotipado para caracterizar el germoplasma de mejoramiento (conjuntos de entrenamiento y predicción) es impulsado principalmente por el costo del genotipado por muestra (y no muy bien captado por el costo por muestra). punto de datos) [144]. De hecho, el costo por muestra con las herramientas disponibles (genotipificación por secuenciación o arreglos de SNP fijos) es a menudo demasiado alto para ser usado de manera rutinaria en un programa de mejoramiento público. en pequeño a programas de mejoramiento de tamaño mediano, el costo por muestra debe ser de alrededor de 10 dólares estadounidenses o menos en para evaluar un número suficiente de individuos. En ese rango de precios, el número de loci que se puede apuntar actualmente es alrededor de 1,000, - 5,000 SNP. Una opción para mantener los costos bajos en el a largo plazo es diseñar un ensayo de genotipado personalizado con SNP seleccionados para ser específicamente informativo en la población reproductora objetivo. Esta sería una opción más económica que GBS o público matrices fijas y permiten una mayor densidad de contenido de información en el conjunto de datos de genotipo. Una costumbre El panel SNP tiene el beneficio adicional de estudiar potencialmente marcadores de rasgos específicos de relevancia para un programa de reproducción además de los marcadores de todo el genoma incluidos en el conjunto; permitiendo así perfiles QTL más extensos de líneas para alelos conocidos que no necesariamente se priorizan para MAS.

De hecho, dependiendo de la capacidad del proveedor de servicios de genotipado, no es descabellado

ahorrar costos de muestreo y extracción de ADN al combinar MAS y toma de huellas dactilares de manera que el cohorte se examina con unos pocos marcadores destinados a MAS, luego para tener el ADN de seleccionado líneas reorganizadas en una nueva placa para la toma de huellas dactilares de todo el genoma.

También es posible lograr un bajo costo de genotipado mediante el uso de genotipado de baja cobertura. secuenciación [145]. Dada la limitación del genotipado por secuenciación cuando la profundidad de secuenciación se reduce (alta tasa de datos faltantes, alta tasa de error para loci heterocigotos), este enfoque no capturar loci heterocigotos de manera eficiente y debe usarse para genotipificar líneas fijas, junto con un marco de imputación eficiente basado en datos de secuencia de alta calidad de líneas ancestrales en el árbol genealógico. Por lo tanto, esto requiere experiencia en bioinformática y acceso a alto rendimiento. recursos de cómputo.

#### *Recomendaciones clave:*

- un. Considere reducir el número de pasos de fenotipado, solo fenotipando un subconjunto de un ensayo, o el uso de características correlacionadas más baratas o de mayor rendimiento.
- B. Diseñar una plataforma de genotipado con un conjunto de marcadores seleccionados específicamente para el germoplasma administrado en el programa de mejoramiento e implementarlo en un proveedor de servicios.

## 4. Un ejemplo del programa de mejoramiento del IRRI para sistemas irrigados

Aquí damos un ejemplo práctico de la integración y el uso de predicciones genómicas en un programa de mejoramiento de arroz. El programa de mejoramiento recientemente rediseñado para sistemas irrigados en el IRRI ofrece un contexto ideal para comprender los elementos clave de un programa de mejoramiento aplicado utilizando predicciones genómicas [146, 147]. De hecho, con su mandato global del Sudeste Asiático, Asia Meridional y África Oriental como las principales áreas de intervención, representa la derivación directa de la los primeros esfuerzos de reproducción que resultaron en la Revolución Verde en Asia. Como tal, es lo mejor

representación posible de un esfuerzo por producir materiales que combinen alto potencial de rendimiento y adaptación a diversas condiciones ambientales.

## 4.1 La transición del mejoramiento genético a la selección genómica recurrente

Las aplicaciones de la selección genómica al programa de mejoramiento del IRRI se dieron en dos grandes

categorías: dentro de las predicciones de cohorte (hermanos completos y medio hermanos que predicen otros hermanos completos y medio hermanos) para

optimizar nuestra estrategia de prueba y predicciones de cohortes (abuelas y madres)

predecir hijas y nietas) para acelerar nuestros ciclos de reproducción, los cuales

cambios necesarios en la estrategia de cría. En primer lugar, ambas aplicaciones requerían la

implementación rentable de una tecnología de genotipado que permitió la toma de huellas dactilares de rutina

del material de cría. Este conjunto de marcadores (conocido como el panel de amplicón 1k-RiCA [67]) tenía

recientemente desarrollado y poblado con marcadores que fueron específicamente informativos en nuestro

germoplasma. La tecnología de genotipado de matriz fija disponible públicamente no habría servido para esto

propósito, ya que muchos de los marcadores en estas matrices se eligen para diferenciar el germoplasma

a nivel mundial [148] y, a menudo, eran muy costosos con relativamente pocos (o peor, sesgados)

polimorfismos.

Con el panel marcador en su lugar y desplegado en un proveedor de servicios, en el plazo inmediato, el

La aplicación más útil de la selección genómica fue permitir que se hicieran selecciones basadas en

rendimiento en los entornos de destino en lugar de depender de una respuesta correlacionada a

selección con entornos filipinos (donde se encuentra la sede del IRRI). El programa

como tiene los recursos actuales, genera una prueba de rendimiento de etapa 1 de aproximadamente 2,000 líneas nuevas cada una

año. Como todos los ensayos de rendimiento del IRRI son realizados por socios nacionales de investigación agrícola, el

capacidad para probar 2000 líneas en pruebas de rendimiento en múltiples ubicaciones en África, el sur de Asia y el sudeste asiático

era extremadamente limitado. Hasta este punto, se seleccionó el material de cría de primera generación.

basado en el desempeño en Filipinas y se envió una pequeña cantidad de líneas avanzadas al

ubicaciones regionales para pruebas y evaluación (Figura 2). Selección genómica usando full y half sibs se empleó para permitir la selección directa basada en el entorno de destino y evitar necesidad de confiar en la selección indirecta. Al seleccionar un subconjunto optimizado de la cohorte y enviar para ser probado en la región de interés, los datos de fenotipo de la región específica de interés podrían utilizarse para predecir el desempeño de la cohorte restante en esa región. De esta forma, la totalidad cohorte se prueba en alguna parte, pero ningún individuo se prueba en todas partes y, por lo tanto, un avance de líneas superiores se pueden enviar a socios que se adaptan a sus condiciones únicas. Para hacer esto, sin embargo, requirió que se identificaran los fondos para tomar las huellas dactilares de la cohorte completa de aproximadamente 2,000 nuevas líneas todos los años. Para que esta forma de selección genómica sea neutral en cuanto a costos, se observó que la La estrategia de prueba en Filipinas fue probar líneas durante tres años (Figura 2, esquema anterior). Por eliminando la fase de prueba intermedia y seleccionando un conjunto de líneas específico de la región para avanzar pruebas, se recuperaron fondos suficientes para cubrir el costo de la toma de huellas dactilares.

La aplicación de predicción genómica con más valor a largo plazo para el programa fue permitir a través de predicciones de cohorte para que las líneas superiores en cada región puedan reciclarse de nuevo en el tubería de reproducción antes de las pruebas regionales, y así acelerar el ciclo de reproducción (Figura 2, esquema futuro). Sin embargo, este tipo de predicción requiere un conjunto de datos de varios años más sólido. que consiste en datos de fenotipo regional en líneas ancestrales, como datos de fenotipo de completo y medio los hermanos de los candidatos emergentes no estarían disponibles en el momento en que la predicción necesita ser hecha. Con la primera aplicación de predicción genómica en su lugar, el programa ahora está bien posicionado para comenzar a generar conjuntos de datos de varios años con observaciones fenotípicas específicas de la región necesarios para predecir nuevos padres. Sin embargo, para hacer posible este tipo de predicción, un era necesario implementar la manipulación dirigida de la estrategia de cruce. El más importante La decisión que toma un criador es seleccionar y cruzar progenitores sobre la base de los valores genéticos para rasgos relevantes. Como esta métrica no se calculaba de manera rutinaria en el IRRI, nuestro primer paso fue recopilar nuestros datos históricos juntos en un solo modelo y generar las mejores estimaciones posibles para

valores genéticos y confiabilidad para el rendimiento, la madurez y la altura de la planta. Crianza de valores para otros rasgos importantes como la calidad del grano, la resistencia a enfermedades y otros rasgos agronómicos no fueron recolectados de forma rutinaria o en suficientes lugares para proporcionar estimaciones significativas de reproducción valor. Este proceso se aceleró sustancialmente debido a los esfuerzos realizados para migrar datos a el sistema de gestión de datos B4R. Como los datos de huellas dactilares de ADN no estaban disponibles en la gran mayoría de nuestras líneas históricas, los datos de pedigrí almacenados en el sistema de gestión de genealogía fueron utilizado para estimar los coeficientes de parentesco. Esta evaluación plurianual de nuestros datos históricos permitió la identificación de un conjunto central único de líneas con valores genéticos altos y confiables para el rendimiento, que formaría la base de futuros esfuerzos de mejoramiento y caracterización de germoplasma. Una vez identificado, este conjunto de líneas de alto valor genético se tomaron las huellas dactilares y los datos luego se se usa para estimar el tamaño efectivo de la población y se usa para estimar las frecuencias de los principales genes para otras características (como el contenido de amilosa o la resistencia al añublo). Estas métricas serían utilizado para guiar las estrategias de selección entre la progenie y evaluar el riesgo/beneficio de introducir nueva genética en el programa.

Este paso, aunque no motivado específicamente por la selección genómica, fue de vital importancia porque junto con el desarrollo y la caracterización del germoplasma central vino un compromiso del programa de cruzar principalmente dentro de este nuevo acervo genético para impulsar la ganancia genética. Esta relación entre generaciones (y la aversión a la introducción frecuente de nuevo germoplasma en el programa) crea una continuidad genética sobre cohortes multigeneracionales que permite la capacidad de usar datos de fenotipos de antepasados para predecir el desempeño de recién creados descendientes. En correspondencia con esa relación fue el desarrollo de reglas de negocio para cruce y desarrollo de la población. Estas reglas aseguran que los nuevos cruces generados por el El programa de reproducción maximizó la variación genética en la próxima generación en la medida de lo posible. También permitieron generar un número suficiente de hermanos completos y medio hermanos en cada cohorte, a partir de la cual se podría obtener el poder predictivo. Entre éstas, las reglas de negocio incluían una

compromiso de cruzar con las líneas de las cohortes más recientes siempre que sea posible (en lugar de líneas liberadas más antiguas), impidiendo el uso de cualquier línea en más del 10% de los cruces para evitar el cuello de botella la variación, la evitación completa del sub-revestimiento para que cada planta F2 genera una línea F6 única, y asegurando que se crearon suficientes líneas fijas nuevas de cada cruce entró en el ensayo de rendimiento de la etapa 1 de tal manera que había una probabilidad razonable de identificar una nueva línea que fue al menos una desviación estándar mejor que el rendimiento promedio de la cruz.

Con estas dos aplicaciones de predicción genómica en marcha, el programa pasó de un largo vivero de pedigrí de ciclo a una estrategia de mejoramiento habilitada por genómica de ciclo rápido. Esta estrategia involucró hacer cruces y establecer objetivos de tamaño de población de acuerdo con negocios predefinidos reglas, generando nuevas líneas a través de enfoques RGA, empleando MAS después de la fijación de la línea, y uso de cosechas a granel de las filas principales seleccionadas para crear semillas para enviar a ubicaciones regionales para las pruebas. Las predicciones de toda la cohorte en todas las regiones asegurarían que cada línea tuviera ya sea una observación o una predicción en cada región, de la cual un conjunto central de región superior se identificaron líneas específicas y se enviaron a los socios para la evaluación de la prueba de rendimiento de la etapa 2 y pruebas. Como datos acumulados en las regiones sobre cohortes de líneas, y como la progenie y gran la progenie del conjunto básico original de líneas comienza a llenar la tubería, la capacidad de predecir el desempeño regional entre las cohortes crecerá hasta que haya suficientes datos disponibles para permitir para la identificación de nuevos progenitores antes de la prueba de rendimiento de la etapa 1.

## 4.2 Descripción de los esquemas de mejoramiento e integración de la predicción genómica

El mapeo del esquema de cría es un componente clave para el uso óptimo de la cría. recursos del programa y comprender dónde podrían estar los puntos de entrada para la selección genómica metido. La estrategia de reproducción actual resumida en la Figura 2 se inició en 2017 en el IRRI en para reducir el tiempo de ciclo y optimizar las evaluaciones multientorno gracias a la introducción de predicción genómica. En esta estrategia, la mayoría de las actividades tienen lugar en la sede del IRRI en

las Filipinas. El primer año se hacen los cruces (80-100) y se validan las plantas F1 usando marcadores SNP dedicados. El segundo año las familias segregantes pasan por SSD de F2 a F6 a través de RGA. En esa etapa, se adelantan de 7.500 a 10.000 líneas: esto corresponde a 200 - 400 líneas por cruz. Los tamaños de población para cada cruce se determinan con base en el anticipado segregación de genes principales. El tercer año, las líneas se evalúan en el campo en hileras de panículas para el aumento de semillas y para la evaluación de la uniformidad, la arquitectura de la planta y la madurez. En el Al mismo tiempo, las líneas se genotipan para la selección asistida por marcadores para los principales loci priorizados para cada conducto de cría. Estos incluyen el gen ceroso para el contenido de amilosa y una serie de genes de resistencia a enfermedades para las principales plagas y enfermedades (añublo, tizón bacteriano de la hoja, ...) [10]. Los La segunda temporada del tercer año se dedica a la preparación de las semillas para ser enviadas en el regiones. El cuarto año, las líneas avanzaron en base a MAS y selección de hileras de cabecera (1,500 - 2000) se genotipan utilizando una plataforma de baja densidad con menos de 1000 marcadores SNP [67]. Los Las mismas líneas también se evalúan en la prueba de rendimiento de la primera etapa en la sede del IRRI en Filipinas. Paralelamente, un subconjunto de la cohorte (250 - 300 líneas) se envía a los socios regionales en el sur de Asia. y África Oriental para la evaluación multiambiental de características agronómicas clave (altura de la planta, tiempo para floración, rendimiento de grano). Este subconjunto (conjunto de entrenamiento) se utiliza para construir el modelo de predicción genómica que luego se utiliza para seleccionar una clase avanzada de líneas superiores entre toda la cohorte. Desde que no datos históricos estaban disponibles para la construcción de modelos fiables de predicción genómica, la integración de la predicción genómica en este esquema se basa en el uso de medios hermanos o hermanos completos para maximizar la precisión con entrenamiento altamente relacionado y conjuntos predichos [142, 149]. La predicción genómica Se utilizan modelos para seleccionar líneas parentales para el siguiente ciclo y para seleccionar líneas promisorias (30 - 40) para los ensayos de rendimiento de segunda etapa que se realizan en el quinto año del esquema de mejoramiento. Las líneas con mejor desempeño al final de esta etapa pueden pasar por pruebas avanzadas en el sistema nacional de liberación de variedades o puede ser utilizado por socios en las regiones en su mejoramiento programa para enriquecer sus acervos genéticos.

En esta estrategia, el ciclo de mejoramiento abarca más de cinco años con el reciclaje de líneas avanzadas como padres ocurriendo durante el cuarto año (Figura 2). En comparación con esquemas de reproducción anteriores que existían en el IRRI, el tiempo del ciclo se acorta en 2 años [147]. Reducción del tiempo de ciclo es un factor clave para aumentar la tasa de ganancia genética [109]. En este esquema, una de las principales herramientas para reducción de ciclo es RGA. Este enfoque, conocido desde hace mucho tiempo [150, 151], fue optimizado en 2013 e implementado a gran escala en el IRRI en 2014 [118]. Actualmente, la predicción genómica no es utilizada para disminuir el tiempo del ciclo y se utiliza principalmente para aumentar la intensidad y la precisión de selección en ambientes regionales, especialmente para el rendimiento. La razón principal de esto es la falta de datos históricos en el programa de mejoramiento adecuado para la predicción genómica. De hecho muy pocos las líneas de mejoramiento han sido consistentemente genotipadas y fenotipadas para construir una base de datos confiable. Por lo tanto, la fase actual es una fase de transición donde los datos actualmente generados alimentan una base de datos que se utilizará para predecir el rendimiento de la progenie futura (a través de la cohorte predicciones). Esto se destaca en la Figura 2 como el esquema futuro. Esta capacidad de predecir directamente el desempeño de los candidatos a la selección antes de evaluarlos en el campo nos permitirá disminuir el tiempo del ciclo por dos años adicionales resultando en un ciclo reproductivo de dos años. Sin embargo, esto viene con desafíos operativos tales como: asegurar cuatro generaciones por año en un establecimiento durante RGA, la producción de suficiente semilla al final de la RGA para permitir múltiples ensayos ambientales, y navegar el proceso de importación/exportación lo suficientemente rápido para asegurar que la semilla llega a los socios a tiempo para la siembra en la temporada principal.

### 4.3 Un ejemplo práctico de la canalización analítica

En esta sección, presentamos la canalización de análisis que utilizamos actualmente en el IRRI para realizar análisis genómicos. Esto corresponde a las actividades asignadas al cuarto año de la cría actual. La tubería de análisis se divide en tres pasos principales (Figura 3):



un. *La selección del conjunto de entrenamiento.* Este paso se basa en marcadores SNP elegidos específicamente ser informativo en el germoplasma élite utilizado en el programa de mejoramiento [67] y el método de optimización de Akdemir et al. [39] que minimiza la varianza del error de predicción (PEV) en el conjunto predicho.

B. *El análisis del ensayo único.* En este paso, los datos fenotípicos (altura de la planta, días a la floración y rendimiento de grano) se miden en el conjunto de entrenamiento en varios lugares regionales, que son analizados por separado para evaluar la calidad de los datos en cada ubicación y estimar espacial ajustes a los valores genotípicos con un modelo mixto, teniendo en cuenta el diseño experimental consideración.

C. *El análisis de predicción genómica.* En este último paso, un modelo GBLUP entrenado con el Los datos genotípicos y fenotípicos del conjunto de entrenamiento se utilizan para predecir la estimación genómica valores genéticos (GEBV) para todas las líneas no probadas.

Para ilustrar el proceso de análisis, datos reales del programa de mejoramiento del IRRI para sistemas de riego se utiliza como ejemplo. Los análisis se realizaron dentro del entorno R y utilizaron el Paquetes R *asreml* (bajo licencia) o *sommer* (disponible gratuitamente) para análisis de modelos mixtos y funciones desarrolladas específicamente para la canalización de análisis y de la literatura. hemos optado dar al usuario la posibilidad de elegir entre *asreml* y *sommer* según su preferencias Todos los scripts de R se proporcionan en el archivo adicional 1.

#### 4.3.1 Selección del conjunto de entrenamiento

En el esquema de reproducción actual, la predicción genómica se usa para predicciones dentro de la cohorte. En Para identificar el mejor subconjunto (conjunto de entrenamiento) para ser fenotipado en MET regional, usamos un método de optimización basado en la teoría de modelos mixtos que minimiza la varianza del error de predicción [39]. Este método disponible en el paquete R STPGA (for Selection of Training Populations by Algoritmo genético) requiere la matriz de relación genómica (matriz G) como entrada. En el Por ejemplo, toda la cohorte de 1722 líneas está genotipada con 1079 marcadores SNP. usamos el Paquete rrBLUP para calcular la matriz G basada en la matriz genotípica (geno\_data) que contiene información de marcador codificada como [-1, 0, 1]. La matriz G se utiliza luego como parámetro para el *OptiTS* función junto con el tamaño deseado del conjunto de entrenamiento (STS = 300) y el número de repeticiones (repetición = 5). El número de repeticiones permite la selección de los individuos más representados en las diferentes ejecuciones que se incluirán en el conjunto de entrenamiento para evitar soluciones subóptimas de el algoritmo genético [39]. Evaluar la representatividad del conjunto de entrenamiento en comparación con toda la cohorte, los individuos se grafican utilizando los dos primeros componentes principales de la G matriz (Figura 4).

#### 4.3.2 Análisis de ensayo único

Una vez que se identifica el conjunto de entrenamiento, se envía a los socios regionales para que lo evalúen en MET. Para esto estudio de caso, se utilizaron datos de ensayos reales de cinco lugares diferentes en Bangladesh. Estos ensayos se realizaron en la época seca de 2020 (enero - mayo). Cada ensayo comprende 362 líneas de mejoramiento de los cuales 299 son líneas de entrenamiento, y el resto son líneas avanzadas de la cohorte anterior y consultar variedades. Todos los ensayos usaron un diseño parcialmente replicado con 20% de líneas replicadas. Tres En este ejemplo se utilizan las características: altura de la planta (cm), días hasta la floración y rendimiento de grano (t/ha). La prueba los datos se cargan en la base de datos B4R, que ha sido adoptada por el IRRI para gestionar todos datos de ensayos de cría. Los datos exportados de la base de datos B4R para cada ubicación se utilizan para realizar

análisis de ensayos individuales individuales (objeto `pheno_data`). El objetivo de este paso es eliminar error potencial en el conjunto de datos y para ajustar la variación espacial utilizando el diseño experimental.

El siguiente modelo mixto (*asreml* o *sommer*) se utiliza para obtener el BLUP y el desregresión

BLUP para cada línea:

```
modelo <- asreml( fijo = rasgo ~ 1 aleatorio = ~ DISEÑO_X ,
                 + DISEÑO_Y + GID,
                 na.acción = na.método(x = "incluir"),
                 datos = conjunto de datos)

modelo <- sommer::mmer(fijo = rasgo ~ 1,
                      aleatorio = ~ DISEÑO_X + DISEÑO_Y + GID,
                      rcov = ~ unidades,
                      datos = conjunto de datos,
                      detallado = FALSO)
```

Las variables `DESIGN_X` y `DESIGN_Y` representan las coordenadas de las parcelas dentro del campo.

La variable `GID` representa la ID de los genotipos. Los valores BLUP y BLUP sin regresión

luego se calculan. El análisis de prueba única está integrado en una función llamada *single\_trial\_asreml*

o *single\_trial\_sommer* que toma los datos crudos fenotípicos formateados como entrada y devuelve un

marco de datos con varias variables que incluyen: ubicación, rasgo, ID de genotipo, BLUP, sin regresión

BLUP y repetibilidad ( $H^2$ ). Luego, la función se usa para todas las ubicaciones y rasgos para ejecutar el modelo.

y recuperar los BLUP (Figura 5A).

#### 4.3.3 Predicciones genómicas

El valor BLUP sin regresión de las líneas del conjunto de entrenamiento del análisis de prueba única y el

datos de genotipo de marcadores de todo el genoma de toda la cohorte (conjunto de entrenamiento y conjunto predicho) que consisten

de 1.722 líneas se utilizan en el modelo de predicción genómica. Los datos del marcador de todo el genoma se utilizan

para construir la matriz de relaciones aditivas con el paquete *sommer*. El inverso del aditivo

Luego se construye la matriz de relaciones en el caso de que *asreml* se utilice en el análisis GBLUP. los

GEBV para cada línea se calcula usando el modelo GBLUP donde el BLUP retrocedido de cada

la ubicación es la variable de respuesta, la ubicación como efecto fijo, la línea de cría (gid) y la inversa de

la matriz G (invG) se utilizan como efectos aleatorios.

```
modelo <- asreml(fijo = rasgo ~ 1 + ubicación,
                aleatorio = ~ vm(gid, invG),
                datos = conjunto de datos)
```

```
modelo <- sommer::mmer( fijo = rasgo ~ 1 + ubicación,
                       aleatorio = ~ vs(gid, Gu = G),
                       rcov = ~ vs(unidades),
                       datos = conjunto de datos,
                       detallado = FALSO)
```

De manera similar al análisis de ubicación única, este modelo está incrustado en una función (*gblup\_asreml* o

*gblup\_sommer*) con dos parámetros: el primero es el resultado del análisis de ubicación única y

la segunda es la inversa de la matriz G. La salida de la función es una tabla que contiene los

GEBV en toda la cohorte (Figura 5B). Los valores de GEBV luego se combinan con el marcador de rasgo

información y utilizada por el mejorador para seleccionar líneas para pruebas avanzadas y, también, seleccionar

padres para el próximo ciclo reproductivo.

## 5. Otras aplicaciones de la predicción genómica para el mejoramiento del arroz

En las partes anteriores del capítulo, vimos que la selección genómica requiere tanto

investigación metodológica y un programa de mejoramiento cuidadosamente diseñado para ser implementado

eficientemente. En esta última parte, presentamos los desarrollos en curso con respecto al uso de la genómica

predicciones para el mejoramiento del arroz. Creemos que es importante que los criadores estén al tanto de las próximas

enfoques y herramientas para estar listos cuando sean lo suficientemente maduros para ser integrados en el mejoramiento

programas cuando corresponda.

### 5.1 Caracterización de la diversidad genética para el premejoramiento

La caracterización y el uso de la diversidad genética es importante para cumplir con el mejoramiento a largo plazo.

objetivos y mantener el potencial adaptativo de las poblaciones reproductoras [152]. En el caso de

selección recurrente en germoplasma élite, la adición de nuevo material amenaza la ganancia genética en a corto plazo al diluir el impacto de los alelos de alto valor cuidadosamente acumulados a través de ciclos sucesivos de selección. Sin embargo, a largo plazo, la pérdida de diversidad genética debido a selección, sino también al arrastre de enlace negativo o neutral o la deriva genética puede ser compensado por introducción cuidadosa de la variación genética en el grupo de élite [153]. La identificación de los mejores accesiones para objetivos particulares de mejoramiento es laborioso, ya que requiere un fenotipado preciso de un gran número de líneas diversas que a menudo enmascaran haplotipos valiosos en condiciones de bajo valor genético antecedentes. En este contexto, la predicción genómica se puede utilizar para identificar accesiones superiores en colecciones de germoplasma y aplicarse al mejoramiento previo, cuyo objetivo es identificar genotipos entre un gran número de accesiones [154-156]. En arroz, la disponibilidad de grandes recursos genómicos como los 3000 genomas de arroz [30] o el panel de matriz de arroz de alta densidad [157] ofrecen una oportunidad única de utilizar la predicción genómica para apuntar a genotipos valiosos relativos a los objetivos de cría.

## 5.2 Definición de grupos heteróticos para mejoramiento híbrido

En la cría de híbridos, los grupos heteróticos suelen ser necesarios para utilizar de manera óptima la heterosis dentro de un especie [158]. Para ello, la selección de híbridos hace que el germoplasma se estructure en grupos genéticamente distintos que muestran un rendimiento híbrido superior cuando los individuos de se cruzan grupos complementarios. A diferencia de otros cultivos importantes (p. ej., maíz [159]), heterótico Los grupos en el arroz se definen en gran medida de acuerdo con la complementariedad con un sistema de esterilidad particular. y no de acuerdo con los grupos de genes definidos por el potencial heterótico complementario. esto es mas complicada en el arroz debido a la fuerte estructura poblacional que caracteriza la diversidad del arroz siendo confundido como diferenciación heterótica de grupos de genes complementarios [29, 30]. Esfuerzos para coaccionar subpoblaciones ancestrales en grupos heteróticos, como en el caso de los dos tipos principales (*indica* y *japonica*), tienen limitaciones debido a la esterilidad, adaptaciones contrastantes y muy diferentes

distribuciones de los principales parámetros de calidad del grano [160]. Se requiere más investigación para identificar patrones naturales de heterosis [161] y, en algunos casos, la predicción genómica puede ayudar a esto . exploración. Recientemente, el uso de predicciones para definir grupos heteróticos basados en complementarios Se ha propuesto rendimiento de rendimiento en arroz [162]. En este estudio basado en datos reales, los autores aplicó el enfoque desarrollado por Zhao et al. [163] para detectar patrones heteróticos para el rendimiento por combinando los rendimientos previstos de todos los híbridos únicos únicos con un simulado algoritmo de recocido con diferentes tamaños de grupo.

### 5.3 Integración de fenotipado de alto rendimiento y ambiental información

El importante progreso logrado con la genómica en los programas de mejoramiento ha reforzado la idea de que el fenotipado sigue siendo un cuello de botella para la mejora genética [164]. Esto puede parecer paradójico ya que una de las ventajas de la selección genómica radica en la reducción de algunos fenotipado pasos. Sin embargo, el fenotipado de campo preciso para rasgos importantes (p. ej., rendimiento de grano) en MET es incluso más importante entrenar eficientemente el modelo de predicción y capturar G×E. Además, selección para rasgos más caros o difíciles (resistencia a la sequía, tolerancia al encamado, calidad del grano, etc... ) puede integrarse antes en el esquema de mejoramiento gracias a la predicción genómica y, por lo tanto, aumentar la intensidad de la selección. Estas observaciones han llevado a un interés cada vez mayor en métodos de fenotipado de alto rendimiento [165, 166]. Varias herramientas (RGB y multiespectral cámaras, sensor térmico, etc.) y plataformas (fenomóviles, vehículos aéreos no tripulados, etc.) están disponibles para el fenotipado de campo y laboratorio con una amplia gama de aplicaciones. Cuando integrados en un modelo de predicción genómica, los datos fenotípicos de alto rendimiento pueden aumentar la precisión de la predicción [167, 168]. En el caso de la selección fenómica, el alto rendimiento los datos de espectroscopia de infrarrojo cercano pueden incluso reemplazar los datos genotípicos y ofrecer una precisión similar [169, 170]. Sin embargo, para ser útil en un contexto de mejoramiento, la gran cantidad de datos generados por

las técnicas de fenotipado de alto rendimiento deben almacenarse en un sistema de gestión de datos, examinado adecuadamente en relación con los costos y las precisiones de selección disponibles a partir de fenotipos manuales, y asociado con datos de genotipo correctos si se trata de mejorar el proceso de toma de decisiones.

Si bien las herramientas y los canales de análisis han evolucionado en los últimos años, aún existen importantes limitaciones para el uso rutinario de estos enfoques: la adquisición de campo multiambiente datos y no solo datos de una estación central de investigación, la disponibilidad de gestión de datos sistemas que pueden manejar grandes conjuntos de datos de series temporales y el costo inicial de los equipos relacionados. Es

A medida que maduran las tecnologías y las reglamentaciones, se espera que las empresas dedicadas que ofrecen alta Surgirán servicios de fenotipado de rendimiento, al igual que ha ocurrido con el genotipado.

Además del fenotipado de alto rendimiento, una mejor caracterización de los factores ambientales afectar el rendimiento de las plantas de cultivo mejorará nuestra capacidad para explicar las fuentes no genéticas de variación. Tal "envirotyping" es un área de investigación activa que muestra una gran promesa [171]. Para convertirse en tecnologías verdaderamente útiles que permitan la recopilación de alto rendimiento de datos ambientales en el tiempo real necesita continuar madurando, así como la gestión de datos y las estrategias analíticas para extraer inteligencia de estos conjuntos de datos.

## 6. Conclusión: un punto de vista de un mejorador de arroz

Basado en la literatura en arroz y en otras especies, la capacidad de hacer predicción genómica y la El valor de aplicar la selección genómica a los programas de mejoramiento de arroz está fuera de toda duda. La capacidad estimar los valores de predicción y los conjuntos de datos y modelos clave que subyacen a la estimación de GEBVs también se entiende muy bien. Los recursos de marcadores y la capacidad de fenotipado en arroz son presente y disponible en este momento incluso para las organizaciones de mejoramiento más remotas. Es más, las reglas que describen cómo se hereda la variación cuantitativa de los rasgos en las poblaciones se entienden bien y parecería que el modelo infinitesimal se aplica a los rasgos cuantitativos en el arroz en la mayoría de los casos. Qué

Lo que queda por capturar el valor total de esta tecnología es la reorientación de los programas de mejoramiento de arroz en torno a una estrategia de selección recurrente de ciclo corto dentro de un acervo genético definido. Durante esa transición, Además, la predicción genómica puede ser útil para mejorar la selección dentro de las cohortes y ahorrar dinero en evaluación de campo. Como resultado, la generación de datos de genotipos o la construcción de una tubería analítica a menudo no es el punto de partida para la implementación de la selección genómica en la mayoría de los programas. Claro reglas comerciales para la recopilación y gestión de datos, mejores prácticas claramente definidas para los padres la selección y el compromiso de trabajar dentro de los acervos genéticos de élite deben ser lo primero. En segundo lugar a estos actividades fundacionales, los programas de mejoramiento deben estandarizar y sistematizar sus operaciones en de tal manera que los recursos se optimicen, los flujos de trabajo sean claros y los criadores no gasten cantidades excesivas de tiempo en la gestión de la logística. El trabajo de campo debe centrarse más en la calidad de los datos y recopilación de datos, reservando las decisiones de selección para después de que los datos hayan sido recopilados, analizados y interpretados. Los sistemas de marcadores para el genotipado de rutina también son necesarios, pero deben desarrollarse tal que los datos del genotipo sean específicamente informativos para el germoplasma de mejoramiento de interés.

La literatura pública sobre el arroz hasta la fecha se ha centrado en gran medida en preguntas relacionadas con si las predicciones funcionan en arroz o cómo optimizar la precisión de la predicción. Muy pocas publicaciones sobre el arroz abordan cómo las predicciones se puede aplicar prácticamente a tasas mejoradas de ganancia genética. Como resultado, en un intento de modernizar muchos criadores se quedan atrapados en el 'purgatorio de prueba de concepto' al intentar replicar los análisis hecho por otros. Los criadores que busquen mejorar su estrategia se beneficiarían, en cambio, de considerando si se han sentado las bases apropiadas en sus programas y luego considerando cuidadosamente cuáles son los puntos de entrada para la predicción en su estrategia de reproducción declarada. Comercial Los programas de mejoramiento pueden tener la ventaja de tener la libertad de invertir recursos en capital adicional o gastos operativos por adelantado para capturar valor a largo plazo.

Sin embargo, dado que los presupuestos suelen ser ajustados, fijos o sujetos a la aprobación del Congreso para proyectos financiados con fondos públicos, programas, ajustes de ahorro de costos a la estrategia de reproducción (como la aplicación de una escasa prueba diseño o implementación de avance de generación rápida para la fijación de línea) puede liberar recursos en el



a corto plazo que se puede aplicar para sentar las bases adecuadas para una predicción genómica completa

estrategia de reproducción habilitada.

## Referencias

1. Ragot M, Bonierbale M, Weltzien E (2018) De la demanda del mercado a las decisiones de reproducción: un marco
2. Gallais A (2011) Méthodes de création de variétés en amélioration des plantes -, Quae
3. Brown J, Caligari P (2011) Introducción al fitomejoramiento. John Wiley e hijos
4. Rutkoski JE (2019) Capítulo cuatro: una guía práctica para la ganancia genética. En: Sparks DL (ed.) Avances en Agronomía. Prensa académica, págs. 217–249
5. Lynch M, Walsh B (1998) Genética y análisis de rasgos cuantitativos. Sinauer Sunderland, MA
6. Cooper M, Hammer GL (1996) Adaptación de plantas y mejoramiento de cultivos. Taxi internacional, Wallingford
7. Chenu K (2015) Capítulo 13: Caracterización del entorno de cultivo: naturaleza, importancia y aplicaciones. En: Sadras VO, Calderini DF (eds) Crop Physiology (Segunda Edición). Prensa académica, San Diego, págs. 321–348
8. Xu Y, Li P, Zou C, et al (2017) Mejorando la ganancia genética en la era del mejoramiento molecular. j Exp. Bot 68:2641–2666. <https://doi.org/10.1093/jxb/erx135>
9. Lande R, Thompson R (1990) Eficiencia de la selección asistida por marcadores en la mejora de rasgos cuantitativos. Genética 124:743–756
10. Cobb JN, Biswas PS, Platten JD (2018) Regreso al futuro: revisión de MAS como herramienta para el fitomejoramiento moderno. Theor Appl Genet. <https://doi.org/10.1007/s00122-018-3266-4>
11. Meuwissen TH, Hayes BJ, Goddard ME (2001) Predicción del valor genético total usando Mapas de marcadores densos de todo el genoma. Genética 157: 1819–1829
12. Goddard ME, Hayes BJ (2007) Selección genómica. J Anim Breed Genet 124: 323–330. <https://doi.org/10.1111/j.1439-0388.2007.00702.x>
13. Goddard M (2008) Selección genómica: predicción de precisión y maximización de la respuesta a largo plazo. Genética 136:245–257. <https://doi.org/10.1007/s10709-008-9308-0>
14. Heffner EL, Sorrells ME, Jannink JL (2009) Selección genómica para la mejora de cultivos. Cosechar Sci 49:1–12. <https://doi.org/10.2135/cropsci2008.08.0512>
15. Sallam AH, Endelman JB, Jannink JL, Smith KP (2015) Evaluación de la precisión de la predicción de selección genómica en una población dinámica de mejoramiento de cebada. Genoma vegetal 8:. <https://doi.org/10.3835/plantgenome2014.05.0020>
16. VanRaden PM, Van Tassell CP, Wiggans GR, et al (2009) Revisión invitada: Confiabilidad de predicciones genómicas para toros Holstein norteamericanos. J Dairy Sci 92:16–24. <https://doi.org/10.3168/jds.2008-1514>
17. Hickey JM, Chiurugwi T, Mackay I, et al (2017) La predicción genómica unifica los animales y programas de fitomejoramiento para formar plataformas para el descubrimiento biológico. Nat Genet 49:1297–1303. <https://doi.org/10.1038/ng.3920>
18. Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017) Selección genómica en el mejoramiento de plantas: métodos, modelos y perspectivas. Tendencias Plant Sci 22: 961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
19. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Revisión invitada: Genomic

- selección en ganado lechero: Avances y desafíos. *J Dairy Sci* 92:433–443. <https://doi.org/10.3168/jds.2008-1646>
20. de los Campos G, Hickey JM, Pong-Wong R, et al (2013) Métodos de predicción y regresión del genoma completo aplicados a la reproducción de plantas y animales. *Genética* 193:327–345. <https://doi.org/10.1534/genetics.112.143313>
  21. Izawa T, Shimamoto K (1996) Convertirse en una planta modelo: La importancia del arroz para la ciencia de las plantas. *Tendencias Plant Sci* 1: 95–99. [https://doi.org/10.1016/S1360-1385\(96\)80041-0](https://doi.org/10.1016/S1360-1385(96)80041-0)
  22. Peng S, Khush G (2003) Four Decades of Breeding for Varietal Improvement of Irrigated Lowland Rice en el Instituto Internacional de Investigación del Arroz. *Plant Prod Sci* 6: 157–164. <https://doi.org/10.1626/ppp.6.157>
  23. Chandler RF (1982) Una aventura en ciencias aplicadas: una historia del Instituto Internacional de Investigación del Arroz. IIRI
  24. Breth S (1985) Investigación internacional del arroz: 25 años de asociación
  25. Guimaraes EP (2009) Mejora del arroz. En: *Cereales*. Springer, págs. 99-126
  26. Jena KK, Mackill DJ (2008) Marcadores moleculares y su uso en la terapia asistida por marcadores. Selección en Arroz Todos los derechos reservados. Ninguna parte de este periódico puede ser reproducida o transmitida de ninguna forma o por ningún medio, electrónico o mecánico, incluyendo fotocopias, grabaciones o cualquier sistema de recuperación y almacenamiento de información, sin el permiso por escrito del editor. El editor ha obtenido el permiso para imprimir y reimprimir el material contenido en este documento. *Crop Sci* 48:1266–1276. <https://doi.org/10.2135/cropsci2008.02.0082>
  27. Ismail AM, Singh US, Singh S, et al (2013) La contribución de las variedades de arroz tolerantes a la inmersión (Sub1) a la seguridad alimentaria en áreas bajas de secano propensas a inundaciones en Asia. *Cultivos extensivos Res* 152:83–93. <https://doi.org/10.1016/j.fcr.2013.01.007>
  28. Steele KA, Price AH, Shashidhar HE, Witcombe JR (2006) Selección asistida por marcadores para introducir QTL de arroz que controlan los rasgos de la raíz en una variedad de arroz de tierras altas de la India. *Theor Appl Genet* 112: 208–221. <https://doi.org/10.1007/s00122-005-0110-4>
  29. Glaszmann JC (1987) Isoenzimas y clasificación de variedades de arroz asiático. *Theor Appl Genet* 74:21–30. <https://doi.org/10.1007/BF00290078>
  30. Wang W, Mauleon R, Hu Z, et al (2018) Variación genómica en 3010 accesiones diversas de arroz asiático cultivado. *Naturaleza* 557: 43–49. <https://doi.org/10.1038/s41586-018-0063-9>
  31. Brar D, Khush G (2002) Transferencia de genes de especies silvestres al arroz. En: Kang MS (ed) *Genética cuantitativa, genómica y fitomejoramiento*. pág. 197
  32. Brar DS, Khush GS (2018) Parientes silvestres del arroz: un recurso genético valioso para la investigación en genómica y mejoramiento. En: Mondal TK, Henry RJ (eds) *The Wild Oryza Genomes*. Springer International Publishing, Cham, págs. 1–25
  33. Breseghello F, de Moraes OP, Pinheiro PV, et al (2011) Resultados de 25 años de Upland Mejoramiento de Arroz en Brasil. *Crop Sci* 51:914–923. <https://doi.org/10.2135/cropsci2010.06.0325>
  34. Spindel J, Iwata H (2018) Selección genómica en el mejoramiento de arroz. En: Sasaki T, Ashikari M (eds) *Rice Genomics, Genetics and Breeding*. Springer Singapur, Singapur, págs. 473–496
  35. Ahmadi N, Bartholomé J, Tuong-Vi C, Grenier C (2020) Selección genómica en arroz: resultados empíricos e implicaciones para el mejoramiento. En: *Genética cuantitativa, genómica y fitomejoramiento*. CABI, Wallingford, págs. 243–258
  36. Guo Z, Tucker DM, Basten CJ, et al (2014) El impacto de la estructura de la población en la predicción genómica en poblaciones estratificadas. *Theor Appl Genet* 127:749–762. <https://doi.org/10.1007/s00122-013-2255-x>
  37. Xu SH, Zhu D, Zhang QF (2014) Predicción del rendimiento híbrido en arroz utilizando la mejor predicción genómica lineal imparcial. *Proc Natl Acad Sci EE. UU.* 111: 12456–12461. <https://doi.org/10.1073/pnas.1413750111>

38. Zhang Z, Ober U, Erbe M, et al (2014) Mejora de la precisión del genoma completo predicción de rasgos complejos utilizando los resultados de estudios de asociación del genoma completo. PLoS ONE 9:. <https://doi.org/10.1371/journal.pone.0093017>
39. Akdemir D, Sanchez JI, Jannink JL (2015) Optimización de poblaciones de entrenamiento de selección genómica con un algoritmo genético. Genet Sel Evol GSE 47:38. <https://doi.org/10.1186/s12711-015-0116-6>
40. Blondel M, Onogi A, Iwata H, Ueda N (2015) Un enfoque de clasificación para la selección genómica. PLOS ONE 10:e0128570. <https://doi.org/10.1371/journal.pone.0128570>
41. Grenier C, Cao TV, Ospina Y, et al (2015) Precisión de la selección genómica en una población sintética de arroz desarrollada para el mejoramiento de selección recurrente. POR FAVOR UNO 10:. <https://doi.org/10.1371/journal.pone.0136594>
42. Isidro J, Jannink JL, Akdemir D, et al (2015) Optimización del conjunto de entrenamiento bajo población estructura en la selección genómica. TAG Theor Appl Genet Theor Angew Genet 128:145–158. <https://doi.org/10.1007/s00122-014-2418-4>
43. Iwata H, Ebana K, Uga Y, Hayashi T (2015) Predicción genómica de la forma biológica: Análisis elíptico de Fourier y regresión de mínimos cuadrados parciales (PLS) de Kernel aplicados a la predicción de la forma del grano en arroz (*Oryza sativa* L.). Por favor uno 10:. <https://doi.org/10.1371/journal.pone.0120610>
44. Onogi A, Ideta O, Inoshita Y, et al (2015) Exploración de las áreas de aplicabilidad de métodos de predicción del genoma del arroz asiático (*Oryza sativa* L.). Theor Appl Genet 128:41–53. <https://doi.org/10.1007/s00122-014-2411-y>
45. Spindel J, Begum H, Akdemir D, et al (2015) Selección y asociación genómica Mapeo en arroz (*Oryza sativa*): efecto de la arquitectura genética de rasgos, composición de la población de entrenamiento, número de marcador y modelo estadístico sobre la precisión de la selección genómica del arroz en líneas de mejoramiento de arroz tropical élite. PLOS Genet 11:e1004982. <https://doi.org/10.1371/journal.pgen.1004982>
46. Bustos-Korts D, Malosetti M, Chapman S, et al (2016) Mejora de la capacidad predictiva mediante la cobertura uniforme del espacio genético objetivo. G3-Genes Genomas Genet 6:3733–3747. <https://doi.org/10.1534/g3.116.035410>
47. Jacquin L, Cao TV, Ahmadi N (2016) Una visión unificada y comprensible de los métodos paramétricos y de kernel para la predicción genómica con aplicación al arroz. Frente Genet 7:. <https://doi.org/10.3389/fgene.2016.00145>
48. Onogi A, Watanabe M, Mochizuki T, et al (2016) Hacia la integración de la selección genómica con el modelado de cultivos: el desarrollo de un enfoque integrado para predecir las fechas de espiga del arroz. Theor Appl Genet 129: 805–817. <https://doi.org/10.1007/s00122-016-2667-5>
49. Spindel JE, Begum H, Akdemir D, et al (2016) Los modelos de predicción de todo el genoma que incorporan GWAS de novo son una herramienta nueva y poderosa para la mejora del arroz tropical. Herencia 116:395–408. <https://doi.org/10.1038/hdy.2015.113>
50. Campbell MT, Du Q, Liu K, et al (2017) A Comprehensive Image-based Phnomic El análisis revela la compleja arquitectura genética de la dinámica de crecimiento de los brotes en el arroz (*Oryza sativa*). Genoma vegetal 10:. <https://doi.org/10.3835/plantgenome2016.07.0064>
51. Gao N, Martini JWR, Zhang Z, et al (2017) Incorporating gene annotation into genomic predicción de fenotipos complejos. Genética 207: 489–501. <https://doi.org/10.1534/genetics.117.300198>
52. Matias FI, Galli G, Granato ISC, Fritsche-Neto R (2017) Predicción genómica de plantas autógamas y alógamas por SNPs y haplotipos. Crop Sci 57:2951–2958. <https://doi.org/10.2135/cropsci2017.01.0022>
53. Morais OP, Duarte JB, Breseghello F, et al (2017) Relevancia de la relación genética aditiva y no aditiva para la predicción genómica en la población de arroz bajo selección recurrente . Genet Mol Res 16:. <https://doi.org/10.4238/gmr16039849>
54. Wang X, Li L, Yang Z, et al (2017) Predicción del rendimiento de híbridos de arroz usando

- y modelos GBLUP multivariantes basados en el diseño de apareamiento II de Carolina del Norte. *Herencia* 118:302–310. <https://doi.org/10.1038/hdy.2016.87>
55. Xu S (2017) Suma de errores residuales prevista de cuadrados de modelos mixtos: una aplicación para la predicción genómica. *G3 GenesGenomesGenetics* 7:895–909. <https://doi.org/10.1534/g3.116.038059>
  56. Ben Hassen M, Bartholome J, Vale G, et al (2018) La predicción genómica que da cuenta de la interacción entre el genotipo y el medio ambiente ofrece un marco eficaz para la reproducción simultánea para la adaptación a un estrés abiótico y el rendimiento en condiciones normales de cultivo en arroz. *G3-Genes Genomas Genet* 8:2319–2332. <https://doi.org/10.1534/g3.118.200098>
  57. Ben Hassen M, Cao TV, Bartholome J, et al (2018) El panel de diversidad de arroz proporciona predicciones genómicas precisas para rasgos complejos en las progenies de cruces biparentales que involucran a miembros del panel. *Theor Appl Genet* 131: 417–435. <https://doi.org/10.1007/s00122-017-3011-4>
  58. Campbell M, Walia H, Morota G (2018) Utilizando modelos de regresión aleatoria para análisis genómico predicción de un rasgo longitudinal derivado del fenotipado de alto rendimiento. *Planta directa* 2:. <https://doi.org/10.1002/pld3.80>
  59. Du C, Wei JL, Wang SB, Jia ZY (2018) Selección genómica usando componente principal regresión. *Herencia* 121:12–23. <https://doi.org/10.1038/s41437-018-0078-x>
  60. Gao N, Teng J, Ye S, et al (2018) Predicción genómica de fenotipos complejos usando gen matriz de parentesco basada en la similitud. *Frente Genet* 9:. <https://doi.org/10.3389/fgene.2018.00364>
  61. Mathew B, Léon J, Sillanpää MJ (2018) Impacto de las estructuras de covarianza residual en la capacidad de predicción genómica en ensayos multientorno. *PLoS UNO* 13:. <https://doi.org/10.1371/journal.pone.0201181>
  62. Monteverde E, Rosas JE, Blanco P, et al (2018) Los modelos multientorno aumentan Precisión de la predicción de rasgos complejos en líneas avanzadas de mejoramiento de arroz. *Crop Sci* 58: 1519–1530. <https://doi.org/10.2135/cropsci2017.09.0564>
  63. Morais Júnior OP, Breseghello F, Duarte JB, et al (2018) Evaluación de modelos de predicción para diferentes rasgos en una población de arroz derivados de un programa de selección recurrente. *Crop Sci* 58:2347–2359. <https://doi.org/10.2135/cropsci2018.02.0087>
  64. Morais Júnior OP, Duarte JB, Breseghello F, et al (2018) Norma de reacción de un solo paso modelos para la predicción genómica en ensayos de selección recurrente multiambiente. *Crop Sci* 58:592–607. <https://doi.org/10.2135/cropsci2017.06.0366>
  65. Xu Y, Wang X, Ding XW, et al (2018) Selección genómica de rasgos agronómicos en arroz híbrido utilizando una población NCII. *Arroz* 11:. <https://doi.org/10.1186/s12284-018-0223-4>
  66. Yabe S, Yoshida H, Kajiya-Kanegae H, et al (2018) Descripción del peso del grano distribución que conduce a la selección genómica de las características de llenado de grano en el arroz. *Plós Uno* 13:. <https://doi.org/10.1371/journal.pone.0207627>
  67. Arbelaez JD, Dwiyantri MS, Tandayu E, et al (2019) 1k-RiCA (1K-Rice Custom Amplicon) un nuevo ensayo de SNP basado en amplicón de genotipado para aplicaciones genéticas y de mejoramiento en arroz. *Arroz* 12:55. <https://doi.org/10.1186/s12284-019-0311-0>
  68. Azodi CB, Bolger E, McCarren A, et al (2019) Benchmarking Parametric and Machine Modelos de aprendizaje para la predicción genómica de rasgos complejos. *G3 GenesGenomesGenetics* 9:3691–3702. <https://doi.org/10.1534/g3.119.400498>
  69. Berro I, Lado B, Nalin RS, et al (2019) Optimización de la población de entrenamiento para la selección genómica. *Genoma vegetal* 12:. <https://doi.org/10.3835/plantgenome2019.04.0028>
  70. Bhandari A, Bartholomé J, Cao-Hamadoun TV, et al (2019) Selección de rasgo específico Los marcadores y los modelos multiambientales mejoran la capacidad de predicción genómica en el arroz. *PLOS UNO* 14:e0208871. <https://doi.org/10.1371/journal.pone.0208871>
  71. e Sousa MB, Galli G, Lyra DH, et al (2019) Aumentando la precisión y reduciendo los costos de

- predicción genómica por selección de marcadores. *Euphytica* 215:.. <https://doi.org/10.1007/s10681-019-2339-z>
72. Frouin J, Labeyrie A, Boissard A, et al (2019) La predicción genómica ofrece la mayor enfoque efectivo de mejoramiento asistido por marcadores para la capacidad de prevenir la acumulación de arsénico en los granos de arroz. *Plós uno* 14:.. <https://doi.org/10.1371/journal.pone.0217516>
  73. Guo T, Yu X, Li X, et al (2019) Diseños óptimos para la selección genómica en cultivos híbridos. *Planta Mol* 12: 390–401. <https://doi.org/10.1016/j.molp.2018.12.022>
  74. Hu X, Xie W, Wu C, Xu S (2019) Una estrategia de aprendizaje dirigida que integra múltiples datos ómicos mejora la predicción genómica. *Plant Biotechnol J* 17:2011–2020. <https://doi.org/10.1111/pbi.13117>
  75. Huang M, Balimponya EG, Mgonja EM, et al (2019) Uso de la selección genómica en el cultivo de arroz (*Oryza sativa* L.) para la resistencia al tizón del arroz (*Magnaporthe oryzae*). *Raza Mol* 39:.. <https://doi.org/10.1007/s11032-019-1023-2>
  76. Lima LP, Azevedo CF, De Resende MDV, et al (2019) New insights into genomic selección a través de métodos de predicción no paramétricos basados en la población. *Ciencia Agrícola* 76:290–298. <https://doi.org/10.1590/1678-992X-2017-0351>
  77. Monteverde E, Gutierrez L, Blanco P, et al (2019) Integración de marcadores moleculares y covariables ambientales para interpretar el genotipo por interacción ambiental en arroz (*Oryza sativa* L.) cultivado en áreas subtropicales. *G3-Genes Genomas Genet* 9: 1519–1531. <https://doi.org/10.1534/g3.119.400064>
  78. Ou JH, Liao CT (2019) Determinación del conjunto de entrenamiento para la selección genómica. *Theor Appl Genet* 132:2781–2792. <https://doi.org/10.1007/s00122-019-03387-0>
  79. Suela MM, Lima LP, Azevedo CF, et al (2019) Índice combinado de métodos de predicción genómica aplicados a rasgos de productividad en arroz. *Cienc Rural* 49:.. <https://doi.org/10.1590/0103-8478cr20181008>
  80. Wang S, Wei J, Li R, et al (2019) Identificación de modelos de predicción óptimos utilizando datos multiómicos para seleccionar arroz híbrido. *Herencia*. <https://doi.org/10.1038/s41437-019-0210-6>
  81. Wang X, Xu Y, Li PC, et al (2019) Eficiencia del índice de selección lineal para predecir el rendimiento de los híbridos de arroz. *Raza Mol* 39:.. <https://doi.org/10.1007/s11032-019-0986-3>
  82. Baba T, Momen M, Campbell MT, et al (2020) Modelos de regresión aleatoria de rasgos múltiples aumentar la precisión de la predicción genómica para un rasgo fisiológico temporal derivado del fenotipado de alto rendimiento. *PLoS UNO* 15:.. <https://doi.org/10.1371/journal.pone.0228118>
  83. Banerjee R, Marathi B, Singh M (2020) Selección genómica eficiente mediante el aprendizaje de conjunto y la reducción de características de conjunto. *J Crop Sci Biotechnol* 23:311–323. <https://doi.org/10.1007/s12892-020-00039-4>
  84. Cui YR, Li RD, Li GW, et al (2020) Mejora híbrida de arroz mediante selección genómica. *Planta Biotecnología J* 18:57–67. <https://doi.org/10.1111/pbi.13170>
  85. Grinberg NF, Orhobor OI, King RD (2020) Una evaluación del aprendizaje automático para predicción del fenotipo: estudios en levadura, arroz y trigo. *Mach Learn* 109: 251–277. <https://doi.org/10.1007/s10994-019-05848-5>
  86. Jarquin D, Kajiya-Kanegae H, Taishen C, et al (2020) Coupling day length data and herramientas de predicción genómica para predecir rasgos relacionados con el tiempo en escenarios complejos. *Representante científico* 10:.. <https://doi.org/10.1038/s41598-020-70267-9>
  87. Schrauf MF, Martini JWR, Simianer H, et al (2020) Epistasis fantasma en la selección genómica: sobre la capacidad predictiva de los modelos epistáticos. *G3-Genes Genomas Genet* 10:3137–3145. <https://doi.org/10.1534/g3.120.401300>
  88. Toda Y, Wakatsuki H, Aoike T, et al (2020) Predicción de la biomasa de arroz con rasgos intermedios: método de modelado que combina modelos de crecimiento de cultivos y modelos de predicción genómica. *Plós Uno* 15:.. <https://doi.org/10.1371/journal.pone.0233951>
  89. Xu Y, Zhao Y, Wang X, et al (2020) La incorporación de datos fenotípicos parentales en modelos multiómicos mejora la predicción de rasgos relacionados con el rendimiento en arroz híbrido. *biotecnología vegetal j*.

- <https://doi.org/10.1111/pbi.13458>
90. Piepho HP (2009) Ridge Regression and Extensions for Genomewide Selection in Maize. *Crop Sci* 49:1165–1176. <https://doi.org/10.2135/cropsci2008.10.0595>
  91. de los Campos G, Naya H, Gianola D, et al (2009) Predicting Quantitative Traits With Modelos de regresión para marcadores moleculares densos y pedigrí. *Genética* 182: 375–385. <https://doi.org/10.1534/genetics.109.101501>
  92. Zhong S, Dekkers JCM, Fernando RL, Jannink JL (2009) Factores que afectan la precisión de la selección genómica en poblaciones derivadas de múltiples líneas consanguíneas: un estudio de caso de cebada. *Genética* 182: 355–364. <https://doi.org/10.1534/genetics.108.098277>
  93. Fukuoka S, Ebana K, Yamamoto T, Yano M (2010) Integración de la genómica en el mejoramiento del arroz. *Arroz* 3:131–137. <https://doi.org/10.1007/s12284-010-9044-9>
  94. Zhao K, Tung CW, Eizenga GC, et al (2011) El mapeo de asociación de todo el genoma revela una rica arquitectura genética de rasgos complejos en *Oryza sativa*. 2:467. <https://doi.org/10.1038/ncomms1467>
  95. Hua JP, Xing YZ, Xu CG, et al (2002) La disección genética de un híbrido de arroz de élite reveló que los heterocigotos no siempre son ventajosos para el rendimiento. *Genética* 162:1885–1895
  96. VanRaden PM (2008) Métodos eficientes para computar predicciones genómicas. *J Dairy Sci* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>
  97. Henderson CR (1975) Mejor estimación y predicción lineal imparcial bajo un modelo de selección. *Biometría* 31:423–447. <https://doi.org/10.2307/2529430>
  98. Michel S, Ametz C, Gungor H, et al (2016) Selección genómica a través de reproducción múltiple ciclos en el mejoramiento aplicado de trigo harinero. *Theor Appl Genet* 129: 1179–1189. <https://doi.org/10.1007/s00122-016-2694-2>
  99. Runcie D, Cheng H (2019) Dificultades y remedios para la validación cruzada con múltiples rasgos. *Métodos de predicción genómica. G3 GenesGenomesGenetics* 9:3727–3741. <https://doi.org/10.1534/g3.119.400598>
  100. Gianola D, Schön CC (2016) Validación cruzada sin realizar una validación cruzada en la predicción habilitada para el genoma. *G3 Bethesda Md* 6:3107–3128. <https://doi.org/10.1534/g3.116.033381>
  101. Habier D, Fernando RL, Dekkers JCM (2007) El impacto de la información sobre la relación genética en los valores de reproducción asistida por el genoma. *Genética* 177:2389–2397. <https://doi.org/10.1534/genetics.107.081190>
  102. Technow F, Schrag TA, Schipprack W, et al (2014) Genome Properties and Prospects of Genomic Prediction of Hybrid Performance in a Breeding Program of Maize. *Genética* 197: 1343–1355. <https://doi.org/10.1534/genetics.114.165860>
  103. González-Diéguez D, Legarra A, Charcosset A, et al (2021) La predicción genómica de cultivos híbridos permite desentrañar la dominancia y la epistasis. *Genética* 218:. <https://doi.org/10.1093/genetics/iyab026>
  104. Crossa J (2012) De la interacción genotipo × entorno a gen × entorno Interacción. *Curr Genomics* 13:225–244. <https://doi.org/10.2174/138920212800543066>
  105. Voss-Fels KP, Cooper M, Hayes BJ (2019) Aceleración de las ganancias genéticas de cultivos con selección genómica. *Theor Appl Genet* 132: 669–686. <https://doi.org/10.1007/s00122-018-3270-8>
  106. Bassi FM, Bentley AR, Charmet G, et al (2016) Esquemas de mejoramiento para la implementación de la selección genómica en trigo (*Triticum spp.*). *Plant Sci* 242: 23–36. <https://doi.org/10.1016/j.plantsci.2015.08.021>
  107. Würschum T, Maurer HP, Weissmann S, et al (2017) Precisión de la predicción genómica dentro y entre familias en triticale. *Raza vegetal* 136: 230–236. <https://doi.org/10.1111/pbr.12465>
  108. Edwards SM, Buntjer JB, Jackson R, et al (2019) Los efectos del diseño de la población de entrenamiento en la precisión de la predicción genómica en el trigo. *Theor Appl Genet* 132: 1943–1952.

- <https://doi.org/10.1007/s00122-019-03327-y>
109. Cobb JN, Juma RU, Biswas PS, et al (2019) Mejora de la tasa de ganancia genética en los programas de fitomejoramiento del sector público: lecciones de la ecuación del fitomejorador. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-019-03317-0>
  110. Dreisigacker S, Crossa J, Pérez-Rodríguez P, et al (2021) Implementación de Genomic Selección en el Programa Global de Trigo del CIMMYT, Hallazgos de los últimos 10 años. *Cultivo Breed Genet Genomics* 3:. <https://doi.org/10.20900/cbagg20210005>
  111. Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME (2010) Fitomejoramiento con selección genómica: ganancia por unidad de tiempo y costo. *Crop Sci* 50: 1681–1690. <https://doi.org/10.2135/cropsci2009.11.0662>
  112. Bernardo R (2020) Reinventando la genética cuantitativa para el fitomejoramiento: algo viejo, algo nuevo, algo prestado, algo AZUL. *Herencia*. <https://doi.org/10.1038/s41437-020-0312-1>
  113. García-Ruiz A, Cole JB, VanRaden PM, et al (2016) Cambios en la selección genética diferenciales e intervalos generacionales en ganado lechero Holstein de EE. UU. como resultado de la selección genómica. *Proc Natl Acad Sci* 113:E3995–E4004. <https://doi.org/10.1073/pnas.1519061113>
  114. Bardhan Roy SK, Pateña GF, Vergara BS (1982) Factibilidad de selección por rasgos asociado con tolerancia al frío en arroz bajo el método de avance de generación rápida. *Eufítica* 31:25–31. <https://doi.org/10.1007/BF00028303>
  115. NIIZEKI H, OONO K (1968) Inducción de plantas de arroz haploides a partir de cultivos de anteras. *Proc Jpn Acad* 44: 554–557. <https://doi.org/10.2183/pjab1945.44.554>
  116. Watson A, Ghosh S, Williams MJ, et al (2018) El mejoramiento rápido es una herramienta poderosa para acelerar la investigación y el mejoramiento de cultivos. *Plantas naturales* 4:23–29. <https://doi.org/10.1038/s41477-017-0083-8>
  117. Yan G, Liu H, Wang H, et al (2017) Generación acelerada de plantas de línea pura autofecundadas para identificación de genes y mejoramiento de cultivos. *Planta frontal Sci* 8:. <https://doi.org/10.3389/fpls.2017.01786>
  118. Collard BCY, Beredo JC, Lenaerts B, et al (2017) Revisando los métodos de mejoramiento del arroz – evaluar el uso del avance de generación rápida (RGA) para el mejoramiento de arroz de rutina. *Plant Prod Sci* 20:337–352. <https://doi.org/10.1080/1343943X.2017.1391705>
  119. Bonnacerrere V, Rosas J, Ferraro B (2019) Impacto económico de la asistencia con marcadores selección y avance rápido de la generación en los programas de mejoramiento. *Euphytica* 215:197. <https://doi.org/10.1007/s10681-019-2529-8>
  120. Gaynor RC, Gorjanc G, Bentley AR, et al (2017) Una estrategia de dos partes para usar Genomic Selección para desarrollar líneas puras. *Crop Sci* 57:2372–2386. <https://doi.org/10.2135/cropsci2016.09.0742>
  121. Muleta KT, Pressoir G, Morris GP (2019) Optimización de la selección genómica para un sorgo Programa de mejoramiento en Haití: un estudio de simulación. *G3 GenesGenomesGenetics* 9:391–401. <https://doi.org/10.1534/g3.118.200932>
  122. Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) El impacto de la genética Arquitectura sobre métodos de evaluación de todo el genoma. *Genética* 185: 1021–1031. <https://doi.org/10.1534/genetics.110.116855>
  123. Goddard ME, Hayes BJ, Meuwissen THE (2011) Uso de la matriz de relación genómica para predecir la precisión de la selección genómica. *J Anim Breed Genet* 128: 409–421. <https://doi.org/10.1111/j.1439-0388.2011.00964.x>
  124. Elsen JM (2017) Un marco analítico para derivar la precisión esperada de la genómica selección. *Genet Sel Evol* 49:95. <https://doi.org/10.1186/s12711-017-0366-6>
  125. Norman A, Taylor J, Edwards J, Kuchel H (2018) Optimización de la selección genómica en Trigo: efecto de la densidad del marcador, el tamaño de la población y la estructura de la población en la precisión de la predicción. *G3 GenesGenomesGenetics* 8:2889. <https://doi.org/10.1534/g3.118.200311>

126. Tayeh N, Klein A, Le Paslier MC, et al (2015) Predicción genómica en guisante: efecto de Densidad de marcadores y tamaño y composición de la población de entrenamiento en la precisión de la predicción. *Planta delantera Sci* 6:941. <https://doi.org/10.3389/fpls.2015.00941>
127. Rincet R, Laloë D, Nicolas S, et al (2012) Maximización de la confiabilidad de la selección genómica al optimizar el conjunto de calibración de individuos de referencia: Comparación de métodos en dos grupos diversos de endogamia de maíz (*Zea mays* L.). *Genética* 192:715–728. <https://doi.org/10.1534/genetics.112.141473>
128. Rincet R, Charcosset A, Moreau L (2017) Predecir la eficiencia de la selección genómica para optimizar el conjunto de calibración y evaluar la precisión de la predicción en poblaciones altamente estructuradas. *Theor Appl Genet* 130:2231–2247. <https://doi.org/10.1007/s00122-017-2956-7>
129. Mangin B, Rincet R, Rabier CE, et al (2019) Optimización del conjunto de entrenamiento de genómica predicción mediante EthAcc. *PLOS UNO* 14:e0205629. <https://doi.org/10.1371/journal.pone.0205629>
130. Pszczola M, Strabel T, Mulder HA, Calus MPL (2012) Confiabilidad de los valores genómicos directos para animales con diferentes relaciones dentro y con la población de referencia. *J Dairy Sci* 95:389–400. <https://doi.org/10.3168/jds.2011-4338>
131. Habier D, Tetens J, Seefried FR, et al (2010) El impacto de la relación genética información sobre valores genéticos genómicos en ganado Holstein alemán. *Genet Sel Evol* 42:5. <https://doi.org/10.1186/1297-9686-42-5>
132. Lorenz AJ, Smith KP (2015) Agregar individuos genéticamente distantes al entrenamiento Las poblaciones reducen la precisión de la predicción genómica en la cebada. *Crop Sci* 55:2657–2667. <https://doi.org/10.2135/cropsci2014.12.0827>
133. Lorenz AJ (2013) Asignación de recursos para maximizar la precisión de la predicción y la genética Ganancia de selección genómica en fitomejoramiento: un experimento de simulación. *G3 GenesGenomesGenetics* 3:481–491. <https://doi.org/10.1534/g3.112.004911>
134. Jarquin D, Howard R, Crossa J, et al (2020) Pruebas escasas mejoradas de predicción genómica para ensayos multientorno. *G3 GenesGenomesGenetics* 10:2725. <https://doi.org/10.1534/g3.120.401349>
135. Grattapaglia D, Resende MV (2011) Selección genómica en el mejoramiento de árboles forestales. *Árbol Genet Genomas* 7: 241–255. <https://doi.org/10.1007/s11295-010-0328-4>
136. Hickey JM, Dreisigacker S, Crossa J, et al (2014) Evaluación de diseños de población de entrenamiento de selección genómica y estrategias de genotipado en programas de mejoramiento de plantas usando simulación. *Crop Sci* 54: 1476–1488. <https://doi.org/10.2135/cropsci2013.03.0195>
137. Meuwissen TH (2009) Precisión de los valores genéticos de individuos "no relacionados" predichos por genotipificación densa de SNP. *Genet Sel Evol* 41:35. <https://doi.org/10.1186/1297-9686-41-35>
138. Mackay IJ, Caligari PDS (1999) Principales errores en los datos y su efecto en la respuesta a la selección. *Recortar Sci* 39:cropsci1999.0011183X003900020016x. <https://doi.org/10.2135/cropsci1999.0011183X003900020016x>
139. Israel C, Weller JI (2000) Efecto de la identificación errónea sobre la ganancia genética y la estimación del valor de cría en poblaciones de ganado lechero. *J Dairy Sci* 83:181–187. [https://doi.org/10.3168/jds.S0022-0302\(00\)74869-7](https://doi.org/10.3168/jds.S0022-0302(00)74869-7)
140. Breseghello F, Mello RN de, Pinheiro PV, et al Construyendo el conjunto de datos de mejoramiento de arroz de Embrapa para una reutilización eficiente de datos. *Ciencia de cultivos n/a*: <https://doi.org/10.1002/csc2.20550>
141. Juanillas V, Dereeper A, Beaume N, et al (2019) Rice Galaxy: un recurso abierto para la ciencia de las plantas. *Gigaciencia* 8:. <https://doi.org/10.1093/gigascience/giz028>
142. Akdemir D, Isidro-Sánchez J (2019) Diseño de poblaciones de entrenamiento para fenotipado en la predicción genómica. *Sci Rep* 9:1446. <https://doi.org/10.1038/s41598-018-38081-6>
143. Ben-Sadoun S, Rincet R, Auzanneau J, et al (2020) Optimización económica de un esquema de reproducción mediante fenotipado selectivo del conjunto de calibración en un contexto de múltiples rasgos:



- Aplicación a la calidad de la panificación. *Theor Appl Genet* 133:2197–2212. <https://doi.org/10.1007/s00122-020-03590-4>
144. Rasheed A, Hao Y, Xia X, et al (2017) Chips de mejoramiento de cultivos y plataformas de genotipado: progreso, desafíos y perspectivas. *Mol Planta* 10: 1047–1064. <https://doi.org/10.1016/j.molp.2017.06.008>
  145. Gorjanc G, Dumasy JF, Gonen S, et al (2017) Potencial de genotipado de baja cobertura por secuenciación e imputación para la selección genómica rentable en poblaciones segregantes biparentales. *Crop Sci* 57:1404–1420. <https://doi.org/10.2135/cropsci2016.08.0675>
  146. Cobb J, Rafiqul M, Kumar Katiyar S, et al (2020) La evolución de una revolución: re diseñar programas de reproducción de la revolución verde en Asia y África para aumentar las tasas de ganancia genética. [W020]. PAG, público, págs. 9–9
  147. Collard BCY, Gregorio GB, Thomson MJ, et al (2019) Transforming Rice Breeding: Re Designing the Irrigated Breeding Pipeline en el Instituto Internacional de Investigación del Arroz (IRRI). *Cultivo Raza Genet Genomics* 1:e190008. <https://doi.org/10.20900/cbagg20190008>
  148. Thomson MJ, Singh N, Dwiyanti MS, et al (2017) Implementación a gran escala de una matriz SNP 6 K de arroz para aplicaciones de genética y mejoramiento. *Arroz* 10:40. <https://doi.org/10.1186/s12284-017-0181-2>
  149. Habier D, Fernando RL, Garrick DJ (2013) Genomic BLUP Decoded: A Look into the Caja Negra de Predicción Genómica. *Genética* 194: 597–607. <https://doi.org/10.1534/genetics.113.152207>
  150. Maruyama K (1989) Uso de avance de generación rápida con descendencia de semilla única en arroz cría. Instituto Internacional de Investigación del Arroz, págs. 253–259
  151. De Pauw RM, Clarke JM (1976) Aceleración del avance generacional en trigo de primavera. *Euphytica* 25: 415–418. <https://doi.org/10.1007/BF00041574>
  152. McCouch S, Baute GJ, Bradeen J, et al (2013) Alimentando el futuro. *Naturaleza* 499:23. <https://doi.org/10.1038/499023a>
  153. Cowling WA (2013) Fitomejoramiento sostenible. *Raza vegetal* 132: 1–9. <https://doi.org/10.1111/pbr.12026>
  154. Gorjanc G, Jenko J, Hearne SJ, Hickey JM (2016) Iniciando programas de premejoramiento de maíz usando selección genómica para aprovechar la variación poligénica de las poblaciones de razas locales. *BMC Genómica* 17:30. <https://doi.org/10.1186/s12864-015-2345-z>
  155. Yu X, Li X, Guo T, et al (2016) Predicción genómica que contribuye a un futuro global prometedor estrategia para impulsar los bancos de genes. *Plantas naturales* 2:16150. <https://doi.org/10.1038/nplants.2016.150>
  156. Tanaka R, Iwata H (2018) Optimización bayesiana para la selección genómica: un método para descubrir el mejor genotipo entre un gran número de candidatos. *Theor Appl Genet* 131:93–105. <https://doi.org/10.1007/s00122-017-2988-z>
  157. Wang DR, Agosto-Pérez FJ, Chebotarov D, et al (2018) Una plataforma de imputación para mejorar la integración de los recursos genéticos del arroz. *Nat Commun* 9:3519. <https://doi.org/10.1038/s41467-018-05538-1>
  158. Melchinger AE, Gumber RK (1998) Resumen de heterosis y grupos heteróticos en cultivos agronómicos. *Conceptos Raza Heterosis Cultivo Plantas* 25:29–44
  159. Reif JC, Melchinger AE, Xia XC, et al (2003) Uso de SSR para establecer grupos heteróticos en maíz subtropical. *Theor Appl Genet* 107:947–957. <https://doi.org/10.1007/s00122-003-1333-x>
  160. Ouyang Y, Liu YG, Zhang Q (2010) Esterilidad híbrida en plantas: historias de arroz. *Curr Opin Plant Biol* 13: 186–192. <https://doi.org/10.1016/j.pbi.2010.01.002>
  161. Xie F, He Z, Esguerra MQ, et al (2014) Determinación de grupos heteróticos para germoplasma de arroz híbrido Indica tropical. *Theor Appl Genet* 127: 407–417. <https://doi.org/10.1007/s00122-013-2227-1>

162. Beukert U, Li Z, Liu G, et al (2017) Genome-Based Identification of Heterotic Patterns in Arroz. *Arroz* 10:22. <https://doi.org/10.1186/s12284-017-0163-4>
163. Zhao Y, Li Z, Liu G, et al (2015) Establecimiento basado en el genoma de un patrón heterótico de alto rendimiento para el mejoramiento de trigo híbrido. *Proc Natl Acad Sci* 112:15624–15629. <https://doi.org/10.1073/pnas.1514547112>
164. Araus JL, Kefauver SC, Zaman-Allah M, et al (2018) Traducción de fenotipado de alto rendimiento en ganancia genética. *Tendencias Plant Sci* 23: 451–466. <https://doi.org/10.1016/j.tplants.2018.02.001>
165. Araus JL, Cairns JE (2014) Fenotipado de alto rendimiento de campo: la nueva frontera de mejoramiento de cultivos. *Tendencias Plant Sci* 19:52–61. <https://doi.org/10.1016/j.tplants.2013.09.008>
166. Pauli D, Chapman SC, Bart R, et al (2016) The Quest for Understanding Phenotypic Variación a través de enfoques integrados en el entorno de campo. *Physiol vegetal* 172: 622–634. <https://doi.org/10.1104/pp.16.00592>
167. Rutkoski J, Poland J, Mondal S, et al (2016) La temperatura del dosel y los índices de vegetación del fenotipado de alto rendimiento mejoran la precisión del pedigrí y la selección genómica para el rendimiento de grano en trigo. *G3 GenesGenomesGenetics* 6:2799–2808. <https://doi.org/10.1534/g3.116.032888>
168. Juliana P, Montesinos-López OA, Crossa J, et al (2019) Integrating genomic-enabled predicción y fenotipado de alto rendimiento en el mejoramiento de trigo harinero resistente al clima. *Theor Appl Genet* 132: 177–194. <https://doi.org/10.1007/s00122-018-3206-3>
169. Rincen R, Charpentier JP, Faivre-Rampant P, et al (2018) La selección fenómica es un método de bajo costo y alto rendimiento basado en predicciones indirectas: prueba de concepto en trigo y álamo. *G3 Genes Genomas Genet* g3.200760.2018. <https://doi.org/10.1534/g3.118.200760>
170. Lane HM, Murray SC, Montesinos-López OA, et al (2020) Selección fenómica y predicción del rendimiento del grano de maíz a partir de la espectroscopia de reflectancia del infrarrojo cercano de los granos. *Fenómeno vegetal J* 3:e20002. <https://doi.org/10.1002/ppj2.20002>
171. Xu Y (2016) Envirotyping para descifrar los impactos ambientales en las plantas de cultivo. *Theor Appl Genet* 129: 653–673. <https://doi.org/10.1007/s00122-016-2691-5>

## Agradecimientos

Los autores agradecen a Adam Famoso y Flavio Breseghello por sus valiosos comentarios y la revisión integral del capítulo. También nos gustaría agradecer al equipo de arroz de regadío del IRRI: Rose Imee Zhella Morante, Vitaliano Lopena, Holden Verdeprado y Juan David Arbelaez, por su ayuda con la adquisición y gestión de datos con respecto al ejemplo proporcionado en el programa de mejoramiento del IRRI. Agradecemos al equipo de IRRI Bangladesh y en particular a Rafiqul M. Islam, así como a nuestros socios en Bangladesh por su apoyo en la obtención de datos fenotípicos para el conjunto de entrenamiento presentado en el ejemplo.

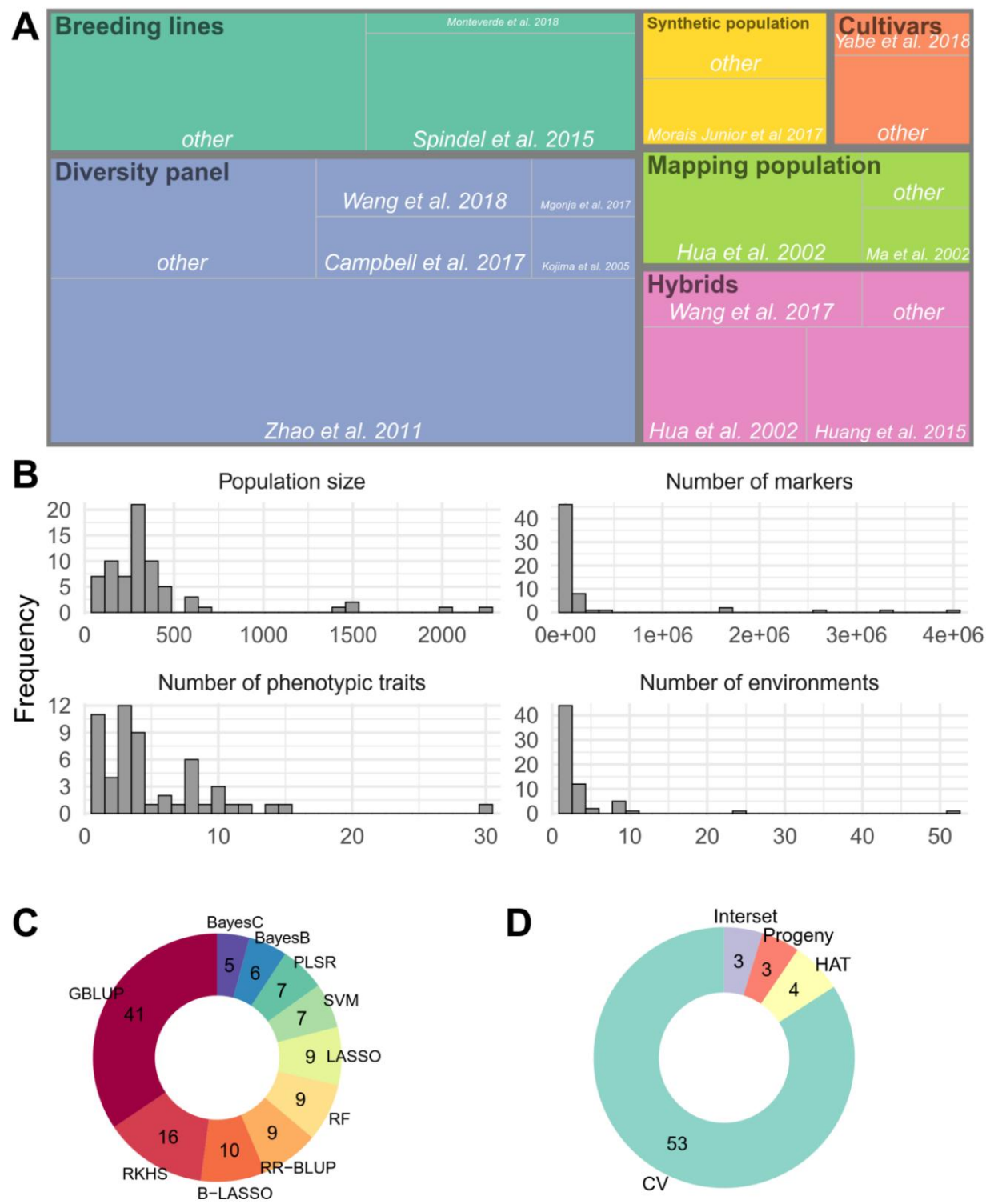
## Fondos

La Fundación Bill y Melinda Gates a través de la Ganancia Genética Acelerada en Arroz (AGGRi)  
El proyecto Alliance patrocinó y financió este trabajo.

## Información adicional

Archivo adicional 1: R scripts para la canalización de análisis de predicción genómica que se utiliza actualmente en el IRRI. Los datos del programa de cría en regadío se proporcionan como ejemplo de un caso real.

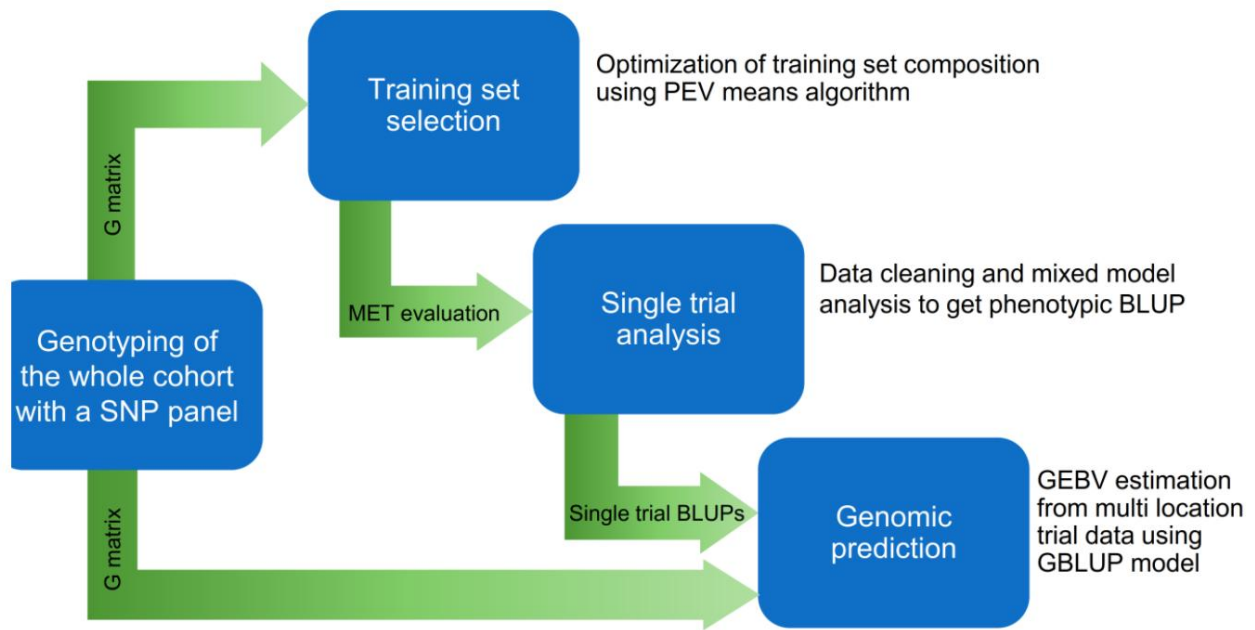
Cifras



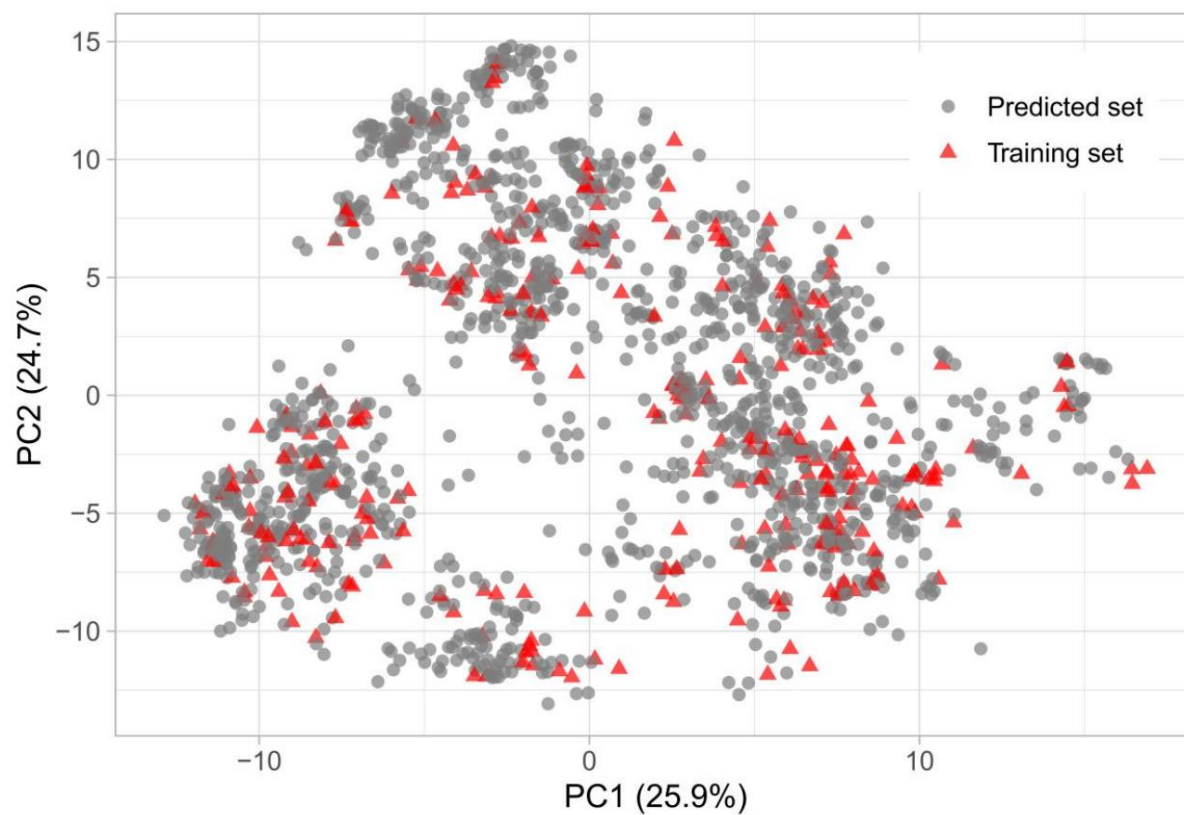
**Figura 1:** Resumen de la literatura sobre predicción genómica de arroz. representa la información se detalla en la Tabla 1. (A) Diagrama de árbol de los tipos de poblaciones utilizadas para entrenar la predicción genómica modelos y las referencias asociadas para estudios que se basaron en ya publicados conjuntos de datos (B) Histogramas de las características importantes de los conjuntos de datos: el tamaño del población, el número de rasgos fenotípicos, el número de ambientes en los que los rasgos fueron medido (año, temporada o ubicación) y el número de marcadores moleculares utilizados para la genómica predicciones (C) Diagrama circular de los diez modelos de predicción más utilizados en los 54 estudios. (D) Diagrama circular de la estrategia de validación utilizada para evaluar la precisión de los modelos de predicción: cruce validación (CV), método HAT, validación entre conjuntos y validación de progenie.



**Figura 2.** Esquemas de mejoramiento anteriores, actuales y futuros en IRRI para sistemas irrigados. los la evolución entre los esquemas se caracteriza por la integración de la predicción genómica (GP) y una reducción de la duración del ciclo reproductivo. La predicción genómica se indica en rojo con el número asociado de individuos fenotipados en las regiones para actualizar el modelo. El color de los pasos corresponde a la ubicación de las actividades: verde en Filipinas y amarillo en el regiones con los socios. Los años y las estaciones (WS: estación húmeda, DS: estación seca) son indicado en el lado izquierdo. Los números a la derecha indican el tamaño de la población de cada paso. los las flechas gruesas negras indican el reciclaje de las mejores líneas como padres. MAS: asistido por marcador selección utilizando 10-20 marcadores de rasgos, en su mayoría relacionados con la resistencia a enfermedades. INGER: Internacional red de evaluación genética del arroz liderada por el IRRI.

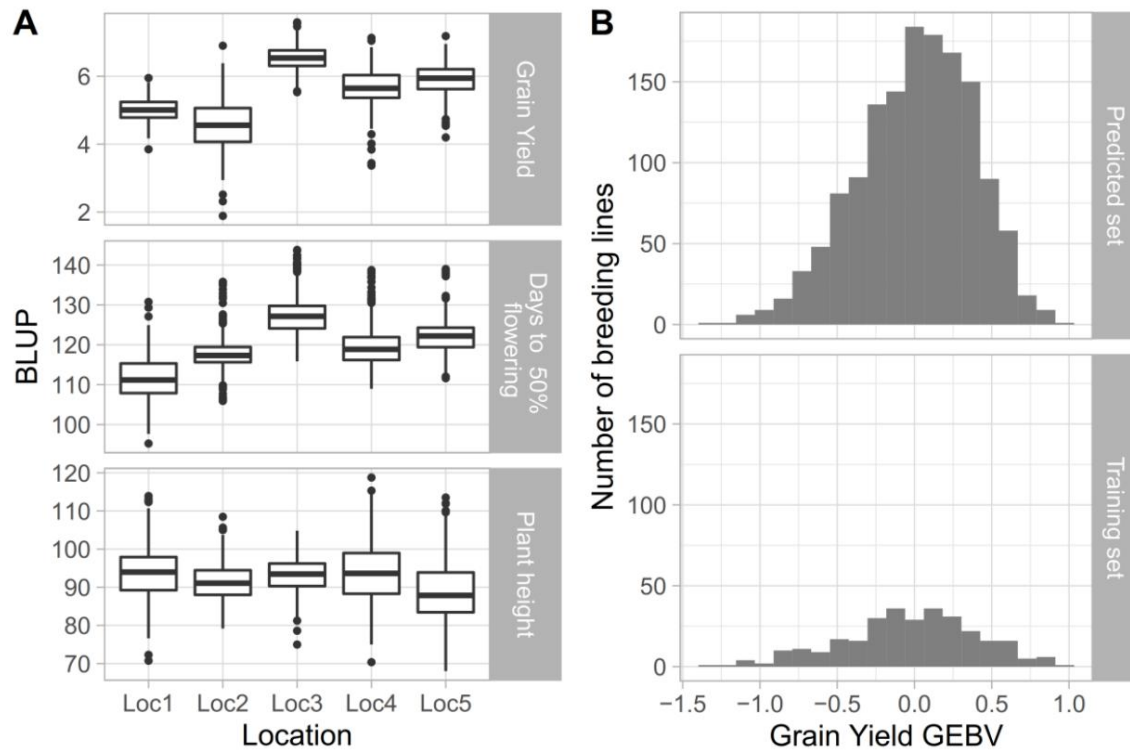


**Figura 3.** El diagrama de flujo del análisis de datos representa los pasos de rutina que se realizan para cada ciclo de mejoramiento en el programa de mejoramiento del IRRI. Toda la cohorte (prueba de rendimiento de la primera etapa) es la primera genotipado con un panel SNP y los datos se utilizan para seleccionar una población de entrenamiento (subconjunto de la toda la cohorte). Luego, la población de entrenamiento se evalúa en ensayos multiambientales (MET). los los ensayos individuales se analizan con un modelo mixto que tiene en cuenta el diseño experimental. A continuación, se utilizan los BLUP de un solo ensayo combinados con la información del marcador de toda la cohorte. para calcular los valores genéticos estimados genómicos (GEBV) de las líneas.



**Figura 4.** Análisis de componentes principales usando los datos de marcadores moleculares en todas las líneas de mejoramiento utilizado para la predicción genómica. Los triángulos negros representan las líneas seleccionadas para formar el entrenamiento. establecido utilizando el método de optimización de Akdemir et al. [39]. Las líneas restantes (en círculos rojos) componen el conjunto predicho.





**Figura 5.** Resultados del análisis de ensayo único y análisis de predicción genómica. El panel A muestra

el diagrama de caja de los valores BLUP para rendimiento de grano, días al 50% de floración y altura de la planta desde el 5 ubicaciones de prueba de los socios. El Panel B presenta la distribución del GEBV de rendimiento de grano del pronóstico y líneas de entrenamiento. Los resultados se obtuvieron usando *asreml*.

# Mesas

**Tabla 1:** Estudios de predicción genómica en arroz. Cuando se usaron múltiples conjuntos de datos en un estudio, la información se reporta solo para el conjunto de datos de arroz.

Referencia	Población	Número de		Modelos de predicción	Tipo de validación	Exactitud	Objetivo principal
	Escribe	Marcadores de rasgos de tamaño					
Guo et al. [36]	Panel de diversidad [94]	413 30		36901 GBLUP	Validación cruzada (k-fold)	0.21 - 0.84	Evaluar empíricamente el impacto de la estructura de la población en la precisión de la predicción genómica utilizando experimentos de validación cruzada en el modelo de predicción genómica (GBLUP)
Xu et al. [37]	Híbridos [95]	278 - 105	4	1619 GBLUP, LASSO, SSVS	Validación cruzada (k-fold) - Validación entre conjuntos	0 - 0,69	Investigar el efecto de las variaciones no aditivas en la eficiencia de la predicción genómica para el rendimiento híbrido
Zhang et al. [38]	Panel de diversidad [94]	413 11		36901 GBLUP, BayesB, BLUP GA	Validación cruzada (k-fold)	0.51 - 0.85	Incorpore matrices de relaciones genómicas específicas de rasgos utilizando el conocimiento existente de arquitecturas genéticas en forma de regiones QTL significativas obtenidas en estudios de asociación independientes en modelos de predicción genómica para mejorar la exactitud
Akdemir et al. [39]	Panel de diversidad [94]	413 6		36901 RR-BLUP	Validación cruzada (submuestreo)	0,2 - 0,8	Diseño de una población de entrenamiento para maximizar la precisión de los modelos de predicción genómica utilizando un algoritmo genético
Blondel et al. [40]	Panel de diversidad [94]	335 14		1311 RF, McRank ordinal, RangoSVM, GBRT, RKHS RR, LambdaMART, B-LASSO EB-LASSO, MEZCLA, SSVS, BayesC, wBSR	Validación cruzada (k-fold)	0,68 - 0.72	Formular la predicción genómica como el problema de clasificar a los individuos según su valor genético para emplear métodos de aprendizaje automático para clasificar
Grenier et al. [41]	Población reproductora sintética	343 4		8336 B-LASSO, B-RR, GBLUP, RR-BLUP, LAZO	Validación cruzada (k-fold)	0.12 - 0.54	Investigar el efecto de los factores clave (tamaño y composición de la población de entrenamiento, número de marcadores, modelo) en la precisión de la predicción genómica

Isidro et al. [42]	Panel de diversidad [94]	413 4		36901 RR-BLUP	Validación cruzada (submuestreo)	0.22 - 0.73	Compare el rendimiento de diferentes criterios de optimización en presencia de estructura de población y evalúe cómo la estructura de población interactúa con estos criterios en la elección de la población de entrenamiento	
Iwata et al. [43]	Dos paneles de diversidad	179 - 386	1	3,254 - 36,901	GBLUP, RKHS, PLSR, KPLSR	Validación cruzada (omitir uno, k-fold)	0.4 - 0,64	Proponer un método para predecir la forma del grano de arroz delineado por descriptores elípticos de Fourier basados en polimorfismos de marcadores de todo el genoma
Onogi et al. [44]	Cultivares	110 - 8		3102	GBLUP, RKHS, LASSO, Red elástica, RF, B-LASSO, EB-LASSO, BSR	Validación cruzada (k-fold)	0,40 - 0,84*	Evaluar el rendimiento de ocho métodos diferentes sobre la precisión de la predicción utilizando datos reales y simulados
Spindel et al. [45]	Líneas de cría	332 3		73147	RR-BLUP, B-LASSO, RKHS, RF, MLR, PBLUP	Validación cruzada (submuestreo)	0 - 0,63*	Investigar el efecto del número de marcadores, el modelo y la arquitectura de rasgos en la precisión de la predicción genómica
bustos Korts et al. [46]	Panel de diversidad [94]	413 3		26259	GBLUP, QGBLUP, RKHS	Validación cruzada (submuestreo)	0.28 - 0.81	Evaluar diferentes métodos para optimizar la población de entrenamiento y su interacción con los modelos de predicción
Jacquín et al. [47]	Líneas de cría y Panel de diversidad	230 - 167 - 188	15	22,691 - 16,444 - 38,390	LASSO, GBLUP, SVM, RKHS	Validación cruzada (k-fold)	0.12 - 0.70	Proporcionar una comprensión clara y unificada de los métodos estadísticos paramétricos y del núcleo, utilizados para la predicción genómica, y para comparar algunos de estos en el contexto de la mejora del arroz.
Onogi et al. [48]	Mapeo de la población (Ma et al. 2002)	174 1		162	EB-LASSO, EB-LASSO + modelo de cultivo	Validación cruzada (omitir uno)	0.87 - 0.97	Prediga la fecha del rumbo combinando la predicción genómica y el modelo de cultivo
Spindel et al. [49]	Líneas de cría	332 3		58318	GBLUP, RR-BLUP, B LASSO, RKHS, RF, MLR	Validación cruzada (submuestreo)	0 - 0,65*	Evaluar el potencial de introducir variables fijas identificadas utilizando GWAS de novo en modelos GS para mejorar la precisión de la predicción, y también considerar la contribución de las pruebas de campo en múltiples ubicaciones a la precisión de la predicción GS
Campbell et al. [50]	Panel de diversidad [94]	360 1		36901	GBLUP	Validación cruzada (k-doblar)	0.39 - 0.73	Evalúa la precisión de la predicción genómica de la dinámica de crecimiento de los brotes
Gao et al. [51]	Líneas de cría [45]	315 3		58227	GBLUP (10 matrices de relación)	Validación cruzada (k-fold)	0.24 - 0.57	Incorpore anotaciones de genes en matrices de relaciones para mejorar la precisión de la predicción genómica
Matías y Alabama. [52]	Líneas de cría [45]	270 2		39915	B-RR, BayesB, B-LASSO	Validación cruzada (submuestreo)	0.26 - 0.42	Uso de bloques de haplotipos como marcadores multialélicos para mejorar la precisión de la predicción genómica

Morais y <i>Alabama</i> . [53]	Población reproductora sintética	174 8	6174	GBLUP (5 matrices de relaciones)	Validación cruzada (submuestreo)	0.31 - 0.68	Evaluar la relevancia de aditivos y no aditivos efectos genéticos en la precisión predictiva	
Wang et al. [54]	Híbridos	575 8	329915 0	GBLUP (univariado y multivariado)	Validación cruzada (k-fold)	0.40 – 0.86	Investigar el rendimiento de modelos multivariados, incluida la dominancia, para predecir fenotipos de híbridos de arroz que se benefician del análisis conjunto con rasgos auxiliares o con los fenotipos observados en otros entornos.	
Xu et al. [55]	Híbridos	1495 10	165403 0	GBLUP	HAT, validación cruzada (k veces) Validación cruzada	0.40 - 0.88	Desarrollar un método alternativo, el método HAT, para reemplazar la validación cruzada en el contexto de la predicción genómica	
Ben Hassen et al. [56]	Panel de diversidad Líneas de cría	284 - 97	3	43686	GBLUP, RKHS (univariado y multivariado)	(submuestreo) - Validación de progenie	-0.12 - 0.96	Explorar la viabilidad de la selección genómica para la adaptación del arroz a la alternancia de humedecimiento y secado en el marco de un esquema de mejoramiento genético
ben hassen et al. [57]	Panel de diversidad Líneas de cría	284 - 97	3	43686	GBLUP, RKHS, BayesB	(submuestreo) - Validación de progenie	0.23 - 0.65	Investigar el impacto del tamaño y la composición de la población de entrenamiento que maximiza la precisión de la predicción del fenotipo de las líneas de progenie.
Campbell et al. [58]	Panel de diversidad [50]	357 1		33674	GBLUP	Validación cruzada (k-fold, submuestreo)	0.4 - 0.89	Examinar la ventaja de utilizar modelos de regresión aleatoria para fenotipos longitudinales sobre la medición de punto final único en el contexto de la predicción genómica
Du et al. [59]	Mapeo de población [95]	210	4 - 1,000 - 24,97 3	1619	RR-BLUP, PCR, PLSR	Validación cruzada (k-fold) - SOMBRERO	0.12 - 0.76	Evaluar las ventajas de la regresión de componentes principales sobre la regresión de mínimos cuadrados parciales para la predicción genómica de rasgos agronómicos, metabolómicos y transcriptómicos
Gao et al. [60]	Líneas de cría [45]	315 3		58227	GBLUP (7 matrices de relación)	Validación cruzada (k-fold)	0.24 - 0.56	Incorpore información de anotaciones de genes en modelos de predicción genómica mediante la construcción de haplotipos con SNP asignados a regiones génicas para mejorar la precisión.
Mateo et al. [61]	Panel de diversidad [94]	371 1		36901	GBLUP (multivariado)	Validación cruzada (k-fold)	0.49 - 0.77	Estudie el impacto de diferentes estructuras de covarianza residual en la capacidad de predicción genómica utilizando diferentes modelos para analizar datos de ensayos de múltiples entornos.

Monteverd et al. [62]	Líneas de cría	309 - 327	5	44,598 - 92,430	GBLUP, RHKS (multivariante)	Validación cruzada (submuestreo)	0,30 - 0.88	Compare el efecto sobre la precisión de la predicción de diferentes modelos multientorno y diferentes poblaciones de entrenamiento
Morais Júnior et al. [63]	Población reproductora sintética [53]	174	8	6174	ABLUP, GBLUP, AGBLUP HBLUP, BayesC, B-LASSO, PLSR, RF, RKHS	Validación cruzada (submuestreo)	0.23 - 0.76	Comparar modelos de predicción para identificar los métodos de selección genómica más precisos y desarrollar métodos de selección genómica de bajo riesgo para su uso en la mejora del arroz.
Morais Júnior et al. [64]	Población reproductora sintética [53]	667 - 174	3	6174	HBLUP bayesiano (multivariante con covariables ambientales)	Validación cruzada (submuestreo)	-0.15 - 0.9	Evaluar modelos de un solo paso que incorporen covariables ambientales y la importancia de los efectos principales y los componentes de interacción para la predicción de respuestas fenotípicas.
Xu et al. [sesenta y cinco]	Híbridos [54]	575	8	256188 9	GBLUP, PLSR, LASSO, BayesB, SVM, RKHS	Validación cruzada (k-fold)	0.15 - 0.88	Evaluar los efectos de los métodos estadísticos, la heredabilidad, la densidad de marcadores y el tamaño de la población de entrenamiento en la predicción del rendimiento híbrido
Yabe et al. [66]	Cultivares	123	1	42508	GBLUP, PLSR	HAT, Validación cruzada (omitir uno)	0.22 - 0.53	Desarrollar un método para describir la distribución del peso del grano y evaluar la eficiencia de la predicción genómica para los parámetros específicos del genotipo de la distribución del peso del grano.
Arbeláez et al. [67]	Líneas de cría	353	3	965	ABLUP, RR-BLUP, BayesA, BayesB, BayesC, B-LASSO, RKHS	Validación cruzada (k-fold)	0.36 - 0.71	Evaluar la efectividad de una plataforma de genotipado de mil sitios SNP altamente informativos para la predicción genómica en la reproducción basada en <i>indica</i> programas
Azodi et al. [68]	Líneas de cría	327	3	73147	RR-BLUP, B-RR, BayesA, BayesB, B-LASSO, SVM, RF, GTB, ANN, CNN	Validación cruzada (k-fold)	0.25 - 0,65	Compare el rendimiento de diferentes modelos de predicción, incluidas las redes neuronales artificiales, utilizando conjuntos de datos disponibles
Berro et al. [69]	Líneas de cría [62]	317 - 327	1	44,598 92,430	GBLUP	Validación cruzada (k-doblar)	0.37 - 0.80	Comparar estrategias para optimizar el conjunto de entrenamiento para modelos de predicción genómica
Bhandari et al. [70]	Panel de diversidad	280	3	215242	GBLUP, RKHS	Validación cruzada (submuestreo)	0.23 - 0.81	Investigar sobre la efectividad de la selección de marcadores específicos de rasgos y de los modelos de predicción de múltiples entornos para mejorar la precisión de las predicciones genómicas para la tolerancia a la sequía en el arroz.
E Sousa et al. [71]	Líneas de cría [45]	270	2	39811	GBLUP, RKHS	Validación cruzada (submuestreo)	0.18 - 0.31	Comparar el efecto de dos estrategias para obtener subconjuntos de marcadores y su efecto sobre la precisión de la predicción, el sesgo y la eficiencia relativa de un modelo de efecto genotípico principal
Frouin et al. [72]	Panel de diversidad	228 - 95	2	22370	GBLUP, BayesA, RKHS	Validación cruzada (submuestreo) -	0.23 - 0.54	Explora la viabilidad de la selección genómica para mejorar la capacidad del arroz para prevenir la absorción y acumulación de arsénico en los granos comestibles.

	Líneas de cría					Intersect validación		
Guo et al. [73]	Híbridos	1439 4		165403 0	GBLUP	Validación cruzada (submuestreo)	0.59 - 0,77*	Optimice la población de entrenamiento para la predicción genómica del rendimiento híbrido utilizando técnicas de pensamiento de diseño y minería de datos. Evaluar una estrategia novedosa de predicción genómica llamada operador de selección y contracción mínima absoluta
Hu et al. [74]	Mapeo de población [95]	210	4 - 1,000 - 24,973	1619	multicapa-LASSO	Validación cruzada (k-fold)	0.16 - 0.76	multicapa (ML-LASSO) mediante la integración de múltiples datos ómicos en un solo modelo que aprende iterativamente tres capas de características genéticas supervisadas por el transcriptoma y el metaboloma observados
Huang et al. [75]	Panel de diversidad [94]	161 - 162	1	66,109 29,030	RR-BLUP, GBLUP (multivariado), BayesA, BayesC	Validación cruzada (k-fold)	0.15 - 0.80	Evaluar la utilidad de la predicción genómica para mejorar la resistencia al añublo del arroz  Proponer el Delta-p (método basado en la distancia genética entre dos subpoblaciones, utilizando los conceptos de cambios en la frecuencia alélica debido a la selección y la teoría de la ganancia genética) y el índice Delta p/G-BLUP, y compararlo con el tradicional G -Método BLUP
Lima et al. [76]	Panel de diversidad [94]	370 7		36901	GBLUP, Delta-p	Validación cruzada (k-fold)	0.27 - 0.83	
Monteverd et al. [77]	Líneas de cría	309 - 327	4	44,598 92,430	GBLUP, PLSR (covariables ambientales multivariadas)	Validación cruzada (submuestreo)	0.10 - 0,90*	Use datos de marcadores moleculares y covariables ambientales simultáneamente para predecir el rendimiento del arroz y las características de calidad de la molienda en entornos no probados
Ou et al. [78]	Panel de diversidad [94]	404 10		30315	RR-BLUP	Validación cruzada (submuestreo)	0.09 - 0,78*	Proponer un nuevo criterio derivado de la correlación de Pearson entre los GEBV y los valores fenotípicos de un conjunto de prueba para determinar un conjunto de entrenamiento para la predicción genómica
Suela et al. [79]	Panel de diversidad [94]	352 9		36901	Delta-p, GBLUP, BayesC, B-LASSO	Validación cruzada (k-doblar)	0.10 - 0.83	Evaluar los índices genómicos Delta-p/BLASSO y Delta-p/BayesCpi y compararlos con el índice Delta p/G-BLUP en términos de eficiencia de predicción de valores genómicos aditivos
Wang et al. [80]	híbridos, Mapeo de población [95]	210 - 278	4 - 1,000 - 24,973	1619	LASSO, GBLUP, SVM, PLSR	Validación cruzada (k-doblar)	0.1 - 0,70*	Demostrar el concepto de que la previsibilidad de los rasgos se puede optimizar mediante el uso de modelos de predicción superiores y conjuntos de datos ómicos selectivos

Wang et al. [81]	Híbridos [54]	575 8	61836	GBLUP	Validación cruzada (k-fold)	0.07 - 0.15	Combine el índice de selección con el método de predicción genómica para predecir el arroz híbrido para una selección más precisa y completa
Baba et al. [82]	Panel de diversidad [50, 95]	357 2	34993	Regresión aleatoria (uni y multivariante)	Validación cruzada (submuestreo)	0.17 - 0,91*	Demostrar la utilidad de un modelo de regresión aleatoria de características múltiples para la predicción genómica del uso diario de agua en el arroz mediante el modelado conjunto con biomasa de brotes
Banerjee et al. [83]	Líneas de cría [45]	315 3	73147	RR, LASSO, SVM, embolsado, RF, AdaBoost, XGBoost	Validación cruzada (k-fold)	0.10 - 0,67	Compare métodos de predicción lineales y no lineales y evalúe la eficiencia de los enfoques de reducción de dimensionalidad utilizando el arroz como ejemplo.
Cui et al. [84]	Híbridos	1,49 5 - 100	10 6	102795 GBLUP (multivariante)	Validación cruzada (k-doblar) - Intersect	0,35 - 0,92	Utilice la mejor predicción genómica lineal imparcial para predecir el rendimiento híbrido mediante la validación cruzada y la validación entre conjuntos
Grinberg et al. [85]	Panel de diversidad [30]	2265 12	101595	LASSO, RR, GBLUP, GBM, RF, MVS	Validación cruzada (k-fold)	0.14 - 0.70	Compare los métodos estándar de aprendizaje automático con el método de genética estadística clásica de última generación: GBLUP
Jarquín et al. [86]	Cultivares [66]	112 1	408372	GBLUP	Validación cruzada (submuestreo)	0.41 - 0,93	Proponer dos métodos novedosos para predecir los días hasta la espiga en arroz de genotipos probados y no probados en ambientes no observados de manera precisa y exacta. Explorar cómo la diferencia en la capacidad predictiva de los modelos epistáticos y los modelos aditivos se relaciona con la densidad de los marcadores utilizados para las predicciones. y poner las observaciones en el contexto de la epistasis fantasma
Schrauf et al. [87]	Panel de diversidad [30]	2018 1	400000 0	GBLUP 3 matrices de relación)	Validación cruzada (k-fold)	0.16 - 0,83*	
Toda et al. [88]	Mapeo de población	123 1	315	GBLUP, LASSO, RR, RKHS, RF (integración con modelo de cultivo)	Validación cruzada (k-fold, submuestreo)	0,40 - 0,68*	Desarrolle modelos para predecir la biomasa del arroz con la integración de los datos fenotípicos observados de los rasgos relacionados con el crecimiento, el genotipo marcador del genoma completo y los datos ambientales.
Xu et al. [89]	híbridos, Mapeo de población [95]	210 - 278	4 - 1,000 - 24,97 3	1619 GBLUP (diferentes matrices de relación)	HAT - Validación de progenie	0.20 - 0,80*	Integre la información fenotípica de los padres en varios modelos de predicción multiómicos aplicados en el mejoramiento híbrido de arroz y compare las previsibilidades de 15 combinaciones de cuatro conjuntos de predictores de los padres, es decir, genoma, transcriptoma, metaboloma y fenoma.

\* leído de gráficos no mencionados en el texto