

# Ancient and modern colonization of North America by hemlock woolly adelgid, *Adelges tsugae* (Hemiptera: Adelgidae), an invasive insect from East Asia

NATHAN P. HAVILL,\* SHIGEHICO SHIYAKE,† ASHLEY LAMB GALLOWAY,‡<sup>1</sup>  
 ROBERT G. FOOTTIT,§ GUOYUE YU,¶ ANNIE PARADIS,\*\* JOSEPH ELKINTON,\*\*  
 MICHAEL E. MONTGOMERY,\* MASAKAZU SANO†† and ADALGISA CACCONE‡‡

\*Northern Research Station, USDA Forest Service, Hamden, Connecticut 06514, USA, †Osaka Museum of Natural History, Osaka 546-0034, Japan, ‡Department of Entomology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA, §Canadian National Collection of Insects, Agriculture and Agri-Food Canada, Ottawa, Ontario K1A 0C6, Canada, ¶Institute of Plant & Environmental Protection, Beijing Academy of Agricultural & Forestry Science, Beijing 100097, China, \*\*Department of Environmental Conservation, University of Massachusetts, Amherst, Massachusetts 01003, USA, ††Systematic Entomology, Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Hokkaido, Japan, ‡‡Department of Ecology & Evolutionary Biology, Yale University, New Haven, Connecticut 06520, USA

## Abstract

Hemlock woolly adelgid, *Adelges tsugae*, is an invasive pest of hemlock trees (*Tsuga*) in eastern North America. We used 14 microsatellites and mitochondrial COI sequences to assess its worldwide genetic structure and reconstruct its colonization history. The resulting information about its life cycle, biogeography and host specialization could help predict invasion by insect herbivores. We identified eight endemic lineages of hemlock adelgids in central China, western China, Ulleung Island (South Korea), western North America, and two each in Taiwan and Japan, with the Japanese lineages specializing on different *Tsuga* species. Adelgid life cycles varied at local and continental scales with different sexual, obligately asexual and facultatively asexual lineages. Adelgids in western North America exhibited very high microsatellite heterozygosity, which suggests ancient asexuality. The earliest lineages diverged in Asia during Pleistocene glacial periods, as estimated using approximate Bayesian computation. Colonization of western North America was estimated to have occurred prior to the last glacial period by adelgids directly ancestral to those in southern Japan, perhaps carried by birds. The modern invasion from southern Japan to eastern North America caused an extreme genetic bottleneck with just two closely related clones detected throughout the introduced range. Both colonization events to North America involved host shifts to unrelated hemlock species. These results suggest that genetic diversity, host specialization and host phylogeny are not predictive of adelgid invasion. Monitoring non-native sentinel host trees and focusing on invasion pathways might be more effective methods of preventing invasion than making predictions using species traits or evolutionary history.

**Keywords:** complex life cycle, cyclical parthenogenesis, host range, invasive species

Received 4 November 2015; revision received 5 February 2016; accepted 9 February 2016

## Introduction

Invasions by non-native herbivorous insects continue to threaten forest ecosystems (Leibold *et al.* 1995; Aukema *et al.* 2010) as a consequence of accelerating arrival rates with international trade (Levine & D'Antonio 2003;

Correspondence: Nathan P. Havill, Fax: 203-230-4315;

E-mail: nphavill@fs.fed.us

<sup>1</sup>Present address: Department of Biology, Roane State Community College, Oak Ridge, Tennessee 37830, USA

Lockwood *et al.* 2005; Brockerhoff *et al.* 2014; Blackburn *et al.* 2015). Fortunately, most introductions of non-native insects do not result in successful establishment (Williamson & Fitter 1996; Aukema *et al.* 2010), while a few become serious invasive pests. This division between innocuous and invasive introductions has been difficult to predict, so ongoing research seeks to identify species traits that correlate with invasive potential. Some candidate characteristics include high genetic diversity, broad host range and the ability to reproduce asexually (e.g. Kolar & Lodge 2001; Jeschke & Strayer 2006; Liebhold & Tobin 2008; Dlugosch *et al.* 2015). However, there is little evidence that any one trait can reliably predict invasiveness, especially for herbivorous insects (Williamson 1996; Sakai *et al.* 2001; National Research Council 2002; Hayes & Barry 2008).

An alternative to the trait-based approach is to examine the evolutionary history of herbivores and their host plants, as a context in which to evaluate the potential for invasion. Many invasive herbivores have resistant host plant species in their native range and susceptible congeneric host species in their introduced range (e.g. Granett *et al.* 2001; Rebek *et al.* 2008; Desurmont *et al.* 2011; Nielsen *et al.* 2011). This pattern led to the concept of 'defence free space', where invasion is enabled by a lack of effective defences in host plants outside a herbivore's native geographic range (Gandhi & Herms 2010). But not every herbivore finds susceptible host plants after being introduced. For example, 86% of the 455 species of non-native herbivores that feed on trees in the United States have no record of causing damage (Aukema *et al.* 2010). Therefore, a lack of evolutionary association is not by itself predictive of invasion, but perhaps other patterns in the evolutionary history of herbivores and their host plant could be informative. For example, a species might be more likely to invade a region with host plants that are phylogenetically closely related to those in its native range.

Hemlock woolly adelgid, *Adelges tsugae* Annand, is a non-native pest that severely impacts the two hemlock (*Tsuga*) species native to eastern North America: *T. canadensis* and *T. caroliniana* (Orwig & Foster 1998; Siderhurst *et al.* 2010; Brantley *et al.* 2013). The native geographic range of *A. tsugae* includes the distributions of the other nine hemlock species found in East Asia and western North America (Havill *et al.* 2014). Mitochondrial and nuclear DNA sequence data from previous studies indicate that the lineage of *A. tsugae* introduced to eastern North America originated from a population in southern Japan that specializes on *T. sieboldii* (Havill *et al.* 2006, 2007), but fine-scale, worldwide population genetic studies have not been completed for this species.

Hemlock adelgids have a complex life cycle, with cyclical parthenogenesis and migration between

alternate hosts (Fig. 1; Havill & Footitt 2007). Populations can be either holocyclic, with host alternation between hemlock and spruce (*Picea*) where there is a sexual generation, or anholocyclic, with no host alternation and only parthenogenetic reproduction on hemlock. The adelgid holocycle takes 2 years to complete and they can overwinter on hemlock or spruce making host alternation facultative.

In this study, we use data from nuclear and mitochondrial DNA to examine the patterns of genetic variation in *A. tsugae* throughout its native and introduced ranges to elucidate patterns of life cycle variation, historical biogeography and host association. We discuss our findings in the context of testing the ability to predict invasion by insect herbivores.

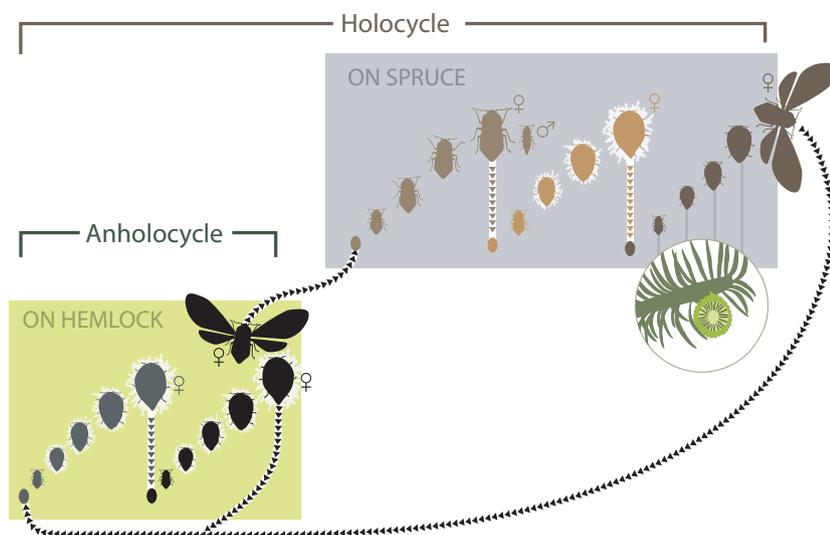
## Methods

### *Insect sampling and genotyping*

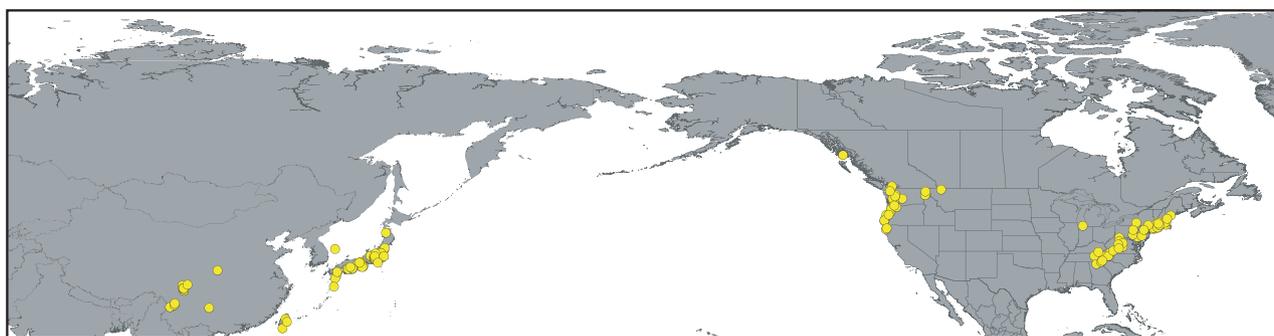
Samples of *A. tsugae* were collected between 2002 and 2015 from host trees growing in their native ranges (Fig. 2, Fig. S1–S4 & Table S1, Supporting information). Sample sites were pooled across years and among localities <10 km apart. For samples on hemlock, we genotyped a single insect per tree, because asexual reproduction on hemlock makes it likely that individuals from the same tree will share the same genotype (Fig. 1). In contrast, on spruce, where each gall is founded by an individual with sexual parents, we genotyped multiple individuals per tree, each from a different gall.

Adelgids were genotyped with 14 microsatellite loci following previously published protocols (Abdoulaye *et al.* 2009). Only samples with complete 14-locus genotypes were used for analyses. For most samples, a 651-base pair portion of the mitochondrial cytochrome oxidase subunit I (COI) gene was sequenced using standard protocols (deWaard *et al.* 2008). Additional details for genotyping and sequencing can be found in Appendix S1, Supporting information.

GENODIVE v 2.0b25 (Meirmans & Van Tiernderen 2004) was used to assign microsatellite genotypes to clonal multilocus lineages (MLLs; Arnaud-Haond *et al.* 2007), based on a threshold distance of eight steps between individuals, chosen by examining the frequency distribution of pairwise distances between genotypes (Fig. S5, Supporting information). Species with both sexual and asexual reproduction are expected to have a multimodal frequency distribution, with the first peak near zero representing clonal individuals and scoring errors, and subsequent peaks representing somatic mutations, sibling matings or population structure (Meirmans & Van Tiernderen 2004).



**Fig. 1** Hemlock adelgid life cycle. The holocycle takes 2 years to complete and includes migration between hemlock and spruce, on which the first generation is sexual and the last generation lives inside galls. Anholocyclic populations have two asexual generations per year, only on hemlock. [Diagram created by Vince D'Amico and Nathan P. Havill (Northern Research Station, USDA Forest Service)].



**Fig. 2** Hemlock adelgid sampling sites. Details can be found in Table S1 and Fig. S1-S4, Supporting information.

### Microsatellite and mitochondrial clusters

To identify the number of different genetic clusters in the microsatellite data, discriminant analysis of principal components (DAPC; Jombart *et al.* 2010) was performed using ADEGENET v 1.3.9 (Jombart 2008) in R v 3.1.1 (R Core Team 2014). The evolutionary relationship among all unique COI haplotypes was reconstructed with MRBAYES v 3.1.2 (Ronquist & Huelsenbeck 2003). Sequences from five other adelgid species were used to root the tree. COI haplotype clustering was also examined by constructing a network based on the statistical parsimony method of Templeton *et al.* (1992) using the software TCS v 1.21 (Clement *et al.* 2000) with a 95% confidence limit. Additional details of clustering analyses are provided in Appendix S1, Supporting information.

### Clonal diversity, molecular diversity, and reproductive mode

Clonal diversity was calculated for sites with 15 or more samples as the ratio of the number of MLLs to the

number of individuals, and estimated using Simpson's diversity index calculated with GENODIVE. To detect signatures of sexual reproduction, the standardized index of multilocus linkage disequilibrium ( $\bar{r}_d$ ; Agapow & Burt 2001) was calculated for sampling sites with 15 or more unique MLLs and tested for deviation from zero with 1000 random permutations, using POPPR v 1.1.2 (Kamvar *et al.* 2014). ARLEQUIN v 3.5.1.3 (Excoffier *et al.* 2005) was used to calculate the mean number of alleles per locus, expected and observed heterozygosities, and to test for differentiation ( $F_{ST}$ ) among sampling sites with 15 or more samples using the infinite allele model and 1000 permutations. SMOGD v 1.2.5 (Crawford 2010) was used to calculate the pairwise harmonic mean of differentiation ( $D_{est}$ ) among sampling sites (Jost 2008). This additional method of quantifying genetic differentiation is less sensitive than  $F_{ST}$  to underestimating differentiation when allelic diversity is high (Jost 2008). A one-tailed Wilcoxon sign-rank test, implemented in BOTTLENECK v 1.2.02 (Cornuet & Luikart, 1996), was used to test for an excess of observed heterozygotes across all loci.

### History of colonization from Asia to western North America

Four scenarios of colonization from Asia to western North America (Fig. 3) were tested using approximate Bayesian computation (ABC; Beaumont *et al.* 2002), implemented with DIYABC v 2.0 (Cornuet *et al.* 2014). The software assumes standard sexual reproduction (Cornuet *et al.* 2010), so a single individual per unique MLL that had both microsatellite and mitochondrial data was included to focus on the sexual generation in the adelgid life cycle. Confidence in model choice (fit to the observed data) and scenario choice (type I and type II error rates) were evaluated using the analyses provided by DIYABC. The divergence times in generations, output by DIYABC, were converted to years by dividing the number of generations by 2.25 based on the adelgid life cycle (Fig. 1). Additional details of the ABC analysis are provided in Appendix S1, Supporting information.

## Results

### Clonal diversity

Microsatellites were genotyped for 748 individuals, from 133 sampling sites across the *A. tsugae* native and introduced ranges (Fig. 2, S1–S4, Supporting information). Genotypes were assigned to 442 MLLs (Table 1). All but one sample in eastern North America were members of the same MLL, which was also found in Nakahata (Osaka Prefecture) and Kobe (Hyogo Prefecture), Japan. A second eastern North American MLL, differing from the dominant MLL by ten dinucleotide repeats in one allele at one locus, was found in one sample from Manchester in eastern Massachusetts. Western North America had higher clonal diversity than eastern North America, with 13 MLLs. The most common and widespread MLL in western North America was shared by 43 individuals from 19 of the 25 sampling sites (Table S1, Supporting information).

### Microsatellite and mitochondrial clusters

DAPC analysis of microsatellite genotypes separated the samples into  $k = 4$  major clusters (Fig. 4A). These correspond to broad geographic structure and host use. Subsequent DAPC analysis of nested population structure in Cluster 4, which includes samples from several regions, split them into  $k = 6$  subclusters (Fig. 4B), with MLLs from Taiwan and from western North America forming discrete clusters. Pairwise differentiation ( $F_{ST}$  and  $D_{est}$ ) between sampling sites showed the same hierarchical pattern with higher values among the major clusters that were identified with DAPC than among

sampling sites within these clusters (Table S3, Supporting information).

COI sequences were generated for 446 individuals (423 with microsatellite genotypes), resulting in 105 unique haplotypes (Table S1, Supporting information): 22 haplotypes in continental China, five in Taiwan, one in Ulleung Island (South Korea), 22 in western North America, 54 in Japan and one in both Japan and all eastern North American samples. The eastern North American haplotype was also found in samples collected in southern Japan on *T. sieboldii* in Kobe (Hyogo Prefecture) and Mt. Koya-san (Wakayama Prefecture) and on *P. torano* in Odai (Nara Prefecture) (Table S1, Supporting information).

The clustering pattern of COI haplotypes was consistent with those from DAPC analysis of microsatellite genotypes. In the COI phylogeny (Fig. 5; detailed in Fig. S8, Supporting information), the haplotype from central China (Guizhou) was in a basal position to all other samples in our study, followed by a clade with samples from western China. The Taiwan samples formed two clades that were basal to the Japanese, Ulleung Island, and North American samples. The haplotype network had five major clusters: the sample from central China (Guizhou), two clusters from Taiwan, a cluster from western China and a large cluster with the haplotypes from Japan, Ulleung Island, western North American and eastern North America (Fig. 6). This last cluster included a distinct subcluster of samples collected from *T. diversifolia*. The haplotypes from *T. sieboldii* and *P. torano* fell in two subclusters with a western North American haplotype between them.

### Clonal diversity and life cycle

Major lineages and sampling sites varied in the amount of clonal reproduction (Table 1). Samples from eastern North America and western China were the least and the most diverse, respectively. Adelgids sampled at sites in western China, Taiwan and southern Japan on *T. sieboldii* had genetic signatures of sexual reproduction ( $\bar{r}_d$ ) that were not statistically rejected, and/or collections with matching mitochondrial haplotypes from both hemlock and spruce, which also implies sexual reproduction (Table 2). In contrast, samples from Japan on *T. diversifolia* and from western North America did not show evidence of host alternation or sexual reproduction. Observed heterozygosity ( $H_o$ ) was much higher than expected heterozygosity ( $H_e$ ) in western North America: 0.942 vs. 0.591, respectively (Wilcoxon sign-rank test,  $P = 0.003$ ).

### Colonization from Asia to western North America

ABC analysis showed that the best-fit scenario of colonization from Asia to western North America involved

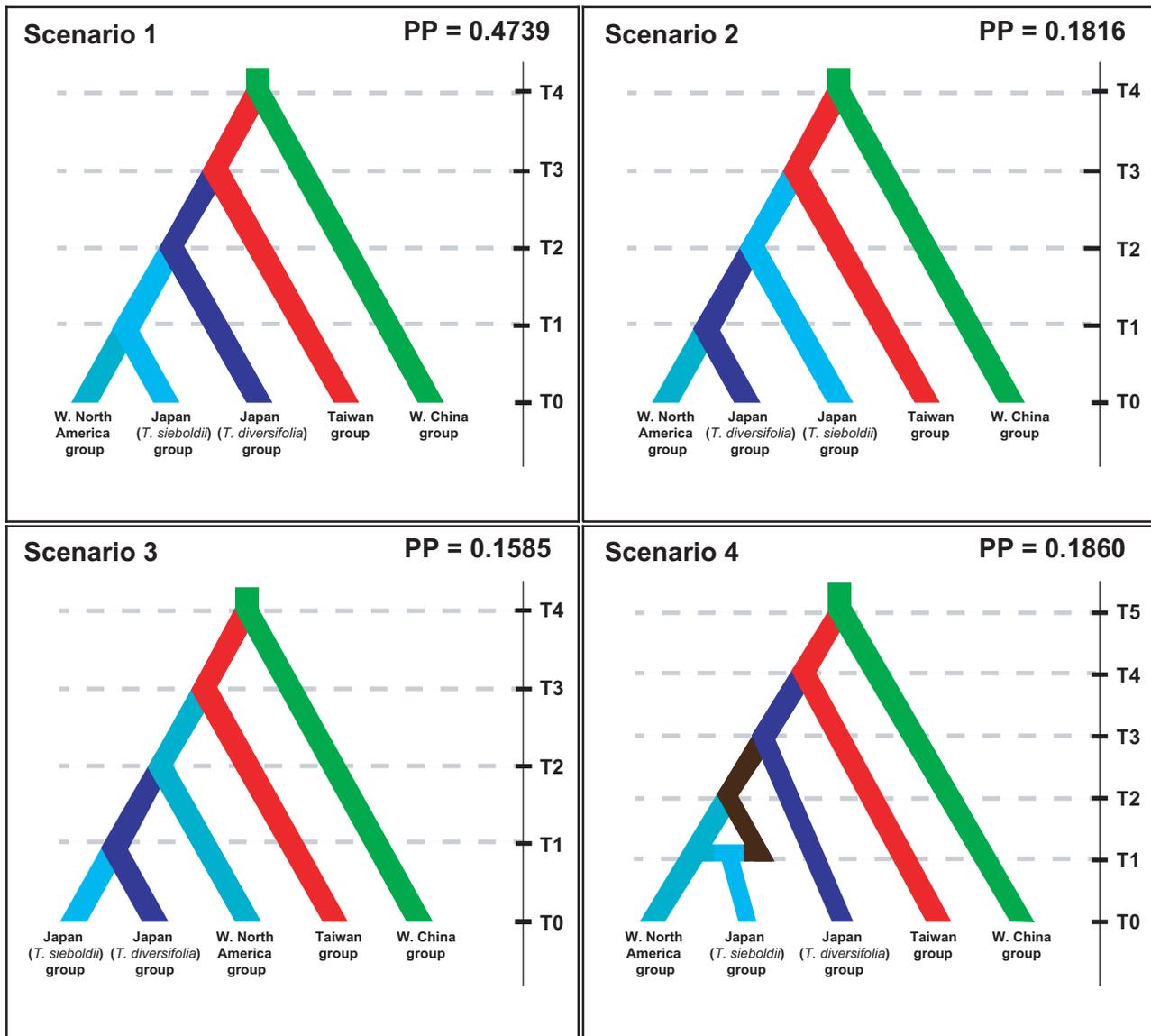


Fig. 3 Scenarios of hemlock adelgid colonization of from East Asia to western North America, tested with approximate Bayesian computation (ABC) using combined mitochondrial and microsatellite data. Posterior probabilities (PP) are shown for each scenario. Scenario 1 had the best fit.

a split from a direct ancestor of the adelgid lineage on *T. sieboldii* and *P. torano* in Japan (Fig. 3; Scenario 1). A principal component plot for evaluating model choice showed that the summary statistics generated using posterior parameters were a good approximation of the observed data (Fig. S9, Supporting information). The mutation rates for the most likely scenario were estimated to be  $6.23 \times 10^{-8}$  COI sequence mutations per generation and  $9.46 \times 10^{-6}$  microsatellite mutations per generation. These values are within the ranges estimated for other insects (Zhang & Hewitt 2003; Papadopoulou *et al.* 2010). Posterior distributions of divergence times in generations are shown in Fig. S10 and listed in

Table S2, Supporting information. Table 3 shows these estimates converted to years.

## Discussion

We describe a range of diversity within hemlock adelgids that was unexpected given that they are all considered a single species. Previous studies of hemlock adelgid phylogeny using DNA sequence data suggested five distinct lineages of hemlock adelgids: in western China, Taiwan, western North America and two in Japan that specialize on each Japanese hemlock species, *T. diversifolia* and *T. sieboldii* (Havill *et al.* 2006, 2007).

Lineage	Site	Host species	N*	MLL <sup>†</sup>	P <sub>d</sub> <sup>‡</sup>	S <sup>§</sup>
Western China		<i>T. chinensis</i> , <i>T. forrestii</i> , <i>T. dumosa</i> , <i>P. brachytyla</i> , <i>P. likiangensis</i>	66	65	0.985	1.00
	Huoditang	<i>T. chinensis</i>	18	18	1.000	1.000
	Yulong	<i>T. forrestii</i> , <i>P. likiangensis</i>	17	16	0.941	0.993
Taiwan		<i>T. formosana</i>	35	28	0.800	0.970
	Tayuling <sup>¶</sup>	<i>T. formosana</i>	24	19	0.774	0.961
Northern Japan ( <i>T. diversifolia</i> )		<i>T. diversifolia</i>	161	73	0.453	0.966
	Azuma	<i>T. diversifolia</i>	36	11	0.306	0.897
	Konsei	<i>T. diversifolia</i>	28	9	0.321	0.796
	Nyukasa	<i>T. diversifolia</i>	31	23	0.742	0.978
	Odai	<i>T. diversifolia</i>	24	4	0.167	0.239
	Tsurugi	<i>T. diversifolia</i>	17	2	0.118	0.485
Southern Japan ( <i>T. sieboldii</i> )		<i>T. sieboldii</i> , <i>P. torano</i>	329	260	0.790	0.991
	Ishizuchi	<i>T. sieboldii</i>	29	27	0.931	0.993
	Kirishima	<i>T. sieboldii</i>	31	31	1.000	1.000
	Kobe	<i>T. sieboldii</i>	31	6	0.194	0.353
	Mishotai	<i>T. sieboldii</i> , <i>P. torano</i>	28	28	1.000	1.000
	Mitsutoge	<i>T. sieboldii</i>	33	32	0.970	0.998
	Miyoshi	<i>T. sieboldii</i>	18	12	0.667	0.895
	Nakahata	<i>T. sieboldii</i>	15	13	0.867	0.981
	Nikko	<i>T. sieboldii</i> , <i>P. torano</i>	40	39	0.975	0.999
	Yakushima	<i>T. sieboldii</i>	30	9	0.300	0.759
	Yokote	<i>T. sieboldii</i>	24	23	0.958	0.996
Western North America		<i>T. heterophylla</i> , <i>T. mertensiana</i>	89	13	0.146	0.683
	Port Orford	<i>T. heterophylla</i>	23	4	0.174	0.325
	Seattle	<i>T. heterophylla</i>	15	4	0.267	0.442
Eastern North America		<i>T. canadensis</i> , <i>T. caroliniana</i>	61	2	0.033	0.033

\*Number of individuals.

<sup>†</sup>Number of unique multilocus lineages.

<sup>‡</sup>Clonal diversity (MLL/N).

<sup>§</sup>Simpson's diversity index.

<sup>¶</sup>Only MLLs from the larger lineage from Tayuling, as shown with DAPC analysis.

Our study confirms those lineages and, by employing more samples and more markers, suggests that there are additional lineages in Taiwan, central China and Ulleung Island, for a total of eight distinguishable lineages (Figs 4–6) that vary in life cycle, historical biogeography and host specialization.

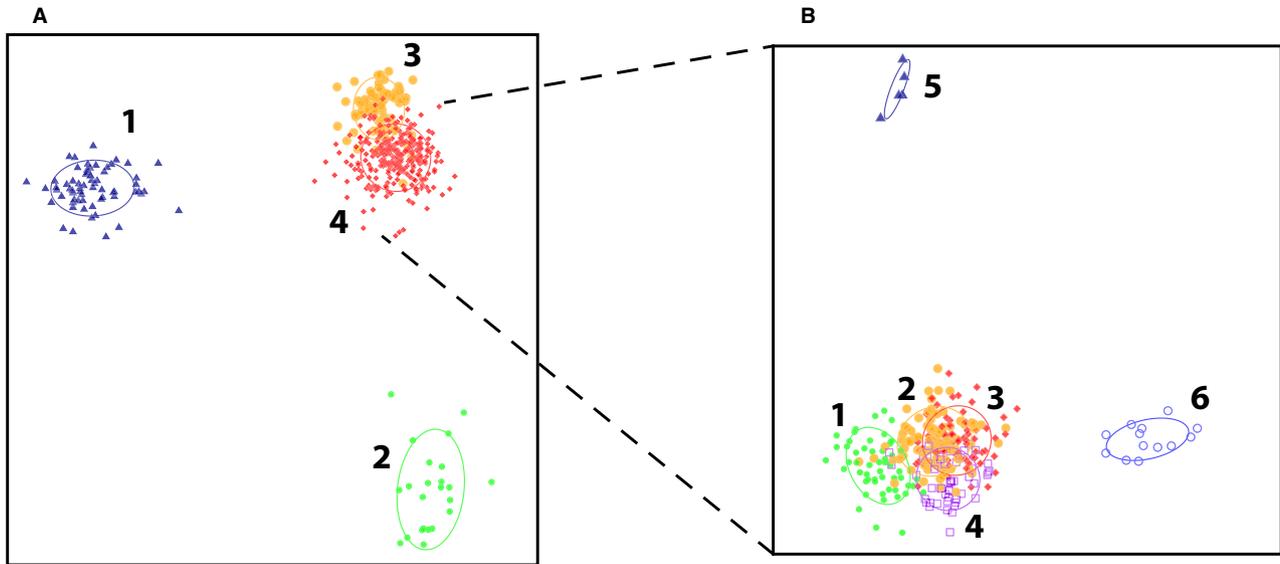
### Life cycle

At both the local and continental scales, *A. tsugae* includes a patchwork of sexual and asexual lineages. Sexual reproduction and host alternation to spruce (*Picea*) are strictly linked in adelgids, such that migration from hemlock to spruce is required for the sexual generation for all species with known life cycles (Havill & Footitt 2007). Our data confirm that sexual, host alternating hemlock adelgids use *P. torano* as a primary host in Japan (Inouye 1945; Sato 1999), and *P. likiangensis*

and *P. brachytyla* in China (Footitt *et al.* 2009) (Table 1). The data also suggest that hemlock adelgids migrate to spruce in Taiwan because we find a signature of sexual reproduction in one of the lineages sampled there. *Picea morrisonicola* is the likely alternate host because we observed dried galls on this species that were similar in morphology to those induced by *A. tsugae* in western China and Japan (S. Shiyake and M. Sano, personal observation). In addition, Chen *et al.* (2014) identified galls on *P. morrisonicola* as belonging to *A. tsugae*.

Within the lineages that are capable of host alternation, some populations are facultatively asexual, simply because they are not in the vicinity of spruce hosts. For example, adelgids from Yakushima Island in southern Japan, where *P. torano* does not grow, exhibit low clonal diversity and no signature of sexual reproduction. On the other hand, *P. torano* is common at the Mt. Mishotai site where there is high adelgid clonal diversity and a

**Table 1** Hemlock adelgid clonal diversity for major lineages and sampling sites with 15 or more individuals



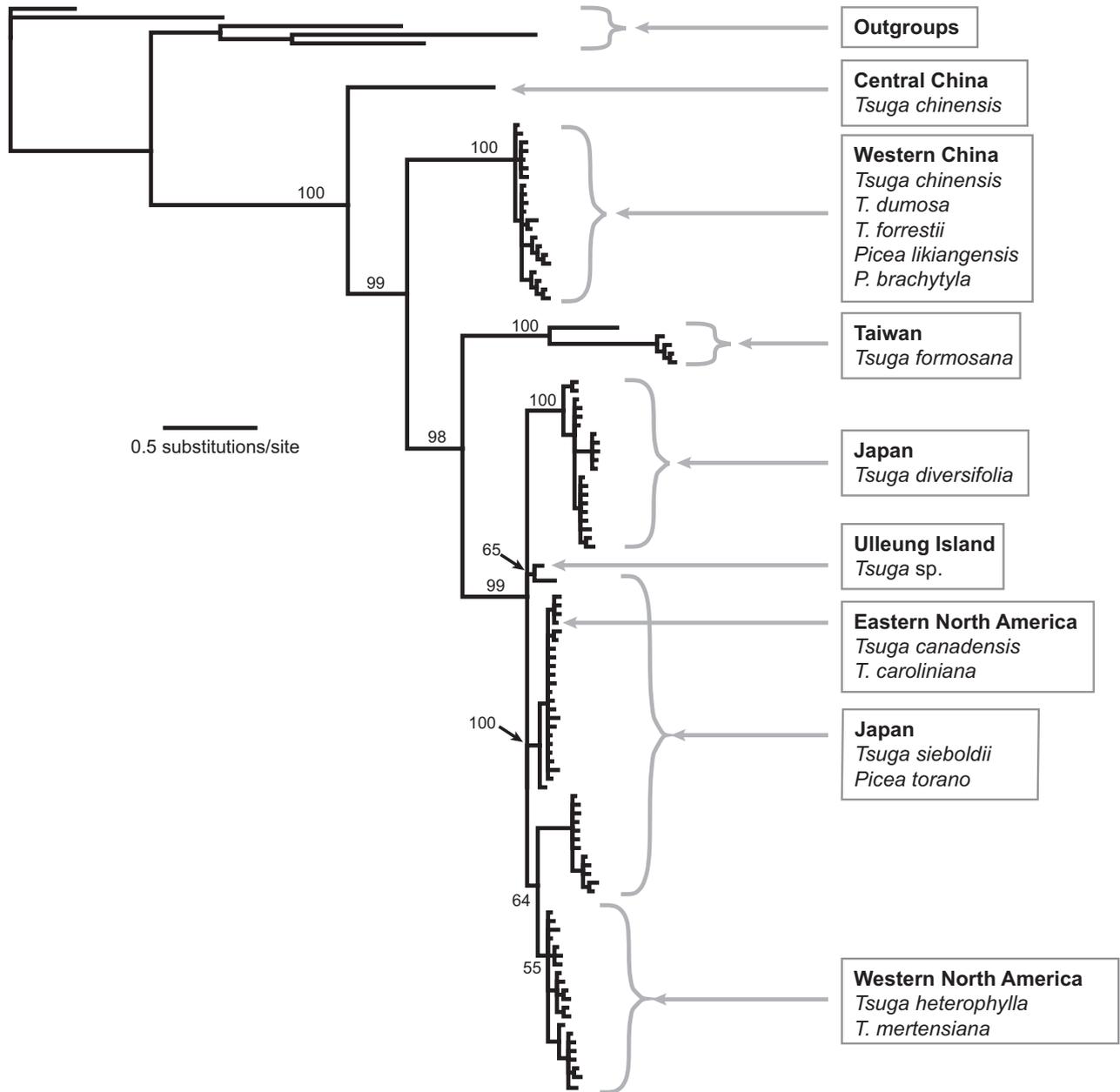
**Fig. 4** Discriminant analysis of principal components (DAPC) scatterplot of the first two principal components from hemlock adelgid microsatellite genotypes. (A) All unique multilocus lineages (MLLs). Cluster 1 included all MLLs from western China. Cluster 2 consisted of 23 of the 28 MLLs from Taiwan. Cluster 3 included all but one from *T. diversifolia* in northern Japan plus one collected on *T. sieboldii* in southern Japan. Cluster 4 included all but one from *T. sieboldii*, one from *T. diversifolia* and all from *P. torano* in Japan, the MLL from central China, the MLL from Ulleung Island, the remaining five from Taiwan, all from western North America and the two MLLs from eastern North America. (B) Subsequent nested analysis of Cluster 4 distinguished the five MLLs from Taiwan (Subcluster 5), and the thirteen MLLs from western North America (Subcluster 6). The four remaining subclusters corresponded weakly with geography, although a few sites had the majority of their MLLs in the same subcluster – for example 24 of 27 from Ishizuchi were in Subcluster 1, and 25 of 31 from Kirishima were in Subcluster 3. The MLL collected from *T. diversifolia* was in Subcluster 1, the central China and Ulleung Island MLLs were in Subcluster 2, and the two eastern North American MLLs were in Subcluster 4.

signature of sexual reproduction. Ulleung Island also lacks *P. torano*, but more evidence is needed to determine whether the hemlock adelgids there are capable of host alternation. On a broader geographic scale, the introduced adelgids in eastern North America originated from a sexual population in Japan (Figs 4–6), but they are asexual in their introduced range because they cannot survive on eastern North American spruce species (McClure 1989).

With just two closely related MLLs and one COI haplotype detected in eastern North America (Figs 4–6; Table S1, Supporting information), this is an example of an extreme genetic bottleneck resulting from invasion. The second MLL differed by mutations in just one microsatellite allele, and it was found in just a single sample collected in Massachusetts, far from the suspected site of original introduction in Virginia (Stoetzel 2002). These mutations therefore likely occurred after establishment. The lack of sexual reproduction and the extreme bottleneck resulting from introduction might be expected to limit the ability to adapt to the new environment (Agashe *et al.* 2011). This limitation could have been offset by two factors. First, the ability to reproduce asexually may have made hemlock adelgids more likely to establish because they would not be constrained by

Allee effects resulting from the difficulty of finding mates at low population densities (Liebhold & Tobin 2008). Tobin *et al.* (2013) provide support for this by showing that inoculation of eastern hemlock trees with a single *A. tsugae* individual can result in successful establishment. Second, clones introduced from a sexual lineage might have higher fitness in a novel environment than those from an obligately asexual lineage due to accumulation of deleterious mutations in the latter. This is consistent with the invasion of eastern North America having originated from a cyclically parthenogenetic population in southern Japan, and no evidence of invasion by asexual western North American adelgids, even though they can reproduce on eastern hemlocks planted in a western arboretum (Mausel 2005).

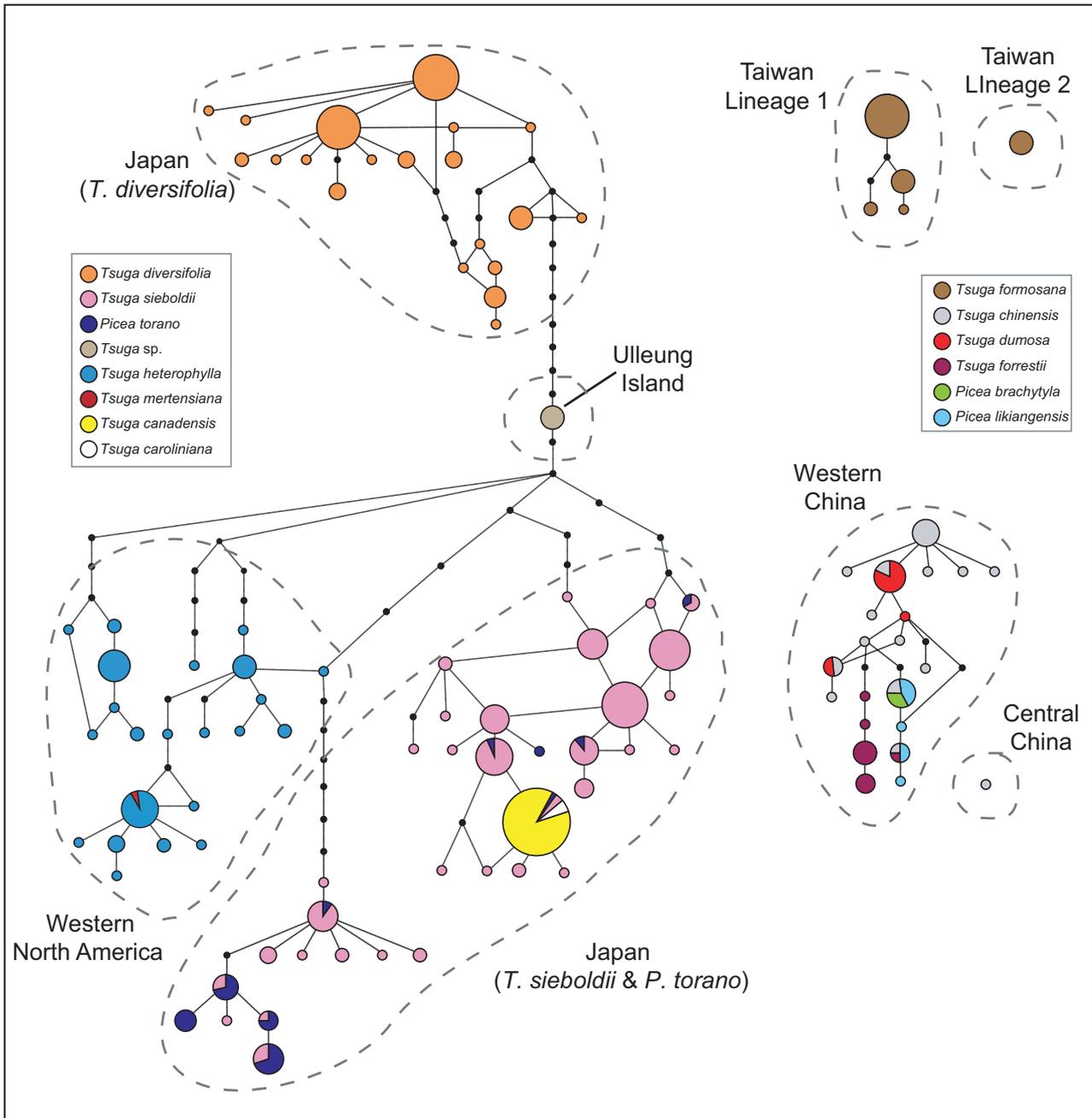
Other adelgid lineages might be obligately asexual because they no longer have the developmental ability to migrate to spruce. These include the lineage on *T. diversifolia* in Japan, the second lineage in Taiwan and the lineage in western North America. The adelgids on *T. diversifolia* do not have a genetic signature of sexual reproduction (Table 2), and they do not have a genetic match to samples collected on *P. torano* (Table 1) or on other Japanese spruce species (Footitt *et al.* 2009). We are also not aware of any winged migrants reported on



**Fig. 5** Bayesian 50% majority-rule consensus tree for all unique hemlock adelgid mitochondrial haplotypes. Posterior probabilities associated with major clades are indicated. The same tree with detailed sample information is shown in Fig. S8, Supporting information.

*T. diversifolia*. For the second lineage in Taiwan (Fig. 4B, Subcluster 5), we had too few samples to test for a genetic signature of sexual reproduction, but as they were collected in the same stand of trees as the sexual lineage (Table S1, Supporting information) with no evidence of admixture (Fig. 4), it could represent an isolated asexual lineage. Finally, in western North America, no winged migrants or generations on spruce are known (Annand 1924), and we find no genetic evidence of sexual reproduction, which suggest that this

lineage could have been asexual from the time it diverged from its Asian ancestor. We estimate that this occurred tens of thousands of years ago (Table 3). The high observed heterozygosity in this lineage (0.942) compared to others (Table 2) is also consistent with ancient asexuality, which could result in mutations accumulating independently on paired chromosomes in the absence of recombination (Birky 1996). Further analyses that explore whether other genetic signatures of ancient asexuality are present (Boschetti *et al.* 2012; Flot



**Fig. 6** Network showing relationships among hemlock adelgid mitochondrial haplotypes. The size of each circle is proportional to the frequency of the haplotype. Small black dots represent mutational steps separating observed haplotypes. Different colours correspond to the host plant from which the samples were collected, and dashed lines surround different lineages. The haplotype introduced to eastern North America on *T. canadensis* (yellow) and *T. caroliniana* (white) matches Japanese samples collected from *T. sieboldii* and *P. torano*.

*et al.* 2013) could be used to confirm this result, and to infer whether frequent extinction of asexual lineages due to accumulation of deleterious mutations (Lynch *et al.* 1993) explains the low species diversity in Adelgidae, despite being a very old group (Havill & Footitt 2007).

*Biogeography*

Lineage diversity within *A. tsugae* can be explained by dispersal and vicariance of adelgids with their host plants during Pleistocene glacial periods. The order of divergence for the major hemlock adelgid lineages can

Lineage	Site	MLL <sup>†</sup>	A <sup>‡</sup>	Ho <sup>§</sup>	He <sup>¶</sup>	$\bar{r}_d$ <sup>††</sup>
Western China	Huoditang	18	6.000	0.702	0.655	0.517**
	Yulong	16	6.929	0.661	0.730	2.127**
Taiwan	Tayuling	19 <sup>‡‡</sup>	6.154	0.506	0.562	-0.094
Northern Japan ( <i>T. diversifolia</i> )	Nyukasa	31	7.929	0.640	0.722	0.367*
Southern Japan ( <i>T. sieboldii</i> )	Ishizuchi	27	6.929	0.630	0.671	0.654**
	Kirishima	31	8.786	0.740	0.791	-0.088
	Mishotai	28	8.857	0.760	0.760	0.723**
	Mitsutoge	32	9.786	0.795	0.784	-0.019
	Nikko	39	9.000	0.727	0.775	0.232*
	Yokote	23	8.857	0.733	0.785	-0.153

<sup>†</sup>Number of unique multilocus lineages.

<sup>‡</sup>Mean alleles per locus.

<sup>§</sup>Observed heterozygosity.

<sup>¶</sup>Expected heterozygosity.

<sup>††</sup>Multilocus index of linkage disequilibrium.

<sup>‡‡</sup>only MLLs from the larger lineage from Tayuling, as shown with DAPC analysis.

\* $P < 0.05$ , \*\* $P < 0.001$ .

**Table 3** Divergence times of hemlock adelgid lineages from the most likely scenario of colonization to western North America (Fig. 3, Scenario 1). Time is converted from generations to years based on the adelgid life cycle

Time point (Fig. 3)	Divergence time (kya, with 90% confidence interval)
T4	323.1 (147.1–684.4)
T3	293.8 (155.5–421.3)
T2	60.4 (29.6–44.4)
T1	23.0 (10.2–47.6)

be inferred from the rooted COI phylogeny (Fig. 5). However, the colonization from Asia to western North America is not fully resolved by the COI data, so ABC analysis was completed to test different scenarios for this event, and to estimate lineage divergence times.

The first split in *A. tsugae* is indicated by the basal position of the sample from central China (Guizhou) in the COI phylogeny. This sample was collected on the eastern end of the Yungui Plateau, an area of rugged topology, which has been implicated as a barrier to gene flow in other insects (e.g. Ye *et al.* 2014; Zhang *et al.* 2015). Cun & Wang (2015) showed that Chinese hemlocks (*T. chinensis*) in southeast China are genetically distinct from those farther west, with a divergence dating perhaps to the Middle Pleistocene [c. 680 thousand years ago (kya)], when glaciation divided western and eastern groups of trees. Hemlock adelgids could have experienced the same event leading to comparable divergence. This date is consistent with our estimates of the remaining divergence times occurring c. 23–323 kya (Table 3). Glacial cycles during this period caused

**Table 2** Hemlock adelgid molecular diversity indices and likelihood of sexual reproduction for sampling sites with 15 or more unique MLLs. Sexual reproduction is inferred by multilocus linkage disequilibrium values ( $\bar{r}_d$ ) that are not significantly different than zero

repeated connection and separation among continental China, Taiwan and Japan (Vorisi 2000). These cycles are typically dated using marine oxygen isotope ratios and assigned to Marine Isotopic Stages (MIS), with odd numbers being warm interglacial periods and even numbers being glacial periods (Lisiecki & Raymo 2005). The divergence between adelgids in Taiwan from those in western China (c. 323 kya; Table 3) could be associated with the glacial cycle defined by MIS 9-10 (300–374 kya), and the divergence between adelgids in Japan from those in Taiwan (c. 294 kya; Table 3) with glacial cycle MIS 8-9 (243–337 kya).

Similarly, the split between the two Japanese hemlock adelgid lineages on *T. sieboldii* and *T. diversifolia* (c. 60 kya; Table 3) could be associated with the penultimate glacial period, MIS 4 (57–71 kya), during which Japan experienced extensive glacial advance (Sawagaki & Aoki 2011). Data to reconstruct the vegetation during this glacial period in Japan are scarce, but the situation could have been similar to the last glacial maximum (LGM; c. 18 kya), when the distributions of *T. sieboldii* and *P. torano* broadly overlapped, as they do now, while *T. diversifolia* was found only in isolated refugia in northern Honshu (Tsukada 1983). During this time, the adelgids feeding on *T. diversifolia* might have lost the ability to migrate to spruce, where the sexual generation occurs, thereby becoming isolated from the adelgids feeding on *T. sieboldii* and *P. torano*. The prolonged lack of gene flow between the lineages could have permitted them to each adapt to feeding on only one hemlock host species.

The ABC results suggested that western North America was colonized by a direct ancestor of the lineage that specializes on *T. sieboldii* in southern Japan (Fig. 3).

This direct divergence suggests that the reticulate pattern in the haplotype network (Fig. 6) is the result of incomplete lineage sorting (where haplotypes have not had enough time to coalesce within a lineage), and not a result of a complex scenario of divergence, admixture and subsequent divergence, as was tested with Scenario 4 (Fig. 3). The divergence time for this colonization event is estimated at *c.* 23 kya (Table 3) in association with the most recent glacial period, MIS 2 (14–29 kya). Vegetation reconstructions for MIS2 suggest that *Tsuga* was not continuous across the Bering land bridge when it connected Asia to western North America. The region was probably dominated by mesic shrub tundra, while hemlock had retreated south to the California coast (Roberts & Hamann 2015). There is also no signal for continuity across Beringia in the phylogenies of the host plants, as the divergences between Asian and western North American species are estimated to have occurred well before this time period [*c.* 42 million years ago (mya) for hemlock (Havill *et al.* 2008); *c.* 20–25 mya for spruce (Lockwood *et al.* 2013)]. Several studies suggest that there could have been small stands of spruce on the Bering land bridge during or near the LGM (Brubaker *et al.* 2005; Anderson *et al.* 2006; Zazula *et al.* 2006), which could have facilitated the adelgid colonization, but there is no similar evidence for hemlocks (Roberts & Hamann 2015). Adelgid colonization would therefore have had to involve crawlers or winged migrants travelling between disconnected Asian and North American forests.

The interglacial period MIS 3 (57–29 kya) is also within the 90% confidence limit of the date for this event (Table 3), and it might be a more suitable time for adelgid colonization than the LGM. This period is characterized by warm phases that lasted about 1–2 kya and oscillations in sea level that kept the Bering land bridge intermittently intact (Rabassa & Ponce 2013). During these warm phases, boreal forests in Asia migrated towards the Bering land bridge (Anderson & Lozhkin 2001; Bigham-Grette *et al.* 2003), and the presence of hemlock pollen in Hokkaido and Sakhalin during MIS 3 suggests that hemlock adelgids might have survived in northeast Asia (Igarashi & Zharov 2011; Leipe *et al.* 2015). Hemlock forest in western North America may also have been closer to the Bering land bridge than during the LGM with pollen records in interior British Columbia and coastal Washington (Clague *et al.* 2003; Jiménez-Moreno *et al.* 2010).

As adelgids would have had to travel long distances across the Bering land bridge to colonize western North America, this might have been mediated by migrating birds passively transporting adelgid eggs or crawlers, rather than active dispersal of winged adelgid adults which are not strong fliers. The furthest an adelgid

species (*Pineus pinifoliae*) is known to travel and colonize new trees is about 30 miles (48 km), when convective winds carried them to high altitudes, and most migrations are much shorter than that (Lowe 1966). On the other hand, eggs and crawlers of *A. tsugae* are known to be dispersed by birds and can survive up to 14 days after being dislodged from a tree (McClure 1990). Moreover, the existence of recently diverged species of forest birds in western North America and Siberia (e.g. Zink *et al.* 1995; Alström *et al.* 2011) supports a scenario of trans-Beringian colonization with migrating birds around the last glacial period.

An alternative date of 6 mya for adelgid colonization from Asia to western North America was estimated using DNA sequence data (Havill *et al.* 2007). The climatic conditions at this time were also suitable for conifer forests around Beringia (Wolfe & Leopold 1967; Matthews 1980; Tiffney 1985), and hemlock pollen has been recorded in both northeastern Asia and western North America during this time (Shilo & Minyuk 2006; Andreev *et al.* 2014; LePage 2003). This older divergence time estimate depended on calibration with only a few fossils of insects in amber hypothesized to be adelgid relatives (Havill *et al.* 2007), making this result questionable. A more recent colonization date, similar to the one that we estimate in this study, seems more likely, given the lack of coalescence of COI sequences for adelgids in western North America and Japan (Fig. 6).

It should be noted, however, that there is also some uncertainty in our ABC results. The type I error rate (0.418; Table S2, Supporting information) was fairly high, and while Scenario 1 had a posterior probability (0.4739) that was 2.5 – 3× higher than the others (Fig. 3), marginal support for the other scenarios indicates that they cannot be completely ruled out. This uncertainty does not seem to be a consequence of ABC model choice because the PCA plot to assess the model shows the observed data set surrounded by a tight cluster of points generated from the posterior distribution of parameters (Fig. S9, Supporting information). Rather, the statistical error likely results from a combination of the low sample size of unique MLLs in the western North American lineage, and from two of the five lineages being obligately asexual, which violates the assumption of standard sexual reproduction in the analysis (Cornuet *et al.* 2010). Future ABC analyses that explicitly account for a cyclically parthenogenetic life cycle and molecular dating methods that use additional fossils for calibration might help to refine our conclusions.

#### Host specialization

The degree of host specialization varies among different hemlock adelgid lineages. The two Japanese lineages

are each confined to just one hemlock species in their native range (Figs. 4–6), while the lineage in western China is a relative generalist, feeding on three hemlock and two spruce host species (Table 1). Moreover, phylogenetic relatedness of hemlock species does not make them more likely to be suitable hosts to the same adelgid lineage. We have identified at least four host shifts between unrelated hemlock species: the modern colonization from *T. sieboldii* to the eastern North American species (*T. canadensis* and *T. caroliniana*), and the ancient colonization from an ancestor of *T. sieboldii* to the western North American species [*T. heterophylla* and *T. mertensiana* (Note, however, that we are not aware of confirmed adelgid specimens from *T. mertensiana* growing in a natural setting, making the host switch to this species uncertain)]. None of the four North American hemlock species are closely related to *T. sieboldii* (Havill *et al.* 2008). On the other hand, the adelgid lineage that specializes on *T. sieboldii* in Japan cannot survive on *T. chinensis* (Del Tredici & Kitajima 2004; Montgomery *et al.* 2009), even though *T. sieboldii* and *T. chinensis* are closely related (Havill *et al.* 2008; Holman 2014). A similar pattern has also been shown in other groups of insects where cospeciation is rare (de Vienne *et al.* 2013), and plant defences play a larger role in determining host suitability than phylogeny, at least when considering more recently diverged host species (e.g. Janike 1990; Desurmont *et al.* 2011).

## Conclusions

Hemlock adelgids have traits that would seem to make them unlikely invaders. Host specialization in Japan did not prevent the modern invasion to new hemlock host species in eastern North America. In addition, this invasion, as well as the ancient colonization of western North America, involved host shifts to hemlock species unrelated to their natal hosts. Therefore, phylogenetic distance between native and non-native hemlock species cannot be used to predict invasion for hemlock adelgids. We might also have expected adaptation to the new environment to have been limited by an extreme genetic bottleneck in the introduced range. This was also not the case, although low genetic diversity could have been offset by asexual reproduction alleviating the difficulty in finding mates in the introduced range, and by increased fitness resulting from sexual reproduction in the source population.

Our data therefore show that the invasive potential of different adelgid lineages may be determined by complex interactions among life history traits, evolutionary history and the introduced environment, making prediction of invasion very difficult. A more reliable way to determine their potential to invade might be to

experimentally expose different lineages to novel hosts, or to monitor non-native sentinel hosts growing in their native ranges (Britton *et al.* 2010; Roques *et al.* 2015).

As a lack of predictive patterns continues to be found for groups of insect herbivores, it might be more effective to focus on invasion pathways rather than on traits or evolutionary history to prevent invasions. Propagule pressure, or the rate of arrival of non-native species and individuals, is currently the most supported predictor of insect invasions (Lockwood *et al.* 2005; Brockerhoff *et al.* 2014; Blackburn *et al.* 2015). By analysing interception rates at points of entry as an estimate of propagule pressure, it is possible to determine high-risk pathways for introduction (Work *et al.* 2005; Liebhold *et al.* 2012). Targeted regulations to limit introduction of non-native insects via known pathways can then be effective, but require frequent re-evaluation based on updated interception data (Brockerhoff *et al.* 2014; Haack *et al.* 2014).

For adelgids, switching to unrelated host species within a genus might be common, while switching to host species in other genera is rare (Havill *et al.* 2007). For each group of herbivores that exhibit some level of host specificity, there could be a threshold divergence time beyond which host phylogeny begins to play a more dominant role in determining host suitability than other factors. Determining whether this could be predictive of invasion for insect herbivores would require more comprehensive analysis than is currently available, but considering the pace at which phylogenies for plants and insects are accumulating (Hinchliff *et al.* 2015), it may soon be possible.

## Acknowledgements

This project was supported by USDA NIFA Award 2007-35302-18097, the US Forest Service Northern and Southern Research Stations and the Yale Institute for Biospheric Studies. Andrea Gloria-Soria, Talbot Trotter and Sandy Liebhold provided comments on an early draft. We are grateful to Eric Maw who helped with laboratory work, to Darcy Nelson, with US Forest Service International Programs, who facilitated collecting logistics, and to our many colleagues who collected samples for this project.

## References

- Abdoulaye D, Acevedo I, Adebayo AA *et al.* (2009) Permanent genetic resources added to molecular ecology resources database 1 August 2009 – 30 September 2009. *Molecular Ecology Resources*, **10**, 232–236.
- Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, **1**, 101–102.
- Agashe D, Falk JJ, Bolnick DI (2011) Effects of founding genetic variation on adaptation to a novel resource. *Evolution*, **65**, 2481–2491.

- Alström P, Saitoh T, Williams D *et al.* (2011) The Arctic Warbler *Phylloscopus borealis* – three anciently separated cryptic species revealed. *Ibis*, **153**, 395–410.
- Anderson PH, Lozhkin AV (2001) The Stage 3 interstadial complex (Karginiskii/middle Wisconsinan interval) of Beringia: variations in paleoenvironments and implications for paleoclimatic interpretations. *Quaternary Science Reviews*, **20**, 93–125.
- Anderson LL, Hu FS, Nelson DM *et al.* (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12447–12450.
- Andreev AA, Tarasov PE, Wennrich V *et al.* (2014) Late Pliocene and Early Pleistocene vegetation history of northeastern Russian Arctic inferred from the Lake El'gygytgyn pollen record. *Climate of the Past*, **10**, 1017–1039.
- Annand PN (1924) A new species of *Adelges* (Hemiptera, Phylloxeridae). *Pan-Pacific Entomologist*, **1**, 79–82.
- Arnaud-Haond S, Duarte CM, Alberto F *et al.* (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology*, **16**, 5115–5139.
- Aukema JE, McCullough DG, Von Holle B *et al.* (2010) Historical accumulation of nonindigenous forest pests in the continental United States. *BioScience*, **60**, 886–897.
- Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
- Bigham-Grette J, Lozhkin AV, Anderson PM *et al.* (2003) Paleoenvironmental conditions in western Beringia before and during the last glacial maximum. In: *Entering America: Northeast Asia and Beringia Before the Last Glacial Maximum* (ed. Masden D), pp. 29–62. University of Utah Press, Salt Lake City, Utah.
- Birky CW (1996) Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics*, **144**, 427–437.
- Blackburn TM, Lockwood JL, Cassey P (2015) The influence of numbers on invasion success. *Molecular Ecology*, **24**, 1942–1953.
- Boschetti C, Carr A, Crisp A *et al.* (2012) Biochemical diversification through foreign gene expression in bdelloid rotifers. *PLoS Genetics*, **8**, e1003035.
- Brantley S, Ford CR, Vose JM (2013) Future species composition will affect forest water use after loss of eastern hemlock from southern Appalachian forests. *Ecological Applications*, **23**, 777–790.
- Britton KO, White P, Kramer A *et al.* (2010) A new approach to stopping the spread of invasive insects and pathogens: early detection and rapid response via a global network of sentinel plantings. *New Zealand Journal of Forestry Science*, **40**, 109–114.
- Brockerhoff EG, Kimberley M, Liebhold AM *et al.* (2014) Predicting how altering propagule pressure changes establishment rates of biological invaders across species pools. *Ecology*, **95**, 594–601.
- Brubaker LB, Anderson PM, Edwards ME, Lozhkin AV (2005) Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *Journal of Biogeography*, **32**, 833–848.
- Chen C-Y, Chiu M-C, Yeh W-B, Kuo M-H (2014) The effect of accumulated temperature on the development of gall-forming adelgids on Taiwan spruce. *Formosan Entomologist*, **34**, 59–71.
- Clague JJ, Mattherw RW, Ager TA (2003) Environments of northwestern North America before the Last Glacial Maximum. In: *Entering America: Northeast Asia and Beringia Before the Last Glacial Maximum* (ed. Masden D), pp. 63–94. University of Utah Press, Salt Lake City, Utah.
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.
- Cornuet JM, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.
- Cornuet JM, Pudlo P, Veyssier J *et al.* (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.
- Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources*, **10**, 556–557.
- Cun Y-Z, Wang X-Q (2015) Phylogeography and evolution of three closely related species of *Tsuga* (hemlock) from subtropical eastern Asia: further insights into speciation of conifers. *Journal of Biogeography*, **42**, 315–327.
- Del Tredici P, Kitajima A (2004) Introduction and cultivation of Chinese hemlock (*Tsuga chinensis*) and its resistance to hemlock woolly adelgid (*Adeles tsugae*). *Journal of Arboriculture*, **30**, 282–287.
- Desurmont GA, Donoghue MJ, Clement WL *et al.* (2011) Evolutionary history predicts plant defense against an invasive pest. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 7070–7074.
- Dlugosch KM, Anderson SR, Braasch J *et al.* (2015) The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular Ecology*, **24**, 2095–2111.
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Flot JF, Hespels B, Li X *et al.* (2013) Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature*, **500**, 453–457.
- Footitt RG, Maw HEL, Havill NP *et al.* (2009) DNA barcodes to identify species and explore diversity in the Adelgidae (Insecta: Hemiptera: Aphidoidea). *Molecular Ecology Resources*, **9**(Suppl. 1), 188–195.
- Gandhi KJK, Herms DA (2010) Direct and indirect effects of alien insect herbivores on ecological processes and interactions in forests of eastern North America. *Biological Invasions*, **12**, 389–405.
- Granett J, Walker MA, Kocsis L *et al.* (2001) Biology and management of grape phylloxera. *Annual Review of Entomology*, **46**, 387–412.
- Haack RA, Britton KO, Brockerhoff EG *et al.* (2014) Effectiveness of the international phytosanitary standard ISPM No. 15 on reducing wood borer infestation rates in wood packaging material entering the United States. *PLoS ONE*, **9**, e96611.
- Havill NP, Footitt RG (2007) Biology and evolution of Adelgidae. *Annual Review of Entomology*, **52**, 325–349.

- Havill NP, Montgomery ME, Yu G *et al.* (2006) Mitochondrial DNA from hemlock woolly adelgid (Hemiptera: Adelgidae) suggests cryptic speciation and pinpoints the source of the introduction to eastern North America. *Annals of the Entomological Society of America*, **99**, 195–203.
- Havill NP, Footitt RG, von Dohlen CD (2007) Evolution of host specialization in the Adelgidae (Insecta: Hemiptera) inferred from molecular phylogenetics. *Molecular Phylogenetics and Evolution*, **44**, 357–370.
- Havill NP, Campbell CS, Vining TF *et al.* (2008) Phylogeny and biogeography of *Tsuga* (Pinaceae) inferred from nuclear ribosomal ITS and chloroplast DNA sequence data. *Systematic Botany*, **33**, 478–489.
- Havill NP, Vieira LC, Salom SM (2014) *Biology and Control of Hemlock Woolly Adelgid*. USDA Forest Service, Forest Health Technology Enterprise Team, Morgantown, West Virginia.
- Hayes KR, Barry SC (2008) Are there any consistent predictors of invasion success? *Biological Invasions*, **10**, 483–506.
- Hinchliff CE, Smith SA, Allman JF *et al.* (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 12764–12769.
- Holman G (2014) *Phylogenetic systematics in Pinaceae (Pine Family)*. PhD Dissertation, University of Maine, Orono, Maine.
- Igarashi Y, Zharov AE (2011) Climate and vegetation change during the late Pleistocene and early Holocene in Sakhalin and Hokkaido, northeast Asia. *Quaternary International*, **237**, 24–31.
- Inouye M (1945) Monographische Studie über die japanischen Koniferen-Gallenläuse (Adelgidae). *Bulletin of the Sapporo Branch of the Government Forest Experiment Station*, **15**, 1–105.
- Jaenike J (1990) Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics*, **21**, 243–273.
- Jeschke JM, Strayer DL (2006) Determinants of vertebrate invasion success in Europe and North America. *Global Change Biology*, **12**, 1608–1619.
- Jiménez-Moreno G, Anderson RS, Desprat S *et al.* (2010) Millennial-scale variability during the last glacial in vegetation records from North America. *Quaternary Science Reviews*, **29**, 2865–2881.
- Jombart T (2008) *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.
- Jost L (2008)  $G_{ST}$  and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.
- Kamvar ZN, Tabima JF, Grünwald NJ (2014) *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281.
- Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution*, **16**, 199–204.
- Leipe C, Nakagawa T, Gotanda K *et al.* (2015) Late Quaternary vegetation and climate dynamics at the northern limit of the East Asian summer monsoon and its regional and global-scale controls. *Quaternary Science Reviews*, **116**, 57–71.
- LePage BA (2003) A new species of *Tsuga* (Pinaceae) from the middle Eocene of Axel Heiberg Island, Canada, and an assessment of the evolution and biogeographical history of the genus. *Botanical Journal of the Linnean Society*, **141**, 257–296.
- Levine JM, D'Antonio CM (2003) Forecasting biological invasions with increasing international trade. *Conservation Biology*, **17**, 322–326.
- Liebholt AM, Tobin PC (2008) Population ecology of insect invasions and their management. *Annual Review of Entomology*, **53**, 387–408.
- Liebholt AM, Macdonald WL, Bergdahl D, Maestro VC (1995) Invasion by exotic forest pests - a threat to forest ecosystems. *Forest Science*, **41**, 1–49.
- Liebholt AM, Brockerhoff EG, Garrett LJ *et al.* (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.
- Lisiecki LE, Raymo ME (2005) A Pliocene-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}O$  records. *Paleoceanography*, **20**, PA1003.
- Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution*, **20**, 223–228.
- Lockwood JD, Aleksic JM, Zou J *et al.* (2013) A new phylogeny for the genus *Picea* from plastid, mitochondrial, and nuclear sequences. *Molecular Phylogenetics and Evolution*, **69**, 717–727.
- Lowe JH (1966) *Biology and dispersal of *Pinus pinifoliae* (Fitch)*. PhD Dissertation, Yale University, New Haven, Connecticut.
- Lynch M, Burger R, Butcher D *et al.* (1993) The mutational meltdown in asexual populations. *Journal of Heredity*, **84**, 339–344.
- Matthews JV (1980) Tertiary land bridges and their climate: backdrop for development of the present Canadian insect fauna. *Canadian Entomologist*, **112**, 1089–1103.
- Mausel DL (2005) Observations on fecundity and natural enemies of hemlock woolly adelgid, *Adelges tsugae* Annand (Hemiptera: Adelgidae), in Seattle, Washington. *Pan-Pacific Entomologist*, **81**, 97–98.
- McClure MS (1989) Evidence of a polymorphic life cycle in the hemlock woolly adelgid, *Adelges tsugae* (Homoptera: Adelgidae). *Annals of the Entomological Society of America*, **82**, 50–54.
- McClure MS (1990) Role of wind, birds, deer, and humans in the dispersal of hemlock woolly adelgid (Homoptera: Adelgidae). *Environmental Entomology*, **19**, 36–43.
- Meirmans PG, Van Tierderden PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.
- Montgomery ME, Bentz SE, Olsen RT (2009) Evaluation of hemlock (*Tsuga*) species and hybrids for resistance to *Adelges tsugae* (Hemiptera: Adelgidae) using artificial infestation. *Journal of Economic Entomology*, **102**, 1247–1254.
- National Research Council (2002) *Predicting Invasions of Nonindigenous of Plants and Plant Pests*. National Academies Press, Washington, District of Columbia.
- Nielsen DG, Muilenburg VL, Herms DA (2011) Interspecific variation in resistance of Asian, European, and North American birches (*Betula* spp.) to bronze birch borer (Coleoptera: Buprestidae). *Environmental Entomology*, **40**, 648–653.
- Orwig DA, Foster DR (1998) Forest response to the introduced hemlock woolly adelgid in southern New England, USA. *Journal of the Torrey Botanical Society*, **125**, 60–73.
- Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution*, **27**, 1659–1672.

- R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Rabassa J, Ponce JF (2013) The Heinrich and Dansgaard-Oeschger climatic events during Marine Isotopic Stage 3: searching for appropriate times for human colonization of the Americas. *Quaternary International*, **299**, 94–105.
- Rebek EJ, Herms DA, Smitley DR (2008) Interspecific variation in resistance to emerald ash borer (Coleoptera: Buprestidae) among North American and Asian ash (*Fraxinus* spp.). *Environmental Entomology*, **37**, 242–246.
- Roberts DR, Hamann A (2015) Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proceedings of the Royal Society B-Biological Sciences*, **282**, 20142903.
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- Roques A, Fan JT, Courtial B *et al.* (2015) Planting sentinel European trees in eastern Asia as a novel method to identify potential insect pest invaders. *PLoS ONE*, **10**, e0120864.
- Sakai AK, Allendorf FW, Holt JS *et al.* (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics*, **32**, 305–332.
- Sato S (1999) Galls on tiger-tail spruce made by *Aphrastasia tsugae* at Mt. Tsurugi. *Gensei*, **73**, 1–26.
- Sawagaki T, Aoki T (2011) Late quaternary glaciations in Japan. *Developments in Quaternary Sciences*, **15**, 1013–1021.
- Shilo NA, Minyuk PS (2006) Magnetostratigraphy of the climatic optimum at the early–middle Miocene transition in Northeast Russia. *Doklady Earth Sciences*, **410**, 1014–1016.
- Siderhurst LA, Griscom HP, Hudy M *et al.* (2010) Changes in light levels and stream temperatures with loss of eastern hemlock (*Tsuga canadensis*) at a southern Appalachian stream: implications for brook trout. *Forest Ecology and Management*, **260**, 1677–1688.
- Stoetzel MB (2002) History of the introduction of *Adelges tsugae* based on voucher specimens in the Smithsonian Institution National Collection of Insects. In: *Proceedings: Hemlock Woolly Adelgid in the Eastern United States Symposium* (eds Onken B, Reardon R, Lashomb J), p. 12. Rutgers University, East Brunswick, New Jersey.
- Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III: cladogram estimation. *Genetics*, **132**, 619–633.
- Tiffney BH (1985) Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. *Journal of the Arnold Arboretum*, **66**, 73–94.
- Tobin P, Turcotte R, Snider D (2013) When one is not necessarily a lonely number: initial colonization dynamics of *Adelges tsugae* on eastern hemlock, *Tsuga canadensis*. *Biological Invasions*, **15**, 1925–1932.
- Tsukada M (1983) Vegetation and climate during the last glacial maximum in Japan. *Quaternary Research*, **19**, 212–235.
- de Vienne DM, Refregier G, Lopez-Villavicencio M, *et al.* (2013) Cospeciation vs. host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist*, **198**, 347–385.
- Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, **27**, 1153–1167.
- deWaard JR, Ivanova NV, Hajibabaei M, Hebert PDN (2008) Assembling DNA barcodes: analytical protocols. In: *Environmental Genomics, Methods in Molecular Biology*, vol. 410 (ed. Martin CC), pp. 275–293. Humana Press, Totowa, New Jersey.
- Williamson M (1996) *Biological Invasions*. Chapman and Hall, New York.
- Williamson M, Fitter A (1996) The varying success of invaders. *Ecology*, **77**, 1661–1666.
- Wolfe JA, Leopold EB (1967) Neogene and early Quaternary vegetation of northwestern North America and northeastern Asia. In: *The Bering Land Bridge* (ed. Hopkins D. M.), pp. 193–206. Stanford University Press, Stanford, California.
- Work TT, McCullough DG, Cavey JF *et al.* (2005) Arrival rate of nonindigenous insect species into the United States through foreign trade. *Biological Invasions*, **7**, 323–332.
- Ye Z, Zhu GP, Chen PP *et al.* (2014) Molecular data and ecological niche modelling reveal the Pleistocene history of a semi-aquatic bug (*Microvelia douglasi douglasi*) in East Asia. *Molecular Ecology*, **23**, 3080–3096.
- Zazula GD, Telka AM, Harington CR *et al.* (2006) New spruce (*Picea* spp.) macrofossils from Yukon Territory: implications for Late Pleistocene refugia in Eastern Beringia. *Arctic*, **59**, 391–400.
- Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.
- Zhang L, Li H, Li S *et al.* (2015) Phylogeographic structure of cotton pest *Adelphocoris suturalis* (Hemiptera: Miridae): strong subdivision in China inferred from mtDNA and rDNA ITS markers. *Scientific Reports*, **5**, 14009.
- Zink RM, Rohwer S, Andreev AV *et al.* (1995) Trans-Beringia comparisons of mitochondrial-DNA differentiation in birds. *Condor*, **97**, 639–649.

---

N.P.H., M.E.M and A.C. designed the study. All authors were involved in the research and N.P.H conducted the analyses. All authors contributed to the manuscript.

---

### Data accessibility

DNA sequences: Sampling locations, voucher accession numbers and GenBank accession numbers for COI sequences are listed in Table S1, Supporting information. Microsatellite genotypes have been archived with Dryad (doi:10.5061/dryad.375dk).

### Supporting information

Additional supporting information may be found in the online version of this article.

**Appendix S1** Supplementary Materials and Methods.

**Table S1** Sample information for hemlock adelgids.

**Table S2** Prior and posterior distribution of parameters for approximate Bayesian computation to test colonization from Asia to western North America using *DIYABC* (Cornuet *et al.* 2014).

**Table S3** Genetic differentiation ( $F_{ST}$ ; above diagonal) and harmonic mean of differentiation (Jost 2008) ( $D_{est}$ ; below diagonal) between hemlock adelgid sites with ten or more MLLs.

**Fig. S1** Sampling sites for hemlock adelgids in continental China and Taiwan.

**Fig. S2** Sampling sites for hemlock adelgids in Japan and Ulleung Island.

**Fig. S3** Sampling sites for hemlock adelgids in western North America.

**Fig. S4** Sampling sites for hemlock adelgids in eastern North America.

**Fig. S5** Histogram showing genetic distance under a stepwise mutation model vs. number of pairwise comparisons for hemlock adelgid microsatellite genotypes generated using *GENODIVE* (Meirmans & Van Tiernderen 2004).

**Fig. S6** Bayesian Information Criterion (BIC) vs. the number of clusters ( $k$ ) generated for discriminant analysis of principal components (DAPC; Jombart *et al.* 2010) using *ADEGENET* (Jombart 2008) for all hemlock adelgid microsatellite MLLs.

**Fig. S7** Bayesian Information Criterion (BIC) vs. the number of clusters ( $k$ ) generated for discriminant analysis of principal components (DAPC; Jombart *et al.* 2010) using *ADEGENET* (Jombart 2008) for Group 4 (Fig. 4). A  $k$  value of six was chosen to describe the data.

**Fig. S8** Detailed Bayesian tree of hemlock adelgid mitochondrial COI sequences.

**Fig. S9** Model check plot of the first two principal components of summary statistics generated using the prior distribution of model parameters, the posterior distribution of model parameters, and the observed data set.

**Fig. S10** Prior and posterior distributions of effective population sizes and divergence times for the most likely scenario of colonization from Asia to western North America (Fig. 3) resulting from approximate Bayesian computation (ABC) analysis with *DIYABC* (Cornuet *et al.* 2014).