



**HAL**  
open science

# A stochastic dispersal-limited trait-based model of community dynamics

Franck Jabot

► **To cite this version:**

Franck Jabot. A stochastic dispersal-limited trait-based model of community dynamics. *Journal of Theoretical Biology*, 2010, 262 (4), pp.650. 10.1016/j.jtbi.2009.11.004 . hal-00564087

**HAL Id: hal-00564087**

**<https://hal.science/hal-00564087>**

Submitted on 8 Feb 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Author's Accepted Manuscript

A stochastic dispersal-limited trait-based model of community dynamics

Franck Jabot

PII: S0022-5193(09)00530-X  
DOI: doi:10.1016/j.jtbi.2009.11.004  
Reference: YJTBI5769

To appear in: *Journal of Theoretical Biology*

Received date: 4 May 2009  
Revised date: 29 September 2009  
Accepted date: 6 November 2009

Cite this article as: Franck Jabot, A stochastic dispersal-limited trait-based model of community dynamics, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2009.11.004](https://doi.org/10.1016/j.jtbi.2009.11.004)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



[www.elsevier.com/locate/jtbi](http://www.elsevier.com/locate/jtbi)

A stochastic dispersal-limited trait-based model of community dynamics

Franck Jabot<sup>a,b</sup>

<sup>a</sup>Laboratoire Evolution et Diversité Biologique, CNRS, Université Paul Sabatier

Bâtiment 4R3

31062 Toulouse cedex 4

France

<sup>b</sup>AgroParisTech ENGREF

19 avenue du Maine

75732 Paris

France

E-mail: [franck.jabot@m4x.org](mailto:franck.jabot@m4x.org)

Tel: (+33) 5 61 55 67 60

Fax: (+33) 5 61 55 73 27

Running Head: Trait-based model of community dynamics

Words in abstract: 204

Words in main text: 6486

**Abstract**

I present a model of stochastic community dynamics in which death occurs randomly in the community, propagules disperse randomly from a regional pool, and recruitment of new individuals of a species is proportional to the species local abundance multiplied by its local competitive ability. The competitive ability of a species is assumed to be determined by a function of one trait of the species, and I call this function the environmental filtering function. I show that information on local species abundances in a network of plots, together with trait data for each species, enables the inference of both the immigration rate and the environmental filtering function in each plot. I further study how the diversity patterns produced by this model deviate from the neutral predictions, and how this deviation depends on the characteristics of the environmental filtering function. I show that this inference framework is more powerful at detecting trait-based environmental filtering than existing statistical approaches based on trait distributions, and discuss how the predictions of this model could be used to assess environmental heterogeneity in a plot, to detect functionally meaningful trade-offs among species traits, and to test the assumption that there exists a simple relationship between species traits and local competitive ability.

Keywords: Community model, functional trait, environmental filtering, Approximate Bayesian Computation.

## Introduction

Understanding the dynamics of communities is a challenging task for ecologists due to the diversity of ecological mechanisms and the ways in which they potentially interact (Chesson 2000). Different models of community dynamics have primarily been compared through the analysis of diversity patterns (Chave et al. 2002, 2006, Etienne and Olf 2005, Volkov et al. 2005). Due to the limited information content of such data (McGill et al. 2007), there is a need for additional sources of information to robustly contrast community dynamics models. Two main sources of information have been the focus of the recent efforts: phylogenies (Webb et al. 2002, Jabot and Chave 2009), and species traits (Lavorel and Garnier 2002, McGill et al. 2006, Westoby et al. 2002), both lines of research being intimately linked (Cavender-Bares et al. 2004, Webb et al. 2008). In this study, I focus on how one may make better use of species functional traits to compare the predictions of different community models in terms of trait distribution in local communities.

Species traits are thought to capture components of species ecological strategies (Westoby et al. 2002, Wright et al. 2004). If two species have similar traits, they are thus likely to share similar environmental affinities (Diaz et al. 1998, Weiher and Keddy 1995), and to be in competition for the same resources (Dayan and Simberloff 2005, Fargione et al. 2003). Data on species traits have been used to disentangle patterns attributed to stochastic or mass effects (Mouquet and Loreau 2003), from those more likely due to niche effects in community assembly (Cadotte 2007, Cornwell et al. 2006, Fukami et al. 2005, Kraft et al. 2008, Mouillot et al. 2007, Ricklefs and Travis 1980). If the environment acts as a filter for species by preventing them from establishing in a community, then the traits of coexisting species should be more similar than in randomly assembled communities. On the other hand, competition for resources may

lead to the exclusion of the most similar species (limiting similarity, MacArthur and Levins 1967). This limiting similarity should produce a statistical overdispersion of the traits present in a community. In species-rich communities like tropical forests, where understanding community dynamics is particularly challenging, the inclusion of data on species traits has proven fruitful: easily-measurable traits have been shown to explain roughly half of the variation in growth and mortality of trees (Poorter et al. 2008), and a significant part of their ranges and local abundances (Baltzer et al. 2008, Engelbrecht et al. 2007). This indicates that traits do bring relevant information in our understanding of community dynamics (Grime 2006, McGill et al. 2006, Westoby and Wright 2006).

In this quest for a trait-based ecology, various statistical methods have been developed to relate the trait composition of a community to putative ecological mechanisms of community assembly. Tests for competition (Cornwell and Ackerly 2009, Kraft et al. 2008, Stubbs and Wilson 2004), and for environmental filtering (Cornwell et al. 2006, Villéger et al. 2008), enable to detect significant departures from a null model of random community assembly. The fourth-corner method (Dray and Legendre 2008, Legendre et al. 1997) relates traits to environmental features in order to explain species presence or abundances in given environmental conditions. Despite the rich insights brought by these approaches, they all suffer from a common shortcoming. Because they are not based on an explicit model of community dynamics, they are unable to quantify the impact of environmental constraints on basic properties of community dynamics such as immigration rate and difference in competitive ability among species. Also, they ignore stochasticity (Hubbell 2001), although the real community dynamics is certainly stochastic (Ellner & Fussmann 2003, Erjonaes et al. 2006, Fukami et al. 2005).

Consequently, current methods for analyzing trait distributions in communities do not enable ecologists to respond adequately to the claim of others that differences among

species may have a minor role in community dynamics compared to stochastic factors (Hubbell 2001). Indeed, detecting subtle non-random patterns of trait distribution in natural communities does not imply that the effect of traits on community dynamics is strong. If so, taking into account trait differences among species in such communities may not bring any key improvement.

To tackle these limitations, I present a method to quantify both immigration rate and environmental filtering in ecological communities. This approach is based on the stochastic modeling of a local community dynamics at the individual level. In this model, the community is subject to stochastic birth-death-immigration events, as in the neutral model of Hubbell (2001), but each individual has a local competitive ability determined by one of its traits (Schwilk and Ackerly 2005). This integration of two frameworks, trait-based competition and neutral theory, aims at capturing the compound effects of stochasticity, dispersal limitation and species differences in the structure of natural communities (Adler et al. 2007, Gravel et al. 2006, Holyoak and Loreau 2006). I assume that each individual in the community is experiencing the same environment, thus I neglect the effects of habitat variation within a local community. I call the function that relates a trait value to local competitive ability the environmental filtering function. This function is a property of the local environment, which acts as a filter for species depending on their traits. This environmental filtering function can be seen as emerging from classical species niche models on an environmental axis (e.g. Schwilk and Ackerly 2005, see Appendix A for more details). If the filtering function is flat in trait space, then the community will be subject to purely neutral dynamics. If the environmental filtering function is unimodal in trait space, then the community will experience environmental filtering in which species with traits close to the peak experience a maximal recruitment rate (Fig. 1a). In the particular case where the peak of

the filtering function is positioned either on the minimal or on the maximal trait value, then the filtering function is modeling environmental filtering for an extreme trait value. The resulting trait-local abundance relationship then depends on the relative strength of this environmental filtering and of the immigration rate and demographic stochasticity (Fig. 1b, c, d). By modeling the environmental filtering function with a Gaussian distribution, I develop an inference method for estimating the immigration rate and the height, width and position of the filtering function in a local community with data on species abundances and traits. I test this method on simulated data, and show that all the parameters of this model may be inferred. I explore how the height of the filtering function – a measure of the maximal competitive differences between species, and the width of this filtering function – a measure of the specificity of the filtering, contribute to producing either neutral or non-neutral diversity patterns, and how the inference method proposed here compares with other approaches for detecting such deviations from neutrality. Finally, I test whether a multimodal filtering function, which may arise if several micro-habitats are filtering the species within a plot, may be detected with this method.

## Methods

### *The Model*

To model local community dynamics, each species are given a local competitive ability as determined by a filtering function  $F$ . More specifically,  $F$  is modeled by a Gaussian function of one quantitative trait  $t$ :  $F(t) = 1 + A * \exp[-(t-h)^2/2\sigma^2]$ , with the following three parameters:  $A$  is the maximal competitive advantage (see below) possible for a species, compared to the worst competitive species;  $h$  is the optimal trait value in the local environment; and  $\sigma$  describes the width of the filtering function in the local environment (the greater  $\sigma$ , the larger the range of trait values leading to enhanced

competitive ability). This model neglects intraspecific variability in trait value and competitive ability. This serious simplification is in line with many field studies, where traits data are often collected at the species level rather than at the individual level (but see Paine et al. in prep).

At each time step, one individual drawn at random dies in a local community of size  $J$ . It is replaced either by an immigrant from the regional pool with probability  $I/(H+J-1)$  (Etienne and Olff 2004) or by the descendant of a local individual. Parameter  $I$  is the fourth parameter of the model, and it measures the amount of immigration from the regional pool into the local community (assuming that each local individual produces one propagule,  $I$  is the number of additional propagules coming from the regional pool and competing for the empty site of the dead individual). The probability that the replacing individual is of species  $i$  is proportional to the abundance of this species in the local community (or in the regional pool if it is an immigrant), multiplied by its local competitive ability  $F_i$ . Hence, the environmental filtering is solely acting at the recruitment stage, and the propagules of a species having a trait value close to  $h$  have a greater recruitment probability. Regional species' relative abundances  $\chi_i$  are assumed to be known and they vary little over ecological time-scales. In applications, regional abundances may be approximated by the species abundances pooled over all community samples available (Jabot et al. 2008). This local dynamics leads to a dynamic equilibrium (data not shown) between immigration, competition, and ecological drift – ie. the random fluctuation of abundances due to demographic stochasticity. The proposed approach models environmental filtering as genic selection is modeled in population genetics.

*Likelihood formula*

Assuming that we know the species relative regional abundances  $\chi_i$ , we aim at inferring the parameters of the model ( $I$ ,  $A$ ,  $h$  and  $\sigma$ ) from data on species traits and on their local abundances in a sample. For this model, a likelihood formula is available in the population genetics literature. In noting  $\chi_i$ ,  $x_i$  and  $F_i$  the regional relative abundance, the local relative abundance and the local competitive ability of species  $i$ , and  $S$  the total number of species in the regional pool, one obtain by analogy with Wright's formula (Wright 1931, 1937) the likelihood:

$$L(I, A, h, \sigma) = C \bar{w}^J \prod_{i=1}^S x_i^{I \chi_i}$$

where  $\bar{w} = \sum_{i=1}^S x_i F_i$ , and  $C = 1 / \int_{x_i} \bar{w}^J \prod_{i=1}^S x_i^{I \chi_i}$ , the integral being over all the possible  $x_i$  distributions.

The value of  $C$  must be computed in order to calculate the likelihood for each parameter. One approach is to estimate  $C$  by Monte-Carlo simulations. However a preliminary implementation of this approach showed that this solution is time-consuming. A simplification of this formula is available when only two competitive ability values are possible in the population (Mark Beaumont, pers. comm.). To my knowledge, no such simplification is available in the general case of interest here. I therefore used an approximate method.

#### *Parameter inference using Approximate Bayesian Computation (ABC)*

I used a computer-intensive approach called ABC. The parameters are drawn at random from a prior distribution. Communities are then simulated according to the model using these parameter values, and a set of summary statistics are computed for these simulated communities. Only those simulations for which the summary statistics are close to the

empirical ones (ie computed from the data) are retained. The parameters associated with these selected simulations are used to construct an approximate posterior distribution (Beaumont et al. 2002, and Jabot & Chave 2009 for an application to ecology).

In each simulation, local communities were initialized by randomly drawing  $J$  individuals from the regional pool with the probability of drawing species  $i$  proportional to  $\chi_i * F_i$ . A forward-in-time dynamics was then simulated during  $\mathcal{P}$  time steps, a time sufficient for ensuring that the dynamic equilibrium was reached (see Appendix B). I used four summary statistics to represent the communities: the species richness  $S$ , and Shannon's index  $H$  (already used in ecological applications of ABC, Jabot & Chave 2009), the mean trait value  $MTV$  among individuals in the community, and the skewness of the distribution of trait values  $STV$  among individuals. These summary statistics were chosen based on a preliminary analysis:  $S$  was found to capture the total amount of filtering due to both dispersal limitation and environmental filtering.  $H$  provides relevant information on the relative importance of these two effects.  $MTV$  informs on the optimal trait value and  $STV$  on the width of the environmental filtering function (see Box 1 for more details).

#### *Test of the method by simulations*

To test whether the ABC method retrieved the parameters of numerically generated datasets, I simulated 100,000 communities of size  $J=500$ , and species richness  $S$  ranging between 90 and 110. These values of sample size and species richness are typical of tropical forest tree inventories of 1 ha, a major source of biological data and of inspiration for community ecology theory (e.g. Hubbell 2001). Due to the stochasticity of the dynamics, the species richness of a simulated community is a random number. I therefore simulated many more communities than needed and only retained those that

had the targeted richness. A regional pool of 1,000 species was defined. Species had equal regional abundances and were evenly spaced on the trait axis between 0.1 and 100. I used a uniform prior distribution for the four parameters:  $\ln(J)$  in  $[3 ; 5]$ ,  $\ln(A)$  in  $[\ln(0.1) ; \ln(5)]$ ,  $\ln(\sigma)$  in  $[\ln(0.5) ; \ln(25)]$ , and  $h$  in  $[-25 ; 125]$ . I extended the prior on  $h$  beyond the minimal and maximal trait values in the communities, so that there is no boundary effect for the estimation of communities filtered for trait values around 0 or 100. The prior on  $\ln(J)$  was chosen so that the targeted richness was likely to be obtained in simulations (see Box 1, Fig. 2a). I then picked at random 100 out of these 100,000 communities, and considered them as artificial datasets on which to test the inference method. I used for each of these artificial datasets the 99,999 remaining simulated communities to infer the four model parameters by ABC. I then computed a correlation coefficient between the values of the parameters used to simulate these 100 artificial datasets and the parameters values inferred by ABC for these 100 datasets.

The robustness of these results to the various model assumptions was also assessed: the number of communities was varied in the ABC procedure (10,000 , 20,000 , 50,000 , and 200,000), as well as the size of the regional pool (300 and 3,000 species), and the local community richness ( $S=20, 50, \text{ and } 200$ ). The regional relative abundances were also modeled by a logseries distribution with parameter  $\alpha=200$ , using the function “fisher.ecosystem” of the R package `untb` (Hankin 2007, R Development Core Team 2008), and relative abundances were either assigned randomly, or so that regionally abundant species were aggregated in the trait space. This aggregation was performed by picking at random a position in the trait space, and then placing the species around this trait value in decreasing order of abundances so as to form a peak in the trait-regional abundance plot. Two other shapes of the environmental filtering function were also tested: a rectangular function and a triangular one (Appendix C).

*Test of environmental filtering at the individual level*

Previously, most trait-based tests of environmental filtering have ignored the information on relative species abundances in a local community. To take into account this information together with the information on species traits, one can use the dispersal-limited neutral model as a null model, and compare some summary statistics of neutral and filtered communities to look for differences. To determine which summary statistics was more able to distinguish neutral and filtered communities, I simulated 1000 neutral communities with  $J=500$  and  $S=100$  with parameter  $\ln(J)$  drawn at random in  $[3 ; 5]$ , and 1000 communities simulated with environmental filtering with  $h$  equal to 25,  $\ln(J)$  randomly drawn in  $[3 ; 5]$ ,  $\ln(A)$  in  $[\ln(0.1) ; \ln(5)]$ , and  $\ln(\sigma)$  in  $[\ln(0.5) ; \ln(25)]$ . I compared the value of four summary statistics among neutral and filtered communities: Shannon's index  $H$ , the range of trait values in the local community  $Range$ , the range of trait values in the 5 locally most abundant species  $Range5$ , and the variance of trait values among individuals  $VTV$ . Comparing neutral and filtered communities with the statistics  $Range$  actually corresponds to the method of Cornwell et al. (2006). It does not make use of species abundances, only of species traits and of their presence or absence in the local community. The statistics  $H$  uses solely information on species abundances. The two last statistics combine information on species traits and abundances. The same analysis was performed in other situations: the size of the regional pool was varied (300 and 3,000 species), as well as the local community richness ( $S=20, 50, \text{ and } 200$ ), the regional abundance structure, and the shape of the environmental filtering function (Appendix C).

*Detection of micro-habitats*

The model assumes that the local environment is homogeneous. This assumption is

likely to be violated in real communities where habitats are heterogeneous at different spatial scales (Palmer and Dixon 1990). In the presence of micro-habitats within a sample, the environmental filtering function  $F$  may be multiply peaked. In this case, the studied single-peak model should be unable to precisely reproduce the trait-abundance relationship observed in real communities, and this discrepancy between simulated single-peak communities and real datasets might be captured in some summary statistics. One potentially good summary statistic to assess the presence of multiple micro-habitats in real communities is the variance of trait values among individuals  $VTV$ , since multiple peaks in the filtering function should increase  $VTV$ , everything else being equal. To test this idea, I simulated 100 artificial datasets using two peaks in the filtering function at  $h_1=25$  and  $h_2=75$ . The filtering function is then equal to  $F = 1 + A * (\exp[-(t - h_1)^2/2\sigma^2] + \exp[-(t - h_2)^2/2\sigma^2])$ . I used a value of  $A=3$ ,  $\sigma=1$ , and  $\ln(I)$  randomly drawn in  $[3:5]$ . In each of these 100 artificial datasets, I computed  $VTV$ . Then, for each artificial dataset, the four model parameters were inferred by ABC, and the 200 best-fit communities, which were simulated with single-peak environmental filtering functions, were retained.  $VTV$  was computed in all these best-fit (single-peak) communities, and compared with  $VTV$  values observed in artificial (2-peak) datasets. More precisely, a statistical test was designed, consisting in computing for each artificial dataset the percentage of best-fit communities having  $VTV$  values larger than in the artificial dataset. This percentage constitutes the  $p$ -value for the rejection of the null hypothesis: “the artificial dataset is well described by a single-peak environmental filtering function”.

## Results

### *ABC inference*

Based on 100 communities of richness  $S$  equal to 100, simulated with parameter values

drawn from the prior distribution, the correlation coefficients  $R^2$  between simulated and inferred parameter values were equal to 0.64, 0.70, 0.57 and 0.54 for parameters  $\ln(J)$ ,  $\ln(A)$ ,  $h$  and  $\ln(\sigma)$  respectively (Fig. 3). It was difficult to properly infer parameters in the communities that were either weakly filtered (Fig. 4a), or filtered for trait values close to the regional mean (Fig. 4c). Once these problematic communities were taken out of the correlation analysis, correlation coefficients rose to 0.65, 0.99, 0.73 and 0.82 for the parameters  $\ln(J)$ ,  $\ln(A)$ ,  $h$  and  $\ln(\sigma)$  respectively (see Appendix C for more details on the detection of outliers).

In situations where parameter inference is unreliable, the width of the approximate posterior distribution is larger (Fig. 4b, d, e). It is thus possible to assess the reliability of any parameter inference from real data by comparing the width of their approximate posterior distribution to the ones of well-estimated simulated communities. These results were robust to variations in the regional pool richness, in the structure of regional species abundances, and in the shape of the environmental filtering function (Appendix C). Similar results were also obtained in communities of local richness  $S=50$ , and  $S=200$ . For communities with small local richness  $S=20$ , parameter inference performed poorly (Appendix C). This inference method is thus better suited for species-rich communities. Additional tests with communities with random local species richness are reported in Appendix C, and corroborate these results.

#### *Test of environmental filtering at the individual level*

When environmental filtering was simulated in artificial communities, the values of  $H$ ,  $Range5$ , and  $VTV$  were lower than in neutral communities, but not the range of trait values (Fig. 5). Filtered communities had increasingly non-neutral patterns with increasing  $A$  (Fig. 5). The effect of increasing  $\sigma$  was more complex, and depended on

the statistics considered. Figure 5 shows that, for the simulated communities,  $H$ -based statistical tests were the most efficient for detecting deviation from neutrality. Similar results were found with different regional pool richness, local community richness, and shapes of environmental filtering functions (see Appendix C for more details). When species regional abundances were unequal (modeled by a logseries distribution with parameter  $\alpha=200$ ) and uncorrelated with species traits, similar results were again obtained (Fig. C7). When species regional abundances were correlated with species trait values, different results were found, except for  $H$ -based tests of environmental filtering which were quite robust (Fig. C8).

#### *Detection of micro-habitats*

$VTV$  was generally larger in the artificial datasets simulated with two peaks in the environmental filtering function than in the 1-peak communities that best fitted these artificial datasets. This increase in  $VTV$  value was found to be significant, at an  $\alpha$  level of 5%, in 60% of the artificial datasets (Fig. 6).

## **Discussion**

#### *Integrating niche and neutral processes in an inferential framework*

Hubbell's neutral theory of biodiversity is advocated to be an ideal anchoring point for building a quantitative theory of community dynamics (Adler et al. 2007, Alonso et al. 2006, Gravel et al. 2006, Holyoak and Loreau 2006). A natural next step for the stochastic modeling of communities is to add environmental filtering to the mechanisms of dispersal limitation and ecological drift already present in neutral theory (Schwilk and Ackerly 2005), in order to see whether species differences need to be taken into account to understand natural communities. In the present work, I contribute to this progress by modeling local asymmetrical competitive ability among species in a

background of ecological drift, just as selection is modeled in population genetics against a background of genetic drift (Wright 1931, 1937). I assume that the local competitive abilities of species are determined by one trait, and an environmental filtering function is defined, which relates the trait of a given species to its local competitive ability (Fig. 1). Using simulated data, I show that it is possible to infer the parameters of this filtering function – position of the peak in the trait space, height and width – together with the immigration rate from the regional pool, from the knowledge of species traits and of their abundances.

*Understanding the influence of environmental filtering on communities*

Previous attempts to measure the intensity of environmental filtering from species traits were mainly based on the study of the convex volume formed by the positions in the traits space of the species present in the focal community (Cornwell et al. 2006). The present approach describes environmental filtering by three parameters:  $h$  measures the position of the optimal trait value,  $A$  measures the strength of the asymmetry among species (Alonso et al. 2008), and  $\sigma$  measures the specificity of the environmental filtering. Using simulations, the overall intensity of filtering was found to be a compound effect of these three parameters, and mostly of the parameters  $A$  and  $\sigma$ . If one considers a species different from the most competitive species, the higher  $A$ , the larger is the difference in competitive ability between the most competitive species and the focal species, the reverse being true for  $\sigma$ . Thus an increase in  $A$  tends to make species abundances less even, while an increase in  $\sigma$  produces the reverse effect (Fig. 5a). Hence, a given level of abundance unevenness can be achieved either by a strong competitive advantage of a large number of species, or by a weaker advantage of a fewer number of species, as long as the ratio  $A/\sigma$  is roughly constant (Fig. 2b). While this particular criterion of constant  $A/\sigma$  (Box 1) is only valid for the particular model

developed here, the fact that filtering intensity and filtering specificity have opposite effects on species abundance distributions is likely to be much more general (Walker 2007, Whittaker 1965). The present study further points out that although two filters, one with large  $A$  and  $\sigma$ , and the other one with small  $A$  and  $\sigma$ , can lead to similar species abundance distributions, the trait distributions associated with these different filters should be different, hence providing a way to understand how environment is filtering species in ecological communities.

This observation is likely to have important consequences for the investigation of species habitat associations in species-rich communities. Such investigations are generally based on the study of species occurrence and/or abundance in networks of plots (or in partitions of large plots into many subplots), whose abiotic environmental properties are characterized. In such studies, a species is considered to be associated with a given habitat if it is significantly over-represented in this habitat (Dufrene and Legendre 1997, Harms et al. 2001, Plotkin et al. 2000). It is shown here that even if a species has a strong competitive advantage compared to poorly adapted species, it may fail to be over-represented locally, if many other species present similar adaptations. In this case, the role of historical contingency will be primordial in explaining the presence or absence of over-representation in habitats where a species is a good competitor (Purves and Pacala 2005).

To decide whether a particular locally rare species is adapted or not to this local habitat, one needs to increase the sample size until this species has a sufficient total abundance across samples. This is immensely challenging for most species in mega-rich communities which are rare at the regional scale (eg. Pitman et al. 2001). This study suggests that an alternative strategy may be employed. For instance, one may look

where the trait value of the locally rare species is located compared to the inferred environmental filtering function: if it is close to the peak of the filtering function, then the focal species is likely to be locally rare due to historical contingency rather than environmental filtering.

#### *Applicability of the inference method*

The data required to apply the present model are species abundances and species-level trait data in a network of plots belonging to the same region. The present model provides a sampling theory for analyzing such biodiversity data (Alonso et al. 2006). It enables filtering effects to be quantified, taking into account the fact that only a limited sample of the all community is available, and thus that some species may be missed in the sample not because they are absent from the community, but because the sample is of limited size. In the tests of the method, I used sample sizes typically encountered in 1-ha tropical forest tree plots, thus the method should be powerful enough to make robust inferences with already existing datasets (Kleyer et al. 2008, Kraft et al. 2008, Prinzig et al. 2008). The present inference method does not work well for species-poor communities (Appendix C), but for species richness typically encountered in tropical forests, it has been shown to be efficient (Fig. 3). Although parameter inference is not robust in some parts of the parameter space (Fig. 4), it is possible to know where the method will fail to apply for a real dataset. Indeed, by using a subset of the ABC simulations as reference communities, one can obtain the typical volume of the posterior distribution in well-estimated communities, and compare the volume obtained with real data to these reference volumes: if the posterior distribution obtained with real data has a larger volume, then the inference should not be trusted. Logically, the inference does not work well when the filtering intensity is weak. More interestingly, the inference does not work well when environmental filtering selects species for trait

values close to the regional mean (ie in habitats which are widespread). In those cases, both the signals in the trait distribution produced by immigration and by environmental filtering are symmetric around the mean regional trait value, and it is therefore not possible to know the level of specificity of this environmental filtering (Box 1). For these problematic common habitats, the consideration of additional summary statistics measuring the similarity in traits among plots might help disentangle how specificity and intensity of filtering are interacting.

*Comparison with other approaches for analyzing the trait composition of communities*

The more widely used approach for assessing environmental filtering from the distribution of species traits in a community consists in computing the convex hull volume which contains all locally present species positioned in the traits space (Cornwell et al. 2006, Kraft et al. 2008). The more communities are filtered, the smaller this volume. However, rare species with atypical traits largely contribute to the volume obtained, and can drastically decrease the power of this approach to detect environmental filtering. More refined analyses based not only on convex volumes but also on the distribution of individuals in this volume have been proposed (eg. Villéger et al. 2008). This may increase the power of such analyses for detecting and characterizing environmental filtering. Indeed, I found here that the structure of species abundances, and its correlation with species traits, was far more informative than the sole consideration of the range of species traits in a community for detecting environmental filtering (Fig. 5). Furthermore, approaches based on convex volumes and their extensions do not provide any robust quantification of this filtering in terms of community dynamics. The present approach provides a clear quantification of environmental filtering coupled with a more detailed description of this ecological

mechanism. It is based on a modeling of community dynamics, based at the individual level. Being parameterizable from field data, this model can be used to predict the dynamics of ecological communities in a more data-informed way.

The results of this study further question the rationale of range-based (or its multidimensional convex-volume equivalent) approaches for detecting environmental filtering in species-rich communities. Indeed, regardless of its intensity, environmental filtering did not produce any range contraction compared to ranges observed in neutrally-assembled communities of similar richness, and this result was robust to variations in local and regional richness, in regional abundance structure, and in the shape of the filtering function (Appendix C). The reason for this is that, when observing a community with a given species richness, if this community is strongly environmentally filtered, then it has to have a larger immigration rate from the regional pool than a neutrally assembled community (to counter-balance the effect of filtering on species richness, see Box 1). This immigration brings continually less adapted species into the community, and this leads to wide trait ranges, which are undistinguishable from those found in neutral communities. This observation is likely linked to the modeling choice of setting the minimal competitive ability to 1. If this minimal competitive ability had been set equal to 0, range contraction might have certainly been observed in filtered communities as suggested by the results of Schwilck and Ackerly (2005). What distinguishes these two modeling choices? Setting minimal competitive ability to 0 implies that some species will be completely excluded from some habitats, and consequently that the study is encompassing several ecological guilds (*sensu* Hubbell 2001). On the contrary, setting the minimal competitive ability to 1 implies that every species will be likely to be present (although possibly rare) in every habitat. This is what seems to be observed in most tropical forest habitats (Hubbell 2001). This

distinction might explain why trait range-based studies of community assembly were able to detect environmental filtering in communities with limited species richness, and strong habitat heterogeneity (eg. Cornwell and Ackerly 2009, Cornwell et al. 2006, Pausas and Verdu 2008), while studies in more diverse communities, and in more homogeneous habitats, like tropical forests, have difficulties to detect reduction in trait ranges, even when differences in trait community means are detected (Kraft et al. 2008, Paine et al. in prep).

#### *Detection of local heterogeneity*

In real communities, the assumption that environmental conditions are homogeneous is unlikely to be valid (Palmer & Dixon 1990). The presence of different micro-habitats should enable different ecological strategies to coexist, and should decrease the detectability of environmental filtering. Other ecological mechanisms such as frequency-dependence (Connell 1971, Janzen 1970), limiting similarity (MacArthur & Levins 1967), storage effect (Chesson 2000), or disturbance (Grime 2006) may have similar effects on the detectability of environmental filtering, by decreasing the correlation between trait value and competitive advantage. These problems are not specific to this particular approach but common to all attempts to detect environmental filtering from trait data. A particular advantage of the present method is that it is possible to detect anomalies in the distribution of individuals in the trait space that are not explained by the model. For instance, bimodality in the trait distribution may be due to the presence of two micro-habitats, or it may only be due to stochasticity. I show here that the present approach enables to disentangle these two explanations (Fig. 6). The detection of such anomalies may inform us on the more complex features of local community structuring.

*Perspectives*

The present method enables the environmental filtering associated to a single trait (or a PCA axis) to be quantified. Hence, one can assess the importance of each trait or trait axis in the filtering of species from the community. A drawback of treating each trait independently is that it does not lead to a dynamical community model parameterized with multiple traits, which could be used for predicting community dynamics from static data. An extension of this model to use multiple traits simultaneously would thus be very valuable. Preliminary works on this topic suggest that such an extension might be performed with the same ABC approach, but this will require a careful choice of informative summary statistics. Such a choice is likely to greatly depend on the specificity of the dataset studied, such as the regional species richness, the distribution of individuals in the traits space, and the correlations between traits. Future studies investigating community modeling based on multiple traits will thus need to be tailored to each dataset. This will require many summary statistics to be simulated, to see which ones are the most correlated to the different model parameters, with the dataset studied, and to use these best summary statistics in an ABC framework.

Another possible extension of this approach is to model the intraspecific variability in traits and competitive abilities. Indeed, some recent studies measured traits for each individual rather than for each species (Ackerly and Cornwell 2007, Paine et al. in prep). For instance, one could attribute to each species the variability observed in its traits values across the dataset. Each new recruited individual of a given species would be provided traits values drawn randomly in this species-specific traits distribution. This would enable to study how species traits vary between locations in response to different environmental filters (Ackerly and Cornwell 2007), and how this intraspecific variation is affecting community dynamics (Aarssen 1992, Chave 2004, Vellend and Geber

2005).

A possible obstacle for trait-based studies is the existence of equalizing trade-offs (Chesson 2000) among species traits (Grime et al. 1997, Kneitel and Chase 2004, Reich et al. 2003, Tilman 1990, Westoby et al. 2002, Wright et al. 2004). These trade-offs may imply that different combinations of traits lead to a similar competitive ability in given local conditions. In this case, multiple peaks in the filtering function may occur for each trait, even in a homogeneous local environment. If those trade-offs are low dimensional in the sense that they are based on a limited set of traits, the different ecological strategies with similar competitive ability are likely to be connected by a ridge in trait space (Wainwright 2007). Hence, a way to limit the impact of such effects could be to decompose this set of traits in principal components. The present method would then infer a very large value of  $\sigma$  and/or a small value of  $A$  for the axis describing the trade-offs, but may detect environmental filtering on perpendicular axis, if they capture competitive differences among species once the effect of equalizing trade-offs is sorted out. In this vein, non-linear principal component analysis may be more flexible than the more widely used linear PCA (e.g. Gorban and Zinovyev 2007). The method presented in this contribution may thus be able to deal with low dimensional trade-offs which are abundantly described in the literature (Wright et al. 2004).

Recent modeling works however suggest that real trade-offs may involve much more dimensions (Marks and Lechowicz 2006). Although such highly dimensional trade-offs may explain the commonly observed low dimensional trade-offs (Marks and Lechowicz 2006), they are also likely to produce filtering functions which are highly irregular, as in the metaphor of holey adaptive landscapes in the theory of speciation (Gavrilets 1999, 2004). If the simulation results of Marks and Lechowicz apply to real ecological

systems, the proposed modeling approach does not provide a satisfactory description of the relationship between traits and species competitive abilities, since the “real filtering function” would have to present a great number of peaks, the high trait diversity observed in local communities stemming from the abundance of alternative designs leading to similar competitive abilities (Marks and Lechowicz 2006). However, simulation experiments suggest that traits measuring organism-level performance like shade tolerance or drought resistance may be less influenced by such alternative designs, and more correlated to environmental conditions than organ-level traits such as leaf nitrogen concentration or specific leaf area (Marks 2007). This argues for the use of such more informative yet more difficult to measure traits in trait-based studies of community assembly and dynamics to decrease the noise associated with the presence of alternative designs (Valladares et al. 2002, Engelbrecht et al. 2007).

The problem of alternative designs is not unique to this particular approach for using species traits but is common to all methods aiming at using species traits to detect environmental filtering. It is thus surprising that this limitation has received so little attention in the present literature. A key advantage of the present approach is that it produces clear predictions of trait distributions in communities. Hence, by looking at more subtle patterns in species abundances and co-occurrences, it may be possible to detect whether species with similar traits tend to share similar ecological strategies or whether there is no continuous association between traits and strategies. Although such tests need further research to be developed, the present approach is the first to potentially enable to answer the question of whether local trait diversity is maintained through source-sink dynamics via immigration of locally less competitive species from the regional pool (Pulliam 1988), or to the presence of a great number of alternative designs leading to similar competitive abilities. To answer such difficult questions, it

may be necessary to make use of additional data in this framework, such as growth or mortality data, together with the use of multiple traits possibly linked by trade-offs. The flexibility of the ABC approach should enable the integration of such new data, but may necessitate more efficient ABC implementation (Beaumont et al. 2009, Blum and François *in press*).

### **Acknowledgments**

I thank Jérôme Chave, Nathan Kraft, Christian Marks, Tim Paine and Christopher Baraloto for stimulating discussions and, for their hospitality, the members of David Ackerly's lab where a part of this work was performed. The manuscript has been greatly improved by the comments of the Editor and two anonymous reviewers. This work was funded by the French Ministry of Agriculture, the ATUPS grant of the University of Toulouse, and the ANR Bridge project.

## References

- Aarssen, L.W., 1992. Causes and consequences of variation in competitive ability in plant communities. *J. Veg. Sci.* 3, 165-174.
- Ackerly, D.D., Cornwell, W.K., 2007. A trait-based approach to community assembly : partitioning of species trait values into within- and among-community components. *Ecol. Lett.* 10, 135-145.
- Adler, P.B., HilleRisLambers, J., Levine, J.M., 2007. A niche for neutrality. *Ecol. Lett.* 10, 95-104.
- Alonso, D., Etienne, R.S., McKane, A.J., 2006. The merits of neutral theory. *Trends Ecol. Evol.* 21, 451-457., [doi:10.1016/j.tree.2006.03.019](https://doi.org/10.1016/j.tree.2006.03.019).
- Alonso, D., Ostling, A., Etienne, R.S., 2008. The implicit assumption of symmetry and the species abundance distribution. *Ecol. Lett.* 11, 93-105.
- Baltzer, J.L., Davies, S.J., Bunyavejchewin, S., Noor, N.S.M., 2008. The role of desiccation tolerance in determining tree species distributions along the Malay-Thai Peninsula. *Funct. Ecol.* 22, 221-231.
- Beaumont, M.A., Zhang, W.Y., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. *Genetics* 162, 2025-2035.
- Beaumont, M.A., Cornuet, J.-M., Marin, J.-M., Robert, C.P., 2009. Adaptivity for ABC algorithms : the ABC-PMC scheme. *Biometrika in press*.
- Blum, M.G.B., François, O., 2009. Non-linear regression models for Approximate Bayesian Computation. *Stat. Comput. in press*.
- Cadotte, M.W., 2007. Concurrent niche and neutral processes in the competition-colonization model of species coexistence. *Proc. Roy. Soc. B.* 274, 2739-2744.
- Cavender-Bares, J., Ackerly, D.D., Baum, D.A., Bazzaz, F.A., 2004. Phylogenetic

- overdispersion in Floridian oak communities. *Am. Nat.* 163, 823-843.
- Chave, J., 2004. Neutral theory and community ecology. *Ecol. Lett.* 7, 241-253.
- Chave, J., Muller-Landau, H.C., Levin, S.A., 2002. Comparing classical community models: Theoretical consequences for patterns of diversity. *Am. Nat.* 159, 1-23.
- Chesson, P., 2000. Mechanisms of maintenance of species diversity. *Ann. Rev. Ecol. Sys.* 31, 343-366.
- Connell, J.H., 1971, On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In: Den Boer, P.J., Gradwell, G. (Eds.), *Dynamics of populations*. PUDOC, Wageningen, pp. 298-312.
- Cornwell, W.K., Ackerly D.D., 2009. Community assembly and shifts in the distribution of trait values across an environmental gradient in coastal California. *Ecol. Monogr.* 79, 109-126.
- Cornwell, W.K., Schilck, D.W., Ackerly, D.D., 2006. A trait-based test for habitat filtering: Convex hull volume. *Ecology* 87, 1465-1471.
- Dayan, T., Simberloff D., 2005. Ecological and community-wide character displacement: the next generation. *Ecol. Lett.* 8, 875-894.
- Diaz, S., Cabido, M., Casanoves, F., 1998. Plant functional traits and environmental filters at a regional scale. *J. Veg. Sci.* 9, 113-122.
- Dray, S., Legendre, P., 2008. Testing the species traits-environment relationships : the fourth-corner problem revisited. *Ecology* 89, 3400-3412.
- Dufrene, M., Legendre, P., 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol. Monogr.* 67, 345-366.
- Ellner, S.P., Fussmann, G., 2003. Effects of successional dynamics on metapopulation persistence. *Ecology* 84, 882-889.

- Engelbrecht, B.M.J., Comita, L.S., Condit, R., Kursar, T.A., Tyree, M.T., Turner, B.L., Hubbell, S.P., 2007. Drought sensitivity shapes species distribution patterns in tropical forests. *Nature* 447, 80-82.
- Erjinaes, R., Bruun, H.H., Graae, B.J., 2006. Community assembly in experimental grasslands : suitable environment or timely arrival? *Ecology* 87,1225-1233.
- Etienne, R.S., Olff, H., 2004. A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.* 7, 170-175.
- Etienne, R.S., Olff, H., 2005. Confronting different models of community structure to species-abundance data: a Bayesian model comparison. *Ecol. Lett.* 8, 493-504,
- Fargione, J., Brown, C.S., Tilman, D., 2003. Community assembly and invasion: an experimental test of neutral versus niche processes. *Proc. Natl. Acad. Sci. USA* 100, 8916-8920.
- Fukami, T., Bezemer, T.M., Mortimer, S.R., van der Putten, W.H., 2005. Species divergence and trait convergence in experimental plant community assembly. *Ecol. Lett.* 8, 1283-1290.
- Gavrilets, S., 1999. A dynamical theory of speciation on holey adaptive landscapes. *Am. Nat.* 154, 1-22.
- Gavrilets, S., 2004. *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton, NJ.
- Gorban, A.N., Zinovyev, A.Y., 2007. Elastic maps and nets for approximating principal manifolds and their application to microarray data visualization. In: Gorban, A.N., Kégl, B., Wunsch, D.C., Zinovyev, A.Y. (Eds.), *Principal Manifolds for Data Visualization and Dimension Reduction. Lecture Notes in Computational Science and Engineering*, Springer, pp. 97-131.

- Gravel, D., Canham, C.D., Beaudet, M., Messier, C., 2006. Reconciling niche and neutrality: the continuum hypothesis. *Ecol. Lett.* 9, 399-409.
- Grime, J.P., Thompson, K., Hunt, R., Hodgson, J.G., Cornelissen, J.H.C., Rorison, I.H., Hendry, G.A.F., Ashenden, T.W., Askew, A.P., Band, S.R., Booth, R.E., Bossard, C.C., Campbell, B.D., Cooper, J.E.L., Davison, A.W., Gupta, P.L., Hall, W., Hand, D.W., Hannah, M.A., Hillier, S.H., Hodgkinson, D.J., Jalilio, A., Liu, Z., Mackey, J.M.L., Matthews, N., Mowforth, M.A., Real, A.M., Reader, R.J., Reiling, K., Ross-Fraser, W., Spencer, R.E., Sutton, F., Tasker, D.E., Thorpe, P.C., Whitehouse, J., 1997. Integrated screening validates primary axes of specialisation in plants. *Oikos* 79, 259-281.
- Grime, J.P., 2006. Trait convergence and trait divergence in herbaceous plant communities: mechanisms and consequences. *J. Veg. Sci.* 17, 255-260.
- Hankin, R.K.S., 2007. *untb*: ecological drift under the UNTB. R package version 1.3-2. <http://www.r-project.org>.
- Harms, K.E., Condit, R., Hubbell, S.P., Foster, R.B., 2001. Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *J. Ecol.* 89, 947-959.
- Holyoak, M., Loreau, M., 2006. Reconciling empirical ecology with neutral community models. *Ecology* 87, 1370-1377.
- Hubbell, S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.
- Jabot, F., Chave, J., 2009. Inferring the parameters of the neutral theory of biodiversity using phylogenetic information, and implications for tropical forests. *Ecol. Lett.* 12, 239-248.
- Jabot, F., Etienne, R.S., Chave, J., 2008. Reconciling neutral community models and environmental filtering : theory and an empirical test. *Oikos* 117, 1308-1320.

- Janzen, D.H., 1970. Herbivores and the number of tree species in tropical forests. *Am. Nat.* 104, 501-529.
- Kleyer, M., Bekker, R.M., Knevel, I.C., Bakker, J.P., Thompson, K., Sonnenschein, M., Poschod, P., van Groenendael, J.M., Klimes, L., Klimesova, J., Klotz, S., Rusch, G.M., Hermy, M., Adriaens, D., Boedeltje, G., Bossuyt, B., Dannemann, A., Endels, P., Gotzenberger, L., Hodgson, J.G., Jackel, A.K., Kuhn, I., Kunzmann, D., Ozinga, W.A., Romermann, C., Stadler, M., Schlegelmilch, J., Steendam, H.J., Tackenberg, O., Wilmann, B., Cornelissen, J.H.C., Eriksson, O., Garnier, E., Peco, B., 2008. The LEDA Traitbase: a database of life-history traits of the Northwest European flora. *J. Ecol.* 96, 1266-1274.
- Kneitel, J.M., Chase, J.M., 2004. Trade-offs in community ecology : linking spatial scales and species coexistence. *Ecol. Lett.* 7, 69-80.
- Kraft, N.J.B., Valencia, R., Ackerly, D.D., 2008. Functional traits and niche-based tree community assembly in an amazonian forest. *Science* 322, 580-582.
- Lavorel, S., Garnier, E., 2002. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Funct. Ecol.* 16, 545-556.
- Legendre, P., Galzin, R., Harmelin-Vivien, M.L., 1997. Relating behavior to habitat: Solutions to the fourth-corner problem. *Ecology* 78, 547-562.
- MacArthur, R., Levins, R., 1967. The limiting similarity, convergence, and divergence of coexisting species. *Am. Nat.* 101, 377-385.
- Marks, C.O., 2007. The causes of variation in tree seedling traits: the roles of environmental selection versus chance. *Evolution* 61, 455-469.
- Marks, C.O., Lechowicz, M.J., 2006. Alternative designs and the evolution of functional diversity. *Am. Nat.* 167, 55-66.

- McGill, B.J., Enquist, B.J., Weiher, E., Westoby, M., 2006. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* 21, 178-185., [doi:10.1016/j.tree.2006.02.002](https://doi.org/10.1016/j.tree.2006.02.002).
- McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F.L., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I., White, E.P., 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* 10, 995-1015.
- Mouillot, D., Mason, N.W.H., Wilson, J.B., 2007. Is the abundance of species determined by their functional traits? A new method with a test using plant communities. *Oecologia* 152, 729-737.
- Mouquet, N., Loreau, M., 2003. Community patterns in source-sink metacommunities. *Am. Nat.* 162, 544-557.
- Paine, C.E.T., Baraloto, C., Chave, J., Hérault, B., Jabot, F., in prep. The neutrality of functional trait distributions in French Guiana rain forests.
- Palmer, M.W., Dixon, P.M., 1990. Small-scale environmental heterogeneity and the analysis of species distributions along gradients. *J. Veg. Sci.* 1, 57-65.
- Pausas, J.G., Verdu, M., 2008. Fire reduces morphospace occupation in plant communities. *Ecology* 89, 2181-2186.
- Pitman, N.C.A., Terborgh, J.W., Silman, M.R., Nunez, P., Neill, D.A., Ceron, C.E., Palacios, W.A., Aulestia, M., 2001. Dominance and distribution of tree species in upper Amazonian terra firme forests. *Ecology* 82, 2101-2117.
- Plotkin, J.B., Potts, M.D., Leslie, N., Manokaran, N., LaFrankie, J., Ashton, P.S., 2000. Species-area curves, spatial aggregation, and habitat specialization in tropical forests. *J. Theor. Biol.* 207, 81-99., [doi:10.1006/jtbi.2000.2158](https://doi.org/10.1006/jtbi.2000.2158).

- Poorter, L., Wright, S.J., Paz, H., Ackerly, D.D., Condit, R., Ibarra-Manriques, G., Harms, K.E., Licona, J.C., Martinez-Ramos, M., Mazer, S.J., Muller-Landau, H.C., Pena-Claros, M., Webb, C.O., Wright, I.J., 2008. Are functional traits good predictors of demographic rates? Evidence from five Neotropical forests. *Ecology* 89, 1908-1920.
- Prinzing, A., Reiffers, R., Braakhekke, W.G., Hennekens, S.M., Tackenberg, O., Ozinga, W.A., Schaminee, J.H.J., van Groenendael, J.M., 2008. Less lineages - more trait variation: phylogenetically clustered plant communities are functionally more diverse. *Ecol. Lett.* 11, 809-819.
- Pulliam, H.R., 1988. Sources, sinks, and population regulation. *Am. Nat.* 132, 652-661.
- Purves, D.W., Pacala, S.W., 2005. Ecological drift in niche-structured communities: neutral pattern does not imply neutral process. In: Burslem, D., Pinard, M., Hartley, S. (Eds.), *Biotic interactions in the tropics*. Cambridge University Press, Cambridge, pp. 107-138.
- R Development Core Team, 2008. *R: A language and environment for statistical computing*. R foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.
- Reich, P.B., Wright, I.J., Cavender-Bares, J., Craine, J.M., Oleksyn, J., Westoby, M., Walters, M.B., 2003. The evolution of plant functional variation : traits, spectra, and strategies. *Int. J. Plant. Sci.* 164, S143-S164.
- Ricklefs, R.E., Travis, J., 1980. A morphological approach to the study of avian community organization. *Auk* 97, 321-338.
- Schwilk, D.W., Ackerly, D.D., 2005. Limiting similarity and functional diversity along environmental gradients. *Ecol. Lett.* 8, 272-281.
- Stubbs, W.J., Wilson, J.B., 2004. Evidence for limiting similarity in a sand dune

- community. *J. Ecol.* 92, 557-567.
- Tilman, D., 1990. Constraints and tradeoffs : toward a predictive theory of competition and succession. *Oikos* 58, 3-15.
- Valladares, F., Skillman, J.B., Pearcy, R.W., 2002. Convergence in light capture efficiencies among tropical forest understory plants with contrasting crown architectures: a case of morphological compensation. *Am. J. Bot.* 89, 1275-1284.
- Vellend, M., Geber, M.A., 2005. Connections between species diversity and genetic diversity. *Ecol. Lett.* 8, 767-781.
- Villéger, S., Mason, N.W.H., Mouillot, D., 2008. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* 89, 2290-2301.
- Volkov, I., Banavar, J.R., He, F., Hubbell, S.P., Maritan, A., 2005. Density and frequency dependence explains tree species abundance and diversity in tropical forests. *Nature* 438, 658-661.
- Walker, S.C., 2007. When and why do non-neutral metacommunities appear neutral? *Theor. Popul. Biol.* 71, 318-331., [doi:10.1016/j.tpb.2006.12.008](https://doi.org/10.1016/j.tpb.2006.12.008).
- Wainwright, P.C., 2007. Functional versus morphological diversity in macroevolution. *Ann. Rev. Ecol. Evol. Sys.* 38, 381-401.
- Webb, C.O., Ackerly, D.D., McPeck, M.A., Donoghue, M.J., 2002. Phylogenies and community ecology. *Ann. Rev. Ecol. Evol. Sys.* 33, 475-505.
- Webb, C.O., Cannon, C.H., Davies, S.J., 2008. Ecological organization, biogeography, and the phylogenetic structure of tropical forest tree communities. In: Carson, W.P., Schnitzer, S.A. (Eds.), *Tropical Forest Community Ecology*. Blackwell, pp. 79-97.
- Weiher, E., Keddy, P.A., 1995. Assembly rules, null models, and trait dispersion: new

- questions from old patterns. *Oikos* 74, 159-164.
- Westoby, M., Wright, I.J., 2006. Land-plant ecology on the basis of functional traits. *Trends Ecol. Evol.* 21, 261-268., [doi:10.1016/j.tree.2006.02.004](https://doi.org/10.1016/j.tree.2006.02.004).
- Westoby, M., Falster, D.S., Moles, A.T., Vesk, P.A., Wright, I.J., 2002. Plant ecological strategies: Some leading dimensions of variation between species. *Ann. Rev. Ecol. Evol. Sys.* 33, 125-159.
- Whittaker, R.H., 1965. Dominance and diversity in land plant communities. *Science*, 147, 250-260.
- Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J.H.C., Diemer, M., Flexas, J., Garnier, E., Groom, P.K., Gulias, J., Hikosaka, K., Lamont, B.B., Lee, T., Lee, W., Lusk, C., Midgley, J.J., Navas, M.L., Niinemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V.I., Roumet, C., Thomas, S.C., Tjoelker, M.G., Veneklaas, E.J., Villar, R., 2004. The worldwide leaf economics spectrum. *Nature* 428, 821-827.
- Wright S., 1931. Evolution in mendelian populations. *Genetics* 16, 97-159.
- Wright S., 1937. The distribution of gene frequencies in populations. *Proc. Natl. Acad. Sci. USA* 23, 307-320.

**Table 1**

Parameter definitions

Symbol	Definition
$I$	Immigration rate
$h$	Optimal trait value = Mode of the environmental filtering function
$A$	Height of the environmental filtering function
$\sigma$	Width of the environmental filtering function

Accepted manuscript

**Fig. 1.** Illustration of the filtering function.

Caption: Panel a: Filtering function. Each species has a trait value which determines its local competitive ability. This competitive ability is given by the filtering function, which has three parameters:  $h$  the position of the optimal trait value,  $A$  the maximal competitive advantage associated with this optimal trait value compared to the least competitive species, and  $\sigma$  the standard deviation of the filtering function which encapsulates the width of the filtering function. Panels b, c, d: Examples of trait distributions in simulated communities with various strength of filtering. Each '+' represents the abundance of one species in the local community. Panel b: Strong filtering. The distribution of trait values is highly peaked; species with a trait value different from the optimal trait value are at very low abundances. Panel c: Intermediate filtering. The distribution of trait values is still peaked, but more species have large abundances, and species with a trait value different from the optimal trait value can have medium abundances. Panel d: Low filtering. The distribution of trait values is only slightly peaked and species with very different trait values coexist at medium abundances.

**Box 1.** Identifiability of the model parameters with four summary statistics.

The immigration rate  $I$  greatly influences the species richness  $S$  of the local community (Fig. 2a), but it is not the sole factor in determining the local richness. A given local species richness ( $S=100$  in Fig. 2b) can be produced by different combinations of parameter values (Fig. 2b). What matters is the total amount of filters from the regional pool into the local community. In the present model, both dispersal limitation (the lower  $I$ , the greater the dispersal limitation) and environmental filtering act as filters. The intensity of environmental filtering is positively correlated with the height  $A$  of the environmental filtering function, since a greater  $A$  implies a greater asymmetry in

competitive ability among species, hence a greater chance of competitive exclusion of poorly competitive species from the local community. The intensity of environmental filtering is also negatively correlated with the width  $\sigma$  of the environmental filtering function, since a larger  $\sigma$  implies a greater number of species with relatively good competitive ability in the local community. In sum, environmental filtering is positively correlated with the ratio  $A/\sigma$ . The knowledge of  $S$  enables to restrict the most likely parameters values through the information of total filtering intensity (Fig. 2b), but not to know the relative contribution of environmental filtering and dispersal limitation (Jabot et al. 2008). Among those simulated communities of equal richness ( $S=100$ ), Shannon's index  $H$  is negatively correlated with the immigration rate  $I$  (Fig. 2c). Hence, the knowledge of  $S$  and  $H$  enables to know the relative contribution of dispersal and environmental filtering in this model, i.e. to know the value of  $I$  and  $A/\sigma$ . At fixed  $S$  and  $H$  (here I consider a narrow stripe of  $H$  values to have still a large number of simulated communities), the mean trait value in the community ( $MTV$ ) is strongly correlated with the position of the environmental filtering function on the trait axis  $h$  (Fig. 2d). Hence, this additional summary statistic enables to estimate parameter  $h$ . Finally, for fixed  $S$ ,  $H$ , and  $MTV$  values (here again, I take a narrow stripe of  $MTV$  values), the skewness of the trait value distribution in the community ( $STV$ ) is correlated with the height of the environmental filtering function  $A$  (Fig. 2e). Hence  $STV$  enables to determine  $A$  and consequently  $\sigma$ , since the ratio  $A/\sigma$  was already known from the value of  $H$ . This final correlation between  $STV$  and  $A$  can be understood as follows. For a given level of environmental filtering intensity ( $A/\sigma$ ), this filtering can be either due to the strong competitive advantage of a large number of species, or by a lower competitive advantage of a smaller number of species. In the second case,  $A$  and  $\sigma$  will be smaller, and the community will be composed of a few species with high local competitive ability, and of a lot of species nearly randomly drawn from the regional

pool. Here, communities filtered around a value of trait  $h=30$  were selected, hence a large number of species drawn between 0 and 100 in the regional pool will have trait values above 30, and the distribution of trait values among individuals will tend to have a high skewness. On the contrary, communities simulated with larger  $A$  and  $\sigma$  value will be composed of a greater number of species symmetrically drawn around 30 according to the environmental filtering function (which is symmetric), and of a smaller number of species randomly drawn from the regional pool, hence the skewness of the trait distribution will be lower. If communities had been selected around a value of  $h$  above 50, an opposite correlation between  $STV$  and  $A$  would have been observed, but it would have been possible to infer the parameters as well, since what matters is the existence of this correlation which ensures that the summary statistics give information on the parameter values. A potential problem can be envisaged: for communities filtered around the center of the trait distribution ( $h=50$ ), the value of  $STV$  might not be anymore correlated with  $A$ , hence the estimation of the parameters may not be possible anymore, at least with this approach. This is what is observed (see Results).

**Fig. 2. TO BE PUT IN BOX 1** Summary statistics and estimation of the model parameters

Caption: In all panels, crosses represent communities of size  $J=500$  simulated with the present model of community dynamics with a regional pool composed of 1000 species of equal abundances and evenly spaced on the trait axis (values between 0 and 100).

Panel a: local species richness ( $S$ ) is positively correlated with the immigration parameter  $\ln(I)$ . Panel b: only communities with  $S$  equal to 100 are retained. A positive correlation between the parameters  $\ln(I)$  and  $\ln(A/\sigma)$  can be observed. In other words, to keep local richness constant, communities simulated with larger  $\ln(I)$  need to be also simulated with larger  $\ln(A/\sigma)$ . Panel c: at constant local richness value  $S$ , communities

simulated with larger  $\ln(I)$  have lower Shannon's index  $H$ . The vertical bars delimit the range of  $H$  values retained for panels d and e, so that communities next examined have similar  $H$  values. Panel d: mean trait values in local communities  $MTV$  are positively correlated with parameter  $h$ . Vertical bars delimit the range of  $MTV$  retained for panel e. Panel e: skewness of trait values in local communities  $STV$  are negatively correlated with parameter  $\ln(A)$ . Altogether, variations in  $S$ ,  $H$ ,  $MTV$  and  $STV$  enable to capture variations in the four parameters of the model.

**Fig. 3.** Identifiability of the four model parameters

Caption: Each panel represents the correlation between one parameter used in 100 simulated communities and the corresponding parameter value inferred by ABC. The black line represents the line of equation  $y=x$ . Each simulated community is of size  $J=500$ , and species richness  $S=100$ . The regional pool used in the simulation contains 1,000 species of equal abundances and evenly spaced on the trait axis between 0 and 100. Panel a: immigration parameter  $I$ , panel b: parameter  $h$ , panel c: parameter  $A$ , panel d: parameter  $\sigma$ . In each panel, red circles highlight communities for which the parameter inference is not efficient for the parameter represented in the panel. The properties of these badly-inferred communities are described in Fig. 4.

**Fig. 4.** Area of the parameter space with poor parameter identifiability

Caption: Panel a: parameters  $A$  and  $\sigma$  used for the 25 communities that have at least one parameter poorly inferred. There are two distinct groups of communities. The first group (depicted by crosses) correspond to communities which are not strongly filtered, either because  $A$  is low, or because  $\sigma$  is large. The second group (depicted by stars) corresponds to communities which are more strongly filtered, but for trait values around the mean trait value in the regional pool (see Panel c). Panel b: Volume of the

approximate posterior distribution of the three parameters  $A$ ,  $\sigma$  and  $I$  in the 100 communities, measured by the convex hull volume of the 200 first points of the approximate posterior distribution. The values of the 25 communities which are not well inferred are in red. They are larger than for well-inferred communities, i.e. communities which are not well inferred have a posterior distribution with a larger volume. Panel c: Values of the parameter  $h$  used in the 100 simulated communities. The values for the 4 communities which are not well inferred and which are strongly filtered are in red. They are close to the mean value of the regional pool ( $h=50$ ). Panel d (resp. e): Example of the approximate posterior distribution (without the  $h$  axis) of a community which is not well inferred (resp. well inferred) by our method.

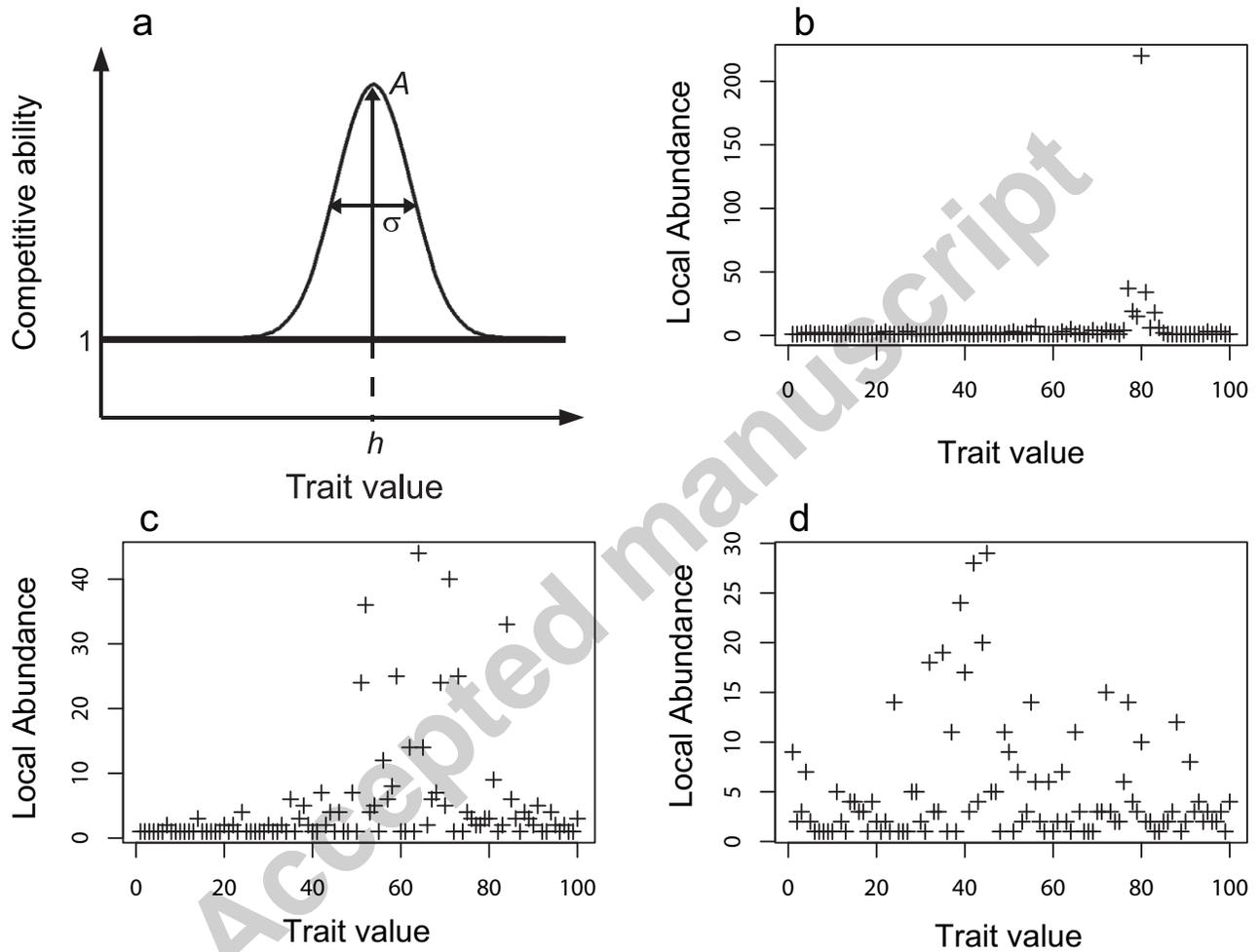
**Fig. 5.** Filtering shape and detectability of non-neutral patterns.

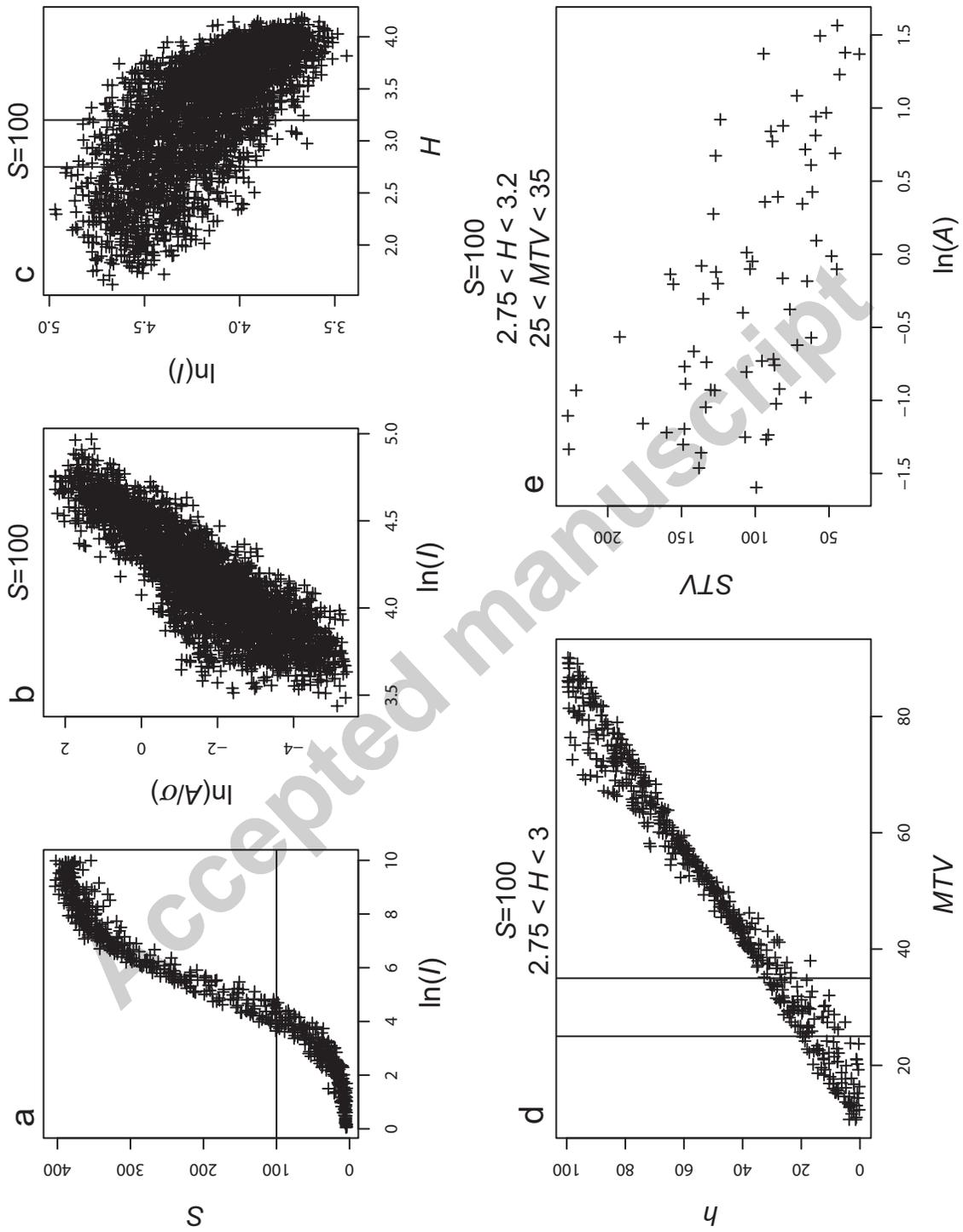
Caption: Panel a: Color-coded values of Shannon's index  $H$  as a function of the parameters  $A$  and  $\sigma$  in 1000 simulated communities. To obtain a continuous representation instead of 1000 colored points, values of  $H$  at each point of the  $(A, \sigma)$  surface were interpolated with a kriging technique from the 1000 known values scattered in the  $(A, \sigma)$  surface, and then plotted. The black line corresponds to the minimal value of  $H$  obtained in 1000 simulated neutral communities. Hence, all combination of parameters  $A$  and  $\sigma$  at the right of this line lead to species abundances that are not compatible with neutrality. The maximal value of  $H$  obtained in neutral simulations is not depicted, since  $H$  values in filtered communities never exceed neutral values. Indeed, environmental filtering as modeled here can only decrease  $H$  values compared to those found in neutral communities. Panels b, c, d: same as Panel a, for the three summary statistics  $Range$ ,  $Range5$ , and  $VTV$ . All filtered communities had values of  $Range$  compatible with neutrality.

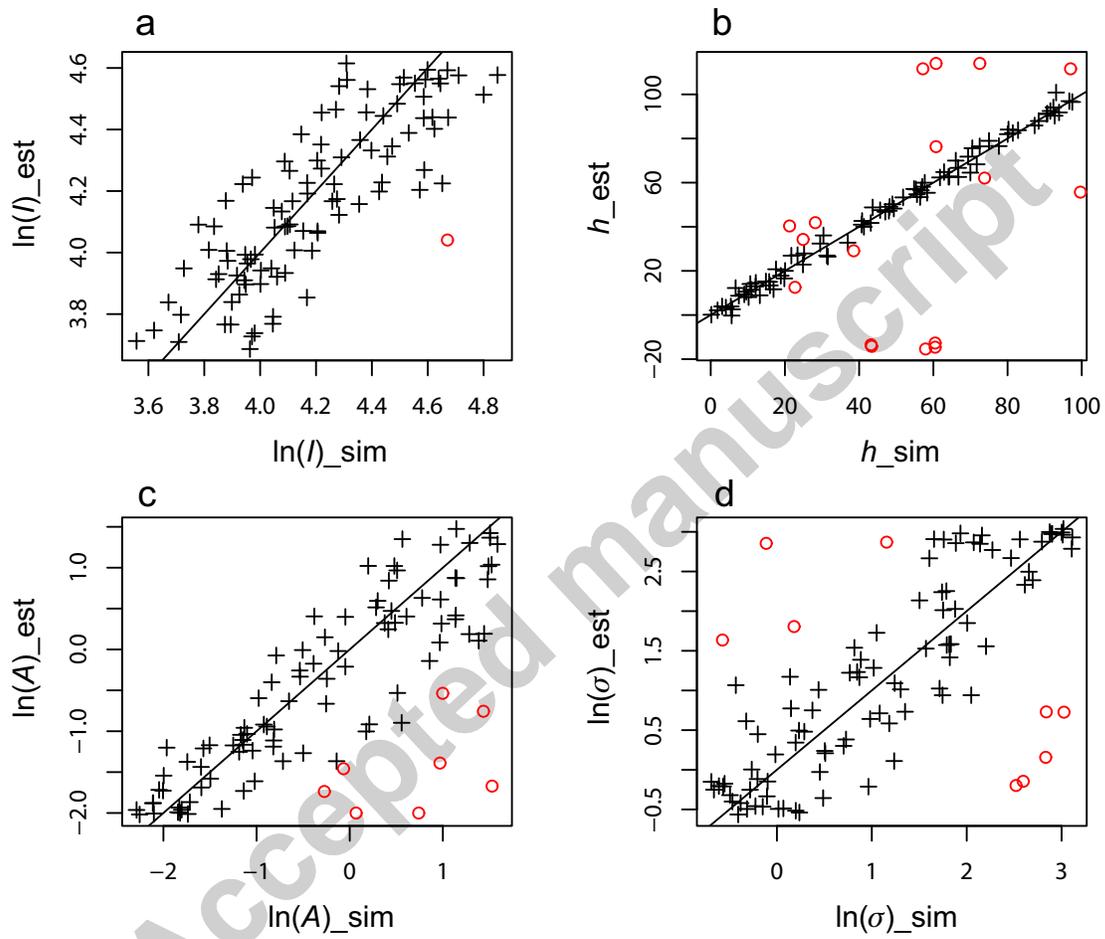
**Fig. 6.** Detection of multiple peaks in the filtering function.

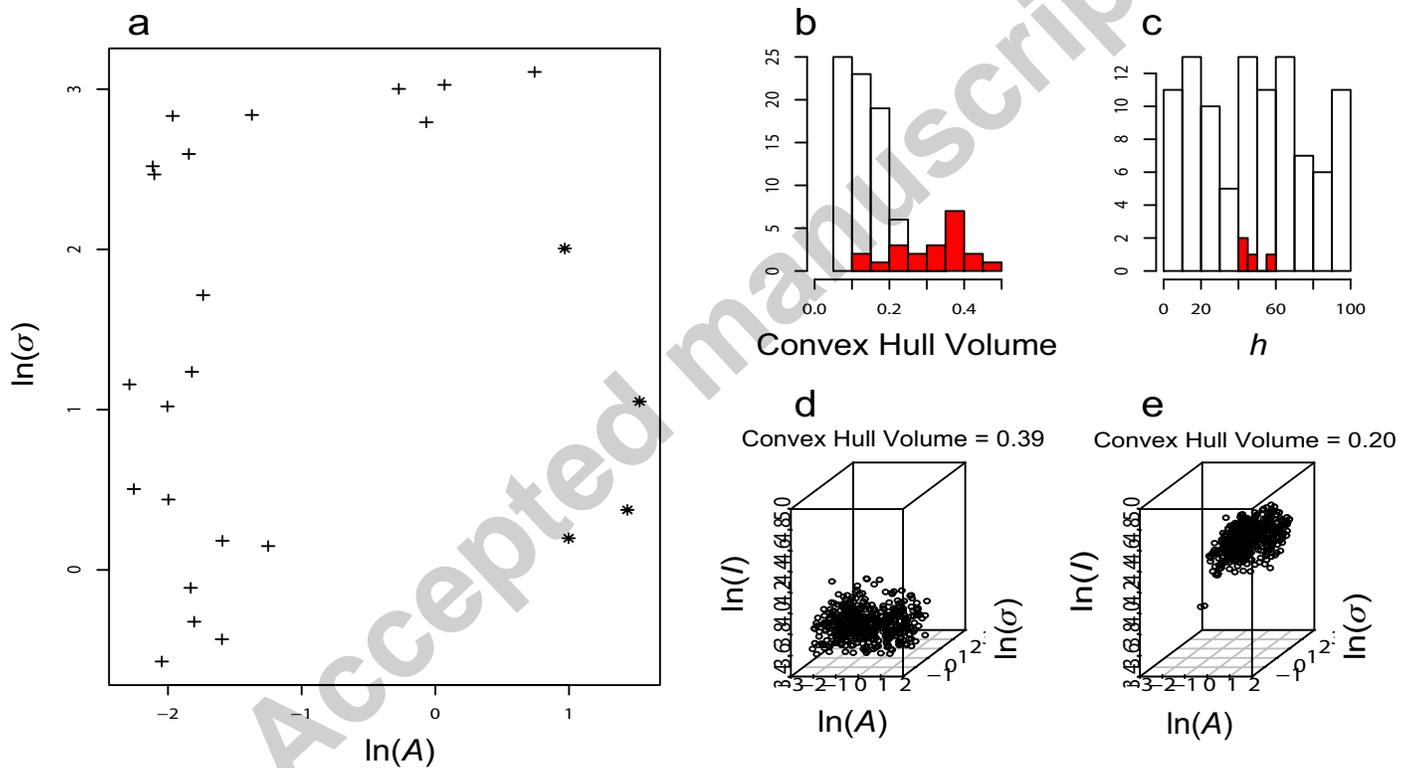
Caption: Panel a: Example of the test. The vertical bar stands for the  $VTV$  value in the simulated 2-peaks community. The histogram represents the distribution of  $VTV$  values in the 200 best-fit 1-peak communities. Panel b:  $p$ -values obtained in 100 simulated 2-peaks communities.

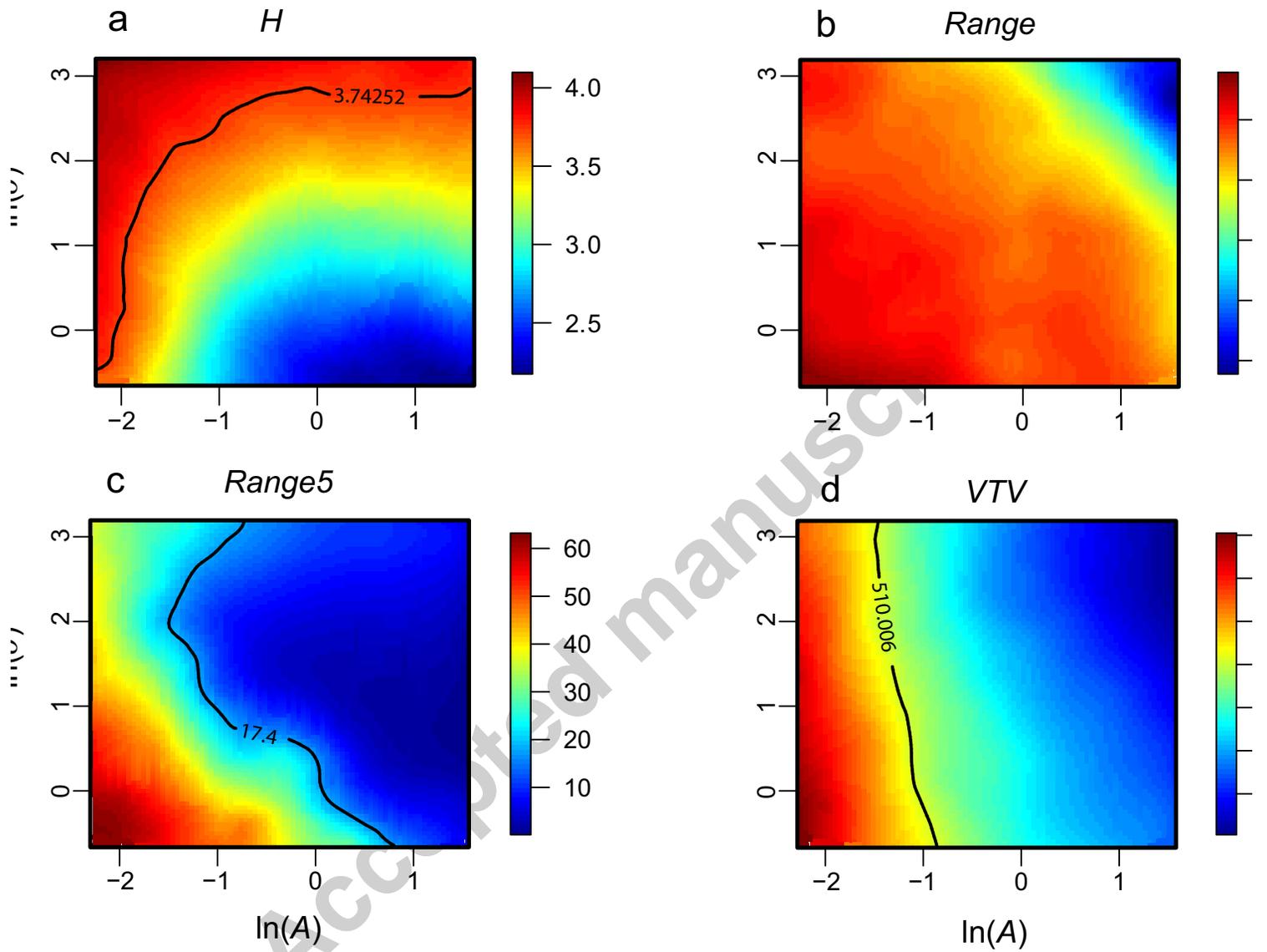
Accepted manuscript

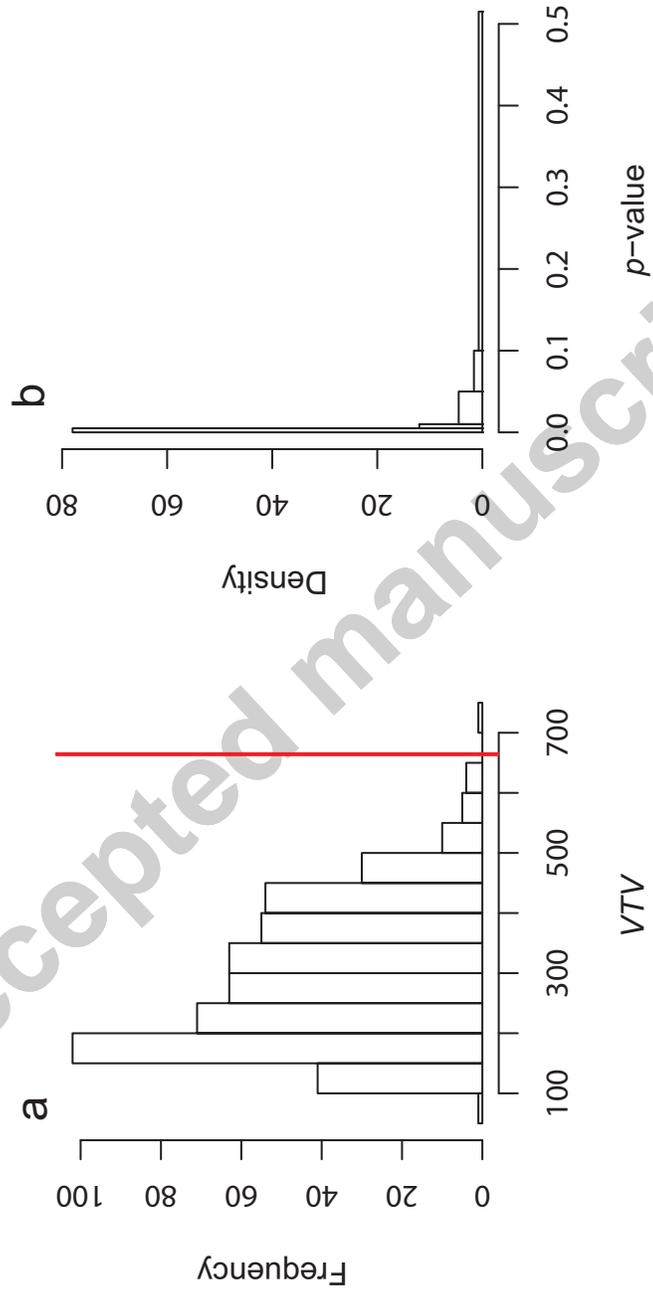








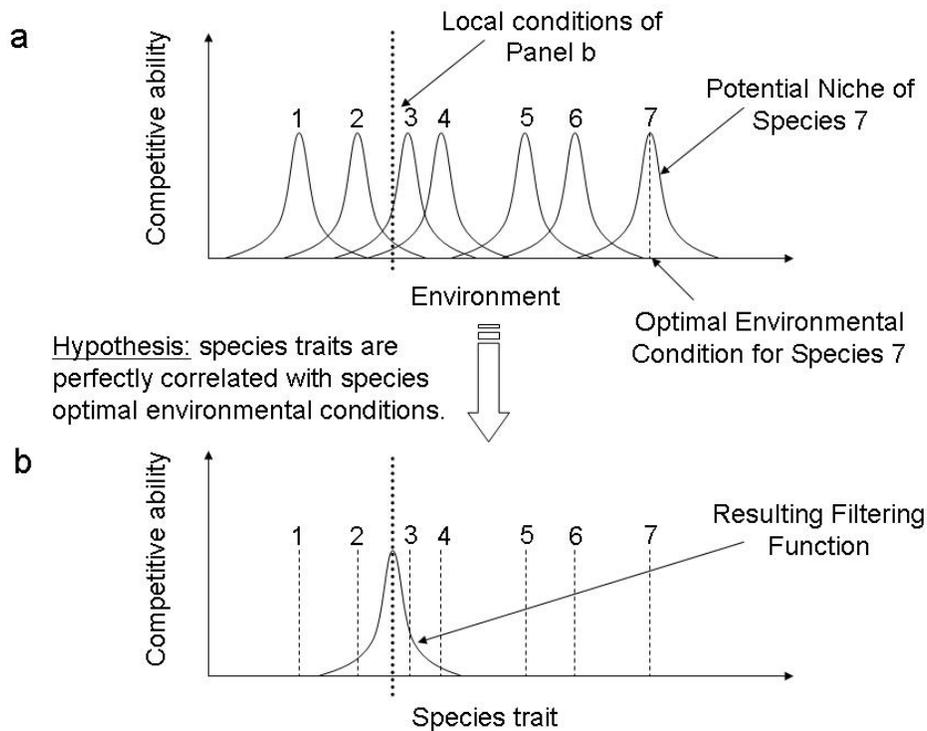




## Supplementary Material

Appendix A: Connection between the environmental filtering function modeling and classical approaches in species niche modeling.

Theoretical modeling of species niches often makes use of Gaussian functions around an optimal environmental position for a species (e.g. Schwilk and Ackerly 2005). This Gaussian function describes the competitive ability of the focal species as a function of the environment (Fig. A1a). If the variance and height of these niche functions are constant among species, and if the species optimal environmental conditions are perfectly correlated to species traits, the filtering function at a given environmental condition is a Gaussian function of same variance and height centered on the trait value corresponding to the local environmental condition (Fig. A1b). This result keeps true for all types of symmetric niche functions (ie such that  $f(h-x)=f(h+x)$  ), not just the Gaussian one.



**Fig. A1.** Panel a: Classical species niche modeling on a one-dimensional environmental axis. Panel b: Resulting filtering function as a function of species traits, under the assumptions that all species niche functions are symmetric and identical in shape among species, and that species traits are perfectly correlated to their optimal environmental conditions.

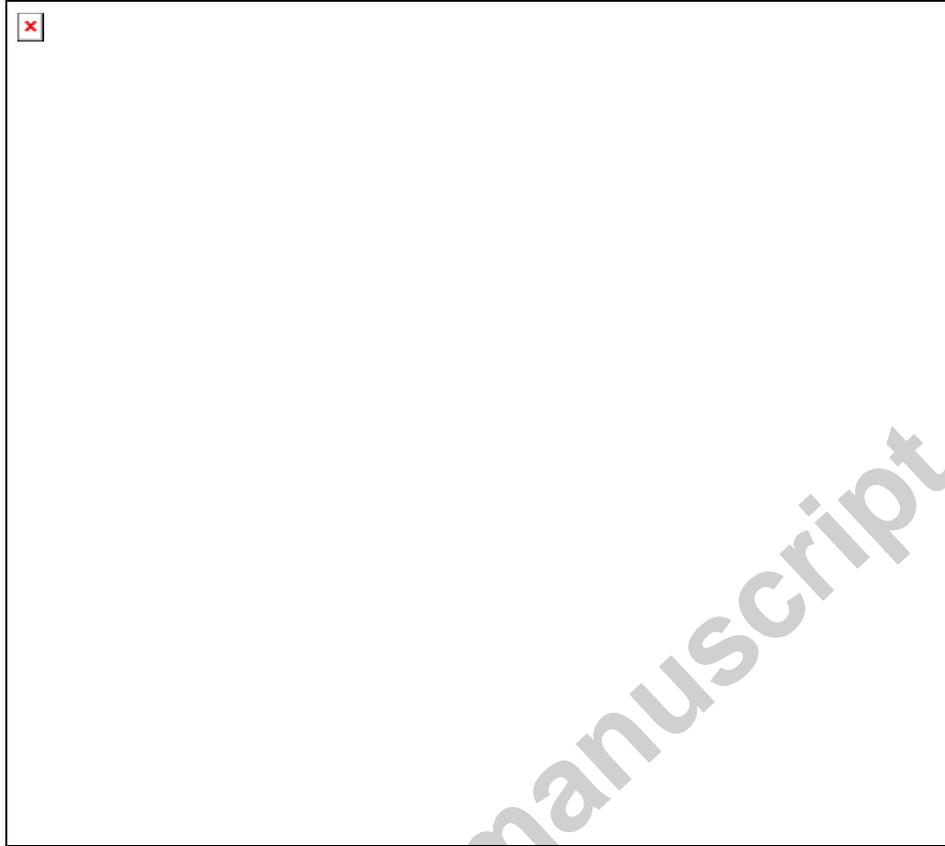
Reference:

Schwilk, D.W., Ackerly, D.D., 2005. Limiting similarity and functional diversity along environmental gradients. *Ecol. Lett.* 8, 272-281.

## Appendix B: Details on the ABC implementation

### 1-Length of the forward-in-time simulation:

In the model considered in this article, the community dynamics necessitates to be simulated forward-in-time. Starting from a random draw from the regional pool, each species being pondered by its local competitive ability given the parameters values, the dynamics consist in a series of death and replacement of individuals. After some time, a dynamic equilibrium is reached. In the present work, the forward-in-time dynamics was run for  $\mathcal{J}$  time steps, each time step consisting of a death event followed by a recruitment event ( $J$  stands for the local sample size and is equal to 500 in the present paper). To check whether this simulation length was sufficient so that a dynamic equilibrium was reached, I simulated communities with forward-in-time dynamics of length  $20 * J$ ,  $100 * J$ ,  $500 * J$  and  $2500 * J$ . The same regional pool as in the main text was used. 24 combinations of parameters values were explored:  $\ln(J)$  in  $\{2,6\}$ ,  $A$  in  $\{0.5,2\}$ ,  $h$  in  $\{5,20,50\}$ ,  $\sigma$  in  $\{5,15\}$ . For each combination of parameters and each forward-in-time dynamics length, 100 communities were simulated (hence 9600 communities were simulated in total). For each of these communities, the four summary statistics used in the ABC implementation were recorded:  $S$ ,  $H$ ,  $MTV$  and  $STV$ . Although there was an effect of the simulation length for small length values (data not shown), no effect was detectable between the lengths  $500 * J$  and  $2500 * J$ . This was assessed by 96 t-tests (one for each parameters combination and summary statistics). The distribution of  $p$ -values among these 96 t-tests (Fig. B1) was not different from a uniform distribution between 0 and 1 ( $p$ -value=0.74). Furthermore, the few parameters combinations which lead to  $p$ -values below 0.05 were not aggregated in the parameter space. Hence they do not suggest that in some parts of the parameter space, a greater simulation length needs to be used.



**Fig. B1.**  $p$ -values in 96 t-tests for differences between communities simulated with  $500 \cdot J$  time steps and communities simulated with  $2500 \cdot J$  time steps in the forward-in-time algorithm.

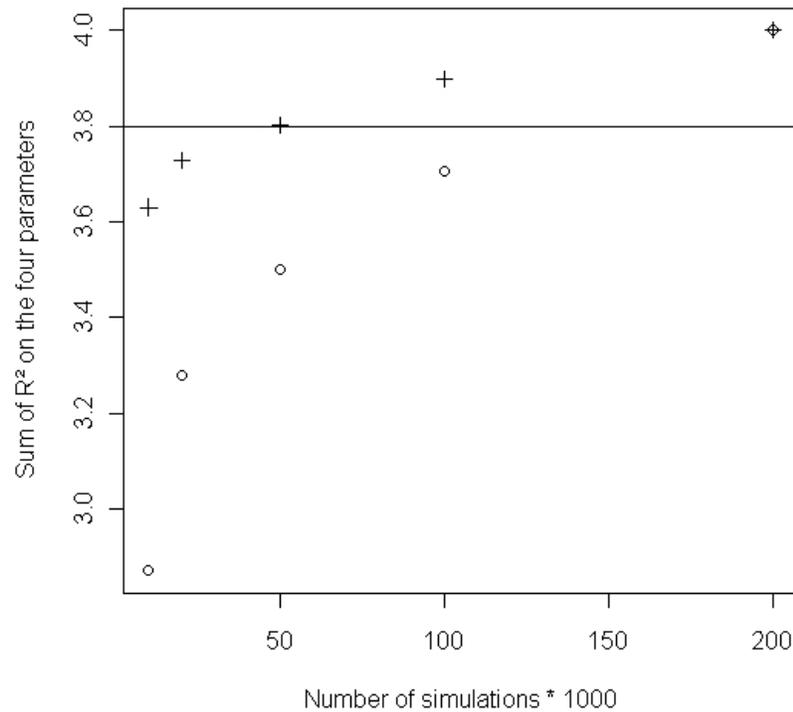
### 2-Posterior mode estimation:

We used a simple R function (“density”) to estimate the posterior marginal mode for each parameter of the model. This estimation was performed on the set of the 500 closest ABC simulations, each one being weighted with an Epanechnikov function of its distance to the data (Beaumont et al. 2002).

### 3-Number of ABC simulations for parameter inference:

In any ABC inference, the greater the number of ABC simulations, the better is the inference. However, for computing limitations, there is a need to restrain the number of

simulations. To check whether the number of ABC simulations retained in this study of 100,000 communities with species richness comprised between 90 and 110 (for estimating parameters of communities with species richness equal to 100) was sufficiently large, I performed the same estimation with a number of ABC simulations varying from 20,000 to 200,000. I then computed the correlation coefficient  $R^2$  between parameters estimated with 200,000 ABC simulations, and parameters estimated with fewer ABC simulations, and summed them over the four parameters to get a global picture of the similarity among estimation procedures. For numbers of ABC simulations above 50,000, the  $R^2$  is above 0.95 on average (Fig. B2). The number used of 100,000 ABC simulations thus seems to be a good compromise between inference quality and inference speed.



**Fig. B2.** Circles stand for correlation coefficients between the parameters inferred in 100 simulated communities with a number of ABC simulations equal to 200,000 and other smaller numbers of ABC simulations. Crosses stand for the same correlation coefficients on the 75 simulated communities where the ABC inference works well. The horizontal line stands for an average  $R^2$  of 0.95 over the four model parameters.

Reference:

Beaumont, M.A., Zhang, W.Y., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. *Genetics* 162, 2025-2035.

## Appendix C: Robustness of the results to various simulation choices.

### A/Inference method

To assess whether the proposed inference method was still working in other conditions than the one intensively explored in the main text, I reproduced similar inferences with different species richness in the local community, species pool sizes, regional abundances structure and filtering function shapes. In each case, I simulated 100,000 ABC communities and inferred the parameters of 100 out of these 100,000 communities. I computed the correlation coefficient  $R^2$  between simulated and inferred parameters values with the function “rlm” in the package “MASS” of the R software, excluding the outliers with the arbitrary criterion “outliers = points with a weight in the robust regression below 0.5”. This arbitrary criterion was found visually to select accurately the outliers in the present case. An additional analysis was performed without constraint on local species richness: 1,000,000 ABC communities were simulated without constraint on local species richness, and the parameters of 100 out of these 1,000,000 communities were inferred. These 100 reference communities were randomly picked in the 1,000,000 communities. They had local richness ranging from  $S=11$  to  $S=243$ .

### 1-Effect of the number of species in the community

Local species richness equal to 20, 50, and 200 were considered, in addition to the case  $S=100$  described in the main text. Uniform prior distributions for parameters  $h$ ,  $\ln(A)$ , and  $\ln(\sigma)$  were the same as in the main text ( $h$  in  $[-25;125]$ ,  $\ln(A)$  in  $[\ln(0.1) ; \ln(5)]$ ,  $\ln(\sigma)$  in  $[\ln(0.5) ; \ln(25)]$ ). Parameter  $\ln(J)$  had different uniform prior distributions in each case, so as to obtain communities with targeted local species richness: for  $S=20$ ,  $\ln(J)$  was drawn in  $[0; 4]$ , for  $S=50$ ,  $\ln(J)$  was drawn in  $[1.5; 4.5]$ , for  $S=200$ ,  $\ln(J)$  was

drawn in [4.5; 6], and when  $S$  was not constrained,  $\ln(I)$  was drawn in [0;7].  $R^2$  and percentages of outliers were similar in each case, except in the case with lowest  $S=20$ , where parameter  $\ln(\sigma)$  was poorly inferred (Table C1). Hence, the present approach is better suited for the analysis of species-rich communities. When  $S$  was not constrained,  $R^2$  and percentages of outliers were similar to the results presented in the main text for  $S=100$ . One can note that  $R^2$  for parameter  $\ln(I)$  is higher in this last case. This is due to the fact that a wider range of values for  $\ln(I)$  are examined in this case compared to cases with constrained  $S$ .

Model parameters	$R^2$ (without outliers) $S=20$	% of outliers $S=20$	$R^2$ (without outliers) $S=50$	% of outliers $S=50$	$R^2$ (without outliers) $S=100$	% of outliers $S=100$	$R^2$ (without outliers) $S=200$	% of outliers $S=200$	$R^2$ (without outliers) $S$ not constrained	% of outliers $S$ not constrained
$\ln(I)$	0.28	3	0.65	0	0.65	1	0.65	1	0.97	2
$h$	0.99	7	0.99	9	0.99	17	0.96	26	0.99	17
$\ln(A)$	0.76	10	0.84	10	0.73	8	0.87	6	0.81	10
$\ln(\sigma)$	0.33	4	0.61	3	0.82	9	0.54	4	0.68	14

**Table C1.**  $R^2$  and percentages of outliers for different local species richness ( $S$ ).

## 2-Effect of the number of species in the regional pool

Species pool sizes equal to 300 and 3,000 were considered. The species were still regularly distributed on the trait axis. In the robust regressions between simulated and inferred parameters values,  $R^2$  and percentages of outliers were similar in each case (Table C2). Thus species pool size does not seem to influence much the results.

Model parameters	$R^2$ (without outliers) $S_{tot}=1000$	% of outliers $S_{tot}=1000$	$R^2$ (without outliers) $S_{tot}=300$	% of outliers $S_{tot}=300$	$R^2$ (without outliers) $S_{tot}=3000$	% of outliers $S_{tot}=3000$
$\ln(I)$	0.65	1	0.67	3	0.65	1
$h$	0.99	17	0.99	11	0.98	4
$\ln(A)$	0.73	8	0.61	9	0.76	7
$\ln(\sigma)$	0.82	9	0.80	11	0.76	8

**Table C2.**  $R^2$  and percentages of outliers for different species pool sizes ( $S_{tot}$ ).

### 3-Effect of the structure in regional abundances

In the main text, the species pool has a size of 1000 species with equal regional abundances. Here, for the same species pool size, I considered two additional cases. In both cases, the regional abundances were generated with the use of a Fisher logseries with parameter  $\alpha$  equal to 200 using the function “fisher.ecosystem” of the R package “untb”. In the first case, regional abundances were randomly attributed to each species, independently of their traits, while in the second case, species regional abundances were not independent of species traits: I picked at random a position in the trait space, and then placed the species around this trait value in decreasing order of abundances so as to form a peak in the trait-regional abundance plot. In the robust regressions between simulated and inferred parameters values,  $R^2$  and percentages of outliers were similar in each case, except for the  $R^2$  of parameter  $\sigma$  in the case where abundances and traits were correlated (Table C3). Thus species regional abundances structure does not seem to influence much the results.

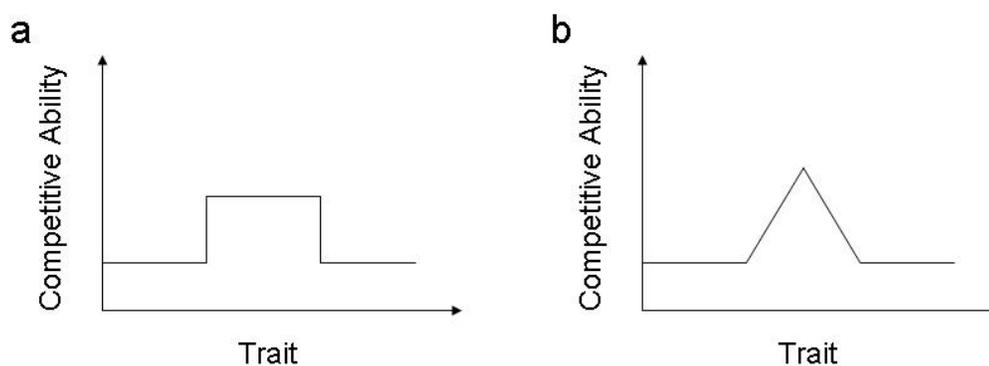
Model parameters	$R^2$ (without outliers) Equal abundances	% of outliers Equal abundances	$R^2$ (without outliers) Unequal abundances	% of outliers Unequal abundances	$R^2$ (without outliers) Unequal abundances correlated with trait values	% of outliers Unequal abundances correlated with trait values
$\ln(I)$	0.65	1	0.70	3	0.63	1
$h$	0.99	17	0.99	9	0.98	19
$\ln(A)$	0.73	8	0.86	9	0.81	11
$\ln(\sigma)$	0.82	9	0.86	16	0.47	3

**Table C3.**  $R^2$  and percentages of outliers for different regional abundances structures.

### 4-Effect of the shape of the filtering function

Two additional shapes of filtering function were considered: rectangular and triangular functions (Fig. C1). In the robust regressions between simulated and inferred parameters

values,  $R^2$  were similar in each case, but there were more outliers in the rectangular and triangular cases than in the Gaussian case treated in the main text (Table C4). A possible explanation for this is that the summary statistics used might perform less well for other shapes of filtering functions. For such other shapes, other summary statistics may be more adapted. Anyway, the method presented in this article is still working honorably for these alternative shapes of filtering function.



**Fig. C1.** Panel a: rectangular filtering function. Panel b: triangular filtering function.

Model parameters	$R^2$ (without outliers) Gaussian function	% of outliers Gaussian function	$R^2$ (without outliers) Rectangular function	% of outliers <sup>a</sup> Rectangular function	$R^2$ (without outliers) Triangular function	% of outliers <sup>a</sup> Triangular function
$\ln(I)$	0.65	1	0.92	16	0.87	9
$H$	0.99	17	0.89	36	0.59	19
$\ln(A)$	0.73	8	0.91	27	0.91	30
$\ln(\sigma)$	0.82	9	0.51	13	0.77	24

**Table C4.**  $R^2$  and percentages of outliers for different shapes of filtering function.

<sup>a</sup> For rectangular and triangular functions, the criterion for deciding that a point was an outlier was modified in “outliers = points with a weight in the robust regression below 1” since the former criterion was not picking all the visually-detected outliers.

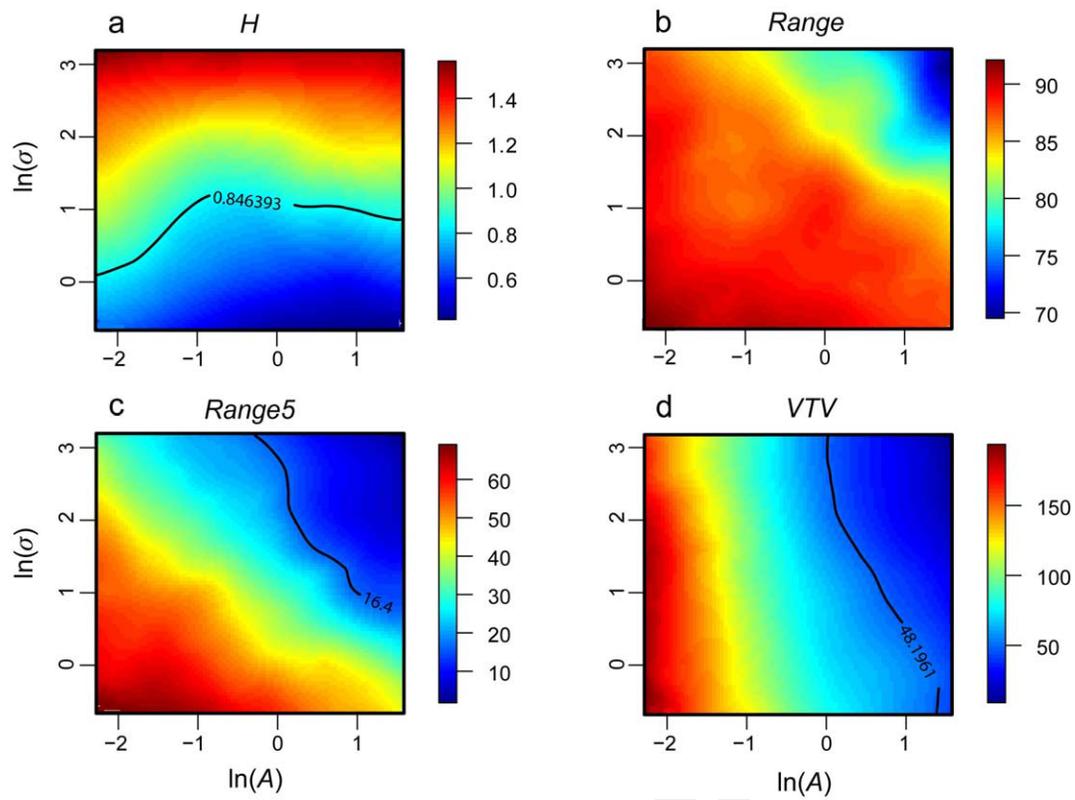
### **B/Test of environmental filtering at the individual level**

The robustness of the results presented in the main text was assessed with different

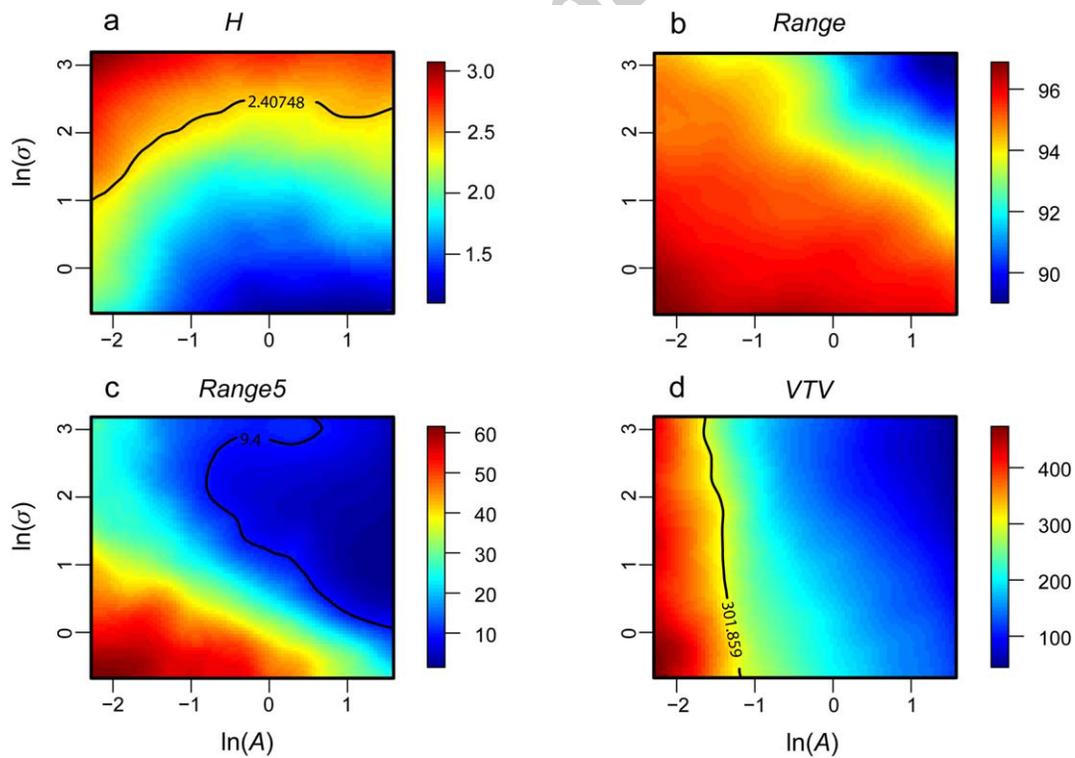
values of local species richness and regional species richness, and with different regional abundance structures and shapes of the environmental filtering functions. For each situation, the neutral reference consisted in 1000 simulations with flat filtering functions, and the same values of local and regional species richness, together with the same regional abundance structure. I compared the value of four summary statistics among neutral and filtered communities: Shannon's index  $H$ , the range of trait values in the local community  $Range$ , the range of trait values in the 5 locally most abundant species  $Range5$ , and the variance of trait values among individuals  $VTV$ . I then compared the relative power of tests of environmental filtering based either on the distribution of a trait at the species level using Cornwell et al. (2006)'s method (ie the summary statistics  $Range$ ), or on the distribution of a trait at the individual level using either  $VTV$  or  $Range5$  (see also Villéger et al. 2008 for other approaches), or on the species abundances evenness via Shannon's index  $H$  (without trait information).

#### 1-Effect of the number of species in the community

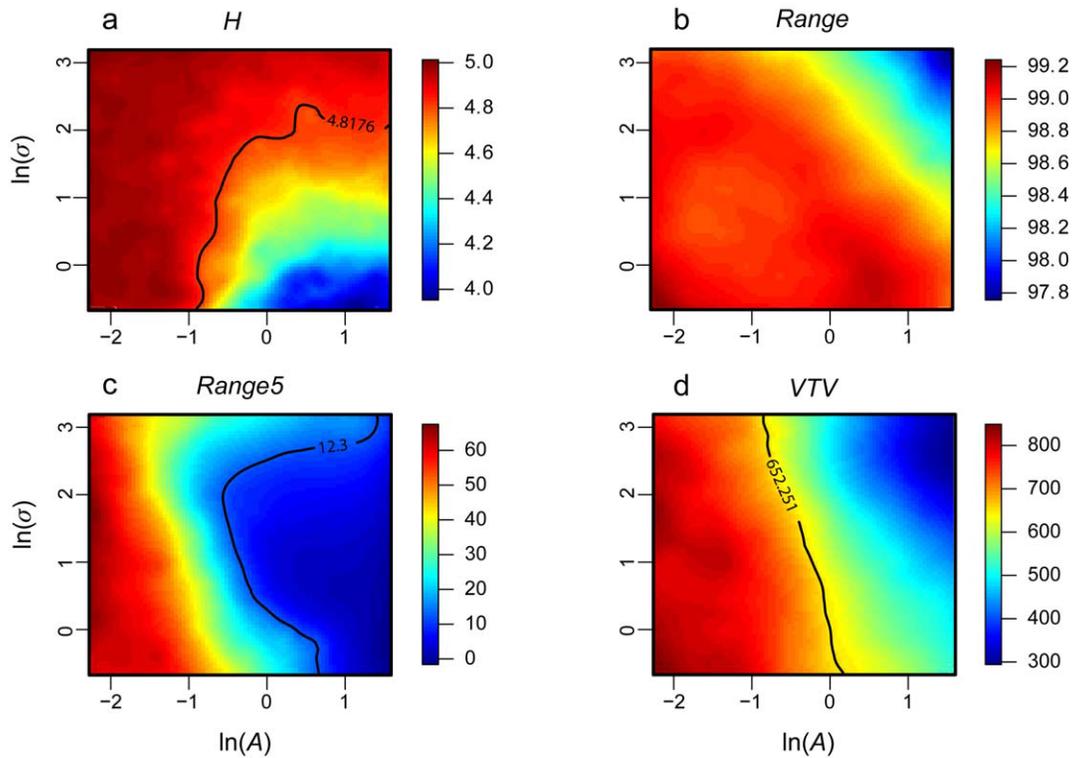
Three additional values of local species richness were considered:  $S=20$ , 50, and 200. 1000 neutral communities were simulated with parameter  $\ln(I)$  randomly drawn in  $[0 ; 4]$  for  $S=20$  (resp.  $[1.5 ; 4.5]$  for  $S=50$ , and  $[4.5 ; 6]$  for  $S=200$ ). 1000 filtered communities were simulated with the same prior for  $\ln(I)$  as in neutral communities, and with  $h$  equal to 25,  $\ln(A)$  in  $[\ln(0.1) ; \ln(5)]$ , and  $\ln(\sigma)$  in  $[\ln(0.5) ; \ln(25)]$ . The results were similar to those obtained in the main text (Fig. 5), with the statistics  $Range$  failing to detect environmental filtering, while the use of the three other statistics managed to detect it (Fig. C2, C3, C4). Furthermore, in communities with intermediate values of local richness  $S$  ( $S=50$ , and  $S=100$ ), environmental filtering was easier to detect when of low intensity (ie. low  $A$ , and large  $\sigma$ ).



**Fig. C2.** Filtering intensity and detectability of non-neutral patterns –  $S=20$ .



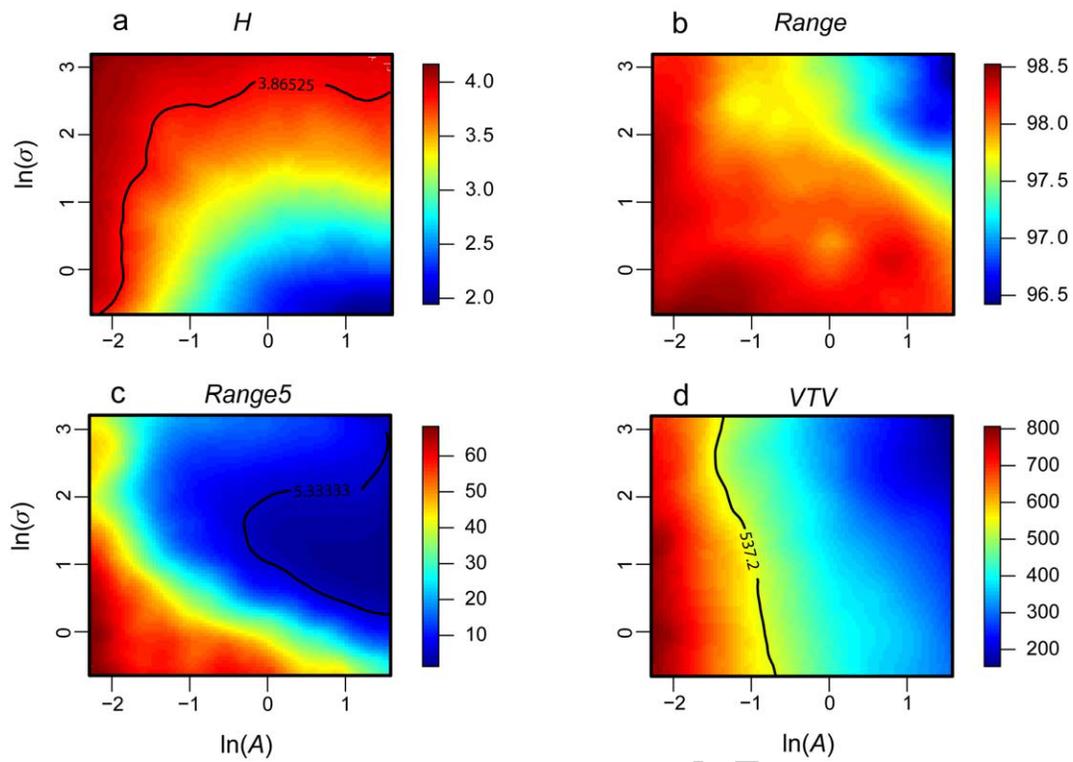
**Fig. C3.** Filtering intensity and detectability of non-neutral patterns –  $S=50$ .



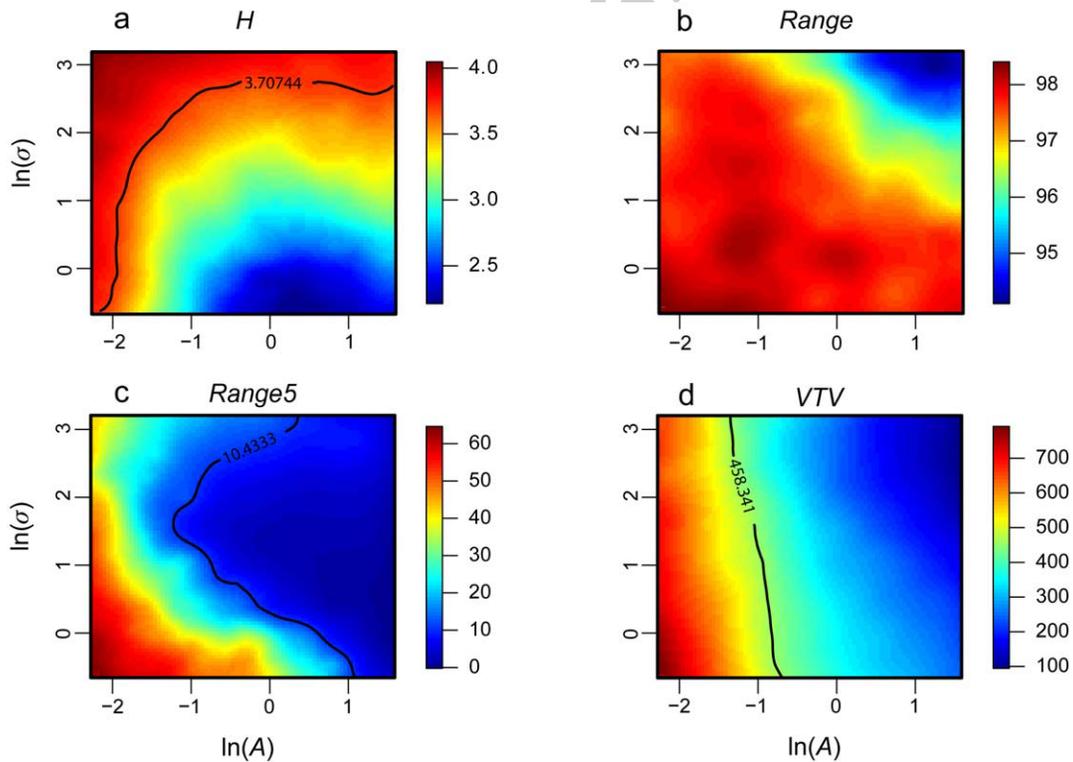
**Fig. C4.** Filtering intensity and detectability of non-neutral patterns –  $S=200$ .

## 2-Effect of the number of species in the regional pool

Two additional values of regional species richness were considered:  $S_{\text{tot}}=300$ , and  $S_{\text{tot}}=3000$ . 1000 neutral communities were simulated with parameter  $\ln(I)$  randomly drawn in  $[3 ; 5.5]$  for  $S_{\text{tot}}=300$  (resp.  $[3 ; 5]$  for  $S_{\text{tot}}=3000$ ). 1000 filtered communities were simulated with the same prior for  $\ln(I)$  as in neutral communities, and with  $h$  equal to 25,  $\ln(A)$  in  $[\ln(0.1) ; \ln(5)]$ , and  $\ln(\sigma)$  in  $[\ln(0.5) ; \ln(25)]$ . The results were similar to those obtained in the main text (Fig. 5), with the statistics *Range* failing to detect environmental filtering, while the use of the three other statistics managed to detect it (Fig. C5, C6). *Range5* was a little less efficient for  $S_{\text{tot}}=300$ , while the effect of  $S_{\text{tot}}$  on the performance of the three other statistics for detecting environmental filtering was very small.



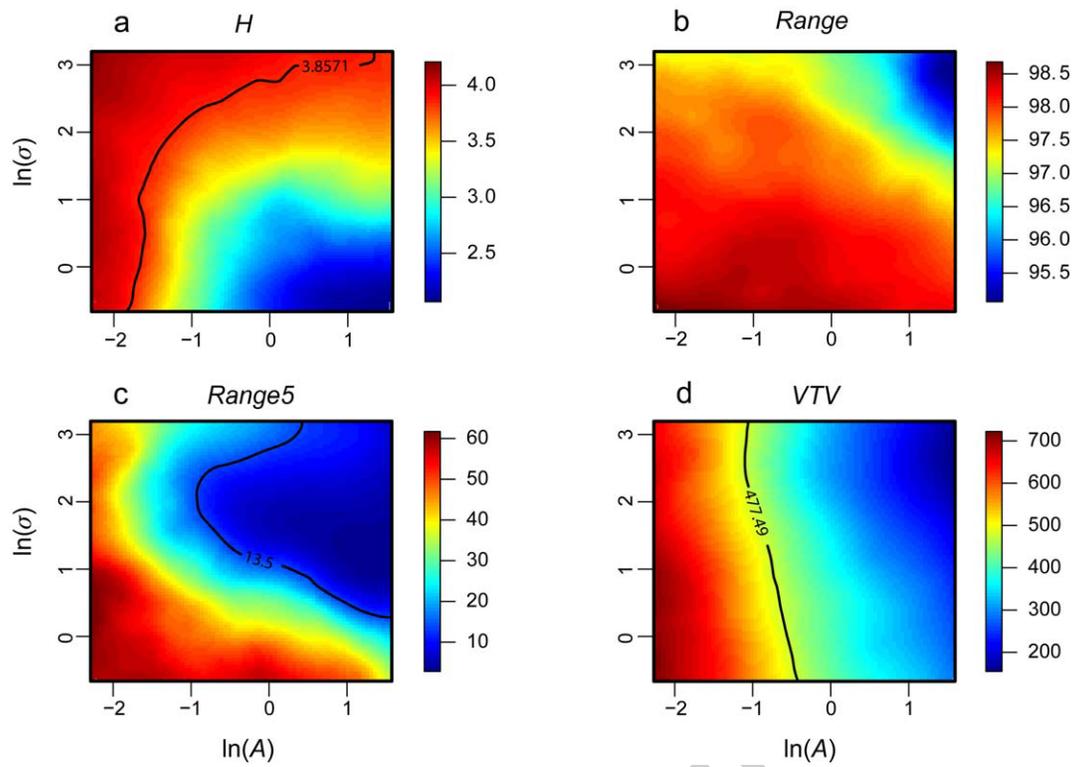
**Fig. C5.** Filtering intensity and detectability of non-neutral patterns –  $S_{\text{tot}}=300$ .



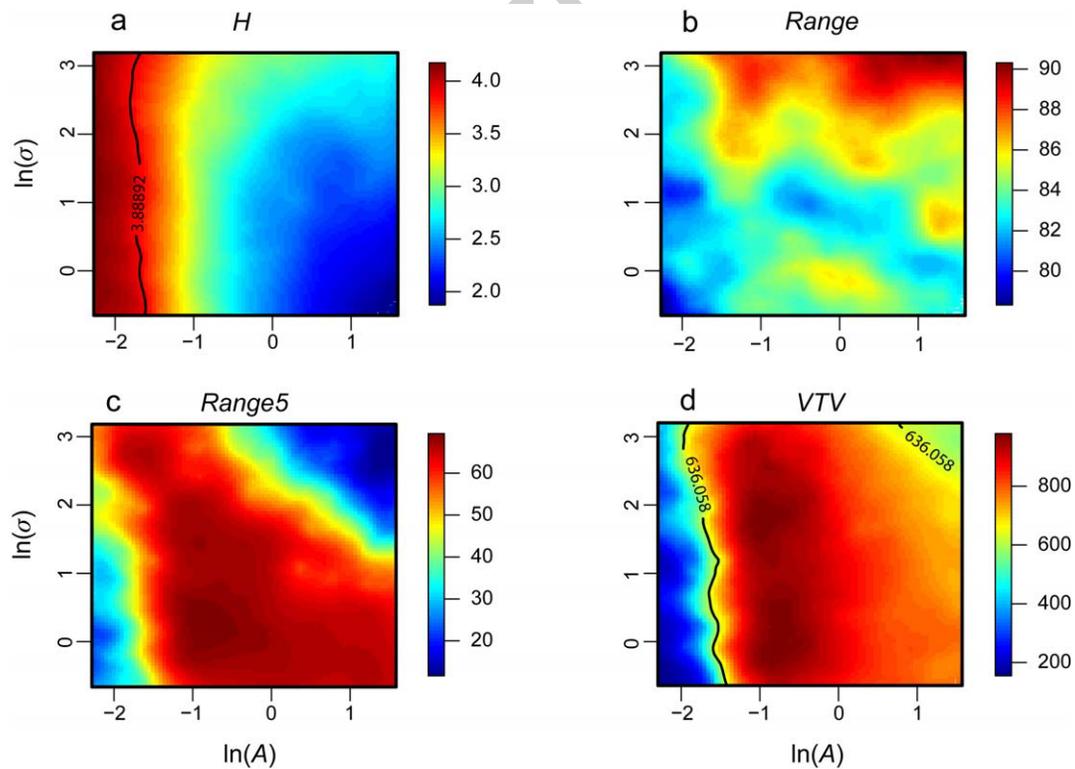
**Fig. C6.** Filtering intensity and detectability of non-neutral patterns –  $S_{\text{tot}}=3000$ .

### 3-Effect of the structure in regional abundances

In the main text, regional abundances are all equal. Two additional cases were considered. In both cases, regional richness was equal to 1000, and the regional abundances were generated with the use of a Fisher logseries with parameter  $\alpha$  equal to 200 using the function “fisher.ecosystem” of the R package “untb”. In the first case, regional abundances were randomly attributed to each species, independently of their traits, while in the second case, species regional abundances were not independent of species traits: I picked at random a position in the trait space, and then placed the species around this trait value in decreasing order of abundances so as to form a peak in the trait-regional abundance plot. 1000 neutral communities were simulated with parameter  $\ln(J)$  randomly drawn in  $[3.5 ; 5.5]$ . 1000 filtered communities were simulated with the same prior for  $\ln(J)$  as in neutral communities, and with  $h$  equal to 25,  $\ln(A)$  in  $[\ln(0.1) ; \ln(5)]$ , and  $\ln(\sigma)$  in  $[\ln(0.5) ; \ln(25)]$ . When regional abundances were uncorrelated with trait values, the results were similar to those obtained in the main text (Fig. 5), with the statistics *Range* failing to detect environmental filtering, while the use of the three other statistics managed to detect it (Fig. C7). When regional abundances were correlated with trait values, different results were obtained (Fig. C8). In the case studied, regional abundances were maximal for trait values equal to 90, while environmental filtering was favoring species with trait values equal to 25. Hence, immigration from the regional pool and environmental filtering were favoring species with different trait values. In this case, it was only possible to detect environmental filtering with the use of  $H$  or  $VTV$ : in the parameter space situated at the right of the black line,  $H$  (resp.  $VTV$ ) was smaller (resp. larger) in filtered communities than in neutral communities (Fig. C8).



**Fig. C7.** Filtering intensity and detectability of non-neutral patterns – Unequal regional abundances, not correlated with trait values.

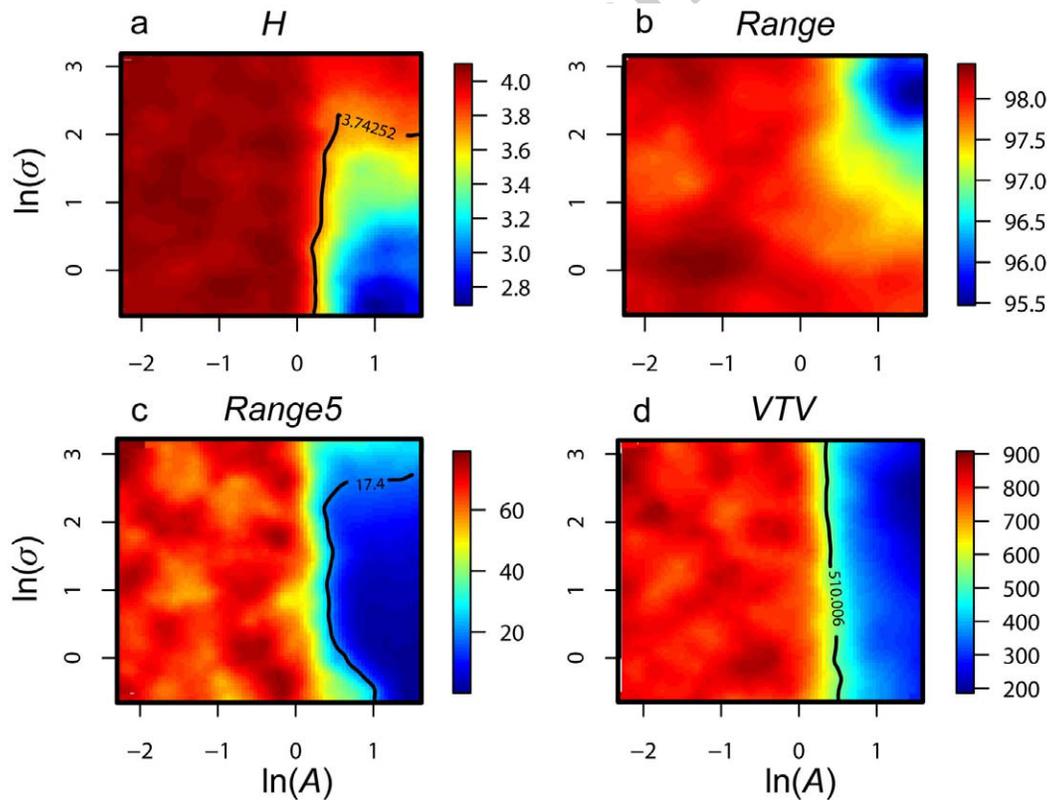


**Fig. C8.** Filtering intensity and detectability of non-neutral patterns – Unequal regional

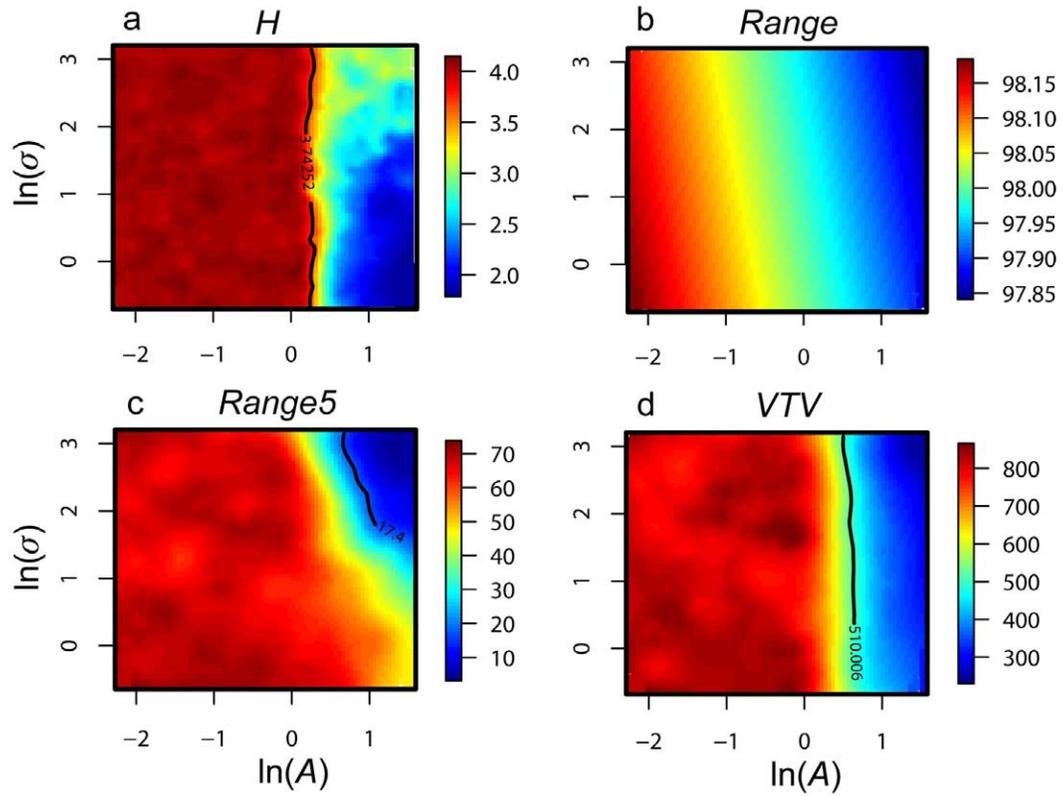
abundances, correlated with trait values (maximal regional abundance for trait value equal to 90).

#### 4-Effect of the shape of the filtering function

Two other shapes of environmental filtering functions (rectangular and triangular) were used (Fig. C1). 1000 neutral communities were simulated with parameter  $\ln(I)$  randomly drawn in  $[3 ; 5]$ . 1000 filtered communities were simulated with the same prior for  $\ln(I)$  as in neutral communities, and with  $h$  equal to 25,  $\ln(A)$  in  $[\ln(0.1) ; \ln(5)]$ , and  $\ln(\sigma)$  in  $[\ln(0.5) ; \ln(25)]$ . The results were similar to those obtained in the main text (Fig. 5), with the statistics *Range* failing to detect environmental filtering, while the use of the three other statistics managed to detect it (Fig. C9, C10).



**Fig. C9.** Filtering shape and detectability of non-neutral patterns – Rectangular functions.



**Fig. C10.** Filtering shape and detectability of non-neutral patterns – Triangular functions.