Taxonomy, comparative genomics of Mullein (Verbascum, Scrophulariaceae), with implications for the evolution of Verbascum and Lamiales

Xiang Dong1,2,3†, Elijah Mbandi Mkala1,2,3,4†, Elizabeth Syowai Mutinda1,2,3†, Jia-Xin Yang1,2,3, Vincent Okelo Wanga1,2,3, Millicent Akinyi Oulo1,2,3, Victor Omondi Onjolo1,2,3, Guang-Wan Hu1,2,3* and Qing-Feng Wang1,2

Abstract

Background: The genus Verbascum L. (Scrophulariaceae) is distributed in Africa, Europe, and parts of Asia, with the Mediterranean having the most species variety. Several researchers have already worked on the phylogenetic and taxonomic analysis of Verbascum by using ITS data and chloroplast genome fragments and have produced different conclusions. The taxonomy and phylogenetic relationships of this genus are unclear.

Results: The complete plastomes (cp) lengths for V. chaixii, V. songaricum, V. phoeniceum, V. blattaria, V. sinaicum, V. thapsus, and V. brevipedicellatum ranged from 153,014 to 153,481 bp. The cp coded 114 unique genes comprising of 80 protein-coding genes, four ribosomal RNA (rRNA), and 30 tRNA genes. We detected variations in the repeat structures, gene expansion on the inverted repeat, and single copy (IR/SC) boundary regions. The substitution rate analysis indicated that some genes were under purifying selection pressure. Phylogenetic analysis supported the sister relationship of (Lentibulariaceae + Acanthaceae + Bignoniaceae + Verbenaceae + Pedaliaceae) and (Lamiaceae + Phrymaceae + Orobanchaceae + Paulowniaceae + Mazaceae) in Lamiales. Within Scrophulariaceae, Verbascum was sister to Scrophularia, while Buddleja formed a monophyletic clade from (Scrophularia + Verbascum) with high bootstrap support values. The relationship of the nine species within Verbascum was highly supported.

Conclusion: Based on the phylogenetic results, we proposed to reinstate the species status of V. brevipedicellatum (Engl.) Hub.-Mor. Additionally, three genera (Mazus, Lancea, and Dodartia) placed in the Phrymaceae family formed a separate clade within Lamiaceae. The classification of the three genera was supported by previous studies. Thus, the current study also suggests the circumscription of these genera as documented previously to be reinstated. The divergence time of Lamiales was approximated to be 86.28 million years ago (Ma) (95% highest posterior density (HPD),
Introduction
Scrophulariaceae family, generally known as the “figwort” consist of angiosperm plants that are herbs and with one genus of shrubs [1]. The family has about 62 genera and approximately 1830 recognized species [2]. The phylogenetic relationship within Scrophulariaceae remains unresolved topic in angiosperm systematics [3]. To date, only chloroplast plastomes of three genera; Buddleja L. [4], Scrophularia L. [5], and Verbascum L. [6, 7], have been studied in this family. The genus Verbascum (Scrophulariaceae, Lamiales), commonly recognized as “Mullein”, comprises about 360 species that are widely distributed in temperate regions in Europe, Africa, and Asia [8–11]. The main diversity hotspot for this genus is Turkey with 235 existing species [12]. The genus Verbascum is still poorly understood, and new species are described regularly, particularly in Turkey [13]. The taxonomy of Verbascum has been a source of contention, and it varies depending on the treatment. This genus is morphologically characterized by rodulate, biennial, or perennial herbs with yellow-flowered, thyrse, or racemose inflorescences [1, 14]. Verbascum extracts, decoctions, and infusions have been utilized in traditional medicine for a long time. The leaves have been used as a diuretic, sudorific, expectorant, sedative, and the flowers have mucolytic and expectorant properties [15–17].

The generic circumscriptions, and placement of the genus Verbascum is a challenge [18], and this has been the focus of contentious debate among the researchers. The initial description and classification of Verbascum was done by Linnaeus [19], grouped species with 4 stamens in the genus Celsia L. and species with 5 stamens in the genus Verbascum. However, Fischer and Meyer [20] separated Verbascum nacolicum to a new genus Staurophragma because of the oblong cylindrical capsule, a distinguishable feature that he observed. In addition, species which were grouped under the genus Celsia were considered as Scrophularia by wilder [21], Allosinosa by (Rui and Pav. 1786), Janthe and Thapsandra by Griseb., in 1844, Trigura by Dunal (1852), and lastly Alects by Schintz (1889). Bentham’s treatment adopted the classification of Celsia and Verbascum by Linneaus’s, based on the presence of stamens [22]. Kuntze (1891) combined the genus Celsia with Verbascum, based on overall resemblance [23]. However, Murbeck (1925) distinguished Celsia and Verbascum by the number of stamens, the presence of a sessile or stipitate placenta, and the number of blooms in each bract [24]. Distinguishing the species based on morphological features (four or five stamens) was found unreliable. Thus, two genera Celsia and Staurophragma were transferred to Verbascum [18, 25, 26]. The findings of molecular phylogenetic investigations later verified the classification [13]. Verbascum’s systematics has had little attention in contemporary to Scrophulariaceae research; hence, it is one of the least understood in terms of infrageneric categorization. Murbeck (1933), separated this genus into two sections using seed morphology, Aulacospermae Murb. with a longitudinally grooved seed, and Bothrospermae (Murb.) Kamelin with diagonally grooved and alveolate seeds [27, 28]. Huber-Morath (1978) presented an alternate classification system for Verbascum in his review of the Turkish members of the genus [29]. In his treatment of the genus, he erected 13 artificial groups A to M among 243 Verbascum species (including 129 hybrids) and claimed that all Turkish Verbascum species belonged to sect. Bothrospermae Murb. In another regional analysis published for the U.S.S.R separated Verbascum and Celsia and documented 51 species [28]. In 1981, another regional classification published for the Iranian plateau, merged the genus Celsia with Verbascum, but the classification did not suggest any infrageneric classification for the forty-nine species and 4 additional hybrids [29], while classification for Europe documented nearly a hundred species of Verbascum, including Celsia [25].

For the Verbascum species distributed in East Africa, one species is only recognized from the region V. brevipedicellatum and one new genus Pelidium from the family Scrophulariaceae was documented [30–32], while V. siniaticum is distributed in Africa and Arabian Peninsula. Verbascum brevipedicellatum is a perennial herb that is well defined morphologically [31]. Currently, the regional flora of Somalia and plants of the world online database (https://powo.science.kew.org/) shows the infrageneric classification of the V. brevipedicellatum first published in 1973 was a synonym of Rhabdotosperma brevipedicellata (Engl.) Hartl [33–35]. However, in the Flora of Tropical East Africa, R. brevipedicellata is documented as a synonym of V. brevipedicellatum with a note description that it is an extremely variable species, and field studies would be necessary to study the morphological variations among populations. Currently, with the material studied, it was difficult to distinguish the species, even to classify them to lower infraspecific ranks that have been
described around this species. Though documented as a synonym of *R. brevipedicellata* in the plants of the world online (https://powo.science.kew.org/). It is still unclear if the recognition of this species is warranted or supported by other lines of evidence, such as phylogenetic data.

Recently, molecular analyses have been conducted to understand the phylogeny of the genus *Verbascum*. The first study was conducted using a nuclear internal transcribed spacer (ITS) and three non-coding chloroplast DNA regions (*trnY-T, psbA-trnH, and trn-G*) using 41 taxa representing two different genera [13]. This study supported the monophyly of the genus *Verbascum*, which agreed with the previous morphological studies of Scrophulariaceae [27]. However, the study established no supportive evidence for the subgeneric classification of the genus *Verbascum*. Previous molecular analysis conducted by Ghahremaninejad et al. [13] suggested that the genus could be monophyletic. Sotoodeh [36], further investigated the genus in Iran and his finding was similar. However, recent studies using plastomes of *V. phoeniceum* L. and *V. chinense* L. and twenty-eight species in the order Lamiales revealed the placement of the genus *Verbascum* [6, 7]. Bi et al. [6], showed that *Verbascum* formed a monophyletic branch sister to Scrophularia while another study by He et al. [7], placed *Verbascum* as sister to Scrophularia. Buddleja formed a sister clade to (*Verbascum + Scrophularia*). Besides, the study by He et al. [7], was not in agreement with the previous studies which indicated the monophyly of the genus. Therefore, the classification of this genus within the family Scrophulariaceae is controversial, and more molecular data is required to solve the existing problem. In comparison to chloroplast DNA markers, plastome studies provide more comprehensive genetic data [5, 37, 38]. Plastomes have been recently used as a useful tool in phylogenetic studies and genetic diversity [39–41]. Additionally, the variable regions of plastomes can be used as molecular markers for future phylogenetic and genetic diversity studies [41]. In this study, we sequenced and analyzed the whole plastomes of *V. siniticum, V. brevipedicellatum, V. thapsus, V. songaricum, V. blattaria, V. chaixii, V. pho -niceum* including two additional sequences from NCBI (NC050920 and MT610040). The main goals of this study are (i) to annotate and compare the plastomes of the eight species of genus *Verbascum*; (ii) to confirm if the genus *Verbascum* is monophyletic or nested within Scrophularia; (iii) to evaluate the infraspecific classification of *Verbascum brevipedicellatum* to *Rhabdotosperma brevipedicellatum* (Engl.) Hartl using phylogenetic analysis; (iv) to identify the fast-evolving cpDNA markers for species identification, phylogenetic construction, and phylorueologic studies in future; (v) to understand the evolutionary relations of *Verbascum* species and other species in order Lamiales.

## Results

### Plastomes features

Genome sequencing was performed using the Illumina paired-end technology platform at the Novogene Company (Beijing, China), the total genomic DNA of seven (one accession (*V. phoeniceum*) is similar to the one available in the NCBI) *Verbascum* species were sequenced (Table 1). The plastomes of these seven species were typical circular double-stranded structures and ranged from 153,014 bp in *V. blattaria* to 153,291 bp in *V. chaixii* (Fig. 1). The plastome sequences of the seven *Verbascum* species were deposited in GenBank (Table 1). All seven plastomes show a quadripartite structure, including LSC region (84,263–84,833 bp) and SSC region (17,811–17,884 bp), which are separated by a pair of IRs (25,426 to 25,467 bp) regions. Comparative analysis among the seven species plastomes showed that *V. chaixii* was the largest in size compared to others. All plastomes encoded 114 unique genes, comprising 80 protein-coding genes, 30 tRNA genes, and four rRNA genes (Fig. 1, Table 2).

The IR regions contained 20 duplicated genes including nine protein-coding genes (*rpl12, rpl23, ycf2, ycf15, ndhB, rps7, rps12, rps19, and rpl22*), seven tRNA genes

### Table 1: Voucher number, collection place, pairs of reads used and average base-coverage of the sequenced species

| species                  | Voucher number | Collection     | Pairs of reads used | Average base-coverage |
|--------------------------|----------------|----------------|--------------------|-----------------------|
| *V. chaixii*             | CPG-73024      | Xinjiang, China| 3,784,286          | 543.2                 |
| *V. songaricum*          | CPG-72889      | Xinjiang, China| 7,697,759          | 444.7                 |
| *V. siniticum*           | SAJIT-003840   | Mt. Kenya, Kenya| 13,682,874         | 519.1                 |
| *V. thapsus*             | HGW-2029       | Yunan, China   | 4,096,754          | 431.5                 |
| *V. brevipedicellatum*   | SAJIT-001330   | Lake Nkuga, Kenya| 6,223,865         | 492.3                 |
| *V. phoeniceum*          | CPG-72157      | Xinjiang, China| 4,475,832          | 430.6                 |
| *V. blattaria*           | CPG-73143      | Xinjiang, China| 5,470,460          | 1,742.6               |
(trnH-GUG, trnL-CAG, trnA-UGC, trnR-ACG, trnN-GUU, and trnV-AC) and four rRNAs (rrn4.5, rrn5, rrn16, rrn23). The SSC region contained 13 genes of which 12 were CDs and one tRNA, while the LSC region contained 59 protein-coding genes and 22 tRNA genes (Fig. 1). In total, 15 genes (rps16, atpF, rpoC1, petB, petD, rpl16, rpl2, ndhB, ndhA, trnI-GAU, trnA-UGC, trnV-UAC, trnL-UAA, trnG-UCC, trnK-UUU) had one intron with rpl2 and ndhB duplicated in the IR, whereas two genes, clpP and ycf3 had two introns. The rps12 gene was trans-spliced; the 3' intron and exon were duplicated in the inverted repeat region and the 5'-exon was found in the large single-copy region (Table S1). The total GC content for seven Verbascum species was 38% similar to
the previous two published Verbascum species (Verbascum phoeniceum (NC_050992) and Verbascum chinense (MT610040) (Table 3) [6, 7]. In the IR guanine content was 43.2% much higher when compared to LSC 36.1%, and SSC 32.3% region, similar to a study performed under the family Oleaceae [42].

Repeat analysis
Repeat sequences in the seven sequenced Verbascum species were analyzed using Tandem and REPuter. A total of 209 tandem repeats were detected in all seven Verbascum species (Fig. 2). The number of tandem repeats was 30 in V. blattaria, 26 in V. brevipedicellatum, 34 in V. chaixii, 31 in V. phoeniceum, 32 in V. sinnaticum, 28 in V. songaricum, and 28 in V. thapsus (Fig. 2). A total of 264 long repeats had a length of 30–41 base pairs, 7 motifs with 44 base pairs, a length of 52 base pairs, and a length of 61 base pairs was found in the long repeats analysis. Verbascum blattaria and V. brevipedicellatum each contained 41 long repeats with 21 palindromic (P), and 20 forward (F) (Fig. 2, Table S2-S3). Verbascum chaixii had 40 long repeats comprising of 19 palindromic (P), 20 forward (F), and 1 reverse (R), and had the longest repeats sequence of 61 bp (Fig. 2, Table S4). Verbascum phoeniceum and V. sinnaticum each recorded 42 and 39 repeats. Of these, 21 were palindromic (P), 21 forward (F) long repeats, and 19 palindromic (P), 20 were forward (F) long repeats, respectively (Fig. 2, Table S5-6). Verbascum songaricum was found to have 38 long repeats, of which 20 were palindromic (P), 17 were forward (F), and 1 was reverse (R) (Fig. 2, Table S7). Verbascum thapsus had only a minimum of 34 long repeats, of which both palindromic (P) and forward (F) had 17 each, and it had a longer 52 bp long repeat (Fig. 2, Table S9). No complement repeats were found in any of these species, except for one reverse repeat sequence identified in each of V. chaixii and V. songaricum. Repeats regions were found in four genes in all seven species: psaB, psaA, ycf2, and ycf3. The number of tandem repeats was 30 in V. blattaria, 26 in V. brevipedicellatum, 34 in V. chaixii, 31 in V. phoeniceum, 32 in V. sinaticum, 28 in V. songaricum, and 28 in V. thapsus (Fig. 2).

In the MISA analyses, a total of 348 microsatellites regions (>= 10 bp) were identified in the seven plastomes (Table 2). Each Verbascum species plastome was found to contain 45–53 SSRs, of which 53 SSRs were shared among the three plastomes, 49 SSRs in two species. No pentanucleotide motifs were found in seven Verbascum species while hexanucleotide motifs were only found in V. blattaria. Among these repeats, mononucleotide motifs were the most abundant and most SSRs contributed to the AT richness in all seven
Table 3 Comparison of all Verbascum species characteristics

| Feature            | V. sinaiticum | V. thapsus | V. brevipedicellatum | V. chaixii | V. songaricum | V. phoeniceum | V. blattaria | V. phoenicium (NC_050920) | V. chinense (MT610040) |
|--------------------|---------------|------------|----------------------|------------|---------------|---------------|--------------|----------------------------|----------------------------|
| Genome size        | 153,481       | 153,392    | 153,467              | 153,291    | 153,273       | 153,014       | 153,348      | 153,618                     |                             |
| Large single copy  | 84,715        | 84,666     | 84,678               | 84,779     | 84,534        | 84,571        | 84,263       | 84,601                      | 84,834                     |
| Small single copy  | 17,832        | 17,844     | 17,811               | 17,844     | 17,839        | 17,850        | 17,791       | 17,884                      |                             |
| Inverted repeat    | 25,467        | 25,441     | 25,444               | 25,436     | 25,459        | 25,426        | 25,478       | 25,440                      |                             |

A-T Percentage

| Feature            | V. sinaiticum | V. thapsus | V. brevipedicellatum | V. chaixii | V. songaricum | V. phoeniceum | V. blattaria | V. phoenicium (NC_050920) | V. chinense (MT610040) |
|--------------------|---------------|------------|----------------------|------------|---------------|---------------|--------------|----------------------------|----------------------------|
| Guanine content %  | 38            | 38         | 38                   | 38         | 38            | 38            | 38           | 38                         | 38                         |
| LSC                | 36.1          | 36         | 36                   | 36         | 36            | 36            | 36           | 36                         | 36                         |
| SSC                | 32.3          | 32.3       | 32.3                 | 32.3       | 32.3          | 32.3          | 32.3         | 32.3                       | 32.3                       |
| IR                 | 43.2          | 43.2       | 43                   | 43         | 43            | 43            | 43           | 43                         | 43                         |
| Number of genes    | 80            | 80         | 80                   | 80         | 80            | 80            | 80           | 80                         | 80                         |
| Number of tRNAs    | 30            | 30         | 30                   | 30         | 30            | 30            | 30           | 30                         | 30                         |
| Number of rRNAs    | 4             | 4          | 4                    | 4          | 4             | 4             | 4            | 4                          | 4                           |
| Duplicated genes   | 20            | 20         | 20                   | 20         | 20            | 20            | 20           | 20                         | 20                         |

Fig. 2 Total number of repeats found in V. sinaiticum, V. brevipedicellatum, and V. thapsus

*Verbascum* species (Fig. 3; Table S9-S10). Most of the motifs were found to be located in the non-coding regions and within ycf2 of the *Verbascum* followed by SSC and IR regions (Table S9). A high resemblance pattern was discovered when comparing the types of SSRs in the seven species. SSRs found in the plastomes of seven *Verbascum* species are generally comparable. These SSRs could be employed as molecular markers for species discrimination, genetic diversity, and evolution research.
Comparative analysis

The *Verbascum* plastomes divergence results showed that IR regions have higher similarity compared to the single copy (SC) regions. More genes were conserved in the coding regions than in the noncoding regions, which is a common phenomenon in most angiosperms [42]. Significantly, the most conserved regions were observed across all species in the tRNA and rRNA regions. High variation was observed in the intergenic spacer regions of *trnH-GUG—psbA, atpH – atpI, rps16 – trnQ-UIUG, petN-psbM, psaA – ycf3, ycf4 – cemA*. Non-coding regions in *ccsA – psaC, rpl32—trnL-UAG* reported high divergence in the SSC. The coding regions with the highest variation include *accD, rpoC2* and *matK*. Divergence was also detected in introns of *trnK-UIUU, rps16, ycf3, petD, rpl16, clpP, rps12 and ndhA*. These are regions of rapid evolutionary changes and therefore are essential sites for the development of molecular markers that could be useful in population genetics and phylogenetic studies.

Divergence hotspot identification

To estimate the divergence of among the *Verbascum* species, a total of 5 cp genome sequences of *Verbascum* were chosen to calculate nucleotide diversity (Pi), including *V. sinaticum, V. thapsus, V. brevipedicellatum, V. phoeniceum* and *V. chinense*. Based on this analysis, we identified five remarkably divergent regions among the five plastomes, which were higher than the 0.012 Pi value (*rps16-trnQ-UIUG, rpl32-trnL-UAG, ndhD-psaC, trnH-GLUG*, and *petD* (Fig. 4)). Gene *trnH-GLUG* was the most divergent region with the highest Pi value of (0.013) and is located in the LSC region. These highly divergent regions could be used as potential molecular markers for phylogenetic reconstruction of the genera *Verbascum*. Overall, the result of this study revealed that sequence divergence was concentrated in the LSC and SSC regions, whereas IR regions presented less divergence, consistent with the mVISTA results (Fig. S1).

Inverted repeats

The eight *Verbascum* plastomes compared LSC/IRs and IRs/SSC borders and their adjacent genes (Fig. 5). The *rps19* gene was located in all *Verbascum* species in the LSC/IRb region. In *V. brevipedicellatum rps19* gene prolonged to the IR with 43 bp and *V. chinense* with 41 bp. While in the other remaining six *Verbascum* species *rps19* gene extended with 37 bp in length (Fig. 5). The *ycf1* gene was situated at the IRb/SSC border junction, extending 824 bp into the IRb region in all *Verbascum* species. The *ndhF* gene was situated at the junction of SSC/IRa, extending 2 bp into the IRa region. Overall, the structure and gene content of the eight plastomes were consistent, no significant expansion or contraction of IR regions were found in the *Verbascum* species.
Selective pressure analysis
Rates of synonymous (Ks) and non-synonymous (Ks) substitution rates were calculated using a total of 119 regions extracted (coding and non-coding regions) from the plastomes of *V. sinaiticum*, *V. brevipedicellatum*, and *V. thapsus* as follows: 78 protein-coding genes, four rRNAs, and 37 tRNAs with intergenic spacer locations. In all the extracted regions, the Ka and Ks values of *trnD-trnY* (3.75887, 3.24071) and *rpoc1* gene (3.76133, 3.24071) were the highest. The Ka/Ks was also evaluated because it is used to determine the effect of selective pressure imposed on specific genes. The Ka/Ks value of 15 extracted regions (*atpI*, *atpH*, *CssA*, *matK*, *ndhE*, *psaI*, *psbE*, *psbN*, *psbT*, *rpl23*, *rps14*, *ycf15*, and *trnH-psbA*) was above 1 indicating positive selection (adaptive evolution) while the rest of all the other genes were less than 1 indicating purifying selection in the genes, suggesting that there has been evolutionary pressure to conserve the ancestral state (negative selection) (Fig. S2) (Ka/Ks = 1, neutral selection, Ka/Ks < 1, purifying selection and Ka/Ks > 1 positive selection) [43]. Purifying selection is common in many protein-coding regions [44]. In this study, 90% of the genes had a Ka/Ks ratio of less than 1. Among the extracted regions, the genes with the highest Ka/Ks variability can be used as candidate barcodes to differentiate among *Verbascum* species and in the future applied to perform phylogenetic and phylogeography analysis.

To further verify the results, we calculated the non-synonymous (dN) and synonymous (dS) substitution rates. dN/dS method was used to compare in the selection analysis to check for putative bias for the same functional protein-coding sequences in 80 genes of seven species, including *V. brevipedicellatum*, *V. sinaiticum*, *V. thapsus*, *V. phoeniceum*, *V. chinense*, *B. colvilei*, and *S. henryi* in EasyCodeML software [45]. The ratio (ω = dN/dS) were calculated based on four site-specific models (M0 vs. M3, M1a vs. M2a, M7 vs. M8, and M8a vs. M8) with likelihood ratio test (LRT) threshold of *p* < 0.05 elucidating adaptation signatures within the genome. Among the four models, the comparative LRT of M7 vs. M8 was positive in determining the chi-square *p*-value < 0.05 and the selection strength. Bayes Empirical Bayes (BEB) [46] analysis was implemented in model M8, two sites were detected as the site of positive selection which represented one photosynthesis-related gene *ndhF*, and hypothetical gene *ycfl* (Table 4).

Phylogenetic analysis
To gain a further insight into the phylogenetic position within *Verbascum* species and their relationship to other closely related species and other families in order Lamiales, two data sets including 80 coding sequence (CDS) genes and nrDNA (ITS1 + 5.8S + ITS2) sequences were aligned to construct the phylogenetic tree.
A phylogenetic study using 80 coding sequence (CDS) genes formed a well-supported tree. The Maximum Likelihood (ML) and Bayesian Inference (BI) inferences produced similar trees with a high support value (95% of nodes with bootstrap support > 90 in ML, 97% of nodes with bootstrap support > 0.9 in BI) (Fig. 6). The families, Scrophulariaceae, Oleaceae, Gesneriaceae, and Plantaginaceae were the first basal angiosperms of Lamiales which branched early. Within the higher core, Lamiales (Lentibulariaceae + Acanthaceae + Bignoniaceae + Verbenaceae + Pedaliaceae) formed a sister clade with (Lamiaceae + Mazaceae + Orobancheaceae + Paulowniaceae + Phrymaceae) families (Mazaceae noted with steric because currently is treated differently within Phrymaceae and in our study, it formed a sister clade to Lamiaceae). Within Scrophulariaceae, *Verbascum* formed a sister clade to *Scrophularia*, while the genus Buddleja formed a monophyletic clade. Within Scrophulariaceae, *Verbascum*, *V. blattaria* formed a sister clade to *Scrophularia*, while the genus *Verbascum* formed a sister clade to *Scrophularia*, *V. blattaria* formed a sister clade to *Scrophularia*, and the genus *Verbascum* formed a sister clade to *Scrophularia*.

![Fig. 5](image-url) Comparison of the LSC, IR, and SSC boundaries between *Verbascum* cp plastomes.

Table 4 Positively selected sites were detected in the cp genome of the Scrophulariaceae

| Gene Name | M8 | Selected Sites | Pr (w > 1) |
|-----------|----|---------------|------------|
| ycf1      |    | 380 K         | 0.954*     |
|           |    | 846 L         | 0.953*     |
| ndhF      | 7951 S | 0.957*       |
|           | 7954 L | 0.955*       |

within the higher core, Lamiales (Lentibulariaceae + Acanthaceae + Bignoniaceae + Verbenaceae + Pedaliaceae) formed a sister clade with (Lamiaceae + Mazaceae + Orobancheaceae + Paulowniaceae + Phyrymaceae) families (Mazaceae noted with steric because currently is treated differently within Phyrymaceae and in our study, it formed a sister clade to Lamiaceae). Within Scrophulariaceae, *Verbascum* formed a sister clade to *Scrophularia*, while the genus *Buddleja* formed a monophyletic clade. Within Scrophulariaceae, *Verbascum*, *V. blattaria* formed a sister clade to *Scrophularia*, while the genus *Verbascum* formed a sister clade to *Scrophularia*, *V. blattaria* formed a sister clade to *Scrophularia*, and the genus *Verbascum* formed a sister clade to *Scrophularia*.
values indicating they are closely related and belong to the genus *Verbascum* (Fig. 7).

**Divergence time estimation**

The divergence time of Lamiales (node 0) was approximated at 86.28 million years ago (Ma) (HPD% 85.12–89.76) (Fig. 8). The Orobanchaceae species diverged at different clades, the first clade diverged at an estimated 83.578 Ma (node 0), same as Lamiales, (95% HPD 85.12–89.76), which relates closely with the cretaceous junction. The second clade diverged at an estimated 60.13 Ma (node 2), (95% HPD 58.01–62.98) which relates closely with Paleocene-Eocene junctions, the third clade diverged at 21.59 Ma (node 12), (95% HPD 20.54–22.88) from Paulowniaceae. Scrophulariaceae diverged at 24.26 Ma (node 10) (95% HPD 23.10-25.67) from Pedaliaceae. Moreover, within the family Scrophulariaceae, *Buddleja* was estimated to diverge at 13.32 Ma (95% HPD 12.59–14.21) sister to *Scrophularia* and *Verbascum*, while *Verbascum* diverged at 9.35 Ma (95% HPD 8.77–10.00) sister to *Scrophularia*. Whereas within *Verbascum* the four species diverged at 2.41 Ma (95% HPD 2.17–2.65). However, Phyramaceae and Mazaceae species diverged at different times, Phyramaceae species (*Erythranthe lutea* + *Phryma leptostachya*) diverged at an estimated 23.03 Ma (node 11), (95% HPD 20.54–22.88) sister to Paulowniaceae and Orobanchaceae which relates with the Oligocene and Miocene junctions. While Mazaceae diverged at an estimated 46.23 Ma (node 6), (95% HPD 44.55–48.49) sister to Lamiales.
Discussion

Comparative plastome analysis

Plastomes are used in taxonomic and evolutionary studies to assess evolutionary relationships and compare genome structure, particularly closely related species. Comparative study of Verbascum plastomes indicated that they are highly conserved. The complete plastomes of seven Verbascum species were quadripartite in structure and composed of the LSC, SSC, and (IRa/IRb) repeat regions [47, 48]. The plastome sizes of seven Verbascum species ranged from 153,014 to 153,481 bp. The total gene content and arrangement of seven species of Verbascum were almost similar. The seven species confined the equal number of CDs, rRNA, and tRNA genes. Due to the low substitution rate in their genomes and their recent estimated time divergence, this kind of conservatism is common. Previous studies performed in different closely related taxa, for example, tribes and genera have reported similar findings [49, 50].

The complete plastomes of most angiosperm plants are highly conserved, but contraction and expansion of the SC and IR regions are believed to be the major cause of plastome size variations [51, 52]. For instance, inversions, loss of genes, absence of IR region, contraction/expansion in IR, and duplication of the rps19 gene in the IR have been reported in different species [53, 54]. In this study, we observed that the boundary region between the two inverted repeats and single copy region were highly conserved, and gene distribution and location specification were consistent. Additionally, a comparison of genes near the IR region of eight different Verbascum species showed that genes exhibit different degrees of contraction/expansion at the boundary of the IR region. These results are consistent with previous conclusions, which showed that most angiosperm plants plastomes are usually highly conserved, but little differences occur due to contraction and expansion of the IR and SC junction regions [55–57].

Repeat analysis

Normally, most repeats are found dispersed in noncoding regions and within the ycf2 gene [58]. Comparative analyses of different complete plastomes have indicated that repeated motifs are the factors that cause gene deletion, insertion, and replacement [59, 60]. The repeat analyses of the seven Verbascum species detected 264 repeats, most of which are 30–41 bp long, with 7 repeats of 44 bp, 52 bp, and 61 bp long. Among the seven Verbascum species, V. chaixii contained the largest number of repeated sequences. Plastome combination and sequence differences mostly occur because of untimely recombination of repetitive sequences and slipped-strand mismatches [56]. These repeats are the basis of genetic markers for phylogenetic and population analyses, being applied widely because of their highly polymorphic and simplicity to be amplified [61, 62]. In this analysis, 348 simple sequence repeat motifs were found, most of them were poly-A and T as previously supported by other studies [63, 64]. In addition, dispersed repeats have been increasingly recognized as a potential genetic variation and regulation source. The repeats found in seven Verbascum species indicate genetic variation among the species. Together with these regulatory roles, a structural role of repeated DNA in shaping the 3D folding of genomes has also been proposed [65].
Selective pressure analysis

The synonymous (Ks) and non-synonymous (Ka) substitution rates and their corresponding ratio (Ka/Ks) also known as (dN/dS) have been used widely in calculating nucleotide evolution rates and natural selection pressure [46]. Generally, studies done previously indicated that Ka/Ks ratios mostly are lower than one [66], due to synonymous nucleotide substitutions rates that occur more often compared to non-synonymous substitutions rates. The genes with the highest Ka/Ks variability can be used as candidate barcodes to differentiate species and in the future applied to perform phylogenetic and phyleogeographic analyses. In this study, 15 extracted regions (atpI, atpH, CssA, matK, ndhE, psaJ, psbE, psbN, psbT, rpf23, rps14, rps19, ycf15, and trnH-psbA) were above 1 indicating positive selection.

To further confirm the results, we calculated the non-synonymous (dN) and synonymous (dS) substitution rates using EasyCodeML [45] of seven species, including V. brevipedicellatum, V. sinaiticum, V. thapsus, V. phototheum, V. chinense, B. colvilei, S. henryi within Scrophulariaceae species. We identified two sites that were
detected as the positive selection site, which represented one photosynthesis-related gene ndhF, and hypothetical gene ycf1 (Table 4). All 17 sites can be used to differentiate species and applied in phylogenetic and phylogeography analyses of the genus Verbascum. The varying results of the Ka/Ks ratio obtained in our study are evidence that evolutionary rates of chloroplast genomes vary among genes. This supported by the previous conclusions drawn by Manezes et al., (2018) [67].

**Phylogenetic analysis**

The sequencing of complete plastomes has increased recently due to the advancement in sequencing technology [68]. Plastomes are important in molecular, phylogenetic, and evolutionary studies as well as solving relationships in plants and can be applied to plants DNA barcodes [69]. The phylogenetic relationships within Scrophulariaceae is challenging to taxonomists, these difficulties in phylogenetic reconstruction of Scrophulariaceae are likely due to reticulate evolution caused by hybridization and polyploidization [5, 13]. Phylogenetic analysis using 80 protein-coding genes produced a well-supported phylogenetic tree (Fig. 6). The first basal angiosperm classification of Lamiales (Oleaceae, Gesneriaceae, Plantaginaceae, and Scrophulariaceae) was confirmed, and the other families (Phrymaceae + Orobanchaceae + Paulowniaceae) which were in concordance with the previous studies [5, 70–73]. However, the most problematic of the families’ placement in the previous studies were (Lentibulariaceae + Acanthaceae + Bignoniaceae + Pedaliaceae) + (Lamiaceae + Thymianaceae + Verbascum) + Menispermaceae and epiphytic ferns time divergence causes this problem. In this study, our crown age estimation caused by hybridization and polyploidization [5, 13]. More chloroplast genes may result in this difference. Within all higher core, Lamiales (Lentibulariaceae + Acanthaceae + Bignoniaceae + Pedaliaceae) formed a sister clade to (Lamiaceae + Phrymaceae + Orobanchaceae + Paulowniaceae + Mazaceae) (Mazus, Lancea, and Dodartia) families, (Mazaceae noted with steric because currently it is treated separately differently within Phrymaceae, in this study, it formed a sister clade with Lamiales). The results of this study confirm the results of a previous study that formed a separate clade [74], and consistent with the AGP IV classification of the Mazaceae family [75]. Therefore, in this study, we suggest the classification of the Mazaceae family to be reinstated back to formal recognition as suggested by James L. Revel [76]. Interestingly, the relationships of the genera within the family Scrophulariaceae is well solved and confirmed.

In the plastomes of 80 CDS tree species, within Verbascum, V. blattaria formed a sister clade to (V. chinense + V. phoeniceum) + (V. brevipedicellatum + V. thapsus + V. siniticum + V. songaricum + V. chaixii) with high bootstrap support values indicating they are closely related and belong to the same genera. But the species from Kenya and China were not well separated into two monophyletic lineages. Even two species from Kenya, V. siniticum and V. brevipedicellatum, were clustered with species from China, V. siniticum was instead more closely related to V. thapsus from Yunnan Province, China (Fig. 6). This suggests that research on the origin and biogeography of Verbascum species is still underexplored. In addition, the phylogenetic analysis based on nrDNA (ITS1 + 5.8S + ITS2) sequence within Scrophulariaceae (Verbascum + Scrophularia + Buddleja) suggests that V. brevipedicellatum formed a sister clade to V. phoeniceum + (V. nudicaule + Verbascum sp.) with high bootstrap support values indicating they are closely related and belong to the genus Verbascum (Fig. 7). However, in our study V. macrocarpum (KP738154) did not cluster into the genus Verbascum, but formed a monophyletic branch, which conforms with the previous study that the species are unresolved [13]. More chloroplast genome data on V. macrocarpum may be needed to solve this problem. In this study, our crown age estimation of Lamiales of 86.26 (Ma) (HPD% 85.1–89.9 Ma) was in concordance with the previous studies divergence time estimation (95% HPD, 67–101 Ma, 95% HPD, 70–99.8 Ma) [5, 77–79]. We selected two reliable fossils that represent the old dating period of order Lamiales as constraints for calibrating the age of our angiosperm tree from the well-reported previous analyses [5]. However, Menispermaeae and epiphytic ferns time divergence was estimated around the cretaceous and Paleocene periods which is linked to the development of tropical rain
forest angiosperms [80, 81]. The study provides a more current, comprehensive and detailed framework to the evolution of the family Scrophulariaceae and other families in order Lamiales with additional data of complete chloroplast plastomes.

**Taxonomic treatment of Verbascum brevipedicellatum**

Our phylogenetic results of nrDNA sequences and plastomes showed *Verbascum brevipedicellatum* formed a sister relationship with other *Verbascum* species. Although several names have been associated with this species in East Africa. In the FTEA it was documented as:

*Verbascum brevipedicellatum* (Engl.) Huber-Morath [family SCROPHULARIACEAE], in Bauhinia 5: 11 (1973); Blundell, Wild Fl. East Afr.: 378, pl. 393 (1987); U.K.W.F.: 254 (1994). Type: Tanzania, Kilimanjaro [Kilmandschar], Meyer 286 (B†, holotype). Neotype: Tanzania, Kilimanjaro, SE of Bismark hut, Bigger 2012 (K!).

*Celsia brevipedicellata* Engl. [family SCROPHULARIACEAE], in Abhandl. Akad. Wissensch. Berlin 1891(2): 237 (1892); Skan in F.T.A.: 4(2): 285 (1906); Murb. in Lunds Univ. Arsskr. 22, 1: 65 (1925); A.V.P.: 164 (1957); F.P.U. ed. 2: 133 (1971).

*Celsia brevipedicellata* Murb. var. homostemon [family SCROPHULARIACEAE], in Lunds Univ. Arsskr. 22, 1: 67 (1925). Type: Kenya, Aberdare Mts, 13 March 1922, R.E. & T.C.E. Fries 2262 (UPS!, holotype).

*Celsia brevipedicellata* Murb. var. heterostemon [family SCROPHULARIACEAE], in Lunds Univ. Arsskr. 22, 1: 68 (1925). Type: East Africa, unspecified.

*Celsia keniensis* Murb. [family SCROPHULARIACEAE], in Lunds Univ. Arsskr. 22, 1: 70, t. 2 (1925); Glover, Prov. Check-List Brit. & It. Somal.: 244 (1947). Type: Kenya, Mt Kenya, R.E. & T.C.E. Fries 458 (K!, UPS, syn.). 1869 (UPS!, synotypy).

*Celsia floccosa* [family SCROPHULARIACEAE], [sensu Agnew UKWF: 550 (1974), non Benth.]

*Rhabdotosperma brevipedicellata* (Engl.) D. Hartl [family SCROPHULARIACEAE], in Beitr. Biol. Pfl. 53(1): 58 (1977); Fischer, F. A.C. Scrophulariaceae: 12, pl. 2 (1999) & in Fl. Ethiop. & Eritr. 5: 252 (2006).

*Rhabdotosperma keniensis* (Murb.) D. Hartl [family SCROPHULARIACEAE], in Beitr. Biol. Pfl., 53(1): 58 (1977); Fischer in Thulin (ed.), Fl. Somal.: 3: 266 (2006).

*Verbascum* is a Linnean genus [19], while *Rhabdotosperma* has been designated by Hartl [82], although in the flora of Somalia Rhabdotosperma is the conserved name, so we consider it as superfluous because according to article 14.5 of the International Code of Nomenclature for algae, fungi, and plants [83], specifies that if a conserved name competes with an earlier name against which it has not explicitly been conserved, that the early name is adopted. Therefore, in this regard, *Verbascum* has priority over *Rhabdotosperma* and also over *Celsia*. Additionally, based on the phylogenetic results and type specimens checked, we propose to reinstate the species status of *Verbascum brevipedicellatum* (Engl.) Huber-Morath.

**Conclusion**

In this study, the plastomes of seven *Verbascum* species (including all the *Verbascum* species distributed in China and two from Kenya) were sequenced and analyzed, determined their phylogenetic placement as well as their dating in the order Lamiales. The chloroplast plastomes indicated that gene organization, GC content, and plastome size are highly conserved. Tandem repeats were highest in *V. chaixii* followed by *V. sinaiticum* and then *V. phoeniceum* (>31). The other remaining five *Verbascum* species recorded the lower number of tandem repeats (<31), while simple sequence repeats were highest in *V. chaixii, V. sinaiticum* and then *V. phoeniceum* respectively. The phylogenetic analysis based on 89 taxa indicated that *Verbascum* formed a sister clade to Scrophularia. The Scrophulariaceae family diverged at 24.26 Ma (node 9) (95% HPD 23.10—25.67) from Pedaliaceae. The three *Verbascum* species diverged at 9.35 Ma (95% HPD 8.77-10.00) from Scrophularia. Classification of *V. brevipedicellatum* to Rhabdotosperma was not supported by our phylogenetic analysis hence suggesting the reinstatement of the species name. Additionally, we suggest the reinstatement of the families Phyrpamaceae and Mazaceae because they formed separate clades with high bootstrap support values. Notably, their divergence period was also different; Mazaceae diverged early compared to Phyrpamaceae species. This study is essential because it indicates the relationship of various families as well as the divergence time estimate in order Lamiales. In addition, phylogenetic results of nrDNA sequences and plastomes supported *Verbascum brevipedicellatum* formed a sister clade with other *Verbascum* species.

**Materials and methods**

Plant material, DNA extraction, and library preparation

The samples of the seven *Verbascum* species, *V. sinaiticum* (voucher Number SAJIT-003840), *V. brevipedicellatum* (Voucher number SAJIT-001330) were collected from Mt. Kenya and Lake Nkuga, Kenya. *Verbascum chaixii* (voucher Number CPG-73024), *V. songaricum* (voucher Number CPG-73143) were collected from Xinjiang province, China, while *V. thapsus* (Voucher number HGW-2029) was collected from Yunnan province, China, and the samples were preserved using Silica gel (Table 1). The voucher specimens were deposited at the East African
Plastome and nrDNA assembly, annotation

The high-quality reads were used for de novo assembly to reconstruct *Verbascum* chloroplast genomes using GetOrganelle v.1.7.2. with wordsize of 150 and K-mer sizes of 105 [85–88]. The visualization of the final assembled graphs was done using Bandage to authenticate the produced plastid plastome [89]. The quality of the newly assembled plastomes was confirmed according to the reading level by aligning the trimmed raw reads to the de novo assemblies using Geneious mapper, Geneious prime 2021 [90], with medium- to low-sensitivity option and iteration up to five times [91]. The annotation of the sequences was performed using CpgAVAS2 [92]. The online blast software version 2.2.25 (https://blast.ncbi.nlm.nih.gov/ Blast. cgi) software was used to match the cp plastomes CDs sequences on NCBI and manual edit them correctly. The tRNAscan-SE software was used to annotate tRNA [93]. The complete circular plastomes were mapped using the OGDRAW program (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) [94], and the complete genomes were deposited in the Gene bank, Accession numbers: *V. sinaiticum* (MW751818), *V. thapsus* (MW751819), and *V. brevipedicellatum* (MW751817), *V. chaixii* (ON121985), *V. songaricum* (ON121987), *V. phoeniceum* (ON121986) and *V. blattaria* (ON121984). The nrDNA sequence (18S-ITS1-5.8S-ITS2-26S) of *V. sinaiticum* (OL763335), *V. thapsus* (OL763334), and *V. brevipedicellatum* (OL763333) were assembled using GetOrganelle v.1.7.2 with default parameters.

Repeat analysis

Online REPuter software (https://babiserv.cebitec.uni-bielefeld.de/reputer) was used to identify and locate forward, palindromic, reverse and complement sequences with minimum repeat size of 20 bp, maximum repeat sequences number of 200 and the E-value below 0.01 [95]. Additionally, tandem repeats were identified using the Tandem repeat finder program [96], with two for the matching alignment and seven for mismatch and indels. Finally, SSRs were identified using the MISA online software (https://webblast.ipk-gatersleben.de/misa/) with the minimum repeat parameters set as 12, 6, 4, 3, 3, 3 repeat units for mono-, di-, tri- tetra-, penta-, and hexanucleotide SSRs, respectively [97].

Comparative analysis

To compare the structural differences and similarities between *Verbascum* species, the mVISTA online tool was applied to analyze the five representative species (*V. sinaiticum*, *V. phoeniceum*, *V. brevipedicellatum*, *Verbascum chinense*, *V. thapsus*) plastome sequences with the Shuffle LAGAN model [98], using the reference of the annotation of *V. chinense*. The chloroplast online tool was used to indicate the comparison between the boundaries between, Large Single regions (LSC), Small Single copy (SSC), and the inverted repeats (IR) regions among the eight (*V. sinaiticum*, *V. phoeniceum*, *V. brevipedicellatum*, *Verbascum chinense*, *V. thapsus*, *V. chaixii*, *V. songaricum*, *V. blattaria*) complete cp plastomes. To explore the highly divergent regions of the cp genomes within eight *Verbascum* species, the software DnaSP version 6.12.03 [99] was used to calculate nucleotide diversity (Pi). The step size and window length were set as 200 bp and 600 bp, respectively. Geneious prime 2021 [90] was used to detect the contraction/expansion of the inverted repeat regions (IRs), and the final graph of expansions/contractions was visualized using Adobe Illustrator.

Selective pressure analysis

To analyze the substitution rate within *Verbascum* species, a total of 119 regions were extracted from the three representative species (*V. sinaiticum*, *V. brevipedicellatum*, *V. thapsus*) cp plastomes. Geneious prime was
### Table 5  Species Table Retrieved from NCBI of complete chloroplast genome sequences

| Species name                 | Accession no | References |
|------------------------------|--------------|------------|
| Verbascum sinaticum          | MW751818     | The present study assembled |
| Verbascum thapsus            | MW751819     | The present study assembled |
| Verbascum brevipedicellatum  | MW751817     | The present study assembled |
| Verbascum chaixii            | ON121985     | The present study assembled |
| Verbascum sorianicum         | ON121987     | The present study assembled |
| Verbascum phoeniceum         | ON121986     | The present study assembled |
| Verbascum blattaria          | ON121984     | The present study assembled |
| Verbascum phoeniceum         | NC_050920    | Bi et al., 2020 (10.1080/23802359.2020.1715880) |
| Verbascum chinense           | MT610040     | He et al., 2020 (10.1080/23802359.2020.1715880) |
| Genlisea violacea            | MF593126     | Silva et al., 2018 (10.1371/journal.pone.0190321) |
| Origanum vulgare             | JX880022     | Lukas et al., 2013 (10.1016/j.gene.2013.07.026) |
| Mentha longifolia            | NC_032054    | Vining et al., 2017 (10.1016/j.molp.2016.10.018) |
| Dracocephalum palatum        | NC_031874    | Unpublished |
| Salvia miltiorrhiza          | NC_020431    | Qian et al., 2013 (10.1371/journal.pone.0057607) |
| Lavandula angustifolia       | NC_029370    | Unpublished |
| Rosmarinus officinalis       | NC_027259    | Unpublished |
| Penilla frutescens           | NC_030757    | Unpublished |
| Ocimum basilicum             | NC_035143    | Unpublished |
| Stenogyne bifida             | NC_029818    | Unpublished |
| Haplastachys haplastachya    | NC_029819    | Unpublished |
| Phyllostegia velutina        | NC_029820    | Unpublished |
| Scutellaria baicalensis      | MFS21633     | Jiang et al., 2017 (10.3390/genes800227) |
| Ajuga reptans                | NC_023102    | Zhu et al., 2014 (10.1093/molbev/msu079) |
| Caryopteris mongholica       | NC_035729    | Liu et al., 2018 (10.1007/s12686-017-0802-5) |
| Premna microphylla           | KM981744     | Unpublished |
| Tectona grandis              | NC_020098    | Unpublished |
| Mazus miqelli                | MW238406     | Unpublished |
| Mazus xiiiningensis          | MW238409     | Unpublished |
| Mazus pumilus                | NC_042444    | Unpublished |
| Mazus lancefolius            | MW238405     | Unpublished |
| Dodartia orientalis          | MW238404     | Unpublished |
| Lancea hirsuta               | NC_037506    | Chi et al., 2018 (10.3390/molecules23030602) |
| Lancea tibetica              | NC_037693    | Unpublished |
| Orobanche densiflora         | KT387723     | Cusimano et al., 2016 (10.1111/nph.13784) |
| Bouardia latissquama         | NC_025641    | Unpublished |
| Phelipanche purpurea         | NC_023132    | Wicke et al., 2013 (10.1105/tpc.113.113373) |
| Orobanche californica        | NC_025651    | Unpublished |
| Conopholis americana         | HG514459     | Unpublished |
| Lindenbergeria philippensis  | NC_022859    | Unpublished |
| Pedicularis muscicola         | NC_046853    | Unpublished |
| Pedicularis oederi           | NC_046854    | Unpublished |
| Castilleja ramamensis        | NC_031805    | Unpublished |
| Melampyrum koreanum          | MW463054     | Unpublished |
| Lathrea squararia            | NC_027838    | Samigullin et al., 2016 (10.1371/journal.pone.0150718) |
| Schwalbea americana          | NC_023115    | Unpublished |
| Siphonostegia chinensis      | NC_046038    | Gao et al., 2019 (10.1080/23802359.2018.1564384) |
| Triaenophora shennongjiaensis| MH071405     | Xia et al., 2018 (10.1080/23802359.2018.1467242) |
| Rehmannia chinii             | NC_033534    | Zeng et al., 2016 (10.1007/s12686-016-0577-0) |
used to extract the regions. Stop codons were cut and removed manually, and the gaps were removed within the sequences using the Gblocks server online. Mega 7 was used to align combined files after the removal of stop codons and gaps. We converted the aligned files manually by saving them into axt format. The non-synonymous (Ka), synonymous (Ks) substitution rates of 80 protein coding genes and non-coding from the Verbascum species and Ka/Ks ratio of each region were calculated using Ka/Ks calculator v 2.0 with maximum likelihood methods Ma (a model that averages parameters across 14 candidate models) and Ms method (a model that has the smallest AICC among 14 candidate models) in a standard code [100].

Another method was employed to determine positive selective pressure within shared genes of seven species

| Species name                  | Accession no | References                                                                 |
|-------------------------------|--------------|----------------------------------------------------------------------------|
| Paulownia coreana             | KP718622     | Yi et al., 2016 (10.1080/23802359.2016.1214546)                           |
| Paulownia tomentosa           | KP718624     |                                                                            |
| Scrophularia buergeriana       | KP718626     |                                                                            |
| Erythranthe lutea             | KU705476     | Vallego-Marín et al., 2016 (10.3732/ajb.1500471)                          |
| Phryma leptostachya           | MK381317     | Xia et al., 2019 (10.3389/fpls.2019.00528)                                |
| Utricularia foliosa           | KY025562     | Silva et al., 2017 (10.1007/s12686-016-0653-5)                            |
| Utricularia gibba             | NC_021449    | Ibara-Laclette et al., 2013 (10.1038/nature12132)                        |
| Verbena officinalis           | MW348926     |                                                                           |
| Pinguicula alpina             | MT740255     | Unpublished                                                                |
| Pinguicula ehlersiae          | NC_023463    | Unpublished                                                                |
| Andrographis paniculata       | KF150644     | Unpublished                                                                |
| Echinocanthes lophouensis     | MF490441     | Unpublished                                                                |
| Sesamum indicum               | JN637766     | Yi et al., 2012 (10.1371/journal.pone.0035872)                            |
| Adenocalymma aurantiacum      | NC_036495    | Fonseca et al., 2017 (10.3389/fpls.2017.01875)                            |
| Neojobertia candelleana       | NC_036503    |                                                                           |
| Tanaecium tetragonolobum      | NC_027955    | Nazareno et al., 2015 (10.1371/journal.pone.0129930)                     |
| Glandularia tenera            | MW538952     | Unpublished                                                                |
| Aloysia citrodora             | KY085903     | Unpublished                                                                |
| Scrophularia henryi           | MF861203     | Unpublished                                                                |
| Scrophularia takesimensis     | KP718628     | Unpublished                                                                |
| Scrophularia dentata          | MF861202     | Unpublished                                                                |
| Buddleja colvilei             | MH411147     | Ge et al., 2018 (10.3390/molecules23061248)                               |
| Buddleja sessilifolia         | MH411151     |                                                                           |
| Veronica nakaiana             | KTE33216     | Choi et al., 2016 (10.3359/fpls.2016.00355)                               |
| Veronicastrum sibiricum       | KTE324053    |                                                                           |
| Plantago media                | NC_028520    | Zhu et al., 2016 (10.1111/nph.13743)                                      |
| Digitalis lanata              | NC_034688    | Unpublished                                                                |
| Chinita brachytricha          | MF177037     | Unpublished                                                                |
| Primulina linearifolia        | MF472013     | Unpublished                                                                |
| Boea hygrometrica             | JN107811     | Zhang et al., 2011 (10.1186/1746–4811-7–38)                               |
| Haberlea rhodopensis          | NC_031852    | Unpublished                                                                |
| Olea woodiana                 | NC_015608    | Besnard et al., 2011 (10.1186/1471–2229-11–80)                            |
| Chionanthus retusus           | NC_035000    | He et al., 2017 (10.1007/s12686-017-0704-6)                               |
| Hesperoaceae palmieri         | NC_025787    | Unpublished                                                                |
| Jasminum nudiflorum           | NC_008407    | Lee et al., 2007 (10.1093/molbev/msm036)                                  |
| Forsythia suspensa            | MF579702     | Wang et al., 2017 (10.3390/ijms.18112288)                                 |
| Abelophyllum distichum        | MN127986     | Min et al., 2019 (10.1080/23802359.2019.1679678)                         |
| Coffea arabica                | NC_008535    | Unpublished                                                                |
| Coffea canephora              | NC_030053    | Unpublished                                                                |
was evaluated using the PAML v4.7 [45] package imple-
mented in EasyCodeML software. Non-synonymous
(dN) and synonymous substitution (dS) substitution
rates, and their ratio (ω = dN/dS) were calculated based
on four site-specific models (M0 vs. M3, M1a vs. M2a,
M7 vs. M8 and M8a vs. M8) with likelihood ratio test
(LRT) threshold of p < 0.05 elucidating adaptation signa-
tures within the genome. The models permit dN/dS vari-
ation within sites while keeping the ω ratio fixed within
branches. Selective pressure analysis was conducted
along ML tree in plain Newick format based on protein-
coding sites used in the generation of phylogenetic rela-
tionship of the selected seven species. Here, individual
CDS sequences were aligned in correspondence to their
amino acids, and their selection was evaluated based on
both ω and LRTs values.

### Phylogenetic analysis

To infer the phylogenetic tree, we used 80 protein-coding
genes of 87 taxa in order Lamiales and two outgroups,
which all but seven of our newly sequenced data were downloaded from NCBI. (Table 5), combined in a single file and aligned using MAFFT [101]. Geneious prime version 2021 was used to concatenate the sequence into various readable formats for other analyses. Performed maximum likelihood (ML) analyses using IQ tree [102], integrated with Phylosuite [103], under the best-fitting model according to Akaike Information Criterion (AIC) of all protein-coding genes under the GTR + R6 + F model for 5000 ultrafast bootstraps [104], as well as the Shimodaira–Hasegawa–like approximate likelihood-ratio test [105]. To ascertain the results, BI analysis was performed using the same data in 80 protein-coding genes of 87 taxa in order Lamiales in MrBayes 3.2.7a and using the Markov chain Monte Carlo (MCMC) method, with two independent runs for 10 million generations (Number of Chains is four). We also used second data set to confirm the taxonomic treatment of Verbascum brevipedicellatum within the family Scrophulariaceae. We used nrDNA ITS1 (5.8S + ITS2) of 36 taxa in the family Scrophulariaceae (Verbascum + Scrophularia + Buddleja), and two outgroups all but three of our newly sequenced data were downloaded from NCBI. (Table 6), sequence alignment was performed using MAFFT v7.308 with default parameters setting. Phylogenetic relationships reconstructions were performed based on maximum likelihood (ML) analysis using the program IQ-TREE with 1000 bootstrap replications. The best-fit model GTR + F + G4 was chosen according to Bayesian Information Criterion (BIC).

Divergence time estimation

BEAST software was used to perform the analysis. Two secondary calibration fossils and one fossil record were used to constrain the nodes; node (0) the 95% highest posterior density (HPD) the old (limit)divergence time for Lamiales (95% HPD, 67–101 Ma; Barba-Montoya et al. [77]), constraining the crown age of Lamiales (95% HPD, 85.05 Ma (Lognormal priors, 1.0 Mean (M), Sigma (S) 0.5). Constraining the most common ancestor of Verbascum (lognormal priors, Off set 65.64, M 1.0, S $\frac{1}{2}$ 0.5) [5]. The chronogram and branch intervals were linked to partitions and other constraints were unlinked. We used the Yule process prior for speciation and uncorrelated lognormal relaxed clock model. Prior’s constraints time of the nodes were selected using the Log-Normal distribution of mean and standard deviation set at the mean and median limits, and the GTR + I + G substitution model was set as the nucleotide substitution model. Two calibration points were selected to estimate the divergence time of Verbascum (Scrophulariaceae): (0) the 95% highest posterior density (HPD) limit of crown age for Lamiales (85.1, 1.0 Mean (M), Sigma (S) 0.5), and to estimate the most common recent ancestor to Verbascum and their age divergence (Log-Normal priors, Offset 65.64, M 1.0, S $\frac{1}{2}$ 0.5). To estimate the dating time, used BEAST2 on XSEDE version 1.8.0 on the CIPRES web server. The MCMC analyses were run for 10 million generations, with sampling every 1000 generations. To check for repeatability, uniformity, and coalescent model parameters in each run three separate BEAST analyses were performed. Tracer version 1.7.1 was used to check the burn-in of trees and chains distribution. Tree Annotator version 1.8.0, was used to summarize and obtain a readable file to show the post-burn-in trees and produce maximum clade credibility. Figtree version 1.4.4 was also used to show the mean divergence time approximates with 95% HPD intervals and to visualize the tree.
Funding
This work was financed by grants from the International Partnership Program of the Chinese Academy of Sciences (151835KYSB20190027), the National Natural Sciences Foundation of China (31970211), and Sino-Africa Joint Research Center, CAS, (SA.C.E202101). We thank the National Wild Plant Germplasm Resource Center for all kinds of support. We also thank Shuai Peng, who helped collect plant materials used for the experiments and technical support, and Florі Mkangombo Kilio for her comments and suggestions. All authors have read and agreed on the manuscript version to be published.

Availability of data and materials
All data generated or analyzed during this study are included in this published article and the complete chloroplast genome sequences (nrDNA sequences) of V. sinaiticum, V. brevipedicellatum, V. thapsus, V. chaixii, V. songaricum, V. phoeniceum, and V. blattaria are deposited in the GenBank with ID no: MW751818, MW5751819 (OL763334), MW751817 (OL763335), ON21985, ON21987, ON21986, and ON21984, respectively. The accession numbers corresponding to the additional datasets used and analyzed in this study can be found in Table 5 and Table 6. These were retrieved from National Center for Biotechnology Information database.

Declarations

Ethics approval and consent to participate
The authors have complied with the relevant institutional, national and international guidelines in collecting biological materials for the study. The relevant permits for this research were granted by National Commission of Science, Technology and Innovation (NACOSTI) of Kenya (NACOSTI/P/19/20003/28091), Kenya Wildlife Service (KWS), Kenya Forest Service (KFS), Chinese Academy of Sciences, and The Wuhan Botanical Garden of the Chinese Academy of Sciences. No materials from animal or human were used in this research.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1. CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China. 2. Sino-Africa Joint Research Center, Chinese Academy of Sciences, Wuhan CN-430074, China. 3. University of Chinese Academy of Sciences, Beijing CN-100049, China. 4. East African Herbarium, National Museums of Kenya, PO Box 451660-0100, Nairobi, Kenya.

Received: 3 August 2021 Accepted: 28 July 2022
Published online: 08 August 2022

References
1. Fischer E. Scrophulariaceae. In: Kadereit JW, editor. Flowering Plants Dicotyledons. The Families and Genera of Vascular Plants, vol. 7. Berlin, Heidelberg: Springer; 2004. p. 333–432.
2. Christenhusz MJ, Byng JW. The number of known plant species in the world and its annual increase. Phytotaxa. 2016;261(3):201–17.
3. Judd WS, Olmstead RG. A survey of tricoplate (eudicot) phylogenetic relationships. Am J Bot. 2004;91(10):1627–44.
4. Ge J, Cai L, Bi G-Q, Chen G, Sun W. Characterization of the Complete Chloroplast Genomes of Buddleja colvilei and B. sessilifolia: Implications for the Taxonomy of Buddleja L. Molecules. 2018;23(6):1248.
5. Xu WQ, Losh J, Chen C, Li P, Wang RH, Zhao YP, Qiu YX, Fu CX. Comparative genomics of figworts (Scrophularia, Scrophulariaceae), with implications for the evolution of Scrophularia and Lamiales. J Syst Evol. 2019;57(1):55–65.
6. Bi Y, Deng P, Li U. The complete chloroplast genome sequence of purple mullein (Verbascum phoeniceum L.). Mitochondrial DNA Part B. 2020;5(1):819–20.
7. He Y, Ma Y, Li Z, Liu P, Yuan W. The complete chloroplast genome of Verbascum chinense (L.) Santapau. Mitochondrial DNA Part B. 2020;5(1):3021–2.
8. Hasler M. Synonymic Checklists of the Vascular Plants of the World. In: Blanks Q, Roskov Y, Vanegrette L, DeWalt RE, Rennes D, Schach P, Orell T, Kepping M, Miller J, Aalbu R, Adaldr R, Adriaenssens E, Aedo C, Aescht E, Akkar N, Alonso-Zarazaga MA, Alvarez B, Alvarez F, Anderson G, et al, editors. Catalogue of Life Checklist (Version 2021-08-06). 2021.
9. Georgiev MI, Ali K, Alipieva K, Verpoorte R, Choi YH. Metabolic differentiations and classification of Verbascum species by NMR-based metabolomics. Phytochemistry. 2011;72(16):2045–51.
10. Press BRIT. Mabberley’s Plant-Book: A Portable Dictionary of Plants, their Classification and Uses. J Bot Res Inst Tex. 2018;12(1):2578.
11. Riahi M, Ghahremaninejad F. The tribe Scrophularieae (Scrophulariaceae): A Review of Phylogenetic Studies. Hacquetia. 2019;18(2).
12. Yilmaz G, Dane F. The genus Verbascum L. in European Turkey. Botanica Sertlica. 2012;26:9–13.
13. Ghahremaninejad F, Riahi M, Babaei M, Attar F, Behzad L, Sonboli A. Monophyly of Verbascum (Scrophulariaceae: Scrophulariaceae): evidence from nuclear and plastid phylogenetic analyses. Aust J Bot. 2015;62(8):638–46.
14. Oxelmann B, Kornhall P, Olmstead RG, Bremer B. Further disintegration of Scrophulariaceae. Taxon. 2005;54(2):411–25.
15. Lust J, Tierra M. The Natural Remedy Bible. Simon and Schuster; 2003.
16. Vogl S, Picker P, Mihaly-Bison J, Fakhruddin N, Atanasov AG, Heiss EH, Wawrosch C, Renzick G, Dirsch VM, Sauer J. Ethnopharmacological in vitro studies on Austria’s folk medicine—an unexplored lore in vitro anti-inflammatory activities of 71 Austrian traditional herbal drugs. J Ethnopharmacol. 2011;143(9):750–71.
17. Kaur V, Upadhyaya K. Antibacterial activity of Verbascum chinense (Scrophulariaceae) extracts. Int J Curr Microbiol App Sci. 2016;5(4):578–84.
18. Karavelioğulları F, Aytac Z. Revision of the genus Verbascum L (Group A) in Turkey. Botany Res J. 2008;1(1):9–32.
19. Limnæus C. Species plantarum, vol. II. Stockholm; 1753.
20. Fischer M, Meyer C. Index Hort Petrop. 1843;9:78–9.
21. Wydler H. Essai monographique sur le genre Scrofularia. Barbezae et Delarue; 1828.
22. Bentham G, Hooker J. Scrophulariaceae. Prodomus systematis naturalis regni vegetabilis. 1846;10(186):586.
23. Kunze C. Orchidaceae. Revisio generum plantarum. 1891;2:645–82.
24. Merbeck S. Monographie der gattung Celsia. Gleerup, 1925.
25. Ferguson I. Verbascum L. Flora europaea. 1972;2:305–16.
26. JTW. Flora of Turkey and the East Aegean Islands, vol. 3. 1972.
27. Murbeck SS. Monographie der gattung Verbascum. 1933.
28. Fedchenko BA. Verbascum L. In: Shishkin BKS, Bobrow EG, editors. Flora U. S. S. R. Izdatelstvo Akademi Nauk S.S.S.R.Leningrad. 1955;22:132–97.
29. Huber Morath A. Verbascum. In: Rechinger Kh, Flora Iranica. No: 20. Fischer E. Scrophulariaceae. In: Kadereit JW, editor. Flowering Plants Dicotyledons. The Families and Genera of Vascular Plants, vol. 7. Berlin, Heidelberg: Springer; 2004. p. 333–432.
30. Christenhusz MJ, Byng JW. The number of known plants species in the world and its annual increase. Phytotaxa. 2016;261(3):201–17.
31. Judd WS, Olmstead RG. A survey of tricoplate (eudicot) phylogenetic relationships. Am J Bot. 2004;91(10):1627–44.
32. Ge J, Cai L, Bi G-Q, Chen G, Sun W. Characterization of the Complete Chloroplast Genomes of Buddleja colvilei and B. sessilifolia: Implications for the Taxonomy of Buddleja L. Molecules. 2018;23(6):1248.
33. Xu WQ, Losh J, Chen C, Li P, Wang RH, Zhao YP, Qiu YX, Fu CX. Comparative genomics of figworts (Scrophularia, Scrophulariaceae), with implications for the evolution of Scrophularia and Lamiales. J Syst Evol. 2019;57(1):55–65.
34. Bi Y, Deng P, Li U. The complete chloroplast genome sequence of purple mullein (Verbascum phoeniceum L.). Mitochondrial DNA Part B. 2020;5(1):819–20.
35. Thulin M: Flora of Somalia: Volume 3: Royal Botanic Gardens, Kew. Published on the Internet; http://wcsp.science.kew.org/ Retrieved. 2006.
36. IPNI. International Plant Names Index. Published on the Internet: http://www.ipni.org; The Royal Botanic Gardens, Kew, Harvard University Herbaria & Libraries and Australian National Botanic Gardens. Retrieved 03 August 2022.
37. Thulin M. Flora of Somalia: Volume 3. Royal Botanic Gardens, 2006.
38. Sotoodeh A. Histoire biogéographique et évolution des genres Verbascum et Artemisia en Iran à l’aide de la phylogénie moléculaire. Université Toulouse III-Paul Sabatier. Université de Toulouse; 2015.
39. Besnard G. Rubio de Casas R, Christin A-P, Vargas P. Phylogenetcs of Olea (Oleaceae) based on plastid and nuclear ribosomal DNA.
sequences: tertiary climatic shifts and lineage differentiation times. Ann Bot. 2009;104(1):143–60.

38. Wanga VO, Dong X, Oulo MA, Mkala EM, Yang J-X, Onjalilaina GE, Gichua MK, Kirka PM, Gituru RW, Hu G-W. Complete Chloroplast Genomes of Acanthochlamys bracteata (China) and Xerophyta (Africa) (Velloziaceae): Comparative Genomics and Phylogenomic Placement. Front Plant Sci. 2021;12:691833.

39. Choi KS, Chung MG, Park S. The complete chloroplast genome sequences of three Veronicaceae species (Plantaginaceae): comparative analysis and highly divergent regions. Front Plant Sci. 2016;7:355.

40. Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, Mei W, Thomas PT, Gardner MF, Ennos RA. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in Araucaria? Mol Ecol Resour. 2015;15(5):1067–78.

41. Kyalo CM, Li Z-Z, Mkila EM, Malombe J, Hu G-W, Wang Q-F. The first glimpse of Steptocarpus ionanthus (Gesneriaceae) phylogenomics: Analysis of five subspecies’ chloroplast genomes. Plants. 2020;9(4):456.

42. Wang W, Hu H, Wang J, Lei W, Gao J, Qiu X, Wang J. The complete chloroplast genome sequences of the medicinal plant Forsythia suspensa (Oleaceae). Int J Mol Sci. 2017;18(11):2288.

43. Nei M. and S. Kumar. Molecular evolution and phylogenetics: Oxford University Press, 2000.

44. Nielsen R. Molecular signatures of natural selection. Annu Rev Genet. 2005;39(1):171–201.

45. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 2007;24(8):1856–91.

46. Yang Z, Wong WSW, Rasmus N. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005;4:1107–18.

47. Wickie S, Schneewess GM, Depamwills CW, Muller KF, Qandu D. The evolution of the plastid chromosome in land plants: gene content, order, gene function, plant Mol Biol. 2011;76(3):273–97.

48. Zhang Y, Du L, Liu A, Chen J, Wu L, Hu W, Zhang W, Kim K, Lee S-C, Yang T-J. The complete chloroplast genome sequences of five Epimedium species: lights into phylogenetic and taxonomic analyses. Front Plant Sci. 2016;7:306.

49. Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. The first complete chloroplast genome sequences in Actinidiaceae: genus structure and comparative analysis. PLoS ONE. 2015;10(6):e0129347.

50. Munyao JN, Dong X, Yang J-X, Mbandi EM, Wanga VO, Oulo MA, Saina JK, Li Z-Z, Gichira AW, Liao Y-Y. The complete chloroplast genome sequences of five Epimedium species: lights into phylogenetic and taxonomic analyses. Front Plant Sci. 2021;11:3390.

51. Luo J, Hou B-W, Niu Z-T, Liu W, Xue Q-Y, Ding X-Y. Comparative chloroplast genomes of photosynthetic orchids: insights into the evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. PLoS ONE. 2014;9(6):e99016.

52. Rono PC, Dong X, Yang J-X, Mutie FM, Oulo MA, Malombe I, Kirka PM, Hu G-W, Wang Q-F. Initial complete chloroplast genomes of Alchemilla (Rosaceae): comparative analysis and phylogenetic relationships. Front Genet. 2021;12:11390.

53. Sun Y-x, Moore MJ, Meng A-p, Solits PS, Solits DE, Li J-q, Wang H-c. Complete plastid genome sequencing of Tribocandaleae reveals a significant expansion of the inverted repeat and suggests a Paleogene divergence between the two extant species. PLoS ONE. 2013;8(4):e60929.

54. Lei W, Ni D, Wang Y, Shao J, Wang X, Yang D, Wang J, Chen H, Liu C. Intraspecific and heteroplasmic variations, gene losses, and inversions in the chloroplast genome of Astragalus membranaceus. Sci Rep. 2016;6(1):1–13.

55. Saina JK, Li Z-Z, Gichira AW, Liao Y-Y. The complete chloroplast genome sequence of the medicinal plant Salvia miltiorrhiza. PLoS ONE. 2013;8(2):e57607.

56. Shapiro JA, Von SR. Why repetitive DNA is essential to genome function. Riv Biol Camb Philos Soc. 2010;80(2):227–50.

57. Yang J, Kang G-H, Pak J-H, Kim S-C. Characterization and comparison of two complete plastomes of Rosaceae species (Potentilla dichnisii var. glabrata and Spiraea insularis) endemic to Ulleung Island, Korea. Int J Mol Sci. 2020;21(14):4933.

58. Apa Menezes, et al. Chloroplast genomes of Byrsonima species (Malpighiaceae): comparative analysis and screening of high divergence sequences. Sci Rep. 2018;8(1):2210.

59. Zhou T, Ruhamia M, Wang J, Zhu H, Li W, Zhang X, Xu Y, Xu F, Wang X. The complete chloroplast genome of Euphorbia regeli, pseudogenization of ndh genes and the phylogenetic relationships within Orobancheae. Front Genet. 2019;10:444.

60. Oulo MA, Yang J-X, Dong X, Wanga VO, Mkala EM, Munyao JN, Onjolo VO, Rono PC, Hu G-W, Wang Q-F. Complete chloroplast genome of Rhipsalis baccifera, the only cactus with natural distribution in the old world: Genome rearrangement, intron gain and loss, and implications for phylogenetic studies. Plants. 2020;9(8):979.

61. Schäferhoff B, et al. Towards resolving Lamiales relationships: insights from rapidly evolving chloroplast sequences. BMC Evol Biol. 2010;10(1):352.

62. Wortley AH, Rudall PJ, Harris DJ, et al. How much data are needed to resolve a difficult phylogeny? case study in lamiaceae. Syst Biol. 2005;54(5):697–709.

63. Albach DC, Yan K, Jensen SR, et al. Phylogenetic placement of Trianeophora (formerly Scrophulariaceae) with some implications for the phylogeny of Lamiaceae. Taxon. 2009;58(3):749–56.

64. Refullo-Rodriguez NF, Olmstead RG. Phylogeny of Lamiidae. Am J Bot. 2014;101(2):287–99.

65. Chi X, Wang J, Gao Q, Zhang F, Chen S. The complete chloroplast genomes of two Lantana species with comparative analysis. Molecules. 2018;23(3):602.

66. Chase MW, Christenhusz M, Fay M, Byng J, Judd WS, Solits D, Mabberley D, Sennikov A, Solits PS, Stevens PF. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181(1):1–20.

67. Reveal J. Summary of recent systems of angiosperm classification. Kew Bull. 2011;66(1):1–48.

68. Barba-Montoya J, Dos Reis M, Schneider H, Donoghue PC, Yang Z. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. New Phytol. 2018;218(2):819–34.

69. Li H-T, Yi T-S, Gao L-M, Ma P-F, Zhang Z, Yang J-B, Gitzendanner MA, Fritsch PW, Cai J, Luo Y. Origin of angiosperms and the puzzle of the Jurassic gap. Nature plants. 2019;5(5):461–70.
