

Article

Identification of Candidate Gene-Based Markers for Girth Growth in Rubber Trees

Gunlayarat Bhusudsawang¹, Ratchanee Rattanawong², Thitaporn Phumichai³, Wirulda Pootakham⁴, Sithichoke Tangphatsornruang⁴  and Kittipat Ukoskit^{1,*}

¹ Department of Biotechnology, Faculty of Science and Technology, Rangsit Campus, Thammasat University, Pathum Thani 12120, Thailand; g33bh@tu.ac.th

² Nong Khai Rubber Research Center, Rubber Research Institute of Thailand, Rubber Authority of Thailand, Rattanawapi District, Nong Khai 43120, Thailand; ratchaneerattanawong@gmail.com

³ Rubber Research Institute of Thailand, Rubber Authority of Thailand, Bangkok 10900, Thailand; thitaporns@gmail.com

⁴ National Science and Technology Development Agency, Thailand Science Park, Khlong Luang, Pathum Thani 12120, Thailand; wirulda.poo@nstda.or.th (W.P.); sithichoke.tan@nstda.or.th (S.T.)

* Correspondence: ku@tu.ac.th; Tel.: +66-2564-4444 (ext. 2458)

Abstract: Girth growth is an important factor in both latex and timber production of the rubber tree. In this study, we performed candidate gene association mapping for girth growth in rubber trees using intron length polymorphism markers (ILP) in identifying the candidate genes responsible for girth growth. The COBL064_1 marker developed from the candidate gene (*COBL4*) regulating cellulose deposition and oriented cell expansion in the plant cell wall showed the strongest association with girth growth across two seasons in the Amazonian population and was validated in the breeding lines. We then applied single molecule real-time (SMRT) circular consensus sequencing (CCS) to analyze a wider gene region of the *COBL4* to pinpoint the single nucleotide polymorphism (SNP) that best explains the association with the traits. A SNP in the 3' UTR showing linkage disequilibrium with the COBL064_1 most associated with girth growth. This study showed that the cost-effective method of ILP gene-based markers can assist in identification of SNPs in the candidate gene associated with girth growth. The SNP markers identified in this study added useful markers for the improvement of girth growth in rubber tree breeding programs.

Keywords: candidate gene association mapping; intron length polymorphism; single nucleotide polymorphism; girth growth; rubber tree



Citation: Bhusudsawang, G.; Rattanawong, R.; Phumichai, T.; Pootakham, W.; Tangphatsornruang, S.; Ukoskit, K. Identification of Candidate Gene-Based Markers for Girth Growth in Rubber Trees. *Plants* **2021**, *10*, 1440. <https://doi.org/10.3390/plants10071440>

Academic Editor: Aakash Chawade

Received: 21 June 2021

Accepted: 12 July 2021

Published: 14 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Girth growth of the rubber tree, *Hevea brasiliensis* (Willd.ex ADR.de Juss.) Muell. Arg., is the most important indicator of the maturity of the plantation, on which the harvesting of latex is based. Wood derived from the trunk of the rubber tree, harvested during periodic replanting, has emerged as an important raw material, especially in counties where logging of the natural forest is restricted. Breeding for girth growth is time-consuming, as the rubber tree has a long juvenile period, non-synchronous flowering, low fruit setting, and a heterozygous nature. Propagation of new varieties requires a minimum of 20–25 years of experimentation. To improve the rubber productivity, a deeper knowledge of the genetic factors responsible for variation in girth growth is necessary.

Genome-wide association mapping for girth growth in Amazonian accessions grown in a suboptimal climate zone has identified candidate genes for adaptation to suboptimal climate zone [1]. However, none of the functional genes associated with secondary growth had been in the SNP set analyzed previously. The low linkage disequilibrium (LD) was also reported in the Amazonian population. Successful application of genome scanning to outcrossing species such as rubber tree will require a very high number of markers, making

the candidate gene method more attractive [1,2]. The candidate gene approach has proven successful in many instances, such as in Arabidopsis [3] wheat [4,5], pea [6], potato [7], and perennial ryegrass [8]. Secondary growth has been extensively studied, and several of the genes involved have been identified. Candidate genes for cellulose and lignin biosynthesis and cell wall development, and transcription factors involved in wood development, have been reported for loblolly pine [9] and eucalyptus [10–13]. The allelic variation of these candidate genes may be used to establish associations between targeted genes with known function and girth growth in rubber tree.

Gene-based or functional markers are derived from sequence polymorphisms within genes that are directly associated with phenotypic variation. Gene-based markers accurately discriminate between traits associated with alleles of a target gene and are ideal molecular markers for marker assisted selection in breeding [14]. Although high throughput sequencing techniques are being developed nowadays, the use of variable markers present in candidate genes is still an interesting alternative for association studies. An approach known as intron length polymorphism (ILP), based on the alignment of EST cDNA sequences against known genomic sequences, has been applied to a range of crop species for genetic diversity, evolutionary and association genetic studies [15–21]. As ILP markers are gene-based, co-dominant, neutral, convenient, reliable and cost effective, and can play a vital role in regulation of gene expression [22], this is a promising approach to genetic studies such as candidate gene association mapping, which seeks to identify allelic variation in the genes that influence girth growth.

In the present study, we developed a set of ILP markers from candidate genes involved in secondary growth and performed candidate gene association mapping for girth growth in Amazonian accessions. Rubber tree breeding lines were used to validate the effectiveness of the ILP markers identified by association mapping. We applied targeted sequencing using Pacific Biosciences single molecule real-time (SMRT) circular consensus sequencing (CCS) technology to determine the sequence variation of the strongest associated candidate gene influencing girth growth. We then conducted candidate gene association mapping in order to exploit SNPs for downstream applications to support rubber tree breeding programs. Our study showed that the cost-effective method of ILP gene-based markers can assist in identification of SNPs in the candidate gene associated with girth growth. The SNP markers identified in this study added useful markers for the improvement of girth growth in rubber tree breeding programs.

2. Results

2.1. Development and Identification of ILP Markers

ILP variation was investigated in 26 genes from four function classes. A total of 45 ESTs (Table S1) were aligned with whole-genome shotgun contigs of the rubber tree to predict the position and number of the introns. The 45 ESTs showed significant hits with 158 contigs of the whole-genome shotgun sequences. Of 530 predicted introns, 235 introns with sizes smaller than 800 bp were selected for primer synthesis and used in the evaluation of length polymorphism. Of the 235 ILP primer pairs examined, 228 pairs successfully amplified genomic DNA samples. We selected 115 polymorphic ILP markers (Table S2) based on their predetermined intronic-Indel size and PIC. The average size of the amplicon was 284 bp, ranging from 150 (CesA004) to 759 bp (PAL117). The selected ILP markers yielded 470 alleles, with an average of 4.20 alleles per locus. The number of alleles ranged from 2 to 12. PIC values were between 0.06 and 0.88, with an average of 0.44. The observed heterozygosity ranged from 0.02 to 1 and the expected heterozygosity from 0.10 to 0.92, with average values of 0.59 and 0.47, respectively.

2.2. Population Structure and LD Analysis of ILP Markers

We further tested the 115 polymorphic ILP markers to evaluate the genetic population structure of the Amazonian accessions used in the present study. The model-based Bayesian analysis, combined with computation of the Evanno ΔK statistics (Figure S1a), suggested

two clusters ($K = 2$) with admixed accessions. Cluster 1 (green bars) comprised 13 accessions from Acre, one from Mato Grosso, and 71 from Rondonia. Cluster 2 (red bars) comprised 52 accessions from Mato Grosso and eight from Rondonia (Figure S1b). The accession with unknown origin (BRAZIL_UN) was placed into Cluster 2. Twenty-four accessions had a membership probability (Q value) below 0.7 and were defined as admixtures. Of these, one originated from Acre, 11 from Mato Grosso, and 12 from Rondonia. The PCA results (Figure S1c) were consistent with those from the admixture model. The accessions from Mato Grosso formed a group that was distinct from the Acre and Rondonia cluster.

The population LD was analyzed using the 115 polymorphic ILP markers. Based on r^2 estimates, 59 of the marker pairs showed $r^2 \geq 0.1$ (Table S3). Of these pairs, 15 pairwise comparisons were intragenic LD with an average r^2 of 0.18. The seven highest r^2 ($0.30 < r^2 < 0.64$) with an average of 0.46 represented intragenic LD between ILP markers of four genes: *LIM*, *CesA3*, *APL*, and *MYB*. Forty-four marker pairs were intergenic LD, with an average r^2 of 0.14. Among intergenic LD, 35 marker pairs were LD between ILP developed from transcription factors and functional genes and nine were LD between ILP developed from functional genes.

2.3. Association Mapping of ILP Markers

The results of phenotype statistics of GiD and GiW were described in the previous work [1]. For GiC, the mean was 4.49 cm, with the maximum of 6.96 cm and the minimum of 2.30 cm. Significant positive phenotypic correlations ($p < 0.01$) were found between GiD and GiC, and between GiW and GiC, with r values of 0.51 and 0.61, respectively. Different association mapping models were compared, and the proportion of significant results was calculated. Q and PCA yielded similar Q-Q plots (Figure S2), only the three models GLM, GLM + Q, and MLM + K + Q are presented for association analysis. The naive GLM model detected the largest number (25) of significant associations for nine candidate genes and three traits studied (Table S4). This model did not take account of possible confounding effects and the associated false positives so that the cumulative distribution of p values was greatest from the observed p values. The GLM + Q models performed better than the naive model but still deviated from the expected values. These models yielded 13 significant associations for six candidate genes and all traits. The MLM + Q + K model showed the smallest departure from expectations in the Q-Q plots. It detected nine marker–trait associations for four candidate genes and three traits. All three models displayed similar distributions of p values for GiC, but the MLM + Q + K model still deviated less from the expected values than the GLM or GLM + Q models and detected a single marker–trait association for GiC. As the MLM + Q + K model was superior at accounting for spurious associations resulting from population structure and/or familial relatedness, the results from this model are discussed next (Table 1). A total of nine marker–trait associations involving four ILP markers developed from three candidate genes were detected. Marker–trait associations of COBL064_1, and PIN227_1 were significantly associated with GiD and GiW, using a Bonferroni corrected cut-off (with $p \leq 4.3 \times 10^{-4}$). These marker–trait associations explained the phenotypic variation, with ranges from 4.9% (CaS108) to 14.6% (COBL064_1)

Table 1. Markers and trait association from the MLM + Q + K model.

Trait ¹	Marker ²	p Value	Phenotypic Variation (%)
GiD	COBL064_1 *	2.0×10^{-7}	14.0
	PIN227_1 *	1.0×10^{-5}	10.2
	PIN220 *	1.1×10^{-4}	7.7
	CaS108	9.2×10^{-4}	4.9
GiW	COBL064_1 *	4.32×10^{-7}	14.6
	PIN227_1 *	4.82×10^{-5}	10.0
	PIN220	4.7×10^{-4}	7.3
	CaS108	1.8×10^{-3}	4.9
GiC	CAD076	1.6×10^{-3}	8.9

¹ GiD = girth in the dry season, GiW = girth in the wet season, and GiC = total girth increment; ² * Significant after Bonferroni correction p value = 4.3×10^{-4} .

2.4. Validation of the Associated Markers in Rubber Tree Breeding Lines

For the breeding lines used to validate the markers identified by association mapping, girth was lower in the dry season than in the wet season. The mean GiD was 22.37 cm, with the largest girth of 28.66 cm and the smallest girth of 13.33 cm. The mean GiW was 26.15 cm, with a maximum of 33.28 cm and a minimum of 15.73 cm. The mean GiC was 13.49 cm, with the maximum of 21.62 cm and the minimum of 7.40 cm. Significant positive phenotypic correlations ($p < 0.01$) were found between GiD and GiC, and between GiW and GiC, with r values of 0.63 and 0.54, respectively.

ILP markers identified in the Amazonian population as associated with the girth growth traits were tested by looking for significant markers in the rubber tree breeding lines, to determine whether the association between marker and trait was maintained over the breeding lines. The results revealed that one allele of COBL064_1 and three of PIN227_1 remained significantly ($p < 0.05$ or 0.01) associated with GiD and GiW in the rubber tree breeding lines, and one allele of CAD76 with GiC (Table 2). In the case of COBL064_1, individuals carrying the 380-bp allele showed a significant ($p < 0.05$) reduction in GiD and GiW. In the case of PIN227_1, individuals carrying the 261-bp allele showed a significant increase in GiD and GiW, whereas individuals carrying the 289- and 298-bp alleles showed a significant reduction in both. In the case of CAD076, individuals carrying the 263-bp allele showed a significant reduction in GiC.

Table 2. Allelic effect on girth growth traits of each significant ILP marker in rubber tree breeding lines.

Marker	Size (bp)	GiD		GiW		GiC	
		p-Value	Effect	p-Value	Effect	p-Value	Effect
COBL4	380	0.033	−1.52	0.044	−1.68	ns	−0.08
	261	0.01	1.22	0.026	1.16	ns	−0.20
PIN227	289	0.005	−1.66	0.01	−1.74	ns	−0.29
	298	0.013	−2.02	0.027	−2.17	ns	−0.41
PIN220	461	0.025	3.98	0.023	4.70	ns	3.32
CAD76	263	ns	−0.93	ns	−1.07	0.017	−1.49
	312	ns	0.41	ns	0.49	0.043	1.25

ns = non-significant.

2.5. Identification of Polymorphism and Linkage Disequilibrium within COBL Gene

Given the strongest association between the ILP marker in the COBL gene with GiD and GiW, we further investigated this particular variant using SMRT CCS technology. Three specific primer pairs yielded high-quality sequences. The sequencing of the COBL gene with a total length of 4236 bp revealed 128 SNPs and 178 indels. The amplified regions covered 6 exons, 5 introns, and the 3' UTR. Of 306 polymorphic sites, 72 and 83 were

located in exons, and introns, respectively, and 151 polymorphic sites within the 3' UTR. The LD between pairs of polymorphic sites ranged from 6.7×10^{-7} to 1.0 with an average $r^2 = 0.24$. Plotting r^2 values against physical distance between polymorphic pairs (Figure 1) indicated the LD decay of 650 bp. i.e., the maximum physical distance for genetically linked markers. This result suggested that the LD of the SNPs within this gene did not extend over the entire gene region.

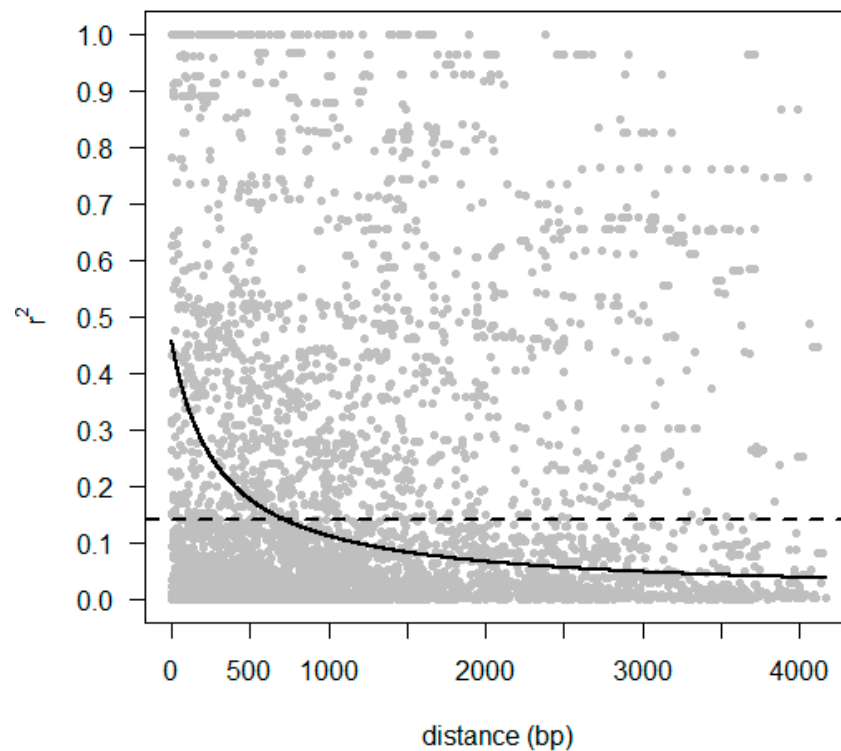


Figure 1. Pairwise LD (r^2) values were plotted against the physical distance between all pairs of polymorphic sites (SNP/Indel) within the *COBL* gene. The trend line of the nonlinear regressions against physical distance is given by the solid line. The horizontal dotted line indicates the 95% percentile of the distribution of the unlinked r^2 , which gives the critical value of r^2 . The intersection point with the nonlinear regression curve, which determines LD decay, marks the threshold for the maximum physical distance between genetically linked markers.

2.6. Association between SNP/Indel within *COBL* and Growth Girth

When pairs of adjacent loci were found to demonstrate LD, only one of the two SNPs from each pair was used in marker–trait association analysis, leaving 55 SNPs and 162 Indels in the candidate gene. Of these, 63, 55 and 99 polymorphic sites were located in exons, introns, and the 3' UTR, respectively, of this gene. The FarmCPU model showed a straight line close to 1:1 with a slightly deviated tail in the Q-Q plots, suggesting that this model controlled both false positives and false negatives (Figure 2). FarmCPU showed a trend of association with GID and GIW toward 3' UTR of the *COBL* gene. The most associated position with GID (Table 3 and Figure 3) was the SNP at 5589 bp ($p = 6.50 \times 10^{-5}$; $q = 0.014$) explaining 11.1% of the phenotypic variance. The suggestive association ($p < 0.01$) with GID and GIW was the SNP at 3221 in intron5.

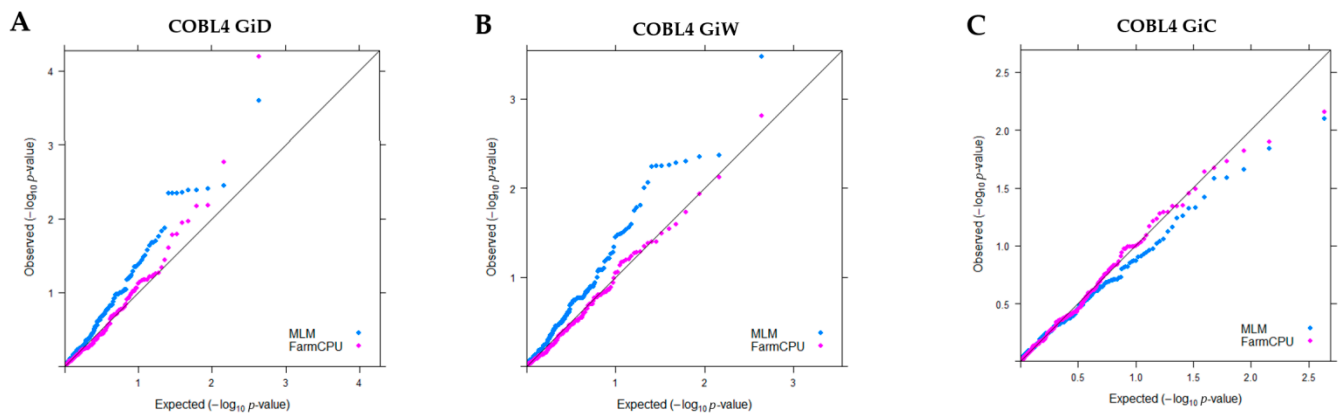


Figure 2. Quantile-quantile plots of estimated $-\log_{10}(p)$ from association analysis using MLM model and FarmCPU and three traits: (A) girth in the dry season (GiD) (B) girth in the wet season (GiW) and (C) total girth increment (GiC). The black line is the expected line under the null distribution. The blue line represents the observed p values using the MLM model, and the red line using FarmCPU.

Table 3. Markers and trait association mapping from FarmCPU model.

Trait ¹	Marker	Position ²	p -Value ³	Phenotypic Variation (%)
GiD	3' UTR	5589	6.50×10^{-5} *	11.1
	intron5	3221	1.69×10^{-3}	7.0
	intron1	1475	6.63×10^{-3}	6.1
	3' UTR	4406	6.69×10^{-3}	10.1
GiW	3' UTR	5589	1.56×10^{-3}	9.8
	intron5	3221	7.64×10^{-3}	6.4
GiC	3' UTR	5526	6.89×10^{-3}	3.6

¹ GiD = girth in the dry season, GiW = girth in the wet season, and GiC = total girth increment; ² Position in GenBank accession AJJZ010406963.1; ³ * Significant after FDR correction.

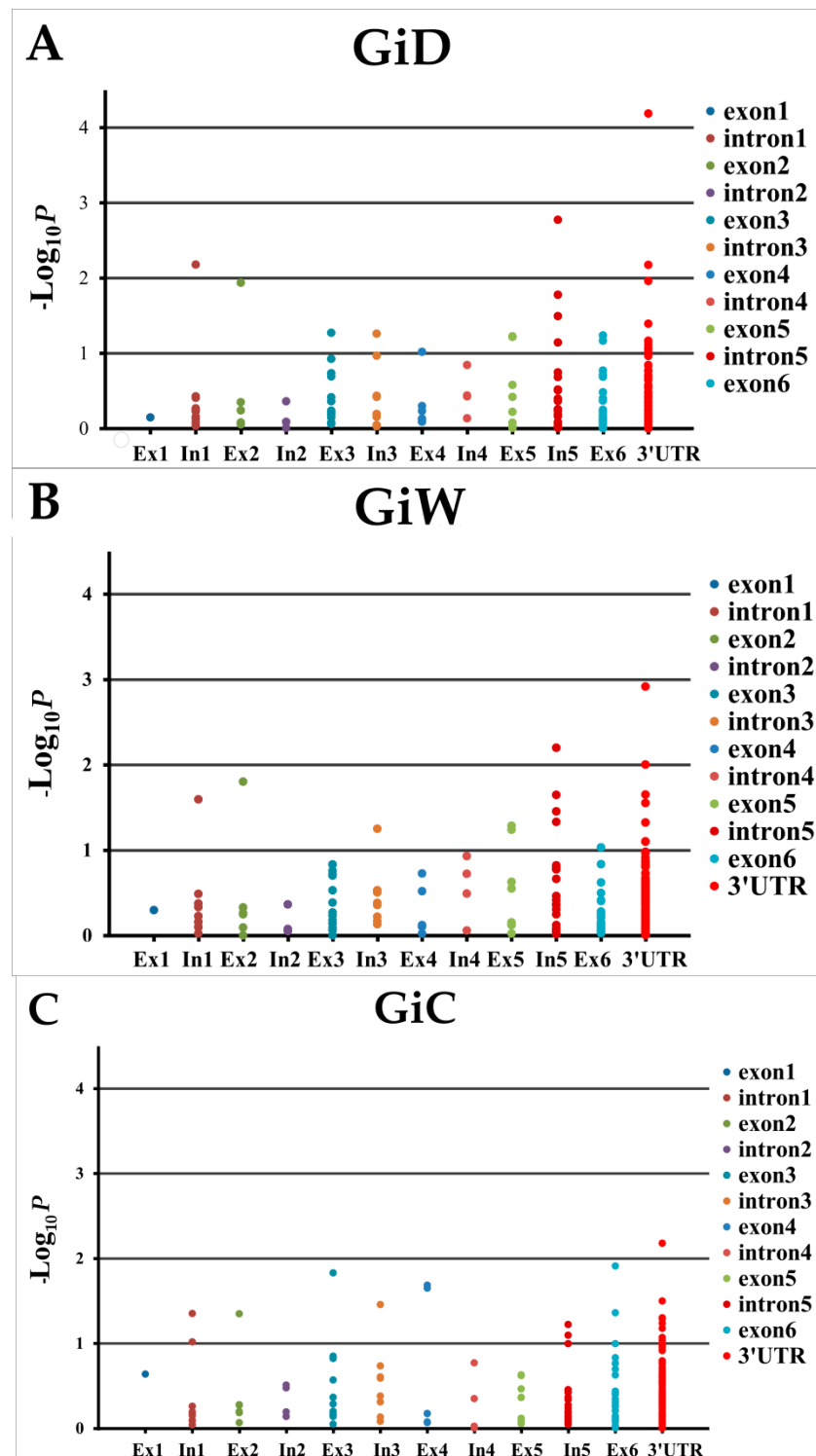


Figure 3. Marker–trait associations across the COBL gene. Associations were based on FarmCPU. The p-value of the marker is given at each position (p -value was decimal log transformed). (A) girth in the dry season (GiD), (B) girth in the wet season (GiW), and (C) total girth increment (GiC).

3. Discussion

Increasing the girth growth is one important step to further improve economic values of the rubber tree. In the present study, we demonstrated for the first time an association between girth growth traits and allelic variations at various candidate gene loci using the ILP markers. Comparison of the EST sequences with the whole-genome shotgun contigs

of the rubber tree enabled the design of 235 primer pairs for ILP markers. This primer design strategy explicitly focused on single-copy genes due to complexities generated by the presence of paralogous copies [23,24].

The high PCR success rate (97%) confirmed that the predicted exon/intron junction sites could be utilized to develop ILP markers in this study. A total of 115 polymorphic ILPs from 26 candidate genes involved in secondary growth, were selected. The PIC varied from 0.06 to 0.88, with an average of 0.44. The results were similar to the findings from EST-SSR markers [25] but higher than the results from ILP markers developed from a previous study [26]. Based on these ILP markers, model-based Bayesian analysis and PCA suggested that the accessions of Amazonian rubber trees fell into two clusters. These results were consistent with previous reports of two distinct clusters within this population, using SNPs [1] and comparable with those reported for the IRRDB collection [27–29]. Our results suggested that the intronic regions were sufficiently varied to act as marker resources when determining the population structure in rubber tree.

Gene-based markers are a type of marker with the most potential to bridge the gap between structural polymorphisms and functional diversity, since these gene-based markers are related to phenotypic variations. Here, we demonstrate the utility of ILP markers in identifying the best candidate gene for girth growth. COBL064_1 derived from *COBL4* showed the strongest association with girth across two seasons and explained the relatively high portion of the phenotypic variance. This marker also remained significantly associated with GiD and GiW in the rubber tree breeding lines. The *COBL4* gene is a member of the COBRA family, which encodes glycosyl phosphatidylinositol (GPI)-anchored proteins that regulate cellulose deposition and oriented cell expansion in the plant cell wall [30]. Linkage analysis in a full-sib family revealed that *COBL4* was the associated gene in a quantitative trait locus (QTL) region for cellulose content in *Eucalyptus* [10,31]. In the present study, the associated polymorphism of the COBL064_1 marker is located in the intron 5. The fact that sequence variants located within a noncoding region are associated with the variation in girth growth traits could be explained by two hypotheses. Firstly, the detected variants are in LD with causal polymorphisms in regulatory sequences that are responsible for the trait variance. Secondly, the nucleotide change within the intron is involved in regulatory functions like alternative splicing, and therefore, can affect the structure, function, and expression level of proteins [32,33]. The important next step to understanding the genetic contributions to the girth growth traits is identifying which SNP(s) are likely to be causal at this candidate gene.

In order to analyze whether genetic variants in *COBL4* might be mediating gene-based marker–trait association, we applied candidate gene association mapping to identify SNPs with the strongest association with the traits. Several studies using this approach in plants have been able to successfully identify associations between SNPs in candidate genes and phenotypic traits [3,7,34]. In the present study, the *COBL4* gene was selected for candidate gene association mapping due to the known functional relevance to growth properties, and its association with the girth growth traits in the natural population and the breeding lines. We applied targeted sequencing using SMRT technology to assess sequence variation. The long-read sequencing technologies such as PacBio SMRT can allow for the assembly of multigene family clusters and serve to more accurately characterize patterns of gene copy variation in gene families [35]. The CCS technology derives a consensus sequence from multiple passes of a single template molecule, producing accurate reads from noisy individual subreads [36–38].

The power of identifying SNP and mapping resolution for complex traits depend on the LD exploited in the population by the statistical model [39]. The fast LD decay over physical distance was observed in the *COBL4* in the Amazonian population.

From our results, FarmCPU, the multilocus model, provides a robust model for association mapping, effectively suppressing false positives. This result corresponds to the previous study demonstrating that the FarmCPU model could reduce both false positives and false negatives in soybean and maize with varying LD decays rates [40].

SMRT CCS of the candidate gene identified a SNP (CA5589) in the 3' UTR most associated with the girth growth traits. The detected SNP is in LD with COBL064_1 ($p = 0.006$). The intragenic LD decayed within 650 bp in this gene. Thus, the SNP showing genetic association is likely to be the causal variant or located in close proximity to the causative polymorphisms. The 3' UTR of messenger RNAs may contain cis-regulatory elements that affect gene expression by altering mRNA stability and translation [41]. Additional studies are needed to better elucidate the mechanisms underlying these putative associations. Overall, this study showed that the cost-effective method of ILP gene-based markers can assist in identification of SNPs in the candidate gene associated with girth growth. A combination of the SNP markers identified in this study and our previous report [1] can be applied to rubber tree breeding programs for improvement of girth growth by marker-assisted selection.

4. Materials and Methods

4.1. Plant Materials and Field Data

The Amazonian population comprised 170 accessions from three Brazilian states divided as follows: 14 from Acre, 91 from Rondoma, 64 from Mato Grosso and one of unknown origin. The details of the materials and field experiments were described in the previous study of Chanroj et al. [1]. Briefly, all trees were propagated by grafting with five replicates, and planted at a spacing of 3×7 m. Measurements were carried out individually on each tree. The experiment was conducted in 1994 at the Nong Khai Rubber Research Center (NKRR) of Nong Khai Province, the northeast region of Thailand. To evaluate the growth of individual trees, the girth (circumference at 1.7 m) was measured in centimeters (cm). Means were then calculated based on the total replicates. Measurements were taken in March (dry) and September (wet) in 2000, 2002, and 2004. The trees generally reached a latex tappable size. The total change in girth (GiC) was also calculated by subtraction from the circumferences in the 2000 and 2004 data. Three traits were used for association analysis, girth in the dry (GiD), and wet (GiW) seasons and GiC.

4.2. Candidate Gene Selection

Gene selection (Table 4) was based on earlier reports of secondary growth development and differential gene expression in *Eucalyptus* spp. [11,13,42,43] and of the genes involved in xylem/wood development in loblolly pine (*Pinus taeda* L.) [9]. Twenty six genes were selected from four function classes: four from cellulose synthesis (*CesA2*, *CesA3*, *CaS* and *COBL4*), eight from lignin synthesis (*4CL*, *C4H*, *CAD*, *CCoAMT*, *CCR*, *COMT*, *PAL*, and *Peroxidase2*), 12 associated with the transcription factors regulating secondary cell wall development (*BTF3*, *FRA2*, *HD-zip*, *KNAT*, *LIM*, *MOR1*, *MYB*, *NAC1*, *NtLIM1*, *PIN1*, *RIC1* and *APL*), and two from cell expansion (*COB*, and *KORRIGAN*).

Table 4. Candidate genes for secondary growth development used in the study.

Functional Class		Gene	Full Name	No. of ESTs	References
Cellulose synthesis	1	<i>CesA2</i>	Cellulose synthase2	1	[9,13]
	2	<i>CesA3</i>	Cellulose synthase3	1	[9,13]
	3	<i>COBL4</i>	COBRA-like protein 4	1	[10,13]
	4	<i>CaS</i>	Callose synthase 10	3	[9]
Lignin synthesis	5	<i>4CL</i>	4-coumarate:CoA ligase	1	[9,11–13]
	6	<i>CAD</i>	Cinnamyl alcohol dehydrogenase	3	[9–13]
	7	<i>COMT1</i>	Caffeic O-methyltransferase1	2	[9,11,13]
	8	<i>CCoAMT</i>	Caffeoyl-CoA O-methyltransferase	3	[9,10,13]
	9	<i>CCR</i>	Cinnamoyl-CoA reductase	2	[9–12]
	10	<i>C4H1</i>	Cinnamate 4-hydroxylase 1	2	[10,11]
	11	<i>PAL</i>	Phenylalanine ammonia-lyase	1	[9,11]
	12	<i>Peroxidase2</i>	Peroxidase 2	1	[9]

Table 4. Cont.

Functional Class		Gene	Full Name	No. of ESTs	References
Transcription	13	<i>LIM</i>	LIM gene for LIM transcription factor	2	[9,13]
factors	14	<i>MYB1</i>	MYB1 gene for MYB transcription factor 1	5	[9,10,13]
	15	<i>BTF3</i>	Transcription factor BTF3 homolog 4	2	[13]
	16	<i>HD-zip</i>	Homeobox-leucine zipper protein ANTHOCYANINLESS 2	1	[13]
	17	<i>NAC1</i>	NAC domain-containing protein 100-like	4	[13]
	18	<i>APL</i>	myb family transcription factor APL	1	[9]
	19	<i>KNAT</i>	Homeobox protein knotted-1-like	1	[9]
	20	<i>NHLIM1</i>	Eglim1 gene for transcription factor lim1	2	[10]
	21	<i>MOR1</i>	Protein MOR1	1	[9]
	22	<i>PIN1</i>	Auxin efflux carrier component 1-like	1	[9]
	23	<i>RIC1</i>	CRIB domain-containing protein RIC7	1	[9]
	24	<i>FRA2</i>	Katanin p60 ATPase-containing subunit A1	1	[9]
Cell expansion	25	<i>COB</i>	COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family	1	[9]
	26	<i>KORRIGAN</i>	Korrigan	1	[9,10]
Total	26			45	

4.3. ILP Marker Development

Gene transcripts were used as queries in a BLASTN program search of the whole-genome shotgun sequences of rubber tree cultivar RRIM 600 [44] in the NCBI database. Regions of similarity (E-value $\leq 10^{-4}$ and an identity > 65%) between selected sequences and whole-genome shotgun contigs were identified using BLAST. The position and length of introns within the rubber tree transcripts were predicted from the alignment results. The specific primers flanking the predicted intron positions were designed using the Primer3 program [45] with the default settings. Amplification by PCR was performed in a 20 μ L reaction mixture containing 10 ng template DNA, 1 \times PCR buffer (20 mM Tris pH 9.0, 100 mM KCl, 3.0 mM MgCl₂), 200 μ M of each of the four dNTPs, 1.5 μ M of each of the forward and reverse primers, and 0.5 U Taq DNA polymerase. Thermal cycling conditions were as follows: 3 min initial denaturation at 95 °C, 30 cycles of 30 s for denaturation at 95 °C, 30 s of annealing at 52–60 °C (depending on the combination of primers), a 30 s extension at 72 °C, and a final 5 min extension at 72 °C. PCR products were separated on 6% denaturing polyacrylamide gels and silver-stained. Since large intron fragments can produce inaccurate scores when differences in fragment size are small, and can reduce the rate of PCR amplification, the polymorphic ILP markers were selected between 150 and 800 bp in length. Allelic variation was calculated from the genotypic profiles of the ILP markers based on the number of alleles and the polymorphism information content (PIC; [46]), using PowerMarker 3.25 [47]. The genetic diversity parameters included the number of alleles (Na), expected heterozygosity (He), and observed heterozygosity (Ho).

4.4. Population Structure and Linkage Disequilibrium Analysis of ILP Markers

To test for the potential applications, the polymorphic ILP markers were used to evaluate the genetic population structure of the Amazonian accessions used in the present study. Two methods were used to determine the population structure. First, a two-dimensional diagram of principal coordinates analysis (PCA) was produced based on a genetic distance (GD) matrix by NTSYS-pc v. 2.0 [48]. The second method used the model-

based Bayesian analysis in the software STRUCTURE v. 2.3.4 [49]. The analysis was run 10 times for each K value from K = 1 to K = 10. The burn-in length was set to 100,000 and the number of iterations to 100,000. The ad hoc statistics defined by Evanno et al. [50] as ΔK were used. Based on the posterior probability of membership (Q) of a given accession, the accession was classified as an admixture with a membership probability of $Q < 0.70$. The squared correlation coefficients (r^2) between ILP markers were used to quantify intragenic and intergenic LD, computed using TASSEL 2.1 [51]. The significance of this measure was assessed using p -values after Bonferroni correction for multiple tests ($\alpha = 0.05$).

4.5. Association Analysis of ILP Markers

To assess the effect of population structure (Q) and relative kinship (K) on traits in the Amazonian germplasm accessions, Q and K were inferred from an independent set of 1820 SNPs as previously described [1] to avoid dependency among terms in the model and to prevent the structure from absorbing the QTL effects from the model [52,53]. ILP markers were used to perform association analyses through general linear models (GLM) and the mixed linear model (MLM). The TASSEL V2.1 [51] software package was used to analyze the multiallelic genotypic data of ILP markers and the phenotypic data of girth traits using the GLM without correction, the GLM model corrected with the Q-matrices (GLM + Q) or PCA (GLM + PCA), and the MLM incorporated both K and Q matrices (MLM + K + Q) or PCA (MLM + K + PCA). All models were assessed for their ability to control for type I errors by plotting the Quantile-Quantile (Q-Q) p values for the markers, where uniformly distributed p values indicate proper control. The best-fit model was selected for association analysis. For multiple comparison adjustment, the marker-trait association was considered significant if the adjusted p -value following FDR (q value) was < 0.05 .

4.6. Validation of the Identified ILP Markers

To confirm the effectiveness of the ILPs detected, markers identified by association mapping were further validated using 95 rubber tree breeding lines with a range of growth girths. These breeding lines were generated from 23 parents, propagated by grafting with three replicates, and planted in seven-tree row plots at a spacing of 3×7 m. The experiment was conducted in the year 2000 at NKRR. Girth measurements were taken collected continuously in March (GiD) and September (GiW) between 2003 and 2006. GiC was calculated by subtraction from the circumferences of years 2003 and 2006. These evaluation periods were treated as independent environments. DNA isolation and PCR amplification were performed as described previously for ILP marker development. The amplified product was resolved using a ZAGTM DNA Analyser.

4.7. Library Preparation and SMRT Sequencing of the Candidate Gene

The candidate gene, *COBL*, that encodes glycosyl phosphatidylinositol (GPI)-anchored proteins regulating cellulose deposition and oriented cell expansion in the plant cell wall, was selected for sequencing in the Amazonian population. We applied targeted sequencing using Pacific Biosciences (PacBio) single-molecule real-time (SMRT) circular consensus sequencing (CCS) to assess sequence variation. Two-step barcoding for each individual was performed with M13-tailed primers. The target regions were first amplified with pairs of gene-specific primers with M13 forward and reverse sequence tails (Table S5). The *COBL*-specific primers were used to generate a 4236 kb fragment covering 6 exons, 5 introns, and 3' UTR (GenBank: AJJZ010406963.1). Thermal cycling conditions were as follows: 3 min initial denaturation at 95 °C, 35 cycles of 30 s for denaturation at 95 °C, 60 s of annealing at 52–60 °C (depending on the combination of primers), a 45 s extension at 72 °C, and a final 5 min extension at 72 °C. Sample-specific barcodes were introduced in a second PCR with identical conditions as the first, but using 3 μ L of the purified product from the first step PCR. A SMRT bell library and sequencing were performed as recommended by Pacific Biosciences.

4.8. Linkage Disequilibrium and Candidate Gene Association Mapping of COBL4

The raw reads generated from SMRT sequencing were processed using the Pacific Biosciences (CA, USA), SMRT analysis (v2.3.0) software package. SNPs within the candidate gene sequence were selected for the linkage disequilibrium analysis based on a minor allele frequency threshold of 5%. The extent of LD was estimated by nonlinear regression analysis on the basis of intragenic r^2 values [1]. When pairs of adjacent loci were found to demonstrate LD, one of the two SNP/Indels from each pair was used for marker–trait association analysis. Associations between the biallelic genotypic data of SNP/Indel markers and the mean girth traits were analyzed using the mixed linear model (a single locus model) incorporated both K, and Q matrices using TASSEL v. 5.2.54 [51], and by applying FarmCPU (Fixed and random model Circulating Probability Unification; a multilocus model) [54] in R. The relative distribution of $-\log_{10} p$ -values was observed for each SNP/Indel association and compared individually with the expected distribution using the Q-Q plot. The best-fit model was selected for association analysis. To control for type I errors due to multiple testing, the marker–trait association was considered significant if the adjusted p -value following FDR was <0.05 .

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/plants10071440/s1>, Figure S1: Population structure of 170 Amazonian accessions based on 115 ILP markers. (a) The most probable number of clusters (K), estimated using ad-hoc ΔK statistics. (b) Bayesian analysis of population structure based on $K = 2$. Each accession is represented by a thin horizontal bar partitioned into two colored segments whose length is proportional to the estimated membership coefficient (Q). (c) Principal component analysis (PCA). The triangles, squares, and circles represent accessions from Acre, Mato Grosso, and Rondonia, respectively.; Figure S2: Quantile-quantile plots of estimated $-\log_{10}(p)$ from association analysis using three models and three traits: girth in the dry (GiD) and wet (GiW) seasons and total girth increment (GiC). The black line is the expected line under the null distribution. The blue line represents the observed p values using the GLM model, the red line those using the GLM with Q or PCA model, and the green line those using the MLM model with Q or PCA and K.; Table S1: Characteristics of selected 115 ILP markers.; Table S2: Linkage disequilibrium of 59 pairs based on estimated $r^2 \geq 0.1$.; Table S3: Association analysis of 115 ILP markers.; Table S4: Primers used for the amplification of COBL gene.; Table S5. Primers used for the amplification of COBL gene.

Author Contributions: K.U. conceived and designed the research. G.B., and K.U. conducted the experiments and analyzed the data. R.R., and T.P. conducted the field experiments. W.P., and S.T. performed SMRT sequencing. K.U. drafted the manuscript. W.P., and S.T. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Rubber Authority of Thailand, and Thammasat University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant data can be found within the manuscript and Supplementary Materials.

Conflicts of Interest: The authors declare no competing interests.

References

1. Chanroj, V.; Rattanawong, R.; Phumichai, T.; Tangphasornruang, S.; Ukoskit, K. Genome-wide association mapping of latex yield and girth in Amazonian accessions of *Hevea brasiliensis* grown in a suboptimal climate zone. *Genomics* **2017**, *109*, 475–484. [CrossRef] [PubMed]
2. De Souza, L.M.; Dos Santos, L.H.B.; Rosa, J.R.B.F.; Da Silva, C.C.; Mantello, C.C.; Conson, A.R.O.; Scaloppi, E.J., Jr.; Fialho, J.D.F.; De Moraes, M.L.T.; Gonçalves, P.D.S.; et al. Linkage disequilibrium and population structure in wild and cultivated populations of rubber tree (*Hevea brasiliensis*). *Front. Plant Sci.* **2018**, *9*, 815. [CrossRef]
3. Ehrenreich, I.; Hanzawa, Y.; Chou, L.; Roe, J.; Kover, P.; Purugganan, M. Candidate Gene Association Mapping of Arabidopsis Flowering Time. *Genetics* **2009**, *183*, 325–335. [CrossRef] [PubMed]

4. Alomari, D.Z.; Eggert, K.; Von Wiren, N.; Alqudah, A.M.; Polley, A.; Plieske, J.; Ganal, M.W.; Pillen, K.; Röder, M.S. Identifying Candidate Genes for Enhancing Grain Zn Concentration in Wheat. *Front. Plant Sci.* **2018**, *9*, 1313. [[CrossRef](#)]
5. Sehgal, D.; Mondal, S.; Guzman, C.; Garcia Barrios, G.; Franco, C.; Singh, R.; Dreisigacker, S. Validation of Candidate Gene-Based Markers and Identification of Novel Loci for Thousand-Grain Weight in Spring Bread Wheat. *Front. Plant Sci.* **2019**, *10*, 1189. [[CrossRef](#)] [[PubMed](#)]
6. Carpenter, M.A.; Shaw, M.; Cooper, R.D.; Frew, T.J.; Butler, R.C.; Murray, S.R.; Moya, L.; Coyne, C.J.; Timmerman-Vaughan, G.M. Association mapping of starch chain length distribution and amylose content in pea (*Pisum sativum* L.) using carbohydrate metabolism candidate genes. *BMC Plant Biol.* **2017**, *17*, 132. [[CrossRef](#)]
7. Álvarez, M.F.; Angarita, M.; Delgado, M.C.; García, C.; Jiménez-Gomez, J.; Gebhardt, C.; Mosquera, T. Identification of Novel Associations of Candidate Genes with Resistance to Late Blight in *Solanum tuberosum* Group Phureja. *Front. Plant Sci.* **2017**, *8*, 1040. [[CrossRef](#)] [[PubMed](#)]
8. Yu, X.; Pijut, P.M.; Byrne, S.; Asp, T.; Bai, G.; Jiang, Y. Candidate gene association mapping for winter survival and spring regrowth in perennial ryegrass. *Plant Sci.* **2015**, *235*, 37–45. [[CrossRef](#)] [[PubMed](#)]
9. Palle, S.R.; Seeve, C.M.; Eckert, A.J.; Cumbie, W.P.; Goldfarb, B.; Loopstra, C.A. Natural variation in expression of genes involved in xylem development in loblolly pine (*Pinus taeda* L.). *Tree Genet. Genomes* **2011**, *7*, 193–206. [[CrossRef](#)]
10. Freeman, J.S.; Potts, B.M.; Downes, G.M.; Pilbeam, D.; Thavamanikumar, S.; Vaillancourt, R.E. Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. *New Phytol.* **2013**, *198*, 1121–1134. [[CrossRef](#)]
11. Harakava, R. Genes encoding enzymes of the lignin biosynthesis pathway in *Eucalyptus*. *Genet. Mol. Biol.* **2005**, *28* (Suppl. S3), 601–607. [[CrossRef](#)]
12. Kirst, M.; Myburg, A.A.; De Leon, J.P.; Kirst, M.E.; Scott, J.; Sederoff, R. Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. *Plant Physiol.* **2004**, *135*, 2368–2378. [[CrossRef](#)]
13. Southerton, S.G.; MacMillan, C.P.; Bell, J.C.; Bhuiyan, N.; Dowries, G.; Ravenwood, I.C.; Joyce, K.R.; Williams, D.; Thumma, B.R. Association of allelic variation in xylem genes with wood properties in *Eucalyptus nitens*. *Aust. For.* **2010**, *73*, 259–264. [[CrossRef](#)]
14. Liu, Y.; He, Z.; Appels, R.; Xia, X. Functional markers in wheat: Current status and future prospects. *Theor. Appl. Genet.* **2012**, *125*, 10–14. [[CrossRef](#)]
15. Badoni, S.; Das, S.; Sayal, Y.K.; Gopalakrishnan, S.; Singh, A.K.; Rao, A.R.; Agarwal, P.; Parida, S.K.; Tyagi, A.K. Genome-wide generation and use of informative intron-spanning and intron-length polymorphism markers for high-throughput genetic analysis in rice. *Sci. Rep.* **2016**, *6*, 23765. [[CrossRef](#)] [[PubMed](#)]
16. Cai, C.; Wu, S.; Niu, E.; Cheng, C.; Guo, W. Identification of genes related to salt stress tolerance using intron-length polymorphic markers, association mapping and virus-induced gene silencing in cotton. *Sci. Rep.* **2017**, *7*, 528. [[CrossRef](#)] [[PubMed](#)]
17. Gupta, S.K.; Bansal, R.; Gopalakrishna, T. Development of intron length polymorphism markers in cowpea [*Vigna unguiculata* (L.) Walp.] and their transferability to other *Vigna* species. *Mol. Breed.* **2012**, *30*, 1363–1370. [[CrossRef](#)]
18. Jayaswall, K.; Sharma, H.; Bhandawat, A.; Sagar, R.; Yadav, V.K.; Sharma, V.; Mahajan, V.; Roy, J.; Singh, M. Development of intron length polymorphic (ILP) markers in onion (*Allium cepa* L.), and their cross-species transferability in garlic (*A. sativum* L.) and wild relatives. *Genet. Resour. Crop Evol.* **2019**, *66*, 1379–1388. [[CrossRef](#)]
19. Li, D.; Xia, Z.; Deng, Z.; Liu, X.; Dong, J.; Feng, F. Development and characterization of intron-flanking EST-PCR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). *Mol. Biotechnol.* **2012**, *51*, 148–159. [[CrossRef](#)] [[PubMed](#)]
20. Saisug, W.; Ukoskit, K. Comparative analysis of EST-derived markers for allelic variation in *Jatropha curcas* L. and cross transferability among economically important species of Euphorbiaceae. *Genes Genom.* **2013**, *35*, 1–12. [[CrossRef](#)]
21. Yan, W.; Zhao, H.; Yu, K.; Wang, T.; Khattak, A.N.; Tian, E. Development of a multiparent advanced generation intercross (MAGIC) population for genetic exploitation of complex traits in *Brassica juncea*: Glucosinolate content as an example. *Plant Breed.* **2020**, *139*, 779–789. [[CrossRef](#)]
22. Parra, G.; Bradnam, K.; Rose, A.B.; Korf, I. Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. *Nucleic Acids Res.* **2011**, *39*, 5328–5337. [[CrossRef](#)]
23. Pootakham, W.; Sonthirod, C.; Naktang, C.; Ruang-Areerate, P.; Yoocha, T.; Sangsrakru, D.; Theerawattanask, K.; Rattanawong, R.; Lekawipat, N.; Tangphatsornruang, S. De novo hybrid assembly of the rubber tree genome reveals evidence of paleotetraploidy in *Hevea* species. *Sci. Rep.* **2017**, *7*, 41457. [[CrossRef](#)]
24. Shearman, J.; Pootakham, W.; Tangphatsornruang, S. The BPM 24 Rubber Tree Genome, Organellar Genomes and Synteny within the Family Euphorbiaceae. In *Compendium of Plant Genomes: The Rubber Genome*; Springer International Publishing: New York, NY, USA, 2020; pp. 67–79.
25. Feng, S.P.; Li, W.G.; Huang, H.S.; Wang, J.Y.; Wu, Y.T. Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). *Mol. Breed.* **2009**, *23*, 85–97. [[CrossRef](#)]
26. Li, D.; Xia, Z.; Deng, Z.; Liu, X.; Feng, F. Development, characterization, genetic diversity and cross-species/genera transferability of ILP markers in rubber tree (*Hevea brasiliensis*). *Genes Genom.* **2013**, *35*, 719–731. [[CrossRef](#)]
27. De Souza, L.M.; Guen, V.L.; Cerqueira-Silva, C.B.M.; Silva, C.C.; Mantello, C.C.; Conson, A.R.O.; Vianna, J.P.G.; Zucchi, M.I.; Scaloppi Junior, E.J.; Fialho, J.D.F.; et al. Genetic diversity strategy for the management and use of rubber genetic resources: More than 1000 wild and cultivated accessions in a 100-genotype core collection. *PLoS ONE* **2015**, *10*, e0134607. [[CrossRef](#)] [[PubMed](#)]

28. Le Guen, V.; Doare, F.; Weber, C.; Seguin, M. Genetic structure of Amazonian populations of *Hevea brasiliensis* is shaped by hydrographical network and isolation by distance. *Tree Genet. Genomes* **2009**, *5*, 673–683. [[CrossRef](#)]
29. Pootakham, W.; Shearman, J.; Tangphatsornruang, S. Development of Molecular Markers in *Hevea brasiliensis* for Marker-assisted Breeding. In *Compendium of Plant Genomes: The Rubber Genome*; Springer International Publishing: New York, NY, USA, 2020; pp. 55–66.
30. Brady, S.M.; Song, S.; Dhugga, K.S.; Rafalski, J.A.; Benfey, P.N. Combining expression and comparative evolutionary analysis. *Plant Physiol.* **2007**, *143*, 172–187. [[CrossRef](#)]
31. Thumma, B.R.; Matheson, B.A.; Zhang, D.; Meeske, C.; Meder, R.; Downes, G.M.; Southerton, S.G. Identification of a Cis-acting regulatory polymorphism in a Eucalypt COBRA-like gene affecting cellulose content. *Genetics* **2009**, *183*, 1153–1164. [[CrossRef](#)]
32. Chamary, J.V.; Hurst, L.D. The price of silent mutations. *Sci. Am.* **2009**, *300*, 46–53. [[CrossRef](#)] [[PubMed](#)]
33. Hunt, R.; Sauna, Z.E.; Ambudkar, S.V.; Gottesman, M.M.; Kimchi-Sarfaty, C. Silent (synonymous) SNPs: Should we care about them? *Methods Mol. Biol.* **2009**, *578*, 23–39. [[PubMed](#)]
34. Ingvarsson, P.K.; Garcia, M.V.; Luquez, V.; Hall, D.; Jansson, S. Nucleotide polymorphism and phenotypic associations within and around the *phytochrome B2* Locus in European aspen (*Populus tremula*, Salicaceae). *Genetics* **2008**, *178*, 2217–2226. [[CrossRef](#)]
35. Larsen, P.A.; Heilman, A.M.; Yoder, A.D. The utility of PacBio circular consensus sequencing for characterizing complex gene families in non-model organisms. *BMC Genom.* **2014**, *15*, 720. [[CrossRef](#)]
36. Travers, K.J.; Chin, C.-S.; Rank, D.R.; Eid, J.S.; Turner, S.W. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res.* **2010**, *38*, e159. [[CrossRef](#)] [[PubMed](#)]
37. Wenger, A.M.; Peluso, P.; Rowell, W.J.; Chang, P.-C.; Hall, R.J.; Concepcion, G.T.; Ebler, J.; Functamman, A.; Kolesnikov, A.; Olson, N.D.; et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **2019**, *37*, 1155–1162. [[CrossRef](#)]
38. Shearman, J.R.; Tangphatsornruang, S. Nanopore sequencing in agricultural and food applications. In *Handbook of Nanotechnology Applications*; Elsevier: Oxford, UK, 2021; pp. 443–459.
39. Yu, J.; Pressoir, G.; Briggs, W.H.; Vroh, B.I.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B.; et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38*, 203–208. [[CrossRef](#)]
40. Kaler, A.S.; Gillman, J.D.; Beissinger, T.; Purcell, L.C. Comparing different statistical models and multiple testing corrections for association mapping in Soybean and Maize. *Front. Plant Sci.* **2020**, *10*, 1794. [[CrossRef](#)]
41. Zhao, W.; Pollack, J.L.; Blagev, D.P.; Zaitlen, N.; McManus, M.T.; Erle, D.J. Massively parallel functional annotation of 3' untranslated regions. *Nat. Biotechnol.* **2014**, *32*, 387–391. [[CrossRef](#)] [[PubMed](#)]
42. Vanholme, R.; Demedts, B.; Morreel, K.; Ralph, J.; Boerjan, W. Lignin biosynthesis and structure. *Plant Physiol.* **2010**, *153*, 895–905. [[CrossRef](#)] [[PubMed](#)]
43. Yoon, J.; Choi, H.; An, G. Roles of lignin biosynthesis and regulatory genes in plant development. *J. Integr. Plant Biol.* **2015**, *57*, 902–912. [[CrossRef](#)]
44. Rahman, A.Y.A.; Usharraj, A.O.; Misra, B.B.; Thottathil, G.P.; Jayasekaran, K.; Feng, Y.; Hou, S.; Ong, S.Y.; Ng, F.L.; Lee, L.S.; et al. Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genom.* **2013**, *14*, 75. [[CrossRef](#)] [[PubMed](#)]
45. Rozen, S.; Skaletsky, H.J. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **2000**, *132*, 365–386. [[PubMed](#)]
46. Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331. [[PubMed](#)]
47. Liu, K.; Muse, S.V. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* **2005**, *21*, 2128–2129. [[CrossRef](#)]
48. Rohlf, F.J. *NTSYSpc Numerical Taxonomy and Multivariate Analysis System Version 2.0 User Guide*, 2nd ed.; Applied Biostatistics Inc.: New York, NY, USA, 1998.
49. Pritchard, J.K.; Stephens, M.; Rosenberg, N.A.; Donnelly, P. Association mapping in structured population. *Am. J. Hum. Genet.* **2000**, *67*, 170–181. [[CrossRef](#)]
50. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [[CrossRef](#)] [[PubMed](#)]
51. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)]
52. Bernardo, R. Genomewide markers as cofactors for precision mapping of quantitative trait loci. *Theor. Appl. Genet.* **2013**, *126*, 999–1009. [[CrossRef](#)] [[PubMed](#)]
53. Patterson, N.; Price, A.L.; Reich, D. Population Structure and Eigenanalysis. *PLoS Genet.* **2006**, *2*, e190. [[CrossRef](#)]
54. Liu, X.; Huang, M.; Fan, B.; Buckler, E.S.; Zhang, Z. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLoS Genet.* **2016**, *12*, e1005767. [[CrossRef](#)] [[PubMed](#)]