bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Isolation-by-Drift: Quantifying the Respective Contributions of Genetic Drift and Gene Flow in Shaping Spatial Patterns of Genetic Differentiation

Short title: Isolation-by-drift

Jérôme G. Prunier[1¶*], Vincent Dubut[2&], Lounès Chikhi[3,4&], Simon Blanchet[1,3¶]

[1] Station d'Ecologie Expérimentale du CNRS à Moulis, USR 2936, Centre National de la Recherche Scientifique (CNRS), Moulis, France

[2] UMR 7263 – IMBE, Équipe EGE ; Aix Marseille Université, CNRS, IRD, Avignon Université ; Marseille, France.

[3] UMR 5174 EDB (Laboratoire Evolution & Diversité Biologique), Centre National de la Recherche Scientifique (CNRS), Ecole Nationale de Formation Agronomique (ENFA), Université Paul Sabatier ; Toulouse, France.

[4] UMR 5174 (EDB) ; Université de Toulouse, UPS ; Toulouse, France

*** Corresponding author**

Email: jerome.prunier@gmail.com (JGP)

[¶] These authors contributed equally to this work.

[&] These authors also contributed equally to this work.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Spatial patterns of neutral genetic diversity are often investigated to infer gene flow in wild populations. However, teasing apart the influence of gene flow from the effect of genetic drift is challenging given that both forces are acting simultaneously on patterns of genetic differentiation. Here, we tested the relevance of a distance-based metric -based on estimates of effective population sizes or on environmental proxies for local carrying capacities- to assess the unique contribution of genetic drift on pairwise measures of genetic differentiation. Using simulations under various models of population genetics, we demonstrated that one of three metrics we tested was particularly promising: it correctly and uniquely captured variance in genetic differentiation that was due to genetic drift when this process was modelled. We further showed that (i) the unique contribution of genetic drift on genetic differentiation was high (up to 20 %) even when gene flow was high and for relatively high effective population sizes, and (ii) that this metric was robust to uncertainty in the estimation of local effective population size (or proxies for carrying capacity). Finally, using an empirical dataset on a freshwater fish (*Gobio occitaniae*), we demonstrated the usefulness of this metric to quantify the relative contribution of genetic drift and gene flow in explaining pattern of genetic differentiation in this species. We conclude that considering Isolation-by-Drift metrics will substantially improve the understanding of evolutionary drivers of observed spatial patterns of genetic variation.

Author Summary

Genetic drift, a major evolutionary process compounded in small populations, is the random alteration of allelic frequencies over generations. It ultimately leads to a local loss of genetic diversity and to an increase in genetic differentiation among populations. Genetic drift is however often overlooked in spatial genetic studies, in which measures of genetic differentiation are thus equated to gene flow only. In this study, we reviewed the distance-based metrics having been proposed to account for genetic drift. We then used simulations and an empirical dataset to evaluate the relevance of these metrics. These distances-based metrics are based on estimates of (or proxies for) population sizes, and is easily implemented in any regression-like analysis. We showed that these metrics can be efficiently used to quantify the contributions of both gene flow and genetic drift in measures of genetic differentiation in a large panel of realistic situations, making them a promising tool for geneticists aiming at better understanding processes underlying genetic differentiation in wild populations.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

The maintenance of dispersal capacities among demes has long been recognized as being of tremendous importance for the viability of spatially structured populations [1,2], especially in the current context of increasing habitat fragmentation worldwide [3-5]. Given the technical challenges of directly monitoring individual movements [6,7], indirect estimates of gene flow (i.e. effective dispersal rates [8]) are now routinely assessed using molecular tools [9-16]. When investigating the influence of landscape features on gene flow [17-20], many spatial genetic studies directly rely on pairwise measures of genetic differentiation (i.e. genetic distances such as *Fst*; [21,22]) as measures of functional connectivity [23-26]. Direct gradient analyses [27-29] are then used to investigate the relative contribution of spatial predictors of *isolation-by-distance* (IBD [30]), *isolation-by-resistance* (IBR [31]) and *isolation-by-environment* (IBE [20,32]) to the variance in pairwise genetic distances. However, in spatial genetics studies, an important additional source of genetic variation is often neglected: genetic drift [21,33]. Gene flow and genetic drift indeed interact as opposing forces, the former decreasing and the latter increasing genetic divergence among populations as their respective influences increase [34-38]. As a result, pairwise measures of genetic differentiation are directly impacted by the balance between these two processes [9,34,39] and dispersal rates and functional connectivity may actually be misinterpreted when genetic drift is overlooked because of biased estimates. Disentangling the respective contributions of these two evolutionary forces to the variance in genetic differentiation is thus of crude importance, as omitting the potential influence of genetic drift may strongly affect the interpretation of spatial genetic analyses [40,41].

There have been several attempts to relate pairwise measures of genetic differentiation to gene flow while accounting for genetic drift (e.g. [21,35,42-44]), although most of them do not allow segregating explicitly the respective effects of gene flow and drift [7,34]. The relative contribution of genetic drift to the variance in pairwise measures of genetic differentiation may be directly quantified using estimates of census (N) or effective (N_e) population sizes [21,33] with the two following basic assumptions: (i) N is positively correlated to N_e, and (ii) the lower N_e of a population, the higher the effects of genetic drift on allelic frequencies. These estimates of population sizes can be used to compute distance-based measures of genetic drift, that are then used as any other predictor (such as environmental, resistance or geographic distances) in regression-like analyses to decompose sources of variation in pairwise measures of genetic differentiation [28]. To our knowledge, the use of distance-based measures of genetic drift was first proposed by Relethford [45], computed then as the sum of the inverses of each (census or effective) population size. More recently, Serrouya et al [40] proposed a similar metric, based on the harmonic mean of (census or effective) population sizes. The use of such metrics to account for the

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

contribution of genetic drift in spatial patterns of genetic variation was rarely considered in spatial genetics (but see [46,47]) and their efficiency in capturing the effects of genetic drift has never been tested theoretically, although it deserves full attention.

Directly accounting for genetic drift through the use of N or N_e is probably the most straightforward approach, but implies a major difficulty: estimating these demographic parameters, a task that may turn out to be tricky [48-50]. Alternatively, we here propose to consider environmental estimates of local carrying capacities (K) as a proxy for population sizes. Carrying capacity reflects the upper asymptote of the logistic growth curve of a population given the distribution and abundance of resources determined by local environmental conditions [20,51,52] and can be approximated using specific environmental variables such as habitat patch size or habitat quality (e.g. [53,54]). Local environmental characteristics are most often used in the framework of IBE to compute pairwise ecological distances and hence ecological resistance to gene flow [20,32,55-58]. However, we argue that metrics habitually used in the framework of IBE do not meet the objective of disentangling the respective contributions of gene flow and genetic drift (see Box 1). Rather, distance-based metrics of genetic drift could be computed from environmental proxies in place of effective population sizes N_e using the formulae proposed by Relethford [45] and Serrouya et al [40] respectively.

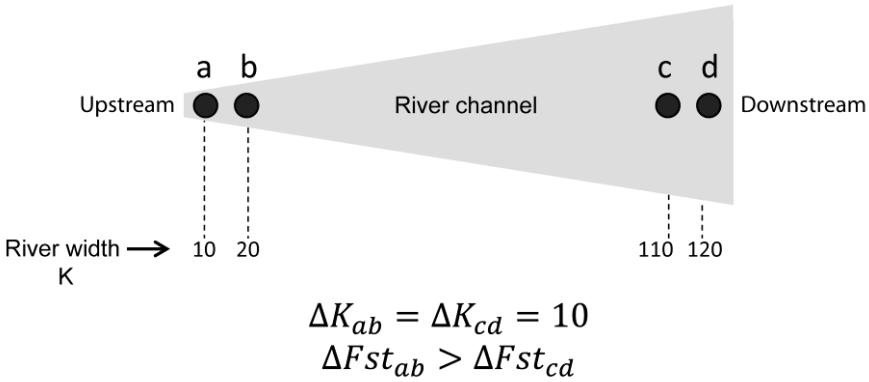
BOX 1. Why do Euclidian environmental distances fail to capture genetic drift?

Several metrics can be used as ecological distances in the framework of IBE, such as metapopulation connectivity indices [20,52] or simple univariate/multivariate Euclidean distances [32,47,59-61]. However, they may not be appropriate to quantify the effect of genetic drift in spatial patterns of genetic variation. For instance, metapopulation connectivity indices will not allow disentangling the respective contributions of gene flow and genetic drift as these metrics include both local patch and interpatch data in the same formula. The case of Euclidean distance metrics is subtler, and can be better understood through a concrete example. Consider four populations in a river channel (Fig 1). Two of them are located 1km apart in the upstream portion of the channel, and are respectively characterized by a river width of 10m and 20m. The two other populations are located 1km apart but in the downstream portion of the channel, and are respectively characterized by a river width of 110m and 120m. The (Euclidian) difference in river width between upstream populations and between downstream populations is the same, i.e. 10m. However, as river width may directly drive local carrying capacities [62-64], we may expect higher genetic differentiation between upstream than between downstream populations simply because of higher genetic drift upstream. Assuming that degrees of spatial connectivity between upstream and

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

between downstream populations are the same, one may fail to explain variability in pairwise measures of genetic differentiation when using a simple Euclidean distance between environmental variables.

Fig 1. The absolute difference between proxies for local carrying capacities *K* (here, river width) may fail to explain spatial patterns of neutral genetic variation arising from genetic drift.



In this study, we compared the relative efficiency of three genetic drift distance metrics based on both N_e and environmental proxies for local carrying capacity (K) in explaining spatial patterns of genetic variation, using regression commonality analyses as a statistical procedure of variance decomposition [28,65]. The first metric (*ds*; distance based on the subtraction, i.e. Euclidian distance) is inspired from common measures of IBE. The second one (*di*; distance based on the inverse) was proposed by Relethford [45]. The third one is the opposite to the harmonic mean (preferred over the arithmetic mean [66]) of carrying capacities (*dhm*; distance based on the harmonic mean). Because the metric based on the harmonic mean shows negative relationships with *Fst* [40,46] and thus does not behave as a classical distance-based metric (an increase in the dissimilarity between populations is supposed to lead to an increase in genetic differentiation), we rather considered its opposite.

$$ds = |K_1 - K_2| \quad (1)$$

$$di = \frac{1}{K_1} + \frac{1}{K_2} = \frac{K_1+K_2}{K_1K_2} \quad (2)$$

$$dhm = -\frac{2}{\frac{1}{K_1}+\frac{1}{K_2}} = -\frac{2K_1K_2}{K_1+K_2} \quad (3)$$

We first investigated the ability and efficiency of each metric to account for genetic drift when they are directly computed from N_e, using simulations in simple two-deme situations and in more complex genetic models of population structure. Secondly, we used similar simulations to test whether these metrics are still

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

efficient when based on environmental proxies of carrying capacities, assuming that K is an imperfect proxy of N_e. Thirdly, we assessed the efficiency of each metric in an empirical case study involving a fish species (*Gobio occitaniae*) living in a river landscape and using environmental estimates of population sizes. We finally discussed how and why these simple metrics measuring *isolation-by-drift* (IBDr) should be recurrently used in spatial genetics studies, notably in the framework of regression commonality analyses. We notably expected the *ds* metric inspired from the IBE framework to perform poorly (Box 1), contrary to the *di* and *dhm* metrics that are expected to efficiently account for the effect of genetic drift in spatial patterns of genetic variation.

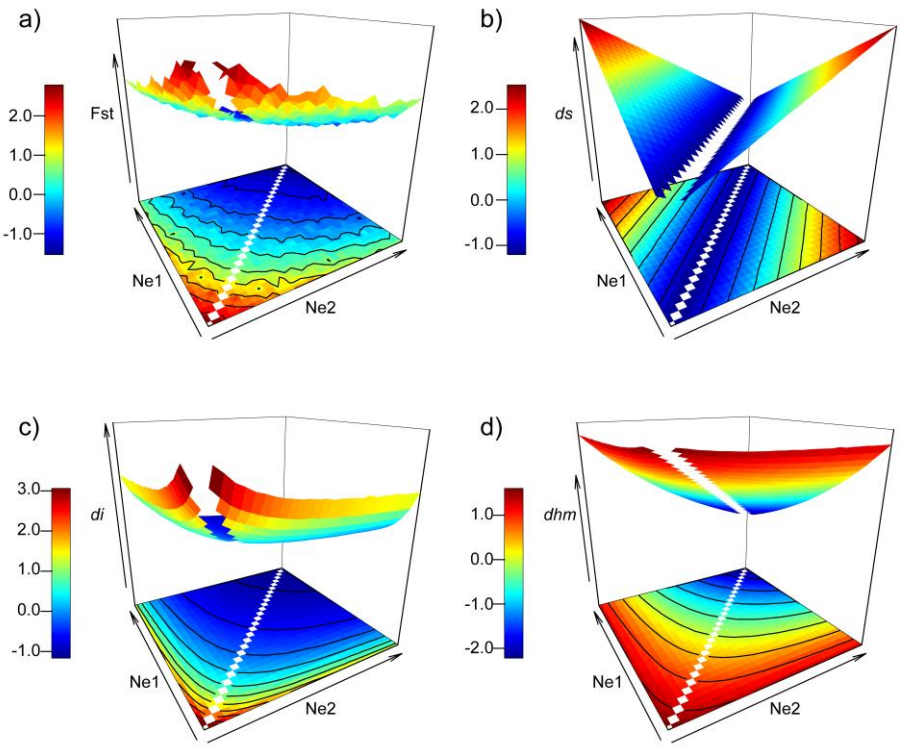
bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Simulated datasets

A simple two-deme model. As expected, when the migration rate *m* between two demes of effective population sizes Ne_1 and Ne_2 was low (*m* = 0.005), mean *Fst* values were highest when both Ne_1 and Ne_2 were low, decreased when the N_e of one of the two demes increased, and were lowest when both Ne_1 and Ne_2 were high (Fig 2a). As expected, the *ds* metric poorly mirrored patterns of *Fst* (Fig 2b; *r* = 0.125). Contrastingly, the *di* and *dhm* metrics followed patterns very similar to that observed for *Fst* (Fig 2c, d), although *di* tends to better fit the general *Fst* pattern than *dhm* (*r* = 0.962 and *r* = 0.939 for *di* and *dhm* respectively), especially when at least one of the demes had low to intermediate *Ne* (Fig 2d).

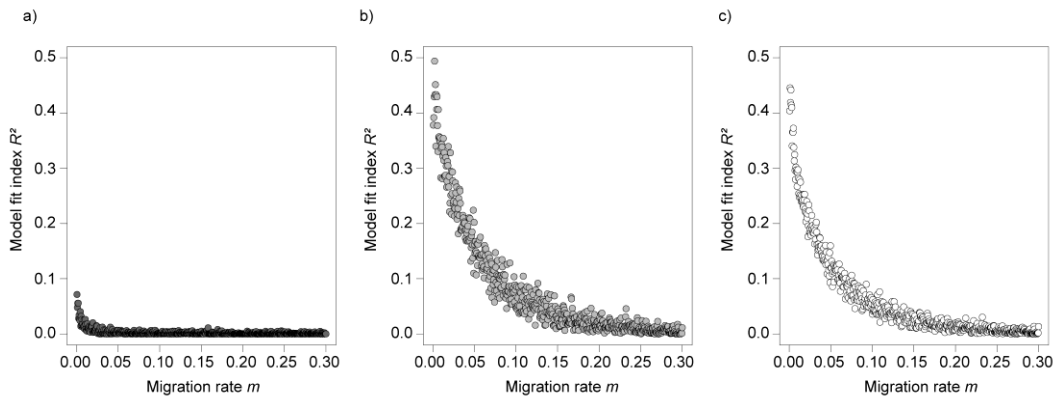
Fig 2. Behaviour of *Fst* (a) and each genetic drift metric (b: *ds*; c: *di*; d: *dhm*) in a simple two-deme system as a function of effective population sizes Ne_1 and Ne_2, expressed as haploid numbers of genes, both ranging from 30 to 300. *Fst* values and genetic drift metrics were all standardized. Projections of the 3D perspective surfaces are shown at the base of each plot.



bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

When migration rates *m* between the two demes were picked from an uniform distribution ranging from 0.0001 to 0.3, *ds* systematically failed to explain a substantial proportion of variance in pairwise genetic distances (Fig 3a). On the contrary, *di* (Fig 3b) and *dhm* (Fig 3c) respectively explained up to 50% and 45% of variance in pairwise genetic distances at low migration rates ($m < 0.005$). This relative contribution of these two metrics decreased exponentially as *m* increased; for instance, the contribution of these metrics fell below 5% for $0.1 < m < 0.15$ (Fig 3b-c). Noteworthy, the dispersion around the model fit was slightly larger for *di* than for *dhm* (Fig 3b-c).

Fig 3. Contribution of metrics *ds* (a), *di* (b) and *dhm* (c) to the variance in genetic differentiation as a function of the migration rate *m* in a simple two-deme system.



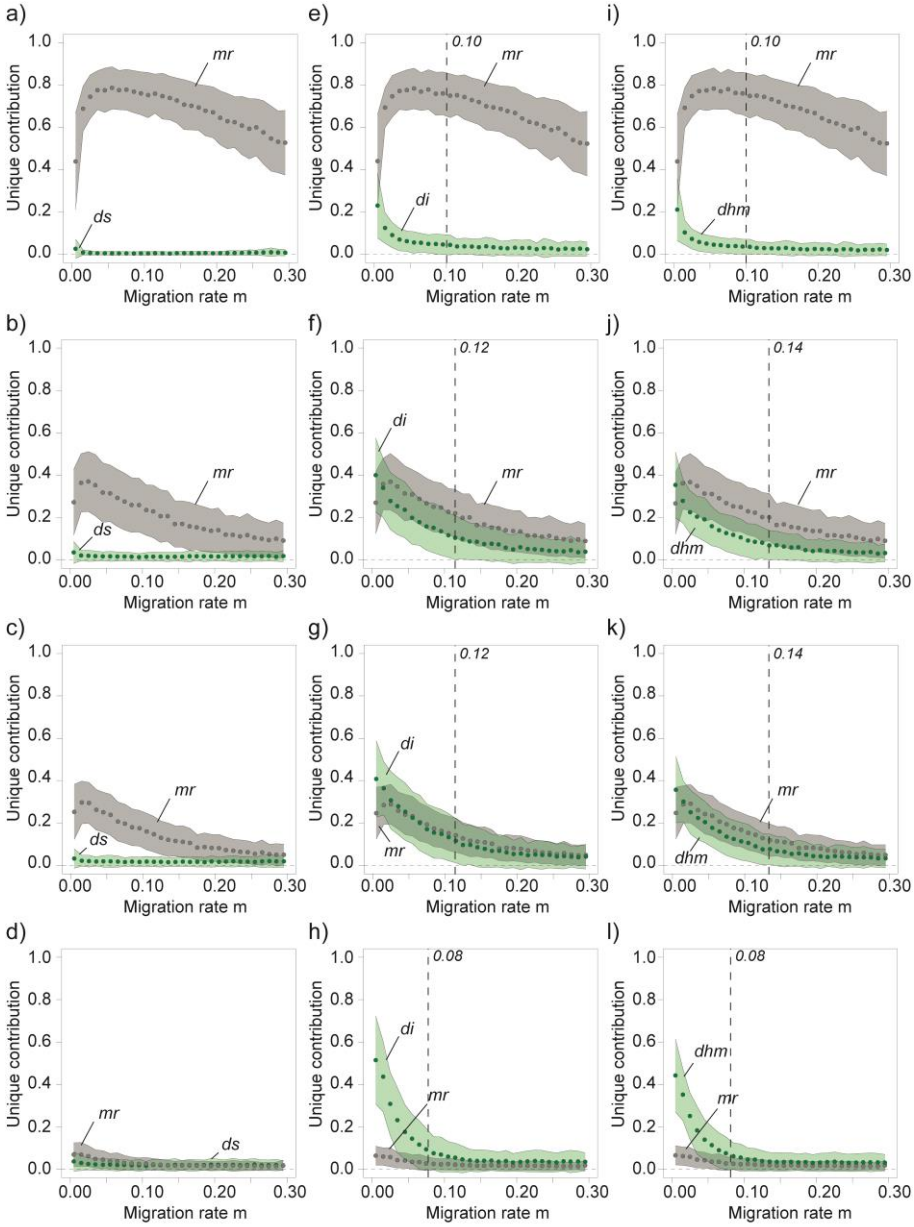
Realistic genetic models. To evaluate the efficiency of each metric in more realistic situations, gene flows were simulated in a one-dimensional linear network or a two-dimensional lattice network. Each network was composed of 16 demes of varying effective population sizes Ne_i and exchanging migrants at a rate *m* following either a stepping-stone or a spatially limited (IBD) migration model (S2 Fig). For all four genetic models (and whatever the model parameters), the unique contribution of *ds* was close to zero (Fig 4a-d), that is, the amount of variance uniquely explained by this metric irrespective of the possible collinear effects of the other predictor *mr* coding for inter-deme matrix resistance was negligible. Overall, both *di* and *dhm* explained a non-negligible part of the total variance in pairwise *Fst* (i.e. from 5% to more than 50%) as soon as *m* was lower than 0.15 (Fig 4e-l). Under these conditions ($m < 0.15$), the relative unique contributions of *di* and *dhm* differed among genetic models: they were the lowest for a linear network with stepping-stone migration and were the highest in a lattice network with spatially limited dispersal. In this later case, *di* and *dhm* explained (for $m < 0.05$) much more variance than the traditionally used *mr* (Fig 4h, l). For other genetic models, the unique contributions of *di* and

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

dhm were as high as the unique contribution of *mr* as far as *m* remained low. *di* and *dhm* behaved very similarly with no noticeable differences, although dispersion around the mean was still higher for *di* than for *dhm* for all genetic models.

Fig 4. Unique contribution of metrics *mr* and *ds* (a-d), or *mr* and *di* (e-h) or *mr* and *dhm* (i-l) as a function of the migration rate m for $\alpha = 0$. Results are for a linear network with stepping-stone migration (a, e, i), a linear network with spatially limited dispersal (b, f, j), a lattice network with stepping-stone migration (c, g, k) and a lattice network with spatially limited dispersal (d, h, l). Circles represent the average unique contribution of each variable and coloured areas represent the dispersion of unique contributions around the mean, as defined by standard deviations (in grey for *mr* and in green for genetic drift distance metrics). Vertical dashed lines indicate the migration rate *m* above which the unique contribution of genetic drift distance metrics become negligible.

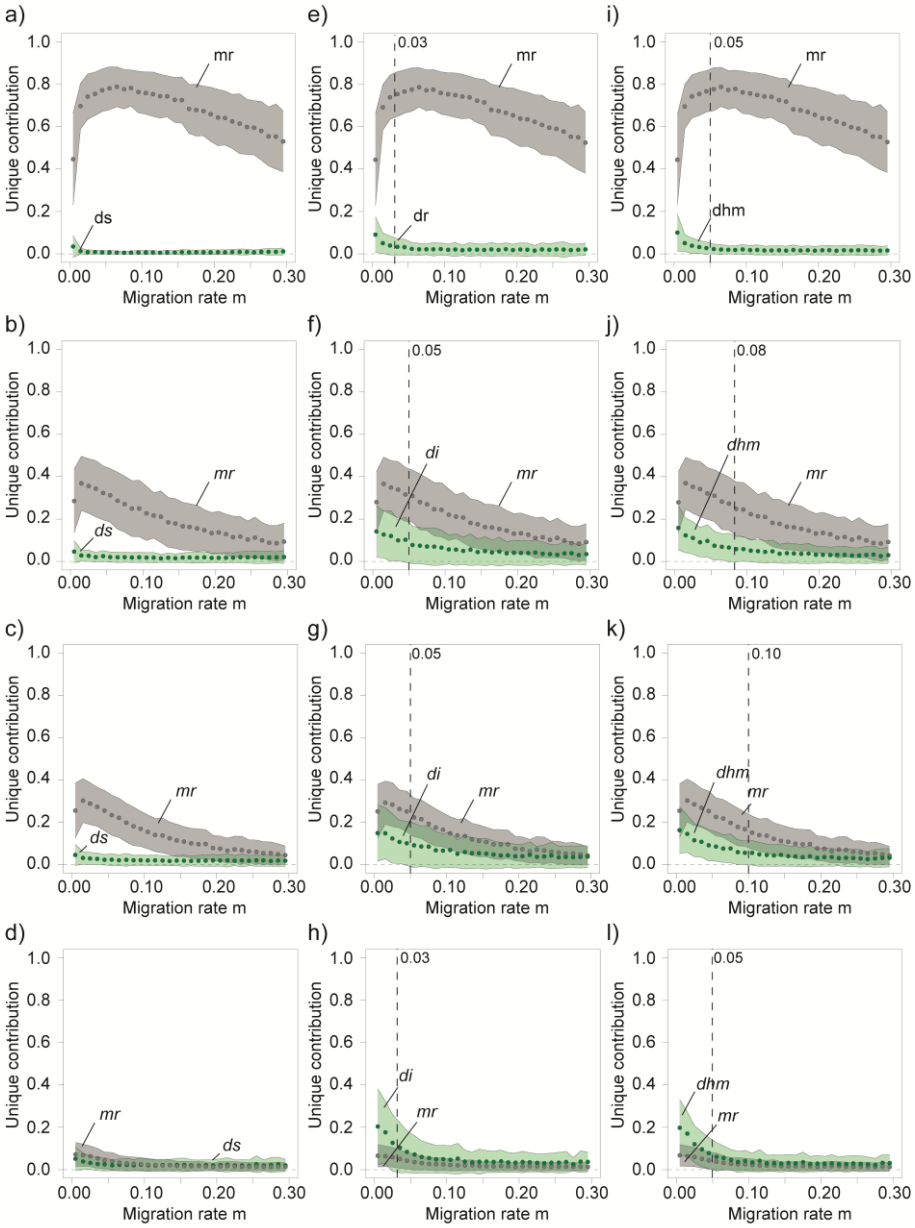
bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



When uncertainty was included in the estimation of demes' effective population size so as to mimic an environmental proxy for K (using $\alpha \in [-0.9\,, 0.9]$), *mr* and *ds* showed similar patterns to those in absence of uncertainty (Fig 4); *ds* systematically failed to explain variance in genetic differentiation. *di* and *dhm* behaved similarly to situations where true estimates of *Ne* were used, although unique contributions were systematically lower (but still substantial for low *m* values) in average (Fig 5). Furthermore, dispersion around the mean was noticeably larger for *di* than for *dhm*, except for a linear network with stepping-stone migration (Fig 5).

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig 5. Unique contribution of metrics *mr* and *ds* (a-d), or *mr* and *di* (e-h) or *mr* and *dhm* (i-l) as a function of the migration rate *m* for an uncertainty parameter α picked from a uniform distribution ranging from -0.9 to 0.9. See legend in Fig 4 for other details.



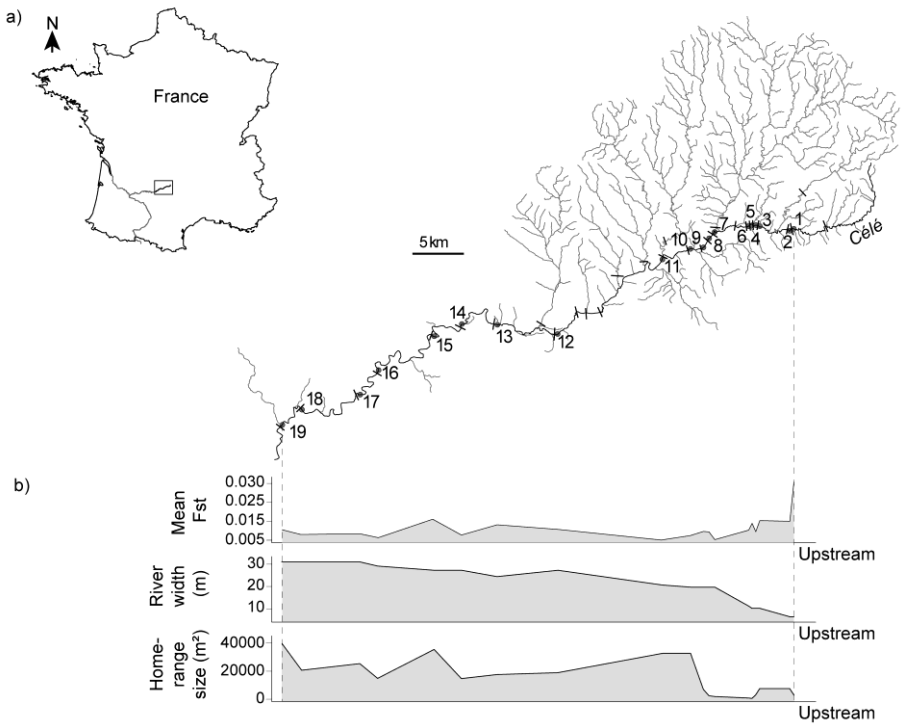
Empirical dataset

Concerning the empirical dataset involving the freshwater fish *Gobio occitaniae* in a river network (Fig 6a), neither null alleles nor linkage disequilibrium were detected, and all populations were at Hardy-Weinberg equilibrium after sequential Bonferroni corrections. Highest mean pairwise *Fst* values involved population 1 (Fig 6b), which was situated upstream of the main channel. This population was characterized by the smallest river width but was not associated with lowest estimates of home-range size (estimated as the product of length and

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

width of the river network (including tributaries) delimited by any downstream or upstream weir), suggesting that river width may be a better proxy of *K* (and hence genetic drift) than the home range size.

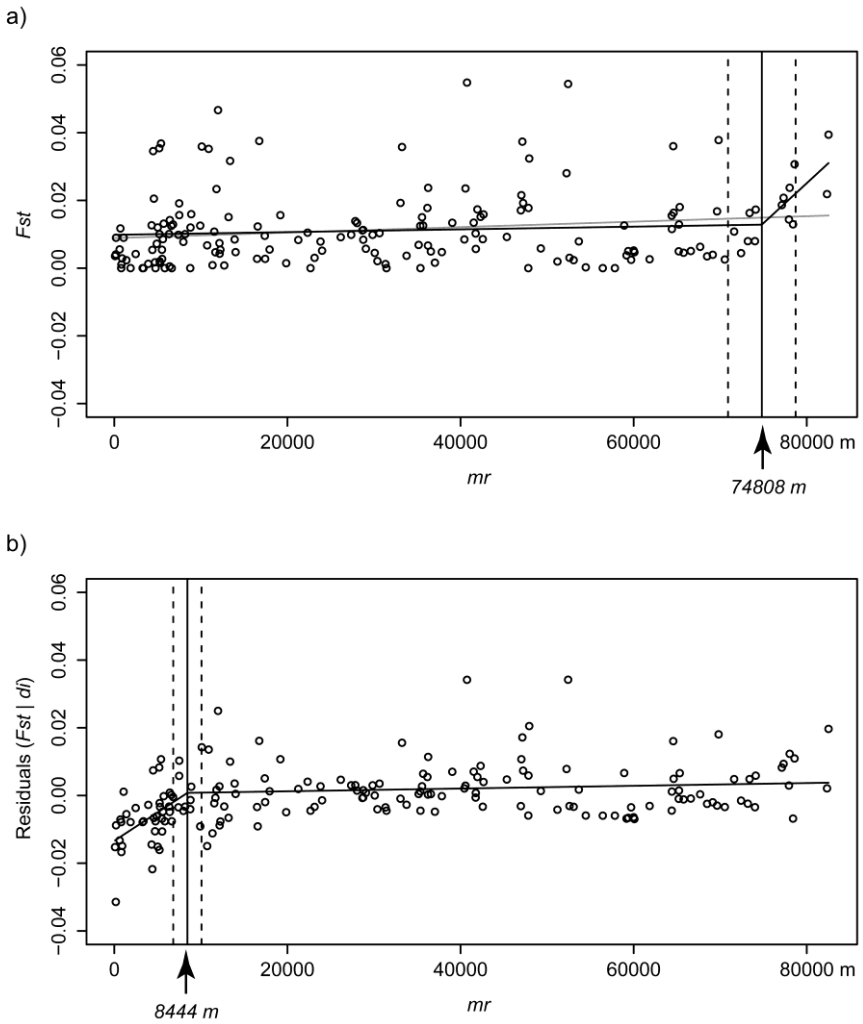
Fig 6. **Main characteristics of the empirical dataset.** (a) Geographical position of the river Célé (in black) and its tributaries (in grey). Grey dots represent the 19 sampling sites (numerated along the upstream-downstream gradient) while small black lines indicate the position of obstacles (weirs) considered in the study (that is, obstacles located on the mainstream channel and obstacles from tributaries used to compute home-range sizes). (b) Downstream-upstream profiles of mean *Fst* values, river width and home-range sizes.



The pattern of IBD was characterized by a slightly positive relationship between *Fst* and *mr* (here coded as the between-deme riparian distance; Fig 7a). Piecewise regression explained a higher proportion of the variance in Fst (6.7%) that linear regression (3.2%), but the threshold value, located at about 75 km (Fig. 7a), probably stemmed from a border effect and had no true biological meaning. Overall, the IBD pattern was characterized by a wide degree of scatter for all distance classes, suggesting a lack of migration-drift equilibrium with genetic drift being more influential than gene flow (case III in [34]).

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig 7. Scatterplot of pairwise *Fst* against (a) pairwise riparian distances (*mr*), (b) pairwise *di* values and (c) pairwise *dhm* values in the empirical dataset. The *di* and *dhm* metrics were computed from river width. All explanatory variables (*mr*, *di* and *dhm*) were rescaled from 0 to 1 to facilitate comparison of panels.



bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Whatever the proxy used for *K*, the unique contribution of *mr* was rather weak, ranging from 1.8 to 7.8% (Table 1). This variability in unique contributions of *mr* stemmed from collinearity with distance-based metrics of genetic drift, as revealed by common contributions C [28,65]: indeed, the highest unique contribution of *mr* (U = 7.8%) was also associated with the highest (negative) common contribution (C = - 4.8%). When *K* values were estimated from home-range sizes, the model including *ds* only explained 3.2% of variance in genetic differentiation (Table 1), with a negligible unique contribution of *ds* (0.3%). The model including *di* did not perform better (unique contribution of 1.1%) whereas the model including *dhm* explained up to 12.7% of variance in genetic differentiation. This substantial increase in model fit stemmed from *dhm*'s unique contribution (U = 9.8%) while the unique contribution of *mr* was comparable to previous models. When *K* values were estimated from river width, the unique contribution of *di* and *dhm* strongly increased to reach 41.3% and 36.7% respectively (Table 1). These results confirm that river width was a better proxy for *K* than home-range size in this dataset.

Table 1. **Results of both MRDM and commonality analyses performed on empirical data.**

K	Model	R^2	Pred	β	p	U	C	T
Home-Range	$mr + ds$	0.032	*mr*	0.147	0.105	0.018	0.011	0.029
			ds	0.059	0.488	0.003	0.011	0.014
	$mr + di$	0.040	*mr*	0.185	0.038	0.034	-0.004	0.029
			di	0.104	0.173	0.011	-0.004	0.006
	$mr + dhm$	0.127	*mr*	0.192	0.026	0.037	-0.008	0.029
			dhm	0.313	0.001	0.098	-0.008	0.090
River width	$mr + di$	0.442	*mr*	0.283	0.001	0.078	-0.048	0.029
			di	0.652	0.001	0.413	-0.048	0.365
	$mr + dhm$	0.396	*mr*	0.219	0.003	0.048	-0.018	0.029
			dhm	0.607	0.001	0.367	-0.018	0.348

For each type of environmental proxy for carrying capacities (*K*) and each model (Model), the table provides the model fit index (R^2) and, for each predictor (Pred), the beta weight (β), the p-value (p) and finally unique (U), common (C) and total (T) contributions to the variance in the dependent variable.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

When exploring the relationship between residuals of the linear regression between *Fst* and the *di* metric (based on measures of river width) and *mr*, piecewise regression explained a substantially higher proportion of the variance in Fst (23.8%) than linear regression (12.5%). The scatterplot showed an increase in residual values up to 8.5km and a clear-cut plateau beyond this threshold (Fig 7b). This pattern clearly suggests a lack of migration-drift equilibrium with gene flow being more influential than genetic drift up to 8.5km (case IV in [34]). This result indicates that accounting for the confounding contribution of genetic drift to pairwise measures of genetic differentiation may provide more precise information regarding the spatial extent of gene flow.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

Our study demonstrates that considering IBDr metrics, that is, distance-based metrics based on estimates of effective population sizes or on environmental proxies for local carrying capacities, is a highly relevant approach to disentangle and thoroughly quantify the respective contribution of two major evolutionary processes, gene flow and genetic drift, to the variance in pairwise measures of genetic differentiation.

Comparison of IBDr metrics

The distance-based metrics of genetic drift (*ds*, *di* and *dhm*) tested here did not perform equally. As expected, *ds* -which is a simple subtraction between *Ne* or proxies of *Ne* - was a poor estimator of genetic drift. This observation has important implications for geneticists interested in testing hypothesis regarding isolation-by-environment. Many local variables may adequately reflect the influence of environment on gene flow processes such as emigration and immigration [20]: this is the case of variables embodying information about habitat quality such as resource availability (mating partners, shelters, food), predatory risk or intraspecific competition [67-69]. For instance, some patches act as sources while others may attract dispersal individuals (i.e. sink patches), depending on the perception of local patch quality by individuals [70-72]. In this situation, a simple pairwise Euclidean distance between habitat quality variables may properly mirror the perceptual distinction driving individuals' dispersal or settlement decisions. However, the case of local variables embodying information about patch size is more intricate [3]. Of course, effective dispersal events are often density-dependent ([69]) and gene flow may thus be altered by spatial heterogeneity in patch sizes. But patch sizes may ultimately reflect local carrying capacities, and thus the possible effects of genetic drift rather than gene flow on genetic differentiation. In this situation, the use of a simple pairwise Euclidean distance between variables is irrelevant (see Box 1). Considering local environmental variables for isolation-by-environment studies thus requires a thorough understanding of their possible influence on effective population sizes and, when required, the choice of a specific metric such as *di* or *dhm* to quantify the relative influence of genetic drift. It is noteworthy that a single variable can be used to estimate both IBE and IBDr through the use of complementary metrics.

When compared to *ds*, the *di* and *dhm* metrics demonstrated a much higher efficiency in quantifying genetic drift. They both exhibited patterns very similar to *Fst* in a simple two-deme situation, properly rendering the influence of genetic drift on deme differentiation as effective population sizes decreased (Fig 2c-d), and explaining up to 50% of variance in measures of genetic differentiation for low migration rates (Fig 3b-c). These

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

contributions to genetic variance followed an expected negative exponential as the migration rate increased, reflecting migration-drift equilibrium [7]. The same conclusions can be drawn from genetic data simulated under a series of more realistic situations (Fig 4). However, amounts of unique contributions at low migration rates strikingly depended on population structures and migration models, which reflected variation in the dynamics of migration-drift equilibriums. Overall, the use of drift metrics allowed explaining substantial amounts of variance in measures of genetic differentiation (up to 50%) at low migration rates.

These metrics were still highly efficient in the fish empirical dataset (Table 1) despite an evident lack of migration-drift equilibrium (Fig 7a), making them particularly promising even for empirical case studies for which a deviation from a migration-drift equilibrium is detected [34]. For *G. occitaniae*, the use of *di* and *dhm* indeed allowed explaining substantial amounts of variance in genetic differentiation (Table 1) when river width was used as a proxy for local carrying capacities: only 2.9% of variance in measures of genetic differentiation would have been accounted for if *mr* had been considered as the only predictor (data not shown), whereas up to 44.2% is explained when either with *di* or *dhm* are used as additional explanatory variables (Table 1). Finally, plotting the residuals of the linear regression *Fst* ~ *dhm* (or *Fst* ~ *di*) against Euclidean distances may help identify the spatial scale at which the amount of gene flow is counterbalanced by genetic drift, thus providing further insight into the extent of effective dispersal rate (Fig. 7b). This empirical case study nicely exemplifies the added value for geneticists of integrating distance-based metrics accounting for isolation-by-drift.

Direct versus environmental estimates of Ne

Interestingly, when a small level of uncertainty in the relationship between effective population sizes and estimates of local carrying capacities was considered, our simulations showed that both *di* or *dhm* were still efficient at detecting genetic drift. However, *dhm* slightly outperformed *di* at intermediate migration rates (i.e., for *m* ranging from 0.05 to 0.1) as the degree of uncertainty increased (S3 Appendix) since *di*'s unique contribution to the variance in genetic differentiation showed higher dispersion around the mean than *dhm* (Fig 5). This trend was confirmed by the empirical dataset; when local carrying capacities were estimated from home-range sizes, *di* failed to detect any genetic drift contribution to genetic differentiation, whereas *dhm* -though less efficient than with river width as a proxy- still explained ~ 10% of variance (Table 1). This difference stems from the inner characteristics of each metric (Fig 2c-d): While *di* values are highest for lowest effective population sizes and show a rapid decrease as soon as effective population sizes increase, the decrease in *dhm* values is much softer, thus still allowing the detection of genetic drift effects on genetic differentiation despite

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

higher uncertainty in environmental estimates of local carrying capacities, and eventually effective population sizes. Although the *dhm* metric appeared less efficient at perfectly reflecting the effect of drift in simple simulated datasets, it may actually be more robust when using environmental proxies for local carrying capacities and should therefore be preferred (or compared) to *di*. It is noteworthy that the two metrics can easily be used in competing models and, provided collinearity patterns are inspected [28], the best at fitting the dataset be selected according to model fit criteria such as the R^2 or the Akaike Information Criteria.

Biologically relevant metrics

Simulations indicated that the influence of genetic drift was still perceptible for migration rates up to [0.1 - 0.15], irrespective of the genetic model being considered (Figs 3-4-5). Interestingly, this range of values is higher than migration rates likely to be encountered in natural systems. Indeed, summary statistics from 49 recent empirical studies that used BAYESASS [12] to estimate interpatch migration rates (collected from a literature survey by [73]; see S2 Text and S2 Table for details) indicated that the median value of average migration rates was 0.023, with more than 95% of studies showing average estimates lower than 0.1 (S4 Fig). For instance, the average estimate of migration rates in *G. occitaniae* in our empirical dataset was 0.02 (unpublished data). These observations suggest that genetic drift is likely to be an important driver of spatial genetic variation in many empirical datasets. We thus argue that considering IBDr in future spatial genetic studies through the use of distance-based metrics such as *di* or *dhm* may thoroughly improve our understanding of observed spatial patterns of genetic variation, at least in situations where genetic drift is the actual main driver of genetic differentiation.

Limitations of IBDr metrics

Considering the difficulties in accurately estimating N_e from genetic data [50,74], the use of alternative estimates of population size such as observed local densities (e.g. [75,76]) or habitat patch size (e.g. [53,54]) to compute IBDr metrics is particularly appealing, but has yet to be considered with caution. The validity of such metrics indeed proceed from the assumption that effective population sizes have remained constant over time [45]. This assumption theoretically limits the practical use of distance-based metrics of genetic drift to systems in which populations are only subject to continuous drift, that is, to the evolutionary process of random fluctuations in allelic frequencies naturally occurring in all populations, whatever their size (although compounded in small ones [21,36,37]). For populations having suffered from bottleneck events [37,77] or from founder effects [78],

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

observed densities or local environmental variables may constitute inaccurate proxies for effective population sizes, thus making genetic drift metrics poor predictors of spatial patterns of genetic differentiation. In these situations, estimating effective population sizes from molecular data -although a delicate exercise- probably remains the best option (see [48] for a review synthesizing methods used to estimate *Ne*). More generally, integrating the demographic processes affecting effective population size over time will be an important challenge to overcome so as to make spatial genetics an integrative discipline accounting for the complexity of spatially and temporally dynamic populations [79].

Conclusion

Considering the ineluctable interplay between evolutionary forces such as gene flow and genetic drift, the combined use of distance-based metrics of genetic drift and classical landscape predictors such as matrix resistance metrics or ecological distances in regression commonality analyses [28] may substantially improve our understanding of how each process respectively contributes to observed spatial patterns of genetic variation. This approach is all the more relevant as it may provide accurate information about the contribution of each process even when a lack of migration-drift regional equilibrium has been identified, which constitutes a substantial advantage over many other methods.

More generally, habitats modifications by humans have two components [3,80]; one acting on decreasing connectivity (fragmentation) and another acting on habitat and resource availability (habitat loss and degradation). By reducing the size of available habitats and by decreasing connectivity among habitats, humans are rapidly making the ground more and more fertile for genetic drift to becoming an increasingly influential evolutionary process. We therefore believe that the time is ripe to systematically quantify the influence of genetic drift on the spatial genetic structure of wild populations.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Simulated datasets

For all simulations, we used a computational pipeline including the programs ABCsampler [81], Simcoal2.1.2 [82] and arlsumstat [83] to simulate and analyse microsatellite genetic datasets, with 15 independent loci following a stepwise mutation model and a unique mutation rate μ = 0.0005. Parameter values (local demes' effective population sizes Ne_i and symmetrical migration rates m_i) were picked from prior distributions using ABCsampler and were then used as inputs in Simcoal2.1.2 to simulate genetic data based on a coalescent approach. In all simulations, 30 haploid genotypes (that is, 15 diploid individuals) were sampled from each deme at the end of simulations and were used to compute pairwise *Fst* among demes using arlsumstat. *ds*, *di* and *dhm* metrics were computed on the basis of demes' local carrying capacities *K*, with $K_i = Ne_i + \alpha Ne_i$. The parameter α represents the uncertainty in the estimate of local demes' population size through an environmental proxy such as habitat patch size. The estimates of local demes' effective population size were considered as unbiased for $\alpha = 0$, or uncertain for $\alpha \neq 0$. *ds*, *di* and *dhm* metrics computation as well as all statistical analyses were performed in R 3.1.2 [84].

We first investigated the match between *Fst* values and each metric in a simple two-deme situation. Local demes' population sizes Ne_i (expressed as haploid numbers of genes) were randomly picked from a uniform distribution ranging from i = [30 - 300] while the migration rate *m* was fixed at 0.005 and α was set to 0. We simulated 1×10^5 genetic datasets and computed, for each dataset, pairwise *Fst* as well as *ds*, *di* and *dhm* metrics. Both Ne_1 and Ne_2 were then divided into 30 classes of equal size and datasets were pooled for each Ne_1xNe_2 combination of classes (i.e., 900 pools). For each pool, we computed mean values of *Fst*, *ds*, *di* and *dhm* and eventually created four standardized pairwise matrices of size 30x30: a dependent matrix of mean *Fst* values and three explanatory matrices of mean ds, *di* and *dhm* values respectively. In addition to plotting each matrix as a function of Ne_1 and Ne_2 classes, we computed the respective Pearson's correlation coefficients *r* between the *Fst* matrix and each explanatory matrix.

Secondly, we assessed the explanatory power of each metric as a function of the migration rate *m* in a simple two-deme situation. Demes' population sizes were randomly picked from a truncated normal distribution with a mean of 40 and a standard deviation of 200, values being bounded between 40 and 1000 (S1 Fig), while migration rates *m* were picked from a uniform distribution ranging from 0.0001 to 0.3. As previously, the parameter α was set to 0. We simulated 6×10^5 genetic datasets and computed, for each dataset, pairwise *Fst* as well as *ds*, *di* and *dhm* metrics. Datasets were then pooled according to their migration rate into 600 classes

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

defined every 0.0005 units (about 1000 datasets per class). For each class, we computed the model fit indices R^2 of the univariate linear regressions between *Fst* and each metric.

We then assessed the strength of each metric (compared to a resistance metric) in four complex situations differing according to both the network structure and the migration model used for simulations (S2 Fig). We considered two different network structures: a one-dimensional 16-deme linear network and a two-dimensional 16-deme lattice network. Euclidean distances between adjacent demes were arbitrarily set to 100m. We considered two distinct migration models: a spatially structured island model (or IBD model [21]) for which migration rates *m* decreases with Euclidean distance following an inverse square function, and a spatially structured stepping-stone model [85] where demes can only exchange migrants with adjacent demes. For each situation, 10000 genetic datasets were simulated. As previously, the parameter α was set to 0, demes' population sizes *N* were randomly picked from a truncated normal distribution with a mean of 40 and a standard deviation of 200, values being bounded between 40 and 1000, and migration rates *m* were picked from a uniform distribution ranging from 0.0001 to 0.3. For each simulated dataset, we computed five pairwise matrices (*Fst*, *ds*, *di*, *dhm* and Euclidean distances *mr*, the latter acting as a simple measure of inter-deme matrix resistance) and performed three multiple linear regressions between *Fst* and each metric, with *mr* as a unique covariate: (1) $Fst = (mr + ds)$, (2) $Fst = (mr + di)$ and (3) $Fst = (mr + dhm)$. All explanatory variables were z-transformed (by subtracting the mean and dividing by the standard deviation of the variable) before regressions so as to standardize parameter estimates. Commonality analysis, a variance partitioning procedure that is particularly suited to identify collinearity issues likely to arise in most spatially structured dataset [28,65], was used in R (package yhat [86]) to estimate the respective unique contribution of each explanatory variable to the variance in the dependent variable. This unique contribution (U) is the part of the total variance in the dependent variable that is explained by the sole effect of the predictor being considered (*mr* or one of the three distance-based metric of genetic drift). This approach improved the accuracy of the estimate of the contribution of each predictor when collinearity occurs. Datasets were finally pooled according to their migration rate into 30 classes defined every 0.01 units (about 330 datasets per class). For each class, we computed the mean and the standard deviation of unique contributions of each predictor. A unique contribution was considered as negligible as soon as the dispersion around the mean included zero.

Finally, we investigated the influence of uncertainty in the estimation of demes' population size through an environmental proxy. We used the same approach as described above (the same four genetic models with the

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

same initial sets of *Ne* and *m* parameters) but with the parameter α picked from a uniform distribution ranging from -0.9 to 0.9 (S3 Fig).

Empirical dataset

As en empirical example, we considered neutral genetic data collected in the gudgeon (*Gobio occitaniae*), a small benthic freshwater fish. Fieldwork was conducted in accordance with French laws and with the approval of the Prefecture du Lot. Five hundred and sixty-two individuals were caught in 2011 using an electro-fishing in 19 sampling sites scattered along the mainstream channel of the river Célé (South-western France; Fig. 6a). Sites were scattered so as to cover the whole upstream-downstream gradient. For each individual, we collected and preserved in 70% ethanol a small piece of pelvic fins, before releasing the fish in their original site. Genomic DNA was extracted from samples using a salt-extraction protocol [87]. Genotyping was performed using a subset of 11 autosomal microsatellite loci chosen among those described in Grenier et al [88] (S1 Table). Polymerase chain reactions (PCR) and genotyping were performed as in Grenier et al [88]. The presence of null alleles was assessed at each locus by analysing homozygote excess in each population using MICROCHECKER 2.2.3 [89]. We also checked for linkage disequilibrium among loci and for Hardy-Weinberg equilibrium at each population using GENEPOP 4.2.1 [90] after sequential Bonferroni correction to account for multiple related tests [91]. Pairwise *Fst* were computed between all pairs of site following Weir and Cockerham [22] using the MATLAB software-coding environment [92,93].

Environmental variables were extracted from national vector datasets (BDCarthage from National Geographic Institute, France) and from the theoretical hydrographical network for France (RHT [94]) using ARCGIS 9.3. The riparian distance among sites was used as a measure of matrix resistance *mr* among sites. We first graphically inspected the linear relationship between pairwise *Fst* and the riparian distance *mr* between pairs of demes as proposed by Hutchison and Templeton [34] to assess migration-drift equilibrium (or lack of it) in our empirical dataset. We further investigated the observed pattern by using piecewise regression [95,96] to identify the distance threshold at which different linear relationships could be observed.

As proxies for local carrying capacities, we used the width of the river at each station (which supposedly reflect the total amount of available habitat [62-64]), as well as the estimated home-range size of each population. The home-range size of each deme was computed as the product of length and width of the river network (including tributaries) delimited by any downstream or upstream weir (see Blanchet et al. 2010 for a description of weirs in this river), that is the water area in which a gudgeon may freely move without

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

encountering any obstacle. Matrices of pairwise *ds*, *di* and *dhm* were then computed from these estimates and were independently confronted to the matrix pairwise *Fst* using multiple regression on distance matrices (MRDM [97]; R package *ecodist*) with 1000 permutations and with *mr* as a covariate. Commonality analyses were then used to disentangle the relative contribution of each predictor to the variance in measures of genetic differentiation. Note that the *ds* matrix based on measures of river width was highly collinear with the *mr* matrix (r = 0.72): the corresponding model was thus omitted because collinearity among predictors are likely to cause important interpretation issues [28]. Details are however provided in Supporting Information (S1 Text).

Finally, we plotted the residuals of the linear regression between *Fst* and the *di* metric (based on measures of river width) against *mr* and inspected the scatterplot to assess migration-drift equilibrium (or lack of it). We further investigated the observed pattern by using piecewise regression to identify the distance threshold at which different linear relationships could be observed. All piecewise regressions were performed using the R package *segmented* [98].

Acknowledgments

The authors thank G. Loot, I. Paz-Vinas, O. Rey and C. Veyssière for their help on the field and in laboratory, as well as K. Saint-Pé for proofreading. The authors also thank the Office Nationale de l'Eau et des Milieux Aquatiques (ONEMA) for their support. Data used in this work were partly produced through the technical facilities of the Centre Méditerranéen Environnement Biodiversité.

References

1. Thomas CD, Kunin WE (1999) The spatial structure of populations. Journal of Animal Ecology 68: 647-657.
2. Hanski I, Gilpin M (1997) Metapopulation Biology: Ecology, Genetics and Evolution. New York: Academic Press.
3. Fahrig L (2003) Effects of habitat fragmentation on biodiversity. Annual Review of Ecology Evolution and Systematics 34: 487-515.
4. Ingvarsson PK (2001) Restoration of genetic variation lost - The genetic rescue hypothesis. Trends in Ecology & Evolution 16: 62-63.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5. Keyghobadi N (2007) The genetic implications of habitat fragmentation for animals. Canadian Journal of Zoology-Revue Canadienne De Zoologie 85: 1049-1064.

6. Broquet T, Ray N, Petit E, Fryxell JM, Burel F (2006) Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). Landscape Ecology 21: 877-889.

7. Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1). Heredity 82: 117-125.

8. Ronce O (2007) How does it feel to be like a rolling stone? Ten questions about dispersal evolution. Annual Review of Ecology Evolution and Systematics. Palo Alto: Annual Reviews. pp. 231-253.

9. Bohonak AJ (1999) Dispersal, gene flow, and population structure. Quarterly Review of Biology 74: 21-45.

10. Sork VL, Nason J, Campbell DR, Fernandez JF (1999) Landscape approaches to historical and contemporary gene flow in plants. Trends in Ecology & Evolution 14: 219-224.

11. Vandewoestijne S, Baguette M (2004) Demographic versus genetic dispersal measures. Population Ecology 46: 281-285.

12. Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163: 1177-1191.

13. Broquet T, Yearsley J, Hirzel AH, Goudet J, Perrin N (2009) Inferring recent migration rates from individual genotypes. Molecular Ecology 18: 1048-1060.

14. Wang JL (2014) Estimation of migration rates from marker-based parentage analysis. Molecular Ecology 23: 3191-3213.

15. Rousset F (2001) Genetic approaches to the estimation of dispersal rates. In: Clobert J, Danchin E, Dhont, Nichols JD, editors. Dispersal. New York: Oxford University Press. pp. 18-28.

16. Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. Molecular Ecology 16: 1149-1166.

17. Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution 18: 189-197.

18. Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, et al. (2007) Putting the 'landscape' in landscape genetics. Heredity 98: 128-142.

19. Holderegger R, Wagner HH (2008) Landscape genetics. Bioscience 58: 199-207.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

20. Pflüger FJ, Balkenhol N (2014) A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. Molecular Ecology 23: 2146-2156.

21. Wright S (1931) Evolution in Mendelian populations. Genetics 16: 0097-0159.

22. Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of population structure. Evolution 38: 1358-1370.

23. Jaquiéry J, Broquet T, Hirzel AH, Yearsley J, Perrin N (2011) Inferring landscape effects on dispersal from genetic distances: how far can we go? Molecular Ecology 20: 692-705.

24. Vos CC, Antonisse-De Jong AG, Goedhart PW, Smulders MJM (2001) Genetic similarity as a measure for connectivity between fragmented populations of the moor frog (Rana arvalis). Heredity 86: 598-608.

25. Zeller KA, McGarigal K, Whiteley AR (2012) Estimating landscape resistance to movement: a review. Landscape Ecology 27: 777-797.

26. Spear SF, Balkenhol N, Fortin MJ, McRae BH, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. Molecular Ecology 19: 3576-3591.

27. ter Braak CJF, Prentice IC (1988) A theory of gradient analysis. Advances in Ecological Research 18: 271-317.

28. Prunier JG, Colyn M, Legendre X, Nimon KF, Flamand MC (2015) Multicollinearity in spatial genetics: Separating the wheat from the chaff using commonality analyses. Molecular Ecology 24: 263-283.

29. Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography 32: 818-830.

30. Wright S (1943) Isolation by distance. Genetics 28: 114-138.

31. Knaapen JP, Scheffer M, Harms B (1992) Estimating habitat isolation in landscape planning. Landscape and Urban Planning 23: 1-16.

32. Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. Evolution 67: 3403-3411.

33. Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting Fst. Nature Reviews Genetics 10: 639-650.

34. Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 53: 1898-1914.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

35. Slatkin M (1985) Gene flow in natural-populations. Annual Review of Ecology and Systematics 16: 393-430.

36. Wright S (1951) The genetical structure of populations. Annals of Eugenics 15: 323-354.

37. Allendorf FW (1986) Genetic drift and the loss of alleles versus heterozygosity. Zoo Biology 5: 181-190.

38. Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. Theoretical Population Biology 12: 253-262.

39. Whitlock MC (2011) G $'_{ST}$ and D do not replace F_{ST}. Molecular Ecology 20: 1083-1091.

40. Serrouya R, Paetkau D, McLellan BN, Boutin S, Campbell M, et al. (2012) Population size and major valleys explain microsatellite variation better than taxonomic units for caribou in western Canada. Molecular Ecology 21: 2588-2601.

41. Graves TA, Beier P, Royle JA (2013) Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. Molecular Ecology 22: 3888-3903.

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

43. Rousset F (2000) Genetic differentiation between individuals. Journal of Evolutionary Biology 13: 58-62.

44. Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity 83: 145-154.

45. Relethford JH (1991) Genetic drift and anthropometric variation in ireland. Human Biology 63: 155-165.

46. Weckworth BV, Musiani M, DeCesare NJ, McDevitt AD, Hebblewhite M, et al. (2013) Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. Proceedings of the Royal Society B-Biological Sciences 280.

47. Mager KH, Colson KE, Groves P, Hundertmark KJ (2014) Population structure over a broad spatial scale driven by nonanthropogenic factors in a wide-ranging migratory mammal, Alaskan caribou. Molecular Ecology 23: 6045-6057.

48. Wang JL (2005) Estimation of effective population sizes from data on genetic markers. Philosophical Transactions of the Royal Society B-Biological Sciences 360: 1395-1409.

49. Whigham PA, Dick GC, Spencer HG (2008) Genetic drift on networks: Ploidy and the time to fixation. Theoretical Population Biology 74: 283-290.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

50. Ryman N, Allendorf FW, Jorde PE, Laikre L, Hossjer O (2014) Samples from subdivided populations yield biased estimates of effective size that overestimate the rate of loss of genetic variation. Molecular Ecology Resources 14: 87-99.

51. Sayre NF (2008) The genesis, history, and limits of carrying capacity. Annals of the Association of American Geographers 98: 120-134.

52. Hanski I (1994) A practical model of metapopulation dynamics. Journal of Animal Ecology 63: 151-162.

53. Gregr EJ, Nichol LM, Watson JC, Ford JKB, Ellis GM (2008) Estimating carrying capacity for sea otters in British Columbia. Journal of Wildlife Management 72: 382-388.

54. Verboom J, Schotman A, Opdam P, Metz JAJ (1991) European nuthatch metapopulations in a fragmented agricultural landscape. Oikos 61: 149-156.

55. Peterman WE, Anderson TL, Ousterhout BH, Drake DL, Semlitsch RD, et al. (2015) Differential dispersal shapes population structure and patterns of genetic differentiation in two sympatric pond breeding salamanders. Conservation Genetics 16: 59-69.

56. Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Molecular Ecology 22: 5983-5999.

57. Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution 68: 1-15.

58. Wang IJ, Bradburd GS (2014) Isolation by environment. Molecular Ecology 23: 5649-5662.

59. Andrew RL, Ostevik KL, Ebert DP, Rieseberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. Molecular Ecology 21: 2078-2091.

60. Papadopulos AST, Kaye M, Devaux C, Hipperson H, Lighten J, et al. (2014) Evaluation of genetic isolation within an island flora reveals unusually widespread local adaptation and supports sympatric speciation. Philosophical Transactions of the Royal Society B-Biological Sciences 369.

61. Gray MM, St Amand P, Bello NM, Galliart MB, Knapp M, et al. (2014) Ecotypes of an ecologically dominant prairie grass (*Andropogon gerardii*) exhibit genetic divergence across the US Midwest grasslands' environmental gradient. Molecular Ecology 23: 6011-6028.

62. Raeymaekers JAM, Raeymaekers D, Koizumi I, Geldof S, Volckaert FAM (2009) Guidelines for restoring connectivity around water mills: a population genetic approach to the management of riverine fish. Journal of Applied Ecology 46: 562-571.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

63. Raeymaekers JAM, Maes GE, Geldof S, Hontis I, Nackaerts K, et al. (2008) Modeling genetic connectivity in sticklebacks as a guideline for river restoration. Evolutionary Applications 1: 475-488.

64. Frankham R (1996) Relationship of genetic variation to population size in wildlife. Conservation Biology 10: 1500-1508.

65. Ray-Mukherjee J, Nimon K, Mukherjee S, Morris DW, Slotow R, et al. (2014) Using commonality analysis in multiple regressions: a tool to decompose regression effects in the face of multicollinearity. Methods in Ecology and Evolution 5: 320-328.

66. Hartl DL, Clark AG (2007) Principles of population genetics. Sunderland, Mass: Sinauer Associates. 652 p.

67. Visconti P, Elkin C (2009) Using connectivity metrics in conservation planning - when does habitat quality matter? Diversity and Distributions 15: 602-612.

68. Fahrig L, Merriam G (1994) Conservation of Fragmented Populations. Conservation Biology 8: 50-59.

69. Matthysen E (2005) Density-dependent dispersal in birds and mammals. Ecography 28: 403-416.

70. Stamps JA (2001) Habitat selection by dispersers: integrating proximate and ultimate approaches. In: Clobert J, Danchin E, Dhondt AA, Nichols JD, editors. New York: Oxford University Press. pp. 230–242.

71. Bonte D, Lukac M, Lens L (2008) Starvation affects pre-dispersal behaviour of Erigone spiders. Basic and Applied Ecology 9: 308-315.

72. Mathieu J, Barot S, Blouin M, Caro G, Decaens T, et al. (2010) Habitat quality, conspecific density, and habitat pre-use affect the dispersal behaviour of two earthworm species, *Aporrectodea icterica* and *Dendrobaena veneta*, in a mesocosm experiment. Soil Biology & Biochemistry 42: 203-209.

73. Meirmans PG (2014) Nonconvergence in Bayesian estimation of migration rates. Molecular Ecology Resources 14: 726-733.

74. Wang JL, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. Genetics 163: 429-446.

75. Blanchet S, Rey O, Etienne R, Lek S, Loot G (2010) Species-specific responses to landscape fragmentation: implications for management strategies. Evolutionary Applications 3: 291-304.

76. Joly P, Miaud C, Lehmann A, Grolet O (2001) Habitat matrix effects on pond occupancy in newts. Conservation Biology 15: 239-248.

77. Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution 29: 1-10.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

78. Ellstrand NC, Elam DR (1993) Population genetic consequences of small population-size - Implications for plant conservation. Annual Review of Ecology and Systematics 24: 217-242.

79. Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? Molecular Ecology 19: 3038-3051.

80. Fischer J, Lindenmayer DB (2007) Landscape modification and habitat fragmentation: a synthesis. Global Ecology and Biogeography 16: 265-280.

81. Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. BMC Bioinformatics 11.

82. Laval G, Excoffier L (2004) SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. Bioinformatics 20: 2485-2487.

83. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10: 564-567.

84. R Development Core Team (2014) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing.

85. Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49: 561-576.

86. Nimon K, Oswald FL, Roberts JK (2013) Interpreting regression effects. R package version 2.0-0. Available: http://cran.r-project.org/web/packages/yhat/index.html

87. Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research 25: 4692-4693.

88. Grenier R, Costedoat C, Chappaz R, Dubut V (2013) Two multiplexed sets of 21 and 18 microsatellites for *Phoxinus phoxinus* (L.) and *Gobio gobio* (L.) developed by cross-species amplification. European Journal of Wildlife Research 59: 291-297.

89. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4: 535-538.

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

91. Rice WR (1989) Analysing tables of statistical tests. Evolution 43: 223-225.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

92. Prunier JG, Kaufmann B, Fenet S, Picard D, Pompanon F, et al. (2013) Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: towards a better assessment of functional connectivity using an individual-based sampling scheme. Molecular Ecology 22: 5516-5530.

93. MATLAB and Statistics Toolbox Release 2010a TM, Inc., Natick, Massachusetts, United States.

94. Pella H, Lejot J, Lamouroux N, Snelder T (2012) Le réseau hydrographique théorique (RHT) français et ses attributs environnementaux. Géomorphologie : relief, processus, environnement 3.

95. Toms JD, Lesperance ML (2003) Piecewise regression: A tool for identifying ecological thresholds. Ecology 84: 2034-2041.

96. Rey O, Turgeon J (2007) Influence of historical events and contemporary estuarine circulation on the genetic structure of the banded killifish (Fundulus diaphanus) in the St. Lawrence River (Quebec, Canada). Canadian Journal of Zoology-Revue Canadienne De Zoologie 85: 891-901.

97. Lichstein JW (2007) Multiple regression on distance matrices: a multivariate spatial analysis tool. Plant Ecology 188: 117-131.

98. Muggeo VMR (2008) segmented: an R Package to Fit Regression Models with Broken-Line Relationships. R News 8: 20-25.

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supporting Information

S1 Fig. Theoretical distribution of 10000 demes' population sizes randomly picked from a truncated normal distribution with a mean of 40 and a standard deviation of 200, values being bounded between 40 and 1000.

S2 Fig. Network structures and migration models used in simulated datasets.

S3 Fig. Theoretical distribution of Pearson's r correlation values between the effective population sizes N used for simulations and local carrying capacities K used to compute genetic drift metrics for an uncertainty parameter α picked from an uniform distribution ranging from -0.9 to 0.9, in 10000 sixteen-deme datasets.

S4 Fig. Distribution of average estimates of migration rates in 49 recent empirical studies (See S2 Table).

S1 Table. Summary data for the 11 microsatellite loci used in the empirical dataset. Number of alleles (A; in brackets, effective number of alleles), observed and expected heterozygosity (*Ho* and *He*) and fixation index (*f*) are given for each locus for the *N* individuals collected from each of the 19 sampled populations. Genebank ID are provided in brackets, below the locus name.

S2 Table. Summary statistics of the 49 empirical studies considered to estimate the range of migration rates likely to be encountered in natural systems (from [73]).

S1 Text. Empirical results from the model $Fst = ds + mr$, with *ds* computed from river width.

S2 Text. Selection criteria for empirical studies cited in the literature survey by Patrick G. Meirmans [73] (See also S2 Table).

S1 Dataset. Raw data used to investigate the match between *Fst* values and each IBDr metric in a simple two-deme situation (Fig 2). For each simulation (Sim), the file provides effective population sizes of population 1 (N1) and population 2 (N2), as well as pairwise FST (FST).

S2 Dataset. Raw data used to assess the explanatory power of each IBDr metric as a function of the migration rate *m* in a simple two-deme situation (Fig 3). For each simulation (Sim), the file provides the migration rate (MIG), effective population sizes of population 1 (N1) and population 2 (N2) as well as pairwise FST (FST).

S3 Dataset. Structure of pairwise matrices of migration rates used to assess the strength of each IBDr metric compared to a resistance metric in four complex situations differing according to both the network structure and the migration model.

S4 Dataset. Raw data used to assess the strength of each IBDr metric compared to a resistance metric in a one-dimensional 16-deme linear network with stepping-stone migration and the α parameter set to 0 (i.e., N = K; Fig4 a, e, i). For each simulation (Sim), the file provides the migration rate (MIG), the effective population size

bioRxiv preprint doi: https://doi.org/10.1101/031633; this version posted November 13, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(N1-N16) and the estimated carrying capacity (K1-K16) of each population, as well as pairwise FST (from FST_2_1 to FST_16_15).

S5 Dataset. Raw data used to assess the strength of each IBDr metric compared to a resistance metric in a one-dimensional 16-deme linear network with spatially limited dispersal and the α parameter set to 0 (i.e., N = K; Fig4 b, f, j). Details as in S4 Data.

S6 Dataset. Raw data used to assess the strength of each IBDr metric compared to a resistance metric in a two-dimensional 16-deme lattice network with stepping-stone migration and the α parameter set to 0 (i.e., N = K; Fig4 c, g, k). Details as in S4 Data.

S7 Dataset. Raw data used to assess the strength of each IBDr metric compared to a resistance metric in a two-dimensional 16-deme lattice network with spatially limited dispersal and the α parameter set to 0 (i.e., N = K; Fig4 d, h, l). Details as in S4 Data.

S8 Dataset. Raw data used to assess the strength of each IBDr metric compared to a resistance metric in a one-dimensional 16-deme linear network with stepping-stone migration and the α parameter picked from a uniform distribution ranging from -0.9 to 0.9 (i.e., N ≠ K; Fig5 a, e, i). For each simulation (Sim), the file provides the migration rate (MIG), the effective population size (N1-N16), the α parameter (NOISE1-NOISE16) and the estimated carrying capacity (K1-K16) of each population, as well as pairwise FST (from FST_2_1 to FST_16_15).

S9 Dataset. Raw data used to assess the strength of each IBDr metric compared to a resistance metric in a one-dimensional 16-deme linear network with spatially limited dispersal and the α parameter picked from a uniform distribution ranging from -0.9 to 0.9 (i.e., N ≠ K; Fig5 b, f, j). Details as in S8 Data.

S10 Dataset. Raw data used to assess the strength of each IBDr metric compared to a resistance metric in a two-dimensional 16-deme lattice network with stepping-stone migration and the α parameter picked from a uniform distribution ranging from -0.9 to 0.9 (i.e., N ≠ K; Fig5 c, g, k). Details as in S8 Data.

S11 Dataset. Raw data used to assess the strength of each IBDr metric compared to a resistance metric in a two-dimensional 16-deme lattice network with spatially limited dispersal and the α picked from a uniform distribution ranging from -0.9 to 0.9 (i.e., N ≠ K; Fig5 d, h, l). Details as in S8 Data.

S12 Dataset. Empirical data. For each population, the "landscapedata" sheet provides the width of the river at sampling point (RiverWidth, in m), the size of the home-range at sampling point (HomeRange, in m^2), the distance from the source, used to computed riverine distances (DistFromSource, in m). The "Geneticdata" sheet provides the pairwise matrix of FST.