# Sequencing, de novo assembly of *Ludwigia* plastomes, and comparative analysis within the Onagraceae family

Barloy-Hubler F.[3], Le Gac A.-L.[1], Boury C.[2], Guichoux E.[2], and Barloy D.[1*]

1-DECOD (Ecosystem Dynamics and Sustainability), Institut Agro, IFREMER, INRAE, Rennes, France.

2- Université de Bordeaux, INRAE, BIOGECO, Cestas, France

3- Université de Rennes 1, CNRS, UMR 6553 ECOBIO, Rennes, France.

* Corresponding author: dominique.barloy@agrocampus-ouest.fr

**Abstract**

The Onagraceae family, which belongs to the order Myrtales, consists of approximately 657 species and 17 genera. This family includes the genus *Ludwigia* L., which is comprised of 82 species. In this study, we focused on the two aquatic invasive species *Ludwigia grandiflora* subsp. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp *montevidensis* (*Lpm*) largely distributed in aquatic environments in North America and in Europe. Both species have been found to degrade major watersheds leading ecological and economical damages. Genomic resources for Onagraceae are limited, with only *Ludwigia octovalvis* (Lo) plastid genome available for the genus *Ludwigia* L. at the time of our study. This scarcity constrains phylogenetic, population genetics, and genomic studies. To brush up genomic ressources, new complete plastid genomes of *Ludwigia grandiflora* subps. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp. *montevidensis* (*Lpm*) were generated using a combination of MiSeq (Illumina) and GridION (Oxford Nanopore) sequencing technologies. These plastomes were then compared to the published *Ludwigia octovalvis* (*Lo*) plastid genome, which was re-annotated by the authors. We initially sequenced and assembled the chloroplast (cp) genomes of *Lpm* and *Lgh* using a hybrid strategy combining short and long reads sequences. We observed the existence of two *Lgh* haplotypes and two potential *Lpm* haplotypes. *Lgh*, *Lpm,* and *Lo* plastomes were similar in terms of genome size (around 159 Kb), gene number, structure, and inverted repeat (IR) boundaries, comparable to other species in the Myrtales order. A total of 45 to 65 SSRs (simple sequence repeats), were detected, depending on the species, with the majority consisting solely of A and T, which is common among angiosperms. Four chloroplast genes (*matK*, *accD*, *ycf2* and *ccsA)* were found under positive selection pressure, which is commonly associated with plant development, and especially in aquatic plants such as *Lgh,* and *Lpm.* Our hybrid sequencing approach revealed the presence of two *Lgh* plastome haplotypes which will help to advance phylogenetic and evolutionary studies, not only specifically for *Ludwigia*, but also the Onagraceae family and Myrtales order. To enhance the robustness of our findings, a larger dataset of chloroplast genomes would be beneficial.

**Keywords**

Water primrose, *Ludwigia* sp., Onagraceae, chloroplast genome, long and short reads, hybrid assembly, haplotype

## Introduction

The Onagraceae family belongs to the order Myrtales which includes approximately 657 species of herbs, shrubs, and trees across 17 genera grouped into two subfamilies: subfam. Ludwigioideae W. L. Wagner and Hoch, which only has one genus (*Ludwigia* L.), and subfam. Onagroideae which contains six tribes and 21 genera [1]. *Ludwigia* L. is composed of 83 species[2][3] . The current classification for *Ludwigia* L., which are composed of several hybrid and/or polyploid species, lists 23 sections. A recent molecular analysis is clarified and supported several major relationships in the genus but has challenged the complex sectional classification of *Ludwigia* L.[4].

The diploid species *Ludwigia peploides* (Kunth) Raven subsp. *montevidensis* (Spreng.) [5] (named here *Lpm*) (2n=16), and the decaploid species, *Ludwigia grandiflora* (Michx) Greuter & Burdet subsp. *hexapetala* (Hook. & Arn) Nesom & Kartesz (named here *Lgh*) (2n=80), reproduce essentially by clonal propagation, which suggests that there is a low genetic diversity within the species [6]. *Lgh* and *Lpm* are native to South America and are considered as one of the most aggressive aquatic invasive plants [7]. Largely distributed in aquatic environments in North America and in Europe [8], both species have been found to degrade major watersheds as well as aquatic and riparian ecosystems [9] leading ecological and economical damages. In France, both species occupied aquatic habitats, such as static or slow-flowing waters, riversides, and have recently been observed in wet meadows [10]. The transition from an aquatic to a terrestrial habitat has led to the emergence of two *Lgh* morphotypes [11]. The appearance of metabolic and morphological adaptations could explain the ability to acclimatize to terrestrial conditions, and this phenotypic plasticity involves various genomic and epigenetic modifications [12].

Adequate genomic resources are necessary in order to be identify the genes and metabolic pathways involved in the adaptation process leading to plant invasion [13] with genomic information making it possible to predict and control invasiveness [14]. However, even though the number of terrestrial plant genomes has increased considerably over the last 20 years, only a small fraction (~ 0.16%) have been sequenced, with some clades being significantly more represented than others [15]. Thus, for the Onagraceae family (which includes *Ludwigia* sp.), only a handful of chloroplast sequences (plastomes) are available, and the complete genome has not yet been sequenced. If *Lpm* is a diploid species (2n=2x=16) with a relatively small genome size (262 Mb), *Lgh* is a decaploid species (2n=10x=80) with a large size genome of 1419 Mb [16]. Obtaining a reference genome for these two non-model species without having a genome close to the *Ludwigia* species is challenging and development of plastome and/or

mitogenome will be a first step to generate genomic resource. As of April 2023, there are 10,712 reference plastomes listed on GenBank (Release 255: April 15 2023), with the vast majority (10,392 genomes) belonging to Viridiplantae (green plants). However, in release 255, the number of plastomes available for the Onagraceae family is limited, with only 36 plastomes currently listed. Among these, 15 plastomes are from the tribe Epilobieae, with 11 in the Epilobium genus and 4 in the Chamaenerion genus. Additionally, there are 23 plastomes from the tribe Onagreae, with 17 in the Oenothera genus, 5 in the Circaea genus, and only one in the *Ludwigia* genus. The *Ludwigia octovalvis* chloroplast genome was released in 2016 as a unique haplotype of approximately 159 kb [17]. *L. octovalvis* belongs to sect. *Macrocarpon* (Micheli) H.Hara while *Lpm* and *Lgh* belong to *Jussieae* section [18][19]. Generally, the inheritance of chloroplast genomes is considered to be maternal in angiosperms. However, biparentally inherited chloroplast genomes could potentially exist in approximately 20% of angiosperm species [20][21]. Both maternal and biparental inheritance are described in the Onagraceae family. In tribe Onagreae, *Oenothera* subsect. *Oenothera* are known to have biparental plastid inheritance [22][23]. In tribe Epilobieae, biparental plastid inheritance was also reported in *Epilobium* L. with mainly maternal transmission, and very low proportions of paternally transmitted chloroplasts [24].

The chloroplast is the symbolic organelle of plants and plays a fundamental role in photosynthesis. Chloroplasts evolved from cyanobacteria through endosymbiosis and thereby inherited components of photosynthesis reactions (photosystems, electron transfer and ATP synthase) and gene expression systems (transcription and translation)[25]. In general, chloroplast genomes (plastomes) are highly conserved in size, structure, and genetic content. They are rather small (120-170 kb,[26]), with a quadripartite structure comprising two long identical inverted repeats (IR, 10–30 kb) separated by large and a small single copy regions (LSC and SSC, respectively). They are also rich in genes, with around 100 unique genes encoding key proteins involved in photosynthesis, and a comprehensive set of ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs)[27]. Plastomes are generally circular but linear shapes also exist [28]. Chloroplast DNA usually represents 5-20% of total DNA extracted from young leaves and therefore low-coverage whole genome sequencing can generate enough data to assemble an entire chloroplast genome [29].

If we refer to their GenBank records, more than 95% of these plastomes were sequenced by so-called short read techniques (mostly Illumina). However, in most seed plants, the plastid genome exhibits two large inverted repeat regions (60 to 335 kb,[29]), which are longer than the short read lengths (< 300 bp). This leads to incomplete or approximate assemblies [30].

113 Recent long-read sequencing (> 1000 bp) provides compelling evidence that terrestrial plant
114 plastomes exhibit two structural haplotypes. These haplotypes are present in equal proportions
115 and differ in their inverted repeat (IR) orientation [31]. This shows the importance of using the
116 so-called third generation sequence (TGS, PacBio or Nanopore) to correctly assemble the IRs
117 of chloroplasts and to identify any different structural haplotypes. The current problem with
118 PacBio or Nanopore long read sequencing is the higher error rate compared to short read
119 technology [32][33][34]. Thus, a hybrid strategy which combines long reads (to access the
120 genomic structure) and short reads (to correct sequencing errors) could be effective [30][35].
121 Here, we report the newly sequenced complete plastid genomes of *Ludwigia grandiflora*
122 subps. *hexapetala* (*Lgh*) and *Ludwigia peploides* subsp *montevidensis* (*Lpm*), using a
123 combination of different sequencing technologies, as well as a re-annotated comparative
124 genomic analysis of the published *Ludwigia octovalvis* (*Lo*) plastid. The main objectives of this
125 study are (1) to assemble and annotate the plastomes of two new species of *Ludwigia* sp., (2) to
126 reveal the divergent sequence hotspots of the plastomes in this genus and in the Onagraceae (3)
127 to identify the genes under positive selection.
128 To achieve this, we utilized long read sequencing data from Oxford Nanopore and short read
129 sequencing data from Illumina to assemble the *Lgh* plastomes and compared these assemblies
130 with those obtained solely from long reads of *Lpm*. We also compared both plastomes to the
131 published plastome of *Lo*. Our findings demonstrated the value of *de novo* assembly in reducing
132 assembly errors and enabling accurate reconstruction of full heteroplasmy. We also evaluated
133 the performance of a variety of software for sequence assembly and correction in order to define
134 a workflow that will be used in the future to assemble *Ludwigia* sp. mitochlondrial and nuclear
135 genomes. Finally, the three new *Ludwigia* plastomes generated by our study make it possible
136 to extend the phylogenetic study of the Onagraceae family and to compare it with previously
137 published analyses [4][36][37].
138

139 **Material and Methods**

140 **Plant sampling and experimental design**

141 The original plant materials were collected in June of 2018 near to Nantes (France) and
142 formal identified by D. Barloy. *L. grandiflora* subsp. *hexapetala (Lgh)* plants were taken from
143 the Mazerolles swamps (N47 23.260, W1 28.206), and *L. peploides* subsp. *montevidensis* (*Lpm*)
144 plants from La Musse (N 47.240926, W -1.788688)*)*. Plants were cultivated in a growth
145 chamber in a mixture of $^{1}/_{3}$ soil, $^{1}/_{3}$ sand, $^{1}/_{3}$ loam with flush water level, at 22°C and a 16 h/8

146 h (light/dark) cycle. A single stem of 10 cm for each species was used for vegetative
147 propagation in order to avoid potential genetic diversity. *De novo* shoots, taken three
148 centimeters from the apex, were sampled for each species. Samples for gDNA extraction were
149 pooled and immediately snap-frozen in liquid nitrogen, then lyophilized over 48 h using a
150 Cosmos 20K freeze-dryer (Cryotec, Saint-Gély-du-Fesc, France) and stored at room
151 temperature. All the plants were destroyed after being used as required by French authorities
152 for invasive plants (article 3, prefectorial decree n°2018/SEE/2423).

153 Due to high polysaccharide content and polyphenols in *Lpm* and *Lgh* tissues and as no
154 standard kit provided good DNA quality for sequencing, genomic DNA extraction was carried
155 out using a modified version of the protocol proposed by Panova et al in 2016, with three
156 purification steps [38].

157 40 mg of lyophilized buds were ground at 30 Hz for 60 s (Retsch MM200 mixer mill,
158 FISHER). The ground tissues were lysed with 1 ml CF lysis buffer (MACHEREY-NAGEL)
159 supplemented with 20 µl RNase and incubated for 1 h at 65°C under agitation. 20 µl proteinase
160 K was then added before another incubation for 1 h at 65°C under agitation. To avoid breaking
161 the DNA during pipetting, the extracted DNA was recovered using a Phase-lock gel tube as
162 described in Belser [39]. The extracts were transferred to 2 ml tubes containing phase-lock gel,
163 and an equal volume of PCIA (Phenol, Chloroform, Isoamyl Alcohol; 25:24:1) was added.
164 After shaking for 5 min, tubes were centrifuged at 11000 g for 20 min. The aqueous phase was
165 transferred into a new tube containing phase-lock gel and extraction with PCIA was repeated.
166 DNA was then precipitated after addition of an equal volume of binding buffer C4
167 (MACHEREY-NAGEL) and 99% ethanol overnight at 4°C or 1 h in ice then centrifuged at 800
168 rpm for 10 min. After removal of the supernatant, 1 ml of CQW buffer was added then the
169 pellet of DNA was re-suspended. Next, DNA purification was carried out by adding a 2 ml
170 mixture of wash buffer PW2 (MACHEREY-NAGEL), wash buffer B5 (MACHEREY-
171 NAGEL), and ethanol at 99% in equal volumes, followed by centrifugation at 800 rpm for 10
172 min. This DNA purification step was carried out twice. Finally, the DNA pellet was dried in
173 the oven at 70°C for 30 min then re-suspended in 100 µl elution buffer BE (MACHEREY-
174 NAGEL) (5 mM Tris solution, pH 8.5) after 10 min incubation at 65°C under agitation.

175 A second purification step was performed using a PCR product extraction from gel agarose
176 kit from Macherey-Nagel (MN) NucleoSpin® Gel and PCR Clean-up kit and restarting the
177 above protocol from the step with the addition of CQW buffer then PW2 buffer.

178 The third purification step consisted of DNA purification using a Macherey-Nagel (MN)
179 NucleoMag kit for clean-up and size selection. Finally, the DNA was resuspended after a 5 min
180 incubation at 65°C in 5 mM TRIS at pH 8.5.
181 The quantity and quality of the gDNA was verified using a NanoDrop spectrometer,
182 electrophoresis on agarose gel and ethidium bromide staining under UV light and Fragment
183 Analyzer (Agilent Technologies) of the University of Rennes1.

184 **Library preparation and sequencing**

185 MiSeq (Illumina) and GridION (Oxford Nanopore Technologies, referred to here as
186 ONT) sequencing were performed at the PGTB (doi:10.15454/1.5572396583599417E12). *Lgh*
187 and *Lpm* genomic DNA were re-purified using homemade SPRI beads (1.8X ratio). *Lgh* has a
188 large genome size of 1419 Mb, 5-fold larger than *Lpm* genome 262 Mb [16]. SR (Illumina, one
189 run) and LR (Oxford Nanopore, three runs) sequencing were therefore carried out for *Lgh* and
190 only LR sequencing for *Lpm* (one run). For Illumina sequencing, 200 ng of *Lgh* DNA was used
191 according to the QIAseq FX DNA Library Kit protocol (Qiagen). The final library was checked
192 on TapeStation D5000 screentape (Agilent Technologies) and quantified using a QIAseq
193 Library Quant Assay Kit (Qiagen). The pool was sequenced on an Illumina MiSeq using V3
194 chemistry and 600 cycles (2x300bp). For ONT sequencing, around 8 µg of *Lgh* and *Lpm* DNA
195 were size selected using a Circulomics SRE kit (according to the manufacturer's instructions)
196 before library preparation using a SQK-LSK109 ligation sequencing kit following ONT
197 recommendations. Basecalling in High Accuracy - Guppy version: 4.0.11 (MinKNOW
198 GridION release 20.06.9) was performed for the 48 h of sequencing. Long reads (LR) and short
199 reads (SR) were available for *Lgh* and only LR for *Lpm*.

200 **Chloroplast assemblies**

201 Quality controls and preprocessing of sequences were conducted using Guppy v4.0.14 for
202 long reads (via Oxford Nanopore Technology Client access) and fastp v0.20.0 [40] for short
203 reads, using Q15, since increasing the Phred quality to 20 or higher has no effect on the number
204 of sequences retained (66%). A preliminary draft assembly was performed using *Lgh* short-
205 reads (SR, 2*23,067,490 reads) with GetOrganelle v1.7.0 [41] and NOVOPlasty v4.2.1[42],
206 and chloroplastic short and long reads were extracted by mapping against this draft genome.
207 Chloroplastic short reads were then *de novo* assemble using Velvet (version 1.2.10) [43],
208 ABySS (version 2.1.5 [44][45]), MEGAHIT (1.1.2,[46]), and SPAdes (version 3.15.4,[47]),
209 without and with prior error correction. The best k-mer parameters were tested using kmergenie
210 [48] and k=99 was found to be optimal. For ONT reads, *Lgh* (550,516 reads) and *Lpm* (68,907
211 reads) reads were self-corrected using CANU 1.8 [49] or SR-corrected using Ratatosk [50] and

212 *de novo* assembly using CANU [49] and FLYE 2.8.2 [51] run with the option --meta and –
213 plasmids. For all these assemblers, unless otherwise specified, we used the default parameters.

214 **Post plastome assembly validation**

215 As we used many assemblers and different strategies, we produced multiple contigs that
216 needed to be analyzed and filtered in order to retain only the most robust plastomes. For that,
217 all assemblies were evaluated using the QUality ASsessment Tool (QUAST) for quality
218 assessment [52] and visualized using BANDAGE [53], both using default parameters.
219 BANDAGE compatible graphs (.gfa format) were created with the megahit_toolkit for
220 MEGAHIT [46] and with gfatools for ABySS [45]. Overlaps between fragments were manually
221 checked and ambiguous "IUPAC or N" nucleotides were also biocured with Illumina reads
222 when available.

223 **Chloroplast genome annotation**

224 Plastomes were annotated via the GeSeq [54] using ARAGORN and tRNAscan_SE to
225 predict tRNAs and rRNAs and tRNAscan_SE to predict tRNAs and rRNAs and via Chloe
226 prediction site [55]. The previously reported *Lo* chloroplast genome was also similarly re-
227 annotated to facilitate genomic comparisons. Gene boundaries, alternative splice isoforms,
228 pseudogenes and gene names and functions were manually checked and biocurated using
229 Geneious (v.10). Finally, plastomes were represented using OrganellarGenomeDRAW
230 (OGDRAW)[56]. These genomes were submitted to GenBank at the National Center of
231 Biotechnology Information (NCBI) with specific accession numbers (for *Lgh* haplotype 1,
232 (LGH1) OR166254 and *Lgh* haplotype 2, (LGH2) OR166255; for *Lpm* haplotype, (LPM)
233 OR166256) using annotation tables generated through GB2sequin [57].

234 **SSRs and Repeat Sequences Analysis**

235 Simple Sequence Repeats (SSRs) were analyzed through the MISA web server [58], with
236 parameters set to 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides,
237 respectively. Direct, reverse and palindromic repeats were identified using RepEx [59].
238 Parameters used were: for inverted repeats (min size 15 nt, spacer = local, class = exact); for
239 palindromes (min size 20 nt); for direct repeats (minimum size 30 nt, minimum repeat similarity
240 97%). Tandem repeats were identified using Tandem Repeats Finder[60], with parameters set
241 to two for the alignment parameter match and seven for mismatches and indels. The IRa region
242 was removed for all these analyses to avoid over representation of the repeats.

243 **Comparative chloroplast genomic analyses**

244 *Lgh* and *Lpm* plastomes were compared with the reannotated and biocurated *Lo* plastome
245 using mVISTA program [61], with the LAGAN alignment algorithm [62] and a cut-off of 70%
246 identity.
247 Nucleotide diversity (Pi) was analyzed using the software DnaSP v.6.12.01 [63] [64]with
248 step size set to 200 bp and window length to 300 bp. IRscope [65] was used for the analyses of
249 inverted repeat (IR) region contraction and expansion at the junctions of chloroplast genomes.
250 To assess the impact of environmental pressures on the evolution of these three *Ludwigia*
251 species, we calculated the nonsynonymous (Ka) and synonymous (Ks) substitutions and their
252 ratios ($\omega$ = Ks/Ks) using TBtools [66] to measure the selective pressure. Genes with $\omega < 1$, $\omega =$
253 1, and $1 < \omega$ were considered to be under purifying selection (negative selection), neutral
254 selection, and positive selection, respectively.

255 **Phylogenetic analysis of *Ludwigia* based on MatK sequences**

256 We performed a phylogenetic analysis on the *Ludwigia* genus using the MatK, only protein
257 coding barcode available for a large number of *Ludwigia* species. All MatK amino acid
258 sequences were aligned with the FFT-NS-2 (Fast Fourier Transform-based Narrow Search)
259 algorithm and BLOSUM62 scoring matrix using MAFFT 7 [67]. The phylogenetic tree analysis
260 was conducted using the rapid hill-climbing algorithm (command line: -f d) in RAxML 8.2.11
261 [68], with GAMMA JTT (Jones-Taylor-Thornton) protein model. Node support was assessed
262 through fast bootstrapping (-f a) with 1,000 non-parametric bootstrap pseudo-replicates.
263 *Circaea* MatK were selected as outgroup, and all accession numbers are indicated on the
264 phylogenetic tree labels.

265 **Graphic representation**

266 Statistical analyses were performed using R software in RStudio integrated development
267 environment (R Core Team, 2015, RStudio: Integrated Development for R. RStudio, Inc.,
268 Boston, MA, http://www.rstudio.com/). Figures were realized using ggplot2, ggpubr, tidyverse,
269 dplyr, gridExtra, reshape2, and viridis packages. SNPs were represented using trackViewer [69]
270 and genes represented using gggenes packages.
271

272 **Results**

273 **Plastome short read assembly**

274 The chloroplastic fraction of *Lgh* short reads (SR) was extracted by mapping against the two
275 draft haplotypes generated by GetOrganelle, which differ only by a "flip-flop" of the SSC region
276 (Figure 1). Since the assembly by NOVOplasty did not provide any additional information
277 compared to GetOrganelle, it was not retained. This subset (1,360,507 reads) was assembled

278 using ABySS, Velvet, MEGAHIT and SPAdes in order to identify the best assembler for this
279 plant model. As shown in Figure 2, both the number and size of contigs depend greatly on the
280 algorithms used and the correction step. The effect of prior read correction is notable for
281 MEGAHIT and Velvet, especially concerning the increase in the size of the large alignment
282 (Add. Figure 1A), loss of misassemblies, and reduction of the number of mismatches (Add.
283 Figure 1B). Investigating results via BANDAGE (Add. Figure 2), we observed that ABySS and
284 SPAdes suggest the tripartite structure with the long single-copy (LSC) region as the larger
285 circle in the graph (blue), joined to the small single-copy region (green) by one copy of the
286 inverted repeats (IRs, red), both IRs being collapsed in a segment of approximately twice the
287 coverage. For Velvet and MEGAHIT, graphs confirm the significant fragmentation of the
288 assemblies, which is improved by prior correction of the reads.
289 In conclusion, none of the short-read assemblers tested in our study produced a complete
290 plastome. The best result was achieved by SPAdes using corrected short reads (mean coverage
291 1900 X) to assemble a plastome consisting of three contigs: 90,272 bp (corresponding to LSC),
292 19,788 bp (corresponding to SSC), and 24,762 bp (corresponding to one of the two copies of
293 the IR).
294 **Plastome long read assembly**
295 Chloroplast fractions of *Lgh* long reads (28,882 reads) were assembled using CANU or
296 FLYE. With raw data, CANU generates a unique contig (NGA50 112648) corresponding to
297 haplotype 2, whereas FLYE makes two contigs (NGA50 133687) that reconstruct haplotype 1.
298 Self-corrected LR leads to fragmentation into two (CANU) or three (FLYE) contigs which both
299 reconstruct haplotype 1, with an large gap corresponding to one of the IR copies for CANU.
300 Finally, SR-correction by RATATOSK allows CANU to assemble two redundant contigs
301 reproducing haplotype 2 while FLYE makes two contigs corresponding to haplotype 1 (Add.
302 Figure 3A). In conclusion, the two *Lgh* haplotypes were reconstructed (average coverage 700X)
303 and the most complete and accurate hybrid assemblies (99.94% accuracy, Additional Figure
304 3B) were submitted to GenBank.
305 Unfortunately, due to the absence of short read data, we could only perform self-corrected
306 long read assembly for *Lpm* using CANU. We also compared CANU and FLYE assembler
307 efficiency, and found that assembly using CANU produces 13 contigs whereas FLYE produces
308 12 contigs. In both cases, only three contigs are required to reconstitute a complete cpDNA
309 assembly (no gap, no N), with an SSC region oriented like those of the *Lgh* haplotype 2 and the
310 *Lo* plastome. Although it is more than likely that these two SSC region orientations also exist
311 for *Lpm*, the low number of nanopore sequences generated (68907 reads) and absence of

312 Illumina short reads prevented us from demonstrating the existence of both haplotypes. As a
313 result, only the "haplotype 2" generated sequence was deposited to Genbank.

314 **Annotation and comparison of *Ludwigia* plastomes**

315 **1. General Variations**

316 Plastomes of the three species of *Ludwigia* sp., *Lgh*, *Lpm* and *Lo,* are circular double-
317 stranded DNA molecules (Figure 3) which are all (as shown in Table 1) approximately the same
318 size: *Lo* is 159,396 bp long, making it the smallest, while *Lgh* is the largest with 159,584 bp,
319 and *Lpm* is intermediate at 159,537 bp. The overall GC content is almost the same for the three
320 species (37.4% for *Lo*, 37.3 % for *Lgh* and *Lpm*) and the GC contents of the IR regions are
321 higher than those of the LSC and SSC regions (approximately 43.5 % compared to 35% and
322 *ca.*32% respectively). Between the three species, the lengths of the total chloroplasts, LSC,
323 SSC, and IR are broadly similar (approximately 90.2 kb for LSC, 19.8 kb for SSC and 24.8 kb
324 for IB, see details Table 1) and the three plastomes are perfectly syntenic if we orient the SSC
325 fragments the same way.

326 All three *Ludwigia* sp. plastomes contain the same number of functional genes (134 in total)
327 encoding 85 proteins (embracing 7 duplicated in the IR region: *ndh*B, *rpl2*, *rpl23*, *rps7*, *rps12*,
328 *ycf2*, *ycf15*), 37 tRNAs (including *trn*K-UUU which contains *mat*K), and 8 rRNAs (16S, 23S,
329 5S, and 4.5S as duplicated sets in the IR). Among these genes, 18 contain introns, of which six
330 are tRNAs (Table 2). Only the *rps12* gene is a trans-spliced gene. A total of 46 genes are
331 involved in photosynthesis, and 71 genes related to transcription and translation, including a
332 bacterial-like RNA polymerase and 70S ribosome, as well as a full set of transfer RNAs
333 (tRNAs) and ribosomal RNAs (rRNAs). Six other protein-coding genes are involved in
334 essential functions, such as *acc*D*,* which encodes the β-carboxyl transferase subunit of acetyl-
335 CoA carboxylase, an important enzyme for fatty acid synthesis; *mat*K encodes for maturase K,
336 which is involved in the splicing of group II introns; *cem*A, a protein located in the membrane
337 envelope of the chloroplast is involved in the extrusion of protons and thereby indirectly allows
338 the absorption of inorganic CO2 in the plastids; *clp*P1 which is involved in proteolysis, and;
339 *ycf1*, *ycf2*, two ATPases members of the TIC translocon. Finally, a highly pseudogenized *ycf15*
340 locus was annotated in the IR even though premature stop codons indicate loss of functionality.

341 **2. Segments Contractions/Expansion**

342 The junctions between the different chloroplast segments were compared between three
343 *Ludwigia* sp. (*Lpm*, *Lgh* and *Lo),* and we found that the overall resemblance of *Ludwigia* sp.
344 plastomes was confirmed at all junctions (Figure 4A). In all three genomes, *rpl22, rps19, and*
345 *rpl2* were located around the LSC/IRb border, and *rpl2*, *trn*H, and *psb*A were located at the

346 IRa/LSC edge. The JSB (junction between IRb and SSC) is either located in the *ndh*F gene or
347 the *ycf1* gene depending on the orientation of the SSC region (Figure 4B). The *ycf1* gene was
348 initially annotated as a 1139 nt pseudogene that we biocurate as a larger gene (5302 nt) with a
349 frameshift due to a base deletion, compared to *Lg* and *Lo* which both carry a complete *ycf1*
350 gene.
351 If we compare *Ludwigia* sp. chloroplastic LSC/SCC/IR junctions (via IRscope) with
352 representative Onagraceae plastomes of *Chamaenerion sp. conspersum* (MZ353638) and *sp.*
353 *angustifolium* (NC_052848)**,** *Circaea sp. cordata* (NC_060876) and *sp. alpina* (NC_061010),
354 *Epilobium amurense* (NC_061015) and *Oenothera villosa* subsp. *strigosa* (NC_061365) and
355 *Oenothera lindheimeri* (MW538951) (Figure 5), we can observe that the gene positions at the
356 JLB (junction of LSC/IRb) and JLA (junction of IRa/LSC) boundary regions are well-preserved
357 throughout the entire family, whereas those at the JSB and JSA regions differ. Concerning JSB
358 (junction of IRb/SSC), in the five Onagraceae genera studied, *ndh*F is duplicated, with the
359 exception of *Circaea* sp. and *Ludwigia* sp. For *Oenothera villosa*, the first copy of *ndh*F, which
360 is located in the IRb, overlaps the JSB border, whereas for *Oenothera lindheimeri*, *Epibolium*
361 *amurense* and Chamaenerion sp., *ndh*F is only located in inverted repeats. Only *Circaea* sp. and
362 *Ludwigia* sp. have a unique copy of this locus, and it is found in the SSC segment (Figure 5).
363 At the JSA border (junction of SSC/Ira), in *Circaea* sp., the *ycf1* gene crosses the IRa/SSC
364 boundary and extends into the IRa region.
365 When comparing the respective sizes of chloroplast fragments (IR/SSC/LSC) in Onagraceae, it
366 can be observed that *Ludwigia* species exhibit expansions in the SSC and LSC regions which
367 are not compensated by significant contractions in the IR regions. This is likely due to the
368 relocation of the *ndh*F in the SSC region and *rps19* in the LSC region. Additionally, there may
369 be significant size variations in the intergenic region between *trnI* and *ycf2*, as well as the
370 intergenic segment containing the *ycf15* pseudogene (Add. Figure 4).

371 **3. Repeats and SSRs analysis**

372 In this study, we analyzed the nature and distribution of single sequence repeats (SSR), as
373 their polymorphism is an interesting indicator in phylogenetic analyses. A total of 65 (*Lgh*), 48
374 (*Lpm*) and 45 (*Lo*) SSRs were detected, the majority being single nucleotide repeats (38–21),
375 followed by tetranucleotides (12–10) and then di-, tri- and penta-nucleotides (Add. Figure 5A).
376 Mononucleotide SSRs are exclusively composed of A and T, indicating a bias towards the use
377 of the A/T bases, which is confirmed for all SSRs (Add. Figure 5B). In addition, the SSRs are
378 mainly distributed in the LSC region for the three species, which is probably biased by the fact
379 that LSC is the longest segment of the plastome (Add. Figure 5C). The analysis of SRR

380 locations revealed that most were distributed in non-coding regions (intergenic regions and
381 introns, Add. Figure 5D).
382 The chloroplast genomes of the three *Ludwigia* species were also screened for long repeat
383 sequences. They were counted in a non-redundant way (if smaller repetitions were included in
384 large repeats, only the large ones were considered). Four types of repeats (tandem, palindromic
385 inverted and direct) were surveyed in the three *Ludwigia* sp. plastomes. No inverted repeats
386 were detected with the criteria used.
387 For the three other types of repeats, here are their distributions:
388 ***Tandem repeats* (Table 3A)**: Perfect tandem repeats (TRs) with more than 15 bp were
389 examined. Twenty-two *loci* were identified in the three *Ludwigia* sp. plastomes (*Lgh*, *Lpm*, *Lo*),
390 heterogeneously distributed as shown in Table 3A: 13 loci (plus one imperfect) in *Lo*, nine loci
391 (plus one imperfect) in *Lgh* and seven loci (plus two imperfect) in *Lpm*. It can therefore be seen
392 that the TR distributions (occurrence and location) are specific to each plastome, since only
393 four pairs are common to the three species. Thus, nine TRs are unique to *Lo*, three to *Lpm* and
394 three to *Lgh.* Two pairs are common to *Lgh* and *Lpm* and one is common to *Lo* and *Lgh.* TRs
395 are mainly intergenic or intronic but are detected in two genes (*acc*D and *ycf1)*. These genes
396 have accelerated substitution rates, although this does not generate a large difference in their
397 lengths. This point will be developed later in this article.
398 ***Direct repeats* (Table 3B)**: There are few direct (non-tandem) repeats (DRs) in the
399 chloroplast genomes of *Ludwigia* sp. A single direct repeat of 41 nt is common to the three
400 species, at 2 kb intervals, in *psaB* and *psaA* genes. This DR corresponds to an amino acid repeat
401 [WLTDIAHHHLAIA] which corresponds to a region predicted as transmembrane. We then
402 observe three direct repeats conserved in *Lpm* and *Lgh* in *ycf1*, *acc*D and *clp*P1 respectively,
403 two unique DRs in *Lo* (in the *acc*D gene and *rps12-clp*P1 intergene) and one in *Lgh* (in the
404 *clp*P1 intron 1 and *clp*P1 intron 2).
405 ***Palindromes* (Table 3C)**: Palindromic repeats make up the majority of long repetitions,
406 with the numbers of perfect repeats varying from 19, 24 and 26 in *Lo*, *Lgh* and *Lpm*,
407 respectively, and the number of quasi-palindromes (1 mutation) varying between 8, 3 and 6.
408 They are mainly found in the intronic and intergenic regions, with the exception of six genic
409 locations in *psb*D, *ndh*K, *ccs*A and *rpl22,* and two palindromic sequences in *ycf2.* These gene
410 palindromic repeats do not seem to cause genetic polymorphism in *Ludwigia* and can be
411 considered as silent.
412 Thirteen palindromes are common to the three species (including 2 with co-variations in
413 *Lo*). 13 others present in *Lpm* and *Lgh* correspond to quasi-palindromes (QPs) in *Lo* due to

414 mutated bases, and conversely, three *Lo* perfect palidromes are mutated in *Lpm* and *Lgh*.
415 Finally, only five palindromes are species specific. Two in particular are located in the
416 hypervariable intergenic spacer *ndh*F-*rpl32*, and are absent in *Lo* due to a large deletion of 160
417 nt.

418 **4. Repeat distribution in LSC, SSC and IR segments**

419 In the IRa/IRb regions, repeats are only identified in the first 9 kb region between *rpl2* and
420 *ycf2*: a tandem repeat in the *Lpm rpl2* intron, and a tetranucleotide repeat, [TATC]*3, located
421 in the *ycf2* gene in the 3 species. In *ycf2* we also found 1 common palindrome (16 nt), a single
422 palindrome in *Lo* (20 nt, absent following an A:G mutation in the 2 other species), as well as a
423 shared tandem repeat (24 nt), and an additional 15 nt tandem repeat in *Lo* which adds 4 amino
424 acids to protein sequence.

425 In the SSC region, the repeats are almost all located in the intergenic and/or intronic
426 regions, with a hotspot between *ndh*F and *ccs*A. There is also a shared microsatellite in *ndh*F,
427 and a palidrome (16 nt) in *ccs*A which is absent in *Lo* (due to an A:C mutation), resulting in a
428 synonymous mutation (from isoleucine to leucine). We also observed multiple and various
429 repeats in the *ycf1* gene: 3 common poly-A repeats (from 10 to 13 nt), 3 species-specific
430 microsatellites (ATAG)*3 and (ACCA)*4 in *Lgh* and (CAAC)*3 in *Lo,* as well as two direct
431 repeats of 32 nt (37 nt spacing), which were absent from *Lo* due to a G:T SNP. Two tandem
432 repeats were also observed in *Lo* and *Lgh*. Neither of these repeats are at the origin of the
433 frameshift causing the pseudogenization of *ycf1* in *Lo*, this latter being due to a single deletion
434 of an A at position 3444 of the gene.

435 Finally, in the LSC region, the longest segment, which consequently contains the maximum
436 number of repeats, we still observed a preferential localization in the intergenic and intronic
437 regions since only genes *atp*A, *rpo*C2, *rpo*B, *psb*D, *psb*A, *psb*B, *ndh*K and *clp*P1 contain either
438 mononucleotic repeats (poly A and T), palindromes, or microsatellites (most often common to
439 the three species and without affecting the sequences of the proteins produced). As mentioned
440 earlier, the only exception is the *acc*D gene, which contains several direct and tandem repeats
441 in *Lgh* and *Lpm,* corresponding to a region of 174 nt (58 amino acids) missing in *Lo* and,
442 conversely, a direct repeat of 40 nucleotides, in a region of 147 nt (49 aa), which is present in
443 *Lo* and missing in the other two species. These tandem repeats lead to the presence of four
444 copies of 9 amino acids [DESENSNEE] in *Lgh* and *Lpm*, two of which form a larger duplication
445 of 17 aa [FLSDSDIDDESENSNEE]. Similarly, the TRs present only in *Lo* generate two perfect
446 9 amino acid repeats [EELSEDGEE], included in two longer degenerate repeats of 27 nt (Add.
447 Figure 6). It should be noted that though these TRs do not disturb the open reading phases, it is

448 still possible for them to form an intron which is not translated. Different functional studies will
449 be necessary to clarify this point. The presence of polymorphisms of the *acc*D gene between
450 *Lo* and the two species (*Lpm*, *Lgh*) is interesting because *acc*D, that encodes a subunit of acetyl-
451 CoA carboxylase (EC 6.4.1.2). This enzyme is essential in fatty acid synthesis and also
452 catalyzes the synthesis of malonyl-CoA, which is necessary for the growth of dicots, plant
453 fitness and leaf longevity, and is involved in the adaptation to specific ecological niches [70].
454 Large *acc*D expansions due to TRs have also been described in other plants such as *Medicago*
455 [71] and *Cupressophytes* [72]. Some authors have suggested that these inserted repeats are not
456 important for acetyl-CoA carboxylase activity as the reading frame is always preserved, and
457 they assume that these repeats must have a regulatory role [73].

458 **5. Sequence Divergence Analysis and Polymorphic Loci Identification**

459 Determination of divergent regions by MVista, using *Lo* as a reference, confirmed that the
460 three *Ludwigia* sp. plastomes are well preserved if the SSC segment is oriented in the same way
461 (Add. Figure 7). Sliding window analysis (Figure 6) indicated variations in definite coding
462 regions, notably *clp*P, *acc*D, *ndh5*, *ycf1* with high Pi values, and to a lesser extent, *rps16*, *mat*K,
463 *ndh*K, *pet*A, *ccs*A and four tRNAs (*trn*H, *trn*D, *trn*T and *trn*N). These polymorphic *loci* could
464 be suitable for inferring genetic diversities in *Ludwigia* sp.

465 A comparative analysis of the sizes of protein coding genes sizes also shows that the *rps11*
466 gene initially annotated in *Lo* is shorter than those which have been newly annotated in *Lgh* and
467 *Lpm* (345 bp instead of 417 bp). Comparative analysis by BLAST shows that it is the long form
468 which is annotated in other Myrtales, and the observation of the locus in *Lo* shows a frameshift
469 mutation (deletion of a nucleotide in position 311). Functional analysis would be necessary to
470 check whether the *rps11* frameshift mutation produces shorter proteins that have lost their
471 function. And only obtaining the complete genome will verify whether copies of some of these
472 genes have been transferred to mitochondrial or nuclear genomes. Such *rps11* horizontal
473 transfers have been reported for this gene in the mitochondrial genomes of various plant
474 families[74]. This also applies to *ycf1*, found as a pseudogene in *Lo* (as specified previously),
475 although it is not known if this reflects a gene transfer or a complete loss of function [75][76].
476 Moreover, there is a deletion of nine nucleotides in the 3' region of the *rpl32* gene in *Lgh* and
477 *Lpm*, leading to a premature end of the translation and the deletion of the last 4 amino acids
478 [QRLD], which are replaced by a K. However, if we look carefully at the preserved region as
479 defined by the RPL32 domain (CHL00152, member of the superfamily CL09115), we see that
480 the later amino acids are not important for *rpl32* function since they are not found in the
481 orthologs.

482 Our results show that the Ka/Ks ratio is less than 1 for most genes (Figure 7). This indicates
483 adaptive pressures to maintain the protein sequence except for *mat*K (1.17 between *Lgh* and
484 *Lpm*), *acc*D (2.48 between *Lgh* and *Lo* and 2.16 between *Lpm* and *Lo*), *ycf2* (4.3 between both
485 *Lgh-Lp* and *Lo*) and *ccs*A (1.4 between both *Lgh-Lpm* and *Lo*), showing a positive selection for
486 these genes, and a possible key role in the processes of the species' ecological adaptations. As
487 we have already described the variability in the *acc*D sequence, we will focus on *ycf2*, *mat*K,
488 and *ccs*A variations.
489 Concerning *ccs*A, the variations observed, although significant, concern only five amino
490 acids, and modifications do not seem to affect the C-type cytochrome synthase gene function.
491 Concerning *ycf2*, our analysis shows that this gene is highly polymorphic with 256 SNPs
492 that provoke 10 deletions, 7 insertions, 21 conservative and 49 non-conservative substitutions
493 in *Lo* (Add. Figure 8), compared to *Lgh* and *Lpm* (100 % identical). This gene has been shown
494 as "variant" in other plant species such as *Helianthus tuberosus* [77].
495 The *mat*K gene has been used as a universal barcoding locus to enable species discrimination
496 of terrestrial plants [78], and is often, together with the *rbc*L gene, the only known genetic
497 resource for many plants. Thus, we propose a phylogenetic tree from *Ludwigia mat*K sequences
498 (Figure 8). It should however be noted that this tree contains only 149 amino acids common to
499 all the sequences (out of the 499 in the complete protein). As only three complete *Ludwigia*
500 plastomes are available at the time of our study, we cannot specify whether these barcodes are
501 faithful to the phylogenomic history of *Ludwigia* in the same way as the complete plastome. In
502 any case, for this tree, we can see that *Lo* stands apart from the
503 other *Ludwigia* sp., *Lpm* and *Lgh*, and that the *L. grandiflora* subsp. *hexapetala* belongs to the
504 same branch as the species *L. ovalis* (aquatic taxon used in aquariums [79]), *L. stolonifera*
505 (native to the Nile, found in a variety of habitats, from freshwater wetlands to brackish and
506 marine waters) [80] and *L. adscendens* (common weed of rice fields in Asia) [81]. *Lpm* is in a
507 sister branch, close to the *L. grandiflora* subsp. *hexapetala*, forming a phylogenetic group
508 corresponding to subsect Jussiaea (in green, Figure 8).
509
510 **Discussion**
511 In the present study, we first sequenced and *de novo* assembled the chloroplast (cp) genomes
512 of *Ludwigia peploides* (*Lpm*) and *Ludwigia grandiflora* (*Lgh*), two species belonging to the
513 Onagraceae family. We employed a hybrid strategy and demonstrated the presence of two cp
514 haplotypes in *Lgh* and one haplotype in *Lpm*, although the presence of both haplotypes in *Lpm*
515 is likely. Furthermore, we compared these genomes with those of other species in the

516 Onagraceae family to expand our knowledge of genome organization and molecular evolution
517 in these species.
518 Our findings demonstrate that the utilization of solely short reads has failed to produce complete
519 *Ludwigia* plastomes, likely due to challenges posed by long repeats and rearrangements. On the
520 other hand, relying solely on long reads resulted in a lower quality sequence due to insufficient
521 coverage and sequencing errors. After conducting our research, we discovered that, for *Lgh*
522 plastomes, hybrid assembly, which incorporates both long and short read sequences, resulted
523 in the most superior complete assemblies. This innovative approach capitalizes on the
524 advantages of both sequencing technologies, harnessing the accuracy of short read sequences
525 and the length of long read sequences. In the case of our study on *Lgh* plastome reconstruction,
526 hybrid assembly was the most complete and effective, similarly to studies on other chloroplasts,
527 such as those in *Eucalyptus* [82], *Falcataria* [83], *Carex* [84] or *Cypripedium* [85].
528 In our study, we were able to identify the presence of two haplotypes in *Lgh*, which is a first
529 for *Ludwigia* (and more broadly within Onagraceae), as the plastome of *L. octovalvis* was only
530 delivered in one haplotype [86].
531 Due to the unavailability of sequence data for *Ludwigia octovalvis* and the fact that we only
532 have long reads for *Ludwigia peploides*, none of which large enough to cover the SSC/IR
533 junctions, we are unable to conclusively identify the presence of these two forms in the
534 *Ludwigia* genus. However, we believe that they are likely to be present. Unfortunately, the
535 current representation of plastomes in GenBank primarily consists of short-read data, which
536 may result in an underrepresentation of this polymorphism. It is unfortunate that structural
537 heteroplasmy, which is expected to be widespread in angiosperms, has been overlooked.
538 Existence of two plastome haplotypes has been identified in the related order of Myrtales
539 (Eucalyptus sp.), in 58 species of Angiosperms, [87], Asparagales (*Ophrys apifera* orchid [88]),
540 Brassicales (*Carica papaya*, *Vasconcellea pubescens* [89]), Solanales (*Solanum tuberosum*
541 [90]), Laurales (Avocado *Persea americana* [91]) and Rhamnaceae (*Rhamnus crenata* [92]).
542 However, the majority of reference plastomes in the current GenBank database (Release 260:
543 April 15, 2024) are described as a single haplotype, indicating an underrepresentation of
544 structural heteroplasmy in angiosperm chloroplasts. This underscores the importance of
545 sequencing techniques, as the database is predominantly composed of short-read data (98%),
546 which are less effective than long reads or hybrid assemblies at detecting flip-flop phenomena
547 in the LSC region.
548 The chloroplast genome sizes for the three genera of Onagraceae subfam. Onagroideae varied
549 as follows: *Circaea* sp. ranged from 155,817 bp to 156,024 bp, *Chamaenerion* sp. ranged from

550 159,496 bp to 160,416 bp, and Epilobium sp. ranged from 160,748 bp to 161,144 bp [93]. Our
551 study revealed that the size of the complete chloroplast of *Ludwigia* (Onagraceae subfamily
552 Ludwigioideae) ranged from 159,369 bp to 159,584 bp, which is remarkably similar to other
553 Onagraceae plants (average length of 162,030 bp). Furthermore, *Ludwigia* plastome sizes are
554 consistent with the range observed in Myrtales (between 152,214 to 171,315 bp [94]). In the
555 same way, similar overall GC content was found in *Ludwigia sp*. (from 37.3 to 37.4%), *Circaea*
556 *sp.* (37.7 to 37.8%), *Chamaenerion sp.* and *Epilobium sp.* (38.1 to 38.2%,[93]) and more
557 generally for the order Myrtales (36.9–38.9%, with the average GC content being 37%,[94]).
558 Higher GC content of the IR regions (43.5%) found in *Ludwigia* sp. has already been shown in
559 the Myrtales order (39.7–43.5%) and in other families/orders such as Amaranthaceae (order
560 Caryophyllales *[95]*) or Lamiaceae (order Lamiales [96]), and is mainly due to the presence of
561 the four GC rich rRNA genes.
562 The complete chloroplast genomes of the three *Ludwigia* species encoded an identical set of
563 134 genes including 85 protein-coding genes, 37 tRNA genes and eight ribosomal RNAs,
564 consistent with gene content found in the Myrtales order, with a gene number varying from 123
565 to 133 genes with 77–81 protein-coding genes, 29–31 tRNA gene and four rRNA genes [94].
566 Chloroplast genes have been selected during evolution due to their functional importance[97].
567 In our current study, we made the noteworthy discovery that *mat*K, *acc*D, *ycf2*, and *ccs*A genes
568 were subjected to positive selection pressure. These genes have frequently been reported in
569 literature as being associated with positive selection, and are known to play crucial roles in
570 plant development conditions. *Lgh* and *Lpm* are known to thrive in aquatic environments, where
571 they grow alongside rooted emergent aquatic plants, with their leaves and stems partially
572 submerged during growth, as reported by Wagner et al. in 2007 [1]. Both species possess the
573 unique ability of vegetative reproduction, enabling them to establish themselves rapidly in
574 diverse habitats, including terrestrial habitats, as noted by Haury et al [98]. Additionally, *Lo* is
575 a wetland plant that typically grows in gullies and at the edges of ponds, as documented by
576 Wagner *et al.* in 2007 [1]. Given their ability to adapt to different habitats, these species may
577 have evolved specialized mechanisms to cope with various abiotic stresses, such as reduced
578 carbon and oxygen availability or limited access to light in submerged or emergent conditions.
579 Concerning *matK*, Barthet et al [99] demonstrated the relationship between light and
580 developmental stages, and MatK maturase activity, suggesting important functions in plant
581 physiology. This gene has recently been largely reported to be under positive selection in an
582 aquatic plant (*Anubias* sp.,[100]), and more generally in terrestrial plants (*Pinus sp* [101]or
583 *Chrysosplenium sp.* [102]). The *acc*D gene has been described as an essential gene required for

leaf development [103] and longevity in tobacco (*Nicotiana tabacum*)[104]. Under drought stress, plant resistance can be increased by inhibiting *acc*D [105], and conversely, enhanced in response to flooding stress by upregulating *acc*D accumulation [106]. Hence, we can hypothesize that the positive selection observed on the *acc*D gene can be explained by the submerged and emerged constraints undergone by *Ludwigia* species. The *ycf2* gene seems to be subject to adaptive evolution in *Ludwigia species*. Its function, although still vague, would be to contribute to a protein complex generating ATP for the TIC machinery (proteins importing into the chloroplasts [107][108]), as well as plant cell survival [109][110]. The *ccs*A gene positive selection is found in some aquatic plants such as *Anubia sp.*[100], marine flowering plants as *Zostera* species [111], and some species of Lythraceae [105]. The *ccs*A gene is required for cytochrome c biogenesis [112] and this hemoprotein plays a key role in aerobic and anaerobic respiration, as well as photosynthesis [113]. Furthermore, we showed that *Lgh* colonization is supported by metabolic adjustments mobilizing glycolysis and fermentation pathways in terrestrial habitats, and the aminoacyl-tRNA biosynthesis pathway, which are key components of protein synthesis in aquatic habitats [114]. It can be assumed that the ability of *Ludwigia* to invade aquatic and wet environments, where the amount of oxygen and light can be variable, leads to a high selective pressure on genes involved in respiration and photosynthesis.

Molecular markers are often used to establish population genetic relationships through phylogenetic studies. Five chloroplasts (*rps16*, *rpl16*, *trn*L-*trn*F, *trn*L-CD, *trnG*) and two nuclear markers (ITS, *waxy*) were used in previous phylogeny studies of *Ludwigia sp*.[115]. However, no SSR markers had previously been made available for the *Ludwigia* genus, or more broadly, the Onagraceae. In this study, we identified 45 to 65 SSR markers depending on the *Ludwigia* species. Most of them were AT mononucleotides, as already recorded for other angiosperms [116][117]. In addition, we identified various genes with highly mutated regions that can also be used as SNP markers. Chloroplast SSRs (cpSSRs) represent potentially useful markers showing high levels of intraspecific variability due to the non-recombinant and uniparental inheritance of the plastomes [118][119]. Chloroplast SSR characteristics for *Ludwigia sp*. (location, type of SSR) were similar to those described in most plants. While the usual molecular markers used for phylogenetic analysis are nuclear DNA markers, cpSSRs have also been used to explore cytoplasmic diversity in many studies [120][121][122]. To conclude, the 13 highly variable loci and cpSSRs identified in this study are potential markers for population genetics or phylogenetic studies of *Ludwigia* species, and more generally, Onagraceae.

618 Concerning the MatK-based phylogenetic tree, its topology is generally congruent with the
619 first molecular classification of Liu *et al.* [115] as all *Ludwigia* from sect *Jussiaea* (clade B1)
620 and sect. *Ludwigia* (clade A1) and sect. *Isnardia* (clade A2) branched together. In this MatK-
621 based tree, *Ludwigia prostrata*, a species absent from previously published phylogenetic
622 studies, positions itself alone at the root of the *Ludwigia* tree. This species, sole member of
623 section *Nematopyxis*, is related as having no close relatives [123], finding supported by our
624 work. We also observed that *Ludwigia ovalis* branches within sect. *Jussiaea*, as its 258 amino
625 acids partial MatK sequence (ca. half of the complete sequence) is identical to the MatK
626 proteins of *L. grandiflora*, *L. stolonifera* and *L. adscendens*. Its phylogenetic placement remains
627 unresolved: classified alone by Raven (1963) [5] and Wagner (2017) [22] in sect. *Miquelia*,
628 later positioned by Liu et al. (2017)[4] within the *Isnardia-Microcarpium* section (using nuclear
629 DNA) or as sister to it (using plastid DNA). For this reason, conducting a whole plastome
630 analysis would be valuable to provide insights into *L. ovalis* phylogenetic positioning. Another
631 species positioned on the margins of sect. *Isnardia* (clade A2) is *Ludwigia suffruticosa*
632 (previously classified in sect. *Microcarpium*), which branches within sect. *Ludwigia* (clade A1).
633 This positioning raises questions about the current grouping of sections *Isnardia*, *Michelia*, and
634 *Microcarpium* into a single section *Isnardia* as proposed by Liu et al. (2023) [124] and
635 highlights that plastid protein coding markers can provide differing phylogenetic insights.
636 Finally, the last species positioned differently of this clade (clade B4) is *Ludwigia decurrens*
637 (sect. *Pterocaulon*) which clusters with *L. leptocarpa* (clade B3) and *L. bonariensis* (clade B4a).
638 However, it is important to note that in their study, Liu et al. (2017) indicate that clade B4 is
639 moderately supported and that the two members of sect. *Pterocaulon*, *L. decurrens* and *L.*
640 *nervosa*, diverge in all trees [4]. In summary, acquiring complete plastomes for *Ludwigia* sp.
641 could significantly enhance our understanding of the phylogeny of this complex genus.
642 Furthermore, comparing nuclear and plastid phylogenies would help determine if they reflect
643 the same evolutionary history and whether plastid phylogeny alone can accurately reconstruct
644 the phylogeny of *Ludwigia* genus.
645
646 **Conclusion**
647 In this study, we conducted the first-time sequencing and assembly of the complete plastomes
648 of *Lpm* and *Lgh*, which are the only available genomic resources for functional analysis in both
649 species. We were able to identify the existence of two haplotypes in *Lgh*, but further
650 investigations will be necessary to confirm their presence in *Lo* and *Lpm*, and more broadly,
651 within the *Ludwigia* genus. Comparison of all 10 Onagraceae plastomes revealed a high degree

652 of conservation in genome size, gene number, structure, and IR boundaries. However, to further
653 elucidate the phylogenetic analysis and evolution in *Ludwigia* and Onagraceae, additional
654 chloroplast genomes will be necessary, as highlighted in recent studies of Iris and Aristidoideae
655 species [125].
656
657 **Declarations**
658 - Availability of data and materials

659 The datasets generated and/or analysed during the current study were available in GenBank (for
660 *Lgh* haplotype 1, (LGH1) OR166254 and *Lgh* haplotype 2, (LGH2) OR166255; for *Lpm*
661 haplotype, (LPM) OR166256). Chloroplastic short and long reads are available upon request.
662 - Conflict of interest disclosure

663 The authors declare that they comply with the PCI rule of having no financial conflicts of
664 interest in relation to the content of the article

665 - Funding

666 The post-doctoral research grant of Anne-Laure Le Gac was supported by the Conseil regional
667 Bretagne (SAD18001).

668 - Acknowledgements

669 We are grateful to Luis Portillo-Lemus for developing the high molecular weight genomic DNA
670 extraction protocol. All sequencing experiments were performed at the PGTB
671 (doi:10.15454/1.5572396583599417E12).
672
673
674

**Figure 1:** Two structural haplotypes of *L. grandiflora* plastomes representing the flip-flop organization of SSC segment.

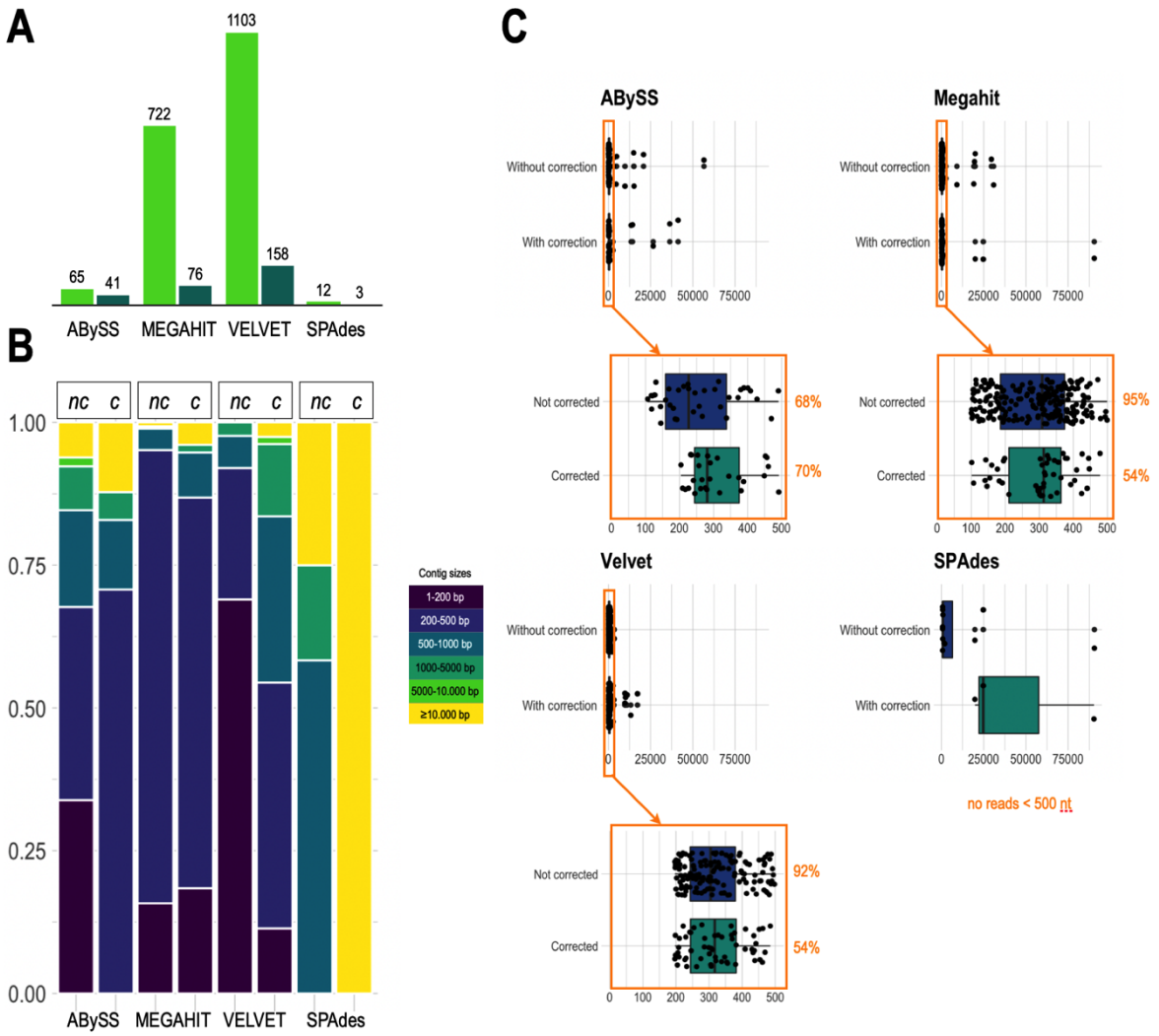


**Figure 2:** Comparative results of *L. grandiflora* short read (SR) assemblies. **A:** Total number of contigs obtained with the uncorrected (dark green) and corrected (light green) chloroplast SRs for the 4 assemblers (ABySS, MEGAHIT, Velvet and SPAdes). **B:** Comparison of the size of contigs assembled by the 4 tools using corrected or uncorrected SRs. **C:** Boxplot showing the distribution of these contigs by size and the improvement brought by the prior correction of the SRs with the long reads for each tool.

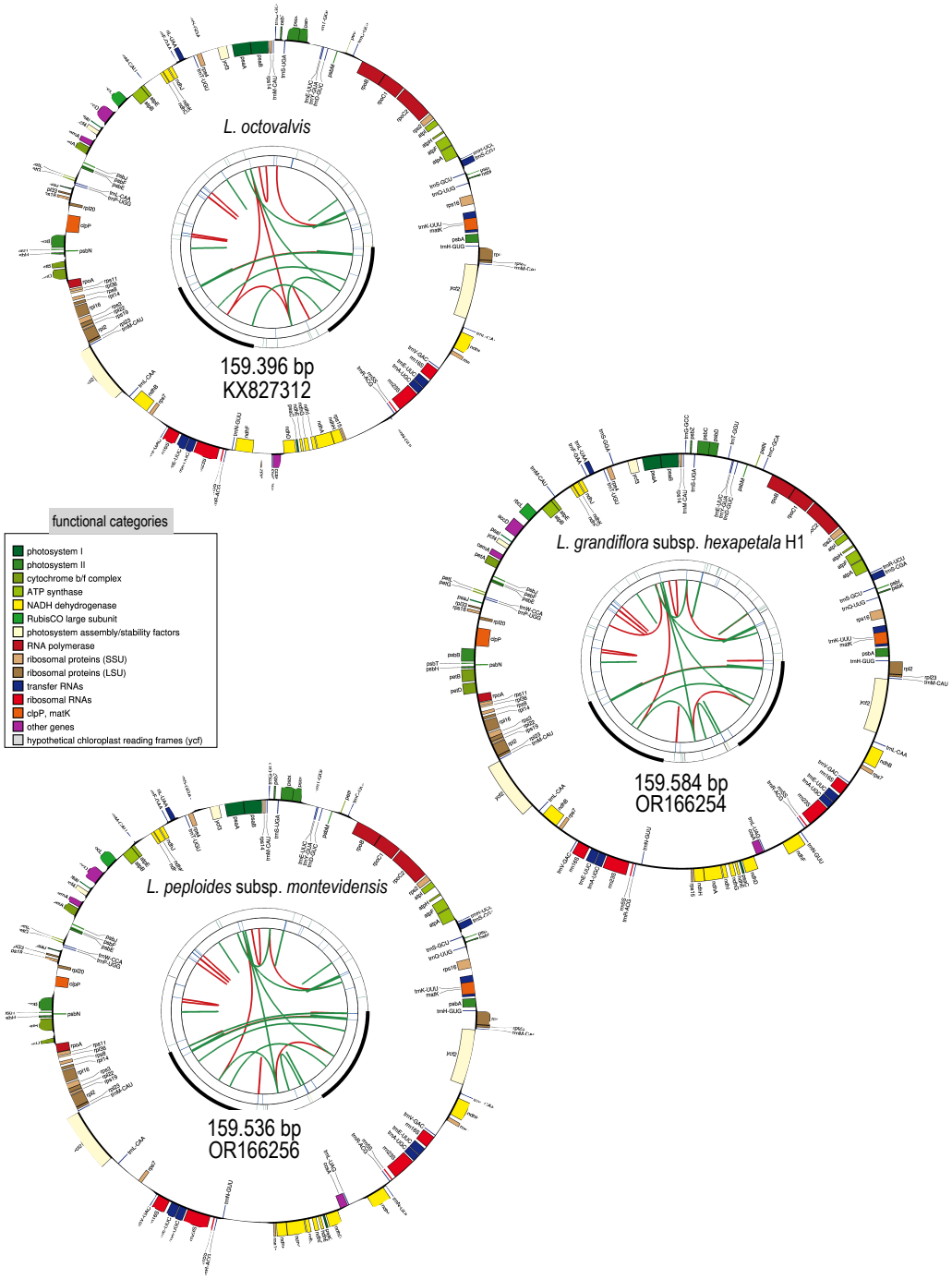


**Figure 3:** Circular representation of annotations plastomes in *Ludwigia octovalis*, *Ludwigia grandiflora* and *Ludwigia peploides* using ogdraw. Each card contains four circles. From the center outwards, the first circle shows forward and reverse repeats (red and green arcs, respectively). The next circle shows tandem repeats as bars. The third circle shows the microsatellite sequences. Finally, the fourth and fifth circles show the genes colored according to their functional categories (see colored legend). Only the haplotype 1 of *L. grandiflora* is represented as haplotype 2 only diverge by the orientation of the SSC segment. Accession numbers are indicated for each plastome.

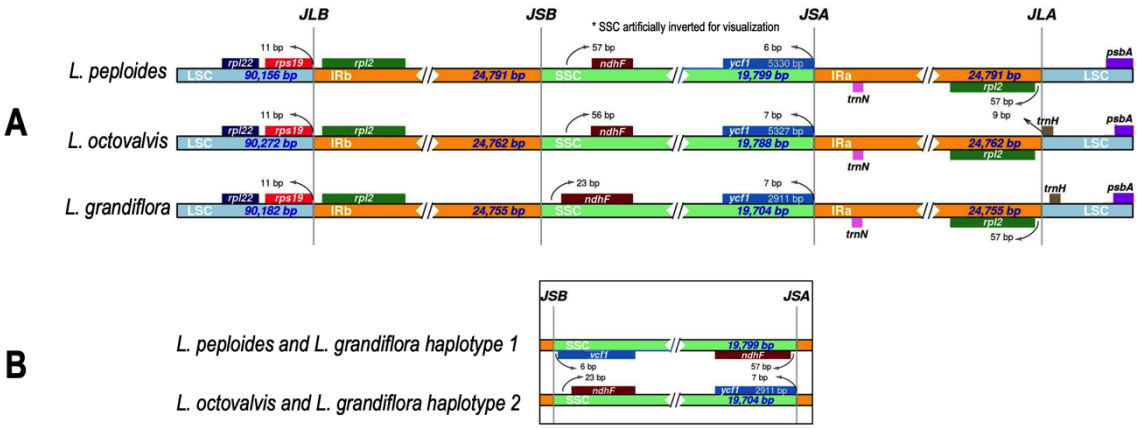


**Figure 4:** Comparison of the borders of LSC, SSC, and IR regions in Onograceae plastomes. **A:** Comparison of the junction between large single-copy (LSC, light blue), inverted repeat (IR, orange) and short single-copy (SSC, light green) regions among the chloroplast genomes of *L. octovalvis*, *L. peploides* and *L. grandiflora* (both haplotypes). Genes are denoted by colored boxes and the gaps between genes and boundaries are indicated by base lengths (bp). JLB: junction line between LSC and IRb; JSB: junction line between IRb and SSC; JSA: junction line between SSC and IRa; JLA: junction line between IRa and LSC. **B:** Comparison of SSC boundaries in haplotype 1 (*L. peploides* and *L. grandiflora* haplotype 1) and haplotype 2 (*L. octovalvis* and *L. grandiflora* haplotype 2) plastomes.

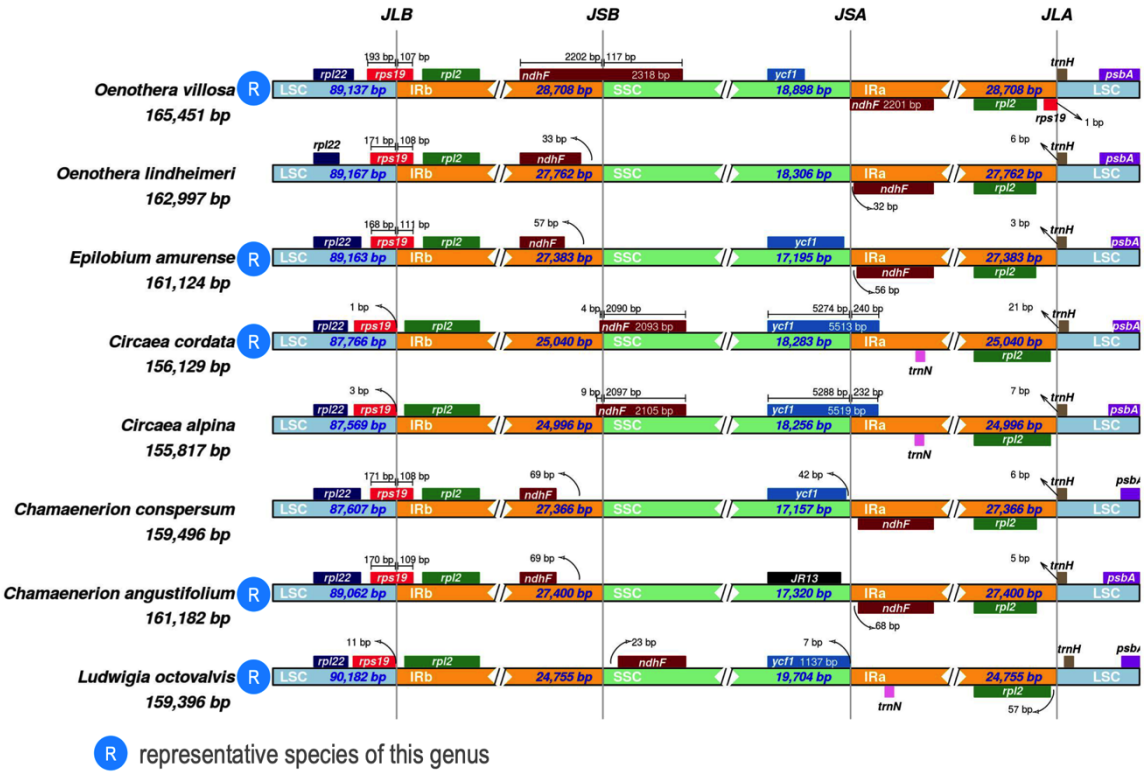


**Figure 5:** Comparison of LSC, SSC and IR regions boundaries in Onograceae chloroplast genomes. Representative sequences from each genus have been chosen (noted R on the diagram) except for *Oenothera lindheimeri* (only 89.35 % identity with others *Oenothera*), *Circaea alpina* (99.5 % identity but all others *Circaea* are 99.9% identical) and *Chamaenerion conspersum* (99% but all others *Chamaenerion* are ca. 99.7 identical). As shown in Figure 7, the 3 *Ludwigia* plastomas had the same structure, *L. octovalvis* was chosen as a representative of this genus. JLB: junction of LSC/IRb; JSB: junction of IRb/SSC; JSA: junction of SSC/IRa; JLA: junction of IRa/LSC. Accession numbers : *Chamaenerion sp. conspersum* (MZ353638), *Chamaenerion sp. angustifolium* (NC_052848)**,** *Circaea sp. cordata* (NC_060876), *Circaea sp. alpina* (NC_061010), *Epilobium amurense* (NC_061015), *Oenothera villosa* subsp. *strigosa* (NC_061365) and *Oenothera lindheimeri* (MW538951).

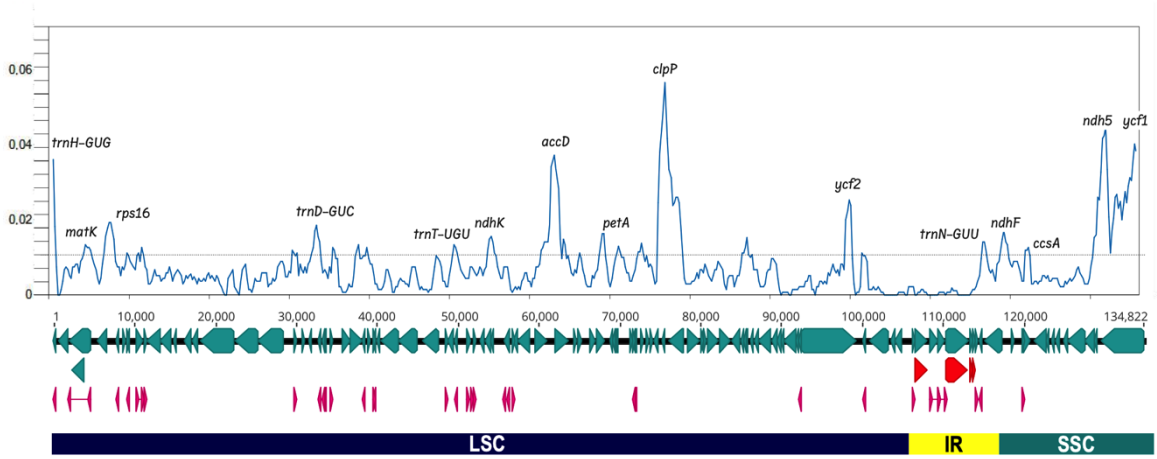


**Figure 6:** Illustration of nucleotide diversity of the three *Ludwigia* chloroplast genome sequences. The graph was generated using DnaSP software version 6.0 (windows length: 800 bp, step size: 200 bp) [64][63]. The x-axis corresponds to the base sequence of the alignment, and the y-axis represents the nucleotide diversity (π value). LSC, SSC and IR segments were indicated under the line representing the genes coding the proteins (in light blue) the tRNAs (in pink) and the rRNAs (in red). The genes marking diversity hotspots are noted at the top of the peaks.

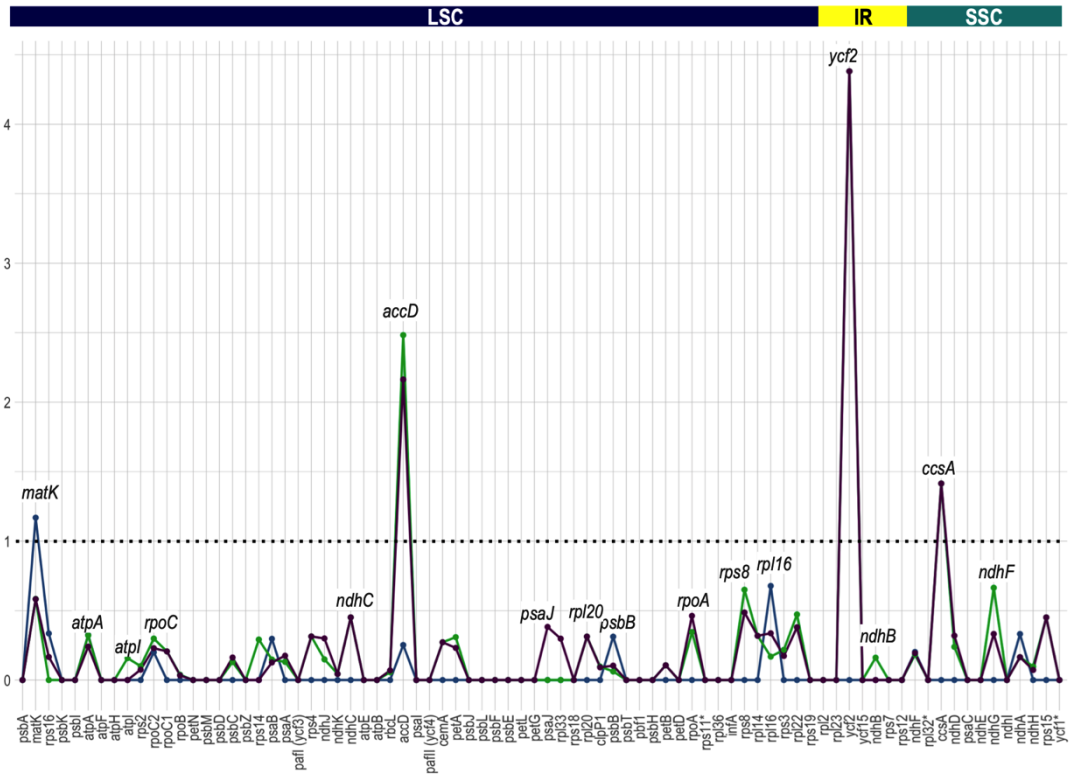


**Figure 7:** The Ka/Ks ratios of the 80 protein-coding genes of *Ludwigia* plastomes. The blue curve represents *L. grandiflora* versus *L. peploides*, purple curve denotes *L. grandiflora* versus *L. octovalvis* and green curve *L. peploides* versus *L. octovalvis*. Four genes (*matK*, *accD*, *ycf2* and *ccsA*) have Ka/Ks ratios greater than 1.0, whereas the Ka/Ks ratios of the other genes were less than 1.0.

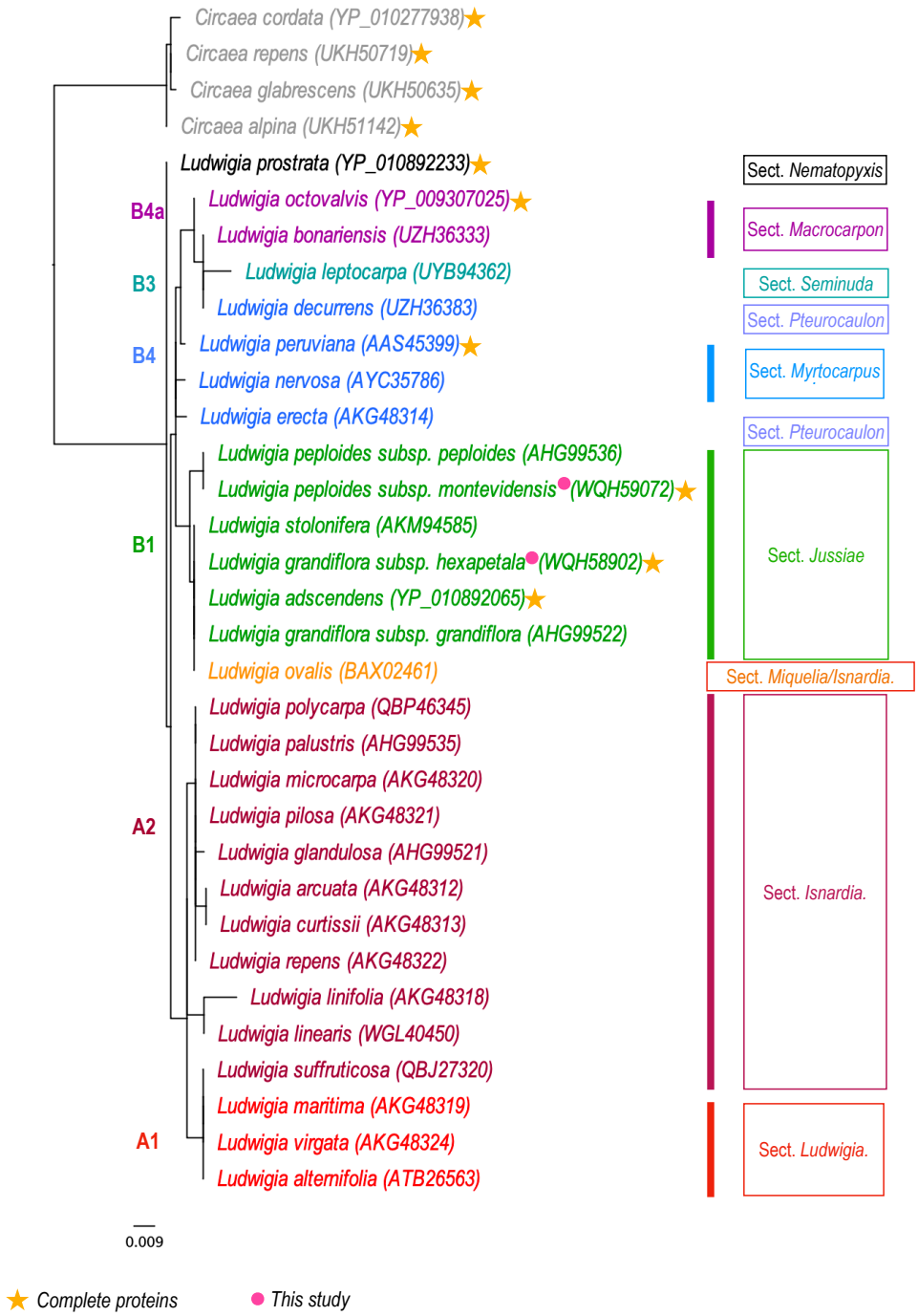


**Figure 8:** Phylogenetic tree based on *Ludwigia* MatK protein sequences. Only six *Ludwigia* sequences are complete (yellow star), the others correspond to amino acids ranging from 128 to 289 aa, with an average of 244 aa. Clades are named and colored regarding the *Ludwigia* phylogeny proposed by Liu et al. (2017) [4]. The sections are based on the works of Raven (1963) [5], Wagner et al (2017) [22] and Liu et al. (2023) [124]. The scale bar indicates the branch length.

749 Table 1. The general characteristics of the 3 *Ludwigia* plastomes

750

| | *L.octovalvis** | *L.grandiflora* | *L.peploides* |
|---|---|---|---|
| **Size (bp)** | | | |
| | 159,396 | 159,584 | 159,537 |
| LSC | 90,183 | 90,272 | 90,156 |
| SSC | 19,703 | 19,788 | 19,799 |
| IR | 24,755 | 24,762 | 24,791 |
| **GC%** | | | |
| | 37.4 | 37.3 | 37.3 |
| LSC | 35.2 | 35.1 | 35.1 |
| SSC | 32 | 31.7 | 31.7 |
| IR | 43.5 | 43.5 | 43.4 |

* KX827312 (ref)

751

752

753 Table 2. Genes present in the plastomes of *Ludwigia*

754

| FUNCTION | NAME |
|---|---|
| Photosynthesis | |
| **Rubisco** | *rbc*L |
| **Photosystem I (PSI)** | *psa*A, *psa*B, *psa*C, *psa*I, *psa*J |
| **PSI assembly factors** | *ycf*3# (*paf*I), *ycf*4 (*paf*II) |
| **Photosystem II** | *psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *pbf*1 (*psb*N) *psb*T, *psb*Z |
| **ATP synthase** | *atp*A, *atp*B, *atp*E, *atp*F#, *atp*H, *atp*I |
| **Cytochrome *b6f*** | *pet*A, *pet*B#, *pet*D#, *pet*G, *pet*L, *pet*N |
| **Cytochrome biogenesis** | *ccs*A |
| **NADPH dehydrogenase** | *ndh*A#, *ndh*B**#, *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J |
| Transcription and translation | |
| **Transcription** | *rpo*A, *rpo*B, *rpo*C1#, *rpo*C2 |
| **Small ribosomal proteins** | *rps*2, *rps*3, *rps*4, *rps*7**, *rps*8, *rps*11, *rps*12**#, *rps*14, *rps*15, *rps*16#, *rps*18, *rps*19 |
| **Large ribosomal proteins** | *rpl*2**#, *rpl*14, *rpl*16#, *rpl*20, *rpl*22, *rpl*23**, *rpl*32, *rpl*33, *rpl*36 |
| **Translation initiation** | *inf*A |
| **Ribosomal RNA** | *rrn*5**, *rrn*4,5**, *rrn*16**, *rrn*23** |
| **Transfer RNA** | *trn*A-UGC**#,*trn*C-GCA,*trn*D-GUC,*trn*E-UUC,*trnF*-GAA,*trn*fM-CAU,*trn*G-GCC,*trn*G-UCC#,*trn*H-GUG,,*trn*I-CAU**,*trn*I-GAU**#,*trn*K-UUU#,*trn*L-CAA**,*trn*L-UAA#,*trn*L-UAG,*trn*M-CAU,*trn*N-GUU**,*trn*P-UGG,*trn*Q-UUG,*trn*R-ACG**,*trn*R-UCU,*trn*S-GCU,*trn*S-GGA,*trn*S-UGA,*trn*T-GGU,*trn*T-UGU,*trn*V-GAC**,*trn*V-UAC#,*trn*W-CCA,*trn*Y-GUA |
| Other functions | |
| **Group II intron splicing** | *mat*K |
| **Inorganic carbon uptake** | *cem*A |
| **Protease** | *clp*P1# |
| **Fatty acid synthesis/Heat tolerance** | *acc*D |
| **TIC machinery (protein import)** | *ycf*1 (Tic214), *ycf*2** |
| ***Unknown function pseudogene*** | *ycf*15** |
| | |
| | ** duplicated in IR region, # spliced genes |

755

756

## Table 3A : Tandem repeats

| Sequence | *Lo* | *Lgh* | *Lpm* | Length | Region | Locus | Comments |
|---|---|---|---|---|---|---|---|
| TTGTAGTCAGGGGTGTAGTACTAT | | | | 24 | IRs | *ycf*2 | |
| TAGAAGAGAGTGCAG | | X | X | 15 | IRs | *ycf*2 | 15 nt deletion in *Lgh* and *Lpm* |
| ATGAAATATCGTATAATGAAGTACCACACGAGTGGATAT | X | X | | 39 | IRs | *rpl* 2 intron | 39 nt deletion in *Lgh* and *Lo* |
| AAAAATAGGATAGGAT | | X | X | 16 | LSC | *ycf* 1-*trn* H-GUG | 56 nt deletion in *Lgh* and *Lpm* |
| TAAATTAATATCTATATA | | X | X | 18 | LSC | *psb* Z-*trn* G-GCC | 18 nt deletion in Lgh and Lpm |
| TTTTCTATCTATCTTATATCAA | | X | X | 22 | LSC | *trn* K-UUU-*rps* 16 | 22 nt deletion in *Lgh* and *Lpm* |
| AGATCCATAACATCATCAAA | | X | X | 20 | LSC | *rps* 16 intron | 22 nt deletion in *Lgh* and *Lpm* |
| TATTAGTTATTAATATTATTAGA | | X | X | 23 | LSC | *trn* P-UGG-*psa* J | 23 nt deletion in *Lgh* and *Lpm* |
| AATAATATATAATAACTTAAATA | | X | X | 23 | LSC | *rpl* 33-*rps* 18 | 33 et 44 nt nt deletion in in *Lgh* et *Lpm*, respectively |
| TTTTTATTTAACATGCTATCAAATCAACAATGCCATACCGTAGGGCATCTGTT | | X | X | 53 | LSC | *rpl* 20-*clp* P1 | 107 nt deletion in *Lgh* and *Lpm* |
| ATATATTTCGATTCAATTC | X | | X | 19 | LSC | *trn* H-GUG-*psb* A | 3 copies in a 57 nt deletion in *Lo* and *Lpm* |
| ATAGAAATATCAGTATTTGAGTG | X | | X | 23 | LSC | *atp* H-*atp* I | 23 nt deletion in *Lo* and *Lpm* |
| TTAATTTTAATTGAAGAA | X | | X | 18 | LSC | *psb* J-*psb* L | 17 and 24 nt deletion in *Lo* and *Lpm*, respectively |
| TTAAAGAATATTAATATTC | imperfect TR | | | 19 | LSC | *trn* R-UCU-*atp* A | A -> C mutation in second copy in *Lo* |
| TATTATTATTATTAAT | X | X | | 16 | LSC | *atp* H-*atp* I | 16 nt deletion in *Lgh* and *Lo* |
| TCTAAGGCTGAAATAAGG | X | X | | 18 | LSC | *paf*I intron | 18 nt deletion in *Lgh* and *Lo* |
| TGTGAATCTATCTAT | | | X | 15 | LSC | *trn* S-UGA-*psb* Z | 8 nt deletion in *Lpm* |
| TTTTTTCTAGTA | | | | 12 | LSC | *paf*I intron | |
| CTAGTTATTGACATGG | | imperfect TR | imperfect TR | 16 | LSC | *psa* J-*rpl* 33 | G -> A mutation in second in *Lpm* et *Lgh* |
| ATTTTTATTAACTCT | X | | imperfect TR | 15 | SSC | *ycf*1 | T->A mutation in first copy in *Lpm*, other sequence in first copy in *Lo* |
| AATCAAATAGTTGAT | | X | X | 15 | SSC | *ycf*1 | other sequence in first copy of *Lpm* and *Lgh* |
| ATAATAATATATTTATTATTAATTAATA | X | | | 28 | SSC | *ndh* F-*rpl* 32 | 160 nt deletion in *Lo* |

## Table 3B : Direct repeats

| Sequence | *Lo* | *Lgh* | *Lpm* | Size (nt) | Spacers (nt) | Region | Locus | Comments |
|---|---|---|---|---|---|---|---|---|
| TTCAATTGGAACGGACGATTCGTCAATCATCT | | | | 32 | 37 | SSC | *ycf*1 | 2 copies. In *Lo*, one mutation (G->A) in the second copie |
| CATCGATGATGAAAGTGAAAACAGTAATGAAGAGG | X | | | 35 | 28 - 22 - 11 | LSC | *acc*D | 3 perfects copies and 1 mutated (G->A) copie in *Lgh* and *Lpm*. Region of 174 nt deleted in *Lo* |
| AGATGGTGAAGAACCTTATGAAGATGGTGAAGAACCTTATG | | X | X | 41 | 22 | LSC | *acc*D | Region of 147 nt deleted in Lgh and Lpm |
| TATCAAATCAACAATGCCATACCGTAGGGCAT | | X | X | 32 | 22 - 21 | LSC | *rps*12-*clp* P1 | 3 copies |
| TTAAGAGCCGTACAGGCACCTTTTGATGCATACGG | X | | | | 408 in *Lpm*, 406 in | LSC | *clp* P1 | 2 copies. In *Lgh*, one mutation (C->T) in the second copie |
| TTAAGAGCCGTACAGGCACTTTTTGATGCATACGG | X | | X | 35 | 811 | LSC | *clp* P1 *intron 1- intron 2* | |
| TGCAATAGCCAAATGATGATGAGCAATATCAGTCAGCCATA | | | | 41 | 2178 | LSC | *psa* B-*psa* A | |

## Table 3C : Palindromic repeats

| Common perfect palidromic repeats | Locus | Comment |
|---|---|---|
| AGACTCTCATGAGAGTCT ATTAAATAGAATATTCTATTTAAT | *trn*C-GCA - *pet*N *trn*E-UUC-*trn*T-GGU | |
| TTGGTAAATTTAGCAA | *pst*D | |
| TTCATTTCAATTTCAATTGAAATTGAAATGAA | *trn*I-CAU-*ycf*2 | 2 copies in IR |
| GAAAAAGGCCTTTTTC | *ycf*2 | 2 copies in IR |
| TCTCAAATGATTAATCATTTGAGA | *trn*L-*UAA* intron | |
| GGATTACTAGTAATCC | *trn*D-GUC-*trn*Y-GUA | |
| TTTGAATGCATTCAAA | *trn*G-UCC intron | |
| ATATATTCGAATATAT | *trn*G-UCC -*trn*R-UCU | |
| TAGTAATTAATTACTA | *trn*G-GCC-*trn*M-CAU | |
| CCAGTATGCATACTGG | *ndh*K | |

| Common palidromic repeats with covariation | | Locus |
|---|---|---|
| *in L. octovalvis* | *in L. grandiflora et L. peploides* | |
| ATAGAATCTATATTCTATTAGAATATAGATTCTAT | ATCGAATCTATATTCTATTAGAATATAGATTCGAT | *ndh*C-*trn*V-UAC |
| ATGTATATATATCGAT | ATCTATATATATAGAT | *trn*E-UUC-*trn*T-GGU |

| Common palindromic and quasi-palidromic repeats | | Locus | Comment |
|---|---|---|---|
| *in L. octovalvis* | *in L. grandiflora and L. peploides* | | |
| TTTAACGAATATTAATATT t GTTAAA TTAA c GAATATTAATATTCTTTAA | TTTAACGAATATTAATATTCGTTAAA TTAAAGAATATTAATATTCTTTAA | *trn*R-UCU-*atp*A | |
| AATTGTA c TTACAATT | AATTGTAATTACAATT | *ccs*A | |
| AGGAAGATTGATCAATCTT t CT | AGGAAGATTGATCAATCTTCCT | | |
| TTA c TAATATTACTAA | TTAGTAATATTACTAA | | |
| ATATAGAATAT c CTATAT | ATATAGAATATTCTATAT | | |
| ACATATCATGATA g GT | ACATATCATGATATGT | *rpl*22 | |
| AATTACTAATTTCTATTACTATGTTCAATTGAACATAGTAATAGAAATTAGTAATT | AATTACTAATTTCTATTACT t TGTTCAATTGAACATAGTAATAGAAATTAGTAATT | *atp*H-*atp*I | |
| TAGTTAGAATTCTAACTA | TAGTT c GAATTCTAACTA | *trn*T-UGU-*trn*L-UAA | |
| TATTTTTTCTAGAAAAAATA | TATTTTTTCTAGAA g AAATA | *ycf*2 | 2 copies in IR |
| *in L. octovalvis and L. peploides* | *in L. grandiflora* | | |
| CCCATCAATCATGATTG t TGGG | CCCATCAATCATGATTGATGGG | *psb*N-*trn*D-GUC | |
| *in L. octovalvis and L. grandiflora* | *in L. peploides* | | |
| ATGAAAAAAATCGATTTTTTTCAT | ATGATAAAAATAGATTTTT a TCAT | *trn*K-UUU-*rps*16 | |
| ATGAAAAAAATCGATTTTTTTCAT- ATGATAAAAATCGATTTTTATCAT | ATGATAAAAATA n ATTTTTATCAT | *trn*K-UUU-*rps*16 | |

| Unique palidromic repeats | | |
|---|---|---|
| *L. peploides* | | |
| TTATATATATATATATATAA | *rpl*32-*ndh*F | Full deletion in *L. octovalvis*, 6 bases deletion in *L. grandiflora* |
| *L. octovalvis* | | |
| ATTGAAATTCGAATTTCAAT | *psb*Z-*trn*G-GCC | Full deletion in *L. grandiflora and L. peploides* |
| *L. peploides and L. grandiflora* | | |
| AAAAAATGGATCCATTTTTT | *trn*L-UAG-*rpl*32 | 3 bases deleted and 3 bases mutated in *L. octovalvis* |
| AATATATTATTATAATAATATATT | *rpl*32-*ndh*F | Full deletion in *L. octovalvis* |
| TATATTTATTATTAATTAATAATAAATATA | *rpl*32-*ndh*F | Full deletion in *L. octovalvis* |
| *L. octovalvis* | | |
| ATTGAAATTCGAATTTCAAT | *psb*Z-*trn*G-GCC | Full deletion in *L. grandiflora and L. peploides* |
| *L. peploides and L. grandiflora* | | |
| AAAAAATGGATCCATTTTTT | *trn*L-UAG-*rpl*32 | 3 bases deleted and 3 bases mutated in *L. octovalvis* |
| AATATATATTATTATAATAATATATT | *rpl*32-*ndh*F | Full deletion in *L. octovalvis* |
| TATATTTATTATTAATTAATAATAAATATA | *rpl*32-*ndh*F | Full deletion in *L. octovalvis* |

*Lo* = *Ludwigia octovalvis; Lgh* = *L. grandiflora* subsp. *hexapetala; Lpm* = *L. peploides* subsp. *montevidensis*.

## REFERENCES


[1] W. L. Wagner, P. C. Hoch, and P. H. Raven, "Revised classification of the Onagraceae," *Systematic Botany Monographs*, 2007.

[2] R. A. Levin *et al.*, "Family-level relationships of Onagraceae based on chloroplast rbcL and ndhF data," *Am J Bot*, vol. 90, no. 1, 2003, doi: 10.3732/ajb.90.1.107.

[3] R. A. Levin *et al.*, "Paraphyly in Tribe Onagreae: Insights into Phylogenetic Relationships of Onagraceae Based on Nuclear and Chloroplast Sequence Data," *Syst Bot*, vol. 29, no. 1, 2004, doi: 10.1600/036364404772974293.

[4] S. H. Liu, P. C. Hoch, M. Diazgranados, P. H. Raven, and J. C. Barber, "Multi-locus phylogeny of ludwigia (Onagraceae): Insights on infra- generic relationships and the current classification of the genus," *Taxon*, vol. 66, no. 5, 2017, doi: 10.12705/665.7.

[5] Raven P.H., "The_Old_World_species_of_Ludwigia_including_Jussia," *Reinwardtia*, vol. 6, pp. 327–427, 1963.

[6] S. Dandelot, R. Verlaque, A. Dutartre, and A. Cazaubon, "Ecological, dynamic and taxonomic problems due to Ludwigia (Onagraceae) in France," in *Hydrobiologia*, 2005. doi: 10.1007/s10750-005-4455-0.

[7] A. M. Reddy *et al.*, "Biological control of invasive water primroses, Ludwigia spp., in the United States: A feasibility assessment." J. Aquat. Plant Manage. 59s: 2021

[8] A. Hussner, M. Windhaus, and U. Starfinger, "From weed biology to successful control: an example of successful management of Ludwigia grandiflora in Germany," *Weed Res*, vol. 56, no. 6, 2016, doi: 10.1111/wre.12224.

[9] B. J. Grewell, M. D. Netherland, and M. J. Skaer Thomason, "Establishing research and management priorities for invasive water primroses (Ludwigia spp.)," *Aquatic Plant Control Research Program, US Army Corps of Engineers, Engineer Research and Development Center, Environmental Laboratory Technical Report ERDC/ELTR-15-X*, no. February, 2016.

[10] E. Lambert, A. Dutartre, J. Coudreuse, and J. Haury, "Relationships between the biomass production of invasive Ludwigia species and physical properties of habitats in France," *Hydrobiologia*, vol. 656, no. 1, 2010, doi: 10.1007/s10750-010-0440-3.

[11] J. Haury, A. Druel, T. Cabral, Y. Paulet, M. Bozec, and J. Coudreuse, "Which adaptations of some invasive Ludwigia spp. (Rosidae, Onagraceae) populations occur in contrasting hydrological conditions in Western France?," *Hydrobiologia*, vol. 737, no. 1, 2014, doi: 10.1007/s10750-014-1815-7.

[12] K. Billet, J. Genitoni, M. Bozec, D. Renault, and D. Barloy, "Aquatic and terrestrial morphotypes of the aquatic invasive plant, Ludwigia grandiflora, show distinct morphological and metabolomic responses," *Ecol Evol*, vol. 8, no. 5, 2018, doi: 10.1002/ece3.3848.

[13] M. Gioria, P. E. Hulme, D. M. Richardson, and P. Pyšek, "Annual Review of Plant Biology Why Are Invasive Plants Successful?," *Annu. Rev. Plant Biol. 2023*, vol. 74, p. 2023, 2023, doi: 10.1146/annurev-arplant-070522.

[14] L. Moravcová, P. Pyšek, V. Jarošík, and J. Pergl, "Getting the right traits: Reproductive and dispersal characteristics predict the invasiveness of herbaceous plant species," *PLoS One*, vol. 10, no. 4, Apr. 2015, doi: 10.1371/journal.pone.0123634.

[15] R. A. Marks, S. Hotaling, P. B. Frandsen, and R. VanBuren, "Representation and participation across 20 years of plant genome sequencing," *Nat Plants*, vol. 7, no. 12, 2021, doi: 10.1038/s41477-021-01031-8.

[16] D. Barloy, L. Portillo-Lemus, S. Krueger-Hadfield, V. Huteau, and O. Coriton, "Genomic relationships among diploid and polyploid species of the genus Ludwigia L. section Jussiaea using a combination of molecular cytogenetic, morphological, and

crossing investigations," *Peer Community Journal*, vol. 4, 2024, doi: 10.24072/pcjournal.364.

[17] S. H. Liu, C. Edwards, P. C. Hoch, P. H. Raven, and J. C. Barber, "Complete plastome sequence of ludwigia octovalvis (Onagraceae), a globally distributed wetland plant," *Genome Announc*, vol. 4, no. 6, 2016, doi: 10.1128/genomeA.01274-16.

[18] E. Zardini and P. H. Raven, "A New Section of Ludwigia (Onagraceae) with a Key to the Sections of the Genus," *Syst Bot*, vol. 17, no. 3, 1992, doi: 10.2307/2419486.

[19] P. C. Hoch, W. L. Wagner, and P. H. Raven, "The correct name for a section of Ludwigia L. (Onagraceae)," *PhytoKeys*, vol. 50, no. 1, 2015, doi: 10.3897/phytokeys.50.4887.

[20] Y. Hu, Q. Zhang, G. Rao, and Sodmergen, "Occurrence of plastids in the sperm cells of caprifoliaceae: Biparental plastid inheritance in angiosperms is unilaterally derived from maternal inheritance," *Plant Cell Physiol*, vol. 49, no. 6, 2008, doi: 10.1093/pcp/pcn069.

[21] Q. Zhang and Sodmergen, "Why does biparental plastid inheritance revive in angiosperms?," *J Plant Res*, vol. 123, no. 2, 2010, doi: 10.1007/s10265-009-0291-z.

[22] W. L. Wagner, P. C. Hoch, and P. H. Raven, *Systematic botany monographs: Revised classification of the Onagraceae*. 2007.

[23] K. Jones and R. E. Cleland, "Oenothera, Cytogenetics and Evolution," *Kew Bull*, vol. 29, no. 1, 1974, doi: 10.2307/4108389.

[24] U. K. Schmitz and K. V. Kowallik, "Plastid inheritance in Epilobium," *Curr Genet*, vol. 11, no. 1, 1986, doi: 10.1007/BF00389419.

[25] N. Sato, "Are cyanobacteria an ancestor of chloroplasts or just one of the gene donors for plants and algae?," *Genes (Basel)*, vol. 12, no. 6, 2021, doi: 10.3390/genes12060823.

[26] J. M. Gualberto, D. Mileshina, C. Wallet, A. K. Niazi, F. Weber-Lotfi, and A. Dietrich, "The plant mitochondrial genome: Dynamics and maintenance," *Biochimie*, vol. 100, no. 1. 2014. doi: 10.1016/j.biochi.2013.09.016.

[27] J. Tonti-Filippini, P. G. Nevill, K. Dixon, and I. Small, "What can we do with 1000 plastid genomes?," *Plant Journal*, vol. 90, no. 4, 2017, doi: 10.1111/tpj.13491.

[28] D. J. Oldenburg and A. J. Bendich, "The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication," *Curr Genet*, vol. 62, no. 2, 2016, doi: 10.1007/s00294-015-0548-0.

[29] A. D. Twyford and R. W. Ness, "Strategies for complete plastid genome sequencing," *Mol Ecol Resour*, vol. 17, no. 5, 2017, doi: 10.1111/1755-0998.12626.

[30] W. Wang, M. Schalamun, A. Morales-Suarez, D. Kainer, B. Schwessinger, and R. Lanfear, "Assembly of chloroplast genomes with long- and short-read data: A comparison of approaches using Eucalyptus pauciflora as a test case," *BMC Genomics*, vol. 19, no. 1, 2018, doi: 10.1186/s12864-018-5348-8.

[31] W. Wang, R. Lanfear, and B. Gaut, "Long-Reads Reveal That the Chloroplast Genome Exists in Two Distinct Versions in Most Plants," *Genome Biol Evol*, vol. 11, no. 12, 2019, doi: 10.1093/gbe/evz256.

[32] M. Ferrarini *et al.*, "An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome," *BMC Genomics*, vol. 14, no. 1, 2013, doi: 10.1186/1471-2164-14-670.

[33] M. Jain *et al.*, "Nanopore sequencing and assembly of a human genome with ultra-long reads," *Nat Biotechnol*, vol. 36, no. 4, 2018, doi: 10.1038/nbt.4060.

[34] F. J. Rang, W. P. Kloosterman, and J. de Ridder, "From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy," *Genome Biology*, vol. 19, no. 1. 2018. doi: 10.1186/s13059-018-1462-9.

[35] A. Scheunert, M. Dorfner, T. Lingl, and C. Oberprieler, "Can we use it? On the utility of de novo and reference-based assembly of Nanopore data for plant plastome sequencing," *PLoS One*, vol. 15, no. 3, 2020, doi: 10.1371/journal.pone.0226234.

[36] A. M. Bedoya and S. Madriñán, "Evolution of the aquatic habit in Ludwigia (Onagraceae): Morpho-anatomical adaptive strategies in the Neotropics," *Aquat Bot*, vol. 120, no. PB, 2015, doi: 10.1016/j.aquabot.2014.10.005.

[37] S. H. Liu *et al.*, "Disentangling Reticulate Evolution of North Temperate Haplostemonous Ludwigia (Onagraceae)," *Annals of the Missouri Botanical Garden*, vol. 105, no. 2, 2020, doi: 10.3417/2020479.

[38] M. Panova *et al.*, "DNA extraction protocols for whole-genome sequencing in marine organisms," in *Methods in Molecular Biology*, vol. 1452, 2016. doi: 10.1007/978-1-4939-3774-5_2.

[39] C. Belser *et al.*, "Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps," *Nature Plants*, vol. 4, no. 11. 2018. doi: 10.1038/s41477-018-0289-4.

[40] S. Chen, Y. Zhou, Y. Chen, and J. Gu, "Fastp: An ultra-fast all-in-one FASTQ preprocessor," in *Bioinformatics*, 2018. doi: 10.1093/bioinformatics/bty560.

[41] J. J. Jin *et al.*, "GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes," *Genome Biol*, vol. 21, no. 1, 2020, doi: 10.1186/s13059-020-02154-5.

[42] N. Dierckxsens, P. Mardulyn, and G. Smits, "NOVOPlasty: De novo assembly of organelle genomes from whole genome data," *Nucleic Acids Res*, vol. 45, no. 4, 2017, doi: 10.1093/nar/gkw955.

[43] D. R. Zerbino and E. Birney, "Velvet: Algorithms for de novo short read assembly using de Bruijn graphs," *Genome Res*, vol. 18, no. 5, pp. 821–829, May 2008, doi: 10.1101/gr.074492.107.

[44] J. T. Simpson, K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones, and I. Birol, "ABySS: A parallel assembler for short read sequence data," *Genome Res*, vol. 19, no. 6, 2009, doi: 10.1101/gr.089532.108.

[45] S. D. Jackman *et al.*, "ABySS 2.0: Resource-efficient assembly of large genomes using a Bloom filter," *Genome Res*, vol. 27, no. 5, 2017, doi: 10.1101/gr.214346.116.

[46] D. Li *et al.*, "MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices," *Methods*, vol. 102. 2016. doi: 10.1016/j.ymeth.2016.02.020.

[47] A. Bankevich *et al.*, "SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing," *Journal of Computational Biology*, vol. 19, no. 5, 2012, doi: 10.1089/cmb.2012.0021.

[48] R. Chikhi and P. Medvedev, "Informed and automated k-mer size selection for genome assembly," *Bioinformatics*, vol. 30, no. 1, 2014, doi: 10.1093/bioinformatics/btt310.

[49] S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, and A. M. Phillippy, "Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation," *Genome Res*, vol. 27, no. 5, 2017, doi: 10.1101/gr.215087.116.

[50] G. Holley *et al.*, "Ratatosk: hybrid error correction of long reads enables accurate variant calling and assembly," *Genome Biol*, vol. 22, no. 1, 2021, doi: 10.1186/s13059-020-02244-4.

[51] M. Kolmogorov, J. Yuan, Y. Lin, and P. A. Pevzner, "Assembly of long, error-prone reads using repeat graphs," *Nat Biotechnol*, vol. 37, no. 5, 2019, doi: 10.1038/s41587-019-0072-8.

[52] A. Gurevich, V. Saveliev, N. Vyahhi, and G. Tesler, "QUAST: Quality assessment tool for genome assemblies," *Bioinformatics*, vol. 29, no. 8, pp. 1072–1075, Apr. 2013, doi: 10.1093/bioinformatics/btt086.

[53] R. R. Wick, M. B. Schultz, J. Zobel, and K. E. Holt, "Bandage: Interactive visualization of de novo genome assemblies," *Bioinformatics*, vol. 31, no. 20, 2015, doi: 10.1093/bioinformatics/btv383.

[54] M. Tillich *et al.*, "GeSeq - Versatile and accurate annotation of organelle genomes," *Nucleic Acids Res*, vol. 45, no. W1, 2017, doi: 10.1093/nar/gkx391.

[55] X. Zhong, "Assembly, annotation and analysis of chloroplast genomes," 2020. [Doctoral Thesis, The University of Western Australia].

[56] S. Greiner, P. Lehwark, and R. Bock, "OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes," *Nucleic Acids Res*, vol. 47, no. W1, 2019, doi: 10.1093/nar/gkz238.

[57] P. Lehwark and S. Greiner, "GB2sequin - A file converter preparing custom GenBank files for database submission," *Genomics*, vol. 111, no. 4, 2019, doi: 10.1016/j.ygeno.2018.05.003.

[58] S. Beier, T. Thiel, T. Münch, U. Scholz, and M. Mascher, "MISA-web: A web server for microsatellite prediction," *Bioinformatics*, vol. 33, no. 16, 2017, doi: 10.1093/bioinformatics/btx198.

[59] M. Gurusaran, D. Ravella, and K. Sekar, "RepEx: Repeat extractor for biological sequences," *Genomics*, vol. 102, no. 4, pp. 403–408, Oct. 2013, doi: 10.1016/j.ygeno.2013.07.005.

[60] G. Benson, "Tandem repeats finder: A program to analyze DNA sequences," *Nucleic Acids Res*, vol. 27, no. 2, 1999, doi: 10.1093/nar/27.2.573.

[61] K. A. Frazer, L. Pachter, A. Poliakov, E. M. Rubin, and I. Dubchak, "VISTA: Computational tools for comparative genomics," *Nucleic Acids Res*, vol. 32, no. WEB SERVER ISS., 2004, doi: 10.1093/nar/gkh458.

[62] M. Brudno *et al.*, "LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA," *Genome Research*, vol. 13, no. 4. 2003. doi: 10.1101/gr.926603.

[63] J. Rozas and R. Rozas, "DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis," *Bioinformatics*, vol. 15, no. 2. 1999. doi: 10.1093/bioinformatics/15.2.174.

[64] J. Rozas *et al.*, "DnaSP 6: DNA sequence polymorphism analysis of large data sets," *Mol Biol Evol*, vol. 34, no. 12, 2017, doi: 10.1093/molbev/msx248.

[65] A. Amiryousefi, J. Hyvönen, and P. Poczai, "IRscope: an online program to visualize the junction sites of chloroplast genomes," *Bioinformatics*, vol. 34, no. 17, 2018, doi: 10.1093/bioinformatics/bty220.

[66] C. Chen *et al.*, "TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data," *Mol Plant*, vol. 13, no. 8, 2020, doi: 10.1016/j.molp.2020.06.009.

[67] K. Katoh, K. Misawa, K. I. Kuma, and T. Miyata, "MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform," *Nucleic Acids Res*, vol. 30, no. 14, 2002, doi: 10.1093/nar/gkf436.

[68] A. Stamatakis, "RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies," *Bioinformatics*, vol. 30, no. 9, pp. 1312–1313, May 2014, doi: 10.1093/bioinformatics/btu033.

[69] J. Ou and L. J. Zhu, "trackViewer: a Bioconductor package for interactive and integrative visualization of multi-omics data," *Nature Methods*, vol. 16, no. 6. Nature Publishing Group, pp. 453–454, Jun. 01, 2019. doi: 10.1038/s41592-019-0430-y.

[70] T. Konishi and Y. Sasaki, "Compartmentalization of two forms of acetyl-CoA carboxylase in plants and the origin of their tolerance toward herbicides," *Proc Natl Acad Sci U S A*, vol. 91, no. 9, pp. 3598–3601, 1994, doi: 10.1073/pnas.91.9.3598.

[71] S. Wu *et al.*, "Extensive genomic rearrangements mediated by repetitive sequences in plastomes of Medicago and its relatives," *BMC Plant Biol*, vol. 21, no. 1, p. 421, 2021, doi: 10.1186/s12870-021-03202-3.
[72] J. Li, Y. Su, and T. Wang, "The Repeat Sequences and Elevated Substitution Rates of the Chloroplast accD Gene in Cupressophytes," *Front Plant Sci*, vol. 9, p. 533, 2018, doi: 10.3389/fpls.2018.00533.
[73] C. Gurdon and P. Maliga, "Two distinct plastid genome configurations and unprecedented intraspecies length variation in the accD coding region in Medicago truncatula," *DNA Res*, vol. 21, no. 4, pp. 417–427, 2014, doi: 10.1093/dnares/dsu007.
[74] A. O. Richardson and J. D. Palmer, "Horizontal gene transfer in plants," *J Exp Bot*, vol. 58, no. 1, pp. 1–9, 2007, doi: 10.1093/jxb/erl148.
[75] J. de Vries, F. L. Sousa, B. Bolter, J. Soll, and S. B. Gould, "YCF1: A Green TIC?," *Plant Cell*, vol. 27, no. 7, pp. 1827–1833, 2015, doi: 10.1105/tpc.114.135541.
[76] E. Filip and L. Skuza, "Horizontal Gene Transfer Involving Chloroplasts," *Int J Mol Sci*, vol. 22, no. 9, 2021, doi: 10.3390/ijms22094484.
[77] Q. Zhong, S. Yang, X. Sun, L. Wang, and Y. Li, "The complete chloroplast genome of the Jerusalem artichoke (Helianthus tuberosus L.) and an adaptive evolutionary analysis of the ycf2 gene," *PeerJ*, vol. 7, p. e7596, 2019, doi: 10.7717/peerj.7596.
[78] S. Antil *et al.*, "DNA barcoding, an effective tool for species identification: a review," *Mol Biol Rep*, vol. 50, no. 1, pp. 761–775, 2023, doi: 10.1007/s11033-022-08015-7.
[79] J. Li *et al.*, "Removal effects of aquatic plants on high-concentration phosphorus in wastewater during summer," *J Environ Manage*, vol. 324, p. 116434, 2022, doi: 10.1016/j.jenvman.2022.116434.
[80] A. T. Soliman, R. S. Hamdy, and A. B. Hamed, "Ludwigia stolonifera (Guill. & Perr.) PH Raven, insight into its phenotypic plasticity, habitat diversity and associated species," *Egyptian Journal of Botany*, vol. 58, no. 3, pp. 605–626, 2018.
[81] A. Kamoshita, H. Ikeda, J. Yamagishi, B. Lor, and M. Ouk, "Residual effects of cultivation methods on weed seed banks and weeds in Cambodia," *Weed Biol Manag*, vol. 16, no. 3, pp. 93–107, 2016.
[82] W. Wang, M. Schalamun, A. Morales-Suarez, D. Kainer, B. Schwessinger, and R. Lanfear, "Assembly of chloroplast genomes with long-and short-read data: a comparison of approaches using Eucalyptus pauciflora as a test case," *BMC Genomics*, vol. 19, pp. 1–15, 2018.
[83] V. P. D. Anita, D. D. Matra, and U. J. Siregar, "Chloroplast genome draft assembly of Falcataria moluccana using hybrid sequencing technology," *BMC Res Notes*, vol. 16, no. 1, p. 31, 2023, doi: 10.1186/s13104-023-06290-6.
[84] S. Xu *et al.*, "Chloroplast genomes of four Carex species: Long repetitive sequences trigger dramatic changes in chloroplast genome structure," *Front Plant Sci*, vol. 14, p. 1100876, 2023, doi: 10.3389/fpls.2023.1100876.
[85] Y. Y. Guo, J. X. Yang, H. K. Li, and H. S. Zhao, "Chloroplast Genomes of Two Species of Cypripedium: Expanded Genome Size and Proliferation of AT-Biased Repeat Sequences," *Front Plant Sci*, vol. 12, p. 609729, 2021, doi: 10.3389/fpls.2021.609729.
[86] S.-H. Liu, C. Edwards, P. C. Hoch, P. H. Raven, and J. C. Barber, "Complete plastome sequence of Ludwigia octovalvis (Onagraceae), a globally distributed wetland plant," *Genome Announc*, vol. 4, no. 6, pp. e01274-16, 2016.
[87] W. Wang and R. Lanfear, "Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants," *Genome Biol Evol*, vol. 11, no. 12, pp. 3372–3381, 2019.
[88] R. M. Bateman, P. J. Rudall, A. R. M. Murphy, R. S. Cowan, D. S. Devey, and O. A. Perez-Escobar, "Whole plastomes are not enough: phylogenomic and morphometric

exploration at multiple demographic levels of the bee orchid clade Ophrys sect. Sphegodes," *J Exp Bot*, vol. 72, no. 2, pp. 654–681, 2021, doi: 10.1093/jxb/eraa467.

[89] Z. Lin *et al.*, "Comparative analysis of chloroplast genomes in Vasconcellea pubescens A.DC. and Carica papaya L," *Sci Rep*, vol. 10, no. 1, p. 15799, 2020, doi: 10.1038/s41598-020-72769-y.

[90] G. A. Lihodeevskiy and E. P. Shanina, "The Use of Long-Read Sequencing to Study the Phylogenetic Diversity of the Potato Varieties Plastome of the Ural Selection," *Agronomy*, vol. 12, no. 4, p. 846, 2022.

[91] O. Nath *et al.*, "A haplotype resolved chromosomal level avocado genome allows analysis of novel avocado genes," *Hortic Res*, vol. 9, p. uhac157, 2022, doi: 10.1093/hr/uhac157.

[92] K. Wanichthanarak *et al.*, "Revisiting chloroplast genomic landscape and annotation towards comparative chloroplast genomes of Rhamnaceae," *BMC Plant Biol*, vol. 23, no. 1, p. 59, 2023.

[93] Y. Luo *et al.*, "Comparative Analysis of Complete Chloroplast Genomes of 13 Species in Epilobium, Circaea, and Chamaenerion and Insights Into Phylogenetic Relationships of Onagraceae," *Front Genet*, vol. 12, p. 730495, 2021, doi: 10.3389/fgene.2021.730495.

[94] X. F. Zhang, J. B. Landis, H. X. Wang, Z. X. Zhu, and H. F. Wang, "Comparative analysis of chloroplast genome structure and molecular dating in Myrtales," *BMC Plant Biol*, vol. 21, no. 1, p. 219, 2021, doi: 10.1186/s12870-021-02985-9.

[95] J. Xu, X. Shen, B. Liao, J. Xu, and D. Hou, "Comparing and phylogenetic analysis chloroplast genome of three Achyranthes species," *Sci Rep*, vol. 10, no. 1, p. 10818, 2020, doi: 10.1038/s41598-020-67679-y.

[96] C. Lian *et al.*, "Comparative analysis of chloroplast genomes reveals phylogenetic relationships and intraspecific variation in the medicinal plant Isodon rubescens," *PLoS One*, vol. 17, no. 4, p. e0266546, 2022, doi: 10.1371/journal.pone.0266546.

[97] T. K. Mohanta, A. K. Mishra, A. Khan, A. Hashem, E. F. Abd Allah, and A. Al-Harrasi, "Gene Loss and Evolution of the Plastome," *Genes (Basel)*, vol. 11, no. 10, 2020, doi: 10.3390/genes11101133.

[98] J. Haury, A. Druel, T. Cabral, Y. Paulet, M. Bozec, and J. Coudreuse, "Which adaptations of some invasive Ludwigia spp.(Rosidae, Onagraceae) populations occur in contrasting hydrological conditions in Western France?," *Hydrobiologia*, vol. 737, pp. 45–56, 2014.

[99] M. M. Barthet and K. W. Hilu, "Expression of matK: functional and evolutionary implications," *Am J Bot*, vol. 94, no. 8, pp. 1402–1412, 2007.

[100] L. Li, C. Liu, K. Hou, and W. Liu, "Comparative Analyses of Plastomes of Four Anubias (Araceae) Taxa, Tropical Aquatic Plants Endemic to Africa," *Genes (Basel)*, vol. 13, no. 11, 2022, doi: 10.3390/genes13112043.

[101] U. Zeb *et al.*, "Comparative genome sequence and phylogenetic analysis of chloroplast for evolutionary relationship among Pinus species," *Saudi J Biol Sci*, vol. 29, no. 3, pp. 1618–1627, 2022, doi: 10.1016/j.sjbs.2021.10.070.

[102] Z. Wu *et al.*, "Analysis of six chloroplast genomes provides insight into the evolution of Chrysosplenium (Saxifragaceae)," *BMC Genomics*, vol. 21, no. 1, p. 621, 2020, doi: 10.1186/s12864-020-07045-4.

[103] V. Kode, E. A. Mudd, S. Iamtham, and A. Day, "The tobacco plastid accD gene is essential and is required for leaf development," *Plant J*, vol. 44, no. 2, pp. 237–244, 2005, doi: 10.1111/j.1365-313X.2005.02533.x.

[104] Y. Madoka, K. Tomizawa, J. Mizoi, I. Nishida, Y. Nagano, and Y. Sasaki, "Chloroplast transformation with modified accD operon increases acetyl-CoA carboxylase and causes

extension of leaf longevity and increase in seed yield in tobacco," *Plant Cell Physiol*, vol. 43, no. 12, pp. 1518–1525, 2002, doi: 10.1093/pcp/pcf172.

[105] H. Gu *et al.*, "Drought stress triggers proteomic changes involving lignin, flavonoids and fatty acids in tea plants," *Sci Rep*, vol. 10, no. 1, p. 15504, 2020, doi: 10.1038/s41598-020-72596-1.

[106] B. Bharadwaj *et al.*, "Physiological and genetic responses of lentil (Lens culinaris) under flood stress," *Plant Stress*, p. 100130, 2023.

[107] S. Kikuchi *et al.*, "A Ycf2-FtsHi Heteromeric AAA-ATPase Complex Is Required for Chloroplast Protein Import," *Plant Cell*, vol. 30, no. 11, pp. 2677–2703, 2018, doi: 10.1105/tpc.18.00357.

[108] T. B. Schreier *et al.*, "Plastidial NAD-Dependent Malate Dehydrogenase: A Moonlighting Protein Involved in Early Chloroplast Development through Its Interaction with an FtsH12-FtsHi Protease Complex," *Plant Cell*, vol. 30, no. 8, pp. 1745–1769, 2018, doi: 10.1105/tpc.18.00121.

[109] A. Drescher, S. Ruf, T. Calsa Jr., H. Carrer, and R. Bock, "The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes," *Plant J*, vol. 22, no. 2, pp. 97–104, 2000, doi: 10.1046/j.1365-313x.2000.00722.x.

[110] J. Xing *et al.*, "The plastid-encoded protein Orf2971 is required for protein translocation and chloroplast quality control," *Plant Cell*, vol. 34, no. 9, pp. 3383–3399, 2022, doi: 10.1093/plcell/koac180.

[111] J. Chen *et al.*, "Chloroplast genomic comparison provides insights into the evolution of seagrasses," *BMC Plant Biol*, vol. 23, no. 1, p. 104, 2023, doi: 10.1186/s12870-023-04119-9.

[112] Z. Xie and S. Merchant, "The plastid-encoded ccsA gene is required for heme attachment to chloroplast c-type cytochromes," *J Biol Chem*, vol. 271, no. 9, pp. 4632–4639, 1996, doi: 10.1074/jbc.271.9.4632.

[113] R. Kranz, R. Lill, B. Goldman, G. Bonnard, and S. Merchant, "Molecular mechanisms of cytochrome c biogenesis: three distinct systems," *Mol Microbiol*, vol. 29, no. 2, pp. 383–396, 1998, doi: 10.1046/j.1365-2958.1998.00869.x.

[114] K. Billet, J. Genitoni, M. Bozec, D. Renault, and D. Barloy, "Aquatic and terrestrial morphotypes of the aquatic invasive plant, Ludwigia grandiflora, show distinct morphological and metabolomic responses," *Ecol Evol*, vol. 8, no. 5, pp. 2568–2579, 2018, doi: 10.1002/ece3.3848.

[115] S.-H. Liu, P. C. Hoch, M. Diazgranados, P. H. Raven, and J. C. Barber, "Multi-locus phylogeny of Ludwigia (Onagraceae): insights on infra-generic relationships and the current classification of the genus," *Taxon*, vol. 66, no. 5, pp. 1112–1127, 2017.

[116] P. Maheswari, C. Kunhikannan, and R. Yasodha, "Chloroplast genome analysis of Angiosperms and phylogenetic relationships among Lamiaceae members with particular reference to teak (Tectona grandis L.f)," *J Biosci*, vol. 46, 2021, [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/34047286

[117] Y. Zhang *et al.*, "The Complete Chloroplast Genome Sequences of Five Epimedium Species: Lights into Phylogenetic and Taxonomic Analyses," *Front Plant Sci*, vol. 7, p. 306, 2016, doi: 10.3389/fpls.2016.00306.

[118] L. S. Huang *et al.*, "Development of high transferability cpSSR markers for individual identification and genetic investigation in Cupressaceae species," *Ecol Evol*, vol. 8, no. 10, pp. 4967–4977, 2018, doi: 10.1002/ece3.4053.

[119] P. Leontaritou, F. N. Lamari, V. Papasotiropoulos, and G. Iatrou, "Exploration of genetic, morphological and essential oil variation reveals tools for the authentication and breeding of Salvia pomifera subsp. calycina (Sm.) Hayek," *Phytochemistry*, vol. 191, p. 112900, 2021, doi: 10.1016/j.phytochem.2021.112900.

[120] M. Snoussi, L. Riahi, M. Ben Romdhane, A. Mliki, and N. Zoghlami, "Chloroplast DNA Diversity of Tunisian Barley Landraces as Revealed by cpSSRs Molecular Markers and Implication for Conservation Strategies," *Genet Res (Camb)*, vol. 2022, p. 3905957, 2022, doi: 10.1155/2022/3905957.

[121] S. L. Song, P. E. Lim, S. M. Phang, W. W. Lee, D. D. Hong, and A. Prathep, "Development of chloroplast simple sequence repeats (cpSSRs) for the intraspecific study of Gracilaria tenuistipitata (Gracilariales, Rhodophyta) from different populations," *BMC Res Notes*, vol. 7, p. 77, 2014, doi: 10.1186/1756-0500-7-77.

[122] G. L. Wheeler, H. E. Dorman, A. Buchanan, L. Challagundla, and L. E. Wallace, "A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology," *Appl Plant Sci*, vol. 2, no. 12, 2014, doi: 10.3732/apps.1400059.

[123] P. H. Raven and W. Tai, "Observations of Chromosomes in Ludwigia (Onagraceae)," 1979. [Online]. Available: https://about.jstor.org/terms

[124] S. H. Liu, K. H. Hung, T. W. Hsu, P. C. Hoch, C. I. Peng, and T. Y. Chiang, "New insights into polyploid evolution and dynamic nature of Ludwigia section Isnardia (Onagraceae)," *Bot Stud*, vol. 64, no. 1, Dec. 2023, doi: 10.1186/s40529-023-00387-8.

[125] J. L. Feng *et al.*, "Comparison Analysis Based on Complete Chloroplast Genomes and Insights into Plastid Phylogenomic of Four Iris Species," *Biomed Res Int*, vol. 2022, 2022, doi: 10.1155/2022/2194021.