scientific reports

OPEN

Molecular confirmation of the hybrid origin of *Sparganium longifolium* (Typhaceae)

Yinjiao Yu[1,3], Fengxia Li[1,3], Eugeny A. Belyakov[2], Weidong Yang[1], Alexander G. Lapirov[2]✉ & Xinwei Xu [1]✉

Sparganium longifolium* was reported as a hybrid between *S. emersum* and *S. gramineum* based on its intermediate type or the common characteristics of its parent species. Its hybrid origin needs to be confirmed using molecular technology. We investigated the origin of *S. longifolium* based on 10 populations of *S. emersum*, *S. gramineum* and *S. longifolium* from five lakes in European Russia, using sequences of six nuclear loci and one chloroplast DNA fragment. Haplotype network, principal coordinate analysis and genetic clustering based on data of nuclear loci confirmed that *S. longifolium* is the hybrid between *S. emersum* and *S. gramineum*. We found that the natural hybridization between *S. emersum* and *S. gramineum* is bidirectional but asymmetrical, and the latter mainly acts as maternal species. We also found that all samples of *S. longifolium* were F1 generations, and thus hypothesized that *S. emersum* and *S. gramineum* could likely maintain their species boundary through the post-zygote reproductive isolation mechanism of F1 generation sterility.

Sparganium L. (Typhaceae) is an ecologically important group of aquatic plants, comprising about 14 species, widely distributed in the temperate and cold regions of the northern hemisphere[1,2]. Natural hybridization between *Sparganium* species is common. Cook and Nicolls reviewed previous studies and listed different interspecific hybrids in *Sparganium*. All these studies identified hybrids using morphological characteristics. Due to phenotypic plasticity and subtle morphological differences, however, hybrids are often difficult to distinguish from their parents[3]. Therefore, molecular identification is an essential method for the study of natural hybridization in *Sparganium*. Recently, the presence of hybrids, such as *S. angustifolium* Michx. × *S. emersum* Rehmann, *S. fallax* Graebn. × *S. japonicum* Rothert, *S. acaule* Rydb. × *S. fluctuans* B.L.Rob. and *S. glomeratum* Laest. ex Beurl. × *S. gramineum* Georgi was verified based on DNA sequences in two phylogenetic studies[3,4].

Sparganium longifolium Turcz. was reported as a hybrid between *S. emersum* and *S. gramineum* based on their morphological characteristics[1]. This hybrid occurs commonly in regions where both parents grow together and often occupies habitats where *S. gramineum* is absent due to its high adaptability under eutrophic conditions[1]. The hybrid origin of *S. longifolium* was emphasized by[5] based on their detailed biomorphological investigation. The hybrid was considered fertile[1] and backcrossing with its parent species was used to explain the phenomenon that some populations of *S. longifolium* were rich in terate forms[5]. So far, only one case study at the molecular level has been conducted for *S. longifolium*. Belyakov et al. sequenced the internal transcribed spacer and found similar and identical ribotypes in *S. emersum*, *S. longifolium*, *S. gramineum* and *S. hyperboreum*, which did not provide directly molecular evidence to clarify the origin of *S. longifolium*[6]. Further molecular studies are still necessary to confirm the hybrid origin of *S. longifolium*. In addition, the direction of hybridization is unknown. If a bidirectional hybridization exists, two distinct life forms of *S. longifolium*, emergent and floating-leaved, which are similar to *S. emersum* and *S. gramineum* respectively, are likely correlated to the direction of hybridization. All these hypotheses need to be verified using molecular technology.

In this study, we collected samples of *S. emersum*, *S. gramineum* and *S. longifolium* from five lakes in European Russia, using sequences of six nuclear loci and one chloroplast DNA fragment to (1) test whether *S. longifolium* is the hybrid between *S. emersum* and *S. gramineum*, and (2) detect the direction of hybridization. The study will deepen our understanding of interspecific hybridization in *Sparganium*.

[1]National Field Station of Freshwater Ecosystem of Liangzi Lake, College of Life Sciences, Wuhan University, Wuhan 430072, China. [2]Papanin Institute for Biology of Inland Waters, Russian Academy of Sciences, Borok, Nekouz District, Yaroslavl Region, Russia 152742. [3]These authors contributed equally: Yinjiao Yu and Fengxia Li. ✉email: a_lapir@ibiw.ru; xuxw@whu.edu.cn
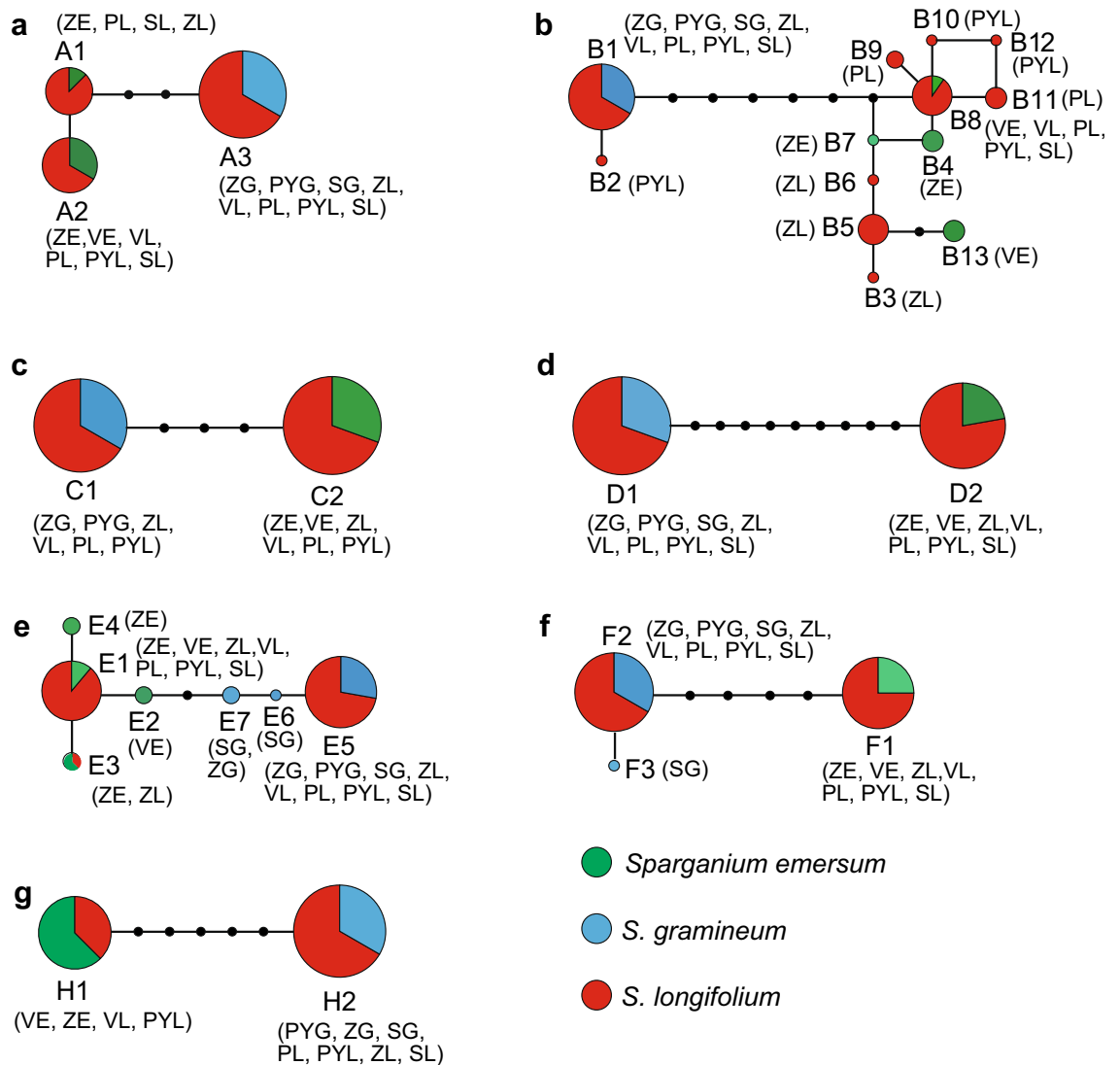
1

Figure 1. Haplotype networks of six nuclear loci Tran05 (**a**), Tran57 (**b**), Tran59 (**c**), Tran66 (**d**), Tran83 (**e**), and Tran93 (**f**), and cpDNA fragment *trn*H-*psb*A (**g**). Node size is proportional to the number of each haplotype. Small black circles represent unsampled or hypothetical haplotypes. Population codes indicated beside haplotypes are the same as Table 1.

Results

Sequence variations.   Sequences of the six nuclear loci were obtained from 85, 83, 62, 85, 84 and 86 individuals, respectively. Their aligned lengths were 579, 492, 501, 465, 326 and 413 bp with 4, 14, 4, 10, 7 and 6 variable sites, respectively (Supplementary Table S1). The numbers of haplotypes at the six nuclear loci were 3, 13, 2, 2, 7 and 3, respectively. The haplotype networks of the six nuclear loci showed the same pattern: haplotypes of *S. emersum* and *S. gramineum* separated well and formed two clades, while all individuals of *S. longifolium* were heterozygous and consisted of two alleles from different haplotype clades (Fig. 1). Of the six nuclear loci, only at the Tran57 locus *S. longifolium* had private haplotypes (Fig. 1).

Sequences of *trn*H-*psb*A were obtained from 82 individuals, including 15 of *S. emersum*, 20 of *S. gramineum* and 47 of *S. longifolium*. The aligned length was 670 bp with four substitutions and two 4-bp indels. Sequences of all samples collapsed into two haplotypes H1 and H2 corresponding to *S. emersum* and *S. gramineum*, respectively (Fig. 1g). In *S. longifolium*, nine individuals from VL and PYL populations shared H1 with *S. emersum* and 38 individuals from PL, PYL, ZL and SL populations shared H2 with *S. gramineum* (Fig. 1g).

Genetic grouping.   The PCoA analysis revealed that all samples were divided into three groups on principal coordinate 1, which explained 38.73% of the total variation. The three groups corresponded to *S. emersum*, *S. gramineum* and *S. longifolium*, respectively, and the hybrid group was located between two parent groups (Fig. 2a). STRUCTURE analysis suggested K = 2 as the optimal number of clusters based on the value of ΔK (Supplementary Fig. S1) and inferred two genetic clusters that consisted of *S. emersum* and *S. gramineum* respectively and genetic admixture for all samples of *S. longifolium* with intermediate admixture coefficient (0.3863–0.5279, Fig. 2b).
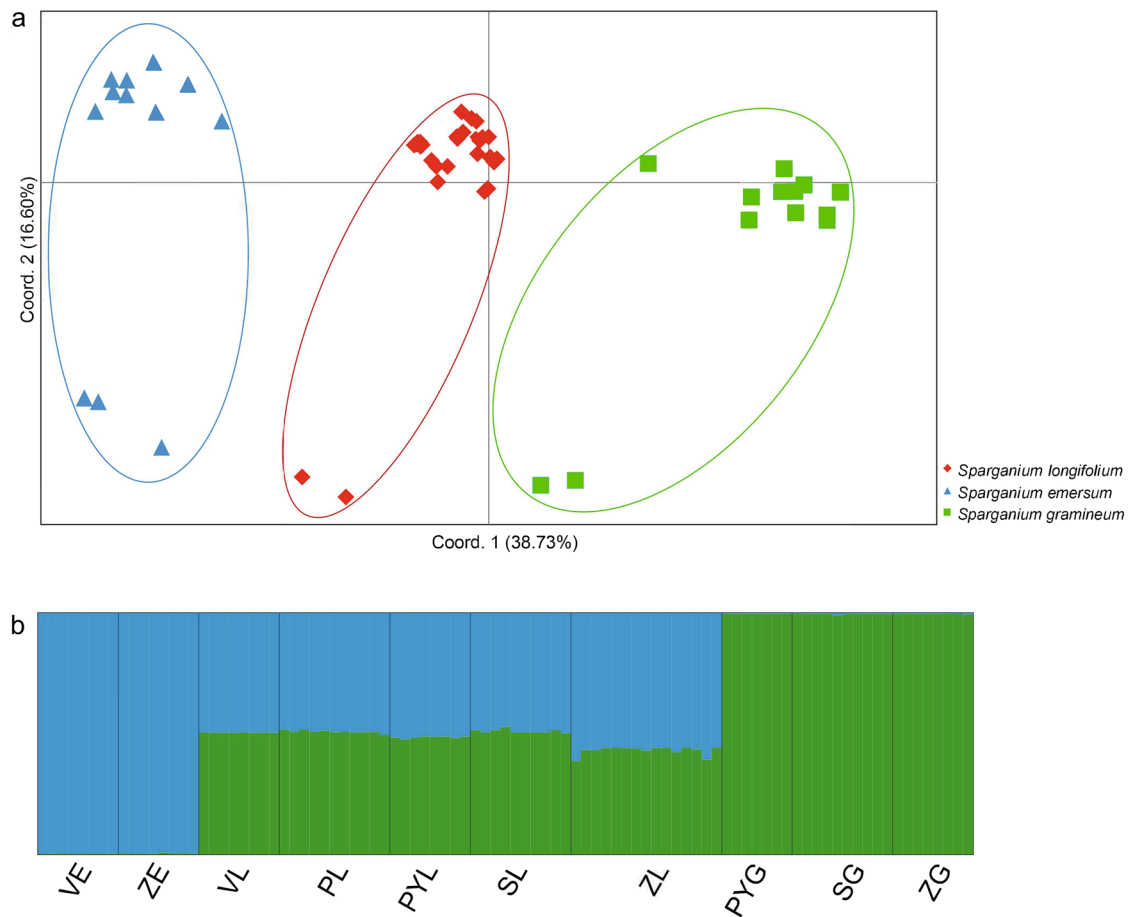
Figure 2. Genetic clustering for 10 populations of *Sparganium emersum*, *S. longifolium* and *S. gramineum* using principal coordinate analysis (**a**) and STRUCTURE (**b**) based on six nuclear loci. In (**b**), blue and green represent the geneic clusters of *S. emersum* and *S. gramineum*, respectively, and a single vertical bar displays the membership coefficients of each individual. Population codes follow Table 1.

Discussion

The haplotype networks, PCoA analysis and STRUCTURE analysis based on the six nuclear loci confirm that *S. longifolium* is a hybrid between *S. emersum* and *S. gramineum*, providing molecular support for previous morphological analyses[5]. Furthermore, all individuals with intermediate admixture coefficient (Fig. 2b) and private haplotypes only present in one out of six nuclear loci (Fig. 1) suggest that *S. longifolium* is most likely a F1 hybrid. We thus hypothesized that *S. emersum* and *S. gramineum* could likely maintain their species boundary through the post-zygote reproductive isolation mechanism of F1 generation sterility. This hypothesis is possible based on the observations from hybrids in European Russia. The pollen viability was checked in *S. longifolium* samples from Vysokovskoe Lake and Sabro Lake, and the vast majority of checked pollens were sterile[5]. In addition, flowering plants of *S. longifolium* often do not form seeds, or the seeds are puny and significantly inferior to normal seeds in size[5]. However, the hypothesis is only based on our limited sampling, which is contrary to the conclusion inferred from morphological characteristics that it is fertile and may backcross with parental species[1]. Further studies with extensive sampling are necessary to test our hypothesis.

The chloroplast DNA fragment *trn*H-*psb*A was used to infer the direction of hybridization between *S. emersum* and *S. gramineum* because chloroplast DNA is maternal inheritance in *Sparganium*[3,4]. The hybrid *S. longifolium* shared haplotypes with *S. emersum* and *S. gramineum* simultaneously (Fig. 1). This finding clearly indicates that bidirectional hybridization exists between *S. emersum* and *S. gramineum*. At the same time, the different frequency of these two haplotypes in the hybrid (H1, 19.1% vs. H2, 80.9%) means that the direction of hybridization is asymmetric. A variety of factors can lead to asymmetry in natural hybridization, such as flowering time, preference of pollinators, quality and quantity of pollen, cross incompatibility and the abundance of parent species[7,8]. Rare species usually act as maternal species relative to abundant species[9,10]. *S. gramineum* is confined to oligotrophic lakes and its abundance is obviously lower than that of *S. emersum*[1,11]. The relatively scarcity combined with the ecology of *S. gramineum* make it more often act as maternal species when hybridizing with *S. emersum*.

As described by[5], the morphological diversification of *S. longifolium* was also observed in this study. For example, individuals of *S. longifolium* with emergent and floating-leaved life forms occur concurrently in Zaozer'ye Lake (Supplementary Fig. S2). However, all individuals had the same haplotype H2 as *S. gramineum* (Fig. 1),

Code	Species	n	Collection no.	Location	Latitude/longitude	Collection date
VE	*S. emersum*	8	Be1	Vysokovskoe Lake, Ivanovo	57.1726/40.9405	2018/8/31
VL	*S. longifolium*	8	Be2	Vysokovskoe Lake, Ivanovo	57.1726/40.9405	2018/8/31
PL	*S. longifolium*	11	Be4	Polevo Lake, Ivanovo	56.5509/41.5945	2018/9/1
PYG	*S. gramineum*	7	Be5	Pyrskoe Lake, Nizhny Novgorod	56.3970/43.3074	2018/9/7
PYL	*S. longifolium*	8	Be6	Pyrskoe Lake, Nizhny Novgorod	56.3970/43.3074	2018/9/7
ZG	*S. gramineum*	8	Ru114	Zaozer'ye Lake, Yaroslavl	56.8248/39.3571	2018/8/21
ZL	*S. longifolium*	15	Ru115	Zaozer'ye Lake, Yaroslavl	56.8248/39.3571	2018/8/21
ZE	*S. emersum*	8	Ru116	Zaozer'ye Lake, Yaroslavl	56.8248/39.3571	2018/8/21
SG	*S. gramineum*	10	Ru096	Sabro Lake, Tver	57.1634/32.9105	2018/8/17
SL	*S. longifolium*	10	Ru097	Sabro Lake, Tver	57.1634/32.9105	2018/8/17

Table 1. Sampling sites and number of samples for *Sparganium emersum*, *S. longifolium* and *S. gramineum*.

suggesting that the direction of hybridization do not determine life form of *S. longifolium*. In addition, all individuals of *S. longifolium* sampled here are likely F1 hybrid. Their variable phenotypes could not be associated with traits segregation due to F2 generation or backcross. Detailed ecological investigation combining with research at the genomic level are essential to find out the potential factors leading to morphological diversification of *S. longifolium*.

Here, using sequences of six nuclear loci and one chloroplast DNA fragment, we confirmed that *S. longifolium* is the hybrid between *S. emersum* and *S. gramineum*. The natural hybridization between *S. emersum* and *S. gramineum* is bidirectional but the latter mainly acts as maternal species. We also found that all samples of *S. longifolium* were F1 generations, indicating that *S. emersum* and *S. gramineum* could maintain their species boundary through the post-zygote reproductive isolation mechanism of F1 generation sterility.

Methods

Sample collection and DNA extraction. A total of 93 individuals from 10 populations of *S. emersum*, *S. gramineum* and *S. longifolium* were collected from five lakes in European Russia (Table 1). Individuals of each population were collected randomly at intervals of at least 10 m. The collection of plant materials was approved by Papanin Institute for Biology of Inland Water (IBIW), Russian Academy of Sciences. Voucher specimens were kept in the herbarium of IBIW and identified by Dr. Eugeny A. Belyakov (IBIW). Fresh leaves were sampled and dried in silica gel for subsequent DNA extraction. Total genomic DNA was extracted using the DNA Secure Plant Kit (Tiangen Biotech, Beijing, China) following the manufacturer's protocol.

Amplification, sequencing and cloning. We sequenced one chloroplast DNA fragment *trn*H-*psb*A[12] and six nuclear loci developed from transcriptome sequences of *S. fallax* (Supplementary Table S1). PCR reactions, sequencing and cloning were performed following[13]. All individuals of *S. longifolium* were heterozygous at the six nuclear loci with multi-point mutations or insertions/deletions, and their alleles were obtained by cloning. Sequence data were aligned using MAFFT v7.3.1[14].

Data analyses. Haplotypes of each locus were identified using DNASP v5.0[15]. The obtained haplotypes were deposited in GenBank (see Supplementary Table S1 for accession numbers). A median-joining network[16] to interpret relationships among haplotypes of each locus was generated using NETWORK v4.0 (http://www.fluxus-engineering.com). The following analyses were performed based on the dataset of six nuclear loci. The genetic lineage proportion of each individual was identified using a Bayesian clustering method implemented in STRUCTURE v2.4[17]. We performed 10 replicate runs with a burn-in period of 20,000 iterations and 100,000 Markov Chain Monte Carlo (MCMC) iterations under the admixture model at the number of clusters from one to eight. Principal coordinate analysis (PCoA) implemented in GenALEx v6.5[18], was also used to examine the genetic clusters of all individuals.

All methods were performed in accordance with the relevant guidelines and regulations.

Data availability

The data that support the findings of this study are available in the GenBank with accession numbers ON015932-ON015962, [https://www.ncbi.nlm.nih.gov/].

References

1. Cook, C. D. K. & Nicholls, M. S. A. Monographic study of the genus *Sparganium* (Sparganiaceae). I: Subgenus Xanthosparganium Holmberg. *Bot. Helv.* **96**, 213–267 (1986).
2. Cook, C. D. K. & Nicholls, M. S. A. Monographic study of the genus *Sparganium* (Sparganiaceae). II: Subgenus Sparganium. *Bot. Helv.* **97**, 1–44 (1987).

3. Sulman, J. D., Drew, B. T., Drummond, C., Hayasaka, E. & Sytsma, K. J. Systematics, biogeography, and character evolution of *Sparganium* (Typhaceae): Diversification of a widespread, aquatic lineage. *Am. J. Bot.* **100**, 2023–2039. https://doi.org/10.3732/ajb.1300048 (2013).
4. Ito, Y., Tanaka, N., Kim, C., Kaul, R. B. & Albach, D. C. Phylogeny of *Sparganium* (Typhaceae) revisited: Non-monophyletic nature of *S. emersum* sensu lato and resurrection of *S. acaule*. *Plant Syst. Evol.* **302**, 129–135. https://doi.org/10.1007/s00606-015-1245-7 (2016).
5. Belyakov, E. A., Shcherbakov, A. V., Lapirov, A. G. & Shilov, M. P. Morphology and ecological characteristics of *Sparganium × longifolium* (Typhaceae) in the Central part of European Russia. *Biosyst. Divers.* **25**, 154–161. https://doi.org/10.15421/011723 (2017).
6. Belyakov, E. A., Machs, E. M., Mikhailova, Y. V. & Rodionov, A. V. The study of hybridization processes within genus *Sparganium* L. subgenus *Xanthosparganium* Holmb. based on data of Next Generation Sequencing (NGS). *Ecol. Genet.* **17**(4), 27–35. https://doi.org/10.17816/ecogen17427-35 (2019).
7. Carney, S. E., Gardner, K. A. & Rieseberg, L. H. Evolutionary changes over the fifty-year history of a hybrid population of sunflowers (*Helianthus*). *Evolution* **54**, 462–474. https://doi.org/10.1111/j.0014-3820.2000.tb00049.x (2000).
8. Muranishi, S., Tamaki, I., Setsuko, S. & Tomaru, N. Asymmetric introgression between *Magnolia stellata* and *M. salicifolia* at a site where the two species grow sympatrically. *Tree Genet. Genomes* **9**, 1005–1015. https://doi.org/10.1007/s11295-013-0612-1 (2013).
9. Lepais, O. *et al.* Species relative abundance and direction of introgression in oaks. *Mol. Ecol.* **18**, 2228–2242. https://doi.org/10.1111/j.1365-294X.2009.04137.x (2009).
10. Zhang, N., Yu, J., Wang, Y. & Gong, X. Molecular evidence for asymmetric hybridization in three closely related sympatric species. *AoB Plants* **10**, ply011. https://doi.org/10.1093/aobpla/ply011 (2018).
11. Belyakov, E. A. & Lapirov, A. G. Morphological and ecological cenotic features of the relict species *Sparganium gramineum* Georgi (Typhaceae) in waterbodies of European Russia. *Inland Water Biol.* **11**, 417–424. https://doi.org/10.1134/S199508291804003X (2018).
12. Shaw, J. *et al.* The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* **92**, 142–166. https://doi.org/10.3732/ajb.92.1.142 (2005).
13. Zhou, B., Yu, D., Ding, Z. & Xu, X. Comparison of genetic diversity in four *Typha* species (Poales, Typhaceae) from China. *Hydrobiologia* **770**, 117–128. https://doi.org/10.1007/s10750-015-2574-9 (2016).
14. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298. https://doi.org/10.1093/bib/bbn013 (2008).
15. Librado, P. & Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452. https://doi.org/10.1093/bioinformatics/btp187 (2009).
16. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036 (1999).
17. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
18. Peakall, R. & Smouse, P. E. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460 (2012).

Acknowledgements

Author contributions

X.X., E.A.B. and A.G.L. designed the research; X.X., A.G.L. and E.A.B. carried out the field collections; Y.Y., F.L. and W.Y. carried out the experiments and performed the data analysis; X.X., E.A.B. and A.G.L. wrote and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-11222-8.

Correspondence and requests for materials should be addressed to A.G.L. or X.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.