# GenoCline: On the trail of spatial patterns of genetic variation

Jose Peña<sup>1</sup>, Luis Gómez-Pérez<sup>1</sup>, and Miguel Alfonso-Sánchez<sup>1</sup>

<sup>1</sup>Universidad del País Vasco Facultad de Ciencia y Tecnología

June 9, 2020

## Abstract

The accurate determination of the spatial trends on the variability of the gene pool of a species is essential to elucidate the underlying demographic-evolutionary events and, ultimately, to unravel the microevolutionary history of the species or population under examination. In this work, we present a new software tool called GenoCline, which is essentially devised to assist in detecting genetic clines from allele and phenotypic frequency data, and even from genome-wide data. This program package allows identifying the geographic orientation of clinal genetic variation through a system of iterative rotation of a virtual coordinate axis. At the same time, GenoCline can also be used to carry out complementary statistical analyses aimed at clarifying the potential origin of the genetic clines found, among which stand out spatial autocorrelation, isolation by distance, centroid method, multidimensional scaling and Sammon projection. Among the main advantages of this software are those related to ease in data entry and potential interconnection with other programs. Data entry is user-friendly. Genetic frequencies and geographic coordinates can be easily entered in spreadsheet table formatting, while genome-wide data can be imported in Eigensoft format. Genetic frequencies can also be exported in a format compatible with other programs dealing with population genetics and evolutionary biology. All illustrations of results are saved in eps format so that there will be high quality and easily editable vectorial graphs available for the researcher. GenoCline is implemented in Java, so it can be used with different operating systems.

## Introduction

The genetic variability of a species is not uniformly represented across its distribution range. If a species spread over a wide region, it will probably be subdivided into subpopulations so that there will be different mating probabilities between individuals from distinct subpopulations (Kimura & Weiss, 1964). It is known that the existence of geographical barriers can promote local genetic differentiation among subpopulations. However, even if such physical barriers do not exist, a larger size of the distribution range of a species as compared with the average migration distance of the individuals may prevent the species from being a single panmictic unit, and this could lead to a sort of seclusion called "isolation by distance" (Wright, 1943). This phenomenon can generate a trend towards local genetic differentiation between the different isolated subgroups, which will accumulate frequency differences in some polymorphisms at random, and therefore, without a specific geographic pattern. However, genetic markers with non-random spatial variation patterns are often observed as well, including those that show gradual variation along a spatial axis. This phenomenon is called a cline or clinal variation, a term introduced by Huxley (1938).

The origin of a clinal variation can be linked to the existence of random genetic sampling drift, resumed contact between formerly isolated populations, spatially discontinuous changes of the environment, and continuous environmental gradients (Endler, 1973). Founder effects and other stochastic processes, acting together or separately are thought to be major causes of random genetic drift (Takahashi, 2015). Genetic clines can be observed in both quantitative and qualitative traits (Hoffmann & Weeks, 2006), and can affect variable proportions of the genome.

For a polymorphic trait, a cline is a temporally stable gradient of genotype or gene frequencies (Endler,

1973). Once the genetic cline arises, it will become stable if allele (or genotype) frequencies reach genetic equilibrium. Overall, most genetic clines are originated by the interplay between the diversifying effects of selection and the homogenizing effects of dispersal (Barton & Hewitt, 1985; Barton & Gale, 1993; Sotka & Palumbi, 2006). Consequently, one way to measure the strength of selective pressures in nature is by measuring the rate of change in gene frequencies in a cline (Slatkin, 1973). Likewise, the combination with other analytical methods such as estimates of linkage disequilibrium among different loci allows for analyzing the impact of simultaneous action of both evolutionary forces (Mallet et al, 1990; Barton & Gale 1993; Sotka & Palumbi, 2006).

A cline is usually defined as the gradual variation of a genetic trait regarding altitude, longitude, or latitude (Roy, Blanckenhorn & Rohner, 2018) or also concerning the distance to a point of origin (Pérez-Losada & Fort, 2018). In this way, refined studies of the spatial orientation of clinal genetic variation are thought to be very helpful to assist in the interpretation of the potential evolutionary forces that have shaped the gene pool of a species.

In this paper, we present a new software package called GenoCline, which has been developed to detect spatial trends of genetic variation from genetic marker frequencies and genome-wide databases. The program also provides a variety of statistical tools to assist in the interpretation of clinal variation. This software has been recently used to identify and determine the geographic orientation of a frequency cline for the MAPT\*H2 haplotype in Western Eurasia (Alfonso-Sánchez et al., 2018), whereas an earlier version of the algorithm was utilized to detect clines of microsatellite markers of the nonrecombining region of chromosome Y (Y-STRs) in the Iberian Peninsula (Peña et al., 2006), as well as of polymorphic *Alu* insertions in the province of Jujuy, Northwestern Argentina (Gómez-Pérez et al., 2011a).

## Methods

Bearing in mind that the main aim of GenoCline is to identify potential genetic clines, geographic and genetic data are required to running it.

To explore spatial patterns of clinal genetic variation, the GenoCline software makes a geographic coordinate axis rotate in consecutive iterations of one degree until completing 360 degrees. At the starting point (zeroiteration step), the axis represents the latitude. Similarly, the 90-degrees iteration step coincides with the longitudinal axis. In every iteration, those points representing the geographic positions of the targeted populations are orthogonally projected on the axis, thus obtaining a new coordinate vector (Fig. 1). Then the program computes Pearson product-moment correlation coefficient (r) between the allelic frequency vector and each one of the 360 geographical coordinate vectors generated by iterations. A clinal genetic variation is considered to exist if at least one of the geographic coordinate vectors shows a significant correlation (p < p(0.05) with the allele frequency vector. The spatial orientation of the gene frequency cline will then be taken as the direction of the significantly correlated geographical coordinate vector. It is not uncommon to find statistically significant associations for several consecutive degrees of rotation. In this case, the program will choose the orientation providing the highest value for the r correlation coefficient. No statistical significance tests 'a posteriori' are performed since such correlations are not statistically independent, and there is no procedure to set what the statistical significance threshold should be in these cases. It is also possible to carry out a cline analysis by using the angular transformation  $\arcsin[?]p$  (where p is the allele frequency) to separate the variance of the estimate from the estimate itself (Hartl & Clark, 1997).

As, particularly in hybrid zones, the behavior of clinal variations can often be better explained through a sigmoid function, (Barton & Gale, 1993; Fitzpatrick, 2013), the Genocline program allows to perform cline analyses by applying this specific regression model. In this case, the statistical criterion to determine the existence of genetic clines lies in F ratio probability values instead of Pearson's correlation coefficient.

When this program is employed to assess spatial genetic patterning using genome-wide databases, the analysis focused on individuals instead of populations. In this kind of study, there might be involved hundreds of thousands of SNP loci, so the one-by-one locus analysis is virtually impossible. In such circumstances, GenoCline detects patterns of clinal variation in the whole genome or the specific genomic region examined.

In doing so, a multidimensional scaling (MDS) analysis is performed from the IBS (Identity By State) distance matrix between each pair of individuals. After the MDS, we select the 10 eigenvectors showing the highest eigenvalues. Each eigenvector is then treated like an allele frequency vector of a genetic marker so that it can be correlated with the geographical coordinates of the individuals' sampling site. In this case, a rotation of coordinates is also carried out, and each eigenvector generated by MDS is tested for statistical association with 360 coordinate vectors. As discussed above, there will be a genetic cline if at least one coordinates' vector features a significant correlation coefficient (p < 0.05) with a eigenvector.

When the software detects the existence of clinal genetic variation, it gives the possibility to perform further analyses to assist in the interpretation of the hypothetical clines, including spatial autocorrelation (Moran, 1950), isolation by distance (Malecot, 1973), centroid method (Harpending & Ward, 1982), MDS (Kruskal, 1964) or the Sammon projection (Sammon, 1969).

Spatial autocorrelation may be roughly described as the property of random variables of featuring values, at pairs of locations a certain distance apart, that will be more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of observations. It poses a problem for statistical testing since autocorrelated data violate the assumption of independence of most standard statistical procedures (Legendre, 1993). For this reason, some authors have devised and implemented specific methods to test whether the observed value of a variable at one locality is significantly dependent on the values of the variable at neighboring localities (Sokal & Oden, 1978).

GenoCline affords the possibility of carrying out a spatial autocorrelation analysis either for a set of allele or haplotype frequencies. The spatial autocorrelation aims to examine, in a given territory, whether neighboring populations tend to be more genetically related than more distant populations. Among several indexes that have been devised to estimate the spatial autocorrelation level, Moran's I (Moran, 1950) has been widely accepted and used in biological studies. Moran's index values can be represented against the geographic distance in a graph known as autocorrelogram. According to the "isolation-by-distance" model, it should be expected a decreasing trend of the Moran's index as geographic distance among populations increases. More recently, Epperson & Li (1996) characterized the statistical properties of Moran's spatial autocorrelation parameters and also developed estimators of gene dispersal based on standing patterns of genetic variation. According to the findings of their study, Moran's statistics proved to be highly informative and predictable, which makes it possible to formulate robust hypothesis tests for neutral theory. Along these lines, the GenoCline software also tests the statistical significance of the autocorrelation values. Predictably, the statistical robustness of the autocorrelogram is greater when most of the Moran's index values are statistically significant.

The "isolation-by-distance" model developed by Malecot (1973) considers that the genetic kinship depends on the geographic distance, as mirrored in the mathematical function of the mean kinship coefficient  $\mathscr{O}(d)$ ,

$$\phi(d) = \mathrm{ae}^{-bd}$$

where d is the distance, a is a measure of local kinship, and b specifies the rate of exponential decline of  $\mathscr{D}(d)$ .

Another method to examine the isolation by distance is based on the analysis of the relationship between the quotient,

$$\frac{F_{\rm ST}}{(1 - F_{\rm ST})}$$

and the geographic distance through linear regression (considering all population pairs), thus estimating the slope and intercept of this relationship (Wright, 1943; Rousset, 1997).

An analysis like this can also be accomplished by using the logarithmic transformation of the geographic distance in the x-axis, as indicated in the paper by Fuselli et al. (2003).

The centroid method was conceived to estimate the relative intensity of gene flow on a group of populations (Harpending & Ward, 1982). With this purpose, the centroid method computes the expected Hardy-Weinberg proportions of heterozygotes of each subpopulation against their genetic variance  $r_i$ 

$$r_i = \sum_k \frac{\left(p_{ik} - p_k\right)^2}{p_k(1 - p_k)}$$

where  $p_{ik}$  is the frequency of allele k in population i and, is the mean frequency of allele k.

In the graph generated by the centroid method, those subpopulations plotting above the line that represents the expected heterozygosity values are interpreted to have experienced a higher impact of gene flow. On the contrary, a position below the cited line indicates a deficit of gene flow. An example of this issue can be consulted in Gómez-Pérez et al. (2011b).

Multidimensional scaling analysis, MDS (Kruskal, 1964), and the Sammon projection (Sammon, 1969) are two methods designed to obtain a small number of dimensions thus making the interpretation of a complex matrix of distances easier. In the first case, GenoCline builds up the corresponding graphical representation from  $F_{ST}$  genetic distances (Reynolds, Weir & Cockerman, 1983), whereas the Sammon projection utilizes the Euclidean distances.

#### Features

For Genocline to run on different operating systems, this software was programmed in Java. No installation is required. The execution of the different available options is easily done, via drop-down menus.

## Input/output data

Data entry is user-friendly. Genetic (either of alleles, haplotypes, or phenotypes) frequencies and geographic coordinates can be easily entered in spreadsheet table formatting. Genome-wide data can be imported in Eigensoft format. Genetic frequencies can also be exported in a format compatible with other program packages dealing with population genetics and evolutionary biology, such as Arlequin, Phylip, Poptree, and Past. All illustrations of results (graphs) are saved in eps formatting so that there are high quality and easily editable vectorial graphs available for the researcher.

## Genetic clines

Once obtained the gene frequency gradients, GenoCline groups them in 10-degrees intervals (10-degrees circumference arcs) and then generates a radar chart (also called spider chart) to represent the absolute frequency of clines per interval.

For the sake of illustrating how the Genocline software works to detect patterns of clinal variation, we employed the database in the paper by Moreira et al. (2015) in their study on European populations of the buff-tailed bumblebee (*Bombus terrestris*). Data are allelic frequencies for a set of microsatellite or short tandem repeats (STRs) loci in several populations. The results obtained with the GenoCline program are displayed in Fig. 2. It can be noticed a large number of clines with a north-south (N-S), and mostly, north-north-west to south-east (NNW-SE) orientation. According to these authors, the Irish Sea would have acted as a geographical barrier to gene flow among the Irish populations and the rest of the European populations, which might have been likely the cause behind the emergence of allele frequency clines.

In the case of the user choosing a specific cline, the GenoCline software can generate a two-dimensional plot with the regression line and their 95% confidence limits. In such a graph, the x-axis represents the coordinates concerning the virtual axis (rotating axis) and the y-axis represents the corresponding allele frequencies/eigenvector values. With this option Genocline also affords: i) orientation of the genetic cline, ii) parameters defining the regression line, iii) coefficient of determination of the regression line ( $r^2$ ), and iv) Pearson's r correlation coefficient with its statistical significance. This analysis is available for both the allele frequencies and the angular transformations of the allele frequencies.

In a study by our research team (Alfonso-Sánchez et al., 2018) on the frequency of Tau haplotypes in the Roma population settled in the Basque Country, we found a frequency cline for the MAPT\*H2 haplotype in a cluster of Western Eurasian human populations. Specifically, we could identify a trend for growing haplotype frequencies as we moved from India to Europe (Fig. 3).

In the quest for gene frequency clines, it may be on occasion more appropriate to perform sigmoid regression than linear regression. Thus, for instance, in the study by Chlaida et al (2009), the authors found clines' traces at the SOD\* locus along the North African coast, in a series of samples of the European pilchard (*Sardina pilchardus*). They could identify two differentiated groups separated by the Agadir Bay. The sigmoid cline detected by GenoCline for allele SOD\*80 is presented in Fig. 4. The midpoint of the threshold is at 0.61, between Sidi Ifni and Agadir populations.

GenoCline can also analyze patterns of clinal variation from genome-wide databases. To illustrate so, we chose a small group of individuals from the 'Human Origins' database in Eigensoft format (Lazaridis et al., 2014). Firstly, we obtained eigenvector values through a multidimensional scaling analysis of 620,000 SNPs from 65 current European individuals. Then, we represented the clinal pattern found for eigenvector 1 values (Fig. 5). As can be noticed, through this analysis we could identify a trend of significant variation (p < 0.0001) in the northeast-southwest (NE-SW) direction.

## Spatial autocorrelation

Based on the populations' geographic coordinates, GenoCline calculates the distances between all pairs of populations and assigns them to seven distance categories or classes. Subsequently, it runs spatial autocorrelation analysis and calculates the Moran's index for each distance class. The program further generates the autocorrelogram, displaying the Moran's index for each class.

Figure 6 illustrates the autocorrelogram with Moran's index (I) values. Such I values were calculated from a set of polymorphic Alu insertions in a set of human populations from the American continent (Gómez-Pérez et al., 2011b).

## Isolation by distance

Intending to assess whether the allele frequency distribution in a given region fits into the Malécot's isolationby-distance model, GenoCline computes both the intercept (a) and slope (b) values either in equations of linear regression models or those of exponential models.

# Centroid method

With GenoCline, carrying out an analysis based on the centroid method is possible as well. The program is designed to represent population values for both the heterozygosity and the genetic variance, as well as a line segment with the mean values of heterozygosity expected.

Buhler et al (2012) surveyed the frequencies of five human *HLA*genes in 11 cantons of Switzerland. Fig. 7 shows the graph generated by the GenoCline program when analyzing the database of this study. The line represents the expected heterozygosity, which is computed based on the variance contributed by each population to the total Wahlund variance. It can be noticed a much lower than expected gene flow in Coira, which is the most geographically isolated population. On the other hand, Bern and Zurich behave as target populations of internal (within Swiss population) gene flow considering their positions, remote from the centroid line but with a rather low contribution to the Wahlund variance (homogenizing effects of gene flow). Likewise, Geneve seems to be the population with the highest level of external gene flow, as per distant position from the centroid and highest contribution to the Wahlund variance.

# $\mathbf{MDS}$

Multidimensional scaling analysis (Kruskal, 1964) and Sammon projection (Sammon, 1969) are further options available in the program presented herein, intending to facilitate the interpretation of the genetic relationships among populations.

# Alternatives to GenoCline

There are very attractive and helpful tools to study genetic clines, but they are focused on goals other than those of GenoCline. These are the cases, for example, of bgc (Gompert & Buerkle, 2012), which is intended to detect genomic regions involved in the emergence of clines, and also of HZAR (Derryberry, Derryberry, Maley, & Brumfield, 2014), which requires prior knowledge of the gradient orientation to fit the results to a cline model.

## Download and usage

The current version (v1.3) of the software, as well as the corresponding User's Manual, are available at http://genocline.sourceforge.net and http://www.didac.ehu.es/genocline. The normal running of this program does not depend on a specific operating system or platform, but merely requires a runtime environment of Java 1.8 or higher to be installed on the system. Finally, we have created the email account genoclinesoft-ware@gmail.com with an eye to facilitate feedback with potential users.

# **Conflict of Interest**

The authors declare no conflict of interest in this paper.

## References

Alfonso-Sánchez, M. A., Espinosa, I., Gómez-Pérez, L., Poveda, A., Rebato, E., Peña, J. A. 2018. Tau haplotypes support the Asian ancestry of the Roma population settled in the Basque Country. Heredity, 120(2): 91.

Barton, N. H., Gale, K. S., 1993. Genetic analysis of hybrid zones (in R. Harrison, editor). Hybrid zones and the evolutionary process. Oxford University Press, New York, New York, USA. pp. 13-45

Barton, N. H., Hewitt, G. M., 1985. Analysis of hybrid zones. Annual Review of Ecology and Systematics, 16(1): 113-148

Buhler S, Nunes JM, Nicoloso G, Tiercy JM, Sanchez-Mazas A. 2012. The heterogeneous *HLA* genetic makeup of the Swiss population. PLoS One 7 (7): e41400.

Chlaida, M., Laurent, V., Kifani, S., Benazzou, T., Jaziri, H., Planes, S. 2009. Evidence of a genetic cline for Sardina pilchardus along the Northwest African coast. ICES Journal of Marine Science 66 (2): 264-271.

Derryberry, E. P., Derryberry, G. E., Maley, J. M., Brumfield, R. T. 2014. HZAR: hybrid zone analysis using an R software package. Molecular Ecology Resources, 14(3) : 652-663.

Endler, J. A., 1973. Gene flow and population differentiation: studies of clines suggest that differentiation along environmental gradients may be independent of gene flow. Science, 179(4070) : 243-250.

Epperson, B.K., Li, T. 1996. Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. Proceedings of the National Academy of Science USA 93: 10528-10532

Fitzpatrick, B. M. 2013. Alternative forms for genomic clines. Ecology and evolution, 3(7), 1951-1966.

Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. 2003. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. Mol Biol Evol 20: 1682-91.

Gómez-Perez, L., Alfonso-Sanchez, M. A., Dipierri, J. E., Alfaro, E., Garcia-Obregon, S., De Pancorbo, M. M., Bailliet, G., Pena, J. A. 2011a. Microevolutionary processes due to landscape features in the province of Jujuy (Argentina). American Journal of Human Biology, 23(2) : 177-184.

Gomez-Perez, L., Alfonso-Sanchez, M. A., Sanchez, D., Garcia-Obregon, S., Espinosa, I., Martinez-Jarreta, B., de Pancorbo, M.M., Pena, J. A. 2011b. *Alu* polymorphisms in the Waorani tribe from the Ecuadorian Amazon reflect the effects of isolation and genetic drift. American Journal of Human Biology 23(6): 790-795.

Gompert, Z., Buerkle, C. A. 2012. bgc: Software for Bayesian estimation of genomic clines. Molecular Ecology Resources, 12(6):1168-1176.

Harpending, H.C., Ward, R.H. 1982. Chemical systematics and human populations. In Biochemical aspects of evolutionary biology (M Nitecki, ed.). Chicago: University of Chicago. pp. 213-56.

Hartl, D. L., Clark, A. G. 1997. Principles of population genetics. Sunderland: Sinauer associates. 542 pp.

Hoffmann, A.A., Weeks, A.R. 2006. Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in Drosophila melanogaster from eastern Australia. Genetica 129:133–147

Huxley, J. S. 1938. Clines: an auxiliary taxonomic principle. Nature, 142(3587), 219-220.

Kimura, M., Weiss, G.H. 1964. The stepping-stone model of population structure and the decrease of genetic correlation with distance. Genetics 49 : 561-576

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29(1): 1-27.

Lazaridis, I., Patterson, N., Mittnik, A., *et al*. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature, 513(7518) : 409.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm?. Ecology, 74(6): 1659-1673.

Malecot, G. 1973. Isolation by distance, in Genetic Structure of Populations (NE Morton, ed.), 72-75. Honolulu: University of Hawaii Press.

Mallet, J., Barton, N., Lamas, G., Santisteban, J., Muedas, M., Eeley, H., 1990. Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in Heliconius hybrid zones. Genetics 124: 921–936.

Moran, P. A. P. 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37: 17-23.

Moreira, A. S., Horgan, F. G., Murray, T. E., Kakouli-Duarte, T. 2015. Population genetic structure of Bombus terrestris in Europe: Isolation and genetic differentiation of Irish and British populations. Molecular Ecology, 24(13) : 3257-3268.

Pena, J. A., Garcia-Obregon, S., Perez-Miranda, A. M., De Pancorbo, M. M., Alfonso-Sanchez, M. A. 2006. Gene flow in the Iberian Peninsula determined from Y-chromosome STR loci. American Journal of Human Biology, 18(4) : 532-539.

Perez-Losada, J., Fort, J. 2018. A serial founder effect model of phonemic diversity based on phonemic loss in low-density populations. PloS One, 13(6) : e0198346.

Reynolds J., Weir B.S., Cockerman C.C. 1983. Estimation of the coancestry coefficient: bases for a short term genetic distance. Genetics 105:767–779

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145: 1219-28.

Roy, J., Blanckenhorn, W. U., Rohner, P. T. 2018. Largely flat latitudinal life history clines in the dung fly Sepsis fulgens across Europe (Diptera: Sepsidae). Oecologia : 1-12.

Sammon, J.W. 1969. A nonlinear mapping for data structure analysis. IEEE Transactions on Computers 18: 401-9.

Slatkin, M., 1973. Gene flow and selection in a cline. Genetics, 75(4): 733-756

Sokal, R. R., Oden, N. L. 1978. Spatial autocorrelation in biology: 2. Some biological implications and four applications of evolutionary and ecological interest. Biological Journal of the Linnean Society, 10(2): 229-249.

Sotka, E. E., Palumbi, S. R., 2006. The use of genetic clines to estimate dispersal distances of marine larvae. Ecology, 87(5): 1094-1103

Takahashi, Y. 2015. Mechanisms and tests for geographic clines in genetic polymorphisms. Population ecology, 57(2): 355-362.

Wright, S. 1943. Isolation by distance. Genetics 28: 114-38.

## Author Contributions

JAP devised the computer software ; JAP, MAA-S and LG-P designed the research, analyzed data to test the program, revised and approved the submitted manuscript ; JAP and MAA-S wrote the paper and changed the manuscript according to criticisms.

# FIGURE LEGENDS

Fig. 1. Graphic representation of the rotation of the Y axis to the Y' axis. The new coordinate is calculated as  $b' = c \propto \cos(\alpha')$ . For calculations, a clockwise axis rotation is set.

**Fig. 2**. Orientation of the genetic clines detected in a set of microsatellite or STR markers in European populations of the buff-tailed bumblebee *Bombus terrestris* (Moreira et al., 2015). Concentric circles indicate the number of clines detected. Radii delineate the orientation of genetic clines in 10-degree intervals.

**Fig. 3**. Frequency cline for MAPT\*H2 haplotypes in Western Eurasian populations. Regression line is displayed along with their 95% confidence limits (dotted lines).

**Fig. 4**. Allele frequency cline for SOD\*80 on the North African coast. Curve represents the statistically significant sigmoidal regression found for the study by Chlaida et al. (2009) on the European pilchard (*Sardina pilchardus*). The orientation of the cline, south southwest-north east (SSW-NE), is also shown.

**Fig. 5**. Clinal pattern detected for eigenvector 1 values, obtained through MDS over a database of 620,000 SNPs in 65 current European individuals. In the graph, each point represents an individual. Regression line along with 95% confidence limits (dotted lines) are shown.

Fig. 6. Autocorrelogram obtained from a set of polymorphic Alu loci in several human populations from the Americas. Variations in the Moran's index owing to the distance among populations are shown, along with the corresponding 95% confidence limits.

Fig. 7. Centroid graph obtained from frequencies of five human HLA loci in 11 cantons of Switzerland. Line is the expected heterozygosity value for this group of populations, which is computed based on  $r_i$ , being  $r_i$  the variance contributed by each population to the total Wahlund variance.













