# Strong human association with plant invasion success for *Trifolium* introductions to New Zealand

**Kelly Gravuer*†, Jon J. Sullivan*, Peter A. Williams‡, and Richard P. Duncan***

*National Centre for Advanced Bio-Protection Technologies and Bio-Protection and Ecology Division, P.O. Box 84, Lincoln University, Lincoln 7647, New Zealand; and ‡Landcare Research, Private Bag 6, Nelson 7042, New Zealand

Edited by James H. Brown, University of New Mexico, Albuquerque, NM, and approved March 11, 2008 (received for review December 20, 2007)

**It has proven difficult, when focused only on biological determinants, to explain why some plant species become naturalized in or invade new locations, whereas others fail. We analyzed the invasion of *Trifolium* (true clover) species into New Zealand, assessing a range of human, biogeographic, and biological influences at three key invasion stages: introduction, naturalization, and spread. We used sparse principal component analysis (SPCA) to define suites of related attributes and aggregated boosted trees to model relationships with invasion outcomes. Human and biogeographic attributes were strongly associated with success at all stages. Whereas biogeographic attributes, notably large native range, were consistently associated with success, different human factors appeared to favor success at different stages, such as presence in early trade/immigration hotspots (introduction), intentional large-scale planting (naturalization), and frequent presence as a seed contaminant (relative spread rate). Biological traits were less strongly associated with success for introduction and spread and little if at all for naturalization; we found that tall perennials with long flowering periods were more frequently selected for introduction, whereas species with extended flowering in New Zealand spread more rapidly. In addition to causal relationships, the importance of human factors may reflect indirect associations, including ecological traits associated with both human use and invasion. Nevertheless, our results highlight key roles that humans can play in facilitating plant invasion via two pathways: (*i*) commercial introduction leading to widespread planting and concomitant naturalization and spread and (*ii*) unintentional introduction and spread of species associated with human activities, such as seed contaminants.**

alien | naturalization | spread | transition | stage

Because some nonnative species have detrimental economic and ecological impacts, a major research goal is to identify potentially harmful species before their introduction or spread (1, 2). Previous studies of why some introduced species are more widespread or problematic than others have focused on biological traits (e.g., ref. 3) and interactions with recipient environments (e.g., ref. 4). These approaches have provided useful insights, but the resulting models typically leave much variation unexplained (5).

Recent studies have emphasized that human influences should be explicitly incorporated into invasion models (6–8). For key processes such as dispersal, anthropogenic variation, because of human preference for transporting and cultivating particular species (9), can be orders of magnitude greater than natural variation. Other species, such as agricultural seed contaminants, can benefit from human activities through unintended transport (10). Given the scale of human influences, it may not be surprising that models based on biological attributes alone can leave much variation unexplained. Such models may even erroneously identify biological traits as direct causes of invasion success when these traits are instead associated with human activities (8). Consequently, understanding the role of human influences in invasions, and how these influences interact with invaders' biogeographic and biological attributes, could substantially improve our understanding of the invasion process (6, 11).

Invasions can be understood as a series of stages (including transport/introduction, naturalization, spread, and integration/impact), with different factors important in determining success at each stage (12–14). This stage-based approach has improved our understanding of animal invasions (15–18), but its application to plants has been hindered by a lack of comprehensive data on which species have succeeded or failed at each stage.

Here, we use a stage-based approach to examine the relationship of human, biogeographic, and biological attributes to the introduction, naturalization and spread of *Trifolium* (true clover) species in New Zealand. We chose this study system for three reasons. First, *Trifolium* plays a vital role in New Zealand's pastoral economy (e.g., ref. 19) and, because New Zealand has no native *Trifolium*, many *Trifolium* species were introduced. There is a remarkably detailed record of these introductions and their outcome, allowing quantitative comparisons of success at key invasion stages (Fig. 1). Second, because of their global economic importance, a wealth of biological, biogeographic, and human-use data on *Trifolium* species could be obtained. Finally, because congeneric species share many attributes, focus on a single genus can more precisely identify the factors responsible for differences in behavior (20).

## Results

To identify factors associated with success at each stage (Fig. 1), we considered a range of human, biogeographic, and biological variables. To reduce multicollinearity issues, we assembled variables into sparse principal components (PCs), with one set of PCs for each invasion stage, and analyzed the relationship of sparse PCs to invasion success by using aggregated boosted trees (ABT) (21) [Tables 1 and 2 and supporting information (SI) Figs. S1–S4]. See SI for complete variable listing (Table S1) and full set of PCs for each stage (Tables S2, S5, and S7). The sparse PC analysis method improves PC interpretability at the expense of optimality (22); sparse PCs can be moderately correlated. We discuss the strongest relationships between PCs and success at each invasion stage below; PC correlations are considered in Tables S3, S4, S6, and S8. We also constructed ABT models using the raw variables for comparison (Tables S9 and S10). We report ABT prediction error as misclassification rate for binary outcomes (introduction and naturalization) and as percentage of variance unexplained for the continuous outcome (relative

---

Author contributions: K.G., P.A.W., and R.P.D. designed research; K.G. and J.J.S. performed research; K.G., J.J.S., and R.P.D. analyzed data; and K.G., J.J.S., and R.P.D. wrote the paper.

This article is a PNAS Direct Submission.

†To whom correspondence should be sent at the present address: NatureServe, 1101 Wilson Boulevard 15th floor, Arlington, VA 22209. E-mail: Kelly_Gravuer@natureserve.org.

This article contains supporting information online at www.pnas.org/cgi/content/full/0712026105/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.0712026105

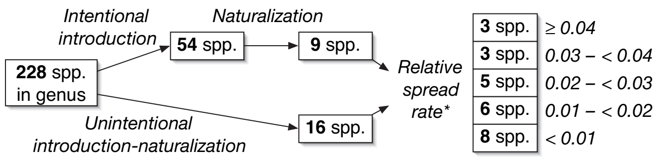

**Fig. 1.** Stages in the invasion process, with the number of *Trifolium* species successful at each stage in New Zealand. Intentional introduction: reported at least once as intentionally planted outdoors in a location suitable for reproduction. Naturalization: intentionally introduced species observed at least once (and usually frequently) in the reproductive state; circumstances of record suggest no direct intervention (i.e., cultivation) by humans. Unintentional introduction-naturalization: observed at least once in the reproductive state, with either no record of intentional introduction or before any record of intentional introduction; circumstances of record suggest no direct intervention by humans. Relative spread rate: for species naturalized after intentional or unintentional introduction, the relative rate of cumulative occupancy of the landscape, 1840–2000. *, units represent relative cumulative occupancy per year of 40-km × 30-km grid cells comprising New Zealand. Units are relative rather than absolute because of a resampling procedure used to standardize for variation in collection effort over time; see *Materials and Methods* for details.

spread rate). Prediction errors were derived by using cross-validation to avoid overly optimistic estimates (21).

The best ABT model of intentional introduction included main effects of three PCs (Table 1 and Fig. S1) and had a prediction error of 17.7% (40 of 228 species misclassified). From the global pool of *Trifolium* species, those intentionally introduced to New Zealand were characterized by large native ranges, tolerance of diverse conditions (many biomes in native range), presence in climates well matched to New Zealand, and extensive opportunity for transport and use by humans, including presence in Britain, a Mediterranean center of origin, and global economic use (PC1) (Table 1). To a lesser extent (lower PC relative importance), intentionally introduced species were tall with an erect habit and a long flowering period, which was also associated with global economic use (PC4), and were characterized by "perennial" traits, including perennial life span, inability to self-pollinate, long corollas, vegetative reproduction, and polyploidy (PC3) (Table 1).

The best ABT model of naturalization of these intentionally introduced species included just one PC (Table 1 and Fig. S2). This model had a prediction error of 1.9% (1 of 54 species misclassified) and suggested that intentionally introduced *Trifolium* that have naturalized were British species that were introduced early to New Zealand. These were generally species used for commercial agriculture, rather than exclusively for experimental plots or horticulture (Table 1).

The best ABT model of unintentional introduction–naturalization also included just one PC (Table 1 and Fig. S3), the same PC having the greatest relative importance for inten-

**Table 1. Sparse PCs retained in the best ABT models for introduction (Intro.) and naturalization (Nat.) of intentionally introduced species and for combined introduction and naturalization (Intro.–Nat.) of unintentionally introduced species**

| | Intentional | | | | Unintentional |
|---|---|---|---|---|---|
| | Intro. | | | Nat. | Intro.–Nat. |
| Explanatory variable | PC1$^i$ | PC4$^i$ | PC3$^i$ | PC1$^n$ | PC1$^i$ |
| Center of origin | −**0.261** | 0.118 | | | −**0.261** |
| Present in Britain | −**0.393** | 0.067 | | −**0.477** | −**0.393** |
| Global economic use | −**0.249** | −**0.292** | −0.035 | −0.169 | −**0.249** |
| Introduction date | n.a. | n.a. | n.a. | **0.637** | n.a. |
| Commercial agriculture use | n.a. | n.a. | n.a. | −**0.528** | n.a. |
| Compatible *Rhizobium* | n.a. | n.a. | n.a. | −0.014 | n.a. |
| Native-range area | −**0.407** | | | −0.161 | −**0.407** |
| Conditions tolerated | −**0.352** | −0.118 | | | −**0.352** |
| New Zealand climate match | −**0.641** | 0.001 | | | −**0.641** |
| Urban/rural habitats | −0.100 | | | −0.049 | −0.100 |
| Elevation of native range | | | −0.143 | | |
| Native canopy cover | | | | −0.124 | |
| Life span | | | −**0.429** | | |
| Vegetative reproduction | | | −**0.346** | | |
| Self-pollination capable | | | −**0.593** | | |
| Corolla length | | −0.190 | −**0.300** | 0.119 | |
| Seed size | | | | 0.020 | |
| Long-distance dispersal | | 0.103 | | | |
| Seed shape like *T. repens* | | 0.033 | −**0.371** | | |
| Polyploidy | 0.008 | | −**0.233** | | 0.008 |
| Reduced chromosome number | | | −0.142 | | |
| Maximum height | | −**0.728** | | | |
| Habit | 0.047 | −**0.410** | 0.144 | 0.024 | 0.047 |
| Flowering time (native) | | | | | |
| Flowering length (native) | −0.048 | −**0.369** | | | −0.048 |
| PC relative importance, % (effect direction) | 58.4 (−) | 22.4 (−) | 19.2 (−) | 100.0 (−) | 100.0 (−) |

PC loadings ≥|0.200| are (arbitrarily) bolded for emphasis. Blank cells indicate loadings ≈0. PC superscripts denote separate PC analyses for introduction (i) (Table S2) and naturalization (n) (Table S5), respectively (note that the same set of introduction PCs was used in both the intentional introduction and unintentional introduction–naturalization analyses). n.a. indicates variables that are not applicable to the introduction stage (intentional or unintentional); these variables were not included in the introduction PC analysis. For an individual tree created in an ABT model, a PC's relative importance is the sum of squared improvements at all splits determined by it. A PC's overall relative importance (shown here) is the average of these values over all trees, expressed as a percentage. Graphs of the relationship between PCs and response variables are in Figs. S1–S3 , and prediction errors of ABT models are given in *Results*.

**Table 2. Sparse PCs retained in the best ABT model of relative spread rate**

| Explanatory variable | PC4 | PC1 | PC6 | PC2 |
|---|---|---|---|---|
| Present in Britain | | −0.003 | **−0.645** | 0.016 |
| Global economic use | −0.098 | 0.161 | | |
| Naturalization date | | | **0.555** | |
| Commercial agriculture use | **−0.212** | 0.116 | | 0.144 |
| Compatible *Rhizobium* | **−0.273** | 0.159 | | |
| Seed-contaminant frequency | **−0.569** | | | |
| Cultivated habitats (NZ) | **−0.351** | | | |
| Native-range area | | **0.386** | | |
| Conditions tolerated | | **0.477** | | **−0.318** |
| New Zealand climate match | | **0.394** | | 0.050 |
| Urban/rural habitats | **−0.298** | | **0.365** | |
| Elevation of native range | **0.207** | | | 0.058 |
| Native canopy cover | **−0.350** | −0.131 | −0.195 | |
| Life span | | **0.243** | | |
| Vegetative reproduction | | **0.458** | | |
| Self-pollination capable | | | | **0.424** |
| Corolla length | 0.051 | 0.140 | | **0.618** |
| Seed size | | | | **0.369** |
| Long-distance dispersal | **−0.352** | −0.043 | | 0.134 |
| Seed shape like *T. repens* | | 0.137 | | 0.125 |
| Polyploidy | −0.104 | 0.124 | | |
| Reduced chromosome number | | | | 0.070 |
| Maximum height | | | | **0.328** |
| Habit | | | **0.286** | 0.032 |
| Flowering time (native) | | | | |
| Flowering length (native) | | **0.251** | −0.149 | −0.025 |
| Flowering time (NZ) | −0.057 | | | −0.028 |
| Flowering length (NZ) | −0.172 | | | −0.161 |
| PC relative importance, % (effect direction) | 39.9 (−) | 26.2 (+) | 17.8 (−) | 16.1 (−) |

PC loadings ≥|0.200| are (arbitrarily) bolded for emphasis. Blank cells indicate loadings ≈0. PC analysis details are in Table S7, graphs of the relationship between PCs and the response variable are in Fig. S4, and prediction errors of ABT models are given in *Results*. The Table 1 legend describes how relative importance values were calculated.

tional introduction (PC1). Like intentional introductions, species that were unintentionally introduced and succeeded in naturalizing were characterized by large native ranges, tolerance of diverse conditions (many biomes in native range), presence in climates well matched to New Zealand, and extensive opportunity for transport and use by humans (Table 1). The relative importance of PC1 was greater for unintentional relative to intentional introductions; the ABT model including only this PC had a prediction error of 2.6% (5 of 174 species misclassified).

The best ABT model for relative spread rate included main effects of four of the seven PCs (Table 2 and Fig. S4), explaining 56.0% of the variation (prediction error = 44.0%). Species that spread most rapidly in New Zealand tended to be frequent seed contaminants; to occupy low-elevation, open, urban/rural habitats in their native ranges and cultivated habitats in New Zealand; to lack morphological adaptations for long-distance dispersal; and to be used for commercial agriculture with associated introduction of compatible *Rhizobium* (PC4) (Table 2). PC1 was almost as important, including large native range, tolerance of diverse conditions (many biomes in native range), and presence in climates well matched to New Zealand; a perennial life span with capability for vegetative reproduction; and a long flowering period in the native range (Table 2). PC6, of lesser importance, suggested that British species that naturalized early spread faster (Table 2). PC2, also of lesser importance, suggested that self-pollinating species with short corollas, small seeds, and reduced height, which also tended to tolerate diverse conditions (many biomes in native range), spread more quickly (Table 2). Finally, length of the New Zealand flowering period had high relative importance in the raw variable analysis (Table S10), although it did not load strongly on any one PC (Table S7).

## Discussion

*Trifolium* introductions to New Zealand provide an ideal system for understanding why some plants succeed through key stages in the invasion process. Two important findings emerge. First, human and biogeographic factors were more strongly associated with success than biological attributes across all stages. We found a relationship with biological traits only for the selection of species for introduction and relative rates of spread. Second, whereas some factors were consistently associated with success across all stages, several were unique to particular stages, implying that different factors determine the probability of plant introduction, naturalization, and spread.

New Zealand lacks native *Trifolium* species, and the first wave of *Trifolium* introductions began in the mid-1800s with an influx of European, predominantly British, immigrants (23). European colonists intentionally introduced *Trifolium* species that were of known economic utility and available in their home countries (PC1), a pattern that parallels bird introductions to New Zealand (24). Starting in the 1950s, efforts to increase agricultural production led to trialing of additional clover species from other locations (25). Economic use and a large native range (PC1, PC4) doubtless increased species' availability; many accessions used in trials were foreign cultivars. Collection sites were often chosen based on their climatic similarity with New Zealand (PC1) (25). Here, tall species with long flowering periods would have been conspicuous (PC4) and may have been more commonly collected. Disproportionate introduction of perennials

(PC3) likely derives from the prior success of such species, and their suitability for many marginal environments (26). Finally, native-range abundance, which we did not measure, may have mediated some processes at this stage, because it is often correlated with range size (27) and likely both a cause and effect of economic use.

Intentional introduction of *Trifolium* species occurred with great variation in timing and effort. The mean (±SE) year of first introduction for commercial species intended for immediate sowing was 1881 (±11 yr), by which time the area planted in "grasses and clovers" exceeded 1.4 million hectares (ha) (28). In contrast, the introduction of species for agricultural trials had a mean first introduction year of 1964 (±3), and the area planted averaged only 0.01 (±0.007) ha (e.g., ref. 29). The success of commercial species in naturalizing (8 of 10 succeeded) relative to species introduced for agricultural trials or horticulture (1 of 44 succeeded) very likely reflects their early introduction and high propagule pressure, two factors repeatedly identified as promoting plant naturalization (30–34). Two commercial species (*Trifolium alexandrinum* and *Trifolium ambiguum*) failed to naturalize, but even these largely fit the pattern; they were minor commercial species with small native ranges, and, for *T. ambiguum*, late introduction. However, biological idiosyncrasies such as unique rhizobial requirements may also have contributed. In addition to the propagule pressure advantage of commercial species, biological attributes that made species commercially desirable, including rapid growth and ecological versatility, may have contributed to naturalization success. Our naturalization results parallel those for animal introductions where propagule pressure exerts a strong influence (7), with native range attributes further contributing (16, 35).

From the pool of species that were not intentionally introduced, those that succeeded in arriving and naturalizing were characterized by the same set of human transport/use and biogeographic attributes as intentional, especially commercial, introductions. This surprising result may reflect the importance of commercial pasture seed contamination as an unintentional introduction pathway (36, 37). This early pathway would have facilitated naturalization through both high propagule pressure and deposition of species into well distributed, favorable environments (38, 39). Pre-1900 European seed contaminants also contributed significantly to the U.S. naturalized flora (40). Although successful species shared certain biological and habitat attributes—nearly all were self-pollinated annuals of lowland, urban/rural habitats—their large native ranges, climatic tolerances, and related presence in Britain were what critically distinguished them from otherwise similar Mediterranean species. Additionally, ecological attributes such as ecological versatility and superior reproduction and dispersal in human-modified environments likely enabled species to both widely colonize Britain and naturalize in New Zealand. Other studies of plant introduction–naturalization have found strong associations with native-range attributes (41–43).

Species that subsequently spread most rapidly included nearly equal numbers of intentional and unintentional introductions. These species shared an affinity for low-elevation, open, urban/rural habitats and a strong association with agricultural systems in New Zealand, being either commercially planted, occurring as seed contaminants, or both (PC4). Association with agricultural systems conferred a significant propagule-pressure advantage, including introduction over a wide area (44). This would have facilitated widespread and numerous naturalization events functioning as invasion foci and resulting in rapid spread (45, 46). Moreover, both the timing of these early escapees relative to other *Trifolium* species (PC6) and their particular habitat preferences (PC4) may have provided a head start in occupying available habitat (47). The inclusion of PC1 and PC2 suggests that rapidly spreading species were characterized by large native ranges, presence in a large number of biomes, and climates well matched to New Zealand. Species with larger native ranges may have achieved greater intentional and unintentional propagule pressure (43), facilitating rapid spread. Alternatively or additionally, such species may have spread quickly because of ecological versatility (41, 42) or preadaptation to the New Zealand climate (15, 48). Finally, species with longer flowering periods, particularly in New Zealand, tended to spread more rapidly (Table S10); this attribute is often associated with invasion success (49), via mechanisms such as associated longer fruiting period and increased seed transport probability. Here, it may also reflect species' versatility and ability to exploit favorable conditions. The ABT model for relative spread explained less variation than the models for earlier invasion stages. We may have missed key variables that characterize rapidly spreading species, but this may also reflect the different pathways by which species with different characteristics can achieve similar relative rates of spread.

Overall, our results suggest that certain *Trifolium* species have become widespread in New Zealand by one of two main pathways: (*i*) economically useful species introduced early by Europeans were widely planted, with most subsequently naturalizing and several spreading rapidly and (*ii*) species cooccurring with commercially important species or present in nearby urban/rural habitats were unintentionally introduced and naturalized, with some spreading rapidly via pathways linked to human activities. Species in each of these groups share different biological traits; the intentional species were predominantly perennials, with traits such as inability to self-pollinate, whereas the unintentional species were predominantly annuals, with traits such as small seeds. These two distinct pathways, favoring species with very different trait suites, may explain why, overall, few biological traits were associated with relative rate of spread.

A large native range was strongly associated with success across all stages and pathways. Our study does not isolate the mechanism of this association, but it suggests several possibilities, such as overlap with a trade/immigration hotspot; greater abundance, human encounter frequency, and therefore propagule pressure; overlap with a similar climate; ecological versatility; and biological traits that facilitate large range size. The consistent importance of this attribute in invasion studies (32) may be a product of several mutually reinforcing processes acting at different invasion stages.

Human-associated factors were strongly tied to invasion success in our study, an association that is increasingly evident in natural as well as modified landscapes (6, 11, 50). However, several unique features of our study system also probably contributed. First, as noted, *Trifolium* species selected for introduction likely also possessed ecological attributes facilitating success. It may be easier to untangle these influences in groups where these two factors are less closely associated, such as horticultural introductions selected for flower form. Moreover, in New Zealand, *Trifolium* species have spread mostly within highly modified habitats. The "barriers" to overcome between our naturalization and spread stages (51) were therefore not great, probably contributing to the association of spread with human dispersal. Biological traits may play a larger role in influencing rate of spread in natural habitats, or in ecological impact (50, 52, 53). Additionally, the scale of *Trifolium* planting in New Zealand far exceeds that of a typical introduced plant; by the late 20th century, *Trifolium* species were planted on 10 million ha (54), nearly 38% of New Zealand's land area. Finally, the importance of human influences may be greater when comparing species with similar life histories (32). Despite this, our overall findings are consistent with studies of other naturalized floras (40, 55). Whereas the economic importance of *Trifolium* facilitated our study, similar stage-based analyses should be feasible for other plant groups, particularly in the New

World where historical introduction and naturalization records are more available (e.g., refs. 31 and 40).

Inferences from retrospective invasion analyses will always be limited, as was our study, by close interrelationships among factors promoting success. Experimental studies that break such correlations (e.g., ref. 56) are therefore critical to understanding invasions. Nevertheless, retrospective studies contribute significantly to understanding of real-world situations, particularly given increasing recognition of human influences as a key invasion driver. Our study illustrates the value of stage- and pathway-specific analyses in which human factors are explicitly considered.

## Materials and Methods

To identify which of the world's 228 *Trifolium* species (57) were introduced intentionally to New Zealand, we consulted the NZ ALLWEEDS database (58), herbarium specimens, historical publications (particularly agricultural trials), and experts, yielding a list of 54 species. Review of herbarium specimens and floras (59) identified 25 species naturalized in New Zealand, of which 9 were intentional introductions. The remaining 16 naturalized species apparently arrived unintentionally, deriving from the pool of 174 *Trifolium* species that were never intentionally introduced. We therefore split the invasion process into four stages and analyzed each separately (Fig. 1): intentional introduction, naturalization of intentionally introduced species, unintentional introduction–naturalization, and spread.

We created a database of occurrence records (1850–2000) for the 25 naturalized *Trifolium* species, including locations from herbarium specimens and the National Vegetation Survey (60). Our database totaled 1,817 records with a mean of 73 per species (range 1–434). The New Zealand Map Series (NZMS260) defines a national grid with cells 40 km × 30 km; each database record was assigned to a grid cell. Because of temporal variation in survey effort, we used a resampling procedure to estimate species' relative spread rates. Dividing our study period into six 25-year intervals, we found that the minimum number of records for all species in any one interval was 12. Each iteration of our resampling procedure therefore selected 12 records at random from each of the six time intervals and then counted the cumulative number of unique grid cells occupied by each species over the six intervals. From 10,000 iterations, we took the mean of these counts for each species–interval combination and assigned each species a zero for the time interval before its first database record. To calculate spread rate from such data, previous studies have used a variety of models (48, 61–63). Simple linear models provided the best fit to our relative data, and we therefore used linear slopes as our measure of relative spread rate.

We identified a set of biological, biogeographic, and human-related variables that could explain variation in invasion success (Table S1). Center-of-origin and economic-use information was taken from the GRIN database,[§] whereas mid-1800s presence in Britain was derived from British floras (64, 65). A variety of sources including seed catalogs, government publications, and newspaper advertisements were consulted to determine whether species had been introduced for commercial-scale agriculture; other reasons for introduction included research and horticulture. Dates of first introduction and naturalization were derived from these sources and historical accounts (66, 67). To estimate species' compatibility with introduced *Rhizobium* strains, we assumed that compatible strains had been introduced for commercial agricultural species and then used published effectiveness groups (68–70) to identify other *Trifolium* species compatible with these strains. Species identified as having compatible strains were coded 1, those for which compatible strains were not identified were coded 0, and species not evaluated by our sources were coded 0.5. After searching the literature for records of *Trifolium* species as pasture seed contaminants, we assigned each naturalized species an ordinal value (0–4) describing its relative frequency as a contaminant. Species' habitat descriptions (59) were used to determine whether or not species were present in New Zealand's cultivated areas.

For each of the world's 228 *Trifolium* species, we digitized GRIN native-range information into a geographic information system (GIS) to calculate native-range area. The number of biomes in which a species occurred (diversity of conditions tolerated) was determined by overlaying the native range onto a World Wildlife Fund biome map (71). A climate match index was calculated as the number of 10′ × 10′ latitude–longitude grid cells in New Zealand that had a 95% or better match to the climate of any 10′ × 10′ grid cell in the species' native range (15). Climate match was calculated by using the Euclidean distance-based CLIMATE algorithm (72) and a global meteorological dataset (73). Habitat attributes and biological traits were recorded from floras and genus monographs (57, 59, 74).

Statistical analyses were performed in R (75). Sixteen of the 29 variables contained at least one missing value (mean percentage missing = 5%). To handle missing values, we chose imputation rather than casewise deletion to avoid bias toward better-known species (76). We imputed missing values using the mix package (77). We specified an estimation model with no interactions, the most complex model we could reasonably estimate. We ran the imputation procedure 1,000 times and used the mean of these runs as our final values; mean values stabilized well before the final iteration.

Many of the explanatory variables were correlated, and the correlation matrix suggested that linear combination might usefully summarize them. To generate meaningful combinations, we used sparse principal component analysis (SPCA) (22), a variant of principal component analysis (PCA) that seeks to simultaneously improve interpretability and maintain high explained variance. SPCA formulates PCA as a regression-type optimization problem in which the lasso penalty ($\lambda_1$) is directly integrated into the regression criterion (22). The user chooses the $\lambda_1$ penalty parameter for each PC, with higher values yielding greater loading sparsity but lower explained variance. Sparse PCs were estimated by using the elasticnet package (78). We performed a separate SPCA for each invasion stage because of differences in the pattern of variation and the addition of variables at later stages. We first chose the number of PCs to retain based on a standard PCA (princomp) and traditional ad hoc criteria (79). We then used elasticnet to plot the percentage of explained variance as a function of $\lambda_1$ and chose $\lambda_1$ values to maintain reasonably high explained variance. Our sparse PCs explained 87%, 83%, and 77% of the variation explained by the original PCs for the introduction, naturalization, and spread stages, respectively, or 61%, 60%, and 60% of the variation at these stages overall (Tables S2, S5, and S7).

Having generated sparse PCs for each stage, we sought to determine their relationship to species' success. We used ABT modeling (21) because of its focus on accurate prediction and interpretable explanation, its suitability for exploratory studies, and its accommodation of nonlinear relationships. However, boosted-tree methods still assume that explanatory variables are independent; we discuss PC correlations in Tables S3, S4, S6, and S8. We used the gbmplus package (80) for the ABT analysis, following Ridgeway (81) for metaparameter values and basing model comparisons on 10 runs of fivefold cross-validation fits (21). We adopted De'ath's (21) approach of selecting a ''best'' model for each stage by removing interaction orders and variables with low relative importance, retaining parsimonious models whose prediction error was not substantially worse than the original. We also constructed ABTs for the raw variables for comparison.

**ACKNOWLEDGMENTS.** We thank A. V. Stewart, R. J. Lucas, W. M. Williams, H. Gatehouse, S. L. Bithell, and A. L. Miller for helpful discussion and B. Case for GIS assistance. New Zealand herbaria personnel, especially A. D. Wilton and M. Breach, provided assistance with specimens. We acknowledge the use of data drawn from the National Vegetation Survey Database (NVS). P. J. Bellingham, J.H. Brown, and two anonymous reviewers provided very helpful comments on the manuscript. This work was supported by Fulbright New Zealand and the Lincoln University Bio-Protection and Ecology Division.

---

[§] U.S. Department of Agriculture, Agricultural Research Service, National Genetic Resources Program. *Germplasm Resources Information Network—(GRIN)* (online database). National Germplasm Resources Laboratory, Beltsville, MD.

1. Pheloung PC, Williams PA, Halloy SR (1999) A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. *J Environ Manage* 57:239–251.
2. Committee on the Scientific Basis for Predicting the Invasive Potential of Nonindigenous Plants and Plant Pests in the United States (2002) *Predicting Invasions of Nonindigenous Plants and Plant Pests* (Natl Acad Press, Washington, DC).
3. Baker HG (1965) Characteristics and modes of origin of weeds. *The Genetics of Colonizing Species*, eds Baker HG, Stebbins GL (Academic, New York), pp 147–172.
4. Newsome AE, Noble IR (1986) Ecological and physiological characters of invading species. *Ecology of Biological Invasions: An Australian Perspective*, eds Groves RH, Burdon JJ (Cambridge Univ Press, Cambridge), pp 1–20.
5. Smith CS, Lonsdale WM, Fortune J (1999) When to ignore advice: Invasion predictions and decision theory. *Biol Invasions* 1:89–96.
6. Kowarik I (2003) Human agency in biological invasions: Secondary releases foster naturalisation and population expansion of alien plant species. *Biol Invasions* 5:293–312.
7. Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. *Trends Ecol Evol* 20:223–228.
8. Colautti RI, Grigorovich IA, MacIsaac HJ (2006) Propagule pressure: A null model for biological invasions. *Biol Invasions* 8:1023–1037.
9. Mack RN (2001) Motivations and consequences of the human dispersal of plants. *The Great Reshuffling: Human Dimensions of Invasive Alien Species*, ed McNeely JA (International Union for Conservation of Nature, Gland, Switzerland), pp 23–34.

10. Mack RN (1991) The commercial seed trade: An early disperser of weeds in the United States. *Econ Bot* 45:257–273.
11. Thuiller W, Richardson DM, Rouget M, Proches S, Wilson JRU (2006) Interactions between environment, species traits, and human uses describe patterns of plant invasions. *Ecology* 87:1755–1769.
12. Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends Ecol Evol* 16:199–204.
13. Heger T, Trepl L (2003) Predicting biological invasions. *Biol Invasions* 5:313–321.
14. Williamson M (2006) Explaining and predicting the success of invading species at different stages of invasion. *Biol Invasions* 8:1561–1568.
15. Duncan RP, Bomford M, Forsyth DM, Conibear L (2001) High predictability in introduction outcomes and the geographical range size of introduced Australian birds: A role for climate. *J Anim Ecol* 70:621–632.
16. Forsyth DM, Duncan RP, Bomford M, Moore G (2004) Climatic suitability, life-history traits, introduction effort, and the establishment and spread of introduced mammals in Australia. *Conserv Biol* 18:557–569.
17. Kolar CS, Lodge DM (2002) Ecological predictions and risk assessment for alien fishes in North America. *Science* 298:1233–1236.
18. Marchetti MP, Moyle PB, Levine R (2004) Invasive species profiling? Exploring the characteristics of non-native fishes across invasion stages in California. *Freshwater Biol* 49:646–661.
19. Saxby SH (1940) Grasses and clovers of New Zealand: Clovers are the key to high production. *N Z J Agric* 60:271–276.
20. Mack RN (1996) Predicting the identity and fate of plant invaders: Emergent and emerging approaches. *Biol Conserv* 78:107–121.
21. De'ath G (2007) Boosted trees for ecological modeling and prediction. *Ecology* 88:243–251.
22. Zou H, Hastie T, Tibshirani R (2006) Sparse principal component analysis. *J Comput Graph Stat* 15:265–286.
23. King M (2003) *The Penguin History of New Zealand* (Penguin Books, Auckland).
24. Duncan RP, Blackburn TM, Cassey P (2006) Factors affecting the release, establishment and spread of introduced birds in New Zealand. *Biological Invasions in New Zealand*, eds Allen RB, Lee WG (Springer, Berlin), pp 137–154.
25. Burt RL, Forde MB (1987) The provision of pasture legumes for difficult environments. *Proc N Z Grassl Assoc* 48:83–88.
26. Scott D (1985) Plant introduction trials: Genotype–environment analysis of plant introductions for the high country. *N Z J Exp Agric* 13:117–127.
27. Gaston KJ, *et al.* (2000) Abundance–occupancy relationships. *J Appl Ecol* 47:39–59.
28. New Zealand Registrar-General's Office (1881) *Statistics of the Colony of New Zealand* (Government Printer, Wellington, New Zealand).
29. Taylor AO, Hughes KA, Hunt BJ (1979) Annual cool-season legumes for forage. II. Seasonal growth patterns and effects of cutting frequency and cutting height on yield. *N Z J Exp Agric* 7:149–152.
30. Kowarik I (1995) Time lags in biological invasions with regard to the success or failure of alien species. *Plant Invasions—General Aspects and Special Problems*, eds Pyšek P, Prach K, Rejmánek M, Wade M (SPB Academic Publishing, Amsterdam), pp 15–38.
31. Mulvaney M (2001) The effect of introduction pressure on the naturalization of ornamental woody plants in south-eastern Australia. *Weed Risk Assessment*, eds Groves RH, Panetta FD, Virtue JG (Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia), pp 186–193.
32. Rejmánek M (2000) Invasive plants: Approaches and predictions. *Aust Ecol* 25:497–506.
33. Krivánek M, Pyšek P, Jarošík V (2006) Planting history and propagule pressure as predictors of invasion by woody species in a temperate region. *Conserv Biol* 20:1487–1498.
34. Dehnen-Schmutz K, Touza J, Perrings C, Williamson M (2007) A century of the ornamental plant trade and its impact on invasion success. *Diversity Distrib* 13:527–534.
35. Duncan RP, Blackburn TM, Sol D (2003) The ecology of bird introductions. *Annu Rev Ecol Evol Syst* 34:71–98.
36. Healy AJ (1952) The introduction and spread of weeds. *Proc N Z Weed Control Conf* 5:5–16.
37. Esler AE (1987) The naturalisation of plants in urban Auckland, New Zealand. 1. The introduction and spread of alien plants. *N Z J Bot* 25:511–522.
38. Crawley MJ (1989) Chance and timing in biological invasions. *Biological Invasions: A Global Perspective*, eds Drake JA, *et al.* (Wiley, New York), pp 407–435.
39. Mack RN (2000) Cultivation fosters plant naturalization by reducing environmental stochasticity. *Biol Invasions* 2:111–122.
40. Mack RN, Erneberg M (2002) The United States naturalized flora: Largely the product of deliberate introductions. *Ann Mo Bot Gard* 89:176–189.
41. Goodwin BJ, McAllister AJ, Fahrig L (1999) Predicting invasiveness of plant species based on biological information. *Conserv Biol* 13:422–426.
42. Prinzing A, Durka W, Klotz S, Brandl R (2002) Which species become aliens? *Evol Ecol Res* 4:385–405.
43. Forcella F, Wood JT (1984) Colonization potentials of alien weeds are related to their ''native'' distributions: Implications for plant quarantine. *J Aust Inst Agric Sci* 50:35–40.
44. Levy EB (1923) The grasslands of New Zealand: Principles of pasture establishment. Grass–seed mixtures for various soils and conditions. *N Z J Agric* 26:263–279.
45. Auld BA, Coote BG (1980) A model of a spreading plant population. *Oikos* 34:287–292.
46. Moody ME, Mack RN (1988) Controlling the spread of plant invasions: The importance of nascent foci. *J Appl Ecol* 25:1009–1021.
47. Baker HG (1974) The evolution of weeds. *Annu Rev Ecol Syst* 5:1–24.
48. Mihulka S, Pyšek P (2001) Invasion history of *Oenothera* congeners in Europe: A comparative study of spreading rates in the last 200 years. *J Biogeogr* 28:597–609.
49. Pyšek P, Richardson DM (2007) Traits associated with invasiveness in alien plants: Where do we stand? *Biological Invasions*, ed Nentwig W (Springer, Berlin), pp 97–125.
50. Pyšek P, Prach K, Šmilauer P (1995) Relating invasion success to plant traits: An analysis of the Czech alien flora. *Plant Invasions–General Aspects and Special Problems*, eds Pyšek P, Prach K, Rejmánek M, Wade M (SPB Academic, Amsterdam), pp 39–60.
51. Richardson DM, *et al.* (2000) Naturalization and invasion of alien plants: concepts and definitions. *Diversity Distrib* 6:93–107.
52. Kornas J (1990) Plant invasions in central Europe: Historical and ecological aspects. *Biological Invasions in Europe and the Mediterranean Basin*, eds di Castri F, Hansen AJ, Debussche M (Kluwer, Dordrecht, The Netherlands), pp 19–36.
53. Gurvich DE, Tecco PA, Díaz S (2005) Plant invasions in undisturbed ecosystems: The triggering attribute approach. *J Veg Sci* 16:723–728.
54. Ministry of Agriculture and Fisheries (1982) *New Zealand Agricultural Statistics* (Ministry of Agriculture and Fisheries, Wellington, New Zealand).
55. Hodkinson DJ, Thompson K (1997) Plant dispersal: The role of man. *J Appl Ecol* 34:1484–1496.
56. Maron JL (2006) The relative importance of latitude matching and propagule pressure in the colonization success of an invasive forb. *Ecography* 29:819–826.
57. Gillett JM, Taylor NL (2001) *The World of Clovers* (Iowa State Univ Press, Ames, IA).
58. Williams PA, Wilton A, Spencer N (2002) *A Proposed Conservation Weed Risk Assessment System for the New Zealand Border* (Department of Conservation, Wellington, New Zealand).
59. Webb CJ, Sykes WR, Garnock-Jones PJ (1988) *Flora of New Zealand Vol IV: Naturalized Pteridophytes, Gymnosperms, and Dicotyledons* (DSIR Botany, Christchurch, New Zealand).
60. Guralnick R, Van Cleve J (2005) Strengths and weaknesses of museum and national survey datasets for predicting regional species richness: Comparative and combined approaches. *Diversity Distrib* 11:349–359.
61. Weber E (1998) The dynamics of plant invasions: A case study of three exotic goldenrod species (*Solidago* L.) in Europe. *J Biogeogr* 25:147–154.
62. Delisle F, Lavoie C, Jean M, Lachance D (2003) Reconstructing the spread of invasive plants: Taking into account biases associated with herbarium specimens. *J Biogeogr* 30:1033–1042.
63. Pyšek P, Prach K (1993) Plant invasions and the role of riparian habitats: A comparison of four species alien to Central Europe. *J Biogeogr* 20:413–420.
64. Lindley J (1835) *A Synopsis of the British Flora* (Longman, Rees, Orme, Brown, Green, and Longman, London).
65. Bentham G (1865) *Handbook of the British Flora* (Lovell Reeve, London).
66. Thomson GM (1922) *The Naturalisation of Animals and Plants in New Zealand* (Cambridge Univ Press, London).
67. Webb CJ (1980) Checklist of dicotyledons naturalised in New Zealand 5. Leguminosae. *N Z J Bot* 18:463–472.
68. Burton JC (1985) *Rhizobium* relationships. *Clover Science and Technology*, ed Taylor NL (American Society of Agronomy, Madison, WI), pp 161–184.
69. Smith RS, Thome S, Randall R (1987) *Rhizobia Inoculants for Various Leguminous Species* (The Nitragin Company, Milwaukee).
70. Bio-Care Technology (2003) *Nodulaid Group Chart* (Bio-Care Technology Pty. Limited, Somersby, New South Wales, Australia).
71. Olson DM, *et al.* (2001) Terrestrial ecoregions of the world: A new map of life on earth. *BioScience* 51:933–938.
72. Pheloung PC (1996) *CLIMATE: A System to Predict the Distribution of an Organism Based on Climate Preferences* (Agriculture Western Australia, South Perth, Australia).
73. New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. *Climate Res* 21:1–25.
74. Zohary M, Heller D (1984) *The Genus Trifolium* (Israel Academy of Sciences and Humanities, Jerusalem).
75. R Development Core Team (2007) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
76. Schafer JL (1997) *Analysis of Incomplete Multivariate Data* (Chapman and Hall/CRC, Boca Raton, FL).
77. Schafer JL (2007) *mix: Estimation/Multiple Imputation for Mixed Categorical and Continuous Data* (R package version 1.0–6).
78. Zou H, Hastie T (2005) *elasticnet: Elastic Net Regularization and Variable Selection* (R package version 1.0–3).
79. Jolliffe IT (2002) *Principal Component Analysis* (Springer, New York).
80. Ridgeway G, modifications by G. De'ath (2006) *gbmplus: Generalized Boosted Regression Models* (R package version 1.5–17).
81. Ridgeway G (2006) *Generalized Boosted Models: A Guide to the gbm Package* (R vignette).