Quantifying invasion risk: the genus *Pinus* as a model system

A thesis
submitted in partial fulfilment
of the requirements for the Degree of
Doctor of Philosophy
at
Lincoln University

by

Kirsty F. McGregor

Lincoln University, New Zealand
2012

Declaration

Data used in part of Chapter 5 was collected by K. Gravuer (Gravuer, 2004; Gravuer *et al.*, 2008). Permission to use these data was given by R. Duncan who owns the data and references are provided to the source of these data in Chapter 5.

The following manuscripts forming Chapter 2 and Chapter 3 have been published at the time of printing:

McGregor, K. F., Watt, M. S., Hulme, P. E. & Duncan, R. P. (2012) How robust is the Australian weed risk assessment protocol? A test using pine invasions in the northern and southern hemispheres. *Biological Invasions* **14**, 987–998.

McGregor, K. F., Watt, M. S., Hulme, P. E. & Duncan, R. P. (2012) What determines pine naturalisation: species traits, climate suitability or forestry use? *Diversity and Distributions* **18**, 1013–1023.

12:51 pm Tuesday 22 February 2011

Kia Kaha

Abstract

Introduction: The New Zealand Government has committed to a 250,000 ha expansion of plantation forests by 2020 in order to diversify the forestry sector and capture carbon to mitigate climate change. Whilst introducing novel alien species can bring economic benefits, the risks of future invasion problems have not been fully quantified at the appropriate scale for many species. This is because invasions are a complex mix of species-traits, biogeographic factors, human actions, and also because many long-lived woody species have lag-phases between initial introduction, naturalisation and invasion.

This thesis investigates why some species become invasive whilst others do not using the genus *Pinus* as a model system, and New Zealand (NZ) and Great Britain (GB) as study regions. I improve on previous studies that have addressed this question by accounting for successes and failures across the entire invasion process (which incorporates the stages introduction, naturalisation and invasion).

Methodology: I compare four methods of quantifying invasion risk by: (a) testing how robust the Australian weed risk assessment tool (WRA) is to methodological issues including taxonomic range, region and knowledge of invasive behaviour elsewhere; (b) quantifying the relative contribution of species, biogeographic, and human factors to invasion success using boosted regression trees (BRT); (c) assessing whether phylogenetic relationships can predict invasion risk, and whether controlling for phylogeny in Markov chain Monte Carlo generalised linear mixed models (MCMCglmm) changes the importance of species, biogeographic and human factors in invasion success; and (d) dissecting the causal relationships between species, biogeographic and human factors using a novel Bayesian method for exploratory path analysis.

Results: I found that the WRA performed well at discriminating between successful and failed species at the introduction and naturalisation stages (AUC ≥ 0.80) but not at the spread stage, and these results were consistent between NZ and GB. When I repeated the procedure without information of species' prior invasion behaviour, the WRA was less accurate at distinguishing among species (area under the reciever operating characteristics curve or “AUC” ≤ 0.73). Thus the WRA may not be a viable approach to risk assessment when this crucial information is unavailable.

Boosted regression tree analysis indicated that human (high forestry use index) and biogeographic factors (closer climate match; NZ only) were the strongest predictors of introduction success. Human (a high forestry use index, large area planted and longer residence time) and biogeographic attributes (a close climate match and larger native range size) were the strongest contributors to naturalisation (NZ and GB). Species attributes (including the Z-score, a composite measure of pine invasiveness) contributed relatively little compared to other factors at all stages. The BRT method was reliable (introduction stage AUC ≥ 0.86; naturalisation stage AUC ≥ 0.98), relatively straightforward, and could be used as an alternative approach to risk assessment when the WRA may fail.

I found that there was no phylogenetic signal in introductions, naturalisations, invasions, or in any traits that might determine success for *Pinus*. Consequently, phylogeny may not be a useful predictor of invasion risk for pines. Phylogenetically controlled MCMCglmm produced the same results as non-phylogenetically controlled models with a similar level of reliability (introduction AUC $= 0.92$; naturalisation AUC $= 1.00$). These results suggest that non-phylogenetic models produced reliable results and that including phylogeny will not bias results even when no phylogenetic signal is present.

Exploratory path analysis suggested that introduction success was determined directly by a close climate match and high forestry use index. In contrast to previous results at the naturalisation stage I found that *Pinus* introductions were also highly influenced by the Z-score (species attributes) as well as direct links with human and biogeographic effects. Propagule pressure (residence time and area planted) was a common mechanism for *Pinus* and the additional study genus *Trifolium*, highlighting the importance of propagule pressure as a null model of invasions. Path analysis also performed well at the introduction (AUC $= 0.93$) and naturalisation stages (AUC $= 1.00$).

Conclusions: The novel aspects of this thesis include: quantifying failures at the introduction stage; comparing the relative importance of species, biogeographic and human factors on success at each stage of invasion; and comparing how the importance of these factors varies for the same taxonomic group across two regions. The results of this thesis suggest that there is an inherent conflict between introducing species for forestry and their invasion risk. This conflict requires measures such as plant breeding and landscape management in order to uncouple utility from risk. Risk assessments such as the WRA may not be suitable for all species when they have no history of introduction outside their native range. Therefore an adaptive approach to risk assessment is needed that includes both the costs and the benefits of introduction and utilises alternative approaches to risk assessment when standard approaches such as the WRA may fail.

Keywords: alien; biogeography; climate match; forestry; hemisphere; human use; introduction; invasion; life-history traits; naturalisation; path analysis; phylogeny; phylogenetic signal; *Pinus*; propagule pressure; spread; *Trifolium*; weed risk assessment.

Acknowledgements

I would first like to thank my supervisors Richard Duncan, Philip Hulme and Michael Watt for the diverse skills they brought to the project. Specifically, for providing many patient hours of discussion, advice on different analytical techniques and guidance on writing for scientific publication. Mike also facilitated my work at Scion collecting archived data. So, thank you all.

This project involved a great deal of databasing and digitisation which I could not have done alone. Thanks are due to: Graham Banton for compiling part the GBIF distribution data; Corinne Staley for digitising the working plans obtained from Scion; Carolin Weser for digitising Scion's archived seed register; and Melanie Harsch for gathering and databasing archived nursery newspaper adverts for New Zealand. Thanks also to Hazel Gatehouse for data on introductions and naturalisation to NZ; Jon Sullivan for loaning nursery catalogues; Brad Case for discussion on GIS; and Rupert Collins for proof reading, help typesetting in LaTeX and advice on phylogenetic analysis. I also wish to thank the staff at Scion in Rotorua, in particular the staff in the print and copy centre for their assistance accessing the archives and allowing me to use their copy equipment.

A special thanks also to my friends in the Plant Biosecurity Group (a.k.a. The Weeds Lab) past and present. In particular Melanie Harsch, who has proved to be a wonderful source of advice on all things from writing manuscripts to Bayesian data analysis. Also my fellow 2009 cohort: Federico Tomasetto, Jennifer Pannell and Elizabeth Wandrag for their moral support!

Finally, I would like to thank my Mother, Susan McGregor, for proof-reading parts of this thesis and for always encouraging me to do my best. My Mother spent a year living in New Zealand and her stories encouraged me to be interested in travel and ultimately to take on this PhD project a world away from home.

Abbreviations

Abbreviation	Definition
AIC	Akaike information criterion
AUC	Area under the receiver operator characteristics curve
BIC	Bayesian information criterion
BRT	Boosted regression tree
DIC	Deviance information criterion
GB	Great Britain
glmm	Generalised linear mixed models
MCMC	Markov Chain Monte Carlo
NZ	New Zealand
ROC	Receiver operating characteristic curve
SEM	Structural equation models
WRA	Australian weed risk assessment

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Commercial trees as invasive aliens

Alien trees introduced for commercial forestry, agroforestry, erosion control and ornamental purposes have become invaders in ecosystems around the world (e.g. CAB International, 2010; Essl *et al.*, 2011, 2010; Křivánek & Pyšek, 2008; Nuñez & Medley, 2011; Procheş *et al.*, 2012; Richardson & Rejmánek, 2004, 2011; Richardson *et al.*, 1994). Such deliberate range expansion circumvents natural dispersal barriers and creates an opportunity for trees to persist outside cultivation (Richardson & Rejmánek, 2004). Forestry is an efficient pathway for invasion, because it introduces individuals from provenances suitable for particular climates and implements large-scale planting, creating massive propagule pressure (Essl *et al.*, 2011, 2010; Křivánek *et al.*, 2006). Forestry is known to be a significant pathway for introduction (Richardson *et al.*, 2000b), and commercial scale plantations are a key factor in their escape and naturalisation globally (Essl *et al.*, 2011, 2010). Because invasive organisms represent a significant threat to biodiversity (Wilcove *et al.*, 1998), it is essential that the risks involved in introducing new species for commercial purposes are fully understood.

Invasive commercial trees, particularly conifers, are especially prevalent in the southern hemisphere, where large areas of land in Australia, New Zealand, southern Africa, and more recently South America, have been converted to plantations (Richardson & Higgins, 1998). Due to their tendency to form dense stands, plantation trees can affect many ecosystem processes. Some tree species can alter natural ecosystems, for example by increasing water loss (Dye, 1996; Zavaleta, 2009), overgrowing tussock grasslands (Ledgard, 2001), increasing fuel loads (Brooks *et al.*, 2004) and nutrient enrichment (Richardson & Higgins, 1998). Conversely, similar species introduced outside their native ranges to the northern hemisphere seem less likely to establish outside cultivation (Adamowski, 2004; Carrillo-Gavilàn & Vilà, 2010; Mortenson & Mack, 2006; Richardson & Rejmánek, 2004). The invasive spread of plantation trees was first noticed in New Zealand and southern Africa in the late nineteenth and early twentieth centuries, and in Australia in 1950 (Richardson *et al.*, 2008). For example in New Zealand, conifer invasions are one of the most visible and costly weed problems (Harding, 2001, Figure 1.1), and conifers make up 70% of the woody species listed on the consolidated list of environmental weeds in New Zealand (Howell, 2008).

Figure 1.1. Pines are highly visible invaders, here pictured spreading across high country grasslands near State Highway 8, close to Lake Pukaki, Canterbury, New Zealand.

Afforestation of marginal land is an attractive option for many governments to meet targets for carbon emission reductions in order to mitigate climate change. Governments of industrialised nations that signed up to the Kyoto Protocol are committed to “the promotion of afforestation and reforestation" (Article 2 of the Kyoto Protocol; United Nations, 1998) as long-term carbon sinks. Alien species already exploited commercially in a region, or species not yet introduced that have proven valuable elsewhere, are likely to be the preferred choice for afforestation schemes. The New Zealand government has committed to a 250,000 hectare expansion of planted forest area by 2020 to mitigate climate change impacts and as the basis for sustainable growth in the forestry sector; and has already achieved a 2,783 hectare increase in forest area since 2007 (Emmissions Trading Scheme Review Panel, 2011). Ninety percent of the forest area in New Zealand is planted with *Pinus radiata* (MAF, 2011). However most of the sites available for this expansion are outside the suitable area where *P. radiata*, can be grown productively. Therefore, there is an urgent need to identify suitable species to diversify planted forests in New Zealand, and elsewhere. While diversification could bring enormous economic benefits, these benefits need to be weighed against the risk that more widespread planting of some species may facilitate their escape and spread.

1.2 Defining invasion by stages

Lack of standardisation around terms such as "introduction", "naturalisation" and "invasion" can result in studies being non-comparable or ambiguous. Recently the concepts and definitions around the invasion process have been formalised in an attempt to standardise terminology (e.g. Blackburn *et al.*, 2011; Richardson *et al.*, 2000b). Invasions are now viewed as a series of stages that species must pass through in order to become invasive (Blackburn *et al.*, 2011, Figure 1.2). The definitions of each stage used throughout this thesis are as follows (*sensu* Richardson *et al.*, 2000b):

Introduction Introduced species must have been selected from the global pool for transport and introduced to a new region through human agency, overcoming a major geographical barrier.

Naturalisation From the pool of species that are introduced to a region, some go on to naturalise. Naturalised species (often called "established species") will have established self-sustaining populations in the wild without direct intervention from humans, recruiting offspring freely, usually close to the source populations. Throughout this thesis, species that are present in the new environment as "casual" (alien species that are present and reproducing occasionally in an area, but which do not form self-replicating populations, relying on repeated introductions for their persistence) are classified as introduced but not naturalised.

Invasion Invasive species are naturalised species that produce reproductive offspring, often in large numbers, and at a considerable distance from the parent plants (at scales: $>$ 100m; $<$ 50 years for taxa spreading by seeds and other propagules). Invasive species thus have the potential to spread over a large area.

Alien Once a species has been transported to a new region, the species is referred to as "alien" in that new region. Thus the term alien refers to introduced, naturalised and invasive species.

In addition to defining the terminology I use relating the the status of a species, I also use the term "propagule pressure" throughout this thesis. Propagule pressure in this thesis is analogous to "introduction effort". That is, I consider propagule pressure as an estimate of the introduction effort experienced by each species, measured as the total planting effort experienced by each species based on historical records.

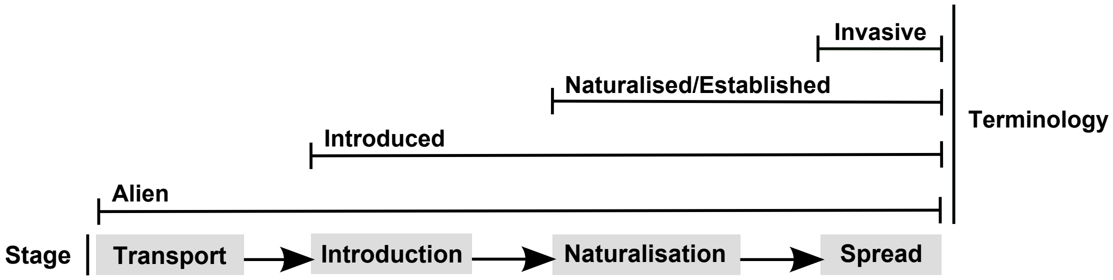

Figure 1.2. Framework of invasions as a stage-based process, and the terminology that describes species in each stage in this thesis, following Blackburn *et al.* (2011).

The definition of invasive used in this thesis deliberately does not include any quantification of the impacts that species may have in the new environment. Terms describing harmful species (species having negative impacts) include: "weed" (plants grown in sites where they are not wanted, usually having detectable economic or ecological impact, not necessarily an alien species); "environmental weed" (an alien species that is unwanted and causing impacts in unmanaged non-agricultural systems); and "transformers" (an invasive plant that changes the character, condition or form of ecosystems over a substantial area relative to that of the non-impacted ecosystem). Whilst these terms may be used to describe perceived problem species, I make no explicit attempt to quantify impact other than as the degree of spatial spread of a species in the landscape.

Using this stage-based framework has three main advantages. First, it provides unambiguous criteria for categorising species as introduced, naturalised or invasive. Second, knowing which pool a species belongs to is important when drawing conclusions about the factors that determine success or failure at a given stage (Cassey *et al.*, 2004). For example, when attempting to identify why some species are invasive (have spread), comparing the traits of all species in the global pool with the traits of those that have become invasive will miss out the introduction and naturalisation stages. This misspecification lumps all the traits that controlled success or failure at the introduction and naturalisation stage into an analysis of the invasion stage, and may lead to spurious results (Cassey *et al.*, 2004). Finally, this framework forces us to consider the traits of species that failed at a given stage, which can be as informative as the traits of the species that made it all the way though to become invasive (Diez *et al.*, 2009).

Whilst there is a great deal of information available quantifying factors that determine naturalisation success, relatively few studies have included success and failures at the introduction stage (Puth & Post, 2005). However, non-random selection at the introduction stage can bias the pool of species available for naturalisation (e.g. Blackburn & Duncan, 2001). Studies quantifying introduction success and failure remain rare for plants. One study used introductions, naturalisations and invasions of the genus *Trifolium* (true clover) in New Zealand (Gravuer *et al.*, 2008). An additional study examined introductions to, and spread from, a botanical garden in Tanzania (Dawson *et al.*, 2009a,b).

1.3 The genus *Pinus* as a model system

The genus *Pinus* (pine trees) has become a model system for studying invasions (Richardson, 2006). Members of the genus are among the most-common forestry invaders, with 21 out of 115 species being invasive or naturalised globally (Richardson & Rejmánek, 2004). Due to their economic importance and obvious impacts in native ecosystems, there is a wealth of information available about species introduction and naturalisation histories, life-history traits, native and naturalised distributions and phylogenetic relationships (e.g. Gernandt *et al.*, 2005).

A suite of species, biogeographic attributes, and human factors have been linked with pine invasions globally: (1) A Z-score derived from three life-history traits (seed mass, minimum juvenile period, and minimum interval between large seed crop years) is able to distinguish between invasive and non-invasive species (Rejmánek & Richardson, 1996). Species with small seeds, short juvenile periods, and short intervals between large seed crops are most likely to be invasive. However, the Z-score was originally derived from a sub-set of the genus (29 species) and it is not known how generalisable the Z-score is to the whole genus, and across regions with different introduction histories; (2) A close climate match between the native and introduced range increases naturalisation and invasion success in pines (e.g. Essl *et al.*, 2011, 2010; Nuñez & Medley, 2011). The size of a species' native range size correlates with its naturalised range size, such that species with a wider native distribution also have a larger naturalised distribution (Procheş *et al.*, 2012). This trend has been found in other conifers (Essl *et al.*, 2011, 2010); (3) Finally, human factors may contribute to invasion risk. An index of global forestry use (measured as the number of citations a species received in the CABI Forestry Compendium) correlates with naturalised range size in pines (Procheş *et al.*, 2012) and use in commercial forestry is linked with conifer naturalisations globally (Essl *et al.*, 2011).

Despite this knowledge, there has not yet been a systematic test of how these factors act at each stage of the invasion process (explicitly including the introduction stage) across different regions to determine invasion risk for pines. Understanding pine invasions is likely to inform our knowledge of the risk from other woody genera used for forestry, such as *Acacia* (Richardson *et al.*, 2011), *Abies*, and *Cupressus* (Essl *et al.*, 2011), and *Eucalyptus* (Rejmànek & Richardson, 2011). Therefore, a quantitative understanding of the factors determining pine invasions can move the discipline on from documenting invasions to predicting invasions.

1.4 Quantifying invasion risk

Since the first synthesis of invasion biology was published in 1958 (Elton, 1958), a large body of work has aimed to identify and predict why some species become invasive whilst others fail, the aim of this work being to better predict and thus prevent future invasions. Methods used for identifying potentially invasive species and attempting to predict invasions largely fall into two categories. First, the formulation of structured risk assessment procedures that involve answering a set of questions, the outcome of which is optimised based on prior information to identify potentially invasive species. Second, researchers have used statistical procedures to identify how a range of factors determine invasion success, with the aim of finding general rules. Finally, there is growing interest in how phylogenetic relationships between species could be used to inform risk, a factor that is often suggested as being important but rarely formally tested (e.g. Miller *et al.*, 2011). Therefore, this thesis focuses on: (1) assessing a widely used risk assessment scheme, (2) using multiple statistical techniques to identify factors determining success, and (3) quantifying the ability of phylogenetic relatedness to predict invasive species.

1.4.1 Weed risk assessment

Formal risk assessments are a popular way to quantify invasion risk. The most widely used and tested risk assessment available is the Australian weed risk assessment (WRA) scheme, which was developed in Australia and New Zealand (Pheloung *et al.*, 1999) and is in use to screen new introductions. The WRA uses the answers to 49 questions concerning the species' biology, biogeography and behaviour elsewhere to classify a plant species according to its risk of becoming invasive (Pheloung *et al.*, 1999). The WRA classifies species into: a) those that pose little risk of becoming invasive in the new location and could be accepted for introduction; b) those that pose a high risk and should be rejected; and c) an intermediate group that require further evaluation. These classifications are assigned using a pre-defined threshold score which was optimised using data from Australia and New Zealand to reject all historically serious weeds, 10% or fewer non-weeds, and recommend no more than 30% of species for further evaluation (Pheloung *et al.*, 1999).

The WRA has been tested for a wide range of plant species and habitats, and found to have a high degree of reliability (reviewed in Gordon *et al.*, 2008b; Roberts *et al.*, 2011). However, the WRA has several limitations that need to be fully explored before it can be accepted as the *de facto* risk assessment scheme. First, it is known to be less reliable when information on previous invasive behaviour of a species is not available (Caley & Kuhnert, 2006). For many species, particularly those that are used widely for commercial purposes, such information may be readily available in databases (e.g. the Forestry Compendium; CAB International, 2010). However, in situations where a species has never been introduced outside its native range or has no history of invasion elsewhere, as is the case for New Zealand where 20% of recently naturalised alien plant species have no history of invasion elsewhere (Williams *et al.*, 2000), the reliance of the WRA on this criterion may lead to inaccurate assessment of potential invasion risk. Second, the repeatability of the WRA has also not been tested in different regions for the same taxonomic group despite the WRA score being able to vary between regions due to region-specific questions related to climate suitability and the presence of pests and deseases. And finally, the way that the WRA has been retrospectively tested using ambiguous definitions of invasiveness that do not conform to the model of invasions as a series of stages (Blackburn *et al.*, 2011) means that it is not known at exactly which stage in the invasion process the assessment is most accurate.

1.4.2 Factors determining invasions

The alternative to using structured risk assessment schemes such as the WRA for assessing invasion risk is statistical analysis of the factors that confer success and failure through retrospectively testing which species succeed or fail at different stages of the invasion process. Using a variety of analytical methods, species traits, biogeographic and human attributes have been linked with success at the naturalisaion and invasion stages of the invasion process. For example, species traits such as rapid maturation, small seeds and frequent reproduction, have been linked with invasiveness in pines and conifers more generally (Grotkopp *et al.*, 2002; Rejmánek & Richardson, 1996; Richardson & Rejmánek, 2004; Richardson *et al.*, 1994). More recent evidence points to the role that human and biogeographic attributes play in determining naturalisation and invasion success. Biogeographic factors that increase naturalisation/invasion probability in woody species include a closer climate match between a species' native and introduced range (Nuñez & Medley, 2011), and a large native range size (e.g. Procheş *et al.*, 2012). Human factors such as introduction effort (or "propagule pressure") can increase the chance of an introduced species naturalising and invading (Boulant *et al.*, 2009; Gassò *et al.*, 2010; Křivánek *et al.*, 2006; Medawatte *et al.*, 2010; Pyšek *et al.*, 2009b), as can residence time (Pyšek *et al.*, 2009b; Richardson *et al.*, 1994) and economic use (Essl *et al.*, 2011, 2010). This large body of previous reserch was used to guide the data collection in this thesis to focus data collection.

Despite these findings, there is little research on how these three main groups of variables (species, biogeographic and human) interact to determine invasion outcomes. Specifically, the relative importance of these variables on success across the invasion process (incorporating failures at the introduction stage) has only been assessed for one group of species (*Trifolium* or true clovers) that were intentionally introduced to New Zealand (Gravuer *et al.*, 2008), and has not previously been assessed for other species. This study found that human factors were important at the introduction stage, biogeographic variables were important at all stages, and factors acted differently at each stage. Thus it is unclear which factors are repeatedly the most important, and how this can vary between different regions, stages of invasion, and with taxonomic group. Furthermore, no study has attempted to quantify the causal linkages between these multiple variables and success at different stages of the invasion process, despite methods such as path analysis being well established in the ecological and invasion literature (e.g. Grace, 2006; Shipley, 2000).

Understanding the relative importance of variables and the causal structure linking them together would be useful for informing managers about avenues for managing future risk and could potentially be used to improve current risk assessments such as the WRA.

1.4.3 The role of phylogeny

Phylogenetic relationships between species have been recognised as an important factor that could be used to identify risk. Because related species may share similar traits through shared ancestry (Harvey & Pagel, 1991), related species may also share traits that promote invasiveness. This theory has been incorporated into the WRA where species with congeneric invaders are given a higher risk score (Pheloung *et al.*, 1999). Several invasion studies have investigated this idea empirically by classifying species to higher taxonomic levels such as family, order and subclass, and assessed whether some groups have more invasive members than expected by chance (Alcaraz *et al.*, 2005; Blackburn & Duncan, 2001; Daehler, 1998; Lloret *et al.*, 2005; Pyšek, 1998; Tingley *et al.*, 2010; Vázquez & Simberloff, 2001; Vilà & Muñoz, 1999). However, only three families (Amaranthaceae, Papareraceae, and Polygonaceae) have been identified by more than one study as being over-represented by invasive species (Daehler, 1998; Pyšek, 1998; Vilà & Muñoz, 1999) and there are methodological issues, such as the non-comparability of Linnaean taxonomic ranks, that may confound such analyses.

Extinction risk is a trait analogous but opposite to invasion risk, and has been investigated in a similar way with the aim of identifying future at-risk species in order to prioritise conservation efforts (Purvis, 2008). However work in this field has gone further by using quantitative measures of phylogenetic non-randomness that utilise phylogenetic trees rather than comparing over-representation in taxonomic groups. Extinction risk is known to cluster on a phylogeny producing a “phylogenetic signal” (defined as the statistical non-independence among species trait values due to their phylogenetic relatedness) in risk at the family and genera level for birds and mammals (Purvis *et al.*, 2000; Russell *et al.*, 1998), amphibians (Corey & Waite, 2008; Stuart *et al.*, 2004) and plants (Davies *et al.*, 2011; Pilgrim *et al.*, 2004; Schwartz & Simberloff, 2001; Vamosi & Wilson, 2008). However, identifying high taxonomic levels, which potentially encompasses large groups of species as risky may not provide any specific management advice. Therefore, the extent to which phylogenetic signal is present in invasion risk, and at which taxonomic level any signal is evident, is an area that requires further investigation.

Knowledge of species phylogenies can also be useful for statistical analyses into the factors determining invasion. Conventional statistical tests assume that species are independent units for analysis, though this is rarely the case because species are linked through shared evolutionary history (Harvey & Pagel, 1991). However, many techniques are now available to control for phylogeny, including independent contrasts (Felsenstein, 1985), variance partitioning (Desdevises *et al.*, 2003), phlyogenetic generalised least squares regression (Grafen, 1989), and phylogenetic mixed models (Housworth *et al.*, 2004). Recently, invasion studies have begun to incorporate phylogenetic information in order to account for this non-independence (Alcaraz *et al.*, 2005; Dawson *et al.*, 2009a, 2011b; Jeschke & Strayer, 2006; Küster *et al.*, 2008; Pyšek *et al.*, 2009a), generally using variance partitioning, independent contrasts and mixed models. These studies have shown that although in general the variance attributed to phylogeny has minor explanatory power, the significance of other explanatory variables can change when phylogeny is included (Alcaraz *et al.*, 2005; Dawson *et al.*, 2009a, 2011b; Jeschke & Strayer, 2006) and that the effect of phylogeny may be greater at lower taxonomic levels and at the later stages of the invasion process (Pyšek *et al.*, 2009a). Taken together, this evidence suggests that accounting for phylogenetic relationships between taxa is important for invasion studies, and that some assessment of the level of phylogenetic autocorrelation of traits should be undertaken.

1.5 Rationale and project aims

As outlined above, introducing alien tree species for commercial purposes can bring substantial economic benefits. However, there is a considerable risk of alien species naturalising and spreading when introducing new species. This risk has not yet been quantified at an appropriate spatial scale and across regions for most species, and is likely to be underestimated for several reasons: (1) The probability that a species will naturalise is strongly linked with introduction effort or “propagule pressure”. More widespread planting increases the amount of seed entering the environment and increases the probability that wild populations will establish. The scale of planting for commercial purposes is often orders of magnitude higher than for experimental plots, and commercial-scale planting is continued over long time periods. (2) The characteristics that make a good commercial species under controlled conditions, such as fast growth rate, may contribute to invasive spread if species escape cultivation on a large scale (Puth & Post, 2005). (3) Commercial trees show well documented lag phases, of up to 100 years, from initial introduction date to becoming invasive (Křivánek & Pyšek, 2008; Richardson *et al.*, 1994). Many species with the potential to become invasive may not yet have done so because their plantings have been localised or of small scale. Consequently, the invasion potential of many long-lived tree species used in both amenity and commercial planting has not yet been realised.

The aim of my thesis is to understand and quantify these risks by identifying factors that determine why some tree species, but not others, escape cultivation and become invasive. This area has been identified by Scion (a New Zealand Crown Research Institute dedicated to improving the international competitiveness of the New Zealand forest industry and building a stronger biobased economy) as a major knowledge gap in the FRST funded "Diverse forests for a sustainable New Zealand" programme. Whilst this research has clear applicability locally, it will also address fundamental questions in plant invasion ecology, including:

1. How robust is the Australian weed risk assessment (WRA) system, and does this risk assessment perform equally well in different regions for the same group of species?
2. Which species, biogeographic, and human factors determine success and failure at each stage of the invasion process, what is their relative importance?
3. Is there a phylogenetic signal in invasion risk within the genus *Pinus*, and the traits linked to invasion success? Does controlling for phylogenetic relationships among species change the importance of factors?
4. Given the relative importance of key species, biogeographic and human factors, do these factors determine success or failure at each stage directly or indirectly, mediated through one or more other variables? Are these causal pathways the same for diverse taxonomic groups (*Pinus* and *Trifolium*)?
5. Given the results of previous chapters, is there an intrinsic conflict between further afforestation using alien species for carbon capture and a diverse forestry sector with future invasion risk?

This thesis uses two regions (New Zealand and Great Britain) in order to provide replication and to assess whether trends found in one region translate to another.

These regions provide excellent case-studies because they both have a wealth of information on *Pinus* introduction, naturalisation and inasion histories, and detailed records outlining the extent of planting for all species in the genus.

1.6 Thesis outline

All chapters have been written as self-contained research papers and consequently there is repetition in the introductions and discussions of some chapters. Where chapters are manuscripts of published research papers, the author contributions are stated at the end of the chapter, and a full citation is given as a footnote on the first page of the chapter. All literature cited in the thesis is given at the end to avoid unnecessary repetition. Chapter 2, Chapter 3 and Chapter 4 use the genus *Pinus* as a model system, and New Zealand and Great Britain as comparative regions. Chapter 5 uses *Pinus* and *Trifolium* introductions to New Zealand as case studies. Chapter 2 examines how robust the popular WRA system is to taxonomic range, region and knowledge of invasive behaviour elsewhere. Chapter 3 quantifies the relative contribution of a suite of species, biogeographic, and human factors to success and failure at the introduction and naturalisation stages of the invasion process in New Zealand and Great Britain. Chapter 4 assesses whether there is phylogenetic signal in invasion risk for pines, and whether controlling for phylogeny changes the conclusions about which factors determine introduction, naturalisation and invasion success. Chapter 5 applies a novel Bayesian method for exploratory path analysis, to quantify the direct and indirect causal effects of variables determining *Pinus* and *Trifolim* introduction and naturalisation to New Zealand. Finally, in Chapter 6, I discuss the implications of my results from all previous chapters for quantifying invasion risk in pines, the implications beyond pines, and suggest avenues for future study. Because each chapter is written as a self-contained research paper, Chapter 6 deals only briefly with the findings from specific chapters and focuses on assessing the implications of the thesis as a whole.

Chapter 2 was published in *Biological Invasions* in 2012[1] and is co-authored with Richard Duncan, Philip Hulme and Michael Watt. Chapter 3 was published in *Diversity and Distributions*[2] co-authored by Richard Duncan, Philip Hulme and

---

[1] McGregor, K. F., Watt, M. S., Hulme, P. E. & Duncan, R. P. (2012) How robust is the Australian weed risk assessment protocol? A test using pine invasions in the northern and southern hemispheres. *Biological Invasions* **14**, 987–998.

[2] McGregor, K. F., Watt, M. S., Hulme, P. E. & Duncan, R. P. (2012) What determines pine naturalisation: species traits, climate suitability or forestry use? *Diversity and Distributions* **18**, 1013–1023.

Michael Watt. A version of Chapter 5 is in preparation as two manusripts for future submission to appropriate journals, co-authored by Richard Duncan, Philip Hulme and Michael Watt.

Chapter 2

How robust is the Australian weed risk assessment protocol? A test using pine invasions in the northern and southern hemispheres[1]

2.1 Abstract

The Australian weed risk assessment protocol (WRA) is often considered the standard approach for pre-border screening of new plant introductions. Here we assess its robustness against three key criteria: ability to discriminate success or failure of species at three stages of the invasion process (introduction, naturalisation and spread); sensitivity to taxonomic range and target region; and dependence on knowledge of invasive behaviour elsewhere. We address these issues by retrospectively testing the WRA using pine (*Pinus*) introductions to New Zealand and Great Britain. For both regions we calculated WRA scores for 115 species, and classified all species according to whether they had been introduced, which of these had naturalised, and the extent of their naturalised distribution (spread). Using regression models, we assessed whether WRA scores could predict success at each stage. We repeated this procedure using WRA scores calculated without information on species naturalisation behaviour elsewhere. In both regions, the WRA could discriminate among species in the same genus at the introduction and naturalisation stages, but not at the spread stage. The outcome at the naturalisation stage depended on prior knowledge of naturalisation behaviour elsewhere. Without this information the WRA may be unable to distinguish among closely related species, and should be used cautiously where data on invasive behaviour elsewhere is lacking. Human selection played a strong role in the invasion

---

[1] McGregor, K. F., Watt, M. S., Hulme, P. E. & Duncan, R. P. (2012) How robust is the Australian weed risk assessment protocol? A test using pine invasions in the northern and southern hemispheres. *Biological Invasions* **14**, 987–998.

process both through introducing pine species likely to naturalise in New Zealand and Great Britain in the first instance, and subsequent use of many of these species for forestry in the target regions.

Keywords: climate matching; risk assessment; exotic species; forestry; spread; weed

2.2 Introduction

Weed risk assessment protocols use information on a species' biology, environmental preferences and known tendency to become a weed to determine the risk that an alien plant species will become invasive following its introduction to a new location. The Australian weed risk assessment protocol (WRA) has been widely evaluated (see Roberts *et al.*, 2011, for a recent review) and is often perceived as the *de facto* standard in weed risk assessment. The WRA uses the answers to 49 questions concerning the species' biology, biogeography and behaviour elsewhere to classify a plant species according to its risk of becoming invasive (Pheloung *et al.*, 1999). Using a pre-defined threshold score, the WRA classifies species into: a) those that pose little risk of becoming invasive in the new location and could be accepted for introduction; b) those that pose a high risk and should be rejected; and c) an intermediate group that require further evaluation. Although initially developed in Australia and New Zealand, the protocol has been adapted and tested for use with a wide range of life-forms in temperate (Gordon *et al.*, 2010; Jefferson *et al.*, 2004; Kato *et al.*, 2006; Křivánek & Pyšek, 2006; Nishida *et al.*, 2009), Mediterranean (Crosti *et al.*, 2010; Gassò *et al.*, 2010); subtropical (Gordon *et al.*, 2008a) and tropical (Daehler *et al.*, 2004; Dawson *et al.*, 2009b) biomes. Although evaluations have shown the WRA to correctly classify species as invasive more than 80% of the time (Gordon *et al.*, 2008b), critical issues with the protocol require assessment before it can be accepted as a robust general approach to weed risk assessment (Hulme, 2012).

First, species risk must be assessed against an objective risk of invasiveness (Hulme, 2010, 2012). Most WRA evaluations to date have used expert opinion to classify the species already present in a region as invasive or non-invasive. Given that the term invasive can be interpreted differently, and may mean naturalisation to some but negative impact to others (Colautti & Richardson, 2009), a more objective measure is needed. For example, Dawson *et al.* (2009b) assessed the WRA not only against its ability to discriminate between naturalised and non-naturalised species but also its value in explaining how widespread an alien species becomes at the landscape scale.

Second, most WRA evaluations assess the existing pool of alien species in a region with the aim of distinguishing between invasive and non-invasive species. This methodology ignores recent developments in invasion biology which stress that the process of becoming invasive involves passing through at least three stages: 1) a species must be introduced to a new region; 2) the species must establish a self-sustaining wild population (naturalise); and 3) the species spreads from its point of naturalisation, at which point it becomes invasive (Blackburn *et al.*, 2011). Species classified by experts as invasive will have passed through all three stages of this process, while those classed as non-invasive could be a mixture of introduced species that have failed to naturalise, naturalised species that have failed to spread, and even widespread species that are perceived to have little impact. Thus previous studies may have used different source pools for comparison (Cassey *et al.*, 2004) and ignored the introduction stage. A more thorough evaluation of the WRA could include the global source-pool of potential introductions to determine whether the WRA can identify which species were introduced, then from the pool of introduced species which have naturalised, and finally, of the naturalised species which have become invasive. This ensures a realistic base-rate is factored into the assessment of the reliability and accuracy of weed risk assessment (Hulme, 2012). A test of this kind would also assess the WRA's ability to discriminate amongst a broader range of species characteristics than those represented by subset of species already in the target region (Hulme, 2012).

Third, retrospective tests of the WRA have invariably drawn on species from a wide taxonomic range, typically including many different families. This taxonomic range may improve the accuracy of assessments because it makes it more likely to include species from groups with very different characteristics that may be linked to naturalisation success or failure (Onderdonk *et al.*, 2010). However, little is known about the ability of the WRA to discriminate among closely related species that may differ in their probabilities of naturalisation but are likely to share many characteristics in common.

Fourth, tests of the WRA have been limited to single regions (with the exception of Pheloung et al. [1999] who compared scores for species shared between New Zealand and Australia) and, while these often achieve high accuracy, we do not know how transferable the results are from one region to another (Onderdonk *et al.*, 2010). Such comparisons have been limited by the often region specific assemblages of alien plants examined using the WRA. However, by repeating the WRA for a common group of species in more than one region, the reproducibility of WRA outcomes can be assessed.

Finally, previous studies have identified a subset of questions in the WRA that are critical to determining the classification outcome (Caley & Kuhnert, 2006; Weber *et al.*, 2009), particularly those related to whether the species is known to be invasive elsewhere. While knowledge of invasive behaviour elsewhere is informative in assessing whether a widely introduced species is likely to naturalise in a new region, it provides no information for species introduced outside their native range for the first time, where there has been no opportunity to assess invasive behaviour. This situation may be common: in New Zealand, for example, over 20% of recent naturalised alien plants have no history of invasion elsewhere in the world (Williams *et al.*, 2000). While reliance on information about invasive behaviour elsewhere has been recognised, we do not know how sensitive the WRA is in situations where we lack this information, or at what stage in the invasion process this information becomes critical.

Our aim in this study is to evaluate the WRA with regard to these five issues, and thus to assess how robust the WRA is at differentiating stages in the invasion process, and its sensitivity to taxonomic range, region and knowledge of naturalisation elsewhere. To do this, we narrowed the taxonomic range by selecting a single, well-studied genus, *Pinus*, and examined how well the WRA could retrospectively predict pine success at three stages in the invasion process (introduction, naturalisation and spread) in two regions: New Zealand and Great Britain (hereafter referred to as NZ and GB respectively). We chose the genus *Pinus* because pines are a model group for studying invasions having been widely introduced and planted, with the subsequent naturalisation and spread of species being well documented (Essl *et al.*, 2011, 2010; Procheş *et al.*, 2012; Rejmánek & Richardson, 1996; Richardson, 2006). The global distribution of invasive pines suggests that locations in the southern hemisphere are more readily invaded than those in the northern hemisphere (Carrillo-Gavilàn & Vilà, 2010; Richardson & Higgins, 1998; Richardson & Rejmánek, 2004). Our choice of NZ and GB as study locations allows us to assess this by comparing locations with a similar area, climate and history of pine introductions, but in different hemispheres.

2.3 Materials & Methods

2.3.1 Study genus

Pines are widely planted for forestry and ornamental purposes, and information on their history of introduction along with data on species attributes are widely available (e.g. Boulant *et al.*, 2009; Grotkopp *et al.*, 2002; Rejmánek & Richardson, 1996; Richardson & Bond, 1991; Richardson, 1998b). We compiled a list of all 115 *Pinus* species (see Appendix A) following comprehensive taxonomic treatments of the genus (Farjon, 2005; Price *et al.*, 1998), and research into commonly delimited species (Earle, 2008; GBIF, 2011; IPIN, 2004; Perry, 1991; USDA, 2011). We did not include recently described species that are not widely recognised, and included only species for which two or more sources supported their recognition at the species level. This was necessary because we used historical sources to identify species that had been introduced, and newly described taxa would not have appeared in these sources. We did not include subspecies or varieties due to disagreements about the delineation of these taxa, and because the data used for weed risk assessment does not allow us to differentiate among taxa within species.

2.3.2 Introduction, naturalisation and invasion

From the global pool, we identified which pine species had been introduced to NZ and GB and the date they were first recorded as introduced using historical records that included the horticultural and scientific literature (Appendix A). Within each region, introduced species were classed as naturalised if they had established wild populations outside of cultivated areas (*sensu* Richardson *et al.*, 2000b); we excluded from this category species regenerating naturally only in areas where the species is currently cultivated. Although *P. sylvestris* is native to GB it has naturalised in parts of GB that are outside its native range but to ensure comparability with NZ we treated it as native to the region and excluded it from the GB analysis. To provide an objective measure of the level of invasion, we used the extent of a species naturalised distribution in each region. In NZ, we collated records of pine species established in the wild from: a) the Department of Conservation's weeds database; b) the results of a questionnaire which we sent to all Department of Conservation area offices asking which of the 13 pine species known to have naturalised in NZ were present as wild populations in their area; c) the spatial locations of herbarium specimens (from `http://www.nzherbaria.org.nz/virtherb.asp` and Scion

National Forestry Herbarium Database); d) the New Zealand Biodiversity Recording Network (`http://www.nzbrn.org.nz`); e) location data in the New Zealand Flora (Webb *et al.*, 1988); and f) literature searches. Subsequently, each record was assigned to one of the NZ Department of Conservation's 43 administrative areas (see Appendix B) and the number of areas in which each species was recorded was used as a measure of invasiveness (sensu Richardson *et al.*, 2000b). For GB we obtained data on the presence of wild populations of each pine species in each of the 107 Watsonian vice-counties in GB (mainland England, Scotland and Wales) from Preston *et al.* (2002), and used the number of vice-counties where present as the measure of invasiveness. We excluded offshore islands in estimating distribution for both NZ and GB.

To determine which of the introduced species had been planted for forestry purposes, we searched archival forest working plans that covered the majority of state forests for NZ (housed at Scion, Rotorua) and GB (housed at the Forestry Commission, Alice Holt, Surrey). We define a forestry species as any species that was listed as planted by the state forest service.

2.3.3 Weed Risk Assessment

Following standard protocols (Gordon *et al.*, 2010), we calculated a separate WRA score in each region for each of the 115 species in the global pool of the genus *Pinus* (excluding data on invasion history derived from the target region; see Appendix C). To reduce the number of species placed in the "evaluate" category after the initial assessment, all species were screened for a second time using a widely used decision tree for secondary classification (Daehler *et al.*, 2004; Křivánek & Pyšek, 2006). Evidence of the presence of effective natural enemies (question 8.05) was treated on a case-by-case basis using information gathered following standard protocols (Gordon *et al.*, 2010) excluding information from the proposed country of importation, and the answer could therefore vary between NZ and GB.

Because climate match is known to be a predictor of pine naturalisations (Nuñez & Medley, 2011) we assessed the goodness of climate match for each species. We obtained records of the presence of each species in their native geographic range from the Global Biodiversity Information Facility (GBIF, 2011), online databases (Burns & Honkala, 1990; CAB International, 2010; Earle, 2008; USDA, 2010), standard reference books (Perry, 1991; Richardson, 1998b), floras for each region, and online literature searches in Google Scholar. We searched for records using both accepted names and recognised synonyms. Duplicate records were removed and records without geographical coordinates were assigned coordinates from location information using Google Earth and Fuzzy Gazetteer (`http://isodp.fh-hof.de/fuzzyg/query/`). All occurrence records were then displayed in ArcMap Version 9.3 (ESRI, 2008), and incorrect coordinates were checked and corrected or removed. Five species (*Pinus hakkodensis*, *P. henryi*, *P. squamata*, *P. stankewiczii* and *P. wangii*) had fewer than three global occurrence records and a specific climate match could not be obtained. As a result, the degree of climate match in the WRA was set to the default score of +2 for these five species.

In contrast to previous WRA analyses that have not used formal models of climate suitability, we quantified the climate match between the native range and the two target regions (NZ and GB) using a global meteorological dataset that gridded the world into $10' \times 10'$ latitude-longitude grid cells (New *et al.*, 2002). Each grid cell has data for the mean, maximum and minimum monthly values for a range of meteorological variables including temperature and precipitation for the period 1961 to 1990. We converted the monthly values into 16 climate parameters that are commonly used in climate matching studies to characterise the climate of a given location (see Duncan *et al.*, 2001, for full list). For each species we identified the grid cells that contained at least one native range occurrence record. For each of the 16 climate variables we then calculated the difference between the value in a native range grid cell and the value in each of the target region grid cells, and divided the difference by the global standard deviation of each variable to generate standard scores. An overall measure of the match between a native range grid cell and the NZ and GB grid cells was computed as the square root of the sum of the squares of the standard scores for each of the 16 climatic variables, divided by 16. The resultant value was then compared to a normal distribution of reference scores that partition the normal distribution into percentage categories based on the area under the normal distribution. Scores within 10% of the mean score are those with a close climate match, while scores of 80% or more, which fall in the tails of the distribution, have the lowest climate match. We then selected the lowest score for each NZ and GB grid cell (i.e. we identified the native range grid cell with the closest match) and used this as a measure of climate match for that grid cell. To produce an overall value for the climate match for each species, we calculated the mean of all grid cells for that species in NZ and the mean for GB and subtracted the mean values from 100 to ensure that higher values indicated a better match. Each species' mean climate match score was then assigned to one of three categories ($0 - 32\% =$ “low”,

$33 - 65\%$ = “intermediate”, $66 - 100\%$ = “high”) that matched WRA categories for question 2.01 (degree of climate match).

2.3.4 Statistical analysis

For each target region, we used the numerical WRA score as an explanatory variable for three response variables: a) whether a species in the global pool was introduced or not (binary score); b) whether, having been introduced, a species had naturalised or not (binary score); and c) the distribution range of a species once naturalised having accounted for date of introduction (interval scale).

Logistic regression was used to model the two binary response variables, and their performance assessed using four measures of goodness of fit: accuracy (the proportion of species that succeeded that the model predicted would succeed); reliability (the proportion of species the model predicted would succeed that actually succeeded); area under the receiver operating characteristics curve (AUC); and the Kappa statistic. To calculate accuracy and reliability, a value of 0.5 was used as the threshold for classifying species as predicted to succeed or fail from the model probabilities. Rather than specifying a threshold for converting predicted probabilities into either successes or failures, AUC provides a measure of how well the model discriminates success and failure across all possible thresholds. An AUC value of 0.5 indicates a model has no ability to discriminate among classes (i.e. it performs no better than chance), a value of 1 indicates a model always correctly assigns success a higher probability than failure, while a value of -1 indicates the opposite. Hosmer & Lemeshow (2000) suggest interpreting AUC values as follows: $0.7 \leq \text{AUC} < 0.8$ = acceptable discriminatory power; $0.8 \leq \text{AUC} < 0.9$ = excellent; $0.9 \leq \text{AUC}$ = outstanding. Kappa is a measure of the difference between observed agreement and agreement expected by chance, standardised to a scale of -1 to 1. Kappa can be interpreted as follows (Landis & Koch, 1977): < 0 no better than chance; $0.01 - 0.20$ slight agreement; $0.21 - 0.40$ fair agreement; $0.41 - 0.60$ moderate agreement; $0.61 - 0.80$ substantial agreement; $0.81 - 0.99$ almost perfect agreement.

Multiple regression was used to determine whether WRA score could predict the number of locations within a region in which pine species had naturalised. We included date of introduction as a covariate in the model to control for the effect of residence time because species that have been resident for longer have had more time to spread to new locations (e.g. Castro *et al.*, 2005; Pyšek *et al.*, 2009b).

To assess how well the WRA performed in the absence of knowledge about the invasive behaviour of a species elsewhere, we repeated the WRA assessment for each species but answered as "unknown" the following questions, identified by Caley & Kuhnert (2006) and Weber *et al.* (2009) as important: 2.05 (history of repeated introductions outside its native range); 3.01 (naturalised beyond native range); 3.02 (garden/amenity/disturbance weed); 3.03 (agricultural/forestry/horticultural weed); 3.04 (environmental weed); and 3.05 (congeneric weed). We then repeated the statistical analyses described above using these modified WRA scores.

All analyses were undertaken in R (R Development Core Team, 2010) and we used the packages ROCR (Sing *et al.*, 2009) to calculate accuracy, reliability and AUC, and vcd (Meyer *et al.*, 2010) to calculate Cohen's Kappa statistic.

Table 2.1. Summary of WRA results for all *Pinus* species ($n = 115$ in NZ and $n = 114$ in GB) classed by whether they were not introduced, introduced and naturalised in New Zealand and Great Britain. Values include: WRA score, number of questions answered, number of species recorded as naturalising elsewhere, WRA classification after the secondary screening procedure (accept, evaluate or reject), mean introduction date, number of species at each stage that are forestry species, and mean climate match score for the stage. Differing upper case letters indicate significantly different mean values determined by a Wilcoxon Signed-Rank tests where $P < 0.05$

		Not introduced	Introduced	Naturalised
New Zealand	WRA score (±SE)	-0.62 (0.42)^{A}	5.46 (0.76)^{B}	14.15 (0.93)^{C}
	Number questions answered (±SE)	29.02 (0.78)^{A}	34.45 (0.43)^{B}	37.15 (0.72)^{B}
	Recorded naturalised elsewhere (% ±SE)	5 (10.20)	40 (60.61)	13 (100.00)
	Accept (% ±SE)	37 (75.55)	20 (30.30)	0
	Evaluate (% ±SE)	11 (22.45)	15 (22.73)	0 (0.00)
	Mean introduction date	-	1892^{A}	1869^{B}
	Forestry species (% ±SE)	-	44 (66.67)	13 (100.00)
	Climate match (±SE)	37.78 (1.98)^{A}	59.35 (1.95)^{B}	77.57 (1.82)^{C}
	Total number of species	49	66	13
Great Britain	WRA score (±SE)	-1.63 (0.42)^{A}	5.14 (0.77)^{B}	14.14 (0.94)^{C}
	Number questions answered (±SE)	29.15 (0.75)^{A}	34.96 (0.45)^{B}	38.00 (1.11)^{B}
	Recorded naturalised elsewhere (% ±SE)	4 (9.76)	38 (52.10)	7 (100.00)
	Accept (% ±SE)	36 (87.81)	28 (38.36)	0 (0.00)
	Evaluate (% ±SE)	5 (12.20)	11 (15.07)	0 (0.00)
	Reject (% ±SE)	0 (0.00)	34 (46.57)	7 (100.00)
	Mean introduction date	-	1834^{A}	1760^{B}
	Forestry species (% ±SE)	-	11 (15.07)	7 (100.00)
	Climate match (±SE)	44.29 (2.34)^{A}	67.53 (1.77)^{B}	88.07 (0.89)^{C}
	Total number of species	41	73	7

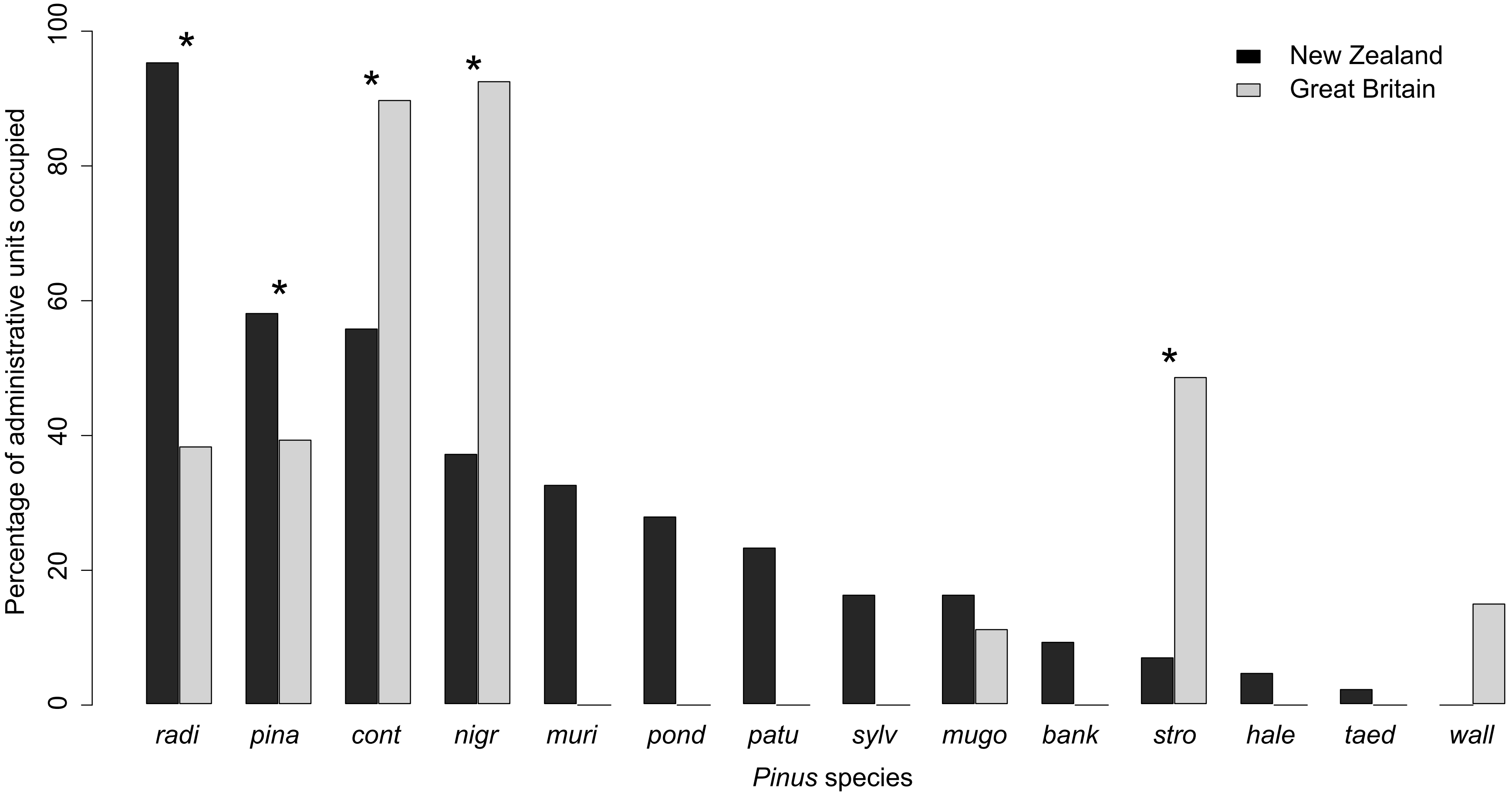

Figure 2.1. The percentage of administrative units with known occurrence of naturalised *Pinus* species in New Zealand (black bars, $n = 43$ regions) and Great Britain (grey bars, $n = 107$ regions). For the seven species present in both countries except for *P. mugo*, the proportion of regions occupied is significantly different ($P < 0.05$ and marked with *) between New Zealand and Great Britain, as determined by two-proportion Z-tests for each species comparison. Species are coded: radi = *P. radiata*; pina = *P. pinaster*; cont = *P. contorta*; nig = *P. nigra*; muri = *P. muricata*; pond = *P. ponderosa*; patu = *P. patula*; sylv = *P. sylvestris*; mugo = *P. mugo*; bank = *P. banksiana*; stro = *P. strobus*; hale = *P. halepensis*; taed = *P. taeda*; wall = *P. wallichiana*

2.4 Results

2.4.1 Introduction, naturalisation and invasion

From the global pool of 115 *Pinus* species, a similar number were scored as potentially invasive (WRA score > 6) in NZ (32 species) and GB (34 species). With the exception of *P. henryi* in NZ, all pines in the global pool assessed as potentially invasive have been introduced into each of the two target regions. In both regions, species that had been introduced had a higher mean WRA score than those not introduced (Table 2.1), with species rejected by the WRA more likely to be introduced than those accepted (χ^2 test, $\chi^2 = 29.50$ and 28.55 for NZ and GB respectively, both $df = 1$, $P < 0.0001$). Date of introduction was strongly negatively correlated with WRA score in NZ (Spearman's rank correlation $r = -0.519$, $P < 0.001$), such that species scored as more invasive in the WRA tended to be introduced earlier. This relationship was weaker, but still negative, for GB (Spearman's rank correlation $r = -0.107$, $P = 0.527$). In both regions, species that had naturalised had, on average, earlier dates of introduction (Table 2.1). Similar numbers of pines were introduced into NZ (66 species) as in GB (73 species). Although the naturalisation rate in NZ was almost twice that of GB, the proportion of species naturalising was not significantly different (20% and 10% for NZ and GB respectively, Z-test, $Z = 1.692$, $P = 0.216$), probably due to small sample size. Six species had naturalised in both regions, all associated with forestry, but each species' rank order of distribution in NZ and GB was significantly different, except for *P. mugo* (Figure 2.1). Nevertheless, for these six species, there was no tendency for them to be, on average, more widespread in NZ than GB ($44.95 \pm 13.20\%$ vs. $53.27 \pm 13.01\%$ in NZ and GB respectively; paired t-test, $t = 0.48$, $df = 5$, $P = 0.652$).

2.4.2 Utility of weed risk assessment for screening *Pinus* species

The WRA scores of the 65 species introduced to both regions were highly correlated (Spearman's rank correlation $r = 0.96$, $P < 0.001$). Since only four questions are regionally specific 2.01, 2.02, 2.04 (relating to climate match) and 8.05 (effective natural enemies present), variation could arise only through region specific scores related to these questions. Mean climate match scores between NZ and GB were highly correlated (Spearman's rank correlation $r = 0.92$, $P < 0.001$). In both regions climate match scores were significantly different between invasion stages (Table 2.1), with naturalised species having the best climate match and species that were not introduced having the worst match. The number of regions occupied was significantly positively correlated with mean climate match in NZ (Spearman's rank correlation $r = 0.62$, $P = 0.03$), but not significantly related in GB (Spearman's rank correlation $r = -0.71$, $P = 0.09$).

The WRA would have rejected slightly under half of the pines that were introduced to both NZ and GB, while it would have rejected all of the introduced species that have naturalised in either region (Table 2.1). Moreover, the WRA score was a highly significant predictor of both introduction and naturalisation success in both regions (Table 2.2). The AUC scores rated the discrimination between successful and unsuccessful species at both the introduction and naturalisation stages in both regions as "excellent" (Table 2.3). Reliability of the WRA was high for the introduction stage in both regions (≥ 0.97) but poor for the naturalisation stage (≤ 0.42, Table 2.3) due to the high proportion of false positives in NZ (42%) and GB (39%). Accuracy was generally lower than reliability at the introduction stage in both regions (≤ 0.66), and higher than reliability at the naturalisation stage (≥ 0.63). Kappa statistics suggested that the WRA was moderately to substantially correct at both stages in both regions (Kappa > 0.36), except at the naturalisation stage for GB (Kappa $= 0.12$) where WRA performance appeared to be only slightly better than chance (Table 2.3). The number of naturalisations in GB was small, with only seven species from a pool of 73, so this low value of Kappa may be due to properties of the Kappa statistic, which is known to be sensitive to low prevalence.

Dates of introduction for the naturalised species ranged from 1830 to 1974 in NZ, and from 1500 to 1930 in GB, reflecting NZ's relatively recent colonisation by Europeans. The number of locations occupied was only weakly associated with introduction date in NZ, and showed no clear relationship in GB (Table 2.4). Having controlled for any effect of introduction date, WRA score was not a significant predictor of the number of locations occupied in either region. These results should be interpreted cautiously due to the small number of naturalised species in both NZ ($n = 13$) and GB ($n = 7$).

Logistic regression models showed that the revised WRA score (excluding information on invasive behaviour elsewhere) was still a significant predictor of introduction success in both countries (Table 2.2). However, the revised WRA score was only significant at the naturalisation stage for NZ, and the degree of significance was much less than with the original WRA score that included information on prior invasiveness (Table 2.2). All other measures of WRA performance indicated that the revised WRA score performed worse than the original WRA (Table 2.3; see Appendix D for full WRA results). The revised WRA failed to explain significant variation in the number of locations occupied in both regions (Table 2.4). This poorer performance reflects the fact that all naturalised species in both countries had records of invasion elsewhere and that this characteristic was a powerful means to discriminate these species.

2.5 Discussion

Our results show that the WRA had high accuracy but low reliability when applied to a group of closely related species that share many attributes in common, and that this performance was repeatable across regions. The WRA correctly identified as invasive 100% of the introduced pine species that had naturalised in both NZ and GB; a higher proportion than previous tests (e.g. Dawson *et al.*, 2009b; Gordon *et al.*, 2008a). Unfortunately, the high number of false positives at the naturalisation stage suggests this accuracy comes at a cost of rejecting potentially useful species. Furthermore, whilst the WRA performed well in discriminating among successful and unsuccessful species at the introduction and naturalisation stages, it performed poorly at predicting the subsequent distribution of species in each region. This finding appears to contrast with Dawson *et al.* (2009b), who found that WRA score correlated with how widespread a species had become at a landscape scale. However, Dawson *et al.* (2009b) examined spread from relatively localised propagule sources, whereas our study measured distribution over a larger region in which naturalised populations most likely arose from multiple sources (Aikio *et al.*, 2010). In these circumstances, how widespread a species becomes may be more strongly linked to the availability of propagule sources resulting from differences in planting effort for forestry. Thus at the landscape scale, attributes related to natural dispersal may be important in determining spread (Dawson *et al.*, 2009a), but at much larger spatial scales human influenced propagule pressure may shape the pattern of invasion. Nevertheless, we used a fairly coarse measure of distribution in both regions (the proportion of occupied administrative units) and this, coupled with the relatively small numbers of naturalised species, undoubtedly limited our ability to detect differences.

We found that the ability of the WRA to discriminate among successful and unsuccessful species in the first two stages of the invasion process was reduced when information about invasive behaviour elsewhere was excluded. This highlights a potential limitation of the WRA: that its performance depends critically on obtaining

Table 2.2. Results of logistic regression models with WRA score as the explanatory variable at the introduction and naturalisation stages for *Pinus* species in New Zealand and Great Britain, with and without information on prior invasion history included in the WRA assessment. Significant results ($P < 0.05$) are highlighted in bold.

	New Zealand				Great Britain			
	Estimate	SE	z	P	Estimate	SE	z	P
Invasion history included								
Introduction	0.294	0.064	4.631	<**0.001**	0.317	0.072	4.392	<**0.001**
Naturalisation	0.612	0.182	3.356	<**0.001**	0.319	0.115	2.783	**0.005**
Invasion history excluded								
Introduction	0.278	0.089	3.114	**0.002**	0.333	0.094	3.548	<**0.001**
Naturalisation	0.344	0.161	2.143	**0.032**	0.265	0.187	1.414	0.157

Table 2.3. Performance measures for logistic regression models with WRA as the explanatory variable in discriminating between (I) introduced and not introduced species, and (N) introduced and naturalised species, both including and excluding information on prior invasion history, for *Pinus* in New Zealand and Great Britain.

		Prior invasion history			
		Included		Excluded	
		I	N	I	N
New Zealand	AUC	0.80	0.97	0.67	0.70
	Kappa	0.36	0.71	0.26	−0.03
	Accuracy	0.69	0.73	0.43	0.79
	Reliability	0.97	0.42	1.00	0.00
Great Britain	AUC	0.83	0.91	0.73	0.65
	Kappa	0.47	0.12	0.27	0.00
	Accuracy	0.66	0.63	0.37	0.92
	Reliability	1.00	0.21	1.00	1.00

knowledge of a species' prior invasive behaviour (Caley & Kuhnert, 2006; Weber *et al.*, 2009). It implies that, within a group of closely related species that share many attributes in common, invasive behaviour elsewhere may be the key attribute distinguishing species with high WRA scores that are likely to naturalise from the rest, and that reliable assessments are feasible only for species with a known history of introduction. On this basis it could be argued that full WRA assessments will waste resources if data on naturalisation elsewhere is the key criterion defining invaders, and that time could be saved by focussing only on this aspect of the assessment. We found, however, that in the absence of information about invasive behaviour the WRA still distinguished between successful and failed species at the introduction (NZ and GB) and naturalisation (NZ only) stages with higher accuracy than expected by chance (Table 2.3). A full WRA assessment can thus provide useful information about the risk posed by species whose invasive behaviour is unknown because they have not been introduced outside their native range. However, such a conclusion is highly dependent on there being no costs to rejecting "false positives", which may not be true where these species are commercially valuable, such as in the case of forestry (Hulme, 2012).

The pool of introduced species was significantly biased towards more invasive species in each country, and had the WRA been implemented to screen *Pinus* prior to importation in the past it would have rejected the most high risk taxa. This bias arises from the fact that conifers selected for forestry often share many characteristics (such as fast growth rate, typically associated with short generation time and high

Table 2.4. Multiple regression results for all naturalised pine species with number of regions naturalised as the response variable, and both WRA score and introduction date as the explanatory variables. Shown are the slopes for both predictors, along with the standard errors, z values, and P values. Overall model results for WRA including information on prior invasion history: New Zealand (Multiple $R^2 = 0.274$, $F = 1.89$, $P = 0.201$), Great Britain (Multiple $R^2 = 0.007$, $F = 0.011$, $P = 0.989$); and excluding information on prior invasion history: New Zealand (Multiple $R^2 = 0.378$, $F = 3.047$, $P = 0.093$), Great Britain (Multiple $R^2 = 0.157$, $F = 0.280$, $P = 0.774$)

	New Zealand				Great Britain			
	Estimate	SE	z	P	Estimate	SE	z	P
Invasion history included								
Introduction	−0.279	0.146	−1.905	0.086	−0.008	0.149	−0.056	0.959
Naturalisation	0.571	0.934	−0.611	0.555	0.762	6.823	0.112	0.918
Invasion history excluded								
Introduction	−0.286	0.135	−2.115	0.061	−0.175	0.253	−0.692	0.539
Naturalisation	−2.385	1.640	−1.454	0.177	−13.179	17.792	−0.741	0.513

reproductive output) associated with invasive behaviour (Essl *et al.*, 2011, 2010). In the case of NZ and GB, around half of all pines were introduced for forestry purposes, including all of the species that subsequently naturalised. This contrasts with other studies that have found forestry species to have lower naturalisation rates than other introduction pathways (Lambdon *et al.*, 2008). Potential conflicts may therefore arise between forestry and environment sectors where screening tools such as WRA would likely limit or prevent introduction of the most economically valuable conifer species. Whether the reduction in WRA score for introductions over time is indicative of a change in introduction policy in each region or simply that the most invasive species had already been introduced, is unclear.

In addition to selection for traits that may increase the chance of naturalisation and invasion, forestry may compound the problem through initial cultivation, provenance selection and the widespread planting of high risk species with good climate match, which could further facilitate their spread (e.g. Procheş *et al.*, 2012). Indeed, all of the pine species with WRA scores of six or more (those the WRA would have rejected for importation) have already been introduced to both countries, with the exception of *P. henryi* in NZ. We would recommend this species not be considered for introduction to NZ. Nevertheless, WRA scores were not good predictors of pine distribution in either NZ or GB. While species selected for forestry tend to have higher WRA scores, among this group higher scores may not translate to wider planting and a potentially wider naturalised distribution.

This study is the first to explicitly use climate matching in the WRA analysis and, as a result, identified the critical role climate plays in the introduction, naturalisation and, at least for NZ, the spread stage of invasion. Essl *et al.* (2011) found that simple climate parameters (mean annual temperature and precipitation) were insufficient to predict the global success of conifer naturalisation and a more integrated measure, such as zonobiome, was necessary to capture the suitability of the recipient environment. Our results suggest that incorporation of more detailed climate variables may indeed usefully discriminate between naturalisation success or failure. While formal climate matching has been advocated in WRA (Pheloung *et al.*, 1999) this has not been undertaken: default scores have been applied (Gordon *et al.*, 2008a), mean climate parameters (Křivánek & Pyšek, 2006), or latitudinal ranges (Dawson *et al.*, 2009b) compared or climate classification maps have been used (Jefferson *et al.*, 2004). Our results indicate that formal climate matching should be an integral element of WRA and that studies failing to do so should be viewed as incomplete.

Such formal climate matching can future-proof WRA assessments by building in changes in risk under predicted future climate change (Walther *et al.*, 2009).

Global assessments of conifer invasions indicate that regions of the southern hemisphere are more invaded than those of the northern hemisphere (Essl *et al.*, 2011, 2010; Richardson & Rejmánek, 2004; Simberloff *et al.*, 2010). GB is somewhat unusual for the northern hemisphere, having few native conifers and a forestry sector based on introduced species. Indeed, by northern hemisphere standards GB has a large number of naturalised conifers (Carrillo-Gavilàn & Vilà, 2010; Essl *et al.*, 2011, 2010). The strong correlation among WRA scores between the two regions suggests that there was little difference in the potential for species to invade each region, possibly due to the broadly equivalent climates. We highlighted many similarities between NZ and GB in the patterns of introduction of pines; of the 73 species introduced, 61 species were introduced to both regions. However, we found that almost twice the number of pines had naturalised in NZ than in GB (although this was not statistically significant, potentially due to the small number of naturalising species). Thus any differences between regions may be more related to the intensity of planting rather than inherent species traits or climate characteristics of the recipient region (also see Essl *et al.*, 2011, 2010). This is probably the reason we found no correlation in the distribution of the six pine species shared between both regions.

2.6 Conclusions

Overall, we have shown that the WRA is able to discriminate among closely related species at the introduction and naturalisation stages of the invasion process, but not at the spread stage, and that these findings are repeatable across regions. Nevertheless, the outcome at the critical naturalisation stage was dependent on prior knowledge of a species invasive behaviour elsewhere; without this information the WRA may be unable to reliably distinguish among closely related species as to their naturalisation potential, and should therefore be used cautiously where data on invasive behaviour elsewhere is lacking. We also document the potential conflict between forestry usage and pine naturalisations in NZ and GB. Finally, our results highlight the key role that humans can play in the first stage of the invasion process through the preferential introduction of species with a high risk of invasive behaviour in the first place.

2.7 Author contributions and publication details

Author contributions were as follows: KFM, RPD and PEH designed the research. KFM undertook the data collection and statistical analyses with normal supervisory guidance from RPD, PEH and MSW. KFM wrote the manuscript, with comments and feedback from RPD, PEH and MSW. Dave Richardson and two anonymous reviewers provided comments that improved the manuscript. This manuscript was published in *Biological Invasions* **14**, 987–998 (2012).

Chapter 3

What determines pine invasions: species traits, climate suitability or forestry use?[1]

3.1 Abstract

AIM: Species attributes, biogeographic features and human factors have all been shown to discriminate between invasive and non-invasive plant species. However, the relative importance of these factors, their generality in determining invasion outcomes across different regions, and their ability to discriminate success and failure at different stages of the invasion process, have not been established.

LOCATION: New Zealand (NZ) and Great Britain (GB).

METHODS: For species in the genus *Pinus*, we used boosted regression trees to identify factors associated with success or failure at the introduction and naturalisation stages in each region.

RESULTS:Human factors, notably the forestry use index, were the strongest determinants of which species from the global pool were introduced to both NZ and GB. Species with a close climate match were also more likely to be introduced to NZ but not to GB. Human factors and climate match were also the strongest determinants of which introduced species became naturalised in both NZ and GB, although the order of importance differed (human factors followed by climate match for NZ and vice versa for GB). Species attributes (life-history traits and the Z-score), had much less ability to discriminate successful and failed species at both the introduction and naturalisation stages in these two regions.

MAIN CONCLUSIONS: We show for the first time that human factors are more important than either species or biogeographic traits in determining the likelihood

---

[1] McGregor, K. F., Watt, M. S., Hulme, P. E. & Duncan, R. P. (2012) What determines pine naturalisation: species traits, climate suitability or forestry use? *Diversity and Distributions* **18**, 1013–1023.

of a species being introduced or becoming naturalised. The similarity between two different regions in the factors found to be important in success at both these invasions stages points to potentially general mechanisms underlying these processes. The strong human component to introduction and naturalisation highlights a potential conflict between future afforestation using alien species with conservation and management aims in the surrounding landscape, given that the factors desirable for forestry species are also those that may promote invasion.

Keywords: alien species; biological invasions; climate match; forestry; *Pinus*; propagule pressure; traits, weeds.

3.2 Introduction

Worldwide, pines (*Pinus* spp.) have been extensively used in plantation forestry in areas outside their native ranges (CAB International, 2010), yet they are also recognised as some of the most widespread and significant invasive plants in the world (Richardson & Higgins, 1998; Richardson & Rejmánek, 2011). It is therefore essential to understand the determinants of pine invasion to ensure forestry practices do not pose a risk of introducing further invasive species. Three key life-history traits (short juvenile period, small seeds and frequent reproduction) have been tightly linked to invasion outcomes globally and have been suggested as a suitable screening mechanism for sustainable forestry using *Pinus* spp. (Grotkopp *et al.*, 2002; Rejmánek & Richardson, 1996; Richardson & Rejmánek, 2004; Richardson *et al.*, 1994). However, recent evidence also points to the interplay between life-history traits and both biogeographic and human factors as key determinants of the likelihood that a pine species will become invasive. Biogeographic factors such as native range size (Procheş *et al.*, 2012) and climate match (Nuñez & Medley, 2011), and human factors such as propagule pressure (Boulant *et al.*, 2009; Gassò *et al.*, 2010; Křivánek *et al.*, 2006; Medawatte *et al.*, 2010), residence time (Richardson *et al.*, 1994) and economic use (Essl *et al.*, 2011, 2010) have all been identified as important correlates of pine invasion success. Furthermore, life-history traits on their own do not appear to explain the apparently greater success of pines in invading regions of the southern hemisphere relative to the northern hemisphere (Carrillo-Gavilàn & Vilà, 2010; Essl *et al.*, 2010; Simberloff *et al.*, 2010), which further suggests that the interaction between species attributes, biogeographic factors and human use may be critical in explaining invasion outcomes globally (Richardson *et al.*, 1994). This complex set of interactions underpinning pine invasions is likely to have parallels in other woody genera known to have naturalised around the world e.g. *Acacia* (Richardson *et al.*, 2011), *Abies*, *Cupressus* (Essl *et al.*, 2010) and *Eucalyptus* (Rejmànek & Richardson, 2011). Thus applying methods to tease apart the relative contribution of species attributes, biogeographic and human factors in plant invasions is an essential step in developing sustainable forestry practices (Chornesky *et al.*, 2005; Moore, 2005).

Recent evidence has highlighted that invasive conifers worldwide are biased towards species used in commercial forestry (Essl *et al.*, 2011, 2010) implying an important role for this sector in the dissemination of invasive species. Yet such analyses have not taken into account the many species deliberately introduced that have never naturalised, which might paint a different picture of the risks posed by commercial forestry (Hulme, 2012). The initial introduction stage of most invasions is poorly studied because there is often no reliable record of the species that were introduced but failed to establish (Diez *et al.*, 2009; Puth & Post, 2005). Nevertheless, this stage is important because non-random patterns of introduction may result in the pool of introduced species being a biased subset with regard to key species attributes and life-history traits, potentially biasing our perception of which factors are important in determining naturalisation outcomes (Blackburn & Duncan, 2001; Cassey *et al.*, 2004; Lambdon *et al.*, 2008). In contrast to the many studies that have examined correlates of naturalisation and invasion in pines, an appreciation of the role of biases at the introduction stage in shaping invasion outcomes is only just emerging (McGregor *et al.*, 2012).

Given this background, our aim was to assess the relative roles of species attributes (life-history traits and the Z-score), biogeographic and human factors on the likelihood of introduction, naturalisation and invasion of pines in New Zealand and Great Britain. New Zealand (NZ) and Great Britain (GB) are island regions of similar size and broadly similar climates, for which pines are economically important, resulting in a well-documented history of introductions, including records of which species were introduced, their residence time, and how widely different species were planted for forestry. These two countries are the major foci of pine introductions and naturalisations in the southern and northern hemispheres respectively (Carrillo-Gavilàn & Vilà, 2010; Essl *et al.*, 2010; Richardson & Rejmánek, 2004), and thus provide an opportunity to test whether similar processes underpin the patterns of introduction and invasion in both regions (McGregor *et al.*, 2012).

3.3 Materials & Methods

3.3.1 Study genus

We used the list of all 115 species in the genus *Pinus* compiled by McGregor *et al.* (2012), which is based on recent taxonomic treatments of the genus excluding recently described species that are not widely recognised in the literature along with subspecies/varieties (see Appendix A). One pine species is native to one of our study locations: *P. sylvestris* is native to GB and we excluded this species from the GB analysis.

3.3.2 Introduction, naturalisation and invasion histories

We identified which pine species had been introduced to NZ, and which had been introduced to GB (excluding all off-shore islands), and the date of first recorded introduction to each region from historical records that included the horticultural, forestry, and scientific literature (see Appendix A). Within each region, an introduced species was classed as naturalised if it had established new self-sustaining populations outside of cultivated areas (sensu Richardson *et al.*, 2000b); and invasive if it produced reproductive offspring often in large numbers, at considerable distances from the parent plants (scale: > 100m; < 50 years; sensu Richardson *et al.*, 2000b). Using these definitions, all naturalised pine species in NZ were classed as invasive, and in the UK, *P. contorta*, *P. nigra*, and *P. pinaster* were classed as invasive. It was not possible to test the transition from naturalisation to invasion in either region, because in NZ there was no source pool for comparisons to be made, and in GB there were too few species to provide robust results. Therefore, we examine the introduced to naturalised transition, and for simplicity refer to all species succeeding in this transition as having naturalised.

3.3.3 Species attributes

We examined two classes of species attribute: life-history traits and the Z-score. For each species we recorded a suite of life-history traits (Table 3.1) by searching standard reference texts (Farjon, 2005; Perry, 1991; Price *et al.*, 1998; Richardson & Bond, 1991; Richardson, 1998b), scientific literature, online databases (e.g. Earle, 2008; GBIF, 2011; IPIN, 2004; USDA, 2011), and horticultural books and websites. We chose these traits for three reasons. First, we recorded the traits required to calculate the Z-score, which uses seed mass, minimum juvenile period and interval between large seed crops to derive an index that has been shown to reliably distinguish invasive from non-invasive pine species (Grotkopp *et al.*, 2002; Rejmánek & Richardson, 1996). Second, we hypothesised that factors related to the desirability of a species for forestry, such as diameter at breast height (DBH), height and life-span (fast growing species with short or intermediate life-spans may be more suited to commercial forestry than long lived slow growing species) should influence the likelihood of introduction, which may in turn influence relationships at the naturalisation stage. Third, we collected traits linked to dispersal ability, including dispersal mode, seed wing length, serotiny (a species with serotinous cones may be less likely to be dispersed in NZ and GB because neither regions are highly fire prone), and plant height (taller trees being more likely to disperse propagules further by wind [Thomson et al. 2011]) because these may influence the ability of plants to colonise areas outside of cultivation. We limited our study to traits that could be obtained for most species. Although pines are a well studied genus, additional traits that have been linked to invasion success, such as relative growth rate and specific leaf area (Grotkopp *et al.*, 2002) were available only for a much smaller subset of species.

While we chose life-history traits that we considered readily obtainable, our life-history data nevertheless had 20% of entries missing (Table 3.1). To handle missing life-history trait data we used imputation rather than case-wise deletion because the latter results in information loss, reduced statistical power, and potentially biased estimates of parameters (Rubin, 1976). These problems become more important if the data are missing because of some underlying reason linked to the biology of the species or are difficult to measure (Nakagawa & Freckleton, 2011). To prevent imputation of negative values, we first added one to all entries for variables that contained true zeros and then log-transformed all continuous variables (Table 3.1). We performed imputation using the `mix` package (Schafer, 2010) in R (R Development Core Team, 2010), specifying an unrestricted general location model (which uses maximum likelihood estimates for the parameters) with a uniform prior. After log-transformation some variables were not normally distributed, but simulations by Graham & Schafer (1999) showed that imputation performs robustly even for a set of highly non-normal variables under an assumption of multivariate normality.

Life-history traits in plants are frequently correlated such that plants exhibit trait syndromes (Grime *et al.*, 1997). For the 115 pine species, many of the traits we measured were strongly correlated with each other (see Appendix E). In order to reduce this collinearity, we selected eight traits to include in the analysis that we hypothesised to be important in pine invasions, and that were not highly correlated ($r < 0.60$) with each other. These included six continuous traits: diameter at breast height (DBH), seed mass, interval between seed crop years, minimum juvenile period, life-span, seed wing length; and two ordinal traits: dispersal mode (wind/animal/both) and serotiny (serotinous/non-serotinous/both).

The Z-score has been identified as an integrative measure of species life-history that is able to discriminate among invasive and non-invasive pine species (Grotkopp *et al.*, 2002; Rejmánek & Richardson, 1996). We calculated Z-scores for each species using the values for seed mass, minimum juvenile period, and interval between large seed crop years, using the discriminant function described by Rejmánek & Richardson (1996).

3.3.4 Biogeographic factors

For each species, we quantified native range size, geographic area of origin, and climate match with both NZ and GB. We chose these three variables because aspects of both the native range and climate match have been shown to predict naturalisation success in pines (Nuñez & Medley, 2011).

We collated presence only occurrence records for each species from online databases and published scientific literature (see Chapter 2). We calculated native range size as a convex hull around all native occurrence points (see Chapter 2), which yielded a value for the extent of occurrence (EOO) for each species. For simplicity we refer to EOO as "native range size" throughout the following text.

To quantify climate match between the native range and the two target regions (NZ and GB), we used the automated matching method outlined in Chapter 2 which used the native range occurrence points. This yielded a value for the percentage of grid cells in each country that had $\geq$ 90% match to the native range. Five species (*Pinus hakkodensis*, *P. henryi*, *P. squamata*, *P. stankewiczii* and *P. wangii*) had fewer than three global occurrence records and neither a specific climate match or native range size could be obtained.

We hypothesised that, given the history of British and European exploration of North America and subsequent British colonisation of NZ, species native to North America and Europe were more likely to be introduced to both GB and NZ than species native to either Central America or Asia. We classed species as Asian or European based on an accepted division at the Ural-Caucasus Mountains: if the majority of their native range was west of the division, they were classed as European, if the majority was east, they were classed as Asian.

3.3.5 Human factors

To quantify the extent to which species used widely for forestry purposes were more likely to have been introduced to NZ and GB, we used an index derived by Procheş *et al.* (2012): the number of citations to each species in the CABI Forestry Compendium (CAB International, 2010). Procheş *et al.* (2012) found that this bibliometric measure explained naturalised range size in pines better than other measures such as ISI Web of Science searched using species names and search terms relevant to forestry, or when searches were restricted to forestry journals. The number of CABI Forestry Compendium citations was significantly correlated with the number of experimental plantings in southern Africa, and explained 71% of the variance in a species' global naturalised range size (Procheş *et al.*, 2012). We therefore used this variable as a measure of the utility of a species in worldwide forestry and have subsequently termed this variable "forestry use index".

For each species introduced into a region, we used the total area planted (ha), the duration of planting (in years), and the first date of planting as measures of introduction effort or propagule size. Data were obtained from regional forestry organisations that collate and archive both private and state forestry data in NZ (housed at Scion, Rotorua) and GB (housed at the Forestry Commission, Alice Holt, Surrey). For each species in each region we calculated the area planted by plotting the number of hectares of standing stock against planting year, and calculated the area under the curve using the areapl function in the splancs library (Rowlingson *et al.*, 2010) in R (R Development Core Team, 2010). To assess the effect of minimum residence time on naturalisation success, we recorded the earliest date of introduction to NZ and GB by searching the scientific and horticultural literature (see Appendix A).

3.3.6 Statistical analysis

All analyses were preformed in R (R Development Core Team, 2010). Because our proposed analytical approach (see below) has some capacity to handle missing values, we included the five species for which we were unable to calculate climate match or native range size.

To quantify the relative importance of variables in explaining outcomes at the introduction and naturalisation stages, we used boosted regression tree (BRT) models. BRT models are a form of regression that use a boosting algorithm to fit many models (hundreds to thousands), resulting in one tree that is a linear combination of many trees, improving the overall model accuracy (e.g. Elith *et al.*, 2008). We used BRTs because they perform well in model discrimination (e.g. Elith *et al.*, 2006; Keller *et al.*, 2011; Schmidt & Drake, 2011a), and because they can readily incorporate non-linear relationships and potential interactions among explanatory variables (see Elith *et al.*, 2008).

For each region we fitted two BRT models at each invasion stage (introduction and naturalisation). First, we used the eight life-history traits, and the biogeographic and human factors as explanatory variables. Since the Z-score is an integrative measure of (and thus highly correlated with) three life-history traits, we fitted a second model that included the Z-score as the only species attribute as well as biogeographic and human factors as explanatory variables. At the introduction stage, our response variable was whether species were introduced or not to each region from the global pool of 115 species. At the naturalisation stage our response variable was whether the species introduced to each region had naturalised or not.

We fitted BRT models with the gbm package using the methods and supplemental functions described in Elith *et al.* (2008). We fitted models with a tree complexity of five (which automatically detected and fitted up to five-way interactions among explanatory variables), a learning rate of 0.001, and a Bernoulli error structure. Because using the same data for model testing and validation leads to overfitting and deflates the estimated error rate, the optimal numbers of trees were determined using ten-fold cross-validation with a bag fraction of 0.5. The relative contribution of predictor variables is estimated in the gbm package (Friedman, 2001). The contribution of a variable is based on the number of times the variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Friedman & Meulman, 2003). The contribution of each variable is scaled so that the sum adds to 100, with higher numbers indicating stronger influence on the response.

Model performance was assessed during cross-validation using the area under the receiver operating curve (AUC, Hanley & Mcneil, 1982), which provides a measure of the degree to which the fitted values discriminate between observed outcomes. An AUC value of 1 indicates that the model perfectly discriminates between outcome 1 (e.g. introduced) and outcome 0 (e.g. not introduced), while a value of 0.5 indicates that the model performs no better than chance. Models containing the full set of explanatory variables were simplified by dropping uninformative variables using the default settings, which allowed model simplification to continue until the average change in predictive deviance exceeded the original standard error (see Elith *et al.*, 2008).

3.4 Results

3.4.1 Determinants of introduction

A similar number of pine species were introduced into each region (NZ: $n = 66$; GB: $n = 73$) with 61 common to both (see Appendix A). The forestry use index (classed as a human factor) had the strongest influence on whether pine species were introduced to a region or not, having a relative contribution of over 50% and 70% in NZ and GB respectively (Figure 3.1). Biogeographic factors had the next largest contribution, with life-history traits contributing relatively little. The relatively small contribution of species attributes in both regions (NZ: 15.7%; GB: 9%) was true whether we included the eight life-history traits or used the Z-score (Figure 3.1). No individual life-history trait made a relative contribution of more than 5% (Table 3.2; Table 3.3). In NZ, a single life-history trait (seed wing length) had a higher contribution than the Z-score to model outcomes, though in GB the Z-score had a higher contribution than any single life-history trait.

Species use widely in global forestry, with a good climate match, large native range size, native ranges in Europe and North America, shorter juvenile period and larger DBH had a higher probability of being introduced to NZ or GB (Table 3.2). All human and biogeographic variables acted in the same direction in both regions, but there appeared to be no consistent directional effect among the different life-history traits (Table 3.2). When the Z-score was used in place of life-history traits a similar pattern emerged; human and biogeographic variables had similar effects in both regions but the effect of the Z-score differed between regions. Species with a tendency to naturalise widely (high Z-score) were more likely to be introduced to NZ but the reverse was true for GB (Table 3.3).

Table 3.1. Characteristics of the variables used in analysis: the transformation applied to each variable prior to imputation and analysis; n the number of species for which empirical data were available (life-history trait data for the remaining species were imputed to the full sample size; $n = 115$). Variables were classified as either categorical (cat.) or continuous (cont.). In order to assess the relative influence of all life-history traits, biogeographic, and human factors, all variables were included in all models except: life-history traits marked with a †(which were not included in any models due to multicollinearity); and human factors marked with a $*$ (which were included for the naturalisation stage only).

Category	Variable	Transformation	n	Type	Coding or units
Species	Diameter at breast height (DBH)	log	96	cont.	cm; maximum encountered
	Dispersal mode	-	84	cat.	1 = animal; 2 = wind; 3 = both
	Cone length†	log	108	cont.	cm
	Genome size†	log	41	cont.	pg DNA
	Height†	log	111	cont.	m
	Interval between large seed crop	log + 1	74	cont.	years
	Juvenile period	log	69	cont.	years
	Life-span	log	43	cont.	years
	Mean seed mass	log	88	cont.	g
	Seed length†	log	108	cont.	mm
	Seed wing length	log + 1	110	cont.	mm
	Serotiny	-	94	cat.	1 = yes; 2 = no; 3 = yes/no
	Z-score	-	69	cont.	Higher score = more invasive
Biogeographic	Native range size	log	110	cont.	km^2
	Climate match	-	112	cont.	-
	Geographic origin	-	115	cat.	1 = North America; 2 = Asia; 3 = Europe; 4 = Central America

Continued on next page...

Continued

Category	Variable	Transformation	n	Type	Coding or unites
Human	Forestry use index	log + 1	115	cont.	Number CABI Forestry Compendium references (from Procheş *et al.*, 2012)
	Area planted*	log + 1	115	cont.	hectares
	Date of first state forestry planting∗	-	-	cont.	years
	Duration of state forestry planting∗	-	-	cont.	years
	Minimum residence time∗	-	-	cont.	years
Response variables	Introduced	-	115	cat.	1 = yes; 0 = no
	Naturalised	-	115	cat.	1 = yes; 0 = no

3.4.2 Determinants of naturalisation

Almost twice as many introduced pine species had naturalised in NZ than in GB (13 and 7 respectively, with six species naturalised in both regions). The relative importance of species attributes, biogeographic and human factors on naturalisation differed between regions. In both regions, life-history traits had the least influence on naturalisation success (maximum relative contribution NZ: 12.3%; GB: 22.6%). No individual life-history trait made a relative contribution of more than 10% (Table 3.2; Table 3.3). In NZ, a single life-history trait (minimum juvenile period) had a higher contribution than the Z-score to model outcomes though in GB the Z-score had a higher contribution than any single life-history trait. In NZ, human factors had the greatest influence on naturalisation success with a relative contribution $\sim 50\%$, whereas in GB biogeographic variables had the greatest influence with a relative contribution $\geq 51\%$ (Figure 3.1).

In both regions, species with greater forestry use index were more likely to naturalise (maximum relative contribution NZ: 21.1%; GB: 23.6%). In NZ species planted over a wider area were also more likely to naturalise, but this was not the case is GB. In both regions, species with a closer climate match, larger native range size, longer residence time, shorter juvenile period, smaller seed mass, and smaller DBH, were more likely to naturalise (Table 3.2). In GB longer lived species originating from Europe and North America were more likely to naturalise than species from Asia and Central America. In both regions, species with a higher Z-score (more invasive) were more likely to naturalise than species with a low Z-score, and the relative contribution of Z-score was similarly low in both regions (Table 3.3).

To explore whether the differences in the correlates of introduction and naturalisation success found for NZ and GB simply reflected the different species composition of pines in each region we repeated the analyses described above with only the species that had been introduced to both regions ($n = 61$). The results from this analysis were similar to those using all species in the genus and the quantitative conclusions did not change (results not presented). Thus differences in regions are likely due to other factors e.g. forestry practices, landscape structure rather than being attributable to individual species.

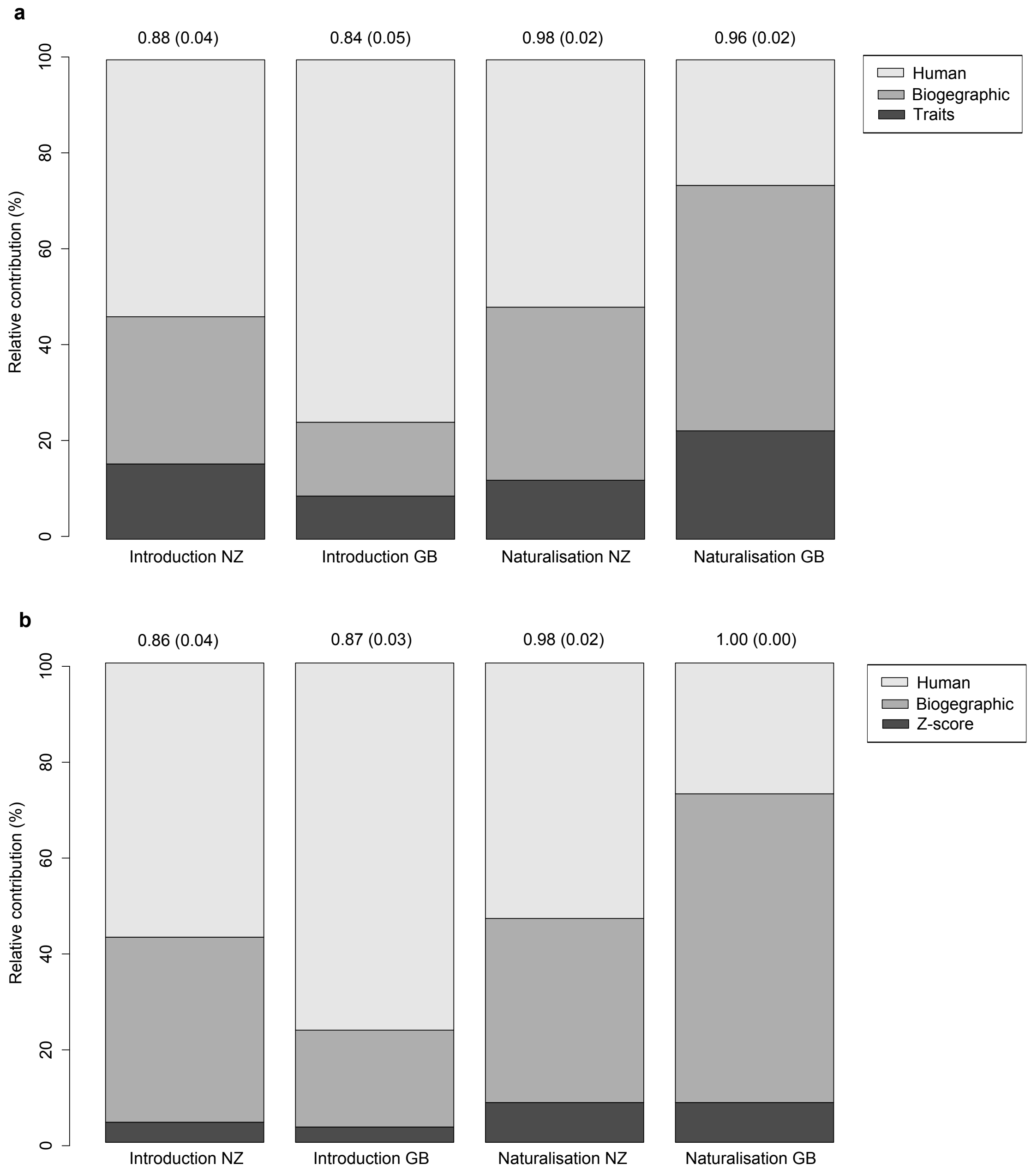

Figure 3.1. Stacked bars showing the percent relative contribution of variables in three classes (human, biogeographic and species attributes) to introduction and naturalisation success for *Pinus* spp. in New Zealand (NZ) and Great Britain (GB) when: (a) life-history traits are used; (b) when the Z-score is used in place of life-history traits. AUC (±SE) values indicate the performance of the model at each stage are given above the stacked bars.

Table 3.2. Results of boosted regression tree (BRT) models at the introduction and naturalisation stages for the invasion of *Pinus* in New Zealand (NZ) and Great Britain (GB), when life-history traits were included as explanatory variables. Contributing variables highlighted in bold were retained after model simplification following Elith *et al.* (2008). n.a. indicates variables that were not included in the models at a given stage. For an individual tree created in the BRT model, a variables' relative importance is the sum of squared improvements at all splits determined by it. A variables' overall relative importance (shown here) is the average of these values over all trees, expressed as a percentage. The direction of the effect of each continuous variable is indicated in brackets. AUC (±SE) values indicate model performance before model simplification

		Introduction		Naturalisation	
Variable class	Explanatory variable	NZ	GB	NZ	GB
Species	Diameter breast height	2.1 (+)	1.0 (+)	0.7 (−)	1.8 (−)
	Dispersal mode	0.1	0.2	0	0.2
	Interval large seed crops	1.7 (−)	1.7 (+)	0.2 (+)	6.5 (+)
	Juvenile period	2.4 (−)	1.6 (−)	**9.5 (−)**	3.6 (−)
	Life-span	3.1 (−)	1.3 (+)	0.8 (−)	5.5 (+)
	Seed mass	1.7 (−)	1.6 (+)	0.4 (−)	2.6 (−)
	Seed wing length	4.3 (+)	1.2 (−)	0.1 (−)	2.2 (+)
	Serotiny	0.2	0.4	0.6	0.2
Biogeographic	Climate match	**25.1 (+)**	1.3 (+)	**29.1 (+)**	**44.4 (+)**
	Geographic are of origin	1.4	**8.4**	0.6	2.1
	Native range size	4.2 (+)	**5.8 (+)**	**6.4 (+)**	4.7 (+)
Human	Forestry use index	**53.6 (+)**	**75.6 (+)**	**21.1 (+)**	**22.0 (+)**
	Area planted	n.a.	n.a.	**19.1 (+)**	2.9 (+)
	First date state planting	n.a.	n.a.	1.2 (−)	0
	Duration state planting	n.a.	n.a.	**6.1 (+)**	0
	Introduction date	n.a.	n.a.	**4.1 (−)**	1.3 (−)
	AUC	0.88 (0.04)	0.84 (0.05)	0.98 (0.02)	0.96 (0.02)

Table 3.3. Results of boosted regression tree (BRT) models at the introduction and naturalisation stages for the invasion of *Pinus* in New Zealand and Great Britain,when the Z-score was used in place of life-history traits. Contributing variables highlighted in bold were retained after model simplification following Elith *et al.* (2008). n.a. indicates variables that were not included in the models at a given stage. For an individual tree created in the BRT model, a variables' relative importance is the sum of squared improvements at all splits determined by it. A variables' overall relative importance (shown here) is the average of these values over all trees, expressed as a percentage. The direction of the effect of each continuous variable is indicated in brackets. AUC (±SE) values indicate model performance before model simplification

		Introduction		Naturalisation	
Variable class	Explanatory variable	NZ	GB	NZ	GB
Species	Z-score	4.1 (+)	3.2 (−)	**8.3 (+)**	8.3 (+)
Biogeographic	Climate match	**29.2 (+)**	**4.8 (+)**	**28.6 (+)**	**53.9 (+)**
	Geographic are of origin	2.4	**8.3**	**1.2**	3.6
	Native range size	**7.1 (+)**	**7.1 (+)**	**8.6 (+)**	6.9 (+)
Human	Forestry use index	**57.2 (+)**	**76.6 (+)**	**19.9 (+)**	**23.6 (+)**
	Area planted	n.a.	n.a.	**21.8 (+)**	2.6 (+)
	First date state planting	n.a.	n.a.	**1.8 (−)**	0
	Duration state planting	n.a.	n.a.	**5.6 (+)**	0
	Introduction date	n.a.	n.a.	**4.2 (−)**	1.1 (−)
	AUC	0.86 (0.04)	0.87 (0.03)	0.98 (0.02)	1.00 (0.00)

3.5 Discussion

Our goal was to assess the relative importance of species attributes (life-history traits and the Z-score), biogeographic, and human factors in determining success at the introduction and naturalisation stages in two regions. An increasing number of studies attempt to assess the correlates of introduction, naturalisation and invasion success (Dawson *et al.*, 2009a; Gravuer *et al.*, 2008). However, we present the first evidence for consistent roles of species attributes, biogeographic, and human factors in introduction and naturalisation success for the same global pool of species in different regions. We found that human factors most strongly influenced the likelihood that a pine species from the global pool would be introduced into either region, followed by a good climate match and large native range size. However, human factors acted differently between regions on naturalisation, being more important in NZ than in GB. In both regions, life-history traits and the Z-score had much less influence on both introduction and naturalisation, being ranked below other factors.

Our results show how bias in the pool of introduced species may affect outcomes in subsequent stages of the invasion process. Pines introduced to NZ and GB were a non-random subset of species in the global pool: they tended to be species that were widely used for forestry elsewhere, had a good climate match to the target region, a large native range and fast maturation (Table 3.2). These are also traits that have been linked to a greater probability of invasion in pines (Grotkopp *et al.*, 2002; Nuñez & Medley, 2011; Procheş *et al.*, 2012; Rejmánek & Richardson, 1996) and other taxa (Gallagher *et al.*, 2011; Gravuer *et al.*, 2008; Hanspach *et al.*, 2008; Hui *et al.*, 2011). The predominance of these traits in the pool of introduced species thus likely increases the base rate of naturalisation and invasion relative to that expected had species been chosen at random. In short, the features of species selected for introduction and commercial planting for forestry, such as good climate match and fast maturation, will in turn favour success at naturalisation and invasion (Essl *et al.*, 2011, 2010).

Our finding that species attributes contribute little to naturalisation success relative to human and biogeographic factors, is contrary to previous studies that identified small seed mass, short juvenile period and short interval between large seed crop years as key factors discriminating invasive and non-invasive pines (Grotkopp *et al.*, 2002; Rejmánek & Richardson, 1996). There are several possible reasons for this discrepancy.

First, previous studies on pines have focused solely on species attributes, and not considered the potential role of biogeographic and human factors (Grotkopp *et al.*, 2002; Rejmánek & Richardson, 1996). Second, it is possible that the forestry use index we used, derived from the CABI Forestry Compendium, was itself highly correlated with traits related to likelihood of naturalisation (and invasion in the case of NZ). This might be expected if foresters biased the pool of commercially used species towards those with specific traits linked to invasion success. However, the only trait with which this index was even weakly correlated was tree height (Spearman's rank correlation, $r = 0.50$, $P < 0.001$). Third, previous studies have analysed only subsets of the genus, typically fewer than 30 species selected on data availability. Our analysis of introduction and naturalisation outcomes in NZ and GB uses all known introduced species (66 and 73 species in NZ and GB respectively) and the inclusion of a wider range of species selected on an objective basis may influence the outcome.

Finally, previous studies have used different measures of "invasion success" to our study and this might lead to differences in the variables identified as important (Hulme, 2012). For example, Rejmánek & Richardson (1996) classified species as invasive if they were spontaneously spreading on at least two continents and as non-invasive if they were planted on at least three continents but never reported as spreading. Invasion success in this context encompasses both naturalisation and subsequent spread in multiple regions, while our study considers naturalisation in specific regions. Other studies have quantified invasion outcomes using the rate or extent of spread within a region and it is possible that life-history traits become more important in influencing these distributional patterns (Gravuer *et al.*, 2008; Lloret *et al.*, 2005). The Z-score has previously been used to discriminate invasive from non-invasive pines and its failure to predict outcomes in our study could be a consequence of the focus on naturalisation. Nevertheless, all pine species naturalised in NZ were invasive and we would therefore expect it to discriminate among these. In contrast, we found that the Z-score had low explanatory power in NZ and was ranked below human and biogeographic factors (Table 3.2). We also found that there was only a weak correlation between the Z-score and area planted (≤ 0.26 in both regions; see Appendix S2). Our findings support recent work by Dawson *et al.* (2011a) who also highlighted that human factors (e.g. propagule pressure) were more important determinants of invasion success than species traits in a wide range of tropical woody taxa.

We found that a close climate match between the native range and the target region was a good predictor of success at both the introduction and naturalisation stages. This result supports several studies that have found one or more measures of climate match to be important determinants of pine naturalisation (Essl *et al.*, 2011, 2010; Nuñez & Medley, 2011), and naturalisation and invasion of other plant taxa (Diez *et al.*, 2009; Gravuer *et al.*, 2008; Hayes & Barry, 2008). However, for twelve pine species used in plantation forestry, Nuñez & Medley (2011) showed that climate match could identify only the invasive species (*sensu* Richardson *et al.*, 2000b) and not those which failed to invade. Given that, for pines, selection for good climate match is linked to forestry use (Essl *et al.*, 2011, 2010; McGregor *et al.*, 2012), it may be that the species examined by Nuñez & Medley (2011) were already preselected to have a good climate match with the recipient region, weakening the discriminatory power of this variable.

Our study is the first to quantitatively compare the outcome of pine introductions to different regions and we show that, perhaps not surprisingly, the forestry use index was a strong predictor of introduction success, supporting previous work that found economic use was a strong predictor of *Trifolium* introductions to New Zealand (Gravuer et al., 2008). Human factors (forestry use index and area planted for forestry) were likewise strong determinants of pine naturalisation (and in NZ, invasion success; Essl *et al.*, 2011, 2010; Procheş *et al.*, 2012). The forestry use index, while not strongly correlated with species traits, was more strongly associated ($r > 0.50$) with native range size and degree of climate match in each region. It was also strongly correlated with the area planted in NZ, which was another important predictor of invasion success in this region. However, in GB, area planted was not a good predictor of naturalisation and was less strongly correlated with the forestry use index. These contrasting patterns in the role of the forestry use index and area planted between regions could be due to differences in the temporal coverage of the data available from forest working plans in each region. In NZ, which was only settled by Europeans in the mid-late 1800s, the working plans covered the entire period of commercial forestry planting. In GB, however, early private plantings may not have been documented in the Forestry Commission archives thus underestimating their extent, particularly for European species. Nuñez *et al.* (2011) found conflicting evidence for the role of propagule pressure in pine invasion success, concluding that its effects could be modified by ecological filters in the area surrounding plantations. This could apply to GB, where the higher human population density relative to NZ, and more intensive land-use could mediate the effect of propagule pressure. In NZ large areas of extensively grazed tussock grasslands are particularly prone to pine invasion (Harding, 2001), and widespread similar habitats do not occur in GB.

We found that a combination of species attributes, biogeographic and human factors were able to accurately distinguish success and failure of pines at the two stages we examined (AUC > 0.8). This discriminatory ability was similar to that found when the same pool of species was assessed using the Australian weed risk assessment (WRA) to distinguish success and failure at these stages in the same regions (McGregor *et al.*, 2012). The WRA was able to classify species with a similar level of accuracy and reliability to our BRT models. This suggests that a full weed risk assessment is no more accurate than using a small but targeted subset of variables, in this case: global forestry use, climate match, native range size and minimum juvenile period (Hulme, 2012). However, for species where good information is available, the WRA is faster than the BRT approach, though this time advantage is reduced for poorly known species that require more research. In addition, when information about invasive behaviour elsewhere is missing, the WRA becomes unreliable (McGregor *et al.*, 2012), and the BRT approach has the advantage that it does not rely on this information for accurate risk assessment.

3.6 Conclusion

We have shown that for two climatically similar island regions in different hemispheres with a similar history of forestry introductions, outcomes at the introduction and naturalisation stages are explained by a similar set of variables. This implies that similar processes may underlie pine invasions in different regions, and that differences in the level of invasion between regions or hemispheres may reflect differences in the total number of species introduced or area planted (the total “invasion pressure”), rather than fundamental differences in the processes underlying invasion or the susceptibility of regions to invasion. Our results suggest that forestry usage and biogeography are better predictors of naturalisation than species attributes. Risk assessment, at least for pines, needs to especially incorporate knowledge of how widespread a species is likely to be planted and biogeographic factors because these predict invasion outcomes better than species attributes. Given the strong economic pressures to introduce alien trees in plantation forestry, uncoupling utility from invasion could require a novel approaches that might include plant breeding (Anderson *et al.*, 2006) to reduce seed dehiscence or viability, and/or landscape management (Buckley *et al.*, 2005; Hulme, 2006) to limit spread beyond the plantation neighbourhood.

3.7 Author contributions and publication details

Author contributions were as follows: KFM, RPD and PEH designed the research. KFM undertook the data collection and statistical analyses with normal supervisory guidance from RPD, PEH and MSW. KFM wrote the manuscript, with comments and feedback from RPD, PEH and MSW. John Wilson, Martin Nuñez, and one anonymous reviewer provided comments that improved the manuscript. This manuscript was published in *Diversity and Distributions* **18**, 1013–1023.

Chapter 4

Is phylogeny a useful predictor of *Pinus* invasion risk?

4.1 Abstract

Predicting which species are likely to be invasive is essential for effective risk management. The potential to predict risk based on species relatedness has been widely recognised and investigated for extinction risk with the aim of identifying future at-risk species in order to prioritise conservation efforts, yet has been relatively poorly investigated for invasion risk. Models of how factors determine invasion success have recently begun to control for phylogenetic autocorrelation and have shown that controlling for phylogeny can change analytical results, but this practice is not widespread in invasion biology. I address these issues and test whether there are phylogenetic signals in introduction, naturalisation and invasion risk, and in factors that could determine risk in the genus *Pinus*. I infer phylogenetic relationships using DNA sequence data (*matK* and *rbcL*) analysed in the Bayesian Markov chain Monte Carlo (MCMC) program BEAST. I then incorporate phylogenetic uncertainty into phylogenetic signal calculations (Blomberg's K and the D test) on the factors that determine introduction and naturalisation success in New Zealand and Great Britain, and on introduction and naturalisation success themselves. In order to assess the effect of accounting for phylogeny, I use phylogenetic and non-phylogenetic MCMC generalised linear mixed models (glmm) to compare the importance of factors that determine success for introduction and naturalisation in New Zealand (NZ) and Great Britain (GB). I found that there was no significant phylogenetic signal in success at any stage or for the factors that determine success, and that controlling for phylogeny does not change the significance of factors influencing success. Introductions were predicted by a high forestry use index (both regions) and close climate match (NZ only). Naturalisation success was increased by a larger area planted (both regions), high forestry use index and close climate match (NZ only). The Z-score was not a significant predictor at either stage or in either region, but played a larger role for naturalisations in NZ than GB. These results suggest that for pines phylogenetic relationships are not a useful predictor of invasion risk and that models produce the same results with and without controlling for phylogeny. Investigating whether phylogenies can predict invasion risk in a wider group of species, for example all Pinaceae, would be a useful broader test of the ideas developed in this chapter.

Keywords: climate match; forestry; native range size; phylogenetic signal; phylogenetic uncertainty; traits; weed; Z-score

4.2 Introduction

Predicting which species are likely to become invasive and which are not is a central aim in invasion biology. It is becoming clear that a suite of species attributes, biogeographic and human factors each contribute to the probability that a species will transition through each stage of the invasion process. For plants, the probability of naturalisation is increased by having life-history traits related to fast and prolific reproduction (Procheş *et al.*, 2012; Pyšek *et al.*, 2009b; Rejmánek & Richardson, 1996), a closer climate match to the introduced regions (Essl *et al.*, 2011; Nuñez & Medley, 2011), a larger native range size (Procheş *et al.*, 2012), greater economic importance (Essl *et al.*, 2011, 2010; Hulme, 2009; Křivánek & Pyšek, 2008; Pyšek *et al.*, 2010) and higher propagule pressure (Essl *et al.*, 2011, 2010; Křivánek & Pyšek, 2006; Pyšek *et al.*, 2009b). However another crucial factor is the role of evolutionary relationships among species, and how these relationships may allow us to predict invaders or confound traditional statistical analyses into the factors determining invasions.

Phylogenetic relationships alone could predict naturalisation and invasion risk because related species may share similar traits through shared ancestry, including those traits linked to risk (Harvey & Pagel, 1991). The potential to predict risk based on species relatedness has been widely recognised and investigated for extinction risk with the aim of identifying future at-risk species in order to prioritise conservation efforts (Purvis, 2008). Quantitatively investigating the strength of phylogenetic clustering (or “phylogenetic signal” defined as the statistical non-independence among species trait values due to their phylogenetic relatedness) in extinction risk is an on-going area of research and generally assigns extinction risk based on IUCN assessment results. Phylogenetic signal in risk is evident for birds and mammals (Purvis *et al.*, 2000; Russell *et al.*, 1998), amphibians (Corey & Waite, 2008; Stuart *et al.*, 2004) and plants (Davies *et al.*, 2011; Pilgrim *et al.*, 2004; Schwartz & Simberloff, 2001; Vamosi & Wilson, 2008), such that some taxonomic groups have more threatened species than is expected by chance. Using the same logic, phylogenetic non-randomness of invasive species could be used to predict invasion risk.

Several studies have investigated phylogenetic non-randomness in invasion risk, although this non-randomness has not been quantified using measures of phylogenetic signal. Species have been assigned to high-level taxonomic groups (such as families, orders, and subclasses) to assess whether certain groups contain more invasive species that expected by chance (Alcaraz *et al.*, 2005; Blackburn & Duncan, 2001; Daehler, 1998; Lloret *et al.*, 2005; Pyšek, 1998; Richardson & Rejmánek, 2004; Tingley *et al.*, 2010; Vázquez & Simberloff, 2001; Vilà & Muñoz, 1999). For plants, invasive species are often over-represented in some taxonomic groups, although there appear to be only three families that are identified by more than one study, including Amaranthaceae, Papareraceae and Polygonaceae (Daehler, 1998; Pyšek, 1998; Vilà & Muñoz, 1999). The usefulness of this approach for informing risk is limited by three factors. First, all grouping levels within the Linnaean hierarchy (e.g. orders, family, genus) are human-made categories that are not equivalent units for comparison (Fisher, 1991; Forey *et al.*, 2004). Second, the higher taxonomic levels used are often large and mean that any groups identified as risky are not specific enough to be used for practical mitigation actions (e.g. Pyšek *et al.*, 2009a; Schmidt & Drake, 2011b). Finally, identifying invasive species from within a taxonomic group without accounting for the opportunity that each species had to become invasive (i.e. whether a species was introduced outside its native range or not) produces a biased sample of invasive species for analysis (Cassey *et al.*, 2004). Consequently, new approaches such as quantifying phylogenetic signal that deal with these methodological issues are needed to assess how phylogeny can inform risk assessment.

In addition to testing whether a species phylogeny alone is a useful proxy for risk, knowledge of how phylogeny affects trait distributions is useful when implementing statistical analyses that make comparisons across taxa. Conventional statistical tests assume that species are independent units for analysis. However, species are part of a hierarchically structured phylogeny and may not be biologically or statistically independent from one another (Harvey & Pagel, 1991; Martins & Hasnes, 1996; Miles & Dunham, 1993). The assumption of phylogenetic independence can lead to spuriously narrow confidence limits and inaccurate correlation and regression parameter estimates (Felsenstein, 1985; Harvey & Pagel, 1991; Martins & Garland Jr, 1991). Consequently, several statistical methods have been developed that incorporate phylogenies and account for non-independence among taxa. These include techniques such as independent contrasts (Felsenstein, 1985), phylogenetic autocorrelation (Cheverud *et al.*, 1985), phylogenetic generalised least squares regression (Grafen, 1989; Martins & Hansen, 1997) and phylogenetic mixed models (Housworth *et al.*, 2004).

Several studies have incorporated phylogeny into analyses of the factors related to invasion risk in order to control for relatedness between species. Studies of invasions have used variance partitioning (Blackburn & Duncan, 2001; Dawson *et al.*, 2009a; Küster *et al.*, 2008; Pyšek *et al.*, 2009a), phylogenetic independent contrasts (Alcaraz *et al.*, 2005; Jeschke & Strayer, 2006) and phylogenetic mixed models (Dawson *et al.*, 2011b). Controlling for phylogeny sometimes changes the importance or significance levels of some variables when compared to a non-phylogenetically controlled analysis (Alcaraz *et al.*, 2005; Dawson *et al.*, 2009a, 2011b; Jeschke & Strayer, 2006). For example, for mammals the effect of hunting (a variable related to human use and reason for introduction) disappeared when phylogeny was included (Jeschke & Strayer, 2006), and several trait differences between native and invasive fish species were significant when phylogeny was included but not significant when phylogeny was excluded (Alcaraz *et al.*, 2005). Phylogeny may also be more important at lower taxonomic levels and the later stages of invasion (Pyšek *et al.*, 2009a). However in general, phylogeny has a minor explanatory role in most studies (Dawson *et al.*, 2009a, 2011b; Küster *et al.*, 2008; Pyšek *et al.*, 2009a). Therefore, although phylogeny itself may have a minor explanatory role, failing to control for its effects can lead to spurious conclusions.

In this study I aim to assess whether phylogenetic relationships can be useful for predicting invasion risk at the species level. Specifically, I assess the strength of phylogenetic signal in introduction and naturalisation success for *Pinus* in New Zealand (NZ) and Great Britain (GB) using quantitative measures which incorporate phylogenetic information from DNA sequence data. Because of the detailed knowledge of pine introductions to NZ and GB (Appendix A), I account for the opportunity of each species to naturalise (or invade) in these regions. Additionally, because *Pinus* is a well-studied genus (e.g. Richardson, 2006), DNA sequence data is available for most species from which phylogenetic relationships can be inferred (Gernandt *et al.*, 2005). Finally, patterns of extinction risk in plants have been found to differ between regions (Davies *et al.*, 2011) and by assessing phylogenetic signal for the same genus in two regions, I can identify whether the pattern in invasion risk differs between regions for the same group of species. This study also aims to assess whether phylogenetically controlled analyses of the factors determining invasions change the conclusions from previous chapters, and whether including phylogeny in analyses changes the significance levels of variables compared to non-phylogenetic models.

4.3 Methods

4.3.1 Species and variables

A list of all species in the genus was compiled and each species was categorised as either not introduced, introduced or naturalised in NZ and GB (Chapter 3 and Appendix A). All naturalised species in NZ are classed as invasive (Richardson & Rejmánek, 2004), so for NZ I effectively test for phylogenetic signal in invasion risk as well. However, not all species are invasive in GB (Richardson & Rejmánek, 2004), so for GB I am testing the phylogenetic signal in naturalisation risk. I refer to this stage as "naturalisation" in both regions to keep terminology simple.

Because a small suite of variables determine introduction and naturalisation success in pines and represent the three main classes of variables—species, biogeographic and human (Chapter 3)—these were the variables that I selected to test for phylogenetic signal. All variables are listed in Table 4.1 and further details of how these variables were collected or calculated are given in Chapter 2 and Chapter 3.

Table 4.1. Table of *Pinus* variables used in phylogenetic MCMC generalised linear mixed models of introduction and naturalisation success for New Zealand and Great Britain; the class of variable; variable type and units or coding used. Variables with a $*$ were only included at the naturalisation stage; all other variables were included at both stages (introduction and naturalisation).

Variable class	Variable	Type	Unit or code
Species	Z-score	cont.	-
Biogeographic	Climate match	cont.	%
	Native range size	cont.	km^2
Human	Forestry use index	cont.	-
	Residence time$*$	cont.	year
	Area planted$*$	cont.	ha

4.3.2 Phylogeny

I used data from Gernandt *et al.* (2005, TREEBASE study accession number S1143, matrix accession number M1964) to construct a phylogenetic hypothesis for the genus *Pinus*. The phylogeny was based on chloroplast DNA regions *matK* and *rbcL*. Since the study by Gernandt *et al.* (2005) further sequence data have become available on GENBANK (Table 4.2). These sequences were incorporated into the matrix, resulting in 109 species (out of 115 that I recognised; see Appendix A) being represented in the phylogeny. DNA sequence data were concatenated and aligned visually in MEGA5 (Tamura *et al.*, 2011) using the translated protein sequence.

Table 4.2. Table showing the GENBANK accession numbers for additional *Pinus matK* and *rbcL* sequence data that were not included in the Gernandt *et al.* (2005), but were used in this study.

Taxon	*matK*	*rbcL*
Pinus arizonica	DQ156484.1	FJ580056.1
Pinus henryi	EU369312.1	EU369313.1
Pinus juarezensis	AY115770.1	AY115752.1
Pinus kwangtungensis	EF546713.1	AB019802.1
Pinus lagunae	AY115783.1	AY115752.1
Pinus orizabensis	AY115753.1	AY115785.1
Pinus strobiformis	AB455588.1	AB455829.1
Pinus washoensis	DQ156490.1	FJ580082.1

Bayesian tree building programs search and optimise over many thousands of possible tree topologies (e.g. Drummond & Rambaut, 2007; Huelsenbeck & Ronquist, 2001). Rather than summarising this variation in tree topology into a single consensus tree, performing multiple analyses using many trees allows us to incorporate the uncertainty associated with the molecular data and tree-building process (Arnold *et al.*, 2010; Huelsenbeck *et al.*, 2000; Pagel & Lutzoni, 2002). Incorporating this uncertainty is important because analytical results have been shown to depend on the tree topology (Lutzoni *et al.*, 2001). Therefore, in order to quantify the uncertainty surrounding the phylogenetic relationships for *Pinus* and because subsequent statistical analyses (see below) required an ultrametric tree (rooted tree where branch lengths correspond to time), I used the Bayesian analysis program BEAST v1.6.1 (Drummond & Rambaut, 2007) because it generates a sample of many thousands of trees from the posterior tree distribution (the “posterior tree block”) and produces ultrametric trees. BEAST is a cross-platform program for Bayesian Markov chain Monte Carlo (MCMC) analysis of molecular sequences that can produce rooted, time-measured phylogenies inferred using strict or relaxed molecular clock models (Drummond *et al.*, 2006). MCMC procedures ensure that trees are sampled in proportion to their probability of occurrence under the model of gene-sequence evolution (see below).

In order to establish the best model of gene sequence evolution for my data, I used the program JMODELTEST (Guindon & Gascuel, 2003; Posada, 2008; Posada & Buckley, 2004) that identified (from those available for implementation in BEAST) the general time reversible substitution model plus invariant sites and gamma rates (GTR+I+G) as the best model, based on Akaike information criterion (AIC; Akaike, 1974) values. Invariant site heterogeneity allows some sites to never undergo evolutionary change, and gamma rate heterogeneity allows the substitution rate to vary so that some sites evolve more slowly and some more quickly. Because previous BEAST runs demonstrated that the data were not clock-like, I used an uncorrelated lognormal relaxed molecular clock where the rate for each branch is drawn from a lognormal distribution and assumes independent rates on different branches (Drummond *et al.*, 2006). Other settings were as follows: base frequencies estimated; no partitioning; the tree prior was set to the Yule speciation process prior (a pure birth process); and all other priors were left as their default settings (see the BEAST documentation).

Bayesian MCMC was run twice from random starting trees over 50 million generations, sampling every 5,000 generations. In order to avoid any trees that might have been sampled before convergence of the Markov chains, the last 5,000 trees from these two analyses (effectively a 50% burn-in) were combined with LOGCOMBINER v1.6.1. and inspected with TRACER v1.5 (all programs included in the BEAST download, see Rambaut & Drummond, 2007), resulting in 10,000 posterior trees (the posterior tree block). In order to obtain the single tree shown in Figure 4.1, I created a median clade credibility tree on the basis of the 10,000 saved trees using TREEANNOTATOR v1.6.1 (Figure 4.1).

4.3.3 Statistical analysis

4.3.3.1 Phylogenetic signal

In order to quantify the effect of variation in the posterior tree block (the effect of phylogenetic uncertainty) on analytical results, I performed multiple statistical tests across a random selection of 1,000 trees from the posterior tree block. This number of trees was selected because preliminary analyses showed that averaging over 1,000 random trees produced similar results to averaging over all 10,000 trees but resulted in a significantly shorter time for analyses to run.

Analyses were carried out in R version 2.13.1 (R Development Core Team, 2011). First, the data and tree block were matched, reordered and pruned to insure that only species present in both the tree and data were included. In order to quantify phylogenetic signal for all continuous variables I calculated Blomberg's K (Blomberg *et al.*, 2003) using the phylosig function in the phytools package in R (Revell, 2011). Blomberg's K has some noted advantages and disadvantages (Münkemüller *et al.*, 2012). I used K because in preliminary tests it produced the same significant and non-significant results as the other comparable alternative measure of phylogenetic signal, Pagel's λ (Pagel, 1999), but was easier to implement and interpret.

The K statistic can take values from $0 \rightarrow \infty$ and can be thought of as the degree of similarity among species in the phylogeny compared with the degree of similarity expected from a Brownian model (BM) of evolution (Felsenstein, 1985). Traits in the BM follow a random walk along the branches of a phylogenetic tree, with the variance in the distribution of trait values being directly proportional to branch length. To test for the null hypothesis of no phylogenetic signal, the observed value of the K statistic is compared to the values expected under random trait distribution (Münkemüller *et al.*, 2012). Low values indicate low phylogenetic dependence and vice versa. When $K = 1$ the traits follow a BM. When $K < 1$ there is low phylogenetic signal and closely related species are more distant than expected by a BM. When $K > 1$ there is a higher degree of phylogenetic signal than expected under a BM and closely related species are more similar to one another than a BM. The phylosig function assesses the significance of K by randomly shuffling occupancy values among species 1,000 times and calculates 95% confidence intervals, returning a P-value.

Blomberg's K cannot be applied to binary traits. Therefore to quantify phylogenetic signal for the binary traits (introduced or not; naturalised or not) I used the D test (Fritz & Purvis, 2010) calculated using the phylo.d function in the caper package (Orme *et al.*, 2011). When $D = 1$ the trait is distributed randomly on the phylogeny. When $D = 0$ the trait is clumped as expected by a BM. Values of D can fall outside the range of 0–1. Because the phylo.d function scales D to be comparable between phylogenies using two null distributions, it can also calculate two P-values (P_1 and P_0). If $P_1 \leq 0.05$ then the trait is significantly different from a random distribution, i.e. phylogenetic signal is present. If $P_0 \leq 0.05$ then the trait is significantly different from a clumped distribution, i.e. phylogenetic signal is not present.

I first calculated K or D for all factors determining introduction success and for introduction success itself, for all 109 taxa in NZ and GB over a random sample of 1,000 trees. Then, because invasions are a stage-based process and some species were not introduced to a region, all 1,000 trees randomly sampled from the posterior tree block were pruned to remove species that were not introduced to a given region. Phylogenetic signal was then calculated on this pruned tree block for the traits determining naturalisation success, and then again for naturalisation success (0 = not naturalised; 1 = nautralised) itself. Therefore phylogenetic signal in traits determining naturalisation was calculated on two different pruned tree blocks: one for those species that were introduced to NZ, and one for those introduced to GB.

4.3.3.2 Congruence of trees

In order to assess whether the method of inference I used (Bayesian) resulted in a similar tree topology to previous trees generated from the same data using different inference software (Gernandt *et al.*, 2005), I tested for the level of congruence between a consensus tree of the BEAST runs and the consensus tree published by Gernandt *et al.* (2005). Gernandt *et al.* (2005) derived their consensus tree using equally weighted parsimony, which may have optimised on a slightly different tree topology. However, if the majority of the major clades are similar between the two different phylogenetic analysis methods it is more likely that the clades present in both trees are an accurate representation of the data. BEAST does not produce a consensus tree, therefore I computed a 50% majority rule consensus tree in R using the ape package (Paradis *et al.*, 2004) so that the BEAST result was comparable with the Gernandt *et al.* (2005) tree. To further make both trees comparable, the BEAST consensus tree was pruned to remove the eight additional species sequence data (Table 4.2) that were not present in the Gernandt *et al.* (2005) data set. I then used the tree.comp function in the spider package (Brown *et al.*, 2012) to compare the proportion of clades that are the same between the two trees.

4.3.3.3 Phylogenetic mixed models

In order to determine whether including phylogeny changed the importance of factors determining success at each stage, I used Bayesian generalised linear mixed models (glmm) implemented in the package MCMCglmm (Hadfield, 2010) using R. MCMCglmm allows phylogenies to be optionally included in models as a component of variance (Hadfield & Nakagawa, 2010) and can handle binary response variables as well as explanatory variables from different sampling distributions. The phylogeny I used in all MCMC glmm models was the median clade credibility tree (Figure 4.1) calculated from the posterior tree block (i.e. phylogenetic mixed model results are not averaged over the posterior tree block like phylogenetic signal was).

To assist with model convergence and ensure that parameter estimates for all continuous and binary explanatory variables were on a comparable scale, all continuous explanatory variables were first standardised by subtracting their mean and dividing by two standard deviations (Gelman, 2008). The phylogenetic tree and data were matched and pruned to ensure that only species present in both data and tree were included. Species with missing values (for climate match and native range size; $n = 3$) were then removed because MCMCglmm cannot incorporate varaibles with missing data if those varaibles are fixed effects. In total 107 species were used in the analyses at the introduction stage. For the naturalisation stage these 107 species were subsetted to analyse only those species that had been introduced to a given region, resulting in 62 species for NZ and 30 species in GB (this low number for GB was due to the high number of missing values for residence time in GB).

To quantify the effect of phylogeny on the explanatory power of each variable, I fitted a model that included factors potentially determining success as fixed effects, and specified phylogeny as a random effect. Effectively, the variance explained by phylogeny was modelled as follows:

$$S_i = \beta_{1_i} + \beta_{2_i} v_i + \beta_{3_i} Phy \tag{4.1}$$

Where S_i is success (1) or failure (0) at a given stage (repeated for introduction or naturalisation); β_{x_i} are parameters to be estimated from the data; v_i is one of the explanatory variables listed in Table 4.1; and Phy is the phylogenetic relationships between species (as a phylogenetic covariance matrix). β_{3_i} measures the variance explained by phylogeny. I then fitted the models a second time without $\beta_{3_i} Phy$. Then, to determine whether phylogeny explained any variation in introduction and naturalisation success (and to assess whether phylogenetic mixed models agreed with measures of phylogenetic signal) I fitted a model with just success, an intercept term and phylogeny as a random effect:

$$S_i = \beta_{1_i} + \beta_{2_i} Phy \tag{4.2}$$

The models were then fitted without phylogeny.

Finally, to determine whether phylogeny made a difference to the overall model results, I fitted a model that included all species, biogeographic and human factors as fixed effects and phylogeny (Phy) as a random effect:

$$I_i = \beta_{1_i} + \beta_{2_i} z_i + \beta_{3_i} n_i + \beta_{4_i} c_i + \beta_{5_i} f_i + \beta_{6_i} Phy \tag{4.3}$$

$$N_i = \beta_{1_i} + \beta_{2_i} z_i + \beta_{3_i} n_i + \beta_{4_i} c_i + \beta_{5_i} f_i + \beta_{6_i} + \beta_{6_i} a_i + \beta_{7_i} r_i + \beta_{8_i} Phy \tag{4.4}$$

Where I_i is introductions success (1) or failure (0) and N_i is naturalisation success (1) or failure (0); β_{x_i} are parameters to be estimated from the data; z_i is the Z-score, n_i is native range size, c_i is climate match, f_i is the forestry use index, a_i is area planted, and r_i is residence time. This model fitting process was repeated for both NZ and GB and then repeated again for both regions without phylogeny.

Because MCMCglmm fits Bayesian models, I had to specify priors. To let the data drive parameter estimates I specified uninformative prior distributions for all model parameters. In MCMCglmm this equated to priors for the response variables I_i and N_i being set to a variance of 1, the degree of belief in prior was set to 1 and residual variance was fixed at 1 because all models were binomial and there is no residual variation term. Parameter expanded priors for the random effect (phylogeny) were used in order to speed chain convergence, with a variance of 1, degree of belief in prior of 1, a mean of 0 and the covariance matrix parameter set to 1000. To avoid problems in model fitting associated with separation but still allow the data to drive parameter estimates, I specified a weakly informative prior that equated to a uniform prior on the probability scale (which provides some constraint of the fixed effects; Gelman *et al.*, 2008) as having a variance of $j_i(1 + \pi^2/3)$, where j_i is the number of fixed effects plus the intercept term (Hadfield, 2012).

Each model was run three times in MCMCglmm in order to obtain three MCMC chains for inspection. I set the number of iterations for each chain to 50,000 with a burn-in of 10,000 and sampled every 10 iterations. Convergence of the three chains was judged by both visual inspection of the chain histories and by using the Gelman–Rubin convergence diagnostic measure (Brooks & Gelman, 1998; Gelman & Rubin, 1992) in the coda package (Plummer *et al.*, 2006). The Gelman–Rubin diagnostic measures whether there is a significant difference between the variance within and between several chains by giving the scale reduction factors for each parameter. A value of 1 indicates that between and within chain variances are equal; values substantially above 1 indicate lack of convergence. As a rule of thumb anything above 1.05 indicates chains have not converged. To increase the efficiency of mixing between chains I used a slice sampling method (Damlen *et al.*, 1999) by passing slice=TRUE to MCMCglmm.

MCMCglmm calculates P-values indicating whether a parameter was significant in a given model. To visualise the significance of each variable I plotted the median, upper and lower 95% credibility intervals. Parameters with credibility intervals including zero are not significant (at the level $P = 0.05$). To determine whether including phylogeny improved model fit I compared the deviance information criterion (DIC; Spiegelhalter *et al.*, 1993) for all models. The DIC aims to assess the trade-off between the fit of the data to the model and the complexity of the model. Models with a smaller DIC value are better supported by the data. Model complexity is measured by estimating the effective number of parameters. The best model provides a good fit but is not overly complicated and loses the least amount of information when the model is used to approximate reality. In models with negligible prior information the DIC will be approximately the same as AIC. The philosophy behind DIC and AIC is that there is no “true” model. As a general rule of thumb, models within 10 of the model with the smallest DIC value might be regarded as the best model while those within 2–4 might be regarded as the more likely candidates (Burnham & Anderson, 2002; Spiegelhalter *et al.*, 2002). Very small differences in DIC values can arise because the MCMC chains introduce stochasticity, therefore small differences might not necessarily indicate a better model.

4.4 Results

4.4.1 Phylogenetic tree

The phylogenetic tree comprised 109 taxa that were divided into the two main subgenera and four sections commonly recognised for pines. All major clades (subgenera, sections and subsections) had a $\geq$ 95% posterior probability (Figure 4.1). When displayed on the tree, species introduced and naturalised in NZ and GB showed no strong tendency to be clumped so that there were no monophyletic groups of naturalised species. However, subsections Contortae, Pinus and Pinaster each had two or more naturalised species (Figure 4.2). The congruence between the BEAST majority rule consensus tree (pruned) and the original tree from Gernandt *et al.* (2005) was 94%.

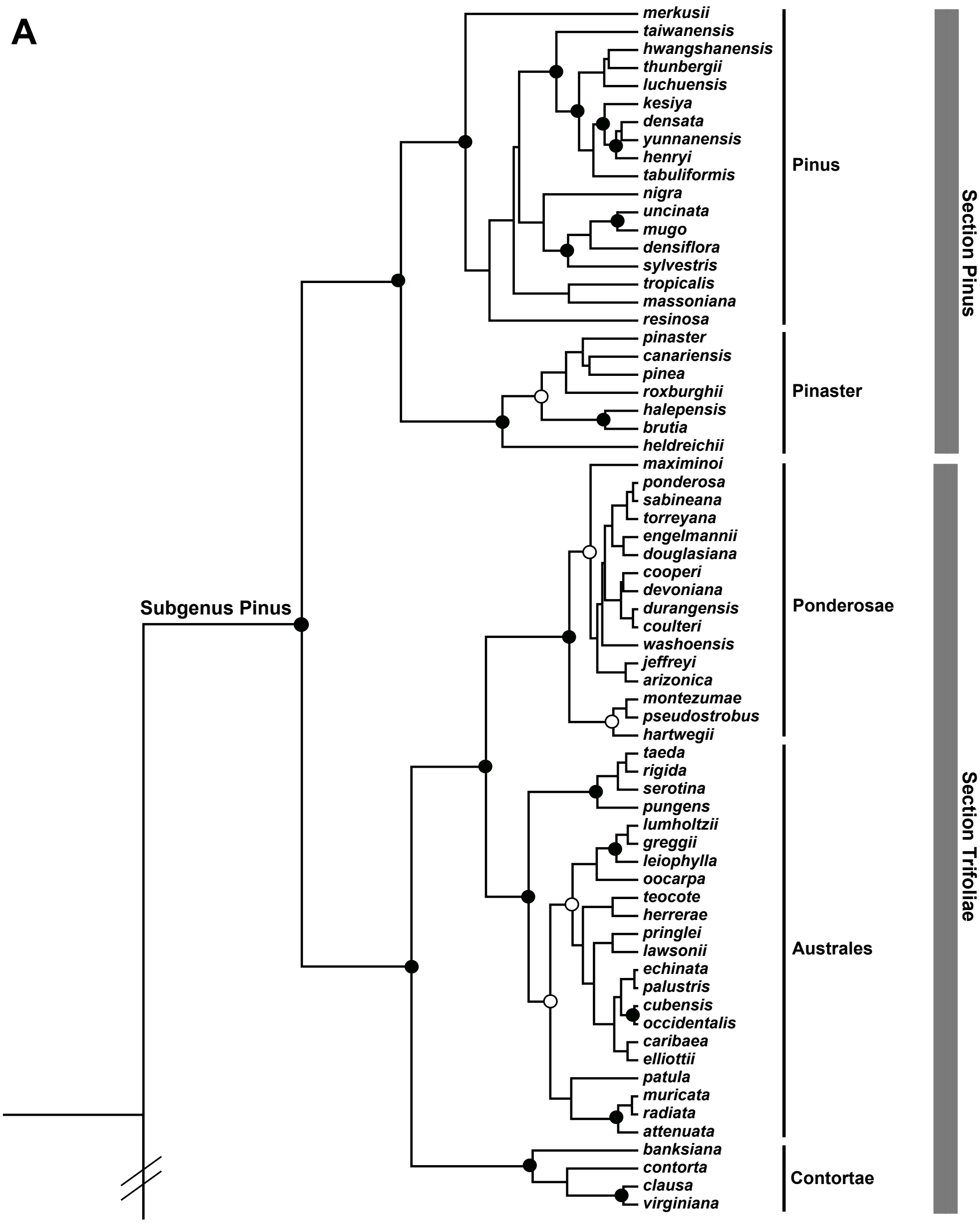

Figure 4.1. Median clade credibility tree produced from Beast for 109 species in the genus *Pinus*. Clades with a posterior probability (support) ≥ 95% are shown with black, those with ≥ 75% posterior probability are shown with white circles, clades with < 75% are not marked with circles. Within each subgenera, clades are labelled by sections (thick bars) and subsections (thin bars). Tree is divided into (A) subgenus Pinus, and (B) subgenus Strobus (on next page...)

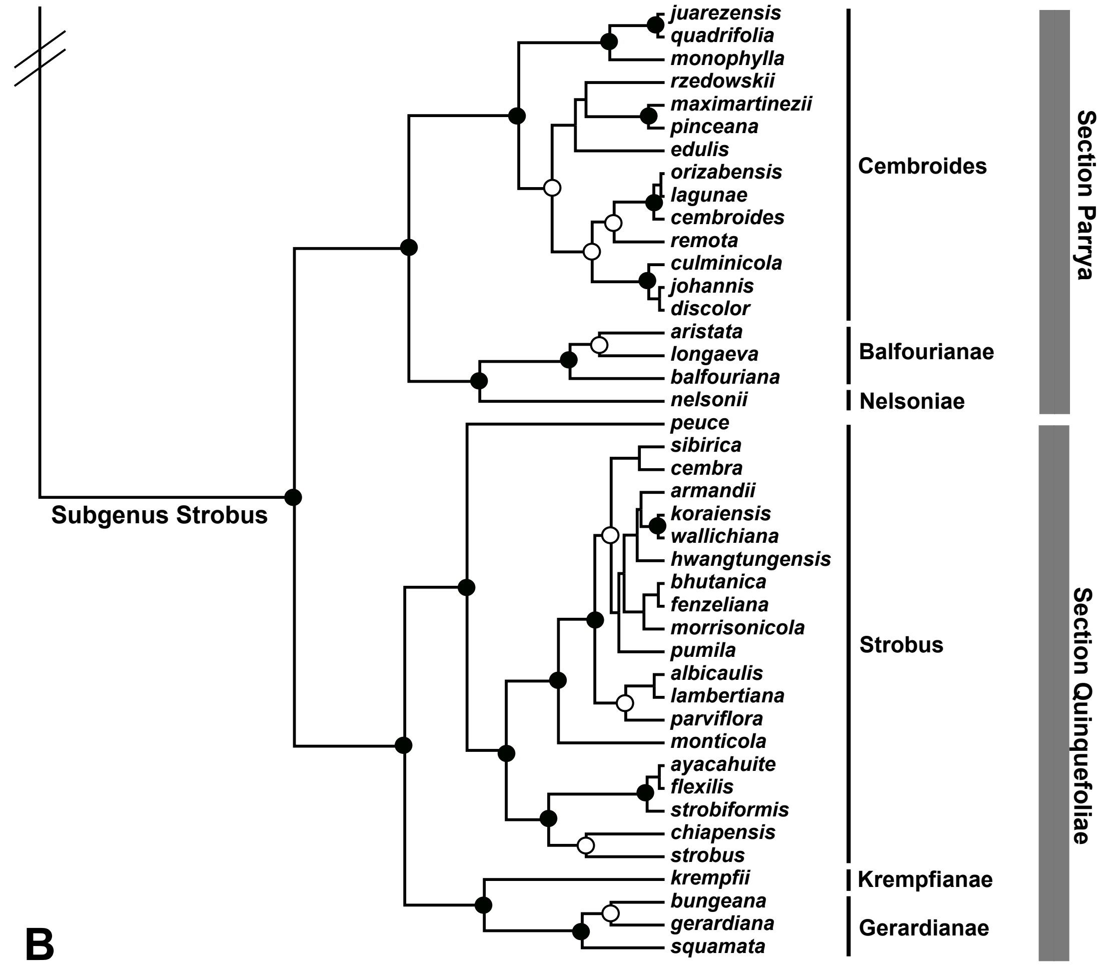

Figure 4.1. . . . continued from previous page

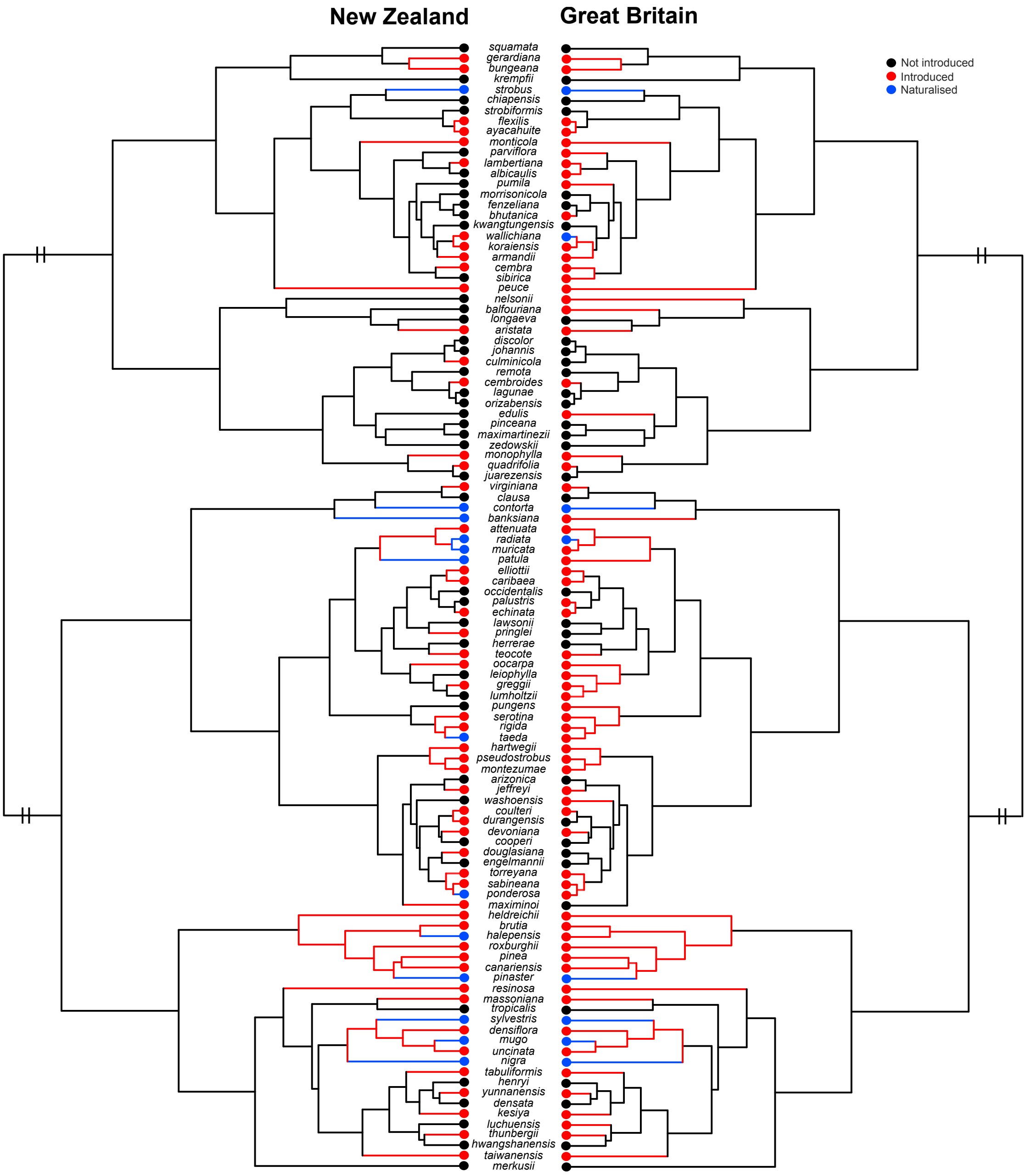

Figure 4.2. Phylogenetic mean clade credibility tree produced from Beast for 109 taxa in the genus *Pinus*, showing the status of each species (not introduced; introduced; naturalised) in New Zealand (left) and Great Britain (right).

4.4.2 Phylogenetic signal

There was no significant phylogenetic signal in either of the binary variables (introduction or naturalisation success (Table 4.3) in NZ or GB. *P*-values for *D* suggested that the distribution of introduction and naturalisation success was not significantly different from a random distribution of introductions and naturalisations on the phylogeny. There was also no significant phylogenetic signal in any of the traits determining introduction or naturalisation success in either NZ or GB (Table 4.4).

Table 4.3. Mean values of *D* test for phylogenetic signal, and mean values for P_1 and P_0. Mean values were calculated by averaging over all results using a sample of 1,000 trees from the Bayesian posterior distribution obtained from BEAST, for *Pinus* introductions and naturalisations in NZ and GB. If $P_1 \leq 0.05$ then the trait is significantly different from a random distribution, thus has phylogenetic signal. If $P_0 \leq 0.05$ then the trait is significantly different from a clumped distribution, thus has no phylogenetic signal. Significant values are highlighted in **bold**.

Variable	*D* (±SE)	P_1 (±SE)	P_0 (±SE)
Introduction NZ	0.860 (0.002)	0.109 (0.002)	**0.000 (0.000)**
Introduction GB	0.876 (0.002)	0.153 (0.004)	**0.000 (0.000)**
Naturalisation NZ	0.849 (0.002)	0.208 (0.003)	**0.004 (0.000)**
Naturalisation GB	1.046 (0.003)	0.547 (0.005)	**0.003 (0.000)**

4.4.3 Phylogenetic mixed models

At the introduction stage in both regions, including or excluding phylogeny did not change the explanatory power of any individual variables (Figure 4.3a), nor the significance level of variables in the full model (Figure 4.3b). Similar results were also seen at the introduction stage for NZ (Figure 4.3c and d). The difference in DIC values between the full model with phylogeny and the full model without phylogeny at the introduction in both regions was < 2, therefore both models had equal explanatory power.

The individual explanatory power of the Z-score, forestry use index and climate match on naturalisation success in GB were significant when phylogeny was not included as a random effect, but became non-significant when phylogeny was included (Figure 4.4a). However including phylogeny in the full model of naturalisation in GB did not change the significance of any variables (Figure 4.4b). Including or excluding phylogeny had no effect on either individual variable importance or their significance in full model of naturalisation success in NZ (Figure 4.4c and d). As at

Table 4.4. Mean values for Blomberg's *K* measure of phylogenetic signal and mean *P*-values. Mean values were calculated by averaging over all results using a sample of 1,000 trees from the Bayesian posterior distribution obtained from Beast, for *Pinus* introductions and naturalisations in NZ and GB. High values (>1) indicate strong phylogenetic signal, whereas small values (<1) indicate a lack of phylogenetic signal. *P*-values ≤ 0.05 indicate significant phylogenetic signal.

Stage	Variables	*K* (±SE)	*P* (±SE)
Introduction	Z-score	0.054 (0.00)	0.308 (0.009)
	Climate match to GB	0.047 (0.001)	0.454 (0.009)
	Climate match to NZ	0.065 (0.00)	0.349 (0.001)
	Native range size	0.045 (0.001)	0.425 (0.001)
	Forestry use index	0.095 (0.001)	0.302 (0.008)
	Area planted GB	0.183 (0.002)	0.203 (0.004)
	Area planted NZ	0.074 (0.001)	0.181 (0.008)
	Residence time GB	0.236 (0.002)	0.098 (0.004)
	Residence time NZ	0.048 (0.001)	0.662 (0.009)
Naturalisation	Z-score GB	0.085 (0.002)	0.303 (0.009)
	Z-score NZ	0.102 (0.021)	0.183 (0.006)
	Climate match to GB	0.05 (0.001)	0.640 (0.007)
	Climate match to NZ	0.065 (0.001)	0.545 (0.008)
	Native range size GB	0.092 (0.001)	0.236 (0.008)
	Native range size NZ	0.100 (0.001)	0.127 (0.008)
	Forestry use index GB	0.097 (0.001)	0.443 (0.009)
	Forestry use index NZ	0.095 (0.001)	0.467 (0.007)
	Area planted GB	0.182 (0.002)	0.300 (0.005)
	Area planted NZ	0.088 (0.001)	0.251 (0.008)
	Residence time GB	0.307 (0.002)	0.080 (0.002)
	Residence time NZ	0.048 (0.001)	0.663 (0.008)

the introduction stage, DIC values suggested that phylogenetic and non-phylogenetic models had the same explanatory power.

Overall, full models including species, biogeographic and human factors suggested that introduction success was significantly increased in both regions by a higher forestry use index, and in NZ a closer climate match with the native range (Figure 4.3b and d). Probability of naturalisation in GB was only significantly increased by a larger area planted (Figure 4.4b). Naturalisation success in NZ was significantly increased by a closer climate match, and greater area planted (Figure 4.4d).

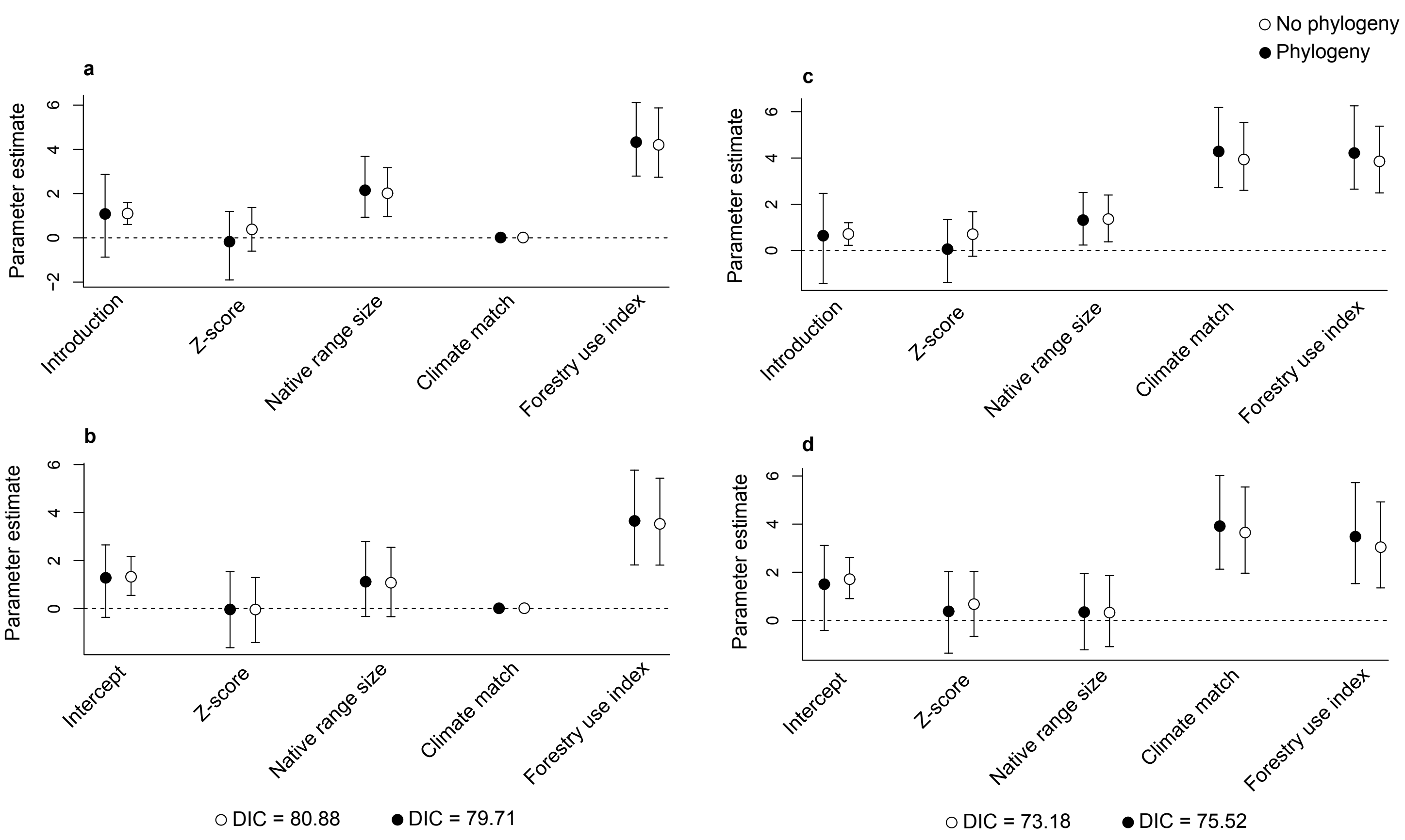

Figure 4.3. Median values ±97.5% credibility intervals for all parameters from Markov chain Monte Carlo generalised linear mixed models with (filled circles) and without (open circles) phylogenetic control, for *Pinus* introduction success to Great Britain (a) when each variable was assessed individually and (b) all variables were used in models; and for introduction success to New Zealand when (c) when each variable was assessed individually and (d) all variables were used in models. Variables with credibility intervals that do not cross the dashed line at $y = 0$ are considered to be significant at the $P = 0.05$ level. DIC values are given for the full models at the bottom of each plot. Lower DIC values indicate a better fitting model, but DIC values differing by < 4 are not considered to be substantially different.

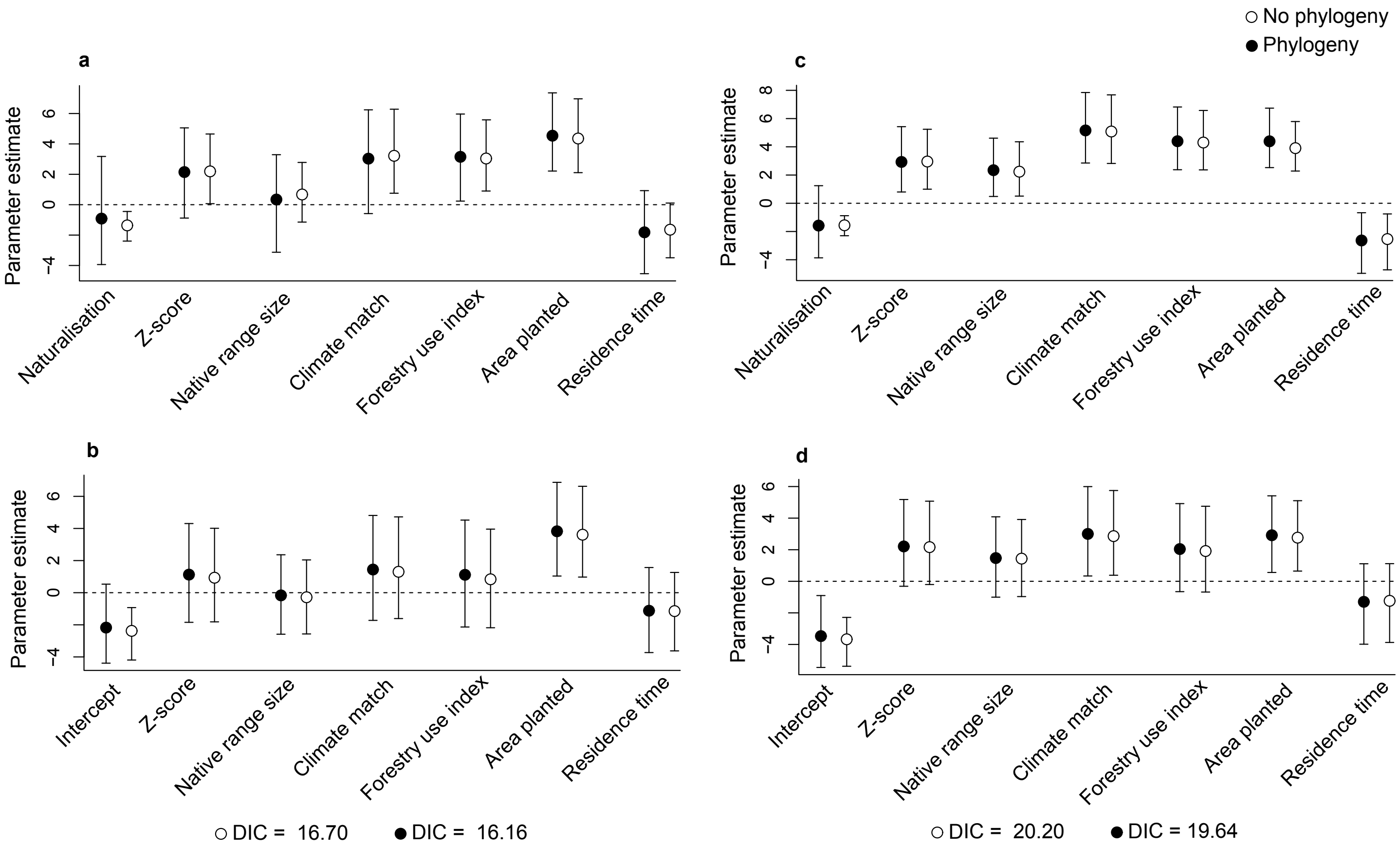

Figure 4.4. Median values ±97.5% credibility intervals for all parameters from Markov chain Monte Carlo generalised linear mixed models with (filled circles) and without (open circles) phylogenetic control, for *Pinus* naturalisation success to Great Britain (a) when each variable was assessed individually and (b) all variables were used in models; and for *Pinus* naturalisation success to New Zealand when (c) when each variable was assessed individually and (d) all variables were used in models. Variables with credibility intervals that do not cross the dashed line at $y = 0$ are considered to be significant at the $P = 0.05$ level. DIC values are given for the full models at the bottom of each plot. Lower DIC values indicate a better fitting model, but DIC values differing by < 4 are not considered to be substantially different.

4.5 Discussion

There was no phylogenetic signal in introduction or naturalisation success, or any of the traits that could determine success for *Pinus*. Phylogenetically controlled models produced the same conclusions as non-phylogenetic models and supported findings from Chapter 3 and Chapter 5. I found that pine introductions are largely determined by a high forestry use index and, in NZ only, better climatic suitability between the native and introduced ranges. Pine naturalisation success was significantly increased in both regions by a large area planted and in NZ by greater climatic suitability (e.g. Essl *et al.*, 2010; Nuñez & Medley, 2011). Although not significant, naturalisation probability in NZ was increased by a longer residence time (e.g. Castro *et al.*, 2005; Křivánek *et al.*, 2006; Pyšek *et al.*, 2009b), greater forestry use index (e.g. Procheş *et al.*, 2012), larger native range size (e.g. Procheş *et al.*, 2012) and higher Z-score (Grotkopp *et al.*, 2002; Rejmánek & Richardson, 1996).

The results of both phylogenetically controlled and non-controlled analyses are consistent with previous invasion studies that found phylogeny had only a minor contribution to results (Dawson *et al.*, 2009a, 2011b; Küster *et al.*, 2008; Pyšek *et al.*, 2009a). I also found that there was no phylogenetic signal in traits potentially linked to invasion risk, including the Z-score, which is a species trait that we may expect to be phylogenetically non-random due to shared evolutionary pressures among closely related species for small seed mass, short juvenile period and large interval between seed crop years. These results suggest that phylogenetic relatedness alone is not a good predictor of invasiveness at lower taxonomic levels (i.e. between species) and that most of the variation in invasion risk is determined by factors that do not necessarily correlate with phylogenetic relatedness, such as propagule pressure. My results suggest that it is unnecessary to control for phylogeny at the species level for *Pinus* when assessing factors determining invasions.

I did not find any evidence that phylogenetic relatedness was more important (or traits more phylogenetically correlated) at the later stage of the invasion process (Pyšek *et al.*, 2009a). However, it is too early to draw any conclusions regarding whether phylogeny is more important at different stages of the invasion process based on these results, which only include one genus in just two regions (NZ and GB). *Pinus* naturalisations in NZ and GB may not be an ideal system in which to test these questions because of the limited number of species being examined and the limited number of regions used. A wider analysis that included several genera from within the Pinaceae and compared these with other taxonomic groups (for example grasses and herbs) and examined introduction, naturalisation and invasion globally, would have more power to detect broad trends. However, such an analysis would require detailed knowledge of success and failure at all stages of the invasion process for hundreds of taxa (Cassey *et al.*, 2004; Diez *et al.*, 2009), which could prove problematic to obtain (particularly data on failed introductions).

My results suggest that there was no significant clustering of naturalised or invasive species within the genus *Pinus* and that identifying species-level invasion risk using phylogeny is not viable. This somewhat contrasts with a study by Miller *et al.* (2011) who found that invasive *Acacia* species cluster together on a phylogeny to some extent. However, my results are not directly comparable with Miller *et al.* (2011) for several reasons. First, their study did not account for the opportunity for species to become invasive by identifying which species were introduced to South Africa, whereas my study did account for failures at the introduction stage (Cassey *et al.*, 2004). In total they classed 16 of the 110 species they examined as "invasive" because they had been recorded as invasive in South Africa, but did not state whether they restricted their 110 species to those that had been introduced to South Africa. Second, whereas I used the entire genus (as far as possible, given available DNA sequence data), Miller *et al.* (2011) used a selection of 110 species in the *Acacia* genus (the genus has over 1000 recognised species) split into three data sets representing all major lineages within the genus and the groups containing the invasive species *A. mernsii* and *A. melanoxylon*. Thus their study may contain unforseen biases related to species selection. Finally, Miller *et al.* (2011) did not aim to quantify phylogenetic signal in these species using measures such as K or D (Blomberg *et al.*, 2003; Fritz & Purvis, 2010), thus their conclusions are necessarily qualitative.

Although there are many advantages of using the genus *Pinus* as a model system (Richardson, 2006), pines may not be the ideal case study when examining phylogenetic patterns in invasion risk because several factors could favour evolutionary divergence even among closely related taxa. *Pinus* is a particularly genetically and ecologically diverse genus (Ledig, 1998). The genus arose around 130 million years ago in the Late Cretaceous period and is a basal member of the Pinaceae (Price *et al.*, 1998). The genetic and ecological strategies of pines favours the creation and recombination of genetic variation, which results in species being able to adapt to a wide variety of ecological conditions. Pines have a high number of chromosomes ($n = 12$) compared to other conifers, a monoecious mating system with sex separation in the tree crown and hybridise readily within many species-complexes–all factors that favour recombination and outcrossing, leading to high genetic diversity

(Ledig, 1998). Since their origins in the late Cretaceous, pines have both expanded and contracted their geographic ranges in response to climatic changes and now occupy a wide variety of diverse habitats including lowland taiga forest, sea level tropical swamp, Mediterranean environments, dry desert and mountain tree line (Farjon, 2005; Richardson & Rundel, 1998). Both the large evolutionary time-scales and this genetic and ecological diversity could have resulted in closely related species being different enough from one another that phylogenetic signal in some traits has been removed. Testing for phylogenetic signal in invasion risk on a wider taxonomic group could avoid the biases associated with using *Pinus* as a model group.

4.6 Conclusion

I found that there was little or no phylogenetic signal in introductions, naturalisations, or the traits determining introductions and naturalisations for the genus *Pinus*, suggesting that phylogeny is not a useful proxy for risk at the species-level in these taxa. When models controlling for phylogeny were compared with non-phylogenetically controlled models, there was no difference in the significance levels of variables between models. Mixed models indicated that factors that were not correlated with phylogeny, and would not necessarily be predicted to correlate with phylogeny including the forestry use, planting effort and climate match, determined introductions and naturalisations in both regions. Factors that were more likely to correlate with phylogeny such as native range size and the Z-score were not significant predictors at either stage of invasion. These results suggest that phylogenetic relationships may have a limited role in predicting invasions at the species level for conifers, because the factors that determine success are largely governed by human selection, which may not correlate with phylogeny. Quantitative tests of how invasion risk is distributed across phylogenies at the species level are needed to assess whether there are any generalities that could be useful for assessing the risk of species becoming invasive in the future.

Chapter 5

Mapping the path to plant naturalisation: identifying causal relationships with exploratory path analysis

5.1 Abstract

1. Understanding the mechanisms driving alien plant invasions is essential for developing effective prevention and ecosystem management strategies. Several species, biogeographic and human factors have been identified that correlate with invasion success, though no research has examined how these factors link together in a causal structure to establish the mechanism of invasion. Path analysis can identify causal links between variables. However, invasions are a complex process, involving multiple variables from different sampling distributions, where there may be no strong prior hypothesis about how factors link to determine success, presenting a challenge to traditional path analysis methods.

2. This chapter introduces a Bayesian method for path analysis that overcomes many of these issues. I use this method to test how species, biogeographic and human factors link in a causal structure to determine outcomes at the introduction and naturalisation stages of the invasion process, for two non-native genera (*Pinus* and *Trifolium*) in New Zealand.

3. I found that human factors (area planted and residence time) acted directly to increase naturalisation success, and that these mechanisms were repeated in both genera, suggesting they are general. However, the factors determining area planted and residence time were different between the two genera. Species attributes had relatively weak direct and indirect links compared to human and biogeographic variables, at both the introduction and naturalisation stages. This analysis revealed the differing causal mechanisms and strength of important biogeographic variables between taxonomic groups.

4. I found that path analysis uncovered more detailed relationships than standard regression methods, and also uncovered instances of potential latent (unmeasured) variables.

5. ***Synthesis*:** My results suggest that variables identified as determining invasion outcomes in traditional regression studies can act via different causal mechanisms between taxonomic groups, and that the results from one group may not apply directly to others. I also found evidence of latent variables, which are likely to be common but unidentified in previous studies, and could be targeted for further research. There is great scope for path analysis to be more widely applied to identifying the drivers of invasions, with the aim of improving prevention and management of invasive species.

Keywords: Bayesian method; biological invasions; BRugs; causation; climate; propagule pressure; path analysis; structural equation; traits; weeds.

5.2 Introduction

Invasions are determined by a range of species attributes, biogeographic, and human factors that can link together in a causal structure. For example, plant invasions have been associated with factors such as climate suitability within a new region (Essl *et al.*, 2011; Gravuer *et al.*, 2008; Nuñez & Medley, 2011; Scott & Panetta, 1993), a large native range size (Goodwin *et al.*, 1999; Gravuer *et al.*, 2008; Procheş *et al.*, 2012), propagule pressure (Essl *et al.*, 2010; Křivánek *et al.*, 2006; Lockwood *et al.*, 2005; Pyšek *et al.*, 2009b), economic use (Essl *et al.*, 2010; Gravuer *et al.*, 2008; Křivánek & Pyšek, 2008; Reichard & Hamilton, 1997), and a range of life-history traits (Gravuer *et al.*, 2008; Pyšek *et al.*, 2009b; Rejmánek & Richardson, 1996). Given that invasions are a process incorporating the stages of introduction, naturalisation, and invasion (Blackburn *et al.*, 2011) there are several possible ways that these variables could link together in a causal structure to determine invasion success. Some factors may be both explanatory and dependent variables in a causal model, such as propagule pressure. While propagule pressure may determine invasion success, this variable may be directly dependent upon climatic suitability to a region and economic use. Such causal linkages have not been explored for large-scale determinants of invasion success, yet could advance our understanding of invasions.

Structural equation models ("SEM"; Bollen, 1989; Haavelmo, 1943; Koller & Friedman, 2009; Neapolitan, 2004; Pearl, 2000; Spiegelhalter *et al.*, 1993) and path analysis (Wright, 1921, a subset of SEM with no unmeasured or "latent" variables), provide a means of determining causal links between variables when experimental manipulation is not possible, as is the case when examining regional- or global-scale determinants of invasion success. Path analysis in particular has become widely used in ecological studies (e.g. Cariveau *et al.*, 2004; Dunham & Mikheyev, 2010; Farris & Lechowicz, 1990; Grace, 2006; Mitchell, 1992; Mysterud *et al.*, 2008; Schemske & Horvitz, 1988; Shipley, 1997, 2000; Sikes *et al.*, 2010; Thomas *et al.*, 2007), yet its application to invasion biology has been limited to assessing the effect of variables on alien species richness (Atwater *et al.*, 2011; Harrison *et al.*, 2006; Hulme, 2011; Keeley *et al.*, 2005; Seabloom *et al.*, 2006; Taylor & Irwin, 2004). Most research has focused on a single study region (Atwater *et al.*, 2011; Harrison *et al.*, 2006; Keeley *et al.*, 2005; Seabloom *et al.*, 2006) or species (Atwater *et al.*, 2011), and only rarely take a continental or global approach (García-Berthou *et al.*, 2005; Hulme, 2011; Taylor & Irwin, 2004). Thus, using path analysis to disentangle the causal relationships between large-scale determinants of success at each stage of the invasion process could yield new insights into the processes determining invasion outcomes.

Despite the potential use of path analysis in invasion studies, two features have limited its wider application. Firstly, standard approaches often assume that variation in the data due to unobserved causes follows a multivariate normal distribution (Bollen, 1989; Shipley, 2000). This makes it difficult to incorporate variables with different sampling distributions (such as binary outcomes or count data) despite these being frequently encountered in invasion biology (e.g. invasive or not invasive). Whilst traditional correction and bootstrapping methods exist that help deal with this problem (Bollen & Stine, 1993; Satorra & Bentler, 1999), they are often difficult to implement and may not be effective for small sample sizes (Hoogland & Boomsma, 1998).

Secondly, identifying suitable path models can be problematic when there are many variables and multiple potential causal models, with no strong prior hypotheses about the causal linkages. Path analysis often aims to identify a "best" model or at least a selection of equally well-performing models using measures of model fit or adequacy such as the AIC (Taylor & Irwin, 2004), BIC (Taylor & Irwin, 2004), and Chi-squared (Harrison *et al.*, 2006; Keeley *et al.*, 2005; Seabloom *et al.*, 2006), or to narrow the set of plausible models by eliminating those which do not meet assumptions given by a probability threshold using a test of independence claims (Shipley, 2009). These approaches work well when there are strong prior hypotheses about the causal relationships between variables or when there are few variables. However, when there is no reason to strongly suspect any particular links over others this method can be difficult or impossible to implement efficiently. For example, given no constraints, the number of potential models with three variables is 64; with four variables this increases to 4,096; and by five variables there are 1,048,576 potential path models to assess (Shipley, 2000).

In this chapter, I describe a new exploratory approach for path analysis to investigate how multiple factors both directly (causally) and indirectly determine invasion outcomes. This approach has two main advantages that overcome the limitations of standard approaches. First, it is straightforward to include variables from different sampling distributions. Second, the approach is not restricted to comparing a subset of all potential path models representing different *a priori* hypotheses, and is thus exploratory.

I apply this approach to disentangle how factors determine success at the introduction and naturalisation stages of the invasion process, for two diverse genera of plant species introduced to New Zealand (NZ): *Pinus* and *Trifolium*. These genera are excellent candidates for partitioning the influence of factors determining success and failure at different stages of invasions for a number of reasons. First, detailed information on success and failure at each stage, together with a large number of human, biogeographic, and life-history traits are available for both genera (Gravuer, 2004; Gravuer *et al.*, 2008; McGregor *et al.*, 2012); however it has proven difficult to isolate the causal relationships underpinning the success of these genera. Second, both genera are naturally absent from NZ and intentionally introduced. For many intentionally introduced plant taxa it is likely to be the case that humans favour species with certain traits (e.g. fast growth rate, broad environmental tolerance) that are in turn linked to invasion outcomes, potentially confounding the traits favouring introduction and traits favouring naturalisation (Chapter 3). Finally, for each stage of the invasion process, there are several alternative and plausible models, and numerous variables potentially determining success, which lends itself to an exploratory approach.

Table 5.1. Characteristics of all variables used in path models for *Pinus* and *Trifolium* introduction and naturalisation in New Zealand. Shown in the table are the response variables and explanatory variables which are categorised into three groups (species, biogeographic and human). For all variables a description of the type of variable (continuous, binary or counts), the coding used or units of the variable; and the source of the data for each variable is given.

Genus	Variable category	Variable	Type	Coding or units	Source
Pinus	Species	Z-score	Cont.	Higher score = more invasive	Appendix A
	Biogeographic	Climate match	Cont.	Number of 10′ lat-long grid cells with ≥ 90% climate match	Chapter 2
		Native range size	Cont.	log km^2	Chapter 2
	Human	Forestry use index	Cont.	Number CABI Forestry Compendium references	Procheş *et al.* (2012)
		Residence time	Cont.	years	Chapter 2
		Area planted	Cont.	hectares (ha)	Chapter 3
	Response	Introduction success	Binary	0 = not introduced; 1 = introduced	Appendix A
		Naturalisation success	Binary	0 = introduced but not naturalised; 1 = naturalised	Appendix A
Trifolium	Species	Height	Cont.	cm; maximum recorded	Gravuer (2004)
		Life-span	Binary	1 = annual or biennial; 2 = perennial	Gravuer (2004)
	Biogeographic	Climate match	Cont.	Number of 10′ lat-long grid cells with ≥ 90% climate match	Gravuer (2004)
		Native range size	Cont.	log km^2	Gravuer (2004)

Continued on next page. . .

Genus	Variable category	Variable	Type	Coding or units	Source
		Conditions tolerated	Count.	–	Gravuer (2004)
		Residence time	Cont.	year	Gravuer (2004)
	Human	Presence in GB in mid-1800s	Binary	0 = no; 1 = yes	Gravuer (2004)
		Area planted	Cont.	hectares (ha)	Gravuer (2004)
	Response	Introduction success	Binary	0 = not introduced; 1 = introduced	Gravuer (2004)
		Naturalisation success	Binary	0 = introduced but not naturalised; 1 = naturalised	Gravuer (2004)

5.3 Methods

5.3.1 Data collection

I used data from two previous studies that have examined factors associated with introduction and naturalisation success in *Pinus* and *Trifolium* species introduced to NZ (Gravuer *et al.*, 2008; McGregor *et al.*, 2012). Details of data collection are given in those studies; here I outline key features of the data that are relevant to the analyses I present below.

5.3.1.1 *Pinus*

Pinus is a genus of *c*. 115 woody species that are naturally absent from NZ, with most species native to the northern hemisphere (Earle, 2008; Farjon, 2005; Price *et al.*, 1998). Pines have been widely introduced around the world for forestry (e.g. Essl *et al.*, 2010; Nuñez & Medley, 2011; Procheş *et al.*, 2012) and are now amongst the worst forestry invaders, with 21 species being considered invasive globally (Richardson & Rejmánek, 2004). Consequently, pines make an excellent model system for studying invasions (Richardson, 2006).

Each pine species in the global pool was classified as having been introduced to NZ or not (0 = no, 1 = yes) by searching historical, horticultural and scientific literature (Appendix A): from the global pool, 66 pine species were intentionally introduced to NZ for commercial forestry and horticulture. These 66 species were then classified as having naturalised in NZ or not (0 = no, 1 = yes), with a species classed as naturalised if it had established new self-sustaining populations outside of cultivated areas (*sensu* Richardson *et al.*, 2000b): 13 pine species have naturalised in NZ. Following the definition of invasion given by Richardson *et al.* (2000b) all naturalised species can also be considered as invasive in NZ.

I used variables that might directly or indirectly determine success at the introduction stage (Table 5.1) including: climate match (a measure of how well matched the climate in NZ is to climate in the species native range on a scale from 0–100; Chapter 2); native range size (in km^2, calculated from range maps; Chapter 2); the forestry use index (the number of citations in the CABI Forestry Compendium [CABI, 2010] for each species following Procheş *et al.*, [2012]); and the Z-score (a composite measure derived from a discriminate analysis of three life-history traits: seed mass, minimum juvenile period and minimum interval between large seed crop years; Rejmánek & Richardson [1996]) in analyses at the introduction stage. At the naturalisation stage I tested the effect of climate match, native range size, area planted (calculated using data from archival working forest plans held at Scion, Rotorua), residence time (number of years since first introduction) and Z-score. I chose these variables for several reasons. First, they represented three different categories of variables (species, biogeographic and human) commonly identified as determining invasion success (Chapter 3). Second, many have previously been identified as important determinants of pine introduction and invasion success (Essl *et al.*, 2011, 2010; Křivánek & Pyšek, 2008; Křivánek *et al.*, 2006; Nuñez & Medley, 2011; Procheş *et al.*, 2012; Pyšek *et al.*, 2009b; Rejmánek & Richardson, 1996). Finally, these variables have the potential to interact via important indirect pathways that have not been previously quantified.

5.3.1.2 *Trifolium*

Trifolium is a genus of *c*. 228 species that are naturally absent from NZ (Gillett & Taylor, 2001; Gravuer *et al.*, 2008). Following Gravuer *et al.* (2008) I classed species as intentionally introduced to NZ or not (0 = no; 1 = yes). Those that were introduced were classed as naturalised or not (0 = no; 1 = yes). Between the years 1843 and 1993, 54 species of clover were introduced intentionally to NZ for agricultural purposes, of which nine species have naturalised. Of the naturalised species, only three are environmental weeds (Howell, 2008). Therefore my test cannot test the transition from naturalised to invasive for *Trifolium* because there are too few data points.

Following (Gravuer *et al.*, 2008) I used variables that might directly or indirectly determine success at the introduction stage (Table 5.1), including: climate match (a measure of how well matched the climate in NZ is to climate in the species native range, on a scale from 0–100); native range size (in km^2, calculated from range maps); the range of conditions tolerated (measured as the number of WWF Ecoregions a species' native range covered); a species presence in Great Britain in the mid-1900's ("GB presence"); and the species traits height and life-span. At the naturalisation stage I used: climate match; native range size; area planted (calculated from historical records that included, seed and plant catalogues, newspapers, government reports/research, Agricultural and Pastoral Show catalogues, advisory documents for farmers, Acclimatisation Society reports, herbarium specimens and sources from the ALLWEEDS database; Gravuer, 2004); residence time (the number of years since first introduction); GB presence; height; and lifespan. I chose these variables for the same reasons as for *Pinus*.

5.3.2 Exploratory path analysis

Path analysis involves a series of linked regression equations specifying the hypothesised causal relationships among variables commonly shown as a path diagram, with arrows specifying the links and direction of causality among variables. A standard approach is to identify one or more path models that represent *a priori* hypothesis about the relationships between variables, and identify the most likely (best-fitting) model from this model set. However, when there are many variables and no strong *a priori* hypotheses about the causal relationships, an exploratory approach that includes all potential hypothesised links between variables is useful.

I fitted path models in a Bayesian framework (e.g. Arhonditsis *et al.*, 2006; Lee, 2007; Rupp *et al.*, 2004; Scheines *et al.*, 1999) using a variable selection method outlined by Duncan *et al.* (2011) to determine: the probability that a given link between variables is present in the best model; and the model-averaged weightings (or "effect size") for each link (calculated as the probability of a link being present in the best model multiplied by the parameter value of a link). The process of fitting a path model in this way involves four steps: (1) construct the path diagram including all biologically plausible links between variables; (2) translate the path diagram into a series of linked regression equations; (3) fit the full model using uninformative priors for the regression coefficients that define the links between variables; (4) refit the model, this time including binary indicator variables that specify whether regression coefficients should be included in the model or not, and using the posterior distributions of the regression coefficients from the full model as priors.

5.3.2.1 Steps 1 and 2: path diagrams and linked regression equations

Based on my hypotheses of how the different factors could influence one another and introduction and naturalisation success (below), I developed the path models shown in Figure 5.1. These models include variables that are binary (introduced or not; naturalised or not), continuous (e.g. the Z-score), and categorical (e.g. life-span). I hypothesised that:

At the introduction stage for *Pinus*: all variables could directly influence introduction success, such that species with a closer climate match, that are common (large native range), with fast growth and reproduction, that are widely used in forestry, might be preferentially introduced because these traits make them more suitable of forestry and horticultural species. I also considered that species with higher Z-scores may have a larger native range size because they have traits that may make it easier to spread widely, and were more likely to be used in forestry because of their fast grown and prolific reproduction; species with large native ranges were more likely to be used in forestry because they are more commonly available for introduction; and species with a larger native ranges could be more climatically suited to NZ because their range may cover a larger number of climatic zones.

At the naturalisation stage for *Pinus*: all the links between explanatory variables possible at the introduction stage were also possible at the naturalisation stage; all variables were able to directly influence naturalisation success, although the effect of native range size *per se* could not plausibly be direct and thus any direct influence identified by the model would suggest the presence of an unmeasured variable between native range size and naturalisation success; all variables except area planted were able to affect residence time, because I might expect that climatically suited, common species with a large native range, that were widely used in forestry and had high Z-scores, would be preferred and thus introduced earlier; all variables were able to determine area planted, for the same reasons as residence time, and a longer residence time was likely to result in a larger area planted.

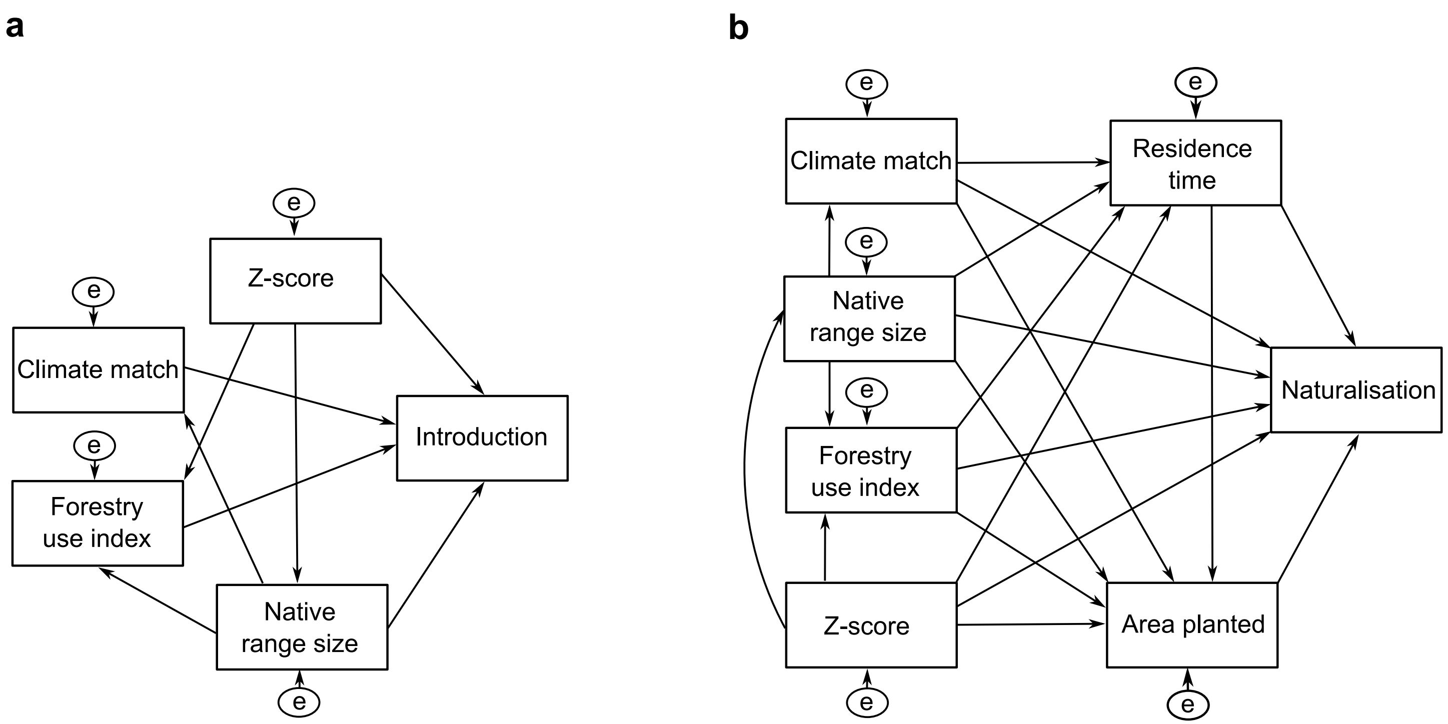

Figure 5.1. Path diagrams showing hypothesised causal links between measured variables (square boxes) in the most complex possible models for (a) introduction and (b) naturalisation of *Pinus*; and (c) introduction and (d) naturalisation of *Trifolium* in New Zealand. Arrows represent the direction of dependencies between variables; the absence of an arrow between two variables indicates that the variables are not causally linked. Oval boxes indicate that unobserved variation ("e") was modelled.

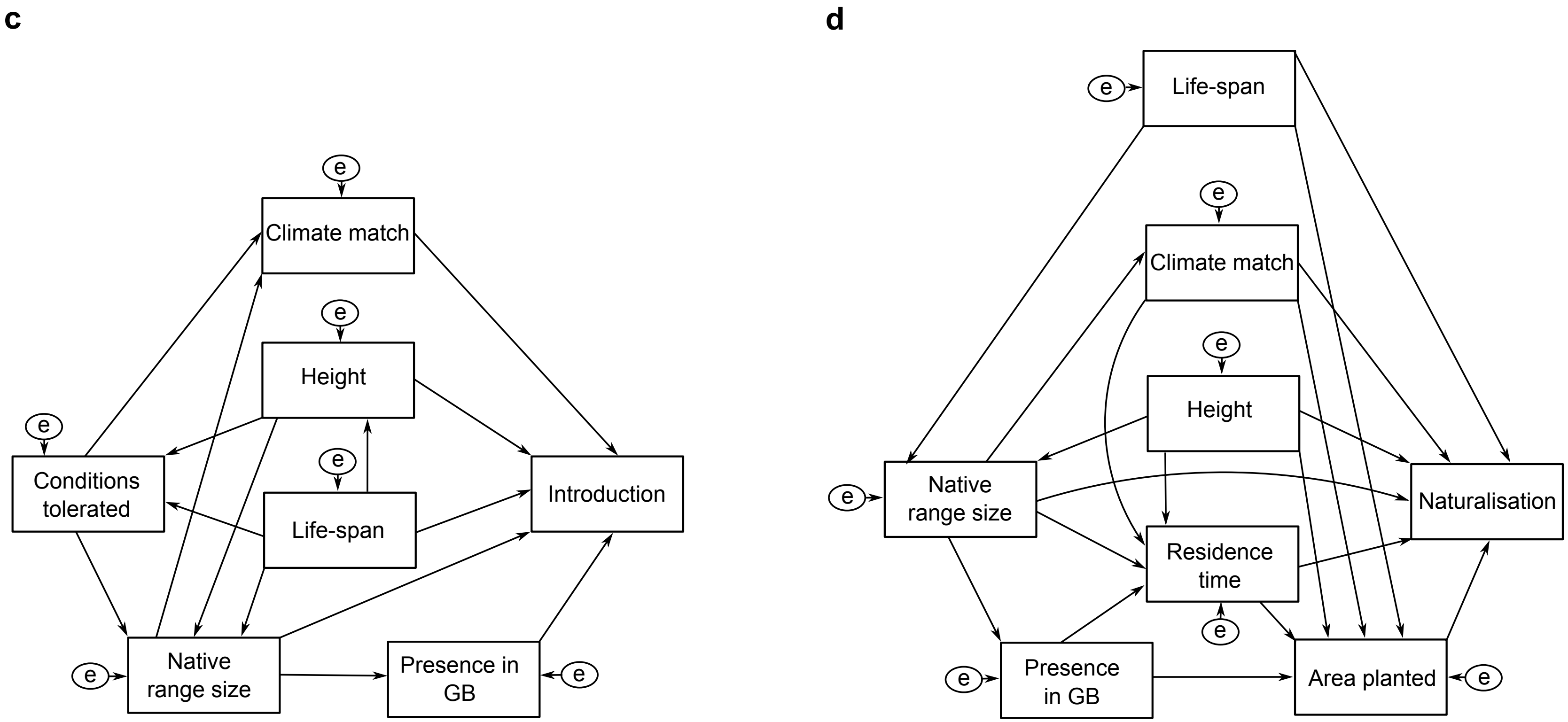

Figure 5.1. **Continued from previous page. . .** Path diagrams showing hypothesised causal links between measured variables (square boxes) in the most complex possible models for (a) introduction and (b) naturalisation of *Pinus*; and (c) introduction and (d) naturalisation of *Trifolium* in New Zealand. Arrows represent the direction of dependencies between variables; the absence of an arrow between two variables indicates that the variables are not causally linked. Oval boxes indicate that unobserved variation ("e") was modelled.

At the introduction stage for *Trifolium*: all variables could directly influence success given that species with a closer climate match, that are common (large native range), already present in GB (thus likely to be known to early NZ settlers), with greater biomass (height) and life-spans suited for commercial agricultural use may be preferentially introduced; the range of conditions tolerated may determine native range size and climate match; life-history traits (height and life-span) may determine the range of conditions tolerated and native range size of a species such that small short lived species are more widespread; being common (having a large native range) could have determined presence in GB such that widespread species were more likely to be in GB; and a large native range may increase the chance of being from a region that is climatically suited to NZ.

At the naturalisation stage for *Trifolium*: all the links between explanatory variables possible at the introduction stage were also possible at the naturalisation stage; all variables were able to directly influence naturalisation success except for presence in GB; as for *Pinus* the effect of native range size on naturalisation success was included to indicate whether there was a missing variable between native range size and success; all variables except area planted and life-span were able to affect residence time, because one might expect that climatically suited, common species with a large native range already present in GB and known to NZ settlers, with large biomass (height) would be preferred and thus introduced earlier; all variables were able to determine area planted for the same reasons as residence time, and a longer residence time was likely to result in a larger area planted.

The path models shown in Figure 5.1 translate into four sets of linked regression equations for introductions and naturalisations which are given in Appendix F.

5.3.2.2 Step 3: fitting the full path model

Continuous explanatory variables were standardised by subtracting their mean and dividing by two standard deviations to assist with model convergence and ensure that parameter estimates for continuous and binary explanatory variables were on a comparable scale (see Gelman, 2008). Being a Bayesian model, I had to specify priors for each unknown parameter reflecting my prior belief regarding the value each parameter is likely to take. The benefits of being able to include prior information in this way include objectively stating any prior beliefs in a model that can then be formally tested against models with different priors, and being able to fit more complex biologically realistic models.

I chose to include uninformative priors to allow the data to drive parameter estimation. I specified flat normal prior distributions for all regression coefficients as having a mean 0 and variance 1000. I specified uniform priors on all standard deviations in the range of 0–100 for the variance terms following Gelman (2006). To avoid problems in model fitting associated with separation (when a linear combination of predictors perfectly predicts the outcome) but still allow the data to drive parameter estimation, I followed Gelman *et al.* (2008) and applied a weakly informative prior to all constant terms on the logit for binary variables. These priors specified values as being drawn from a Cauchy distribution centred on zero (i.e. a t-distribution) with one degree of freedom, with a scale of 10 for regression intercepts and a scale of 2.5 for binary predictors. These priors apply a low probability to changes of 10 on the logistic scale (which equates to the probability of outcome x going from 0.01 to 0.99), which were reasonable assumptions for my data.

Bayesian path models were fitted using Markov chain Monte Carlo (MCMC) methods implemented in the OpenBugs software (Thomas *et al.*, 2006) in R version 2.13.1 (R Development Core Team, 2011). I ran three MCMC chains (Clark & Gelfand, 2006) for 10,000 iterations after a burn-in of 10,000, which was sufficient to achieve convergence (judged by inspection of the chain histories). In order to produce a set of priors necessary for subsequent variable selection models, I sampled and saved the posterior distribution (the post-burn-in iterations) of each chain, for all regression coefficients (β_{x_i}) and variances ($\sigma^2_{x_i}$) (see Appendix G for the R code).

5.3.2.3 Step 4: refitting the model

In order to quantify the probability that a link between any two variables would be present in the best model of introduction and naturalisation, and the model-averaged weightings (effect sizes) of each causal link, I followed Duncan *et al.* (2011) and expanded the linked regression equations outline in Appendix F by assigning each explanatory variable a binary indicator variable ($w_{x,j}$) that specifies whether regression coefficients associated with that link should be included in the model or not. This binary indicator variable takes the value of 1 if the link between variables x and j is included in the model and 0 if it is not. The variable effectively switches linkages on and off. The binary indicator variables were given prior distributions that equated to a uniform prior on the number of links to include in a model (Ley &

Steel, 2007):

$$w_{x,j} \sim \text{Bernoulli}(\Theta); \Theta \sim \text{Beta}(1,1) \tag{5.1}$$

After all MCMC iterations, a distribution of values (either 0 or 1) was obtained for each binary indicator variable. The mean of this distribution can be interpreted as the probability that a link between variables x to j would be included in the most probable model defined by all direct and indirect links. Linkages that are consistently selected (mean $w_{x,j}$ close to 1) are more likely to determine introduction or naturalisation than variables that are rarely selected (mean $w_{x,j}$ near 0). Finally, the binary indicator variable ($w_{x,j}$) and the regression coefficient that the indicator variable is switching on and off were multiplied, which gives the model averaged value ('effect size') of a given linkage.

The priors for the refitted model (step 4) were different from the initial full model (step 3). Uninformative priors are useful when fitting initial models, because they allow the data to drive parameter estimation. However, uninformative priors are usually not uninformative when they are used to assess model probability (Link & Barker, 2006). Therefore I followed Duncan *et al.* (2011) and Aitkin (1991) and used the posterior distributions for regression coefficients and variances saved from the initial model (step 3) as priors for the variable selection model (step 4).

The refitted models (step 4) were fitted in the same way as the initial full models (step 3), using OpenBugs (Thomas *et al.*, 2006), with three MCMC chains, run for 10,000 iterations after a burn-in of 10,000 iterations, which achieved convergence (see Appendix G for R code).

Finally, I calculated the area under the receiver operating curve (AUC, Hanley & Mcneil, 1982), which provides a measure of the degree to which the fitted values discriminate between observed outcomes. An AUC value of 1 indicates that the model perfectly discriminates between outcome 1 (e.g. introduced) and outcome 0 (e.g. not introduced), while a value of 0.5 indicates that the model performs no better than chance.

5.4 Results

5.4.1 Determinants of introductions

Pinus species were more likely to be introduced if they had a close climate match, were used widely in forestry, had a large native range size and higher Z-scores

(Figure 5.2). Both climate match and the forestry use index had a 100% probability of being present in the best model of introduction success (see Appendix H). The effect size of native range size (0.60) and the Z-score (0.26) on introduction success was two orders of magnitude smaller than climate match (14.17) and the forestry use index (10.58). Species with a large native range size were slightly more likely to have a higher forestry use index although this effect was also relatively small (Figure 5.2). All of the strong links determining pine introduction success were direct. The AUC value for *Pinus* introductions was 0.93 indicating that the model performance was "outstanding" (Hosmer & Lemeshow, 2000).

In contrast to *Pinus*, more indirect links were present for *Trifolium* introductions. Native range size had the strongest effect in the model and this was indirect, such that species with larger native ranges were less likely to be present in GB. *Trifolium* species with a larger native range size, not present in GB in the mid-1800's that were taller, lived longer, and had a closer climate match to NZ, had an increased probability of introduction to NZ (Figure 5.2). All variables that were modelled as having a direct causal effect on introduction success had $\geq$ 49% probability of being included in the best model of introduction success (see Appendix H). The AUC value for *Trifolium* introductions was 0.82 indicating that the model performance was "excellent" (Hosmer & Lemeshow, 2000).

5.4.2 Determinants of naturalisations

The naturalisation success of pines was determined by (in order of importance) direct links with the forestry use index, climate match, the Z-score, area planted, residence time and native range size (Figure 5.2). The direction of these effects was positive such that species with a closer climate match, greater forestry use index, larger area planted, longer residence time, higher Z-score and larger native range size were more likely to naturalise. Climate match also had indirect effects on other explanatory variables such that a closer climate match increased the area planted, with a relatively small effect size but a 99% probability of being included in the best model (see Appendix H). All other indirect links were relatively weak (Figure 5.2). As at the introduction stage the majority of links were directly between explanatory variables and naturalisation success. The AUC value for *Pinus* naturalisations was 1.00 indicating that the model produced near perfect predictions.

In contrast to the introduction stage, *Trifolium* naturalisation success was increased mostly through direct links between explanatory variables and naturalisation success, however the strong indirect effect of native range size on presence in GB was still evident (Figure 5.2). Naturalisation success was increased, in order of importance, by a closer climate match, larger native range, longer residence time and larger area planted (Figure 5.2). Weaker links between height and life-span with naturalisation success suggested that smaller, short-lived species were more likely to naturalise than larger, long-lived species. Presence in GB increased both the residence time and the area planted and species with longer residence times were more likely to be planted widely. Species with larger native ranges were more likely to have a close climate match to NZ. All direct links with naturalisation success were present in $\geq$ 52% of the best models of naturalisation (see Appendix H). The AUC value for *Trifolium* naturalisations was 0.99 indicating that this model produced near perfect predictions.

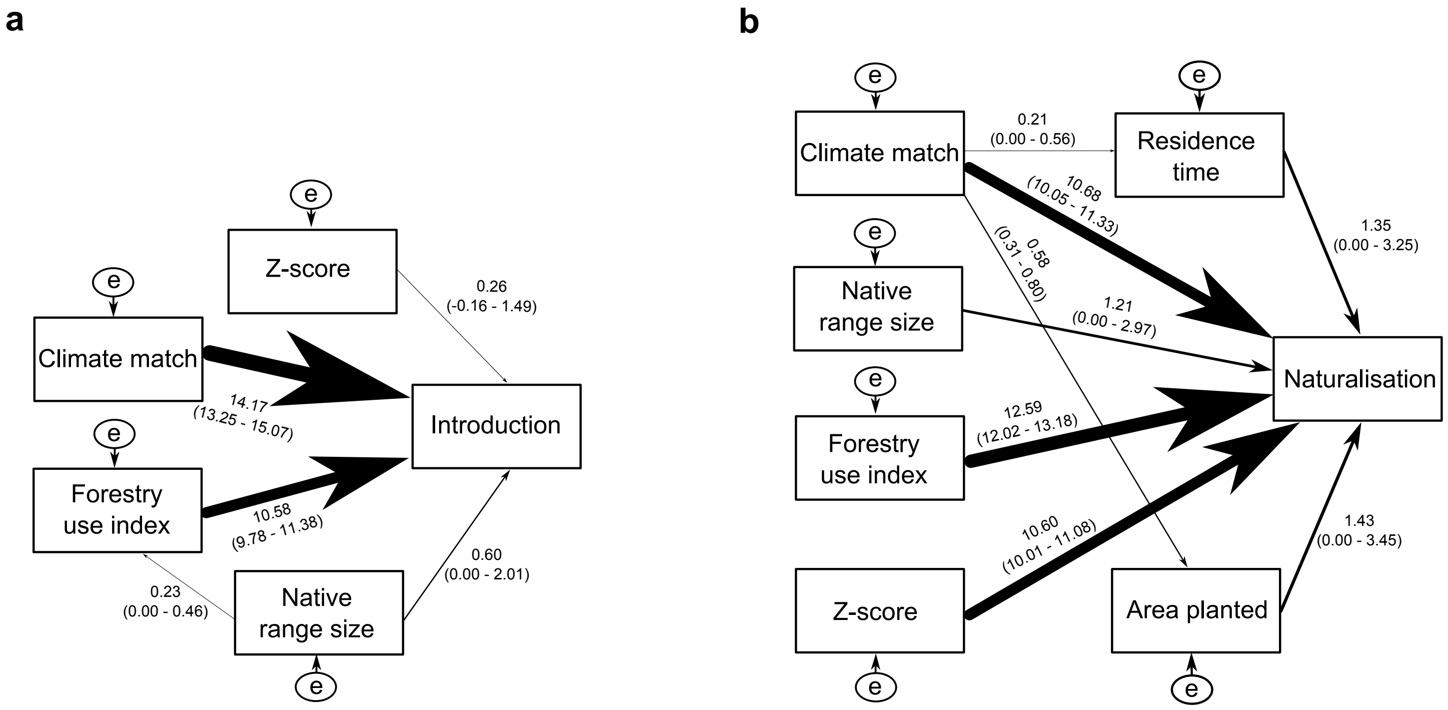

Figure 5.2. Results of exploratory path analyses for (a) introduction and (b) naturalisation of *Pinus*; and on the following page (c) introduction and (d) naturalisation of *Trifolium* in New Zealand. Square boxes indicate measured variables. Oval boxes indicate that unobserved variation ("e") was modelled. Arrows represent the direction of dependencies between variables; the absence of an arrow between two variables indicates that the variables are statistically independent in the model (model averaged weightings were zero or the link was not modelled; see Appendix H). Arrows between measured variables are sized according to the model averaged effect size for a given link (the probability of a link being present in the best model multiplied by the parameter value for the link); the model averaged effect size for a given link ±95% credibility intervals are written next to arrows (only those with a model averaged effect size of ≥ 0.20 are labelled). **Continued on following page. . .**

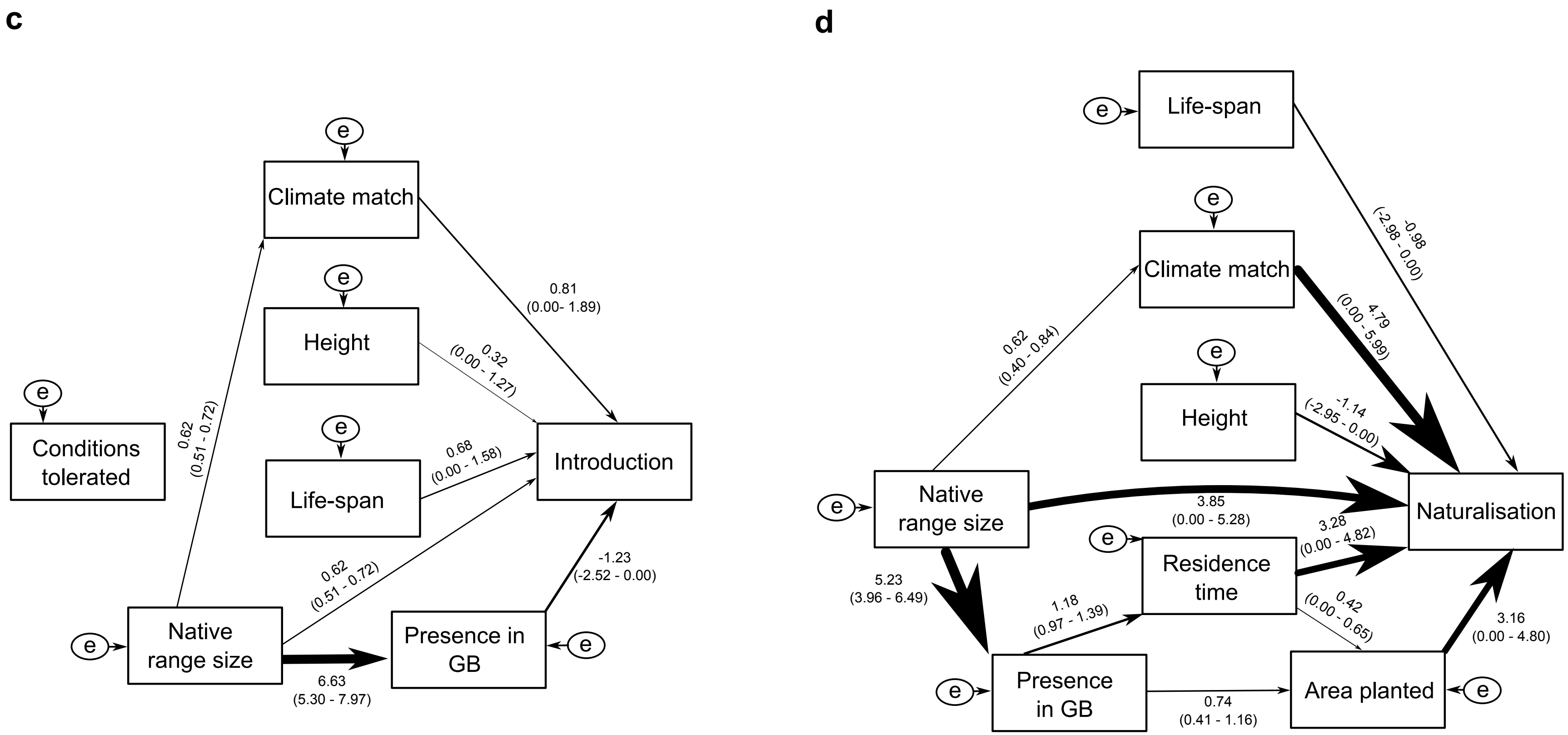

Figure 5.2. **Continued from previous page. . .** Results of exploratory path analyses for (c) introduction and (d) naturalisation of *Trifolium* in New Zealand.

5.5 Discussion

At the introduction stage species traits had relatively weak links with all other variables and with success compared to human and biogeographic factors, and that *Pinus* introductions were determined by direct links whereas *Trifolium* introductions had more indirect links (Appendix H). Naturalisations were largely determined by direct links in both genera and at this stage life-history traits were more important determinants of success than at the introduction stage, although human and biogeographic factors were still the strongest predictors (Appendix H). Climate match was a strong direct determinant of pine introduction and naturalisation success and also acted indirectly to increase the area planted, but for *Trifolium* climate match did not affect the area planted. Propagule pressure (both area planted and residence time), a close climate match, large native range size and traits related to fast reproduction were a common direct mechanisms determining naturalisations in both genera. However the causal links modifying these variables were more complex for *Trifolium* than *Pinus*, particularly at the introduction stage, suggesting different mechanisms operate in these taxonomic groups.

Despite some difference there were several broad similarities between the causal links for *Trifolium* and *Pinus* that could suggest commonalities in the invasion process for different plant genera. First, a greater area planted and longer residence time (both measures of propagule pressure) were directly linked to increased naturalisation success in both genera. This supports previous studies highlighting the consistent importance of these variables in naturalisation and invasion success (Castro *et al.*, 2005; Essl *et al.*, 2010; Křivánek *et al.*, 2006; Pyšek *et al.*, 2009b), and suggests that these may be generally applicable variables which should always be accounted for in models of invasion success (Colautti *et al.*, 2006). However, I have gone further and shown that propagule pressure is itself directly increased by other key variables that are often identified as determining naturalisation and invasion outcomes such as climate match (Essl *et al.*, 2011; Nuñez & Medley, 2011) and a human variable related to previous human familiarity to the species (presence in GB). My results suggest that standard regression approaches that simply partition the variation explained by each variable in a single step could over-emphasise the importance of a given variable's direct influence on success.

Second, the influence of species attributes on both introduction and naturalisation were relatively weaker than the influence of biogeographic and human variables. Although research directly comparing the effects of species attributes against other factors is limited, my findings align with previous research suggesting that species attributes may be relatively less important predictors of success than other factors (e.g. Forsyth *et al.*, 2004). These results suggest that further studies describing determinants of invasion success should include a complete range of biogeographic, human/economic, and landscape-scale factors as well as species attributes, in order to correctly determine the importance of individual factors.

I showed some marked differences in the causal mechanisms between the two genera used. First, in all models for both genera, biogeographic variables had the largest effect size but these were different between genera. For *Pinus* climate match was the strongest variable whereas for *Trifolium* native range size was the strongest variable and climate match had a relatively low rank. I also found that for *Pinus* climate match had the strongest direct effect on success, whereas for *Trifolium* native range size acted most strongly indirectly through presence in GB. For pines, my findings are consistent with previous research on naturalisation and invasion success in woody species that show climate suitability between the native and introduced range is a key predictor of naturalisation success (Essl *et al.*, 2011; Nuñez & Medley, 2011). However, my results suggest that the mechanism of action of climate suitability may vary for different taxonomic groups. Second, although I found that propagule pressure acted the same way in both genera, the variables that in turn determined the amount of planting and residence time were different. For *Pinus*, a good climate match increased the residence time and also the area planted. For *Trifolium*, however, climate match was not linked to either measures of propagule pressure.

Path analysis allows for causal links between variables to be inferred, which suggested the presence of latent (unmeasured) variables in my dataset between native range size and naturalisation success. Native range size was modelled in my study as having potentially direct effects on naturalisation success for both genera as well as indirect effects through links with other explanatory variables. I included direct links in order to test for the presence of unmeasured variables in the system, given that a species having a large native range size cannot *per se* increase the chance of naturalisation occurring. Therefore native range size must act indirectly though other variables. Indeed for *Trifolium* I saw indirect effects of a large native range increasing residence time and the chance that a species was present in GB. However, the relatively strong direct link between native range size and naturalisation success seen in both genera suggests that there is a variable missing between native range size and naturalisation success. This shows how path analysis could be used to initially identify areas where there are potentially missing explanatory variables and thus be used to guide future data collection. This method for exploratory path analysis could be extended to explicitly include latent variables (SEM; Grace, 2006; Shipley, 2000).

When comparing my findings to the studies that the explanatory variables were drawn from (Gravuer, 2004; Gravuer *et al.*, 2008, Chapter 3) it is clear that path analysis produced a more detailed picture of the invasion process than simpler regression based methods. For example, Gravuer *et al.* (2008) used boosted regression trees (BRT) which assigned a "relative contribution" of a variable to the outcome of *Trifolium* introductions to NZ. Gravuer *et al.* (2008) found that only one PC-axis which was highly correlated with human traits (including presence in GB and residence time) predicted naturalisation success. Using path analysis, I was able to dissect these relationships. Path analysis revealed that the effect of presence in GB acts indirectly on introduction success such that a species that was present in GB was likely to have been introduced to NZ earlier and planted more widely. The original study was also not able to detect the relatively strong effects of both native range size and climate match on presence in GB (NZ and GB have relatively similar climates, such that the climate match for a species to NZ is highly correlated with its match to GB), thus biogeographic variables appeared to be of little influence on *Trifolium* naturalisations. However, the strong effect of presence in GB identified by this and the original study appears to be a product of biogeographic factors, but acts through other human factors; a mechanism that could not have been detected using the original method.

Similarly, in Chapter 3 I used a BRT approach for *Pinus* that highlighted broadly the same relative importance of variables at each stage as the path analysis presented in this study (human and biogeographic variables being ranked more highly than species attributes). However, the analysis in Chapter 3 found that the Z-score had a much smaller relative importance (∼8%; Table 3.3) on naturalisation success than climate match (∼30–50%; Table 3.3), whereas path analysis ranked them as almost equal (Figure 5.2). These contrasting results could be due to the different statistical methods used to analyse the relationship between Z-score and naturalisation success in each chapter. Given that all naturalised pines are also considered invasive in NZ, the results from the present chapter support previous studies that also found the Z-score to be able to discriminate between invasive and non-invasive pines (Grotkopp *et al.*, 2002; Rejmánek & Richardson, 1996). Nevertheless, the Z-score is not a better predictor of naturalisation success than the forestry use index (how widely used a species in for forestry globally) or climate match, highlighting that a good understanding of how different variables influence invasion success is important. Path analysis showed that pines with a closer climate match to NZ were planted over a larger area and tended to have a longer residence time. Thus, the strong effect of climate identified in Chapter 3 is multifaceted and more complex than originally identified. These findings suggest that standard regression approaches may be missing detail and falsely identifying the level of influence of some variables.

The present study represents a step forward in the analysis of factors determining success across the invasion process, and thus preventing and managing invasions, by identifying mechanistic relationships. I have highlighted the scope for a much wider application of path analysis/SEM to invasion questions. While I used several variables in my models that represented three major groups (species, biogeographic and human) there are other potential explanatory factors that I did not include, such as relative growth rate (Grotkopp *et al.*, 2002), human population density, per capita gross domestic product and the number of native *Pinus* species (Essl *et al.*, 2011). More insights into the generality of the links between different factors could be gained by applying this method to multiple regions for which data is already available for the same group of species (Essl *et al.*, 2011). Additionally, assessing how common various links are between variables for different functional or taxonomic groups may provide generalities that could inform future risk assessment and management of introduced species. Finally, extending this exploratory path analysis method (where latent variables are not accounted for) to be a full SEM (where latent variables can be included) would be a logical extension (Grace, 2006; Shipley, 2000).

5.6 Conclusions

I have presented a method for exploratory path analysis that is flexible and can easily incorporate variables from different sampling distributions, including binary variables. This method is also able to assess the probability of linkages being present in models where there are no strong *a priori* hypotheses and where the potential number of models makes standard model comparison approaches impractical. This study is the first to apply path analysis to assess how factors determine success at different stages of the invasion process. I used the genera *Pinus* and *Trifolium* as case studies which are both economically important and intentionally introduced. The extent to which findings from these genera, where some members have experienced large propagule pressure, are generalisable to accidentally introduced species is not clear. I found that human factors (area planted and residence time) acted directly to increase naturalisation success, and that these mechanisms were repeated in two genera, suggesting they are general. However, the factors determining area planted and residence time were different between the two genera. My analysis revealed the differing causal mechanisms and strength of important biogeographic variables between taxonomic groups. My results suggest that variables identified as determining invasion outcomes in traditional regression studies can act via different causal mechanisms between taxonomic groups and that the results from one group may not apply directly to others. I also found evidence of latent variables, which are likely to be common but unidentified in previous studies, that could be targeted for further research. There is scope for path analysis to be more widely applied to identifying the drivers of invasions, with the aim of improving prevention and management of invasive species.

Chapter 6

Conclusions

6.1 Thesis aims

The aim of this thesis was to quantify the risk of invasions from introducing alien species. This was achieved by using the genus *Pinus* as a model system, which has parallels with other commercial tree genera, to: (1) assess how robust the WRA is to taxonomic range, region, and knowledge of invasive behaviour elsewhere; (2) determine the relative contribution of species, biogeographic, and human attributes to introduction and naturalisation/invasion success; (3) quantify phylogenetic signal in introduction and naturalisation success, and the traits that determine success to assess whether phylogeny is a useful predictor of invasion risk; and (4) disentangle the direct and indirect effects of a range of variables on introduction and naturalisation success through a novel exploratory method for path analysis.

6.2 Main results

6.2.1 Weed risk assessment

Chapter 2 provided a novel test of how robust the WRA is at distinguishing between species at each stage of the invasion process from within a genus of relatively similar species, across two climatically similar regions. The performance of the WRA at the introduction and naturalisaion stages as measured by AUC values was good, which supports previous studies that also found the WRA performs well at retrospectively classifying species as invasive or not (Gordon *et al.*, 2008b; Roberts *et al.*, 2011). However, when critical information on prior invasiveness was not available the WRA was no longer able to identify successfully introduced or naturalised species and yielded a success rate no better than chance. The WRA was also not able to predict the number of regions that a species had naturalised in, as it was found to do in previous studies (Dawson *et al.*, 2009b). This finding could be partly explained by the necessarily coarse resolution of our data compared to the previous study (Dawson *et al.*, 2009b).

Given these results, it is debatable how useful the WRA protocol is for application to forestry species (Hulme, 2012). It could be argued that the WRA is a useful tool for identifying potential future species that are likely to be invasive because many of the candidate forestry species will already be in use or have been trialled in other countries. Therefore any invasive behaviour would have had an opportunity to manifest itself. On the other hand, the WRA could be a waste of time if the species being considered for introduction do not have any history of introductions and cultivation outside their native ranges. Additionally, if their history of cultivation has been short invasive behaviour may not have had sufficient time to manifest itself given that long-lived woody species can have lag-phases of over 100 years (Křivánek & Pyšek, 2008; Richardson *et al.*, 1994).

One clear drawback of a full weed risk assessment is that the WRA can be time consuming to complete with each of the pine assessments taking between 5–20 hours of research and no guarantee that this work will result in an accurate assessment. Furthermore, for poorly known species the assessment takes longer to complete and information for the critical questions is less likely to be available–effectively the assessment is least accurate for the species that are of most interest. This suggests that the application of the WRA should be considered on a case-by-case basis after considering the introduction and cultivation histories of the species of interest, and that an adaptive approach to weed risk assessment could be implemented (e.g. Hulme, 2012; McGregor *et al.*, 2012). Such adaptive management could include grading the reliability of the WRA based on how many questions were answered and how long it took to complete, and flagging assessments where the key questions about invasion elsewhere could not be answered. For species where this information does not exist, obtaining the required data through experimentation is unlikely to be feasible for long-lived woody species and the WRA may not be a feasible approach at all. In such cases, quantifying risk through other factors that are important determinants of invasion success may be the most sensible approach.

6.2.2 Factors determining introduction and naturalisation

In Chapter 3 I identified a set of factors that determined success and failure at the introduction and naturalisation stages for pines in two climatically similar regions and showed that these factors were similar between regions. I found that species with greater climate suitability, a large native range size, that were widely used globally for forestry (had a higher forestry use index) were the top predictors of introduction success. These factors, as well as greater propagule pressure, determined naturalisation success. Species attributes contributed relatively little at both stages.

My results largely support recent work on pines that found that greater global forestry use (as measured by a forestry use index; Procheş *et al.*, 2012) and a close climate match between the native and introduced regions (e.g. Nuñez & Medley, 2011) increase naturalisation and invasion success. A greater area planted for forestry was also important in predicting naturalisation success in the southern hemisphere region, which supports previous studies that found propagule pressure to be a key determinant of invasion success (e.g. Pyšek *et al.*, 2009b). In contrast with previous studies, I found that species attributes were not an important predictor of success or failure for pines (Grotkopp *et al.*, 2002; Rejmánek & Richardson, 1996). My study is the first to quantify the factors controlling introduction success for pines and I showed that global forestry use and a close climate match were also good predictors of introduction success. The novel aspects of this research were in quantifying the *relative* importance of these factors, in accounting for failures at the introduction stage and in comparing the relative importance of factors for the same group of species in more than one region.

The statistical approach I used (BRT) was able to classify which species succeeded and which failed at each stage as well or better than the WRA (based on AUC values; Table 6.1). Given that the WRA for lesser-known species can require a large amount of time to be spent on research to answer each question, the time spent on a WRA becomes comparable to the time spent collecting and analysing data in a BRT approach. Therefore, the BRT approach could be a viable option when quantifying risk for lesser known species where key WRA questions my be impossible to answer. The advantage of the BRT method is that it doesn't rely on questions such as a species having a history of invasion elsewhere. However, the disadvantage of the BRT approach is that it does rely on good data being available to quantify native range size in order to calculate climate match, which may not be achievable for every species (as was the case in my study, where several species had too few data points to calculate a climate match). If a species was a high priority for assessment, information about the native distribution could be collected on the ground or may become available in the future, whereas if a species has no history of invasiveness elsewhere because it has never been introduced elsewhere there is no way to answer that key WRA question.

6.2.3 Phylogenetic signal in pine invasions

In Chapter 4 I calculated phylogenetic signal for introduction and naturalisation success for *Pinus* and a range of traits known to determine success in order to determine if phylogeny could be a useful proxy for measuring risk. There was weak and non-significant phylogenetic signal of introduction and naturalisation success across the phylogeny for *Pinus*. Likewise, there was no phylogenetic signal in any of the traits that had been previously identified as determining success at each stage.

When considered at higher taxonomic levels such as family and order, there are only three families that have been identified by more than one study as having more invasive members than expected by chance (Daehler, 1998; Pyšek, 1998; Vilà & Muñoz, 1999). As discussed in Chapter 4, risk identification at these high taxonomic levels may not be useful for informing management actions where information on individual species in necessary (e.g. Pyšek *et al.*, 2009a; Schmidt & Drake, 2011b). For phylogenetic relationships to be useful for risk assessment, species-level variation in invasion success must be correlated with phylogeny. Currently there is insufficient literature available to assess how likely it is that traits determining invasion risk show phylogenetic signal at the species-level, although for *Pinus* and *Acacia* (Miller *et al.*, 2011) it appears that it is not present. These results suggest that phylogeny alone may be a poor predictor of invasion risk for some woody groups and that too little is known about the distribution of risk across phylogenies at the appropriate level to be informative. Currently, phylogeny is not a viable alternative approach to other forms of risk assessment. However, in order to draw and general conclusions about the role of phylogeny in predicting risk at a useful taxonomic level, further studies are needed that include a wider taxonomic range (e.g. all members of several genera).

Although these conclusions may seem somewhat pessimistic, it is still essential for invasion biologists to incorporate phylogenetic information into analyses of invasion risk in order to control for possible statistical non-independence among taxa (Harvey & Pagel, 1991). When determinants of both introduction and naturalisation success were analysed with both phylogenetically and non-phylognetically controlled mixed models, there were no differences in the effect size or significance of any variables. These results support previous invasion studies that have compared phylogenetic and non-phylogenetic models and found that including phylogeny makes little difference to the results (e.g. Dawson *et al.*, 2009a, 2011b; Küster *et al.*, 2008; Pyšek *et al.*, 2009a). This suggests that traits linked to phylogeny such as life-history traits and native range size may not determine invasion risk and instead traits such as propagule pressure, which is less likely to be determined by evolutionary pressures, determine risk.

6.2.4 Casual links between factors

In Chapter 5 I used a novel method for exploratory path analysis to identify the causal links between variables and introduction and naturalisation success for *Pinus* and *Trifolium* in New Zealand. I aimed to assess how the causal links differed for two genera introduced to the same region, where the genera being considered were functionally very different but both economically important (i.e. long-lived woody species used for long-term forestry and horticulture vs. short-lived herbs used to enrich pasture with nitrogen). This is the first example of path analysis applied to the question of how factors determine success through the invasion process.

The key findings from Chapter 5 were that *Pinus* introduction and naturalisation success was determined by direct effects of human, biogeographic and species factors. In contrast, *Trifolium* introduction and naturalisation success was a mixture of indirect effects of factors on one another and direct links between human, biogeographic and species factors with success. For *Pinus* climate match was the strongest variable, whereas for *Trifolium* native range size was the strongest variable and climate match had a relatively low rank of importance. There were also broad similarities between the causal links for *Trifolium* and *Pinus* that could suggest commonalities in the invasion process for different plant genera. I found that a greater area planted and longer residence time (both measures of propagule pressure) were directly linked to increased naturalisation success in both genera. Propagule pressure in turn was directly increased by other key variables that are often identified as determining naturalisation and invasion outcomes such as climate match (Essl *et al.*, 2011; Nuñez & Medley, 2011) and a human variable related to previous human familiarity to the species (presence in GB). I also found that the effect of the Z-score in pine naturalisations was as large as the effect sizes of climate match and the forestry use index. This contrasted with the results from Chapter 3 where Z-score had relatively little ability to discriminate between naturalised and non-naturalised species. The results from Chapter 5 suggest that standard regression approaches that simply partition the variation explained by each variable in a single step could over- or under-emphasise the importance of a given variable's direct influence on success.

The main advantage of using path analysis over other methods that are commonly used to quantify factors determining success, such as regression trees or standard linear and linear mixed regression models, is that path analysis can be used to identify causal linkages. I found that path analysis revealed a more detailed picture of *Pinus* and *Trifolium* introductions and naturalisations. The original *Trifolium* study that my data were drawn from (Gravuer *et al.*, 2008) identified human variables as determinants of naturalisation success. However, path analysis revealed that biogeographic variables mediated through human variables played a role in naturalisation success. Furthermore, the novel method for exploratory path analysis that I used was able to handle two common scenarios encountered in multivariate data sets addressing a complex problem, namely multiple alternative models with no strong prior hypotheses and variables from different sampling distributions, in an objective and flexible way. Whilst model comparison methods such as AIC, DIC, BIC and Chi-squared tests are available, systems with more than three variables and no strong hypothesised links can quickly become unmanageable using model comparison methods.

The exploratory path analysis method presented in this thesis could be used more widely by invasion biologists because: (a) all use of path analysis in the invasion literature to date has focused on explaining native-alien species richness rather than invasion drivers (e.g. Atwater *et al.*, 2011; Harrison *et al.*, 2006; Hulme, 2011; Keeley *et al.*, 2005; Seabloom *et al.*, 2006; Taylor & Irwin, 2004); and (b) all previous studies have used a model comparison method to determine the “best” models, without any standardised approach to dealing with the uncertainty around these “best” models or how many models are needed in the *a priori* model set under comparison in order to be sure of identifying a suitable model (Atwater *et al.*, 2011; Harrison *et al.*, 2006; Hulme, 2011; Keeley *et al.*, 2005; Seabloom *et al.*, 2006; Taylor & Irwin, 2004). The exploratory method I presented could easily be expanded to include latent (unmeasured) variables and thus be a full SEM approach (Grace, 2006; Shipley, 2000) but specifying latent varaibles in the model as being drawn from an appropriate distribution with uninformative priors. Developing SEMs could facilitate the identification of missing information for a system and focus future data collection towards defined goals. Future work using path analysis and full SEM for analysing the determinants of invasions could extend our knowledge of how general the causal mechanisms of invasion success are by comparing the causal relationships between numerous taxonomic or functional groups, and by comparing the determinants of success or failure at each stage of invasion across multiple regions. This would provide new insights into the similarity in transition between stages across taxonomic and geographic boundaries.

Chapter 5 demonstrates how invasion biologists could move towards using statistical approaches such as path analysis and SEM that can provide mechanistic insights into invasion rather than just describing patterns. Using Bayesian approaches and exploratory analyses can provide estimates of uncertainty in model selection and confidence that all potential explanations involving a set of variables have been considered.

6.3 Comparison of alternative approaches to risk assessment

In order to directly compare the performance of the four different methods I used to quantify risk in *Pinus* (WRA, BRT, path analysis, and MCMC glmm) I calculated AUC values for all models at the introduction and naturalisation stages in NZ (see methods sections of relevant chapters). I compared performance for NZ only because I did not use path analysis to assess success for *Pinus* in GB.

First, at both introduction and naturalisation AUC values were ≥ 0.80, which is considered as "excellent" prediction according to Hosmer & Lemeshow (2000). For quantifying the risk of introduction in the first instance the WRA was the poorest model. This is perhaps not surprising given that the WRA is not designed to discriminate between introduced and not introduced species. However at the naturalisation stage all models performed "outstandingly" (Hosmer & Lemeshow, 2000). Thus, it appears that no one method gives a significant advantage over any others as far as performance goes. However, as demonstrated in Chapter 2 the WRA is reliant on knowledge of invasive behaviour elsewhere. Without this information the WRA performs no better than chance (AUC ≤ 0.50). If a species has never been introduced outside its native range it would never be possible to collect the required data. Therefore this drawback represents a serious issue that should be addressed by using an additional method of risk assessment. When knowledge of prior invasions is not available the WRA should not be used because it is likely to be unreliable and alternative methods could be considered.

One advantage of the WRA over the other three methods tested is that the WRA provides a clear management recommendation in most cases. Only species that fall into the "evaluate" category are not either directly accepted or rejected (Pheloung *et al.*, 1999). The WRA assessment also requires no specialist statistical analysis skills to complete and interpret and is therefore usable by anyone who may need to perform weed risk assessments. However, the three statistical modelling methods outlined in this thesis could also be used to make predictions if a cut-off probability of naturalisation or invasion was selected. Such a cut-off would depend on the costs of potential invasions as well as the benefits of introduction (Yokomizo *et al.*, 2012), which is outside the scope of this thesis.

Out of the statistical analysis methods (BRT, path analysis, and MCMCglmm) I judged the BRT models to be the easiest to implement in R. BRTs were easy to use because a comprehensive tutorial is available (Elith *et al.*, 2008) that includes R code, and because this method uses machine learning no *a priori* hypotheses needed to be specified. In contrast, the method for exploratory path analysis outlined in Chapter 5 required more complex Bayesian statistics and the development of mechanistic models of introduction and naturalisation success. Phylogenetic mixed models were also more complicated to implement that BRTs, requiring a phylogeny (which in this case was developed directly from molecular sequence data) and *a priori* specification of priors and hypothesised models. Therefore, BRTs offer a relatively straightforward alternative approach when the WRA is unreliable due to missing information on invasion elsewhere.

6.4 Implications beyond pines

6.4.1 Conflict between forestry and invasion risk

My results suggest that there is an intrinsic risk of future invasions when introducing new alien species or more widely planting existing alien species for forestry. Species introduced to NZ and GB had higher WRA scores, were the best climatically suited, had larger native range sizes, and had been used more often for forestry elsewhere than species that were not introduced. Those species which naturalised also had the closest climate matches, which translated into a higher level of propagule pressure. Therefore, these circumstances create a dilemma for foresters and for invasion ecologists: species with desirable traits for forestry, such as fast grwoth rate (measured in this thesis using the syndrome of life-history traits describing rapid development and reproduction) also have traits that are likely to lead to invasions, but species may become invasive because they are planted widely. Therefore traits associated with invasion in forestry trees may be an artefact of widespread planting. There is likely to be a trade-off between the requirements foresters have for selecting suitable species

Table 6.1. Comparison of alternative risk assessment methods: Australian weed risk assessment (WRA) with all knowledge included; boosted regression trees (BRT; AUC values from Z-score models); MCMC phylogenetically controlled generalised linear mixed models (MCMC glmm); and path analysis for assessing the risk of *Pinus* introductions and naturalisations in New Zealand; AUC values at the introduction (AUCi) and naturalisation (AUCn) stage indicating the predictive performance of each method at the naturalisation stage (1 = perfect classification; 0.5 = no better than chance; < 0.5 = model is worst than chance); a summary of main advantages and disadvantages of each approach; and the R package it is implemented in.

Method	AUCi	AUCn	Advantages	Disadvantages	R Package
WRA	0.80	0.97	No specialist skills required; can be relatively rapid; provides clear management recommendation	Potentially unreliable (– knowledge of invasiveness elsewhere); biased towards well-known species; no procedure to deal with missing data	NA
BRT	0.86	0.98	Robust to missing data; relatively straightforward; could utilise relatively few key traits; handles variables from different sampling distributions	No management recommendation; good statistical analysis skills required; relatively time consuming	dismo (Hijmans *et al.*, 2012)
MCMC glmm	0.92	1.00	Can control for phylogeny; gives clear significance values; handles variables from different sampling distributions	Needs molecular data; complicated phylogenetic analysis; does not handle missing data in fixed effects; no management recommendation	MCMCglmm (Hadfield, 2010)
Path analysis	0.93	1.00	Could utilise relatively few key traits; gives mechanistic insights; models missing data; handles variables from different sampling distributions	High level of statistical analysis skills required; relatively time consuming; no management recommendation	BRugs (Thomas *et al.*, 2006)

with future invasion risks. Ultimately, managing the risks from forestry invasions around plantations more consistently may be the only practical solution.

6.4.2 Potential solutions to forestry–invasion conflict

The problem of future forestry invasions is unlikely to be solved by risk assessments as they currently operate for three reasons. First, as noted above foresters introduce species that are good for forestry and intrinsically high-risk. Second, in NZ for example, a large number of alien tree species have been introduced historically and exist in the landscape in low numbers in gardens and arboreta (see Appendix A). There is currently no regulation of the circulation of almost all of these existing species (Howell, 2008) meaning that many potentially problematic species, so called "sleeper weeds", are unlikely to legally require risk assessment. Implementing retrospective risk assessments for species already in a region and regulating the further circulation of sleeper weeds would reduce this risk. Third, current assessment methods do not take a cost-benefit approach to decision making when considering which species to import and may not be practical for assessing commercial species. Optimal decisions for introduction and containment of commercial species depends not only on their probability of their escape but also on the cost of escape and the potential economic benefits of introduction (Yokomizo *et al.*, 2012). Including such cost-benefit parameters in assessments could allow more pragmatic assessment of risk that are suitable for commercial species.

An alternative solution to simply banning the circulation or importation of potentially invasive species based on risk assessment results would be to set industry standards on the layout of plantations and the management of surrounding land. Indeed, a preliminary version of such a national strategy for NZ has been drafted by the Ministry for the Environment (2010). Other solutions put forward to reduce the amount of wilding pine spread in NZ have been outlined by Ledgard and Langer (1999) and these could be applied in other regions. These solutions include planting a belt of low-risk species around the outside of afforested areas to act as a buffer, sighting plantations away from "take-off points" such as the top of hills or in the prevailing wind and managing the land around a plantation with grazing or targeted weed control. Another potential avenue for preventing weed incursions from forestry would be to adopt the "polluter pays" principle (Perrings *et al.*, 2005; Richardson, 1998a) by making plantation owners legally obliged to pay for weed control that results from their forest. Currently in NZ regional councils generally remove wilding trees through a pest-management strategies despite 92% of plantation forests in NZ being privately owned (MAF, 2011). However, if the cost of weed control was passed on to the commercial forestry company the incentive for companies to implement good management practice could reduce the risk of future invasions from plantations.

6.4.3 Implication beyond commercial trees

Biofuel crops have been put forward as a “green” renewable energy source, and their cultivation is gaining momentum. However, biofuel plants have the potential to invade because many of the essential traits for a biofuel crop are also the traits associated with invasiveness such as large stature, high biomass production, efficient resources use and the ability to tolerate a wide range of environments (Raghu *et al.*, 2006). Using an adapted version of the WRA Buddenhagen *et al.* (2009) found that 70% of regionally suitable biofuel crops have a high risk of becoming invasive and are four times more likely to establish in the wild compared to non-biofuel plants in Hawaii.

The situation for biofuels is remarkably similar to commercial afforestation using alien tree species because biofuel species are likely to be planted throughout large areas over long time-scales, creating massive propagule pressure. This risk has led to a call for genotype-specific screening protocols early on in to developing a biofuel species (Barney & DiTomaso, 2008; Low *et al.*, 2011), standards for plantation lay-out and sighting, and development of global databases that can be used as a tool for decision makers (Davis *et al.*, 2010; Koh *et al.*, 2009; Richardson & Blanchard, 2011). Such databases could be one one-stop location to store downloadable up-to-date information on the native range, places of introduction, naturalisation and invasion, a suite of life-history trait data, and completed weed risk assessments with scores.

Lessons learned from commercial forestry invasions could provide a valuable guide for future biofuel crops. However, as I have shown in Chapter 2 the effectiveness of the WRA relies heavily on knowledge of a species behaviour elsewhere. Therefore the WRA may be of little use when screening potential biofuel species that have never been cultivated outside their native range. And as demonstrated in Chapter 3 and Chapter 5, species life-history traits traditionally identified as being important in determining invasions do not appear to predict invasions in some commercial species, and the effect of other factors such as propagule pressure can differ somewhat between regions. Caution and high-quality species-level data are needed before extrapolating any general lessons from commercial forestry to other species and agricultural systems.

6.5 Recommendations for future research

Although the importance of quantifying failures at the introdution stage when examining the invasion process has been stated several years ago (Blackburn & Duncan, 2001; Cassey *et al.*, 2004) and insights gained in plant invasion biology when this has been adopted (Dawson *et al.*, 2009a; Diez *et al.*, 2009; Gravuer *et al.*, 2008), the majority of invasion studies still do not account for the failures at each stage as well as successes or quantify the introduction stage (e.g. Essl *et al.*, 2011, 2010; Procheş *et al.*, 2012). I recommend that more attention is paid to this issue. One reason the introduction stage is only rarely included (Puth & Post, 2005) is that data on introductions is often perceived by researchers as difficult and time-consuming to obtain, requiring searches of archive material which may be available only in specialist holdings. However, obtaining such data ought to be attempted and in many cases may not be any more time consuming than searching for species life-history traits in databases, research papers or through field collection and experimentation. If the data are not available then it may be worthwhile questioning if the insights that the given system can provide are in fact likely to be novel or robust to the issue of the "opportunity to invade". Universally adopting the framework of invasions as a stage-based framework (Blackburn *et al.*, 2011) can help researchers produce insights that are based on robust definitions and are comparable to other studies.

Future work quantifying invasion risk should focus on obtaining detailed datasets, such as the one used in this thesis, for several geographic areas and a wide taxonomic range. I found that there were common patterns between NZ and GB but it remains to be seen how generally applicable these patterns are to other region, particularly regions in the tropics, regions with larger land areas and regions where the forestry sector is not based on the Anglo-European model that applied to NZ and GB. Establishing whether patterns documented in temperate regions translate to tropical regions (e.g. Dawson *et al.*, 2009a) and what lessons can be learnt is critical to preventing invasions in tropical regions.

Further research is needed to determine the shape of the relationship between planting intensity and invasion risk. The potential shapes of the relationship between propagule pressure and invasions has been outlined by Lockwood et al. (2005), yet testing these hypotheses has proven difficult, even in my study where detailed data were available for forestry planting. Detailed information on the amount and timing of planting is difficult to obtain for many taxa because of the historical nature of most introductions and will depend on the use of the species (e.g. commercial crops are often inventoried, and some horticultural species hold special interest and are thus documented). When such detailed information has been available, the results from different regions for the same species can be conflicting. For example, Chapter 3 and Chapter 5 suggest that there is a direct relationship between naturalisation at a regional-scale and propagule pressure for pines in NZ and GB. However, Nuñez *et al.* (2011) found no relationship between propagule pressure and pine naturalisations at a landscape-scale in Isla Victoria, Argentina. These differences may be due to the scale of propagule pressure in each study region: commercial-scale and temporally sustained pressure for NZ and GB versus more limited experimental plantings in Isla Victoria. This suggests a critical threshold level of planting after which it is inevitable that a species will escape cultivation and invade, given that the climate is suitable for that species. Establishing the nature of the relationship between propagule pressure and invasions is thus important for quantifying risk and could facilitate the development of plantation management strategies based on total propagule pressure. Quantification of propagule pressure could be improved by searching for case-study systems like the one presented in this thesis, where detailed historical documentation exists.

One issue that is rarely addressed in retrospective tests of the WRA and in factors determinng risk is that trees can experience time-lags between introduction, naturalisation and invasion in the order of 100 years (Křivánek & Pyšek, 2008; Richardson *et al.*, 1994). This could be a critical limitation to many such studies. In my study, the long time period since introduction of the majority of the genus *Pinus* meant that it is likely that species have had sufficient time to naturalise. Additionally, i included resident time as a covariate in models and thus controlled for this factor. However, future studies would benefit from explicitly confronting this issue and investigating how it could bias results and control ling for it statistically by including introduction date as a null model approach.

When plants are introduced to new regions they may lose the interractions with mutualist, natural enemies and competitors present in their native ranges, but also gain interractions with new species encountered in the introduced range (this topic has been reviewed in Mitchell *et al.*, 2006; Richardson *et al.*, 2000a). The effect of this change in the dynamics of biotic interractions in the new range can be significant. This can lead to advantage for the introduced plant species, such as enemy release, and dissadvantages such as inhibition from lack of mutualistic interractons. Integrating data on the factors dermining invasions including life-history traits, biogeographic and human factors identified in this thesis with data on the presence and effect of biotic interractions could unify the two areas of invasion biology. Taking a case-study approach and focusing on a small taxonomic group such as a famiy or group of genera could be a manageable way to begin addressing the integration of these different factors.

Finally, historical documents such as forestry inventories, plans and nursery catalogues are valuable sources of data on propagule pressure, yet few are digitised. A public database of such documents similar to `www.gbif.org`, along with a wiki that allowed researches to input introduction dates (backed with citations to original sources) could greatly facilitate future invasion studies across several regions. As many such historical documents are held in institutions that often have no budgets for digitisation of such material, this task falls to individual researchers and the invasion biology community.

Appendix A

List of *Pinus* species, their introduction status in New Zealand and Great Britain, WRA results, and Z-scores

Table A.1. Table listing 115 species in the genus *Pinus* used in this study along with: (A) invasion status in NZ (NI = not introduced, I = introduced, N = naturalised); (B) number of WRA questions answered for NZ assessments; (C) NZ WRA scores; (D) NZ WRA result; (E) NZ introduction date; (F) invasion status in GB (NI = not introduced, I = introduced, N = naturalised); (G) number of WRA questions answered for GB assessments; (H) GB WRA scores; (I) GB WRA result; (J) GB introduction date; (K) Z-score. Species epithets are from the International Plant Names Index (IPNI) http://www.ipni.org, though there is some disagreement between the data sources used by the IPNI. Introduction dates for species marked with † came from the forest working plans held in the archives at Scion, Rotorua, New Zealand; and those with a * came from nursery catalogues held at the library at Lincoln University, Lincoln, Canterbury, New Zealand. Numbers in brackets indicate the literature source for the data.

Species	A	B	C	D	E	F	G	H	I	J	K
albicaulis Engelm.	NI	40	4	Accept	-	I [15,16]	40	3	Accept	1853 [26]	−5.61
aristata Engelm.	I [1]	41	−1	Accept	1915 [1]	I [15,16]	41	−3	Accept	-	0.35
arizonica Engelm.	NI	36	−2	Accept	-	NI	37	−2	Accept	-	4.03
armandii Franch.	I [1]	35	−4	Accept	1909 [1]	I [15,16]	36	−5	Accept	1897 [24]	−3.74
attenuata Lemmon.	I [1,8]	36	0	Accept	1860 [8]	I [15,16]	36	3	Accept	-	11.03
ayacahuite Ehrenb. ex Schltdl.	I	35	1	Evaluate	1974†	I [14–16]	35	0	Accept	1872 [24]	7.04
balfouriana S.Watson	NI	35	−4	Accept	-	I [15]	36	−5	Accept	1853 [26]	−1.09
banksiana Lamb.	N [9]	40	16	Reject	1904 [1]	I [14–16]	41	15	Reject	1897 [24]	9.17
bhutanica D.G.Long & C.N.Page	NI	20	3	Evaluate	-	I [20]	21	2	Accept	1930 [22]	0.38
brutia Ten.	I [1]	39	6	Reject	1915 [1]	I [16]	39	6	Reject	-	7.49
bungeana Zucc. ex Endl	I [1]	34	−1	Accept	1903 [1]	I [15,16]	35	−1	Accept	-	−1.38
canariensis C.Sim	I [1,3,6]	37	12	Reject	1861*	I [15,16]	38	11	Reject	-	−12.25
caribaea Morelet	I [1]	34	8	Reject	-	I [15,16]	35	7	Reject	-	0.46
cembra L.	I [1]	35	0	Accept	1872*	I [14–16]	36	-1	Accept	1745 [18]	4.9 0
cembroides Zucc.	I [1]	39	−1	Accept	1872*	I [15,16]	40	−2	Accept	-	−7.79

Continued on next page...

Continued

Species	A	B	C	D	E	F	G	H	I	J	K
chiapensis (Martínez) Andresen	NI	34	−4	Accept	-	NI	35	−5	Accept	-	−2.72
clausa Vasey	NI	35	3	Accept	-	NI	36	2	Accept	-	8.72
contorta Douglas ex Loudon	N[9]	42	15	Reject	1873*	N[9]	43	14	Reject	1851[25]	10.58
cooperi C.E.Blanco	NI	23	−3	Accept	-	NI	24	−3	Accept	1832[23]	7.03
coulteri D.Don	I[1]	38	6	Reject	1870*	I[15,16]	39	10	Reject	-	−3.67
culminicola Andresen & Beaman	I[1]	34	−4	Accept	-	NI	35	−5	Accept	-	−2.96
dalatensis Ferre	NI	26	−4	Accept	-	NI	27	−5	Accept	-	8.05
densata Masters	NI	38	2	Accept	-	NI	39	2	Accept	1926[24]	1.38
densiflora Siebold & Zucc.	I[1,7]	37	14	Reject	1872*	I[14–16]	38	13	Reject	-	7.54
devoniana LIdl.	I	28	1	Evaluate	1872*	I[16]	29	0	Evaluate	-	5.97
discolor D.K.Bradley & Hawksw.	NI	27	−3	Accept	-	NI	28	−3	Accept	-	−2.09
douglasiana Martínez	NI	27	−2	Accept	-	NI	28	−2	Accept	-	6.11
durangensis Martínez	I	31	2	Evaluate	1960†	NI	32	1	Evaluate	1821[24]	−6.63
echinata Mill.	I[1,7]	36	5	Accept	1881[1]	I[15–17]	37	4	Accept	-	4.87
edulis Engelm.	NI	41	1	Evaluate	-	I[16]	42	0	Evaluate	-	−2.71
elliottii Engelm.	I[10]	39	7	Reject	1915[1]	I (16)	40	6	Reject	-	4.79
engelmannii Carrière	NI	35	−1	Accept	-	NI	36	−1	Accept	-	−2.71
fenzeliana Hand.-Mazz.	NI	18	−1	Accept	-	NI	19	−2	Accept	1861[24]	−0.55
flexilis E. James	I[1]	35	0	Accept	1872*	I[15,16]	36	−1	Accept	-	−2.27
gerardiana Wall. ex Lamb.	I[1]	34	−1	Accept	1915[1]	I[15,16]	35	−1	Accept	-	−11.03

Continued on next page…

Continued

Species	A	B	C	D	E	F	G	H	I	J	K
glabra Walter	NI	36	−6	Accept	-	I [15]	37	−7	Accept	-	7.07
greggii Engelm. ex Parl.	I [2,4]	32	0	Evaluate	1960 [6]	I [15,16]	33	−1	Accept	-	4.45
hakkodensis Makino	NI	20	3	Evaluate	-	NI	21	2	Evaluate	-	11.37
halepensis Mill.	N [9]	38	11	Reject	1868 [6]	I [15,16]	39	17	Reject	-	11.76
hartwegii Lindl.	I [1,6]	29	5	Reject	1868 [6]	I [16]	30	4	Evaluate	1855 [24]	4.92
heldreichii Christ	I [1]	28	4	Evaluate	1904 [1]	I [15,16]	29	3	Evaluate	1890 [18]	12.34
henryi Masters	NI	29	6	Reject	-	NI	30	5	Accept	-	7.56
herrerae Martínez	NI	25	-1	Accept	-	NI	26	-1	Accept	-	8.72
hwangshanensis W.Y.Hsia	NI	18	3	Evaluate	-	NI	19	2	Evaluate	-	8.64
jaliscana Pérez de la Rosa	NI	28	−2	Accept	-	NI	29	−3	Accept	-	11.29
jeffreyi A.Murr.	I [6]	37	9	Reject	1861*	I [16]	37	14	Reject	1853 [27]	3.39
johannis M.-F.Robert	NI	28	−6	Accept	-	NI	29	−6	Accept	-	−7.90
juarezensis Lanner	NI	24	1	Evaluate	-	NI	25	0	Evaluate	-	−23.45
kesiya Royal ex Gordon	I [6]	30	7	Reject	1865 [6]	I [15,16]	32	12	Reject	-	9.45
koraiensis Siebold & Zucc.	I [1]	36	4	Evaluate	1906 [1]	I [15,16]	37	3	Reject	1880 [24]	−6.47
krempfii Lecomte	NI	28	−2	Evaluate	-	NI	29	−3	Accept	-	10.33
kwangtungensis Chun ex Tsiang	NI	27	4	Accept	-	NI	28	3	Accept	-	−0.64
lagunae (Passini) Passini	NI	30	−3	Accept	-	NI	31	−4	Accept	-	−1.37
lambertiana Douglas ex R.Taylor & R.Philips	I [1,7]	36	10	Reject	1872*	I [15,16]	37	13	Reject	1827 [23]	−28.63
lawsonii Rozel ex Gordon	NI	28	−4	Accept	-	NI	29	−4	Accept	-	6.92

Continued on next page. . .

Continued

Species	A	B	C	D	E	F	G	H	I	J	K
leiophylla Schltdl. & Cham.	NI	33	−1	Accept	-	I [15,16]	34	−1	Accept	-	7.43
longaeva D.K.Bailey	NI	34	−1	Accept	-	NI	35	−2	Accept	-	6.31
luchuensis Mayr	NI	26	5	Evaluate	-	I [15,16]	27	4	Evaluate	-	7.12
lumholtzii B.L.Rob. & Fernald	NI	26	−1	Accept	-	NI	27	−2	Accept	-	4.91
massoniana Lamb.	I [1]	28	0	Accept	1873*	I [15,16]	29	1	Evaluate	-	6.64
maximartinezii Rzed.	NI	32	−5	Accept	-	NI	33	−5	Accept	-	−19.57
maximinoi H.E.Moore	I	29	2	Evaluate	1949†	NI	30	1	Evaluate	-	7.94
merkusii Jungh. & de Vriese	NI	32	2	Accept	-	NI	32	2	Accept	-	1.36
monophylla Torr. & Frém.	I [1]	38	−2	Accept	1876*	I [15,16]	39	−2	Accept	-	−11.73
montezumae D.Don I Lamb.	I [1]	34	4	Evaluate	1911 [1]	I [15,16]	34	6	Reject	1881 [24]	−0.67
monticola Douglas ex Lamb.	I [1]	35	3	Evaluate	1915 [1]	I [15,16]	35	7	Reject	1823 [23]	3.32
morrisonicola Hayata	NI	24	0	Evaluate	-	NI	25	-1	Accept	-	4.73
mugo Turra	N [9]	36	16	Reject	1861*	N [9,17]	36	16	Reject	1774 [25]	11.63
muricata D. Don	N [9,21]	35	7	Reject	1861**	I [15]	35	7	Reject	1851 [24]	8.65
nelsonii Shaw	NI	28	−2	Accept	-	I [15]	29	−3	Accept	-	−3.85
nigra J.F.Arnold	N [9]	35	14	Reject	1861*	N [21]	35	14	Reject	1759 [18]	4.24
occidentalis Sw.	NI	28	−1	Accept	-	NI	28	−3	Accept	-	−0.16
oocarpa Schiede ex Schltdl.	NI	34	3	Evaluate	-	I [16]	34	3	Evaluate	-	6.18
orizabensis D.K.Bailey & Hawksw.	NI	26	−3	Accept	-	NI	27	−3	Accept	-	−15.11
palustris Mill.	I [1,3,7]	35	6	Reject	1866 [3]	I [15,16]	36	9	Reject	-	−5.08

Continued on next page...

Continued

Species	A	B	C	D	E	F	G	H	I	J	K
parviflora Siebold & Zucc.	NI	28	1	Accept	-	I [15,16]	29	1	Accept	1879 [24]	0.81
patula Schiede ex Schltdl. & Cham.	N [9]	33	14	Reject	1877 [1]	I [15,16]	33	14	Reject	-	10.06
peuce Griseb.	I [1]	31	4	Accept	1876*	I [14−−16]	31	6	Reject	1864 [25]	2.75
pinaster Aiton	N [9]	38	18	Reject	1830 [27]	N [9]	38	18	Reject	1596 [18]	7.89
pineana Gordon	NI	29	−4	Accept	-	NI	30	−4	Accept	1870 [24]	−5.55
pinea L.	I [1,3,7]	33	8	Reject	1861*	I [15,16]	33	15	Reject	1500 [18]	−8.37
ponderosa P.Lawson &. C.Lawson	N [9]	37	19	Reject	1861*	I [15,16]	37	19	Reject	1827 [23]	4.10
pringlei Shaw I Sarg.	I [7,8]	26	−3	Accept	1896 [1]	NI	27	−3	Accept	-	12.45
pseudostrobus Lindl.	I	32	1	Evaluate	1949†	I [15,16]	32	3	Evaluate	1913 [24]	6.05
pumila Regel	NI	30	2	Accept	-	I [15,16]	31	2	Accept	-	2.23
pungens Lamb.	NI	34	0	Accept	-	I [15,16]	35	−1	Accept	-	9.64
quadrifolia Parry ex Parl.	I [7,8]	34	−2	Accept	1896 [7]	I [15]	35	−2	Accept	-	−2.10
radiata	N [9]	41	14	Reject	1858*	N [9,14,16]	41	14	Reject	1832 [25]	8.75
remota D.K.Bailey & Hawksw.	NI	29	−4	Accept	-	NI	30	−4	Accept	-	0.98
resinosa Aiton	I [1]	37	11	Reject	1861*	I [15,16]	38	10	Reject	-	4.72
rigida Mill.	I [6,7]	34	12	Reject	1868 [6]	I [14−16]	35	11	Reject	1895 [24]	8.12
roxburghii Sarg.	I	36	6	Reject	1876*	I [15,16]	36	11	Reject	-	−4.68
rzedowskii Madrigal & M.Caball.	NI	25	−3	Accept	-	NI	26	−4	Accept	-	−1.12
sabiniana Douglas	I [1,7]	35	9	Reject	1861*	I [15,16]	36	8	Reject	1832 [25]	−7.70
serotina Michx.	I	33	−1	Accept	1862*	I [15]	33	−1	Accept	-	10.82

Continued on next page…

Continued

Species	A	B	C	D	E	F	G	H	I	J	K
sibirica Ledeb. Turcz.	NI	34	1	Evaluate	-	I[19]	35	0	Evaluate	-	−3.02
squamata X.W.Li	NI	23	0	Evaluate	-	NI	24	-1	Accept	-	4.65
stankewiczii Sukaczev Fomi	NI	21	−1	Accept	-	NI	22	−2	Accept	-	−13.77
strobiformis Engelm.	NI	35	5	Evaluate	-	NI	36	4	Accept	-	−3.72
strobus L.	N[9]	37	13	Reject	1861*	I[14–16]	37	13	Reject	1605[25]	7.01
sylvestris L.	N[9]	35	10	Reject	1861*	Native	35	10	Reject	1746[24]	7.75
tabuliformis Carrière	I	29	0	Evaluate	1957†	I[15,16]	30	-1	Accept	1910[24]	9.20
taeda L.	N[9,10]	36	17	Reject	1915[1]	I[15,16]	36	17	Reject	1912[24]	6.54
taiwanensis Hayata	I	34	1	Evaluate	1963†	I[16]	34	1	Evaluate	-	2.55
tecunumanii F.Schwerdtf. ex Eguiluz & J.P.Perry	NI	32	−1	Accept	-	NI	33	−2	Accept	-	9.29
teocote Cham. & Schltdl.	I[1]	31	2	Evaluate	1915[1]	I[15]	32	1	Evaluate	-	3.52
thunbergii Parl.	I[1]	33	6	Reject	1890[1]	I[14–16]	33	6	Reject	1880[24]	8.98
torreyana Parry ex Carrière	I[1,7]	33	−2	Accept	1870[1]	I[15,16]	33	−2	Accept	-	−6.90
tropicalis Morelet	NI	31	−2	Accept	-	NI	32	−3	Accept	-	10.11
uncinata Ramond ex DC.	I[10]	33	0	Accept	1962[5]	I[17]	34	0	Accept	-	9.75
virginiana Mill.	I[1]	36	9	Reject	1900[1]	I[15,16]	37	8	Reject	-	10.1
wallichiana A.B.Jacks.	I[1,6]	36	10	Reject	1868[6]	N[17]	36	10	Reject	1823[25]	0.98
wangii Hu & W.C.Cheng	NI	25	−2	Accept	-	NI	26	−3	Accept	-	−2.73
washoensis Mason & Stockw.	NI	28	1	Accept	-	I[16]	29	0	Accept	-	−0.58
yunnanensis Franch.	I[1]	30	2	Accept	1909[1]	I[16]	31	1	Accept	-	7.09

References

1. Adams, T.W. (1915) The species of the genus *Pinus* now growing in New Zealand, with some notes on their introduction and growth. *Transactions and Proceedings of the Royal Society of New Zealand*, **48**, 216-233.

2. Dvorak, W.S., Kietzka, J.E. & Donahue, J.K. (1996) Three-year survival and growth of provenances of *Pinus greggii* in the tropics and subtropics. *Forest Ecology and Management*, **83**, 123-131.

3. Gillies, T.B. (1897) Notes on the Growth of certain Trees on Scoria Soil near Mount Eden, Auckland. *Transactions and Proceedings of the Royal Society of New Zealand*, **12**, 357-358.

4. Kietzka, J. E., Denison, N. P. & Dvorak,W. S. (1996) *Pinus greggii*, a promising new species, pp. 42–45. In: Dieters, M. J., Matheson, A. C., Nikles, D. G., Harwood, C. E. & Walker, S. M. (Eds), *Proc. IUFRO: Breeding Tropical Trees*. Queensland, Australia.

5. Ledgard, N.J. & Ecroyd, C.E. (1988) Introduced forest trees in New Zealand: recognition, role and seed source. 4. *Pinus mugo* Turra - dwarf mountain pine; *Pinus uncinata* Mirbel - mountain pine. *Forest Research Institute Bulletin No. 124,* P.O. Box 3020, Rotorua, New Zealand. 10 pp.

6. Ludlam, A. (1868) Essay on the cultivation and acclimatization of trees and plants. *Transactions and Proceedings of the Royal Society of New Zealand*, **1**, 1-23.

7. Mason, T. (1896) An account of the plants growing at “The Gums” Taita. *Transactions and Proceedings of the Royal Society of New Zealand*, **29**, 393-421.

8. Miller, J.T. & Knowles, F.B. (1988) Introduced forest trees in New Zealand: recognition, role and seed source: Part 5 *Pinus attenuata* Lemmon - knobcone pine - *Forest Research Institute Bulletin 124*. P.O. Box 3020, Rotorua, New Zealand.

9. Richardson, D.M. & Rejmákek, M. (2004) Conifers as invasive aliens: a global survey and predictive framework. *Diversity and Distributions*, **10**, 321-331.

10. Webb, C.J., Sykes, W.R. & Garnock-Jones, P.J.; with contributions by D.R. Given, P.J. Brownsey, B.H. Mackmillan & L.B. Moore (1988). *Flora of New Zealand. Volume IV. - Naturalised Pteridophytes, Gymnosperms, Dicotyledons*. Botany Division, D.S.I.R., Christchurch, New Zealand.

11. Landcare Research (2010) Ecological Traits of New Zealand Flora Database Lancare Research - listed as having a naturalisation period of 1940-1970. Available at: `http://www.landcareresearch.co.nz/databases/`

12. Randall, R. P. (2002) A Global Compendium of Weeds. 905 pp. Available at: `http://www.hear.org/gcw`

13. CABI (2005) Forestry Compendium. CAB International, Wallingford, UK. Available at: http://www.cabi.org/fc

14. Crook, S. (1997) A (very) provisional list of conifers in the British Isles. *BSBI News*. April 1975. No 75. Edited by Gwynn Ellis, 41 Marlborough Road, Roath, Cardiff, CF2 5BU. Available at: http://www.watsonia.org.uk/html/bsbi_news_2.html

15. Chittenden, F.J. (1932) *Conifers in cultivation: The report of the conifer conference held by the Royal Horticultural Society*. Nov. 10-12, 1931.

16. Hillier and Sons (1981) *Hillier's Manual of Trees and Shrubs*. Fifth Edition. David & Charles (Publishrs) Limited, Newton Abbot, Devon. 575 pp.

17. Clapham, A.R., Tutin, T.G. & Moore, D.M. (1987) *Flora of the British Isles*. Third edition. Cambridge University Press, Cambridge. 688 pp.

18. Mitchell, A.F. (1972) *Conifers from Europe and Asia*. p 9-10 In: Napeir, E. (ed) *Conifers in the British Isles*. The Royal Horticultural Society. Proceeding of the third conifer conference arranged by The Royal Horticultural Society held in London, 5 to 8 October 1970. 304 pp.

19. Hunt, D.R. (1972) *Reference list of conifers and conifer allies growing out of doors in the British Isles*. p 109-122 In: Napeir, E. (ed) *Conifers in the British Isles*. The Royal Horticultural Society. Proceeding of the third conifer conference arranged by The Royal Horticultural Society held in London, 5 to 8 October 1970. 304 pp.

20. Flanagan, M. (2003) *History of conifers in horticulture: An overview*. In Proc 4th IS on Conifer Ed: R.R. Mill Acta Hort 615, ISHS 2004. Available at: http://www.actahort.org/members/showpdf?session=27820

21. Richardson, D.M. (2006) *Pinus*: a model group for unlocking the secrets of alien plant invasions? *Preslia*, **78**, 375-388.

22. Flanagan M. (2003) *History of conifers in horticulture: An overview*. Proc 4th IS on Conifer Ed: R.R. Mill Acta Hort 615, ISHS 2003. Available at: http://www.actahort.org

23. House, S. (2003) *The Man Who Made Pines: David Douglas and his Conifer Introductions from the Pacific North West*. Proc 4th IS on Conifer Ed: R.R. Mill Acta Hort 615, ISHS 2003. Available at: http://www.actahort.org

24. Mitchell, A.F. (1972) *Conifer statistics*. In Napeir, E. (ed) *Conifers in the British Isles*. The Royal Horticultural Society. Proceeding of the third conifer conference arranged by The Royal Horticultural Society held in London, 5 to 8 October 1970.

25. Hill, M., Baker, R., Broad, G., Chandler, P.J., Copp, G.H., Ellis, J., Jones, D., Hoyland, C., Laing, I., Longshaw, M., Moore, N., Parrott, D., Pearman, D., Preston1, C., Smith, R.M. & Waters, R. (2005) *Audit of non-native species in England*. English

Nature Research Reports. Number 662. Available at: http://jncc.defra.gov.uk/page-4011 and data behind the report available at: http://www.brc.ac.uk/resources.htm

26. Woods, P. & Woods, J. (2003) *The Oregon Expedition 1850-1854: John Jeffrey and His Conifers*. Proc 4th IS on Conifer Ed: R.R. Mill Acta Hort 615, ISHS 2003. Available at: http://www.actahort.org

27. Knowles F.B. & J.T. Miller J.T. (1989) *Introduced forest trees in New Zealand: recognition, role, and seed source. 8,* Pinus pinaster *Aiton, maritime pine.* Ministry of Forestry, Forest Research Institute, Rotorua, N.Z.

28. Burdon, R.D. & Low, C.B. (1991) *Performance of* Pinus ponderosa *and* Pinus jeffreyi*provenances in New Zealand*. Canadian Journal of Forest Research. 21: 1401–141

Appendix B

List of New Zealand regions

Table B.1. List of 43 regions used in regression analysis for New Zealand, organised into their wider Department of Conservation conservancy (`http://www.doc.govt.nz/by-region/`). A list of Watsonian vice-counties used in this study can be found online at `http://herbariaunited.org/gridrefVC/`

Conservancy	Areas included
Northland	Bay of Islands, Kaitai, Kauri coast, Whangarei
Auckland	Auckland, Warkworth
Waikato	Hauraki, Maniapoto, Waikato
East Coast	Gisborne, Rotorua, Te Urewera, Tauranga
Tongariro	Ruapehu, Taurangi
Wanganui	Palmerston North, Taranaki
Wellington	Whanganui, Hawkes Bay, Kapiti, Wairarapa, Poneke
Nelson	Golden Bay, Motueka, Nelson Lakes, Sounds Area, South Marlborough
West Coast	Buller, Franz Josef, Greymouth, Hokitika, South Westland
Canterbury	Aoraki, Mahaanui, Raukapuka, Twizel, Waimakariri
Otago	Central Otago, Coastal Otago, Wakatipu, Wanaka
Southland	Murihiku, Te Anau

Appendix C

WRA questions and summary of answers

Table C.1. A list of weed risk assessment questions, the default answer used in this study where applicable, and the percentage of assessments that each questions was answered for New Zealand (NZ) and Great Britain (GB).

WRA question		Default	NZ	GB
Domestication/cultivation				
1.01	Is the species highly domesticated	No	100	100
1.02	Has the species become naturalised where grown	NA	100	100
1.03	Does the species have weedy races	NA	100	100
Climate and distribution				
2.01	Species suited to New Zealand/Great Britain climates (0–low; 1–intermediate; 2–high)	-	100	100
2.02	Quality of climate match data (0–low; 1–intermediate; 2–high)	2	100	100
2.03	Broad climate suitability	-	87	87
2.04	Native or naturalised in regions with equable climates	-	94	94
2.05	Does the species have a history of repeated introductions outside its native range	-	97	97
Weed elsewhere (interacts with 2.01 to give a weighted score)				
3.01	Naturalised beyond native range	-	88	88
3.02	Garden/amenity/disturbance weed	-	96	96
3.03	Agricultural/forestry/horticultural weed	-	96	96
3.04	Environmental weed	-	98	98
3.05	Congeneric weed	Yes	100	100

Continued on next page...

Continued

WRA question		Default	NZ	GB
Undesirable traits				
4.01	Produces spines, thorns or burrs	No	100	100
4.02	Allelopathic	-	17	17
4.03	Parasitic	No	100	100
4.04	Unpalatable to grazing animals	-	38	38
4.05	Toxic to animals	-	42	42
4.06	Host for recognised pests and pathogens	-	82	82
4.07	Causes allergies or otherwise toxic to humans	No	100	100
4.08	Created a fire hazard in natural ecosystems	-	36	36
4.09	Is a shade tolerant plant at some stage of its life cycle	-	65	65
4.10	Tolerates a wide range of soil conditions	-	77	77
4.11	Climbing or smothering growth habit	No	100	100
4.12	Forms dense thickets	-	72	72
Plant type				
5.01	Aquatic	No	100	100
5.02	Grass	No	100	100
5.03	Nitrogen fixing woody plant	No	100	100
5.04	Geophyte	No	100	100
Reproduction				
6.01	Evidence of substantial reproductive failure in native habitat	-	97	97
6.02	Produces viable seed	-	97	97
6.03	Hybridises naturally	-	65	65
6.04	Self-compatible or apomictic	-	29	29
6.05	Requires specialist pollinators	No	100	100
6.06	Reproduction by vegetative fragmentation	-	96	96
6.07	Minimum generative time (years)	-	61	61
Dispersal mechanisms				
7.01	Propagules likely to be dispersed unintentionally (plants growing in heavily trafficked areas)	-	95	95

Continued on next page. . .

Continued

WRA question		Default	NZ	GB
7.02	Propagules dispersed intentionally by people	-	86	86
7.03	Propagules likely to be dispersed as a produce contaminant	-	97	97
7.04	Propagules adapted to wind dispersal	-	96	96
7.05	Propagules water dispersed	-	97	97
7.06	Propagules bird dispersed	-	41	41
7.07	Propagules dispersed by other animals (externally)	-	15	15
7.08	Propagules survive passage through the gut	Unknown	0	0
Persistence attributes				
8.01	Prolific seed production (>1000/m^2)	-	20	20
8.02	Evidence that a persistent propagule bank is formed (>1 yr)	-	26	26
8.03	Well controlled by herbicides	-	6	6
8.04	Tolerates, or benefits from mutilation, cultivation, or fire	-	71	71
8.05	Effective natural enemies present in New Zealand or Great Britain	-	25	100

Appendix D

WRA results without information of prior invasiveness

Table D.1. Summary of WRA results for all *Pinus* species classed by whether they were not introduced, introduced and naturalised in New Zealand and Great Britain. These analyses differ from Table 2.1 in Chapter 2, in that they do not include information on prior invasion history. Values include: WRA score, number of questions answered, number of species recorded as naturalising elsewhere, WRA classification after the secondary screening procedure (accept, evaluate or reject), mean introduction date, number of species at each stage that are forestry species, and mean climate match score for the stage. Differing upper case letters indicate significantly different mean values determined by a Wilcoxon Signed-Rank tests where $P < 0.05$

		Not introduced	Introduced	Naturalised
New Zealand	WRA score (±SE)	$-0.63\ (0.34)^{A}$	$0.83\ (0.28)^{B}$	$2.08\ (0.49)^{B}$
	Number questions answered (±SE)	$22.02\ (0.78)^{A}$	$27.45\ (0.43)^{B}$	$30.15\ (0.72)^{C}$
	Accept (% ±SE)	32 (65.34)	30 (45.56)	4 (30.82)
	Evaluate (% ±SE)	17 (34.75)	35 (53.04)	9 (69.23)
	Reject (% ±SE)	0 (0.00)	1 (1.52)	0 (0.00)
	Total number of species	49	66	13
Great Britain	WRA score (±SE)	$-1.78\ (0.34)^{A}$	$0.06\ (0.29)^{B}$	$1.29\ (0.64)^{B}$
	Number questions answered (±SE)	$22.15\ (0.75)^{A}$	$27.96\ (0.45)^{B}$	$31.00\ (1.11)^{B}$
	Accept (% ±SE)	38 (92.7)	44 (60.27)	4 (57.14)
	Evaluate (% ±SE)	3 (7.3)	28 (38.36)	2 (28.57)
	Reject (% ±SE)	0 (0.00)	1 (1.37)	1 (14.29)
	Total number of species	41	73	7

Appendix E

Correlations between all variables

Table E.1. Correlation matrix of all continuous variables collected for the genus *Pinus* in New Zealand and Great Britain. Correlation coefficients were calculated using Spearman's correlation coefficient.

	DBH	Interval large crop	Juvenile period	Life-span	Seed mass	Seed wing length	Cone length	Genome size	Height	Seed length	Z-score	Native range size	Climate match NZ	Climate match GB	Global forestry use	Area planted NZ	Area planted GB	First date NZ	First date GB	Range planting NZ	Range planting GB	Residence time NZ	Residence time GB
DBH		-0.10	-0.05	0.27	-0.05	0.30	0.33	-0.12	0.55	0.02	0.11	-0.02	0.17	0.14	0.26	0.13	0.03	-0.11	-0.47	0.19	0.32	-0.06	-0.04
Interval large crop	-0.10		0.29	0.20	0.28	-0.06	0.25	0.03	-0.03	0.16	-0.56	0.17	-0.07	-0.03	0.05	0.07	-0.05	-0.06	0.18	-0.18	-0.12	0.00	0.03
Juvenile period	-0.05	0.29		0.16	0.55	-0.16	0.17	0.52	-0.09	0.60	-0.84	-0.09	-0.26	-0.17	-0.21	-0.28	-0.17	-0.03	0.24	0.04	-0.41	0.03	0.02
Life-span	0.27	0.20	0.16		0.07	-0.04	0.06	0.16	-0.01	0.00	-0.14	0.04	-0.18	-0.17	-0.11	-0.08	-0.03	0.05	0.06	0.05	-0.18	-0.02	0.10
Seed mass	-0.05	0.28	0.55	0.07		-0.32	0.26	0.70	-0.31	0.85	-0.80	-0.10	-0.12	-0.11	-0.24	-0.34	-0.21	-0.09	-0.06	0.05	-0.33	-0.10	-0.18
Seed wing length	0.30	-0.06	-0.16	-0.04	-0.32		0.51	-0.16	0.47	-0.20	0.29	0.02	0.28	0.31	0.32	0.42	0.25	-0.27	-0.07	0.17	-0.27	-0.26	-0.01
Cone length	0.33	0.25	0.17	0.06	0.26	0.51		0.29	0.31	0.38	-0.24	-0.01	0.13	0.14	0.19	0.16	0.04	-0.07	-0.07	0.12	-0.23	-0.15	-0.07
Genome size	-0.12	0.03	0.52	0.16	0.70	-0.16	0.29		-0.30	0.69	-0.58	-0.05	-0.25	-0.23	-0.28	-0.39	-0.18	0.03	-0.20	-0.10	-0.12	-0.02	0.03
Height	0.55	-0.03	-0.09	-0.01	-0.31	0.47	0.31	-0.30		-0.27	0.20	0.10	0.28	0.28	0.49	0.42	0.18	-0.11	-0.13	0.18	-0.06	-0.05	-0.17
Seed length	0.02	0.16	0.60	0.00	0.85	-0.20	0.38	0.69	-0.27		-0.74	0.00	-0.09	-0.08	-0.15	-0.30	-0.21	0.01	0.04	-0.04	-0.42	-0.12	-0.07
Z-score	0.11	-0.56	-0.84	-0.14	-0.80	0.29	-0.24	-0.58	0.20	-0.74		0.05	0.18	0.13	0.22	0.26	0.20	0.07	-0.23	0.03	0.46	0.01	0.14
Native range size	-0.02	0.17	-0.09	0.04	-0.10	0.02	-0.01	-0.05	0.10	0.00	0.05		0.21	0.26	0.58	0.37	0.28	-0.03	-0.09	0.09	0.12	-0.11	-0.02
Climate match NZ	0.17	-0.07	-0.26	-0.18	-0.12	0.28	0.13	-0.25	0.28	-0.09	0.18	0.21		0.86	0.57	0.61	0.50	-0.32	-0.61	0.44	0.54	-0.31	-0.30
Climate match GB	0.14	-0.03	-0.17	-0.17	-0.11	0.31	0.14	-0.23	0.28	-0.08	0.13	0.26	0.86		0.56	0.59	0.51	-0.36	-0.69	0.51	0.59	-0.42	-0.34
Global forestry use	0.26	0.05	-0.21	-0.11	-0.24	0.32	0.19	-0.28	0.49	-0.15	0.22	0.58	0.57	0.56		0.63	0.43	-0.42	-0.65	0.49	0.73	-0.34	-0.24
Area planted NZ	0.13	0.07	-0.28	-0.08	-0.34	0.42	0.16	-0.39	0.42	-0.30	0.26	0.37	0.61	0.59	0.63		0.45	-0.65	-0.60	0.77	0.65	-0.35	-0.12
Area planted GB	0.03	-0.05	-0.17	-0.03	-0.21	0.25	0.04	-0.18	0.18	-0.21	0.20	0.28	0.50	0.51	0.43	0.45		-0.46	-0.62	0.41	0.68	-0.43	-0.31
First date NZ	-0.11	-0.06	-0.03	0.05	-0.09	-0.27	-0.07	0.03	-0.11	0.01	0.07	-0.03	-0.32	-0.36	-0.42	-0.65	-0.46		-0.03	-0.66	0.04	0.55	0.15
First date GB	-0.47	0.18	0.24	0.06	-0.06	-0.07	-0.07	-0.20	-0.13	0.04	-0.23	-0.09	-0.61	-0.69	-0.65	-0.60	-0.62	-0.03		-0.37	-0.89	0.87	0.79
Range planting NZ	0.19	-0.18	0.04	0.05	0.05	0.17	0.12	-0.10	0.18	-0.04	0.03	0.09	0.44	0.51	0.49	0.77	0.41	-0.66	-0.37		0.41	-0.42	-0.39
Range planting GB	0.32	-0.12	-0.41	-0.18	-0.33	-0.27	-0.23	-0.12	-0.06	-0.42	0.46	0.12	0.54	0.59	0.73	0.65	0.68	0.04	-0.89	0.41		-0.63	-0.57
Residence time NZ	-0.06	0.00	0.03	-0.02	-0.10	-0.26	-0.15	-0.02	-0.05	-0.12	0.01	-0.11	-0.31	-0.42	-0.34	-0.35	-0.43	0.55	0.87	-0.42	-0.63		0.56
Residence time GB	-0.04	0.03	0.02	0.10	-0.18	-0.01	-0.07	0.03	-0.17	-0.07	0.14	-0.02	-0.30	-0.34	-0.24	-0.12	-0.31	0.15	0.79	-0.39	-0.57	0.56	

Appendix F

Full linked regression equations for exploratory path analysis

The process of fitting a path model was divided into four stages: (1) construct a path diagram that includes all of the possible links among variables (Figure 5.1 in main text); (2) translate this path diagram into a series of regression equations; (3) fit the full model to the data using uninformative priors for the regression coefficients that define the links between variables; (4) refit the model, this time including binary indicator variables that specify whether regression coefficients should be included in the model or not, and using the posterior distributions of the regression coefficients from the full model as priors in this analysis. Below is stage 2 in this process, firstly for introductions of *Pinus* to NZ:

$$\text{introduced}_i \sim \text{Bernoulli}(P_i) \tag{F.1}$$

$$\text{logit}(P_i) = \beta_0 + \beta_1\text{Z-score}_i + \beta_2\text{climate match}_i + \beta_3\text{native range}_i + \beta_4\text{forestry use index}_i \tag{F.2}$$

$$\text{climate match}_i \sim \text{Normal}(\mu_{ci}, \sigma_c^2) \tag{F.3}$$

$$\mu_{ci} = \beta_5 + \beta_6\text{native range}_i \tag{F.4}$$

$$\text{native range}_i \sim \text{Normal}(\mu_{ni}, \sigma_n^2) \tag{F.5}$$

$$\mu_{ni} = \beta_7 + \beta_8\text{Z-score}_i \tag{F.6}$$

$$\text{forestry use index}_i \sim \text{Normal}(\mu_{fi}, \sigma_f^2) \tag{F.7}$$

$$\mu_{fi} = \beta_9 + \beta_{10}\text{native range}_i + \beta_{11}\text{Z-score}_i \tag{F.8}$$

Where introduced_i specifies whether the ith species was introduced to NZ (1) or not (0), which is modelled as drawn from a Bernoulli distribution with probability P_i of naturalisation, which is in turn modelled (on the logit scale) as a function of Z-score, climate match, native range size and the forestry use index. The explanatory variables climate match, native range size and the forestry use index are dependent on other variables (i.e. the ones have arrows going towards them), and were modelled as drawn from separate normal distributions; with climate match modelled as a linear function of native range size, native range size modelled as a linear function of Z-score, and the forestry use index was modelled as a linear function of native range size and the Z-score. The β_0–β_{11} are regression coefficients to be estimated from the data, and σ_c^2, σ_n^2, and σ_f^2 are unknown variances (the variation in the data due to unobserved causes) that are also estimated from the data. While Z-score and native range size causally influence other variables in the model we have not explicitly considered their causes.

The probability of naturalisation for *Pinus* was modelled thus:

$$\text{naturalised}_i \sim \text{Bernoulli}(P_i) \tag{F.9}$$

$$\begin{aligned}\text{logit}(P_i) = \beta_0 + \beta_1\text{Z-score}_i + \beta_2\text{climate match}_i + \beta_3\text{native range}_i \\ + \beta_4\text{forestry use index}_i + \beta_5\text{residence time}_i + \beta_6\text{area planted}_i\end{aligned} \tag{F.10}$$

$$\text{climate match}_i \sim \text{Normal}(\mu_{ci}, \sigma_c^2) \tag{F.11}$$

$$\mu_{ci} = \beta_7 + \beta_8\text{native range}_i \tag{F.12}$$

$$\text{native range}_i \sim \text{Normal}(\mu_{ni}, \sigma_n^2) \tag{F.13}$$

$$\mu_{ni} = \beta_9 + \beta_{10}\text{Z-score}_i \tag{F.14}$$

$$\text{forestry use index}_i \sim \text{Normal}(\mu_{fi}, \sigma_f^2) \tag{F.15}$$

$$\mu_{fi} = \beta_{11} + \beta_{12}\text{native range}_i + \beta_{13}\text{Z-score}_i \tag{F.16}$$

$$\text{residence time}_i \sim \text{Normal}(\mu_{ri}, \sigma_r^2) \tag{F.17}$$

$$\mu_{ri} = \beta_{14} + \beta_{15}\text{forestry use index}_i + \beta_{16}\text{Z-score}_i + \beta_{17}\text{climate match}_i + \beta_{18}\text{native range}_i \tag{F.18}$$

$$\text{area planted}_i \sim \text{Normal}(\mu_{ai}, \sigma_a^2) \tag{F.19}$$

$$\begin{aligned}\mu_{ai} = \beta_{19} + \beta_{20}\text{native range}_i + \beta_{21}\text{climate match}_i + \beta_{22}\text{Z-score}_i \\ + \beta_{23}\text{forestry use index}_i + \beta_{24}\text{residence time}_i\end{aligned} \tag{F.20}$$

Where naturalised_i specifies whether the ith species naturalised in NZ (1) or not (0), which is modelled as drawn from a Bernoulli distribution with probability P_i of naturalisation, which is in turn modelled (on the logit scale) as a function of Z-score, climate match, native range size, the forestry use index, residence time and area planted. The explanatory variables climate match, native range size, the forestry use index, residence time and area planted are dependent on other variables, thus were modelled as drawn from separate normal distributions. Climate match modelled as a linear function of native range size; native range size was modelled as a function of Z-score; the forestry use index was modelled as a linear function of native range size and Z-score; residence time was modelled as a linear function of the forestry use index, Z-score, climate match and native range size; and area planted was modelled as a linear function of native range size, climate match, Z-score, the forestry use index and residence time. As before, β_0–β_{24} are unknown regression coefficients to be estimated from the data, and $\sigma_c^2, \sigma_n^2, \sigma_f^2, \sigma_r^2$ and σ_a^2 are unknown variances that are also estimated from the data. While Z-score causally influence other variables in the model we have not explicitly considered its cause.

The probability of *Trifolium* introductions were modelled as follows:

$$\text{introduced}_i \sim \text{Bernoulli}(P_i) \tag{F.21}$$

$$\text{logit}(P_i) = \beta_0 + \beta_1 \text{native range}_i + \beta_2 \text{climate match}_i + \beta_3 \text{GB}_i + \beta_4 \text{height}_i + \beta_5 \text{life-span}_i \tag{F.22}$$

$$\text{native range}_i \sim \text{Normal}(\mu_{ni}, \sigma_n^2) \tag{F.23}$$

$$\mu_{ni} = \beta_6 + \beta_7 \text{conditions}_i + \beta_8 \text{height}_i + \beta_9 \text{life-span}_i \tag{F.24}$$

$$\text{climate match}_i \sim \text{Normal}(\mu_{ci}, \sigma_c^2) \tag{F.25}$$

$$\mu_{ci} = \beta_{10} + \beta_{11} \text{native range}_i + \beta_{12} \text{conditions}_i \tag{F.26}$$

$$\text{height}_i \sim \text{Normal}(\mu_{hi}, \sigma_h^2) \tag{F.27}$$

$$\mu_{hi} = \beta_{13} + \beta_{14} \text{life-span}_i \tag{F.28}$$

$$\text{life-span}_i \sim \text{Bernoulli}(P_{li}) \tag{F.29}$$

$$\text{logit}(P_{li}) = \beta_{15} \tag{F.30}$$

$$\text{GB}_i \sim \text{Bernoulli}(P_{GBi}) \tag{F.31}$$

$$\text{logit}(P_{GBi}) = \beta_{16} + \beta_{17} \text{native range}_i \tag{F.32}$$

$$\text{conditions}_i \sim \text{Poisson}(P_{Coi}) \tag{F.33}$$

$$\log(P_{Coi}) = \beta_{18} + \beta_{19} \text{height}_i + \beta_{20} \text{life-span}_i \tag{F.34}$$

Where introduced_i specifies whether the ith clover species was intentionally introduced to NZ (1) or not (0), which is modelled as drawn from a Bernoulli distribution with probability

I_i of introduction, which is in turn modelled (on the logit scale) as a function of native range size, climate match, presence in GB, centre of origin, and economic use. Continuous explanatory variables were modelled as being drawn for separate normal distributions; binary variables were modelled as being drawn from a Bernoulli distribution with a probability P_{GBi} and P_{li} which was in turn modelled on a logit scale; and the count variable was modelled as being drawn from a Poisson distribution P_{Coi} on a log scale. Native range size was modelled as a linear function of range of conditions tolerated, height and life-span; climate match was modelled as a linear function of native range size and range of conditions tolerated. Height was modelled as a linear function of life-span. Presence in GB was modelled (on a logit scale) as a linear function of native range size; range of conditions tolerated was modelled (on a log scale) as a linear function of height and life-span. As previously β_0–β_{20} are regression coefficients to be estimated from the data, and σ_n^2, σ_c^2 and σ_h^2 are unknown variances that are also estimated from the data.

Trifolium naturalisation was modelled thus:

$$\text{naturalisaed}_i \sim \text{Bernoulli}(P_i) \tag{F.35}$$

$$\begin{aligned}\text{logit}(P_i) = \beta_0 + \beta_1\text{residence time}_i + \beta_2\text{climate match}_i + \beta_3\text{native range}_i \\ + \beta_4\text{life-span}_i + \beta_5\text{height}_i + \beta_6\text{area planted}_i\end{aligned} \tag{F.36}$$

$$\text{residence time}_i \sim \text{Normal}(\mu_{ri}, \sigma_r^2) \tag{F.37}$$

$$\mu_{ri} = \beta_7 + \beta_8\text{climate match}_i + \beta_9\text{native range}_i + \beta_{10}\text{height}_i + \beta_{11}\text{life-span}_i + \beta_{12}\text{GB}_i \tag{F.38}$$

$$\text{climate match}_i \sim \text{Normal}(\mu_{ci}, \sigma_c^2) \tag{F.39}$$

$$\mu_{ci} = \beta_{13} + \beta_{14}\text{native range}_i \tag{F.40}$$

$$\text{native range}_i \sim \text{Normal}(\mu_{ni}, \sigma_n^2) \tag{F.41}$$

$$\mu_{ni} = \beta_{15} + \beta_{16}\text{height}_i + \beta_{17}\text{life-span}_i \tag{F.42}$$

$$\text{height}_i \sim \text{Normal}(\mu_{hi}, \sigma_h^2) \tag{F.43}$$

$$\mu_{ni} = \beta_{18} + \beta_{19}\text{life-span}_i \tag{F.44}$$

$$\text{life-span}_i \sim \text{Bernoulli}(P_{li}) \tag{F.45}$$

$$\text{logit}(P_{li}) = \beta_{20} \tag{F.46}$$

$$\text{GB}_i \sim \text{Bernoulli}(P_{GBi}) \tag{F.47}$$

$$\text{logit}(P_{GBi}) = \beta_{21} + \beta_{22}\text{native range}_i \tag{F.48}$$

$$\text{area planted}_i \sim \text{Normal}(\mu_{ai}, \sigma_a^2) \tag{F.49}$$

$$\begin{aligned}\mu_{ai} = \beta_{23} + \beta_{24}\text{residence time}_i + \beta_{25}\text{climate match}_i \\ + \beta_{26}\text{native range}_i + \beta_{27}\text{height}_i + \beta_{28}\text{life-span}_i + \beta_{29}\text{GB}_i\end{aligned} \tag{F.50}$$

Where naturalised$_i$ specifies whether the ith *Trifolium* species has naturalised in NZ (1) or not (0), which is modelled as drawn from a Bernoulli distribution with probability P_i of naturalisation, which is in turn modelled (on the logit scale) as a function of residence time, climate match, native range size, life-span, height, and area planted. Continuous explanatory variables were modelled as being drawn for separate normal distributions, and the binary variables (presence in GB and life-span) were modelled as being drawn from a Bernoulli distribution with some probability, which was in turn modelled on a logit scale. Residence time was modelled as a linear function of range of climate match, native range size, height, life-span, and presence in GB. Climate match was modelled as a linear function of native range size; native range size was modelled as a linear function of height and life-span; height was modelled as linear function of life-span; area planted was modelled as linear function of residence time, climate match, native range size, height, life-span and presence in GB. Presence in GB was modelled (on a logit scale) as a function of native range size. β_0–β_{29} are regression coefficients to be estimated from the data, and $\sigma_r^2, \sigma_c^2, \sigma_n^2, \sigma_h^2$ and σ_a^2 are unknown variance that was estimated from the data.

Appendix G

R tutorial: exploratory path analysis

```
# Call the BRugs library for this session
  library(BRugs)

# Set a working directory
  setwd("E:\\Uni.backup\\Thesis_Chapter_4\\New_idea\\draft_May")

# Read pine data
  pine <- read.csv("C:\\your\\file\\path\\Pine.csv")

# Subset pine data to include only those that were introduced
  nz <- subset(pine, int.nz==1)

# Read Trifolium introduction data
  tri.int <- read.csv("TrifoliumInt.csv")

# Read Trifolium naturalisation data
  tri.nat <- read.csv("TrifoliumNat.csv")

# Write a function to standardise your variables by subtracting the mean and dividing
# by two standard deviations (z-scores)
  stan <- function(x) ((x - mean(x, na.rm=T)) / sd(x, na.rm=T))/2


### PINUS INTRODUCTION

# Specify the model variables, standardise the continuous onse using z-scores
nz.int <- pine$int.nz
clim <- stan(pine$Nzclim.perc)
range <- stan(pine$EOO)
zscore <- stan(pine$zscore)
cabi <-stan(pine$CABI.abstracts)
N <- length(nz.int)

# Fit the full model containing all hypothesised causal links
mod <- "  model {
    for(i in 1:N) {
      nz.int[i] ~ dbern(p[i])
      logit(p[i]) <- b[1] + b[2]*range[i] + b[3]*clim[i] + b[4]*zscore[i] + b[5]*cabi[i]

      cabi[i] ~ dnorm(mu.cabi[i], tau[1])
      clim[i] ~ dnorm(mu.clim[i], tau[2])
      zscore[i] ~ dnorm(mu.zscore[i], tau[3])
      range[i] ~ dnorm(mu.range[i], tau[4])

      mu.cabi[i] <- b[6] + b[7]*range[i] + b[8]*zscore[i]
      mu.clim[i] <- b[9] + b[10]*range[i]
      mu.zscore[i] <- b[11]
      mu.range[i] <- b[12] + b[13]*zscore[i]
    }

# Set priors. Gelman et al (2008) suggest for regression parameters to use a prior
# with a scale (=standard deviation) of 2.5, except that BUGS uses the precision (= 1/(sd*sd)).
# The other thing Gelman suggests is to use a Cauchy (a t-distribution with one degree
# of freedom) rather than a t-distribution with 7 degrees of freedom.

    b[1] ~ dt(0, 0.01, 1)
    for(i in 2:5) {
      b[i] ~ dt(0, 0.16, 1)
    }
```

```
    for(i in 6:13) {
      b[i] ~ dnorm(0, 0.0001)
    }
    for(i in 1:4) {
      tau[i] <- pow(sigma[i], -2)
      log(sigma[i]) <- l.sigma[i]
      l.sigma[i] ~ dnorm(0, 0.0001)
    }
  }"

# Write model
  write(mod, "bugs model.txt")

# Write data and intial values to files
  bugsData(list(N=N, nz.int=nz.int, clim=clim, range=range, zscore=zscore, cabi=cabi),
  fileName="data.txt")
  bugsData(list(b=rnorm(13), l.sigma=rnorm(4)), fileName="inits1.txt")
  bugsData(list(b=rnorm(13), l.sigma=rnorm(4)), fileName="inits2.txt")
  bugsData(list(b=rnorm(13), l.sigma=rnorm(4)), fileName="inits3.txt")

  modelCheck("bugs model.txt")
  modelData("data.txt")
  modelCompile(numChains=3)
  modelInits(c("inits1.txt", "inits2.txt", "inits3.txt"))
  modelGenInits()
  modelUpdate(10000)
  samplesSet(c("b", "sigma", "l.sigma"))
  modelUpdate(10000)
  post.b.nzint <- samplesStats("b")
  post.sigma.nzint <- samplesStats("l.sigma")

# Extract the mean and sd of the posterior estimates for the parameters (b and sigma)
  mean.b.nzint <- post.b.nzint[, 1]
  sd.b.nzint <- post.b.nzint[, 2]
  mean.sigma.nzint <- post.sigma.nzint[, 1]
  sd.sigma.nzint <- post.sigma.nzint[, 2]

# Refit the full model with variable selection indicators, including the posterior
# distributions as priors

mod <- "  model {
    for(i in 1:N) {
      nz.int[i] ~ dbern(p[i])
      logit(p[i]) <- b[1] + w[1]*b[2]*range[i] + w[2]*b[3]*clim[i] + w[3]*b[4]*zscore[i] +
      w[4]*b[5]*cabi[i]

      cabi[i] ~ dnorm(mu.cabi[i], tau[1])
      clim[i] ~ dnorm(mu.clim[i], tau[2])
      zscore[i] ~ dnorm(mu.zscore[i], tau[3])
      range[i] ~ dnorm(mu.range[i], tau[4])

      mu.cabi[i] <- b[6] + w[5]*b[7]*range[i] + w[6]*b[8]*zscore[i]
      mu.clim[i] <- b[9] + w[7]*b[10]*range[i]
      mu.zscore[i] <- b[11]
      mu.range[i] <- b[12] + w[8]*b[13]*zscore[i]
    }
    for(i in 1:13) {
      b[i] ~ dnorm(mean.b.nzint[i], sd.b.nzint[i])
    }
    for(i in 1:4) {
      tau[i] <- pow(sigma[i], -2)
      log(sigma[i]) <- l.sigma[i]
      l.sigma[i] ~ dnorm(mean.sigma.nzint[i], sd.sigma.nzint[i])
    }

# Priors for w
    for(i in 1:8) {
      w[i] ~ dbern(theta[i])
      theta[i] ~ dbeta(1,1)
     }

# Combine w and b
    v[1] <- w[1]*b[2]
```

```
    v[2] <- w[2]*b[3]
    v[3] <- w[3]*b[4]
    v[4] <- w[4]*b[5]
    v[5] <- w[5]*b[7]
    v[6] <- w[6]*b[8]
    v[7] <- w[7]*b[10]
    v[8] <- w[8]*b[13]

  }"

# Write model
  write(mod, "bugs model.txt")

# Write data and intial values to files
  bugsData(list(N=N, nz.int=nz.int, clim=clim, range=range, zscore=zscore, cabi=cabi,
  mean.b.nzint=mean.b.nzint,
  sd.b.nzint=sd.b.nzint,mean.sigma.nzint=mean.sigma.nzint,
  sd.sigma.nzint=sd.sigma.nzint), fileName="data.txt")
  bugsData(list(b=rnorm(13), l.sigma=rnorm(4), theta=runif(8)), fileName="inits1.txt")
  bugsData(list(b=rnorm(13), l.sigma=rnorm(4), theta=runif(8)), fileName="inits2.txt")
  bugsData(list(b=rnorm(13), l.sigma=rnorm(4), theta=runif(8)), fileName="inits3.txt")

  modelCheck("bugs model.txt")
  modelData("data.txt")
  modelCompile(numChains=3)
  modelInits(c("inits1.txt", "inits2.txt", "inits3.txt"))
  modelGenInits()
  modelUpdate(10000)
  samplesSet(c("b", "sigma", "w", "v"))
  modelUpdate(10000)
  out.b.nzint <- samplesStats("b")
  out.w.nzint <- samplesStats("w")
  out.v.nzint <- samplesStats("v")

  out.nzint <- rbind(out.b.nzint, out.w.nzint, out.v.nzint)
  write.csv(out.nzint, "IntPine.csv")


##### PINUS NATURALISATION

# Naturalisation stage varaibles NZ
  nz.nat <- nz$nat.nz
  range <- stan(nz$EOO)
  clim <- stan(nz$Nzclim.perc)
  area <- stan(nz$area.log1)
  zscore <- stan(nz$zscore)
  residence <- 2011-nz$nz.intro.date
  resid <- stan(residence)
  cabi <- stan(nz$CABI.abstracts)

  N <- length(nz.nat)

# Fit the full model containing all hypothesised causal links
mod <- "  model {
    for(i in 1:N) {
      nz.nat[i] ~ dbern(p[i])
      logit(p[i]) <- b[1] + b[2]*area[i] + b[3]*clim[i] + b[4]*zscore[i] + b[5]*cabi[i]
      + b[6]*resid[i] + b[7]*range[i]

      area[i] ~ dnorm(mu.area[i], tau[1])
      clim[i] ~ dnorm(mu.clim[i], tau[2])
      zscore[i] ~ dnorm(mu.zscore[i], tau[3])
      range[i] ~ dnorm(mu.range[i], tau[4])
      resid[i] ~ dnorm(mu.resid[i], tau[5])
      cabi[i] ~ dnorm(mu.cabi[i], tau[6])

      mu.area[i] <- b[8] + b[9]*clim[i] + b[10]*zscore[i] +b[11]*cabi[i] + b[12]*range[i]
      + b[13]*resid[i]
      mu.clim[i] <- b[14] + b[15]*range[i]
      mu.zscore[i] <- b[16]
      mu.range[i] <- b[17] + b[18]*zscore[i]
      mu.resid[i] <- b[19] + b[20]*clim[i] + b[21]*zscore[i] + b[22]*cabi[i] +
      b[23]*range[i]
```

```
      mu.cabi[i] <- b[24] + b[25]*zscore[i] + b[26]*range[i]
    }
    b[1] ~ dt(0, 0.01, 1)
    for(i in 2:7) {
      b[i] ~ dt(0, 0.16, 1)
    }
    for(i in 8:26) {
      b[i] ~ dnorm(0, 0.0001)
    }
    for(i in 1:6) {
      tau[i] <- pow(sigma[i], -2)
      log(sigma[i]) <- l.sigma[i]
      l.sigma[i] ~ dnorm(0, 0.0001)
    }
  }"

# Write model
  write(mod, "bugs model.txt")

# Write data and intial values to files
  bugsData(list(N=N, nz.nat=nz.nat, area=area, clim=clim, range=range,
  zscore=zscore,
  resid=resid, cabi=cabi), fileName="data.txt")
  bugsData(list(b=rnorm(26), l.sigma=rnorm(6)), fileName="inits1.txt")
  bugsData(list(b=rnorm(26), l.sigma=rnorm(6)), fileName="inits2.txt")
  bugsData(list(b=rnorm(26), l.sigma=rnorm(6)), fileName="inits3.txt")

  modelCheck("bugs model.txt")
  modelData("data.txt")
  modelCompile(numChains=3)
  modelInits(c("inits1.txt", "inits2.txt", "inits3.txt"))
  modelGenInits()
  modelUpdate(10000)
  samplesSet(c("b", "sigma", "l.sigma"))
  modelUpdate(10000)
  post.b.nznat <- samplesStats("b")
  post.sigma.nznat <- samplesStats("l.sigma")

# Extract the mean and sd of the posterior estimates for the parameters (b and sigma)
  mean.b.nznat <- post.b.nznat[, 1]
  sd.b.nznat <- post.b.nznat[, 2]
  mean.sigma.nznat <- post.sigma.nznat[, 1]
  sd.sigma.nznat <- post.sigma.nznat[, 2]

# Refit the full model with variable selection indicators, including the posterior
# distributions as priors
mod <- "  model {
    for(i in 1:N) {
      nz.nat[i] ~ dbern(p[i])
      logit(p[i]) <- b[1] + w[1]*b[2]*area[i] + w[2]*b[3]*clim[i] + w[3]*b[4]*zscore[i] +
      w[4]*b[5]*cabi[i] + w[5]*b[6]*resid[i] + w[6]*b[7]*range[i]

      area[i] ~ dnorm(mu.area[i], tau[1])
      clim[i] ~ dnorm(mu.clim[i], tau[2])
      zscore[i] ~ dnorm(mu.zscore[i], tau[3])
      range[i] ~ dnorm(mu.range[i], tau[4])
      resid[i] ~ dnorm(mu.resid[i], tau[5])
      cabi[i] ~ dnorm(mu.cabi[i], tau[6])

      mu.area[i] <- b[8] + w[7]*b[9]*clim[i] + w[8]*b[10]*zscore[i] + w[9]*b[11]*cabi[i]
      +  w[10]*b[12]*range[i] + w[11]*b[13]*resid[i]
      mu.clim[i] <- b[14] + w[12]*b[15]*range[i]
      mu.zscore[i] <- b[16]
      mu.range[i] <- b[17]+ w[13]*b[18]*zscore[i]
      mu.resid[i] <- b[19] + w[14]*b[20]*clim[i] + w[15]*b[21]*zscore[i] +
      w[16]*b[22]*cabi[i] +
      w[17]*b[23]*range[i]
      mu.cabi[i] <- b[24] + w[18]*b[25]*zscore[i] + w[19]*b[26]*range[i]
    }
    for(i in 1:26) {
      b[i] ~ dnorm(mean.b.nznat[i], sd.b.nznat[i])
    }
    for(i in 1:6) {
```

```
      tau[i] <- pow(sigma[i], -2)
      log(sigma[i]) <- l.sigma[i]
      l.sigma[i] ~ dnorm(mean.sigma.nznat[i], sd.sigma.nznat[i])
    }

# Priors for w
    for(i in 1:19) {
      w[i] ~ dbern(theta[i])
      theta[i] ~ dbeta(1,1)
     }

# Combine w and b
    v[1] <- w[1]*b[2]
    v[2] <- w[2]*b[3]
    v[3] <- w[3]*b[4]
    v[4] <- w[4]*b[5]
    v[5] <- w[5]*b[6]
    v[6] <- w[6]*b[7]
    v[7] <- w[7]*b[9]
    v[8] <- w[8]*b[10]
    v[9] <- w[9]*b[11]
    v[10] <- w[10]*b[12]
    v[11] <- w[11]*b[13]
    v[12] <- w[12]*b[15]
    v[13] <- w[13]*b[18]
    v[14] <- w[14]*b[20]
    v[15] <- w[15]*b[21]
    v[16] <- w[16]*b[22]
    v[17] <- w[17]*b[23]
    v[18] <- w[18]*b[25]
    v[19] <- w[19]*b[26]
  }"

# Write model
  write(mod, "bugs model.txt")

# Write data and intial values to files
  bugsData(list(N=N, nz.nat=nz.nat, area=area, clim=clim, range=range,
  zscore=zscore,
  resid=resid, cabi=cabi, mean.b.nznat=mean.b.nznat, sd.b.nznat=sd.b.nznat,
  mean.sigma.nznat=mean.sigma.nznat, sd.sigma.nznat=sd.sigma.nznat),
  fileName="data.txt")

  modelCheck("bugs model.txt")
  modelData("data.txt")
  modelCompile(numChains=3)
  modelInits(c("inits1.txt", "inits2.txt", "inits3.txt"))
  modelGenInits()
  modelUpdate(10000)
  samplesSet(c("b", "sigma", "w", "v"))
  modelUpdate(10000)
  out.b.nznat <- samplesStats("b")
  out.w.nznat <- samplesStats("w")
  out.v.nznat <- samplesStats("v")

  out.nznat <- rbind(out.b.nznat, out.w.nznat, out.v.nznat)
  write.csv(out.nznat, "NatPine.csv")


 ### TRIFOLIUM INTRODUCTION

# Specify the model variables, standardise the continuous ones using z-scores
  nz.int <- tri.int$intro # Binary
  clim <- stan(tri.int$class05) # Continuous
  range <- stan(tri.int$logarea) # Continuous
gb <- tri.int$britpres # Binary
  height <- stan(tri.int$maxht)
lifesp <-tri.int$lifespan ## Categorical
  conditions <- tri.int$wwf.totnumbio  ## Count

N <- length(nz.int)

# Fit the full model containing all hypothesised causal links
```

```
mod <- "  model {
    for(i in 1:N) {
      nz.int[i] ~ dbern(p[i])
      logit(p[i]) <- b[1] + b[2]*range[i] + b[3]*clim[i] + b[4]*height[i]
      + b[5]*lifesp[i] + b[6]*gb[i]

      clim[i] ~ dnorm(mu.clim[i], tau[1])
      mu.clim[i] <- b[7] + b[8]*range[i] + b[9]*conditions[i]

      range[i] ~ dnorm(mu.range[i], tau[2])
      mu.range[i] <- b[10] + b[11]*conditions[i] + b[12]*height[i] +b[13]*lifesp[i]

      height[i] ~ dnorm(mu.height[i], tau[3])
      mu.height[i] <- b[14] + b[15]*lifesp[i]

      lifesp[i] ~ dbern(p.lifesp[i])
      logit(p.lifesp[i]) <- b[16]

      gb[i] ~ dbern(p.gb[i])
      logit(p.gb[i]) <- b[17] + b[18]*range[i]

      conditions[i] ~ dpois(mu.conditions[i])
      log(mu.conditions[i]) <- b[19] + b[20]*height[i] + b[21]*lifesp[i]

    }

    b[1] ~ dt(0, 0.01, 1)
    b[17] ~ dt(0, 0.01, 1)
    b[16] ~ dt(0, 0.01, 1)
    b[19] ~ dt(0, 0.01, 1)

    for(i in 2:6) {
      b[i] ~ dt(0, 0.16, 1)
    }

    for(i in 7:15) {
      b[i] ~ dnorm(0, 0.0001)
    }
    for(i in 18:18) {
      b[i] ~ dt(0, 2.5, 7)
    }

    for(i in 20:21) {
      b[i] ~ dt(0, 2.5, 7)
    }

    for(i in 1:3) {
      tau[i] <- pow(sigma[i], -2)
      log(sigma[i]) <- l.sigma[i]
      l.sigma[i] ~ dnorm(0, 0.0001)
    }

  }"

# Write model
  write(mod, "bugs model.txt")

# write data and intial values to files
  bugsData(list(N=N, nz.int=nz.int, clim=clim, range=range, gb=gb, height=height,
  lifesp=lifesp, conditions=conditions), fileName="data.txt")
  bugsData(list(b=rnorm(21), l.sigma=rnorm(3)), fileName="inits1.txt")
  bugsData(list(b=rnorm(21), l.sigma=rnorm(3)), fileName="inits2.txt")
  bugsData(list(b=rnorm(21), l.sigma=rnorm(3)), fileName="inits3.txt")

  modelCheck("bugs model.txt")
  modelData("data.txt")
  modelCompile(numChains=3)
  modelInits(c("inits1.txt", "inits2.txt", "inits3.txt"))
  modelGenInits()
  modelUpdate(10000)
  samplesSet(c("b", "sigma", "l.sigma"))
  modelUpdate(10000)
  post.b.nzint <- samplesStats("b")
```

```
  post.sigma.nzint <- samplesStats("l.sigma")

# Extract the mean and sd of the posterior estimates for the parameters (b and sigma)
  mean.b.nzint <- post.b.nzint[, 1]
  sd.b.nzint <- post.b.nzint[, 2]
  mean.sigma.nzint <- post.sigma.nzint[, 1]
  sd.sigma.nzint <- post.sigma.nzint[, 2]

# Refit the full model with variable selection indicators, including the posterior
# distributions as priors

mod <- "  model {
    for(i in 1:N) {
      nz.int[i] ~ dbern(p[i])
      logit(p[i]) <- b[1] + w[1]*b[2]*range[i] + w[2]*b[3]*clim[i] + w[3]*b[4]*height[i] +
      w[4]*b[5]*lifesp[i] + w[5]*b[6]*gb[i]

      clim[i] ~ dnorm(mu.clim[i], tau[1])
      mu.clim[i] <- b[7] + w[6]*b[8]*range[i] + w[7]*b[9]*conditions[i]

      range[i] ~ dnorm(mu.range[i], tau[2])
      mu.range[i] <- b[10] + w[8]*b[11]*conditions[i] + w[9]*b[12]*height[i]
      + w[10]*b[13]*lifesp[i]

      height[i] ~ dnorm(mu.height[i], tau[3])
      mu.height[i] <- b[14] + w[11]*b[15]*lifesp[i]

      lifesp[i] ~ dbern(p.lifesp[i])
      logit(p.lifesp[i]) <- b[16]

      gb[i] ~ dbern(p.gb[i])
      logit(p.gb[i]) <- b[17] + w[12]*b[18]*range[i]

      conditions[i] ~ dpois(mu.conditions[i])
      log(mu.conditions[i]) <- b[19] + w[13]*b[20]*height[i] + w[14]*b[21]*lifesp[i]

    }
    for(i in 1:21) {
      b[i] ~ dnorm(mean.b.nzint[i], sd.b.nzint[i])
    }
    for(i in 1:3) {
      tau[i] <- pow(sigma[i], -2)
      log(sigma[i]) <- l.sigma[i]
      l.sigma[i] ~ dnorm(mean.sigma.nzint[i], sd.sigma.nzint[i])
    }

# Priors for w
    for(i in 1:14) {
      w[i] ~ dbern(theta[i])
      theta[i] ~ dbeta(1,1)
     }

# Combine w and b
    v[1] <- w[1]*b[2]
    v[2] <- w[2]*b[3]
    v[3] <- w[3]*b[4]
    v[4] <- w[4]*b[5]
    v[5] <- w[5]*b[6]
    v[6] <- w[6]*b[8]
    v[7] <- w[7]*b[9]
    v[8] <- w[8]*b[11]
    v[9] <- w[9]*b[12]
    v[10] <- w[10]*b[13]
    v[11] <- w[11]*b[15]
    v[12] <- w[12]*b[18]
    v[13] <- w[13]*b[20]
    v[14] <- w[14]*b[21]


  }"

# Write model
  write(mod, "bugs model.txt")
```

```
# Write data and intial values to files
  bugsData(list(N=N, nz.int=nz.int, clim=clim, range=range, gb=gb, height=height, lifesp=lifesp,
  conditions=conditions,mean.b.nzint=mean.b.nzint, sd.b.nzint=sd.b.nzint,
  mean.sigma.nzint=mean.sigma.nzint,
  sd.sigma.nzint=sd.sigma.nzint), fileName="data.txt")
  bugsData(list(b=rnorm(21), l.sigma=rnorm(3)), fileName="inits1.txt")
  bugsData(list(b=rnorm(21), l.sigma=rnorm(3)), fileName="inits2.txt")
  bugsData(list(b=rnorm(21), l.sigma=rnorm(3)), fileName="inits3.txt")

  modelCheck("bugs model.txt")
  modelData("data.txt")
  modelCompile(numChains=3)
  modelInits(c("inits1.txt", "inits2.txt", "inits3.txt"))
  modelGenInits()
  modelUpdate(10000)
  samplesSet(c("b", "sigma", "w", "v"))
  modelUpdate(10000)
  out.b.nzint <- samplesStats("b")
  out.w.nzint <- samplesStats("w")
  out.v.nzint <- samplesStats("v")

  out.nzint <- rbind(out.b.nzint, out.w.nzint, out.v.nzint)
  write.csv(out.nzint, "IntTrifolium.csv")


#### TRIFOLIUM NATURALISATION

# Naturalisation stage varaibles NZ
  nz.nat <- tri.nat$nat # Binary
  residence <- 2011-tri.nat$intdate
  resid <- stan(residence) # Cont
  clim <- stan(tri.nat$class05) # Cont
  gb <- tri.nat$britpres # Binary
  height <- stan(tri.nat$maxht) # cont
  lifesp <- tri.nat$lifespan # Binary
  range <- stan(tri.nat$logarea)
  ha <- stan(tri.nat$ltotha)
  N <- length(nz.nat)

# Fit the full model containing all hypothesised causal links
mod <- "  model {
    for(i in 1:N) {
      nz.nat[i] ~ dbern(p[i])
      logit(p[i]) <- b[1] + b[2]*resid[i] + b[3]*clim[i] + b[4]*range[i] + b[5]*lifesp[i]
      +  b[6]*height[i] + b[7]*ha[i]

      resid[i] ~ dnorm(mu.resid[i], tau[1])
      mu.resid[i] <- b[8] + b[9]*clim[i] + b[10]*range[i] + b[11]*height[i] +
     b[12]*lifesp[i] + b[13]*gb[i]

      clim[i] ~ dnorm(mu.clim[i], tau[2])
      mu.clim[i] <- b[14] + b[15]*range[i]

      range[i] ~ dnorm(mu.range[i], tau[3])
      mu.range[i] <- b[16] + b[17]*height[i] + b[18]*lifesp[i]

      height[i] ~ dnorm(mu.height[i], tau[4])
      mu.height[i] <- b[19] + b[20]*lifesp[i]

      lifesp[i] ~ dbern(p.lifesp[i])
      logit(p.lifesp[i]) <- b[21]

      gb[i] ~ dbern(p.gb[i])
      logit(p.gb[i]) <- b[22] + b[23]*range[i]

      ha[i] ~ dnorm(mu.ha[i], tau[5])
      mu.ha[i] <- b[24] + b[25]*resid[i] + b[26]*clim[i] + b[27]*range[i] + b[28]*height[i]
     + b[29]*lifesp[i] + b[30]*gb[i]

    }
    b[1] ~ dt(0, 0.01, 1)
    b[21] ~ dt(0, 0.01, 1)
```

```
    b[22] ~ dt(0, 0.01, 1)

    for(i in 2:7) {
      b[i] ~ dt(0, 0.16, 1)
    }

    for(i in 23:23) {
      b[i] ~ dt(0, 0.16, 1)
    }

    for(i in 8:20) {
      b[i] ~ dnorm(0, 0.0001)
    }

    for(i in 24:30) {
      b[i] ~ dnorm(0, 0.0001)
    }

    for(i in 1:5) {
      tau[i] <- pow(sigma[i], -2)
      log(sigma[i]) <- l.sigma[i]
      l.sigma[i] ~ dnorm(0, 0.0001)
    }
  }"

# Write model
  write(mod, "bugs model.txt")

# Write data and intial values to files
  bugsData(list(N=N, nz.nat=nz.nat, resid=resid, clim=clim, gb=gb, height=height,
  lifesp=lifesp, range=range, ha=ha), fileName="data.txt")
  bugsData(list(b=rnorm(30), l.sigma=rnorm(5)), fileName="inits1.txt")
  bugsData(list(b=rnorm(30), l.sigma=rnorm(5)), fileName="inits2.txt")
  bugsData(list(b=rnorm(30), l.sigma=rnorm(5)), fileName="inits3.txt")

  modelCheck("bugs model.txt")
  modelData("data.txt")
  modelCompile(numChains=3)
  modelInits(c("inits1.txt", "inits2.txt", "inits3.txt"))
  modelGenInits()
  modelUpdate(10000)
  samplesSet(c("b", "sigma", "l.sigma"))
  modelUpdate(10000)
  post.b.nznat <- samplesStats("b")
  post.sigma.nznat <- samplesStats("l.sigma")

# extract the mean and sd of the posterior estimates for the parameters (b and sigma)
  mean.b.nznat <- post.b.nznat[, 1]
  sd.b.nznat <- post.b.nznat[, 2]
  mean.sigma.nznat <- post.sigma.nznat[, 1]
  sd.sigma.nznat <- post.sigma.nznat[, 2]

# Refit the full model with variable selection indicators, including the posterior
# distributions as priors
mod <- "  model {
    for(i in 1:N) {
      nz.nat[i] ~ dbern(p[i])
      logit(p[i]) <- b[1] + w[1]*b[2]*resid[i] + w[2]*b[3]*clim[i] + w[3]*b[4]*range[i]
      + w[4]*b[5]*lifesp[i] +  w[5]*b[6]*height[i] + w[6]*b[7]*ha[i]

      resid[i] ~ dnorm(mu.resid[i], tau[1])
      mu.resid[i] <- b[8] + w[7]*b[9]*clim[i] + w[8]*b[10]*range[i] + w[9]*b[11]*height[i]
      + w[10]*b[12]*lifesp[i] + w[11]*b[13]*gb[i]

      clim[i] ~ dnorm(mu.clim[i], tau[2])
      mu.clim[i] <- b[14] + w[12]*b[15]*range[i]

      range[i] ~ dnorm(mu.range[i], tau[3])
      mu.range[i] <- b[16] + w[13]*b[17]*height[i] + w[14]*b[18]*lifesp[i]

      height[i] ~ dnorm(mu.height[i], tau[4])
      mu.height[i] <- b[19] + w[15]*b[20]*lifesp[i]
```

```
      lifesp[i] ~ dbern(p.lifesp[i])
      logit(p.lifesp[i]) <- b[21]

      gb[i] ~ dbern(p.gb[i])
      logit(p.gb[i]) <- b[22] + w[16]*b[23]*range[i]

      ha[i] ~ dnorm(mu.ha[i], tau[5])
      mu.ha[i] <- b[24] + w[17]*b[25]*resid[i] + w[18]*b[26]*clim[i] + w[19]*b[27]*range[i]
      + w[20]*b[28]*height[i] + w[21]*b[29]*lifesp[i] + w[22]*b[30]*gb[i]

    }

    for(i in 1:30) {
      b[i] ~ dnorm(mean.b.nznat[i], sd.b.nznat[i])
    }
    for(i in 1:5) {
      tau[i] <- pow(sigma[i], -2)
      log(sigma[i]) <- l.sigma[i]
      l.sigma[i] ~ dnorm(mean.sigma.nznat[i], sd.sigma.nznat[i])
    }

# Priors for w
    for(i in 1:22) {
      w[i] ~ dbern(theta[i])
      theta[i] ~ dbeta(1,1)
     }

# Combine w and b
    v[1] <- w[1]*b[2]
    v[2] <- w[2]*b[3]
    v[3] <- w[3]*b[4]
    v[4] <- w[4]*b[5]
    v[5] <- w[5]*b[6]
    v[6] <- w[6]*b[7]
    v[7] <- w[7]*b[9]
    v[8] <- w[8]*b[10]
    v[9] <- w[9]*b[11]
    v[10] <- w[10]*b[12]
    v[11] <- w[11]*b[13]
    v[12] <- w[12]*b[15]
    v[13] <- w[13]*b[17]
    v[14] <- w[14]*b[18]
    v[15] <- w[15]*b[20]
    v[16] <- w[16]*b[23]
    v[17] <- w[17]*b[25]
    v[18] <- w[18]*b[26]
    v[19] <- w[19]*b[27]
    v[20] <- w[20]*b[28]
    v[21] <- w[21]*b[29]
    v[22] <- w[22]*b[30]

  }"

# Write model
  write(mod, "bugs model.txt")

# Write data and intial values to files
  bugsData(list(N=N, nz.nat=nz.nat, resid=resid, clim=clim, gb=gb, height=height, lifesp=lifesp,
  range=range, ha=ha, mean.b.nznat=mean.b.nznat, sd.b.nznat=sd.b.nznat,
  mean.sigma.nznat=mean.sigma.nznat, sd.sigma.nznat=sd.sigma.nznat), fileName="data.txt")

  modelCheck("bugs model.txt")
  modelData("data.txt")
  modelCompile(numChains=3)
  modelInits(c("inits1.txt", "inits2.txt", "inits3.txt"))
  modelGenInits()
  modelUpdate(10000)
  samplesSet(c("b", "sigma", "w", "v"))
  modelUpdate(10000)
  out.b.nznat <- samplesStats("b")
  out.w.nznat <- samplesStats("w")
  out.v.nznat <- samplesStats("v")
```

```
out.nznat <- rbind(out.b.nznat, out.w.nznat, out.v.nznat)
write.csv(out.nznat, "NatTrifolium.csv")
```

Appendix H

Full results for path analyses

Table H.1. Full results of path models describing introduction and naturalisation of *Pinus* and *Trifolium* to New Zealand. Shown are species and the stage that the model represents; the parameter from the relevant equation that the results relate to (see Appendix F); the causal link between variables that the result relates to; the probability (P±SD) of a link being present in the best model for that stage; and the model averaged weight ($W \pm$ 95% credibility intervals) for the effect size of each link (the probability of a link being present in the best model multiplied by the parameter value for the link). Links with ≥50% probability of being present in the best models are highlighted in **bold**.

Species and stage	Parameter	Casual link	P (±SD)	W (±95% CI)
Pinus introductions	β_1	Z-score → Introduction	0.43 (0.49)	0.26 (−0.16–1.49)
	β_2	Climate match → Introduction	**1.00 (0.00)**	14.17 (13.25–15.07)
	β_3	Native range size → Introduction	**0.61 (0.49)**	0.60 (0.00–2.01)
	β_4	Forestry use index → Introduction	**1.00 (0.00)**	10.58 (9.78–11.38)
	β_6	Native range size → Climate match	0.09 (0.29)	0.01 (0.00–0.21)
	β_8	Z-score → Native range size	0.03 (0.18)	0.00 (0.00–0.00)
	β_{10}	Native range size → Forestry use index	**0.79 (0.41)**	0.23 (0.00–0.46)
	β_{11}	Z-score → Forestry use index	0.11 (0.32)	0.02 (0.00–0.24)
Pinus naturalisations	β_1	Z-score → Naturalisation	**0.99 (0.01)**	10.60 (10.01–11.18)
	β_2	Climate match → Naturalisation	**0.99 (0.03)**	10.68 (10.05–11.33)
	β_3	Native range size → Naturalisation	**0.60 (0.49)**	1.21 (0.00–12.97)
	β_4	Forestry use index → Naturalisation	**1.00 (0.02)**	12.59 (12.02–13.18)
	β_5	Residence time → Naturalisation	**0.61 (0.79)**	1.35 (0.00–3.25)
	β_6	Area planted → Naturalisation	**0.57 (0.49)**	1.43 (0.00–3.45)
	β_8	Native range size → Climate match	0.06 (0.24)	0.01 (0.00–0.12)
	β_{10}	Z-score → Native range size	0.11 (0.32)	0.02 (0.00–0.27)
	β_{12}	Native range size → Forestry use index	**0.52 (0.50)**	0.16 (0.00–0.51)

Continued on next page…

Continued. . .

Species and stage	Parameter	Casual link	P (±SD)	W (±95% CI)
	β_{13}	Z-score → Forestry use index	0.11 (0.32)	0.02 (0.00–0.28)
	β_{15}	Forestry use index → Residence time	0.18 (0.38)	0.04 (0.00–0.38)
	β_{16}	Z-score → Residence time	0.14 (0.35)	−0.03 (−0.32–0.00)
	β_{17}	Climate match → Residence time	**0.62 (0.48)**	0.21 (0.00–0.56)
	β_{18}	Native range size → Residence time	0.05 (0.22)	0.00 (0.00–0.06)
	β_{20}	Native range size → Area planted	0.20 (0.40)	0.04 (0.00–0.32)
	β_{21}	Climate match → Area planted	**0.99 (0.08)**	0.58 (0.31–0.80)
	β_{22}	Z-score → Area planted	0.03 (0.17)	0.00 (0.00–0.00)
	β_{23}	Forestry use index → Area planted	0.23 (0.42)	0.05 (0.00–0.40)
	β_{24}	Residence time → Area planted	0.06 (0.24)	0.01 (0.00–0.15)
Trifolium introductions	β_{1}	Native range size → Introduction	**1.00 (0.02**)	2.84 (1.58–4.16)
	β_{2}	Climate match → Introduction	**0.78 (0.42)**	0.81 (0.00–1.89)
	β_{3}	Presence in GB → Introduction	**0.86 (0.35)**	−1.23 (−2.52–0.00)
	β_{4}	Height → Introduction	0.49 (0.50)	0.32 (0.00-1.27)
	β_{5}	Life-span → Introduction	**0.77 (0.42)**	0.68 (0.00–0.58)
	β_{7}	Conditions tolerated → Native range size	**1.00 (0.00)**	0.18 (0.16–0.19)
	β_{8}	Height → Native range size	0.02 (0.13)	0.00 (0.00–0.00)
	β_{9}	Life-span → Native range size	0.01 (0.10)	0.00 (0.00–0.00)
	β_{11}	Native range size → Climate match	**1.00 (0.00)**	0.61 (0.51–0.72)
	β_{12}	Conditions tolerated → Climate match	0.00 (0.13)	0.00 (0.00–0.00)

Continued on next page. . .

Continued...

Species and stage	Parameter	Casual link			P (±SD)	W (±95% CI)	
	β_{14}	Life-span	→	Height	0.07 (0.26)	−0.01	(−0.15–0.00)
	β_{17}	Native range size	→	Presence in GB	**1.00 (0.00)**	6.63	(5.30–7.97)
	β_{19}	Height	→	Conditions tolerated	**0.75 (0.43)**	0.14	(0.00–0.30)
	β_{20}	Life-span	→	Conditions tolerated	0.05 (0.22)	−0.01	(−0.10–0.00)
Trifolium naturalisations	β_{1}	Residence time	→	Naturalisation	**0.85 (0.36)**	3.28	(0.00–4.82)
	β_{2}	Climate match	→	Naturalisation	**0.91 (0.28)**	4.79	(0.00–5.99)
	β_{3}	Native range size	→	Naturalisation	**0.88 (0.32)**	3.85	(0.00–5.28)
	β_{4}	Life-span	→	Naturalisation	**0.52 (0.50)**	−0.98	(−2.98–0.00)
	β_{5}	Height	→	Naturalisation	**0.65 (0.48)**	−1.14	(−2.95–0.00)
	β_{6}	Area planted	→	Naturalisation	**0.83 (0.37)**	3.16	(0.00–4.80)
	β_{8}	Climate match	→	Residence time	0.04 (0.21)	0.00	(0.00–0.08)
	β_{9}	Native range size	→	Residence time	0.05 (0.21)	0.00	(−0.07–0.00)
	β_{10}	Height	→	Residence time	0.03 (0.16)	0.00	(0.00–0.00)
	β_{11}	Life-span	→	Residence time	0.02 (0.15)	0.00	(0.00–0.00)
	β_{12}	Presence in GB	→	Residence time	**1.00 (0.00)**	1.18	(0.97–1.39)
	β_{14}	Native range size	→	Climate match	**1.00 (0.003)**	0.62	(0.40–0.84)
	β_{16}	Height	→	Native range size	0.14 (0.34)	0.03	(0.00–0.34)
	β_{17}	Life-span	→	Native range size	0.06 (0.24)	0.01	(0.00–0.12)
	β_{19}	Life-span	→	Height	0.09 (0.28)	−0.01	(−0.23–0.00)
	β_{22}	Native range size	→	Presence in GB	**0.87 (0.34)**	1.76	(0.00–3.82)

Continued on next page...

Continued. . .

Species and stage	Parameter	Casual link			P (±SD)	W (±95% CI)	
	β_{24}	Residence time	→	Area planted	**1.00 (0.01)**	5.23	(3.96–6.49)
	β_{28}	Climate match	→	Area planted	0.02 (0.14)	0.00	(0.00–0.00)
	β_{26}	Native range size	→	Area planted	0.02 (0.15)	0.00	(0.00–0.00)
	β_{27}	Height	→	Area planted	0.02 (0.13)	0.00	(0.00–0.00)
	β_{28}	Life-span	→	Area planted	0.02 (0.13)	0.00	(0.00–0.00)
	β_{29}	Presence in GB	→	Area planted	**1.00 (0.00)**	0.74	(0.41–1.16)

Appendix I

Biological Invasions paper

McGregor *et al.* (2012) *Biological Invasions* **14**, 987–998.

The following pages have been removed for copyright reasons.

This page has been removed for copyright reasons.

This page has been removed for copyright reasons.

This page has been removed for copyright reasons.

This page has been removed for copyright reasons.

This page has been removed for copyright reasons.

This page has been removed for copyright reasons.

This page has been removed for copyright reasons.

This page has been removed for copyright reasons.

This page has been removed for copyright reasons.

This page has been removed for copyright reasons.

This page has been removed for copyright reasons.

References

Adamowski, W. (2004). Why don't alien conifers invade the Bialowieża Forest? *Weed Technology*, 18: 1453–1456.

Aikio, S. and Duncan, R. P. and Hulme, P. E. (2010). Herbarium records identify the role of long-distance spread in the spatial distribution of alien plants in New Zealand. *Journal of Biogeography*, 37: 1740–1751.

Aitkin, M. (1991). Posterior Bayes factors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53: 111–142.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, 16: 3–14.

Alcaraz, C. and Vila-Gispert, A. and García-Berthou, E. (2005). Profiling invasive fish species: the importance of phylogeny and human use. *Diversity and Distributions*, 11: 289–298.

Anderson, N. O. and Galatowitsch S. M. and Gomez, N. (2006). Selection strategies to reduce invasive potential in introduced plants. *Euphytica*, 148: 203–216.

Arhonditsis, G. B. and Stow, C. A. and Steinberg, L. J. and Kenney, M. A. and Lathrop, R. C. and McBride, S. J. and Reckhow, K. H. (2006). Exploring ecological patterns with structural equation modeling and Bayesian analysis. *Ecological Modelling*, 192: 385–409.

Arnold, C. and Matthews, L. J. and Nunn, C. L. (2010). The 10ktrees website: a new online resource for primate phylogeny. *Evolutionary Anthroplogy*, 19: 112–118.

Atwater, D. Z. and Bauer, C. M. and Callaway, R. M. (2011). Indirect positive effects ameliorate strong negative effects of *Euphorbia esula* on a native plant. *Plant Ecology*, 212: 1–8.

Barney, J. N. and DiTomaso, J. M. (2008). Nonnative species and bioenergy: are we cultivating the next invader? *BioScience*, 58: 64–70.

Blackburn, T. M. and Duncan, R. P. (2001). Establishment patterns of exotic birds are constrained by non-random patterns in introduction. *Journal of Biogeography*, 28: 927–939.

Blackburn, T. M. and Pyšek, P. and Bacher, S. and Carlton, J. T. and Duncan, R. P. and Jarošík, V. and Wilson, J. R. U. and Richardson, D. M. (2011). A proposed unified framework for biological invasions. *Trends in Ecology & Evolution*, 26: 333–339.

Blomberg, S. P. and Garland Jr, T. and Ives, A .R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, 57: 717–745.

Bollen, K. A. (1989). Structural equations with latent variables. A Wiley-Interscience Publication.

Bollen, K. A. and Stine, R. A. (1993). *Testing Structural Equation Models*, chapter Bootstrapping goodness-of-fit measures in structural equation models, (pp. 111–134). John Wiley and Sons, Inc.: New Work, USA.

Boulant, N. and Garnier, A. and Curt, T. and Lepart, J. (2009). Disentangling the effects of land use, shrub cover and climate on the invasion speed of native and introduced pines in grasslands. *Diversity and Distributions*, 15: 1047–1059.

Brooks, M. L. and D'Antonio, C. M. and Richardson, D. M. and Grace, J. B. and Keeley, J. E. and DiTomaso, J. M. and Hobbs, R. J. and Pellant, M. and Pyke, D. (2004). Effects of invasive alien plants on fire regimes. *BioScience*, 54: 677–688.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7: 434–455.

Brown, S. D. J. and Collins, R. A. and Boyer, S. and Lefort, M-C. and Malumbres-Olarte, J. and Vink, C. J. and Cruickshank, R. H. (2012). SPIDER: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, 12: 562–565.

Buckley, Y. M. and Brockerhoff, E. and Langer, L. and Ledgard, N. and North, H. and Rees, M. (2005). Slowing down a pine invasion despite uncertainty in demography and dispersal. *Journal of Applied Ecology*, 42: 1020–1030.

Buddenhagen, C. E. and Chimera, C. and Clifford, P. (2009). Assessing biofuel crop invasiveness: a case study. *PLoS ONE*, 4: e5261.

Burnham, K. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York, USA: Springer-Verlag New York Inc.

Burns, R. M. and Honkala, B. H. (1990). Silvics of North America. `http://www.na.fs.fed.us/spfo/pubs/silvics_manual/table_of_contents.htm`.

CAB International (2010). The Forestry Compendium. `http://www.cabi.org/fc/`.

Caley, P. and Kuhnert, P. M. (2006). Application and evaluation of classification trees for screening unwanted plants. *Austral Ecology*, 31: 647–655.

Cariveau, D. and Irwin, R. E. and Brody, A. K. and Garcia-Mayeya, L. S. and Von Der Ohe, A. (2004). Direct and indirect effects of pollinators and seed predators to selection on plant and floral traits. *Oikos*, 104: 15–26.

Carrillo-Gavilàn, M. A. and Vilà, M. (2010). Little evidence of invasion by alien conifers in Europe. *Diversity and Distributions*, 16: 203–213.

Cassey, P. and Blackburn, T. M. and Jones, K. E. and Lockwood, J. L. (2004). Mistakes in the analysis of exotic species establishment: source pool designation and correlates of introduction success among parrots (aves: Psittaciformes) of the world. *Journal of Biogeography*, 31: 277–284.

Castro, S. A. and Figueroa, J. A. and Muñoz-Schick, M. and Jaksic, F. M. (2005). Minimum residence time, biogeographical origin, and life cycle as determinants of the geographical extent of naturalized plants in continental Chile. *Diversity and Distributions*, 11: 183–191.

Cheverud, J.M. and Dow, M.M. and Leutenegger, W. (1985). The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution*, 39: 1335–1351.

Chornesky, E. A. and Bartuska, A. M. and Aplet, G. H. and Britton, K. O. and Cummings-Carlson, J. and Davis, F. W. and Eskow, J. and Gordon, D. R. and Gottschalk, K. W. and Haack, R. A. and Hansen, A. J. and Mack, R. N. and Rahel, F. J. and Shannon, M. A. and Wainger, L. A. and Wigley, T. B. (2005). Science priorities for reducing the threat of invasive species to sustainable forestry. *BioScience*, 55: 335–348.

Clark, J. S. and Gelfand, A. E. (2006). *Heirarchical Modelling for the Environmental Science: Statistical Methods and Applications*. Oxford, UK: Oxford University Press.

Colautti, R. I. and Grigorovich, I. A. and MacIsaac, H. J. (2006). Propagule pressure: a null model for biological invasions. *Biological Invasions*, 8: 1023–1037.

Colautti, R. I. and Richardson, D. M. (2009). Subjectivity and flexibility in invasion terminology: too much of a good thing? *Biological Invasions*, 11: 1225–1229.

Corey, S. J. and Waite, T. A. (2008). Phylogenetic autocorrelation of extinction threat in globally imperilled amphibians. *Diversity and Distributions*, 14: 614–629.

Crosti, R. and Cascone, C. and Cipollaro, S. (2010). Use of a weed risk assessment for the Mediterranean region of central Italy to prevent loss of functionality and biodiversity in agro-ecosystems. *Biological Invasions*, 12: 1607–1616.

Daehler, C. C. (1998). The taxonomic distribution of invasive angiosperm plants: ecological insights and comparison to agricultural weeds. *Biological Conservation*, 84: 167–180.

Daehler, C. C. and Denslow, J. S. and Ansari, S. and Kuo, H. C. (2004). A risk-assessment system for screening out invasive pest plants from Hawaii and other Pacific Islands. *Conservation Biology*, 18: 360–368.

Damlen, P. and Wakefield, J. and Walker, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61: 331–344.

Davies, T. J. and Smith, G. F. and Bellstedt, D. U. and Boatwright, J. S. and Bytebier, B. and Cowling, R. M. and Forest, F. and Harmon, L. J. and Muasya, A. M. and Schrire, B. D. and Steenkamp, Y. and Van Der Bank, M. and Savolaninen, V. (2011). Extinction risk and diversification are linked in a plant biodiversity hotspot. *PLoS Biology*, 9: e1000620.

Davis, A. S. and Cousens, R. D. and Hill, J. and Mack, R. N. and Simberloff, D. and Raghu, S. (2010). Screening bioenergy feedstock crops to mitigate invasion risk. *Frontiers in Ecology and the Environment*, 8: 533–539.

Dawson, W. and Burslem, D. F. R. P. and Hulme, P. E. (2009a). Factors explaining alien plant invasion success in a tropical ecosystem differ at each stage of invasion. *Journal of Ecology*, 97: 657–665.

Dawson, W. and Burslem, D. F. R. P. and Hulme, P. E. (2009b). The suitability of weed risk assessment as a conservation tool to identify invasive plant threats in East African rainforests. *Biological Conservation*, 142: 1018–1024.

Dawson, W. and Burslem, D. F. R. P. and Hulme, P. E. (2011a). The comparative importance of species traits and introduction characteristics in tropical plant invasions. *Diversity and Distributions*, 17: 1111–1121.

Dawson, W. and Fischer, M. and van Kleunen, M. (2011b). The maximum relative growth rate of common UK plant species is positively associated with their global invasiveness. *Global Ecology and Biogeography*, 20: 299–306.

Desdevises, Y. and Legendre, P. and Azouzi, L. and Morand, S. (2003). Quantifying phylogenetically structured environmental variation. *Evolution*, 57: 2647–2652.

Diez, J. M. and Williams, P. A. and Randall, R. P. and Sullivan, J. J. and Hulme, P. E. and Duncan, R. P. (2009). Learning from failures: testing broad taxonomic hypotheses about plant naturalization. *Ecology Letters*, 12: 1174–1183.

Drummond, A. and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7: 214.

Drummond, A. J. and Ho, S. Y. W. and Phillips, M. J. and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4: e88.

Duncan, R. P. and Bomford, M. and Forsyth, D. M. and Conibear, L. (2001). High predictability in introduction outcomes and the geographical range size of introduced Australian birds: a role for climate. *Journal of Animal Ecology*, 70: 621–632.

Duncan, R. P. and Clemants, S. E. and Corlett, R. T. and Hahs, A. K. and McCarthy, M. A. and McDonnell, M. J. and Schwartz, M. W. and Thompson, K. and Vesk, P. A. and Williams, N. S. G. (2011). Plant traits and extinction in urban areas: a meta-analysis of 11 cities. *Global Ecology and Biogeography*, 20: 509–519.

Dunham, A. E. and Mikheyev, A. S. (2010). Influence of an invasive ant on grazing and detrital communities and nutrient fluxes in a tropical forest. *Diversity and Distributions*, 16: 33–42.

Dye, P. J. (1996). Climate, forest and streamflow relationships in South African afforested catchments. *Commonwealth Forestry Review*, 75: 31–387.

Earle, C. J. (2008). The Gymnosperm Database. `www.conifers.org`.

Elith, J. and Graham, C. H. and Anderson, R. T. and Dudík, M. and Ferrier, S. and Guisan, A. and Hijmans, R. J. and Huettmann, F. and Leathwick, J. R. and Lehmann, A. and Li, J. and Lohmann, L. G. and Loiselle, B. A. and Manion, G. and Moritz, C. and Nakamura, M. and Nakazawa, Y. and Overton, J. M. and Townsend Peterso, A. and

Phillips, S. J. and Richardson, K. and Scachetti-Pereira, R. and Schapire, R. E. and Soberón, J. and Williams, S. and Wisz, M. S. and Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29: 129–151.

Elith, J. and Leathwick, J. R. and Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77: 802–813.

Elton, C.S. (1958). *The Ecology of Invasions by Animals and Plants*. London, UK: Methuen.

Emmissions Trading Scheme Review Panel (2011). *Doings New Zealand's fair share. Emissions Trading Scheme review 2011 / Final report*. Technical report, Wellington: Ministry for the Environment.

ESRI (2008). ArcMap 9.3 1999-2008. Environmental Systems Research Institute, Inc.

Essl, F. and Mang, T. and Dullinger, S. and Moser, D. and Hulme, P. E. (2011). Macroecological drivers of alien conifer naturalizations worldwide. *Ecography*, 34: 1076–1084.

Essl, F. and Moser, D. and Dullinger, S. and Mang, T. and Hulme, P. E. (2010). Selection for commercial forestry determines global patterns of alien conifer invasions. *Diversity and Distributions*, 16: 911–921.

Farjon, A. (2005). *Drawings and Descriptions of the Genus* Pinus. Leiden, Netherlands: E. J. Brill and W. Backhuys.

Farris, M. A. and Lechowicz, M. J. (1990). Functional interactions among traits that determine reproductive success in a native annual plant. *Ecology*, 71: 548–557.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125: 1–15.

Fisher, D. C. (1991). *Analytical Paleobiology*, chapter Phylogenetic analysis and its application in evolutionary paleobiology. Paleontological Society: Knoxville, TN, USA.

Forey, P. L. and Fortey, R. A. and Kenrick, P. and Smith, A. B. (2004). Taxonomy and fossils: a critical appraisal. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359: 639–653.

Forsyth, D. M. and Duncan, R. P. and Bomford, M. and Moore, G. (2004). Climatic suitability, life-history traits, introduction effort, and the establishment and spread of introduced mammals in Australia. *Conservation Biology*, 18: 557–569.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29: 1189–1232.

Friedman, J. H. and Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, 22: 1365–1381.

Fritz, S. A. and Purvis, A. (2010). Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, 24: 1042–1051.

Gallagher, R. V. and Leishman, M. R. and Miller, J. T. and Hui, C. and Richardson, D. M. and Suda, J. and Trávníček, P. (2011). Invasiveness in introduced Australian acacias: the role of species traits and genome size. *Diversity and Distributions*, 17: 884–897.

García-Berthou, E. and Alcaraz, C. and Pou-Rovira, Q. and Zamora, L. and Coenders, G. and Feo, C. (2005). Introduction pathways and establishment rates of invasive aquatic species in europe. *Canadian Journal of Fisheries and Aquatic Sciences*, 62: 453–463.

Gassò, N. and Basnou, C. and Vilà, M. (2010). Predicting plant invaders in the Mediterranean through a weed risk assessment system. *Biological Invasions*, 12: 463–476.

GBIF (2011). Global Biodiversity Information Facility. `www.gbif.org`.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1: 515–533.

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27: 2865–2873.

Gelman, A. and Jakulin, A. and Pittau, M. G. and Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2: 1360–1383.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7: 457–472.

Gernandt, D. S. and López, G. G. and García, S. O. and Liston, A. (2005). Phylogeny and classification of *Pinus*. *Taxon*, 54: 29–42.

Gillett, J. M. and Taylor, N. L. (2001). *The Wold of Clovers*. 2121 State Avenue, Ames, IA, USA: Iowa State University Press.

Goodwin, B. J. and McAllister, A. J. and Fahrig, L. (1999). Predicting invasiveness of plant species based on biological information. *Conservation Biology*, 13: 422–426.

Gordon, D.R. and Onderdonk, D.A. and Fox, A.M. and Stocker, R.K. and Gantz, C. (2008a). Predicting invasive plants in Florida using the Australian Weed Risk Assessment. *Invasive Plant Science and Management*, 1: 178–195.

Gordon, D. R. and Mitterdorfer, B. and Pheloung, P. C. and Ansari, S. and Buddenhagen, C. and Chimera, C. and Daehler, C. C. and Dawson, W. and Denslow, J. S. and LaRosa, A. M. and Nishida, T. and Onderdonk. D. A. and Panetta, F. and Pyšek, P. and Randall, R. P. and Richardson, D. M . and Tshidada, N. J. and Virtue, J. G. and Williams, P. A. (2010). Guidance for addressing the Australian Weed Risk Assessment questions. *Plant Protection Quarterly*, 25: 56–74.

Gordon, D. R. and Onderdonk, D. A. and Fox, A. M. and Stocker, R. K. (2008b). Consistent accuracy of the Australian weed risk assessment system across varied geographies. *Diversity and Distributions*, 14: 234–242.

Grace, J. B. (2006). *Structural Equation Modeling and Natural Systems*. Cambridge, UK: Cambridge University Press.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 326: 119–157.

Graham, J. W. and Schafer, J. L. (1999). *Statistical Strategies for Small Sample Research*, chapter On the performance of multiple imputation for multivariate data with small sample size, (pp. 1–29). Thousand Oaks: CA, USA.

Gravuer, K. (2004). *Determinants of the introduction, naturalisation, and spread of Trifolium species in New Zealand*. M.Sc Thesis, Lincoln Univeristy, Lincoln 7647, Canterbury, New Zealand. `http://researcharchive.lincoln.ac.nz/dspace/bitstream/10182/340/1/gravuer_msc.pdf`.

Gravuer, K. and Sullivan, J. J. and Williams, P. A. and Duncan, R. P. (2008). Strong human association with plant invasion success for *Trifolium* introductions to New Zealand. *Proceedings of the National Academy of Sciences*, 105: 6344–6349.

Grime, J. P. and Thompson, K. and Hunt, R. and Hodgson, J. G. and Cornelissen, J. H. C. and Rorison, I. H. and Hendry, G. A. F. and Ashenden, T. W. and Askew, A. P. and Band, S. R. and Booth, R. E. and Bossard, C. C. and Campbell, B. D. and Cooper, J. E. L. and Davison, A. W. and Gupta, P. L.and Hall, W. and Hand, D. W. and Hannah, M. A. and Hillier, S. H. and Hodkinson, D. J. and Jalili, A. and Liu, Z. and Mackey, J. M. L. and Matthews, N. and Mowforth, M. A. and Neal, A. M. and Reader, R. J. and Reiling, K. and Ross-Fraser, W. and Spencer, R. E. and Sutton, F. and Tasker, D. E. and Thorpe, P. C. and Whitehouse, J. (1997). Integrated screening validates primary axes of specialisation in plants. *Oikos*, 79: 259–281.

Grotkopp, E. and Rejmánek, M. and Rost, T. L. (2002). Toward a causal explanation of plant invasiveness: seedling growth and life-history strategies of 29 pine (*Pinus*) species. *The American Naturalist*, 159: 396–419.

Guindon, S. and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52: 696–704.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11: 1–12.

Hadfield, J. (2010). MCMC Methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 32: 1–22.

Hadfield, J. (2012). Mcmcglmm course notes. `http://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf`.

Hadfield, J. D. and Nakagawa, S. (2010). General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, 23: 494–508.

Hanley, J. A. and Mcneil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143: 29–36.

Hanspach, J. and Kühn, I. and Pyšek, P. and Boos, E. and Klotz, S. (2008). Correlates of naturalization and occupancy of introduced ornamentals in Germany. *Perspectives in Plant Ecology, Evolution and Systematics*, 10: 241–250.

Harding, M. (2001). *South Island wilding conifer strategy*. Technical report, Deptment of Conservation, Wellington, New Zealand.

Harrison, S. and Grace, J. B. and Davies, K. F. and Safford, H. D. and Viers, J. H. (2006). Invasion in a diversity hotspot: exotic cover and native richness in the Californian serpentine flora. *Ecology*, 87: 695–703.

Harvey, P. H. and Pagel, M. D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford, UK: Oxford University Press.

Hayes, K. R. and Barry, S. C. (2008). Are there any consistent predictors of invasion success? *Biological Invasions*, 10: 483–506.

Hijmans, R. J. and Phillips, S. and Leathwick, J. R. and Elith, J. (2012). dismo: Species distribution modeling. R package version 0.7-17. `http://CRAN.R-project.org/package=dismo`.

Hoogland, J. J. and Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods and Research*, 26: 329–367.

Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*. Hoboken, NJ, USA: John Wiley and Sons, Inc.

Housworth, E. A. and Martins, E. P. and Lynch, M. (2004). The phylogenetic mixed model. *The American Naturalist*, 163: 84–96.

Howell, C. (2008). *Consolidated list of environmental weeds in New Zealand*. DOC Research & Development Series 292, Department of Conservation, Science & Technical Publishing, PO Box 10420, The Terrace, Wellington, New Zealand.

Huelsenbeck, J.P. and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17: 754–755.

Huelsenbeck, J. P. and Rannala, B. and Masly, J. P. (2000). Accommodating phylogenetic uncertainty in evolutionary studies. *Science*, 288: 2349–2350.

Hui, C. and Richardson, D. M. and Robertson, M. P. and Wilson, J. R. U. and Yates, C. J. (2011). Macroecology meets invasion ecology: linking the native distributions of Australian acacias to invasiveness. *Diversity and Distributions*, 17: 872–883.

Hulme, P.E. (2006). Beyond control: wider implications for the management of biological invasions. *Journal of Applied Ecology*, 43: 835–847.

Hulme, P.E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46: 10–18.

Hulme, P.E. (2011). Addressing the threat to biodiversity from botanic gardens. *Trends in Ecology & Evolution*, 26: 168–174.

Hulme, P. E. (2010). *Fifty Years of Invasion Ecology*, chapter Biosecurity: the changing face of invasion biology, (pp. 301–314). Wiley-Blackwell: Oxford, UK.

Hulme, P. E. (2012). Weed risk assessment: a way forward or a waste of time? *Journal of Applied Ecology*, 49: 10–19.

IPIN (2004). The International Plant Names Index. `www.ipni.org`.

Jefferson, L. and Havens, K. and Ault, J. (2004). Implementing invasive screening procedures: the Chicago Botanic Garden model. *Weed Technology*, 18: 1434–1440.

Jeschke, J. M. and Strayer, D. L. (2006). Determinants of vertebrate invasion success in Europe and North America. *Global Change Biology*, 12: 1608–1619.

Kato, H. and Hata, K. and Yamamoto, H. and Yoshioka, T. (2006). *Assessment and Control of Biological Invasion Risks*, chapter Effectiveness of the weed risk assessment system for the Bonin Islands, (pp. 65–72). Shoukadoh Book Sellers: Kyoto, Japan and IUCN, Gland, Switzerland.

Keeley, J. E. and Baer-Keeley, M. and Fotheringham, C. J. (2005). Alien plant dynamics following fire in Mediterranean-climate California shrublands. *Ecological Applications*, 15: 2109–2125.

Keller, R. P. and Kocev, D. and Džeroski, S. (2011). Trait-based risk assessment for invasive species: high performance across diverse taxonomic groups, geographic ranges and machine learning/statistical tools. *Diversity and Distributions*, 17: 451–461.

Koh, L. P. and Levang, P. and Ghazoul, J. (2009). Designer landscapes for sustainable biofuels. *Trends in Ecology & Evolution*, 24: 431–438.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: The MIT Press.

Küster, E. C. and Kühn, I. and Bruelheide, H. and Klotz, S. (2008). Trait interactions help explain plant invasion success in the German flora. *Journal of Ecology*, 96: 860–868.

Křivánek, M. and Pyšek, P. (2006). Predicting invasions by woody species in a temperate zone: a test of three risk assessment schemes in the Czech Republic (Central Europe). *Diversity and Distributions*, 12: 319–327.

Křivánek, M. and Pyšek, P. (2008). *Plant Invasions: Human Perception, Ecological Impacts and Management*, chapter Forestry and horticulture as pathways of plant invasions: a database of alien woody plants in the Czech Republic, (pp. 21–38). Leiden, The Netherlands: Backhuys Publishers.

Křivánek, M. and Pyšek, P. and Jarošík, V. (2006). Planting history and propagule pressure as predictors of invasion by woody species in a temperate region. *Conservation Biology*, 20: 1487–1498.

Lambdon, P. W. and Lloret, F. and Hulme, P. E. (2008). How do introduction characteristics influence the invasion success of Mediterranean alien plants? *Perspectives in Plant Ecology, Evolution and Systematics*, 10: 143–159.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33: 159–174.

Ledgard, N. (2001). The spread of lodgepole pine (*Pinus contorta*, dougl.) in New Zealand. *Forest Ecology and Management*, 141: 43–57.

Ledgard, N. J. and Langer, E. R. (1999). Wilding prevention – guidelines for minimising the risk of unwanted spread from new plantings of introduced conifers. Forest Research, Box 29237, Fendalton, Christchurch.

Ledig, F. T. (1998). *Ecology and Biogeographyof* Pinus, chapter Genetic variation in *Pinus*, (pp. 251–280). Cambridge University Press.

Lee, S. Y. (2007). *Structural Equation Eodelling: a Bayesian Approach*. New York, NY, USA: John Wiley & Sons Ltd.

Ley, E. and Steel, M . F. J. (2007). *On the effect of prior assumptions in Bayesian model averaging with applications to growth regression*. Technical report, World Bank Policy Research Paper No. 4238, World Bank, Washington, D.C.

Link, W. A. and Barker, R. J. (2006). Model weights and the foundations of multimodel inference. *Ecology*, 87: 2626–2635.

Lloret, F. and Médail, F. and Brundu, G. and Camarda, I. and Moragues, E. and Rita, J. and Lambdon, P. and Hulme, P. E. (2005). Species attributes and invasion success by alien plants in Mediterranean islands. *Journal of Ecology*, 93: 512–520.

Lockwood, J. L. and Cassey, P. and Blackburn, T. (2005). The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution*, 20: 223–228.

Low, T. and Booth, C. and Sheppard, A. (2011). Weedy biofuels: what can be done? *Current Opinion in Environmental Sustainability*, 3: 55–59.

Lutzoni, F. and Pagel, M. and Reeb, V. (2001). Major fungal lineages are derived from lichen symbiotic ancestors. *Nature*, 411: 937–940.

MAF (2011). *National Exotic Forest Description as at 1 April 2011*. `http://www.maf.govt.nz/news-resources/statistics-forecasting/statistical-publications`. Technical Report Edition 28, Ministry of Agriculture and Forestry, Pastoral House, 25 The Terrace, PO Box 2526, Wellington 6140, New Zealand.

Martins, E.P. and Hansen, T.F. (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist*, 149: 646–667.

Martins, E.P. and Hasnes, T.F. (1996). *Phylogenies and the Comparative Method in Animal Behaviour*, chapter The statistical analysis of interspecific data: a review and evaluation of phylogenetic comparative methods. Oxford University Press: Oxford, UK.

Martins, E. P. and Garland Jr, T. (1991). Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution*, 45: 534–557.

McGregor, K. F. and Watt, M. S. and Hulme, P. E. and Duncan, R. P. (2012). How robust is the Australian weed risk assessment protocol? A test using pine invasions in the northern and southern hemispheres. *Biological Invasions*, 14: 987–998.

Medawatte, W. W. M. A. B. and Tennakoon, K. U. and Hulme, P. E. and Gunatilleke, C. V. S. and Gunatilleke, I. A. U. N. (2010). *Invasive Alien Species–Strengthening Capacity to Control Introduction and Spread in Sri Lanka*, chapter Is Caribbean pine invading grasslands in the Knuckles Range?, (pp. 131–140). Biodiversity Secretariat of the Ministry of Environment: Sri Lanka.

Meyer, D. and Zeilesis, A. and Hornik, K. (2010). vcd: Visualising Catagorical Data. R package version 1.2-8. http://cran.r-project.org/web/packages/vcd/index.html.

Miles, D. B. and Dunham, A. E. (1993). Historical perspectives in ecology and evolutionary biology: the use of phylogenetic comparative analyses. *Annual Review of Ecology, Evolution and Systematics*, 24: 587–619.

Miller, J. T. and Murphy, D. J. and Brown, G. K. and Richardson, D. M. and González-Orozco, C. E. (2011). The evolution and phylogenetic placement of invasive Australian *Acacia* species. *Diversity and Distributions*, 17: 848–860.

Ministry for the Environment (2010). *Proposed National Environmental Standards for Plantation Forestry: Discussion Document. 2010.* http://www.mfe.govt.nz/publications/rma/proposed-nes-plantation-forestry/proposed-nes-plantation-forestry.pdf. Technical report, Ministry for the Environment, Manatü Mö Te Taiao, PO Box 10362, Wellington 6143, New Zealand.

Mitchell, C. E. and Agrawal, A. A. and Bever, J. D. and Gilbert, G. S. and Hufbauer, R. A. and Klironomos, J. N. and Maron, J. L. and Morris, W. F. and Parker, I. M. and Power, A. G. and Seabloom, E. W. and Torchin, M. E. and Vázquez, D. P. (2006). Biotic interactions and plant invasions. *Ecology Letters*, 9: 726–740.

Mitchell, R. J. (1992). Testing evolutionary and ecological hypotheses using path analysis and structural equation modelling. *Functional Ecology*, 6: 123–129.

Moore, R.A. (2005). *Alien invasive species: impacts on forests and forestry*. Technical report, Food and Agriculture Organization of the United Nations, Forestry Depatrment, Forest Resources Division, Forest Resources Development Service, Working Paper FBS/8E.

Mortenson, S. G. and Mack, R. N. (2006). The fate of alien conifers in long-term plantings in the USA. *Diversity and Distributions*, 12: 456–466.

Münkemüller, T. and Lavergne, S. and Bzeznik, B. and Dray, S. and Jombart, T. and Schiffers, K. and Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, DOI: 10.1111/j.2041-210X.2012.00196.x.

Mysterud, A. and Yoccoz, N. G. and Langvatn, R. and Pettorelli, N. and Stenseth, N. C. (2008). Hierarchical path analysis of deer responses to direct and indirect effects of climate in northern forest. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363: 2357–2366.

Nakagawa, S. and Freckleton, R. (2011). Model averaging, missing data and multiple imputation: a case study for behavioural ecology. *Behavioral Ecology and Sociobiology*, 65: 103–116.

Neapolitan, R. E. (2004). *Learning Bayesian Networks*. Upper Saddle River, NJ, USA: Pearson Prentice Hall.

New, M. and Lister, D. and Hulme, M. and Makin, I. (2002). A high-resolution data set of surface climate over global land areas. *Climate Research*, 21: 1–25.

Nishida, T. and Yamashita, N. and Asai, M. and Kurokawa, S. and Enomoto, T. and Pheloung, P. C. and Groves, R. H. (2009). Developing a pre-entry weed risk assessment system for use in Japan. *Biological Invasions*, 11: 1319–1333.

Nuñez, M. A. and Medley, K. A. (2011). Pine invasions: climate predicts invasion success; something else predicts failure. *Diversity and Distributions*, 17: 703–713.

Nuñez, M. A. and Moretti, A. and Simberloff, D. (2011). Propagule pressure hypothesis not supported by an 80-year experiment on woody species invasions. *Oikos*, 120: 1311–1316.

Onderdonk, D. A. and Gordon, D. R. and Fox, A. M. and Stocker, R. K. (2010). Lessons learned from testing the Australian weed risk assessment system: the devil is in the details. *Plant Protection Quarterly*, 25: 79–85.

Orme, D. and Freckleton, R. and Thomas, G. and Petzolt, T. and Fritz, S. and Isaac, N. (2011). caper: Comparative Analysis of Phylogenetics and Evolution in R. R package version 0.4. `http://CRAN.R-project.org.package=caper`.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401: 877–884.

Pagel, M. and Lutzoni, F. (2002). *Biological Evolution and Statistical Physics*, chapter Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation, (pp. 148–161). Springer-Verlag: Berlin, Germany.

Paradis, R. and Julien, C. and Korbinian, S. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20: 289–290.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.

Perrings, C. and Dehnen-Schmutz, K. and Touza, J. and Williamson, M. (2005). How to manage biological invasions under globalization. *Trends in Ecology & Evolution*, 20: 212–215.

Perry, J. P. (1991). *The Pines of Mexico and Central America*. Portland, OR, USA: Timber Press.

Pheloung, P. C. and Williams, P. A. and Halloy, S. R. (1999). A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. *Journal of Environmental Management*, 57: 239–251.

Pilgrim, E. S. and Crawley, M. J. and Dolphin, K. (2004). Patterns of rarity in the native British flora. *Biological Conservation*, 120: 161–170.

Plummer, M. and Best, N. and Cowles, K. and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News*, 6: 7–11.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, 25: 1253–1256.

Posada, D. and Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, 53: 793–808.

Preston, C. D. and Pearman, D. A. and Dines, T. D. (2002). *New Atlas of the British and Irish flora: an Atlas of the Vascular Plants of Britain, Ireland, the Isle of Man, and the Channel Islands*. Oxford, UK: Oxford University Press.

Price, R. A. and Liston, A. and Strauss, S. H. (1998). *Ecology and Biogeographyof* Pinus, chapter Phylogeny and systematics of *Pinus*, (pp. 49–68). Cambridge University Press.

Procheş, Ş. and Wilson, J. R. U. and Richardson, D. M. and Rejmánek, M. (2012). Native and naturalized range size in *Pinus*: relative importance of biogeography, introduction effort and species traits. *Global Ecology and Biogeography*, 21: 513–523.

Purvis, A. (2008). Phylogenetic approaches to the study of extinction. *Annual Review of Ecology, Evolution and Systematics*, 39: 301–319.

Purvis, A. and Agapow, P. M. and Gittleman, J. L. and Mace, G. M. (2000). Nonrandom extinction and the loss of evolutionary history. *Science*, 288: 328–330.

Puth, L. M. and Post, D. M. (2005). Studying invasion: have we missed the boat? *Ecology Letters*, 8: 715–721.

Pyšek, P. (1998). Is there a taxonomic pattern to plant invasions? *Oikos*, 82: 282–294.

Pyšek, P. and Jarošík, V. and Hulme, P. E. and Kühn, I. and Wild, J. and Arianoutsou, M. and Bacher, S. and Chiron, F. and Didžiulis, V. and Essl, F. and Genovesi, P. and Gherardi, F. and Hejda, M. and Kark, S. and Lambdon, P.W. and Desprez-Loustau, M. and Nentwig, W. and Pergl, J. and Poboljšaj, K. and Rabitsch, W. and Roques, A. and Roy, D. B. and Shirley, A. and Solarz, W. and Vilà, M. and Winterd, M. (2010). Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proceedings of the National Academy of Sciences*, 107: 12157–12162.

Pyšek, P. and Jarošík, V. and Pergl, J. and Randall, R. and Chytrỳ, M. and Kühn, I. and Tichỳ, L. and Danihelka, J. and Sádlo, J. (2009a). The global invasion success of Central European plants is related to distribution characteristics in their native range and species traits. *Diversity and Distributions*, 15: 891–903.

Pyšek, P. and Křivánek, M. and Jarošík, V. (2009b). Planting intensity, residence time, and species traits determine invasion success of alien woody species. *Ecology*, 90: 2734–2744.

R Development Core Team (2010). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.r-project.org.

R Development Core Team (2011). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.r-project.org.

Raghu, S. and Anderson, R. C. and Daehler, C. C. and Davis, A. S. and Wiedenmann, R. N. and Simberloff, D. and Mack, R. N. (2006). Adding biofuels to the invasive species fire? *Science*, 313: 1742–1742.

Rambaut, A. and Drummond, A. J. (2007). Tracer v1.4. http://beast.bio.ed.ac.uk/Tracer.

Reichard, S. H. and Hamilton, C. W. (1997). Predicting invasions of woody plants introduced into North America. *Conservation Biology*, 11: 193–203.

Rejmánek, M. and Richardson, D. M. (1996). What attributes make some plant species more invasive? *Ecology*, 77: 1655–1661.

Rejmànek, M. and Richardson, D. M. (2011). *Encyclopedia of Biological Invasions*, chapter Eucalypts, (pp. 203–209). University of California Press: Berkeley, USA.

Revell, L. J. (2011). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3: 217–223.

Richardson, D.M. (1998a). Forestry trees as invasive aliens. *Conservation Biology*, 12: 18–26.

Richardson, D.M. and Blanchard, R. (2011). Learning from our mistakes: minimizing problems with invasive biofuel plants. *Current Opinion in Environmental Sustainability*, 3: 36 –42.

Richardson, D.M. and Bond, W.J. (1991). Determinants of plant distribution: evidence from pine invasions. *The American Naturalist*, 137: 639–668.

Richardson, D.M. and Higgins, S.I. (1998). *The Ecology and Biogeography of* Pinus, chapter Pines as invaders in the southern hemisphere, (pp. 450–473). Cambridge University Press.

Richardson, D.M. and Rejmánek, M. (2004). Conifers as invasive aliens: a global survey and predictive framework. *Diversity and Distributions*, 10: 321–331.

Richardson, D. M. (1998b). *Ecology and Biogeography of* Pinus. Cambridge, UK: Cambridge University Press.

Richardson, D. M. (2006). *Pinus*: a model group for unlocking the secrets of alien plant invasions? *Preslia*, 78: 375–388.

Richardson, D. M. and Allsopp, N. and D'Antonio, C. M. and Milton, S. J. and Rejmánek, M. (2000a). Plant invasions—the role of mutualisms. *Biological Reviews*, 75: 65–93.

Richardson, D. M. and Carruthers, J. and Hui, C. and Impson, F. A. C. and Miller, J. T. and Robertson, M. P. and Rouget, M. and Le Roux, J. J. and Wilson, J. R. U. (2011). Human-mediated introductions of Australian acacias–a global experiment in biogeography. *Diversity and Distributions*, 17: 771–787.

Richardson, D. M. and Pyšek, P. and Rejmánek, M. and Barbour, M. G. and Panetta, F. D. and West, C. J. (2000b). Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distributions*, 6: 93–107.

Richardson, D. M. and Rejmánek, M. (2011). Trees and shrubs as invasive alien species–a global review. *Diversity and Distributions*, 17: 788–809.

Richardson, D. M. and Rundel, P. W. (1998). *Ecology and Biogeographyof* Pinus, chapter Ecology and Biolgeography of *Pinus*: an introduction, (pp. 3–46). Cambridge University Press.

Richardson, D. M. and van Wilgen, B. and Nuñez, M. (2008). Alien conifer invasions in South America: short fuse burning? *Biological Invasions*, 10: 573–577.

Richardson, D. M. and Williams, P. A . and Hobbs, R. J. (1994). Pine invasions in the southern hemisphere: determinants of spread and invadability. *Journal of Biogeography*, 21: 511–527.

Roberts, W. and Harrod, O. and Mitterdorfer, B. and Pheloung, P. C. (2011). Regulating invasive plants and use of weed risk assessments. *Current Opinion in Environmental Sustainability*, 3: 60–65.

Rowlingson, B. and Diggle, P and adapted and packaged for R by Roger Bivand, pcp functions by Giovanni Petris, goodness of fit by Stephen Eglen (2010). splancs: Spatial and space-time point pattern analysis. R package version 2.01-27. `http://CRAN.R-project.org/package=splancs`.

Rubin, D. (1976). Inference and missing data. *Biometrika*, 63: 581–592.

Rupp, A. A. and Dey, D. K. and Zumbo, B. D. (2004). To Bayes or not to Bayes, from whether to when: applications of Bayesian methodology to modeling. *Structural Equation Modeling*, 11: 424–451.

Russell, G. J. and Brooks, T. M. and McKinney, M. M. and Anderson, C. G. (1998). Present and future taxonomic selectivity in bird and mammal extinctions. *Conservation Biology*, 12: 1365–1376.

Satorra, A. and Bentler, P. (1999). *Scaling corrections for statistics in covariance structure analysis*. Technical report, Department of Statistics Papers, Department of Statistics, UCLA, UC Lost Andeles.

Schafer, O. B. J. L. (2010). mix: Estimation/multiple imputation for mixed categorical and continuous data. `http://cran.r-project.org/web/packages/mix/mix.pdf`.

Scheines, R. and Hoijtink, H. and Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, 64: 37–52.

Schemske, D. W. and Horvitz, C. C. (1988). Plant-animal interactions and fruit production in a neotropical herb: a path analysis. *Ecology*, 69: 1128–1137.

Schmidt, J. P. and Drake, J. M. (2011a). Time since introduction, seed mass, and genome size predict successful invaders among the cultivated vascular plants of Hawaii. *PLoS ONE*, 6: e17391.

Schmidt, J. P. and Drake, J. M (2011b). Why are some plant genera more invasive than others. *PLoS ONE*, 6: e18654.

Schwartz, M. W. and Simberloff, D. (2001). Taxon size predicts rates of rarity in vascular plants. *Ecology Letters*, 4: 464–469.

Scott, J. K. and Panetta, F. D. (1993). Predicting the Australian weed status of southern African plants. *Journal of Biogeography*, 20: 87–93.

Seabloom, E. W. and Williams, J. W. and Slayback, D. and Stoms, D. M. and Viers, J. H. and Dobson, A. P. (2006). Human impacts, plant invasion, and imperiled plant species in California. *Ecological Applications*, 16: 1338–1350.

Shipley, B. (1997). Exploratory path analysis with applications in ecology and evolution. *The American Naturalist*, 149: 1113–1138.

Shipley, B. (2000). *Cause and Correlation in Biology: a User's Guide to Path Analysis, Structural Equations and Causal Inference*. Cambridge, UK: Cambridge University Press.

Shipley, B. (2009). Confirmatory path analysis in a generalized multilevel context. *Ecology*, 90: 363–368.

Sikes, B. A. and Powell, J. R. and Rillig, M. C. (2010). Deciphering the relative contributions of multiple functions within plant-microbe symbioses. *Ecology*, 91: 1591–1597.

Simberloff, D. and Nuñez, M. A. and Ledgard, N. J. and Pauchard, A. and Richardson, D. M. and Sarasola, M. and van Wilgen, B. W. and Zalba, S. M. and Zenni, R. D. and Bustamante, R. and Peña, E. and Ziller, S. R. (2010). Spread and impact of introduced conifers in South America: lessons from other southern hemisphere regions. *Austral Ecology*, 35: 489–504.

Sing, T. and Sander, O. and Beerenwinkel, N. and Lengauer, T. (2009). ROCR: Visualising the performance of scoring classifiers. R package version 1.0-4. `http://CRAN.R-project.org/package=ROCR`.

Spiegelhalter, D.J. and Dawid, A.P. and Lauritzen, S.L. and Cowell, R.G. (1993). Bayesian analysis in expert systems. *Statistical Science*, 8: 219–247.

Spiegelhalter, D. J. and Best, N. G. and Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64: 583–639.

Stuart, S. N. and Chanson, J. S. and Cox, N. A. and Young, B. E. and Rodrigues, A. S. L. and Fischman, D. L. and Waller, R. W. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science*, 306: 1783–1786.

Tamura, K. and Peterson, D. and Peterson, N. and Stecher, G. and Nei, M. and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28: 2731–2739.

Taylor, B. W. and Irwin, R. E. (2004). Linking economic activities to the distribution of exotic plants. *Proceedings of the National Academy of Sciences*, 101: 17725–17730.

Thomas, A. and O'Hara, B. and Ligges, U. and Sturtz, S. (2006). Making BUGS open. *R News*, 6: 12–17.

Thomas, D. W. and Shipley, B. and Blondel, J. and Perret, P. and Simon, A. and Lambrechts, M. M. (2007). Common paths link food abundance and ectoparasite loads to physiological performance and recruitment in nestling blue tits. *Functional Ecology*, 21: 947–955.

Thomson, F. J. and Moles, A. T. and Auld, T. D. and Kingsford, R. T. (2011). Seed dispersal distance is more strongly correlated with plant height than with seed mass. *Journal of Ecology*, 99: 1299–1307.

Tingley, R. and Romagosa, C. M. and Kraus, F. and Bickford, D. and Phillips, B. L. and Shine, R. (2010). The frog filter: amphibian introduction bias driven by taxonomy, body size and biogeography. *Global Ecology and Biogeography*, 19: 496–503.

United Nations (1998). Kyoto protocol to the United Nations framework convention on climate change. `http://unfccc.int/resource/docs/convkp/kpeng.pdf`.

USDA (2010). Fire effects information system. `www.fs.fed.us/database/feis`.

USDA (2011). The plants database. `http://plants.usda.gov/java`.

Vamosi, J.C. and Wilson, J. R. U. (2008). Nonrandom extinction leads to elevated loss of angiosperm evolutionary history. *Ecology Letters*, 11: 1047–1053.

Vázquez, D. P. and Simberloff, D. (2001). *Biotic Homogenization*, chapter Taxonomic selectivity in surviving introduced insects in the United States, (pp. 103–124). Kluwer Academic/Plenum: New York, USA.

Vilà, M. and Muñoz, I. (1999). Patterns and correlates of exotic and endemic plant taxa in the Balearic Islands. *Ecologia Mediterranea*, 25: 153–161.

Walther, G. R. and Roques, A. and Hulme, P. E. and Sykes, M. T. and Pyšek, P. and Kuhn, I. and Zobel, M. and Bacher, S. and Botta-Dukát, Z. and Bugmann, H. and Czúcz, B. and Dauber, J. and Hickler, T. and Jarošík, V. and Kenis, M. and Klotz, S. and Minchin, D. and Moora, M. and Nentwig, W. and Ott, J. and Panov, V. E. and Reineking, B. and Robinet, C. and Semenchenko, V. and Solarz, W. and Thuiller, W. and Vilà, M. and Vohland, K. and Settele, J. (2009). Alien species in a warmer world: risks and opportunities. *Trends in Ecology & Evolution*, 24: 686–693.

Webb, C. J. and Sykes, W. R. and Garnock-Jones, P. J. (1988). *Flora of New Zealand, Volume IV, Naturalised Pteridophytes, Gymnosperms, Dicotyledons*. Christchurch, NZ: Botany Division, DSIR.

Weber, J. and Panetta, D. F. and Virtue, J. G. and Pheloung, P. C. (2009). An analysis of assessment outcomes from eight years' operation of the Australian border weed risk assessment system. *Journal of Environmental Management*, 90: 798–807.

Wilcove, D. S. and Rothstein, D. and Dubow, J. and Phillips, A. and Losos, E. (1998). Quantifying threats to imperiled species in the United States. *BioScience*, 48: 607–615.

Williams, P. A. and Nicol, E. and Newfield, M. (2000). *Assessing the Risk to Indigenous Biota from New Exotic Plant Taxa and Genetic Material*. Wellington, New Zealand: Science for Conservation 143, Department of Conservation, New Zealand.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20: 557–585.

Yokomizo, H. and Possingham, H. and Hulme, P. E. and Grice, A. and Buckley, Y. M. (2012). Cost-benefit analysis for intentional plant introductions under uncertainty. *Biological Invasions*, 14: 839–849.

Zavaleta, E. (2009). The economic value of controlling an invasive shrub. *AMBIO: A Journal of the Human Environment*, 29: 462–467.