

# Ph.D. Showcase: Very High Resolution Digital Elevation Models (VHR DEMs) and multiscale landscape genomics analysis applied to an alpine plant species

PhD Student Leempoel Kevin PhD candidate Geographic Information Systems Laboratory EPFL - LASIG, Station 18, CH-1015 Lausanne, Switzerland +41 21 693 57 84 kevin.leempoel@epfl.ch	Stucki Sylvie PhD candidate Geographic Information Systems Laboratory EPFL - LASIG, Station 18, CH-1015 Lausanne, Switzerland +41 21 693 57 84 sylvie.stucki@epfl.ch	Parisod Christian PhD Laboratory of evolutionary botany University of Neuchâtel CH-2000 Neuchâtel, Switzerland +41 32 718 22 44 christian.parisod@unin e.ch	PhD Supervisor Joost Stéphane PhD Geographic Information Systems Laboratory EPFL - LASIG, Station 18, CH-1015 Lausanne, Switzerland +41 21 693 57 82 stephane.joost@epfl.ch
---	--	--	--

## Categories and Subject Descriptors

J.3 [Life and medical sciences]: Biology and genetics

## General Terms

Measurements, Theory

## Keywords

DEM derivatives, Landscape genomics, multiscale, spatial analysis

## 1. ABSTRACT

In landscape genomics, choosing the appropriate environmental parameter is primordial. While most studies are looking at large scale associations between genetic markers and environmental variables, adaptation of plants can also happen on a fine scale. However, the predictive ability of large scale predictors is weak when considering microclimates and landscape heterogeneity in mountainous areas. In that context, DEM's derivatives have been proven useful but remain confidential. This research focuses on the alpine herb *Biscutella laevigata* subsp. *laevigata* L., an autotetraploid perennial plant, which occurs in small patches in a heterogeneous landscape. The objectives are firstly, to explore correlations between environmental variables, exclusively derived from two DEMs of different spatial resolutions, respectively 25 and 2 meters, and genetic markers. Secondly, to analyze the effect of different spatial resolutions on significant correlations. The most important result obtained is the large congruency between

results from a population genomics and a spatial coincidence approach. In addition, three variables show interesting results that is altitude, duration of insolation and morphometric protection index, which is rarely used. And finally, better significances were found most of the times with the 25m DEM, which raises questions about the most appropriate scale of computation of geomorphological and environmental processes.

## 2. INTRODUCTION

Environmental parameters computed from remote sensors have often been used in the context of ecological and landscape genomics studies (Manel *et al.*, 2003; Joost *et al.*, 2007). However, these large scale predictors are useless when considering microclimates and landscape heterogeneity in mountainous areas. Nevertheless, environmental variables derived from Digital Elevation Models (DEM) remain underexploited (Manel *et al.*, 2010). DEM's derivatives such as slope, aspect and curvature have been used in prediction models of plant species distribution (Guisan & Harrel, 2000; Jelaska *et al.*, 2003; Lassueur *et al.*, 2006) but rarely for gene distribution, despite signs of fine-scale adaptation in heterogeneous landscapes (Linhart & Grant, 1996; Vekemans & Hardy, 2004). In addition, the coarse resolution of the previous generation of DEMs restrained their utilization. The predictive ability of these derivatives depends mostly on the resolution of the DEM and on the size of the calculation window, used to take into account the surrounding values. In fact, the presence/absence of a species or gene does not depend only on the environmental properties at their specific location but are also likely to depend on their immediate environment. This aspect was developed in Lassueur *et al.* (2006) for species distribution in the Swiss Alps but with a limited number of DEM's derivatives.

The objectives of this study are firstly, to explore correlations between environmental variables, exclusively derived from two DEMs of different spatial resolutions, respectively 25 and 2 meters, and genetic markers (AFLPs; Amplified Fragment Length Polymorphisms) developed for the alpine plant species *Biscutella laevigata* L. (Brassicaceae), in order to detect potential signatures of natural selection within its genome. We expect from a previous

study on *B.laevigata* a correlation with total solar radiation, altitude and surrogates of humidity, such as the wetness index (Lassueur *et al.* 2006; Parisod & Joost, 2010). Secondly, to analyze the effect of different spatial resolutions on significant and nearly significant correlations. Previous studies show an enhanced prediction with more accurate elevation models (Lassueur *et al.*, 2006).

### 3. MATERIAL AND METHODS

#### 3.1 Species and molecular data

The study zone is situated in the Swiss Prealps (external Alps) at “Les Rochers-de-Naye” (N46°26’00’’ E6°58’50’’). Altitude ranges between 1440 and 2042m. This research focuses on the alpine herb *Biscutella laevigata* subsp. *laevigata* L., an autotetraploid perennial plant, which occurs in small patches, exclusively in warm and dry areas of the European Alps. Gene dispersal is commonly restricted, because pollination is achieved by generalist Diptera and Lepidoptera while seeds are heavy and mainly dispersed by gravity and strong wind (Parisod & Bonvin, 2008). *B. laevigata* survived the last glacial maximum in several locations before recolonizing the Alps. At les Rochers-de-Naye, *B. laevigata* forms a natural hybrid zone between closely related lineages and for which divergent genotypes are found in contrasted habitats related to radiation regimes (Parisod & Christin, 2008). This particular population was thus surmised as being adapted to its local environment. The genomes of 232

individuals were sampled in 2 x 2m plots every 12.5m in 2003, scanned using AFLPs to generate 152 anonymous markers. Part of this dataset was analyzed in Parisod & Christin (2008). This systematic sampling strategy provides its natural distribution.

#### 3.2 DEM datasets

Two DEMs acquired by the Swiss Federal Office of Topography (swisstopo) were used. The 25m resolution model was interpolated from contour lines of the national map at 1:25 000. The accuracy is estimated at ±2m in the Prealps (swisstopo, 2011). The 2m resolution model is based on LIDAR data (Light Detection and Ranging) obtained from an airborne platform since 2000 with an altitude accuracy of ±50 cm (swisstopo, 2011). This DEM is only available under 2000 meters. Both elevation models are projected in CH-1903 (New Swiss grid), which will be used for all the derivatives as well.

#### 3.3 Computation of environmental variables

SAGA (Böhner, 2006) was used to derive environmental variables from the two DEM subsets. Derivatives shown in table 1 were computed.

To assess correlations between the different DEM’s derivatives at sampling plots, R<sup>2</sup> were computed for each pair of variables. Bonferroni correction for multiple tests was applied here as well.

**Table 1 List of derivatives computed in SAGA GIS**

Variable	Description	Formula/Units
Slope	Proxy for water flow, snow movements, erosion and organical deposition	radians
Curvature	Affect snow accumulation which influences directly the growing season duration and water availability (Zevenbergen and Thorne, 1987).	radians/m
Profile curvature		
Plan curvature		
Terrain ruggedness index (TRI)	Quantitative measure of topographic heterogeneity (Riley et al., 1999).	no unit
Saga Wetness Index	Quantifies topographic control on hydrological processes	$W = \frac{a}{\ln S}$ <p>Where <i>a</i> is the specific catchment area and <i>S</i> the slope (Olaya, 2004).</p>
Morphometric protection index (MPI)	Expresses the protection of a point from the surrounding relief. It is based on the maximum angle found at the zenith or at nadir from the point over a defined radius (Yokohama, 2002).	no unit  Value is negative when the point is not protected and positive when it is.

Total Solar radiation	Sum of direct and diffuse insolation. Computed over a period of 243 days (1-Mars to 31-October) with a time-step of 5 days (Boehner & Antonic, 2009; Wilson & Gallant, 2000).	kwh/m <sup>2</sup>
Duration of insolation	Sum of hours of direct insolation.	hours

Environmental variable	AFLP Marker	$\beta_1$ p-value	Efron R <sup>2</sup>
Protection Index (25)	v25	1.03E-11	0.28
Altitude (25)	v25	1.04E-11	0.46
Altitude (2)	v25	1.09E-11	0.45
Protection Index (2)	v25	2.43E-09	0.38
Duration of insolation (25)	v25	8.03E-09	0.27
Profile curvature (25)	v25	2.43E-08	0.14
Duration of insolation (2)	v25	3.48E-08	0.30
Protection Index (25)	v42	7.88E-08	0.13
Altitude (2)	v15	1.91E-07	0.15
Altitude (25)	v15	1.92E-07	0.14
Altitude (2)	b50	3.06E-07	0.16
Altitude (25)	b50	3.14E-07	0.16
Protection Index (25)	b80	3.42E-07	0.12
Wetness Index (25)	v25	5.50E-07	0.20
Plan curvature (2)	b80	1.14E-06	0.13
Duration of insolation (25)	v34	2.12E-06	0.11
Protection Index (25)	v15	2.21E-06	0.08

### 3.4 Computing association models between environmental variables and genetic markers

MatSAM ([www.econogene.eu/software/sam/](http://www.econogene.eu/software/sam/)) is a Matlab® program designed to process many logistic regressions between explanatory variables and genetic markers in parallel (Joost *et al.*, 2007). It is based on the spatial coincidence between the different levels of information (Goodchild, 1996) and is similar to other population genomics approaches that aim at identifying genes associated with environmental parameters (i.e. signature of the action of natural selection on DNA sequences). Unlike corresponding methods, such as BayeScan (Foll and Gaggiotti, 2008), it is however independent from any theoretical model of population genetics, such as the Hardy-Weinberg equilibrium.

Several methods exist in MatSAM to assess the significance of association models, among which the Wald test (Hosmer & Lemeshow, 2000) whose results are shown here. The null hypothesis is that the model with the examined variables does not explain the observed distribution better than a model with a constant only. The conservative Bonferroni correction for multiple tests was applied to assess p-values. In parallel, the Efron pseudo R<sup>2</sup> (Freese *et al.*, 2006) was computed to compare the percentage of variance explained by the variables derived from the two DEMs used.

The same data set was analyzed with a software implementing a theoretical approach in population genomics. BayeScan was developed to directly estimate the probability for each locus to be under natural selection, using a Bayesian method. This approach is not able to provide indications on the nature of a possible selective pressure.

## 4. RESULTS

### 4.1 Correlations between environmental variables at sampling plots

The correlation between DEM25 and DEM2 is very high (>0.999). However, the first derivatives (i.e. slope, curvature, plan curvature and profile curvature) show important differences between DEMs. MPI (25) shows a better correlation with altitude (-0.93) than MPI (2) (-0.81). For insolation variables, correlations with altitude are higher with the 2m derivatives (Total insolation - (DEM25) 0.70; 0.61 - (DEM2) 0.76; 0.79; Duration of insolation - (DEM25) 0.84; 0.84 - (DEM2) 0.88; 0.88).

On 200 correlations tests between environmental variables, 93 significant tests were counted. Six of the 10 variables extracted from DEM2 are significantly correlated with their counterpart from DEM25.

**Table 2 Sorted significant values of  $\beta_1$  with resolution between brackets expressed in meters**

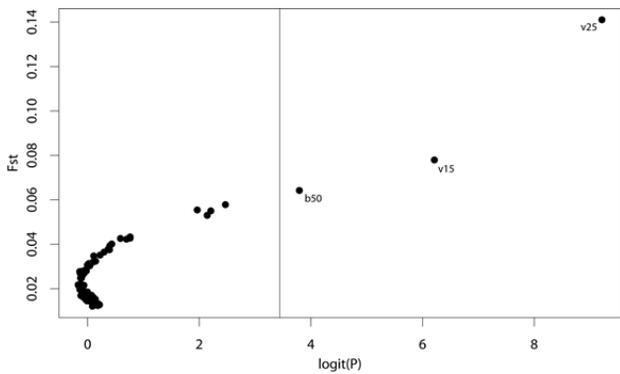
### 4.2 Correlations between molecular markers and environmental parameters

As shown in Table 2, the most significant models associate marker 'v25' with altitude, morphometric protection index and duration of insolation. These 3 pairs of variables are significant for both DEMs, with smaller p-values for the DEM25. Profile curvature and wetness index are also significant with this marker, but only for the DEM25.

There is no particular trend for R<sup>2</sup>. However, the highest R<sup>2</sup> values are observed with marker 'v25' (e.g. Altitude (25), Altitude (2) and Protection Index (2)).

### 4.3 Validation with a population genomics theoretical approach

Seven AFLP bands were detected to be possibly under natural selection by BAYESCAN, of which two loci (v25 and v15) showed a Bayes Factor > 99, meaning a decisive evidence for selection. Moreover, marker b50 showed a very strong evidence for selection.



**Figure 1** BayeScan analysis with the posterior probability for a locus to be under natural selection on the x-axis (log scale). The vertical line is the threshold beyond which there is a decisive evidence for selection

## 5. DISCUSSION

The most important result obtained in this study is that both MatSAM and BayeScan approaches are largely congruent. Indeed, marker V25 is significantly associated with several environmental variables.

MPI may be a very interesting factor in landscape genomics studies over small scales. In fact, in mountain areas, protection from meteorological phenomenon, which are stressing factors, such as strong winds, is very likely to play an important role in the adaptation of species. In this case, individuals on the ridge grow in locations with low MPI value and are very exposed to wind compared to individuals of the lower population. In addition to total solar radiation that turned out to be significant on a smaller dataset on the ridge (Parisod & Christin, 2008), significant association was found with duration of insolation. This shows that deeper investigations on the influence of radiation processes on plants and genes distributions are required.

The influence of altitude on the spatial distribution of species and genes is well-documented. However, this parameter is a very general proxy that is associated with a multitude of proximal ecological factors. Indeed, altitude is a surrogate for temperature, pluviometry and radiation properties (Körner, 2003). Here, altitude and duration of insolation show a strong correlation and are inversely correlated with MPI. In fact, markers v25, v15, b50 and b80 - showing a significant association with these variables - are present in individuals located on the ridge only, at a higher altitude and poorly protected. This implies a high duration of insolation. It shows that these four genomic regions are likely to be selected by the three environmental variables mentioned here above, in dry and sunny areas, as mentioned by Parisod & Christin (2008). In contrast, individuals from the lower population are protected by surrounding mountains resulting in a lower duration of insolation.

Lassueur *et al.* (2006) found that the 25m DEM had an overall similar predictive ability for species distribution than a Laser DEM (<2m). With one exception for aspect, a rough alternative to energetic processes, which provided a better prediction when using a Laser DEM. On the contrary, in this study better significances were most of the times found with the 25m DEM. This result is interesting in the context of the generalization of environmental processes—and raises questions about the most appropriate scale of computation of geomorphological and

environmental processes. In fact not only the resolution is likely to play an important role in the enhancement of p-values, the window size (i.e. the neighborhood of the point of interest) and the parameterization of the derivatives' computation are not trivial to assess and will be the subject of further analysis.

## 6. PERSPECTIVES

Using SenseFly® technology (Zufferey *et al.*, 2010), a new Ultra High Resolution (UHR) DEM (25x25 cm) will soon be generated. This DEM will be processed by well-known image analysis approaches to compute multiscale derivatives and to process additional association models (Kalbermatten, 2010). Indeed, wavelet transforms and structure tensors permit to process new terrain indicators such as energy, coherency and orientation (Kalbermatten, 2010), whose possible correlation with molecular markers' frequency may reveal new insights useful to improve our understanding of adaptation processes.

## 7. REFERENCES

- [1] Bonin A, Miaud C, Taberlet P, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution* 23, 773-783.
- [2] Böhner, J., McCloy, K.R., Strobl, J. [Eds.] (2006): SAGA – Analysis and Modelling Applications. Göttinger Geographische Abhandlungen, Vol.115, 130pp
- [3] Boehner, J., Antonic, O. (2009): Land Surface Parameters Specific to Topo-Climatology. in Hengl, T. & Reuter, H.I. [Eds.]: *Geomorphometry - Concepts, Software, Applications*.
- [4] Foll, M. & O. Gaggiotti (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics*, 180, 977-993.
- [5] Freese, Jeremy and J. Scott Long. *Regression Models for Categorical Dependent Variables Using Stata*. College Station: Stata Press, 2006.
- [6] Goodchild, M. F. 1996. Geographic Information Systems and spatial analysis in the social sciences. In *Anthropology, space, and Geographic Information Systems*, eds. M. Aldenderfer & H. D. G. Maschner, 214-250. New York: Oxford University Press.
- [7] Guisan, A., Harrell, F.E., 2000. Ordinal response regression models in ecology. *J. Veg. Sci.* 11, 617–626.
- [8] Hosmer D.W. & Lemeshow S. (2000) *Applied Logistic Regression*. John Wiley & Sons, New York.
- [9] Jelaska, S.D., Antonic, O., Nikolic, T., et al., 2003. Estimating plant species occurrence in MTB/64 quadrants as a function of DEM-based variables—a case study for Medvednica Nature Park Croatia. *Ecol. Model.* 170, 333–343.
- [10] Joost, S., Bonin, A., Bruford, M.W., Després, L., Conord, C., Erhardt, G., Taberlet, P., (2007) A Spatial Analysis Method (SAM) to detect candidate loci for

- selection: towards a landscape genomics approach to adaptation, *Molecular Ecology*, 18:3955–3969.
- [11] Kalbermatten, M., Multiscale analysis of high resolution digital elevation models using the wavelet transform. Thèse EPFL, no 4610 (2010). Dir.: François Golay.
- [12] Körner, C. 2003. *Alpine Plant Life*, 2nd edition. Heidelberg, Germany Springer.
- [13] Lassueur, T., Joost, S. & Randin, C.F. (2006) Very high resolution digital elevation models: do they improve models of plant species distribution? *Ecological Modelling*, 198, 139–153.
- [14] Linhart YB, Grant MC. 1996. Evolutionary significance of local genetic differentiation in plants. *Annual Reviews in Ecology and Systematics* 27: 237–277.
- [15] Manel S, Schwartz MK, Luikart G and Taberlet P (2003) Landscape genetics: Combining landscape ecology and population genetics. *Trends Ecol Evol* 15:189-197.
- [16] Manel, S., Joost, S., Epperson, B. K., Holderegger, R., Storfer, A., Rosenberg, M. S., Scribner, K. T., Bonin, A. and Fortin, M.-J. (2010), Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, 19: 3760–3772. doi: 10.1111/j.1365-294X.2010.04717.x
- [17] Olaya, V. F., 2004. A gentle introduction to SAGA GIS. The SAGA User Group e.V., Gottingen, Germany, p. 208.
- [18] Parisod C, Bonvin G. 2008. Fine-scale genetic structure and marginal processes in an expanding population of *Biscutella laevigata* L. (Brassicaceae). *Heredity* 101: 536-542
- [19] Parisod C, Christin (2008) Genome-wide association to finescale ecological heterogeneity within a continuous population of *Biscutella laevigata* (Brassicaceae). *New Phytologist*, 178, 436–447.
- [20] Parisod C, Joost S. 2010. Divergent selection in trailing- versus leading-edge populations of *Biscutella laevigata*. *Annals of Botany* 105: 655-660.
- [21] Riley, S.J., De Gloria, S.D., Elliot, R. (1999): A Terrain Ruggedness that Quantifies Topographic Heterogeneity. *Intermountain Journal of Science*, Vol.5, No.1-4, pp.23-27
- [22] Swisstopo, 2011. <http://www.swisstopo.admin.ch/internet/swisstopo/fr/home/products/height.html> . Accessed May 31, 2011, last modification May 30, 2011.
- [23] Vekemans X, Hardy OJ. 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* 13: 921–935.
- [24] Wilson, J.P., Gallant, J.C. [Eds.] (2000): *Terrain Analysis - Principles and Applications*. New York, John Wiley & Sons, Inc. pp.91-106
- [25] Zevenbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topography. *Earth Surf. Process.* 12, 47–56.
- [26] Zufferey JC, Beyeler A, Floreano D. 2010. Autonomous flight at low altitude with vision-based collision avoidance and GPS-based path following, *Proceeding of the IEEE International Conference on Robotics and Automation Anchorage Convention District May 3-8, 2010, Anchorage, Alaska, USA.*