REVIEW AND SYNTHESIS

A practical guide to measuring local adaptation

François Blanquart,[1]* Oliver Kaltz,[2] Scott L. Nuismer[3] and Sylvain Gandon[1]

Abstract

Patterns of local adaptation are expected to emerge when selection is spatially heterogeneous and sufficiently strong relative to the action of other evolutionary forces. The observation of local adaptation thus provides important insight into evolutionary processes and the adaptive divergence of populations. The detection of local adaptation, however, suffers from several conceptual, statistical and methodological issues. Here, we provide practical recommendations regarding (1) the definition of local adaptation, (2) the analysis of transplant experiments and (3) the optimisation of the experimental design of local adaptation studies. Together, these recommendations provide a unified approach for measuring local adaptation and understanding the adaptive divergence of populations in a wide range of biological systems.

INTRODUCTION

Adaptation to local environmental conditions plays a fundamental role in the generation and maintenance of biodiversity (Levene 1953; Gavrilets 2003), the contraction and expansion of species geographical ranges (Kirkpatrick & Barton 1997) and the ecological and evolutionary dynamics of species interactions (Kaltz & Shykoff 1998). In addition, because local adaptation measures the match between adaptive genetic variation and environmental variation, its magnitude depends on the interaction among evolutionary forces such as selection and gene flow (Kawecki & Ebert 2004) such that quantifying levels of local adaptation may provide important insight into the relative strengths of these evolutionary forces. Because of its many applications, numerous experimental approaches have been developed to estimate the amount of local adaptation in natural populations. Despite a long history of empirical investigation, however, little consensus exists regarding the best methodology for measuring local adaptation.

In cases where adaptive traits have been identified, several methods can be used to test for local adaptation. For instance, adaptation to local environmental conditions can be inferred if greater divergence is observed for a candidate trait than can be explained by drift alone (Hendry *et al.* 2001; McKay & Latta 2002). Alternatively, support for local adaptation could be provided by evidence of strong correlations between a candidate trait and functionally relevant environmental variables (Fumagalli *et al.* 2011; Hancock *et al.* 2011). Although potentially insightful, such methods are limited in scope because they depend on the existence of a modest number of relevant and measurable fitness traits and cannot quantify the full extent of local adaptation, which may be the cumulative result of small contributions by myriad underlying unknown phenotypic traits.

More generally, local adaptation can be estimated by measuring the fitness of populations in their own habitat and when transplanted to other habitats. The history of such transplant experiments can be traced back to studies of spatial phenotypic variation in plants (von Marilaun 1895, Bonnier 1895). It was Turesson (1922), however, who first used these experiments to demonstrate that the spatial heterogeneity in populations' traits was adaptive and genetically based. He coined the term 'ecotype' to describe a genetically distinct population adapted to specific environmental conditions. Following Turesson's pioneering work, a large number of studies have estimated local adaptation and identified factors that shape its magnitude in a broad range of organisms (Bradshaw 1960; Edmunds & Alstad 1978; Hoeksema & Forde 2008; Leimu & Fischer 2008; Hereford 2009).

Although most local adaptation studies based on transplant experiments employ similar experimental designs, various methods are used to calculate local adaptation and evaluate its significance. Consequently, it is challenging to generate meaningful comparisons across different studies. In a seminal review on the subject, Kawecki & Ebert (2004) raised important issues regarding the appropriate definitions and criteria for the detection of local adaptation in transplant experiments. In this article, we provide a formal treatment of these conceptual questions, using analytical and evolutionary simulation approaches. We attempt to standardise measures of local adaptation and to establish practical recommendations for the design and analysis of transplant experiments.

An appropriate measure of local adaptation should quantify the proportion of spatial variation in mean fitness caused by adaptation to local conditions. Thus, local adaptation is not a property of a single population, but rather of a metapopulation where multiple environments and populations are sampled. In this article, we focus mainly on a class of methods that defines local adaptation within a

[1]*Centre d'Ecologie Fonctionnelle et Evolutive, Unité Mixte de Recherche 5175, 1919 route de Mende, 34293, Montpellier Cedex 5, France*

[2]*Institut des Sciences de l'Evolution (ISEM), UMR 5554 (CC065), Université Montpellier 2, Place Eugène Bataillon, Montpellier Cedex, 34095, France*

[3]*Department of Biological Sciences, University of Idaho, Moscow, ID, 83844, USA*

Correspondence: E-mail: francois.blanquart@normalesup.org

metapopulation as the difference between the fitness of populations on their home sites (in sympatry) and fitness of populations when transplanted to other sites (in allopatry). This 'sympatric vs. allopatric' contrast has the appealing property of quantifying the extent to which the genotypic composition of individual populations matches their local environmental conditions. Throughout the text, we will explore how this approach based on the sympatric vs. allopatric contrast differs from the more stringent criterion advocated by Kawecki & Ebert (2004), which considers a metapopulation locally adapted only if the fitness of each population at its local site is superior to the average fitness of foreign populations transplanted to this site.

We begin by reviewing theoretical work on the processes that generate a pattern of local adaptation in spatially heterogeneous environments. Theoretical studies have the luxury of calculating local adaptation by averaging over a very large number of simulated generations and/or populations. In contrast, experimental studies are often limited to sampling a handful of populations at only a single point in time. In these cases, we show that different tests of local adaptation may yield qualitatively different conclusions on the existence of local adaptation. We clarify the link between these different tests and make statistical recommendation for the analysis of transplant experiments. Finally, we provide practical guidelines to optimise the experimental design in order to detect and quantify local adaptation.

PROCESSES GENERATING LOCAL ADAPTATION

Local adaptation results from the interaction between multiple evolutionary forces (e.g. selection, genetic drift, mutation, migration) and has been the focus of many theoretical studies. Here, we briefly review this literature to better grasp the interplay between these different evolutionary forces (see Kawecki & Ebert 2004 for a more thorough review). A key prerequisite for the emergence of local adaptation is the existence of a spatially heterogeneous environment generating a heterogeneous selective pressure. Whether local adaptation actually evolves, and to what extent, however, depends on the balance among different evolutionary forces. First, local adaptation is sensitive to the balance between gene flow and local selection (Levene 1953; Nagylaki 1980, Gavrilets & Gibson 2002; Whitlock & Gomulkiewicz 2005; Yeaman & Otto 2011; Blanquart *et al.* 2012). When gene flow is limited, specialised genotypes can be maintained in isolated populations and this favours local adaptation. When gene flow is very large, however, the genotype that is, on average, the best, invades the population and local adaptation vanishes ('gene swamping', Lenormand 2002). Second, the amount of genetic drift may also act on local adaptation. Genetic drift is expected to reduce local adaptation by reducing additive genetic variance, and by causing the random fixation of a reduced number of genotypes (Yeaman & Otto 2011; Blanquart *et al.* 2012).

Interestingly, some of the above predictions may be modified in a temporally variable environment. In particular, local adaptation may be maximal for intermediate levels of gene flow (Gandon 2002; Blanquart & Gandon 2011). If selection pressures change, gene flow can facilitate adaptation by augmenting local genetic variation. Rapidly changing selective pressures are often found in the context of antagonistic interactions, where coevolution results in 'arms race' dynamics with constantly escalating traits in both partners, or 'Red Queen' dynamics with periodical fluctuations in allele frequencies. In such systems, it is expected that the most rapidly evolving part-

ner of the interaction is locally adapted while the other is not. Local adaptation experiments are thus frequently used as a practical tool to detect which partner is adapted to the other, which is indicative of a shorter generation time or higher mutation or migration rate (Gandon 2002).

To illustrate how these processes influence local adaptation, we developed evolutionary simulations of a metapopulation evolving in response to spatially heterogeneous selection, gene flow, mutation and genetic drift (details in Appendix A). We use these evolutionary simulations to show how the average value of local adaptation changes as a function of the migration rate, in a constant or in a temporally changing environment (Fig. 1). Local adaptation is calculated as the difference between fitness of populations in sympatry and the fitness of populations in allopatry, averaged over all populations in the metapopulation and over time. Local adaptation is present when fitness is higher in sympatry than in allopatry. These simulations illustrate the erosion of local adaptation by migration in a constant environment, and the non-monotonous relationship between local adaptation and migration in a periodically fluctuating environment.

In summary, the evolutionary forces responsible for shaping the magnitude of local adaptation are well understood within the context of theoretical studies capable of sampling all populations within the metapopulation at multiple points in time. This theoretical framework does not, however, address the more practical issue of estimating local adaptation and testing significance when only a finite number of populations can be sampled from the entire metapopulation. Within this context, different definitions of local adapta-
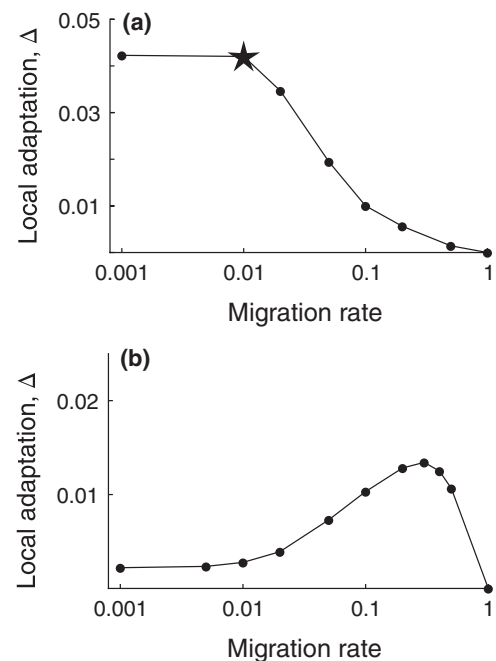

Figure 1 Local adaptation in the evolutionary simulations (described in Appendix A) as a function of the migration rate. Each dot is the difference between fitness in sympatry and in allopatry averaged over the whole metapopulation and the 500 final generations. Top panel: the environment is constant. The star identifies the expected level of local adaptation at the metapopulation level for the parameters chosen for the robustness tests used in parts 3.4 and 4.3 (Figs 4 and 6). Bottom panel: the environment is periodically changing with period $T = 10$ generations, such that an intermediate migration rate maximises local adaptation.

tion have been proposed and applied. In the following sections, we show that these different definitions quantify very different evolutionary phenomena and explain why this is the case.

CONCEPTUAL ISSUES IN LOCAL ADAPTATION

Empirical measures of local adaptation

The most straightforward approach to estimating local adaptation in a metapopulation is to calculate the difference between the average fitness in sympatric combinations of populations and sites and the average fitness in allopatric combinations (the SA contrast). As explained in the introduction, this is typically the measure used in most studies to analyse the effect of a specific factor (e.g. the effect of migration in Fig. 1). Yet, in practice, a transplant experiment taking this approach yields only a single value of local adaptation. Clearly, any significance testing requires estimating the amount of variation around this average measure. A natural way to proceed is to define a measure of local adaptation for each population and to estimate the amount of variation among populations. Two popular definitions of local adaptation at the population level lend themselves naturally to such an approach (reviewed in Kawecki & Ebert 2004). First, the 'home vs. away' (HA) definition is calculated as the mean fitness of the population at home minus the average mean fitness of the population when transplanted in all other habitats. Second, the 'local vs. foreign' (LF) definition is calculated as the mean fitness of a focal population at home minus the average mean fitness of all other populations when transplanted into the focal

patch. Both definitions provide estimates of local adaptation for each population within the metapopulation and thus some estimate of variation in levels of local adaptation. These two definitions can be expressed mathematically as follows:

$$\Delta_{HA}(i) = \bar{W}_{i \to i} - \frac{1}{P-1} \sum_{\substack{j=1 \\ j \neq i}}^{P} \bar{W}_{i \to j}$$

$$\Delta_{LF}(i) = \bar{W}_{i \to i} - \frac{1}{P-1} \sum_{\substack{j=1 \\ j \neq i}}^{P} \bar{W}_{j \to i}$$

(1)

where each $\bar{W}_{i \to j}$ is the mean fitness of population i transplanted into the habitat of population j; e.g. in a full factorial experimental design with P populations, experimentalists measure $P \times P$ mean fitnesses.

In Fig. 2, we present different hypothetical outcomes generated by the simplest possible local adaptation experiment, with a 2×2 population matrix and discuss definitions of local adaptation at the population level. As discussed previously by Kawecki & Ebert (2004), with such diverse patterns the two definitions of local adaptation can lead to contrasting conclusions. For instance, in Fig. 2a, the two definitions agree and show that there is local adaptation. In contrast, in Fig. 2b–d, the estimates of local adaptation differ quantitatively (the two definitions do not lead to the same magnitude of local adaptation) and qualitatively (the sign of local adaptation may differ between the two definitions in panels b and c, populations 2).

In spite of these different patterns, the average local adaptation over the two populations is the same in all panels of Fig. 2, regardless of what definition of local adaptation is used (HA or LF). The
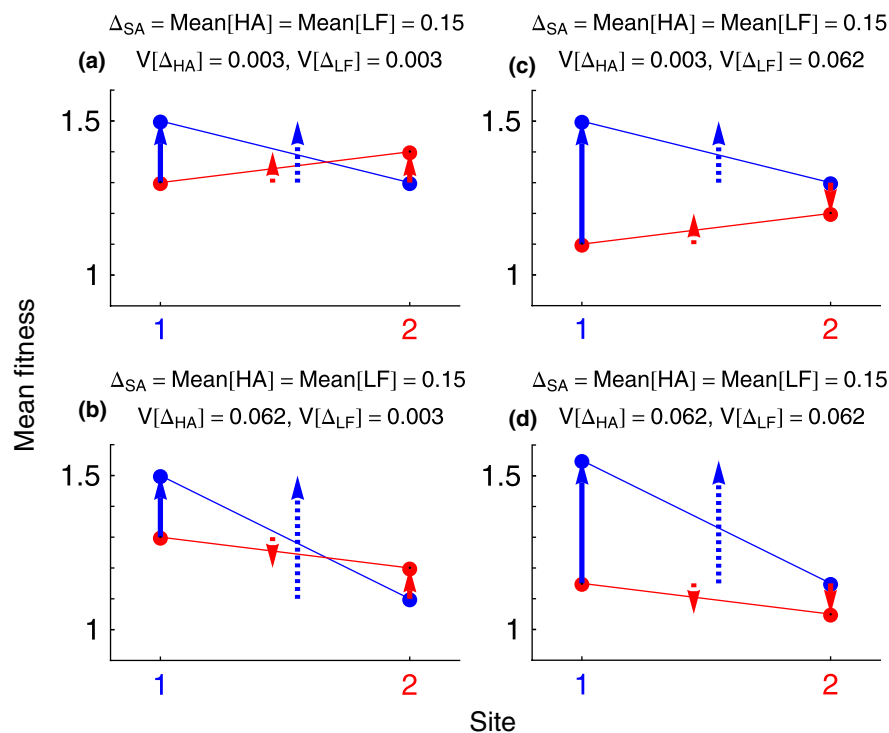

Figure 2 Mean fitness of populations is shown as a function of the site they are transplanted to in four hypothetical and heuristic scenarios. Mean fitness of population 1 (originating from site 1) is shown in blue and mean fitness of population 2 (originating from site 2) in red. The arrows show the local adaptation value for each population according to the 'home vs. away' (dashed arrow) or to the 'local vs. foreign' (plain arrow) definitions. The direction of the arrow indicates positive (upward) or negative (downward) value of local adaptation. Δ: average local adaptation at the scale of the 2 by 2 metapopulation (SA contrast). V[]: variance of local adaptation across populations. (adapted from Kawecki & Ebert 2004).

expected level of local adaptation is simply equal to the SA contrast mentioned above, i.e. the difference between the average fitness in sympatric combinations and the average fitness in allopatric combinations (Morgan *et al.* 2005; Appendix B). This average has the appealing property of quantifying the extent to which genotypic composition 'fit' local environmental conditions (Nuismer & Gandon 2008; Blanquart *et al.* 2012). This latter property can be formalised by assuming that each deme is characterised by the habitat quality and each individual by a genotype belonging to a discrete set of possible genotypes. The fitness of an individual of genotype k in deme i is defined as $w_{k,i} = g_k + h_i + \alpha_{k,i}$ where g_k represents the quality of genotype k, h_i represents the quality of habitat i, and $\alpha_{k,i}$ is the interaction between genotype and environment. As experimenters often seek to minimise the impact of phenotypic plasticity and maternal effects on the results of transplant experiments, we neglect environment-of-origin effects in this model. The frequency of genotype k in deme i is noted $X_{k,i}$. In such a framework, the average local adaptation in the metapopulation is (Appendix C):

$$\Delta_{SA} = E[\Delta_{HA}(i)] = E[\Delta_{LF}(i)] = \frac{P}{P-1} \sum_{k=1}^{n_{geno}} \text{Cov}[\alpha_k, X_k] \qquad (2)$$

where $E[]$ is the average over all demes of the metapopulation, $\text{Cov}[]$ is the spatial covariance, X_k (α_k respectively) is the variable taking value $X_{i,k}$ ($\alpha_{i,k}$ respectively) in deme i (Appendix C). The covariance in eqn (2) quantifies the matching between the environments and the genotypic frequencies. The $P/(P-1)$ factor emerges because we consider 'vs. away' measures, and would disappear were we to consider 'vs. global' measures (Nuismer & Gandon 2008).

Although the expected value of local adaptation is the same under both definitions, the *variance* across populations may be very different under the HA and the LF definitions (Fig. 2b–d). Consequently, tests of statistical significance may yield different results depending on the definition used (Morgan *et al.* 2005; Vogwill *et al.* 2010). This situation calls for a conceptual clarification of the diversity of outcomes illustrated in Fig. 2.

Home vs. away local adaptation

We first elaborate on the model presented in 'Empirical measures of local adaptation' to better understand the underlying factors generating the diversity of patterns in Fig. 2. Using this framework (Appendix C), the HA definition for a single population i is

$$\Delta_{HA}(i) = \frac{P}{P-1} \left[\underbrace{h_i - E[h]}_{\text{Habitat effect}} + \underbrace{\sum_{k=1}^{n_{geno}} X_{k,i}(\alpha_{k,i} - E[\alpha_k])}_{\text{Local adaptation}} \right] \qquad (3)$$

where the operator $E[]$ is the average over all demes. HA local adaptation of deme i depends on the intrinsic quality of habitat i relative to the average habitat quality $h_i - E[h]$. Indeed, if habitat i happens to be particularly good (e.g. because of a rich environment), local adaptation will be strong because fitness away will be on average lower. The second term, $\sum_{k=1}^{n_{geno}} X_{k,i}(\alpha_{k,i} - E[\alpha_k])$ reflects the adaptation of population i to its specific habitat conditions (the $G \times E$ interaction for fitness). A single population has positive HA local adaptation, if it resides on a particularly favourable site ($h_i - E[h] > 0$)) or if this population is adapted to the particular ecological conditions of its

local site ($\sum_{k=1}^{n_{geno}} X_{k,i}(\alpha_{k,i} - E[\alpha_k]) > 0$). For instance, in Fig. 2b, population 1 appears strongly locally adapted when the HA definition of local adaptation is used because habitat 1 has intrinsically higher quality than habitat 2.

Local vs. foreign local adaptation

In the terms of the model we developed above, the LF definition for a single population i is

$$\Delta_{LF}(i) = \frac{P}{P-1} \left[\underbrace{\sum_{k=1}^{n_{geno}} g_{k,i}(X_{k,i} - E[X_k])}_{\text{Deme quality effect}} + \underbrace{\sum_{k=1}^{n_{geno}} \alpha_{k,i}(X_{k,i} - E[X_k])}_{\text{Local adaptation}} \right] \qquad (4)$$

LF local adaptation of deme i depends on the genetic quality of individuals in deme i relative to the average individuals, $\sum_{k=1}^{n_{geno}} g_{k,i}(X_{k,i} - E[X_k])$. Indeed, if genotypes of individuals in deme i happen to be of high quality, LF local adaptation will be strong because the fitness of foreign populations will be on average lower. The second term, $\sum_{k=1}^{n_{geno}} \alpha_{k,i}(X_{k,i} - E[X_k])$ reflects the adaptation of population i to its specific habitat conditions relative to the average populations (the $G \times E$ interaction for fitness). A single population has positive LF local adaptation, if it is of high genetic quality $\left(\sum_{k=1}^{n_{geno}} g_{k,i}(X_{k,i} - E[X_k]) > 0 \right)$ or if this population is adapted to the particular ecological conditions of its local site $\left(\sum_{k=1}^{n_{geno}} \alpha_{k,i}(X_{k,i} - E[X_k]) > 0 \right)$. For instance, in Fig. 2c, population 1 appears strongly locally adapted when the LF definition of local adaptation is used because population 2 is of lower quality than population 1. These genetic differences in fitness which are independent of the habitat may be due, e.g. to different levels of inbreeding in the two populations or, in general, to differences in fitness caused by different evolutionary histories of the populations.

In summary, population-level measures such as HA and LF depend on the interaction between genotypes and local environments (which is the very essence of local adaptation after all), but they also include a term that has nothing to do with the covariation between the distribution of adaptive genetic variation and the environment but rather measures the overall habitat quality (with HA definition) or deme quality (with LF definition) of the focal population. As a consequence, the variance across populations of the HA and LF may be wildly different (Fig. 2) because the variance of HA is partly explained by habitat effects, while the variance of LF is partly explained by deme quality effects. In contrast, the metapopulation-level SA contrast estimates the quantity of interest with no contributions made by confounding factors such as habitat or deme quality.

Measures of local adaptation in evolutionary simulations

In this section, we use the evolutionary simulations (see section "Processes generating local adaptation") to illustrate the essential findings from the above analytical approach. In particular, we show how overall differences in fitness due to genetic or environmental effects can produce diverse outcomes for measurements of local adaptation in a transplant experiment. To this end, we use the final

generation of these simulations to estimate local adaptation for the case where the migration rate is m = 0.01, which generates an average local adaptation of approximately 0.04 (Fig. 1, Appendix A). The metapopulation is composed of 100 demes such that 4950 distinct pairs of populations may be sampled.

In the evolutionary simulations, the environmental quality differs across demes, which increases variation in HA measures among the different populations. Similarly, the genetic quality of populations differs across demes, which increases variation in LF measures. These deme quality effects are caused by stochastic fixation of deleterious mutations in some demes. Thus, as mentioned above, some of the variability in HA and LF measures is independent of the average measure of local adaptation at the metapopulation level. To illustrate this point, we generated 'null models' allowing only genetic variability (null model 1) or only habitat variability (null model 2). These models do not generate any pattern of local adaptation at the metapopulation level (i.e. no SA contrast). However, variation is generated for the LF contrast when there is variation of deme quality across populations, and, similarly, variation is generated for the HA contrast when there is variation in habitat quality across populations (Fig. S2). It is important to note that although habitat effects are constant and may be large, deme quality differences across populations are evolving and are expected to be eroded by natural selection. For example, in Fig. S1, although we allowed for potentially large fitness differences across the different genotypes (Appendix A), the among-population variation of local adaptation is higher with the HA definition than with the LF definition.

Furthermore, when subsampling the metapopulation for a transplant experiment, we are also likely to obtain different estimates for the mean level of local adaptation (SA) across different samples (Fig. S1). It is thus important to keep in mind that in transplant experiments, sampling a very limited number of populations may not necessarily give a representative picture of the true level of local adaptation in the metapopulation. Clearly, deviations from the true picture will diminish with an increasing number of populations used in the transplant experiment.

So far, we have focused on the *average* sympatric vs. allopatric measure of local adaptation. We have shown that this average summarises a diversity of underlying patterns and explained what processes generate these patterns. Another way of summarising this diversity into a single metapopulation-level criterion for local adaptation, however, was proposed by Kawecki & Ebert (2004). They suggest that a metapopulation is locally adapted only when *all populations* are locally adapted following a local vs. foreign criterion, as shown in Fig. 2a,b only. Equation (4) makes clear that a metapopulation is likely to be locally adapted according to this criterion only in situations where deme quality effects are very reduced. Moreover, the criterion becomes increasingly stringent as the number of sampled populations increases. Thus, all else being equal, those experiments which sample a greater number of populations are less likely to detect local adaptation than those which sample only a small number of populations. Put differently, this means the criterion proposed by Kawecki & Ebert (2004) has the very undesirable property of actually causing statistical power to decline as the sample size (number of populations) is increased. In contrast, the sympatric vs. allopatric measure has a clear conceptual interpretation (eqn 2) and does not depend on effects not related to local adaptation, such as deme and habitat quality effects. Consequently, we propose that a

metapopulation is locally adapted if the sympatric vs. allopatric difference is greater than 0. In the following, we discuss different ways of testing for the significance of the sympatric vs. allopatric difference by comparing the average value of this measure to its underlying variability.

STATISTICAL ISSUES IN LOCAL ADAPTATION

Three statistical tests for local adaptation

In experimental work, statistical tests are very often based on the HA or LF definitions. As both definitions do not yield the same variability across populations, it is not clear which one should be used. In their review, Kawecki & Ebert (2004) argue that the LF definition should be 'regarded as diagnostic for the pattern of local adaptation' because it is more 'relevant to the driving force of natural selection – divergent natural selection – which acts on genetic differences in relative fitness within each habitat'. This recommendation has been very influential and led to widespread use of the LF definition of local adaptation (Morgan *et al.* 2005; Greischar & Koskella 2007; Sicard *et al.* 2007; Becker *et al.* 2008; Leimu & Fischer 2008; Hereford 2009; Schulte *et al.* 2011; Garrido *et al.* 2012). However, a large number of studies have also used a third statistical test, the 'sympatric vs. allopatric' (SA) test (Turkington & Harper 1979; Bever 1994; Kaltz *et al.* 1999; Mutikainen *et al.* 2000; Lajeunesse & Forbes 2002; Thrall *et al.* 2002; Lively *et al.* 2004; Ganz & Washburn 2006; Greischar & Koskella 2007; Hoeksema & Forde 2008; Adiba *et al.* 2010; Franceschi *et al.* 2010; Vogwill *et al.* 2010; Vogwill *et al.* 2010). This test is not based on the variability of a population-level measure, but rather on the residual variability that remains once deme and habitat effects are accounted for. In the following, we evaluate the statistical properties of these candidate definitions and their respective statistical power.

In a first step, we use classical linear models. These models have the advantage of being commonly used and relatively simple. We will discuss later how more complex statistical models can be used to deal with more realistic situations. To test for local adaptation using the first two definitions, we test whether the mean of the distribution of LF estimates [the $\Delta_{LF}(i)$] or HA estimates [the $\Delta_{HA}(i)$] is significantly different from 0. Let us assume for the time being that the HA and LF measures are independent: we can thus use a two-sided, one sample *t*-test (actually, we will use the *F* statistic instead of the *t*, but the approaches are identical). To test for local adaptation directly using the sympatric vs. allopatric contrast, we test whether the means of two distributions (sympatry and allopatry) are significantly different from each other. To this end, we use the *F*-test corresponding to an ANOVA including habitat effects, deme quality effects and the sympatric vs. allopatric effect, and look at the significance of the sympatric vs. allopatric contrast. The null hypothesis we are seeking to reject is that the sympatric vs. allopatric difference does not explain significant variation in fitness in the transplant experiment. The *F* statistics quantify the amount of variation in the data attributable to the factor considered, relative to the amount of variation that cannot be attributed to this factor (Appendix D1). Assuming a full-factorial experimental design, the level of significance of the *F* statistics for the three tests, F_{HA}, F_{LF} and F_{SA} can be found using the *F*-distributions with degrees of freedom $(1, P - 1)$ for the first two tests, and $(1, P^2 - 2 (P - 1) - 2)$ for the third (where P is the number of populations used for the experiment).

Power of the three tests

To evaluate each test rigorously, we derive analytical predictions for their sensitivity (or power), which is the probability of detecting local adaptation when there is true local adaptation. Note that the specificity of the tests, which is the probability of a 'true negative' (local adaptation is not detected, and, indeed, there is no local adaptation) is in theory set to 0.95 in all three tests because the significance level is set at 0.05.

We derive the analytical predictions under the assumption that the performance of individual k of deme i transplanted into deme j can be modeled as:

$$W_{i \to j,k} = constant + \gamma_i + \psi_j + \delta_{i,j}\alpha + \epsilon_{i,j} + \epsilon_{i,j,k} \qquad (5a)$$

where γ_i is the deme effect of population i, ψ_j is the habitat effect of environment j, $\delta_{i,j}$, is an indicator variable which is 1 if $i = j$ and 0 otherwise, α is the magnitude of the fitness advantage of being in sympatry relative to allopatry, ϵ_{ij} accounts for variation in the $G \times E$ interaction at the level of the population (with variance $\sigma^2_{\epsilon,pop}$), and $\epsilon_{i,j,k}$ is an additional error term that accounts for individual-level error (with variance $\sigma^2_{\epsilon,ind}$), including both the experimental error and variability in the quality of individuals within deme j. In some cases, data at the level of individuals are not accessible, (e.g. experimental evolution using bacteria and phages), in which case we assume a simplified version of (5a), where the mean fitness of population i transplanted to deme j is given by the following:

$$\bar{W}_{i \to j} = constant + \gamma_i + \psi_j + \delta_{i,j}\alpha + \epsilon_{i,j} \qquad (5b)$$

and the notation is similar to that used in eqn (5a), except that ϵ_{ij} captures both variation in the population × environment interaction across pairs of populations ij, and random error. We assume that γ, ψ and ϵ are drawn from distributions with mean 0 and true variances σ^2_γ, σ^2_ψ and σ^2_ϵ respectively. Our goal is to assess statistical significance of local adaptation as measured by the difference between sympatric and allopatric fitness α. For simplicity, we derive analytical predictions for power in the population-level model (5b); we will come back to the individual-level model (5a) in section "Methodological issues in local adaptation". To match the assumptions of the linear model, we assume the error is normally distributed, and the habitat, deme quality, and sympatric vs. allopatric effects are independent.

Computing power requires the distribution of the statistic when there is true local adaptation, which, in our case, is the noncentral *F*-distribution. We found expressions for the power of the three tests under the model defined in eqn (5) (Appendix D2). These expressions bring four insights. First, as expected, the power of all three tests increases with the magnitude of local adaptation α and decreases with variance associated with the error σ^2_ϵ. Second, the sympatric vs. allopatric test is only affected by these two parameters, but the power of the HA and the LF tests decreases with the variance of the habitat and the deme effects respectively. Third, the HA test has greater power than the LF test if $\sigma^2_\psi < \sigma^2_\gamma$, but LF has greater power if the reverse is true. Fourth, the SA test always has higher power than the two other tests because it has more error degrees of freedom.

Numerical approach

We test the precision of our analytical results by numerically computing power using data corresponding to a fully factorial reciprocal
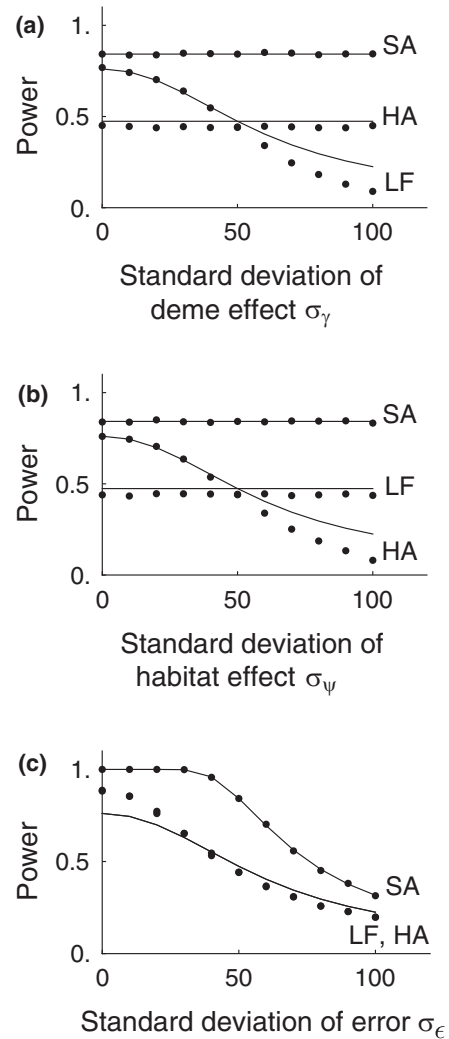

Figure 3 Power of HA, LF and SA tests as a function of the standard deviation in the deme quality effect σ_γ (a), of the habitat effect σ_ψ (b), and random error σ_ϵ (c). Dots show the power calculated numerically by generating 5000 replicates of the data set. Lines show the analytical result given by eqn (D6) (Appendix D2). In all graphs, $P = 10$, local adaptation is set to 50, and σ_γ, σ_ψ, and σ_ϵ are set to 50 (except when they vary along the x-axis).

transplant generated according to the model of eqn (5). We evaluate the power of each of the three tests as a function of the error standard deviation σ_ϵ, the variances of habitat and deme quality effects σ_ψ and σ_γ, and the number of sampled populations P. Simulation results confirm the conclusions drawn from the analytical model (Fig. 3). The analytical result [eqn (C9)] performs very well for the SA test, confirming that this test is the most powerful. But slight discrepancies between analytical prediction and simulation exist for the HA and LF tests. These discrepancies are greatest in the presence of large amounts of habitat or deme quality variance and arise because the data violates the assumption of independence across populations on which the *F*-test relies. Specifically, the measures of local adaptation are not independent across populations because deme effects or habitat effects are shared across all populations for LF and HA designs respectively [eqns 3 and 4, Appendix C]. In addition to influencing the accuracy of our analytical predictions for statistical power, the lack of independence among estimates of local

adaptation in studies taking the LF or HA definitions causes the specificity of these tests to be artificially inflated (Fig. S3). Consequently, using the LF or HA approach is too statistically conservative, making it likely that using these approaches will cause many cases of true local adaptation to be missed.

Although we made the specific assumptions that the deme and habitat effects are independent, and that the errors are normally distributed, it is possible to use more elaborate statistical tests to relax these assumptions. These alternative statistical approaches do not affect our basic conclusions. Tests based on mixed models, with habitat and deme quality effects modeled as random effects, are
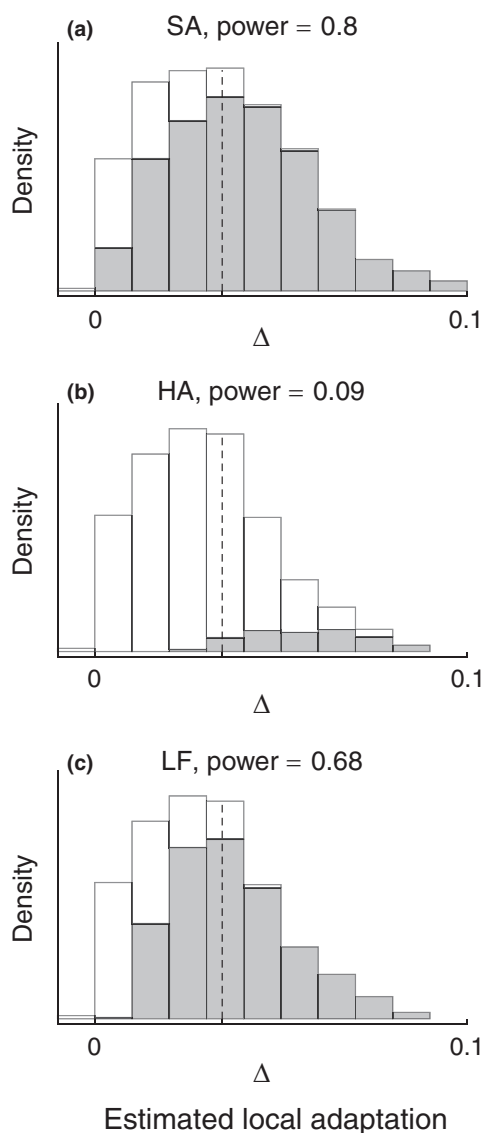

Figure 4 Comparison of the power of SA (panel a), HA (panel b) and LF (panel c) tests using the final generation of the evolutionary simulations described in Appendix A. Distribution of estimated local adaptation in the SA, HA and LF tests is shown over 1000 replicates of eight populations (full factorial experimental design) sampled out of the 100 populations of the metapopulation. The grey part of the histogram indicates the fraction of replicates that lead to rejection of the null hypothesis of no local adaptation (i.e. significantly positive local adaptation). True local adaptation (the SA contrast based on the whole metapopulation) and the power of the tests, defined as the fraction of replicates in which local adaptation was detected, are also indicated.

slightly more powerful than classical linear models, but this increased power comes at the cost of increased rate of false positive (Figs S4 and S5, Appendix E). A test based on a Bayesian method, where we sampled the posterior distribution of the parameters of the mixed model *with* local adaptation and recorded significant local adaptation when 0 was not in the 95% credible interval, performed poorly (Fig. S4, Appendix E). Thus, we found the classical linear model is actually the most relevant and the most straightforward approach to test for local adaptation.

Robustness of the methods

To evaluate the robustness of our recommendations in cases where the assumptions of classical linear models are violated, we applied the three tests HA, LF and SA to data generated with the evolutionary simulation model introduced in part 1 (described in Appendix A). In spite of possible violations of the assumptions of the linear model (non-homogeneity of variance, non-normal error), we found that the sympatric vs. allopatric test continued to outperform the two others (Fig. 4). It is noteworthy that the home vs. away test performs very poorly, and in particular much worse than the LF test. Once again, this is because habitat effects are quite strong in our simulations, while deme quality effects are relatively weak because they are generated by weak forces (mutation, drift) and constantly eroded by selection. The low power of HA and LF tests makes them often unable to detect local adaptation, except when it is quite strong (Fig. 4).

To summarise, the SA test assesses significance of the sympatric vs. allopatric measure while correctly accounting for the variability generated by habitat and deme quality effects, which makes it a more powerful test than tests based on HA and LF measures. In addition, the SA test also gives a clear picture of the processes shaping fitness in the metapopulation. In particular, the deme quality effects reflect fitness variation across populations which is independent of habitat heterogeneity (e.g. caused by drift), while the sympatric vs. allopatric contrast reflects the magnitude of divergent natural selection relative to migration and drift. These effects are entangled in the HA and LF definitions, which makes their interpretation more difficult.

As a corollary, it is worth pointing out that we do not recommend testing for local adaptation in a metapopulation using only two populations, as in the examples of Fig. 2. Indeed, with four measures of fitness there are not enough degrees of freedom to assess significance of the sympatric vs. allopatric contrast while accounting for habitat and deme quality effects. Therefore, any pattern of non-parallel reaction norms as observed in Fig. 2 could be generated by experimental error or some form of $G \times E$ interaction independent of divergent selection (Kawecki & Ebert 2004).

Up to this point, we have focused on the experimental design where fitness of populations was measured in a full-factorial design for all $P \times P$ transplants. Because of practical limitations, however, many incomplete experimental designs may actually be employed. In the next section, we examine how the experimental design can be optimised to maximise the power to detect local adaptation.

METHODOLOGICAL ISSUES IN LOCAL ADAPTATION

Optimisation of the experimental design

Transplanting populations and measuring fitness can be tedious and expensive. Consequently, identifying sampling strategies that maxi-
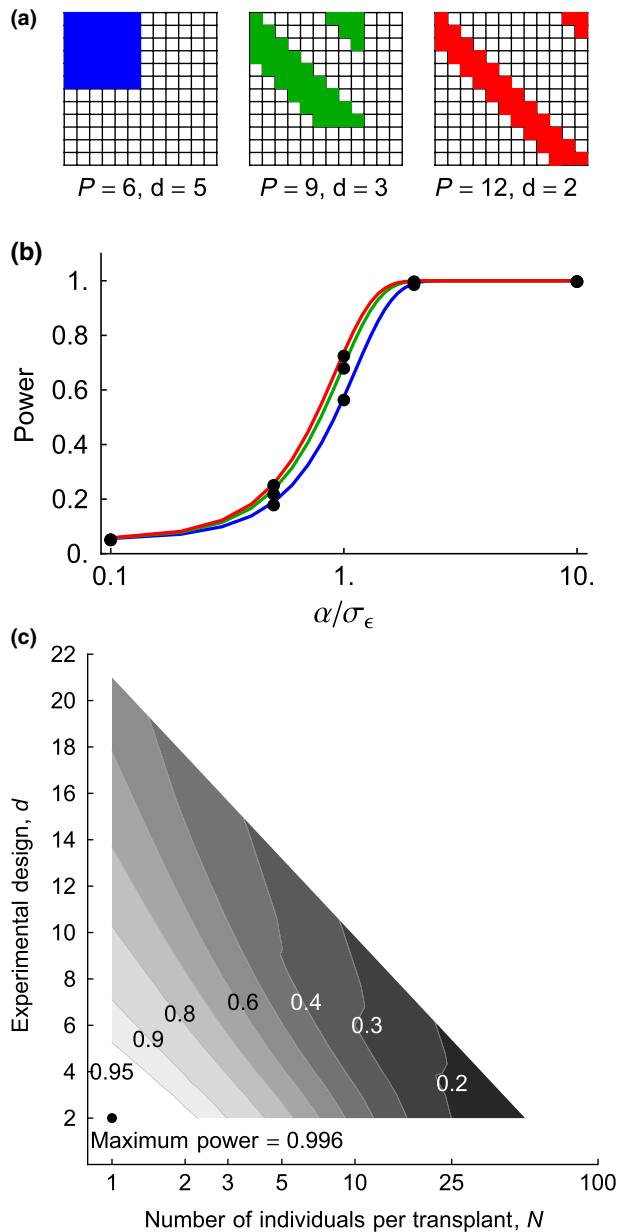
Figure 5 Power of various experimental designs. Panel a: experimental designs with six sympatric and 30 allopatric transplants (blue), nine sympatric and 27 allopatric transplants (green) and 12 sympatric and 24 allopatric transplants (red). Panel b: power for the three designs when no individual-level data is available, as a function of the ratio α/α_ϵ. The points show power as calculated numerically using 5000 replicates, and the lines are analytical results (Appendix D3). Panel c: power of experiments when individual-level data is available, with various number of individuals per transplant N and experimental designs d. Across all these designs, the total number of individuals tested is kept constant to $T = 500$. Local adaptation is $\alpha = 50$, individual-level error $\alpha_{ind} = 80$ and population-level error $\alpha_\epsilon = 80$.

Figure 6 Power of the SA test as a function of the experimental design, using the final generation of the evolutionary simulations described in Appendix A. Panel a: power of the $P/2P$ design (with twice as many allopatric transplants as sympatric transplants) (squares) and of the fully reciprocal design (circles) as a function of the number of transplants. Panel b: power of the $P/2P$ design when the number of individual per transplant and the number of transplants is varied, keeping the total number of individuals constant. 144 individuals are allocated in a variable number of populations from 8 (six individuals per transplant) to 48 (one individual per transplant). In both panels, power is calculated over 1000 random samples of populations randomly sampled in the metapopulation obtained with the evolutionary simulations.

the total number of transplants constant, as illustrated on Fig. 5a (comparing these designs make sense only when the number of populations is strictly greater than three). In addition, we may want to know how power is affected by different allocation strategies of individuals to transplants. Should we sample many individuals from only a small number of populations, or should we instead sample only a small number of individuals from a large number of populations?

With individual level data, testing for the presence of local adaptation with the SA test requires the use of a linear model where we fit habitat and deme quality effects, the SA effect, and the remainder of the interaction (the $\in_{ij}$). It is important to note that the SA effect must be tested against the remainder of the interaction and not against the individual error. This is because the population, and not the individual, is the relevant unit of replication when testing for local adaptation. We derive analytical results (and check those results with simulations) for the power of the SA statistical test as a function of the number of populations and the number of individuals assayed per population (Appendix D3). More specifically, we fixed the total number of transplants T and distributed this number over (1) the number of individuals used per population N, (2) the number of populations sampled P and (3) the number allopatric transplants Pd, with the constraint $T = (1 + d)PN$ (Appendix D3, Fig. 5c). Note that the perfectly balanced design with equal number of sympatric and allopatric transplants ($d = 1$) is not feasible if one wants to estimate habitat, deme quality and SA effects within the same statistical model.

mise power to detect local adaptation is essential. To this point, we have considered only a full-factorial design where the fitness of P sympatric and $P(P-1)$ allopatric transplants is measured. With this design, a total of P^2 transplants is required; a number which may be prohibitive in many systems for even a modest number of populations. For this reason, we study the consequences for power of varying the number of sympatric and allopatric transplants while holding
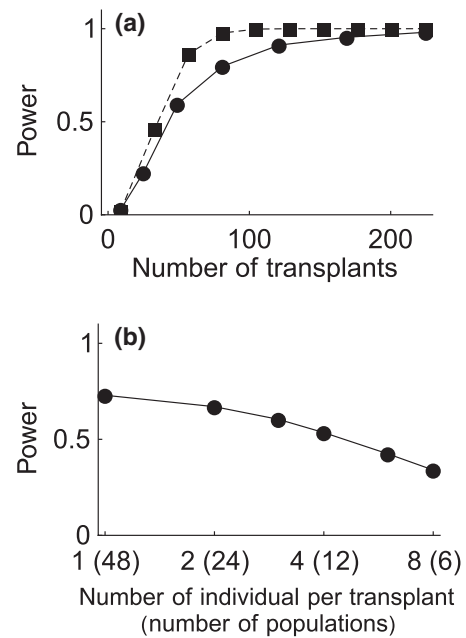
This analysis reveals that the most powerful design is always the one with fewer individuals sampled from a greater number of populations: $N = 1$, $d = 2$ and $P = T/3$ (Appendix D3; Fig. 5c). First, balancing the number of sympatric and allopatric transplants as much as possible ($d = 2$) increases power. Although these more balanced designs also have fewer error degrees of freedom (because the statistical model fits more habitat and deme quality effects), simulations reveal that the beneficial impact of balanced designs outweigh the loss of degrees of freedom (see Fig. 5c). Second, it is best to allocate only one individual per transplant ($N = 1$). This is explained by the strong advantage of having many populations sampled, which overcomes the disadvantages of having fewer degrees of freedom (more habitat and deme quality effects are estimated) and less precise estimates of the mean fitness of each transplanted population. In cases where no individual data is available, it is also best to sample P sympatric and $2P$ allopatric populations ($d = 2$). Interestingly, if a limited number of populations P is available for the experiment, it is of little interest in terms of power to increase the number of allopatric transplants from $2P$ to $P(P - 1)$ (Fig. S6). Moreover, in more realistic cases where local adaptation varies across populations (see Fig. S1), it is always better to increase the number of sampled populations P because the estimated local adaptation will also be closer to the true local adaptation of the whole metapopulation.

Robustness of the methods

To assess the robustness of our recommendations in more realistic settings, we tested the fully reciprocal design and the $P/2P$ design on the result of the evolutionary simulations (Fig. 6a). We found that the $P/2P$ design consistently performs better than the fully factorial design. Moreover, we also tested various sampling strategies in cases where individual-level data is available, and confirmed that the most powerful experimental design is the one which samples only one individual per population (Fig. 6b).

DISCUSSION

Local adaptation is the subject of an abundant theoretical and experimental literature. However, a rift exists between these two perspectives. Theoretical studies interested in the interplay between different evolutionary forces often focus on the average measure without discussing the underlying variability of the overall pattern. In contrast, the experimental literature must confront this variability to determine the statistical significance of the level of local adaptation obtained in a given transplant experiment. Here, we show that the estimation of this variability is very sensitive to the definition used to characterise local adaptation at the population level. Besides, a diversity of statistical and methodological approaches have been used in the past, such that a synthetic understanding of the strength and underlying causes of local adaptation has failed to emerge. Our study is an attempt to bridge the gap between these perspectives. We began with a review of the processes that contribute to local adaptation and used this review to formalise the statistical and methodological issues relevant to local adaptation studies. We hope to standardise estimation procedures in a way that will facilitate comparisons among studies and improve our ability to tease apart the causes of local adaptation using meta-analytic approaches.

Our formal investigation leads to three practical recommendations. First, local adaptation – defined as the component of the $G \times E$ interaction explained by the sympatric vs. allopatric contrast – is most clearly elucidated using a measure at the scale of the metapopulation. The reason for this is primarily that at least several populations and several habitats are necessary to reveal a $G \times E$ interaction in a statistical model. Second, a linear model describing the pattern of mean fitness as a sum of a habitat effect, a deme quality effect and a sympatric vs. allopatric effect is the most powerful and straightforward way to detect local adaptation. Such analysis gives insights into various processes that shape the fitness of the metapopulation. Specifically, the habitat effect reveals intrinsic differences in fitness due to environmental variations, the deme quality effects inform on processes that shape fitness regardless of environmental heterogeneity, such as inbreeding, while the sympatric vs. allopatric contrast informs on the strength of heterogeneous selection relative to other forces such as migration or drift. Third, as pointed out by Hoeksema & Forde (2008) unambiguous measures of local adaptation do not require a full factorial design but only that each population be measured both in sympatry and in allopatry (i.e. the reciprocal designs in Hoeksema & Forde 2008). Our analysis formalises this recommendation by comparing the statistical power among different reciprocal designs. If there are constraints on the number of individuals that can be transplanted, it is best to (1) assay twice as many allopatric transplants as sympatric transplants and (2) minimise the number of individuals sampled per population (i.e. $N = 1$). This design maximises statistical power because it maximises the number of populations sampled for the experiment. Hence, from a practical standpoint, one of our most important results is that a local adaptation experiment need not be fully reciprocal.

It is worth mentioning that alternative designs may be imagined. For example, it is possible to test for local adaptation by focusing on a set of independent sympatric vs. allopatric contrasts. These contrasts could be obtained as the average of sympatric fitnesses minus the average of allopatric fitnesses for a small 2×2 transplant matrix (e.g. 10 independent contrasts are obtained using the result of 40 transplants). In this particular design, it is possible to test for local adaptation using as many sympatric as allopatric transplants, but this comes at the cost of not estimating habitat or deme quality effects. A t-test may be used on these independent contrasts to test for significance of local adaptation. We calculated the power of this test and found that, in some cases (high number of transplants and weak local adaptation), it can perform better than the best SA test we propose above (Appendix D4, Fig. S7). Nevertheless, we recommend the SA test, because it estimates not only local adaptation but also the magnitude of habitat and deme quality effects (i.e. E and G main effects in the ANOVA), providing a better picture of the relative importance of the processes that shape adaptation of the metapopulation. Estimating habitat and deme quality effects is also extremely important because they elucidate other processes occurring within the metapopulation, such as habitat heterogeneity, genetic drift or deleterious mutations. Even if local adaptation is detected, it is possible that it explains little variation in population mean fitness, which may be primarily determined by the direct effect of the habitat or evolutionary processes other than adapting to local conditions. Thus, the linear model with habitat, deme quality and sympatric vs. allopatric effects gives a complete picture of the metapopulation and allows the relative importance of various evolutionary processes to be assessed.

In the linear model that we recommend, habitat and deme quality effects are supposed to be independent across populations. This lin-

ear model is most appropriate when the sampled populations are genetically isolated. However, our evolutionary simulations demonstrate that the linear model performs correctly even when populations are not strictly independent because of gene flow across populations. Non-independence of habitat or demes may be dealt with in several ways. Quantifying habitat effects may allow demes to be grouped in several types of habitats, e.g. polluted vs. unpolluted soil. In such cases, our practical recommendations may be adapted by modifying the structure of the habitat effect. For the sympatric vs. allopatric effect, the matter is slightly more complex. It may be modeled either as a 'same type of habitat' vs. 'different type of habitat' effect (i.e. some allopatric transplant may be classified as 'same type of habitat'). Alternatively, if other features of the habitat are suspected to be important for local adaptation, the sympatric vs. allopatric effect can be broken down into 'sympatric – same type', 'allopatric – same type', and 'allopatric – other type' (Adiba *et al.* 2010). In other words, total local adaptation is broken down into the part due to the identified environmental factor (e.g. soil pollution), given by the 'allopatric – same type' vs. 'allopatric – other type' contrast, and the part due to non-identified environmental factors given by the 'sympatric – same type' vs. 'allopatric – same type' contrast. Similar linear models may be used if populations may be grouped according to similarities in their genetic qualities (i.e. populations originating from different locations). Finally, it is also possible to take into account the distance between sites by partitioning the sympatric vs. allopatric contrast into 'sympatric vs. near allopatric' and 'sympatric vs. far allopatric', as in Adiba *et al.* (2010) (see also Kaltz *et al.* 1999).

In the context of host–parasite interactions, in particular, heterogeneities in the abiotic environment (i.e. selection mosaics) may affect the speed of coevolution and structure adaptation at different spatial scales (Thompson 1994, 2005; Gandon & Nuismer 2009). Each partner may adapt both to the abiotic and biotic environment. Nuismer & Gandon (2008) have shown that a local adaptation experiment in a common-garden context captures the part of local adaptation due to adaptation to the biotic environment only. In this case, there is no particular methodological problem, and our recommendations should hold. In contrast, a local adaptation experiment done with transplants (e.g. in the field) includes both adaptation to the biotic AND abiotic environments. Again, if there is no interaction between adaptation to biotic and abiotic environment, our recommendations may also readily apply. More complex situations may arise if the abiotic environment also conditions adaptation to the biotic environment (e.g. Lopez Pascua *et al.* 2012).

CONCLUSION

Among the various experimental designs mentioned in the introduction, transplant experiments provide the most complete picture of adaptive differentiation. Indeed, experiments based on trait-environment correlations or on differentiation of adaptive vs. neutral markers are likely not to sample exhaustively the set of relevant traits/loci. As adaptation may rely on myriad loci of very small effect (e.g. Fournier-Level *et al.* 2011), quantifying local adaptation with transplant experiments will likely remain a favoured method to understand adaptation in natural populations. The recommendations we make here regarding the design and the analysis of such experiments apply to a broad range of biological systems and may thus facilitate comparison among studies. This comparative approach is critical to investigating the relative importance of different evolutionary forces (selection, gene flow, genetic drift) on adaptive dynamics in spatially heterogeneous environments (Hoeksema & Forde 2008; Leimu & Fischer 2008; Hereford 2009).

AUTHORSHIP

FB, OK, SN, SG conceived the study. FB did the analysis and simulations; FB, OK, SN, SG wrote the manuscript.

REFERENCES

Adiba, S., Huet, M. & Kaltz, O. (2010). Experimental evolution of local parasite maladaptation. *J. Evol. Biol.*, 23, 1195–1205.

Becker, U., Dostal, P., Jorritsma-Wienk, L. & Matthies, D. (2008). The spatial scale of adaptive population differentiation in a wide-spread, well-dispersed plant species. *Oikos*, 117, 1865–1873.

Bever, J. (1994). Feeback between plants and their soil communities in an old field community. *Ecology*, 75, 1965–1977.

Blanquart, F. & Gandon, S. (2011). Evolution of migration in a periodically changing environment. *Am. Nat.*, 177, 188–201.

Blanquart, F., Gandon, S. & Nuismer, S. (2012). The effects of migration and drift on local adaptation to a heterogeneous environment. *J. Evol. Biol.*, 25, 1351–1363.

Bonnier, G. (1895), Recherches expérimentales sur ladaptation des plantes au climat alpin. *Ann. Sci. Nat., Bot.*, 7th ser. 20, 217–360.

Bradshaw, A. (1960). Population differentiation in agrostis tenuis sibth. iii. populations in varied environments. *New Phytol.*, 59, 92–103.

Edmunds, G. & Alstad, D. (1978). Coevolution in insect herbivores and conifers. *Science*, 199, 941.

Fournier-Level, A., Korte, A., Cooper, M., Nordborg, M., Schmitt, J. & Wilczek, A. (2011). A map of local adaptation in *Arabidopsis thaliana*. *Science*, 334, 86–89.

Franceschi, N., Cornet, S., Bollache, L., Dechaume-Moncharmont, F., Bauer, A., Motreuil, S. *et al.* (2010). Variation between populations and local adaptation in acanthocephalan-induced parasite manipulation. *Evolution*, 64, 2417–2430.

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L. & Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet.*, 7, e1002355.

Gandon, S. (2002). Local adaptation and the geometry of host-parasite coevolution. *Ecol. Lett.*, 5, 246–256.

Gandon, S. & Nuismer, S. (2009). Interactions between genetic drift, gene flow, and selection mosaics drive parasite local adaptation. *Am. Nat.*, 173, 212–224.

Ganz, H. & Washburn, J. (2006). Relative migration rates and local adaptation in a mosquito-protozoan interaction. *J. Evol. Biol.*, 19, 816–824.

Garrido, E., Andraca-Gómez, G. & Fornoni, J. (2012). Local adaptation: simultaneously considering herbivores and their host plants. *New Phytol.*, 193, 445–453.

Gavrilets, S. (2003). Perspective: models of speciation: what have we learned in 40 years? *Evolution*, 57, 2197–2215.

Gavrilets, S. & Gibson, N. (2002). Fixation probabilities in a spatially heterogeneous environment. *Popul. Ecol.*, 44, 51–58.

Greischar, M. & Koskella, B. (2007). A synthesis of experimental work on parasite local adaptation. *Ecol. Lett.*, 10, 418–434.

Hancock, A., Brachi, B., Faure, N., Horton, M., Jarymowycz, L., Sperone, F. et al. (2011). Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*, 334, 83–86.

Hendry, A., Day, T. & Taylor, E. (2001). Population mixing and the adaptive divergence of quantitative traits in discrete populations: a theoretical framework for empirical tests. *Evolution*, 55, 459–466.

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.*, 173, 579–588.

Hoeksema, J. & Forde, S. (2008). A meta-analysis of factors affecting local adaptation between interacting species. *Am. Nat.*, 171, 275–290.

Kaltz, O. & Shykoff, J. (1998). Local adaptation in host–parasite systems. *Heredity*, 81, 361–370.

Kaltz, O., Gandon, S., Michalakis, Y. & Shykoff, J. (1999). Local maladaptation in the anther-smut fungus *Microbotryum violaceum* to its host plant *Silene latifolia*: evidence from a cross-inoculation experiment. *Evolution*, 53, 395–407.

Kawecki, T. & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecol. Lett.*, 7, 1225–1241.

Kirkpatrick, M. & Barton, N. (1997). Evolution of a species range. *Am. Nat.*, 150, 1–23.

Lajeunesse, M. & Forbes, M. (2002). Host range and local parasite adaptation. *Proc. R. Soc. Lond. B Biol. Sci.*, 269, 703–710.

Leimu, R. & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS ONE*, 3, 4010.

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

Levene, H. (1953). Genetic equilibrium when more than one ecological niche is available. *Am. Nat.*, 87, 331–333.

Lively, C., Dybdahl, M., Jokela, J., Osnas, E. & Delph, L. (2004), Host sex and local adaptation by parasites in a snail-trematode interaction. *Am. Nat.*, 164 (S5), S6–S18.

Lopez Pascua, L., Gandon, S. & Buckling, A. (2012). Abiotic heterogeneity drives parasite local adaptation in coevolving bacteria and phages. *J. Evol. Biol.*, 25, 187–195.

McKay, J. & Latta, R. (2002). Adaptive population divergence: markers, QTL and traits. *Trends Ecol. Evol.*, 17, 285–291.

Morgan, A., Gandon, S. & Buckling, A. (2005). The effect of migration on local adaptation in a coevolving host-parasite system. *Nature*, 437, 253–256.

Mutikainen, P., Salonen, V., Puustinen, S. & Koskela, T. (2000). Local adaptation, resistance, and virulence in a hemiparasitic plant-host plant interaction. *Evolution*, 54, 433–440.

Nagylaki, T. (1980). The strong-migration limit in geographically structured populations. *J. Math. Biology*, 9, 101–114.

Von Marilaun, A.K & Oliver, F.W. (1895). *The Natural History of Plants, Their Forms, Growth, Reproduction, and Distribution*. Henry Holt, New York.

Nuismer, S. & Gandon, S. (2008). Moving beyond common-garden and transplant designs: insight into the causes of local adaptation in species interactions. *Am. Nat.*, 171, 658–668.

Schulte, R., Makus, C., Hasert, B., Michiels, N. & Schulenburg, H. (2011). Host–parasite local adaptation after experimental coevolution of *Caenorhabditis elegans* and its microparasite *Bacillus thuringiensis*. *Proc. R. Soc. B: Biol. Sci.*, 278, 2832–2839.

Sicard, D., Pennings, P., Grandclément, C., Acosta, J., Kaltz, O. & Shykoff, J. (2007). Specialization and local adaptation of a fungal parasite on two host plant species as revealed by two fitness traits. *Evolution*, 61, 27–41.

Thompson, J. (1994). *The Coevolutionary Process*. University of Chicago Press, Chicago.

Thompson, J.N. (2005). *The Geographic Mosaic of Coevolution*. University of Chicago Press, Chicago.

Thrall, P., Burdon, J. & Bever, J. (2002). Local adaptation in the linum marginale —melampsora lini host-pathogen interaction. *Evolution*, 56, 1340–1351.

Turesson, G. (1922). The genotypical response of the plant species to the habitat. *Hereditas*, 3, 211–350.

Turkington, R. & Harper, J. (1979). The growth, distribution and neighbour relationships of trifolium repens in a permanent pasture. *J. Ecol.*, 67, 245–254.

Vogwill, T., Fenton, A. & Brockhurst, M. (2010). How does spatial dispersal network affect the evolution of parasite local adaptation? *Evolution*, 64, 1795–1801.

Whitlock, M. & Gomulkiewicz, R. (2005). Probability of fixation in a heterogeneous environment. *Genetics*, 171, 1407.

Yeaman, S. & Otto, S. (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution*, 65, 2123–2129.

SUPPORTING INFORMATION

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