ORIGINAL ARTICLE

Genetic structure of duckweed population of *Spirodela*, *Landoltia* and *Lemna* from Lake Tai, China

Jie Tang · Fei Zhang · Weihua Cui · Jiong Ma

Abstract

Main conclusion **Presenting a basic framework for using MLST to characterize *Spirodela*, *Landoltia* and in particular *Lemna* strains at the species level, and to study population genetics and evolution history of natural duckweed populations.**

Abstract Duckweed is widely used in environmental biotechnology and has recently emerged as a potential feedstock for biofuels due to its high growth rate and starch content. The genetic diversity and composition of a natural duckweed population in genera *Spirodela*, *Landoltia* and *Lemna* from Lake Tai, China, were investigated using probabilistic analysis of multilocus sequence typing (MLST). The 78 strains were categorized into five lineages, among which strains representing *L. aequinoctialis* and *S. polyrhiza* were predominant. Among the five lineages, interlineage transfers of markers were infrequent and no recombination was statistically detected. Tajima's *D* tests determined that all loci are subject to population bottlenecks, which is likely one of the main reasons for the low genetic diversity observed within the lineages. Interestingly, strains of *L. turionifera* are found to contain small admixture from *L. minor*, providing rare evidence of transfer of genetic materials in duckweed. This was discussed with respect to the hypothesis that a cross of these two gave rise to *L. japonica*. Moreover, the conventional maximum-likelihood phylogenetic analysis clearly recognized all the species in the three genera with high bootstrap supports. In conclusion, this work offers a basic framework for using MLST to characterize *Spirodela*, *Landoltia* and in particular *Lemna* strains at the species level, and to study population genetics and evolution history of natural duckweed populations.

Abbreviations

atpF–atpH	Intergenic spacer between ATPase subunit I (*atpF*) and ATPase subunit III (*atpH*)
matK	Maturase *K* gene
ML	Maximum likelihood
MLST	Multilocus sequence typing
rpoB	RNA polymerase beta subunit gene
ST	Sequence type
I_{cong}	Index of congruence

J. Tang and F. Zhang contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00425-014-2053-y) contains supplementary material, which is available to authorized users.

J. Tang · F. Zhang · J. Ma (✉)
School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen 518055, China
e-mail: majiong@pkusz.edu.cn

W. Cui
School of Water Resources and Environment, China University of Geosciences, Beijing 100083, China

Introduction

Lemnaceae (Duckweed) is a free-floating, flowering, widely distributed aquatic plant. As the smallest angiosperm, duckweed mainly consists of a leaf-like body (frond) and varying amounts of roots. Under suitable growth conditions, duckweed doubles its biomass in 1–3 days and produces a continued biomass supply for 9–12 months annually (Cheng and Stomp 2009; Xu et al.

2011, 2012). The biomass of duckweed primarily consists of carbohydrates, proteins, crude fibers, lipids and minerals. The most attractive advantage of duckweed is its high starch content, which can reach about 50 % of dry weight, depending on duckweed species and growing conditions used (Reid and Bieleski 1970; Landolt and Kandeler 1987; Cheng and Stomp 2009). By extrapolating from field-study results, biomass yields of 39.1–105.9 t ha^{-1} year^{-1} (dry biomass) could be achieved for duckweed using wastewater as a nutrient source. These yields are substantially higher than most other potential energy crops such as switchgrass, miscanthus, willow and Bermuda grass (Xu et al. 2012). Therefore, duckweed is a promising feedstock for various applications including animal feed, bioproducts, biofuels, and waste-to-energy utilization of nutrient-rich wastewater.

Duckweed is a monocotyledonous family of five genera, including *Spirodela*, *Landoltia*, *Lemna*, *Wolffia* and *Wolffiella* (Landolt and Kandeler 1987; Les et al. 1997). Duckweed taxonomy, however, has been controversial due to its extremely simplified morphologies. The conventional identification method depends on a few morphological characteristics, such as size, shape and pigmentation of fronds, which are susceptible to environmental effects. More recently, molecular techniques, especially sequencing technologies with increasing accuracy and decreasing costs, have become widely available and practical for genotyping purposes. Further, the use of chloroplast sequences (Cabrera et al. 2008; Wang et al. 2010; Appenroth et al. 2013; Bog et al. 2013) has proved to be more informative and provide more robust support to phylogenetic analyses of duckweed.

It remains largely unknown whether duckweed reproduces sexually in the wild and how sexual reproduction affects its population structure and genetic integrity and/or diversity. Normally, duckweed propagates clonally by forming daughter fronds from the mother frond. It was reported, however, that flowering of duckweed can be induced. Induction of flowering by chemicals was achieved in *S. polyrhiza*, *L. aequinoctialis* and *W. microscopica* (Maheshwari and Gupta 1967; Seth et al. 1970; Marinc-Hrzenjak et al. 2008); it was achieved with hormones in *S. polyrhiza* and *L. aequinoctialis* (Marinc-Hrzenjak et al. 2008; Shimakawa et al. 2012; Miyawaki et al. 2013). In addition, *L. perpusilla* was induced to flower by controlling photoperiod (Purves 1961) and *L. aequinoctialis* by starvation stress (Shimakawa et al. 2012). The consequences of flowering might contribute to the exchange of genetic materials. Furthermore, genetic variations have been found among duckweed strains from different geographic regions (Bog et al. 2010, 2013; Xue et al. 2012). Thus, the genetic structure of natural duckweed populations could be uniquely shaped due to distinct propagation mechanisms and the diverse environmental influences over time.

Although duckweeds have been studied in the laboratory for many years, limited information is available on the biology of natural populations. Such data is essential for understanding the reproduction mechanisms in the natural population and should provide useful insights into the evolutionary processes at the population level. In this report, we present genetic characterization of duckweed strains collected from numerous sites around Lake Tai, China. We used multilocus sequence typing (MLST) (Maiden 2006) of one non-coding intergenic spacer (*atpF–atpH*) and two housekeeping genes (*matK*, *rpoB*) in our analyses. Evolutionary histories were inferred using model-based Bayesian approach as well as conventional maximum-likelihood (ML) methods. We further analyzed the genetic diversity and composition of these duckweed populations.

Materials and methods

Sampling sites and duckweed collection

Lake Tai, China's third largest freshwater lake with an area of 2,250 km^2 and an average depth of 2 m, is located in the Yangtze Delta plain, Eastern China. Duckweed strains are distributed over the entire lake as well as the water channel systems connected to the lake. Fifty-five sites were randomly chosen for sampling, covering most regions in Lake Tai (Supplementary Fig. S1; Supplementary Table S1). We focused only on collecting strains of *Spirodela*, *Landoltia* and *Lemna*, given our research interest in bioenergy. A total of 78 strains of *Spirodela*, *Landoltia* and *Lemna* were collected between July 6 and 10, 2013, and named by a combination of numbers (sampling site) and letters (different morphologies). All samples were maintained at 4 °C before further treatment.

In the laboratory, aseptic duckweed strains were established by rinsing fronds with 1 % sodium hypochlorite solution (NaClO) for 1 min. Then, single fronds were picked separately and cultured in Schenk and Hildebrandt basal salt mixture (Sigma, S6765) liquid medium at pH 5.8. The plants were grown in a controlled climate chamber under a photoperiod of 16 h light (100 μmol m^{-2} s^{-1}; 23 °C) and 8 h dark (15 °C).

DNA extraction, amplification and sequencing

Approximately 100 mg of powdered duckweed tissue was used for nucleic acid isolation. Total genomic DNA was isolated by Plant DNA Isolation Reagent (Takara, Dalian, China). The quality and quantity of genomic DNA were examined by agarose gel electrophoresis and Nanodrop 1000 (Nanodrop Technologies, Wilmington, USA), respectively.

Partial sequences of two coding genes (*matK*, *rpoB*) and complete sequences of one non-coding intergenic spacer (*atpF–atpH*) were generated using polymerase chain reactions (PCR) for all 78 strains collected from Lake Tai. DNA amplifications were performed using a TProfessional thermocycler (Biometra, Goettingen, Germany) with 20-µL reaction mixtures containing 1–1.5 ng DNA, 0.1 mmol/L each dNTP, 0.2 µmol/L each primer, 10 µL Premix Taq™ (Takara) and sterile double-distilled water. Primers for amplification and sequencing were all derived from the literature and, together with temperature and PCR cycling conditions, are shown in Supplementary Table S2. PCR products were sequenced by Beijing Genome Institute (BGI, Shenzhen, China). Nucleotide sequences generated in this study were all deposited in GenBank and their accession numbers are listed in Supplementary Table S1.

Analysis of sequence data

Coding sequences were read in-frame and aligned using RevTrans v1.4 (Wernersson and Pedersen 2003), taking into account corresponding amino acid alignments. Non-coding sequences were aligned by Muscle implemented in Mega5 (Tamura et al. 2011). Manual editing of alignments was done using Mega5. Sequences of each alignment were trimmed to the same length. Gene sequences that differed from each other by one or more polymorphisms were identified using the unique.seqs command implemented in Mothur v1.3 (Schloss et al. 2009). Sequences were concatenated using BioEdit 7 (http://www.mbio.ncsu.edu/bioedit/bioedit.html) and each unique allelic profile was assigned a sequence type (ST) number.

Population structure and ancestry

Population structure of the 78 strains was inferred using the admixture model with correlated allele frequencies (Pritchard et al. 2000) implemented in STRUCTURE v2.3. Sequence data (3 loci) were formatted using xmfa2struct (http://www.xavierdidelot.xtreemhost.com/clonalframe.htm). STRUCTURE uses a Bayesian clustering framework and assumes that the observed data are derived from *K* ancestral populations (lineages). The admixture model allows for mixed ancestry in more than one of the *K* populations. Five replicate Markov Chain Monte Carlo (MCMC) runs were performed for each value of *K* ranging from 2 to 7 using 100,000 burn-in and 100,000 sampling iterations. A value of *K* was selected on the basis of the ad hoc approach described in the software documentation as well as the additional criterion that an ancestral population must contribute >50 %

genetic material to at least one individual to be recognized. Structure results of the selected *K* were visualized by CLUMPP (Jakobsson and Rosenberg 2007) for label switching.

Phylogenetic analysis

Besides sequences representing the 78 strains from Lake Tai, reference sequences of *Spirodela*, *Landoltia* and *Lemna* species, and *Wolffia globosa* 6919 as outgroup were also retrieved from GenBank and included in phylogenetic analysis (Supplementary Table S3). Maximum-likelihood (ML) phylogenetic analyses were carried out using PhyML v3.0 (Guindon et al. 2010) and the substitution model GTR + G + I, chosen on the basis of the Akaike information criterion implemented in jModelTest2 (Darriba et al. 2012). Settings used in PhyML were as follows: BioNJ starting tree, four substitution rate categories, estimated proportion of invariable sites, and SPR and NNI tree improvement algorithms. ML trees were reconstructed using unique STs for concatenated sequences; 1,000 non-parametric bootstrap replications were used to assess support. Trees were drawn using Mega5 software. Data for 1,000 replicate bootstrap trees from the ML analysis of the three loci sequences were imported into SplitsTree version 4.1 (Huson and Bryant 2006) and used to construct a consensus network graph.

An index, I_{cong}, was used to assess the extent of congruence between ML trees reconstructed for each of the three loci using an online calculator (http://max2.ese.u-psud.fr/icong/). The topological congruence between the trees was assessed using MAST (maximum agreement subtrees) and the index is based on a limit law for the distribution of the size of the MAST when trees are chosen at random (De Vienne et al. 2007). An associated confidence level was also provided by the testing.

Population genetics analysis

Summary statistics for sequences representing the 78 strains from Lake Tai were calculated by DnaSP v5.1 (Librado and Rozas 2009), including G + C content, the number of polymorphic sites (*S*), and the number of haplotypes (*H*). The average number of pairwise nucleotide differences per site (π), number of synonymous substitutions per synonymous site (π_S), number of nonsynonymous substitutions per nonsynonymous site (π_N), and the ratio of nonsynonymous to synonymous substitutions (d*N*/d*S*) were calculated with Jukes–Cantor correction. Tajima's *D* test of neutrality was calculated based on segregating sites. In addition, a phi test implemented in SplitsTree (Huson and Bryant 2006) was used to test whether recombination exists among sequences.

Results

Summary statistics of duckweed populations at Lake Tai

In total, 78 duckweed strains of the genera *Spirodela*, *Landoltia* and *Lemna* from Lake Tai were collected; details of the origins are listed in Supplementary Table S1. PCR amplification of *atpF–atpH*, *matK* and *rpoB* was successfully conducted for all 78 strains and statistics for the nucleotide sequence data are summarized in Table 1. The *rpoB* gene showed a higher G − C content (41 %) relative to the average of the three loci (34 %). The length of *atpF–atpH* spacer varied from 581 to 622 bp, while the alignment was 646 bp in length and included 13 insertions/deletions. These length variations in *atpF–atpH* spacer suggest a low level of conservation, compared to the other two housekeeping genes.

There were five allelic types (*H*) for each locus. Values of d*N*/d*S* for *matK* and *rpoB* loci were <1 indicating that both were subjected to purifying selection. This conclusion is also supported by significant values of the *Z* test statistics rejecting the null hypothesis of d*N* = d*S* (data not shown). Strong purifying selection is consistent with essential functions of housekeeping genes. Tajima's *D* values were significantly different from 0 for all loci (Table 1), suggesting significant departure from a standard neutral model (Feil 2010). Furthermore, the phi test for the 78 sequence data showed *P* values of 0.33, 0.62 and 1.00 for *atpF–atpH*, *matK* and *rpoB*, respectively, indicating no significant recombination was detected.

The three concatenated loci sequences (1,750 bp) were classified into 5 unique sequence types (STs) (Table 2). At Lake Tai, strains of ST1 and ST2 were predominant and accounted for 79.5 % of the total collection, followed by 14.1 % of ST5. Both ST3 and ST4 are under-represented at below 4 %. When compared to reference sequences, two-thirds of the strains had STs nearly identical to the reference strains (Table 2). Sequences of ST1, ST4 and ST5 exhibited 99.9 % similarities to those of reference strains, namely *S. polyrhiza* 7498, *L. minor* 7210 and *L. punctata* 7449. A minority of strains representing ST2 and ST3 showed relatively more divergence (99.3 and 99.8 %, respectively), compared to reference strains *L. aequinoctialis* 6612 or *L. turionifera* 8760. The differences between the two STs and reference strains were contributed by *atpF–atpH* spacer for ST2, and *matK* plus *rpoB* for ST3, respectively.

Population structure

Different STRUCTURE models were explored with *K* (number of ancestral lineages) ranging from 2 to 7. Data for multiple STRUCTURE runs using the admixture model with correlated allele frequencies indicated that *K* = 5 was optimal based on criteria described in Methods. Data for ancestry and admixture levels of the 78 strains are shown in Fig. 1. Strains with >50 % genetic materials from one of the five ancestral lineages are considered to be representative of that lineage. Strains assigned to lineage II were most abundant (38 strains), whereas those in lineage I (2 strains)

Table 1 Summary statistics for loci sequences of the 78 duckweed strains from Lake Tai

| Locus	Length (bp)	GC content (%)	*H*	*S*	π	π_S	π_N	d*N*/d*S*	*D*
atpF–atpH	646	31.6	5	61	0.0395	n.d.	n.d.	n.d.	2.388*
matK	717	32.9	5	79	0.0491	0.1071	0.0338	0.316	3.627***
rpoB	387	40.9	5	20	0.0206	0.0546	0.0105	0.192	2.762**
Concatenated	1,750	34.3	5	160	0.0391	n.d.	n.d.	n.d.	3.168**

Length the length of alignment, *H* number of haplotypes (alleles), *S* number of polymorphic (segregating) sites; nucleotide diversity estimated for all sites (π), synonymous sites (π_S), and nonsynonymous sites (π_N) with Jukes-Cantor correction; d*N*/d*S*, ratio of nonsynonymous to synonymous substitutions, *D* Tajima's *D* based on segregating sites, *n.d* not determined

Significance level: * 0.01 < *P* < 0.05; ** 0.001 < *P* < 0.01; *** *P* < 0.001

Table 2 Frequencies and lineage assignment of strains representing 5 unique sequence types (STs) from Lake Tai

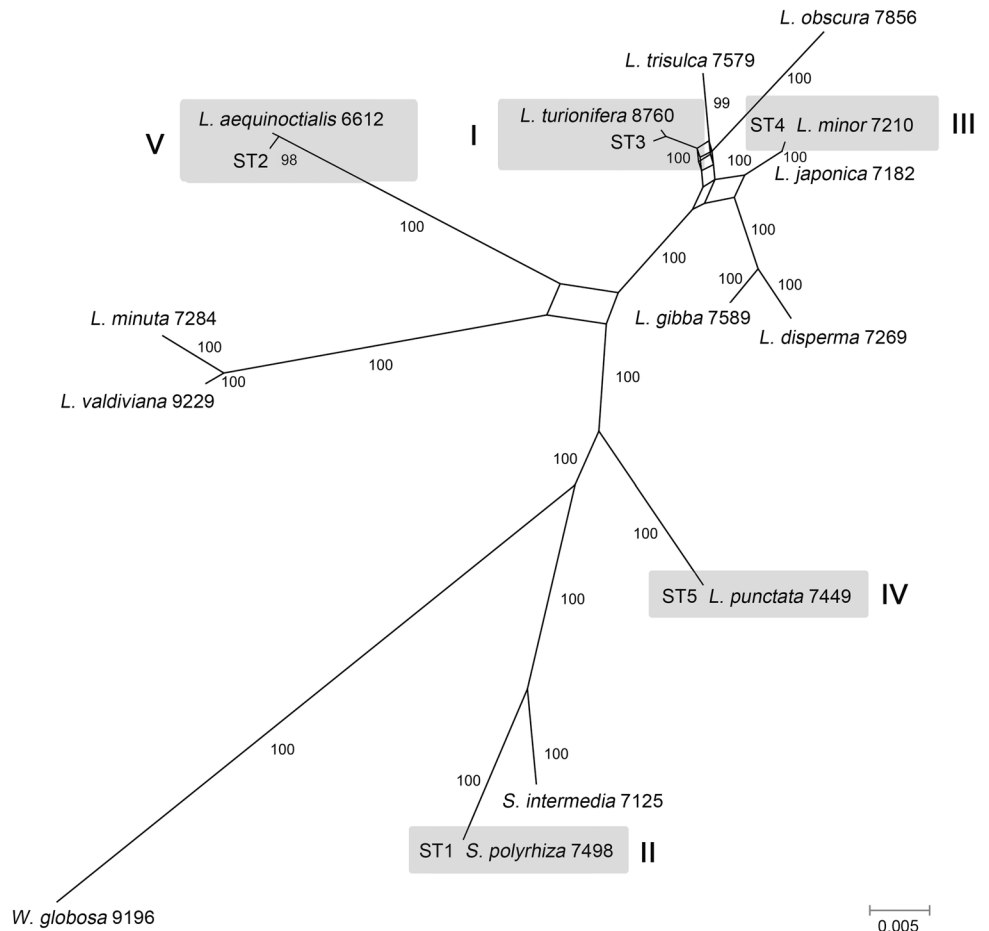| Lineage	Sequence type	Strain counts	Frequency	Closest reference species (similarity)
I	ST3	2	0.026	*Lemna turionifera* 8760 (99.8 %)
II	ST1	38	0.487	*Spirodela polyrhiza* 7498 (99.9 %)
III	ST4	3	0.038	*Lemna minor* 7210 (99.9 %)
IV	ST5	11	0.141	*Landoltia punctata* 7449 (99.9 %)
V	ST2	24	0.308	*Lemna aequinoctialis* 6612 (99.3 %)

were the least. Strains of lineage II, III, IV and V derived 100 % of their genetic material from their respective ancestral lineages, suggesting that very limited interlineage flow of loci had taken place. Only the 2 strains representing lineage I exhibited mixed ancestries derived from lineage III, although for likely a very small percentage (Fig. 1).
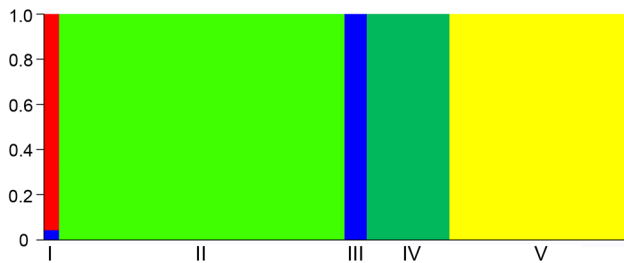
Fig. 1 Ancestries of loci from the 78 strains as inferred by STRUCTURE. Each lineage is represented by a *color block*, whose width is proportional to the number of strains assigned to it: I (*red*; 2), II (*green*; 38), III (*blue*; 3), IV (*dark green*; 11) and V (*yellow*; 24). Vertical lengths of each strain is proportional to each of the 5 inferred lineages

Furthermore, each lineage contained only one sequence type, indicating no genetic variations within lineage.

Phylogeny

The consensus network graph was reconstructed on 1,000 replicate bootstrap trees from ML analysis of concatenated sequences representing strains from Lake Tai and reference species (Fig. 2); the corresponding ML tree is presented in Fig. 3. The network graph showed well-defined clusters of strains corresponding to the five lineages inferred by STRUCTURE. The reference strains were also grouped with high bootstrap support (99–100 %). Using a relatively low value of 0.1 for the consensus network threshold, no reticulation was evident within clusters corresponding to the five lineages. But small reticulations were shown within the genus *Lemna*, particularly on inner edges. Reticulation reflects phylogenetic uncertainty and may indicate recombination events.

As shown in Fig. 2, the nearest neighbor of ST1 strains (lineage II) was *S. polyrhiza* 7498, whereas that of ST5 strains (lineage IV) was *L. punctata* 7449. Lineage I, III and V were classified into genus *Lemna*, and the

Fig. 2 Consensus network graph of concatenated *atpF–atpH + matK + rpoB* partial sequences (1,777 bp) for duckweed reference strains and the 5 unique STs from Lake Tai. *Scale bar* indicates estimated substitutions per site

Fig. 3 Maximum-likelihood tree of concatenated *atpF–atpH* + *matK* + *rpoB* partial sequences (1,777 bp) for duckweed reference strains and the 5 unique STs from Lake Tai. *Scale bar* indicates estimated substitutions per site
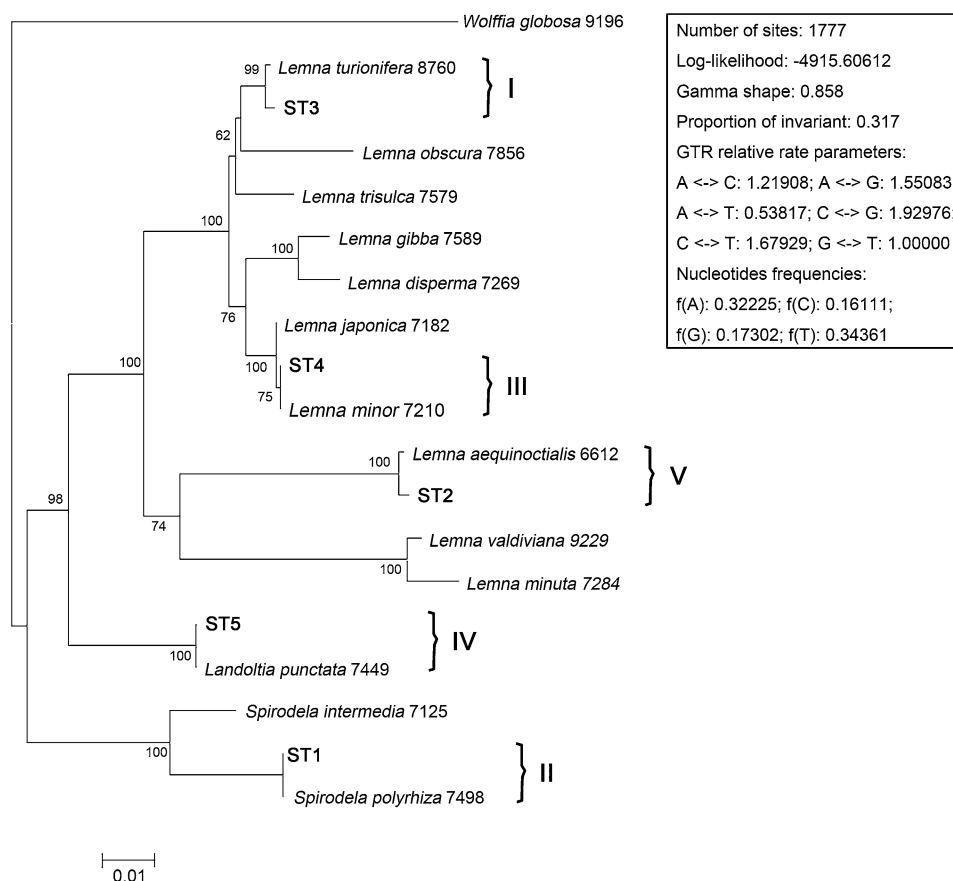
corresponding closest relative of the STs of each lineage was *L. turionifera* 8760, *L. minor* 7210 and *L. aequinoctialis* 6612, respectively.

Further, the I_{cong} index was used to assess the extent of congruence between ML trees reconstructed for each of the three loci (Supplementary Fig. S2). The I_{cong} indexes of all associations between ML trees were statistically significant (Table 3), indicating that the three single locus trees were congruent with each other, and further suggesting that little or no intragenic recombination had occurred. This result is in agreement with the phi test for sequence data including 5 STs and reference strains, which detected no significant recombination at each locus (*P* values of 0.07, 0.28 and 0.41 for *atpF–atpH*, *matK* and *rpoB*, respectively).

Table 3 Statistics of tree congruence index (I_{cong})

Association	No. of taxa	I_{cong}	*P*
atpF–atpH/matK	19	1.262	0.0366*
atpF–atpH/rpoB	19	1.577	0.0004***
matK/rpoB	19	1.419	0.0038**

P, *P* value associated with the method used to test for congruence

Significance level: * $0.01 < P < 0.05$, ** $0.001 < P < 0.01$, *** $P < 0.001$

Discussion

In this study, we investigated natural duckweed populations collected from Lake Tai, China by multilocus sequence typing (MLST) of three loci, followed by STRUCTURE and conventional ML phylogenetic analyses. This is the first time such a framework was used to analyze duckweed species. Our data and analysis validated the framework.

The 78 duckweed strains were divided into five lineages corresponding to *L. turionifera* (lineage I), *L. aequinoctialis* (lineage II), *L. minor* (lineage III), *L. punctata* (lineage IV) and *S. polyrhiza* (lineage V). No genetic variation was observed within lineages (Fig. 1; Table 2), and interlineage transfer of loci occurred infrequently as suggested by the STRUCTURE analysis (Fig. 1). This observation was supported by both the congruence among ML trees (Table 3; Supplementary Fig. S2) and the Phi tests. These data provide supporting evidence to the prevailing hypothesis that duckweeds primarily inherit genetic materials by asexual cloning. Although rare frequency of flowering and seed setting may increase the level of genotypic and genetic diversity (Vasseur et al. 1993; Santamaría 2002; Stomp 2005), it appears that genetic exchanges, such as recombination and gene flow, are limited for duckweed populations in the wild. Some thought somatic mutations may provide the

main source of genetic variations in clonal species with rare periods of sexual reproduction (Orive 2001); however, such variations were not observed in the loci studied in this work. It has also been hypothesized that influx of new genotypes by zoochory may contribute to the genetic diversification of natural duckweed populations (Silvertown 2008).

The paucity of variations within lineages cannot be completely explained as the results of low mutation rates, however. We hypothesize that the genetic landscape could be shaped by several possible processes. First, the low level of diversity is often associated with genetic bottlenecks (Chiang et al. 2001; Huang et al. 2001), as evidenced by our data which exhibited significant deviations from the neutral model by the Tajima's *D* test (Table 1). Genetic bottlenecks likely had led to the fixation of dominant alleles within the population and contributed to depletion of genetic diversity (Tajima 1983). Second, the clonal reproduction mode usually leads to much smaller sizes of effective population (Gliddon et al. 1987), although census sizes of duckweeds are often in the order of millions of individuals. Furthermore, fluctuations in population size due to seasonal changes in the wild (Landolt 1986) may also affect the genetic variability. Third, due to clonal reproduction, linkage disequilibria increases the "hitchhiking" effects where selection on a single locus lead to selection on its adjacent linked loci as well (Jordan et al. 1996; Maruyama and Birky 1996), thus decreasing genetic diversity. It is also suggested that selection may favor clonal organisms to generate a single genotype that is competitive under diverse environmental conditions (Lynch 1984).

Our data also point to remarkable genetic stability in these wild duckweed strains. Based on the three genetic markers analyzed, these strains from sea level in Eastern China are almost identical to the reference strains (Table 2), which were from diverse geographic locations including North America, Europe, Asia (in India), and Africa, and from different altitudes ranging from sea level to more than 1,000 m above sea level (Supplementary Table S3). This stability holds for a more extreme case of the ST3/lineage I (99.8 % similarities) and its reference strain, *Lemna turionifera* 8760, which was collected from the Czech Republic at an altitude of 1,239 m. Interestingly, the small variations of ST3 were mostly contributed by the two housekeeping gene markers *matK* and *rpoB* rather than the more variable intergenic *atpF–atpH* marker. Furthermore, the two strains representing lineage I exhibited mixed ancestries with lineage III (Fig. 1). This suggests that the genetic materials exchange between lineages I and III might have occurred before their separate dispersions on Earth. This hypothesis will need to be investigated carefully in future work using molecular clock theories.

In summary, the genetic structure of natural duckweed populations might be shaped by complex interactions

between gene flow, genetic drift and various selection pressures (Jordan et al. 1996; Wang et al. 2009). However, given their primary reproduction mechanism of asexual cloning, genetic conservation and stability appear to be the main theme for duckweed during their evolutionary history.

The three genetic markers used in this study were selected based on their successful applications in phylogenetic studies of duckweed (Cabrera et al. 2008; Wang et al. 2010; Wang and Messing 2011). Although previous study showed that *atpF–atpH* was a universal barcode for species-level identification of duckweed, its sequence polymorphisms might not be sufficient to distinguish closely related species, such as *L. minor*/*L. japonica*, *L. minuta*/*L. valdiviana* (Wang et al. 2010). Further, the significant length variations of the intergenic *atpF–atpH*, caused by deletion/insertion, simple sequence repeats and rearrangements, were also problematic for accurate alignment and might result in incorrect phylogenies. To remedy these drawbacks, analyses of more markers are needed to provide a better understanding of the genetic relationships (Pennisi 2007; Fazekas et al. 2008; Chase and Fay 2009). The two housekeeping genes, *matK* and *rpoB*, were widely used and sufficiently variable for phylogenetic study (Hilu et al. 2003; Chase et al. 2005). Owing to more functional constrains and higher selection pressures as compared to *atpF–atpH*, the housekeeping genes should complement *atpF–atpH*, and they together should provide a more complete framework for the analyses. Using the three loci, the MLST method definitively classified the 78 strains from Lake Tai into five well-separated lineages (Figs. 1, 2). In addition, the network graph based on 1,000 ML trees of concatenated sequence of the three loci gave a cleaner delineation of each species supported by high bootstrap values (Fig. 2), compared to the ML trees inferred by any single locus (Supplementary Fig. S2). These results further proved that MLST might be an effective tool for distinguishing closely related species.

The markers used in this study were all from chloroplast, which possesses distinct patterns of inheritance compared to the nuclear genome (Petit et al. 2005). This potentially poses a challenge since results generated with chloroplast markers could misrepresent genetic information in the nuclear genome. It is possible that incompatible signals exist among markers from the nuclear genome. However, most previous studies indicated that great conservation and insufficient variations were found among nuclear markers in duckweeds (Les et al. 2002; Bog et al. 2010, 2013; Xue et al. 2012). In few cases, such as incongruent events in *W. globosa*, genetic exchange might have occurred among individuals (Yuan et al. 2011), and its mechanisms remain unknown.

Based on phylogenetic analysis of these three markers, we conclude that the collected duckweed strains from

Lake Tai were separated into five species out of three genera (Fig. 2). All identified species of the three genera (Supplementary Table S3), except for *L. perpusilla*, *L. tenera* and *L. yungensis* (sequence data unavailable), were included and placed on well-supported branches (Fig. 2). This analysis framework was helpful in answering several questions concerning the phylogeny of duckweed. One is on classification between *L. minor* and *L. gibba*, which had been an impossible task based on morphological data (De Lange 1975; Kandeler 1975; Landolt 1975). The distinction between the two is essential for their utilizations for different purposes: *L. gibba* is preferred in the OECD-based investigations (Brain and Solomon 2007) whereas *L. minor* is often used for bio-tests based on the ISO 20079 protocol (Naumann et al. 2007). In this report, our analyses provided unequivocal classification of the two species.

Another question is regarding whether *L. japonica* originated from the hybridization of *L. minor* and *L. turionifera*. Previous attempts failed to reproduce the hybridization event (Landolt 1986), mainly due to the rare events of flowering and the small size of flower which made outcrossing extremely difficult (Stomp 2005). In this report, we found that *L. japonica* is phylogenetically much more closely related to strains of *L. minor* than to *L. turionifera* (Fig. 2; Supplementary Fig. S2). Furthermore, our STRUCTURE analysis also showed that representative strains of lineage I (*L. turionifera*) has a small admixture from those of lineage III (*L. minor*) (Fig. 1), and reticulations were evident on the inner edges of these branches (Figs. 2, 3), suggesting genetic flow or recombination events within *Lemna* genus. Taken together, our data suggest that *L. japonica* is not a new species from a cross of *L. minor* and *L. turionifera* (Landolt 1986), but a geographical subspecies of *L. minor* in lineage III, much similar to the case of ST3 and lineage I (Figs. 2, 3). However, it should be cautioned that such kind of small admixtures sometimes could result from data artifacts. Thus this hypothesis requires more supporting evidence when sequences of multiple *L. japonica* strains become available.

In conclusion, our results from MLST analyses not only established a solid basis for characterization of *Landoltia*, *Spirodela* and especially *Lemna* strains at the species level, but also provided insights into the genetic structure of natural duckweed populations. Moreover, our work highlighted the potential of using duckweed as a model system to study population genetics and the effects of asexual cloning on genetic diversity and evolution. The framework used in this work, combined with sequencing of more genes and eventually genome-wide single nucleotide polymorphism typing in duckweeds, should prove to be valuable tools for future studies in this field.

References

Appenroth KJ, Borisjuk N, Lam E (2013) Telling duckweed apart: genotyping technologies for the Lemnaceae. Chin J Appl Environ Biol 19:1–10

Bog M, Baumbach H, Schween U, Hellwig F, Landolt E, Appenroth KJ (2010) Genetic structure of the genus *Lemna* L. (Lemnaceae) as revealed by amplified fragment length polymorphism. Planta 232:609–619

Bog M, Schneider P, Hellwig F, Sachse S, Kochieva E, Martyrosian E, Landolt E, Appenroth KJ (2013) Genetic characterization and barcoding of taxa in the genus *Wolffia* Horkel ex Schleid. (Lemnaceae) as revealed by two plastidic markers and amplified fragment length polymorphism (AFLP). Planta 237:1–13

Brain RA, Solomon KR (2007) A protocol for conducting 7-day daily renewal tests with *Lemna gibba*. Nat Protocols 2:979–987

Cabrera LI, Salazar GA, Chase MW, Mayo SJ, Bogner J, Dávila P (2008) Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. Am J Bot 95:1153–1165

Chase MW, Fay MF (2009) Barcoding of plants and fungi. Science 325:682–683

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V (2005) Land plants and DNA barcodes: short-term and long-term goals. Philos Trans R Soc Lond B Biol Sci 360:1889–1895

Cheng JJ, Stomp AM (2009) Growing duckweed to recover nutrients from wastewaters and for production of fuel ethanol and animal feed. Clean Soil Air Water 37:17–26

Chiang TY, Chiang YC, Chen YJ, Chou CH, Havanond S, Hong TN, Huang S (2001) Phylogeography of *Kandelia candel* in east asiatic mangroves based on nucleotide variation of chloroplast and mitochondrial DNAs. Mol Ecol 10:2697–2710

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Meth 9:772

De Lange L (1975) Gibbosity in the complex *Lemna gibba/Lemna minor*: literature survey and ecological aspects. Aquat Bot 1:327–332

De Vienne DM, Giraud T, Martin OC (2007) A congruence index for testing topological similarity between trees. Bioinformatics 23(23):3119–3124. doi:10.1093/bioinformatics/btm500

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SCH (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3:e2802

Feil EJ (2010) Linkage, selection, and the clonal complex. In: Robinson DA, Falush D, Feil EJ (eds) Bacterial population genetics in infectious disease. John Wiley & Sons, Hoboken, pp 19–35. doi:10.1002/9780470600122.ch2

Gliddon C, Belhassen E, Gouyon PH (1987) Genetic neighbourhoods in plants with diverse systems of mating and different patterns of growth. Heredity 59:29–32

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate

maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321

Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS, Chatrou LW (2003) Angiosperm phylogeny based on *matK* sequence information. Am J Bot 90:1758–1776

Huang S, Chiang YC, Schaal BA, Chou CH, Chiang TY (2001) Organelle DNA phylogeography of *Cycas taitungensis*, a relict species in Taiwan. Mol Ecol 10:2669–2681

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806

Jordan WC, Courtney MW, Neigel JE (1996) Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA Locus in North American duckweeds (Lemnaceae). Am J Bot 83:430–439

Kandeler R (1975) Species delimitation in the genus *Lemna*. Aquat Bot 1:365–376

Landolt E (1975) Morphological differentiation and geographical distribution of the *Lemna gibba*, *Lemna minor* group. Aquat Bot 1:345–363

Landolt E (1986) The family of Lemnaceae: a monographic study, vol 1. Veroffentlichungen des Geobotanischen Institutes Stiftung Rubel, ETH, Zurich

Landolt E, Kandeler R (1987) The family of Lemnaceae: a monographic study. Phytochemistry, physiology, application, bibliography. Veröffentlichungen des Geobotanischen Institutes ETH Stiftung Rubel, Zurich

Les D, Landolt E, Crawford DJ (1997) Systematics of the Lemnaceae (duckweeds): inferences from micromolecular and morphological data. Plant Syst Evol 204:161–177

Les DH, Crawford DJ, Landolt E, Gabel JD, Kimball RT (2002) Phylogeny and systematics of Lemnaceae, the duckweed family. Syst Bot 27:221–240

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Lynch M (1984) Destabilizing hybridization, general-purpose genotypes and geographic parthenogenesis. Q Rev Biol 59:257–290

Maheshwari S, Gupta S (1967) Induction of flowering in *Lemna paucicostata*, a short-day plant, by chelating agents and iron. Planta 77:95–98

Maiden MC (2006) Multilocus sequence typing of bacteria. Annu Rev Microbiol 60:561–588

Marinc-Hrzenjak V, Kristl J, Krajncic B (2008) The effects of jasmonic acid and ethylene-di-o-hydroxyphenylacetic acid on floral induction and induction of turions in *Spirodela polyrrhiza* (L.) Schleiden. Acta Botanica Croatica 67:131–138

Maruyama T, Birky JRC (1996) Effects of periodic selection on gene diversity in organelle genomes and other systems without recombination. Genet 127:449–451

Miyawaki T, Matsumoto S, Takahashi W, Tanaka O (2013) Effect of heat-treated noradrenaline on flowering in *Lemna*. Biosci Biotechnol Biochem 77:1586–1588

Naumann B, Eberius M, Appenroth KJ (2007) Growth rate based dose-response relationships and EC-values of ten heavy metals using the duckweed growth inhibition test (ISO 20079) with *Lemna minor L.* clone St. J Plant Physiol 164:1656–1664

Orive ME (2001) Somatic mutations in organisms with complex life histories. Theor Popul Biol 59:235–249

Pennisi E (2007) Wanted: a barcode for plants. Science 318:190–191

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 14:689–701

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genet 155:945–959

Purves WK (1961) Dark reactions in the flowering of *Lemna perpusilla* 6746. Planta 56:684–690

Reid MS, Bieleski RL (1970) Response of *Spirodela oligorrhiza* to phosphorus deficiency. Plant Physiol 46:609–613

Santamaría L (2002) Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. Acta Oecologica 23:137–154

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microb 75:7537–7541

Seth PN, Venkataraman R, Maheshwari SC (1970) Studies on the growth and flowering of a short-day plant, *Wolffia microscopica*. Planta 90:349–359

Shimakawa A, Shiraya T, Ishizuka Y, Wada KC, Mitsui T, Takeno K (2012) Salicylic acid is involved in the regulation of starvation stress-induced flowering in *Lemna paucicostata*. J Plant Physiol 169:987–991

Silvertown J (2008) The evolutionary maintenance of sexual reproduction: evidence from the ecological distribution of asexual reproduction in clonal plants. Int J Plant Sci 169:157–168

Stomp AM (2005) The duckweeds: a valuable plant for biomanufacturing. Biotechnol Annu Rev 11:69–99

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genet 105:437–460

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Vasseur L, Aarssen LW, Bennett T (1993) Allozymic variation in local apomictic populations of *Lemna minor* (Lemnaceae). Am J Bot 80:974–979

Wang W, Messing J (2011) High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. PLoS ONE 6:e24670

Wang R, Ai B, Gao BQ, Yu S, Li YY, Chen XY (2009) Spatial genetic structure and restricted gene flow in a functionally dioecious fig, *Ficus pumila* L. var. *pumila* (Moraceae). Popul Ecol 51:307–315

Wang W, Wu Y, Yan Y, Ermakova M, Kerstetter R, Messing J (2010) DNA barcoding of the Lemnaceae, a family of aquatic monocots. BMC Plant Biol 10:205

Wernersson R, Pedersen AG (2003) RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res 31:3537–3539

Xu J, Cui W, Cheng JJ, Stomp AM (2011) Production of high-starch duckweed and its conversion to bioethanol. Biosyst Eng 110:67–72

Xu J, Zhao H, Stomp AM, Cheng JJ (2012) The production of duckweed as a source of biofuels. Biofuels 3:589–601

Xue H, Xiao Y, Jin Y, Li X, Fang Y, Zhao H, Zhao Y, Guan J (2012) Genetic diversity and geographic differentiation analysis of duckweed using inter-simple sequence repeat markers. Mol Biol Rep 39:547–554

Yuan JX, Pan J, Wang BS, Zhang DM (2011) Genetic differentiation of *Wolffia globosa* in China. J Syst Evol 49:509–517