

PENDEKATAN SELEKSI GENOMIK DALAM Mendukung PROGRAM PEMULIAAN KELAPA SAWIT

Fathur Rachman*, Sujadi, Heri Adriwan Siregar, dan Yurna Yenni

Abstrak - Kelapa sawit merupakan salah satu tanaman penghasil minyak nabati terbesar di dunia. Produktivitas yang tinggi pada kelapa sawit tidak lepas dari peran penyediaan bahan tanaman unggul yang dilakukan melalui program pemuliaan kelapa sawit dengan skema *reciprocal recurrent selection* (RRS). Akan tetapi, skema tersebut memiliki kelemahan dalam hal siklus waktu yang lama dan intensitas seleksi yang rendah. Perkembangan teknologi molekuler yang masif melalui seleksi genomik dapat mengatasi permasalahan tersebut. Beberapa faktor pertimbangan yang harus diperhatikan dalam mengadopsi pendekatan seleksi genomik antara lain karakteristik dari sifat yang akan diperbaiki, tipe dan jumlah marka molekuler, ukuran dan keterkaitan populasi pelatihan dan pengujian, serta model statistik yang akan digunakan. Penelitian mengenai adopsi seleksi genomik pada kelapa sawit telah dilakukan untuk menjawab permasalahan yang ada. Pada program pemuliaan kelapa sawit, seleksi genomik berhasil memberikan efisiensi biaya yang baik (26%-65%) dan mempercepat siklus hingga 35%. Selain itu, permasalahan intensitas seleksi yang rendah pada RRS juga dapat diatasi melalui strategi pra-seleksi genomik.

Kata kunci: GBLUP, RRGs, seleksi berbantu marka

PENDAHULUAN

Kelapa sawit (*Elaeis guineensis* Jacq.) merupakan salah satu tanaman yang banyak dibudidayakan sebagai tanaman penghasil minyak (Corley & Tinker, 2016), dengan produksi minyak yang lebih tinggi secara komparatif dibandingkan tanaman penghasil minyak nabati lainnya (FAOSTAT, 2020). Produksi minyak sawit dunia didominasi oleh Indonesia, kemudian diikuti oleh negara lainnya seperti Malaysia, Thailand, Kolombia, dan Nigeria (Goggin & Murphy, 2018). Hingga saat ini, produktivitas tandan buah segar di Indonesia tergolong cukup tinggi dengan kisaran 16,3-21,0 ton ha⁻¹ (Ariesca et al., 2023). Peningkatan produktivitas kelapa sawit tidak lepas dari penyediaan dan penggunaan bahan tanaman unggul melalui program pemuliaan. Pemuliaan pada kelapa sawit diawali dengan kegiatan seleksi massa pada tahun 1920, hingga pada akhirnya ditemukan pewarisan sifat ketebalan cangkang yang menjadi titik

awal pengembangan kelapa sawit tipe *tenera* (Corley & Tinker 2016). Pengembangan hibrida tipe *tenera* menjadikan skema seleksi massa digantikan oleh skema seleksi berulang resiprokal (*reciprocal recurrent selection*, RRS) (Corley & Tinker, 2016; Soh et al., 2017).

Konsep breeding scheme RRS adalah memilih individu terbaik berdasarkan keragaan keturunannya (*tenera*). Selanjutnya individu terpilih dapat digunakan sebagai tetua pada produksi benih komersial sekaligus bahan genetik pada siklus berikutnya. Awal mula skema RRS pada pemuliaan kelapa sawit dilakukan oleh Institut de Recherche pour les Huile et Oléagineux (IRHO, saat ini CIRAD) di Côte d'Ivoire dengan memberikan hasil yang baik dalam menghasilkan kelapa sawit unggul dan mengeksplorasi heterosis (Meunier & Gascon, 1972). Pada skema ini, kelapa sawit dikelompokkan pada dua grup heterotik, yaitu grup A (jumlah tandan sedikit, ukuran tandan besar) yang umumnya merupakan populasi Deli, dan grup B (jumlah tandan banyak, ukuran tandan kecil) yang terdiri atas populasi Afrika (Meunier & Gascon, 1972). Walaupun skema seleksi ini memiliki keuntungan dalam akurasi pendugaan nilai genetik, skema tersebut masih memiliki kelemahan, seperti membutuhkan waktu yang lama (satu siklus

Penulis yang tidak disertai dengan catatan kaki instansi adalah peneliti pada Pusat Penelitian Kelapa Sawit

Fathur Rachman* (✉)
Pusat Penelitian Kelapa Sawit
Jl. Brigjen Katamso No. 51 Medan 20158, Indonesia

Email: rachmanfathur.fr@gmail.com

membutuhkan waktu sekitar 20 tahun) (Yusop et al., 2020) dan menggunakan intensitas seleksi yang rendah (Nyouma et al., 2019).

Pemanfaatan informasi genomik menjadi titik awal dalam mengatasi tantangan pada kegiatan pemuliaan kelapa sawit. Informasi genomik dimanfaatkan dalam kegiatan seleksi berbantu marka (*marker-assisted selection*, MAS) melalui deteksi *quantitative trait loci* (QTL) sebagai marka seleksi, sehingga intensitas seleksi dapat ditingkatkan dan siklus pemuliaan dapat diperpendek (Muranty et al., 2014). Akan tetapi, skema MAS sulit dilakukan pada karakter kuantitatif yang dikendalikan oleh banyak QTL dengan efek kecil (Grattapaglia et al., 2018; Muranty et al., 2014). Oleh karena itu, Meuwissen et al. (2001) mengenalkan seleksi genomik (*genomic selection*, GS) sebagai salah satu pendekatan yang efisien untuk perbaikan karakter kuantitatif melalui integrasi informasi fenotipik dan genotipik.

Pada kelapa sawit, Wong & Bernardo (2008) menunjukkan kemampuan GS yang lebih efisien dalam hal waktu dan biaya untuk menghasilkan kemajuan genetik tinggi dibandingkan seleksi berbasis fenotipe. Penelitian GS pada kelapa sawit telah banyak dilakukan dalam studi keterkaitan populasi pelatihan dan pengujian (Cros et al., 2015a), perbandingan model statistik dan tipe marka (Kwong et al., 2017a), pemanfaatan model multivariat (Marchal et al., 2016), pendekatan *genotyping* (Cros et al., 2017; Kwong et al., 2017b; Nyouma et al., 2022), seleksi klon (Nyouma et al., 2020), pengembangan hibrida *E. guineensis* × *E. oleifera* (Garzón-Martínez et al., 2022), dan strategi integrasi GS dalam skema RRS (*reciprocal recurrent genomic selection*, RRGs) (Cros et al., 2015a; Cros et al., 2017; Cros et al., 2018). Selain kelapa sawit, skema seleksi genomik juga telah banyak dilakukan pada pemuliaan ternak (Wiggans et al., 2017), tanaman semusim (Edwards et al., 2019; Kadam et al., 2016; Y. Zhao et al., 2013), dan tanaman tahunan (Cros et al., 2019; Durán et al., 2017; Muranty et al., 2015; Yadav et al., 2020). Pada tulisan ini, topik yang akan dibahas terdiri atas selayang pandang pendekatan seleksi genomik, faktor-faktor yang perlu dipertimbangkan dalam seleksi genomik, dan kondisi terkini pemanfaatan pendekatan seleksi genomik pada pemuliaan kelapa sawit.

PENDEKATAN SELEKSI GENOMIK

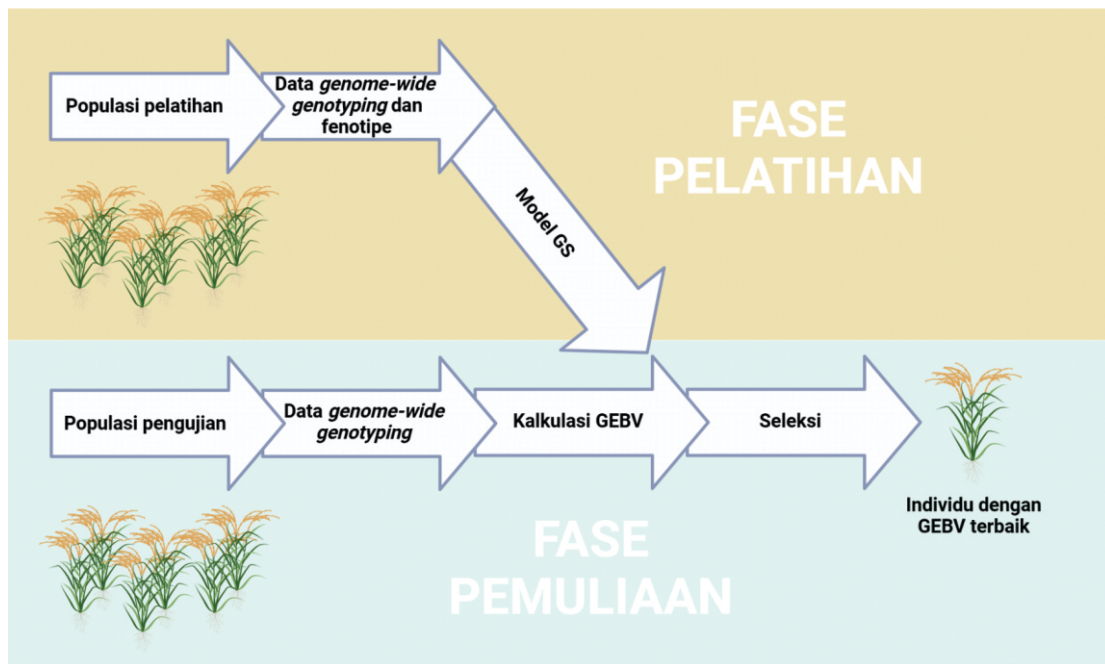
Seleksi genomik, atau biasa disebut juga sebagai *genome-wide selection*, merupakan pendekatan baru dalam kegiatan seleksi berbantu marka molekuler. Menurut Goddard & Hayes (2007), seleksi genomik sebagai bentuk lain dari MAS yang menggunakan marka molekuler pada keseluruhan genom, sehingga seluruh QTL berada dalam *linkage disequilibrium* (LD) dengan paling sedikit satu marka. Metode ini pertama kali dikembangkan oleh Meuwissen et al. (2001) untuk mengidentifikasi nilai genetik atau nilai pemuliaan suatu karakter kuantitatif. Parameter nilai pemuliaan (*breeding value*, BV) diduga berdasarkan pengaruh genomik (*genomic estimated breeding value*, GEBV), yang selanjutnya digunakan sebagai parameter dalam kegiatan seleksi. Secara tradisional, pendugaan nilai pemuliaan telah dilakukan menggunakan informasi silsilah (*pedigree*) melalui prediksi linier tak bias terbaik (*best linear unbiased prediction*, BLUP) (Piepho et al., 2008; Sood et al., 2020).

Pelaksanaan seleksi genomik secara garis besar dilakukan pada dua tahapan, yaitu (1) fase pelatihan (*training phase*) untuk prediksi model pada populasi pelatihan (*training population*); serta (2) fase pemuliaan (*breeding phase*) untuk penentuan GEBV pada populasi pengujian (*test population*) dan seleksi (Gambar 1; Meuwissen et al., 2001). Pengumpulan data fenotipe dan genotipe dilakukan pada individu-individu dalam populasi pelatihan, yang selanjutnya digunakan untuk membangun model penduga GEBV. Model tersebut kemudian diuji tingkat akurasi, sehingga model dapat digunakan pada proses berikutnya. Data yang dikumpulkan pada populasi pengujian hanya berupa informasi genotipe, dimana data tersebut akan dimasukkan pada model untuk menghitung GEBV. Proses ini umumnya disebut sebagai prediksi genomik (*genomic prediction*, GP) (Elias, 2022). Selanjutnya, proses seleksi pada populasi pengujian dilakukan berdasarkan GEBV.

Seperti penjelasan sebelumnya, pendekatan GS merupakan pengembangan dari pendekatan MAS. Oleh karena itu, kerangka kerja yang digunakan memiliki kemiripan dalam hal sistem kegiatan yang dibagi menjadi fase pelatihan dan fase pemuliaan. Walau demikian, terdapat perbedaan antara GS dan MAS jika ditinjau dari beberapa faktor. Pada fase pelatihan, informasi QTL diidentifikasi untuk skema MAS,

sedangkan model GEBV untuk skema GS. Pada fase pemuliaan, informasi genotipe yang digunakan pada MAS hanya pada daerah genom tertentu, sedangkan informasi genotipe pada GS berdasarkan keseluruhan genom. Selain itu, seleksi pada MAS didasarkan pada marka molekuler terkait QTL, sedangkan seleksi pada GS didasarkan pada GEBV (Nakaya & Isobe, 2012). Beberapa studi telah dilakukan untuk membandingkan dua pendekatan tersebut. Heffner et al. (2010) menunjukkan

bahwa pendekatan GS mampu memberikan kemajuan genetik dua dan tiga kali lebih besar dibandingkan pendekatan MAS pada program pemuliaan gandum dan jagung. Pada pemuliaan kelapa sawit, Wong & Bernardo (2008) menemukan penggunaan GS memberikan tingkat efisiensi yang baik dibandingkan dengan MAS apabila ditinjau dari biaya dan waktu. Beberapa ulasan riset terkait pemanfaatan GS pada kegiatan pemuliaan kelapa sawit akan dibahas terperinci pada tulisan ini.



Gambar 1. Tahapan kegiatan seleksi genomik secara umum (Diadaptasi dari Heffner et al., 2009)

FAKTOR PERTIMBANGAN DALAM SELEKSI GENOMIK

Karakteristik Sifat

Karakteristik dari suatu sifat seperti heritabilitas (Muranty et al., 2015) dan arsitektur genetik dapat mempengaruhi akurasi GS (Wang et al., 2013). Sifat-sifat dengan nilai heritabilitas yang tinggi dapat menghasilkan peningkatan akurasi prediksi. Zhang et al. (2019) menemukan bahwa sifat dengan heritabilitas yang tinggi dapat meningkatkan akurasi prediksi. Lin et al. (2014) menjelaskan bahwa sebagian besar karakter yang bernilai ekonomi memiliki heritabilitas yang tinggi, sehingga GS menjadi pendekatan yang potensial dalam perakitan varietas baru dengan nilai ekonomi yang baik. Karakter dengan heritabilitas

rendah juga dapat dipertimbangkan dalam pendekatan GS. Kizilkaya et al. (2014) menginvestigasi kemampuan GS pada karakter dengan heritabilitas yang rendah, seperti resistensi penyakit, dimana akurasi GS dapat ditingkatkan dengan penggunaan model ambang (*threshold model*).

Selain heritabilitas, arsitektur genetik seperti jumlah QTL dan distribusinya juga dapat berperan dalam menentukan tingkat akurasi GS (Davoudi et al., 2018; Grattapaglia, 2017). Peningkatan jumlah QTL dapat menurunkan akurasi GS akibat partisi ragam genetik aditif ke sejumlah besar QTL, sehingga efisiensi estimasi efek yang kecil dari QTL akan menurun (Abdollahi-Arpanahi et al., 2013). Selain itu, penurunan akurasi GS akibat peningkatan jumlah QTL

cenderung terlihat pada kondisi kerapatan marka yang rendah dan ukuran populasi efektif yang lebih besar (Grattapaglia, 2017). Model statistik juga memegang peran penting yang berkaitan dengan jumlah QTL dan akurasi GS. Akurasi GS akan sangat sensitif terhadap jumlah QTL apabila menggunakan model statistik Bayesian, tetapi hal sebaliknya terjadi pada model GBLUP (Wientjes et al., 2015). Walau demikian, beberapa studi menunjukkan pengaruh jumlah QTL tidak begitu signifikan dalam menentukan tingkat akurasi GS (Ahmadi et al., 2021; Ashoori-Banaei et al., 2021; Kasnavi et al., 2018).

Di sisi lain, efek dari QTL atau gen juga dapat memberikan pengaruh terhadap tingkat akurasi prediksi. Karakter kualitatif umumnya dikendalikan oleh satu atau dua gen mayor, sedangkan karakter kuantitatif dikendalikan oleh banyak gen yang bersifat minor. Akurasi prediksi karakter yang dikendalikan oleh sejumlah kecil gen mayor menunjukkan akurasi prediksi yang tinggi (Gao et al., 2015; Z. Zhang et al., 2015). Walau demikian, hubungan arsitektur genetik terhadap akurasi prediksi juga ditentukan oleh model statistik yang digunakan.

Marka Molekuler

Linkage disequilibrium (LD) didefinisikan sebagai tingkat keterpautan antar lokus yang terjadi akibat asosiasi tak acak antar lokus (Klug et al., 2018). Prinsip ini penting dimanfaatkan pada GS karena menggambarkan keterpautan suatu marka atau penanda terhadap lokus yang diteliti. Semakin tinggi intensitas LD antar dua lokus, maka semakin tinggi tingkat keterpautannya (Ramakrishnan, 2013). Intensitas LD umumnya diestimasi dengan menghitung deviasi (D) antara frekuensi haplotipe terhadap frekuensi alel harapan pada kondisi kawin acak. Nilai D berkisar antara $-0,25$ hingga $0,25$, dimana $D = 0$ menggambarkan tidak adanya LD antar dua lokus yang diteliti (Ramakrishnan, 2013).

LD yang baik pada model GS dapat dicapai dengan mempertimbangkan jumlah marka dan distribusinya pada keseluruhan genom. Jumlah marka dapat ditentukan dengan membagi ukuran genom terhadap jarak antar marka yang memiliki LD signifikan (Nakaya & Isobe, 2012). Jumlah marka atau kerapatan marka dapat mempengaruhi tingkat akurasi, dimana peningkatan kerapatan marka dapat meningkatkan

akurasi GS (Cao et al., 2017; Lee et al., 2017; Xu et al., 2018). Walau demikian, Kriaridou et al. (2020) mengungkapkan bahwa penggunaan kerapatan marka rendah masih dapat memberikan akurasi yang baik pada GS akuakultur. Di sisi lain, kerapatan marka akan sangat beragam pada setiap populasi yang digunakan pada program pemuliaan. Populasi yang berasal dari populasi liar (*wild-type population*) membutuhkan kerapatan marka yang lebih tinggi dibandingkan populasi yang berasal dari persilangan biparental (Hao et al., 2019; Liu et al., 2018). Selain itu, kerapatan marka akan lebih dipertimbangkan pada tanaman menyerbuk silang dibanding tanaman menyerbuk sendiri (Juliana et al., 2019).

Selain kerapatan marka, hal yang juga perlu dipertimbangkan adalah tipe marka yang digunakan pada aplikasi GS. Secara umum, marka molekuler dapat dibagi menjadi dua tipe, yaitu marka dominan dan marka kodominan. Marka dominan merupakan marka yang tidak dapat mengidentifikasi perbedaan alel suatu gen pada kondisi heterozigot, sedangkan marka kodominan dapat mengidentifikasi perbedaan alel tersebut. Beberapa contoh marka yang tergolong dalam marka dominan adalah *random amplified polymorphic DNA* (RAPD), *inter simple sequence repeat* (ISSR), dan *amplified fragment length polymorphism* (AFLP), sedangkan contoh marka kodominan yaitu *restriction fragment length polymorphism* (RFLP), *single sequence repeat* (SSR), dan *single nucleotide polymorphism* (SNP) (Adhikari et al., 2017). Tipe marka juga dapat mempengaruhi tingkat akurasi GS (Lin et al., 2014), dimana marka kodominan dapat memberikan akurasi pendugaan GEBV yang lebih baik dibandingkan dengan marka dominan. Hal ini disebabkan oleh adanya tingkat kekuatan deteksi LD yang baik dan perbaikan akurasi dengan mempertimbangkan haplotipe yang terdeteksi marka kodominan. Penggunaan marka multi-alelik juga dapat memberikan akurasi dan efisiensi yang baik dibandingkan marka bi-alelik (Sallam et al., 2020). Selain berkaitan dengan akurasi prediksi, penentuan tipe marka juga akan berkaitan dengan penggunaan biaya riset. Perkembangan *next generation sequencing* (NGS) memberikan peningkatan efisiensi dalam menghasilkan data yang besar dengan biaya yang kecil. *Genotyping-by-sequencing* (GBS) dan *SNP arrays* merupakan metode NGS yang umum digunakan pada GS untuk mendapatkan informasi SNP pada keseluruhan genom dengan kerapatan yang tinggi (Lan et al., 2020; Lin et al., 2014).

Populasi Pelatihan dan Pengujian

Tingkat akurasi GS dapat dipengaruhi oleh karakteristik dari suatu populasi, seperti hubungan antara populasi pelatihan dan pengujian serta ukuran populasi. Hubungan genetik antara populasi pelatihan dan pengujian menjadi hal yang penting untuk dipertimbangkan karena hubungan tersebut akan mempengaruhi akurasi prediksi pada GS (Fraslin et al., 2022). Hubungan yang rendah antar populasi akan menurunkan akurasi prediksi (Liu et al., 2015; Lozada et al., 2019). Hal ini disebabkan karena penanda yang berada pada kondisi LD dengan suatu QTL pada populasi pelatihan tidak dapat mewakili latar belakang genetik pada populasi pengujian (Sun et al., 2016). Oleh karena itu, populasi pelatihan dan pengujian yang digunakan terkadang diasumsikan berasal dari moyang yang sama (Nakaya & Isobe, 2012). Di sisi lain, ukuran populasi pelatihan juga dapat mempengaruhi akurasi prediksi GEBV (Nakaya & Isobe, 2012), dimana peningkatan ukuran populasi pelatihan akan meningkatkan akurasi prediksi. Akan tetapi, peningkatan tersebut akan mencapai *plateau* pada ukuran tertentu (Cericola et al., 2017), sehingga ukuran populasi optimum dapat ditentukan. Walau demikian, peningkatan ukuran populasi akan berimbas pada peningkatan biaya riset, sehingga penelitian mendalam terkait hal tersebut perlu dilakukan. Nakaya & Isobe (2012) menjelaskan bahwa rasio populasi pelatihan terhadap pengujian dapat beragam pada setiap spesies akibat perbedaan keragaman genetik, heritabilitas, dan jumlah QTL.

Model Statistik

Keberhasilan kegiatan pemuliaan dengan pendekatan GS ditentukan oleh konstruksi model statistik dalam memprediksi informasi atau parameter genetik secara akurat. Tantangan yang ditemukan pada GS adalah jumlah data *genotype* yang sangat banyak dibandingkan jumlah individu uji ("*large-p small-n*" problems). Kondisi tersebut dapat menyebabkan permasalahan statistik seperti tingkat multikolinearitas antar marka yang tinggi, sehingga terjadi *overfitting* pada model statistik yang digunakan. Pada tahap berikutnya, *overfitting* dapat menyebabkan penurunan akurasi prediksi (Ying, 2019). Oleh karena itu, model statistik yang digunakan harus dapat menanggulangi

permasalahan tersebut. Secara umum, model statistik yang digunakan dalam metode GS dapat dibagi menjadi metode parametrik, semi-parametrik, dan non-parametrik (Lin et al., 2014; Wang et al., 2018). Beberapa model statistik tersebut antara lain sebagai berikut:

1. *Ridge-regression* BLUP (RR-BLUP), merupakan model linier parametrik yang pertama kali dikenalkan dalam aplikasi GS oleh Meuwissen et al. (2001) untuk menduga efek marka melalui model campuran linier aditif sebagai efek acak. Melalui model ini, penyusutan (*shrinkage*) efek marka menuju nilai nol dilakukan secara merata melalui suatu parameter koreksi (λ) (Gouy et al., 2013; Lin et al., 2014) untuk memenuhi asumsi efek kecil dari setiap marka (Budhlakoti et al., 2022). Nilai pemuliaan dapat diduga dengan akurat melalui akumulasi efek marka pada setiap individu. Model ini memiliki kelemahan dalam menduga nilai pemuliaan pada karakter yang dikendalikan oleh gen mayor (Meuwissen et al., 2001).
2. *Least absolute shrinkage and selection operator* (LASSO), merupakan model linier parametrik untuk menduga efek marka yang mengkombinasikan pemilihan variabel dan penyusutan (Tibshirani, 1996). LASSO memiliki kemiripan dengan RR-BLUP, tetapi penyusutan koefisien lebih kuat terjadi pada model LASSO, sehingga beberapa koefisien memiliki nilai nol (Lorenz et al., 2011). Model ini memiliki keunggulan dalam efisiensi dan kecepatan komputasi data besar. Akan tetapi, model ini kurang baik apabila terdapat korelasi yang besar antar prediktor (Friedman et al., 2010).
3. *Genomic* BLUP (GBLUP), merupakan model linier parametrik yang memanfaatkan matriks hubungan genomik (G) untuk menduga nilai pemuliaan. Matriks G dibangun melalui evaluasi kovarian marka antar individu, dimana matriks ini digunakan sebagai pengganti matriks hubungan aditif (A) yang disusun berdasarkan silsilah pada BLUP. Model GBLUP memiliki keunggulan dalam hal efisiensi komputasi yang lebih sederhana (Wang et al., 2015). Pendekatan GBLUP dapat membantu program pemuliaan tanaman semusim yang tidak memiliki informasi silsilah lengkap (Hayes et al., 2013).
4. Bayesian, merupakan model non-linier parametrik yang mengasumsikan efek marka menyebar *t-student*, ragam yang tidak sama pada setiap efek

marka, dan beberapa efek marka tertentu bernilai nol. Model Bayesian sangat baik untuk karakter yang dikendalikan oleh sejumlah kecil QTL dengan efek besar (Heffner et al., 2011). Pada awalnya, terdapat dua jenis model Bayesian yang diperkenalkan pada GS, yaitu BayesA dan BayesB (Meuwissen et al., 2001). BayesA mengasumsikan setiap marka memiliki efek tertentu dengan ragam yang berbeda-beda, dimana efek marka pada model Bayes A akan menyusut ke nilai nol, tetapi tidak mencapai nilai nol itu sendiri (Heffner et al., 2009). Di sisi lain, BayesB memiliki kemiripan dengan BayesA, tetapi model ini dapat mengasumsikan sejumlah efek marka bernilai nol. Asumsi ini lebih realistis karena tidak seluruh marka memiliki asosiasi dengan suatu QTL (Lin et al., 2014). Beberapa model Bayesian lain yang telah dikembangkan hingga saat ini yaitu Bayes-LASSO (Park & Casella, 2008), BayesC π , BayesD π (Habier et al., 2011), dan BayesR (Erbe et al., 2012).

5. *Partial least squares regression* (PLSR) dan *principal component regression* (PCR), merupakan model regresi yang dikembangkan dengan prinsip reduksi dimensional untuk mengatasi permasalahan "*large-p small-n*" dan multikolinearitas (da Costa et al., 2021). Model ini awalnya dikembangkan dalam bidang ilmu kemometrika untuk mengkonstruksi model prediksi suatu variabel yang hubungannya kurang dipahami (Tobias, 1995). Variabel laten digunakan sebagai penduga variabel, dimana variabel tersebut dipilih untuk menjelaskan sebanyak mungkin ragam dalam prediktor asli (pada PCR) atau memperkuat hubungan antara variabel terhadap variabel (pada PLSR) (Jannink et al., 2010).
6. Model semi-parametrik dan non-parametrik. Model ini juga dikembangkan untuk pendekatan GS di samping model parametrik. Lin et al. (2014) menjelaskan bahwa model semi- dan non-parametrik dapat menggambarkan pengaruh genetik non-aditif, seperti dominan dan epistasis. Kemampuan tersebut menjadi keunggulan apabila pengaruh genetik non-aditif memegang peranan penting dalam kendali suatu sifat. Salah satu pendekatan semi-parametrik yang umum digunakan adalah *reproducing kernel Hilbert space* (RKHS), yaitu dengan menambahkan fungsi kernel pada model untuk mengevaluasi jarak genomik

antar individu (Gianola et al., 2006). Selain RKHS, pendekatan *machine-learning* tergolong ke dalam metode non-parametrik yang umum digunakan pada GS, seperti *support vector regression* (SVR) (de Oliveira et al., 2014; W. Zhao et al., 2020), *random forest* (RF) (Annicchiarico et al., 2015; Hu et al., 2019), dan *artificial neural network* (NN) (Montesinos-López et al., 2021; Peters et al., 2020).

PEMANFAATAN SELEKSI GENOMIK PADA KELAPA SAWIT

Awal mula riset seleksi genomik pada kelapa sawit dilakukan oleh Wong & Bernardo (2008) yang melakukan simulasi perbandingan respon selama empat siklus akibat seleksi fenotipik, seleksi berulang berbantu marka (*marker-assisted recurrent selection*, MARS), dan GS. Simulasi dilakukan dengan beberapa pertimbangan, seperti (i) jumlah individu (N) dan ulangan (N_{Rep}) yang berkaitan dengan ukuran populasi; dan (ii) model genetik berupa jumlah QTL dan heritabilitas. Penelitian tersebut menunjukkan kemampuan GS dalam memberikan efisiensi biaya sebesar 26%-65% dan menghasilkan kemajuan genetik yang baik dibandingkan seleksi fenotipik dan MARS. Selain itu, kondisi ukuran populasi yang relatif kecil (N = 50 dengan dua ulangan) dapat memberikan respon seleksi yang baik. Hal ini menunjukkan potensi GS pada program pemuliaan kelapa sawit yang memiliki ukuran populasi kecil pada pengujian. Walau demikian, hasil simulasi tersebut tidak dapat diaplikasikan secara praktis pada program pemuliaan kelapa sawit, terkhusus pada program RRS. Hingga saat ini, beberapa penelitian terus dilakukan untuk menginvestigasi potensi GS pada kelapa sawit.

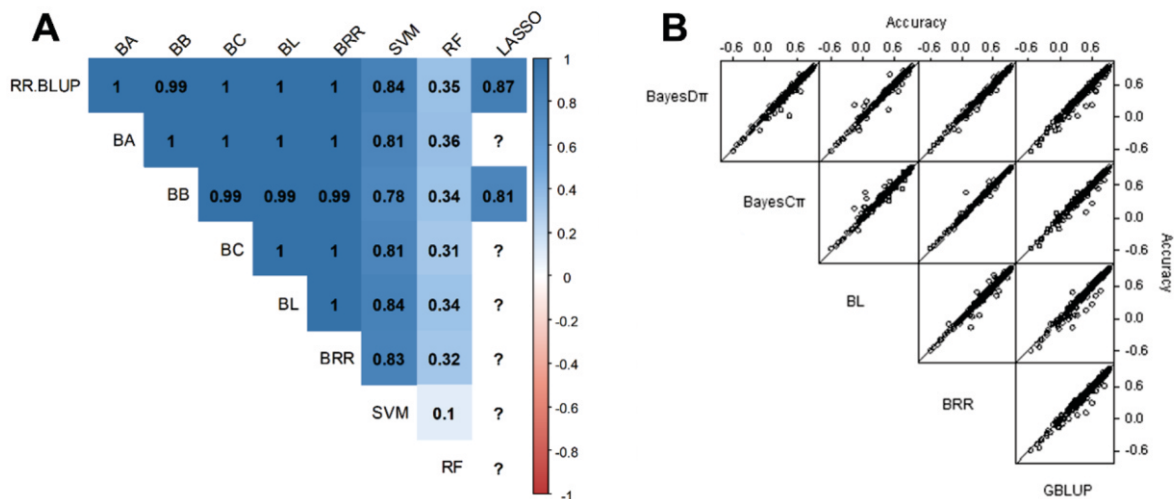
Penentuan tipe marka molekuler menjadi hal yang penting untuk memaksimalkan akurasi GS. Penelitian empiris GS kelapa sawit pertama kali dilakukan oleh Cros et al. (2015b) dengan menggunakan marka SSR, dimana tingkat akurasi prediksi yang dihasilkan berkisar antara 0,47 hingga 0,64 untuk karakter produksi dan kualitas tandan. Marchal et al. (2016) juga melakukan hal yang serupa, dimana akurasi prediksi untuk karakter produksi berada pada kisaran 0,61-0,72. Walaupun marka SSR menunjukkan tingkat akurasi yang cukup baik dalam prediksi model, marka tersebut memiliki kelemahan dalam hal penggunaan sumberdaya yang relatif tinggi. Penggunaan marka

high-throughput seperti SNP diperlukan untuk menurunkan biaya dan waktu *genotyping*. Pendekatan GS pada kelapa sawit terkini telah menggunakan marka SNP dengan keunggulan dalam hal kerapatan marka dan akurasi yang lebih baik (Cros et al., 2017; Kwong et al., 2017a; Kwong et al., 2017b; Nyouma et al., 2020). Pada kelapa sawit, kerapatan marka untuk mendapatkan tingkat akurasi prediksi yang baik tergolong relatif lebih rendah (± 500 SNP) (Cros et al., 2017) jika dibandingkan dengan kerapatan marka pada tanaman menyerbuk silang lainnya, misalnya tanaman jagung (Yu et al., 2023).

Investigasi pengaruh model statistik GS terhadap akurasi prediksi nilai pemuliaan telah banyak dilakukan pada kelapa sawit. Kwong et al. (2017a) dan Ithnin et al. (2017) menyelidiki tingkat akurasi prediksi beberapa model statistik berbasis BLUP, LASSO, dan Bayesian pada beberapa karakter kualitas tandan.

Selain itu, Kwong et al. (2017b) juga menambahkan model statistik yang berbasis non-parametrik. Tingkat akurasi pada setiap model statistik tidak berbeda ditunjukkan dengan koefisien korelasi antar model statistik yang tinggi (0,78-1,00), kecuali model *random forest* (Gambar 2a). Hal serupa juga ditemukan pada penelitian Cros et al. (2015b), dimana akurasi model Bayesian dan GBLUP tidak menunjukkan perbedaan dan memiliki koefisien korelasi akurasi antar model berkisar 0,982-0,995 (Gambar 2b). Temuan ini mengindikasikan model-model statistik tersebut dapat memprediksi karakter kuantitatif pada kelapa sawit dengan baik. Oleh karena itu, Cros et al. (2015b) dan Nyouma et al. (2019) mengusulkan penggunaan GBLUP sebagai model statistik untuk GS pada kelapa sawit dengan pertimbangan kemudahan komputasi pada perangkat lunak yang ada.

Strategi pelatihan model menjadi hal penting



Gambar 2. Korelasi akurasi prediksi antara model statistik GS pada karakter kualitas tandan kelapa sawit. (A) korelasi akurasi prediksi pada Ithnin et al. (2017), Kwong et al. (2017a), dan Kwong et al. (2017b); (B) korelasi akurasi prediksi pada Cros et al. (2015b).

untuk membangun model prediksi GEBV yang akurat, terkhusus pada kelapa sawit. Beberapa penelitian yang menginvestigasi pendekatan untuk membangun model prediksi telah dilakukan pada kelapa sawit (Tabel 1). Pendekatan pertama yang dilakukan adalah memprediksi nilai pemuliaan tetua pada masing-masing grup heterotik tetua secara terpisah. Strategi yang mengkombinasikan informasi

genotipe (molekuler) pada tetua dan data fenotipe tetua untuk pemodelan menunjukkan akurasi prediksi yang relatif rendah (0,03-0,17; Tabel 1). Hal ini diakibatkan oleh ketidakmampuan fenotipe tetua untuk merepresentasikan pewarisan sifat pada keturunannya. Strategi lain yaitu mengkombinasikan informasi genotipe dari tetua dan daya gabung umum (DGU) tetua untuk membangun model GS

(Cros et al., 2015b). Metode ini cukup baik digunakan karena menghasilkan akurasi prediksi yang baik pada Grup B (populasi Afrika), tetapi hal sebaliknya ditunjukkan pada Grup A (populasi Deli).

Pendekatan kedua yang diajukan adalah menggabungkan dua grup tetua pada analisis secara simultan, sehingga pendekatan tersebut lebih relevan dengan skema RRS yang dilakukan pada kelapa sawit. Pendekatan ini memanfaatkan informasi molekuler dari tetua persilangan atau hibrida (atau kombinasinya) dan fenotipe hibrida keturunannya. Pendekatan ini menghasilkan akurasi

prediksi yang sangat baik dibandingkan pendekatan sebelumnya. Selain itu, kombinasi informasi tetua dan hibrida memberikan akurasi prediksi yang relatif lebih baik dibandingkan dengan hanya menggunakan informasi genotipe tetua (Tabel 1). Dalam rangka meningkatkan akurasi prediksi, Nyouma et al. (2022) meneliti pengaruh asal grup tetua dari alel yang diwariskan. Model dengan asumsi efek alel yang sama pada Grup A dan Grup B dapat meningkatkan akurasi prediksi sebesar 3% dibandingkan dengan asumsi yang membedakan efek alel dari Grup A dan Grup B.

Tabel 1. Perbandingan akurasi prediksi beberapa pendekatan pelatihan model GS pada kelapa sawit

Pendekatan	Informasi genotipe	Data	Akurasi prediksi	Sumber
Grup tetua secara terpisah	Tetua	Fenotipe tetua	0,03-0,17	Ithnin et al. (2017)
		DGU tetua	0,18-0,43	Kwong et al. (2017b)
			0,28-0,71	Cros et al. (2015b)
Gabungan grup tetua	Tetua		0,27-0,78	Cros et al. (2017)
		Fenotipe hibrida	0,61-0,72	Marchal et al. (2016)
	Hibrida		0,16-0,88	Nyouma et al. (2022)
		Fenotipe hibrida	0,43-0,75	Kwong et al. (2017a)
		0,17-0,66	Nyouma et al. (2020)	
	Tetua + hibrida	Fenotipe hibrida	0,21-0,89	Nyouma et al. (2022)

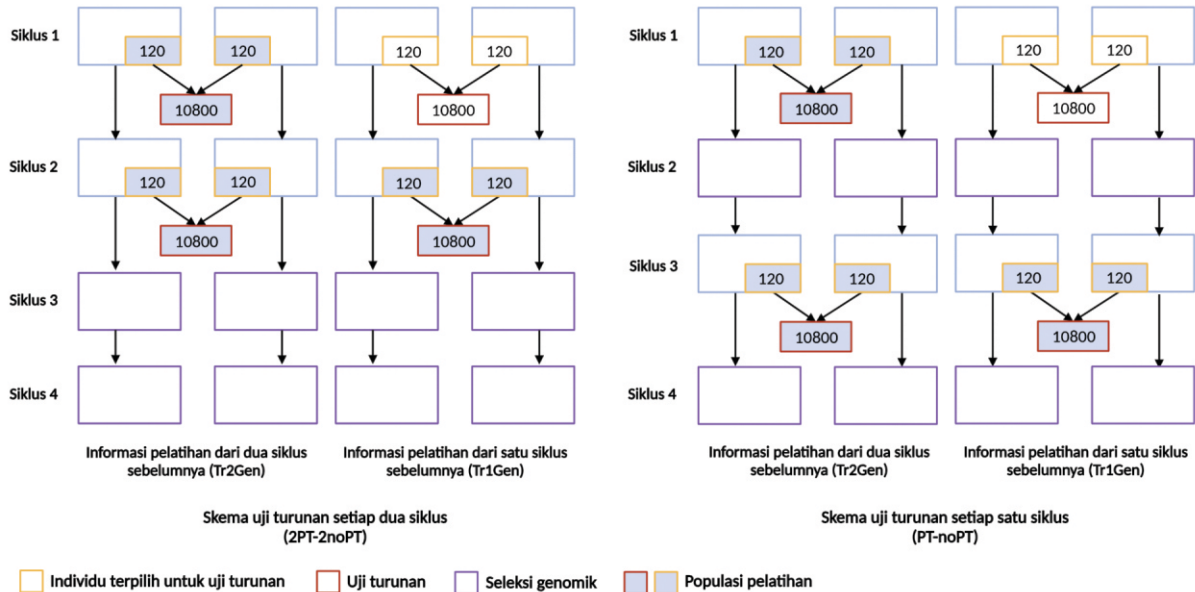
Skema RRS pada kelapa sawit memiliki permasalahan berupa intensitas seleksi yang rendah dan siklus waktu yang panjang. Oleh karena itu, GS diharapkan dapat menanggulangi permasalahan tersebut. Cros et al. (2015b) menelusuri peran GS pada program RRS kelapa sawit, dimana akurasi prediksi pada populasi Deli relatif rendah akibat sempitnya keragaman genetik pada populasi tersebut yang selanjutnya menyebabkan tidak terpenuhinya prinsip *mendelian sampling*. Hal ini menyebabkan potensi GS dalam mempersingkat siklus tidak dapat dilakukan secara optimal. Oleh karena itu, penelitian ini mengajukan metode pra-seleksi genomik pada karakter komponen produksi untuk memilih tetua sebelum pengujian turunan. Validasi metode tersebut dilakukan oleh Cros et al. (2017), dimana pra-seleksi genomik sebelum pengujian turunan dapat

meningkatkan produksi tandan buah segar hingga 10% pada hibrida yang dibentuk akibat peningkatan intensitas seleksi. Hal serupa juga ditemukan oleh Nyouma et al. (2020) pada seleksi ortet sebelum percobaan klonal tenera elit.

Pencarian solusi untuk permasalahan siklus waktu yang panjang pada RRS terus dilakukan dalam rangka optimalisasi potensi GS pada kelapa sawit. Cros et al. (2018) melakukan simulasi selama empat siklus dengan beberapa pendekatan, yaitu (1) skema GS, dimana GS dilakukan setiap siklus (PT-noPT) atau dua siklus (2PT-2noPT) berseling dengan seleksi fenotipik (*phenotypic selection*, PS); dan (2) penggunaan informasi model pelatihan dari satu siklus (Tr1Gen) atau dua siklus sebelumnya (Tr2Gen) (Gambar 3). Hasil simulasi menunjukkan

bahwa pendekatan PT-noPT dan Tr2Gen memberikan peningkatan kemajuan genetik tahunan yang baik (36,6%-57,5%) serta mampu mengurangi waktu untuk menyelesaikan empat siklus pemuliaan hingga 35%. Hasil simulasi tersebut

mengimplikasikan adanya potensi untuk mengurangi siklus pemuliaan kelapa sawit yang cukup panjang. Oleh karena itu, penelitian lebih mendalam sangat diperlukan untuk validasi hasil simulasi tersebut.



Gambar 3. Skema strategi integrasi GS pada RRS untuk menanggulangi waktu pemuliaan (Diadaptasi dari Cros et al., 2018)

KESIMPULAN

Keberhasilan pendekatan GS pada pemuliaan ternak dan pemuliaan tanaman lainnya telah menunjukkan adanya potensi penerapan GS pada kelapa sawit. Permasalahan intensitas seleksi yang rendah dapat ditanggulangi melalui penerapan pra-seleksi genomik. Selain itu, GS juga berpotensi untuk mengurangi siklus pemuliaan kelapa sawit melalui peniadaan pengujian turunan. Adaptasi pendekatan-pendekatan yang telah diteliti dapat dilakukan untuk melihat respon pada bahan genetik yang berbeda. Selain itu, optimasi perlu dilakukan pada beberapa aspek seperti pendekatan pelatihan model, keterkaitan dan ukuran populasi, metode *genotyping*, serta karakter-karakter lainnya (vegetatif, toleransi cekaman).

DAFTAR PUSTAKA

- Abdollahi-Arpanahi, R., Pakdel, A., Nejati-Javaremi, A., & Shahrbabak, M. M. (2013). Comparison of genomic evaluation methods in complex traits with different genetic architecture. *Animal Production*, 15(1), 65–77. <https://doi.org/10.22059/jap.2013.36604>
- Adhikari, S., Saha, S., Biswas, A., Rana, T. S., Bandyopadhyay, T. K., & Ghosh, P. (2017). Application of molecular markers in plant genome analysis: a review. *The Nucleus* 2017 60 : 3 , 60 (3) , 283 – 297 . <https://doi.org/10.1007/S13237-017-0214-7>
- Ahmadi, Z., Ghafouri-Kesbi, F., & Zamani, P. (2021). Assessing the performance of a novel method for genomic selection: rrBLUP-method6. *Journal of Genetics*, 100(2), 24.

- <https://doi.org/10.1007/s12041-021-01275-5>
- Annicchiarico, P., Nazzicari, N., Li, X., Wei, Y., Pecetti, L., & Brummer, E. C. (2015). Accuracy of genomic selection for alfalfa biomass yield in different reference populations. *BMC Genomics*, *16*(1), 1020. <https://doi.org/10.1186/s12864-015-2212-y>
- Ariesca, R., Sau, A. A. W. T., Adinugroho, W. C., Setiawan, A. A. R., Ahamed, T., & Noguchi, R. (2023). Land swap option for sustainable production of oil palm plantations in Kalimantan, Indonesia. *Sustainability*, *15*(3), 2394. <https://doi.org/10.3390/su15032394>
- Ashoori-Banaei, S., Ghafouri-Kesbi, F., & Ahmadi, A. (2021). Comparison of regression tree-based methods in genomic selection. *Journal of Genetics*, *100*(2), 85. <https://doi.org/10.1007/s12041-021-01334-x>
- Budhlakoti, N., Kushwaha, A. K., Rai, A., Chaturvedi, K. K., Kumar, A., Pradhan, A. K., Kumar, U., Kumar, R. R., Juliana, P., Mishra, D. C., & Kumar, S. (2022). Genomic selection: a tool for accelerating the efficiency of molecular breeding for development of climate-resilient crops. *Frontiers in Genetics*, *13*, 832153. <https://doi.org/10.3389/fgene.2022.832153>
- Cao, S., Loladze, A., Yuan, Y., Wu, Y., Zhang, A., Chen, J., Huestis, G., Cao, J., Chaikam, V., Olsen, M., Prasanna, B. M., Vicente, F. S., & Zhang, X. (2017). Genome-wide analysis of tar spot complex resistance in maize using genotyping-by-sequencing SNPs and whole-genome prediction. *The Plant Genome*, *10*(2). <https://doi.org/10.3835/PLANTGENOME2016.10.0099>
- Cericola, F., Jahoor, A., Orabi, J., Andersen, J. R., Janss, L. L., & Jensen, J. (2017). Optimizing training population size and genotyping strategy for genomic prediction using association study results and pedigree information: a case of study in advanced wheat breeding lines. *PLOS ONE*, *12*(1), e0169606. <https://doi.org/10.1371/journal.pone.0169606>
- Corley, R. H. V., & Tinker, P. B. (2016). The Oil Palm: Fifth Edition. *The Oil Palm: Fifth Edition*. <https://doi.org/10.1002/9781118953297>
- Cros, D., Bocs, S., Riou, V., Ortega-Abboud, E., Tisné, S., Argout, X., Pomiès, V., Nodichao, L., Lubis, Z., Cochard, B., & Durand-Gasselin, T. (2017). Genomic preselection with genotyping-by-sequencing increases performance of commercial oil palm hybrid crosses. *BMC Genomics*, *18*(1), 839. <https://doi.org/10.1186/s12864-017-4179-3>
- Cros, D., Denis, M., Bouvet, J.-M., & Sánchez, L. (2015a). Long-term genomic selection for heterosis without dominance in multiplicative traits: case study of bunch production in oil palm. *BMC Genomics*, *16*(1), 651. <https://doi.org/10.1186/s12864-015-1866-9>
- Cros, D., Denis, M., Sánchez, L., Cochard, B., Flori, A., Durand-Gasselin, T., Nouy, B., Omoré, A., Pomiès, V., Riou, V., Suryana, E., & Bouvet, J.-M. (2015b). Genomic selection prediction accuracy in a perennial crop: case study of oil palm (*Elaeis guineensis* Jacq.). *Theoretical and Applied Genetics*, *128*(3), 397–410. <https://doi.org/10.1007/s00122-014-2439-z>
- Cros, D., Mbo-Nkoulou, L., Bell, J. M., Oum, J., Masson, A., Soumahoro, M., Tran, D. M., Achour, Z., Le Guen, V., & Clement-Demange, A. (2019). Within-family genomic selection in rubber tree (*Hevea brasiliensis*) increases genetic gain for rubber production. *Industrial Crops and Products*, *138*, 111464. <https://doi.org/10.1016/j.indcrop.2019.111464>
- Cros, D., Tchounke, B., & Nkague-Nkamba, L. (2018). Training genomic selection models across several breeding cycles increases genetic gain in oil palm in silico study. *Molecular Breeding*, *38*(7), 89. <https://doi.org/10.1007/s11032-018-0850-x>
- da Costa, J. A., Azevedo, C. F., Nascimento, M., E Silva, F. F., de Resende, M. D. V., & Nascimento, A. C. C. (2021). A comparison of regression methods based on dimensional reduction for genomic prediction. *Genetics and Molecular Research*, *20*(2). <https://doi.org/10.4238/GMR18877>
- Davoudi, P., Abdollahi-Arpanahi, R., & Nejati-Javaremi, A. (2018). The impact of QTL allele frequency distribution on the accuracy of genomic prediction. *Archives Animal Breeding*,

- 61(2), 207–213. <https://doi.org/10.5194/AAB-61-207-2018>
- de Oliveira, F. C., Borges, C. C. H., Almeida, F. N., e Silva, F. F., da Silva Verneque, R., da Silva, M. V. G., & Arbex, W. (2014). SNPs selection using support vector regression and genetic algorithms in GWAS. *BMC Genomics*, 15(S7), S4. <https://doi.org/10.1186/1471-2164-15-S7-S4>
- Durán, R., Isik, F., Zapata-Valenzuela, J., Balocchi, C., & Valenzuela, S. (2017). Genomic predictions of breeding values in a cloned Eucalyptus globulus population in Chile. *Tree Genetics and Genomes*, 13(4), 1–12. <https://doi.org/10.1007/S11295-017-1158-4/METRICS>
- Edwards, S. M., Buntjer, J. B., Jackson, R., Bentley, A. R., Lage, J., Byrne, E., Burt, C., Jack, P., Berry, S., Flatman, E., Poupard, B., Smith, S., Hayes, C., Gaynor, R. C., Gorjanc, G., Howell, P., Ober, E., Mackay, I. J., & Hickey, J. M. (2019). The effects of training population design on genomic prediction accuracy in wheat. *Theoretical and Applied Genetics*. <https://doi.org/10.1007/s00122-019-03327-y>
- Elias, A. A. (2022). Genomic selection models for plant breeding schemes: the power of choice. In A. A. Elias & S. Goel (Eds.), *Genomic Selection in Plants: A Guide for Breeders* (pp. 1–22). CRC Press.
- Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., Mason, B. A., & Goddard, M. E. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science*, 95(7), 4114–4129. <https://doi.org/10.3168/jds.2011-5019>
- FAOSTAT. (2020). *Food and agricultural data*. Food and Agriculture Organization of the United Nations. <https://www.fao.org/faostat/en/#home>
- Fraslin, C., Yáñez, J. M., Robledo, D., & Houston, R. D. (2022). The impact of genetic relationship between training and validation populations on genomic prediction accuracy in Atlantic salmon. *Aquaculture Reports*, 23, 101033. <https://doi.org/10.1016/j.aqrep.2022.101033>
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22.
- Gao, N., Li, J., He, J., Xiao, G., Luo, Y., Zhang, H., Chen, Z., & Zhang, Z. (2015). Improving accuracy of genomic prediction by genetic architecture based priors in a Bayesian model. *BMC Genetics*, 16(1), 120. <https://doi.org/10.1186/s12863-015-0278-9>
- Garzón-Martínez, G. A., Osorio-Guarín, J. A., Moreno, L. P., Bastidas, S., Barrero, L. S., Lopez-Cruz, M., & Enciso-Rodríguez, F. E. (2022). Genomic selection for morphological and yield-related traits using genome-wide SNPs in oil palm. *Molecular Breeding*, 42(12), 71. <https://doi.org/10.1007/s11032-022-01341-5>
- Gianola, D., Fernando, R. L., & Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*, 173(3), 1761–1776. <https://doi.org/10.1534/genetics.105.049510>
- Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*, 124(6), 323–330. <https://doi.org/10.1111/J.1439-0388.2007.00702.X>
- Goggin, K. A., & Murphy, D. J. (2018). Monitoring the traceability, safety and authenticity of imported palm oils in Europe. *OCL*, 25(6), A603. <https://doi.org/10.1051/ocl/2018059>
- Gouy, M., Rousselle, Y., Bastianelli, D., Lecomte, P., Bonnal, L., Roques, D., Efile, J.-C., Rocher, S., Daugrois, J., Toubi, L., Nabeneza, S., Hervouet, C., Telismart, H., Denis, M., Thong-Chane, A., Glaszmann, J. C., Hoarau, J.-Y., Nibouche, S., & Costet, L. (2013). Experimental assessment of the accuracy of genomic selection in sugarcane. *Theoretical and Applied Genetics*, 126(10), 2575–2586. <https://doi.org/10.1007/s00122-013-2156-z>
- Grattapaglia, D. (2017). Status and perspectives of genomic selection in forest tree breeding. In R. K. Varshney, M. Roorkiwal, & M. E. Sorrells (Eds.), *Genomic Selection for Crop Improvement* (pp. 199–249). Springer International Publishing.

- https://doi.org/10.1007/978-3-319-63170-7_9
- Grattapaglia, D., Silva-Junior, O. B., Resende, R. T., Cappa, E. P., Müller, B. S. F., Tan, B., Isik, F., Ratcliffe, B., & El-Kassaby, Y. A. (2018). Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science*, 9.
- <https://doi.org/10.3389/fpls.2018.01693>
- Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12(1), 186. <https://doi.org/10.1186/1471-2105-12-186>
- Hao, Y., Wang, H., Yang, X., Zhang, H., He, C., Li, D., Li, H., Wang, G., Wang, J., & Fu, J. (2019). Genomic prediction using existing historical data contributing to selection in biparental populations: a study of kernel oil in maize. *The Plant Genome*, 12(1), 180025. <https://doi.org/10.3835/plantgenome2018.05.0025>
- Hayes, B. J., Cogan, N. O. I., Pembleton, L. W., Goddard, M. E., Wang, J., Spangenberg, G. C., & Forster, J. W. (2013). Prospects for genomic selection in forage plant species. *Plant Breeding*, 132(2), 133–143. <https://doi.org/10.1111/pbr.12037>
- Heffner, E. L., Jannink, J., & Sorrells, M. E. (2011). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome*, 4(1), plantgenome2010.12.0029. <https://doi.org/10.3835/plantgenome2010.12.0029>
- Heffner, E. L., Lorenz, A. J., Jannink, J., & Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Science*, 50(5), 1681–1690. <https://doi.org/10.2135/cropsci2009.11.0662>
- Heffner, E. L., Sorrells, M. E., & Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Science*, 49(1), 1–12. <https://doi.org/10.2135/cropsci2008.08.0512>
- Hu, X., Carver, B. F., Powers, C., Yan, L., Zhu, L., & Chen, C. (2019). Effectiveness of Genomic Selection by Response to Selection for Winter Wheat Variety Improvement. *The Plant Genome*, 12(3). <https://doi.org/10.3835/plantgenome2018.11.0090>
- Ithnin, M., Xu, Y., Marjuni, M., Serdari, N. M., Amiruddin, M. D., Low, E.-T. L., Tan, Y.-C., Yap, S.-J., Ooi, L. C. L., Nookiah, R., Singh, R., & Xu, S. (2017). Multiple locus genome-wide association studies for important economic traits of oil palm. *Tree Genetics & Genomes*, 13(5), 103. <https://doi.org/10.1007/s11295-017-1185-1>
- Jannink, J.-L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics*, 9(2), 166–177. <https://doi.org/10.1093/bfgp/elq001>
- Juliana, P., Poland, J., Huerta-Espino, J., Shrestha, S., Crossa, J., Crespo-Herrera, L., Toledo, F. H., Govindan, V., Mondal, S., Kumar, U., Bhavani, S., Singh, P. K., Randhawa, M. S., He, X., Guzman, C., Dreisigacker, S., Rouse, M. N., Jin, Y., Pérez-Rodríguez, P., ... Singh, R. P. (2019). Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nature Genetics* 2019 51:10, 51(10), 1530–1539. <https://doi.org/10.1038/s41588-019-0496-6>
- Kadam, D. C., Potts, S. M., Bohn, M. O., Lipka, A. E., & Lorenz, A. J. (2016). Genomic prediction of single crosses in the early stages of a maize hybrid breeding pipeline. *G3 Genes Genomes Genetics*, 6(11), 3443–3453. <https://doi.org/10.1534/g3.116.031286>
- Kasnavi, S. A., Aminafshar, M., Shariati, M. M., Emam Jomeh Kashan, N., & Honarvar, M. (2018). The effect of kernel selection on genome wide prediction of discrete traits by Support Vector Machine. *Gene Reports*, 11, 279–282. <https://doi.org/10.1016/J.GENREP.2018.04.006>
- Kizilkaya, K., Fernando, R. L., & Garrick, D. J. (2014). Reduction in accuracy of genomic prediction for ordered categorical data compared to continuous observations. *Genetics Selection Evolution*, 46(1), 1–13. <https://doi.org/10.1186/1297-9686-46-37/TABLES/5>
- Klug, W. S., Cummings, M. R., Spencer, C. A.,

- Palladino, M. A., & Killian, D. J. (2018). *Concept of Genetics* (Twelfth Edition). Pearson Education.
- Kriaridou, C., Tsairidou, S., Houston, R. D., & Robledo, D. (2020). Genomic prediction using low density marker panels in aquaculture: performance across species, traits, and genotyping platforms. *Frontiers in Genetics, 11*. <https://doi.org/10.3389/fgene.2020.00124>
- Kwong, Q. B., Ong, A. L., Teh, C. K., Chew, F. T., Tammi, M., Mayes, S., Kulaveerasingam, H., Yeoh, S. H., Harikrishna, J. A., & Appleton, D. R. (2017a). Genomic selection in commercial perennial crops: applicability and improvement in oil palm (*Elaeis guineensis* Jacq.). *Scientific Reports, 7*(1), 2872. <https://doi.org/10.1038/s41598-017-02602-6>
- Kwong, Q. B., Teh, C. K., Ong, A. L., Chew, F. T., Mayes, S., Kulaveerasingam, H., Tammi, M., Yeoh, S. H., Appleton, D. R., & Harikrishna, J. A. (2017b). Evaluation of methods and marker systems in genomic selection of oil palm (*Elaeis guineensis* Jacq.). *BMC Genetics, 18*(1), 107. <https://doi.org/10.1186/s12863-017-0576-5>
- Lan, S., Zheng, C., Hauck, K., McCausland, M., Duguid, S. D., Booker, H. M., Cloutier, S., & You, F. M. (2020). Genomic prediction accuracy of seven breeding selection traits improved by QTL identification in flax. *International Journal of Molecular Sciences, 21*(5), 1577. <https://doi.org/10.3390/ijms21051577>
- Lee, S. H., Clark, S., & van der Werf, J. H. J. (2017). Estimation of genomic prediction accuracy from reference populations with varying degrees of relationship. *PLOS ONE, 12*(12), e0189775. <https://doi.org/10.1371/journal.pone.0189775>
- Lin, Z., Hayes, B. J., & Daetwyler, H. D. (2014). Genomic selection in crops, trees and forages: a review. *Crop and Pasture Science, 65*(11), 1177. <https://doi.org/10.1071/CP13363>
- Liu, H., Zhou, H., Wu, Y., Li, X., Zhao, J., Zuo, T., Zhang, X., Zhang, Y., Liu, S., Shen, Y., Lin, H., Zhang, Z., Huang, K., Lübberstedt, T., & Pan, G. (2015). The Impact of Genetic Relationship and Linkage Disequilibrium on Genomic Selection. *PLOS ONE, 10*(7), e0132379. <https://doi.org/10.1371/journal.pone.0132379>
- Liu, X., Wang, H., Wang, H., Guo, Z., Xu, X., Liu, J., Wang, S., Li, W. X., Zou, C., Prasanna, B. M., Olsen, M. S., Huang, C., & Xu, Y. (2018). Factors affecting genomic selection revealed by empirical evidence in maize. *The Crop Journal, 6*(4), 341–352. <https://doi.org/10.1016/J.CJ.2018.03.005>
- Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., Smith, K. P., Sorrells, M. E., & Jannink, J.-L. (2011). *Genomic selection in plant breeding* (pp. 77–123). <https://doi.org/10.1016/B978-0-12-385531-2.00002-5>
- Lozada, D. N., Mason, R. E., Sarinelli, J. M., & Brown-Guedira, G. (2019). Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. *BMC Genetics, 20*(1), 82. <https://doi.org/10.1186/s12863-019-0785-1>
- Marchal, A., Legarra, A., Tisné, S., Carasco-Lacombe, C., Manez, A., Suryana, E., Omoré, A., Nouy, B., Durand-Gasselín, T., Sánchez, L., Bouvet, J.-M., & Cros, D. (2016). Multivariate genomic model improves analysis of oil palm (*Elaeis guineensis* Jacq.) progeny tests. *Molecular Breeding, 36*(1), 2. <https://doi.org/10.1007/s11032-015-0423-1>
- Meunier, J., & Gascon, J. (1972). Le schéma général d'amélioration du palmier à huile à l'IRHO. *Oléagineux, 27*, 1–12.
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819–1829. <https://doi.org/10.1093/GENETICS/157.4.1819>
- Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W. R., Fajardo-Flores, S. B., Gaytan-Lugo, L. S., Santana-Mancilla, P. C., & Crossa, J. (2021). A review of deep learning applications for genomic selection. *BMC Genomics, 22*(1), 19. <https://doi.org/10.1186/s12864-020-07319-x>
- Muranty, H., Jorge, V., Bastien, C., Lepoittevin, C., Bouffier, L., & Sanchez, L. (2014). Potential for marker-assisted selection for forest tree

- breeding: lessons from 20 years of MAS in crops. *Tree Genetics and Genomes*, 10(6), 1491–1510. <https://doi.org/10.1007/S11295-014-0790-5/METRICS>
- Muranty, H., Troggio, M., Sadok, I. Ben, Rifaï, M. Al, Auwerkerken, A., Banchi, E., Velasco, R., Stevanato, P., van de Weg, W. E., Di Guardo, M., Kumar, S., Laurens, F., & Bink, M. C. A. M. (2015). Accuracy and responses of genomic selection on key traits in apple breeding. *Horticulture Research*, 2(1), 15060. <https://doi.org/10.1038/hortres.2015.60>
- Nakaya, A., & Isobe, S. N. (2012). Will genomic selection be a practical method for plant breeding? *Annals of Botany*, 110(6), 1303–1316. <https://doi.org/10.1093/aob/mcs109>
- Nyouma, A., Bell, J. M., Jacob, F., & Cros, D. (2019). From mass selection to genomic selection: one century of breeding for quantitative yield components of oil palm (*Elaeis guineensis* Jacq.). *Tree Genetics & Genomes*, 15(5), 69. <https://doi.org/10.1007/s11295-019-1373-2>
- Nyouma, A., Bell, J. M., Jacob, F., Riou, V., Manez, A., Pomiès, V., Domonhede, H., Arifiyanto, D., Cochard, B., Durand-Gasselín, T., & Cros, D. (2022). Improving the accuracy of genomic predictions in an outcrossing species with hybrid cultivars between heterozygote parents: a case study of oil palm (*Elaeis guineensis* Jacq.). *Molecular Genetics and Genomics*, 297(2), 523–533. <https://doi.org/10.1007/s00438-022-01867-5>
- Nyouma, A., Bell, J. M., Jacob, F., Riou, V., Manez, A., Pomiès, V., Nodichao, L., Syahputra, I., Affandi, D., Cochard, B., Durand-Gasselín, T., & Cros, D. (2020). Genomic predictions improve clonal selection in oil palm (*Elaeis guineensis* Jacq.) hybrids. *Plant Science*, 299, 110547. <https://doi.org/10.1016/j.plantsci.2020.110547>
- Park, T., & Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482), 681–686. <https://doi.org/10.1198/016214508000000337>
- Peters, S. O., Sinecen, M., Kizilkaya, K., & Thomas, M. G. (2020). Genomic Prediction With Different Heritability, QTL, and SNP Panel Scenarios Using Artificial Neural Network. *IEEE Access*, 8, 147995–148006. <https://doi.org/10.1109/ACCESS.2020.3015814>
- Piepho, H. P., Möhring, J., Melchinger, A. E., & Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, 161(1–2), 209–228. <https://doi.org/10.1007/S10681-007-9449-8/METRICS>
- Ramakrishnan, A. P. (2013). Linkage disequilibrium. *Brenner's Encyclopedia of Genetics: Second Edition*, 252–253. <https://doi.org/10.1016/B978-0-12-374984-0.00870-6>
- Sallam, A. H., Conley, E., Prakapenka, D., Da, Y., & Anderson, J. A. (2020). Improving Prediction Accuracy Using Multi-allelic Haplotype Prediction and Training Population Optimization in Wheat. *G3 Genes|Genomes|Genetics*, 10(7), 2265–2273. <https://doi.org/10.1534/g3.120.401165>
- Soh, A. C., Mayes, S., Roberts, J., Cros, D., & Purba, R. (2017). Breeding plans and selection methods. In A. C. Soh, S. Mayes, & J. Roberts (Eds.), *Oil Palm Breeding: Genetics and Genomics* (pp. 143–164). CRC Press. <https://doi.org/10.1201/9781315119724-6>
- Sood, S., Bhardwaj, V., Kaushik, S. K., & Sharma, S. (2020). Prediction based on estimated breeding values using genealogy for tuber yield and late blight resistance in auto-tetraploid potato (*Solanum tuberosum* L.). *Heliyon*, 6(11), e05624. <https://doi.org/10.1016/j.heliyon.2020.e05624>
- Sun, X., Fernando, R., & Dekkers, J. (2016). Contributions of linkage disequilibrium and co-segregation information to the accuracy of genomic prediction. *Genetics Selection Evolution*, 48(1), 77. <https://doi.org/10.1186/s12711-016-0255-4>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tobias, R. D. (1995). An introduction to partial least

- squares regression. In H. Neil (Ed.), *Proceedings of the twentieth annual SAS users group international conference* (pp. 1250–1257). SAS Institute Inc.
- Wang, C. L., Ding, X. D., Wang, J. Y., Liu, J. F., Fu, W. X., Zhang, Z., Yin, Z. J., & Zhang, Q. (2013). Bayesian methods for estimating GEBVs of threshold traits. *Heredity*, *110*(3), 213–219. <https://doi.org/10.1038/hdy.2012.65>
- Wang, X., Xu, Y., Hu, Z., & Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal*, *6*(4), 330–340. <https://doi.org/10.1016/j.cj.2018.03.001>
- Wang, X., Yang, Z., & Xu, C. (2015). A comparison of genomic selection methods for breeding value prediction. *Science Bulletin*, *60*(10), 925–935. <https://doi.org/10.1007/s11434-015-0791-2>
- Wientjes, Y., Veerkamp, R. F., Bijma, P., Bovenhuis, H., Schrooten, C., & Calus, M. (2015). Empirical and deterministic accuracies of across-population genomic prediction. *Genetics Selection Evolution*, *47*(1), 5. <https://doi.org/10.1186/s12711-014-0086-0>
- Wiggans, G. R., Cole, J. B., Hubbard, S. M., & Sonstegard, T. S. (2017). Genomic selection in dairy cattle: the USDA experience. *Annual Review of Animal Biosciences*, *5*(1), 309–327. <https://doi.org/10.1146/annurev-animal-021815-111422>
- Wong, C. K., & Bernardo, R. (2008). Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theoretical and Applied Genetics*, *116*(6), 815–824. <https://doi.org/10.1007/s00122-008-0715-5>
- Xu, Y., Wang, X., Ding, X., Zheng, X., Yang, Z., Xu, C., & Hu, Z. (2018). Genomic selection of agronomic traits in hybrid rice using an NCII population. *Rice*, *11*(1), 1–10. <https://doi.org/10.1186/S12284-018-0223-4/FIGURES/5>
- Yadav, S., Jackson, P., Wei, X., Ross, E. M., Aitken, K., Deomano, E., Atkin, F., Hayes, B. J., & Voss-Fels, K. P. (2020). Accelerating genetic gain in sugarcane breeding using genomic selection. *Agronomy*, *10*(4), 585. <https://doi.org/10.3390/agronomy10040585>
- Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*, 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>
- Yu, G., Cui, Y., Jiao, Y., Zhou, K., Wang, X., Yang, W., Xu, Y., Yang, K., Zhang, X., Li, P., Yang, Z., Xu, Y., & Xu, C. (2023). Comparison of sequencing-based and array-based genotyping platforms for genomic prediction of maize hybrid performance. *The Crop Journal*, *11*(2), 490–498. <https://doi.org/10.1016/j.cj.2022.09.004>
- Yusop, M. R., Sukaimi, J., Amiruddin, M. D., Jalloh, M., Swaray, S., Yusuff, O., & Chukwu, S. C. (2020). *Genetic improvement of oil palm through recurrent selection*. 35–46. https://doi.org/10.1007/978-3-030-22549-0_4
- Zhang, H., Yin, L., Wang, M., Yuan, X., & Liu, X. (2019). Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. *Frontiers in Genetics*, *10*(MAR), 441967. <https://doi.org/10.3389/FGENE.2019.00189/BIOTEX>
- Zhang, Z., Erbe, M., He, J., Ober, U., Gao, N., Zhang, H., Simianer, H., & Li, J. (2015). Accuracy of whole-genome prediction using a genetic architecture-enhanced variance-covariance matrix. *G3 Genes|Genomes|Genetics*, *5*(4), 615–627. <https://doi.org/10.1534/g3.114.016261>
- Zhao, W., Lai, X., Liu, D., Zhang, Z., Ma, P., Wang, Q., Zhang, Z., & Pan, Y. (2020). Applications of Support Vector Machine in Genomic Prediction in Pig and Maize Populations. *Frontiers in Genetics*, *11*. <https://doi.org/10.3389/fgene.2020.598318>
- Zhao, Y., Zeng, J., Fernando, R., & Reif, J. C. (2013). Genomic prediction of hybrid wheat performance. *Crop Science*, *53*(3), 802–810. <https://doi.org/10.2135/cropsci2012.08.0463>

