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. A preliminary draft assembly was performed using *Lgh* short-reads (SR, 2*23,067,490
204 reads) with GetOrganelle v1.7.0 [41] and NOVOPlasty v4.2.1[42], and chloroplastic short and
205 long reads were extracted by mapping against this draft genome .
206 Chloroplastic short reads were then *de novo* assemble using Velvet (version 1.2.10) [43],
207 ABySS (version 2.1.5 [44][45]), MEGAHIT (1.1.2,[46]), and SPAdes (version 3.15.4,[47]),
208 without and with prior error correction. The best k-mer parameters were tested using kmergenie
209 [48] and k=99 was found to be optimal. For ONT reads, *Lgh* (550,516 reads) and *Lpm* (68,907
210 reads) reads were self-corrected using CANU 1.8 [49] or SR-corrected using Ratatosk [50] and

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

213 **Post plastome assembly validation**

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

222 **Chloroplast genome annotation**

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

233 **SSRs and Repeat Sequences Analysis**

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

242 **Comparative chloroplast genomic analyses**

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

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

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

264 **Graphic representation**

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

270 **Results**

271 **Plastome short read assembly**

272 The chloroplastic fraction of *Lgh* short reads (SR) was extracted by mapping against the two
273 draft haplotypes generated by GetOrganelle, which differ only by a "flip-flop" of the SSC region
274 (Figure 1). This subset (1,360,507 reads) were assembled using ABySS, Velvet, MEGAHIT
275 and SPAdes in order to identify the best assembler for this plant model. As shown in Figure 2,
276 both the number and size of contigs depend greatly on the algorithms used and the correction

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

310 **Annotation and comparison of *Ludwigia* plastomes**

311 **1. General Variations**

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

322 All three *Ludwigia* sp. plastomes contain the same number of functional genes (134 in total)
323 encoding 85 proteins (embracing 7 duplicated in the IR region: *ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*,
324 *ycf2*, *ycf15*), 37 tRNAs (including trnK-UUU which contains *matK*), and 8 rRNAs (16S, 23S,
325 5S, and 4.5S as duplicated sets in the IR). Among these genes, 18 contain introns, of which six
326 are tRNAs (Table 2). Only the *rps12* gene is a trans-spliced gene. A total of 46 genes are
327 involved in photosynthesis, and 71 genes related to transcription and translation, including a
328 bacterial-like RNA polymerase and 70S ribosome, as well as a full set of transfer RNAs
329 (tRNAs) and ribosomal RNAs (rRNAs). Six other protein-coding genes are involved in
330 essential functions, such as *accD,* which encodes the β-carboxyl transferase subunit of acetyl-
331 CoA carboxylase, an important enzyme for fatty acid synthesis; *matK* encodes for maturase K,
332 which is involved in the splicing of group II introns; *cemA*, a protein located in the membrane
333 envelope of the chloroplast is involved in the extrusion of protons and therby indirectly allows
334 the absorption of inorganic CO2 in the plastids; *clpP1* which is involved in proteolysis, and;
335 *ycf1*, *ycf2*, two ATPases members of the TIC translocon. Finally, a highly pseudogenized *ycf15*
336 locus was annotated in the IR even though premature stop codons indicate loss of functionality.

337 **2. Segments Contractions/Expansion**

338 The junctions between the different chloroplast segments were compared between three
339 *Ludwigia* sp. (*Lpm*, *Lgh* and *Lo),* and we found that the overall resemblance of *Ludwigia* sp.
340 plastomes was confirmed at all junctions (Figure 4A). In all three genomes, *rpl22, rps19, and*
341 *rpl2* were located around the LSC/IRb border, and *rpl2*, *trnH*, and *psbA* were located at the
342 IRa/LSC edge. The JSB (junction between IRb and SSC) is either located in the *ndhF* gene or
343 the *ycf1* gene depending on the orientation of the SSC region (Figure 4B). The *ycf1* gene was

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

367 **3. Repeats and SSRs analysis**

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

378 The chloroplast genomes of the three *Ludwigia* species were also screened for long repeat
379 sequences. They were counted in a non-redundant way (if smaller repetitions were included in
380 large repeats, only the large ones were considered). Four types of repeats (tandem, palindromic
381 inverted and direct) were surveyed in the three *Ludwigia* sp. plastomes. No inverted repeats
382 were detected with the criteria used.

383 For the three other types of repeats, here are their distributions:

384 ***Tandem repeats* (Table 3A)**: Perfect tandem repeats (TRs) with more than 15 bp were
385 examined. Twenty-two *loci* were identified in the three *Ludwigia* sp. plastomes (*Lgh*, *Lpm*, *Lo*),
386 heterogeneously distributed as shown in Table 3A: 13 loci (plus one imperfect) in *Lo*, nine loci
387 (plus one imperfect) in *Lgh* and seven loci (plus two imperfect) in *Lpm*. It can therefore be seen
388 that the TR distributions (occurrence and location) are specific to each plastome, since only
389 four pairs are common to the three species. Thus, nine TRs are unique to *Lo*, three to *Lpm* and
390 three to *Lgh.* Two pairs are common to *Lgh* and *Lpm* and one is common to *Lo* and *Lgh.* TRs
391 are mainly intergenic or intronic but are detected in two genes (*accD* and *ycf1)*. These genes
392 have accelerated substitution rates, although this does not generate a large difference in their
393 lengths. This point will be developed later in this article.

394 ***Direct repeats* (Table 3B)**: There are few direct (non-tandem) repeats (DRs) in the
395 chloroplast genomes of *Ludwigia* sp. A single direct repeat of 41 nt is common to the three
396 species, at 2 kb intervals, in *psaB* and *psaA* genes. This DR corresponds to an amino acid repeat
397 [WLTDIAHHHLAIA] which corresponds to a region predicted as transmembrane. We then
398 observe three direct repeats conserved in *Lpm* and *Lgh* in *ycf1*, *accD* and *clpP1* respectively,
399 two unique DRs in *Lo* (in the *accD* gene and *rps12-clpP1* intergene) and one in *Lgh* (in the
400 *clpP1* intron 1 and *clpP1* intron 2).

401 ***Palindromes* (Table 3C)**: Palindromic repeats make up the majority of long repetitions,
402 with the numbers of perfect repeats varying from 19, 24 and 26 in *Lo*, *Lgh* and *Lpm*,
403 respectively, and the number of quasi-palindromes (1 mutation) varying between 8, 3 and 6.
404 They are mainly found in the intronic and intergenic regions, with the exception of six genic
405 locations in *psbD*, *ndhK*, *ccsA* and *rpl22,* and two palindromic sequences in *ycf2.* These gene
406 palindromic repeats do not seem to cause genetic polymorphism in *Ludwigia* and can be
407 considered as silent.

408 Thirteen palindromes are common to the three species (including 2 with co-variations in
409 *Lo*). 13 others present in *Lpm* and *Lgh* correspond to quasi-palindromes (QPs) in *Lo* due to
410 mutated bases, and conversely, three *Lo* perfect palidromes are mutated in *Lpm* and *Lgh*.
411 Finally, only five palindromes are species specific. Two in particular are located in the

412 hypervariable intergenic spacer *ndhF-rpl32*, and are absent in *Lo* due to a large deletion of 160
413 nt.

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

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

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

431 Finally, in the LSC region, the longest segment, which consequently contains the maximum
432 number of repeats, we still observed a preferential localization in the intergenic and intronic
433 regions since only genes *atpA*, *rpoC2*, *rpoB*, *psbD*, *psbA*, *psbB*, *ndhK* and *clpP1* contain either
434 mononucleotic repeats (poly A and T), palindromes, or microsatellites (most often common to
435 the three species and without affecting the sequences of the proteins produced). As mentioned
436 earlier, the only exception is the *accD* gene, which contains several direct and tandem repeats
437 in *Lgh* and *Lpm,* corresponding to a region of 174 nt (58 amino acids) missing in *Lo* and,
438 conversely, a direct repeat of 40 nucleotides, in a region of 147 nt (49 aa), which is present in
439 *Lo* and missing in the other two species. These tandem repeats lead to the presence of four
440 copies of 9 amino acids [DESENSNEE] in *Lgh* and *Lpm*, two of which form a larger duplication
441 of 17 aa [FLSDSDIDDESENSNEE]. Similarly, the TRs present only in *Lo* generate two perfect
442 9 amino acid repeats [EELSEDGEE], included in two longer degenerate repeats of 27 nt (Add.
443 Figure 6). It should be noted that though these TRs do not disturb the open reading phases, it is
444 still possible for them to form an intron which is not translated. Different functional studies will
445 be necessary to clarify this point. The presence of polymorphisms of the *accD* gene between

446 *Lo* and the two species (*Lpm*, *Lgh*) is interesting because *accD*, that encodes a subunit of acetyl-
447 CoA carboxylase (EC 6.4.1.2). This enzyme is essential in fatty acid synthesis and also
448 catalyzes the synthesis of malonyl-CoA, which is necessary for the growth of dicots, plant
449 fitness and leaf longevity, and is involved in the adaptation to specific ecological niches [70].
450 Large *accD* expansions due to TRs have also been described in other plants such as *Medicago*
451 [71] and *Cupressophytes* [72]. Some authors have suggested that these inserted repeats are not
452 important for acetyl-CoA carboxylase activity as the reading frame is always preserved, and
453 they assume that these repeats must have a regulatory role [73].

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

455 Determination of divergent regions by MVista, using *Lo* as a reference, confirmed that the
456 three *Ludwigia* sp. plastomes are well preserved if the SSC segment is oriented in the same way
457 (Add. Figure 7). Sliding window analysis (Figure 6) indicated variations in definite coding
458 regions, notably *clpP*, *accD*, *ndh5*, *ycf1* with high Pi values, and to a lesser extent, *rps16*, *matK*,
459 *ndhK*, *petA*, *ccsA* and four tRNAs (*trnH*,*trnD*, *trnT* and *trnN*). These polymorphic *loci* could be
460 suitable for inferring genetic diversities in *Ludwigia* sp.

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

478 Our results show that the Ka/Ks ratio is less than 1 for most genes (Figure 7). This indicates
479 adaptive pressures to maintain the protein sequence except for *matK* (1.17 between *Lgh* and

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

Our findings demonstrate that the utilization of solely short reads has failed to produce complete *Ludwigia* plastomes, likely due to challenges posed by long repeats and rearrangements. On the other hand, relying solely on long reads resulted in a lower quality sequence due to insufficient coverage and sequencing errors. After conducting our research, we discovered that hybrid assembly, which incorporates both long and short read sequences, resulted in the most superior complete assemblies. This innovative approach capitalizes on the advantages of both sequencing technologies, harnessing the accuracy of short read sequences and the length of long read sequences. In the case of our study on *Ludwigia* plastomes reconstruction, hybrid assembly was the most complete and effective, similarly to studies on other chloroplasts, such as those in *Eucalyptus* [82], *Falcataria* [83], *Carex* [84] or *Cypripedium* [85].

In our study, we were able to identify the presence of two haplotypes in *Lgh*, which is a first for *Ludwigia* (and more broadly within Onagraceae), as the plastome of *L. octovalvis* was only delivered in one haplotype [86]. Due to the unavailability of sequence data for *Ludwigia octovalvis* and our exclusive use of long reads for *Ludwigia peploides*, we are unable to conclusively identify the presence of these two forms in the *Ludwigia* genus. However, we believe that they are likely to be present. Unfortunately, the current representation of plastomes in GenBank primarily consists of short-read data, which may result in an underrepresentation of this polymorphism. It is unfortunate that structural heteroplasmy, which is expected to be widespread in angiosperms, has been overlooked. Existence of two plastome haplotypes has been identified in the related order of Myrtales (Eucalyptus sp.), in 58 species of Angiosperms, [87], Asparagales (*Ophrys apifera* orchid [88]), Brassicales (*Carica papaya*, *Vasconcellea pubescens* [89]), Solanales (*Solanum tuberosum* [90]), Laurales (Avocado *Persea americana* [91]) and Rhamnaceae (*Rhamnus crenata* [92]). However, the majority of reference plastomes in the current GenBank database (Release 260: April 15, 2024) are described as a single haplotype, indicating an underrepresentation of structural heteroplasmy in angiosperm chloroplasts. This underscores the importance of sequencing techniques, as the database is predominantly composed of short-read data (98%), which are less effective than long reads or hybrid assemblies at detecting flip-flop phenomena in the LSC region.

The chloroplast genome sizes for the three genera of Onagraceae subfam. Onagroideae varied as follows: *Circaea* sp. ranged from 155,817 bp to 156,024 bp, *Chamaenerion* sp. ranged from 159,496 bp to 160,416 bp, and Epilobium sp. ranged from 160,748 bp to 161,144 bp [93]. Our study revealed that the size of the complete chloroplast of *Ludwigia* (Onagraceae subfamily Ludwigioideae) ranged from 159,369 bp to 159,584 bp, which is remarkably similar to other Onagraceae plants (average length of 162,030 bp). Furthermore, *Ludwigia* plastome sizes are

548 consistent with the range observed in Myrtales (between 152,214 to 171,315 bp [94]). In the
549 same way, similar overall GC content was found in *Ludwigia sp*. (from 37.3 to 37.4%), *Circaea*
550 *sp.* (37.7 to 37.8%), *Chamaenerion sp.* and *Epilobium sp.* (38.1 to 38.2%,[93]) and more
551 generally for the order Myrtales (36.9–38.9%, with the average GC content being 37%,[94]).
552 Higher GC content of the IR regions (43.5%) found in *Ludwigia* sp. has already been shown in
553 the Myrtales order (39.7–43.5%) and in other families/orders such as Amaranthaceae (order
554 Caryophyllales *[95]*) or Lamiaceae (order Lamiales [96]), and is mainly due to the presence of
555 the four GC rich rRNA genes.
556 The complete chloroplast genomes of the three *Ludwigia* species encoded an identical set of
557 134 genes including 85 protein-coding genes, 37 tRNA genes and eight ribosomal RNAs,
558 consistent with gene content found in the Myrtales order, with a gene number varying from 123
559 to 133 genes with 77–81 protein-coding genes, 29–31 tRNA gene and four rRNA genes [94].
560 Chloroplast genes have been selected during evolution due to their functional importance[97].
561 In our current study, we made the noteworthy discovery that *matK*, *accD*, *ycf2*, and *ccsA* genes
562 were subjected to positive selection pressure. These genes have frequently been reported in
563 literature as being associated with positive selection, and are known to play crucial roles in
564 plant development conditions. *Lgh* and *Lpm* are known to thrive in aquatic environments, where
565 they grow alongside rooted emergent aquatic plants, with their leaves and stems partially
566 submerged during growth, as reported by Wagner et al. in 2007 [1]. Both species possess the
567 unique ability of vegetative reproduction, enabling them to establish themselves rapidly in
568 diverse habitats, including terrestrial habitats, as noted by Haury et al [98]. Additionally, *Lo* is
569 a wetland plant that typically grows in gullies and at the edges of ponds, as documented by
570 Wagner *et al.* in 2007 [1]. Given their ability to adapt to different habitats, these species may
571 have evolved specialized mechanisms to cope with various abiotic stresses, such as reduced
572 carbon and oxygen availability or limited access to light in submerged or emergent conditions.
573 Concerning *matK*, Barthet et al [99] demonstrated the relationship between light and
574 developmental stages, and MatK maturase activity, suggesting important functions in plant
575 physiology. This gene has recently been largely reported to be under positive selection in an
576 aquatic plant (*Anubias* sp.,[100]), and more generally in terrestrial plants (*Pinus sp* [101]or
577 *Chrysosplenium sp.* [102]). The *accD* gene has been described as an essential gene required for
578 leaf development [103] and longevity in tobacco (*Nicotiana tabacum*)[104]. Under drought
579 stress, plant resistance can be increased by inhibiting *accD* [105], and conversely, enhanced in
580 response to flooding stress by upregulating *accD* accumulation [106]. Hence, we can
581 hypothesize that the positive selection observed on the *accD* gene can be explained by the

582 submerged and emerged constraints undergone by *Ludwigia* species. The *ycf2* gene seems to
583 be subject to adaptive evolution in *Ludwigia species*. Its function, although still vague, would
584 be to contribute to a protein complex generating ATP for the TIC machinery (proteins importing
585 into the chloroplasts [107][108]), as well as plant cell survival [109][110]. The *ccsA* gene
586 positive selection is found in some aquatic plants such as *Anubia sp.*[100], marine flowering
587 plants as *Zostera* species [111], and some species of Lythraceae [105]. The *ccsA* gene is
588 required for cytochrome c biogenesis [112] and this hemoprotein plays a key role in aerobic
589 and anaerobic respiration, as well as photosynthesis [113]. Furthermore, we showed that *Lgh*
590 colonization is supported by metabolic adjustments mobilizing glycolysis and fermentation
591 pathways in terrestrial habitats, and the aminoacyl-tRNA biosynthesis pathway, which are key
592 components of protein synthesis in aquatic habitats [114]. It can be assumed that the ability of
593 *Ludwigia* to invade aquatic and wet environments, where the amount of oxygen and light can
594 be variable, leads to a high selective pressure on genes involved in respiration and
595 photosynthesis.

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

612 Concerning the MatK-based phylogenetic tree, its topology is generally congruent with the
613 first molecular classification of Liu *et al.* [115] as all *Ludwigia* from sect *Jussiaea* (clade B1)
614 and sect. *Ludwigia* (clade A1) and sect. *Isnardia* (clade A2) branched together. In this MatK-
615 based tree, *Ludwigia prostrata*, a species absent from previously published phylogenetic

616 studies, positions itself alone at the root of the *Ludwigia* tree. This species, sole member of
617 section *Nematopyxis*, is related as having no close relatives [123], finding supported by our
618 work. We also observed that *Ludwigia ovalis* branches within sect. *Jussiaea*, as its 258 amino
619 acids partial MatK sequence (ca. half of the complete sequence) is identical to the MatK
620 proteins of *L. grandiflora*, *L. stolonifera* and *L. adscendens*. Its phylogenetic placement remains
621 unresolved: classified alone by Raven (1963) [5] and Wagner (2017) [22] in sect. *Miquelia*,
622 later positioned by Liu et al. (2017)[4] within the *Isnardia-Microcarpium* section (using nuclear
623 DNA) or as sister to it (using plastid DNA). For this reason, conducting a whole plastome
624 analysis would be valuable to provide insights into *L. ovalis* phylogenetic positioning. Another
625 species positioned on the margins of sect. *Isnardia* (clade A2) is *Ludwigia suffruticosa*
626 (previously classified in sect. *Microcarpium*), which branches within sect. *Ludwigia* (clade A1).
627 This positioning raises questions about the current grouping of sections *Isnardia*, *Michelia*, and
628 *Microcarpium* into a single section *Isnardia* as proposed by Liu et al. (2023) [124] and
629 highlights that plastid protein coding markers can provide differing phylogenetic insights.
630 Finally, the last species positioned differently of this clade (clade B4) is *Ludwigia decurrens*
631 (sect. *Pterocaulon*) which clusters with *L. leptocarpa* (clade B3) and *L. bonariensis* (clade B4a).
632 However, it is important to note that in their study, Liu et al. (2017) indicate that clade B4 is
633 moderately supported and that the two members of sect. *Pterocaulon*, *L. decurrens* and *L.*
634 *nervosa*, diverge in all trees [4]. In summary, acquiring complete plastomes for *Ludwigia* sp.
635 could significantly enhance our understanding of the phylogeny of this complex genus.
636 Furthermore, comparing nuclear and plastid phylogenies would help determine if they reflect
637 the same evolutionary history and whether plastid phylogeny alone can accurately reconstruct
638 the phylogeny of *Ludwigia* genus.
639

640 **Conclusion**

641 In this study, we conducted the first-time sequencing and assembly of the complete plastomes
642 of *Lpm* and *Lgh*, which are the only available genomic resources for functional analysis in both
643 species. We were able to identify the existence of two haplotypes in both *Lpm* and *Lgh*, while
644 the absence of the *Lo* genome precluded further investigation for this species. Comparison of
645 all 10 Onagraceae plastomes revealed a high degree of conservation in genome size, gene
646 number, structure, and IR boundaries. However, to further elucidate the phylogenetic analysis
647 and evolution in *Ludwigia* and Onagraceae, additional chloroplast genomes will be necessary,
648 as highlighted in recent studies of Iris and Aristidoideae species [125].
649

Declarations

- Availability of data and materials

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

- Conflict of interest disclosure

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

- Funding

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

- Acknowledgements

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

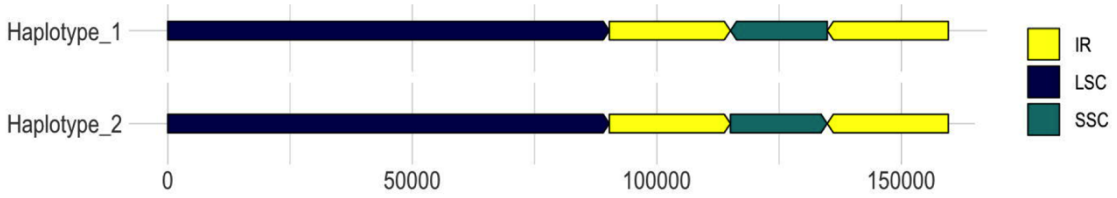


Figure 1: Two structural haplotypes of *L. grandiflora* plastomes representing the flip-flop organization of SSC segment.

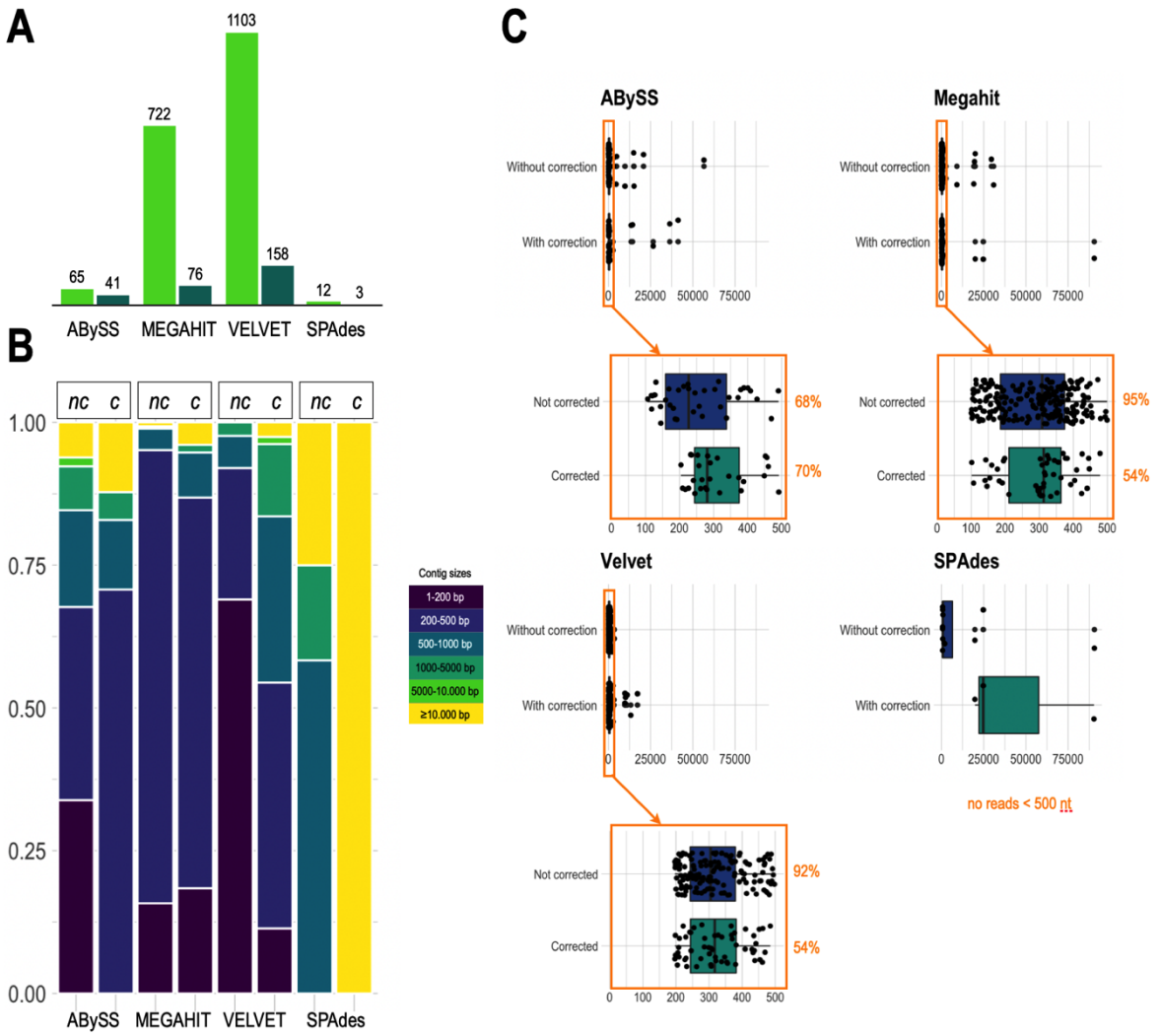


Figure 2: Comparative results of *L. grandiflora* short read (SR) assemblies. **A:** Total number of contigs obtained with the uncorrected (dark green) and corrected (light green) chloroplast SRs for the 4 assemblers (ABySS, MEGAHIT, Velvet and SPAdes). **B:** Comparison of the size of contigs assembled 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.

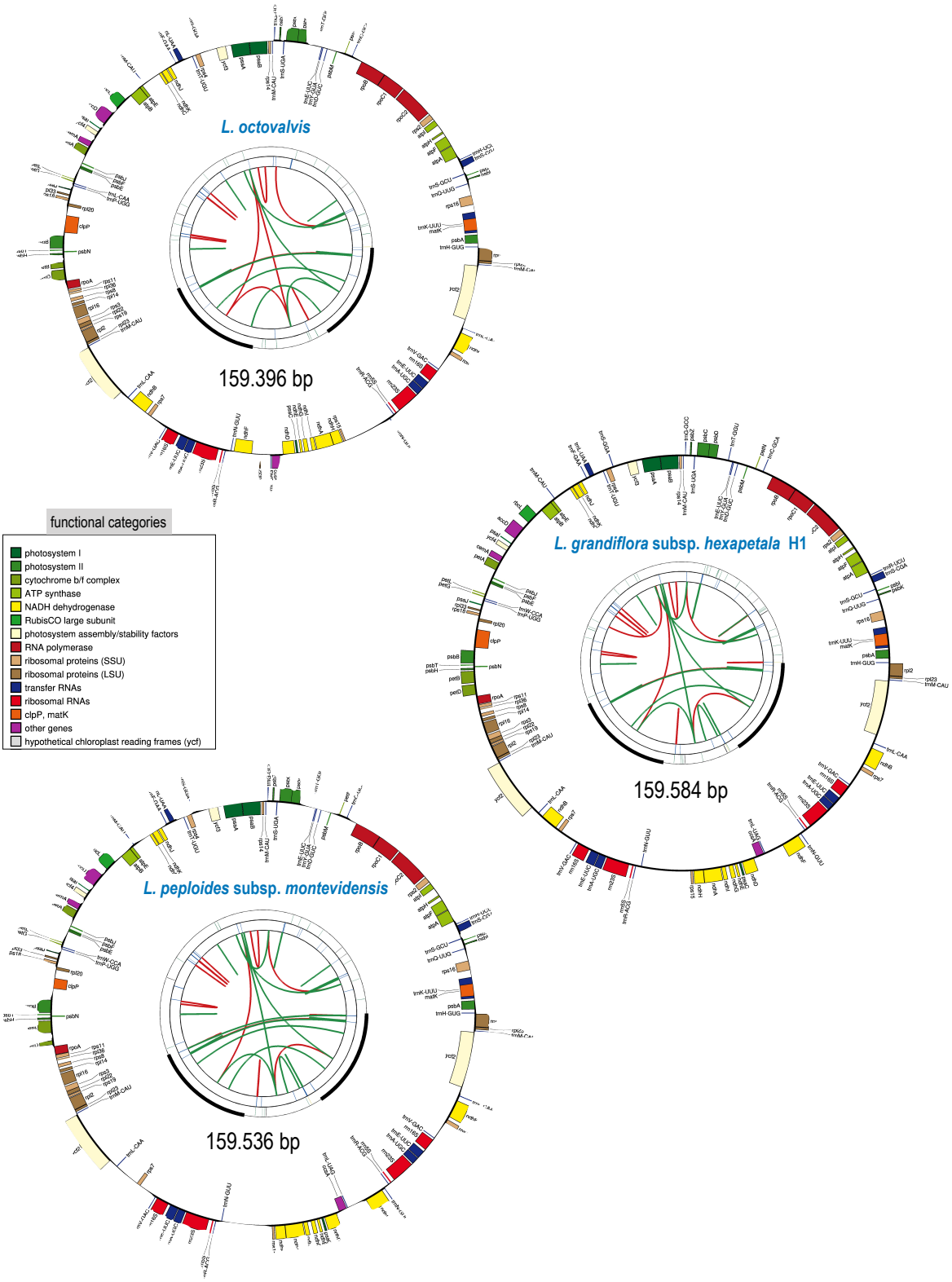


Figure 3: Circular representation of annotations plastomes in *Ludwigia octovalis*, *Ludwigia grandiflora* and *Ludwigia peploides* using ogdraw. Each card contains four circles. From the center outwards, the first circle shows forward and reverse repeats (red and green arcs, respectively). The next circle shows tandem repeats as bars. The third circle shows the

689 microsatellite sequences. Finally, the fourth and fifth circles show the genes colored according
690 to their functional categories (see colored legend). Only the haplotype 1 of *L. grandiflora* is
691 represented as haplotype 2 only diverge by the orientation of the SSC segment.
692

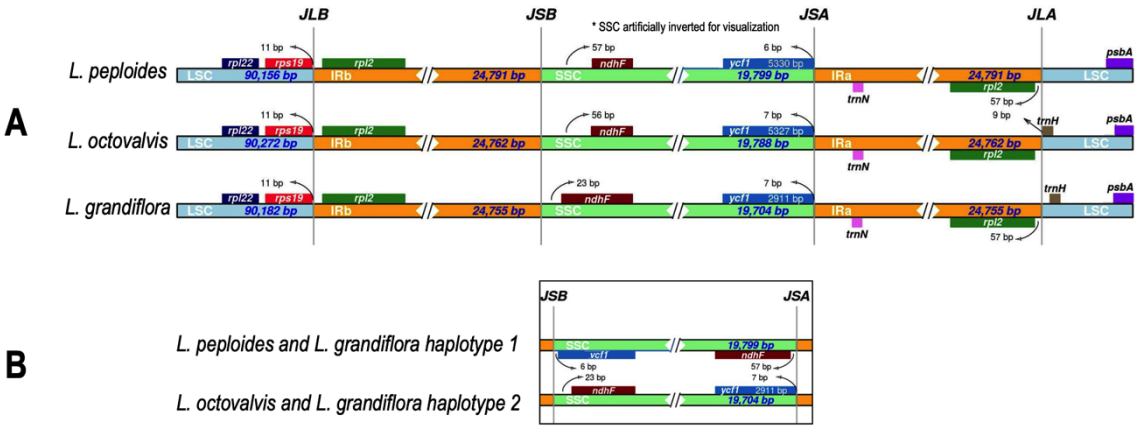


Figure 4: Comparison of the borders of LSC, SSC, and IR regions in Onograceae plastomes. **A:** Comparison of the junction between large single-copy (LSC, light blue), inverted repeat (IR, orange) and short single-copy (SSC, light green) regions among the chloroplast genomes of *L. octovalvis*, *L. peploides* 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.

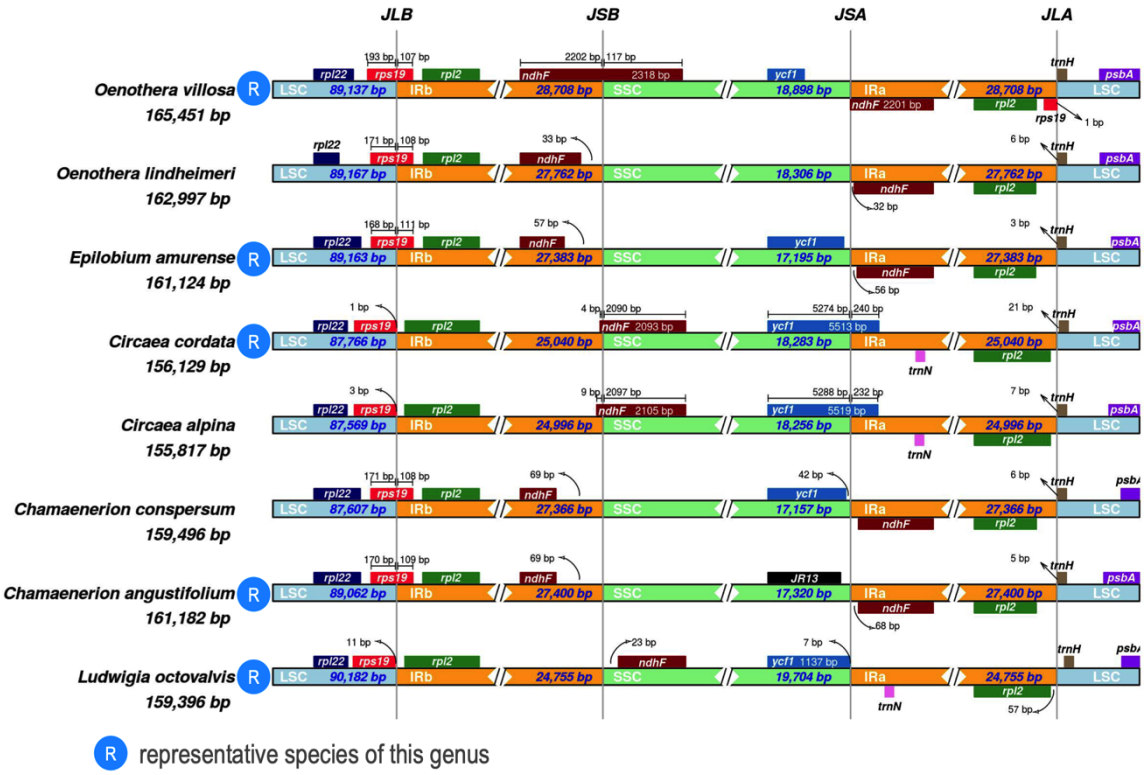


Figure 5: Comparison of LSC, SSC and IR regions boundaries in Onograceae chloroplast genomes. Representative sequences from each genus have been chosen (noted R on the diagram) except for *Oenothera lindheimeri* (only 89.35 % identity with others *Oenothera*), *Circaea alpina* (99.5 % identity but all 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).

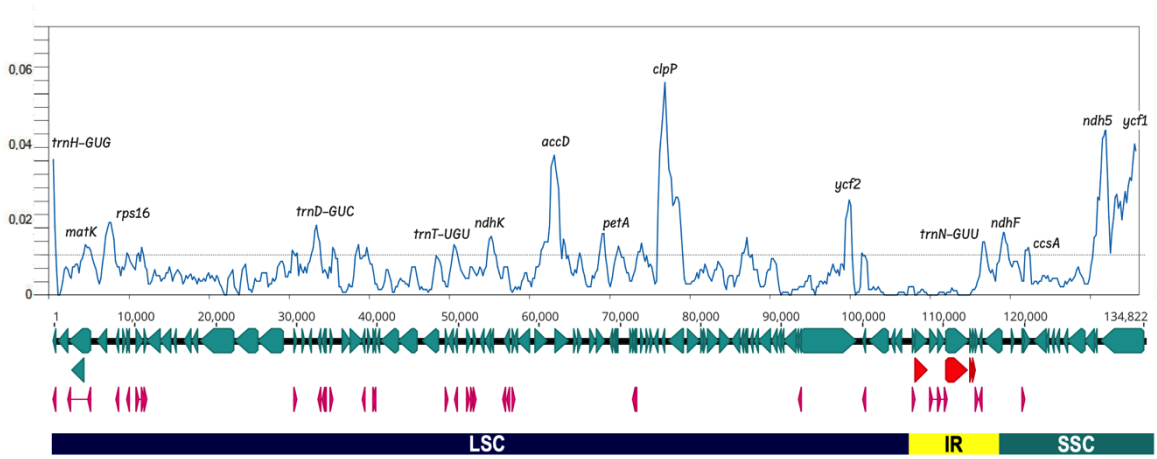


Figure 6: Illustration of nucleotide diversity of the three *Ludwigia* chloroplast genome sequences. The graph was generated using DnaSP software version 6.0 (windows length: 800 bp, step size: 200 bp) [64][63]. The x-axis corresponds to the base sequence of the alignment, and the y-axis represents the 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.

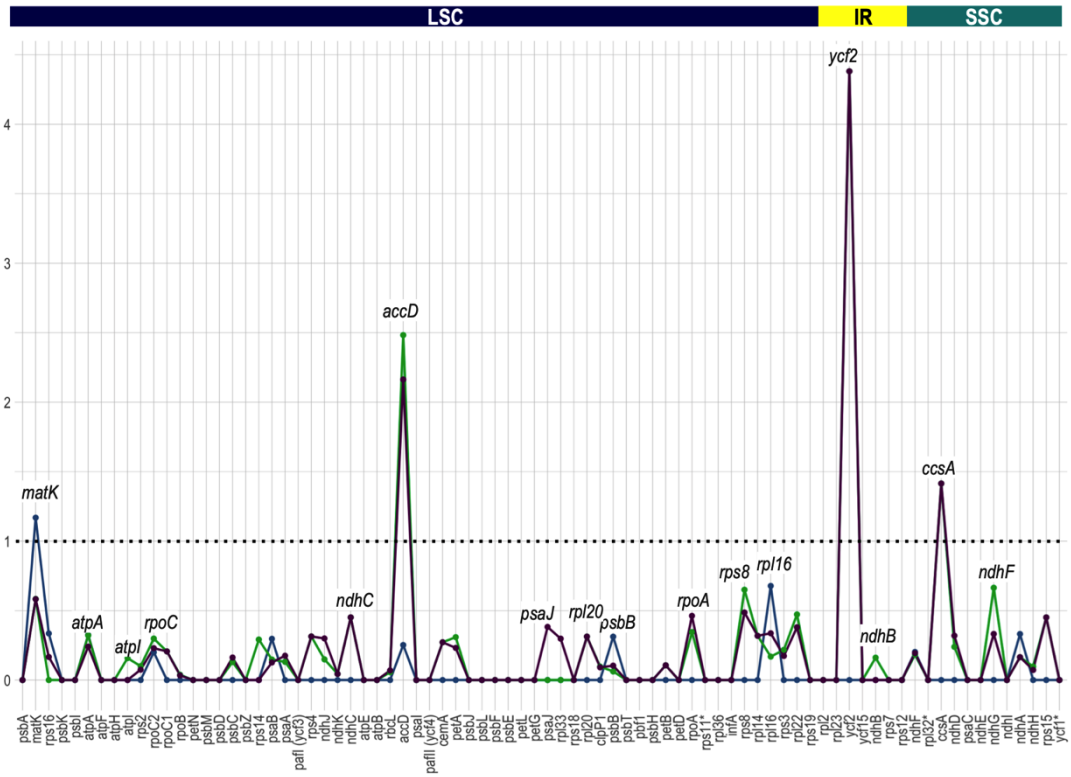


Figure 7: The Ka/Ks ratios of the 80 protein-coding genes of *Ludwigia* plastomes. The blue curve represents *L. grandiflora* versus *L. peploides*, purple curve denotes *L. grandiflora* versus *L. octovalvis* and green curve *L. peploides* versus *L. octovalvis*. Four genes (*matK*, *accD*, *ycf2* and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

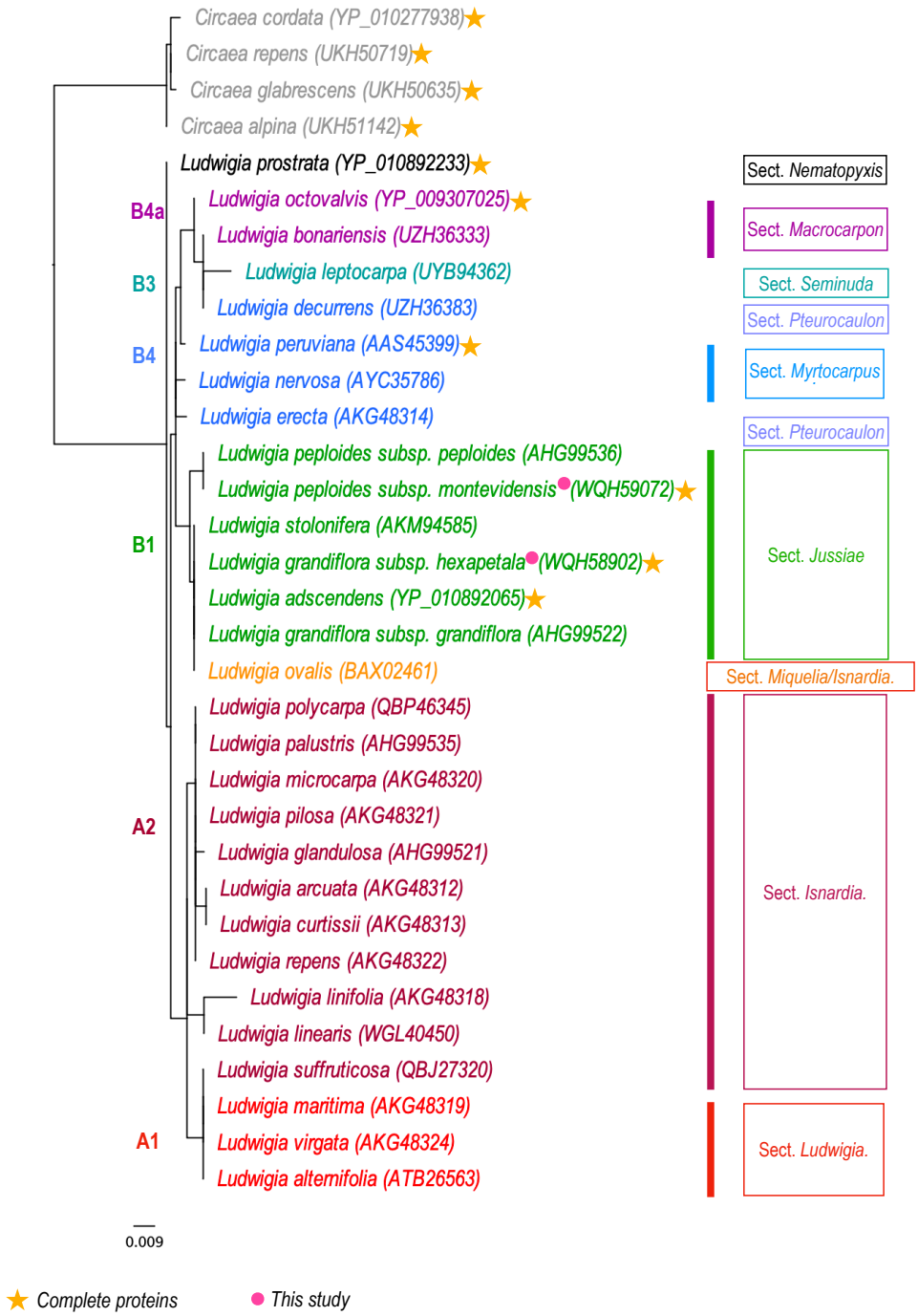


Figure 8: Phylogenetic tree based on *Ludwigia* MatK protein sequences. Only six *Ludwigia* sequences are complete (yellow star), the others correspond to amino acids ranging from 128 to 289 aa, with an average of 244 aa. Clades are named and colored regarding the *Ludwigia* phylogeny proposed by Liu 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.

741

Table 1. The general characteristics of the 3 Ludwigia plastomes

	*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)

742

743

744

745

Table 2 : Genes present in the plastomes of *Ludwigia*

FUNCTION	NAME
Photosynthesis	
Rubisco	rbcL
Photosystem I (PSI)	psaA; psaB; psaC; psaI; psaJ
PSI assembly factors	ycf3# (pafI); ycf4 (pafII)
Photosystem II	psbA; psbB; psbC; psbD; psbE; psbF; psbH; psbI; psbJ; psbK; psbL; psbM; pbf1 (psbN) psbT; psbZ
ATP synthase	atpA; atpB; atpE; atpF#; atpH; atpI
Cytochrome *b6f*	petA; petB#; petD#; petG; petL; petN
Cytochrome biogenesis	ccsA
NADPH dehydrogenase	ndhA#; ndhB**#; ndhC; ndhD; ndhE; ndhF; ndhG; ndhH; ndhI; ndhJ
Transcription and translation	
Transcription	rpoA; rpoB; rpoC1#; rpoC2
Small ribosomal proteins	rps2; rps3; rps4; rps7**; rps8; rps11; rps12**#; rps14; rps15; rps16#; rps18; rps19
Large ribosomal proteins	rpl2**#; rpl14; rpl16#; rpl20; rpl22; rpl23**; rpl32; rpl33; rpl36
Translation initiation	infA
Ribosomal RNA	rrn5**; rrn4;5**; rrn16**; rrn23**
Transfer RNA	trnA-UGC**#;trnC-GCA;trnD-GUC;trnE-UUC;trnF-GAA;trnfM-CAU;trnG-GCC;trnG-UCC#;trnH-GUG;;trnI-CAU**;trnI-GAU**#;trnK-UUU#;trnL-CAA**;trnL-UAA#;trnL-UAG;trnM-CAU;trnN-GUU**;trnP-UGG;trnQ-UUG;trnR-ACG**;trnR-UCU;trnS-GCU;trnS-GGA;trnS-UGA;trnT-GGU;trnT-UGU;trnV-GAC**;trnV-UAC#;trnW-CCA;trnY-GUA
Other functions	
Group II intron splicing	matK
Inorganic carbon uptake	cemA
Protease	clpP1#
Fatty acid synthesis/Heat tolerance	accD
TIC machinery (protein import)	ycf1 (Tic214); ycf2**
Unknown function	ycf15**
	** duplicated in IR region; # spliced genes

746

747

Table 3:

Table 3A : Tandem repeats

Sequence	*L. octovalis (L.o)*	*L. grandiflora (L.g)*	*L. peploides (L.p)*	Length	Region	Locus	Comments
TTGTAGTCAGGGGTGTAGTACTAT				24	IRs	*ycf2*	
TAGAAGAGAGTGCAG		X	X	15	IRs	*ycf2*	15 nt deletion in L.g and L.p
ATGAAATATCGTATAATGAAGTACCACACGAGTGGATAT	X	X		39	IRs	*rpl2 intron*	39 nt deletion in L.g and L.o
AAAAATAGGATAGGAT		X	X	16	LSC	*ycf1-trnH-GUG*	56 nt deletion in L.g and L.p
TAAATTAATATCTATATA		X	X	18	LSC	psbZ-trnG-GCC	18 nt deletion in L.g and L.p
TTTTCTATCTATCTTATATCAA		X	X	22	LSC	*trnK-UUU-rps16*	22 nt deletion in L.g and L.p
AGATCCATAACATCATCAAA		X	X	20	LSC	*rps16 intron*	22 nt deletion in L.g and L.p
TATTAGTTATTAATATTATTAGA		X	X	23	LSC	trnP-UGG-psaJ	23 nt deletion in L.g and L.p
AATAATATATAATAACTTAAATA		X	X	23	LSC	*rpl33-rps18*	33 et 44 nt nt deletion in in L.g et L.p, respectively
TTTTTATTTAACATGCTATCAAATCAACAATGCCATACCGTAGGGCATCTGTT		X	X	53	LSC	*rpl20-clpP1*	107 nt deletion in L.g and L.p
ATATATTTCGATTCAATTC	X		X	19	LSC	*trnH-GUG-psbA*	3 copies in a 57 nt deletion in L.o and L.p
ATAGAAATATCAGTATTTGAGTG	X		X	23	LSC	*atpH-atpI*	23 nt deletion in L.o and L.p
TTAATTTTAATTGAAGAA	X		X	18	LSC	*psbJ-psbL*	17 and 24 nt deletion in L.o and L.p, respectively
TTAAAGAATATTAATATTC	imperfect TR			19	LSC	*trnR-UCU-atpA*	A -> C mutation in second copy in L.o
TATTATTATTATTAAT	X	X		16	LSC	*atpH-atpI*	16 nt deletion in L.g and L.o
TCTAAGGCTGAAATAAGG	X	X		18	LSC	pafI intron	18 nt deletion in L.g and L.o
TGTGAATCTATCTAT			X	15	LSC	*trnS-UGA-psbZ*	8 nt deletion in L.p
TTTTTTCTAGTA				12	LSC	*pafI intron*	
CTAGTTATTGACATGG		imperfect TR	imperfect TR	16	LSC	*psaJ-rpl33*	G -> A mutation in second in L.p et L.g
ATTTTTATTAACTCT	X		imperfect TR	15	SSC	*ycf1*	T->A mutation in first copy in L.p, other sequence in first copy in L.o
AATCAAATAGTTGAT		X	X	15	SSC	*ycf1*	other sequence in first copy of L.p and L.g
ATAATAATATATTTATTATTAATTAATA	X			28	SSC	*ndhF-rpl32*	160 nt deletion in L.o

Table 3B : Direct repeats

Sequence	*L. octovalis (L.o)*	*L. grandiflora (L.g)*	*L. peploides (L.p)*	Size (nt)	Spacers (nt)	Region	Locus	Comments
TTCAATTGGAACGGACGATTCGTCAATCATCT				32	37	SSC	*ycf1*	2 copies. In L.o, one mutation (G->A) in the second copie
CATCGATGATGAAAGTGAAAACAGTAATGAAGAGG	X			35	28 - 22 - 11	LSC	*accD*	3 perfects copies and 1 mutated (G->A) copie in *L.g* and *L.p*. Region of 174 nt deleted in *L.o*
AGATGGTGAAGAACCTTATGAAGATGGTGAAGAACCTTATG		X	X	41	22	LSC	*accD*	Region of 147 nt deleted in *L.g* and *L.p*
TATCAAATCAACAATGCCATACCGTAGGGCAT		X	X	32	22 - 21	LSC	*rps12-clpP1*	3 copies
TTAAGAGCCGTACAGGCACCTTTTGATGCATACGG	X				408 in *L.p*, 406 in *L.g*	LSC	*clpP1*	2 copies. In L.g, one mutation (C->T) in the second copie
TTAAGAGCCGTACAGGCACTTTTTGATGCATACGG	X		X	35	811	LSC	*clpP1 intron 1-intron 2*	
TGCAATAGCCAAATGATGATGAGCAATATCAGTCAGCCATA				41	2178		*psaB* & *psaA*	

Table 3C : Palindromic repeats

Common perfect palidromic repeats			
AGACTCTCATGAGAGTCT		*trnC-GCA - petN*	
ATTAAATAGAATATTCTATTTAAT		*trnE-UUC-trnT-GGU*	
TTGGTAAATTTACCAA		*psbD*	
TTCATTTCAATTTCAATTGAAATTGAAATGAA		*trnI-CAU-ycf2*	2 copies in IR
GAAAAAGGCCTTTTTC		*ycf2*	2 copies in IR
TCTCAAATGATTAATCATTTGAGA		*trnL-UAA* intron	
GGATTACTAGTAATCC		*trnD-GUC-trnY-GUA*	
TTTGAATGCATTCAAA		*trnG-UCC* intron	
ATATATTCGAATATAT		*trnG-UCC -trnR-UCU*	
TAGTAATTAATTACTA		*trnG-GCC-trnfM-CAU*	
CCAGTATGCATACTGG		*ndhK*	
Common palidromic repeats with covariation			
in L. octovalvis	*in L. grandiflora et L. peploides*		
ATAGAATCTATATTCTATTAGAATATAGATTCTAT	ATCGAATCTATATTCTATTAGAATATAGATTCGAT	*ndhC-trnV-UAC*	
ATGTATATATCGAT	ATCTATATATATAGAT	*trnE-UUC-trnT-GGU*	
Common palindromic and quasi-palidromic repeats			
in L. octovalvis	*in L. grandiflora and L. peploides*		
TTTAACGAATATTAATATT t GTTAAA	TTTAACGAATATTAATATTCGTTAAA	*trnR-UCU-atpA*	
TTAA c GAATATTAATATTCTTTAA	TTAAAGAATATTAATATTCTTTAA	*trnR-UCU-atpA*	
AATTGTA c TTACAATT	AATTGTAATTACAATT	*ccsA*	
AGGAAGATTGATCAATCTT t CT	AGGAAGATTGATCAATCTTCCT	*trnL-UAG-rpl32*	
TTA c TAATATTACTAA	TTAGTAATATTACTAA	*trnK-UUU* intron	
ATATAGAATAT c CTATAT	ATATAGAATATTCTATAT	*psbZ-trnG-GCC*	
ACATATCATGATA g GT	ACATATCATGATATGT	*rpl22*	
AATTACTAATTTCTATTACTATGTTCAATTGAACATAGTAATAGAAATTAGTAATT	AATTACTAATTTCTATTACT t TGTTCAATTGAACATAGTAATAGAAATTAGTAATT	*atpH-atpI*	
TAGTTAGAATTCTAACTA	TAGTT c GAATTCTAACTA	*trnT-UGU-trnL-UAA*	
TATTTTTTCTAGAAAAAATA	TATTTTTTCTAGAA g AAATA	*ycf2*	2 copies in IR
in L. octovalvis and L. peploides	*in L. grandiflora*		
CCCATCAATCATGATTG t TGGG	CCCATCAATCATGATTGATGGG	*psbN-trnD-GUC*	
in L. octovalvis and L. grandiflora	*in L. peploides*		
ATGAAAAAAATCGATTTTTTTCAT	ATGATAAAAATAGATTTTT a TCAT	*trnK-UUU-rps16*	
ATGAAAAAAATCGATTTTTTTCAT- ATGATAAAAATCGATTTTTATCAT	ATGATAAAAATA g ATTTTTATCAT	*trnK-UUU-rps16*	
Unique palidromic repeats			
L. peploides			
TTATATATATATATATATAA		*rpl32-ndhF*	Full deletion in *L. octovalvis*, 6 bases deletion in *L. grandiflora*
L. octovalvis			
ATTGAAATTCGAATTTCAAT		*psbZ-trnG-GCC*	Full deletion in *L. grandiflora and L. peploides*
L. peploides and L. grandiflora			
AAAAAATGGATCCATTTTTT		*trnL-UAG-rpl32*	3 bases deleted and 3 bases mutated in *L. octovalvis*
AATATATTATTATAATAATATATT		*rpl32-ndhF*	Full deletion in *L. octovalvis*
TATATTTATTATTAATTAATAATAAATATA		*rpl32-ndhF*	Full deletion in *L. octovalvis*

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

755

A

	ABySS		MEGAHIT		VELVET		SPAdes	
	not corrected	corrected	not corrected	corrected	not corrected	corrected	not corrected	corrected
Using all contigs								
Genome fraction (%)	86.868	85.279	86.428	85.158	91.927	86.796	84.682	84.483
Duplication ratio	1.047	1.042	1.796	1.041	2.002	1.128	1.042	1
Largest alignment	56 588	41 262	30 904	90 352	3531	17 235	90 399	90 272
Using contigs > 200 nt								
Genome fraction (%)	86.419	85.279	86.377	85.057	76.589	86.181	84.682	84.483
Duplication ratio	1.028	1.042	1.681	1.029	1.177	1.11	1.042	1
Largest alignment	56 588	41 262	30 904	90 352	3531	17 235	90 399	90 272
Using contigs > 500 nt								
Genome fraction (%)	85.564	84.517	83.503	84.774	45.468	79.279	84.682	84.483
Duplication ratio	1.009	1.012	1.041	1.004	1.015	1.054	1.042	1
Largest alignment	56 588	41 262	30 904	90 352	3531	17 235	90 399	90 272
Using contigs > 1000 nt								
Genome fraction (%)	83.701	84.199	81.256	84.545	22.194	66.438	84.563	84.483
Duplication ratio	1	1.002	1.007	1.001	1	1.011	1.026	1
Largest alignment	56 588	41 262	30 904	90 352	3531	17 235	90 399	90 272

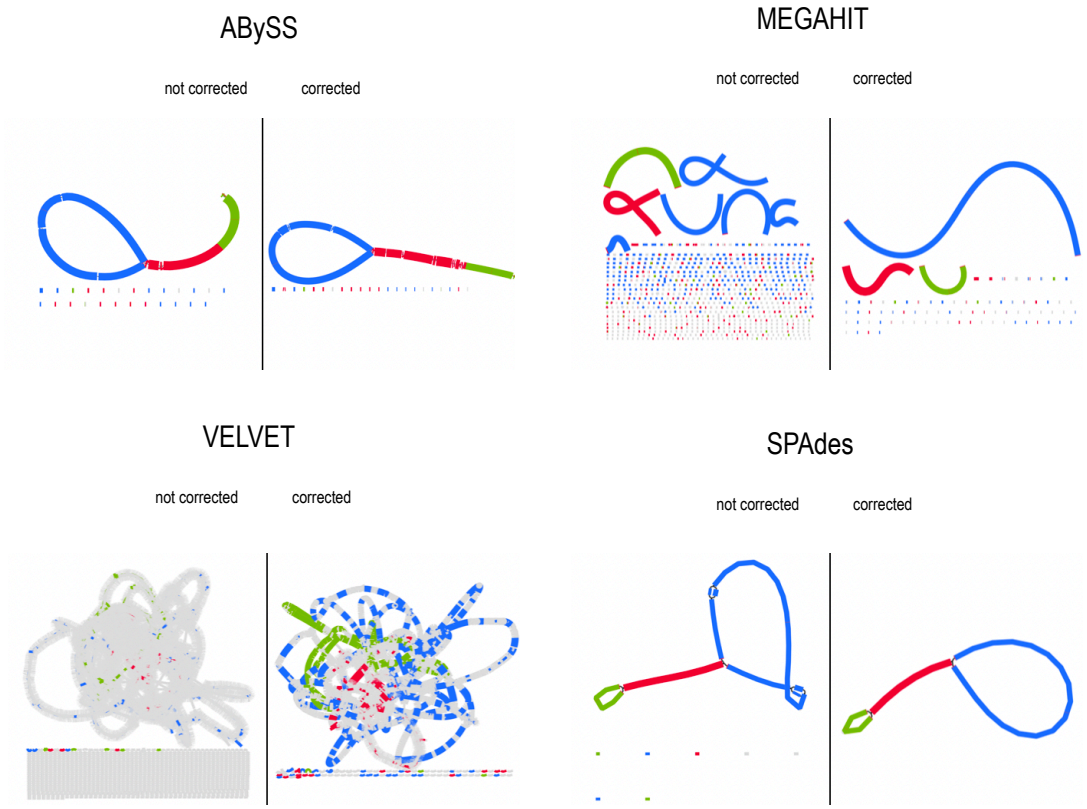
B

	ABySS		MEGAHIT		VELVET		SPAdes	
	not corrected	corrected	not corrected	corrected	not corrected	corrected	not corrected	corrected
Using all contigs								
NGA50	15 215	26 577	19 986	90 352	469	2796	90 399	90 272
LGA50	3	3	3	1	93	9	1	1
Misassemblies								
# misassemblies	0	0	4	0	0	0	0	0
Misassembled contigs length	0	0	1595	0	0	0	0	0
Mismatches								
# mismatches per 100 kbp	109.53	107.19	1036.93	45.24	499.16	229.11	96.57	0
# indels per 100 kbp	12.4	10.58	62.99	16.26	27.92	74.88	19.17	0
# N's per 100 kbp	0	0	0	0	0	6.1	0	0
Using contigs > 500 nt								
NGA50	15 215	26 577	19 986	90 352	-	2796	90 399	90 272
LGA50	3	3	3	1	-	9	1	1
Misassemblies								
# misassemblies	0	0	1	0	0	0	0	0
Misassembled contigs length	0	0	665	0	0	0	0	0
Mismatches								
# mismatches per 100 kbp	62.39	46.17	123.32	8.1	221.33	148.48	96.57	0
# indels per 100 kbp	2.9	2.2	4.33	1.47	28.51	63.74	19.17	0
# N's per 100 kbp	0	0	0	0	0	7.22	0	0
Using contigs > 1000 nt								
NGA50	15 215	26 577	19 986	90 352	-	2796	90 399	90 272
LGA50	3	3	3	1	-	9	1	1
Misassemblies								
# misassemblies	0	0	0	0	0	0	0	0
Misassembled contigs length	0	0	0	0	0	0	0	0
Mismatches								
# mismatches per 100 kbp	0	0	37.51	0.74	64.94	61.6	67.87	0
# indels per 100 kbp	1.5	0.74	0	0.74	25.41	56.93	19.49	0
# N's per 100 kbp	0	0	0	0	0	9.25	0	0

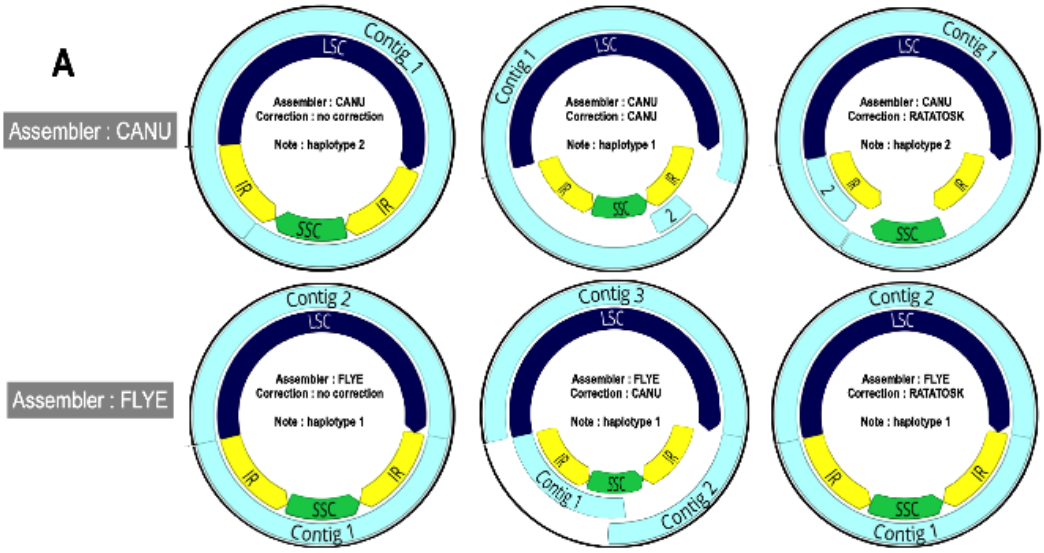
756

757 **Supp. Figure 1:** QUAST evaluation of performance of the four assembly tools (using corrected or
758 uncorrected SRs). **A:** Comparison of plastome fraction, duplication rate and size of the largest alignment
759 obtained. **B:** Comparison of classic metrics (NGA50 and LGA50), number of errors (misassemblies and
760 mismatches) produced.

761



Supp. Figure 2: BANDAGE visualization of the *L. grandiflora* plastome assembly graphs on corrected or uncorrected SRs. Contigs are colored according to their BLAST match to the LSC (blue), SSC (green), and IR (red) segments



B

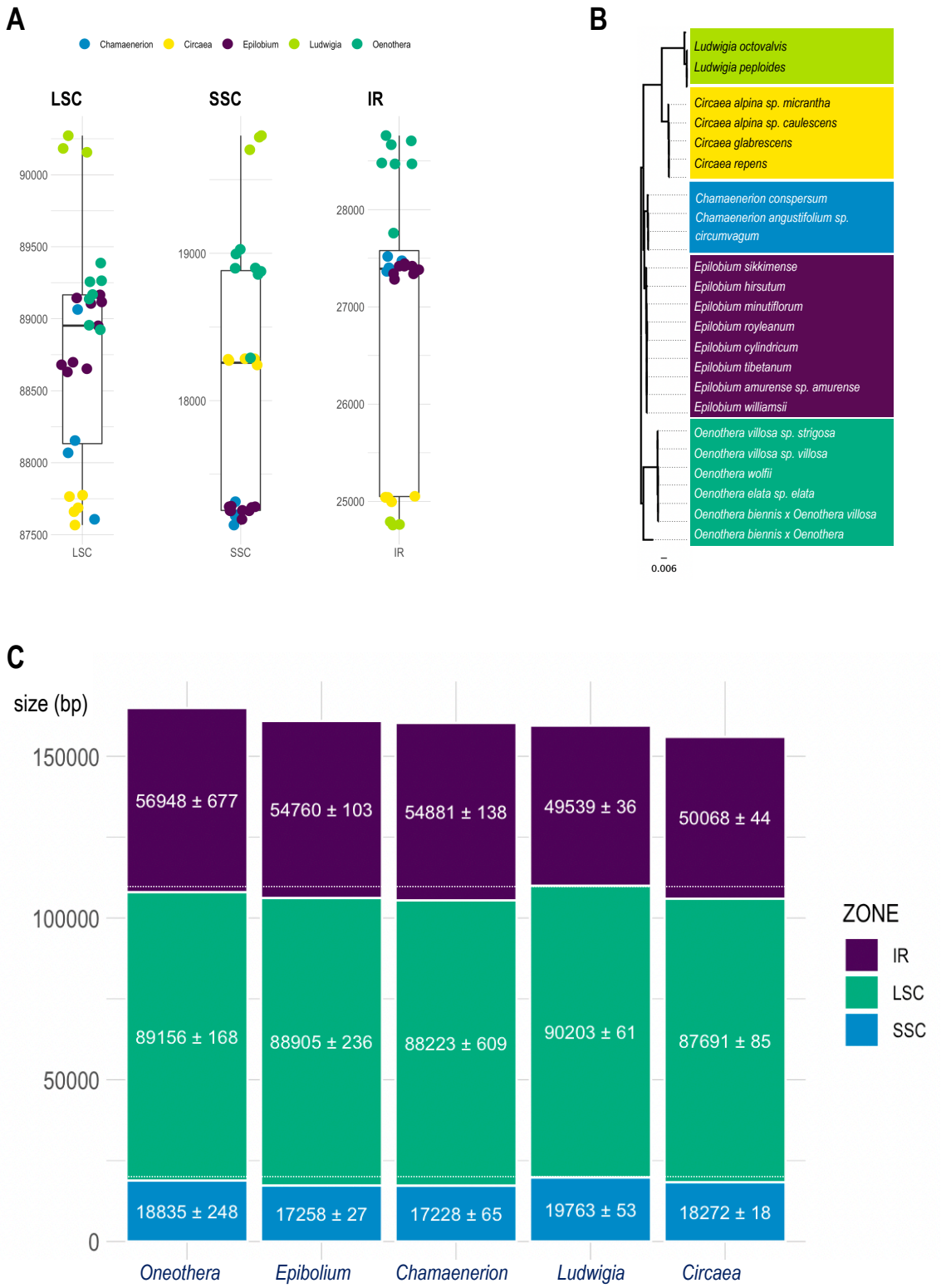
Assembler : CANU

	LSC	IR	SSC
No correction	99.793 (187 nt)	99.960 (30 nt)	99.778 (27 nt)
Canu-correction	99.811 (171 nt)	99.899 (25 nt)	99.828 (34 nt)
Ratatosk-correction	99.952 (43 nt)	99.879 (10 nt)	99.864 (27 nt)

Assembler : FLYE

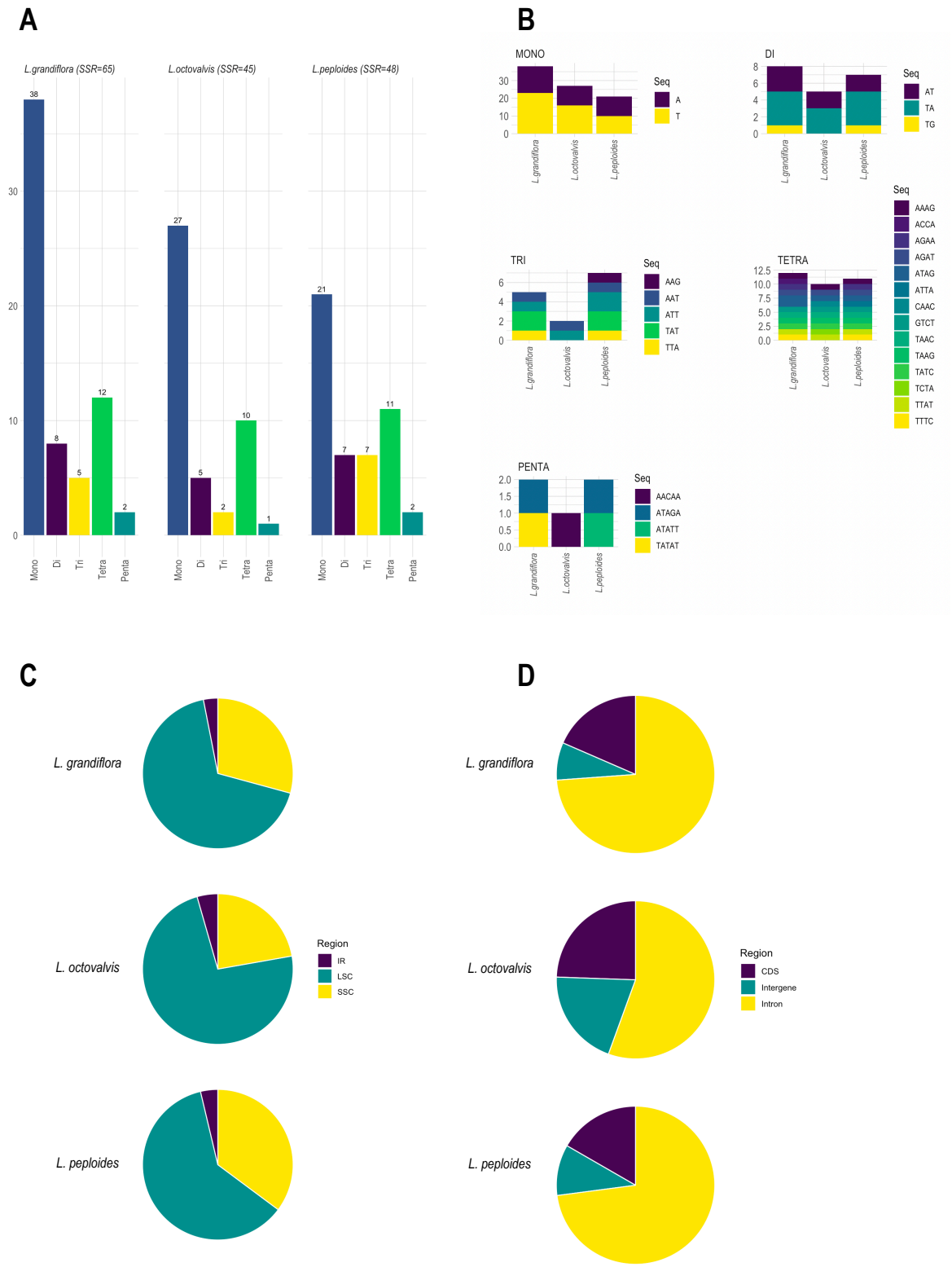
	LSC	IR	SSC
No correction	99.798 (182 nt)	99.903 (24 nt)	99.783 (50 nt)
Canu-correction	99.831 (153 nt)	99.907 (23 nt)	99.843 (31 nt)
Ratatosk-correction	99.975(23 nt)	99.988 (3 nt)	99.995 (3 nt)

Supp. Figure 3: Graphs representing the assemblies of *L. grandiflora* long reads. **A:** Contigs are represented in light blue and the three segments (LSC, SSC and IR) in dark blue, green and yellow, respectively. B: Comparative effectiveness of CANU and RATATOSK correctors.

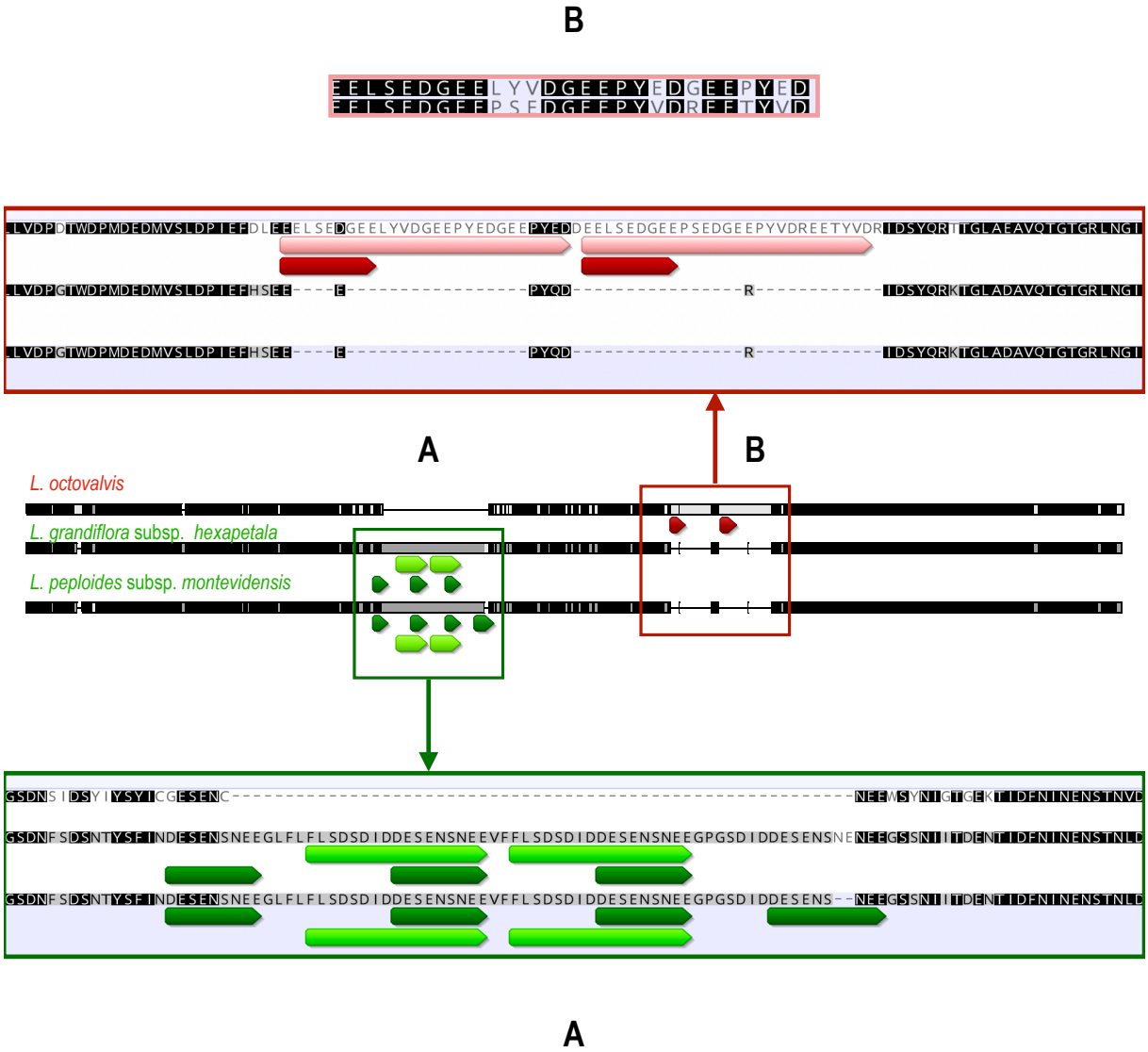


Supp. Figure 4: Comparison of LSC, SSC and IR sizes in the Onagraceae. **A:** Comparison of the sizes of LSC, SSC and IR segments in the Onograceae family (*Chamaenerion* in blue, *Circaea* in yellow, *Epibolium* in dark purple, *Ludwigia* in light green and *Oenothera* in dark green). **B:** Maximum likelihood tree made using RAxML (model GTR-GAMMA, algorithm Rapid Hill-climbing) on multiple sequences alignment of

779 Onograceae plastomes made using MAFFT. **C:** Average size of the different chloroplast segments (LSC,
780 SSC and IR) for the 5 genres of Onograceae. IR size corresponds to the sum of the two copies.
781

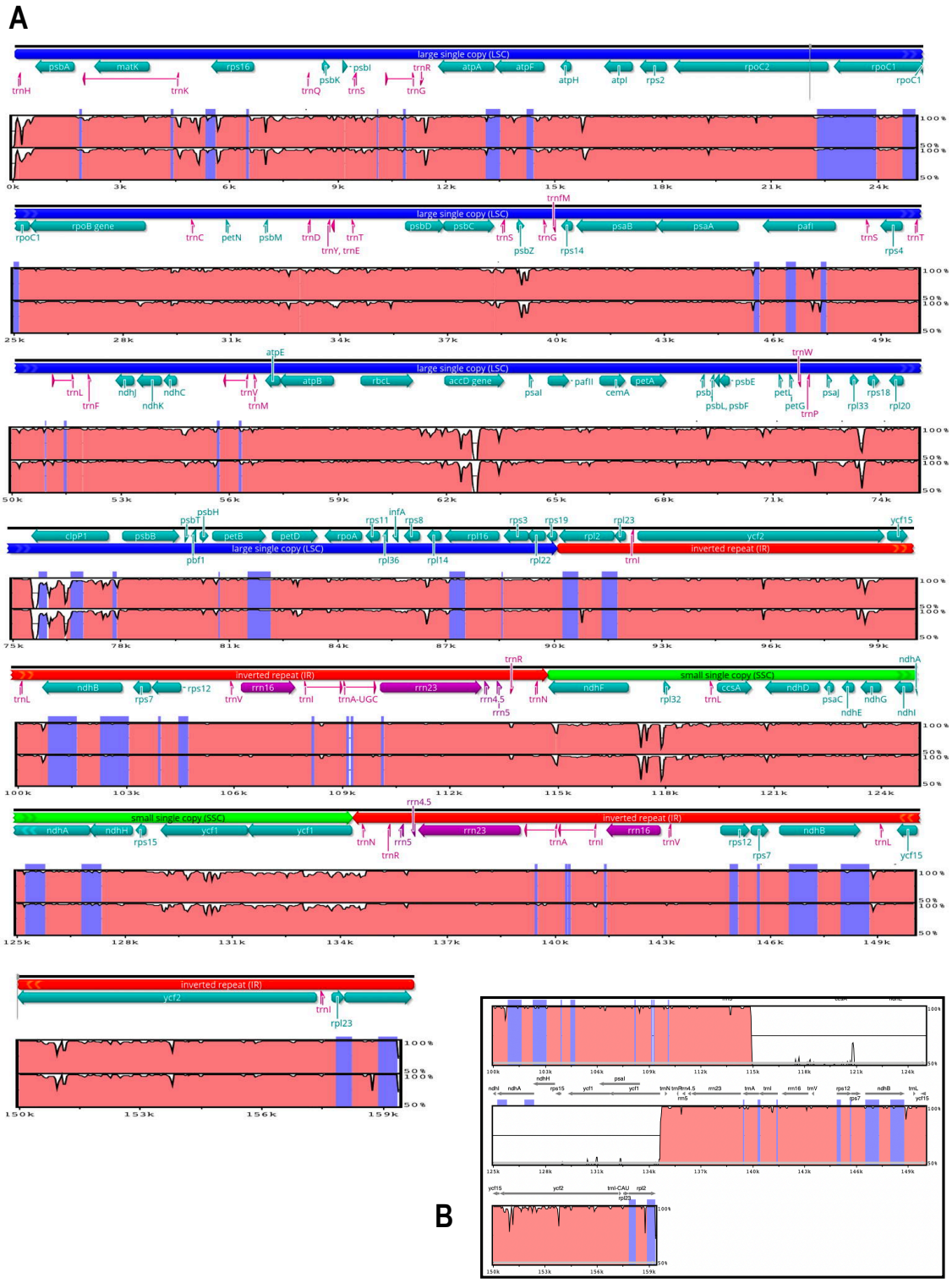


Supp. Figure 5: Comparative analysis of Simple-Sequence Repeats (SSRs) in *Ludwigia* chloroplast genomes. A: SSR numbers detected in the three species, by repeat class types (mono, di-, tri-, tetra and pentanucleotides). **B:** Frequency of SSR motifs by repeat class types. **C:** Frequency of SSRs in LSC, SSC and IR regions. **D:** Repartition of SSRs in intergenic, protein-coding and intronic regions.



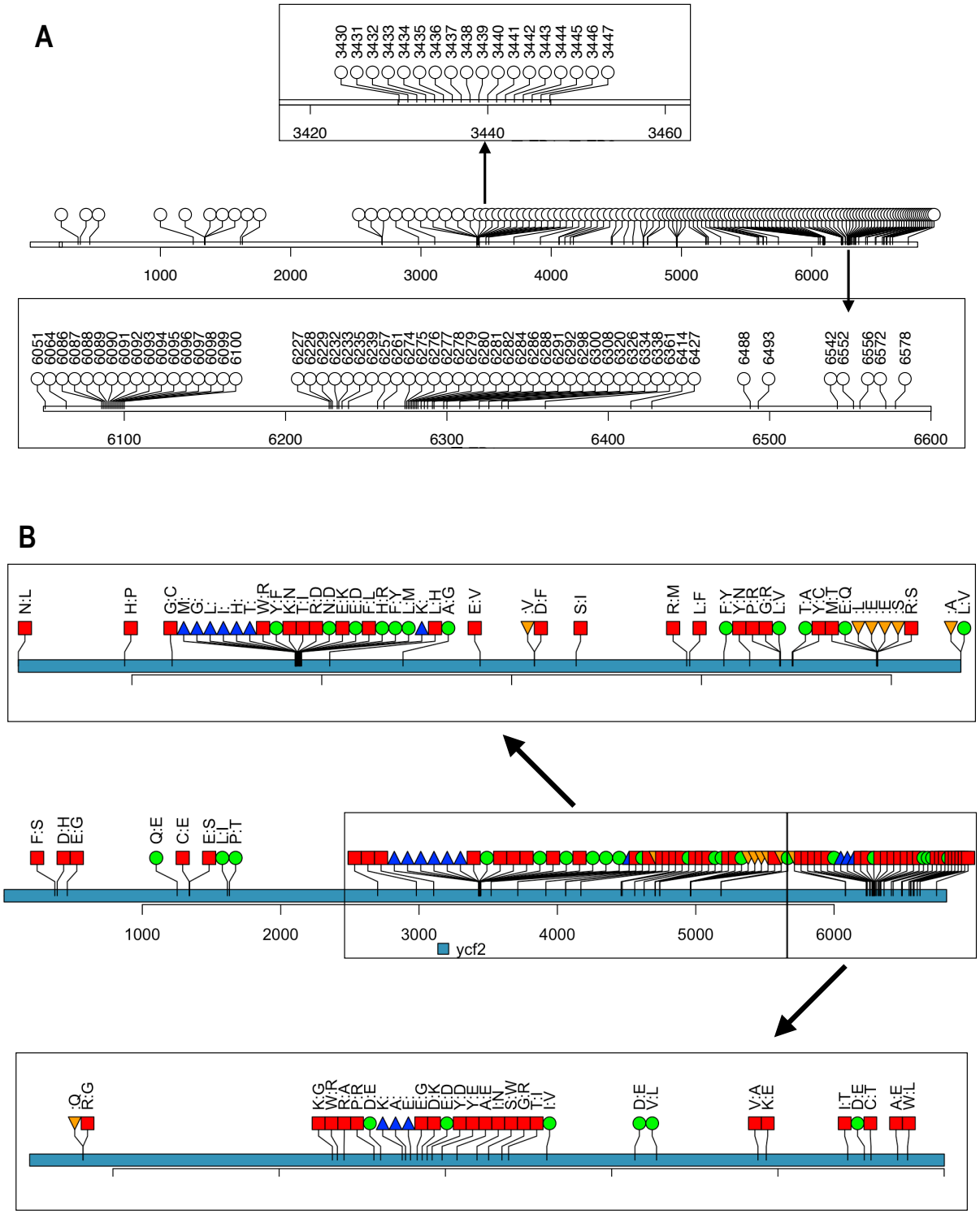
Supp. Figure 6: Diagram showing the position of tandem repeats in the *accD* gene. *L. octovalis* (in red) and *L. peploides* and *L. grandiflora* (in green). We also observe the consequences of these repetitions on the insertion of amino acids, also repeated.

796



Supp. Figure 7: Comparison of the three *Ludwigia* plastomes using mVISTA, with the *L. octovalvis* as a reference. **A:** The y-axis represents the identity percentage (between 50 and 100%). The arrows show the genes (in green: proteins genes, in purple: rRNAs and in fuchsia: tRNAs). Blue blocks indicate exonic regions. LCS, IR and SSC regions are also distinguished (in dark blue, red and green, respectively). The second line corresponds to *L. grandiflora* haplotype 2 (For this haplotype, SSC segment is oriented like *L.*

803 *octovalvis*) and the third line corresponds to *L. peploides* for which the SSC region has been artificially
804 oriented in the same way as the two other plastomes to allow comparison. **B:** Small box showing a part of
805 the alignment and presenting the consequences if we do not artificially orient the SSC segments in the same
806 direction for the analysis.
807



Supp. Figure 8: Lollipop diagram allowing the visualization of SNPs and their translational effects on the *ycf2*. **A:** localization of the 256 single nucleotide polymorphisms (SNP) observed by comparing *L. grandiflora-L. peploides* with *L. octovalvis.* Two regions particularly dense in SNPs (between 3420 and 3460 and between 6100 and 6600) have been zoomed into to allow better reading. **B:** Effect of these SNPs on the translated sequence of *L. octovalvis*, compared to Ycf2 of the other two species: non conservative mutation:

red square; conservative mutation: circle green; deletion: triangle_point_up blue and insertion: triangle_point_down, orange. As for A, two regions were zoomed into in order to distinguish each mutation.

REFERENES

[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.