

Two-Step Contractions of Inverted Repeat Region and *psaI* Gene Duplication from the Plastome of *Croton tiglium* (Euphorbiaceae)

Sangjin Jo and Ki-Joong Kim

School of Life Sciences, Korea University, Seoul 02841, Korea

Corresponding author

Ki-Joong Kim

Anam-ro, Seoul, 02841, Korea

Email address: kimkj@korea.ac.kr

Abstract

Croton L. (Euphorbiaceae) is a very speciose genus and consists of about 1,250 species, mainly distributed in the New World. The first complete plastome sequence from the genus, *Croton tiglium*, is reported in this study (NCBI acc. no. MH394334). The plastome is 150,021 bp in length. The lengths of LSC and SSC are 111,654 bp and 18,167 bp, respectively. However, the length of the IR region is only 10,100 bp and includes only four *rrn* and four *trn* genes, and a small part of the *ycf1* gene. We propose two-step IR contractions to explain this unique IR region of the *C. tiglium* plastome. First, the IR contracted from *rps19-rpl2* to *ycf2-trnL-CAA* on the LSC/IRb boundary. Second, the IR contracted from *ycf2-trnL-CAA* to *rrn16-trnV-GAC* on the LSC/IRa boundary. In addition, duplicated copies of *psaI* genes were discovered in the *C. tiglium* plastome. Both copies were located side by side between *accD* and *ycf4* genes, but one copy was pseudogenized because of a five-basepair (TAGCT) insertion in the middle of the gene and following frameshift mutation. The plastome contains 112 genes, of which 78 are protein-coding genes, 30 are tRNA genes, and four are rRNA genes. Sixteen genes contain one intron and two genes have two introns. The *infA* gene is lost. Twelve large repeats were detected in the plastome. All large repeats are located in the LSC region. Also, 272 simple sequence repeats (SSRs) were identified. The penta-SSRs accounted for 45% of total SSRs, followed by mono- (32%), di- (12%), tetra (6%) and tri-SSRs (5%). Most of them were distributed in the large single copy (LSC) region (85%). In addition, 76% of the SSRs were located in the intergenic spacer (IGS). Phylogenetic analysis suggested that *C. tiglium* is a sister group of *Jatropha curcas* with 100% bootstrap support. Seven Euphorbiaceae species formed one clade with 100% bootstrap support.

Introduction

The Euphorbiaceae belongs to Malpighiales and is the seventh-largest family in the flowering plants that contains 6,252 species in 209 genera (APG IV 2016; Christenhusz & Byng 2016). The Euphorbiaceae was classified into five subfamilies at first (Tokuoka 2007; Webster 1994a; Webster 1994b; Wurdack et al. 2005), but was later treated to include four subfamilies (Acalyphoideae, Cheilosoideae, Crotonoideae, and Euphorbioideae) and the subfamily Peroideae was treated as Peraceae, which is an independent family, now (Wurdack & Davis 2009; Xi et al. 2012).

The genus *Croton* is the second-largest genus after *Euphorbia* in the Euphorbiaceae, comprising more than 1,250 species (Frodin 2004; Govaerts et al. 2000). This genus belongs to the subfamily Crotonoideae (Tokuoka 2007; Webster 1994a; Webster 1994b; Wurdack et al. 2005). Webster (1993) classified *Croton* into 40 sections (Webster 1993). Van Ee et al. (2011) divided the genus *Croton* into four subgenera, *Adenophylli*, *Croton*, *Geiseleria*, and *Quadrilobi* (van Ee et al. 2011). *Croton* is distributed all over the world, ranging from America to Africa, Asia, and Oceania, but particularly many species of it are intensively distributed in the New World (van Ee & Berry 2016; van Ee et al. 2011). Various species of *Croton* are used for wound healing, rheumatism treatment, cancer treatment, or are widely cultivated as ornamental plants (Loureiro 2008; Salatino et al. 2007).

Croton tiglium L. is known as purging croton, and originated in the tropical Asia and China (Loureiro 2008). *Croton tiglium* belongs to the genus *Croton* subgenus *Croton* section *Tiglium* (van Ee et al. 2011; Webster 1993). As with other *Croton* species, *C. tiglium* is used to treat diseases such as rheumatism and cancer, and is also used as a laxative in China (Loureiro 2008; Salatino et al. 2007).

With the development of genome sequencing technology, studies on the evolution of plants using complete plastomes have been actively in progress. Currently, over 2,200 complete plastome sequences can be downloaded from the NCBI database. Among them, in the case of Malpighiales to which Euphorbiaceae belongs, the plastomes of a total of 105 species in nine families have been reported (retrieved July 10, 2018). However, Chrysobalanaceae with 50 species and Salicaceae with 41 species account for most of the foregoing species (Bardon et al. 2016; Huang et al. 2014; Huang et al. 2017). Also, the plastomes of 14 species of Passifloraceae (Cauz-Santos et al. 2017; Rabah et al. 2018), three species of Malpighiaceae (Menezes et al. 2018), one species of Clusiaceae (Jo et al. 2017), one species of Erythroxylaceae, one species of Linaceae (de Santana Lopes et al. 2018), and one species of Violaceae (Cheon et al. 2017) have been reported. In the case of Euphorbiaceae, the plastomes of six species have been reported. The six species belong to three subfamilies; Acalyphoideae (*Ricinus communis* L.) (Rivarola et al. 2011), Crotonoideae (*Hevea brasiliensis* (Willd. ex A. Juss) Müll. Arg., *Jatropha curcas* L., *Manihot esculenta* Crantz and *Vernicia fordii* (Hemsl.) Airy Shaw) (Asif et al. 2010; Daniell et al. 2008; Li et al. 2017; Tangphatsornruang et al. 2011) and Euphorbioideae (*Euphorbia esula* L.) (Horvath et al. 2018). There is no published plastome sequences for *Croton*. The plastome

structure to Euphorbiaceae species, published to date, is almost identical to that of the general angiosperm plastomes. However, a 30 kb inversion has been reported for *Hevea brasiliensis* (Tangphatsornruang et al. 2011). Given that the complete plastomes of only six species, which are around 0.1% of the 6,252 species of Euphorbiaceae, have been reported, if more species are studied hereafter, many variations may be discovered.

Therefore, this study reports for the first time the complete plastome for a *Croton* species. Through this study, it was indicated that the IR region in the plastome of *C. tiglium* was shortened by two-step contractions. In addition, we also found *psaI* gene duplication and following pseudogenization in the *C. tiglium* plastome. The amounts and distributions of the large repeats and the simple sequence repeats (SSRs) were compared among seven Euphorbiaceae plastomes. These results are expected to be used for the development of molecular identification markers for *Croton* species and for studies of Euphorbiaceae plastome evolution.

Materials and Methods

DNA extraction, Sequencing and Annotation

Leaves of *C. tiglium* used in this study were collected from the Korea University greenhouse, where we grew the plants from seeds that were originally collected in Indonesia. The plants flowered and fruited in the greenhouse. A voucher specimen was deposited in the Korea University Herbarium (KUS acc. no. 2014-0242). Fresh leaves were ground into powder in liquid nitrogen and total DNAs were extracted using the CTAB method (Doyle 1987). The DNAs were further purified by ultracentrifugation and dialysis (Palmer 1986). The genomic DNA was deposited in the Plant DNA Bank in Korea (PDBK) under the accession number PDBK2014-0242.

Approximately 200 ng of DNA were used for library construction and raw sequence reads were generated using an Illumina HiSeq 2000 system (Illumina, Inc., San Diego, CA). A total of 26,602,848 raw reads was generated. Among them, 1,396,325 plastome reads were filtered and collected using Geneious v.6.1.8 (Biomatters, Inc., Auckland, New Zealand) (Kearse et al. 2012). The plastome sequences of *Manihot esculenta* (NC_010433) were used as filtering reference. The collected plastome reads were subjected to generate a single contig to cover a whole plastome.

The IR contractions and their boundaries were confirmed by PCR method. We synthesized four following primers: *ycf2*F-ACGCAAGAAAAGCACTTCGA, *trnLR*-GAATGGGGAGTCCGCTTTGA, *trnHF*-AAGATGAAGAAGCACCCGCA and *rrn16R*-ATTCGGAAAATGGGGCTGGT. PCR amplifications were carried out using all possible combinations of primers. The PCR condition was 3 min initial denaturation at 95°C; followed by 30 cycles of 1 min annealing at 60°C, 1 min extension at 72°C and 1 min denaturation at 95°C; and 7 min final extension at 72°C. The PCR products were sequenced by Sanger method. The IR boundary were confirmed by sequence comparison.

Gene annotations were performed using the National Center for Biotechnology Information (NCBI) BLAST and tRNAscan-SE programs (Lowe & Eddy 1997). The circular map was drawn using the OGDRAW program (Lohse et al. 2013). The raw read sequence data and the completely annotated plastome sequence data were deposited on NCBI GenBank database under acc. no. SAMN09982495 and acc. no. MH394334, respectively.

Comparison of structural differences in the Euphorbiaceae plastomes

The Mauve alignment (Darling et al. 2004) of Geneious v.6.1.8 was used to compare *C. tiglium* with six Euphorbiaceae plastomes. In addition, MUSCLE v.3.8.425 alignments (Edgar 2004) with the *psaI* genes of six Euphorbiaceae were carried out to identify the *psaI* and Ψ *psaI* genes existing between *accD-ycf4* in *C. tiglium*.

Repeat Analysis in the *Croton tiglium* Plastomes

Large repeats in *C. tiglium* were analyzed using the REPuter (bibiserv2.cebitec.uni-bielefeld.de/reputer) (Kurtz et al. 2001). Forward, reverse and palindromic repeats were identified. In addition, we confirmed the large repeats in the six Euphorbiaceae plastomes. For all repeat types, the minimal size and similarity were 26 bp and 100%, respectively.

Simple sequence repeats (SSRs) were identified using Phobos v.3.3.12 (Leese et al. 2008) from Geneious v.6.1.8. The detected SSRs were five types (mono-, di-, tri-, tetra- and penta-SSRs), which had more than 10 bp (Table S1). In addition to *C. tiglium*, the SSRs were also analyzed and compared among six published Euphorbiaceae species.

Phylogenetic Analysis

In order to identify the phylogenetic position among Euphorbiaceae taxa, we reconstruct the phylogenetic tree. The whole-plastome sequences of 27 species belonging to fabids were downloaded from NCBI (Table S2). Seventy-six protein-coding and four rRNA gene sequences from plastome were used for alignments. Three plastome genes including *infA*, *rpl32*, and *rps16* were excluded from the analysis because these genes were deleted from several species. Each gene sequences were aligned using the MUSCLE v.3.8.425 program (Edgar 2004) and all gene sequences were concatenated. Maximum likelihood (ML) analysis of the concatenated sequences (79,798 bp) was conducted using RAxML v 7.7.1 (Stamatakis et al. 2008).

Results and Discussion

Structure and Gene Contents of *Croton tiglium* Complete Plastome

The plastome of *C. tiglium* was sequenced by Illumina HiSeq 2000 platform. The data was summarized in Table 1. The plastome of *C. tiglium* shows a quadripartite structure similar to the plastome structure of common flowering plants (Fig. 1). The plastome is 150,021 bp in length. In previous studies, the lengths of Euphorbiaceae plastome ranged from 160,512 bp (*Euphorbia esula*) to 163,856 bp (*Jatropha curcas*) (Fig. 2). The length of *C. tiglium* plastome was shorter by

around 10 ~ 13kb compared to them. This is because the length of the IR region was shortened and the length of the LSC region was increased (Fig. 3). Despite the length of the *C. tiglium* plastome being shorter than that of other species, the length of SSC is around 1 kb longer compared to *Euphorbia. esula* and *Jatropha curcas*.

The plastome comprises 112 unique genes (78 protein-coding genes, 30 tRNA genes and four rRNA genes). Four tRNA and four rRNA genes are duplicated in the IR regions (Fig. 1, Table 2). Protein-coding genes only appeared in the LSC and SSC regions. The 16 genes have one or two introns (Table 2). Two genes (*clpP* and *ycf3* genes) have two introns. The others have one intron. Both the *psaI* and the Ψ *psaI* are present between *accD* and *ycf4*. Ψ *psaI* was pseudogenized as TAGCT was inserted (Fig. 4). In the six previously reported Euphorbiaceae plastomes, only *psaI* exists between *accD* and *ycf4* (Asif et al. 2010; Daniell et al. 2008; Horvath et al. 2018; Li et al. 2017; Rivarola et al. 2011; Tangphatsornruang et al. 2011). *Jatropha curcas* is 9 bp longer than other species. The reason for this is that whereas the poly A of other species is 7 bp, that of *Jatropha curcas* became 6 bp, leading to an increase in the length of *psaI* (Fig. 2). The *infA* gene was lost in *C. tiglium*, which is identical to the results of previous Euphorbiaceae plastome studies.

Comparison of Plastome Structures in Euphorbiaceae

We compared the plastome structures among seven plastomes in Euphorbiaceae. Unlike six other complete plastomes, the length of the IR region of *C. tiglium* was only 10,100 bp and it included only four *rrn* and four *trn* genes, and a small part of the *ycf1* gene. The IR region in other Euphorbiaceae are usually more than 26,000 bp in length (Fig. 2). In order to explain the current short IR in the *C. tiglium* plastome, we propose a two-step contraction mechanism based on comparative analyses with other plastomes (Fig. 3). The *rps19-rrn16* region is generally located at IR in angiosperms (Kim & Lee 2004; Shinozaki et al. 1986; Yi & Kim 2012), and in other Euphorbiaceae plastomes, too (Table 3). However, it is located in the LSC region in *C. tiglium*. This is because the first IR contraction occurred in the region from *rps19-rpl2* to *ycf2-trnL-CAA* and then the second IR contraction occurred in the region from *ycf2-trnL-CAA* to *rrn16-trnV-GAC* (Figs. 1 and 3). As a result, the *rps19-trnL-CAA* region is located at the 3'-LSC and the *ycf2-rrn16* region is located at the 5'-LSC. These forms of IR contractions are reported for the first time. More studies are necessary on whether these IR contractions are a phenomenon occurring in certain taxa in the genus *Croton* or are observed in other genera as well.

The large inversion of 30 kb previously reported in the LSC region (*trnS-GCU-trnT-GGU*) of *Hevea brasiliensis* (Tangphatsornruang et al. 2011) was identified as not being present in *Croton*.

Distribution of Large and Simple Sequence Repeats

Dispersed repeat sequences in plant plastomes have been analyzed in order to understand the evolutionary roles of the sequences (Gupta & Varshney 2000; Muriira et al. 2018; Powell et al. 1995; Provan et al. 2001; Timme et al. 2007). A total of 12 large repeats were found in *C. tiglium*

plastome (Table 4). Among them, forward repeats occupied the majority with nine, and the numbers of reverse and palindromic repeats were one and two, respectively. The large repeats of seven Euphorbiaceae plastomes, including *C. tiglium*, were compared in Fig. 4. In the results, the number of large repeats of *Jatropha curcas* was the largest with 37 and the number of large repeats of *Euphorbia esula* was the smallest with eight (Fig. 5A). In all seven species, most of the large repeats were located in the LSC, and in particular, unlike other species, in the case of *C. tiglium*, all large repeats were present in the LSC (Fig. 5B). The large repeats of *Euphorbia esula* and *Hevea brasiliensis* were not present in the SSC. In the case of *Ricinus communis* no large repeat was present in the IR.

With regard to simple sequence repeats (SSRs), a total of 272 SSRs were found in the *C. tiglium* plastome. In order to compare the numbers of SSRs, the SSR numbers were also calculated from six other published Euphorbiaceae taxa. The SSR numbers are 214 in *Euphorbia esula*, 245 in *Hevea brasiliensis*, 250 in *Jatropha curcas*, 201 in *Manihot esculenta*, 283 in *Ricinus communis* and 259 in *Vernicia fordii*. The SSR numbers are similar each other among the Euphorbiaceae plastomes, but they are 2-3 times higher than the plastomes from other families (Kim & Lee 2004; Shinozaki et al. 1986; Yi & Kim 2012). Among them, the numbers of penta-SSRs and mono-SSRs were 123 and 87, respectively, and accounted for 77% of the entire SSRs (Fig. 6A). These results are in sharp contrast to previous results regarding other taxa, in which mono-SSRs were the most abundant and penta-SSRs were rarely present (Cauz-Santos et al. 2017; Gu et al. 2018; Saina et al. 2018). The largest number of SSRs was located in the LSC (232), followed by the SSC (37) and the IR (3) (Fig. 6B). This is because the length of the LSC is as long as 111,654 bp and the rate of variations is high. Unlike the length of a general angiosperm LSC, which is around 85 kb, *C. tiglium* has a larger LSC due to the two-step IR contractions, so that the distribution ratios of SSRs increased in the LSC and decreased in the IR. Most SSRs were distributed in the IGS and 34 and 32 SSRs were distributed in the intron and the CDS, respectively (Fig. 6C). This is one of the reasons why the IGS has higher sequence divergences than the CDS. These SSR loci may be usefully used in interspecific or intergroup comparative studies.

Phylogenetic Analysis

The phylogenetic position of *C. tiglium* was identified using 28 complete plastome sequences belonging to the Fabids (Fig. 7). *Croton tiglium* formed a clade with *Jatropha curcas* (BS = 100%) and *Vernicia fordii* were located as a sister group of them (BS = 100%). They share inaperturate seed traits and belong to the subfamily Crotonoideae (Tokuoka 2007; Wurdack et al. 2005). *Manihot esculenta* and *Hevea brasiliensis* form a clade (BS = 100%) in our tree. They formed a sister clade of the *Croton-Jatropha-Vernicia* clade. They also belong to the subfamily Crotonoideae but they share articulated seed traits (Tokuoka 2007; Wurdack et al. 2005). *Euphorbia esula*, which belongs to the subfamily Euphorbioideae, was located as a sister group of the *Manihot-Hevea* clade (BS = 85%). Our results indicate the subfamily Crotonoideae is

paraphyletic group because it included the subfamily Euphorbioideae. *Ricinus communis*, which belongs to Acalyphoideae, was located as a basal clade in Euphorbiaceae (BS = 84%).

The results of this study indicate that Euphorbiaceae are sister group of the Salicaceae-Passifloraceae-Violaceae-Linaceae clade. However, their relationships are uncertain because there is no information on the plastomes of Rafflesiaceae and Peraceae (Davis et al. 2007; Wurdack & Davis 2009; Xi et al. 2012), which are thought to be closely related to Euphorbiaceae. Hereafter, clear relationships will be established if the whole plastome sequences of Rafflesiaceae and Peraceae are added.

Phylogenetic studies of *Croton* have been conducted several times using partial sequences such as *matK*, *rbcL*, *trnL-F*, and ITS (Berry et al. 2005; Haber et al. 2017; van Ee & Berry 2010; van Ee et al. 2011; Wurdack et al. 2005). Phylogenetic trees made using *rbcL* had poor resolutions of *Croton*. However, monophyly was formed in the phylogenetic tree made using *trnL-F* (BS = 100%). Monophyly was also formed along with 100% support in the phylogenetic tree made using combined *rbcL* and *trnL-F* data. In *Croton*'s biogeography studies, *C. tiglium* formed a clade with *C. acutifolius*, *C. argyratus*, *C. cascarilloides*, and *C. megalobotrys* (Haber et al. 2017).

The first results on complete plastome of *Croton* obtained in this study will be useful reference sequences for the further phylogenetic and biogeographic studies.

Conclusions

In this study, we reported the complete plastome of *Croton tiglium* (Euphorbiaceae). The length of *C. tiglium* plastome was shorter by around 10 ~ 13kb compared to the plastomes of other Euphorbiaceae members. It is primarily due to the short invert repeat (IR). The length of the invert repeat (IR) region is 10,100 bp and it includes only four *rrn* and four *trn* genes, and a small part of the *ycf1* gene. In order to explain this unique IR structure, we propose two-step IR contraction model. First, the IR contracted from *rps19-rpl2* to *ycf2-trnL-CAA* on the LSC/IRb boundary. Second, the IR contracted from *ycf2-trnL-CAA* to *rrn16-trnV-GAC* on the LSC/IRa boundary. The first contraction must occur prior to the second contraction. In addition, duplicated copies of *psaI* genes were discovered in the *C. tiglium* plastome. Both copies were located side by side between *accD* and *ycf4* genes, but one copy was pseudogenized because of a five-basepair (TAGCT) insertion and following frameshift mutation in the middle of the gene.

Acknowledgments

We thank the curator of the Korea University Herbarium (KUS) for voucher specimen preparations. The genomic and chloroplast DNAs are deposited in the Plant DNA Bank in Korea (PDBK).

Additional Information and Declaration

Funding

This work was supported by the National Research Foundation of Korea (NRF). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

National Research Foundation of Korea (NRF): NRF-2015M3A9B8030588

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Sangjin Jo performed the experiments and data analysis, and wrote the materials and methods section of the manuscript.
- Ki-Joong Kim conceived the idea for the work, guided the experiment and data analysis, wrote the remaining sections of the manuscript, and wrote the final version of manuscript.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

Total DNA of *Croton tiglium* was deposited in the Plant DNA Bank in Korea (PDBK) under the accession number PDBK2014-0242.

Data Availability

The following information was supplied regarding data availability: The raw sequence data of NGS is deposited in the NCBI BioSample under the accession number SAMN09982495 and the annotated plastid genome sequences are deposited in the NCBI GenBank under the accession number MH394334, respectively.

Supplemental Information

Supplemental information for this article can be found online at

References

- APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181:1-20.
- Asif MH, Mantri SS, Sharma A, Srivastava A, Trivedi I, Gupta P, Mohanty CS, Sawant SV, and Tuli R. 2010. Complete sequence and organisation of the *Jatropha curcas* (Euphorbiaceae) chloroplast genome. *Tree Genetics & Genomes* 6:941-952.

- Bardon L, Sothers C, Prance GT, Malé P-JG, Xi Z, Davis CC, Murienne J, García-Villacorta R, Coissac E, and Lavergne S. 2016. Unraveling the biogeographical history of Chrysobalanaceae from plastid genomes. *American Journal of Botany* 103:1089-1102.
- Berry PE, Hipp AL, Wurdack KJ, Van Ee B, and Riina R. 2005. Molecular phylogenetics of the giant genus *Croton* and tribe Crotonaeae (Euphorbiaceae sensu stricto) using ITS and *trnL-trnF* DNA sequence data. *American Journal of Botany* 92:1520-1534.
- Cauz-Santos LA, Munhoz CF, Rodde N, Cauet S, Santos AA, Penha HA, Dornelas MC, Varani AM, Oliveira GC, and Bergès H. 2017. The chloroplast genome of *Passiflora edulis* (Passifloraceae) assembled from long sequence reads: Structural organization and phylogenomic studies in Malpighiales. *Frontiers in Plant Science* 8:334 doi:10.3389/fpls.2017.00334.
- Cheon K-S, Yang J-C, Kim K-A, Jang S-K, and Yoo K-O. 2017. The first complete chloroplast genome sequence from Violaceae (*Viola seoulensis*). *Mitochondrial DNA Part A* 28:67-68.
- Christenhusz MJ, and Byng JW. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa* 261:201-217.
- Daniell H, Wurdack KJ, Kanagaraj A, Lee S-B, Saski C, and Jansen RK. 2008. The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of *atpF* in Malpighiales: RNA editing and multiple losses of a group II intron. *Theoretical and Applied Genetics* 116:723-737.
- Darling AC, Mau B, Blattner FR, and Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14:1394-1403.
- Davis CC, Latvis M, Nickrent DL, Wurdack KJ, and Baum DA. 2007. Floral gigantism in Rafflesiaceae. *Science* 315:1812-1812.
- de Santana Lopes A, Pacheco TG, dos Santos KG, do Nascimento Vieira L, Guerra MP, Nodari RO, de Souza EM, de Oliveira Pedrosa F, and Rogalski M. 2018. The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. *Plant Cell Reports* 37:307-328.
- Doyle JJ. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19:11-15.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792-1797.
- Frodin DG. 2004. History and concepts of big plant genera. *Taxon* 53:753-776.
- Govaerts R, Frodin DG, Radcliffe-Smith A, and Carter S. 2000. *World checklist and bibliography of Euphorbiaceae (with Pandaceae)* Royal Botanic Gardens, Kew, UK: Kew Publishing.
- Gu C, Dong B, Xu L, Tembrock LR, Zheng S, and Wu Z. 2018. The Complete Chloroplast Genome of *Heimia myrtifolia* and Comparative Analysis within Myrtales. *Molecules* 23:846 doi:10.3390/molecules23040846.
- Gupta PK, and Varshney R. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113:163-185.
- Haber EA, Kainulainen K, Van Ee BW, Oyserman BO, and Berry PE. 2017. Phylogenetic relationships of a major diversification of *Croton* (Euphorbiaceae) in the western Indian Ocean region. *Botanical Journal of the Linnean Society* 183:532-544.

- Horvath DP, Patel S, Dođramaci M, Chao WS, Anderson JV, Foley ME, Scheffler B, Lazo G, Dorn K, and Yan C. 2018. Gene Space and Transcriptome Assemblies of Leafy Spurge (*Euphorbia esula*) Identify Promoter Sequences, Repetitive Elements, High-Quality Markers, and a Full-Length Chloroplast Genome. *Weed Science* 66:355-367.
- Huang DI, Hefer CA, Kolosova N, Douglas CJ, and Cronk QC. 2014. Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, *Populus balsamifera* and *P. trichocarpa* (Salicaceae). *New Phytologist* 204:693-703.
- Huang Y, Wang J, Yang Y, Fan C, and Chen J. 2017. Phylogenomic Analysis and Dynamic Evolution of Chloroplast Genomes in Salicaceae. *Frontiers in plant science* 8:1050 doi:10.3389/fpls.2017.01050.
- Jo S, Kim H-W, Kim Y-K, Sohn J-Y, Cheon S-H, and Kim K-J. 2017. The complete plastome of tropical fruit *Garcinia mangostana* (Clusiaceae). *Mitochondrial DNA Part B* 2:722-724.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, and Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647-1649.
- Kim K-J, and Lee H-L. 2004. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Research* 11:247-261.
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, and Giegerich R. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* 29:4633-4642.
- Leese F, Mayer C, and Held C. 2008. Isolation of microsatellites from unknown genomes using known genomes as enrichment templates. *Limnology and Oceanography: Methods* 6:412-426.
- Li Z, Long H, Zhang L, Liu Z, Cao H, Shi M, and Tan X. 2017. The complete chloroplast genome sequence of tung tree (*Vernicia fordii*): Organization and phylogenetic relationships with other angiosperms. *Scientific Reports* 7:1869 doi:10.1038/s41598-017-02076-6.
- Lohse M, Drechsel O, Kahlau S, and Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* 41:W575-W581.
- Loureiro T. 2008. *Croton* Linnaeus. In: Wu Z, Raven PH, and Hong D, eds. *Flora of China Volume 11*. Beijing, China: Science Press, 258-264.
- Lowe TM, and Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25:955-964.
- Menezes AP, Resende-Moreira LC, Buzatti RS, Nazareno AG, Carlsen M, Lobo FP, Kalapothakis E, and Lovato MB. 2018. Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Scientific Reports* 8:2210 doi:10.1038/s41598-018-20189-4.
- Muriira NG, Muchugi A, Yu A, Xu J, and Liu A. 2018. Genetic Diversity Analysis Reveals Genetic Differentiation and Strong Population Structure in *Calotropis* Plants. *Scientific Reports* 8:7832 doi:10.1038/s41598-018-26275-x.

- Palmer JD. 1986. Isolation and structural analysis of chloroplast DNA. *Methods in enzymology* 118:167-186.
- Powell W, Morgante M, McDevitt R, Vendramin G, and Rafalski J. 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proceedings of the National Academy of Sciences of the United States of America* 92:7759-7763.
- Provan J, Powell W, and Hollingsworth PM. 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in ecology & evolution* 16:142-147.
- Rabah SO, Shrestha B, Hajrah NH, Sabir MJ, Alharby HF, Sabir MJ, Alhebshi AM, Sabir JS, Gilbert LE, and Ruhlman TA. 2018. *Passiflora* plastome sequencing reveals widespread genomic rearrangements. *Journal of Systematics and Evolution* doi:10.1111/jse.12425.
- Rivarola M, Foster JT, Chan AP, Williams AL, Rice DW, Liu X, Melake-Berhan A, Creasy HH, Puiu D, and Rosovitz M. 2011. Castor bean organelle genome sequencing and worldwide genetic diversity analysis. *PloS One* 6:e21743 doi:10.1371/journal.pone.0021743.
- Saina JK, Li Z-Z, Gichira AW, and Liao Y-Y. 2018. The Complete Chloroplast Genome Sequence of Tree of Heaven (*Ailanthus altissima* (Mill.)(Sapindales: Simaroubaceae), an Important Pantropical Tree. *International Journal of Molecular Sciences* 19:929 doi:10.3390/ijms19040929.
- Salatino A, Salatino MLF, and Negri G. 2007. Traditional uses, chemistry and pharmacology of *Croton* species (Euphorbiaceae). *Journal of the Brazilian Chemical Society* 18:11-33.
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, and Yamaguchi-Shinozaki K. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO Journal* 5:2043-2049.
- Stamatakis A, Hoover P, and Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* 57:758-771.
- Tangphatsornruang S, Uthaipaisanwong P, Sangsrakru D, Chanprasert J, Yoocha T, Jomchai N, and Tragoonrung S. 2011. Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene* 475:104-112.
- Timme RE, Kuehl JV, Boore JL, and Jansen RK. 2007. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. *American Journal of Botany* 94:302-312.
- Tokuoka T. 2007. Molecular phylogenetic analysis of Euphorbiaceae sensu stricto based on plastid and nuclear DNA sequences and ovule and seed character evolution. *Journal of Plant Research* 120:511-522.
- van Ee BW, and Berry PE. 2010. Taxonomy and phylogeny of *Croton* section *Heptallon* (Euphorbiaceae). *Systematic Botany* 35:151-167.
- van Ee BW, and Berry PE. 2016. *Croton* Linnaeus. In: Flora of North America Editorial Committee, ed. *Flora of North America Volume 12*. New York, NY, USA: Oxford University Press on Demand.
- van Ee BW, Riina R, and Berry PE. 2011. A revised infrageneric classification and molecular phylogeny of New World *Croton* (Euphorbiaceae). *Taxon* 60:791-823.
- Webster GL. 1993. A provisional synopsis of the sections of the genus *Croton* (Euphorbiaceae). *Taxon* 42:793-823.

- Webster GL. 1994a. Classification of the Euphorbiaceae. *Annals of the Missouri Botanical Garden*:3-32.
- Webster GL. 1994b. Synopsis of the genera and suprageneric taxa of Euphorbiaceae. *Annals of the Missouri Botanical Garden* 81:33-144.
- Wurdack KJ, and Davis CC. 2009. Malpighiales phylogenetics: gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. *American Journal of Botany* 96:1551-1570.
- Wurdack KJ, Hoffmann P, and Chase MW. 2005. Molecular phylogenetic analysis of uniovulate Euphorbiaceae (Euphorbiaceae sensu stricto) using plastid *rbcL* and *trnL-F* DNA sequences. *American Journal of Botany* 92:1397-1420.
- Xi Z, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, Wurdack KJ, Endress PK, Matthews ML, Stevens PF, and Mathews S. 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences of the United States of America* 109:17519-17524.
- Yi D-K, and Kim K-J. 2012. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PloS One* 7:e35872 doi:10.1371/journal.pone.0035872.