MDPI

Article

The Chloroplast Genome of the Lichen Photobiont *Trebouxiophyceae* sp. DW1 and Its Phylogenetic Implications

Lidan Wang [1], Shenglu Zhang [2], Jinjin Fang [1], Xinjie Jin [3], Reyim Mamut [1,*] and Pan Li [2,*]

1 College of Life Sciences and Technology, Xinjiang University, Urumchi 830046, China
2 College of Life Sciences, Zhejiang University, Hangzhou 310058, China
3 College of Life and Environmental Science, Wenzhou University, Wenzhou 325035, China
* Correspondence: reyim_mamut@xju.edu.cn (R.M.); panli_zju@126.com (P.L.)

Abstract: Lichens are symbiotic associations of algae and fungi. The genetic mechanism of the symbiosis of lichens and the influence of symbiosis on the size and composition of the genomes of symbiotic algae have always been intriguing scientific questions explored by lichenologists. However, there were limited data on lichen genomes. Therefore, we isolated and purified a lichen symbiotic alga to obtain a single strain (*Trebouxiophyceae* sp. DW1), and then obtained its chloroplast genome information by next-generation sequencing (NGS). The chloroplast genome is 129,447 bp in length, and the GC content is 35.2%. Repetitive sequences with the length of 30–35 bp account for 1.27% of the total chloroplast genome. The simple sequence repeats are all mononucleotide repeats. Codon usage analysis showed that the genome tended to use codon ending in A/U. By comparing the length of different regions of Trebouxiophyceae genomes, we found that the changes in the length of exons, introns, and intergenic sequences affect the size of genomes. Trebouxiophyceae had an unstable chloroplast genome structure, with IRs repeatedly losing during evolution. Phylogenetic analysis showed that Trebouxiophyceae is paraphyletic, and *Trebouxiophyceae* sp. DW1 is sister to the clade of *Koliella longiseta* and *Pabia signiensis*.

Keywords: algae; genome size; *Peltigera*; Prasiolales; phylogeny

1. Introduction

Green plants, which include green algae and embryonic plants, are one of the most important primary producers on Earth [1]. Green algae are ubiquitous in the world's marine and freshwater ecosystems and play an important role in adapting to changing environmental conditions [2]. Green algae can become symbiotic with other species such as animals, bacteria, fungi, and plants to achieve mutualism [3–5]. It is estimated that green algae originated more than 1.8 billion years ago and split early into two lineages: the Charophyta and the Chlorophyta [6–8]. In Chlorophyta, most taxonomic advances have been made at the level of species and genera, but there are few studies above this taxonomic level and there is often a lack of consensus [9–11]. In the early days, Chlorophyta was traditionally divided into four groups, Ulvophyceae, Trebouxiophyceae, Chlorophyceae and Prasinophyceae, according to the characteristics of flagellar apparatus configuration and cell division process [12,13]. Later on, based on molecular data, morphology and ecological diversity, the core Chlorophyta was divided into three major classes, Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC), plus two smaller lineages, Chlorodendrophyceae and Pedinophyceae [14–16]. Among the three major classes, Chlorophyceae has five orders (Oedogoniales, Chaetophorales, Chaetopeltidales, Volvocales, and Sphaeropleales), and their phylogeny are well-understood. In contrast, the phylogenic relationships within Ulvophyceae and especially Trebouxiophyceae remain unresolved [17]. Trebouxiophyceae are widely distributed throughout the world. It was thought to have seven major lineages [18]; however, as more and more new species are discovered, it is now shown that

Trebouxiophyceae has as many as 16 distinct lineages [1,19,20]. Phylogenetic establishment of Trebouxiophyceae by 18s RNA and *rbc*L, however, is difficult to solve with limited data and a few gene loci [20,21]. Of the 40 or so green algae found in Trebouxiophyceae, 22 are known to form lichen clusters [22]. With the wide application of genome in phylogeny, organelle genomes have been widely used for phylogenetic construction, and it has become a common practice for the phylogenetic analysis of the entire genome (phylogenomics) [23]. From algae to higher plants, chloroplast genomes have been used to study phylogeny, providing researchers with new ideas and insights. Genome-level analysis has led to profound changes in our understanding of the evolution of green algae and has greatly improved our understanding of the deepest relationships in the phylogeny of green algae [24–27].

The chloroplast genome data from Trebouxiophyceae are very limited compared with higher plants, and its chloroplast genome has a highly variable structure [28]. Usually, the chloroplast genome is made up of four parts: the large single copy region (LSC), the small single copy region (SSC), and the two reverse repeats (IR_A and IR_B). However, some algal genomes show different characteristics, including the deletion/shrinkage–disapperance/expansion of IR regions, gene rearrangement, and large intergenic spacers [29]. In previous studies, only one chloroplast genome was reported from a *Trebouxia* lichen phycobiont [30], which made it difficult to further study the structure of symbiotic algae.

In this study, symbiotic alga was isolated from *Peltigera rufescens* and then sequenced by high-throughput sequencing for the first time. The complete chloroplast genome sequence of *Trebouxiophyceae* sp. DW1 was reported, and its size, structure and gene content were compared with other published genomes of green algae. Phylogenetic analysis of Trebouxiophyceae was conducted based on 36 chloroplast genes, including 20 associated with photosynthesis, 14 ribosomal genes, and also *tuf*A and *ycf*3 genes.

2. Materials and Methods

2.1. Phycobiont Isolation and Culture Conditions

Trebouxiophyceae sp. DW1 was isolated from the lichen *P. rufescens* (Figure 1), the sample of *P. rufescens* was collected from Bayi Forest Farm, in Xinjiang Province. This voucher specimen was deposited in the Herbarium of College of Life Science and Technology at Xinjiang University in Urumchi, China, under the voucher number BY201830, by grinding the sterilized lichens and culturing it in BG11 medium [31] in a growth chamber at 22 °C under a 10:14 h light:dark cycle.

Figure 1. Morphological structure of *P. rufescens*.

2.2. DNA Extraction, Sequencing and Raw Data Preprocessing

DNA was extracted using the CTAB method [32]. The sequencing was performed in Hiseq PE150 mode (pair-end sequencing). Before the quality control, the original data were evaluated with BBTools [33] statistical information, and some basic information was counted and visualized to determine the data quality. In order to improve the quality and reliability of the follow-up data analyses, we carried out quality control on the original data. Using BBduk [33], to remove possible joint sequences in reading sequence; quality trimming: sliding window method to remove continuous low-quality sequences (Q value < 20) at both ends of reading sequence; to remove reading sequence and its paired reading sequence with length less than 35 bp after quality control. To evaluate the pollution of reading sequences we compared BLAST+, using 10,000 sequences randomly selected from QC (quality control). The comparison database was NCBI NT. The results of comparison were evaluated with e value $\leq 1 \times 10^{-10}$ and similarity > 90%, coverage > 80%. This step of analysis is to determine whether the sample is contaminated during the preparation and sequencing of the library. Possible contamination may include microbial contamination of the environment, and contamination of human genetic material that may be introduced into the laboratory operation, etc.

2.3. Genome Assembly

SPAdes (http://bioinf.spbau.ru/SPAdes/) (accessed on 11 September 2020) was first used to correct the sequence errors of original sequence and then to assemble by multiple Kmer values. Finally, the best results were obtained by synthesizing the Kmer values (K-mer = 39). Therefore, the assembly effect and accuracy of SPAdes were high. Gapcloser v1.12 [34] and GapFiller v2.1.2 [35], the gap-filling tools, use a large number of pairs of short sequences to complete the gap-filling process.

2.4. Gene Prediction

The chloroplast genomes were annotated by GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) (accessed on 27 January 2021) [36] and then the wrong genes were manually corrected by Sequin software, which manually adjusts the position of start and stop codons. Lastly, the results were placed in OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) (accessed on 23 March 2021) [36] to generate the cpDNA map.

2.5. Codon Usage

With the exception of Met and Trp, amino-acid residues are encoded by two or more synonyms, and genes have a tendency to use synonyms nonrandomly to encode amino acids, which is known as codon usage bias. Codon usage bias analysis could provide clues for revealing the law of genetic evolution [37]. Therefore, we analyzed codon usage bias of *Trebouxiophyceae* sp. DW1 in Mega v7.0.26 [38].

2.6. Identification of Repeat and Simple Sequence Repeat

Repeat sequences were detected with the online version of REPuter [39]. We included forward and palindromic repeat sequences with a minimum repeat size of 30 bp. Meanwhile, online Tandem Repeats Finder was used to search for Tandem Repeats in chloroplast DNA sequences [40]. Simple sequence repeats (SSR) were performed by using MISA (https://webblast.ipk-gatersleben.de/misa/) (accessed on 19 December 2020) for prediction, and the parameter was set as follows: the minimum SSR sequence segment length of 10 bp [41].

2.7. Phylogenetic Analysis

Thirty-six chloroplast genes (atpA, atpB, atpE, petA, petB, psaA, psaB, psbA, psbB, psbC, psbD, psbE, psbH, psbJ, psbK, psbL, psbN, psbT, psbZ, rbcL, rpl2, rpl5, rpl14, rpl16, rpl20, rpl23, rps2, rps3, rps7, rps8, rps11, rps12, rps18, rps19, tufA and ycf3) from 22 species of green algae were downloaded from GenBank. These sequences, together with those of

Trebouxiophyceae sp. DW1, were used for phylogenetic analysis (Table 1). The sequences were aligned with MAFFT version 7 [42] and trimmed with Gblocks version 0.91b [43]. Maximum-likelihood (ML) method was used to construct phylogenetic tree by RAxML v7.2.8 [44], bootstrap repeats were 1×10^6 generations, and the best model was selected with ModelFinder [45]. FigTree v3.2 was used to exhibit the consensus tree.

Table 1. Species and GenBank accession number used for phylogenetic analysis.

Species	GenBank No.
Botryococcuus braunii	*NC_025545*
Coccomyxa sp.	*NC_015084*
Paradoxia multiseta	*NC_025540*
Trebouxiophyceae sp.	*NC_018569*
Elliptochloris bilobata	*NC_025548*
K. longiseta	*NC_025531*
P. signiensis	*NC_025529*
Chlorella vulgaris	*NC_001565*
Trebouxiophyceae sp. DW1	***MW_255987***
Myrmecia israelensis	*NC_025525*
Trebouxia sp. TR9	*MK_643158*
Chlorella sorokiniana	*NC_023835*
Chlorella variabilis	*NC_015359*
Micractinium conductrix	*NC_036806*
Dicloster acuatus	*NC_025546*
Parachlorella kessleri	*NC_012978*
Pedinomonas minor	*NC_016733*
Pedinomonas tuberculata	*NC_025530*
Marsupiomonas sp.	*KM_462870*
Nephroselmis astigmatica	*NC_024829*
Nephroselmis olivacea	*NC_000927*
Chara vulgaris	*NC_008097*
Chlorokybus atmophyticus	*NC_008822*

Note: The new sequences generated in this study are indicated in boldface.

3. Results

3.1. Structural Characteristics of Chloroplast Genome

The length of *Trebouxiophyceae* sp. DW1 chloroplast genome is 129,447 bp, of which the LSC is 98,101 bp, the SSC is 16,602 bp, and the two inverted repeats (IR_A and IR_B) are 7372 bp (Figure 2). The chloroplast genome of *Trebouxiophyceae* sp. DW1 contains 103 genes (Table S1), including 67 protein-coding genes, 32 transport RNAs (tRNAs), and 2 ribosomal RNAs (rRNAs) with exons and intergenic sequences occupy a large proportion (Figure 3). *Rpo*C2 has two introns, while *rpo*B, *rpo*C1, and *trn*L(uaa) each have one intron. The length of the introns ranged from 404 to 1574 bp. The IRs region contains one protein-coding gene *pet*D, two tRNAs and two rRNAs, among which *rrn*23 is a pseudogene.
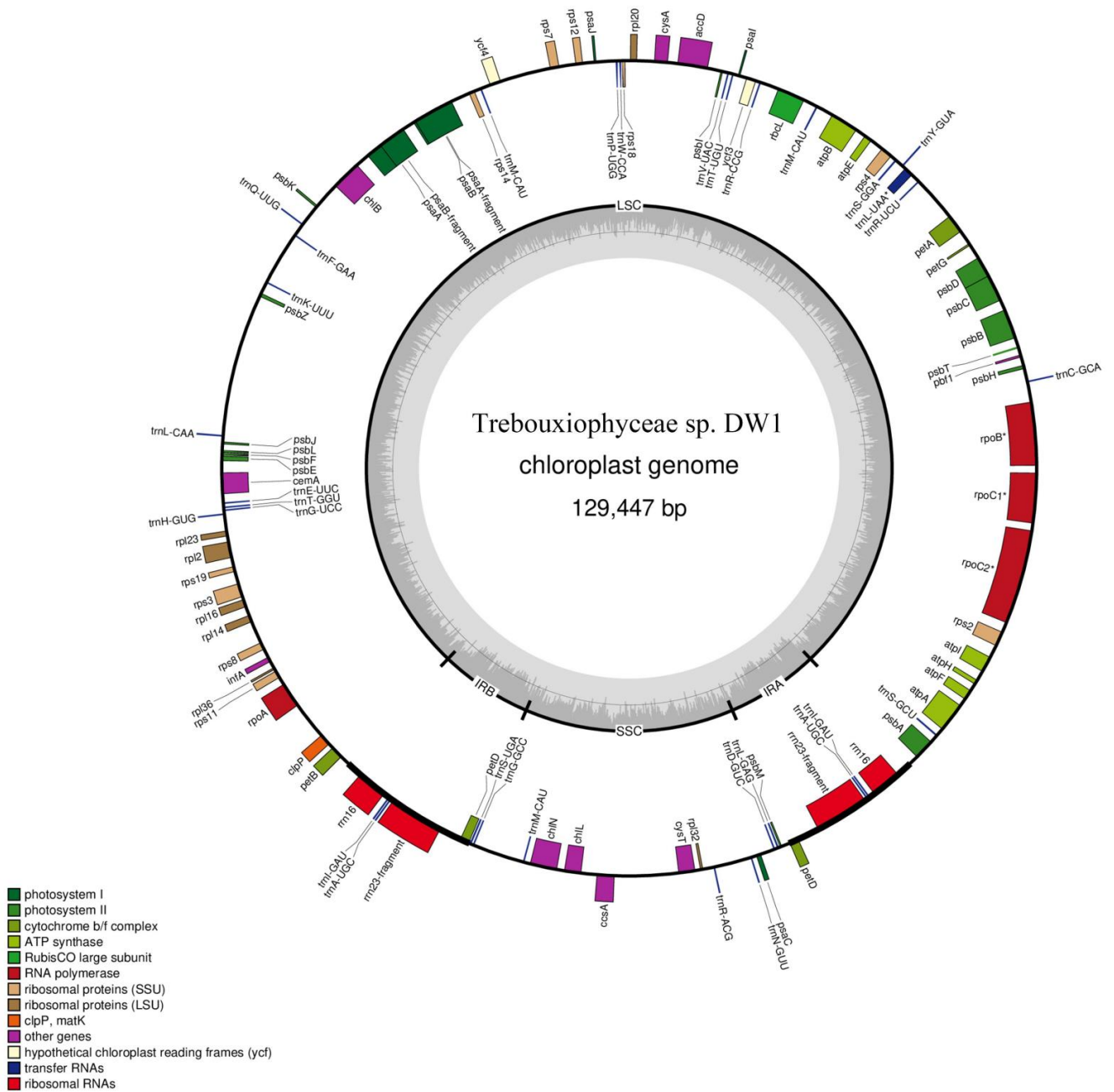
Figure 2. Gene map of the complete chloroplast genome of the green algae. *Trebouxiophyceae* sp. DW1. The genes inside the circle are transcribed clockwise, and the genes outside the circle are transcribed counterclockwise. Genes with introns are marked with an asterisk. The dark gray and light gray of the outer circle correspond to the LSC and SSC regions, respectively. The two regions are, respectively, represented as IRA and IRB, between LSC and SSC regions.
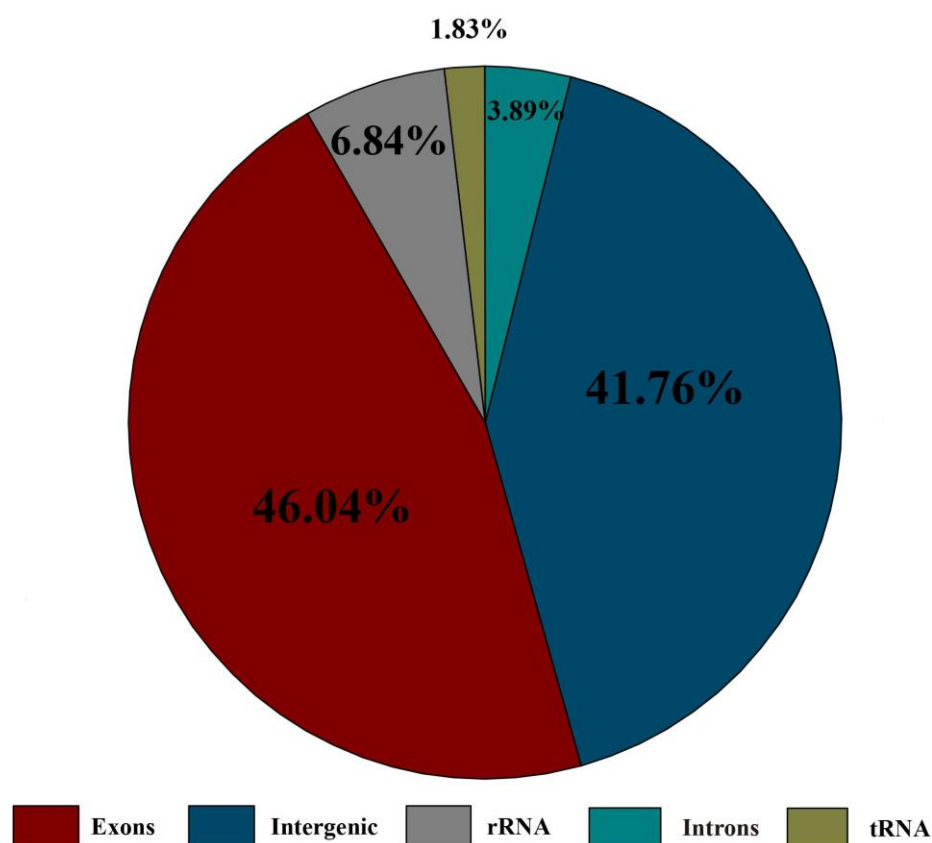
Figure 3. Proportions of the whole chloroplast genome of *Trebouxiophyceae* sp. DW1. The exons, introns, intergenic sequences, tRNA, and rRNA genes region.

3.2. Repeat Elements in the Trebouxiophyceae sp. DW1 Chloroplast Genome

A total of 49 repetitive sequences were detected in the chloroplast genome of *Trebouxiophyceae* sp. DW1, including 30 palindromic and 19 forward (Table 2). The size of these repetitive sequences ranged from 30 bp to 50 bp. The longest repetitive sequences were located between the *rps*7 and *ycf*4 genes. A total of 1645 bp repetitive sequences were detected in the *Trebouxiophyceae* sp. DW1 chloroplast genome, which accounted for 1.27% of the whole chloroplast genome.

Table 2. Repetitive sequences detected in the *Trebouxiophyceae* sp. DW1 chloroplast genome using the REPuter.

Number of Bases.	Type	Number
30–35	P	20
	F	14
36–40	P	7
	F	4
41–45	P	2
	F	1
46–50	P	1
	F	0

Note: forward repeats were denoted by F; palindromic repeats were denoted by P.

Only one kind of SSRs, the mononucleotide repeats, was detected for the *Trebouxiophyceae* sp. DW1 chloroplast genome. We counted all SSRs with lengths greater than 10 bp. In the chloroplast genome of *Trebouxiophyceae* sp. DW1 (Table S2), the main single nucleotide (p1) was A/T (about 94.74%), and no C/G was detected. Similar to other plant chloroplast genomes reported previously, simple repetitive sequences were mainly by poly

thymine nucleotides (polyT) or poly adenine nucleotides (polyA), which rarely appear in a series of cytosine and guanine repeats.

　　We also performed tandem repeats analysis in the chloroplast genome of *Trebouxiophyceae* sp. DW1 (Table 3), which accounted for 4.64% of the whole chloroplast genome. The longest tandem sequence in the *Trebouxiophyceae* sp. DW1 chloroplast genome is 77 bp, which is located in the intergenic region between neighboring genes *pet*G and *pet*A. The highest copy numbers detected were 2.1 and 2.8 in the *Trebouxiophyceae* sp. DW1 chloroplast genome.

Table 3. Tandem repeats detected in the *Trebouxiophyceae* sp. DW1 chloroplast genome using the Tandem Repeats Finder.

Indices	Period Size	Copy Number	Consensus Size	Percent Matches	Percent Indels	Score	A	C	G	T	Entropy (0–2)
28,773–28,933	77	2.1	77	87	5	227	37	14	9	37	1.79
106,572–106,604	12	2.8	12	100	0	66	33	18	9	39	1.82

3.3. Codon Usage Analysis

　　Codon usage analysis indicated that TTT (for phenylalanine; Phe), AAA (for lysine; Lys), ATT (for isoleucine; Ile), CAA (for glutamine; Gln), and TTA/TTG (for leucine; Leu) in the chloroplast genome of *Trebouxiophyceae* sp. DW1 were used more frequently (Figure 4). The genomic codons have 32 codons with relative synonymous codon usage (U_{RSC}) > 1, of which 28 codons ended in A/U, indicating that the chloroplast genome prefers to use codons ending in A/U.
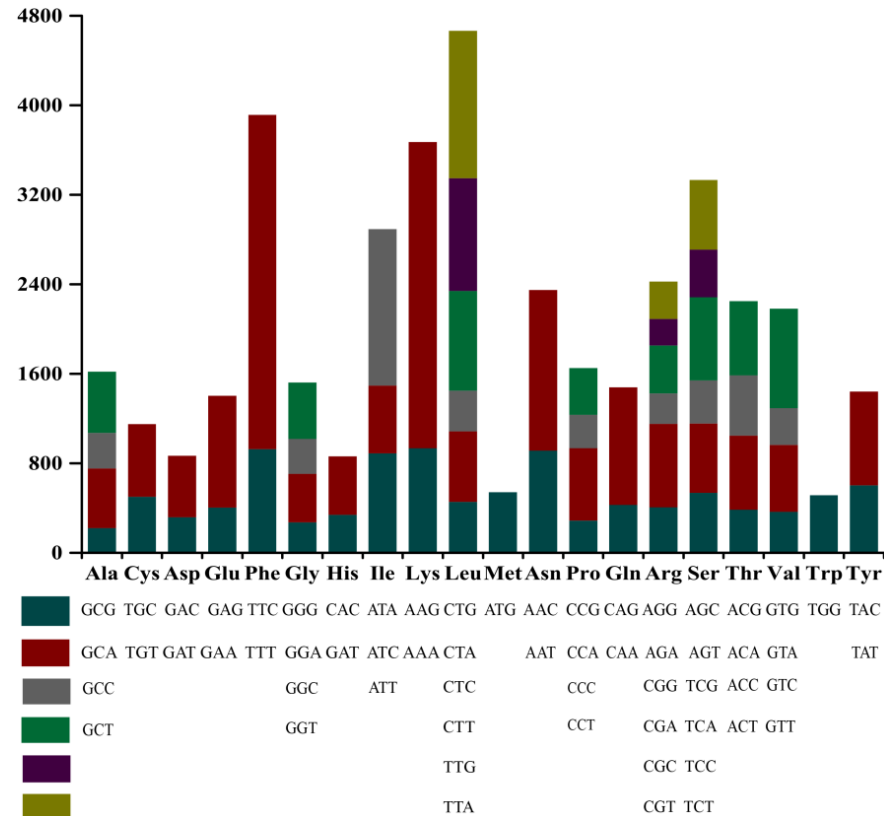

Figure 4. Codon usage in the whole chloroplast genome of *Trebouxiophyceae* sp. DW1. The *x*-axis represents amino acids, frequency of codon usage is plotted on the *y*-axis.

We calculated the GC content at positions 1, 2, and 3 of the codons and recorded them as GC_1, GC_2, and GC_3, respectively. GC_1 is 35.25%, GC_2 is 35.20%, and GC_3 is 34.26%. The value of GC_3 is slightly lower than GC_1 and GC_2. The high frequency of A and T in the codons contributes to the high AT content of the *Trebouxiophyceae* sp. DW1 chloroplast genome (64.80%).

3.4. Phylogenetic Analysis

Phylogeny of Trebouxiophyceae and relatives was reconstructed based on an alignment of 23,288 nt (the DNA sequences of 36 protein-coding genes), with two Streptophyta (*C. atmophyticus* and *C. vulgaris*) as outgroups (Figure 5). Trebouxiophyceae is paraphyletic, with Chlorellales more closely related with Pedinophyceae (BS = 93). *Trebouxiophyceae* sp. DW1 is sister to the clade of *K. longiseta* and *P. signiensis*, with high support (BS = 90).
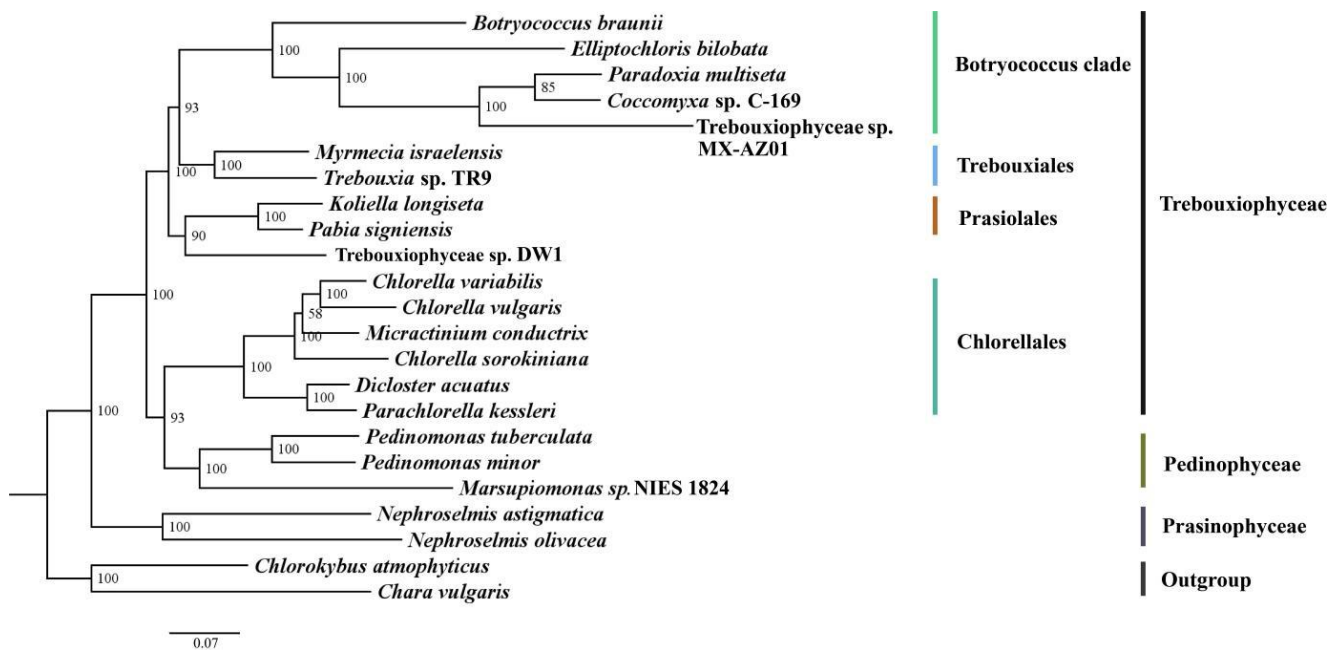

Figure 5. Maximum-likelihood (ML) phylogeny of 21 Trebouxiophyceae and Prasinophyceae species based on an alignment of 23,288 nt of 36 protein-coding genes. Bootstrap values are indicated in the nodes. The scale bar indicates substitutions/site.

4. Discussion

4.1. Features That Affect the Size of the Chloroplast Genome

To date, 153 chloroplast genomes of Chlorophyta have been published in NCBI. Among them, *Haematococcus lacustris* UTEX 2505 has the largest genome, with a length of 1,352,306 bp [46], while *Prototheca zopfii* strain SAG 2021 has the smallest genome, with a length of 28,638 bp [47]. The size of chloroplast genomes in Chlorophyta varies greatly. We compared proportions of exons, introns, and intergenic sequences and the proportions of IRs, LSC and SSC regions in some algae chloroplast genomes (Figure 6). *Trebouxia* sp. TR9 and *Prasiolosis* sp. SAG 84.81 have a larger size than others; the proportions of their exons, intergenic spacers and introns in the genomes were similar. The difference is that the latter lacks the complete chloroplast four-part chloroplast structure. *Trebouxiophyceae* sp. DW1 has a smaller genome. By comparing the genomic data, we found that the reasons for the differences in genome length variation are mainly as follows: (1) differences in the number and size of genes; (2) the length of the introns; (3) the size of the intergenic spacers; and (4) the expansion/contraction or absence of IR regions. This conclusion is consistent with previous studies [48–50].
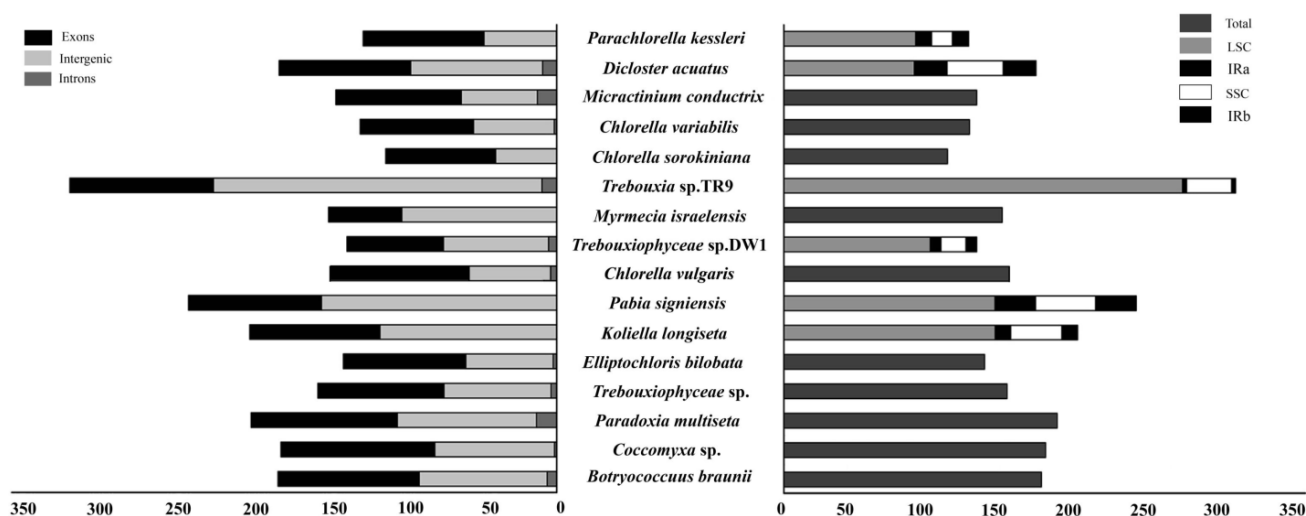
Figure 6. Total lengths of different regions of the cpDNAs in Trebouxiophyceae. The proportions of exons, introns and intergenic sequences in the chloroplast genome are shown on the left, and the proportions of IRS, LSC and SSC regions in the chloroplast genome are shown on the right.

There are two hypotheses about genome size. One hypothesis is the adaptive theory and the other is the junk DNA theory. The adaptive theory says that a very small genome may favor organisms with a short reproductive cycle, but a large genome may be more acceptable if the reproductive cycle is long, and changes in genome size reflect the adaptive needs of different organisms or the effectiveness of natural selection. The junk DNA theory states that extra DNA is indeed superfluous and useless so that maladaptive DNA is fixed by random drift and is passively carried on the chromosomes [51–53]. As the genome size increases, the ratio of insertion-sequence to coding-sequence size increases more rapidly [54,55]. The final genome size is set at a tolerable maximum, depending on the specific ecological and developmental needs of the organism. At present, the research mainly addresses the process of the evolution of genome size, and more experimental data are needed to explore the functional significance of genome size. Both genome expansion and contraction have been proposed as an evolutionary strategy to achieve the optimal balance between genomic stability and plasticity [53]. In *Trebouxiophyceae* sp. DW1, compared with the longer green-algae genome, there are fewer repeats with shorter gene segments, which may be the reason why *Trebouxiophyceae* sp. DW1 has a smaller genome.

4.2. Variations in Chloroplast Genome Structure

Structurally, variations are usually due to the expansion, contraction, or absence of the IRs. IRs are hot spots of genomic instability in prokaryotes and eukaryotes [56]. In eukaryotes, they break dsDNA and make chromosomes brittle, stimulate homologous recombination, and induce total chromosomal rearrangements [57–61]. IRs plays an important role in genetic instability, gene splicing, and replication delay, and there is a tendency to reduce genomic stability in various organisms [62–66].

Trebouxiophyceae sp. DW1 has a complete four-part structure including one SSC, one LSC, and two IRs, while *Coccomyxa* sp., *B. braunii*, *M. israelensis* and other algae lost IRs. The high propensity of IR in size, gene content and gene order, and the repetitive loss is experienced during the evolution of the Trebouxiophyceae [67,68]. Most of the IR-less genomes are rearranged, and the sequence evolves faster. The chloroplast genomes of Trebouxiophyceae show great plasticity [28,69]. Higher plants usually have complete four parts; however, IRs loss has also occurred in a few gymnosperms and angiosperms, such as *Cryptomeria japonica*, *Cephalotaxus wilsoniana*, *Taiwania cryptomerioides*, *Carnegiea gigantea* and *Wisteria* [70–73]. Longer IRs generally increase the stability of the genome, whereas the shrinkage and loss of IRs generally increase the likelihood of gene rearrangement and gene loss, affecting genome size [74,75]. The IRs of algae were smaller than higher plants, which

tend to gradually degenerate. Perhaps IRs were not a necessary part of the chloroplast energy of algae.

4.3. Repeat Elements: Significant Mononucleotide Repeats in SSRs

SSRs are widely found in prokaryotes and eukaryotes. They are inevitable and highly variable products of genome replication. They exist in coding and non-coding regions of the genome and play an important role in genome evolution and recombination [76]. They were thought to participate in the gene expression, regulation, and function of components as transcription activation [77].

The 19 SSRs detected in *Trebouxiophyceae* sp. DW1 mainly existed in the non-coding region, and there was only one complex polynucleotide, and the rest were all mononucleotide. The most significant repetition pattern in the Chlorophyta chloroplast genome was mononucleotide poly A/T repetition [78], which was consistent with our results. The composition of SSRs varies greatly among different species. It was thought in the past that plants were mainly composed of trinucleotide repeats; differently, in angiosperms mononucleotide repeats were more prominent, followed by dinucleotide repeats, such as *Panax ginseng*, *Elodea canadensis*, *Helianthus annuus*, and *Olea europaea* [79,80]. *Physcomitrella* consisted mainly of dimer repeats [79]. In Brassicaceae, trinucleotide repeats were more common in the coding region, while other repeats were more common in the non-coding DNA [81]. Although the repetition types varied among different species, similarly, the number of SSRs in non-coding regions was usually greater than the number of coding regions.

4.4. Codon Usage

Most amino acids, except methionine and tryptophan, are encoded by 2–6 different codons [82]. Different species also use different synonymous codon. The main reason for this result is mutation and/or natural selection [83]. Although research on genetic mechanisms are limited, the selection of different codon can improve elongation and/or translation accuracy [84,85]. It has been found to affect protein expression, structure and function, as well as translation elongation [86].

The GC content of the third base of codon (GC_3) is an important part of the GC content of genotype. GC content is an important indicator of the base composition of an organism's genome. As the selection pressure on the third position of the codon is less than that on the first two, GC_3 is often used as an indicator to measure the preference of codons. Dicotyledons tend to use A/U, while monocotyledons tend to use C/G [87,88]. The results showed that A/U was the main encoding with high frequency in *Trebouxiophyceae* sp. DW1.

4.5. Chloroplast Phylogenomics

Chloroplast phylogenomics has become an effective way to elucidate the mysterious evolutionary relationships at different taxonomic levels of plants [89]. The phylogenetic relationship of Chlorophyta, especially at higher taxonomic levels (order, class), has been a topic of debate. Therefore, we need to analyze more Trebouxiophyceae chloroplast genomes in order to reveal the deeper relationships of Trebouxiophyceae. There has been considerable debate regarding whether Trebouxiophyceae is a monophyly. Several studies showed that the Trebouxiophyceae is monophyletic [90,91], while others suggested that Trebouxiophyceae is paraphyletic [26,30,92]. Here, we used 36 protein-coding genes to reconstruct a more reliable phylogeny, which showed that Trebouxiophyceae is paraphyletic.

Author Contributions: Conceived and designed the experiments: R.M. and P.L. Performed the experiments: L.W., S.Z., J.F., X.J. and R.M. Sequenced the data: L.W. and X.J. Analyzed the data: L.W.,

S.Z., J.F. and P.L. Wrote the paper: L.W., R.M. and P.L. Provided guidance on the whole study: P.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The *Trebouxiophyceae* sp. DW1 isolated from the lichen *P. rufescens* are available from the corresponding author (reyim_mamut@xju.edu.cn) upon reasonable request. The lichen materials are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Leliaert, F.; Smith, D.R.; Moreau, H.; Herron, M.D.; Verbruggen, H.; Delwiche, C.F.; De Clerck, O. Phylogeny and Molecular Evolution of the Green Algae. *Crit. Rev. Plant. Sci.* **2012**, *31*, 1–46. [CrossRef]
2. Brodie, J.; Chan, C.X.; Clerck, O.D.; Cock, J.M.; Bhattacharya, D. The algal revolution. *Trends Plant Sci.* **2017**, *22*, 726–738. [CrossRef] [PubMed]
3. Muscatine, L.; Trench, R.R.P.K. Symbiosis of algae and invertebrates: Aspects of the symbiont surface and the host-symbiont interface. *Trans. Am. Microsc. Soc.* **1975**, *94*, 450–469. [CrossRef]
4. Wooldridge, S.A. Is the coral-algae symbiosis really 'mutually beneficial' for the partners? *Bioessays* **2010**, *32*, 615–625. [CrossRef] [PubMed]
5. Kouzuma, A.; Watanabe, K. Exploring the potential of algae/bacteria interactions. *Curr. Opin. Biotechnol.* **2015**, *33*, 125–129. [CrossRef]
6. Lewis, L.A.; McCourt, R.M. Green algae and the origin of land plants. *Am. J. Bot.* **2004**, *91*, 1535–1556. [CrossRef]
7. Leliaert, F.; Verbruggen, H.; Zechman, F.W. Into the deep: New discoveries at the base of the green plant phylogeny. *Bioessays* **2011**, *33*, 683–692. [CrossRef]
8. Moczydlowska, M.; Landing, E.D.; Zang, W.; Palacios, T. Proterozoic phytoplankton and timing of chlorophyte algae origins. *Palaeontology* **2011**, *54*, 721–733. [CrossRef]
9. Boedeker, C.; O' Kelly, C.J.; Star, W.; Leliaert, F. Molecular phylogeny and taxonomy of the Aegagropila clade (Cladophorales, Ulvophyceae), including the description of *Aegagropilopsis* gen. nov. and *Pseudocladophora* gen. nov. *J. Phycol.* **2012**, *48*, 808–825. [CrossRef]
10. Carlile, A.; O'Kelly, C.J.; Sherwood, A.R. The green algal genus *Cloniophora* represents a novel lineage in the Ulvales: A proposal for *Cloniophoraceae* fam. nov. *J. Phycol.* **2011**, *47*, 1379–1387. [CrossRef]
11. Fucikova, K.; Leliaert, F.; Cooper, E.D.; Skaloud, P.; Verbruggen, H. New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. *Front. Ecol. Evol.* **2014**, *2*, 63.
12. Mattox, K.R.; Stewart, K.D. Classification of the green algae: A concept based on comparative cytology. In *Systematics of the Green Algae*; Academic Press: London, UK, 1984; Volume 1, pp. 29–72.
13. Sluiman, H.J. The green algal class Ulvophyceae an ultrastructural survey and classification. *Crypt. Bot.* **1989**, *1*, 83–94.
14. Arora, M.; Anil, A.C.; Leliaert, F.; Delany, J.; Mesbahi, E. *Tetraselmis indica* (Chlorodendrophyceae, Chlorophyta), a new species isolated from salt pans in Goa, India. *Eur. J. Phycol.* **2013**, *48*, 61–78. [CrossRef]
15. Norris, R.E.; Hori, T.; Chihara, M. Revision of the genus Tetraselmis (Class Prasinophyceae). *J. Plant. Res.* **1980**, *93*, 317–339.
16. Marin, B. Nested in the Chlorellales or independent class? Phylogeny and classification of the Pedinophyceae (Viridiplantae) revealed by molecular phylogenetic analyses of complete nuclear and plastid-encoded rRNA operons. *Protist* **2012**, *163*, 778–805. [CrossRef] [PubMed]
17. Fucikova, K.; Lewis, P.O.; Lewis, L.A. Putting incertae sedis taxa in their place: A proposal for ten new families and three new genera in Sphaeropleales (Chlorophyceae, Chlorophyta). *J. Phycol.* **2014**, *50*, 14–25. [CrossRef]
18. Fucikova, K.; Lewis, P.O.; Lewis, L.A. Widespread desert affiliations of Trebouxiophycean algae (Trebouxiophyceae, Chlorophyta) including discovery of three new desert genera. *Phycol. Res.* **2014**, *62*, 294–305. [CrossRef]
19. Gaysina, L.; Nemcova, Y.; Skaloud, P.; Sevcikova, T.; Elias, M. *Chloropyrula uraliensis* gen. et sp. nov. (Trebouxiophyceae, Chlorophyta), a new green coccoid alga with a unique ultra structure, isolated from soil in South Urals. *J. Syst. Evol.* **2013**, *51*, 476–484. [CrossRef]
20. Neustupa, J.; Elias, M.; Skaloud, P.; Nemcova, Y.; Sejnohova, L. *Xylochloris irregularis* gen. et sp. nov. (Trebouxiophyceae, Chlorophyta), a novel subaerial coccoid green alga. *Phycologia* **2011**, *50*, 57–66. [CrossRef]
21. Rindi, F.; Mcivor, L.; Sherwood, A.R.; Friedl, T.; Guiry, M.D.; Sheath, R.G. Molecular phylogeny of the green algal order Prasiolales (Trebouxiophyceae, Chlorophyta)1. *J. Phycol.* **2010**, *43*, 811–822. [CrossRef]
22. Lucia, M.; Steven, L.; Eva, B. The hidden diversity of lichenised Trebouxiophyceae (Chlorophyta). *Phycologia* **2018**, *57*, 503–524.

23. Spatafora, J.W.; Bushley, K.E. Phylogenomics and evolution of secondary metabolism in plant-associated fungi. *Curr. Opin. Plant Biol.* **2015**, *26*, 37–44. [CrossRef]

24. Monique, T.; Cambiaire, J.D.; Otis, C.; Lemieux, C. Distinctive architecture of the chloroplast genome in the Chlorodendrophycean green algae *Scherffelia dubia* and *Tetraselmis* sp. CCMP 881. *PLoS ONE* **2016**, *11*, e0148934.

25. Parks, M.; Cronn, R.; Liston, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* **2009**, *7*, 84. [CrossRef] [PubMed]

26. Sun, L.; Ling, F.; Zhang, Z.; Xin, C.; Penny, D.; Zhong, B. Chloroplast phylogenomic inference of green algae relationships. *Sci. Rep.* **2016**, *6*, 20528. [CrossRef]

27. Ruhfel, B.R.; Gitzendanner, M.A.; Soltis, P.S.; Soltis, D.E.; Burleigh, J.G. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* **2014**, *14*, 23. [CrossRef]

28. Cambiaire, J.; Otis, C.; Turmel, M.; Lemieux, C. The chloroplast genome sequence of the green alga *Leptosira terrestris*: Multiple losses of the inverted repeat and extensive genome rearrangements within the Trebouxiophyceae. *BMC Genom.* **2007**, *8*, 213. [CrossRef]

29. Lemieux, C.; Otis, C.; Turmel, M. Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC Evol. Biol.* **2014**, *14*, 211. [CrossRef]

30. Fernando, M.A.; Eva, B.; Leonardo, M.C.; Francisco, G.; Arantzazu, M.; Patricia, M.; Maria, G.H.; Eva, M.C. The chloroplast genome of the lichen-symbiont microalga Trebouxia sp. Tr9 (Trebouxiophyceae, Chlorophyta) shows short inverted repeats with a single gene and loss of the rps4 gene, which is encoded by the nucleus. *J. Phycol.* **2020**, *56*, 170–184.

31. Rippka, R. Photoheterotrophy and chemoheterotrophy among unicellular blue-green algae. *Arch. Microbiol.* **1972**, *87*, 93–98. [CrossRef]

32. Cota-Sánchez, J.H.; Remarchuk, K.; Ubayasena, K. Ready-to-use DNA extracted with a CTAB method adapted for herbarium specimens and mucilaginous plant tissue. *Plant Mol. Biol. Rep.* **2006**, *24*, 161–167. [CrossRef]

33. Khan, M.A.; Bhatia, P.; Sadiq, M. Bbtool: A tool to generate the test cases. *IJRTE* **2012**, *1*, 192–197.

34. Xu, M.; Guo, L.; Gu, S.; Wang, O.; Zhang, Y. Tgs-Gapcloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience* **2020**, *9*, gia094. [CrossRef] [PubMed]

35. Boetzer, M.; Pirovano, W. Toward almost closed genomes with GapFiller. *Genom. Biol.* **2012**, *13*, R56. [CrossRef] [PubMed]

36. Michael, T.; Pascal, L.; Tommaso, P.; Ulbricht-Jones, E.S.; Axel, F.; Ralph, B.; Stephan, G. Geseq-versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *5*, W6–W11.

37. Qin Cai, Z.Q.; Xia, G.M.; Wang, M.C. Synonymous codon usage bias is correlative to intron number and shows disequilibrium among exons in plants. *BMC Genom.* **2013**, *14*, 56. [CrossRef]

38. Sudhir, K.; Glen, S.; Koichiro, T. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *7*, 1870.

39. Stefan, K.; Choudhuri, J.V.; Enno, O.; Chris, S.; Jens, S.; Robert, G. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *22*, 4633–4642.

40. Gary, B. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *2*, 573–580.

41. Sebastian, B.; Thomas, T.; Thomas, M.; Uwe, S.; Martin, M. Misa-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *16*, 2583–2585.

42. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2017**, *20*, 1160–1166. [CrossRef] [PubMed]

43. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [CrossRef] [PubMed]

44. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [CrossRef] [PubMed]

45. Subha, K.; Bui, Q.M.; Wong, T.K.; Von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Met.* **2017**, *14*, 587–589.

46. Bauman, N.; Akella, A.S.; Hann, A.E.; Morey, A.R.; Schwartz, A.A.S. Next-Generation Sequencing of *Haematococcus lacustris* Reveals an Extremely Large 1.35-Megabase Chloroplast Genome. *Genome Announc.* **2018**, *6*, e00181-18. [CrossRef]

47. Hideto, S.; Rui, K.; Takaaki, I.; Masanobu, O.; Koichi, M.; Atsuhiko, H.; Hiroshi, K. In vitro susceptibility of *Prototheca zopfii* genotypes 1 and 2. *Med. Mycol.* **2011**, *49*, 222–224.

48. Brouard, J.S.; Otis, C.; Lemieux, C.; Turmel, M. The exceptionally large chloroplast genome of the green alga *Floydiella terrestris* illuminates the evolutionary history of the Chlorophyceae. *Genome Biol. Evol.* **2010**, *2*, 240–256. [CrossRef]

49. Cremen, M.; Frederik, L.; Marcelino, V.R.; Heroen, V. Large diversity of nonstandard genes and dynamic evolution of chloroplast genomes in siphonous green algae (Bryopsidales, Chlorophyta). *Genome Biol. Evol.* **2018**, *10*, 1048–1061. [CrossRef] [PubMed]

50. Turmel, M.; Otis, C.; Lemieux, C. The complete chloroplast DNA sequences of the charophycean green algae Staurastrum and Zygnema reveal that the chloroplast genome underwent extensive changes during the evolution of the Zygnematales. *BMC Biol.* **2005**, *3*, 22. [CrossRef]

51. Ohno, S. So much "junk" DNA in our genome. *Brookhaven Symp. Biol.* **1972**, *23*, 366–370.

52. Petrov, D.A. Evolution of genome size: New approaches to an old problem. *Trends Genet.* **2001**, *17*, 23–28. [CrossRef]

53. Schubert, I.; Vu, G. Genome stability and evolution: Attempting a holistic view. *Trends Plant Sci.* **2016**, *9*, 749–757. [CrossRef]

54. Doolittle, W.F.; Sapienza, C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **1980**, *284*, 601–603. [CrossRef] [PubMed]

55. Vinogradov, A.E. Intron-genome size relationship on a large evolutionary scale. *J. Mol. Evol.* **1999**, *49*, 376–384. [CrossRef] [PubMed]

56. Voineagu, I.; Narayanan, V.; Lobachev, K.S.; Mirkin, S.M. Replication stalling at unstable inverted repeats: Interplay between DNA hairpins and fork stabilizing proteins. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9936–9941. [CrossRef] [PubMed]

57. Lemoine, F.J.; Degtyareva, N.P.; Lobachev, K.; Petes, T.D. Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell* **2005**, *120*, 587–598. [CrossRef]

58. Lobachev, K.S.; Shor, B.M.; Tran, H.T.; Taylor, W.; Keen, J.D.; Resnick, M.A.; Gordenin, D.A. Factors affecting inverted repeat stimulation of recombination and deletion in Saccharomyces cerevisiae. *Genetics* **1998**, *148*, 1507–1524. [CrossRef]

59. Lobachev, K.S.; Gordenin, D.A.; Resnick, M.A. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rear rangements. *Cell* **2002**, *108*, 183–193. [CrossRef]

60. Lobachev, K.S.; Stenger, J.E.; Kozyreva, O.G.; Jurka, J.; Resnick, M.A. Inverted Alu repeats unstable in yeast are excluded from the human genome. *Embo J.* **2000**, *19*, 3822–3830. [CrossRef]

61. Nag, D.K.; Manisha, S.; Stenson, E.K. Both CAG repeats and inverted DNA repeats stimulate spontaneous unequal sister-chromatid exchange in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **2004**, *32*, 5677–5684. [CrossRef]

62. Lacomme, C.; Hrubikova, K.; Hein, I. Enhancement of virus-induced gene silencing through viral-based production of inverted-repeats. *Plant J.* **2003**, *34*, 543–553. [CrossRef] [PubMed]

63. Leach, D. Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *Bioessays* **2010**, *16*, 893–900. [CrossRef] [PubMed]

64. Mizuno, K.; Miyabe, I.; Schalbetter, S.A.; Carr, A.M.; Murray, J.M. Recombination-restarted replication makes inverted chromosome fusions at inverted repeats. *Nature* **2013**, *493*, 246–249. [CrossRef] [PubMed]

65. Smith, N.A.; Singh, S.P.; Wang, M.B.; Stoutjesdijk, P.A.; Green, A.G.; Waterhouse, P.M. Total silencing by intron-spliced hairpin RNAs. *Nature* **2000**, *407*, 319–320. [CrossRef] [PubMed]

66. Zhang, R.; Ge, F.; Li, H.; Chen, Y.; Yang, L. PCIR: A database of plant chloroplast inverted repeats. *Database* **2019**, *2019*, baz127. [CrossRef] [PubMed]

67. Wakasugi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, M.; Ito, S.; Tsudzuki, J.; Nakashima, K.; Tsudzuki, T.; Suzuki, Y.; et al. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: The existence of genes possibly involved in chloroplast division. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 5967–5972. [CrossRef]

68. Koning, A.; Keeling, P.J. The complete plastid genome sequence of the parasitic green alga *Helicosporidium* sp. is highly reduced and structured. *BMC Biol.* **2006**, *4*, 12. [CrossRef] [PubMed]

69. Monique, T.; Christian, O.; Claude, L. Dynamic evolution of the chloroplast genome in the green algal classes Pedinophyceae and Trebouxiophyceae. *Genome Biol. Evol.* **2015**, *7*, 2062–2082.

70. Sanderson, M.J.; Copetti, D.; Burquez, A.; Bustamante, E.; Charboneau, J.; Eguiarte, L.E.; Kumar, S.; Lee, H.O.; Lee, J.; Mcmahon, M. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the ndh gene suite and inverted repeat. *Am. J. Bot.* **2015**, *102*, 1115–1127. [CrossRef]

71. Tomonori, H.; Atsushi Manabu, K.; Teiji, K.; Katsuhiko, T. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: Diversified genomic structure of coniferous species. *BMC Plant Biol.* **2008**, *8*, 70.

72. Wu, C.S.; Wang, Y.N.; Hsu, C.Y.; Lin, C.P.; Chaw, S.M. Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and Cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Genome Biol. Evol.* **2011**, *3*, 1284–1295. [CrossRef] [PubMed]

73. Zhang, Y.; Ma, J.; Yang, B.; Li, R.; Zhu, W.; Sun, L.; Tian, J.; Zhang, L. The complete chloroplast genome sequence of *Taxus chinensis* var. mairei (Taxaceae): Loss of an inverted repeat region and comparative analysis with related species. *Gene* **2014**, *540*, 201–209. [CrossRef] [PubMed]

74. Hao, D.C.; Xiao, P.G.; Huang, B.; Ge, G.B.; Yang, L. Interspecific relationships and origins of Taxaceae and Cephalotaxaceae revealed by partitioned Bayesian analyses of chloroplast and nuclear DNA sequences. *Plant Syst. Evol.* **2008**, *276*, 89–104. [CrossRef]

75. Yang, M.; Zhang, X.; Liu, G.; Yin, Y.; Chen, K.; Yun, Q.; Zhao, D.; Al-Mssallem, I.S.; Yu, J. The Complete Chloroplast Genome Sequence of Date Palm (*Phoenix dactylifera* L.). *PLoS ONE* **2010**, *5*, e12762. [CrossRef] [PubMed]

76. Subramanian, S.; Mishra, R.K.; Singh, L. Genome-wide analysis of microsatellite repeats in humans: Their abundance and density in specific genomic regions. *Genome Biol.* **2003**, *4*, R13. [CrossRef]

77. Legendre, M.; Pochet, N.; Pak, T.; Verstrepen, K.J. Sequence-based estimation of minisatellite and microsatellite repeat variability. *Genome Res.* **2007**, *17*, 1787–1796. [CrossRef]

78. Kuntal, H.; Sharma, V.; Daniell, H. Microsatellite analysis in organelle genomes of Chlorophyta. *Bioinformation* **2012**, *8*, 255–259. [CrossRef]

79. Stackelberg, M.V.; Rensing, S.A.; Reski, R. Identification of genic moss SSR markers and a comparative analysis of twenty-four algal and plant gene indices reveal species-specific rather than group-specific characteristics of microsatellites. *BMC Plant Biol.* **2006**, *6*, 9. [CrossRef]

80. Huotari, T.; Korpelainen, H. Complete chloroplast genome sequence of *Elodea canadensis* and comparative analyses with other monocot plastid genomes. *Gene* **2012**, *508*, 96–105. [CrossRef]
81. Gandhi, S.G.; Awasthi, P.; Bedi, Y.S. Analysis of SSR dynamics in chloroplast genomes of Brassicaceae family. *Bioinformation* **2010**, *5*, 16–20. [CrossRef]
82. Lehmann, J.; Libchaber, A. Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon. *RNA* **2008**, *14*, 1264–1269. [CrossRef] [PubMed]
83. Zhang, W.-J.; Zhou, J.; Li, Z.-F.; Wang, L.; Gu, X.; Zhong, Y. Comparative analysis of codon usage patterns among mitochondrion, chloroplast and nuclear genes in *Triticum aestivum* L. *J. Integr. Plant Biol.* **2007**, *49*, 246–254. [CrossRef]
84. Akashi, H. Synonymous codon usage in Drosophila melanogaster: Natural selection and translational accuracy. *Genetics* **1994**, *136*, 927–935. [CrossRef] [PubMed]
85. Dong, H.; Nilsson, L.; Kurland, C.G. Co-variation of tRNA abundance and codon usage in *Escherichia coliat* different growth rates. *J. Mol. Biol.* **1996**, *260*, 649–663. [CrossRef]
86. Yu, C.-H.; Dang, Y.; Zhou, Z.; Wu, C.; Zhao, F.; Sachs, M.S.; Liu, Y. Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Mol. Cell* **2015**, *59*, 744–754. [CrossRef]
87. Chen, L.; Liu, T.; Yang, D.; Nong, X.; Xie, Y.; Fu, Y.; Wu, X.; Huang, X.; Gu, X.; Wang, S.; et al. Analysis of codon usage patterns in *Taenia pisiformis* through annotated transcriptome data. *Biochem. Biophys. Res. Commun.* **2013**, *430*, 1344–1348. [CrossRef] [PubMed]
88. Yao, Z.; Hanmei, L.; Yong, G. Analysis of characteristic of codon usage in waxy gene of *Zea mays*. *J. Maize Sci.* **2008**, *2*, 432–436.
89. Jansen, R.K.; Cai, Z.; Raubeson, L.A.; Daniell, H.; Depamphilis, C.W.; Leebens-Mack, J.; Müller, K.F.; Guisinger-Bellian, M.; Haberle, R.C.; Hansen, A.K.; et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19369–19374. [CrossRef]
90. Fang, L.; Leliaert, F.; Novis, P.M.; Zhang, Z.; Zhu, H.; Liu, G.; Penny, D.; Zhong, B. Improving phylogenetic inference of core Chlorophyta using chloroplast sequences with strong phylogenetic signals and heterogeneous models. *Mol. Phylogenet. Evol.* **2018**, *127*, 248–255. [CrossRef]
91. Turmel, M.; Gagnon, M.-C.; O'Kelly, C.J.; Otis, C.; Lemieux, C. The Chloroplast Genomes of the Green Algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* Shed New light on the Evolutionary History of Prasinophytes and the Origin of the Secondary Chloroplasts of Euglenids. *Mol. Biol. Evol.* **2008**, *26*, 631–648. [CrossRef]
92. Huss, V.A.R.; Frank, C.; Hartmann, E.C.; Hirmer, M.; Kloboucek, A.; Seidel, B.M.; Wenzeler, P.; Kessler, E. Biochemical taxonomy and molecular phylogeny of the genus *Chlorella sensu* Lato (Chlorophyta). *J. Phycol.* **1999**, *35*, 587–598. [CrossRef]