



Article

Genome Assembly and Analysis of the Flavonoid and Phenylpropanoid Biosynthetic Pathways in Fingerroot Ginger (*Boesenbergia rotunda*)

Sima Taheri ¹, Chee How Teo ¹ , John S. Heslop-Harrison ^{2,3,*} , Trude Schwarzacher ^{2,3}, Yew Seong Tan ⁴, Wei Yee Wee ⁵, Norzulaani Khalid ⁶, Manosh Kumar Biswas ², Naresh V. R. Mutha ⁷ , Yusmin Mohd-Yusuf ^{1,8}, Han Ming Gan ⁹ and Jennifer Ann Harikrishna ^{1,4,*}

- ¹ Centre for Research in Biotechnology for Agriculture, University of Malaya, Kuala Lumpur 50603, Malaysia; sima.taheri100@gmail.com (S.T.); cheehow.teo@um.edu.my (C.H.T.); yusmin_y@um.edu.my (Y.M.-Y.)
- ² Department of Genetics and Genome Biology, University of Leicester, Leicester LE1 7RH, UK; ts32@leicester.ac.uk (T.S.); mkb35@leicester.ac.uk (M.K.B.)
- ³ Key Laboratory of Plant Resources Conservation and Sustainable Utilization/Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
- ⁴ Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur 50603, Malaysia; tyewseong@yahoo.com
- ⁵ School of Science, Monash University Malaysia, Subang Jaya 47500, Malaysia; wee.weiyee@monash.edu
- ⁶ Department of Biology, International University of Malaya-Wales, Kuala Lumpur 50603, Malaysia; norzulaani@iumw.edu.my
- ⁷ Division of Infectious Diseases, Vanderbilt University Medical Center, Nashville, TN 37203, USA; nareshmvr@gmail.com
- ⁸ Biology Division, Centre for Foundation Studies in Science, University of Malaya, Kuala Lumpur 50603, Malaysia
- ⁹ Department of Biological Sciences, Sunway University, Bandar Sunway, Petaling Jaya 47500, Malaysia; hxg2760@gmail.com
- * Correspondence: phh4@leicester.ac.uk (J.S.H.-H.); jennihari@um.edu.my (J.A.H.)



Citation: Taheri, S.; Teo, C.H.; Heslop-Harrison, J.S.; Schwarzacher, T.; Tan, Y.S.; Wee, W.Y.; Khalid, N.; Biswas, M.K.; Mutha, N.V.R.; Mohd-Yusuf, Y.; et al. Genome Assembly and Analysis of the Flavonoid and Phenylpropanoid Biosynthetic Pathways in Fingerroot Ginger (*Boesenbergia rotunda*). *Int. J. Mol. Sci.* **2022**, *23*, 7269. <https://doi.org/10.3390/ijms23137269>

Academic Editor: Jae Youl Cho

Received: 2 June 2022

Accepted: 27 June 2022

Published: 30 June 2022

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



Copyright: © 2022 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/>).

Abstract: *Boesenbergia rotunda* (Zingiberaceae), is a high-value culinary and ethno-medicinal plant of Southeast Asia. The rhizomes of this herb have a high flavanone and chalcone content. Here we report the genome analysis of *B. rotunda* together with a complete genome sequence as a hybrid assembly. *B. rotunda* has an estimated genome size of 2.4 Gb which is assembled as 27,491 contigs with an N50 size of 12.386 Mb. The highly heterozygous genome encodes 71,072 protein-coding genes and has a 72% repeat content, with class I TEs occupying ~67% of the assembled genome. Fluorescence in situ hybridization of the 18 chromosome pairs at the metaphase showed six sites of 45S rDNA and two sites of 5S rDNA. An SSR analysis identified 238,441 gSSRs and 4604 EST-SSRs with 49 SSR markers common among related species. Genome-wide methylation percentages ranged from 73% CpG, 36% CHG and 34% CHH in the leaf to 53% CpG, 18% CHG and 25% CHH in the embryogenic callus. Panduratin A biosynthetic unigenes were most highly expressed in the watery callus. *B. rotunda* has a relatively large genome with a high heterozygosity and TE content. This assembly and data (PRJNA71294) comprise a source for further research on the functional genomics of *B. rotunda*, the evolution of the ginger plant family and the potential genetic selection or improvement of gingers.

Keywords: *Boesenbergia rotunda*; DNA methylation; genome assembly; ginger; panduratin A; SSR; TE

1. Introduction

Boesenbergia rotunda (L.) Mansf. (*syn.* *B. pandurata* (Roxb.) Schltr.) (ITIS Taxonomic Serial No.: 506504), commonly known as Fingerroot ginger and as a type of galanga or galangal, is a member of the family Zingiberaceae in the order Zingiberales. The common name for the plant, Fingerroot, is due to its finger-shaped rhizomes (Figure 1B). With

50 genera and 1600 species, the Zingiberaceae is the largest family in the order, along with other families of ginger (Zingiberaceae, Costaceae, Marantaceae, and Cannaceae) and banana (Musaceae, Strelitziaceae, Lowiaceae, and Heliconiaceae), that include many economically important plant species [1,2]. The Zingiberaceae family consists of herbaceous perennial plants distributed over tropical and subtropical regions with the highest diversity in Southeast Asia (especially in Indonesia, Malaysia and Thailand), India and Southern China [3–6]. The leaves, flowers and in particular the rhizomes of many of the Zingiberaceae family members are used as flavouring agents and for herbal medicine [4,7].

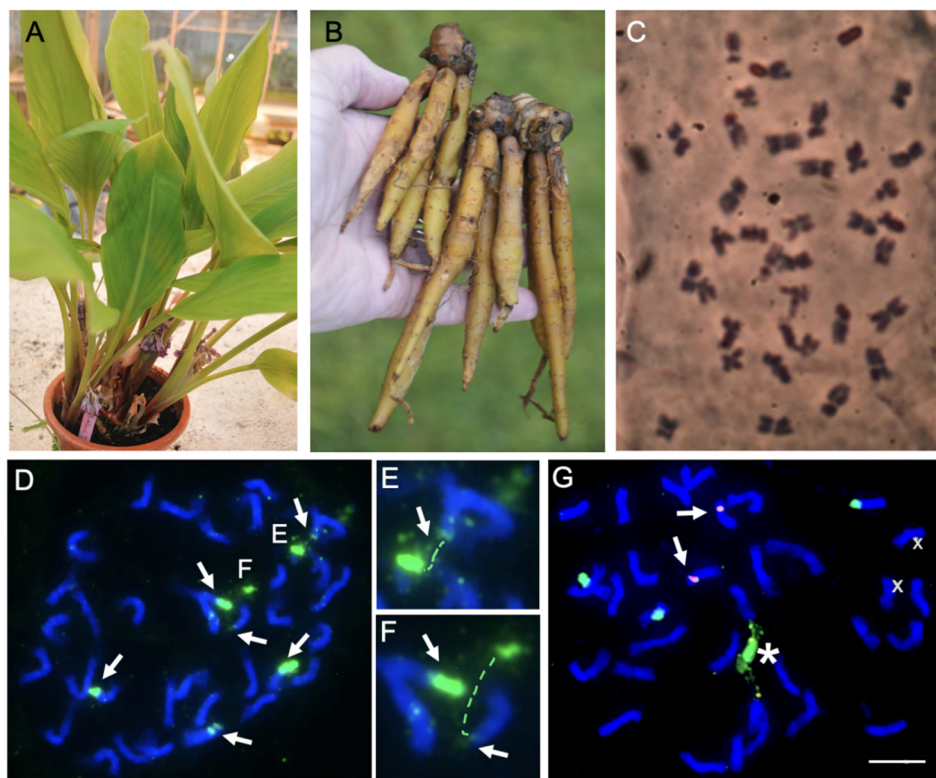


Figure 1. *Boesenbergia rotunda* leaves, rhizomes and number and location of 45S and 5S rDNA sites on metaphase chromosomes ($2n = 36$). (A) Leaves. (B) Rhizomes. (C) Chromosome preparation using fresh root tips. (D–G) Fluorescent in situ hybridization with clone pTa71 (45S rDNA of wheat), labelled with digoxigenin and detected with FITC (green) and clone pTa794 (5S rDNA of wheat) labelled with biotin and detected with Alexa 674 (shown in red). (D–F) early metaphase showing six sites of 45S rDNA (arrows) of variable strength at ends of 3 pairs of chromosomes. In some cases, the rDNA is extended, and the satellite is separated from the main chromosomes shown enlarged in (E,F). (G) Two 5S rDNA sites (arrows) were detected on a chromosome pair not bearing 45S rDNA. The star indicates the fusion of 2 or 3 45S rDNA sites.

Boesenbergia is a genus that has around 80 different species and can be found all the way from India to Southeast Asia [3,8–10]. *B. rotunda* is a perennial herb propagated via rhizomes and widely cultivated commercially for its rhizomes and shoots to flavour food and for ethno-medicinal use [11,12]. Antioxidant activity is among the most important bioactivities of plant flavonoid compounds [13], and research on the secondary metabolites of *B. rotunda* has focused on the medicinal properties of the rhizome extracts, in particular the flavanones and chalcones including panduratin A, pinocembrin, pinostrobin, alpinetin, boesenbergin, cardamonin, naringenin, quercetin, and kaempferol [8,14–20]. Of these, the flavonoid compounds, panduratin A/DI, 4-hydroxypanduratin, and cardamonin show the clearest biological and pharmacological effects such as being anti-inflammatory [21,22], anti-tumour activity against human breast and lung cancers [23–27], and antimicrobial activity

against HIV protease [28], Dengue-2 (DEN-2) virus NS3 protease [29,30], SARS-CoV-2 in human airway epithelial cells [31], the oral bacteria *Streptococcus mutans* [32,33], *Helicobacter pylori* [34], and against the spoilage bacteria *Lactobacillus plantarum* (Lactiplantibacillus) [35]. A recent patent claimed that panduratin derivatives from *B. rotunda* have the potential for preventing, ameliorating, or treating bone loss disease [36], while 4-hydroxypanduratin was reported to have the most potent vasorelaxant activity among the major flavonoids of the *B. rotunda* extracts [37].

The ethno-medicinal and potential pharmaceutical importance of *B. rotunda* have led to interest in exploring cell and tissue cultures for secondary metabolite production. In commercial farms, the plant is propagated clonally from rhizomes, and several protocols for multiplication via in vitro culture have been reported including plantlet regeneration via somatic embryogenesis from callus cultures [38,39], shoot bud explants [40] and embryogenic cell suspension cultures [38]. Cell suspensions of *B. rotunda* [18,41] and various types of callus [42,43] have been explored as potential sources for alpinetin, cardamonin, pinocembrin, pinostrobin and panduratin A. Reproducible methods for the in vitro cell culture of *B. rotunda*, have led to protocols for genetic transformation [39], that could facilitate the metabolic engineering of cell materials for specific desirable metabolite production; however, current knowledge of the underlying biosynthetic pathways is sparse. Other than biochemical profiling [17,43], the application of current technologies for determining the deep sets of the genetic sequences expressed in various tissue and cell types can deliver useful information.

Genomic level studies improve the understanding of the biology and biochemistry of the plant and can be applied in breeding for improved agronomy and plant products. Whole genome sequencing identifies the genes and regulatory sequences for complex biological processes such as secondary metabolite biosynthesis [44–46], while transcriptional profiling provides information for functional studies. Structural genomic studies have been undertaken in other Zingiberales including turmeric (*Curcuma longa*; genome size of 1.24 Gb) [47] and for several Musaceae species and cultivars, which have genome sizes ranging from 462 Mb to 598 Mb (Banana Genome Hub <https://banana-genome-hub.southgreen.fr/> accessed on 28 May 2022) [48], while the Pan-genome of *Musa Ensete* has a genome size of 951.6 Mb [49]. Larger plant genomes have now been sequenced including those of important monocot species such as wheat ~17 Gb (International Wheat Genome Sequencing Consortium), *Aegilops tauschii* ~4.3 Gb [50], oil palm ~1.8 Gb [51] and maize ~2.6 Gb [52], in addition to species known for their unique metabolites such as tea (*Camellia sinensis*) ~2.98 Gb [53,54] and ginseng (*Panax ginseng*) ~ 3.2 Gb [55]; however, even with the recent advances in long sequence technology, large plant genomes can be challenging to assemble due to a high repeat content and high levels of heterozygosity [56,57].

The availability of an assembled genome sequence expands the functional biological questions that can be asked, since regulatory and variable elements, many of which may be involved in epigenetic regulation, cannot be seen purely using expression data. Consequently, while transcriptome [41] and proteome [58] data for *B. rotunda* are available, the lack of a previously published genome assembly is a limitation for functional studies. Genome assemblies also facilitate the exploration of genomic repeats which can not only be a source for genetic markers but are also drivers of genome size, gene content and order, centromere function and reflect genome evolution [59,60]. Last but not least, the epigenetic dynamism in genomes mainly involves “non-coding DNA”, thus, a genome assembly provides the framework for epigenetic studies; therefore, in the current investigation we performed the first complete genome sequence for *B. rotunda* made with a hybrid assembly strategy using the Pacific Biosciences (PacBio) and Illumina HiSeq platforms. We explored the sites of 45S rDNA and 5S rDNA on metaphase chromosomes observed by Fluorescence in situ hybridization (FISH). In addition, we carried out a deep transcriptome (RNA-seq data) assembly from five *B. rotunda* samples, including various types of callus cultures, and leaves. Gene expression profiles and bisulfite seq DNA methylation data from these tissues and samples were used for a co-expression analysis to identify any association of gene

expression and local DNA methylation of unigenes related to methylation, somatic embryogenesis, and the pathways for flavonoid and phenylpropanoid biosynthesis. We also report novel expressed sequence tags-SSR (EST-SSR) and genomic SSR markers for *B. rotunda* and the estimated cross-transferability of the designed primers between *B. rotunda* and closely related species to provide deeper genetic resources to support further study of the biology and biodiversity in this genus. Genomic information and complete sequence data for this less investigated herb should provide a solid foundation as a vital step in genetic analysis to facilitate *B. rotunda* improvement and to reach a deeper understanding of the metabolic pathways of its natural products.

2. Results

2.1. Chromosomes and Location of rDNA Sites

Boesenbergia rotunda ($2n = 36$; 18 pairs of submetacentric chromosomes) has three pairs of 45S rDNA sites near the ends of three pairs of chromosomes (Figure 1A). One pair of the 5S rDNA sites (Figure 1D) is on a chromosome pair not bearing 45S rDNA.

2.2. Genome Assembly

Genomic DNA from the leaves of a single, clonal *B. rotunda* plant was sequenced using multiple approaches (Table S1), with 114 Gb PacBio long reads, 260 Gb of Illumina HiSeq 2500 250 bp paired-end reads, and 90 Gb of mate-paired reads with 2, 5, 10, 20 and 40 kb insert sizes. Based on the k-mer analysis ($k = 17$, GenomeScope), the estimated haploid genome size of *B. rotunda* was 2.4 Gb (Figure S1), consistent with the flow cytometry (Figure S2). The heterozygosity was estimated as 3.01%. A hybrid genome assembly pipeline combining the Illumina data and PacBio data was adopted (Figure S3). The final assembled genome size was 2.347 Gb, characterized by 27,491 contigs and 10,627 scaffolds, with a contig N50 of 123.86 kb and a scaffold N50 of 394.68 kb (Table 1). Based on a benchmarking universal single-copy orthologs (BUSCO) analysis [61] mapping the *B. rotunda* genome against a set of 1440 core eukaryotic genes, 1232 (85.6%) were present (Table S2). An assembly quality assessment showed that over 95% of the Illumina PE250 reads were mapped to the contig assembly (Table 2).

Table 1. Statistics of the final genome assembly of the *B. rotunda*.

	No.	Scaffolds		Contigs		
		Size (bp)		No.	Size (bp)	Gaps
		With gaps	Without Gaps			
Total Number	10,627			27,491		16,864
Min	-	5830	5830	-	5198	25
Median	-	136,187	131,005	-	55,047	2415
Mean	-	220,901	213,350	-	82,473	4758
Max	-	2,848,924	2,758,809	-	1,033,476	38,914
Total size	-	2,347,517,452	2,267,274,222	-	2,267,274,222	80,243,230
N50	-	394,682	379,106	-	123,867	11,038
N90	-	107,821	103,307	-	37,045	2551
N95	-	69,101	66,089	-	27,170	1540
GC content (%)				40.1		

Table 2. Evaluation of completeness of the final assembly.

Species	Read Length (bp)	Data	Sequence Depth (X)	Mapped (%)	Properly Paired (%)	Singletons (%)	Reference Total Length (Gb)	Reads Covered Length (Gb)	Coverage (%)
<i>B. rotunda</i>	250_250	260 (Gb)	104	95.24	84.47	0.20	2.35	2.25	96

2.3. Annotation of the *B. rotunda* Genome

Five sets of RNA-seq datasets were generated from three cell culture types, in vitro and ex vitro leaves of *B. rotunda*, given the importance for secondary metabolites production. Individual transcriptomes were assembled from these RNA-seq reads using different de novo transcriptome assemblers (Table 3, Figure S4). The assembled transcriptome size ranged from 31 to 71 million base pairs with 72,085 to 158,465 contigs for the Oases, SOAPdenovo-Trans, TransAbyss, and Trinity (Table 3, Figure S4). Oases had the largest N50 size and average contig length. The BUSCO quantitative measure of the completeness transcriptomes in terms of expected gene content scores, also showed Oases (36.7%) and TransAbyss (36.6%) to give the assemblies with higher numbers of complete and single-copy contigs compared to the SOAPdenovo-Trans and Trinity (31.7%) (Figure S5). The non-redundant transcript sequences formed from the Oases followed by TGICL were used to annotate the *B. rotunda* genome and for the downstream expression analysis.

Table 3. Comparison of *denovo* transcriptome assembly results for four different assembly software: SOAP-*denovo*, Oases, TransAbyss, and Trinity.

Features	SOAP- <i>denovo</i> (K25)	Oases (K21)	TransAbyss (K25)	Trinity (K25)
N50 size (bp)	410	1019	495	487
N50 No.	22,910	14,286	28,234	36,730
Contig number	78,492	72,085	111,327	158,465
Transcript's size (bp)	30,869,274	51,258,323	50,358,442	70,949,809
Avg. transcript length (bp)	393	711	452	448
Min. contig length(bp)	200	200	200	200
Max. contig length (bp)	15,760	12,523	33,886	13,325
Assessment assembly after merged assembly of non-redundant contigs from different k-mers via TGICL				
N50 size (bp)	572	1013	607	536
N50 no.	18,528	17,329	28,419	26,034
Contig number	78,963	95,847	132,572	115,096
Transcriptome size (bp)	38,503,434	66,535,881	67,286,353	54,131,258
Average length (bp)	488	694	508	470
Min. contig length(bp)	200	200	200	200
Max. contig length (bp)	43,900	88,053	100,968	19,761

Based on a combination of de novo and homology-based gene prediction methods, 72.51% of the genome (1.70 Gb) was annotated as repeats including 6.94% tandem repeats. Among the Class I TEs (Retroelements), long terminal repeats (LTRs) constituted the greatest proportion of the genome (67.16%) while DNA TE made up 3.29 % of the genome (Figure S6, Table 4).

From the 10,627 assembled contigs and 95,847 assembled transcriptome sequences searched for SSRs, (Table 5, Figure 2), the density of the microsatellites was 102 SSR loci per Mbp in the genomic and 69 SSR loci per Mbp in the transcriptome sequences. Among the identified repeat motif types, trinucleotides were the most abundant in both genomic (35.62%) and transcriptome (51.67%) sequences, followed by mono- and dinucleotide repeats (Table 5, Figure 2a). The class II type SSR-loci (<30 bp) were two-fold higher than the class I type in the genomic sequences, whereas the class II type SSR-loci were four-fold higher than the class I type SSR loci in the transcriptome sequences (Figure 2b). The number of AT rich microsatellites was significantly higher than that of the GC rich and microsatellites with a balanced GC content.

Table 4. TEs Content in the assembled *B. rotunda* genome.

Type		Repeat Size (bp)	% of Genome
TRF		162,927,183	6.94
RepeatMasker (RepBase TEs)	DNA	23,107,771	0.98
	LINE	7,269,664	0.31
	LTR	308,993,979	13.16
	SINE	50,226	0.00
	Other	1305	0.00
	Unknown	0.00	0.00
	Total	339,001,341	14.44
RepeatProteinMask (TE proteins)	DNA	30,071,545	1.28
	LINE	15,711,684	0.67
	LTR	449,069,450	19.13
	SINE	0.00	0
	Other	0.00	0
	Unknown	0.00	0
	Total	494,297,946	21.05
De novo	DNA	49,253,612	2.10
	LINE	9,765,795	0.42
	LTR	1,524,782,230	64.95
	SINE	789,305	0.03
	Other	0	0.00
	Satellite	8,247,215	0.351316
	Simple_repeat	6,742,687	0.287226
	Unknown	2,202,787	0.09
	Total	1,591,591,610	67.80
Combined TEs	DNA	77,273,965	3.29
	LINE	23,221,220	0.99
	LTR	1,576,612,191	67.16
	SINE	832,585	0.04
	Other	1305	0.00
	Unknown	2,202,787	0.09
	Total	1,653,717,174	70.45
Total		1,702,210,889	72.51

Note: Repbase TEs: the result of RepeatMasker based on Repbase; TE proteins: the result of RepeatProteinMask based on Repbase; De novo: result of RepeatMasker by using library predicted through De novo; Combined: combine the results of Repbase TEs, TE proteins and De novo.

The mapping of *B. rotunda* SSR to close relatives using newly designed primer sequences showed that from 93.81% of the genomic SSR and 73.12% of the transcriptome sequences suitable for the SSR primer design, only a low number of primers mapped to the selected relatives, *Musa acuminata*, *Musa balbisiana*, *Musa itinerans* and *Ensete ventricosum* (Table 5). Overall, 224 G-SSR and 65 EST-SSR primers showed transferability into any of the four related species (with slightly more in the Ensete), while only 42 genomic SSRs and 7 transcript SSRs were common to all five genomes (Figure 2c,d). A subset of 14 *B. rotunda* SSR primer pairs (Table S3) were tested for their marker potentiality and all showed amplified bands of the expected sizes for each species (Figure S7).

The annotation of the predicted protein-coding genes was a combination of homology-based and de novo prediction in addition to comparison with the *B. rotunda* transcriptome data (Table S4). After consolidation, 73,102 protein-coding genes were predicted in the *B. rotunda* genome with an average transcript length of 4312 bp (excluding UTR), CDS length of 1360 bp, average exon and intron lengths of 303 bp and 812 bp, respectively, and 4.49 exons per gene (Table S4). For the homology-based protein-coding gene predictions, the protein sequences from four species (*M. acuminata*, *Phoenix dactylifera*, *Oryza sativa* and *Arabidopsis thaliana*) were mapped onto the *B. rotunda* genome. From these alignments, *B. rotunda* had the highest number of matches with *P. dactylifera* followed by *O. sativa*,

A. thaliana and *M. acuminata* (Figure S8). Functional annotation of the 73,102 predicted proteins from *B. rotunda* against seven databases enabled functional predictions for 97.8% of the predicted genes (Table 6). Non-coding RNA analysis of the assembly identified 213 microRNA (miRNA), 2727 transfer RNA (tRNA), 486 ribosomal RNA (rRNA), and 2136 small nuclear RNA (snRNA) genes (Table 7).

Table 5. Genome and transcriptome-wide microsatellite identification and characterization in *B. rotunda*.

Item	Genome-Wide	%	Transcriptome-Wide	%
Total number of sequences examined	10,627		95,847	
Total size of examined sequences (bp)	2,347,517,452		66,535,881	
Total number of identified microsatellites	238,441		4579	
Number of microsatellites containing sequences	10,381		4032	
Sequences contain more than one microsatellite	9803		384	
Microsatellites in compound formation	4309		27	
Microsatellite's density (per Mbp)	102		69	
Class I microsatellites	82,414	35.20	949	20.85
Class II microsatellites	151,718	64.80	3603	79.15
AT rich microsatellites	176,052	75.19	2778	61.03
GC rich microsatellites	43,155	18.43	1275	28.01
AT/GC balance microsatellites	14,925	6.37	499	10.96
Mono-nucleotide repeats	68,961	28.92	1137	24.83
Di-nucleotide repeats	61,439	25.77	574	12.54
Tri-nucleotide repeats	84,932	35.62	2366	51.67
Tera-nucleotide repeats	5330	2.24	148	3.23
Penta-nucleotide repeats	9917	4.16	185	4.04
Hexa-nucleotide repeats	7862	3.30	169	3.69
Primer modelling was successful	223,678	93.81	3348	73.12
Primer modelling failed	14,763	6.60	1231	36.77
Non redundant primer	132,792	59.37	1888	56.39
No. of Primer Mapped on <i>M. acuminata</i> genome	100	0.075	30	1.59
No. of Primer Mapped on <i>M. balbisiana</i> genome	105	0.079	25	1.32
No. of Primer Mapped on <i>M. itinerans</i> genome	102	0.077	32	1.69
No. of Primer Mapped on <i>Ensete ventricosum</i> genome	121	0.091	27	1.43
No. of primer tested	6	100	8	100
No. of primer amplified	6	100	8	100

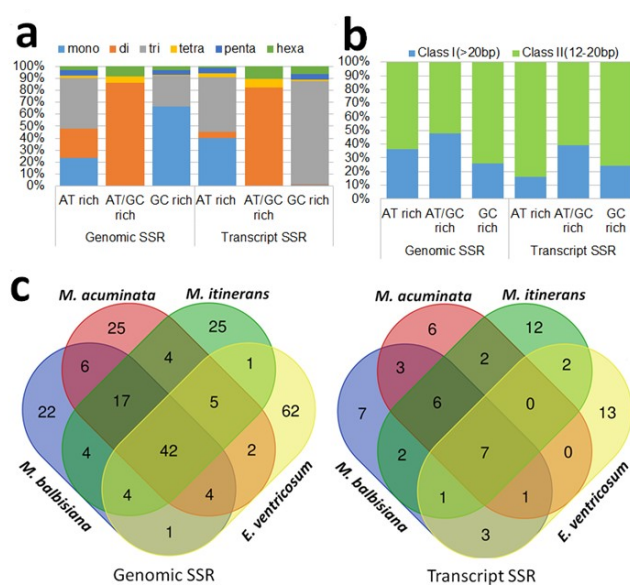


Figure 2. (a,b) frequency distribution of SSR motif; (c) transferability of genomic and transcript SSR markers in four relatives of *B. rotunda*.

Table 6. Statistics of function annotation.

Database	Number	Percent (%)
NR	71,072	97.22
InterPro	69,525	95.11
GO	45,256	61.91
KEGG	59,649	81.60
Swissprot	57,622	78.82
COG	24,851	33.99
TrEMBL	70,990	97.11
Total annotated	73,102	97.81
Unannotated	1602	2.19

Table 7. Non-coding RNA genes in the genome of *B. rotunda*.

Type	Copy	Average Length (bp)	Total Length (bp)	% of Genome
miRNA	213	119	25,384	0.001081
tRNA	2727	75	205,538	0.008756
rRNA	486	232	112,876	0.004808
18S	105	666	69,922	0.002979
28S	147	119	17,441	0.000743
5.8S	40	148	5931	0.000253
5S	194	101	19,582	0.000834
snRNA	2136	154	329,909	0.014054
CD-box	600	105	62,771	0.002674
HACA-box	53	134	7091	0.000302
splicing	1483	175	260,047	0.011078

A final genome annotation was performed by using MAKER together with de novo assembled non-redundant transcripts, predicted proteins, non-coding RNAs and repeats.

2.4. Functional Classification by Gene Ontology

From a total of 95,847 unigenes derived from the *B. rotunda* transcriptome, 41,550 unigenes (43.35%) were found significantly scoring BLASTX hits against the NR protein database. Of these, 6850 (7.15% of the total unigenes) returned significant sequence alignments that could not be linked to Gene Ontology entries; 6038 (6.3%) of the GO mapped dataset did not obtain an annotation assignment and we could assign functional labels to 28,662 (29.9%) of the input sequences (Figure S9). Species distribution among the BLASTX matches showed that *M. acuminata* subsp. *malaccensis* had a very high similarity score with 87,000 top BLASTX hits from *B. rotunda*. Other species matches included *Ethiopian banana*, *Ensete ventricosum* (Musaceae) with 70,000 hits, African oil palm, *Elaeis guineensis* (Arecaceae) with 62,500 BLASTX hits and date palm, *P. dactylifera* (Arecaceae) with 62,000 BLASTX hits (Figure S10). Biological processes were the most represented functional group based on the categorization of the GO classes for the Nr annotated sequences (Figure 3a).

A Blast2GO enzyme code (EC) annotation showed the distribution of *B. rotunda* predicted proteins among six main enzyme classes of oxidoreductases (1400), transferases (3500), hydrolases (2250), lyases (450), isomerases (250), and ligases (270) (Figure S11). The KOG function classification produced Nr hits for 18,767 unigenes annotated and classified functionally into 25 KOG functional categories (Figure 3b). General function prediction was the most represented group (2396) followed by signal transduction mechanism (2178) and posttranslational modification, protein turnover, and chaperons' with 2031 genes. Comparison of all the *B. rotunda* unigenes with the KEGG pathway database resulted in 1494 out of 28,662 annotated unigene sequences (5.21%) being assigned to 145 predicted metabolic pathways.

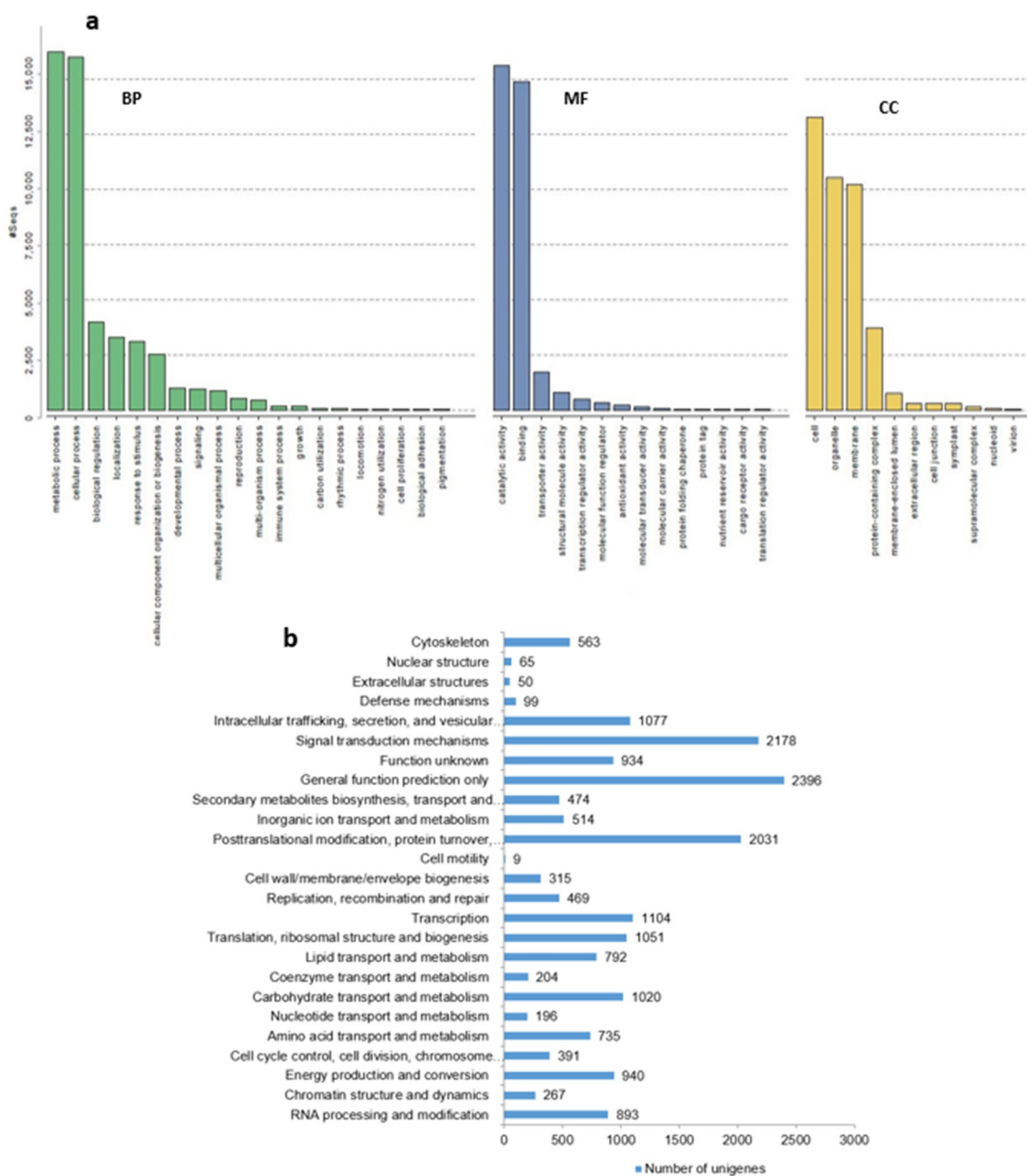


Figure 3. (a) Gene ontology (GO) classification of assembled unigenes of *B. rotunda*. Results are summarized in three main categories: biological process (BP), molecular function (MF), and cellular component (CC). The x-axis indicates the subgroups in the GO annotation while the y-axis indicates the percentage of specific categories of genes in each main category; (b) distribution of Eukaryotic Orthologous Groups (KOG) classification. A total of 18,767 assembled unigenes were annotated and assigned to 25 functional categories. The vertical axis indicates subgroups in the KOG classification and the x-axis represents the number of genes in each main category.

2.5. Phylogenetic Orthology Inference of *B. rotunda* Genes

A total of 62,520 orthogroups were found with Orthofinder [62] (Table S5) with matches of the genes from *B. rotunda* to 979,315 genes from 13 other species (*Glycine max*, *Cucumis melo*, *Gossypium raimondii*, *Brassica napus*, *Arabidopsis thaliana*, *Solanum tuberosum*, *Solanum lycopersicum*, *Musa acuminata*, *Zea mays*, *Oryza sativa* subsp. japonica, *Hordeum vulgare*, *Phoenix dactylifera* and *Brachypodium distachyon*). Of these, 7276 orthogroups were shared among all species and there were no single copy orthogroups (Table S5). The species tree inferred by STAG [63] and rooted by STRIDE [64] indicated that *B. rotunda* has the

closest relationship with *M. acuminata* (order Zingiberales) and *P. dactylifera* (order Arecales) followed by members of the Poaceae family (*Z. mays*, *O. sativa* subsp. japonica, *H. vulgare*, and *B. distachyon*), and was distant from the plant species from the Solanaceae, Brassicaceae, Malvaceae, Cucurbitaceae, and Fabaceae (Figure 4a, Table S5). UpSet plotting showed 7276 orthogroups shared between the *B. rotunda* and 13 selected reference genomes (Figure 4b). Additionally, 1849 protein orthologs are specific for *B. rotunda* and 274 orthogroups are shared among the 13 selected reference genomes except for *B. rotunda*.

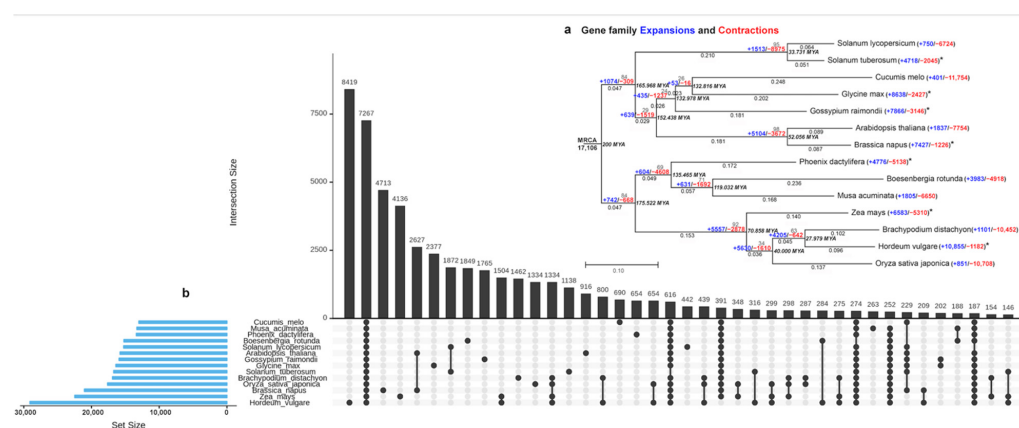


Figure 4. (a) cross-genera phylogenetic analysis of *B. rotunda* and 13 other species. Asterisks after the expansion and contraction figures indicate genomes with reported ancient whole genome duplication or massive segmental duplications or major chromosomal duplications; (b) UpSet plot showing unique and shared protein ortholog clusters of *B. rotunda* and 13 selected reference genomes. Connected dots represent the intersections of overlapping orthologs with the vertical black bars above showing the number of orthogroups in each intersection.

Gene Family Expansion and Contraction

Using the data generated from OrthoFinder [62], we explored the gene family expansion and contractions in *B. rotunda* (Figure 4a). In total, there are 17,106 gene families shared by the most recent common ancestor (MRCA). There were large numbers of gene families expanding (53–10,855) or contracting (16–11,754) between 14 plant genomes (Figure 4a). Our results show the substantial expansion of gene families in the Poaceae (5557) followed by the Brassicaceae (5104) and the Pooideae subfamilies (4205). A large gene family contraction was observed in Solanaceae (8975). Interestingly, the majority of the genomes with reported ancient whole genome duplication or massive segmental duplications or major chromosomal duplications show a higher number of gene family duplications than gene family losses (indicated by asterisks in Figure 4a).

2.6. Transcriptome Changes of *B. rotunda* Unigenes Related to Flavonoid and Phenylpropanoid Biosynthesis Pathways

A Transcriptome analysis showed that, in total, 167 unigenes from *B. rotunda* were mapped to five different classes of enzymes including oxidoreductase, transferase, ligase, lyase, and hydrolase in the flavonoid and phenylpropanoid pathways. Of these, only 23 enzymes showed differential expression in the different samples, i.e., in vitro leaf (IVL), embryogenic callus (EC), and non-embryogenic calli (dry callus (DC) and watery callus (WC)) using ex vitro leaf (EVL) samples as the comparator (Figure 5, Table S6). Phenylalanine ammonia-lyase (PAL), the first enzyme in the phenylpropanoid pathway, turns phenylalanine to cinnamic acid. The PAL was expressed at the lowest levels among all the samples in the IVL, with the highest expression level in the WC (indicated by the dark red squares in Figure 5). Then, the coenzyme A (CoA) would attach to cinnamic acid or p-coumaric acid by 4-coumarate–CoA ligase (4CL) and form cinnamoyl-CoA or p-coumaroyl-CoA. This enzyme showed a relatively higher expression in all the samples except for the EC.

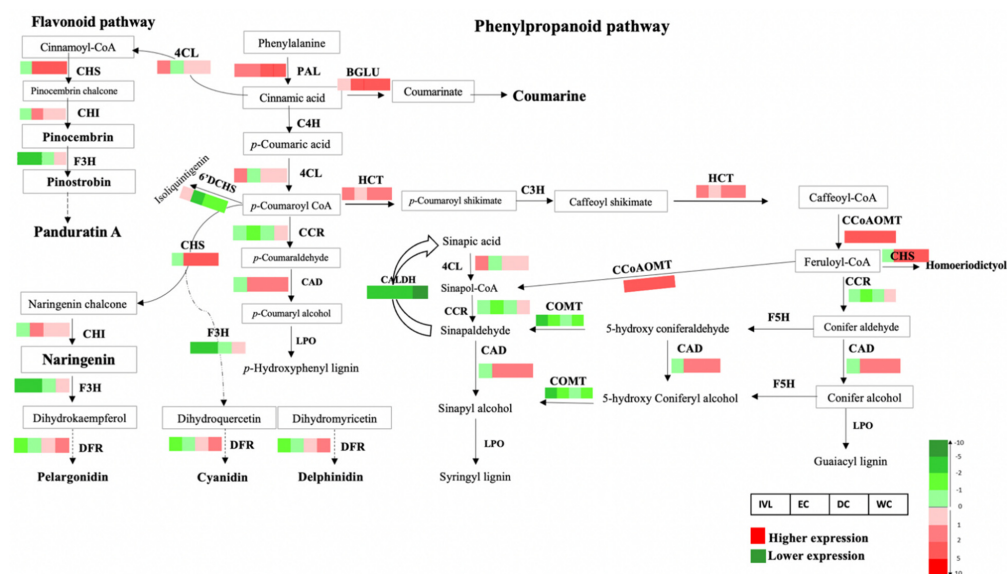


Figure 5. Scheme of the flavonoid and phenylpropanoid biosynthetic pathways in *B. rotunda* based on KEGG pathways. Genes encoding enzymes for each step are indicated as follows: CAD, cinnamyl alcohol dehydrogenase; BGLU, Beta-glucosidase; CALDH, coniferyl-aldehyde dehydrogenase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate–CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; CCoAOMT, caffeoyl-CoA 3-*O*-methyltransferase; C3H, *p*-coumarate 3-hydroxylase; CCR, cinnamoyl-CoA reductase; COMT, caffeic acid *O*-methyltransferase; 6'DCHS, 6'-deoxychalcone synthase; DFR, dihydroflavonol 4-reductase; F3H, flavonoid 3-hydroxylase; F5H, ferulate 5-hydroxylase; HCT, Hydroxycinnamoyl-CoA shikimate; LPO, Lactoperoxidase; PAL, phenylalanine ammonia lyase. Beside each enzyme, four boxes are shown (from left to right): in vitro leaf (IVL), embryogenic callus (EC), dry callus (DC) and watery callus (WC). Red boxes indicate relatively higher mRNA expression compared to the leaf sample with the highest levels in darker red. Green boxes indicate relatively lower expression compared to the leaf sample. The colour box is based on log₂FC values.

In the phenylpropanoid pathway, cinnamic acid is also converted to coumarinate by Beta-glucosidase (BGLU) to produce coumarin. BGLU was expressed in all samples, with the highest expression level in the non-embryogenic calli (NEC). Then CHS, chalcone synthase (CHS) converts cinnamoyl-CoA to pinocembrin chalcone and *p*-coumaroyl CoA to naringenin chalcone. CHS was expressed in all the samples except the IVL, with the highest expression level in the WC. In the next step, the two flavanones of pinocembrin and naringenin are synthesised by chalcone isomerase (CHI). CHI was expressed in all the samples except the IVL, with the highest expression level in the EC. Pinocembrin is converted to pinostrobin by flavanone-3-hydroxylase (F3H) which serves as a precursor of panduratin A synthesis. Expression analysis of the unigenes related to the F3H enzyme and dihydroflavonol 4-reductase (DFR), which are involved in the synthesis of anthocyanidins such as pelargonidin, cyanidin, and delphinidin, showed the DFR to be more highly expressed in the DC and WC compared to the other samples, while the F3H was only relatively up-regulated in the WC. Other enzymes in the phenylpropanoid pathway include hydroxycinnamoyl-CoA shikimate (HCT), cinnamoyl-CoA reductase (CCR), cinnamyl alcohol dehydrogenase (CAD; EC1.1.1.195), caffeoyl-CoA *O*-methyltransferase (CCoAOMT), and lactoperoxidase enzyme (LPO), involved in monolignols synthesis such as *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S). Among them, the HCT, CCoAOMT, and CAD showed a higher expression in all the samples, except the IVL for the CAD, while CCR showed a higher expression in the WC. The gene expression differences between the tissue samples for cinnamic acid 4-hydroxylase (C4H), *p*-coumarate 3-hydroxylase (C3H), ferulate 5-hydroxylase (F5H), caffeic acid *O*-methyltransferase (COMT), and cinnamyl alcohol dehydrogenase (CAD) were below the threshold of the FPKM without any differential expression in the studied samples.

2.7. DNA Methylation Analysis Using Bisulfite Sequencing

Genome wide methylation percentages determined from the bisulfite sequence data from the leaf and four tissue cultured samples, were higher in all methylated cytosine contexts for the samples from the EVL (CpG 73.2%, CHG 36.2% and CHH 33.7%) and IVL (CpG 71.3%, CHG 35.4% and CHH 33.5%). The lowest levels were for the EC (CpG 53.4%, CHG 18.5% and CHH 25.3%), followed by the WC (CpG 63.8%, CHG 21.9% and CHH 28.1%) and the DC (CpG 68.4%, CHG 25.9% and CHH 28.6%). We also evaluated the DNA methylation levels of three groups of genes (30 genes in total) including DNA methyltransferase-related genes across the genome of *B. rotunda* (Figure 6a–d). In general, the CHH methylation levels were higher than the methylation levels in the CpG and CHG context, and out of 30 genes, 22 genes (73.3%) showed low methylation levels (<0.1) in the CHG and CHH cytosine contexts, whereas only 30% of the genes showed low methylation levels in the CpG context. Cytosine methylation of the methylation-related genes in all cytosine contexts (CpG, CHG and CHH) was the highest for the DRM2, followed by MET1 and CMT3 (Figure 6a–d). Among the somatic embryogenesis-related genes, the WOX gene was heavily methylated in the CpG and CHG contexts compared to the other somatic embryogenesis-related genes (LEC2, BBM, SERK) (Figure 6a,d). For the pathway-related genes, the LPOs methylated more in all the studied samples compared to the other genes.

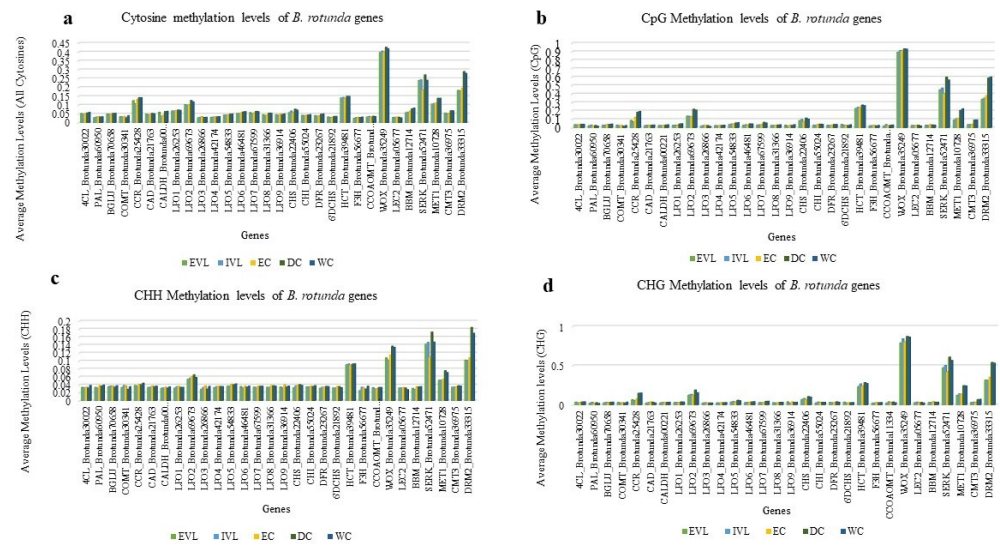


Figure 6. (a–d) average methylation levels of DNA methyltransferase-related genes (MET1, CMT3, and DRM2), somatic embryogenesis genes (SERK, BBM, LEC2, and WOX), and genes involved in flavonoid and phenylpropanoid biosynthesis pathways in different samples of *B. rotunda*; ex-vitro leaf (EVL), in vitro leaf (IVL), embryogenic callus (EC), dry callus (DC), and watery callus (WC). (a) cytosine methylation; (b) CpG methylation; (c) CHG methylation; (d) CHH methylation. CAD, cinnamyl alcohol dehydrogenase; BGLU, Beta-glucosidase; CALDH, coniferyl-aldehyde dehydrogenase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate–CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; CCoAOMT, caffeoyl-CoA 3-O-methyltransferase; C3H, ρ -coumarate 3-hydroxylase; CCR, cinnamoyl-CoA reductase; COMT, caffeic acid O-methyltransferase; 6'DCHS, 6'-deoxychalcone synthase; DFR, dihydroflavonol 4-reductase; F3H, flavonoid 3-hydroxylase; F5H, ferulate 5-hydroxylase; HCT, Hydroxycinnamoyl-CoA shikimate; LPO, Lactoperoxidase; PAL, phenylalanine ammonia lyase; WOX, Wuschel; LEC3, Leafy cotyledon 2; BBM, Baby boom; SERK, Somatic embryogenesis receptor-like kinase; MET1, Methyltransferase 1; CMT3, Chromomethylase 3; DRM2, Domain rearranged methyltransferase 2.

2.8. Correlation between Gene Expression Levels and DNA Methylation Levels of Genes Related to Methylation, Somatic Embryogenesis and Secondary Metabolite Pathway

From the gene expression analysis, we observed that the expression level of the DNA methyltransferase genes, MET 1, CMT 3, and DRM2 was higher in the callus than the leaf

samples and was highest in the embryogenic callus (EC) for all three genes, but lowest in the in vitro leaf (IVL) (Figure 7a). DRM2 showed the lowest level of expression and the highest level of DNA methylation. The DNA methylation levels of these genes at the CpG, CHG, CHH cytosine contexts were the highest in the DC and WC with similar and lower methylation levels in the embryogenic callus and leaf samples. Overall, the expression of methylation-related genes was higher in the samples of EC, DC, and WC but other than for the CMT3, which showed an inverse relationship between the expression level and methylation levels, there was no clear correlation between the level of DNA methylation and the level of gene expression (Figure 7b).

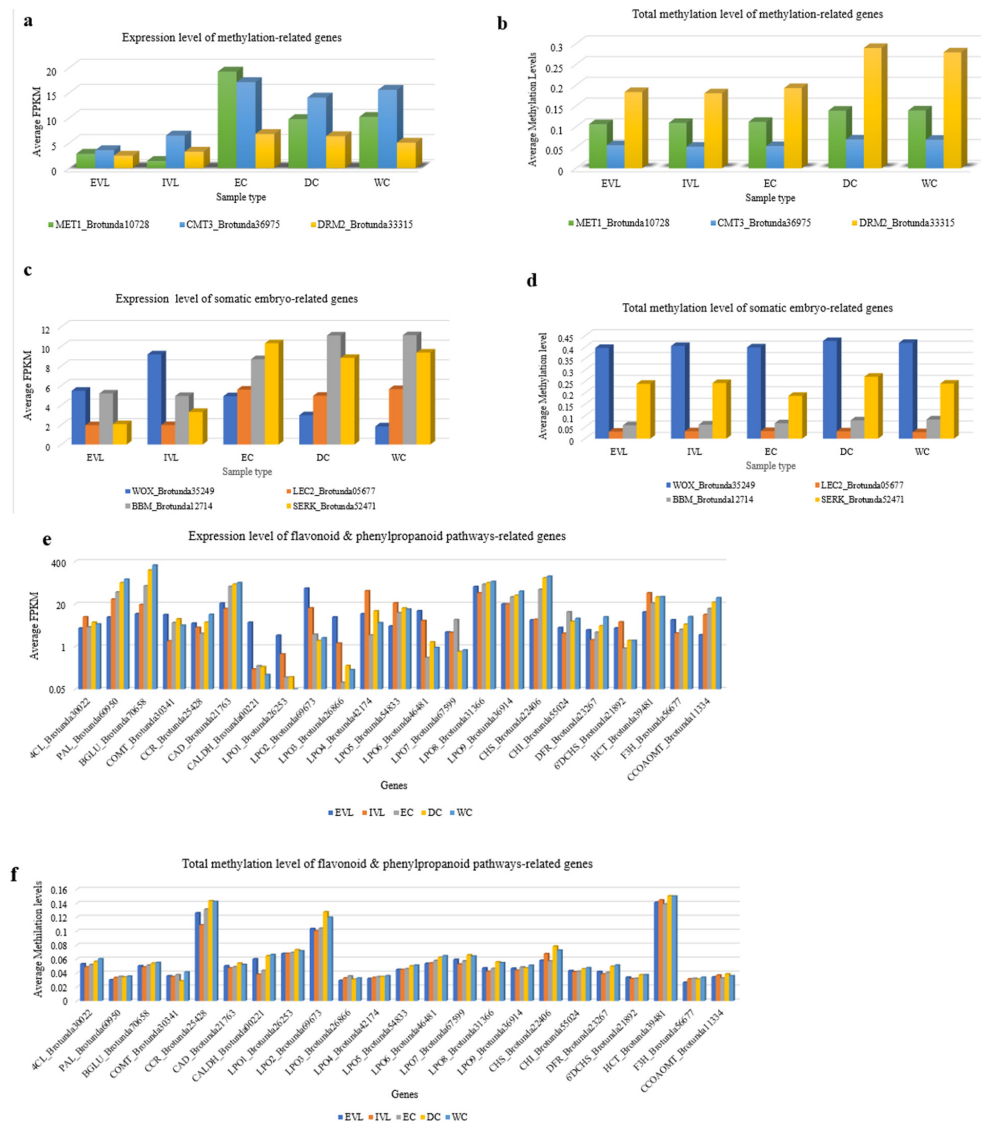


Figure 7. Expression level and total methylation level in all cytosine contexts of methylation-related genes (a,b), somatic embryogenesis-related gene (c,d) and flavonoid and phenylpropanoid biosynthesis pathways-related genes (e,f) in ex vitro leaf (EVL), in vitro leaf (IVL), embryogenic callus (EC), dry callus (DC), and watery callus (WC). CAD, cinnamyl alcohol dehydrogenase; BGLU, Beta-glucosidase; CALDH, coniferyl-aldehyde dehydrogenase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate–CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; CCoAOMT, caffeoyl-CoA 3-O-methyltransferase; C3H, ρ -coumarate 3-hydroxylase; CCR, cinnamoyl-CoA reductase; COMT, caffeic acid O-methyltransferase; 6'DCHS, 6'-deoxychalcone synthase; DFR, dihydroflavonol 4-reductase; F3H, flavonoid 3-hydroxylase; F5H, ferulate 5-hydroxylase; HCT, Hydroxycinnamoyl-CoA shikimate; LPO, Lactoperoxidase; PAL, phenylalanine ammonia lyase.

Similarly, while there were different expression patterns for the four somatic embryogenesis-related genes of SERK, BBM, LEC2, and WOX between the different leaf and callus samples (Figure 7c), the DNA methylation level of each gene across the different leaf and callus samples was largely unchanged (Figure 7d). A comparison of 23 of the *B. rotunda* genes involved in the flavonoid and phenylpropanoid pathways showed them to be expressed differentially in the *B. rotunda* leaf and callus samples. Among them, the BGLU, CAD, CHS, LPO8, LPO9 and PAL were expressed more highly in the callus than in the leaf samples (Figure 7e). The highest level of DNA methylation was observed for the HCT, CCR, and LPO2 genes in all the studied samples and again, there was no general correlation between the gene expression levels and methylation levels for these samples (Figure 7f).

3. Discussion

We present a genome assembly of *Boesenbergia rotunda* ($2n = 36$) with an estimated genome size of 2.4 Gb. The genome of the plant we sequenced, when in cultivation a largely vegetatively propagated species, shows an unusually high heterozygosity of 3.01%, suggesting that the cultivar may be of hybrid origin or may have undergone whole genome duplication events. This is also suggested based on the large number of unigenes, 408 in *B. rotunda*, notably more than twice that of *Ensete glaucum* [57], and 46,765 duplicated genes (65.8% of the *B. rotunda* genome, with at least 50% support). As noted in Citrus limon [65], high levels of heterozygosity complicate the assembly process and due to the clonal propagation nature of the fingerroot ginger, offspring resulting from sexual hybridization is rather limited. Thus, we applied a similar approach as reported by Chin et al. (2016) and Baek et al. (2018), for the assembly of the *B. rotunda* genome [66,67]. The sequencing assembly of *B. rotunda* using long PacBio reads, in addition to the Illumina short-reads, and followed by assembly using the FALCON assembler resulted in a scaffold number of 10,627. This relatively high scaffold number is not unexpected considering the high repeat content (72.51%) of the *B. rotunda* genome, coupled with the relatively high level of heterozygosity (3.01%), and the lack of any molecular marker and breeding data for *B. rotunda*. Future mapping and marker studies could help to resolve an assembly into the anticipated 18 chromosomes, as could more recent technologies such as single chromosome sequencing and optical mapping [56].

Sequence information for other *Boesenbergia* species is not yet available, with the closest relative of *B. rotunda* from sequenced genomes at the time of our study being *M. acuminata*, based on previous analyses using amino acid data from single genes including chalcone isomerase (CHI) [68] and phytyltransferase (BrPT2) [69]. Our phylogeny analysis also showed *M. acuminata* as the closest relative among those compared with *Z. mays*, *O. sativa*, *H. vulgare*, and *B. distachyon* from the Poaceae family, though more distantly related, as expected.

The repeat content of the *B. rotunda* at ~72% of the assembled genome is high compared to many other plant genomes in this order such as *Musa itinerans* (38.95%) [70] and *M. acuminata* (35.43%) [71], but similar to that of *Z. officinale* (ginger official) at 81% [72]. A higher level of repeat content has been observed to correlate with the larger genome sizes in Fabaceae [73] and Melampodium [74]. Both of those reports suggest the greater genome size to be largely driven by Ty3/gypsy LTR-retrotransposons and it is interesting to note that *B. rotunda* also has a high LTR content of 64%. While data for the genome sizes and content are not yet available for other *Boesenbergia* species, the *Z. officinale* genome has a similarly high value of 61% LTR which was also suggested to contribute to the high genome size [75]. Studies in other plant species have reported that plant genomes generally have over 50% transposable elements content (e.g., maize) while some small plant genomes such as *Arabidopsis* may have as low as 10% repeat content [76–78]. Cytosine methylation is usually much denser in transposons than in genes [79–81] and this has also been correlated with the evolution of genome size in angiosperms [77]. The large genome size and high repeat content of *B. rotunda* with relatively low gene body cytosine methylation levels of the genes selected for observation in the current study, fit well with this model and it will

be interesting to compare this with other *Boesenbergia* species in the future when similar data becomes available.

As DNA methylation is dynamic, we saw variations in the global DNA methylation levels in the different samples. Unmethylated DNA has been shown to demarcate expressed genes [82] and so to be able to examine this in the context of gene expression in *B. rotunda* and to add depth to our genome data, we included deep sequencing of the leaf and callus transcriptomes from *B. rotunda*. There are several alternative tools for the de novo assembly of RNA-seq short reads into a reference transcriptome and we compared analysis from four assemblers. The quality of the assembly was noticeably affected by both the k-mer size and assembler tool, with Oases delivering the highest N50 size and average contig length at k-mer 21 compared to at k-mer 24 or other assemblers (Figure S4, Table 3), indicating a more effective and accurate assembly. In comparison to a previous transcriptome assembly of *B. rotunda* by the SOAPdenovo-Trans de novo assembler, our study obtained a longer N50 size (1019) compared to an N50 value of 236 reported by [40]. An Oases assembly of the genome sequence data from a Fern, *Lygodium japonicum*, was also found to give the best mean transcript length and N50 size when compared to assemblies using Trinity and SOAPdenovo-Trans [83]. The BUSCO assessment of the *B. rotunda* transcriptome data also showed that Oases had higher numbers of complete and single copy contigs and less fragmented contigs. Based on this, the transcriptome assembly using Oases offered an improved resource for genome annotation and the gene expression study in *B. rotunda*.

We focused the functional aspects of the *B. rotunda* genome study on the methylation and the flavonoid and phenylpropanoid pathways, as the chalcone, panduratin A, is considered one of the most promising bioactive compounds from *B. rotunda* and previous studies from our research group had indicated that DNA methylation may influence the gene expression in tissue cultured materials [84,85]. From the 23 flavonoid and phenylpropanoid pathway genes that showed differential expression between the leaf and any of the callus samples, most were more highly expressed in the EC, DC, and WC, including PAL, CHS, CHI, DFR, BGLU, HCT, CCOAOMT, and CAD (Figure 7), with the highest expression level in the non-embryogenic callus (DC and WC). This aligns with previous ultra performance liquid chromatography-mass spectrometry (UPLC-MS) data showing the WC followed by the DC to have a higher concentration of panduratin, pinocembrin, pinostrobin, cardamonin and alpinetin [43]. Based on this, the unigenes identified in the genome assembly that correspond to CHS and CHI, encode key enzymes in the biosynthesis of panduratin A in *B. rotunda*. Although DNA methylation plays an important role in the regulation of gene expression, comparison of the methylation of the differentially expressed flavonoid and phenylpropanoid pathway genes, with their cytosine methylation, showed no obvious patterns to indicate any correlation for this gene set.

As our samples included embryogenic and non-embryogenic callus tissue, we also evaluated the expression level of DNA methylase genes (MET1, CMT3, DRM2) and genes related to somatic embryogenesis (SERK, BBM, LEC2, WUS) with the DNA methylation levels across the genome of *B. rotunda* based on a bisulfite sequence analysis. An earlier study with some quantitative qRT-PCR validation suggested that the higher level of expression of the methyltransferase-related genes and the lower CG, CHG and CHH sequence contexts in the EC samples was negatively correlated with the total methylation level of the DNA methyltransferase-related genes [85]. We did observe a similar pattern for CMT3 in all five sample types in the current study (Figure 7); however, no similar correlation between the expression level and cytosine methylation was observed in the current data for the other genes examined. The lack of correlation between the transcript expression and the respective gene body methylation from our data may be due to the limitations of the current genome assembly, such that the cis regions could not be well annotated. In the future, a higher resolution genome assembly for *B. rotunda* would be useful to examine the methylation data from the current study.

Although only a minor portion of the *B. rotunda* genome at around 0.35%, microsatellites are the key elements in plant genomes. Among these, short sequence repeat mi-

cro-satellites (SSRs) have found wide utility as co-dominant markers useful in breeding and diversity studies [86,87]. In this study, we identified the genomic and EST-SSRs from *B. rotunda*, designing primers and showing several to have a transferability to *Musa* and *Ensete* genomes, mostly in silico analysis, but with 14 tested in PCR experiments. *Boesenbergia*, *Musa* and *Ensete* are members of the same plant family Zingiberales, and all have abundant AT-rich SSR sequences; however, they are not from the same genus, so are phylogenetically somewhat distant as reflected in the fairly low numbers with potential as markers across these species. Nevertheless, these newly developed SSR markers enhance the genetic resources for *B. rotunda* as well as the plant family Zingiberales, and these markers could be utilized for genotyping, population structure analysis, association studies and cultivar identification as well as any other breeding application of the *Boesenbergia* spp.

In conclusion, the genome assembly of *B. rotunda* covers some 2300 Mbp divided among 18 relatively similar submetacentric chromosomes. The cultivated accession sequenced was highly heterozygous. The genome assembly, transcriptome, gene expression, SSR analysis and DNA methylation data from this study are resources that will allow further understanding of the unique secondary metabolite properties and their biosynthetic pathways in the genus *Boesenbergia* and for functional genomics of *B. rotunda* characteristics. As this data represent a first report of a *Boesenbergia* genome, this, together with the existing data from sister families such as Musacea, can contribute to the understanding of the pan-genome of the Zingiberales, evolution of the ginger plant family and the potential genetic selection or improvement of gingers.

4. Materials and Methods

4.1. Ethics

The conduct of this study was approved by the University of Malaya's grant management committee, chaired by Professor Noorsaadah Abdul Rahman (noorsaadah@um.edu.my), the Director of the Institute of Research Management and Monitoring, and no human, animal, or endangered or protected plant species were used as materials.

4.2. Plant Materials and Samples Preparation

Rhizomes of *B. rotunda* (L.) Mansf. were collected from a commercial farm in Temerloh, Pahang, Malaysia (Latitude: 3.27° N, Longitude: 102.25° E) and propagated in the laboratory using the steps outlined by Karim et al. (2018b) [85]. To promote rhizome sprouting, first the rhizomes were cleaned under running tap water for ten minutes, then they were allowed to air dry before being planted in the black polybags.

Each day, water was sprayed on the samples to induce the growth of shoots and leaves. Four weeks after planting, young ex vitro leaf (EVL) samples were collected from the rhizome-derived plants. To generate the callus samples, meristematic block explants were subcultured on MS medium supplemented with 30 g L⁻¹ of sucrose and 2 g L⁻¹ of Gelrite[®] with 2,4-dichlorophenoxy acetic acid (2,4-D) at concentrations of 1 mg L⁻¹ (4.5 µM) for the watery callus (WC), 3 mg L⁻¹ (13.5 µM) for the embryogenic callus (EC) and 4 mg L⁻¹ (18 µM) for the dry callus (DC). The WC, EC and DC samples were collected after four weeks on the respective media (eight weeks after initial culturing from the explant). Plants grown from the embryogenic calli were cultured on regeneration media (MS0) and the in vitro leaves (IVL) were collected after 8 weeks, when the leaves were large enough for nucleic acid extraction (16 weeks after the initial culturing from meristematic block explants) [84].

4.3. DNA Extraction and Sequencing for Genome and Bisulfite Sequence (BS-Seq) Analysis

Total genomic DNA was isolated using a modified cetyl trimethyl ammonium bromide (CTAB) method from the ex vitro leaf (EVL) of *B. rotunda* [88]. The quality and quantity of the extracted genomic DNA were determined by a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and Qubit[®] 2.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). The DNA sample was sent to BGI

Shenzhen (Shenzhen, China) for library construction and de novo sequencing on the Illumina HiSeq2000 and HiSeq2500 platforms (Illumina Inc., San Diego, CA, USA) and the PacBio RS II platform (PacBio Inc., Menlo Park, CA, USA) [89]. The BS sequence analysis was performed using the extracted DNA from the EVL, EC, DC, WC, and IVL treated by sodium bisulfite. The paired-end reads were generated using an Illumina HiSeq™ 2000 platform (Illumina Inc., San Diego, CA, USA) by Sengenics Sdn. Bhd., Malaysia from a total of five samples with three biological replicates for each sample.

4.4. RNA Extraction and Sequencing for Transcriptome (RNA-Seq) Analysis

The total RNA was extracted from the EVL, EC, DC, WC, and IVL samples using a CTAB method [90] and three independent sets of RNAs for each sample were generated. High quality RNA samples were sent to BGI-Shenzhen (Shenzhen, China) for library construction and sequencing using the Illumina Genome Analyzer IIX (GAIIx) platform (Illumina Inc., San Diego, CA, USA) to generate single-end reads.

4.5. Determination of Chromosome Number and Location of 45S and 5S rDNA Sites on Metaphase Chromosomes of *B. rotunda* ($2n = 36$) Using Fluorescent In Situ Hybridization (FISH)

The FISH technique was performed according to Schwarzacher and Heslop-Harrison (2000) [91]. ‘Fingers’ of *B. rotunda* were placed in shallow dishes with soil to initiate root growth and kept in the glasshouse at the University of Leicester, UK. Newly grown root tips of 1–2 cm length were treated with 2 mM 8-hydroxyquinoline at growth temperature for 2 h followed by incubation overnight at 4 °C, and then fixed with 96% ethanol:glacial acetic acid (3:1). The roots were digested for 1–3 h at 37 °C with a mixture of cellulose (32 U/mL, Sigma-Aldrich, St. Louis, MO, USA, C1184), ‘Onozuka’ RS cellulose (20 U/mL), pectinase (from *Aspergillus niger*, Sigma-Aldrich P4716) and Viscozyme (20 U/mL, Sigma-Aldrich V2010) in a 10 mM citric acid/sodium citrate buffer (pH4.6). Chromosome preparations of dissected meristems were made in 60% acetic acid by squashing under a cover slip. Slides were stored at –20 °C until the FISH.

The 45S rDNA and 5S rDNA probes were labelled by random priming (Invitrogen Inc., Carlsbad, CA, USA) with digoxigenin 11-dUTP or biotin 11-dUTP (Roche, Basel, Switzerland) using the linearised clone pTa71 or the PCR amplified insert of clone pTa794 (from *Triticum aestivum*, Gerlach and Bedbrook 1979), respectively [92]. For hybridization, 50–100 ng of the labelled probes were prepared in a 40–50 µL mixture of 40% (*v/v*) formamide, 20% (*w/v*) dextran sulphate, 2xSSC (sodium chloride sodium citrate), 0.03 µg of salmon sperm DNA, 0.12% SDS (sodium dodecyl sulphate) and 0.12 mM EDTA (ethylenediamine-tetra acetic acid). The chromosomes and probe mixtures were denatured together at 70 °C for 6–8 min, before cooling down slowly to 37 °C and hybridized for 16 h at 37 °C. The slides were washed at 42 °C in 0.1xSSC and the hybridization sites were detected with anti-digoxigenin-FITC (2 µg/mL; Roche) and Streptavidin-Alexa594 (1 µg/mL; Molecular Probes Inc., Eugene, OR, USA). The chromosomes were counterstained with DAPI (4',6-diamidino-2-phenylindole, 4 µg/mL) and mounted in CitifluorAF. The slides were examined with a Nikon Eclipse 80i microscope and images were captured using NIS-Elements v2.34 (Nikon, Tokyo, Japan), and a DS-QiMc monochrome camera. The images were pseudocoloured and the final figures were prepared with Adobe Photoshop CC2018 using enhancements that treat all pixels of the image [90].

4.6. Genome Size Estimation

Flow cytometry was used to determine the genome size of the *B. rotunda* as an unknown sample on a MACSQuant Analyzer (Miltenyl Biotec Inc., Bergisch Gladbach, Germany), using soybean (*Glycine max* cv. Polanka (G) 2C = 2.50 pg DNA) and pea (*Pisum sativum* cv. Ctirad (P) 2C = 9.09 pg DNA) as the internal standards and propidium iodide as the stain. Each plant (sample and comparator) was compared using an average of four biological replicates [53,93,94]. We also performed a K-mer analysis to estimate the *B. rotunda* genome size and heterozygosity rate using Jellyfish [95] and GenomeScope [96].

4.7. Genome Assembly

A combination of sequencing technologies of the PacBio RSII platform, Illumina HiSeq 2500 paired-end reads (PE) with a 450 bp insert size library, and Illumina HiSeq 2000 mate-pair reads (MP) with insert size libraries of 2, 5, 10, 20, and 40 kb was performed for the genome assembly. Before assembly, the Illumina HiSeq sequence reads were filtered by removing the adaptors and low-quality nucleotides. PacBio reads were filtered to remove the short reads of less than 500 bp or a quality score lower than 0.8, then error correction for the long reads was completed by FALCON [67,97], following the general principles proposed by [98]. We tried to use several de novo assemblers to construct the assembly with both the Illumina and PacBio reads. Finally, we chose the SMARTdenovo software [99]. Corrected PacBio reads were assembled with the SMARTdenovo software (<https://github.com/ruanjue/smartdenovo> accessed on 1 June 2018) to construct the contigs. For the PacBio data, constructed contigs were subsequently polished by stand-alone consensus modules [98] and Pilon software [100] for the Illumina PE reads. Polished contigs were used as input for the scaffolding. The scaffolds were constructed by SOAP scaffolding, and the SSPACE tool [101] with Illumina mate-pair reads (2–40 k) with default parameters to extend the length of the scaffolds for the raw assembly. The gaps within the scaffolds, and consensus sequences generated from the PacBio sub-reads were filled using PBJelly2 [102]. Finally, the scaffolds were corrected by Pilon [100] with Illumina PE reads to correct the assembly errors and to obtain the final genome assembly.

The completeness of the genome assembly was tested by searching for 1440 core eukaryotic genes using benchmarking universal single-copy orthologs (BUSCO) (v2.0) [61]. The quality of the genome assembly was assessed by mapping of the Illumina paired-end 250 bp read data to the contigs using BWA-MEM (version 0.7.15-r1142) [103].

4.8. Repeat Annotation

Tandem repeats were identified with the tandem repeat finder (TRF) [104] (version 4.0.4). The transposable elements (TE) were identified with integrated homology-based and de novo methods [53]. The homology-based prediction was completed at the DNA and protein levels by comparing the assembly to the RepBase v.20.04 [105] database as a query library using RepeatMasker v.4.0.7 (<http://www.repeatmasker.org/> accessed on 1 June 2018) and ProteinRepeatMask v.4.0.7 (<http://www.repeatmasker.org/> accessed on 1 June 2018). To search those absent TEs in the RepBase library, a de novo repeat library was constructed using RepeatModeler v.1.0.10 (<http://www.repeatmasker.org/> accessed on 1 June 2018) to run against the *B. rotunda* genome assembly using RepeatMasker v.4.0.7 (<http://www.repeatmasker.org/> accessed on 1 June 2018).

4.9. Gene Annotation

The protein-coding genes prediction was completed by homology, de novo, and RNA-Seq-based approaches. For generation of the homology-based predictions, the gene sets from four species, i.e., *M. acuminata* (<http://www.promusa.org/Musa+acuminata> accessed on 28 May 2019), *P. dactylifera*, *O. sativa* (<http://rice.plantbiology.msu.edu/> accessed on 28 May 2019) and *A. thaliana* (<https://www.arabidopsis.org/> accessed May 2019) were downloaded. The nonredundant protein sequences for each gene set was searched by TBLASTN. For generation of the expression-based evidence, RNA-Seq short reads originated from the EVL, IVL, EC, DC, and WC tissues were mapped to the ginger genome with the Hisat2 v.2.0.4 [106] alignment program. For the de novo gene annotation, well-supported transcripts identified both by the homology-based and the RNA-Seq based predictions were selected for ab initio prediction using AUGUSTUS v.3.2.3 program [107,108]. The exon–intron structure of the genes was predicted using Genscan [109] and SNAP [110]. The results from the three approaches were consolidated using MAKER v.2.31.9 [111] to generate a protein-coding gene set. For the functional information, in silico translated products of the coding genes were aligned to the seven known protein databases of NR [112], InterPro [113], GO [114], KEGG [115], Swissprot and TrEMBL [116], COG [117].

4.10. Non-Coding RNAs Annotation

Four types of ncRNA, including microRNA (miRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), and small nuclear RNA (snRNA), were annotated in the assembled genome based on de novo/or homology-based search methods. The tRNAs genes were annotated using tRNAScan-SE v.1.3.1 [118] with default parameters and filtered to remove the pseudo annotated tRNA genes. To identify the rRNA genes, the *B. rotunda* genome assembly was searched against the rRNA template sequences (Rfam database, release 13.0) [119] of *A. thaliana*, *ca* and *M. acuminata* with BLASTN, with an identity cutoff of $\geq 90\%$ and a coverage at 80% or more. Using Infernal v.1.1.2 [120], mapping of the *B. rotunda* genome sequences to the Rfam database was performed to identify the miRNA and snRNA genes [53,89].

4.11. Construction of Phylogenetic Trees

The conserved orthologs genes (COS) in the *B. rotunda* genome and 13 other species were identified using the Orthofinder program [62]. Using identified single-copy orthologous genes, a neighbour joining (NJ) tree was constructed using a MEGAX [121] and UpSet plot using UpSetR [122].

4.12. Gene Family Expansion and Contraction Analysis

The data generated from OrthoFinder were used as input for the computational Analysis of gene Family Evolution (CAFE) [123] to estimate the gene family expansion and contraction. The phylogenetic tree from OrthoFinder was converted to an ultrametric tree using the `make_ultrametric.py` in OrthoFinder. Gene families with a large variance (≥ 100 gene copies) were removed using the `clade_and_size_filter.py` in the CAFE package. Divergence times in the phylogenetic tree were estimated using PATHd8 [124] and were calibrated using the divergence time between *Brachypodium* and *Oryza* (40–45 million years ago) (The International Brachypodium Initiative 2010) [125] and *Arabidopsis* and *Oryza* (130–200 million years ago) [126]. CAFE version 5 [123] was used to determine the stochastic birth and death processes and for modelling of the gene family evolution. The parameters for CAFE5 were “`cafe5 -i orthofinder_gene_families.txt -t orthofinder_ultrametric.tre -p -e`”.

4.13. DNA Methylation Analysis Using Bisulfite Sequencing (BS-Seq)

For the bisulfite sequencing analysis, low quality reads and adapters were trimmed by the Trim-Galore [127] tool specific for bisulfite sequencing. The trimmed reads were mapped to draft the ginger genome with the Bismark [128] tool by choosing the Bowtie aligner with options set to the best, minimum map length of 50 bp and insert size of 500 bp. Mapping duplicates were removed by the Methpipe [129] tool. The Methcounts program from the Methpipe software package was used for the mapping of methylated and unmethylated cytosines where the methylation level at a single base resolution was calculated based on the number of 5-methylated cytosines (5 mC) in the reads, divided by the sum of the C and thymines (T) in the CG, CHG and CHH sequence contexts within the coding sequences of all the selected genes from *B. rotunda* [84].

4.14. De Novo Transcriptome Assembly of *B. rotunda* and Functional Annotation

To gather information related to the secondary metabolites, expression of the genes involved in the flavonoid and phenylpropanoid pathways of *B. rotunda* was based on the deep transcriptome sequencing of three cell culture types, in vitro and ex vitro leaves of *B. rotunda*. Based on our previous studies of embryogenesis related genes [84,85] and on the levels of metabolites in the cell cultures [43], we generated deep transcriptome data from five tissue types (each with three replicates) to investigate the gene regulation patterns in the phenylpropanoid and flavonoid pathways, to identify the metabolite producing cells in the *B. rotunda* in vitro cultured cells. The quality of the RNA-Seq raw reads were checked by the FastQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> accessed on 28 May 2018) to remove the low-quality reads and adapter sequences. Four

different short-read assemblers: Oases [130], TransABySS [131], SOAPdenovo-Trans [132] and Trinity [133], were used to assemble the high-quality reads to the contigs. Using a best k-mer strategy, the high-quality reads were assembled at different k-mer lengths 21–51 using Oases, TransABySS and SOAPdenovo-Trans, whereas the assembly by Trinity used default parameters (K-mer 25). The contigs generated from all the k-mers by each respective assembler were merged and the redundancy was removed using CD-HIT [134]. Then, the TGICL clustering tool [135] was used to assemble the non-redundant contigs with a maximum identity of 90 and a minimum overlap length of 40. The completeness of the transcriptome assemblies was measured using the BUSCO [61] software.

High-throughput functional annotation was performed with the Blast2GO Command Line [136]. The BLASTX algorithm was performed to obtain a list of the potential homologous for each input sequence. To evaluate the functional annotation for the query sequences [137], gene ontology (GO) terms associated with the obtained BLAST hits were mapped by Blast2GO. GO mapping and Enzyme Commission (EC) classification were performed based on the annotation parameters of a Cut-off 55, E-Value 1×10^{-6} , GO Weight of 5, and HSP-Hit Coverage Cut-off 0. In order to identify the functional enrichment categories among the differentially expressed genes (DEGs), a Fisher exact test with a false discovery rate (FDR) cutoff of 0.05 was used. Classification of the *B. rotunda* transcripts into functional categories was performed using the Eukaryotic Orthologous Groups (KOG) [117] protein database. Biological pathways mapping of the *B. rotunda* transcripts was completed using the KEGG database [115]. Unigenes potentially related to the Panduratin A and other secondary metabolites biosynthesis were identified as those with a unigene annotated function matching to the enzymes assigned to the flavonoid and phenylpropanoid biosynthesis pathways in the KEGG pathway database.

4.15. Estimation of Transcript Abundance and Differential Expression

The RSEM software package [138] was used for the estimation of the gene expression level with a mean fragment length of 200 bp and fragment length standard deviation of 80 bp. Then, the FPKMs (fragments per feature kilobase per million reads mapped) were used to normalize the expression level for each gene and comparison between the samples. The Bioconductor tool (EdgeR) [139] was used for the differential expression analysis with a *p*-value threshold of ≤ 0.05 and $|\log_2(\text{Fold Change})| \geq 1$ used to identify the significant differential expression of the transcripts.

4.16. Mining of Simple Sequence Repeats (SSRs) from *B. rotunda* Transcriptome and Genome Assembly

The whole genome assembly and assembled transcriptome sequences were searched for SSRs using a modified Liliaceae simple sequence analysis tool (LSAT) pipeline [140]. The searches were standardized for mining SSRs from mono to 20 bp with a minimum repeat loci of 12 nucleotides. The SSRs were classified based on the SSR locus length (class I > 20 nt and class II 12–20 nt) and the nucleotide base composition of the SSR loci (AT-rich, GC-rich and AT-GC balance). Primer pair sequences were developed for each identified SSR loci using the default parameters of the primer 3 (<http://bioinfo.ut.ee/primer3> accessed on 28 May 2020) software [141]. Redundant primer pairs were eliminated using the perl script developed by Biswas et al. [87]. An electronic polymerase chain reaction (ePCR) [142] strategy was applied for mapping and estimating the transferability of the designed primers. The primers were mapped on four genomes viz. *M. acuminata*, *M. balbisiana*, *M. itinerans* and *E. ventricosum*, as those are the most related plant species of *B. rotunda*. A maximum 2 nt mismatch with two gaps was set as a cut off value for the ePCR result filter.

Wet lab validation of the transcriptome SSR (EST-SSR) and genomic SSR (G-SSR) markers

A total of 14 (8 EST-SSR and 6 G-SSR) primer pairs were selected based on their in silico transferability result to assess their marker potentiality. Three *B. rotunda*, two *Ensete* and three *Musa* species were used to validate the selected primer sets. Fresh leaf samples were harvested from the greenhouse grown plants and the total genomic DNA was extracted

following the CTAB methods. PCR amplifications were carried out for the SSR primer validation under the following conditions: 95 °C for 2 min, 35 cycles at 95 °C for 1 min, 60 °C for 1 min, and 72 °C for 1 min, followed by a final elongation at 72 °C for 10 min. The amplified DNA fragments were run on 2% agarose gels in a 1 × Tris–Borate–EDTA (TBE) buffer with 80v for 90 min. A 100 bp molecular ladder was used to estimate the amplicon size.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23137269/s1>.

Author Contributions: J.A.H., N.K., J.S.H.-H. and T.S. conceived and designed the study; S.T., C.H.T., Y.S.T., M.K.B., N.V.R.M., W.Y.W. and H.M.G. performed the data acquisition, genome sequence assembly and bioinformatics analyses; T.S. and Y.M.-Y. performed the chromosome analysis; S.T., C.H.T., J.A.H. and J.S.H.-H. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the University of Malaya: High Impact Research Chancellery Grant UM.C/625/1/HIR/MOE/SC/15, CEBAR grants TU002G-2018 and RU004A-2020; and by the Biotechnology & Biological Sciences Research Council (BBSRC) Grant No. BB/P02307X/1. The APC was funded by the University of Leicester Open access publishing fund.

Data Availability Statement: Raw sequence data used for genome assembly, mRNA sequencing (RNA-Seq) and whole-genome bisulfite sequencing (BS-Seq) are available at NCBI under BioProject ID PRJNA712941.

Acknowledgments: The work was supported by the High Impact Research Chancellery Grant UM.C/625/1/HIR/MOE/SC/15 from the University of Malaya, Kuala Lumpur, Malaysia. MKB, TS and JSHH acknowledge partial support of GCRF Foundation Awards for Global Agricultural and Food Systems Research, entitled, ‘Modelling and genomics resources to enhance exploitation of the sustainable and diverse Ethiopian starch crop enset and support livelihoods’ [Grant No. BB/P02307X/1] from the Biotechnology & Biological Sciences Research Council (BBSRC). We acknowledge our colleagues at the Centre for Research in Biotechnology for Agriculture (CEBAR) and the Plant Biotechnology Research Laboratory (PBRL), University of Malaya and from the Department of Genetics and Genome Biology, University of Leicester, for their kind support and guidance during this research and CEBAR grants TU002G-2018 and RU004A-2020 (for support in research facilities and maintenance).

Conflicts of Interest: The authors declare no conflict to interest.

References

1. Benedict, J.C.; Smith, S.Y.; Specht, C.D.; Collinson, M.E.; Leong-Škorničková, J.; Parkinson, D.Y.; Marone, F. Species diversity driven by morphological and ecological disparity: A case study of comparative seed morphology and anatomy across a large monocot order. *AoB Plants* **2016**, *8*, plw063. [[CrossRef](#)] [[PubMed](#)]
2. Christenhusz, M.J.; Byng, J.W. The number of known plants species in the world and its annual increase. *Phytotaxa* **2016**, *261*, 201–217. [[CrossRef](#)]
3. Baker, J.G. The Flora of British India. In *Scitamineae*; Hooker, J.D., Ed.; Reeve & Co.: London, UK, 1890; Volume 6.
4. Burkill, I.H. A dictionary of the economic products of the Malay Peninsula. In *A Dictionary of the Economic Products of the Malay Peninsula*; Ministry of Agriculture and Co-Operatives: Kuala Lumpur, Malaysia, 1966; Volume 2.
5. Larsen, K. A preliminary checklist of the Zingiberaceae of Thailand. *Thai For. Bull.* **1996**, *24*, 35–49.
6. Larsen, K.; Ibrahim, H.; Khaw, S.; Saw, L. *Gingers of Peninsular Malaysia and Singapore*; Natural History Borneo: Kota Kinabalu, Malaysia, 1999.
7. Rachkeeree, A.; Kantadoung, K.; Suksathan, R.; Puangpradab, R.; Page, P.A.; Sommano, S.R. Nutritional compositions and phytochemical properties of the edible flowers from selected Zingiberaceae found in Thailand. *Front. Nutr.* **2018**, *5*, 3. [[CrossRef](#)]
8. Jing, L.J.; Mohamed, M.; Rahmat, A.; Bakar, M.F.A. Phytochemicals, antioxidant properties and anticancer investigations of the different parts of several gingers species (*Boesenbergia rotunda*, *Boesenbergia pulchella* var *attenuata* and *Boesenbergia armeniaca*). *J. Med. Plant Res.* **2010**, *4*, 027–032.
9. Larsen, K.; Lock, J.; Maas, H.; Maas, P. Zingiberaceae. In *Flowering Plants Monocotyledons*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 474–495.
10. Mood, J.; Veldkamp, J.; Dey, S.; Prince, L. Nomenclatural changes in Zingiberaceae: Caulokaempferia is a superfluous name for Monolophus and Jirawongsea is reduced to Boesenbergia. *Bull. Singap.* **2014**, *66*, 215–231.

11. Eng-Chong, T.; Yean-Kee, L.; Chin-Fei, C.; Choon-Han, H.; Sher-Ming, W.; Li-Ping, C.T.; Gen-Teck, F.; Khalid, N.; Abd Rahman, N.; Karsani, S.A. *Boesenbergia rotunda*: From ethnomedicine to drug discovery. *Evid. Based Complement. Altern. Med.* **2012**, *2012*, 25. [[CrossRef](#)]
12. Jirakiattikul, Y.; Rithichai, P.; Prachai, R.; Itharat, A. Elicitation enhancement of bioactive compound accumulation and antioxidant activity in shoot cultures of *Boesenbergia rotunda* L. *Agric. Nat. Resour* **2021**, *55*, 456–463.
13. Shen, N.; Wang, T.; Gan, Q.; Liu, S.; Wang, L.; Biao, J. Plant flavonoids: Classification, distribution, biosynthesis, and antioxidant activity. *Food Chem.* **2022**, *383*, 132531. [[CrossRef](#)]
14. Trakoontivakorn, G.; Nakahara, K.; Shinmoto, H.; Takenaka, M.; Onishi-Kameyama, M.; Ono, H.; Yoshida, M.; Nagata, T.; Tsushida, T. Structural analysis of a novel antimutagenic compound, 4-hydroxypanduratin A, and the antimutagenic activity of flavonoids in a Thai spice, fingerroot (*Boesenbergia pandurata* Schult.) against mutagenic heterocyclic amines. *J. Agric. Food Chem.* **2001**, *49*, 3046–3050. [[CrossRef](#)]
15. Jaipetch, T.; Kanghae, S.; Pancharoen, O.; Patrick, V.; Reutrakul, V.; Tuntiwachwuttikul, P.; White, A. Constituents of *Boesenbergia pandurata* (syn. *Kaempferia pandurata*): Isolation, crystal structure and synthesis of (±)-Boesenbergin A. *Aust. J. Chem.* **1982**, *35*, 351–361.
16. Tan, S.K. *Flavanoids from Boesenbergia rotunda (L.) Mansf.: Chemistry, Bioactivity and Accumulation*; Universiti Malaya: Kuala Lumpur, Malaysia, 2005.
17. Tan, B.C.; Tan, S.K.; Wong, S.M.; Ata, N.; Rahman, N.A.; Khalid, N. Distribution of flavonoids and cyclohexenyl chalcone derivatives in conventional propagated and in vitro-derived field-grown *Boesenbergia rotunda* (L.) Mansf. *Evid. Based Complement. Altern. Med.* **2015**, *2015*, 451870. [[CrossRef](#)] [[PubMed](#)]
18. Yusuf, N.A.; M Annuar, M.S.; Khalid, N. Existence of bioactive flavonoids in rhizomes and plant cell cultures of '*Boesenbergia rotunda*' (L.) Mansf. *Kulturpfl. Aust. J. Crop Sci.* **2013**, *7*, 730.
19. Rosdianto, A.M.; Puspitasari, I.M.; Lesmana, R.; Levita, J. Determination of Quercetin and Flavonol Synthase in *Boesenbergia rotunda* Rhizome. *Pak. J. Biol. Sci.* **2020**, *23*, 264–270. [[CrossRef](#)] [[PubMed](#)]
20. Ching, A.Y.L.; Wah, T.S.; Sukari, M.A.; Cheng Lian, G.E.; Rahmani, M.; Khalid, K. Characterization of flavonoid derivatives from *Boesenbergia rotunda* (L.). *Malays. J. Anal. Sci.* **2007**, *11*, 154–159.
21. Tuchinda, P.; Reutrakul, V.; Claeson, P.; Pongprayoon, U.; Sematong, T.; Santisuk, T.; Taylor, W.C. Anti-inflammatory cyclohexenyl chalcone derivatives in *Boesenbergia pandurata*. *Phytochemistry* **2002**, *59*, 169–173. [[CrossRef](#)]
22. Rosdianto, A.M.; Puspitasari, I.M.; Lesmana, R.; Levita, J.J.J. Bioactive compounds of *Boesenbergia* sp. and their anti-inflammatory mechanism: A review. *J. Appl. Pharm. Sci.* **2020**, *10*, 116–126.
23. Break, M.K.B.; Chiang, M.; Wiart, C.; Chin, C.-F.; Khoo, A.S.B.; Khoo, T.-J. Cytotoxic Activity of *Boesenbergia rotunda* Extracts against Nasopharyngeal Carcinoma Cells (HK1). Cardamonin, a *Boesenbergia rotunda* Constituent, Inhibits Growth and Migration of HK1 Cells by Inducing Caspase-Dependent Apoptosis and G2/M-Phase Arrest. *Nutr. Cancer* **2021**, *73*, 473–483. [[CrossRef](#)]
24. Kirana, C.; Jones, G.P.; Record, I.R.; McIntosh, G.H. Anticancer properties of panduratin A isolated from *Boesenbergia pandurata* (Zingiberaceae). *J. Nat. Med.* **2007**, *61*, 131–137. [[CrossRef](#)]
25. Liu, Q.; Cao, Y.; Zhou, P.; Gui, S.; Wu, X.; Xia, Y.; Tu, J. Panduratin A inhibits cell proliferation by inducing G0/G1 phase cell cycle arrest and induces apoptosis in breast cancer cells. *Biomol. Ther.* **2018**, *26*, 328. [[CrossRef](#)]
26. Tanigaki, R.; Takahashi, R.; Nguyen, M.T.T.; Nguyen, N.T.; Do, T.V.N.; Nguyen, H.X.; Kataoka, T. 4-Hydroxypanduratin A and isopanduratin A inhibit tumor necrosis factor α -stimulated gene expression and the nuclear factor κ B-dependent signaling pathway in human lung adenocarcinoma A549 cells. *Biol. Pharm. Bull.* **2019**, *42*, 26–33. [[CrossRef](#)] [[PubMed](#)]
27. Win, N.N.; Awale, S.; Esumi, H.; Tezuka, Y.; Kadota, S. Panduratin D—I, novel secondary metabolites from rhizomes of *Boesenbergia pandurata*. *Chem. Pharm. Bull.* **2008**, *56*, 491–496. [[CrossRef](#)] [[PubMed](#)]
28. Cheenpracha, S.; Karalai, C.; Ponglimanont, C.; Subhadhirasakul, S.; Tewtrakul, S.J.B. Anti-HIV-1 protease activity of compounds from *Boesenbergia pandurata*. *Bioorg. Med. Chem.* **2006**, *14*, 1710–1714. [[CrossRef](#)] [[PubMed](#)]
29. Chee, C.F.; Abdullah, I.; Buckle, M.J.; Rahman, N.A. An efficient synthesis of (±)-panduratin A and (±)-isopanduratin A, inhibitors of dengue-2 viral activity. *Tetrahedron Lett.* **2010**, *51*, 495–498. [[CrossRef](#)]
30. Kiat, T.S.; Pippen, R.; Yusof, R.; Ibrahim, H.; Khalid, N.; Rahman, N.A. Inhibitory activity of cyclohexenyl chalcone derivatives and flavonoids of fingerroot, *Boesenbergia rotunda* (L.), towards dengue-2 virus NS3 protease. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3337–3340. [[CrossRef](#)]
31. Kanjanasirirat, P.; Suksatu, A.; Manopwisedjaroen, S.; Munyoo, B.; Tuchinda, P.; Jearawuttanakul, K.; Seemakhan, S.; Charoensutthivarakul, S.; Wongtrakoongate, P.; Rangkasenee, N. High-content screening of Thai medicinal plants reveals *Boesenbergia rotunda* extract and its component Panduratin A as anti-SARS-CoV-2 agents. *Sci. Rep.* **2020**, *10*, 19963. [[CrossRef](#)]
32. Hwang, J.-K.; Chung, J.-Y.; Baek, N.-I.; Park, J.-H. Isopanduratin A from *Kaempferia pandurata* as an active antibacterial agent against cariogenic *Streptococcus mutans*. *Int. J. Antimicrob. Agents* **2004**, *23*, 377–381. [[CrossRef](#)]
33. Mazlan, R.R.; Zakaria, M.; Rukayadi, Y. Antimicrobial activity of fingerroot [*Boesenbergia rotunda* (L.) Mansf. A.] Extract against *Streptococcus mutans* and *Streptococcus sobrinus*. *J. Pure. Appl. Microbiol.* **2016**, *10*, 1755–1762.
34. Bhamarapravati, S.; Juthapruth, S.; Mahachai, W.; Mahady, G. Antibacterial activity of *Boesenbergia rotunda* (L.) Mansf. and *Myristica fragrans* Houtt. against *Helicobacter pylori*. *Songklanakarini J. Sci. Technol.* **2006**, *28*, 157–163.
35. Pattaratanawadee, E.; Rachtanapun, C.; Wanchaitanawong, P.; Mahakarnchanakul, W. Antimicrobial activity of spice extracts against pathogenic and spoilage microorganisms. *Kasetsart J. Nat. Sci.* **2006**, *40*, 159–165.

36. Hwang, J.K.; Kim, S.Y.; Kim, M.B. Composition Comprising Panduratin or Fingerroot (*Boesenbergia pandurata*) Extract for Treating, Preventing, or Ameliorating Bone Loss Disease. Patent Application No. WO2017150934A1, 3 March 2017.
37. Adhikari, D.; Gong, D.-S.; Oh, S.H.; Sung, E.H.; Lee, S.O.; Kim, D.-W.; Oak, M.-H.; Kim, H.J.J.P. Vasorelaxant Effect of *Boesenbergia rotunda* and Its Active Ingredients on an Isolated Coronary Artery. *Plants* **2020**, *9*, 1688. [[CrossRef](#)] [[PubMed](#)]
38. Tan, S.; Pippen, R.; Yusof, R.; Ibrahim, H.; Rahman, N.; Khalid, N. Simple one-medium formulation regeneration of fingerroot [*Boesenbergia rotunda* (L.) Mansf. Kulturpfl.] via somatic embryogenesis. *Vitr. Cell. Dev. Biol. Plant* **2005**, *41*, 757–761. [[CrossRef](#)]
39. Wong, S.M.; Salim, N.; Harikrishna, J.A.; Khalid, N. Highly efficient plant regeneration via somatic embryogenesis from cell suspension cultures of *Boesenbergia rotunda*. *Vitr. Cell. Dev. Biol.-Plant* **2013**, *49*, 665–673. [[CrossRef](#)]
40. Yusuf, N.A.; Annuar, M.S.; Khalid, N. Rapid micropropagation of *Boesenbergia rotunda* (L.) Mansf. Kulturpfl. (a valuable medicinal plant) from shoot bud explants. *Afr. J. Biotechnol.* **2011**, *10*, 1194–1199.
41. Md-Mustafa, N.D.; Khalid, N.; Gao, H.; Peng, Z.; Alimin, M.F.; Bujang, N.; Ming, W.S.; Mohd-Yusuf, Y.; Harikrishna, J.A.; Othman, R.Y. Transcriptome profiling shows gene regulation patterns in a flavonoid pathway in response to exogenous phenylalanine in *Boesenbergia rotunda* cell culture. *BMC Genom.* **2014**, *15*, 984. [[CrossRef](#)]
42. Wu, M.; Li, Q.; Hu, Z.; Li, X.; Chen, S. The complete *Amomum kravanh* chloroplast genome sequence and phylogenetic analysis of the commelinids. *Molecules* **2017**, *22*, 1875. [[CrossRef](#)]
43. Ng, T.L.M.; Karim, R.; Tan, Y.S.; Teh, H.F.; Danial, A.D.; Ho, L.S.; Khalid, N.; Appleton, D.R.; Harikrishna, J.A. Amino acid and secondary metabolite production in embryogenic and non-embryogenic callus of *Fingerroot ginger* (*Boesenbergia rotunda*). *PLoS ONE* **2016**, *11*, e0156714. [[CrossRef](#)]
44. Chen, D.-x.; Pan, Y.; Wang, Y.; Cui, Y.-Z.; Zhang, Y.-J.; Mo, R.-y.; Wu, X.-l.; Tan, J.; Zhang, J.; Guo, L.-a. The chromosome-level reference genome of *Coptis chinensis* provides insights into genomic evolution and berberine biosynthesis. *Hort. Res.* **2021**, *8*, 121. [[CrossRef](#)]
45. Xia, Z.; Huang, D.; Zhang, S.; Wang, W.; Ma, F.; Wu, B.; Xu, Y.; Xu, B.; Chen, D.; Zou, M. Chromosome-scale genome assembly provides insights into the evolution and flavor synthesis of passion fruit (*Passiflora edulis* Sims). *Hort. Res.* **2021**, *8*, 14. [[CrossRef](#)]
46. Xu, X.; Yuan, H.; Yu, X.; Huang, S.; Sun, Y.; Zhang, T.; Liu, Q.; Tong, H.; Zhang, Y.; Wang, Y. The chromosome-level *Stevia* genome provides insights into steviol glycoside biosynthesis. *Hort. Res.* **2021**, *8*, 189. [[CrossRef](#)]
47. Chakraborty, A.; Mahajan, S.; Jaiswal, S.K.; Sharma, V.K. Genome sequencing of turmeric provides evolutionary insights into its medicinal properties. *Commun. Biol.* **2021**, *4*, 1193. [[CrossRef](#)] [[PubMed](#)]
48. Droc, G.; Lariviere, D.; Guignon, V.; Yahiaoui, N.; This, D.; Garsmeur, O.; Dereeper, A.; Hamelin, C.; Argout, X.; Dufayard, J.F. The banana genome hub. *Database* **2013**, *2013*, bat035. [[CrossRef](#)] [[PubMed](#)]
49. Rijzaani, H.; Bayer, P.E.; Rouard, M.; Doležal, J.; Batley, J.; Edwards, D. The pangenome of banana highlights differences between genera and genomes. *Plant Genome* **2021**, *15*, e20100. [[CrossRef](#)] [[PubMed](#)]
50. Luo, M.-C.; Gu, Y.Q.; Puiiu, D.; Wang, H.; Twardziok, S.O.; Deal, K.R.; Huo, N.; Zhu, T.; Wang, L.; Wang, Y. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* **2017**, *551*, 498–502. [[CrossRef](#)] [[PubMed](#)]
51. Singh, R.; Ong-Abdullah, M.; Low, E.-T.L.; Manaf, M.A.A.; Rosli, R.; Nookiah, R.; Ooi, L.C.-L.; Ooi, S.E.; Chan, K.-L.; Halim, M.A. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* **2013**, *500*, 335–339. [[CrossRef](#)]
52. Schnable, P.S.; Ware, D.; Fulton, R.S.; Stein, J.C.; Wei, F.; Pasternak, S.; Liang, C.; Zhang, J.; Fulton, L.; Graves, T.A. The B73 maize genome: Complexity, diversity, and dynamics. *Science* **2009**, *326*, 1112–1115. [[CrossRef](#)]
53. Wei, C.; Yang, H.; Wang, S.; Zhao, J.; Liu, C.; Gao, L.; Xia, E.; Lu, Y.; Tai, Y.; She, G. Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4151–E4158. [[CrossRef](#)]
54. Xia, E.-H.; Tong, W.; Wu, Q.; Wei, S.; Zhao, J.; Zhang, Z.-Z.; Wei, C.-L.; Wan, X.-C. Tea plant genomics: Achievements, challenges and perspectives. *Hort. Res.* **2020**, *7*, 7. [[CrossRef](#)]
55. Jayakodi, M.; Choi, B.-S.; Lee, S.-C.; Kim, N.-H.; Park, J.Y.; Jang, W.; Lakshmanan, M.; Mohan, S.V.; Lee, D.-Y.; Yang, T.-J. Ginseng genome database: An open-access platform for genomics of *Panax ginseng*. *BMC Plant Biol.* **2018**, *18*, 62. [[CrossRef](#)]
56. Michael, T.P.; VanBuren, R. Building near-complete plant genomes. *Curr. Opin. Plant Biol.* **2020**, *54*, 26–33. [[CrossRef](#)]
57. Wang, Z.; Rouard, M.; Biswas, M.K.; Droc, G.; Cui, D.; Roux, N.; Baurens, F.-C.; Ge, X.-J.; Schwarzacher, T.; Heslop-Harrison, P.J. A chromosome-level reference genome of *Ensete glaucum* gives insight into diversity and chromosomal and repetitive sequence evolution in the Musaceae. *GigaScience* **2022**, *11*, giac027. [[CrossRef](#)] [[PubMed](#)]
58. Tan, E.C.; Karsani, S.A.; Foo, G.T.; Wong, S.M.; Rahman, N.A.; Khalid, N.; Othman, S.; Yusof, R. Proteomic analysis of cell suspension cultures of *Boesenbergia rotunda* induced by phenylalanine: Identification of proteins involved in flavonoid and phenylpropanoid biosynthesis pathways. *Plant Cell Tissue Organ Cult. (PCTOC)* **2012**, *111*, 219–229. [[CrossRef](#)]
59. Bennetzen, J.L.; Wang, H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* **2014**, *65*, 505–530. [[CrossRef](#)]
60. Liu, N.; Niu, Y.; Zhang, G.; Feng, Z.; Bo, Y.; Lian, J.; Wang, B.; Gong, Y. Genome sequencing and population resequencing provide insights into the genetic basis of domestication and diversity of vegetable soybean. *Hort. Res.* **2022**, *9*, uhab052. [[CrossRef](#)] [[PubMed](#)]
61. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)] [[PubMed](#)]
62. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [[CrossRef](#)]
63. Emms, D.; Kelly, S. STAG: Species tree inference from all genes. *BioRxiv* **2018**, 267914. [[CrossRef](#)]
64. Emms, D.M.; Kelly, S. STRIDE: Species tree root inference from gene duplication events. *Mol. Biol. Evol.* **2017**, *34*, 3267–3278. [[CrossRef](#)]

65. Guardo, M.D.; Moretto, M.; Moser, M.; Catalano, C.; Troggio, M.; Deng, Z.; Cestaro, A.; Caruso, M.; Distefano, G.; Malfa, S.L. The haplotype-resolved reference genome of lemon (*Citrus limon* L. Burm. f.). *Tree Genet. Genom.* **2021**, *17*, 46. [[CrossRef](#)]
66. Baek, S.; Choi, K.; Kim, G.-B.; Yu, H.-J.; Cho, A.; Jang, H.; Kim, C.; Kim, H.-J.; Chang, K.S.; Kim, J.-H. Draft genome sequence of wild *Prunus yedoensis* reveals massive inter-specific hybridization between sympatric flowering cherries. *Genome Biol.* **2018**, *19*, 127. [[CrossRef](#)]
67. Chin, C.-S.; Peluso, P.; Sedlazeck, F.J.; Nattestad, M.; Concepcion, G.T.; Clum, A.; Dunn, C.; O'Malley, R.; Figueroa-Balderas, R.; Morales-Cruz, A. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **2016**, *13*, 1050–1054. [[CrossRef](#)] [[PubMed](#)]
68. Chia, Y.C.; Teh, S.-H.; Mohamed, Z. Isolation and characterization of Chalcone Isomerase (CHI) Gene from *Boesenbergia rotunda*. *S. Afr. J. Bot.* **2020**, *130*, 475–482. [[CrossRef](#)]
69. Liew, Y.J.M.; Lee, Y.K.; Khalid, N.; Abd Rahman, N.; Tan, B.C. Enhancing flavonoid production by promiscuous activity of prenyltransferase, BrPT2 from *Boesenbergia rotunda*. *PeerJ* **2020**, *8*, e9094. [[CrossRef](#)]
70. Wu, W.; Yang, Y.-L.; He, W.-M.; Rouard, M.; Li, W.-M.; Xu, M.; Roux, N.; Ge, X.-J. Whole genome sequencing of a banana wild relative *Musa itinerans* provides insights into lineage-specific diversification of the *Musa* genus. *Sci. Rep.* **2016**, *6*, 31586. [[CrossRef](#)] [[PubMed](#)]
71. D'hont, A.; Denoeud, F.; Aury, J.-M.; Baurens, F.-C.; Carreel, F.; Garsmeur, O.; Noel, B.; Bocs, S.; Droc, G.; Rouard, M. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* **2012**, *488*, 213–217. [[CrossRef](#)] [[PubMed](#)]
72. Cheng, S.-P.; Jia, K.H.; Liu, H.; Zhang, R.G.; Li, Z.C.; Zhou, S.S.; Shi, T.L.; Ma, A.C.; Yu, C.W.; Gao, C. Haplotype-resolved genome assembly and allele-specific gene expression in cultivated ginger. *Hort. Res.* **2021**, *8*, 188. [[CrossRef](#)]
73. Macas, J.; Novak, P.; Pellicer, J.; Čížková, J.; Koblížková, A.; Neumann, P.; Fukova, I.; Doležel, J.; Kelly, L.J.; Leitch, I.J. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabaeae. *PLoS ONE* **2015**, *10*, e0143424. [[CrossRef](#)]
74. McCann, J.; Macas, J.; Novák, P.; Stuessy, T.F.; Villaseñor, J.L.; Weiss-Schneeweiss, H. Differential genome size and repetitive DNA evolution in diploid species of *Melampodium* sect. *Melampodium* (Asteraceae). *Front. Plant Sci.* **2020**, *11*, 362. [[CrossRef](#)]
75. Li, H.-L.; Wu, L.; Dong, Z.; Jiang, Y.; Jiang, S.; Xing, H.; Li, Q.; Liu, G.; Tian, S.; Wu, Z. Haplotype-resolved genome of diploid ginger (*Zingiber officinale*) and its unique gingerol biosynthetic pathway. *Hort. Res.* **2021**, *8*, 189. [[CrossRef](#)]
76. Bennetzen, J.L. The structure and evolution of angiosperm nuclear genomes. *Curr. Opin. Plant Biol.* **1998**, *1*, 103–108. [[CrossRef](#)]
77. Alonso, C.; Pérez, R.; Bazaga, P.; Herrera, C.M. Global DNA cytosine methylation as an evolving trait: Phylogenetic signal and correlated evolution with genome size in angiosperms. *Front. Genet.* **2015**, *6*, 4. [[CrossRef](#)] [[PubMed](#)]
78. Fedoroff, N.V. Transposable elements, epigenetics, and genome evolution. *Science* **2012**, *338*, 758–767. [[CrossRef](#)] [[PubMed](#)]
79. Rabinowicz, P.D.; Citek, R.; Budiman, M.A.; Nunberg, A.; Bedell, J.A.; Lakey, N.; O'Shaughnessy, A.L.; Nascimento, L.U.; McCombie, W.R.; Martienssen, R.A. Differential methylation of genes and repeats in land plants. *Genome Res.* **2005**, *15*, 1431–1440. [[CrossRef](#)] [[PubMed](#)]
80. Rabinowicz, P.D.; Palmer, L.E.; May, B.P.; Hemann, M.T.; Lowe, S.W.; McCombie, W.R.; Martienssen, R.A. Genes and transposons are differentially methylated in plants, but not in mammals. *Genome Res.* **2003**, *13*, 2658–2664. [[CrossRef](#)] [[PubMed](#)]
81. Vaughn, M.W.; Tanurdžić, M.; Lippman, Z.; Jiang, H.; Carrasquillo, R.; Rabinowicz, P.D.; Dedhia, N.; McCombie, W.R.; Agier, N.; Bulski, A. Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol.* **2007**, *5*, e174. [[CrossRef](#)]
82. Crisp, P.A.; Marand, A.P.; Noshay, J.M.; Zhou, P.; Lu, Z.; Schmitz, R.J.; Springer, N.M. Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 23991–24000. [[CrossRef](#)]
83. Aya, K.; Kobayashi, M.; Tanaka, J.; Ohyanagi, H.; Suzuki, T.; Yano, K.; Takano, T.; Yano, K.; Matsuoka, M.J.P.; Physiology, C. De novo transcriptome assembly of a fern, *Lygodium japonicum*, and a web resource database, Ljtrans DB. *Plant Cell Physiol.* **2015**, *56*, e5. [[CrossRef](#)]
84. Karim, R.; Tan, Y.S.; Singh, P.; Khalid, N.; Harikrishna, J.A. Expression and DNA methylation of SERK, BBM, LEC2 and WUS genes in in vitro cultures of *Boesenbergia rotunda* (L.) Mansf. *Physiol. Mol. Biol. Plants* **2018**, *24*, 741–751. [[CrossRef](#)]
85. Karim, R.; Tan, Y.S.; Singh, P.; Nuruzzaman, M.; Khalid, N.; Harikrishna, J.A. Expression and DNA methylation of MET1, CMT3 and DRM2 during in vitro culture of *Boesenbergia rotunda* (L.) Mansf. *Philipp. Agric. Sci.* **2018**, *101*, 261–270.
86. Biswas, M.K.; Bagchi, M.; Nath, U.K.; Biswas, D.; Natarajan, S.; Jesse, D.M.I.; Park, J.-I.; Nou, I.-S. Transcriptome wide SSR discovery cross-taxa transferability and development of marker database for studying genetic diversity population structure of *Lilium* species. *Sci. Rep.* **2020**, *10*, 18621. [[CrossRef](#)]
87. Biswas, M.K.; Darbar, J.N.; Borrell, J.S.; Bagchi, M.; Biswas, D.; Nuraga, G.W.; Demissew, S.; Wilkin, P.; Schwarzacher, T.; Heslop-Harrison, J. The landscape of microsatellites in the enset (*Ensete ventricosum*) genome and web-based marker resource development. *Sci. Rep.* **2020**, *10*, 15312. [[CrossRef](#)] [[PubMed](#)]
88. Devi, K.D.; Punyarani, K.; Singh, N.S.; Devi, H.S. An efficient protocol for total DNA extraction from the members of order Zingiberales-suitable for diverse PCR based downstream applications. *SpringerPlus* **2013**, *2*, 669. [[CrossRef](#)] [[PubMed](#)]
89. Xie, M.; Chung, C.Y.-L.; Li, M.-W.; Wong, F.-L.; Wang, X.; Liu, A.; Wang, Z.; Leung, A.K.-Y.; Wong, T.-H.; Tong, S.-W. A reference-grade wild soybean genome. *Nat. Commun.* **2019**, *10*, 1216. [[CrossRef](#)] [[PubMed](#)]
90. Kiefer, E.; Heller, W.; Ernst, D. A simple and efficient protocol for isolation of functional RNA from plant tissues rich in secondary metabolites. *Plant Mol. Biol. Rep.* **2000**, *18*, 33–39. [[CrossRef](#)]
91. Schwarzacher, T.; Heslop-Harrison, P. *Practical In Situ Hybridization*; BIOS Scientific Publishers Ltd.: Oxford, UK, 2000; p. 216.

92. Gerlach, W.; Bedbrook, J. Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucleic Acids Res.* **1979**, *7*, 1869–1885. [[CrossRef](#)]
93. Yan, L.; Wang, X.; Liu, H.; Tian, Y.; Lian, J.; Yang, R.; Hao, S.; Wang, X.; Yang, S.; Li, Q. The genome of *Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb. *Mol. Plant* **2015**, *8*, 922–934. [[CrossRef](#)]
94. Lin, E.; Zhuang, H.; Yu, J.; Liu, X.; Huang, H.; Zhu, M.; Tong, Z. Genome survey of Chinese fir (*Cunninghamia lanceolata*): Identification of genomic SSRs and demonstration of their utility in genetic diversity analysis. *Sci. Rep.* **2020**, *10*, 4698. [[CrossRef](#)]
95. Marçais, G.; Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **2011**, *27*, 764–770. [[CrossRef](#)]
96. Ranallo-Benavidez, T.R.; Jaron, K.S.; Schatz, M.C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **2020**, *11*, 1432. [[CrossRef](#)]
97. Pendleton, M.; Sebra, R.; Pang, A.W.C.; Ummat, A.; Franzen, O.; Rausch, T.; Stütz, A.M.; Stedman, W.; Anantharaman, T.; Hastie, A. Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat. Methods* **2015**, *12*, 780. [[CrossRef](#)]
98. Chin, C.-S.; Alexander, D.H.; Marks, P.; Klammer, A.A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E.E. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **2013**, *10*, 563. [[CrossRef](#)] [[PubMed](#)]
99. Liu, H.; Wu, S.; Li, A.; Ruan, J. SMARTdenovo: A de novo assembler using long noisy reads. *Gigabyte* **2021**, *2021*, 1–9. [[CrossRef](#)]
100. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [[CrossRef](#)]
101. Boetzer, M.; Henkel, C.V.; Jansen, H.J.; Butler, D.; Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **2011**, *27*, 578–579. [[CrossRef](#)] [[PubMed](#)]
102. English, A.C.; Richards, S.; Han, Y.; Wang, M.; Vee, V.; Qu, J.; Qin, X.; Muzny, D.M.; Reid, J.G.; Worley, K.C. Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS ONE* **2012**, *7*, e47768.
103. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
104. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)]
105. Bao, W.; Kojima, K.K.; Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **2015**, *6*, 11. [[CrossRef](#)]
106. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [[CrossRef](#)]
107. Stanke, M.; Diekhans, M.; Baertsch, R.; Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **2008**, *24*, 637–644. [[CrossRef](#)]
108. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **2006**, *34* (Suppl. S2), W435–W439. [[CrossRef](#)] [[PubMed](#)]
109. Burge, C.; Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **1997**, *268*, 78–94. [[CrossRef](#)] [[PubMed](#)]
110. Korf, I. Gene finding in novel genomes. *BMC Bioinform.* **2004**, *5*, 59. [[CrossRef](#)] [[PubMed](#)]
111. Cantarel, B.L.; Korf, I.; Robb, S.M.; Parra, G.; Ross, E.; Moore, B.; Holt, C.; Alvarado, A.S.; Yandell, M. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **2008**, *18*, 188–196. [[CrossRef](#)]
112. Deng, Y.; Li, J.; Wu, S.; Zhu, Y.; Chen, Y.; He, F.; Chen, Y.; Deng, L.Y.; LI, J.; WU, S. Integrated nr database in protein annotation system and its localization. *Comput. Eng.* **2006**, *32*, 71–74.
113. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [[CrossRef](#)]
114. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)]
115. Kanehisa, M. The KEGG database. In *'In Silico' Simulation of Biological Processes*; Bock, G., Goode, J.A., Eds.; Novartis Foundation Symposium; John Wiley & Sons Ltd.: London, UK, 2002; Volume 247, pp. 91–100.
116. Bairoch, A.; Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **2000**, *28*, 45–48. [[CrossRef](#)]
117. Tatusov, R.L.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Kiryutin, B.; Koonin, E.V.; Krylov, D.M.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N. The COG database: An updated version includes eukaryotes. *BMC Bioinform.* **2003**, *4*, 41. [[CrossRef](#)]
118. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [[CrossRef](#)]
119. Kalvari, I.; Argasinska, J.; Quinones-Olvera, N.; Nawrocki, E.P.; Rivas, E.; Eddy, S.R.; Bateman, A.; Finn, R.D.; Petrov, A.I. Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **2018**, *46*, D335–D342. [[CrossRef](#)] [[PubMed](#)]
120. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **2013**, *29*, 2933–2935. [[CrossRef](#)] [[PubMed](#)]
121. Kumar, S.; Stecher, G.; Li, M.; Niyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
122. Conway, J.R.; Lex, A.; Gehlenborg, N. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **2017**, *33*, 2938–2940. [[CrossRef](#)]

123. Mendes, F.K.; Vanderpool, D.; Fulton, B.; Hahn, M.W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **2021**, *36*, 5516–5518. [[CrossRef](#)] [[PubMed](#)]
124. Britton, T.; Anderson, C.L.; Jacquet, D.; Lundqvist, S.; Bremer, K. Estimating divergence times in large phylogenetic trees. *Syst. Biol.* **2007**, *56*, 741–752. [[CrossRef](#)]
125. Jeremy, S.; Dan, R.; Kerrie, B.; Susan, L.; Miranda, H.-S.; Kathleen, L.; Hope, T.; Jane, G.; Neil, M.; Naxin, H. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **2010**, *463*, 7282.
126. Vandepoele, K.; Saeys, Y.; Simillion, C.; Raes, J.; Van de Peer, Y. The automatic detection of homologous regions (ADHoRe) and its application to microcolinearity between Arabidopsis and rice. *Genome Res.* **2002**, *12*, 1792–1801. [[CrossRef](#)]
127. Krueger, F. Trim Galore: A Wrapper Tool around Cutadapt and FastQC to Consistently Apply Quality and Adapter Trimming to FastQ Files. Available online: <https://github.com/FelixKrueger/TrimGalore> (accessed on 1 June 2020).
128. Krueger, F.; Andrews, S.R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **2011**, *27*, 1571–1572. [[CrossRef](#)]
129. Song, Q.; Decato, B.; Hong, E.E.; Zhou, M.; Fang, F.; Qu, J.; Garvin, T.; Kessler, M.; Zhou, J.; Smith, A.D. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS ONE* **2013**, *8*, e81148. [[CrossRef](#)]
130. Schulz, M.H.; Zerbino, D.R.; Vingron, M.; Birney, E. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **2012**, *28*, 1086–1092. [[CrossRef](#)] [[PubMed](#)]
131. Simpson, J.T.; Wong, K.; Jackman, S.D.; Schein, J.E.; Jones, S.J.; Birol, I. ABySS: A parallel assembler for short read sequence data. *Genome Res.* **2009**, *19*, 1117–1123. [[CrossRef](#)]
132. Xie, Y.; Wu, G.; Tang, J.; Luo, R.; Patterson, J.; Liu, S.; Huang, W.; He, G.; Gu, S.; Li, S. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **2014**, *30*, 1660–1666. [[CrossRef](#)] [[PubMed](#)]
133. Henschel, R.; Lieber, M.; Wu, L.-S.; Nista, P.M.; Haas, B.J.; LeDuc, R.D. Trinity RNA-Seq Assembler Performance Optimization. In Proceedings of the XSEDE '12 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the Extreme to the Campus and Beyond, Chicago, IL, USA, 16–20 July 2012; ACM: Chicago, IL, USA, 2012; p. 45.
134. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [[CrossRef](#)] [[PubMed](#)]
135. Pertea, G.; Huang, X.; Liang, F.; Antonescu, V.; Sultana, R.; Karamycheva, S.; Lee, Y.; White, J.; Cheung, F.; Parvizi, B. TIGR Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. *Bioinformatics* **2003**, *19*, 651–652. [[CrossRef](#)] [[PubMed](#)]
136. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [[CrossRef](#)]
137. Götz, S.; García-Gómez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **2008**, *36*, 3420–3435. [[CrossRef](#)]
138. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)]
139. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)]
140. Biswas, M.; Nath, U.; Howlader, J.; Bagchi, M.; Natarajan, S.; Kayum, M.A.; Kim, H.-T.; Park, J.-I.; Kang, J.-G.; Nou, I.-S. Exploration and exploitation of novel SSR markers for candidate transcription factor genes in Liliaceae species. *Genes* **2018**, *9*, 97. [[CrossRef](#)]
141. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [[CrossRef](#)] [[PubMed](#)]
142. Schuler, G.D. Sequence mapping by electronic PCR. *Genome Res.* **1997**, *7*, 541–550. [[CrossRef](#)] [[PubMed](#)]