

Genetic diversity and population structure inferred from the partially duplicated genome of domesticated carp, *Cyprinus carpio* L.

Lior DAVID^{a,d*}, Noah A. ROSENBERG^b, Uri LAVI^c,
Marcus W. FELDMAN^d, Jossi HILLEL^e

^a Department of Animal Sciences, Faculty of Agricultural, Food and Environmental Quality Sciences, The Hebrew University of Jerusalem, P.O. Box 12, Rehovot, 76100, Israel

^b Department of Human Genetics, Bioinformatics Program, and the Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109-2218, USA

^c Institute of Horticulture, ARO-Volcani Center, P.O. Box 6, Bet-Dagan, 50250, Israel

^d Department of Biological Sciences, Stanford University, Stanford, CA 94305, USA

^e The Robert H. Smith Institute of Plant Sciences & Genetics, Faculty of Agricultural, Food and Environmental Quality Sciences, The Hebrew University of Jerusalem, P.O. Box 12, Rehovot, 76100, Israel

(Received 10 May 2006; accepted 8 December 2006)

Abstract – Genetic relationships among eight populations of domesticated carp (*Cyprinus carpio* L.), a species with a partially duplicated genome, were studied using 12 microsatellites and 505 AFLP bands. The populations included three aquacultured carp strains and five ornamental carp (koi) variants. Grass carp (*Ctenopharyngodon idella*) was used as an outgroup. AFLP-based gene diversity varied from 5% (grass carp) to 32% (koi) and reflected the reasonably well understood histories and breeding practices of the populations. A large fraction of the molecular variance was due to differences between aquacultured and ornamental carps. Further analyses based on microsatellite data, including cluster analysis and neighbor-joining trees, supported the genetic distinctiveness of aquacultured and ornamental carps, despite the recent divergence of the two groups. In contrast to what was observed for AFLP-based diversity, the frequency of heterozygotes based on microsatellites was comparable among all populations. This discrepancy can potentially be explained by duplication of some loci in *Cyprinus carpio* L., and a model that shows how duplication can increase heterozygosity estimates for microsatellites but not for AFLP loci is discussed. Our analyses in carp can help in understanding the consequences of genotyping duplicated loci and in interpreting discrepancies between dominant and co-dominant markers in species with recent genome duplication.

genetic structure / *C. carpio* / AFLP / genome duplication

* Corresponding author: davidl@agri.huji.ac.il

1. INTRODUCTION

The common carp (*Cyprinus carpio* L.) is an important freshwater food fish, produced in temperate climatic regions around the world. The wild ancestor of domesticated carp probably lived in the Caspian and Aral Sea basins, from where it was dispersed both to Western Europe and to East Asia [1]. Archaeological findings and documentation of fish consumption suggest that the Romans probably domesticated European races, early in the 1st century.

In Asia, at roughly the same time as domestication in Europe, wild carp were stocked and grown in reservoirs. More recently, in the early 19th century, the Japanese ornamental carp (koi), a colored variant of the common carp, was developed in Japan, as documented by breeders in the Niigata Prefecture [1]. Several koi variants have been bred specifically for different color patterns, leading to considerable phenotypic variability within the ornamental carp group.

Despite a long-standing interest in carp genetics, only a few studies have used genetic markers to evaluate genetic relationships among carp populations. Genetic variability has been estimated in a variety of feral and aquacultured populations using allozymes, microsatellites, AFLP and PCR-RFLP markers for mtDNA [2, 3, 5, 11, 13, 14, 34]. Differentiation between the European/Central Asian carps and the East Asian carps has been established, and smaller differences have also been observed between populations within each of these geographic clades.

Previous cytological [23] and molecular [15] studies have suggested that the common carp recently experienced whole-genome duplication. Indeed, using microsatellite markers, we found that paralogous fragments can frequently be amplified by the same primer-pair, and we estimated that about 60% of the carp genome remains duplicated [4]. Segregation patterns of PCR amplicons in families allowed us to distinguish between alleles and paralogs, and suggested disomic inheritance of alleles. Molecular variation between paralogs was higher than between alleles of the same ortholog, and in several cases, paralogous loci fixed with different microsatellite alleles were observed [4]. Though evidence for whole-genome duplications has been seen for many species (including several vertebrates), the common carp provides a rare opportunity to study a relatively recent event. Accumulating data suggest that copy number variations and variability within and between duplicated regions have important functional consequences, including in human diseases (*e.g.* [30]). However, outside of humans, relatively little experimental data exist that address problems associated with genotyping of multiple copies and that help in

interpreting such genotypes on the population level in situations for which no knowledge of relatedness among individuals is available.

Here, for eight carp populations and a grass carp outgroup, using amplified fragment length polymorphism (AFLP) patterns and microsatellite loci that we have developed [3], we study levels of genetic variability and genetic relationships among groups. We then discuss how different marker types perform in assaying these two aspects of genetic variation in the context of the recent whole-genome duplication of the common carp and the breeding histories of the populations.

2. MATERIALS AND METHODS

2.1. Populations

Individuals were randomly sampled from each of three aquacultured and five ornamental carp populations. The aquacultured carp strains used were Yugoslavian, Dor-70 and their commercial hybrid. The two parental strains originated in Europe and were bred to better fit Israeli carp production practices [18]. Ornamental carp (koi) varieties included white (Ohgon), black (Karasu) and three color-patterned varieties (Kohaku, Sanke and Showa). Figure 1 shows pictures of representative fish. All of the aquacultured and ornamental populations are from one species (*Cyprinus carpio* L.). An albino mutant of the grass carp (*Ctenopharyngodon idella*) was chosen as an outgroup population, because it is one of the few species that can hybridize with common carp (producing infertile offspring), and because microsatellite markers in common carp have previously been shown to have a high rate of cross-amplification in grass carp [3]. Grass carp is considered a diploid species that likely did not experience the genome duplication believed to have occurred recently in common carp. Blood samples were obtained from the Gan-Shmuel Fish Breeding Center (Gan-Shmuel, Israel), and five individuals were sampled per population. This is a small sample size, but similarly to a study of chickens (Fig. 5 of [26]), our study finds that while small, it can be sufficient for inference about some aspects of population structure. Where possible, we restricted our analysis to methods that are still informative when relatively few individuals are sampled.

2.2. Genotyping of microsatellite and AFLP markers

Twelve microsatellite markers (*Koi3-4*, *Koi13-14*, *Koi17-18*, *Koi21-22*, *Koi29-30*, *Koi51-52*, *Koi55-56*, *Koi57-58*, *Koi69-70*, *Koi85-86*, *Koi89-90* and

Koi111-112) that were found to be highly polymorphic in a previous study [3], and that amplify DNA fragments from the grass carp, were genotyped. At the time of genotyping, no information on duplicated loci was available. Seven combinations of AFLP selective primers (AAC/CTA, ACA/CAT, ACA/CTG, ACG/CAA, ACG/CTC, ACG/CTT, AGC/CAG) were used to generate 505 visually scored AFLP bands. Each combination is named after the three selective nucleotides used for *Eco RI/Mse I* adapters, respectively. Most of the AFLP bands that were scored were present both in grass carp and in common carp. Since the two are different species, we trusted bands that were present in both species more than bands that were absent in one of them. However, since the grass carp is a very monomorphic population, it is possible that bands that are strong and reliable in the common carp would be absent from all individuals of grass carp. We included several such reliable bands in our AFLP results. For microsatellites, fluorescently labeled primers were used to PCR amplify the loci, and fragments were resolved on an ABI Prism 377 DNA sequencer to determine the sizes of the alleles. For AFLP loci, a radioactively labeled *Eco RI* primer was used in the selective PCR amplification step, and fragments were resolved on a 5% polyacrylamide sequencing gel. For further details on genotyping procedures of both marker types see David *et al.* [3].

2.3. Microsatellite data irregularities

Owing to technical problems or null alleles, genotypes of 24 individual-locus combinations (4.4%) were not obtained. Showa koi individual no. 45 had genotypes at only three of the 12 loci, and was the only individual with missing data at more than two loci. No genotypes were available at locus *Koi13-14* in any of the Showa koi individuals, due to either technical failure or to the presence of a null allele.

Among the 540 individual-locus combinations, three distinct alleles were observed in 37 combinations obtained from seven loci, and four alleles were observed in six such combinations obtained from three of these loci. In David *et al.* [4] we analyzed segregation of alleles in families and found these seven loci (*Koi3-4*, *Koi17-18*, *Koi29-30*, *Koi55-56*, *Koi69-70*, *Koi89-90* and *Koi111-112*) to be duplicated. Since the present study is based not on families but on random samples from populations, allelic relationships at these duplicated loci could not be determined. Thus, adjustments were performed in order to determine genetic diversity or population structure (see details in specific sections below).

2.4. Diversity indices

AFLP loci were scored for the presence or absence of a band. Consequently, gene diversity measures the proportion of loci at which two individuals differ in a score. For AFLP patterns, the unbiased estimator of gene diversity [21] was calculated using *Arlequin* [29]. For microsatellite data, observed heterozygosity of a population was the proportion of individuals with two or more alleles, averaged across loci. Similarly, for each population the number of alleles was averaged across loci. In a population study, if duplicated loci are amplified, it is impossible to say which amplicons are alleles and which are paralogs; thus, the actual heterozygosity is difficult to determine. Therefore, in addition to the proportion of heterozygotes, we also used only the duplicated loci to calculate an estimate of heterozygosity based on all possible combinations of genotypes compatible with the number of distinct alleles that were observed. Given a disomic inheritance pattern, an individual with two amplicons in a duplicated locus could be a homozygote if different alleles were in different duplicates. Alternatively, if both alleles exist in only one duplicate then the individual would have a heterozygosity of 0.5, and if both alleles were present in both duplicates then its heterozygosity would equal 1. In cases of two amplicons in a duplicated locus, seven combinations of genotypes are possible in two unlinked loci: (A/A, A/B), (A/A, B/B), (A/B, A/A), (A/B, A/B), (A/B, B/B), (B/B, A/A) and (B/B, A/B), where A and B are the two observed alleles, and (A/A, A/B) refers to a situation in which the first duplicate has an AA genotype and the second has an AB genotype. An estimate of the “expected heterozygosity of duplicated loci” can be obtained as the product of the probability and the heterozygosity of a combination of genotypes, summed over possible combinations. We assumed equal probability for each of these combinations to be the actual genotype of an individual. Therefore, the heterozygosity of an individual at this locus was estimated to be 0.43. In case of three amplicons, 12 possible genotypes exist – (A/A, B/C), (A/B, A/C), (A/B, B/C), (A/B, C/C), (A/C, A/B), (A/C, B/B), (A/C, B/C), (B/B, A/C), (B/C, A/A), (B/C, A/B), (B/C, A/C), and (C/C, A/B) – and the expected heterozygosity is 0.75. For four amplicons, both loci must be heterozygous and the expected heterozygosity is 1. The estimates for individuals were averaged across duplicated loci and across individuals to estimate heterozygosities of populations. We also calculated the possible boundaries of these expected heterozygosities by assigning either the lowest or the highest value to individuals, considering all possible allelic combinations. Finally, we calculated the proportion of heterozygotes based only on the five markers that amplify single (that is, nonduplicated) loci.

2.5. Molecular variance analysis

To partition the genetic variance, analysis of molecular variance (AMOVA) was carried out on both AFLP and microsatellite data using *Arlequin* [29]. The AMOVA for microsatellites was performed under the assumption that all loci are non-duplicated. In the few individuals for which more than two alleles were detected, we eliminated the alleles with the lowest frequencies in the population of this individual, assuming that their contribution to the genetic profile of the population was minor. Analysis was performed with two models: (1) One-way, in which individuals were classified into one of the nine carp populations; and (2) Nested, in which populations were nested within carp types (aquacultured, ornamental, or grass carp). Each of the two models was applied twice, once including and once omitting the outgroup (grass carp). The *F*-statistics produced in the analysis were tested for difference from zero using 10 000 permutations of individual labels.

2.6. Genetic structure of populations

An iterative model-based clustering algorithm implemented in the *structure* program was used with the microsatellite data to find optimal subdivisions of the data into genetic clusters [25]. Without using prior information on the populations of origin of individuals, the algorithm places individuals into *K* genetic clusters in such a way that each individual has a membership coefficient for each cluster, corresponding to the proportion of its genome estimated to have ancestry in the cluster. Many replicate runs of the *structure* correlated allele frequency model were performed using 30 000 iterations following a burn-in period of 100 000. Similar results were obtained in all runs, and a representative example was chosen for display in Figure 1.

For the 37 individual-locus pairs in which more than two alleles were found, two of the three or four observed alleles were chosen at random for the cluster analysis. The analysis was repeated using 100 random data sets constructed in this manner, and 98 of these were found to produce essentially the same result, in that a measure of similarity for *structure* runs [27] exceeded 0.96 in all comparisons of pairs taken from among the 98 data sets. The remaining two data sets produced noticeably different results for only two of the nine populations.

2.7. Population differentiation test

Neighbor-joining trees were constructed for all pairs of populations separately for the microsatellite and AFLP data. The shapes of pairwise trees were assessed as follows: if the individuals from two populations could be partitioned into two clades, each of which contained all the individuals from one of the populations [20], the tree was labelled “consistent with population affiliation”. If the partition was made by cutting the tree along its longest internal branch, the tree was labeled “strongly consistent”; if the cut was not necessarily along this branch, the tree was labeled “weakly consistent” [26]. For the microsatellite data, we used the same representative dataset that was used to generate the results shown for the *structure* clustering. For microsatellites, genetic distances between each pair of individuals were computed as one minus the proportion of shared alleles [17], and for the AFLP data, the simple matching distance was used [33]. In each pairwise calculation based on microsatellites, only loci for which neither individual had missing data were used. Neighbor-joining trees [28] were obtained using the *neighbor* program in the *phylip* package [8]. Showa koi individual 45 was not included in the microsatellite-based pairwise population trees.

2.8. Clustering cladograms for individuals

Trees of all 45 individuals were constructed, separately for the microsatellite and AFLP data, as described above for pairwise population trees. Because of missing data, Showa koi individual 45 was omitted from the microsatellite tree. The consensus neighbor-joining tree [28] based on 1000 bootstrap distance matrices across loci [7] was obtained using the *consense* program in the *phylip* package [8]. To assess the clustering results, we identified the three branches that had the highest bootstrap support, among those that subtended at least five individuals. The same three groupings (with similar bootstrap support) were identified in all 100 data sets formed by random elimination of alleles from individuals with three or four distinct alleles; the plot in Figure 2b is based on the same representative random elimination data set as was used for Figure 1 and Table III.

3. RESULTS

3.1. Genetic diversity of carp populations

First we estimated the genetic diversity based on the two types of DNA markers. The average gene diversity across loci, based on AFLP data, varied

Table I. Genetic diversity indices of carp populations.

Population	Type	Gene diversity ^a	Observed frequency of heterozygotes ^b	Number of alleles per locus	Heterozygosity in duplicated and single loci ^c	Range of expected heterozygosity in duplicated loci ^d
		AFLP	Microsats.	Microsats.	Microsats.	Microsats.
Grass carp	Grass carp	0.05	0.51	2.08	0.56 ^e (0.44)	N.A.
Yugoslavian carp	Aquacultured	0.21	0.55	3.25	0.33 (0.38)	0.08–0.67
Dor-70 carp	Aquacultured	0.26	0.57	4.00	0.35 (0.45)	0.11–0.66
Commercial hybrid carp	Aquacultured	0.26	0.63	3.42	0.34 (0.52)	0.06–0.71
White koi	Ornamental	0.13	0.58	4.08	0.34 (0.48)	0.09–0.66
Black koi	Ornamental	0.12	0.62	3.50	0.40 (0.40)	0.11–0.77
Kohaku koi	Ornamental	0.30	0.60	3.25	0.37 (0.31)	0.04–0.80
Sanke koi	Ornamental	0.32	0.47	2.83	0.27 (0.35)	0.06–0.56
Showa koi	Ornamental	0.32	0.43	2.33	0.29 (0.25)	0.06–0.59
Average across populations		0.22	0.55	3.19	0.33 (0.40)	0.07–0.66
Coefficient of variation (%)		44.7	12.4	21.3	23.3 (21.6)	33.6–12.1

Values are based on genotypes of five individuals per population at 12 microsatellite loci and 505 AFLP bands. For each population, diversity statistics were averaged across loci.

^a Gene diversity for AFLP data is the probability that two individuals have different states for a given band.

^b Observed proportion of individuals with two or more amplicons per microsatellite marker, averaged across loci.

^c Expected heterozygosity based on the possible combinations of microsatellite amplicons, averaged across seven duplicated loci (see Materials and Methods). In parentheses, observed proportion of heterozygotes averaged across five single-copy microsatellites.

^d Range of expected heterozygosity based on possible combinations of microsatellite amplicons in seven duplicated loci (see Materials and Methods).

^e Proportion of heterozygotes in the fully diploid grass carp based on the seven loci that are duplicated in the common carp.

considerably among populations, with an average of 0.22 and a coefficient of variation (CV) of 44.7% (Tab. I). The grass carp exhibited very little polymorphism (0.05), white and black koi exhibited low polymorphism (0.13 and 0.12, respectively), and the color-patterned koi populations varied between 0.30 and 0.32. The aquacultured carp breeds showed intermediate values of diversity (0.21–0.26). The observed proportion of heterozygotes, based on all microsatellite loci, was relatively similar across all populations (average of

0.55 and CV of 12.4%). The number of microsatellite alleles was more variable across populations than was this observed heterozygosity, with an average of 3.19 alleles per locus per population and a CV of 21.3%. Grass carp had the lowest number of alleles (2.08 per locus). The expected heterozygosity, based on all possible allelic combinations at each of the seven duplicated loci, ranged from 0.27 in Sanke koi to 0.40 in black koi. The seven loci that were duplicated in the common carp produced a heterozygosity of 0.56 in the (fully diploid) grass carp. The observed heterozygosity for the five single (*i.e.* nonduplicated) loci ranged between 0.25 in Showa koi and 0.52 in commercial carp. The coefficient of variation associated with heterozygosity estimates was similar whether estimated from the five single loci from the seven duplicated ones. The range of expected heterozygosity based on duplicated loci was large for all populations, demonstrating the difficulty in estimating the actual values. Heterozygosities, whether estimated from duplicates or observed in single loci, were lower than the observed heterozygosity based on all loci. Variable levels of diversity were found with AFLP markers, while regardless of the estimate type and the set of markers used, microsatellites produced diversity levels that were more similar among populations.

3.2. Analysis of molecular variance

The partitioning of molecular variation was similar for both marker types in the one-way design. Exclusion of the grass carp reduced the among-populations component, producing F_{ST} estimates of 20% and 26% for AFLP and microsatellites, respectively (Tab. II). In the nested design without the grass carp, variance partitioning differed for microsatellites and AFLP markers. For AFLP and microsatellite markers, respectively, 8% and 26% of the molecular variance resided between the two carp types, indicating that larger differences were detected by microsatellites. In this nested design (excluding grass carp), 77.5% and 64.8% of the variation were within populations for AFLP and microsatellites, respectively. Thus, analysis of molecular variation identified a substantial amount of variation between aquacultured and ornamental carps.

3.3. Genetic structure of carp populations

To better understand the genetic structure of the populations, we applied two methods that do not make use of population information; *structure* and

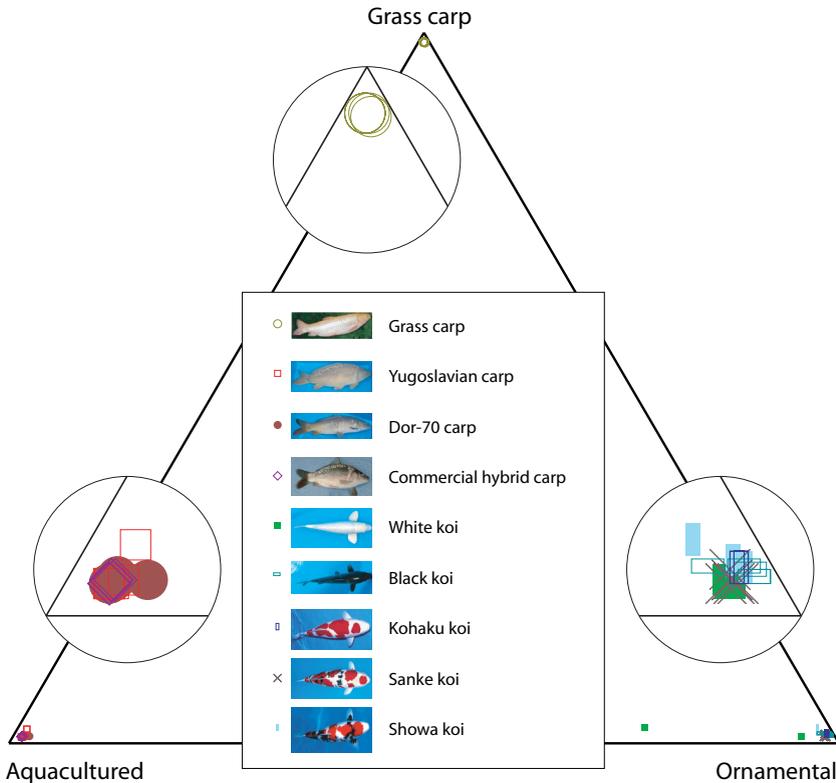


Figure 1. Clustering of individuals based on microsatellite data using *structure*. Each of the vertices represents a cluster and is magnified in a circle to clarify its composition. Individuals are shown in different colors and shapes, denoting their populations of origin. For each individual, the fraction of the genome corresponding to a given cluster is represented by the distance to the triangle side opposite to the vertex associated with the cluster. The sum of the distances from any point to all three sides equals one. Although pairwise genetic distances between clusters are not equal, clusters are graphically represented as equidistant from each other.

tree-based clustering. The most likely partition inferred by *structure* analysis of the microsatellite data had $K = 3$ genetic clusters, which corresponded closely to the three carp types: grass carp, aquacultured carp and ornamental carp (Fig. 1). Individuals from the same population had similar membership coefficients in the three clusters, and nearly all individuals had membership coefficients greater than 0.96 for the cluster to which they were most closely affiliated; in 39 of 45 cases this membership coefficient was greater than 0.99. The only individual that did not have a very high membership coefficient for the cluster that contained other members of its population was white koi individual

Table II. Analysis of molecular variance (AMOVA).

Design ^a	Source of variation	Percent of variation			<i>F</i> -statistic ^b		
		Statistic	AFLP	Microsats.	Statistic	AFLP	Microsats.
One-way (grass carp included)	Among populations	V_a	36.6	31.5	F_{ST}	0.366***	0.315***
	Within populations	V_b	63.4	68.5			
One-way (grass carp excluded)	Among populations	V_a	19.7	26.2	F_{ST}	0.197***	0.262***
	Within populations	V_b	80.3	73.8			
Nested (grass carp included)	Among carp types	$V_{a'}$	30.5	62.8	F_{CT}	0.305**	0.628***
	Among populations within carp types	$V_{b'}$	13.0	2.6	F_{SC}	0.187***	0.070***
	Within populations	V_c	56.4	34.6	$F_{ST'}$	0.440***	0.650***
Nested (grass carp excluded)	Among carp types	$V_{a'}$	7.6	26.3	F_{CT}	0.076	0.263*
	Among populations within carp types	$V_{b'}$	14.9	8.9	F_{SC}	0.162***	0.120***
	Within populations	V_c	77.5	64.8	$F_{ST'}$	0.225***	0.352***

^a In the one-way design, all populations are treated as one group while in the nested design, groups are defined as carp types (grass carp, aquacultured and ornamental).

^b $F_{ST} = V_a/V_{Total}$, $F_{CT} = V_{a'}/V_{Total}$, $F_{SC} = V_{b'}/(V_{b'} + V_c)$, $F_{ST'} = (V_{a'} + V_{b'})/V_{Total}$. For each design, V_{Total} is the sum of all variance components.

* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$.

25 (77% for the ornamental carp cluster). This robust analysis divided individuals into three major clusters with perfect concordance to carp types.

The tree of individuals based on AFLP markers showed a three-branched structure (Fig. 2a). The three branches of the tree that had the highest bootstrap support separated: (1) grass carp, (2) black and white koi and (3) Yugoslavian carp, together with some Dor-70 and some commercial hybrid individuals. Bootstrap support for these groupings was extremely high (96%–100%). The center of the tree contained mostly Kohaku, Sanke and Showa koi individuals as well as individuals from the commercial hybrid strain (Fig. 2a).

The three main branches of the microsatellite-based tree separated the three carp types perfectly (Fig. 2b), although with less bootstrap support than the three main branches of the AFLP-based tree. The aquacultured carp individuals sub-clustered considerably but not perfectly into the three populations representing this group. Among ornamental carp, black and white koi individuals were clustered separately from Kohaku, Sanke and Showa koi individuals, which were mixed together. In both trees, white koi individual 25 (the slightly ambiguous individual in the *structure* analysis) was placed with other white koi individuals. In accordance with the partition of molecular variance, microsatellites separated the carp types in both *structure* and tree-based clustering while the AFLP-based tree separated groups with less genetic variability (grass carp, black and white koi).

3.4. Pairwise differentiation test

Pairwise neighbor-joining trees based on AFLP showed that individuals from grass carp, black koi and white koi always clustered consistently with their population affiliation (Tab. III). Dor-70 carp individuals could not be completely separated from Yugoslavian carp or from commercial hybrid carp. Kohaku, Sanke and Showa koi could not be completely separated from Dor-70 and commercial hybrid carps, or from each other. Thus, AFLP separated most easily the populations with low levels of diversity. In general, trees based on microsatellites strongly separated pairs of populations from different carp types. Koi populations were clustered either weakly-consistently with population affiliation or not consistently at all (Tab. III). Such weakly consistent clustering was also found between aquacultured carps. Among the 13 differences in clustering results between AFLP and microsatellites, namely between the *ij* and *ji* elements of Table III, in only three cases did AFLP markers cluster individuals “more consistently” with their population affiliation than did microsatellites.

4. DISCUSSION

4.1. Diversity of carp populations

The history of each population is expected to be reflected in its genetic variability, and despite the small sample size, gene diversity values (based on 505 AFLP bands) agreed with our predictions based on the different histories of these populations and on previous results [3]. Grass carp represents a different

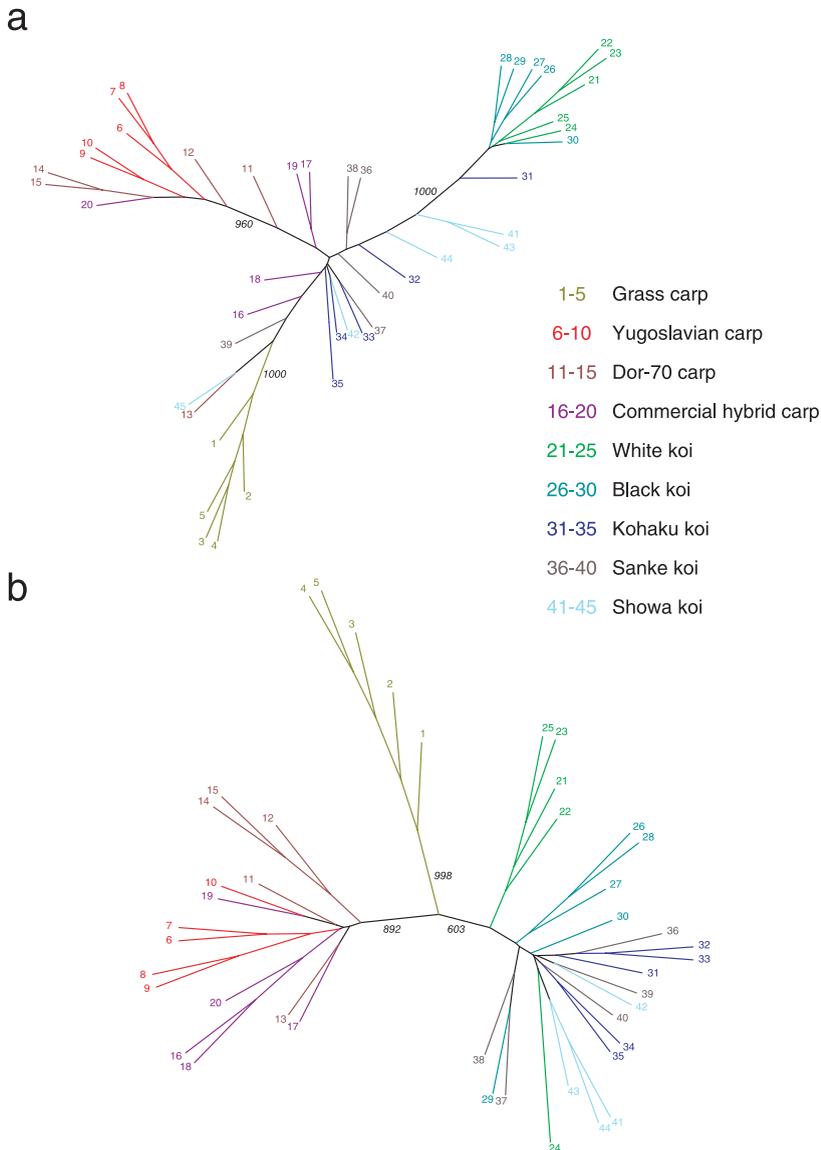


Figure 2. Neighbor-joining trees of individuals. **(a)** Tree based on AFLP markers, using the simple matching distance. **(b)** Tree based on 12 microsatellite markers, using one-minus-the-proportion-of-shared-alleles distance. Both trees represent a consensus of 1000 bootstraps across loci, constructed using the neighbor-joining clustering method. Branches and individual labels are color-coded by population. Branches that do not correspond to any particular population are colored black. For each tree, we show the three branches with the highest bootstrap values among branches that separate the tree into two components of at least five individuals each.

Table III. Population differentiation: shapes of neighbor-joining trees based on individuals from pairs of populations.

	1	2	3	4	5	6	7	8	9
1. Grass carp		S	S	S	S	S	S	S	S
2. Yugoslavian carp	S		<u>W</u>	W	S	S	S	S	<u>S</u>
3. Dor-70 carp	S	N		<u>W</u>	S	S	<u>S</u>	<u>S</u>	<u>S</u>
4. Commercial hybrid carp	S	W	N		S	S	<u>S</u>	<u>S</u>	<u>S</u>
5. White koi	S	S	S	S		N	W	S	W
6. Black koi	S	S	S	S	<u>W</u>		N	N	W
7. Kohaku koi	S	S	N	N	W	<u>W</u>		N	N
8. Sanke koi	S	S	N	N	S	<u>S</u>	N		<u>W</u>
9. Showa koi	S	W	N	N	W	W	N	N	

Below the diagonal: trees based on the simple-matching distance using AFLP data. Above the diagonal: trees based on one-minus-the-proportion-of-shared-alleles distance using microsatellite data.

S: tree strongly consistent with population affiliation; W: tree weakly consistent with population affiliation; N: tree not consistent with population affiliation.

There are 13 cases where AFLP and microsatellites differ. In these cases the shape that is “more consistent” with population affiliation is underlined and italicized.

Showa koi individual 45 was omitted from the analysis with microsatellites.

species from the other carp populations in this study. Its broodstock suffered heavy mortality several years ago, and consequently, the living fish descend from a very small number of founders. This bottleneck predicts a very low level of polymorphism, as was observed. The close clustering of grass carp individuals probably results from this low level of polymorphism and from the relative genetic divergence of grass carp from *Cyprinus carpio* L.

The Yugoslavian and the Dor-70 strains are kept as parental lines and are crossed by breeders to produce the commercial hybrid. The Yugoslavian line descends from the inbred *Našice* strain from Yugoslavia. The Dor-70 line was developed between 1965 and 1970 by five generations of individual selection for growth performance, beginning with a crossbred line between European strains present in Israel [18]. The Yugoslavian strain was found to be less polymorphic than Dor-70, as expected from its history as an inbred strain. These aquacultured carps have an intermediate level of gene diversity (0.21–0.26 based on AFLP). This relatively low level of polymorphism is likely to be the outcome of a recent history in which parental lines were kept as separate populations, maintained by inbreeding since the 1970s.

Monochromatic koi (such as black or white) are bred within strains to produce similar progeny, from which the most desirable individuals are selected

for future broodstock. The very high fecundity of carp enables broodstocks to be maintained as small populations. This small population size, together with strong positive assortative mating, may explain the low genetic variability within black and white koi variants, as well as their relative distinctiveness among the five koi populations.

Kohaku, Sanke and Showa are usually crossed within and among variants. Due to high demand for these varieties, larger populations are maintained as broodstocks. Relatively large population sizes and outbreeding among these variants are reflected in higher levels of polymorphism and in the genetic similarity among these three breeds (Fig. 2 and Tab. III).

The patterns of polymorphism observed in this study demonstrate the major impact that domestication and breeding practices have had on the genetic diversity and structure of these nine carp populations. Intra-population diversity as measured by AFLP can be well explained by the histories and breeding practices of the different populations, suggesting that diversity was reliably measured on a scale that is typical for AFLP. Because broodstock can be held as small populations, breeders of highly fecund species such as *C. carpio* should take special care in order to avoid the loss of genetic diversity. Loss of genetic diversity in the short term can cause the performance of the broodstock to deteriorate due to increased inbreeding depression, and in the long term can prevent the generation of new genotypes with greater commercial value [18, 32].

4.2. Genetic structure

In a one-way AMOVA-model that included *Cyprinus carpio* only, the among-population component of genetic variance was 26% based on microsatellites and 20% for AFLP, indicating a genetic difference between the two carp types. In the nested design, much less variation (9% and 15%) was found among populations within carp types, suggesting that differentiating populations within each type is considerably more difficult.

As expected, the differentiation test with both marker types confirmed that grass carp is genetically different from each of the other populations. Pairwise-population trees and the tree of all individuals, based on microsatellite data, supported the distinctiveness of aquacultured and ornamental carps. These analyses differentiated weakly among the aquacultured strains and between the monochromatic and the color-patterned koi. Both *structure* and neighbor-joining analyses used multilocus genotypes and generated distinct genetic clusters for the two carp types, without utilizing the population labels of individuals. The difference between European/Central Asian and East Asian carps

is consistent with this distinction [11, 14], suggesting that ornamental carps originated from the latter. Crosses between German carp and koi have been documented [1], but we did not find significant admixture with European carps among the koi that we tested. The distinctive signatures are clearer in analyses based on microsatellites, perhaps due to the relatively high mutation rate of these loci that allows discrimination even between recently diverged populations. A similar result was obtained by Estoup *et al.* [6], who showed that microsatellites, in contrast to allozymes, could detect differences among recently diverged populations of brown trout (*Salmo trutta*).

The measures we took to test the bias introduced by duplicated loci (*i.e.* re-analyzing the data after randomly eliminating “surplus alleles” and analyses with only single loci) suggest that the genetic structure analysis was not substantially biased by inclusion of duplicated loci. Additionally, although levels of polymorphism are somewhat different between the five nonduplicated loci and the seven duplicated loci, *structure* and neighbor-joining analyses based only on the five nonduplicated loci yielded very similar results to those presented (not shown).

4.3. Polymorphism of microsatellites and duplicated loci

In general, observed population levels of microsatellite heterozygosity were higher than the corresponding estimates based on AFLP markers. However, unlike AFLP and what was expected from population history, microsatellite-based diversity estimates were mostly similar across all populations. Discrepancies between AFLP and microsatellite results have also been seen in studies of plants, where microsatellites revealed subtler differences among populations than did AFLP markers [12, 16]. Furthermore, differences in diversity estimates, both in magnitude and in the order of estimates across populations, were observed between the marker types.

The higher level of microsatellite-based diversity can result from the presence of several alleles in a population as well as from the higher rate of mutation typical of microsatellites. The microsatellites for this study were selected because they were highly polymorphic in aquacultured and ornamental carp populations [3], which might lead to higher and more similar levels of polymorphism across populations or to lower values in the grass carp. However, we hypothesize that the relatively uniform heterozygosity of microsatellites can result in part from the genotyping of duplicated loci. It is believed that *Cyprinus carpio* is an ancestral allotetraploid that has since become diploid in parts of its genome [4, 15, 23]. Among the 12 markers we used, seven were

found duplicated [4]. We propose that in an extreme case, fixation of alleles at some of the duplicated loci would increase the estimate of heterozygosity. If fixation of different alleles in each of the paralogs occurs, this differential fixation arrangement will be inherited across generations and will be interpreted through a co-dominant genotyping process as a single heterozygous locus (Fig. 3). In such cases, fixation, instead of causing a decrease in heterozygosity, will maintain it – but only when co-dominant markers are used. The genotypes we scored suggest that the model of differential fixation applies in some of the population-locus combinations, as in various cases, only one type of heterozygote was found, at a frequency higher than expected by free segregation of alleles. We have also observed such differential fixation scenarios at a frequency of 28% in analysis of carp families [4]. Therefore, observing one type of heterozygote in a proportion significantly larger than expected based on free segregation of alleles may be a good indicator of differential fixation, suggesting that a duplicated locus has been studied and that heterozygosity has been overestimated.

Under differential fixation of duplicated loci, populations with lower polymorphism levels, such as the inbred white koi (gene diversity of 0.13 using AFLP), can still be assessed as polymorphic by microsatellites (observed heterozygosity of 0.58 and 4.08 alleles per locus). The same underlying heterozygosity assessed by the dominant AFLP markers would appear as monomorphism for the presence of a band, since allelic relations cannot be identified in this marker system (Fig. 3). Thus, in such low-polymorphism populations, AFLP markers would reflect more accurately the genotypes at these loci by showing monomorphism that would result in lower estimates of polymorphism levels.

To better define the heterozygosity levels of populations, we estimated the expected heterozygosity for duplicated loci separately from the observed heterozygosity for single loci alone. Both subsets of markers showed values lower than the “proportion of heterozygotes” based on all loci, and estimates based on single loci were as variable as those of duplicated loci. Nevertheless, none of the estimates based on microsatellites correlated well with the AFLP-based estimates and with the history of populations. This result reflects the difficulty in estimating the polymorphism levels of populations in partially polyploid species, when duplicated loci are genotyped. This difficulty is also evidenced by the wide range of possible heterozygosities expected from combinations of alleles at duplicated loci (Tab. I).

In addition to difficulties in measuring population-genetic parameters, genotyping of duplicated loci might obscure the real genotypes in studies that

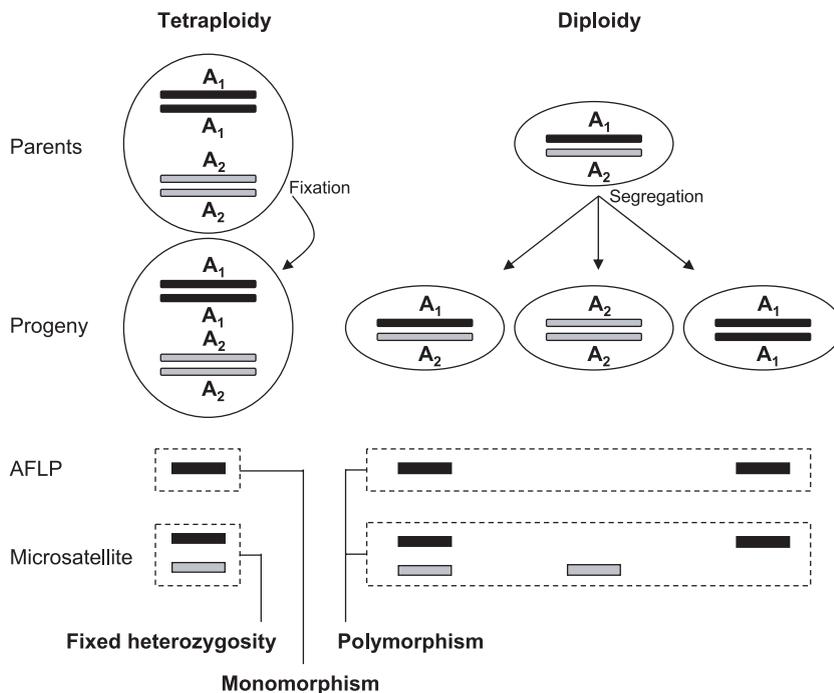


Figure 3. Banding pattern of AFLP and microsatellite markers at duplicated and single loci. Crossing two parents with the same apparent genotype (A_1A_2) results in the following: (1) Fixed heterozygosity in progeny at a duplicated locus, with fixation of different alleles. (2) Free segregation of alleles in progeny at a diploid locus. Below are the genotypes of the corresponding progeny, as detected by the dominant AFLP marker (assuming dominance of A_1) and the co-dominant microsatellite marker.

construct genetic maps and look for phenotype-genotype associations. For the duplicated genome of the rainbow trout, a genetic map was constructed using androgenetic doubled haploid progeny [22, 35]. However, such methods are not applicable for most other species with partially duplicated genomes, and are not useful for assessment of natural variation. Given the considerable fraction of duplicated regions in many genomes and the growing body of evidence on effects of duplicated regions on phenotypes such as human diseases [30,31], the consequences of genotyping duplicated loci need to be considered more carefully [9]. Because numerous species, including the common carp, have large portions of their genomes still duplicated, genotyping only unique genomic regions might miss the biological properties that are encoded and affected by duplications, as suggested by Gahr *et al.* [10] for the *ID* gene family in the rainbow trout (*Oncorhynchus mykiss*).

Finally, a population with differential fixation at duplicated microsatellite loci can appear more heterozygous than it really is. Though true heterozygosity is difficult to estimate, this partial duplication might have important evolutionary implications. One of the evolutionary advantages suggested for polyploidy is that it enables the maintenance of a higher level of heterozygosity [24]. In recently duplicated regions, paralogous genes may keep their sequence similarity and thus may have some overlap in function [19]. In such cases, an individual can have four different functional alleles, and there are many more possible combinations at the population level. We previously showed that comparison of sequence from flanking regions of microsatellite repeats allows detection of paralogous loci [4] and thus assists in the interpretation of the genotypes. This study suggests how genotyping duplicates with co-dominant markers increases estimates of genetic diversity even in populations that are expected to have low polymorphism levels, and thus makes it difficult to evaluate the actual heterozygosity. However, genotyping duplicates with co-dominant markers did not have a major effect on estimation of the genetic structure of populations, even for recently diverged populations. These observations together with the possibility of observing differential fixation in paralogous loci, make *Cyprinus carpio* L. an interesting model to study the consequences of such genetic arrangements in noncoding and coding regions.

ACKNOWLEDGEMENTS

We would like to specifically thank Dr. Shmuel Rothbard for his assistance in obtaining the fish samples and for helpful discussions on the history of the populations. We also thank the team of Gan-Shmuel Fish Breeding Center for making fish samples available. This research was supported in part by NIH grant GM28428, by a Burroughs Wellcome Fund Career Award in the Biomedical Sciences (to NAR), and by Research Fellow Award No. FI-329-2002 from BARD, the United States - Israel Binational Agricultural Research and Development Fund (to LD). The experimental work was carried out in compliance with the laws for animal and genetic research in Israel.

REFERENCES

- [1] Balon E.K., Origin and domestication of the wild carp, *Cyprinus carpio*: from Roman gourmets to the swimming flowers, *Aquaculture* 129 (1995) 3–48.

- [2] Crooijmans R.P.M.A., Bierbooms V.A.F., Komen J., Van der Poel J.J., Groenen M.A.M., Microsatellite markers in common carp (*Cyprinus carpio* L.), *Anim. Genet.* 28 (1997) 129–134.
- [3] David L., Rajasekaran P., Fang J., Hillel J., Lavi U., Polymorphism in ornamental and common carp strains (*Cyprinus carpio* L.) as revealed by AFLP analysis and a new set of microsatellite markers, *Mol. Genet. Genomics* 266 (2001) 353–362.
- [4] David L., Blum S., Feldman M.W., Lavi U., Hillel J., Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci, *Mol. Biol. Evol.* 20 (2003) 1425–1434.
- [5] Desvignes J.F., Laroche J., Durand J.D., Bouvet Y., Genetic variability in reared stocks of common carp (*Cyprinus carpio* L.) based on allozymes and microsatellites, *Aquaculture* 194 (2001) 291–301.
- [6] Estoup A., Rousset F., Michalakis Y., Cornuet J.M., Adriamanga M., Guyomard R., Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*), *Mol. Ecol.* 7 (1998) 339–353.
- [7] Felsenstein J., Confidence limits on phylogenies: an approach using the bootstrap, *Evolution* 39 (1985) 783–791.
- [8] Felsenstein J., *PHYLIP (Phylogeny Inference Package)*, Version 3.6, Department of Genetics, University of Washington, Seattle (1993) <http://evolution.genetics.washington.edu/phylip.html>
- [9] Fredman D., White S.J., Potter S., Eichler E.E., Den Dunnen J.T., Brookes A.J., Complex SNP-related sequence variation in segmental genome duplications, *Nat. Genet.* 36 (2004) 861–866.
- [10] Gahr S.A., Rodriguez M.F., Rexroad C.E. III, Identification and expression profile of the ID gene family in the rainbow trout (*Oncorhynchus mykiss*), *Biochim. Biophys. Acta.* 1729 (2005) 64–73.
- [11] Gross R., Kohlmann K., Kersten P., PCR-RFLP analysis of the mitochondrial *ND-3/4* and *ND-5/6* gene polymorphisms in the European and East Asian subspecies of common carp (*Cyprinus carpio* L.), *Aquaculture* 204 (2002) 507–516.
- [12] Jakse J., Kindlhofer K., Javornik B., Assessment of genetic variation and differentiation of hop genotypes by microsatellite and AFLP markers, *Genome* 44 (2001) 773–782.
- [13] Kohlmann K., Kersten P., Genetic variability of German and foreign common carp (*Cyprinus carpio* L.) populations, *Aquaculture* 173 (1999) 435–445.
- [14] Kohlmann K., Gross R., Murakaeva A., Kersten P., Genetic variability and structure of common carp (*Cyprinus carpio*) populations throughout the distribution range inferred from allozyme, microsatellite and mitochondrial DNA markers, *Aquat. Living Resour.* 16 (2003) 421–431.
- [15] Larhammar D., Risinger C., Molecular genetic aspects of tetraploidy in the common carp *Cyprinus carpio*, *Mol. Phylogenet. Evol.* 3 (1994) 59–68.
- [16] Mariette S., Chagne D., Lezier C., Pastuszka P., Raffin A., Plomion C., Kremer A., Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers, *Heredity* 86 (2001) 469–479.

- [17] Minch E., Ruiz-Linares A., Goldstein D.B., Feldman M.W., Cavalli-Sforza L.L., *Microsat*: a program for calculating statistics on microsatellite data, version 2 alpha, Department of Genetics, Stanford University, Stanford, CA (1998) <http://hpgl.stanford.edu/projects/microsat/>
- [18] Moav R., Hulata G., Wohlfarth G., Genetic differences between the Chinese and European races of the common carp. I. Analysis of genotype-environment interactions for growth rate, *Heredity* 34 (1975) 323–340.
- [19] Moore R.C., Grant S.R., Purugganan M.D., Molecular population genetics of redundant floral-regulatory genes in *Arabidopsis thaliana*, *Mol. Biol. Evol.* 22 (2005) 91–103.
- [20] Mountain J.L., Cavalli-Sforza L.L., Multilocus genotypes, a tree of individuals, and human evolutionary history, *Am. J. Hum. Genet.* 61 (1997) 705–718.
- [21] Nei M., *Molecular evolutionary genetics*, Columbia University Press, New York, 1987.
- [22] Nichols K.M., Young W.P., Danzmann R.G., Robison B.D., Rexroad C., Noakes M., Phillips R.B., Bentzen P., Spies I., Knudsen K., Allendorf F.W., Cunningham B.M., Brunelli J., Zhang H., Ristow S., Drew R., Brown K.H., Wheeler P.A., Thorgaard G.H., A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*), *Anim. Genet.* 34 (2003) 102–115.
- [23] Ohno S., *Evolution by gene duplication*, Allen and Unwin, London, 1970.
- [24] Otto S.P., Whitton J., Polyploid incidence and evolution, *Annu. Rev. Genet.* 34 (2000) 401–437.
- [25] Pritchard J.K., Stephens M., Donnelly P., Inference of population structure using multilocus genotype data, *Genetics* 155 (2000) 945–959.
- [26] Rosenberg N.A., Burke T., Elo K., Feldman M.W., Freidlin P.J., Groenen M.A.M., Hillel J., Mäki-Tanila A., Tixier-Boichard M., Vignal A., Wimmers K., Weigend S., Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds, *Genetics* 159 (2001) 699–713.
- [27] Rosenberg N.A., Pritchard J.K., Weber J.L., Cann H.M., Kidd K.K., Zhivotovsky L.A., Feldman M.W., Genetic structure of human populations, *Science* 298 (2002) 2381–2385.
- [28] Saitou N., Nei M., The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.* 4 (1987) 406–425.
- [29] Schneider S., Roessli D., Excoffier L., *Arlequin*. A software for population genetics data analysis, Version 2.000, Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva (2000) <http://lgb.unige.ch/arlequin/>
- [30] Sebat J., Lakshmi B., Troge J., Alexander J., Young J., Lundin P., Maner S., Massa H., Walker M., Chi M., Navin N., Lucito R., Healy J., Hicks J., Ye K., Reiner A., Gilliam T.C., Trask B., Patterson N., Zetterberg A., Wigler M., Large-scale copy number polymorphism in the human genome, *Science* 305 (2004) 525–528.
- [31] Sharp A.J., Locke D.P., McGrath S.D., Cheng Z., Bailey J.A., Vallente R.U., Pertz L.M., Clark R.A., Schwartz S., Segraves R., Oseroff V.V., Albertson D.G., Pinkel D., Eichler E.E., Segmental duplications and copy-number variation in the human genome, *Am. J. Hum. Genet.* 77 (2005) 78–88.

- [32] Shikano T., Chiyokubo T., Taniguchi N., Temporal changes in allele frequency, genetic variation and inbreeding depression in small populations of the guppy, *Poecilia reticulata*, *Heredity* 86 (2001) 153–160.
- [33] Sokal R.R., Michener C.D., A statistical method for evaluating systematic relationships, University of Kansas, *Scientific Bulletin* 38 (1958) 1409–1438.
- [34] Tanck M.W.T., Baars H.C.A., Kohlmann K., Van der Poel J.J., Komen J., Genetic characterization of wild Dutch common carp (*Cyprinus carpio* L.), *Aquac. Res.* 31 (2000) 779–783.
- [35] Young W.P., Wheeler P.A., Coryell V.H., Keim P., Thorgaard G.H., A detailed linkage map of rainbow trout produced using doubled haploids, *Genetics* 148 (1998) 839–850.