Population genetics and symbiont assemblages support different invasion scenarios for the red turpentine beetle (*Dendroctonus valens*)

Stephen J. Taerum[1], Adam Konečný[2], Z. Wilhelm de Beer[3], David Cibrián-Tovar[4], Michael J. Wingfield[1]

[1] Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa

[2] Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 2, Brno 61137, Czech Republic

[3] Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa

[4] División de Ciencias Forestales, Universidad Autónoma Chapingo, Km 38.5 Carretera México-Texcoco, Chapingo, Estado de México, México

Running title: *D. valens*' genetic structure and invasion history

Abstract

Exotic forest insects and their symbionts pose an increasing threat to forest health. This is apparently true for the red turpentine beetle (*Dendroctonus valens*), which was unintentionally introduced to China, where the beetle has killed millions of healthy native pine trees. Previous population genetics studies that used cytochrome oxidase I as a marker concluded that the source of *D. valens* in China was western North America. In contrast, surveys of fungi associated with *D. valens* demonstrated that more fungal species are shared between China and eastern North America than between China and western North America, suggesting that the source population of *D. valens* could be eastern North America. In this study, we used microsatellite markers to determine population structure of *D. valens* in North America as well as the source population of the beetle in China. The analyses revealed that four genetically-distinct populations (herein named the West, Central, Northeast and Mexico) represent the native range of *D. valens*. Clustering analyses and a simulation-based Approximate Bayesian Computation (ABC) approach supported the hypothesis that western North America is the source of the invasive *D. valens* population. This study provides a demonstration of non-congruence between patterns inferred by studies on population genetics and symbiont assemblages in an invasive bark beetle.

Keywords: Approximate Bayesian Computation - biogeography - biological invasion - forest insect - population genetics - symbiosis

Introduction

Over the past century, rapid growth in global trade combined with environmental changes caused by human activities have led to a dramatic increase in the movement of forest insects around the world (Liebhold *et al.*, 2012; Weed *et al.*, 2013; Wingfield *et al.*, 2015). Although most forest insects that become established outside their native ranges either through range expansion or translocation to non-contiguous locations have minor impacts on invaded forests, a few exotic insect species can form tree-killing epidemics that can dramatically transform forest ecosystems (Aukema *et al.*, 2011). To confound matters, forest insects can co-invade with symbionts or acquire new symbionts that are themselves damaging to trees, or that otherwise facilitate the mortality caused by their vectors (Hulcr & Dunn, 2011; Ploetz *et al.*, 2013). Alarmingly, several of the most damaging associations between exotic forest insects and their symbionts are relatively non-aggressive in their native ranges (e.g., Haack & Rabaglia 2013; Slippers *et al.* 2015), demonstrating that the movement of any forest insect poses a threat to forest health. As global trade increases and environmental conditions continue to change, effective management strategies are required to limit the damage caused by current and future outbreaks of invasive insects and their symbionts.

Management of an exotic forest insect and its symbionts requires substantial knowledge of the insect's invasion history (Estoup & Guillemaud, 2010). This information can help researchers to identify high risk source populations (Ciosi *et al.*, 2008) and pathways (Hulme *et al.*, 2008) for future invasions, origins of damaging symbionts vectored by the invasive pest (Boissin *et al.*, 2012), environmental conditions or anthropogenic activities that are conducive to the movement of the invasive species (Pyšek *et al.*, 2010), evolutionary processes such as bottlenecks, selection, and admixture that influence the spread of the insect in its invaded ranges

(Dlugosch & Parker, 2008), potential biological control agents from the pest's native range (Roderick & Navajas, 2003), and other factors that can influence the success and impact of invading forest insects and their symbionts. To test hypotheses on the source populations, movement patterns, demographic changes and evolutionary trajectories of an invasive insect, researchers often supplement observational data, such as trap capture data in the insect's invasive ranges, with genetic data (Estoup & Guillemaud, 2010).

The invasive red turpentine beetle (*Dendroctonus valens* LeConte) is an economically and ecologically important invasive forest insect (Yan *et al.*, 2005) for which the previously collected data may be inadequate to accurately assess its invasion history. The insect was introduced to Shanxi province, China, in the 1980s (Yan *et al.*, 2005; Sun *et al.*, 2013). Since 1999, *D. valens* has rapidly expanded its range in China (Fig. 1), while killing millions of otherwise healthy native pine trees (Yan *et al.*, 2005; Sun *et al.*, 2013). The aggressive behaviour of *D. valens* in China is a major behavioural shift, as the beetle primarily colonizes stressed or dying pine trees and is rarely the primary cause of host mortality in its native range in North America (Fig. 1; Owen *et al.*, 2010). The aggressive population of *D. valens* poses a further threat to Eurasia's extensive pine forests if its invasive range continues to expand (Sun *et al.*, 2013).

Historical records suggest that *D. valens* was introduced to China on untreated timber imported from the western USA (Yan *et al.*, 2005). Molecular studies using the mitochondrial cytochrome oxidase I (COI) gene as a marker have largely supported a western North American origin, as these studies have concluded that the source of *D. valens* in China was either the Pacific Northwest (Cognato *et al.*, 2005) or California (Cai *et al.*, 2008). However, recent surveys of ophiostomatalean (Wingfield *et al.*, 1993) fungi (Ascomycota: Ophiostomatales)

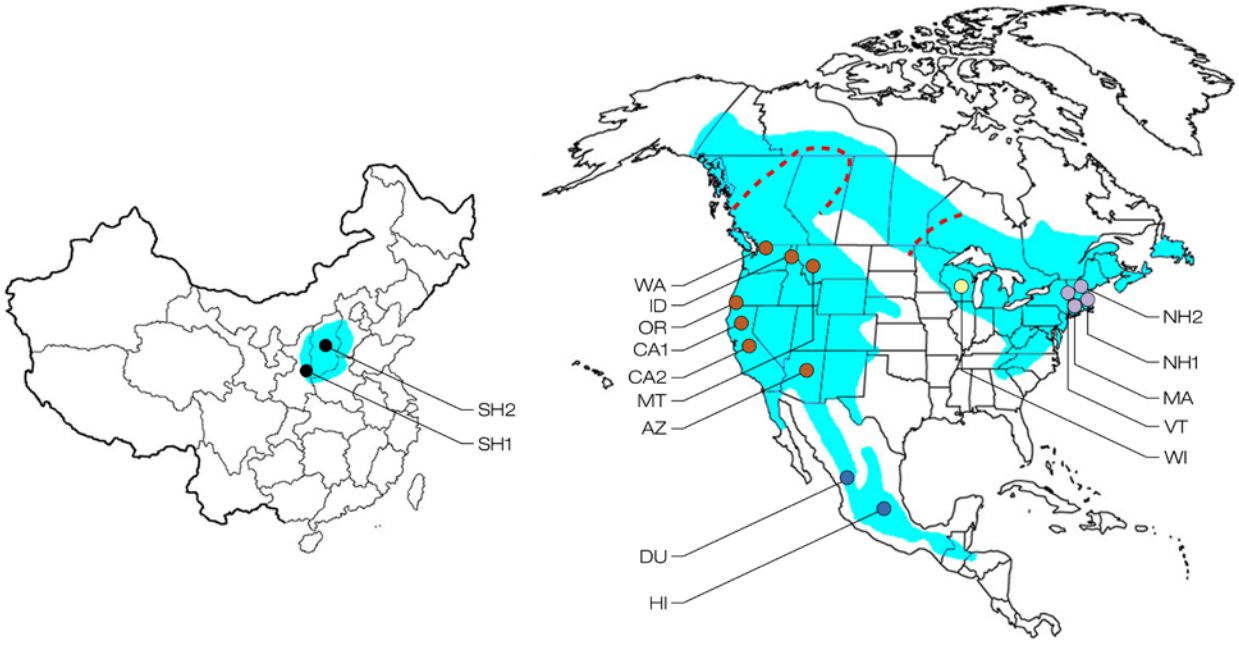

Fig. 1: Distributions of Dendroctonus valens in China and North America, shown with the sampling locations for this study. Red dashed lines demarcate the eastern and western North American ranges (Taerum et al., 2013). Circles represent sampling locations, and the colours of the circles represent the population to which the sampling locations belong (based on the genetic clusters shown in Fig. 2): brown for the West population, yellow for the Central population, purple for the Northeast population, blue for the Mexico population, and black for the China population. Abbreviations are as follows: AZ - Arizona, CA1 - California 1, CA2 - California 2, DU - Durango, HI - Hidalgo, ID - Idaho, MA - Massachusetts, MT - Montana, NH1 - New Hampshire 1, NH2 - New Hampshire 2, OR - Oregon, SH1 - Shaanxi, SH2 - Shanxi, VT - Vermont, WA - Washington, WI - Wisconsin.

associated with *D. valens* in China and North America have supported the hypothesis of an eastern North American origin for the bark beetle (Lu *et al.*, 2009a,b; Taerum *et al.*, 2013). Taerum *et al.* (2013) determined that three ophiostomatalean species (*Grosmannia koreana* (J.J. Kim & G.H. Kim) Masuya, J.J. Kim & M.J. Wingf., *Leptographium procerum* (W.B. Kend.) M.J. Wingf., and *Ophiostoma abietinum* Marm. & Butin) are associated with *D. valens* in both eastern North America and China while only one species (*Ophiostoma floccosum* Math.-Käärik) was associated with *D. valens* in western North America and China.

Based on the studies by Lu *et al.* (2009a,b) and Taerum *et al.* (2013), *L. procerum* was the most commonly isolated associate of *D. valens* in both China (61.0% of the isolates) and eastern North America (45.6%). This commensalist fungus is externally-vectored between pine trees by numerous bark beetle and weevil species, and is considered to be only weakly pathogenic on trees in North America (Jacobs & Wingfield, 2001). As *L. procerum* has never been isolated in western North America and has been reported in China only as an associate of *D. valens*, this fungus has been hypothesized to have co-invaded China along with *D. valens* that may have originated from eastern North America (Lu *et al.*, 2011; Sun *et al.*, 2013; Taerum *et al.*, 2013). Interestingly, *L. procerum* may be pathogenic on Chinese pines, and has been suggested to contribute to the aggressive behaviour of *D. valens* in China by causing tree hosts to produce greater quantities of the monoterpene 3-carene, which is a primary attractant of *D. valens* (Erbilgin *et al.* 2007; Lu *et al.*, 2011). The synergy between *D. valens* and *L. procerum* in China accentuates the importance of an accurate understanding of the invasion history of *D. valens*.

The apparent conflict between the molecular studies and the symbiont surveys may be in part due to the fact that eastern North America was under sampled in the molecular studies. Only

eight individuals were included from the eastern North American range, specifically from one locality in the U.S. state of Michigan (Cognato *et al.*, 2005; Cai *et al.*, 2008). Because the range of *D. valens* in eastern North America extends from the Atlantic coast to the Great Plains (Fig. 1; Owen *et al.*, 2010), a small number of samples from one location in eastern North America is unlikely to represent the genetic diversity of the bark beetle in eastern North America.

The use of COI as a molecular marker for *D. valens* is also problematic because ambiguous base pairs are commonly found when sequencing the COI gene of the insect (Cai *et al.*, 2008; Cai *et al.*, 2011; S.J. Taerum, unpublished data). Cai *et al.* (2011) determined that the ambiguous sites are likely caused by contamination from cryptic nuclear mitochondrial DNA (NUMTs), or insertions of mitochondrial DNA into nuclear genomes. Cryptic NUMTs, which are common in bark beetle species (Jordal & Kambestad, 2014), are difficult to detect because they differ from mitochondrial sequences by a small number of base pairs and frequently lack stop codons. NUMT contamination is a serious problem for population genetics studies because NUMT sequences can lead to overestimates of genetic diversity and false phylogeographical inferences (Bertheau *et al.*, 2011). Although Cai *et al.* (2008) excluded sequences with ambiguous sites from their analyses, some of the sequences that they included may have been NUMT sequences that amplified more efficiently than the true COI sequences (Bertheau *et al.*, 2011). Ambiguous base pairs have also been found in other mitochondrial gene regions such as cytochrome oxidase II and cytochrome b (S.J. Taerum, unpublished data). This suggests that nuclear markers may be more suitable for testing hypotheses on the population structure and invasion history of *D. valens*.

Microsatellite markers provide an alternative to mitochondrial sequences to determine the genetic diversity of *D. valens* and possibly the source of the invasive population in China. These

highly variable nuclear markers are effective for estimating population diversities, assessing population structure over geographical ranges (Sunnucks, 2000), and determining the movement histories of invasive species (e.g., Boissin *et al.*, 2012; Konečný *et al.*, 2013). In addition, microsatellites are not impacted by NUMT contamination. Although results from studies using microsatellite and mitochondrial data are usually congruent, microsatellites can sometimes infer genetic structure and phylogeographical patterns that differ from mitochondrial data (e.g., Borden & Stepien, 2006; Zarza *et al.*, 2011). Although microsatellites have drawbacks (reviewed in Brito & Edwards, 2008), these markers might provide a more accurate estimation of the source population and overall invasion history of *D. valens* in China.

In this study, we used microsatellite markers to discern the genetic diversity and structure of *D. valens* in its native range in North America and its invasive range in China. We analyzed these data to test the following hypotheses on the source population of *D. valens* in China: 1) the source was western North America, supporting historical records (Yan *et al.*, 2005) as well as the previous population genetics studies (Cognato *et al.*, 2005; Cai *et al.*, 2008); 2) the source was eastern North America, as suggested by the surveys of ophiostomatalean fungi associated with *D. valens* in China and North America (Lu *et al.*, 2009a,b; Taerum *et al.*, 2013); or 3) the Chinese population was the result of admixture between individuals introduced from both sources.

Methods

Collections

Table 1: Collection and locality information for *Dendroctonus valens* used in this study. * indicates sampling locations where coordinates were estimated from Google Maps (maps.google.com). - indicates that collection details are unknown.

Code	Country	State/Province	Locality	Coordinates	# collected	Collector	Collection year	Collection method
SH1	China	Shaanxi	Yaopin Forest station	N35.76868 E109.26798*	45	Min Lu	2011	Trap collection
SH2	China	Shanxi	Qinyuan County	N36.72904 E112.31812*	17	Min Lu	2011	Trap collection
WA	U.S.A.	Washington	Blewett Pass	N47.33189 W120.67556	32	Darci Carlson	2012	Trap collection
OR	U.S.A.	Oregon	Applegate Lake	N42.04174 W123.10596	32	Bill Schaupp	2012	Trap collection
ID	U.S.A.	Idaho	Spirit Lake	N47.95201 W116.86853	31	Lee Pederson	2012	Trap collection
MT	U.S.A.	Montana	Blue Mountain	N46.81665 W114.09602	27	Diana Six	2012	Hand collection
CA1	U.S.A.	California	McCloud	N41.27213 W122.11021*	28	-	-	-
CA2	U.S.A.	California	El Dorado National Forest	N38.73667 W120.74167	38	Carline Carvalho	2010	Trap collection
AZ	U.S.A.	Arizona	Centennial Forest	N35.16259 W111.76207	28	Rich Hofstetter	2013	Trap collection
WI	U.S.A.	Wisconsin	Pettenwell Lake	N44.14028 W089.83806	34	Aaron Adams	2010	Hand collection
MA	U.S.A.	Massachusetts	Oakham State Forest	N42.42297 W072.02901	21	Stephen Taerum	2011	Trap collection
NH1	U.S.A.	New Hampshire	Durham	N43.14929 W070.93566	27	Garret Dubois	2012	Trap collection
NH2	U.S.A.	New Hampshire	White Mountain National Forest	N43.84322 W071.87319	26	Garret Dubois	2012	Trap collection
VT	U.S.A.	Vermont	Green Mountain National Forest	N43.21133 W072.85205	28	Stephen Taerum	2011	Trap collection
DU	Mexico	Durango	Parque Natural Puerto de los Ángeles	N23.70286 W105.43458	33	Wilhelm de Beer	2013	Hand collection
HI	Mexico	Hidalgo	Jacala	N20.92822 W099.23133	17	Wilhelm de Beer	2013	Hand collection

We collected *D. valens* either by using funnel traps baited with α-pinene or 3-carene to capture adults, or by collecting adults or larvae directly from *D. valens*' galleries (collection details in Table 1). Only one larva was collected from each individual gallery to ensure that siblings were not sampled. Adults collected in traps or galleries were assumed to represent a random sampling of beetles from the area. Beetles were collected from multiple locations in China, Mexico and the USA (Fig. 1; locality information and abbreviations in Table 1), and were stored at -20°C or -80°C immediately after collection.

DNA extractions and microsatellite amplification

We extracted DNA using a CTAB protocol modified from Möller *et al.* (1992). DNA was stored at -20°C until utilization. Thirty three microsatellite markers developed for other *Dendroctonus* spp. were screened to determine if they were informative for *D. valens*: eight markers originally developed for *Dendroctonus frontalis* Zimmerman (Schrey *et al.*, 2007), 16 for *Dendroctonus ponderosae* Hopkins (Davis *et al.*, 2009), and nine for *Dendroctonus rufipennis* Kirby (Maroja *et al.*, 2007). Two individuals each from four different sampling locations were selected to screen the markers using unlabeled primers. Polymerase chain reaction (PCR) amplifications were conducted in 12 µL reactions containing 3 µL template DNA, 4.9 µL dH_2O, 2 µL 5x MyTaq reaction buffer (Bioline, London, UK), 1 µL $MgCl_2$ (25 mM; Roche, Basel, Switzerland), 0.5 µL of each primer (10 pmol/µL; Inqaba biotec, Pretoria, South Africa), and 0.1 µL MyTaq (Bioline). PCR conditions were as follows: 96°C for 2 min, followed by 30 cycles of 94°C for 15 sec, 54°C or 58°C (depending on the primer pair used; Table 1) for 30 sec, and 72°C for 25 sec, followed by a final extension step of 72°C for 30 mins. Because *D. valens* is diploid, the amplicons were cloned into the pGEM-T Vector (Promega, Madison, WI, USA) to

Locus	Previous name	Source	Repeat motif	Annealing temp. (°C)	# alleles	Size range (bp)	Repeat range	F_{ST}	F_{IS}
DV-01	D06	Schrey et al. 2007	AC	58	7	114 - 126	4 - 10	0.269	0.186
DV-02	D10	Schrey et al. 2007	GT	58	2	128 - 130	4 - 5	0.064	0.053
DV-04	D17	Schrey et al. 2007	GT	58	9	113 - 129	6 - 14	0.160	0.004
DV-05	D24	Schrey et al. 2007	AC	58	6	186 - 195$_A$	4 - 8	0.268	0.127
DV-07	D453	Davis et al. 2009	AC$_D$	58	12	169 - 191	13 - 24	0.073	0.008
DV-08	D566	Davis et al. 2009	GT	58	18	125 - 161	7 - 25	0.061	-0.003
DV-09	D793	Davis et al. 2009	CA	54	7	147 - 159	7 - 13	0.083	0.063
DV-10	L54	Maroja et al. 2007	AGG$_E$	58	7	99 - 123$_B$	4 - 12	0.135	0.004
DV-11	L76	Maroja et al. 2007	GT	58	4	192 - 197$_C$	5	0.180	0.228
Global								0.144	0.074

Table 2: Sources and properties of the microsatellite markers used in this study. $_A$One allele (186 bp) has a one base deletion in the microsatellite flanking region. $_B$One allele (11 5 bp) has a two base deletion in the microsatellite flanking region. $_C$The microsatellite in this maker is invariable, and all of the size variation in this locus is due to indels in the flanking region. $_D$The eighth repeat has a G instead of a C based on microsatellite sequences. $_E$The second last repeat has a ATG instead of a AGG motif.

separate the two copies of each microsatellite marker. Sequencing reactions were as described by Yin *et al.* (2015).

From the 33 screened markers, we selected nine markers for analyses of *D. valens* populations, because those markers amplified readily, contained microsatellite regions, and were polymorphic in *D. valens* (Table 2; see Results). The same PCR protocol with the labeled primers as with the unlabeled primers was used. Labeled amplicons were scored using GeneMarker 2.2.0 (SoftGenetics, State College, PA, USA).

To test the reliability of the microsatellite markers, we estimated the frequencies of null alleles, the presence of large allele drop-out, and errors due to stuttering using MICRO-CHECKER 2.2.3 (Van Oosterhout *et al.*, 2004). Deviations from Hardy-Weinberg equilibrium within each sampling location were calculated and we tested for linkage disequilibrium between loci using GENEPOP 4.2.2 (Rousset, 2008). Observed and expected heterozygosity of each locus was determined for every location using Arlequin 3.5 (Excoffier *et al.*, 2005). Mean F_{ST} and F_{IS} values for each locus were calculated using GenAlEx 6.5 (Peakall & Smouse, 2012).

Population structure

We tested for population structure using two complementary methods: 1) the software STRUCTURE 2.3 (Pritchard *et al.*, 2000); and 2) discriminant analysis of principal components (DAPC) implemented in the R package, adegenet (Jombart, 2008). STRUCTURE utilizes Bayesian methods to assign individuals to genetic clusters, which can be used to determine putative populations. The software calculates the likelihood that the dataset can be divided into a set number of genetic clusters (i.e., the K-value) that are in Hardy-Weinberg and linkage equilibrium, while assigning individuals within the dataset to one of the clusters, or to multiple

clusters if an individual appears to be admixed. In contrast, DAPC uses a multivariate model to assign individuals to clusters whereby the variance among clusters is maximized, and the variance within clusters is minimized. The optimal number of clusters is estimated based on the Bayesian Information Criterion (BIC), which compares the likelihoods of the different K-values within the dataset. Both STRUCTURE and DAPC allow the assignment of individuals with unknown origins to pre-determined genetic clusters, which in this study was used to estimate the ancestry of the invasive *D. valens* population in China.

STRUCTURE was first used to determine the optimal number of clusters in *D. valens*' native range in North America (i.e., excluding the Chinese samples). After repeated trials using different model settings, analyses were conducted using an admixture model for ancestry, and a correlated allele frequencies model. One million Markov Chain Monte Carlo (MCMC) replicates followed a burn-in of 300000, for K-values ranging from one to ten. One hundred iterations were run for each K-value. The STRUCTURE output was submitted to STRUCTURE HARVESTER (http://taylor0.biology.ucla.edu/structureHarvester/; Earl & vonHoldt, 2012), which estimated the most likely K-value for the dataset following the methods described in Evanno *et al.* (2005). After selecting a range of K-values (see Results), symmetric similarity coefficients were generated and the membership coefficients were averaged among the 100 runs for each selected K-value using CLUMPP 1.1.2 (Jakobsson & Rosenberg, 2007). Bar charts showing the average membership coefficients were produced using DISTRUCT 1.1 (Rosenberg, 2004). To test for further subdivisions, the above protocol was run on each individual cluster. This was repeated until there was no evidence of further subdivisions.

For comparison with the STRUCTURE results, we used DAPC to identify the optimal number of genetic clusters to which *D. valens* from North America can be assigned. Data were

transformed using principal component analyses (PCA), after which the likelihoods of the different K-values were compared using BIC. The optimal number of principal components to retain was determined by generating an α-score plot. After selecting a suitable range of K-values, DAPC was used to assign individuals to K clusters and to generate membership probabilities. DISTRUCT was used to generate bar charts showing the membership probabilities of individuals within genetic clusters.

After selecting the K-value that best explained the geographical subdivision of the North American dataset (K = 4; see Results) we assigned individuals from China to the North American clusters using STRUCTURE and DAPC. The North American individuals were first pre-specified into the geographical populations determined when K = 4. The USEPOPINFO model in STRUCTURE was then run to assign the genotypes of *D. valens* collected in China to the North American clusters, thereby estimating the origin of the Chinese population. STRUCTURE was run using the same ancestry and allele correlation models as above, along with the same number of burn-in and MCMC replicates, for 100 iterations. Average membership coefficients for the Chinese population were determined using CLUMPP, and visualized using DISTRUCT. STRUCTURE was also run using the above model parameters on the Chinese individuals alone to test for subdivision within China.

We then used the adegenet package to assign the Chinese individuals to the pre-determined North American genetic clusters. The DAPC analyses were run on the North American dataset to generate a clustering model. The Chinese *D. valens* population was then added as supplementary individuals, and the North American clustering model was used to assign the Chinese individuals to North American genetic clusters. Bar charts showing the membership probabilities were obtained using DISTRUCT.

Geographical patterns and diversity

Diversity indices were calculated for each sampling location using PopGene 1.31 (Yeh *et al.*, 1997). In addition, differentiation among sampling locations was determined by calculating pairwise F_{ST} values following Weir & Cockerham (1984) using the R packages diveRsity (Keenan *et al.*, 2013). Bias-corrected 95% confidence intervals around the F_{ST} values were calculated from 1000 bootstrap replicates. Diversity indices and pairwise F_{ST} values were also calculated on the geographical populations determined by STRUCTURE and DAPC at K = 4 (see Results).

To test for geographical patterns in the data, Nei's unbiased genetic distances among sampling locations were determined using GenAlEx. Principal coordinate analyses (PCoA) were conducted to visualize the distance data.

Invasion history

To determine the most likely source (or sources) of *D. valens* in China, the probabilities of potential invasion scenarios were compared using Approximate Bayesian Computation (ABC; Beaumont *et al.*, 2002). Analyses were conducted using the software DIYABC 2.0.4 (Cornuet *et al.*, 2014). This coalescent-based software makes it possible to define and compare a number of hypothetical demographic and evolutionary scenarios that may explain how present day populations of a study species arose from a single ancestral population. DIYABC generates simulated datasets for each scenario, and compares selected summary statistics (in this study: NAL, HET, VAR, N2P, H2P, V2P, FST; see DIYABC user manual for descriptions of each statistic) of the actual dataset to the simulated datasets in order to calculate the posterior

probabilities of the different scenarios. In addition, the ABC approach allows for the estimation of parameters that are relevant to the scenarios, such as the effective population sizes of the populations examined, the numbers of generations in the past at which historical events (i.e., changes in effective population size, divergence events or admixture between populations) occurred, rates of admixture between populations, and the mutation rates of loci examined. DIYABC software also makes it possible to scrutinize the analyses through model-checking and evaluation of confidence in scenarios.

DIYABC analyses were conducted using eight of the microsatellite loci, because one of the loci (DV-11; Table 2) did not vary in size because of variation in the microsatellite region. DIYABC requires that size variation be in the microsatellite region for estimates of mutation rates. In the analysis, separate mutation rates were obtained for markers with di- and tri-nucleotide repeats. The ranges of the parameter prior distributions for the analyses are listed in Table S9.

A step-wise approach was used to determine the most likely invasion history of *D. valens* in China using DIYABC. In the first step, we compared five different scenarios regarding the most likely origin of the majority of *D. valens* that invaded China. The North American dataset was divided into four genetically- and geographically-distinct populations based on the results of the clustering analyses: West, Central, Northeast, and Mexico (Fig. 2; see Results). Because the relationships between the four populations were indeterminate, the four populations were assumed to have arisen at the same time as a polytomy from an ancestral population in all scenarios. The scenarios are listed in Table S8.

In the second step, we built upon the best (i.e., the most probable) scenario from the first step. Four scenarios were compared to test the hypothesis that the population in China arose from

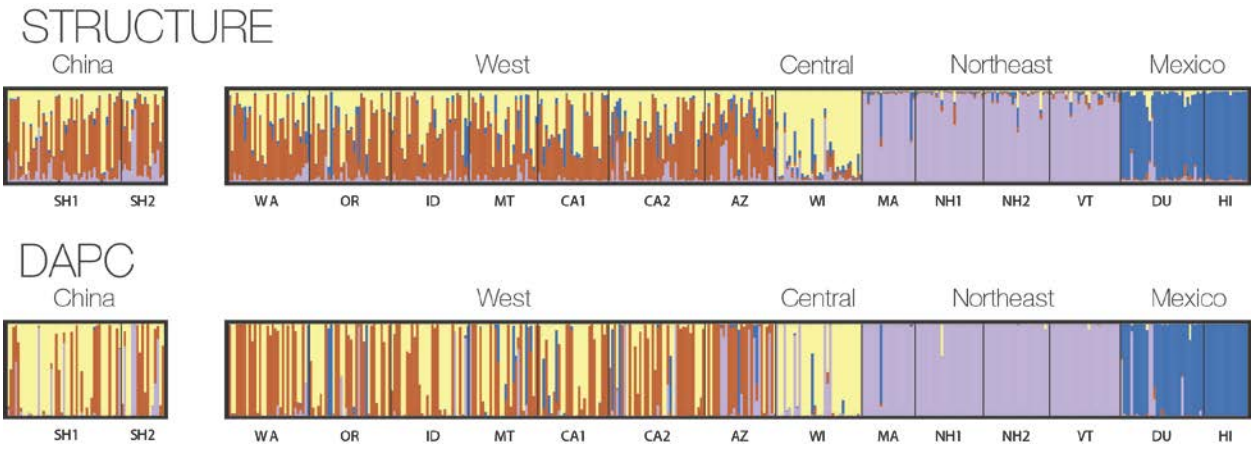

Fig. 2: Bar plots of the average membership coefficients (STRUCTURE) and the membership probabilities (DAPC) of the Dendroctonus valens individuals collected in China as well as North America for K = 4. The genetic clusters are represented by purple bars, brown bars, blue bars, yellow bars, and grey bars.

admixed North American populations. In scenario 1.1, the invasive population arose only from the source with the highest probability (the West; see Results) in step one, while in the remaining scenarios the invasive population arose from admixture between the source with the highest probability and another population (Table S9).

After the best scenario(s) was determined in step two, the means, medians, modes and 95% confidence intervals of the parameters were estimated. Because *D. valens* has between one and three generations per year (depending on ambient temperature; Sun *et al.*, 2013; Yan *et al.*, 2005), a possible range of times in years was estimated for the timing of the historical events.

The type I and II errors were calculated for the best scenarios in each step to evaluate the confidence in scenario choice. Five hundred datasets were simulated for each scenario. The proportion of times the selected scenario did not have the highest posterior probability when it was in fact the true scenario (i.e., type I error) was calculated, along with the number of times the selected scenario had the highest posterior probability when it was in fact the false scenario (type II error). In addition, model checking was conducted using summary statistics that were not utilized in the initial analyses (MGW, LIK, DAS, DM2, AML; see DIYABC user manual). Model checking is a way to test the goodness-of-fit of a model (i.e. scenario) by comparing the observed dataset with the posterior predictive distribution of the scenario model (Cornuet *et al.*, 2010).

Results

Collections

In total, 464 *D. valens* specimens were collected: 50 from Mexico (see Table 1 for breakdown by sampling location), 352 from the United States, and 62 from China.

Microsatellites

Of the 33 microsatellite markers screened, nine loci were polymorphic based on size (Table 2; GenBank accession numbers KU183606 - KU183614). One locus (DV-11) varied in size because of insertions in the flanking regions rather than in the microsatellite. However, this marker was retained for the majority of analyses because it remained informative (F_{ST} = 0.1660). In addition, an allele of DV-05 had a single base deletion in the flanking region. Finally, an allele of DV-10 had a two base deletion in the flanking region. In the remaining loci, all of the size variation was due to repeat variation in the microsatellite region.

Six loci (DV-01, DV-04, DV-05, DV-07, DV-09 and DV-11) violated the assumption of Hardy-Weinberg equilibrium in at least one of the 16 sample locations (Table S1). In addition, there was evidence of null alleles in at least one sample location for four markers (DV-01, DV-05, DV-09 and DV-11), and scoring error due to stutter in at least one sample location for three markers (DV-01, DV-05 and DV-11) based on MICRO-CHECKER. However, the violations of Hardy-Weinberg equilibrium were minor, and the proportions of predicted null alleles were small. In addition, microsatellites were scored a second time to ensure that no scoring errors occurred because of stuttering. None of the pairs of loci showed linkage disequilibrium (at $p < 0.01$).

Clustering

According to the ΔK value (Evanno *et al.*, 2005), the most likely number of genetic clusters based on STRUCTURE analyses of *D. valens* in North America was three (Fig. S1). However, there was a relatively small range in ΔK between K = 2 and K = 5. In addition, the plot of the posterior probabilities, or L(K), for each K-value did not plateau until K = 5 (Fig. S1). We, therefore, analyzed the dataset at all four K-values from two to five. The symmetric similarity coefficients between STRUCTURE runs were larger than 0.90 from K = 2 to K = 5, therefore all runs were included in the subsequent analyses.

For the DAPC analyses, we determined that the K-value with the lowest BIC was 11 (Fig. S2a), and retained 24 principal components based on the α-score plot (Fig. S2b). The optimal K-value was not clear based on the relationship between K and BIC, therefore, a range of K-values from K = 2 (the lowest number of clusters tested with STRUCTURE) to K = 5 (the point at which BIC stops sharply declining with additional genetic clusters) were selected. The mean membership coefficients from STRUCTURE and the mean membership probabilities from DAPC at each sampling location are summarized for each K-value in Table S2.

After comparing the analyses from STRUCTURE and DAPC (summarized for K = 2 to K = 5 in Fig. S3), we selected four as the most likely number of genetic clusters of *D. valens* in North America. There was no evidence for further subdivision based on STRUCTURE analyses (Fig. S4). The four genetic clusters (represented by yellow, purple, brown, and blue bars) largely supported four geographically-distinct populations: the West population (which was largely assigned to the brown cluster) contained the individuals from AZ, CA1, CA2, ID, MT, OR, and WA, the Central population (yellow cluster) contained the individuals from WI, the Northeast population (purple cluster) contained the individuals from MA, NH1, NH2 and VT, and the Mexico population (blue cluster) contained the individuals from DU and HI. The mean

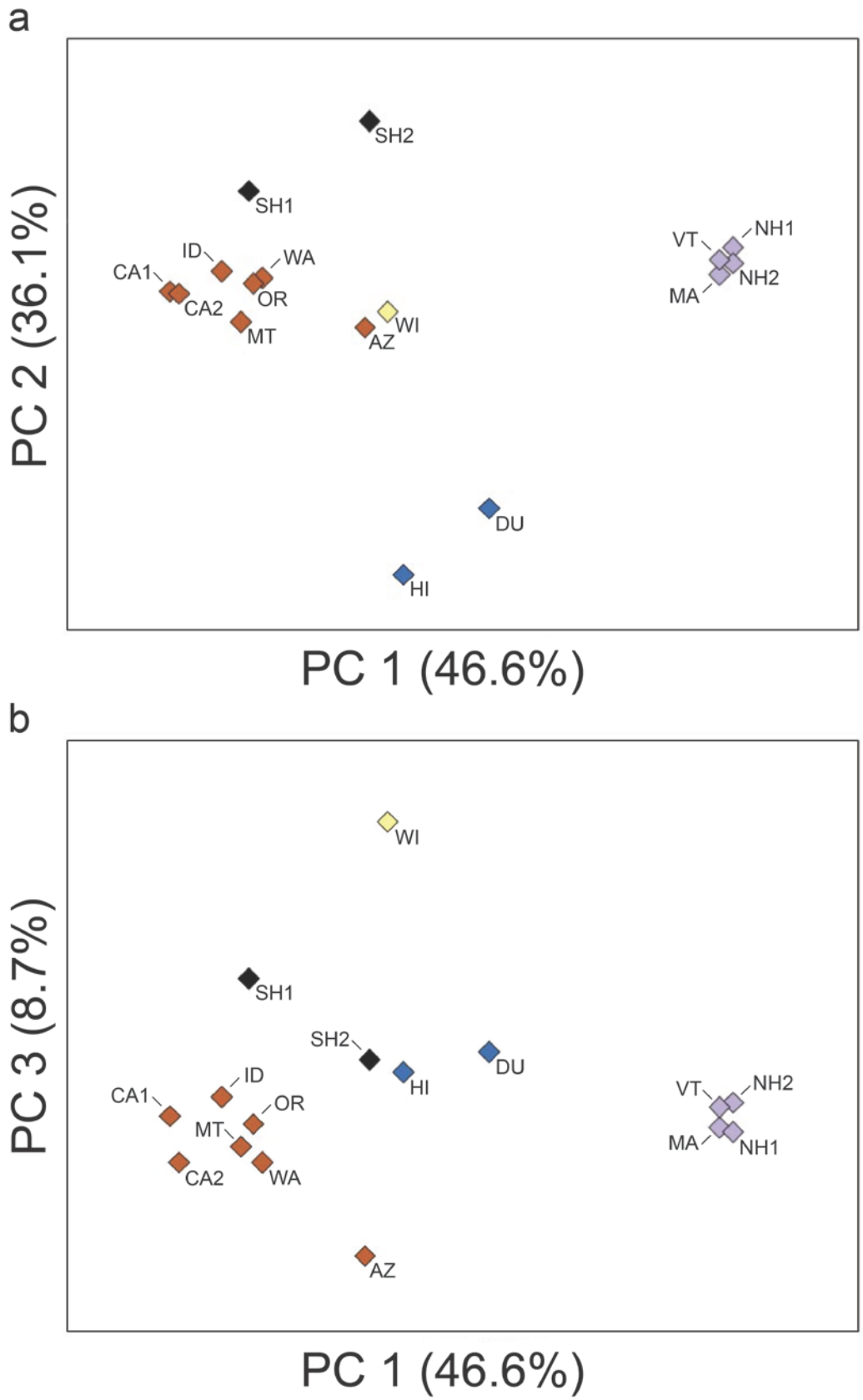

Fig. 3: Principal coordinate analysis plots showing the distance among the sampling locations based on Nei's genetic distance. The diamonds represent sampling locations, and the colours of the circles represent the population to which the sampling locations belong (based on the genetic clusters shown in Fig. 2): brown for the West population, yellow for the Central population, purple for the Northeast population, blue for the Mexico population, and black for the China population. a) The first two principal coordinates. b) The first and third principal coordinates.

membership coefficients and membership probabilities for each population, as well as the percentages of individuals from each population that were clearly assigned (i.e., the membership coefficients or membership probabilities > 0.8) or mostly assigned (the membership coefficients or membership probabilities > 0.5) to each cluster are given in Table S3.

Using the above genetic groupings, STRUCTURE clearly assigned 24.2% of individuals from China to the brown cluster (Fig. 3, Table S3), while it assigned 11.3% of individuals to the yellow cluster. No individuals were clearly assigned to the purple or blue clusters. The average membership coefficients within the China population were most similar to those from the West population as the average membership coefficients were 0.523 for the brown cluster (compared to 0.532 in the West population) and 0.332 for the yellow cluster (compared to 0.339 in the West population). In contrast, DAPC clearly assigned 53.2% of the individuals from China to the yellow cluster, while it assigned 21.0% to the brown cluster. The average membership probabilities of the China population were intermediate between the West and Central populations as they were 0.586 for the yellow cluster (compared to 0.379 in the West and 0.775 in the Central population) and 0.285 for the brown cluster (compared to 0.482 in the West and 0.008 in the Central population). There were no subdivisions within the China population based on STRUCTURE (Fig. S4).

Geographical patterns and diversity

The West population had the largest mean number of alleles (N_A = 6.778; Table 3), followed by the Northeast, Mexico, China, and Central populations. In addition, the West population had the highest mean expected heterozygosity (H_E = 0.55677), followed by the

Table 3: Diversity statistics for the different populations of *Dendroctonus valens*. A = mean number of alleles, EA = effective number of alleles, H = Shannon's information index, ML = monomorphic loci, PA = private allele, Ho = observed heterozygosity, He = expected heterozygosity, s.d. = standard deviation.

	A	s.d.	EA	s.d.	H	s.d.	ML	PA	Mean Ho	s.d.	Mean He (unbiased)	s.d.	Mean He (Nei 1973)	s.d.
China	4.333	2.121	2.206	0.987	0.888	0.476	0	0	0.446	0.219	0.468	0.224	0.464	0.224
West	6.778	3.898	3.102	1.900	1.150	0.641	0	7	0.496	0.265	0.557	0.281	0.556	0.281
Central	3.111	1.616	1.885	0.814	0.672	0.487	2	0	0.371	0.284	0.379	0.268	0.374	0.264
Northeast	5.778	5.142	2.278	1.978	0.779	0.730	1	8	0.351	0.299	0.362	0.309	0.360	0.307
Mexico	4.111	2.315	2.268	1.280	0.823	0.596	1	2	0.378	0.279	0.434	0.292	0.430	0.289

Mexico, Central, China, and Northeast populations. Diversity indices of sampling locations are summarized in Table S4.

All of the alleles found in the China populations were also present in the West population (Table S5), while fewer alleles were shared between China and the Northeast (79.5% of alleles present in China were also present in the Northeast), Mexico (74.4%) and Central (64.1%) populations. Within China, SH1 and SH2 shared 25 alleles (Table S6), while 13 alleles were found only in SH1 and one allele was found only in SH2.

Based on pairwise F_{ST} values, the China population was the least differentiated from the West population (F_{ST} = 0.0551; Table 4), followed by the Central population (F_{ST} = 0.1207), the Northeast population (F_{ST} = 0.2024), and the Mexico population (F_{ST} = 0.2499). Within North America, the Central and West populations were the least differentiated (F_{ST} = 0.0740), while the Mexico and Northeast populations were the most differentiated (F_{ST} = 0.2090).

Within the China population, the genetic differentiation between sampling locations was low as the pairwise F_{ST} between SH1 and SH2 was 0.0239 (Table S7). Within the Northeast population and within the West population, the majority of pairwise F_{ST} values were less than 0.0100. The major exceptions were the pairwise F_{ST} values between AZ and the other sampling locations within the West, as the pairwise F_{ST} values ranged from 0.176 (AZ and MT) to 0.0448 (AZ and CA1). Within the Mexico cluster, the pairwise F_{ST} value between DU and HI was 0.0460.

The PCoA analyses supported the geographical clustering of STRUCTURE and DAPC, as well as the pairwise F_{ST} values (Fig. 4). The sampling locations within the Northeast population were well separated from the remaining sampling locations based on the first principal coordinate (which explained 46.6% of the total variation), the sampling locations

Table 4: Pairwise FST values among populations following Weir & Cockerham (1984). The upper right corner shows the 95% confidence intervals.

	China	West	Central	Northeast	Mexico
China		0.0432 to 0.0680	0.0873 to 0.1561	0.1773 to 0.2271	0.2211 to 0.2803
West	0.0551		0.0740 to 0.1162	0.1505 to 0.1797	0.1174 to 0.1565
Central	0.1207	0.094		0.1622 to 0.2441	0.1563 to 0.2259
Northeast	0.2024	0.1649	0.2037		0.2090 to 0.2823
Mexico	0.2499	0.1368	0.1912	0.2443	

within the Mexico population were well separated from the others based on the second principal coordinate (36.1%), and the sampling location within the Central population was well separated from the others based on the third principal coordinate (8.7%). The samples from China were closest to the West population based on all three principal coordinates.

Invasion history

In step one, scenario 1 (i.e., the China population originated from the West population) had the highest posterior probability (p = 0.9675, 95% C.I. = 0.9479 - 0.9870; Table S8). In step two, scenario 1.1 (i.e., no admixture occurred) had the highest posterior probability (p = 0.5103, 95% C.I. = 0.4170 - 0.6035). However, scenario 1.2 in step two (i.e., the population in China was the result of admixture between the West and Central populations) had a fairly high posterior probability (p = 0.3926, 95% C.I. = 0.2913 - 0.4940).

The mean, median and mode of the posterior probability distributions of most parameters in the best scenario for step two (scenario 1.1) were very similar, except for the estimates of the effective population size of the ancestral population (AN; the estimated parameters are listed in Table S9). The West population had the highest estimated effective population size, followed by the Northeast, Mexico, Central, and ancestral populations respectively. The introduction to China was estimated to have occurred 226 generations ago (95% C.I. = 40.2 - 385). The estimated time range for the invasion of China by *D. valens* was between 75 (95% C.I. = 13 - 128) and 226 (95% C.I. = 40.2 - 385) years ago. The parameter estimates were very similar between scenarios 1.1 and 1.2. The admixture rate from the West cluster was estimated to be 0.840 (95% C.I. = 0.460 - 0.970), suggesting that the admixture rate from the Central cluster was ~0.160.

In step one, the best scenario (scenario 1) had low type I and II errors (0.088 and 0.009 respectively), supporting its selection as the most probable scenario. However, in step two, the best scenario (scenario 1.1) had a fairly high type I error (0.212), suggesting that when the best scenario was simulated, there was a reasonably good chance that a different scenario would be selected. Scenario 1.1 had a low type II error (0.085).

Based on model checking for the scenario selected in step two, only three out of 96 statistics had an excessively low probability (Table S10). In addition, according to the model checking principal component analysis plot, the actual dataset was close to the middle of the group of datasets predicted from the posterior distribution (Fig. S5). The best scenario therefore appeared to fit the observed data.

Discussion

The results of this study show that the main source of *D. valens* in China is most likely the West population sampled in western North America. These results, based on an extensive sampling of *D. valens*, generally support the conclusions of Cognato *et al.* (2005) and Cai *et al.* (2008), as well as historical records of timber imports into China (Yan *et al.*, 2005). The results also reveal that the native range of *D. valens* can be divided into at least four genetically- and geographically-distinct clusters.

The China population of *D. valens* had lower diversity than the West population, which is consistent with most invasive populations that have undergone a genetic bottleneck (Estoup & Guillemaud, 2010). A similar bottleneck was detected by Cai *et al.* (2008). Interestingly, the results of the present study showed the diversity of the China population was high relative to the other invasive populations, suggesting that the founding population was large, or that the

invasive population is the result of multiple introduction events. Within China, the two sampling locations were somewhat differentiated based on pairwise F_{ST} values. This could be evidence of gene surfing (Hallatschek & Nelson, 2008) following the range expansion of the beetle after its introduction.

The population of *D. valens* from western North America is the most parsimonious source of the insect in China because the majority of exports from North America to Asia are shipped from western North America. However, this study did not disprove the possibility that some invasive propagules could have originated from an additional population. Direct trade also occurs between China and both eastern North America and Mexico. In addition, substantial trade occurs within North America. Logs from eastern North America or Mexico that were infested with *D. valens* could have been moved to China via western North America. If admixture occurred between beetles from genetically distinct populations, this would have contributed to the high genetic diversity of *D. valens* in China relative to the Northeast, Central, or Mexico populations. However, previous breeding experiments with *D. valens* have shown that although *D. valens* from eastern North America (specifically from New York state, which we predict is part of the Northeast population) will readily mate with *D. valens* from western North America (from Arizona and California), the beetles will not produce brood, at least under the experimental conditions used in the breeding study (Pajares & Lanier, 1990). The results of Pajares & Lanier (1990) suggest that admixture between the different populations is unlikely, although their experiments did not include beetles from the Central or Mexico populations.

The introduction of beetles from the Central or Northeast populations along with beetles from the West would provide a possible explanation for the association between *D. valens* and *L. procerum* in China. This is because *L. procerum* is associated with the beetle in both the Central

and Northeast populations (Taerum *et al.*, 2013). The membership probabilities of the Chinese population based on DAPC could indicate that admixture occurred between the West and Central populations. This is because the membership probabilities of *D. valens* in China for the brown and yellow clusters were intermediate between the West and Central populations (Table S3). In addition, the DIYABC scenario where admixture occurred between the West and Central populations to form the China population (step two, scenario 1.2) had relatively strong support. However, the strong support for this scenario could have arisen from the fact that the West and Central populations are genetically similar. Alternatively, if the source population for *D. valens* was the West population as this study suggests, the observed membership probabilities in China, as well as the non-negligible support for scenario 1.2, could be due to genetic drift. Even if the *D. valens* population currently in China originated entirely from the West population, we cannot discount the possibility that some *D. valens* were introduced from eastern North America carrying spores of *L. procerum*. A small number of beetles may have originated from eastern North America, after which the fungus might have undergone a vector jump onto *D. valens* originating from the West.

If introductions from the eastern North American range did not occur, alternative explanations for the association between *D. valens* and *L. procerum* in China must be considered. For example, *L. procerum* may in fact be present in western North America, but is rare or absent in the locations from which *D. valens* was collected in the study by Taerum *et al.* (2013). However, if *L. procerum* is an infrequent associate of *D. valens* in western North America, it seems unlikely that this fungus would have successfully co-invaded China rather than more common fungal associates of *D. valens* from western North America. For example, *Leptographium* sp. 1 and *Ophiostoma* sp. 1 made up 36.3% and 33.1% respectively of the fungi

isolated from *D. valens* in western North America but they have not been found associated with the insect in China. This is unless *L. procerum* has a greater level of fitness in China's forests relative to the other fungal associates.

Another possible explanation for the association between *D. valens* and *L. procerum* in China is that *L. procerum* was already present in China, and made a host jump onto *D. valens* after the beetle became established in China. Although *L. procerum* has never been reported from China except as a symbiont of *D. valens* (Lu *et al.*, 2009a,b), several species in the *L. procerum* species complex appear to be indigenous to Asia. This suggests that the continent may be the centre of diversity for the species complex (Yin *et al*., 2015), and potentially the origin of *L. procerum*. In addition, the surveys in China have tended to focus on associates of especially damaging forest insects (Zhou *et al.*, 2013), while *L. procerum* is more commonly associated with relatively non-aggressive bark beetles (Jacobs & Wingfield, 2001).

Some putative isolates of *L. procerum* have been identified in Europe and Japan (Jankowiak, 2012; Masuya *et al.*, 2013). Most studies have based identifications on morphology and/or the internal transcribed spacer sequence, which is invariable for most species within the *L. procerum* species complex (Yin *et al*., 2015). However, recent studies have confirmed the presence of *L. procerum* in Poland by using sequences of the beta-tubulin region (Jankowiak & Bilański, 2013a,b,c), which is sufficient to differentiate *L. procerum* from the other fungi in the *L. procerum* species complex. Additional surveys of ophiostomatalean fungi associated with bark beetles in Asia and North America, as well as population genetics analyses on the currently collected fungi may help clarify the origin of *L. procerum* associated with *D. valens* in China.

This study presents evidence that the North American range of *D. valens* can be divided into at least four geographically- and genetically-distinct populations. More populations may

exist, especially in areas that were unsampled in this study such as Canada, Central America, and the southern Appalachian Mountains. The West population was the most diverse, supporting the findings of Cai *et al.* (2008). Our estimation of population diversity may have been biased for the West population because this population included the most sampling locations (nine) and covered the largest geographical range. However, the diversities of individual sampling locations tended to be higher within the West than within other populations supporting a higher overall diversity within the West population compared to other populations.

The results of this study supported the existence of barriers to gene flow between the four populations. The two nearest sampling locations in the Mexico (DU) and West (AZ) populations are separated by the Sierra Madre Occidental mountain range, which has been demonstrated to be a barrier to gene flow for other species (Pfeiler *et al.*, 2013; Ruiz-Sanchez & Specht, 2013). This finding is largely in agreement with the results of Cognato *et al.* (2005) and Cai *et al.* (2008), as their studies supported the placement of most *D. valens* haplotypes from central and southern Mexico into separate clades from the other populations. Additional collections of *D. valens* between the two locations are required to determine the exact location of the boundary between the two populations, and if any admixture occurs between the West and Mexico. In addition, additional collections within central and southern Mexico are required to adequately understand the diversity of *D. valens* in this location.

Surprisingly, the Northeast population of *D. valens* was very divergent from the Central population, despite the presence of pine forests connecting the two populations. The high differentiation between the two populations may be due to colonization from different glacial refugia. Similar patterns of divergence have been observed with other organisms in eastern North America (e.g., Zamudio & Savage, 2003; Griffin & Barrett, 2004 because of colonization from

different refugia around the Appalachian mountains. The two populations may be genetically incompatible due to isolation, and restricted to their ranges by competitive exclusion or adaptation to specific hosts or other environmental conditions. Additional samples collected between WI and the sampling locations within the Northeast population will clarify the boundary between the two populations. Alternatively, the two populations may be highly differentiated because they represent collections from different edges of the eastern North American range, in which case the observed geographical structure may be due to a high level of isolation by distance. Samples collected between the two populations should therefore have intermediate genotypes to the ones observed in the range edges. This is an unlikely explanation, however, as the Northeast and Central populations showed very high genetic divergence relative to the furthest apart sampling locations in the West population (AZ and WA), which were separated by a comparable distance (~1500 km).

The relatively low divergence between the West and Central populations, despite their separation by the Great Plains was unexpected. The Great Plains has been demonstrated to act as a barrier to gene flow for other species of forest inhabiting organisms (Hamelin *et al.*, 2000; Kelly & Hutto, 2005). However, the study results may indicate gene flow between the two locations. This appears to contradict the findings of Cognato *et al.* (2005) and Cai *et al.* (2008), who indicated a high degree of divergence between the mitochondrial sequences of their samples from western North American and Michigan, which is relatively close (~400 km) to our collection site, WI. It is unknown whether beetles from their Michigan site and our WI sampling location belong to the same population based on shared mitochondrial sequences and microsatellite genotypes, or if the two locations are separated by a barrier to gene flow.

Additional analyses of both the mitochondrial and microsatellite diversity of *D. valens* in and around WI are needed to clarify the diversity of the Central population.

Dendroctonus valens from the West and Central populations may be genetically similar because of migration across the Great Plains, as bark beetles can be moved long distances by wind (Byers, 2000). In addition, small plots of pine trees or other conifers in the Great Plains may act as a corridor between the two populations. Also, the Boreal Forest may be a corridor between the two populations. Another possible explanation for the genetic similarity between the West and Central populations is that *D. valens* from one population was introduced to the other via the movement of plant material by humans. The accidental transport of forest insects within North America has been observed before, as human movement of fire wood has greatly contributed to the range expansion of the emerald ash borer (Poland & McCullough, 2006). Human movement of woody material has already resulted in the successful establishment of a highly diverse *D. valens* population in China (Table 3). It is thus not inconceivable that one of the observed *D. valens* populations in North America was recently established from a population elsewhere in the country. Finally, the genetic similarities between the West and Central populations could be due to recent divergence, if the two populations were established, for example, from the same glacial refugia. Coalescent analyses would help clarify the genetic divergence between *D. valens* populations in the West and Central populations, as well as the evolutionary history of the bark beetle in North America in general.

This study generally supported the previous population genetics studies on *D. valens* in concluding that western North America (specifically the West population) is the source of the beetle in China, while providing more data on the diversity of the insect in China. The results suggest that the West population is a potential source for future invasions, along with the

individuals already present in China. In addition, our findings provided details regarding the diversity and genetic structure of *D. valens* in North America. The geographical patterns observed in North America are most likely the result of geographical barriers to gene flow combined with recolonization from distinct glacial refugia. Future work should aim to clarify the origin of the fungal symbiont, *L. procerum*, associated with *D. valens*. Additional research should be conducted on *D. valens* in North America to expand the available body of knowledge regarding the evolutionary history of this insect within its native range.

Acknowledgements

We thank Carline Carvalho, Kevin Dodds, Nancy Gillette, Rich Hofstetter, Lu Min, Ken Raffa, Diana Six, Omar Alejandro Pérez Vera, and numerous other forestry researchers across the United States, Mexico and China who provided invaluable support in the collection of *D. valens* and without which this study would not have been possible. We also thank Felix Sperling and two anonymous reviewers for their constructive comments. Financial support was provided by the Department of Science and Technology (DST)/National Research Foundation (NRF) Centre of Excellence in Tree Health Biotechnology. This project was supported by multiple grants from the NRF, South Africa, including the grant specific unique reference number (UID) 83924. The grant holders acknowledge that opinions, findings and conclusions or recommendations expressed in publications generated by NRF supported research are that of the authors, and the NRF accepts no liability whatsoever in this regard.

References

Aukema JE, Leung B, Kovacs K, Chivers C, Britton KO, Englin J, Frankel SJ, Haight RG, Holmes TP, Liebhold AM, McCullough DG, Von Holle B. 2011. Economic impacts of non-native forest insects in the continental United States. *Plos One* **6:** e24587.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* **162:** 2025-2035.

Bertheau C, Schuler H, Krumbock S, Arthofer W, Stauffer C. 2011. Hit or miss in phylogeographic analyses: the case of the cryptic NUMTs. *Molecular Ecology Resources* **11:** 1056-1059.

Boissin E, Hurley B, Wingfield MJ, Vasaitis R, Stenlid J, Davis C, de Groot P, Ahumada R, Carnegie A, Goldarazena A, Klasmer P, Wermelinger B, Slippers B. 2012. Retracing the routes of introduction of invasive species: the case of the *Sirex noctilio* woodwasp. *Molecular Ecology* **21:** 5728-5744.

Borden WC, Stepien CA. 2006. Discordant population genetic structuring of smallmouth bass, *Micropterus dolomieu* Lacepede, in Lake Erie based on mitochondrial DNA sequences and nuclear DNA microsatellites. *Journal of Great Lakes Research* **32:** 242-257.

Brito PH, Edwards SV. 2008. Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetics* **135:** 439-455.

Byers JA. 2000. Wind-aided dispersal of simulated bark beetles flying through forests. *Ecological Modelling* **125:** 231-243.

Cai Y, Cheng XY, Duan D, Xu R. 2011. Mitochondrial COI gene transfers to the nuclear genome of *Dendroctonus valens* and its implications. *Journal of Applied Entomology* **135:** 302-310.

Cai YW, Cheng XY, Xu RM, Duan DH, Kirkendall LR. 2008. Genetic diversity and biogeography of red turpentine beetle *Dendroctonus valens* in its native and invasive regions. *Insect Science* **15:** 291-301.

Ciosi M, Miller NJ, Kim KS, Giordano R, Estoup A, Guillemaud T. 2008. Invasion of Europe by the western corn rootworm, *Diabrotica virgifera virgifera*: multiple transatlantic introductions with various reductions of genetic diversity. *Molecular Ecology* **17:** 3614-3627.

Cognato AI, Sun JH, Anducho-Reyes MA, Owen DR. 2005. Genetic variation and origin of red turpentine beetle (*Dendroctonus valens* LeConte) introduced to the People's Republic of China. *Agricultural and Forest Entomology* **7:** 87-94.

Cornuet J-M, Ravigné V, Estoup A. 2010. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *Bioinformatics* **11:** 401.

Cornuet J-M, Puldo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J-M, Estoup A. 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30:** 1187-1189.

Davis CS, Mock KE, Bentz BJ, Bromilow SM, Bartell NV, Murray BW, Roe AD, Cooke JEK. 2009. Isolation and characterization of 16 microsatellite loci in the mountain pine beetle, *Dendroctonus ponderosae* Hopkins (Coleoptera: Curculionidae: Scolytinae). *Molecular Ecology Resources* **9:** 1071-1073.

Dlugosch KM, Parker IM. 2008. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* **17:** 431-449.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4:** 359-361.

Erbilgin N, Mori SR, Sun J-H, Stein JD, Owen DR, Merrill LD, Bolanos RC, Raffa KF, Montiel TM, Wood DL, Gillette NE. 2007. Response to host volatiles by native and introduced populations of *Dendroctonus valens* (Coleoptera: Curculionidae, Scolytinae) in North America and China. *Journal of Chemical Ecology* **33**: 131-146.

Estoup A, Guillemaud T. 2010. Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology* **19:** 4113-4130.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611-2620.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1:** 47-50.

Griffin SR, Barrett SCH. 2004. Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in Eastern North America: Inferences from phylogeography. *American Journal of Botany* **91:** 465-473.

Haack RA, Rabaglia RJ. 2013. Exotic bark and ambrosia beetles in the USA: potential and current invaders. In: Peña J, ed. *Potential Invasive Pests of Agricultural Crops*. Wallingford, UK: CABI, 48-74

Hallatschek O, Nelson DR. 2008. Gene surfing in expanding populations. *Theoretical Population Biology* **73:** 158-170.

Hamelin RC, Hunt RS, Geils BW, Jensen GD, Jacobi V, Lecours N. 2000. Barrier to gene flow between eastern and western populations of *Cronartium ribicola* in North America. *Phytopathology* **90:** 1073-1078.

Hulcr J, Dunn RR. 2011. The sudden emergence of pathogenicity in insect–fungus symbioses threatens naive forest ecosystems. *Proceedings of the Royal Society B* **278:** 2866-2873.

Hulme PE, Bacher S, Kenis M, Klotz S, Kuehn I, Minchin D, Nentwig W, Olenin S, Panov V, Pergl J, Pyšek P, Roques A, Sol D, Solarz W, Vilà M. 2008. Grasping at the routes of biological invasions: a framework for integrating pathways into policy. *Journal of Applied Ecology* **45:** 403-414.

Jacobs K, Wingfield MJ. 2001. Leptographium *Species - Tree Pathogens, Insect Associates, and Agents of Blue-Stain*. St. Paul, MN, USA: APS Press.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:** 1801-1806.

Jankowiak R. 2012. Ophiostomatoid fungi associated with *Ips sexdentatus* on *Pinus sylvestris* in Poland. *Dendrobiology* **68:** 43-53.

Jankowiak R, Bilański P. 2013a. Diversity of ophiostomatoid fungi associated with the large pine weevil, *Hylobius abietis*, and infested Scots pine seedlings in Poland. *Annals of Forest Science* **70:** 391-402.

Jankowiak R, Bilański P. 2013b. Association of the pine-infesting *Pissodes species* with ophiostomatoid fungi in Poland. *European Journal of Forest Research* **132:** 523-534.

Jankowiak R, Bilański P. 2013c. Ophiostomatoid fungi associated with root-feeding bark beetles on Scots pine in Poland. *Forest Pathology* **43:** 422-428.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24:** 1403-1405.

Jordal BH, Kambestad M. 2014. DNA barcoding of bark and ambrosia beetles reveals excessive NUMTs and consistent east-west divergence across Palearctic forests. *Molecular Ecology Resources* **14:** 7-17.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodoehl PA. 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* **4:** 782-788.

Kelly JF, Hutto RL. 2005. An east-west comparison of migration in North American wood warblers. *Condor* **107:** 197-211.

Konečný A, Estoup A, Duplantier JM, Bryja J, Bâ K, Galan M, Tatard C, Cosson J-F. 2013. Invasion genetics of the introduced black rat (*Rattus rattus*) in Senegal, West Africa. *Molecular Ecology* **22:** 286-300.

Liebhold AM, Brockerhoff EG, Garrett LJ, Parke JL, Britton KO. 2012. Live plant imports: the major pathway for forest insect and pathogen invasions of the US. *Frontiers in Ecology and the Environment* **10:** 135-143.

Lu M, Wingfield MJ, Gillette NE, Sun J-H. 2011. Do novel genotypes drive the success of an invasive bark beetle-fungus complex? Implications for potential reinvasion. *Ecology* **92:** 2013-2019.

Lu M, Zhou XD, De Beer ZW, Wingfield MJ, Sun J-H. 2009a. Ophiostomatoid fungi associated with the invasive pine-infesting bark beetle, *Dendroctonus valens*, in China. *Fungal Diversity* **38:** 133-145.

Lu Q, Decock C, Zhang XY, Maraite H. 2009b. Ophiostomatoid fungi (Ascomycota) associated with *Pinus tabuliformis* infested by *Dendroctonus valens* (Coleoptera) in northern China and an assessment of their pathogenicity on mature trees. *Antonie van Leeuwenhoek* **96:** 275-293.

Maroja LS, Bogdanowicz SM, Wallin KF, Raffa KF, Harrison RG. 2007. Phylogeography of spruce beetles (*Dendroctonus rufipennis* Kirby) (Curculionidae: Scolytinae) in North America. *Molecular Ecology* **16:** 2560-2573.

Masuya H, Yamaoka Y, Wingfield MJ. 2013. Ophiostomatoid fungi and their associations with bark beetles in Japan. In: Seifert KA, De Beer ZW and Wingfield MJ, eds. *The Ophiostomatoid Fungi: Expanding Fronteirs*. CBS-KNAW Fungal Biodiversity Centre: Utrecht, The Netherlands, 77-90.

Möller EM, Bahnweg G, Sandermann H, Geiger HH. 1992. A simple and efficient protocol for isolation of high-molecular-weight DNA from filamentous fungi, fruit bodies, and infected-plant tissues. *Nucleic Acids Research* **20:** 6115-6116.

Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **70:** 3321-3323.

Owen DR, Smith SL, Seybold SJ. 2010. Red turpentine beetle. *US Department of Agriculture, Forest Pest Leaflet* **55**.

Pajares JA, Lanier GN. 1990. Biosystematics of the turpentine beetles *Dendroctonus terebrans* and *D. valens* (Coleoptera, Scolytidae). *Annals of the Entomological Society of America* **83:** 171-188.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28:** 2537-2539.

Pfeiler E, Richmond MP, Riesgo-Escovar JR, Tellez-Garcia AA, Johnson S, Markow TA. 2013. Genetic differentiation, speciation, and phylogeography of cactus flies (Diptera: Neriidae: *Odontoloxozus*) from Mexico and south-western USA. *Biological Journal of the Linnean Society* **110:** 245-256.

Ploetz RC, Hulcr J, Wingfield MJ, de Beer ZW. 2013. Destructive tree diseases associated with ambrosia and bark beetles: black swan events in tree pathology? *Plant Disease* **97:** 856-872.

Poland TM, McCullough DG. 2006. Emerald ash borer: invasion of the urban forest and the threat to North America's ash resource. *Journal of Forestry* **104:** 118-124.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945-959.

Pyšek P, Jarošík V, Hulme PE, Kűhn I, Wild J, Arianoutsou M, Bacher S, Chiron F, Didžiulis V, Essl F, Genovesi P, Gherardi F, Hejda M, Kark S, Lambdon PW, Desprez-Loustau M-L, Nentwig W, Pergl J, Poboljšaj K, Rabitsch W, Roques A, Roy DB, Shirley S, Solarz W, Vilà M, Winter M. 2010. Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 12157-12162.

Roderick GK, Navajas M. 2003. Genes in new environments: genetics and evolution in biological control. *Nature Reviews Genetics* **4:** 889-899.

Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4:** 137-138.

Rousset F. 2008. GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8:** 103-106.

Ruiz-Sanchez E, Specht CD. 2013. Influence of the geological history of the Trans-Mexican Volcanic Belt on the diversification of *Nolina parviflora* (Asparagaceae: Nolinoideae). *Journal of Biogeography* **40:** 1336-1347.

Schrey NM, Schrey AW, Heist EJ, Reeve JD. 2007. Microsatellite loci for the southern pine beetle (*Dendroctonus frontalis*) and cross-species amplification in *Dendroctonus*. *Molecular Ecology Notes* **7:** 857-859.

Slippers B, Hurley BP, Wingfield MJ. 2015. *Sirex* woodwasp: a model for evolving management paradigms of invasive forest pests. *Annual Review of Entomology* **60:** 601-619

Sun J-H, Lu M, Gillette NE, Wingfield MJ. 2013. Red turpentine beetle: innocuous native becomes invasive tree killer in China. *Annual Review of Entomology* **58:** 293-311.

Sunnucks P. 2000. Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15:** 199-203.

Taerum SJ, Duong TA, de Beer ZW, Gillette N, Sun J-H, Owen DR, Wingfield MJ. 2013. Large shift in symbiont assemblage in the invasive red turpentine beetle. *Plos One* **8:** e78126.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4:** 535-538.

Weed AS, Ayres MP, Hicke JA. 2013. Consequences of climate change for biotic disturbances in North American forests. *Ecological Monographs* **83:** 441-470.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38:** 1358-1370.

Wingfield MJ, Seifert KA, Webber JF. 1993. *Ceratocystis and Ophiostoma: Taxonomy, Ecology, and Pathogenicity*. St. Paul, MN, USA: APS Press.

Wingfield MJ, Brockerhoff EG, Wingfield BD, Slippers B. 2015. Planted forest health: The need for a global strategy. *Science* **349:** 832-836.

Yan ZL, Sun JH, Don O, Zhang ZN. 2005. The red turpentine beetle, *Dendroctonus valens* LeConte (Scolytidae): an exotic invasive pest of pine in China. *Biodiversity and Conservation* **14:** 1735-1760.

Yeh FC, Yang RC, Boyle TBJ, Ye Z, Mao JX. 1997. *POPGENE, the user-friendly shareware for population genetic analysis.* Edmonton: Molecular Biology and Biotechnology Centre, University of Alberta.

Yin M, Duong TA, Wingfield MJ, Zhou XD, De Beer ZW. 2015. Taxonomy and phylogeny of the *Leptographium procerum* complex, including *L. sinense* sp. nov. and *L. longiconidiophorum* sp. nov. *Antonie van Leeuwenhoek*, **107:** 547-563.

Zamudio KR, Savage WK. 2003. Historical isolation, range expansion, and secondary contact of two highly divergent mitochondrial lineages in spotted salamanders (*Ambystoma maculatum*). *Evolution* **57:** 1631-1652.

Zarza E, Reynoso VH, Emerson BC. 2011. Discordant patterns of geographic variation between mitochondrial and microsatellite markers in the Mexican black iguana (*Ctenosaura pectinata*) in a contact zone. *Journal of Biogeography* **38:** 1394-1405.

Zhou XD, De Beer ZW, Wingfield MJ. 2013. Ophiostomatoid fungi associated with conifer-infesting bark beetles in China. In: Seifert KA, De Beer ZW and Wingfield MJ, eds. *The Ophiostomatoid Fungi: Expanding Frontiers*. Utrecht, The Netherlands: CBS-KNAW Biodiversity Centre.

Supplementary tables and figures

Fig. S1: Variation in ΔK and log probabilities over different K-values (numbers of genetic clusters) based on STRUCTURE analyses of the North American dataset of *D. valens*. ΔK values are represented by brown circles, and log probabilities are represented by blue bars.

Fig. S2: Plots for optimal model selection in the discriminate analysis of principal component analysis using the *Dendroctonus valens* dataset for North America. a) The BIC plot used to select the optimal number of clusters. b) The α-score plot used to select the optimal number of retained principal components.

Fig. S3: Bar plots of the average membership coefficients (STRUCTURE) and the membership probabilities (DAPC) of the *Dendroctonus valens* individuals collected in North America. K-values between two and five are shown. The genetic clusters are represented by purple bars, brown bars, blue bars, yellow bars, and grey bars.

Fig. S4: Bar plots of the average membership coefficients for each geographical population (see Results, Fig. S3) at K = 2.

Fig. S5: Principal component analysis plot of the model checking analysis for scenario 1.1. The two principal components with the majority of the variation are shown. The large yellow circle represents the actual dataset. The small hollow green circles represent the simulated datasets

based on the prior distributions. The large solid green circles represent the simulated datasets based on the posterior distributions.

Table S1: Summary statistics of the microsatellite markers in each sampling location. Bolded text indicates a sampling location-marker combination that violates Hardy-Weinberg equilibrium. $_{A}$The sampling location-marker combination shows evidence of null alleles based on homozygote excess. $_{B}$The sampling location-marker combination shows evidence of scoring error due to stuttering.

Table S2: Average membership coefficients (STRUCTURE) and membership probabilities (DAPC) for each sampling location in North America at the different K-values between two and five. s.d. = standard deviation.

Table S3: Average membership coefficients (STRUCTURE) and membership probabilities (DAPC) for each population at K = 4, followed in brackets by percentages of the individuals within the populations that were clearly assigned (i.e., membership coefficient or memberships probability > 0.8) or mostly assigned (membership coefficient or memberships probability > 0.8) to a cluster.

Table S4: Diversity statistics for the different sampling locations within this study. A = mean number of alleles, EA = effective number of alleles, H = Shannon's information index, ML = monomorphic loci, PA = private allele, Ho = observed heterozygosity, He = expected heterozygosity, s.d. = standard deviation.

Table S5: Allele frequencies within each *Dendroctonus valens* population. Rows highlighted in grey indicate alleles that are present in China.

Table S6: Allele frequencies within each sampling location. Rows highlighted in grey indicate alleles that are present in China.

Table S7: Pairwise F_{ST} values among sampling locations following Weir & Cockerham (1984). The upper right corner shows the 95% confidence intervals.

Table S8: The posterior probabilities with 95% confidence intervals of each scenario tested in steps 1 and 2.

Table S9: Ranges of parameter priors and posterior estimates of parameter values for scenarios 1.1 and 1.2 in step two. AN = the effective population size (Ne) of a hypothetical ancestral population for all of the populations in North America; CH = the Ne of the China population; WE = the Ne of the West population, CT = the Ne of the Central population; NE = the Ne of the Northeast population; MX = the Ne of the Mexico population; t1 = the time at which the four North American populations diverged from an ancestral population; t2 = the time at which the China population originated from the West population (in scenario 1.1) or the time at which the China population resulted from admixture between the West and Central populations (scenario 1.2); r1 = the percentage of admixture from the West population in scenario 1.2; μ1 = the

mutation rate of the microsatellites with tri-nucleotide repeats; μ2 = the mutation rate of the microsatellites with di-nucleotide repeats.

Table S10: Summary statistics used in model-checking of selected scenario. Observed values are given, along with probabilities that simulated < observed, calculated from 10000 simulated data sets. Statistics where $p < 0.05$ or > 0.95 are shown in bold.