Temperature is a cryptic factor to shape the geographical pattern of genetic variation in *Ceratophyllum demersum* across a subtropical freshwater lake

Yixian Li¹, Manli Xia¹, Xuyao Zhao¹, Jingjing Yang¹, Gaojie Li¹, Zuoliang Sun¹, Wei Xinzeng², and Hongwei Hou¹

¹Institute of Hydrobiology Chinese Academy of Sciences
²Chinese Academy of Sciences Wuhan Botanical Garden

October 6, 2022

Abstract

Macrophyte habitats exhibit remarkable heterogeneity encompasses the spatial variation of abiotic and biotic components such as water condition changes, climates and anthropogenic stressors. Environmental factors have been proposed as important drivers in shaping genetic and epigenetic variation of aquatic plants, yet the linking between genetic diversity, epigenetic variation and environmental variables remain largely unclear, especially in clonal aquatic plants. Here, we applied population genetic and epigenetic analysis, in conjunction with the habitats discriminations to detail the environmental factors of which drive intraspecies genetic and epigenetic variations of *Ceratophyllum demersum* from a subtropical lake. Our results demonstrated that environmental factors were highly correlated to the genetic and epigenetic variation of *C. demersum*, temperature was a key driver in generating the genetic variation of this aquatic herb. Genetic and epigenetic variation were positively driven by water temperature, climate temperature was defined to exert negative effects on genetic and epigenetic variation. These findings indicate that the genetic and epigenetic variations of this clonal aquatic herb could not be related to the geographic feature, but driven by environmental hierarchal, which confers new benefits of temperature to local genetic and epigenetic variation in aquatic systems.

Introduction

Environment water conditions constantly drive the aquatic plants to cope with the dynamic variations of their habitats (Hedrick, 1976). While on the process of adaption to the environment shifts, plants usually generate necessitate genetic and epigenetic variations to compensating the environment oscillations (Robertson, Schrey, Shayter, Moss, & Richards, 2017). Plants genetic and epigenetic diversity in natural populations are generally found to be related to environmental gradients (Richards et al., 2017). In clonal aquatic herbs, species usually developed diverse phenotypic traits to fit the long-term environmental influences (M.-Z. Wang, Li, Li, & Yu, 2020). However, as the epigenetic induced phenotypic changes can be inherited through meiosis, epigenetic modifications have been considered as a rapid response to the random environment changes and anthropogenic stressors in comparing with genetic variation (Schulz, Eckstein, & Durka, 2014).

Investigating the relative contributions of geographical and environmental variation to genetic divergence is a critical role to understand species adaptive differentiation in ecological speciation. Associations between environmental characteristics and species adaptive variation and details of their interactions has been widely investigated in plants species (Ortego, Riordan, Gugger, & Sork, 2012; Barajas-Barbosa, Weigelt, Borregaard, Keppel, & Kreft, 2020; Shen et al., 2022). Such studies have made considerable progress in understanding the relative roles of adaptive and nonadaptive processes in shaping patterns of genomic variation and the effects
of environmental variables on adaptive differentiation. Although researchers have primarily recognized the importance of environmental variables on genetic structure and epigenetic variation, few have focus on the aquatic plants, especially clonal macrophyte.

*Ceratophyllum demersum* L. (Ceratophyllaceae) is a submerged, rootless free-floating aquatic macrophyte with cosmopolitan distribution. This species mainly produced vegetatively and occurring offshore about 10 meters, the stems usually up to 1-3 m in length and branches can be modified as rhizoids (Cronk & Fennessy, 2016). *C. demersum* is sensitive to water pollution and now it is established as a reference specie to detected trace element pollution in freshwater ecosystem (Polechońska & Klink, 2021). Considering that *C. demersum* occurs widely in China and worldwide, where the environment habitats are usually heterogeneous, previous studies have only characterized the genotypes differentiate among geographical regions, the intraspecies variation and the associations between genetic variation and habitats variation is still uncharacterized (Hyldgaard, Lambertini, & Brix, 2017). Meanwhile, since the 10-years fishing ban in Yangtze river was implemented in the year 2020, natural populations of *C. demersum* is experiencing cascading and extirpation in numbers of lakes during our field monitoring. Thereby, exploring how heterogeneous environments shape its genetic and epigenetic variation can provide useful information for the restoration of this species, not only in our study areas, but also on a wide scale.

In this study, we investigated the levels of genetic and epigenetic variation of *C. demersum* over the range of Liangzi Lake, of which conserved the highest level of macrophyte diversity in Yangtze river basin, and measured the environment components among the habitats. Using amplified fragment polymorphism length (AFLP) and methylated-sensitive amplified polymorphism (MSAP) markers, we focus on: 1) investigating the genetic diversity and epigenetic variation of *C. demersum* across the Liangzi Lake, 2) elucidating the correlation between environment factors, genetic and epigenetic variations of *C. demersum*, 3) suggesting suitable areas for the specie restoration.

Materials and methods

Sampling and DNA extraction

Plant samples of *C. demersum* was exhaustively searched across the study area in December 2019. A total of 110 individuals were obtained from 12 sites (Figure 1 and Table 1). The sampled populations were separated apart approximately at Euclidean distance of 5 kilometers and individuals were separated by each other at 20 meters. Same development stage of young leaves was collected and dried immediately in silica gel then stored at -20 for further use. Total genomic DNA was extracted from approximate 20 mg of dried tissue using plant genomic DNA rapid extraction kit (Tsingke, Beijing, China). The genomic DNA was quantified on 1% TAE-agarose gel using λ DNA labeled as the template for molecular analysis.

AFLP and MSAP genotyping

For genetic and epigenetic analysis, AFLP and MSAP molecular markers were used to evaluate the genetic and epigenetic diversity level of *C. demersum*. Detailed procedure of AFLP and MSAP and statistical analysis were described in Appendix S1 (Supporting information).

An initial selective polymerase chain reaction of eight individuals cross four populations was carried out with 12 primer combinations, primers only provided clear and reproducible bands with sufficient polymorphic variations between populations were used in the analysis. We finally screened six most informative primer combinations (E-ACA/M-CAA, E-ACT/M-CAC, E-ACT/M-CTT, E-AGC/M-CAA, E-AGC/M-CTC, E-AGC/M-CTT) for the AFLP research and six primer pairs (E-ACA/M-TTA, E-ACA/M-TTG, E-ACT/M-TTA, E-ACT/M-TTG, E-AGC/M-TTA, E-AGC/M-TTG) for the MSAP analysis.

Environmental variables

In the process of habitats heterogeneity assessment, a total of 43 environmental variables were collected to fit the analysis. The stratum of environment gradient was examined based on the Euclidean distance of environmental variables at each sampling site. Elevation, pH, waterbody temperature (WT), Secchi depth
(SD), water conductivity (Cond), total dissolved solids (TDS), salinity (Sal) and dissolved oxygen (DO) were determined in-situ with a multiparameter probe (YSI Company, Ohio, USA). Total nitrogen in water (TN) and sediments (STN), total phosphorus in water (TP) and sediments (STP), chemical oxygen demand (COD) and chlorophyll-a (Chl-a) were measured according to the method recommended by national groundwater analysis standard procedure (GB 3838-2002). Concentrations of the six heavy metal elements (As, Cd, Pb, Se, Zn, Cu) in sampled waters were analysis on inductively coupled plasma-mass spectrometry (ICP-MS, NexION2000, PerkinElmer, Ohio, USA). Distance to the nearest village (DtV), distance to the road (DtR) and number of populations in village (NPV) were acquired through QGIS v3.22 (http://www.qgis.org) and field survey by the locals. Moreover, we access the global climate and weather database from Worldclim v2.1 (http://worldclim.org) and extract 19 bio-climates attributes at a 30 arc second resolution. Previous studies indicated that multicollinearity in variables might lead to the overfitting of models, to reduce the multicollinearity, we divided the temperature (Bio1-11) and precipitation (Bio12-19) into two groups and then calculated the Pearson coefficients. One of the variables in each pair with Pearson’s correlation coefficients ($r > 0.8$) was eliminated. As topographic slope strongly influences the local temperature thereby influences the thermal condition of the habitats, we analyzed the slope aspect of each sampling sites based on the attributes of topography of Liangzi Lake from the Advanced Spaceborne Thermal Emission and Reflection Radiometer Global digital elevation Model V3 (ASTER GDAM V3).

**Data analysis**

The AFLP and MSAP genotypes were obtained from ABI sequence prism 377 using GENSCAN-v3.7 (Thermo Fisher Scientific, Massachusetts, USA). Fragments between 100 and 500 bp was detected by using GeneMarker v2.2.2 software (Hulce, Li, Snyder-Leiby, & Liu, 2011). The polymorphism loci were recon-confirmed manually and then transformed into a binary 0/1 matrix as present ‘1’ or absent ‘0’ of the bands. For MSAP datasets, R-package msap (Pérez-Figueroa, 2013) was used to calculate the types of genome methylations. Locus were classified as either methylated-susceptible loci (MSL) or non-methylated loci, and a 0/1 binary matrix was also constructed where ‘0’ represented as non-methylated locus and ‘1’ represented as methylated locus. Indices of genetic diversity and epigenetic variation within populations were obtained from POPGENE-v1.3.2 (Yeh, Boyle, Yang, Ye, & Xiyan, 1997) assuming Hardy-Weinberg equilibrium as (i) expected heterozygosity ($H_E$ and $eH_E$), (ii) mean Shannon’s information index ($I$ and $eI$), (iii) percentage of polymorphic loci ($PPL$ and $ePPL$ )

Based on the Jaccard genetic similarity coefficient, an UPGMA dendrogram was constructed to devide the genetic groups among populations with ‘ape’ package (Paradis et al., 2019) in R program. The portion of population genetic variation among groups ($o_{CT}$), among populations within groups ($o_{SC}$) and within populations ($o_{ST}$) was examined by the Analysis of molecular variance (AMOVA) in R-package poppr (Kamvar, Tabima, & Grunwald, 2014). We also used GenAlEx v6.503 (Peakall & Smouse, 2012) to characterized the genetic and epigenetic differentiation at each locus. The statistical significant levels were determined after 9999 permutations.

We then inferred the potential population structure of *C. demersum* from AFLP dataset using Bayesian clustering approach in STRUCTURE v2.3.4 software (Pritchard, Stephens, & Donnelly, 2000). The analysis was performed using admixture ancestry model with correlated allele frequencies, the number of clusters (K) was set to vary from 1-12 with each run having 5x10^6 Markov chain Monte Carlo (MCMC) iterations following a burn-in period of 10^6 steps. To identify the appropriate number of clusters of individuals, the likelihood values were calculated to partitioning the populations into different numbers of clusters (K) in STRUCTURE HARVESTER v0.6.94 (Earl & VonHoldt, 2012).

To further investigate the potential relationship between genetic diversity, epigenetic variation, geographic distance and environment distance, we constructed a pair-wise distance matrix complied genetic diversity indices ($H_E$, $I$, $PPL$) and epigenetic indices ($eH_E$, $eI$, $ePPL$) and calculated the Bray-curtis distance among samples. The Euclidean geographic distance and Euclidean environmental distance were also computed using the coordinates of sampling sites and environmental variables respectively. Mantel test was performed using R-package vegan (Oksanen et al., 2007) with 9999 permutations for a significance tests to detect the
The importance of environmental variables which correlated to genetic diversity and epigenetic variation was accessed by Boruta feature selection “random forest” analysis ($P < 0.05$) in R-package Boruta (Kursa, Rudnicki, & Kursa, 2020). Random forest (RF) is a robust supervised learning algorithm that can be used for a variety task including evaluate the importance of variance and classification problems. The RF model was popular used to predict the accuracy of classification when the aim variances were factors, otherwise, can be applied to measure the relative importance of each feature on the prediction. Environmental variables ranked by the RF model in order of genetic and epigenetic variation importance were determined over 1000 iterations and collected for the downstream analysis.

In order to quantified pure and combied effects of environmental variables on genetic and epigenetic variation, we performed the distance-based redundancy analysis (db-RDA) in R-package vegan. This method is a multivariate ordination technique used to test whether the variation of independent variables explains one to another. Furthermore, the contributions of environmental variables were derived from variance partitioning analysis (VPA) and hierarchical partitioning (HP) applying the algorithm of multiple regression in R-package rdacca.hp (Lai, Zou, Zhang, & Peres-Neto, 2021).

To identified the putative environment factors correlated to genetic clusters, we analyzed whether the environmental variables can be used as predictor to portion the genetic clusters of C. demersum populations. Random forest model with 10-fold cross-validation was constructed with the `rfcv` function in R-package randomForest (Liaw & Wiener, 2001). The minimum cross-validation error was obtained when using the correlated environmental variables detected above, according to this, we choose it as potential predictor in clustering the genetic groups of all populations.

Furthermore, as the pattern of genetic clusters may showed a non-liner correlation with the environmental factors, we then performed the generalized additive models (GAMs) to fit the environmental variables to the population clusters detected above. The generalized additive models (GAMs) was applied using 'ordisurf' function in R-package vegan. GAM models got the function of allowing for both a linear or nonlinear fit of the environmental variables to the detected genetic structure, and without a transformed environmental distance to fit the analysis, these models fitted the environmental data as a smooth response over the genetic structure accounting for both axes.

Next, to abtain the degree of environmental contributions at genetic diversity and epigenetic variation level. The partial least squares path model (PLS-PM), which has been widely used to study complex multivariate relations among variables, was performed to quantified direct and indirect effects of environment variables on genetic and epigenetic variation. The PLS-PM dose not require any distributional assumptions on the data which is usually difficult to meet in natural ecosystem. We established a model based on the expected relationships and key drivers among environmental variables, genetic and epigenetic variation with R-package plspm (Sanchez, Trinchera, Sanchez, & FactoMineR, 2013). In the model we compiled variables those highly linear to bio-climates as latent variable, a nonparametric bootstraping validation (1000 resamples) was performed to estimate the precision of parameters. The 95% bootstrap confidence interval was used to judge whether estimated path coefficients were significant. The final model was chose of all constructed models based on the goodness of fit ($\text{GoF} > 0.7$).

For the restoration purpose, we set up the areas of habitats for specie restoration mainly based on genetic diversity and epigenetic variation level in R-package prioritizr (Hanson et al., 2020). This package was used to help making a suggestion for suitable patches during species restorations. The 'prioritizr' package uses mixed integer linear programming (MILP) techniques to provide a flexible interface for building and solving restoration planning problems. It supports a broad range of constraints, that can be used to restoration planning problems to the specific needs of a restoration planning exercise. In addition, it possesses the function to find solutions in a much shorter period of time than other programs. We develop trade-offs
restoration prioritizations to identify priority areas for protected area establishment according to the degree of contributions from environmental variables to the genetic diversity and epigenetic variation. The aim of our restore purpose was to ensure that 20% of total genetic diversity and epigenetic variation among the surveyed populations were conserved (Faieller, Touron-Gardic, & Traore, 2019), thereby, environmental variables correlated to genetic and epigenetic indices were used as penalty and constraint factors to solve the restoration planning problems.

Results

Environmental gradient over sampling sites

We captured a subsential environment gradient over the samples area (Stress = 0.00004, Figure 2) according to the results of NMDS analysis. In addition, water temperature (WT), annual mean temperature (AMT), mean temperature of wettest quarter (MTWeQ) and mean temperature of coldest quarter (MTCoQ) were documented to show an obvious gradient at a relatively fine geographic waterbody scale.

Population genetic diversity and epigenetic variation

Based on six AFLP primers, a total of 483 bands were obtained from 12 populations of *C. demersum*. The number of bands generated by different primer combinations varied from 70 to 89 (Table S1). At species level, the percentage of polymorphic loci (*PPL*), mean Nei’s gene diversity (*H_E*) and mean Shannon’s information index of genetic diversity (*I*) was 63.94%, 0.219 and 0.331 respectively. At the population level, the *PPL* of each population varied from 43.48% (C5) to 91.93% (C10), with an average of 63.94%. Population C10 (*H_E* = 0.316, *I* = 0.476, *PPL* = 91.93%) exhibited the highest level of genetic diversity and population C5 (*H_E* = 0.123, *I* = 0.195, *PPL* = 43.48%) showed the lowest level of genetic diversity (Table 1).

Epigenetic variation was relatively higher than genetic diversity among each populations (mean *eH_E* = 0.229, *eI* = 0.350, *ePPL* = 69.27%), of which population C12 (*eH_E* = 0.335, *eI* = 0.495, *ePPL* = 88.67%) and C8 (*eH_E* = 0.151, *eI* = 0.240, *ePPL* = 52.53%) maintained the highest vice versa lowest level of genomic methylation (Table 1).

The population-based UPGMA tree revealed two genetic clades among populations of *C. demersum* based on the Jaccard genetic similarity coefficient (Figure 3). In addition, the methylated type of loci varied obviously among the genetic clades: population C9, C10, C11 and C12 exhibited lower full methylation type (mean FML = 68.61%) and higher unmethylated type (mean NMSL= 7.78%) (Figure. 3). Conversely, population C1- C8 showed lower FML (mean FML = 81.42%) and NMSL (mean NMSL = 2.37%) (Table S2). The non-hierarchical analysis of molecular variance (AMOVA) of genetic data revealed the component of genetic differentiation within populations was 89%, which indicating a strong genetic variation among populations. Hierarchical AMOVA attributed 9.1% of total variance to differences among two groups, and 6.2% genetic variance resided between groups among populations, where most variance was partitioned within populations (84.6%) (Table S3).

Population genetic structure

In population structuring analysis, the highest likelihood at K = 2 was revealed according to the Bayesian assignment (Figure 1 and Figure S1), therefore, we retained an optimal ancestry clades division within *C. demersum* populations at two clusters: populations from low temperature area (C9, C10, C11, C12) were fixed as cluster 1, whereas populations from high temperature area (C1, C2, C3, C4, C5, C6, C7, C8) were proportioned to the second cluster.

Correlations among genetic diversity, epigenetic variation with environmental variables

Mantel test revealed a significant pattern of correlation among genetic and epigenetic variation with environmental variables (*r* = 0.42, *P* < 0.01) (Figure. 4A). We filtered 13 environmental variables that account for over 1% importance of genetic and epigenetic indices with the RF model (Figure S2A and S2B). Environmental variables including water temperature (WT), annual mean temperature (AMT), mean temperature of
wettest quarter (MTWeQ) and mean temperature of coldest quarter (MTCoQ) were remarkably correlated to the genetic diversity and epigenetic variation indices (mantel’s $r > 0.2$, $P < 0.05$) (Figure 4B). Additionally, Procrust analysis according to the NMDS results which detected above further confirmed the strong associations between genetic clusters and the environmental stratum ($M^2 = 0.514$, $P = 0.005$) (Figure S3).

In redundancy analysis (db-RDA) results, the first two axes explained 99.77% of the variances, with db-RDA axes 1 and 2 accounted for 95.96% and 3.81% of the genetic and epigenetic variation, respectively (Figure 5A). The WT presented a vital effect (individual effect = 15.12%) and made the largest contribution to genetic diversity. In fact, WT, MTWeQ and AMT shared the effects on genetic diversity through variance partitioning analysis (explained variations = 32.89%). However, although MTCoQ exhibits the highest positive individual effect on epigenetic variation through hierarchical partitioning (individual effect = 29.05%), AMT overshadowed the influence (explained variation = 31.88%) and explained the largest portion of measured epigenetic variation (Figure 5B).

Non-negligibly, the genetic clades retained among populations were essentially matched the environment gradient measured in our NMDS analysis, that is, shown a coordinate between genetic clades and environment groups. Furthermore, mean temperature of coldest quarter (MTCoQ) was defined as environment predictor to distinguish the two genetic clusters as it showed the lowest cross-validation error through our random forest model (Figure S3C). The GAMs model further confirmed that MTCoQ was a strongly determinant of the genetic clusters among populations, which explained over half amount of deviance within our model (Figure 6).

Our PLS-PM model provided an expected robustness to our data, as suggest by the goodness of fit (GoF = 0.82). The PLS-PM analysis identified that WT exerts a positive influence with account for 65.8% ($R^2 = 0.658$) of variations on genetic diversity (path coefficient = 0.354, $P < 0.05$) and epigenetic variation (path coefficient = 0.483, $P < 0.05$) (Figure 7). In contrast, we observed that bio-climates negatively influenced the genetic diversity (path coefficient = -0.348) and epigenetic variance (path coefficient = -0.459), which contributed 71.4% of variations on them ($R^2 = 0.714$). Besides, genetic diversity was found to drive epigenetic variation positively (path coefficient = 0.278, $P < 0.05$) according to our model.

Species restorations area

Based on the restoration results, the lake area was divided into prioritize and relative importance areas for specie restoration based on genetic diversity and epigenetic variation level. Owing to the correlations among temperature, genetic and epigenetic variations, our solution of restoration scenarios was determined not only by genetic and epigenetic variation level, but also by temperature variables, that is, both the genetic data and the temperature are used as constraint factors during the construction of restoration patches (Figure 8).

Discussion

The geographic pattern of genetic and epigenetic variation

Unrooted aquatic plants usually possess a capacity of high level of dispersal, but can show explicit genetic variation among populations. The pattern of genetic variation and epigenetic variation is often driven by geographic and environmental influences exerted by geographic events, anthropogenic activities and climatic oscillations (Orsini, Vanoverbeke, Swillen, Mergeay, & De Meester, 2013; I. J. Wang & Bradburd, 2014). Geographic isolation is usually considered to be a necessity process during allopatric speciation, for example, geographic isolated populations usually occupy unique habitat specialist genotype since their gradual adaptions to the local conditions (Coyne, 2007; Karbstein et al., 2021). Due to this reason, genetic variation can not only demonstrate differentiation among populations, but also reflect ecological divergent among them. At fine geographic scale, especially in lakes, species dispersal is usually overriding the restriction of geographic barriers, particularly for some unrooted aquatic plants.

In our study, genetic and epigenetic variation doesn’t show a correlation against geographic distance, but reflects a positive correlation to the environmental distance, thereby, at least partially, meets the hypothesis
of isolation by environment (IBE) and make a congruent with empirical genetic studies on terrestrial plants (Herrera, Medrano, & Bazaga, 2017).

Correlations among genetic diversity, epigenetic variation and environment variables

Fine scale water environment has been proved to be hierarchical and highly vulnerable to climate changes, as the ever-changing of environment play a pivotal role in driving genetic diversity and epigenetic variation of aquatic plant species (Santamaria, 2002; Schulz et al., 2014). Due to the potential inheritance and rapid adjustment of adaption, it is essential to evaluate the species genetic diversity as well as epigenetic variation level even at fine-scale topographic and microclimatic variation (Foust et al., 2016; Zheng, Payne, & Wagner, 2019). Much of the previous genetic studies on aquatic plants neglected to characterize the correlation among environmental variables, species genetic diversity and epigenetic variation, such evidence that association abiotic factors to genetic and epigenetic variation remains scarce (Hughes & Stachowicz, 2009). Indeed, our study inferred a clear environmental gradient in Liangzi Lake which shaped by temperature factors. Expectedly, genetic and epigenetic variations of *C. demersum* populations in our analysis shown a correlative relationship which in line with this environmental gradient. The temperature scale has rendered a scenario where higher temperature area maintained a lower genetic and epigenetic variation, and lower temperature area conserved higher level of genetic and epigenetic variation.

For clonal macrophytes, although studies have revealed a concave-up and significant relationship between intra-population genetic diversity and epigenetic variation (M.-Z. Wang et al., 2020), the contributions of genetic diversity to epigenetic variation is still not been described clearly as epigenetic variation is, at least in part a subsidiary effect of genetic variation (Richards, Schrey, & Pigliucci, 2012). Our PLS-PM analysis revealed the direct effects of genetic diversity which exerts to epigenetic variation. However, while comparing the effect that driven by environmental variables, such influence can be overwhelmed during the slow adaption to the variation of environmental conditions.

The role of temperature on genetic variation and epigenetic variation

One of the major challenges in population genetic analysis lies in the distinguish of genetic divergence caused by historical events or anthropogenic disturbances (McKown et al., 2014; Inostroza et al., 2016). Although the Liangzi Lake conserved the highest diverse flora of aquatic plants due to its complex topography in Yangtze River basin, the environment conditions were constantly in a rapid changing process in recent years (Zhang, Dong, Yang, Odgaard, & Jeppesen, 2019). Seasonality of flood was recorded to cause the diminish of aquatic plant species, especially in the year 2016. Our field investigating on aquatic plants diversity in this area expanding 2019-2021 demonstrated that the population scale of aquatic plants was expeditiously recovered. Besides, the population scales of *C. demersum* decreased sharply in the year 2021 when compared to 2019, right at the starting of a 10-years fishing ban in key areas of Yangtze river basin. As *C. demersum* was feed by herbivorous fish species and crabs, we deduced that the diminished scenario may cause by the disequilibrium between aquatic herbs and herbivorous fish in this waterbody environment system.

Temperature-imposed selection usually leads to species local adaption across large geographic scale, where gradients of temperature often results in the spatial divergent thereby leads to the population divergence (Medina et al., 2021). Moreover, populations throughout the thermal regime exhibit different level of genetic diversity can confer benefits to the consumers, which further influence their genetic diversity and help to shape the population structure in return (Ruff et al., 2011; Farleigh et al., 2021). It is therefore an open question that how intraspecific genetic variation is being shaped in the context of temperature range where such environmental gradient was existed.

In our analysis, patterns of genetic variation across the range of the species illustrated how temperature affects the genetic and epigenetic variation. More importantly, water temperature was found to exert positive effects on genetic and epigenetic variation, while air temperature was identified to exert negative effects on genetic and epigenetic variation. As demonstrated, populations located in the lower temperature area exhibits higher genetic and epigenetic variation, where populations located in relative higher temperature shown
lower genetic and epigenetic variation. This distribution scenarios of genetic and epigenetic variation in lakes challenged the results of which high temperature areas maintained larger biodiversity than cold areas. However, our findings make a consistent with the analysis of the evolutionary potential-amount of total genetic variation that responds to natural selection of two reef-building corals, of which warm temperature drove the corals to rapid functional extinction in the absence of evolutionary potential (Walsworth et al., 2019).

In lake areas, macrophytes expands usually leads to the dissolved oxygen rise to a high range, then maintain the water temperature at a lower level than the sites where dissolved oxygen content is low, in other words, higher biodiversity areas in aquatic systems can keep the water temperature at low level (Hamberg, Fraser, Robinson, Trant, & Murphy, 2020; Jane, Hansen, Kraemer, Leavitt, & Rose, 2021). In addition, air and water temperature does not allow to have relations for all lakes in all climate regions, sites temperature usually depends on the depth of lake, especially where the lake is extremely shallow. Our study area located in the subtropical region in China, which secured the water temperature well above 0, meanwhile the air temperature can be lower than 0 in winter, thereby, the linear and non-linear relations between air and water temperature would disappear in this area. Since global average surface air temperature was expected to increase 3 at the end of this century (Friedlingstein et al., 2010), surface water temperature in lakes will be likely increased as air temperature increases. Although the change in water temperature may not be as large as the change in air temperature, these slight changes in water temperatures might affect the genetic and epigenetic variation of aquatic plants to some extent.

Moreover, as the detected genetic structure among *C. demersum* populations was not divergent by geographic but separated by temperature, our results gain a further emphasis on the importance of temperature in shaping the pattern of aquatic plants genetic variation at a fine geographic scale. Meanwhile, empirical studies mainly focus on large geographic scale and mitigate the temperature influences as detailing the habitats variables can be a daunting task (Liang, Zhang, & Wei, 2022), our analysis largely deprecated the geographic impact on genetic analysis and fills a knowledge gap in the field of temperature effects on aquatic plants genetic analysis, which further advanced the understandings of the robust relationships between genetic, epigenetic variation and environmental variables.

**Restoration implications**

Generally, species genetic diversity which reflects the reservoir of species for short-term ecological adaptation and long-term evolutionary changes was applied to constructed the restoration siting (Teixeira & Huber, 2021). When target on aquatic species restoration, little is known about genetic variation and species-sites interactions, especially restoration projects accounted for extreme climate events or anthropogenic influences (Forester, Beever, Darst, Szymanski, & Funk, 2022). Furthermore, genetic diversity is considered crucial for species adaptation to unforeseen climate changes and maintained resilience to abiotic and biotic stresses. The Liangzi Lake is now encompassed to the protected area and buffered from the anthropogenic threats since the year 2020. As protected area should not only conserve the intraspecific genetic diversity but also habitats heterogeneity, scaling potential relationships between genetic and epigenetic variation and environment variables, and linking those to fitness would offers an option for constructing restoration portfolios when the environmental pressures are difficult to measure (Bay, Rose, Logan, & Palumbi, 2017). In our analysis, the suitable areas for species restorations were not only depends on the genetic diversity and epigenetic variation level, but also taking into account the incremental temperature benefits. Trade-offs restoration approach between genetic diversity, epigenetic variation and temperature was made when generating prioritized restoration areas.

**Conclusion**

Our genetic analysis of *C. demersum* in the Liangzi lake delineated a cryptic correlation between genetic diversity, epigenetic variation and environment variables, and set the bias for estimate temperature in shaping the geographical pattern of genetic variation of clonal aquatic herbs in thermal gradients. Environmental variables play a congruently role in driving the species adaption congruently with genetic diversity and
epigenetic variation. Genetic and epigenetic variations among populations can override the influences of geographic isolation when significant environmental gradients were existing. Moreover, species restoration strategy should take the temperature into account even at a fine scale, especially for the aquatic plants, as their suitable habitats were usually confined to a restricted water environment scale.

Acknowledgments

Sample collection was supported by Liangzi lake reserves. We thank Wenjun Guo for the help in sampling. Founding was provided by the National Key R & D Program of China (2018YFD0900801). Author contributions: HW and YX conceived and designed the research; YX conducted the experiments and data analysis, generated the figures and write the manuscript with the contributions from XY, ZX, JJ, GJ, XZ and ZL; all authors approved the final manuscript.

Data accessibility

All codes and polymorphism raw datas are openly available from the Github repository: yixian185/For-FEE: Data and code (github.com)

References:


Table 1. Details of sample sites, genetic diversity and epigenetic variation for 12 populations of *C. demersum* in Liangzi lake

<table>
<thead>
<tr>
<th>Code</th>
<th>Sample numbers</th>
<th>Longitude</th>
<th>Latitude</th>
<th>$H_E$ (%)</th>
<th>$I$ (%)</th>
<th>PPL (%)</th>
<th>$eH_E$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>8</td>
<td>114.562</td>
<td>30.249</td>
<td>0.163</td>
<td>0.243</td>
<td>44.93</td>
<td>0.224</td>
</tr>
<tr>
<td>C2</td>
<td>11</td>
<td>114.474</td>
<td>30.366</td>
<td>0.233</td>
<td>0.355</td>
<td>71.01</td>
<td>0.252</td>
</tr>
<tr>
<td>C3</td>
<td>6</td>
<td>114.623</td>
<td>30.176</td>
<td>0.173</td>
<td>0.261</td>
<td>48.65</td>
<td>0.216</td>
</tr>
<tr>
<td>C4</td>
<td>7</td>
<td>114.637</td>
<td>30.098</td>
<td>0.190</td>
<td>0.292</td>
<td>57.35</td>
<td>0.214</td>
</tr>
<tr>
<td>C5</td>
<td>9</td>
<td>114.563</td>
<td>30.144</td>
<td>0.123</td>
<td>0.195</td>
<td>43.48</td>
<td>0.212</td>
</tr>
<tr>
<td>C6</td>
<td>8</td>
<td>114.530</td>
<td>30.280</td>
<td>0.208</td>
<td>0.318</td>
<td>62.53</td>
<td>0.247</td>
</tr>
<tr>
<td>C7</td>
<td>7</td>
<td>114.467</td>
<td>30.280</td>
<td>0.201</td>
<td>0.306</td>
<td>59.21</td>
<td>0.157</td>
</tr>
<tr>
<td>C8</td>
<td>9</td>
<td>114.456</td>
<td>30.112</td>
<td>0.198</td>
<td>0.304</td>
<td>61.28</td>
<td>0.151</td>
</tr>
<tr>
<td>C9</td>
<td>6</td>
<td>114.511</td>
<td>30.361</td>
<td>0.278</td>
<td>0.410</td>
<td>72.05</td>
<td>0.263</td>
</tr>
<tr>
<td>C10</td>
<td>14</td>
<td>114.555</td>
<td>30.364</td>
<td>0.316</td>
<td>0.476</td>
<td>91.93</td>
<td>0.275</td>
</tr>
<tr>
<td>C11</td>
<td>15</td>
<td>114.614</td>
<td>30.268</td>
<td>0.322</td>
<td>0.481</td>
<td>90.89</td>
<td>0.308</td>
</tr>
<tr>
<td>C12</td>
<td>10</td>
<td>114.587</td>
<td>30.325</td>
<td>0.336</td>
<td>0.497</td>
<td>88.41</td>
<td>0.335</td>
</tr>
</tbody>
</table>

Mean 0.219 0.331 63.94 0.299

Abbreviation: $H_E$, expected heterozygosity of genetic diversity; $eH_E$, expected heterozygosity of epigenetic variation; $I$, Shannon’s information index of genetic diversity; $eI$, Shannon’s information index of epigenetic variation; PPL (%), percentage of polymorphic loci; ePPL(%), percentage of epigenetic polymorphic loci.

Hosted file

[Figure.docx available at](https://authorea.com/users/512852/articles/589228-temperature-is-a-cryptic-factor-to-shape-the-geographical-pattern-of-genetic-variation-in-ceratophyllum-demersum-across-a-subtropical-freshwater-lake)