Applied Biological Research 20(1): 10-26; (2018)
DOI: 10.5958/0974-4517.2018.00002.2

BIOCOMPUTATIONAL MAPPING OF INTRASPECIES EVOLUTIONARY DISTANCES USING GEOGRAPHICAL INFORMATION SYSTEM

Mohammad Mousaei Sanjerehei[1*] and Philip W. Rundel[2]

[1]*Department of Plant Protection, Yazd Branch, Islamic Azad University, Yazd, Iran*
[2]*Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, California, United States of America*
*e-mail: *mmusaei@iauyazd.ac.ir*

(Received 29 January, 2017; accepted 20 December, 2017)

ABSTRACT

The evolutionary distances from an individual to other individuals of the same plant species were mapped in geographical information system (GIS). To do this, the relationship between the intraspecies evolutionary distances and the bioclimatic, geographic and elevation distances were determined for 17 plant species through multiple linear regression (MLR). Evolutionary distances were calculated by constructing Neighbor-joining and Maximum likelihood-based phylogenetic tree. Power, inverse-power, log10 and Box-Cox transformations were used for both evolutionary and environmental distances. The use of bioclimatic, geographic and elevation predictors and power transformation for MLR models resulted in explaining more than 60% of variation in intraspecies evolutionary distances for 12 plant species. Power transformation method was more effective than other methods for increasing models R^2. Comparison of MLR and distance-based redundancy analysis (db-RDA) indicated the efficiency of MLR in predicting intraspecies evolutionary distances based on environmental distances.

Key words: db-RDA, GIS, intraspecies evolutionary distance, power transformation

INTRODUCTION

Intraspecies evolutionary distance is defined in terms of the proportion of sites that differ between the sequences of the taxa (i.e. individuals representing same species). Determination of relationship between the evolutionary/genetic divergence and environmental variables is important for evaluating environmental processes driving population structure (Diniz-Filho *et al*., 2013), assessing the response of species to ecogeographic variations, studying the influence of geographic and climatic variables on the genetic structure and differentiation of species (Yang *et al.*, 2013), analysing the adaptation of organisms to different habitats (Yang *et al*., 2013) and prioritization of areas of high value for conservation (Bagnoli *et al*., 2011). A variety of methods including distanced-based redundancy analysis [db-RDA] (Legendre and Anderson, 1999), Mantel test (Mantel, 1967) and its derived forms, partial Mantel test (Smouse *et al.*, 1986) and Mantel correlogram (Oden and Sokal, 1986), the test of congruence among several distance matrices (Legendre and Lapointe, 2004) and linear regression (Donoso, 2012) have been developed and applied for modelling the relationship between the matrix of evolutionary/genetic distances and the matrix of geographic and/or climatic

distances. These methods have been used to determine whether the species and populations that are geographically or climatically close, are either genetically or phylogenetically similar (James *et al.*, 2015). Numerical simulations show that the power of linear correlation, regression and canonical analysis is far greater than that of the Mantel test and derived forms (Legendre and Fortin, 2010). The db-RDA and linear regression are more powerful and informative methods of spatial analysis than Mantel tests conducted with distance matrices. The fitted values of regression or RDA models can be mapped, thus providing a visual representation of the structure at different spatial scales (Legendre *et al.*, 2015). Another potential advantage of linear regression and RDA is that they provide measures of variance around parameter estimates, which allows for improved interpretation of results over Mantel tests that only provide a correlation coefficient and P-value (Legendre and Fortın, 2010; Kierepka and Latch, 2015).

Geographical Information System (GIS) is one of the powerful tools for modelling the effective environmental factors in studies of genetic and environmental relationships (Michels *et al.*, 2001). It can be used to visualize and map the spatial distribution of genetic indices and the environmental drivers of spatial genetic patterns (Manel *et al.*, 2003). The objectives of this research were to evaluate the performance of multiple linear regression in modelling the relationship between the evolutionary distances and bioclimatic, geographic and elevation distances in comparison to db-RDA and to produce GIS-based map of intraspecies evolutionary distances using regression model based on a new method.

MATERIALS AND METHODS

Plant species and environmental data

The relationship between the evolutionary distances and the bioclimatic, geographic (latitude and longitude) and elevation distances was determined for 17 plant species throughout the world. The name of the plant species, geographic coordinates of the location of plant individuals and the GenBank accession numbers are presented as under:

Species name	Latitude and longitude	GenBank accession No.	Reference	Species name	Latitude and longitude	GenBank accession No.	Reference
A. cristatum	39.40 N, 37.08 E	GQ373300	Dizkirici *et al.* (2010)	*L. microphyllum*	5.00 N, 2.00 W	DQ845212	Madeira *et al.* (2008)
A. cristatum	40.73 N, 43.42 E	GQ373302		*L. microphyllum*	26.96 S, 32.83 E	DQ845213	
A. cristatum	39.57 N, 33.44 E	GQ373304		*L. microphyllum*	16.3 S, 145.2 E	DQ845211	
A. cristatum	39.70 N, 42.29 E	GQ373305		*L. microphyllum*	26.8 N, 87.30 E	DQ845210	
A. cristatum	38.61 N, 34.74 E	GQ373306		*L. microphyllum*	22.4 N, 114.1 E	DQ546209	
A. vitaliana	37.05 N, 3.35 W	EU109186	Dixon *et al.* (2009)	*L. microphyllum*	8.54 N, 77.36 E	DQ845205	
A. vitaliana	40.07 N, 1 W	FJ654497		*L. microphyllum*	12.7 S, 143.2 E	DQ845206	
A. vitaliana	40.22 N, 5.59 W	EU109184		*L. microphyllum*	27 N, 80.1 E	DQ845207	
A. vitaliana	43.02 N, 4.38 W	EU109183		*L. microphyllum*	28.1 N, 82.3 E	DQ845208	
A. vitaliana	42.81 N, 0.65 W	FJ654496		*L. japonica*	29.05 N, 120.43 E	KF160912	Wang *et al* (2014)
A. vitaliana	42.8 N, 0.44 W	FJ654495		*L. japonica*	27.62 N, 119.05 E	KF160890	
A. vitaliana	42.77 N, 0.4 W	EU109188		*L. japonica*	27.45 N, 119.5 E	KF160892	
A. vitaliana	42.68 N, 0.39 W	FJ654494		*L. japonica*	28.43 N, 119.9 E	KF160893	
A. vitaliana	42.77 N, 0.12 E	FJ654493		*L. japonica*	27.97 N, 119.62 E	KF160896	
A. vitaliana	42.59 N, 0.69 E	FJ654492		*L. japonica*	28.13 N, 120.28 E	KF160904	
A. vitaliana	42.63 N, 1.12 E	EU109187		*L. jonesii*	37.75 N,111.71W	EF367726	Larson *et al.* (2010)
A. vitaliana	42.37 N, 2.15 E	FJ654491		*L. jonesii*	37.9 N,111.45 W	EF367728	
A. vitaliana	44.08 N, 6.64 E	EU109190		*L. jonesii*	38.17 N,110.6 W	EF367730	
A. vitaliana	44.34, N, 6.85 E	FJ654501		*L. jonesii*	38.25 N,110.61W	EF367732	
A. vitaliana	44.56 N, 6.76 E	EU109193		*L. jonesii*	38.46 N,110.89W	EF367734	
A. vitaliana	44.69 N, 6.98 E	FJ654503		*L. jonesii*	38.86 N,110.36W	EF367736	
A. vitaliana	44.84 N, 7.01 E	FJ654505		*L. jonesii*	39.45 N110.37 W	EF367737	

A. vitaliana	44.86 N, 7.09 E	FJ654500	
A. vitaliana	45.05 N, 6.94 E	EU109194	
A. vitaliana	44.63 N, 6.04 E	FJ654504	
A. vitaliana	45.06 N, 6.37 E	FJ654502	
A. vitaliana	45.98 N, 7.78 E	EU109196	
A. vitaliana	46.36 N, 7.66 E	EU109195	
A. vitaliana	46.26 N, 8.06 E	FJ654499	
A. vitaliana	46.5 N, 11.66 E	FJ654498	
A. vitaliana	46.48 N, 11.82 E	EU109198	
A. vitaliana	46.65 N, 12.75 E	EU109197	
A. vitaliana	42.1 N, 14.06 E	EU109199	
C. aprica	55.47 N, 61.22E	AY546109	
C. aprica	47.67 N, 92.5 E	AY546121	
C. aprica	51.17 N, 71.5 E	AY546117	
C. aprica	45.5 N, 97.08 E	AY546124	
C. aprica	55.00 N, 100.32 E	AY546132	
C. aprica	49.50 N, 96.00 E	AY546131	
C. aprica	49.55 N, 106.22 E	AY546130	
C. aprica	48.88 N, 95.75 E	AY546129	
C. aprica	48.27N, 102.97 E	AY546135	
C. aprica	48.27 N, 106.55 E	AY546123	
C. aprica	49.2 N, 86.35 E	AY546126	
C. aprica	55.47 N, 61.22 E	AY546111	
C. aprica	48.02 N, 91.62 E	AY546133	
C. aprica	62.42 N, 131.8 E	AY546125	
C. aprica	50.35N, 90.5 E	AY546118	
C. aprica	53.82 N, 91.4 E	AY546127	
C. aprica	53.25N,112.77 E	AY546122	
C. aprica	52.37N, 115.57E	AY546134	
C. aprica	41.7 N, 44.92 E	AY546110	
C. aprica	52.88 N, 56.6 E	AY546112	
C. aprica	53.42 N, 49.5 E	AY546128	
C. aprica	51.67 N, 49.38 E	AY546113	
C. aprica	51.67 N, 39.47 E	AY546114	
C. aprica	53.37 N, 76.30 E	AY546115	
C. aprica	52.47 N, 61.87 E	AY546116	
C. aprica	67.95 N, 134.1 E	AY546136	
C. aprica	54.38 N, 64.78 E	AY546119	
C. aprica	53.73 N, 77.65 E	AY546120	Franzke *et al.* (2004)
D. polylepis	35.30 N, 58.94 E	KC595350	
D. polylepis	37.44 N, 58.25 E	KC595314	
D. polylepis	36.15 N, 59.69 E	KC595317	
D. polylepis	36.17 N, 59.60 E	KC595351	
D. polylepis	37.47 N, 58.73 E	KC595352	
D. polylepis	36.84 N, 59.61 E	KC595336	
D. polylepis	35.50 N, 59.19 E	KC595339	Farsi *et al.* (2013
F. crassifolia	35.27 N, 46.19 E	KM435178	
F. crassifolia	37.72 N, 47.70 E	KM435179	
F. crassifolia	37.76 N, 47.64 E	KM435180	
F. crassifolia	37.64 N, 47.47 E	KM435182	
F. crassifolia	37.70 N, 47.51 E	KM435181	
F. crassifolia	37.60 N, 47.53 E	KM435184	
F. imperialis	37.33 N, 55.38 E	KM435147	
F. imperialis	33.15 N, 50.40 E	KM435156	
F. imperialis	33.19 N, 50.33 E	KM435157	
F. imperialis	32.54 N, 50.23 E	KM435158	
F. imperialis	32.81 N, 50.27 E	KM435159	
F. imperialis	31.78 N, 51.02 E	KM435160	
F. imperialis	32.18 N, 50.36 E	KM435161	
F. imperialis	35.30 N, 46.20 E	KM435162	
F. imperialis	35.28 N, 46.20 E	KM435163	
F. imperialis	34.33 N, 46.12 E	KM435164	Khourang *et al.* (2014)
L. jonesii	39.88 N,110.0 W	EF367739	
L. jonesii	40.07 N, 110.0 W	EF367741	
L. jonesii	40.35 N,109.67W	EF367743	
L. jonesii	40.87 N,109.11W	EF367745	
L. jonesii	40.2 N, 110.78 W	EF367747	
L. montanum	38.45 N,113.27W	EF367914	
L. montanum	39.13 N,113.1 W	EF367916	
L. montanum	43.04 N,112.54W	EF367943	
L. montanum	42.12 N,113.4 W	EF367945	
L. montanum	42.06 N,113.43W	EF367946	
L. montanum	42.09 N, 113.5 W	EF367947	
L. montanum	43.88 N, 116.6 W	EF367948	
L. montanum	42.68 N,117.8 W	EF367949	
L. montanum	42.01 N,117.8 W	EF367950	
L. montanum	40.97 N,114.1 W	EF367952	
L. montanum	41.78 N,113.49 W	EF367954	
L. montanum	41.1 N, 115 W	EF367956	
L. montanum	41.43 N,114.65 W	EF367958	
L. montanum	41.39 N,117.24 W	EF367966	
L. montanum	41.43 N, 117.4 W	EF367967	
L. montanum	41.18 N,105.72W	EF367969	Larson *et al.* (2010)
P. duclouxii	26.15 N, 100.28 E	JF746388	
P. duclouxii	26.25 N, 100.52 E	JF746390	
P. duclouxii	24.8 N, 103.43 E	JF746384	
P. duclouxii	28.42 N, 100.92 E	JF746395	
P. duclouxii	27.48 N, 100.38 E	JF746393	
P. duclouxii	28.48 N, 102.43 E	JF746396	Dong *et al.* (2011)
P. multisectum	36.46 N, 108.03 E	FJ752662	
P. multisectum	38.8 N, 101.08 E	FJ752660	
P. multisectum	37.08 N, 104.02 E	FJ752658	
P. multisectum	36.02 N, 103.96 E	FJ752659	
P. multisectum	37.49 N, 107.49 E	FJ752665	Zhao *et al.* (2011)
Q. coccifera	38.9 N, 1.25 E	DQ342349	
Q. coccifera	39.49 N, 2.95 E	DQ342347	
Q. coccifera	39.6 N, 19.93 E	DQ342352	
Q. coccifera	39.94 N, 18.27 E	AY226834	
Q. coccifera	36.59 N, 4.85 W	DQ342348	
Q. coccifera	40.26 N, 3.45 W	DQ342346	
Q. coccifera	41.5 N, 1.6 W	DQ342345	de Heredia *et al.* (2007)
Q. ilex	42.61 N, 2.75 W	DQ342360	
Q. ilex	39.84 N, 2.9 E	DQ342356	
Q. ilex	40.26 N, 18.4 E	DQ342358	
Q. ilex	33.73 N, 4.12 W	DQ342357	
Q. ilex	37.95 N, 5.75 W	DQ342354	
Q. ilex	42.23 N, 12.67 E	AY226836	
Q. ilex	39.92 N, 4.25 E	DQ342355	
Q. ilex	42.3 N, 9.15 E	DQ342350	
Q. ilex	38.98 N, 0.28 W	DQ342351	
Q. ilex	43.4 N, 6.07 E	DQ342353	de Heredia *et al.* (2007)
Q. suber	39.91 N, 5.29 W	AY827130	
Q. suber	40.5 N, 3.77 W	AY827129	
Q. suber	36.9 N, 7.25 E	AY827137	
Q. suber	41.42 N, 13.82 E	AY827111	
Q. suber	39.66 N, 2.51 E	AY827145	
Q. suber	39.12 N, 16.17 E	AY827121	
Q. suber	36.97 N, 8.89 E	AY827136	
Q. suber	39.35 N, 7.4 W	AY827128	
Q. suber	42.51 N, 3.06 E	AY827126	
Q. suber	38.48 N, 8.58 W	AY827132	
Q. suber	40.73 N, 8.56 E	AY827115	
Q. suber	43.77 N, 1.32 W	AY827125	
Q. suber	39.22 N, 6.18 W	AY827143	de Heredia *et al.* (2007)

F. imperialis	34.35 N, 46.16 E	KM435165	*Q. suber*	33.77 N, 4.1 W	AY827134	
F. imperialis	33.11 N, 49.42 E	KM435166	*Q. suber*	39.9 N, 4.24 E	AY827142	
F. imperialis	33.14 N, 47.53 E	KM435167	*Q. suber*	40.63 N, 17.93 E	AY827138	
F. imperialis	33.13 N, 47.53 E	KM435169	*Q. suber*	39.89 N, 0.37 W	AY827141	
F. persica	33.97 N, 49.55 E	KM435148	*S. latifolia*	55.37 N, 14.05 E	FN821124	Rautenberg *et al.* (2010)
F. persica	32.50 N, 50.26 E	KM435171	*S. latifolia*	38.78 N, -9.45 W	FN821126	
F. persica	32.18 N, 50.36 E	KM435172	*S. latifolia*	53.98 N, 23.1 E	FN821127	
F. persica	35.54 N, 48.74 E	KM435173	*S. latifolia*	41 N, 39.7 E	FN821129	
F. persica	32.22 N, 50.26 E	KM435174	*S. latifolia*	49.02 N, 1.15 E	FN821130	
F. persica	35.69 N, 48.67 E	KM435175	*S. latifolia*	37.68 N, 15.08 E	FN821131	
F. persica	35.65 N, 48.69 E	KM435176	*S. latifolia*	61.1 N, 21.78 E	FN821133	
			S. latifolia	43.32 N, 2.98 W	FN821134	

The gene regions of plant samples studied for sequences: 18 S rRNA gene, partial sequence; ITS1, 5.8 S rRNA gene, and ITS2, complete sequence; and 28 S rRNA gene, partial sequence for *Lonicera japonica, Silene latifolia, Fritillaria crassifolia, F. imperialis, F. persica* and *Lepidium montanum* var. *montanum* species individuals, tRNA-Leu (trnL) gene and trnL-trnF intergenic spacer; partial sequence, chloroplast for *Lygodium microphyllum*, ITS1, 5.8 S rRNA gene, and ITS2, complete sequence for *Pterygiella duclouxii*, ITS1, partial sequence for *Clausia aprica*, ITS1, partial sequence; 5.8 S rRNA gene, complete sequence; and ITS2, partial sequence for *Androsace vitaliana* and *Agropyron cristatum*, ITS1, 5.8 S rRNA gene, and ITS2, complete sequence for *Quercus coccifera, Q. ilex* and *Q. suber*, tRNA-His (trnH) gene and trnH-PsbA intergenic spacer, partial sequence; chloroplast for *Peganum multisectum*, 1 gene, partial CDS for *Dianthus polylepis*, and tRNA-Leu (trnL) gene, partial sequence; trnL-trnF intergenic spacer, complete sequence; and tRNA-Phe (trnF) gene, partial sequence; chloroplast for *Lepidium montanum* var. *jonesii*.

The 20 bioclimatic variables used in this study included annual mean temperature (AMT), mean diurnal range (MDR), isothermality (Iso), temperature seasonality (TS), maximum temperature of warmest month (MTWM), minimum temperature of coldest month (MTCM), temperature annual range (TAR), mean temperature of wettest quarter (MTWeQ), mean temperature of driest quarter (MTDQ), mean temperature of warmest quarter (MTWaQ), mean temperature of coldest quarter (MTCQ), annual precipitation (AP), precipitation of wettest month (PWM), precipitation of driest month (PDM), precipitation seasonality (PS), precipitation of wettest quarter (PWeQ), precipitation of driest quarter (PDQ), precipitation of warmest quarter (PWaQ), precipitation of coldest quarter (PCQ) and extraterrestrial radiation (Ra). The (30 arc-seconds) map of 19 bioclimatic variables (1950-2000) and elevation of the world was obtained from Worldclim website (*http://www.worldclim.org*). The map of extraterrestrial radiation was produced using the latitude of the pixels of world map based on the FAO equation (Allen *et al*., 1998). To produce the world raster map of latitude and longitude, the world raster map was converted to a point layer and the decimal latitude and longitude of points was used as the pixel value of raster map. The rasterized point layer of the sampling location of plants was crossed to the map of bioclimatic and elevation variables to obtain the value of bioclimatic and elevation variables at the sampling location of the plants. The analyses were performed with ILWIS 3.2.

Calculation of intraspecies evolutionary distances

To obtain the matrix of intraspecies evolutionary distance between the individuals of each plant species, the phylogenetic tree was constructed by maximum likelihood and Neighbor joining methods. The evolutionary distance between two individuals was calculated by summing the lengths of branches connecting each pair of individuals according to the scale at the bottom of the phylogenetic tree (Shimada *et al.*, 2009). To construct the phylogenic tree using Neighbor joining method, the nucleotide sequences were obtained from the GenBank (*http://www.ncbi.nlm.nih.gov*) using the GenBank accession number. The sequences were aligned using Muscle (*http://www.ebi.ac.uk/Tools/msa/muscle*), and the aligned sequences were used to generate the phylogenetic tree based on the Neighbor joining method from the ClustalW2 tool (*http://www.ebi.ac.uk/Tools/phylogeny/clustalw2_phylogeny*). To generate the phylogenetic tree based on maximum likelihood method, the sequences were aligned by Muscle and the tree was constructed using PhyML program at the phylogeny.fr platform (*http://www.phylogeny.fr*).

Linear regression analysis and data transformation

The relationship between the intraspecies evolutionary distances and the environmental (bioclimatic, geographic and elevation) distances for each plant species was determined using multiple linear regression model and distance-based redundancy analysis (db-RDA). To relate the intraspecies evolutionary distances and the environmental distances using a multiple linear regression analysis, the environmental distances were calculated as the difference of the value of each explanatory variable between the sampling locations of plant individuals for each species. For example, 91 distances ($n \times n$-1)/2) for mean annual temperature was calculated using the difference of the value of this variable between 14 sampling points of *F. imperialis* individuals. All data were tested for normality of distribution using Shapiro–Wilk test.

To normalize the distribution of the data and to increase the value of R-squared, 4 methods of data transformation were applied. The methods included power transformation (X^P) and inverse-power transformation $1/(X+1)^P$ where P was 0.1, 0.2, ……5, $\log_{10}$ (X+1) and Box-Cox transformation. Power, inverse power and log transformation were applied for both evolutionary and environmental distances, while Box-Cox method was used for evolutionary distances and also both evolutionary and environmental distances. The optimal lambda was obtained in the Box-Cox method and the data were transformed as ($X^{Lambda-1}$/Lambda).

To avoid collinearity between the predictors, a correlation matrix for the 23 predictors' distances was constructed for each plant species, and the predictors with the correlation coefficient (r) of less than 0.8 (Riordan and Rundel, 2014) were selected for multiple regression analysis. Among the predictors with $r > 0.8$, the one with the highest correlation with the intraspecies evolutionary distance data was selected for the regression analysis. This was done for both untransformed and transformed data. In spite of keeping one variable among the highly correlated variables (with $r >$ 0.8), a collinearity may be present in the regression model specifically, following the data transformation. To prevent this problem, the predictor(s) with a high variance inflation factor (VIF > 3) and a low influence on the variation of the dependent variable in the regression model were excluded from the analysis. Since a large number of predictors in the regression model can cause overfitting and a decrease in the prediction accuracy and thereby model misidentification (Fan and Lv, 2010), for the plant species with a small sample size (i.e., *P. multisectum* with a sample size of 10), maximum two explanatory variables were selected for the regression model. To determine the significant predictors, the backward method was used for multiple regression analysis. SPSS software was used for regression analysis.

Due to the effect of sample size on the power of regression models, the power of all regression models was calculated using G*Power 3.1 program. The power of regression models with more than one predictor in the model, was calculated based on F-test using α error probability (0.05), total sample size and the effect size. The effect size was calculated as $R^2/(1- R^2)$, where R^2 is the partial R^2 of the predictors. The power of regression models with one predictor was calculated based on t-test using correlation coefficient, standard deviation of the dependent and independent variables, total sample size and α error probability (0.05).

Validation of the models

To compare the observed evolutionary distance and predicted distance by the regression model, the parametric paired sample t-test and the Wilcoxon signed rank nonparametric test were calculated. The nonparametric test was used because the distribution of evolutionary distance data for some plant species specifically those with a small sample size may not be normalized following the data transformation.

Distance-based redundancy analysis

The results of MLR were compared to db-RDA. To perform the db-RDA, principal coordinate analysis (PCoA) was performed on the matrix of evolutionary distance and the resulting sample scores was entered into RDA after correction for negative eigenvalues. The matrix of explanatory variables was constructed using the value of the bioclimatic, latitude, longitude and elevation

variables at the sampling points of the plants. The forward selection method was used for reducing the large set of explanatory variables and selecting the best and uncorrelated predictors on the basis of maximum extra fit. The statistical significance of each selected variable was evaluated by a Monte-Carlo permutation test (499 permutations). Canoco for Windows 4.5 was used for the analysis.

GIS map of intraspecies evolutionary distances

The linear regression model with a high and significant R^2 can be reliably used to produce the GIS map of intraspecies evolutionary distances. We produced the world map of intraspecies evolutionary distances for two plant species (as an example) using the raster map of the significant predictor(s) in the regression model. To produce the map of evolutionary distances from an individual to other individuals of the same species based on multiple linear regression, the following steps were used:

1. A) To produce the map of environmental distance (i.e., temperature distance), the absolute of the difference between the value of each pixel and the value of pixel at the location of one of the plant individuals was calculated for the environmental variable (i.e., temperature map). B) If a transformation (i.e., power) on both evolutionary distance and the environmental distance (i.e., temperature distance) was used for obtaining the MLR, this transformation was applied on the value of pixels of the environmental distance map produced in step A.
2. C) The map of evolutionary distances from the plant individual to other individuals of the same species was produced using the environmental distance map(s) and the MLR. D) If a transformation (i.e., power) was used on both evolutionary and environmental distances for obtaining MLR, the value of pixels of the evolutionary distance map produced in step C was then transformed back to the original scale using inverse of the transformation (i.e., 1/power) applied to the data.

RESULTS

Maximum likelihood and Neighbor joining-based phylogenetic trees

Fig. 1 and 2 depict the phylogenetic trees constructed by Maximum likelihood and Neighbor joining techniques, respectively, for 17 studied plant species. The ratio of length of tree branches between individuals as a measure of intraspecies evolutionary distance was equal by Maximum likelihood and Neighbor joining techniques for *A. cristatum, F. crassifolia, P. multisectum, L. jonesii, L. montanum* and *P. duclouxii* and highly significantly correlated ($r>0.9$, Sig.= 0.000) by these two techniques for other plant species.

Multiple linear regression models

Table 1 shows the results of multiple linear regression models in which the evolutionary distances were calculated by constructing the Neighbor joining-based phylogenetic tree (MLR_{NJ}). The R-squared based on untransformed and transformed data, the significant predictor(s) based on untransformed data and the value of the power and inverse-power by which the highest R^2 was obtained for the MLR_{NJ} models are presented. The linear scaling was applied as $(X-X_{min}/ X_{max}- X_{min})$ for transforming the data to a 0-1 scale. The test of Shapiro-Wilk indicated a normal distribution of intraspecies evolutionary distance data for *D. polylepis*, *Q. coccifera* and *Q. ilex* ($P>0.05$) and an abnormal distribution of evolutionary distances for other species. The distribution of evolutionary distances obtained by both maximum likelihood and Neighbor joining-based phylogenetic tree was similar. R^2 indicating the proportion of variation in the evolutionary distances, explained by the predictors, was 0.2-0.4 for 4, 0.4-0.6 for 6, 0.6-0.8 for 5 and 0.8-1 for 2 plant species using MLR_{NJ} models. The power transformation could increase the value of R^2 for 10 plant species models from 15% (for *F. imperialis*) to 77% (for *F. crassofolia*). However, R^2 was not changed for 7 plants species by power transformation. The power by which the highest R^2 was obtained was in range of 0.3-3.0. The significant predictor(s) selected by the backward regression procedure were same based on

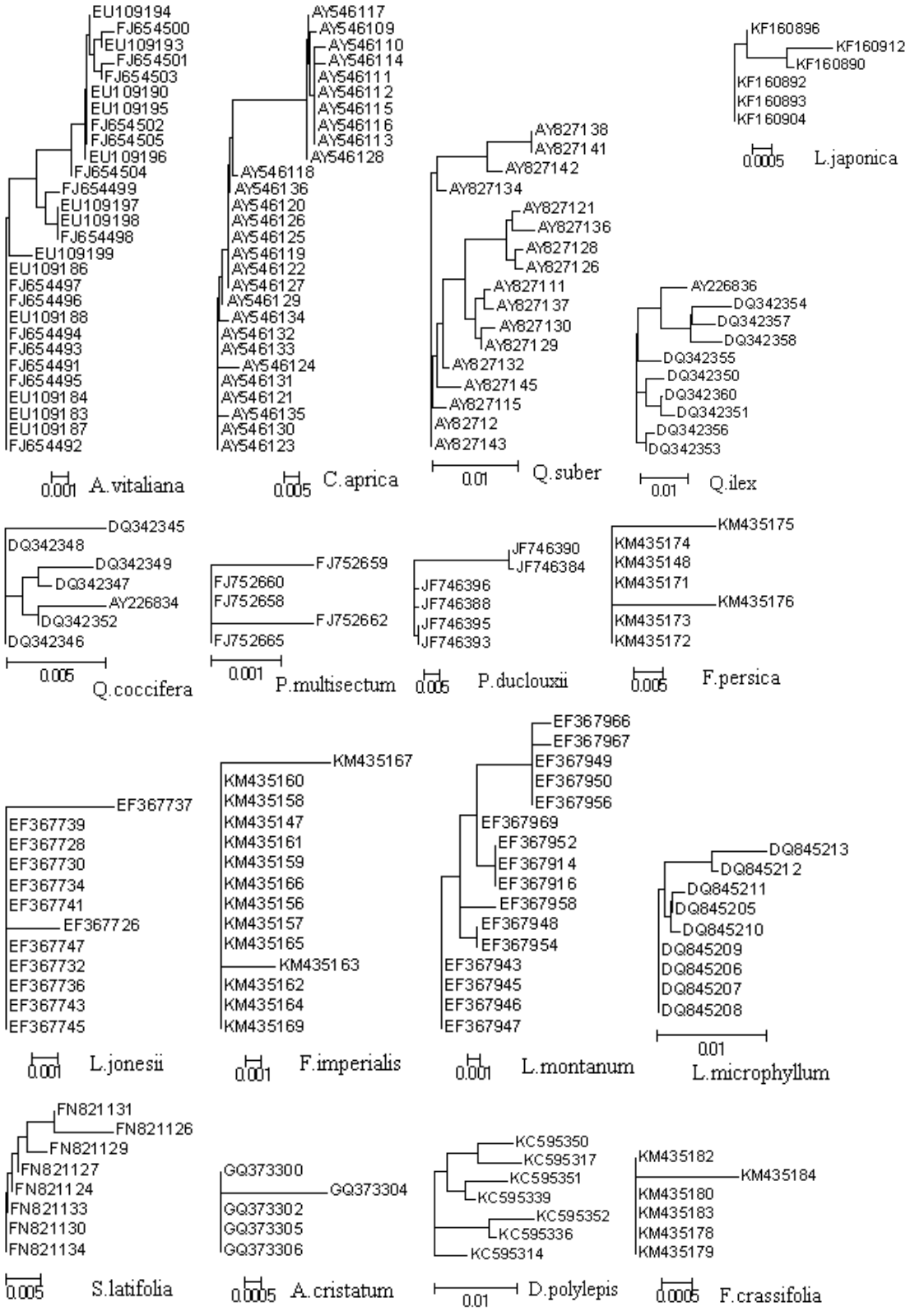

Fig. 1: The intraspecies phylogenetic trees for 17 studied plant species constructed by Neighbor joining method. The phylogenetic trees were visualized from their Newick formats using MEGA 6.06.

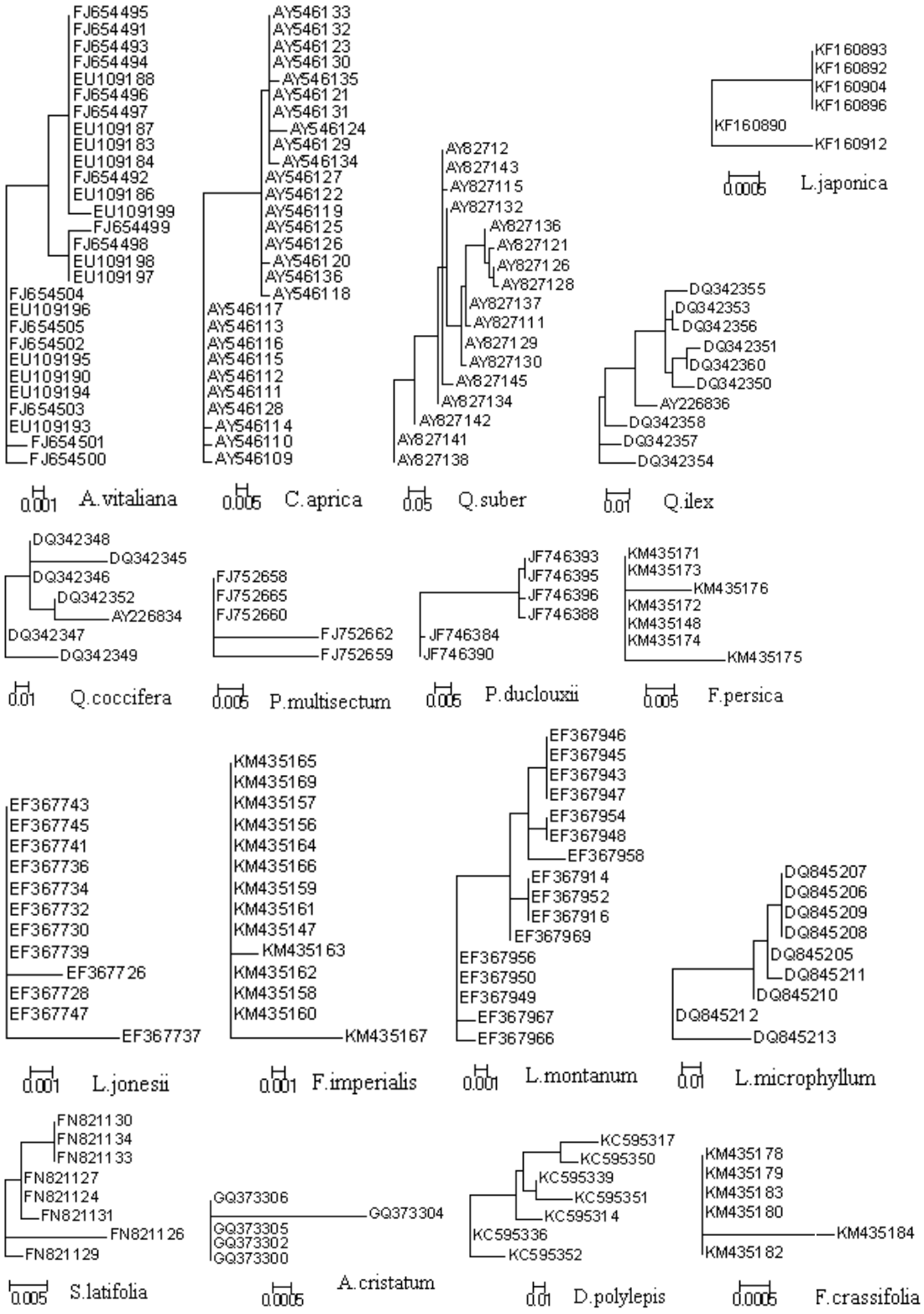

Fig. 2: The intraspecies phylogenetic trees for 17 studied plant species constructed by Maximum likelihood method. The phylogenetic trees were visualized from their Newick formats using MEGA 6.06.

Table 2: Multiple regression analysis in which evolutionary distances (dependent variable) were calculated using the constructed Neighbor joining-based phylogenetic tree (MLR_{NJ}) and maximum likelihood-based phylogenetic tree (MLR_{ML}). The model R^2 based on untrans-formed, power: (X^P), inverse-power: $1/(X+1)^P$, $\log_{10}(X+1)$ and Box-Cox transformed data, the significant predictor (SP) ($p<0.05$) based on untransformed data and the value of power (P) and inverse power (P) by which the highest R^2 was obtained are presented. All transformations were used on both dependent and independent variables, but Box-Cox transformation was used for dependent (1) and for both dependent & independent variables (2)

	MLR_{NJ}								
Species	Untransformed data		Power-transformed data		$\log_{10}$-transformed data	Inver power-transformed data		Box-Cox (1)	Box-Cox (2)
	SP	R^2	P	R^2	R^2	P	R^2	R^2	R^2
A. cristatum	MTWaQ	0.83	1	0.83	0.62	0.1	0.58	0.83	0.78
A. vitaliana	Lon, Ra, PCQ, MTDQ, TAR, TS	0.50	0.4	0.69	0.53	0.1	0.5	0.51	0.51
C. aprica	Lon, PDM, MTWaQ, Ra, MTDQ, Iso	0.64	1	0.64	0.53	0.1	0.51	0.64	0.6
D. ploylepis	TAR, Lon	0.42	2.4	0.55	0.26	0.1	0.26	0.42	0.36
F. crassifolia	MTCQ, Iso	0.34	2.5	0.60	0.38	0.1	0.38	0.34	0.3
F. imperialis	MTWM	0.20	1.8	0.23	0.12	0.1	0.11	0.2	0.15
F. persica	PDQ, AP, E	0.67	0.5	0.8	0.76	0.1	0.76	0.67	0.73
L. japonica	PS, Lat, PCQ	0.58	0.3	0.79	0.72	0.1	0.72	0.58	0.65
L. jonesii	Iso, MDR, TAR	0.25	3	0.32	0.15	0.1	0.14	0.25	0.17
L. microphyllum	PWeQ, PWaQ, Lon	0.60	1	0.60	0.25	0.1	0.24	0.6	0.55
L. montanum	Iso, E, PDM, MTCM	0.54	1	0.54	0.54	0.1	0.53	0.54	0.52
P. duclouxii	PS, MDR, Lat	0.69	0.4	0.76	0.69	0.1	0.68	0.68	0.59
P. multisectum	PDM, PWaQ	0.52	3	0.73	0.53	0.1	0.53	0.52	0.53
Q. coccifera	MTWaQ, PWM, MDR, MTWeQ	0.8	1	0.80	0.75	0.2	0.7	0.79	0.76
Q. ilex	TS, PDM, PWaQ, Lon	0.72	1	0.72	0.58	0.1	0.55	0.72	0.68
Q. suber	TS, MTWeQ, PCQ, E	0.28	1	0.28	0.19	0.1	0.17	0.28	0.25
S. latifolia	PWaQ, PS, E	0.57	3	0.90	0.22	0.1	0.21	0.57	0.53
	MLR_{ML}								
Species	SP	R^2	P	R^2	R^2	P	R^2	R^2	R^2
A. cristatum	MTWaQ	0.83	1	0.83	0.62	0.1	0.58	0.83	0.78
A. vitaliana	Lon, Ra, PCQ, MTDQ, TAR, TS	0.50	0.4	0.67	0.52	0.1	0.5	0.5	0.5
C. aprica	Lon, PDM, MTWaQ, Ra, MTDQ, Iso	0.64	1	0.64	0.53	0.1	0.52	0.64	0.6
D. ploylepis	TAR, Lon	0.39	4	0.61	0.37	0.1	0.36	0.39	0.31
F. crassifolia	MTCQ, Iso	0.34	2.5	0.6	0.38	0.1	0.38	0.34	0.3
F. imperialis	MTWM	0.32	1	0.32	0.23	0.1	0.22	0.32	0.26
F. persica	PDQ, AP, E	0.81	0.7	0.86	0.86	0.1	0.85	0.81	0.71
L. japonica	PS, Lat, PCQ	0.67	0.5	0.82	0.80	0.1	0.81	0.67	0.76
L. jonesii	Iso, MDR, TAR	0.25	3	0.32	0.15	0.1	0.14	0.25	0.17
L. microphyllum	PWeQ, PWaQ, Lon	0.64	2	0.68	0.28	0.1	0.27	0.58	0.61
L. montanum	Iso, E, PDM, MTCM	0.54	1	0.54	0.54	0.1	0.53	0.54	0.52
P. duclouxii	PS, MDR, Lat	0.69	0.4	0.76	0.69	0.1	0.68	0.68	0.59
P. multisectum	PDM, PWaQ	0.52	3	0.73	0.53	0.1	0.53	0.52	0.53
Q. coccifera	MTWaQ, MTWeQ, PWM, MDR	0.60	1	0.6	0.68	0.1	0.68	0.59	0.61
Q. ilex	PWaQ, TS, PDM	0.67	1	0.67	0.60	0.1	0.56	0.67	0.67
Q. suber	TS, PCQ, MTWeQ, Lat	0.29	1.6	0.32	0.20	0.1	0.2	0.22	0.2
S. latifolia	PS, Iso, MTWeQ	0.57	1	0.57	0.51	0.1	0.43	0.57	0.52

untransformed and power-transformed data. Log_{10} transformation could increase the model R^2 from 6 to 24% for 4, did not change it for 3, and decreased it for 10 plant species from 6 to 62%. The changes in R^2 by inverse power-transformed data were approximately similar to that by log-transformed data. The power by which the highest R^2 was obtained by inverse-power transformed data was 0.1 for all the plant species models. The Box-Cox transformation on evolutionary distances did not change R^2 for any of the models, but the use of this method on both evolutionary and environmental distances resulted in an increase of R^2 for 2, a decrease of R^2 for 13 and a lack of change in R^2 for 2 plant species models.

The R^2 for multiple linear regression models in which evolutionary distances were calculated by constructing the maximum likelihood-based phylogenetic tree (MLR_{ML}) was 0.2-0.4, 0.4-0.6, 0.6-0.8 and 0.8-1 for 5, 4, 6 and 2 plant species, respectively (Table 1). The R^2 was increased from 6 (for *F. persica*) to 76% (for *F. crassofolia*) using the power-transformed data for 10 plant species models, but it was not changed for 7 plant species. The significant predictors selected by the backward regression analysis were the same using both power-transformed and untransformed data for all plant species. The power by which the highest R-squared was obtained was 0.4-4. Log_{10} transformation resulted in an increase of 4-19%, a decrease of 5-56% and a lack of change in R^2 for 5, 9 and 3 plant species models, respectively. The changes in R^2 by inverse-power and log_{10} transformed data were approximately similar. The power by which the highest R^2 was obtained by inverse-power transformed data was 0.1 for all plant species models (Table 1). Box-Cox transformation on evolutionary distances did not change the R^2, but on both evolutionary and environmental distances led to decrease, increase and a lack of change in R^2 for 11, 1 and 5 plant species models, respectively. For both MLR_{NJ} and MLR_{ML}, the increase in R^2 by power transformation was higher than and the same as that by the other methods for 10 and 7 plant species models, respectively. The MLR_{ML} gave higher R^2 for 8, lower R^2 for 4 and equal R^2 for 5 plant species models in comparison to the $MLR_{NJ.}$ The significant predictors were same by both MLR_{ML} and MLR_{NJ} models for all studied plant species except for *L. jonesii, Q. ilex* and *S. latifolia* which were different in most influential predictor.

MLR models with the best fit

Of the two regression models MLR_{NJ} and $MLR_{ML,}$ made for each plant species, the one with higher R^2 based on untransformed or power-transformed data is presented in Table 2, with the significant predictor(s), coefficient, significance and VIF of the predictors, the value of the power by which the highest R^2 was obtained, the R^2 by adding predictor, sample size, the power of regression model and the result of paired t test and Wilcoxon signed rank test. The VIF of the explanatory variables was 1-2 for 15 models and 2-3 for 2 models indicating no collinearity in the models. The power of regression was >0.9 for 16 models and 0.64 for 1 model representing no overfitting and approximately enough sample size for confidence in the results. The result of Paired t-test and Wilcoxon signed rank test was similar for 15 plant species models and indicated a non-significant difference ($p>0.05$) between the observed evolutionary distances and the predicted evolutionary distances using the regression analysis. For two plant species (*F. imperialis* and *L. jonesii*), the paired t-test showed insignificant difference, but Wilcoxon signed rank test indicated a significant difference ($p<0.05$) between the observed and predicted values. The results of Wilcoxon signed rank test appeared to be more confident than that of paired t-test because of low R^2 of 0.32 for the models of 2 species and an abnormal distribution of data in spite of applying the transformed data.

The map of evolutionary distances from one individual to other individuals of the same species can be produced based on the MLR with a high and significant R^2. The map of evolutionary distance from an individual of *Q. coccifera* based on MLR_{NJ} and from an individual of *L. japonica* based on MLR_{ML} produced using the raster map of the significant predictor(s) in the regression models (Fig. 3 and 4).

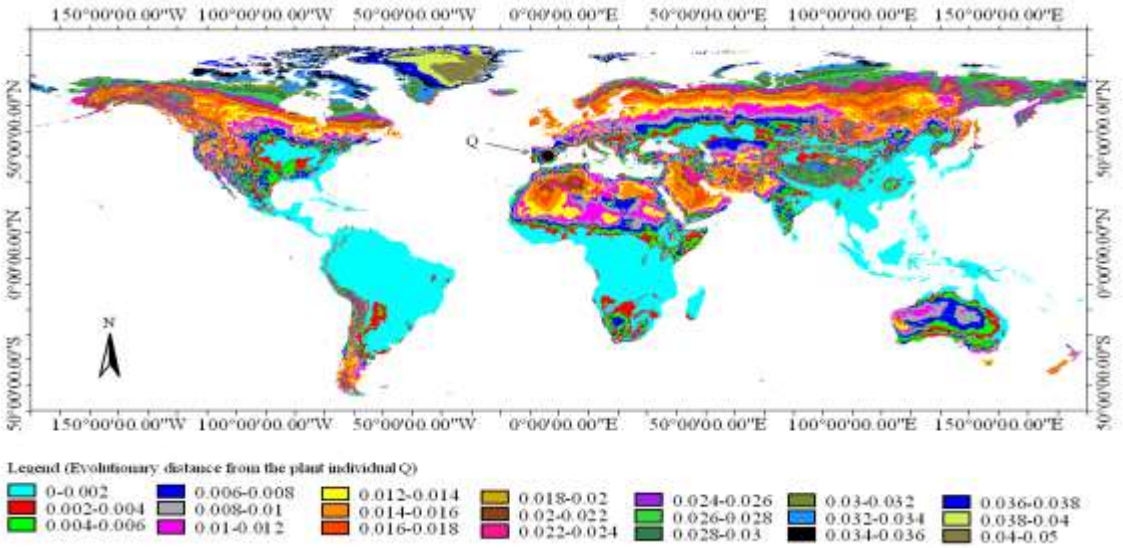

Fig. 3: The map of evolutionary distances from an individual of *Q. coccifera* to other individuals of this species based on the linear regression: (Evolutionary distance = 0.00016 MTWaQ -3.13× 10^{-5} PWM+7.33× 10^{-5}MDR-5.04× 10^{-5} MTWeQ + 0.004). This equation was obtained using untransformed data. The map of the environmental distance (MTWaQ, PWM, MDR and MTWeQ distance map) was produced using calculating the absolute of difference between the value of each pixel and the value of pixel at the location of one of the *Q. coccifera* individuals in the predictor map.

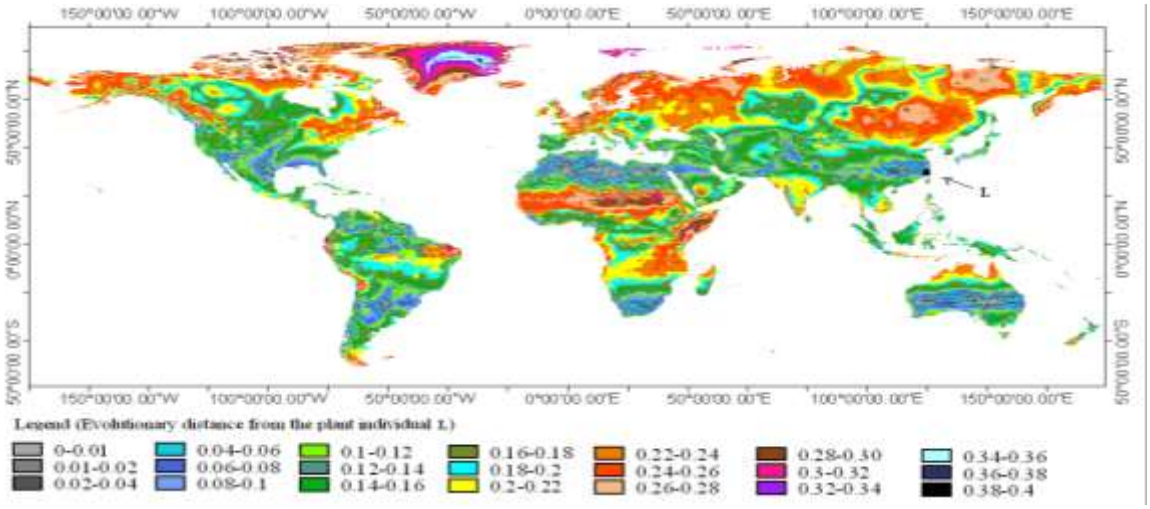

Fig. 4: The map of evolutionary distances from an individual of *L.japonica* to other individuals of this species based on linear regression: Evolutionary distance = [(0.0206×PS-0.0054×PCQ+0.0388 ×Lat-0.0106)]. This equation was obtained using power transformed data (power = 0.5) on evolutionary, PS, PCQ and lat distances. To produce PS, PCQ and Lat distance map, the value of each pixel was subtracted from value of the pixel at the location of one of the *L.japonica* individuals in PS, PCQ and Lat map, and the absolute difference was calculated and square rooted. The map of evolutionary distances was then produced by transforming back the value of the pixels of produced map to original scale using squaring (1/power) the values.

Db-RDA models

The db-RDA in which matrix of evolutionary distances was calculated using construction of maximum likelihood-based phylogenetic tree (db-RDA_{ML}) provided an R^2 of 0.2-0.4 for 1, 0.4-0.6 for 3, 0.6-0.8 for 5 and 0.8-1 for 3 plant species (Table 3). For 5 plant species (*F. crassifolia, F. persica, L. japonica, L. jonesii* and *Q. coccifera*) models, no significant variable ($p<0.05$) was

Table 3: The model with best fit for predicting evolutionary distances. Of MLR_{ML} and MLR_{NJ}, the one with higher R^2 is presented 1) MLR_{NJ}, (2) MLR_{ML} and (1,2) MLR_{ML} and MLR_{NJ} gave the same R^2, but the result of MLR_{ML} are presented). The significant predictor (SP, P<0.05), coefficient (C), significance and VIF of the predictors, the value of power by which highest R^2 was obtained (VP), the R^2 by adding predictor, sample size (SS), the power of regression model (PR) and the result of paired t test (PT) and Wilcoxon signed rank (W) test are presented

Species	SP	C	Sig.	VIF	VP	R^2	SS	PR	PT	W
A. cristatum (1,2)	MTWaQ	0.009	0.001	1	1	0.83	10	0.999	0.84	0.72
		C = -0.085								
A. vitaliana (1)	Lon	0.04	0.000	1.99	0.4	0.28	378	0.999	0.06	0.08
	Ra	0.064	0.000	2.61		0.40				
	PCQ*	0.007	0.000	1.63		0.51				
	MTDQ*	-0.008	0.000	1.85		0.60				
	TAR*	-0.009	0.000	1.90		0.66				
	TS*	-0.003	0.000	1.97		0.69				
		C = -0.002								
C. aprica (1,2)	Lon	0.0003	0.000	2.98	1	0.20	378	0.999	0.22	0.57
	PDM	0.0007	0.000	2.26		0.38				
	MTWaQ*	0.0001	0.000	1.35		0.47				
	Ra*	-0.0016	0.000	2.4		0.61				
	MTDQ*	-3.8×10^{-5}	0.000	2.05		0.63				
	Iso*	0.00057	0.010	2.05						
		C = 0.002								
D. ploylepis (2)	TAR	1.07×10^{-12}	0.000	1.78	4	0.38	21	0.91	1	0.64
	Lon*	-1.96×10^{-7}	0.004	1.78		0.61				
		$C = 2.15\times10^{-7}$								
F. crassifolia (1,2)	MTCQ	2×10^{-11}	0.007	1.12	2.5	0.43	15	0.64	0.99	1
	Iso*	-7.2×10^{-8}	0.050	1.12		0.60				
		$C= 4.5\times10^{-8}$								
F. imperialis (2)	MTWM	0.007	0.000	1	1	0.32	91	0.999	0.84	0.007
		C= -0.073								
F. persica (2)	PDQ	0.036	0.000	1.38	0.7	0.58	21	0.999	0.99	0.64
	AP*	-0.0029	0.001	1.73		0.73				
	E*	0.00058	0.001	1.30		0.86				
		C = -0.0097								
L. japonica (2)	PS	0.0206	0.004	1.05	0.5	0.55	15	0.960	0.56	0.89
	Lat*	0.0388	0.013	1.20		0.71				
	PCQ*	-0.0054	0.002	1.19		0.82				
		C =-0.0106								
L. jonesii (1,2)	Iso	4.65×10^{-10}	0.001	1.09	3	0.17	66	0.92	0.92	0.02
	MDR*	1.3×10^{-12}	0.027	1.17		0.26				
	TAR*	-6.06×10^{-14}	0.031	1.23		0.32				
		$C = 4.87\times10^{-9}$								
L. microphyllum (2)	PWeQ	2.99×10^{-10}	0.000	1.36	2	0.47	36	0.99	0.99	0.4
	PWaQ*	-1.14×10^{-10}	0.004	1.35		0.60				
	Lon*	3.88×10^{-9}	0.022	1.02		0.68				
		$C = 1\times10^{-5}$								
L. montanum (1,2)	Iso	0.00055	0.000	1.09	1	0.35	120	0.999	0.65	0.42
	E*	-2.65×10^{-6}	0.006	2.20		0.43				
	PDM*	0.00022	0.006	1.40		0.50				
	MTCM*	4.69×10^{-5}	0.050	2.60		0.54				
		C = 0.002								
P. duclouxii (1,2)	PS	0.093	0.001	1.37	0.5	0.45	15	0.93	0.96	0.78
	MDR*	-0.074	0.009	1.31		0.68				
	Lon*	0.084	0.05	1.15		0.76				
		C = 0.05								
P. multisectum (1,2)	PDM	7.88×10^{-6}	0.004	1.90	3	0.50	10	0.999	0.98	0.96
	PWaQ	-2.4×10^{-11}	0.014	1.90		0.73				
		$C= 7.27\times10^{-6}$								

Q. coccifera (1)	MTWaQ	0.00016	0.000	1.2	1	0.41	21	0.95	0.47	0.46
	PWM	-3.13×10^{-5}	0.020	1.2		0.62				
	MDR*	7.33×10^{-5}	0.001	1.08		0.72				
	MTWeQ*	-5.04×10^{-5}	0.02	1.06		0.80				
		C = 0.004								
Q. ilex (1)	TS	4.2×10^{-6}	0.006	1.17	1	0.38	45	0.999	0.93	0.89
	PDM	-0.0003	0.001	1.61		0.58				
	PWaQ	0.00017	0.000	1.76		0.67				
	Lon	0.00025	0.033	1.04		0.72				
		C=0.008								
Q. suber (2)	TS	-1.24×10^{-8}	0.000	1.03	1.6	0.15	136	0.98	0.99	0.34
	PCQ	1.62×10^{-7}	0.006	1.03		0.22				
	MTWeQ*	1.44×10^{-6}	0.02	1.02		0.28				
	Lat*	-5.26×10^{-5}	0.011	1.02		0.32				
		C= 0.002								

*tested predictor for calculating the power of the regression model.

Table 4: The results of db-RDA in which the evolutionary distances (dependent variable) were calculated using constructing maximum likelihood (db-RDA$_{ML}$) and neighbor joining (db-RDA$_{NJ}$)-based phylogenetic tree. The significant predictor (SP), R^2 by adding predictor, significance of the predictor and the power of the db-RDA models (P) are presented.

Species	db-RDA$_{ML}$				db-RDA $_{NJ}$			
	SP	R^2	Sig.	P	SP	R^2	Sig.	P
A. cristatum	MTWaQ	0.88	0.002	0.999	MTWaQ	0.88	0.002	0.999
A. vitaliana	Iso,	0.25	0.002	0.999	Iso	0.18	0.002	0.999
	MTDQ	0.40	0.002		MTDQ	0.28	0.002	
	Ra*	0.48	0.016		Ra	0.34	0.014	
	MDR*	0.56	0.006		MDR*	0.40	0.006	
	PS*	0.63	0.018		PS*	0.46	0.05	
					TAR*	0.54	0.024	
C. aprica	PDM	0.57	0.002	0.999	PDM	0.57	0.002	0.999
	MTWaQ*	0.62	0.05		MTWaQ*	0.62	0.036	
D. ploylepis	MTCM	0.47	0.006	0.977	MTCM	0.5	0.004	0.991
F. crassifolia	-	-	-	-	-	-	-	-
F. imperialis	MTWM	0.45	0.05	0.999	MTWM	0.39	0.048	0.999
F. persica	-	-	-	-	-	-	-	-
L. japonica	-	-	-	-	-	-	-	-
L. jonesii	-	-	-	-	Iso	0.32	0.022	0.994
L. microphyllum	Lon	0.55	0.014	0.999	Lon	0.45	0.004	0.999
	Lat	0.86	0.004		Lat	0.65	0.01	
	Iso*	0.93	0.008		Iso*	0.79	0.02	
L. montanum	Iso	0.42	0.002	0.999	Iso	0.38	0.002	0.999
	PDQ	0.53	0.012		PDQ	0.50	0.01	
	Lat*	0.64	0.032		Lat*	0.60	0.032	
P. duclouxii	PS	0.76	0.046	0.999	PS	0.76	0.046	0.999
P. multisectum	PS	0.92	0.004	0.999	PS	0.92	0.004	0.999
Q. coccifera	-	-	-	-	MTWaQ	0.40	0.016	0.94
Q. ilex	TS	0.55	0.002	0.998	TS	0.41	0.002	0.999
	Iso*	0.71	0.046		Lon	0.60	0.05	
					PWaQ*	0.77	0.016	
Q. suber	PCQ	0.26	0.042	0.999	PCQ	0.26	0.042	0.93
S. latifolia	PDM	0.41	0.026	0.984	PS	0.46	0.02	0.880
					PWaQ*	0.61	0.005	

*****tested predictor for calculating the power of the regression model; - = no significant predictor was identified by db-RDA.

identified in db-RDA_{ML}. The power of db-RDA_{ML} models was > 0.977 for all plant species. R^2 for db-RDA in which the matrix of evolutionary distances was calculated using construction of Neighbor joining-based phylogenetic tree (db-RDA_{NJ}) was 0.2-0.4, 0.4-0.6, 0.6-0.8 and 0.8-1.0 for 3, 3, 6 and 2 plant species, respectively (Table 3). No significant variable was found by db-RDA_{ED-NJ} for 3 plant species (*F. crassifolia*, *F. persica* and *L. japonica*). The power of db-RDA_{NJ} models was >0.88 for all plant species models. As compared to the untransformed data-based MLR models, the db-RDA models provided a higher R^2 for 9 and a lower R^2 for 5 plant species. But in comparison to the power transformed data-based MLR, the db-RDA models gave a higher, lower and similar R^2 for 5, 6 and 3 plant species, respectively. The significant predictor(s) of each model identified by both MLR and db-RDA were similar for 3 and different for 1 plant species. For 10 plant species, some of the significant predictors in each model identified by MLR and db-RDA were similar, among which the predictor with most influence on the explained proportion of variance in evolutionary distances was same for 6 and different for 4 plant species.

DISCUSSION

Integration of genetic indices (i.e. evolutionary distances) and environmental factors through GIS-based mapping can provide in detail the distribution and evolutionary history of a species in the context of geoclimatic history of planet (Avise, 2000). We tried to obtain the best fitted linear regression model to explain a high proportion of variation in intraspecies evolutionary distances for 17 plant species. This was done by using geographic (latitude and altitude), elevation and 20 bioclimatic predictors, 4 methods of data transformation, 2 methods of phylogenetic tree construction, and considering multicollinearity and the power of regression models. The ratio of length of tree branches between individuals as an indicator of intraspecies evolutionary distance was equal by Maximum likelihood and Neighbor joining methods for *A. cristatum, F. crassifolia, P. multisectum, L jonesii, L. montanum* and *P. duclouxii* and highly significantly correlated for other plant species. The difference in the results of Maximum likelihood and Neighbor joining-based phylogenetic tree results from a drastic variation in the evolutionary rate among branches and in the number of sequences compared (Saitou and Imanishi, 1989).

Linear regression model was used to produce the map of evolutionary distances from an individual to other individuals of the same species by presenting a new method. Most of the studies for predicting genetic/evolutionary differentiation have used a small number of predictors such as geographic, precipitation and/or temperature (Vargas-Mendoza *et al*., 2015). Explanation of more than 60% variance in the intraspecies evolutionary distances for 12 (out of 17) species in this study revealed the efficiency of geographic, elevation and bioclimatic predictors in an accurate prediction of evolutionary distances. Power transformation resulted in an increase of R-squared for 10 models of MLR_{NJ} and MLR_{ML}. The performance of square root and $\log_{10}$ transformation has been proved in normalizing the distribution of genetic distance data and in increasing the model fit (James *et al*., 2015). Although $\log_{10}$ as well as Box-Cox and inverse-power transformation resulted in an increase of R^2 for a number of regression models in this study, these methods caused a decrease in R^2 for others. In addition, R^2 increase by the use of these methods was not higher than that by the power transformation for any of the models (except *Q. coccifera* for which MLR_{ML} model R^2 using inverse-power and $\log_{10}$ transformed data was 13% higher than that using power transformed data). MLR_{NJ} and MLR_{ML} gave same results for 5 plant species, but MLR_{ML} R^2 was higher for 8 and lower for 4 plant species than MLR_{NJ} R^2. Although a higher performance of maximum likelihood method than Neighbor joining method has been discussed for constructing phylogenetic tree and calculation of genetic distances (Yang and Rannala, 2012), it is not reliable to use only this method for all plant species and also the other organisms. Comparison of MLR and db-RDA indicated that MLR is as

efficient as db-RDA (Donoso, 2012; Legendre *et al.*, 2015). In spite of the difference in MLR and db-RDA R^2 for a number of plant species models, the species whose evolutionary distance between individuals was well modeled by db-RDA was also well modeled by MLR and vice-versa (except *L. microphyllum* and *Q. coccifera* for which db-RDA and MLR gave highly different results). However, db-RDA_{ML} and db-RDA_{NJ} identified no significant predictor for 5 and 3 plant species models, respectively. We suggest the use of different methods of constructing phylogenetic tree such as maximum likelihood, Neighbor joining, UPGMA and maximum parsimony for calculating evolutionary distance matrix and the use of power transformation with different powers (X^P) on both evolutionary and environmental distance data in multiple linear regressions. Soil attributes in addition to bioclimatic, geographic and elevation predictors can be used to increase the explained proportion of variation in genetic differentiation. Remote sensing images can be applied for determining a variety of soil properties. The steps for obtaining a reliable and highly fitted regression model and the new GIS-based mapping in this study can be used as a guideline for producing the map of intra- and inter-species evolutionary/genetic distances at different scales for plants, animals and microorganisms based on environmental predictors.

REFERENCES

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M. 1998. *Crop Evapotranspiration - Guidelines for Computing Crop Water Requirements - FAO Irrigation and drainage. Paper 56.* FAO, Rome, Italy.

Avise, J.C. 2000. *Phylogeography: The History and Formation of species*. Harvard University Press, London, England.

Bagnoli, F., Fady, B., Fineschi, S., Oddou-Muratorio, S., Piotti, A., Sebastiani, F., and Vendramin, G.G. 2011. Neutral patterns of genetic variation and applications to conservation in conifer species. pp. 141-195. **In**: *Genetics, genomics and breeding of conifers,* CRC Press and Science Publishers, United States.

de Heredia, U.L., Jiménez, P., Collada, C., Simeone, M.C., Bellarosa, R., Schirone, B., Cervera, M.T. and Gil, L. 2007. Multi-marker phylogeny of three evergreen oaks reveals vicariant patterns in the Western Mediterranean. *Taxon*, **56**: 1209-1209.

Diniz-Filho, J.A.F., Soares, T.N., Lima, J.S., Dobrovolski, R., Landeiro, V.L., Telles, M.P.D.C., Rangel, T.F. and Bini, L.M. 2013. Mantel test in population genetics. *Genetics and Molecular Biology*, **36**: 475-485.

Dixon, C.J., Schönswetter, P., Vargas, P., Ertl, S. and Schneeweiss, G.M. 2009. Bayesian hypothesis testing supports long-distance Pleistocene migrations in a European high mountain plant (Androsace vitaliana, Primulaceae). *Molecular Phylogenetics and Evolution*, **53**: 580-591.

Dizkirici, A., Kaya, Z., Cabi, E. and Dogan, M. 2010. Phylogenetic relationships of Elymus L. and related genera (Poaceae) based on the nuclear ribosomal internal transcribed spacer sequences. *Turkish Journal of Botany*, **34**: 467-478.

Dong, L.N., Wortley, A.H., Wang, H., LI, D.Z. and Lu, L. 2011. Efficiency of DNA barcodes for species delimitation: A case in Pterygiella Oliv.(Orobanchaceae). *Journal of Systematics and Evolution*, **49**: 189-202.

Donoso, D.A. 2012. Additions to the taxonomy of the *Armadillo* ants (Hymenoptera, Formicidae, Tatuidris. *Zootaxa*, **3503**: 61-81.

Fan, J. and Lv, J. 2010. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, **20**: 101-148.

Farsi, M., Behroozian, M., Vaezi, J., Joharchi, M.R. and Memariani, F. 2013. The evolution of Dianthus polylepis complex (Caryophyllaceae) inferred from morphological and nuclear DNA sequence data: one or two species?. *Plant Systematics and Evolution*, **299**: 1419-1431.

Franzke, A., Hurka, H., Janssen, D., Neuffer, B., Friesen, N., Markov, M. and Mummenhoff, K. 2004. Molecular signals for late tertiary/early quaternary range splits of an Eurasian steppe plant: *Clausia aprica* (Brassicaceae). *Molecular Ecology*, **13**: 2789-2795.

James, P., Cooke, B., Brunet, B.M., Lumley, L.M., Sperling, F.A., Fortin, M.J., Quinn, V.S. and Sturtevant, B.R. 2015. Life-stage differences in spatial genetic structure in an irruptive forest insect: Implications for dispersal and spatial synchrony. *Molecular Ecology*, **24**: 296-309.

Khourang, M., Babaei, A., Sefidkon, F., Naghavi, M.R., Asgari, D. and Potter, D. 2014. Phylogenetic relationship in *Fritillaria* spp. of Iran inferred from ribosomal ITS and chloroplast trnL-trnF sequence data. *Biochemical Systematics and Ecology*, **57**: 451-457.

Kierepka, E.M. and Latch, E.K. 2015. Performance of partial statistics in individual-based landscape genetics. *Molecular Ecology Resources*, **15**: 512-525.

Larson, S.R., Culumber, C.M., Schweigert, R.N. and Chatterton, N.J. 2010. Species delimitation tests of endemic *Lepidium papilliferum* and identification of other possible evolutionarily significant units in the *Lepidium montanum* complex (Brassicaceae) of western north America. *Conservation Genetics*, **11**: 57-76.

Legendre, P. and Anderson, M.J. 1999. Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, **69**: 1-24.

Legendre, P. and Fortin, M.J. 2010. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**: 831-844.

Legendre, P. and Lapointe, F.J. 2004. Assessing congruence among distance matrices: Single malt Scotch whiskies revisited. *Australian and New Zealand Journal of Statistics*, **46**: 615-629.

Legendre, P., Fortin, M.J. and Borcard, D. 2015. Should the Mantel test be used in spatial analysis?. *Methods in Ecology and Evolution*, **6**: 1239-1247.

Madeira, P.T., Pemberton, R.W. and Center, T.D. 2008. A molecular phylogeny of the genus *Lygodium* (Schizaeaceae) with special reference to the biological control and host range testing of *Lygodium microphyllum. Biological Control*, **45**: 308-318.

Manel, S., Schwartz, M.K., Luikart, G. and Taberlet, P. 2003. Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**: 189-197.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**: 209–220.

Michels, E., Cottenie, K., Neys, L., De Gelas, K., Coppin, P. and De Meester, L. 2001. Geographical and genetic distances among zooplankton populations in a set of interconnected ponds: A plea for using GIS modelling of the effective geographical distance. *Molecular Ecology*, **10**: 1929-1938.

Oden, N.L. and Sokal, R.R. 1986. Directional autocorrelation: an extension of spatial correlograms to two dimensions. *Systematic Zoology*, **35**: 608-617.

Rautenberg, A., Hathaway, L., Oxelman, B. and Prentice, H.C. 2010. Geographic and phylogenetic patterns in Silene section *Melandrium* (Caryophyllaceae) as inferred from chloroplast and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, **57**: 978-991.

Riordan, E.C. and Rundel, P.W. 2014. Land use compounds habitat losses under projected climate change in a threatened California ecosystem. *PloS one*, **9**: 1-15.

Saitou, N., and Imanishi, T. 1989. Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution and Neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Molecular Biology and Evolution*, **6**: 514-525.

Shimada, E., Kinoshita, M. and Murata, K. 2009. Expression of cardiac myosin light chain 2 during embryonic heart development in medaka fish, *Oryzia slatipes* and phylogenetic relationship with other myosin light chains. *Development, Growth and Differentiation*, **51**: 1-16.

Smouse, P.E. Long, J.C. and Sokal, R.R. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**: 627–632.

Vargas-Mendoza, C.F., Ortegón-Campos, I., Marrufo-Zapata, D., Herrera, C.M. and Parra-Tabla, V. 2015. Genetic diversity, outcrossing rate and demographic history along a climatic gradient in the ruderal plant *Ruellia nudiflora* (Acanthaceae). *Revista Mexicana de Biodiversidad*, **86**: 508-520.

Wang, Q., Quan, Q., Zhou, X., Zhu, Y., Lan, Y., Li, S., Yu, Y. and Cheng, Z. 2014. A comparative study of *Lonicera japonica* with related species: morphological characteristics, ITS sequences and active compounds. *Biochemical Systematics and Ecology*, **54**: 198-207.

Yang, J., Cushman, S.A., Yang, J., Yang, M. and Bao, T. 2013. Effects of climatic gradients on genetic differentiation of *Caragana* on the Ordos Plateau, China. *Landscape Ecology*, **28**: 1729-1741.

Yang, Z. and Rannala, B. 2012. Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, **13**: 303-314.

Zhao, T., Wang, Z.T., Branford-White, C.J., Xu, H. and Wang, C.H. 2011. Classification and differentiation of the genus *Peganum* indigenous to China based on chloroplast trnL-F and psbA-trnH sequences and seed coat morphology. *Plant Biology*, **13**: 940-947.