Article

Validation of *forward-in-time* method using artificial neural network: the application in a biological system

Abdurahim Kalajdzic[1], **Samim Konjicija**[2], **Belma Kalamujic Stroil**[1], **Naris Pojskic**[1]

[1]Institute for Genetic Engineering and Biotechnology, University of Sarajevo, Zmaja od Bosne 8 (Kampus), Sarajevo 71000, Bosnia and Herzegovina

[2]Faculty of Electrical Engineering, University of Sarajevo, Zmaja od Bosne bb (Kampus), Sarajevo 71000, Bosnia and Herzegovina

E-mail: abdurahim.kalajdzic@ingeb.unsa.ba

Received 23 October 2020; *Accepted 25 November 2020*; *Published 1 March 2021*

Abstract

Simulation studies in population genetics play a crucial role in better understanding of different evolution scenarios and effects of different genetic models on genetic diversity. *Forward-in-time* method starts with an initial population and follows the entire evolution under various genetic models within multiple generations. Artificial neural networks represent a formidable method for genetic simulation and prediction. In this study, we wanted to compare and corroborate results obtained with *forward-in-time* simulation with results attained from a specially designed strategy based on artificial neural networking. As input data, alleles of 13 microsatellite loci from 187 specimens representing autochthonous Adriatic haplotype of *Salmo trutta* L. from the Neretva River were used. The main goal of this study was to compare precision and reliability of these two methods. Our results are in concordance with other reports from literature which indicate that both of these approaches can be used as a reliable simulation tools. However, it is believed that artificial neural networks can represent more powerful simulation tools.

Keywords forward-in-time; artificial neural network; simulation; conservation.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Unavailability of ancestral information and complexity of real genetic factors led to development of computational simulation methods used for testing hypothesis and studying evolutionary genetic parameters (Balloux and Goudet, 2002). Computer simulators are in fact programs that were developed for the purpose of simulating abstract models of certain systems. In the last decade, these simulation programs became available to a wider circle of users thus allowing scientists that are not specialized in software programming to analyze evolution and conservation.Genetic simulation methods use two main principles: *backward-in-time*

(coalescent) (Kingman, 1982) and *forward-in-time*. *Backward-in-time* method starts from a sample of unknown genotype. This method identifies common ancestors of individuals and connects them based on stochastic processes which are characterized by evolutionary parameters such as mutation, recombination and migration. After a common ancestor is established, process goes forward and assigns genetic information to individuals. *Forward-in-time* method is much simpler and is based on individuals. Each individual in simulated population goes through a life cycle and this method starts with initial population and follows its evolution generation by generation. Since it follows each individual, this method is considerably slower; however modeling is much more complex which makes it better suitable for predictive questions. Also, one of the main differences between these two methods is that *forward-in-time* method requires defining initial parameters (Cyran and Myszor, 2008).

Application of Artificial Neural Network (ANN) in population genetics represents one of the most interesting areas in regards to simulation and prediction of different type of data. Artificial neural networks with multiple layers have many applications and, at the same time, represent general frames for presenting non-linear functional mapping. First layer of nodes is marked as an input layer and the last one as an output layer. Layers between these two are marked as hidden layers. Number of neurons in input and output layers is usually determined by application. Number of neurons in hidden layers, as well as the number of hidden layers, depends on a specific task of an ANN, and is usually determined during the training and validation process. These hidden layers enable the network to learn complex information. Number of hidden neurons is very significant because small number of these neurons prevents regular training and huge number leads to an exaggerated learning process. There are three possible ways of learning process: *supervised* learning which provides networks a correct output for every input pattern; *unsupervised* learning that does not require correct answer connected to each input parameter (the network itself explores patterns and correlation between data); *hybrid* (or *semi-supervised*) represents a combination between the two above mentioned methods. It is also important to point out that there are two principles related to neural networks: *feed-forward* and *back-propagation*. First of these two principles allows information to travel in only one direction, from input to output. *Back-propagation* is used for training the multi-layer *feed-forward* networks in which case every error calculated at the network's outputs is propagated backwards through layers, from neurons of output layer to neurons of input layer. Exactly this method was used in our study. Also one of the most important elements in training networks such as this one is regularization, performed in order of avoiding *overfitting*, which finally secures appropriate performance of the network itself (Krogh, 2008). Artificial neural networks have many applications in analyzing different biological data. Nurul Amerah and Afnizanfaizal (2017) applied neural networks in classification of DNA sequence. Also, Kang et al. (2017) developed a neural network that predicts phenotypes from transcriptome data. Finally, Aurelle et al. (1999) created an artificial neural network that uses microsatellite data as input to successfully separate contemporary and ancestral populations of different forms of brown trout. They used *feed-forward* network with one hidden layer and *back-propagation* algorithm for training. Results of the learning process itself gave high values of well-classified individuals (up to 95%), proving that artificial neural networks can be used in validating genetic data (Aurelle et al., 1999). Main goal of this study was to perform validation of *forward-in-time* population using ANN, based on microsatellite data from autochthonous Adriatic *Salmo trutta* L. from the Neretva River.

2 Material and Methods

Input data used in this research were retrieved from REBIDA (*REgional BIodiversity DAtabase*) database (Kalamujić Stroil et al., 2017). Model organism in this study was Neretvan (Adriatic) autochthonous brown

trout (*Salmo trutta* L.) (187 individuals with Adriatic haplotype from upper and middle flow of the Neretva River). EASYPOP software was used for *forward-in-time* simulation (Balloux, 2001). First step in this simulation method is to provide the software with some initial criteria. These criteria were: one population of 10,000 diploid individuals with equal number of males and females; free exchange of genetic material; *Stepwise Mutation Model* (SMM); mutation rate was set at zero due to a possible questionable validity of determining mutation rate in artificial neural network; free recombination and 150 generations. It is important to point out that a number of possible allelic states was set for each locus separately on the basis of the highest detected number of alleles in the contemporary brown trout population. After the *forward-in-time* method was performed, GenAlEx v6.5 (Peakall and Smouse, 2012) was used to calculate allele frequencies which were further used for validation analysis.

For creating ANN script, R programming language was used (R Core Team, 2020). This script was based on *neuralnet* package (Fritsch et al., 2019) and the neural network was trained with *sag* (*sag – smallest absolute gradient*) algorithm which is a modified *grprop* (*globally convergent algorithm*). This algorithm is based on *backpropagation* model with a modified learning rate. ANN structure (number of layers and neurons) and *threshold* values were set for each locus individually. This was done because of the fact that for loci with higher number of allelic variants, simpler ANN structure and higher *threshold* value was used and for loci with smaller number of allelic variants, more complex ANN structure and lower *threshold* value was used (Table 1).

Table 1 ANN structure and *threshold* value for each locus individually

Locus	Number of layers	Number of neurons	*Threshold* values
SsaD190	1	20	$1*10^{-5}$
SsaD71	2	10;10	$1*10^{-5}$
SSsp2213	1	20	$1*10^{-7}$
OMM1064	1	20	$1*10^{-7}$
Ssa85	2	20;10	$1*10^{-7}$
Str73	2	20;10	$1*10^{-9}$
Ssa410	1	20	$1*10^{-6}$
Str60	2	20;10	$1*10^{-7}$
Ssa408	2	10;10	$1*10^{-5}$
SsoSL438	2	10;10	$1*10^{-8}$
SSsp2216	1	20	$1*10^{-5}$
Str15	2	10;10	$1*10^{-7}$

Trained data were allele frequencies of contemporary population and test data were allele frequencies of *forward-in-time* simulation population. For comparison of predicted allele frequencies between *forward-in-time* and ANN we used *Chi-squared* and *Fisher exact* test. For the same purpose, in a case where results were treated as a ranking data, *Wilcoxon signed rank* test (*paired* test) and *Kendall-Tau correlation coefficient* were used. Special aspect of our comparison analysis was calculating the proportion of correctly predicted allele frequencies of ANN compared to *forward-in-time*. In this case, we used *equality-proportion* test with correction (*Chi-squared* test), with which we determined if there were any statistically significant differences in proportions between correctly and incorrectly predicted values. For all of these five analyses,

the level of statistical significance was $p<0.05$ and they were all implemented in our ANN R script. If there were a statistically significant higher proportion of adequately predicted allele frequencies, the ANN was considered to have confirmed the *forward-in-time* results.

Since the created *neural net* script gives values larger than one, relative allele frequencies were multiplied by 100 in order of obtaining integers. Values were rounded to one for those alleles which frequencies were below one after a given approach. Prediction was considered accurate in two cases: 1) if the neural network predicted the same frequency as *forward-in-time* method; 2) if difference of +/-1 was noticed, considering the fact of rounding the predictive values by the neural network. Difference of two and more was considered inadequate. Therefore, in Fig. 2, values of correct prediction were included within the lines. All values outside these lines were predicted incorrectly.

3 Results

Forward-in-time method was performed using parameters which were described in detail in the Materials and Methods section. After given analysis, allele frequencies (Table 2) were obtained using GenAlEx v6.5 software (Peakall and Smouse, 2012). This was done because of the fact that these frequencies were used for validating the accuracy of *forward-in-time* method with artificial neural network.

Table 2 Allele frequencies of *forward-in-time* simulation population (N- sample size).

Allele	SsaD190	SsaD71	SSsp2213	OMM1064	Ssa85	Str73	Ssa410	Str60	Ssa408	SsoSL438	SSsp2216	Str15
N	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000
1	0.056	0.040	0.060	0.014	0.223	0.292	0.022	0.297	0.030	0.227	0.036	0.127
2	0.081	0.071	0.018	0.008	0.220	0.381	0.041	0.382	0.024	0.104	0.043	0.120
3	0.050	0.043	0.041	0.029	0.142	0.327	0.030	0.321	0.047	0.191	0.066	0.172
4	0.103	0.013	0.039	0.031	0.163		0.070		0.033	0.218	0.033	0.127
5	0.044	0.052	0.029	0.032	0.252		0.060		0.019	0.260	0.065	0.151
6	0.049	0.052	0.015	0.014			0.043		0.063		0.054	0.164
7	0.026	0.044	0.076	0.029			0.054		0.060		0.078	0.139
8	0.041	0.039	0.049	0.010			0.048		0.063		0.049	
9	0.081	0.107	0.029	0.012			0.021		0.040		0.035	
10	0.120	0.012	0.083	0.029			0.022		0.079		0.034	
11	0.037	0.034	0.048	0.044			0.043		0.047		0.063	
12	0.084	0.049	0.023	0.037			0.048		0.051		0.040	
13	0.059	0.044	0.041	0.044			0.030		0.021		0.067	
14	0.050	0.053	0.041	0.002			0.050		0.034		0.050	
15	0.031	0.050	0.027	0.024			0.078		0.035		0.030	
16	0.085	0.058	0.021	0.011			0.067		0.047		0.050	
17		0.059	0.059	0.034			0.080		0.040		0.046	
18		0.067	0.037	0.055			0.035		0.055		0.039	
19		0.006	0.051	0.004			0.026		0.078		0.067	
20		0.054	0.025	0.016			0.037		0.017		0.055	
21		0.053	0.041	0.022			0.030		0.023			
22			0.032	0.046			0.066		0.051			
23			0.059	0.022					0.044			

24	0.058	0.037
25		0.022
26		0.011
27		0.024
28		0.046
29		0.042
30		0.038
31		0.016
32		0.019
33		0.010
34		0.033
35		0.018
36		0.023
37		0.013
38		0.005
39		0.028
40		0.021
41		0.027

In order to check the accuracy of the *forward-in-time* simulation we created an artificial neural network, which we thoroughly described in section Material and Methods. The structure of ANN for each locus and comparison of the allele frequencies results of *forward-in-time* method and the predicted values of the ANN are given in Figs 1-2 and Table 3.

As for locus SsaD190, 16 alleles were detected with ANN. The reliability of prediction by an artificial neural network in this case can be clearly concluded because 75% of allele frequencies were adequately predicted. In a case of SsaD71 locus, 21 alleles were detected and the reliability of prediction can also be concluded because 85.71% of allele frequencies were adequately predicted. As for OMM1064 locus, 41 alleles were detected. In this case also, the reliability of prediction can be concluded because 85.36% of allele frequencies were adequately predicted. Five alleles were detected within the Ssa85 locus. Once again the reliability of prediction can be concluded due to the fact that 80% of allele frequencies were adequately predicted. *The equality – proportions* test in all of these cases showed that there were statistically significantly more accurate predictions than incorrect ones. Also, for all of these loci, the *Pearson Chi squared* test and the *Fisher exact* test showed no statistically significant differences in predicted allele frequencies between *forward-in-time* and ANN prediction. The *Kendall Tau* test in all of these cases showed a statistically significant correlation between the two analyzed methods, while the *Wilcoxon paired* test showed no statistically significant difference between the two predictions.

Three alleles were detected within Str73 and Str60 loci and five and seven alleles were detected within SsoSL438 and Str15 loci, respectively. The reliability of prediction by an artificial neural network in all of these cases can be clearly concluded because 100% of allele frequencies were adequately predicted. The *Pearson Chi squared* test and the *Fisher exact* test showed no statistically significant differences in predicted allele frequencies. The *Kendall Tau* test did not show a statistically significant correlation for Str73 and Str60. However, in a case of SsoSL438 and Str15 loci it did. *Wilcoxon paired* test showed no statistically significant difference in all of these loci.

In a case of Ssa410 locus, 22 alleles were detected. The reliability of prediction by an artificial neural network in this case can be clearly concluded because 63.63% of allele frequencies were adequately predicted. As for Ssa408 locus, 23 alleles were detected. In this case also, the reliability of prediction can be concluded because 78.26% of allele frequencies were adequately predicted. Twenty alleles were detected within SSsp2216 locus and the reliability of prediction can also be concluded because 80% of allele frequencies were adequately predicted. The *Pearson Chi squared* test and the *Fisher exact* test for these three loci showed no statistically significant differences in predicted allele frequencies. The *Kendall Tau* test in all of these cases showed a statistically significant correlation between the two analyzed methods, while the *Wilcoxon paired* test showed no statistically significant difference between the two predictions.

As for locus SSsp2213, 24 alleles were detected. It can be clearly concluded that in this case the absolute reliability of prediction by artificial neural networks cannot be claimed because 58.33% of allelic frequencies were adequately predicted. The *equality – proportions* test showed that there were no statistically significantly more accurately predicted frequencies than incorrect ones. The *Pearson Chi squared* test and the *Fisher exact* test showed no statistically significant differences in predicted allele frequencies between *forward-in-time* and ANN prediction, however the *Kendall Tau* test did not show a statistically significant correlation between the two analyzed methods. *Wilcoxon paired* test showed no statistically significant difference between the two predictions.

4 Discussion

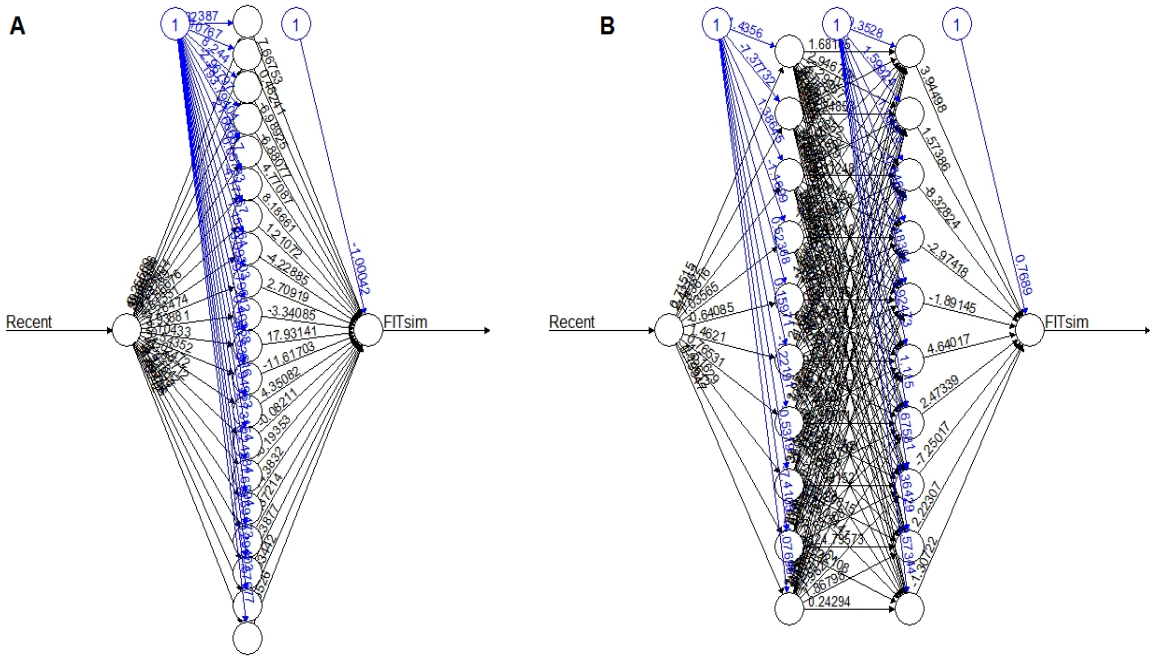
From the above results of allele frequency predictions using neural networks in relation to *forward-in-time* simulations, we can clearly conclude a very strong match of allele frequency values between the two simulation methods. For certain loci such as Str73, Str60, Str15, SsoSL438 the predictions were fully consistent with *forward-in-time* simulations. For the SsaD190, SsaD71, OMM1064, Ssa85, Ssa410, Ssa408 and SSsp2216, the predictions were almost completely correct. Even though OMM1064 has 41 alleles, only 14.63% of frequencies were not adequately predicted. The worst prediction was for locus SSsp2213 where 58.33% was true and 41.66% was false, so the adequacy of prediction accuracy in this case is questionable. It should be borne in mind that different adjustments of threshold, number of neurons and number of layers were used in order to achieve optimal prediction (Table 1). The only common parameter was the number of iterations. In the case of optimization for the SSsp2213 locus, the application of various adjustments showed no change in prediction. In the case of loci with 100% accuracy, the change of parameters did not change the results, and for other loci it was insignificant, so the most favorable parameters were taken.

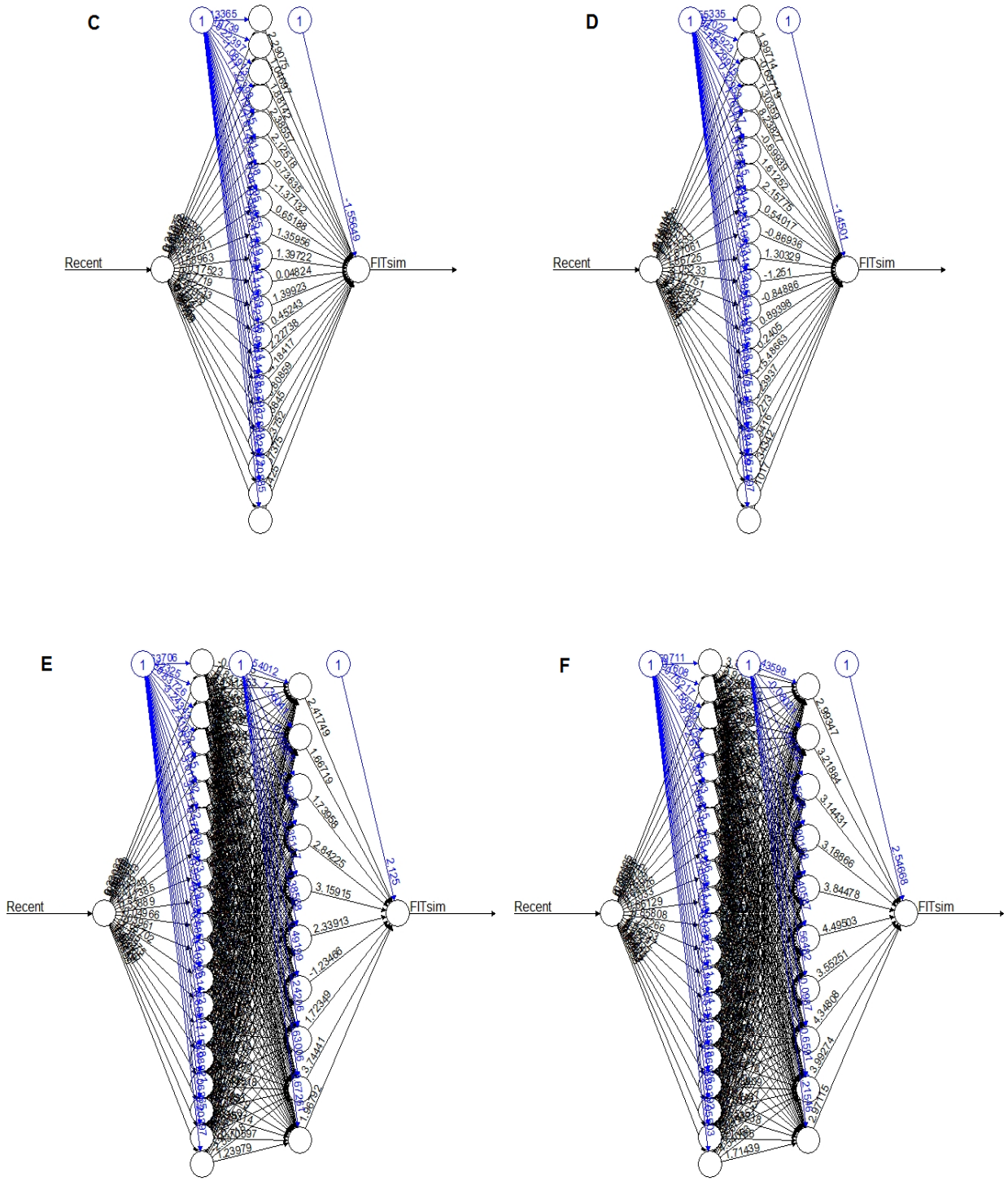
The application of a neural network, where the trained data would be allele frequency values within the contemporary population and test data allele frequency values of the *forward-in-time* population, would provide a possible answer to the question of adequacy of *forward-in-time* simulation. We can safely conclude that an artificial neural network is an adequate method in predicting allele frequencies based on training and test data. Barbosa et al. (2011) determined the importance of neural network optimization especially in the case of inadequate estimation of threshold, number of iterations as well as number of neurons and layers. In our case, the neural network aimed to confirm or deny the *forward-in-time* simulation and thus determine the validity of this method in relation to the starting population.

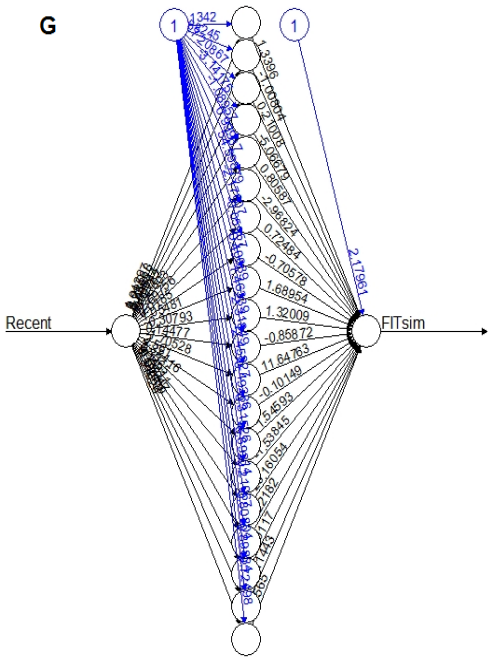
The genetic aspect of these simulations is multiple. Kalamujić (2013) and Pojskić and Kalamujić (2015) point to the fact of rapid reduction of the Adriatic (autochthonous) haplotype of the Neretva brown trout, due to various factors such as: inadequate restocking (restocking with allochthonous Danube and Atlantic haplotypes), overfishing, interspecies hybridization (with soft-mouthed trout (*Salmo obtusirostris*, Heckel) and marble trout (*Salmo marmoratus*, Cuvier)) and anthropogenic factor (building dams that prevent gene

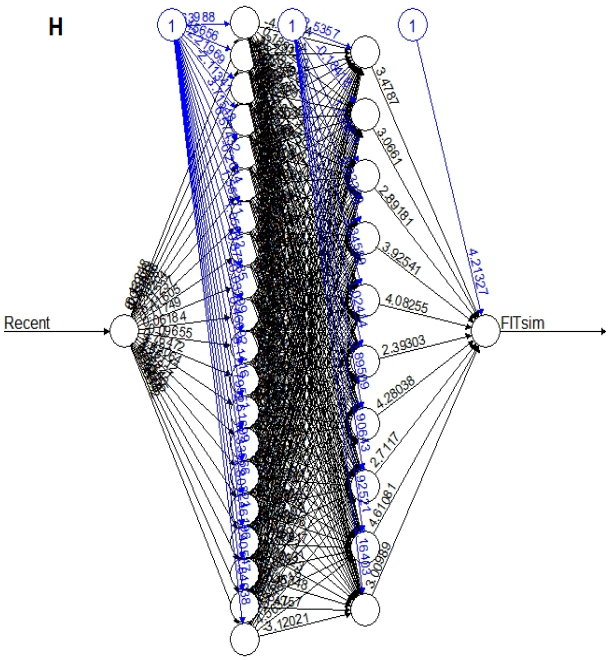
flow). *Forward-in-time* simulation clearly shows that even if the starting parameters of a contemporary population with high inbreeding, reduced observed heterozygosity, but with a sufficient number of alleles are applied, it can show the "appearance" and structure of a hypothetical population that is devoid of greater anthropogenic influence. Given that the artificial neural network has largely confirmed the *forward-in-time* simulation, when it comes to allele frequencies, the conclusions of Kalamujić (2013) about the great influence of anthropogenic factors on the reduction of autochthonous Neretva brown trout can be reasonably confirmed. The application of artificial neural network is multiple. In a study by Pojskić et al. (2018), an algorithm for prediction of missing variants in the Y-STR profile of archaeological human skeletal remains was created. According to the authors, the predictability is high on the test data, and the size of the training base is very important because the network can better understand the patterns. Their results also indicate the fact that the number of neurons and layers is very important in optimizing the prediction itself.

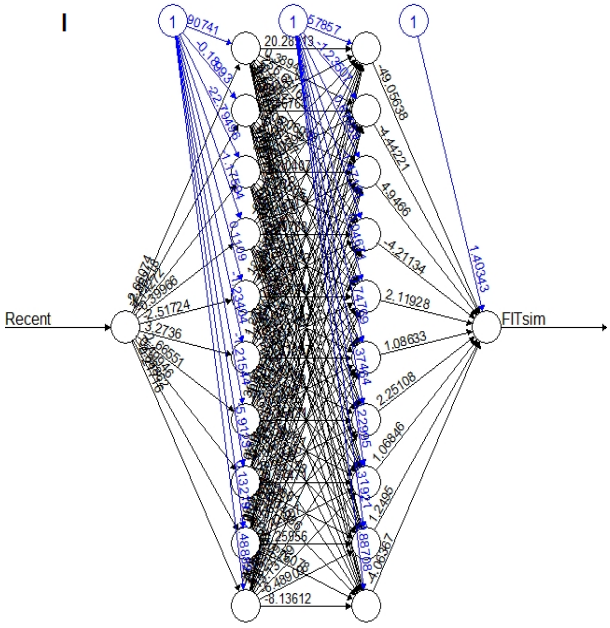
In any case, different approaches should be applied in assessing the genetic heterogeneity and differentiation of populations, whether they are plant, animal or even human populations. Testing the validity of these results using ANN certainly contributes to the strength of the conclusions derived from classical population-genomic analyses. However, the use of neural networks alone cannot answer all the questions, and therefore the best approach would be the use of classical and modern population-genomic analyses with additional analyses using artificial neural networks.




C
Recent
FITsim
D
Recent
FITsim
E
Recent
FITsim
F
Recent
FITsim


G
Recent
FITsim


H
Recent
FITsim


I
Recent
FITsim

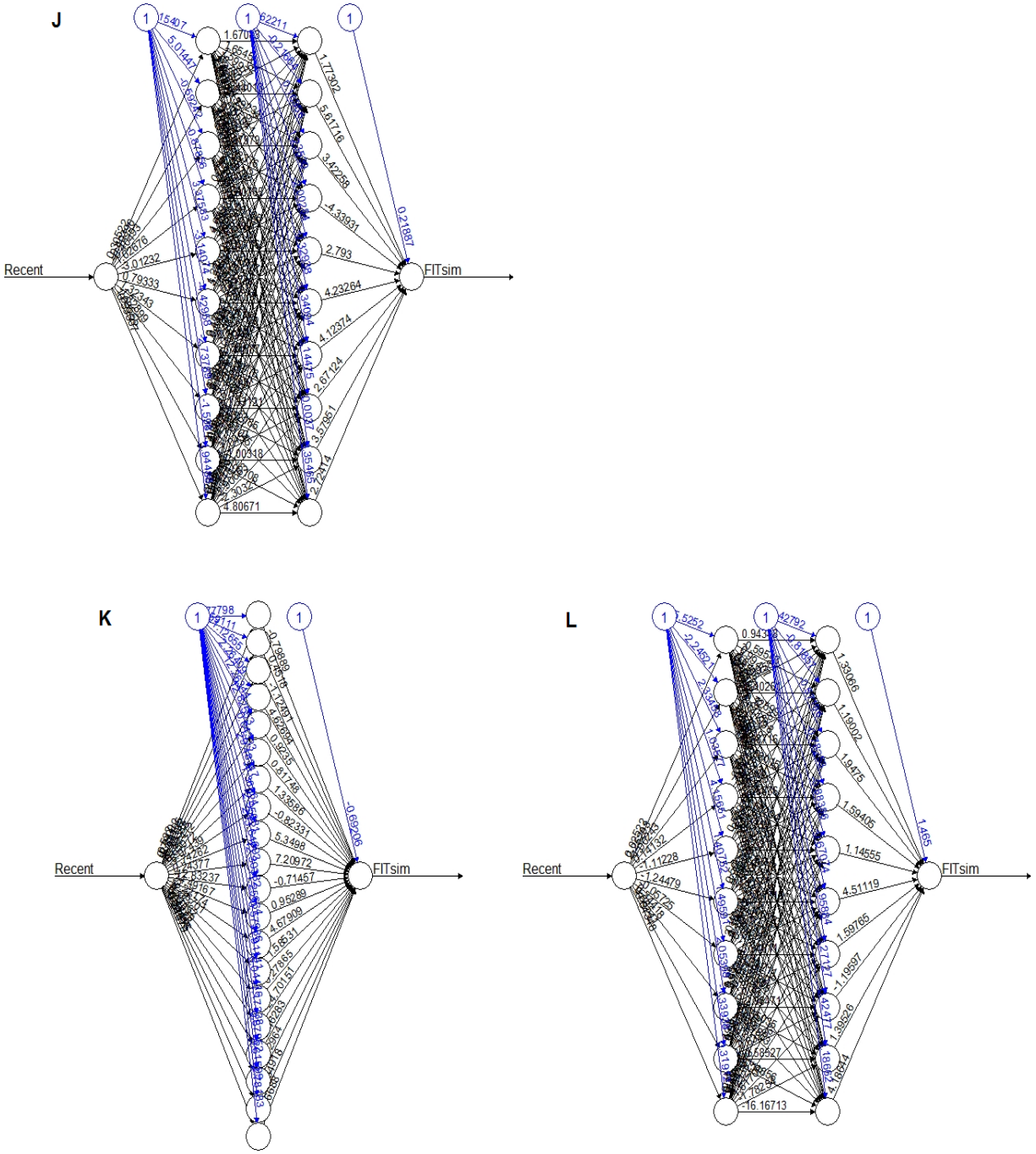
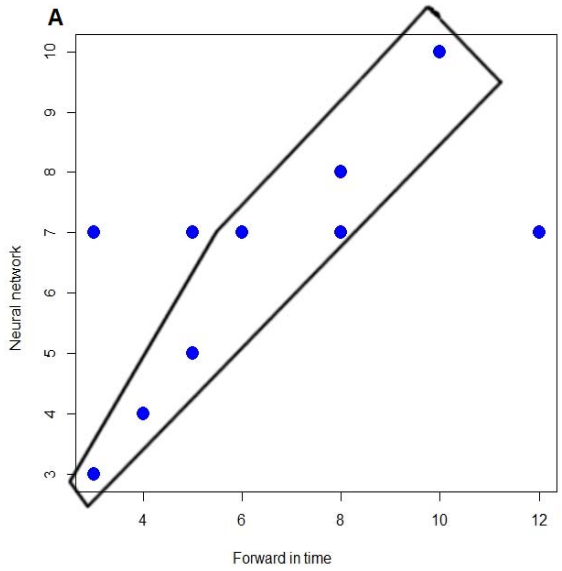
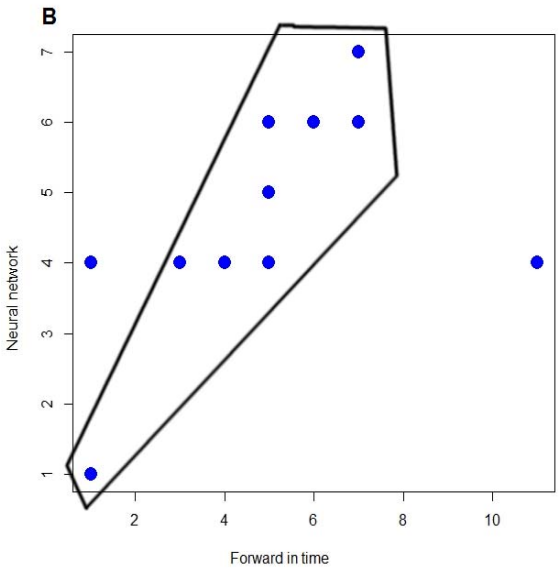
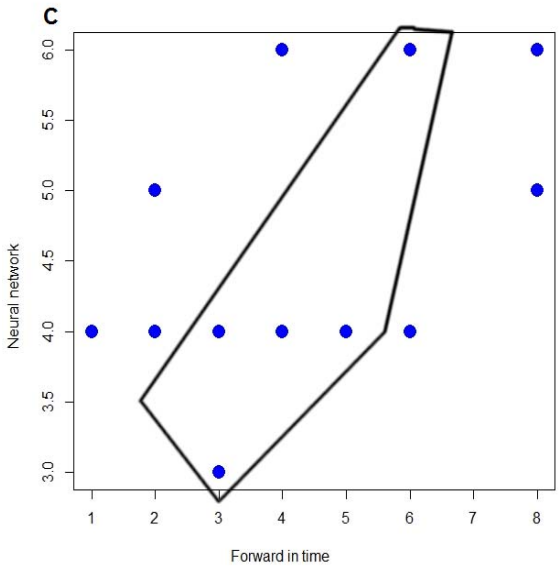
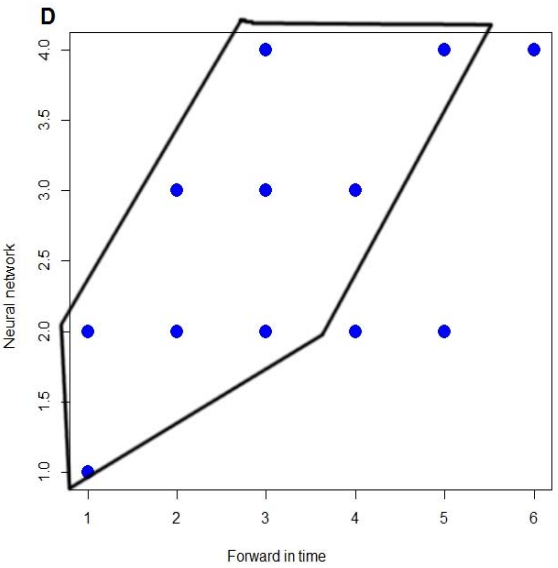


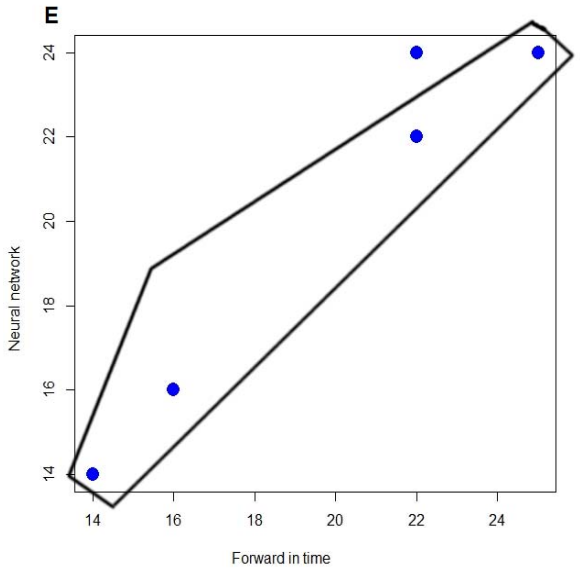
Fig. 1 Structure of the ANN network for each locus; A - SsaD190; B - SsaD71; C - SSsp2213; D - OMM1064; E - Ssa85; F - Str73; G - Ssa410; H - Str60; I - Ssa408; J - SsoSL438; K - SSsp2216; L - Str15.

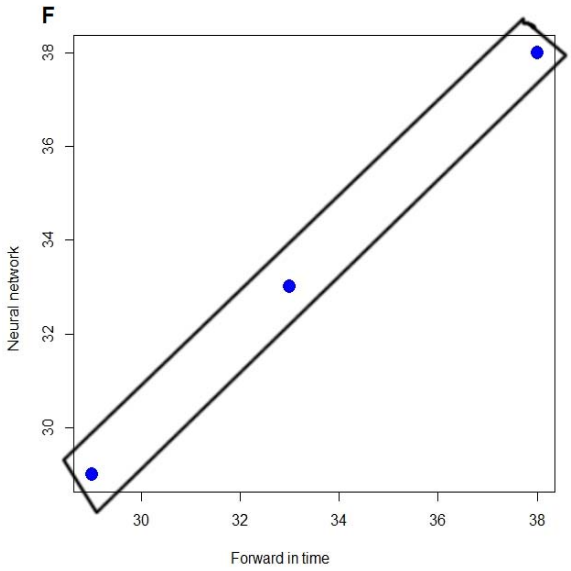

A
Neural network
Forward in time
3
4
5
6
7
8
9
10
4
6
8
10
12

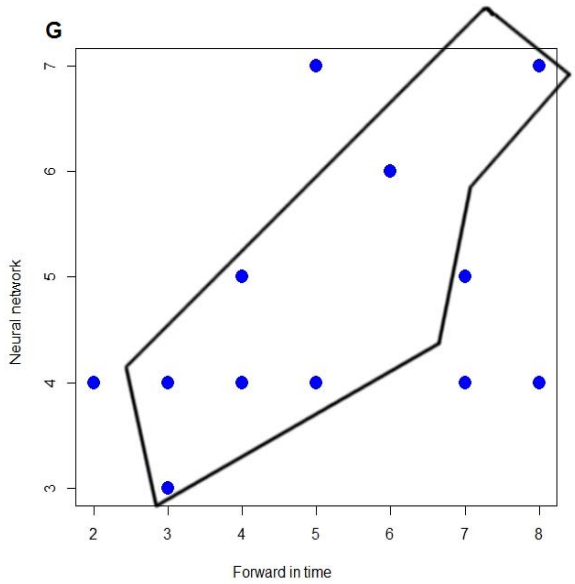

B
Neural network
Forward in time
1
2
3
4
5
6
7
2
4
6
8
10

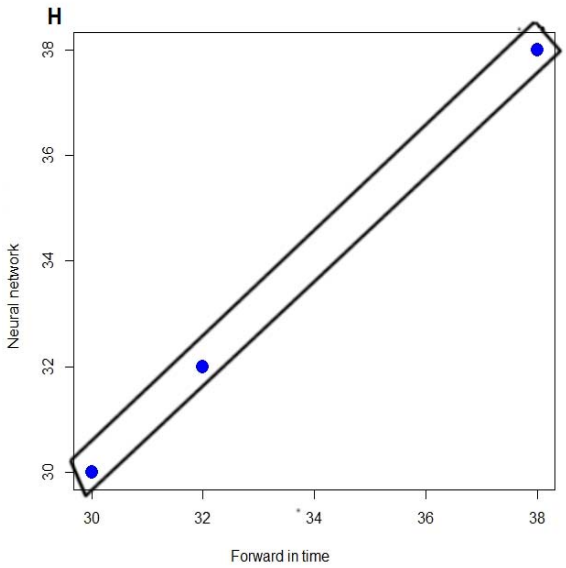

C
Neural network
Forward in time
3.0
3.5
4.0
4.5
5.0
5.5
6.0
1
2
3
4
5
6
7
8


D
Neural network
Forward in time
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1
2
3
4
5
6


E
Neural network
Forward in time
14
16
18
20
22
24


F
Neural network
Forward in time
30
32
34
36
38


G
Neural network
Forward in time
2
3
4
5
6
7
8


H
Neural network
Forward in time
30
32
34
36
38

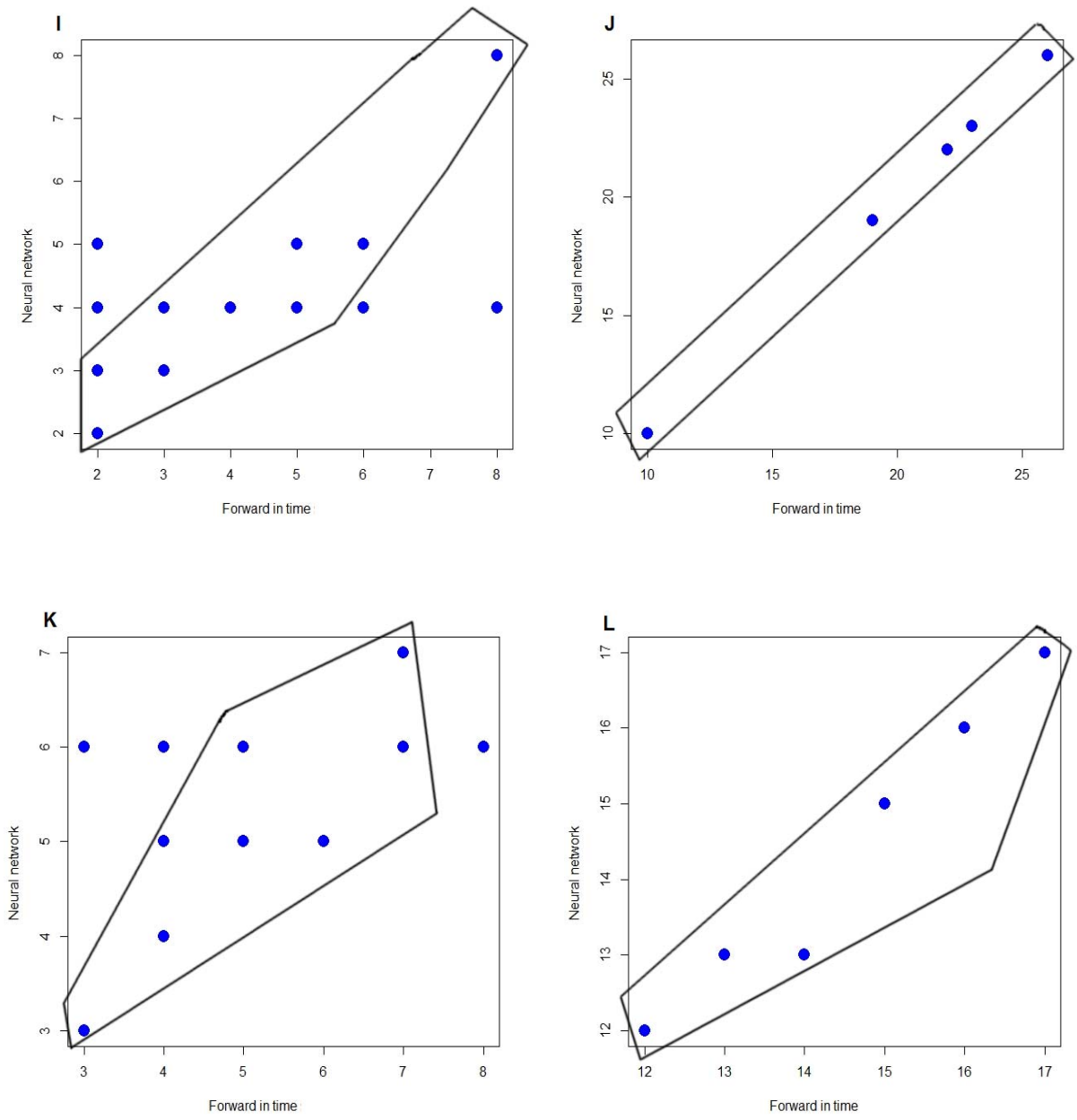


Fig. 2 Comparison of the results of *forward-in-time* method and the predicted values of the ANN for each locus (points within the marked area - accurate prediction; points outside the marked area - inaccurate prediction); A - SsaD190; B - SsaD71; C - SSsp2213; D - OMM1064; E - Ssa85; F - Str73; G - Ssa410; H - Str60; I - Ssa408; J - SsoSL438; K - SSsp2216; L - Str15

Table 3 ANN prediction of allele frequencies compared to *forward-in-time* method (statistical parameters; *Wilcoxon signed rank* test (*paired* test) and *Kendall-Tau* correlation coefficient were used for comparing predicted allele frequencies when results were treated as ranking data)

		SsaD190	**SsaD71**	**SSsp2213**	**OMM1064**	**Ssa85**	**Str73**	**Ssa410**	**Str60**	**Ssa408**	**SsoSL438**	**SSsp2216**	**Str15**
Pearson's Chi-squared	χ^2	3.8257	7.5417	7.883	9.6548	0.10234	0	6.3403	0	5.4443	0	2.6709	0.032013
	df	15	20	23	40	4	2	21	2	22	4	19	6
	p-value	0.9983	0.9945	0.9986	1	0.9987	1	0.9991	1	0.9999	1	1	1
Fisher's Exact	p-value	0.9985	0.9956	0.9910	1	0.9994	1	0.9994	1	0.9999	1	1	1
Kendall Tau	p-value	0.0009296	0.0007219	0.09429	0.001317	0.03736	0.333	0.01716	0.3333	0.005547	0.01667	0.00593	0.004333
	z-value	3.311	3.3811	1.6732	3.2123	2.0818	3	2.3833	3	2.7735	10	2.7516	2.8529
	Tau -value	0.6937141	0.6373797	0.2971037	0.4406182	0.8888889	1	0.4331685	1	0.4929593	1	0.5279306	0.9486833
Wilcoxon paired test	p-value	0.523	0.8322	0.8473	0.2588	1	NA‡	0.9644	NA‡	0.8969	NA‡	0.3936	1
	v-value	13	30	72	267	1	0	84	0	55	0	28	1
Equality proportion test	χ^2	6.125	18.667	0.75	38.244	1.6	2.6667	2.2727	2.6667	12.522	6.4	12.1	10.826
	df	1	1	1	1	1	1	1	1	1	1	1	1
	p-value	0.01333	$1.557*10^{-5}$	0.3865	$6.243*10^{-10}$	0.2059;	0.1025	0.1317	0.1025	0.0004022	0.01141	0.0005042	0.001341
	%†	75%;25%	85.71%;14.29%	58.33%;41.67%	85.36%;14.64%	80%;20%	100%	63.63%;36.37%	100%	78.26%;21.74%	100%	80%;20%	100%

†Proportion of adequately; inadequately predicted allele frequencies; ‡- not applicable

5 Conclusions

Artificial neural network is a good method in assessing the validity of a *forward-in-time* simulation. The selection of optimizing parameters for ANN is essential for the application of an artificial neural network in allele frequency estimates.

References

Aurelle D, Lek S, Giraudel JL, Berrebi P. 1999. Microsatellites and artificial neural networks: tools for the discrimination between natural and hatchery brown trout (*Salmo trutta*, L.) in Atlantic populations. Ecological Modelling, 120: 313-324

Balloux F. 2001. EASYPOP (Version 1.7): A Computer Program for Population Genetics Simulations. Journal of Heredity, 92(3): 301-302

Balloux F, Goudet J. 2002. Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. Molecular Ecology, 11: 771-783

Barbosa CE, Viana AP, Quintal SSR, Pereira MG. 2011. Artificial neural network analysis of genetic diversity in *Carica papaya* L. Crop Breeding and Applied Biotechnology, 11:224-231

Cyran KA, Myszor D. 2008. Coalescent vs. Time-forward Simulations in the Problem of the Detection of Past Population Expansion. International Journal of Applied Mathematics and Informatics, 2(1): 10-17

Fritsch S, Guenther F, Wright M, Suling M, Mueller SM. 2019. Neuralnet: Training of Neural Networks. R package v1.44.2

Kalamujić B. 2013. Genetički diverzitet recentnih populacija Salmo trutta L. rijeke Neretve u odnosu na rekonstruisani simulacioni model početnog stanja. Doktorski rad. Prirodno-matematički fakultet, Univerzitet u Sarajevu, COBISS.BH-ID: 20817670 [in Bosnian]

Kalamujić Stroil B, Dorić S, Hanjalić J, Lasić L, Pojskić N. 2017. REgional BIodiversity DAtabase (REBIDA) – The First Comprehensive Database of Biological Diversity of Bosnia and Herzegovina. Genetics and Applications, 1(2): 59-65

Kang T, Ding W, Zhang L, Ziemek D, Zarringhalam K. 2017. A biological network - based regularized artificial neural network model for robust phenotype prediction from gene expression data. BMC Bioinformatics, 18: 565-576

Kingman JFC. 1982. The coalescent. Stochastic Processes and their Application, 13: 235-248

Krogh A. 2008. What are artificial neural networks? Computational Biology, 26(2): 195-197

Nurul Amerah K, Afnizanfaizal A. 2017. Classification of DNA Sequences Using Convolutional Neural Network Approach. UTM Comp Proceedings, 2: 12-18

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics, 28: 2537-2539

Pojskić N, Kalamujić B. 2015. Simulations based on molecular-genetic data in detection of expansion *Salmo trutta* allochtonous population in Neretva River's tributaries. In: Proceedings of 27th International Congress for Conservation Biology and 4th European Congress for Conservation Biology (Visconti P, Game E, Mathevet R, Wilkerson M, eds). 539-540, Montpellier, France

Pojskić N, Konjicija S, Osmanković D, Dorić S. 2018. Digital reconstruction of partial DNA profiles of human skeletal remains. Book of Abstracts: Belgrade Bioinformatics Conference 2018. Biologia Serbica, 40(1): 45

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.