

Bayesian Analysis of Population Structure and Gene Flow in *Chara* (Chryophyceae) Species

Akram Ahmadi¹, Masoud Sheidai^{1*}

Received: 2017.08.23

Revised and accepted: 2017.11.25

Abstract

Chara is a morphologically variable genus. Molecular analysis (ISSR) of genetic diversity and population structure was performed on *Chara* species. Population groups identified based on geographical provinces showed a significant genetic difference, but Mantel test did not show isolation by distance in the studied species. Bayesian analysis of population structure grouped *Chara* species in 2 distinct genetic groups differing in allelic composition and frequency. STRUCTURE analysis revealed genetic admixture among species which was supported by reticulation analysis. Coalescence analysis showed the occurrence of gene duplication and extinction as possible evolutionary changes along with polyploidy as the main forces of speciation in the genus *Chara*.

Key words: *Chara*, Gene exchange, Isolation by Distance.

Introduction

Characeae is a large and unique family of algae with about 300 species, characterized by the complexity of their morphological features, including the structure of their gametangia and their axis differentiated

into nodes and internodes (Picelli-Vicentini et al., 2004). They grow in freshwater, in brackish and semi-terrestrial environments. They range in size from a few millimeters to over a meter in length. The filament internodes are unicellular (sometimes may be covered by subsequent growth of corticated filaments), while filament nodes have a complex, parenchyma-like organization.

The Charales have an excellent fossil record extending back far over 420 million years, and two extant lineages. The Characeae and Nitellaceae can be traced back to roughly 200 million years ago (Mattox and Stewart, 1984). These algae are considered the closest living relatives of land plants (Karol et al., 2001) and recent molecular study of mitochondrial DNA and chloroplast DNA showed that despite important differences in size and intron composition, *Chara* mtDNA strikingly resembles *Marchantia* mtDNA and genome comparisons and phylogenetic analyses based on mtDNA and cpDNA provided unequivocal support for a sister-group relationship between the *Charales* and the land plants (Turmel et al., 2003, 2006). These green algae also play an important ecological role in aquatic ecosystems as their assemblage increases water transpar-

¹- Faculty of Life Sciences and Biotechnology, University of Shahid Beheshti, Evin, Tehran, Iran.
*email: msheidai@yahoo.com

ency and they act as efficient nutrient sinks and influence zooplankton and phytoplankton biomasses (Meurer and Bueno, 2012). In some places, Characeae species constitute much of the submerged aquatic plant life, and thus contribute through photosynthesis to the oxygen balance in natural waters.

Charophytes are a taxonomically difficult group, both with regards to species identification and in relation to other algal groups (Griffin, 1963). High variability within species separated by small morphological differences can prevent the determination of species and subspecies, varieties and forms. Further, the morphology of members of Characeae are often affected by their ecological environment (Urbaniak, 2010). Extensive reshuffling occurs in the taxonomic status of Charophytes; large number of species were reduced to the status of subspecies, variety or forma; while some distant species were merged together (Abrol and Bhatnagar, 2006).

Few molecular studies have been performed to study genetic diversity among and within Characeae species; for instance Mannschreck et al. (2002) used amplified fragment length polymorphism (AFLP) to study biosystematics of *C. hispida*, *C. intermedia* and *C. tomentosa* from the genus *Chara*, while Abrol and Bhatnagar (2006) used Random amplified polymorphic DNA (RAPD) markers to investigate biodiversity of 12 Indian charophyte taxa.

Schaible et al. (2011) used microsatellite markers (SSRs) and Bayesian assignment method to determine population genetic

structure in sympatric sexually and Parthenogenetically reproducing population of *C. canescens* (Charophyta) in Europe; while Lewis and Lewis (2005) performed molecular phylodiversity of *Chara* species growing in desert area by using 18S rDNA and Bayesian analysis showing substantial molecular diversity in these taxa and that desert lineages are distantly related to their nearest aquatic relatives. Schneider et al. (2015) used barcodes of the ITS2, matK and rbcL regions to test if the distribution of barcode haplotypes among individuals of 91 specimens from 10 European countries, Canada and Argentina. They found out that herbarium specimens also for aquatic plants like *Chara* are useful as a source of material for genetic analyses and rbcL and matK had highest sequence recoverability, but matK and ITS2 had higher discriminatory power than rbcL.

Limited systematic studies are available on charophytes of Iran only recently reports concerned with morphological and micromorphological diversity of *Chara* species have been published (Ahmadi et al., 2012a,b,c). The present study considers cytology, genome size and molecular diversity of some *Chara* species of Iran for the first time and tries to use these data to investigate species relationship.

Materials and Methods

Plant material

Studies were performed on 18 populations of *Chara* species: *Chara gymnophylla* (A. Br.) A. Br. var. *gymnophylla*, *C. gymno-*

phylla var. *rohlena* (Vilh.) Fil., *C. vulgaris* var. *longibracteata* (Kutz.) J. Gr. & B.-W., *C. vulgaris* L. var. *vulgaris*, *C. contraria* A. Br. ex Kutz., *C. kirghisorum* Less., *C. crassicaulis* Schl. ex A. Br., *C. socotrensioides* R.D.W., *C. tomentosa* L., *C. connivens* Salz. ex A. Br., *C. pedunculata* Kutz., *C. fibrosa* C. Ag. ex Br., *C. kohrangiana* Ahmadi et al. and *Nitella hyalina* (DC.) C. Ag., as out-group.

ISSR assay

Ten ISSR primers: (AGC)5GT, (CA)7GT, (AGC)5GG, UBC 810, (CA)7AT, (GA)9C, UBC 807, UBC 811, (GA)9A and (GT)7CA commercialized by UBC (the University of British Columbia) were used. PCR reactions were performed in a 25 µl volume containing 10 mM Tris- HCl buffer at pH 8; 50 mM KCl; 1.5 mM MgCl₂; 0.2 mM of each dNTP (Bioron, Germany); 0.2 µM of a single primer; 20 ng genomic DNA and 3U of Taq DNA polymerase (Bioron, Germany). Amplification reactions were performed in Techne thermocycler (Germany) with the following program: 5 min initial denaturation step 94°C, 30 s at 94°C; 1 min at 50°C and 1 min at 72°C. The reaction was completed by final extension step of 7 min at 72°C. Amplification products were visualized by running on 2% agarose gel, following ethidium bromide staining. Fragment size was estimated by using a 100 bp molecular size ladder (Fermentas, Germany). The experiment was replicated 3 times and constant ISSR bands were used for further analyses.

Data analyses

ISSR bands obtained were treated as bina-

ry characters and coded accordingly (presence = 1, absence = 0). Based on geographical distribution of species and populations studied, they were divided into 6 population groups for genetic diversity analysis.

Genetic diversity parameters determined were percentage of allelic polymorphism, allele diversity (Weising et al., 2005), Nei's gene diversity (H), Shannon information index (I) (Weising et al., 2005; Freeland et al., 2011), number of effective alleles and percentage of polymorphism. Furthermore, AMOVA (Analysis of molecular variance) test (with 1000 permutations) as performed in GenAlEx 6.4 (Peakall and Smouse, 2006), was used to show molecular difference among the population groups. Mantel test (Podani, 2000) was performed to study association between molecular distance and geographical distance of the populations.

Dice as well as Nei's genetic distance (Weising et al., 2005; Freeland et al., 2011) was determined among trees and used for the grouping of the genotypes by unweighted paired group method with arithmetic average (UPGMA) and Neighbor Joining (NJ) clustering methods after 100 times bootstrapping (Freeland et al., 2011). Similarly, ordination plot was utilized based on principal co-ordinate analysis (PCoA), as well as Multidimensional scaling (MDS) (Podani, 2000), using PAST ver. 2.17 (2012, Hammer et al., 2001) and DARwin ver. 5 (2012).

Bayesian clustering method was performed to elucidate the populations, genetic structure by using STRUCTURE v. 2.3 (Pritchard et al., 2000). The program structure imple-

ments a model-based clustering method for inferring population structure using genotype data consisting of unlinked markers. For this reason, we first performed linkage disequilibrium test for SSR loci as implemented in POPGENE ver. 1.32 (2000).

The model applied in STRUCTURE analysis assumes the existence of K clusters (Pritchard et al., 2000). The Markov chain Monte Carlo simulation was run 20 times for each value of K (5) for 10^6 iterations after a burn-in period of 10^5 . All other parameters were set at their default values. Data were entered as suggested by Falush et al. (2007) and data sample provided in STRUCTURE home page. STRUCTURE Harvester web site (Earl and von Holdt, 2012) was used to visualize the STRUCTURE results and also to perform Evanno method to identify proper number of K (Evanno et al., 2005). The choice of the most likely number of clusters (K) was carried out comparing log probabilities of data $[\Pr(X|K)]$ for each value of K (Pritchard et al., 2000), as well as by calculating an *ad hoc* statistic ΔK based on the rate of change in the log probability of data between successive K values, as described by Evanno et al. (2005). Genetic differentiation of population subgroups was determined by F_{st} determined by STRUCTURE. Reticulation was performed by T-REX (Tree and Reticulogram Reconstruction) ver. 3 (2000), and DARwin ver. 5 (2012) which infer reticulogram from distance matrix. For reticulation, we first built a supporting phylogenetic tree using Neighbor Joining (NJ), followed by a reticulation branch that

minimizes the least-squares at each step of the algorithm (Legendre and Makarenkov, 2002). Furthermore, coalescence analysis was performed for molecular data after 1000 reiterations (Liu et al., 2009), as suggested for SSR polymorphic data by Wilson and Balding (1998) and performed in Mesquite (Maddison and Madson, 2011). Gene tree heterogeneity and discordance with the species/population tree was checked by parameters provided in Mesquite (Maddison and Madson, 2011), including deep coalescence, gene duplication and extinction.

Results

Genetic diversity analysis

All ISSR primers used produced 67 bands in 18 *Chara* species and populations studied. ISSR band No. 17 was present in all *Chara* species studied but was absent in outgroup species *Nitella*, while band 20 occurred only in *Nitella hyalina* and was absent in all *Chara* species. ISSR loci 35, 50 and 69 were private band for *C. fibrosa.*, while band 49 occurred only in *C. crassicaulis*.

Dentrented correspondence analysis plot obtained including Convex hulls as well as 95% Ellipses methods (Fig. 1), showed scattered distribution of all ISSR loci in the plot, indicating they are well scattered in the genome and possibly are not correlated to each other. G_{st} analysis revealed that ISSR loci have G_{st} values of 0.05 to 0.80 with mean G_{st} value of 0.42, which is considered a low to moderate value. This indicates the presence of shared common alleles and possible gene exchange among *Chara* species

(which will be discussed more in following paragraphs).

Genetic diversity analysis performed among 5 population groups based on geographical provinces (Table 1) revealed a higher degree of genetic polymorphism in population group 2 (Markazi province) (83.58%), followed by population group 4 (Chaharmahal and Bakhtiari province) (71.64%), compared to other population groups. Similarly, Markazi province had a higher value

of effective alleles (N_e) (1.41) and expected heterozygosity (H_e , 0.26) compared to other provinces. Nei's genetic identity determined among 5 provinces varied from 0.70 between populations 3 (Kohgiluyeh and Boyer-Ahmad) and 5 (Kurdistan) to 0.87 between populations 1 (Esfahan) and 3 (Kohgiluyeh and Boyer-Ahmad). Similarly, these populations had the highest and lowest Nei's genetic distance values (0.35 and 0.14 respectively). When UPGMA tree based

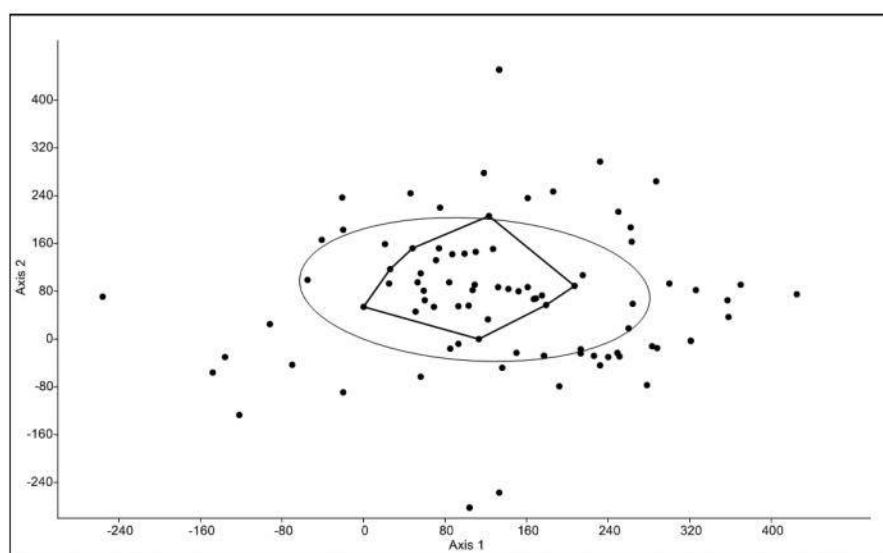


Fig. 1. Dentrented correspondence analysis of ISSR loci.

Table 1. Genetic diversity parameters in studied provinces.

Pop	Province	N	Na	Ne	I	He	%P
Pop1	Esfahan	4.000	1.373	1.314	0.298	0.194	59.70%
Pop2	Markazi	6.000	1.672	1.429	0.401	0.261	83.58%
Pop3	Kohgiluyeh and Boyer-Ahmad	2.000	0.970	1.253	0.217	0.148	35.82%
Pop4	Chaharmahal and Bakhtiari	5.000	1.478	1.347	0.337	0.216	71.64%
Pop5	Sanandaj-Kurdistan	1.000	0.299	1.000	0.000	0.000	0.00%
Total		3.600	1.158	1.269	0.250	0.164	50.15%

N = No. of populations/species in the group, Na = No. of Different Alleles,
 Ne = No. of Effective Alleles = $1 / (p^2 + q^2)$, I = Shannon's Information Index
 = $-1 * (p * \ln(p) + q * \ln(q))$, He = Expected Heterozygosity = $2 * p * q$,
 %p = Polymorphism percentage.

on Nei's genetic distance of provinces was drawn (Fig. 2), Markazi and Charmahal and Bakhteiri (pops 2 and 4) were placed close to each other, followed by Esfahan and Kuhgiluyeh and Boyer-Ahmad provinces (pops 1 and 3), while Kurdistan (pop 5) differed from other provinces and joined the others with a greater distance.

AMOVA test showed significant genetic difference ($p < 0.05$) among provinces and revealed that 11% of total genetic variation is due to among group difference, while 89% is due to within group genetic variation. This indicates the presence of high degree of genetic difference among as well as within each province.

Mantel test (Fig. 3) showed no significant correlation ($p = 0.40$) between geographical distance and genetic distance. Therefore, isolation by distance does not occur in *Chara* species studied. This is also supported by F_{st} values of population groups determined by STRUCTURE. These values varied from 0.00-0.07, which are very low F_{st} values showing gene flow among geographical regions. F_{st} values obtained for population groups are in agreement with G_{st} results of ISSR alleles presented earlier, both indicating presence of shared common alleles and gene exchange among populations.

Genetic affinity of Chara species and populations

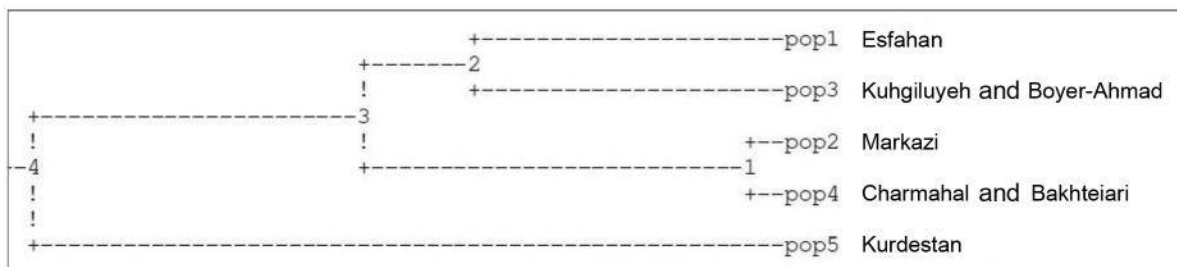


Fig. 2. UPGMA tree of provinces based on Nei's genetic distance.

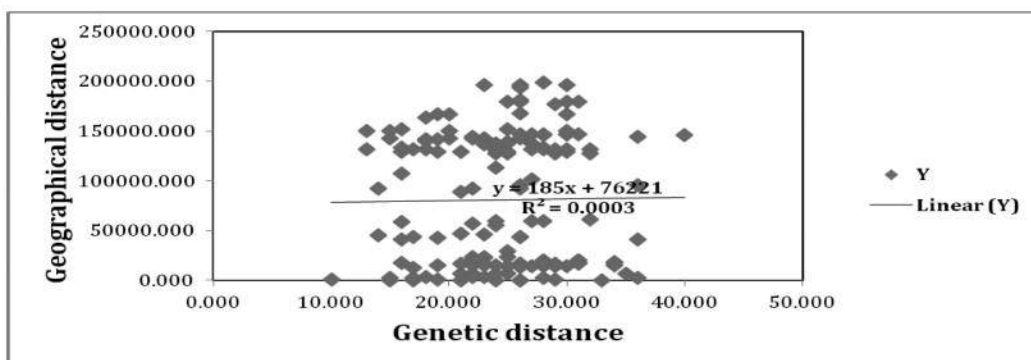


Fig. 3. Mantel plot between geographical and genetic distance.

NJ and UPGMA tree of ISSR data produced similar results, but since NJ tree had a higher cophenetic correlation (0.93), it is discussed here (Fig. 4). Outgroup species, *Nitella hyallina*, is placed separate from *Chara* species studied, while our ingroup taxa are placed in 2 major clusters. In the first major cluster, *C. vulgaris* var. *vulgaris* and *C. vulgaris* var. *longibracteata*, and populations of *C. tomentosa* are placed close to each other and along with *C. crassicaulis*, *C. connivens* and *C. pedunculata* form this major cluster. The second major cluster contains 3 sub-clusters, showing closer affinity between *C. gymnophylla* var. *rohlena* and *C. gymnophylla* var. *gymnophylla*, *C. contraria* and *C. kohrangiana* which along with 1 population of *C. vulgaris* var. *longibracteata* comprise the first sub-cluster. *C. fibrosa* and *C. kirghisorum* form the second sub-cluster, while *C. socotrensioides* and 1 population of *C. gymnophylla* comprise the third sub-cluster.

Close affinity between *C. vulgaris* and *C. crassicaulis*, *C. contraria* and *C. kohrangiana*, as well as between *C. gymnophylla* and *C. contraria* is in agreement with our earlier morphometrical study (Ahmadi et al., 2012b). Moreover separation of *C. fibrosa*, *C. kirghisorum* and *C. socotrensioides* from *C. vulgaris* and *C. gymnophylla* and their position in a separate cluster is also supported by morphology.

MDS and PCoA plots separated different species and populations in 2 major groups (Figures not given), supporting NJ tree. *Nitella hyallina*, the outgroup species was separated from the other *Chara* species and populations of *C. vulgaris* and *C. gymnophylla* were placed far from each other due to genetic variation. Test for linkage disequilibrium performed for ISSR loci, as implemented in POPGENE, did not show any significant association between them supporting Dentrended correspondence anal-

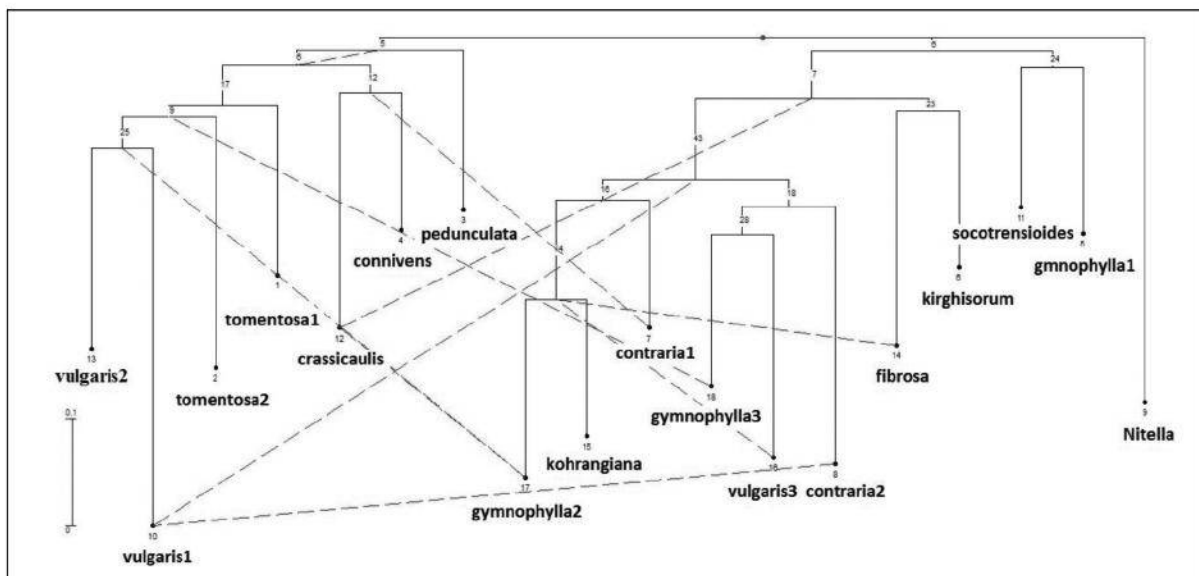


Fig. 4. NJ tree and reticulation of *Chara* species and populations. (numbers above branches are bootstrap values, dashed lines indicate possible gene exchange).

ysis. Therefore, these molecular markers were considered to meet the assumptions for applying the Bayesian method implemented in the program STRUCTURE to assign individuals to population groups.

Q-matrix plot of STRUCTURE analysis (Fig. 5) showed presence of 2 major subgroups. Evanno method and ad hoc statistic ΔK also showed $k = 2$ as the best number of population subgroups (figure not given). STRUCTURE plot showed that populations of *C. tomentosa* (No. 1 and 2) and *C. contraria* (No. 9 and 10) populations differed only in allele frequencies (different proportion of segments with similar colors), while *C. gymnophylla* differed in allelic composition (different colours). This holds true particularly for *C. gymnophylla* var. *rohlena* collected from Markazi province (No.

5 in Fig. 5). Similarly, *C. vulgaris* populations differed in both allelic composition and frequency. Mannschreck et al. (2002) used AFLP markers to study 33 samples of 3 *Chara* species, *C. hispida*, *C. intermedia* and *C. tomentosa* and found great interspecific molecular diversity.

CCA plot (Fig. 6), showed that all 3 ecological factors studies contribute to distribution of *Chara* species, however, we may say that latitude has prominent effect on *C. tomentosa*, while combination of both latitude and altitude affect *C. gymnophylla* and in particular combination of longitude and altitude affect *C. contraria* and *C. vulgaris*.

Reticulogram (Fig. 4) showed gene exchange/shared common genetic loci among most of the *Chara* species and populations studied. Gene exchange occurred between

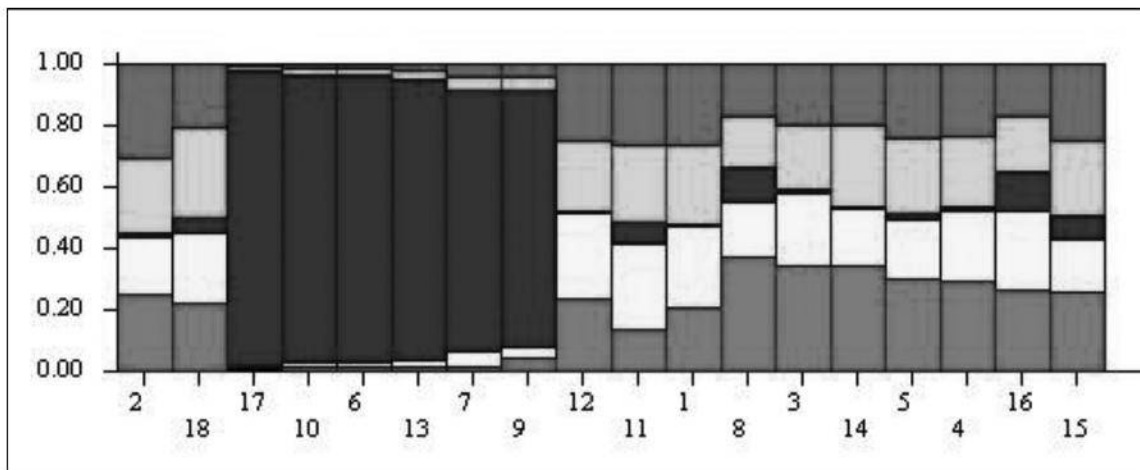


Fig. 5. Q-matrix plot of STRUCTURE.

Populations abbreviations: 1 & 2 = *C. tomentosa*, 3 = *C. pedunculata*, 4 = *C. connivens*, 5-7 = *C. gymnophylla*, 8 = *C. kirghisorum*, 9 & 10 = *C. contraria*, 11-13 = *C. vulgaris*, 14 = *C. socotrensioides*, 15 = *C. crassicaulis*, 16 = *C. fibrosa*, 17 = *C. kohrangiana*, and 18 = *Nitella hyalina*.

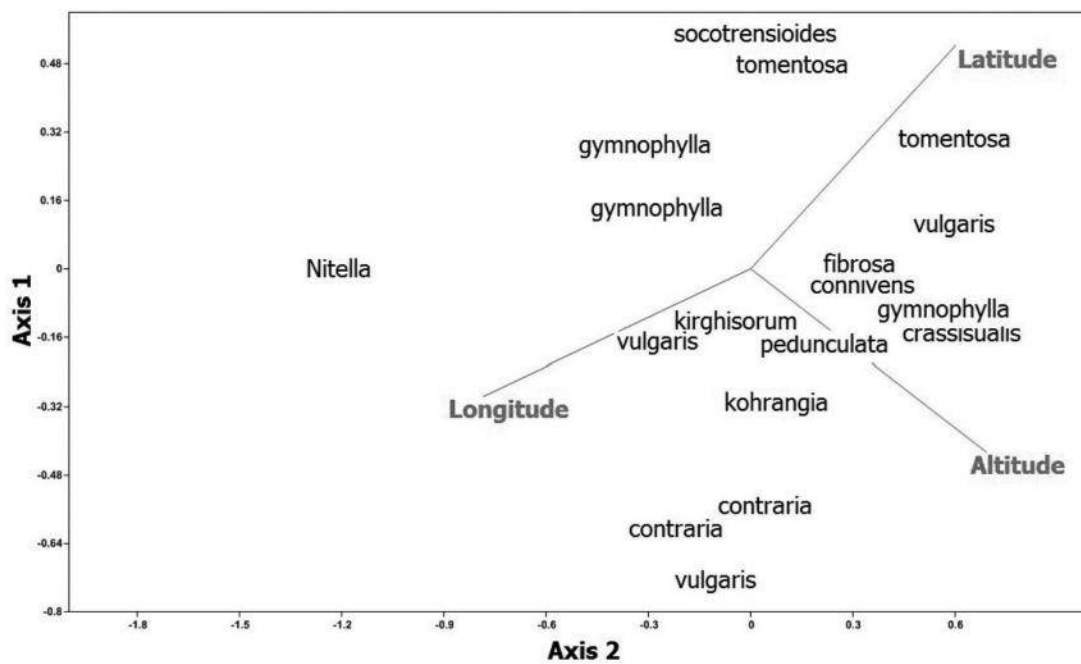


Fig. 6. CCA plot of ISSR data.

species from both clusters of NJ tree. These results support STRUCTURE analysis results and further illustrated the source of common allelic composition identified in STRUCTURE plot.

Coalescence analysis produced gene trees which grouped *Chara* species somewhat different from NJ tree and when gene trees were contained in population tree, the best result obtained showed deep coalescence cost of 23 with 8 duplication and 39 extinction (Fig. 7).

To summarize the findings, we may say that ISSR markers showed genetic diversity among and within *Chara* species. These molecular markers discriminate *Nitella hyalina* from *Chara* species and also can reveal *Chara* species relationship. Bayesian analysis of data showed genetic differences of the species studied and grouped them in 2

separate genetic groups and indicated great genetic admixture among species as also supported by reticulation analysis. Coalescence results showed that along with cytological change in *Chara* species which is a main evolutionary mechanism, gene duplication and extinction have played a role in species diversification.

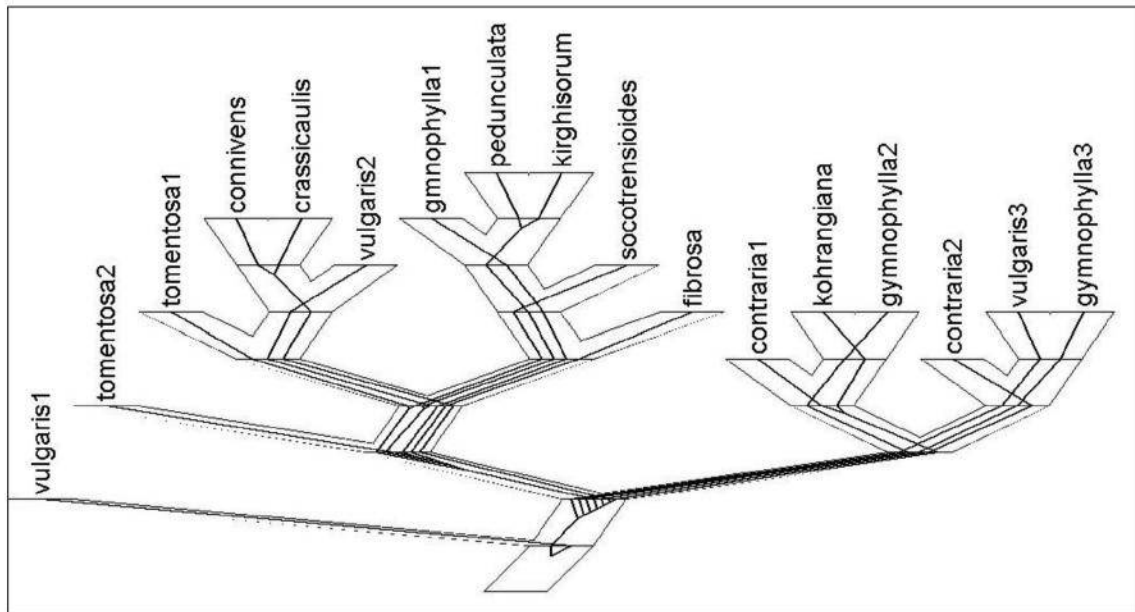


Fig. 7. Coalescence tree of *Chara* species. (gene tree is contained in species tree).

References

- Abrol D and Bhatnagar SK. (2006). Biodiversity of few Indian charophyte taxa based on molecular characterization and construction of phylogenetic tree. *African Journal of Biotechnology*. 5 (17): 1511-1518.
- Ahmadi A, Riahi H, Sheidai M, Van Raam J. (2012a). Some Charophytes (Characeae, Charophyta) from Central and Western of Iran Including *Chara kohrangiana* species nova. *Cryptogamie, Algologie*. 33 (4): 359-390.
- Ahmadi A, Riahi H, Sheidai M, VanRaam J. (2012b). Numerical taxonomy of some charophytes in Iran. *Nordic journal of botany*. 30: 206-214.
- Ahmadi A, Riahi H, Sheidai M, VanRaam J. (2012c). A study of the oospore characteristics in some Charophytes (Characeae) of Iran. *Nova Hedwigia*. 94: 487-504.
- Earl DA. and von Holdt BM. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*. 4 (2): 359-361. Doi: 10.1007/s12686-011-9548-7
- Evanno G, Regnaut S, Goudet J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*. 14: 2611-2620.
- Falush D, Stephens M, Pritchard JK. (2007). Inference of population structure using multi-locus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*. 7 (4): 574-578.
- Freeland JR, Kirk H, Peterson SD. (2011). *Molecular Ecology* 2nd edition. Wiley-Blackwell, UK. pp 449.
- Griffin DG. (1963). Variation in oospores of six species of *Chara*. *Bulletin of Torrey Botanical Society*. 90 (6): 400-402.
- Hammer Ø, Harper DAT, Ryan PD. (2001). PAST: Paleontological statistics software

- package for education and data analysis. *Palaeontologia Electronica*. 4 (1): 9 pp.
- Karol KG, Mccourt RM, Cimino MT, Delwiche CF. (2001). The closest living relatives of land plants. *Science*. 294: 23-51.
- Legendre P. and Makarenkov V. (2002). Reconstruction of biogeographic and evolutionary networks using reticulograms. *Systematic Biology*. 51 (2): 199-216.
- Lewis LA and Lewis PO. (2005). Unearthing the molecular phylogeny of desert soil green algae (Chlorophyta). *Systematic Biology*. 54 (6): 936-047.
- Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV. (2009). Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution*. 53: 320-328.
- Maddison WP and Madson DR. (2011). Mesquite: a modular system for evolutionary analysis. Version 2.75. <http://mesquiteproject.org>.
- Mannschreck B, Fink T, Melzer A. (2002). Biosystematics of selected *Chara* species (Charophyta) using amplified fragment length polymorphism. *Phycologia*. 41 (6): 657-666.
- Mattox KR and Stewart KD. (1984). Classification of the green algae: a concept based on comparative cytology. pp. 29-72 in DEG Irvine and DM John, *Systematics of the Green Algae*, Systematics Association Special Volume. Academic Press, London and Orlando.
- Peakall R, Smouse PE. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*. 6: 288-295.
- Picelli-Vicentim MM and Bicudo CEM. (1993). *Criptógamos do Parque Estadual das Fontes do Ipiranga, São Paulo, SP*. Algas, 4: Charophyceae. *Hoehnea*. 20: 9-22.
- Podani J. (2000). Introduction to the Exploration of Multivariate Data [English translation], Backhuyes, Leide., Ltd.
- Pritchard JK, Stephens M, Donnelly P. (2000). Inference of population structure using multilocus genotype Data. *Genetics*. 155: 945-959
- Schaible R, Bergmann I., Schubert H. (2011). Genetic Structure of Sympatric Sexually and Parthenogenetically Reproducing Population of *Chara canescens* (Charophyta). *ISRN Ecology*, pp. 13. Doi:10.5402/2011/501838
- Schneider CS, Rodrigues A, Moe TF, Ballot A. (2015). DNA Barcoding the Genus *Chara*: Molecular Evidence Recovers Fewer Taxa Than the Classical Morphological Approach. 51: 367-380.
- Urbaniak J. (2010). Analysis of morphological characters of *Chara baltica*, *C. hispida*, *C. horrida* and *C. rudis* from Europe. *Plant Systematics and Evolution*. 286: 209-221.
- Weising K, Nybom H, Wolff K, Kahl G. (2005). *DNA Fingerprinting in Plants. Principles, Methods and Applications*. (2nd ed) Taylor & Francis. pp 444.
- Wilson IJ and Balding DJ. (1998). Genealogical inference from Microsatellite data. *Genetics*. 150 (1): 499-510.