New Artificial Neural Network based Test for the Detection of Past Population Expansion using Microsatellite Loci

Krzysztof A. Cyran, Dariusz Myszor

***Abstract*—**Detection of the past population growth is one of the crucial issues in contemporary population genetics. The importance of the problem is especially well understood in the context of neutral theory of evolution at molecular level proposed by Kimura. This theory often serves as neutral hypothesis in the search for genes which underwent natural selection. The conclusions in such studies can be false if population expansion was present but not detected and therefore not introduced into the model. In the paper we present novel statistical test which emerged from application of artificial neural networks theory. The test is designed to detect past population growths based on genetic microsatellite data. In experimental part of our research we created set of samples, using forward in time simulation methods. These samples were picked at random from simulated populations that had undergone growths of different types and intensities. Then, we created and trained series of different artificial neural networks and checked power of new tests based on these networks. We also compared powers of new tests with powers obtained by known methods based on microsatellites. Our studies showed that proposed by us new test provides better power in detection of population growth than the best currently available tests based on microsatellites i.e. Kimmel's and King's imbalance indices.

***Keywords*—** stochastic computer simulations, population growth detection tests, artificial neural networks, microsatellite loci, single step mutation model

I. Introduction

THIS is a well known fact that results of the search for the natural selection operating at molecular level are affected by population history. Therefore the estimation of the probable long-term demographic history of a population, and in particular, the detection of the past population growth has become one of the main problems in statistical genetics. In the last decade, with the advances of new numerical methods and the more and more productive computers the forward in time simulations started to play the role reserved earlier for coalescent methods.

On the other hand artificial neural networks have been successfully used for years in many scientifically sound problems. Neural networks have ability for adaptation and generalization of knowledge, and can find hidden patterns in input data by inductive machine learning process [1]. Therefore they might be successfully used in solving problems that are often hard to describe by rule-based algorithms. The crucial is point is only the availability the training data representing properly the problem considered.

In the studies dedicated for the detection of population growth the researchers often use various statistics computed for the same sample and then they try to analyze the results and draw conclusions [2]. Our goal was to create a method which would be able to encompass knowledge gained from a few statistics based on microsatellites.

Microsatellites are short tandem repeats, STRs [3-5]) which are quite abundant in genomes and undergo relatively fast mutations. Therefore they are suitable for testing the evolution of populations rather than species, and no doubt have found applications in various tests for population detection. Using such data we propose new statistical test that has a greater power for detection of population growth than other available microstallite based methods.

II. Computer Simulations

In order to create samples used as training data for artificial neural networks we applied time forward computer simulation. We used Wright-Fisher [6] model with provided dynamic description of the evolution and modified this model to allow changes between amounts of individuals in generations of population.

In time forward simulation method we must simulate creation of every generation from the second generation to the final generation. Number of individuals might vary between generations. Each member has constant amount of microsatellites. Members of each generation are based on members of the previous generation. During creation of generation's member we draw parent of the individual, take parent's microsatellites and run mutation process based on single-step stepwise mutation model (SSMM) [7] then we assign microsatellites to the child. First generation has N members, each member has M unlinked microsatellites, each

The scientific work financed by Ministry of Science and Higher Education in Poland from funds for supporting science in 2008-2010, as a research project number N N519 319035

K. A. Cyran is with the Institute of Informatics, Silesian University of Technology, Gliwice, 44-100 Poland, phone: +48-32-237-2500; fax: +48-32- -237-2733; e-mail: krzysztof.cyran@ polsl.pl.

D. Myszor is with the Institute of Informatics, Silesian University of Technology, Gliwice, 44-100 Poland, phone: +48-32-237-2500; fax: +48-32- -237-2733; e-mail: dariusz.myszor@ polsl.pl.

microsatellite has the same initial value. For first 2*N* to 4*N* generations we should simulate population of constant size [8] in order to achieve genetic drift – mutation equilibrium (we call it initializing generations) we don't use samples from this period in our researches. After this time we can simulate population growth.

Our initial population size was 2 500, each individual had $M = 30$ microsatellites and mutation rate was $v = 5\times10^{-4}$. Each simulation comprised at least $8N = 20\,000$ initializing generations (we choose rather conservative number 8*N* to be sure about achieving the equilibrium) before we start simulating growth.

We simulated two types of population growth:

a) exponential growth from $N = 2\,500$ individuals to 5 000, 25 000 and 250 000. For the same final size of population, rate of exponential growth varied because of different times of achieving final population size. We used up to 11 different time scales to simulate growths as fast as lasting only 625 generations to as slow as lasting even 640 000 generations. Unique connection of final population size and time of reaching final population size is described in this paper as a scenario.

b) stepwise growth from N = 2 500 individuals to 5 000, 25 000 and 250 000.

For each scenario of population growth (if not said different in experiment description) we created 100 independent histories. Because in real life we usually don't have all members of the generation, but just a few representatives we took 100 samples from every final generation. Each sample contains 40 individuals (one individual couldn't be found twice in one sample, but might be found in several different samples). From each individual we gained 30 unlinked microsatellites, and as was mentioned the mutation rate was set at $v = 5\text{x}10^{-4}$.

For ten fold stepwise growth from 2 500 to 25 000 individuals, we created a series of samples with different values of simulation parameters like: mutation rate (2.5×10^{-4}, 5×10^{-4}, 7.5×10^{-4}) amount of individual's microsatellites (10, 30, 40), amount of individuals in examined sample (10, 40, 70). Then we counted power of neural classifier for this cases and compared with power obtained by $\ln\hat{\beta}_1$ (for the definition of this coefficient see section 3.4) for the same samples.

III. Artificial Neural Networks

As it was said in the introduction, artificial neural networks can use knowledge obtained from several sources feeded to network inputs by learning from examples the strength of the influence of given input data source on the quantity under consideration. The description of formal operation of neural network which is relevant for the study id presented in following subsections

A. Architecture

Fundamental element of artificial neural network is neuron. Neuron is a cell with many inputs and one output. Each input value xk is multiplied by coefficient wjk signed to this input. There is additional input value x0 with fixed value (usually 1) and coefficient wj0 usually set at the beginning to -1.The artificial neuron output y is given by

$$y_j = \frac{1}{1+e^{-t}} \tag{1}$$

where

$$t = \sum_{k=0}^{i} w_{jk} x_k \tag{2}$$

Neurons in networks are organized in layers. Single layer may contain one or more neurons. In our researches we used one and two layers networks. In our networks, all outputs from previous layer are inputs for every neuron in next layer (feed-forward neural networks) all signals are always going forward (there is no recursion). Size of network (amount of layers and neurons) is significant, if network is too small it might be unable to achieve desired global error value during learning. Networks with too many neurons or layers might remember how to recognize all learning samples and lost generalization capabilities.

Inputs of first layer are fed by samples coefficients, values on the input should be normalized.

Network can have many neurons in last layer, each such neuron is one element of output vector. In our networks last layer contains always one neuron. We assume that 0 on the network output means that population size is constant and 1 that population experienced expansion.

Depending on activation function of the last layer we can get discrete or continuous value on the network output. In our studies we gained continuous values and then compared them with the cut off value.

B. Learning

To train our networks we used supervised learning: in loop, we presented learning samples on the network input and modified weights w_{jk} using the steepest descend method (back propagation) until we reached desired error value or number of iterations.

It is important to choose correct learning set. Samples in learning set should cover uniformly all cases. If some rule have significant more samples in learning set than other rules, it might have negative influence on the network, because during training network can learn how to recognize this predominant rule faster than other rules, therefore achieve low global error. But during using network wouldn't be able to recognize all other rules correctly.

Our learning set contains similar amount of samples from constant and from growing populations. Samples from growing populations came from populations that underwent stepwise growth and exponential growth.

We allowed our networks to learn for 100 000 iterations (in each learning session, we saved network better than previous

best one in that session). We performed in this way a few hundred sessions obtaining several hundred networks. From this set of networks we choose the one with the greatest power.

When the network learning was finished we calculated the cut off value as 0.95 quantile of the output values generated by network for histories with constant population size. Set of samples which we used to count power of classifiers contained neural network learning set, but this learning set was just little fraction (about 5%) of all samples set used in further studies.

C. Detection of Growth

In order to detect possible growth of the population we supplied the trained network with growth detection statistics described in the next Section.

These statistics are designed to detect different histories of samples drawn from populations with constant size (Fig. 1a) and those which underwent in past a substantial growth (Fig. 1b).

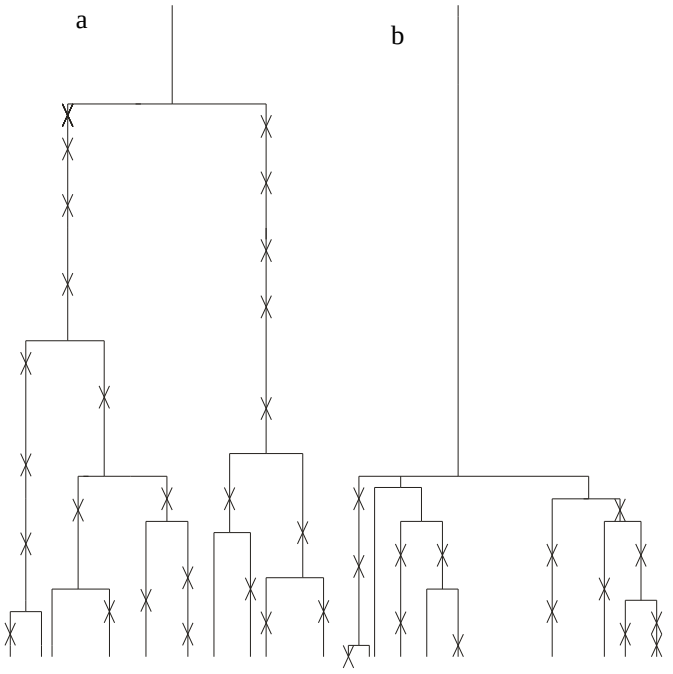


Fig. 1. Genealogy trees for genes of 10 individuals drawn at random from a population having constant size of 20 000 (a) and population of 20 000 individuals which however underwent 8 000 generations ago 100 fold growth. The crosses represent mutation events.

The qualitatively different lengths of the branches leading from the most recent common ancestor in both genealogies presented in Fig. 1 are the reason why the distributions of the length of alleles are also different. For the constant population size the old branches are long and therefore accumulate a lot of mutations what is reflected in two or three modal distributions of the allele length (Fig. 2). This is not so for the population evolving after significant growth. The corresponding genealogy has got short branches leading from the most common ancestor, so the mutations accumulate in young branches yielding unimodal distributions (Fig. 3).

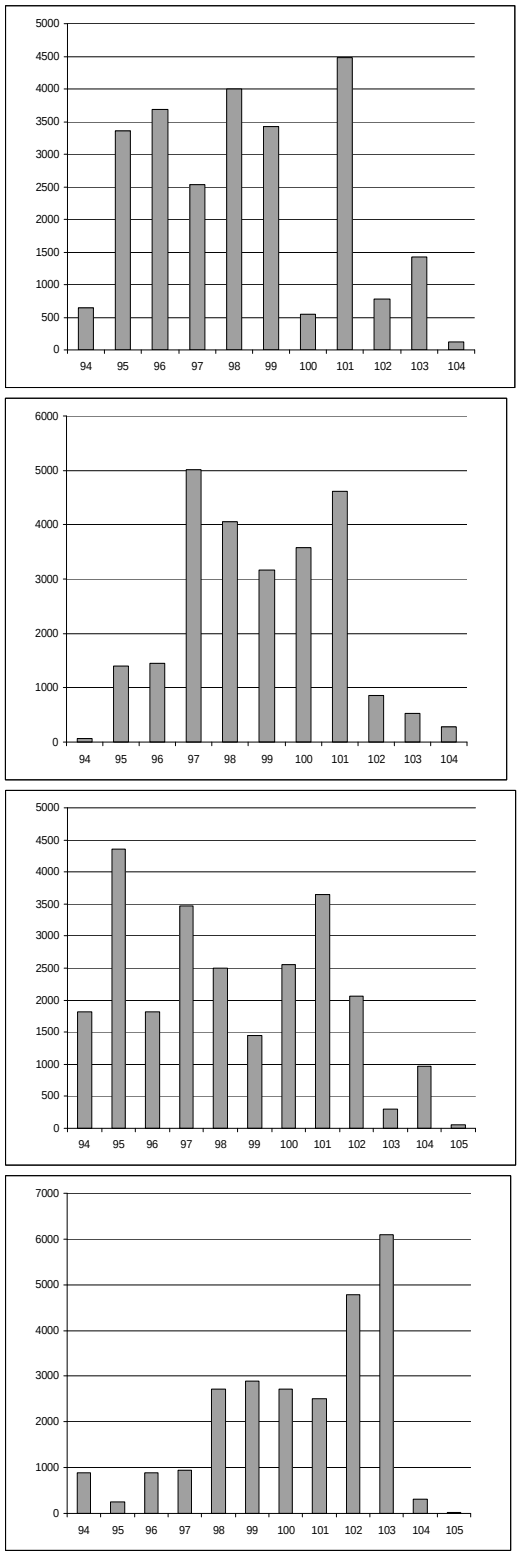


Fig. 2 Histograms of the allele length for constant population size of 25 000 individuals and with typical for mucrosatellites mutation rate equal 0,0005.

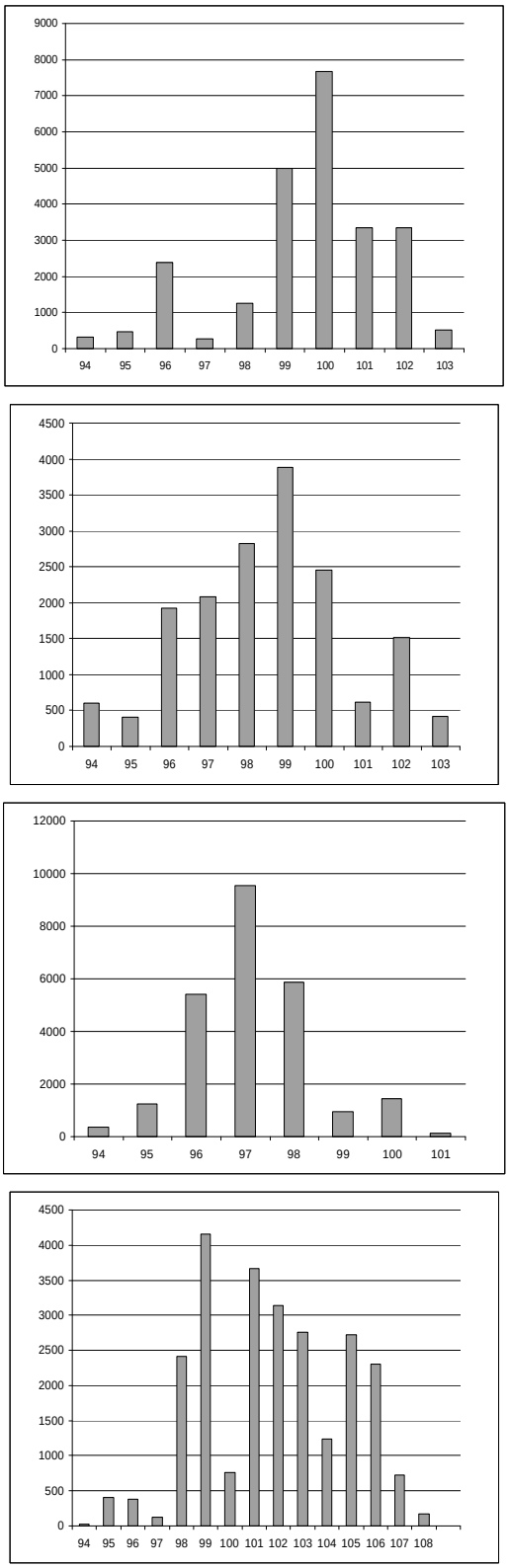


Fig. 3 Histograms of the allele length for population of the size 25 000 individuals which underwent 10 fold increase 20 000 generations ago and having typical for mucrosatellites mutation rate equal 0,0005.

D. Inputs of the Network

As inputs of our network we used four growth coefficients based on microsatellites:

a) two estimators of imbalance indices: Kimmel's imbalance index

$$\ln \hat{\beta}_1 = \ln \hat{\theta}_{\overline{V}} - \ln \hat{\theta}_{\overline{Po}} \tag{3}$$

and King's and Kimmel's imbalance index

$$\ln \hat{\beta}_2 = \frac{1}{m}\sum_{i=1}^{m}\left((\ln \hat{\theta}_V)_i - (\ln \hat{\theta}_{Po})_i\right) \tag{4}$$

In the aforementioned formulas m is the amount of microsatellites of each individual, $\hat{\theta}_V$ denotes allele size variance estimator of composite parameter $\theta = 4Nv$ and $\hat{\theta}_{Po}$ is allele size homozygosity estimator of θ. Moreover,

$$\hat{V} = \frac{1}{n(n-1)}\sum_{i\neq j}(X_i - X_j)^2 = \frac{2}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{5}$$

where n is the amount of individuals in the sample, X_i length of microsatellite of i^{th} individual and $\overline{X}$ is the mean of the length of microsatellites among individuals

$$\overline{V} = \frac{1}{m}\sum_{i=1}^{m}\hat{V}_i \tag{6}$$

and

$$\hat{P}_0 = \frac{\left(n\sum_{k\in K} p_k^2 - 1\right)}{n-1} \tag{7}$$

where K is a set of allele length in the sample, and

$$p_k = \frac{n_k}{n} \tag{8}$$

with n_k denoting the amount of alleles with length equal to k.

Additionally,

$$\overline{P_0} = \frac{1}{m}\sum_{i=1}^{m}\hat{P}_{0i} \tag{9}$$

and

$$\hat{\theta}_{Po} = \frac{1/\hat{P}_0^2 - 1}{2}. \tag{10}$$

These equations are further described in [7].

b) inter locus estimator *g* being the ratio of observed and predicted variance of the allele length

$$g = \frac{Var(\hat{V})}{\frac{4}{3}\overline{\hat{V}}^2 + \frac{1}{6}\overline{\hat{V}}} \quad (11)$$

Observe that in the above formula the nominator denotes the observed variance of the allele length given by

$$ObservedVariance(\hat{\theta}_V) = Var(\hat{V}) = \frac{1}{n-1}\sum_{j=1}^{m}\left(\hat{V}_j - \overline{\hat{V}}\right)^2 \quad (12)$$

and the denominator has got the meaning of the variance value predicted in drift-mutation equilibrium

$$ExpectedVariance(\hat{\theta}_V) = \frac{4}{3}\overline{\hat{V}}^2 + \frac{1}{6}\overline{\hat{V}} \quad (13)$$

In these formulas $\hat{V}_j$ is an unbiased estimator of variance of the allele length distribution at locus *j*, and $\overline{\hat{V}}$ is the mean of unbiased estimators of variance of allele length distributions. More detailed description of these equations the reader can find in [9].

c) within locus estimator *k*

$$k = 2{,}5 * Sig^4 + 0{,}28 * S^2 - 0{,}95/n - Gam_4 \quad (13)$$

where

$$Sig^4 = \frac{(n^2 - 3n + 3)}{n(n-1)(n-2)(n-3)}\left(\sum_{i=1}^{n}(X_i - \overline{X})^2\right)^2 - \frac{1}{(n-2)(n-3)}\sum(X_i - \overline{X})^4 \quad (14)$$

and

$$Gam_4 = \frac{(n^2 - 2n + 3)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}(X_i - \overline{X})^4 - \frac{(6n-9)}{n(n-1)(n-2)(n-3)}\left(\sum_{i=1}^{n}(Xi - \overline{X})^2)\right)^2 \quad (15)$$

whereas

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad (16)$$

In above equations described in detail in [10], S^2 has got the meaning of an unbiased estimator of the variance, Sig^4 is unbiased estimator for the variance squared, and Gam_4 is the fourth central moment. Of allele length distribution

Estimators of imbalance indices and values of inter locus estimator *g* usually receive values from range 0 to 1 so we can put them directly on the network input. Value of within locus coefficient *k* should be divided by a number of microsatellites *n* that we used to count this indicator.

IV. Results

In our experiments to implement random number generator we tried the same algorithm which was formerly used by the first author for simulating branching processes in the problem of Mitochondrial Eve dating [11]. However, finally, mainly due to easier implementation reasons we used so called Mersenne Twister generator.

We used also feed-forward neural network. Namely, one layer and two layers perceptrons were utilized because these networks are universal, contrary to probabilistic neural networks which learn much faster but are dedicated primarily for classification [12]. Also, due to their fast learning probabilistic neural networks can be applied as a criterion in optimization of feature space [13], but they are not a good choice in approximation problems considered in our study.

Having decided to use multilayer perceptrons we tested with different amounts of neurons in layers. Interestingly, we obtained the greatest power with single layer neural network containing only one neuron. For such a simple network there is a possibility to give the equation realized by network and therefore to define analytically a new test γ defined as

$$\gamma = \left(1 + e^{4{,}201\ln\hat{\beta}_1 + 2{,}417\ln\hat{\beta}_2 + 3{,}842\frac{k}{n} - 1{,}247g + 1{,}511}\right)^{-1} \quad (17)$$

Test γ returns values from a range (0,1) with cut off value equal to 0,797 at significance level 0.05 (if the test returns greater value we assume that sample comes from population that experienced growth).

We compared power of γ, with power achieved by one of the most powerful growth detectors, namely estimators of imbalance index. Based on empirical distribution of imbalance index estimators values for constant population size histories, we determined critical values for the tests, $\ln\hat{\beta}_1$ = -0,51 and $\ln\hat{\beta}_2$ = -0,787.

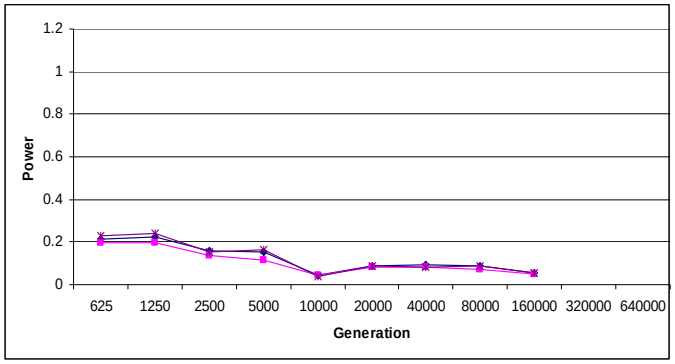
To obtain these values we created 150 histories of constant population size $N = 2\,500$ individuals and we simulated 100 000 generations. Starting from the 50 000th generation we

took samples from every generation divisible by 10 000. From each history we took 100 samples, each containing 40 individuals.

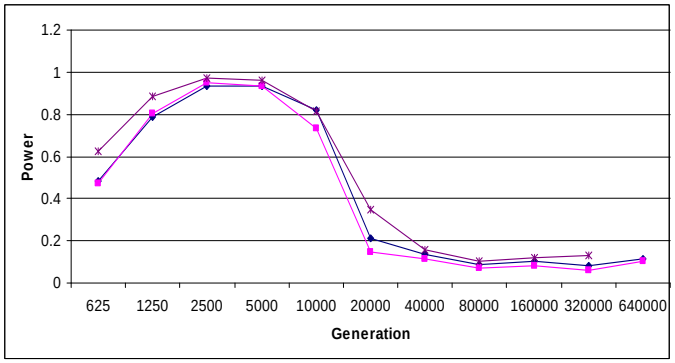
A. *Stepwise Growth*

For small stepwise growth (like two fold growth) there is no difference in power of detection between $\ln \hat{\beta}_1$ and γ (Fig 4a), actually all tests have low power.

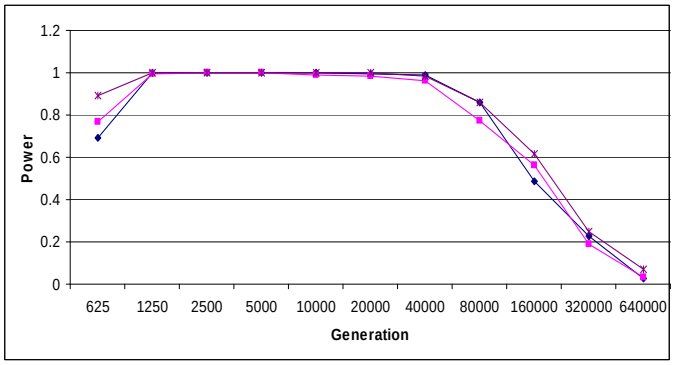
For ten fold growth there is a visible difference in power of tests (Fig 4b), neural network based test γ is able to detect growth earlier than methods based on imbalance index and signal of expansion stays longer. For 100 fold growth (Fig 4c), we can detect growth with γ even earlier and for longer time.



(a)



(b)

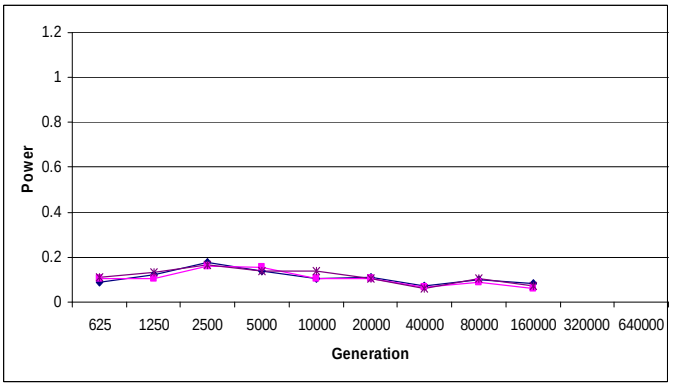


(c)

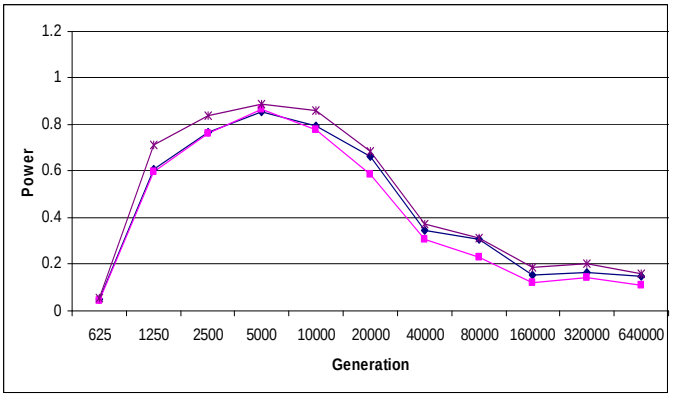
Fig. 4. Powers of $\ln \hat{\beta}_1$ (♦), $\ln \hat{\beta}_2$ (■) and neural network γ (*) tests. Populations experienced stepwise growth from N = 2 500 individuals to (a) 5000, (b) 25 000 and (c) 250 000 individuals.

B. *Exponential Growth*

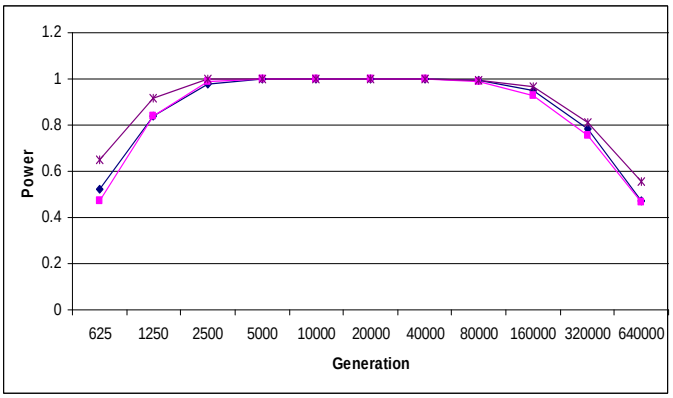
For small growths powers of all tests are low (Fig. 5a). In case of greater exponential expansion γ has greater power and might detect growth for longer time (Fig. 5 b and c). Fig. 5 shows us that γ usually give us outcomes better than other available tests based on microsatellites.



(a)



(b)



(c)

Fig. 5. Powers of $\ln \hat{\beta}_1$ (♦), $\ln \hat{\beta}_2$ (■) and neural network γ (*) tests. Populations experienced exponential growth from N=2 500 individuals to (a) 5000, (b) 25 000 and (c) 250 000 in different time

periods. For each generation marked on the graph we created set of 100 unlinked histories which final size was reached at marked time.

Differences in power of tests are especially visible for populations undergoing growth of bigger rate. In the Fig. 6. it is visible that for small growths (left side of the graph, little difference in amounts of individuals between two generations) powers of $\ln \hat{\beta}_1$ and γ are similar but for bigger (right side of the graph) growths γ has a greater power.

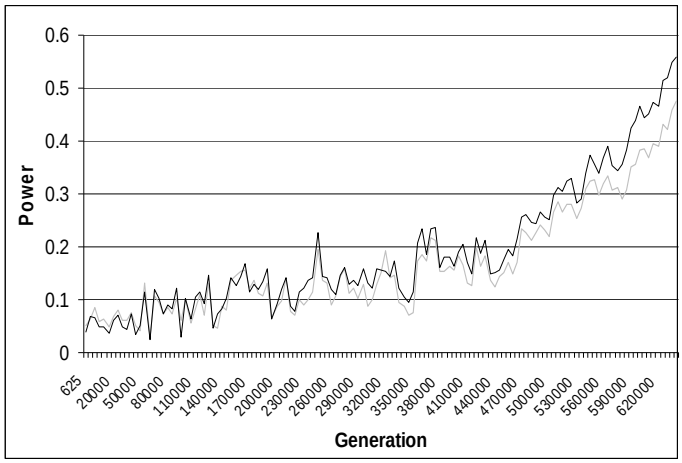
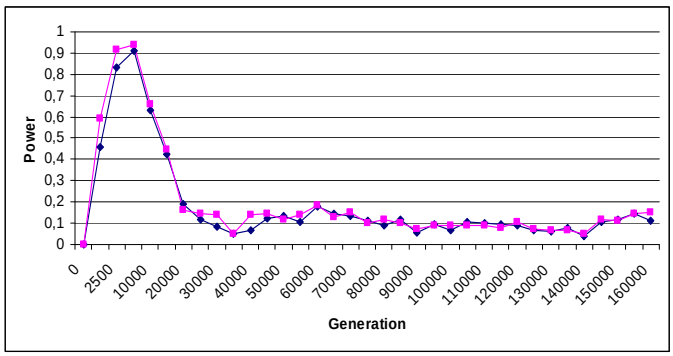
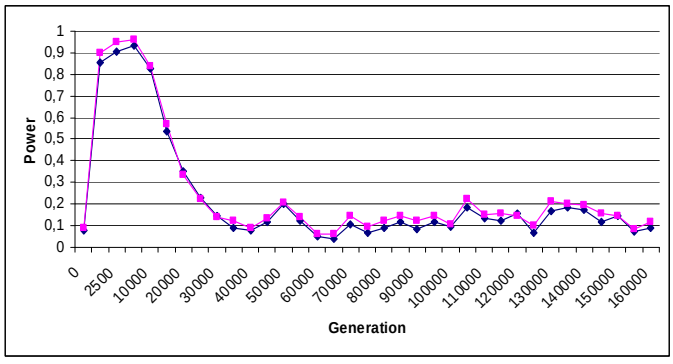


Fig. 6. Power of γ (black) and $\ln \hat{\beta}_2$ (gray) for population which undergoes exponential growth from N=2 500 to 250 000 individuals during 640 000. generations. Statistics are counted for number of generations marked on horizontal axis.



(a)



(b)

Fig. 7. Comparison of power of $\ln \hat{\beta}_1$ (♦) and γ (■) for different mutation rates. Mutation rate was (a) $v = 2.5\times10^{-4}$ and (b) $v = 7.5\times10^{-4}$.

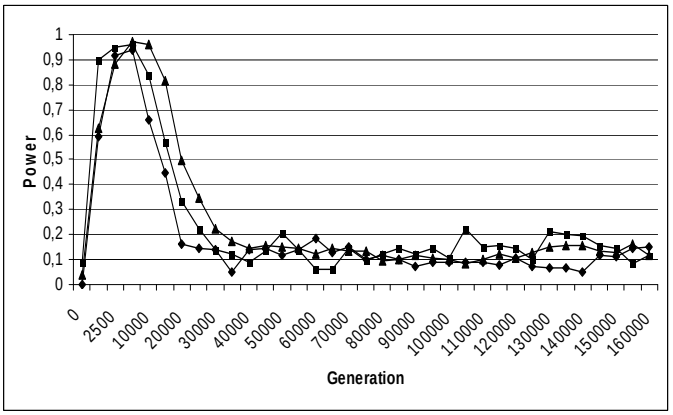
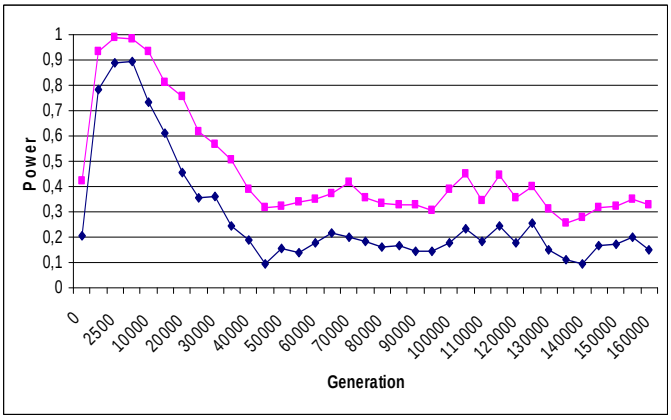
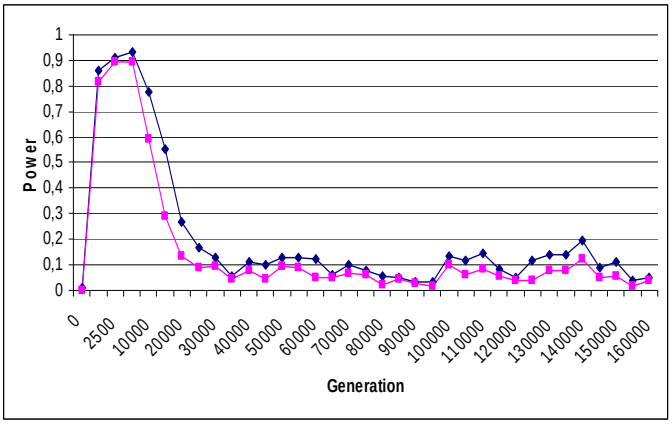


Fig. 8. Comparison of power of γ for different mutation rates during stepwise growth from 2 500 to 25 000 individuals. Mutation rate was $v = 2.5\times10^{-4}$ (♦), $v = 5\times10^{-4}$ (■) and $v = 7.5\times10^{-4}$ (▲).



(a)



(b)

Fig. 9. Comparison of power of $\ln \hat{\beta}_1$ (♦) and γ (■) for different amounts of microsatellites possessed by individuals in the sample. Each individual has (a) 10 or (b) 40 microsatellites.

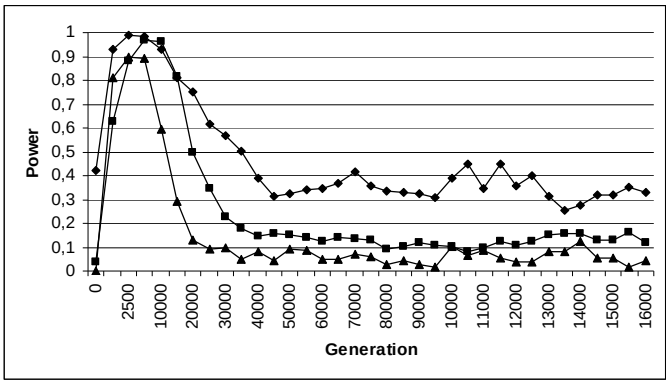
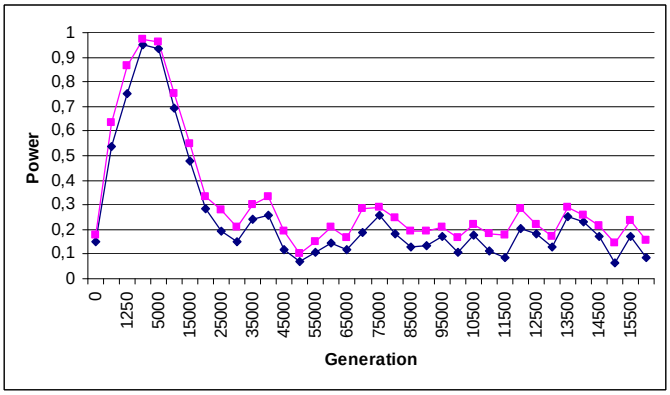
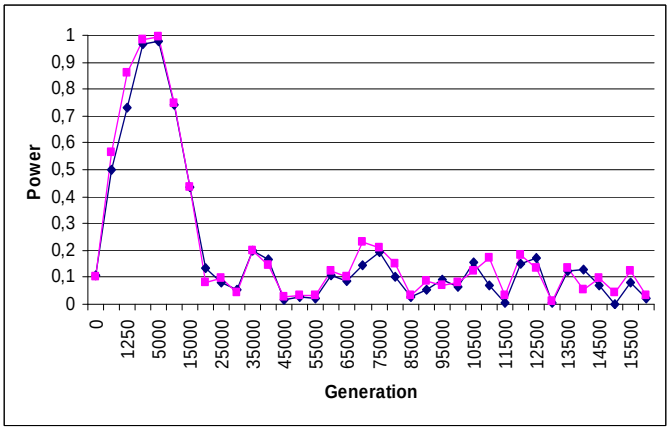


Fig. 10. Comparison of power of γ for different amounts of microsatellites in single individual equal to 10 (♦), 30 (■), 40 (▲).



(a)



(b)

Fig. 11. Comparison of power of $\ln \hat{\beta}_1$ (♦) and γ (■) for different amounts of individuals in the sample. From each unlinked history we took (a) 10, (b) 70 individuals.

C. Changes in the Mutation Rates

For samples coming from populations with different mutation rate power of growth detection is lower that for populations with mutation rate equal to $\nu = 5\times10^{-4}$ (compare Fig. 7a, 7b and Fig 8). However power of neural classifier is still higher than power of method based on imbalance index (Fig. 7).

D. Changes in the Amount of Microsatellites

Observe that for 40 microsatellites the power of γ is even slightly lesser than that of $\ln \hat{\beta}_1$ (Fig. 9b). However for small number of microsatellites the power of γ is substantially greater than that of $\ln \hat{\beta}_1$ (Fig. 9a). With the increasing number of microsatellites the power of γ is falling (Fig. 10) .

E. Changes in the Amount of Individuals in the Sample

In general the power of the novel test γ is greater than that of $\ln \hat{\beta}_1$ (Fig. 11). Note however, that with the increasing amount of individuals in the sample falling (Fig. 12) it is falling in similar fashion to imbalance to the fall of the power of index based test.

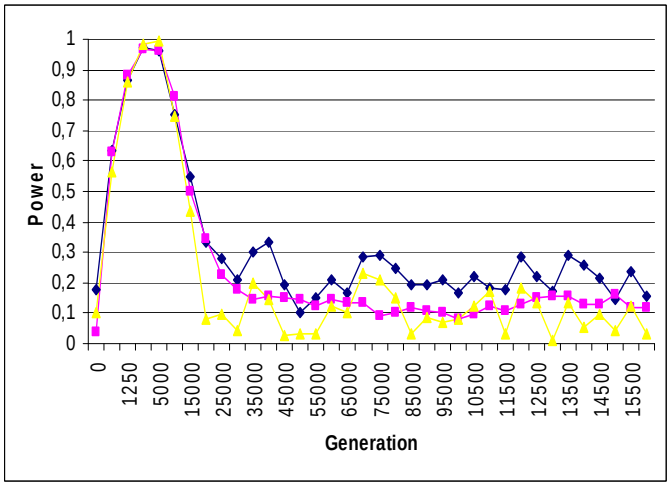


Fig. 12. Comparison of power of γ for different amounts of individuals in the sample: 10 (♦), 40 (■), 70 (▲).

V. Discussion

In the paper we demonstrated that artificial neural networks may be helpful in detection of population growth. Our study proved that properly trained neural network defines a novel statistical test γ given by (17) which achieve greater power in the detection of population growth, than any other tests based on microsatellites as it is showed in Fig. 4, 5. and 6.

For the greatest possible power of detection, the change in sample parameters (for example different amount of individuals in the sample taken from population or different number of microsatellites) requires in some cases the training of the new artificial neural network. Then, as a learning set we should use samples from populations with corrected parameters.

However, even without that in majority of cases the power of defined in the paper test γ (17) remains greater than that of known tests. It can be easy to understand if we take in mind that the test γ uses the information involved in other tests and the importance of information in any particular test is

weighted by the neural network according to the rule learned from training data obtained from computer simulations

It was proved by King, Kimmel, and Chakraborty [7] that power of imbalance indices $\ln \hat{\beta}_1$ and $\ln \hat{\beta}_2$ is greater than that of *k* and *g* statistics defined by Reich et al. in [9, 10], however we showed in this study that addition of information covered in these two latter tests can further increase the power of resulting γ statistic.

Therefore, the definition of this test, obtained by the use of neural networks techniques we consider as the main novelty of the paper. Except for the definition of the formula for γ test, we also present in the articlce its critical value at significance level 0.05, so the test can be used directly by others researches in problems of population growth detection inferred from microsatellites.

ACKNOWLEDGMENT

The authors would like to thank Prof. M. Kimmel from Department of Statistics at Rice University in Houston TX, USA, for advice and long discussions concerning population genetics problems

REFERENCES

[1] G Reibnegger, G Weiss, G Werner-Felmayer, G Judmaier, and H Wachter, "Neural networks as a tool for utilizing laboratory information: Comparison with linear discriminant analysis and with classification and regression trees", *Proceedings of the National Academy of Science USA*, vol 88, no. 24, pp. 11426-11430, 1991.

[2] M. C. Fisher, G. L. Koenig, Th. J. White, G. San-Blas, R. Negroni, I. Gutiérrez Alvarez, B. Wanke, and J. W. Taylor, "Biogeographic range expansion into South America by Coccidioides immitis mirrors New World patterns of human migration", *Proceeding of National Academy of Science USA*, vol. 98, no.8, pp. 4558-4562, 2001.

[3] A. Renwick, L. Davison, H. Spratt, J. P. King, and M. Kimmel, "DNA Dinucleotide Evolution in Humans: Fitting Theory to Facts", *Genetics*, vol. 159, no. 2, pp. 737-747, 2001.

[4] I. Agrafioti and M. P. H. Stumpf, "SNPSTR: a database of compound microsatellite-SNP markers", *Nucleic Acids Res.*, vol. 35, supplement 1, pp. D71-D75, 2007.

[5] E. J. Vowles and W. Amos, "Quantifying Ascertainment Bias and Species-Specific Length Differences in Human and Chimpanzee Microsatellites Using Genome Sequences", *Mol. Biol. Evol.*, vol. 23, no. 3, pp. 598-607, 2006.

[6] J. Hein, M. H. Schierup, and C. Wiuf, *Gene Genealogies: Variation and Evolution a Premier In Coalescent Theory*, Oxford University Press, 2005.

[7] J. P. King, M. Kimmel, and R. Chakraborty: "A Power Analysis of Microsatellite-Based Statistics for Inferring Past Population Growth", Mol. Biol. Evol., vol. 17, no. 12, pp. 1859-1868, 2000.

[8] M. J. Donnelly, M. C. Licht, and T. Lehmann: "Evidence for recent population expansion in the evolutionary history of the malaria vectors Anopheles arabiensis and Anopheles gambiae", *Mol. Biol. Evol.*, vol. 18, no. 7, pp. 1353-1364, 2001.

[9] D. E. Reich, M. W. Feldman, and D. B. Goldstein, "Statistical Properties of Two Tests that Use Multilocus Data Sets to Detect Population Expansions", *Mol. Biol. Evol.*, vol. 16, pp. 453-466, 1999.

[10] D. E. Reich and D. B. Goldstein, "Genetic evidence for a Paleolithic human population expansion in Africa", *Proceedings of National Academy of Science U S A.*, vol. 95, no. 14, pp. 8119-8123, 1998.

[11] K. A. Cyran, "Simulating branching processes in the problem of Mitochondrial Eve dating based on coalescent distributions", *International Journal of Mathematics and Computers in Simulation*, vol. 1, no. 3, pp. 268-274, 2007.

[12] K. A. Cyran, "Image recognition with a diffractive optical variable device and probabilistic neural network: improvements of feature space", *WSEAS Trans. on Information Science and Applications*, vol. 2, no. 12, pp. 2212-2219, 2005.

[13] K. A. Cyran, "Comparison of neural network and rule-based classifiers used as selection determinants in evolution of feature space", *WSEAS Trans. on Systems*, vol. 6, no. 3, pp. 549-555, 2007.

Krzysztof A. Cyran was born in Cracow, Poland, in 1968. He received MSc degree in computer science (1992) and PhD degree (with honours) in technical sciences with specialty in computer science (2000) from the Silesian University of Technology SUT, Gliwice, Poland. His PhD dissertation addresses the problem of image recognition with the use of computer generated holograms applied as ring-wedge detectors.

He has been an author and co-author of more than 70 technical papers in journals (several of them indexed by Thomson Scientific) and conference proceedings. These include scientific articles like: K. A. Cyran and A. Mrózek, "Rough sets in hybrid methods for pattern recognition," *Int. J. Intel. Syst.*, vol. 16, 2001, pp. 149-168, and K. A. Cyran and M. Kimmel, "Interactions of Neanderthals and modern humans: what can be inferred from mitochondrial DNA?" *Math. Biosci. Eng.*, vol. 2, 2005, pp. 487-498, as well as a monograph: U. Stańczyk, K. Cyran, and B. Pochopień, *Theory of Logic Circuits*, vol 1 and 2, Gliwice: Publishers of the Silesian University of Technology, 2007. Dr. Cyran (in 2003-2004) was a Visiting Scholar in Department of Statistics at Rice University in Houston, US. He is currently the Assistant Professor and the Vice-Head of the Institute of Informatics at Silesian University of Technology, Gliwice, Poland. His current research interests are in image recognition and processing, artificial intelligence, digital circuits, decision support systems, rough sets, computational population genetics and bioinformatics.

Dr. Cyran has been involved in numerous statutory projects led at the Institute of Informatics and some scientific grants awarded by the State Committee for Scientific Research. He also has received several awards of the Rector of the Silesian University of Technology for his scientific achievements. In 2004-2005 he was a member of International Society for Computational Biology. Currently he is a member of the Editorial Board of Journal of Biological Systems, member of the Scientific Program Committee of WSEAS international conferences in Malta (ECC'08), Rodos (AIC'08, ISCGAV'08, ISTASC'08) and multiconference in Crete (CSCC'08) as well as a reviewer for Studia Informatica and such journals indexed by Thompson Scientific as: Optoelectronic Review, Mathematical Biosciences and Engineering, and Journal of Biological Systems.

Dariusz Myszor was born in Pszczyna, Poland in 1983. He received MSc degree in computer science (2007) from the Silesian University of Technology SUT, Gliwice, Poland.

Currently he is a PhD student at Silesian University of Technology and is involved in scientific and didactiv activities of the Institute of Informatics at SUT.

His current research interests are focused in such areas as: neural networks, computational population genetics, computer simulations and bioinformatics.