On the use of temporal variation in neutral genetic clines to estimate gene flow: a case study in a fungal plant pathogen.

A. RIEUX, J. CARLIER, L. DE LAPEYRE and V. RAVIGNE

UMR biologie & génétique des interactions plantes/parasites - Campus international de Baillarguet - Montpellier

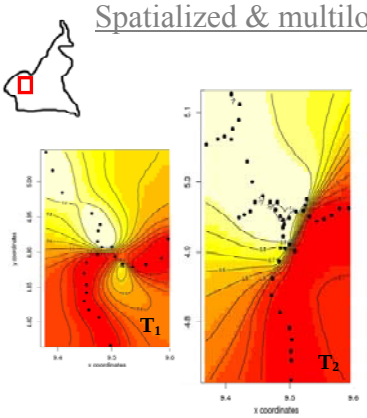
Why studying plant pathogen's dispersal ?

- Drive the colonization processes of species (**ie spread of diseases** in the special case of pathogens)
- Determine the level of gene flow between populations (which **affects their evolutionary potential**)

A good situation to estimate dispersal parameters at the agrosystem scale using neutral genetic clines:

- ***Mycosphaerella fijiensis*** is a haploid pathogen fungus causing black leaf streak disease (BLSD), the most destructive leaf disease of banana. Populations have been shown to be panmictic and the fungus mainly spread through **wind dispersed sexual spores**.

Spatialized & multilocus genetic clustering analyses (*[1])



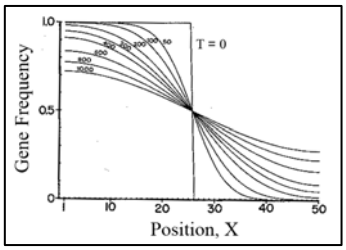
- 2 temporal sampling (T2 – T1 = 15 generations)
- 30 Km long transect
- 400 individuals (4 - 8 isolates/site)
- 19 microsatellites

2 genetic entities without correlation between ecological/landscape features and genetic structure

Locus	F_{ST} (T_1)	F_{ST} (T_2)
424	-0.012	0.015
430	-0.010	-0.010
N203	-0.005	0.034
428	-0.001	0.100
F62	0.001	-0.008
417	0.007	0.060
401	0.016	0.008
403	0.017	-0.009
N137	0.022	0.028
ff09	0.032	-0.005
N194b	0.058	0.005
Fe09	0.060	0.109
F26b	0.103	0.092
434	0.116	0.047
405	0.132	-0.001
Fe05B	0.201	0.033
407	0.220	0.229
413	0.228	0.123
425	0.268	0.262
All loci	**0.077**	**0.058**

Large variance in F_{ST} values among loci.

Historical secondary contact of two expanding populations with high effective size



In theory, the rate of decay is supposed only to be affected by quasi-deterministic spatial **diffusion** processes (σ).

Method of neutral clines analysis

- We only considered loci with 2 alleles and we analyzed the spatial distribution of our data by fitting locus by locus and multilocus allelic frequencies along the transect to classical mathematical functions using a simulated annealing algorithm (*[2]). We compared different models using AIC criterion. Once we obtained the likelihood of the best model (*L*), we calculated the likelihood of two nested models (see below) in order to test for two different hypotheses.

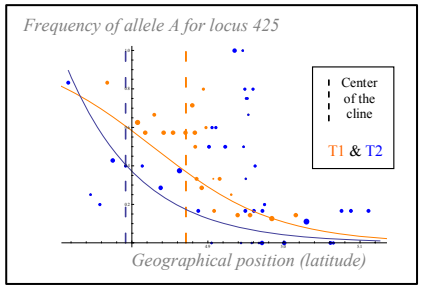
a) We looked for for both locus/locus and multi-locus significant genetic cline pattern in our data by testing if slopes of fitted functions were significantly different from 0. We calculated the likelihood of our data (L_0) under the H_0 hypothesis: "slope of the cline = 0" and we used a X^2 test $_{(ddl=df1-df2)}$ with df1 and df2 the degree of freedom of respectively the global and the nested model to test for significant values of the likelihood ratio test $D=-2\ln(L/L_0)$.
b) In the cases of significant clines patterns (rejected H_0), we tested for a significant variation in the slopes (concordance) of genetic clines between the two temporal samples. For that we compared the likelihood of the sum of the two independent temporal sampling fitted cline models ($L1=L_{1st\ sampling} + L_{2nd\ sampling}$) with the likelihood of a model in which we constrained the slope to be the same in both fitted clines (L_2) using likelihood ratio test $D=-2\ln(L_1/L_2)$.

- If a barrier to gene flow is absent or weak and involves few genes, allelic clines at neutral loci are expected to decay after the secondary contact. In this simple scenario, the width of a neutral cline (w) depends only on the dispersal rate (σ) and the number of generation since contact (T) (*[3]) as (1) $w = \sqrt{(2\pi)}.\sigma\sqrt{T}$. Considering that the width of a cline w = Δp/s with p = allelic frequency of allele *a* calculated in each sites and s = the slope of the cline we can use the temporal variation in cline slopes to infer dispersal as (2) $\sigma = \dfrac{w_2 - w_1}{\sqrt{2\pi}(T_2 - T_1)}$ with w_1 and w_2 = width of clines of the first and second sampling and T_2-T_1 = the time in between the two temporal samplings.

Results

1- When fitting all 19 loci simultaneously, the best model to fit our data was a logit function in both temporal sampling. Slopes of both clines were significantly different from 0 and the slope in the second temporal cline was significantly different and smaller than the slope in the first one).

2- When fitting all loci independently, the best models were the logit function. Only 3 loci out of 19 (425, 407 & Fe05b) showed both slopes significantly different from 0 and significant differences in slopes between the two temporal sampling.



Locus	Slope T1	Slope T2	Centre T1	Centre T2	σ (m/generation)
407	-14.86	-3.95	4.871025	4.79166	**649.2**
425	-18.17	-7.64	5.002994	5.122082	**200.76**
Fe05b	-12.25	-3.39	4.827258	4.772044	**700.09**
All loci	-7.85	-4.34			**322.09**

Magnitude of σ infered from equation (2) (see Method)

Conclusion: Our results show that combining genetic clustering approaches to detect sharp breaks in allelic frequencies distribution with analyses of genetic clines to better characterise their spatial variation offers a convenient way to gain solid information on gene movements.

References:

*[1] Guillot *et al*. (2005) A spatial statistical model for landscape genetics. *Genetics* ;
*[2] Gay *et al*. 2008 Comparing genetic and phenotypic clines in hybrid zones: a window on tension zone models. *Evolution*;
*[3] Endler (1977) Geographic variation, speciation, and clines. *Monogr Popul Biol*