Recent and rapid diversification of the small carpenter bees in eastern North America

WYATT A. SHELL and SANDRA M. REHAN*

Department of Biological Sciences, University of New Hampshire, 46 College Road, Durham, NH, 03824, USA

Received 16 June 2015; revised 22 August 2015; accepted for publication 23 August 2015

Bees fulfil a critical ecological role as pollinators, significantly contributing to the reproductive success of myriad angiosperm species. Although increasingly appreciated for their agricultural contributions, relatively little is yet known about the natural history of the vast majority of the more than 20 000 wild bee species worldwide. The small carpenter bee genus *Ceratina* occurs globally, and is represented in North America by its most recently diverged subgenera, *Ceratinula* and *Zadontomerus*. Recent genetic analysis of eastern *Ceratina* (*Zadontomerus*) supports the existence of five closely related, yet genetically distinct species living in sympatry. This phylogeographical study employs molecular barcoding of the most comprehensive specimen collection yet assembled to confirm the identities of these recently diverged eastern North American *Ceratina* (*Zadontomerus*) species. Delineation of extant population structure, evolutionary history and known range of this emerging model native pollinator are greatly improved by this study. We consider ecological and behavioural factors potentially contributing to the maintenance of genetic identity among these sympatric species.


INTRODUCTION

The study of speciation is a powerful approach to exploring evolutionary, climatic and ecological change (Hampe & Petit, 2005). By understanding the elements contributing to the vicariance, dispersal and displacement of ecologically essential species groups, we gain important insight into current and historical phylogeography (Davies et al., 2013; López-Uribe et al., 2014; Rehan & Schwarz, 2015). Pollinator species are of significant ecological importance, and are highly sensitive to changes in their environment; they are thus among the most informative taxa for gauging biological responses to ecological dynamics (Potts et al., 2010; Winfree et al., 2011).

Bees have coevolved with angiosperms for over 100 Myr and are represented by more than 20 000 species worldwide (Engel, 2000; Michener, 2007). Bees are thus regarded as important pollinators both for their efficiency and for their ubiquity (Kremen, Williams & Thorp, 2002; Winfree et al., 2011; Britain et al., 2012; Rogers, Tarpy & Burrack, 2014; Ascher & Pickering, 2015). Directly tied to trophic productivity, bees are vitally important to sustainable commercial agriculture and insect-pollinated ecosystems (Klein et al., 2007). The value of bee pollination to agriculture is estimated at $200 billion per year worldwide, and is thought to be even greater in natural ecosystems (Gallai et al., 2009). As bees are strongly influenced by climatic and environmental conditions, understanding their phylogenetic relationships and current and historical species ranges can reveal a great deal about climate change, pollinator dispersal patterns and the ongoing evolution of pollinator–plant relationships (Conte & Navajas, 2008; Bartomeus et al., 2011; Russo et al., 2013; López-Uribe et al., 2014; Rehan & Schwarz, 2015).

The small carpenter bees, genus *Ceratina* (Hymenoptera: Apidae: Xylocopinae), are a globally dispersed and highly diverse group of ubiquitous pollinators, occurring on every continent except Antarctica (Rehan, Richards & Schwarz, 2010b; Rehan et al., 2010a; Kennedy et al., 2013). Originating...
in Africa ~55 Mya, Ceratina dispersed north and eastward to cover the entirety of Eurasia before arriving in the New World some 43 Mya (Rehan & Schwarz, 2015). Given a cooler climate, these initial individuals probably took to refugia in the Neotropics before dispersing north into the Nearctic as glacial periods came to a close (Rehan & Schwarz, 2015). The extant Nearctic lineages are thought to have originated in the Neotropics, rapidly dispersing north-west as far as Oregon and British Columbia, and north-east into the Great Lakes region and Maritime Canada (Daly, 1973; Sheffield et al., 2009; Rehan & Sheffield, 2011). Ceratina has diversified into at least 26 known species in North America, and is currently represented by subgenera Ceratinula and the most recently diverged Zadontomerus (Daly, 1973; Rehan & Schwarz, 2015).

Although 21 of the 26 described Ceratina (Zadontomerus) species occur west of the Rocky Mountain range, the few known eastern North American species are taxonomically challenging to differentiate from one another (Daly, 1973; Rehan & Richards, 2008; Rehan & Sheffield, 2011). Because of this lack of diagnostic morphology, there were originally thought to be just three eastern species: Ceratina calcarata, C. strenua and C. dupla (Mitchell, 1962; Daly, 1973). It was exactly this type of taxonomic challenge that inspired the development of genetic ‘barcoding’, in which rapidly mutating mitochondrial DNA, specifically the cytochrome c oxidase subunit I (COI) locus, is targeted to assess population structure with high resolution (Hebert, Ratnasingham & deWaard, 2003b; Hebert et al., 2003a, 2004; Beebee & Rowe, 2004). This method of DNA barcoding has been used with great success in both bee and wasp phylogeographical studies (Packer et al., 2008; Raychoudhury et al., 2010; Groom, Stevens & Schwarz, 2015).

Recent genetic analyses of eastern North American Ceratina support the existence of two additional species, C. mikmaqi and C. floridana, both genetically distinct from, but still closely related to, C. dupla (the triad are considered a species-complex; Rehan & Sheffield, 2011). Interestingly, while C. calcarata, C. strenua, C. mikmaqi and C. dupla have broadly overlapping ranges, C. floridana has been primarily identified in the south-east (Rehan & Sheffield, 2011). Taken together, the five species are semi-sympatric east of the Mississippi River and known ecology and nesting biology are similar (Grothaus, 1962; Daly, 1973; Kislow, 1976).

DNA barcoding has aided the taxonomic revision of the eastern Ceratina species-complex and revealed cryptic species and population structure, but the exact phylogenetic relationships within the eastern species-complex remain unknown (Rehan & Richards, 2008; Rehan & Sheffield, 2011). Here we provide the first phylogeography of this native pollinator genus in North America, adding new records expanding the known distributions of all eastern Ceratina species, and provide population genetic resolution of recently described species C. mikmaqi and C. floridana. Dates of species divergence are estimated and used to reconstruct the historical biogeography of the eastern North American Ceratina (Zadontomerus) clade. Factors influencing and maintaining distinct species despite high ecological overlap are considered.

METHODS

SAMPLE SELECTION AND IDENTIFICATION

While previous morphological taxonomic study has established expansive regional ranges for these species, recent taxonomic revisions and molecular genetic characterizations have focused on north-eastern populations around the Great Lakes and east to maritime Canada (Daly, 1973; Rehan & Richards, 2008; Rehan & Sheffield, 2011). To compare genetic divergences across their continental range, specimens of C. calcarata, C. strenua, C. mikmaqi, C. dupla and C. floridana were barcoded from all states for which samples were available: east of the Mississippi as far south as Florida and west to Nebraska (Figs 1, 2). In total, 172 specimens were morphologically identified to species following Rehan & Sheffield (2011) and voucher specimens are retained in the University of New Hampshire Insect Collections.

Figure 1. Ceratina calcarata (grey squares) and C. strenua (white squares) collection locations. Black frames indicate north-east (NE), north-central (NC), mid-Atlantic (MA) and south-east (SE) ecoregions.
Figure 2. *Ceratina mikmaqi* (grey circles), *C. dupla* (white circles) and *C. floridana* (grey triangles) collection locations. Black frames north-east (NE), north-central (NC), mid-Atlantic (MA) and south-east (SE) ecoregions.

**DNA extraction, amplification and sequencing**

Three legs from the right side of each specimen were used for DNA extraction. DNA was extracted following a modified phenol chloroform-isoamylalcohol procedure (Kirby, 1956). Genetic material was amplified at barcoding region COI, as in Hebert *et al.* (2003a, b), using the Lep1 (F + R) primer pair (Lep1F, 5′-ATCCAACCA ATCAAAGATATTG-3′; Lep1R, 5′-TAAACTTCTGGATGTCAAAAATCA-3′). PCRs were mixed as follows: 7.2 µL double distilled H2O, 2.0 µL 10× buffer, 2.0 µL MgCl2, 1.0 µL Lep1-F, 1.0 µL Lep1-R, 0.4 µL dNTPs, 0.4 µL Taq, 6.0 µL DNA for a 20-µL total reaction volume. Reactions were run in an Eppendorf Mastercycler gradient thermocycler following cycling procedures from Hebert *et al.* (2004): 94 °C for 1 min; followed by six cycles of 94 °C for 1 min, 45 °C for 90 s and 72 °C for 75 s; followed by 36 cycles of 94 °C for 1 min, 51 °C for 90 s and 72 °C for 75 s; followed by a final extension period of 72 °C for 5 min. Amplification success and product size were confirmed via gel electrophoresis (3 µL sample load, 1% agarose gel, 87 V run, 25–40 min GelRed bath stain). Ninety-six-well plates (twin.tec, semi-skirted, blue, cat. no. 951020362) were loaded with successful reactions and 17 µL of PCR product for each sample was sent to Eurofins Genomics in Louisville, Kentucky, for PCR cleanup and Sanger sequencing.

**Sequence quality and alignment**

COI sequence data were visually inspected for quality in BioEdit (Hall, 1999), and chromatogram residues were manually edited for base call accuracy. A set of 172 new sequences (GenBank accession numbers KP747134–KP747305) were combined with 194 sequences previously published in Rehan & Sheffield (2011). This total set of 366 sequences was aligned via ClustalW using default settings (Thompson, Higgins & Gibson, 1994), then trimmed to a consensus region of 575 unambiguous, gap-free base pairs for further analysis. These sequences were then screened via BLAST database search for Wolbachia contamination following Groom, Stevens & Schwarz (2013). Three outgroup COI haplotypes were included to root the ingroup, including two western *Ceratina* (*Zadontomerus*) species, *C. acantha* and *C. nanula*, and one species from a sister subgenus, *Ceratina* (*Ceratinula*) *cockerelli* (Rehan *et al.*, 2010a, b). *Ceratina cockerelli* is sympatric in southern distribution to the eastern *Ceratina* (*Zadontomerus*) species group (Daly, 1973).

**Haplotype diversity and population genetic analyses**

Minimum spanning trees (MSTs) were constructed using Haploview (Salzburger, Ewing & VonHaeuseler, 2011) and were partitioned both by species and by regional collection location (i.e. north-east ‘NE’, north-central ‘NC’, mid-Atlantic ‘MA’, and south-east ‘SE’; Figs 1, 2). These regional assignments are representative of biologically distinct ecoregions, which are well defined by the US Environmental Protection Agency (http://www.epa.gov/wed/pages/ecoregions/na_eco.htm; McMahon *et al.*, 2001) and utilized in other studies (White *et al.*, 2009). All population genetic analyses were performed in Arlequin v. 3.5.1.2 (Excoffier & Lischer, 2010). At the intra-population level we ran Tajima’s *D* and Fu’s *F* tests of neutrality, based on 1000 simulations, for each species. Analysis of molecular variance (AMOVA) was performed to compare genetic variation within and among each species, and within and among the regional populations of each species. Species regional population genetic structure was assessed using Wright’s F-statistics calculated over the course of 1000 permutations, capturing *F*ST and pairwise differences between populations. To test whether populations were significantly differentiated from each other based on haplotypes, an exact test of sample differentiation was performed.

**Molecular clock estimates**

Fossil records for samples of North American *Ceratina* do not exist, and thus we estimated species divergence times using an approximate mitochondrial mutation rate. Because base composition can...
strongly influence mutation rate (Montooth & Rand, 2008), we followed Groom et al. (2015), who performed a similar analysis, and compared the AT percentage of our Ceratina dataset with that of their Lasioglossum population. Ceratina AT content comprised ~74% of its dataset, identical to that of Lasioglossum (Groom et al., 2015). As in Groom et al. (2015) we used D. melanogaster’s mitochondrial mutation rate (6.2 \times 10^{-8} mutations per site per generation; Haag-Liautard et al., 2008) to aid in our divergence time estimations. We assumed one generation per year as reported for eastern North American Ceratina species (reviewed by Rehan & Richards, 2010b).

**Molecular phylogenetic analysis**

Phylogenetic analyses were performed using Bayesian inference implemented in BEAST v.1.8.2 (Drummond et al., 2012) with a lognormal relaxed molecular clock (normal rate = 6.2 \times 10^{-8}, SD = 1 \times 10^{-6}) for 60 million iterations, sampling every 6000th iteration under a Yule Process specification tree model. A lognormal relaxed clock was employed for the data set to allow for rate variation among species (Duchène, Lanfear & Ho, 2014). Data included 102 unique ingroup haplotypes obtained from the total trimmed sequences (Supporting Information, Table S1), as well as two western Ceratina (Zadontomerus) species, C. acantha and C. nanula, and one Ceratina (Ceratinula) cockerellie sequence as outgroups. The most suitable substitution prior was determined as HKY + I + G by an Akaike information criterion (AIC) in jModelTest v.2.1.7 (Darriba et al., 2012). Codon positions were partitioned to place unique weight on the third position [i.e. (1 + 2), 3]. With no calibration data, the ‘treeModel.rootHeight’ prior was left unassigned, such that branch lengths would be proportional to age, given an estimated COI mutation rate (i.e. clock.rate). Analyses were performed in triplicate to ensure congruence. Subsequent log files were analysed in Tracer v.1.6 (Rambaut et al., 2014) to assess acceptability of estimated sample size (ESS > 200). TreeAnnotator (included in the BEAST software package) was used to select a Maximum Clade Credibility (MCC) phylogeny from 10 000 trees (following a one thousand tree burn-in) generated during the Markov chain Monte Carlo (MCMC) process. This MCC tree was then annotated in FigTree (http://tree.bio.ed.ac.uk/software/figtree/) to display posterior probability.

**Haplotype diversification estimation**

Both a lineage through time (LTT) and Bayesian Skyline Plot (BSP) were generated for eastern North American Ceratina following a modification of the run used to secure initial phylogeny. First, outgroup sequences were trimmed from the dataset, such that resolution would be constrained to ingroup. The HKY + I substitution prior was most strongly supported by jModelTest for the ingroup dataset. Under the Bayesian skyline coalescent tree model, we assigned a tree model root height using time to most recent common ancestor (tMRC) estimated in initial MCMC phylogenetic analysis (treeModel-rootHeight = 1.80 \times 10^5, SD = 2.5 \times 10^4 Normal prior). Clock.rate prior was again set following previously published mtDNA mutation rates (lognormal clock, normal rate = 6.2 \times 10^{-8}, SD = 1 \times 10^{-6}; Haag-Liautard et al., 2008). MCMC was again run, in triplicate, for 60 million generations, sampling every 6000th generation, and subsequent log files were assessed in Tracer for acceptability of estimated sample size (ESS > 200). These data were then used to generate an eastern North America Ceratina (Zadontomerus)-specific MCC tree in TreeAnnotator, subsequently annotated in FigTree to display posterior probability (PP) scores and 95% highest posterior density values. Tracer was then used to perform BSP and LTT analyses (Rambaut et al., 2014).

**Results**

A total of 366 mtDNA barcode sequences were obtained for the eastern North America Ceratina species group across each study species’ known range (Figs 1, 2). While C. calcara, C. dupla and C. strenua were sampled across all defined study regions (i.e. NE, NC, MA, SE), C. mikmaqi was not collected in the south-east, and C. floridana was found only in the south-east and mid-Atlantic. BEAST analysis recovered eastern North American Ceratina as a monophyletic clade (PP = 1.0; Supporting Information, Fig. S1). Within the eastern North American C. (Zadontomerus) species group, monophyly of each species was strongly supported [C. dupla (PP = 0.999), C. floridana (PP = 0.987), C. calcara (PP = 0.999), C. strenua (PP = 0.995) and C. mikmaqi (PP = 0.997)]. Ceratina mikmaqi was strongly supported as basal to the rest of the eastern Ceratina (Zadontomerus) species group (PP = 1.0). Ceratina calcara and C. strenua were recovered as sister species (PP = 0.799), as were C. dupla and C. floridana (PP = 0.901), although the C. calcara + C. strenua and C. dupla + C. floridana clades were not well resolved as direct sister clades (PP = 0.37; Supporting Information, Figs S2, S3).

In the 366 COI sequences there were 96 polymorphic sites, with between 14 and 28 unique
haplotypes within any one species (Supporting Information, Table S1). Pairwise comparisons among all individuals revealed significant sequence divergence both between and within species, and between two and six fixed nucleotide differences between species (Supporting Information, Table S2). When applied to our dataset, the Drosophila melanogaster mtDNA mutation rate \( (6.2 \times 10^{-8} \times 575 \text{ bases} \times 1 \text{ generation/year} = 3.565 \times 10^{-5} \text{ fixed mutations/year} \) estimated approximately one fixed mutation every 28,000 years (i.e., 1 fixed mutation/3.575 \( \times 10^{-5} \) mutations/year = 28,050 years/fixed mutation). Estimates of population divergence suggest a steady series of speciation events, starting around 175 kya (C. mikmaqi), and continuing at 132 kya (C. dupla + C. floridana and C. calcarata + C. strenua crown), 92 kya (C. calcarata + C. strenua crown) and finally 83 kya (C. dupla + C. floridana crown); LTT plotting reflects this slow, steady increase in genetic diversity (Fig. 3A). Bayesian skyline analysis supports a relatively recent (between 25 and 10 kya) population expansion to the broad range observed today (Fig. 3B).

MST construction revealed distinct population structuring among species (Fig. 4) and within species by region (Supporting Information, Figs S4, S5). AMOVA revealed an overall \( F_{ST} = 0.76 \) (P < 0.001) for the eastern Ceratina (Zadontomerus) species group, and \( F_{ST} \) values ranging from 0.68 to 0.82 between each species (P < 0.001). Average pairwise base pair differences between species were also significant, with a range of 4.00-9.09 differences between species groups (P < 0.001). Tajima’s D test of neutrality yielded significantly negative D values for C. strenua and C. dupla, with slightly but not significantly negative values for all other species (Supporting Information, Table S3). Fu’s \( F_{S} \) test of neutrality yielded significantly negative \( F_{S} \) scores for all species ranging from −7.75 to −17.5 (P-values ranging from < 0.001 to 0.006; Supporting Information, Table S3).

Each species was assessed by region for analysis of sub-population structure. Within species, C. calcarata had an \( F_{ST} \) of 0.128 (P < 0.001), and between 3.21 and 3.55 average base pair differences between each significantly distinct population (Supporting Information, Table S4). Ceratina strenua had no significant fixation indices or average pairwise differences within or between regional populations [overall within-species \( F_{ST} \) of 0.075 (all P ≥ 0.05; Supporting Information, Table S5)].

Ceratina dupla had an overall within-species \( F_{ST} \) of 0.321 (P < 0.001), with between 1.909 and 3.157 average base pair differences between each significantly distinct population and significant pairwise \( F_{ST} \) ranging from 0.193 to 0.395 (P < 0.05; Supporting Information, Table S6). All regional C. dupla populations were significantly distinct from each other except between mid-Atlantic and south-east populations. Ceratina mikmaqi had no significant fixation indices or average pairwise differences within or between populations [overall within-species \( F_{ST} \) of 0.004 (P = 0.337; Supporting Information, Table S7)]. Ceratina floridana had an overall within-species \( F_{ST} \) of 0.162 (P < 0.001), with a significant average base pair difference of 2.400 (P < 0.001; Supporting Information, Table S8).

DISCUSSION

Our results confirm the distinct identities of five largely sympatric Ceratina (Zadontomerus) species in eastern North America and support the recent description of C. floridana and C. mikmaqi as discrete species (Rehan & Sheffield, 2011). Collectively, the eastern North American Ceratina (Zadontomerus) species are very closely related at the COI locus, with a maximum fixed nucleotide difference of 6 bp between C. mikmaqi and C. calcarata, down to as few as two fixed base pairs between C. dupla and C. floridana; these numbers are consistent with previous studies (Rehan & Richards, 2008; Rehan & Sheffield, 2011). Our expanded sampling has greatly increased our understanding of the incidental range of each of these species, extending C. floridana’s known range into the mid-Atlantic region, and framing C. mikmaqi as occurring primarily in the northeast (Figs 1, 2).

SPECIES DISTINCTIONS

Given the broad sympathy of these five species, it is very likely that they are still in the early stages of divergence. Ceratina dupla and C. calcarata have even been observed to viably hybridize in greenhouse mating experiments (Hung & Norden, 1987). Ceratina dupla and C. mikmaqi have broadly overlapping ranges but distinctly structured populations (\( F_{ST} = 0.77, \ P < 0.001; \) sequence divergence = 1.34%). By contrast, C. mikmaqi and C. floridana were found to have one of the lowest \( F_{ST} \) values of any two species considered in the study (\( F_{ST} = 0.69, \ P < 0.001; \) sequence divergence = 0.69%) despite lowest overlap in species’ ranges.

Our Tajima’s D and Fu’s \( F_{S} \) values were significantly negative on average, suggesting relatively recent population expansion events and genetic drift (Supporting Information, Table S3). Fu’s \( F_{S} \) is calculated using only observed vs. expected allelic diversities, so the comparatively less negative scores measured in C. mikmaqi and C. floridana directly
Figure 3. A, log lineage through time (LTT) plot based on Bayesian skyline chronogram. Maximum time is the root height mean and dotted vertical line represents lower 95% highest posterior density. B, maximum clade credibility (MCC) tree for eastern North American Ceratina (Zadontomerus). Roman numerals indicate estimated times of divergence for (i) C. mikmaqi; (ii) C. calcarata + C. strenua + C. dupla + C. floridana; (iii) C. calcarata + C. strenua; and (iv) C. dupla + C. floridana. Posterior probability support values for the monophyly of each species are provided. C, Bayesian skyline plot displaying estimated effective population size ($N_e$) over time. Dotted grey line indicates 95% highest posterior density limit. Timescale maximum is mean root height. D, change in atmospheric temperature ($^\circ$C) over the past 165 kyr (from ice core data; Petit et al., 1999). Yellow bars indicate 20 kya mark.
support their observed physical range limitations. However, mtDNA is highly susceptible to genetic drift and this tendency to accrue silent site mutations could be contributing, at least in part, to the significantly negative $D$ and $F_S$ values (Beebee & Rowe, 2004).

Species that are morphologically monotonous can be identified by even minimal allelic variation, and DNA barcoding has been used to assess current and historical population structure in other closely related bee species (Packer et al., 2008; Hurtado-Burillo et al., 2013; Groom et al., 2015) as well as within other orders (e.g. Diptera, Smith et al., 2006; Coleoptera, Grebennikov, 2014). However, the barcoding process theoretically relies on relatively high interspecific and relatively low intraspecific variation for best resolution of identities (Packer et al., 2008). As in other bee barcoding studies (Groom et al., 2013) we took measures to confirm our results were not the result of cryptic infection via endosymbiotic bacteria Wolbachia, which co-segregates with an organism’s mtDNA, obscuring multiple species’ true mitochondrial identities (Whitworth et al., 2007). Infection via Wolbachia can be acceptably ruled out, as our phylogeny is strongly supported by morphological identity of voucher samples (Rehan & Sheffield, 2011) as well as by Bayesian analysis. Furthermore, BLAST search yielded no Wolbachia hits for our sequence set. Expanding our suite of mitochondrial loci in combination with highly variable microsatellite loci would only deepen our understanding of current and historical population structure and dynamics of our five highly sympatric Ceratina (Zadontomerus) species.

Given the low frequency of fixed nucleotide differences among eastern Ceratina (Zadontomerus) species, speciation events probably occurred relatively recently. Bayesian analysis supports a crown divergence within eastern Ceratina (Zadontomerus) as recently as 200 kya (mean age of 180 kya; Fig. 3C, Supporting Information, Fig. S2) following an earlier divergence between sister subgenera Zadontomerus and Ceratinula ~30 Mya (Supporting Information, Fig. S1; Rehan & Schwarz, 2015). Bayesian analysis also supports our significantly negative Tajima’s $D$ and Fu’s $F_S$, and suggests an ongoing population expansion event initiated sometime towards the end of the Wisconsinan glacial period (Fig. 3A).

HISTORICAL PHYLOGEOGRAPHY

Over the course of its history, North America has experienced repeated periods of glacial expansions and contractions, involving the growth and recession of massive sheets of ice across the continent (Rand, 1948; Pielou, 1992). This process occurs over many thousands of years, broadly affecting the global climate and, during periods of glacial maxima, forcing...
all terrestrial life south, into areas of refugia (Pielou, 1992; Hewitt, 1996). As the most recent glacial period began its final thaw around 20 kya (Pielou, 1992), the contraction of the Laurentide ice sheet opened up vast regions of habitable terrain (Rand, 1948; Hewitt, 1996). The Pleistocene climate dynamically shaped the phylogeography of bees around the globe, from the Pacific Islands (sweat bees; Groom et al., 2015) to the Americas (orchid bees; López-Uribe et al., 2014), and sparked vicariance events in other Hymenoptera genera (parasitic wasp genus Nasonia; Raychoudhury et al., 2010). In North America, plants and pollinating insects migrated north, their formerly cloistered populations separating widely and diversifying rapidly (Pielou, 1992; Soucy & Danforth, 2002; Hines, 2008; Jaramillo-Correa et al., 2009).

This thawing process reached its zenith from 10 to 6 kya (the 'Hypsithermal'), greatly expanding the habitable range of migrant temperate species in relatively little time, and marking the beginning of our present Holocene (Pielou, 1992).

Prior to the Sangamonian interglacial (~135–120 kya), global temperatures matched, and even temporarily exceeded, those we experience today (Petit et al., 1999). Ceratina (Zadontomerus) populations may have occupied a very similar range to what we presently observe, with C. mikmaqi representing the only extant eastern Zadontomerus species at that time (Fig. 3B; node i). Sometime during the Sangamonian interglacial period the shared common ancestor of C. calcarata and C. strenua began to diverge from the shared common ancestor of C. dupla and C. floridana (Fig. 3B; node ii). North America then slowly and steadily approached the onset of the Wisconsinan glacial maxima as the climate cooled over a span of 75 000 years (~100–25 kya; Pielou, 1992). Prior to any significant glaciation, C. calcarata may have broken away from C. strenua (~92 kya; Fig. 3B; node iii) and C. dupla from C. floridana (~83 kya; Fig. 3B; node iv). During the glacial maxima, these populations would have been forced south, into numerous and potentially ecologically segregated refugia. During the long, final recession of the Laurentide ice sheet (~20–11 kya, initiation indicated by vertical yellow lines in Fig. 3), these groups followed angiosperms north and east as flowering plants rapidly filled in the landscape along distinct natural corridors revealed by the thaw (Hewitt, 1996). Over the course of around 10 000 years, including the Hypsithermal period (Pielou, 1992), these populations probably travelled broadly and rapidly around eastern North America, spreading along the Appalachian Mountain range and eastern seaboard, to what is presently the northernmost extent of their known range in southern Canada (Swenson & Howard, 2005).

**EcoLocial niches**

The ecologies of the eastern Ceratina (Zadontomerus) species (not including C. floridana, considered an ecomorph of C. dupla until only recently) have been studied in detail (Daly, 1973; Michener, 2007). In each studied species, nesting is initiated when a mature female bores out a brood passage from the soft pith of certain preferred plant species (often staghorn sumac; Rau, 1928; Krombein, 1960; Grothaus, 1962; Rehan & Richards, 2010b). Ceratina calcarata begin dispersing around early to mid-May in the north-east, rearing a single brood through June or July (Rehan & Richards, 2010b), but may begin their season as early as March in the southeastern extent of their range (Kislow, 1976). Ceratina dupla have been observed to nest earlier in the season than either C. calcarata or C. mikmaqi, and each species has displayed unique preferences for brooding in particular plant species (e.g. sumac, teasel or raspberry) and for shady or sunny nest-sites (Grothaus, 1962; Vickruck et al., 2011). In addition, because of its larger body size, C. calcarata was found to prefer slightly wider nesting material than C. dupla and C. strenua within their respective sympatric ranges (Grothaus, 1962; Kislow, 1976). Given relatively recent and rapid population expansion, it is not unlikely that these numerous environmental, intra- and interspecies variations in phenology and nest-site quality have contributed significantly to the vicariance of these distinct but still closely related species groups.

**behavioural plasticity**

Bees are primarily composed of solitary species (Michener, 2007), but Ceratina is one of few genera known to contain socially polymorphic groups. Across the tropics species form both solitary and social nests in sympatry (Sakagami & Maeta, 1987; Rehan et al., 2010a, b; Rehan, Tierney & Wcislo, 2015), allowing for direct observation and comparison of solitary and social reproductive strategies within and between Ceratina species (Rehan et al., 2014a). Although solitary as they disperse and establish nests in spring, C. calcarata females shift from primarily foraging to prolonged maternal care once they begin to raise a brood (Rehan & Richards, 2010a). Ceratina calcarata is thus considered ‘subsocial’, its largely solitary lifestyle punctuated by a period of grooming and guarding its brood to maturity (Rehan & Richards, 2010b). Intriguingly, Ceratina females of various species, including C. calcarata and C. japonica, often rear a dwarf eldest daughter (Sakagami & Maeta, 1987; Rehan & Richards, 2010a). As Ceratina mothers directly control the amount of nourishment received...
by their offspring, it is thought this underfed female acts as a non-reproductive worker (Rehan, Berens & Toth, 2014b). However, due to a relatively short reproductive season, C. calcarata are univoltine in the northern US and Canada (Figs 1, 2). While this allows for some generational overlap and reproductive division of labour it is expected that southern populations of each Ceratina species may be operating under a bivoltine reproductive strategy (Rau, 1928; Rehan & Richards, 2010a).

There is growing evidence that bee species occurring at lower, warmer altitudes and latitudes show more developed social structure than even conspecifics living in significantly cooler conditions (Sakagami & Munakata, 1972; Soucy & Danforth, 2002; Cronin & Hirata, 2003; Kocher et al., 2014). In socially polymorphic species (e.g. the halictid bee Lasio glossum baleicum), these variations in environmental conditions may shape conspecific behaviour by limiting phenotypic expression of extant social genes (Hirata & Higashi, 2008). Therefore, while ecological specialization probably plays a role in maintaining species identity, latitudinal variation in season length and quality may be shaping the social structure, and thus evolutionary trajectory, within and between North American Ceratina (Zadontomerus) species. Additional phylogeographical and sociogenomic investigation could thus explore the extent to which environment influences the expression of sociality in C. calcarata across the broad latitudinal gradient of its endemic range.

CONCLUSIONS

The five eastern North American Ceratina (Zadontomerus) species collectively represent a largely sympatric, widely occurring and richly diverse native pollinator group. These species, still largely uniform in morphology and known ecology, emerged in just the last 200 kyr, and experienced a massive population expansion event as recently as 20 kya. Although collectively semi-sympatric, this group appears to occupy species-specific niches, which may be contributing to what is probably ongoing vicariance. Of the eastern species, C. calcarata is emerging as a model organism for the study of social polymorphism and pollinator phylogeography (Rehan & Sheffield, 2011; Rehan et al., 2014b). Further research expanding genetic resources and phenological data will allow for continued exploration of the effects of environmental variation on the dynamics and development of social polymorphism, nesting biology and population structure.

ACKNOWLEDGEMENTS

We thank John Barthell, Sam Droge, April Hamblin, Joan O’Donnell, Scott Prazijner and Tracy Zarillo for providing specimens. Sean Lombard, Nicholas Pizi and Jacob Withee assisted with field collections. We also thank Erika Tucker and two anonymous reviewers for their helpful comments. Funding from the University of New Hampshire to S.M.R. supported this work.

REFERENCES


SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Figure S1. Maximum clade credibility tree for eastern North American Ceratina (Zadontomerus); posterior probability values support eastern Zadontomerus subgenus as monophyletic. Outgroups include C. (Ceratinula) cockerelli, C. (Zadontomerus) acantha and C. (Zadontomerus) nanula.

Figure S2. Maximum clade credibility tree for all eastern North American Ceratina (Zadontomerus) species. Blue bars indicate 95% highest posterior density confidence intervals for each time estimation.

Figure S3. Maximum clade credibility tree for all eastern North American Ceratina (Zadontomerus) species. Posterior probability values are presented for all haplotypes.

Figure S4. Minimum spanning tree constructions of C. calcara (A) and C. strenua (B) haplotypes. Each circle represents a unique haplotype; values indicate numbers of individuals that share that particular sequence. Colours indicate collection location of each individual: north-east (NE), north-central (NC), mid-Atlantic (MA) and south-east (SE).

Figure S5. Minimum spanning trees of C. mikmaqi (C), C. dupla (D) and C. floridana (E) haplotype networks. Each circle represents a unique haplotype; values indicate numbers of individuals that share that particular sequence. Colours indicate region of origin of each individual: north-east (NE), north-central (NC), mid-Atlantic (MA) and south-east (SE).

Table S1. Sequence and haplotype totals for all study species. Previously published sequences are taken from Rehan & Sheffield (2011).

Table S2. Overall population structure of five co-occurring Ceratina (Zadontomerus) species. Diagonal indicates average pairwise differences within species and value in parentheses indicates percentage sequence divergence within those species; above diagonal are average pairwise differences between species and parentheses indicate percentage sequence divergence between those species; below diagonal are pairwise FST values, along with fixed nucleotide differences between species in parentheses. Significant values (P < 0.001) are indicated in bold. Fixation index over all loci FST = 0.76 (P < 0.001).

Table S3. Tajima’s D and Fu’s FS tests of neutrality. Sample size, segregating sites (S), Tajima’s D score and significance value (D P-value), and Fu’s FS value and significance values (FS P-value) are presented. Values in bold are statistically significant (P < 0.05).

Table S4. Ceratina calcara (A) regional population structure. Diagonal indicates average pairwise differences within populations, and number in parentheses indicates total number of sequences for that region; above diagonal are average pairwise differences between populations; below diagonal are pairwise FST values. Significant values (P < 0.05) are indicated in bold. Fixation index over all loci FST = 0.128 (P < 0.001).

Table S5. Ceratina strenua regional population structure. Diagonal indicates average pairwise differences within populations, and number in parentheses indicates total number of sequences for that region; above diagonal are average pairwise differences between populations; below diagonal are pairwise FST values. Values are insignificant (P ≥ 0.05). Fixation index over all loci FST = 0.075 (P = 0.115).

Table S6. Ceratina dupla regional population structure. Diagonal indicates average pairwise differences within populations, and number in parentheses indicates total number of sequences for that region; above diagonal are average pairwise differences between populations; below diagonal are pairwise FST values. Significant values (P < 0.05) are indicated in bold. Fixation index over all loci FST = 0.321 (P < 0.001).

Table S7. Ceratina mikmaqi regional population structure. Diagonal indicates average pairwise differences within populations, and number in parentheses indicates total number of sequences for that region; above
diagonal are average pairwise differences between populations; below diagonal are pairwise $F_{ST}$ values. No values are significant ($P \geq 0.135$). Fixation index over all loci $F_{ST} = 0.004$ ($P = 0.337$).

**Table S8.** *Ceratina floridana* regional population structure. Diagonal indicates average pairwise differences within populations, and number in parentheses indicates total number of sequences for that region; above diagonal are average pairwise differences between populations; below diagonal are pairwise $F_{ST}$ values. Significant values ($P < 0.001$) are indicated in bold. Fixation index over all loci $F_{ST} = 0.162$ ($P < 0.001$).