Industrial Crops & Products 228 (2025) 120908

Contents lists available at ScienceDirect

Industrial Crops & Products

journal homepage: www.elsevier.com/locate/indcrop

Enhancing genomic selection in rubber tree (*Hevea brasiliensis*): Exploring the impact of genetic relatedness and QTL integration

Daouda Koffi Kouassi[a], Aurélie Daval[b,c,d], Vincent Le Guen[c,d], André Clément-Demange[c,d], David Lopez[c,d], Pierre Mournet[c,d], Florelle Bonal[c,d], Jean-Luc Hofs[a], Mouman Soumahoro[f], Doffou Sélastique Akaffou[e], David Cros[c,d,*]

[a] *SOGB, SOCFIN, San Pedro, Côte d'Ivoire*
[b] *CIRAD, UMR AGAP Institut, San Pedro, Côte d'Ivoire*
[c] *CIRAD, UMR AGAP Institut, Montpellier F-34398, France*
[d] *UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France*
[e] *UFR Agroforesterie, Univ Jean Lorougnon Guédé, Daloa, Côte d'Ivoire*
[f] *SAPH (Société Africaine de Plantations d'Hévéas), Abidjan, Côte d'Ivoire*

ARTICLE INFO

Keywords:
Genomic selection
Rubber tree
Prediction accuracy
QTL integration
Full-sib families
Half-sib families
Training populations

ABSTRACT

Rubber tree breeding faces significant challenges, mainly due to the low female fertility of the plant, the long breeding cycles and the complex trait architecture. The advent of genomic selection offers a significant opportunity to explore more effective breeding strategies and accelerate genetic gains. This study investigates genomic prediction strategies for *Hevea brasiliensis* breeding, focusing on two biparental families connected by a common parent and evaluated among four different sites. The objective was to assess the impact of using full-sib and half-sib populations on prediction accuracy for key traits, rubber production and sucrose content. Results confirmed that prediction accuracies are higher when the training and validation populations consist of full-sibs (0.54 for rubber production and 0.36 for sucrose), as compared to half-sibs (0.17 for rubber production and 0.21 for sucrose). Combining full-sibs and half-sibs in the training population yielded prediction performance comparable to intra-family models (0.52 for rubber production and 0.37 for sucrose), providing a more practical option for breeding programs. Additionally, the integration of QTL information into prediction models for rubber production did not improved accuracy in full-sib (0.53) or half-sib (0.16) validation approaches and reduced accuracy in cross-validation, likely due to the polygenic nature of the trait and genotype-by-environment interactions. Rubber tree breeding programs could benefit from constructing training populations composed of multiple related families, simplifying logistics while maintaining prediction accuracy across more diverse populations than single biparental families. This approach offers a promising pathway to enhance the efficiency and genetic gains in rubber tree genomic selection.

1. Introduction

Hevea brasiliensis is the most efficient cultivated species for the production of natural rubber (NR, 1,4-cis-polyisoprene polymer), grown on an area of around 14 million hectares worldwide with a global production of 14.5 million tonnes in 2023, and widely used in the tyre industry (FAO, 2022). Overall, Southeast Asia currently account for 87 % of NR production, with Thailand, Indonesia, Vietnam, India, China and

---

Abbreviations: ANOVA, Analysis of Variance; APROMAC, Association des Professionnels du Caoutchouc Naturel de Côte d'Ivoire; BLUP, Best Linear Unbiased Prediction; CIRAD, Centre de Coopération Internationale en Recherche Agronomique pour le Développement; FAO, Food and Agriculture Organization; FS, Full-Sib; GBLUP, Genomic Best Linear Unbiased Prediction; GBS, Genotyping-by-Sequencing; GS, Genomic Selection; G x E, Genotype-by-Environment interactions; GWAS, Genome-Wide Association Studies; HS, Half-Sib; HSD, Honest Significant Difference; IRCA, Institut des Recherches sur le Caoutchouc en Afrique; IRSG, International Rubber Study Group; MAF, Minor Allele Frequency; MATAB, Mixed Alkyl Trimethylammonium Bromide; Mix, Mixed Relationship Approach; NR, Natural Rubber; QTL, Quantitative Trait Locus; RENL, Rubber Estates Nigeria Limited; SAPH, Société Africaine de Plantations d'Hévéa; SNP, Single Nucleotide Polymorphism; SNP-IP, Single Nucleotide Polymorphism Inclusion Probability; SOGB, Société des Caoutchoucs de Grand-Bereby; SSR, Simple Sequence Repeat.

[*] Corresponding author at: CIRAD, UMR AGAP Institut, Montpellier F-34398, France.
E-mail address: david.cros@cirad.fr (D. Cros).

https://doi.org/10.1016/j.indcrop.2025.120908
Received 1 October 2024; Received in revised form 28 January 2025; Accepted 19 March 2025

Malaysia as the main producers, and 76 % of the consumption (40 % by China). In West Africa, Côte d'Ivoire has emerged as the third largest NR producer in the world in 2023, with 11 % of global production, behind Thailand and Indonesia and ahead of Vietnam (IRSG, 2024).

Rubber cropping was initiated by the agro-industrial sector at the beginning of the 1900s. Currently, 85 % of the total planted area is carried out by smallholders (Fox and Castella, 2013; Wang et al., 2023). In Côte d'Ivoire, most of the stakeholders in the NR sector are organised in an association called APROMAC (Association des Professionnels du Caoutchouc Naturel de Côte d'Ivoire). This association represents the interests of planters, provides them with technical support and registers the planted areas and their owners in order to enable the traceability of the origin of rubber and to comply with national and international standards against deforestation (Bager et al., 2021; Kumeh and Ramcilovic-Suominen, 2023).

Natural rubber is issued from the latex of *Hevea brasiliensis* (hereafter called "rubber tree"), collected through periodic bark tapping, which involves cutting into the laticiferous system located in the outer part of the bark (Hao and Wu, 2000). After planting, the trees grow during the so-called "immature period" until the initiation of tapping at around 6 years of age. Then, the productive period of the plot before felling lasts around 25–30 years. Latex or coagulated rubber from the fields is dried in factories on the production sites, and then transferred in the form of dry blocks to manufacturing sites for the production of vulcanised rubber end-use items such as tires (Webster and Baulkwill, 1995).

Rubber genetic improvement was initially marked by the development of phenotypic selection and clonal trees grafted onto seedling rootstocks (Marattukalam and Saraswathyamma, 1992; Van Helten, 1918). Phenotypic selection emerged as the primary method for extracting new clonal varieties from diverse populations, with an increasing contribution of quantitative genetics and advancements in computing technology playing a crucial role in this process. Clonal selection was first applied to seedlings issued from natural pollination, then to full sib families issued from manual pollination (Simmonds, 1989). In Côte d'Ivoire, a breeding program launched in 1972 by the IRCA (Institut des Recherches sur le Caoutchouc en Afrique), using such methods, has successfully developed high-performing clones. These clones are now extensively cultivated in industrial plantations and, to a lesser extent, by smallholders (Elabo et al., 2019). Genetic selection has improved the yield over the past century, increasing from around 650 kg rubber ha^{-1} to 2500 kg rubber ha^{-1} between 1920s and 1990s.

Due to the long time period required for clonal evaluation, a distinction was done between, on the one hand, early selection on mother-tree seedlings and then on grafted trees in small scale trials with small plots and a short duration of evaluation, and, on the other hand, large scale trials with large plots and a long duration of evaluation. One main limitation is the low female fertility of the plant, which limits genetic recombination and the building of mating designs, and generates unbalanced full-sib families (Clément-Demange et al., 2007). Enhancing the genetic improvement of this species is still at stake to meet increasing global demand and to address challenges such as yield optimization, resistance to multiple diseases and adaptation to the environment.

Recent developments in genomic technologies have significantly advanced breeding methodologies across various crops, including perennial species. The integration of single nucleotide polymorphism (SNP) markers, genomic selection (GS) and quantitative trait loci (QTL) mapping has enabled more precise selection for complex traits, accelerating genetic gain (Cooper et al., 2009; Darkwa et al., 2020; Li et al., 2022; Vignal et al., 2002). Genomic selection involves the use of genome-wide markers to predict the breeding values of individuals, thereby reducing the cost of phenotypic evaluation, increasing selection intensity, and enhancing the selection process efficiency (Meuwissen et al., 2001). This innovative methodology integrates quantitative genetics principles, high-throughput genomic technologies and biostatics methods to analyse genome-wide single-nucleotide polymorphisms (SNP) (Goddard and Hayes, 2007; Grattapaglia et al., 2018). It uses prediction models constructed from a genetically related training population, comprising individuals with both genotypic and phenotypic data, to predict the genetic potential of selection candidates which are genotyped but not phenotyped (Robert-Granié et al., 2011). It has become a cornerstone in animal breeding programs (Hayes et al., 2009; Wiggans et al., 2017) and is increasingly adopted in plant breeding (Voss-Fels et al., 2019). Moreover, ongoing research continues to develop innovative approaches to enhance the utility of GS in diverse breeding contexts.

Genomic selection has been successfully applied in various crop species, in particular tropical perennial crops and plantation trees (Balocchi et al., 2023; Bouvet et al., 2020; Duarte et al., 2024; Seyum et al., 2022). Its potential in *Hevea brasiliensis* has been explored in recent studies. Cros et al. (2019) analysed GS efficiency for rubber production in the full-sib family PB260 x RRIM600 which was phenotyped and genotyped with 332 SSR markers. For the same family, Munyengwa et al. (2021) studied how to manage missing data for genomic predictions made with SNP obtained by genotyping-by-sequencing (GBS). Souza et al. (2019) studied how to manage G x E interactions with GS models, whereas Aono et al. (2022) investigated on the application of machine learning to genomic prediction in rubber tree. So far, these studies considered populations comprising only one or a few full-sib families.

Many parameters affect GS accuracy, i.e. the correlation between the predicted and true genetic value, like the size of the training population, the trait, the type and density of genotypic data, the relatedness between individuals, etc. In rubber tree, despite the advancements reported above, the influence of genetic relatedness between training and validation populations on the efficacy of GS remains an area that requires further exploration. Several studies have demonstrated that genomic selection accuracy is strongly influenced by the degree of genetic relatedness between the training and selection populations (Adeyemo et al., 2020; Asoro et al., 2011; Daetwyler et al., 2013; Grattapaglia, 2014; Isidro y Sánchez and Akdemir, 2021). It is therefore crucial to understand the influence of such relationships on a model population before implementing a breeding cycle based on GS. Additionally, number of studies have emphasized the potential of including quantitative trait locus (QTL) data into prediction models with the aim of enhancing the GS efficiency (Grattapaglia, 2022; Grattapaglia et al., 2018; Nsibi et al., 2020; Zhang et al., 2023). However, this approach has not yet been explored in *Hevea brasiliensis*.

This study aims to address two primary objectives in advancing GS for *Hevea brasiliensis*. The first objective is to evaluate the impact of genetic relatedness between training and validation populations on prediction accuracy for key traits, rubber production and sucrose content. The second objective is to assess the potential of incorporating QTL information into GS models to improve predictive performance.

For this purpose, we used datasets from two related full-sib families, derived from two F1 crosses PB260 × RRIM600 and PB260 × RRIC100, genotyped with 14,143 SNP markers and phenotyped for rubber yield and sucrose content at four trial sites in Côte d'Ivoire and Nigeria.

The influence of genetic relatedness between training and validation populations were examined on the accuracy of genomic predictions by comparing three different training population structures: (1) training and validation populations consisting of full-sibs, (2) training and validation populations consisting of half-sibs, and (3) a training population that was a mixture of full-sibs and half-sibs. To investigate the effect of integrating QTL information into the genomic prediction model, prior QTL detection was conducted in the training population and the information obtained from these analyses was incorporated into the GS model.

2. Materials and methods

2.1. Plant material

The plant material used in this study consists of two families of *Hevea brasiliensis* related by a common female parent, clone PB260. PB260, derived from the cross PB5/51 × PB49, was developed in 1958. Known for its vigour and high latex yield, PB260 is widely used as a female in crosses due to its superior female fertility (Cros et al., 2019). This clone exhibits a relatively rapid growth during the immature period prior to tapping, and is particularly characterised by a rapid increase in latex production once tapping begins, and low sucrose content in latex cells.

The male parents, RRIM600 and RRIC100, are prominent clones in major rubber producing regions and remain widely used as varieties in rubber cropping. RRIM600, derived from the TJIR1 × PB86 cross, was selected by the Rubber Research Institute of Malaysia in 1937. It combines moderate growth, latex yield and sucrose content in the latex cells, making it a stable performer under variety of conditions. This clone has historically been the most widely planted clone in the world due to its adaptability to different environments. RRIC100, selected by the Rubber Research Institute of Sri Lanka, is the result of a cross involving the clone RRIC52 as the female parent, with the male parent being unidentified (CIRAD, 2011). It is characterised by very fast growth during the immature period, allowing tapping to begin before the age of five. While its latex yield and sucrose content in latex cells are moderate, RRIC100 is highly valued for its exceptional tolerance to fungal leaf diseases. This disease tolerance significantly influenced its inclusion as a breeding parent, although it was not addressed in this study.

For each family, individuals were divided into two groups for phenotyping at two different experimental sites.

2.2. Methods

2.2.1. Study Sites

The study was conducted among four sites in Côte d'Ivoire and Nigeria. The PB260 × RRIM600 family was evaluated at two different sites in Côte d'Ivoire as describe by Cros et al. (2019). The first site as to referred as Site 1, is the SOGB plantation (Société des Caoutchoucs de Grand-Bereby), located in the southwest region of Côte d'Ivoire (latitude: 4°40′54″ N, longitude: 7°06′05″ W). At this site, 265 genotypes were planted in 2012. The second site, Site 2, is the Toupah plantation, part of the SAPH group (Société Africaine de Plantations d'Hévéa) and situated in the Grands-Ponts region of Côte d'Ivoire (latitude: 5° 19′ 47.79″ N, longitude: 4° 36′ 39.74″ W). At this location, 176 genotypes planted in July 2013.

The PB260 × RRIC100 family was planted in two additional sites, one in Nigeria and another in Côte d'Ivoire. The first site, Site 3, is the Osse River plantation of Rubber Estates Nigeria Limited (RENL), located in Edo State, Nigeria (latitude: 6°27'24.13"N, longitude: 5°24'50.92"E), where 250 genotypes were planted in 2012. The second site for this family, Site 4, was again at the SOGB plantation in Côte d'Ivoire, where 172 genotypes were planted in 2014.

2.2.2. Experimental design

At all four sites, the trees were planted using complete block designs, with six blocks and individual trees randomised within each block. The density was 1600 trees per hectare, with a spacing of 2.5 m between trees.

2.2.3. Phenotypic data

Phenotypic data were collected for three traits: rubber production and sucrose content, following the protocol outlined by Cros et al. (2019) and Munyengwa et al. (2021). The trees were tapped every two days using a half-spiral system, and latex was collected in 180 ml plastic boxes attached to the trunks. No ethephon stimulation was applied, ensuring natural latex flow. The raw rubber production data over a three-month period were analysed using a linear mixed model and the Best Linear Unbiased Prediction (BLUP) method via the lme4 R package (Bates et al., 2015). This process adjusted the rubber production for experimental design effects (blocks) and variations in tree girth at tapping initiation date. These adjusted values, referred to as phenotypes, represent the genetic performance of each clone. The phenotypic data of each site were standardized by centering and scaling using the following formula (Eq. 1):

$$\mathbf{z}_i = (\boldsymbol{x}_i - \overline{x}) \quad / \quad \boldsymbol{\sigma}_x \tag{1}$$

where $\mathbf{z_i}$ is the standardized value for the i^{th} observation, x_i the raw adjusted value for the i^{th} observation, $\overline{x}$ is the mean of all raw adjusted values and $\boldsymbol{\sigma}_\mathbf{x}$ their standard deviation. This transformation ensures that the phenotypic data have a mean of zero and a standard deviation of one for all sites.

2.2.4. Molecular data

Genomic DNA was extracted from young, healthy leaflets using mixed alkyl trimethylammonium bromide (MATAB) buffer and Macherey-Nagel NucleoMag magnetic beadst (Macherey-Nagel, Düren, Germany) on the Beckman Fxp robot (Beckman Coulter, CA, USA). Subsequent to the DNA extraction process, two enzymes (Pst1 and Mse1) were employed for DNA digestion, in conjunction with barcodes and adapters, and then multiplexed to accommodate 96 individuals per library. These libraries were processed using the CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement) Genotyping Technology Platform and then sent to Azenta Life Sciences (Germany) for next-generation sequencing. The Genotyping-by-Sequencing (GBS) methodology included several key steps: normalization of genomic DNA, digestion with *Pst*I and *Mse*I, ligation with barcoded adapter sequences, pooling of PCR products (combining 96 genotypes into one GBS library), DNA purification, PCR amplification with adapter-specific primers, dual DNA purification, and sequencing on the Novasq 6000 platform, producing paired-end reads of 150 nucleotides for each fragment (Mournet et al., 2020).

Single nucleotide polymorphisms (SNPs) were identified from the sequence reads using the reference genome of PB 260 (Ismawanto et al., 2024; Lopez, 2023) and VcfHunter, a pipeline available at https://github.com/SouthGreenPlatform/VcfHunter/ (Garsmeur et al., 2018).

2.2.5. Marker filtering and quality checking

Molecular markers, encoded using allelic dosage values of 0, 1, and 2 depending on the number of alternative alleles copies present, were contained in VCF (variant call format) files. Before proceeding with the analysis, a quality check of the genotyping data was conducted to identify and remove outliers and other inconsistencies. The quality assessment revealed that, on average, 13 % of data was missing per individual in the initial dataset. From the initial set of 14,143 bi-allelic SNPs, 12,960 markers were selected for further analysis based on a minor allele frequency (MAF) threshold of 0.05 and a maximum of 30 % missing data. All SNPs were positioned on 27 pseudo-chromosomes. This filtering process led to the elimination of 81 individuals. The missing data percentage per marker ranged from 0 % to 28.9 %, with an average of 10.06 %. As a result, the data preparation process yielded a dataset consisting of 674 genotyped individuals and 12,960 SNP markers. Missing data were subsequently imputed using Beagle 5.4 (Browning et al., 2018).

2.2.6. Validation approaches for genomic predictions

Genomic predictions were conducted within and between sites, resulting in 28 distinct prediction scenarios for a given trait (e.g. Site 1 toward Site 2, Site 2 + Site 3 toward Site 4, etc.). Based on the use of a single site or of multiple sites and on the degree of relatedness between the training and validation populations, the different scenarios were grouped into five principal prediction approaches.

2.2.6.1. *Cross-validation within each site.*

The performance of prediction within each site, and thus within full-sibs, was assessed using a cross-validation strategy for the different traits. The population was randomly partitioned into two subsets: 80 % of the set was used as the training population to predict the genetic values of the remaining partition (20 % of the set), whose phenotypes were assumed to be unknown. This procedure was repeated 30 times, with samples drawn with replacement from the set.

2.2.6.2. *Prediction using full-sib data across sites.*

Secondly, the efficacy of the prediction was evaluated using phenotypic data from a full-sib to predict the genetic values of their full-sib planted on disparate sites. The dataset from each of the four sites was successively employed as a training set to predict the GEBVs of their full-sib planted on the other sites (e.g., site 1 toward site 2 or site 3 toward site 4). The approach, designated as "FS", allowed the testing of four distinct scenarios.

2.2.6.3. *Prediction using half-sib data across sites.*

Subsequently, the efficiency of prediction was tested using phenotypic data from a family of full-sibs to predict the genetic values of their half-sibs planted on separate sites (e.g., site 1 toward site 3 or site 4 toward site 2). Designated as "HS", this validation approach allowed the testing of eight different prediction scenarios.

Additionally, the HS scenario was extended into a scenario named "HS2" which allowed for studying the impact of increasing the size of the training population. This was achieved by combining data from two sites planted with clones of the same family for the training of the model and using it to predict the genetic values of half-sibs at other sites. This resulted in four scenarios (e.g., site 1 + site 2 towards site 3 or site 3 + site 4 towards site 2).

2.2.6.4. *Prediction using data from a training set comprising a mixture of full-sib and half-sib relatives in relation to the validation population.*

Finally, the prediction model was trained using a dataset that included both full-sib and half-sib relatives from two different sites. This was done with the aim of evaluating the prediction performance when the training set is composed of mixed familial relationships, which reflect the complexity of genetic relationships in the hevea breeding programmes. The training dataset was constructed by combining data from two families at two different sites with the objective to predict the genetic values of other sites. This allowed for the testing of eight different scenarios, for example, site 1 + site 3 towards site 2 or site 2 + site 4 towards site 3. We referred to this prediction approach as "Mix".

2.2.7. *Statistical methods for genomic predictions*

The Genomic Best Linear Unbiased Prediction (GBLUP) model was utilized for genomic prediction (VanRaden, 2007), evaluating various scenarios based on the relatedness between the training and validation populations. This model links phenotypic variability to the variation explained by all markers, assumed to be in linkage disequilibrium with loci responsible for the traits of interest. The equation of the model is given by Eq. 2:

$$\boldsymbol{y} \quad = \quad \mathbf{1}\mu + \boldsymbol{Xg} + \quad \boldsymbol{e} \tag{2}$$

where y is the vector of adjusted phenotypes of the training set, μ the overall mean, 1 a column vector of 1 s, $g \sim N(0, G\sigma_g^2)$ is the vector of additive genetic effects of the training and validation sets. **X** is the incidence matrix associated with vector **g**. $\mathrm{G}\sigma_{\mathrm{g}}^2$ is the variance-covariance matrix of **g**, σ_{g}^2 the additive variance and **G** the genomic additive relationship matrix. **G** was obtained from the training and validation SNP data as $\mathbf{G} = ZZ'/n_{snp}$, with the matrix with **Z** the matrix of SNP genotypes with SNPs in columns and standardized so that columns have mean zero and standard deviation one, and **Z'** is the transpose of **Z** (Pérez and de los Campos, 2014). **G** was adjusted to be non-singular as $\boldsymbol{G} = (1 - \alpha)\boldsymbol{G} + \boldsymbol{I}\alpha$, with $\alpha = 0.001$ (Legarra, 2017). $\boldsymbol{e} \sim N(0, I\sigma_e^2)$ is the vector of residual effects. Variance parameters were estimated by restricted maximum likelihood (REML) and solutions of the mixed models, i.e. the vector $\hat{\mathbf{g}}$, were obtained by resolving Henderson's mixed model equations (Henderson, 1975), using lme4GS package (Caamal-Pat et al., 2021).

The prediction accuracy was evaluated using Pearson's correlation coefficient between the observed and predicted phenotypes.

To assess the effect of site for cross-validations and validation scenarios for between site predictions, analyses of variance (ANOVA) were performed separately for each trait on the accuracy. A Tukey's HSD (honestly significant difference) test was used for pairwise comparison of prediction accuracies between sites and validation approaches.

2.2.8. *Genomic prediction optimization using QTL*

To enhance the accuracy of genomic selection models, we integrated genetic architecture (i.e. QTL results) into these models. This methodology involves partitioning molecular markers into two groups based on their association with phenotypic variation. SNPs with the highest average probability of inclusion are used as covariates. This approach aims to prioritize markers strongly linked to specific traits and to inform the genomic prediction models, which is expected to enhance their performance. This approach was evaluated on the rubber production trait.

Two strategies were explored to implement this approach effectively. The first strategy, similar to cross-validation approach, involved using a random sample of 80 % of the population from a given site to detect markers associated with QTLs. The information regarding the QTL-derived markers was then used in the GS model to predict the genetic values of the remaining 20 % of the population. This procedure was repeated 30 times for each site to ensure robustness. The second strategy utilized the entire population of one site to detect QTL-associated markers, with the goal of using this QTL information to predict the genetic values of individuals from another site.

The QTL analyses were made for rubber production using the BayesC model from the BGLR package (Pérez and de los Campos, 2014). BayesC is a Bayesian variable selection method that can be used for QTL detection (Amadeu et al., 2021; Pérez-Rodríguez and De Los Campos, 2022). The selection of relevant variables (i.e. SNP) is based on their posterior inclusion probabilities, provided by the function. The inclusion probability indicates the importance of a SNP in trait variability. We conducted 60,000 iterations with a proportion of markers having non-zero effects set at 0.001. For each site, five analyses were performed. An average inclusion probability was calculated over the five analyses to select markers linked to traits. The proportion of phenotypic variation explained by each SNP was measured using the R^2 metric, which was derived from a linear model with the phenotype as the dependent variable and the SNP as a fixed effect, using the lme4GS package.

The GBLUP model which incorporated prior information about the QTL associated with the trait is illustrated by Eq. 3:

$$\boldsymbol{y} \quad = \quad \mathbf{1}\mu + \quad \boldsymbol{X_1 g_1} \quad + \quad \boldsymbol{X_2 g_2} \quad + \quad \boldsymbol{e} \tag{3}$$

Here, g_1 denotes the part of the additive genetic values associated with the selected markers, while g_2 denotes the part associated with the other markers. $\mathrm{g}_1 \sim \mathrm{N}\ (0, \mathrm{G}_1\sigma_{\mathrm{g}1}^2)$ and $\mathrm{g}_2 \sim \mathrm{N}\ (0, \mathrm{G}_2\sigma_{\mathrm{g}2}^2)$, with G1 and G2 being the genomic additive relationship matrix computed with the markers within genomic region of interest or with the markers outside this region, respectively.

ANOVA followed by a Tukey's HSD test was employed for pairwise comparison of prediction accuracies between models.

3. Results

3.1. *Exploration of the phenotypic data*

Table 1 presents the descriptive statistics for the phenotypic traits,

Table 1
Descriptive statistics of phenotypic traits measured across sites.

Site	Trait	Mean	SD	Min	Max	Heritability (%)
Site 1	Latex production (cg)	7578,53	2807,59	1658,22	16618,77	90
	Sucrose content (mM/L)	26,08	3,44	17,6	35,04	56
Site 2	Latex production (cg)	24161,9	6774,81	8290,39	48573	93
	Sucrose content (mM/L)	17,6	1,56	13,66	23,38	69
Site 3	Tree girth (mm)	26,81	2,14	20,56	33,33	65
	Latex production (cg)	11062,99	3608,17	1471,61	22897,21	85
	Sucrose content (mM/L)	14,94	2,92	7,76	29,86	57
Site 4	Tree girth (mm)	31,91	2,08	26,07	36,47	85
	Latex production (cg)	18774,93	4319,56	10739,25	31584,19	94
	Sucrose content (mM/L)	17,23	3,49	9,13	29,97	67

showing notable variation in latex production, sucrose content and tree girth across the four sites. Latex production ranged from 7578.53 cg at site 1–24161.9 cg at site 2. Heritability estimates for latex production ranged from 85 % at Site 3–94 % at Site 4. For sucrose content, Site 1 had the highest mean value (26.08 mM/L), whereas Sites 3 had the lowest (14.94 mM/L). The heritability ranged from 56 % (Site 3) to 69 % (Site 2).

Tree girth was only measured at Sites 3 and 4, with site 4 having a higher girth (31.91 mm) compared to Site 3 (26.81 mm). The heritability for this trait varied from 85 % at Site 4–65 % at Site 3.

3.2. Genomic predictions using cross-validation approach

The predictions for the three traits, namely rubber production, sucrose content and tree girth, among four different sites using cross-validation demonstrated variability in prediction accuracy among sites and traits (Fig. 1).

The prediction accuracy for rubber production varied significantly among the four sites (Table 2). Sites 1 and 3 demonstrated the highest prediction accuracies, with values of 0.63 and 0.57, respectively. In contrast, Sites 4 and 2 exhibited lower prediction accuracies of 0.37 and 0.33. These two sites differed significantly in terms of prediction accuracy, compared to Sites 1 and 3.

For sucrose content, the prediction accuracy varied notably among the four sites. Site 2 exhibited the highest prediction accuracy at 0.51.

Table 2
Comparison of prediction accuracy among sites for each trait using Tukey's HSD test for significant differences. Values followed by the same letter (s) among sites are not significantly different at $P \leq 0.05$.

	Rubber production	Sucrose content	Tree girth
Site 1	0,63 a	0,42 ab	-
Site 2	0,33 b	0,51 a	-
Site 3	0,57 a	0,30c	0,45 a
Site 4	0,37 b	0,41 b	0,34 b

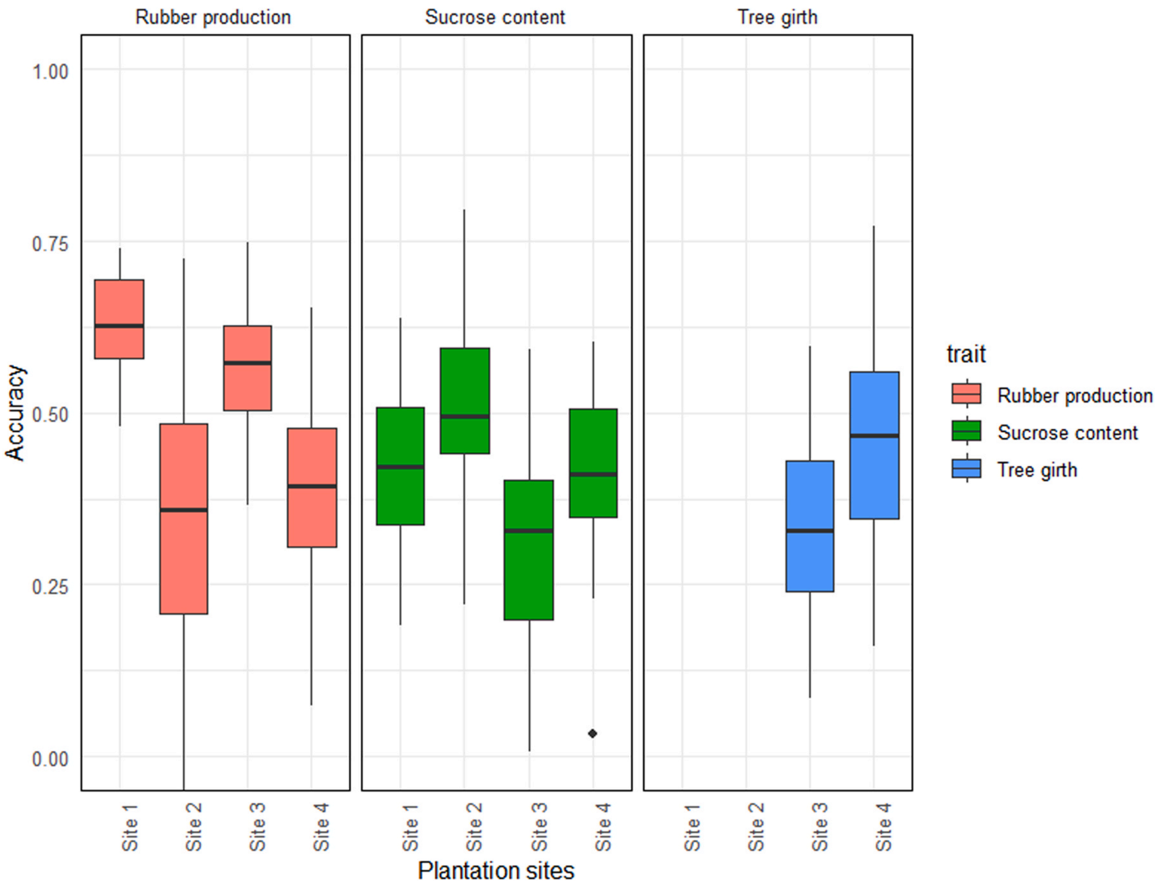

Fig. 1. GS accuracy for rubber production, sucrose content and tree girth among sites obtained by cross-validation.

With an accuracy of 0.42. Site 1 was not significantly different from either Site 2 or Site 4, which had an accuracy of 0.41. Site 3, with the lowest prediction accuracy of 0.30, was significantly different from the other sites.

Tree girth was only evaluated at Sites 3 and 4, with Site 4 showing better prediction performance (mean = 0.453) compared to Site 3 (mean = 0.341).

3.3. Impact of relatedness between training and validation populations on genomic prediction

The genomic prediction accuracy for rubber production and sucrose content varied depending on the validation scenario and the trait (Fig. 2).

For rubber production, the FS approach achieved the highest accuracy of 0.54, closely followed by the Mix approach, at 0.52. The variance analysis indicates that using full-sib data (FS) or combining full-sib with half-sib data (Mix) for model training provide similar prediction performance. In contrast, the HS2 and HS approaches resulted in significantly lower accuracies of 0.20 and 0.17, respectively (Table 3).

For sucrose content, similar trends were observed. Thus, the Mix and FS approaches gave similar prediction accuracies (0.37 and 0.36, respectively), while HS and HS2 led to lower values (although the difference between HS2 and FS/mix was not significant). As with the previous trait, increasing the size of the training population with the HS2 approach increased the prediction accuracy compared to the HS scenario.

3.4. Optimization of genomic prediction with QTL information

3.4.1. QTL detection

The QTL analysis for rubber production across four different sites identified several significant markers, with varying genomic location and phenotypic effect according to sites. The top five markers identified from the QTL detection analyses using the full dataset from each site are presented in Table 4. A small number of common genomic regions were identified among sites of the same family, i.e. a 4.2 Mbp region on pseudo-chromosome ptg000024 for family PB260 x RRIM600 (Sites 1 and 2) and two regions of 2.6 Mbp on ptg000006 and 5.7 Mbp on ptg000017 for PB260 x RIC100 (Sites 3 and 4). However, no common significant SNPs were detected. At Site 1, the top three markers (ptg000024l_22674579, ptg000024l_20455161, and ptg000024l_23205187) were all located on a region covering less than 3Mbp on pseudo-chromosome ptg000024l, explaining 17 %, 15 %, and 14 % of the phenotypic variation, respectively. The subsequent markers were positioned on different pseudo-chromosomes, contributing to lower proportions of the variation (7 % and 5 %, respectively). For Site 2, the most important marker was ptg000024l_18962668, explaining 16 % of the phenotypic variation. Moreover, two markers were found on pseudo-chromosome ptg000005l, contributing 14 % and 10 %. Additionally, two markers were identified on pseudo-chromosome ptg000029l, explaining 11 % and 9 % of the variation, respectively. At Site 3, the marker ptg000006l_87274838 was the most significant, accounting for 15 % of the phenotypic variation. Notably, two markers positioned on pseudo-chromosome ptg000008l explained 11 % and 10 % of the variation, respectively. Furthermore, two other markers on pseudo-chromosome ptg000017l contributed 8 % and 7 % of the variation. Finally, at Site 4, the key marker ptg000014l_44438953 explained 14 % of the phenotypic variation, followed by two markers located on pseudo-chromosome ptg000006l (each accounted for 11 % of the variation).

By using random samples comprising 80 % of the population to detect markers associated with QTLs, the ranking of markers by importance showed variability across different analyses of the same site. In the 30 analyses conducted, between 10 and 20 distinct SNPs were identified as having the highest probability of inclusion for a given site. A single SNP appeared as the top-ranked marker between 3 and 14 times out of the 30 analyses for site 1 (Table 5).

3.4.2. Integration of QTL information into GS models

The Fig. 4 presents the GS accuracy with the cross-validation strategy in the four sites using the GBLUP model and the GBLUP model augmented with QTL information. The prediction accuracy decreased in

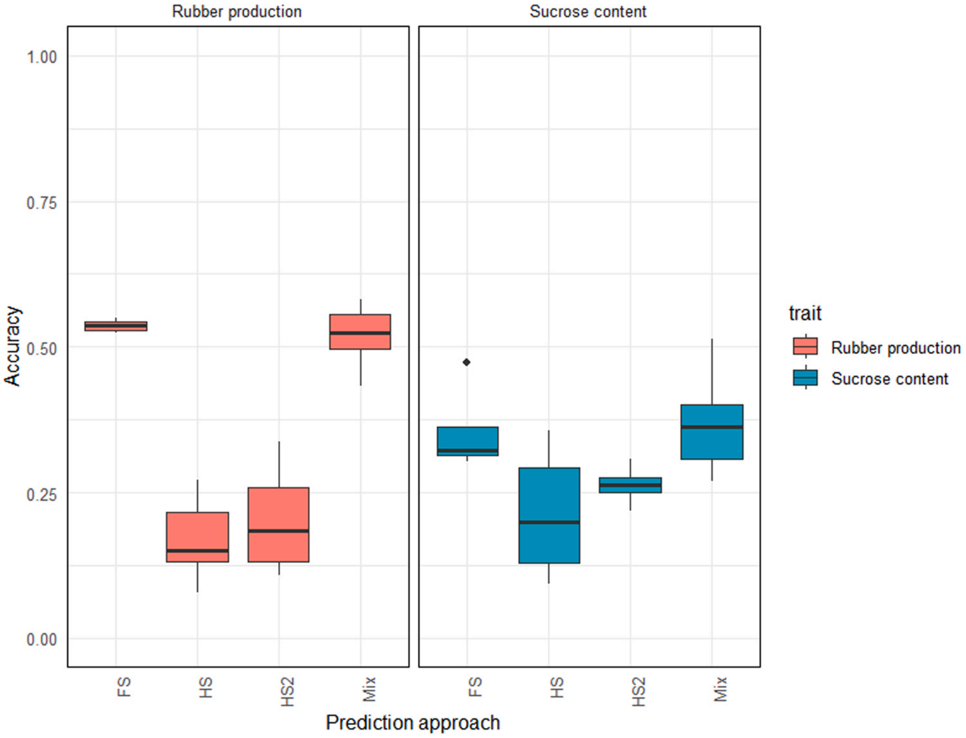

Fig. 2. GS accuracy for rubber production and sucrose content using different validation approaches across sites.

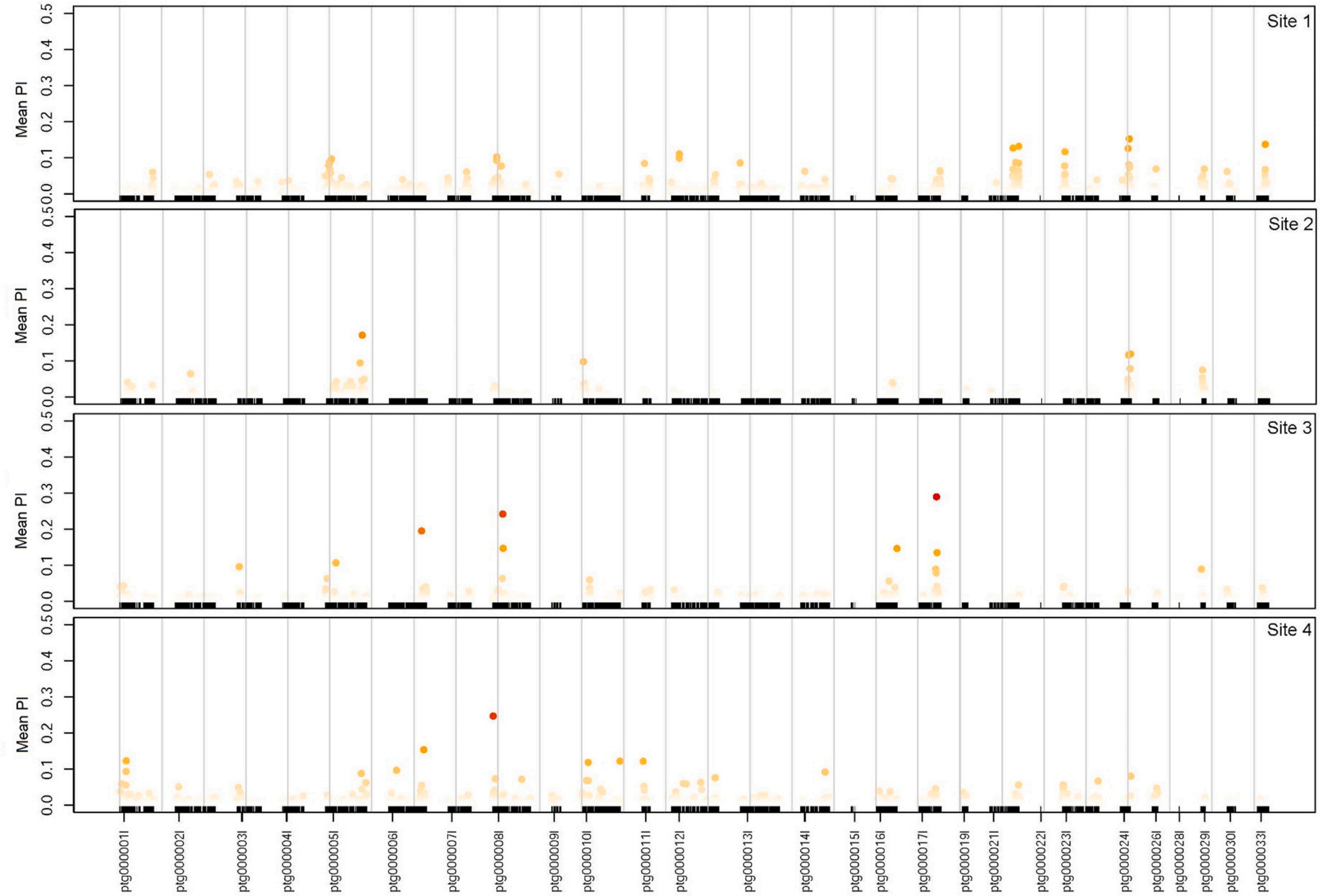

Fig. 3. Genomic regions associated with rubber production for each site. Each dot represents a SNP. PI: probability of inclusion.

Table 3

Comparison of GS accuracy by validation approaches for each trait using Tukey's HSD test for significant differences. Values with the same letter among validation approaches are not significantly different at P = 0.05.

	Rubber production	Sucrose content
FS	0,54 a	0,36 ab
HS	0,17 b	0,21 b
HS2	0,20 b	0,26 ab
Mix	0,52 a	0,37 a

all sites when using the model integrating QTL information. On Site 1, the decrease was small, with the GS accuracy of the GBLUP + QTL model being 0.62 against 0.63 for the GBLUP model, but stronger decreases were found in the other sites. Thus, in Site 2, the accuracy of the GBLUP model was 0.41 against 0.33 for the GBLUP + QTL model at 0.33. On Site 3, the inclusion of QTL information in the GBLUP model significantly reduced the accuracy from 0.52 to 0.42. Similarly, in Site 4, the GBLUP model had an accuracy of 0.40, while the GBLUP + QTL model decreased to 0.34 (Table 6).

The second strategy involved using phenotypic data from a single site, supplemented with trait-associated marker information, to predict the genetic values of individuals at a different site through the implementation of a FS and HS validation approach.

For the FS scenario, the GS accuracy for the GBLUP model was 0.53, while the GBLUP + QTL model showed a slightly lower accuracy of 0.52; however, the difference was not statistically significant (Table 7). In the HS scenario, the GBLUP model and the GBLUP + QTL model also displayed similar accuracies, with values of 0.16 and 0.13, respectively, with no significant differences observed between the models (Fig. 5).

Table 4

Top 5 markers (with the highest probability of inclusion) identified from QTL detection analysis for rubber production using entire dataset of each site.

Site	Marker	Pseudo-chromosome	Position (pb)	phenotypic variation (R^2, in %)
Site 1	ptg000024l_22674579	ptg000024l	22674579	17
	ptg000024l_20455161	ptg000024l	20455161	15
	ptg000024l_23205187	ptg000024l	23205187	14
	ptg000008l_11355608	ptg000008l	11355608	7
	ptg000033l_21322938	ptg000033l	21322938	5
Site 2	ptg000024l_18962668	ptg000024l	18962668	16
	ptg000005l_80904429	ptg000005l	80904429	14
	ptg000029l_2144265	ptg000029l	2144265	11
	ptg000005l_86014406	ptg000005l	86014406	10
	ptg000029l_2424338	ptg000029l	2424338	9
Site 3	ptg000006l_87274838	ptg000006l	87274838	15
	ptg000008l_24778126	ptg000008l	24778126	11
	ptg000008l_25862107	ptg000008l	25862107	10
	ptg000017l_43295569	ptg000017l	43295569	8
	ptg000017l_44457735	ptg000017l	44457735	7
Site 4	ptg000014l_44438953	ptg000014l	44438953	14
	ptg000006l_84657378	ptg000006l	84657378	11
	ptg000008l_1908656	ptg000008l	1908656	11
	ptg000017l_38788405	ptg000017l	38788405	9
	ptg000005l_90912486	ptg000005l	90912486	7

4. Discussion

In genomic selection programs, the choice of the training population is crucial for the accuracy and effectiveness of breeding efforts. In this research, we investigated the impact of genetic relatedness between

Table 5
Top markers identified based on inclusion probability from QTL analysis using 80 % of the population per site. Frequency indicates the number of times a marker had the highest inclusion probability over the 30 replicates of analysis.

Site 1	Frequency	Site 2	Frequency	Site 3	Frequency	Site 4	Frequency
ptg000024l_22674579	14	ptg000024l_18962668	12	ptg000006l_87274838	3	ptg000014l_44438953	12
ptg000012l_103928856	7	ptg000005l_15237515	3	ptg000006l_89299314	3	ptg000008l_1908656	3
ptg000024l_20455161	2	ptg000005l_86014406	2	ptg000008l_23887251	3	ptg000006l_84657378	2
ptg000001l_78905997	1	ptg000005l_86103026	2	ptg000017l_41216414	3	ptg000017l_38788405	2
ptg000002l_90885569	1	ptg000024l_18966905	2	ptg000008l_24778126	2	ptg000002l_9056815	1
ptg000005l_15508981	1	ptg000001l_24891765	1	ptg000030l_5386154	2	ptg000005l_90912486	1
ptg000008l_14352859	1	ptg000001l_37184084	1	ptg000001l_18827744	1	ptg000005l_96828545	1
ptg000017l_49813050	1	ptg000005l_14411091	1	ptg000001l_81389555	1	ptg000006l_79881853	1
ptg000021l_63535779	1	ptg000005l_16764351	1	ptg000006l_75411744	1	ptg000012l_104393027	1
ptg000024l_24628149	1	ptg000005l_84888882	1	ptg000006l_79443043	1	ptg000012l_99873043	1
		ptg000008l_81049708	1	ptg000008l_25862107	1	ptg000013l_55295056	1
		ptg000012l_83460118	1	ptg000008l_9278955	1	ptg000013l_58906799	1
		ptg000014l_31089981	1	ptg000010l_19353106	1	ptg000023l_2355330	1
		ptg000023l_12532345	1	ptg000016l_46345493	1	ptg000023l_38596384	1
				ptg000017l_42053065	1	ptg000023l_85788303	1
				ptg000017l_42490831	1		
				ptg000017l_42614127	1		
				ptg000021l_61579198	1		
				ptg000029l_904508	1		
				ptg000033l_10403229	1		

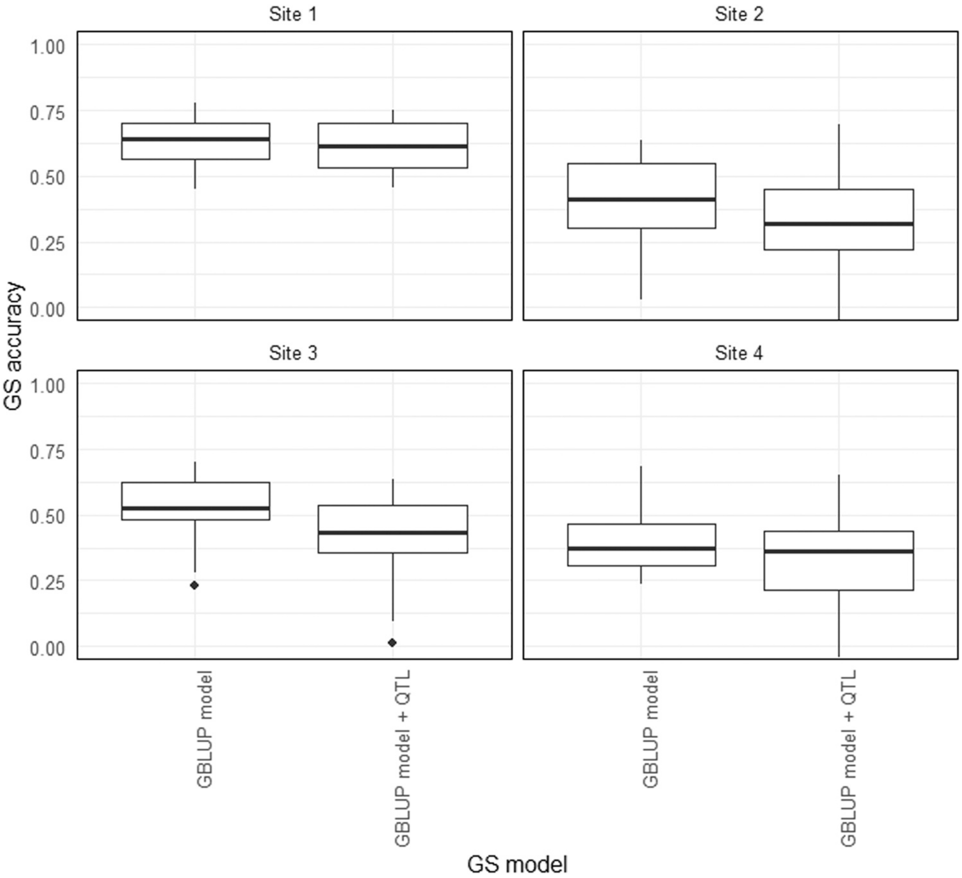

Fig. 4. GS accuracy for rubber production according to prediction model across sites.

Table 6
Comparison of GS accuracy according to prediction model and site using Tukey's HSD test for significant differences. Values followed by the same letter (s) among prediction models are not significantly different at $P \leq 0.05$.

	Site 1	Site 2	Site 3	Site 4
GBLUP model	0,63 a	0,41 a	0,52 a	0,40 a
GBLUP model + QTL	0,62 a	0,33 a	0,42 b	0,34 b

training and validation populations, as well as the integration of prior QTL information, on the efficiency of genomic selection in *Hevea brasiliensis*, using two biparental families planted across four distinct sites. By estimating GS prediction accuracy for rubber production, tree girth, and sucrose content across different sites and families, the present study provided a broader perspective on the performance of GS for this species. Among the traits, rubber production exhibited the highest GS accuracy, suggesting it is influenced by a larger number of genes compared

Table 7
Comparison of GS accuracy according to prediction model and validation approach using Tukey's HSD test for significant differences. Values followed by the same letter (s) among prediction models are not significantly different at $P \leq 0.05$.

	FS	HS
GBLUP model	0,53 a	0,16 a
GBLUP model + QTL	0,52 a	0,13 a

to the other traits. Furthermore, our results revealed significant variations in GS accuracy depending on the site and family, illustrating the need for caution when considering the practical application of the method.

4.1. Impact of relatedness on genomic predictions

The analysis of relatedness between training and validation populations demonstrated that closer genetic relationships, as in the FS approach, resulted in higher prediction accuracies for rubber production. Conversely a substantial drop in accuracy was observed with the HS and HS2 approaches, underscoring the critical role of genetic relatedness in building robust prediction models. In particular, when a genomic prediction model is trained on one family to perform predictions in another family, the accuracy tends to be lower than what can be achieved within each family, despite the fact that the two families are connected by a common parent. These findings are consistent with a number of studies in other plant species (Brauner et al., 2020; Duhnen et al., 2017; Guo et al., 2014; Habier et al., 2007; Lorenz et al., 2012; Rio et al., 2019, 2021; Windhausen et al., 2012) which have shown that genomic selection models perform better when training and validation sets are closely related. This deterioration in the prediction accuracy can be attributed to differences between families in terms of alleles contributing to trait variation, allele frequencies at QTLs (Schopp et al., 2017) and QTL allele effects and differences in linkage disequilibrium between markers and QTLs (de Roos et al., 2009), leading to biases in the effects estimated at markers. While using the same full-sib family for training and validation allowed for the highest prediction accuracies, this approach requires creating as many training sets as there are candidate families, which is resource-intensive and unrealistic for large-scale breeding programs.

This study showed that combining full-sibs and half-sibs in the training population, as in the Mix approach, can achieve prediction performances comparable to intra-family models. A practical approach could involve constructing a single training population composed of multiple families connected by common parents in genomic selection. This strategy would have two practical advantages compared to the use of multiple training populations, each consisting of a single full-sib family. First, it would eliminate the need for artificial pollinations, which require significant labour time, and replace them with open pollinations among a small number of selected individuals. Second, it would allow for the management of a single training population within the breeding program, which could be used generically for all families related to the initially selected individuals, particularly their progenies. Third, the training set would be made of full-sibs and a panel of several half-sib families (not of a single one as in our study), and this can increase the prediction accuracy, as found by Juliana et al. (2022). Such an approach relying on multiple connected families has also been recommended by Brauner et al. (2020). The higher genericity of such a training population would reduce the number of individuals to be phenotyped among several full-sib families, and predictions could be applied to a larger selection population, thus allowing for a higher selection intensity and genetic diversity.

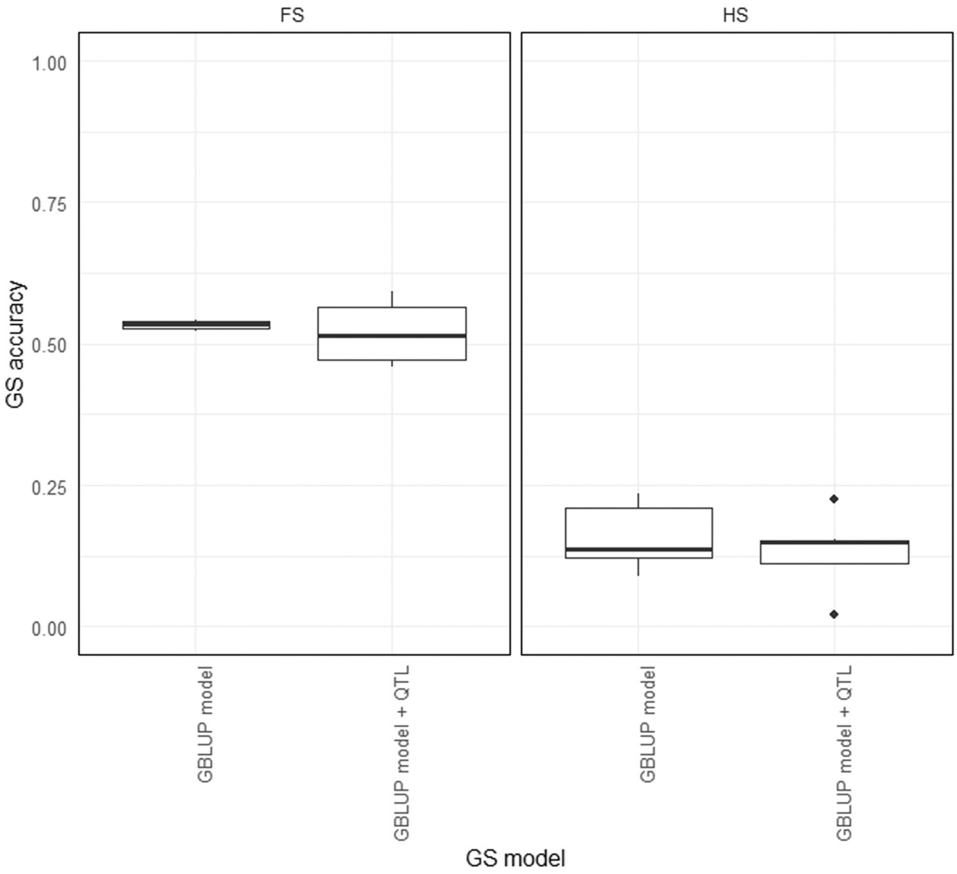

Fig. 5. GS accuracy for rubber production according to prediction model and validation approach across sites.

Before implementing this strategy, it is important to quantify the effect, in terms of accuracy of genomic prediction, of using a training population that includes a mix of full-sibs and half-sibs from the selection family, compared to a training population composed solely of full-sibs. Here, we have shown that adding half-sibs to a training population that includes a full-sib family does not reduce accuracy compared to calibrating with only full-sibs. This demonstrates that the proposed strategy is feasible. Additionally, the results also showed that increasing the size of the training population by adding half-sibs does not increase the accuracy of genomic prediction. This suggests two possibilities. The first is that it might be possible to reduce the size of the different full-sib families, which would increase the efficiency of genomic selection per unit cost and/or allow the reallocation of saved resources to evaluate other families. The second would be to increase the density of molecular markers to enhance segregation within the population and thus make better use of the available population size.

4.2. QTL detection

The objective was to identify QTL with potentially strong effects and integrate them into the GS model, rather than focusing on the segregation of favourable alleles within the family, as in the case of classical QTL mapping approaches. Few methods of QTL detection are well-suited for full-sib families between heterozygous parents. The BayesC model, a bayesian variable selection method of the whole genome association analysis type, was chosen for its simplicity and flexibility among several methods, including single interval mapping with Rqtl (Broman et al., 2003, 2019) and composite interval mapping with FullsibQTL (Gazaffi et al., 2014). BayesC partitions molecular markers based on their posterior inclusion probabilities, enabling the prioritisation of markers strongly associated with trait variation (Wolc and Dekkers, 2022). Unlike traditional QTL mapping approaches, this method does not require a genetic linkage map. By conducting multiple iterations and selecting SNPs with the highest average inclusion probabilities across analyses, we ensured robust identification of genomic regions linked to rubber production. Improving QTL detection could be a key area for future research. Higher-resolution mapping techniques and denser marker arrays may provide better resolution in future studies (Heidari et al., 2011; Wu et al., 2022). Additionally, multi-environment QTL analyses or modelling genotype-by-environment interactions could allow for the identification of more stable QTL, potentially enhancing their utility in genomic predictions (Ahmadi and Bartholomé, 2022).

4.3. Integration of QTL information in genomic prediction models

The integration of QTL information into GS models, while theoretically promising (Grattapaglia, 2022; Grattapaglia et al., 2018; Nsibi et al., 2020; Zhang et al., 2023), resulted in a consistent decrease in prediction accuracy across all cases (cross-validation or other validations approaches) in our study. This outcome may be attributed to the complex genetic architecture of the trait, with multiple small-effect QTLs (Morgante et al., 2018). This is supported by the fact that the highest R^2 for each site accounts for only 14–17 % of the total phenotypic variance, and that several regions were identified depending on site, family and, for cross-validations, sample of training individuals. For instance, in Nsibi et al. (2020), where positive results were obtained when adding QTL information in the GS model, the R^2 values for the QTLs reached up to 51.2 %, and 40 % of their QTLs had R^2 values ≥ 17 %. The inconsistency between the results of the different sites, even for a same family, further reduces the effectiveness of predictions, as it indicates instability in QTL detection, which is likely caused by GxE interaction. The instability of QTLs can further complicate predictions, as noted by Nsibi (2021), especially for polygenic trait, reducing the expected performance from incorporating QTL data. McElroy et al. (2018), working on cocoa, also considered the use of a priori information on the SNPs, by making genomic predictions including as fixed effects the SNPs detected by GWAS. This did not improve prediction accuracies in their study, but this likely occurred as a consequence of the genetic differentiation between the training and application populations, making irrelevant the detected SNPs, and as a consequence of the small number of SNPs identified by GWAS. When considering predictions between half-sib families, the absence of common QTLs between the two families further explains the failure of using QTL information for genomic selection prediction.

The goal of the cross-validation study was to consider a situation without GxE interactions, assuming that this could make effective the genomic prediction approach that incorporated QTL information into the model. However, this was not the result observed. We assume that the explanation lies not only in the low variance explained by the SNPs at the QTLs, but also in populations that were too small for QTL detection using a cross-validation approach, where 20 % of individuals are set aside for validation. This is suggested by the fact that different QTLs were identified when using different samples in a given site.

The negative effect of incorporating QTL information into the GS models in our study may be attributed to the fact that only a single SNP was included in the model, which may not have been sufficient to capture a large enough portion of the genetic variance. Preliminary analyses were conducted using up to 5 SNPs with the highest probability of inclusion, but this did not lead to any improvement in the results (data not shown). The consistent reduction in prediction accuracy when incorporating QTL information suggests that this approach, as implemented in this study, does not yet provide a reliable advantage for improving genomic predictions in *Hevea brasiliensis*. This has consequences for breeding programs, as it demonstrates the need to carefully consider the added value of QTLs in predictive models, especially for polygenic traits.

Finally, integrating functional annotations (e.g. candidate genes) could enhance the effectiveness of QTL-informed genomic selection, representing a valuable addition to our current approach. Prioritising specific genomic regions or gene networks based on prior biological knowledge could improve the accuracy of predictions, particularly for complex traits. Furthermore, integrating omics approaches (transcriptomics, metabolomics, and proteomics) could enhance genomic prediction by capturing intricate biological interactions and non-linear relationships, ultimately leading to more precise and reliable trait predictions (Ahmadi and Bartholomé, 2022; Alemu et al., 2024).

4.4. Phenotypic data

The study was conducted across four diverse sites: three in Côte d'Ivoire (two at the SOGB plantation and one at the Toupah plantation) and one in Nigeria (Osse River plantation). These sites differ in environmental conditions, including soil type, climate, and management practices, which significantly influenced phenotypic variability for traits such as latex production, sucrose content, and tree girth. The observed variation in genomic prediction accuracy across suggests a significant amount of genotype-by-environment interactions. For instance, the higher prediction accuracies for latex production at certain sites could be linked to more stable phenotypic expression under specific environmental conditions. Conversely, the lower accuracies observed at other sites underscore the challenges posed by greater environmental heterogeneity.

5. Conclusion

The present study provided a broader perspective on the performance of GS for rubber tree by considering various traits, sites and families. This study provides a comprehensive evaluation of genomic prediction models in rubber tree breeding, highlighting the importance of environmental factors, genetic relatedness, and careful integration of QTL information. Our results suggest that the construction of a training population composed of several families connected by common parents is an interesting strategy for rubber breeding. Future research should focus on refining genomic selection models by incorporating environmental covariates and modelling genotype by environment interactions to enhance prediction accuracy across diverse conditions. Additionally, the results suggest that leveraging closely related training populations and validating QTL effects across multiple environments could improve the reliability of genomic predictions in rubber tree breeding programs.

CRediT authorship contribution statement

Mournet Pierre: Writing – review & editing, Methodology. **Lopez David:** Formal analysis. **Hofs Jean-Luc:** Writing – review & editing, Conceptualization. **Bonal Florelle:** Methodology. **Soumahoro Mouman:** Conceptualization. **Sélastique Akaffou Doffou:** Writing – review & editing, Project administration. **CROS David:** Writing – review & editing, Supervision, Project administration, Formal analysis, Conceptualization. **Daval Aurélie:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Kouassi Daouda Koffi:** Writing – original draft, Formal analysis, Data curation. **Clément-Demange André:** Writing – review & editing, Conceptualization. **Le Guen Vincent:** Writing – review & editing, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the Institut Français du Caoutchouc (IFC), and the companies Michelin, SIPH, and SOCFIN for funding the part of this research devoted to Hevea rubber in the framework of the IFC project “Hevea breeding in West-Africa”. We also thank the companies SOGB (Société Grand Bereby) and SAPH (Société Africaine de Plantations d'Hévéa) for their logistical assistance in the field experiments in Côte d'Ivoire. We are grateful to the UMR AGAP genotyping technology platform (CIRAD, Montpellier) and the CIRAD-UMR AGAP HPC data centre of the South Green bioinformatics platform (http://www.southgreen.fr/) for their assistance. Furthermore, this project received the support of MESO@LR-Platform at the University of Montpellier.

Data availability

Data will be made available on request.

References

Adeyemo, E., Bajgain, P., Conley, E., Sallam, A.H., Anderson, J.A., 2020. Optimizing training population size and content to improve prediction accuracy of FHB-related traits in wheat. Article 4 Agronomy 10 (4). https://doi.org/10.3390/agronomy10040543.

Ahmadi, N., Bartholomé, J. (Eds.), 2022. Genomic Prediction of Complex Traits: Methods and Protocols, 2467. Springer US. https://doi.org/10.1007/978-1-0716-2205-6.

Alemu, A., Åstrand, J., Montesinos-López, O.A., Isidro y Sánchez, J., Fernández-Gónzalez, J., Tadesse, W., Vetukuri, R.R., Carlsson, A.S., Ceplitis, A., Crossa, J., Ortiz, R., Chawade, A., 2024. Genomic selection in plant breeding: Key factors shaping two decades of progress. Mol. Plant 17 (4), 552–578. https://doi.org/10.1016/j.molp.2024.03.007.

Amadeu, R.R., Muñoz, P.R., Zheng, C., Endelman, J.B., 2021. QTL mapping in outbred tetraploid (and diploid) diallel populations. Genetics 219 (3), iyab124. https://doi.org/10.1093/genetics/iyab124.

Aono, A.H., Francisco, F.R., Souza, L.M., Gonçalves, P. de S., Scaloppi Junior, E.J., Le Guen, V., Fritsche-Neto, R., Gorjanc, G., Quiles, M.G., de Souza, A.P., 2022. A divide-and-conquer approach for genomic prediction in rubber tree using machine learning. Sci. Rep. 12 (1), 18023. https://doi.org/10.1038/s41598-022-20416-z.

Asoro, F.G., Newell, M.A., Beavis, W.D., Scott, M.P., Jannink, J.-L., 2011. Accuracy and training population design for genomic selection on quantitative traits in elite north american oats. Plant Genome 4 (2). https://doi.org/10.3835/plantgenome2011.02.0007.

Bager, S.L., Persson, U.M., Reis, T.N.P. dos, 2021. Eighty-six EU policy options for reducing imported deforestation. One Earth 4 (2), 289–306. https://doi.org/10.1016/j.oneear.2021.01.011.

Balocchi, C., Durán, R., Nuñez, P., Ordoñez, J., Ramírez, M., Zapata-Valenzuela, J., 2023. Genomic selection: an effective tool for operational Eucalyptus globulus clonal selection. Tree Genet. Genomes 19 (4), 32. https://doi.org/10.1007/s11295-023-01609-7.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using **lme4**. J. Stat. Softw. 67 (1). https://doi.org/10.18637/jss.v067.i01.

Bouvet, J.-M., Makouanzi Ekomono, C.G., Brendel, O., Laclau, J.-P., Bouillet, J.-P., Epron, D., 2020. Selecting for water use efficiency, wood chemical traits and biomass with genomic selection in a Eucalyptus breeding program. For. Ecol. Manag. 465, 118092. https://doi.org/10.1016/j.foreco.2020.118092.

Brauner, P.C., Müller, D., Molenaar, W.S., Melchinger, A.E., 2020. Genomic prediction with multiple biparental families. Theor. Appl. Genet. 133 (1), 133–147. https://doi.org/10.1007/s00122-019-03445-7.

Broman, K.W., Wu, H., Sen, S., Churchill, G.A., 2003. R/qtl: QTL mapping in experimental crosses. Bioinforma. (Oxf., Engl.) 19 (7), 889–890. https://doi.org/10.1093/bioinformatics/btg112.

Broman, K.W., Gatti, D.M., Simecek, P., Furlotte, N.A., Prins, P., Sen, Ś., Yandell, B.S., Churchill, G.A., 2019. R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. Genetics 211 (2), 495–502. https://doi.org/10.1534/genetics.118.301595.

Browning, B.L., Zhou, Y., Browning, S.R., 2018. A one-penny imputed genome from next-generation reference panels. Am. J. Hum. Genet. 103 (3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015.

Caamal-Pat, D., Pérez-Rodríguez, P., Crossa, J., Velasco-Cruz, C., Pérez-Elizalde, S., Vázquez-Peña, M., 2021. lme4GS: an R-package for genomic selection. Front. Genet. 12. https://doi.org/10.3389/fgene.2021.680569.

CIRAD. (2011). *Rubberclone.* ⟨https://rubberclones.cirad.fr/index.php/clones/RRIC100⟩.

Clément-Demange, A., Priyadarshan, P.M., Thuy Hoa, T.T., Venkatachalam, P., 2007. Hevea Rubber Breeding and Genetics. In *Plant Breeding Reviews*. John Wiley & Sons, Ltd., pp. 177–283. https://doi.org/10.1002/9780470168035.ch4

Cooper, M., Van Eeuwijk, F.A., Hammer, G.L., Podlich, D.W., Messina, C., 2009. Modeling QTL for complex traits: detection and context for plant breeding. Curr. Opin. Plant Biol. 12 (2), 231–240. https://doi.org/10.1016/j.pbi.2009.01.006.

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. Ind. Crops Prod. 138, 111464. https://doi.org/10.1016/j.indcrop.2019.111464.

Daetwyler, H.D., Calus, M.P.L., Pong-Wong, R., de los Campos, G., Hickey, J.M., 2013. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193 (2), 347–365. https://doi.org/10.1534/genetics.112.147983.

Darkwa, K., Agre, P., Olasanmi, B., Kolade, O., Mournet, P., Lopez-Montes, A., Dekoeyer, D., Adebola, P., Kumar, L.P., Asiedu, R., & Asfaw, A. (2020). *A SNP-Based Linkage Map and QTL Identification for Resistance to Yam Anthracnose Disease (YAD) in Water Yam (Dioscorea alata)* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-26806/v1.

Duarte, D., Jurcic, E.J., Dutour, J., Villalba, P.V., Centurión, C., Grattapaglia, D., Cappa, E.P., 2024. Genomic selection in forest trees comes to life: Unraveling its potential in an advanced four-generation Eucalyptus grandis population. Front. Plant Sci. 15. https://doi.org/10.3389/fpls.2024.1462285.

Duhnen, A., Gras, A., Teyssèdre, S., Romestant, M., Claustres, B., Daydé, J., Mangin, B., 2017. Genomic selection for yield and seed protein content in soybean: a study of breeding program data and assessment of prediction accuracy. Crop Sci. 57 (3), 1325–1337. https://doi.org/10.2135/cropsci2016.06.0496.

Elabo, A.E.A., Atsin, G.J.O., Soumahin, E.F., Okoma, K.M., Gnagne, Y.M., Démange, A.C., Obouayeba, S., 2019. Agronomic performance of Hevea brasiliensis Muell. Arg. Clone IRCA 331 Grown in Southwestern Côte d'Ivoire. Asian J. Res. Agric. For. 1–10. https://doi.org/10.9734/ajraf/2019/v4i330063.

FAO. (2022). *Land statistics 2001–2022. Global, regional and country trends.* Statistics. ⟨https://www.fao.org/statistics/highlights-archive/highlights-detail/land-statistics-2001-2022.-global–regional-and-country-trends/⟩.

Fox, J., Castella, J.-C., 2013. Expansion of rubber (*Hevea brasiliensis*) in Mainland Southeast Asia: What are the prospects for smallholders? J. Peasant Stud. 40 (1), 155–170. https://doi.org/10.1080/03066150.2012.750605.

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., Jenkins, J., Martin, G., Charron, C., Hervouet, C., Costet, L., Yahiaoui, N., Healey, A., Sims, D., Cherukuri, Y., Sreedasyam, A., Kilian, A., Chan, A., Van Sluys, M.-A., 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nat. Commun. 9 (1), 2638. https://doi.org/10.1038/s41467-018-05051-5.

Gazaffi, R., Margarido, G.R.A., Pastina, M.M., Mollinari, M., Garcia, A.A.F., 2014. A model for quantitative trait loci mapping, linkage phase, and segregation pattern estimation for a full-sib progeny. Tree Genet. Genomes 10 (4), 791–801. https://doi.org/10.1007/s11295-013-0664-2.

Goddard, M. e, Hayes, B. j, 2007. Genomic selection. J. Anim. Breed. Genet. 124 (6), 323–330. https://doi.org/10.1111/j.1439-0388.2007.00702.x.

Grattapaglia, D., 2014. Breeding Forest Trees by Genomic Selection: Current Progress and the Way Forward. In: Tuberosa, R., Graner, A., Frison, E. (Eds.), Genomics of Plant Genetic Resources. Springer Netherlands, pp. 651–682. https://doi.org/10.1007/978-94-007-7572-5_26.

Grattapaglia, D., 2022. Twelve years into genomic selection in forest trees: climbing the slope of enlightenment of marker assisted tree breeding. Article 10 Forests 13 (10). https://doi.org/10.3390/f13101554.

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. Front. Plant Sci. 9. ⟨https://www.frontiersin.org/articles/10.3389/fpls.2018.01693⟩.

Guo, Z., Tucker, D.M., Basten, C.J., Gandhi, H., Ersoz, E., Guo, B., Xu, Z., Wang, D., Gay, G., 2014. The impact of population structure on genomic prediction in stratified populations. Theor. Appl. Genet. 127 (3), 749–762. https://doi.org/10.1007/s00122-013-2255-x.

Habier, D., Fernando, R.L., Dekkers, J.C.M., 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177 (4), 2389–2397. https://doi.org/10.1534/genetics.107.081190.

Hao, B.-Z., Wu, J.-L., 2000. Laticifer differentiation in *Hevea brasiliensis*: induction by exogenous jasmonic acid and linolenic acid. Ann. Bot. 85 (1), 37–43. https://doi.org/10.1006/anbo.1999.0995.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard, M.E., 2009. *Invited review*: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92 (2), 433–443. https://doi.org/10.3168/jds.2008-1646.

Heidari, B., Sayed-Tabatabaei, B.E., Saeidi, G., Kearsey, M., Suenaga, K., 2011. Mapping QTL for grain yield, yield components, and spike features in a doubled haploid population of bread wheat. Genome 54 (6), 517–527. https://doi.org/10.1139/g11-017.

Henderson, C.R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31 (2), 423–447. https://doi.org/10.2307/2529430.

IRSG, 2024. IRSG rubber statistical bulletin. 78(7–9).⟨https://www.rubberstudy.org/⟩.

Isidro y Sánchez, J., Akdemir, D., 2021. Training set optimization for sparse phenotyping in genomic selection: a conceptual overview. Front. Plant Sci. 12, 715910. https://doi.org/10.3389/fpls.2021.715910.

Ismawanto, S., Aji, M., Lopez, D., Mournet, P., Gohet, E., Syafaah, A., Bonal, F., Oktavia, F., Taryono, T., Subandiyah, S., Montoro, P., 2024. Genetic analysis of agronomic and physiological traits associated with latex yield revealed complex genetic bases in Hevea brasiliensis. Heliyon 10 (13), e33421. https://doi.org/10.1016/j.heliyon.2024.e33421.

Juliana, P., He, X., Poland, J., Roy, K.K., Malaker, P.K., Mishra, V.K., Chand, R., Shrestha, S., Kumar, U., Roy, C., Gahtyari, N.C., Joshi, A.K., Singh, R.P., Singh, P.K., 2022. Genomic selection for spot blotch in bread wheat breeding panels, full-sibs and half-sibs and index-based selection for spot blotch, heading and plant height. Theor. Appl. Genet. 135 (6), 1965–1983. https://doi.org/10.1007/s00122-022-04087-y.

Kumeh, E.M., Ramcilovic-Suominen, S., 2023. Is the EU shirking responsibility for its deforestation footprint in tropical countries? Power, material, and epistemic inequalities in the EU's global environmental governance. Sustain. Sci. 18 (2), 599–616. https://doi.org/10.1007/s11625-023-01302-7.

Legarra, A., 2017. *Bases for Genomic Prediction* (Version 0.9.6; p. 81). INRA. ⟨http://genoweb.toulouse.inra.fr/~alegarra/GSIP_git.pdf⟩.

Li, Y., Ruperao, P., Batley, J., Edwards, D., Martin, W., Hobson, K., Sutton, T., 2022. Genomic prediction of preliminary yield trials in chickpea: effect of functional annotation of SNPs and environment. Plant Genome 15 (1), e20166. https://doi.org/10.1002/tpg2.20166.

Lopez, D. (2023). *PB260 draft assembly* [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.10281549.

Lorenz, A.J., Smith, K. p, Jannink, J.-L., 2012. Potential and optimization of genomic selection for fusarium head blight resistance in six-row barley. Crop Sci. 52 (4), 1609–1621. https://doi.org/10.2135/cropsci2011.09.0503.

Marattukalam, J.G., Saraswathyamma, C.K., 1992. Propagation and Planting. In: In *Developments in Crop Science*, 23. Elsevier, pp. 164–199. https://doi.org/10.1016/B978-0-444-88329-2.50014-3.

McElroy, M.S., Navarro, A.J.R., Mustiga, G., Stack, C., Gezan, S., Peña, G., Sarabia, W., Saquicela, D., Sotomayor, I., Douglas, G.M., Migicovsky, Z., Amores, F., Tarqui, O., Myles, S., Motamayor, J.C., 2018. Prediction of Cacao (Theobroma cacao) resistance to Moniliophthora spp. diseases via genome-wide association analysis and genomic selection. Front. Plant Sci. 9. ⟨https://www.frontiersin.org/articles/10.3389/fpls.2018.00343⟩.

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.

Morgante, F., Huang, W., Maltecca, C., Mackay, T.F.C., 2018. Effect of genetic architecture on the prediction accuracy of quantitative traits in samples of unrelated individuals. Heredity 120 (6), 500–514. https://doi.org/10.1038/s41437-017-0043-0.

Mournet, P., De Albuquerque, P.S.B., Alves, R.M., Silva-Werneck, J.O., Rivallan, R., Marcellino, L.H., Clément, D., 2020. A reference high-density genetic map of Theobroma grandiflorum (Willd. Ex Spreng) and QTL detection for resistance to witches' broom disease (Moniliophthora perniciosa). Tree Genet. Genomes 16 (6), 89. https://doi.org/10.1007/s11295-020-01479-3.

Munyengwa, N., Le Guen, V., Bille, H.N., Souza, L.M., Clément-Demange, A., Mournet, P., Masson, A., Soumahoro, M., Kouassi, D., Cros, D., 2021. Optimizing imputation of marker data from genotyping-by-sequencing (GBS) for genomic selection in non-model species: Rubber tree (Hevea brasiliensis) as a case study. Genomics 113 (2), 655–668. https://doi.org/10.1016/j.ygeno.2021.01.012.

Nsibi, M. (2021). *Contribution des stratégies de sélection génomique et phénomique aux programmes d'amélioration génétique de l'abricotier (Prunus armeniaca L.) pour quelques traits d'intérêt* [These de doctorat, Montpellier, SupAgro]. ⟨https://theses.fr/2021NSAM0021⟩.

Nsibi, M., Gouble, B., Bureau, S., Flutre, T., Sauvage, C., Audergon, J.-M., Regnard, J.-L., 2020. Adoption and optimization of genomic selection to sustain breeding for apricot fruit quality. G3: Genes|Genomes|Genet. 10 (12), 4513–4529. https://doi.org/10.1534/g3.120.401452.

Pérez, P., de los Campos, G., 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics 198 (2), 483–495. https://doi.org/10.1534/genetics.114.164442.

Pérez-Rodríguez, P., De Los Campos, G., 2022. Multitrait Bayesian shrinkage and variable selection models with the BGLR-R package. Genetics 222 (1), iyac112. https://doi.org/10.1093/genetics/iyac112.

Rio, S., Mary-Huard, T., Moreau, L., Charcosset, A., 2019. Genomic selection efficiency and a priori estimation of accuracy in a structured dent maize panel. Theor. Appl. Genet. 132 (1), 81–96. https://doi.org/10.1007/s00122-018-3196-1.

Rio, S., Gallego-Sánchez, L., Montilla-Bascón, G., Canales, F.J., Isidro y Sánchez, J., Prats, E., 2021. Genomic prediction and training set optimization in a structured Mediterranean oat population. Theor. Appl. Genet. 134 (11), 3595–3609. https://doi.org/10.1007/s00122-021-03916-w.

Robert-Granié, C., Legarra, A., Ducrocq, V., 2011. Principes de base de la sélection génomique. Article 4 INRAE Prod. Anim. 24 (4). https://doi.org/10.20870/productions-animales.2011.24.4.3265.

de Roos, A.P.W., Hayes, B.J., Goddard, M.E., 2009. Reliability of genomic predictions across multiple populations. Genetics 183 (4), 1545–1553. https://doi.org/10.1534/genetics.109.104935.

Schopp, P., Müller, D., Wientjes, Y.C.J., Melchinger, A.E., 2017. Genomic prediction within and across biparental families: means and variances of prediction accuracy and usefulness of deterministic equations. G3 Genes|Genomes|Genet. 7 (11), 3571–3586. https://doi.org/10.1534/g3.117.300076.

Seyum, E.G., Bille, N.H., Abtew, W.G., Munyengwa, N., Bell, J.M., Cros, D., 2022. Genomic selection in tropical perennial crops and plantation trees: a review. Mol. Breed. 42 (10), 58. https://doi.org/10.1007/s11032-022-01326-4.

Simmonds, N., 1989. Rubber breeding. In: Webster, C.C., Baulkwill, W.J. (Eds.), Rubber. Longman London, pp. 85–124.

Souza, L.M., Francisco, F.O., Gonçalves, P.D.S., Scaloppi Junior, E.J., Le Guen, V., Fritsche-Neto, R., Souza, A.P., 2019. Genomic selection in rubber tree breeding: A comparison of models and methods for managing G×E interactions. Front. Plant Sci. https://doi.org/10.3389/fpls.2019.01353.

Van Helten, W. (1918). *Het oculeeren van Hevea*.

VanRaden, P.M. (2007). Genomic measures of relationship and inbreeding. *Interbull Bulletin*, *37*, Article 37. ⟨https://journal.interbull.org/index.php/ib/article/view/1400⟩.

Vignal, A., Milan, D., SanCristobal, M., Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34 (3), 275. https://doi.org/10.1186/1297-9686-34-3-275.

Voss-Fels, K.P., Cooper, M., Hayes, B.J., 2019. Accelerating crop genetic gains with genomic selection. Tag. Theor. Appl. Genet. Theor. Und Angew. Genet. 132 (3), 669–686. https://doi.org/10.1007/s00122-018-3270-8.

Wang, Y., Hollingsworth, P.M., Zhai, D., West, C.D., Green, J.M.H., Chen, H., Hurni, K., Su, Y., Warren-Thomas, E., Xu, J., Ahrends, A., 2023. High-resolution maps show that rubber causes substantial deforestation. Nature 623 (7986), 340–346. https://doi.org/10.1038/s41586-023-06642-z.

Webster, C.C., Baulkwill, W.J., 1995. Rubber. Wiley. ⟨https://books.google.ci/books?id=cvvFQgAACAAJ⟩.

Wiggans, G.R., Cole, J.B., Hubbard, S.M., Sonstegard, T.S., 2017. Genomic selection in dairy cattle: the USDA experience. Annu. Rev. Anim. Biosci. 5 (1), 309–327. https://doi.org/10.1146/annurev-animal-021815-111422.

Windhausen, V.S., Atlin, G.N., Hickey, J.M., Crossa, J., Jannink, J.-L., Sorrells, M.E., Raman, B., Cairns, J.E., Tarekegne, A., Semagn, K., Beyene, Y., Grudloyma, P., Technow, F., Riedelsheimer, C., Melchinger, A.E., 2012. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 Genes|Genomes|Genet. 2 (11), 1427–1436. https://doi.org/10.1534/g3.112.003699.

Wolc, A., Dekkers, J.C.M., 2022. Application of Bayesian genomic prediction methods to genome-wide association analyses. Genet. Sel. Evol. 54 (1), 31. https://doi.org/10.1186/s12711-022-00724-8.

Wu, W., Zhang, X., Deng, Z., An, Z., Huang, H., Li, W., Cheng, H., 2022. Ultrahigh-density genetic map construction and identification of quantitative trait loci for growth in rubber tree (Hevea brasiliensis). Ind. Crops Prod. 178, 114560. https://doi.org/10.1016/j.indcrop.2022.114560.

Zhang, Y., Zhang, M., Ye, J., Xu, Q., Feng, Y., Xu, S., Hu, D., Wei, X., Hu, P., Yang, Y., 2023. Integrating genome-wide association study into genomic selection for the prediction of agronomic traits in rice (Oryza sativa L.). Mol. Breed. 43 (11), 81. https://doi.org/10.1007/s11032-023-01423-y.