Estimation of the dispersal of a major pest of maize by cline analysis of a temporary contact zone between two invasive outbreaks

Gérald Bermond, Aurelie Blin, Elodie Vercken, Virginie Ravigné, Adrien Rieux, Sophie Mallez, Thibaut Morel-Journel, Thomas Guillemaud

Estimation of the dispersal of a major pest of maize by cline analysis of a temporary contact zone between two invasive outbreaks

GÉRALD BERMOND,*†‡ AURÉLIE BLIN,*†‡ ELODIE VERCKEN,*†‡ VIRGINIE RAVIGNÉ,§
ADRIEN RIEUX,¶ SOPHIE MALLEZ,*†‡ THIBAUT MOREL-JOURNEL*†‡
 and THOMAS GUILLEMAUD*†‡

*INRA, UMR 1355, Sophia Antipolis, F-06903, France, †Université de Nice Sophia Antipolis, UMR Institut Sophia Agrobiotech, Sophia Antipolis, France, ‡CNRS, UMR 7254, Sophia Antipolis, France, §CIRAD, UMR BGPI, Campus International de Baillarguet, Montpellier F-34398, France, ¶University College London, London WC1E 6BT, UK

Abstract

Dispersal is a key factor in invasion and in the persistence and evolution of species. Despite the importance of estimates of dispersal distance, dispersal measurement remains a real methodological challenge. In this study, we characterized dispersal by exploiting a specific case of biological invasion, in which multiple introductions in disconnected areas lead to secondary contact between two differentiated expanding outbreaks. By applying cline theory to this ecological setting, we estimated σ, the standard deviation of the parent offspring distance distribution, of the western corn rootworm, *Diabrotica virgifera virgifera*, one of the most destructive pests of maize. This species is currently invading Europe, and the two largest invasive outbreaks, in northern Italy and Central Europe, have recently formed a secondary contact zone in northern Italy. We identified vanishing clines at 12 microsatellite loci throughout the contact zone. By analysing both the rate of change of cline slope and the spatial variation of linkage disequilibrium at these markers, we obtained two σ estimates of about 20 km/generation$^{1/2}$. Simulations indicated that these estimates were robust to changes in dispersal kernels and differences in population density between the two outbreaks, despite a systematic weak bias. These estimates are consistent with the results of direct methods for measuring dispersal applied to the same species. We conclude that secondary contact resulting from multiple introductions is very useful for the inference of dispersal parameters and should be more widely used in other species.

Keywords: *Diabrotica virgifera*, geneflow estimation, genetic clines, hybrid zone, linkage disequilibrium

Introduction

Biological invasions are of considerable significance in applied ecology because they often have major consequences for biodiversity, agriculture and human health (McKinney & Lockwood 1999; Mack *et al.* 2000; Ruiz *et al.* 2000; Pimentel *et al.* 2001; Lodge & Shrader-Frechette 2003; Olden *et al.* 2004). From a strictly scientific perspective, they may also be very useful for investigations of the various evolutionary processes underlying species distributions and abundances, community associations and relationships between native and invasive species (Sax *et al.* 2005; Lockwood, *et al.* 2007). An understanding of biological invasions also appears to be a fundamental element of the fight to prevent population establishment in the early stages of invasion or to optimize the management of the invading species in the areas invaded.

We need to understand the demographic and evolutionary processes occurring in the invaded area in order to outline the mechanisms underlying biological

Correspondence: Gerald Bermond, Fax: (33) 4 92 38 64 01;
E mail: gerald.bermond@sophia.inra.fr

invasions. The key factor in these processes is dispersal (Johnson & Gaines 1990) – that is, the set of active or passive phenomena resulting in the establishment of a geographical distance between the breeding sites of individuals from two successive generations – defined here as effective dispersal. Dispersal is a major determinant of biological invasions (Lockwood *et al.* 2007), as it leads to the expansion of the geographical range of the introduced species after its establishment outside its native area (Wilson *et al.* 2009) and because it determines the level of gene flow between introduced populations. The estimation of dispersal distances can thus be used to characterize the invasiveness of a species (Rehage & Sih 2004), to predict the expansion speed of invasive populations (Lockwood *et al.* 2007) and to evaluate the balance between migration and selection (Lenormand & Raymond 2000) controlling the distribution range of expanding populations (Kirkpatrick & Barton 1997).

Despite the importance of accurate dispersal distance estimates, it remains a real challenge to measure dispersal, due to the complexity of the methods involved and associated technical constraints (Rousset 2001; Broquet & Petit 2009). 'Direct' methods, based on direct observations of the movements of individuals (Dobzhansky & Wright 1943), are logistically very demanding and do not generally provide estimates of effective dispersal. 'Indirect' genetic methods (reviewed in Broquet & Petit 2009) are often based on assumptions, such as migration drift equilibrium (Wright 1943; Rousset 1997), and thus reveal long-term dispersal patterns between populations (Slatkin 1987; Peacock 1997). In addition, the population genetics models on which these methods are based are dependent on highly stochastic processes (such as drift), resulting in a large variance for parameter estimates (Bossart & Pashley Prowell 1998).

Biological invasions are natural experiments of non-equilibrium population dynamics in which dispersal can be estimated directly from transient situations. Appropriate theoretical models can be used to estimate dispersal directly from (i) a knowledge of the expansion speed of expanding populations [e.g. Skellam's model (Skellam 1951) or Fisher's model (Fisher 1937)] or (ii) description of the neutral spatial genetic structure of the contact zone between two expanding populations (Barton 1982; Barton & Gale 1993) and its temporal dynamics (Endler 1977). Many studies have investigated the expansion speeds of invading populations (e.g. Arim *et al.* 2006; Phillips *et al.* 2006), and some have provided estimates of dispersal (e.g. Lubina & Simon 1988; Andow *et al.* 1990). By contrast, transient contact zones between invading populations have seldom been used to estimate dispersal. Only one study to date seems to have used this approach (Rieux *et al.* 2013), even though this biological context is common and the

methodology is easy to implement. Biological invasions frequently begin with multiple independent introductions at different, remote locations (Facon *et al.* 2003; Kolbe *et al.* 2004; Bossdorf *et al.* 2005; Lockwood *et al.* 2005; Miller *et al.* 2005; Sax *et al.* 2005; Roman & Darling 2007; Ciosi *et al.* 2008; Dlugosch & Parker 2008). In cases of bottlenecks during introduction and/or drift during establishment, considerable genetic differentiation between the introduced populations may occur (e.g. Ciosi *et al.* 2008). The spatial expansion of these populations may lead to secondary contact (e.g. Rieux *et al.* 2013) and, eventually, to the formation of transient clines of allelic frequencies (e.g. Bermond *et al.* 2012). In cases of recent introduction, the introduced populations may be isolated for too short a time for selection to act against their hybrids (Orr & Turelli 2001; Edmands 2002). In case of neutral diffusion, allelic frequencies therefore tend to become homogeneous throughout the secondary contact zone over time. Cline slope declines at a rate dependent on dispersal strength and time since contact (Endler 1977; Rieux *et al.* 2013). This rate can therefore be used to estimate dispersal. Furthermore, dispersal between differentiated populations generates linkage disequilibrium, which peaks in the centre of the cline and can be used to obtain another estimate of dispersal (Barton 1982; Barton & Gale 1993). These methods are not completely independent because they both rely on the estimation of the cline slope. They are not dependent on migration drift equilibrium, instead being based on the instantaneous impact of dispersal on the genetic structure of the transient contact zone between two expanding populations.

In this study, we estimated the dispersal of one of the major pests of maize, the western corn rootworm (Coleoptera: Chrysomelidae), *Diabrotica virgifera virgifera*, which is currently invading Europe. *D. v. virgifera* completes its life cycle in a year (Branson & Krysan 1981). Eggs are laid during the summer and they overwinter in the soil, hatching in late spring. They then pupate and the adults emerge in early summer. They mate, disperse and lay eggs during the summer (Branson & Krysan 1981; Spencer *et al.* 2009). *D. v. virgifera* was introduced several times in Europe from the USA (Miller *et al.* 2005; Ciosi *et al.* 2008) and was detected for the first time in Europe near Belgrade Airport in 1992 (Sivcev *et al.* 1994). Population genetics analyses (Miller *et al.* 2005) showed that multiple introductions of *D. v. virgifera* were responsible for the European invasion and that the two principal invasive outbreaks in Europe, in North-West Italy (NW Italy, first observations in 2000) and Central and southeastern Europe (CSE Europe, first observations in 1992), were founded independently. Bermond *et al.* (2012) have recently shown that these two outbreaks have converged, com-

ing into contact in 2008, and that they now form a continuum in the Veneto region of northern Italy. There is considerable genetic differentiation between these two invasive populations because of intense bottleneck during their introduction (Ciosi *et al.* 2008), so this area has become an admixture zone, with allelic frequency clines at microsatellite loci (Bermond *et al.* 2012). It therefore presents an ideal opportunity for the estimation of the dispersal distances of *D. v. virgifera*.

We focused on the contact zone between the two largest invasive populations of *D. v. virgifera* in Europe to estimate the migration variance of this species, σ^2, and the variance of the parent offspring distance distribution. We carried out a spatial analysis of linkage disequilibrium and a temporal analysis of frequency clines on the basis of individual multilocus genotypes for 12 microsatellite loci. We also performed simulations to evaluate the effect of departure from common hypotheses on the estimate obtained.

Materials and methods

Sample collection

Samples were collected in summer 2012, from July 4 to 22, at 13 sites (see details in Table 1) in Italy, Slovenia and Hungary, along a west east transect crossing the outbreaks of NW Italy and CSE Europe (Fig. 1). The transect axis was chosen to match the direction of expansion of the outbreaks in this area (Fig. 1). For each sampling site, between 23 and 41 individuals were collected from the same maize field, and GPS coordinates were recorded. A total of 466 adult beetles were caught by hand (funnel bound with a 50-ml tube filled with
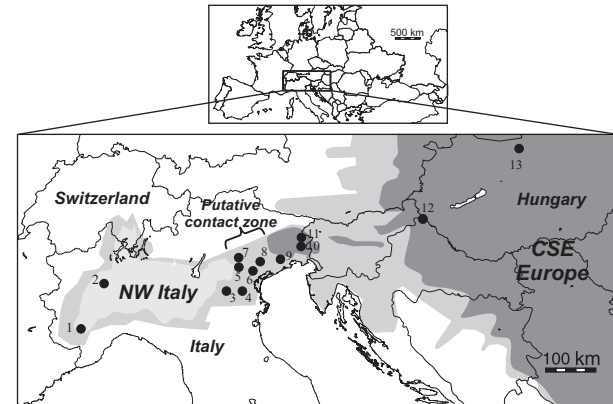
Fig. 1 Location of the European *Diabrotica virgifera virgifera* samples collected in 2012 in North West (NW) Italy, in the contact zone and in Central southeastern (CSE) Europe. Numbers refer to sample names (see in Table 1). The distribution areas of *D. v. virgifera* before contact between the western and eastern outbreak populations (in 2007) are shown in light grey (NW Italy) and dark grey (CSE Europe). The distribution area after contact (in 2012) is shown in medium grey.

96% alcohol) or with a mouth vacuum. All *D. v. virgifera* samples were stored at 20 °C in 96% ethanol until DNA extraction.

DNA extraction and microsatellite analyses

Individuals were washed three times in 0.065% NaCl before extraction, to remove ethanol from the tissues. For all samples, we extracted DNA from the thorax or half body, cut lengthwise, with the DNeasy® tissue kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions, with an elution volume of

Table 1 Description of the *Diabrotica virgifera virgifera* samples from Northern Italy and Central and South Eastern Europe. N is the sample size. x_i and y_i are the two dimensional (2D) geographical coordinates of sample i obtained from its GPS coordinates and converted into km with orthodromic distances

Sample name	Sample number	Region (Country)	Year of 1st observation	N	x_i	y_i	GPS coordinates (Latitude; Longitude)
Cuneo	1	Piedmont (Italy)	2000	41	0	0	(44°28′39.10″N; 07°34′24.10″E)
Olcenengo	2	Piedmont (Italy)	2000	35	58	99	(45°21′59.50″N; 08°18′18.40″E)
Vr24	3	Veneto (Italy)	2002	40	302	85	(45°14′35.50″N; 11°22′40.70″E)
Conselve	4	Veneto (Italy)	2002	35	340	84	(45°14′40.60″N; 11°51′31.40″E)
PD1	5	Veneto (Italy)	2002	40	333	130	(45°38′40.50″N; 11°45′55.80″E)
Scorzè	6	Veneto (Italy)	2002	33	369	120	(45°33′90.29″N; 12°13′27.40″E)
Borso del Grappa	7	Veneto (Italy)	2002	40	334	148	(45°48′26.30″N; 11°46′51.80″E)
Tv22	8	Veneto (Italy)	2002	40	377	137	(45°42′43.90″N; 12°19′49.70″E)
Summaga	9	Veneto (Italy)	2002	24	424	144	(45°46′18.91″N; 12°55′01.60″E)
Medeuzza	10	Friuli (Italy)	2003	23	462	162	(45°55′58.10″N; 13°24′02.50″E)
Premariacco	11	Friuli (Italy)	2003	34	462	175	(46° 03′17.00″N; 13°24′00.00″E)
Pince	12	Prekmurje (Slovenia)	2003	41	710	227	(46°31′15.38″N; 16°31′23.70″E)
Budapest	13	Pest (Hungary)	1992	40	888	338	(47°31′12.90″N; 18°46′33.10″E)

100 μL. Twelve microsatellite loci of *D. v. virgifera* (including di- and trinucleotides) (details available in Bermond *et al.* 2012) were amplified in three separate multiplex PCR performed in a PTC-225 MJ Research thermocycler and were analysed as described by Miller *et al.* (2007).

Cline analysis

Descriptive analysis and cline fitting. We first tested Hardy Weinberg equilibrium for each sample, with the probability test approach implemented in GENEPOP, version 4.0.1 (Raymond & Rousset 1995; Rousset 2008). We used the sequential Bonferroni procedure (Sokal & Rohlf 1995) to take multiple testing into account. For cline analysis, the alleles of each microsatellite were clustered into two groups identified in a multiple corre-spondence analysis (Daguin *et al.* 2001; Bierne *et al.* 2003), allowing maximal frequency variance along the cline. Assuming a scenario of neutral secondary contact, the frequency f_{ijk} of allele group k of locus j in sample i can be described by the scaled cumulative Gaussian distribution (a Mathematica notebook has been sub-mitted to Dryad at doi:10.5061/dryad.10kd4 with full details):

$$f_{ij1}(\rho_{ij}) = h_{1j} + h_{pj}(1 \quad h_{1j}) \int_{-\infty}^{\rho_{ij}} \frac{1}{b_j\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{c_j - \rho_{ij}}{b_j})^2}$$
$$\text{and } f_{ij2}(\rho_{ij}) = 1 \quad f_{ij1}(\rho_{ij}) \qquad \text{(eqn 1)}$$

where ρ_{ij} is the geographical coordinate of sample i along the cline axis of locus j, c_j is the cline centre, b_j is a cline slope parameter (and the standard deviation of the Gaussian), and h_{1j} and $h_{1j} + h_{pj}(1 \quad h_{1j})$ are the frequencies of the first allele group at the two extremities of the cline of locus j (see the Mathematica notebook). For each locus j, the angle θ_j between the parallel and the cline axis was used to calculate ρ_{ij}, the transformed geographical distance of each sample i along the cline axis of locus j, $\rho_{ij} \quad x_i \cos(\theta_j) + y_i \sin(\theta_j)$, where x_i and y_i are the two-dimensional (2D) geographical coordinates

of sample i obtained from its GPS coordinates and con-verted into km with orthodromic distances (see Table 1).

Model selection and parameter estimation. Various models of scaled cumulative Gaussian distributions, differing in the number of parameters included, were fitted to the alle-lic group frequencies of the 12 microsatellites, making it possible to test hypotheses about changes in cline shape. The most complex model (model M_8) had five parameters (h_{1j}, h_{pj}, c_j, b_j and θ_j) per locus j, and the simplest model (model M_1) had two parameters (h_{1j} and h_{pj}) per locus j plus three parameters (c, b and θ) common to all loci (all details of model configurations available in Table 2). Assuming a pure diffusion process, we would expect to select model M_1 with coincident, concordant and parallel clines (i.e. with identical parameters c, b and θ) for all loci. Assuming a diffusion process with selection acting on some genomic regions to which some of the microsatel-lites are linked or strong genetic drift, parameters c_j, b_j and θ_j might be expected to vary between loci, and we would expect to select a model other than model M1. We thus compared the eight models, with various levels of con-straint on parameters c_j, b_j and θ_j, using the information theory approach proposed by Burnham and Anderson (2010). The best model was chosen on the basis of the cor-rected QAIC$_c$ criterion (Burnham & Anderson 2010). This modified Akaike criterion takes into account sample size and overdispersion of the data by correcting the deviance with the variance inflation factor ĉ (Lebreton *et al.* 1992).

Estimates of the scaled cumulative Gaussian distribu-tion parameters of the chosen model were obtained by maximizing the log likelihood of the observed frequencies:

$$\sum_i \sum_j \sum_k n_{ijk} \ln(f_{ijk}) \qquad \text{(eqn 2)}$$

with n_{ijk} the observed number and f_{ijk} the expected fre-quencies of individuals displaying allele group k at locus j in population i of eqn 1. Maximum-likelihood (ML) estimates of model parameters h_{1j}, h_{pj}, c_j, b_j and θ_j

Table 2 Comparison of the quality of fit of the various clinal models of the spatial pattern of microsatellite frequencies for *Diabrotica virgifera virgifera* in the contact zone between the two main European invasive populations. c indicates the centre of the cline, s the slope and θ the angle. QAICc is the modified Akaike criterion for sample size and for overdispersion (see the text for details). The best fit according to QAICc is shown in bold typeface

Model	**M$_1$**	M$_2$	M$_3$	M$_4$	M$_5$	M$_6$	M$_7$	M$_8$
Configuration	**1 c, 1 s, 1 θ**	12 c, 1 s, 1 θ	1 c, 12 s, 1 θ	1 c, 1 s, 12 θ	12 c, 1 s, 12 θ	1 c, 12 s, 12 θ	12 c, 12 s, 1 θ	12 c, 12 s, 12 θ
Maximum Log likelihood	**4395.17**	4359.67	4360.01	4376.41	4348.71	4338.21	4339.00	4337.20
Number of parameters	**27**	38	38	38	49	49	49	60
QAICc	**1178.09**	1205.24	1205.33	1209.46	1246.15	1243.50	1243.70	1304.4
% of Deviance	**86**	87	87	87	88	88	88	88

were obtained simultaneously with the FindMaximum function of Mathematica 8.0 (Wolfram Research, Inc., 2010) (see the Mathematica notebook). At least 3 estimations were obtained with various starting values of the parameters to avoid local likelihood maxima.

Linkage disequilibrium analysis

Descriptive analysis of linkage disequilibrium. The second-order multilocus cumulant $\kappa_{0,2}$ defined by Barton (2000) was used as the average multilocus linkage disequilibrium, taking into account allelic frequencies and their variation. As pointed out by Gay *et al.* (2008), estimates of $\kappa_{0,2}$ are subject to drift and sampling errors. Moreover, no estimate of $\kappa_{0,2}$ was available for the precise centre of the cline because there was no sampling site at this position. We therefore fitted a spatial function to $\kappa_{0,2}$. In secondary contact with pure diffusion and no selection, linkage disequilibrium   here assessed by measuring $\kappa_{0,2}$ should be approximately proportional to a Gaussian distribution (see the Mathematica notebook). We therefore fitted it to a function of the form:

$$\text{LD}(\rho_i) = D_{\max}e^{-f(\rho_i-g)^2}(\text{with } f > 0) \qquad \text{(eqn 3)}$$

where ρ_i is the geographical coordinate of sample i along the cline axis, $D_{\max}$ is the maximum value of linkage disequilibrium at the centre of the cline g, and f is a variance parameter of the Gaussian distribution. Note that in cases of pure diffusion, the clines of all loci have the same centre and are parallel.

Maximum-likelihood estimation of maximal linkage disequilibrium. Multilocus genotypic frequencies depend on allelic frequencies, heterozygote deficits or excesses and linkage disequilibrium. As no deviation from Hardy Weinberg equilibrium was recorded for any locus, multilocus genotype frequencies depended on allelic frequencies and linkage disequilibrium only in this study. For each sample i, an estimate of $\kappa_{0,2i}$ can thus be obtained by maximizing the log-likelihood LL_i of the multilocus genotype frequencies (Barton 2000, see the Mathematica notebook for details). Note that locus Dvv-T3 was removed from this analysis because of its very low variation in space and for the sake of computational feasibility. In addition, the total log-likelihood $\sum \text{LL}_i$ in which each $\kappa_{0,2i}$ was replaced by $\text{LD}(\rho_i)$ (eqn 3) was used to fit the spatial model of linkage disequilibrium. Maximum-likelihood (ML) estimates of the spatial model parameters, including $D_{\max}$, were obtained simultaneously using a simulated annealing algorithm, adapted from the work of Szymura and Barton (1986) and Gay *et al.* (2008) and implemented in Mathematica 8.0 (Wolfram Research, Inc., 2010) using the code of Gregory (2013) (see the Mathematica notebook).

For analyses of both clines and the spatial distribution of linkage disequilibrium, the associated support limits of parameter estimates were defined as the range of values within two units of log likelihood from the maximum (Edwards 1972).

Estimation of dispersal

Estimation of dispersal from cline width analysis. When two genetically differentiated populations meet in the absence of selection against gene flow, clines are formed at neutral markers and progressively vanish at a pace determined by the amount of dispersal between the two genetic entities. Specifically, when the dispersal kernel is well approximated by a diffusion equation (e.g. a Gaussian kernel), the decay of the cline after T generations since contact results in a cline width given by:

$$w = \sigma\sqrt{2\pi T} \qquad \text{(eqn 4)}$$

(Endler 1977; Gay *et al.* 2008), where σ is the diffusion coefficient (which, here, is also equal to the standard deviation of the normally distributed parent offspring distances). If the widths of the clines are all the same, the point estimate of the parent offspring variance distance and its support limits can thus be estimated by calculating the maximum log likelihood of eqn. 2, in which we replaced f_{ijk} by the scaled cumulative Gaussian distribution of eqn. 1 and parameters h_{1j}, h_{pj}, c_j, and θ_j by their point estimates. The maximum slope of the scaled cumulative Gaussian distribution is $\frac{1}{b\sqrt{2\pi}}$ (see the Mathematica notebook). Thus, in cases of a pure diffusion process, the width of the clines is the same for all neutral markers and is given by $w = b\sqrt{2\pi}$. We therefore replaced b with $w/\sqrt{2\pi} = \sigma\sqrt{T}$ in eqn. 2. Finally, historical observations suggest that these two invasive populations first came into contact four generations ago (Bermond *et al.* 2012). We thus used this value for T in eqn. 2. This method is referred to hereafter as Endler's method.

Estimation of dispersal from linkage disequilibrium analysis. Positive linkage disequilibrium is expected in cases of admixture between two genetically differentiated populations. According to Barton (1982) Barton and Gale (1993), the maximum linkage disequilibrium between two loci occurs at the centre of the contact zone and can be expressed as:

$$D_{\max} = \frac{\sigma^2}{r}\frac{1}{w_1 w_2} \qquad \text{(eqn 5)}$$

where w_i is the cline width at locus i, and r is the rate of recombination between the two loci. Even though Barton proposed this method for use with clines maintained by selection (e.g. Barton 1982), the method

assumes an absence of selection at the centre of the cline and is thus suitable for the neutral case presented here. If the widths of all the clines are the same ($w_i = b\sqrt{2\pi}$ for all microsatellites), the joint estimation of w and σ is obtained by maximizing the joint log-likelihood $(\sum_i LL_i + \sum_i \sum_j \sum_k n_{ijk} \ln(f_{ijk}))$ of the data, in which each $\kappa_{0,2i}$ is replaced with $LD(\rho_i)$ (eqn 3) and D_{max} is replaced with $\frac{\sigma^2}{r}\left(\frac{1}{b\sqrt{2\pi}}\right)^2$. This method is referred to hereafter as the LD method.

The combination of eqn 4 and eqn 5 leads to:

$$T = \frac{1}{2\pi r D_{max}} \qquad \text{(eqn 6)}$$

This provides a way, independent of the estimation of σ, to estimate the time since contact T by maximizing the log likelihood of the multilocus genotype frequencies $\sum_i LL_i$ in which each $\kappa_{0,2i}$ is replaced with $LD(\rho_i)$ (eqn 3).

Simulations

The AIC-based comparison described above can be used to compare models with small or large numbers of parameters, when the parameters are common to or different for all loci, respectively. This method can be used to detect the effect of processes, for example selection, affecting the various loci differently. Demographic parameters, such as dispersal and population density, affect all neutral genetic markers equally and therefore cannot be evaluated by AIC-based model comparison. We thus simulated genetic data with a known dispersal kernel and heterogeneity of population density, to evaluate their effect on the estimation of dispersal and time since contact. Simulations used discrete time processes and included various dispersal kernels with various kurtoses and various population density differences across the contact zone. The variance of parent offspring distance was then estimated from these simulated data with Endler's method and the LD method and was compared with its known value. Time since contact was estimated by combining Endler's and LD methods and also compared with its known value.

Genotypic frequencies at two biallelic loci were simulated by a series of 700 demes connected by migration (Mallet *et al.* 1990; Lenormand *et al.* 1998) with reflecting boundaries. Reproduction was simulated by assuming Hardy Weinberg equilibrium in each deme and independence between loci. In each generation, population densities were set to 1000 individuals, except for the simulations with varying densities. For those simulations, population densities were set to K_1

(varying between 200 and 1000) and K_2 (1000) individuals per deme for $x < z$ and $x > z$, respectively, with z the location of the initial contact between the two populations. Allelic frequencies at each locus were set to 1 for $x < z$ and to 0 for $x > z$ just before contact. Diploid individuals dispersed randomly according to an isotropic migration model with various dispersal kernels. These kernels were mixtures of Gaussian distributions or exponential power distributions, including the Laplace kernel, differing in kurtosis and tail heaviness (Table S1, Supporting information). Parameter values of each dispersal kernel were chosen so that all distributions had the same variance, close to the estimated σ^2 for *D. v. virgifera*, but different kurtoses (γ ranging from 1 to 30, Table S1, Fig. 5). For each set of parameters, five replicates of the dynamics of genotypic frequencies were simulated for four generations after contact.

For each parameter set, simulations were performed with R-2.15.1 (R Development Core Team 2012). The R script used for the simulations is available on Dryad at doi:10.5061/dryad.10kd4. As only two loci were simulated, it was not necessary to calculate an average multilocus linkage disequilibrium $\kappa_{0,2}$, but only the usual two-locus index of linkage disequilibrium. As individuals were sampled after dispersal, r was replaced by $r/(r + 1)$ in eqn 5, as suggested by Barton and Gale (1993). We were not interested in evaluating the effect of sampling variance on the precision of the estimations. Therefore, 2 loci, 100 demes and 100 individuals were sampled per deme rather than the 12 loci and 13 samples of about 30 individuals in the real data set. For this reason and for reasons of tractability, we provide only point estimates of σ and T, without support limits, in the simulations.

Results

Cline analysis

We observed no significant deviation from Hardy Weinberg equilibrium (not shown here) after sequential Bonferroni analysis (Sokal & Rohlf 1995), and western (Cuneo and Olcenengo) and eastern populations (Pince and Budapest) displayed significant genetic differentiation ($p < 10^{-5}$, mean value of F_{ST} 0.26). All polymorphic loci except for DVV-T3 displayed a clinal pattern (Fig. 2). According to the QAICc criterion, the model best fitting the allelic frequencies at the 12 microsatellite loci was the simplest model, M_1, which accounted for 86% of the deviance (Table 2 and Fig. 2). All frequency clines are thus coincident (they have the same centre), concordant (they have the same width) and parallel (they are at the same angle to the parallel axis). These results are consistent with a pure diffusion model and thus allow estima-
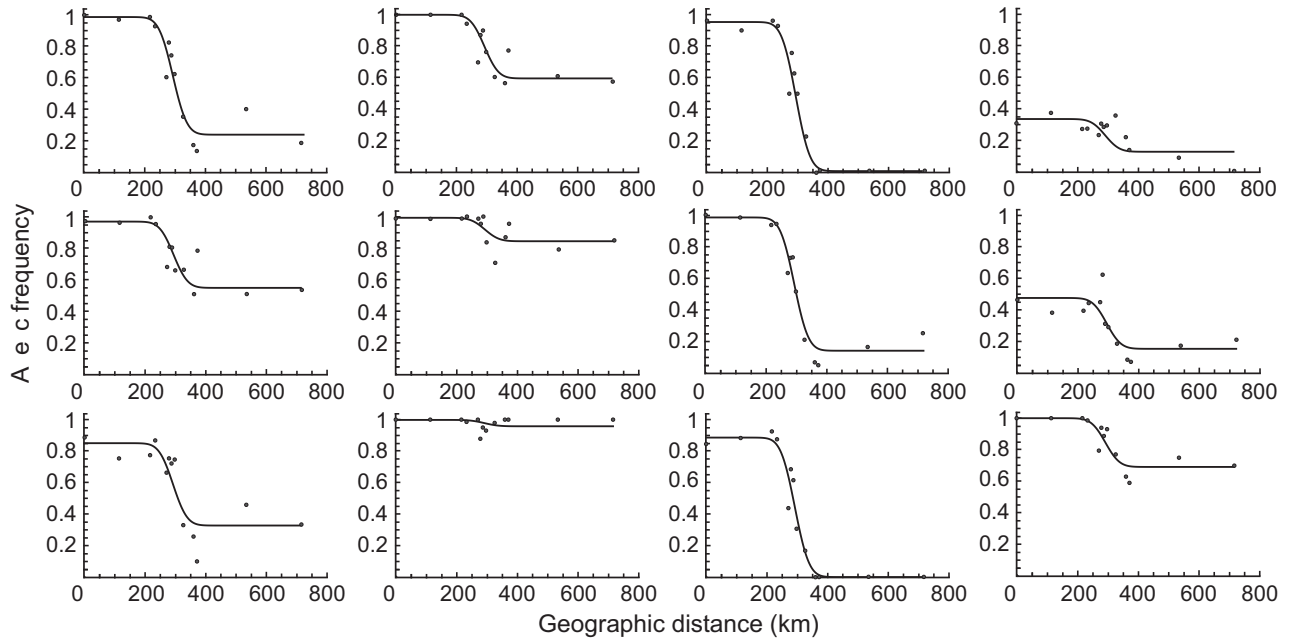
Fig. 2 Observed (dots) and fitted (lines) frequencies of the allelic groups of the 12 polymorphic microsatellites (from top to bottom and left to right: Dvv D2, Dvv D4, Dvv D8, Dvv D9, Dvv D11, Dvv T2, Dvv ET1, Dba 01, Dba 05, Dvv T3, Dvv D12 and Dba 07) from *Diabrotica virgifera virgifera* in the region of contact between the two main invasive populations in Europe. Allelic group frequen cies are represented as functions of geographical distance, along the cline axis (see text for details).

tion of the dispersal of *D. v. virgifera* by Endler's method and the LD method. According to the M_1 model, we obtained the following ML estimations: the centre of the clines is located 293 km (with support limits of 265 321 km) from Cuneo, the width of the clines is 91.6 km (80.8 103.1 km), and the angle between the cline and the *x* axis is 1.08 radians (0.96 1.18 radians).

Assuming that the western and eastern populations came into contact four generations before sampling (i.e. in 2008, Bermond *et al.* 2012), Endler's method provides a ML estimate of the standard deviation σ of the parent offspring distances measured in one dimension of 18.26 km/generation$^{1/2}$ (16.11 20.56 km/generation$^{1/2}$). If uncertainty about the year of contact was taken into account (T varying from 2 to 5 years), then the ML estimate of σ and its support limits ranged between 14.4 and 29.1 km/generation$^{1/2}$ (Fig. 3a).

Linkage disequilibrium analysis

Most ML estimates of average linkage disequilibrium $\kappa_{0,2}$ had support limits including 0, except for five samples close to the centre of the clines (Fig. 4). The maximum estimate of $\kappa_{0,2}$ was 0.11 (0.09 0.12), in Summaga, Veneto (Fig. 4). ML estimation of the parameters of the spatial fit of linkage disequilibrium provided an estimate of 0.113 (0.078 0.136) for maximum linkage disequilibrium D_{max} (Fig. 4).

As a model with concordant clines for all loci was selected, (eqn 5) (Barton & Gale 1993) can be reduced to $\sigma = w\sqrt{D_{max}}r$, where w is the cline width common to all loci. Figure 3b shows the support area of w and σ, assuming that all loci are physically independent (r 0.5). The ML estimate of σ is 21.74 km/generation$^{1/2}$ with a support area bounded by 17.30 25.52 km/generation$^{1/2}$ (Fig. 3b, Mathematica notebook).

By combining Endler's and LD's method, we obtained an estimate of the time since contact T of 2.82 years (2.33 4.08 years).
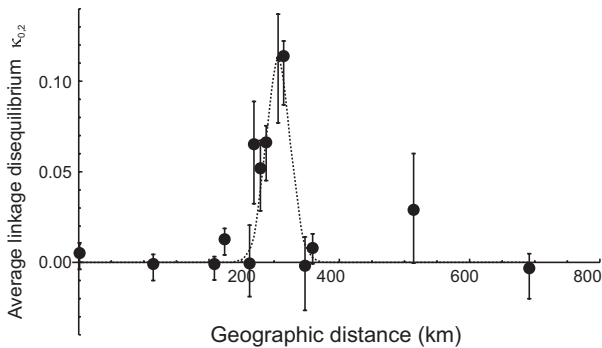
Simulations

With Gaussian and exponential kernels and no difference in population size between the two sides of the contact zone, relative biases of the estimate of the standard deviation of the parent offspring distance were consistently below 12% (Fig. 5) and those of the time since contact were between 10 and 15%. There were systematic differences between estimation methods, with LD method producing the largest estimates of σ and the largest biases. Overall, the estimates of σ and T were only weakly biased, whatever the kurtosis value and population density differences, with errors always below 30%. The largest relative biases of σ estimation were found for extreme values of kurtosis (with biases
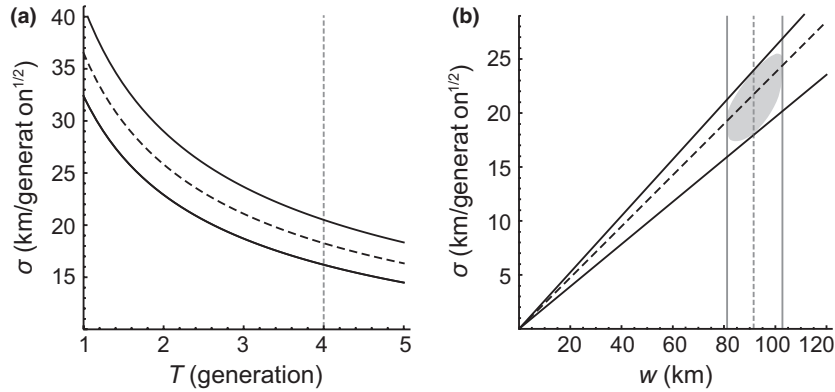
Fig. 3 Maximum likelihood (ML) estimates of σ, the standard deviation of offspring parent dispersal distances, obtained by the Endler's and LD methods. (a) ML estimate of σ as a function of the time since contact T between the NW Italian and CSE European invasive populations of *Diabrotica virgifera virgifera*. Dashed and solid lines indicate the ML estimate of σ obtained by the Endler's method and its associated support limits, respectively. The vertical dashed line in grey corresponds to the value of T (4 years) suggested by historical information. (b) ML estimate of σ obtained by the LD method. The support area of σ and the width of the clines w are shown in grey and are obtained from the joint support limits of D_{max} (the maximum value of linkage disequilibrium at the centre of the cline) and w. The dashed and solid black lines correspond to the ML estimates of D_{max} and its associated support limits, respectively. The dashed and solid grey lines correspond to the ML estimates of w and its associated support limits, respectively.

Fig. 4. Maximum likelihood estimates (full circles) and spatial fit (dotted line) for the average linkage disequilibrium, $\kappa_{0,2}$ (Barton 2000), of *Diabrotica virgifera virgifera* in the contact zone between the two main European invasive populations in North West Italy and Central Europe, as a function of geographical distance along the cline axis. Vertical bars indicate the support limits for the estimated values of $\kappa_{0,2}$ and the maximum average linkage disequilibrium D_{max}.

of about 25% for γ 30) for the mixture of Gaussians and for the highest density differences between the two sides of the contact zone (with biases of about 20% for K_1/K_2 0.2). The largest relative biases of T estimation (of about 30%) were found for the highest density differences between the two sides of the contact zone (K_1/K_2 0.2).

Discussion

This study was carried out on a secondary contact zone between two invading European populations of

D. v. virgifera, to obtain estimates of dispersal distances for this species. Studies of the population genetic structure of the contact zone and of the spatial variation of microsatellite allelic frequency and linkage disequilibrium generated two estimates of the standard deviation of parent offspring distances for *D. v. virgifera*. For both methods, we obtained ML estimates between 18 and 22 km/generation$^{1/2}$, with support limits roughly bounded by 16 and 26 km/generation$^{1/2}$. Assuming a Gaussian dispersal kernel, this standard deviation corresponds to a mean distance of about 13 21 km/generation (the mean of a folded Gaussian distribution is $\sigma\sqrt{2/\pi}$). We also obtained an estimate of the time since contact between the two invasive outbreaks of northern Italy and Central and southeastern Europe of about 3 years (bounded by 2.33 and 4.08 years). This is in fair agreement with historical information that dated the contact 4 years before sampling.

Validity of the estimate

The methods for estimating dispersal used here are based on the following assumptions: genotypic frequencies result from a pure migration process, with migration amounting to a diffusion process. As a consequence, genetic drift is assumed to be negligible, and loci are assumed to be neutral and independent. Finally, population size is assumed to remain constant over space.

Microsatellite loci are expected to be neutral [(Ellegren 2004), but see (Sutherland & Richards 1995; Watkins *et al.* 1995)], but it is possible that they are linked to genomic regions subject to selection (Charles-
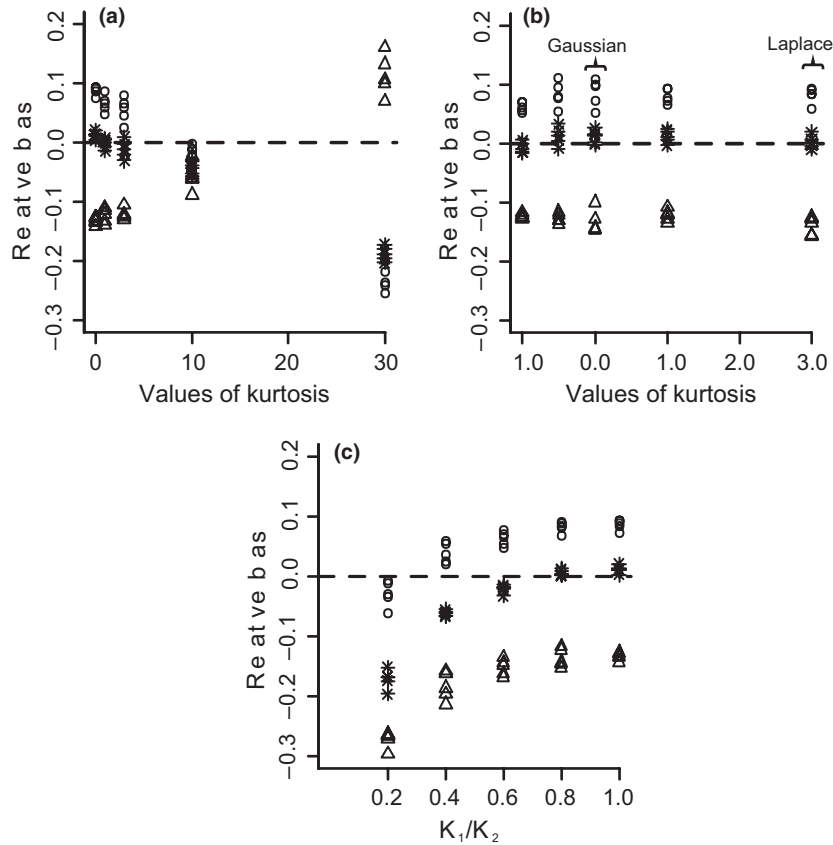
Fig. 5 Influence of dispersal kernels [with mixtures of Gaussian kernels in (a) and exponential power kernels in (b)] and population density differences between the two sides of the contact zone (c) on the relative bias of ML estimates of the dispersal σ and the time since contact T as revealed by simulations. The relative bias of parameter X is defined as $\frac{\hat{X}-X}{X}$, with $\hat{X}$ the estimated value and X the value of the parameter used to simulate the data. Number of demes 700, number of individuals per deme from 200 to 1000, σ 20 km/generation$^{1/2}$, number of simulated generations 4 (see the R script used for the simulations). The characteristics of dispersal kernels are presented in Supporting Information (Table S1). Dashed lines indicate a relative bias of zero. The ratio K_1/K_2 indicates the ratio of deme density between the two outbreaks. For each part (a, b and c), asterisks, circles and triangle represent the relative bias of σ obtained with Endler's and LD method and the relative bias of T, respectively.

worth *et al.* 1993; Zhong *et al.* 1994). Several lines of evidence strongly suggest that selection was not involved in shaping the clines described here: selection on specific genomic regions may affect the various loci differently and create heterogeneity in cline slopes and centres. All the clines in this study were concordant, coincident, smooth sigmoid clines, whereas stepped clines would be expected in case of strong selection on multiple loci (Barton & Gale 1993). Moreover, selection against hybrids is unlikely because both outbreaks (NW Italy and CSE Europe) began too recently (founded in 2000 and 1992, respectively), and their isolation time was short (about 20 generations) (Orr & Turelli 2001; Edmands 2002).

Long-distance dispersal is frequent in many species (Lockwood *et al.* 2007). For instance, Kim and Sappington (2004) found that recurrent large-scale migrations between populations of the boll weevil, *Anthonomus*

grandis, took place in North America. More recently, several authors have shown that *D. v. virgifera* can also disperse over long distances (Miller *et al.* 2005; Ciosi *et al.* 2011). Consequently, dispersal kernels are often not diffusive, instead having heavy tails and positive kurtoses (Lockwood *et al.* 2007; Liebhold & Tobin 2008). We simulated genotypic data resulting from various kernels (mixture of Gaussian or exponential power laws), with light or heavy tails, and with a kurtosis of between 1 and 30. If dispersal is not too leptokurtic, simulations showed that the true value of **σ** is closed to that estimated under diffusion hypothesis. Simulations thus indicated that dispersal estimates were not very sensitive to the type of dispersal kernel and kurtosis, with a relative bias of <25%. These results are consistent with previous findings showing that kurtosis had little effect on the estimation of σ with comparable methods (Rieux *et al.* 2013). If long-distance dispersal is very rare, then the dispersal

parameter estimated here is the same as the diffusion parameter. However, if there is enough long-range movement, the expansion dynamics can significantly change. The co-occurrence of both local diffusion and long-distance dispersal – a phenomenon called stratified dispersal (Shigesada *et al.* 1995) – can lead to a faster expansion than expected under local diffusion alone. With stratified dispersal, remote colonies are stochastically sent outside the expanding population. They eventually merge with the main population and contribute to the advance of the front (Shigesada *et al.* 1995).

In very large study areas, such as that studied here (more than 800 km wide), there is probably environmental heterogeneity rendering population density heterogeneous over space. Moreover, the two invasive populations of *D. v. virgifera* are of different ages, the NW Italian outbreak occurring about 8 years after the CSE European outbreak (Miller *et al.* 2005), and this may account for differences in density. A difference in density between the two colonizing populations may also result from differences in anthropogenic selection intensity (insecticide treatments and crop rotations). When densities are equal, the slope of the cline is a function of dispersal. Stepwise differences in densities across the contact zone would be expected to lead to a steeper slope of the cline, because of the stronger contribution to gene flow of the population with the highest density (Barton 1979). We thus simulated a worst-case scenario in which there were constant differences in population size between the two populations. Simulations showed that these differences had little influence on dispersal estimates, as biases were always below 20%, even for a density ratio of 500%.

We assumed a recombination rate of 0.5 between all microsatellites. This seems reasonable, because all linkage disequilibrium values were zero outside the contact zone. However, this assumption may be false for some microsatellites, resulting in a mean recombination rate slightly below 0.5. As $\sigma^2 \propto r$ with the LD method (Barton & Gale 1993), this may result in an overestimation of σ when considering r 0.5. Note that for a wide range of recombination rates, between 0.3 and 0.5, the assumption of complete independence between loci would result in a limited bias of the estimation of σ of less than 25%.

The methodologies based on Endler's method and LD (Endler 1977; Barton & Gale 1993) are both based on the assumption that drift has a limited effect. We were confident from the outset that this was the case for *D. v. virgifera* in northern Italy, as very large population sizes are often reported for invasive outbreaks (Boriani *et al.* 2006; Szalai *et al.* 2011), and previous studies have shown negligible genetic differentiation between samples within outbreaks (Ciosi *et al.* 2008; Bermond *et al.*

2012). We are also confident, with hindsight, that drift had little effect on the genetic structure of *D. v. virgifera* in the contact zone because (i) we found concordant, coincident and parallel clines, whereas drift would be expected to create some heterogeneity in cline shape (width, centre and angle) between loci, and (ii) 86% of the deviance was explained by the M_1 model, whereas drift would be expected to generate considerable background noise around expected allele frequencies, resulting in a poor fit.

Comparison with other dispersal estimates for *D. v. virgifera*

Diabrotica virgifera virgifera is one of the worst pests of maize worldwide. The control strategies used against it, including the use of GM maize producing *Bacillus thuringiensis* toxins and refuge zones, are based on the hypothesis that *D. v. virgifera* disperses effectively. The combination of different selective pressures (i.e. various pesticides) over space, to increase the degree of treatment heterogeneity (REX 2013) against *D. v. virgifera,* requires a rough knowledge of the distances over which *D. v. virgifera* can disperse. *D. v. virgifera* is invading Europe, and the invasive outbreaks are expanding geographically at a speed dependent on the dispersal capacities of *D. v. virgifera.* Risk analyses and simulations of various scenarios of human defence against *D. v. virgifera* spread (Baufeld & Enzian 2001; Carrasco *et al.* 2010) require the parameterization of *D. v. virgifera* dispersal capacities. This is why many previous studies have focused on the dispersal of *D. v. virgifera* in the USA (Coats *et al.* 1986; Gray *et al.* 2009) and in Europe (Baufeld & Enzian 2001; Ciosi *et al.* 2011).

Carrasco *et al.* (2010) performed capture mark recapture experiments and showed that half the individuals tested in the field moved between 117 and 425 m/day (with a median of 117 188 m/day). Given that the adult lifespan of *D. v. virgifera* is three months (Branson & Johnson 1973; Hill 1975), these daily distances correspond to between 11 and 38 km/generation (or a median of 11 17 km/generation), consistent with our estimates. Baufeld and Enzian (2001) fitted a model to the expansion of the CSE Europe population since the start of this invasion near Belgrade in 1992 and estimated the expansion rate at 60 100 km/year. Assuming a diffusion approximation with a Gaussian dispersal kernel, Fisher's wave of advance model suggests that the expansion rate of a colonizing population with an intrinsic growth rate r is C $2\sigma.r^{1/2}$ (Shigesada & Kawasaki 1997). In 2011, Szalai *et al.* (2011) showed that the mean annual growth rate of the CSE European outbreak was about four, giving a value of σ between 15 and 25 km/year$^{1/2}$. Our estimates are consistent with these findings.

The temporary admixture zone as an opportunity for estimating dispersal

Temporary and vanishing admixture zones following secondary contact are rarely used to estimate dispersal. One recent study made use of the relationship found by Endler (1977) between dispersal and the slope of neutral clines to estimate dispersal distances in secondary contact zones between two colonizing populations of the banana pest *Mycosphaerella fijiensis* (Rieux *et al.* 2013). Rieux *et al.* (2013) fitted a scaled logistic model to their microsatellite frequencies samples at two different times. The slope difference between the two periods was used to estimate **σ** through Endler's equation. In our case, we did not use a temporal sampling because (i) we had a good idea of the time since contact (which was not the case of Rieux *et al.* (2013)) and (ii) we could estimate **σ** from linkage disequilibrium without knowing T with a single year of sampling. Most studies using linkage disequilibrium in a secondary contact zone to estimate σ focus on hybrid zones maintained by selection (e.g. Barton 1982; Barton & Gale 1993; Gay *et al.* 2008; Singhal & Moritz 2012). To our knowledge, this study is also the first to use a spatial fit of the multilocus linkage disequilibrium in a maximum-likelihood framework. This fit was computationally tractable, because all samples were at Hardy Weinberg equilibrium, such that it was not necessary to estimate the between-genome within-gene association parameter $\kappa_{1,1}$ (Barton 2000) along with $\kappa_{0,2}$.

The methodology used here in the context of secondary contact between invasive populations has several advantages: (i) it provides estimates of effective dispersal distance subject to only weak bias, (ii) it provides estimates that are robust to departure from several hypotheses, (iii) it provides narrow support limits, indicating a low level of uncertainty on the estimates, (iv) it is cost- and time-effective as it requires the genotyping of only a few polymorphic genetic (diallelic) markers in a few hundred individuals sampled only once, and (v) there is no need to know precisely the time for which the populations have been in contact, and this factor can be estimated. The main constraint associated with this method is that the location of the secondary contact zone must be roughly known before sampling, to ensure accurate estimations of cline width.

This methodology seems to be suitable for studies of many cases of biological invasions *in natura*. Many invasions originate from multiple introductions in disconnected areas (Urbanelli *et al.* 2000; Barney 2006; Dlugosch & Parker 2008; Hufbauer & Sforza 2008; Lombaert *et al.* 2010) and many lead to secondary contact (e.g. Kolbe *et al.* 2004; Lavergne & Molofsky 2007; Lombaert *et al.* 2010; Lucek *et al.* 2010; Rieux *et al.* 2013).

Temporary and vanishing allelic frequency clines are thus probably very frequent *in natura*, particularly in the context of biological invasions, and could be used to estimate effective dispersal distance accurately in many species provided time since contact is not too large.

References

Andow D, Kareiva P, Levin SA, Okubo A (1990) Spread of invading organisms. *Landscape Ecology*, **4**, 177 188.

Arim M, Abades SR, Neill PE, Lima M, Marquet PA (2006) Spread dynamics of invasive species. *Proceedings of the National Academy of Sciences USA*, **103**, 374 378.

Barney J (2006) North American history of two invasive plant species: phytogeographic distribution, dispersal vectors, and multiple introductions. *Biological Invasions*, **8**, 703 717.

Barton NH (1979) The dynamics of hybrid zones. *Heredity*, **43**, 341 359.

Barton NH (1982) The structure of the hybrid zone in Uroder ma bilobatum (Chiroptera, Phyllostomatidae). *Evolution*, **36**, 863 866.

Barton NH (2000) Estimating multilocus linkage disequilibria. *Heredity*, **84**, 373 389.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zone. In: *Hybrid zones and The Evolutionary Process* (ed Harison RG), pp. 13 45. Oxford University Press, New York.

Baufeld P, Enzian S (2001) Simulations model for spreading scenarios of western corn rootworm (Diabrotica virgifera virgifera) in case of Germany. *IWGO Newsletter*, **22**, 14 15.

Bermond G, Ciosi M, Lombaert E *et al.* (2012) Secondary con tact and admixture between independently invading popula tions of the western corn rootworm, Diabrotica virgifera virgifera in Europe. *PLoS ONE*, **7**, e50129.

Bierne N, Borsa P, Daguin C *et al.* (2003) Introgression patterns in the mosaic hybrid zone between Mytilus edulis and M galloprovincialis. *Molecular Ecology*, **12**, 447 461.

Boriani M, Agosti M, Kiss J, Edwards CR (2006) Sustainable management of the western corn rootworm, *Diabrotica virgif era virgifera* LeConte (Coleoptera: Chrysomelidae), in infested areas: experiences in Italy, Hungary and the USA. *EPPO Bulletin*, **36**, 531 537.

Bossart JL, Pashley Prowell D (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology & Evolution*, **13**, 202 206.

Bossdorf O, Auge H, Lafuma L *et al.* (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia*, **144**, 1 11.

Branson TF, Johnson RD (1973) Adult western corn rootworms: oviposition, fecundity, and longevity in the laboratory. *Journal of Economic Entomology*, **66**, 417 418.

Branson TF, Krysan JL (1981) Feeding and oviposition behavior and life cycle strategies of *Diabrotica*: an evolutionary view with implications for pest management. *Environmental Entomology*, **10**, 826 831.

Broquet T, Petit EJ (2009) Molecular estimation of dispersal for ecology and population genetics. *Annual Review of Ecology Evolution and Systematics*, **40** 193 216.

Burnham KP, Anderson DR (2010) *Model Selection and Multimodel Inference: A Practical Information Theoretic Approach*, 2nd edn. Springer, New York, New York.

Carrasco LR, Harwood TD, Toepfer S *et al.* (2010) Dispersal kernels of the invasive alien western corn rootworm and the effectiveness of buffer zones in eradication programmes in Europe. *Annals of Applied Biology*, **156**, 63 77.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289 1303.

Ciosi M, Miller NJ, Kim KS *et al.* (2008) Invasion of Europe by the western corn rootworm, *Diabrotica virgifera virgifera*: multiple transatlantic introductions with various reductions of genetic diversity. *Molecular Ecology*, **17**, 3614 3627.

Ciosi M, Miller NJ, Toepfer S, Estoup A, Guillemaud T (2011) Stratified dispersal and increasing genetic variation during the invasion of Central Europe by the western corn rootworm, *Diabrotica virgifera virgifera*. *Evolutionary Applications*, **4**, 54 70.

Coats SA, Tollefson JJ, Mutchmor JA (1986) Study of migratory flight in the western corn rootworm (Coleoptera: Chrysomelidae). *Environmental Entomology*, **15**, 620 625.

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.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431 449.

Dobzhansky T, Wright S (1943) Genetics of natural populations. X. Dispersion Rates in *Drosophila Pseudoobscura*. *Genetics*, **28**, 304 340.

Edmands S (2002) Does parental divergence predict reproductive compatibility? *Trends in Ecology & Evolution*, **17**, 520 527.

Edwards AWF (1972) *Likelihood*. Cambridge University Press, Cambridge.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**, 435 445.

Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, New Jersey.

Facon B, Pointier J P, Glaubrecht M *et al.* (2003) A molecular phylogeography approach to biological invasions of the New World by parthenogenetic Thiarid snails. *Molecular Ecology*, **12**, 3027 3039.

Fisher RA (1937) The wave of advance of advantageous genes. *Annals of Eugenics*, **7**, 355 369.

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

Gray ME, Sappington TW, Miller NJ, Moeser J, Bohn MO (2009) Adaptation and invasiveness of western corn rootworm: intensifying research on a worsening pest. *Annual Review of Entomology*, **54**, 303 321.

Gregory P (2013) *"Markov Chain Monte Carlo Simulation Using the Metropolis Algorithm" from the Wolfram Demonstrations Project*: http://demonstrations.wolfram.com/MarkovChainMonteCarloSimulationUsingTheMetropolisAlgorithm/

Hill RE (1975) Mating, oviposition patterns, fecundity and longevity of the western corn rootworm. *Journal of Economic Entomology*, **68**, 311 315.

Hufbauer RA, Sforza R (2008) Multiple introductions of two invasive Centaurea taxa inferred from cpDNA haplotypes. *Diversity and Distributions*, **14**, 252 261.

Johnson ML, Gaines MS (1990) Evolution of dispersal: theoretical models and empirical tests using birds and mammals. *Annual Review of Ecology and Systematics*, **21**, 449 480.

Kim KS, Sappington TW (2004) Boll weevil (Anthonomus grandis Boheman)(Coleoptera: Curculionidae) dispersal in the southern United States: evidence from mitochondrial DNA variation. *Environmental Entomology*, **33**, 457 470.

Kirkpatrick M, Barton NH (1997) Evolution of a species' range. *American Naturalist*, **150**, 1 23.

Kolbe JJ, Glor RE, Schettino LRG *et al.* (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177 181.

Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences USA*, **104**, 3883 3888.

Lebreton JD, Burnham KP, Clobert J, Anderson DR (1992) Modeling survival and testing biological hypotheses using marked animals a unified approach with case studies. *Ecological Monographs*, **62**, 67 118.

Lenormand T, Raymond M (2000) Analysis of clines with variable selection and variable migration. *American Naturalist*, **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.

Liebhold AM, Tobin PC (2008) Population ecology of insect invasions and their management. *Annual Review of Entomology*, **53**, 387 408.

Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution*, **20**, 223 228.

Lockwood JL, Hoopes MF, Marchetti MP (2007) *Invasion Ecology*. Blackwell, Malden, MA.

Lodge DM, Shrader Frechette K (2003) Nonindigenous species: ecological explanation, environmental ethics, and public policy. *Conservation Biology*, **17**, 31 37.

Lombaert E, Guillemaud T, Cornuet JM *et al.* (2010) Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLoS ONE*, **5**, e9743.

Lubina JA, Simon AL (1988) The spread of a reinvading species: range expansion in the California sea otter. *The American Naturalist*, **131**, 526 543.

Lucek KAY, Roy D, Bezault E, Sivasundar A, Seehausen OLE (2010) Hybridization between distant lineages increases adaptive variation during a biological invasion: stickleback in Switzerland. *Molecular Ecology*, **19**, 3995 4011.

Mack RN, Simberloff D, Lonsdale WM *et al.* (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, **10**, 689 710.

Mallet J, Barton N, Lamas G *et al.* (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics*, **124**, 921 936.

McKinney ML, Lockwood JL (1999) Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends in Ecology & Evolution*, **14**, 450 453.

Miller N, Estoup A, Toepfer S *et al.* (2005) Multiple transatlantic introductions of the western corn rootworm. *Science*, **310**, 992.

Miller NJ, Ciosi M, Sappington TW *et al.* (2007) Genome scan of *Diabrotica virgifera virgifera* for genetic variation associated with crop rotation tolerance. *Journal of Applied Entomology*, **131**, 378 385.

Olden JD, LeRoy Poff N, Douglas MR, Douglas ME, Fausch KD (2004) Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology & Evolution*, **19**, 18 24.

Orr HA, Turelli M (2001) The evolution of postzygotic isolation: accumulating Dobzhansky Muller incompatibilities. *Evolution*, **55**, 1085 1094.

Peacock MM (1997) Determining natal dispersal patterns in a population of North American pikas (*Ochotona princeps*) using direct mark resight and indirect genetic methods. *Behavioral Ecology*, **8**, 340 350.

Phillips BL, Brown GP, Webb JK, Shine R (2006) Invasion and the evolution of speed in toads. *Nature*, **439**, 803 803.

Pimentel D, McNair S, Janecka J *et al.* (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture Ecosystems & Environment*, **84**, 1 20.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raymond M, Rousset F (1995) Genepop (version. 1.2), a population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248 249.

Rehage JS, Sih A (2004) Dispersal behavior, boldness, and the link to invasiveness: a comparison of four Gambusia species. *Biological Invasions*, **6**, 379 391.

REX (2013) Heterogeneity of selection and the evolution of resistance. *Trends in Ecology & Evolution*, **28**, 110 118.

Rieux A, Lenormand T, Carlier J, Lapeyre de Bellaire L, Ravigné V (2013) Using neutral cline decay to estimate contemporary dispersal: a generic tool and its application to a major crop pathogen. *Ecology letters*, **16**, 721 730.

Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology & Evolution*, **22**, 454 464.

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 AA & Nichols JD), pp. 18 28. Oxford University Press, New York, New York.

Rousset F (2008) GENEPOP '007: a complete re implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103 106.

Ruiz GM, Rawlings TK, Dobbs FC *et al.* (2000) Global spread of microorganisms by ships Ballast water discharged from vessels harbours a cocktail of potential pathogens. *Nature*, **408**, 49 50.

Sax DF, Stachowicz JJ, Gaines SD (2005) *Species Invasions: Insights into Ecology, Evolution, and Biogeography*. Sinauer Associates, Sunderland, MA.

Shigesada N, Kawasaki K (1997) *Biological Invasions: Theory and Practice*. Oxford University Press, Oxford [etc.].

Shigesada N, Kawasaki K, Takeda Y (1995) Modeling stratified diffusion in biological invasions. *American Naturalist*, **146**, 229 251.

Singhal S, Moritz C (2012) Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic lineages in a rainforest lizard. *Evolution*, **66**, 1474 1489.

Sivcev I, Manojlovic B, Krnjajic S *et al.* (1994) Distribution and harmfulness of Diabrotica virgifera LeConte (Coleoptera, Chrysomelidae), a new maize pest in Yugoslavia. *Zastita Bil ja*, **45**, 19 26.

Skellam JG (1951) Random dispersal in theoretical populations. *Biometrika*, **38**, 196 218.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787 792.

Sokal RR, Rohlf FJ (1995) *Biometry. The Principles and Practice of Statistics in Biological Research*, 3rd edn. W.H. Freeman and Company, New York.

Spencer JL, Hibbard BE, Moeser J, Onstad DW (2009) Behaviour and ecology of the western corn rootworm (Diabrotica virgifera virgifera LeConte). *Agricultural and Forest Entomology*, **11**, 9 27.

Sutherland GR, Richards RI (1995) Simple tandem DNA repeats and human genetic disease. *Proceedings of the National Academy of Sciences USA*, **92**, 3636 3641.

Szalai M, Papp Komáromi J, Bažok R *et al.* (2011) Generational growth rate estimates of *Diabrotica virgifera virgifera* populations (Coleoptera: Chrysomelidae). *Journal of Pest Science*, **84**, 133 142.

Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire bellied toads, *Bombina bombina* and *Bombina variegata*, near Cracow in southern Poland. *Evolution*, **40**, 1141 1159.

Urbanelli S, Bellini R, Carrieri M, Sallicandro P, Celli G (2000) Population structure of *Aedes albopictus* (Skuse): the mosquito which is colonizing Mediterranean countries. *Heredity*, **84**, 331 337.

Watkins WS, Bamshad M, Jorde LB (1995) Population genetics of trinucleotide repeat polymorphisms. *Human Molecular Genetics*, **4**, 1485 1491.

Wilson JRU, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM (2009) Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology & Evolution*, **24**, 136 144.

Wolfram Research Inc (2010) Mathematica Edition: Version 8.0. Wolfram Research, Inc., Champaign, Illinois.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114.

Zhong N, Ye LL, Dobkin C, Brown WT (1994) Fragile X founder chromosome effects linkage disequilibrium or microsat

ellite heterogeneity. *American Journal of Medical Genetics*, **51**, 405 411.

Data accessibility

Sample locations, microsatellite data, Mathematica notebook and R script: DRYAD entry doi:10.5061/dryad. 10kd4.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1. Construction of dispersal distributions from mixtures of Gaussian or exponential power family kernels.