The Open Biotechnology Journal, 2015, *9*, 191-198

Analysis of Correlations Between Climate and Molecular Adaptive Evolution of Wild Barley with Geographical Information Systems (GIS)

Lina Zhao, Tao Pei, Gege Yang and Chenghu Zhou*

State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, China

Abstract: Climate is one of the most important factors determining the adaptive evolution of plants. In this study, 44 different populations of wild barley were used as materials to analyze the diversity of 17 genes (495 sequences), in order to study the influence of different climatic conditions on adaptive evolution of wild barley. A Geographical Information System (GIS) provided tools to visually present and analyse the geographical distribution of number of variable positions in the alignments and the differentiation index. 19 different bio factors were classified into 4 main component groups, and OLS Regression was used to analyze the differentiation coefficient of main climatic factors and the expression of gene differentiation degree. DHN family, a drought and temperature related gene family, showed significant spatial correlation with main climatic factors. Meanwhile, the results of other genes, which reported insignificant effects on drought or heat tolerance, were quite contrary. The phenotypes of plants were observed after the interaction of plant genotype with the environment. Simultaneously, the environment also has influence on some special genes. New analysis methods using GIS could be used to research the complex relationship between plant phenotype, plant genotype and the environment.

Keywords: Gene sequence analysis, biodiversity, molecular evolution, geographical information system (GIS), data meta-analysis, wild barley.

1. INTRODUCTION

Climate is one of the most important reasons for adaptive evolution in plants, and also leads to important selective factors determining intraspecific differentiation [1]. Yongfeng Zhou *et al*. used two closely related pine species growing in southeastern Chinese, Pinus massoniana Lamb and Pinus hwangshanensis Hisa as materials, to analyse 25 climate related genes.They found that variations in climate played an important role in the ecological divergence of the two species [2]. Not only on the woody perennial plants, climate also has a similar impact on the annual herbaceous plants such as Arabidopsis, barley and so on [3]. Wild barley (H. vulgare ssp. Spontaneum) is an annual, diploid grass species with 7 chromosomes and an estimated rate of self-fertilization of 98%. As the progenitor of cultivated barley, wild barley represents an important resource for the study of the adaptive evolution of plants' population [4].

The phenotypes of plants were observed after examining the results of the interaction of plant genotype and the environment. Meanwhile, the environment also had effects on some special genes. Alcohol dehydrogenases (ADH) is a group of Zn-binding enzymes which present in many organisms. The ADH family plays an important role in facilitating the interconversion between alcohols and aldehydes or ketones with nicotinamide adenine dinucleotide (NAD) as a coenzyme [5]. According to the previous reports, ADH family of plants has been observed to be involved in the biotic and abiotic stresses, such as disease resistance, drought tolerance, flooding resistance and so on [6]. Dehydrin (DHN) is a multi-family of proteins belong to the Group II Late Embryogenesis Abundant (LEA) family [7], and plays an important role in cold and drought stresses. Tommashi *et al*. found the expression levels of Dhn1, Dhn2, Dhn3, Dhn4, Dhn7, Dhn9 and Dhn10 in barley, to be highly increased in the germ, mesocotyl and the roots during drought stress. Meanwhile, Dhn5, Dhn8 and Dhn13 were significantly induced by cold and drought stresses [8]. In order to resist environmental change, and propagate the race, many factors such as natural selection, mating system, migration and genetic drift result in genetic diversity in the population. Even the same gene in the same population will have diversity because of the environmental differences. However, the relationship between genetic diversity and environmental differences could be analyzed through geographic information system (GIS) [9].

The base of Geographic Information System (GIS) is mainly the geographical space database. By collecting relevant information about the geographical space, appropriately processing the analysis, operating simulation, and using the geographic modeling, mutiple and dynamic geographic information resources were provided. Recently, GIS technology has been highly developed, and widely applied in many fields of life and production, such as for the

*Address correspondence to this author at the State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, China; Tel: 13911823799; E-mail: 596295719@qq.com

1874-0707/15

Table 1. Loci, length of the aligned DNA sequences and numbers of accessions used in the study.

Locus	Abbrev.	Authors	Length of Alignment	No. of Accessions Investigated
Alcohol dehydrogenase 1	Adh1	Michael P. Cummings and Michael T. Clegg (1998)	1362	19
Alcohol dehydrogenase 2	Adh2	Jing-Zhong Lin *et al.* (2002)	1980	25
Alcohol dehydrogenase 3	Adh3	Jing-Zhong Lin *et al.* (2001)	1873	25
Alpha-amylase type B gene	α-AMYb	Peter L. Morrel *et al.* (2003)	856	25
C-repeat binding factor 3-like protein	Cfb3	Peter L. Morrell and Michael T. Clegg (2006)	1514	44
Dehydrin 1	Dhn1	Peter L. Morrel *et al.* (2004)	1538	23
Dehydrin 4	Dhn4	Peter L. Morrell and Michael T. Clegg (2006)	1047	24
Dehydrin 5	Dhn5	Peter L. Morrel *et al.* (2003)	1088	23
Dehydrin 7	Dhn7	Peter L. Morrel *et al.* (2004)	1400	27
Dehydrin 9	Dhn9	Peter L. Morrel *et al.* (2003)	1011	45
Glutathione-dependent formaldehyde dehydrogenase	Faldh	Peter L. Morrell and Michael T. Clegg (2006)	1092	25
Glyceraldehyde-3-phosphate dehydrogenase	G3pdh	Peter L. Morrel *et al.* (2003)	2010	26
Putative cleaveage stimulation factor	ORF1	Peter L. Morrell and Michael T. Clegg (2006)	1533	45
Phosphoenolpyruvate carboxylase	Pepc	Peter L. Morrel *et al.* (2003)	3173	25
Putative serine/threonine kinase	Stk	Peter L. Morrell and Michael T. Clegg (2006)	1057	25
MADS box transcription factor	Vrn1	Peter L. Morrel *et al.* (2004)	1262	19
Granule bound starch synthase	Waxy	Peter L. Morrel *et al.* (2003)	1232	26
Heat shock protein 17	Hsp17	Liu, Y. and Yang, Z.(2010)	1361	16

prediction of spatial distribution of variety of diseases and insect pests [10], topographic and geomorphic conditions, meteorological factors, and for the regional distribution of diseases [11]. In this article, GIS technology was used to analysz the relationship between climate and genetic diversity in wild barley, in order to reveal the important effect of gene evolution in the adaptive evolution of plants.

2. RELATED THEORY AND RESEARCH STATUS

2.1. The Data of Species Distribution

This study focused on examining the influences of climatic factors on the diversity of genes, especially factors like temperature and humidity . The wild barley was taken as the research object, and "Genetic diversity" and "Wild Barley" were chosen as key words. Web of Science (http://isiknowledge.com) and NCBI PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) databases were used to search the related information. 422 papers were found; 17 genes (Table **1**) and 44 wild barley lines were selected. 495 sequences, related to the 17 genes were obtained from NCBI Genebank.

These sequences were aligned using CLUSTALX [12] and primarily analyzed by DNASP [13]. Geographical Information System Arc/Info was used for all geographical analyses.

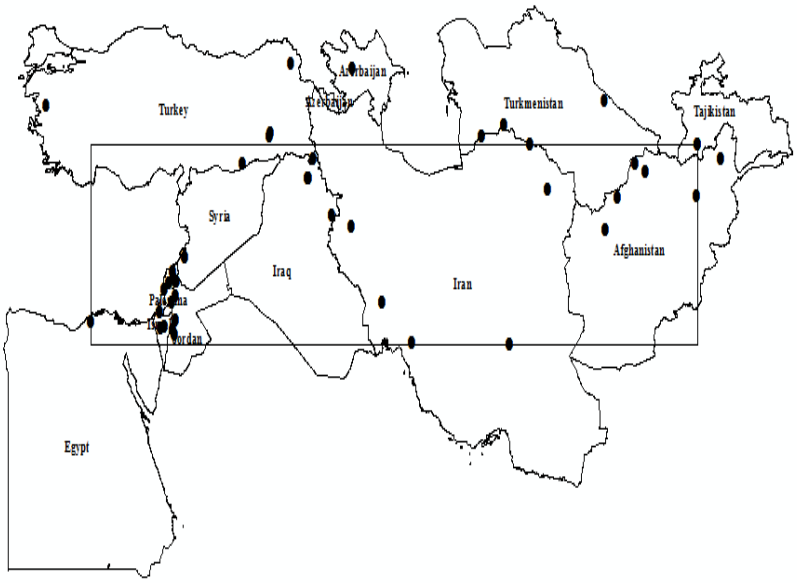


Fig. (1). The distribution of 44 wild barley populations.

The latitude and longitude of 44 wild barley populations were obtained through the GENESYS website. The distribution of these samples involved in this paper was graphically displayed by the GIS system (Fig. **1**). The rectangular area in the middle was chosen as the research scope according to finite sample frequency of 17 genes.

2.2. Climate Data Source

All climatic data were obtained from the world climate database (WORLDCLIM, http://www.worldclim.org/) [14]. From the year 1950 to 2000, climate data of 50 years from different weather stations around the world were pooled together. Using interpolation method, the global climate data was created in which, 19 biological climatic factors were observed to play important roles in the distribution of species . Detailed information is given as follows:

BIO1 = Annual Mean Temperature

BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))

BIO3 = Isothermality (BIO2/BIO7) (* 100)

BIO4 = Temperature Seasonality (standard deviation *100)

BIO5 = Max Temperature of Warmest Month

BIO6 = Min Temperature of Coldest Month

BIO7 = Temperature Annual Range (BIO5-BIO6)

BIO8 = Mean Temperature of Wettest Quarter

BIO9 = Mean Temperature of Driest Quarter

BIO10 = Mean Temperature of Warmest Quarter

BIO11 = Mean Temperature of Coldest Quarter

BIO12 = Annual Precipitation

BIO13 = Precipitation of Wettest Month

BIO14 = Precipitation of Driest Month

BIO15 = Precipitation Seasonality (Coefficient of Variation)

BIO16 = Precipitation of Wettest Quarter

BIO17 = Precipitation of Driest Quarter

BIO18 = Precipitation of Warmest Quarter

BIO19 = Precipitation of Coldest Quarter

A quarter is a period of three months (1/4 of the year)

3. MATERIALS AND METHODS

3.1. The Calculation of Differentiation Index

In a particular gene, the variable positions could be classified as singletons and nonsingletons, in which, singleton is a unique substitution ofa specific DNA sequence.

Considering a set of n aligned sequences, the differentiation index for the r-th sequence was calculated by

$$\pi_r = [\sum_{j=1}^{n} \pi_{rj}] / m_T(n-1) \quad (1)$$

Where , n is the number of homologous sequences in the specific gene. If one sequence r is considered for example, π_{rj} is the number of mismatches of sequence r compared with sequence j, and m_T is the length of the sequence r. Therefore, the differentiation index π_r represents the ratio between the sum of mismatched numbers and product of the length of the sequence and the homologous sequence number.

The interval of differentiation index was between 0 and 1, when the differentiation index was 0 which indicated that the sequence was totally identical with other homologous sequences, otherwise, 1 meant that it was completely different. The index was used in this paper to calculate the differentiation degree of the same gene in different wild barley populations.

3.2. Detection of Geographic Genetic Clines

All the investigated accessions of wild barley were sampled in 13 different countries in Middle-East, including Isreal, Jordan, Turkey *etc*. The Ordinary Kriging Interpolation, a spatial interpolation between sample points which uses information from the values of the sample points (z-axis in a Cartesian coordinate system) and the distances between them (x- and y-axes), was applied to quantify the spatial variation of the 17 genes [15]. In this study, the geographic genetic clines were classified through the Kriging maps, which were created by ArcGIS 10.

3.3. Analysis of the Correlations Between Gene Differentiation Index and Climate

Firstly, the regular meshes (Fishnet) were created in the selected area (Fig. **1**) by ArcGIS software. Using DIVA-GIS software, the point data of climate variables were extracted from the fishnet grid data. Principal component analysis of 19 climatic variables in the region was conducted by SPSS 13 software [16], highlighting that the main climatic factors may affect differentiation index of the wild barley populations in the region. The Ordinary Least Squares (OLS) in ArcGIS was used to construct the local regression model, to analyze the relationship between the differentiation coefficient of the 17 genes with major climatic factors.

4. EXPERIMENTAL RESULTS

4.1. Assessment of Molecular Data with GIS

In this paper, the differentiation indexes of 44 wild barley populations were used as the basis; the gene differentiation indexes in the research field were calculated through Ordinary Kriging Interpolation; the relationship between biological climatic factors and gene differentiation indexes was analyzed.

The distribution of wild barley accessions, number of variable sites in the DNA sequences and the differentiation indexes of 17 genes were visually represented through ArcGIS.

As shown in Fig. (**2**) s, *DHN4* was taken as an example. The geographical distribution of gene population, number of singleton and the different indexes. were graphically displayed. Throughout the analysis, the differentiation index of all genes, except *ADH1*, was observed to be much higher in the wild barley population distributed in Israel and Jordan, which may be greatly associated with the region's Canyon topography and climatic change [17].

4.2. Effect of Main Climate Factors in the Region

The climatic variable data in the research region was extracted using DIVA-GIS, and principal component analysis was conducted using SPSS software. Experimental results showed that cumulative contribution rate of 4 main components in the research area couldbe 91.44%, signifying biological climatic variable information.

The principal component 1 included information on Bio5, Bio10, Bio1 and Bio9, mainly reflected as the temperature in driest and warmest period; the principal component 2 included information on Bio6, Bio11 and Bio8, mainly reflected as the temperature in wettest and coldest period; the principal component 3 included information on Bio19, Bio15, Bio16 and Bio13, mainly reflected as the precipitation in the wettest and coldest period; principal component 4 included information on Bio4 and Bio7, reflected as changes in temperature differences. Based on the principal component analysis, Bio5, Bio6, Bio19 and Bio4 were selected to represent the 4 main components respectively. Meanwhile, the relationship between the gene differentiation indexes and 4 main climate factors was also studied. The principal component analysis of 19 biological climate factors in the research area is shown in Table **2**.

4.3. Spatial Dependence Between Genes and Climate Factors

Using OLS (ordinary least squares) regression method [18], the spatial correlation between the gene differentiation indexes and the main climatic factors is shown in Fig. (**3**).

In the past studies, it was reported that there was a direct relationship between the plant's DHN family and environmental stresses, such as drought and heat. Whereas, the alpha-amylase type B gene was not reported to have any significant relationship with drought or temperature.

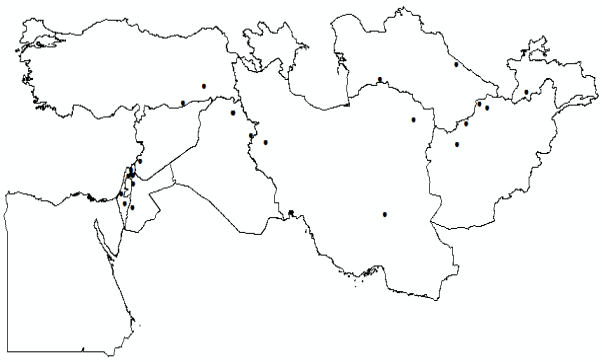
According to the experimental results and analysis, the OLS regression result indicated a significant relationship between the DHN family and 4 main climatic factors, while there was less correlation between alpha-amylase type B gene and the main climate factor.

CONCLUSION

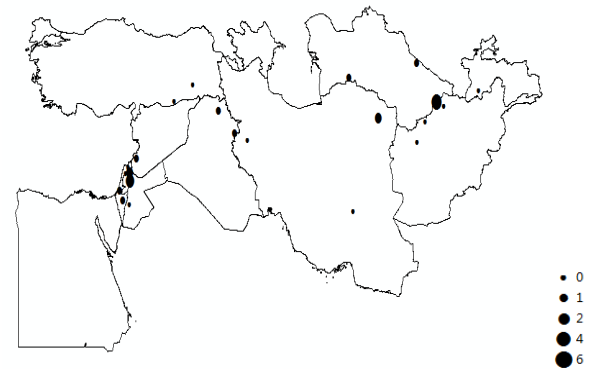
In this article, wild barley was used as the research object, and "Genetic diversity" and "Wild Barley" were chosen as key words; 17 genes and 44 wild barley populations were studied through Meta-analysis. The differentiation indexes of the same gene in different wild barley accessions were calculated, and the geographic distributions of these data were graphicallyrepresented. The information on 19 biological climatic factors was extracted, and principal component analysis was conducted using SPSS software. OLS was used to analyze the relationship between the main climatic factors and gene differentiation indexes, which determined the degree of gene evolution. The study found a significant spatial correlation between the 4 main

Table 2. The principal component analysis of 19 environment factors in the research area.

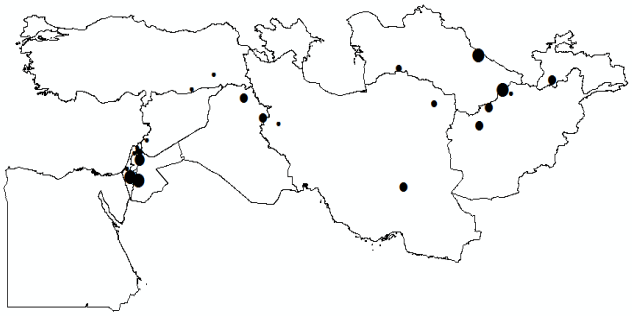
The Total Explained Variance				
Factor	**Initial Eigen Value**			
	Total	**Variance %**	**Total %**	
1	8.843	46.544	46.544	
2	4.232	22.272	68.816	
3	2.470	13.002	81.818	
4	1.828	9.621	91.439	
Environment	Factor			
	1	2	3	4
bio1	.856	.487	.040	.137
bio2	.593	-.599	.097	-.063
bio3	.591	.083	.015	-.635
bio4	.114	-.622	.113	.756
bio5	.889	.190	.105	.370
bio6	.667	.735	.008	.006
bio7	.208	-.808	.127	.477
bio8	.610	.557	-.266	.157
bio9	.856	.349	.084	.319
bio10	.876	.334	.074	.328
bio11	.777	.622	.015	-.029
bio12	-.786	.394	.411	.179
bio13	-.690	.376	.588	.100
bio14	-.655	.405	-.476	.220
bio15	.565	-.142	.628	-.230
bio16	-.693	.380	.591	.102
bio17	-.702	.398	-.449	.216
bio18	-.702	.348	-.453	.161
bio19	-.579	.421	.666	.065



A. Distribution of DHN4 wild barley accessions

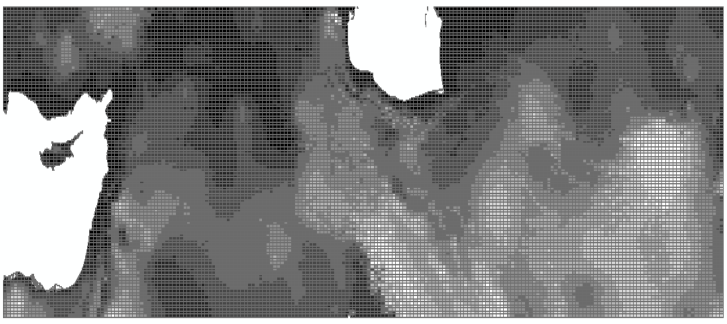


B. Distribution of singletons in DHN4 sequenced genomic region.



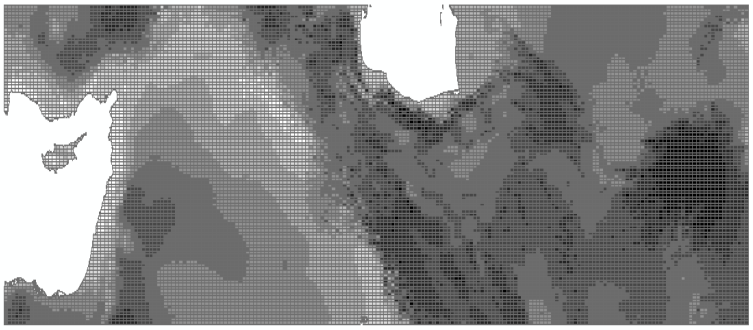
C. Distribution of differentiation index in DHN4 sequenced genomic region.

Fig. (2). The geographical distribution of DHN4 wild barley accessions, singleton and differentiation indexes. The size of the dots is **equivalen**t to the number of singletons/differentiation index.

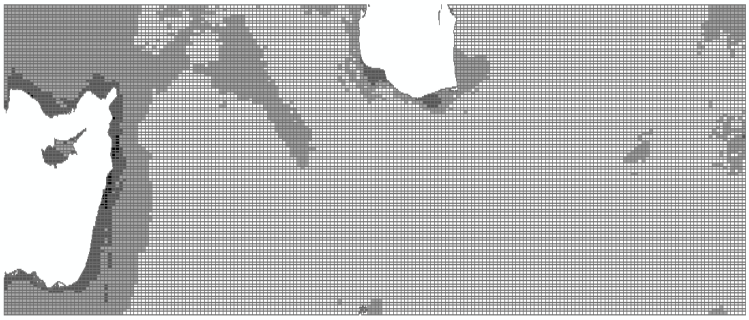


A. The spatial correlation between the DHN1 gene differentiation indexes and the major climatic factors.

Fig. (**3**). Contd…



B. The spatial correlation between the DHN4 gene differentiation indexes and the major climatic factors.



C. The spatial correlation between the alpha-amylase type B gene differentiation indexes and the major climatic factors.

Fig. (3). The spatial correlation between the gene differentiation indexes and the major climatic factors.

climatic factors with the differentiation indexes in stress-related genes, such as DHN family. Whereas, the genes reported not to be involved in drought or heat tolerance were found to have little correlation with main climatic factors. Thus, it can be seen that, Bio4, Bio5, Bio6 and Bio9 play important roles in gene evolution and in drought related stresses. In addition, this method could be used to predict the functions of unknown genes in the future.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Olson MS, Levsen N, Soolanayakanahally RY, *et al*. The adaptive potential of Populus balsamifera L. to phenology requirements in a warmer global climate. Mol Ecol 2013; 22(5): 1214-30.

[2] Zhou Y, Liu J, Savolainen O. Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. Mol Ecol 2014; 23(14): 3504-22.

[3] Wilczek AM, Cooper MD, Korves TM, Schmitt J. Lagging adaptation to warming climate in Arabidopsis thaliana. Proce Nat Acad Sci USA 2014; 111(22): 7906-13.

[4] Cronin JK, Bundock PC, Henry RJ, Nevo E. Adaptive climatic molecular evolution in wild barley at the Isa defense locus. Proc Natl Acad Sci USA 2007; 104(8): 2773-8.

[5] Strommer J. The plant ADH gene family. Plant J 2011; 66(1): 128-42.

[6] de Bruxelles GL, Peacock WJ, Dennis ES, Dolferus R. Abscisic acid induces the alcohol dehydrogenase gene in Arabidopsis. Plant Physiol 1996; 111(2): 381-91.

[7] Yang Y, He M, Zhu Z, *et al*. Identification of the dehydrin gene family from grapevine species and analysis of their responsiveness to various forms of abiotic and biotic stress. BMC Plant Biol 2012; 12: 140.

[8] Tommasini L, Svensson JT, Rodriguez EM, *et al*. Dehydrin gene expression provides an indicator of low temperature and drought stress: transcriptome-based analysis of barley (Hordeum vulgare L.). Funct Integr Genom 2008; 8(4): 387-405.

[9] Hoffmann MH, Glass AS, Tomiuk J, Schmuths H, Fritsch RM, Bachmann K. Analysis of molecular data of Arabidopsis thaliana (L.) Heynh (Brassicaceae) with Geographical Information Systems (GIS). Mol Ecol 2003; 12(4): 1007-19.

[10] Fand BB, Kumar M, Kamble AL. Predicting the potential geographic distribution of cotton mealybug Phenacoccus solenopsis in India based on MAXENT ecological niche model. J Environ Biol 2014; 35(5): 973-82.

[11] Elebead FM, Hamid A, Hilmi HS, Galal H. Mapping cancer disease using geographical information system (GIS) in Gezira State-Sudan. J Commun Health 2012; 37(4): 830-9.

[12] Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997; 25(24): 4876-82.

[13] Rozas J, Rozas R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 1999; 15(2): 174-5.

[14] Cuervo PF, Rinaldi L, Cringoli G. Modeling the extrinsic incubation of Dirofilaria immitis in South America based on monthly and continuous climatic data. Vet Parasitol 2015; 209(1-2): 70-5.

[15] Liu W, Du P, Wang D. Ensemble learning for spatial interpolation of soil potassium content based on environmental information. PLoS One 2015; 10(4): e0124383.

[16] Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods 2013; 18(2): 137-50.

[17] Zhang T, Li GR, Yang ZJ, Nevo E. Adaptive evolution of duplicated hsp17 genes in wild barley from microclimatically divergent sites of Israel. Genet Mol Res 2014; 13(1): 1220-32.

[18] Ugrinowitsch C, Fellingham GW, Ricard MD. Limitations of ordinary least squares models in analyzing repeated measures data. Med Sci Sports Exerc 2004; 36(12): 2144-8.

Received: October 11, 2014 Revised: December 30, 2014 Accepted: December 30, 2014