Development of *Daphnia Magna* SSR Markers and Genetic Diversity Analysis Based on RAD-Seq Technology

Shengman Zhang

Shanghai Ocean University

Yiran Xiong

Shanghai Ocean University

Meng Tan

Shanghai Ocean University

Siwei Chen

Shanghai Ocean University

Zheng Han

Shanghai Ocean University

Wenhui He

Shanghai Ocean University

Peimin He (✉ pmhe@shou.edu.cn)

Shanghai Ocean University

Research Article

Keywords: *Daphnia magna*, SSR polymorphism, RAD-seq, Genetic structure, Genetic diversity

Posted Date: August 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-807257/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Daphnia magna belongs to the Cladocera order and plays an important role in the water ecosystem. With the intensification of water pollution, the wild population of *D. magna* has declined rapidly in recent years, and insufficient molecular markers have limited effective research and conservation of this species. In our research, 26 novel microsatellite (SSR) markers were developed in an artificially domesticated of *D. magna* and 12 wild population of *D. magna* using restriction site-associated DNA sequencing (RAD-seq). The results showed that the observed heterozygosity (H_o) and expected heterozygosity (H_e) ranged from 0.083 to 0.999 and 0.085 to 0.862, respectively. The PIC ranged from 0.368 to 0.805. These results indicate that the developed SSR marker is highly polymorphic. Nei's genetic identity (H) ranged from 0.0926 to 0.3462, with a mean of 0.2233. Shannon's Information index (I) ranged from 0.1333 to 0.4799, with an average of 0.3073; Shanxi province had the highest value and Hunan province had the lowest. Genetic distance and Nei's genetic identity analysis, NJ tree diagram analysis, and PCoA analysis were conducted on populations of *D. magna* from different regions. The results show that the *D. magna* genetic relationship between Liaoning and Shanxi, Hunan and Anhui, and Beijing and Hainan is relatively close, while the genetic structure of *D. magna* in Guangdong, Jiangsu, and Sichuan is quite different from other sampling sites. An analysis of population genetic structure divided the test *D. magna* samples into two major groups. These results indicate that the genetic diversity of *D. magna* is rich, and the genetic structure of *D. magna* differs considerably in different regions. These research results and the newly developed polymorphic SSR markers for *D. magna* are of great significance in terms of the genetic breeding of *D. magna*, identification of wild and artificially domesticated population and conservation genetics research.

Introduction

Daphnia magna belongs to the Daphniidae family. *D. magna* are an important component of the water ecosystem, mainly living in freshwater lakes, rivers, ponds, and reservoirs [1]. *D. magna* feeds directly on phytoplankton [2], and is often used to control the population and community structure of phytoplankton in water bodies [3]. In China, the artificially domesticated *D. magna* has been used to inhibit *Cyanobacteria* blooms and restore submerged vegetation, which has been used in more than 400 ecological restoration projects [4–5]. *D. magna* has the characteristics of wide distribution, strong adaptability, short reproduction period, allowing large-scale cultivation in the laboratory, and is highly sensitive to a variety of toxic substances in the water environment [6], so it is widely used in water pollution monitoring and aquatic organism toxicological research [7–8]. It is also an ideal model organism in the field of life science researches [9].

D. magna is an extremely important zooplankton in lakes, rivers, and reservoirs. It grows fast, reproduces well, and has a strong feeding power on algae, effectively controlling the accumulation of primary products in water bodies [10]. *D. magna* helps to increase the diversity of phytoplankton in the water body and build a stable phytoplankton community structure, as well preventing the occurrence of cyanobacteria blooms. It is the key to balance in the water ecosystem. In recent years, with the rapid expansion of the population and increases in the discharge of domestic sewage, the aquatic ecosystem has been severely damaged; relevant research shows that 54% of Asian, 53% of European, 48% of North American, 41% of South American, and 28% of African rivers are eutrophic water bodies, and eutrophication of water bodies is currently the most serious problem facing rivers and lakes [11]. As a result, the habitat of *D. magna* has been severely damaged, and its genetic diversity is also declining sharply. In this study, we use the microsatellite marker method to understand the genetic variation, genetic structure, and gene distribution of *D. magna*, in order to better protect its genetic diversity. In addition, Shanghai Taihe Water Environment Technology Development Co., Ltd. invented a technology for ecological restoration of *D. magna* guided submerged plants, the artificially domesticated *D. magna* that can quickly filter the high biomass of phytoplankton and other particulate organic matter, increasing the transparency of the water body by 1.0–1.5 m in a short time, so that the planting of submerged plants can quickly form vegetation. *D. magna* was first introduced to water ecological restoration in the world and has achieved good ecological restoration effects in many rivers and water bodies [4–5]. Therefore, it is very important to use the SSR markers to study identification of wild and artificially domesticated species, too.

SSR refers to simple sequence repeat [12], also called microsatellite DNA, which is a DNA molecular marker technology based on PCR with specific primers. SSRs are short tandem repeat sequences composed of one to six nucleotides as repeat units in the genome. They are widely distributed in various regions of the genome, with the length generally between 100 and 200bp. SSRs are currently one of the most commonly used microsatellite markers. Because of their large number, codominance, easy operation, stable results, high polymorphism, and strong reproducibility, they have been widely used in research on gene mapping, genetic map construction, fingerprint analysis, genetic relationship identification, and biodiversity assessment [13–15].

Restriction-site associated DNA sequencing (RAD-seq) is an important sequencing technology for simplifying the genome. It has a high number of markers and high density, and can be applied in many fields [16]. This technology has been widely used in research in the fields of population genetics, genetic map construction, and systematic evolution [17]. Our study used RAD-seq technology to develop and screen polymorphic SSR loci of 12 wild population of *D. magna* and 1 species of artificially bred *D. magna*, and analyze the genetic diversity and genetic structure of *D. magna*. The research results are of great significance for identification of wild and artificially domesticated species, genetic breeding, evaluation of genetic diversity, genetic relationship identification, construction of a genetic map, and protection of germplasm resources of *D. magna*.

Materials And Methods

Sample collection and genomic DNA extraction

The experimental samples used in this research are one artificially bred population and 12 wild populations (Table 1); the artificially bred population was taken from the breeding base of Shanghai Taihe Water Environment Technology Development Co., Ltd. (Shanghai Fengjing). Hundreds of *D. magna* adult individuals were randomly collected from 12 wild population sampling sites and brought back to the laboratory for cultivation. The TIANGEN DP304 kit (TIANGEN BIOTECH Co., Ltd, China) was used to extract 50 parthenogenesis cultured *D. magna* genomic DNA at each sampling site, and the integrity was checked using 1.5% agarose gel electrophoresis. The purity and concentration of the DNA were detected using a UV spectrophotometer to ensure the genomic DNA obtained met the quality conditions for building a database. DNA samples were diluted to a concentration of 30 ng/μL and stored at -20°C (Haier BCD-576WDPU, China) until further use.

Table 1 Information on samples used for genetic marker development				
Sampling sites	Codes	Longitude	Latitude	Dates
Taihe Water Co., Ltd.	DZ	121°21'34.7"E	30°50'50.9"N	2019.09
Yan tai	YT	121°37'44.4"E	37°23'33.6"N	2019.11
Ji ning	JN	116°39'36.5"E	35°0'36.5"N	2019.11
Yun nan	YN	102°39'41.7"E	25°1'20.8"N	2019.12
Hai nan	HNa	110°20'7.9"E	20°3'45.2"N	2019.12
Jiang su	JS	120°36'37.8"E	31°19'18.1"N	2020.01
Bei jing	BJ	116°32'27.2"E	39°46'7.9"N	2020.04
Hu nan	HNb	110°39'59.2"E	26°45'1.9"N	2020.05
Shan xi	SX	113°4'33.4"E	36°19'56.0"N	2020.06
Liao ning	LN	122°6'44.9"E	41°12'51.9"N	2020.06
Guang dong	GD	114°25'17.7"E	23°6'7.3"N	2020.06
An hui	AH	116°58'50.9"E	33°39'21.4"N	2020.07
Si chuan	SC	104°4'58.8"E	30°48'52.5"N	2020.07

Restriction Site-associated Dna Sequencing

The genomic DNA of the four populations (DZ, YN, JN, and YT) was mixed in equal amounts, and sequenced by Shanghai Map Biotech Co., Ltd. (Shanghai, China). Since *D. magna* had no reference genome, it was necessary to use multiple restriction endonucleases to digest the DNA of *D. magna* separately. After digesting the genomic DNA with appropriate restriction endonuclease, P1 adapters were added. A P1 adapter contains the primer sequence required for PCR amplification, the sequence of the Illumina sequencing primer binding site, and the short tag sequence to distinguish different samples. Samples with different P1 adapters were mixed, physically breaking them into sequences of 300-500bp, and then P2 adapters were added. Next, we performed PCR amplification and enrichment of RAD-tags, construction of a paired-end library, and sequencing using the Illumina HiSeq PE150 (Illumina, USA) system. Raw data were filtered using trimmomatic software (<http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic>) to obtain high-quality clean data and ensure smooth subsequent analysis. We removed the linker sequence in the reads and reads without inserts because of factors such as linker interconnection. Bases of lower quality were trimmed at the 3' end of the sequence and entire sequences containing base quality values less than 10 were removed, eliminating reads with a N ratio higher than 10%, and removing adapter and quality modification sequences of less than 75bp in length.

Ssr Loci Analysis And Primer Design

MISA (<http://pgrc.ipkgatersleben.de/misa>) software was used to estimate the qualified SSR, and Primer 3 [18] software was used to design SSR primers for better SSR loci.

Ssr Loci Screening

SSR loci screening mainly includes unitary and polymorphism screening. Unitary screening uses PCR and agarose gel electrophoresis (1.5% agarose gel, 200V, 15 min) to determine whether there is a product and whether the product is single-stranded. Polymorphism screening uses fluorescent primer PCR and capillary electrophoresis to determine whether the peak shape is qualified and whether it is polymorphic. Six samples (DZ, YN, HNa, GD, JN, and SZ) were used for polymorphism screening, with PCR amplification using tailed primers, and the universal typing primer Common-famF (AGTCACGACGTTGTAAAACGAC). Polymerase chain reaction (PCR) amplification was performed in 20μL volumes containing 10μL Taq Master Mix (Vazyme Biotech Co., Ltd, China), 30ng/μL DNA template 1μL, typing primer 1μL, forward primer 0.2μL, and reverse primer 1μL, supplemented with

ddH₂O 7μL. The PCR amplification reaction was carried out on a Mastercycler pro PCR machine (FR-180, China). The PCR thermal cycle consisted of initial denaturation at 94°C for 5 min, followed by 35 cycles at 94°C for 30s, locus-specific annealing temperature for 30s, (Table 2 shows the annealing temperature for each primer reaction), 72°C for 30s, final extension at 72°C for 10 min, and then the temperature was maintained at 4°C. PCR products were electrophoresed with 1% agarose and stained with ethidium bromide (EB), observed, and photographed using the gel imaging system.

Data analysis

Genemapper 4.0 software was used to read the molecular weight data of the SSR amplified products, and then DataFormater software was used to determine the genotype of individual points based on the molecular weight data. The total number and number of effective alleles (N_a and N_e , respectively), and observed and expected heterozygosity (H_o and H_e , respectively) were computed using GenAlEx 6.5 [19]. Polymorphic information content was calculated using Powermaker 3.25. POPGENE software [20] was used to calculate the number of polymorphic loci (A), the percentage of polymorphic loci (P), Nei's genetic identity (H), Shannon's information index (I), and Nei's unbiased measures of genetic identity and genetic distance for each population [21]. A neighbor-joining (NJ) dendrogram was constructed from the calculated genetic distance between populations using POPTREE software [22]. During bootstrapping, 1000 permutations were performed to evaluate the robustness of the clusters. Principal Coordinates Analysis (PCoA) was implemented in GenAlEx 6.5 in order to analyze the genetic relationships across populations. Bayesian model-based clustering analysis was performed using STRUCTURE 2.3.4 software to estimate the most likely number of genetic clusters [23]. This clustering method was used to identify genetically distinct subpopulations based on allele frequencies. The admixture model was applied, and the number of clusters (K value) was set from 2 to 10 with three independent runs for each fixed K number. Each run included a burn-in length of 10,000 followed by 10,000 MCMC (Monte Carlo Markov Chain) repetitions. The most likely K value was determined based on the method described by Evanno et al. [24] by submitting all result files for $K = 2$ to 10 to STRUCTURE HARVESTER [25]. The run with the highest log probability value among the three independent ones was chosen and represented as bar plots [26] using DISTRUCT 1.1 [27].

Results

Polymorphism of newly developed microsatellite markers

A total of 26 polymorphic SSR loci were developed in 13 populations using RAD-seq technology. The primer sequences (5'-3'), repeat motifs, size ranges, and annealing temperatures are shown in Table 2. The value of N_a and N_e across all loci was highest in ZSC_320 and lowest in ZSC_7904, with an average number of 4.654 alleles per locus, and mean effective number of alleles of 3.000. The observed heterozygosity (H_o) ranged from 0.083 to 0.999, with a mean of 0.486. The expected heterozygosity (H_e) was 0.085–0.862, with a mean of 0.645. The Polymorphic information content (PIC) range was 0.368–0.805, with an average of 0.576. The Shannon's information index (I) ranged from 0.173 to 1.936, with an average of 1.187. The highest values for H_e , PIC, and I were found in ZSC_320, while the highest values for H_o were found in ZSC_3833. The lowest values for H_o , H_e , PIC, and I were all found in ZSC_7904 (Table 3). The PIC value is an important index to evaluate the degree of variation of a gene locus. If the PIC value of a gene locus is greater than 0.50, it means it is a highly polymorphic locus; if the PIC value is between 0.25 and 0.50, it means it is a moderately polymorphic locus; and if the PIC value is less than 0.25, it means there is low polymorphism. In our research, the average PIC value of 26 SSR markers in 13 populations was 0.576, with 17 loci having a PIC value higher than 0.5. These results indicate that the developed SSR marker is highly polymorphic.

Table 2
Detailed information on the 26 pairs of microsatellite primers

Locus	Primer sequences (5'-3')	Repeat motif	Size range (bp)	Annealing temperature (°C)
ZSC_320	F:ACAACTACACCTACACCAGCAG R:TAGCTGTGACCTGGTTTAAATG	(CAG)6	222 ~ 239	57.0
ZSC_448	F:TACAGCAATCACATAGCAAATG R:TTTTCAGGTCATGAATAATGGA	(TTG)6	97 ~ 114	56.5
ZSC_546	F:ACATTTCAATTGTAAAAGGGTG R:GGATGTTGCGTTTGTCAAT	(TGT)6	174 ~ 191	55.8
ZSC_679	F:AATTCCTCGAAATCGGGT R:GAAAATTCAAAGTTGCCTAACC	(TTG)7	29 ~ 49	57.0
ZSC_748	F:CCGACTTGATGGATGAGATTAT R:CACCTGTTAATGTGAGTCTTGC	(AAC)7	76 ~ 96	58.0
ZSC_761	F:TTATTGCTTTTTGATGTTGGAC R:GGAAGCTTTTGTATAGGCAGTC	(AAC)5	246 ~ 260	56.9
ZSC_1372	F:CTGACATCCATCAAGAAGAAGA R:AACAAATCCCTCTGCTATTCTC	(GTT)7	198 ~ 215	57.0
ZSC_1701	F:CTGAGCAACAGCATCATTTG R:GATTGTAAAGCGGATAGTGTGA	(CAG)5	108 ~ 122	58.0
ZSC_18335	F:CGCTCTCTCTTTCTCTCTTTGT R:CCGAGTATTTCTTCGATGTTTA	(TC)7	113 ~ 126	57.3
ZSC_875	F:AATTCGATGACGGGAATC R:CAACAGCGTCAGTCCTACAC	(ACC)5	21 ~ 35	55.7
ZSC_1797	F:AAATAAAACAACCGAAAAATGG R:CAAGACGTGTCCTGTGTCACT	(CAA)5	250 ~ 264	57.2
ZSC_1890	F:AGAAACGATGTCTGATCACTGT R:CGGTTAATCACTGCATCACTAT	(CAA)5	39 ~ 53	56.3
ZSC_2050	F:AACGATTTCACTTGACATTAG R:GAGACGGACAAGAACAGTAACA	(TGC)6	165 ~ 128	57.0
ZSC_2586	F:ATTCACCGATCGTCAGAAAT R:GTCACTCCCTCCTAGACAACTC	(GCC)8	26 ~ 49	57.0
ZSC_2799	F:AAAGAAATGCCTCGAGTATCAT R:ATTGTAGGTCATCGGGAGGT	(AGA)5	148 ~ 162	57.0
ZSC_2828	F:CACTGAACAAAGGTGAAAAGAA R:GATGAGTGTCATGCTGACTT	(AGA)5	62 ~ 76	57.0
ZSC_3113	F:ATCTTCATATCGAGACCGTT R:TGAATCAGTTCTCCAATAATCG	(AAG)5	142 ~ 156	56.9
ZSC_3761	F:TCATCATAGCGTCGTTTCC R:GGAAGAGGCCTGAGATCC	(CAG)6	256 ~ 273	57.2
ZSC_3833	F:GCGCCTTTCTATAAACTAA R:GCAAAGAAGAAGGAGAACAAGT	(AAC)6	67 ~ 84	57.0
ZSC_4037	F:CTACGGTCCAGGTAGAACTGAT R:GAGATGGTTACGACTGAATTTG	(ACA)6	141 ~ 158	57.4

Locus	Primer sequences (5'-3')	Repeat motif	Size range (bp)	Annealing temperature (°C)
ZSC_4319	F:GTCCAGGCTATTTCCAGTCTAC R:AAAACTCGCTTGAAAAATTGTA	(CTT)7	87 ~ 107	57.0
ZSC_10218	F:CTCTTGACCAGAACTGGAGAC R:ATTCGACTGGAGGTATGGATT	(GTT)6	40 ~ 57	56.5
ZSC_6980	F:GCGTCACTATGTCACCTCATTA R:GGAAGTCCATACCAAGAAAAAC	(ACA)6	196 ~ 213	57.8
ZSC_7904	F:TCATTTTCCAAAAGAAGAAACC R:GTATATGTCAATGTTTGAGCGG	(AAC)6	117 ~ 134	57.4
ZSC_8717	F:AGTTTCATCAACAGCCACTTC R:AGGAAGGCGTTTTTGGAC	(GCC)6	179 ~ 196	56.9
ZSC_9631	F:AGAAGAGGAAGAGAAGCAGTTG R:TGTTTTTAATGAGCCGATGAT	(CAG)6	143 ~ 160	57.1

Table 3
Characterization of 26 novel microsatellite loci in 13 populations of *Daphnia magna*

Locus	Number of alleles (Na)	Effective number of alleles (Ne)	Observed heterozygosities (Ho)	Expected heterozygosities (He)	Polymorphic information content (PIC)	Shannon's Information index (I)
ZSC_320	9	5.760	0.417	0.862	0.805	1.936
ZSC_448	3	2.198	0.700	0.573	0.441	0.856
ZSC_546	6	3.646	0.583	0.757	0.696	1.537
ZSC_679	6	4.400	0.455	0.810	0.738	1.605
ZSC_748	6	3.507	0.727	0.749	0.679	1.477
ZSC_761	3	2.051	0.778	0.543	0.426	0.828
ZSC_1372	5	2.439	0.600	0.621	0.553	1.175
ZSC_1701	5	2.793	0.231	0.668	0.585	1.229
ZSC_18335	6	2.432	0.308	0.612	0.536	1.172
ZSC_875	6	3.485	0.615	0.742	0.678	1.482
ZSC_1797	3	2.602	0.636	0.645	0.542	1.021
ZSC_1890	6	4.500	0.750	0.812	0.742	1.596
ZSC_2050	3	1.852	0.400	0.484	0.410	0.802
ZSC_2586	4	3.270	0.636	0.727	0.637	1.264
ZSC_2799	3	2.268	0.333	0.583	0.488	0.934
ZSC_2828	3	2.283	0.364	0.589	0.476	0.917
ZSC_3113	3	1.986	0.250	0.518	0.443	0.857
ZSC_3761	6	2.522	0.231	0.628	0.560	1.229
ZSC_3833	4	2.455	0.999	0.628	0.512	1.035
ZSC_4037	3	1.811	0.250	0.467	0.397	0.778
ZSC_4319	3	2.160	0.556	0.569	0.468	0.901
ZSC_10218	5	2.704	0.539	0.655	0.561	1.172
ZSC_6980	6	4.546	0.600	0.821	0.749	1.639
ZSC_7904	2	1.087	0.083	0.085	0.368	0.173
ZSC_8717	5	4.000	0.167	0.783	0.711	1.488
ZSC_9631	7	5.236	0.417	0.844	0.782	1.762
Mean	4.654	3.000	0.486	0.645	0.576	1.187

Genetic diversity of *D. magna* populations

Amplification results for the 26 SSR markers in 13 *D. magna* populations showed high polymorphism. At the population level, the 26 SSR markers detected that the number of polymorphic loci ranged from 2 to 18 in the 13 populations, with an average of 9.923. The percentage of polymorphic loci ranged from 7.69–69.23%, with an average of 41.12%. The SX population amplification results showed the highest level of polymorphism, with a polymorphic loci percentage of 69.23%, followed by GD (65.38%), DZ (61.54%), and LN (61.54%) populations, while the JS population amplification results showed the lowest level of polymorphism, with a polymorphic loci percentage of only 7.69%. Furthermore, Nei's genetic identity (H) ranged from 0.0926 to 0.3462, with a mean of 0.2233. The value for SX was the highest and that for HNb was the lowest. In addition, the values for LN, GD, and DZ also exceeded 3.000. Shannon's Information index (I) ranged from 0.1333 to 0.4799, with an average of 0.3073. SX had the highest value, followed by GD, while HNb had the lowest value (Table 4). The above results show that among the 13 regions, genetic diversity of the *D. magna* populations in Shanxi, Guangdong, Liaoning, and Taihe Water is relatively high. Genetic diversity of the *D. magna* populations in Shanxi Province is the most abundant, while that in Jiangsu Province is relatively low. There are obvious genetic differences among different provinces and cities.

Table 4
Genetic diversity index for 13 populations of *D. magna* analyzed using SSR markers

Populations	Number of polymorphic loci (A)	The percentage of polymorphic loci (P)	Nei's genetic identity (H)	Shannon's Information index (I)
AH	11	42.31%	0.2115	0.2933
BJ	5	19.23%	0.1569	0.2018
DZ	16	61.54%	0.3077	0.4266
HNa	4	15.38%	0.1176	0.1631
GD	17	65.38%	0.3269	0.4532
HNb	5	19.23%	0.0926	0.1333
JN	14	53.85%	0.2692	0.3732
JS	2	7.69%	0.1111	0.1540
SC	7	26.92%	0.1667	0.2310
SX	18	69.23%	0.3462	0.4799
YN	9	34.62%	0.1800	0.2495
YT	5	57.69%	0.2885	0.3999
Mean	9.923	41.12%	0.2233	0.3073

Genetic structure analysis of *D. magna* populations

In order to further explore the genetic structure differences of *D. magna* populations in different regions, we calculated Nei's genetic identity and genetic distance for the 13 sampling sites. Nei's genetic identity and genetic distance is an important indicator that reflects the distance between populations. When the genetic distance is 1 (the genetic identity is 0), it indicates that the *D. magna* of the two populations are completely different and there is no genetic relationship; when the genetic distance is 0 (the genetic identity is 1), it indicates that the germplasm of the two populations is identical [28]. The genetic distances for the 13 different sources of *D. magna* were 0.0034–0.6864, with an average value of 0.3722. The genetic distance between YT and JS was the largest, and the genetic distance between AH and HNb was the smallest. Nei's genetic identity for *D. magna* in the 13 different regions was 0.3220–0.9966, with an average value of 0.6461. AH and HNb had the highest Nei's genetic identity, and JS and AH had the lowest (Table 5). This shows that there are certain differences in the genetic structure of *D. magna* in different regions, and the genetic diversity of *D. magna* is extremely rich. Jiangsu and Anhui populations are relatively distantly related, while Anhui and Hunan populations are relatively closely related.

Table 5
Nei's genetic identity and genetic distance for 13 populations of *Daphnia magna*

Populations	AH	BJ	DZ	HNa	GD	HNb	JN	LN	SC	SX	JS	YN	YT
AH	****	0.4686	0.9650	0.4528	0.5897	0.9966	0.9378	0.9275	0.5434	0.9351	0.3220	0.9891	0.9944
BJ	0.5316	****	0.4495	0.7346	0.6485	0.4230	0.3629	0.4122	0.7469	0.4573	0.3264	0.4343	0.4098
DZ	0.1080	0.5508	****	0.4765	0.5334	0.9303	0.8002	0.9522	0.5904	0.9825	0.3922	0.9591	0.7658
HNa	0.5482	0.3084	0.5324	****	0.6833	0.4387	0.3970	0.4349	0.8229	0.5018	0.3954	0.4426	0.3911
GD	0.4585	0.3994	0.4669	0.3651	****	0.6358	0.6682	0.5118	0.6969	0.4193	0.3943	0.7318	0.5020
HNb	0.0034	0.5771	0.1616	0.5613	0.4026	****	0.8324	0.8778	0.5639	0.8859	0.3759	0.9696	0.9564
JN	0.1497	0.6505	0.2468	0.6035	0.3669	0.2068	****	0.8037	0.7343	0.8038	0.3225	0.9322	0.9736
LN	0.0893	0.5888	0.1274	0.5652	0.4983	0.1513	0.2514	****	0.5976	0.9890	0.3461	0.7763	0.8661
SC	0.3666	0.3145	0.4159	0.2361	0.3584	0.4664	0.3516	0.4158	****	0.6449	0.3226	0.5844	0.5165
SX	0.0802	0.5432	0.0826	0.5205	0.5812	0.1409	0.1518	0.0037	0.4025	****	0.3483	0.7971	0.9185
JS	0.6783	0.6739	0.6080	0.6048	0.6061	0.6243	0.6780	0.6542	0.6779	0.6517	****	0.3896	0.3436
YN	0.0114	0.5657	0.1169	0.5577	0.3340	0.0323	0.1586	0.2532	0.4305	0.2268	0.6174	****	0.9452
YT	0.1648	0.5904	0.2992	0.6089	0.5195	0.1210	0.0951	0.1664	0.4839	0.0903	0.6864	0.1381	****
Note: Nei's genetic identity (above diagonal) and genetic distance (below diagonal).													

In order to more intuitively show the genetic distance differences of different populations of *D. magna*, POPTREE software was used to construct a neighbor connection (NJ) tree diagram. The results showed that AH and HNb clustered first, and then clustered with YN. LN and SX are clustered, and

then clustered with YT, JN, and DZ, respectively. BJ clustered with HNa, and then clustered with SC. Among these 13 sampling sites, the genetic structures of GD and JS are quite different from the genetic structure of the other regions, and the genetic relationship between AH and HNb, and between BJ and HNa, is closer. The clustering results are consistent with Table 5.

Principal coordinates analysis (PCoA) is a powerful tool for evaluating the genetic structure of a population. Based on the screened SSR loci, GenAlEx 6.5 software was used to perform principal coordinates analysis on *D. magna* in different regions, and PCoA cluster maps of experimental samples in 13 regions were obtained (Fig. 3). It can be seen from the PCoA cluster maps that AH, HNb, YN, LN, JN, SX, YT, and DZ are clustered together. BJ and HNa are clustered together but they are obviously scattered from the other populations, while the *D. magna* populations for GD, JS, and SC are distributed separately and relatively scattered. These results are basically consistent with the results from population genetic distance analysis and clustering results, indicating the consistency of the clustering results.

From the tree diagram and PCoA scatter diagram, we can intuitively understand the classification relationship between populations, but in order to determine how many subgroups a certain group has, whether there is gene exchange between groups, and the degree of hybridization of each individual, we need a structure diagram to analyze the structure of the populations. The preset number of population subgroups is equal to 1–10, that is, $K = 1-10$, and the grouping situation and ancestry composition of the population are calculated based on the Bayes algorithm. Simulation results for each K value will produce the maximum likelihood. The larger the value of $\ln$ likelihood, the closer the K value is to the real situation, which means we need to determine the simulation result with the largest likelihood and smallest K value [25]. It can be seen from Fig. 4 that as the value of K increases, the value of $\ln(K)$ also gradually increases, but when $K = 2$, the value of $\ln$ likelihood enters a plateau, and the value of ΔK appears at its maximum with an obvious peak, indicating that *D. magna* in the 13 regions can be divided into two groups. Each individual is represented by a vertical colored bar, and the proportion of the color in each bar represents the probability of membership in the relevant cluster. It can be seen from Fig. 5 that *D. magna* at the eight sampling sites (AH, DZ, HNb, JN, LN, SX, YN, and YT) can be put into one group, and they may have a common ancestor, while BJ, HNa, and JS are divided into another group, their genetic structure is relatively simple. The colored bars for GD and SC are composed of two colors, indicating that the genetic background of the populations for GD and SC is relatively complicated, and it is likely to be derived from a cross of two ancestral subgroups, with the ancestral subgroup of GD accounting for between 40% and 60% of the pedigrees, and SC mostly belonging to the red group.

Discussion

Analysis of genetic characteristics of *D. magna* population by SSR markers

D. magna is a small planktonic cladoceran. *D. magna* has been used as a model species for ecotoxicology, as it is sensitive to environmental stressors and environmental changes [29] (Lee et al. 2019). At present, research on *D. magna* mainly focuses on reproduction ecology, toxic and toxicological effects [30], and water pollution purification [31–32]. Lee et al. [29] assembled and characterized the genome of *D. magna*, which has greatly helped the molecular research of *D. magna*. Robinson et al. [33] used statistics calculated from multilocus microsatellite datasets to estimate population ages in data generated through coalescent simulations and in samples from populations of known age in a metapopulation of *D. magna* in Finland. Our research uses SSR markers to identify wild and artificially domesticated species.

In our study, 26 polymorphic SSR loci were developed using RAD-seq technology, and genetic diversity analysis was performed on 12 wild *D. magna* populations and one artificially selected *D. magna* population in China. The average number of alleles for the 26 pairs of SSR primers is 4.654. With an average of more than four alleles, one can better assess the genetic diversity of a population [34], so the developed microsatellite markers show good polymorphism and can provide help in further molecular research on *D. magna*. Diversity analysis of *D. magna* populations in the 13 regions shows that the mean values of Nei's genetic identity and Shannon's Information index are 0.2233 and 0.3073, respectively, indicating that the genetic marker diversity of *D. magna* is relatively high. The genetic diversity of *D. magna* is relatively high in Shanxi and Guangdong and is the lowest in Jiangsu. In addition, it also showed good genetic diversity in artificially domesticated *D. magna* (DZ). Genetic distance and Nei's genetic identity analysis, NJ cluster analysis, PCoA analysis, and STRUCTURE genetic structure analysis were conducted on germplasm samples from *D. magna* in different regions; the results were basically consistent and complementary. The research results showed that the genetic diversity of *D. magna* from different sources was quite different. The genetic relationship between Anhui and Hunan's *D. magna* germplasm is the closest, and the genetic relationship between Anhui and Jiangsu is the furthest. NJ cluster analysis and genetic distance analysis gave the same results. PCoA analysis results showed that *D. magna* in Anhui, Hunan, Jining, Liaoning, Taiheshui, Yantai, and Yunnan can be clustered into one group. The differences in *D. magna* in Beijing and Hainan are also small, while *D. magna* in Guangdong, Sichuan, and Jiangsu are quite different from other regions, and the clustering results have no correlation with geographic origin. The STRUCTURE genetic structure plot divides *D. magna* from the 13 regions into two groups. The genetic backgrounds of *D. magna* in Anhui, Hunan, Jining, Liaoning, Shanxi, Yunnan, Yantai, and Taihe Water are relatively simple, and the genetic structure of

D. magna in Beijing, Hainan, and Jiangsu is also very simple but significantly different from the previous group, while the genes of *D. magna* in Guangdong and Sichuan are shared with the two groups.

Genetic differentiation of *D. magna* population

From the sampling site distribution and clustering results, there is no absolute correlation between the gene distribution of *D. magna* and its geographic origin, but on the whole, there are still large differences in the genetic structure of *D. magna* in the north and south of China, which may be a result of differences in temperature, climate, precipitation, and water environment. However, there are exceptions; the genetic distance between *D. magna* in Beijing and Hainan is relatively small, which may be due to similarities in the water environment causing *D. magna* in the two places to evolve in a similar direction. Human interference is an extremely significant activity affecting gene exchange in species, with the high similarity of *D. magna* genes in different regions likely to be the result of artificial introduction. Meanwhile, zooplankton have weak swimming ability and are highly susceptible to factors such as water level fluctuations, tidal currents, and water erosion [35]. Ingestion by large aquatic animals such as fish, shrimps, and crabs will also accelerate gene exchange between *D. magna* from different regions. However, based on our research, the genetic structure of *D. magna* in different regions is quite different. Human construction activities destroy the habitats of *D. magna*, which fragment the habitat and spread populations, resulting in low genetic similarity between *D. magna* in different regions. In addition to its weak swimming ability, *D. magna* mainly reproduces asexually, which inevitably affects gene exchange between populations due to long-term geographic isolation, which may lead to obvious genetic differentiation between populations, increase in genetic distance, and decrease in population size, leading to population selfing and genetic drift, which ultimately leads to a decrease in genetic diversity. Therefore, we should raise awareness of environmental protection, protect the safety and integrity of the water environment, and also cultivate and introduce *D. magna* into different regions and rivers in order to increase gene flow between populations, prevent the aggravation of genetic differentiation, and increase genetic diversity. A high level of genetic diversity can provide a source of genes to optimize the germplasm resources of *D. magna*. As a natural open bait for high-level aquatic animals such as fish and shrimps, high-quality *D. magna* germplasm resources can improve the quality of aquatic products, shorten the breeding cycle of aquatic products, and improve the economic benefits of aquaculture [36]. In addition, artificial domestication of *D. magna* (Taihe Water sampling site) can quickly remove algae and other organic particles, increasing the transparency of the water body by 1.0-1.5 m in a short time, so that submerged plants can be transplanted and survive. Submerged plants quickly reduce the content of nutrients such as nitrogen and phosphorus in the water body through nutrient competition, and inhibit the growth and reproduction of blooms and algae[4], playing a major role in the restoration of water ecosystems. Therefore, studying the genetic structure of *D. magna* in different regions can not only protect its genetic diversity at a molecular level, but it can also enrich the *D. magna* gene bank and increase our understanding of its evolutionary potential.

The life history of *D. magna* is unique; it consists of parthenogenetic generation and bisexual reproduction, and these two reproductive methods can alternate with changes in the external environment [37]. When the external environment is suitable, such as suitable temperature, sufficient food, and sufficient biological space, *D. magna* reproduces mainly by parthenogenesis, and the egg cells produced by females can directly develop into female offspring without fertilization. This is the most important way for it to reproduce. Once external conditions deteriorate, such as lack of food, high or low pH, the amount of dissolved oxygen in the water drops sharply, metabolic wastes accumulate, population density increases, or the light cycle shortens, the reproductive mode of *D. magna* will change from asexual to bisexual. When environmental conditions improve again, the reproductive mode of *D. magna* will switch from asexual reproduction to bisexual reproduction [38]. All *D. magna* samples used in this experiment must be the offspring of asexual reproduction, so it is necessary to ensure a suitable culture environment. In wild waters environment, *D. magna* reproduces asexually under most conditions, but in some harsh environments, some individuals will reproduce with both sexes and produce fertilized eggs covered by a saddle to resist the poor external environment. Therefore, there will still be differences in the genetic structure of *D. magna* in the same water area, even if the difference is very small. In addition, the sampling sites in this study are evenly distributed, but only a few sampling sites were chosen, China has a vast territory and a large north-south span. The sampling sites of a certain river cannot fully reflect the genetic diversity of all *D. magna* in a province or city. Although the polymorphic SSR markers developed in this experiment are abundant in quantity, provided reliable results, and co-dominant markers, the SSR markers still have co-migration problems. Today, the rapid development of new molecular markers such as EST, SNP, and gene chips have brought genetic diversity research more efficient and rich results and will be the focus of our next research.

Declarations

Acknowledgements We wish to thank Shanghai Taihe Water Environment Technology Development Co.Ltd for help with experimental materials.

Funding The study was supported by National Water Pollution Control and Treatment Science and Technology Major Project of China (2017ZX07205).

Data availability The data of this study are available from the author upon request.

Declarations

Conflict of interest The authors declares that there is on conflict of interest.

References

1. Forrò L, Korovchinsky NM, Kotov AA, Petrusek A (2008) Global diversity of cladocerans (Cladocera; Crustacea) in freshwater. *Hydrobiologia* 595:177–184
2. Nandini S, Rao TR (1997) Somatic and population growth in selected cladoceran and rotifer species offered the cyanobacterium *Microcystis aeruginosa* as food. *Aquat Ecol* 31:283–298
3. Sarnelle O (2007) Initial conditions mediate the interaction between *Daphnia* and bloom-forming cyanobacteria. *Limnol Oceanogr* 52:2120–2127
4. Huo YZ, He WH, Luo K, Wang YY, Zhang YJ, Tian QT, He PM (2010) Bioremediation efficiency of applying *Daphnia magna* and submerged plants: a case study in Dishui Lake of Shanghai, China. *Chinese J of Appl Ecol* 21:495–499
5. Peng ZR, He WH, Gao JH, Yu YN (2011) Change of water quality during the ecological restoration in the Garden of Yuanmingyuan. *J Shanghai Ocean Uni* 20:457–461
6. Seda J, Petrusek A (2011) *Daphnia* as a model organism in limnology and aquatic biology: introductory remarks. *J Limnol* 70:337–344
7. Dietrich S, Ploessl F, Bracher F, Laforsch C (2010) Single and combined toxicity of pharmaceuticals at environmentally relevant concentrations in *Daphnia magna*—a multigenerational study. *Chemosphere* 79:60–66
8. Fan WH, Wu CG, Zhao CM, Yu T, Zhang Y (2011) Application of enriched stable isotope technique to the study of copper bioavailability in *Daphnia magna*. *J Environ Sci* 23:831–836
9. Miner BE, Knapp RA, Colbourne JK, Pfrender ME (2013) Evolutionary history of alpine and subalpine *Daphnia* in western North America. *Freshwater Biol* 58:1512–1522
10. Grintzalis K, Dai WK, Panagiotidis K, Belavgeni A (2017) Miniaturising acute toxicity and feeding rate measurements in *Daphnia magna*. *Ecotox Environ Safe* 139:352–357
11. Bagousse YL, Liancourt P, Gross N, Straile D (2012) Indirect facilitation promotes macrophyte survival and growth in freshwater ecosystems threatened by eutrophication. *J Ecol* 100:530–538
12. Ding S, Wang S, He K, Jiang M, Li F (2017) Large-scale analysis reveals that the genome features of simple sequence repeats are generally conserved at the family level in insects. *BMC Genomics* 18:848
13. Panesar, Parmjit S, Bera Manab B, Kaur S, Varinder, Shubhneet (2015) Simple sequence repeat markers in genetic divergence and marker-assisted selection of rice cultivars: a review. *Crit Rev Food Sci* 55:41–49
14. Kim KS, Sappington TW (2013) Microsatellite data analysis for population genetics. *Methods Mol Biol* 1006:271–295
15. Chambers GK, Curtis C, Millar CD, Huynen L, Lambert DM (2014) DNA fingerprinting in zoology: past, present, future. *Investigative Genetics* 5:3
16. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* 17:240–248
17. Li B, Liu CL, Liu G, Chen SQ, Zhang LL, Ge JL, Li FH, Tan J (2018) SNP discovery in spotted halibut (*Verasper variegatus*) using restriction site-associated DNA sequencing (RAD-seq). *Conserv Genet Resour* 10:409
18. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595
19. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
20. Mir RR, Kumar J, Balyan HS, Guata PK (2012) A study of genetic diversity among Indian bread wheat (*Triticumaestivum* L.) cultivars released during last 100 years. *Genet Resour Crop Evol* 59:717–726
21. Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590
22. Takezaki N, Nei M, Tamura K (2010) POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Mol Biol Evol* 27:747–752
23. Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
24. Evanno GS, Regnaut SJ, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
25. Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361
26. Sittler V, Zhang D, Harris, DL, Yadav AK, Zee FT, Meinhardt LW, Dhekney SA (2014) Genetic characterization of guava (*Psidium guajava* L.) germplasm in the United States using microsatellite markers. *Genet Resour Crop Evol* 61:829–839
27. Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138
28. Nei M (1972) Nei M Genetic distance between populations. *Am Nat.* *Am Nat* 106:283–292
29. Lee BY, Choi BS, Kim MS, Park JC, Jeong CB, Han J, Lee JS (2019) The genome of the freshwater water flea *Daphnia magna*: a potential use for freshwater molecular ecotoxicology. *Aquat Toxicol.*
30. Rosenfeldt RR, Seitz F, Schulz R (2014) Heavy metal uptake and toxicity in the presence of titanium dioxide nanoparticles: a factorial approach using *Daphnia magna*. *Enviro Sci Technol* 48:6965–6972

31. Berger SA, Diehl S, Stibor H, Trommer G, Ruhenstroth M, Wild A, Weigert A, Jäger CG, Striebel M (2007) Water temperature and mixing depth affect timing and magnitude of events during spring succession of the plankton. *Oecologia* 150:643–654
32. Straile D (2000) Meteorological forcing of plankton dynamics in a large and deep continental European lake. *Oecologia* 122:44–50
33. Robinson JD, Haag CR, Hall DW, Pajunen VI, Wares JP (2012) Genetic estimates of population age in the water flea, *Daphnia magna*. *J Hered* 103: 887–897
34. Wajid A, Wasim M, Yaqub T, Firyal S, Hussain T (2014) Assessment of genetic in Balochi and Rakhshani sheep breeds of Balochistan using microsatellite DNA markers. *J Anim Plant Sci* 24:1348–1354
35. Robinson HE, Finelli CM, Koehl MAR (2013) Interactions between benthic predators and zooplanktonic prey are affected by turbulent waves. *Integr Comp Biol* 5:810–820
36. Feng Y, Huang ZY, Hua XM, He WH, Peng ZR, Cai QJ, Liu YC (2018) Nutrient analysis and potential utilization value of *Daphnia magna* used as aqua-feed raw material. *China J Fish* 42:565–574
37. Buchberger F, Stibor H, Neusius D, Nickelsen J, Stockenreiter M (2020) Transgenic and cell wall-deficient *Chlamydomonas reinhardtii* food affects life history of *Daphnia magna*. *J Appl Phycol* 32:319–328
38. Hebert PDN (2010) The Population Biology of *Daphnia* (Crustacea, Daphnidae). *Biol Rev* 53:387–426

Figures

Figure 1

Map of the sampling locations for the 13 populations of *Daphnia magna* in the present study

Figure 2

Neighbor-joining dendrogram based on genetic distances among different populations of *Daphnia magna*.

Figure 3

Scatter plots of principal coordinates analysis (PCoA) revealing the genetic relationships between *D. magna* in 13 regions

Figure 4
 Distribution of K values with $L(K)$, $L'(K)$, $|L''(K)|$, and ΔK

Figure 5

Estimated genetic structure resulting from Bayesian model-based analysis using STRUCTURE 2.3.4 for populations of *Daphnia magna*. (Each number in the figure represents a population. 1: AH, 2: BJ, 3: DZ, 4: HNa, 5: GD, 6: HNb, 7: JN, 8: LN, 9: SC, 10: SX, 11: JS, 12: YN, 13:YT)