Introduction

To carry out this lab, we use GIS software (Quantum GIS and OpenGeoDa) and a specific landscape genomic software implementing a correlative approach to detect signatures of natural selection and spatial statistics named Samβada. Please refer to their respective documentation as regards their functioning, and as regards the theories and methods they implement. This information is available here: http://lasig.epfl.ch/dgs.

Data used

The main dataset we will use was made available by Andrew Eckert and is about loblolly pine (*Pinus taeda* L.). The dataset was published in two papers:

- Eckert, A. J., A. D. Bower, S. C. González-Martínez, J. L. Wegrzyn, G. Coop and D. B. Neale. 2010. Back to nature: Ecological genomics of loblolly pine (Pinus taeda, Pinaceae). Molecular Ecology 19: 3789-3805.
- Eckert, A. J., J. van Heerwaarden, J. L. Wegrzyn, C. D. Nelson, J. Ross-Ibarra, S. C. González-Martínez and D. B. Neale. 2010. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (Pinus taeda L., Pinaceae). Genetics 185: 969-982.

The two papers are available here: http://lasig.epfl.ch/dgs

Main goal of the exercise

We will mainly use Eckert et al's Genetics paper as a reference in this exercise. Our goal will be to compare the results Eckert et al. obtained with the results we will get using QGIS, Samβada and OpenGeoda.

A. Geographic and environmental context with Quantum GIS

Loblolly pine (*Pinus taeda* L.) is distributed throughout the southeastern United States, ranging from Texas to Delaware. Its range is divided primarily by the Mississippi River Valley, with 60% of the distribution range located east of the Mississippi River.
The samples are geo-referenced by county (n=188). The average number of sampled trees per county is 4 ± 6 (range: 1–67).
Finally, as loblolly pines are only located in the southeastern part of the USA, Eckert et al. focused on aridity as the main environmental parameter considered in the study.

1) Launch Quantum GIS (QGIS), then in the menu **PROJECT** choose **SAVE AS…**, navigate to the data folder, name your project "loblollyPine.qgs" and **SAVE**.
   We will now add the different geo-environmental layers necessary to implement our landscape genomic analysis.
2) In the **LAYER** menu, choose **ADD VECTOR LAYER…**, click on **BROWSE**, navigate to the data folder, select the file SWUScounties.shp and click on **OPEN**. In the Coordinate Reference System Selector, in the window entitled Coordinate reference systems of the world, select **WGS84** and click on **OK**. A map showing the US southeastern counties will be displayed.

3) In the LAYER menu, choose ADD VECTOR LAYER..., click on BROWSE, navigate to the data folder, select the file mississipi.shp and click on OPEN. In the Coordinate Reference System Selector, in the window entitled Coordinate reference systems of the world, select WGS84 and click on OK. A layer containing the Mississipi river has been added in the LAYER MENU on the left.
4) In the LAYER menu, choose ADD RASTER LAYER..., navigate to the data folder, select the file GTOPO30.grd and click on OPEN. A layer containing the Digital Elevation Model GTOPO30 (spatial resolution of ~1km) has been added in the LAYER MENU on the left.
5) In the LAYER menu, choose ADD RASTER LAYER..., navigate to the data folder, select the file prec_6_USA_WGS84.sdat and click on OPEN. A layer containing the Digital Elevation Model GTOPO30 (spatial resolution of ~1km) has been added in the LAYER MENU on the left.
6) With the help of the documentation (or not!), find how to:
    a. Change the superposition order of the layers
    b. Change the color of the counties
    c. Change the color of the river
    d. Change the color of the raster layer

7) Now we need to include the loblolly pine dataset. It is a .txt file made of 4 columns: id, state, latitude and longitude.
In the LAYER menu, choose ADD DELIMITED TEXT LAYER..., browse your files and select loblollyPineSamples.txt. Choose .tab as file format, check FIRST RECORD HAS FIELD NAMES, and click OK. In the Coordinate Reference System Selector, in the window entitled Coordinate reference systems of the world, select WGS84 and click on OK.The individuals are now displayed on the map. Place this layer on the top and change the color of the points to violet and their size to 1.5mm.
8) Now we will learn how to transfer environmental data from the raster layers to the vector loblolly pine dataset. This permits to characterize the sampling location with the environmental information chosen for analysis. To this end, we first need to install an extension to QGIS named "Point sampling tool". In the menu PLUGINS, choose MANAGE AND INSTALL PLUGINS... and in the search field write "Point sampling tool". The extension will be listed in the central column; click on it and then click on INSTALL PLUGIN (bottom right). Once done, click on CLOSE and the icon of the new extension will appear in the menu bar  . Be sure that the loblollyPineSamples layer is selected, and then click on the Point sampling tool icon. In the GENERAL panel, there is only one layer with sampling points and it is correctly selected. In the second window, you have to select the layers from which you want to retrieve data. In our case, we need to keep the id of the samples (**very important**!!!), so select the first line (id sourcepoint), and then go to the bottom of the list and select the raster layer with precipitation (band 1) and the raster layer with elevation (band 1). In the OUTPUT POINT VECTOR LAYER, browse the folders to the right location, indicate the name of the resulting layer (e.g. loblollyEnv) and check the box ADD CREATED LAYER TO THE TOC so that the new layer directly appears in the LAYER MANAGER on the left. Check the fields in (we should have 3 variables: id, precipitation and elevation) and click OK. Then confirm the coordinate reference system (WGS84) and clik OK and close the Point sampling tool extension. The new layer appears then in the LAYER MANAGER.

9) To check if the transfer was successful, select the new layer and click on the OPEN ATTRIBUTE TABLE button  . You should see the 3 columns we selected before.

Now you know how to characterize your sampling locations with geo-environmental data by means of GIS. Key information is environmental data, and the environmental data you look for need to fit the geographical scale of the study (in our case we have individuals geo-referenced at the scale of the US county (=grain) throughout the whole US southeastern area (=extent) , i.e. a large geographical scale, and thus we used environmental data with a coarse spatial resolution.

Data sources
There are now many geo-environmental (including climate data) available on the Internet. An important data source is http://earthexplorer.usgs.gov/. Several other sources are listed here: http://dyerlab.bio.vcu.edu/teaching/dgs2014/.

Projection systems
In this exercise, we use datasets with geographical information expressed in longitude/latitude. It is perfect for the kind of manipulation carried out here (spatial coincidence), but this does not allow for a number of operations requiring a metric system (to calculate distances for example). Indeed, with a system in longitude and latitude, the coordinates are expressed in degrees, minutes and seconds. To use a metric system, we need to project the geodata in a plane surface, considering the resulting deformation. There are many different coordinate projection systems with a number of specificities, among which the location on the Earth. For example, the Universal Transverse Mercator (UTM) system divides the Earth into sixty zones, each a six-degree band of longitude, and uses a secant transverse Mercator projection in each zone.

For more information, see Joost et al. (2010) Animal genetics. The paper is available here: http://lasig.epfl.ch/dgs

B. Identify signatures of natural selection with Samβada

To save time, we prepared the data required to implement the correlative approach with Samβada. The software is available here: http://lasig.epfl.ch/sambada. Together with the different versions of the application, you will find the documentation named ["SamBadoc_v.0.9.pdf]" (please have a look). Samβada is a command line software and you will need the documentation to learn the different parameter names. Samβada is a successive version of the MatSAM software written several years ago (Joost et al. 2007).
We propose that you create a new folder named "sambada", and all necessary files used in this exercise will be stored here. This is also where Samβada will write the results.
Well, what we need to run Samβada are the following files:

a) WIN64_Sambada.v09.exe
b) parameters.txt
c) a file with environmental data (lob_ENV.txt)
d) a file with genetic data (lob_GEN.txt)
e) cmd_dos_sambada.bat
f) Sambada-Error-Codes.txt

Elements a) to d) are compulsory. In fact Samβada is able to work with only 1 input file containing environmental and genetic data (see documentation). But here we chose a clear configuration with 2 datasets. Element e) is proposed to facilitate the launching of the software. It is a batch file with the MS-DOS commands to run Samβada. The information provided by element f) is useful when Samβada lists the error numbers during the process.

You can see that there are other Samβada modules:

- WIN64_RecodePlink.v1.exe is useful to recode data from PLINK to Samβada
- WIN64_RecodePlink-LFMM.v1.exe is useful to recode data from PLINK to LFMM
- WIN64_Supervision.v1.exe is useful to split files to be distributed among a computer cluster in the context of HPC computing

We won't use these 3 modules in this exercise.
The next step is to prepare the different files mentioned above before running Samβada. The files with genetic data (lob_GEN.txt) and environmental data (lob_ENV.txt) are ready. Here is a short description of their content.

Genetic data
lob_GEN.txt contains 9248 columns, one with the ID, and 9247 with SNP genotypes recoded 1 (genotype present) and 0 (genotype absent). Example: the first genotype in the dataset is "0-10037-01-257_AA" where "0-10037-01-257" is the SNP and "AA" the alleles.

Environmental data
lob_ENV.txt contains 20 columns. The id, the US state, latitude, longitude will not be processed and will be declared as such in the parameter file (cf documentation). The other variables are described in the file named "data-codebook.txt".

Parameter file
Here is the parameter file with the basic instructions we will use here.

```
INPUTFILE lob_ENV.txt lob_GEN.txt
HEADERS YES
NUMVARENV 20
NUMMARK 9247
NUMINDIV 622
IDINDIV id
COLSUPENV state latitude longitude
DIMMAX 1
SAVETYPE END BEST 0.01
```

INPUTFILE	You must indicate the name of the environmental data file, and the genetic data file with a space as separator between them
HEADERS	Yes = our files contain a header
NUMVARENV	Number of environmental variables
NUMMARK	Number of genetic markers
NUMINDIV	Number of individuals sampled
IDINDIV	Name of the column with the ID
COLSUPENV	Name of additional variables contained within the environmental data

	file but which won't be processed
DIMMAX	0 = a model with a constant only; 1 = model with a constant and 1 explanatory variable; 2 = model with a constant and 2 explanatory variables; etc.
SAVETYPE END BEST	Instructions related to how and what to save (see documentation). Here we decide to save at the end of the processing and to save only the best models corresponding to a significance level of 0.01, Bonferroni correction included.

Batch file
The content of the batch file named cmd_dos_sambada.bat is as follows:

```
cd E:\Lasig\Cours\SC-congenomics\sambada
WIN64_Sambada.v09.exe parameters.txt lob_ENV.txt lob_GEN.txt
```

So here you have to indicate the correct path to your Samβada folder, indicate the correct version of Samβada (WIN32) and that's it. Once done, you are ready and you can run Samβada by double-clicking the batch file.

Results and comparison with Eckert's results
After the end of the processing, Samβada will write 2 files, lob_GEN-Out-0.txt and lob_GEN-Out-1.txt. The first one contains the results for the constant model. Edit it with Notepad++ and see documentation for a description of the different columns. Do the same with the second file. This one contains the most significant models (cf our criterion of 0.01) ranked according to the Wald score. For information, the Wald statistic determines whether the parameters added to a model containing a constant only are all different from 0. The highest the score, the most significant the model. Then [open lob_GEN-Out-1.txt] and:

- Save your results in Excel or OpenOffice for instance, and have a look, sort the models according to different criteria (R2, AIC, BIC), draw some straightforward conclusion (for instance what is the characteristic of all the best models?).
- Then take Eckert et al.'s paper (Genetics 2010), read the results and find the SNP loci identified by both methods. Imagine a table showing results produced by Samβada and Eckert's approach, in which we can read the SNP loci identified, the environmental variables associated and R2.
- Are your findings in accordance with Eckert's results?

C. Use population membership coefficient as covariate in multivariate models

Now, on the basis of univariate models obtained before, we will process bivariate models integrating the coefficient of membership to populations calculated by Admixture. This resembles the approach implemented within LFMM for which parameter beta1 aggregates variance related to environmental variables, and parameter U aggregates the remaining variance attributed to population structure. To this end, we will reduce the number of environmental variables involved and keep only the ones involved in the most significant univariate models. Based on a ranking according to AIC, we will keep the following variables: ti216, bio6, aiq2, bio12, aiq4.
And in addition we will also use variables mbcp1 and mbcp3, which are the membership coefficients as calculated by Admixture. About these coefficients of membership, we will use

QGIS again to visualize them on a map. In our QGIS project, we will first import a specific file showing the individuals sampled in a way that superposed points (identical geographical coordinates) are all visible. So import the file named LoblollyPineVisu.csv in QGIS (In the LAYER menu, choose ADD DELIMITED TEXT LAYER..., browse your files and select LoblollyPineVisu.csv). Then import the file named membership.txt (same way). Then select the layer LoblollyPineVisu in the LAYER MANAGER, and in the LAYER menu, click on PROPERTIES. Click on the JOINS panel, and then on the  button. Select membership as JOIN LAYER, id as JOIN FIELD, id as TARGET FIELD, and check CACHE JOIN LAYER IN VIRTUAL MEMORY, and click OK. Now if you open the attribute table, you will see that variables mbcp1, mbcp2, mbcp3 and Pop will be displayed. Now you can create a thematic map to show the spatial distribution of these coefficients of membership. To this end, double-click on the LoblollyPineVisu layer, select the STYLE panel, in the pop-up menu on the left, select CATEGORIZED instead of SINGLE SYMBOL, in the COLUMN field, select membership_Pop, and then on the button CLASSIFY. You must have 3 categories, so remove any additional one (with DELETE). You can choose the color of the 3 populations by double-clicking the colored symbol and then selecting another color. Then click OK to visualize the spatial distribution of the 3 populations of loblolly pine.

Then adapt the parameter file to run bivariate models based on these variables, and run Samβada. You can use the COLSUPENV parameter to exclude the environmental variables you don't want to include in the analysis. The processing will take ~20 minutes.

Results and comparison with Eckert's results

Do the same like with univariate models and compare the results obtained by Eckert et al. What can you say of the list of bivariate models obtained? Compare also with univariate models you computed before focusing on those containing the same variables only.
On the basis on these analyses, select a SNP locus of interest for spatial analysis according to one or several arguments of yours. In particular, don't forget to consider the frequency of the marker in question across the area of investigation.

D. Analyse spatial autocorrelation of candidate loci with OpenGeoDa

To provide explanations on the use of the OpenGeoda software and on the use of the measure of spatial autocorrelation, we decided to focus on SNP genotype 0-17251-01-147_GG (its frequency is 83%, it is associated with variable ti216). We will first create a shape file to be used by OpenGeoda.
Open the file lob_GEN.xlsx and extract the column displaying the frequency of 0-17251-01-147_GG and the id. Export a .csv file containing these two columns and 623 lines (622 individuals + header). Name this file 0-17251-01-147_GG.csv. Open the file in Notepad++ and substitute question marks with a space ("?" for missing data in Samβada). In QGIS, import this file (when you import, click on NO GEOMETRY) and make a join so that the column "0-17251-01-147_GG" becomes an attribute of the layer LoblollyPineVisu. Then right click on the layer LoblollyPineVisu and select SAVE AS... Select SHAPE as format, keep the current projection system, name your file 0-17251-01-147_GG.shp, browse to the right folder and click OK. This way you have a first spatial dataset you can use with OpenGeoda, but of course you can include in the analysis other loci of interest you have identified with Samβada. Hereunder, we

refer to markers named "M"; please consider that this "M" includes 0-17251-01-147_GG and the other markers you decided to analyse.

Launch OpenGeoda. Click on the left icon in the tool bar to open the shape file. Make the map window larger. We will first assess the global level of spatial autocorrelation of allele frequencies with different weighting criteria (or spatial lags). We first need to create weighting files in order to define the neighborhood of each sampling point. In other words, with global spatial autocorrelation we want to compare the allele frequency at each sampling location with the mean allele frequency at sampling points included in a given neighborhood. Here we will use several numbers of neighbors (20, 40, 60 and 80) to manually build a spatial correlogram on the basis of the Moran's I obtained for each weighting configuration.
Choose **SELECT TOOLS** > **WEIGHT** > **CREATE** to create the 4 corresponding weighting files. Use id as the ID variable. Then select k-Nearest neighbors and indicate 20 in the corresponding field. Don't forget to indicate the weighting criterion in the name of the file (e.g. 0-17251-01-147_GG_20k), otherwise you will face problems later. Be sure that there is no special character in the path to the folder where you save your weighting files, otherwise OpenGeoda won't work. Then repeat the task with 40, 60, 80 neighbors.
A good idea is to open one of the weighting files (.gwt) to understand how it is structured. After the header, you have the list of the 20 nearest points of each of the 622 individuals in the data set, represented by their ID. The distance is indicated in decimal degrees. To calculate the global Moran's I, for each sampling point the software will calculate the mean frequency of a marker M among the 20 nearest neighbors listed in this file, and compare it to the real frequency. When processing a univariate regression of marker M's frequency on the weighted M's frequency, we obtain a slope and this slope is Moran's I (see Anselin 2006 in for more details).
Now we will calculate the Moran's I for each weighting configuration, and for your selection of SNPs. Select **SPACE** > **UNIVARIATE MORAN'S I** and select your marker, and in the next window select a weighting file. You will obtain a Moran scatterplot showing standardized (centered and reduced) values of your marker frequencies vs weighted frequencies of your marker for the corresponding weighting scheme. Moran's I is an indicator of spatial autocorrelation over the whole area under study and for this particular criterion (20K) only.
An important element to check is the significance of the global Moran's I obtained. The significance level is calculated on the basis of random permutations between all sampling points in the dataset, using Monte-Carlo simulations. The Moran's I obtained for the real observed geographic configuration is compared to the Moran's I calculated on the basis of many other spatial configurations obtained by means of n permutations. These permutations mean that the attribute values (the allele frequencies) at the sampling points are exchanged at random (e.g. allele frequency of sampling location 1 goes to sampling location 18, allele frequency of sampling location 4 goes to sampling location 14, etc.). The number of possible spatial configurations is the number of spatial objects factorial (n!). In our case it would be 622!
To experiment randomization, right-click on the Moran scattergram and select **RANDOMIZATION** and then "999" permutations. On the histogram you will observe that no simulated spatial configuration is larger than the Moran's I measured for the observed situation (the reality measured in the field). Read the explanations in the legend of Figure 1.

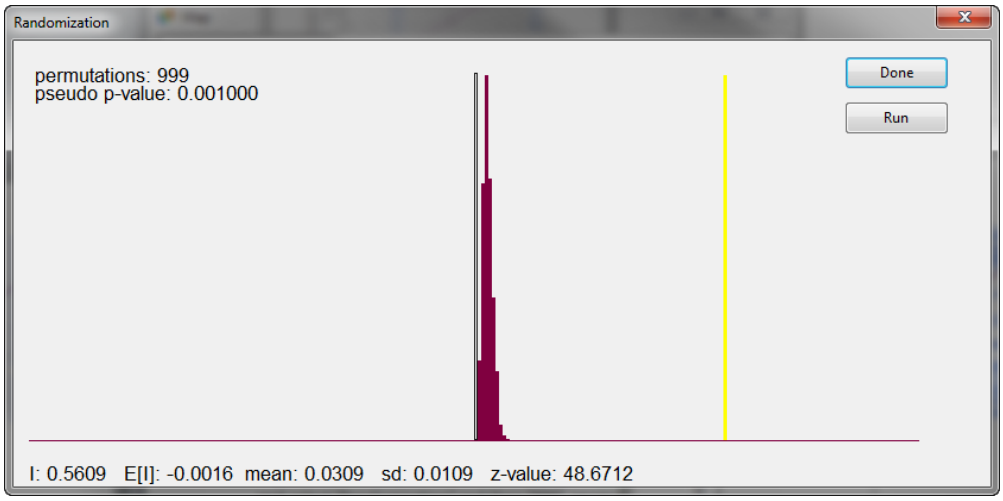


Figure1: In the randomization process, the histogram shows the value of the observed Moran's I with the yellow bar (observed situation), for one loci with K = 20 nearest neighbours. Each time you press the "Run" button, the software runs another set of n permutations. In this case, we ask you to run 999 permutations. The histogram shows the statistical distribution of these 999 Moran's I calculated when the "Run" button was pressed (it means 999 out of 622! configurations). The histogram shows the frequency in Y, and classes of Moran's I in X. Here we can read that the yellow bar (observed Moran's I) = 0.56. The p-value is calculated as the number of simulated Moran's I being larger or equal to the observed Moran's I + 1 divided by the total number of random permutations + 1. Here we have 1/1000 = 0.001. This Moran's I is significant.

Try several times to run 999 permutations (or a larger number of permutations), and check the variation of the Index, to be sure that no spatial configuration will generate a Moran's I larger than the one corresponding to the observed situation.
Then calculate Moran's I for each selected marker and for each weighting configuration; also check the significance level. Then fill a table in Excel (as shown in Figure 2) with all Moran's I for your selected markers and build a spatial correlogram for each of them. Alternatively, in case you master R, you can build the corresponding spatial correlogram (http://cran.r-project.org/web/packages/spatial/spatial.pdf).

Moran's I

	k=20	k=25	k=30	k=35	k=40	k=45	k=50
M1	0.2774						
M2	0.2295						
M3	0.0775						

Significance level for 999 permutations

	k=20	k=25	k=30	k=35	k=40	k=45	k=50
M1	Yes, 0.001						
M2	Yes, 0.001						
M3	No, 0.013						

Figure 2: Table to fill in a spreadsheet in order to build spatial correlograms for analysed markers. For that purpose you will use the upper table named Moran's I. The role of the other table is to collect significance values and to mention if the global spatial autocorrelation measured is significant or not.

Univariate local spatial autocorrelation (LISA)

Choose one of your markers and we will calculate a local index of spatial association (LISA, Anselin 1995), i.e. measure the local spatial autocorrelation of its frequency. Choose SPACE>UNIVARIATE LISA, then select your locus**,** and a weighting file (40 nearest neighbors). Then on the window tick all 4 options. The significance map (Figure 3) shows the significant

sampling locations in green, from dark green (most significant) to light green (less significant).

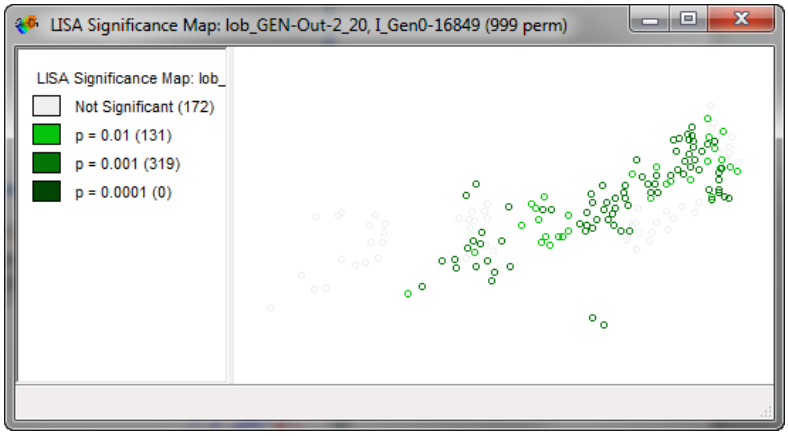


Figure 3: LISA significance map. The colour scale shows light to dark green sampling points. Dark green are the most significant sampling points. White sampling points (almost invisible here) are not significant (there is no spatial dependence, space is neutral).

Sampling locations in white are not significant (=no spatial dependence). For these points, it means that when carrying out random permutations (Monte-Carlo simulations), there is always at least one spatial configuration showing a highest Moran's I than the real situation.

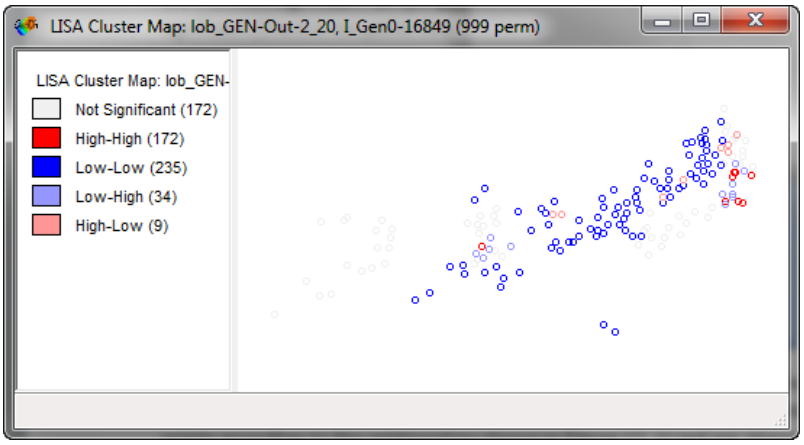


Figure 4: LISA cluster map. We distinctly see significant sampling points (as shown in Figure 3) for which allele frequency is spatially dependent. Here colors are attributed according to the relationships shown in Figure 5. High-High correlations are shown in red: high allele frequency values correlated with high weighted allele frequency values, located in the upper right square of the Moran's scattergram. In blue are shown Low-Low correlations. In light blue are shown low allele frequency values correlated with high weighted allele frequency values and the opposite for light red.

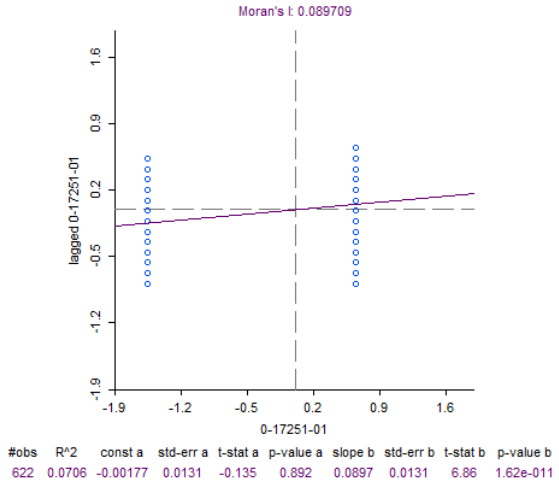
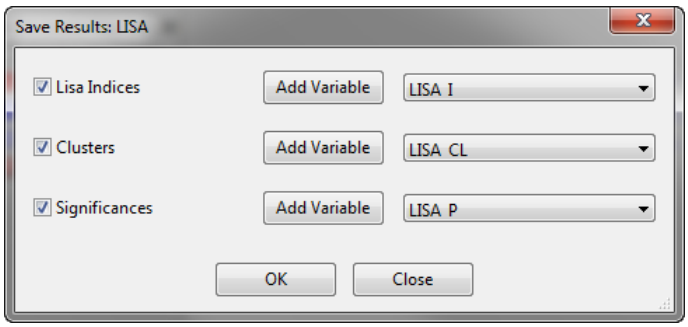


Figure 5: Moran's scatterplot for a marker M and for K = 20 nearest neighbors . 0.089 is the slope and the value of Moran's I (almost global spatial autocorrelation here). We see here the 4 quadrants corresponding to the4 types of LISA clusters (high-high, low-low, high-low and low-high).

Bivariate LISA

To carry out a bivariate LISA analysis, In the menu SPACE, choose BIVARIATE LOCAL MORAN'S I. Use one of the significant environmental variable (e.g. aiq4 also used in Eckert's paper), and use the 40 nearest neighbors as weighting file. Provide a significance map, a cluster map and a Moran's scattergram. Bivariate LISA calculates a local coefficient of correlation between the frequency of your marker and the mean of the selected environmental variable in the considered weighting scheme. It permits to analyse how this relationshsip varies in the landscape. Save the bivariate results by right-clicking on the cluster map and save the LISA variables. The following window opens:

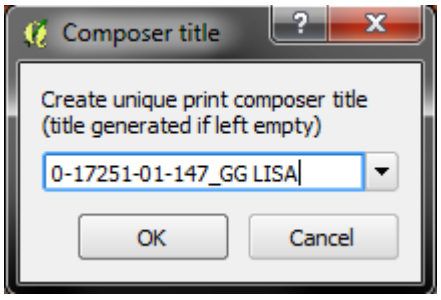


Then, save your modified shapefile (Menu FILE, choose SAVE). We will now use this shapefile in QGIS to edit a final map of these results.

E. Editing maps with Quantum GIS

Open your QGIS project and load your shapefile with LISA indices. Change the style of your data based on categories of LISA clusters. To attribute the correct colors, please refer to the following legend:

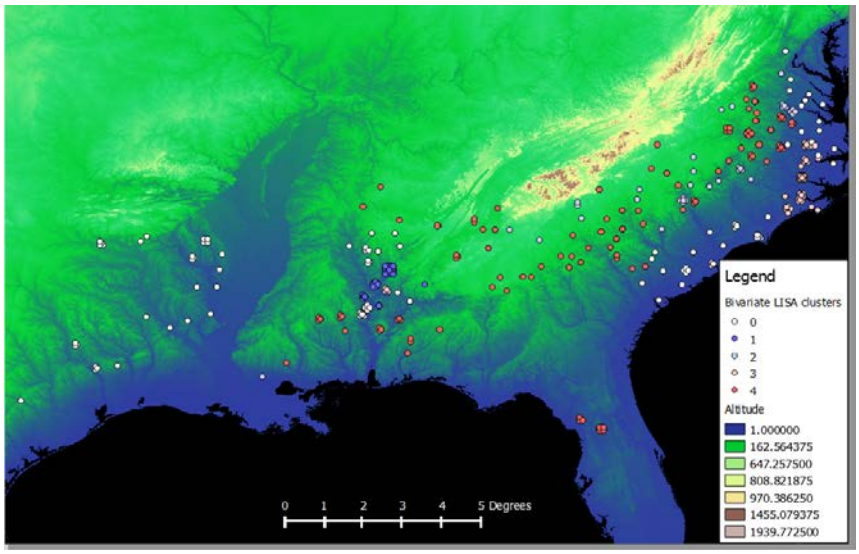
In order to create and export a map as a .pdf or as a .jpg , we need to use the **PRINT COMPOSER**; it is not possible to directly include a legend and a scale bar in the map view. To use the **PRINT COMPOSER**, first activate the layers you want to include, adopt the correct superposition order, set transparency and colours before moving to further steps. Then click on **ADD A NEW PRINT COMPOSER**  , write the title of the composer:



Click on **ADD A NEW MAP**  and draw a rectangle all over the page. On the pannel on the right, you can change the scale and the extent of the map in **ITEM PROPERTIES**. You can also pan the content of the window with the **MOVE ITEM CONTENT** button  .

We now need to add a legend and a scale bar to the map. Click on **ADD A NEW LEGEND**  then click on the map where you want to place it. Corners of the legend object allow you to resize it. In the **ITEM PROPERTIES** of the legend you can change the elements that are shown in the legend, modify their names, play with the transparency, color and frame of the legend background. The following toolbar permits to apply these changes:

To add a scale bar, click on **ADD A NEW SCALE**  . Again you can modify all its propoerties in the **ITEM PROPERTIES** such as defining scale shape, its segments, font size etc. You can also suppress its background by deactivating its frame and putting transparency to zero.
Finally you can export your map in .pdf, .png or any other format available.



F. References

Anselin L. (1995) Local Indicators of Spatial Association - Lisa. Geographical Analysis 27, 93-115.

Anselin L., Syabri I. & Kho Y. (2006) GeoDa: An introduction to spatial data analysis. Geographical Analysis 38, 5-22.

Eckert A.J., van Heerwaarden J., Wegrzyn J.L., Nelson C.D., Ross-Ibarra J., Gonzalez-Martinez S.C. & Neale D.B. (2010) Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (Pinus taeda L., Pinaceae). Genetics 185, 969-82.

Eckert A.J., Bower A.D., Gonzalez-Martinez S.C., Wegrzyn J.L., Coop G. & Neale D.B. (2010) Back to nature: ecological genomics of loblolly pine (Pinus taeda, Pinaceae). Molecular Ecology 19, 3789-805.

Joost S., Bonin A., Bruford M.W., Despres 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 16, 3955-69.

Joost S., Colli L., Baret P.V., Garcia J.F., Boettcher P.J., Tixier-Boichard M., Ajmone-Marsan P. & Consortium G. (2010) Integrating geo-referenced multiscale and multidisciplinary data for the management of biodiversity in livestock genetic resources. Animal Genetics 41, 47-63.