Discovery of cryptic species among North American pine-feeding *Chionaspis* scale insects (Hemiptera: Diaspididae)

RODGER A. GWIAZDOWSKI¹,²*, ISABELLE M. VEA³, JEREMY C. ANDERSEN¹,⁴ and BENJAMIN B. NORMARK¹,²

¹Graduate Program in Organismic and Evolutionary Biology, University of Massachusetts, Amherst, MA 01003, USA  
²Department of Plant, Soil and Insect Science, University of Massachusetts, Amherst, MA 01003, USA  
³Richard Gilder Graduate School, American Museum of Natural History, Division of Invertebrate Zoology, New York, NY 10024, USA  
⁴Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA

Received 19 February 2011; revised 7 April 2011; accepted for publication 7 April 2011

Cryptic species are present in many animal groups and they may be best detected through large sample sizes collected over broad geographic ranges. Fine-scale local adaptation has been hypothesized to occur in armoured scale insects (Hemiptera: Diaspididae) and a consequence of this process may be multiple cryptic species. We estimate species diversity of pine-feeding *Chionaspis* scale insects across North America by inferring species boundaries using genealogical concordance across allele genealogies of two nuclear loci and one mitochondrial locus. Our ingroup sample includes 366 individual insects from 320 localities and 51 host species within the Pinaceae. We also conducted a morphological survey of all insect specimens and assigned them as either one of the two currently recognized pine-feeding species, *Ch. heterophyllae*, *Ch. Pinifoliae*, or with undescribed morphology. Using maximum likelihood allele genealogies in a majority-rule consensus to assess congruence, we conservatively detect ten species in this group. Most of these species are robust to alternative methods of genealogical inference (Bayesian inference, maximum parsimony) and a few are robust to the use of strict consensus to assess congruence. Species show both narrow and more widespread ranges where almost half of the individuals sampled belong to a single very widespread polyphagous species. © 2011 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2011, 104, 47–62.


INTRODUCTION

There are many groups of organisms in which the number of species on earth is currently underestimated (Chapman, 2009; Trontelj & Fiser, 2009). This can occur if many species go unsampled or unrecognized. One group that may contain many unsampled as well as unrecognized species is the armoured scale insects (Hemiptera: Diaspididae). Because armoured scale insects are small and inconspicuous, they are overlooked by all but the most specialized collectors. Because most life stages of armoured scale insects are completely sessile on plant surfaces, they are generally unsampled by mass collecting techniques such as canopy fogging. As a result, they are undersampled in natural habitats in spite of their abundance in agricultural and ornamental settings. Because armoured scales have a simplified morphology with relatively
few usable characters, many species may go unrecognized even by expert identifiers. That is, there may be many cryptic species (Bickford et al., 2007). Because the dispersal stage typically moves only a few meters, scale insects may have strongly structured metapopulations, which may allow host plants and local conditions to exert diversifying selection (Edmunds & Alstad, 1978; Hanks & Denno, 1993). This may result in high rates of host race formation and possibly speciation. Newly originated species are likely to be morphologically similar. Thus, the true level of species diversity may be much greater than is currently recognized.

Here, we investigate the possibility of cryptic species diversity in pine-feeding Chionaspis Signoret (1868) scale insects in North America. Currently, two species are recognized, Chionaspis heterophyllae Cooley (1897) and Ch. pinifoliae Fitch (1856). Although many species of Chionaspis feed on a broad range of angiosperm species, Ch. heterophyllae and Ch. pinifoliae are the only members of the genus that are restricted to conifers (Liu et al., 1989), feeding primarily on Pinus spp. (Shour & Schuder, 1987; Liu et al., 1989; Watson, 2005; Miller & Ben Dov, 2006). Both species are serious economic pests (Miller & Davidson, 2005) and, while thorough sampling from agricultural and ornamental settings has revealed subtle inter- and intra-specific morphological diversity (Liu et al., 1989), native hosts have gone largely undersampled, suggesting the potential for unrecognized, cryptic species.

Here, we use a multi-locus method of species delimitation using several single locus genealogies in a genealogical concordance approach (Dettman, Jacobson & Taylor, 2003), in which we identify congruent internodes as evidence of species boundaries. Genealogical concordance is a pattern that is expected to be present among well-diverged, reproductively isolated lineages (Neigel & Avise, 1986; Harrison, 1998; Kuo & Avise, 2005) and we interpret our results as consistent with the biological species concept (BSC). We will use the results to address several questions: (1) Are multiple morphologically cryptic species present in this group?; (2) What is the geographic distribution of these species?; and (3) Are species associated with broad or narrow ranges of hosts?

MATERIAL AND METHODS

COLLECTION METHODS

Collection sites across the USA, Mexico and parts of Canada were selected according to areas of high host diversity or endemism, using range maps for North American trees in the family Pinaceae from the US Geological Survey (http://esp.cr.usgs.gov/data/atlas/little/index.html, USGS, 1999). Host species identifications were based on foliar and cone morphology (Farjon, de la Rosa & Styles, 1997; Farjon & Styles, 1997; Perry, Graham & Richardson, 1998; Petrides & Petrides, 1998; Petrides & Wehr, 1998; Richardson, 1998; Arévalo & González-Elizondo, 2003). Whenever possible, identifications were made during field collection by Pinaceae specialists (D. Gernandt, A. G. Arévalo, J. A. de la Rosa and L. M. González-Villarreal).

SAMPLING AND VOUCHERING OF SPECIMENS

Based on previous hypotheses that neighbouring trees can have locally adapted armoured scales (Edmunds & Alstad, 1978; Hanks & Denno, 1994), each locality represents a collection from an individual tree. Pine needles with scales from each locality were preserved in the field, with replicate lots stored in both 100% ethanol and also nitrogen vapour. Both scale and host tissue collected during 2006–2007 have been vouchered at the Ambrose Monell Cryo Collection (AMCC) of the American Museum of Natural History and the Insect Collection at the University of Massachusetts, Amherst (UMass, see also Supporting Information, Table S2). For each specimen from which DNA was extracted, the cuticle has been mounted on a labelled microscope slide. Slidemounted specimens and their accompanying DNA are maintained in the Insect Collection, UMass, Amherst. Outgroup species Duplachionaspis noeae (Hall), Chionaspis wisteriae Cooley, Ch. salicis (Linnaeus), Ch. ortholobis Comstock, Ch. americana Johnson, Ch. gleditsiae Sanders were chosen based on family level phylogenies of armoured scales (Andersen et al., 2010b).

DNA EXTRACTION, MORPHOLOGICAL SPECIMEN PREPARATION AND CHARACTER RECOGNITION

Genomic DNA was extracted using the DNEasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA). Whole individual insects were pierced with a size 000 insect pin and immersed in buffer during the initial digestion step. After DNA extraction, the cuticle of each individual was slide mounted using the following protocol: 30 min clearing bath in 10% potassium hydroxide (KOH) heated at 60 ºC, 10 min in distilled water, 10 min in double stain (distilled water added), 10 min in 70% ethanol (EtOH), 10 min in 100% EtOH, 10 min or more in clove oil. After the clove oil bath, specimens were slide mounted in balsam and stored at 45–50 ºC for 1 month. Each mounted specimen was examined using an Olympus CH compound microscope. Specimens were initially examined
**Table 1.** Morphological characters used to characterize an individual scale insect as having a *Ch. heterophyllae*-like or *Ch. pinifoliae*-like morphology. Specimens with previously undescribed morphology are indicated in Figures 2–4

<table>
<thead>
<tr>
<th>Species</th>
<th>Median lobes (L1)</th>
<th>Third lobes (L3)</th>
<th>Gland spines on three last segments</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Chionaspis</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>heterophyllae</em></td>
<td>– slightly protruding yoke</td>
<td>– lobes recessed and less visible</td>
<td>Longer than median lobes</td>
</tr>
<tr>
<td></td>
<td>– medial margin strongly divergent</td>
<td>– presence of notches, especially marked on lateral lobule</td>
<td></td>
</tr>
<tr>
<td></td>
<td>– lateral notches present</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chionaspis</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>pinifoliae</em></td>
<td>– strongly protruding yoke</td>
<td>– lobes protruding</td>
<td>Same length to shorter than median lobes</td>
</tr>
<tr>
<td></td>
<td>– medial margin parallel or convergent</td>
<td>– generally round without notch</td>
<td></td>
</tr>
<tr>
<td></td>
<td>– lateral notches absent</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Independently of the phylogenetic data and identified as having morphology consistent with *Chionaspis pinifoliae*, *Chionaspis heterophyllae* or an undescribed morphology. Identifications were based on complete descriptions of both species from Liu et al. (1989) and Miller & Davidson (2005). The main characters distinguishing these species are located on the pygidium, particularly on the pygidial lobes. The combination of characters used to identify species is presented in Table 1.

**Locus amplification, sequencing and alignment**

We amplified three gene regions: the D2–D3 portion of the large ribosomal subunit rDNA (28S), elongation factor 1-alpha (EF-1α) and a mitochondrial fragment spanning parts of cytochrome c oxidase subunits I and II (COI–COII). Primers for 28S and EF-1α were the same as those of Morse & Normark (2006) and primers for COI–COII were those of Provencher et al. (2005). All PCR reactions were performed using Takara Ex Taq PCR conditions for 28S and EF-1α followed those of Morse & Normark (2006); PCR conditions for COI–COII were similar to Morse & Normark (2006), but modified to have 40 cycles of a denaturation step of 95 °C for 30 seconds, an annealing step of 47 °C for 1 minute and an extension step of 72 °C for 2 minutes.

PCR products were visualized using 1.5% agarose gels in 0.5 X TBE and stained with SYBRsafe (Invitrogen, Carlsbad, CA, USA). PCR products were cleaned using Exo SAP-IT enzymatic digestion (USB Corporation, Cleveland, OH, USA). Sequencing of PCR products in both directions was performed on a 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA) at the Yale University DNA Analysis Facility (New Haven, CT, USA). Sequences were edited using Sequencher 4.2 (Gene Codes Corporation, Ann Arbor, MI, USA). Heterozygous sites present in the nuclear loci were recorded using the IUPAC ambiguity code. Alignment of 28S and EF-1α sequences was unambiguous. COI–COII contains an intergenic region that was difficult to align and was excluded from the analysis. Lastly, wherever multiple individuals from the same locality were identical at all three loci, only one was included in the analysis.

**Allelic genealogies**

All analyses were run using the CIPRES computer resources (Miller et al., 2008) of the San Diego Supercomputer Center (La Jolla, CA, USA). The matrix for each locus was analysed using maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) methods. Maximum likelihood genealogies were used for the data analysis. Parsimony and Bayesian genealogies were used for comparison with the likelihood results.

Maximum likelihood genealogies were inferred using RAxML v7.0.4 (Stamatakis, 2006; Stamatakis, Hoover & Rougemont, 2008). The data were partitioned by codon position in the COI–COII data and by codon position and intron in EF-1α. The GTR+Γ model of nucleotide substitution was used for all loci. All model parameters are estimated by RAxML, where individual α-shape parameters, generalized time reversible (GTR) rates and empirical base frequencies are estimated and optimized for each partition (Stamatakis, 2006; 2008). We conducted 200 replicate maximum likelihood analyses on each of the three loci; each replicate was initiated using a maximum parsimony starting tree generated by RAxML. The genealogy selected for each locus was the tree with the highest likelihood score among the 200 trees. The three genealogies were combined in a majority rule consensus tree using PAUP 4.0b10 (Swofford, 2003) and this consensus tree was used in the data analysis.

Maximum parsimony genealogies and consensus trees were generated using PAUP 4.0b10 (Swofford, 2003). A heuristic search algorithm was used, with 100 random–addition–sequence starting trees and tree–bisection–reconnection branch swapping, maxtrees = 10 000. All maximum parsimony trees for each
locus were combined in a strict consensus tree for each locus. The three resulting strict consensus trees were combined as a majority rule consensus tree and also a strict consensus tree.

Bayesian inference genealogies were generated using MrBayes V. 3.1.2 64bit (Huelsenbeck & Ronquist, 2001). The data were partitioned by codon position in the COI–COII data and by position, codon and intron in EF-1α. The GTR + I + G model was chosen for all partitions as it was either the same or a more complex model than any found using MrModeltest v2.3 (Nylander, 2008) and is comparable with the model used with the maximum likelihood analyses. Model parameters were uncorrelated and rates were allowed to vary under flat priors across partitions. Three independent replicates were performed. Each replicate run included four simultaneous and incrementally heated Markov chains using default heating values. Markov chains were initiated from a random tree and were run for 10 000 000 generations; samples were taken every 1000th generation. Posterior probability log-likelihood values were plotted against generation number and observed visually in Tracer v1.4.1 (Rambaut & Drummond, 2007). All runs were observed to reach stationarity and the number of generations required for ‘burn-in’ was assessed visually in Tracer. All generations sampled prior to burn-in were discarded and the remaining samples, pooled across replicate runs for any one locus, were summed to determine the posterior probability distribution and consensus tree for each locus. The three resulting consensus trees were combined to produce a majority rule consensus tree and also as a strict consensus tree.

Species delimitation
Under the biological species concept, separate species do not exchange genes. Therefore, if biological species are separated for a sufficient length of time following the cessation of gene exchange (speciation), one or both of the daughter species is expected to become monophyletic at most loci (Neigel & Avise, 1986; Hudson & Coyne, 2002). We use a genealogical concordance approach to detecting boundaries between biological species: each internode in the majority-rule consensus tree represents a hypothetical species boundary. This is a relatively conservative species delimitation criterion, because it is biased towards detection of divergence following speciation events (where monophyly has evolved at multiple loci in at least one of the daughter species). Nonetheless, this criterion may still yield artificial hypotheses of species boundaries (a form of type I error) if sampling is poor. Therefore, in applying this species delimitation criterion, we conservatively decline to recognize groups of three or fewer individuals, unless a group of three is supported by all loci. We treat groups of two or three as ‘singletons’ (consisting of one or more individuals from a single locality) and we include them in the species to which they are most closely related, according to the majority-rule consensus tree (Fig. 1).

Inter-specific mitochondrial diversity
Our consensus method of concordance does not take into account a magnitude of sequence divergence. For an overview of mitochondrial interspecific divergence, pairwise uncorrected $P$-values between exemplar specimens from each outgroup species and new species identified in the analysis were calculated using PAUP* 4.0b10 (Swofford, 2003). Additionally, we calculate intraspecific divergence, as above, within one of the newly identified species.

Results
Specimens sampled
Of the approximately 72 native North America Pinus species (sensu Gernandt et al., 2005), 61 species, predominately from natural locations, were sampled for scale insects. Scale insects in the genus Chionaspis were found on 48 of the 61 Pinus species (see also Supporting Information, Table S1). Samples from all 48 of these hosts appear in the analysis, including scales from five non-native Pinus species planted as ornamentals, as well as two species from Picea and one from Pseudotsuga. In total, samples from 56 Pinaceae species included in the analysis comprised 366 individual scale insects, from 320 individual trees.
collected across North America. In addition, eight scale specimens representing outgroup taxa were included for a total of 374 individuals. Collection information for all specimens and their hosts is available in the Supporting Information (Table S2).

**DNA SEQUENCING**

All sequences have been deposited in GenBank and accession numbers per locus per individual are provided in the Supporting Information (Table S2) (28S GU349091–GU349464, EF-1α GU349839–GU350211 and COI–COII GU349465–GU349838). The sequence alignments, genealogies and consensus trees have been deposited in TreeBASE, accessible at http://purl.org/phylo/treebase/phylows/study/TB2:ST0639. The alignment length, proportion of variable sites and number of parsimony informative (PI) sites per locus are reported in Table 2.

**SPEICES DELIMITED VIA GENEALOGICAL CONCORDANCE**

Applying our delimitation criteria to the maximum likelihood consensus tree (Figs 2–4), we recognize ten species of pine-feeding *Chionaspis* in North America. Species are informally designated in Figures 2–4 by a capital S and a number from 1 to 10 (e.g. S9). The numbers of specimens within species range from a maximum of 201 in S10 to a minimum of 3 in S9. Not all species are identified by the same concordances across loci. Strict concordance of all three loci identifies five species: S1, S3, S4 and S9. EF-1α and COI–COII together support four species, S2, S6, S8 and S10. Ribosomal DNA 28S and mtDNA support one species, S7.

Mitochondrial uncorrected pairwise per cent divergences between outgroup specimens and exemplar specimens from species identified in this analysis are presented in Figure 5. The maximum divergence between species identified in our analysis (Fig. 5A) is 10.5% between S1 and S2; S1 is recovered only from the south-eastern USA, whereas S2 is widely distributed across Mexico (Fig. 6). The minimum divergence is 1.4% between S7 and S10. S7 and S10 overlap in range, although S7 was recovered only from eastern North America. Most species pairs are > 7% divergent from each other. Within species S2 (Fig. 5B), the maximum divergence between individuals is 10.5%. The 27 individuals within S2 show a range of divergence from 10.5 to 0%, although most individuals are in a range showing over > 4% to > 6% divergence.

**MAXIMUM LIKELIHOOD COMPARED WITH MAXIMUM PARSIMONY AND BAYESIAN INFERENCE ANALYSES**

Maximum parsimony and Bayesian inference do not find additional species previously identified using maximum likelihood. Species identified by maximum parsimony and Bayesian inference are indicated in Figures 2–4, and all individuals incongruent as to species assignment between maximum likelihood and maximum parsimony or Bayesian inference are lumped into S10. S10 and S2 are both identified by maximum likelihood as widespread species. Maximum parsimony identifies eight of the ten species, lumping S2 and S5 into S10. Bayesian inference identifies nine of the ten species, lumping S5 into S10. The parsimony result represents the most conservative number of species (eight) across all three phylogenetic methods.

According to a strict consensus (Baum & Shaw, 1995), concordant nodes in the maximum likelihood consensus tree identifies six species: S1, S3, S4, S9, a group directly corresponding to morphological group A, and all other specimens would comprise a sixth species. A strict consensus of maximum parsimony genealogies identifies four species: S3, S4, morphological group A and a fourth species comprising all other samples. Curiously, a strict maximum parsimony consensus tree does not identify S1 as it is not present in the 28S strict consensus genealogy, presumably because of the low number of parsimony informative sites at this locus. A strict consensus of Bayesian inference genealogies identifies five species: S1, S3, S4, morphological group A and a fifth species comprising all other samples.

**PREVIOUSLY UNDESCRIPTED MORPHOLOGICAL CHARACTERS**

Morphological evaluation of each specimen (Table 1) reveals morphological groups possessing a previously undescribed morphology unique to that group indicated by the letters A–E in Figures 2–4. Morphological group C directly corresponds to species 4. Morphological groups A and B are nested within species 2; morphological group D is nested within

---

**Table 2.** Trimmed length of most sequences in alignment, percentage of variable sites and percentage of parsimony-informative (PI) sites for loci included in the phylogenetic analysis. Variable and PI sites include all EF-1α introns but exclude the intergenic region of COI–COII.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Alignment length (bp)</th>
<th>Variable sites (%)</th>
<th>PI sites (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>28S</td>
<td>685</td>
<td>7.4</td>
<td>5.1</td>
</tr>
<tr>
<td>EF-1α</td>
<td>1064</td>
<td>25.7</td>
<td>12.4</td>
</tr>
<tr>
<td>COI–COII</td>
<td>789</td>
<td>51.9</td>
<td>43.0</td>
</tr>
</tbody>
</table>

Figure 2. Depicted is one of three figures showing sections of the maximum likelihood consensus cladogram. On the left is a reduced version of the full tree indicating the sections presented in each figure. Branches are displayed according to the combination of loci supporting a given node, as indicted in the key. Taxon names include the accession number (see also Supporting Information, Table S2), along with the host genus and species. The first longitudinal bars on the right identify species delimited in this analysis and are labelled by a capital S and their number. The second bar presents species designated by morphology alone (Morph.) where letters A–E designate groups possessing previously undescribed morphology. Thin black bars represent species identified by maximum parsimony (MP) and Bayesian inference (BI). Open areas with species numbers represent species not identified by that analysis and refer to their placement within species in the maximum likelihood analysis. The distribution of S1 is geographically depicted in part A of Figure 6. Figures 3 and 4 begin with collapsed nodes representing cladogram structure displayed in the previous figures.

© 2011 The Linnean Society of London, Biological Journal of the Linnean Society, 2011, 104, 47–62
species 5; and morphological group E is nested within species 6. Although our analysis conservatively lumps these morphological groups into species, in the context of armoured scale morphology these groups are sufficiently different to merit species status on their own. As a result of this study, five new species are being described based on new morphological characters for the genus Chionaspis (I. M. Vea & R. A. Gwiazdowski, unpubl. data). We do not describe the unique morphology here, however some of the diagnostic characters include a characteristic reduction of the head (group D), more than one gland spine next to the pygidial lobes (B, C and D) or a peculiar median lobe shape (group A).

Figure 3. This section of the consensus cladogram focuses on a clade of species found mostly in Mexico where a hierarchical order of topology occurs across the genealogies, most species are geographically depicted in part C of Figure 6. BI, Bayesian inference; Morph., morphology alone; MP, maximum parsimony.
This portion of the unresolved polytomy contains a further 175 individuals. All are part of 1 species, here labeled S10.
Figure 4. This final portion of the consensus cladogram focuses a polytomy of four geographically widespread species and one with a restricted range. Here, S10 comprises a very large portion of the polytomy and has been truncated. The full extent of S10 specimens present, relative to the rest of the tree, is presented in a reduced cladogram illustrated in Figure 2. Generalized ranges of these species are depicted in parts A and C of Figure 6. All specimens present in S10 are listed in the Supporting Information (Table S2). BI, Bayesian inference; Morph., morphology alone; MP, maximum parsimony.

**Geographic Distribution and Host Affiliation of Species**

Among the ten species, our analysis suggests some occupy vast geographic ranges, while others are more narrowly distributed. Generalized species distributions, encompassing sampled points, are graphically illustrated in Figure 6. Broadly distributed species are found in both the USA and Mexico. Species S10 is found over much of the continent and co-occurs with all species except for S3 and S4. S1 is recovered along the north-east coast and throughout the southeastern USA, where the only other species occurring with it is S7. S8 is recovered across most of the western USA. It is nested within the range of S10 and overlaps with S6, which is recovered from California and Baja California in Mexico. The only narrowly distributed species is S9, occurring only in the sky islands region of southern Arizona.

Species showing both widespread and narrowly distributed geographic ranges are also recovered from both wide and narrow ranges of hosts. Only four scale species are found on more than three host species. Host affiliation by species and host taxonomy are provided in the Supporting Information (Table S1). Most scale species (seven of ten) are widespread (Fig. 6) and feed on at least two host species from two or more subsections in the genus *Pinus* (see also Supporting Information, Table S1). S1 is recovered only from the subgenus *Pinus*, particularly the subsections *Australes* and *Contortae*, which predominate in this species’s range, along with subsections *Pinus* and *Pinaster*, which here represent non-native ornamental plantings within the range of S1. S1 and S10 share very few hosts, including only one native (*P. rigida*) and two ornamental hosts (*P. nigra*, *P. sylvestris*), even although S10 is recovered from the genus *Picea* as well as every subsection of the genus *Pinus*. S10, S8 and S6 share four host species: *P. contorta*, *P. attenuata*, *P. ponderosa* and *P. lambertiana*. These species of pines span both subgenera of *Pinus* and are the predominant trees where the scale insect species ranges intersect. Notably, two species suggest evidence for host specialization where they occupy widespread as well as narrow distributions. S15 occurs over much of eastern North America and feeds predominantly on *P. strobus* (except for the host use of two specimens). Whereas S9 occurs only on mountain ranges in southern Arizona and feeds exclusively on *P. discolor*.

**Discussion**

Here we have used genealogical concordance across allele genealogies to delimit species. We find multiple cryptic species present under the taxonomic designations *Chionaspis heterophyllae* and *Chionaspis pini-foliae*. Using maximum likelihood allele genealogies and a majority-rule consensus to assess genealogical concordance, we detect ten species. Most of these species are robust to the method of inference of allele genealogy (maximum parsimony vs. Bayesian inference vs. maximum likelihood). Even under a strict criterion of complete congruence at all loci, we find multiple cryptic species (4, 5 or 6), depending on method of inference of allele genealogy). Newly detected species contain groups united by undescribed morphological characters as well as host affiliation and single locus markers. This pattern suggests that genealogical concordance may have conservatively grouped species in which it may be possible, using more samples or finer-scale approaches, to uncover further species boundaries.

**Genealogical Congruence as a Species Delimitation Criterion**

Genealogical concordance delimits species based on congruence across multiple genealogies of unlinked loci. In the (majority-rule) consensus tree of allele genealogies, a congruent internode is interpreted as a species boundary. Variations on this method of species delimitation are now widely used (Starrett & Hedin, 2007; Groeneveld et al., 2009). In our view, the best rationale for the method is that it serves as a relatively conservative estimator of long-standing barriers to gene exchange. It has this property because congruent internodes are expected to occur only if a majority of allele genealogies have attained monophyly in at least one daughter lineage following a speciation event, which is not expected to occur until at least 3.80 N generations following the cessation of gene exchange (Hudson & Coyne, 2002).

One weakness of the genealogical congruence method is that it is unclear what the appropriate
Figure 5. A, mitochondrial pairwise divergences calculated from uncorrected P-distances using outgroup specimens and exemplar samples from species identified using genealogical concordance. B, divergences as in (A) calculated from outgroup specimens and all individuals comprising species 2 (S2, Figure 3). Samples are labelled with the species designation in bold and their accession number. Cells are coloured from light to dark according to a scale of increasing divergence in increments of 0–1%, 1–3%, 3–6%, 6–10% and > 10%. The dark vertical line represents the boundary between divergences of outgroup and ingroup specimens on the left and solely ingroup specimens on the right.
Figure 6. Species distributions of widespread species (A), as well as more narrowly distributed western species in California and Baja California (B), and the south-western USA and Mexico (C). Collection points for species with extensive ranges are depicted with different shapes, shown in the key. A given species distribution is indicated by the species designation number directly adjacent to it. Shaded areas represent the range of Pinus spp. in North America based on USGS distributions (USGS, 1999).
majority or supermajority threshold should be for recognition of congruence. Strict (100%) congruence may be too stringent as some loci in undoubted species (e.g. chimpanzees and humans) would not meet it (Hudson & Coyne, 2002) and many authors have advocated for some lower threshold (Geiser, Pitt & Taylor, 1998; Dettman et al., 2003; Balakrishnan, 2005; Baum, 2009). Another weakness of the genealogical concordance method is that it may be subject to false positives if sample sizes are too low; for instance, if only the extremes of variation within a species are sampled.

Our goal is to be conservative in postulating the existence of new species – to tolerate false negatives (erroneous lumping) in order to avoid false positives (erroneous splitting). In using only a 2/3 majority-rule consensus, we may have courted the possibility of false positives. We have sought to minimize this possibility by (1) thorough sampling, (2) use of relatively conservative loci [both of which are more likely to yield false negatives (Beheregaray & Caccone, 2007)] and (3) lumping of three or fewer individuals, unless a group of three is supported by all loci.

There are several very likely cases of false negatives where we may have conservatively failed to detect species boundaries, at least within species S2, S5 and S6. Current taxonomy of armoured scales is based on a morphological species concept (Miller & Davidson, 2005) and some members of these species possess a previously undescribed morphology (morphology A, B, C, D and E, Fig. 3). For example, individuals with morphology E all share a host (Pinus torreyanae) from which no other Chionaspis specimens have been collected and comprise a well-supported mitochondrial clade with mainland (San Diego, CA, USA) and island (Santa Rosa, CA, USA) haplotypes as inferred by all three phylogenetic methods (see also Supporting Information, Table S1). Two of the other morphological groups (A and B, Fig. 3) are also only recovered from a single host and all morphological groups have distinct mitochondrial haplotypes. Additionally, we have identified species where most individuals are morphologically similar, yet show large intraspecific mitochondrial divergence. For example, most pairs of Chionaspis species designated here appear to be 6–10% divergent at COI–COII (Fig. 5A). Species 2 contains a mix of morphologically similar specimens and two morphologically unique groups (groups A and B, Fig. 3). Individuals within species 2 range from 4–10% divergent from each other at COI–COII (Fig. 5B). This difference is consistent with intraspecific divergence within Chionaspis (Fig. 5A), and is similar to congeneric mitochondrial divergence seen in other insects as an average of 5.8% (Pons et al., 2006) (as averaged across several loci) and a range 0.0–7.95% (Hebert et al., 2004). These morphological groups, and species with high levels of intraspecific divergence, appear to be cases where genealogical concordance has conservatively failed to reject the null hypothesis of only one species present, suggesting that increased sampling of this species would reveal even more cryptic species diversity. In these cases where it is suspected that species identified in this way may not be exclusive (Velasco, 2009), then these species can be used as starting positions to investigate species boundaries at lower levels (Davis & Nixon, 1992; Monaghan et al., 2009) where finer scale measures such as gene flow (Porter, 1990), population genetic markers and morphology can be used to tease out signatures of reproductive isolation.

**Geographic distribution of Ch. heterophyllae and Ch. pinifoliae**

S1 predominates in the eastern USA, overlapping with S7 and S10 only in the mid-Atlantic and northeastern USA (Fig. 6). This distribution is consistent with previous observations describing the distribution of Ch. heterophyllae (Ferris, 1937; Ferris, 1942; Shour & Schuder, 1987), where this species is predicted to occur in all states east of the Mississippi river, including Missouri, Louisiana and Texas. Although the morphological survey reveals Ch. heterophyllae-like morphology present in species from Mexico (Fig. 3), no Mexican species sharing this morphology co-occur in the range of S1 (Fig. 6A). Recognizing the distribution of Ch. pinifoliae is less straightforward because of the numerous species that share this morphology. However, the geographic distribution and host breadth reported in the literature for Ch. pinifoliae (Shour, 1986; Miller & Davidson, 2005; Watson, 2005) is consistent with those of species S10.

**Cryptic species and their implications**

The ability to recognize cryptic species has particular implications for future scale insect studies and general implications for species discovery. The presence of cryptic species of armoured scales has been hypothesized based on their intimate relationship with their host plants and the considerable intraspecific molecular divergence observed in some species (Provencher et al., 2005). The sessile armoured scale life history may allow host plants and local conditions to exert diversifying selection within and between populations. This process has been predicted to lead to extreme local adaptation, also called adaptive deme formation (ADF) (Edmunds, 1973; Edmunds & Alstd, 1978; Mopper, 1996). First formulated based on work with armoured scales (Edmunds & Alstd, 1978), the ADF hypothesis proposes that a population
may adapt to an individual host resulting in ‘a series of semi-isolated subpopulations, or demes’ (Edmunds & Alstad, 1978). These demes should show decreased performance when transferred from natal to a non-natal host (Van Zandt & Mopper, 1998) and the main experimental tool used to detect ADF is reciprocal transfer of insects between hosts. Support for ADF has had mixed and somewhat contentious results (Van Zandt & Mopper, 1998; Ruhnke et al., 2006). Although microevolutionary genetic variation is implicit in the ADF hypothesis, very few scale insect studies have taken gene flow into account (Alstad, Hotchkiss & Corbin, 1991; Hanks & Denno, 1998). Although the species in this analysis await formal taxonomic description, their current presence could confound the assumption of a homogenous population on any one host, suggesting future tests for ADF should account for the presence of cryptic diversity.

The same microevolutionary processes involved in ADF could, over longer timescales, lead to multiple closely related cryptic species, and this may be fostering the existence of cryptic species in other groups of armoured scales (Andersen et al., 2010a) and other scale insects (Cook, 2000; 2001; Gwiazdowski et al., 2006; Cook & Rowell, 2007). Beyond exploring evolutionary processes, detecting cryptic species has important practical implications where pest management tools such as biological control (Ngeve, 2003) rely on the ability to recognize and control invasive armoured scales. Current morphological methods which identify armoured scales may fail to recognize cryptic species without the use of molecular markers (Evans, Watson & Miller, 2009; Rugman-Jones, Morse & Stouthamer, 2009).

Beyond scale insect studies, any species discovery efforts will have to contend with identifying cryptic species (Bickford et al., 2007; Schlick-Steiner et al., 2007). Cryptic species occur across the Metazoa (Pfenninger & Schwenk, 2007; also see Trontelj & Fishe, 2009), even among the presumably obvious megafauna (Brown et al., 2007; Oliveira et al., 2008; Murata et al., 2009), and any inability to recognize them will not only confound conventional biodiversity estimates (Bebber et al., 2007) but also undermine effective communication about biological diversity (Vrijenhoek, 2009).

ACKNOWLEDGEMENTS

For valuable comments on this manuscript we thank S. Gillespie, H. Graves, J. Graves, P. Z. Goldstein and A. Okusu, as well as helpful comments from two anonymous reviewers. Financial support for this work was provided by NSF (DEB-0118718, DEB-0447880), USDA (2002-35302-12479) and the University of Massachusetts David J. Klingener Endowment Scholarship. R.A.G. was supported in part by a University of Massachusetts Pearson Entomology Fellowship.


For logistical support we thank the Applied Entomology Group of Martin R. Aluja and Juan Rull of the Instituto de Ecología, David Gernandt, and the Gernandt Laboratory at Universidad Nacional Autónoma de Mexico, Abell García Arévalo of the Ecology and Natural Resource Unit, Instituto de Ecología, Jorge A. Perez de la Rosa, Luz-Marie Gonzalez-Villareal and the Department of Biology and Zoology of the Universidad de Guadalajara, as well as The Penny Gullen and Peter Cranston Laboratory of the University of California, Davis, Dug Miller of the USDA Systematics Entomology Laboratory, Rutgers University Marine Field Station, Michael and Matilda Kosztarab, the Graves family and the Gwiazdowski family. Lindsey Cushing, Amanda Krewinski and Kelsey Robare skilfully mounted many of the slide vouchers. This paper comprises a part of the PhD dissertation of R. A. G. at the University of Massachusetts, Amherst.

REFERENCES


© 2011 The Linnean Society of London, Biological Journal of the Linnean Society, 2011, 104, 47–62


Nylander JAA. 2008. MrModeltest v2.3. *Evolutionary Biology Centre, Uppsala University*. Program distributed by the author.


Starrett J, Hedin M. 2007. Multilocus genealogies reveal multiple cryptic species and biogeographical complexity in...


**Velasco J. 2009.** When monophyly is not enough: exclusivity as the key to defining a phylogenetic species concept. *Biology & Philosophy* 24: 473–486.


**Watson GW. 2005.** Diaspididae of the world. *Arthropods of Economic Importance*: ETI Bioinformatics: World Biodiversity Database.

**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article:

**Table S1.** Geographic locations and species designations according to host taxonomy, number of hosts and number of specimens sampled in this study.

**Table S2.** Accession numbers and collection localities for all specimens and sequences used in this study.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.