Article

Strategies for Intraspecific Genotyping of Duckweed: Comparison of Five Orthogonal Methods Applied to the Giant Duckweed *Spirodela polyrhiza*

Manuela Bog [1,*], Luca Braglia [2], Laura Morello [2], Karen I. Noboa Melo [1], Ingo Schubert [3], Oleg N. Shchepin [1], K. Sowjanya Sree [4], Shuqing Xu [5], Eric Lam [6] and Klaus J. Appenroth [7]

[1] Institute of Botany and Landscape Ecology, University of Greifswald, 17489 Greifswald, Germany
[2] Istituto Biologia e Biotecnologia Agraria, Via Bassini 15, 20131 Milano, Italy
[3] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, 06466 Stadt Seeland, Germany
[4] Department of Environmental Science, Central University of Kerala, Periye 671320, India
[5] Institute of Organismic and Molecular Evolution, Johannes Gutenberg University Mainz, 55128 Mainz, Germany
[6] Department of Plant Biology, Rutgers the State University of New Jersey, New Brunswick, NJ 08901, USA
[7] Matthias Schleiden Institute—Plant Physiology, University of Jena, 07743 Jena, Germany
* Correspondence: manuela.bog@uni-greifswald.de; Tel.: +49-38344204146

Abstract: The predominantly vegetative propagating duckweeds are of growing commercial interest. Since clonal accessions within a respective species can vary considerably with respect to their physiological as well as biochemical traits, it is critical to be able to track the clones of species of interest after their characterization. Here, we compared the efficacy of five different genotyping methods for *Spirodela polyrhiza*, a species with very low intraspecific sequence variations, including polymorphic NB-ARC-related loci, tubulin-gene-based polymorphism (TBP), simple sequence repeat variations (SSR), multiplexed ISSR genotyping by sequencing (MIG-seq), and low-coverage, reduced-representation genome sequencing (GBS). Four of the five approaches could distinguish 20 to 22 genotypes out of the 23 investigated clones, while TBP resolved just seven genotypes. The choice for a particular method for intraspecific genotyping can depend on the research question and the project budget, while the combination of orthogonal methods may increase the confidence and resolution for the results obtained.

Keywords: *Spirodela polyrhiza*; Lemnaceae; duckweed; genotyping; intraspecific variation

1. Introduction

In the last two decades, the evolution of molecular methods has revolutionized phytotaxonomy. These advances have also been applied to the duckweed family, Lemnaceae Martinov [1,2]. The taxonomic investigation of duckweeds using molecular methods started with the genotyping work presented in [3] and was extended with the application of several methods, such as amplified fragment length polymorphism (AFLP) and plastidic and nuclear barcoding (reviewed in [4]). Classification based on morphological markers was the sole option of botanists for centuries (for duckweeds, see [5,6]) and was, to a great extent, confirmed and extended by molecular taxonomic investigations, uncovering phylogenetic relationships. It should be emphasized that there is little in common between these two approaches, with the latter being a more quantitative method.

Presently, 30 out of 36 species of duckweed can be reliably identified by molecular taxonomy [7,8]. Thus, further progress in these methods is still required. Several areas of duckweed research require not only the identification of species but also that of specific

clones from the same species, since some physiological properties of duckweeds are defined at the level of clones rather than the level of species and are sometimes linked with the respective ploidy level [9]. Such features include, e.g., the growth rate [10,11] and starch accumulation under the condition of the deficiency of essential nutrients, such as nitrate, phosphate, and sulphate [12], and other stresses [13,14]. Furthermore, in view of the growing commercial perspective of duckweeds, the patenting of specific clones [15] or their monitoring in commercial products would also require the ability to distinguish clones of interest from others of the same species. However, identification at the level of clones for duckweeds is extremely difficult using solely morphological markers due to their simple structure and abbreviated anatomical features. Therefore, suitable molecular methods for intraspecific genotyping must be developed.

Among the duckweed family, *Spirodela polyrhiza* (L.) Schleid (commonly called greater duckweed, Figure 1) displays unusually low intraspecific genetic variation. This was first suggested by the authors of [16], who found that there is no significant difference in genome size between the 38 clones of this species (in contrast to several other duckweed species; see [17], this Special Issue), and the investigation of three plastidic regions similarly detected hardly any variations. Subsequently, clonal variations between two reference genome quality assemblies of *S. polyrhiza* clones were found to be very low in terms of both intraspecific sequence variations as well as heterozygosity [18], with the single-nucleotide polymorphism positions (SNPs) being approximately six times fewer than those found between the ecotypes of *Arabidopsis* [19,20]. Using a population genomics approach with low-coverage sequencing reads, refs. [21,22] further extended the generality of these characteristics by finding very low genetic diversity between *S. polyrhiza* clones from a large number of locations across the globe. Moreover, the genome-wide spontaneous mutation rate in this species was estimated to be seven times lower than those of other multicellular eukaryotes characterized to date [21]. Using a novel approach to systematically identify and rank the polymorphic loci in the nuclear genome that may enable effective intraspecific genotyping, the authors of [23] used the genome sequences (40X or more coverage) from 10 clones of greater duckweed as a training set to identify loci among the NB-ARC-related gene family that could be used to discriminate between the clones of this species. This gene family, which is known to be involved in plant defence and immunity, was chosen because it displays the highest intraspecific polymorphism among the plant genomes that have been studied [19,24]. Validated primer sets were then used to uncover the intraspecific variations with an additional 13 clones of *S. polyrhiza*, bringing the total number to 23. From this work, 20 genotypes of these 23 clones could be distinguished. All 23 clones were selected from the list of 36 that were studied in [25] for their specific turion yield trait, based on their availability in the RDSC at the time. Three of the tested clones could not be resolved from each other using this genotyping technique but were distinct from each other in terms of the specific turion yield [25]. Nevertheless, the application of NB-ARC-related gene polymorphism represented a leap forward for the identification of intraspecific variations in Lemnaceae.

Here, we used the same clones as those studied by [23] for comparative studies using four additional orthogonal methods: fragment length polymorphisms (FLPs) and/or single-nucleotide polymorphisms (SNPs), tubulin-gene-based polymorphism (TBP; [26]), simple sequence repeat variants (SSR; [27]), multiplexed ISSR genotyping by sequencing (MIG-seq; [28]), and genotyping by genome-wide sequencing (GBS; [7]). In all five cases, we calculated dissimilarity trees to evaluate the different methods in terms of their genotyping efficacy and considered the advantages and limitations of each.
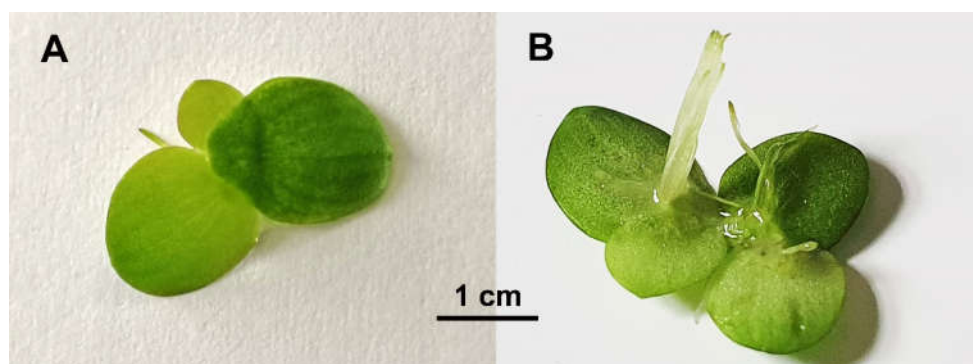
Figure 1. A colony of fronds of *Spirodela polyrhiza*, clone 7498: (**A**) dorsal view; (**B**) ventral view.

2. Results

2.1. NB-ARC-Related Genes (NB-ARC)

The NB-ARC approach is based on FLPs and SNPs of the DNA samples after their amplification with different primer sets, the sequences for which were generated and ranked by an informatic pipeline targeting genomic loci containing annotated NB-ARC genes [23]. Overall, 40 polymorphic characters were observed, with 17 based on FLPs and 23 based on SNPs (Table 1). The ambiguous, i.e., non-homozygous differences amounted to 26%. Figure 2 shows a single-linkage cluster and a heat map representation based on the uncorrected p-distance for the 23 *S. polyrhiza* clones used. All the numeric data, including the absolute and relative number of differences between clones, are given in the Supplementary Data S1. Three clones, 7379, 9503, and 9506, form a single data point and could not be distinguished from each other, but they are distinct from the other 20 clones. Thus, 21 genotypes were resolved from the 23 clones using this method. It should be noted that no error calculation is available for this approach, but based on the experience of one of the authors (E.L.), the repetition of the investigation of the same clone always produces identical results. One advantage of this method, as well as the TBP method, is that the number of fragments per sample tested is relatively small for the FLP and SNP analysis. Thus, the problem that usually affects other generic approaches in terms of the ambiguity in the band assignment becomes less of an issue. In any case, the results of this work revealed a geographic pattern associated with the dendrogram, i.e., clones from one continent were more likely to be similar to each other than clones from different continents (Figure 2).

Table 1. Comparison of the five tested methods for the genotyping of *Spirodela polyrhiza*, concerning the number of characters obtained and the proportion of ambiguous (non-homozygous) characters.

	NB-ARC	TBP	SSR	MIG-Seq	GBS
Marker Type [1]	FLP + SNP	FLP	FLP	SNP	SNP
Number of Characters	40 (17 + 23)	13	95	1292	6170
Percentage of Ambiguous Characters	26	0	0	29	14
Average Number of Ambiguous Characters per Sample ± Standard Deviation	11 ± 5	0 ± 0	0 ± 0	380 ± 69	894 ± 286
Error Rate % (Number of Replicates) [2]	n.d. (0)	0.00 (1)	0.26 (8)	0.00 (3)	0.17 (3)

[1] FLP—fragment length polymorphism; SNP—single-nucleotide polymorphism; [2] n.d.—not determined.
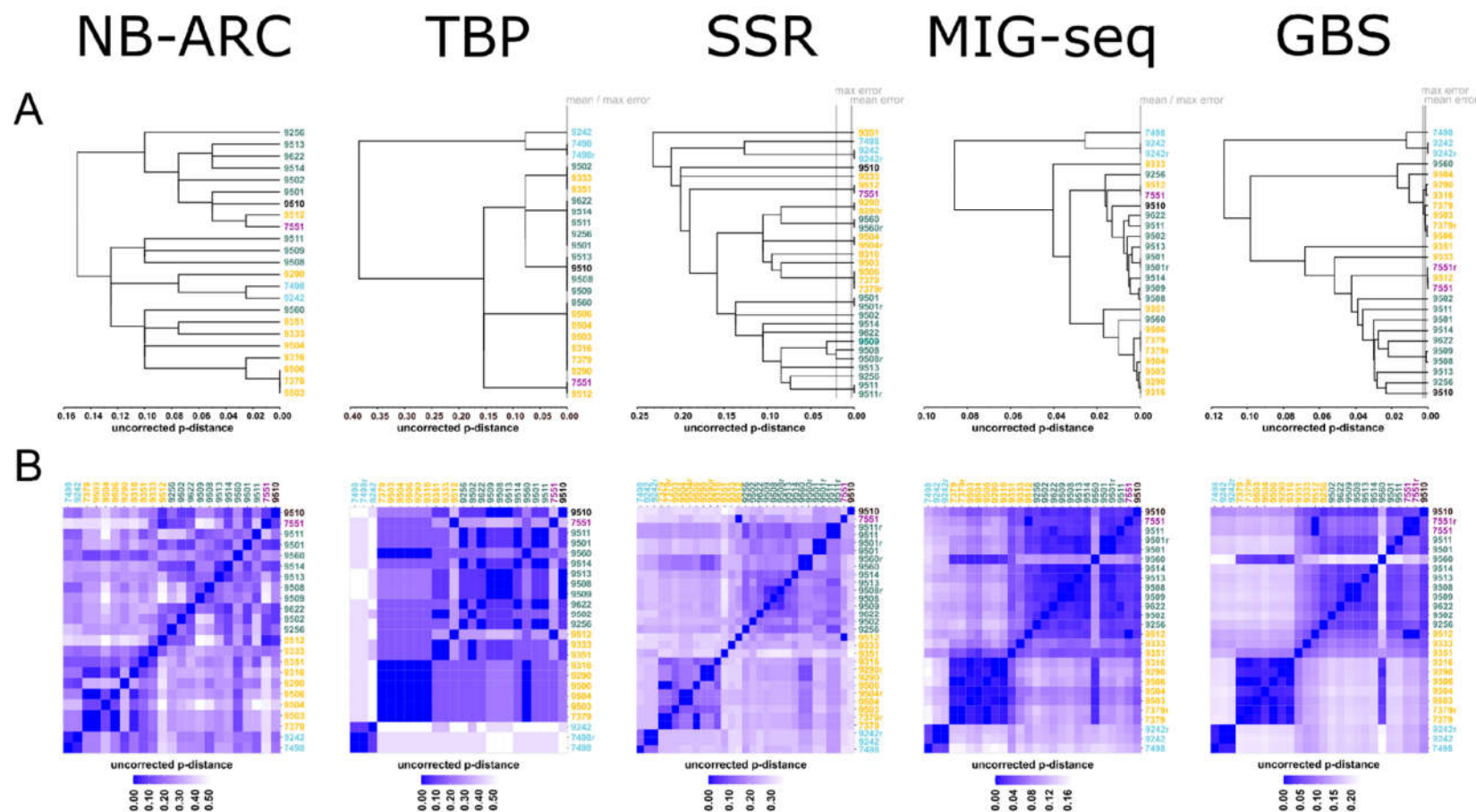
Figure 2. Comparison of five molecular methods for the genotyping of 23 clones of *S. polyrhiza*, including replicates (denoted by r). Colours denote the continent of origin of the clones: black—Africa, light blue—North + South America, yellow—Asia, violet—Australia, blue green—Europe. Upper panel (**A**) single-linkage clustering based on the uncorrected p-distances. The borders of resolution are given by the mean and max errors, calculated from the replicated clones. Lower panel (**B**) heat map representation of the uncorrected p-distances.

2.2. Tubulin-Gene-Based Polymorphism (TBP)

The TBP method uses FLPs from two introns of the conserved, multigene β-tubulin family. Since only a relatively small number of genomic loci were queried using this method, the number of variants available for the differentiation of the clones was also lowest when using this method. A total of 13 characters (fragment lengths), including four monomorphic clones, were detectable, with all of them originating from the first intron and with no polymorphism scored for the second intron (Table 1, Supplementary Data S1). Since the fragments were scored as either present or absent, no ambiguous character was detected. Due to the low number of the polymorphic scores, the resolution of this method for the 23 clones studied was very low, and only seven distinct genotypes could be resolved (Figure 2, Supplementary Data S1). The one sample that was run in replicates in this work yielded an identical result. The same was observed by two of the authors (L.B. and L.M.) for up to four replicates per sample of other *S. polyrhiza* clones not investigated in this study. In spite of the relatively low resolution of this method, a geographic pattern became obvious as well (Figure 2).

2.3. Simple Sequence Repeats (SSR)

The SSR method, like the TBP method, uses FLPs for the differentiation of different genotypes. In total, 95 amplification fragments resolved by their mobility on gels were scored as present or absent, which led, again, to no ambiguous characters (Supplementary Data S1). The error rate for the SSR was the highest of all the tested methods, but at 0.26%, it was still considered low (Table 1). Of the 23 investigated clones, 21 different genotypes could be detected. It was not possible to discriminate clone 7379 from clone 9506 and clone 7551 from clone 9512. A geographical pattern was, again, observable (Figure 2).

2.4. Multiplexed ISSR Genotyping by Sequencing (MIG-Seq)

The MIG-seq method is an SNP-based method. The final dataset (alignment can be found as Supplementary Data S2), chosen from the 20 tested parameter combinations, consisted of 1292 characters, of which 29% or, on average, 380 ± 69 characters per sample were ambiguous due to their heterozygosity and gaps (Supplementary Data S3). Surprisingly, there were no differences in homozygous position between the three samples that were run in replicates, which displayed an error rate of 0% (Table 1). The method revealed 20 genotypes. The clones 7379 and 9506, 7551 and 9512, and 9290 and 9316 could not be distinguished from one another. The geographical pattern was characterised by shorter branch lengths within the geographical subgroups rather than between the groups, in addition to the division between the continents of origin (Figure 2).

2.5. Genotyping-by-Sequencing (GBS)

The SNP-based GBS method yielded 6170 SNPs in total, which is the highest number of characters among all the tested methods (for the alignment, see Supplementary Data S4). The error rate (0.17%) and proportion of ambiguous characters (14%) are intermediate compared to the other methods (Table 1). Depending on the strictness of the error rate treatment, GBS could distinguish 22 genotypes (mean error rate as the threshold) or 17 genotypes (maximum error rate as the threshold) (Supplementary Data S1). Using the mean error rate as a threshold, only the clones 9506 and 9316 could not be distinguished from one another, while using the maximum error rate, the following clone pairs could not be distinguished from each other: 7551/9512, 7379/9506, 9290/9316, 9503/9506, 9506/9316, and 9509/9508. The clustering methods show a clear separation of the American clones from all the other clones (Figure 2). Additionally, the clones from Europe can be found in a separate cluster, although clone 9560 from Hungary clusters with the Asian clones, and this unusual pattern was found using all the investigated methods, including that of TBP. The Asian clones show a paraphyletic clustering, with clones 9333 (Hubei-

China), 9351 (Hanoi-Vietnam), and 9512 (Irkutsk-Russia) being more similar to the clones from Europe and the one from Australia. The other monophyletic Asian clones are from India.

3. Discussion

In order to test the efficacy of the orthogonal molecular methods in distinguishing between the clones of *S. polyrhiza*, we compared the NB-ARC, TBP, SSR, MIG-seq, and GBS approaches. At least four of these methods are known to have a high capacity to distinguish genotypes. We added TBP in this work, because this method is experimentally easy to carry out, as just two PCR reactions, followed by capillary electrophoresis, are required. Moreover, the capacity of TBP for resolving certain clones of the same species has been successfully tested on different *Lemna* species [29].

In most cases, the tested methods detected 20 or 21 genotypes among the 23 investigated clones. By GBS, with a lower stringency error rate treatment, even 22 genotypes may be distinguishable. We were unable to discriminate between all 23 clones using any of these methods alone. Four methods failed to distinguish clones 7379 and 9506 from each other, while with GBS, at a higher stringency, these two clones also became difficult to resolve. The situation was almost the same for the pair of clones 9506 and 9503, where SSR was the only method able to resolve these two genotypes. All three of these clones with unusually high genome sequence identities, 7379, 9506, and 9503, are from India, but they exhibited a more than 3-fold range difference in their specific turion yields of 1.86, 0.97, and 0.51, respectively. We therefore suspect that relatively specific genomic sequence (or epigenome) variations might account for these trait differences, consistent with previous speculations [23]. Our data were analysed quite strictly, and only the homozygous sites were considered for the method comparison. This led to a weakening of the results, because even heterozygous differences could represent genotype-distinguishing features. Nevertheless, we decided to opt for this strict approach, which does not require any complicating assumptions about phasing.

As expected, the methods that yield a high number of total characters and, therefore, differences, i.e., MIG-seq and GBS, showed the best resolution, especially in the phylogeographical studies. The geographical pattern closely fitted the results obtained by whole-genome resequencing in [21], which also separated the Indian *S. polyrhiza* clones from the Southeast Asian ones and clustered clone 9560 from Hungary together with the Asian clones. A great advantage of these two methods is their easy applicability for a wide range of organisms, although both methods require Illumina sequencing and the related equipment [7] (see Table 2 for a general comparison of the applied methods). The costs, especially if a company is asked to perform the analyses, are relatively high compared to the other methods. Furthermore, MIG-seq has an advantage over GBS, because in the case of MIG-seq, degraded DNA can be used as a template, since the amplified fragments between the SSR-based primers are relatively short. In contrast, GBS requires very good-quality DNA, since the first step is a fragmentation using restriction enzymes, and the clustering success of the single short Illumina-sequenced fragments highly depends on the quality of the template.

Table 2. Pros and cons of the five different methods used for genotyping of 23 *Spirodela polyrhiza* clones with a worldwide distribution.

	NB-ARC	**TBP**	**SSR**	**MIG-Seq**	**GBS**
DNA requirements	degraded works as well	degraded works as well	degraded works as well	degraded works as well	high quality necessary
Establishment	elaborate, genome sequence information needed	easy	elaborate, sequence information needed	easy	easy

Universality *	no, primers can be species-specific	yes	(no), low cross-amplification	yes	yes
Resolution capacity	high	low	high	high	high
Costs per sample	low to moderate, depends on whether genome information is available	low	low	moderate	high

* indicates whether the established method can be easily transferred between species. The sequence-guided approach and pipeline created for NB-ARC is indeed universal, but specific primers need to be established for each species. The same applies for SSR.

The NB-ARC polymorphism-based genotyping approach has shown a very good performance in genotyping *S. polyrhiza* [23], as observed here, with many of the original key observations confirmed and extended by the present work. A drawback of this method is the genome assembly resource required at the outset to establish this marker system for the species of interest. A well-assembled reference genome for the species and the nine other sequenced accessions were used as a training set for *S. polyrhiza*, a species known to have unusually low sequence variations, in order to informatically generate a ranked list of suitable primers to amplify the most polymorphic regions among the NB-ARC genes in the genome [23]. To improve the resolving power of this method, one could include additional genome sequence data on challenging clones, such as 7379 and 9503, in our training set for the pipeline, since clone 9506 was already part of the original training set. One apparent limitation of this approach could affect genomes that have a limited number of NB-ARC genes, such as the recently sequenced *Wolffia australiana* genome that has only three to four copies of these genes remaining [30]. In these cases, other large-plant gene families with members containing highly conserved exons could be used in the same informatic pipeline as that established in [23] to generate the requisite primers. For example, we note that in *W. australiana*, there is an amplification of the genes encoding the LRR-RK type of pattern recognition receptors (PRRs) to about 90 copies [1]. This gene family may present an alternative to the NB-ARC genes in *Wolffia australiana* for the generation of intraspecific genotyping primer sets.

SSR is likewise powerful, but it requires at least partial genomic information as well as the selection of useful sequence repeats and the design of primers in the flanking regions, as conducted in [27]. Furthermore, the use of cross-amplification to find useful polymorphic sites in related species is rather inefficient. The authors of [27] demonstrated this for *Lemna perpusilla*, and tests on a clone of *S. intermedia* (the sister species of *S. polyrhiza*) were unsuccessful in the lab of one of the authors (M.B.). The screening of already available whole-genome sequences or shotgun genome information appears to be a straightforward method for developing SSR markers using tools such as GMATA and others ([31] and references therein). Once the proper primers are available, the experimental procedure is not highly demanding financially or with respect to the required equipment or the quality of the template DNA.

The TBP method yielded, in our analysis, only seven genotypes based on the 13 scored characters for the 23 *S. polyrhiza* clones studied here. The same method, however, provided a useful resolution for the presumed *Lemna minor* clones, with 34 characters scored for the first intron and 36 genotypes identified among the 40 analysed clones [32]. Successful intraspecific genotyping was previously reported in crop species such as grape and olive [33,34]. However, because *S. polyrhiza* generally shows very little intraspecific variation [21,22], this species is a tough match for the TBP method and could be an exception rather than the common rule. Thus, for a duckweed species that displays greater intraspecific genome variations, this simple approach could still be a useful genotyping tool.

In summary, we presented five methods, four of which are well-suited for genotyping *S. polyrhiza* and likely other species, even those beyond the duckweed family, with a low intraspecific variability. The decision regarding which of these methods should be

applied depends on the question to be solved (and on the available budget and genome resources). In addition, while the data generated using the four most effective approaches are consistent, to a large extent, and thus provide strong support for their validity, they also suggest the potential advantage of combining two of these orthogonal methods to extend the power of the analysis. For example, using GBS analysis, under a lower stringency of error, only one pair of clones, 9506 and 9316, could not be resolved. These two clones, however, can be distinguished using any of the other three methods. Thus, if data based on GBS and one of these methods can be integrated together in a genotype matrix, the resolution of all 23 clones tested here could become feasible.

4. Materials and Methods

4.1. Plant Material and Cultivation

All 23 clones of *Spirodela polyrhiza* were acquired from the duckweed collection of the Matthias Schleiden Institute—Plant Physiology, University of Jena, Germany, and are available as living materials from the duckweed collections in Jena and at Rutgers State University of New Jersey (New Brunswick, New Jersey, U.S.A.) under the international four-digit code [35] given in Table 3. They were selected by the authors of [23] according to their geographical diversity and the wide range of their specific turion yield [25]. The clones were additionally characterised previously by measurements of their relative growth rate [11] and genome size [16], or these parameters were measured in the present project (Table 3). The species' identity was confirmed by barcoding using the plastidic markers *rpl16*, *rps16*, and *atpF-atpH* [16]. In total, 20 out of the 23 clones were whole-genome sequenced [21,23,36]. All the plants were cultivated under axenic conditions in N medium under standardised conditions for 10 days [37]: 8 mM KNO_3, 0.15 mM KH_2PO_4, 1 mM $MgSO_4$, 1 mM $Ca(NO_3)_2$, 5 μM H_3BO_3, 0.4 μM Na_2MoO_4, 13 μM $MnCl_2$, and 25 μM Fe(III)NaEDTA. The plant material was stored at −80 °C for further use.

Table 3. Origin of investigated *Spirodela polyrhiza* clones and their specific properties.

Clone ID	Country	Area/State	Specific Turion Yield [1]	Specific Growth Rate (h^{-1}) [2]	Genome Size (Mbp/1C) [3]	Genome Sequencing [4]
7379	India	Tamil Nadu	1.86 ± 0.26	0.315 ± 0.008	158 ± 3	n.d.
7498	USA	North Carolina	2.37 ± 0.27	0.401 ± 0.016	157 ± 2	+
7551	Australia	Northern Territory	n.d.	0.376 ± 0.009	n.d.	+
9242	Ecuador	Guayas	1.37 ± 0.12	0.456 ± 0.003	n.d.	+
9256	Finland	Uusimaa	3.73 ± 0.25	0.296 ± 0.011	160 ± 6	+
9290	India	Delhi	1.40 ± 0.13	0.299 ± 0.010	n.d.	+
9316	India	Rajasthan	1.22 ± 0.13	0.281 ± 0.009	159 ± 2	+
9333	China	Hubei	n.d.	0.311 ± 0.033	n.d.	+
9351	Vietnam	Hanoi	n.d.	0.367 ± 0.024	n.d.	n.d.
9501	Albania	Fieri	5.93 ± 0.03	0.357 ± 0.005	164 ± 4	+
9502	Ireland	Leinster	1.64 ± 0.14	0.337 ± 0.008	164 ± 5	+
9503	India	Rajasthan	0.51 ± 0.03	0.312 ± 0.007	170 ± 5	+
9504	India	Rajasthan	0.34 ± 0.03	0.284 ± 0.012	168 ± 6	+
9506	India	Telangana	0.97 ± 0.07	0.313 ± 0.005	161 ± 7	+
9508	Poland	Cracow	0.66 ± 0.04	0.360 ± 0.007	160 ± 4	+
9509	Germany	Thuringia	0.51 ± 0.07	0.323 ± 0.006	157 ± 2	+
9510	Mozambique	Maputo	0.98 ± 0.10	0.289 ± 0.008	n.d.	+
9511	Russia	Moscow	2.25 ± 0.11	0.381 ± 0.009	161 ± 5	+
9512	Russia	Irkutsk	2.92 ± 0.49	0.386 ± 0.005	156 ± 3	+
9513	Czech	Jindřichův Hradec	1.16 ± 0.04	0.324 ± 0.006	161 ± 5	+
9514	Austria	Viena	1.17 ± 0.16	0.333 ± 0.012	159 ± 5	+

| 9560 | Hungary | Bekes | n.d. | 0.367 ± 0.008 | 160 ± 6 | + |
| 9622 | Germany | Baden-Württemberg | n.d. | 0.310 ± 0.011 | 159 ± 5 | + |

[1] [25], [2] [11] or measured in the present project, [3] [16], [4] [21,23,36], n.d.—not determined.

4.2. DNA Isolation and Downstream Lab Work

DNA was isolated from 100–200 mg fresh weight of duckweed plants using the CTAB method [38] and quantified using a NanoVue spectrophotometer (GE Healthcare Europe GmbH, Freiburg, Germany). For GBS, the DNA was extracted using silica columns, according to [39].

The wet lab methods used for the analysis by NB-ARC are described in [23].

TBP amplification was performed by targeting both the 1st and 2nd introns of the multigene β-tubulin family, using two universal primer pairs, Fex-Rex and Fin-Rin, respectively, and PCR conditions described in [29]. FAM-labelled amplified fragments were separated by capillary electrophoresis (CE) using a AB3500 Genetic Analyzer (Thermo Fisher Scientific Inc., Waltham, MA, USA). The numerical output of the CE was exported to Excel format, and the fragments were aligned by size across the samples, considering a one-nucleotide (± 0.5) approximation. The amplicon size was considered as a marker, and its presence/absence was scored in a binary matrix (1/0, respectively), which was used for the subsequent elaboration of the data. Both the TBP 1st and 2nd introns were scored. Further technical details about the data acquisition and analysis are reported in [26]. The sample 7498 was run in duplicate using the same DNA extract for the error estimation.

For the SSR investigations, a total of 12 SSR markers multiplexed in three sets of four primer pairs each were used, according to [27]. The forward primers were labelled with a fluorescence dye (set1: SP12[FAM], SP14[HEX], SP43[CY3], SP51[ROX]; set2: SP6[HEX], SP25[ROX], SP29[FAM], SP42[CY3]; set3: SP36[HEX], SP45[ROX], SP47[CY3], SP53[FAM]). PCR amplification was performed using a thermal cycler (SensoQuest GmbH, Göttingen, Germany) in a total volume of 10 μL containing 45 ng DNA template, 2.5 μL of 2 X KAPA 2G (Kapa Biosystems Pty Ltd., South Africa), and the final concentration of each primer (SP6[0.1 μM], SP12[0.2 μM], SP14[0.2 μM], SP25[0.2 μM], SP29[0.2 μM], SP36[0.1 μM], SP42[0.4 μM], SP43[0.2 μM], SP45[0.4 μM], SP47[0.4 μM], SP51[0.2 μM], SP53[0.1 μM]). The samples were amplified under the following conditions: 3 min at 95 °C, 35 cycles of 15 s at 95 °C, 35 s at 54 °C, and 30 s at 72 °C, and 5 min of the final extension at 72 °C. Subsequently, 2.5 μL of the diluted PCR products (set 1: undiluted, set 2: 1:10, set 3: 1:25) were mixed with 0.16 μL of GeneScan 500 LIZ size standard and 10.5 μL HiDi formamide (both Thermo Fisher Scientific). Finally, the fragment analysis was carried out using a 3130xl Genetic Analyzer (Thermo Fisher Scientific). Eight samples (7379, 9242, 9290, 9501, 9504, 9508, 9511, and 9560) from the same DNA extracts were run in duplicate for the error estimation.

For MIG-seq, 25 μL of each DNA extract at a concentration of 30 ng/μL was sent to LGC Genomics GmbH (Berlin, Germany) for analysis. Sequencing by synthesis was performed with a 150 bp paired-end read chemistry using an Illumina NextSeq500/550 device (Illumina Inc., San Diego, CA, USA). Three samples (7379, 9242, 9501) from the same DNA extracts were run in duplicate for the error estimation. The sequence reads were uploaded to the sequence read archive (SRA) [40], PRJNA888369, accession numbers: SRR21845253-SRR21845278.

For the GBS method, DNA was used for the construction of barcoded libraries, essentially performed as described in [41]. Sequencing by synthesis (single read, 1 × 107 cycles, index read 8 cycles) using the Illumina HiSeq 2500 device was performed according to the protocols provided by the manufacturer (Illumina Inc.). Three samples (7379, 7551, and 9242) from the same DNA extracts were run in duplicate for the error estimation. The sequence reads were uploaded to the sequence read archive (SRA) [40], PRJNA888369, accession numbers: SRR21845363-SRR21845387 and SRR21845392.

4.3. Data Analysis

The dataset for the analysis by NB-ARC was taken from [23], and the scoring of the TBP fragments was performed as described in [26]. The fragment lengths based on the SSR analysis were scored using GeneMapper v5 (Thermo Fisher Scientific) and subsequently rounded using Tandem v1.09 [42]. Finally, the fragment length data were converted into a 0/1 matrix. For MIG-seq, demultiplexing and trimming of sequencing adapters and primers from the raw sequence reads were done by the sequencing facility. Around 8 million pre-processed paired-end reads (mean: 307.5 thousand reads per sample, SD: 45.4 thousand reads) were de novo assembled and analysed using the ipyrad v0.9.84 pipeline [43] run on the HPC Brain Cluster of the University of Greifswald (Supplementary Data S5). We optimized the core assembly parameters 'clustering threshold for de novo assembly' (ct) and 'minimum depth for statistical base calling' (md) by running the analysis with 10 different values of ct (0.81–0.99, with a step of 0.2) and two values of md (6 and 10). Following the method of [44], the datasets obtained with the parameters ct = 0.91 and md = 10 were determined as optimal for the downstream analyses (Supplementary data S5). The SNP identification of the approximately 77 mio pre-processed single-end reads (mean: 3 mio reads per sample, SD: 600 thousand reads) for the GBS data was performed as described in [7], except for the fact that the *S. polyrhiza* genome available in [45] was used for the SNP mapping.

We attempted to make the data analysis for the comparison as similar as possible, although the different methods produced different data types, including FLP (scored as 0/1) and SNP data. Therefore, 0/1 data was recoded as c/a, and each of the five datasets was saved as a FASTA file, which was further processed using R v4.1.0 [46], with the additional libraries ape v5.6-2 [47], stringdist v0.9.8 [48], phangorn v2.10.0 [49], reshape2 v1.4.4 [50], and ggplot2 v3.3.6 [51]. In a first step, the uncorrected p distance and character differences were calculated using the 'dist.hamming' function. Ambiguous characters were not considered. Then, a single-linkage clustering was applied using the 'hclust' function. The resulting trees were saved in newick format and formatted in FigTree v1.4.4 [52]. Heat maps were plotted using the 'ggplot' function after the conversion of the distance matrices into a linear form using the 'melt' function. Ambiguous sites were counted using a custom python script.

References

1. Acosta, K.; Appenroth, K.J.; Borisjuk, L.; Edelman, M.; Heinig, U.; Jansen, M.A.K.; Oyama, T.; Pasaribu, B.; Schubert, I.; Sorrels, S.; et al. Return of the Lemnaceae: Duckweed as a model plant system in the genomics and postgenomics era. *Plant Cell* **2021**, *33*, 3207–3234. https://doi.org/10.1093/plcell/koab189.
2. Tippery, N.P.; Les, D.H.; Appenroth, K.J.; Sree, K.S.; Crawford, D.J.; Bog, M. Lemnaceae and Orontiaceae are phylogenetically and morphologically distinct from Araceae. *Plants* **2021**, *10*, 2639. https://doi.org/10.3390/plants10122639.
3. Les, D.H.; Crawford, D.J.; Landolt, E.; Gabel, J.D.; Kimball, R.T. Phylogeny and systematics of Lemnaceae, the duckweed family. *Syst. Bot.* **2002**, *27*, 221–240. https://doi.org/10.1043/0363-6445-27.2.221.
4. Bog, M.; Appenroth, K.J.; Sree, K.S. Duckweed (Lemnaceae): Its molecular taxonomy. *Front. Sustain. Food Syst.* **2019**, *3*, 117. https://doi.org/10.3389/fsufs.2019.00117.
5. Landolt, E. *The Family of Lemnaceae—A Monographic Study*. *Vol. 1. Biosystematic Investigations in the Family of Duckweeds (Lemnaceae)*; Veröffentlichungen des Geobotanischen Institutes der ETH, Stiftung Ruebel: Zurich, Switzerland, 1986.
6. Bog, M.; Appenroth, K.J.; Sree, K.S. Key to the determination of taxa of Lemnaceae: An update. *Nordic J. Bot.* **2020**, *8*, e02658. https://doi.org/10.1111/njb.02658.
7. Bog, M.; Xu, S.; Himmelbach, A.; Brandt, R.; Wagner, F.; Appenroth, K.J.; Sree, K.S. Genotyping-By-Sequencing for species delimitation in the section *Uninerves* Hegelm. (Genus *Lemna* L.). In *Compendium of Plant Genomes: The Duckweed Genomes*, 1st ed.; Cao, X.H., Fourounjian, P., Wang, W., Eds.; Springer Nature: Cham, Switzerland, 2020; pp. 115–123. https://doi.org/10.1007/978-3-030-11045-1_11.
8. Bog, M.; Sree, K.S.; Fuchs, J.; Hoang, P.T.N.; Schubert, I.; Kuever, J.; Rabenstein, A.; Paolacci, S.; Jansen, M.A.K.; Appenroth, K.J. A taxonomic revision of *Lemna* sect. *Uninerves* (Lemnaceae). *Taxon* **2020**, *69*, 56–66. https://doi.org/10.1002/tax.12188.
9. Vunsh, R.; Heinig, U.; Malitsky, S.; Aharoni, A.; Avidov, A.; Lerner, A.; Edelman, M. Manipulating duckweed through genome duplication. *Plant Biol.* **2015**, *17*, 115–119. https://doi.org/10.1111/plb.12212.
10. Sree, K.S.; Sudakaran, S.; Appenroth, K.J. How fast can angiosperms grow? Species and clonal diversity of growth rates in the genus *Wolffia* (Lemnaceae). *Acta Physiol. Plant.* **2015**, *37*, 204. https://doi.org/10.1007/s11738-015-1951-3.
11. Ziegler, P.; Adelmann, K.; Zimmer, S.; Schmidt, C.; Appenroth, K.J. Relative in vitro growth rates of duckweeds (Lemnaceae), the most rapidly growing higher plants. *Plant Biol.* **2015**, *17*, 33–41. https://doi.org/10.1111/plb.12184.
12. Sree, K.S.; Appenroth, K.J. Starch accumulation in duckweeds (Lemnaceae) induced by nutrient deficiency. *Emir. J. Food Agric.* **2022**, *34*, 204–212. https://doi.org/10.9755/ejfa.2022.v34.i3.2846.
13. Sree, K.S.; Adelmann, K.; Garcia, C.; Lam, E.; Appenroth, K.J. Natural variance in salt tolerance and induction of starch accumulation in duckweeds. *Planta* **2015**, *241*, 1395–1404. https://doi.org/10.1007/s00425-015-2264-x.
14. Appenroth, K.J.; Ziegler, P.; Sree, K.S. Accumulation of starch in duckweeds (Lemnaceae), potential energy plants. *Physiol. Mol. Biol. Plants*. **2021**, *27*, 2621–2633. https://doi.org/10.1007/s12298-021-01100-4.
15. Kaplan, A.; Zelicha, H.; Tsaban, G.; Yaskolka, M.A.; Rinott, E.; Kovsan, J.; Novack, L.; Thiery, J.; Ceglarek, U.; Burkhardt, R.; et al. Protein bioavailability of *Wolffia globosa* duckweed, a novel aquatic plant—A randomized controlled trial. *Clin. Nutr.* **2019**, *38*, 2576–2582. https://doi.org/10.1016/j.clnu.2018.12.009.
16. Bog, M.; Lautenschlager, U.; Landrock, M.F.; Landolt, E.; Fuchs, J.; Sree, K.S.; Oberprieler, C.; Appenroth, K.J. Genetic characterization and barcoding of taxa in the genera *Landoltia* and *Spirodela* (Lemnaceae) by three plastidic markers and amplified fragment length polymorphism (AFLP). *Hydrobiologia* **2015**, *749*, 169–182. https://doi.org/10.1007/s10750-014-2163-3.
17. Hoang, P.T.N.; Fuchs, J.; Schubert, V.; Tran, T.B.N.; Schubert, I. Chromosome Numbers and Genome Sizes of All 36 Duckweed Species (Lemnaceae). *Plants* **2022**, *11*, 2674. https://doi.org/10.3390/plants11202674.
18. Michael, T.P.; Bryant, D.; Gutierrez, R.; Borisjuk, N.; Chu, P.; Zhang, H.; Xia, J.; Zhou, J.; Peng, H.; El Baidouri, M. et al. Comprehensive definition of genome features in *Spirodela polyrhiza* by high-depth physical mapping and short-read DNA sequencing strategies. *Plant J.* **2017**, *89*, 617–635. https://doi.org/10.1111/tpj.13400.
19. Gan, X.; Stegle, O.; Behr, J.; Steffen, J.; Drewe, P.; Hildebrand, K.; Lyngsoe, R.; Schultheiss, S.J.; Osborne, E.J.; Sreedharan, K.L. et al. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* **2011**, *477*, 419–423. https://doi.org/10.1038/nature10414.
20. Lu, P.; Han, X.; Qi, J.; Yang, J.; Wijeratne, A.J.; Li, T.; and Ma, H. Analysis of *Arabidopsis* genome-wide variations before and after meiosis and meiotic recombination by resequencing Landsberg *erecta* and all four products of a single meiosis. *Genome Res.* **2012**, *22*, 508–518.
21. Xu, S.; Stapley, J.; Gablenz, S.; Boyer, J.; Appenroth, K.J.; Sree, K.S.; Gershenzon, J.; Widmer, A.; Huber, M. Low genetic variation is associated with low mutation rate in the giant duckweed. *Nat. Commun.* **2019**, *10*, 1243. https://doi.org/10.1038/s41467-019-09235-5.
22. Ho, E.K.H.; Bartkowska, M.; Wright, S.I.; Agrawal, A.F. Population genomics of the facultatively asexual duckweed *Spirodela polyrhiza*. *New Phytol.* **2019**, *224*, 1361–1371. https://doi.org/10.1111/nph.16056.
23. Chu, P.; Wilson, G.M.; Michael, T.P.; Vaiciunas, J.; Honig, J.; Lam, E. Sequence-guided approach to genotyping plant clones and species using polymorphic NB-ARC-related genes. *Plant Mol. Biol.* **2018**, *98*, 219–231. https://doi.org/10.1007/s11103-018-0774-1.
24. Clark, R.M.; Schweikert, G.; Toomajian, C.; Ossowski, S.; Zeller, G.; Shinn, P.; Warthmann, N.; Gu, T.T.; Fu, G.; Hinds, D.A. et al. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **2007**, *317*, 338–342. https://doi.org/10.1126/science.1138632.

25. Kuehdorf, K.; Jetschke, G.; Ballani, L.; Appenroth, K.J. The clonal dependence of turion formation in the duckweed Spirodela polyrhiza—An ecogeographical approach. *Physiol. Plant.* **2014**, *150*, 46–54. https://doi.org/10.1111/ppl.12065.

26. Braglia, L.; Gavazzi, F.; Morello, L.; Gianì, S.; Nick, P.; Breviario, D. On the applicability of the Tubulin-Based Polymorphism (TBP) genotyping method: A comprehensive guide illustrated through the application on different genetic resources in the legume family. *Plant Methods* **2020**, *16*, 86. https://doi.org/10.1186/s13007-020-00627-z.

27. Xu, N.; Hu, F.; Zhang, W.; Wang, M.; Zhu, M.; Ke, J. Characterization of 19 polymorphic SSR markers in *Spirodela polyrhiza* (Lemnaceae) and cross amplification in *Lemna perpusilla*. *Appl. Plant Sci.* **2018**, *6*, 1–5. https://doi.org/10.1002/aps3.1153.

28. Suyama, Y.; Matsuki, Y. MIG-seq: An effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. *Sci. Rep.* **2015**, *5*, 16963. https://doi.org/10.1038/srep16963.

29. Braglia, L.; Lauria, M.; Appenroth, K.J.; Bog, M.; Breviario, D.; Grasso, A.; Gavazzi, F.; Morello, L. Duckweed species genotyping and interspecific hybrid discovery by tubulin-based polymorphism fingerprinting. *Front. Plant Sci.* **2021**, *12*, 625670. https://doi.org/10.3389/fpls.2021.625670.

30. Michael, T.P.; Ernst, E.; Hartwick, N.; Chu, P.; Bryant, D.; Gilbert, S.; Ortleb, S.; Baggs, E.L.; Sree, K.S.; Appenroth, K.J.; et al. Genome and time-of-day transcriptome of *Wolffia australiana* link morphological minimization with gene loss and less growth control. *Genome Res* **2021**, *31*, 225–238. https://doi.org/10.1101/gr.266429.120.

31. Wang, X.; Wang, L. GMATA: An Integrated Software Package for Genome-Scale SSR Mining, Marker Development and Viewing. *Front. Plant Sci.* **2016**, *7*, 1350. https://doi.org/10.3389/fpls.2016.01350.

32. Braglia, L.; Breviario, D.; Gianì, S.; Gavazzi, F.; De Gregori, J.; Morello, L. New Insights into Interspecific Hybridization in *Lemna* L. L. Sect. *Lemna* (Lemnaceae Martinov). *Plants* **2021**, *10*, 2767. https://doi.org/10.3390/plants10122767.

33. Gavazzi, F.; Braglia, L.; Mastromauro, F.; Gianì, S.; Morello, L.; Breviario, D. The Tubulin-Based-Polymorphism Method Provides a Simple and Effective Alternative to the Genomic Profiling of Grape. *PLoS ONE* **2016**, *11*, e0163335. https://doi.org/10.1371/journal.pone.0163335.

34. Braglia, L.; Manca, A.; Gianì, S.; Hatzopoulos, P.; Breviario, D. A Simplified Approach for Olive (*Olea europaea* L.) Genotyping and Cultivars Traceability. *Am. J. Plant Sci.* **2017**, *8*, 3475–3489. https://doi.org/10.4236/ajps.2017.813234.

35. Zhao, H.; Appenroth, K.J.; Landesman, L.; Salmean, A.A.; Lam, E. Duckweed rising at Chengdu: Summary of the 1st International Conference on Duckweed Application and Research. *Plant Mol. Biol.* **2012**, *78*, 627–632. https://doi.org/10.1007/s11103-012-9889-y.

36. Hoang, P.N.T.; Michael, T.P.; Gilbert, S.; Chu, P.; Motley, S.T.; Appenroth, K.J.; Schubert, I.; Lam, E. Generating a high-confidence reference genome map of the greater duckweed by integration of cytogenomic, optical mapping, and Oxford Nanopore technologies. *Plant J.* **2018**, *96*, 670–684. https://doi.org/10.1111/tpj.14049.

37. Appenroth, K.J.; Teller, S.; Horn, M. Photophysiology of turion formation and germination in *Spirodela polyrhiza*. *Biol. Plant.* **1996**, *38*, 95–106. https://doi.org/10.1007/BF02879642.

38. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.

39. Valledor, L.; Escandón, M.; Meijón, M.; Nukarinen, E.; Canal, M.J.; Weckwerth, W. A universal protocol for the combined isolation of metabolites, DNA, long RNAs, small RNAs, and proteins from plants and microorganisms. *Plant J.* **2014**, *79*, 173–180. https://doi.org/10.1111/tpj.12546.

40. National Center for Biotechnolgy Information (NCBI). Available online: https://www.ncbi.nlm.nih.gov/ (accessed on 10 October 2022).

41. Wendler, N.; Mascher, M.; Nöh, C.; Himmelbach, A.; Scholz, U.; Ruge-Wehling, B.; Stein, N. Unlocking the secondary genepool of barley with next-generation sequencing. *Plant Biotechnol. J.* **2014**, *12*, 1122–1131. https://doi.org/10.1111/pbi.12219.

42. Matschiner, M.; Salzburger, W. TANDEM: Integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **2009**, *25*, 1982–1983. https://doi.org/ 10.1093/bioinformatics/btp303.

43. Eaton, D.A.R.; Overcast, I. ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* **2020**, *36*, 2592–2594. https://doi.org/10.1093/bioinformatics/btz966.

44. McCartney-Melstad, E.; Gidiş, M.; Shaffer, H.B. An empirical pipeline for choosing the optimal clustering threshold in RADseq studies. *Mol. Ecol. Resour.* **2019**, *19*, 1195–1204. https://doi.org/10.1111/1755-0998.13029.

45. *Spirodela polyrhiza* Genome on Joint Genome Institute (JGI). Available online: https://data.jgi.doe.gov/refine-download/phytozome?organism=Spolyrhiza&expanded=290 (accessed on 13 October 2022).

46. R Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2021. Available online https://www.R-project.org/ (accessed on 6 October 2022).

47. Paradis, E.; Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2019**, *35*, 526–528. https://doi.org/10.1093/bioinformatics/bty633.

48. van der Loo, M.P.J. The stringdist package for approximate string matching. *R J.* **2014**, *6*, 111–122.

49. Schliep, K.P. phangorn: Phylogenetic analysis in R. *Bioinformatics* **2011**, *27*, 592–593. https://doi.org/ 10.1093/bioinformatics/btq706.

50. Wickham, H. Reshaping Data with the reshape Package. *J. Stat. Soft* **2007**, *21*, 1–20. https://doi.org/10.18637/jss.v021.i12.

51. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.

52. FigTree—A Graphical Viewer of Phylogenetic Trees and as a Program for Producing Publication-Ready Figures. Available online: http://tree.bio.ed.ac.uk/software/figtree/ (accessed on 6 October 2022).