J. Anim. Breed. Genet. 117 (2000), 105–120
© 2000 Blackwell Wissenschafts-Verlag, Berlin
ISSN 0931–2668

Ms. received: 13.4.1999

Laboratorio de Genética Molecular, Departamento de Producción Animal, Facultad de Veterinaria, Madrid, Spain

Genetic structure in Atlantic brown trout (*Salmo trutta* L.) populations in the Iberian peninsula: evidence from mitochondrial and nuclear DNA analysis

By S. DUNNER, L. ROYO and J. CAÑON

Summary

Brown trout (*Salmo trutta* L.) was sampled in rivers belonging to three different Spanish basins in order to analyse the distribution of genetic variability. The genetic analysis was performed by using two systems and techniques: nuclear DNA was screened through random amplified polymorphic DNAs (screening 2×10^5 bp of the whole genome), and mitochondrial DNA (mtDNA) through sequencing of the hypervariable control region. Genetic distances between the populations were similar using either analysis although some differences arise. For example, some populations of the Tajo basin were very close through nuclear analysis but more distant using mtDNA. Differences between the two DNA sources could be the result of a different evolutionary rate, and the fact that mtDNA is maternally transmitted and differences in sex migration rates will influence the patterns of genetic variation between the transmitted DNAs. Total variation was partitioned using AMOVA showing a clear subdivision among basins although intrapopulation variation remained as high as 62%. A correspondence analysis defined the differences in a three-dimensional way, clustering the populations according to their common basin. When mtDNA was sequenced, higher variability was noted in the segment between 400 and 600 bp of the whole D-loop sequence, suggesting that these 200 bp improved the analysis of the variability more than sequencing the t-RNA ends of the control region. A comparison was made between the t-RNA^{Pro} ends of the 10 populations screened here and the rest of the published sequences found in the literature, leading to a concentration of these populations in group IV which includes all trouts which originate in the Atlantic. The analyses performed suggest that a high genetic variability is present in all populations and that although there has been a probable interference from stocked strains introduced to increase population density, this was only detectable through the variance between rivers which reflect different policies according to the region where the basin is located. However, the genetic analysis using the two approaches allows the control of the natural populations avoiding a loss of their genetic potential.

Zusammenfassung

*Genetische Struktur atlantischer Populationen iberischer Bachforellen (*Salmo trutta *L.): Nachweis mittels mitochondrialer und Kern-DNA-Analysen*

In drei verschiedenen spanischen Wassersystemen wurden Bachforellen (*Salmo trutta* L.) gesammelt, um die genetische Variabilität zu analysieren. Zur Analyse wurden zwei Systeme und zwei Techniken genutzt: Kern-DNA wurde mittels RAPD analysiert (2.10^5bp des Gesamtgenoms), mitochondriale DNA durch das Sequenzieren der hypervariablen Kontrollregion. Beide Techniken ergaben ähnliche genetische Distanzen zwischen den Populationen, obwohl Unterschiede bei Populationen aus dem Tajo-Gebiet auftraten, deren Kern-DNA eine sehr geringe Distanz aufwiesen, während die über Sequenzierung von mtDNA ermittelte Distanz auf größere Unterschiede hinwies. Diese Differenz könnte das Ergebnis unterschiedlicher Mutationsraten während der Evolution sein, oder auf die Tatsache zurückzuführen sein, daß mtDNA über die Mutterlinien weitergegeben wird, und Migrationsunterschiede zwischen den Geschlechtern die Muster der genetischen Variation der weitergegebenen DNA-Stücke beeinflußt. Die Gesamtvarianz wurde mit Hilfe von AMOVA analysiert, wobei zwischen den Wassersystemen eine eindeutige Unterteilung festgestellt werden konnte, obwohl die Varianz innerhalb der Population bei 62% liegt. Eine Korrespondenzanalyse erklärt die Distanzen dreidimensional, wodurch die Populationen gleicher Herkunft in Cluster zusammenfallen. Bei der Sequenzierung von mtDNA wurde eine größere Variabilität im Sequenzbereich zwischen 400 und 600 bp der gesamten

D-Loop Sequenz festgestellt, was darauf schließen läßt, daß diese 200 bp für eine Analyse der Variabilität geeigneter sind als eine Sequenzierung der t-RNA Enden der Kontrollregion. Ein Vergleich der t-RNA^{Pro} Enden der zehn hier untersuchten Populationen und der bereits publizierten Sequenzen führen zu einer Konzentration dieser in der Gruppe IV, zu der alle Forellenpopulationen zählen, die aus dem Atlantik stammen. Die Untersuchung deutet auf eine hohe genetische Variabilität in allen Populationen hin und auf Interferenzen durch Aufstockungsmaßnahmen zur Erhöhung der Populationsdichten hin. Dies ist nur durch die Varianz zwischen den Flußsystemen aufzuklären, die von Region zu Region unterschiedliche Maßnahmen wiederspiegelt. Die genetische Analyse mittels dieser zwei Ansätze erlaubt eine Kontrolle der natürlichen Populationen zwecks Vermeidung eines Verlustes ihres genetischen Potentials.

Introduction

Brown trout (*Salmo trutta* L.) is very highly prized by fishermen. Over-fishing, which has been allowed over the years, has led to a decrease in the numbers of trout in the rivers, and a consequent policy of introducing foreign stocks to increase the size of the available populations.

Native brown trout is difficult to rear in captivity, as it is a slow growing fish that takes a long time to reach the maturity. The introduction of genetically selected stocks from Denmark or Germany has impoverished the autochthonous populations by homogenizing the genetic variability and leading to a presumed loss of individual fitness.

It is important to know the extent to which the different policies have interfered with the natural population dynamics. This is a difficult task as autochthonous trout very likely produce founder effects (strains issuing from a small number of reproducers) and there is usually a high interpopulation diversity as a result of the fragmentation in small populations of the different hydrographic basins (PRESA et al. 1994). Evolutionary processes may be inferred from data on the genetic structure of populations. Such inferences require two phases, first describing patterns of genetic variation, followed by the estimation of parameters quantifying the effects of mutation, selection, inbreeding, drift or gene flow. Gene flow is a powerful homogenizing force that can act to prevent genetic divergence arising among populations by either drift or selection (SLATKIN 1995).

The use of different sources of DNA polymorphism is of concern in any population genetics study: mitochondrial DNA (mtDNA) can provide unique insights into population history (AVISE et al. 1987; AVISE 1994), and can suggest hypotheses about the boundaries of genetically divergent groups. However, mtDNA must be used in conjunction with nuclear markers to identify evolutionarily distinct populations for conservation (CRONIN et al. 1993) because given a lower effective number of genes (BIRKY et al. 1989) or greater dispersal by males than females, mtDNA can diverge whereas nuclear genes do not (MORITZ 1994). Some effort has been devoted to the analysis of both nuclear and mtDNA for the purpose of identifying genetic variation in natural populations. In the special case of *Salmo trutta* L., studies have focused on nuclear DNA using allozymes (PRESA et al. 1994; GIUFFRA et al. 1994; BERNATCHEZ and OSINOV 1995; GARCÍA-MARIN and PLÁ 1996; APOSTOLIDIS et al. 1996a), minisatellites (PRODÖHL et al. 1994), microsatellites (ESTOUP et al. 1993, 1998) or on mitochondrial DNA including the D-Loop region (BERNATCHEZ et al. 1992; GIUFFRA et al. 1994; BERNATCHEZ and OSINOV 1995; APOSTOLIDIS et al. 1996b, 1997).

Although in recent years different authors have described several microsatellite markers in Salmonids (see, e.g. ESTOUP et al. 1993, 1998), the trout nuclear genome has, up to now, only been represented by these sparse markers and by the protein-coding sequences which provide a small insight into the extent of the nuclear DNA (BERNATCHEZ and OSINOV 1995). The use of random amplified polymorphic DNA (RAPD) increases the number of loci surveyed and thus the portion of the genome sampled. Although they are dominant markers, this shows the advantage of the technique when compared with restriction fragment length polymorphism (RFLP) or allozyme studies (KARAKOUSIS and TRIANTAPHYLLIDIS 1990; GARCÍA-MARÍN et al. 1991; GARCÍA-MARÍN and PLA 1996). RAPD analysis is a

polymerase chain reaction (PCR)-based assay capable of analysing genetic relatedness among and within species (WILLIAMS et al. 1990). By using short, single, random primers, this technique is capable of scanning a genome for the many inverted priming sites that are close enough to be amplified.

In recent years, mtDNA has been used to define the matriarchal phylogenies of different animal species. The methods used include restriction analysis of the whole mitochondrial genome, and PCR amplification of determined fragments (cytochrome *b*, ND regions and D-Loop; see for example KOCHER et al. 1989; BERNATCHEZ et al. 1992).

Therefore, the objective of this paper was to study the genetic variation distribution of brown trout populations of different basins located in the Iberian peninsula by using mitochondrial and nuclear DNA through two techniques: sequencing of the control region and generating markers with random primers.

Materials and methods

Samples collection and DNA extraction

Genomic DNA was extracted from either liver or adipose fin, using a standard phenol-chloroform method (SAMBROOK et al. 1989) on samples taken from nine populations of nine rivers belonging to three different Atlantic basins within the Iberian peninsula. The individuals were captured either by electric or by traditional fishing. From these samples, a total of 155 native individuals as well as 20 from a German selected strain, were analysed for RAPD analysis (Fig. 1). To facilitate the selection of markers, two DNA pools per population were set up by mixing equal volumes of DNA from five randomly chosen individuals. After selection of appropriate markers, every individual sampled was screened for each RAPD marker chosen and these data were used in the analysis performed later. For mtDNA sequencing, five randomly chosen individuals were used from each of the populations.

RAPD amplifications

A total of 100 arbitrary primers (10-Mers Set 100/8 RAPD Oligo Project-N°701–800 purchased at British Colombia University) were used for amplifications of the DNA pools in different 25 μl PCR reaction volumes containing 10 pmol of each primer, 0.2 mM dNTP mix, 0.5 unit *Taq* polymerase (Stoffel Fragment; PE Biosystems, Warrington, UK), 20 ng template DNA, and 3.5 mM $MgCl_2$. Amplifications were performed in a programmable thermal controller (PTC-100; MJ Research, Inc., Watertown, MA, USA) with 5 min denaturation, followed by 45 cycles of 1 min at 93°C, annealing at 36°C for 1 min, and extension at 72°C for 2 min. An additional 5 min for extension at 72°C followed the last cycle. The PCR products were electrophoresed in 2% agarose gels (NuSieve 3 : 1; FMC, DK-2665 Vallensback) stained with ethidium bromide. A 100-pb leader was used as a molecular size marker. RAPD fingerprints were visualized under UV light and photographed for recording purposes. Each pool was processed in triplicate to ensure reproducibility, and after marker selection, the same conditions were used for individual amplifications.

D-Loop amplifications

Total DNA (200 ng) of each individual was used as template in 50 *m*l PCR reactions containing 20 mM Tris-HCl pH 8.55; 160 mM $(NH_4)_2SO_4$; 150 μg/ml bovine serum albumin; 0.2 mM dNTP mix; 2.5 mM $MgCl_2$; 1 unit *Taq* polymerase (Epicentre Technologies, Madison, WI, USA) and 50 pmol of each primer. The PCR reactions were carried out in an MJResearch device (PTC-16). The primers used were: (1) Loophe (5′ TTA TGC TTT AGT TAA GCT ACG CCA ACT AG 3′) (nt 328–300 of GenBank Accession *n*° M97974,

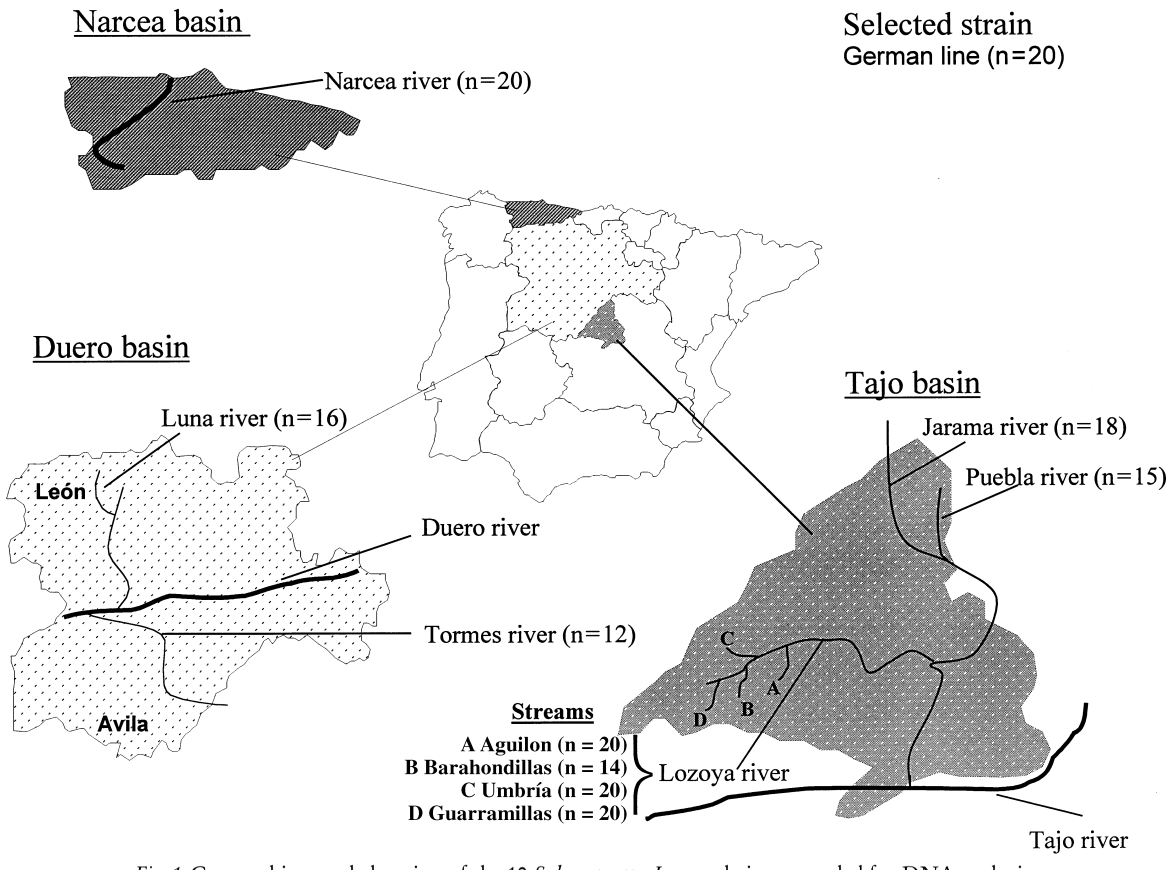


Fig. 1. Geographic sample location of the 10 *Salmo trutta L.* populations sampled for DNA analysis

BERNATCHEZ et al. 1992); (2) Loopend (5′ AAG ATA CCC CGG CTT CTG CGC GGT AAC CCC 3′) (nt 17–47 of GenBank Accession no. M97974, BERNATCHEZ et al. 1992) were initially designed to amplify the 3′ end of the D-Loop; (3) L16007 (KOCHER et al. 1989) in conjunction with Loophe was used to amplify the whole control region ($\approx$1.1 kb fragment). Fragments were cloned in the PGEM-T Vector System (Promega), introduced in bacterial strain *Escherichia coli* sure (Promega Innogenetics, Madrid, Spain) and plated in SOB (Bactotryptone-yeast extract-NaCl) + agar in presence of IPTG (Isopropylthio-β-D-gallactosidase) and X-Gal (5-bromo-4-chloro-3-indolul-β-D-galactoside). Positive clones were sequenced using an ALF DNA sequencer (Pharmacia Biotech, Madrid, Spain). Two additional primers Dloopfor (5′ ATC TAG TTC GAC CTT GTT AGA C 3′) (nt 416–436 GenBank no. STU 62283) and Dlooprev (5′ CGT CTT TAC CCA CCA ACT TTC 3′) (nt 728–708 GenBank no. STU 62283) were used as internal primers to allow the automatic sequencing of the larger fragment. Each of the individual cloning steps were performed in triplicate in order to eliminate artefactual variation.

Statistical methods applied to RAPD's

A measure of genetic similarity using the Nei and Li coefficient (NEI and LI 1979) was computed after scoring '1' for each band present and '0' for each band not present, and a similarity matrix between individuals was then constructed:

$$Coef_{\text{Nei–Li}} = 2n_{xy}/n_x + n_y$$

where $2n_{xy}$ is the number of markers shared by the two animals, n_x and n_y are the numbers of markers observed in individual x and y, respectively. Two methods were used to represent genetic relationship: (1) the unweighted pair-group method with arithmetic mean (UPGMA) algorithm performed by the SAHN (Sequential, Agglomerative, Hierarchical and Nested) program in the PC version of NTSYS (ROHLF 1993) was applied to cluster either, individuals and the 10 populations, according to the genetic distances ($1 - Coef_{\text{Nei–Li}}$); and (2) a correspondence analysis, as a procedure for condensing multivariate data (CAVALLI-SFORZA et al. 1994), where differences between populations using RAPD markers are represented in a three-dimensional figure, each axis explaining the variability by means of the percentage represented by a set of RAPD markers. In addition, a matrix of linear geographic distances was constructed, and the relationship between this and the matrix of similarity was examined with a test of time-space clustering (MANTEL 1967). The resulting r-values were interpreted as correlation coefficients.

The individual data were also used to estimate variance components for RAPD genotypes through application of the AMOVA program (EXCOFFIER et al. 1992). A matrix of genetic distances between individuals is generated using a slightly modified coefficient of NEI and LI (1979):

$$D = 100\,[1 - Coef_{\text{Nei–Li}}]$$

These slight modifications were made to enable use of the AMOVA program, which applies the Euclidean metric of Excoffier (EXCOFFIER et al. 1992) and leads to a series of F-statistics analogues, labelled Φ-statistics by EXCOFFIER et al. (1992). Φ_{ST} and Φ_{SC} are measures of the degree of resemblance between individuals of a river. This resemblance can be regarded as a difference between individuals in different rivers and expressed as the between-river differences, as a proportion of the total variance (Φ_{ST}) or a proportion of the basin variance (Φ_{SC}). A similar situation applies to the parameter Φ_{CT} which is a measure of the degree of resemblance between individuals of a basin expressed as a proportion of the total variance.

Statistical analysis of mtDNA control region sequences

For the analysis of mtDNA sequence data, we first used the GCG package (DEVERAUX et al. 1984) to align and edit the sequences. To estimate genetic distances, a Tamura model

(TAMURA 1992) was adjusted, which makes it possible to take account of the different nucleotide frequencies, as well as the different base substitutions (transitions and transversions). The Kimura two-parameter model (KIMURA 1980) was considered for comparison of the domestic strains and other populations described elsewhere.

As a result, dendrograms were constructed using the UPGMA and the Neighbour-joining algorithms (SAITOU and NEI 1987) with the genetic distances estimate under the Tamura and the Kimura models, respectively. The statistical confidence of each node was estimated by 1000 random bootstrap resamplings of the data using the branch and bound option (FELSENSTEIN 1985). Computing was carried out by MEGA package (KUMAR et al. 1993).

Results

RAPD analysis

Seven of the one hundred primers tested showed reproducible polymorphic bands under optimized reaction conditions among triplicates that could be scored reliably on population gels. These chosen seven primers (BC706; BC709; BC719; BC721; BC722; BC729; BC775) were used next to amplify patterns in all the individuals that composed the population pools. Finally 13 polymorphic RAPD markers bands were found with these seven primers and used to construct UPGMA trees from pairwise distances between the 10 populations (Fig. 2a) and from the pairwise distances between 69 individuals (representing approximately 40% of the total number of individuals) taken at random for each population described in Figure 1 (Fig. 3). The numbers at the nodes in Figure 2 are bootstrapping values for 1000 bootstrap resampling. Of the 64 native individuals in Figure 3 only eight (12.5%) do not cluster with trout from the same basin. Only the stocked population showed a fragmented pattern of clustering with individuals split into the three basins.

Goodness of fit of the clustering to data set was acceptable ($r = 0.82$) reflecting the clear cluster within basins, except for the selected strain which showed a fragmented pattern of clustering.

Populations were spatially clustered using correspondence analysis and the first three components, which accounted for 83.7% of the underlying variation (the first component 49.7%, the second 22.2% and the third 11.8%), plotted in Figure 4. Examination of this figure reveals a clear separation of the three basins. Tajo populations being explained by first inertia, whereas factor 3 is the most important for discrimination of the Duero basin populations.

Components of variance make use of the individual data and permit an explanation of the level of differences among individuals of different hydrographic basins, among individuals of different river populations within basins, and among individuals within rivers populations (Table 1). Of the total genetic variation, 38.5% (Φ_{ST}) was due to differences among rivers. From the rest of the genetic variability, the largest proportion (approximately 31%) occurred among basins. Unbiased estimates of Excoffier's θ_{ST} showed considerable levels of population differentiation. When genetic distances between populations were plotted against geographical distances a significant positive correlation (Mantel test; $p < 0.01$) was observed ($r = 0.72$).

Mitochondrial DNA variation

Internal primers had to be used to sequence the complete control region of the mtDNA. The sequences are deposited in GenBank with the following accession numbers: STU62283–87.

The entire control region sequence was determined for 50 individuals, leading to 1030 bp length as a mean, with Puebla haplotype being the longest (1054 bp) and Lozoya A (Aguilón stream) and Lozoya D (Guarramillas stream) the shortest (1014 bp). This region is A-T rich

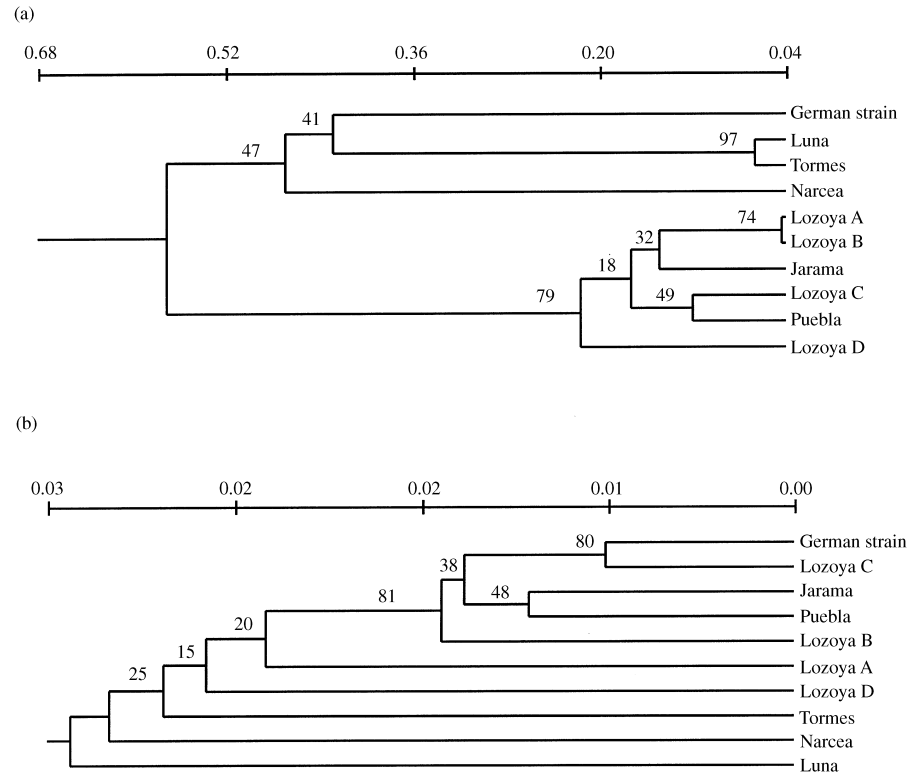


Fig. 2. UPGMA cluster analysis for 10 brown trout populations based on: (a) RAPD markers information and NEI and LI (1979) genetic distances, and (b) sequence variation of the D-Loop and genetic distances computed under the TAMURA (1992) model. The numbers indicated correspond to the percentages of grouping confidence calculated by bootstrap analysis

(64%). Divergence between sequences belonging to individuals of the same population showed an intrapopulation variation that was especially high in the Jarama population when compared with the German strain (data not shown). A total of 172 positions of the 1030 mean bp (approximately 17%) in the sequences were variable involving mutational events, 20 involving the 3′ end and 31 the 5′ end. The remainder (121 positions) were concentrated in the central part of the control region mostly between nt 450 and nt 600 (Table 2).

Alignment of the D-Loop sequences corresponding to a single population was performed, leading to a consensus sequence for each population, and pairwise consensus sequence divergence estimates were computed and used to construct an UPGMA tree relating the 10 populations studied (Fig. 2b). These consensus sequences were also used to compare their $tRNA^{Pro}$ ends to the AT1 sequence described in BERNATCHEZ et al. (1992) for an Atlantic genotype (GenBank Accession N° M97969), and divergences are shown in Table 3. Only four variable nucleotide positions were found for the 21 positions defined by BERNATCHEZ et al. (1992). The 5′ end of the control region was more variable than the 3′ end in comparison to the conclusions of BERNATCHEZ et al. (1992). A Neighbour-joining tree was constructed by using all available sequences for the 310 end fragment (Fig. 5).

Fig. 3. UPGMA dendrogram constructed from the pairwise distances among 69 individuals from 10 populations and three basins

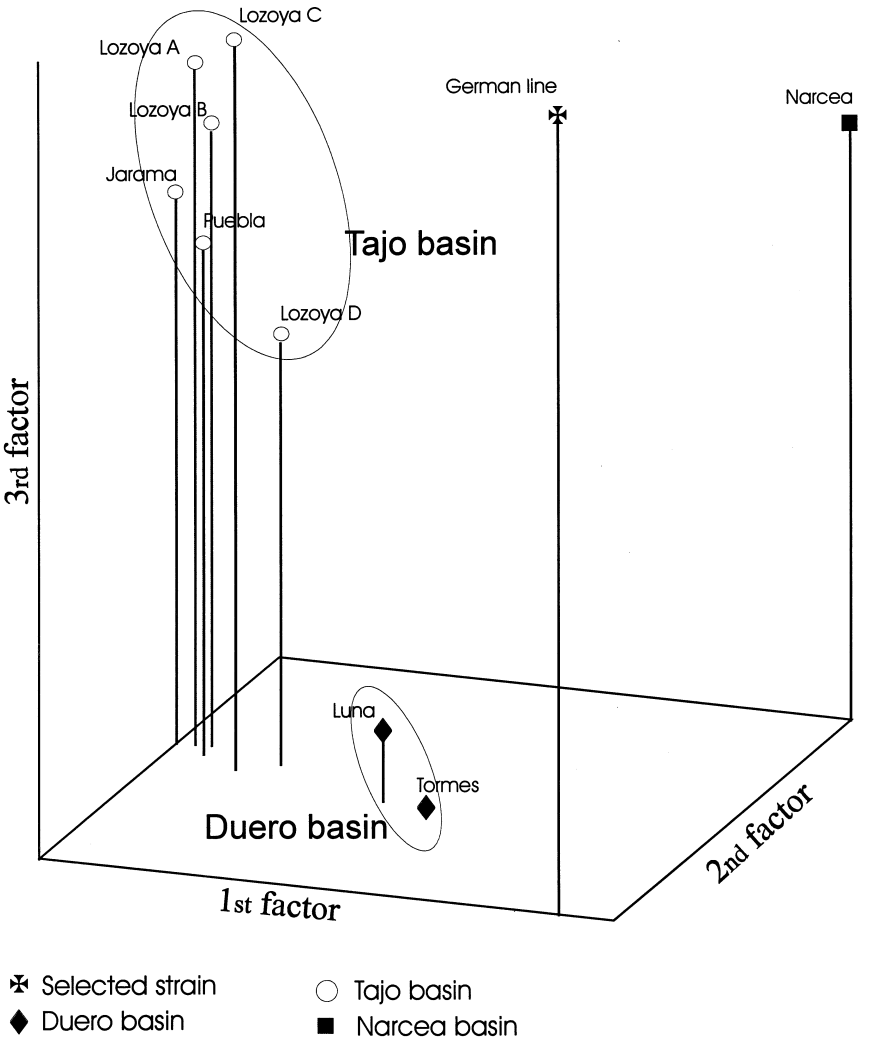
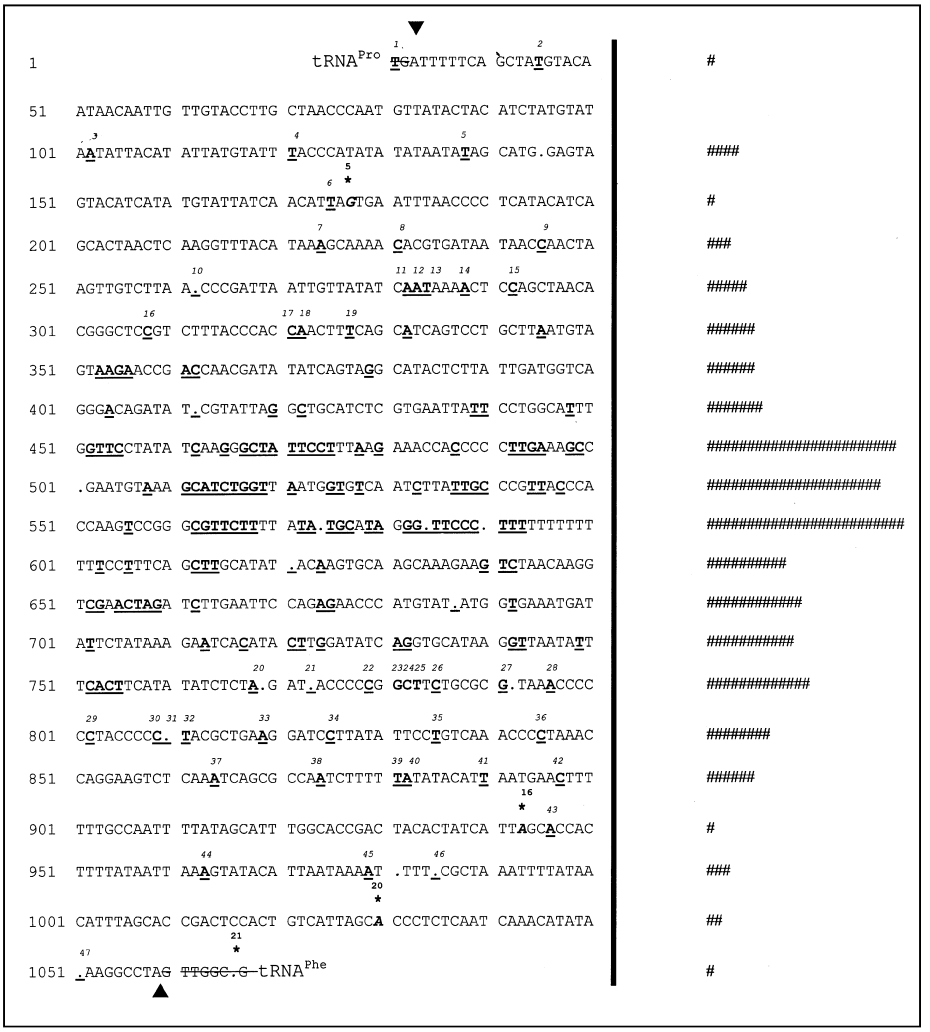


Fig. 4. Correspondence Analysis clustering the 10 populations in a three-dimensional way where each axis explains the variability by means of the percentage represented by some RAPD markers

Table 1. **Genetic variability distribution and Φ-statistics of *Salmo trutta* populations**

Level of subdivision	Variance	% Total	Φ statistics
Among basins (σ_a^2)	8.5	30.8	$\Phi_{CT} = 0.308$
Among rivers/basins (σ_b^2)	2.1	7.7	$\Phi_{SC} = 0.112$
Within rivers (σ_c^2)	16.9	61.5	$\Phi_{ST} = 0.385$

Table 2. **Complete D-Loop sequence including some nucleotides of the proline and phenilalanine t-RNA. The variable sites respective to AT1 (BERNATCHEZ et al. 1992) are shown with * and italics, and letters in bold and underlined show the other variable sites found in this study. The sequence corresponds to the light strain. Arrows show the end of the Proline t-RNA extreme and the beginning of the Phenilalanine t-RNA extreme. The numbers indicated above each variable site corresponds to the same numbers stated in Table 3. On the right side, nucleotide variability in the D-Loop sequence among populations**



Discussion

The nuclear genome scanned in this study is approximately 200 kb (this data is computed by summing the molecular weight of all the bands present in the patterns generated by 100

Table 3. **Variable sites in the complete D-Loop sequence between populations. Columns with numbers in italics correspond to variable sites coincident with the AT1 sequence (Bernatchez et al. 1992)**

	5	*16*	*20*	*21*	1	2	3	4	5	6	*7*	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
AT1	G	A	A	*	T	T	A	T	T	T	A	C	C	–	A	A	T	A	C	C	C	A	T	A	*	
German	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	*	–	
Narcea	A	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	*	–	–	–	–	–	–	–	
Tormes	–	G	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	*	–	*	*	–	*	–	–	
Luna	–	G	–	–	–	*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	
Jarama	A	–	*	–	–	–	*	–	C	–	–	–	*	–	–	–	–	–	–	–	–	–	–	–	–	
Puebla	A	–	–	–	–	–	–	–	C	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	T	
Lozoya A	–	*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	
Lozoya B	*	*	–	–	*	–	–	*	*	–	–	–	–	–	*	*	*	–	–	*	–	–	–	–	–	
Lozoya C	–	G	–	C	–	–	–	–	–	–	–	–	–	–	–	–	–	*	–	–	–	–	–	*	–	
Lozoya D	*	*	–	–	–	–	–	–	*	*	*	*	–	*	–	–	–	–	*	–	–	T	–	–	–	

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
AT1	C	G	C	T	C	G	A	C	C	*	T	A	C	T	C	A	A	T	A	T	C	A	A	A	*	*
German	C	–	–	–	–	*	A	–	*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	A	–	–
Narcea	C	–	–	–	–	–	A	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	*	A	–	–
Tormes	C	–	–	*	*	–	A	–	*	–	*	–	–	–	–	–	–	–	–	*	–	G	–	–	–	–
Luna	*	*	*	–	–	–	A	–	–	–	–	–	–	–	–	–	–	–	–	–	–	G		–	–	–
Jarama	C	–	–	–	–	–	A	*	–	–	–	–	–	–	–	–	–	–	*	–	–	–	–	–	T	–
Puebla	C	–	–	–	–	–	A	–	–	C	–	–	–	–	–	–	–	–	–	–	A	–	–	–	T	–
Lozoya A	C	–	–	–	–	–	A	–	–	–	–	–	–	–	–	–	–	–	–	–	–	*	–	*	–	–
Lozoya B	C	–	–	–	–	–	A	–	–	–	–	–	*	–	–	–	–	–	*	–	–	*	–	–	–	–
Lozoya C	C	–	–	–	–	*	A	–	–	–	–	*	–	–	–	–	–	*	–	–	–	G	–	–	–	–
Lozoya D	C	–	*	*	–	–	A	*	–	–	*	–	–	*	*	*	*	–	–	–	–	*	–	–	–	T

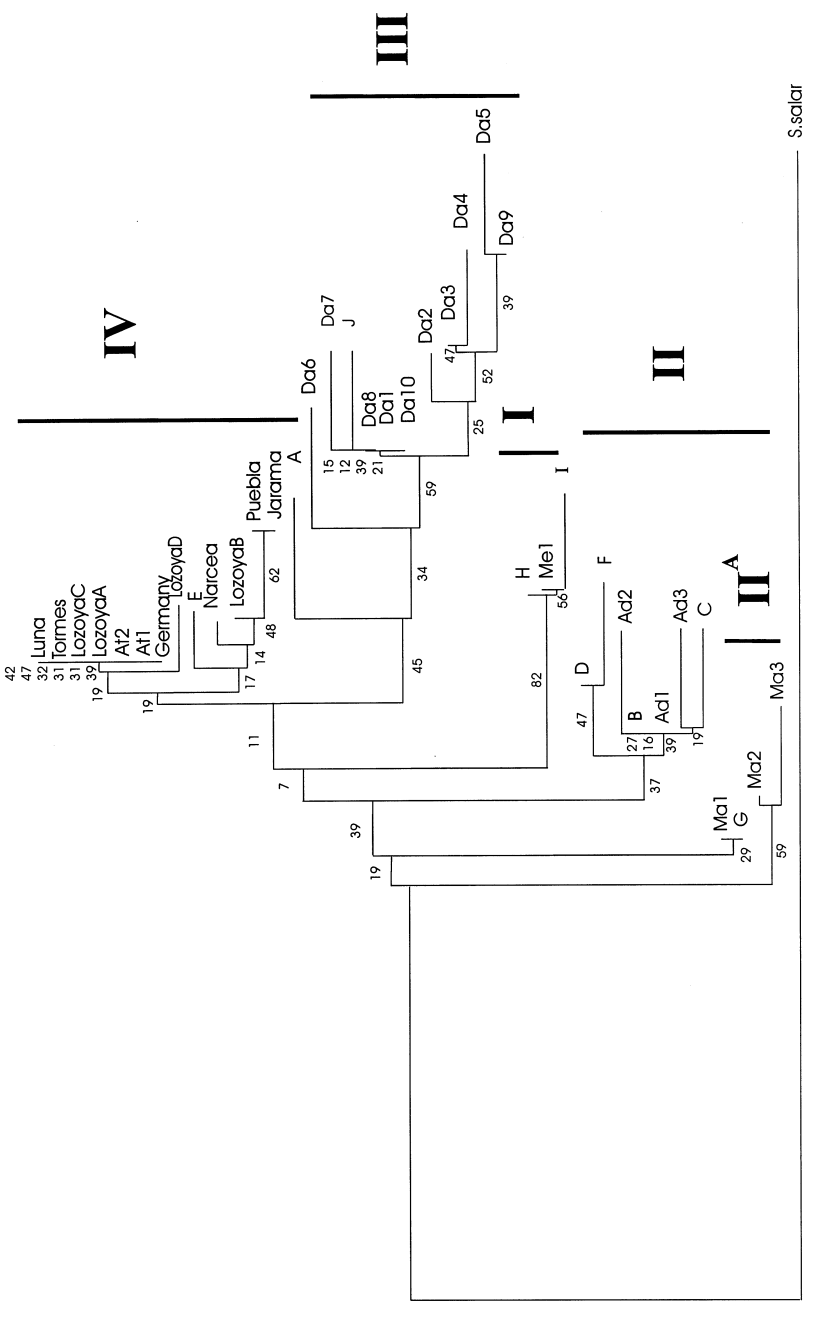


Fig. 5. Neighbour-joining tree relating all the sequences tRNA^{Pro} end (310 pb) available in GenBank including the 10 populations of the study. (A-J Apostolidis et al. (1996); Ad1-Ad3; At1-At3; Me; Ma1-Ma3 (Bernatchez et al. 1992; Giuffra et al. 1994). The numbers at the nodes are bootstrapping values for 1000 bootstrap resampling

primers) which increases the percentage of genome scanned dramatically in comparison with RFLP or allozymes.

The use of DNA pools reduces the effort required to identify polymorphism producing simple patterns with an easy detection of population specific bands. Bulked analysis accentuates overall genetic differences among populations (MICHELMORE et al. 1991) and is useful for reducing the number of candidate primers to be screened in individual trout (PERRON et al. 1995), thus lowering the number of genotypes to analyse.

Marker bands were found to be more frequent in some populations than others. However, no band was potentially a marker for any population; this would suggest a common origin, although differences among basins were great enough to show a clear differentiation and to allow its analysis.

The hierarchical genetic variability analysis of the populations showed a clear subdivision among regions (basins) (Table 1). Around 70% of the overall genetic diversity observed was within drainage ($\Phi_{CT} = 0.31$), which can also be inferred from the high correlation ($r = 0.72$) between geographic and genetic distances. This high proportion of the genetic variability assigned to the basin effect and the high correlation between geographic and genetic distances contrast with previous results obtained by GARCÍA-MARÍN and PLÁ (1996) for Spanish native populations of brown trout. These authors sampled from Mediterranean and Atlantic basins and used isozymes as molecular markers. The geographic genetic structure detected in the present study can presumably be caused by regional radiation and by regional policies of genetic management introducing different hatchery strains to increase the density of trouts in the rivers for fishing purposes.

The intrapopulation variability seems quite large when compared with other studies. In hatchery strains, PRESA et al. (1994) found that 91% of the total variability was explained by intrapopulation differentiation versus 43 or 50% present in the Mediterranean populations (PRESA et al. 1994). However, the fact that we found 62% of the total variability explained by within rivers variance shows the high genetic potential of the native populations.

Population genetics theory for nuclear genes predicts genetic differentiation among populations if the exchange of effective migrants per generation is less than one. The number of migrating individuals is calculated through the formula $N_m = \frac{1}{2}(1/\tilde{\Phi}_{CT} - 1)$ (HUDSON et al. 1992) and is 1.1 of migrants per generation corresponding to a small gene flow between populations. This supports the relatively high percentage of total variation found among basins.

The RAPD analysis allows the differentiation of the populations in two groups (Fig. 2a): the natural strains of the centre of Spain distributed around rivers of the Tajo basin, and the rest performing a different cluster, clearly defining the different geographical locations (Luna and Tormes rivers belonging to the Duero basin, the Narcea basin and the German hatchery strain). This is also shown in the tree constructed from the pairwise inter-individual distances (Fig. 3) and in the correspondence analysis (Fig. 4). Here the three axes show together approximately 84% of the total variability and cluster the populations in the different basins analysed in a very distinct way. This is an interesting figure where a few markers concentrate most of the variance measured and can show it in this three-dimensional way.

The high diversity exhibited in most vertebrate control regions has been detected in brown trout as well (BERNATCHEZ et al. 1992) although this hypervariability seems higher than in other fish but less than in mammals. When the complete D-Loop sequence is used, levels of divergence vary from 0.8 and 3.9% being higher than values that were previously observed when only the t-RNA ends were analysed (0.16–1.92%; GIUFFRA et al. 1994). APOSTOLIDIS et al. (1997) found the NADH dehydrogenase genes to be more variable than the $tRNA^{Pro}$ and the $tRNA^{Phe}$ ends of the control region. In the present study however, the highest variability of the control region is found to be in the segment between 400 bp and 600 bp (Table 2). When comparing the whole sequence, especially the internal segment, populations showed another pattern of differentiation. The cluster analysis performed with

the distance data showed a general divergence between populations making the distinction between basins less clear than that resulting from the RAPD analysis, but somehow different from that resulting from the analysis of the ends of the D-Loop (Fig. 2b).

The type of mutations did not differ much from the rest of observations performed on Salmonids in other studies. GIUFFRA et al. (1994) found a lower transitional bias in the control region of brown trout by using the 5′ t-RNA end. In this study, however, a lower bias was noted, making the ratio between transitional and transversional events 2 : 1. It has already been stated (BROWN et al. 1982) that this ratio is very unusual among vertebrates and seems to be a sign of higher taxonomic differentiation as shown in primates where the ratio varies from 32 : 1 at the intraspecific level, decreasing to 3 : 1 between very divergent species. When this ratio is looked at in the populations of the Tajo basin, the value decreases to approximately 2 : 1 (13 : 6 or 15 : 10) among the rest of the populations. These results, as well as the AT content found in this study (64%), are similar to those found by other authors: BERNATCHEZ and DANZMANN (1993) for Salvelinus fontinalis and APOSTOLIDIS et al. (1997) for brown trout. This suggests that there is a structural or functional need to maintain a given A + T composition in the control region of salmonids which could lead to this lower probability of transitions. The differences between the Iberian populations studied here using the inner part of the control region suggest a high differentiation rate not found with the less variable t-RNA ends which demonstates how more informative the sequences containing nucleotides 400–600 are, when compared to the ends of the control region.

A comparative study of the $tRNA^{Pro}$ extreme (Fig. 5) revealed the inclusion of our 10 populations into the group IV defined in BERNATCHEZ et al. (1992). Group IV was defined by these authors as the Atlantic group and was clearly differentiated from the Mediterranean, Danubian, and Adriatic groups. Our samples belong to rivers which drain into the Atlantic, so the fact that they cluster in this group was highly predictable. Although there are some unexplained clustering APOSTOLIDIS et al. (1997) previously discussed, the 310 bp of the proline extreme does not seem to be sufficiently informative to conclude if there is any hybridization with foreign strains as all samples belong to the same origin group. As no complete sequences of the control region are available for other populations, it has not been possible to verify the extent of this conclusion.

The highest variability of the control region found in the segment between 400 and 600 bp makes the use of internal primers for comparative purposes and phylogeographic assemblages recommendable.

The value for genetic distances found in the populations examined ranged from 0.011 to 0.039 for mtDNA and from 0.04 to 0.73 for nuclear DNA. Distances generated by nuclear or mitochondrial analysis give differences (matrix correlation = 0.44, Mantel test, $p < 0.05$), and small incongruences raise in the differentiation between some Tajo basin populations (Jarama river and Lozoya A (Aguilón stream) stream for example) and between these and the German stocked strain. When nuclear DNA is used, the resulting analysis cluster very clearly the different basins while mtDNA does not. These incongruences are in agreement with APOSTOLIDIS et al. (1996b) who found no mtDNA diversity in two populations which showed among the highest values of nuclear heterozygosity. Lack of congruence between mitochondrial and nuclear-based studies can result from the different modes of transmission (mtDNA is transmitted exclusively by females) and evolution (because of the smaller effective population size of the mtDNA compared with the genomic DNA) of this genetic material, but also from differential introgression rules, e.g. mainly only wild males are used to cross with hatchery females.

Although the introgression level between the foreign stocks and the native ones are difficult to interpret with the present data, there is strong evidence of a high variability between rivers where stocked strains have been routinely introduced for fishing purposes but where the extent of a possible hybridization is not easy to detect, perhaps because of a lack of samples of the possible hatchery strains used in the different provinces within which

each basin is included. Estimates of the stocking effect on wild populations fluctuate greatly between 0 and 50% (PRESA et al. 1994). However, analysis of both RAPD polymorphism and mtDNA sequencing could allow detection of these entries, making possible a controlled introduction of stock with nonintrogressive native populations originated in other parts of the country to avoid a loss of genetic variability.

Recently some methods (SHRIVER et al. 1997; GÖTZ and THALLER 1998) have shown the ability of molecular markers to estimate admixture proportions and to better understand dynamics of the admixture process, which should be applied in future to construct admixture maps.

Acknowledgements

This work has been supported by the Consejeria de Medio Ambiente y Desarrollo Rural of the Comunidad de Madrid. We wish to thank Cristina Cristóbal for her technical assistance.

References

APOSTOLIDIS, A. P.; KARAKOUSIS, Y.; TRIANTAPHYLLIDIS, C., 1996a: Genetic divergence and phylogenetic relationships among Salmo trutta L. (brown trout) populations from Greece and other European countries. Heredity **76:** 551–560.

APOSTOLIDIS, A. P.; KARAKOUSIS, Y.; TRIANTAPHYLLIDIS, C., 1996b: Genetic differentiation and phylogenetic relationships among Greek *Salmo trutta* L. (brown trout) populations as revealed by RFLP analysis of PCR amplified mitochondrial DNA segments. Heredity **77:** 608–618.

APOSTOLIDIS, A. P.; TRIANTAPHYLLIDIS, C.; KOUVATSI, A.; ECONOMIDIS, P. S., 1997: Mitochondrial DNA sequence variation and phylogeography among *Salmo trut*ta L. (Greek brown trout) populations. Mol. Ecol. **6:** 531–542.

AVISE, J. C.; ARNOLD, J.; BALL, R. M. JR; BERMINGHAM, E.; LAMB, T.; NEIGEL, J. E.; REEB, C. A.; SAUNDERS, N. C., 1987: Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. **18:** 489–522.

AVISE, J. C., 1994**.** Molecular Markers, Natural History and Evolution. Chapman & Hall, New York.

BERNATCHEZ, L.; GUYOMARD, R.; BONHOMME, F., 1992: DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. Mol. Ecol. **1:** 161–173.

BERNATCHEZ, L.; OSINOV, A., 1995: Genetic diversity of trout (genus *Salmo*) from its most eastern native range based on mitochondrial DNA and nuclear gene variation. Mol. Ecol. **4:** 285–297.

BERNATCHEZ, L.; DANZMANN, R. G., 1993: Congruence in control region sequence and restriction-site variation in mitochondrial DNA of brook charr (*Salvelinus fontinalis* Mitchill). Mol. Biol. Evol. **10:** 1002–1014.

BIRKY, C. W.; FUERST, P.; MARUYAMA, T., 1989: Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics **121:** 613–627.

BROWN, W. M.; PRAGER, E. M.; WANG, A.; WILSON, A. C., 1982: Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18:** 225–229.

CAVALLI-SFORZA, L. L.; MENOZZI, P.; PIAZZA, A., 1994**.** The History and Geography of Human Genes. Princetown University Press, Princetown, NY.

CRONIN, M. A.; SPEARMAN, W. J.; WILMOT, R. L.; PATTON, J. C.; BICKHAM, J. W., 1993: Mitochondrial DNA variation in Chinook (*Oncorhynchus tshawytscha*) and Chum Salmon (*O. Keta*) detected by restriction enzyme analysis of polymerase chain reaction (PCR) products. Can. J. Fish Aquatique Sci. **50:** 708–715.

DEVERAUX, J.; HAEBERLI, P.; SMITHIES, O., 1984: A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. **12:** 387–395.

ESTOUP, A.; PRESA, P.; KRIEG, F.; VAIMAN, D.; GUYOMARD, R., 1993: (CT) n and (GT) n microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). Heredity **71:** 488–496.

ESTOUP, A.; ROUSSET, F.; MICHALAKIS, Y.; CORNUET, J. M.; ADRIAMANGA, M.; GUYOMARD, R., 1998: Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). Mol. Ecol. **7:** 339–354.

EXCOFFIER, L.; SMOUSE, P. E.; QUATTRO, J. M., 1992: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131:** 479–491.

FELSENSTEIN, J., 1985: Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

GARCÍA-MARÍN, J. L.; JORDE, P. E.; RYMAN, N.; UTTER, F.; PLA, C., 1991: Management implications

of genetic differentiation between native and hatchery populations of brown trout (*Salmo trutta*) in Spain. Aquaculture **95:** 235–249.
García-Marín, J. L.; Plá, C., 1996: Origins and relationships of native populations of *Salmo trutta* (brown trout) in Spain. Heredity **77:** 313–323.
Giuffra, E.; Bernatchez, L.; Guyomard, R., 1994: Mitochondrial control region and protein coding genes sequence variation among phenotypic forms of brown trout *Salmo trutta* from northern Italy. Mol. Ecol. **3:** 161–171.
Götz, K.-U.; Thaller, G., 1998: Assignment of individuals to populations using microsatellites. J. Anim. Breed. Genet. **115:** 53–62.
Hudson, R. R.; Slatin, M.; Maddison, W. P., 1992: Estimation of gene flow from DNA sequence data. Genetics **132:** 583–589.
Karakousis, Y.; Triantaphyllidis, C., 1990: Genetic structure and differentiation among Greek brown trout (*Salmo trutta* L.) populations. Heredity **64:** 297–304.
Kimura, M., 1980: A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.
Kocher, T. D.; Thomas, W. K.; Meyer, A.; Edwards, S. V.; Paabo, S.; Villablanca, F. X.; Wilson, A. C., 1989: Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl Acad. Sci. USA **86:** 6196–6200.
Kumar, S.; Tamura, K.; Nei, M., 1993: The Pennsylvania State University, University Park, PA.
Mantel, N. A., 1967: The detection of disease clustering and a generalized regression approach. Cancer Res. **27:** 209–220.
Michelmore, R. W.; Paran, I.; Kesseli, R. V., 1991: Identification of markers linked to disease-resistance gene by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl Acad. Sci. USA **88:** 9828–9832.
Moritz, C., 1994: Applications of mitochondrial DNA analysis in conservation: a critical review. Mol. Ecol. **3:** 401–411.
Nei, M.; Li, W. H., 1979: Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl Acad. Sci. USA **76:** 5269–5273.
Perron, M.; Gordon, A. G.; Bousquet, J., 1995: Species-specific RAPD fingerprints for the closely related *Picea mariana* and *P. rubens*. Theoret. Appl. Genet. **91:** 142–149.
Presa, P.; Krieg, F.; Estoup, A.; Guyomard, R., 1994: Diversité et gestion génétique de la truite commune: apport de l'étude du polymorphisme des locus protéiques et microsatellites. Genet. Sel. Evol. **26** (suppl 1)**:** 183s–202s.
Prodohl, P. A.; Taggart, J. B.; Ferguson, A., 1994: Single locus inheritance and joint segregation analysis of minisatellite (VNTR) DNA loci in brown trout (*Salmo trutta* L.). Heredity **73:** 556–566.
Rohlf, F. J., 1993: NTSYS-pc. Numerical Taxonomy and Multivariate Analysis System. Exeter Software. New York.
Saitou, N.; Nei, M., 1987: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.
Sambrook, J.; Fritsch, E. F.; Maniatis, T., 1989**.** Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.
Shriver, M. D.; Smith, M. W.; Jin, L.; Marcini, A.; Akey, J. M.; Deka, R.; Ferrell, R. E., 1997: Ethnic-affiliation estimation by use of population-specific DNA markers. Am. J. Hum. Genet. **60:** 957–964.
Slatkin, M., 1995: A measure of population subdivision based on microsatellite allele frequencies. Genetics **139:** 457–462.
Tamura, K., 1992: Estimation of the number of nucleotide substitutions when there are strong transition-transversuions and G+C content biases. Mol. Biol. Evol. **9:** 678–687.
Williams, J. G. K.; Kubelik, A. R.; Livak, K. J.; Rafalski, J. A.; Tingey, S. V., 1990: DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. **18:** 6531–6535.

Address of authors': S. Dunner, Laboratorio de Genética Molecular. Dpto. de Producción Animal. Facultad de Veterinaria. 28040 Madrid. Spain. Tel.: +34 91 394 37 65; fax: +34 91 394 37 72; e-mail: dunner@eucmax.sim.ucm.es