Comparative Chloroplast Genomics of C_3, Kranz type C_4 and Single Cell C_4 photosynthetic members of Chenopodiaceae

Richard M. Sharpe[1*], Bruce Williamson-Benavides[1,2*], Gerald E. Edwards[2,3], and Amit Dhingra[1,2§]

[1]Department of Horticulture, Washington State University, Pullman, WA – 99164

[2]Molecular Plants Sciences, Washington State University, Pullman, WA -99164

[3]School of Biological Sciences, Washington State University, Pullman, WA – 99164

*These authors contributed equally to this work

§Corresponding Author: adhingra@wsu.edu

ABSTRACT

Background

Chloroplast genome information is critical to understanding taxonomic relationships in the plant kingdom. During the evolutionary process, plants have developed different photosynthetic strategies that are accompanied by complementary biochemical and anatomical features. Members of family Chenopodiaceae have species with C_3 photosynthesis and variations of C_4 photosynthesis in which photorespiration is reduced by concentrating CO_2 around Rubisco through dual coordinated functioning of dimorphic chloroplasts. Among dicots, the family has a large number of C_4 species, and greatest structural and biochemical diversity in forms of C_4 including the canonical dual-cell Kranz anatomy, and the recently identified single-cell C_4 with the presence of dimorphic chloroplasts separated by a vacuole. This is the first comparative analysis of chloroplast genomes in species representative of photosynthetic types in the family.

Results

High quality and complete chloroplast genomes of eight species representing C_3, Kranz type C_4, and single-cell C_4 (SSC_4) photosynthesis were obtained using high throughput sequencing complemented with Sanger sequencing of selected loci. Six of the eight chloroplast genome sequences are new, while two represent corrected versions of previously published chloroplast genomes. Comparative genomic analysis with previously sequenced plastid genomes revealed a similar genome organization, gene order, and content with a few revisions. High-quality complete chloroplast genome sequences resulted in correcting the orientation of the LSC region of the published *Bienertia sinuspersici* chloroplast genome, identification of stop codons in the rpl23 gene in *B. sinuspersici* and *B. cycloptera*, and identifying an instance of IR expansion in the *Haloxylon ammodendron* inverted repeat sequence. The rare observation of a mitochondria-to-chloroplast inter-organellar gene transfer event was identified in family Chenopodiaceae.

Conclusions

This study reports complete chloroplast genomes from seven Chenopodiaceae and one Amaranthaceae species. The depth of coverage obtained using high-throughput sequencing complemented with targeted resequencing of certain loci enabled superior resolution of the border junctions, directionality, and repeat region sequences. Therefore, the use of high throughput and Sanger sequencing, in a hybrid method, reaffirms to be rapid, efficient, and reliable for chloroplast genome sequencing.

Keywords: Chenopodiaceae, chloroplast genome, Single Cell C4 photosynthesis, genomics, Amaranthaceae

Background

Plastids convert light energy into chemical energy and are an essential site for the biosynthesis of pigments, lipids, several amino acids, and vitamins [1,2]. Comparative genomics studies have facilitated the understanding of chloroplast genome organization and phylogenetic relationships [3–5]. Additionally, the availability of chloroplast genome sequences can be useful for constructing transformation vectors to enable chloroplast transformation via homologous recombination [6,7].

Higher plant chloroplast genomes possess a characteristic organization comprising a Large Single Copy (LSC), a Small Single Copy (SSC), and two Inverted Repeat (IRa and IRb) regions, with only a few exceptions, e.g., in *Pisum sativum* and some other legumes [8–10]. Several methods have been used to sequence chloroplast genomes in plants, including primer walking [11–14] and high-throughput sequencing (HTS) [15]. HTS, both with isolated chloroplast DNA [16–18] and total cellular DNA [19–21] has been employed to generate physical maps of the chloroplast genome. However, the junctions of LSC/IRa, IRa/SSC, SSC/IRb, and IRb/LSC need to be resolved using additional experimentation [22]. Genome sequencing and subsequent assembly of the chloroplast genome can be challenging due to variable IR borders; the presence of chloroplast genome sequences in the nuclear genome; sequence homology between chloroplast and mitochondrial genes, such as the NAD(P)H and NADH dehydrogenase genes; as well as the NAD(P)H genes being distributed throughout the chloroplast genome [3,23–28].

Chloroplasts, the green plastids in plants, are the site for photosynthesis where Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), captures CO_2 with the synthesis of 3-phosphoglyceric acid (3PGA) in the Calvin-Benson cycle, leading to the synthesis of carbohydrates, and cellular constituents. Three major types of oxygenic photosynthesis are known to date: C_3, C_4, and Crassulacean acid metabolism (CAM). In C_3 plants, Rubisco directly fixes atmospheric CO_2 introducing carbon into the Calvin-Benson cycle. In C_4 and CAM photosynthesis, CO_2 is first captured by phosphoenolpyruvate carboxylase (PEPC) with the synthesis of 4-carbon organic acids, which are sequestered in a spatial manner in C_4 plants and a temporal manner in CAM plants. Decarboxylation of the 4-carbon organic acid generates a CO_2-rich environment around Rubisco [29]. This mechanism suppresses the oxygenation reaction by Rubisco and the subsequent energetically-wasteful photorespiratory pathway. C_4 plants function with spatial separation of two types of chloroplasts; one type supports the fixation of atmospheric CO_2 by PEPC and synthesis of C_4 acids.

In contrast, the other type utilizes the CO_2 generated from the decarboxylation of C_4 acids in the Calvin Benson cycle. In Kranz type C_4 plants, mesophyll chloroplasts support fixation of atmospheric CO_2 by PEPC, while bundle sheath chloroplasts utilize CO_2 generated by decarboxylation of C_4 acids. The unique single-cell C_4 (SCC_4) plants perform C_4 photosynthesis within individual chlorenchyma cells with spatial separation of two types of chloroplasts. One type supports the capture of atmospheric CO_2 by PEPC, and the other assimilates the CO_2 generated by decarboxylation of C_4 acids in the Benson-Calvin cycle [30–32].

Among dicot families, the Chenopodiaceae and Amaranthaceae families have by far the largest number (~800) of C_4 species, with up to 15 distinct lineages [33]. Although they are currently recognized as separate families in a clade, they are known to be closely related [34]. Chenopodiaceae species are acclimated to diverse ecosystems from xeric to more temperate salt marshes, including highly saline soils, while Amaranthus species predominantly occur in tropical and subtropical regions. The Chenopodiaceae family is very diverse, with six structural forms of Kranz anatomy present among its members [35]. Furthermore, it is the only family known to have SCC_4 species [34]. Phylogenetic analyses have identified independent origins of C_4 photosynthesis. In particular, the results allude to the unique independent origins of C_4 in subfamily Suaedoideae, including Kranz C_4 anatomy in *Suaeda* species and two independent origins of the SCC_4 system in *Bienertia* and *Suaeda* [33,36–39]. In general, the causation of these independent events is hypothesized to be a result of the harsh environments induced by global climate change and periodic reductions in CO_2 content over the past 35 million years [40,41].

In this study, complete chloroplast genome sequences for seven Chenopodiaceae species and one Amaranthaceae species were generated using whole leaf tissue genomic DNA (gDNA) via HTS complemented with Sanger sequencing of targeted loci. The species analyzed were: *Bassia muricata* (C_4-Kochioid anatomy, tribe Camphorosmoideae)*, Haloxylon ammodendron* (C_4-Salsoloid anatomy, tribe Salsoleae), *Bienertia cycloptera* (C_4: SCC_4-tribe Suaedeae)*, Bienertia sinuspersici* (C_4: SCC_4-tribe Suaedeae)*, Suaeda aralocaspica* (SCC_4-tribe Suaedeae), *Suaeda eltonica* (C_4-Schoberioid type anatomy, tribe Suaedeae)*, and Suaeda maritima* (C_3-tribe Suaedeae). The chloroplast genome from *Amaranthus retroflexus* (C_4-Atriplicoid type anatomy, family Amaranthaceae, tribe Amarantheae), was also sequenced and used for comparative analysis.

Results and Discussion

Genome sequencing and assembly

A summary of the sequencing data obtained from Illumina sequencing and assembly of *A. retroflexus, B. muricata, B. cycloptera*, *B. sinuspersici, H. ammodendron, S. aralocaspica, S. eltonica,* and *S. maritima* chloroplast genomes is presented in Table 1. Three large contigs with overlapping 5' and 3' regions were generated during genome assembly for *A. retroflexus, B. muricata, B. cycloptera, B. sinuspersici, H. ammodendron, S. aralocaspica,* and *S. maritima.* These three contigs were identified as LSC, SSC, and IR via BLAST homology alignment [42] and DOGMA gene identity prediction [43]. The overlapping regions were present at all four possible junctions when the IR region was reverse complemented (LSC-IR, IR-SSC, SSC-IR, and IR-LSC). These overlapping areas ranged from 19 to 51 nucleotides (nt) (illustrated in Fig. S1 with *B. cycloptera*). The directionality of the LSC, SSC and IR, and all overlapping aligned junctions were validated via Sanger sequencing of both strands of the amplicons generated from these regions (Table S1; Fig. S1). For *S. eltonica*, the LSC-IRa and IRb-LSC overlapping regions were 23 nt long and were validated with Sanger sequencing (Table S1). The IRa-SSC and SSC-IRb sections were both missing a 1,475 nt section in the IRa and IRb borders. The 300 nt sequence contiguous to the 1,475 nt section had a low GC content of 19%. A possible cause of the shortened contig flanking the IR-1475 area may be due to the low GC content value, which could impact the accuracy of the HTS genome assembly [44]. The 1,475 nt section was sequenced by primer walking and Sanger sequencing (Table S1). The GC content in the 1,475 nt region and IR was 31.3% and 42.1%, respectively.

The average base depth of coverage for the eight assembled chloroplast genomes ranged from 1553 to 5998-fold. For accurate assembly, a minimum of 30-40x sequence coverage is recommended [45–47]. In this study, the only areas with less than 40x average coverage were identified in the last 1 to 3 nucleotides of the IRb sequence for each of the eight genomes. This is expected due to the assembler algorithm parameters. The end of the IRb and the beginning of the LSC were concatenated, and these sections were remapped. Remapped coverage results were reported to be above 40x for the IRb ends and surrounding areas. The eight assembled genomes (0.8/0.9 for the read length fraction/similarity fraction mapping) were also compared with a more stringent remapping of the reads to the contigs of 0.99/0.99 length fraction/similarity fraction. Analyses with both levels of stringency show almost identical assembly minimum-coverage and average-coverage for the eight species sequenced in this study (Fig. S2).

147 Overall, the assembly and subsequent Sanger sequencing-based validation generated high
148 quality and complete chloroplast genomes, with all possessing a quadripartite structure as reported in
149 other land plant species.

150 **Size, organization and gene content of the chloroplast genomes**

151 The size of the chloroplast genomes from the eight species ranged from 146,634 to 161,251 nt
152 (Table 2). As expected, each chloroplast genome included a pair of inverted repeat regions, IRa and IRb,
153 separated by an SSC and an LSC region (Table 2 and Fig. S3). With one exception, the size of the IRs
154 ranged from 23,461 to 25,213 nt. (Table 2). The *H. ammodendron* inverted repeat sequence presented
155 an instance of IR length expansion (29,061 nt) compared to the other seven species. The GC content was
156 similar among the eight species, and for all the plastomes, LSC, SSC, and IRs, it ranged from 36.4-36.6,
157 34.1-34.6, 29.1-30.2, and 42.1-43.0%, respectively (Table 2). All chloroplast genomes contained a similar
158 number of protein-coding, ribosomal, and tRNA genes. The number of genes and tRNAs ranged from
159 113 to 116 and 27-29, respectively, in the eight genomes (Table 3 and Fig. S3). For seven out of eight
160 species, 60.1-61.9% of the chloroplast sequence consisted of the coding region, which included 52.7-
161 54.3% of protein-coding genes and 7.4-7.9% of RNA genes. The *S. eltonica* chloroplast genome was
162 composed of 56.8% coding region, including 48.9% of protein-coding genes and 7.9% of RNA genes. This
163 difference between *S. eltonica* and the rest of chloroplast genomes is possibly due to the higher repeat
164 content in intergenic sequences of the *S. eltonica* chloroplast genome (Table 4 and Fig. 1).

165 Gene order and content were largely conserved among the eight chloroplast genomes in this
166 study. However, some structural rearrangements, gene losses, and IR expansions were identified. The
167 genes ycf15, ycf68, and rpl23 were identified as pseudogenes due to the presence of internal stop
168 codons. The ycf15 and ycf68 genes are quite commonly classified as pseudogenes in angiosperms
169 [23,48]. The rpl23 is also classified as a pseudogene in some species such as the *Fagopyrum* spp.,
170 buckwheat, and spinach as well as *Suaeda* and *Haloxylon* species [22,23,49,50]. In *S. eltonica,* rpl23 was
171 not predicted to be in the chloroplast genome by GeSeq, but it was identified as a pseudogene via the
172 BLAST sequence analysis [42]. No stop codons were identified in the rpl23 of a previously published *B.*
173 *sinuspersici* chloroplast genome [51]. In this study, four stop codons were identified at the same
174 locations for *B. sinuspersici* and its close relative *B. cycloptera*.

175 At least one complete copy of the ycf1 gene was identified in the eight chloroplast genomes
176 (total length of 5.3-5.6 Kb). In seven out of the eight chloroplast genomes, a duplicated ycf1 pseudogene

(1,000-1,300 nt) was found at the IRa-SSC boundary. This is a common feature found in other species [23,52]. In the case of *H. ammodendron*, there is a complete duplication of the ycf1 gene. Therefore the *H. ammodendron* chloroplast genome has two full copies in the IR-SSC borders. The complete duplication of the ycf1 gene in *H. ammodendron* leads to the previously mentioned IR expansion (Fig. S3). This phenomenon has also been observed in *Amphilophium, Adenocalymma*, *Anemopaegma,* and *Fagopyrum* species; these species possess an expanded IR region and two full-length copies of ycf1 gene [23,53,54]. The IRs for the other seven species are variable in length. In *A retroflexus, B. muricata, B. cycloptera, B. sinuspersici, S. aralocaspica,* and *S. maritima*, the IR includes the duplicated ycf1 pseudogene (1-1.3 kb) (Fig. S3). A small segment of the ycf1 gene is also duplicated in *V. vinifera, S. oleracea* and *B. vulgaris.* In *S. eltonica,* the IR has expanded to include the trnH-GTG and a fragment of the psbA gene (Fig. S3). The biological significance of this duplication remains unknown.

Annotation of the ycf15 gene with the Dual Organellar Genome Annotator (DOGMA) [43] shows variability in terms of its physical location. In *A. retroflexus, B. vulgaris,* and *S. eltonica,* the ycf15 is located between the rps12 and trnV-GAC. In *B. cycloptera, B. muricata, B. sinuspersici, H. ammodendron, S. aralocaspica,* and *S. maritima,* the ycf15 is located between ycf2 and trnL-CAA. The ycf15, as well as other genes, such as the ycF2, psbA, clpP, and matK, have been reported to have a variable physical location in different plants [55–58].

The genes ycf3, clpP, rpoc1, and rpl2 have been found to have a variable number of introns among and within some taxonomic groups [23]. The gain or loss of introns in these genes has occurred independently in several lineages of flowering plants [23,59]. However, no differences were found in the number of introns among the eight species; the ycf3, clpP, rpoc1, and rpl2 contain 2, 2, 1, and 0 introns, respectively.

The orientation of the SSC region in *A. retroflexus and B. muricata* differs from the orientation of the SSC in *B. cycloptera, B. sinuspersici, H. ammodendron, S. aralocaspica*, *S. maritima* and *S. eltonica* (Fig. S3). The SSC orientation has been shown to exist in the two different states within individual plants [60–63]. Therefore, SSC variation observed among taxa in this study is likely due to alternative states of the SSC region within individual plants. Although there was some variation in the SSC orientation, the number and content of genes were the same among the eight species. The only exception is the presence of a trnU-TCA in the SSC of *H. ammodendron*.

Repeat structures

Seven out of the eight chloroplast genomes had 45 to 58 repeats, which ranged in length from 30 to 73 nt per repeat (Fig 1). The majority of these repeats were shown to be between 30 and 40 nt in length. In the *S. eltonica* chloroplast genome, repeat analysis with REPuter [64] found a total of 174 repeats, which ranged from 30 to 145 nt in length (Fig 1). The number of repeats was similarly distributed among species for repeats found in intergenic regions and intron/exons (Table 4). An exception was *S. eltonica,* in which a majority (80%) of repeats were located in the intergenic regions. Four species possessed reverse repeats; *S. maritima* and *S. aralocaspica* had one, *B. muricata* had two, and *S. eltonica* had four.

The presence of repeats varied for the genes ycf1, ycf2, ycf3, and psaA. Repeats were present in the gene ycf1 except for *A. retroflexus*, *S. aralocaspica,* and *S. maritima*. All chloroplast genomes possessed repeats in the ycf2 gene except for *H. ammodendron*. Repeats in the introns of the ycf3 gene were only present in the *A. retroflexus, B. cycloptera*, and *B. sinuspersici*. All species presented at least one repeat in the psaA gene, and *H. ammodendron* presented the highest number with six repeats.

Comparison of *Amaranthus retroflexus* chloroplast genome with previously sequenced Amaranthus spp. chloroplast genomes

A. retroflexus, commonly known as pigweed, is used as a vegetable for human consumption as well as for fodder. It is the most widely distributed and damaging *Amaranthus* weed in the US and the world [65]. The availability of the *A. retroflexus* chloroplast genome provides an important tool for accurately monitoring the spread of this species and identifying possible hybridizations. Microsatellites were previously identified for *Amaranthus* spp. [66]. Six out of the nine polymorphic microsatellites were shown to be polymorphic between A. *hypochondriacus* and *A. retroflexus* (Table 5). Most of these microsatellites were located in the LSC regions and represented A or T mononucleotide repeats. SSRs can serve as molecular markers for future molecular breeding for *Amaranthus* spp. which are considered as emerging crops [66]. The chloroplast genomes of four Amaranthus spp; A. *hypochondriacus, A. cruentus, A. caudatus,* and *A. hybridus*, have been reported previously [66]. The *A. hypochondriacus* genome (GenBank accession KX279888.1) is 150,725 nt, and the quadripartite regions of LSC, SSC, and 2 IRs consist of 83,873, 17,941 and 24,352 nts, respectively. These sizes are very similar to the lengths of the *A. retroflexus* chloroplast genome reported in this study (Table 2). BLAST analysis showed a 99% sequence similarity between the chloroplast genomes of *A. hypochondriacus* and *A. retroflexus*.

Comparative analysis of the *B. sinuspersici* chloroplast genomes

Kim et al. (2016), and Caburatan et al., (2018) previously reported the chloroplast genome of *B. sinuspersici* (GenBank accession no. KU726550). Compared to our results with *B. sinuspersici* (Table 2), the size of their genome (153,472 nt) is 138 nt larger; the LSC and SSC in their study are 84,560 nt and 19,016 nt in size, respectively which is 70 nt larger than in our study (Table 2). The IR was reported to be 24,948 nt in length, versus 24,949 nt length in this study. The increase in length in the published *B. sinuspersici* chloroplast genome [51] is predominantly located at the LSC-IRa and SSC-IRb junctions, which has a repeat of 72 and 13 nts, respectively. The two repeats are separated by spacer sequences of 1nt in the LSC-IRa junction and 48 nt in the SSC-IR junction. The 72 and 13 nt sequences were present just once in the *B. sinuspersici* chloroplast genome presented in the current study. The presence of a single occurrence of the 72 and 13 nt sequence in the genome was validated by Sanger sequencing of loci in question for both IRb-LSC and LSC-IRa loci (Table S1). Further comparison of the two *B. sinuspersici* genomes identified 18 SNPs and nine indels. In the published *B. sinuspersici* chloroplast genome, the LSC is inverted with respect to the rest of the sequence (IRa+SSC+IRb). In our study, the orientation of the LSC was validated using Sanger sequencing of PCR amplicons spanning the junctions IRb-LSC and LSC-IRa (Table S1). As described above, there were also differences in the presence of stop codons in the rpl23 gene. In the previous study [67], a total of 110 unique genes were reported; a total of a total of 114 genes were identified in the current study (Fig. S3).

Differences between the previously reported chloroplast genome of *B. sinuspersici* compared to the current study likely stems from how the Celera assembler algorithm and the CLC algorithm process the read data. Each of these algorithms has its inherent pros and cons [68]. The assembly parameters for the previous *B. sinuspersici* chloroplast genome were not reported. Also, the chloroplast genome loci that were found to be different within the two previous versions [51,67] were not resequenced. The chloroplast genome of *B. sinuspersici* presented in this study showed a minimum, maximum, and average coverage of 37, 23,533, 3,204.28 nt. Furthermore, areas of ambiguity were validated via Sanger sequencing of PCR amplicons generated from selected loci. The combination of the assembly strategy utilized and resequencing of loci resulted in the generation of an improved version of the *B. sinuspersici* chloroplast genome.

Analysis of the two closest SCC_4 related species, *B. cycloptera* and *B. sinuspersici*, chloroplast genomes showed a 99.70% sequence similarity between both sequences. *B. cycloptera* and *B. sinuspersici* chloroplast genomes differed in overall length by seven nt. *B. sinuspersici* IR, and SSC regions were larger than the *B. cycloptera* by 44 nt and *B. cycloptera*'s LSC region was larger by 51 nt. The

268 difference in size was due to changes in the intergenic region, length, and the number of repeat regions.
269 The number of genes with introns and repeats was the same between the two species. *B. cycloptera* had
270 two larger repeats, one between 40 to 44nt and the second greater than 45nt. *B. sinuspersici* had one
271 smaller repeat of 30 to 34nt. Both species had the same number and identity of protein-coding, tRNA,
272 and rRNA genes.

273 **Comparative analysis of *Haloxylon ammodendron* chloroplast genomes: a case of transfer of**
274 **mitochondrial DNA to the plastid genome**

275 The chloroplast genome of *H. ammodendron* was published recently (GenBank accession no. KF534478)
276 [69]. The size of the chloroplast genome was reported to be 151,570 nt, with an LSC of 84,214 nt, SSC of
277 19,014 nt, and two IRs of 24,171 nt [69]. In our study, the genome assembled to a size of 161,251 nt,
278 which is 9,681 nts larger. BLAST alignment of the two genomes indicated that the additional 9,681 nts
279 were derived from the expansion of the IR, which is 4,868 nt in size. The IRs of the *H. ammodendron*
280 chloroplast genome in our study were 29,061 nt long. This represents an expansion of the IR that is also
281 observed in *S. eltonica* (Table 2). Expansion and gene duplication are a common phenomenon in the IR
282 regions of chloroplast genomes [70,71]. In grasses, the junctions between the IR and SSC regions are
283 highly variable with the ends of genes ndhF, rps19, and ndhH repeatedly migrating into and out of the
284 adjacent IR regions [72]. BLAST alignment between the two genomes revealed that the first 115 nt
285 showed 78% homology with chloroplast sequences of *H. persicum* and *H. ammodendron* present in the
286 IRs of the published genomes [69]. The following region of 671 nt did not show any significant similarity,
287 and the last 4,028 nt showed homology to mitochondrial genome sequences. The highest significant hit
288 (94%; E value = 0.0) for this 4,028 nt section resembled *Beta vulgaris* and *Spinacia oleraceae*.
289 Interestingly, annotation identified the mitochondrial gene Cytochrome b (cob) in this 4,814 nt section,
290 although the plastid copy had a nonsense mutation that resulted in a premature stop codon.

291 Evidence showing the transfer of mitochondrial DNA (mtDNA) or nuclear DNA (nucDNA) to the
292 plastid genome in plants had been lacking until recently. A few recent reports indicate that the plastid
293 genomes of carrot [73], milkweed [74], and bamboo [72] show evidence of gene transfer from
294 mitochondria to the plastid. *Daucus carota* has a 1.5 kb region of mitochondrial origin located in the
295 rps12-trnV intergenic space of the chloroplast genome. Only *Daucus* species and the close relative cumin
296 (*Cuminum cyminum*) show the mitochondrion-to-chloroplast gene transfer [73]. It was concluded that a
297 mitochondria-located DNA segment present in the ancestor of the Apiaceae subsequently moved to the
298 plastid genome in the common ancestor of *Daucus* and *Cuminum*. *Asclepias syriaca,* the common

milkweed, has a 2.4 kb mtDNA-like insert in the chloroplast genome. The mtDNA-like insert contains an intact exon of the mitochondrial ribosomal protein (rpl2) as well as a noncoding region [74]. There was a 92% sequence identity between the mitochondrial and plastid version of rpl2 in *A. syriaca,* whereas the plastid copy had a nonsense mutation resulting in a premature stop codon. Similarly, the IR region in three herbaceous bamboo species of the *Pariana* genus had a 2.7 kb insertion [72]. The insertion was located in the trnI-CAU-trnL-CAA intergenic spacer region. Potential variations of this insertion in another *Pariana* species and species from the sister genus *Eremitis* were also reported. These studies suggest that the transferred sequence may have originated as a single event in a common ancestor; however, the inserted sequence evolved rapidly [72].

In our study, the inserted section in *H. ammodendron* had average coverage of 1,320 x reported from the stringent 0.99-0.99 length fraction/similarity mapped to the assembly. The coverage corresponded well to the average coverage of 1,269 X for other regions. Five kb regions flanking the 4.8kb section had a similar coverage of 929, and 1,066 reads. The Illumina reads from *H. ammodendron* (0.99-0.99 99 length fraction/similarity fraction) were mapped to three randomly selected intronless mitochondrial genes identified from the *H. ammodendron* assembly [72]. The mitochondrial genes ccmFN, matR, and rrn26 showed a much lower average coverage of 242, 211, and 447, respectively. Thus, the mapping results supported the result that the insertion in the *H. ammodendron* chloroplast genome was not an artifact of the assembly.

Since the *H. ammodendron* chloroplast genome reported in this study was assembled from reads obtained using total cellular DNA, the origin of 4.8 kb insert was confirmed using a complementary Sanger sequencing approach. Amplified segments flanking the entire 4,814 nt insertion were 6,607, 7,172, and 8,132 nt long with the forward and the reverse primers flanking the ycf1 and ndhF genes, respectively (Fig. 2; Table S2). Primers flanking both the ycf1 and ndhF genes coupled with a primer annealing to the middle section of the inserted region produced amplicons of predicted sizes of 3,810 and 4,458 nt (Fig. 2; Table S2). The PCR results were the first line of confirmation since no PCR amplification should be expected from the published *H. ammodendron* chloroplast genome due to primer mismatch. Interestingly, expected DNA amplicons were also obtained when PCR was performed on *Haloxylon persicum,* a close relative of *H. ammodendron* (Fig. 2). A total section of 6.2 kb, including the 4,814nt inserted section, was sequenced and validated via primer walking (Table S2). The sequenced amplicon results produced a 100% alignment match to the *H. ammodendron* chloroplast genome assembly obtained in this study. Amplification and sequence homology validation of the 4,814 nt section

confirmed the presence of the insertion in the *H ammodendron* chloroplast genome. The integration of intracellularly transferred DNA into the intergenic region of ycf1 and ndhF would be expected as an insertion in the coding region would have disrupted gene function.

This is the first report to document mitochondria-to-chloroplast inter-organellar gene transfer in the Chenopodiaceae family and the fourth example in angiosperms. However, the mechanisms underlying the transfer of genomic DNA fragments remain to be elucidated [72–74].

Chloroplast genomes among different types of C_4 species versus C_3 species

The eight chloroplast genomes studied, include the C_3 species *S. maritima* and seven forms of C_4 species. The results indicate the chloroplast genomes are very similar in the number (82-84) and type of CDS genes encoding proteins. Despite some differences in gene content and organization among the chloroplast genomes, these differences do not coincide with the type of oxygenic photosynthesis (C3 or C4) that these eight species represent. There is general conservation of genes present in the C_3 species *B. muricata* and the C_4 species. This suggests nuclear genes encode most chloroplast-targeted proteins that are needed to support the C_4 pathway. Both Kranz type and single-cell type C_4 species have dimorphic chloroplasts (relative to function in carbon assimilation, starch synthesis, and in the relative expression of photosystem I and photosystem II for balancing requirements for ATP and NADPH). In carbon assimilation, one type of chloroplast supports the fixation of atmospheric CO_2 by PEPC with the synthesis of C_4 acids. They generate energy to support the conversion of pyruvate to phosphoenolpyruvate utilizing pyruvate, Pi dikinase, adenylate kinase, and inorganic pyrophosphatase, and they support the reduction of oxaloacetate to malate by NADP-malate dehydrogenase. The other type of chloroplast has the Calvin-Benson cycle with Rubisco fixing CO_2 that is generated by decarboxylation of C_4 acids (utilizing plastid-targeted NADP-malic enzyme in some C_4 species). Currently, all enzymes required in chloroplasts to support the C_4 cycle and Calvin-Benson cycle are considered to be nuclear-encoded except the gene for the large subunit of Rubisco, which is in the chloroplast genome, while the small subunit gene is in the nucleus [75–79]. In the dual-cell Kranz type C_4 plants, cell-specific control of transcription of nuclear genes may contribute to the development of dimorphic chloroplasts. Other mechanisms must control the development of dimorphic chloroplasts in SCC4 species [see hypotheses, selective protein import, selective mRNA targeting, selective protein degradation, [76]]. Future studies are needed to determine how dimorphic chloroplasts develop to

coordinate the function of C_4 in carbon assimilation, metabolite transport between chloroplasts, and requirements of energy from photochemistry.

Conclusions

This study reports high quality and complete chloroplast genomes from seven Chenopodiaceae and one Amaranthaceae species. It reaffirms that the hybrid method of using high throughput and Sanger sequencing is rapid, efficient, and reliable for chloroplast genome sequencing [80,81]. While genome organization, gene order, and content were largely conserved, there were a few structural differences, such as the variable location of the ycf15 gene; the high repeat content in the *S. eltonica* genome; the presence of two copies of ycf1 gene in *H. ammodendron* along with the IR expansion; and the IR expansion in *S. eltonica* that includes the trnH-GTG and psbA. The biological significance of these differences remains to be investigated.

The *B. sinuspersici* chloroplast genome presented in this study represents an improved version due to the high sequencing coverage and the validation of the junction regions through Sanger sequencing. The improvement in the *B. sinuspersici* chloroplast genome sequence allowed for the identification of a higher number of chloroplast genes. Interestingly, the *H. ammodendron* chloroplast genome presented in this study is 9,681 nt larger than the previously published genome [69]. This difference originated from a duplicated region of the IR, which is 4,868 nt in size and represented a rare instance of inter-organellar DNA transfer from the mitochondria to the chloroplast genome.

C_4 plants evolved independently from C_3 species more than 60 times (Sage et al. 2011), leading to the development of different forms of Kranz, along with single-cell C_4 species, all of which have dimorphic chloroplasts coordinated in functions to support C_4 photosynthesis. This includes differential expression of enzymes in carbon assimilation, selective expression of metabolite transporters to control the flux of carbon between the two chloroplasts, and expression of photosystem I and II for production of ATP and NADPH. How these dimorphic chloroplasts develop through control of expression of nuclear and chloroplast genes is not known. Complete chloroplast genomic information on different forms of C_4 species should be useful in future studies on the control of its development, and in determining what is required for C_4 photosynthesis. Also, high-quality complete chloroplast genomes should be of value in future phylogenetic analyses towards elucidating phylogenetic relationships among C_4 species, including the relative placement of the single-cell photosynthetic C_4 species within the Angiosperms.

Methods:

Plant Material and DNA extraction

Amaranthus retroflexus, Bassia muricata, Suaeda eltonica, and *Suauaeda maritima* plants were grown in a growth chamber with a 14/10 h photoperiod, the light regime of 525 PPFD and day/night, and temperature of 28°C/18°C. The same photoperiod and light regime were used for *Bienertia cycloptera*, *B. sinuspersici,* and *Suaeda aralocaspica*; however, the day/night temperatures were modified to 35°C/18°C. *Haloxylon ammodendron* plants were grown under natural annual environmental conditions in Pullman, WA. Total cellular DNA was isolated using fresh leaf tissue from each species with a Urea Lysis Buffer Method. Briefly, leaf tissue was flash-frozen in liquid nitrogen and ground to a fine powder, and approximately 100 mg tissue was placed in 600 μl buffer containing 42% w/v Urea, 250 mM NaCl, 50 mM Tris (pH 8.0), 1% sodium dodecyl sulfate (SDS) and 20 mM EDTA. The solution was briefly vortexed, extracted with an equal volume of 1:1 phenol: chloroform, and vortexed for 45 seconds. Samples were then centrifuged at 9,500g for 5 minutes, and the supernatant was added to an equal volume of ice-cold 2-propanol. The tube was rocked gently six times and centrifuged for 10 minutes at 9,500g. The pellet was washed in 1 mL ice-cold 70% ethanol and centrifuged at 9,500g for two minutes, and the supernatant was decanted. The pellet was dried and suspended in 500 μL TE buffer with 20 μg/mL RNAse A and incubated for 30 minutes at 37˚C prior to the addition of 1/10th volume 3M sodium acetate (pH 5.3) and two volumes of 95% ethanol and rocked gently six times. The tube was centrifuged at 9,500g for ten minutes, the supernatant removed, and the pellet was rinsed with 500 μL 70% ethanol, centrifuged for 2 minutes at 9,500g, and the pellet was dried before being suspended in 50 μL TE buffer.

DNA Sequencing, validation, and contig assembly

DNA samples were sequenced on the Illumina HiSeq 2000 utilizing the 100PE chemistry by the Research Technology Support Facility at Michigan State University (East Lansing, MI. USA). Quality control on raw sequence data was performed using CLC Genomics Workbench ver. 6.0.1 (CLC) (QIAGEN, Redwood City, CA. USA). CLC was utilized for read trimming, merging reads, and filtering out low-quality sequences with a phred score below 40. Assembly and mapping of the reads to the contigs were accomplished with CLC software. Mapping of reads to contigs was conducted using the following mapping parameters: mismatch cost 2, insertion cost 3, deletion cost 3, length fraction 0.8, and similarity fraction 0.9. BLASTN searches on NCBI (https://www.ncbi.nlm.nih.gov/) were performed using

the assembled contigs as query sequences to identify contigs with high homology to chloroplast large single copy (LSC), small single copy (SSC) and inverted repeat (IR) for each of the assembled libraries obtained from each of the eight plant species. Identified IR contigs were reverse complemented, and overlapping borders of each of the identified contigs were aligned to assemble a complete chloroplast genome sequence in the following order of LSC+IR+SSC+IR. Chloroplast contig junctions from overlapping border regions were aligned and analyzed with MEGA6 version 6.0.6 (http://www.megasoftware.net/). Flanking primers for chloroplast junctions were designed utilizing Primer 3 Software [82]. PCR amplification was performed using Platinum Taq High-Fidelity DNA polymerase (Invitrogen, CA), and PCR products were purified using the QIAquick PCR purification Kit (QIAGEN, MD). Amplicons, ranging in size from 0.2 to 0.5 kb, were Sanger sequenced to ensure sequence fidelity of the DNA assembly output (Eurofins Genomics, KY). A primer walking and Sanger sequencing method was utilized to identify non-overlapping regions in the LSC+IRa and IRb+LSC junctions of *T. indica* and the IRa+SSC and SSC+IRb junctions of *S. eltonica.* The primer walking and Sanger sequencing method was also employed to validate specific conflicting sequences in the *H. ammodendron* chloroplast genome when compared to the publicly available *H. ammodendron* sequence. A remapping of the Illumina sequenced reads was performed using the final predicted chloroplast genomes from the eight species utilizing CLC software. A length fraction and similarity fraction of 0.99 was chosen as remapping parameters to ensure high stringency alignment. Assemblies generated with 0.80-0.90 and 0.99-0.99 length fraction and similarity fraction were screened to identify regions with coverage below 40x. Sequence data have been deposited to GenBank database under accession numbers MT299584 (*A. retroflexus*), MT316306 (*B. muricata*), MT316305 (*B. cycloptera*), MT316307 (*B. sinuspersici*), MT316308 (*H. ammodendron*), MT316309 (*S. aralocaspica*), MT316310 (*S. eltonica*), and MT316311 (*S. maritima*).

Genome annotation and Visualization

All the chloroplast genomes were annotated and visualized with GeSeq [83], which incorporates the Dual Organellar Genome Annotator (DOGMA) [43].

Comparisons of gene content and gene order

Comparisons for both gene content and order were performed for the eight chloroplast sequences. This comparison included three chloroplast reference genomes: *V. vinifera* (NC_007957.1)*, S.*

oleracea (AJ400848.1), and *B. vulgaris* (EF534108.1). Gene order and content were parsed manually using pair-wise comparisons between species.

Examination of the repeat structure

REPuter [64] was utilized to identify the number and location of forward, reverse, complementary, and palindromic repeats in the sequence of the eight species predicted chloroplast sequences. A minimum repeat size of 30 nt and a Hamming distance of 3 (>90% sequence identity) was utilized. Shared and unique repeats were identified manually and with the use of BLASTN based on intergenomic comparisons.

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: The sequence of the chloroplast genomes generated during the current study are accessible from GenBank accession numbers: MT299584 (*A. retroflexus*), MT316306 (*B. muricata*), MT316305 (*B. cycloptera*), MT316307 (*B. sinuspersici*), MT316308 (*H. ammodendron*), MT316309 (*S. aralocaspica*), MT316310 (*S. eltonica*), and MT316311 (*S. maritima*), [https://www.ncbi.nlm.nih.gov/genbank/]

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

Funding: This work was supported in part by USDA-NIFA Hatch Grant WNP00011 to AD and National Science Foundation Grant IBN-0641232 to GEE. BWB acknowledges graduate research assistantship support from Washington State University Graduate School.

Authors' contributions: RMS, GEE and AD designed the study. GEE and AD supervised the study. RMS and BWB performed the experiments and generated the data. RMS, BWB and AD analyzed the data. All authors read and approved the final manuscript.

Acknowledgments: The authors acknowledge Sequoia Dance for her support with PCR.

Bibliography

1. Neuhaus HE, Emes MJ. Nonphotosynthetic metabolism in plastids. Annu Rev Plant Biol. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA; 2000;51:111–40.

475 2. Namitha KK, Negi PS. Chemistry and biotechnology of carotenoids. Crit Rev Food Sci Nutr. Taylor &
476 Francis; 2010;50:728–60.

477 3. Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, et al. Analysis of 81 genes
478 from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary
479 patterns. Proc Natl Acad Sci. National Acad Sciences; 2007;104:19369–74.

480 4. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to
481 choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J
482 Bot. Wiley Online Library; 2007;94:275–88.

483 5. Green BR. Chloroplast genomes of photosynthetic eukaryotes. Plant J. 2011;66:34–44.

484 6. Svab Z, Hajdukiewicz P, Maliga P. Stable transformation of plastids in higher plants. Proc Natl Acad Sci.
485 National Acad Sciences; 1990;87:8526–30.

486 7. Dhingra A, Daniell H. Chloroplast genetic engineering via organogenesis or somatic embryogenesis.
487 Arab Protoc. Springer; 2006. p. 245–62.

488 8. Palmer JD, Osorio B, Aldrich J, Thompson WF. Chloroplast DNA evolution among legumes: loss of a
489 large inverted repeat occurred prior to other sequence rearrangements. Curr Genet. Springer;
490 1987;11:275–86.

491 9. Wu C-S, Chaw S-M. Large-scale comparative analysis reveals the mechanisms driving plastomic
492 compaction, reduction, and inversions in conifers II (cupressophytes). Genome Biol Evol. Oxford
493 University Press; 2016;8:3740–50.

494 10. Braukmann TWA, Broe MB, Stefanović S, Freudenstein J V. On the brink: the highly reduced
495 plastomes of nonphotosynthetic Ericaceae. New Phytol. Wiley Online Library; 2017;216:254–66.

496 11. Dhingra A, Folta KM. ASAP: amplification, sequencing & annotation of plastomes. BMC Genomics.
497 BioMed Central; 2005;6:176.

498 12. Cahoon AB, Sharpe RM, Mysayphonh C, Thompson EJ, Ward AD, Lin A. The complete chloroplast
499 genome of tall fescue (Lolium arundinaceum; Poaceae) and comparison of whole plastomes from the
500 family Poaceae. Am J Bot. Wiley Online Library; 2010;97:49–58.

501 13. Zhang Y-J, Ma P-F, Li D-Z. High-throughput sequencing of six bamboo chloroplast genomes:
502 phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). PLoS One. Public

503 Library of Science; 2011;6:e20596.

504 14. Wu J, Liu B, Cheng F, Ramchiary N, Choi S-R, Lim YP, et al. Sequencing of chloroplast genome using
505 whole cellular DNA and Solexa sequencing technology. Front Plant Sci. Frontiers; 2012;3:243.

506 15. Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Folta KM, et al. Rapid and accurate
507 pyrosequencing of angiosperm plastid genomes. BMC Plant Biol. BioMed Central; 2006;6:17.

508 16. Nagy E, Hegedűs G, Taller J, Kutasy B, Virág E. Illumina sequencing of the chloroplast genome of
509 common ragweed (Ambrosia artemisiifolia L.). Data Br. Elsevier; 2017;15:606–11.

510 17. Sakaguchi S, Ueno S, Tsumura Y, Setoguchi H, Ito M, Hattori C, et al. Application of a simplified
511 method of chloroplast enrichment to small amounts of tissue for chloroplast genome sequencing. Appl
512 Plant Sci. Wiley Online Library; 2017;5:1700002.

513 18. Sakulsathaporn A, Wonnapinij P, Vuttipongchaikij S, Apisitwanich S. The complete chloroplast
514 genome sequence of Asian Palmyra palm (Borassus flabellifer). BMC Res Notes. BioMed Central;
515 2017;10:740.

516 19. Keller J, Rousseau-Gueutin M, Martin GE, Morice J, Boutte J, Coissac E, et al. The evolutionary fate of
517 the chloroplast and nuclear rps16 genes as revealed through the sequencing and comparative analyses
518 of four novel legume chloroplast genomes from Lupinus. Dna Res. Oxford University Press;
519 2017;24:343–58.

520 20. Schuster TM, Setaro SD, Tibbits JFG, Batty EL, Fowler RM, McLay TGB, et al. Chloroplast variation is
521 incongruent with classification of the Australian bloodwood eucalypts (genus Corymbia, family
522 Myrtaceae). PLoS One. Public Library of Science; 2018;13:e0195034.

523 21. Nock CJ, Hardner CM, Montenegro JD, Termizi AAA, Hayashi S, Playford J, et al. Wild origins of
524 macadamia domestication identified through intraspecific chloroplast genome sequencing. Front Plant
525 Sci. Frontiers; 2019;10:334.

526 22. Dong W, Xu C, Cheng T, Lin K, Zhou S. Sequencing angiosperm plastid genomes made easy: a
527 complete set of universal primers and a case study on the phylogeny of Saxifragales. Genome Biol Evol.
528 Oxford University Press; 2013;5:989–97.

529 23. Logacheva MD, Samigullin TH, Dhingra A, Penin AA. Comparative chloroplast genomics and
530 phylogenetics of Fagopyrum esculentum ssp. ancestrale–a wild ancestor of cultivated buckwheat. BMC

Plant Biol. BioMed Central; 2008;8:59.

24. Wang R-J, Cheng C-L, Chang C-C, Wu C-L, Su T-M, Chaw S-M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol Biol. BioMed Central; 2008;8:36.

25. Peng L, Yamamoto H, Shikanai T. Structure and biogenesis of the chloroplast NAD (P) H dehydrogenase complex. Biochim Biophys Acta (BBA)-Bioenergetics. Elsevier; 2011;1807:945–53.

26. Gu C, Tembrock LR, Li Y, Lu X, Wu Z. The complete chloroplast genome of queen's crape-myrtle (Lagerstroemia macrocarpa). Mitochondrial DNA Part B. Taylor & Francis; 2016;1:408–9.

27. Weng M, Ruhlman TA, Jansen RK. Expansion of inverted repeat does not decrease substitution rates in Pelargonium plastid genomes. New Phytol. Wiley Online Library; 2017;214:842–51.

28. Zuo L-H, Shang A-Q, Zhang S, Yu X-Y, Ren Y-C, Yang M-S, et al. The first complete chloroplast genome sequences of Ulmus species by de novo sequencing: Genome comparative and taxonomic position analysis. PLoS One. Public Library of Science; 2017;12:e0171264.

29. Jurić I, González-Pérez V, Hibberd JM, Edwards G, Burroughs NJ. Size matters for single-cell C4 photosynthesis in Bienertia. J Exp Bot. Oxford University Press UK; 2016;68:255–67.

30. Offermann S, Friso G, Doroshenk KA, Sun Q, Sharpe RM, Okita TW, et al. Developmental and subcellular organization of single-cell C4 photosynthesis in Bienertia sinuspersici determined by large-scale proteomics and cDNA assembly from 454 DNA sequencing. J Proteome Res. ACS Publications; 2015;14:2090–108.

31. Edwards GE, Franceschi VR, Voznesenskaya E V. SINGLE-CELL C 4 PHOTOSYNTHESIS VERSUS THE DUAL-CELL (KRANZ) PARADIGM . Annu Rev Plant Biol. Annual Reviews; 2004;55:173–96.

32. Voznesenskaya Elena V, Franceschi Vincent R, Kiirats O, Artyusheva Elena G, Freitag H, Edwards Gerald E. Proof of C4 photosynthesis without Kranz anatomy in Bienertia cycloptera (Chenopodiaceae). Plant J. 2002;31:649–62.

33. Sage RF, Christin P-A, Edwards EJ. The C4 plant lineages of planet Earth. J Exp Bot. Oxford University Press; 2011;62:3155–69.

34. Hernández-Ledesma P, Berendsohn WG, Borsch T, Von Mering S, Akhani H, Arias S, et al. A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales.

559 Willdenowia. JSTOR; 2015;281–383.

560 35. Edwards GE, Voznesenskaya E V. Chapter 4 C4 Photosynthesis: Kranz Forms and Single-Cell C4 in
561 Terrestrial Plants. 2010. p. 29–61.

562 36. Kadereit G, Borsch T, Weising K, Freitag H. Phylogeny of Amaranthaceae and Chenopodiaceae and
563 the evolution of C4 photosynthesis. Int J Plant Sci. The University of Chicago Press; 2003;164:959–86.

564 37. Schütze P, Freitag H, Weising K. An integrated molecular and morphological study of the subfamily
565 Suaedoideae Ulbr.(Chenopodiaceae). Plant Syst Evol. Springer; 2003;239:257–86.

566 38. Kapralov M V, Akhani H, Voznesenskaya E V, Edwards G, Franceschi V, Roalson EH. Phylogenetic
567 relationships in the Salicornioideae/Suaedoideae/Salsoloideae sl (Chenopodiaceae) clade and a
568 clarification of the phylogenetic position of Bienertia and Alexandra using multiple DNA sequence
569 datasets. Syst Bot. American Society of Plant Taxonomists; 2006;31:571–85.

570 39. Sharpe RM, Offermann S. One decade after the discovery of single-cell C 4 species in terrestrial
571 plants: what did we learn about the minimal requirements of C 4 photosynthesis? Photosynth Res.
572 Springer; 2014;119:169–80.

573 40. Sage RF. The evolution of C4 photosynthesis. New Phytol. Wiley Online Library; 2004;161:341–70.

574 41. Christin P, Sage TL, Edwards EJ, Ogburn RM, Khoshravesh R, Sage RF. Complex evolutionary
575 transitions and the significance of C3–C4 intermediate forms of photosynthesis in Molluginaceae. Evol
576 Int J Org Evol. Wiley Online Library; 2011;65:643–60.

577 42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol.
578 Elsevier; 1990;215:403–10.

579 43. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA.
580 Bioinformatics. Oxford University Press; 2004;20:3252–5.

581 44. Chen Y-C, Liu T, Yu C-H, Chiang T-Y, Hwang C-C. Effects of GC bias in next-generation-sequencing data
582 on de novo genome assembly. PLoS One. Public Library of Science; 2013;8:e62856.

583 45. Rieber N, Zapatka M, Lasitschka B, Jones D, Northcott P, Hutter B, et al. Coverage bias and sensitivity
584 of variant calling for four whole-genome sequencing technologies. PLoS One. Public Library of Science;
585 2013;8:e66621.

586 46. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations
587 in genomic analyses. Nat Rev Genet. Nature Publishing Group; 2014;15:121.

588 47. Lindsey RL, Pouseele H, Chen JC, Strockbine NA, Carleton HA. Implementation of whole genome
589 sequencing (WGS) for identification and characterization of Shiga toxin-producing Escherichia coli (STEC)
590 in the United States. Front Microbiol. Frontiers; 2016;7:766.

591 48. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, et al. Comparative
592 chloroplast genomics: analyses including new sequences from the angiosperms Nuphar advena and
593 Ranunculus macranthus. BMC Genomics. BioMed Central; 2007;8:174.

594 49. Park J-S, Choi I-S, Lee D-H, Choi B-H. The complete plastid genome of Suaeda malacosperma
595 (Amaranthaceae/Chenopodiaceae), a vulnerable halophyte in coastal regions of Korea and Japan.
596 Mitochondrial DNA Part B. Taylor & Francis; 2018;3:382–3.

597 50. Kim Y, Park J, Chung Y. The complete chloroplast genome of Suaeda japonica Makino
598 (Amaranthaceae). Mitochondrial DNA Part B. Taylor & Francis; 2019;4:1505–7.

599 51. Kim B, Kim J, Park H, Park J. The complete chloroplast genome sequence of Bienertia sinuspersici.
600 Mitochondrial DNA Part B. Taylor & Francis; 2016;1:388–9.

601 52. Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, et al. The complete chloroplast genome sequence of
602 date palm (Phoenix dactylifera L.). PLoS One. Public Library of Science; 2010;5:e12762.

603 53. Firetti F, Zuntini AR, Gaiarsa JW, Oliveira RS, Lohmann LG, Van Sluys M. Complete chloroplast
604 genome sequences contribute to plant species delimitation: A case study of the Anemopaegma species
605 complex. Am J Bot. Wiley Online Library; 2017;104:1493–509.

606 54. Fonseca LHM, Lohmann LG. Plastome rearrangements in the “Adenocalymma-Neojobertia” Clade
607 (Bignonieae, Bignoniaceae) and its phylogenetic implications. Front Plant Sci. Frontiers; 2017;8:1875.

608 55. Kim K-J, Lee H-L. Complete chloroplast genome sequences from Korean ginseng (Panax schinseng
609 Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. Oxford
610 University Press; 2004;11:247–61.

611 56. Dong W, Liu J, Yu J, Wang L, Zhou S. Highly variable chloroplast markers for evaluating plant
612 phylogeny at low taxonomic levels and for DNA barcoding. PLoS One. Public Library of Science;
613 2012;7:e35071.

614 57. Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, et al. The complete chloroplast genome sequence of the
615 medicinal plant Salvia miltiorrhiza. PLoS One. Public Library of Science; 2013;8:e57607.

616 58. Asaf S, Khan AL, Khan AR, Waqas M, Kang S-M, Khan MA, et al. Complete chloroplast genome of
617 Nicotiana otophora and its comparison with related species. Front Plant Sci. Frontiers; 2016;7:843.

618 59. Downie SR, Olmstead RG, Zurawski G, Soltis DE, Soltis PS, Watson JC, et al. Six independent losses of
619 the chloroplast DNA rpl2 intron in dicotyledons: molecular and phylogenetic implications. Evolution (N
620 Y). Wiley Online Library; 1991;45:1245–59.

621 60. Palmer JD. Chloroplast DNA exists in two orientations. Nature. Nature Publishing Group;
622 1983;301:92–3.

623 61. Oldenburg DJ, Bendich AJ. Most chloroplast DNA of maize seedlings in linear molecules with defined
624 ends and branched forms. J Mol Biol. Elsevier; 2004;335:953–70.

625 62. Martin G, Baurens F-C, Cardi C, Aury J-M, D'Hont A. The complete chloroplast genome of banana
626 (Musa acuminata, Zingiberales): insight into plastid monocotyledon evolution. PLoS One. Public Library
627 of Science; 2013;8:e67350.

628 63. Walker JF, Zanis MJ, Emery NC. Correction to "Comparative analysis of complete chloroplast genome
629 sequence and inversion variation in Lasthenia burkei (Madieae, Asteraceae)." Am J Bot. Wiley Online
630 Library; 2015;102:1008.

631 64. Kurtz S, Choudhuri J V, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold
632 applications of repeat analysis on a genomic scale. Nucleic Acids Res. Oxford University Press;
633 2001;29:4633–42.

634 65. Tranel PJ, Trucco F. 21st-century weed science: a call for Amaranthus genomics. Weedy and invasive
635 plant genomics. Wiley Online Library; 2009;53–81.

636 66. Chaney L, Mangelson R, Ramaraj T, Jellen EN, Maughan PJ. The complete chloroplast genome
637 sequences for four Amaranthus species (Amaranthaceae). Appl Plant Sci. 2016;4:1–6.

638 67. Caburatan L, Kim JG, Park J. Comparative Chloroplast Genome Analysis of Single-Cell C4 Bienertia
639 Sinuspersici with Other Amaranthaceae Genomes. J Plant Sci. Science Publishing Group; 2018;6:134–43.

640 68. Li Z, Chen Y, Mu D, Yuan J, Shi Y, Zhang H, et al. Comparison of the two major classes of assembly
641 algorithms: overlap–layout–consensus and de-bruijn-graph. Brief Funct Genomics. Oxford University

642 Press; 2012;11:25–37.

643 69. Dong W, Xu C, Li D, Jin X, Li R, Lu Q, et al. Comparative analysis of the complete chloroplast genome
644 sequences in psammophytic Haloxylon species (Amaranthaceae). PeerJ. PeerJ Inc.; 2016;4:e2699.

645 70. Goulding SE, Wolfe KH, Olmstead RG, Morden CW. Ebb and flow of the chloroplast inverted repeat.
646 Mol Gen Genet MGG. Springer; 1996;252:195–206.

647 71. Lee S-B, Kaittanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD, et al. The complete chloroplast
648 genome sequence of Gossypium hirsutum: organization and phylogenetic relationships to other
649 angiosperms. BMC Genomics. BioMed Central; 2006;7:61.

650 72. Ma P-F, Zhang Y-X, Guo Z-H, Li D-Z. Evidence for horizontal transfer of mitochondrial DNA to the
651 plastid genome in a bamboo genus. Sci Rep. Nature Publishing Group; 2015;5:11608.

652 73. Iorizzo M, Grzebelus D, Senalik D, Szklarczyk M, Spooner D, Simon P. Against the traffic: the first
653 evidence for mitochondrial DNA transfer into the plastid genome. Mob Genet Elements. Taylor &
654 Francis; 2012;2:261–6.

655 74. Straub SCK, Cronn RC, Edwards C, Fishbein M, Liston A. Horizontal transfer of DNA from the
656 mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae).
657 Genome Biol Evol. Oxford University Press; 2013;5:1872–85.

658 75. Wimmer D, Bohnhorst P, Shekhar V, Hwang I, Offermann S. Transit peptide elements mediate
659 selective protein targeting to two different types of chloroplasts in the single-cell C4 species Bienertia
660 sinuspersici. Sci Rep. 2017;7:41187.

661 76. Offermann S, Okita TW, Edwards GE. Resolving the compartmentation and function of C4
662 photosynthesis in the single-cell C4 species Bienertia sinuspersici. Plant Physiol. Plant Physiol;
663 2011;155:1612–28.

664 77. Hibberd JM, Covshoff S. The Regulation of Gene Expression Required for C 4 Photosynthesis . Annu
665 Rev Plant Biol. Annual Reviews; 2010;61:181–207.

666 78. Sharpe RM, Offermann S. One decade after the discovery of single-cell C4 species in terrestrial
667 plants: What did we learn about the minimal requirements of C 4 photosynthesis? Photosynth. Res.
668 Photosynth Res; 2014. p. 169–80.

669 79. Dhingra A, Portis Jr. AR, Daniell H. Enhanced translation of a chloroplast-expressed RbcS gene

restores small subunit levels and photosynthesis in nuclear RbcS antisense plants. Proc Natl Acad Sci U S A. 2004;101.

80. Twyford AD, Ness RW. Strategies for complete plastid genome sequencing. Mol Ecol Resour. Wiley Online Library; 2017;17:858–68.

81. Wang W, Schalamun M, Morales-Suarez A, Kainer D, Schwessinger B, Lanfear R. Assembly of chloroplast genomes with long-and short-read data: a comparison of approaches using Eucalyptus pauciflora as a test case. BMC Genomics. BioMed Central; 2018;19:977.

82. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. Nucleic Acids Res. Oxford University Press; 2012;40:e115–e115.

83. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. Oxford University Press; 2017;45:W6–11.

Table 1: Sequencing and assembly data when length fraction and similarity fraction parameters were set to 80 and 90 respectively during read mapping in the chloroplast genomes of eight Chenopod species.

Variable	Species							
	A. retroflexus	***B. cycloptera***	***B. muricata***	***B. sinuspersici***	***H. ammodendron***	***S. aralocaspica***	***S. eltonica***	***S. maritima***
Total number of reads	94,491,120	73,061,587	61,098,096	80,215,373	81,126,072	86,825,544	66,358,800	87,411,736
Mean read length (nt)	78.98	79.04	79.44	79.66	78.5	78.83	78.35	79.88
Minimum coverage (bases)	22	12	37	39	12	55	25	33
Maximum coverage (bases)	29,721	6,135	23,533	15,558	15,284	23,266	14,652	6,991
Average coverage (bases)	3,649.67	1,553.12	3,204.28	5,998.08	1,357.20	4,864.44	1,591.56	4,111.85
Total (%) of reads assemble to genome	7.37	4.12	10.01	14.39	3.44	10.42	4.55	8.95

Table 2: A summary of the complete chloroplast genome, IR, LSC and SSC length (nt) and GC content from *A. retroflexus, B. muricata, B. cycloptera, B. sinuspersici, H. ammodendron, S. aralocaspica, S. eltonica,* and *S. maritima*.

Species	Complete Chloroplast Genome		IR		LSC		SSC	
	Size (bp)	GC Content (%)	IRs size	GC Content (%)	LSC	GC Content (%)	SSC	GC Content (%)
A. retroflexus	150,786	36.67	24,353	42.64	83,963	34.51	18,117	30.20
B. muricata	151,593	36.61	24,355	43.00	84,288	34.50	18,595	29.42
B. cycloptera	153,341	36.50	24,942	42.92	84,541	34.42	18,916	29.69
B. sinuspersici	153,334	36.65	24,949	42.97	84,490	34.57	18,946	29.49
H. ammodendron	161,251	36.42	29,061	42.90	84,236	34.18	18,893	29.42
S. aralocaspica	146,634	36.53	23,461	42.94	81,878	34.42	17,834	29.30
S. eltonica	148,729	36.44	24,585	42.11	80,218	34.69	19,341	29.20
S. maritima	152,011	36.45	25,213	42.72	83,482	34.11	18,103	29.17

Table 3. A summary of the number of genes in the eight Chenopodiaceae chloroplast genomes.

Species name	CDS genes	rRNA	tRNA	genes w Introns	tRNA w Introns	Total genes
A. retroflexus	83	4	29	rps12, rps16, atpF, rpoC1, ycf3, clpP, ndhB, ndhA, ndhB	trnK-UUU, trnS-AGA, trnS-CGA, trnL-UAA, trnV-UAC, trnR-UCU, trnA-UGC, trnE-UUC, trnW-CCA, trnStop-UUA, trnC-ACA, trnD-GUC	116
B. cycloptera	83	4	27	clpP, rps12, ycf3, rpoC1, atpF, rps16, ndhB, ndhA, ndhB	trnT-UGU, trnC-ACA, trnL-UAA, trnF-GAA, trnS-CGA, trnK-UUU, trnE-UUC, trnA-UGC, trnW-CCA, trnW-CCA, trnA-UGC, trnE-UUC	114
B. muricata	84	4	28	rps12, rps16, atpF, rpoC1, ycf3, rps12, clpP, ndhB, ndhA	trnK-UUU, trnS-AGA, trnS-CGA, trnL-UAA, trnV-UAC, trnE-UUC, trnA-UGC, trnR-UCU, trnW-CCA, trnA-UGC, trnE-UUC	116
B. sinuspersici	83	4	27	clpP, rps12, ycf3, rpoC1, atpF, rps16, ndhB, ndhA, ndhB	trnT-UGU, trnC-ACA, trnL-UAA, trnF-GAA, trnS-CGA, trnK-UUU, trnE-UUC, trnA-UGC, trnW-CCA, trnW-CCA, trnA-UGC, trnE-UUC	114
H. ammodendron	82	4	27	clpP, rps12, ycf3, rpoC1, atpF, rps16, ndhB, ndhA, ndhB	trnC-ACA, trnL-UAA, trnS-CGA, trnK-UUU, trnE-UUC, trnA-UGC, trnW-CCA, trnW-CCA, trnA-UGC, trnE-UUC	113
S. aralocaspica	84	4	27	clpP, rps12, ycf3, rpoC1, atpF, rps16, ndhB, ndhA, ndhB	trnV-UAC, trnL-UAA, trnG-CCC, trnK-UUU, trnK-UUU, trnE-UUC, trnA-UGC, trnW-CCA, trnW-CCA, trnA-UGC, trnE-UUC	115
S. eltonica	83	4	28	clpP, rps12, ycf3, rpoC1, atpF, rps16, ndhB, ndhA, ndhB	trnA-GGC, trnV-UAC, trnL-UAA, trnS-CGA, trnK-UUU, trnE-UUC, trnI-GAU, trnA-UGC, trnA-UGC, trnE-UUC	115
S. maritima	84	4	28	clpP, rps12, ycf3, rpoC1, atpF, rps16, ndhB, ndhA, ndhB	trnC-ACA, trnL-UAA, trnK-CUU, trnS-CGA, trnK-UUU, trnE-UUC, trnA-UGC, trnW-CCA, trnW-CCA, trnA-UGC, trnE-UUC	116

Table 4: Distribution of repeated sequences (>/30 nt) among intergenic regions, exons and introns in eight chloroplast genomes.

Location	*A. retroflexus*		*B. cycloptera*		*B. muricata*		*B. sinuspersici*		*H. ammodendron*		*S. aralocaspica*		*S. eltonica*		*S. maritima*	
	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
Intergenic	17	50.00	12	38.71	16	33.33	13	40.63	25	54.35	13	34.21	136	78.16	20	41.67
Exons	14	41.18	16	51.61	32	66.67	16	50.00	21	45.65	22	57.89	35	20.11	25	52.08
Introns	3	8.82	3	9.68	0	0.00	3	9.38	0	0.00	3	7.89	3	1.72	3	6.25

Table 5: Polymorphic simple sequence repeats (SSRs) in *Amaranthus hypochondriacus* and *A. retroflexus*.

SRR location in *A. hypochondriacus* chloroplast genome (nt)	Repeat unit	Number of repeats	
		A. hypochondriacus	*A. retroflexus*
5,572 - 5,583	T	12	10
7,526 - 7,537	T	12	10
46,236 - 46,253	TA	9	8
46,573 - 46,588	AT	8	8
47,532 - 47,543	A	12	13
52,543 - 52,557	T	15	12
54,580 - 54,591	A	12	13
65,482 - 65,496	T	15	18
70,858 - 70,869	A	12	11
79,076 - 79,087	T	12	14
112,930 - 112,944	T	15	14
116,360 - 116,371	TATT	3	4

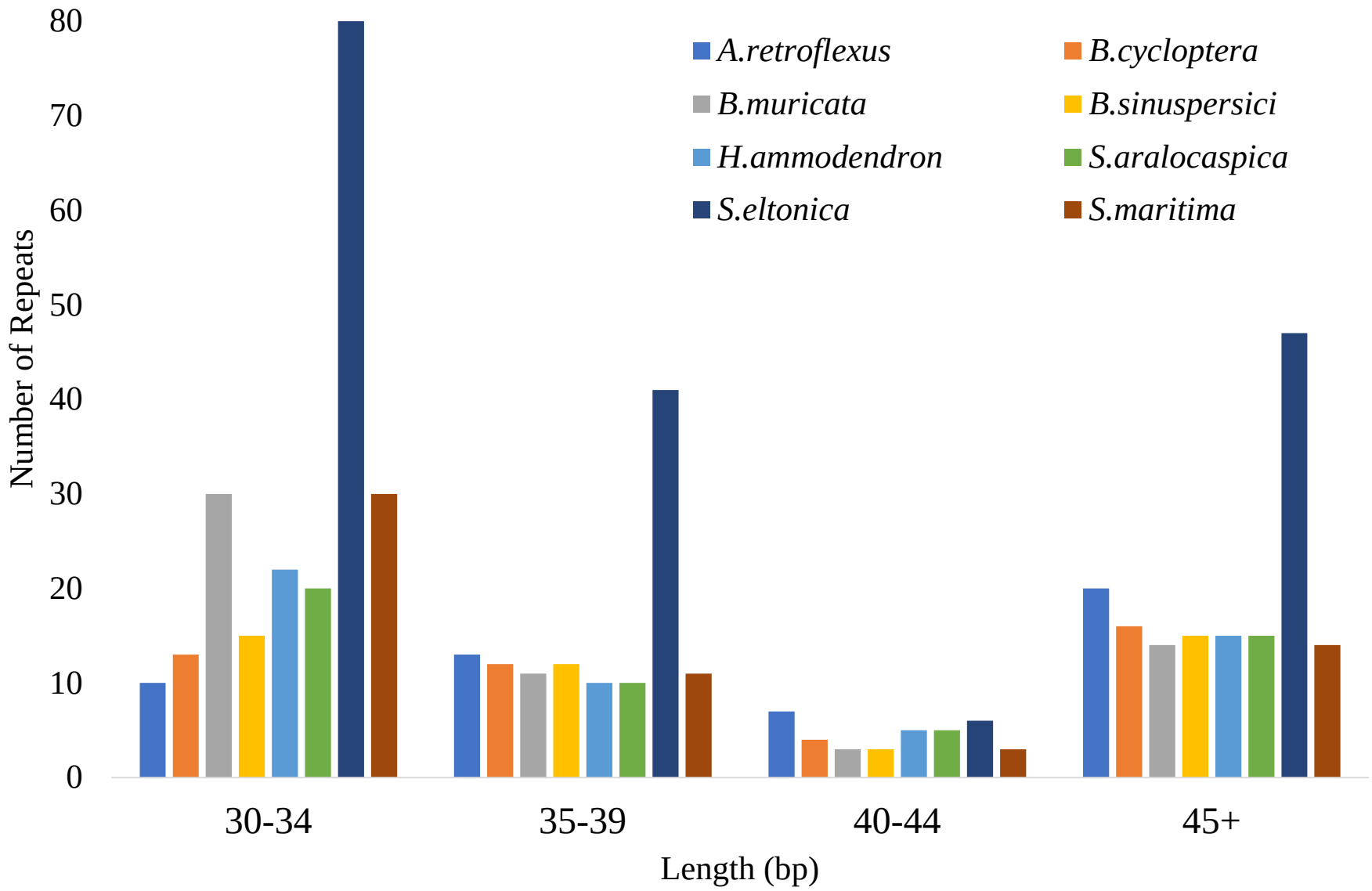


Figure 1. Histogram of number of repeated sequences (>/30 nt) in length identified with REPuter for nine chloroplast genomes. .

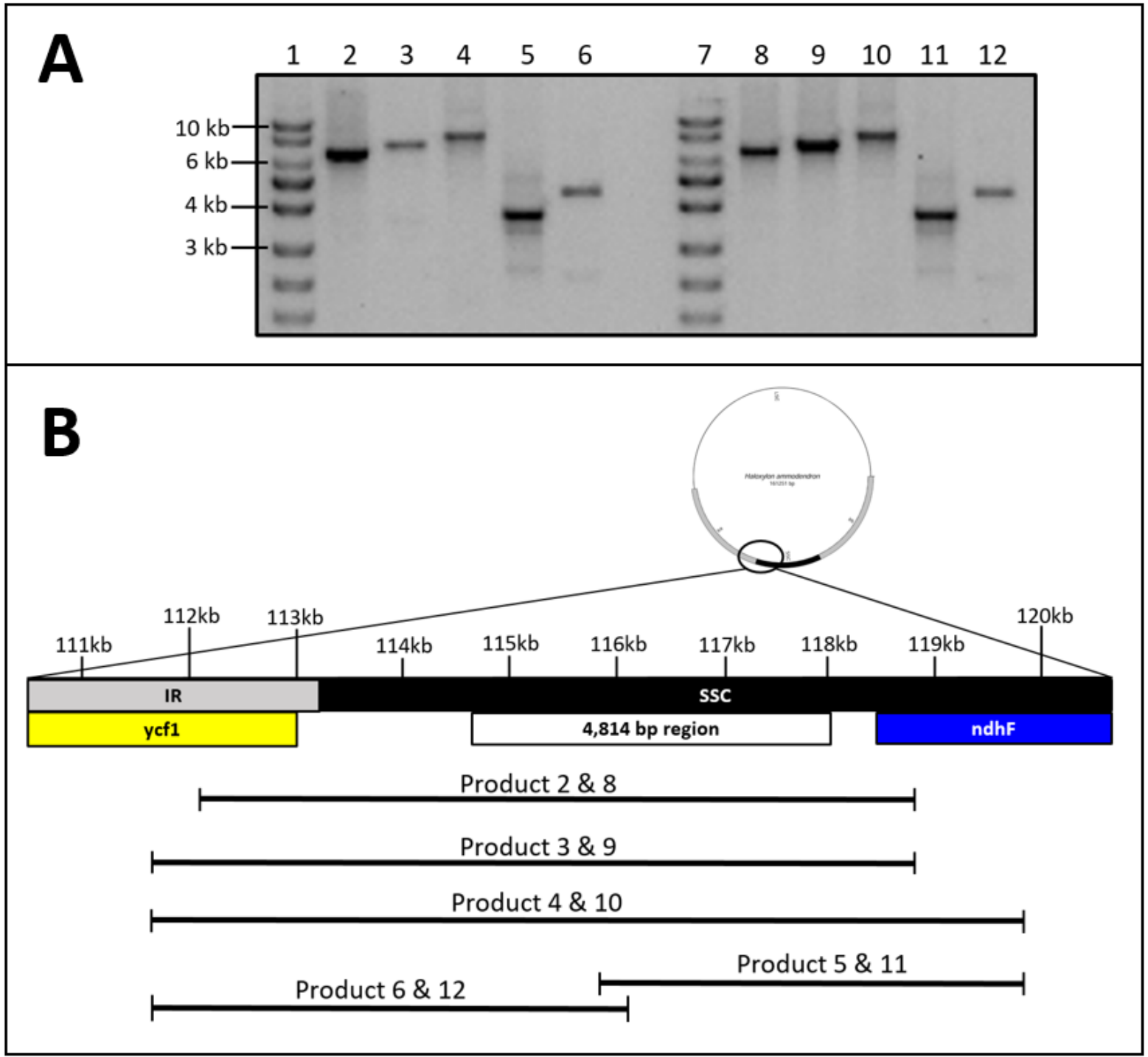


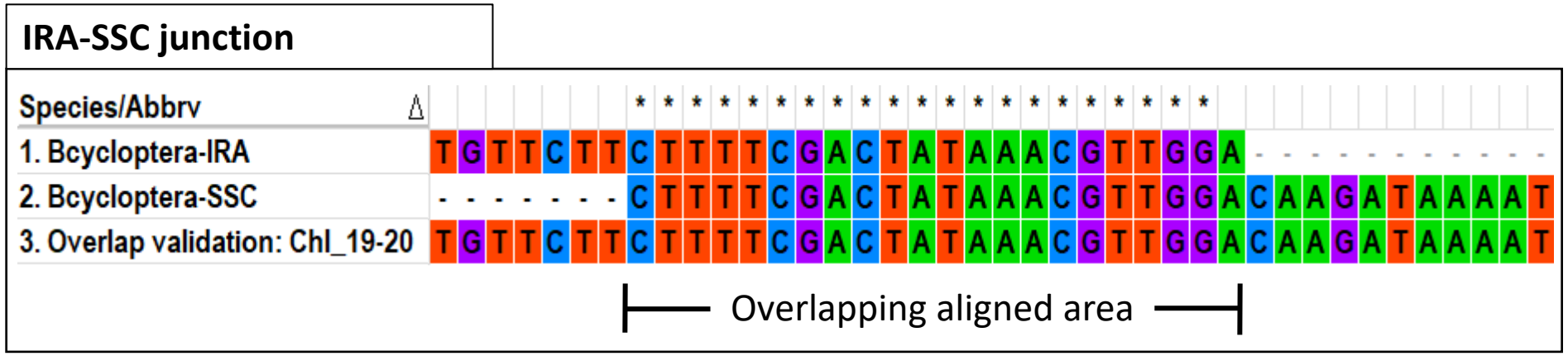
Figure 2. PCR amplicons flanking a 4.8kb insertion in the chloroplast genome of *H. ammodendron* (2 to 6) and *H. persicum* (8 to 12). Agarose gel electrophoresis of PCR products (A); diagram representing location and length of each amplicon (B). Expected amplicon sizes are 6607 (2 and 8), 7172 (3 and 9), 8132 (4 and 6), 3810 (5 and 11), and 4458 nt (6 and 12). Primers for PCRs 2 to 4 and 8 to 10 flank the ycf1 and ndhF genes. Primers for PCRs 5 and 11 flank flank the middle section of the 4.8 kb insertion and the ndhF gene. Primers for PCRs 6 and 12 flank the ycf1 gene and the middle section of the 4.8 kb insertion. 1 and 7: exACTGene DNA Ladders 1kb DNA Ladder..

Supplementary Table 1. Forward and reverse primers used to amplify and validate the overlapping regions present in all four possible junctions (LSC-IR, IR-SSC, SSC-IR, and IR-LSC) of eight chloroplast genomes.

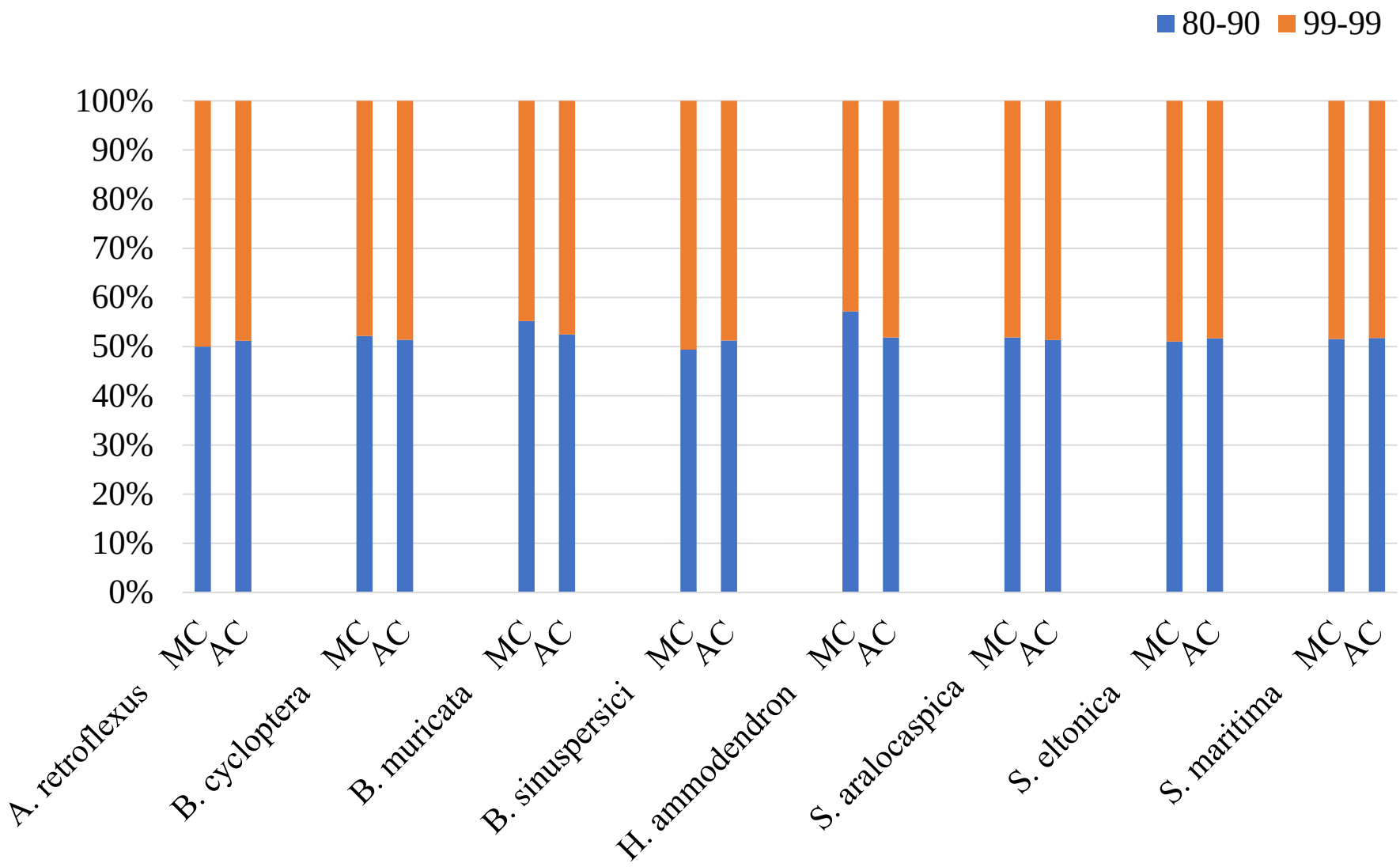
Species	Junction	Primer Name	Directionality	Sequence	Amplicon Size
A. retroflexus	LSC-IRA	Chl_1	Forward	CGGTTTCAACTTCAACCAATC	967
		Chl_2	Reverse	CCAAAAACTGCTCAGCAACA	
	IRA-SSC	Chl_3	Forward	CAAAGGATGAATTCCACTTGC	988
		Chl_4	Reverse	ACGTTTCGATTTCCTTGTGC	
	SSC-IRB	Chl_5	Forward	TCCAATTTGAATGAGTCCGTAG	961
		Chl_6	Reverse	TCCGAGTGAATGGAAAGGAC	
	IRB-LSC	Chl_7	Forward	AGTCGGACAAGTGGGAAATG	952
		Chl_8	Reverse	CATGCTCGATAGGGCATAAG	
B. muricata	LSC-IRA	Chl_9	Forward	CATGCTCGATAAGGCATGAG	877
		Chl_10	Reverse	TGGCTAGGTAAGCGTCCTGT	
	IRA-SSC	Chl_11	Forward	TGAAAGGGGCTGAATTGGTA	944
		Chl_12	Reverse	GAGGTTTGGCTGAACTAATTTCG	
	SSC-IRB	Chl_13	Forward	AAGGATCGTGGTTTTCTTGC	972
		Chl_14	Reverse	TGAAAGGGGCTGAATTGGTA	
	IRB-LSC	Chl_15	Forward	CAAGTGGGGAATGTTGGAGT	992
		Chl_16	Reverse	CTGGGCGGATATCATTAACC	
B. cycloptera	LSC-IRA	Chl_17	Forward	AGCTGCTATTGAAGCTCCATC	979
		Chl_18	Reverse	CCAAAAACTGCTCAGCAACA	
	IRA-SSC	Chl_19	Forward	GCAGAATGCCGTCACCTATT	896
		Chl_20	Reverse	GGGGGTCGAATTTCTTCCTA	
	SSC-IRB	Chl_21	Forward	TCGTTTTGAGTTTTATCGTTGC	968
		Chl_22	Reverse	TTGAGAAATTCTATGGCTCGAA	
	IRB-LSC	Chl_23	Forward	CCAAAAACTGCTCAGCAACA	966
		Chl_24	Reverse	TTGCTGCTGCAGAATAAATCA	
B. sinuspersici	LSC-IRA	Chl_25	Forward	AGCTGCTATTGAAGCTCCATCT	980
		Chl_26	Reverse	CCAAAAACTGCTCAGCAACA	
	IRA-SSC	Chl_27	Forward	GCAGAATGCCGTCACCTATT	895
		Chl_28	Reverse	GGGGGTCGAATTTCTTCCTA	
	SSC-IRB	Chl_29	Forward	CGTTTTGAGTTTTATCGTTGC	804
		Chl_30	Reverse	GCAGAATGCCGTCACCTATT	
	IRB-LSC	Chl_31	Forward	AAACTGCTCAGCAACAGTCG	974
		Chl_32	Reverse	GTGCCTAGCATTTGCTGCT	
H. ammodendrum	LSC-IRA	Chl_33	Forward	GCTCGATAGGGCATGAGTTC	933
		Chl_34	Reverse	CCAAAAACTGCTCAGCAACA	
	IRA-SSC	Chl_35	Forward	TTTCTTTGGCCCAATTATCG	907
		Chl_36	Reverse	TGGTTCGCAATCAATAACATC	
	SSC-IRB	Chl_37	Forward	CCCCTTTTATTTCCCTCCAA	938
		Chl_38	Reverse	CAAAACAAATAGGAAATCGGGTA	
	IRB-LSC	Chl_39	Forward	CTCAGCAACAGTCGGACAAG	981
		Chl_40	Reverse	TGGCCTTCAACCTAAATGGT	
S. aralocaspica	LSC-IRA	Chl_41	Forward	TTGACCGTAAAGGGGCAGTA	986
		Chl_42	Reverse	GCCTATTGGAATTGGTGTGG	
	IRA-SSC	Chl_43	Forward	GCAAGAAGAGGGATTCATCG	867
		Chl_44	Reverse	CTGGGGATTGGTATGAATTTG	
	SSC-IRB	Chl_45	Forward	ATACCACGATGAGATCCGTTT	800
		Chl_46	Reverse	TCTGAAACGGAGAAGACTAAACAGC	
	IRB-LSC	Chl_47	Forward	GCCTATTGGAATTGGTGTGG	902
		Chl_48	Reverse	CCTCCCCAGAGGGTAATTTT	
Suaeda eltonica	LSC-IRA	Chl_49	Forward	AGATTCCAATGGATATTGGTTGAC	805
		Chl_50	Reverse	CTAGAAGTTACCAAGGAACCATGC	
	IRA-SSC	Chl_51	Forward	AAAAGACAGGGAAGGGGTCA	2440
		Chl_52	Reverse	TCCAGAGACGCGACTTGAAT	
		Chl_53	Forward	GAATTGAGATGATTCGTAGTTTCG	1205
		Chl_54	Reverse	AATAGGTGACGGCATTCTGC	
	SSC-IRB	Chl_55	Forward	ACTGTGACATTTCATTTCTTACCG	2315
		Chl_56	Reverse	CTACTCTTAACAGCCAAAGCGAGT	
		Chl_57	Forward	CGAATCGAATCATTTTGCTG	706
		Chl_58	Reverse	TTCGTTTTATTTTCGCCATAG	
	IRB-LSC	Chl_59	Forward	CTAGAAGTTACCAAGGAACCATGC	804
		Chl_60	Reverse	TCGATTAGGGTCGTATTCTATGGT	
S. maritima	LSC-IRA	Chl_61	Forward	TGCTCATAACTTCCCTCTAGACCT	801
		Chl_62	Reverse	AGTTATGAACCCTGTAGACCATCC	
	IRA-SSC	Chl_63	Forward	AACAAAGGATGAATTCCACTTGTA	899
		Chl_64	Reverse	TTTCGATTAGTATAGCTTATGCAGGA	
	SSC-IRB	Chl_65	Forward	TTCTTGATCTTCCCAACCTATTTG	835
		Chl_66	Reverse	GGAATAGAGACAATTCGGAAACAG	
	IRB-LSC	Chl_67	Forward	TTAAAATTACCTTCTGGGGAGGTC	885
		Chl_68	Reverse	TCTGATCAATTACTCTTCGTGCTT	

Supplementary Table 2. Forward and reverse primers used to amplify and validate the mitochondrion-to-plastidial DNA transfer in *Haloxylon ammodendron* and *H. persicum*.

Intersection	Primer Name	Directionality	Sequence	Amplicon Size	Position of the 5' end base in the chloroplast genome
ycf1-ndhf	Chl_69	Forward	CCATCAACGTCAAATGATTTTT	6607	112168
	Chl_70	Reverse	AAACCCGTTTATTCATCCTTAAAA		118775
ycf1-ndhf	Chl_71	Forward	TGATGTTTTATCATCGAAATTTGTTT	7172	111603
	Chl_72	Reverse	AAACCCGTTTATTCATCCTTAAAA		118775
ycf1-ndhf	Chl_73	Forward	TGATGTTTTATCATCGAAATTTGTTT	8132	111603
	Chl_74	Reverse	GCGACCTTAGCTCTTGCTCA		119735
4.8 extra region-ndhF	Chl_75	Forward	GCGGAATCACCCTTTGTTAC	3810	115925
	Chl_76	Reverse	GCGACCTTAGCTCTTGCTCA		119735
ycf1-4.8 extra region	Chl_77	Forward	TGATGTTTTATCATCGAAATTTGTTT	4458	111603
	Chl_78	Reverse	TCACCGATCAGGTGTTCTCA		116061



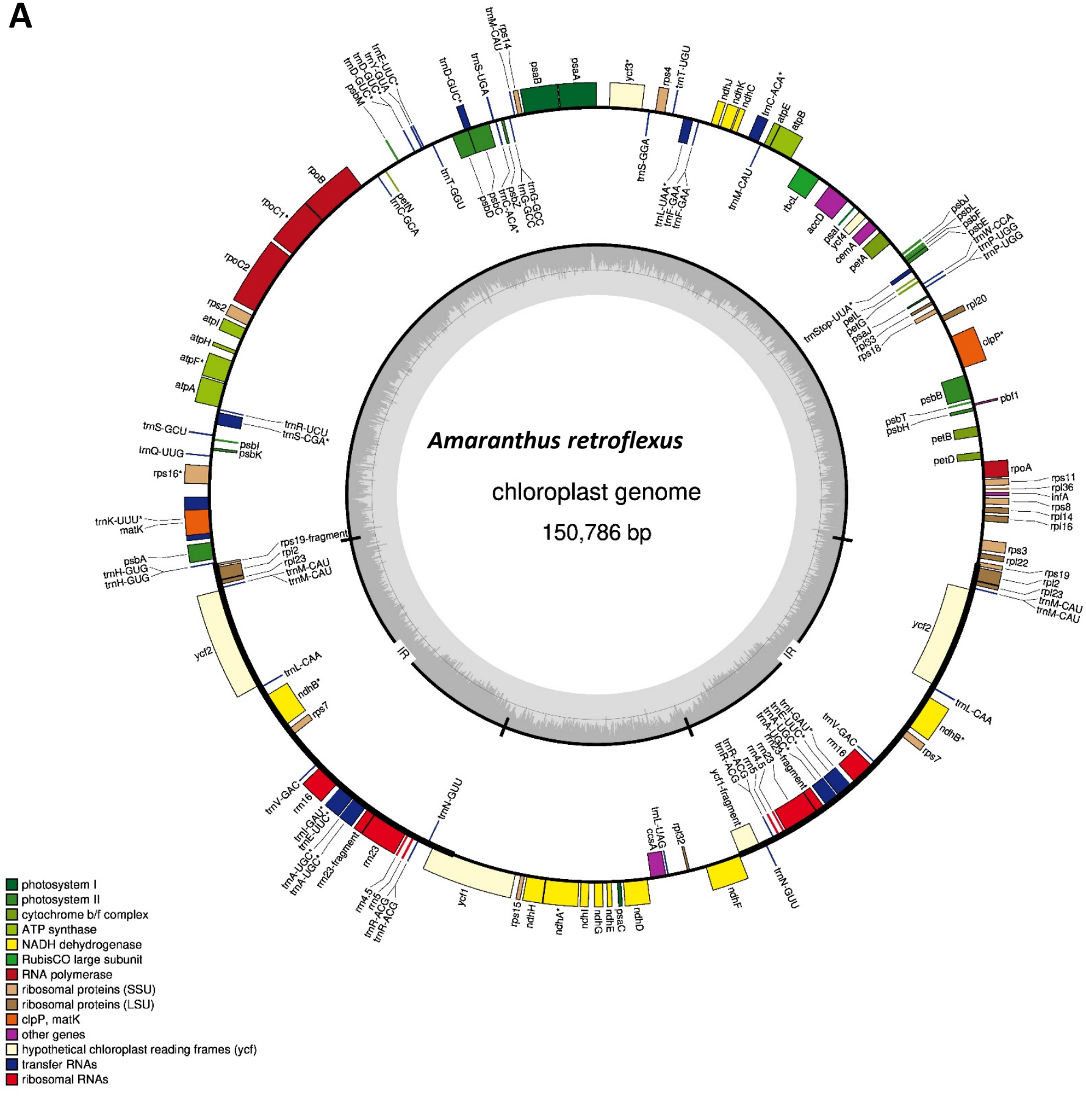
Supplementary Figure 1. Representative example of an overlap region amplicon sequenced with Sanger approch. The IRA-SSC junction showed a 100% match during nucleotide alignment.



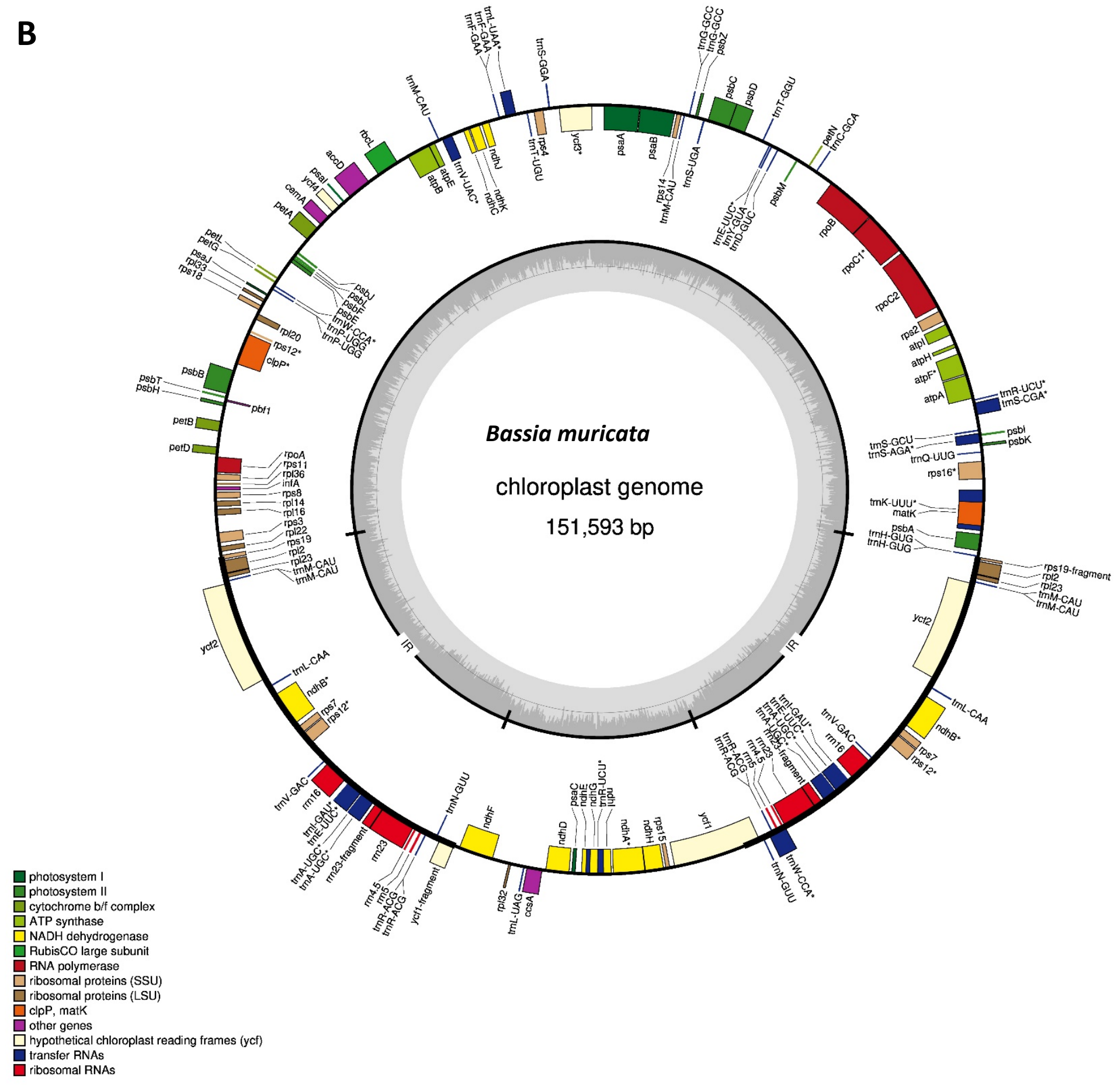
Supplementary Figure 2. Stack column graphs of minimum coverage (MC) and average coverage (AC) for eight chloroplast genomes assembled with 80%-90% (blue) and 99%-99% (orange) length fraction-similarity fraction parameters.

Supplementary Figure 3. Representative maps of the chloroplast genome of A. *Amaranthus retroflexus* , B. *Bassia muricata,* C. *Bienertia cycloptera,* D. *B. sinuspersici,* E. *Haloxylon ammodendron,* F. *Suaeda aralocaspica,* G. *S. eltonica,* and H. *S. maritima*. Genes shown outside the outer circle are transcribed clockwise whereas those represented inside are transcribed counterclockwise. Large single copy (LSC), small single copy (SSC), and inverted repeats (IRa, IRb) regions are indicated.

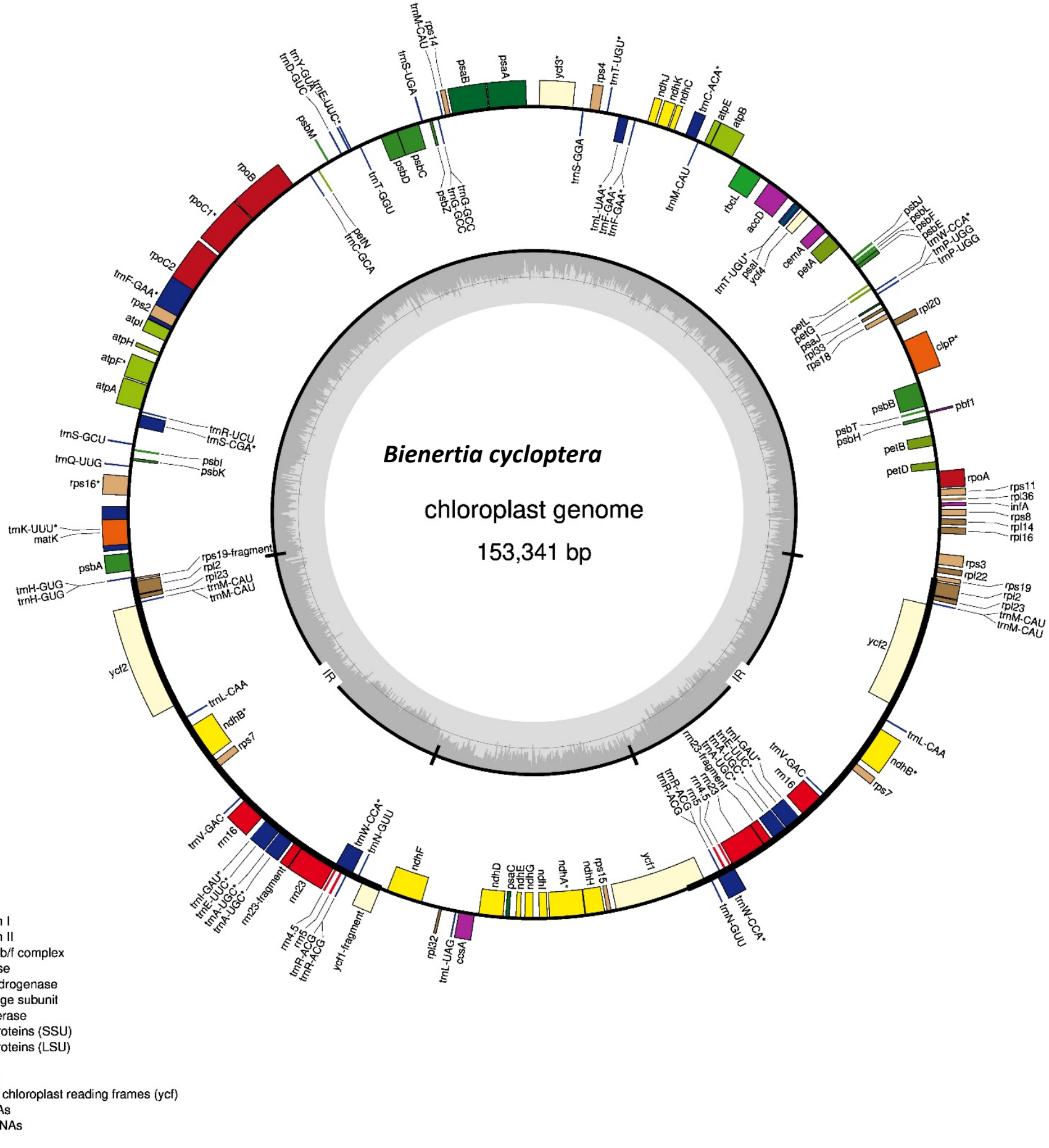
A


Amaranthus retroflexus
chloroplast genome
150,786 bp
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

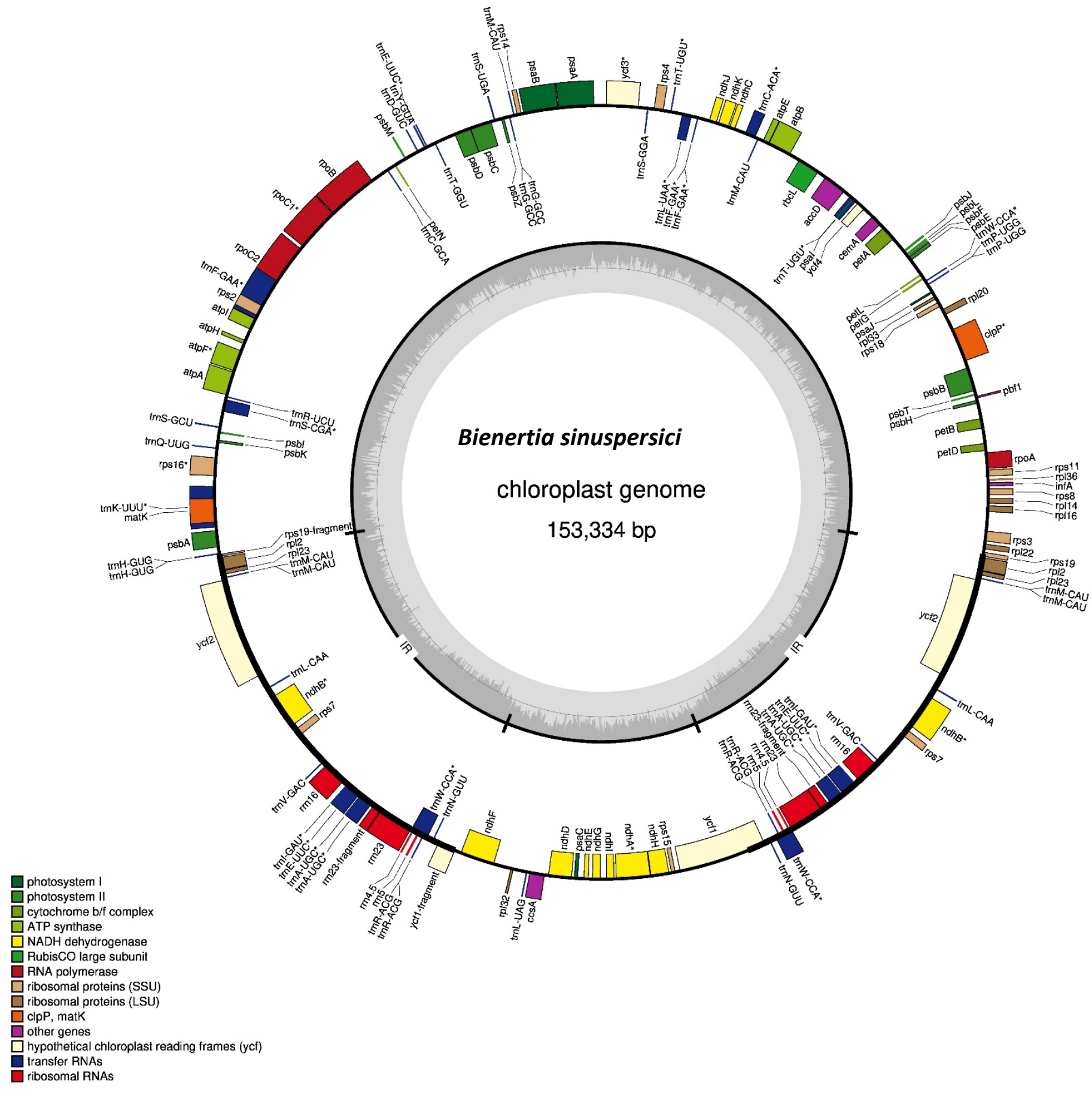
B


Bassia muricata
chloroplast genome
151,593 bp
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

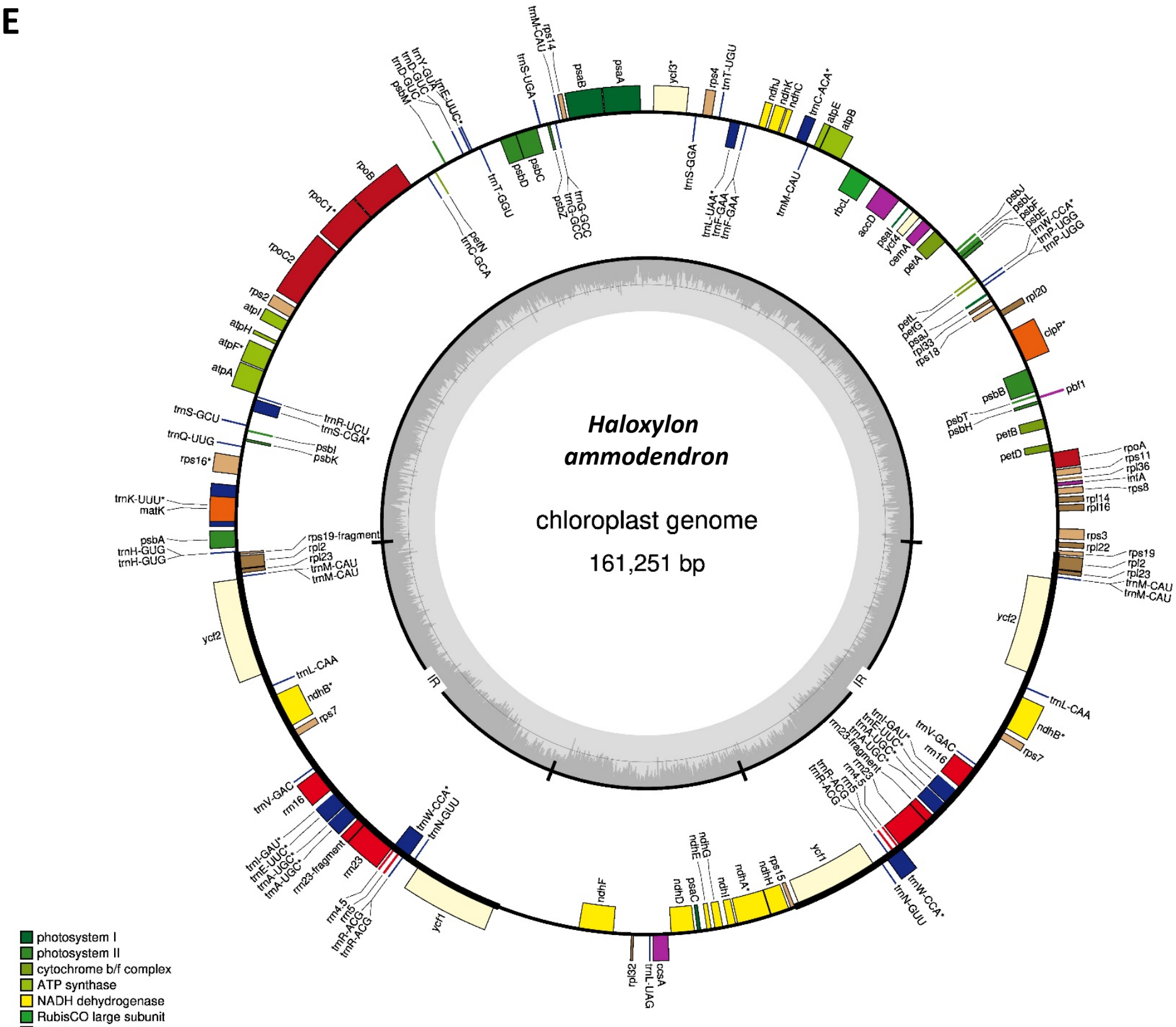
C


Bienertia cycloptera
chloroplast genome
153,341 bp
IR
IR
rpoB
rpoC1*
rpoC2
trnF-GAA*
rps2
atpI
atpH
atpF*
atpA
trnS-GCU
trnQ-UUG
rps16*
trnK-UUU*
matK
psbA
trnH-GUG
trnH-GUG
rps19-fragment
rpl2
rpl23
trnM-CAU
trnM-CAU
ycf2
trnL-CAA
ndhB*
rps7
trnV-GAC
rrn16
trnI-GAU*
trnE-UUC*
trnA-UGC*
trnA-UGC*
rrn23-fragment
rrn23
rrn4.5
rrn5
trnR-ACG
trnR-ACG
ycf1-fragment
trnN-GUU
trnW-CCA*
ndhF
rpl32
trnL-UAG
ccsA
ndhD
psaC
ndhE
ndhG
ndhI
ndhA*
ndhH
rps15
ycf1
trnV-GAC
rrn16
trnI-GAU*
trnE-UUC*
trnA-UGC*
trnA-UGC*
rrn23-fragment
rrn23
rrn4.5
rrn5
trnR-ACG
trnR-ACG
trnN-GUU
trnW-CCA*
ycf2
trnL-CAA
ndhB*
rps7
rps3
rpl22
rps19
rpl2
rpl23
trnM-CAU
trnM-CAU
rps11
rpl36
infA
rps8
rpl14
rpl16
rpoA
petD
petB
psbT
psbH
psbB
pbf1
clpP*
rpl20
petL
petG
psaJ
rpl33
rps18
psbJ
psbL
psbF
psbE
trnW-CCA*
trnP-UGG
trnP-UGG
trnT-UGU*
psaJ
ycf4
cemA
petA
accD
rbcL
trnM-CAU
atpE
atpB
trnC-ACA*
ndhJ
ndhK
ndhC
trnL-UAA*
trnF-GAA*
trnF-GAA*
trnS-GGA
rps4
trnT-UGU*
ycf3*
psaB
psaA
trnM-CAU
rps14
trnS-UGA
psbC
psbD
psbZ
trnG-GCC
trnG-GCC
trnY-GUA
trnD-GUC
trnE-UUC*
psbM
trnT-GGU
petN
trnC-GCA
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

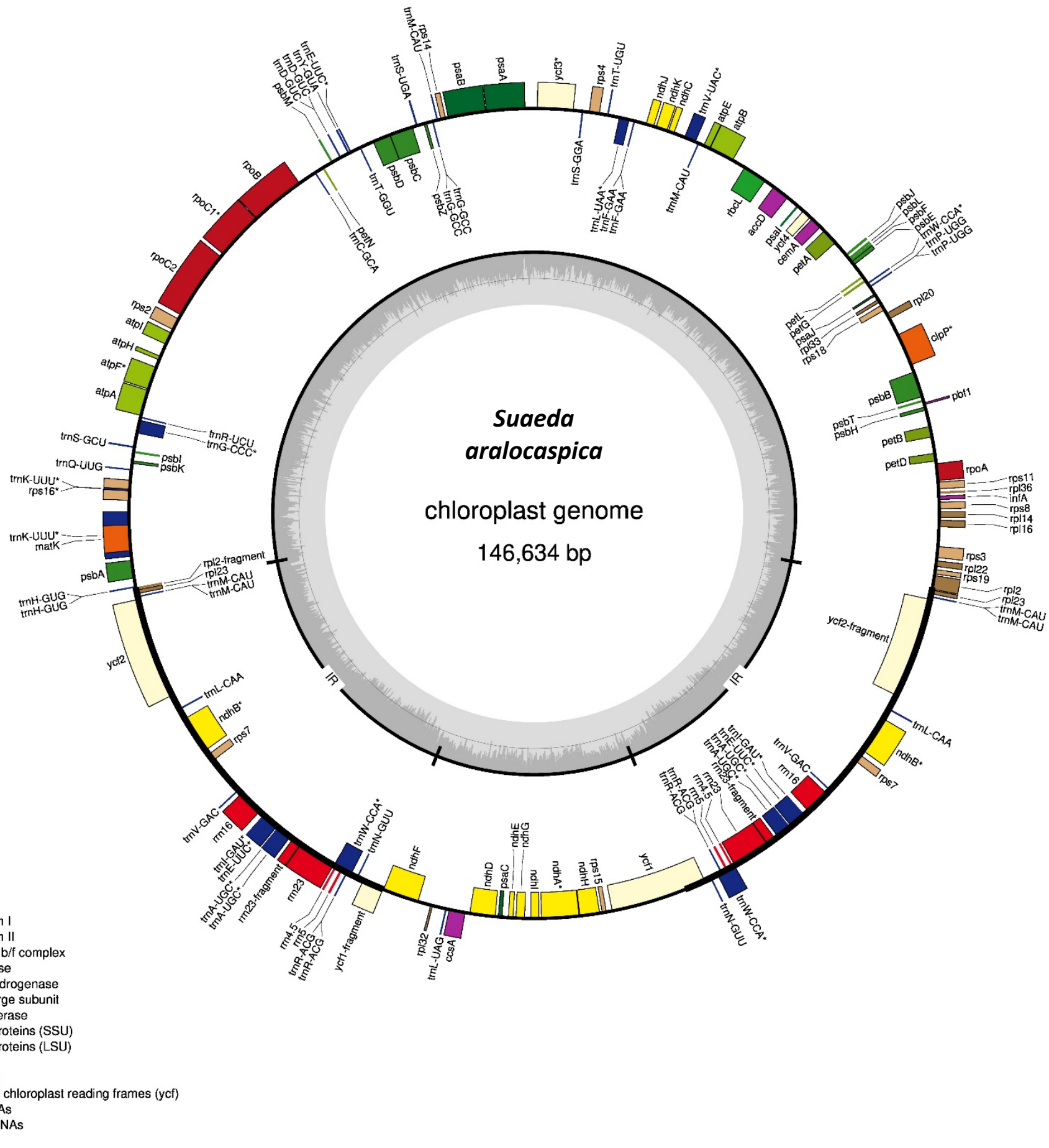
D


Bienertia sinuspersici
chloroplast genome
153,334 bp
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

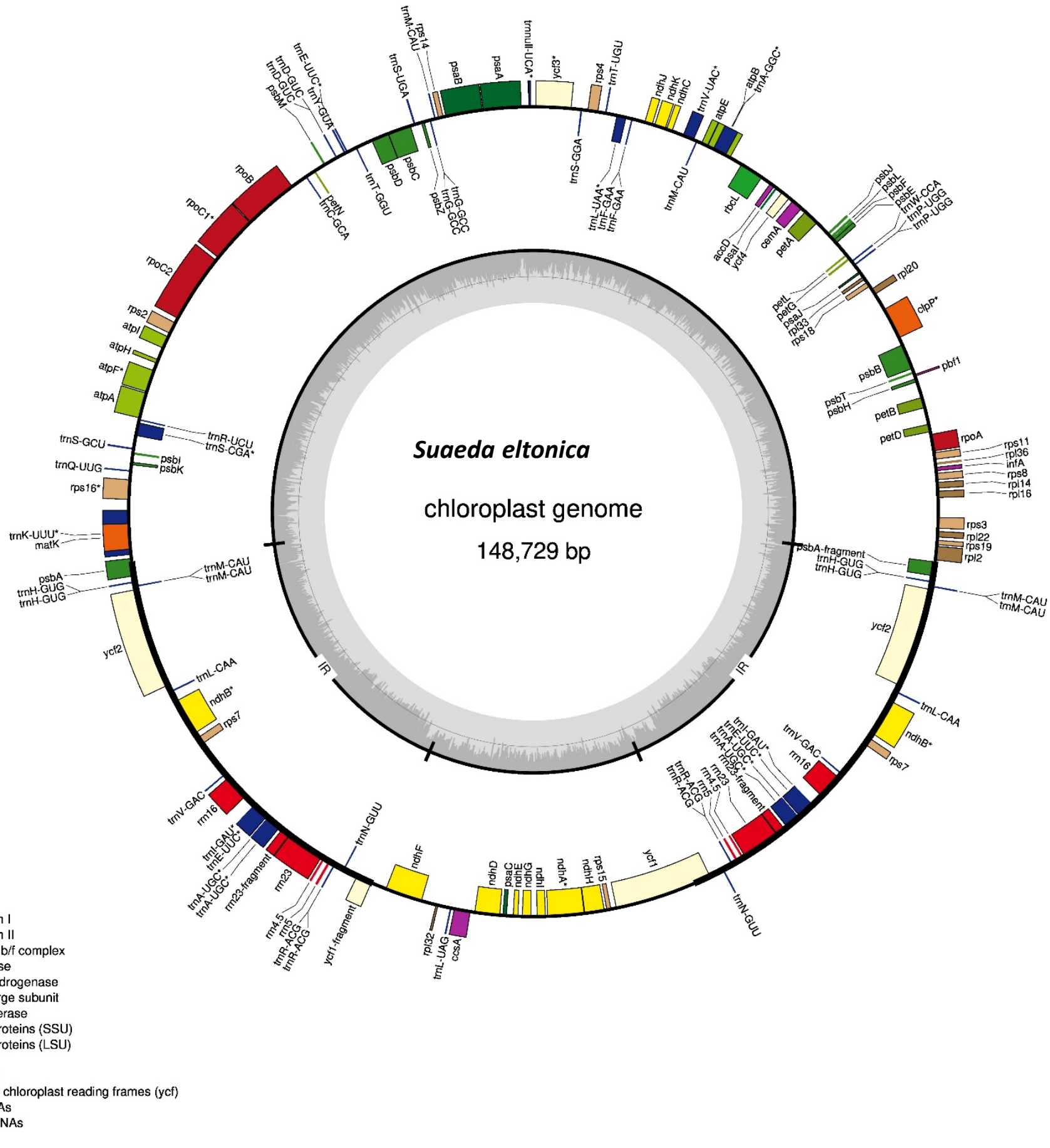
E


Haloxylon ammodendron
chloroplast genome
161,251 bp
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

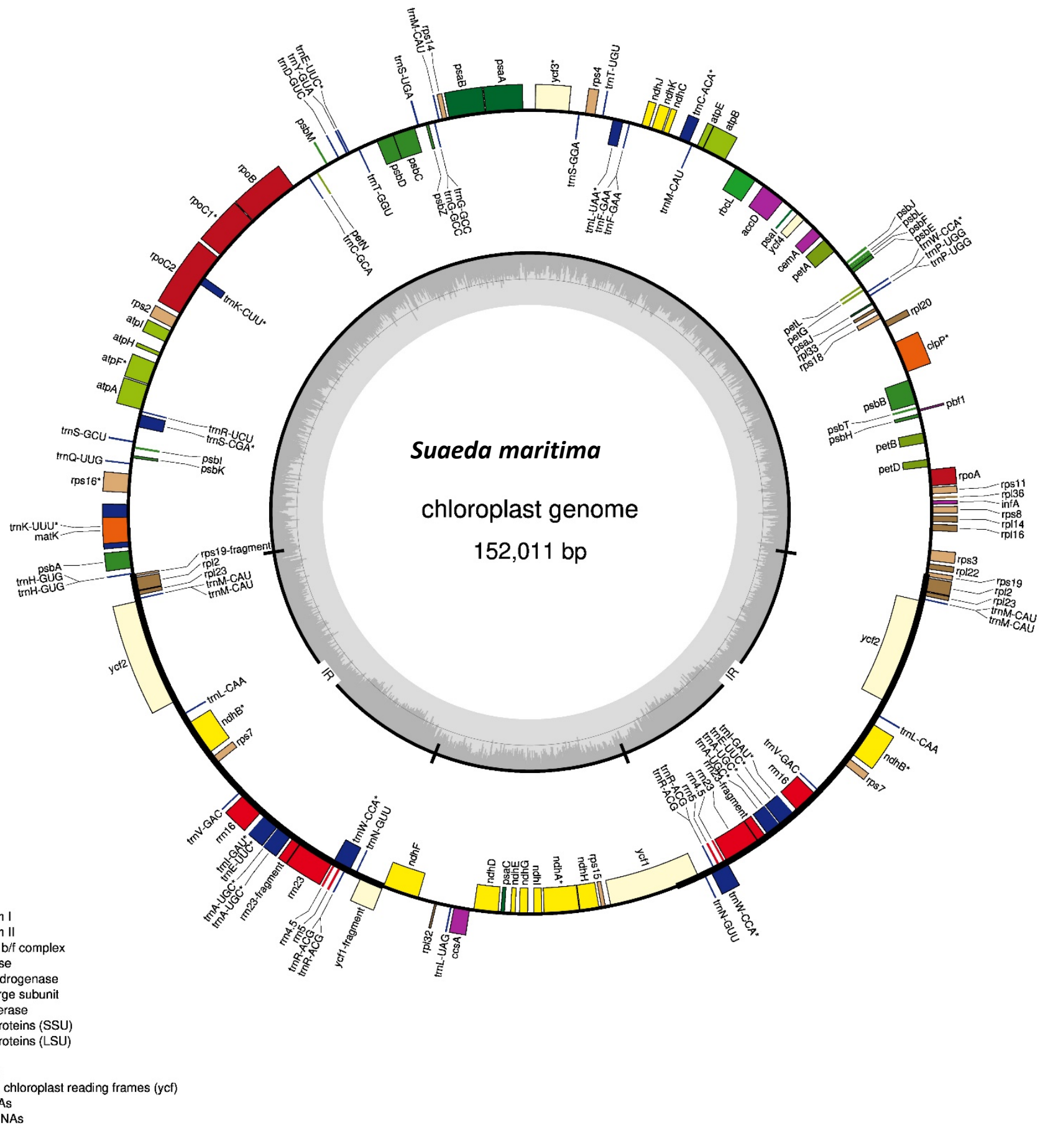
F


Suaeda aralocaspica
chloroplast genome
146,634 bp
IR
IR
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

G


Suaeda eltonica
chloroplast genome
148,729 bp
IR
IR
rpoB
rpoC1*
rpoC2
rps2
atpI
atpH
atpF*
atpA
trnR-UCU
trnS-CGA*
psbI
psbK
trnS-GCU
trnQ-UUG
rps16*
trnK-UUU*
matK
psbA
trnH-GUG
trnH-GUG
trnM-CAU
trnM-CAU
ycf2
trnL-CAA
ndhB*
rps7
trnV-GAC
rrn16
trnI-GAU*
trnE-UUC*
trnA-UGC*
trnA-UGC*
rrn23-fragment
rrn23
rrn4.5
rrn5
trnR-ACG
trnR-ACG
trnN-GUU
ycf1-fragment
ndhF
rpl32
trnL-UAG
ccsA
ndhD
psaC
ndhE
ndhG
ndhI
ndhA*
ndhH
rps15
ycf1
trnN-GUU
trnR-ACG
trnR-ACG
rrn4.5
rrn5
rrn23
rrn23-fragment
trnI-GAU*
trnE-UUC*
trnA-UGC*
trnA-UGC*
trnV-GAC
rrn16
ndhB*
rps7
trnL-CAA
ycf2
trnM-CAU
trnM-CAU
psbA-fragment
trnH-GUG
trnH-GUG
rps3
rpl22
rps19
rpl2
rpoA
rps11
rpl36
infA
rps8
rpl14
rpl16
psbB
pbf1
psbT
psbH
petB
petD
rpl20
clpP*
psbJ
psbL
psbF
psbE
trnM-CCA
trnP-UGG
trnP-UGG
petL
petG
psaJ
rpl33
rps18
rbcL
accD
psaI
ycf4
cemA
petA
trnM-CAU
trnL-UAA*
trnF-GAA
trnF-GAA
trnS-GGA
rps4
trnT-UGU
ndhJ
ndhK
ndhC
trnV-UAC*
atpE
atpB
trnA-GGC*
ycf3*
trnnull-UCA*
psaA
psaB
rps14
trnM-CAU
trnS-UGA
psbD
psbC
trnT-GGU
psbZ
trnG-GCC
trnG-GCC
petN
trnC-GCA
psbM
trnD-GUC
trnD-GUC
trnE-UUC*
trnY-GUA
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

H


Suaeda maritima
chloroplast genome
152,011 bp
IR
IR
rpoB
rpoC1*
rpoC2
rps2
atpI
atpH
atpF*
atpA
trnR-UCU
trnS-CGA*
psbI
psbK
trnS-GCU
trnQ-UUG
rps16*
trnK-UUU*
matK
psbA
trnH-GUG
trnH-GUG
rps19-fragment
rpl2
rpl23
trnM-CAU
trnM-CAU
ycf2
trnL-CAA
ndhB*
rps7
trnV-GAC
rrn16
trnI-GAU*
trnE-UUC*
trnA-UGC*
trnA-UGC*
rrn23-fragment
rrn23
rrn4.5
rrn5
trnR-ACG
trnR-ACG
ycf1-fragment
ndhF
trnW-CCA*
trnN-GUU
rpl32
trnL-UAG
ccsA
ndhD
psaC
ndhE
ndhG
ndhI
ndhA*
ndhH
rps15
ycf1
trnN-GUU
trnW-CCA*
trnV-GAC
rrn16
trnI-GAU*
trnE-UUC*
trnA-UGC*
trnA-UGC*
rrn23-fragment
rrn23
rrn4.5
rrn5
trnR-ACG
trnR-ACG
ycf2
trnL-CAA
ndhB*
rps7
rps19
rpl2
rpl23
trnM-CAU
trnM-CAU
rps3
rpl22
rps11
rpl36
infA
rps8
rpl14
rpl16
rpoA
petD
petB
psbB
psbT
psbH
pbf1
clpP*
rpl20
petL
petG
psaJ
rpl33
rps18
psbJ
psbL
psbF
psbE
trnW-CCA*
trnP-UGG
trnP-UGG
cemA
petA
psaI
ycf4
accD
rbcL
trnM-CAU
atpE
atpB
ndhJ
ndhK
ndhC
trnC-ACA*
trnT-UGU
rps4
ycf3*
trnS-GGA
trnL-UAA*
trnF-GAA
trnF-GAA
psaB
psaA
trnS-UGA
rps14
trnM-CAU
trnE-UUC*
trnY-GUA
trnD-GUC
psbM
petN
trnC-GCA
trnT-GGU
psbC
psbD
psbZ
trnG-GCC
trnG-GCC
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs