Research Article

Investigation of genetic diversity and phylogenetic relationship of *Ctenopharyngodon idella* from different regions of Punjab by using SSR Markers

Muhammad Nasir Khan Babar¹, Tahir Naqqash², Aasma Noureen¹, Aneela Kanwal¹, Muhammad Baber², Nabeela Tariq³, Nadia Noureen⁴ and Tasleem Kausar^{4*}

1. Department of Zoology, Virtual University of Pakistan

2. Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan, Pakistan

3. Department of Zoology, Sardar Bahadar Khan Women University, Quetta, Pakistan

4. Department of Zoology, Government Sadiq College Women University, Bahawalpur, Pakistan

*Corresponding author's email: tasleem.kausar@gscwu.edu.pk

Citation

Muhammad Nasir Khan Babar, Tahir Naqqash, Aasma Noureen, Aneela Kanwal, Muhammad Baber, Nabeela Tariq, Nadia Noureen and Tasleem Kausar. Investigation of genetic diversity and phylogenetic relationship of *Ctenopharyngodon idella* from different regions of Punjab by using SSR Markers. Pure and Applied Biology. Vol. 11, Issue 1, pp209-216. <http://dx.doi.org/10.19045/bspab.2022.110022>

Received: 10/04/2021

Revised: 20/05/2021

Accepted: 24/05/2021

Online First: 26/05/2021

Abstract

Genetic monitoring is a commonly used tool to find the genetic diversity, for this purpose, simple sequence repeat (SSR) markers is best implemented because of its polymorphic nature, low cost, variability and distribution. Ten SSR markers were used for the investigation of genetic variations in grass carp *Ctenopharyngodon idella* (*C. idella*). A total of 15 samples were collected from Khan Garh, Ghazi Ghatt and river Chenab Pakistan. Target loci were amplified and products were resolved on 10 % polyacrylamide gel. Mean value of polymorphism for all ten markers was 77.57%. Within 422 loci scored, 366 were found polymorphic in all samples. The allele number varied from 2 to 6 with average 3.70 values and allelic frequency ranged from 0.6777 to 0.9111 with average value of 0.693. Gene diversity was found from 0.3011 to 0.5066 with average 0.3950 and PIC value ranged from 0.2366 to 0.4039 with average value of 0.2726. The values of Nei's genetic distant among populations of river Chenab were 0.1764, and 0.4411, while values of Jaccard's similarity coefficient were 0.7500 and 0.9167. These values showed that genetic makeup of *C. idella* from different regions of river Chenab varied from each other. Conclusion: The results showed that samples from same study area exhibited more genetic similarity but different from other areas indicating the genetic diversity of grass carp in different areas. It can be concluded that climatic variables, reproductive and geographic isolation influenced the genetic diversity in grass carp.

Keywords: *Ctenopharyngodon idella*; Polymorphism; SSR Markers

Introduction

Ctenopharyngodon idella (*C. idella*) is among the top ten aquaculture species of the world,

commonly known as grass carp is a popular fish in South-East Asian countries such as

India, Pakistan, Myanmar and Bangladesh [1]. Because of its good taste and rapid growth, this species has superior status in Pakistan and Bangladesh [2]. This fish industry is flourishing in Pakistan too. Chenab is the second longest river of Pakistan and a perfect home for all Chinese and Indian minor and major carps such as grass carp, Calbassu, gulfam, khagga, silver carp, mrigala, rohu and thaila. Fish faunal diversity indicates the variety of fish species within a population. High level of genetic variability can give important genetic resources to expand the genotypes [3].

Thus, loss of genetic diversity can cause different issues such as sustainability, gene fixation, resistance against diseases, loss of fitness in terms of fecundity, vigor and even can lead to extermination of local population. Hence, information of genetic variation and genetic diversity can be used for the settings of managements, preserving of wild genetic resources and genetic enhancing programs [4]. There are different strategies used to determine the level of polymorphism, PCR based markers such as simple sequence repeat (SSR), inter simple sequence repeat (ISSR), randomly amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) are most acceptable for the estimation of genetic diversity.

Microsatellite DNA markers that were first recognized in eukaryotic genome, are 2-6 base pair short stretches of DNA, are randomly present in genome of organisms and are integral tool for genetic distinguishing proof of wild and cultured population. Various sets of microsatellite DNA markers have been separated from *C. idella* that were used for genetic description in earlier studies [5-7]. The purpose of this investigation is to determine the intra and inter population genetic differentiation of *C. idella* of Indus and Chenab rivers of Pakistan by using microsatellite DNA markers.

Materials and Methods

Samples collection and extraction of DNA

The study was carried out to find the genetic diversity and polymorphism level in natural populations of *C. idella* from three specific areas. Fifteen samples of *C. idella* were collected from three areas (Khan Garh, Ghazi Ghatt, and River Chenab) of Punjab. A 2-5cm piece of muscle was extracted below the dorsal fin, and DNA was extracted from fish muscle, DNA was quantified with the help of spectrophotometer (PerkinElmer Ltd., UK) and 40ng/μl concentration was used for 20μl PCR reaction mixture.

Simple sequence repeat primers amplification

Ten pairs of SSR markers clone CID14, clone HLJC182, clone CID0131, clone CID0470, clone CID0306, clone CID0487, clone CID0044, clone CID0487, clone CID0476 and clone CID1105 [8, 9], (Table 1) were used in this investigation for the determination of genetic diversity of 15 samples of *C. idella*. Polyacrylamide gel electrophoresis (PAGE) was performed for separation of bands of PCR amplified products by using 10% polyacrylamide gel on a vertical electrophoretic apparatus (Cat#MGV-220-33 CBS scientific UK). After performing PAGE, gel was passed through silver staining, bands were observed and recorded using gel documentation system. After scoring band, excel sheet was created manually, Similarity Matrix was formed by utilizing Jaccard & Dice Similarity Coefficients. Number of effective alleles (NE), Shannon information index (I) and Polymorphism percentage were determined for genetic profile of all *C. idella* samples. Cluster investigation of all the samples was achieved with the help of Un-Weighted Pair Group Method with Arithmetic Mean (UPGMA) in SAHN program of NTSYS – PC version 2.11 J Package. The dendrogram was formulated with the help of MEGA6 software based on

Nei's pair-wise genetic differences. Analogous of heterozygosity which is known as Nei's genetic diversity (HE) was also calculated for the genetic differentiation. PIC (polymorphism information content) value

was calculated by using $PIC_i = 1 - \sum_j p_{ij}^2$ where i demonstrates marker, n and p_{ij} shows the number and frequency of alleles accordingly for that marker.

Table 1. Details of SSR primers

Sr.	Primer	Accession	Sequence 5'—3'	Ref.	Anl Temp.	Product
1	Clone CID14	FJ957942	F:5'AGGGTTTGTGGCTGCTGG-3' (18mer) R:5'CCCAGTGGATTGAAACGGAGA -3' (21mer)	[8].	60.10C	204bp
2	Clone CID0131	FJ883231	F: 5'-TCCAGGATGGCTTGAGCA-3' (18mer) R: 5'-TCACACGCAACCAGACGT-3' (18mer)	[9]	58.80C	144bp
3	Clone CID0470	FJ883332	F: 5'- TCTCCTCCGCTGGCGTAT -3' (21mer) R: 5'- TGCTCGCAGCTGATTGGT-3' (18mer)	[9]	59.80C	153BP
4	clone HLJC182	HM563726	F: 5' CCCACCCATCCACCCAAC -3' (18mer) R:5'-CCAACCGACTGTCTGCGT-3'(18mer)	[10]	59.90C	174bp
5	Clone CID0306	FJ883292	F: 5'- GGGCTGCATCGCTGAGAT -3' (18mer) R:5'-AACACCGACACCTGTGTG-3'(18mer)	[9]	58.60C	107bp
6	Clone CID0487	FJ883337	F: 5'-TGAGGTCAGAGCTGGGGAA-3' (19mer) R:5'-TTCACGACGTCGCTGCTT-3'(18mer)	[9]	59.90C	228bp
7	Clone CID0044	FJ883184	F: 5'-TTCGCCCACATGTCACCG-3' (18mer) R:5'-ACCTGTGTAAATTGTGACAAGC-3' (22mer)	[9]	58.80C	217bp
8	clone CID0487	J883341	F:5'-AGTGTTTCAGTCTAGTGCGCAT-3'(21mer) R: 5'- ACTGTCTCAAAGTGGAGCGG -3' (20mer)	[9]	59.80C	100bp
9	Clone CID0476	FJ883335	F: 5'- GCCTACCCATCATCCCACC-3' (19mer) R:5'-TCCATGACACATCAGAGTCTCT-3 (22mer)	[9]	58.90C	198
10	Clone CID1105	FJ883218	F: 5'-TGCACAGGGACACCTCATG -3' (19mer) R: 5'-GTGCTAGAGGAGGTGTGAGC -3 '(20mer)	[9]	59.70C	119bp

Results

CID14 marker is was used for amplification of fifteen samples of *C. idella* by applying polyacrylamide gel electrophoresis (PAGE) and 42 loci and four alleles were observed. All of the alleles showed 100% polymorphism without any monomorphic allele with PIC value of 0.3142. PAGE results describe that three alleles with 39 loci were present in CID0131 from which 24 are polymorphic and 15 are monomorphic with 38.76% polymorphism. The PIC value of that specific marker is 0. 4039. There were 25 loci from two alleles of CID0470 when applying PAGE. All loci were polymorphic showing 100% polymorphism with PIC value about 0.2366. HLJC18 primer results indicated four alleles and 46 loci from which 30 are monomorphic and 16 showed polymorphism with 34.78% polymorphism having the PIC

value about 0.2891. PAGE results for CID0306 primer, described three alleles with 41 loci. There are 30 monomorphic loci and remaining 11 loci showing polymorphic that were 26.82% polymorphism having PIC value of 0.3146. For CID0487 primer, 40 loci from three alleles were observed all showing 100% polymorphism with no loci showing monomorphic phenomena. That primer has PIC value about 0.3371. By applying PAGE, the CID0044 primer result shows 50 loci from 6 alleles all loci showing 100% polymorphism. There are no loci showing monomorphism with PIC value of primer as 0.3167. CID0487 marker having four alleles regarding 46 loci all loci showing 100% polymorphism with PIC value of 0.3186. PAGE results for CID0476 primer indicated 61 loci from 5 alleles out of which 15 loci were monomorphic and 46 loci were

polymorphic with 75.40% polymorphism. The PIC value of this primer is 0.2499. CID1105 primer described 34 loci from three alleles. All loci show 100% polymorphism with PIC value of 0.2624. Total 37 alleles with 422 loci were detected, 336 loci were polymorphic in all samples. Six primers i.e. clone CID14, clone CID0470, clone CID0487, clone CID0044, clone CID0487 and clone CID1105 were found genetically 100% polymorphic, clone CID0131 demonstrate 38.76% polymorphism, clone HLJC182 shows 34.78% polymorphism, clone CID0306 demonstrate 26.82% polymorphism, and clone CID0476 demonstrate 75.40% polymorphism. According to the polymorphic band estimation in all samples, 77.57% average value of polymorphism for all the primers was achieved. This value demonstrates that all markers have moderate to high level of polymorphism. Percentage for polymorphism in association with total and polymorphic loci number for each primer is given in (Table 2).

PIC value ranges from 0.2366 - 0.4039 with 0.2726 mean values that indicated the moderate allelic ratio for *C. idella* samples. The highest value of PIC was 0.4039 for the primer clone CID0131. On the other way, lowest value was 0.2366 for clone CID0470 (Table 2). PIC value 1.00 was not founded against any SSR markers for any samples of *C. idella* which demonstrates that there is no new genotype.

The genotypes based on Jaccard's similarity, ranges from 1.00 to 0.00 from most extreme to minimum value. Maximum similarity value was 0.9706 between GC3 and GC4 that demonstrates both samples of *C. idella* assembled from river Chenab. Minimum similarity value 0.3236 was observed between GG2 and GK5. In the present study, the basic range of genetic isolation among 15 samples of *C. idella* was 0.00-0.97 that

demonstrates the closed genetic relationship of the samples (Table 3 & 4).

In this study number of alleles ranges from 1 to 6. The number of alleles present for clone CID14, HLJC182, and CID0487 was four, clone CID0131, CID0306, CID0487, CID1105 exhibited three alleles, clone CID0470 demonstrated two alleles, clone CID0476 demonstrate five alleles and clone CID0044 demonstrate six alleles.

Based on genetic differences, phylogenetic tree was formulated utilizing TFPGA for the all amplifying primers. A UPGMA (unweighted pair group method with arithmetic mean) tree known as dendrogram was formulated based on the genetic similarities which can be variable for each primer and it divides all the fifteen samples into three clusters, called cluster 1, 2 and 3. Based on dendrogram, the varieties having same cluster were in close relationship and have common genetic facts. These three clusters contained six sub clusters that showed the relatives differentiation in the same sub cluster (Fig. 1). First cluster contains five samples of fish that were collected from Khan Garh area. These samples were named as GK1, GK2, GK3, GK4 and GK5. Cluster 1 is further divided into sub cluster 1 and 2. Sub cluster 1 contains GK2 and GK3 and sub cluster 2 contains GK4 and GK5. Similar value between GK2 and GK3 was 0.8236 while similar value of GK4 and GK5 was about 0.7648. These similarity values demonstrated that fish samples of sub cluster 1 for example GK2 and GK3 and fish samples of sub cluster 2 for example GK4 and GK5 were genetically similar with one another.

Second cluster also contain five samples of fish that were collected from Chenab and Ghazi ghatt areas. These samples were named as GC1, GC2, GC3, GG4 and GG5. Cluster 2 further divided in to sub cluster 3 and sub cluster 4. Sub cluster 3 contains GC1 and GC2 and sub cluster 4 contains GG4 and

GG5. Similarity value between GC1 and GC2 was 0.7942. Similarity values of GG4 and GG5 was 0.7942 indicating that fish samples of sub cluster 3 and fish samples of sub cluster 4 were genetically similar with one another. Third cluster also contain five samples of fish that were collected from Chenab and Ghazi Ghatt areas. These samples were named as GC1, GG2, GG3, GC4 and GG5. Furthermore, cluster 3 also is divided into sub cluster 5 that contain GG2

and GG3 and sub cluster 6 that contain GC4 and GG1. Similarity value between GG2 and GG3 was 0.7942. Similar values between GC1 and GG4 was 0.5295. So it is clear from the phylogenetic tree that samples collected from similar areas are genetically more similar than those collected from different areas. Thus sample of each area are making their own cluster which demonstrate that there is a clear genetic variability from area to area.

Table 2. Genetic diversity parameters of 15 *C. idella* samples with 10 SSR markers

Markers	Allele Frequency	N _E	H _E	PIC	Total loci	Polymorphic	Polymorphism percentage
CId14	0.6999	4.0000	0.3955	0.3142	42	42	100%
CID0131	0.8666	3.0000	0.5066	0.4039	39	24	38.76%
CID0470	0.8333	2.0000	0.4355	0.2366	25	25	100%
HLJC182	0.6999	4.0000	0.3644	0.2891	42	16	34.78%
CID0306	0.9111	3.0000	0.3911	0.3146	41	11	26.82%
CID0487	0.7555	3.0000	0.3969	0.3371	40	40	100%
CID0044	0.6777	6.0000	0.4029	0.3167	50	50	100%
CID0487	0.7166	4.0000	0.3977	0.3186	48	48	100%
CID0476	0.8399	5.0000	0.3011	0.2499	61	46	75.40%
CID1105	0.7647	3.0000	0.3590	0.2624	34	34	100%
Mean	0.693	3.70	0.3950	0.2726	422	336	77.57%

N_E = Number of effective alleles; H_E = Nei's genetic diversity, PIC = Polymorphism Information content

Figure 1. UPGAMA dendrogram based on Nei (1972) genetic distances between the samples of *C. idella* fish

Table 3. Frequency based genetic differences

	Population														
	GK1	GK2	GK3	GK4	GK5	GC1	GC2	GC3	GC4	GC5	GG1	GG2	GG3	GG4	GG5
GK1	0														
GK2	0.058824	0													
GK3	0.235294	0.176471	0												
GK4	0.352941	0.294118	0.294118	0											
GK5	0.294118	0.294118	0.235294	0.235294	0										
GC1	0.352941	0.294118	0.294118	0.176471	0.176471	0									
GC2	0.205882	0.205882	0.205882	0.147059	0.147059	0.205882	0								
GC3	0.176471	0.176471	0.176471	0.176471	0.176471	0.176471	0.029412	0							
GC4	0.323529	0.323529	0.264706	0.264706	0.205882	0.205882	0.176471	0.147059	0						
GC5	0.323529	0.264706	0.264706	0.205882	0.205882	0.205882	0.176471	0.205882	0.117647	0					
GG1	0.441176	0.382353	0.323529	0.558824	0.441176	0.441176	0.470588	0.441176	0.470588	0.470588	0				
GG2	0.5	0.441176	0.558824	0.676471	0.5	0.5	0.647059	0.617647	0.588235	0.588235	0.264706	0			
GG3	0.529412	0.470588	0.529412	0.647059	0.5	0.529412	0.617647	0.588235	0.558824	0.5	0.294118	0.205882	0		
GG4	0.352941	0.294118	0.411765	0.470588	0.5	0.411765	0.441176	0.411765	0.441176	0.382353	0.382353	0.411765	0.264706	0	
GG5	0.205882	0.205882	0.323529	0.323529	0.441176	0.382353	0.352941	0.323529	0.470588	0.411765	0.411765	0.470588	0.441176	0.205882	0

Table 4. Genetic similarity based on gene frequency (Similarity index)

	Population														
	GK1	GK2	GK3	GK4	GK5	GC1	GC2	GC3	GC4	GC5	GG1	GG2	GG3	GG4	GG5
GK1	0														
GK2	0.9411	0													
GK3	0.7648	0.8236	0												
GK4	0.6471	0.7059	0.7059	0											
GK5	0.7059	0.7059	0.7648	0.7648	0										
GC1	0.6471	0.7059	0.7059	0.8236	0.8236	0									
GC2	0.7942	0.7942	0.7942	0.853	0.853	0.7942	0								
GC3	0.8236	0.8236	0.8236	0.8236	0.8236	0.8236	0.9706	0							
GC4	0.6765	0.6765	0.7353	0.7353	0.7942	0.7942	0.8236	0.853	0						
GC5	0.6765	0.7353	0.7353	0.7942	0.7942	0.7942	0.8236	0.7942	0.8824	0					
GG1	0.5589	0.6177	0.6765	0.4412	0.5589	0.5589	0.5289	0.5589	0.5295	0.5295	0				
GG2	0.5	0.5589	0.4412	0.3236	0.5	0.5	0.353	0.3824	0.4118	0.4118	0.7353	0			
GG3	0.4706	0.5295	0.4706	0.353	0.5	0.4706	0.3824	0.4412	0.4412	0.5	0.7059	0.7942	0		
GG4	0.6471	0.7059	0.58	0.47083	0.5	0.5883	0.5589	0.5883	0.5589	0.6177	0.6177	0.5883	0.7353	0	
GG5	0.7942	0.7942	0.6765	0.6765	0.5589	0.6177	0.6471	0.6765	0.5295	0.5883	0.5883	0.5295	0.5589	0.7942	0

Discussion

In this investigation, ten SSR markers were used for the determination of genetic diversity among fifteen *C. idella* samples of various regions of river Chenab. All the SSR markers produce 422 clearly distinct bands. From ten markers used, six were 100% polymorphic, however average polymorphism showing by all the markers achieved as 77.57%. Shannon information index, that is diversity parameter relay on allele abundance and is joined with allele number and frequency, moderate value represents moderate genetic variations among populations of *C. idella*. PIC value estimates possible effectuality in molecular lineage, linkage map and genetic assessment. Markers showing higher PIC values, have greater chances of showing allelic variations. In our study, Clone CID0131 showed highest PIC value of 0.4039, so it can be stated that SSR markers are one of the best suited genetic markers representing the allelic fluctuations. This is supported by Danish and Singh, while working on genetic diversity of *Labeo gonius*.

It was found that the genetic variation was more obvious between the individuals of different study area. Reason of this variation may be because of fluctuations in natural states of water bodies, broad geographical barriers and restricted gene flow between populations. Another study supports these results where the genetic variations were greater between hatchery and wild stock of *C. idella* and *C. carpio* [11]. Using different genetic markers for the estimation of intra specific genetic diversity resulted in high genetic variation in Indian major carp. It manifests the change in genotype of a specific population which becomes isolated by fragmentation of habitats.

Anthropogenic activities can also result in genetic variation among the individuals of a population by blocking of fish movement and devastation of rearing grounds. The three

study areas in our study also indicated the anthropogenic effect on the population of aquatic organisms. Sahoo [12] worked with *C. idella* and reported genetic relation of parents to their offspring and within siblings representing allelic frequency and polymorphism in a given population, this clearly indicates that over the time the change in gene frequency can result in different genotypes of a population as investigated in our study. Spatial and temporal genetic heterogeneity can be lower among samples within population [13]. This investigation provided an insight into spatial distribution of grass carp in different regions of Punjab, Pakistan. It can be submitted that the wild stock of grass carp in Punjab is more genetically diverse and viable and can be used to enhance the brood stock for increased production, growth and income.

Conclusion

Samples of three selecting regions of river Chenab are genetically different from one another. But this genetic variability is not similar among all the samples. Some are genetically more similar to each other as comparing to the other ones. However, it is much fascinating that sample of the same area is showing genetically similarities to each other and is placing in a same cluster based on dendrogram. It means that genetic similarities differ from area to area. Understanding of genetic similarity and genetic differentiations is useful for choosing the breeders with desired genes. The simple sequence repeat (SSR) markers are best implements for genetic investigation, until required to enhance huge number of highly polymorphic SSR markers for the determination of *C. idella* genome. Marker determination is closely related to targeted genes. Genetic investigation of *C. idella* might be restricted because the absence of abundant informational markers, therefore it is required to identifying a best quality marker.

Authors' contributions

Conceived and designed the experiments: A Noureen, A Kanwal & M Babar, Performed the experiments: MNK Babar, Analyzed the data: T Naqqash & N Tariq, Contributed materials/ analysis/ tools: M Babar & T Naqqash, Wrote the paper: N Noureen & T Kausar.

References

1. Malik AA *et al.* (2020). Drought and plant litter chemistry alter microbial gene expression and metabolite production. *ISME J* 14(9): 2236-2247.
2. Sutton JT *et al.* (2013). Characterization of MHC class II B polymorphism in bottlenecked New Zealand saddlebacks reveals low levels of genetic diversity. *Immunogenetics* 65(8): 619-33.
3. Ouborg NJ *et al.* (2010). Conservation genetics in transition to conservation genomics. *Trends Genet* 26(4): 177-87.
4. Ashley MV *et al.* (2003). Evolutionarily enlightened management. *Biol Conserv* 111(2): 115-123.
5. Ahmed T & Abbas K (2018). Patterns of genetic variability in natural and hatchery populations of Catla catla based on microsatellite DNA markers. *Pak J Agric Sci* 55(4).
6. Haque N & Hoq T (2016). Deterioration of genetic diversity: concern in hatchery populations of Catla (Catla catla, Hamilton 1822: Cypriniformes, Cyprinidae) of Sylhet district in Bangladesh. *Inter J Curr Microbiol Appl Sci* 5(12): 171-181.
7. Sahoo L *et al.* (2019). Population structure and genetic diversity of hatchery stocks as revealed by combined mtDNA fragment sequences in Indian major carp, Catla catla. *Mitochondrial DNA A DNA Mapp Seq Anal* 30(2): 289-295.
8. Guo H *et al.* (2007). Genetic diversity of Nelumbo accessions revealed by RAPD. *Genet Resour Crop Evol* 54(4): 741-748.
9. Xia X & Xie Z (2001). DAMBE: software package for data analysis in molecular biology and evolution. *J Hered* 92(4): 371-3.
10. Liu F *et al.* (2009). High genetic diversity and substantial population differentiation in grass carp (*Ctenopharyngodon idella*) revealed by microsatellite analysis. *Aquacu* 297(4): 51-56
11. Sahoo L *et al.* (2014). Limited genetic differentiation in Labeo rohita (Hamilton 1822) populations as revealed by microsatellite markers. *Biochem Syst Ecol* 57: 427-431.
12. Vieira C *et al.* (2019). Diversity and origin of the genus Lobophora in the Mediterranean Sea including the description of two new species. *Phycol* 58(2): 163-168.
13. Spandana B *et al.* (2012). Development and characterization of microsatellite markers (SSR) in Sesamum (Sesamum indicum L.) species. *Appl Biochem Biotechnol* 168(6): 1594-607.