LETTER

Using neutral cline decay to estimate contemporary dispersal: a generic tool and its application to a major crop pathogen

A. Rieux,[1,4]* T. Lenormand,[2] J. Carlier,[1] L. de Lapeyre de Bellaire[3] and V. Ravigné[1]

Abstract

Dispersal is a key parameter of adaptation, invasion and persistence. Yet standard population genetics inference methods hardly distinguish it from drift and many species cannot be studied by direct mark-recapture methods. Here, we introduce a method using rates of change in cline shapes for neutral markers to estimate contemporary dispersal. We apply it to the devastating banana pest *Mycosphaerella fijiensis*, a wind-dispersed fungus for which a secondary contact zone had previously been detected using landscape genetics tools. By tracking the spatio-temporal frequency change of 15 microsatellite markers, we find that σ, the standard deviation of parent–offspring dispersal distances, is 1.2 km/generation$^{1/2}$. The analysis is further shown robust to a large range of dispersal kernels. We conclude that combining landscape genetics approaches to detect breaks in allelic frequencies with analyses of changes in neutral genetic clines offers a powerful way to obtain ecologically relevant estimates of dispersal in many species.

Keywords

Dispersal ecology, emergent disease, gene flow estimation, genetic clines, landscape genetics, *Mycosphaerella fijiensis*, neutral markers, non-equilibrium populations, secondary contact.

Ecology Letters (2014) **16**: 721–730

INTRODUCTION

Virtually every ecological and evolutionary process is influenced by dispersal. From an ecological perspective, dispersal influences population dynamics and persistence, species distribution and abundance, and community structure (Dieckmann *et al.* 1999). It is particularly critical for biological invasions (Lockwood *et al.* 2005). From an evolutionary perspective, dispersal determines the level of gene flow between populations and therefore affects processes such as local adaptation and species range (Lenormand 2002), speciation (e.g. reviewed in Barton 2001), and the evolution of life history traits (examples in Ronce 2007). For pathogens, dispersal results in disease spread (Brown & Hovmøller 2002) and might condition their evolutionary potential (McDonald & Linde 2002).

Despite its importance, estimating dispersal in natural populations remains challenging (Rousset 2001). There is a variety of techniques ranging from direct observation to indirect (genetic-based) measures of dispersal-related processes. Each method has drawbacks and advantages (Rousset 2001; Broquet & Petit 2009). For instance, only genetic-based inference methods are able to quantify effective gene (rather than individual) movements but contrary to direct methods, which rely on few assumptions, they critically depend on an appropriate modelling of a combination of underlying processes. This occurs in models inferring dispersal from patterns of genetic variation expected at drift–dispersal equilibrium (e.g. isolation by distance, Rousset 1997) or selection–dispersal equilibrium (e.g. tension zones, reviewed in Barton & Hewitt 1985). Models at drift–dispersal equilibrium present the specific drawback that they require independent information on effective population densities to infer dispersal rates

(Slatkin & Barton 1989; Rousset 1997). Besides, the provided estimates reflect both historical and contemporary dispersal events, hampering comparisons with direct measurements. Parentage assignment analyses have been developed to infer contemporary dispersal (reviewed in Broquet & Petit 2009) but they rely on an extensive sampling of both parents and their progeny, which is infeasible in many species. Alternatively, methods at selection–dispersal equilibrium consider clines maintained, either by selection against hybrids between two taxa (tension zones) or by different selection pressures in geographically adjacent areas. In such situations, provided multilocus clines are observed, measures of linkage disequilibria between loci allow estimating short-term dispersal parameters (Barton 1982; Barton & Hewitt 1985 for tension zones, Lenormand *et al.* 1998 for local adaptation). There also have been attempts to exploit non-equilibrium situations such as (1) cline temporal variation (e.g. the speed of a wave of advance of a Wolbachia infection, Turelli & Hoffmann 1991) or (2) cline shape rate of change when selection or migration varies (Lenormand & Raymond 2000; Saccheri *et al.* 2008). Despite its potential for estimating dispersal, genetic cline theory (whether or not at equilibrium) has seldom been used in comparison to other methods. One possible explanation could be that situations near dispersal-selection equilibrium except hybrid zones are difficult to identify in natural populations. The aim of this article was to illustrate how contact zones in a purely neutral situation can be useful to estimate dispersal in various species.

Genetic clines, that are spatial gradients in gene frequencies, may be observed for neutral polymorphic markers when divergent populations meet after having been separated by some climatic or geological change or following introduction into a new region

[1]CIRAD, UMR BGPI, Campus international de Baillarguet, TA A-54K, F-34398, Montpellier Cedex 5, France
[2]CEFE-CNRS, UMR 5175, 1919 route de Mende, F-34293, Montpellier Cedex 5, France
[3]CIRAD, Persyst, UPR Syst. Banan. Ananas, TA B-26/PS4, Blvd. de la Lironde, 34398, Montpellier Cedex 5, France

[4] *Present address: Department of Genetics, Evolution & Environment, Faculty of Life Sciences, UCL Genetics Institute, London, WC1E 6BT, UK*
Correspondence: E-mail: a.rieux@ucl.ac.uk

[Copyright line has been changed on 22 July 2014, after first online publication.]

(secondary contact). Such neutral clines are not stable through time (Endler 1977). Provided that there is no selection (either antagonistic between the areas or against hybrid genotypes), the initial step in allele frequencies should gradually dissolve, leaving only a smooth vanishing cline. Under conditions detailed below, Endler (1977) found a relationship between neutral cline width, migration intensity and the time since contact. Surprisingly, this theory has not been used yet to estimate dispersal in natural populations, probably for several reasons. First, detecting secondary contact zones may be difficult. Accurate localisation of contact zones requires two-dimensional sampling strategies and spatially explicit methods of analysis. With the recent emergence of landscape genetics, such methods now exist (reviewed in Guillot et al. 2009). Second, time since historical contact is rarely known. Third, Endler's formula applies in absence of selection or drift. Its application therefore requires diagnosing whether drift and selection can be neglected in the studied system. Last, an important underlying assumption is that migration amounts to a diffusion process, which may not be true in biological systems where long-distance dispersal events occur.

In this article, we show how the rate of change of neutral cline width over a known number of generations allows estimating dispersal in natural populations in absence of information about contact time. We further show that, using a maximum likelihood framework, one may test whether clines evolve under the effect of migration only. We apply this method to the fungus *Mycosphaerella fijiensis*, a wind-dispersed plant pathogen causing black leaf streak disease (BLSD) of banana. BLSD is the most economically important disease of bananas (Jones 2000) and is among the most important crop diseases in the world (Pennisi 2010). A better knowledge of dispersal is particularly required to improve strategies of disease control and fungicide resistance management (Lenormand & Raymond 1998; de Lapeyre de Bellaire et al. 2010). *Mycosphaerella fijiensis* has recently spread worldwide from South-East Asia (Carlier 2000). It has been first mentioned in 1973 in Africa and in 1980 in Cameroon (Mourichon & Fullerton 1990). All previous attempts to measure its dispersal through indirect methods have failed, because its population structure reflects historical patterns of colonisation rather than current equilibrium between evolutionary forces (Halkett et al. 2010; Rieux et al. 2011). In a previous landscape genetics study in South-West Cameroon, Rieux et al. (2011) have detected a contact area between two expanding populations. Their contact time is unknown. In the following, we use spatially explicit genetic clustering to show that the genetic discontinuity previously detected is still present 2 years later. We find the best description of neutral clines around the contact area for 15 microsatellite loci. We use the spatio-temporal evolution of these clines to estimate dispersal in *M. fijiensis*. A maximum likelihood framework allows measuring estimation quality and the validity of the pure-migration hypothesis. Finally, using theoretical simulations, we assess the robustness of the analysis to deviations from the diffusion approximation. We discuss the validity of our results and the portability of the method to other species.

MATERIALS AND METHODS

Study area and sampling

Mycosphaerella fijiensis samples were collected in accessible areas along the roadside verges of a South-West Cameroonian agricultural landscape. At each sampling site, infected leaf fragments were collected from a few neighbour banana plants, and GPS coordinates were recorded. Two samplings were realised in November 2005 (T_1, included in Rieux et al. (2011)) and February 2007 (T_2) (Fig. 1, Appendix S1). We estimated that 15 pathogen generations of sexual reproduction have occurred between the two samplings (Jones 2000). Three to eight individuals were isolated per sampling site. At T_1, 21 sites were sampled along a 32-km long transect. Mycelium cultures were initiated by single ascospores isolated from necrotic lesions grown on solid medium (Halkett et al. 2010). At T_2, 44 sites were sampled along a 45-km long transect and thanks to new technical developments, the laborious pathogen isolation step was avoided by collecting young lesions directly from banana leaves (Robert et al. 2010).

Genotyping

Total genomic DNA was either extracted from mycelial cultures (T_1) or from young lesions (T_2). Genotyping was carried out using 15 microsatellites loci (Neu et al. 1999; Zapater et al. 2008; Robert et al. 2010). Protocols and polymerase chain reaction conditions are detailed in Robert et al. (2010). Among the 15 markers, two are trinucleotide and four are di-nucleotide microsatellite loci already used in the 2011 study. The nine others are tetra-nucleotide markers developed between the two studies (Appendix S3). Amplified fragments were separated on a 16-capillary Sequencer (ABI Prism
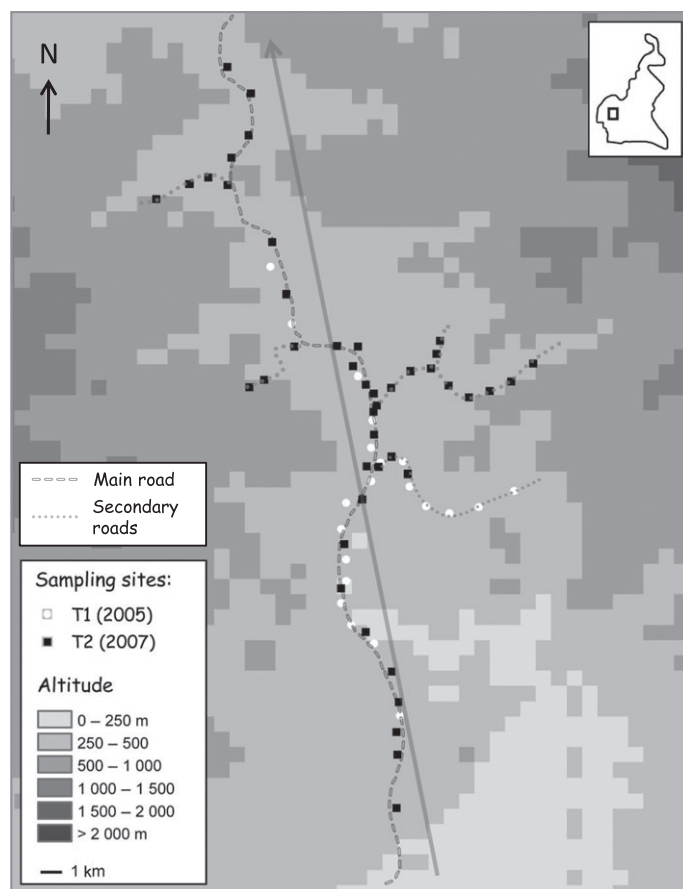

Figure 1 Sample site locations along the transect in South-West Cameroon. The arrow indicates the best-fit axis θ as inferred using the neutral cline approach.

3130XL (Applied Biosystems, Foster City, CA, USA) and analysed for length variation using GeneMapper® Software. In total, 144 individuals (T_1) and 244 individuals (T_2) were genotyped.

Delineating the genetic discontinuity using spatial genetic clustering

Spatial population genetic structure was assessed using the clustering algorithms implemented in GENELAND v3.1.4 (Guillot *et al.* 2005) and TESS v2.3.1 (Chen *et al.* 2007). Both algorithms distribute individual genotypes into K clusters by minimising Hardy–Weinberg disequilibrium and gametic phase disequilibrium within groups, taking into account spatial coordinates. In GENELAND, we used the independent model of allele frequencies. We ran TESS using an admixture model defined according to the BYM model. In both analyses, we performed five independent runs with 10^6 iterations (including 10% of burn-in period) allowing K to vary from 1 to 10 in GENELAND and from 2 to 10 in TESS. K was inferred from the modal value of the run highest posterior probability in GENELAND and from the mean value of the run lowest deviance information criterion (DIC) in TESS. For the best K, posterior individual membership probability to each genetic group or parental population was averaged across best runs' outputs. We used the mba.surf function of R software (R Core Team 2012) to map the distribution of genetic groups over each pixel of the spatial domain considered.

Population structure

On the entire dataset and for both dates, gene diversity was assessed using expected heterozygosity (H_E) and the number of alleles (N_A) using FSTAT2.9.3 (Goudet 1995). We used the nonparametric Wilcoxon–Mann–Whitney U-test to detect differences in heterozygosity between dates. Linkage disequilibrium between all pairs of loci was tested using Fisher's exact tests implemented in GENEPOP (Rousset 2008). The false discovery rate (FDR) procedure implemented in the R package QVALUE (Storey 2002) allowed controlling for multiple testing. Genetic differentiation between clusters was estimated through F_{ST} using GENEPOP. Likelihood ratio G test allowed estimating the significance of F_{ST} values.

Description of neutral genetic clines

Cline fitting tool

To fit genetic clines, we used the software CFit7 (Gay *et al.* 2008, available at http://www.cefe.cnrs.fr/genetique-et-ecologie-evolutive/cfit), which we modified to account for 2D data. It allows fitting bi-allelic allele count data to geographical distance using various cline shapes. Several clines can be fitted to compare their slopes (concordance) and position (coincidence) by constraining shape parameters. Estimation is performed by maximum likelihood. Likelihood is computed considering that allelic counts in a site are drawn from a binomial distribution. It is maximised using a simulated annealing algorithm. Here, most genetic markers had only two alleles (Appendix S3). The others were converted to two-allele systems by assigning each allele to a cluster-specific compound allele according to its coordinates on the first axis of a multiple correspondence analysis (MCA) (Daguin *et al.* 2001). Because sampling sites were not strictly aligned, we used 2-dimensional clines.

Cline description

We first determined how to best describe the data by fitting three different 2D cline shapes to multilocus genetic data independently for T_1 and T_2 without any constraint on parameters. To do so, spatial coordinates (X, Y) were transformed into a single coordinate along the cline axis using:

$$x = X \cos(\theta) + Y \sin(\theta), \tag{1}$$

where θ is the angle between cline axis and the parallel (East–West line). We then fitted three functions of x to allelic count data: two Logit shapes consistent with the neutral hypothesis and a stepped cline expected in tension zones involving strong indirect selection across the genome (details and full equations in Appendix S4). Second, using the best shape, for each locus, we first determined whether a cline was present at T_1 (non-zero slope) using likelihood ratio tests (LRT). Third, for each locus presenting a significant cline pattern at T_1, we tested for a slope change between T_1 and T_2 using LRT again. Model choice was made at the 5% significance level using a chi-squared test at d.f. difference in freedom degrees between the compared models.

Dispersal estimation

Cline width (w) is the ratio between the maximal difference in allele frequencies along the cline and maximal cline slope:

$$w = \frac{|p(+\infty) - p(-\infty)|}{p'(\hat{c})}, \tag{2}$$

where $p(.)$ is allele frequency, $p'(.)$ its derivative with respect to geographical distance and $\hat{c}$ the estimated cline centre. Dispersal estimation was based on the multilocus cline width rate of change. Following a secondary contact, assuming no difference in density between the two groups, negligible drift and diffusion approximation for dispersal, cline width should decay at a rate only determined by the standard deviation of parent–offspring axial dispersal distances (σ) and the time since contact (T) (Endler 1977; Gay *et al.* 2008):

$$w = \sigma\sqrt{2\pi T}. \tag{3}$$

Cline width at T_2 (w_2) thus depends on cline width at T_1 (w_1) and dispersal (σ):

$$w_2 = \sqrt{w_1^2 + \sigma^2 2\pi(T_2 - T_1)}. \tag{4}$$

Because the contact zone may move with time, for example under asymmetric gene flow (Barton & Hewitt 1985), cline centres were also allowed to differ between T_1 and T_2 by a distance shift δ. As detailed in Appendix S6, we reparameterised eqn 4 using the scaled Logit shape and integrated this equation in a new version of CFit7 (available online) thus allowing a maximum likelihood estimation of dispersal (jointly with cline angle, cline centre and centre shift) from the full dataset.

Loci without any cline pattern at T_1 were included in the likelihood computation assuming a constant allele frequency over all sites at T_1. Likelihood was corrected for overdispersion of genetic data using the variance inflation factor (Lebreton *et al.* 1992),

estimated from the residual deviance of the complete model divided by its residual degrees of freedom. Model selection was done using Akaike criterion. Support limits for σ and θ were estimated within 2 units around the maximum likelihood value.

Influence of kernel on dispersal estimation

We investigated whether dispersal kernels deviating from diffusion could bias the estimation of σ. We simulated the evolution of genetic clines under different dispersal distributions with varying kurtosis and conducted the estimation procedure as above. We considered a circular one-dimensional space divided in n sites connected by migration. Migration was modelled with five mixtures of Binomial ($\Delta_0 \to \Delta_4$), as used in previous theoretical studies about kernel-mediated biases in cline approaches, and 11 distributions ($\Delta_5 \to \Delta_{16}$) belonging to three different kernel families commonly used to model dispersal kernels: Exponential power, 2Dt and Geometric (full equations in Appendix S9). For each kernel family, parameter values were chosen so that all distributions have the same variance (σ^2), corresponding to the value inferred from the real data, but different kurtosis (γ ranging from -1 to ~ 100, Fig. 4a). We then simulated the evolution of allele frequency at one bi-allelic neutral locus following a secondary contact under each of these dispersal kernels over 200 generations (Fig. 4b). Producing the next generation amounted to convolve the list of allelic frequencies at generation t with migration kernel so that at each site i, allele frequency at generation $t + 1$ is given by:

$$p_i(t+1) = \sum_k m_{i-k} p_k(t), \tag{5}$$

where m is the migration kernel. Dispersal was then estimated for all dispersal kernels and for each of them by varying first sampling time (T_1) from 1 to 185 generations after secondary contact. All simulations were performed with Mathematica 7.0 (Wolfram Research, Inc., Champaign, IL, USA).

RESULTS

Delineating the genetic discontinuity

In both datasets (T_1 and T_2), GENELAND and TESS clearly detected two genetic groups (Appendix S2). In T_1, applying an assignment threshold of 0.80, GENELAND assigned 86 individuals to the first cluster and 56 to the second one. Only two individuals were not assigned. Applying the same assignment threshold in T_2, only seven individuals were not assigned as 158 individuals were assigned to another first cluster and 79 individuals to a second one. As expected under the admixture model, differences in membership profiles were found with TESS (Appendix S2). In T_1, TESS assigned 30 and 70 individuals to the first and second parental populations, respectively, the other 44 individuals, that is $\sim 30\%$ of the population, being admixed. In T_2, the proportion of admixed individuals increased to $\sim 40\%$ (i.e. 100 individuals). The remaining 93 and 51 individuals were assigned to two other parental populations.

The existence of a break in allelic frequencies was more conspicuous from GENELAND map but both algorithms exhibited consistent spatial genetic structures. Finally, the multilocus genetic discontinuity was located at approximately the same latitude ($\sim 4.90°$, Fig. 2) in both datasets.

Indices of genetic diversity for T_1 and T_2 samples are given in Appendix S3. Expected heterozygosity did not significantly vary between the two samplings (U-test $P = 0.815$). No pair of loci showed significant linkage disequilibrium after FDR control (Q-value < 0.05). Genetic differentiation (F_{ST}) between the two clusters detected in GENELAND was significant in both samplings ($F_{ST} = 0.091$ at T_1 and $F_{ST} = 0.07$ at T_2) and greatly variable among microsatellite loci (Appendix S3).

Cline description

Scaled Logit shape best fitted the data (Appendix S4). This is probably because for most loci, alleles are not fixed within parental populations (Fig. 3) Eight loci (of 15) presented a significant cline pattern (i.e. non-zero slope) (Table 1). Among these loci, cline widths at T_1 ranged between 3 km and 39 km (Table 1), cline centres were coincident and cline widths always increased significantly between T_1 and T_2 (Appendix S5). Hence, these clines tended to vanish, as expected under the homogenising effect of gene flow. The seven remaining loci showed no cline pattern (zero-slopes) at both T_1 and T_2 (Appendix S5).
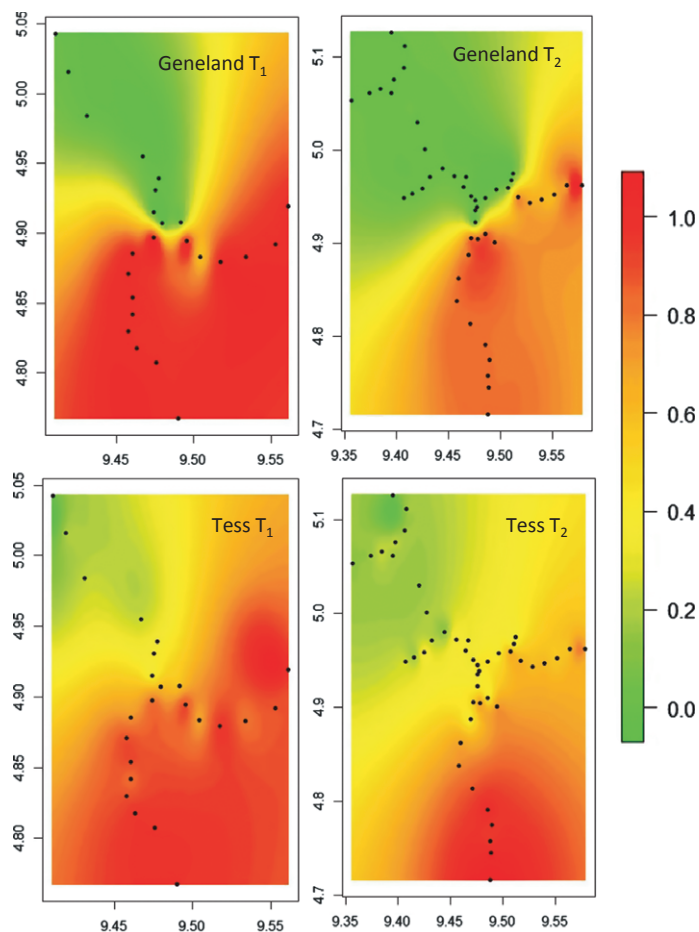

Figure 2 Spatial genetic structure of *Mycosphaerella fijiensis* populations over the study area. The maps either indicate the posterior probability of belonging to cluster 1 (GENELAND – upper panels) or to the first parental population (TESS – lower panels) at both T_1 and T_2 sampling dates. Black dots represent sampled sites and x- and y-axes correspond to geographical coordinates.

Dispersal estimation

First, cline angle (θ = 1.87 radians $\approx$ 107°, support limit interval [1.64–1.98 radians]) was not significantly different between the two samplings (Table 2) and was congruent with the disposition of most sampling sites along the main road (Fig. 1). Second and importantly, the model with a single value of σ common to all loci was significantly better than the model with locus-specific values of σ (Table 2). The best model (model G) considers that centres are coincident between loci at T_1 and that no cline movement has occurred (δ = 0) (Table 2). From the best model, σ was estimated at 1175 metres/generation$^{1/2}$ with a support limit interval of [690 m–1586 m] (Appendix S7).

Influence of kernel on dispersal estimation

Figure 4c shows the effects of the family and kurtosis of underlying dispersal kernels as well as time of first sampling on σ estimation. All estimates felt within the support limits obtained on the real dataset, indicating that the departure from diffusion approximation causes biases well below measurement error in our case. Only the kernel Δ_4, a mixture of Binomial distributions with extremely high kurtosis, combined with a very early sampling date (T_1 < 8 generations) could generate a detectable bias (underestimation of σ). Under these conditions, clines showed non-logit shapes with discontinuities in allele frequencies (Fig. 4b). Interestingly, an exponential power distribution with the same kurtosis value induced smaller biases (kernel Δ_{11}). This suggests that moments of order higher than four (distribution tail) might be at the origin of the observed deviation with Δ_4.

DISCUSSION

This study represents the first use of the decay of neutral clines to quantitatively infer dispersal. This approach basically requires genotyping individuals with neutral markers in a situation of secondary contact between differentiated populations in which drift can be considered as negligible in front of migration. It provided fairly accurate measure of contemporary gene flow and was robust to deviations from diffusion approximation.

Estimation of dispersal in *Mycosphaerella fijiensis*

We provided the first estimate of dispersal in the fungal banana pathogen *M. fijiensis*. It is also among the first quantitative genetic-based estimates of dispersal at a landscape scale in fungal plant pathogens, a group that comprises most emerging infectious plant diseases (Fisher *et al.* 2012), and which ecology is strikingly poorly known. Up to now, population genetics studies have seldom allowed estimating dispersal because pathogen populations are frequently out of genetic equilibrium and/or suspected to have too high effective sizes (Giraud *et al.* 2008; Rieux *et al.* 2011). Most dispersal estimates have therefore relied on spore traps or field epidemiological studies often at very small geographical scales (McCartney *et al.* 2006). Here, we estimate σ, a synthetic diffusion parameter in which gene flow is averaged over 15 generations, which is also equivalent to the standard deviation of parent–offspring axial dispersal distances. Its value (σ = 1175 m/generation$^{1/2}$) suggests very high dispersal abilities in *M. fijiensis*. Despite the absence of previous dispersal estimates in this species, this estimate

Table 1 Testing for an initial cline pattern in 15 microsatellites

MfSSR	407	425	405	322	324	413	434	340	401	430	362	417	428	203	350
LL$_{T1}$ (k = 5)	−84.42	−46.74	−87.17	−83.71	−74.65	−85.89	−83.46	−82.19	−38.45	−16.41	−16.62	−88.39	−91.55	−75.26	−36.14
θ*	1.871	1.878	1.869	1.870	1.870	1.875	1.881	1.873	x	x	x	x	x	x	x
c (km)*	20.06	20.08	19.87	21.07	20.09	21.02	19.94	19.99	x	x	x	x	x	x	x
b*	68.33	140.19	11.20	25.73	17.25	35.72	38.97	22.92	0.10	0.03	0.07	0.09	0.12	0.02	0.06
h_1*	0.13	0.62	0.00	0.08	0.02	0.16	0.17	0.00	0.67	0.99	0.96	0.45	0.46	0.12	0.85
l_p*	0.59	0.94	0.95	0.64	0.98	0.69	0.78	0.70	0.01	1.00	1.00	0.05	0.00	0.10	0.04
b_2†	0.64	0.98	0.95	0.67	0.98	0.74	0.82	0.70	0.67	1.00	1.00	0.47	0.46	0.21	0.85
Slope†	8.79	12.62	2.67	3.77	4.16	5.24	6.29	3.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Width (km)†	6.44	3.14	39.30	17.10	25.51	12.32	11.29	19.20	x	x	x	x	x	x	x
LL$_0$ Worse model (k = 1)	−95.54	−59.02	−93.36	−92.43	−84.69	−94.48	−96.29	−91.61	−38.46	−18.13	−18.13	−88.95	−93.19	−76.76	−37.14
Dev: −2LL (LL$_0$−LL$_{T1}$)	**22.24**	**24.55**	**12.40**	**17.44**	**20.07**	**17.17**	**25.65**	**18.84**	0.02	3.46	3.04	1.11	3.29	3.01	2.00

Clines were fitted using scaled Logit function (Equation in Appendix S6).
*Denotes inferred parameters: cline centre (c), cline slope-related parameter (b), lower asymptotic frequency (b_1), allelic frequency step between the two populations (b_p) and cline angle (θ).
†Holds for a *posteriori* computed parameters. These are b_2, the higher asymptotic frequency ($b_1 + (1 − b_1)b_p$), cline slopes ($−b(1 − b_1)b_p$) and for non-flat clines, cline widths ($4/b$, in km). LL is log likelihood. For each locus, slope significance was tested using likelihood ratio tests against a null model assuming homogeneous allele frequency over the study area. Deviance (Dev) is given with significant values (5% level) highlighted in bold × indicates situations in which parameters could not be computed (see text for details).

is consistent with the biology of the fungus. *Mycosphaerella fijiensis* exhibits two natural modes of dispersal. The first relies on asexual spores (conidia), suspected to disperse at very short distances through rain splash (Gauhl 1994). The second is ensured by light sexual spores (ascospores) actively projected in the air (Gauhl 1994) and wind dispersed (Burt 2003). By analogy to other airborne fungi, *M. fijiensis* ascospores have been suspected to occasionally travel over several kilometres (Burt 2003; Amil *et al.* 2007). Finally, in addition to natural spore dispersal, the disease also spreads through human-mediated transport of infected plant material (Carlier 2000). This propagation mode may occur in the studied valley, as trades of banana plants between villages are common local agricultural practices. So important dispersal combined with high population sizes are also consistent with the recent non-detection of any Isolation By Distance (IBD) pattern along a 33-Km long transect (Rieux *et al.* 2011). The fact that no movement of cline centres was detected suggests isotropic dispersal over the study area (Barton & Hewitt 1985). Wind data are unavailable, but this result suggests either isotropic wind influences and/or isotropic movement of infected plants between villages surrounding cline centre over the last 2 years.
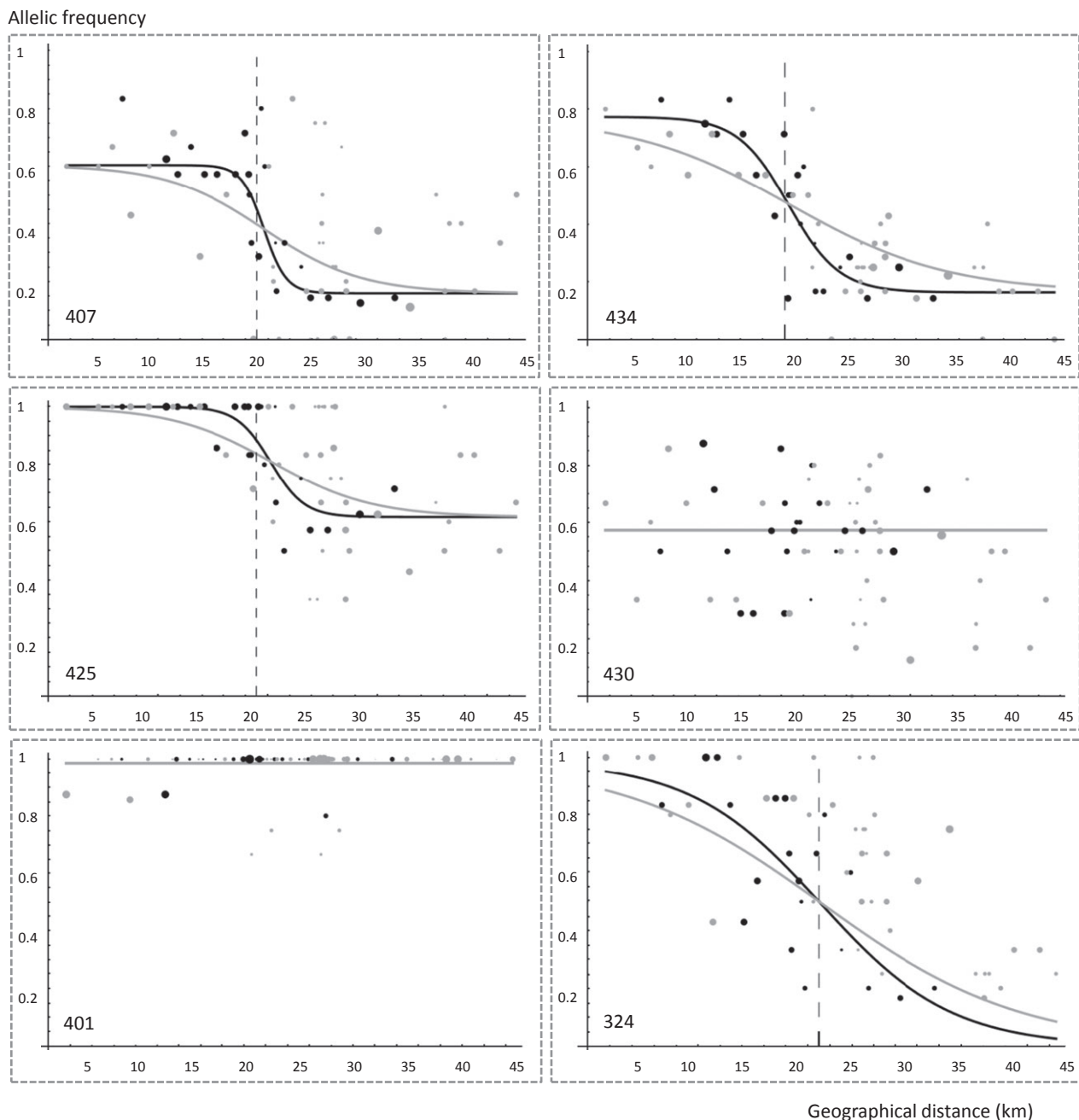

Figure 3 Change in neutral genetic clines for six of the 15 microsatellites considered between the two sampling dates. From top to bottom and left to right, the loci represented are MfSSR-407, 434, 425, 430, 401 and 324. Allelic frequencies at T_1 (black) and T_2 (grey) are plotted against distance along the main axis of the cline. Distances are given in km, 0 being the southernmost site. Dots represent sampling sites, dot size being proportional to the number of individuals sampled. Curves show the fitted clines (scaled Logit shapes), with their centres highlighted through dashed vertical lines.

Table 2 Testing cline coincidence, similarity in cline vanishing rates between loci and isotropy. **A**. Compared models. Parameters constrained and type of constraint operated are highlighted in bold. σ is the standard deviation of parent–offspring axial dispersal distances, b a cline slope parameter, h_1 lower asymptotic frequency, h_p the step in allelic frequency between the two populations, θ cline angle, c cline centre, and δ centre shift between T_1 and T_2. **B**. Model comparison on the whole dataset (both samplings and 15 loci). LL is log likelihood and k the number of parameters. Q_{AIC} corrects Akaike Information Criterion (AIC) for overdispersed frequency data and the best model is highlighted in bold

A

Model	Parameters constrained
A	$\theta T_1 \neq \theta T_2/c, b, h_1, h_p, \delta$ & $\sigma \neq$between locus
B	$\boldsymbol{\theta T_1 = \theta T_2}/c, b, h_1, h_p, \delta$ & $\sigma \neq$between locus
C	$\boldsymbol{\theta T_1 = \theta T_2}/c, b, h_1, h_p$ & $\delta \neq$between locus$/\boldsymbol{\sigma = 0}$
D	$\boldsymbol{\theta T_1 = \theta T_2}/c, b, h_1, h_p$ & $\delta \neq$between locus$/\mathbf{s}$ **same** σ **for each locus**
E	$\boldsymbol{\theta T_1 = \theta T_2}/b, h_1, h_p$ & $\delta \neq$between locus$/$**same** σ **&** $\mathbf{c}$ **for each locus**
F	$\boldsymbol{\theta T_1 = \theta T_2}/b, h_1$ & $h_p \neq$between locus$/$**same** σ, $\mathbf{c}$ **&** δ **for each locus**
G	$\boldsymbol{\theta T_1 = \theta T_2}/b, h_1$ & $h_p \neq$between locus$/$**same** σ **&** $\mathbf{c}$ **for each locus**$/\boldsymbol{\delta = 0}$
H	$\boldsymbol{\theta T_1 = \theta T_2}/c, b, h_1$ & $h_p \neq$between locus$/$**same** σ **&** δ **for each locus**
I	$\boldsymbol{\theta T_1 = \theta T_2}/c, b, h_1$ & $h_p \neq$between locus$/$**same** σ **for each locus**$/\boldsymbol{\delta = 0}$

B

Model	LL	k	AIC	Q_{AIC}	AIC weight
G	**−2637.63**	**34**	**5343.26**	**3731.72**	**0.88**
F	−2638.68	35	5347.35	3735.17	0.11
I	−2636.09	41	5354.18	3743.58	0.00
D	−2635.17	49	5368.33	3758.30	0.00
C	−2638.93	48	5373.86	3761.53	0.00
B	−2627.74	63	5381.48	3775.98	0.00
E	−2649.25	42	5382.51	3763.87	0.00
A	−2627.29	64	5382.57	3777.35	0.00
H	−2652.63	42	5389.26	3768.56	0.00

Validity of the 'pure-migration' assumption

The present analysis relies on the existence of clines vanishing under the action of migration alone (i.e. without any selection or genetic drift). Both cline theory and empirical data have shown that linkage disequilibrium can build up easily in clines maintained by selection (Barton 1983). When this happens, any neutral locus gets more or less linked to loci under selection. Studying (neutral) microsatellite markers therefore does not guarantee that observed clines are not indirectly affected by selection or drift. Here, we had *a priori* reasons to argue that the observed genetic clines are not influenced by selection and drift. First, recent spore countings in Cameroonian populations revealed huge census population sizes probably high enough to avoid genetic drift (Rieux *et al.* 2012). Second, the sampled area is relatively small and characterised by an apparent absence of environmental or exogenous selective gradient (only one host species and no fungicide pressure). Third, contrary to hybrid zones in which distinct diploid races or subspecies meet, all sampled individuals are haploid and belong to the same species, which should limit the occurrence of endogenous selection against hybrids. Beyond *a priori* clues, one major advantage of this method is to provide tools to test for departures from the pure-migration hypothesis through a maximum likelihood framework. Here, the analysis did not reveal any impact of indirect selection. Clines were sigmoid rather than stepped as expected under strong indirect selection. The eight clines identified at T_1 were all shown to significantly

fade 15 generations later, suggesting that the genetic discontinuity is not at migration–selection equilibrium. Importantly, no heterogeneity in the rate of increase in cline width was detected among loci. If physically linked to a selected locus, neutral loci should see their cline width increase at a rate determined by the strength of selection against introgression at the selected locus (Barton & Gale 1993). Selection is therefore expected to result in contrasted rates of cline width increase at different microsatellites, which was not observed. Finally, the seven loci that showed no initial cline pattern behaved consistently with the pure-migration scenario, as no discordant geographical pattern appeared 15 generations later. Drift also appeared negligible here. Cline fits were of good quality as attested by low variance inflation factors (Table 2). Besides, if influenced by genetic drift, neutral loci should see the variance in allele frequency increase at random rates, leading again to heterogeneity in rates of cline width increase among loci, which was not observed.

Robustness of the method

As most previous cline studies, the present approach relies on the underlying assumption that dispersal is a pure diffusion process (but see Saccheri *et al.* 2008 for an exception). Diffusion approximately corresponds with a Gaussian dispersal kernel. However, for many organisms, and especially wind-dispersed species such as *M. fijiensis*, dispersal has been shown to follow a leptokurtic kernel, that is with an excess of small- and long-distance dispersal in comparison to the Gaussian distribution (e.g. Sackett & Mundt 2005). Two previous studies have investigated the effects of varying kurtosis on clines at migration–selection equilibrium (Lenormand 1998; Rousset 2001). In both cases, it has been shown that large kurtoses may significantly increase cline maximum slope in comparison to a Gaussian dispersal distribution, leading to an underestimation of σ. As the present approach relies on a temporal comparison of cline slopes, it was not clear whether leptokurtic dispersal distributions would deteriorate dispersal estimation. Our simulations showed that the method is actually robust to leptokurtic kernels, as most estimated values remained within the study support limits. Results were very congruent among kernel families, showing that other aspects of dispersal distributions (e.g. tail) have low impact on the estimation. The only noticeable deviations were detected for Binomial mixture distributions, with exceptionally strong kurtosis ($\gamma > 100$) and only for early sampling dates. Such high kurtoses have never been reported from experimental studies (Appendix S8) and were only studied for illustration purposes. In this situation, clines present discontinuities in allele frequencies due to an excess of long-distance dispersal and a lack of short-distance mixing. Clines are then poorly fitted using smooth shapes and Endler's formula should not apply. Should the case arise with real data, such signature could be readily detected from the allele frequency pattern.

Desirable species, markers and sampling

The method is formulated in a very general manner potentially applicable to a variety of organisms. One important prerequisite is that effective sizes be high enough for drift to be negligible in front of migration in studied populations. This clearly disqualifies the method for species with very low effective sizes. For these, one may count on other methods, such as mark-recapture methods, to measure dispersal. Our approach may nevertheless represent a pow-
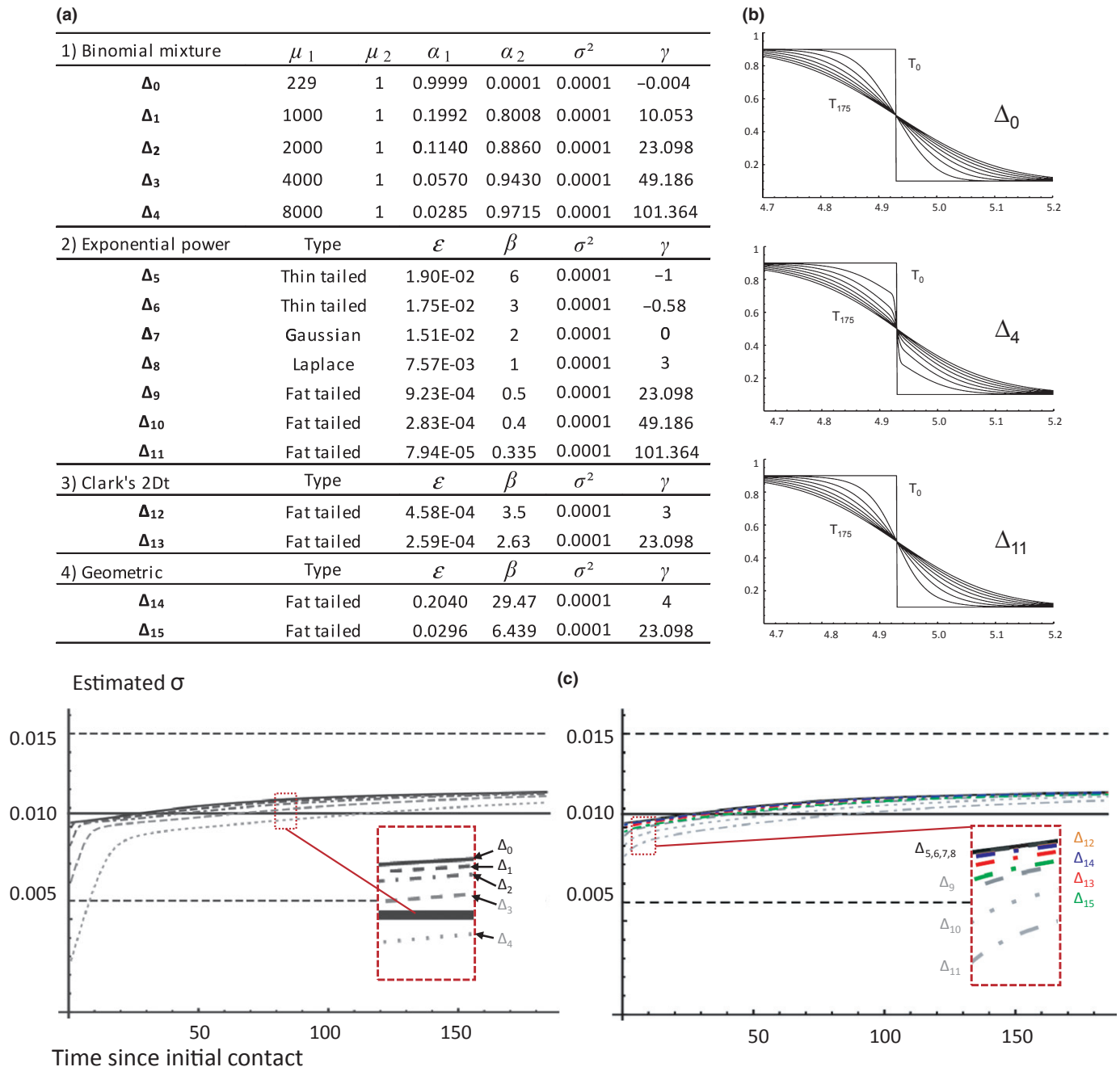
(a)

1) Binomial mixture	μ_1	μ_2	α_1	α_2	σ^2	γ
Δ_0	229	1	0.9999	0.0001	0.0001	−0.004
Δ_1	1000	1	0.1992	0.8008	0.0001	10.053
Δ_2	2000	1	0.1140	0.8860	0.0001	23.098
Δ_3	4000	1	0.0570	0.9430	0.0001	49.186
Δ_4	8000	1	0.0285	0.9715	0.0001	101.364
2) Exponential power	Type	ε	β	σ^2	γ	
Δ_5	Thin tailed	1.90E-02	6	0.0001	−1	
Δ_6	Thin tailed	1.75E-02	3	0.0001	−0.58	
Δ_7	Gaussian	1.51E-02	2	0.0001	0	
Δ_8	Laplace	7.57E-03	1	0.0001	3	
Δ_9	Fat tailed	9.23E-04	0.5	0.0001	23.098	
Δ_{10}	Fat tailed	2.83E-04	0.4	0.0001	49.186	
Δ_{11}	Fat tailed	7.94E-05	0.335	0.0001	101.364	
3) Clark's 2Dt	Type	ε	β	σ^2	γ	
Δ_{12}	Fat tailed	4.58E-04	3.5	0.0001	3	
Δ_{13}	Fat tailed	2.59E-04	2.63	0.0001	23.098	
4) Geometric	Type	ε	β	σ^2	γ	
Δ_{14}	Fat tailed	0.2040	29.47	0.0001	4	
Δ_{15}	Fat tailed	0.0296	6.439	0.0001	23.098	

(b)

(c)

Figure 4 Influence of kernel on σ estimation. (a) Construction of dispersal distributions Δ_i using four different kernel families displaying identical variance (σ^2, in degree^2/generation) but different kurtosis (γ). μ, α, ε and β are kernel parameters (Appendix S9 for equations). (b) Simulated clines obtained from Δ_0, Δ_4, Δ_{11}. Clines are shown every 25 generations from T = 0 to 175 generations after initial contact. (c) Temporal dynamics of σ estimate from simulated data sets. Left panel: mixture of Binomial distributions. Right panel: other dispersal kernels. Horizontal black solid line indicates the expected value (σ = 1175 m/generation$^{1/2}$), and horizontal dashed lines its associated support limits.

erful alternative for various species characterised by moderate to high effective sizes. Importantly, several of its key assumptions (including drift effect) can be internally checked, so that the method should overall provide robust estimates.

Using bi-allelic markers is not a prerequisite, as robust approaches exist to convert multi-allelic markers into a bi-allelic system (see for instance Daguin *et al.* (2001) and references therein). Nevertheless, genetic markers displaying a small number of alleles should be preferred to avoid introducing additional variance in cline fits. We

would thus recommend considering long tandem repeats markers or ideally single-nucleotide polymorphism markers (SNPs).

One other critical aspect of the method is sampling design. First, identifying a local contact zone clear enough so that individuals can be assigned without ambiguity to differentiated clusters is not required, as illustrated here with TESS results. Importantly, sampling should (1) encompass individuals at both extremities of the contact zone so that cline plateaux can be properly estimated and (2) be intensified near the contact zone for cline centre and slope to be accurately

inferred. If more than two populations meet, similar sampling advices apply as the method would allow taking advantage of such design by simultaneously fitting clines from several locations. Second, if dispersal truly occurs in 2D, sampling should also be performed in 2D, for cline main axis to be correctly inferred. Such precaution is necessary because if sampling is designed as a 1D transect along an inappropriate direction that forms an angle θ with cline main axis, inferred cline slopes will be multiplied by a factor $Cos(\theta)$, resulting in overestimating σ by a factor $1/|Cos(\theta)|$. Here, secondary contact occurred in a shallow valley, where pathogen host plants are continuously distributed along roadsides. The valley is not a straight line and secondary axes exist (Fig. 1), so that we considered unsafe to making an *a priori* guess on cline angle and used a 2D approach. Cline angle was weak relative to the direction of the main road with a narrow support limit, which confirms that dispersal actually mainly occurs along roadsides. In this specific situation, a 1D approach would have resulted in a relatively minor error on dispersal estimate.

Back to the historical contact time

In principle, from Endler's model, one may estimate the product $\sigma\sqrt{T}$, where T is the time since historical contact in generations (eqn 3). Using two samplings, we ideally would have been in a position to jointly infer σ and T. We would have written the likelihood of T_1 clines as a function of both parameters. This was not possible here for one reason. Under pure migration, clines formed subsequently to a secondary contact should have the same width across loci. This was not true in our data as cline widths at T_1 differed. This variability in initial cline widths suggests a major role of drift during the initial stages of cline formation and is consistent with the historical expansion of *M. fijiensis* in Cameroon. This species is self-incompatible: the production of sexual spores requires high enough population size for two sexually compatible lesions to germinate, grow and collapse on the same leaf. This Allee effect may have constrained initial population spread to rely on asexual spores, possibly leading to an increased intensity of drift. Such limitation may not hold for other species, for which the method would allow estimating contact time.

CONCLUSION

Our results show that combining genetic clustering approaches to detect breaks in allelic frequencies with neutral genetic cline analyses offers a convenient way to gain solid information on the rate of contemporary dispersal. It combines the desirable properties of both direct methods (e.g. estimation over a known and recent period) and classical indirect approaches (e.g. measure of effective dispersal). With the continuous development of both new sequencing technologies and spatial genetic clustering algorithms (Guillot *et al.* 2009), we expect that opportunities to detect such fine-scale spatial genetic discontinuities will multiply in the future. The neutral cline approach may thus become a useful and cost-effective method to infer dispersal, particularly in emerging or invasive species. By applying it to the fungal pathogen, *M. fijiensis*, we have provided evidence for the high dispersal abilities of this species. Obtaining landscape-scale estimates of contemporary dispersal in a wider range of species is now crucial for ecological and evolutionary principles to guide the design of quarantine and management policies (e.g. rationalisation of pesticide use, deployment of resistant varieties).

AUTHORSHIP

AR, VR, JC and LLB designed the study. LLB was involved in data collection. AR conducted strain isolation and molecular work. TL provided the code of the cline fitting tool. VR and TL modified the code. AR performed all genetic and statistical analyses under the supervision of VR and TL. VR and AR performed the simulations. AR and VR wrote the first draft of the manuscript, and all authors substantially contributed to revisions.

REFERENCES

Amil, A.F., Heaney, S.P., Stanger, C. & Shaw, M.W. (2007). Dynamics of QoI sensitivity in *Mycosphaerella fijiensis* in Costa Rica during 2000 to 2003. *Phytopathology*, 97, 1451–1457.

Barton, N.H. (1982). The structure of the hybrid zone in Uroderma-bilobatum (Chiroptera, Phyllostomatidae). *Evolution*, 36, 863–866.

Barton, N.H. (1983). Multilocus clines. *Evolution*, 37, 454–471.

Barton, N.H. (2001) The evolutionary consequences of gene flow and local adaptation: future approaches. In: *Dispersion: Individual, Population and Community* (eds Clobert, J., Danchin, E., Dhondt, A.A. & Nichols, J.D.). Oxford University Press, Oxford, pp. 329–340.

Barton, N.H. & Gale, K.S. (1993). Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process*. (ed. Harrison, R.G.). Oxford University Press, Oxford, pp. 13–45.

Barton, N.H. & Hewitt, G. (1985). Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.*, 16, 113–148.

Broquet, T. & Petit, E.J. (2009). Molecular estimation of dispersal for ecology and population genetics. *Annu. Rev. Ecol. Evol. Syst.*, 40, 193–216.

Brown, J.K.M. & Hovmøller, M.S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science*, 297, 537–541.

Burt, P.J.A. (2003). Airborne dispersal of *Mycosphaerella fijiensis*. In: *Mycosphaerella Leaf Spot Diseases of Bananas: Present Status and Outlook* (ed. Jacome, L., Lepoivre, P., Marin, D., Ortiz, R., Romero, R., Escalant, J.V.) Proceedings of the Workshop on Mycosphaerella Leaf Spot Diseases, San José, Costa Rica, pp. 111–121.

Carlier, J. (2000). Population genetic structure and dispersal of *Mycosphaerella fijiensis*. *Info. musa*, 13, 17–20.

Chen, C., Durand, E., Forbes, F. & François, O. (2007). Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Mol. Ecol. Notes*, 7, 747–756.

Daguin, C., Bonhomme, F. & Borsa, P. (2001). The zone of sympatry and hybridization of *Mytilus edulis* and *M-galloprovincialis*, as described by intron length polymorphism at locus mac-1. *Heredity*, 86, 342–354.

Dieckmann, U., O'Hara, B. & Weisser, W. (1999). The evolutionary ecology of dispersal. *Trends Ecol. Evol.*, 14, 88–90.

Endler, J.A. (1977). Geographic variation, speciation, and clines. *Monogr. Popul. Biol.*, 10, 1–246.

Fisher, M.C., Henk, D.A., Briggs, C. J., Brownstein, J.S., Madoff, L.C., McCraw, S.L. *et al.* (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature*, 484, 186–194.

Gauhl, F. (1994). *Epidemiology and ecology of black sigatoka*. PhD Thesis, INIBAP edn. INIBAP, Montpellier.

Gay, L., Crochet, P.A., Bell, D.A. & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution*, 62, 2789–2806.

Giraud, T., Enjalbert, J., Fournier, E., Delmotte, F. & Dutech, C. (2008). Population genetics of fungal diseases of plants parasite. *Parasite*, 15, 449–454.

Goudet, J. (1995). FSTAT (Version 1.2): a computer program to calculate F-statistics. *J. Hered.*, 86, 485–486.

Guillot, G., Estoup, A., Mortier, F. & Cosson, J.F. (2005). A spatial statistical model for landscape genetics. *Genetics*, 170, 1261–1280.

Guillot, G., Leblois, R., Coulon, A. & Frantz, A.C. (2009). Statistical methods in spatial genetics. *Mol. Ecol.*, 18, 4734–4756.

Halkett, F., Coste, D., Rivas Platero, G.G., Zapater, M.F.Z., Abadie, C. & Carlier, J. (2010). Isolation by distance and genetic discontinuities in recently established populations of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Mol. Ecol.*, 18, 3909–3923.

Jones, D.R. (2000). *Diseases of Banana, Abaca and Enset*. CABI Publishing, Oxon.

de Lapeyre de Bellaire, L., Fouré, E., Abadie, C. & Carlier, J. (2010). Black leaf streak disease is challenging the banana industry. *Fruits*, 65, 327–342.

Lebreton, J.D., Burnham, K.P., Clobert, J. & Anderson, D.R. (1992). Modeling survival and testing biological hypotheses using marked animals – a unified approach with case-studies. *Ecol. Monogr.*, 62, 67–118.

Lenormand, T. (1998). Dynamique de l'adaptation locale: Résistance aux insecticides chez Culex pipiens. PhD thesis, Ecole Nationale Supérieure Agronomique de Montpellier, 64–65.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends Ecol. Evol.*, 17, 183–189.

Lenormand, T. & Raymond, M. (1998). Resistance management: the stable zone strategy. *Proc. R. Soc. Lond. B Biol. Sci.*, 265, 1985–1990.

Lenormand, T. & Raymond, M. (2000). Analysis of clines with variable selection and variable migration. *Am. Nat.*, 155, 70–82.

Lenormand, T., Guillemaud, T., Bourguet, D. & Raymond, M. (1998). Evaluating gene flow using selected markers: a case study. *Genetics*, 149, 1383–1392.

Lockwood, J.L., Cassey, P. & Blackburn, T. (2005). The role of propagule pressure in explaining species invasions. *Trends Ecol. Evol.*, 20, 223–228.

McCartney, H.A. & Fitt, B.D.L. & West, J. (2006). Dispersal of foliar plant pathogens: mechanisms gradients and spatial patterns. In: *The Epidemiology of Plant Diseases* (eds Cooke, B.M., Jones, D.G. & Kaye, B.). Springer, Netherlands, pp. 159–192.

McDonald, B.A. & Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.*, 40, 349.

Mourichon, X. & Fullerton, R.A. (1990). Geographical distribution of the two species *Mycosphaerella musicola* Leach (*Cercospora musae*) and *M. fijiensis* Morelet (*C. fijiensis*), respectively agents of sigatoka disease and black leaf streak disease in bananas and plantains. *Fruits*, 45, 213–218.

Neu, C., Kaemmer, D., Kahl, G., Fischer, D. & Weising, K. (1999). Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis. Mol. Ecol.*, 8, 523–525.

Pennisi, E. (2010). Armed and dangerous, (pg 804). *Science*, 327, 804–805.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available: http://www.R-project.org/.

Rieux, A., Halkett, F., de Lapeyre de Bellaire, L., Zapater, M.F., Rousset, F., Ravigne, V. *et al.* (2011). Inferences on pathogenic fungus population structures from microsatellite data: new insights from spatial genetics approaches. *Mol. Ecol.*, 20, 1661–1674.

Rieux, A., de Lapeyre de Bellaire, L., Zapater, MF., Ravigne, V. & Carlier, J. (2012). Recent range expansion and agricultural landscape heterogeneity have only minimal effect on the spatial genetic structure of the plant pathogenic fungus *Mycosphaerella fijiensis. Heredity*, 19, 29–38.

Robert, S., Rieux, A., Argout, X., Carlier, J. & Zapater, M.F. (2010). Optimized genotyping with microsatellite markers in the fungal banana pathogen *Mycosphaerella fijiensis. Am. J. Bot.*, 97, 130–132.

Ronce, O. (2007). How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annu. Rev. Ecol. Evol. Syst.*, 38, 231–253.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, 145, 1219–1228.

Rousset, F. (2001). Genetic approaches to the estimation of dispersal rates. In: *Dispersal* (eds Clobert, J. Danchin, E. Dhondt, A.A. & Nichols, J.D.). New York: Oxford Univ. Press, pp. 18–28.

Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.*, 8, 103–106.

Saccheri, I.J., Rousset, F., Watts, P.C., Brakefield, P.M. & Cook, L.M. (2008). Selection and gene flow on a diminishing cline of melanic peppered moths. *Proc. Natl Acad. Sci. USA*, 105, 16212–16217.

Sackett, K.E. & Mundt, C.C. (2005). Primary disease gradients of wheat stripe rust in large field plots. *Phytopathology*, 95, 983–991.

Slatkin, M. & Barton, N. (1989). A comparison of three indirect methods for estimating average levels of gene flow. *Evolution*, 43, 1349–1368.

Storey, J.D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Series B Stat. Methodol.*, 64, 479–498.

Turelli, M. & Hoffmann, A.A. (1991). Rapid spread of an inherited incompatibility factor in california drosophila. *Nature*, 353, 440–442.

Zapater, M.F., Duchemin, M., Dussart, J.F., Coste, D., Brottier, P. & Carlier, J. (2008). Microsatellite markers for the fungal banana pathogens *Mycosphaerella fijiensis, Mycosphaerella musicola* and *Mycosphaerella eumusae. Mol. Ecol. Resour.*, 8, 1121–1125.

SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecologyletters.com).