Ecole de Biologie

ECOLOGICAL AND PHYLOGEOGRAPHICAL APPROACH OF A BIOLOGICAL INVASION: *PRUNUS SEROTINA*, A CASE STUDY

Master of Sciences in Evolutionnary and Conservation Biology

by

Blaise PETITPIERRE

Director : Prof. Antoine Guisan
Supervisors: Dr. Guillaume Besnard, Olivier Broennimann and Prof. Antoine Guisan
Expert: Dr. Alexandre Hirzel

Department of Ecology and Evolution

January 2008

Table of contents

Abstract

Invasion by non-indigenous plant species is currently a major threat for biodiversity, economy and health. Prediction and anticipation of future biological invasions are central to avoid or mitigate such impacts. Here, we combined niche modeling with phylogeography to assess the ongoing invasion of the native North-American tree species *Prunus serotina* in Europe. First, ecological niche models were developed from 7015 occurrences distributed in the native and invasive ranges, by relating observed occurrences to eight climatic variables. We further performed a principal component analysis to investigate the ecological requirements of the species in the space of climatic parameters. Second, the phylogeography of the species was investigated by analyzing cpDNA polymorphisms from 52 populations distributed in both ranges to determine which chlorotypes were introduced in Europe. Then, both approaches were combined to distinguish how chlorotypes differed in their environmental settings. Prediction accuracy of the models calibrated in the native range was low when projected on the invaded range. It was dramatically increased when models were calibrated with data from both ranges, keeping a high predictive ability in the native range. Six chlorotypes were found. Three of them were common to both ranges, supporting multiple introductions. We found a pronounced climatic niche shift between the occurrences in the two ranges, but also between the European and American distribution of every chlorotypes introduced in Europe. Such a climatic shift may limit the utility of ecological niche models in predicting the full extent of biological invasion.

Introduction

Introduction of non-native species in a new habitat is the second cause of biodiversity loss with habitats decline (Wilcove *et al.,* 1998). This phenomenon has exploded with commercial exchanges and all the human activities linked to them (Levine & D'Antonio, 2003). Each year, invasive species cause billions dollars of damages (Pimentel *et al.,* 2000). A lag phase often precedes invasion by introduced species while they remain inconspicuous (Mack *et al.*, 2000) but once established, invasive alien species are very difficult to eradicate. That is why preventing such biological invasions is of great interest (Byers *et al.*, 2002).

Environmental niche modelling (ENM) allows predicting distribution of the species. This approach is based on the climatic niche of the species, defining a climatic space in which the species can grow and survive, and then project it on the geographical space. The latter shows all geographic locations of potentially suitable habitat for the species. Despite several limitations, such climate-matching approach has shown some success in predicting invasive ranges (Peterson & Vieglais, 2001; Thuiller *et al.*, 2005; Zhu *et al.*, 2007).

Several assumptions are necessary to use ENM. First, climatic factors are assumed to be the main force driving the species distribution at the continent scale (Woodward, 1987). Second, occurrences distribution should represent the realized niche of the species and finally, the climatic niche should be conserved between the native and the invasive ranges. Environmental and biotic variables other than climate (such as soil type, disturbance regime or interspecific interactions) and basic species life history traits involved in biological invasion success (such as fecundity or dispersal ability; Higgins & Richardson, 1999) are ignored in climate matching models (Thuiller *et al.*, 2005). That is why ENM necessarily considers occurrence distribution as the realized niche of the species. Physiological limits and ecosystem constraints define the fundamental niche. The realized niche *sensu* Hutchinson (1957) additionally considers all the biotic interactions and competitive exclusions which generally lower the species

performances in their environmental habitats (Ellenberg, 1953; Malanson *et al.*, 1992; Malanson, 1997). Observations derived from the field can thus only inform us about the realized niche because they integrate all effects (such as biotic) other than strictly physiological. This can limit applications of ENM in changing environmental situations (Guisan & Theurillat, 2000). Hence, when applied to biological invasions, ENMs assume niche conservatism (Scott & Panetta, 1993; Peterson, 2003; Thuiller *et al.*, 2005), defined as the tendency of species to maintain their ancestral ecological requirements (Wiens & Graham, 2005). The main arguments are the conservative force of natural selection on the fundamental niche (Holt & Gaines, 1992) and the low level of niche differentiation between sister taxa (Peterson *et al.*, 1999). Recently, evidences for a niche shift between the native and the invasive ranges of the invasive weed *Centaurea maculosa* has been shown (Broennimann *et al.*, 2007). This questions the ability of the ENM approach to predict the full extent of biological invasions, although models still allow predicting the area of likely introduction (Broennimann *et al.*, 2007; Fitzpatrick *et al.*, 2007).

Broennimann *et al.* (2007) suggest that such a shift can result from changes in the species fundamental niche as caused by an evolutionary process (e.g. hybridization or evolution of increased competitive ability; Blossey & Notzold, 1995), changes in the realized niche as caused by a different biotic environment in the introduced range (e.g. enemy-release hypothesis; Keane & Crawley, 2002; Reinhart & Callaway., 2006), or both (Dietz & Edwards, 2006; see also Pearman et al. 2008). Niche shifts have been shown as an important driver of sympatric speciation (Losos *et al.*, 2003; Levin, 2005), whereas allopatric speciation can be explained by niche conservatism (Huntley *et al.* 1989; Peterson & Holt 2003; Wiens & Graham, 2005). Evolutionary components are key factors in the response of a species to an environmental change (Huey *et al.*, 2000; Genton *et al.*, 2005; Frankham, 2005) and investigating population and evolutionary genetics is central to better understand species'invasivenes. (Sakai *et al.*, 2001;

Lee, 2002). This study proposes to integrate a genetic component to the ENM approach.

The founder effect due to the low number of initial colonists results in the loss of genetic diversity in invasive populations (e.g. Husband & Barett, 1991; Amsellem *et al.*, 2000; Parisod *et al.*, 2005; Meimberg *et al.*, 2006; Besnard *et al.*, 2007; Pairon & Jacquemart, 2007). Consequences of reduced genetic diversity can be inbreeding depression that may limit population growth and lower the probability of population persistence. Another effect of a low genetic variability can limit the ability of introduced populations to evolve in their new environments (Allendorf & Lundquist, 2005). However, genetic variability remains a key to respond to rapid environmental changes and it seems paradoxal that invasive species succeed in new environments with a low genetic diversity (Allendorf & Lundquist, 2003; Frankham, 2005), especially knowing all the environmental changes implied by the niche shift problematic.

Several solutions exist for the species to deal with this dilemma, including asexual reproduction, self-fertilisation coupled to high reproductive rates allowing purging deleterious alleles, or ad-mixing differentiated populations introduced from several different sources which all limit the detrimental effects of inbreeding and genetic drift (Amsellem *et al.,* 2000; Sakai *et al.*, 2001; Lee, 2002; Kolbe *et al.*, 2004; Genton *et al.*, 2005; Frankham, 2005).

In the perspective of these complex evolutionary processes driving biological invasions, comparing the niche of native versus invasive occurrences may need to be conducted not at the species but at the haplotype level . Mau-Crimmins *et al.*, (2006) have already underlined the bias due to the limited introduced variant range comparing to the niche of the whole species. This study proposes a comparison of the climatic niche between the occurrence and the haplotype level and its consequence on predicting the potential invasive range of black cherry (*Prunus serotina*). This American native tree has broadly spread out into Europe since nearly four centuries. Occurrences and chloroplast sequences data were

combined to answer the following questions: i) which lineages have been introduced in Europe and what are the general patterns of diversity in the native and invasive ranges? ii) How does the niche behave between the ranges for global occurrences and for lineages introduced in the new range; and finally iii) is the invasive distribution of *P. serotina* predictable by ENM?

Material & Methods

The study species

Prunus serotina Ehrh., also known as "black cherry" in North America or "woodpest" in Europe, was chosen because of its well delimited distribution in its native area (North America). Many invasive plants are linked to human disturbances or activities and have a very broad ecological and geographical distribution, which leads to some difficulties in the use of ecological modeling. On the sight of preliminary investigations, this problem seems limited for *P. serotina* which native distribution seems more delimited by ecological than by human factors.

Prunus serotina is a deciduous tree or shrub from the Rosaceae family. In its native range, it grows up to 38 m, while in its invaded range it is mostly a shrub, rarely a tree up to 20 m tall. Leaves are simple, ovate to lanceolate, 5 – 15 cm long, 2.5 - 5 cm wide, tough, bright green above, pale green below, with finely toothed margins. Inflorescence is a raceme 6 – 5 cm long, with about 30 flowers showing a calyx tube of short lobes with 5 white petals. Fruits are dark red to black berries of 8 – 10 mm in diameters, from which the species has its common name (CPS, 2006; EPPO, 2007; Starfinger, 2007; USDA, 2007).

In its native range, *P. serotina* covers most of the eastern part of the USA. It occurs from Nova Scotia south to Northern Florida and west to Minnesota. Its western range raises Nebraska, Kansas and Oklahoma. In the southern part of its distribution, outlying populations are found in central Texas, west Texas, New

Mexico, Arizona and Mexico to Guatemala (Fig. 1; Little, 1971). It is naturalized in several countries from northern South America (PLANT, 2007), such as Ecuador and Peru (EPPO, 2007). Black cherry is one of the most valued cabinet and furniture wood in North America (PLANT, 2007) but large, high-quality trees suited for a commercial use are found in a restricted commercial range on the Allegheny Plateau of Pennsylvania, New York and West Virginia (Marquis, 1975; Hough, 1965). Smaller quantities of high-quality trees are found scattered along the southern Appalachian Moutains and the upland areas of the Gulf Coastal Plain (Marquis, 1990), in Ohio, Tennessee and Virginia (Walters, 1985). Elsewhere in its native distribution, black cherry is often small and poorly formed (Marquis, 1990) like in its invaded range. Based on phenotypical criterion such as the height of the trees and the thickness of the leaves, five varieties are considered within the *P. serotina* clade: *P. s.* var. *eximia* (Small) Little is confined to the Edwards Plateau of Central Texas, *P. s.* var. *rufula* (Woot. & Standl.) McVaugh and var. *virens* (Woot. & Standl.) McVaugh are distributed in Texas, New Mexico and Arizona, and populations of *P. s.* var. *salicifolia* (Kunth) Koehne (or subsp. *capuli*) are found from Mexico to Guatemala. The most abundant and common variant is *P. s.* var. *serotina* widespread in all the rest of the USA (PLANT, 2007). Because no genetic differentiation has been shown, the important quantity of occurrences in the *P. s.* var. *serotina* range and the lack of accuracy of the data for this phylogeny level, varieties were not considered in the models.

Prunus serotina was first introduced in Europe in the early 17th century. It reached France in Paris around 1630 (Starfinger, 2007), England in 1629 (Hough, 1957 in EPPO, 2007) and Germany in 1685 (Starfinger & Kowarik, 2003) where it was particularly appreciated as ornamental trees in gardens. Since the 19th century, it was massively planted throughout Europe for several reasons. Foresters were interested in the value of *P. serotina* showed in its native range and black cherry is able to grow under conifer plantation. Timber plantations were never achieved because of poor results but the black cherry was used as a wind

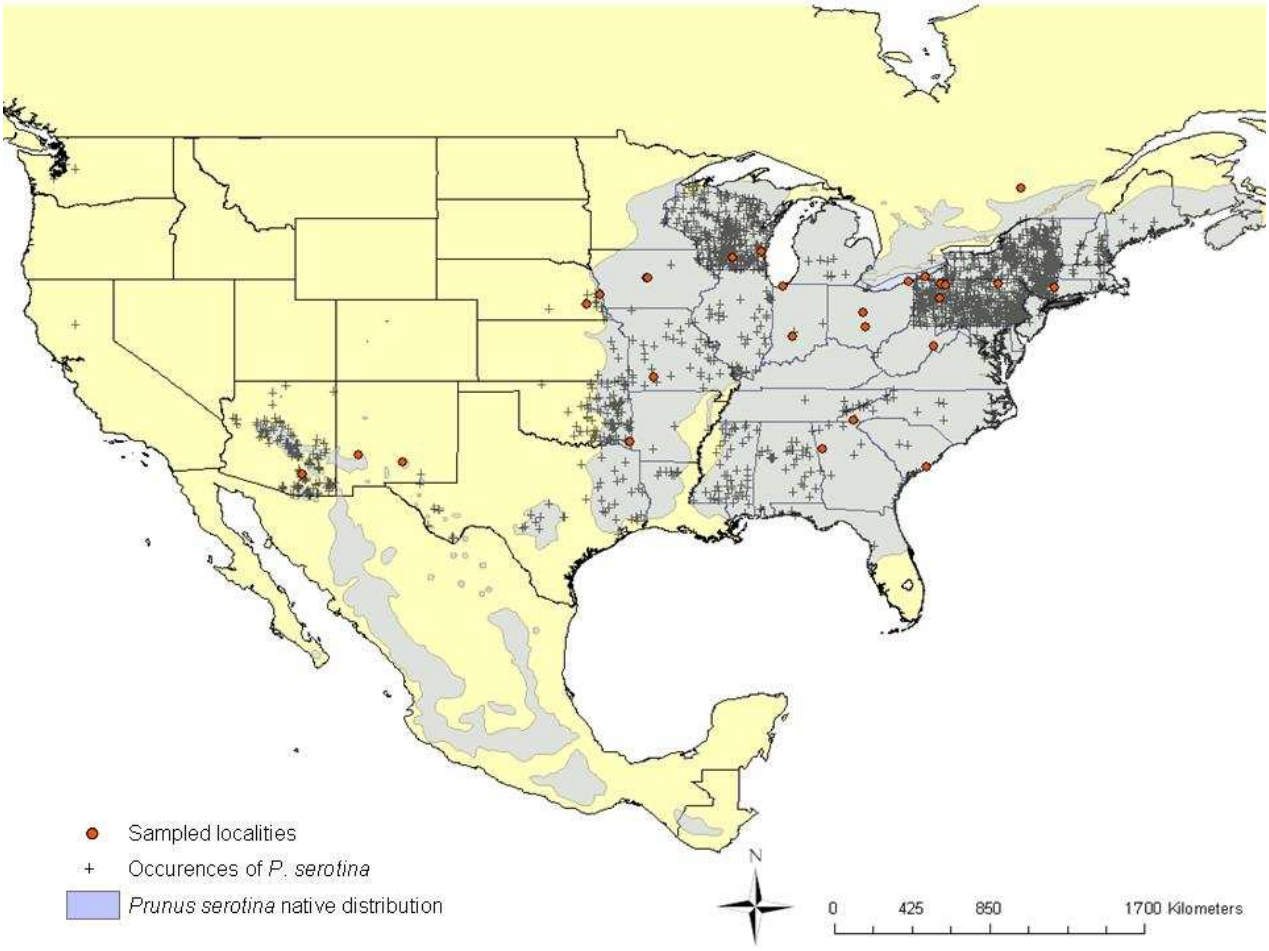


Fig. 1: *P. serotina* distribution in North America. The blue layer shows *P. serotina* native distribution (Little, 1971), crosses are occurrences used in the models and orange points are sampled localities.

and fire breaks and as a soil quality improver. It was thought to speed up soil decomposition due to the low C/N ratio in its leaves. These are the reasons why it is now widespread in France, Germany, Denmark, Netherlands, Belgium and Poland and locally abundant in Lithuania, Romania, Hungary, Switzerland and Italy Starfinger, 2007; Godefroid *et al.*, 2005; CPS 2006; Sartoti, 1988; Fontanetto *et al.,* 2004). Additionally, presence of naturalized specimens of *P. serotina* are signaled in Sweden, Norway, Estonia, Russia, Austria, Bosnia and Herzegovina, Bulgaria, Belarus, Croatia, Czechia, Luxembourg, Serbia and Montenegro, Spain, Slovakia, Slovenia, United Kingdom and Ukraine (EPPO, 2007). It is rare or absent in Mediterranean countries. In Switzerland, it is

especially abundant in the south of Ticino around the area of Lugano in woodlands and borders (CPS, 2007; personal observations).

In its invasive range, black cherry grows in both forests and open habitats. Even if it is classified as a shade intolerant in its native area (Marquis, 1990), *P. serotina* reaches its highest cover value in forests. Seedlings can develop in the shade of an overstory or in a partially cut stands (Godefroid *et al.*, 2005). Thanks to its toleration to a very wide panel of moisture conditions (Starfinger, 2007), black cherry can invade open habitats such as wetlands, e.g. bogs and their degeneration stages, but also dry grassland too dry for most other woody plants. *Prunus serotina* benefits from anthropogenic influence such as soil disturbance or

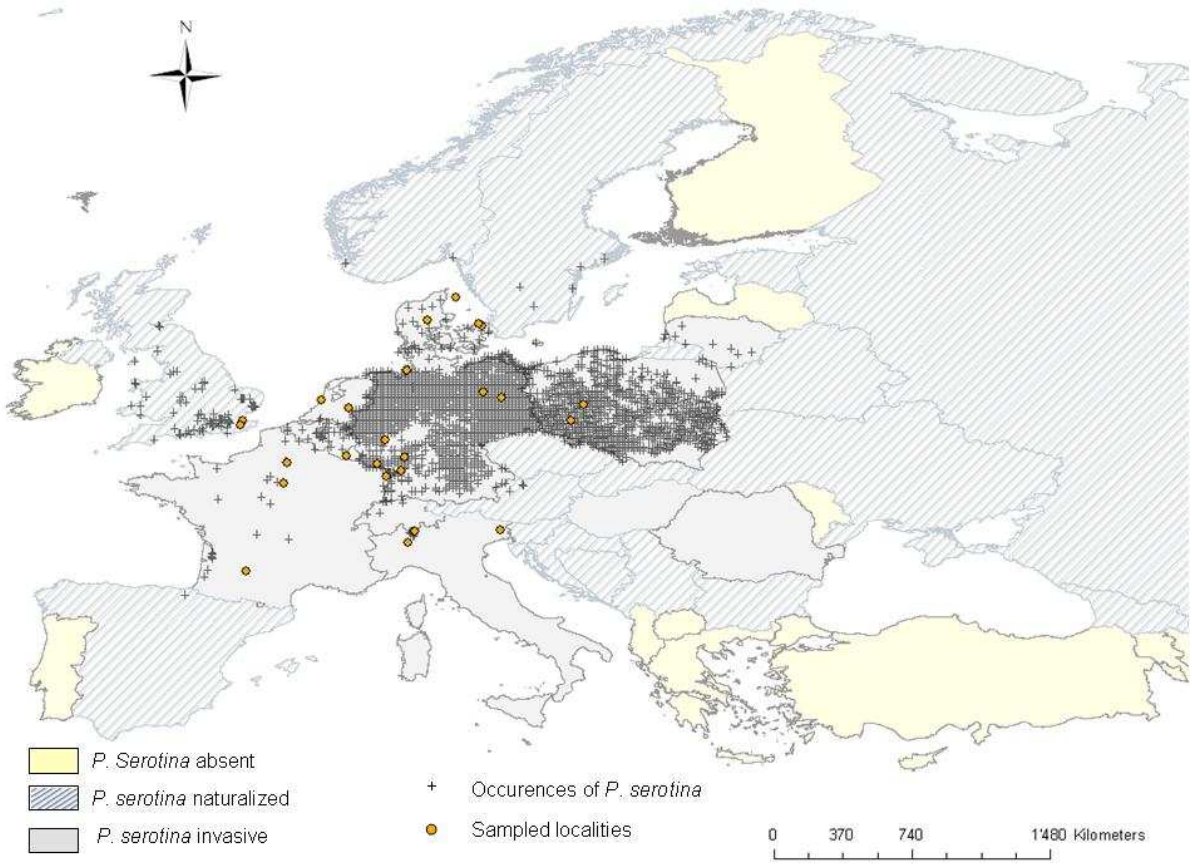


Fig. 2: *P. serotina* invasive range. Black cherry is invasive or locally abundant in grey countries, naturalized but no problematic in streaked countries and not signaled in yellow countries. Crosses represent occurrences used in the models and points indicate sampled localities

eutrophication (Godefroid *et al.*, 2005). It is sometimes planted in hedgerows in the agricultural landscape and can reproduce and disperse here. It is also common in urban areas, parks and gardens, in particular in less intensively managed situations. It mainly grows on sandy soil (Godefroid *et al.*, 2005), but the species can also germinate and establish on a wide range of other soil types (peaty to sandy).

Prunus serotina disperses by seeds but has a very efficient asexually reproduction by suckering and sprouting. If it is sufficiently exposed to light, it produces an abundant quantity of seeds since the age of seven. Seeds are dispersed by birds and frugivore mammals but its dispersion by seeds seems quite limited and the main force of its spread in Europe is planting (Starfinger, 2007).

By overshadowing the ground, forming dense shrub layers, competing for nutrient and water resources, *P. serotina* impedes woodland regeneration and presents an important threat for indigene species even if the conservation value of woodlands species is often quite small. When it colonizes open vegetation habitats such as dry grasslands or bogs containing more rare or endangered species, *P. serotina* becomes a more problematic conservation problem. Cyanhydric acids present in the whole plants may cause allelopathic effects (Starfinger, 2007) and may cause sickness or even death in animals, although many species feed on the fruit (Kingsbury, 1964). The positive expected effect on the litter is generally poor or even can be negative by decreasing the soil pH (Starfinger, 2003). Its sprouting abilities and its abundance make the control over this species quite heavy.

Study area and population sampling

Both ranges, native and invasive, were included in this study (Fig. 1 and 2). Even though *P. serotina* is distributed from South-Eastern Canada to Guatemala in its native range, only USA and Canadian data were used because European

populations were probably introduced from the woody culture in North-America. Moreover, systematic is unclear about several subspecies of *P. serotina* present in South America.

The occurrences data concerning the invaded range are distributed along nine European countries: France, Netherlands, Germany, Denmark, Belgium, Switzerland, Poland where *P. serotina* is considered as invasive (Branquart, 2007; CPS, 2006; EPPO, 2007; Starfinger, 2007; Fontaneto *et al.*, 2004), Austria and England, in which it seems not to cause major problems.

Populations sampling for molecular analysis covered well the native and the invasive ranges of *P. serotina*. Samples were provided by different herbariums, universities, particular sampling and personal sampling. In total, 52 populations (27 and 25 in the native and invasive ranges, respectively) were sampled between 2005 and September 2007. In each population, 2-3 leaves were collected from 5 individuals randomly distributed or chosen along a transect, except for both New-Mexican and Arizonian populations, where only one individual was sampled. Sampled tree individuals were at least 15 meters apart from each other but the population diameter was not greater than 1 kilometer. A total of 238 individuals were analyzed (Appendix 1 & 2).

Plastid DNA characterization

About 20 mg of dry leaf material were ground in a 2-ml Eppendorf tube using a Tissuelyzer (QIAGEN) and three tungsten balls in each tube. Genomic DNA was extracted using a DNA-mini-extraction Kit (QIAGEN) according to the manufacturer's protocol.

In order to characterize maternally inherited polymorphisms, the plastid genome (ptDNA) was investigated. Three non-coding plastid regions (*trnT-trnL, trnD-trnT* and *trnS-trnG*) were sequenced on eight individuals originating from distant populations (NM2, PE7, IO1, SC1, UK1, IT1, FR1, NE2) using either universal primers (for *trnT-trnL* and *trnD-trnT*; Taberlet *et al.*, 1991; Demesure *et*

al., 1995) or specific primers (for *trnS-trnG*, Table 1. Each PCR reaction (50 µl) contained 10 ng DNA template, 1X reaction buffer, 0.2 mM dNTPs, 0.2 µmol of each primer, and 0.75 Units *Taq* DNA polymerase (Promega, Madison, WI, USA). Reaction mixtures were incubated in a thermocycler (T1; Biometra, Göttingen, Germany) for 3 min at 94ºC, followed by 36 cycles of 45 s at 94ºC, 45 s at 53ºC or specified temperature (Table 1), and 120 s at 72ºC.The last cycle was followed by a 10 min extension at 72ºC. PCR products were then purified using a QIAquick Purification Kit (QIAGEN) following the manufacturer protocol. Direct sequencing was performed using the Big Dye 3.1 Terminator sequencing cycle (Applied Biosystem) according to manufacturer's instructions and an ABI PRISM 3100 genetic analyzer (Applied Biosystem).

Based on ptDNA sequences, six loci were developed to characterize our plant sample. Primers (Table 1 were designed in regions flanking polymorphisms (two substitutions and four length polymorphisms). Each PCR reaction (25 µl) contained 10 ng DNA template, 1X reaction buffer, 0.2 mM dNTPs, 0.2 µmol of

Table 1: Characteristics of specific primers used to characterize chloroplast polymorphisms, including temperature of annealing (Ta in °C) and the allele sizes of the variants (in bp). *TrnS-TrnG* and its internal primers were used to find polymorphisms in the *TrnS-TrnG* spacer sequence. The rest of the primers were used to amplify polymorphic loci.

Locus	GenBank	Primers	Ta	Allele sizes	Label
TrnS-TrnG	X0000000	F. GTG'AAA'CTT'TGG'TTT'CAT'C R. GAA'AAA'AGA'GAA'GAC'TAT'GTT'AC	47	-	
TrnS-TrnG internal	X0000000	F.GAT'TCT'TGG'ATA'CAA'ATCACG R. TTA'ATC'CTT'TAC'CTC'TCA'ATG	49	-	
TrnS-G res	X0000000	F.CTA'TAA'TCA'TAG'AAA'TCT'AAA'TAA'ACA R[1)]. TC'GTA'AAT'AAA'CTG'ATT'TAT'TTG'ATT	47	108, 73, 66	FAM
TrnD-T res	X0000000	F. GGA'TAA'TCA'CTCTTT'CAA'TGT R[2)]. AAT'TCT'GAT'CTT'GCT'AAT'GAT'C	50	108, 86, 85	HEX
TrnD-T indel	X0000000	F[1)]. ATC'ATT'AGC'AAG'ATC'AGA'ATT'T R. AGA'GAA'AGG'TTT'TCT'GAT'ATA'T	47	82,77	FAM
TrnT-L poly T	X0000000	F[1)]. ATT'AGC'TTA'ATT'AGA'TAG'TAA'G R. CCG'TTA'ATT'TAT'AAT'TAG'AAG'A	47	132, 128, 127	FAM
TrnT-L poly A	X0000000	F. TAA'TTC'AGA'TCA'TAA'TGA'AAC'A R[1)]. GTG'CAA'TTT'ATA'CTT'GAA	47	106, 105	FAM
TrnT-L indel	X0000000	F. GGG'CAT'ATC'TAA'GTA'TAA'TAA R[2)]. TTG'GTC'CAA'GAA'ATG'ATT	47	164, 126	HEX

[1)] Primers labeled with FAM fluorochrome

[2)] Primers labeled with HEX fluorochrome

each primer (one labelled with a fluorochrome; Table 1), and 0.75 Units Taq DNA polymerase (Promega, Madison, WI, USA). Specific PCR conditions for the selected primer pairs are reported in Table 1. Reaction mixtures were incubated in a thermocycler (T1; Biometra, Göttingen, Germany) for 3 min at 94ºC, followed by 36 cycles of 30 s at 94ºC, 30 s at the defined annealing temperature and 30 s at 72ºC. The last cycle was followed by a 10 min extension at 72ºC. Digestion of two PCR fragments (*TrnDT-res* and *TrnSG-res*, Table 1) was performed using the restriction enzyme *Tas*I (Fermentas, St Leon-Rot, Germany) according to the manufacturer's recommendations. Electrophoresis of PCR products was carried out in denaturing 5% (w/v) polyacrylamide gels using an automated ABI 377 sequencer (Applied Biosystems).

Molecular data analyses

The phylogenetic relationships between *P. serotina* divergent haplotypes were illustrated with a median-joining haplotype network constructed with the software NETWORK (Bandelt *et al.*, 1999). Distribution of the haplotypes was showed as pies indicating the proportion of haplotypes in each locality. Population abbreviations are described in Table 1.

Haplotypes diversity (the probability that two alleles are different), allelic richness (indication of the mean number of different alleles found in a population) and the level of genetic differentiation (Weir & Cockerham, 1984) were compared between the native and invasive ranges to estimate the level of diversity using the software FSTAT 2.9.3.2 (Goudet, 1995). Diversity and allelic richness are affected by the sample size, so as population from Arizona and New-Mexico (AZ, NM1 and NM2) were not taken in account in analyzes. All tests in FSTAT are randomization-based

Species occurrences data collection

Species occurrences of *P. serotina* were gathered from several Herbarium, land management and state agencies. Only occurrences with locational accuracy equal to or finer than the resolution to the climate data were kept. Sampling occurrences were added to the database. All the occurrences were digitalized and treated in the software ArcMap™ 9.1 (Environmental Systems Research Institute, Inc., Redlands, CA, USA). The final database consisted in 3207 occurrences for the USA and 3808 for Europe.

Climate data

Because of their relevance for the biology of the species and their use in similar study (Guisan & Thuiller, 2005; Thuiller *et al.*, 2005; Broennimann *et al.*, 2007), we used the eight climatic predictors proposed by the climatic research unit (CRU; University of East Anglia, UK; http://www.cru.uea.ac.uk) at a resolution of 10'. The eight CRU 10' climatic maps were: ratio of actual to potential evapotranspiration (aet/pet), potential evapotranspiration (pet), annual amount of precipitations (prec), annual variation of precipitation (std_pre), minimum temperature of the coldest month (tmin), annual mean temperature (tmp), maximum temperature of the warmest month (tmax)and growing degree-days above 5°C. All these maps show a resolution of ca. 16 kilometers (10'). All these climatic conditions were sampled at each occurrence with the Spatial Analyst module of ArcMap™ 9.1.

GIS modeling

It has been shown that accuracy of current or future predictions can be very variable depending on the modeling technique used (Thuiller *et al.*, 2004; Elith et

al. 2006, Pearson *et al.*, 2006, Guisan et al., 2007) so that an "ensemble" approach combining several types of modeling techniques is recommended (Araújo & New, 2006). Moreover, it allows comparison with the study of Broennimann *et al.* (2007).

We used eight modeling techniques implemented in the latest version of the BIOMOD tool (Thuiller, 2003) for the version 2.4.0 of the R software (The R Project for Statistical Computing, 2007): artificial neural networks (ANN), boosted regression trees (BRT), classification tree analyses (CTA), generalized linear models (GLM), generalized additive models (GAM), multivariate adaptive regression splines (MARS), mixture discriminant analysis (MDA) and random forests (RF). GLM, GAM, CTA and ANN are described and discussed in the original BIOMOD paper (Thuiller, 2003). BRT and MARS were tested, together with GLM, GAM and CTA in a large study comparing 16 predictive techniques (Elith *et al.*, 2006), with BRT ranking best. MDA (Hastie & Tibshirani, 1996) and RF (Breiman, 2001) were also added as promising modeling methods (see Prasad et al. 2006).

Due to different sampling efforts, occurrences were not homogeneously distributed, and some aggregates may cause spatial autocorrelation, possibly biasing and over-fitting the models (Guisan & Thuiller 2005, Segurado *et al.* 2006). A re-sampling with a fixed minimal distance between occurrences was done. This minimal distance was a consensus based on a minimal correlation between environmental and geographical distance and a consistent number of occurrences (> 400) (Fig. 5). The PLANT county map for the distribution of *P. serotina* in USA was used to determine the absences in USA. Absences in Europe were determined by subtracting the presence map with a 50 kilometers buffer area to the country map for seven countries well covered by occurrences information: France, England, Switzerland, Germany, Belgium, Denmark, Lithuania, Sweden and Poland. Data were fractioned in two independent datasets to fit and evaluate the models: 70 % of the data were randomly allocated to a calibration dataset and the remaining 30 % were used as an evaluation dataset

(Thuiller 2003). The ratio between presences and absences (species' prevalence) was maintained in each partition. Three calibration approaches were done: (1) the invaded range was predicted from native occurrences only; (2) North American (NA) and European (EU) occurrences were pooled and used to predict the native range, keeping the prevalence between both areas; (3) a retro-projection was made from EU occurrences only to predict the NA native range. This latter approach was done to assess what *P. serotina* distribution would be in NA under EU climate indicating the transferability of the models (Fitzpatrick *et al.*, 2007)

The area under the curve (AUC) of a receiver-operating characteristics (ROC) plot was used to evaluate the prediction of the models. It compares the projection of the model with the locations of real occurrences. Following Swets (1988) scale a value of 1 means perfect prediction and a value of 0.5 means that models predictions is not different than random. Values between 0.5 and 0.7 can be interpreted as poor predictions, between 0.7 and 0.9 as fair to good predictions, and above 0.9 as excellent predictions. When the ROC-AUC value is lower than 0.5, it reflects a negative correlation between predictions and observations. The final evaluation was the sum of the eight AUC values, one for each modeling technique, so that the final predictions ranged along a scale from 1 to 8 (Fig.6 & Fig. 7).

For every calibration approach, the whole process was repeated 100 times, by sampling the data set 100 times. The mean of the 100 probabilities of occurrences was retained, summed with each model and projected in USA and Europe with the software ArcMap 9.1

Testing for climatic niche conservatism

Principal component analyses (PCA) were run in the R software to compare the position of occurrences from both the native and invaded ranges in the climatic space. Occurrences were weighted to ensure an equal representation of the two

ranges in the analyses. The magnitude and the significance of the niche shift between the two occurrence clouds in the PCA graph were assessed using a between-class analysis, yielding a between class inertia percentage (Dolédec & Chessel, 1987) with a 1000 permutations Monte Carlo randomization test (Romesburg, 1985). Haplotypes were projected in the same ecoclimatic space than the occurrences.

Results

Plastid DNA polymorphism in Prunus serotinas

The alignments of the sequences from the sampled individuals showed several insertions/deletions (indels) and substitutions. A total of six variable sites were found among the three part of the analyzed chloroplast genome (Table 2). On the *trnD-T* spacer, one indel and one T/A substitution were found at the 207th and 98th bases respectively. On *trnT-L*, two indels and one length ploymorphism were found at the 48th, 645th and 551st bases respectively. On *trnS-G*, one T/A substitution was found at the 496th base.

Table 2: haplotypes profiles. The size (in bp) of the complete amplified fragments (*TrnD-T indel*, *TrnT-L poly T*, *TrnT-L poly A* and *TrnT-L* indel) or of the restricted fragments (*Trn S-G res* and *Trn D-T res*) by the *Tas*I restriction enzyme.

Allele	**Profile**					
	Trn S-G res	*TrnD-T res*	*TrnD-T indel*	*TrnT-L poly T*	*TrnT-L poly A*	*TrnT-L indel*
H1	108	108	*77*	127	105	164
H2	108	86	*77*	127	105	164
H3	73	86	*77*	127	105	164
H4	108	86	*77*	127	106	164
H5	108	108	*77*	128	105	164
H6	66	73	*82*	132	106	126

Using the six developed ptDNA loci, six different haplotypes (named H1, H2, H3, H4, H5, H6) were found (profiles in Table 2, phylogenetic network in Fig. 4). Three haplotypes (H1, H2 and H3) are shared by both native and invasive populations, two haplotyes (H4 and H6) are specific to the native area and one

haplotype (H5) was only found in the invasive range (Appendices 1 & 2). Haplotype H1 represents 60.8% of the sampling, 76% of those from the invaded and 46.6% of those from the native range. Haplotype H2 represents 22.9% of the specimens, 5.6% of those from the invaded and 39.1 % of those from the native range. Haplotype H3 is found in 5.8% of the individuals, 10.4% of those from the invaded and 1.5% of those from the native range. European proportion of

haplotypes differs significantly from the American one (Chi-squared test, p-val < 0.001). Invaded range presents a lower number of alleles, haplotypes diversity, allelic richness and a higher genetic differentiation between populations but only the haplotype diversity and the allelic richness differences are significant ($p < 0.05$) (Table 3).

Table 3: number of haplotypes (N), haplotype diversity (Hs), allelic richness (Ar) and the level of genetic differentiation (Fst) between native and introduced ranges of *P.serotina*. Localities with only one (AZ, NM1, NM2) sample were not included in the calculation of Hs, Ar and Fst.

Group	**N**	**Hs**	**Na**	**Ar**	**Fst**
Native P. serotina	123	0.300	5	1.667	0.466
Introduced P.serotina	125	0.152*	4	1.320*	0.633

* Haplotype diversity and allelic richness in the introduced range of *P.serotina* is significantly lower (two-sided, $p<0.05$, 15000 permutations) than in the native one.

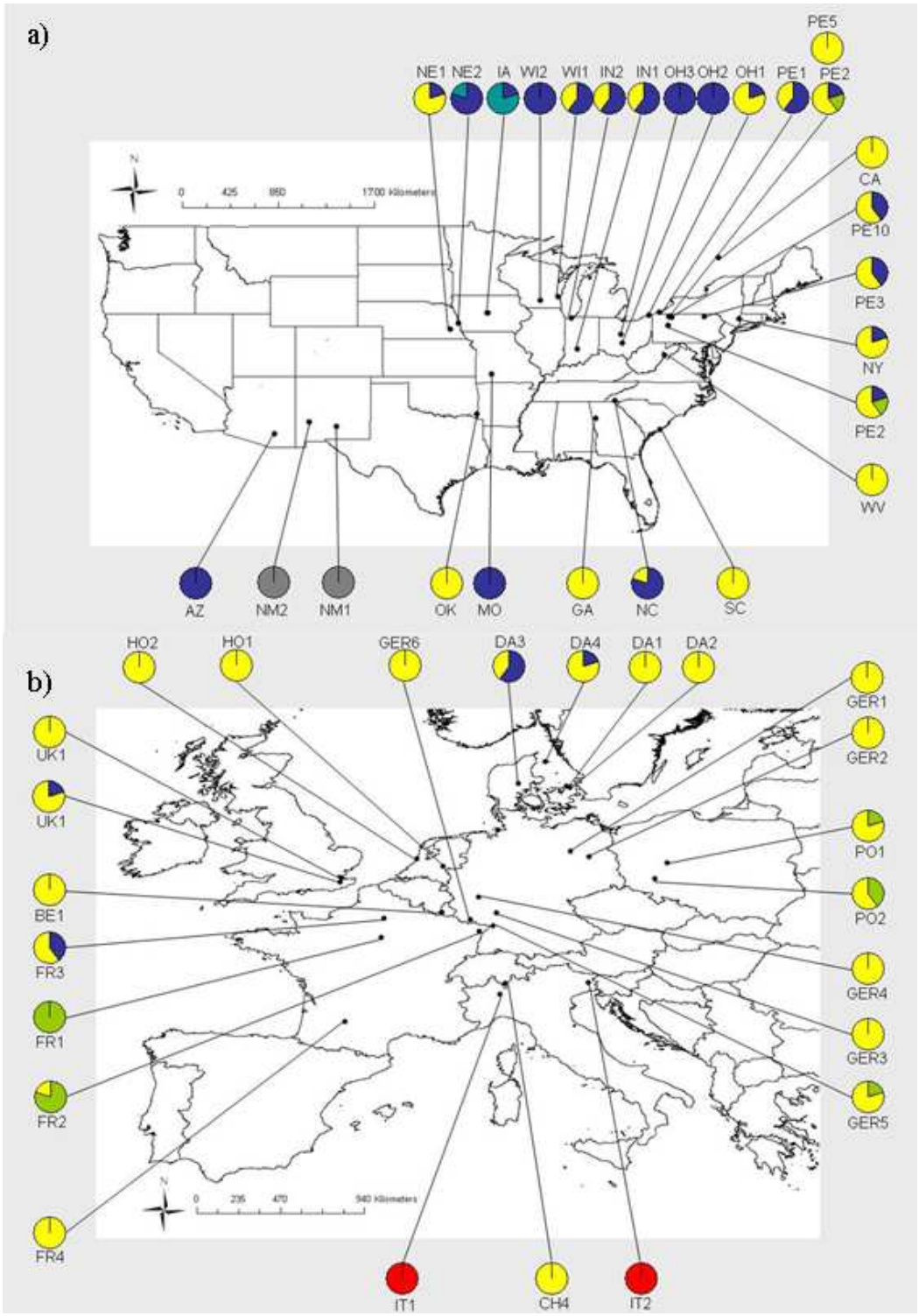


Fig. 3: Haplotypes maps. Every color represents a haplotype. Six different haplotypes were found: H1 (yellow), H2 (blue), H3 (green), H4 (turquoise), H5 (red), and h6 (grey). Three haplotypes (H1, H2, H3) are shared by both native and invasive area, two haplotyes (H4, H6) are specific to the native area and one haplotype (H5) was found only in the invasive range. Every population is composed by five individuals.

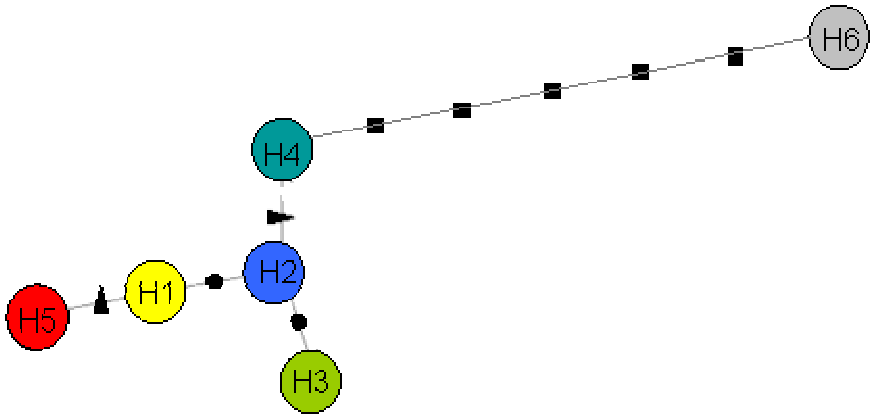


Fig. 4: median joining haplotypes network constructed using the *trnT-trnL*, *trnD-trnT* and *trnS-trnG* intergenic spacers. Indels are considered as only one evolutionary step. Black squares indicate insertion/deletion, black circles indicate A/T substitutions and triangles indicate polyT/A length polymorphisms.

Spatial Autocorrelation and subsampling for models calibration

When native occurrences are distant by a minimal distance of 0.32°, correlation between geographical distance and environmental distance matrixes is significantly reduced (Tukey HSD, pval < 0.001) from 0.5 to 0.37 +/- 0.006 (Fig. 5a). However, spatial autocorrelation increased when the minimal distance between occurrences increased in Europe. Subsamples with a 0.32° minimal distance, providing 428+/- 0.7 (after 100 resamplings) occurrences at each iteration of the modeling, were then used for calibration and evaluation of the models in NA but no subsampling was undertaken in EU because of the increasing spatial autocorrelation with increasing the minimal distance in Europe (Fig. 5b).

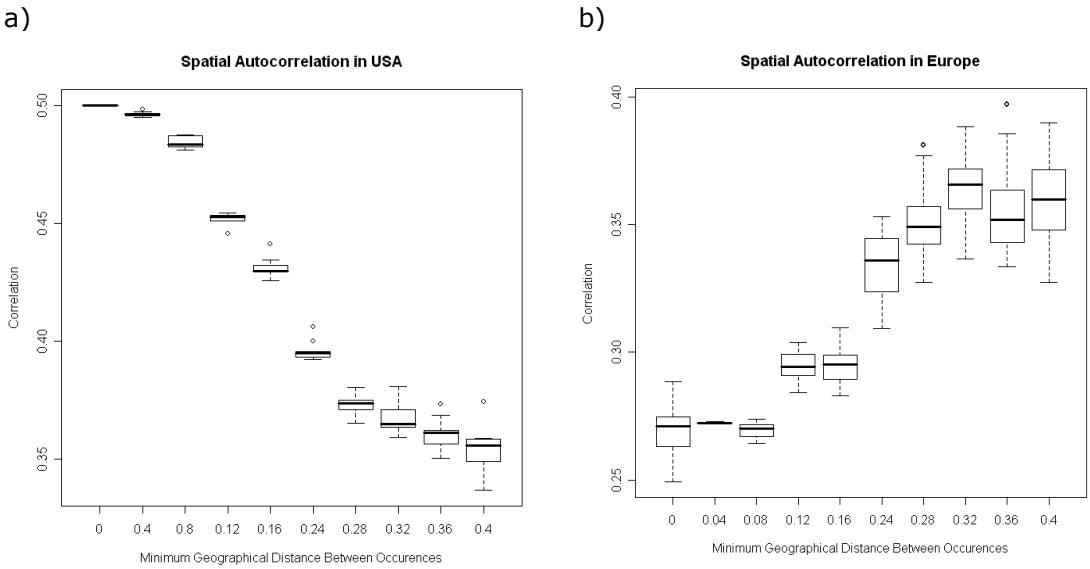


Fig. 5: Correlation between geographical and environmental distances in USA (a) and in Europe (b). On sampling with a minimal 0.32° geographical distance, correlation is significantly reduced (Bartlett test, p-val <0.01)from 0.5 (Mantel-test, 20 perm., p-val<0.05) to 0.37 +/- 0.006 (Mantel-test, 20 permutations, p-val<0.05). As it was impossible to analyze more than 3000 occurences in with a mantel test Europe, a 3000 random sampling was done on the European dataset when there were more than 3000 occurrences (Total 3808), explaining the variation of the first boxplot. Mean of autocorrelation if dist = 0 : 0.26922 +/- 0.01 (Mantel-test, 20 permutations, p-val<0.05).

Prediction of the species distribution

Prediction accuracy of the models calibrated in the native range was high when projected on the native area (AUC = 0.92 +/- 0.03), but low when projected on the invaded range (AUC = 0.55 +/- 0.17) (Fig. 6a & 7a). Calibrating models with data from both ranges dramatically increased AUC on the invaded range (AUC = 0.93 +/-0.04), while keeping as well a high predictive ability in the native range (AUC = 0.89 +/- 0.06) (Fig. 6b & 7b). When models are calibrated with invasive data, prediction accuracy is excellent in Europe (AUC = 0.97 +/- 0.03) but poor in USA (AUC = 0.60 +/- 0.14) (See Appendix 3 & Fig. 7c).

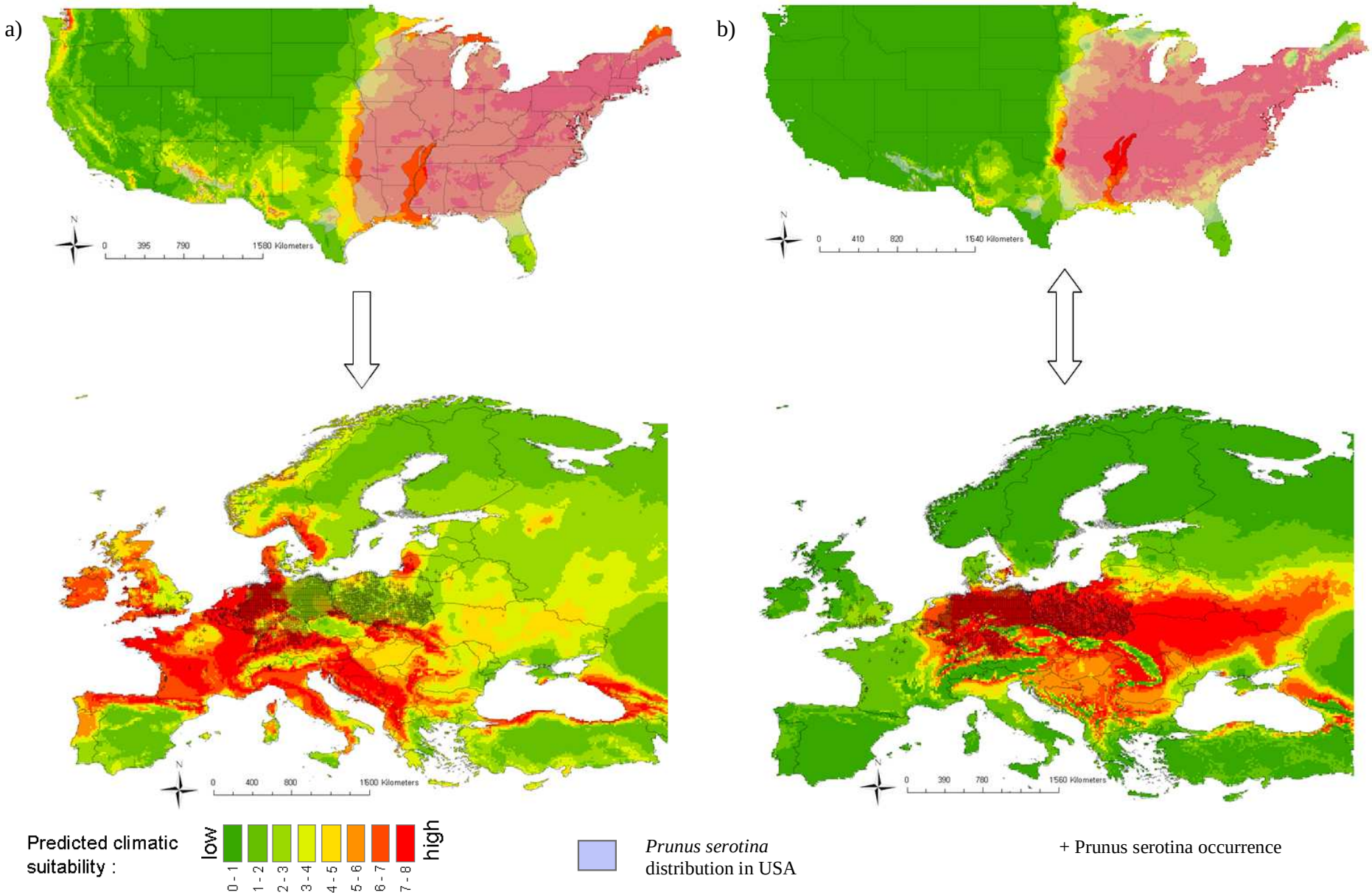


Fig. 6: Prediction maps. The maps illustrate the predicted climatic suitability (mean number of models, among eight modelling techniques, predicting the species present) obtained from models calibrated in USA (a) and through both ranges, Europe and USA (b). *P serotina* distribution in USA (Little, 1971) is represented by the blue layer. Occurences were used to illustrate the partial distritubion of *P. serotina* in Europe.

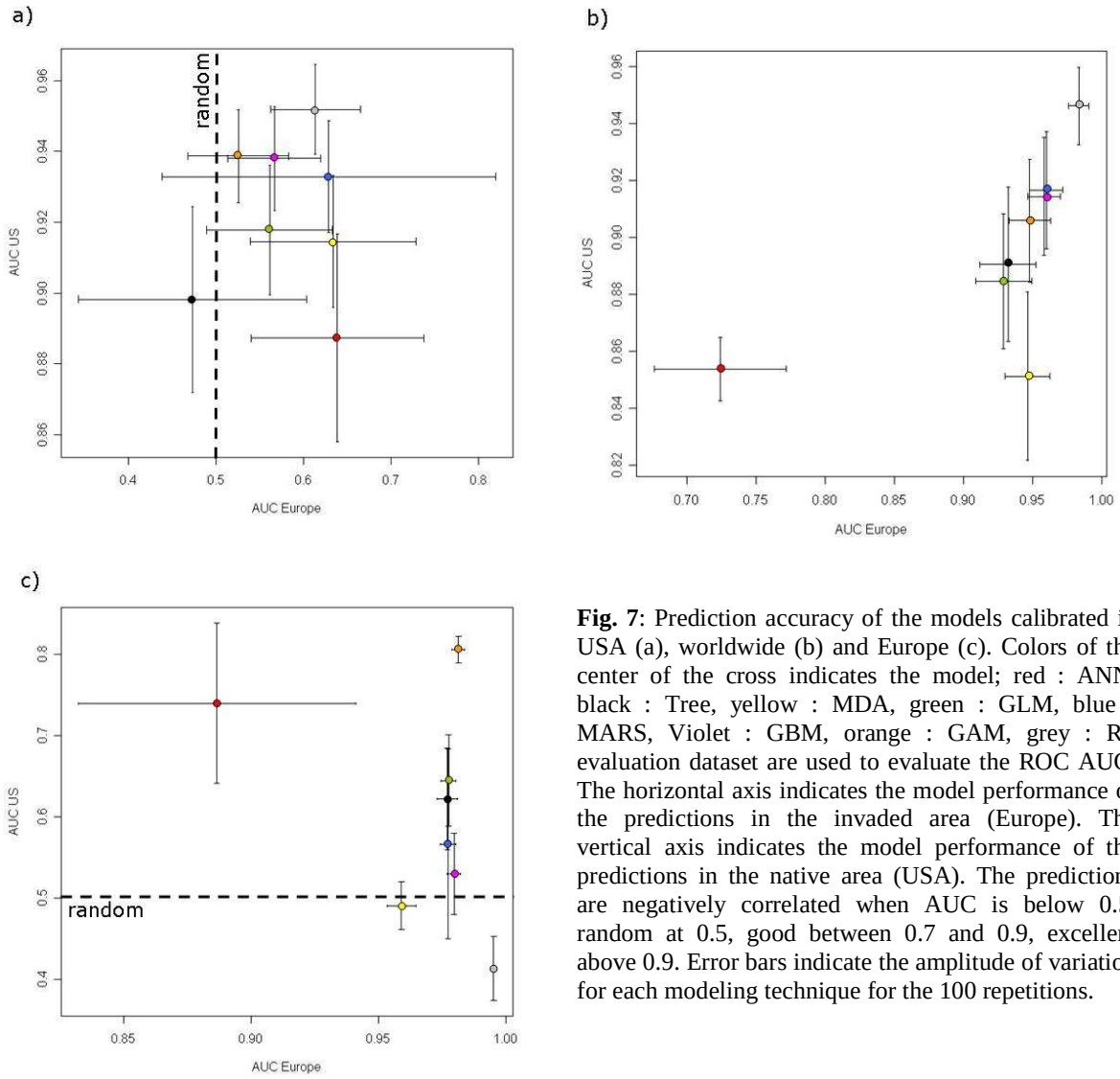


Fig. 7: Prediction accuracy of the models calibrated in USA (a), worldwide (b) and Europe (c). Colors of the center of the cross indicates the model; red : ANN, black : Tree, yellow : MDA, green : GLM, blue : MARS, Violet : GBM, orange : GAM, grey : RF evaluation dataset are used to evaluate the ROC AUC. The horizontal axis indicates the model performance of the predictions in the invaded area (Europe). The vertical axis indicates the model performance of the predictions in the native area (USA). The predictions are negatively correlated when AUC is below 0.5, random at 0.5, good between 0.7 and 0.9, excellent above 0.9. Error bars indicate the amplitude of variation for each modeling technique for the 100 repetitions.

ACP

Principal component analyses of the pooled climatic data revealed two significant axes of climatic variation, defining a realized climate space of reduced dimensionality which allows the investigation of niche conservatism (Fig. 8 & Fig. 9). The enclosed correlation circle (Fig. 8 & Fig. 9) indicates the relative contributions of climatic predictor variables to axis 1 and 2. The two axes are associated closely with water availability and heat energy, respectively. Examination of the position of the species in climate space reveals that niche

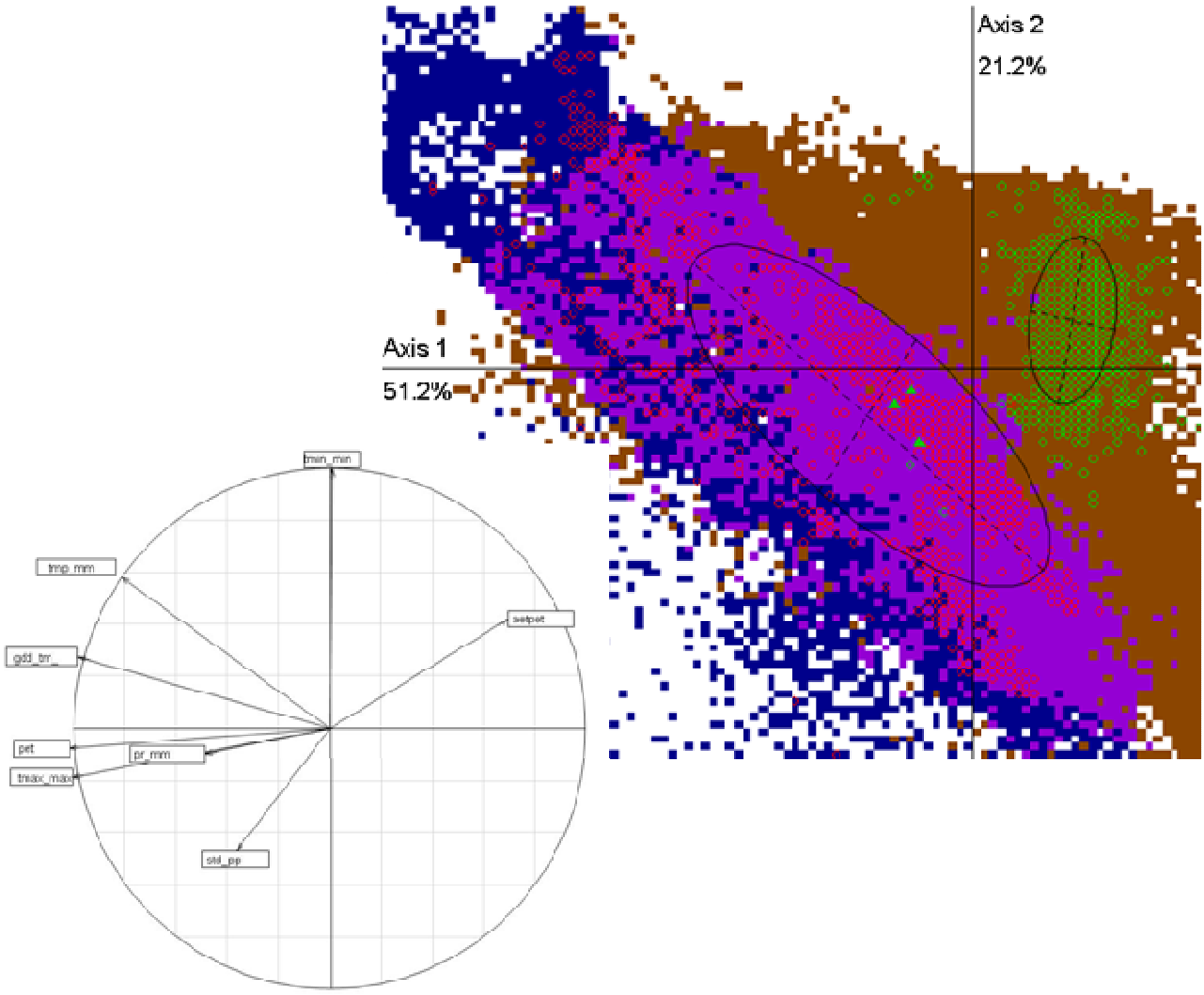


Fig. 8: Bioclimatic space with illustration of niche shift. The position of occurrences, from the native and invaded ranges along the principal climatic gradients is indicated with red and green circles respectively. Green triangles represent EU localities in the native eco-climatic niche. The background represents continental climates. Blue pixels are exclusive NA climates, brown pixels are exclusive EU climates and violet pixels are climate shared by both ranges. The enclosed correlation circle indicates the importance of each bioclimatic variable on the two significant axes of the principal component analysis, which jointly explain 72.4 % of the variance in the data. Climatic predictors are: tmp = annual mean temperature, tmax = maximum temperature of the warmest month, tmin = minimum temperature of the coldest month, prec = annual sum of precipitation, std_prec = annual variation of precipitation, gdd = annual growing-degree days above 5 °C, aet/pet = ratio of actual potential evapotranspiration& pet = annual potential evapotranspiration.

centroids differ strongly between the native and introduced ranges, in spite of extensive overlap of European and western North American climates (Fig. 8). Sampled localities fit very well the observed occurrences in the ecoclimatic space, supporting the good representation of *P. serotina* distribution of our collection. The shift occurs for the observed occurrences (between group inertia: 36.7 %,

p<0.001) as well as for the sampled populations (32.5 %, p<0.001). Fig. 9 shows that the shift occurs for every haplotype.

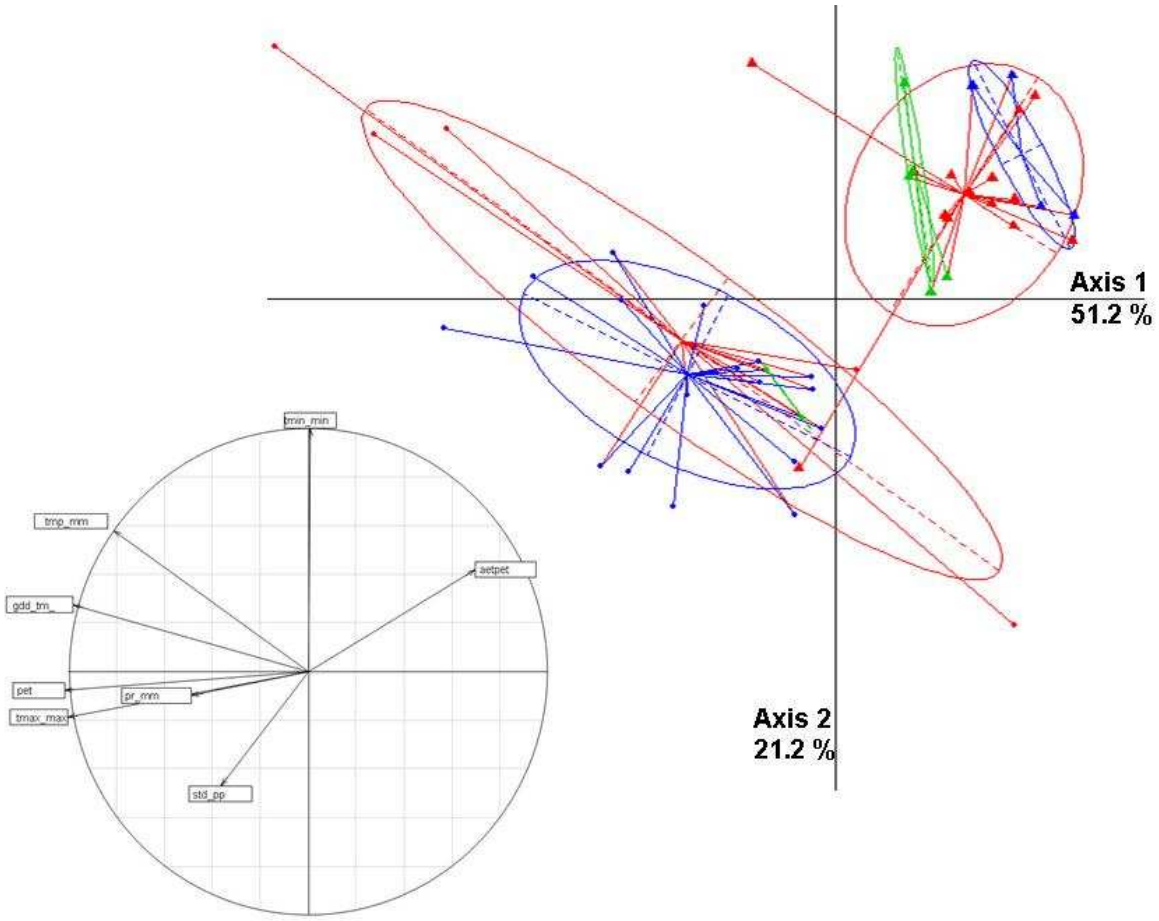


Fig. 9: Principal Components Analysis (ACP) on the haplotypes. European and American populations are indicated with triangles and points respectively. Colors indicates haplotypes : red = H1, blue = H2, green = H3. The climatic niche shift occurs for every haplotypes.

Discussion

Models efficiency and niche conservatism

The very high accuracy of the models predictions in the evaluation dataset of black cherry native range shows that the species responds very well to the eco-climatic factors, which support climate as a primary driver of its distribution. The low level of prediction accuracy for the projection of models calibrated in the

native range and projected in the USA is hypothesized to result from the shift of the realized climatic niche, as observed in the PCA climatic space. This observed niche shift violates one of the pivotal assumptions for the use of ENM in predicting biological invasions, which is the conservatism of the niche in the invaded range. This supports previous evidences of niche observed in *Centaurea maculosa* (Broennimann *et al.*, 2007) and questions most studies using ENM to predict the future extent of invasive species introduced in a new area (Peterson *et al.*, 1999; Peterson & Viéglais, 2001; Iguchi *et al.*, 2004; Thuiller *et al.*, 2005; see Pearman et al. 2008). None of these studies investigated whether such a shift actually occurred in the climatic space (as defined from real observations, not predictions), so that is impossible to assess at this stage whether no shift indeed occurred with these species, or whether the predictions in the new range were not evaluated appropriately.

In the case of *P. serotina*, the frequent human plantations in its invaded range may have biased the quantification of its niche, which may thus not represent the natural realized niche. However, we do not think this is the case here because the same approach, applied to a more accidentally introduced and more naturally disseminated species, yielded the same results (Broennimann *et al.*, 2007). Generalizing the PCA approach to every species already investigated by ENM may help discerning whether such niche shift between native and invaded ranges is a general phenomenon associated with biological invasions.

Nevertheless, even if a niche shift does occur, ENMs remain useful for anticipating biological invasions, by still being able to predict areas most susceptible to introduction (Broennimann *et al.*, 2007). The fact that areas of introductions needs to be within the realized niche of the species in its native range strongly suggests that mechanisms promoting a niche shift must have occurred during the lag phase (e.g. by founder effect followed by genetic drift and/or natural selection, and/or hybridization). In the present study, many of the areas where *P. serotina* poses important problems are robustly predicted. High

prediction accuracy is found in Benelux, Denmark, Western Germany, Lithuania or the Southern Alps, where *P. serotina* has proliferated. The models also predicted the species in Norway and Sweden, where *P. serotina* naturalized successfully (Starfinger, 2007). Interestingly, the only occurrences in Europe corresponding to the native niche in the USA are all located in Ticino, Switzerland. On the other hand, the models failed in predicting the eastern part of the invaded range, and over predicted the species in the western, oceanic-temperate range (France, England and Spain). The fact that the southern part of the native distribution of *P. serotina* in Mexico was not taken in account in the model should not biased its accuracy, because *P. serotina* is very rare or absent in the Mediterranean part of Europe having a climate similar to Mexico. Differences in the biotic environment between the native and introduced range have been documented for *P. serotina* in its introduced range, and may explain the success of this species in its new habitats, by supporting the enemy release hypothesis (Reinhart *et al.*, 2003).

Our results further suggest that the best solution for accurately predicting the spatial distribution of *P. serotina* is neither to calibrating the models in the native range only (like in Peterson & Vieglais, 2001; Peterson, 2003; Thuiller *et al.*, 2005), nor in the invaded range only (like in Mau-Crimmins *et al.,* 2006; Zhu *et al.*, 2007), but to calibrate the models with data from both ranges simulateously. In the present study, this approach of fitting a single model with data from the two ranges allowed obtaining a better consensus, which means a model predicting fairly in the two ranges (the AUC here proved excellent when tested with both native and invasive ranges).

Several hypotheses may explain such improvement of calibrating models in both ranges. First, models may this way fit what one could call a "metaniche", possibly including differing niches between American and European ecotypes. Indeed, the full realized niche of a species across ranges may result from the pooling of several subniches of local populations or ecotypes in the two ranges, exactly as Brown et al. (1996) described for explaining the geographic ranges of a

species as a sum of many ecotypes ranges nested within the overall species range. Investigating the ecological relevance of the predictor used in the model may increase the transferability of the models. If the models over fit the distribution, the transferability may be questioned. The more evident example is the AUC difference within the RF model, excellent when evaluated in the calibrated range but very poor when evaluated in the reciprocal range (Fig. 7). This supports the use of a very large panel of models to check if there is a global tendency in the accuracy of prediction and this way avoid an eventual artefactual explanation. Models performances vary a lot between models projected in their reciprocal range (i.e. model fitted in EU projected in USA, and vice versa), but as expected all models reach highly performance when they are calibrated in both ranges and evaluated in both ranges. The fair performances of the models when calibrated in both ranges may also suggest that black cherry simply may not have reached environmental equilibrium state in its native range. Fig. 8 shows many introduced occurrences in a non-existent climate in USA. *P. serotina* may have a broader climatic tolerance than the climates panel in its native range, but geographical barriers prevent the species to colonize such niches.

Considering the prediction map established with its worldwide distribution, all Eastern Europe shows a great potential to be invaded by black cherry. The last primeval European lowland and deciduous forests are all situated in this part of the continent (e.g. Bialowieski forest in Poland, Strandzha forest in Bulgaria). They represent a huge refuge for the European diversity (Yurukov & Zhelev, 2001; Grodzinska *et al.*, 2002). In NA *P. serotina,* may be problematic outside its distribution. Both European and American calibrated models (Fig. 6a & Appendix 3) are congruent in predicting the State of Washington as a good habitat. Foresters in these area may be warned of the potential threat. Even if its distribution may be mainly due to human plantations, *P. serotina* shows a great success once established in its new range and predicting the potential habitats where it shall not be planted remains very useful to avoid biodiversity threats.

Phylogeography

The three haplotypes common to both native and invaded ranges are all present in Pennsylvania, supporting the assumption that cultivated timber wood from this area was the source of introduced individuals. The very low presences of haplotype H3 in the native populations (two individuals) are only found in Pennsylvania. This suggests that increasing the number of samples in each population, but especially in the Pennsylvanian may allow finding haplotype H5, only present in the invasive range. If haplotype H5 would be absent from our sampled populations, it is not probable that its origin would be from the Mexican distribution. Haplotypes network (Fig. 4) shows that H5 is very closer from H1 than H6. Furthermore, H6 is very distant from the rest of every haplotype. This genetic distance suggests that H6 may be taxonomically moved away from the other haplotypes, which supports the belonging of these populations to the *rufula* variant which distribution fits with these localities (PLANT, 2007).

When compared between both ranges, genetic diversity and allelic richness differences and genetic differentiation trend support the assumption of the loss of diversity due to a founder effect. The fact that genetic differentiation and allelic richness differences are unpronounced between native and invasive populations may be the results of multiples introductions. Ad-mixing populations from differentiated sources may transform the variation among groups in variation within groups (Kolbe *et al.*, 2004; Genton *et al.*, 2005; Frankham, 2005). The distribution of the haplotypes in Europe suggests at least one source providing the Italian populations, one source providing FR1, FR4 and both Polish populations, one source for GB1, DA3, DA4 and FR1 and another origin for the other populations. The pattern also supports that human is the main factor in the dispersion of black cherry. Population at Southern Alps (IT1, IT2, and CH4) illustrates this phenomenon fairly well. As two distant populations display the same and rare haplotype (H5) below such a geographical barrier as the Alps, a third one, nearer but situated in another country presents a widely common

haplotype (H1). This reflects how important the forester policy in the invaded distribution of *P.serotina* was. Both Italian populations probably resulted from the mid-twenty century Italian policy of massive plantation of black cherry in forests and wastelands (Sartotti, 1988). A large number of seeds from the same source may have been broadly introduced. This case figure illustrates the two different types of introduced populations. On one side, massively planted specimens were introduced for forestry keeping a high diversity when admixed with specimen from different sources and on the other side escaped individuals (from horticultural use, natural seed dispersing) founded bottlenecked populations. Only nuclear microsatellites may confirm this assumption.

The low quantity of variations found in the chloroplast sequences does not provide enough information to produce robust analyses on the genetic variation and to provide a fine resolution to detect the origin of the invader. The allotetraploidy of black cherry (Pairon & Jacquemart, 2005) suggests *P. serotina* may be formed by a recent hybridization between two ancestral lineages, which explains the low genetic variation rate on the chloroplast sequences. On the other hand, chloroplast analyzes results are very informative when compared to nuclear microsatellite analyzes performed on the invasive range (Pairon, 2007) and the native range (Pairon, unpublished). In the invasive range, it allows confirming the main differentiated clusters (FR1, FR4 and the Italian one) and assuring differentiation when clustering was unclear (the polish populations). Additionally, it allows discerning differentiated populations in Denmark. In the native range, chloroplast and nuclear analyzes are congruent about the spot of allele richness in Pennsylvania. In the contrary of cpDNA analyses, every nuclear allele present in the invaded range was found in the native range, supporting the assumption that the Italian haplotype may have missed in our sub sampling. On the other side, nuclear analyses do not found enough genetic structure in the native range to discern clusters of populations. The maternally inherited DNA allows detecting H4 confined in the western range of the native distribution and H2 populations more

present at the western side of the Appalachian. Differences in cpDNA and nuclear DNA structure is not an exception (Petit, 2005) and show how complementary are these both approaches.

Phylogeography and niche conservatism

No haplotype present a closer native niche from its invasive niche (Fig 8). The shift occurs for every shared haplotypes. This is a different result than in Mau-Crimmins *et al.* (2006) where the invasive range corresponded more to the sub niche of the introduced variant. The very broad distribution of H1 in the invasive range may suggest that H1 would be more adapted to the European climate. Actually, H1 shows the broader niche in its both ranges, but the shift is the same as for the other haplotypes. The assumption emerging from this result is that H1 is not better adapted to its invaded range but presents more plasticity in its adaptation in general. At the western limit of its distribution in Nebraska, *P. serotina* has expanded its range in recent year becoming invasive in weedy places and forest margins (R. B. Kaul, personal communications). Locality NE1 is one of these problematic populations, as NE2 is situated in a forest where specimens are certainly from native plants (Robert B. Kaul, personal communications). So it wouldn't be a fortune if haplotype H1 is widely represented in population NE1, which displays a very different profile than NE2.

Conclusion

Biological invasions are very complex phenomenon, involving evolutionary, ecological and sociological component. The present study focuses on one problematic species, *P. serotina*, and integrates evolutionary and ecological approaches to better understand and prevent biological invasions. This study shows that i) many introductions from several native populations (probably Pennsylvanian) of black cherry occurred in Europe, ii) a climatic shift occurred

between the native and the invasive distribution affecting every introduced haplotypes and finally iii) despite the limitations imposed by the non-conservatism of the niche, invasive range of *P. serotina* remains predictable by ENM.

Acknowledgements

I firstly wish to thanks Antoine Guisan, Olivier Broennimann and Guillaume Besnard who have supported this project. Many thanks to Robin Engler who georeferrenced ton of occurrences. We are grateful to The National Centers of Competence in Research (NCCR) for their financial contribution to the project. This research was also made possible through a grant to K.O. Reinhart from the National Parks Ecological Research Fellowship Program, a partnership between the National Park Service, the Ecological Society of America and the National Park Foundation. It is funded through a generous grant from the Andrew W. Mellon Foundation.

I wish to thanks every one who collected the leaf samples without them this study would not be possible: Gretchen Meyer (University of Wisconsin-Milwaukee Field Station), Amy K. Buthod (Robert Bebb Herbarium, The University of Oklahoma), Deborah A. Lewis (Ada Hayden Herbarium, Iowa State University), Louis Iverson (Northern Research Station, USDA Forest Service), Robert B. Kaul and David M. Sutherland (Bessey Herbarium University of Nebraska), Theodore S. Cochrane (University of Wisconsin-Madison Herbarium), Kurt O. Reinhart (United State Department Of Agriculture, Agricultural Research Service), Robert A. Klips (Ohio State University), Donovan Bailey and Patrick J. Alexander (New Mexico State University), Phil Jenkins and Joan Tedford (Herbarium ARIZ, The University of Arizona), Laura M. Bowe (Ozarks Regional Herbarium, Missouri State University), David Morgan (University of West Georgia) for the US. Samples from France were provided by Pierre Geldreich, Jean-Marie Ansolabehere, Rémi Rodriguez and Jerôme Jaminon (Office National des Forêts), and Guillaume

Decocq (University of Picardie Jules Verne); from Germany by Thilo Heinken (University of Potsdam), Uwe Starfinger (Technische University of Berlin) and by Jörn Meyer (Forestrevier Altenwalde); from The Netherlands by Wim van der Putten (NIOO-KNAW) and by Antje Ehrenburg (Amsterdam Water Supply); from Italy by Fulvio Caronni and Lisa Hildebrand (Parco Lombardo della Valle del Ticino) and Simone Del Fabbrio; from Denmark by reik Kjaer (Royal Veterinary and Agricutltural University), by Niels Peter Revsbech (University of Aarhus) and Olivier Raspé (Jardin botanique national de Belgique); from Poland by Boleslaw Suszka (Polish academy of sciences) and Aleksandra Halarevicz (University of Agriculture of Wroclaw); from Belgium by Marie Pairon (University Catholique de Louvain-La-Neuve).

Many thanks to G. Besnard for his assistance in the lab, to O. Broennimann for his useful knowledge in R and to Marie Pairon for her sampling and her knowledge about *P. serotina*. Comments on the previous version of the manuscript were contributed by G. Besnard, O. Broennimann and Antoine Guisan. At last, many thanks to all my family and my close friends who support me during this work, especially to Nadia Chennaf who helps me collecting in Ticino.

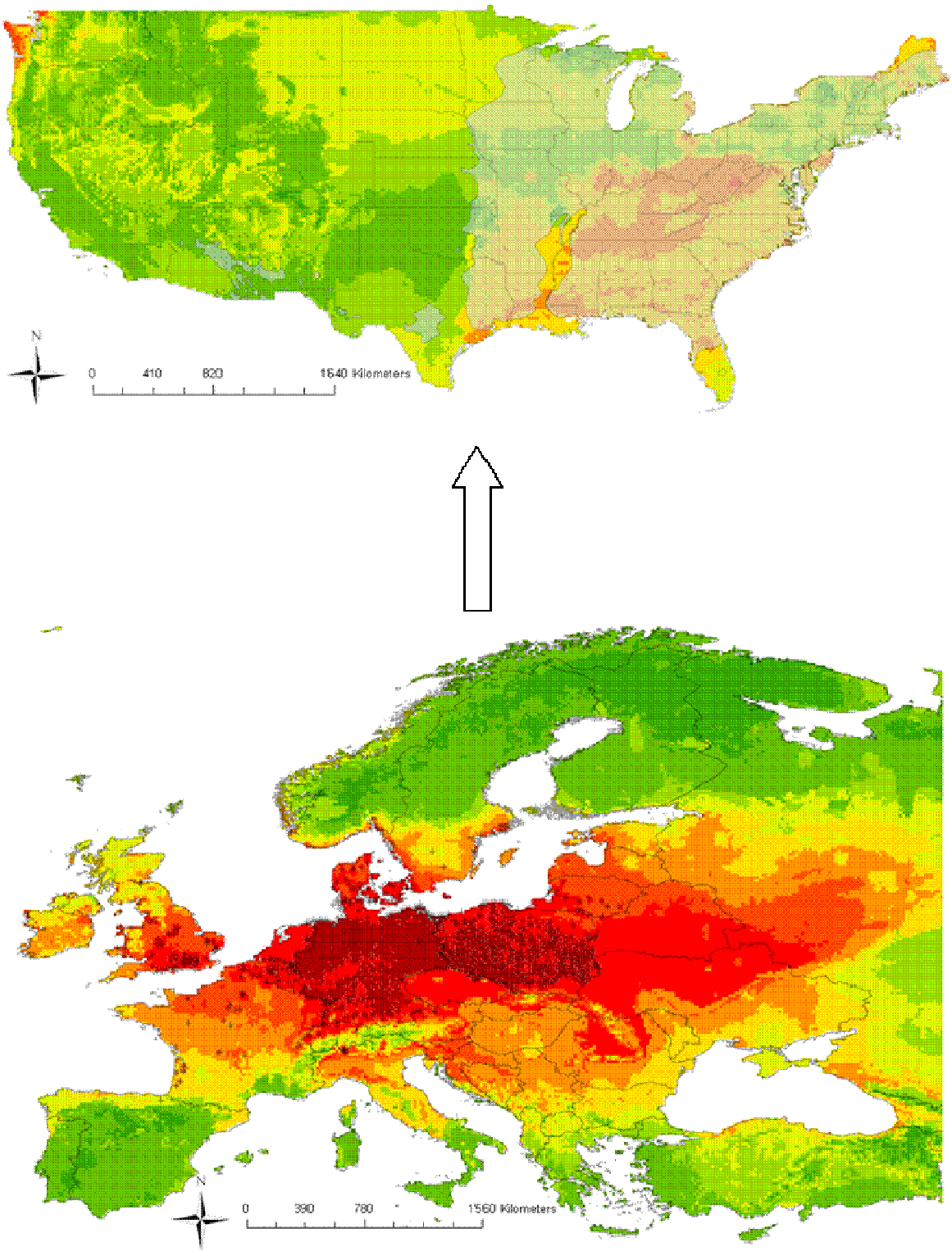
Appendices

Appendix 1: Sampling in the native range. N is the number of samples for each locality, N_A the number of allele found in each locality and H_S the haplotype diversity in each locality. Each of this value is calculated for the whole range. Localities where only one individual was sampled were not taken in account in the calculation fo N_A and H_S

	ID	Location	State	Country	Latitude		Longitude		N	Haplotypes	H_S
Native Range											
	AZ	Tanque Verde	Arizona	USA	32.37°	N	-110.69°	W	1 (-)	H2	0
	CAN	Guénette	Quebec	Canada	46.30°	N	-75.25°	W	5	H1	0
	GA	Carrolton	Georgia	USA	33.65°	N	-85.03°	W	5	H1	0
	IN1	Michigan City	Indiana	USA	41.67°	N	-86.98°	W	5	H1, H2	0.6
	IN2	Bloomington	Indiana	USA	39.20°	N	-86.51°	W	5	H1, H2	0.6
	IO1	Ames	Iowa	USA	42.04°	N	-93.67°	W	5	H2, H4	0.4
	MO1	Springfield	Missouri	USA	37.17°	N	-93.33°	W	5	H2	0
	NC1	Otto	North Carolina	USA	35.03°	N	-83.48°	W	5	H1, H2	0.4
	NE1	Lincoln	Nebraska	USA	40.80°	N	-96.64°	W	5	H1, H2	0.4
	NE2	Omaha	Nebraska	USA	41.25°	N	-96.00°	W	5	H2, H4	0.4
	NM1	Sierra County	New Mexico	USA	33.35°	N	-107.89°	W	1 (-)	H6	0
	NM2	Cloud Croft	New Mexico	USA	32.96°	N	-105.75°	W	1 (-)	H6	0
	OH1	Ashtabula	Ohio	USA	41.89°	N	-80.79°	W	5	H1, H2	0.4
	OH2	Delaware	Ohio	USA	40.37°	N	-83.05°	W	5	H2	0
	OH3	Achville	Ohio	USA	39.67°	N	-82.93°	W	5	H2	0
	OK1	Broken Bow	Oklahoma	USA	33.97°	N	-94.54°	W	5	H1	0
	PEN1	Behrend	Pennsylvania	USA	42.07°	N	-79.98°	W	5	H1, H2	0.6
	PEN2	Buckaloons	Pennsylvania	USA	41.08°	N	-79.25°	W	5	H1, H2, H3	0.7
	PEN3	Burlington	Pennsylvania	USA	41.77°	N	-76.38°	W	5	H1,H2	0.6
	PEN4	Chapman Dam	Pennsylvania	USA	41.75°	N	-79.17°	W	5	H1, H2, H3	0.7
	PEN5	Warren	Pennsylvania	USA	41.76°	N	-79.19°	W	5	H1	0
	PEN7	Pawling	New York	USA	41.57°	N	-73.60°	W	5	H1, H2	0.4
	PEN10	Tionesta	Pennsylvania	USA	41.70°	N	-78.93°	W	5	H1, H2	0.6
	SC	Charleston	South Carolina	USA	32.75°	N	-79.91°	W	5	H1	0
	WI1	Saukville	Wisconsin	USA	43.39°	N	-88.02°	W	5	H1	0
	WI2	Madison	Wisconsin	USA	43.04°	N	-89.44°	W	5	H1, H2	0.4
	WV1	Spruce Knob	West Virginia	USA	38.72°	N	-79.54°	W	5	H1	0
Total									**123 (120)**	**H1, H2, H3, H4, H6**	**0.300**

Appendix 2: Sampling in the invasive range. N is the number of samples for each locality, N_A the number of allele found in each locality and H_S the haplotype diversity in each locality. Each of this value is calculated for the whole range. Localities where only one individual was sampled were not taken in account in the calculation fo N_A and H_S

	ID	Location	State	Country	Latitude		Longitude		N	Haplotypes	H_S
Invasive range											
	BE1	Lagland		Belgium	49.40°	N	5.46°	W	5	H1	0
	CH4	Castelrotto		Switzerland	46.06°	N	9.02°	W	5	H1	0
	DEN1	Copenhagen		Dennmark	55.95°	N	12.27°	W	5	H1, H2	0.6
	DEN2	Tisviled		Dennmark	56.05°	N	12.08°	W	5	H1	0
	DEN3	Aarhus		Dennmark	56.18°	N	9.05°	W	5	H1	0
	DEN4	Laeso		Dennmark	57.28°	N	11.00°	W	5	H1, H2	0.4
	FR1	Fontainebleau		France	48.38°	N	2.75°	W	5	H3	0
	FR2	Alsace		France	7.68°	N	48.72°	W	5	H1, H3	0.4
	FR3	Compiègne		France	2.88°	N	49.37°	W	5	H1, H2	0.6
	FR4	Bordeaux		France	0.90°	N	44.13°	W	5	H1	0
	GER1	PotsdamS		Germany	12.30°	N	52.73°	W	5	H1	0
	GER2	PotsdamT		Germany	13.22°	N	52.47°	W	5	H1	0
	GER3	Neuwig		Germany	7.63°	N	50.43°	W	5	H1	0
	GER4	MannheimN		Germany	8.53°	N	49.65°	W	5	H1	0
	GER5	Karlsruhe		Germany	8.37°	N	48.97°	W	5	H1	0.4
	GER6	Saarbrucken		Germany	7.23°	N	49.30°	W	5	H1, H3	0
	GER7	Cuxhaven		Germany	8.63°	N	53.82°	W	5	H1	0
	HO1	Arnhem		Netherlands	5.83°	N	52.00°	W	5	H1	0
	HO2	Amsterdam		Netherlands	4.55°	N	52.35°	W	5	H1	0
	IT1	Ticino		Italy	8.72°	N	45.52°	W	5	H5	0
	IT2	Udine		Italy	13.17°	N	46.12°	W	5	H5	0
	PO1	Poznan		Poland	17.15°	N	52.17°	W	5	H1, H3	0.4
	PO2	Wroclaw		Poland	16.55°	N	51.37°	W	5	H1, H3	0.6
	UK1	Shakelford		United Kingdom	0.65°	N	51.18°	W	5	H1	0
	UK2	Bagshot		United Kingdom	0.73°	N	51.38°	W	5	H1, H2	0.4
Total									**125**	**H1, H2, H3, H5**	**0.152**



Appendix 3: Retro-projection in the native range. Invaded range is used to calibrate prediction models for the native distribution. This approach allows checking for climatic shift. Legend is the same as in Fig. 6

References

Amsellem, L., Noyer J.L., Le Bourgeois, T. & Hossaert-Mckey, M. 2000. Comparison of genetic diversity of the invasive weed *Rubus alceifolius* Poir. (Rosaceae) in its native range and in areas of introduction, using amplified fragment length polymorphism (AFLP) markers. *Molecular Ecology* 9, 443-455.

Araujo, M. B. & New, M. 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution* 22, 42–47.

Bandelt, H.J, Forster, P. & Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, 37-48.

Besnard G., Henry P., Wille L., Cooke D. & Chapuis E. 2007. On the origin of the invasive olives (*Olea europaea* L., Oleaceae). *Heredity* 99, 608-619.

Blossey, B. & Notzold, R. 1995. Evolution of increased competitive ability in invasive nonindigenous plants – a hypothesis. *Journal of Ecology* 83, 887-889.

Branquart, E., Vanderhoeven, S., Van Landuyt, W., Van Rossum, F. & Verloove F. 2007. Alert, black and watch lists of invasive species in Belgium. Harmonia version 1.2, Belgian Forum on Invasive species, accessed on 12/17/2007 from: http://ias.biodiversity.be

Breiman, L. 2001. Random forests. *Machine Learning* 45, 5-32.

Broennimann, O., Treier, U. A., Müller-Schärer, H., Thuiller, W., Peterson, A. T. & Guisan, A. 2007. Evidence of climatic niche shift during biological invasion. *Ecology Letters* 10, 701-709

Brown, J.H., Stevens, G.C. & Kaufman, D.M. 1996. The geographic range: size, shape, boundaries, and internal structure. *Annual Review Ecological Systematics*, 27, 597– 623.

Byers, J. E., Reichard, S., Randall, J. M., Parker, I. M., Smith, C. S., Lonsdale, W. M., Atkinson, I. A. E., Seastedt, T. R., Williamson, M., Chornesky, E. & Hayes, D., 2002. Directing research to reduce the impacts of nonindigeneous species. *Conservation Biology* 16, 630-640.

CPS-SKEW. 2006. Fact sheet: Cerisier Tardif, accessed on 12/17/2007 from http://www.cps-skew.ch/francais/inva_prun_ser_f.pdf

Dietz, H. & Edwards, P. J. 2006. Recognition that causal processes change during plant invasion helps explain conflicts in evidence. *Ecology* 87, 1359-1367.

Demesure, B., Sodzi, N. & Petit, R. J. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4, 129-131.

Dolédec, S. & Chessel, D. 1987. Rythmes saisonniers et composantes stationnelles en milieu aquatique I – Description d'un plan d'observations complet par projection de variables. *Acta Oecologica. Oecologia Generalis* 8, 403-426.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A. *et al.* 2006. Novel methods improve prediction of species distributions from occurrence data. *Ecography*, 29, 129-151.

Ellenberg, H. 1953. Physiologisches und ökologisches Verhalten derselben Pflanzenarten. *Berliner Deutscher botannik Gesellschaft* 65, 351-362.

EPPO. 2007. EPPO List of invasive alien plants accessed on 12/17/2007 from http://www.eppo.org/QUARANTINE/ias_plants.htm

Fitzpatrick M. C., Weltzin J. F, Sanders N. J. & Dunn R. R. 2007. Why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography,* 16, 24-33.

Fontaneto, D., Gomarasca, S., Castiglione, S., Sartori, F., Vietto, L., Vendramin, G. G. & Caronni F. 2004. Strategie per il contenimento di una specie alloctona invasiva del parco del Ticino : « Il Prugnolo Tardivo ». Convegno Internazionale Il sistema rurale Una sfida per la prgttazione tra salvaguardia, sosteibilità e governo delle trasformazioni. Accessed on 12/17/2007 from http://www.cedat.polimi.it/convegno/en/doc/interventi_pdf/1_99_Castiglione.pdf

Frankham, R. 2005. Invasion biology: Resolving the genetic paradox in invasive species. *Heredity* 94, 385.

Genton, B. J., Shykoff, J. A. & Giraud, T. 2005. High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Molecular Ecology* 14, 4275-4285.

Godefroid S., Phartyal S. S., Weyembergh G., Koedam N. 2005. Ecological factors controlling the abundance of non-native invasive black cherry (*Prunus serotina*) in deciduous forest understory in Belgium. *Forest Ecology and Management*, 210, 91-105.

Goudet, J. 1995. Fstat version 1.2: a computer program to calculate Fstatistics. *Journal of Heredity* 86, 485-486.

Grodzinska K., Godzik B. & Szarek-Lukaszewsa G. 2002. Nature Conservation in Poland, Emphasising the Carpathian Mountains. In *Effects of Air Pollution on Forest Health and Biodiversity in forests of the Carpathian Mountains*. Ed. IOS Press. 319p.

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

Guisan, A. & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8, 993-1009.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C., Phillips, S. & Peterson, A.T. 2007. What matters for predicting spatial distributions of trees: techniques, data, or species' characteristics? *Ecological Monographs* 77, 615-630.

Hastie, T. & Tibshirani, R. 1996. Discriminant analysis by Gaussian mixtures. *Journal of the Royal Statistical Society. Series B*, 58, 155-176.

Higgins, S. I. & Richardson, D. M. 1999. Predicting plant migration rate in a changing World: the role of long-distance dispersal. *The American Naturalist* 153, 464-475.

Holt, R. D. & Gaines, M. S. 1992. Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. *Evolutionary Ecology* 6, 433-447.

Hough, A. F. 1957. Encyclopedia of American Woods. R. Speller and Sons Eds. New York, USA.

Hough, A. F. 1965. Black Cherry (*Prunus serotina* Hehrh.). *Silvics of forest trees of the United States*. 539-545. H. A. Fowells, comp. U. S. Department of Agriculture, Agricultures Handbook 271. Washington, DC.

Huey, R. B., Gilchrist, G.W., Carlson, M. L., Berrigan, D. B. & Serra, L. 2000. Rapid evolution of a geographic cline in size in an introduced population. *Science* 287, 308-309.

Hufbauer, R. A. 2004. Population genetics of invasions: Can we link neutral markers to management? *Weed Technology* 18, 1522-1527.

Huntley, B., Bartlein, P. J. & Prentice, I. C. 1989. Climatic control of the distribution and abundance of beech (*Fagus* L.) in Europe and North-America. *Journal of Biogeography* 16, 551-560.

Husband, B. C. & Barrett, S. C. H. 1991. Colonisation history and population genetic structure of *Eichornia paniculata* in Jamaica. *Heredity* 66, 287-296.

Hutchinson, G. E. 1957. Concluding remarks. In *Population Studies: Animal Ecology and Demography*. Cold Spring Harbor Symposia on Quantitative Biology, Volume 22. Cold Spring Harbor Laboratory Press (NY), pp. 415-427.

Iguchi, K., Matsuura K., Mc Nyset, K.M., Peterson A. T. & Scachetti, R. 2004. Predicting invasions of North American basses in Japan using native range data and a genetic algorithm. *Transactions of the American Fisheries Society* 133, 845-854.

Keane, R. M. & Crawley, M. J. 2002. Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology & Evolution* 17, 164-170.

Kingsbury, J. M. 1964. Poisonous plants of the United States and Canada. Prentice-Hall Inc., Englewood Cliffs, N. J., USA. 626 pages.

Kolbe, J. J., Glor, R. E., Schettino, L. R., Lara, A. C., Larson, A. & Losos, J. B. 2004. Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431, 177-181.

Lee, C. E. 2002. Evolutionary genetics of invasive species. *Trends in Ecology & Evolution* 17, 386-391.

Levin, D. A. 2005. Niche shifts: the primary driver of novelty within angiosperm genera. *Systematic Botany* 30, 9-15.

Levine, J. M. & D'Antonio, C. M. 2003. Forecasting biological invasions with increasing international trade. *Conservation Biology* 17, 322-326.

Little, E. L., Jr. 1971. Atlas of United States trees, volume 1, conifers and important hardwoods: U.S. Department of Agriculture Miscellaneous Publication 1146, 9 pages, 200 maps.

Losos, J. B., Leal, M., Glor, R. E., de Queiroz, K., Hertz, P. E., Schettino, L. R. *et al.* 2003. Niche lability in the evolution of a Caribbean lizard community. *Nature* 424, 542-545.

Mack, R. N., Simberloff, D. & Lonsdale, W. M. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications* 10, 689-710.

. **Malanson, G. P., Westman, W. E. & Yan, Y.L. 1992.** Realized versus fundamental niche functions in a model of chaparral response to climatic change. *Ecological Modeling* 64, 261-277.

Malanson, G. P. 1997. Simulated responses to hypothetical fundamental niches. *Journal of Vegetation Science* 8: 307-316

Marquis, D. A. 1975. The Allegheny hardwood forests of Pennsylvania. USDA Forest Service, General Technical Report NE-15. Northeastern Forest Experiment Station, Upper Darby, PA. 32 pages.

Marquis, D. A. 1990. *Prunus serotina* Ehrh. Black Charry. In: Burns RM, Honkala BH, eds. Silvics of North America. *Agriculture Handbook 654, Volume 2*. Hardwoods. Washington, DC, USA: USDA, pp. 238-249. Accessed on 12/17/2007 http://na.fs.fed.us/spf/pubs/silvics_manual/volume_2/prunus/serotina.htm

Mau-Crimmins T. M., Schussman H. R. & Geiger E. L. 2006. Can the invaded range of a species be predicted sufficiently using only native-range data? Lehmann lovegrass (Eragrostis lehmanniana) in the southwestern United States. *Ecological Modelling*, 93, 736-746.

Pairon, M. C. & Jacquemart, A. L. 2005. Disomic segregation of microsatellites in the tetraploid *Prunus serotina* Ehrh. (Rosaceae). *Journal of the American Society for Horticultural Sciences* 130, 729-734.

Pairon, M. & Jacquemart, A. L. 2008. Molecular Evidences for two contrasted scenarios of colonization and multiple introductions of the invasive black cherry (*Prunus serotina*) to Europe. Submitted.

Parisod, C., Trippi, C. & Galland, N. 2005. Genetic variability and founder effect in the Pitcher Plant (*Sarracenia purpurea*) in populations introduced into Switzerland: from inbreeding to invasion. *Annals of Botany* 95, 277-286.

Pearman, P., Guisan, A., Broennimann, O. & Randin, C. (in press) Niche dynamics in space and time. *Trends in Ecology & Evolution*

Pearson, R. G., Thuiller, W., Araújo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P. & Lees, D.C. 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography* 33, 1704-1711.

Peterson, A. T., Soberon, J. & Sanchez-Cordero, V. 1999. Conservatism of ecological niches in evolutionary time. *Science* 285, 1265-1267.

Peterson, A. T. & Vieglais, D. A. 2001. Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *BioScience* 51, 363-371.

Peterson, A. T. 2003. Predicting the geography of species invasions via ecological niche modeling. *The Quatterly Review of Biology* 78, 419-433.

Peterson, A. T. & Holt, R. D. 2003. Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. Ecology Letters 6, 774-782.

Petit, R. J., Duminil, J., Fineschi, S., Hampe, A., Salivini, D., Vendramin, G. G. 2005. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* 14, 689–701.

Pimentel, D., Lach, L., Zuniga, R. & Morrison, D., 2000. Environmental and economic costs of nonindigenous species in the United States. *BioScience* 50, 53-65

Prasad A.M., Iverson L.R. & Liaw A. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, 9, 181-199.

R Development Core Team. 2007. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. Accessed on 3/21/2007. http://www.R-project.org

Reinhart, K. O., Packer, A., Van der Putten, W. H. & Clay K. 2003. Plant–soil biota interactions and spatial distribution of black cherry in its native and invasive ranges. *Ecology Letters* 6, 1046-1050.

Reinhart, K. O., Callaway, R. M. 2006. Soil biota and invasive plants. *New Phytologist*, 170, 445-457.

Romesburg, H. C. 1985. Exploring, confirming and randomization tests. *Computers and Geosciences* 11, 19-37.

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., Baughman, S., Cabin, R. J., Cohen, J. E., Ellstrand, N. C., McCauley, D. E., O'Neil, P., Parker, I. M., Thompson, J. N. & Weller, S. G. 2001. The population biology of invasive species. *Annual Review of Ecology and Systematics* 32, 305-332.

Sartori, F. 1988. *Prunus serotina* Ehrh. en Italie. *Colloques Phytosociologiques* 14, 185-217.

Segurado P., Araújo M.B. & Kunin W.E. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, 43, 433-444

Starfinger, U., Kowarik, I. 2003. *Prunus serotina* Ehrh. (Rosaceae), Späte Traubenkirsch. *Neoflora fact sheet.* Accessed on 01/22/2008 http://www.floraweb.de/neoflora/handbuch/prunusserotina.pdf

Starfinger, U., Kowarik, I., Rode, M. & Schepker, H. 2003. From desirable ornamental plant to pest to accepted addition to the flora? - The perception of an alien plant species through the centuries. *Biological Invasions* 5, 323-335.

Starfinger U. 2007. NOBANIS – Invasive Alien Species Fact Sheet – *Prunus serotina*. Online Database of the North European and Baltic Network on Invasive Alien Species – NOBANIS accessed on 12/17/2007 from www.nobanis.org

Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285-1293.

Thuiller, W. 2003. BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9, 1353-1362.

Thuiller, W., Araújo, M. B., Pearson, R. G., Whittaker, R. J., Brotons, L. & Lavorel, S. 2004. Biodiversity conservation: Uncertainty in predictions of extinction risk. *Nature* 430, 33.

Thuiller, W., Richardson, D. M., Pysek, P., Midgley, G. F., Hughes, G. O. & Rouge, M. 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Changes Biology* 11, 2234-2250.

USDA, NRCS. 2007. The PLANTS Database. Accessed on 12/17/2007, http://plants.usda.gov

Walters, R. S. 1985. Black Cherry Provenances for Planting in Northwestern Pennsylvania. USDA Forest Service, North Eastern Experiment Station. Research Paper NE-552.

Weir, B. S. & Cockerham, C. C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358-1370.

Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. 1998. Quantifying threats to imperiled species in the United States. *BioScience* 48, 607-615.

Woodward, F. I. & Williams, B. G. 1987. Climate and plant distribution at global and local scales. *Plant Ecology*, 69, 189-197

Yurukov S, Zhelev P. 2001. The Woody Flora of Bulgaria: A Review *Schweizerische Zeitschrift für Forstwesen*, 152, 52-60.

Zhu, L., Sun, O. J., Sang, W., Li, Z. & Ma, K. 2007. Predicting the spatial distribution of an invasive plant species (*Eupatorium adenophorum*) in China. *Landscape Ecology* 22, 1143-1154.