

Combining biotechnologies and GIScience for livestock genetic resources conservation

Stéphane Joost & the Econogene Consortium¹
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Stephane.Joost@epfl.ch

SUMMARY

In agriculture, livestock is presently facing important pressures chiefly due to modern farming. FAO estimates that one breed of traditional livestock dies out every week somewhere in the world. Loss of breeding stock by extinction and appearance of new diseases are in progress and ways have to be found to conserve diversity in livestock genetic resources. The Econogene project, supported by the European Union, banked on Geographical Information Science and biotechnologies complementarity to develop tools to measure, monitor and manage biodiversity. In the context of the analysis of relationships between environment and the genome, this paper describes the use of a method allowing the visual interpretation of a large amount of logistic regression models results (rejection tables). This approach is likely to suggest some avenues worth exploring with numerical analysis. This will then lead to the possible identification of specific roles for molecular markers, and eventually allow to take appropriate breeding measures.

KEYWORDS: *Spatial analysis, biotechnologies, biodiversity conservation, logistic regression*

INTRODUCTION

Under pressure mainly because of commercial production, breeds are losing genetic diversity and this erosion has significant adverse consequences for the development of a sustainable agriculture. Traditional breeds are gradually replaced by a few high production ones (Bruford, 2003) which require high inputs, skilled management and comparatively benign environments (Thrupp, 1998). Actions have to be undertaken to check this trend, aiming at conserving biodiversity. Thus, tools are required to monitor the evolution of genetic diversity. Molecular biology makes it possible to measure genetic diversity by means of identified markers within the genome. Spatial analysis is able to exploit this information taking into account its spatial variability. Moreover, it also gives the opportunity to study the relationship between genetic characteristics and the nature of the environment according to the sampling location. This is likely to provide additional indications when genetic resources management policies have to be introduced. This paper shows how Geographical Information Science (GIScience) and biotechnologies can act in a complementary way to constitute an approach within which data investigation is likely to allow the identification of specific roles for molecular markers.

LOSS OF GENETIC DIVERSITY

Within the branch of agriculture, livestock sector is presently facing important stress due to modern farming. Biodiversity is threatened because artificial selection and controlled reproduction gradually led to a general loss of genetic diversity among species and breeds, which may potentially cause damaging effects like loss of breeding stock (extinction, new diseases) (Bruford *et al.*, 2003). Consequences are of global concern and sustainable ways have to be found to optimally conserve livestock genetic resources diversity as we are confronted to an accelerating extinction crisis (Luikart *et al.*, 2003).

To face the threat, FAO (Food and Agriculture Organization) initiated a global strategy for the management of Farm Animal Genetic Resources (AnGR²) which general purpose is to propose the

¹ <http://lasig.epfl.ch/projets/econogene/>

introduction of adapted policies aiming at conserving livestock biodiversity. FAO suggests to identify and understand the genetic resources of each important farm animal species, and to prioritize and conserve unique AnGR.

Methods and techniques have been developed to reach these goals. Since the beginning of the 1990's, the development of biotechnologies led to the elaboration of an array of different molecular techniques³ able to measure diversity at the DNA level (Karp *et al.*, 1997) and molecular approaches have been progressively recognized to be appropriate tools to measure, monitor and manage genetic diversity (Bruford *et al.*, 2003).

GENETICS AND GISCIENCE

What can be the role of Geographical Information Systems (GIS) in environmental processes like the one described here above? "How best to use GIS to change the way environmental modeling is being done?" to refer to the way Maidment (1996) addressed this issue. GIS have an important role to play in environmental monitoring, as long as it is not restricted to the creation of "pretty maps" (Larsen, 1999). Nevertheless, cartography in all likelihood constituted the first achieved integration of geographical information and any environmental issue.

In a book dedicated to "Environmental modeling with GIS" (Goodchild *et al.*, 1993), several authors acknowledged that GIS and Environmental modeling were well established methods, but that their integration was still an emerging field (Fedra, 1993; Parks, 1993). Today, talking of "Environmental modeling" may even imply the fact that geographical information technologies are involved, this because since the 1990's the requested integration has been massively realized (Hunsaker *et al.*, 1993). Aspinall (1999) classifies the integration of GIS and biodiversity topics in a "Landscape conservation" category within which analyses often consist in evaluating geographical patterns of diversity (McKendry & Machlis, 1991). Since the first steps of ecological modeling in the late 1960's, ecologists gradually appropriated GIScience technologies (Hunsaker *et al.*, 1993). More and more researches were led using GIS in spatial ecology, especially in habitat modeling were considerable developments were made in interaction with advanced spatial statistics (Guisan & Zimmermann, 2000).

Considering Genetics, the study of spatial structures is existing since Wright (1931) developed adaptation models which were incorporating spatial distribution and distance considerations (Epperson, 2003). Distance remains a central issue in spatial genetics as the main reference models directly refer to, or are constrained by it (Epperson, 2003; MacArthur & Wilson, 2001). The Mantel test (1967) was used to compare geographical with genetic distances (Epperson, 2003). On these basis, GIS were introduced to develop dispersal models to simulate animal population migrations in the landscape (Vuilleumier, 2003), or to provide tools for visualization and analysis of geographical population structures (Hoffmann *et al.*, 2003). Spatial analysis methods like kriging were also used to define diversity zones (Hoffmann *et al.*, 2003; Bucci & Vendramin, 2000). Hamann *et al.* (2000) also exploited kriging to detect areas of genetic differentiation. AFLP markers were also used to show association with salt tolerance in wild barley (Pakniyat *et al.*, 1997). And directly related to the reasoning explained in the present paper, Skot *et al.* (2002) investigated interaction between environmental characteristics and molecular information, and identified AFLP markers that correlated in frequency with cold tolerance.

² <http://www.fao.org/ag/cgrfa/AnGR.htm>

³ A glossary of molecular biology related terms appears at the end of the paper

RESEARCH CONTEXT

As a contribution to a EU program that aims at improving the sustainability of European agriculture, the Econogene project⁴ was conceived to promote the sustainable conservation of genetic resources in sheep and goats.

Spatial analysis has been involved to visualize patterns of genetic diversity, and to assess interaction between animal genetic resources and environmental characteristics to identify potential associations between molecular markers and given environmental variables. When detected, any of these associations would be the best evidence for adaptive significance (Luikart *et al.*, 2003) and would allow the identification of specific roles for particular markers. This may therefore constitute a contribution in helping to understand the functioning of the genome and make it possible to prioritize breed populations for conservation.

DATA AND METHODS

We propose an exploratory approach to assess Amplified Fragment Length Polymorphism (AFLP) markers sensitivity to a set of topographic and climatic variables characterizing the area where sampled sheep breeds are raised. The method to reveal associations consists in using the genotyping of the sampled breeds and to look for the markers whose frequency is correlated with selected environmental variables.

The main aim of this work is to emphasize the contribution of the proposed methodology in the general task of understanding the genome functioning, and particularly when trying to identify genes under selection. Special attention has to be turned to the method, and results can be considered as an illustration of some possible ways to exploit it.

DNA markers

Molecular markers is a selection technique of DNA signposts which allows the identification of differences in the nucleotide sequences. This technique allows to locate the genes for a trait of interest on a chromosome. Various classes of markers exist, as DNA is divided into coding and non coding regions, as well as repetitive and non repetitive sequences. Those markers kinds have different scopes and different advantages and disadvantages (specificity, cost, ease of analytical interpretation). Moreover, one can distinct markers which are submitted to selection pressure (adaptive markers), and markers that are not (neutral markers) (Luikart *et al.*, 2003). AFLP are increasingly used to identify markers associated with traits that are under selection (Skøt *et al.*, 2002; Luikart *et al.*, 2003). Though AFLP are neutral markers, they can be used to check potential association with environmental parameters and thus to locate genes under selection according to a population genomics approach (Luikart *et al.*, 2003).

AFLP data consist of binomial information (details of AFLP technology in Ajmone-Marsan *et al.*, 1997). At a given geographic location, a marker is present or not and has a frequency. In the rejection table (figure 2), the ID row for markers correspond to their abundance at sampling locations. The higher the ID, the more abundant the marker.

Sheep biological data were captured in the framework of the Econogene project. 624 animals⁵ were sampled representing 40 breeds⁶, among farms in 10 European countries (figure 1). As several animals (up to 3) could be sampled in the same location, we have 577 distinct locations characterized by 62 AFLP markers presence/absence.

⁴ Quality of life V framework program, <http://europa.eu.int/comm/research/quality-of-life.html>

⁵ To identify conservation priorities, biological samples were collected for over 3'000 animals representing 56 sheep breed and 44 goat breed

⁶ 4 sheep breeds were not used in the context of this research

Georeferenced information

Coordinates of the farms were recorded on the field by partners in charge of animal sampling. They either used a GPS device, or derived the location from local maps. Given the extent of the study area, the location error tolerance was fixed by the CORINE land cover database (250 m)⁷, the spatial information with the highest resolution at our disposal, used for other aspects of the project.

Topographic data

Altitude was also measured on the field by partners which either used an altimeter, or a GPS device. All data were validated with the 30 arc second Digital Elevation Model (DEM) of the Shuttle Radar Topography Mission (SRTM, NASA), and the 3 arc second (90 m) model was used when the altitude of farms which coordinates were missing had to be calculated.

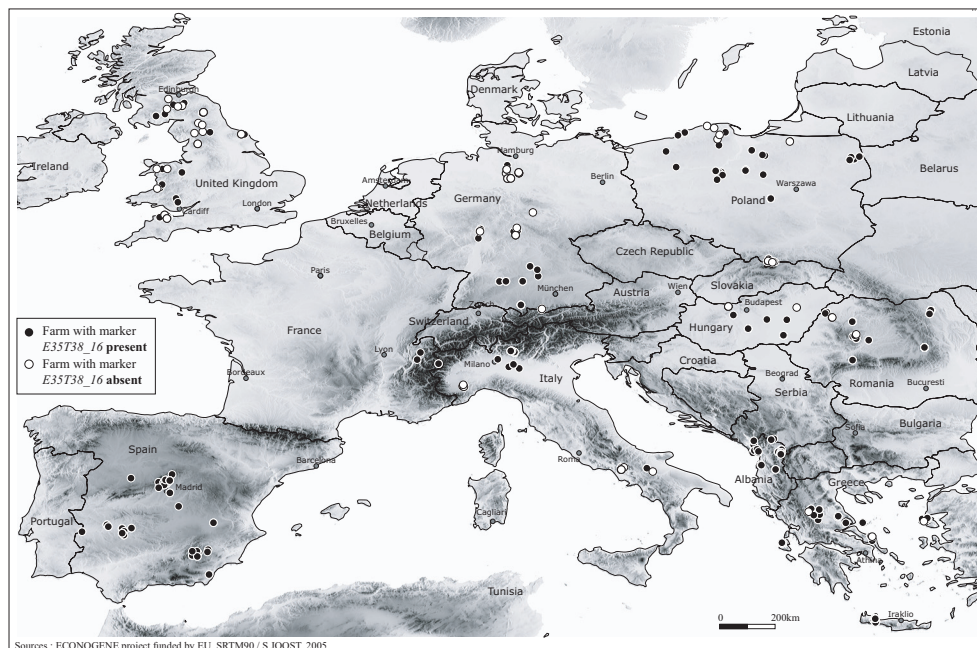


Figure 1 : Spatial distribution of the farms where sheep were sampled and example of the presence/absence of the E35T38_16 marker.

Climatic data

Climatic data consist of 10 minutes latitude/longitude resolution grids over global land areas, for the period 1961 to 1990, and were produced by the Climatic Research Unit of Norwich⁸ (CRU). Monthly variables are mean temperature, diurnal temperature range, relative humidity, sunshine, ground-frost frequency, wet-day frequency, wind speed, and precipitation. Information about the climate station data used to construct the climatology is described in New *et al.* (2002). It is to be noted that those mean values are used to discriminate regions between them and to provide general climatic profiles of reference through the overall sampling area. This is important because the sampling of genetic information was made after the 1961 to 1990 period, and climate continuously evolves.

⁷ <http://terrestrial.eionet.eu.int/>

⁸ <http://www.cru.uea.ac.uk/>

For each variable, yearly means were also calculated. With altitude, a total of 118 topo-climatic variables were made available for association tests.

A Principal Component Analysis was run to synthesize environmental information and to bring out distinct groups (environmental profiles) to be interpreted. This interpretation was based on the first 3 principal components (75% of explained variance) and allowed to identify a humidity factor (component 1), a wind factor (2), and a cold factor (3).

Interaction assessment

Univariate logistic regression models (Hosmer & Lemeshow, 2000) were fitted to the data. Multivariate models were not used because the goal of the analysis is to identify AFLP markers individual specific responses to topo-climatic variables. Moreover, interaction between processes in natural sciences are so complex that it is meaningless to combine only a part of the contributing information.

The logit link was of the simple form $g(x) = \alpha + \beta x$. The significance of the β coefficient in the model was assessed using the Wald test (Hosmer & Lemeshow, 2000). This test compares the likelihood estimator of the β_i parameter with an estimation of its standard error.

$$W = \frac{\hat{\beta}_i}{\sigma(\hat{\beta}_i)}$$

The Wald statistic may belong to the normal distribution or not, and thus the null hypothesis (H0) being rejected or not. In the context of multiple hypothesis testing, it is recognized that when one wishes to simultaneously test several hypotheses at a common significance level α , the generalized Type I error probability (the probability of rejecting at least one of the hypotheses being tested that is in fact true) is typically much in excess of α (Shaffer, 1995). There are a large number of multiple testing procedures available, and we chose to apply the simple Bonferroni correction (Shaffer, 1995) to identify models for which studied parameters were very significant. This correction implies to divide the wanted confidence interval by the number of comparisons to get a significance threshold.

Rejection table

As the important information on which this method focuses is to know whether the null hypothesis is rejected or not, we created a rejection table containing the results of the many tests (62 markers by 118 topo-climatic variables = 7316 models) on one single page (figure 2). The use of rejection tables is effective because of the simplicity of its functioning and its synthesizing power (Lassueur *et al.*, submitted for publication). Indeed, it permits to display information produced by a large number of calculations and provides different ways of analysing results. This tool allows to generate global visual information about a set of predictors and to compare them with the undeniable advantage of concentrating the attention on significant information only. Making use of the same table, it is possible to get detailed differential information between markers or variables and to make specific comparisons between them. On top of the model response, the last cell of all lines and columns contains the total and the percentage of significant models, summarizing the information per variable or per marker.

Sorting orders of molecular markers or topo-climatic variables can be exploited to make any eventual structure appear and to be visually detected (table 2). Several values can be used to sort out the models : the number of significant models per variable or the factorial loads of topo-climatic variables on the principal components which have been attached to the table. If there is a link between the components and the significant models, the latter will form homogeneous groups associated with high correlations on a given component.

ID		1	2	3	4	5	6	48	49	...	50	51	52	53	54	55	56	57	58	59	60	61	62		
		E35T32_25	E35T32_22	E35T38_9	E35T32_32	E35T32_28	E35T38_7	E35T38_19	E35T32_26	...	E45T38_24	E35T38_16	E35T38_1	E35T38_17	E35T32_27	E35T38_5	E35T32_29	E35T38_20	E35T38_31	E35T32_23	E35T38_24	E35T38_25	E35T32_10	#signif. models	%
1	altitude	0	0	0	0	0	0	0	0	...	0	0	0	1	1	0	0	0	0	0	0	0	0	4	6.5
2	wndjan	0	0	0	1	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0	0	0	1	10	14.5
3	wndfeb	0	0	0	1	0	0	0	0	...	0	1	0	1	0	0	0	0	0	0	0	0	1	10	12.9
...	5	8.1
117	sundec	0	0	0	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0	0	0	0	5	8.1
118	sunjear	0	0	1	0	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0	0	0	0	7	11.3
	#signif. models	2.5	3	0.8	1	29.7	35	48.3	57	7.6	9	0.0	0	0	0	0	0	0	0	0	0	0	0	736	
	N																								

↑ Extract of a rejection table

Synthetic view of the 7316 models results →

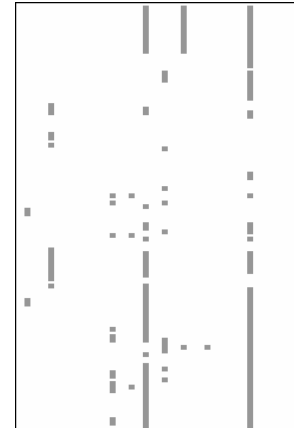


Figure 2 : On the left, detail of a “rejection” table in which each cell contains the result of a H0 test. “1” are indicating significant models. On the right, visual display of a full results table : markers are ordered in an increasing order according to their frequency at sampling sites (from left to right), and topo-climatic variables are grouped together by type (wind, temperature, etc.), from top to bottom. Vertical lines are suggesting that AFLP markers’ response is not exclusive to specific environmental variables. Isolated points will turn out not to be significant with a higher confidence level.

RESULTS

7316 models were run to assess AFLP markers response to environmental stimulus. The overall response of the 62 AFLP markers could first be visualized by sorting out the information contained by the rejection table, according to the different available criteria. The right part of figure 2 shows that – with a 95% confidence interval – the visualization of the whole table is useful to detect that the overall response is distributed among almost all variables types.

Refining the analysis

The results were evaluated according to 4 different confidence levels (table 1). Let’s focus on the 99.9% confidence level where 90 models (1.2% of the total) rejected the null hypothesis. These 90 models are distributed among 3 AFLP markers : A) E35T38_16 to which 55 topo-climatic variables are significantly related, B) E35T32_6 (33) and C) E45T38_26 (2). The distribution of significant topo-climatic variables among the first two markers (A and B) is very similar : only 6 of them are significant for B) and not for A) – mainly wind variables.

Conversely, 28 variables are significant for marker A) and not for B), the main part of them being *mean diurnal temperature range* and *relative humidity* (18/28). The third marker (C) is only reacting to the months of March and April *relative humidity* values.

	95% conf. level	99% conf. level	99.5% conf. level	99.9% conf. level
Number of markers	10	7	6	3
Number of models	205	151	128	90

Table 1 : Number of molecular markers responding to environmental stimulus and number of significant models according to the confidence level.

This being observed, it is difficult to interpret the group of variables causing a reaction as they are spreaded over several climatic information families. Moreover, there is no apparent relationship between them and the environmental profiles obtained thanks to the PCA (table 2). The most interesting and useful point for both A) and B) markers is that the *number of days with ground-frost* and the *precipitation* variables are absent from significant models, with the exception of ground frost in April. Finally, we can observe that the B) marker seems to be more sensitive to wind, and less to humidity in comparison with A).

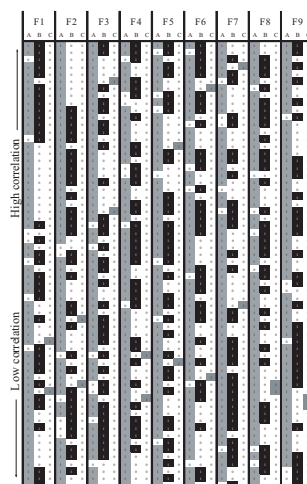


Table 2 : Significant models for markers A (grey), B (black) and C (dark grey) with a 99.9% confidence interval. White cells (0) are non significant models. Markers are sorted out in a decreasing order according to their correlation to each of the first 9 factors of the PCA (93% of the variance), showing that significant models are spreaded out over high and low correlations to all factors, and therefore that there is no apparent association with any of those environmental profiles.

DISCUSSION AND PERSPECTIVES

On one side, this work allowed to experiment an exploratory method designed to study potential associations between the environment and genetic characteristics of sheep. Such an overall approach has to be completed by a detailed statistic analysis of the highlighted relationships. This should permit to classify significant models (exploiting the p-values), to better exploit the information contained in non-significant models and to improve the way of interpreting groups of significant responses : the discriminating power of PCA turned out not to be powerful enough and it would probably have brought more information when associated to non-parametric methods. Moreover, the robustness of the significance of the models can be consolidated by processing other statistics to test null hypothesis in logistic regression. Hosmer and Lemeshow (2000) are proposing several statistics to overcome the possible Wald test weakness. G, C and D statistics are alternative ways to measure the pertinence of the inclusion of variables in a model (Hosmer & Lemeshow, 2000). Parallel to these statistics improvements, investigations have to be led at the genetic level to find out the gene(s) associated with the three identified AFLP markers.

On the other side, and more fundamentally, the 90 occurrences for which environmental characteristics participated in explaining the presence of both markers providing an important number of significant models, made it possible to observe that the response to environmental stimulus was not concentrated on a few homogeneous variables, but distributed among several distinct ones. These markers are probably related to environmental characteristics; in case they are associated with genes,

it could signify that the latter have a generalist role regards to selection processes, and not a specialized one as we could have expected (association with cold variables only for instance). This is likely to correspond to recent developments about genome expression arguing that genes do not have only one assigned task, but take part in several processes being combined between them in different ways.

Other aspects should be developed to pave the way for an efficient detailed numerical analysis :

Geographical improvements

Spatial analysis can be improved, first by performing the same calculations on a reduced area to observe how the significance of models is evolving with a reduction of the number of measurements.

Of course, to further exploit the contribution of geographical information analysis, the spatial distribution of selected markers has to be investigated (figure 1 is an example). The systematic mapping of markers for which functional (selective) hypothesis can be done, is likely to shed additional light on the AFLP markers response to topo-climatic variables. This analysis may highlight situations of barriers or isolated locations for instance (mountains, rivers), or any other spatial pattern.

Additional geodata can also be used to complete the environmental profiles (CORINE land cover, the digital soil map of the world (FAO), etc.).

Other topographical variables could be derived from the SRTM 3 arc second DEM. Considering the effect those variables may indirectly have on milk yield (Msechu, 1995), aspect and slope are likely to be important to identify adaptive molecular markers.

Microsatellites and SNPs

This perspective is the more important : the methodology will soon be applied to other molecular markers types (Microsatellites and SNPs) available for the same animals. Comparing results will provide a reference allowing to evaluate the approach. This will be particularly interesting with SNPs as they are supposed to be adaptive markers. Moreover, this method will be used to assess the behaviour of molecular markers defined as outliers by standard genetic analysis, these markers being often supposed – but not proved – to be associated with selection processes.

This method has of course no sense *per se*, and is only one link in the chain of an interdisciplinary reasoning whose ultimate purpose is to provide information precise enough to identify markers, allowing to find out candidate genes for supposed traits of interest.

Acknowledgement

This work has been supported by the European Commission (Econogene contract QLK5-CT-2001-02461). The content of the publication does not represent the views of the Commission or its services. I would like to thank Régis Caloz, Abram Pointet and Joël Chételat for their valuable comments.

MOLECULAR GENETICS GLOSSARY

<i>AFLP marker</i>	Amplified Fragment Length Polymorphism. A highly sensitive method for detecting polymorphisms in DNA.
<i>Allele</i>	One of the different forms of a gene that can exist at a single locus.
<i>Genotype</i>	The specific allelic composition of a cell, either of the entire cell or more commonly for a certain gene or a set of genes.
<i>Locus</i>	The specific place on a chromosome where a gene is located.
<i>Marker</i>	Any genetic element (DNA sequence) which can be readily detected and used to follow a chromosome or chromosomal segment during genetic analysis.

<i>Microsatellites marker</i>	Microsatellites are defined as regions within DNA sequences where short sequences of nucleotides are repeated in tandem arrays.
<i>Molecular biology</i>	The study of the structure, function, and makeup of biologically important molecules.
<i>Molecular genetics</i>	The study of the molecular processes underlying gene structure and function.
<i>Molecule</i>	A pure substance which results when two or more atoms of a single element share electrons, for example O ₂ .
<i>Nucleotide</i>	The molecules that form the basic modular structure of the double helix of the DNA molecule.
<i>Polymorphism</i>	A difference in DNA sequence among individuals, groups, or populations.
<i>SNP marker</i>	Single Nucleotide Polymorphisms. DNA sequence variations that occur when a single nucleotide (A, C, G, or T) is altered in the genome sequence. Each individual has many SNPs that together create a unique DNA pattern.

BIBLIOGRAPHY

- Ajmone-Marsan, P., Valentini, A., Cassandro, M., Vecchiotti-Antaldi, G., Bertoni, G., Kuiper, M., 1997 AFLP markers for DNA fingerprinting in cattle. *Animal Genetics*, No 28, pp 418-426.
- Aspinall, R.J., 1999 GIS and landscape conservation. In Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (ed.). *Geographical Information Systems, Volume 2*. John Wiley & Sons, pp 967-980.
- Bruford, M.W., Bradley, D.G., Luikart, G., 2003 DNA markers reveal the complexity of livestock domestication. *Nature*, Vol.4, No 11, pp 900-910.
- Bucci, G., Vendramin, G.G., 2000 Delineation of genetic zones in the European Norway spruce natural range: preliminary evidence. *Molecular Ecology*, Vol.9, No 7, 923-934.
- Epperson, K.E., 2003 *Geographical Genetics*. Princeton University Press.
- Fedra, K., 1993 GIS and environmental modeling. In Goodchild, M.F., Parks, B.O., Steyaert, L.T. (ed.). *Environmental Modeling with GIS*. Oxford University Press, pp 36-50.
- Goodchild, M.F., Parks, B.O., Steyaert, L.T., 1993 *Environmental Modeling with GIS*. Oxford University Press.
- Guisan A., Zimmermann, N., 2000 Predictive habitat distribution models in ecology. *Ecological Modelling*, No 135, pp 147-186.
- Hamann, A., Koshy, M.P., Namkoong, G., Ying, C.C., 2000 Genotype x environment interactions in *Alnus rubra*: developing seed zones and seed-transfer guidelines with spatial statistics and GIS. *Forest Ecology and Management*, Vol.136, pp 107-119.
- Hoffmann MH, Glass AS, Tomiuk J, Schmutz H, Fritsch RM, Bachmann K., 2003 Analysis of molecular data of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS). *Molecular Ecology*, Vol.12, No 4, pp 1007-1019.
- Hosmer, D.W., Lemeshow, S., 2000 *Applied logistic regression*. John Wiley & Sons, New York.
- Hunsaker, C. T., Nisbet, R.A., Lam, D.C.L., Browder, J.A., Baker, W. L. Turner, M. G., Botkin, D.B., 1993 Spatial models of ecological systems and processes: The role of GIS. In Goodchild, M.F., Parks, B.O., Steyaert, L.T. (ed.). *Environmental Modeling with GIS*. Oxford University Press, pp 248-264.
- Karp, A., Edwards, K.J., Bruford, M., Funk, S., Vosman, B., Morgante, M., Seberg, O., Kremer, A., Boursot, P., Arcander, P., Tautz, D., Hewitt, G.M., 1997 Molecular technologies for biodiversity evaluation : opportunities and challenges. *Nature Biotechnology*, Vol.5, pp 625-628.

- Larsen, L., 1999 GIS in environmental monitoring and assessment. In Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (ed.). Geographical Information Systems, Volume 2. John Wiley & Sons, pp 999-1007.
- Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W., 1999 Geographical Information Systems, Volume 2. John Wiley & Sons.
- Luikart, G., England, P.R., Tallmon, D., Jordan, S., Taberlet, P., 2003 The power and promise of population genomics: from genotyping to genome typing. *Nature*, Vol.4, No 12, pp 981-994.
- Maidment, D.R., 1996 Environmental Modeling with GIS. In Goodchild, F., Steyaert, L.T., Parks, B.O., *et al.* (ed.). GIS and Environment Modeling : Progress and Research Issues. GIS World Books, pp 315-323.
- MacArthur, R.H., Wilson, E.O., 2001 The theory of island biogeography. Princeton University Press.
- McKendry, J.E., Machlis, G.E., 1991 The role of geography in extending biodiversity GAP analysis. *Applied Geography*, Vol.13, No 2, pp 135-152.
- Msechu, J.K.K., Mgheni, M., Syrstad, O., 1995 Influence of various climatic factors on milk-production in cattle in Tanzania. *Tropical Animal Health and production*, Vol.27, No 2, pp 121-126.
- New, M., Lister, D., Hulme, M., Makin, I., 2002 A high-resolution data set of surface climate over global land areas. *Climate Research*, Vol.21, pp 1-25.
- Pakniyat, H., Powell, W., Baird, E., Handley, L.L., Robinson, D., Scrimgeour, C.M., Nevo, E., Hackett, C.A., Caligari, P.D.S., Forster, B.P., 1997 AFLP variation in wild barley (*Hordeum spontaneum* C. Koch) with reference to salt tolerance and associated ecogeography. *Genome*, Vol.40, No 3, pp 332-341.
- Parks, B.O., 1993 The need for integration. In Goodchild, M.F., Parks, B.O., Steyaert, L.T. (ed.). *Environmental Modeling with GIS*. Oxford University Press, pp 31-34.
- Roekaerts, M., 2002 The Biogeographical Regions Map of Europe : Basic principles of its creation and overview of its development. European Environmental Agency (EEA), European Topic Centre Nature Protection and Biodiversity.
- Shaffer, J.P., 1995 Multiple Hypothesis Testing. *Annual Review of Psychology*. No 46, pp 561-584.
- Skøt, L., Hamilton, N.R.S., Mizzen, S., *et al.*, 2002 Molecular genecology of temperature response in *Lolium perenne*: 2. association of AFLP markers with ecogeography, *Molecular Ecology* Vol.11, No 9, pp 1865-1876.
- Thrupp, L.A., 1998 Linking Biodiversity and Agriculture: Challenges and Opportunities for Sustainable Food Security. World Resources Institute.
- van Tienderen, P., de Haan, A., van der Linden, C., Vosman, B., 2002. Biodiversity assessment using markers for ecologically important traits. *Trends Ecol. Evol.* Vol.17, pp 577-582.
- Vuilleumier, S., 2003 Dispersal modelling : integrating landscape features, behaviour and metapopulations. Phd Thesis, SSIE, EPFL, No 2878.