The complete chloroplast genome of *Diplodiscus trichospermus* and phylogenetic position of Brownlowioideae within Malvaceae

Mingsong Wu[1], Liu He[1], Guangyao Ma[1], Kai Zhang[2*], Haijian Yang[1] and Xinquan Yang[1*]

Abstract

Background Malvaceae is an economically important plant family of 4,225 species in nine subfamilies. Phylogenetic relationships among the nine subfamilies have always been controversial, especially for Brownlowioideae, whose phylogenetic position remains largely unknown due to the lack of samples in previous analysis datasets. To greatly clarify the phylogenetic relationship of Malvaceae, we newly sequenced and assembled the plastome of *Diplodiscus trichospermus* taxonomically located in Brownlowioideae, and downloaded the allied genomes from public database to build a dataset covering all subfamily members of Malvaceae.

Results The annotation results showed that the plastome of *Diplodiscus trichospermus* has a typical quadripartite structure, comprising 112 unique genes, namely 78 protein-coding genes, 30 tRNA genes and 4 rRNA genes. The total length was 158,570 bp with 37.2% GC content. Based on the maximum likelihood method and Bayesian inference, a robust phylogenetic backbone of Malvaceae was reconstructed. The topology showed that Malvaceae was divided distinctly into two major branches which were previously recognized as Byttneriina and Malvadendrina. In the Malvadendrina clade, Malvoideae and Bombacoideae formed, as always, a close sister clade named as Malvatheca. Subfamily Helicteroideae occupied the most basal position and was followed by Sterculioideae which was sister to the alliance of Malvatheca, Brownlowioideae, Dombeyoideae, and Tilioideae. Brownlowioideae together with the clade comprising Dombeyoideae and Tilioideae formed a sister clade to Malvatheca. In addition, one specific conservation SSR and three specific palindrome sequences were observed in Brownlowioideae.

Conclusions In this study, the phylogenetic framework of subfamilies in Malvaceae has been resolved clearly based on plastomes, which may contribute to a better understanding of the classification and plastome evolution for Malvaceae.

Keywords Plastome, Phylogeny, Malvales, RSCU, SSRs, Palindromic sequence

*Correspondence:
Kai Zhang
k_zhang@foxmail.com
Xinquan Yang
yangxinquan@sina.com
Full list of author information is available at the end of the article

Wu *et al. BMC Genomics*   (2023) 24:571

Page 2 of 15

Background

Comprehensive and robust phylogenetic trees can advance our understanding of life origin, species differentiation and evolutionary process [1]. Complete chloroplast (cp) genomes usually inherited maternally have a moderate nucleotide substitution rate [2], and provide variation-rich nucleotide sequences compared with a few plastid or nuclear DNA markers, which have been widely used for plant phylogeny reconstruction, estimating divergence and generating genetic markers in recent years [1, 3–7]. Predominantly, the whole cp genomes contain 5–130 genes and its size ranges from 11 kb [8] to 240 kb (Accession: NC_031206 unpublished) in land plants, and generally exhibits a typical quadripartite structure consisting of two inverted repeats (IRs), one small single copy (SSC) and one large single copy (LSC) [9].

Malvaceae, the largest family in Malvales, incorporates the former separate families Malvaceae s.s., Bombacaceae, Sterculiaceae, and Tiliaceae [10]. It comprises 4,225 species in 244 genera in nine subfamilies [10, 11]. They are distributed more abundantly in tropical and subtropical regions, and were found all over the world except from the Arctic, the Antarctic, and the Gobi Desert [11]. Malvaceae, several members of which are widely used in agriculture, forestry, and horticulture, is an economically important plant family within the order Malvales in rosids. The economic importance includes herbal medicines [12, 13], fibers [14], gums [15, 16], fruits [17], vegetables [18–22], oils [23], beverages [14], timbers [24, 25], and numerous ornamental cultivars [26].

In traditional circumscription, Tiliaceae, Sterculiaceae, Bombacaceae, and Malvaceae s.s. were recognized as the "core Malvales", and the close relationship among these families was generally recognized [27, 28]. The first phylogenetic study focused on this group based on morphological features showed that only Malvaceae s.s. is likely monophyletic, and the other three families are paraphyletic or polyphyletic. Therefore, the "core Malvales" were proposed to be recognized at the familial level, i.e., Malvaceae s.l. [29]. This taxonomic treatment was supported by subsequent both morphological and molecular studies [10, 30–34]. Based on molecular studies encompassing a small number of DNA fragments, Malvaceae have been subdivided into nine subfamilies (i.e., Byttnerioideae, Grewioideae, Helicteroideae, Sterculioideae, Brownlowioideae, Dombeyoideae, Tilioideae, Bombacoideae, and Malvoideae), comprising two sister clades (i.e., Byttneriina and Malvadendrina) [10, 31]. Apart from the fact that Bombacoideae and Malvoideae together formed a well-supported clade named as Malvatheca, and Byttneriina (including Byttnerioideae and Grewioideae) formed a sister clade with the remaining Malvaceae

taxa, relationships of other subfamilies have been poorly resolved [35–41]. Until recent studies based on plastid genomes, the phylogenetic relationships among nine subfamilies have been largely improved [42–46]. However, since the datasets in their analysis lack one to three subfamily members in most cases, the relationships among several of its nine subfamilies still remain unclear. Only Cvetković et al. [47] reported the complete phylogenetic tree based on cp genomes, which exhibited a well-supported topology confirming the split of the family into Byttneriina and Malvadendrina. Defectively, in their topology, the clade including Brownlowioideae, Dombeyoideae, and Tilioideae was supported as sister clade to Malvatheca with moderate bootstrap (bs=72). Moreover, only one species belonging to Brownlowioideae was included in their dataset. Thus, further phylogenetic analysis including more species in Brownlowioideae to clarify the phylogeny within Malvaceae is necessary.

In Malvaceae, many studies have assessed the possibility that cp genomes can be used to clarify the phylogenetic relationships, or to improve topology of phylogenetic tree among its subfamilies [42–44, 46, 48]. Since the cp genome of *Gossypium hirsutum* representing the first plastome in the Malvaceae was reported in 2006 [49], complete cp genome of a large number of species in Malvaceae were sequenced. Up to Oct 21, 2022, a total of 296 records of complete cp genome were retrieved from Genbank, including 132 species from nine subfamilies, but only one complete cp genome of Brownlowioideae has been retrieved. Here, we newly sequenced and assembled the plastome of *Diplodiscus trichospermus* recognized in Brownlowioideae and downloaded the allied genomes from public database to construct the dataset representing all subfamilies of Malvaceae, aiming to greatly clarify the phylogeny of Malvaceae, especially the phylogenetic position of Brownlowioideae.

Results

Plastome structure and RSCU of *Diplodiscus trichosperma*

The complete cp genome of *Diplodiscus trichosperma* was successfully assembled and annotated (Fig. 1). Like the most species of Malvaceae, its plastome has a typical quadripartite structure [44, 50, 51], namely, the two repeat regions (IRs) are separated by a large single copy region (LSC) and a small single copy region (SSC). The total length was 158,570 bp with 37.2% GC content. 112 unique genes were found in complete cp genome of *Diplodiscus trichosperma*, including 78 protein-coding genes, 30 tRNA genes, 4 rRNA genes. LSC (87,808 bp), SSC (19,558 bp), and IR (25,602 bp) included 82, 13, and 17 unique genes respectively.

Codon usage bias, preferential or non-random use for synonymous codons, is a universal phenomenon

Wu *et al. BMC Genomics*   (2023) 24:571

Page 3 of 15
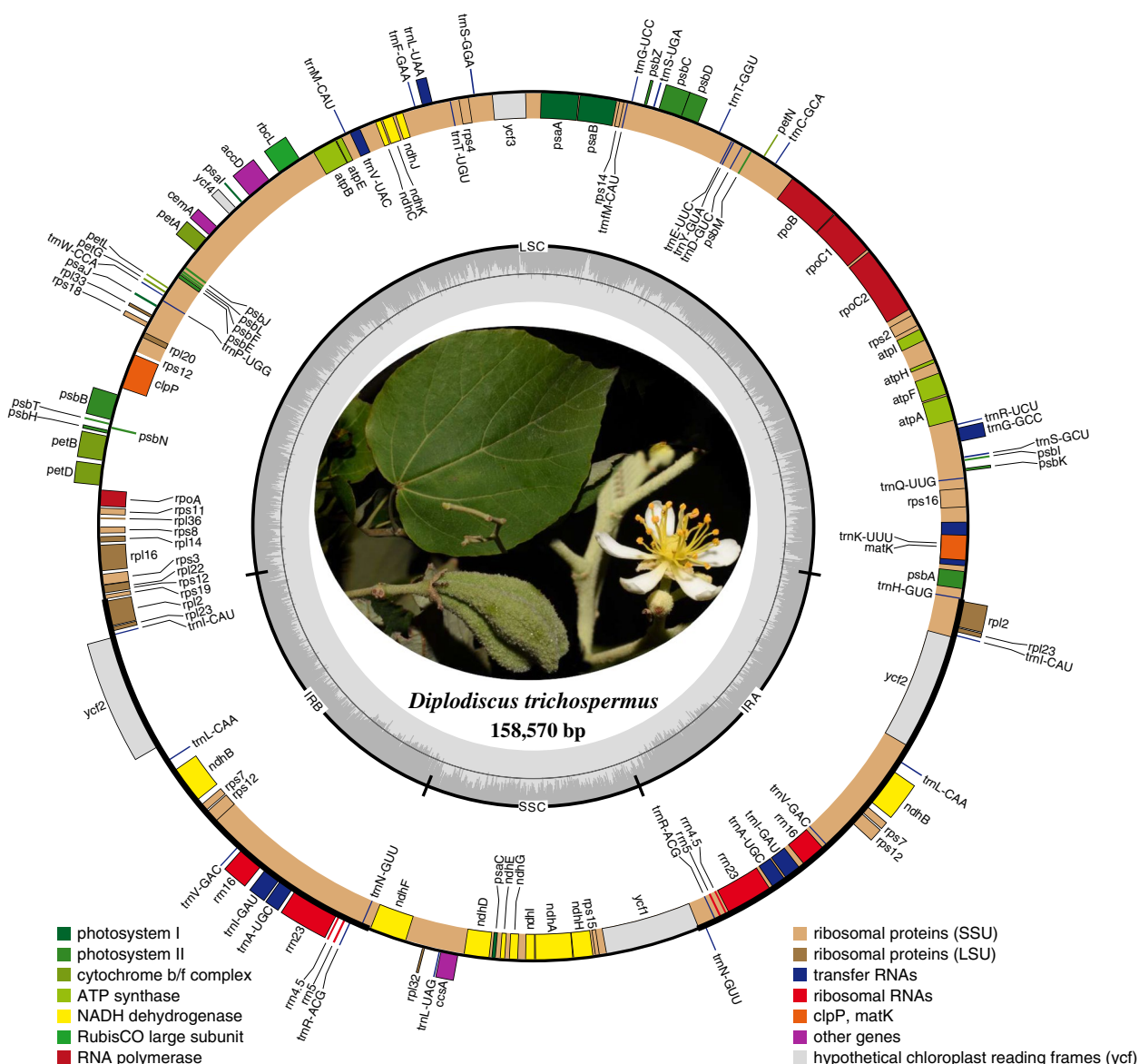
Fig. 1 Plastid genome map of *Diplodiscus trichosperma*. Genes inside the circle are transcribed clockwise, genes outside are transcribed counterclockwise. Genes are color-coded to indicate functional groups. The circle inside the GC content graph marks the 50% threshold

observed in organisms [52–54]. It is generally affected by gene mutation, natural selection, and genetic drift [55, 56]. To analyze the frequency of codon usage, a total of 78 unique coding sequences were extracted from cp genome to calculate the relative synonymous codon usage values (RSCU). All genes began with the codon AUG, except *ndh*D gene with the non-AUG start codon (GUG) (Fig. 2). 31 codons in the 78 unique CDSs have a positive bias (RSCU value > 1), where 29 are A- or U-ending codons, and the most frequent was AUU isoleucine-encoding (987 occurrences). Correspondingly, the negative bias (RSCU value < 1) was found in 33 codons

where 30 are G- or C-ending, and the least frequent codon was GUG methionine-encoding (only 1 occurrence). All stop codons were found, especially the UAA with 42 occurrences showed the strong codon usage bias (RSCU value = 1.62).

Phylogenetic analysis

To reveal the phylogenetic position of Brownlowioideae in Malvaceae, we reconstructed the Bayesian inference (BI) and maximum likelihood (ML) trees based on 148 genomes (including all genes and intergenic spacers) covering all subfamilies in Malvaceae and three outgroups.

Wu *et al. BMC Genomics*   (2023) 24:571

Page 4 of 15

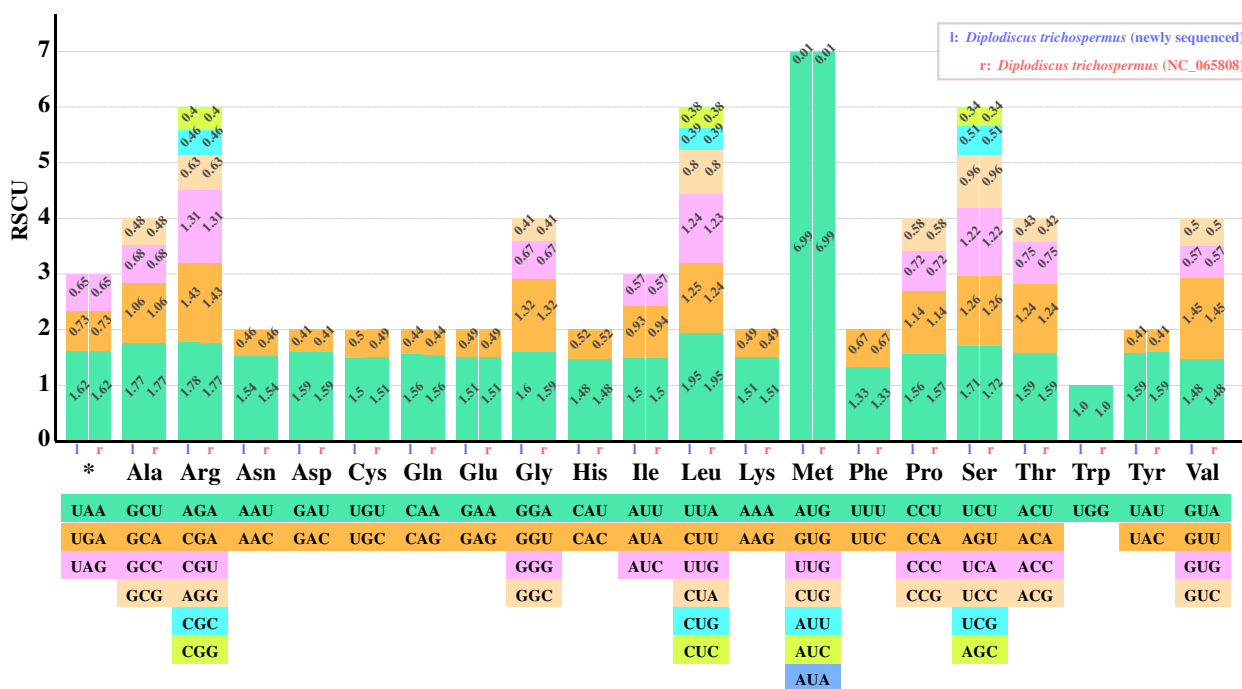**Fig. 2** The RSCU values of the 20 amino acids and one stop codon for the *Diplodiscus trichosperma* cp genome. Bar color corresponds to the codon for each amino acid

All 148 cp genome sequences have a typical quadripartite structure, and its genome size ranges from 157,936 bp to 168,953 bp. The IR length ranges from 23,726 bp to 34,496 bp. By aligning the genome to a reference sequence, we found that SSC in almost half of the 148 sequences have forward read orientation (the SSC orientation of *Malva wigandii* was designed as a reference), while the remains possess the reverse orientation. Thus, we normalized the orientation of all sequences according to LSC-IRb-SSC-IRa for subsequent phylogenetic analyses. The analysis recovered a robust phylogenetic backbone of Malvaceae, and a closer relationship between Brownlowioideae, Tilioideae, and Dombeyoideae (Fig. 3 and Additional file 1). All phylogenetic trees constructed by RAxML, IQ-TREE 2 or MrBayes have strong support values in each node of their topology, and the inferred relationships are completely congruent among these trees (Additional file 1).

Unsurprisingly, Grewioideae and Byttnerioideae formed a clade named as Byttneriina, which was a sister to the clade comprising the residual subfamilies in Malvaceae named as Malvadendrina. In the Malvadendrina clade, Malvoideae and Bombacoideae formed a close sister clade named as Malvatheca; subfamily Helicteroideae occupied the most basal position and was followed by Sterculioideae which was sister to the alliance of Malvoideae, Bombacoideae, Dombeyoideae, Tilioideae, and

Brownlowioideae; Brownlowioideae together with the clade including Dombeyoideae and Tilioideae formed a sister clade to Malvatheca. Overall, the topology we recovered is basically identical to previous analysis based on the cp genomes, except for some results based on molecular fragments where the topology of the phylogenetic tree has weak support, and the phylogenetic position of some subfamilies such as Sterculioideae and Dombeyoideae is unstable (Fig. 4).

Tree topology tests were performed on the phylogenetic trees previously considered controversial (Fig. 4). The statistical tests rejected hypotheses b (Brownlowioideae and Dombeyoideae formed a sister group and Sterculioideae was close to Malvatheca), c (Sterculioideae and Tilioideae formed a close clade which was sister to Malvatheca) and d (Dombeyoideae formed the earliest divergent clade), but failed to reject hypothesis a (Helicteroideae located at the most basal position and Brownlowioideae formed a sister to the clade comprising Tilioideae and Dombeyoideae) which possess high confidence (Table 1). The result indicated a reliable phylogenetic tree of Malvaceae where the systematic position of Brownlowioideae was resolved with great certainty.

Comparison of genome structures

To analyse the junctions of four distinct regions (LSC, IRb, SSL, and IRa) in cp genomes within Malvaceae, a
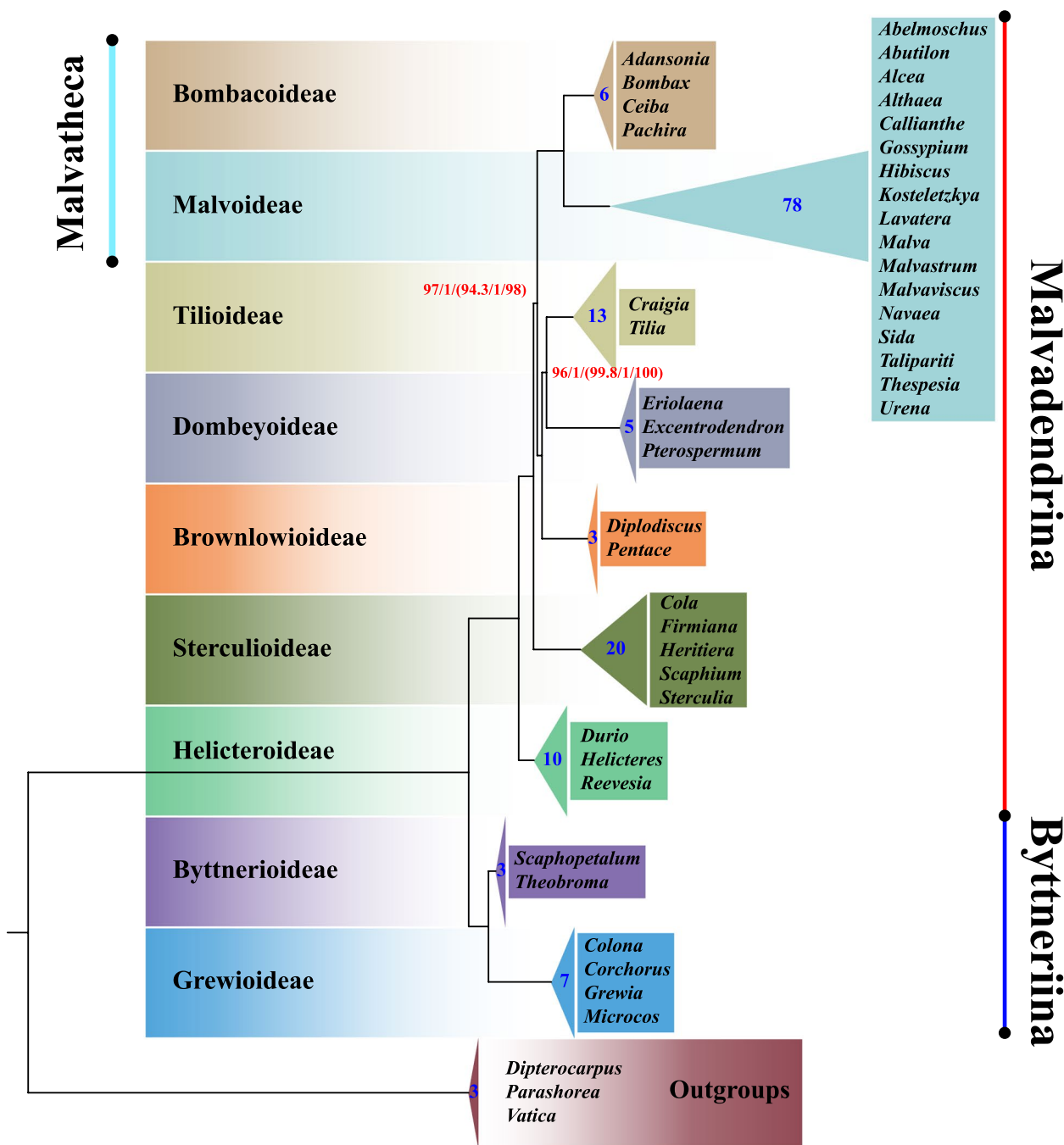
Fig. 3 The plastid phylogeny of the Malvaceae inferred from complete cp genome sequences. The numbers at each node indicate the bootstrap support (BS) / posterior probability (PP) / (SH-aLRT support / aBayes support / ultrafast bootstrap support). The unlabeled nodes indicate 100% / 1.0 / (100% / 1 / 100%) support values. Clades are color-coded according to subfamily. The blue numbers deposited in triangle shows the species number in each subfamily

dataset including the sequence newly generated in this study and another 31 species downloaded from public database was constructed. This dataset comprised all subfamily members of Malvaceae and three outgroups. We visualized the IR boundaries and the gene order of Malvaceae showed as Figs. 5 and 6 respectively. The contraction and expansion of IRs exhibited similar patterns (Fig. 5), namely, *rps*19 and *rpl*2 located in the vicinity of the LSC/IRb junctions (JLB); *trn*N and *ndh*F located in IRb/SSC (JSB); *ycf*1 and *trn*N located in SSC/IRa (JSA);
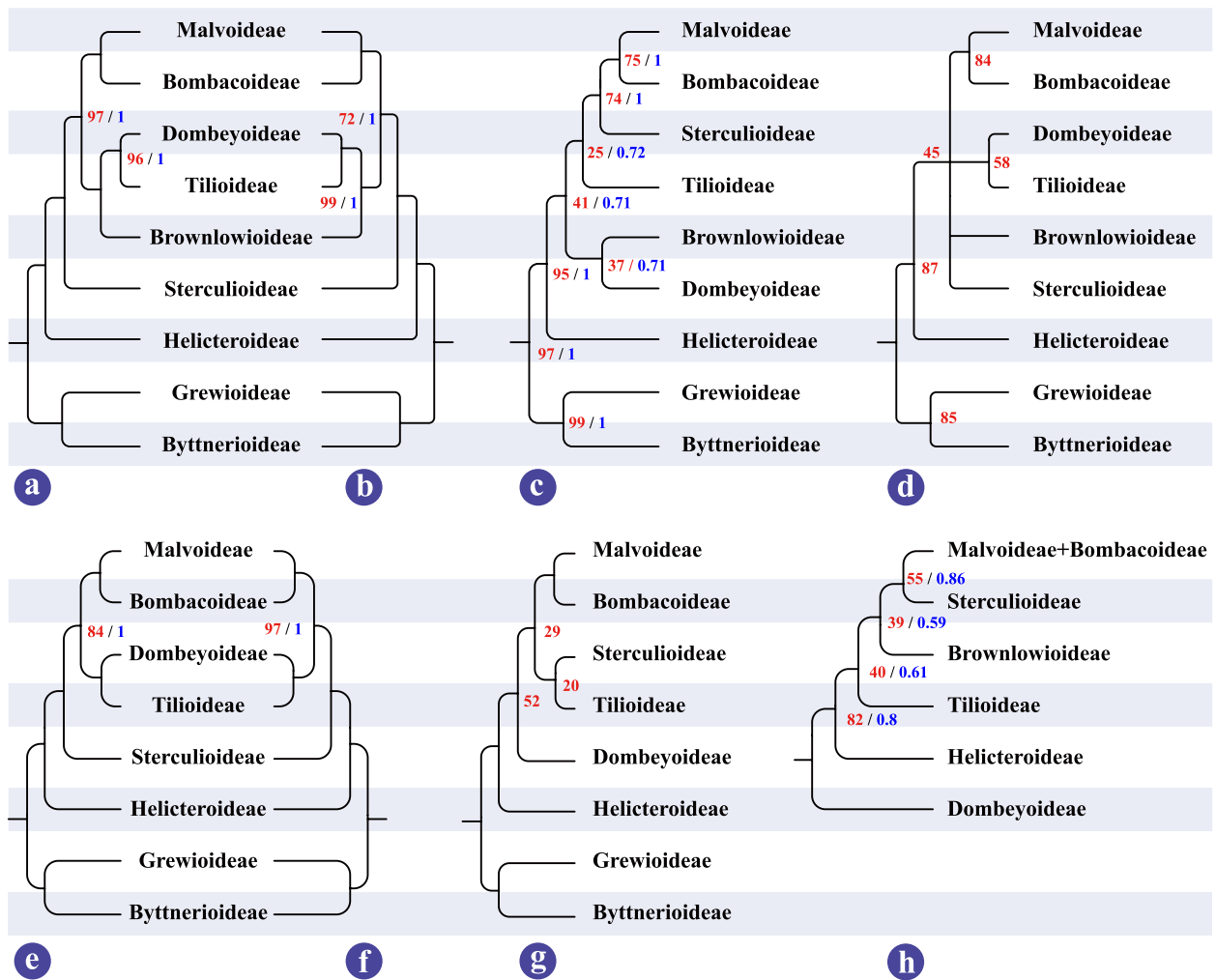
Fig. 4 Phylogenetic relationships among subfamilies of Malvaceae. The number labeled by red and blue font indicates the ML bootstrap values and BI posterior probability respectively. The unlabeled nodes indicate 100% and (or) 1.0 support values. **a** In this study. **b** From Cvetković et al. [47]. **c** From Hernández-Gutiérrez and Magallón [41]. **d** From Alverson et al. [31]. **e** From Wang et al. [46]. **f** From Li et al. [45]. **g** From Conover et al. [42]. **h** From Nyffeler et al. [37]

Table 1 Statistical tests of alternative tree topology hypotheses conducted by IQ-TREE 2

Tree	logL	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
a	-1,014,060.351	0	1+	1+	1+	1+	1+	1+	1+
b	-1,014,228.252	167.9	0-	0-	0.0001-	0-	0-	1.37e-25-	2.29e-05-
c	-1,014,230.306	169.95	0-	0-	0.0001-	0-	0-	1.16e-24-	1.06e-82-
d	-1,014,632.753	572.4	0-	0-	0-	0-	0-	4.61e-159-	0.00018-

a Helicteroideae located at the most basal position and Brownlowioideae formed a sister to the clade comprising the Tilioideae and Dombeyoideae (present study); b Brownlowioideae and Dombeyoideae formed a sister group and Sterculioideae was close to Malvatheca (structurally identical to Hernández-Gutiérrez and Magallón [41]); c Sterculioideae and Tilioideae formed a close clade which was sister to Malvatheca (structurally identical to Conover et al. [42]); d Dombeyoideae formed the earliest divergent clade (structurally identical to Nyffeler et al. [37]). a, b, c, and d corresponding to Additional file 4a, b, c, and d respectively. The value of deltaL indicates that logL differs from the maximal log1 in the comparison. bp-RELL, bootstrap proportion using RELL method [57]; p-KH, *p*-value of one-sided Kishino-Hasegawa test [58]; p-SH, *p*-value of Shimodaira-Hasegawa test [59]; c-ELW, Expected Likelihood Weight [60]. p-AU, *p*-value of approximately unbiased (AU) test [61]. Plus signs denote the 95% confidence sets. Minus signs denote significant exclusion. All test performed on 10,000 replicates using the RELL method

Wu *et al. BMC Genomics*   (2023) 24:571

Page 7 of 15

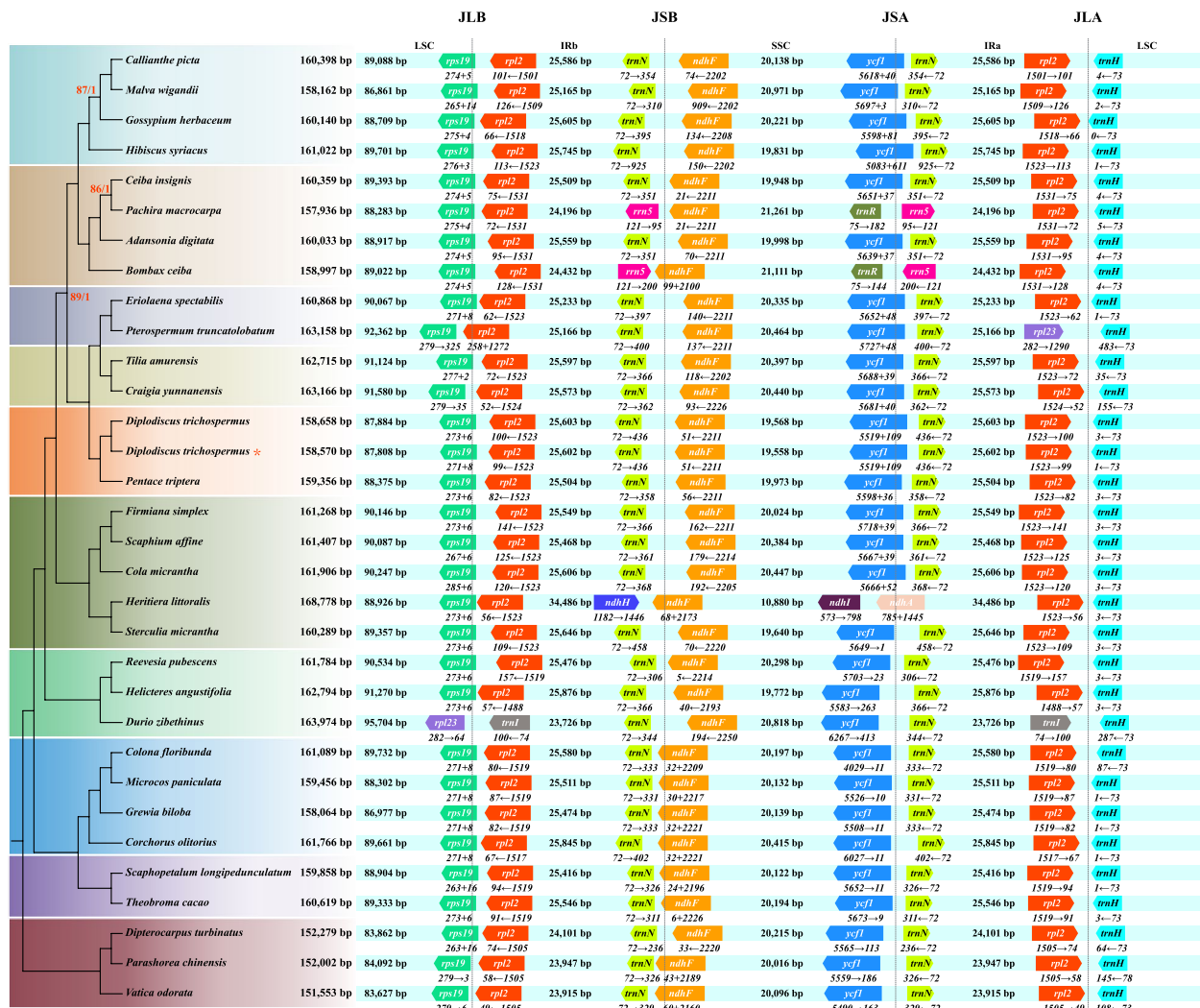| Species	Genome	LSC (JLB)	rps19	rpl2	IRb	trnN (JSB)	ndhF	SSC	ycf1 (JSA)	trnN	IRa	rpl2 (JLA)	trnH / LSC
Callianthe picta	160,398 bp	89,088 bp	rps19 274+5	rpl2 101←1501	25,586 bp	trnN 72→354	ndhF 74←2202	20,138 bp	ycf1 5618+40	trnN 354←72	25,586 bp	rpl2 1501←101	trnH 4←73
Malva wigandii	158,162 bp	86,861 bp	rps19 265+14	rpl2 126←1509	25,165 bp	trnN 72→310	ndhF 909←2202	20,971 bp	ycf1 5697+3	trnN 310←72	25,165 bp	rpl2 1509←126	trnH 2←73
Gossypium herbaceum	160,140 bp	88,709 bp	rps19 275+4	rpl2 66←1518	25,605 bp	trnN 72→395	ndhF 134←2208	20,221 bp	ycf1 5598+81	trnN 395←72	25,605 bp	rpl2 1518←66	trnH 0←73
Hibiscus syriacus	161,022 bp	89,701 bp	rps19 276+3	rpl2 113←1523	25,745 bp	trnN 72→925	ndhF 150←2202	19,831 bp	ycf1 5083+611	trnN 925←72	25,745 bp	rpl2 1523←113	trnH 1←73
Ceiba insignis	160,359 bp	89,393 bp	rps19 274+5	rpl2 75←1531	25,509 bp	trnN 72→351	ndhF 21←2211	19,948 bp	ycf1 5651+37	trnN 351←72	25,509 bp	rpl2 1531←75	trnH 4←73
Pachira macrocarpa	157,936 bp	88,283 bp	rps19 275+4	rpl2 72←1531	24,196 bp	rrn5 121→95	ndhF 21←2211	21,261 bp	trnR 75←142	rrn5 95←121	24,196 bp	rpl2 1531←72	trnH 5←73
Adansonia digitata	160,033 bp	88,917 bp	rps19 274+5	rpl2 95←1531	25,559 bp	trnN 72→351	ndhF 70←2211	19,998 bp	ycf1 5639+37	trnN 351←72	25,559 bp	rpl2 1531←95	trnH 4←73
Bombax ceiba	158,997 bp	89,022 bp	rps19 274+5	rpl2 128←1531	24,432 bp	rrn5 121→200	ndhF 99+2100	21,111 bp	trnR 75←144	rrn5 200←121	24,432 bp	rpl2 1531←128	trnH 4←73
Eriolaena spectabilis	160,868 bp	90,067 bp	rps19 271+8	rpl2 62←1523	25,233 bp	trnN 72→397	ndhF 140←2211	20,335 bp	ycf1 5652+48	trnN 397←72	25,233 bp	rpl2 1523←62	trnH 1←73
Pterospermum truncatolobatum	163,158 bp	92,362 bp	rps19 279→325	rpl2 258+1272	25,166 bp	trnN 72→400	ndhF 137←2211	20,464 bp	ycf1 5727+48	trnN 400←72	25,166 bp	rpl23 282←1290	trnH 483←73
Tilia amurensis	162,715 bp	91,124 bp	rps19 277+2	rpl2 72←1523	25,597 bp	trnN 72→366	ndhF 118←2202	20,397 bp	ycf1 5688+39	trnN 366←72	25,597 bp	rpl2 1523←72	trnH 35←73
Craigia yunnanensis	163,166 bp	91,580 bp	rps19 279→35	rpl2 52←1524	25,573 bp	trnN 72→362	ndhF 93←2226	20,440 bp	ycf1 5681+40	trnN 362←72	25,573 bp	rpl2 1524←52	trnH 155←73
Diplodiscus trichospermus	158,658 bp	87,884 bp	rps19 273+6	rpl2 100←1523	25,603 bp	trnN 72→436	ndhF 51←2211	19,568 bp	ycf1 5519+109	trnN 436←72	25,603 bp	rpl2 1523←100	trnH 3←73
Diplodiscus trichospermus ✳	158,570 bp	87,808 bp	rps19 271+8	rpl2 99←1523	25,602 bp	trnN 72→436	ndhF 51←2211	19,558 bp	ycf1 5519+109	trnN 436←72	25,602 bp	rpl2 1523←99	trnH 1←73
Pentace triptera	159,356 bp	88,375 bp	rps19 273+6	rpl2 82←1523	25,504 bp	trnN 72→358	ndhF 56←2211	19,973 bp	ycf1 5598+36	trnN 358←72	25,504 bp	rpl2 1523←82	trnH 3←73
Firmiana simplex	161,268 bp	90,146 bp	rps19 273+6	rpl2 141←1523	25,549 bp	trnN 72→366	ndhF 162←2211	20,024 bp	ycf1 5718+39	trnN 366←72	25,549 bp	rpl2 1523←141	trnH 3←73
Scaphium affine	161,407 bp	90,087 bp	rps19 267+6	rpl2 125←1523	25,468 bp	trnN 72→361	ndhF 179←2214	20,384 bp	ycf1 5667+39	trnN 361←72	25,468 bp	rpl2 1523←125	trnH 3←73
Cola micrantha	161,906 bp	90,247 bp	rps19 285+6	rpl2 120←1523	25,606 bp	trnN 72→368	ndhF 192←2205	20,447 bp	ycf1 5666+52	trnN 368←72	25,606 bp	rpl2 1523←120	trnH 3←73
Heritiera littoralis	168,778 bp	88,926 bp	ndhH 273+6	ndhF 56←1523	34,486 bp	ndhH 1182→1446	ndhF 68+2173	10,880 bp	ndhA 573→798	ndhA 785+1445	34,486 bp	rpl2 1523←56	trnH 3←73
Sterculia micrantha	160,289 bp	89,357 bp	rps19 273+6	rpl2 109←1523	25,646 bp	trnN 72→458	ndhF 70←2220	19,640 bp	ycf1 5649←1	trnN 458←72	25,646 bp	rpl2 1523←109	trnH 3←73
Reevesia pubescens	161,784 bp	90,534 bp	rps19 273+6	rpl2 157←1519	25,476 bp	trnN 72→306	ndhF 5←2214	20,298 bp	ycf1 5703→23	trnN 306←72	25,476 bp	rpl2 1519←157	trnH 3←73
Helicteres angustifolia	162,794 bp	91,270 bp	rps19 273+6	rpl2 57←1488	25,876 bp	trnN 72→366	ndhF 40←2193	19,772 bp	ycf1 5583→263	trnN 366←72	25,876 bp	rpl2 1488←57	trnH 3←73
Durio zibethinus	163,974 bp	95,704 bp	rpl23 282→64	trnI 100←74	23,726 bp	trnN 72→344	ndhF 194←2250	20,818 bp	ycf1 6267→413	trnN 344←72	23,726 bp	trnI 74←100	trnH 287←73
Colona floribunda	161,089 bp	89,732 bp	rps19 271+8	rpl2 80←1519	25,580 bp	trnN 72→333	ndhF 32+2209	20,197 bp	ycf1 4029→11	trnN 333←72	25,580 bp	rpl2 1519←80	trnH 87←73
Microcos paniculata	159,456 bp	88,302 bp	rps19 271+8	rpl2 87←1519	25,511 bp	trnN 72→331	ndhF 30+2217	20,132 bp	ycf1 5526→10	trnN 331←72	25,511 bp	rpl2 1519←87	trnH 1←73
Grewia biloba	158,064 bp	86,977 bp	rps19 271+8	rpl2 82←1519	25,474 bp	trnN 72→333	ndhF 32+2221	20,139 bp	ycf1 5508→11	trnN 333←72	25,474 bp	rpl2 1519←82	trnH 1←73
Corchorus olitorius	161,766 bp	89,661 bp	rps19 271+8	rpl2 67←1517	25,845 bp	trnN 72→402	ndhF 32+2221	20,415 bp	ycf1 6027→11	trnN 402←72	25,845 bp	rpl2 1517←67	trnH 1←73
Scaphopetalum longipedunculatum	159,858 bp	88,904 bp	rps19 263+16	rpl2 94←1519	25,416 bp	trnN 72→326	ndhF 24+2196	20,122 bp	ycf1 5652→11	trnN 326←72	25,416 bp	rpl2 1519←94	trnH 1←73
Theobroma cacao	160,619 bp	89,333 bp	rps19 273+6	rpl2 91←1519	25,546 bp	trnN 72→311	ndhF 6+2226	20,194 bp	ycf1 5673→9	trnN 311←72	25,546 bp	rpl2 1519←91	trnH 3←73
Dipterocarpus turbinatus	152,279 bp	83,862 bp	rps19 263+16	rpl2 74←1505	24,101 bp	trnN 72→236	ndhF 33←2220	20,215 bp	ycf1 5565→113	trnN 236←72	24,101 bp	rpl2 1505←74	trnH 64←73
Parashorea chinensis	152,002 bp	84,092 bp	rps19 279→3	rpl2 58←1505	23,947 bp	trnN 72→326	ndhF 43+2189	20,016 bp	ycf1 5559→186	trnN 326←72	23,947 bp	rpl2 1505←58	trnH 145←78
Vatica odorata	151,553 bp	83,627 bp	rps19 279→6	rpl2 49←1505	23,915 bp	trnN 72→329	ndhF 60+2160	20,096 bp	ycf1 5490→163	trnN 329←72	23,915 bp	rpl2 1505←49	trnH 108←73

Node support values: 87/1, 86/1, 89/1

Fig. 5 Comparison of the IR-boundaries among species of Malvaceae. The number at the tail and tip of mini arrows showed the gene length and bp distance from the corresponding junction site respectively. The sum of the numbers at both ends of the '+' is the gene length, and each number showed the bp distance from the corresponding junction site. In the phylogenetic tree on the left, clades are color-coded according to subfamily. The ML bootstrap values / BI posterior probability was marked at each node. The unlabeled nodes indicate 100% / 1.0 support values. The sequence newly generated in this study was marked by asterisk

*rpl*2 and *trn*H located in IRa/LSC (JLA). Apparently, the genes enclosing each junction site or bp distance of the genes from the junction site are not coincident with the phylogenetic tree. In subfamily Bombacoideae, although *Ceiba insignis* and *Pachira macrocarpa* showed a close relationship, their junction genes were different. In Byttnerioideae and Grewioideae, their junction patterns were completely identical, namely *rps*19 extended from LSC to IRb; the whole *rpl*2 and *trn*N were included in IRs; *ndh*F mostly existed in SSC and partially in IRb region; *ycf*1 completely located in SSC; unsurprisingly, the *trn*H gene of each cp genome presented completely in LSC region. The similar structural feature was also observed in subfamily Bombacoideae and Malvoideae, except that *ndh*F was mostly localized in SSC and *ycf*1 started from the IRa region and integrated into the SSC region. In conclusion, the junction patterns of most Malvaceae species were similar to Myrtaceae [62], Lythraceae [63], Combretaceae [64] and Brassicaceae [65] in Malvids, namely *rps*19 and *rpl*2 in JLB, *trn*N and *ndh*F in JSB, *ycf*1 and *trn*N in JSA, *rpl*2 and *trn*H in JLA.

In addition, the gene order of cp genome is highly conserved among subfamilies in Malvaceae (Fig. 6), except for partial gene order changes caused by the contraction or expansion in the IR regions or gene duplications in single-copy regions. The contraction of IR regions in
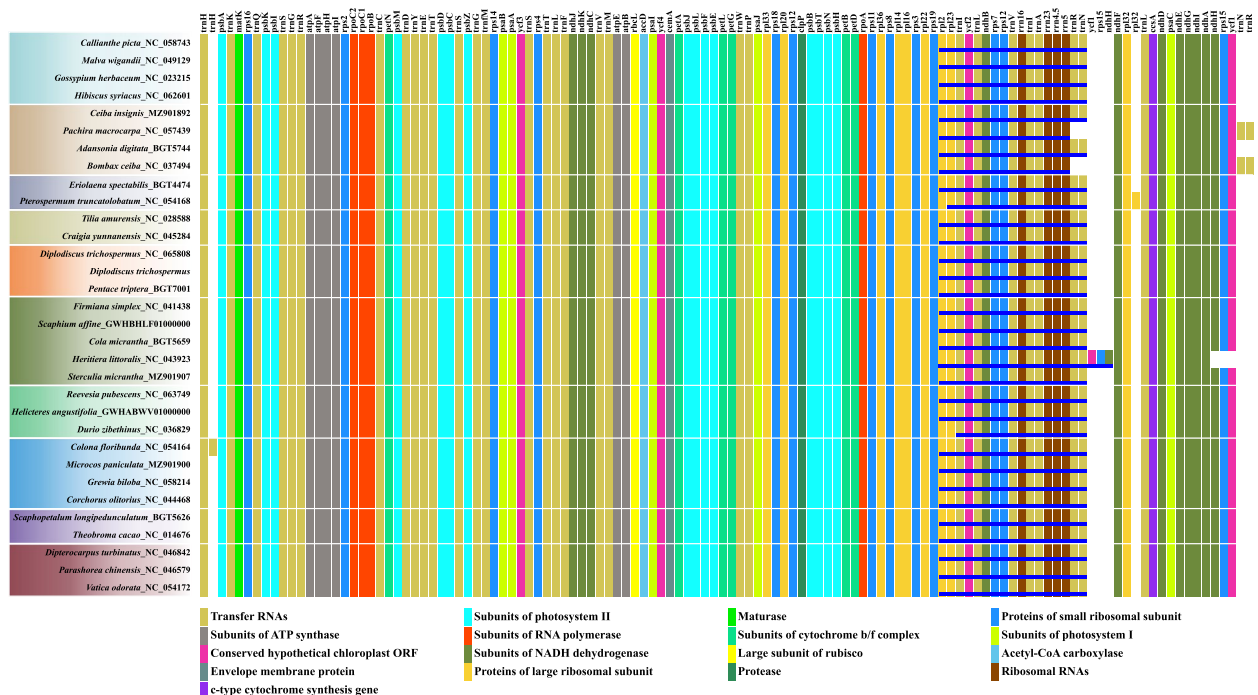
Fig. 6 The gene order map of Malvaceae cp genomes. the genes marked with a blue underline are located in IRb region. The IRa region was removed from analysis

Pachira macrocarpa and *Bombax ceiba* led to the transfer of *trn*N and *trn*R to SSC region, while the contraction of IR regions in *Durio zibethinus* and *Pterospermum truncatolobatum* resulted in the transfer of *rpl*2 and *rpl*2 together with *rpl*23 to LSC respectively. The expansion of the IR regions in *Heritiera littoralis* led to the loss of *ycf*1, *rps*15, and *ndh*H from SSC. Additionally, the duplications of *trn*H in LSC and *rpl*32 in SSC were observed in *Colona floribunda* and *Pterospermum truncatolobatum* respectively.

SSRs, palindromic sequences and conserved sequence analysis

Combining the sequences newly generated in this study and reported in public database, a total of 145 cp genomes, which covered 9 subfamilies, 42 genera, and 145 species (see Additional file 2 for detail), were used to investigate the sequence features. Almost of all subfamilies in Malvaceae have its specifical variable number of tandem repeat (VNTR) sequences, except Malvoideae where we not found complete conservation SSRs (Fig. 7). Only one conservation VNTR was observed in Grewioideae, Byttnerioideae, and Brownlowioideae respectively, while at least two sequences in other subfamilies, especially four and nine in Bombacoideae and Tilioideae respectively. The members in Malvatheca and in the clade comprising Dombeyoideae and Tilioideae

shared the VNTR sequences "TATATGGATAATATA TGGATAA" and "ACTAATGAAACTAATGAA" respectively. Byttnerioideae and Grewioideae were observed to share two conserved VNTR sequences located in the intergenic spacer between *ycf*4 gene and *cem*A gene. Furthermore, the palindrome sequences shared among subfamilies were also analysed. In Brownlowioideae, three palindrome sequences were observed in its members, two of which were situated in IRs and one in *trn*L-UAA gene. A palindrome sequence ("AGATTG CAATCT") which posited in *ndh*A gene was shared by Dombeyoideae, Tilioideae, and Brownlowioideae. The palindrome sequence ("CCGCTATAGCGG") in *rpo*B gene was observed as completely conservation in Sterculioideae. In Grewioideae, Byttnerioideae, and Dombeyoideae, one shared palindromic sequence was found in the intergenic spacer in each subfamily (Fig. 7). The clade comprising Sterculioideae, Brownlowioideae, Tilioideae, Dombeyoideae, and Malvatheca shared two palindromic sequences ("TTGATCGATCAA" and "TTTCTA GAAA") located in IRs. The clade Byttneriina shared "TTGATCATGATCAA". Interestingly, the palindromic sequence "AAAATCGATTTT" and "GAACGTTC" are lost in Malvoideae and Bombacoideae respectively. All Malvaceae members shared four palindromic sequences (Fig. 7). In addition, we found more conserved sequences in the branch nodes containing two or more closely

Wu *et al. BMC Genomics*   (2023) 24:571

Page 9 of 15
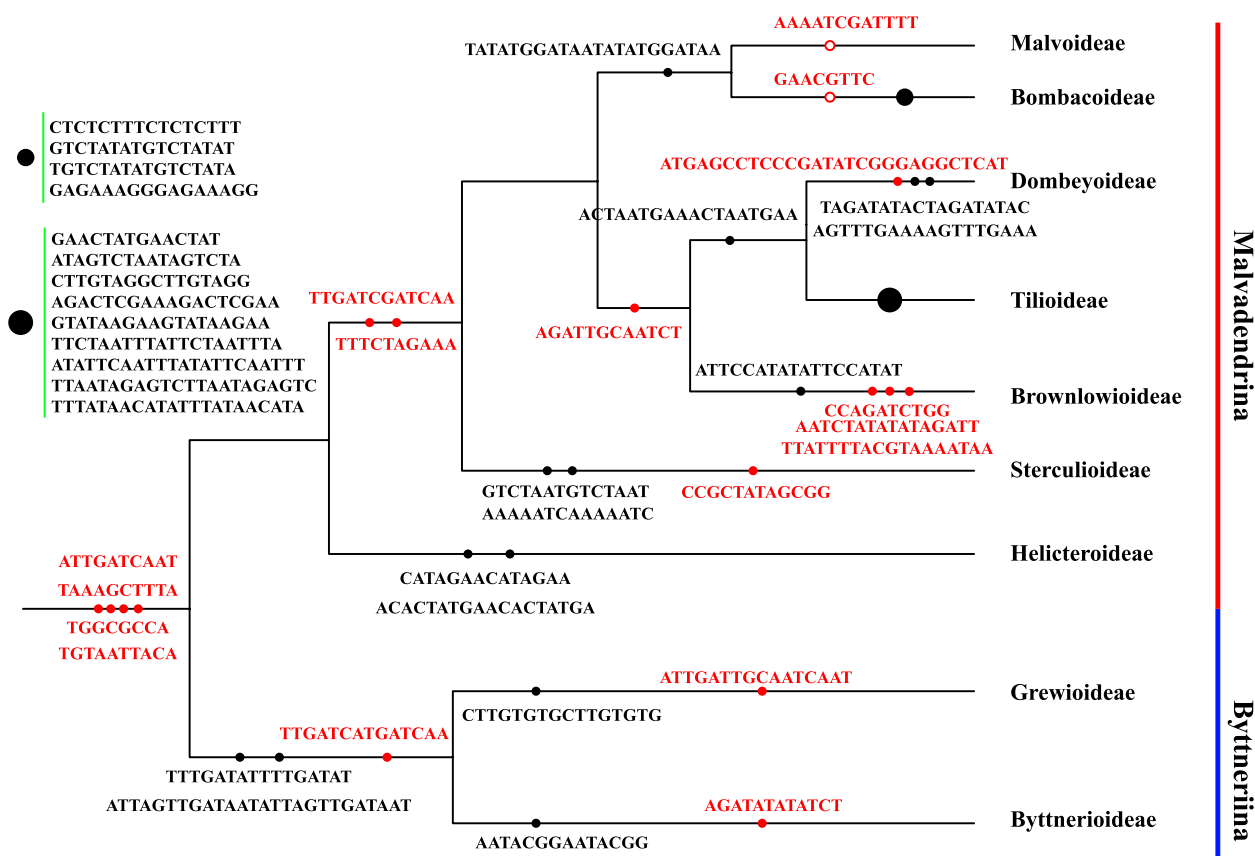
Fig. 7 The SSRs and palindromic sequences shared within or among subfamilies. The black and red sequences indicate SSRs and palindromic sequences respectively. The solid circle indicates the presence of sequence while the hollow circle indicates the absence of sequence

related subfamilies (see Additional file 3 for detail). These sequences may provide evolutionary evidence for the divergence of each subfamily.

Discussion

Chloroplast genome of Malvaceae

In this study, 145 representative taxa from 42 genera of Malvaceae and three outgroups were included in analysis dataset. These cp genomes have wide variations of SSC orientation as claimed by Cheng et al. [51], especially for genus *Hibiscus* as Fig. 8. However, whether this is the case or not, further third-generation sequencing data may be necessary to confirm it. There is no doubt that the inconsistency of the reference selection when annotating plastomes also can result in a variable read orientation [66].

The majority of cp genomes in angiosperm have conservative quadripartite structure, namely two inverted repeats were separated by one small single copy and one large single copy [67], and the genome size ranges from 11 kb [8] to 240 kb (in *Pelargonium transvaalense*, Accession: NC_031206.1). In our dataset, no exception was observed in the quadripartite cp genome structure of Malvaceae, and its genome size which ranged from 157,936 bp to 168,953 bp was also within the general size for angiosperms, suggesting that the species exhibited extremely conserved cp genome size and structure in Malvaceae.

The IR region is important to stabilize cp genome structure, and its slower nucleotide substitution rates compared with single-copy regions can enhance copy-correction activity [68]. About 10,000 bp variations of IR length of cp genome in Malvaceae indicated the noticeable genetic differences generally resulting from the contraction or expansion of the IR regions [69]. The changes in IR regions may result in a rearrangement of their gene order [70]. An obvious expansion in IRs of *Heritiera littoralis* resulted in that three genes (*ycf*1, *rps*15, and *ndh*H) general located in SSC were transferred to IR regions, while *Bombax ceiba* (Bombacoideae), *Pachira macrocarpa* (Bombacoideae), *Durio zibethinus* (Helicteroideae), and *Pterospermum truncatolobatum* (Dombeyoideae) have a contrary case where some genes general in IRs were transferred to sing-copy regions (Fig. 6). The

Wu *et al. BMC Genomics* (2023) 24:571

Page 10 of 15
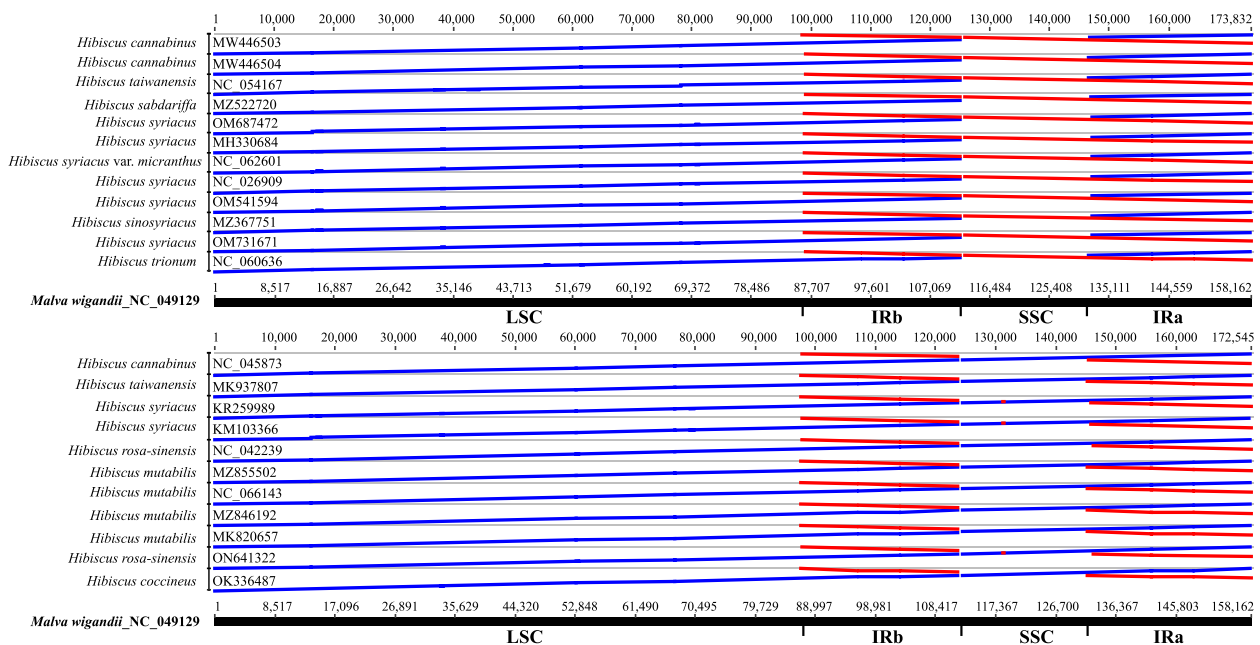
Fig. 8 The orientation of cp genomes for *Hibiscus*. *Malva wigandii* (NC_049129) was designed as the target sequence. The blue line indicates the same direction as the target sequence, while the red line indicates the opposite direction

events of IRs contraction or expansion also occurred in Rutaceae [71], Sapindaceae [72], Meliaceae [73], Onagraceae [74], and Thymelaeaceae, which all belong to the order Malvids. Especially in Thymelaeaceae, the IR length is nearly twice of it in most angiosperms, resulting in about 2–3 kb residue in SSC region which only contains the *ndh*F and *rpl*32 genes [75].

The gene duplication in cp genome is an essential source of organelle evolution, new genes, and new genetic functions [76]. The gene duplication in singlecopy regions was usually caused by the expansion of the IR regions [76], and only a few not involving the IRs have been documented in cp genomes, such as *psb*Z in *Wolffia* [77], *trn*Q-UUG in *Epimedium* and Geraniaceae [78, 79], *psb*A and *trn*T-GGU in *Pinus* [80, 81], and *psb*J in *Trachelium* [82]. Gene duplication events not involving the IRs were detected in a few Malvaceae species (*trn*H occurs in LSC of *Colona floribunda*, and *rpl*32 occurs in SSC of *Pterospermum truncatolobatum*), indicating that the cp genome of Malvaceae may be undergoing an evolution of new genes or new gene functions to further adapt to the changeable environment.

Interestingly, the above genetic events (IR regions contraction and expansion, and gene duplication) in Malvaceae are not exclusive to a single subfamily or several closely related subfamilies, but are scattered into different subfamilies. Thus, it is evident that gene losses or gains in the repeat regions or gene duplication in singlecopy regions may not indicate a phylogenetic signal at the subfamily level, which is similar to the claims of Jansen et al. [83].

Phylogenetic relationships inference

Malvaceae, which provides food, beverage, timber and traditional medicine for humans, especially most important fiber crops, is an importantly economical plant family in rosids [84]. However, the intrafamilial phylogenetic relationships are currently controversial, which may be caused by two primary reasons. One is that no sample of Brownlowioideae was included in the dataset, which thereby led to the fact that its phylogenetic position was unknown [46]; the other is that the phylogenetic trees were generally reconstructed by using one or a few loci, which resulted in different topologies with relatively low supports [37].

The plastome, general conservation, uniparental inheritance, and less prone to recombination between homologous copies [5, 49], is an ideal model for studying gene evolution and phylogenetic relationships. Compared to a limited number of DNA fragments which provide relatively little genetic variation, the whole cp genome sequences contain more integrated and adequate genetic information, and were regarded as an effective tool to investigate the phylogenetic relationships and gene evolution [4, 7, 85]. To produce a high-supported tree and clarify their phylogeny, we employed the whole cp genome sequences to evaluate the phylogenetic relationship among subfamilies in Malvaceae. Maximum

likelihood analysis and Bayesian inference recovered a strongly supported phylogenetic backbone of Malvaceae. Our newly generated phylogenetic tree (Fig. 4a) is structurally identical to the one recently reported by Cvetković, et al. [47] (Fig. 4b). There is a moderate support between Malvatheca and the clade comprising Tilioideae, Dombeyoideae, and Brownlowioideae for the phylogenetic tree recovered by Cvetković, et al., while strongly supported value (BS = 97, PP = 1) for ours. The phylogenetic tree of Malvaceae is distinctly divided into two major branches, namely Byttneriina and Malvadendrina formed a sister group without controversy [31, 41, 46]. In addition, the Byttneriina including Byttnerioideae and Grewioideae shared the largest number of conserved sequences, and these sequences can be up to 182 bp in length (see Additional file 3 for detail), which may result in a distinct divergence from Malvadendrina.

Within the Malvadendrina clade, the close relationship of the Malvoideae and Bombacoideae was firstly identified [31, 36] and had no controversy for a long time. The majority of studies on phylogeny of Malvaceae showed that Helicteroideae located at the base of Malvadendrina (Fig. 4b, c, d, e, f, g), while based on concatenation of *atp*B, *mat*K and *ndh*F or *atp*B, *trn*K-*mat*K, *ndh*F, *rbc*L and ITS showed that Dombeyoideae was the first divergent (Fig. 4h) [37, 86]. The topologies between Tilioideae, Brownlowioideae, and Sterculioideae have been largely incongruent and remain unresolved. Especially for Brownlowioideae, its phylogenetic position remains largely unknown for that no sample in Brownlowioideae was included in analysis dataset [42, 45, 46]. Alverson et al. [31] tried to recover the phylogeny of the "core Malvales" based on the *ndh*F sequences which was the first dataset including the sample of Brownlowioideae, but the relationship between Brownlowioideae, Sterculioideae, Malvatheca, and the clade comprising the Dombeyoideae and Tilioideae had not been resolved (Fig. 4d). The tree reconstructed by concatenation of *atp*B, *mat*K and *ndh*F revealed the Brownlowioideae as sister to the clade comprising Sterculioideae and Malvatheca with weak support (Fig. 4h) [37]. Furthermore, the close relationship between Brownlowioideae and Dombeyoideae was reported by Hernández-Gutiérrez and Magallón [41], but the bootstrap support value is extremely low (Fig. 4c). Up to 2021, a robust relationship between Brownlowioideae, Tilioideae, and Dombeyoideae was confirmed by Cvetković, et al. [47] based on the cp genome data where Brownlowioideae was represented by only one sequence, namely Brownlowioideae is sister group to the other two subfamilies. The identical phylogeny was also recovered by our dataset comprising three sequences in Brownlowioideae. In addition, the systematic position of Sterculioideae is generally controversial. Some studies argued it

formed as a sister to Malvatheca clade (Fig. 4c, h) [37, 41], while others supported it as a close relative to Tilioideae (Fig. 4g) [42]. Recently, cp genomic data resolved Sterculioideae as the base divergent clades after Helicteroideae in Malvadendrina [45–47], which is consistent with our study.

Conclusions

Clarifying the phylogenetic backbone of Malvaceae may contribute to exploiting the alternative food, drink, fiber, and wood resources from this economically important family and protect them better in the future. Here, we recovered a robust phylogenetic tree of "core Malvales", and revealed Brownlowioideae was sister group to Tilioideae and Dombeyoideae. The result exhibited that the cp genomic data not only can improve resolution of phylogenetic relationship among orders, families or even more genera, but also can resolve the phylogeny perfectly at subfamily level. Despite robust support values in every internode among subfamilies, more morphological synapomorphies are still required to support this phylogenetic relationship derived from cp genomes. In addition, the analysis of this study showed that the expansion or contraction of IR regions and gene duplication in single-copy regions are scattered in different subfamilies, so that they may not provide obvious phylogenetic signals at the subfamily level.

Materials and methods
Plant materials and total DNA extraction
Plant samples of *Diplodiscus trichosperma* were collected from Jianfeng Town, Ledong Li Autonomous County, Hainan Province, China (N 18.7001767, E 108.7062028). The voucher specimen (Mingsong Wu, WuMS216) was deposited in the herbarium of Sichuan University (SZ).

Total genomic DNA was extracted from young developing leaf tissues collected from the living plant, and dried immediately by silica gel using the modified CTAB method [87]. Genome skimming was conducted by Novogene Bioinformatics Technology Co. Ltd. (Tianjin, China) using next-generation sequencing technologies on the Illumina NovaSeq 6000 platform with 150 bp paired-end reads and 350 bp insert size.

Genome assembly and annotation
A total of 3.21 Gb paired-end sequencing data was generated to proceed the further analysis. The GetOrganelle pipeline [88], Bandage [89] and Plastid Genome Annotator [90] were employed to assemble the complete plastome, visualize the assemblies and annotate the genome features respectively. The cp genome of *Malva wigandii* (NC_049129) was designated as a reference for annotation. The start/stop codons, intron/exon boundaries, and

tRNA genes for the preliminary annotation result were manually adjusted by Geneious Prime 2020.1.2 (Biomatters Ltd., Auckland, New Zealand).

Relative synonymous codon usage and gene map

The relative synonymous codon usage (RSCU) of protein-coding genes was calculated and visualized using a python script written by Mingsong Wu. The circle gene maps of the plastid genes were drawn by OGDRAW [91].

Comparative analysis of genome structure

To compare the cp genome structural features of all subfamilies in Malvaceae, we downloaded all available cp genomes of Malvaceae from Genbank (https://www.ncbi.nlm.nih.gov/nuccore/), CGIR (https://ngdc.cncb.ac.cn/cgir/), and reported by Cvetković et al. [47]. A typical species was selected for each genus, with the exception of the subfamily Malvoideae, where only 4 species were selected to represent 4 of the 16 genera. A total of 32 cp genomes were employed as the analysis dataset, covering all subfamily members in the Malvaceae and three outgroups. In addition, we reannotated all sequences using the plastome of *Malva wigandii* (NC_049129) as a reference. The LasterZ plugin in Geneious was used to normalize the orientation of all sequences according to LSC-IRb-SSC-IRa. The genome sequences were aligned by MAFFT v.7.308 [92] with default parameters. The gene order and gene content adjacent to the borders of the two single copies were visualized and compared by a python script written by Mingsong Wu.

SSRs, palindromic sequences and completely conserved sequences identification

Simple sequence repeats (SSRs) grouped into four categories (i.e., P-SSRs, C-SSRs, I-SSRs, and VNTRs) were identified and localized using Krait software [93]. The default parameters were set for all SSRs analysis in this study. Palindromic sequences finder in NovoPro online tools (https://www.novoprolabs.com/tools/dna-palindrome) was used to find the palindromic sequences. The completely conserved sequences within or among the subfamilies were identified using a python script written by Mingsong Wu.

Phylogenetic inference and tree topology comparison

We employed *Pentace triptera* together with *Diplodiscus trichosperma* to represent the subfamily Brownlowioideae, and combined 142 allied genomes downloaded from public database to reconstruct the phylogenetic backbone of Malvaceae, and clarify the phylogenetic position of subfamily Brownlowioideae within "core Malves". A total of 148 genomes (see Additional file 2 for detail) were included in this dataset, and all sequences were reannotated with the reference genome. The orientation of all sequences was standardized according to LSC-IRb-SSC-IRa. Whole cp genome sequences were used to construct the data matrix and MAFFT v.7.308 [92] was employed to align the data matrix. The maximum likelihood (ML) and Bayesian inference (BI) phylogenetic trees were reconstructed using RAxML [94] and MrBayes [95] on CIPRES cluster (https://www.phylo.org/) respectively. The parameters for ML were GTR-GAMMA substitution model and 1000 bootstraps, and for BI were as follows: lset nst = 6; rates = gamma; mcmcp ngen = 1000000; relburnin = yes; burninfrac = 0.25; printfreq = 1000; samplefreq = 1000; nchains = 4; savebrlens = yes; other settings = default. IQ-TREE 2 [96] was employed to infer another ML tree and performed the SH-aLRT test, aBayes test, as well as ultrafast bootstrap test with 10,000 replicates. The analyses were run with the command "iqtree2 -s inputfile.phy -m MFP −abayes −alrt 10000 -B 10000 -T AUTO".

The tree topologies were generated using TreeGraph2 [97] to reflect the controversial hypotheses on the phylogeny of subfamilies in Malvaceae. Four aspects of Malvadendrina phylogeny that have been previously considered controversial were tested under a likelihood theory framework (see Additional file 4 for detail): (a) Helicteroideae located at the most basal position and Brownlowioideae formed a sister to the clade comprising the Tilioideae and Dombeyoideae; (b) Brownlowioideae and Dombeyoideae formed a sister group and Sterculioideae was close to Malvatheca; (c) Sterculioideae and Tilioideae formed a close clade which was sister to Malvatheca; and (d) Dombeyoideae formed the earliest divergent clade. Bootstrap proportion (BP) test [57], Kishino-Hasegawa (KH) test [58], Shimodaira-Hasegawa (SH) test [59], approximately unbiased (AU) test [61], weighted KH (WKH), weighted SH (WSH) and expected likelihood weight (ELW) [60] were performed in IQ-TREE 2 [96]. The number of RELL replicates was specified as 10,000. Probability values (p-values) of the KH, SH and AU test smaller than 0.05 indicate that the hypothesis was rejected (marked with a—sign). The command was "iqtree2 -s inputfile.phy -z inputfile.trees -n 0 -zb 10000 -zw -au".

Supplementary Information

Additional file 1. The phylogenetic trees of Malvaceae. Clades are color-coded according to subfamily. a, b and c indicate that ML tree recovered by RAxML, BI tree recovered by MrBayes and ML tree recovered by IQ-TREE 2 respectively. Numbers at each node in a and b indicate the BS and PP values respectively. Numbers at each node in c indicate SH-aLRT support/aBayes support/ultrafast bootstrap supports.

Additional file 2. The species in Malvaceae covered in dataset. The yellow background indicates the plastome of *Diplodiscus trichospermus* newly sequenced in this study. Species marked in red font are designated as outer groups.

Additional file 3. The conserved sequences shared among subfamilies. The blue color indicated the sequences located in the IRs.

Additional file 4. The hypothetical tree topologies generated by TreeGraph 2. a Helicteroideae located at the most basal position and Brownlowioideae formed a sister to the clade comprising Tilioideae and Dombeyoideae (present study); b Brownlowioideae and Dombeyoideae formed a sister group and Sterculioideae was close to Malvatheca; c Sterculioideae and Tilioideae formed a close clade which was sister to Malvatheca; d Dombeyoideae formed the earliest divergent clade.

Authors' contributions
XY, KZ, and MW conceived and designed the research. MW, LH, and GM conducted experiments and analyzed data. MW, KZ, and HY collected the sample. MW wrote the manuscript. All authors revised and approved the manuscript.

Availability of data and materials
The complete annotated sequence of *Diplodiscus trichospermus* is deposited in the NCBI database (https://www.ncbi.nlm.nih.gov/) (GenBank accession number: OP572286). The *D. trichospermus* material was obtained from Ledong, Hainan, China, and the specimen was subsequently deposited in the herbarium of Sichuan University (SZ). The other cp genomes used in this study were downloaded from the NCBI.

Declarations

Ethics approval and consent to participate
The voucher specimen (Mingsong Wu, WuMS216) of *Diplodiscus trichosperma* was collected from Jianfeng Town, Ledong Li Autonomous County, Hainan Province, China (N 18.7001767, E 108.7062028) and deposited in the herbarium of Sichuan University (SZ) under the deposition number SZ02076000. The plant specimen was identified by Kai Zhang who is engaged in research of plant taxonomy. The authors declared that we have been permitted to collect the specimen of *Diplodiscus trichospermus*. The plant materials analyzed in this study followed the Regulations on the Protection of Wild Plants of the People's Republic of China, the IUCN Policy Statement on Research Involving Species at Risk of Extinction, and the Convention on the Trade in Endangered Species of Wild Fauna and Flora.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Hainan Provincial Key Laboratory of Resources Conservation and Development of Southern Medicine, Hainan Branch of the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Haikou 570311, China. [2]Ministry of Education Key Laboratory for Ecology of Tropical Islands, College of Life Sciences, Hainan Normal University, Haikou 571158, China.

References

1. Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, et al. Origin of angiosperms and the puzzle of the Jurassic gap. Nat Plants. 2019;5:461–70.
2. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U S A. 1987;84:9054–8.
3. Barrett CF, Specht CD, Leebens-Mack J, Stevenson DW, Zomlefer WB, Davis JI. Resolving ancient radiations: Can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales)? Ann Bot. 2014;113:119–33.
4. He J, Lyu R, Luo Y, Lin L, Yao M, Xiao J, et al. An updated phylogenetic and biogeographic analysis based on genome skimming data reveals convergent evolution of shrubby habit in *Clematis* in the Pliocene and Pleistocene. Mol Phylogenet Evol. 2021;164:107259.
5. Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, et al. Plastid phylogenomic insights into relationships of all flowering plant families. BMC Biol. 2021;19:232.
6. Xi Z, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, Wurdack KJ, et al. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proc Natl Acad Sci U S A. 2012;109:17519–24.
7. Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, et al. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. New Phytol. 2017;214:1355–67.
8. Bellot S, Renner SS. The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. Genome Biol Evol. 2016;8:189–201.
9. Ravi V, Khurana JP, Tyagi AK, Khurana P. An update on chloroplast genomes. Plant Syst Evol. 2008;271:101–22.
10. Bayer C, Fay MF, De Bruijn AY, Savolainen V, Morton CM, Kubitzki K, et al. Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: a combined analysis of plastid *atp*B and *rbc*L DNA sequences. Bot J Linn Soc. 1999;129:267–303.
11. Christenhusz MJM, Fay MF, Chase MW. Plants of the world: an illustrated encyclopedia of vascular plants. Chicago: The University of Chicago Press; 2017.
12. Islam S. A review study on different plants in Malvaceae family and their medicinal uses. Am J Biomed Sci Res. 2019;3:94–7.
13. Chinese Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China, 2020 Version, Part 1. Beijing: China Medical Science Press; 2020.
14. Bayer C, Kubitzki K. Flowering plants · Dicotyledons: Malvales, Capparales, and Non-betalain Caryophyllales. In: Kubitzki K, editor. The families and genera of vascular plants, vol. 5. Berlin: Springer-Verlag; 2003. p. 225–311.
15. Brito ACF, Silva DA, de Paula RCM, Feitosa JPA. *Sterculia striata* exudate polysaccharide: Characterization, rheological properties and comparison with *Sterculia urens* (karaya) polysaccharide. Polym Int. 2004;53:1025–32.
16. Silva JSF da, Oliveira AC de J, Soares MF de LR, Soares-Sobrinho JL. Recent advances of *Sterculia* gums uses in drug delivery systems. Int J Biol Macromol. 2021;193:481–90.
17. Ketsa S, Daengkanit T. Physiological changes during postharvest ripening of durian fruit (*Durio zibethinus* Murray). J Hortic Sci Biotechnol. 1998;73:575–7.
18. Kalloo G, Bergh BO. Genetic improvement of vegetable crops. Oxford: Pergamon Press; 1992.
19. Kumar S, Dagnoko S, Haougui A, Ratnadass A, Pasternak D, Kouame C. Okra (*Abelmoschus* spp.) in west and central africa: Potential and progress on its improvement. Afr J Agric Res. 2010;5:3590–8.
20. Lim TK. Edible medicinal and non-medicinal plants: Volume 1, Fruits. New York: Springer; 2012.
21. Lim TK. Edible medicinal and non-medicinal plants: Volume 3, Fruits. New York: Springer; 2012.
22. Singh RJ. Genetic resources, chromosome engineering, and crop improvement, Volume 3: Vegetable crops. London: CRC Press; 2007.

Wu *et al. BMC Genomics*　　(2023) 24:571

Page 14 of 15

23. Raj SP, Solomon PR, Thangaraj B. Biodiesel from flowering plants. Singapore: Springer; 2022.

24. Borrega M, Ahvenainen P, Serimaa R, Gibson L. Composition and structure of balsa (*Ochroma pyramidale*) wood. Wood Sci Technol. 2015;49:403–20.

25. Ruffinatto F, Crivellaro A. Atlas of macroscopic wood identification: with a special focus on timbers used in Europe and CITES-listed species. Switzerland: Springer; 2019.

26. Datta SK, Gupta YC. Floriculture and ornamental plants. Singapore: Springer; 2022.

27. Bentham G. Notes on Malvaceae and Sterculiaceae. J Proc Linn Soc London Bot. 1862;6:97–123.

28. Cronquist A. An integrated system of classification of flowering plants. New York: Columbia University Press; 1981.

29. Judd WS, Manchester SR. Circumscription of Malvaceae (Malvales) as determined by a preliminary cladistic analysis of morphological, anatomical, palynological, and chemical characters. Brittonia. 1997;49:384–405.

30. Alverson WS, Karol KG, Baum DA, Chase MW, Swensen SM, McCourt R, et al. Circumscription of the Malvales and relationships to other Rosidae: evidence from *rbc*L sequence data. Am J Bot. 1998;85:876–87.

31. Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA. Phylogeny of the core Malvales: evidence from *ndh*F sequence data. Am J Bot. 1999;86:1474–86.

32. Bayer C. The bicolor unit - homology and transformation of an inflorescence structure unique to core Malvales. Plant Syst Evol. 1999;214:187–98.

33. Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, et al. Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. Bot J Linn Soc. 2000;133:381–461.

34. Ibrahim Z, Hassan S, ElAzab H, Badawi A. Cladistic analysis of some taxa in Malvaceae s.l. "Core Malvales" based on anatomical characteristics. Egypt J Exp Biol. 2018;14:87–105.

35. Whitlock BA, Bayer C, Baum DA. Phylogenetic relationships and floral evolution of the Byttnerioideae ("Sterculiaceae" or Malvaceae s.l.) based on sequences of the chloroplast gene, *ndh*F. Syst Bot. 2001;26:420–37.

36. Baum DA, Smith SD, Yen A, Alverson WS, Nyffeler R, Whitlock BA, et al. Phylogenetic relationships of Malvatheca (Bombacoideae and Malvoideae; Malvaceae sensu lato) as inferred from plastid DNA sequences. Am J Bot. 2004;91:1863–71.

37. Nyffeler R, Bayer C, Alverson WS, Yen A, Whitlock BA, Chase MW, et al. Phylogenetic analysis of the Malvadendrina clade (Malvaceae s.l.) based on plastid DNA sequences. Org Divers Evol. 2005;5:109–23.

38. Wilkie P, Clark A, Pennington RT, Cheek M, Bayer C, Wilcock CC. Phylogenetic relationships within the subfamily Sterculioideae (Malvaceae/Sterculiaceae-Sterculieae) using the chloroplast gene *ndh*F. Syst Bot. 2006;31:160–70.

39. Won H. Phylogenetic position of *Corchoropsis* siebold & zucc. (Malvaceae s.l.) inferred from plastid DNA sequences. J Plant Biol. 2009;52:411–6.

40. Richardson JE, Whitlock BA, Meerow AW, Madriñán S. The age of chocolate: A diversification history of *Theobroma* and Malvaceae. Front Ecol Evol. 2015;3:120.

41. Hernández-Gutiérrez R, Magallón S. The timing of Malvales evolution: Incorporating its extensive fossil record to inform about lineage diversification. Mol Phylogenet Evol. 2019;140:106606.

42. Conover JL, Karimi N, Stenz N, Ané C, Grover CE, Skema C, et al. A Malvaceae mystery: a mallow maelstrom of genome multiplications and maybe misleading methods? J Integr Plant Biol. 2018;61:12–31.

43. Abdullah. Evolutionary dynamics and phylogeny of family Malvaceae. PhD Thesis: Quaid-i-Azam University; 2020.

44. Li J, Ye GY, Liu HL, Wang ZH. Complete chloroplast genomes of three important species, *Abelmoschus moschatus*, *A. manihot* and *A. sagittifolius*: Genome structures, mutational hotspots, comparative and phylogenetic analysis in Malvaceae. PLoS One. 2020;15(11):e0242591.

45. Li R, Cai J, Yang J, Zhang Z, Li D, Yu W. Plastid phylogenomics resolving phylogenetic placement and genera phylogeny of Sterculioideae (Malvaceae s. l.). Guihaia. 2022;42:25–38.

46. Wang JH, Moore MJ, Wang H, Zhu ZX, Wang HF. Plastome evolution and phylogenetic relationships among Malvaceae subfamilies. Gene. 2021;765:145103.

47. Cvetković T, Areces-Berazain F, Hinsinger DD, Thomas DC, Wieringa JJ, Ganesan SK, et al. Phylogenomics resolves deep subfamilial relationships in Malvaceae s.l. G3 Genes|Genomes|Genetics. 2021;11:jkab136.

48. Cai J, Ma PF, Li HT, Li DZ. Complete plastid genome sequencing of four *Tilia* species (Malvaceae): A comparative analysis and phylogenetic implications. PLoS ONE. 2015;10(11):e0142705.

49. Lee SB, Kaittanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD, et al. The complete chloroplast genome sequence of *Gossypium hirsutum*: Organization and phylogenetic relationships to other angiosperms. BMC Genomics. 2006;7:61.

50. Abdullah, Shahzadi I, Mehmood F, Ali Z, Malik MS, Waseem S, et al. Comparative analyses of chloroplast genomes among three *Firmiana* species: Identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. Plant Gene. 2019;19:100199.

51. Cheng Y, Zhang L, Qi J, Zhang L. Complete chloroplast genome sequence of *Hibiscus cannabinus* and comparative analysis of the Malvaceae family. Front Genet. 2020;11:227.

52. Ma QP, Li C, Wang J, Wang Y, Ding ZT. Analysis of synonymous codon usage in *FAD*7 genes from different plant species. Genet Mol Res. 2015;14:1414–22.

53. Liu Y. A code within the genetic code: codon usage regulates co-translational protein folding. Cell Commun Signal. 2020;18:145.

54. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. Mol Biol Rep. 2022;49:539–65.

55. Ermolaeva MD. Synonymous codon usage in bacteria. Curr Issues Mol Biol. 2001;3:91–7.

56. Wong GKS, Wang J, Tao L, Tan J, Zhang J, Passey DA, et al. Compositional gradients in Gramineae genes. Genome Res. 2002;12:851–6.

57. Kishino H, Miyata T, Hasegawa M. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J Mol Evol. 1990;31:151–60.

58. Kishino H, Hasegawa M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J Mol Evol. 1989;29:170–9.

59. Shimodaira H, Hasegawa M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol. 1999;16:1114–6.

60. Strimmer K, Rambaut A. Inferring confidence sets of possibly misspecified gene trees. Proc R Soc B Biol Sci. 2002;269:137–42.

61. Shimodaira H. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 2002;51:492–508.

62. Machado L de O, Vieira L do N, Stefenon VM, Pedrosa F de O, Souza EM de, Guerra MP, et al. Phylogenomic relationship of feijoa (*Acca sellowiana* (O.Berg) Burret) with other Myrtaceae based on complete chloroplast genome sequences. Genetica. 2017;145:163–74.

63. Gu C, Ma L, Wu Z, Chen K, Wang Y. Comparative analyses of chloroplast genomes from 22 Lythraceae species: Inferences for phylogenetic relationships and genome evolution within Myrtales. BMC Plant Biol. 2019;19:281.

64. Zhang Y, Li HL, Zhong J Di, Wang Y, Yuan CC. Chloroplast genome sequences and comparative analyses of Combretaceae mangroves with related species. Biomed Res Int. 2020;2020:5867673.

65. Javaid N, Ramzan M, Khan IA, Alahmadi TA, Datta R, Fahad S, et al. The chloroplast genome of *Farsetia hamiltonii* Royle, phylogenetic analysis, and comparative study with other members of Clade C of Brassicaceae. BMC Plant Biol. 2022;22:384.

66. Gomes Pacheco T, de Santana Lopes A, Monteiro Viana GD, Nascimento da Silva O, Morais da Silva G, do Nascimento Vieira L, et al. Genetic, evolutionary and phylogenetic aspects of the plastome of annatto (*Bixa orellana* L.), the Amazonian commercial species of natural dyes. Planta. 2019;249:563–82.

67. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76:273–97.

68. Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 2016;209:1747–56.

69. Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. 2004;11:247–61.

70. Yue F, Cui L, dePamphilis CW, Moret BME, Tang J. Gene rearrangement analysis and ancestral order inference from chloroplast genomes with inverted repeat. BMC Genomics. 2008;9(Suppl 1):S25.

71. Sun K, Liu QY, Wang A, Gao YW, Zhao LC, Guan W Bin. Comparative analysis and phylogenetic implications of plastomes of five genera in subfamily Amyridoideae (Rutaceae). Forests. 2021;12:277.

72. Dong F, Lin Z, Lin J, Ming R, Zhang W. Chloroplast genome of rambutan and comparative analyses in Sapindaceae. Plants. 2021;10:283.

73. Li Y, Gu M, Lin J, Jiang H, Xiao X, Zhou W. Comparative analysis of the complete chloroplast genomes in *Toona sinensis* and *Toona ciliata*: Phylogenetic relationship of *Toona*. PREPRINT (Version 1) available at Research Square. 2022.

74. Luo Y, He J, Lyu R, Xiao J, Li W, Yao M, et al. Comparative analysis of complete chloroplast genomes of 13 species in *Epilobium*, *Circaea*, and *Chamaenerion* and insights into phylogenetic relationships of Onagraceae. Front Genet. 2021;12:730495.

75. Qian S, Zhang Y, Lee SY. Comparative analysis of complete chloroplast genome sequences in *Edgeworthia* (Thymelaeaceae) and new insights into phylogenetic relationships. Front Genet. 2021;12:643552.

76. Xiong AS, Peng RH, Zhuang J, Gao F, Zhu B, Fu XY, et al. Gene duplication, transfer, and evolution in the chloroplast genome. Biotechnol Adv. 2009;27:340–7.

77. Choi KS, Park KT, Park SJ. The chloroplast genome of *Symplocarpus renifolius*: A comparison of chloroplast genome structure in Araceae. Genes (Basel). 2017;8:324.

78. Weng ML, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol Biol Evol. 2014;31:645–59.

79. Zhang Y, Du L, Liu A, Chen J, Wu L, Hu W, et al. The complete chloroplast genome sequences of five *Epimedium* species: Lights into phylogenetic and taxonomic analyses. Front Plant Sci. 2016;7:306.

80. Lidholm J, Szmidt A, Gustafsson P. Duplication of the *psb*A gene in the chloroplast genome of two *Pinus* species. MGG Mol Gen Genet. 1991;226:345–52.

81. Kang HI, Lee HO, Lee IH, Kim IS, Lee SW, Yang TJ, et al. Complete chloroplast genome of *Pinus densiflora* Siebold & Zucc. and comparative analysis with five pine trees. Forests. 2019;10:600.

82. Haberle RC, Fourcade HM, Boore JL, Jansen RK. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. J Mol Evol. 2008;66:350–61.

83. Jansen RK, Kaittanis C, Saski C, Lee SB, Tomkins J, Alverson AJ, et al. Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: Effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. BMC Evol Biol. 2006;6:32.

84. Mitchell AS. Economic aspects of the Malvaceae in Australia. Econ Bot. 1982;36:313–22.

85. Yang JB, Tang M, Li HT, Zhang ZR, Li DZ. Complete chloroplast genome of the genus *Cymbidium*: Lights into the species identification, phylogenetic implications and population genetic analyses. BMC Evol Biol. 2013;13:84.

86. Barbosa-Silva RG, Coutinho TS, Vasconcelos S, da Silva DF, Oliveira G, Zappi DC. Preliminary placement and new records of an overlooked amazonian tree, *Christiana mennegae* (Malvaceae). PeerJ. 2021;9:e12244.

87. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11–5.

88. Jin JJ, Yu W Bin, Yang JB, Song Y, Depamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:241.

89. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: Interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31:3350–2.

90. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019;15:50.

91. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:W59–64.

92. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–780.

93. Du L, Zhang C, Liu Q, Zhang X, Yue B. Krait: An ultrafast tool for genome-wide survey of microsatellites and primer design. Bioinformatics. 2018;34:681–3.

94. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

95. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539–42.

96. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.

97. Stöver BC, Müller KF. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics. 2010;11:7.

Publisher's Note