

Microsatellites and artificial neural networks: tools for the discrimination between natural and hatchery brown trout (*Salmo trutta*, L.) in Atlantic populations

Didier Aurelle ^{a,*}, Sovan Lek ^b, Jean-Luc Giraudel ^c, Patrick Berrebi ^a

^a *Laboratoire Génome et Populations, CNRS UPR 9060, Cc063, Université Montpellier II, Place Eugène Bataillon, 34095, Montpellier Cedex 05, France*

^b *CESAC, UMR 5576, Bat 4R3, CNRS-Univ. Paul Sabatier, 118 route de Narbonne, 31062, Toulouse Cedex, France*

^c *IUT Périgueux Bordeaux IV, Département Génie biologique, 39 rue Paul Mazy, 24019, Périgueux Cedex, France*

Abstract

Artificial Neural Networks (ANN) were applied to microsatellite data (highly variable genetic markers) to separate genetically differentiated forms of brown trout (*Salmo trutta*) in south-western France. A classic *feed-forward* network with one hidden layer was used. Training was performed using a back-propagation algorithm and reference samples representing the different genetic types. The hold-out and the leave-one-out procedures were used to test the validity of the network. They were chosen according to the populations and the questions analysed. The informative content of the different variables used for the distinction (the alleles of the different loci) was also evaluated using the Garson–Goh algorithm. The results of learning gave high percentages of well-classified individuals (up to 95% for the test with the hold-out analysis). This confirms that ANNs are suitable for such genetic analyses of populations. From a biological point of view, the study enabled evaluation of the genetic composition and differentiation of different river populations and of the impact of stocking. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Artificial Neural Network; Classification; Microsatellites; Stocking; Brown trout

1. Introduction

Salmonids are extensively studied fishes both from a practical point of view (fisheries management) and for some more theoretical aspects (ecology and evolution). The brown trout (*Salmo trutta* L.) displays some interesting biological characteristics for the study of genetic intraspe-

cific differentiation: brown trout lives in the upper part of the rivers and is philopatric. Genetic studies have shown that the species *S. trutta* includes several genetic entities. For example, in the western part of the French Pyrenees, two wild forms are present naturally: ancestral Atlantic and modern Atlantic (the first one was called ancestral according to Hamilton *et al.*, 1989). Moreover, stocking practices led to the introduction there (and more generally in most French rivers) of a third form, the domestic modern Atlantic trout,

* Corresponding author. Fax: + 33-467-144-554..

E-mail address: aurelle@crit.univ-montp2.fr (D. Aurelle)

which does not originate from these rivers (Aurelle and Berrebi, 1998). The three forms may be found in the same river and can hybridise.

Nevertheless, the classification of individuals among the different forms is a prerequisite for the study of genetic interactions. Allozymes separate modern and ancestral forms, but no diagnostic markers are available to distinguish between domestic and wild modern Atlantic trout. However, microsatellites have shown that the distinction is justified as the populations of some rivers appear to be genetically different to hatchery strains (Aurelle and Berrebi, 1998). Because of microsatellite properties, distinction between individuals of the different forms remains difficult. These loci usually display a high mutation rate and are subject to retention of ancestral polymorphism and homoplasy phenomena (Jarne and Lagoda, 1996). There are numerous shared alleles between wild and domestic modern populations and only some differences in allelic frequencies. It is, therefore, necessary to use powerful statistical classification tools to appraise the genetic composition of the populations studied and at the same time to separate natural migration and human manipulations (stocking).

Artificial neural networks (ANNs) seem well-suited to the problem. They have already been used for a wide range of different studies and situations. They are commonly used in physics and chemistry but less so in ecology and population genetics. However, preliminary studies have shown that ANNs are suitable for these topics (Guégan *et al.*, 1998) and more effective than classic discriminant analysis Cornuet *et al.*, 1996; Mastrotrillo *et al.*, 1997). Moreover, no particular assumptions are required concerning the data used for classification. ANNs have proven to be effective in population genetics, at several different taxonomic levels and with highly variable markers such as microsatellites (Cornuet *et al.*, 1996). They are, therefore, expected to be capable of classifying individuals in populations belonging to the same sub-species and genetically relatively similar (e.g. wild and domestic modern trout). Until now, neural networks have been tested with some well separated and genetically differentiated groups (such as bees in Cornuet *et al.*, 1996). In

the work reported here, we applied them to mixed populations where samples may contain several genetic units; this raises the question of the reference samples necessary for training the network (see Section 2) and that of the validation procedures (how can we know if the result is right?). Several training and validation procedures were tested depending on the situation.

Analyses were performed with different purposes. Firstly, we wished to verify using independent markers (microsatellites), the distinction between modern and ancestral fishes which is shown by allozymes at only one locus (*LDH5**); this also enabled us to test the method in a clear, well known situation. We then sought wild modern populations (with no or almost no stocking influence). This enabled us to evaluate the genetic composition of the different populations analysed here. The importance of the different alleles in the classification (and their informative content) is also discussed for the different microsatellite loci used.

2. Materials and methods

2.1. The populations analysed

The populations from nine rivers and three hatchery strains were analysed. The origins and sizes of the samples are provided in Table 1. The numbers refer to Fig. 1, and the percentages of allele *LDH5*90* provide some information about the genetic composition of the populations. The ancestral form is characterised by allele 100 at this locus whereas the two modern forms possess allele 90. A population with 100% *LDH5*90* is then considered as modern, but we do not know whether these fishes are wild or domestic (there is no diagnostic allele for this distinction). Some populations consisting of only a few individuals were analysed because they were genetically and morphologically original (Andurentako) or because they seemed to be mixed (Marcadau which, according to local managers is heavily stocked; moreover hatchery fishes are often easy to recognise thanks to coloration) but we kept in mind the problems of small samples.

According to allozymic data (unpublished) some river samples consisted mainly of modern fishes (Chiroulet, Oussouet and Luz) and certain other samples were almost completely ancestral (Dancharia, Andurentako, Béhérékobentako and Bastan). According to local managers, these populations have not been stocked for several years. Moreover, the morphological characteristics would tend to show that Chiroulet, Oussouet and Luz fishes are mainly wild. Marcadau and Béhérobie contain both modern and ancestral fishes.

The morphology of Marcadau fishes tend to show that the population is quite heavily restocked.

2.2. Microsatellite loci

Four microsatellite loci were analysed. *Strutta* 58 has been cloned by Poteaux (1995). MST 73 and MST 15 have been cloned by Estoup (Estoup et al., 1993). MSU 4 has been published in Genbank under accession number U43694; it was submitted directly by P.T. O'Reilly and has been

Table 1

Origin and characteristics of the samples; bold names refer to the samples names used in the text

No. (map)	Locality	River	Basin	Sample size	% LDH5*90
	La Canourgue	hatchery		50	95
	Brassac	hatchery		30	100
	Suech	hatchery		36	99
1	Cauterets	Marcadau	Adour	15	33
2	Sare	Beherekobentako	Nivelle	24	0
3	Dancharia	Nivelle	Nivelle	30	2
4	Herboure	Andurentako	Untxin	5	0
5	Bidarray	Bastan	Adour	29	4
6	Béhérobie	Nive de Béhérobie	Adour	25	27
7	Chiroulet	Adour de Lesponne	Adour	86	89
8	Bagnères de Bigorre	Oussouet	Adour	86	82
9	Argeles	Luz	Adour	88	95

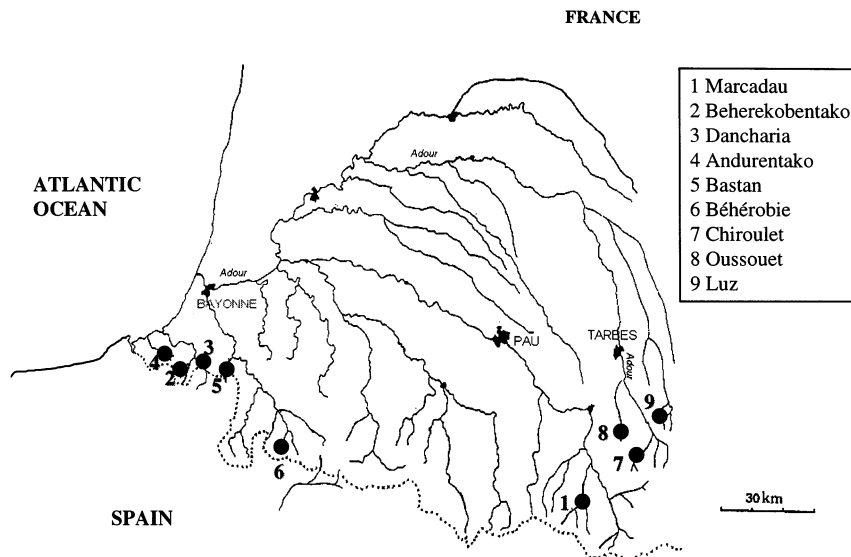


Fig. 1. Location of the sampling points.

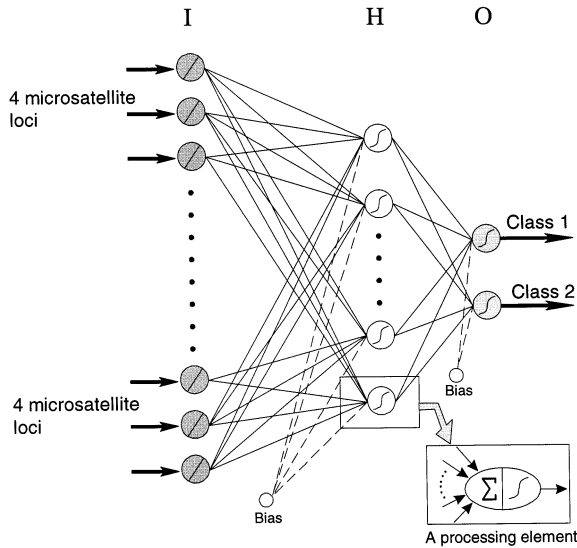


Fig. 2. Structure of an Artificial Neural Network (ANN).

identified in salmon (*Salmo salar*). Two of these four loci were highly variable (*Strutta 58* and *MSU 4* with 38 and 18 alleles respectively). The two others displayed only a few alleles in comparison with the usual microsatellite variability (seven alleles for *MST 73* and eight for *MST 15*). PCR and analyses procedures are described in Aurelle and Berrebi (1998).

2.3. Artificial neural networks

A classic *feed-forward* network (Rumelhart et al., 1986) was used in the study. This network had three layers: an input layer, a hidden layer and an output layer. The input layer was connected with the variables used for discrimination; in our study, these variables were the 71 alleles coded as follows: for each allele, each individual was noted zero if it did not possess it, one if the fish was heterozygotic for the allele and two if it was homozygotic for it. The hidden layer was reduced to two neurones to avoid too large a number of parameters; this choice did not reduce the network efficiency beyond reasonable limits. The number of neurones in the output layer corresponds to the number of categories in which individuals should be classified (depending on the analyses, see Section 2.4).

Each neurone is connected with the neurones of the neighbouring layers; it receives and sends signals through these connections and always from input to output (Fig. 2). Each connection is weighted according to the signal intensity. Each neurone integrates the signals received from the former neurones and sends a new signal to the next ones. This signal is delivered according to a non-linear transfer function applied to the sum of the weighted signals of the former neurones (see Cornuet et al., 1996; Mastorillo et al., 1997). Let w_i and x_i be the weight and the signal outgoing from the former neurone i (layer n); the incoming signal for one neurone in the layer $n + 1$ will be:

$$z = \sum w_i * x_i \quad (1)$$

The outgoing signal for this neurone in layer $n + 1$ will then be:

$$f(z) = [1 + \exp(-z)]^{-1} \quad (2)$$

For the input layer, incoming signals correspond to the variables used to classify individuals (the 71 alleles). The outgoing signals of the output layer designate the category where the studied individual will be assigned by the network. The decision is made in the light of the highest score. Nevertheless, as is mentioned in Section 2.4, absolute output values can and should be discussed. For example, individuals with a score of one in a group can be considered as quite accurately classified in this category but the interpretation of individuals with intermediate scores (0.5 for example) is not as easy. On the other hand, individuals with scores of zero to 0.1 in their original category can be considered to be incorrectly classified.

The network must be trained in order to classify individuals correctly. A training data set (randomly chosen in the global data set) is used to modify the weights of the different connections in order to maximise the percentage of well-classified individuals. We used a ‘back-propagation’ algorithm. First, the initial weights are randomly distributed. They are then modified iteratively depending on the differences between expected and observed output signals (assignment scores; see Cornuet et al., 1996; Mastorillo et al., 1997).

Numerous iterations are usually necessary to obtain a good percentage of well-classified individuals without an over-fit to the training data set. Effectively, if the percentage of well classified individuals is much higher for the learning data than for the test data (see below), we can deduce that the network has learned the training data particularities and cannot be applied to a more general situation.

A hold-out procedure (Kohavi, 1995) can be used to test the validity of the network. For this purpose, a data set with some known categories is divided into two parts. The first part is used for training the network. When the training procedure has been completed, the network is then applied to the second part and we can evaluate the percentage of well classified individuals for data not used for learning. This second part is then used as a test. Once it has been verified that the network is well suited and does not over-fit the learning data, it can be applied to unknown data (application stage).

If the data set is too small to be divided into two parts or if its composition is not well known and possibly heterogeneous, one can use the leave-one-out procedure (Kohavi 1995). For example, for a data set with N individuals, training is performed with $N - 1$ individuals (by assuming that their categories are known) and the network is applied to the N th individual, which is then classified according to its proximity to one of the previously learned categories. This analysis is repeated for the N individuals which are all assigned to one group. Given the high number of training stages (N steps), the number of iterations for each training is limited to 500.

For analysis of the results, each individual was assigned to the category where it showed the highest score. At the population level, it is interesting to study the individual score distributions for the various categories. In order to analyse the contributions of the different alleles to classification, we used the Garson–Goh algorithm (Garson, 1991; Goh, 1995; Lek *et al.*, 1996a,b). This algorithm determines the relative importance of the various input variables by taking into account the weights of the hidden layer neurones connected with these input. Briefly, for each hidden

neurone, the weight of the connection from one input variable to this neurone is multiplied by the weight of one output connection; these products are summed for all the output connections and then expressed relatively as a percentage for the comparison of all input variables. These percentages are intended to express the informative content of each variable.

2.4. Analysis protocols

(1) First, we tested the effectiveness of the method for a situation in which some genetic markers different from microsatellites were able to distinguish between several categories. Here, **modern and ancestral** individuals can be separated with allozymes (especially with the *LDH-5** locus). The training set consisted of four ancestral populations (Bastan, Béhérékobentako, Dancharia and Andurentako) versus four modern populations (the three hatcheries and Luz). This distinction was analysed using a hold-out (1a) and then a leave-one-out (1b) procedure.

2) We then analysed the **hatchery populations**. The different strains are assumed to be genetically quite similar so the sample analysed should be representative of the different hatchery strains used in the country. We tried to verify these assumptions by using a leave-one-out procedure (for the analysis of all individuals and because one strain may be heterogeneous) with three categories corresponding to the three strains analysed.

3) We also sought **wild modern Atlantic** populations. As a modern population may be heterogeneous and contain wild and domestic fishes, we decided to use a leave-one-out procedure with two classes comparing each modern river population to hatcheries which were pooled (according to the results of analysis two showing the genetic homogeneity of these strains). Three tests were performed: Chiroulet versus hatcheries, Oussouet/hatcheries and Luz/hatcheries.

4) The **other river populations** (ancestral and mixed) were also compared to hatcheries by the leave-one-out method to examine the potential influence of domestic fishes in these samples. The leave-one-out procedure is useful for this comparison because each fish is analysed individually

and the presence of a foreign fish (a domestic fish in a river) can theoretically be detected. We compared Bastan with the pooled hatcheries, Béhérobie versus hatcheries and Marcadau versus hatcheries.

3. Results

For each analyses we will give some percentages of so-called ‘incorrectly classified individuals’: this indicates individuals which were not classified by the network in the population where they were sampled. Nevertheless, they can either be classified in the population from which they originate (as for example some domestic fishes classified in the hatchery category but sampled in one river) or they can effectively correspond to some errors of the network.

3.1. The ancestral—modern distinction

(1a) The percentage of incorrectly classified individuals by leave-one-out is 2% in the global comparison between ancestral and modern. This proportion is 1% among supposedly modern in-

dividuals and 7% for populations expected to be ancestral. Analysis of the distribution of the scores within the ancestral category for ancestral populations (Fig. 3) shows that most individuals (65%) score between 0.8 and one; 26% are between 0.5 and 0.8, corresponding to less sharp and correct assignation, like the 2% scoring between 0.3 and 0.5. Finally, 7% should really be classified in the other group (score between 0.1 and 0.3). Conversely, more than 80% of modern individuals scored between zero and 0.1 and were then well classified in their original category. 1% scored between 0.9 and one and were assigned to the ancestral type whereas they were in a modern sample.

(1b) With the hold-out procedure, we observed 1% of incorrectly classified fishes in the learning stage and 5% in the test. When this network is applied to new populations, the percentage of modern individuals can be evaluated and compared to the frequency of the *LDH-5** modern allele (Table 2). There is a reasonably good correlation between the two sets of variables.

For this analysis, the contributions of the different alleles to the network are shown in Fig.

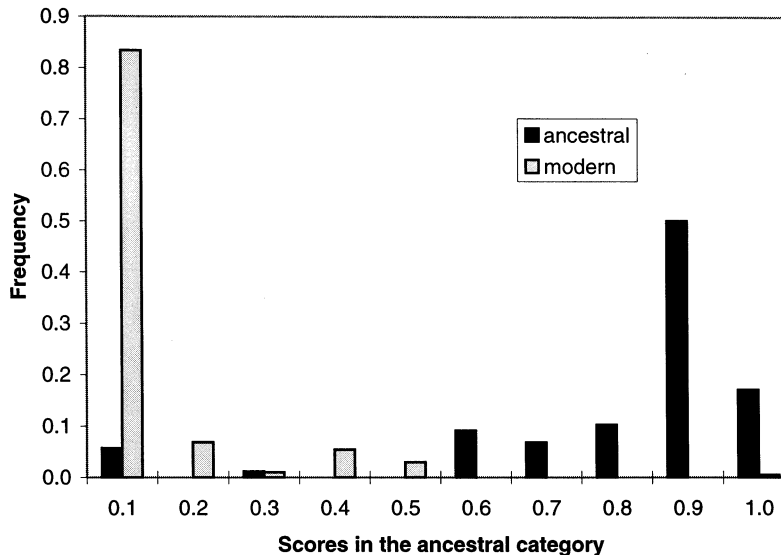


Fig. 3. Score distribution in the ancestral category for the leave-one-out comparison between ancestral and modern. The first category corresponds to scores of between zero and 0.1 and the second to scores of between 0.1 and 0.2,....

Table 2

Percentage of modern individuals in four populations as predicted by artificial neural network compared with the frequency of modern *LDH5** allele

Populations	Neural network predictions (% modern individuals)	Allozyme predictions (% modern alleles)
Béhérobic	32	27
Marcadau	53	33
Chiroulet	73	89
Oussouet	73	82

4. On the x axis, alleles are classified by increasing abundance in the overall data set. The more frequent alleles usually contribute more to analysis

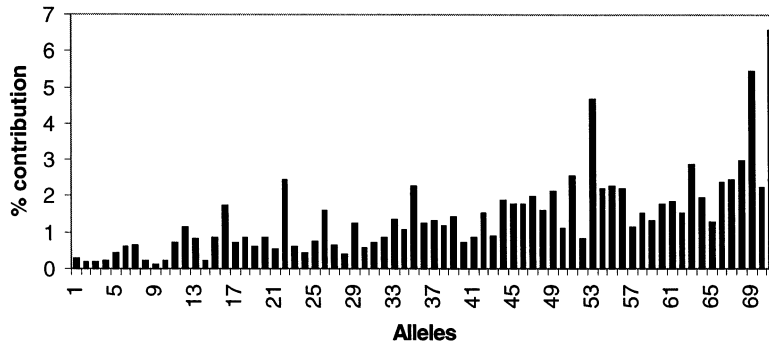


Fig. 4. Contributions of the different alleles to the leave-one-out ancestral/modern. Alleles are set out on the x axis according to their frequency in the overall data set (all loci are included). Contributions are computed with the Garson–Goh algorithm (Garson, 1991; Goh, 1995; Lek *et al.*, 1996a,b).

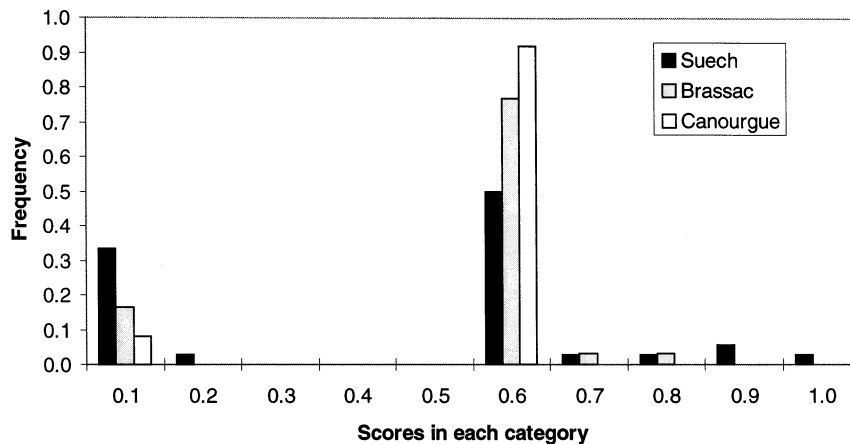


Fig. 5. Scores of the different hatcheries individuals in their own category for the leave-one-out with three groups corresponding to the three hatchery strains: Canourgue, Brassac and Suech.

than those that are fairly rare, but with some exceptions. Some rare alleles can be useful or not, depending on the analysis.

3.2. The different hatchery strains

(2a) In the leave-one-out analysis with three categories corresponding to the three hatchery strains, 19% of the individuals were found to be incorrectly classified, which is high compared to the previous analyses. The score distribution of each strain in its corresponding category (Fig. 5) shows that only a very small proportion of individuals scored between 0.8 and one (2.6%; none for Brassac and Canourgue); most scores are be-

Table 3

Percentage of incorrectly assigned individuals for each of the three comparisons between modern river and hatchery populations. Such individuals are assigned to the opposite category, e.g. 6% of Chiroulet individuals are classified in 'hatcheries'

Comparison	Chiroulet/hatcheries	Oussouet/hatcheries	Luz/hatcheries
river populations	6	8	5
hatcheries	5	3	3

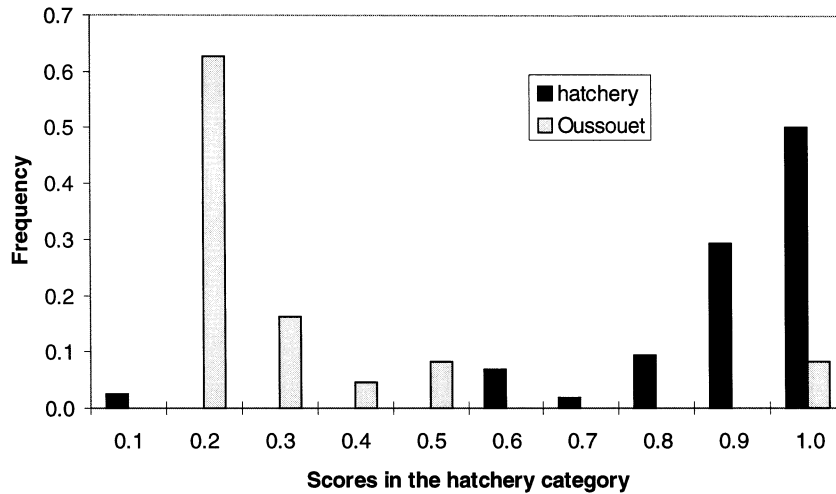


Fig. 6. Distribution of the scores in the hatchery class for the comparison between Oussouet (modern river population) and hatcheries (the three domestic strains analysed have been pooled). 1 = hatchery, 0 = Oussouet.

tween 0.5 and 0.6 (75%) and a large proportion of fishes scored between zero and 0.1 (18%, corresponding to incorrectly classified individuals).

(2b) In the hold-out procedure, 2% of the individuals in the training set were not correctly classified, but the test showed 17% errors. This would tend to show that the network was suited to the features of the learning data set but not well suited to new data. There may be too small an overall difference between the different strains, preventing good application to new data.

3.3. The 'wild' modern populations

The percentages of incorrectly classified individuals for each of the three leave-one-out comparisons with hatchery samples (Luz, Chiroulet and

Oussouet compared with domestic fishes) are given in Table 3. In the three river populations, the percentage of individuals assigned to domestic types varied from 5 (Luz) to 8% (Oussouet).

In the Oussouet/hatcheries comparison, the score distribution of Oussouet individuals in the hatchery category placed most individuals between 0.1 and 0.2 (Fig. 6), but with a large proportion between 0.2 and 0.5. Individuals with a result higher than 0.5 were all in the 0.9–one range and were then well assigned to hatcheries. Almost 80% of domestic individuals, scored between 0.8 and one. Individuals with a score lower than 0.5 were all in the 0–0.1 range and were then classified as Oussouet. It appeared to be more difficult to classify wild trout than domestic ones in this analysis.

3.4. Comparison of the other river populations with hatcheries

In the leave-one-out comparison between an ancestral population (such as Bastan) and hatcheries, we obtained 1% ‘errors’ in the domestic strains and 3% in the ancestral population. However, analysis of the score distribution in the hatcheries category (Fig. 7) shows that 97% of the Bastan fishes scored between 0.4 and 0.5; the remaining 3% corresponded to fishes classified as domestic (score between 0.9 and one in this group). In contrast, hatchery individuals are all well classified with scores between 0.6 and one in the hatchery category. The computation procedure may perhaps explain why no Bastan individual displayed a high score (between 0.8 and one) in its own category: as the time required by this technique is quite long, the number of iterations for the learning of each individual was limited to a maximum of 500. However, there must be a phenomenon making learning more difficult for this comparison than for the former ones. The same analysis was performed with Béhérobic (with some similar results to Bastan) and with Marcadau.

For Marcadau (Fig. 8), 40% of the individuals displayed a hatcheries category score of between 0.4 and 0.5; 60% scored between 0.9 and one and were then assigned to domestic type. These results

agree well with morphological observations and with information from local managers, which tend to show that this population is quite heavily stocked. In this set of analyses, the scores of wild trout are limited to 0.5. However, if we agree that individuals with scores of between zero and 0.5 in the hatcheries category are wild trout, we can deduce that Marcadau is the population analysed that has been most modified by stocking.

4. Discussion

When the trout classes had been previously well defined using allozymes (comparison of ancestral and modern, tests (1a) and (1b)), the first analyses confirm that neural networks give good results when applied to microsatellite data despite all the problems usually associated with these markers, and especially the presence of rare alleles, ancestral polymorphism and homoplasy which means that some alleles of the same size are not always identical by descent (Jarne and Lagoda, 1996). Because of the high mutation rate of microsatellite loci (particularly for loci with a high number of alleles), and because of a possible relatively recent coancestry of the populations analysed (both natural and domestic), it is difficult to find diagnostic alleles separating wild and hatchery Atlantic populations and which could be used for

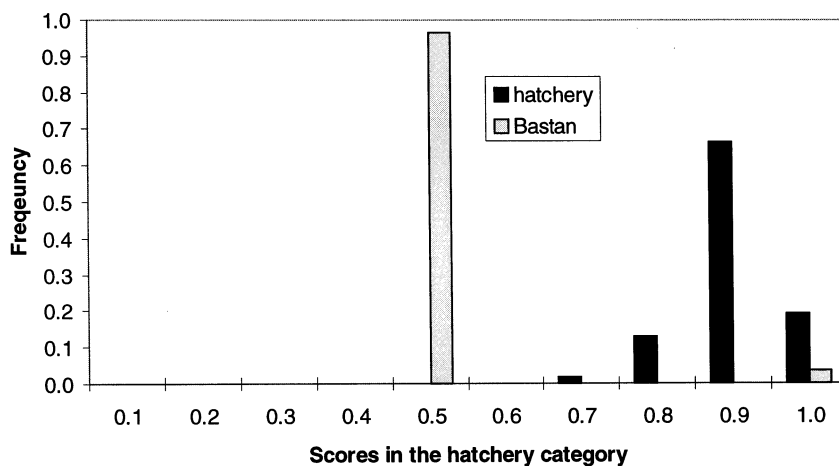


Fig. 7. Distribution of the scores in the hatchery category for the leave-one-out Bastan/hatcheries; 1 = hatcheries, 0 = Bastan.

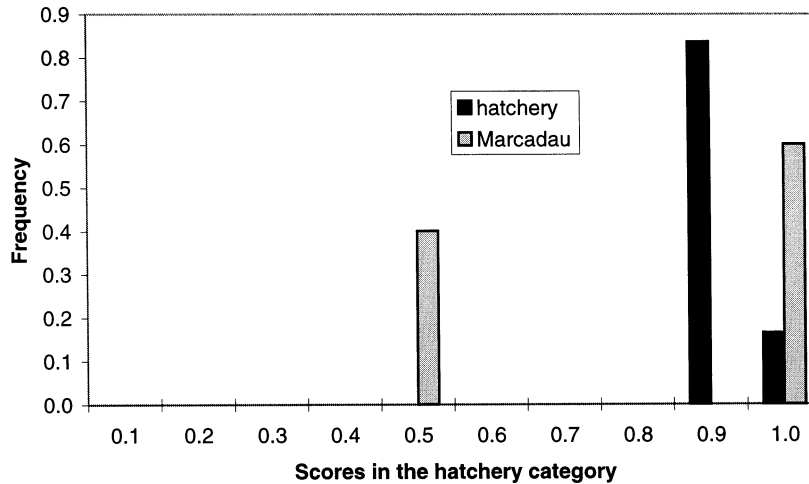


Fig. 8. Score distribution in the hatcheries category for the leave-one-out Marcadau/hatcheries. 1 = hatcheries, 0 = Marcadau.

several river drainage basins. For this reason, multilocus analysis is more useful and particularly ANN which can probably take into account quite small differences in allelic frequencies. Cornuet *et al.*, (1996) have already obtained good results in the classification of certain bee (*Apis mellifera*) lineages with microsatellite data and ANN.

4.1. Efficiency and utilisation of artificial neural networks

Homoplasmy does not appear to drastically reduce the learning capacities of the neural network. For rare alleles, the graph showing the contributions of the different alleles according to their frequencies indicates that the most informative alleles are also often quite frequent; nevertheless, less frequent alleles may provide more information in some comparisons. In all cases, learning appears to be able to recognise the most discriminant information (for a particular comparison) among all the input variables, and this technique does not require any particular adaptation of the data. Neural networks gave some better results than classical discriminant analysis, as is shown by Cornuet *et al.* (1996).

For the first analysis (ancestral/modern comparison), the application of the network to populations other than those used for learning gave

good results. The percentages of modern individuals predicted by the network agree well with the frequencies of modern alleles of *LDH-5**. The differences between these two parameters may be caused by different behaviour of the two markers, with randomly different introgression rates. The four supposed neutral microsatellite markers probably give a better description than a single allozymic (possibly selected) *LDH-5** marker. Moreover, one should keep in mind that these are a different type of information (allelic frequencies versus percentage of individuals).

Caution was required in this study because of the sample characteristics. Some samples (especially river populations) are or might be heterogeneous. Wild and domestic individuals may be found in the same sample of some of the 'modern' populations. For this reason, it was decided to test the leave-one-out procedure. It gave good results for the first comparison (ancestral/modern), and was then used for other comparisons. The technique appears well suited for the study of heterogeneous samples.

With both the leave-one-out and the hold-out procedures, neural networks associated with microsatellites confirm the distinction between modern Atlantic (wild or domestic) and ancestral Atlantic trout, which had previously only been analysed using allozymes.

4.2. Application to hatchery strains

Hatchery samples are needed as reference for assessing the proportion of domestic individuals in rivers. Analysis of these strains is necessary to evaluate the genetic diversity of domestic fishes; this shows whether the domestic samples analysed can be considered as representative of those used for stocking or if there is too much variability among hatcheries. Several studies have shown that these domestic strains were genetically quite similar (Guyomard, 1989; Garcia-Marin *et al.*, 1991), but we tried to verify this assumption using microsatellites and ANN. The high number of incorrectly classified individuals (both in the leave-one-out and hold-out results) underlines this homogeneity. The lack of differences may prevent good learning. It also shows that ANN can indicate when there is not enough differentiation between the categories used for learning as the network will not always give good percentages of correctly classified individuals, whatever is presented for learning. In our study, this homogeneity of domestic samples enabled us to pool them for the next analyses.

4.3. Characterisation of wild modern populations

Discriminating between wild and domestic modern Atlantic trout is an important objective. The identification of populations not or almost-not affected by stocking is useful for the protection and management of the genetic diversity of this species as this is threatened by stocking (Ferguson *et al.*, 1995). The use of the leave-one-out procedure for the comparison of each of the modern populations with hatchery populations gave low percentages of domestic individuals (from 5 to 8%) within the three modern populations (Chiroulet, Oussouet and Luz) which seemed to be mainly wild according to the morphological characteristics of their fishes. This would tend to show that these rivers are only modified by stocking slightly or not at all. Apart from this practical aspect, these results also show that neural networks are efficient even for genetically quite similar (but differentiated) entities.

The comparison of other samples with hatchery populations did not always give such clear results. For example, a large number of Bastan individuals had intermediate scores. This is probably linked with microsatellite properties and shared alleles, which in this case required more time for learning. However, individuals with intermediate scores could also be hybrid individuals and this raises the problem of how they are classified by the network. For example, in the Marcadau population (known to be heavily stocked), 40% of individuals displayed intermediate scores. This is probably the consequences of hybridisation of wild and domestic fishes; the strong impact of stocking on this population is confirmed by the percentage of individuals assigned to the domestic type (60%). This shows that when such individuals are present in a river population, the network is able to recognise them. There may be some hybrids in the Bastan population, (F1 or individuals resulting from backcrosses) even if allozymes indicate that it is a pure ancestral population; effectively, different markers can give different results because of selection and genetic drift. Moreover, as has already been explained, the training procedure may also cause this high proportion of intermediate scores. It should be noted that hardly any individuals in this population are clearly classified in the domestic category, as would have been expected in case of a high stocking impact (e.g. Marcadau). This population is probably not highly introgressed by domestic alleles.

Although the interpretation of these results is not as clear as for the former analyses, ANNs provided important information about the genetic composition of these populations.

5. Conclusion

From a technical point of view, our results confirm that ANNs are well suited to population genetics data. Effective analysis requires reference populations well chosen for the study, relatively balanced sample sizes and an appropriate validation procedure (hold-out or leave-one-out). For example, the leave-one-out procedure seems well

suited for mixed populations whereas the hold-out procedure gives a more precise idea of the prediction capability of the model. From a more fundamental point of view, this study confirms the presence in this area of several trout forms: two wild types (modern and ancestral) and one domestic form, which can coexist in the same river. Moreover, we identified certain pure or almost-pure wild populations. This raises the problem of their management and protection and is a new example of low stocking effectiveness. It is also an example of practical application of ANNs in ecology and population genetics.

Acknowledgements

This research was supported by the Bureau des Ressources Génétiques (grant No. 95011), the Conseil Supérieur de la Pêche (grant No. 9507127) and the Club Halieutique Interdépartemental. The field captures were performed by the local Fédérations de Pêche kindly assisted by scientists and students from ENSAT (Toulouse) and volunteers from Montpellier II University.

References

- Aurelle, D., Berrebi, P., 1998. Microsatellite markers and management of brown trout *Salmo trutta fario* populations in south-western France. *Génétique, Sélection, Evolution* 30, S75–S90.
- Cornuet, J.M., Aulagnier, S., Lek, S., Franck, P., Solignac, M., 1996. Classifying individuals among infra-specific taxa using microsatellites data and neural networks. *C. R. Acad. Sci. Paris, Life sciences* 319, 1167–1177.
- Estoup, A., Presa, P., Krieg, F., Vaiman, D., Guyomard, R., 1993. CTn and (GT)n microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). *Journal of the Genetical Society of Great Britain* 71, 488–496.
- Ferguson, A., Taggart, J.B., Prodöhl, P.A., MacMeel, O., Thompson, C., 1995. The application of molecular markers to the study and conservation of fish populations, with special reference to *Salmo*. *Journal of Fish Biology* 47, 103–126.
- Garcia-Marin, 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.
- Garson, G.D., 1991. Interpreting neural-network connection weights. *Artificial Intelligence Expert* 6, 47–51.
- Goh, A.T.C., 1995. Back-propagation neural networks for modelling complex systems. *Artificial Intelligence Engineering* 9, 143–151.
- Guégan, J.F., Lek, S., Oberdorff, T., 1998. Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature* 391, 382–384.
- Guyomard, R., 1989. Diversité génétique de la truite commune. *Bulletin Francais de Pêche et Pisciculture* 314, 118–135.
- Hamilton, K.E., Ferguson, A., Taggart, J.B., Tomasson, T., Walker, A., Fahy, E., 1989. Post-glacial colonisation of brown trout, *Salmo trutta*: *Ldh-5* as a phylogeographic marker locus. *Journal of Fish Biology* 35, 651–664.
- Jarne, P., Lagoda, P.J.L., 1996. Microsatellites, from molecules to populations and back. *Tree* 11, 424–428.
- Kohavi R., 1995. A study of cross-validation and bootstrap for estimation and model selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. Morgan Kaufmann Publishers Inc., 1137–1143.
- Lek, S., Belaud, A., Baran, P., Dimopoulos, I., Delacoste, M., 1996a. Role of some environmental variables in trout abundance models using neural networks. *Aquatic Living Resource* 9, 23–29.
- Lek, S., Delacoste, M., Baran, P., Lauga, J., Aulagnier, S., 1996b. Application of neural networks to modelling non-linear relationships in ecology. *Ecological Modelling* 90, 39–52.
- Mastrorillo, S., Lek, S., Dauba, F., Belaud, A., 1997. The use of artificial neural networks to predict the presence of small-bodied fish in river. *Freshwater Biology* 38, 237–246.
- Poteaux, C., 1995. Interactions génétiques entre formes sauvages et formes domestiques chez la truite commune (*Salmo trutta fario* L.), Thesis of Université Montpellier II, France, 110 pp.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating error. *Nature* 323, 533–536.