Rates of molecular evolution and genetic diversity in European vs. North American populations of invasive insect species

Robert G. YOUNG [1],*, T. Fatima MITTERBOECK [2],*, Tzitziki LOEZA-QUINTANA [1] and Sarah J. ADAMOWICZ [1]

[1] Department of Integrative Biology & Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada;
e-mails: ryoung04@uoguelph.ca, tloezaqu@uoguelph.ca, sadamowi@uoguelph.ca

[2] Canadian Research Institute for Food Safety, University of Guelph, Guelph, ON, Canada; email: tmitterb@uoguelph.ca

Abstract. Many factors contribute to the 'invasive potential' of species or populations. It has been suggested that the rate of genetic evolution of a species and the amount of genetic diversity upon which selection can act may play a role in invasiveness. In this study, we examine whether invasive species have a higher relative pace of molecular evolution as compared with closely related non-invasive species, as well as examine the genetic diversity between invasive and closely related species. To do this, we used mitochondrial cytochrome *c* oxidase subunit I sequences of 35 species with a European native range that are invasive in North America. Unique to molecular rate studies, we permuted across sequences when comparing each invasive species with its sister clade species, incorporating a range of recorded genetic variation within species using 405,765 total combinations of invasive, sister, and outgroup sequences. We observed no significant trend in relative molecular rates between invasive and non-invasive sister clade species, nor in intraspecific genetic diversity, suggesting that differences in invasive status between closely related lineages are not strongly determined by the relative overall pace of genetic evolution or molecular genetic diversity. We support previous observations of more often higher genetic diversity in native than invaded ranges using available data for this genetic region.

INTRODUCTION

Non-native species are of large concern to natural ecosystems and can have drastic negative impacts on the native flora and fauna if they become established. An introduced species is considered invasive when it can survive introduction, establish a population, and spread or have the potential to spread further in the introduced range (Richardson et al., 2000; Blackburn et al., 2011). Insects introduced into North America are responsible for major ecological damage and result in an economical impact estimated at 32.56 billion dollars per year (Bradshaw et al., 2016). If we can better understand the factors contributing to a successful species invasion, we can better manage high-risk invasion corridors and mitigate negative effects on native populations. Some currently identified factors influencing a successful invasion include available space, available food, and the absence of predatory and parasitic organisms (Sorte et al., 2010). While it can be argued that anthropogenic factors, such as the movement of organisms across vast distances by human-mediated transportation routes, have been largely responsible for the global distribution of non-native species (Lindroth, 1957; Lockwood et al., 2005), organismal traits such as the rate of dispersal and mode of reproduction also play a role in the success of non-native species (Moravcová et al., 2015).

Studies have shown a genetic component of invasion, including the effects of additive genetic variation, selection, or genomic rearrangements (reviewed in Lee, 2002). One aspect of invasion genetics that has not been considered is the general rate of molecular evolution of lineages. We propose that at the most basic level, higher rates of molecular evolution would provide greater substrate for adaptation through introduction of new mutations. In addition, factors that increase the evolutionary pace (e.g. generation time) will affect the speed of both macro and micro evolution. As such, we investigate the rate of interspecific genetic evolution for evidence of association with invasion potential.

Invasive species may differ in the rate of molecular evolution as compared to non-invasive species due to correlation, cause, or consequence (Fig. 1). Firstly, molecular rates and invasive success may correspond because of common biological or ecological parameters that both

---

* Corresponding authors; e-mails: ryoung04@uoguelph.ca, tmitterb@uoguelph.ca. These authors contributed equally.
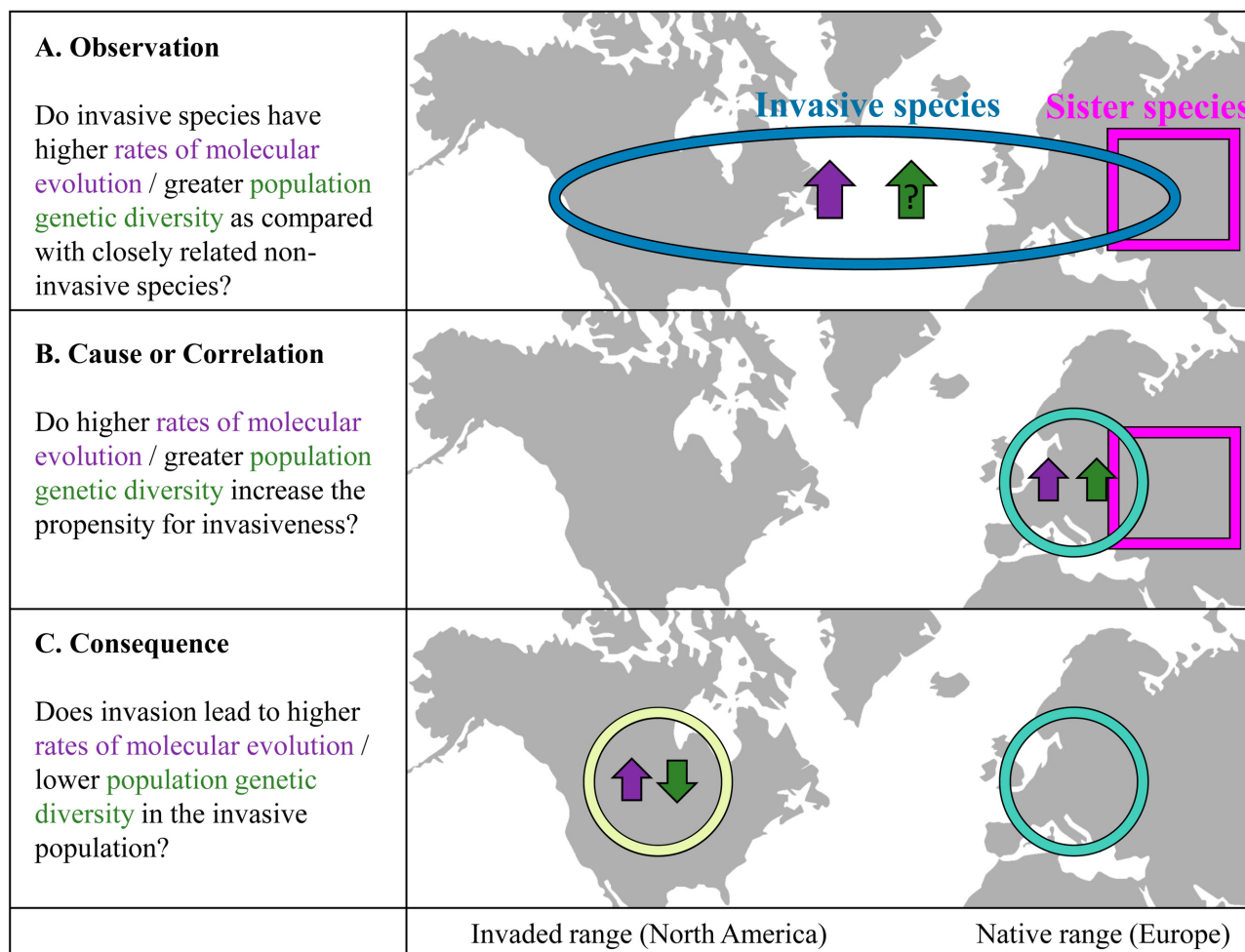
Fig. 1. Hypotheses on the link between invasive potential and molecular evolutionary rates or population genetic diversity in closely related lineages. An example geographic range for the invasive species is represented in circles/ovals while an example geographic range of the sister species is represented by squares. (A) Higher molecular rates may be observed for invasive vs. sister species. If invasive species have consistently higher rates across examples of invasive vs. sister clade species, then the explanation for such an observation could be that (B) higher molecular rates lead to invasive potential, or a common factor increases both rates and the propensity for invasion. Alternatively (C), a successful invasion could lead to higher rates of molecular evolution as compared to related species that have not undergone an invasion. We similarly investigate population genetic diversity; however, population genetic diversity may be lower in the invaded range as a consequence of invasion (C). Note that the sister clade species is not necessarily from Europe but is not known to be invasive.

promote invasion and affect molecular rates (correlation hypothesis; Fig. 1B). Biological traits that could be considered 'invasive traits' are those that favour establishment, including high fecundity and population growth rates (Lee & Bell, 1999) and fast generation time (Sakai et al., 2001). High fecundity and population growth rates are expected to lead to more mutations over time, increasing the pace of molecular evolution at DNA sites with low selective pressures. Faster generation time is expected to lead to both higher mutation rate and faster pace of evolution due to more bouts of fixation per unit time (Bromham, 2009). Thus, the possession of these traits may correspond with a generally faster rate of molecular evolution as well as promote invasion.

Secondly, molecular rates and invasion success could correspond due to a higher mutation rate promoting invasion potential (causation hypothesis, Fig. 1B). Biological traits that are thought to be more common in successful invasive species (e.g. high fecundity, Lee & Bell, 1999)

may also result in higher mutation rates (Bromham, 2011). This higher mutation rate would then provide new variation upon which selection could act, facilitating evolution within a shorter timeframe. Thirdly, biological invasion may lead to increased substitution rates via positive selection or relaxed selective constraints following invasion (consequence hypothesis, Fig. 1C). Invasive species establishment in non-native environments (Prentis et al., 2008) can increase positive selection pressures in specific genes (Roman & Darling, 2007) and could therefore increase fixation in those genes. However, we do not expect that the molecular rates would be observably changed as a consequence of invasion in this study, given the very recent occurrences (within the last approximately 180 to 20 years, Supplement S1-13) of these insect invasions and the use of COI, which has no clear link to adaptive changes in invasion success. Population genetics trends related to population molecular diversity are more likely to display trends as a consequence of invasion, such as due to population

bottlenecks occurring upon colonization of a new environment by few invading individuals.

Published research in invasion genetics, which has tended to focus primarily upon plants, has highlighted the ability for introduced invasive species to respond to selection pressures as an important factor in invasion success (Lee, 2002; Sherman et al., 2016). Additive genetic variation (Lee, 2002), i.e. when phenotypic variation of a given trait is related to the additive, independent effects of multiple genes, is thought to increase invasion success when the characteristics (e.g. phenotype) involved are relevant to the invasion. Often, a high genetic diversity within populations is invoked as facilitating invasive success (Rius & Darling, 2014). The various measures of genetic variation are often assumed to be correlated; however, empirical studies have estimated that molecular genetic variance correlates weakly (r = 0.22) with additive genetic variation (Reed & Frankham, 2001). Genetic diversity has been reported for invasive species compared to similar non-invasive species in specific target taxa (e.g. Pappert et al., 2000), but a general test of the genetic diversity in invasive vs. related non-invasive species is lacking in insects. Following the logic of Fig. 1, we test whether molecular genetic variation corresponds with (Fig. 1A) or promotes invasive success (causation hypothesis, Fig. 1B); however, we acknowledge that molecular genetic variation may not be useful for inferring or predicting processes of adaptation (except in the individual gene investigated), as with additive genetic variation.

There is great interest into how newly introduced non-native species, often with reduced genetic variation in comparison with variation in the native species range, can adapt so rapidly in new environments (Allendorf & Lundquist, 2003; Roman & Darling, 2007; Schrieber & Lachmuth, 2016), but there are few studies that have synthesized the extent of this reduction in genetic variation in insects. Molecular genetic data are useful to infer effects of more neutral factors including population size, gene flow, and population structure (Reed & Frankham, 2001). The Dlugosch & Parker (2008) synthesis of literature included investigations of allelic diversity in nuclear genes for 13 invasive insect species; they concluded reduced allelic diversity and heterozygosity in introduced populations as compared to the native range populations. However, there have been no large studies on population genetic trends in mitochondrial loci between native and introduced ranges in insects that applied consistent methods across all comparisons. To address this gap, we test whether there is a general trend of reduced genetic diversity in the invaded vs. native range of invasive insect species that have invaded North America from Europe (consequence hypothesis, Fig. 1C).

The two measures of molecular evolutionary rates and genetic diversity we will examine here are not directly related and could exhibit different trends. An elevated mutation rate alone could lead to an increase in both intra- and interspecific variation. However, often, intraspecific genetic diversity and interspecific lineage rates do not correspond due to their relationship with effective population size (which affects the rate of mutational fixation), and instead they can be inversely related (Fujisawa et al., 2015). As well, genetic diversity can be increased through other avenues such as repeated introductions and hybridization (Roman & Darling, 2007). Thus, we test both interspecific rates and intraspecific diversity to investigate whether either corresponds with invasion potential and success.

MATERIALS AND METHODS

Study design

We use a widely available region of the cytochrome *c* oxidase subunit I gene (COI) to test 35 European to North American invasive insects for rates of molecular evolution as compared to related non-invasive species from the closest available sister clade. Firstly, we test the rates of molecular evolution in the invasive species as a whole (all geographic locations) vs. related sister clade species (Fig. 1A). To differentiate cause/association vs. consequence in any observed trends in relative molecular rates (Fig. 1), we examine whether members of the invasive species occupying the native range (Europe) have higher molecular evolutionary rates than closely related non-invasive species, which would suggest a causative influence or an association of rates with invasion (Fig. 1B, cause/association). To test for an effect of invasion on rates (consequence hypothesis), we observe whether invasive lineages in the invaded range have higher rates than those in the native range (Fig. 1C). Similar to tests on molecular evolutionary rates (Fig. 1), we also test population genetic diversity for 32 of those invasive insect species, examining the genetic diversity in invasive species (whole range) vs. sister clade species (as in Fig. 1A, correlation), genetic diversity in the invasive species' native ranges vs. sister clade species (as in Fig. 1B, causation/association), and genetic diversity within invasive species in their invaded vs. native range (as in Fig. 1C, consequence).

Invasive insect data acquisition and data verification

A list of European invasive insects into North America was obtained from the Canadian Wildlife Federation website (http://cwf-fcf.org/en/; download performed May 2015). A literature review using a Web of Science search (criteria in S1-1) was conducted in May 2016 to support the selection of target species as invasive in North America, based on peer-reviewed research (References in S1-1). Using the Barcode of Life Data (BOLD) Systems (Ratnasingham & Hebert, 2007), we verified that there were COI sequence data available for each invasive insect species of interest (referred to here as "target" species) by obtaining all Barcode Index Numbers (BINs) associated with the Latin names of the target species. A Barcode Index Number (BIN) references a group of nucleotide sequences of the COI-5P region, which represent species-like units, based on the Refined Single Linkage molecular clustering method (Ratnasingham & Hebert, 2013). BINs associated with our target species were validated for correspondence with our target species: first, BINs were used if the majority of records in the BIN belonged to the target species or a synonymized name. When a single morphological species name was prevalent in multiple BINs, indicating potential cryptic diversity within the named species, all of the obtained BINs were included in analyses (this occurred in 4 of our 35 tested invasive species. See S1-3).

Using the BOLD system, we constructed preliminary phenograms using the neighbour-joining (NJ) method with Kimura-2-parameter (K2P) distances using all publicly available sequence data on BOLD (minimum 400 nucleotides in length, no flags or stop codons) for the entire genus containing each of our identified target species. After visual inspection, if the resulting NJ phenogram did not contain four successive deeper nodes from

the target species lineage, the genus group was not used, and the process was repeated for all sequences within the subfamily or family containing a target species. Using the appropriate genus or family name, we then downloaded all available sequence data from BOLD using the BOLD public data API.

Each genus or family-level data set was globally aligned in MAFFT Ver. 7 (Katoh & Standley, 2013), followed by trimming and verifying the amino acid alignment by eye using MEGA6 (Tamura et al., 2013). The multiple sequence alignment (MSA) was then reduced by eliminating sequences exhibiting ≥ 98% similarity using ElimDupes (https://hcv.lanl.gov/content/sequence/ELIMDUPES/elimdupes.html) to reduce computational demands and facilitate maximum likelihood phylogenetic analysis. In two alignments that had fewer and more closely related sequences (genus-level alignments), the criterion was changed to ≥ 99% similarity in order to keep all BINs in the alignment. Model testing was completed using MEGA6 on the generic or sub/family alignments, and the model with lowest Bayesian Information Criterion (BIC) score was selected (S1-2). Maximum likelihood (ML) trees were constructed using MEGA6, using the best-fit model of nucleotide substitution and 1000 bootstrap pseudoreplicates to indicate node support.

For each genus or subfamily/family containing a target invasive species, sister lineages to the target species were identified using the ML tree. If the node connecting the target and sister clade had a bootstrap support value of 70% or greater, the 2nd-branching lineage to the target+sister clade was used as an outgroup. If a bootstrap support value of less than 70% was present at this node, then the 3rd-branching lineage from the target+sister clade was used as an outgroup. The 3rd-branching outgroup to the target species was selected in cases of low bootstrap values of closer relatives in order to provide confidence that the outgroups were indeed phylogenetically outside the ingroups; otherwise, incorrect conclusions could be drawn about the relative molecular evolutionary rates between the target and sister lineages. If the ML tree did not contain the appropriate number of branching outgroups, a new sequence download and ML tree construction was performed using the next higher taxonomic level (for example moving from a genus download to a subfamily download).

BINs associated with sequences in the sister lineage were obtained. These BINs were checked to ensure that they contained individuals bearing a Linnaean taxonomic identification to species level on the BOLD system. The sister clade BIN species names were then used to complete a Web of Science search to check for evidence of invasiveness using the same process and criteria used to check the target species. Sister clade BINs were removed from analysis if literature evidence was found of introduction and establishment in non-native ranges, as we aim to compare known invasive target species to sister clade species not known to be invasive. We acknowledge that some of the sister clade species could have invasive tendencies not yet reported in the literature or may not have had dispersal opportunities to non-native regions; however, the target species are well-known invasive species, and we expect that they differ, on average, in some element of invasiveness from their sister clade species.

For the target and sister clade BINs selected above from each genus or subfamily/family group ML tree, all publicly available COI sequences were downloaded from BOLD. BINs were used to download sequence data, as opposed to downloading based on taxonomic identifications, to include as many sequences as possible, including those within these BINs that currently lack low-level taxonomic identifications. The outgroup sequences used to construct the ML trees were added to the alignments with the target and sister downloaded sequences, and each sequence set was aligned and trimmed as above. Alignments were verified to be in

reading frame, and alignments lengths were all of a multiple of three for later analysis. To ensure that high-quality and consistent data were present for further analysis, sequences with greater than 1% unknown nucleotides (N's and/or gaps) were removed by a Perl script (S2-6). Unlike above, these alignments were not reduced by eliminating identical or similar sequences, as all target and sister sequences were necessary for further down-stream analyses. These versions of the alignments (hereafter called 'target sequence alignments' identified by the invasive species name, e.g. *Yponomeuta malinellus*) were used in "Genetic diversity analysis" further below.

Relative rates analysis

Three separate analyses (North American regional, European regional, and total) were conducted using each of the 'target sequence alignments'. The first two analyses used geographic regions of collection (North America, Europe) to conduct independent analyses of relative rates of molecular evolution. This enabled comparisons between geographic regions. To accomplish this, the sister clade sequences were reduced to unique sequences in R v.3.2.4 (R Core Team, 2016) for North America and Europe separately (all R scripts provided in S2-1 to 5). The target sequences were reduced to unique sequences by geographic region for conducting a rates analysis for the species by region without duplicate sequences. The third analysis conducted was a total data analysis without considering geographic region. The 'target sequence alignments' were reduced to unique sequences for the entire alignment and used for further analyses to compute molecular evolutionary rates using all available data.

Species trees were assembled as input for rates analysis and included a representative target sequence, sister sequence, and outgroup sequence (Fig. 2). A single outgroup was used at a time as the use of multiple outgroups does not significantly improve rate estimations in relative rate tests (Robinson et al., 1998). A balanced (1 vs. 1) choice of target and sister sequence was used to remove bias arising from the node-density effect (Robinson et al., 1998). The program baseml in package PAML version 4.7 (Yang, 2007) was used to estimate the length of each branch on the 3-species trees. For each of the three analyses — North American, European, and total — all possible combinations of target sequences, sister sequences, and outgroup sequences were permuted for rates analyses using R. The nucleotide model results previously obtained from MEGA6 for conducting the taxon-specific ML trees were used for rates analysis using the target sequence alignments; the best (lowest BIC score) model without G or I parameters was used for PAML analysis (S1-2). Non-synonymous substitution (dN) rates, synonymous substitution (dS) rates, and the ratio between the previous two (dN/dS) were obtained for each target and sister lineage similarly using the *codeml* program in PAML.

Relative molecular rates are uncertain for recently diverged lineages. While the rate results can be trimmed for uncertain estimates (e.g. Welch & Waxman, 2008), in this analysis all of our contrasts are on short time frames, and most branch lengths are necessarily short (but greater than 2% sequence divergence in all cases). In cases where species sampling within a genus was lower, the branch lengths between contrasts are expected to be longer and the relative rates less uncertain; however, the biological character of interest (invasiveness) still only occurs at the species level, and thus comparisons on more highly-divergent lineages more poorly represent a contrast in character state. Due to the recent divergences of all comparisons (i.e. closely related species within genera), we elected not to eliminate contrasts with less certain estimates or to weight results based on divergence. From the rate estimates obtained from the PAML program, rela-
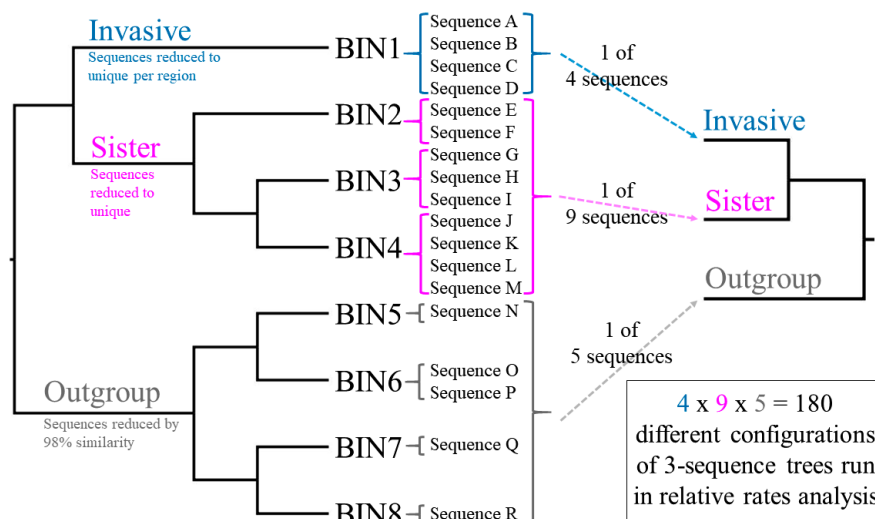
Fig. 2. Comparison set-up and permutation approach for the molecular rates analysis. From the maximum likelihood trees built using COI sequences, BINs of the target invasive species, BINs belonging to the closest sister lineage, and outgroup sequences in the 2nd or 3rd branching lineage from the ingroup were used. One unique sequence per target and sister clade BIN, and an outgroup sequence, formed a 3-species tree used to estimate branch lengths for the invasive vs. sister lineage. All possible 3-sequence combinations were run, with the data points representing the relative invasive:sister substitution rates presented in Fig. 3.

tive rates between target and sister clade species were calculated. Firstly, for each analysis, the larger divided by the smaller substitution rate (minus one to center final values around 0) were examined. These relative rates were signed based on direction whereby positive indicated a higher target species rate than sister rate, and negative was assigned for a higher sister clade species rate than target species rate. Relative dN rates, dS rates, and dN/dS ratios were obtained by a different method where we compressed the relative rates between -1 and 1 through taking '1-smaller/larger' and signing based on direction (target rate > sister rate is positive, sister rate > target rate is negative) (as in Wright et al., 2006; Mitterboeck et al., 2016) in order to account for low values without displaying extreme results.

To determine whether target vs. sister clade species exhibited different relative rates, binomial and Wilcoxon signed-rank tests were performed using the medians of the relative substitution rates, dN rates, dS rates, and dN/dS ratios across the 35 comparisons for the total data set category. This analysis was repeated with the removal of 8 species that were purposefully introduced into North America (indicated in S1-13), since we have no hypothesis for the molecular evolutionary rates of these species as a causative factor in invasion potential. Secondly, we compared the relative position of the North American and European population medians for all comparisons having both sets of data points; this was performed using a binomial and Wilcoxon signed-rank test for their relative position and difference; e.g. if the North America median was a higher value than the European, then the result was signed as positive. The relative breadths of range in the relative rates between regions were also compared using the differences between regions by 2-tailed binomial and Wilcoxon signed-rank tests.

Genetic diversity analysis

The 'target sequence alignments' were used to calculate haplotype numbers, haplotype diversity, and nucleotide diversity in DNAsp version 5.10 (Librado & Rozas, 2009). Haplotypes were recognised as sequences differing in two or more nucleotides, not counting unknown nucleotides. BINs were used for the analysis, though most often the Linnaean species name matched a single BIN. For each set of sequences from invasive BINs, diversity statistics were calculated for the total set of sequences, the European sequences, and the North American sequences where there were two or more sequences available per BIN and geographic location. Two-tailed binomial tests were performed on the number of sequences, haplotype numbers, and diversity measures for North American vs. European sequences in each invasive species BIN to test the hypothesis of consequence (Fig. 1C); two-tailed Wilcoxon signed-rank tests were similarly performed after taking 1-smaller/larger value and signing based on direction (e.g. Europe greater is positive, North America greater is negative). We additionally test whether the species that were purposefully introduced into North America represent a difference in their proportion between North American > European and European > North American genetic diversity results, using a Fisher's Exact test in R.

Similarly, we calculated the number of haplotypes, the haplotype diversity, and the nucleotide diversity for the sets of sequences in each associated sister clade BIN and compared these metrics to the results obtained for each invasive species BIN (corresponding to testing for a general observation between invasive species and molecular genetic diversity, Fig. 1A). We repeated the diversity measures for European vs. sister clade BIN sequences only, since the European geographic region relates to the hypothesis of *cause* (Fig. 1B). Binomial and Wilcoxon signed-rank tests were performed as described above. Since there can be multiple invasive or sister clade BINs within a paired comparison, we compare each invasive species BIN to all possible sister clade BINs (using 1-smaller/larger and sign) and summarise those results per invasive BIN by the median. We then summarise results for multiple BINs within a single invasive species name by taking the median of the genetic diversity metric from the various sister pairings. We repeat all summaries and Wilcoxon signed-rank tests using only those comparisons with higher sample size — 6 or more sequences in each BIN or region — since the subsampling of 6 sequences from a population allows the distinction between high and low population estimates of haplotype and nucleotide diversity in the COI gene in animals (Goodall-Copestake et al., 2012).
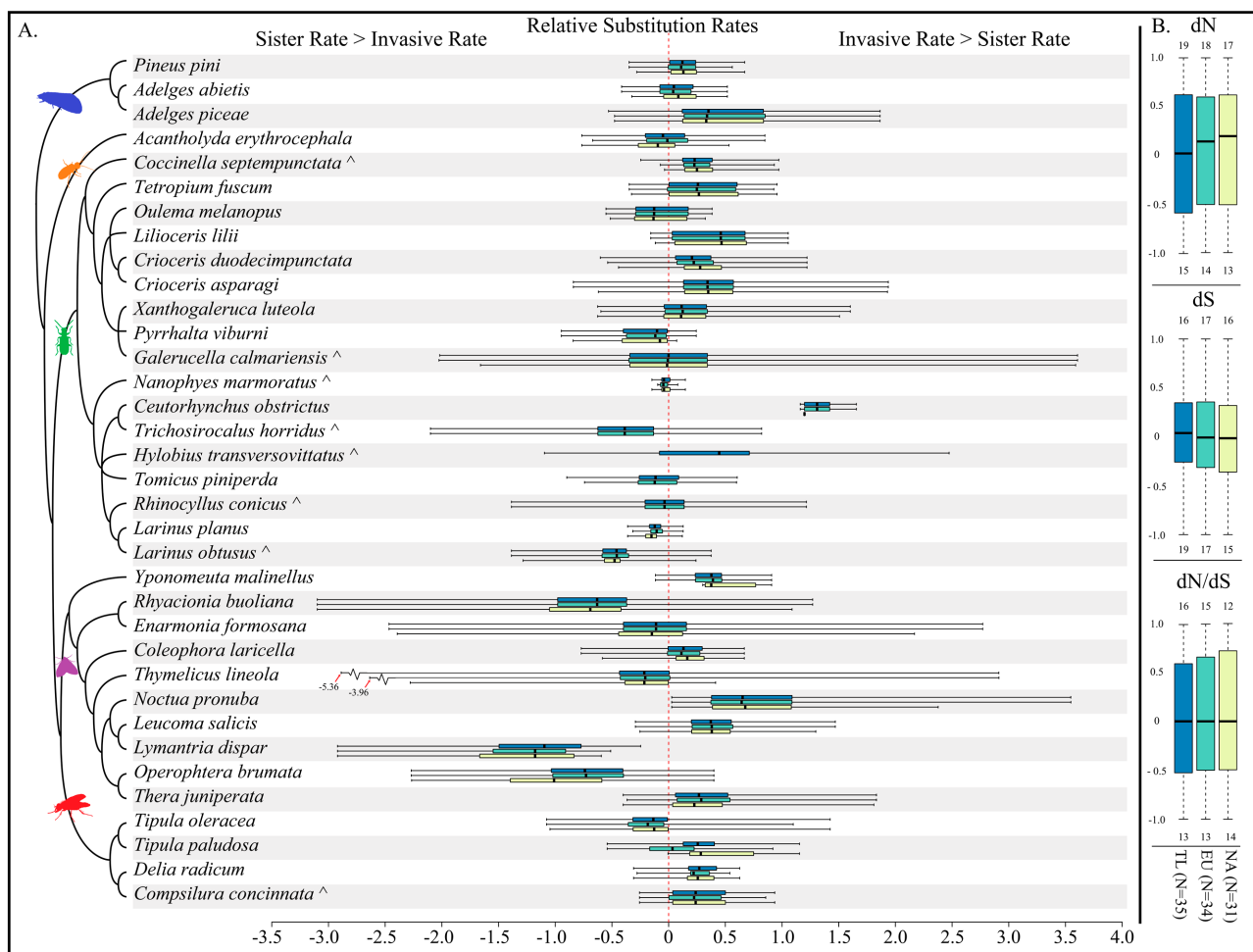
Fig. 3. Thirty-five comparisons of relative molecular evolutionary rates between related invasive and non-invasive insect species. The distribution of data points representing the relative invasive : sister molecular rates are shown in panel A. Positive relative rates represent data points for which the invasive rate > sister rate, and negative where the sister rate > invasive rate. Yellow (light) bars indicate results of sequences of individuals that were collected in North America (NA); green (medium weight shading) indicates European (EU) individuals; blue bars (dark weight shading) indicate the total (TL) data points from all regions and those without any location information. The thick vertical lines on the boxes represent the median relative rate. Phylogenetic relationships shown to the left are based on taxonomy, as well as molecular phylogenetic relationships from Hunt et al. (2007), Regier et al. (2009), McKenna et al. (2009), Wiegmann et al. (2011), and Misof et al. (2014). Panel B shows the distribution of the medians of relative dN rates, dS rates, and dN/dS ratios across the 35 comparisons, where each data point in the boxplot is a median of all relative rates belonging to a single invasive vs. sister comparison. The number of positive (invasive > sister) and negative (sister > invasive) medians are given above and below each boxplot, respectively. Zero values are included on the graph but not tallied. Species purposefully introduced into North America are marked with '^'.

RESULTS

Relative substitution rates

The relative rates for all invasive-sister pairs, along with their medians, are plotted in Fig. 3A. 405,765 combinations of invasive, sister, and outgroup sequences were analysed for relative substitution rates. In 19 out of 35 invasive species analysed, the medians of the substitution rates were positive, indicating that the invasive rate was higher than the sister clade species rate ($p_{binomial[b]}$ = 0.74, $p_{Wilcoxon[W]}$ = 0.23) (Fig. 3A, blue/darkest bars). This result corresponds with the question in Fig. 1A (observation). The median of the rate medians was +0.11, signifying that the invasive species molecular rate was 11% higher (by median) than the sister rate in 50% or more of the invasive species tested. The results were similar when the purposefully-introduced

species were removed (16 out of 27 medians positive, p_b = 0.44, p_W = 0.16).

Similarly, the medians of the European sequences did not differ significantly from the null expectation. This result corresponds with the question in Fig. 1B (cause/association). Neither North American nor European comparisons had more often high median relative rates than the other region for comparisons possessing data from both regions (14 Europe higher, 16 North America higher, p_b = 0.88, p_W = 1.0). This result corresponds with the question in Fig. 1C (consequence). The range in relative substitution rates was larger for European data points than North American data points, with 22 out of the 31 comparisons with data from both regions having a larger range of relative substitution rates for the European data (p_b = 0.029, p_W = 0.0014).

Thirty-two of 35 comparisons had both positive and negative data points for the total data set of permuted se-
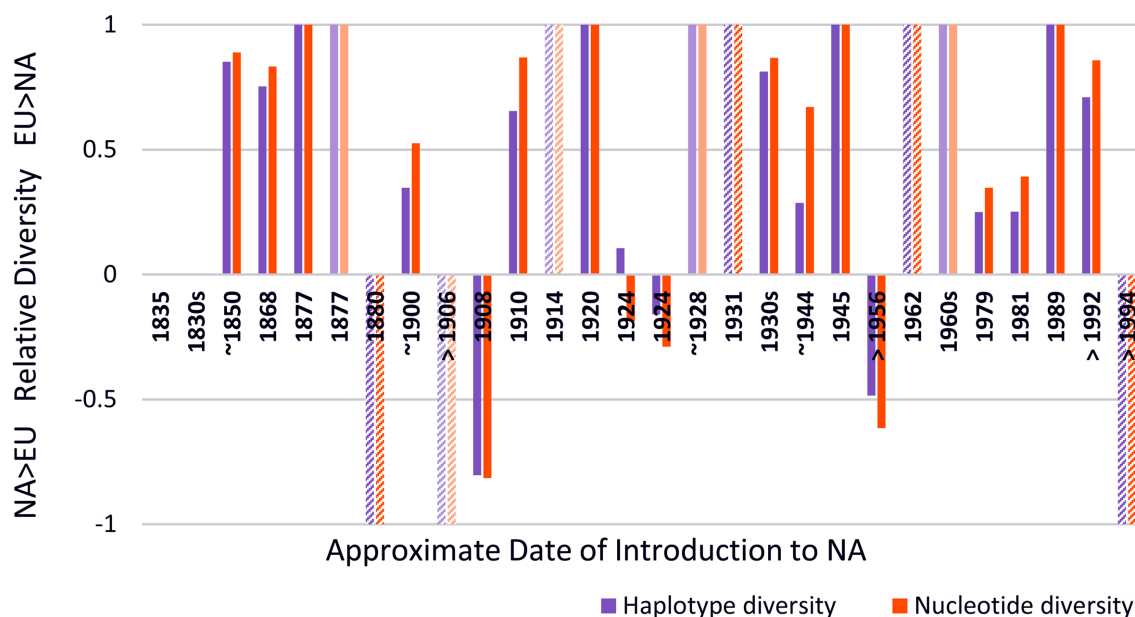
Fig. 4. Haplotype diversity and nucleotide diversity estimates between populations in the native range (Europe [EU]) and the invaded range (North America [NA]) in twenty-eight invasive insect species. The European population had greater haplotype diversity (orange bars above the x axis) in 20 of 26 cases (p_b = 0.0094, p_W = 0.014) and greater nucleotide diversity (purple bars above the x axis) in 19 of 26 cases (p_b = 0.029, p_W = 0.012), with two cases of equal (0) diversity measures, as compared to the population in North America (bars below the x axis). The diversity measures were made relative by taking 1-(smaller/larger) and signing based on direction (EU > NA signed as positive, NA > EU signed as negative). The bars are ordered by approximate date of introduction of the species (earliest to latest, with '~' indicating approximate); the lighter bar colouration indicates that there were fewer than 6 sequences available for one of the regions (5 sets of bars, and 2nd 0 result); the dashed bars indicate cases where the number of sequences collected between regions differed by 5 or more times (6 sets of bars). Species purposefully introduced into North America are marked with '^'.

quence groupings. This signifies that the varying choice of a single sequence each from a target, closest sister, and outgroup BIN can give different directional results, i.e. that the invasive rate was greater than the sister, or vice versa. However, in the 32 cases that span zero, 20 cross by a tail quartile of data. Therefore, this signifies that less than 25% of the data points within each of those comparisons would yield an opposite conclusion.

Relative dN/dS ratios, dN rates, and dS rates

The relative dN rates, dS rates, and dN/dS ratios for all invasive-sister pairs, along with their medians, are plotted in Fig. 3B. There was no significant difference in the frequency of higher rates between invasive vs. sister clade species. Neither the North American nor European region had significantly higher dN/dS ratios, dN rates, or dS rates than the other region (p values in S1-9).

Genetic diversity by region and in invasive vs. sister clade species

The number of invasive species sequences available was similar between geographic regions (Europe and North America), with more sequences in the European region in half (14 of 28) of the invasive species ($p_{binomial}$ = 1.0, $p_{Wilcoxon}$ = 0.57). However, the number of haplotypes (21 of 26, p_b = 0.0025, p_W = 0.0033), the haplotype diversity (20 of 26, p_b = 0.0094, p_W = 0.014), and the nucleotide diversity (19 of 26, p_b = 0.029, p_W = 0.012) were each significantly more often higher in the European region than in North America by both binomial and Wilcoxon signed-rank tests (Fig. 4, corresponding with the question in Fig. 1C). The results

excluded two cases of equal measures. The directional results were relatively consistent among the metrics of genetic diversity, with species exhibiting higher diversity in Europe 81% of the time for haplotype number, 77% of the time for haplotype diversity, and 73% of the time for nucleotide diversity. This finding was similar when examining only pairs represented by 6 or more sequences from each region (full results in S1-11). The species displaying North American > European genetic diversities represented a greater proportion of purposefully introduced species than the species displaying European > North American genetic diversities (3 of 6 vs. 1 of 19 cases, $p_{Fisher's}$ = 0.031, for 25 cases where the haplotype and nucleotide diversity were in the same direction).

Invasive species had more sequences and haplotypes (by medians) than the sister clade species; in 28 of 32 comparisons the invasive had more sequences (p_b = 1.9 × 10^{-5}, p_W = 5.0 × 10^{-7}), and in 27 of 32 comparisons the invasive had more haplotypes (p_b = 0.00011, p_W = 0.00071). Nevertheless, the haplotype diversity and nucleotide diversity metrics, which consider number of sequences, did not differ significantly between invasive and sister clade BINs (corresponding with the question in Fig. 1A) (haplotype diversity: 14 of 32, p_b = 0.60, p_W = 0.29; nucleotide diversity: 13 of 32 with the invasive higher than the sister, p_b = 0.38, p_W = 0.14). When considering only comparisons with higher sequencing sample size (6+ per species), the number of haplotypes did not differ significantly; instead, the haplotype diversity (7 of 27; p_b = 0.019, p_W = 0.0036) and nucleotide diversity (7 of 27; p_b = 0.019, p_W = 0.00074) were signifi-

cantly more often greater in the sister clade species in both the binomial and Wilcoxon signed-rank tests. When considering only the invasive species' European sequences vs. the sister clade species (corresponding with the question in Fig. 1B), the number of sequences was again higher in the invasive species, and again the haplotype and nucleotide diversity did not differ significantly. Similar to the results when including data from all regions, for those pairs with 6+ sequences/species, the haplotype diversity (7 of 23, p_W = 0.017) and nucleotide diversity (8 of 23, p_W = 0.013) of the sister clade species were significantly more often higher by Wilcoxon signed-rank test as compared to the invasive species (S1-11).

DISCUSSION

Relative substitution rates and interpretation of cause or consequence

Invasive species did not display higher molecular substitution rates more often than closely related non-invasive species as seen by the medians of the relative rates (Fig. 3A). These results provide no support that invasion occurs as a consequence of faster rates of molecular evolution (Fig. 1A represented by Fig. 3A 'Total', Fig. 1B in Fig. 3A European region). The North American sequences also did not have consistently higher relative rates than European sequences (Fig. 1C represented by relative position of European [EU] and North American [NA] region bars in Fig. 3A), which suggests rates were not increased as a consequence of invasion.

The short evolutionary timeframes investigated between target and sister clade species (a few million years) has not provided a lot of opportunities for differential substitution accumulation. Because of this, it is difficult to distinguish whether the null result in molecular rates is a general trend in species-level contrasts or due to uncertainty in measurement of rates. A significant directional trend, however, would have suggested that rates of molecular evolution are associated (through cause, association, or consequence) to invasive potential in insects on short (and likely long) timeframes. Furthermore, we elected only to explore overall trends, rather than make more subtle considerations such as estimate the degree of invasiveness for each species. Our results were a first step toward addressing the question of general molecular rates and invasive potential. Future investigations of this question through species-level contrasts may require much larger sample sizes, or investigations could include taxonomically higher clades containing invasive species to provide longer timeframes so that rate differences would be more apparent. For example, testing molecular rate differences between clades differing in their proportion of known invasive species would be an interesting avenue for a future study. Currently, this would be difficult to accomplish accurately and on a worldwide scale as there are few comprehensive lists of invasive species (Rilov & Crooks, 2008; Clout & Williams, 2009) and none in insects (Foottit & Adler, 2009). Our use of the medians of relative rates obtained through multiple rates calculations using all available combinations of invasive,

sister, and outgroup sequences helped to reduce variation introduced through small sample size and the stochasticity of genetic change on short time scales. The greater range in the relative rates of molecular evolution in European vs. North American populations corresponded with a greater number of haplotypes observed for the European population.

Relative dN rates, dS rates, and dN/dS ratios

Examination of the different types of substitutions (synonymous/non-synonymous) provides more details on the effects of selection and mutation in the COI gene of invasive vs. related species, providing a broader understanding on how COI acts both as a representative of genome-wide trends and may experience differing selection pressures itself. The relative synonymous substitution (dS) rate can serve as a proxy for relative mutation rate differences. There was no significant difference between invasive and sister clade species dS rates, suggesting that invasive species do not have a higher mutation rate, either as a causative factor in their success in invasion or as a correlation with other biological or ecological traits that could both promote invasion and impact the mutation rate. Similarly, the invasive species in the native range did not have greater dN rates than the corresponding sister clade species, suggesting that evolutionary pace (such as due to faster generation time) is not higher in the invasive species. The North American sequences did not have significantly more often higher non-synonymous rates than the European sequences. This suggests no trend in adaptive evolution in the COI gene, neither positive selection (e.g. Scott et al., 2011) nor relaxation of selective constraints (e.g. Mitterboeck & Adamowicz, 2013), in conjunction with the population becoming established in the new environment. A trend in COI related to positive selection was not necessarily expected; while metabolism and other gene types have been shown to be under positive selection in certain invasive insect species (Wang et al., 2011), there lacks evidence for selection in COI associated with invasive success. Our interpretations of selection are with respect to a single mitochondrial marker, and we acknowledge that different selective forces may be acting on additional genetic markers relevant for invasion-related traits.

Genetic diversity

The literature suggests additive genetic variance in a source population as a facilitator of successful invasion (Lee, 2002), with additive genetic variance linked weakly with molecular genetic variation (Reed & Frankham, 2001). We observed more haplotypes present in invasive species than related sister clade species. However, the invasive species also had more often a greater number of sequences available than closely related species. Due to the nature of data collection for this study, we cannot be certain if available sequences are representative of the true population make-up for the included species. This difference in research effort could be a direct consequence of a species having 'invasive' status being more available or of higher interest to researchers. Despite these limitations

of sample size, the genetic diversity results displayed the opposite trend compared to what might be expected based upon sample size of sequences alone. When considering haplotype diversity and nucleotide diversity measures, which do consider sample size, the sister clade species typically exhibited higher diversity than the paired invasive species for the subset of comparisons that included 6 or more sequences per BIN. To truly consider whether genetic diversity influenced the propensity for invasive potential, the source population should be examined. While the sequences obtained from the native range may not represent all sequences from which the invasion was derived (due to sampling effort), it provides a point of comparison to evaluate diversity relative to the sister clade species. When considering the European region only, the pattern was similar between invasive species vs. sister clade species, with higher diversity in the latter. Thus, our data do not provide support toward the assertion of genetic variation facilitating invasion, since we did not observe any significant difference in the European region of the invasive species as compared to the sister clade species. However, we used COI as a representation of molecular genetic variation; our analyses did not consider additional genes, which may be more relevant to invasion-related traits.

Genetic diversity for invaded geographic ranges cited in the literature is often observed to be reduced as compared to the native range of the invasive species (e.g. Kliber & Eckert, 2005); there are, however, some exceptions to this reduced genetic diversity in the invaded range (Klobe et al., 2004). We observed more often a greater genetic diversity in the native European geographic range as compared with the invaded North American range, as seen in the number of hapotypes (81% of species), haplotype diversity (77%), and nucleotide diversity (73%). Our work included information for 28 target species having sequences from both geographic regions. There were 6 cases in which the haplotype and nucleotide diversity was higher in the invaded range, contrary to expectation. However, most (4) of these cases appear to correspond with potential explanations: in 3 cases the species were purposefully introduced, as well in 3 cases the sequence availability was highly unbalanced (5 times different), with higher diversity measures corresponding with the region having the much greater sequence availability. While the species in these regions were not exhaustively sampled, the publicly available data from BOLD represents the most wide-spread effort to generate DNA sequence data across as many species as possible, allowing us to perform a broad-scale analysis using consistent methods. Our collated data suggest a general trend of higher intraspecific variation in the source range as compared to the invaded range for invasive insects. This result is similar but less consistent in direction to the synthesis of Dlugosch & Parker (2008), in which 11 of 13 (85%) invasive insect species with various countries of origin and invasion had greater heterozygosity in the native ranges. Our species lists did not overlap with that study.

Taxon representation

Our included invasive species represented 5 of 32 insect orders, and the genera and sub/families included here often contained more than one invasive species analysed. The most common families included in our analysis, the beetle families Curculionidae and Chrysomelidae, are both species rich with over 60,000 and 36,000 species, respectively (Foottit & Adler, 2009). A higher number of invasive species would be expected in these groups, as compared to less-specious insect families, simply by chance. Out of our included insect orders (Diptera, Hymenoptera, Coleoptera, Lepidoptera, and Hemiptera), Hemiptera has been observed to be disproportionally represented by invasives as compared to native species in North America (Yamanaka et al., 2015).

Considerations and limitations of the work

Here we used a simple measure of genetic differentiation between invasive and closely-related non-invasive species. We did not test for genomic selection in various genes. While the genetics of invasion are undoubtedly more complex than simple rates of molecular evolution, the relative molecular rate is one potential correlate of invasiveness that had not been directly addressed to date.

When selecting our molecular marker for this study, we gave preference to COI due to the larger number of publicly available sequences, in terms of both number of sequences and breadth of taxon sampling, which would have been reduced by using a multiple marker data set. Due to the use of a single marker, relationships constructed may be considered gene trees rather than species trees. Given the phylogenetic uncertainty, we were cautious in our choice of outgroup (taking the 2nd or 3rd branching outgroup, based on node support) in order to limit the possibility of our outgroup choice being closer to the target or sister lineages, thereby influencing our analysis results. Furthermore, COI has been shown to produce fairly accurate relationships for lower taxonomic levels when compared with multi-gene trees (e.g. Wilson, 2010; Boyle & Adamowicz, 2015). We chose to collect data for species introduced from Europe into North America to reduce latitudinal differences as a confounding factor on our molecular evolutionary rates analysis. The sister clade species used in our analyses were not necessarily native to Europe, which could introduce noise due to geographic region of origin, such as potential latitudinal effects on genetic diversity or molecular rates. However, since multiple comparisons were used that included data from various individuals and various geographic collection ranges, this aids to mitigate confounding geographic issues by assessing larger trends.

Implications of the permutation method

Permuting through equal choices of individuals within a species and through species within the same phylogenetic range may be useful in two ways: the first is in avoiding node-density effects while still considering information from multiple terminal lineages; the second is in the demonstration of how sequence or lineage choice can influence

the result obtained for molecular rates studies in simple three-taxon trees. Ninety-one percent (32 of 35) of our comparisons using all available (total) data spanned both positive and negative sides of the null result. In 63% (20 of 32) of cases, the crossing was only by a tail quartile. The choice of sequences to represent a target, sister, or outgroup species, such as with multiple sister clade BINs and outgroup sequences, can greatly influence the resulting relative rates calculations and thereby the possible interpretation of the data. Variation in rates results, introduced by species choice, was evident. However, even in the *Ceutorhynchus obstrictus* comparison, which had a single target BIN (6 sequences), sister clade BIN (1 sequence), and outgroup (1 sequence), the relative rates results were variable. Thus, our study suggests that whether or not real biological differences are present, the choice of a single sequence to represent a species may not reflect the majority of the data or demonstrate the variability of results available. Since relative rates studies may opt for a single sequence per sister lineage for analysis, to avoid the node-density effect, we suggest that the issue of sampling effects requires further attention in molecular rates research.

CONCLUSION

We have investigated the relationship between rates of molecular evolution in the COI gene, which is commonly used in identifying specimens to species, and the trait of invasiveness in insect species. No definitive trends were observed in rates; however, it remains possible that trends exist when considering longer time frames. Using permutation in the analysis of rates, we showed that variation in relative rates exists within a species or BIN, and unique sequence choice using simple trees impacts rate estimates. Similar permutative analysis using data at various taxonomic levels and a comparison across these levels is suggested for future efforts. Our intraspecific diversity results also do not support the hypothesis that higher genetic variance promotes invasive potential, since we did not observe a higher genetic diversity in the invasive source population than in closely related non-invasive species. However, our results indicate that the intraspecific diversity in a mitochondrial gene within invasive species is significantly more often higher in the source geographic range than in the invaded ranges, using a large sample size of regional comparisons in insects, consistent with previous findings using nuclear markers.

REFERENCES

ALLENDORF F.W. & LUNDQUIST L.L. 2003: Introduction: population biology, evolution, and control of invasive species. — *Conserv. Biol.* **17**: 24–30.

BLACKBURN T.M., PYŠEK P., BACHER S., CARLTON J.T., DUNCAN R.P., JAROŠÍK V., WILSON J.R.U. & RICHARDSON D.M. 2011: A proposed unified framework for biological invasions. — *Trends Ecol. Evol.* **26**: 333–339.

BOYLE E.E. & ADAMOWICZ S.J. 2015: Community phylogenetics: assessing tree reconstruction methods and the utility of DNA barcodes. — *PLoS ONE* **10**(6): e0126662, 18 pp.

BRADSHAW C.J.A., LEROY B., BELLARD C., ROIZ D., ALBERT C., FOURNIER A., BARBET-MASSIN M., SALLES M.J., SIMARD F. & COURCHAMP F. 2016: Massive yet grossly underestimated global costs of invasive insects. — *Nat. Commun.* **7**: 12986, 8 pp.

BROMHAM L. 2009: Why do species vary in their rate of molecular evolution? — *Biol. Lett.* **5**: 401–404.

BROMHAM L. 2011: The genome as a life-history character: why rate of molecular evolution varies between mammal species. — *Philos. Trans. R. Soc. Lond. (B)* **366**: 2503–2513.

CLOUT M.N. & WILLIAMS P.A. 2009: *Invasive Species Management: A Handbook of Principles and Techniques.* Oxford University Press, New York, 320 pp.

GOODALL-COPESTAKE W.P., TARLING G.A. & MURPHY E.J. 2012: On the comparison of population-level estimates of haplotype and nucleotide diversity: a case study using the gene cox1 in animals. — *Heredity* **109**: 50–56.

DLUGOSCH K. & PARKER I. 2008: Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. — *Mol. Ecol.* **17**: 431–449.

FOOTTIT R.J. & ADLER P.H. 2009: *Insect Biodiversity: Science and Society.* John Wiley & Sons, Chichester, 632 pp.

FUJISAWA T., VOGLER A.P. & BARRACLOUGH T.G. 2015: Ecology has contrasting effects on genetic variation within species versus rates of molecular evolution across species in water beetles. — *Proc. R. Soc. (B)* **282**: 20142476, 9 pp.

HUNT T., BERGSTEN J., LEVKANICOVA Z., PAPADOPOULOU A., JOHN O.S., WILD R., HAMMOND P.M., AHRENS D., BALKE M., CATERINO M.S. ET AL. 2007: A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. — *Science* **318**: 1913–1916.

KATOH K. & STANDLEY D.M. 2013: MAFFT Multiple Sequence Alignment Software version 7: improvements in performance and usability. — *Mol. Biol. Evol.* **30**: 772–780.

KOLBE J.J., GLOR R.E., SCHETTINO L.R., LARA A.C., LARSON A. & LOSOS J.B. 2004: Genetic variation increases during biological invasion by a Cuban lizard. — *Nature* **246**: 177–181.

KLIBER A. & ECKERT C.G. 2005: Interaction between founder effect and selection during biological invasion in an aquatic plant. — *Evolution* **59**: 1900–1913.

LEE C.E. 2002: Evolutionary genetics of invasive species. — *Trends Ecol. Evol.* **17**: 386–391.

LEE C.E. & BELL M.A. 1999: Causes and consequences of recent freshwater invasions by saltwater animals. — *Trends Ecol. Evol.* **14**: 284–288.

LIBRADO P. & ROZAS J. 2009: DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. — *Bioinformatics* **25**: 1451–1452.

LINDROTH C.H. 1957: *The Faunal Connections between Europe and North America.* Wiley, New York, 356 pp.

Lockwood J.L., Cassey P. & Blackburn T. 2005: The role of propagule pressure in explaining species invasions. — *Trends Ecol. Evol.* **20**: 223–228.

McKenna D.D., Sequeira A.S., Marvaldi A.E. & Farrell B.D. 2009: Temporal lags and overlap in the diversification of weevils and flowering plants. — *Proc. Natl. Acad. Sci.* **106**: 7083–7088.

Misof B., Liu S., Meusemann K., Peters R.S., Donath A., Mayer C., Frandsen P.B., Ware J., Flouri T., Beutel R.G. et al. 2014: Phylogenomics resolves the timing and pattern of insect evolution. — *Science* **346**: 763–767.

Mitterboeck T.F. & Adamowicz S.J. 2013: Flight loss linked to faster molecular evolution in insects. — *Proc. R. Soc. (B)* **280**: 20131128, 8 pp.

Mitterboeck T.F., Chen A.Y., Zaheer O.A., Ma E.Y.T. & Adamowicz S.J. 2016: Do saline taxa evolve faster? Comparing relative rates of molecular evolution between freshwater and marine eukaryotes. — *Evolution* **70**: 1960–1978.

Moravcová L., Pysek P., Jarosik V. & Pergl J. 2015: Getting the right traits: reproductive and dispersal characteristics predict the invasiveness of herbaceous plant species. — *PLoS ONE* **10**(4): e0123634, 16 pp.

Pappert R.A., Hamrick J.L. & Donovan L.A. 2000: Genetic variation in *Pueraria lobate* (Fabaceae), an introduced, clonal, invasive plant of the southeastern United States. — *Am. J. Bot.* **87**: 1240–1245.

Prentis P.J., Wilson J.R.U., Dormontt E.E., Richardson D.M., & Lowe A.J. 2008: Adaptive evolution in invasive species. — *Trends Plant Sci.* **13**: 288–294.

R Core Team 2016: *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, URL: http://www.R-project.org/.

Ratnasingham S. & Hebert P.D.N. 2007: BOLD: The Barcode of Life Data System (www.barcodinglife.org). — *Mol. Ecol. Notes* **7**: 355–364.

Ratnasingham S. & Hebert P.D.N. 2013: A DNA-based registry for all animal species: the barcode index number (BIN) system. — *PLoS ONE* **8**(8): e66213, 16 pp.

Reed D.H. & Frankham R. 2001: How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. — *Evolution* **55**: 1095–1103.

Regier J.C., Zwick A., Cummings M.P., Kawahara A.Y., Cho S., Weller S., Roe A., Baixeras J., Brown J.W., Parr C. et al. 2009: Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. — *BMC Evol. Biol.* **9**: 280, 21 pp.

Richardson D.M., Pysek P., Rejmánek M., Barbour M.G., Panetta F.D. & West C.J. 2000: Naturalization and invasion of alien plants: concepts and definitions. — *Divers. Distrib.* **6**: 93–107.

Rilov G. & Crooks J.A. (eds) 2008: *Biological Invasions in Marine Ecosystems: Ecological, Management, and Geographic Perspectives.* Springer Science & Business Media, Berlin, 641 pp.

Rius M. & Darling J.A. 2014: How important is intraspecific genetic admixture to the success of colonising populations? — *Trends Ecol. Evol.* **29**: 233–242.

Robinson M., Gouy M., Gautier C., Mouchiroud D. & Bernard C. 1998: Sensitivity of the relative-rate test to taxonomic sampling. — *Mol. Biol. Evol.* **15**: 1091–1098.

Roman J. & Darling J.A. 2007: Paradox lost: genetic diversity and the success of aquatic invasions. — *Trends Ecol. Evol.* **22**: 454–464.

Sakai A.K., Allendorf F.W., Holt J.S., Lodge D.M., Molofsky J., With K.A., Baughman S., Cabin R.J., Cohen J.E., Ellstrand N.C. et al. 2001: The population biology of invasive species. — *Annu. Rev. Ecol. Syst.* **32**: 305–332.

Schrieber K. & Lachmuth S. 2016: The genetic paradox of invasions revisited: the potential role of inbreeding × environment interactions in invasion success. — *Biol. Rev. Camb. Philos. Soc.* **92**: 939–952.

Scott G.R., Schulte P.M., Egginton S., Scott A.L.M., Richards J.G. & Milsom W.K. 2011: Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. — *Mol. Biol. Evol.* **28**: 351–363.

Sherman C.D.H., Lotterhos K.E., Richardson M.F., Tepolt C.K., Rollins L.A., Palumbi S.R. & Miller A.D. 2016: What are we missing about marine invasions? Filling in the gaps with evolutionary genomics. — *Mar. Biol.* **163**: 1–24.

Sorte C.J., Williams S.L. & Carlton J.T. 2010: Marine range shifts and species introductions: comparative spread rates and community impacts. — *Global Ecol. Biogeogr.* **19**: 303–316.

Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S. 2013: MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. — *Mol. Biol. Evol.* **30**: 2725–2729.

Wang X.-W., Luan J.-B., Li J.-M., Su Y.-L., Xia J. & Liu S.-S. 2011: Transcriptome analysis and comparison reveal divergence between two invasive whitefly cryptic species. — *BMC Genomics* **12**: 458, 12 pp.

Welch J.J. & Waxman D. 2008: Calculating independent contrasts for the comparative study of substitution rates. — *J. Theor. Biol.* **251**: 667–678.

Wiegmann B.M., Trautwein M.D., Winkler I.S., Barr N.B., Kim J.-W., Lambkin C., Bertone M.A., Cassel B.K., Bayless K.M., Heimberg A.M. et al. 2011: Episodic radiations in the fly tree of life. — *Proc. Natl. Acad. Sci. U.S.A.* **108**: 5690–5695.

Wilson J.J. 2010: Assessing the value of DNA barcodes and other priority gene regions for molecular phylogenetics of Lepidoptera. — *PLoS ONE* **5**(5): e10525, 9 pp.

Wright S., Keeling J. & Gillman L. 2006: The road from Santa Rosalia: a faster tempo of evolution in tropical climates. — *Proc. Natl. Acad. Sci. U.S.A.* **103**: 7718–7722.

Yamanaka T., Morimoto N., Nishida G.M., Kiritani K., Moriya S. & Liebhold A.M. 2015: Comparison of insect invasions in North America, Japan and their Islands. — *Biol. Invas.* **17**: 3049–3061.

Yang Z. 2007: PAML 4: phylogenetic analysis by maximum likelihood. — *Mol. Biol. Evol.* **24**: 1586–1591.

Supplementary files:

S1 (http://www.eje.cz/2018/071/S01.xlsx): File that can be viewed in MS Excel and contains supplements 1–13 with information on invasive, sister, and outgroup species and/or BINs; literature search criteria; maximum likelihood trees; BOLD sequence identifiers for sequences used in analysis; relative substitution rates outputs; relative dN rates, dS rates, and dN/dS ratio outputs; estimated intraspecific diversities; and binomial and Wilcoxon signed-rank tests. Fasta files containing the sequences used in analysis are available upon request.

S2 (http://www.eje.cz/2018/071/S02.xlsx): Scripts 1–6. R scripts and Perl script used in analysis.