HAL open science

Modélisation de la niche écologique des fétuques à feuilles fines : quels apports pour la conservation et la valorisation des ressources génétiques ?

Jean Paul J. P. Sampoux, Vincent V. Badeau

To cite this version:

Jean Paul J. P. Sampoux, Vincent V. Badeau. Modélisation de la niche écologique des fétuques à feuilles fines : quels apports pour la conservation et la valorisation des ressources génétiques ?. Innovations Agronomiques, 2009, 7, pp.79-91. 10.17180/agda-j231 . hal-02664746

HAL Id: hal-02664746

https://hal.inrae.fr/hal-02664746v1

Submitted on 31 May 2020

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.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Innovations Agronomiques (2009) 7, 79-91

Modélisation de la niche écologique des fétuques à feuilles fines : quels apports pour la conservation et la valorisation des ressources génétiques ?

J.P. Sampoux[1], V. Badeau[2]

[1] : INRA, Centre Poitou-Charentes, Unité de Recherche Pluridisciplinaire Prairies et Plantes Fourragères, F 86600 Lusignan

[2] : INRA, Centre de Nancy, UMR Ecologie et Ecophysiologie Forestières, F-54280 Champenoux

Correspondance : jpsampoux@lusignan.inra.fr

Pour échantillonner au mieux la diversité spontanée des espèces fourragères et à gazon, il est nécessaire d'avoir une certaine connaissance de la gamme de conditions environnementales convenant à ces espèces, ainsi que de la structuration environnementale de leur diversité intra-spécifique. Nous donnons dans ce document des principes d'analyse statistique permettant d'étudier le déterminisme environnemental de la présence des espèces et leur cadre d'application, et nous montrons en quoi les résultats issus de telles études pourraient contribuer à faciliter le travail de prospection de diversité spontanée et à améliorer la gestion des collections d'accessions.

Résumé :

La modélisation statistique de la niche écologique d'une espèce consiste à construire une fonction de paramètres environnementaux qui prédit la probabilité de présence de l'espèce à partir d'un jeu de données de calibration comprenant des données de présence/absence ou d'abondance de l'espèce et des valeurs de paramètres environnementaux aux sites d'observation. A partir de l'exemple d'une collection de fétuques à feuilles fines, nous montrons que les données de référencement spatial et les données environnementales rassemblées à l'occasion des campagnes de prospection de ressources génétiques peuvent constituer une information pertinente pour la modélisation de la niche écologique des espèces. Nous présentons deux méthodes qui permettent l'étude de la distribution environnementale des espèces et la modélisation de leur niche écologique à partir d'observations de présence/absence : l'analyse canonique des corrélations et la régression logistique. Nous suggérons des représentations spatiales, obtenues à l'aide de logiciel SIG, des patrons de distribution environnementale de la diversité prédits par les modèles statistiques qui devraient permettre de mieux piloter les campagnes de prospection de diversité spontanée. Nous indiquons également comment les modèles de niche prédisant la probabilité de présence des espèces pourraient contribuer à optimiser le choix de core-collections. Les méthodes présentées dans ce document pourraient offrir des avantages identiques pour d'autres espèces d'intérêt agronomique disposant d'un compartiment de diversité naturelle, ainsi que pour des espèces sauvages apparentées à certaines espèces domestiques

Mots-clés : Analyse canonique des corrélations, core-collection, distribution environnementale, diversité intra-spécifique, diversité spontanée, fétuques, modélisation statistique, niche écologique, prospection, régression logistique, ressources génétiques

Abstract: Modelling ecological niche of fine-leaved fescues

Rational sampling of spontaneous diversity of forage and turf species requires an *a priori* knowledge of the range of environmental conditions suitable for these species. We introduce some concepts and methods for investigating the environmental range of species by statistical modelling of species ecological niche, and we suggest how such investigations could help in planning collection campaigns and in improving choice of core-collections. Statistical modelling of ecological niche of a species

consists in building a function of environmental parameters predicting presence of the species from a calibration dataset including observed presence/absence or abundance records of the species and environmental data at observation sites. From the example of a collection of fine-leaved fescue taxa, we showed that spatial references and environmental data recorded in genetic resource collection campaigns can be relevant information for species niche modelling. Two methods for investigating the environmental distribution of species and for niche modelling based on presence/absence data are presented: the canonical correlation analysis and the logistic regression. Examples combining niche model and GIS software are given that may contribute to organize collection campaigns. We suggest that models predicting the probability of presence of species may be useful for the selection of core-collections. Such models may help in delineating geographically isolated areas of presence of species that should be sampled separately for selecting a core-collection. In each isolated area of presence, it is proposed to stratify accessions in clusters according to the predicted probability of presence of the species in collection sites, and to select accessions in each cluster. Methods proposed in this paper may be useful to improve collection and conservation of spontaneous diversity of many crop species and of wild species related to crop species.

Keywords: Canonical correlation analysis, core-collection, environmental distribution, ecological niche, fescues, genetic resource collection, intra-specific diversity, logistic regression, spontaneous diversity, statistical modelling

Introduction

La plupart des espèces végétales faisant l'objet d'un travail d'amélioration pour un usage en fourrage ou en gazon sont présentes de façon spontanée dans une grande variété de conditions environnementales. Au sein de chacune de ces espèces, la sélection naturelle a conduit à la différenciation d'une diversité d'écotypes permettant l'adaptation à une gamme plus ou moins large de conditions environnementales. C'est dans cette diversité spontanée qu'a été puisé le matériel génétique à l'origine des premiers travaux d'amélioration de ces espèces voici une cinquantaine d'années. Depuis lors, la diversité spontanée a contribué de façon récurrente à enrichir la diversité travaillée par les sélectionneurs, et elle constitue encore aujourd'hui une ressource indispensable pour faire face à de nouveaux enjeux agronomiques et environnementaux. Pour se procurer des sources originales de diversité spontanée, les sélectionneurs d'espèces fourragères et à gazon ont procédé, et procèdent encore régulièrement, à des campagnes de prospection couvrant des zones géographiques plus ou moins vastes. Les populations ainsi échantillonnées sont conservées sous forme de lots de graines (ou accessions) regroupés en collections et stockés en chambre froide. A partir de ces lots de graines, des analyses de diversité phénotypique ou moléculaire peuvent être entreprises et les accessions les plus prometteuses peuvent être introduites dans les programmes de sélection. Pour organiser au mieux les campagnes de prospection, il est nécessaire d'avoir une certaine connaissance de la gamme de conditions environnementales convenant aux espèces recherchées, ainsi que de la structuration environnementale de leur diversité intra-spécifique. De telles connaissances étaient évidemment sommaires lors des premières campagnes de collecte de diversité spontanée. Toutefois, l'analyse des informations aujourd'hui disponibles grâce aux collectes déjà réalisées devrait permettre de mieux comprendre et prédire la distribution environnementale des espèces fourragères et à gazon. Certains travaux ont en effet permis de mieux appréhender la distribution spatiale et environnementale de caractères phénotypiques et de marqueurs moléculaires au niveau intra-spécifique pour quelques espèces (Balfourier et Charmet, 1991 ; Warren et al., 1998 ; Prospéri et al., 2005). En revanche, la question de la détermination des conditions environnementales favorables à la présence spontanée des différentes espèces n'avait pas été réellement prise en considération jusqu'à aujourd'hui par la communauté des améliorateurs. Nous donnons dans ce document des principes d'analyse statistique

permettant d'étudier le déterminisme environnemental de la présence des espèces et leur cadre d'application, et nous montrons en quoi les résultats issus de telles études pourraient contribuer à faciliter le travail de prospection de diversité spontanée et à améliorer la gestion des collections d'accessions. Pour cela, nous nous appuyons sur l'exemple d'une collection de fétuques à feuilles fines rassemblée à l'Unité de Recherche Pluri-disciplinaire Prairies et Plantes Fourragères de l'INRA à Lusignan à partir de laquelle nous avons tenté de déterminer les conditions environnementales favorables à la présence de différents taxons à l'aide d'une modélisation statistique de leur niche écologique.

Cadre conceptuel et méthodes

Le concept de niche écologique

Toute espèce vivante ne peut accomplir son cycle biologique et sa reproduction qu'à l'intérieur d'une plage limitée de variations environnementales d'origine abiotique (facteurs physiques et chimiques) et biotique (interactions entre espèces). Grinell (1917, 1924) a proposé le concept de *niche écologique* pour désigner la plage de variations environnementales compatible avec la présence d'une espèce. Hutchinson (1957) a précisé ce concept en définissant la niche écologique d'une espèce comme un hyper-volume multidimensionnel dans lequel chaque dimension représente un paramètre biotique ou abiotique conditionnant la présence de l'espèce. Plus précisément, Hutchinson distinguait la *niche fondamentale* correspondant à l'ensemble de la gamme de variation des paramètres environnementaux qui autorise la présence de l'espèce en l'absence de compétition ou de prédation causées par d'autres espèces, et la *niche réalisée* restreinte à la partie de la niche fondamentale dans laquelle l'espèce est effectivement présente.

La modélisation statistique (ou empirique) de la niche écologique d'une espèce consiste à construire une fonction de paramètres environnementaux qui prédit la probabilité de présence de l'espèce à partir d'un jeu de données de calibration comprenant des données de présence/absence ou d'abondance de l'espèce et des valeurs de paramètres environnementaux aux sites d'observation. Cette modélisation statistique étant basée sur des observations effectives de présence de l'espèce, elle a été considérée comme décrivant sa niche réalisée. En fait, de tels modèles prédisent plutôt des combinaisons de facteurs environnementaux qui sont attendues favorables à la présence de l'espèce ; il serait donc plus approprié de considérer qu'ils prédisent des *habitats potentiels* (Araujo et Guisan, 2006). Il faut de plus souligner que cette modélisation empirique fait l'hypothèse que les espèces étudiées sont en équilibre avec leur environnement, c'est-à-dire qu'elles ne sont plus dans une phase de colonisation de leur domaine environnemental potentiel. Cette hypothèse ne peut pas toujours être tenue pour acceptable ; c'est le cas par exemple de certaines espèces invasives récemment introduites.

Quelles données pour modéliser la niche écologique d'une espèce ?

Pour modéliser la niche écologique d'une espèce, il est nécessaire de disposer d'un jeu de données de calibration associant des données d'occurrence de l'espèce à des valeurs de paramètres environnementaux sur un certain nombre de sites d'observation. Les données d'occurrence de l'espèce peuvent être des données de présence/absence ou des données d'abondance. Dans le domaine des sciences écologiques, de telles données sont souvent obtenues en pratiquant des campagnes de relevé d'espèces sur des zones d'intérêt. Dans le cas des espèces végétales, des données d'herbier sont parfois utilisées pour compléter des données de relevé, voire pour se substituer à celles-ci. Notons ici que le référencement spatial des sites de collecte effectué au cours des campagnes de collecte de ressources génétiques spontanées par les améliorateurs d'espèces végétales est aussi susceptible de fournir des données d'occurrence.

Austin (2007) classe en différentes catégories les paramètres environnementaux physiques et bio-physiques susceptibles de conditionner la présence d'une espèce. Il établit une distinction entre paramètres indirects (comme l'altitude ou la latitude), paramètres directs (comme la température, le rayonnement solaire, les précipitations) et paramètres de ressource (comme la disponibilité en azote pour les plantes). Il distingue également des paramètres distaux (comme les paramètres climatiques) et des paramètres proximaux (c'est-à-dire impliqués dans des processus éco-physiologiques au plus proche de l'interface entre les tissus vivants et leur environnement). Il constate que la modélisation de la niche écologique est en général d'autant plus efficace que les paramètres environnementaux utilisés sont plus directs et plus proximaux. D'autres types de données environnementales peuvent être informatifs, par exemple des données édaphiques ou des données caractérisant la couverture du sol (Thuiller et al., 2004), ou encore des informations sur la présence d'autres espèces (compétiteurs, facilitateurs ou prédateurs de l'espèce à modéliser). Un certain nombre de paramètres environnementaux utiles peut être noté sur le terrain pendant les campagnes de relevé ou de collecte de diversité ; dans le cas des collections de ressources génétiques spontanées d'espèces végétales, les informations environnementales recueillies sur site ainsi que les coordonnées géographiques des lieux de collecte forment ce que l'on appelle les "données passeport". Il faut aussi noter qu'un nombre croissant de bases de données environnementales géo-référencées offre la possibilité d'extraire des données environnementales pertinentes. Des données climatiques interpolées en résolution fine sont ainsi disponibles pour de vastes zones géographiques, par exemple dans la base de données Worldclim (http://www.worldclim.org) qui couvre une grande partie du monde en résolution kilométrique, ou encore dans la base Aurelhy de Météo-France couvrant le territoire français en résolution kilométrique à partir d'une grande densité de points d'observation. Des images satellites peuvent aussi être traitées pour obtenir différents types de données comme le rayonnement solaire global ou des paramètres de couverture du sol. Des données de couverture du sol sont par exemple disponibles pour l'Europe dans la base de données CORINE (http://www.eea.europa.eu/themes/landuse/clc-download).

Il est important de bien choisir l'échelle spatiale du jeu de données de calibration, car elle est déterminante pour la pertinence et la qualité des modèles de niche écologique (Guisan et Thuiller, 2005). La résolution du réseau de sites d'observation doit être suffisamment fine pour capturer la variation des paramètres environnementaux qui déterminent la présence de l'espèce. D'autre part, l'étendue de la zone d'observation doit être adéquate. Si une espèce n'a été observée que sur une partie limitée de sa plage de présence pour un paramètre environnemental important, le modèle obtenu ne sera valide que dans l'intervalle de variation du paramètre environnemental qui aura été exploré.

Les modèles

Différentes méthodes ont été proposées pour la modélisation statistique de la niche écologique des espèces. Nous présenterons brièvement deux méthodes qui peuvent être utilisées dans le cas de données de présence/absence. Supposons que la présence ou l'absence de plusieurs espèces ait été notée dans un certain nombre de sites d'observation. Le jeu de données de calibration peut être organisé en deux tables. Une première table liste les observations de présence/absence des espèces par site sous forme binaire (1 lorsqu'une espèce est présente dans un site, 0 dans le cas contraire). La seconde table contient les valeurs des paramètres environnementaux aux sites d'observation.

Une méthode multivariable : l'analyse canonique des corrélations

A partir des deux tables ainsi constituées, il est possible de réaliser une analyse canonique des corrélations (Hotelling, 1936). Cette méthode recherche des paires successives de combinaisons linéaires de variables appartenant à deux groupes distincts dont la corrélation est maximale, les combinaisons linéaires (ou variables canoniques) successives de chaque groupe de variables étant non

corrélées. Dans le cas des deux tables décrites précédemment, les deux groupes de variables sont les indices de présence/absence des différentes espèces d'une part, et les paramètres environnementaux d'autre part. Notons que les conditions d'utilisation de l'analyse canonique des corrélations requièrent que le nombre de sites d'observation soit nettement supérieur au nombre d'espèces et au nombre de paramètres environnementaux. Les variables canoniques associées aux paramètres environnementaux peuvent alors être considérées comme les gradients environnementaux qui séparent le mieux les niches des espèces (Gimaret-Carpentier et al., 2003), et l'amplitude des niches des espèces peut être appréciée par l'écart-type de leurs scores sur chaque axe canonique (Carnes et Slades, 1982). Le degré d'adéquation des conditions environnementales d'un site quelconque avec la présence d'une espèce peut être apprécié par la distance de ce site au barycentre de l'espèce dans l'espace des premiers axes canoniques des paramètres environnementaux. Il n'est toutefois pas possible d'établir réellement un modèle de prédiction de la probabilité de présence d'une espèce avec l'analyse canonique des corrélations. Cette méthode a essentiellement un intérêt descriptif ; elle trouve par exemple son utilité pour identifier les déterminants environnementaux les plus importants lorsque de nombreux paramètres environnementaux sont disponibles.

La régression logistique

La régression logistique est un type de modèle statistique qui appartient à la famille des Modèles Linéaires Généralisés. Pour les modèles de cette famille, l'espérance de la variable à prédire est modélisée par une fonction (dite de lien) appliquée à une combinaison linéaire de prédicteurs. Différentes lois de distribution de la variable à prédire sont possibles, ainsi que différentes fonctions de lien (McCullagh and Nelder, 1997). La régression logistique, qui suppose une loi de distribution binomiale de la variable à prédire et une fonction de lien logistique, convient pour la modélisation d'événements de présence/absence d'une espèce ; dans ce cas, la variable à prédire est l'indice binaire décrivant la présence ou l'absence de l'espèce et les prédicteurs sont les valeurs des paramètres environnementaux. La régression logistique est alors appliquée de façon à prédire la probabilité de présence de l'espèce. Comme dans une régression linéaire classique, la colinéarité entre les prédicteurs et la surparamétrisation des modèles doivent être évitées. Une approche multivariable préalable, comme l'analyse canonique des corrélations, peut aider à déterminer un nombre limité de prédicteurs non corrélés. Un nombre minimum d'observations et d'événements de présence est également nécessaire pour obtenir un modèle fiable (Pearce et Ferrier, 2000a ; Coudun et Gégout, 2006). Remarquons que de réelles observations d'absence de l'espèce sont en principe nécessaires ; il est toutefois possible d'utiliser des techniques alternatives qui génèrent des observations de pseudo-absence (Zaniewski et al., 2002). La sélection des prédicteurs entrant dans le modèle peut être réalisée par une sélection stepwise classique. Cependant, des méthodes plus efficaces et plus fiables ont été développées (Wintle et al., 2003 ; Johnson et Omland, 2004). La valeur prédictive du modèle doit finalement être appréciée par deux critères, sa calibration et son pouvoir de discrimination. La calibration du modèle (ou qualité de l'ajustement) peut être évaluée par exemple à l'aide du coefficient de détermination généralisé de Nagelkerke (1991). Le pouvoir discriminant du modèle est sa capacité à discriminer les événements de présence et d'absence de l'espèce. Il peut être évalué en comparant les événements prédits aux événements observés à l'aide du jeu de données de calibration. Il est toutefois préférable d'utiliser des jeux de données différents pour la calibration du modèle et pour sa validation.

Un exemple : la modélisation de la niche écologique des taxons de fétuque à feuilles fines

Les fétuques à feuilles fines forment un groupe monophylétique (genre *Festuca* sous-genre *Festuca*) composé de nombreux taxons classés en différentes sections et sous-sections, parmi lesquelles la section *Aulaxyper* (fétuques rouges) et la section *Festuca* sous-section *Festuca* (fétuques ovines) sont

les deux entités les plus importantes en nombre d'espèces et en fréquence. Ces espèces sont utilisées pour le semis de gazon et font l'objet d'une activité de création variétale. Une prospection de fétuques à feuilles fines a été réalisée par l'ACVF (Association des Créateurs de Variétés Fourragères) et l'INRA en zone de plaine et de moyenne montagne sur une grande partie du territoire français et a permis de rassembler 382 populations, appartenant à cinq taxons de fétuques rouges et trois taxons de fétuques ovines (Figure 1). Nous avons utilisé les informations recueillies à l'occasion de cette prospection pour étudier la niche écologique des taxons échantillonnés. La liste des sites de collecte des taxons a permis de constituer une table de présence/absence des taxons ; les prospecteurs avaient en effet pour objectif de collecter tous les taxons présents sur chaque site, mais il s'est avéré que très peu de sites hébergeaient plus d'un taxon. Les sites d'observation ont été caractérisés par une série de paramètres environnementaux afin de constituer la table des données environnementales. Des données mensuelles (normales trentenaires) de température et de précipitation ont été extraites de la base de données Aurelhy de Météo-France au pas kilométrique ; nous avons également utilisé des données de rayonnement solaire global mensuel calculées par Badeau et al. (2009) à partir de données d'image satellite fournies par le SATMOS (Service d'Archivage et de Traitement Météorologique des Observations Spatiales). Enfin, des données de texture du sol et d'utilisation des sites ont été puisées dans les informations recueillies sur les sites de collecte par les prospecteurs (données passeport).

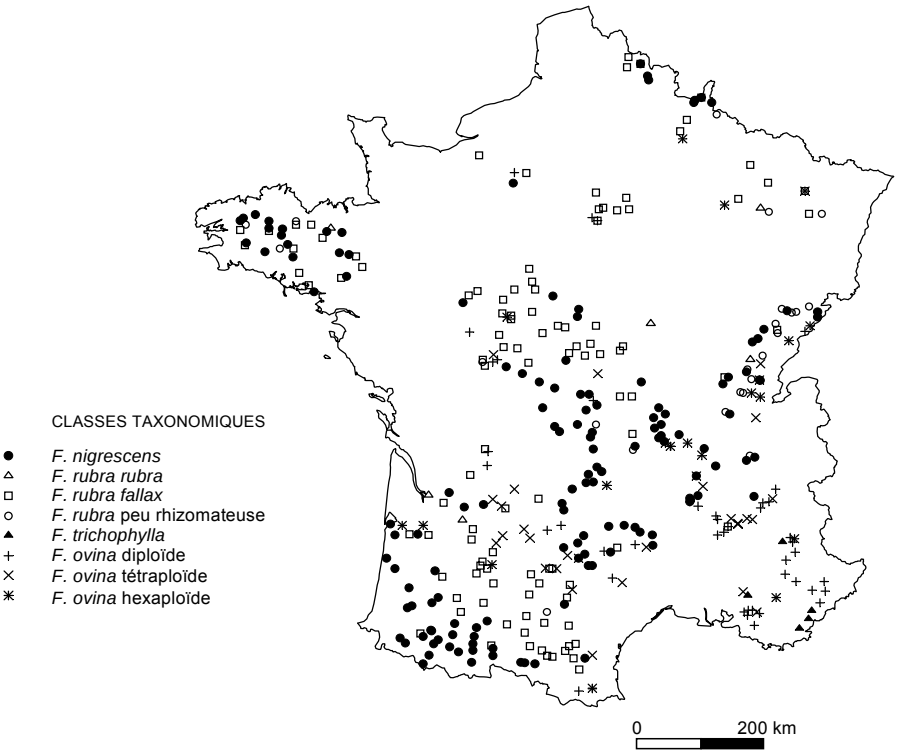

Figure 1. Distribution géographique sur le territoire français des 382 sites de collecte de populations de fétuques à feuilles fines et des huit classes taxonomiques identifiées dans la collection.

La Figure 2 illustre les résultats d'une analyse canonique des corrélations des données d'occurrence des espèces et des paramètres climatiques réalisée à partir de ces informations. Cette analyse met en évidence deux gradients climatiques principaux qui séparent le mieux les niches climatiques des taxons, un gradient de rayonnement solaire global estival et un gradient de précipitations de printemps – automne, associés à la première et à la deuxième directions canoniques, respectivement.

Une modélisation par régression logistique a ensuite été effectuée pour les taxons collectés sur plus de 100 sites en utilisant comme prédicteurs environnementaux potentiels le rayonnement solaire global estival, les précipitations cumulées de printemps et d'automne, la texture du sol et l'utilisation des sites.

La Figure 3 présente les courbes de réponse marginale obtenues pour la probabilité de présence de deux taxons en fonction du rayonnement global estival et selon différentes textures de sol. La courbe de réponse de la probabilité de présence d'une espèce le long d'un gradient environnemental est attendue unimodale et doit tendre vers zéro aux frontières de la plage de présence de l'espèce. Remarquons ici que le territoire français offre toute la plage de rayonnement global estival favorable à la présence de *Festuca rubra* subsp. *fallax*, mais seulement une partie de celle convenant à la présence de *Festuca nigrescens*.

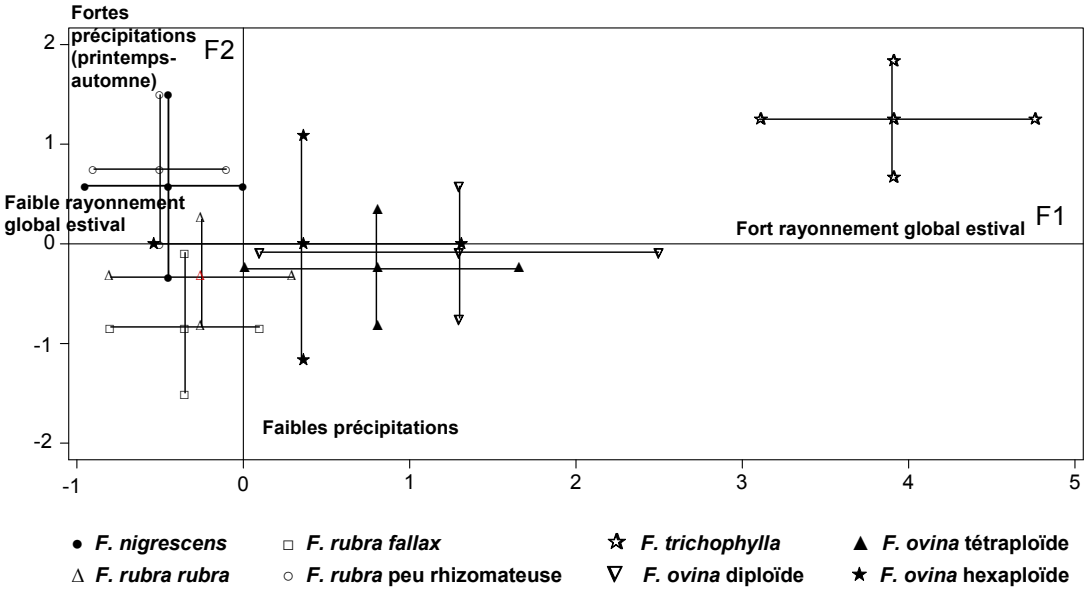

Figure 2. Analyse canonique des corrélations entre les occurrences des classes taxonomiques de fétuque à feuilles fines et une série de paramètres climatiques. Les barres horizontales et verticales représentent un intervalle de ± 1 unité d'écart-type des scores canoniques.

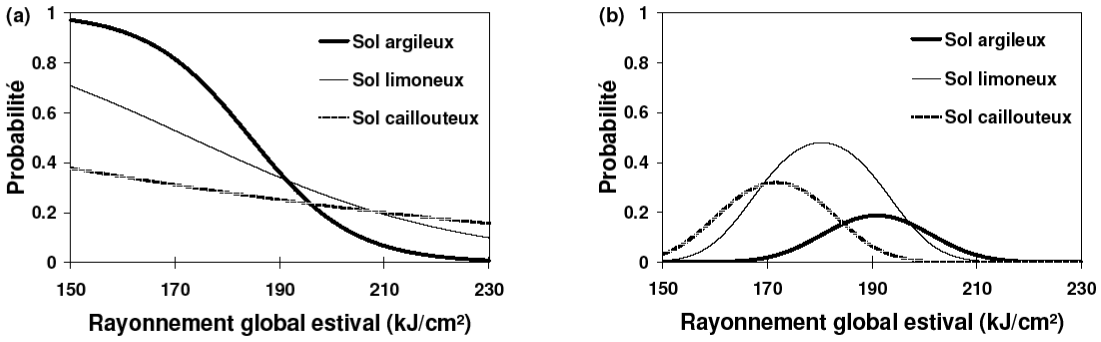

Figure 3. Courbes de réponse marginale de la probabilité de présence de deux taxons de fétuque à feuilles fines obtenues par régression logistique : (a) *Festuca nigrescens*, (b) *Festuca rubra fallax*.

Que peut on attendre des modèles de niche écologique ?

Une contribution aux sciences écologiques

La modélisation statistique de la niche écologique des espèces est un outil d'étude empirique de la distribution environnementale des espèces ; les modèles obtenus peuvent être utilisés pour tester par

une approche statistique des hypothèses en écologie et en biogéographie. Nous avons par exemple utilisé les modèles que nous avons obtenus pour les taxons de fétuque à feuilles fines afin d'évaluer les contributions relatives de la diversité adaptative des traits (ou caractères phénotypiques) et de la variation du niveau de polyploïdie à la diversité des niches réalisées de ces taxons (Sampoux et Huyghe, 2009). Cette étude montre que la diversité adaptative des traits contribue davantage que la variation du niveau de ploïdie à la diversité des niches. Elle suggère également que les cytotypes à haut niveau de polyploïdie ont pu être de vigoureux colonisateurs et compétiteurs, favorisant l'expansion géographique de la lignée des fétuques à feuilles fines, et conduisant à leur prévalence aujourd'hui constatée dans cette lignée.

Une contribution à l'organisation de prospections de ressources génétiques

De la modélisation de la distribution environnementale des espèces à la prédiction de leur aire de présence

Des modèles prédisant la probabilité de présence des espèces peuvent être construits en utilisant comme prédicteurs des paramètres environnementaux disponibles sur des grilles à résolution fine couvrant de vastes surfaces géographiques. Dans de tels cas, il est possible de prédire la probabilité de présence de l'espèce à chaque nœud de grille. Des logiciels SIG (Système d'Information Géographique) peuvent ensuite être utilisés afin de cartographier la probabilité de présence de l'espèce sur une zone d'intérêt. A partir de modèles de régression logistique établis pour plusieurs taxons de fétuque à feuilles fines en utilisant comme prédicteurs des paramètres climatiques disponibles sur une grille au pas kilométrique, nous avons obtenu des cartographies de l'aire probable de présence de ces taxons telles que celle présentée en Figure 4. Nous attendons bien sûr de ces résultats graphiques qu'ils soient un support efficace pour orienter de futures campagnes de prospection visant à compléter nos collections. Cependant, de telles représentations spatiales doivent être utilisées avec prudence lorsqu'elles sont produites pour une région différente de celle où se situent les sites d'observation à l'origine du jeu de données de calibration. Dans de tels cas, il est prudent de valider les prédictions obtenues dans la région pour laquelle une représentation spatiale est souhaitée avec une série d'observations effectives dans cette région.

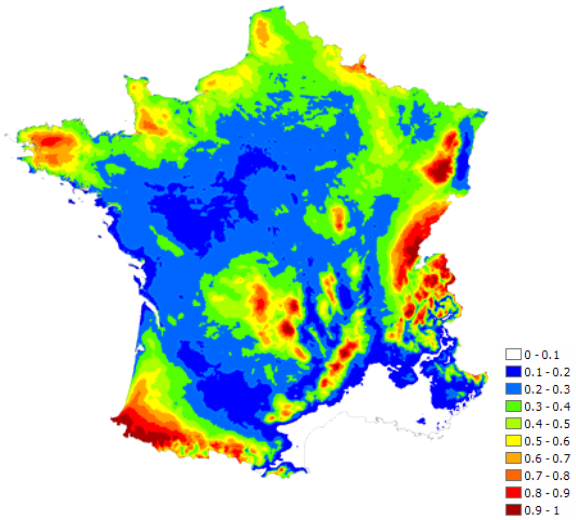

Figure 4. Carte de la probabilité de présence du taxon *Festuca nigrescens*.

A la recherche de phénotypes d'intérêt

Les pressions de sélection naturelle s'exerçant sur une espèce varient selon les conditions environnementales ; les variations phénotypiques intra-spécifiques résultant de la fluctuation des pressions de sélection permettent une certaine adaptation locale des espèces. Cet ajustement adaptatif du phénotype aux variations environnementales est à l'origine de la plupart des covariations constatées entre caractères phénotypiques et paramètres environnementaux. Comme l'on peut supposer que la sélection naturelle au sein d'une espèce tend à sélectionner à des phénotypes similaires dans des zones géographiques distantes mais semblables pour leurs conditions environnementales, il est fondé de chercher à prédire la valeur de caractères phénotypiques à un site donné à partir de la valeur à ce site des paramètres environnementaux avec lesquels ils sont statistiquement liés. Au sein de la diversité intra-taxon des fétuques à feuilles fines, nous avons observé un certain nombre de relations entre caractères phénotypiques et paramètres climatiques qui ont pu être modélisées de façon satisfaisante par des régressions linéaires (Figure 5, par exemple). Pour plusieurs taxons, la combinaison de ces modèles de régression linéaire avec le modèle climatique de probabilité de présence du taxon a permis de calculer, en fonction des conditions climatiques, la probabilité de présence d'une population du taxon pour laquelle la valeur du caractère phénotypique dépasserait un certain seuil. Ces probabilités ont été cartographiées pour différents caractères et pour différents taxons; la Figure 6 en donne un exemple pour la résistance à une maladie foliaire. Nous utiliserons ces représentations spatiales pour rechercher des populations de fétuques à feuilles fines présentant certaines caractéristiques phénotypiques particulièrement souhaitables.

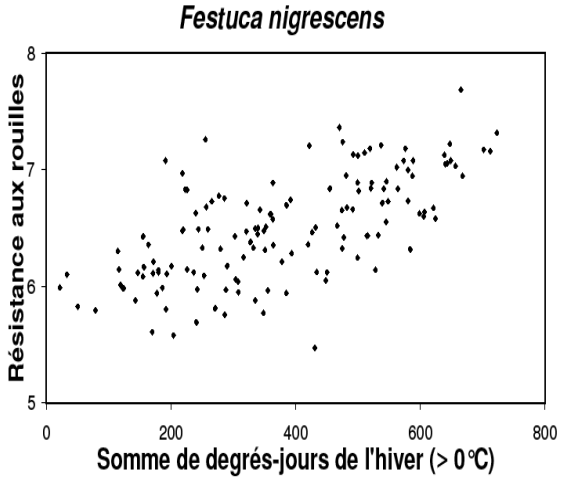

Figure 5. Distribution observée de la note de résistance aux rouilles des populations du taxon *Festuca nigrescens* en fonction de la somme de degrés-jours de l'hiver.

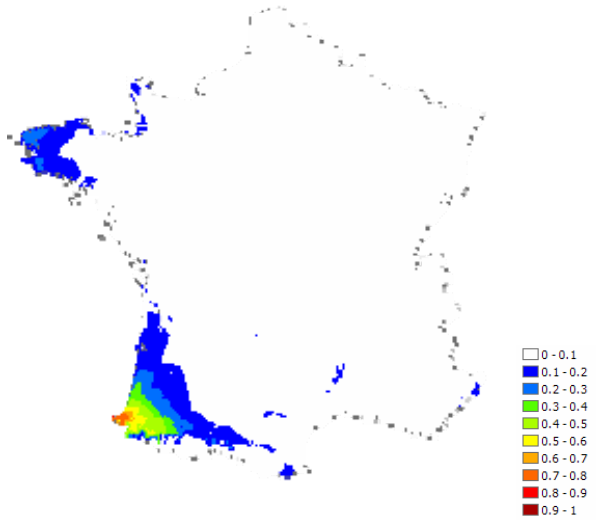

Figure 6. Carte de probabilité de présence de populations de *Festuca nigrescens* résistantes aux rouilles (résistance aux rouilles attendue au moins dans le premier tiers de la gamme de variation des populations déjà en collection).

Une contribution à la sélection de core-collections dans les programmes de conservation de ressources génétiques

Si l'intérêt des collections de ressources génétiques est bien de conserver la diversité la plus large possible, la conservation d'accessions dans les centres de recherche ou les stations d'amélioration génétique pose des problèmes logistiques et techniques croissant à mesure que le nombre d'accessions augmente. Frankel et Brown (1984) ont donc proposé le concept de core-collection qui consiste à définir, au sein d'une vaste collection, une sous-collection représentant au mieux la diversité génétique de l'ensemble de la collection avec un nombre limité d'accessions. Pour constituer une core-collection, Brown (1989a) a d'abord suggéré de procéder à un échantillonnage aléatoire de 10% des accessions de la large collection initiale (collection de base). Il a également proposé de stratifier la collection de base en classes à partir de leur origine géographique ou de marqueurs génétiques, et de tirer aléatoirement des accessions dans chaque classe (Brown, 1989b). Shoen et Brown (1993) ont ensuite proposé des méthodes visant à maximiser la diversité neutre dans la core-collection. Cependant, de telles méthodes ne sont efficaces pour maximiser la diversité sélectionnée que s'il existe un déséquilibre de liaison entre diversités neutre et sélectionnée et/ou si le flux de gènes entre populations d'où sont issues les accessions est faible (Bataillon et al., 1996). Le choix d'une core-collection peut également être effectué de façon à représenter au mieux la diversité de caractères phénotypiques préalablement observés. Dans ce cas, la collection de base peut être stratifiée en classes en fonction de la distance phénotypique entre accessions ; la core-collection est alors constituée en tirant aléatoirement des accessions dans chaque classe. Noirot et al. (1996) ont proposé une approche plus directe, connue sous le nom de méthode PCS (Principal Component Score method), qui consiste à sélectionner les accessions de façon à maximiser un critère de diversité phénotypique dans la core-collection. Toutefois, certaines différences d'adaptation environnementale entre accessions peuvent ne pas être détectées quand les accessions sont évaluées dans un réseau expérimental n'incluant que quelques lieux. De plus, il n'est possible d'observer qu'un nombre limité de caractères phénotypiques dans une grande collection. Maximiser la diversité phénotypique observée n'implique donc pas que la diversité génétique sélectionnée soit maximisée. De façon à limiter l'apparentement génétique entre accessions dans la sélection d'une core-collection de populations géo-référencées, Balfourier et al. (1998) ont proposé plusieurs méthodes basées sur la prise en compte des patrons d'auto-corrélation spatiale des caractères phénotypiques, en supposant que ceux-ci sont causés par les variations spatiales de pression de sélection environnementale et par l'isolement génétique croissant avec la distance.

Dans le cas de collections de populations spontanées échantillonnées au sein d'une espèce sur une vaste aire géographique, nous allons maintenant indiquer comment un modèle statistique de la niche écologique de l'espèce pourrait contribuer à optimiser le choix des accessions entrant dans une core-collection.

Les cartes de probabilité de présence des espèces produites à l'aide de SIG peuvent parfois mettre en évidence que certaines zones de présence de l'espèce sont très probablement disjointes. Ainsi, la Figure 4 suggère que plusieurs zones de présence de *Festuca nigrescens* sont géographiquement isolées entre elles par des variations bioclimatiques. Lorsqu'une telle fragmentation macro-géographique de l'aire de présence apparaît probable, nous proposons de stratifier la collection de base en groupes correspondant aux zones géographiques prédites comme disjointes desquelles les accessions proviennent, et de sélectionner des accessions pour la core-collection indépendamment dans chaque groupe. Ce procédé devrait permettre d'assurer la capture de la diversité génétique de plusieurs groupes dont la différenciation génétique serait maintenue par l'isolement géographique. L'isolement par la distance est de plus reconnu comme un facteur favorisant le développement de barrières reproductives entre les groupes isolés. Lorsqu'il existe plusieurs groupes d'accessions non inter-fertiles au sein d'une collection, Noirot et al. (1996) attirent l'attention sur le fait que la méthode PCS devrait être utilisée séparément sur chaque groupe ; dans ce cas, la stratification en groupes

isolés géographiquement que nous proposons pourrait contribuer à remplir cette condition d'utilisation de la méthode PCS.

De plus, la sélection d'accessions au sein de chaque groupe géographiquement isolé pourrait également prendre en compte avec avantage la probabilité de présence prédite de l'espèce aux sites où les accessions ont été collectées. Dans les sites où la valeur prédite de la probabilité de présence de l'espèce est faible, la présence de l'espèce peut être supposée assez précaire, car de faibles changements environnementaux sont susceptibles de rendre le site impropre à la présence de l'espèce. De plus, dans les zones géographiques où les principaux facteurs environnementaux conditionnant la présence d'une espèce sont sous-optimaux, la présence de l'espèce est davantage conditionnée par le niveau de facteurs environnementaux secondaires. Par exemple, les Figures 3(a) et 3(b) montrent que la présence des taxons de fétuque à feuilles fines dépend davantage de la texture du sol dans la partie du gradient de rayonnement global estival qui leur est la moins favorable. Par conséquent, la prospection d'une espèce peut se révéler difficile, et donc coûteuse en temps et en moyens, dans une zone géographique où ses déterminants environnementaux principaux sont sous-optimaux. La sélection d'une core-collection devrait donc délibérément retenir une part de la diversité provenant de sites où la probabilité de présence attendue de l'espèce est faible ; la diversité présente dans de tels sites étant la plus menacée par les changements environnementaux et la plus coûteuse à collecter. Pour atteindre cet objectif, nous suggérons que la collection de base (ou les groupes d'accessions correspondant à des zones de présence géographiquement isolées) soit stratifiée en classes en fonction de la probabilité de présence attendue de l'espèce aux sites de collecte, puis que les accessions soient sélectionnées de façon indépendante au sein de chaque classe. Comme le modèle statistique à l'origine des prédictions a toujours une fiabilité limitée par les choix techniques, nécessairement simplificateurs, de la modélisation et par la qualité et la taille du jeu de données de calibration, nous suggérons de ne pas distinguer plus de deux ou trois classes d'intervalles de même amplitude de probabilité de présence de l'espèce. Au sein d'une espèce, l'hypothèse peut être envisagée que certains caractères phénotypiques soient distribués selon les mêmes gradients environnementaux que ceux qui déterminent la présence de l'espèce ; dans ce cas, les classes de probabilité de présence attendue de l'espèce devraient structurer la diversité de tels caractères, bien que les classes de plus faible probabilité de présence devraient rassembler à la fois les valeurs phénotypiques les plus faibles et les plus élevées. Toutefois, certains caractères phénotypiques peuvent aussi être distribués selon d'autres gradients environnementaux que ceux qui conditionnent la présence de l'espèce (c'est le cas de l'exemple de la Figure 5 où la résistance aux rouilles de *Festuca nigrescens* se distribue selon un gradient de somme de degrés-jours de l'hiver alors que ce paramètre environnemental n'a pas été identifié comme déterminant pour la présence du taxon). Pour de tels caractères, une large gamme de variation peut être attendue au sein de chaque classe de probabilité de présence. Il semblerait donc judicieux d'établir une core-collection en s'assurant que la diversité la plus grande possible soit sélectionnée au sein de chaque classe de probabilité de présence, par exemple en utilisant la méthode PCS de Noirot et al. (1996) qui maximise la diversité phénotypique observée.

Conclusion et perspectives

Dans le cas de collections de diversité spontanée des espèces fourragères et à gazon, la modélisation statistique de la niche écologique des espèces se révèle un outil utile pour piloter géographiquement les prospections et pour définir des core-collections. Cette modélisation de la niche des espèces pourrait offrir des avantages identiques pour d'autres espèces d'intérêt agronomique disposant d'un compartiment de diversité naturelle, ainsi que pour des espèces sauvages apparentées à certaines espèces domestiques. La modélisation de la niche écologique des espèces est aujourd'hui facilitée par le développement de méthodes statistiques adaptées à différents types de situation et par la disponibilité croissante de bases de données environnementales. De plus, les logiciels SIG offrent de nouvelles opportunités de représentation spatiale des patrons de distribution environnementale de la

diversité. Dans un autre registre de la conservation des ressources génétiques, les prédictions de distribution environnementale basées sur les modèles statistiques de niche devraient aussi pouvoir rendre des services importants pour l'optimisation des réseaux de conservation *in situ*.

En plus de leur intérêt pour la gestion des ressources génétiques, les modèles de niche écologique pourraient éventuellement être valorisés pour un autre type d'objectif agronomique. Ces modèles pourraient en effet être utilisés afin d'évaluer l'adéquation des conditions environnementales d'un site avec la présence d'une espèce que l'on souhaiterait y implanter. C'est d'ailleurs dans cet objectif que fut réalisée l'une des premières tentatives de modélisation de niche d'espèce végétale (Nix et al., 1977, in Guisan et Thuiller, 2005). Souvenons-nous toutefois que la modélisation statistique de la niche des espèces ne prédit pas la niche fondamentale qui représente les conditions environnementales favorables à la présence d'une espèce en l'absence de compétition et de prédation, mais une niche plus restreinte basée sur des observations de présence réelle.

Les données associées aux collections de ressources génétiques d'origine spontanée constituent une information pertinente pour la calibration des modèles de niche écologique. Mais les vecteurs de reproduction conservés dans les collections de ressources génétiques offrent de plus la possibilité d'évaluer en conditions contrôlées la diversité de caractères phénotypiques et de paramètres éco-physiologiques à partir de l'observation d'un nombre important d'échantillons, puis d'en étudier la distribution dans les environnements d'origine. Les collections de ressources génétiques d'origine spontanée forment donc des outils qui pourraient être davantage utilisés dans un questionnement d'écologie ou de biogéographie. L'étendue souvent vaste des territoires prospectés se prête bien à l'étude de la distribution de la diversité à l'échelle macro-environnementale ; elle est en particulier bien adaptée à la prise en compte des gradients bio-climatiques, et devrait donc être favorable à l'estimation des paramètres de distribution utiles à l'étude de l'impact des changements climatiques sur la diversité.

Références bibliographiques

Araujo M., Guisan A., 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography 33, 1677-1688.

Austin M., 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling 200, 1-19.

Badeau V., Dupouey J.L., Cluzeau C., Drapier J., Le Bas C., 2009. Climate change and French tree species biogeography: first results and perspectives. *Carbon sequestration in major French ecosystems* (ed. by D. Loustau), in press. Editions Quae, Versailles, France**.**

Balfourier F., Charmet G., 1991. Relationships between agronomic characters and ecogeographical factors in a collection of French perennial ryegrass populations. Agronomie 11, 645-657.

Balfourier F., Charmet G., Prosperi J.M., Goulard M., Monestiez P., 1998. Comparison of different spatial strategies for sampling a core collection of natural populations of fodder crops. Genetics Selection Evolution 30 (Suppl. 1), S215-S235.

Bataillon T.M., David J.L., Schoen D.J., 1996. Neutral genetic markers and conservation genetics: simulated germplasm collections. Genetics 144, 409-417.

Brown A.D.H., 1989a. Core collections: a practical approach to genetic resources management. Genome 31, 818-824.

Brown A.D.H., 1989b. Size and structure of collection: the case for core collection, in: Hodgkin T., Brown A.D.H., Hintum T.J.L. van, Morales E.A.V. (eds), The Use of Plant Genetic Resources, John Wiley & Sons, Baffins Lane, Chichester, UK, pp. 136-156.

Carnes B.A., Slade N.A., 1982. Some comments on niche analysis in canonical space. Ecology 63, 888–893.

Coudun C., Gégout J.C., 2006. The derivation of species response curves with Gaussian logistic regression is sensitive to sampling intensity and curve characteristics. Ecological Modelling 199, 164–175.

Frankel O.H., Brown A.D.H., 1984. Current plant genetic resources – a critical appraisal, in: Genetics New Frontiers, Proceedings of the 15th International Congress of Genetics, vol. 4, Oxford and IBH Publishing Co, pp. 3-13.

Gimaret-Carpentier C., Dray S., Pascal J., 2003. Broad-scale biodiversity pattern of the endemic tree flora of the Western Ghats (India) using canonical correlation analysis of herbarium records. Ecography 26, 429–444.

Guisan A., Thuiller W., 2005. Predicting species ditribution: offering more than simple habitat models. Ecology letters 8, 993-1009.

Grinnell J., 1917. Field tests of theories concerning distributional control. The American Naturalist 51, 115-128.

Grinnell J., 1924. Geography of evolution. Ecology 5, 225-229.

Hotelling H., 1936. Relations between two sets of varieties. Biometrika 28, 321–377.

Hutchinson G.E., 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 22, 145–159.

Johnson J.B., Omland K.S., 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19, 101-102.

McCullagh P., Nelder J.A., 1997. Generalized linear models. Monographs on statistics and applied probability. Chapman & Hall, London.

Nagelkerke N.J.D., 1991. A note on a general definition of the coefficient of determination. Biometrika 78, 691–692.

Nix H., McMahon J., Mackenzie D., 1977. Potential areas of production and the future of pigeon pea and other grain legumes in Australia. In: Wallis E.S. & Whiteman P.C. (eds) *The potential for pigeon pea in Australia. Proceedings of Pigeon Pea (Cajanus Cajan (L.) Millsp.) Field Day*. University of Queensland, Queensland, Australia, pp.5/1-5/12.

Noirot M., Hamon S., Anthony F., 1996. The principal component scoring: a new method of constituting a core collection using quantitative data. Genetic Resources and Crop Evolution 43, 1-6.

Pearce J., Ferrier S., 2000a. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecological Modelling 128, 127–147.

Pearce J., Ferrier S., 2000b. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling 133, 225–245.

Prosperi J.M., Jenczewski E., Angevain M., Ronfort J., 2005. Morphologic and agronomic diversity of wild genetic resources of Medicago sativa L. collected in Spain. Genetic Resources and Crop Evolution 53, 843-856.

Sampoux J.P., Huyghe C., 2009. Contribution of ploidy-level variation and adaptive trait diversity to the environmental distribution of taxa in the 'fine-leaved fescue' lineage (genus *Festuca* subg. *Festuca*). Journal of Biogeography 36, 1978–1993.

Schoen D.J., Brown A.D.H., 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc. Natl. Acad. Sci. USA 90, 10623-10627.

Thuiller W., Araújo M.B., Lavorel S., 2004. Do we need land-cover data to model species distributions in Europe? Journal of Biogeography 31, 353–361.

Warren J.M., Raybould A.F., Ball T., Gray A.J., Hayward M.D., 1998. Genetic structure in the perennial grasses *Lolium perenne* and *Agrostis curtisii*. Heredity 81, 556-562.

Wintle B.A., McCarthy M.A., Volinsky C.T., Kavanagh R.P., 2003. The use of bayesian model averaging to better represent uncertainty in ecological models. Conservation Biology 17, 1579-1590.

Zaniewski A.E., Lehmann A., Overton J.M.C., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling 157, 261–280.