Diversity and Distributions, (Diversity Distrib.) (2008) **14**, 786–798

BIODIVERSITY RESEARCH

Distribution of parasitic fungal species richness: influence of climate versus host species diversity

Corinne Vacher[1]*, Denis Vile[2], Emmanuelle Helion[1], Dominique Piou[1,3] and Marie-Laure Desprez-Loustau[1]

[1]*UMR1202 Biodiversité Gènes et Communautés, INRA, F-33883 Villenave d'Ornon, France,* [2]*UMR759 Laboratoire d'Ecophysiologie des Plantes sous Stress Environnementaux, INRA, F-34060 Montpellier, France,* [3]*Département de la Santé des Forêts, Ministère de l'Agriculture et de la Pêche, F-33150 Cenon, France*

*Correspondence: Corinne Vacher, INRA, UMR1202 Biodiversité, Gènes et Communautés (BioGeCo), Equipe de Pathologie Forestière, Centre de Bordeaux, BP 81-71, Avenue E. Bourlaux, F-33883 Villenave d'Ornon Cedex, France. E-mail: cvacher@bordeaux.inra.fr

ABSTRACT

One major challenge in parasitology and epidemiology is determining whether the richness of parasitic and infectious diseases simply tracks host diversity or is largely determined by exogenous factors, such as climate-forced variables. We addressed this issue by analysing a 30-year survey of fungal diseases in French forests. We first combined generalized linear models and stepwise analyses to select the habitat descriptors that may account for variations in parasitic fungal species richness. Our results suggest that host species diversity is not a major determinant of parasite richness. Temperature seasonality, host abundance, and the composition of host species assemblages may play a greater role. Then we used structural equation modelling to investigate the links between these habitat descriptors and parasitic fungal species richness. Our results showed that climatic and host species descriptors had not only direct effects on species richness, but also indirect effects (via host species and sampling effort, respectively). Our results also showed that the direct effects of climate and host species were roughly equal. We therefore conclude that it is important to take into account exogenous factors when investigating the potential causes of spatial variation in the richness of parasitic diseases, in particular for plant diseases.

Keywords

Biodiversity, climate, disease, fungi, host–parasite relationships, trees.

INTRODUCTION

A habitat can be defined as a physical place, on a particular spatial and temporal scale, inhabited or potentially inhabited by an organism (Kearney, 2006). It can be described in terms of biotic and abiotic variables without reference to the traits of the organism concerned (Kearney, 2006). This definition therefore applies to both free-living and parasitic organisms. This concept of 'habitat' is useful for accounting for or predicting the spatial distributions of species. The most common approach involves statistically associating the presence or absence of a species, or abundance data, with spatial variables describing the habitat (Kearney, 2006). Correlative approaches of this type have been widely used in free-living organisms to study the spatial distribution of a given species (e.g. Osborne *et al.*, 2001; Kuhn *et al.*, 2002; Kearney *et al.*, 2003; Gebre-Michael *et al.*, 2004) or the spatial distribution of species richness within a taxonomic group (e.g. O'Brien, 1998; Wohlgemuth, 1998; Real *et al.*, 2003; Lobo *et al.*, 2004; Moser *et al.*, 2005; Cabrero-Sanudo & Lobo, 2006). In contrast, they have seldom been used to interpret the spatial distribution of a single parasite species or the spatial distribution of parasite species richness (but see Cocu *et al.*, 2005; Cumming & Guegan, 2006).

Indeed, a slightly different definition of habitat is commonly used in parasitology. Parasitologists usually consider the parasite to live in or on discrete habitats, or hosts (Poulin, 1997), thereby excluding the matrix surrounding host individuals from the habitat. This definition of habitat matches that of Kearney for obligatory endoparasites (the physical place inhabited or potentially inhabited being the internal environment of the host). However, it is of limited suitability for ectoparasites, which live in direct contact with the biotic or abiotic elements surrounding host individuals, and for facultative parasites, which may live freely outside of their host species during part of their life cycle. In particular, the definition of habitat commonly used in parasitology is not appropriate for plant parasites. Even parasites living within a plant host are subject to climatic variations, as plants do not regulate their internal temperature and only partially regulate their water status.

As a consequence of the host-centred definition of habitat used in parasitology, many studies have searched for correlates

DOI: 10.1111/j.1472-4642.2008.00479.x

of parasite species richness in host-related traits, such as host population size, host population geographical isolation, or host species diversity (see Guégan *et al.*, 2005 for review). Only a few studies have demonstrated an effect of abiotic conditions on the spatial distribution of parasite species richness. The global-scale study by Guernier *et al.* (2004) showed that climatic factors, including annual variations in rainfall, are key elements in the negative relationship between latitude and species richness for several taxa responsible for human parasitic and infectious diseases (PID). Interestingly, the relationship between latitude and PID species richness was not significant for taxa found entirely within the host (e.g. virus, directly transmitted bacteria and fungi) and therefore less affected by climatic variability (Guernier *et al.*, 2004). Similarly, a significant latitudinal gradient in the parasite species richness of wild primates has been reported for protozoan parasites, but not for helminths and viruses. This difference may be related to the higher percentage of arthropod-transmitted species among protozoan parasites, together with a positive effect of temperature on vector activity and parasite development rate (Nunn *et al.*, 2005). Krasnov *et al.* (2006) showed that global flea species richness is positively correlated with that of their mammalian host species. The significant deviation of some regions of the world from the confidence interval of this positive relationship may be due to abiotic factors, such as precipitation or substrate, because the pre-imaginal development of fleas almost always occurs off the host (Krasnov *et al.*, 2006). Finally, a pan-African analysis of tick-transmitted human pathogens revealed a weak but significant correlation between the abiotic environment and pathogen species richness (Cumming & Guegan, 2006), and the study of helminth communities in red-legged partridge across Spain showed that helminth species richness was directly correlated with mean temperature (Calvete, 2003).

Based on these results, parasitologists and epidemiologists clearly need to determine whether the richness of parasitic and infectious diseases simply tracks host diversity or is largely determined by exogenous factors, such as climate-forced variables (Guégan *et al.*, 2005). Here we addressed this issue by analysing a 30-year survey of fungal diseases in French forests. Besides climate variables and host-related variables, we included the spatial variations in sampled area and sampling effort when modelling the richness of parasitic fungal species, as advised by several authors (Poulin, 1997; Guégan *et al.*, 2005). We also took into account the potential effects of host-related variables on sampling effort (Guégan & Kennedy, 1996). The following questions were addressed: (1) What are the relative contributions of climate variables, host-related variables and sampling variables to the observed richness of parasitic fungal species? (2) What are the relative contributions of direct versus indirect (via host-related variables) effects of climate variables on parasitic fungal species richness? (3) What are the relative contributions of direct versus indirect (via sampling effort) effects of host-related variables on parasitic fungal species richness? In order to answer those questions, we first combined generalized linear models (GLM) and stepwise analyses to identify variables that may be important and which may be mechanistically linked to the presence of fungal species (Mitchell, 2005). Then we used structural equation modelling (Shipley, 2000; Pugesek *et al.*, 2003) to investigate the direct and indirect effects of the selected variables on parasitic fungal species richness.

METHODS

The data

Fungal species records

For this study, we used the data base managed by the Forest Health Department (Département de la Santé des Forêts, DSF; French Ministry of Agriculture and Fisheries). This data base contains information supplied by a network of foresters trained in the observation of pathological and entomological phenomena and of all types of tree decline. The primary aim of this monitoring network is to prevent disease spread, pest outbreaks, and other types of damage by alerting the authorities as soon as a threat to forest health is identified. The foresters report damages they have noticed during their daily work in the forest, when they consider that these latter may reduce the survival or the economical value of the trees. Over 67,000 cases of insect attack, disease, abiotic stresses, or decline were reported between 1972 and 2005. Most of them (*c.* 97%) were reported after 1989, when the monitoring network experienced a deep reorganization at the national level. We retained 11,341 observations of forest tree diseases from these reports, for which the causal agent was a parasitic fungus identified to species level (fungal species identification was based on disease symptoms or, when necessary, microscopy and isolation by a specialist laboratory (Laboratoire National de la Protection des Végétaux, Nancy)). After correction for species names synonyms, we obtained a data set of 161 fungal species (see Appendix S1 in Supplementary Material).

Geographical area

The observed parasitic fungal species richness was calculated for 88 geographical units covering the whole of France (Fig. 1). These geographical units corresponded roughly to French administrative *départements* and were defined so as to be of approximately equal size. The *départements* of Paris, Yvelines, Essonne, Hauts-de-Seine, Seine-Saint-Denis, Val-de-Marne, and Val-d'Oise were grouped together into a single geographical unit because of their small sizes. Corsica was also considered as a single unit, although actually composed of two *départements*. Disease records from the Territoire de Belfort, the smallest French *département*, were added to those of the surrounding *départements*. As a result, the average area of the geographical units was 6191.1 ± 1410.9 km^2 (mean ± standard deviation).

Sampling bias estimates

Spatial variations in the observed parasitic fungal species richness may reflect spatial variations in the recording intensity of DSF foresters. We assumed that the average risk of tree damage

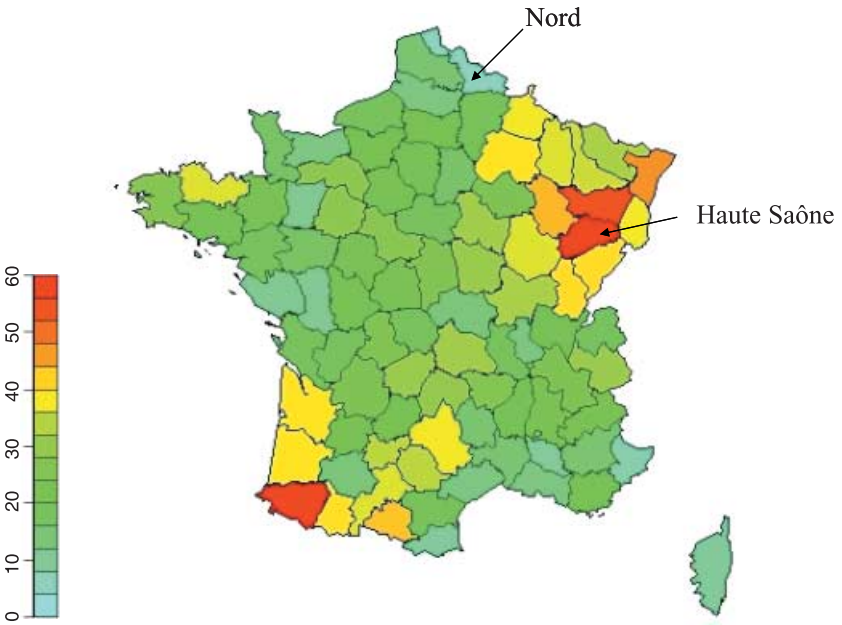

Figure 1 Recorded species richness of parasitic forest fungi, ranging from five species in the Nord to 58 species in Haute-Saône.

was equal over the entire French territory and consequently used the total number of DSF records per geographical unit as an estimate of sampling intensity. All types of records except fungal diseases records were included in this calculation. In this way, estimates of sampling intensity and parasitic fungal species richness were based on independent data sets. The sampling intensity was partitioned into two components: the sampled area (Area) and the sampling effort per area unit (Seffort). We used the area of the geographical unit as an estimate of the area sampled and the ratio of the number of observations of tree damage caused by factors other than parasitic fungi to the area of the geographical unit as an estimate of sampling effort. Alternatively, we could have used the forest area in each geographical unit as an estimate of the area sampled and the ratio of the number of observations of tree damage caused by factors other than parasitic fungi to the forest area as an estimate of sampling effort. Both measures of sampling effort were highly correlated (Pearson correlation test on log-transformed variables; $r = 0.790$; $P < 0.001$). However, the second definition presented the disadvantage of basing the calculation of Seffort on a variable related to host species (the forest area). Since our aim in this study was to tease apart the effects of sampling effort and host-related factors on parasitic fungal species richness, we chose the first definition of Area and Seffort.

Climatic descriptors

We used the predictions from the AURELHY (Analysis Using the Relief for Hydrometeorology) model (Benichou & Le Breton, 1987) supplied by the French National Meteorological Office (Météo-France) to compile data for 29 climatic variables for each geographical unit. This model allows the extrapolation of meteorological records made at 65 meteorological stations to a 4×4 km grid by taking into account the local topography. We first used the monthly predictions given for the centroids of the *départments*, averaged on the period 1961–90, to calculate 24 variables (Table 1) describing seasonal variations in temperature, precipitation, and potential evapotranspiration. Potential evapotranspiration was calculated with Turc's method (Turc, 1963). Since we expected parasitic fungal species richness to be highly dependant on air and soil moisture, we calculated five additional variables to characterize water balance seasonality. Water balance was defined as the difference between precipitation and evapotranspiration (Table 1).

Host species descriptors

We used the 2000 census report of the Inventaire Forestier National – the French governmental agency in charge of forest resources assessment – to estimate the wood volume and area covered by 55 taxa of forest trees within each geographical unit. As all but five of the tree taxa corresponded to species, we will refer to them hereafter as species (see Appendix S2). Moreover, as all but four species were reported in the DSF data base as being the host of one or several fungal species, we refer to them hereafter as host species. For each geographical unit we used eight variables (Table 1) to characterize the diversity of host species, the composition of host species assemblages, and the abundance of host species. The diversity of host species was characterized by the total number of host species (Stree) and the evenness of host species abundance (Jtree). The composition of host species assemblages was characterized by focusing on the 20 most represented host species (see Appendix S2). Together, these species covered more than 95% of the surface area of French forests. We carried out principal component analysis (PCA) of the relative areas covered by these dominant species in the 88 geographical units, using the ade4 package in R (Chessel *et al.*, 2005). The first and second axis of PCA accounted for 17.6 and 14.3% of the observed variation, respectively. Geographical units with negative scores on the first axis were characterized by a high abundance of two typically

Table 1 List of climatic and host species descriptors. Variables belonging to the same group were highly correlated (correlation coefficient > |0.65|). Variables that did not belong to any group were lowly correlated to all the other variables (correlation coefficient < |0.65|).

Abbreviation	Explanation	Group
Climatic descriptors		
Temperature		
Tspring	Mean temperature in the spring (March, April, May)	1
Tsummer	Mean temperature in the summer (June, July, August)	2
Tautumn	Mean temperature in the autumn (September, October, November)	1
Twinter	Mean temperature in the winter (December, January, February)	1
Tmin	Mean minimal temperature of the coldest month	3
Tmax	Mean maximal temperature of the warmest month	2
Trange	Within-year variation of temperature (Tmax–Tmin)	–
Tmean	Mean annual temperature	1
Tfrost	Average number of days of frost per year	3
Precipitation		
Pspring	Mean monthly precipitation in the spring	4
Psummer	Mean monthly precipitation in the summer	–
Pautumn	Mean monthly precipitation in the autumn	4
Pwinter	Mean monthly precipitation in the winter	4
Pmin	Mean precipitation of the driest month	–
Pmax	Mean precipitation of the rainiest month	–
Prange	Within-year variation of precipitation (Pmax–Pmin)	–
Ptot	Mean annual precipitation	4
Potential evapotranspiration		
PETspring	Mean monthly potential evapotranspiration in the spring	5
PETsummer	Mean monthly potential evapotranspiration in the summer	6
PETautumn	Mean monthly potential evapotranspiration in the autumn	5
PETwinter	Mean monthly potential evapotranspiration in the winter	7
PETmin	Minimum value of monthly potential evapotranspiration within a year	7
PETmax	Maximum value of monthly potential evapotranspiration within a year	6
PETrange	Within-year variation of potential evapotranspiration (PETmax–PETmin)	6
Water balance		
WBspring	Mean monthly water balance in the spring (Pspring–PETspring)	8
WBsummer	Mean monthly water balance in the summer (Psummer–PETsummer)	9
WBautumn	Mean monthly water balance in the autumn (Pautumn–PETautumn)	8
WBwinter	Mean monthly water balance in the winter (Pwinter–PETwinter)	4
WBmin	Minimum value of monthly water balance within a year (maximal monthly water deficit)	9
Host species descriptors		
Diversity of tree species		
Stree	Species richness of forest trees	–
Jtree	Relative diversity of tree species (Jtree = Htree/log(Stree), where Htree is the Shannon–Wiener diversity index calculated with the relative area under tree species)	–
Composition of tree species assemblages		
PCA1	PCA first axis scores	–
PCA2	PCA second axis scores	–
Pintro	Percentage of exotic species (ratio of the area under exotic species to the area of the geographical unit)	–
Pgymn	Percentage of gymnosperm species (ratio of the area under gymnosperms to the area of the geographical unit)	–
Abundance of forest trees		
Fcover	Forest cover (ratio of the area under forest to the area of the geographical unit)	–
Fvol	Forest volume (ratio of wood volume to forest area)	–

Mediterranean species (*Quercus ilex* and *Pinus halepensis*) and a supra-Mediterranean species (*Quercus pubescens*), whereas geographical units with negative scores on the second axis were characterized by a high abundance of two alpine species (*Abies alba* and *Picea excelsa*). In order to characterize the composition of host species assemblages, we also calculated the percentage of the surface area under forest covered by exotic host species (Pintro) and the percentage of the surface area under forest covered by

gymnosperms (Pgymn). A species was considered exotic if introduced into France after about 1500. Finally, the abundance of hosts was characterized for each geographical unit by the percentage of surface area under forest (Fcover) and the wood volume per unit of forest area (Fvol).

Selection of explanatory variables

We first selected a combination of sampling, climatic, and host-related variables that accounts well for differences in parasitic fungal species richness between most of the geographical units, by fitting GLMs (McCullagh & Nelder 1989) to the data. As expected for count data (Vincent & Haworth, 1983; McCullagh & Nelder, 1989; Guisan & Zimmermann, 2000), parasitic fungal species richness did not follow a normal distribution (Shapiro–Wilk test, $P = 0.002$), whereas the assumption of a Poisson distribution, which we tested using the VCD package in R (Meyer *et al.*, 2006), could not be rejected (goodness-of-fit test; $P = 0.055$ when species richness counts were grouped into six classes; $P = 0.078$ for 12 classes). We therefore used a logarithm link function to develop the GLM (McCullagh & Nelder, 1989; Guisan & Zimmermann, 2000), assuming a Poisson distribution of errors (see Crawley, 1993; Lobo *et al.*, 2004). All the explanatory variables were continuous and were standardized to eliminate the effects of differences in measurement scale. The five steps of the procedure of selection of explanatory variables are described below.

Transformation of explanatory variables

The effects of some explanatory variables on parasitic fungal species richness were expected to be nonlinear, whereas the GLM assumes linear relationships. Temperatures, in particular, were expected to have positive effects on parasitic fungal species richness in their medium range but negative effects in their low and high ranges (Petzoldt & Seaman, 2006). These possible curvilinear relationships were accounted for by separately testing quadratic and cubic functions for each explanatory variable and transforming the variables before inclusion in the GLM (see Nicholls, 1989; Heikkinen & Neuvonen, 1997; Lobo *et al.*, 2004; Moser *et al.*, 2005; Cabrero-Sanudo & Lobo, 2006). The cubic function of the explanatory variable was selected if it gave a significantly lower residual deviance than the quadratic function (chi-square test, $P < 0.001$). The quadratic function was selected if it gave a significantly lower deviance than the linear function (chi-square test, $P < 0.001$). The linear function was selected if it gave a significantly lower deviance than the null model (chi-square test, $P < 0.001$). The variable was not included in the GLM if it did not give a significant lower deviance than the null model (chi-square test, $P < 0.001$).

Removal of correlated explanatory variables

Besides the explanatory variables which did not cause any significant change in deviance when tested separately, we did not include the variables which were highly correlated to other explanatory variables (see Moser *et al.*, 2005). Within a group of highly correlated variables, only the variable accounting for the greatest change in deviance when tested separately was included in the GLM, in its selected form (i.e. linear, quadratic or cubic). Groups of highly correlated variables (Table 1) were defined as groups within which $|\tau| > 0.65$ for each pair of explanatory variables (see Moser *et al.*, 2005), where τ is the Kendall's tau correlation coefficient. Groups were identified by using a hierarchical cluster analysis based on the matrix of distances between variables, with the distance between two variables defined as $(1-|\tau|)$. Cluster analysis was performed in R (R Development Core Team, 2006), using the Cluster package (Maechler *et al.*, 2005).

Stepwise selection of a subset of explanatory variables

A GLM was then fitted to parasitic fungal species richness, using this selected set of transformed variables. A stepwise selection analysis based on Akaike's information criterion (AIC) (Hilborn & Mangel, 1997) was carried out to identify the most parsimonious and most accurate combination of explanatory variables. The decrease in deviance associated with each selected variable was tested for significance with a chi-square test, with a 5% significance level (Guisan & Zimmermann, 2000).

Assessment of the overall quality of the model

The goodness-of-fit of the GLM was assessed using AIC (a good model being one with low AIC value) and deviance reduction (a good model being one giving a large decrease in deviance). Deviance reduction was defined as described by Guisan & Zimmermann (2000):

$$D^2 = (\text{null deviance} - \text{residual deviance})/\text{null deviance}$$

And its value was adjusted for the number of observations (n) and predictors (p), using the following formula:

$$\text{Adj-}D^2 = 1 - [(n-1)/(n-\text{p})]^*[1 - D^2]$$

We also tested a key assumption of our modelling approach – the underlying Poisson distribution of errors – by examining the ratio of residual deviance to residual degrees of freedom of the model. The value obtained, the dispersion parameter (d), should equal 1.0. If not, the data would have been better modelled with another distribution law (e.g. negative binomial; see Stevens & Carson, 1999). We tested for overdispersion ($d > 1$) – the most common violation of the Poisson distribution assumption – with a chi-square goodness-of-fit test (see Stevens & Carson, 1999). We then investigated the overall appropriateness of the model in more detail, using diagnostic plots (McCullagh & Nelder, 1989; Guisan & Zimmermann, 2000; Everitt & Hothorn, 2006). We first plotted the residuals on a normal probability plot and checked the normality of the distribution of residuals with a Shapiro–Wilk test. We then plotted the residuals against the fitted values to evaluate the homogeneity of variance of the residuals across the range of predicted values. Finally, a jackknife procedure

(see Guisan & Zimmermann, 2000; Lobo *et al.*, 2004) was used to estimate the predictive power of the model. This procedure recalculated model coefficients as many times as there were geographical units, leaving out one geographical unit each time. The new coefficients were used to predict parasitic fungal species richness in the excluded geographical unit. Predictive power was estimated by examining the correlation between recorded and jackknife-predicted species richness with a Kendall's rank test, and by calculating the percentage error in the jackknife-predicted value for each geographical unit. The mean prediction error of the model was defined as the mean percentage error for the various geographical units (Lobo *et al.*, 2004).

Detection and removal of isolated discrepancies

As suggested by Everitt & Hothorn (2006), the analysis was finally re-conducted after removal of the geographical units which appeared as isolated discrepancies. A geographical unit was considered an isolated discrepancy if the associated residual value departed significantly from the normal distribution or if the associated Cook's distance was high (Everitt & Hothorn, 2006). The Cook's distance is a measure of the influence of an observation on estimation of the coefficients for a model, and several cut-off values have been proposed for the detection of overly influential observations. Here we used a cut-off value of 1 (Montgomery & Peck, 1982; Neter *et al.*, 1985). The five steps of the procedure of selection of explanatory variables were applied until the removal of geographical units ceased to improve the overall quality of the model. Indeed, it should be noted again that our aim here was not to realize a predictive model of parasitic fungal species richness to be applied to new geographical areas, but to select a combination of explanatory variables that accounts well for differences in species richness between most of the geographical units which were sampled. The variables of the final GLM were considered as the selected set of explanatory variables.

Assessment of direct and indirect effects of selected variables on species richness

The dependency relationships between the selected explanatory variables, and between the selected variables and parasitic fungal species richness, were then investigated by developing a structural equation model (SEM). It is noteworthy that all the geographical units were taken into account in this modelling step. The model was defined by following the procedure described below.

Model calibration

The selected explanatory variables and the response variables were first transformed to approach linearity of the bivariate relationships and normal distribution of the residuals. Those transformations helped to satisfy the assumption multivariate normality, which is required for structural equation modelling. Pearson's correlations between all transformed variables were calculated. The initial structural model assumed that climatic descriptors may have direct effects on parasitic fungal species richness and on host species descriptors, whereas host species descriptors may have direct effects on parasitic fungal species richness and on sampling effort. Sampled area and sampling effort only had direct effects on observed parasitic fungal species richness. The initial dependency relationships between the variables were chosen among all the possible relationships based on our knowledge of the system and on the observed covariance structure. The model was then improved step by step, by removing the dependency relationships that were not significant and by adding other relationships that seemed ecologically plausible. At each step, the structural model was estimated using the MPLUS program version 2.14 (Muthén & Muthén, 1998, 2003).

Model evaluation

The fit of the predicted covariance matrix to the observed covariance matrix was evaluated using the standard maximum likelihood (ML) estimator and the robust maximum likelihood estimator (MLM). This latter is more appropriated than the standard ML under non-normal data conditions. The goodness-of-fit between the observed and predicted covariance structures was expressed as a goodness-of-fit chi-square statistic. A significant goodness-of-fit statistic indicates that the model does not fit the data. Once a model has not been rejected as ecologically plausible, parameter estimates can be tested for significance using the z statistics. Two other indices (RMSEA and CFI) were also used to assess the closeness of fit. Good models have a RMSEA < 0.05 and a CFI > 0.95.

Direct and indirect effects assessment

The final SEM model was used to assess the direct and indirect effects of all the included explanatory variables on parasitic fungal species richness. The direct effect of an explanatory variable was equal to the standardized coefficient along the direct path going from the variable to parasitic fungal species richness. The indirect effect of an explanatory variable along a given indirect path was calculating by multiplying the standardized coefficients along this path. The total indirect effect of an explanatory variable was the sum of all its indirect effects. Only the indirect effects through directed paths were taken into account in this calculation (free or unexplained correlations were excluded). The total effect of an explanatory variable on parasitic fungal species richness was the sum of its direct and indirect effects. The relative contribution of the direct effect to the total effect was calculated as $|DE|/(|DE| + |IE|)$, where DE is the direct effect and IE is the total indirect effect (see Menendez *et al.*, 2007).

RESULTS

Selection of explanatory variables

We first removed 15 climatic variables from the set of 39 explanatory variables because they were found to be strongly correlated with other variables. Three variables (Pintro, Jtree, Pmax) were also removed because they did not significantly affect deviance

Table 2 Summary of the selection of explanatory variables during the three modelling steps.

Category	Variable (correlation group)	Model A			Model B			Model C		
		P	D^2	D^2_step	*P*	D^2	D^2_step	*P*	D^2	D^2_step
Sampling	Area	1	3.57	3.57***	1	2.89	2.89***	–	–	–
	Seffort	3	64.66	66.39***	3	66.55	68.03***	3	67.72	67.72***
Temperature	Twinter (1)	3	23.74	–	3	23.35	3.23**	3	23.40	5.22***
	Tsummer (2)	–	–	–	3	10.92	4.42***	3	9.55	6.05***
	Tmax (2)	2	8.85	–	–	–	–	–	–	–
	Tmin (3)	3	18.84	1.26	3	16.21	–	3	14.82	–
	Trange	3	9.63	4.41***	3	18.57	–	3	21.63	–
Precipitation	Pspring (4)	–	–	–	1	9.66	0.58	1	9.10	0.05
	Ptot (4)	2	8.61	2.56*	–	–	–	–	–	–
	Psummer	2	28.72	–	2	29.38	–	2	26.23	–
	Pmin	2	23.82	0.56	2	22.31	–	2	18.83	–
	Pmax	–	–	–	2	4.37	–	2	4.12	–
	Prange	1	3.36	–	3	8.51	–	3	10.11	–
Potential evapotranspiration	PETspring (5)	–	–	–	3	8.92	–	3	6.94	–
	PETautumn (5)	1	5.45	0.15	–	–	–	–	–	–
	PETrange (6)	1	11.21	2.20**	1	12.20	–	1	13.33	–
	PETwinter (7)	1	2.36	0.20	–	–	–	–	–	–
	PETmin (7)	–	–	–	–	–	–	2	3.90	–
Water balance	WBspring (8)	2	13.69	–	1	11.47	–	1	11.50	–
	WBsummer (9)	3	23.91	–	2	21.58	–	2	22.04	–
Host diversity	Stree	1	10.25	–	1	11.52	–	2	11.99	–
Host assemblages	PCA1	3	18.54	–	2	15.12	2.21**	3	19.15	5.68***
	PCA2	1	8.04	–	1	6.46	–	1	6.30	–
	Pintro	–	–	–	–	–	–	3	7.57	–
	Pgymn	3	5.30	1.44	3	5.84	2.17*	3	5.41	1.82*
Host abundance	Fcover	2	25.71	1.35	1	27.58	0.25	1	25.14	0.03
	Fvol	1	24.88	–	1	20.37	–	1	18.38	–

P: Polynomial form selected for inclusion in the generalized linear model (1, linear; 2, quadratic; 3, cubic).

D2: Decrease in deviance (%) caused by the variable when tested separately.

D2_step: Decrease in deviance (%) caused by the variable when included in the model after selection by the stepwise procedure. Significant changes in deviance are reported (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).

when tested separately. Then a combination of 11 variables was selected among the 21 remaining variables by stepwise analysis (AIC = 555. 1823, $D^2 = 84.12\%$, Adj-$D^2 = 82.06\%$, d = 1.17). Although the dispersion parameter of the model (called Model A and described in Table 2) exceeded 1, overdispersion was not significant (goodness-of-fit test; $P = 0.17$). Moreover, the model was of satisfactory overall quality: the hypothesis of normality could not be rejected for the distribution of residuals (Shapiro–Wilk test, $P = 0.6$), the variance of the residuals was uniform across the range of predicted values and the correlation between observed and jackknife-predicted values was significant and positive ($\tau = 0.63$, $P < 0.0001$). However, a few observations indicated abnormal behaviour with respect to the theoretical assumptions of the model: three geographical units (Marne, Var, and Pyrénées Orientales) were identified as outliers in diagnostic plots and three geographical units (Vosges, Cantal, and Alpes de Haute-Provence) had a Cook's distance greater than 1. In addition, the mean jackknife percentage error of the model was quite high (31.20%). In particular, we found that four geographical units had a jackknife percentage error exceeding 100%: three of them had been identified as overly influential observations (Vosges, Cantal, and Alpes de Haute Provence) and the fourth one was the Nord *département*. We were not surprised by this latter result because we know that the DSF foresters have encountered some problems in the Nord *département* (a densely populated, very industrialized area with a very few forests). The recorded number of parasitic fungal species in the Nord *département*, which is the lowest among all geographical units (Fig. 1), may therefore be largely misrepresentative. We first removed the six geographical units identified as isolated discrepancies (Marne, Var and Pyrénées Orientales, Vosges, Cantal and Alpes de Haute-Provence) from the data set to improve the goodness-of-fit of the model for the remaining units. The resulting model also failed in predicting parasitic fungal species richness in the Nord *département*: the jackknife percentage error associated with the Nord (167%) was well above the percentage errors for all others geographical units (< 100%). Therefore we chose to remove the Nord *département* from the data set, in addition to the six

Table 3 Pearson's correlations between all the variables included in the structural equation model. See Table 1 for variables names.

	Sfungi§	Twinter	Tsummer	Pspring§	PCA1	Pgymn§	Fcover	Fvol	Seffort§	Area§
Sfungi§	1									
Twinter	–0.325**	1								
Tsummer	–0.281**	0.712***	1							
Pspring§	0.305**	–0.320**	–0.301**	1						
PCA1	0.311**	–0.348***	–0.500***	–0.107 ns	1					
Pgymn§	0.094 ns	–0.103 ns	0.068 ns	0.205 ns	–0.496***	1				
Fcover	0.465 ***	–0.318**	0.090 ns	0.370***	–0.280**	0.450***	1			
Fvol	0.466 ***	–0.589***	–0.455***	0.103 ns	0.645***	0.067 ns	0.142 ns	1		
Seffort§	0.779 ***	–0.411***	–0.189 ns	0.364***	0.108 ns	0.237*	0.641***	0.466***	1	
Area§	0.217 *	0.151 ns	–0.002 ns	0.037 ns	0.087 ns	–0.068 ns	–0.053 ns	–0.068 ns	–0.084 ns	1

§log-transformed. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, ns, not significant.

geographical units identified as outliers, in order that this misrepresentative observation does not influence the selection of explanatory variables. The resulting model was called Model B.

As in the previous step, the number of explanatory variables was initially reduced from 39 to 21. In total, eight variables (Table 2) were selected by the stepwise procedure (AIC = 501.05, D^2 = 83.79%, Adj-D^2 = 82.23%, d = 1.03). The model fulfilled most of the quality criteria: the dispersion parameter was close to 1 (goodness-of-fit test for overdispersion; $P = 0.40$), the hypothesis of normality could not be rejected for the distribution of residuals (Shapiro–Wilk, $P = 0.5$), the variance of the residuals was uniform across the range of predicted values, all the geographical units had a Cook's distance below 1 and a jackknife percentage error below 100%, and the correlation between observed and jackknife-predicted values was significant and positive ($\tau = 0.69$, $P < 0.0001$). The mean jackknife percentage error was reduced by more than 10% with this model (from 31.20% to 20.85%). However, four geographical units (Bouches du Rhône, Creuse, Aveyron, and Haute-Saône) departed from the assumptions underlying the model in diagnostic plots. We therefore performed a third modelling step. The four geographical units in question were removed from the data set, and the resulting model is called Model C.

In this third step, 22 variables were initially selected (Table 2). The stepwise procedure retained seven explanatory variables (AIC = 460.16, D^2 = 87.35%, Adj-D^2 = 86.26%, d = 0.76), all of which were present in Model B. The partial effects of these seven variables on predicted parasitic fungal species richness were qualitatively similar in models B and C. The removal of the four geographical units did not therefore change the model qualitatively. We chose to retain Model B as the final model, because Model C displayed slight underdispersion and had only slightly greater predictive power than Model B (the mean jackknife percentage error was decreased by only 3.10%).

The final model had eight explanatory variables: two sampling variables (Area and Seffort), three host-related variables (Fcover, Pgymn and PCA1), and three climatic variables (Twinter, Tsummer, and Pspring). The model equation was:

$$\begin{aligned}\text{Log (Sfungi)} = {} & 0.102^{*}\text{Area} + 0.430^{*}\text{Seffort} - 0.240^{*}\text{Seffort}^2 \\ & + 0.045^{*}\text{Seffort}^3 + 0.160^{*}\text{Fcover} + 0.046^{*}\text{Pgymn} \\ & - 0.026^{*}\text{Pgymn}^2 - 0.028^{*}\text{Pgymn}^3 + 0.132^{*}\text{PCA1} \\ & + 0.012^{*}\text{PCA1}^2 + 0.159^{*}\text{Twinter} \\ & + 0.049^{*}\text{Twinter}^2 + 0.003^{*}\text{Twinter}^3 \\ & + 0.035^{*}\text{Tsummer} + 0.057^{*}\text{Tsummer}^2 \\ & - 0.083^{*}\text{Tsummer}^3 - 0.050^{*}\text{Pspring} + 3.250\end{aligned}$$

The partial effects of each variable on predicted parasitic fungal species richness are shown in Fig. 2.

Assessment of direct and indirect effects of selected variables on species richness

The eight explanatory variables retained in the final GLM were included in the structural equation model. In addition, we included the forest volume (Fvol) per unit area because this variable caused high change in deviance (*c.* 20%) in all three GLMs but was never selected (Table 2), and because there was some evidence in the literature that host body size may influence parasite species richness (see Discussion). Pearson's correlations between the nine explanatory variables are given in Table 3. The final structural model explained 74% of the variation in parasitic fungal species richness. It was accepted with both the standard chi-square statistic (ML estimator; $\chi^2 = 32.0$; d.f. = 26; $P = 0.19$) and the scaled chi-square statistic (MLM estimator; $\chi^2 = 28.9$; d.f. = 26; $P = 0.32$). Other closeness-of-fit indices confirmed that the model provided a very good fit to the data (RMSEA = 0.036; CFI = 0.99). The complete structure of the model is shown in Fig. 3, whereas the total, direct, and indirect effects of the explanatory variables on parasitic fungal species richness are given in Table 4.

DISCUSSION

Correlative approaches to habitat modelling have one major drawback: they cannot be used to establish a causal relationship between species distribution and the selected habitat descriptors (Mitchell, 2005; Kearney, 2006). Nevertheless, these approaches

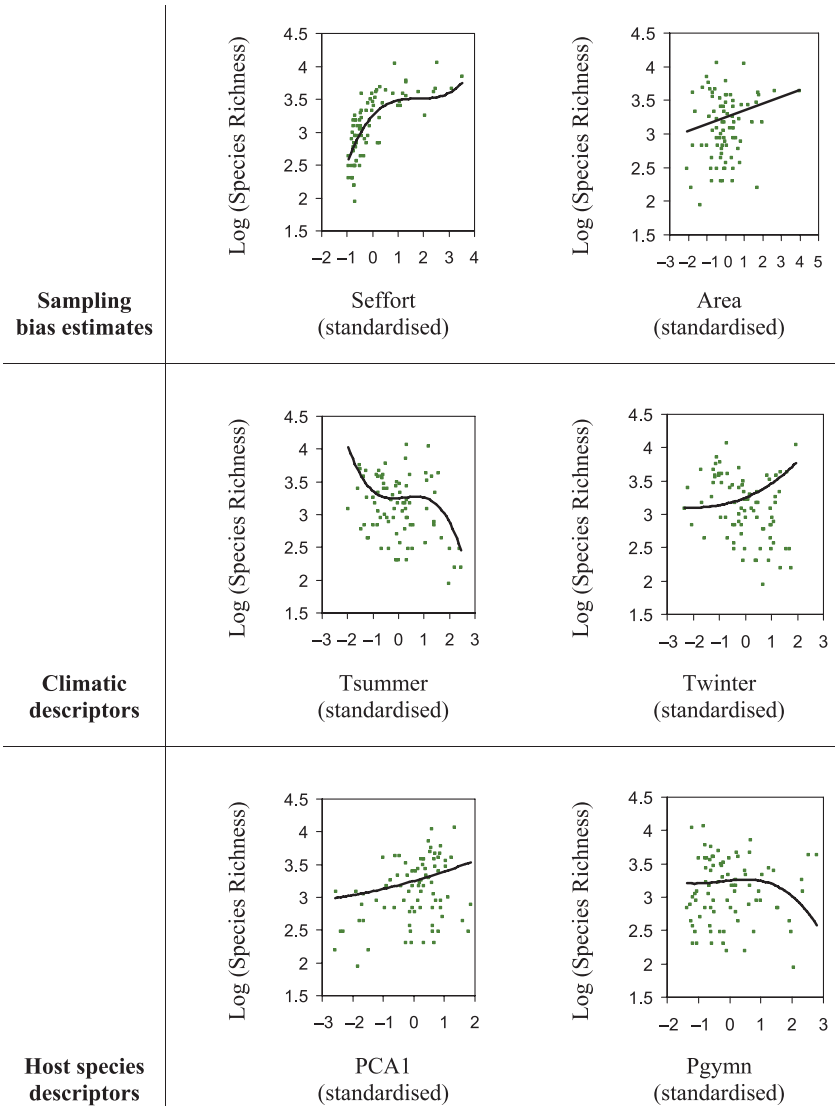

Figure 2 Partial effects of explanatory variables for generalized linear model B. The modelled relationships between explanatory variables (standardized) and parasitic fungal species richness (log-transformed) are shown with a line. Dots indicate the recorded values of parasitic fungal species richness (log-transformed) as a function of the explanatory variable (standardized). We show only variables causing a significant change in deviance when included in the model.

Table 4 Effects of explanatory variables on parasitic fungal species richness predicted by the structural equation model.

Variable	DE	IE	TE	RC
Climatic descriptors				
Twinter	0.203	–0.510	–0.307	28.5%
Tsummer	–0.187	0.263	0.076	41.5%
Pspring	–	0.174	0.174	–
Host species descriptors				
PCA1	0.232	–	0.232	–
Pgymn	–	0.169	0.169	–
Fcover	0.160	0.406	0.566	28.3%
Fvol	–	0.264	0.264	–
Sampling bias estimates				
Seffort	0.709	–	0.709	–
Area	0.239	–	0.239	–

DE, direct effect; IE, total indirect effect; TE, total effect; RC, relative contribution of the direct effect to the total effect. See Table 1 for variables names.

are a valuable guide, identifying habitat descriptors that may be important and which may be mechanistically linked to the presence or abundance of a species (Mitchell, 2005). In this study, we used a correlative approach, based on GLMs and stepwise analyses, to select the habitat descriptors that may account for the species richness distribution of parasitic fungi growing on trees. Then we used structural equation modelling to evaluate the direct and indirect effects of the selected habitat descriptors on parasitic fungal species richness.

Our results suggest that host species diversity may not be an important factor of the species richness of parasitic fungi growing on trees. Indeed, neither tree species richness (Stree) nor tree species relative diversity (Jtree) was selected as an explanatory variable. Only the selection of the relative area under gymnosperm species (Pgymn) was suggestive of a possible effect of tree species diversity on parasitic fungal species diversity (Table 2, Model B). Indeed, French forests with a large area under gymnosperm species are often dominated by a single tree species responsible for most of the annual wood production. The GLM indicates a slight tendency for these gymnosperm-dominated, monospecific

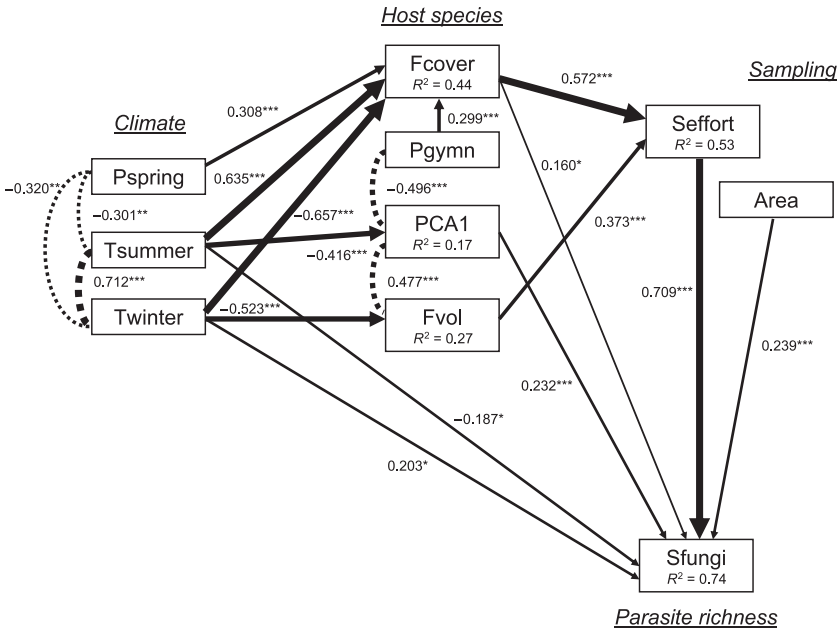

Figure 3 Structural equation model of the influence the selected explanatory variables on fungal species richness. Directed path are shown with straight simple-headed arrows, whereas undirected paths (free or unexplained correlations) are shown with dotted curves. Significant standardized path coefficients are shown for directed paths, with their level of statistical significance indicated by asterisk ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$). Arrow and curve widths are proportional to the standardized path coefficients. Parts of the variance (R^2) of the endogenous variables are given under the variables names. See Table 1 for variables names and description.

forests to contain fewer species of parasitic fungi (Fig. 2). However, the structural equation model did not confirm such a direct, negative effect of Pgymn on parasitic fungal species richness (Fig. 3). The coefficient for the total effect of Pgymn on fungal species was positive (Table 4), which suggests that the number of parasitic fungal species increases with the relative area under gymnosperm species (Table 4). Given these discrepancies, we cannot conclude from our results that host species diversity has an effect on parasitic fungal species richness.

Our results suggest that parasitic fungal species richness is rather dependent on the abundance of hosts and the composition of host species assemblages. Indeed, besides Pgymn, the selection procedure of explanatory variables retained two other host-related variables (Table 2): the variable Fcover, corresponding to the relative area under forest within each geographical unit, and the variable PCA1, corresponding to the distance to the Mediterranean zone in terms of tree species composition. The positive coefficient of Fcover in the GLM suggests that geographical units with only small patches of forest tend to have a lower number of parasitic fungal species. This positive, direct effect of Fcover on parasitic fungal species richness is confirmed by the structural equation model (Fig. 3). Such a negative effect of habitat fragmentation on species richness has already been reported in several studies (see Guégan *et al*., 2005 for review). The GLM also predicts a positive relationship between the variable PCA1 and parasitic fungal species richness (Fig. 2), which suggests that Mediterranean tree species assemblages do not sustain a great diversity of parasitic fungi. This direct, positive effect of PCA1 on parasitic fungal species richness is also confirmed by the structural equation model (Fig. 3). The most straightforward explanation for this relationship would be a higher resistance of Mediterranean tree species to fungal parasites. However, there are no published data to suggest such a trend. Alternatively, PCA1 may have been selected in the GLM because of its correlation with other habitat descriptors. In particular, PCA1 is highly correlated with forest wood volume (Fvol) per unit area (Table 3), and the low wood volume per unit area of the Mediterranean forests (involving low tree density or small tree size) may account for the small number of parasitic fungal species in this area. Indeed, it has been suggested that host features associated with the probability of colonization by new parasite species, such as body size, should covary with parasite species richness. It has also been suggested that host features increasing the success of parasite transmission among host individuals, such as high population density, should sustain a greater diversity of parasites. Both these theoretical predictions are supported by the results of comparative studies of parasite diversity in fish and mammalian host species (see Poulin, 2004 for review). However, the structural equation model does not confirm this hypothesis: the correlation between Fvol and PCA1 is retained in the model but the forest wood volume per unit area has no direct, significant effect on parasitic fungal species richness (Fig. 3). Hence, at this stage of the study, the cause for Mediterranean tree species assemblages to contain fewer parasitic fungal species remains uncertain.

Moreover, our results suggest that parasitic fungal species richness is directly affected by climate variables, including winter and summer mean temperatures (Twinter and Tsummer) in particular (Table 2, Model B). Both the GLM (Fig. 2) and the structural equation model (Fig. 3) predict that high winter temperatures and low summer temperatures (i.e. a narrow range of temperatures throughout the year) favour the diversity of fungal diseases in forests. These findings are consistent with the recent report by Petzoldt & Seaman (2006) on the impact of global warming, showing that the fungi responsible for plant diseases grow best in moderate temperature ranges. It is also consistent with studies showing that low winter temperatures, and frost in particular, are limiting for the survival of some parasitic species growing on trees (e.g. *Phytophthora cinnamomi* (Marçais *et al*.

2004)). The quasi-absence of water-related variables among the climate variables selected was unexpected, because moisture is an essential factor for the growth and reproduction of fungi (Agrios, 2005). Only one water-related variable, the mean monthly precipitation in the spring (Pspring), was retained in the GLM. However, it had no significant effect on deviance (Table 2). The structural equation model confirmed the absence of direct effect of Pspring on parasitic fungal species richness while it does report significant indirect effects of Pspring through its positive impact on the relative area under forest cover (Fig. 3).

However, none of the habitat descriptors described above was anywhere near as influential as sampling effort (Seffort). In the GLM, Seffort was positively associated with parasitic fungal species richness (Fig. 2) and accounted for a decrease of more than 60% in deviance (Table 2, Model B). As expected (see Guégan *et al.*, 2005 for review), we also observed a positive relationship between recorded species richness and the area sampled (Table 2 and Fig. 2). The structural equation model confirmed this high influence of sampling variables on the observed parasitic fungal species richness (Fig. 3). Our results thus confirm the importance of taking into account spatial variations in the sampled area and sampling effort when studying the distribution of species richness (Poulin, 1997; Guégan *et al.*, 2005).

Overall, the structural equation model suggests that four habitat descriptors may have a direct effect on parasitic fungal species richness (Fig. 3): the winter mean temperature (Twinter), the summer mean temperature (Tsummer), the tree species composition (PCA1), and the relative area under forest cover (Fcover). The direct effects of these four variables were roughly equal (their coefficients were close to 0.200 in absolute value (Table 4)). It is noteworthy that they were much lower than the direct effect of sampling effort (*c.* 0.700 (Table 4)). The structural equation model also indicates that all the four habitat descriptors, except PCA1, may have indirect effects on parasitic fungal species richness through their influence on other variables. These indirect effects were important: they represented 60–70% of the variables contribution to parasitic fungal species richness (Table 4). The indirect effects of Twinter and Tsummer, through their influence on host abundance (Fcover and Fvol) and host species composition (PCA1), counter-balanced their direct effects on parasitic fungal species richness (Table 4). In contrast, the indirect effect of Fcover via its effect on Seffort strongly reinforced its positive influence on parasitic fungal species richness (Table 4). The second descriptor of host species abundance (Fvol) also had a positive influence on parasitic fungal species richness via its influence on sampling effort. These results suggest that geographical units with only small patches of forest, or with low wood volume, were monitored less thoroughly and consequently had a lower observed parasitic fungal species richness.

These spatial variations in sampling effort are obviously a major limitation to our study of the ecological determinants of parasitic fungal species richness. We indeed used data of parasitic fungal species occurrence which were initially collected for epidemiological rather than ecological purposes. These data were obtained through spontaneous reporting, as it is often the case in medical epidemiology (Goldman, 1998; Edwards, 1999). Two major advantages of spontaneous reporting are its power to detect rare occurrences and its possible use at large scales (e.g. national or international level). However, its reliability depends on the active and equal participation of reporters. A statistically designed data collection, involving a regular monitoring of forest patches of equivalent size evenly distributed across France, may have been more relevant to study the ecological determinants of parasitic fungal species richness (e.g. Kuffer & Senn-Irlet, 2005 for saprophytic fungal species richness in Switzerland). The survey of forest pathogens which has been initiated recently at the European scale, based on a uniform 16×16 km grid of 4800 plots of 20 trees each (Lorenz, 1995; UN/ECE, 2006), will provide such data. However, studies of parasitic fungal species richness based on these data will also have some limitations. In particular, because of the use of a wide-mesh sampling grid, rare fungal parasites or fungal parasites inducing disease foci (e.g. *Armillaria ostoyae*) may pass unnoticed. Therefore we believe that there is no perfect design to identify the determinants of the species richness of parasitic fungi growing on trees. Only the convergence of several studies will give a solid picture of these determinants.

In conclusion, our results suggested that the richness of parasitic fungal species in France is directly affected by temperature seasonality, host abundance, and the composition of host species assemblages. They also suggested that host species diversity may not be a factor of importance for parasitic fungal species richness. The structural equation model revealed that the direct effects of climatic descriptors and host species descriptors were roughly equal, and that they were three times lower than the direct effects of sampling effort. The model also indicated that the direct effects of climatic and host species descriptors constituted only 30% to 40% of their total effect on parasitic fungal species richness. Our study therefore demonstrates the importance of taking into account the indirect effects of climatic variables, via host-related variables, and the indirect effects of host-related variables, via sampling effort, when investigating the potential causes of spatial variation in parasite richness. Our study also shows the importance of considering abiotic factors when investigating the potential causes of variation in parasite richness, especially for plant parasites. In the case of fungal parasites of trees, our study suggested that winter and summer temperatures were important determinants of parasite richness. This result surprised us because we expected water-related variables to be more influential than temperatures for fungal species. We will investigate the relative impacts of water-related variables and temperatures in a next study, by distinguishing aerial and soil species, in particular.

ACKNOWLEDGEMENTS

We thank the Département Santé des Forêts (DSF) for having allowed us to use their data base, and all members of the Forest Pathology team (INRA, UMR BioGeCo) and of the Groupe Français de Pathologie Forestière (GFPF) for helpful discussions and suggestions. We also thank three anonymous reviewers and the Associate Editor for helpful comments.

REFERENCES

Agrios, G.N. (2005) *Plant pathology*, 5th edn. Elsevier Academic Press, Amsterdam, The Netherlands.

Benichou, P. & Le Breton, O. (1987) Prise en compte de la topographie pour la cartographie de champs pluviométriques statistiques. *La Météorologie*, **19**, 23–34.

Cabrero-Sanudo, F.J. & Lobo, J.M. (2006) Determinant variables of Iberian peninsula Aphodiinae diversity (Coleoptera, Scarabaeoidea, Aphodiidae). *Journal of Biogeography*, **33**, 1021–1043.

Calvete, C. (2003) Correlates of helminth community in the red-legged partridge (*Alectoris rufa* L.) in Spain. *Journal of Parasitology*, **89**, 445–451.

Chessel, D., Dufour, A.-N. & Dray, S. (with contributions from Jean R. Lobry, Sebastien Ollier, Sandrine Pavoine and Jean Thioulouse) (2005) ade4: analysis of environmental data: exploratory and Euclidean methods in environmental sciences. *R package version 1.4-0*. URL http://www.R-project.org.

Cocu, N., Harrington, R., Rounsevell, M.D.A., Worner, S.P. & Hulle, M. (2005) Geographical location, climate and land use influences on the phenology and numbers of the aphid, *Myzus persicae*, in Europe. *Journal of Biogeography*, **32**, 615–632.

Crawley, M.J. (1993) *GLIM for ecologists*, Blackwell Science, Cambridge, MA.

Cumming, G.S. & Guegan, J.F. (2006) Food webs and disease: Is pathogen diversity limited by vector diversity? *Ecohealth*, **3**, 163–170.

Edwards, I.R. (1999) Spontaneous reporting – of what? Clinical concerns about drugs. *British Journal of Clinical Pharmacology*, **48**, 138–141.

Everitt, B.S. & Hothorn, T. (2006) *A Handbook of Statistical Analyses using R*. Chapman & Hall/CRC, Boca Raton, Florida.

Gebre-Michael, T., Malone, J.B., Balkew, M., Ali, A., Berhe, N., Hailu, A. & Herzi, A.A. (2004) Mapping the potential distribution of *Phlebotomus martini* and *P. orientalis* (Diptera: Psychodidae), vectors of kala-azar in East Africa by use of geographic information systems. *Acta Tropica*, **90**, 73–86.

Goldman, S.A. (1998) Limitations and strengths of spontaneous reports data. *Clinical Therapeutics*, **20**, C40–C44.

Guégan, J.-F. & Kennedy, C.R. (1996) Parasitic richness/Sampling effort/Host range: the fancy three-piece jigsaw puzzle. *Parasitology Today*, **12**, 367–369.

Guégan, J.-F., Morand, S. & Poulin, R. (2005) Are there general laws in parasite community ecology? The emergence of spatial parasitology and epidemiology. *Parasitim and ecosystems* (ed. by F. Thomas, F. Renaud and R. Poulin), p. 221. Oxford University Press, Oxford, UK.

Guernier, V., Hochberg, M.E. & Guegan, J.F.O. (2004) Ecology drives the worldwide distribution of human diseases. *Plos Biology*, **2**, 740–746.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Heikkinen, R.K. & Neuvonen, S. (1997) Species richness of vascular plants in the subarctic landscape of northern Finland: modelling relationships to the environment. *Biodiversity and Conservation*, **6**, 1181–1201.

Hilborn, R. & Mangel, M. (1997) *The ecological detective: confronting models with data*. Princeton University Press, Princeton, New Jersey.

Kearney, M. (2006) Habitat, environment and niche: what are we modelling? *Oikos*, **115**, 186–191.

Kearney, M., Moussalli, A., Strasburg, J., Lindenmayer, D. & Moritz, C. (2003) Geographic parthenogenesis in the Australian arid zone: I. A climatic analysis of the *Heteronotia binoei* complex (Gekkonidae). *Evolutionary Ecology Research*, **5**, 953–976.

Krasnov, B.R., Stanko, M., Miklisova, D. & Morand, S. (2006) Habitat variation in species composition of flea assemblages on small mammals in central Europe. *Ecological Research*, **21**, 460–469.

Kuffer, N. & Senn-Irlet, B. (2005) Influence of forest management on the species richness and composition of wood-inhabiting basidiomycetes in Swiss forests. *Biodiversity and Conservation*, **14**, 2419–2435.

Kuhn, K.G., Campbell-Lendrum, D.H. & Davies, C.R. (2002) A continental risk map for malaria mosquito (Diptera: Culicidae) vectors in Europe. *Journal of Medical Entomology*, **39**, 621–630.

Lobo, J.M., Jay-Robert, P. & Lumaret, J.P. (2004) Modelling the species richness distribution for French Aphodiidae (Coleoptera, Scarabaeoidea). *Ecography*, **27**, 145–156.

Lorenz, M. (1995) International co-operative programme on assessment and monitoring of air pollution effects on forests – ICP forests. *Water Air and Soil Pollution*, **85**, 1221–1226.

Maechler, M., Rousseeuw, P., Struyf, A. & Hubert, M. (2005) *Cluster: analysis basics and extensions. R package*. URL http://www.R-project.org.

Marçais, B., Bergot, M., Pérarnaud, V., Levy, A. & Desprez-Loustau, M.-L. (2004) Prediction and mapping of the impact of winter temperature on the development of *Phytophthora cinnamomi*-induced cankers on red and pedunculate oak in France. *Phytopathology*, **94**, 826–831.

McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*. Chapman & Hall/CRC, New York.

Menendez, R., Gonzales-Megias, A., Collingham, Y., Fox, R., Roy, D.B., Ohlemüller, R. & Thomas, C.D. (2007) Direct and indirect effects of climate and habitat factors on butterfly diversity. *Ecology*, **88**, 605–611.

Meyer, D., Zeileis, A. & Hornik, K. (2006) *vcd: visualizing categorical data. R package* version 1.0-2. URL http://www.R-project.org.

Mitchell, S.C. (2005) How useful is the concept of habitat? A critique. *Oikos*, **110**, 634–638.

Montgomery, D.C. & Peck, E.A. (1982) *Introduction to linear regression analysis*. John Wiley and Sons, New York.

Moser, D., Dullinger, S., Englisch, T., Niklfeld, H., Plutzar, C., Sauberer, N., Zechmeister, H.G. & Grabherr, G. (2005) Environmental determinants of vascular plant species richness in the Austrian Alps. *Journal of Biogeography*, **32**, 1117–1127.

Muthén, L.K. & Muthén, B. (1998) *Mplus user's guide*. Mplus, Los Angeles, California.

Muthén, L.K. & Muthén, B. (2003) *Mplus user's guide*. Mplus, Version 2.14. Los Angeles, California.

Neter, J., Wasserman, W. & Kutner, M.H. (1985) *Applied linear statistical models*, 2nd edn. Richard. D. Irwin, Homewood, Illinois.

Nicholls, A.O. (1989) How to make biological surveys go further with generalised linear models. *Biological Conservation*, **50**, 51–75.

Nunn, C.L., Altizer, S.M., Sechrest, W. & Cunningham, A.A. (2005) Latitudinal gradients of parasite species richness in primates. *Diversity and Distributions*, **11**, 249–256.

O'Brien, E.M. (1998) Water-energy dynamics, climate, and prediction of woody plant species richness: an interim general model. *Journal of Biogeography*, **25**, 379–398.

Osborne, P.E., Alonso, J.C. & Bryant, R.G. (2001) Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *Journal of Applied Ecology*, **38**, 458–471.

Petzoldt, C. & Seaman, A. (2006) Climate change effects on insects and pathogens. Climate change and agriculture: promoting practical and profitable responses, SARE project ENE05-091. Cooperative State Research, Education and Extension Service, US Department of Agriculture, Washington, DC.

Poulin, R. (1997) Species richness of parasite assemblages: evolution and patterns. *Annual Review of Ecology and Systematics*, **28**, 341–358.

Poulin, R. (2004) Macroecological patterns of species richness in parasite assemblages. *Basic and Applied Ecology*, **5**, 423–434.

Pugesek, B.H., Tomer, A. & von Eye, A. (2003) *Structural equation modeling: applications in ecological and evolutionary biology*. Cambridge University Press, Cambridge, UK.

R Development Core Team (2006) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Real, R., Barbosa, A.M., Porras, D., Kin, M.S., Marquez, A.L., Guerrero, J.C., Palomo, L.J., Justo, E.R. & Vargas, J.M. (2003) Relative importance of environment, human activity and spatial situation in determining the distribution of terrestrial mammal diversity in Argentina. *Journal of Biogeography*, **30**, 939–947.

Shipley, B. (2000) *Cause and correlation in biology: a user's guide to path analysis, structural equations, and causal inference*. Cambridge University Press, Cambridge, UK.

Stevens, M.H.H. & Carson, W.P. (1999) The significance of assemblage-level thinning for species richness. *Journal of Ecology*, **87**, 490–502.

Turc, L. (1963) Evaluation des besoins en eau d'irrigation, évapotranspiration potentielle, formulation simplifié et mise à jour. *Annales Agronomiques*, **12**, 13–49.

United Nations Economic Commission for Europe (UNECE) (2006) International Cooperative Programme on Assessment and Monitoring of Air Pollution Effects on Forests. Part II: Visual assessment of crown condition. URL http://www.icp-forests.org/pdf/Chapt2_compl06.pdf.

Vincent, P.J. & Haworth, J.M. (1983) Poisson regression models of species abundance. *Journal of Biogeography*, **10**, 153–160.

Wohlgemuth, T. (1998) Modelling floristic species richness on a regional scale: a case study in Switzerland. *Biodiversity and Conservation*, **7**, 159–177.

Editor: Brendan Wintle

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Appendix S1 List of parasitic fungal species.

Appendix S2 List of forest tree taxa.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1472-4642.2008.00479.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.