Applied Medical Informatics
Original Research Vol. 28, No. 1/2011, pp: 45-52

Determining the Similarities between Stations using the Haplotypes of the Species *Cerastoderma glaucum* (Poiret, 1789) from the Romanian Black Sea Infralittoral

Dragomir-Cosmin DAVID[1], Ştefan I. ŢIGAN[2]

[1]Babeş-Bolyai University, Institute for Interdisciplinary Research in Nano-Bio-Sciences, Molecular Biology Center, 42 Treboniu Laurian Street, 400271 Cluj-Napoca, Romania.
[2]"Iuliu Hatieganu" University of Medicine and Pharmacy Cluj-Napoca, 13 Emil Isac, 400023 Cluj-Napoca, Romania.
E-mails: cosmin.david@hasdeu.ubbcluj.ro; stigan@umfcluj.ro

* Author to whom correspondence should be addressed; Tel.: +4-0749942417.

Received: 28 November 2010 /Accepted: 21 January 2011 / Published online: 7 March 2011

Abstract

Cerastoderma glaucum (Poiret, 1789) represents one of the marine mollusk species found on the Romanian Black Sea soft bottom infralittoral. This species had been investigated in the past by morphological methods. Some problems concerning its biological, ecological and even behavioral aspects have remained unresolved. At the moment, molecular technique studies allow the completion of these gaps. In the present article we are aiming to identify this species by molecular methods and to perform a statistical analysis of the obtained data in order to determine the distribution of this species' haplotypes on the littoral. With the aid of the Nucleospin® Tissue kits, through the PCR technique, the extraction and amplification of DNA has been performed to obtain the necessary data. The obtained sequences have been compared to the sequences from databases such as National Center for Biotechnology Information (NCBI) to establish similarities. After the identification of the similar sequences, we have interpreted their distribution in different stations using the Kendall rank correlation coefficient. A number of 12 haplotypes which belong to 32 individuals from 5 stations have been identified. Using the Kendall rank correlation coefficient the identified haplotypes show different degrees of concordance with the standard haplotypes of this species. A significant statistical similarity appears in 81.14% out of the total 493 compared haplotype pairs. Out of these, 18% present perfect similarity, 76.25% good and 5.75% acceptable. The stations are similar regarding haplotypes, with some exceptions that are commented in the paper, similarities established by calculating the Kendall rank correlation coefficient.

Keywords: *Cerastoderma glaucum*; Haplotypes; Statistics; Kendall's tau coefficient; Black Sea.

Introduction

Cerastoderma glaucum (Poiret, 1789) is a euryhaline and eurythermic species which has a large distribution, being present on the European seacoasts. It can be found in the Mediterranean Sea, Caspian Sea, Black Sea, Sea of Azov and in the eastern part of the Baltic Sea. The distribution of this species is limited due to exposure to waves. This is the reason why it is never found in loose sediments, having a reduced burrowing capacity [1,2].

C. glaucum also has an economical importance because it has been a source of food ever since

the antiquity, according to Xenocrates. He stated that the soup prepared from this species had a laxative effect and the meat was not easily digestible. At present, even though this species is not consumed in our country, it represents a source of food for the coastal communities of the Nestos and Vistonis lagoons from Greece [3]. Marine filtering organisms have the capacity to accumulate heavy metals from aquatic environments, resulted from industrialization. *C. glaucum* is one of these species and accumulates heavy metals like Pb, Cr, Hg, having a favorable impact upon the environment [4]. Because of this accumulative capacity, consuming it as a source of food may have some harmful effects on health.

Past studies upon the *Cerastoderma glaucum* species approached morphological characteristics that involved biological and ecological investigations. Thus, its reproduction [5], temperature and salinity influence upon the species [6], exposure to air and respiration [7], byssus structure [8] have been investigated. A transition towards the investigation of this species through molecular means has been done gradually, by studying the DNA variation [9,10], gene flow [11], phylogeography [12] and use of genetic markers [13-15].

The methods employed in molecular studies show the importance of the nuclear and mitochondrial genetic markers because of the multitude of data that they provide. The most frequently used genetic markers in molecular studies are the mitochondrial and ribosomal genes.

A mitochondrial marker must fulfill certain conditions: it must be present in a single copy in the genome, otherwise the obtained sequences could come from different copies of the genes; the sequences must be easy to align; the composition of nitrogen bases must be approximately equal and the primers must be universal [16]. Some examples of mitochondrial markers are: 16S rRNA, 12S rDNA, *COI* (cytochrome oxidase subunit I) and some nuclear markers: ITS1 (internal transcribed spacer 1), ITS2 (internal transcribed spacer 2), 18S rRNA (small subunit of the ribosomal ARN), 28S rRNA (large subunit of the ribosomal ARN). Out of all these genetic markers, cytochrome oxidase subunit I (*COI*) has been chosen as a universal genetic marker used in order to identify species on a molecular basis (DNA barcoding) [17].

The aim of this paper is to identify the existent haplotypes from the Romanian Black Sea infralittoral and to compare them by calculating the Kendall rank correlation coefficient between stations where these haplotypes are found to establish their similarity.

Material and Method

Sample Collection and DNA Extraction

The *C. glaucum* samples used in this study have been collected by autonomous diver from 7 stations situated on the Romanian Black Sea infralittoral from a depth of maximum 25 m (fig. 1, tab. 1). After sampling, the individuals have been preserved in ethanol 90% to ensure proper conservation of the DNA and stored for later analysis. Individuals from each station have been numbered as follows: Navodari 4, 5, 6, 11 ,12, 13; Mamaia 1, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 17, 19, 20, 21; Agigea 5, 6, 8, 9, 17; Eforie 3, 5; Vama Veche 7, 10, 11, 13.

Table 1. Sampling stations from the Romanian Black Sea infralittoral

Collecting area	Number of individuals	Geographical coordinates	
Vadu (V)	4	Latitude: 44°26'56.25"N	Longitude: 28°44'6.16"E
Navodari (N)	13	Latitude: 44°19'16.00"N	Longitude: 28°36'48.00"E
Mamaia (M)	23	Latitude: 44°12'41.29"N	Longitude: 28°38'39.56"E
Constanta (Ct)	6	Latitude: 44°10'24.00"N	Longitude: 28°38'18.00"E
Agigea (AG)	9	Latitude: 44° 5'30.00"N	Longitude: 28°36'41.00"E
Eforie (E)	4	Latitude: 44° 2'56.81"N	Longitude: 28°39'9.82"E
Vama Veche (VV)	10	Latitude: 43°45'10.53"N	Longitude: 28°34'21.18"E

DNA extraction has been performed using the classical method of phenol-chloroform-isoamyl alcoohol (25:24:1) and Nucleospin® Tissue, (Macherey-Nagel, Düren, Germany) commercial

extraction kit in some cases to obtain a better DNA quality.

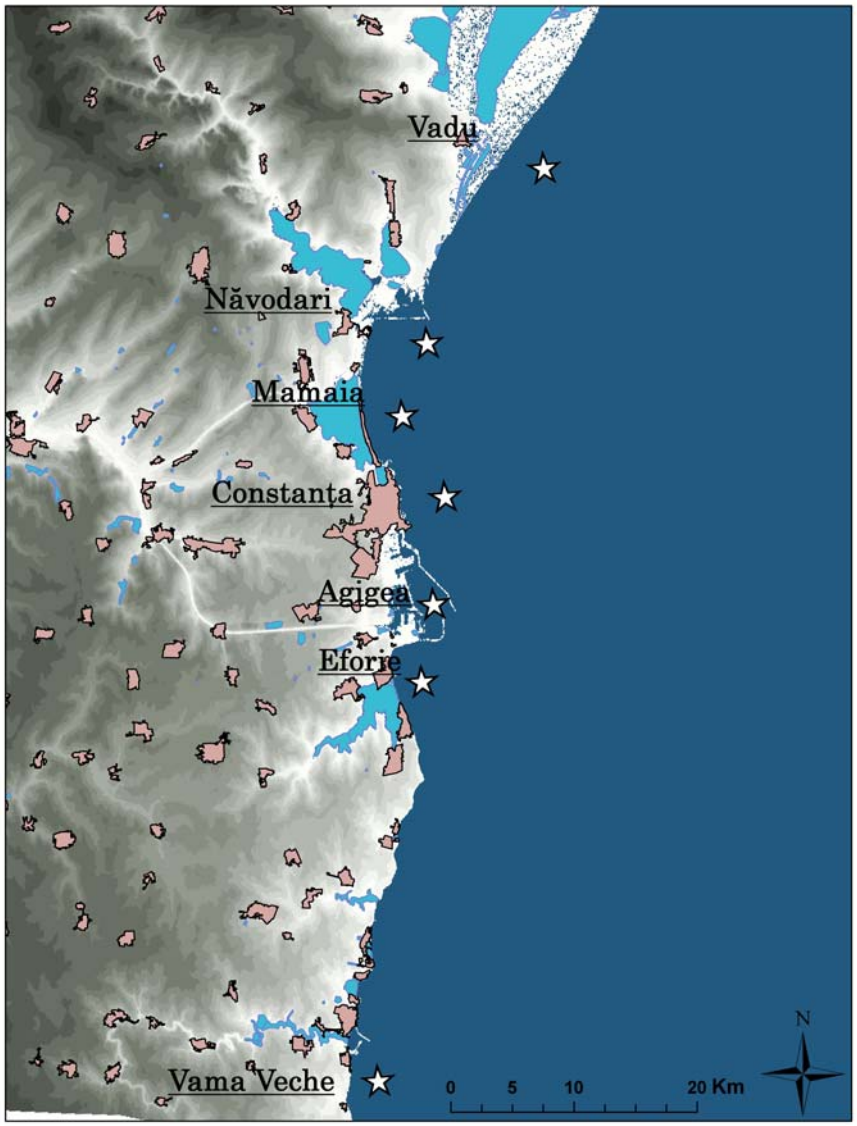

Figure 1. Sampling points from the Romanian Black Sea infralittoral

Sample Amplification and DNA Sequencing

The gene or gene fragments were amplified using universal primers. The genetic marker for which the amplification was performed is *COI* (cytochrome oxidase subunit I), a mitochondrial marker. Universal primers have been used to amplify the gene for cytochrome oxidase as follows: LCO1490: 5'-GGTCAACAAATCATAAAGATATTGG-3' and HCO2198: 5'-TAAACTTCAGGGTGACCAAAAAATCA-3' [18]. Mastercycler ep Authorized Thermal Cycler and Gradient Palm-Cycler™ machines were used in order to amplify DNA.

The PCR reaction mixture of 25µl (12.5 µl Mangomix, 0.5 µl $MgCl_2$ 50 mM concentration, 0.25 µl of each primer, 2.5 µl of DNA template and 9.5 µl H_2O UP/UV) was subjected to amplification.

The amplification process for cytochrome oxidase subunit I (*COI*) consisted of total denaturation of DNA for 5 minutes at 95°C, followed by a number of 35 cycles. Each of these cycles followed three steps: a denaturation at 95°C for 30 seconds, an alignment at 47°C for 30 seconds and an elongation at 72°C also for 30 seconds. The final elongation was done at 72°C for 10 minutes. The amplification results have been verified on an electrophoresis gel obtained with a

Tris-Acetate-EDTA (TAE) buffer solution of 2% concentration. The PCR products have been sequenced with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) using the Genetic Analysis System CEQ™ 8800 (Beckman Coulter) sequencer or have been sent for sequencing to Macrogen Inc. (South Korea).

The resulted sequences have been processed with the Biological sequence alignment editor (BioEdit version 7.0.5.3) program and aligned by ClustalW multiple alignment algorithm [19] implemented in this program [20]. The obtained sequences were compared to other sequences available in international databases like National Center for Biotechnology Information (NCBI). The BLAST (Basic Local Alignment Tool) program was utilized [21].

The Kendall rank correlation coefficient and the Statistica program were used in order to analyze the relations between stations [22]. The Kendall rank correlation coefficient was determined through the Statistica program [23].

Results

All the sampled individuals from the 7 stations were analyzed on a molecular basis. DNA sequences were obtained after the amplification with species-specific universal primers for individuals from the following 5 stations: Navodari, Mamaia, Agigea, Eforie Nord and Vama Veche. In the cases of Vadu and Constanta, DNA was less concentrated or too concentrated resulting in the impossibility of amplifying DNA with *COI*, although a lot of protocols have been tried. The results have been compared to those in the NCBI database.

Following the BLAST analysis 12 haplotypes were identified, with a minimum of 9 haplotypes/individual. These are: 24, 25, 26, 27, 28, 29, 30, 31, 32, 175, 179, 188. The analysis requires the identification of similar sequences from the NCBI database. The investigation has been carried out by searching a maximum of 10 identical sequences in the NCBI database in order to obtain a very high similarity. All the *C. glaucum COI* sequences from the database have been taken into consideration.

The high number of identified haplotypes belonging to 32 *C. glaucum* individuals from the 5 stations shows a high genetic variability.

The haplotypes that were identified in the analyzed samples present similarities to the standard haplotypes in various proportions, between 98% and 100%. In this way, two groups regarding the haplotype similarity are differentiated. The haplotypes of all the individuals from Eforie Nord and Vama Veche, one from Navodari and one from Mamaia have a similarity of 99 – 100% with the standard ones. The rest of the individuals from Navodari, Mamaia and Agigea show a lower percentage similarity (98 – 100%).

According to the number of present haplotypes, there are 9 identical haplotypes (24, 25, 26, 27, 28, 29, 30, 31, and 32) for the next stations: Agigea, Eforie and Vama Veche, whereas in Navodari and Mamaia there are a few haplotypes which are not found in the other stations. For example, one individual from Navodari has in addition the haplotype 179 and one individual from Mamaia – the haplotypes 175 and 188.

Based on the described haplotypes, the comparison of all individuals by pairs has been done with the help of the Kendall rank correlation coefficient. The significance of the test is relevant if the p-level < 0.001. Following the comparison of 493 possible combinations, in 383 cases (77.69%) the similarities are significant (Figure 2). Table number 2 illustrates the distribution of concordant significant haplotype pairs into 3 groups established by the Kendall coefficient. Thus, perfect concordance has been distinguished in the case of 18.80% from haplotype pairs, good concordance in 79.63% and acceptable concordance in 1.57%. The ratio between concordant and discordant haplotype pairs are presented in Figure 2. All of these cases indicate a similarity of haplotypes present in the majority of the investigated stations. A relatively even distribution of haplotypes from the Romanian Black Sea infralittoral is observed by comparing individuals according to this statistical method. A more profound characterization of the genetic structure of *C. glaucum* is possible through diversification of mathematical methods.

The concordance of the analyzed haplotypes was weak where the value of $K < 0.6$ and was

absent where $p \geq 0.05$.

Comparing individual 13 from Navodari to individual 19 from Mamaia, the values of p-level = 0.1181416 and K= - 0.344828.

These low values indicate their discordance, also suggested by the differences in haplotype distribution. In addition to the common haplotypes, haplotypes 175 and 188 from Mamaia and 179 from Navodari are present. Haplotype 31 is absent from Mamaia.

Table 2. Concordance distribution according to statistical significance

Number of compared haplotype pairs	p-level	Kendall rank correlation coefficient	K values	Concordance
72	0.000006	1.000000	K=1	perfect
122	0.000070	0.878310	0.74< K <1	good
4	0.000153	0.836660	0.74< K <1	
18	0.000240	0.811503	0.74< K <1	
31	0.000049	0.897538	0.74< K <1	
130	0.000774	0.742857	0.74< K <1	
6	0.000882	0.734847	0.6<K< 0.74	acceptable

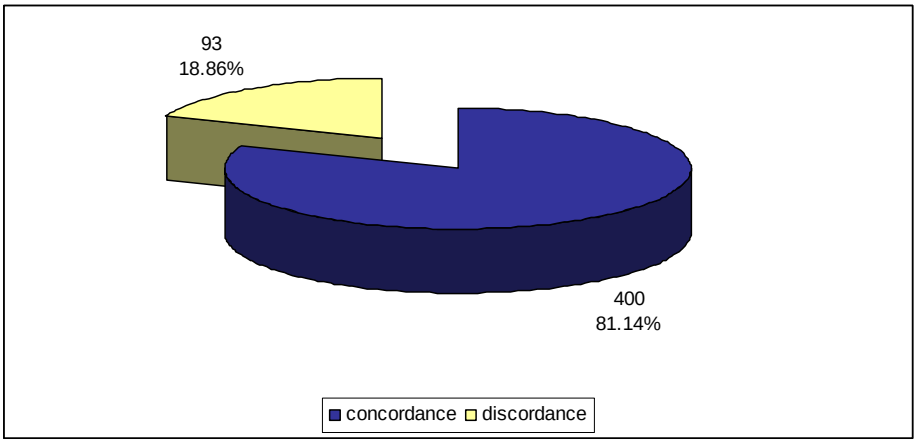

Figure 2. Percentage of the compared haplotype pairs

Even though in most of the cases there is a similarity of compared haplotypes, there are some statistically insignificant results as well. If taking into consideration the insignificant data, the percentage of significant concordance modifies as follows: perfect concordance 14.60%, good concordance 61.87% and acceptable concordance 1.22%.

The total number of statistically insignificant haplotype pairs is of 110 haplotype pairs (22.31%). These negative values indicate discordance due to the differences that appear between the haplotypes of the two individuals from Mamaia and Navodari.

Confidence intervals have been calculated for the significant concordance situations (Table 3).

Table 3. Confidence interval for concordance percentage

Concordance	Haplotype pairs	Percentage of the compared haplotype pairs (%)	95%CI (Confidence interval) (%)
Perfect	72	14.60	11.90-18.47
Good	305	61.87	58.67-67.46
Acceptable	6	1.22	0.50-2.80
Discordance	110	22.31	19.19-26.82
Perfect + Good	377	76.47	74.04-81.61
Perfect + Good + Acceptable	383	77.69	75.34-82.76

Discussion

The total number of identified haplotypes was 12, out of which 9 were common to all investigated individuals.

The high number of identified haplotypes belonging to 32 *C. glaucum* individuals from the 5 stations indicated a high genetic variability.

There is a similarity between the identified haplotypes of the studied material and the standard ones in a percentage of 98 - 100%.

By calculating the confidence interval, there is a 95% probability that there is significant concordance between pairs of identified haplotypes, with a variation of 75.34%-82.76% in reality.

According to *COI*, two main phylogroups were determined regarding the genetic distribution of the species *C. glaucum*: the Atlanto-Mediterranean phylogroup and the Ponto-Caspian phylogroup which also includes the haplotypes present in the Black Sea [15]. Phylogeographically, the identified haplotypes belong to the Ponto-Caspian phylogroup.

Comparing individuals by the Kendall rank correlation coefficient, there is a relatively even distribution of haplotypes along the Black Sea littoral. A more profound characterization of the structure of *C. glaucum* is possible through diversification of mathematical means.

Conclusions

The haplotypes of all individuals from Eforie Nord and Vama Veche, one individual from Navodari and one from Mamaia have a 99 – 100% similarity to the standard haplotypes.

The rest of the individuals from Navodari, Mamaia and Agigea have a lower percentage (98 – 100%) regarding similarity.

Out of the 12 identified haplotypes, 9 are found in all of the 5 investigated stations, suggesting a similarity between them.

Higher genetic variability found in 2 individuals (3 different haplotypes) does not significantly modify the similarity between stations.

By comparing the haplotypes of the individuals using the Kendall rank correlation coefficient a relatively even distribution along the Black Sea littoral results.

Three significant concordance ranks of identified haplotypes compared to standard ones have been obtained using the Kendall coefficient.

The results have been interpreted with a single genetic marker and represent preliminary data upon investigating the species *Cerastoderma glaucum*.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

The authors wish to thank for the financial support provided from programs co-financed by The SECTORAL OPERATIONAL PROGRAMME HUMAN RESOURCES DEVELOPMENT, Contract POSDRU 6/1.5/S/3 „Doctoral studies: through science towards society".

The authors also wish to thank for the help of biologist Brînduşa Bocoş and PhD student Sergiu Chira.

The study was supported partially by the PNII_ID_2040/2008. The samples used in this study were collected by Dr. Bea Kelemen and analyzed in the molecular biology laboratory of Dr. Octavian Popescu.

References

1. Russell PJC. Biological studies on *Cardium glaucum,* based on some Baltic and Mediterranean populations. Mar Biol 1972;16:290-6.
2. Brock V. Habitat selection of two congeneric bivalves, *Cardium edule* and *C. glaucum* in sympatric and allopatric populations. Mar Biol 1979;54:149-56.
3. Voultsiadou E, Koutsoubas D, Achparaki M. Bivalve mollusc exploitation in Mediterranean coastal communities: an historical approach. J Biol Res-Thessalon 2010;13:35-45.
4. Szefer P, Wolowicz M, Kusak A, Deslous-Paoli J-M, Czarnowski W, Frelek K, Belzunce M-J. Distribution of Mercury and Other Trace Metals in teh cockle Cerastoderma glaucum from the Mediterranean Lagoon Etang de Thau. Arch Environ Contam Toxicol 1999;36:56-63.
5. Boyden CR. A comparative study of the reproductive cycles of the cockles *Cerastoderma edule* and *C. glaucum*. J Mar Biol Ass UK 1971;51:605-22.
6. Kingston P. Some observations on the effects of temperature and salinty upon the growth of *Cardium edule* and *Cardium glaucum* larvae in the laboratory. J Mar Biol Ass UK 1974;54:309-17.
7. Boyden CR. The behaviour, survival and respiration of the cockles *Cerastoderma edule* and *C. glaucum* in air. J Mar Biol Ass UK 1972;52:661-80.
8. Yankson K. Observations on byssus systems in the spat of *Cerastoderma glaucum* and *C. edule*. J Mar Biol Ass UK 1986;51:277-92.
9. Brock V. Genetic relations between the bivalves *Cardium* (*Cerastoderma*) *edule*, *Cardium lamarcki* and *Cardium glaucum*, studied by means of crossed immunolectrophoresis. Mar Biol 1987;93:493-8.
10. Brock V. Christiansen G. Evolution of *Cardium* (*Cerastoderma*) *edule*, *C. lamarcki* and *C. glaucum*: studies of DNA-variation. Mar Biol 1989;102:505-11.
11. Mariani S, Ketmaier V, de Matthaeis, E. Genetic structuring and gene flow in *Cerastoderma glaucum* (Bivalvia: Cardiidae): evidence from allozymes variation at different geographic scales. Mar Biol 2002;140:687-97.
12. Nikula R, Vainola R. Phylogeography of *Cerastoderma glaucum* (Bivalvia: Cardiidae) across Europe: a major break in the Eastern Mediterranean. Mar Biol 2003;143:339-50.
13. Freire R, Arias A, Mendez J, Insua A. Sequence variation of the internal transcribed spacer (ITS) region of ribosomal DNA in *Cerastoderma* species (Bivalvia: Cardiidae). J Moll Stud 2010;76:77-86.
14. Ladhar-Chaabouni L, Hamza-Chaffai A, Hardivillier Y, Chenais B, Denis F. A pilot study of genetic differentiation betweeen two phenotypes of a Mediterranean population of the bivalve *Cerastoderma glaucum* and genetic discrimination with other *Cerastoderma glaucum* and *Cerastoderma edule* populations outside the Mediterranean. Mar Ecol 2010;31:355-63.
15. Tarnowska K, Chenuil A, Nikula R, Feral J-P, Wolowicz M. Complex genetic population structure of the bivalve *Cerastoderma glaucum* in a highly fragmented lagoon habitat. Mar Ecol Prog Ser 2010;406:173-84.
16. Cruickshank RH. Molecular markers for the phylogenetics of mites and ticks. Syst Appl Acarol 2002;7:3-14.
17. Remigio EA, Hebert PDN. Testing the utility of partial COI sequences for phylogenetic estimates of gastropod relationships. Mol Phylogenet Evol 2003;29:641-7.
18. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan inveretebrates. Mol Mar Biol Biotechnol 1994;3:294-9.
19. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res 1994;22:4673-80.
20. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 1999;41:95-8.
21. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 1997;25:3389-402.

22. Kirkwood BR, Sterne AC. Essential Medical Statistics. Blackwell Science; Oxford, 2003.
23. . *******, Statistica for Windows, Second Edition, Vol.1, StatSoft, http://www.statsoft.com Tulsa. USA, 1998.