POPULATION STRUCTURE AND GENE FLOW IN SOLITARY BEES: IMPLICATIONS FOR THEIR CONSERVATION AND MANAGEMENT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Margarita María López Uribe

August 2014

POPULATION STRUCTURE AND GENE FLOW IN SOLITARY BEES: IMPLICATIONS FOR THEIR CONSERVATION AND MANAGEMENT

Margarita María López Uribe, Ph. D.

Cornell University 2014

The study of genetic population structure is a central issue in evolutionary biology because it determines the distributional pattern and amount of genetic variation that is available for evolution within a species. Therefore, species adaptability to environmental change operates within the constraints imposed by population structure. Bees have recently become the focus of conservation concern due to the increasing evidence of population declines worldwide drawing attention to the ecological and economic consequences of pollinator loss. Results from multiple studies have identified three major drivers of bee decline: (1) Environmental stressors, including habitat loss and pesticide exposure; (2) Pests and pathogens; and (3) Loss of genetic diversity. Despite its importance, the levels and distribution of genetic diversity remain poorly understood in bees. Here, I investigate the patterns of genetic diversity in five solitary bee species to understand how climate, crop domestication and landscape features drive changes in bee population demography and genetic structure. I developed *de novo* genetic markers and spatial models for all species, to determine how populations are structured and connected through gene flow at different geographic scales. I find evidence that changes in climatic conditions, range shifts in host-plants due to crop domestication, and the distribution of suitable nesting sites are important predictors of levels of genetic diversity and population structure in solitary bees. My results reveal that male-biased dispersal may be common in bees and the production of high

frequencies of sterile diploid males are not a necessary outcome of populations with low genetic variability. I also provide evidence that drier climates, tillage and low-density of suitable nesting areas may diminish bee population abundance. In summary, I find that investigating patterns of genetic variability in bees using molecular tools provides significant insights into how environmental stressors affected past and current population demography. My results have implications for bee conservation and the effective management of wild bee populations for crop pollination.

BIOGRAPHICAL SKETCH

Margarita María López Uribe was born in Bogotá, Colombia on February 22nd of 1981. She grew up and lived in Cúcuta with her parents and sister until she graduated from high school in 1997. The following year, she moved to Canada for one year, and in 1999, Margarita started pursuing her B.A. degree in Biology at Universidad de los Andes in Bogotá. During a trip to the Colombian Amazon for an undergraduate Entomology course, she encountered the first orchid bee in the field and, since then, she fell in love with bees and their natural history. She obtained her Bachelor's degree in 2004 and moved to Brazil where she started her Master's studies in Genetics and Evolution at Universidade Federal de São Carlos in the laboratory of Dr. Marco Antonio Del Lama. After obtaining her Master's degree with high honors, she moved to Panama for a short-term internship at the Smithsonian Tropical Research Institute to work with Bill Wcislo on nocturnal bees. In August of 2007, she came to Cornell University to pursue her Ph.D in Entomology with Dr. Bryan Danforth.

To my family

For their love and support through all these years.

ACKNOWLEDGMENT

I especially thank my advisor Bryan Danforth for giving me the opportunity to be part of her research group and for supporting all the crazy ideas that I had during graduate school. He provided invaluable support and encouragement throughout all these years at Cornell. I also thank my committee members Kelly Zamudio, Steve Morreale and Matt Hare for providing a lot more than just feedback on manuscripts, but giving me support, advise on professional development and life-work balance.

Many thanks to my collaborators Kelly Zamudio, Steve Morreale, Robert Paxton, Antonella Soro, Robert Minckley, Jim Cane and Carolina Cardoso for their help with ideas and executing projects that later became part of my dissertation. All this work could have not been possible without the two undergraduate students that helped me in the laboratory: Amy Green and Christine Santiago.

I am indebted to all the members of the Danforth Lab that overlapped with me during my time at Cornell: Sophie Cardinal, Jesse Litman, Mia Park, Christophe Praz, Dennis Michez, Jason Gibbs, Shannon Hedtke, Kristin Brochu, Laura Russo, Mary Centrella and Achik Dorchin. You made a huge impact on my development as a graduate student. Thanks to my friends Stuart Campbell, Jacob Crawford, Rayna Bell and the members of the Zamudio lab for providing a great intellectual environment to discuss ideas.

Linda Rayor through the Naturalist Outreach Practicum, and Derina Samuels through the Center for Teaching Excellence, gave the opportunity to be associated with their programs and be actively involved in teaching while improving my public communication skills.

I would like to thank my family, my parents Martha and Juan, and my little sister Pily, for being the foundation of everything that I am and giving me the emotional support throughout my career. Finally, thanks to my husband Luis for his love and support in every step of my graduate work: field work, lab work, teaching and writing, and to my beautiful son Tiago from bringing immense happiness and the perfect balance into my life.

The major funding sources for my dissertation came from National Science Foundation Research Grants to Bryan Danforth (DEB-0814544 and DEB-0742998). Funding for field trips and laboratory reagents were provided by the Grace Griswold Fund, Rawlins Endowment, Cornell Graduate Travel Research Grant, Einaudi Center Travel Grant, Organization for Tropical Studies (OTS), Lewis and Clark Exploration Fund, Explorer's Club Exploration Fund, Andrew W. Mellon and Sarah Bradley Fellowship. My graduate studies were funded by teaching assistantships, the Ollin and Palmer Fellowships from Cornell University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The intrinsic factors driving bee responses to environmental change

Insect pollinators play a critical role in pollination ecosystem services, as they are responsible for the increased quantity, quality, and stability of over 75% our crops (Ollerton *et al.* 2011). Furthermore, bees are the main pollinators of flowering plants in natural areas (Kearns *et al.* 1998; Klein *et al.* 2007) playing a key role in the maintenance of biodiversity. Therefore, declines in bee populations worldwide have raised concerns about the ecological and economic consequences of losing the pollination services they provide (Burkle *et al.* 2013).

The most important managed pollinator is the European honey bee (*Apis mellifera*), which is capable of increasing yield in 96% of animal-pollinated crops (Klein *et al.* 2007). However, clear evidence of declines in the number of honey bee colonies in North America and Europe have uncovered the possibility of future unavailability of pollination services. This scenario is even worse given that the proportion of agricultural crops that relies on bee pollination is rapidly increasing, and so are the demands for pollination services (Aizen & Harder 2009). The interaction between environmental stressors, pest and pathogens, and loss of genetic diversity are attributed as the main drivers of honey bee declines during the last decade (Becher *et al.* 2013).

Native bees can provide insurance against a potential pollination crisis caused by declines in the number of honey bee colonies (Winfree et al. 2007). However, native bees are also in decline as a result of a combination of

habitat loss, agricultural intensification, increased pathogen loads, and climate change (Potts *et al.* 2010). Numerous studies have demonstrated the negative effects of environmental stressors on native bees [land-use change (Kremen *et al.* 2007); pesticide exposure (Whitehorn *et al.* 2012); pathogen pressure (Fürst *et al.* 2014)]. Therefore, the main drivers of bee decline on bee communities are now well established but the mechanisms behind these patterns are not well understood.

However, the role of *intrinsic factors* in responses to stress has received little attention (Brown & Paxton 2009; Jauker *et al.* 2012; Kremen *et al.* 2007) (Fig. I.1). The investigation of taxon-specific responses can shed light into the mechanisms behind adaptation and survival to disturbance and environmental change in bees. A summary of some of the main life-history traits that can predict species responses to environmental stress are listed below.

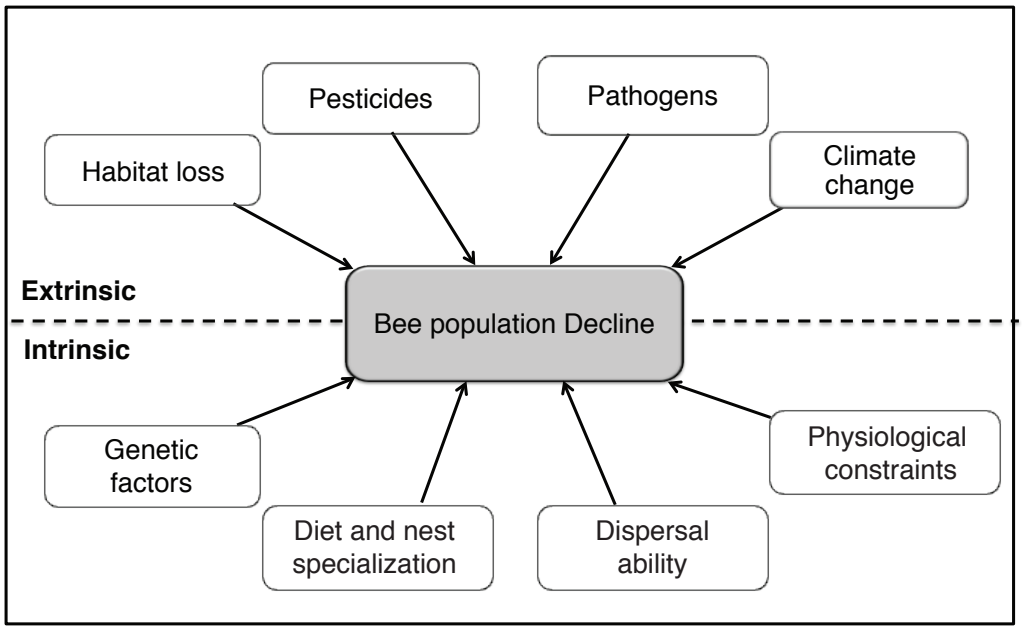


Figure I.1. Conceptual framework of extrinsic and intrinsic drivers to bee population decline.

Genetic factors

Under demographic scenarios of reduction of population size and isolation, some genetic features particular to bees predispose them to extinction at the population level (Packer & Owen 2001). First, as haplodiploid insects, bees have a smaller effective population size than diplodiploid insects (for a sex ratio of ¾ Ne instead of 4Ne). This predicts that the negative effects of genetic drift in small populations would be stronger than in diploid organisms. Second, sex in bees is determined at a complementary sex determination (csd) locus which, when homozygous, causes production of sterile diploid males instead of females (Beye *et al.* 2003b). Diploid male production further reduces effective population size and increases genetic load (Hedrick *et al.* 2006). This feedback loop between small effective population size and increased diploid male production can lead to an "extinction vortex" according to theoretical models (Zayed & Packer 2005). Therefore, small effective population sizes and low genetic variability are hypothesized to have strong detrimental effects in bees at the demographic level.

Even though no empirical data is available to support the relationship between genetic diversity and pesticide/pathogen exposure in wild bees, the same patterns of increased immunity with higher genetic diversity observed in other insects are also expected. Thus, bee populations with higher levels of genetic variability and larger effective population sizes are predicted to be buffered against the negative impacts of climate change, pesticide exposure and pathogen attack.

Diet and nesting requirements

Comparative studies on the effect of habitat fragmentation on bee communities show that the abundance and richness of bees with specialized diets are more negatively affected than other groups within the bee community (Alfert *et al.* 2001; Cane *et al.* 2006b). Floral host specialization is a trait that may increase vulnerability to extinction because 1) specialization is related to decreased species abundance (Lawton *et al.* 1994) and 2) availability of suitable habitat/resources for specialists is generally more restricted than for generalists (McKinney 1997; Polus *et al.* 2007). However, bee's responses are directly related to the abundance of their host plants. In some cases, pollen specialists can be positively affected by agriculture if crops are within the diet requirements of the bee (e.g. *Peponapis pruinosa*). On the other hand, bee species that exclusively rely on pollen from crops could be more severely affected by long-term pesticide exposure from contaminated pollen. However, this question has not been investigated.

Habitat specialization for nesting requirements can make some bee species susceptible to local extinctions due to land-use change for agriculture or urbanization. For example, the sand dune specialist *Colletes floralis* is now an endangered species in Great Britain due to extreme reduction of population sizes and isolation driven by urbanization of previously suitable habitats (Davis *et al.* 2010). Therefore, the prediction is that responses to environmental change from bee species that exhibit diet or habitat specialization will be partially driven by changes in the availability of resources given unless species can adapt and broaden their niches (Briggs *et al.* 2013).

Dispersal abilities

Bees are highly mobile organisms with potential for long distance dispersal (Wikelski *et al.* 2010). Dispersal ability in bees is believed to be positively correlated with body size (Greenleaf et al. 2007). Bee sizes vary between 2 to 40 mm, which translates into flying distances that vary between ~100 m in the smallest bees to over 5 km in the largest bees (even though there are records of flying distance of over 20 km in one day, see Janzen 1971). Studies of population structure in bees general indicate low genetic differentiation at regional and even continental geographic scales suggesting high levels of gene flow among populations (Beveridge & Simmons 2006; Exeler *et al.* 2008; López-Uribe *et al.* 2014; Zayed & Packer 2002). However, the wide variation in flight capacity among species raises the question of how different bees respond to increasing amounts of impervious habitat. In fact, some bee species show moderate to high levels of genetic differentiation in highly fragmented habitats (Darvill *et al.* 2006; Davis *et al.* 2010; Jha & Kremen 2013) but some show low genetic structure even in highly modified habitats such as urban areas (Cerantola *et al.* 2010). In spite of its critical importance, dispersal remains one of the least understood aspects of bee biology and it is a major gap for the establishment of effective conservation and management plans to enhance bee populations and their pollination services. Bees with greater dispersal capacities would have the ability to better respond to environmental stress because they could (1) search for resources in low quality habitats or greater areas, or (2) disperse and colonize new areas if local environmental conditions are below what it is sustainable.

On the other hand, rapid pathogen spillover and interspecies transmission across large geographic areas is a great threat in bees due to their great dispersal capacities. For example, the dramatic decline in multiple

bumble bee species in North America is attributed to an exotic pathogen (*Nosema bombi*) introduced from Europe in the mid 90's that rapidly infected and spread through different native species (Cameron *et al.* 2011). Evidence of this was also recently demonstrated for bumble bees in Europe that have higher pathogen prevalence in areas where they share floral resources with managed honey bees that have high pathogen loads (Fürst *et al.* 2014). Therefore, high rates of pathogen spill over is a realistic prediction for bee species with higher dispersal rates.

Physiological constraints

Bees are ectothermic organisms strongly dependent on ambient temperature for the daily activities that require flight. Bees use behavioral strategies to regulate body temperature, such as searching for shade when they are overheated or basking when they need to warm up (Willmer & Stone 2004). However, some bees, specially those with larger body size, can use endothermic mechanisms to increase their body temperature using indirect flight muscles without flapping their wings (Heinrich 1974). Nevertheless, bees are at risk of overheating, especially in environments where they are close to their upper critical body temperatures (e.g. deserts, tropical regions). In the face of global climate change, it is unclear how different bee species will adapt by modifying their behavior or increasing their temperature tolerance (Potts *et al.* 2010). Even though physiological constraints have been investigated in various groups of bees (Willmer & Stone 2004), the phylogenetic conservatism of their thermal properties is not clear. Because foraging and dispersal capacity are constrained by physiological limitations, it is key to integrate physiology into the life-history traits predicting how a taxon will respond to

environmental stress. Furthermore, species-specific responses to the toxic elements present in contaminated pollen in agricultural fields are unknown and may predict which species could be more tolerant to these conditions.

Relevance for conservation and management

Even though most of the research on environmental drivers of bee decline has focused on community responses, it is necessary to investigate species-specific responses to environmental stress to understand the mechanisms behind the observed patterns at the community level. An integrative approach that includes genetic factors, diet and habitat specialization, dispersal ability and physiological constraints, may provide the necessary framework to understand these mechanisms. Continued work on these aspects is a possible approach to provide recommendations that enhance bee populations in natural, agricultural and urban ecosystems. Furthermore, in the face of loss of ecosystem services in some parts of the world (Morreale & Sullivan 2010), the understanding of these intrinsic aspects of bee biology will better inform practices for pollinator management in highly intensified agricultural landscapes.

REFERENCES

Aizen MA, Harder LD (2009) The global stock of domesticated honey bees is growing slower than agricultural demand for pollination. *Current Biology* ***19***, 915-918.

Alfert T, Steffan-Dewenter I, Tscharntke T (2001) Bienen und wespen (Hymenoptera, Aculeata) in Kalksteinbrüchen: Der effect von flächengröße und flächenalter (Bees and wasps (Hymenoptera, Aculeata) in limestone quarries: the effects of area and age). *Mitt Dtsch Ges Allg Angew Ent* **13**, 579–582.

Becher MA, Osborne JL, Thorbek P, Kennedy PJ, Grimm V (2013) Towards a systems approach for understanding honeybee decline: a stocktaking and synthesis of existing models. *Journal of Applied Ecology* **50**, 868-880.

Beveridge M, Simmons L (2006) Panmixia: an example from Dawson's burrowing bee (*Amegilla dawsoni*) (Hymenoptera: Anthophorini). *Molecular Ecology* **15**, 951-957.

Beye M, Hasselmann M, Fondrk MK, Page RE, Omholt SW (2003) The gene *csd* is the primary signal for the sexual development in the honeybee and encodes as SR-type protein. *Cell* **114**, 419-429.

Briggs HM, Perfecto I, Brosi BJ (2013) The role of the agricultural matrix: coffee management and euglossine bee (Hymenoptera: Apidae: Euglossini) communities in southern Mexico. *Environmental Entomology* **42**, 1210-1217.

Brown MJ, Paxton RJ (2009) The conservation of bees: a global perspective. *Apidologie* **40**, 410-416.

Burkle LA, Marlin JC, Knight TM (2013) Plant-pollinator interactions over 120 years: loss of species, co-occurrence and function. *Science* **339**, 1611-1615.

Cameron S A, Lozier JD, Strange JP, Koch JB, Cordes N, Solter LF, Griswold TL (2011) Patterns of widespread decline in North American bumble bees. *Proceedings of the National Academy of Sciences* **108**, 662-667.

Cane JH, Minckley RL, Kervin LJ, Roulston TH, Williams NM (2006) Complex responses within a desert bee guild (Hymenoptera: Apiformes) to urban habitat fragmentation. *Ecological Applications* **16**, 632-644.

Cerantola NCM, Oi CA, Cervini M, Del Lama MA (2010) Genetic differentiation of urban populations of *Euglossa cordata* from the state of São Paulo, Brazil. *Apidologie* **39**, 410-418.

Darvill B, Ellis J, Lye G, Goulson D (2006) Population structure and inbreeding in a rare and declining bumblebee, *Bombus muscorum* (Hymenoptera: Apidae). *Molecular Ecology* **15**, 601-611.

Davis ES, Murray TE, Fitzpatrick U, Brown MJF, Paxton RJ (2010) Landscape effects on extremely fragmented populations of a rare solitary bee, *Colletes floralis*. *Molecular Ecology* **19**, 4922-4935.

Exeler N, Kratochwil A, Hochkirch A (2008) Strong genetic exchange among populations of a specialist bee, *Andrena vaga* (Hymenoptera: Andrenidae). *Conservation Genetics* **9**, 1233-1241.

Fürst MA, McMahon DP, Osborne JL, Paxton RJ, Brown MJF (2014) Disease associations between honeybees and bumblebees as a threat to wild pollinators. *Nature* **506**, 364-366.

Hedrick PW, Gadau J, Page Jr RE (2006) Genetic sex determination and extinction. *Trends in Ecology and Evolution* **21**, 55-57.

Heinrich B (1974) Thermoregulation in endothermic insects. *Science* **185**, 747-756.

Janzen DH (1971) Euglossine bees as long-distance pollinators of tropical plants. *Science* **171**, 203-205.

Jauker B, Krauss J, Jauker F, Steffan-Dewenter I (2012) Linking life history traits to pollinator loss in fragmented calcareous grasslands. *Landscape Ecology* **28**, 107-120.

Jha S, Kremen C (2013) Urban land use limits regional bumble bee gene flow. *Molecular Ecology* **22**, 2483-2495.

Kearns CA, Inouye DW, Waser NM (1998) Endangered mutualisms: the conservation of plant-pollinator interactions. *Annual Review of Ecology and Systematics* **29**, 83-112.

Kremen C, Williams NM, Aizen MA*, et al.* (2007) Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change. *Ecology Letters* **10**, 299-314.

Lawton JH, Nee S, Letcher A, Harvey P (1994) Animal distributions: patterns and processes. In: *Large-scale ecology and conservation biology* (eds. Edwards PJ, May RM, Webb N), pp. 41-58. Blackwell, New York.

López-Uribe MM, Oi CA, Del Lama MA (2008) Nectar-foraging behavior of Euglossine bees (Hymenoptera: Apidae) in urban areas. *Apidologie* **39**, 410-418.

López-Uribe MM, Zamudio KR, Cardoso CF, Danforth BN (2014) Climate, physiological tolerance and sex-biased dispersal shape genetic structure of Neotropical orchid bees. *Molecular Ecology* **23**, 1874–1890.

McKinney ML (1997) Extinction vulnerability and selectivity: Combining ecological and paleontological view. *Annual Review of Ecology and Systematics* **28**, 495-516.

Morreale SJ, Sullivan KL (2010) Community-level enhancements of biodiversity and ecosystem services. *Frontiers of Earth Science in China* **4**, 14-21.

Ollerton J, Winfree R, Tarrant S (2011) How many flowering plants are pollinated by animals? *Oikos* **120**, 321-326.

Packer L, Owen RE (2001) Population genetic aspects of pollinator decline. *Conservation Ecology* **5**, 1-36.

Polus E, Vandewoestijne S, Choutt J, Baguette M (2007) Tracking the effects of one century of habitat loss and fragmentation on calcareous grassland butterfly communities. *Biodiversity and Conservation* **16**, 3423-3436.

Potts SG, Biesmeijer JC, Kremen C*, et al.* (2010) Global pollinator declines: trends, impacts and drivers. *Trends In Ecology and Evolution*, 1-9.

Whitehorn PR, O'Connor S, Wackers FL, Goulson D (2012) Neonicotinoid pesticide reduces bumble bee colony growth and queen production. *Science* **336**, 351-352.

Wikelski M, Moxley J, Eaton-Mordas A*, et al.* (2010) Large-range movements of neotropical orchid bees observed via radio telemetry. *PLoS ONE* **5**, e10738.

Willmer P, Stone G (2004) Behavioral, ecological, and physiological determinants of the activity patterns of bees. *Advances in the Study of Behavior* **34**, 347-466.

Zayed A, Packer L (2002) Genetic differentiation across a behavioural boundary in a primitively eusocial bee, *Halictus poeyi* Lepeletier(Hymenoptera, Halictidae). *Insectes Sociaux* **49**, 282-288.

Zayed A, Packer L (2005) Complementary sex determination substantially increases extinction proneness of haplodiploid populations. *PNAS* **102**, 10742-10746.

CHAPTER 1

Climate, physiological tolerance, and sex-biased dispersal shape genetic structure of Neotropical orchid bees†

Margarita M. López-Uribe[1], Kelly R. Zamudio[2], Carolina F. Cardoso[3] and Bryan N. Danforth[1]

[1] Department of Entomology, Cornell University, Ithaca, NY 14853

[2] Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853

[3] Laboratório de Sistemática e Ecologia de Abelhas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil, Caixa Postal 486, 30123-970

Abstract

Understanding the impact of past climatic events on the demographic history of extant species is critical for predicting species' responses to future climate change. Paleoclimatic instability is a major mechanism of lineage diversification in taxa with low dispersal and small geographic ranges in tropical ecosystems. However, the impact of these climatic events remains questionable for the diversification of species with high levels of gene flow and large geographic distributions. In this study, we investigate the impact of Pleistocene climate change on three Neotropical orchid bee species (*Eulaema bombiformis*, *E. meriana* and *E. cingulata*) with transcontinental distributions and different physiological tolerances. We first generated ecological niche

models to identify species-specific climatically stable areas during Pleistocene climatic oscillations. Using a combination of mitochondrial and nuclear markers, we inferred calibrated phylogenies and estimated historical demographic parameters to reconstruct the phylogeographic history of each species. Our results indicate species with narrower physiological tolerance experienced less suitable habitat during glaciations and currently exhibit strong population structure in the mitochondrial genome. However, nuclear markers with low and high mutation rates show lack of association with geography. These results combined with lower migration rate estimates from the mitochondrial than the nuclear genome suggest strong male-biased dispersal. We conclude that despite large effective population sizes and capacity for long-distance dispersal, climatic instability is an important mechanism of maternal lineage diversification in orchid bees. Thus, these Neotropical pollinators are susceptible to disruption of genetic connectivity in the event of large-scale climatic changes.

Introduction

Climatic and anthropogenic landscape changes are two major threats to the persistence of biodiversity in the 21st century (Heller & Zavaleta 2009; Potts 2010). Climate change during the past 50 years has already altered the geographic distribution, developmental time and reproductive rates of many organisms (Parmesan & Yohe 2003; Bartomeus et al. 2011). Concurrently, habitat loss and agricultural intensification have fragmented and degraded natural areas, causing severe animal and plant species declines worldwide (Pimm & Kevan 2000). In the face of this biodiversity crisis (Sala et al. 2000; Forister et al. 2010), an enduring challenge for biologists is to develop a

predictive framework to guide appropriate conservation and management plans (Dawson et al. 2011). Investigating microevolutionary responses to past geologic and climatic events can provide insights into species vulnerability and predictions about the biodiversity consequences of future environmental changes.

Maintaining evolutionary processes is especially important in areas with high species diversity, such as tropical regions, that are the result of complex ecological and evolutionary histories (Myers et al. 2000). For instance, Neotropical lineages have experienced a dynamic geologic history over the past 25my, including the progressive uplift of the Andes (~23 - ~4.5 mya), the formation of the Amazon basin (~7 mya) and the closure of the Panama Isthmus (~3.5 mya) (Hoorn et al. 2010). These tectonic events during the Neogene contributed to species diversification in Neotropical lineages of frogs (Santos et al. 2009), bats (Ditchfield et al. 2000) and birds (Fjeldsa 1994) through topographic complexity and repeated vicariance/dispersal episodes. More recently over the past 2.6 mya, the geology of the Neotropics has remained relatively unchanged, but climate has been unstable (Bennett 1990). Repeated periods of cold/dry and warm/wet climate altered forest composition, driving shifts in species distributions and intraspecific lineage diversification (Carnaval et al. 2009; Mulcahy et al. 2006; Prado et al. 2012).

Patterns of responses to climatic fluctuations during Pleistocene glaciations vary among Neotropical terrestrial organisms (Hewitt 2004). Phylogeographic studies in vertebrate species with relatively small effective population sizes, low dispersal capacity, and restricted distributions often show deep mitochondrial divergences that are temporally congruent with the Pleisteocene, but spatially idiosyncratic, presumably due to differences in

species-specific ecological traits (Carnaval et al. 2009; Kieswetter & Schneider 2013; Prado et al. 2012). In contrast, the occurrence of strong phylogeographic breaks driven by climatic events in Neotropical species with large population sizes, long-distance dispersal, and widespread distributions remains unclear (Cheviron et al. 2005; Martins et al. 2009; Solomon et al. 2008). Therefore, we do not yet have consensus on expected genetic patterns, and potential diversifying mechanisms, for highly mobile terrestrial organisms in the Neotropics.

Understanding the effects of past climatic changes on species' distributions requires the investigation of their demographic responses to environmental change. For explicit hypothesis testing, it is necessary to understand spatial distributions of species in past and present times (Richards et al. 2007). The incorporation of ecological niche modeling in phylogeographic studies provides a methodological framework to identify climatically stable areas (CSAs) that served as potential 'refugia' during Quaternary climatic oscillations (Hugall et al. 2002; Carnaval et al. 2009). This approach generates species-specific spatial models with demographic predictions that can be validated with molecular data (Knowles et al. 2007). Furthermore, niche models provide information about species niche breadth and can be used as proxies for animal physiological tolerance (Bonier et al. 2007). Although niche conservatism, ecological competition, spatial variation of the environment and dispersal may be the limiting factors to species ranges, geographic distributions are largely determined by climate (Chown et al. 2010). Therefore, correlational approaches between species presence data and climate data are used to infer the limits of physiological tolerance for species distributions.

In this study, we investigate the impact of Pleistocene climatic instability on pollinating insects with great dispersal capacities and large population sizes, but varying levels of physiological tolerance to dry climatic conditions. Orchid bees (Apidae: Euglossini) are fast-flying insects endemic to the New World that exhibit an interesting courtship behavior. Male orchid bees collect volatile compounds from floral and non-floral resources throughout their lives, and present a chemical bouquet to females at display sites enabling mate recognition and choice (Eltz et al. 2005). Ecological and behavioral data indicate that dispersal is likely male-mediated in orchid bees because males have large home ranges where they search for volatiles (Wikelski et al. 2010). In contrast, females are central-place foragers that exhibit strong philopatric behavior (Augusto & Garófalo 2011). Orchid bees are remarkably abundant in low or mid-elevation forests where they comprise up to 25% of bee communities (Roubik & Hanson 2004). Many orchid bee species have a strong association to wet forested areas and show the highest community richness in forests where precipitation exceeds 2000 mm per year (Dressler 1982). Thus, orchid bees are an excellent system for investigating species responses to the dry-wet Pleistocene climatic cycles in tropical pollinating insects.

We examine the demographic history of three widespread orchid bees of the genus *Eulaema* (*E. bombiformis, E. meriana* and *E. cingulata*) using an integrative approach that includes ecological niche modeling and multi-locus population genetic data. Specifically, we test whether Pleistocene climate instability led to intraspecific diversification in three orchid bee species with different physiological tolerances. The three focal species of this study have similar widespread distributions but they display different habitat associations and elevation ranges. Both *E. bombiformis* and *E. meriana* are exclusively

found in wet forests but *E. bombiformis* is distributed from sea level up to 950 m.a.s.l., whereas *E. meriana* is found up to 1700 m.a.s.l. (Ramírez et al. 2002). *Eulaema cingulata* is present in both wet and dry lowland forests and its altitudinal distribution reaches 2600 m.a.s.l. (Ramírez et al. 2002). Therefore, we predict these species display a continuum of physiological tolerance to climate variation, with lowest in *E. bombiformis* and *E. meriana* and highest in *E. cingulata*. To test this prediction, we developed a novel method that uses response curves to environmental variables to quantify the relative difference in niche-breadth between species.

Based on the shared ecological and behavioral traits, yet differing physiological tolerances among our three focal species, we have three predictions about their responses to historical climatic instability. First, we predict that the amount of suitable habitat in refugia during dry periods of the Pleistocene was reduced for all species, however, not equally. Specifically, *E. cingulata*, which unlike the other two species now inhabits drier forests, should have experienced the largest amount of suitable habitat during the Pleistocene, as wet forests became drier. Second, we predict that species with low physiological tolerance should show strongest phylogeographic structure spatially congruent with hypothesized CSAs because lower physiological tolerance restricted gene flow outside suitable areas. Third, we predict that the spatial patterns of genetic diversity in these species are the result of isolation/colonization events during repeated cycles of forest contraction/expansion, thus are temporally congruent with Pleistocene climatic instability. The alternative scenario is that these species were widespread throughout the Neotropics pre-Pleistocene and experienced vicariance during the geologic changes of the Pliocene and contraction/expansions events

during the Pleistocene. In that case, the deepest phylogeographic breaks in species phylogenies should be temporally congruent with Pliocene geologic events. Testing these predictions has important implications for species persistence under scenarios of future climate change. We discuss the need for further investigation of these questions in species with high levels of gene flow and the value of integrative studies for future conservation plans of orchid bee species.

Materials and methods

Ecological niche modeling

We identified CSAs for the three *Eulaema* species by overlaying distribution models based on current (averaged over the past 50 years) and Last Glacial Maxima (LGM; 21kya) climate data through ecological niche models (ENM). We used the current and LGM ENMs as proxies for species-specific distributions during interglacial and glacial periods, respectively. We compiled current presence data points for the three focal species from published articles, the Discover Life Website, the Global Biodiversity Information Facility, and data points from individuals collected during our own fieldwork. In total, we collected 101, 127 and 123 location data points for *E. bombiformis*, *E. meriana* and *E. cingulata*, respectively (Fig. S1). Species distributions were modeled in Maxent v. 3.3.3 (Phillips et al. 2006, Phillips & Dudík 2008) using 19 environmental data layers at a 30-arc-second (~1 km) resolution from the WorldClim database. All bioclimatic variables were included in analyses despite correlation among them because the exclusion of correlated variables usually has little effect on ENMs (Knowles & Alvarado-Serrano 2010). We evaluated the predictive power of the model in two ways. First, we used a

binomial test based on omission and predicted areas to test if our model predicted species distributions better than random. Second, the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curve was used to measure model performance based on the sensitivity and specificity of the model. A random sample of 25% of the initial dataset over 10 replicates was used as 'testing data' to check for the spatial independence of the locality points used to build the models. We collected data from ten independent runs to verify the replicability of our models. To define binary presence-absence suitability models, we selected the Maxent threshold that minimized the difference between the omission rate and predicted area (equal training sensitivity and specificity threshold) (Pearson et al. 2004).

Assuming species niches have not changed considerably during the past 21ky, we characterized species distributions during LGM by projecting results from the current distribution models onto the same environmental layers from the LGM. We overlaid the binary predicted distributions of the current and LGM models using the raster calculator tool in ArcGIS 10 (ESRI 2011) to outline areas that have been climatically stable throughout the repeated glacial cycles of the past 1.5 My. We defined major 'refugia' as contiguous areas with a minimum size of 2000 km2 and separated by a minimum linear distance of 300 km, a distance 10-times greater than the maximum home range distance reported for orchid bees (Janzen 1971).

Niche-breadth quantification

We used response curves to all bioclimatic variables to calculate the relative ecological niche breadth for our focal species. Response curves depict the range of species' predicted suitable conditions in ecological space based on individual responses to one bioclimatic variable (Phillips & Dudík 2008).

Response curves can provide information on what niche space a species can tolerate but they are uninformative about what conditions species cannot tolerate because ranges may also be restricted by biotic factors (Chown et al. 2010). Therefore, we developed the Relative Niche Breadth (*RNB*) index to serve as a proxy for the species-specific breadth of physiological tolerance and quantify the relative difference in tolerance between species. We calculated this index as follows,

$$RNB = \frac{\sum_{j}^{m} \Delta_{ij}}{\max\left\{\sum_{j}^{m} \Delta_{ij}\right\}_{i=1}^{j=n}} \quad (1.1)$$

where D_{ij} is the difference between the minimum and the maximum value of the environmental variable *j* in species *i* for which the probability of presence is at least 0.5 and $\max\left\{\sum_{j}^{m} \Delta_{ij}\right\}_{i=1}^{j=n}$ is the maximum sum of all D_{ij} or all environmental variables (*m*) among all species (*n*) compared (Fig. S2). The rationale for this index as a proxy physiological tolerance breadth is that the greater the difference between the minimum and maximum probability of species' presence for a given environmental variable, the wider the range of environmental conditions that a species can tolerate. The 0.5 probability value is used as a benchmark for the delta calculation because it is a standard cut-off threshold for the probability of presence of a species (Trevan 1927). This cut-off also avoids the potential ambiguity of making measurements from the extreme of the curves where values tend to converge.

Sampling

We collected a total of 454 male individuals from throughout the geographic distribution of *E. bombiformis* (113), *E. meriana* (201), and *E. cingulata* (137) from 14, 16 and 16 localities, respectively (Fig. 1). In the field, we attracted

male orchid bees using five chemical baits (cineol, methyl salicylate, vanillin, eugenol and benzyl acetate) from several sites at each locality to avoid sampling related individuals. Male bees were collected using entomological nets and preserved in 95% ethanol. Geographic positioning was recorded for each locality using a Garmin GPSmap 76CSx unit. Front legs from all bees were removed for DNA extraction before individuals were mounted on entomological pins, labeled, and accessioned in the Cornell University Insect Collection (CUIC). Field collections were supplemented with museum specimens from localities in Guatemala, Bolivia, Colombia, French Guiana and Brazil (Appendix Table S1.1).

Genetic data collection

Leg tissue was ground in liquid nitrogen and digested overnight in 1µg/µL of Proteinase K. DNA was extracted using the DNeasy blood and tissue extraction kit (QIAGEN) for field collected tissues, or a standard phenol-chloroform DNA extraction protocol in the case of museum specimens (Danforth 1999). We sequenced a total of ~3750 base pairs from seven genes: two mitochondrial genes (COI: Cytochrome oxidase I, and Cytb: Cytochrome b), two nuclear protein coding genes (CAD: Carbamoyl-phosphate synthetase 2, Aspartate transcarbamylase, and Dihydroorotase; and EF1-α: Elongation Factor 1 alpha) and three anonymous single copy nuclear loci (ASCN) that included a microsatellite region (EM8, EM70 and EM106) (Table 1.1). We chose these markers because they all showed clear sequencing results and comprise a group of genes with a wide range of mutation rates. All mitochondrial and nuclear coding genes were sequenced in both directions, while the ASCN loci were sequenced only in the forward strand using a 24bp

Table 1.1. Summary of PCR primers and genetic diversity indices per locus for *Eulaema bombiformis* (Ebom), *E. meriana* (Emer) and *E. cingulata* (Ecin). N_P, number of polymorphic sites; N_S, number of singletons; N_A, number of alleles.

Locus	Primers	Species	Length	N_P	N_S	N_A	Reference
COI	F: 5'-CAA CAT TTA TTT TGA TTT TTT GG-3' R: 5'-TCC AAT GCA CTA ATC TGC CAT ATT A-3'	Ebom	838	51	15	38	Simon *et al.* 1994
		Emer	847	53	12	59	
		Ecin	835 – 838	23	9	25	
Cytb	F: 5'-TAT GTA CTA CCA TGA GGA CAA ATA TC-3' R: 5'-ATT ACA CCT CCT AAT TTA TTA GGA AT-3'	Ebom	433	18	4	24	Crozier *et al.* 1991
		Emer	433	39	11	45	
		Ecin	433	18	5	26	
CAD	F: 5'-GGW TAT CCC GTD ATG GCB MGW GC-3' R: 5'-GCA THA CYT CHC CCA CRC TYT TC-3'	Ebom	673	40	16	63	Danforth *et al.* 2006
		Emer	673	17	6	69	
		Ecin	673	46	30	69	
EF1α	F: 5'-G GGY AAA GGW TCC TTC AAR TAT GC-3' R: 5'-A ATC AGC AGC ACC TTT AGG TGG -3'	Ebom	991 – 994	33	26	10	Danforth & Ji 1998
		Emer	991 – 996	17	14	10	
		Ecin	991 – 994	29	25	13	
EM8	F: 5'-CAG CGT CGC GAT TGG TTC TAC A-3' R: 5'-TCA GCT TTG TCA CCG GCA CTG T-3'	Ebom	210 – 238	5	5	6	López-Uribe *et al.* 2011
		Emer	260 – 282	12	3	13	
		Ecin	254 – 268	10	6	9	
EM70	F: 5'-GTA CCA CTG CGA GAG CGA AGA AAA-3' R: 5'-CCA GTG GCC CGA AGT AGA AAC A-3'	Ebom	229 – 240	13	9	12	López-Uribe *et al.* 2011
		Emer	231 – 254	15	8	11	
		Ecin	231 – 251	27	15	33	
EM106	F: 5'-GAC GTG GAT GAG CCG CAG AAG AC-3' R: 5'-TCC GAC GAT GTA CGA GCA CGA A-3'	Ebom	207 – 234	15	1	24	López-Uribe *et al.* 2011
		Emer	197 – 257	6	1	11	
		Ecin	219 – 267	12	7	12	

tag added to the forward primers (López-Uribe et al. 2011). Fragments were sequenced using the Big Dye terminator kit (version 3.1) (Applied Biosystems) on an ABI 3730 capillary sequencer.

Genetic diversity

Haploid male sequences were assembled and individually edited in Sequencher v.5.0 (DNASTAR). Haplotype sequences were deposited in GenBank (accession numbers KF895558-KF970454; Appendix Table S1.1). We used Clustal W to generate a multiple sequence alignment in MegAlign v.8.0.2 (DNASTAR), which was then manually edited in MacClade v.4.08 (Maddison & Maddison 2005). We calculated the number of polymorphic sites (NP), number of singletons (NS) and number of alleles (NA) of each gene per species in the software DNAsp v.5 (Librado & Rozas 2009). We also estimated the average number of nucleotide differences (k), nucleotide diversity (π) and haplotype diversity (h) of the populations grouped by the CSAs identified based on ENMs. Because mitochondrial recombination is possible (Galtier et al 2009), we tested for recombination in each protein-coding gene and in the concatenated mitochondrial dataset using Sawyer's statistical test, maximum chi-squared, sister scanning method and automated bootscanning as implemented in the software RDP (Martin et al. 2005).

Phylogenetic inference

We inferred the best-fit models of molecular evolution for the concatenated dataset of protein-coding genes using jModelTest v.0.1.1 (Posada 2008) based on the corrected Akaike Information Criterion (AICc). Bayesian analyses were performed separately for each species in MrBayes v.3.1.2 (Huelsenbeck & Ronquist 2001) using 2 independent runs each with three heated and one cold Markov chains sampling every 1000 generations. Twenty

million generations were sufficient to ensure the average standard deviation of split frequencies was below 0.01. Stationarity of likelihood scores and parameter convergence was assessed using Tracer v. 1.4. Posterior probabilities were estimated from the distribution of trees after discarding the first 20000 as burn-in. We used sequences from the other two *Eulaema* spp. included in this study as outgroups for phylogenetic reconstructions.

A time-calibrated phylogeny of the mitochondrial dataset was estimated in BEAST 1.6.2 (Drummond & Rambaut 2007) under the coalescent model of expansion growth. We estimated the time to the most recent common ancestor (tMRCA) using the COI substitution rate of 1.2% - 1.5% per million years calibrated for other insects (Farrell 2001). Node ages were estimated using a strict clock, which is appropriate for intraspecific data analyzed under a coalescent model. We performed two independent runs for 20 million generations sampling every 1000 generations. Convergence was assessed by examining run log files in Tracer v. 1.4 and confirming that effective sampling size (ESS) for parameters were greater than 200. To reconstruct the maximum clade credibility tree, we discarded the first 10% of the sampled trees as burn-in.

Haplotype networks

Mitochondrial and nuclear protein coding haplotypes were also visualized with a neighbor-net network constructed in SplitsTree4 v. 4.12.3 (Huson & Bryant 2006). We used the Tamura-Nei distance to estimate the net sequence divergence between major phylogeographic groups in the program DNAsp v.5 (Librado & Rozas 2009). To detect spatial structuring in the mitochondrial vs. nuclear data, we used the global and local tests of the spatial principal component analysis (sPCA) in the R package ADAGENET v.1.3-9.2 (Jombart

et al. 2008; R Core Team 2013). sPCA uses allele frequencies to investigate patterns of genetic variability in a spatially explicit context. This test differentiates between global structures, that indicate presence of coarse-scale spatial patterns (e.g. clines, patches), and local structures, that detect differentiation between neighboring sites. This approach is appropriate for detecting spatial patterns unidentifiable with tree-based methods (Jombart et al. 2008).

Population structure

We tested phylogeographic hypotheses based on the CSAs identified for each species (Table 1.2) using two different methods. In all cases, mitochondrial sequence data and nuclear haplotype data were analyzed separately due to potential for incongruent phylogeographic signals from the different markers. First, we used full and partial Mantel tests, as implemented in IBDWB v.3.23 (Jensen et al. 2005), to investigate patterns of isolation-by-distance (linear geographic distance) and isolation-by-barrier (matrix between hypothesized CSA) as predictors of the level of genetic differentiation between populations. Inter-population differentiation was tested using Rousset's distance method (PhiST/(1-PhiST)). Second, we implemented an analysis of molecular variance (AMOVA) in Arlequin v. 3.5.1.2 to calculate the percentage of genetic variation explained by populations grouped based on the identified CSAs (Excoffier & Lischer 2010).

Demographic inference

We calculated Fu's F in Arlequin v. 3.5.1.2 (Excoffier & Lischer 2010) to detect departures from neutrality or constant population size. The demographic history of each phylogeographic lineage was reconstructed using an Extended

Table 1.2. Summary table showing population assignment to CSA and sample sizes per population.

Species	CSA	Population	Sample size
E. bombiformis	CA	CRIC	7
		CRIS	1
		PANI	12
		PANP	1
	CR	COLQ	10
		ECUW	5
	NA	BRAA	32
		COLE	2
		ECUE	3
		GUFA	1
		PERN	7
		PERS	9
	NAF	BAFN	18
	SAF	BAFS	5
E. meriana	CA	CRIC	47
		CRIS	11
		PANI	16
		PANP	4
	CR	COLQ	17
		COLW	4
		ECUW	26
	WA + CAm	BOLA	3
		ECUE	9
		COLE	4
		PERN	12
		PERS	13
	NA	BRAA	16
		GUFA	3
	AF	BAFN	10
E. cingulata	CA	CRIC	5
		CRIN	5
		CRIS	2
		GUAC	1
		PANI	13
		PANP	3
	CR	COLN	9
		COLQ	11
		COLW	4
		ECUW	11

NA	BRAA	16
	COLB	7
	GUFA	2
SA	BOLA	8
	ECUE	15
	PERN	14
	PERS	5

Bayesian Skyline Plot (EBSP) as implemented in BEAST 1.6.2 (Drummond & Rambaut 2007). All nuclear and mitochondrial markers were pooled in this analysis and the haplodiploid effective population size of nuclear markers was incorporated in our model as x-linked genes. To test for sex-biased dispersal, we estimated migration rates between CSAs for nuclear and mitochondrial data independently by fitting a model of isolation-with-migration as implemented in the software IMa2 (Hey and Nielsen 2007). Greater migration rates in the nuclear than in the mitochondrial DNA would support the male-biased dispersal hypothesis. For loci EM8 and EM70, information about the microsatellite and flanking regions was provided separately. Locus EM106 was not included because the flanking region violated the infinite sites mutation model. We first simultaneously analyzed all populations of *E. bombiformis* and *E. meriana*. However, due to the computational challenges of estimating a high number of parameters from multiple populations, the IMa2 analyses only reached convergence for two pairs of sister populations from neighboring CSAs in *E. bombiformis* (pair 1: “CR2”and CA; pair 2: NA2+NA3 and ”AF”). Final analyses consisted of two runs of 10-40 geometrically heated chains with a burnin 100k steps and sampled for 20000k steps. We report population migration rate calculated as the mutation-scaled migration rate multiplied by population size.

Results

Niche modeling

The Maxent ENMs accurately predicted the current distribution of the three species (AUC_{Ebom}=0.983±0.001; AUC_{Emer}=0.989±0.002; AUC_{Ecin}=0.968±0.013) (Fig 1.1A-C). However, we identified some areas of under-prediction likely due to limited locality presence records from these regions (Fig 1.1A-C). Performance of the test and training datasets were similar, indicating that the data points used for model building were not spatially autocorrelated. The permutation importance test indicated that temperature seasonality was one of the variables contributing the most to model gain in all three species. Mean temperature of the coldest quarter was important for *E. bombiformis* and *E. meriana*; and precipitation of the coldest quarter was important for *E. meriana*.

For *E. bombiformis*, five major CSAs were identified (Fig. 1.2A) with two major distribution shifts during LGM: (1) a reduction of the suitable habitat in the Amazon basin to one large suitable area located on the Northern Amazon and the foothills of Guiana highlands (NA) and (2) a subdivision of the Atlantic forest into two smaller refugia (NAF, SAF) spatially congruent with earlier models of Pleistocene refugia (Carnaval & Bates 2007; Carnaval et al. 2009). The distribution of *E. meriana* was more fragmented during LGM and six major CSAs were identified (Fig. 1.2B). For *E. meriana*, the suitable habitat in the Atlantic forest was reduced to one small area restricted to the coast of Pernambuco (AF). The Amazon Basin was reduced to small pockets of forest located to the west and center of the Amazon (WA, CAm) and the

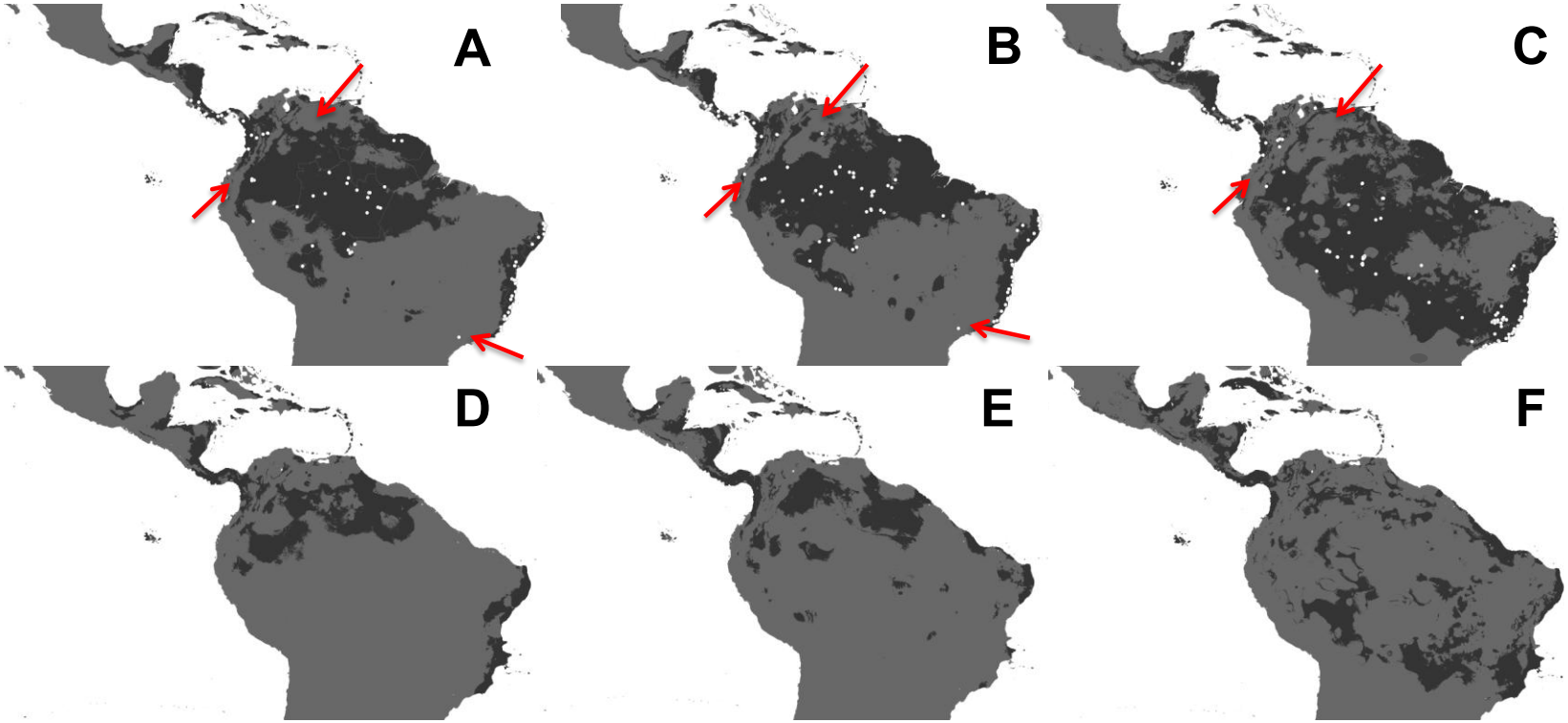


Figure 1.1. Habitat suitability binary models and sampling localities used for niche modeling analysis for *E. bombiformis* (A and D), *E. meriana* (B and E) and *E. cingulata* (C and F) based on current (A-C) and Last Glacial Maxima (D-F) environmental data. Red arrows indicate areas of under-prediction.

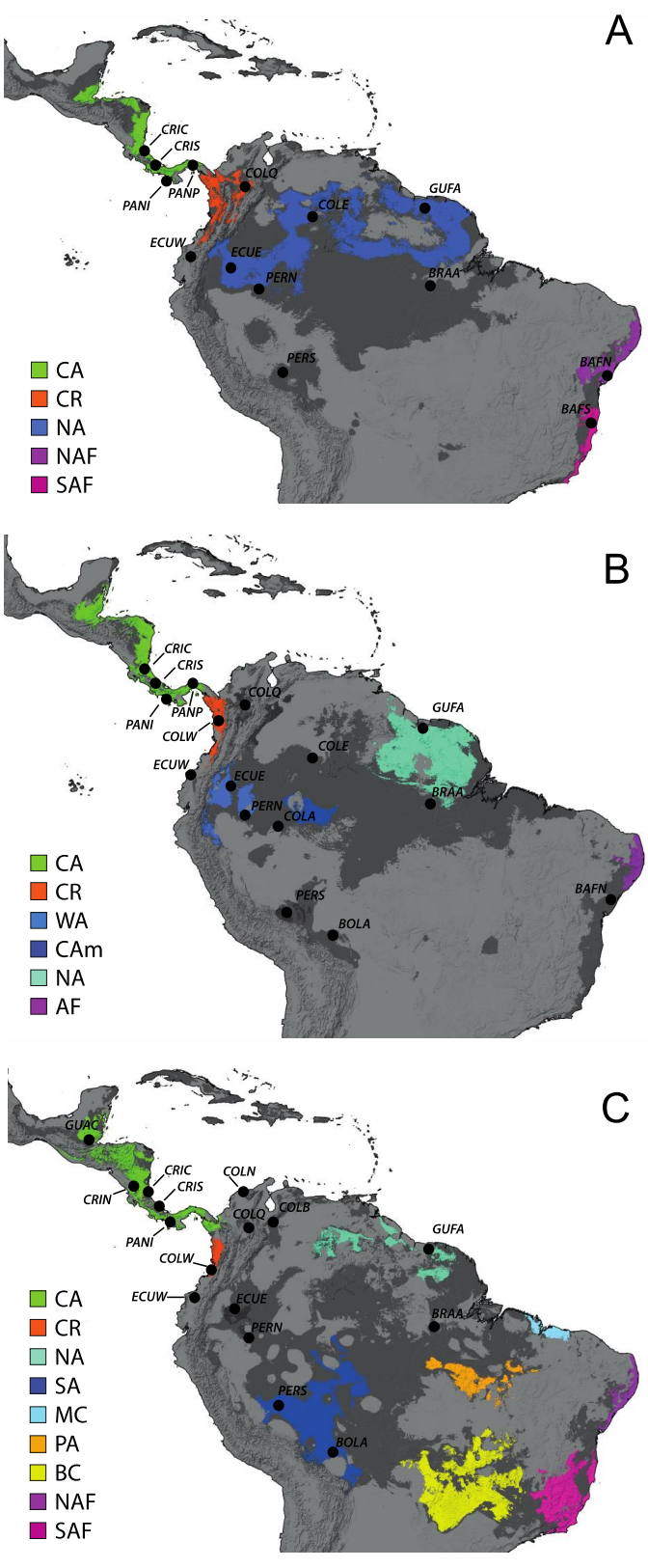


Figure 1.2. Climatically stable areas (CSA) and distribution of sampling localities for *Eulaema* bombiformis (A), *E. meriana* (B) and *E. cingulata* (C). CSA are the result of overlapping current and LGM climatic suitability models for each species. Colored areas indicate contiguous CSA: Central America (CA), Chocó Region (CR), Northern Amazon (NA), Northern Atlantic Forest (NAF), Southern Atlantic Forest (SAF), West Amazon (WA), Central Amazon (CAm), Southern Amazon (SA), Brazilian Cerrado (BC), Para Amazon (PA) and Maranhão Coast (MC). Shaded grey areas indicate suitable areas for the three species under current climatic conditions.

foothills of Guiana highlands (GF). Interestingly, the paleomodel for this species revealed that Central America (CA) and Chocó (CR), which are currently separated by the Darien gap, were fully connected by a stretch of suitable habitat that extended from the Atlantic to the Pacific coast along the Isthmus of Panama (Fig. S1D). For *E. cingulata*, we identified eight CSAs with the largest areas located in Central America and the southern part of the current distribution of this species (Fig. 1C). In contrast to *E. bombiformis* and *E. meriana*, CSAs for *E. cingulata* included current dry forests and savanna areas, such as the Brazilian Cerrado (BC).

Niche-breadth index

RNB indices showed that the species with the broadest niche breadth is *E. cingulata* (RNB=1; $\Sigma\Delta_{Ecin}$=54131) followed by *E. bombiformis* (RNB=0.34; $\Sigma\Delta_{Ebom}$=18396) and *E. meriana* (RNB=0.33; $\Sigma\Delta_{Emer}$=17748) (Table 1.3). More specifically, *E. meriana* and *E. bombiformis* showed one-third the niche breadth of *E. cingulata*. This difference is in accordance with the exclusive association of *E. meriana* and *E. bombiformis* to wet forests, while *E. cingulata* inhabits both wet and dry forests (Ramírez et al. 2002). Response curves to environmental variables revealed that *E. bombiformis* has broader physiological tolerance than *E. meriana* based on 16 out of the 19 climatic variables. However, *E. meriana* has a wider physiological tolerance to colder temperatures (ΔBIO6 -Table 1.2) suggesting that this physiological trait may explain its greater altitudinal range. Overall, paleomodels indicate that suitable habitat was reduced during the dry period of the Pleistocene for all three species, especially in the Amazon Basin (Fig. 1.1D–F). As expected, the species with the broadest physiological tolerance (*E. cingulata*) had the most widespread distribution during dry climatic periods of the Pleistocene despite

Table 1.3. Calculation of the relative niche breadth index (*RNB*) for *E. bombiformis* (Ebom), *E. meriana* (Emer) and *E. cingulata* (Ecin).

	Bioclimatic variable	Ebom	Emer	Ecin
ΔBIO1	Annual Mean Temperature	38	27	43.6
ΔBIO2	Mean Diurnal Range (Mean monthly (max tº - min tº))	27.9	14.48	47.5
ΔBIO3	Isothermality (BIO2/BIO7) (* 100)	34	32.06	30.97
ΔBIO4	Temperature Seasonality (SD *100)	754.1	551.6	1030.4
ΔBIO5	Max Temperature of Warmest Month	45.2	39.9	58.8
ΔBIO6	Min Temperature of Coldest Month	65.8	143.6	157.7
ΔBIO7	Temperature Annual Range (BIO5-BIO6)	39.44	33.13	65.46
ΔBIO8	Mean Temperature of Wettest Quarter	35.2	26.1	38.6
ΔBIO9	Mean Temperature of Driest Quarter	38.7	35.2	55.9
ΔBIO10	Mean Temperature of Warmest Quarter	32.5	21.6	35.3
ΔBIO11	Mean Temperature of Coldest Quarter	136.1	121.5	144.4
ΔBIO12	Annual Precipitation	6235	6225	7094
ΔBIO13	Precipitation of Wettest Month	1379.7	1132.3	1399
ΔBIO14	Precipitation of Driest Month	421.19	412.18	462.5
ΔBIO15	Precipitation Seasonality (Coefficient of Variation)	36.39	40.21	34.63
ΔBIO16	Precipitation of Wettest Quarter	3007.9	3038.5	37276.3
ΔBIO17	Precipitation of Driest Quarter	1337.8	1316.4	1546.5
ΔBIO18	Precipitation of Warmest Quarter	1978.4	1836.8	1766.9
ΔBIO19	Precipitation of Coldest Quarter	2753.1	2700.7	2843
Σ		**18396.42**	**17748.26**	**54131.46**
RNB		**0.34**	**0.33**	**1.00**

experiencing the most severe reduction in overall suitable habitat (47%). Reduction of suitable habitat was 30% and 40% for *E. bombiformis* and *E. meriana*, respectively.

Patterns of genetic diversity

The degree of mitochondrial genetic diversity within CSAs was concordant with predictions based on the age of each phylogeographic group (see below). Generally, geographic areas with older lineages showed greater levels of genetic diversity. Geographic regions that showed the highest mitochondrial genetic diversity were the Northern Amazon (NA) in *E. bombiformis* (=0.00751), Chocó Region (CR) in *E. meriana* (π_{mt}=0.00982) and Northern Amazon (NA) in *E. cingulata* (=0.00472) (Table 1.4). Average nucleotide diversity was higher for mitochondrial than for nuclear genes in *E. meriana* (=0.00548; =0.00286), but similar in *E. bombiformis* (=0.00454; =0.00423) and *E. cingulata* (=0.00337; =0.00312). We found no evidence of recombination for the mitochondrial or nuclear genes using four different recombination detection tests.

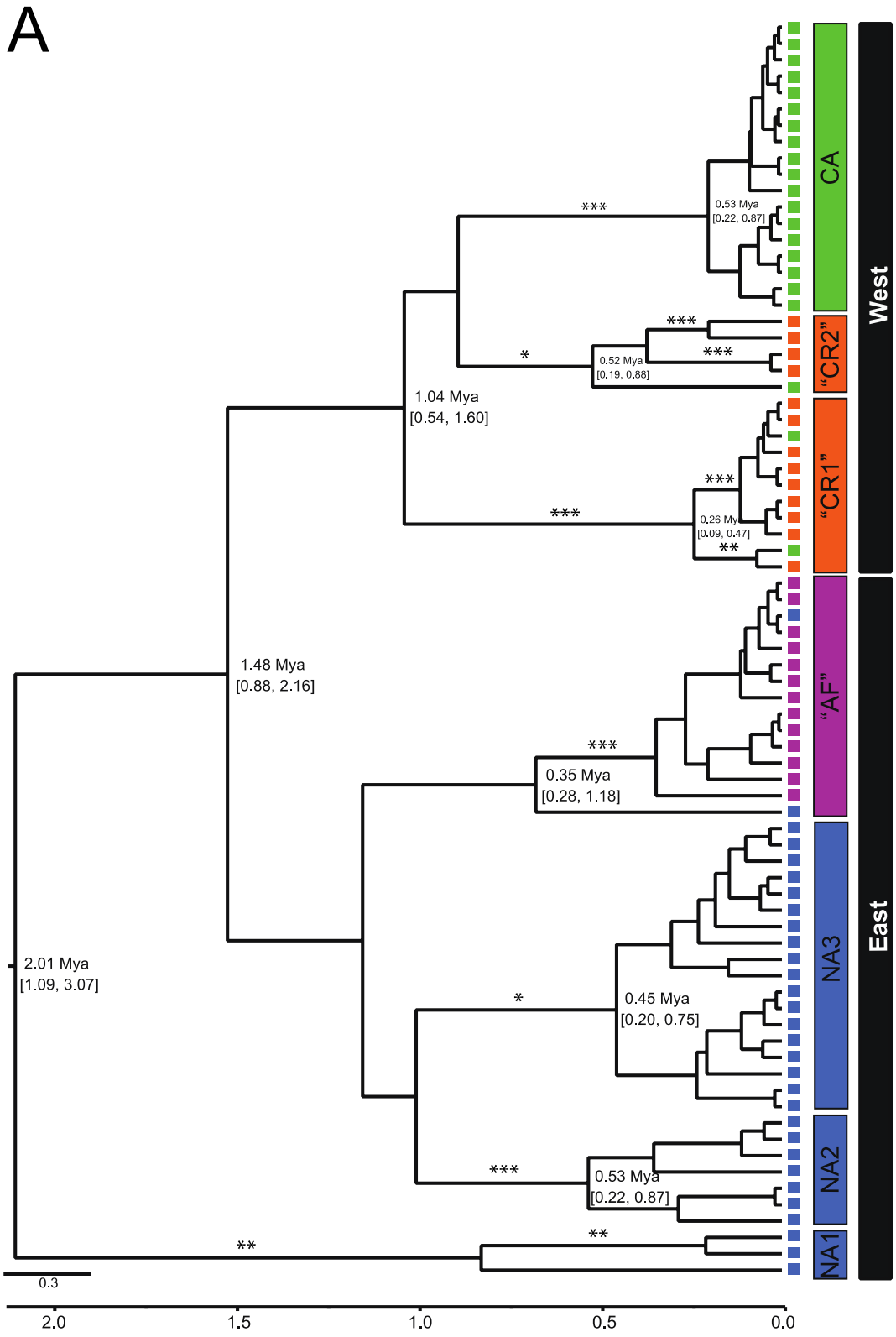
Time calibrated phylogenies

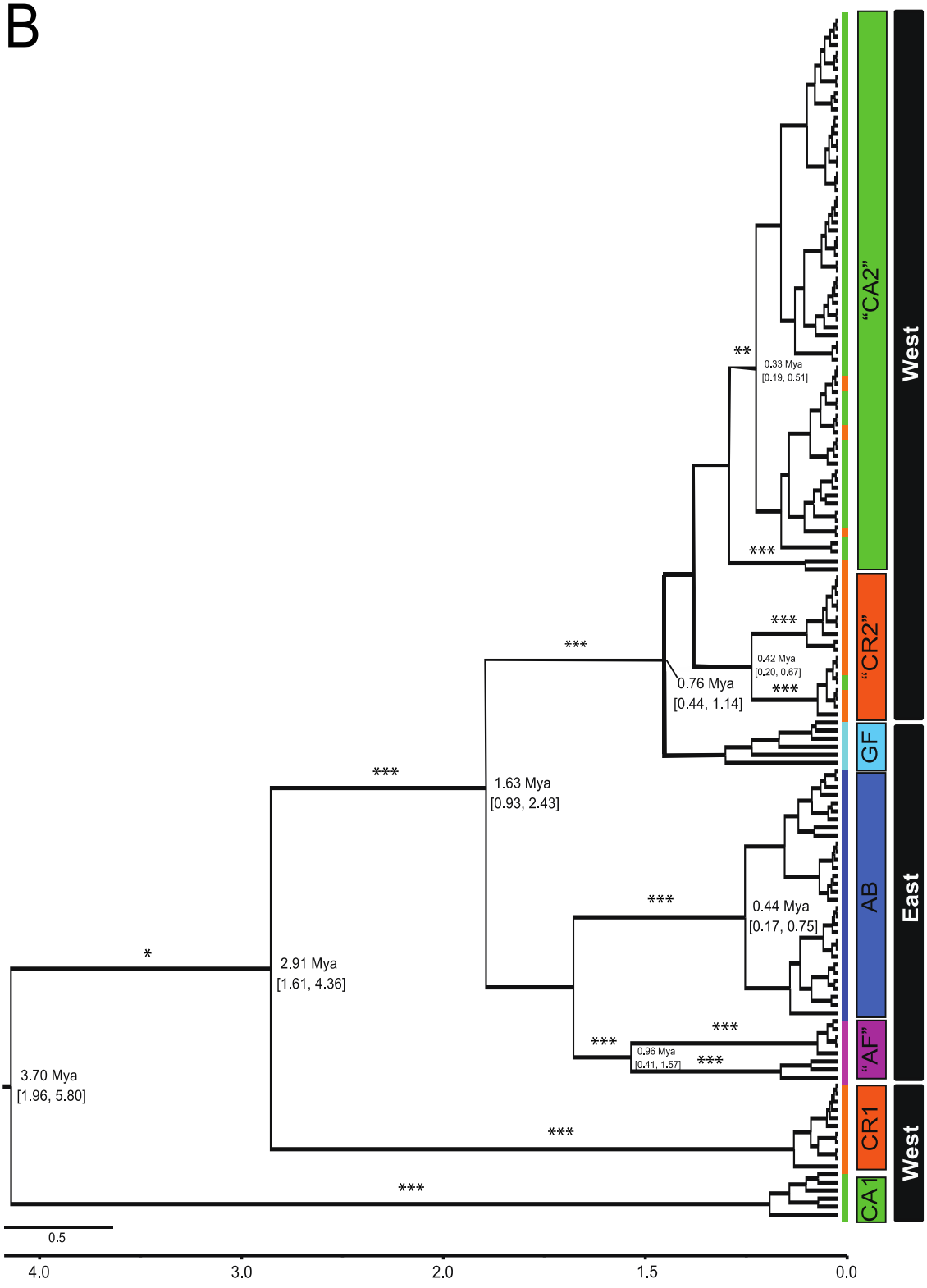
Bayesian tree topologies of concatenated mtDNA phylogenies were congruent between the MrBayes and BEAST analyses. Mitochondrial Bayesian phylogenies recover strong phylogeographic structure for *E. bombiformis* and *E. meriana*, the two focal species with narrower physiological tolerances, but not for *E. cingulata* (Fig. 1.3). Furthermore, the majority of the lineages in *E. bombiformis* and *E. meriana* were spatially congruent with predicted CSAs and date to the Pleistocene (Fig. 1.3). For *E. bombiformis*, we analyzed both mitochondrial genes concatenated under the HKY+I+G model. The maximum clade credibility (MCC) tree recovered one old lineage from the Amazon basin

Table 1.4. Genetic diversity indices for the mitochondrial (mtDNA) and nuclear (nuDNA) protein coding genes of populations grouped by the closest CSA in *E. bombiformis* (Ebom), *E. meriana* (Emer) and *E. cingulata* (Ecin). Average nucleotide differences (*k*), nucleotide diversity (π), and haplotype diversity (*h*).

Species	CSA	mtDNA				nuDNA			
		n	***k***	**π**	***h***	**n**	***k***	**π**	***h***
Eulaema bombiformis	CA	19	3.520	0.00277	0.708	10	7.667	0.00330	1
	CR	10	6.578	0.00518	0.956	8	13.071	0.00563	1
	NA	24	9.543	0.00751	0.989	18	10.261	0.00444	1
	AF*	10	3.444	0.00271	0.911	9	8.194	0.00354	1
Eulaema meriana	CA	67	6.316	0.00493	0.95	49	6.276	0.002669	0.999
	CR	31	12.568	0.00982	0.916	29	7.320	0.003082	1
	Am**	31	3.596	0.00290	0.94	38	6.434	0.002715	1
	GF	5	6.2	0.00484	1	6	6.667	0.002805	1
	AF	7	6.286	0.00491	0.857	4	7.167	0.003011	1
Eulaema cingulata	CA	62	4.303	0.00339	0.887	44	8.309	0.00351	1
	NA	9	5.99	0.00472	0.989	4	7.5	0.00314	1
	WA	25	3.94	0.00311	0.903	16	7.283	0.00306	0.992
	SA	12	2.879	0.00227	0.955	9	6.556	0.00275	1

* AF groups populations from NAF and SAF; ** Am groups populations from WA and CAm.


A
CA
"CR2"
"CR1"
"AF"
NA3
NA2
NA1
West
East
0.53 Mya
[0.22, 0.87]
0.52 Mya
[0.19, 0.88]
1.04 Mya
[0.54, 1.60]
0.26 Mya
[0.09, 0.47]
1.48 Mya
[0.88, 2.16]
0.35 Mya
[0.28, 1.18]
2.01 Mya
[1.09, 3.07]
0.45 Mya
[0.20, 0.75]
0.53 Mya
[0.22, 0.87]
0.3
2.0
1.5
1.0
0.5
0.0


B
3.70 Mya
[1.96, 5.80]
2.91 Mya
[1.61, 4.36]
1.63 Mya
[0.93, 2.43]
0.76 Mya
[0.44, 1.14]
0.44 Mya
[0.17, 0.75]
0.42 Mya
[0.20, 0.67]
0.33 Mya
[0.19, 0.51]
0.96 Mya
[0.41, 1.57]
*
**
***
0.5
4.0
3.0
2.0
1.5
0.0
"CA2"
"CR2"
GF
AB
"AF"
CR1
CA1
West
East
West

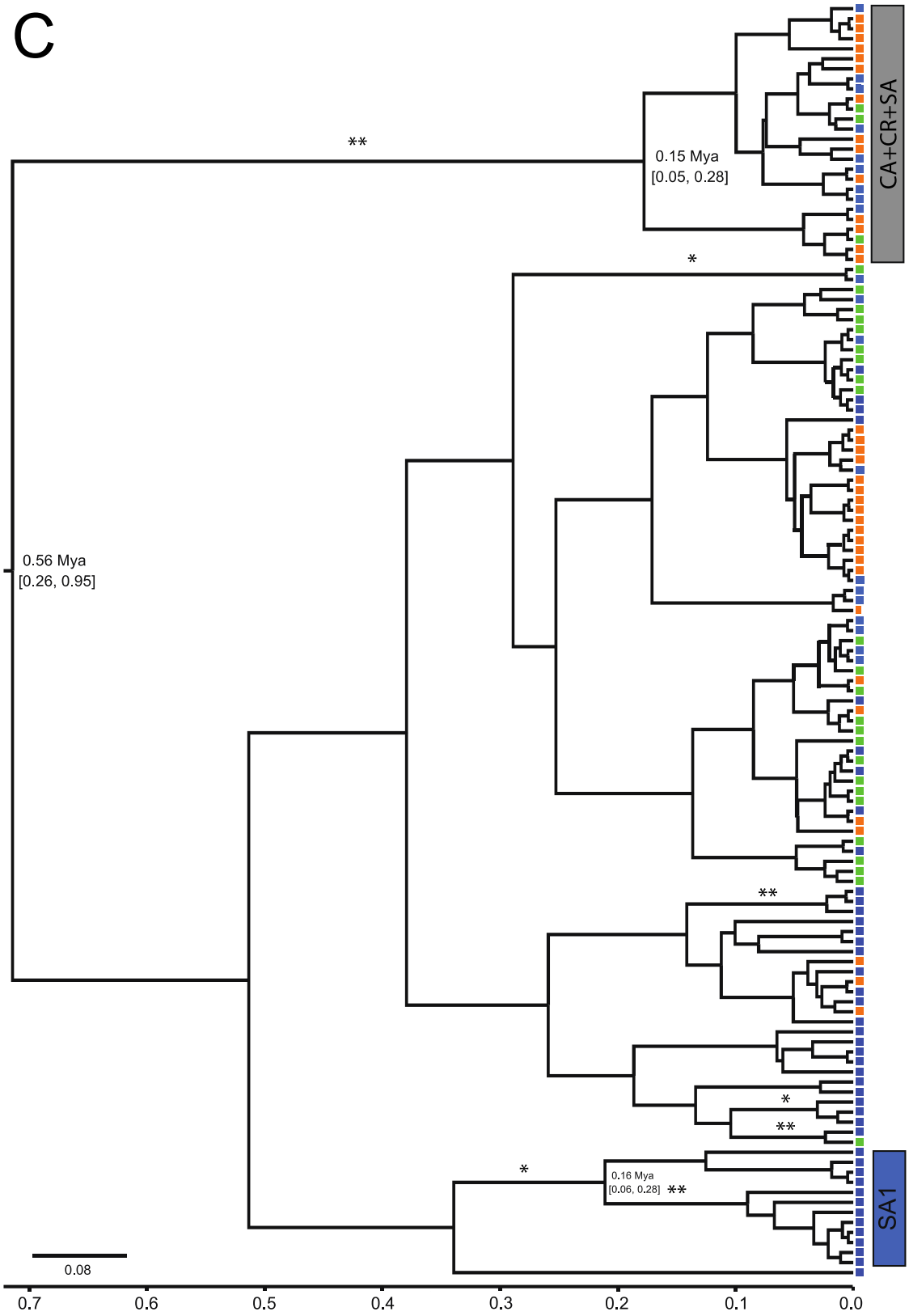

C
CA+CR+SA
SA1
**
0.15 Mya
[0.05, 0.28]
*
0.56 Mya
[0.26, 0.95]
**
*
**
*
0.16 Mya
[0.06, 0.28] **
0.08
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

Figure 1.3. Calibrated mitochondrial phylogenies for *E. bombiformis* (A)*, E. meriana* (B) and *E. cingulata* (C). Asterisks represent Bayesian posterior probability branch supports (*>70%, **>80%, ***>90%). Bayesian estimates for the time to the most recent common ancestor (rMRCA) in million years (my) of crown groups are shown in Table 3. Colors correspond to the CSAs identified from the ecological niche models (Fig. 1).

(Am1) and two reciprocally monophyletic groups (~1.5 Mya; 95% highest posterior density [HPD]= 0.88, 2.16) spatially concordant with an east-west split due to the presence of the Andes cordillera (Fig. 1.3A). Within the east clade, there are two monophyletic lineages from the Amazon Basin (Am2, Am3) and one lineage that groups all the Atlantic Forest haplotypes and two individuals from the Amazon basin ("AF") (Fig. 1.3A). Within the west clade, samples from Central America (CA) are monophyletic (~0.2 Mya; HPD= 0.09, 0.39) while individuals from the Chocó Region ("CR1", "CR2") form two highly supported clades that include some individuals from Central America.

The MCC tree for *E. meriana*, based on the GTR+I+G model, revealed two old lineages that diverged early in the history of this species: one from Central America (CA1) (~4.3 Mya; HPD= 1.96, 5.80) and one from Chocó region (CR1) (~2.9 Mya; HPD= 1.61, 4.36) (Fig. 1.3B). These early-diverged lineages are sister to a clade (~1.6 Mya; HPD= 0.93, 2.43) that comprises two lineages: one from the Amazon Basin (Am) (~0.44 Mya; HPD= 0.17, 0.75) + Atlantic Forest ("AF") (~0.96 Mya; HPD= 0.41, 1.57) sister to a monophyletic clade comprising individuals from Guiana's foothills (GF), Chocó Region ("CR2") (~0.42 Mya; HPD= 0.20, 0.67) and Central America ("CA2") (~0.33 Mya; HPD= 0.19, 0.51).

The Bayesian phylogenetic trees for *E. cingulata* were built using the HKY+I molecular evolution model. Unlike the patterns found in *E. bombiformis*

and *E. meriana*, the mitochondrial dataset recovered no geographically structured lineages in *E. cingulata* (Fig. 1.3C). The only obvious features of the tree topology are the presence of: (1) only one highly supported phylogenetic group with mixed haplotypes from Central America, Chocó Region and Amazon basin and (2) a monophyletic lineage that includes haplotypes from the Southern Amazon basin refugia (Am).

The Bayesian phylogenetic reconstruction from the protein-coding nuclear genes did not recover phylogeographic signal for any of the species (not shown). Identical nuclear haplotypes were found in individuals from Costa Rica and Bolivia located ~4000 km apart. Although not informative for phylogenetic inference, the haplotype genetic data from the nuclear loci were used for the population and phylogeographic level analyses (see below).

Average mitochondrial nucleotide divergence between phylogeographic groups was 1.13% for *E. bombiformis*, 0.96% for *E. meriana* and 0.39% for *E. cingulata*. Average nuclear nucleotide divergence was lower than mitochondrial divergence in *E. bombiformis* and *E. meriana* (0.53% and 0.3%, respectively) but similar in *E. cingulata* (0.31%).

Haplotype networks and microsatellite trees

Neighbor-net haplotype networks for the mitochondrial data strongly supported the spatial grouping based on the ENMs for *E. bombiformis* and *E. meriana* (Fig. 1.4A-B). The sPCA global test indicated highly significant structure for all species in the mitochondrial dataset ($p_{Ebom}<0.001$; $p_{Emer}<0.001$; $p_{Ecin}<0.001$). These results reveal cryptic genetic structure that is not explained by the hypothesized CSAs in *E. cingulata*. The nuclear haplotype network data for

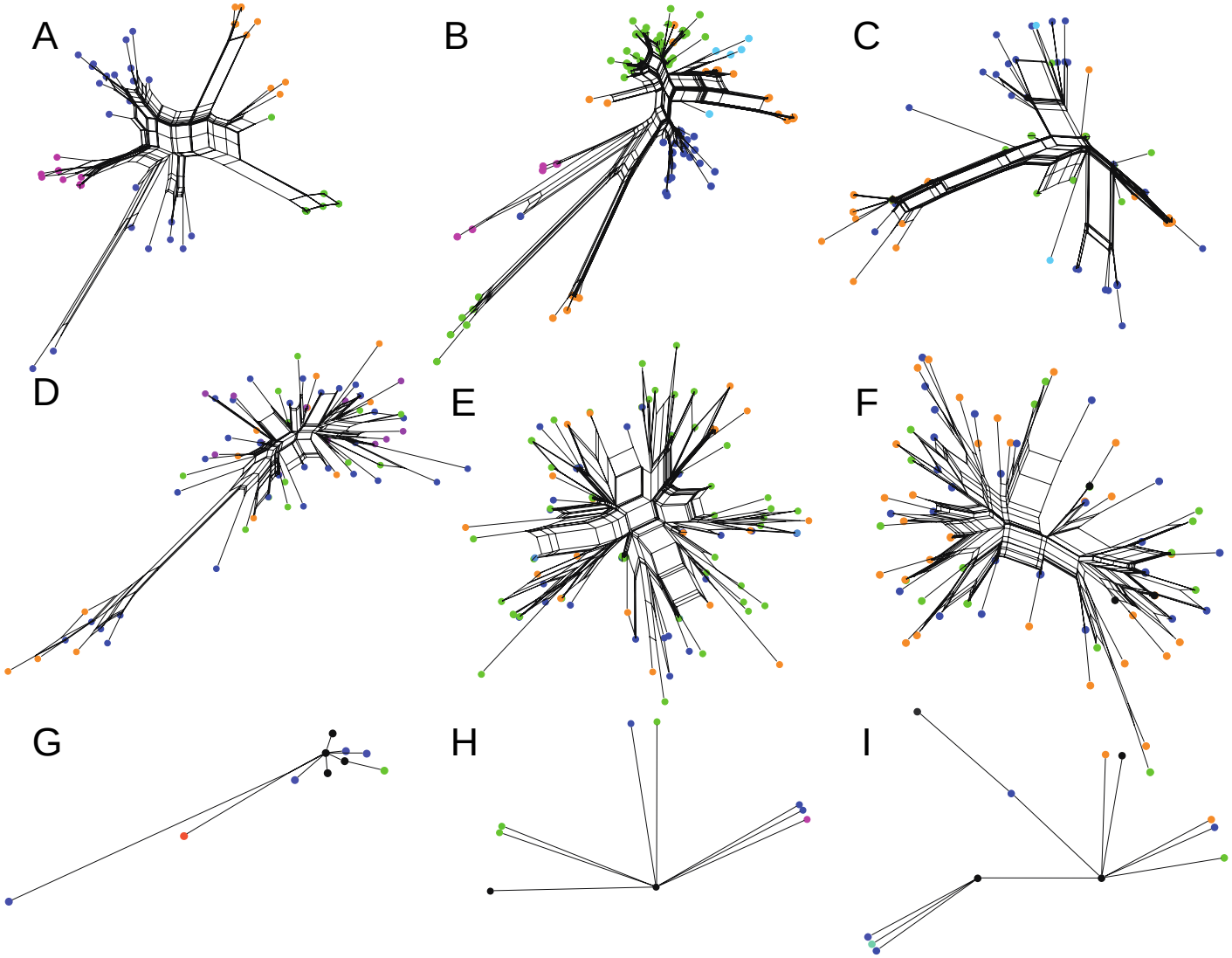

A
B
C
D
E
F
G
H
I

Figure 1.4. Multi-locus mitochondrial (A-C), CAD (D-F) and EF1α(G-I) networks under the neighbor-net algorithm for *E. bombiformis* (A, D, G), *E. meriana* (B ,E, H) and *E. cingulata* (C, F, I). Haplotypes are colored by the closest CSAs (Figure 1.2, Table 1.2) from their geographic origin. Black dots indicate haplotypes that were found in multiple CSAs.

CAD and EF1α showed a random distribution of haplotypes based on their geographic origin (Fig. 1.4D-I). These results supported by the non-significant sPCA global test results for all species (p_{Ebom}=0.131; p_{Emer}=0.208, pEcin=0.347). Local structure was detected only for *E. meriana* in both the mitochondrial and nuclear data (mt: p_{Ebom}=1; p_{Emer}=0.026, p_{Ecin}=0.097; nu: p_{Ebom}=0.12; p_{Emer}=0.007, p_{Ecin}=0.671). This significant local structure is likely the result of genetic differentiation between *E. meriana* individuals co-distributed in Central America and the Chocó Region but belonging to the basal and derived lineages in this species (Fig. 1.3B).

Phylogeographic structure and signatures of population expansion

The hierarchical F-statistics results for the mitochondrial and nuclear datasets were incongruent across all three focal species. The AMOVA analysis of mtDNA data indicated that CSAs explained 24% and 34% of the variation in *E. bombiformis* and *E. meriana*, respectively, but did not explain any variation in *E. cingulata* (Table 3). Likewise, Fct values based on mtDNA were highly significant in *E. bombiformis* and *E. meriana*, but not in *E. cingulata*. For nuDNA, most of the genetic variation was found within populations (Vc > 90% for all three species) (Table 1.4). We found significant patterns of isolation-by-distance for the mitochondrial data in *E. bombiformis* and *E. meriana* (r_{Ebom}=0.32, p=0.031; r_{Emer}=0.52, p=0.001), but not in *E. cingulata* (r_{Ecin}=0.0082, p=0.28). However, isolation-by-barrier showed a higher

correlation with genetic differentiation in *E. bombiformis* and *E. meriana* (r_{Ebom}=0.65, p=0.001; r_{Emer}=0.67, p<0.001). Neither IBD nor IBB were significant for the nuclear haplotype data (IBD: r_{Ebom}=0.19, p=0.12; r_{Emer}=0.12, p=0.22; r_{Ecin}=0.16, p=0.14; IBB: r_{Ebom}=-0.018, p=0.47; r_{Emer}=0.13, p=0.88; r_{Ecin}=0.0078, p=0.53).

Fu's Fs values were negative and highly significant in all phylogeographic groups except for NA and AF in *E. meriana* (Table 1.5). These results indicate that most of the sampled populations have not reached mutation-drift equilibrium due to demographic population expansion or purifying selection. The Extended Bayesian Skyline Plots (EBSPs) detected one population size increase for all lineages except for CR in *E. bombiformis* and AF in *E. meriana* (Fig. 1.5). EBSPs were not calculated in *E. cingulata* due to lack of monophyletic lineages in this taxon. Estimates of current effective population sizes based on EBSPs vary between 60k and 3700k individuals per lineage (Table 1.4). IMa2 runs for two pairs of populations in *E. bombiformis* (pair 1: "CR2"and CA; pair 2: NA2+NA3 and "AF") reached convergence with high ESS values across all parameters. However, parameters could not be estimated with certainty (posterior density did not reach low levels at the lower limit of the prior), thus these results should be interpreted with caution. Peak posterior distribution of migration rate estimates indicated that the number of migrants per generation was at least 3 times lower in the mtDNA than in the nuDNA (Table 1.6, Fig. 1.6). The highest estimated migration rate was detected from Central America (CA) to the Chocó Region ("CR2") (mtDNA; HiPt 0.0125, 95% HPD=0.0075 – 97.49; nuDNA: HiPt 0.563, 95% HPD=0.168

Table 1.5. Summary of the analysis of molecular variance (AMOVA) for the mitochondrial (mtDNA) and nuclear (nuDNA) datasets. Populations are grouped based on the identified CSAs from the ecological niche modeling. Significance of p-values shown as < 0.05 (*), p < 0.01 (**), p < 0.005 (***), or non-significant (n.s).

		Among CSA		Among populations within CSA		Within populations	
Species		**Va**	**Fct**	**Vb**	**Fsc**	**Vc**	**Fst**
Eulaema bombiformis	mtDNA	23.71%	0.24**	8.56%	0.11 $^{n.s}$	67.73%	0.32***
	nuDNA	5.23%	0.053*	-2.76%	-0.025$^{n.s}$	97.53%	0.025$^{n.s}$
Eulaema meriana	mtDNA	34.33%	0.34***	11.31%	0.17***	54.36%	0.46***
	nuDNA	2.65%	0.026 $^{n.s.}$	6.27%	0.064**	91.07%	0.089***
Eulaema cingulata	mtDNA	-3.13%	-0.031 $^{n.s.}$	39.95%	0.39***	63.18%	0.37***
	nuDNA	0.005%	0.002$^{n.s.}$	-0.02%	-0.009$^{n.s.}$	100%	-0.008$^{n.s.}$

Table 1.6. Demographic parameter estimates for lineages and neutrality tests for geographic groups in *E. bombiformis, E. meriana* and *E. cingulata.* Current median effective population size (*Ne* in thousands) and number of changes in population size (Δ*Ne*) were estimated through Extended Bayesian Skyline Plots. Significance for the neutrality tests shown as p < 0.02 (*), p < 0.005 (**), p < 0.001 (***), or non-significant (n.s).

	Lineage	***Ne***	**Δ*Ne***	**CSA**	**Fu's *Fs***
Eulaema bombiformis	**CA**	377 (±242)	1 (±0.042)	**CA**	-21.834***
	'CR1'	419 (±291)	0 (±0.001)	**CR**	-4.846**
	NA2	2049 (±169)	1 (±0.0087)	**NA**	-6.4571*
	NA3	3714 (±1563)	1 (±0.013)		
	AF	371 (±96.4)	1 (±0.096)	**AF**	-7.609***
Eulaema meriana	**CA1**	246 (±5.6)	1 (±0.005)	**CA**	-25.184***
	'CA2'	670 (±121)	1 (±0.021)		
	CR1	29 (±16.4)	1 (±0.019)	**CR**	-22.562***
	'CR2'	243(±198)	1 (±0.033)		
	NA	66 (±16)	1 (±0.009)	**NA**	-1.0114
	WA	1622 (±558)	1 (±0.018)	**WA**	-26.035***
	AF	689 (±260)	0 (±0.011)	**AF**	-2.482
Eulaema cingulata	**CA**	N.A.	N.A.	**CA**	-26.495***
	CR	N.A.	N.A.	**CR**	-25.231***
	NAm	N.A.	N.A.	**NAm**	-8.627***
	SAm	N.A.	N.A.	**SAm**	-25.626***

– 22120.40) and the lowest from the Atlantic Forest (”AF”) to the Amazon (NA2+NA3) (mtDNA: HiPt 0.015, 95% HPD=0.0084 – 244.76; nuDNA: HiPt 0.619, 95% HPD=0.219 – 296.41) (Table 1.6, Fig. 1.6).

Discussion

Demographic effects of climatic stability

Ours is the first study to investigate the phylogeographic history of widespread pollinators at a transcontinental geographic scale using a combination of spatial models and multi-locus molecular markers. Our results confirmed our hypothesis that climatic instability during the Pleistocene played an important role on the intraspecific lineage diversification in these Neotropical pollinators, and that the impact of historical climate variability varied between taxa with different niche breadths. Our initial hypothesis was supported based on three results. First, species-specific paleomodels indicated a reduction in the geographic distribution of all three species. As predicted, *E. cingulata* was the most widely distributed species during the LGM but this species also experienced the greatest reduction in distribution suggesting that during glaciations the dry areas in the Neotropics were unsuitable even for the species with the widest physiological tolerance. Second, our results revealed strong geographic structuring of mitochondrial genetic diversity in the two species with narrower physiological tolerance, consistent with predictions of population persistence in refugia and low female dispersal between CSAs. This pattern was not found in *E. cingulata*, the species with wide physiological tolerance. The lack of structuring in *E. cingulata* may be the result of high historical and contemporary dispersal between CSAs, given the broad physiological tolerance of this species. Even though the ENM for *E. cingulata*

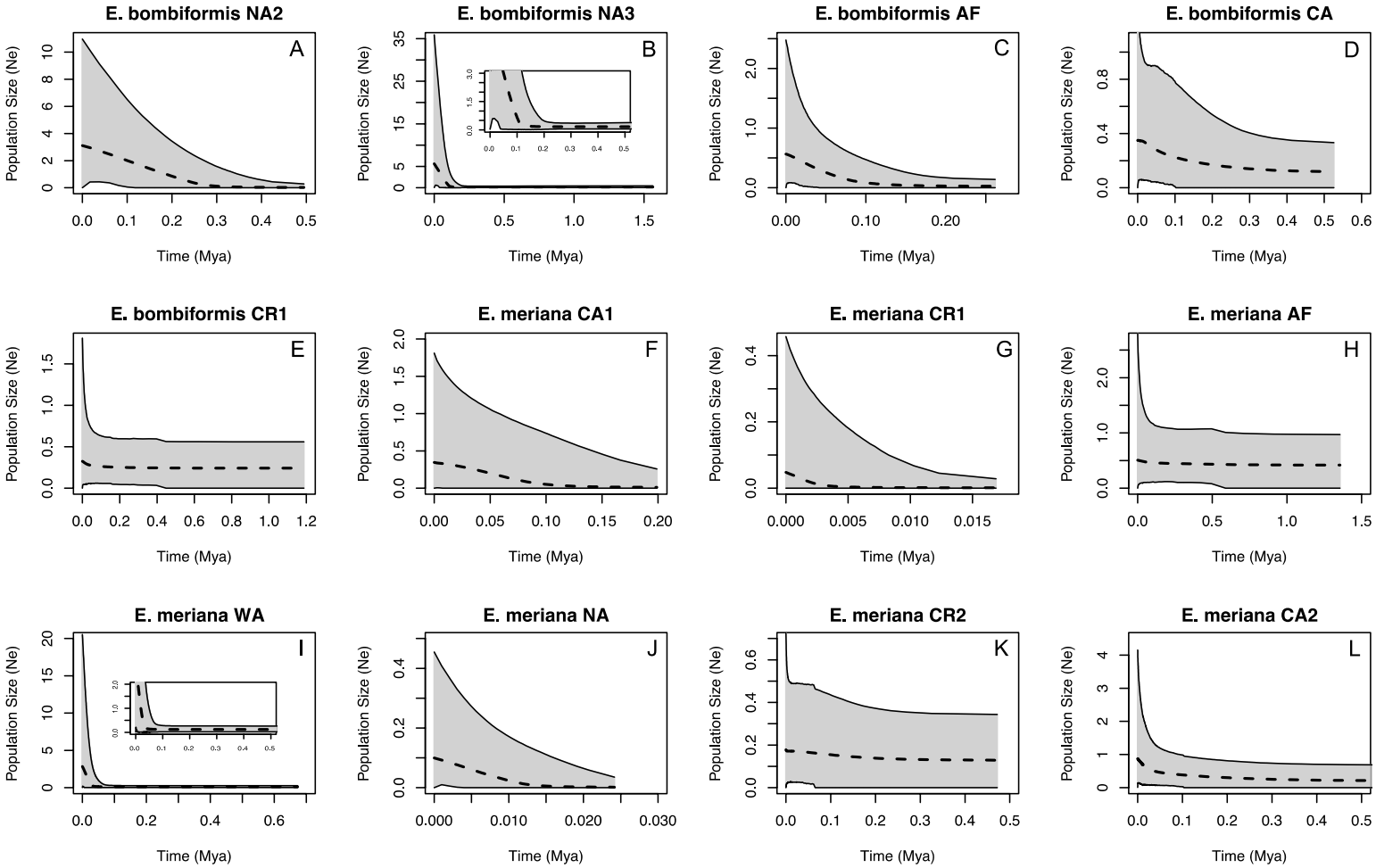


Figure 1.5. Effective population size (Ne) through time (Mya) for monophyletic lineages in *E. bombiformis* (A-E) and *E. meriana* (F-L). Dotted lines indicate median Ne. Upper and lower lines indicate 95% highest posterior density intervals.

Table 1.7. Estimates of effective population migration rate for *E. bombiformis* from the IMa2 analyses. Values shown are peak probabilities with highest posterior densities at the lower and upper 95%.

Populations[¶]	Genome	$m_{0>1}$[§]			$m_{1>0}$[¥]		
		High Pt	**HPD95Lo**	**HPD95Hi**	**High Pt**	**HPD95Lo**	**HPD95Hi**
CR – CA	mt	0.0125	0.0075	97.49	0.509	0.193	89.16
	nu	0.0875	0.194	46708	1.618	0.218	43518
Am - AF	mt	0.0129	0.0061	48.55	0.015	0.0084	244.76
	nu	0.563	0.168	22120	0.619	0.219	296.41

[¶]Populations CR, CA, Am, AF correspond to clades “CR2”, CA, NA2+NA3, “AF” on Figure 2A

[§]Rate at which population 0 (CR, Am) receives genes from population 1 (CA, AF)

[¥]Rate at which population 1 (CA, AF) receives genes from population 0 (CR, Am)

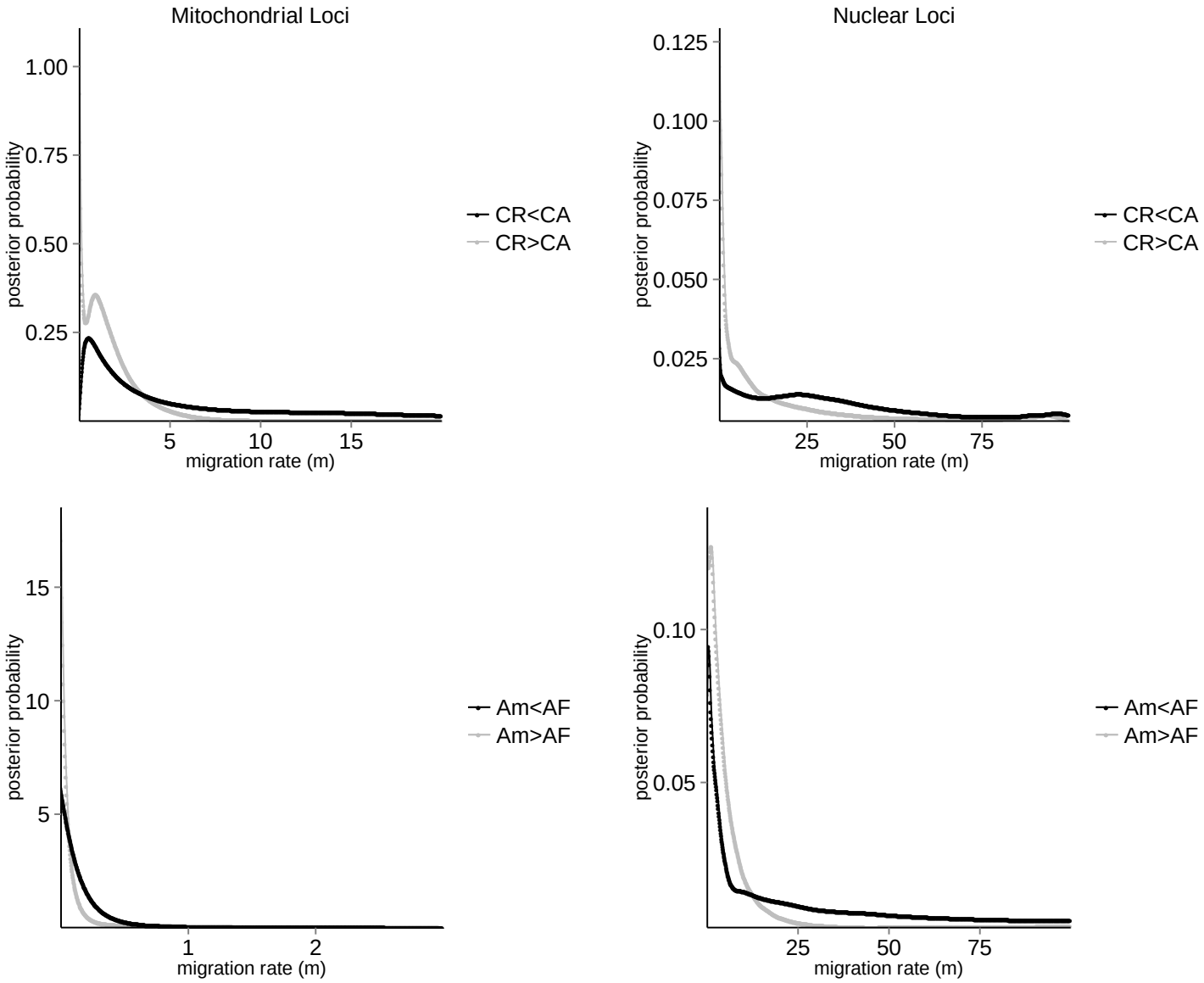

Mitochondrial Loci
posterior probability
1.00
0.75
0.50
0.25
5
10
15
migration rate (m)
CR<CA
CR>CA
Nuclear Loci
posterior probability
0.125
0.100
0.075
0.050
0.025
25
50
75
migration rate (m)
CR<CA
CR>CA
posterior probability
15
10
5
1
2
migration rate (m)
Am<AF
Am>AF
posterior probability
0.10
0.05
25
50
75
migration rate (m)
Am<AF
Am>AF

Figure 1.6. Posterior probability densities for mitochondrial and nuclear mutation-scaled migration rates between populations from the east and west clades of *Eulaema* bombiformis. Parameters denote migration rate estimated as number of migrants per population. Populations CR, CA, Am, AF correspond to clades “CR2”, CA, NA2+NA3, “AF” on Figure 2A.

indicated a dramatic reduction in habitat suitability during the Pleistocene, we cannot dismiss the possibility that this species was capable of dispersal through unsuitable areas during that time. Third, phylogeographic structuring was temporally consistent with the inferred CSAs. Our dating analysis indicated that both *E. bombiformis* and *E. meriana* originated during the late Pliocene and, except for the two basal lineages in *E. meriana*, all lineages date from the Pleistocene suggesting that the diversification of these lineages is recent, and not likely a result of earlier Pliocene geological and tectonic events in South America.

We found striking incongruence in the patterns of distribution of genetic diversity from nuclear and mitochondrial markers. MtDNA showed a significant association between geographic and genetic distance in all three focal species, a pattern absent in the nuDNA. Discordant cyto-nuclear patterns are common in recently diverged lineages (Gómez-Zurita & Vogler 2003; Bryja et al. 2010) and different processes may explain this pattern. First, discordance can result from long-term male-biased dispersal resulting in panmixia in the nuDNA and genetic structure in the mtDNA. Alternatively, cyto-nuclear discordance can arise due to slower mutation rate and larger effective population size of the nuclear genome (Prugnolle & Meeus 2002). Another possible explanation is that female traits, such as mtDNA, are adaptive resulting in mitochondrial haplotype frequencies selected by environmental

conditions (Pavlova et al. 2013). In this study, we targeted two mitochondrial protein-coding markers and five nuclear markers with slow and fast mutation rates. Both types of nuclear markers, the slow-evolving protein-coding and the fast-evolving ASCN loci, lacked genetic structure, while the mtDNA loci consistently showed high population structure. Our sampling within CSAs included populations separated by more than 2500 km and with latitudinal variation of 23 degrees. Therefore, adaptive selection seems an unlikely explanation for the cyto-nuclear discordance given the great environmental variation within sampled CSAs. Estimates of migration rates for *E. bombiformis* indicate lower gene flow between phylogeographic groups with the mitochondrial than the nuclear genome. Therefore, our results validate previous ecological and behavioral studies that suggested that gene flow is highly male-biased in orchid bees (López-Uribe et al. 2008).

Overall, paleomodels and demographic inferences indicated that climatic instability, physiological tolerance and sex-biased dispersal shaped the current distribution of genetic diversity in our focal orchid bee species. Furthermore, our results corroborate that current wet forested areas in the Neotropics, particularly in the Amazon basin, were drier and supported smaller populations during the LGM (Solomon et al. 2008, Alfaro et al. 2012). This study provides clear evidence that climatic instability can be an important driver of intraspecific lineage diversification even in highly mobile organisms with large effective population sizes.

Phylogeographic history

Our mitochondrial results provide new insights into the origins and colonization histories of our focal orchid bee species. Because a single mated female bee can colonize a new area and establish a population (Zayed et al. 2007), the

mtDNA is an informative marker to make inferences about colonization events even though this marker exclusively depicts the evolutionary history of maternal lineages. The mitochondrial topology for *E. bombiformis* indicates an origin in the Amazon Basin and later colonization west of the Andes. Our data indicate that the colonization of the Atlantic Forest from the Amazon basin occurred during a wet climatic period of the Pleistocene (~0.35 Mya; HPD= 0.28, 1.18) (Fig. 2A). The presence of two individuals from the northern Brazilian Amazon within the Atlantic forest clade suggests a connection between these two regions through the Caatinga in northeastern Brazil (Wang et al. 2004; Batalha-Filho et al. 2012) and not through the Paraguayan corridor in the south (Ramírez et al. 2010). Furthermore, a recent collection of *E. bombiformis* in mid-elevation forested islands in the middle of the Caatinga (Ubajara, Ceará, Brazil) (Nemésio & Ferrari 2012) supports the hypothesis of a forest connection between the Amazon and the Atlantic forest through historical gene flow corridors in northern South America. We found a recent divergence between populations east and west of the Andes (~1.48 Mya; HPD= 0.88, 2.16) that post-dates the uplift of this Cordillera (Hoorn et al. 2010) indicating cross-Andean colonization rather than vicariance (Fig. 2A).

Eulaema meriana originated in Central America about ~3.7 Mya (HPD= 1.96, 5.8) and colonized the Chocó Region during the late Pliocene (~2.9 Mya; HPD= 1.61, 4.36) (Fig. 2B). Eastern South America was subsequently colonized after the formation of the Andes, and those populations diverged into two lineages (Atlantic Forest [“AF”] and Amazon Basin [WA+CAm]), congruent with CSAs identified by the paleomodels in that region. During mid-Pleistocene, *E. meriana* colonized the Guiana’s foothills in northern South American and recolonized the lowland forests west of the Andes where two

different lineages diverged: one in the Chocó Region and one in Central America. Results from *E. meriana* corroborate the evidence of cross-Andean dispersal found in *E. bombiformis* and other orchid bee species with transcontinental distributions (Dick et al. 2004).

The lack of phylogenetic signal in the mtDNA of *E. cingulata* precludes inference of the phylogeographic history of this species (Fig. 2C). From the dating analysis, it is clear that *E. cingulata* has a more recent origin than *E. bombiformis* and *E. meriana*. Therefore, we interpret the absence of phylogeographic resolution for this species as one or a combination of the following: (1) incomplete lineage sorting due to the recent origin of the species; (2) broader physiological tolerance to dry climatic conditions allowing gene flow through unsuitable habitats between CSAs; (3) recent colonization by a maternal lineage from a single refugium. We caution that the estimated divergence times were based on mutation rates calibrated for mitochondrial genes in other insects. Therefore, these estimates should not be interpreted as absolute but relative times of divergence. Nonetheless, the low sequence divergence detected between lineages supports the recent dates estimated based on the mitochondrial mutation rates.

The phylogeographic groups we identified in this study are congruent with major biogeographic breaks that shaped the interspecific diversification of other widespread Neotropical taxa (Martins et al. 2009; Solomon et al. 2008). These studies report monophyletic lineages from Central America, the Amazon basin, and the Atlantic Forests but the presence/absence of lineages within these major phylogeographic areas is not concordant among all taxa examined to date. Phylogeographic studies in the Neotropics suggest that paleoclimatic instability was an important driver of intraspecific diversification

in highly mobile organisms as evidenced by the spatial and temporal congruence between genetic lineages and predicted CSAs. However, general spatial patterns of vicariance and dispersal are difficult to predict due to taxon-specific physiological niche breadths that drive different responses to climatic changes.

Persistence and species vulnerability to future climate change

Our multi-locus comparative study suggests that orchid bees have large effective population sizes and experience long-distance male-mediated gene flow that maintains high levels of genetic diversity and connectivity among populations. Our niche models show that the geographic distribution of these *Eulaema* species can be significantly reduced under dry climatic conditions, nonetheless, orchid bees seem resilient to the detrimental effects of climatic instability due to their long-distance dispersal capabilities and large effective population sizes. However, we observed different species-specific responses depending on physiological tolerance. Species with narrower niches are more susceptible to isolation as a result of climate change due to their inability to disperse through unsuitable habitat. In the face of continued climate change, those more susceptible species will be the first to suffer demographic consequences of geographic isolation due to reduction of suitable habitat. In addition, habitat degradation and severe reduction of wet forested areas due to agricultural intensification are likely to exacerbate the reduction in geographic distribution and increase in isolation. Our results shows that understanding species-specific responses to historical climate change, and combining those data with estimates of physiological tolerances, can predict which species will be most affected by future environmental changes.

Acknowledgements

We thank F. Campos, C. Keil, and C. Castillo for their assistance with collecting permits; S. Ramírez, D. Roubik, S. Cardinal, M. Oliveira and M. Sánchez for providing museum specimens included in this study; L. Duque for his assistance in the field; S. Bogdanowicz, A. Green and C. Santiago for their assistance in the laboratory; and members of the Danforth lab, Zamudio lab and two anonymous reviewers for feedback on earlier versions of the manuscript. Funding was provided by grants from the Organization of Tropical Studies (OTS), Grace Griswold Endowment, Mario Einaudi Center, Lewis and Clark Exploration Fund, Explorer's Club Exploration Fund (to M.M.L-U), and by awards from the National Science Foundation (DEB-0814544 and DEB-0742998 to B.N.D.). Samples were collected and exported under the following permits: 614-2008 Costa Rica, 2009-027-QCAZ-E Ecuador, 304-2010 Peru and 48/2010 Brazil.

REFERENCES

Alfaro JWL, Boubli JP, Olson LE, Di Fiore A, Wilson B, Gutiérrez-Espeleta GA, Chiou KL, Schulte M, Neitzel S, Ross V, Schwochow D, Nguyen MTT, Farias I, Janson CH, Alfaro ME (2012) Explosive Pleistocene range expansion leads to widespread Amazonian sympatry between robust and gracile capuchin monkeys. Journal of Biogeography, 39, 272–288.

Augusto SC, Garófalo CA (2011) Task allocation and interactions among females in Euglossa carolina nests (Hymenoptera, Apidae, Euglossini). Apidologie, 42, 162–173.

Bartomeus I, Ascher JS, Wagner D, Danforth BN, Colla S, Kornbluth S, Winfree R (2011) Climate-associated phenological advances in bee pollinators and bee-pollinated plants. Proceedings National Academy Sciences USA, 108, 20645–20649.

Batalha-Filho H, Fjeldså J, Fabre P-H, Miyaki CY (2012) Connections between the Atlantic and the Amazonian Forest avifaunas represent distinct historical events. Journal of Ornithology, 154, 41−50.

Bennet KD (1990) Milankovitch cycles and their effects on species in ecological and evolutionary time. Paleobiology, 16, 11-21.

Bonier F, Martin PR, Wingfield JC (2007) Urban birds have broader environmental tolerance. Biology Letters, 3, 670–673.

Bryja J, Smith C, Konečný A, Reichard M (2010) Range-wide population genetic structure of the European bitterling (*Rhodeus amarus*) based on microsatellite and mitochondrial DNA analysis. Molecular Ecology, 19, 4708–4722.

Carnaval A, Bates J (2007) Amphibian DNA shows marked genetic structure and tracks Pleistocene climate change in northeastern Brazil. Evolution, 61, 2942-2957.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues, MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic Forest hotspot. Science, 323, 785–789.

Cheviron ZA, Hackett SJ, Capparella AJ (2005) Complex evolutionary history of a Neotropical lowland forest bird (*Lepidothrix coronata*) and its implications for historical hypotheses of the origin of Neotropical avian diversity. Molecular Phylogenetics and Evolution, 36, 338–357.

Chown SL, Hoffmann AA, Kristensen TN, Angilletta Jr MJ, Stenseth NC, Pertoldi C (2010) Adapting to climate change: a perspective from evolutionary physiology. Climate Research, 43, 3–15.

Crozier YC, Koulianos S, Crozier RH (1991) An improved test for Africanized honeybee mitochondrial DNA. Experientia, 47, 968-969.

Danforth BN, Ji S (1998). Elongation factor-1α occurs as two copies in bees: Implications for phylogenetic analysis of EF-1α sequences in insects. Molecular Biology and Evolution, 15, 225-235.

Danforth BN (1999) Phylogeny of the bee genus *Lasioglossum* (Hymenoptera: Halictidae) based on mitochondrial COI sequence data. Systematic Entomology, 24, 377–393.

Danforth BN, Fang J, Sipes SD (2006) Analysis of family level relationships in bees (Hymenoptera: Apiformes) using 28S and two previously unexplored nuclear genes: CAD and RNA polymerase II. Molecular Phylogenetics and Evolution, 39, 358-372.

Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM (2011) Beyond predictions: biodiversity conservation in a changing climate. Science, 332, 53–58.

Dick CW, Roubik DW, Gruber KF, Bermingham E (2004) Long-distance gene flow and cross-Andean dispersal of lowland rainforest bees (Apidae: Euglossini) revealed by comparative mitochondrial DNA phylogeography. Molecular Ecology, 13, 3775−3785.

Ditchfield AD (2000) The comparative phylogeography of Neotropical mammals: patterns of intraspecific mitochondrial DNA variation among bats contrasted to nonvolant small mammals. Molecular Ecology, 9, 1307–1318.

Dressler RL (1982) Biology of the orchid bees (Euglossini). Annual Review of Ecology and Systematics, 13, 373–394.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology, 7, 214.

Eltz T, Sager A, Lunau K (2005). Juggling with volatiles: Exposure of perfumes by displaying male orchid bees. Journal of Comparative Physiology. A, Neuroetholy, Sensory, Neural and Behavioral Physiology 191, 575–581.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, 10, 564–567.

Farrell BD (2001) Evolutionary assembly of the milkweed fauna: Cytochrome oxidase I and the age of Tetraopes beetles. Molecular Phylogenetics and Evolution, 18, 467–478.

Fjeldsa J (1994) Geographical patterns for relict and young species of birds in Africa and South-America and implications for conservation priorities. Biodiversity and Conservation 3, 207-226.

Forister ML, McCall AC, Sanders NJ, Fordyce JA, Thorne JH, O'Brien J, Waetjen DP, Shapiro AM (2010) Compounded effects of climate change and habitat alteration shift patterns of butterfly diversity. Proceedings National Academy Sciences USA, 107, 2088–2092.

Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular Ecology, 18, 4541–4550.

Gómez-Zurita J, Vogler AP (2003) Incongruent nuclear and mitochondrial phylogeographic patterns in the *Timarcha goettingensis* species complex (Coleoptera, Chrysomelidae). Journal of Evolutionary Biology, 16, 833–843.

Heller NE, Zavaleta ES (2009) Biodiversity management in the face of climate change: a review of 22 years of recommendations. Biological Conservation, 142, 14–32.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society B, 359, 183–195.

Hey J, Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proceedings National Academy Sciences USA, 104, 2785–2790.

Hoorn C, Wesselingh FP, Ter Steege H, Bermudez MA, Mora A, Sevink J, Sanmartín I, Sanchez-Meseguer A, Anderson CL, Figueiredo JP, Jaramillo C, Riff D, Negri FR, Hooghiemstra H, Lundberg J, Stadler T, Särkinen T,

Antonelli A (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science, 330, 927-931.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 17, 754–755.

Hugall A, Moritz C, Moussalli A, Stanisic J (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875) Proceedings National Academy Sciences USA, 99, 6112–6117

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution, 23, 254–267.

Janzen DH (1971) Euglossine bees as long-distance pollinators of tropical plants. Science, 171, 203-205.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. BMC Genetics, 6, 13.

Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity, 101, 92-103.

Kieswetter CM, Schneider CJ (2013) Phylogeography in the northern Andes: Complex history and cryptic diversity in a cloud forest frog, *Pristimantis w-nigrum* (Craugastoridae) Molecular Phylogenetics and Evolution 69:417-429.

Knowles LL, Carstens BC, Keat ML (2007) Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. Current Biology, 17, 940–946.

Knowles LL, Alvarado-Serrano DF (2010) Exploring the population genetic consequences of the colonization process with spatio-temporally explicit

models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. Molecular Ecology, 19, 3727–3745.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25, 1451–1452.

López-Uribe MM, Oi CA, Del Lama MA (2008) Nectar-foraging behavior of Euglossine bees (Hymenoptera: Apidae) in urban areas. Apidologie, 39, 410–418.

López-Uribe MM, Green AN, Ramírez SR, Bogdanowicz SM, Danforth BN (2011) Isolation and cross-species characterization of polymorphic microsatellites for the orchid bee *Eulaema meriana* (Hymenoptera: Apidae: Euglossini). Conservation Genetics Resources, 3, 21–23.

Maddison DR, Maddison WP (2005) MacClade 4: Analysis of phylogeny and character evolution. Version 4.08a. http://macclade.org.

Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. Bioinformatics, 21, 260–262.

Martins FM, Templeton AR, Pavan ACO, Kohlbach BC, Morgante JS (2009) Phylogeography of the common vampire bat (*Desmodus rotundus*): Marked population structure, Neotropical Pleistocene vicariance and incongruence between nuclear and mtDNA markers. BMC Evolutionary Biology, 9, 294.

Mulcahy DG, Morrill BH, Mendelson JR (2006) Historical biogeography of lowland species of toads (*Bufo*) across the Trans-Mexican Neovolcanic Belt and the Isthmus of Tehuantepec. Journal of Biogeography, 33, 1889-1904.

Myers M, Mittermeier RA, Mittermeier, CG, Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature, 403, 853-858

Nemésio A, Ferrari RR (2012) The species of *Eulaema* (*Eulaema*) Lepeletier, 1841 (Hymenoptera: Apidae: Euglossina) from eastern Brazil, with description of *Eulaema quadragintanovem* sp. n. from the state of Ceará. Zootaxa, 3478, 123–132.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature, 421, 37–42.

Pavlova A, Amos JN, Joseph L, Loynes K, Austin JJ, Keogh JS, Stone GN, Nicholls JA, Sunnucks P (2013) Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. Evolution, 67, 3412-3428.

Pearson RG, Dawson TP, Liu C (2004) Modelling species distributions in britain: A hierarchical integration of climate and land-cover data. Ecography, 27, 285–298.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190, 231–259.

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography, 31, 161–175.

Pimm SL, Kevan P (2000) Biodiversity: extinction by numbers. Nature, 403, 843–845.

Posada D (2008) jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution, 25, 1253–1256.

Potts SG, Biesmeijer JC, Kremen C, Neumann P, Schweiger O, Kunin WE (2010) Global pollinator declines: trends, impacts and drivers. Trends in Ecology and Evolution, 25, 345–353.

Prado CPA, Haddad CFB, Zamudio KR (2012) Cryptic lineages and Pleistocene population expansion in a Brazilian Cerrado frog. Molecular Ecology, 21, 921–941.

Prugnolle F, Meeus T de (2002) Inferring sex-biased dispersal from population genetic tools: a review. Heredity, 88, 161–165.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Ramírez S, Dressler RL, Ospina M (2002) Abejas euglosinas (Hymenoptera: Apidae) de la región Neotropical: listado de especies con notas sobre su biología. Biota Colombiana, 3, 7–118.

Ramírez SR, Roubik DW, Skov C, Pierce NE (2010) Phylogeny, diversification patterns and historical biogeography of euglossine orchid bees (Hymenoptera: Apidae). Biological Journal of the Linnean Society, 100, 552–572.

Richards CL, Carstens BC, Knowles LL (2007) Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. Journal of Biogeography, 34, 1833–1845.

Roubik DW, Hanson PE (2004) Orchid bees: Biology and field guide. INBio, San Jose, Costa Rica.

Sala OE, Chapin FS III, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Sanwald E, Huenneke LF, Jackson RB, Kinzig A, Leemans R, Lodge DM, Mooney HA, Oesterheld M, Poff NL, Sykes MT, Walker BH, Walker M, Wall DH (2000) Global biodiversity scenarios for the year 2100. Science, 287, 1770–1774.

Santos J, Coloma L, Summers K, Caldwell J, Ree R, Cannatella D (2009) Amazonian amphibian diversity is primarily derived from late Miocene Andean lineages. PLoS Biol, 7, 448-461.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annual Review of Entomology, 87, 651-701.

Solomon SE, Bacci M Jr, Martins J Jr, Vinha GG, Mueller UG (2008) Paleodistributions and comparative molecular phylogeography of leafcutter ants (Atta spp.) provide new insight into the origins of amazonian diversity. PLoS ONE, 3, e2738.

Trevan JW (1927) The error of determination of toxicity. Proceedings of the Royal Society of London, 101, 483–514.

Wang X, Auler AS, Edwards RL, Cheng H, Cristalli PS, Smart PL, Richards DA, Shen C-C (2004) Wet periods in northeastern Brazil over the past 210 kyr linked to distant climate anomalies. Nature, 432, 740−743.

Wikelski M, Moxley J, Eaton-Mordas A, López-Uribe MM, Holland R, Moskowitz D, Roubik DW, Kays R (2010) Large-range movements of Neotropical orchid bees observed via radio telemetry. PLoS ONE, 5, e10738.

Zayed A, Constantin SA, Packer L (2007) Successful biological invasion despite a severe genetic load. PLoS ONE, 2:e868.

CHAPTER 2

Genetic signatures of range expansion in the squash specialist bee, *Peponapis pruinosa*, support evidence of multiple domestication of *Cucurbita pepo*

Margarita M. López-Uribe[1], Robert L. Minckley[2], James H. Cane[3] and Bryan N. Danforth[1]

[1] Department of Entomology, Cornell University, Ithaca, NY 14850

[2] Department of Biology, University of Rochester, Rochester, NY 14627

[3] USDA-ARS Bee Biology and Systematics Laboratory, Utah State University, Logan, UT 84322

Abstract

Range expansions are common demographic processes in the history of most species. However, over the past 10,000 years humans have facilitated major range expansions of many species as a consequence of regional and global human movements. *Peponapis pruinosa* is a solitary, host-plant bee specialist of plants of the genus *Cucurbita*. The bee's geographic range was restricted to the distribution of its wild host-plant, *Cucurbita foetidissima,* which occurs in the warm deserts of Mexico and the southwestern of the United States. However, *P. pruinosa* is currently widely distributed throughout North America supporting the hypothesis that its range has greatly expanded as a consequence of the domestication and widespread cultivation of *C. pepo* by Native Americans. In this study, we investigated the impact of human

domestication and cultivation of *C. pepo* on the demographic history of its wild pollinator *P. pruinosa* based on a total of 942 individuals collected from 23 populations in Mexico, United States and Canada. We found clear genetic signatures of a spatial range expansion: decreasing genetic diversity with increasing geographic distance from Mexico. However, we did not detect evidence of population growth as the species expanded its range. On the contrary, coalescent-based analyses indicated severe bottlenecks in populations throughout the distribution of the species. Approximate Bayesian Computation analysis showed that the colonization of North America occurred from at least two independent events: one lineage colonized the west of NA (west Texas, Arizona, California, Utah and Idaho) while the other lineage colonized the center and east of the continent. Our data strongly support the hypothesis that the colonization of the northeast of North America occurred after the second domestication of *C. pepo* in the Midwest of the United States and not after the first domestication event in northeastern Mexico. This study supports the view that some bee species can prevail over extremely severe bottlenecks and yet successfully colonize and persist in large geographic areas. Our results also have direct implication for the potential management of *P. pruinosa* as a wild pollinator of cucurbits.

Introduction

Biological range expansions, contractions and shifts are common events in the demographic history of most species on earth. Bouts of range expansion and contraction are likely to occur repeatedly between the origination and extinction of most species. Changes in species ranges are also associated with the formation of new community assemblies that lead to changes in

species interactions and ecosystem functions. Perhaps the best-studied examples of species range shifts are the historical changes that occurred in the cold-warm periods of the Quaternary glacial cycles, during which some species migrated to new areas and other remained in refugia that persisted in their original range (Hewitt 2004). Also, land bridges during glaciations, as a result of reductions in sea levels, served as the colonization routes for intercontinental exchange of flora and fauna that later speciated and diversified (Ran *et al.* 2006; Xiang *et al.* 2005). Therefore, knowledge about past range shifts is relevant for understanding current community assemblage composition and biogeographic patterns.

Even though these demographic events are common in species histories, the proportion of species and the rate at which range shifts occur is accelerating due to human activities (Miller-Rushing *et al.* 2010). Introduction of non-native species into new geographic areas outside their potential range (Wilson *et al.* 2009) and latitudinal and altitudinal shifts as a consequence of the current rapid global climate change (Parmesan & Yohe 2003) are the two mechanisms thought to be the primary anthropogenic drivers of changes in species distributions. Specifically, increased air, land and water transport of commodities has greatly facilitated spread of invasive species in the last century (Richardson and Pysek 2007). Concurrently, atmospheric warming, greater severity and frequency of extreme climatic events and altered biogeochemical cycles have changed considerably in the last 50 years (Oreskes 2004).

Largely unstudied have been human-mediated changes in organism distributions that have been ongoing since the Holocene (~11kya) with the beginning of plant and animal domestication for food production (Smith 2006).

For many species, human activities have been a major force shaping demographic changes (at the species level) and the distribution of biodiversity (at the community level) in their recent evolutionary history. These species include not only those that have undergone domestication, but others that are distributed sympatrically with species that have been domesticated. Crop species provide dramatic examples of changes in species distributions among extant taxa because domestication typically occurs in a restricted geographic area and human transport later results in a subsequent, often exponential, increase in abundance and distribution. Archeological information on domestication events often provide evidence which, in combination with genetic studies, can help reconstruct the evolutionary history of these domesticated species and provide information about their center of origin and routes of expansion (Vavilov 1992).

Of particular interest is the domestication and range expansion of squash (*Cucurbita pepo*) that has been the center of debate due to the possible presence of two independent centers of domestication. Archeological evidence from seeds and fruit rind fragments suggest *C. pepo* was domesticated around 10,000 BP in southern Mexico, before other major crops such as maize and beans (Smith 1997). For a long time, the Oaxaca valley in Mexico was considered the only center of domestication of *C. pepo* (Wilson 1990; Kirkpatrick and Wilson 1988). Under this single domestication hypothesis, the appearance of domesticated *C. pepo* in eastern North America by 3,000 BP was considered the result of a human-mediated northward expansion of this “container crop” from Mesoamerica. However, the present-day records of wild *C. pepo* var. *texana* in the Midwest and pre-4,000 BP records of wild cucurbits in Illinois (USA) suggest that *C. pepo* was possibly

brought into domestication in North America about 5,000 years after the first domestication event in Mexico (Smith 2006). Allozyme, PCR-RFLPs and mitochondrial sequence data indicate deep divergence between Mexican mid-sized cucurbits and the ornamental gourds, acorns and crooknecks abundant in eastern NA (Decker and Wilson 1987; Sanjur *et al.* 2002; Wilson *et al.* 1992). This divergence at the genetic level further supports the hypothesis of two independent domestications of *C. pepo*. The first domestication event presumably gave rise to *C. pepo* var. *pepo,* the lineage of pumpkins and marrows; while the second domestication gave rise to *C. pepo* var. *ovifera*, the lineage of acorns and crooknecks (Smith 2006).

The impact of *C. pepo* range shifts on the distribution of native pollinators has been disregarded as an independent piece of evidence to better understand the domestication history of cucurbits (but see Bischoff *et al.* 2009). Species of *Cucurbita* are obligate bee-pollinated, and are visited by a diverse set of bee species (Petersen *et al.* 2013), including two related genera, *Xenoglossa and Peponapis,* of which all species are pollen-specialists of cucurbits. The 7 species of *Xenoglossa* are entirely found in North and Central America. The 13 species of *Peponapis* extend from South America throughout NA. However, *Peponapis pruinosa* is the only pollen-specialist species of *Cucurbita* that occurs outside the native range of its host plant. At present, *P. pruinosa* is currently distributed throughout most of NA from southern Mexico to southern Quebec (northeast) and central Idaho (northwest) (Fig. 2.1). The primary wild host of *P. pruinosa* is *C. foetidissima,* the wild buffalo gourd, which is native to xeric areas from Guanajuato (Mexico) to southern California to the west, Nebraska to the north and the Mississippi river to the east (Whitaker & Bemis 1964) (Fig. 2.1). On the east coast of NA, *P. pruinosa*

relies on the domesticated species (mostly *C. pepo* but also *C. moschata* and *C. maxima*) for pollen foraging. Thus, this narrowly oligolectic bee is hypothesized to have followed the cultivation of domesticated *Cucurbita* far beyond the range limits of its native warm-desert pollen host, *C. foetidissima*. However, the colonization route of *P. pruinosa* into the northeast of NA is unclear (Fig. 2.1). Under the single *C. pepo* domestication hypothesis, the range expansion to the northeast started 10,000 years ago (yra), and accompanied the spread of cultivated cucurbits through the east coast by Native Americans (Smith 2006). However, domesticated *C. pepo* only appears in the fossil record outside of Mexico until 3,000 B.P. Thus, the host shift of *P. pruinosa* between *C. foetidissima* and cultivated *C. pepo* could have occurred after the second domestication event in the Midwest.

Furthermore, some evidence suggests the distributional range of *P. pruinosa* is moving northwards today with its domesticated host. In western NA, the bee's range has extended as far as Boise (Idaho, United States) from the four-corner region (UT) in just 150 years, presumably following squash cultivation. In eastern NA, *P. pruinosa* reached its northern most distribution in the province of Quebec (Canada) only about 50 yra (Payette & Payette 2003) (Fig. 2.1). Thus, archeological and historical records of *Cucurbita* cultivation suggest that *P. pruinosa* started shifting its geographic range during the Holocene but its range is still expanding at the limits of the distribution.

Understanding the evolutionary consequences of range expansions on population demographics and adaptation is of critical importance to predict species responses to human-mediated range shifts. When species expand their ranges, individuals on the advancing edge of the distribution experience unique selective pressures as a result of the interaction with new biotic and

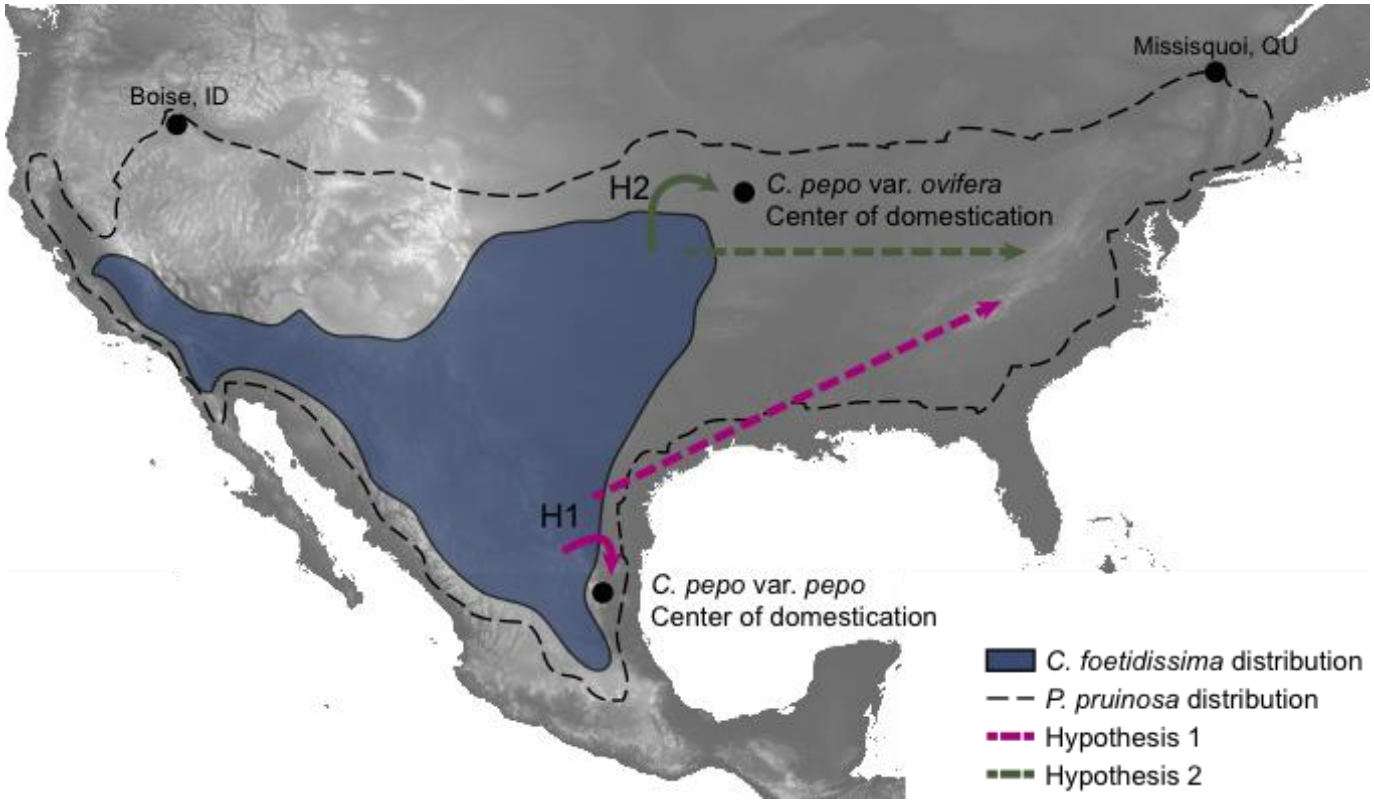


Figure 2.1. Hypotheses about the colonization routes of *Peponapis pruinosa* to the northeast of North America. Black points show the two centers of domestication of *Cucurbita pepo,* and the northernmost points of distribution. Map shows current geographic range of *P. pruinosa* (black dashed line) and *C. foetidissima* (blue area). Hypothesis 1 (H1) indicates the colonization of the northeast occurred by shifting host plants from *C. foetidissima* to *C. pepo* var. *pepo* after the first domestication event. Hypothesis 2 (H2) suggests the host shift occurred after the second domestication event from *C. foetidissima* to *C. pepo* var. *ovifera.*

abiotic agents (Suarez & Tsutsui 2008). Therefore, it is necessary to understand the role that different evolutionary processes play before, during and after species move from their native range into new habitats. During spatial range expansions, two demographic processes repeatedly occur. First, the source population usually increases in size (pure demographic expansion), predicting weak influence of genetic drift and stable allele frequencies at the source population (Kimura & Crow 1963). Second, the colonization of a new area by migrants from the source population causes a founder effect in the new population with an expected reduction of genetic diversity. Thus, for range expansions over large geographic areas, repeated founder effects produce a pattern of genetic diversity steadily decreasing along the expansion axis (Austerlitz *et al.* 1997). However, the ability to detect these bottlenecks from genetic data depends on the size of the founder population, the levels of gene flow with the source population and the population growth rate of the species (Excoffier *et al.* 2009). Despite abundant studies empirically showing the genetic consequences of range expansions in domesticated crops (Zeder *et al.* 2006) or their pathogens (Fontaine *et al.* 2013), few have investigated the genetic consequences of expansions in mutualistic species (Rosendahl *et al.* 2009).

The aim of this study was to reconstruct the evolutionary history of the cucurbit-specialist bee *P. pruinosa,* to test three hypotheses regarding the presumed history of range expansion. Our first hypothesis is that *P. pruinosa* expanded its geographic range northwards from the hypothesized center of origin of its wild host *C. foetidissima* in the xeric areas of Mexico and the southwestern United States (Hurd & Linsley 1964). If the ancestral range of *P. pruinosa* was located in the southern range of its current distribution, we

expect to find decreasing genetic diversity with increasing latitude. Our second hypothesis is that *P pruinosa* colonized the east coast of NA by switching host plant species from the wild *C. foetidissima* to the cultivated *C. pepo* var. *ovifera* in the Midwest of NA (Bischoff 2003). If this hypothesis is supported, *P. pruinosa* may have shifted from its native to its domesticated host about 5,000 ybp. The alternative hypothesis is that *P. pruinosa* colonized the east coast after the domestication of *C. pepo* var. *pepo* in northeastern Mexico, approximately 10,000 ybp (Fig. 1) (Whitaker & Bemis 1964). Under the hypothesis of a Midwest origin, populations from the northeast of NA would be more genetically related to populations from the Midwest than to populations from northeastern Mexico and east Texas. Our third hypothesis is that *P. pruinosa* populations at the extremes of the distribution show evidence of ongoing range expansion. This hypothesis would be supported by very recent population changes of population size around the periphery of the current distribution. We predict that the time of the most recent bottleneck is more recent in populations at the extremes of the distribution that in any other population throughout its geographic range. We use a combination of microsatellite, demographic parameter estimation, and Approximate Bayesian Computation (ABC) analysis to test these three hypotheses.

Our results provide unique insights into the genetic consequences of range shifts over a millennial scale (5,000 to 10,000 years) in a continuously expanding population, and the effects of agricultural expansion on a pollinator mutualist. Our results give evidence that the current widespread distribution of *P. pruinosa* in NA is likely due to the range expansion of two lineages from Mexico into the United States: one to the west (western Texas, Arizona, California, Utah and Idaho) and one to the center and east of NA. Moreover,

the ABC analysis indicates that the most likely scenario of colonization of the northeast of NA corresponds to the host-plant switch between the wild *C. foetidissima* and the cultivated *C. pepo* in the Midwest and not in northeastern Mexico. This study supports the view that some bee species can colonize large geographic areas and have long-term persistence in spite of repeated population bottlenecks and surprisingly low genetic diversity.

Materials and Methods

Sampling

We collected 942 individuals of *P. pruniosa* (438 ♂ and 504 ♀) from 23 populations (33 sites) in Mexico, the United States and Canada (Fig. 2.2). Individuals sampled from sites within the same state in the northeast of NA were treated as the same population. All specimens were collected in crop fields of *Cucurbita* plants except for samples from Austin, TX collected from a home garden and samples from El Paso, TX and Douglas, AZ that were collected on *C. foetidissima*. Individuals were kept in alcohol for DNA preservation until DNA analyses.

Microsatellite development and variability

We built genomic libraries enriched for microsatellites using DNA extracted from two individuals collected in Ithaca (NY, USA). For microsatellite discovery, we used two different methods: cloning and pyrosequencing (López-Uribe *et al.* 2013). For the first method, we cloned DNA fragments into pUC 19 plasmids that were inserted into *Escherichia coli* cells through electroporation. Cultures of transformed bacteria were screened for microsatellites through hybridization to oligonucleotides with the same microsatellite motifs used for the enrichment process. Bacterial colonies that

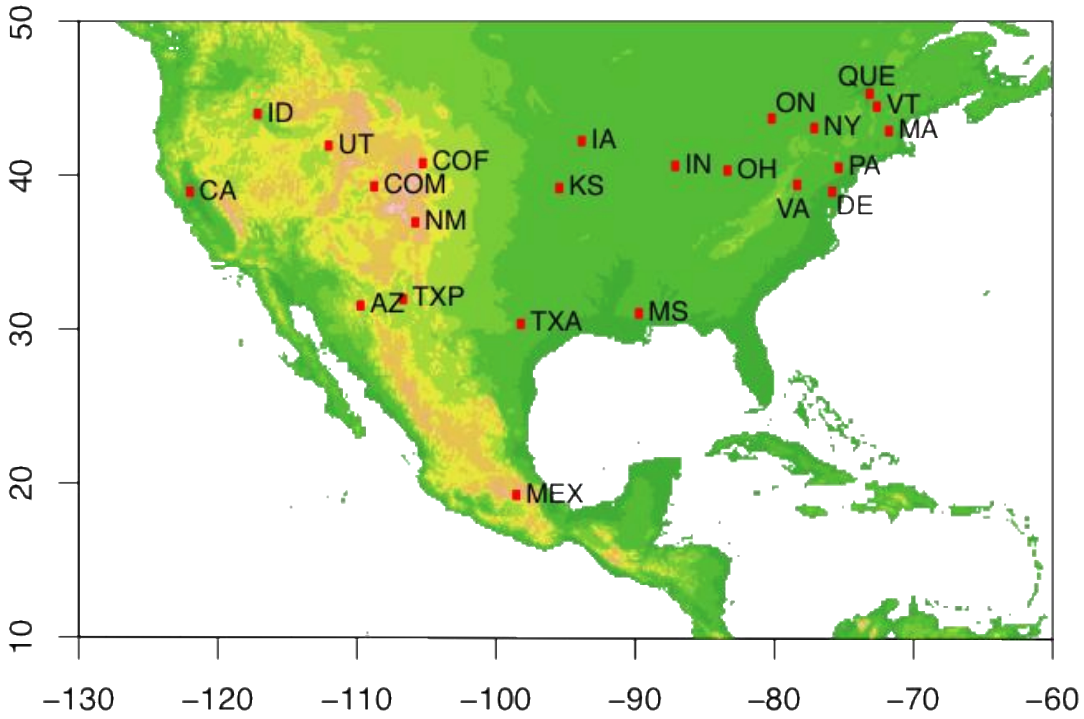


Figure 2.2. Geographic distribution of sampled sites in Mexico, United States and Canada. Abbreviations show populations from Mexico (MX), Texas 'El Paso' (TXP), Arizona (AZ), California (CA), Utah (UT), Idaho (ID), New Mexico (NM), Colorado 'Montrose' (COM), Colorado 'Fort Collins' (COF), Texas 'Austin' (TXA), Iowa (IA), Kansas (KA), Indiana (IN), Delaware (DE), Mississippi (MS), Ohio (OH), Massachusetts (MA), Virginia (VA), Quebec (QU), Vermont (VT), Ontario (ON), New York (NY), Pennsylvania (PA).

were positive for the presence of microsatellite were extracted and Sanger sequenced using M13 primers. For the pyrosequencing method, we ligated post-enrichment PCR products to 454 A and B adapters for sequencing in half 454 run shared with libraries from seven other taxa. Each library was identified with a unique sequence called multiplex identifier (MID). Adapters and MIDs were trimmed and cleaned reads were assembled into unique contigs using CodonCode Aligner 4.0.2 (CodonCode Co.). We scanned all sequences forsimple tandem repeats using the software Repeat-Masker (Smit et al. 2010). Primer designed and testing were conducted as described in López-Uribe *et al.* (2013).

We amplified microsatellite markers using the Qiagen PCR Multiplex Kit with forward primers fluorescently labeled with one of the following dyes: 6-FAM (blue), VIC (green), NED (black), and PET (red). All reverse primers had the six-base **"**pigtail**"** (GTTTCT) added to the 5'-end to reduce stutter in electropherograms. We designed two microsatellite multiplex sets of three loci based on optimal annealing temperatures (Table 2.1). Multiplexed PCR products were diluted and mixed with GeneScan-500 LIZ and Hi-Di Formamide for genotyping on an Applied Biosystems 3730xl DNA Analyzer. Electropherograms were visualized and genotyped in the software PeakScanner v.1.0 (Applied BioSystems). We assessed repeatability of microsatellite genotyping by reanalyzing 5% of our samples.

To estimate the genetic variability of the developed microsatellite markers, we calculated allele richness (*Ar*) and expected heterozygosity (*He*) in GenoDive 2.0b24 (Meirmans & van Tienderen 2004). We tested for the presence of Hardy-Weinberg within loci and linkage disequilibrium between loci using the permutation tests implemented in Genepop v.4.2 (Raymond and

Table 2.1. Descriptive summary of the amplification conditions and variability of the six microsatellite markers isolated from *Peponapis pruinosa*. For amplification conditions we describe: Primer sequences, repeat motif, annealing temperature (Opt Ta), reference allele size (Size), fluorescent tab (Tag), microsatellite multiplex set (Set). For genetic varibility we describe: percent missing data (NA%), number of alleles (N_a), expected heterozygosity (H_E) and F_{IS} statistic.

Locus	Primers	Motif	Opt T_a	Size	Tag	Set	NA%	N_a	H_E	F_{IS}
PP4	F: AAC GCA TCG CCG CTT TGT G R: CGA CCT CTC ATT AAT TCC CCT CAA	$(TC)_{13}$	55.0	213	VIC	II[a]	0.030	11	0.452	-0.833
PP8	F: GGC CCT CGA CTA TCC CCA CTC T R: GTA AAG AAC GCC AAA GAA ATC CCT AAA G	$(TTC)_9$	54.5	146	NED	I[b]	0.028	7	0.049	-12.925
PP11	F: GGA CGA AAC GAA CAC GAT AGG ATT R: GAC CGT GGAAAG ATA AGC GAG AAG	$(AC)_{11}$	54.1	191	FAM	I	0.044	7	0.355	-1.130
PP356	F: CAA GGT GGC GAA GAA AAG AAA G R: TCG AAG GGT GAC CAG AGA AAA C	$(GA)_{11}$	56.4	128	VIC	I	0.029	10	0.354	-1.133
PP420	F: CGT TCG GTC TCG TAC TTT ATC GTT R: TGC CAA GGT CAC AGT ATT ATT CGT T	$(CA)_{10}$	54.4	375	PET	II	0.089	9	0.452	-0.810
PP588	F: CGG CAT TTA CGA ACG GAG GAA R: AGT GCT AAG GAG TGA TTG CTG TCT ACG	$(GT)_9$ $(TTGT)_3$	56.2	423	FAM	II	0.049	14	0.584	-0.371

[a] Optimal annealing temperature for multiplex I was 56ºC

[b] Optimal annealing temperature for multiplex II was 58ºC

Rousset 1995). Presence of null alleles was investigated as the percentage of missing data in the haploid males.

Within population genetic diversity

We assessed population genetic diversity estimates as allele richness (*Ar*), expected heterozygosity (*He*) and Shannon diversity index (*H'*) standardizing for unequal sample sizes in MSA v4.05 (Dieringer and Schlötterer 2003). We analyzed both males and females in the same dataset but treated males asdiploid inbred lines. We visualized geographic patterns of genetic diversity by spatially interpolating *Ar*, *He*, and *H'* using a thin plate spline as implemented in the R package 'fields' (Furrer et al. 2011). To test for statistical significance between genetic diversity and geographic distance, we regressed population genetic diversity estimates onto linear geographic distance from the area with highest genetic diversity using the R package 'lm' (R Core Team 2013).

Population structure

We used 10k permutations to detect deviations from Hardy-Weinberg equilibrium by calculating the Fis statistic in the software FSTAT 2.9.3.2 (Goudet 1995). To determine if dispersal is restricted by geographic distance, we tested the hypothesis of isolation by distance using a Mantel test with a Reduced Major Axis regression method as implemented in IBDWS (Jensen *et al.* 2005). Population differentiation was estimated as Fst/(1-Fst), using the unbiased Weir & Cockerham estimator, and regressed onto log-transformed linear geographic distance as recommended by Rousset (1997) for a two-dimension isolation by distance analysis.

To determine the number of clusters that best describe our data, we used the discriminant analysis of principal components (DAPC) implemented

in the package ADEGENET v.1.3-9.2 for R (Jombart *et al.* 2010). DAPC is a multivariate approach that identifies clusters of genetically related organisms by partitioning genetic variability into clusters that maximize between-group and minimize within-group differentiation. Unlike Bayesian clustering algorithms such as STRUCTURE that use an admixture model to partition each genotype into membership to populations, DAPC derives membership probability from the position of the genotypes on the discriminant factors. Thus, this multivariate approach does not assume Hardy-Weinberg equilibrium and it is more powerful at finding the existing population structure. Because individual membership probability changes with the number of PCA retained, we used the alpha score function to chose the optimal number of principal components for the analysis of our dataset (Jombart *et al.* 2010). We performed the DAPC analysis using the number of sampled populations as the prior representing the maximum number of possible clusters.

Demographic parameter estimation

We used the coalescent-based approach incorporated in MSVAR to estimate demographic parameters of changes in population size of *P. pruinosa* across North America. Using an MCMC approach, MSVAR estimates the posterior probability distribution of three demographic parameters (1) effective population size of the ancestral population (N0), (2) effective population size of the current population (N1), and (3) time since the population started changing in size. Wide uniform priors were chosen for all parameters (Table 2.2) to allow us to compare parameter estimates from each run in different populations. We assumed a linear change in population size and a step-wise mutation model for the microsatellite. We excluded locus PP588 from this analysis due to presence of irregular mutation steps.

Table 2.2. Prior distribution for model parameters of MSVAR v. 1.3 analysis. Model parameters are current population size (*N0*), ancestral population size (*N1*), mutation rate (*μ*), and time since population size change (*ta)*. Prior of model parameters have a log-normal distribution (*M* and *V*). Hyper-prior distributions for *M* (α and σ) and *V* (β and τ) have normal distributions.

	M	***V***	***α***	***σ***	***β***	***τ***
N0	10	2	10	4	0	0.5
N1	10	2	10	4	0	0.5
μ	-3.5	1	-3.5	2	0	0.5
ta	4	2	4	3	0	0.5

We conducted one analysis with the complete dataset excluding populations from California, Utah and Idaho that were clearly differentiated based on the DAPC analysis. For the entire dataset, we randomly resampled alleles to reduce the number of sampled chromosomes to 180. We also independently analyzed populations where we sampled more than 30 chromosomes using the same priors to be able to compare relative values of the estimated demographic parameters. For each data set, we ran 4 independent chains of 10^8 recording parameter values every 10^3. In theabsence of convergence, we ran longer chains of 10^9 with parameter values recorded every 2500. The first 10% of steps of all chains were discarded as burn-in. We analyzed the MSVAR outputs using the R packages ‘locfit’, ‘coda’ and ‘runjags’. Chain convergence was assessed by computing the multivariate Gelman and Rubin**’**s diagnostic (Brooks and Gelman 1998) on at least three independent Markov chains. A Gelman and Rubin**’**s diagnostic >1.1 indicates poor convergence, thus longer chains were run. In addition, we plotted and visually inspected posterior densities of each parameter for consistency in mode and shape. For parameter estimation, we combined all chains that reached

convergence to obtain the mode and 90% highest probability density (90%HPD) limits.

Reconstruction of colonization history

We built a neighbor-joining (NJ) tree to investigate the relationship between populations based on genetic distance as a first test to differentiate between the proposed hypotheses about the colonization history of *P. pruinosa*. We used the pairwise Cavalli-Sforzar chord distance (Dce) between populations in MSA v4.05 (Dieringer and Schlötterer 2003) as the measurement of genetic distance. Using the population from Mexico as the root, we built a rooted NJ tree with 10000 bootstrap replicated in the R package 'ape' (Paradis et al 2004).

We tested different hypotheses about the colonization of *P. pruinosa* from Mexico across North America using an Approximate Bayesian Computation (ABC) framework in the software DIYABC v.2.0.4 (Cornuet *et al.* 2008; Cornuet *et al.* 2010). After the DAPC analysis, we identified the populations from California and Idaho+Utah as highly divergent from the rest of the populations. Therefore, besides investigating the routes of colonization to the northeast of NA, we aimed to differentiate the number of colonization events from Mexico to western North America. We tested three different scenarios: (SI.1) California and Idaho+Utah were independently colonized from individuals in Mexico; (SI.2) California was colonized by individuals from Idaho+Utah that diverged from the population in Mexico; and (SI.3) Idaho+Utah was colonized by individuals from California that diverged from the population in Mexico (Fig. 2.3 A-C). The second set of demographic scenarios (SII) had the goal of differentiating between the alternative hypothesis about

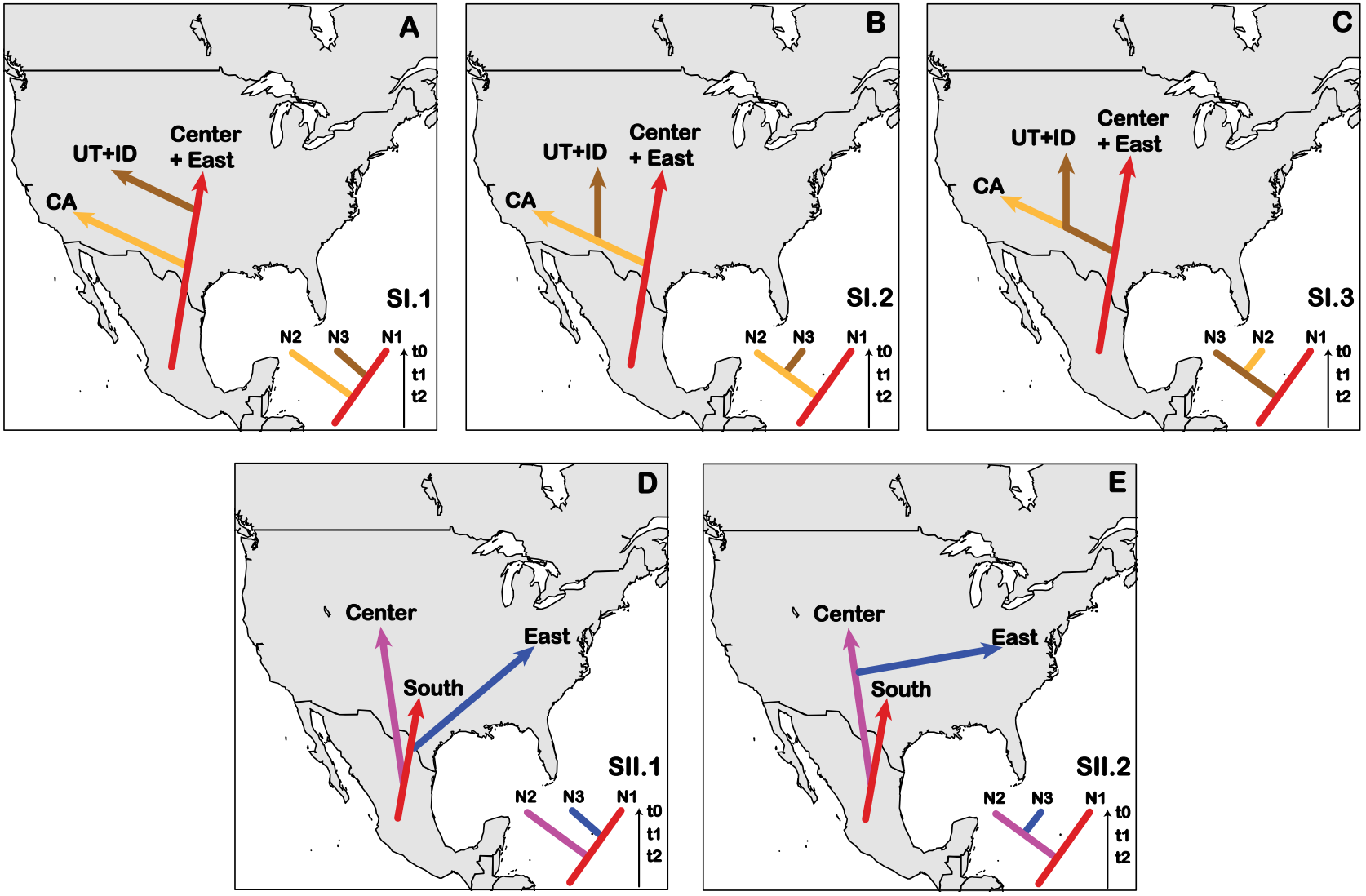

A
UT+ID
Center
+ East
CA
SI.1
N2
N3
N1
t0
t1
t2
B
UT+ID
Center
+ East
CA
SI.2
N2
N3
N1
t0
t1
t2
C
UT+ID
Center
+ East
CA
SI.3
N3
N2
N1
t0
t1
t2
D
Center
East
South
SII.1
N2
N3
N1
t0
t1
t2
E
Center
East
South
SII.2
N2
N3
N1
t0
t1
t2

Figure 2.3. Hypothesized scenarios of colonization of North America by *Peponapis pruinosa* tested by the approximate Bayesian computation (ABC) approach. Maps of the routes of colonization and the drawings of the demographic scenarios are shown for each hypothesis. (A-C) Demographic scenarios of colonization events from Mexico to the west of North America. (A - SI.1) California and Idaho+Utah were independently colonized from individuals in Mexico; (B - SI.2) California was colonized by individuals from Idaho+Utah that diverged from the population in Mexico; and (C - SI.3) Idaho+Utah was colonized by individuals from California that diverged from the population in Mexico. (D – E) Demographic scenarios of the hypothesized colonization of the northeast of North America: (D - SII.1) the colonization of the northeast was an independent event as a result of the range expansion of the lineage of the south (Mexico); or (SII.2) the colonization of the northeast was a result of the range expansion of the lineage that colonized the center of the US from the south.

the colonization of *P. pruinosa* to the northeast of North America: (SII.1) the colonization of the northeast of NA was the result of a range expansion from from the Midwest; or (SII.2) the northeast of NA was colonized through a range expansion from northeastern Mexico (Fig. 2.3 D-E).

We assumed a Generalized Stepwise Mutation (GSM) model (Estoup *et al.* 2002) to simulate mutations at the microsatellite loci. Mean mutation rate (μ) was drawn from a uniform prior distribution ranging from 10^{-5} to 10^{-3}. Number of repeats added or removed from the microsatellite motif in each mutation step (*P*) was drawn from a uniform distribution between 0.1 and 0.3.For each individual microsatellite loci, mutation rate and number of repeats added per mutation step were allowed to vary using a gamma distribution (Table 2.3). Single nucleotide polymorphisms introduced in microsatellite repeat motifs were introduced in the model using the parameter SNI described

Table 2.3. Prior distribution for demographic parameters in scenarios I and II of ABC analysis in DIYABC. Mutation rates priors were the same for both sets of scenarios.

Parameter	Distribution	Min	Max	Shape
Priors for demographic parameters – Scenarios I				
N1	Uniform	10	10000	-
N2	Uniform	10	10000	-
N3	Uniform	10	10000	-
t1	Uniform	10	10000	-
t2	Uniform	10	10000	-
Priors for demographic parameters – Scenarios I				
N1	Uniform	10	50000	
N2	Uniform	10	20000	
N3	Uniform	10	20000	
t1	Uniform	1000	10000	
db	Uniform	100	10000	
Nf1	Uniform	10	10000	
t2	Uniform	1000	10000	
Nf2	Uniform	10	10000	
Priors for microsatellite mutation rate model				
μ	Uniform	10^{-5}	10^{-3}	-
μ	Gamma	10^{-5}	10^{-2}	2
P	Uniform	10^{-1}	$3x10^{-1}$	-
P	Gamma	10^{-2}	$9x10^{-1}$	2
SNI	Log-Uniform	10^{-8}	10^{-5}	-
SNI	Gamma	10^{-9}	10^{-4}	2

with a log-normal distribution ranging from 10^{-5} to 10^{-3}. Each allele had 40 possible allelic states with contiguous numbers of repeats. To investigate the number of colonization events to the western distribution of *P. pruinosa* (Fig. 2.3 A-C), we used wide uninformative priors due to the lack of information about the dates of these events. Prior distribution for population sizes (*N1, N2, N3*) and time of divergence (*t1, t2*) were drawn from a uniform distribution (101-104). To differentiate between the two possible routes of colonization to the northeast of NA (Fig. 2.3 D-E), priors for demographic parameters were defined based on information from the archeological evidence of domestication of *C. pepo* and information from the MSVAR analysis. Ancestral and current population size of the population from the south were bounded between 101 and 5x104, while the population sizes of the Midwest and the northeast populations were bounded between 101 and 2x104. Priors for the time of divergence were set to wide distributions (101-104) to allow for distinctions between the different scenarios. We bounded population sizes during bottlenecks between 101-103. Priors for all parameters were drawn from uniform distributions.

For each scenario we simulated 2x106 datasets. Within populations, we compared the following summary statistics: mean number of alleles per locus (NAL), and mean expected heterozygosity (He). Between populations, we compared NAL, He, Fst and shared allele distance (DAS). The posterior probability of each competing scenario was estimated using a logistic regression on 10000 simulated datasets. We chose the best scenario based on the highest significant probability values with non-overlapping 95% confidence intervals. Type-I and Type-II errors were estimated by simulating 500 datasets of the best and each alternative scenario, respectively (Cornuet

et al. 2010). Posterior distributions for demographic parameters under the best scenario were estimated carrying out a linear regression on the 1% closest of 105 datasets. We tested the performance of the best demographic scenario by trying to reproduce the observed data with 105 pseudo-replications using the model checking procedure implemented in DIYABC v.2.0.4 (Cornuet *et al.* 2008; Cornuet *et al.* 2010).

Results

Microsatellite development and variability

We designed primers for 24 DNA sequences (11 from the cloning protocol and 13 from the 454 protocol). However, six of these primer pairs did not produce detectable PCR products and 12 were monomorphic. The six variable microsatellite loci showed allele richness ranging from 7 to 14 and expected heterozygosity ranging from 0.046 to 0.584 (Table 2.4). Missing data was below 0.1% for all loci (Table 2.4). We detected very low average genotyping error 0.02%. All loci deviated from Hardy-Weinberg equilibrium (p=0.0001) but no significant linkage disequilibrium was detected between loci (all p-value<0.000). Due to the haplodiploid nature of our dataset, we were unable to check for the presence of null alleles using tests based on Hardy-Weinberg equilibrium. However, the low rate of missing data (~0.5%) in our male dataset suggests that the presence of null alleles was negligible.

Within population genetic diversity

The populations from Mexico, Arizona and Texas ‘El Paso’ showed the highest expected genetic diversity among the sampled populations (Table 2.4). In contrast, the population from California displayed the lowest genetic diversity

Table 2.4. Summary table of genetic diversity indices per population. Allele richness, expected heterozygosity, Shannon diversity index (H'), and inbreeding coefficient. All estimates were corrected for unequal sample size.

Population	Abbreviation	N	Ar	Hs	H'	Gis
Arizona	AZ	7	3.997	0.6188	1.1152	-0.332
California	CA	132	1.26	0.0844	0.1491	0.866
Colorado (Fort Collins)	COF	31	2.716	0.4629	0.8316	-0.812
Colorado (Montrose)	COM	32	1.916	0.3135	0.5026	-2.383
Delaware	DE	30	2.037	0.3903	0.5967	-1.375
Iowa	IA	45	2.075	0.4365	0.6452	-1.146
Idaho	ID	29	1.259	0.0649	0.1218	-20.379
Indiana	IN	22	2.003	0.3438	0.5394	0.537
Kansas	KS	9	1.667	0.3669	0.4364	-2.296
Massachusetts	MA	40	2.048	0.4218	0.6311	-0.367
Mexico	MEX	43	3.557	0.6241	1.1971	-0.767
Mississippi	MS	27	1.956	0.3714	0.5563	-0.538
New Mexico	NM	16	2.079	0.4354	0.6249	-1.153
New York	NY	111	2.003	0.3399	0.5508	-1.193
Ohio	OH	44	2.06	0.4063	0.6169	-1.473
Ontario	ON	32	1.822	0.2853	0.445	-1.825
Pennsylvania	PA	121	2.132	0.394	0.6356	-0.746
Quebec	QU	60	2.033	0.3312	0.5463	-1.094
Texas (Austin)	TXA	15	2.583	0.4211	0.7383	-0.809
Texas (El Paso)	TXP	13	3.711	0.6915	1.2279	-2.674
Utah	UT	36	1.957	0.2864	0.4859	-0.699
Vermont	VT	14	2.86	0.5222	0.8712	-1.119

(*Ar* = 1.26; *He*= 0.084; *H'*=0.15) and all loci characterized by a high frequency allele varying between 0.7 and 1. The spatial interpolation of the genetic diversity indices showed a clear pattern of decreasing genetic diversity with increasing latitude (Fig. 2.4). However, among the three populations with the highest genetic diversity, we only found a significant decrease between geographic distance and genetic diversity from the population in Mexico (*Ar*: r2=0.345, p<0.005; *He:* r2=0.309, p<0.01; *H'*: r2=0.385, p<0.005) but not from the populations in Arizona and Texas 'El Paso' (Fig. 2.5). Slopes were steeper on allele richness (r2=0.385; p=0.001) than on expected heterozygosity (Fig. 2.5).

Population genetic structure

Out of the 138 possible population-locus combinations, 19 Hardy–Weinberg tests could not be calculated because of the presence of a fixed allele. Forty-nine of the remaining tests showed a significantly higher Fis value indicating deviations from Hardy–Weinberg equilibrium due to deficiency of heterozygotes. In the population from Mexico, all loci showed Hardy–Weinberg equilibrium (p=0.0001). The overall Fst value across all populations 0.457. We detected a significant pattern of isolation-by-distance (Pearson's r=0.328, slope b= 0.8; p<0.001; Fig. 2.6), even though the linear relationship between the variables was low (r2= 0.1).

The DAPC results show the number of clusters that best describe our data (Fig. 2.6). The spline interpolation of the alpha scores indicated that the optimal number of principal components was 35 (Fig. 2.7). The DAPC analysis identified three genetic groups within our data. One group includes all the individuals from California, which was the population with the least genetic diversity (*He*= 0.085) and most inbreeding (Fis= 0.866; Table 2.4). The second

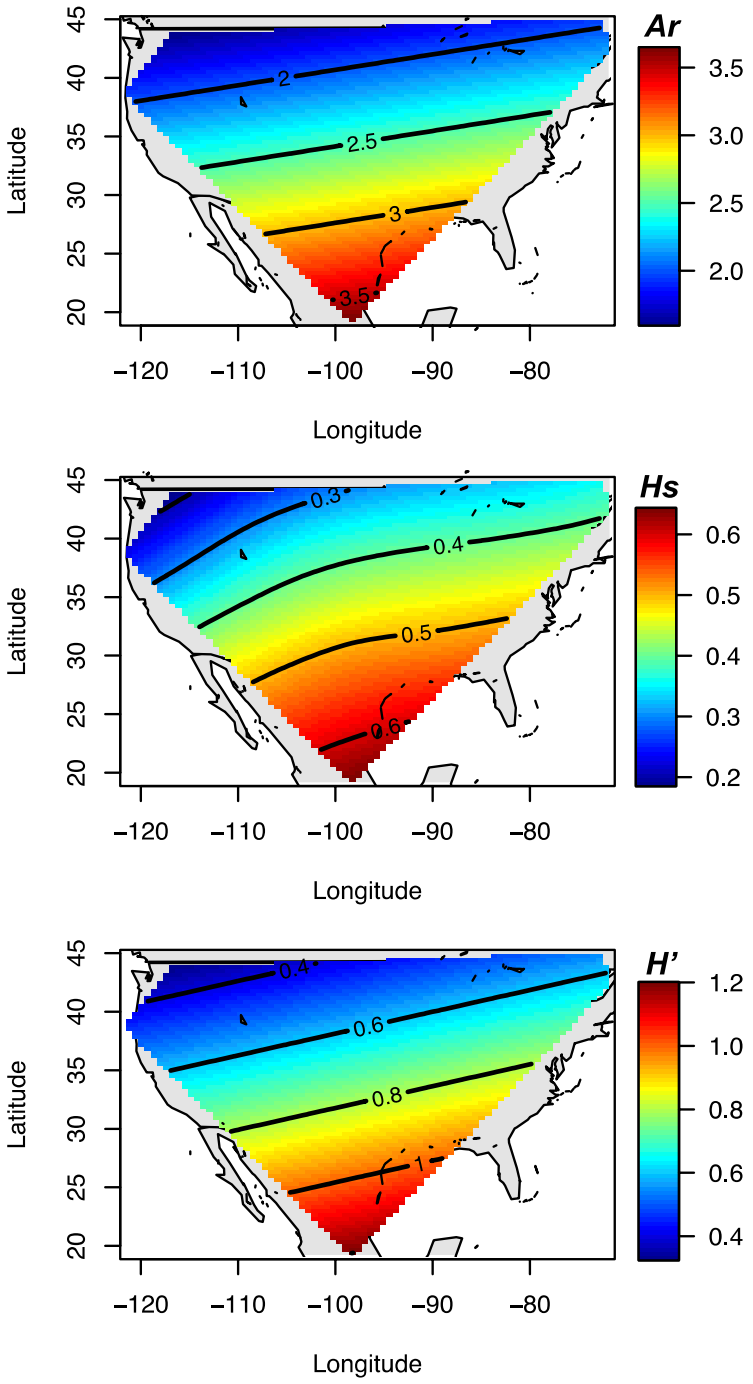


Figure 2.4. Genetic diversity indices of *Peponapis pruinosa* in North America. Thin spline interpolated values for allele richness (A_r), expected heterozygosity (H_e) and Shannon Diversity Index (H').

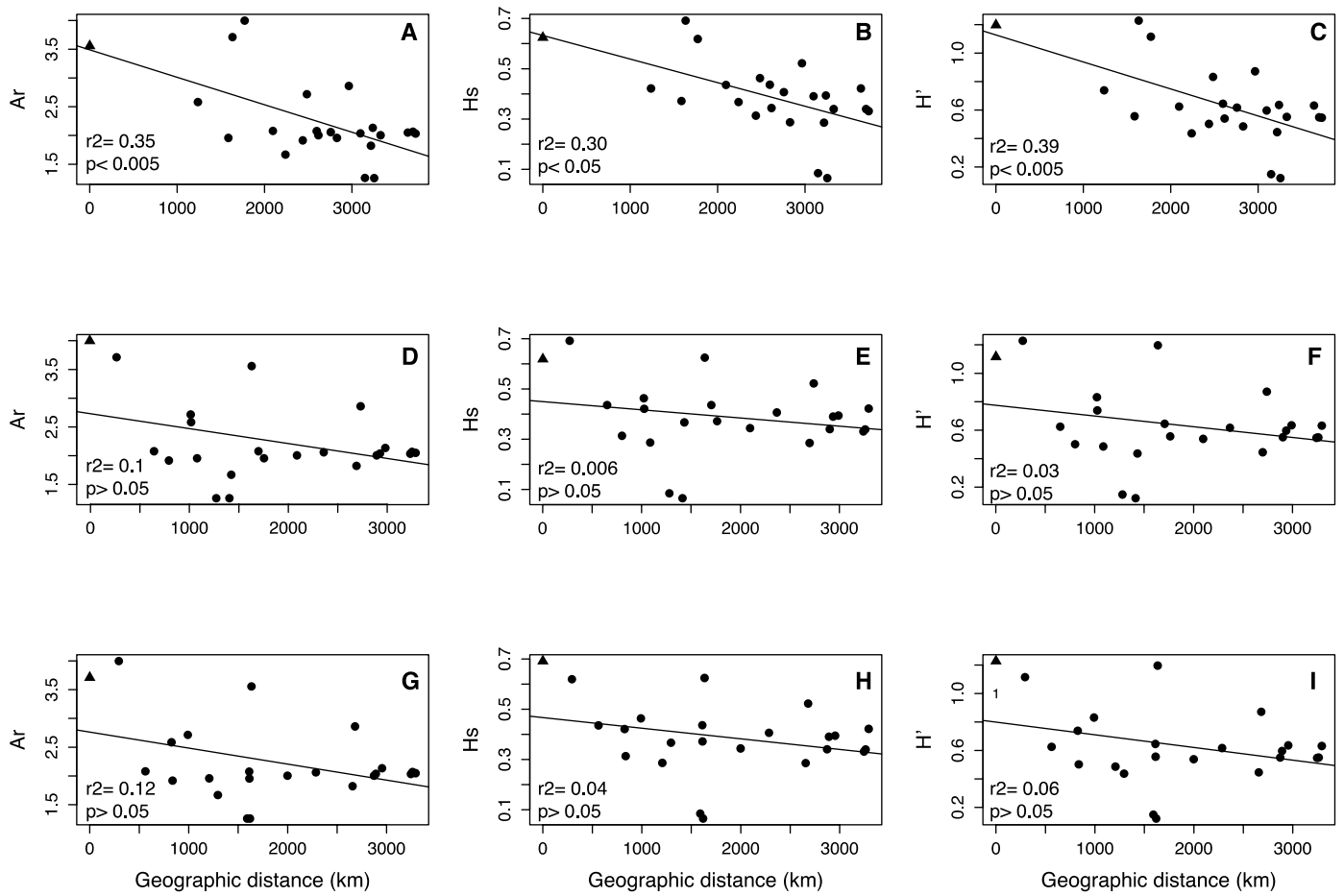


Figure 2.5. Linear regression showing decreasing genetic diversity with geographic distance from two hypothesized centers of origin: Mexico (upper panel; A-C), Arizona (middle panel; D-F) and Texas El Paso (lower panel; G-I). Genetic diversity measured as allele richness (Ar), expected heterozygosity (Hs), and Shannon Diversity Index (H').

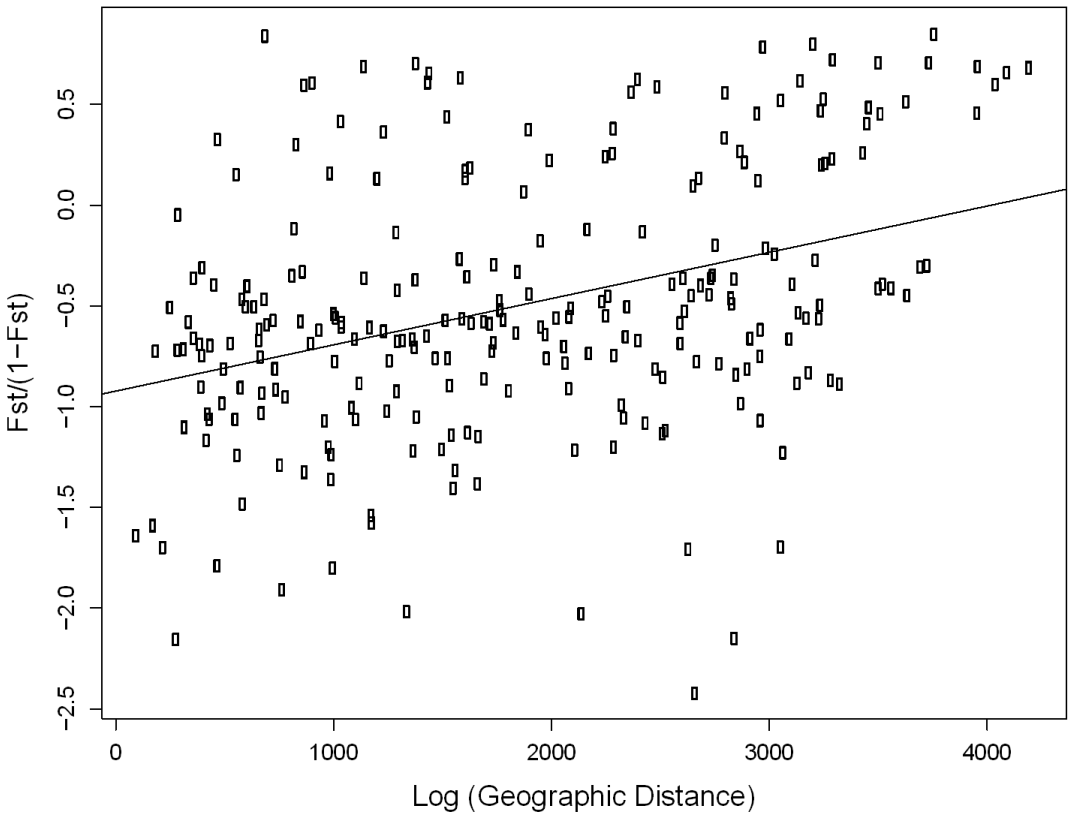


Figure 2.6. Patterns of isolation-by-distance for populations *Peponapis pruinosa*. Genetic diversity measured as pairwise Fst/(1-Fst) and linear geographic distances (km) is in the log scale (Pearson's r=0.328, slope b= 0.8; p<0.001).

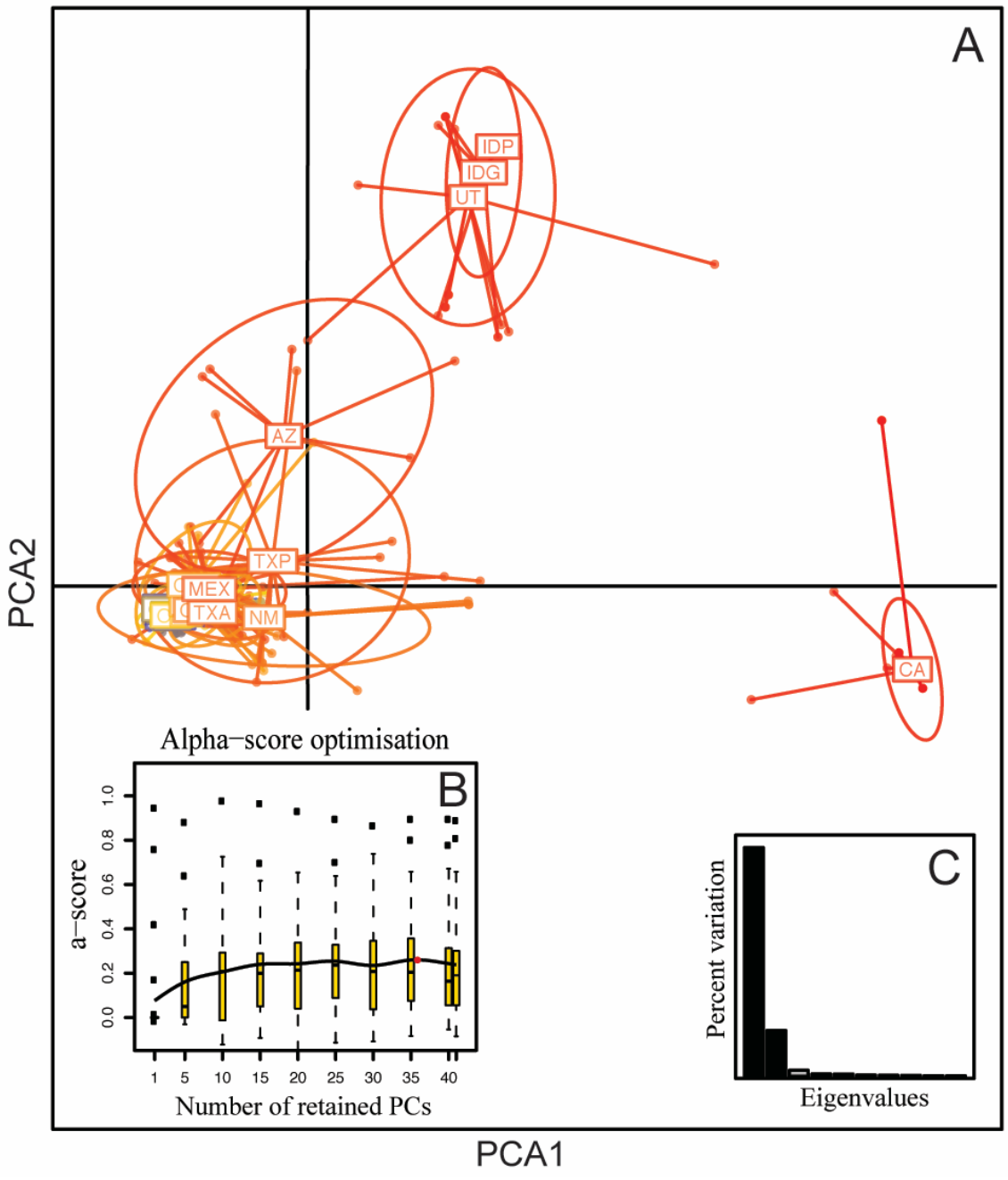


Figure 2.7. (A) Scatterplot of the discriminant analysis of principal components (DAPC). The scatterplot shows the first principal component, grouping individuals by populations. Right inset (B) shows the trend of the alpha scores use to choose the optimal number principal components (n=35). Left inset (B) indicates the eigenvalues included in the analysis.

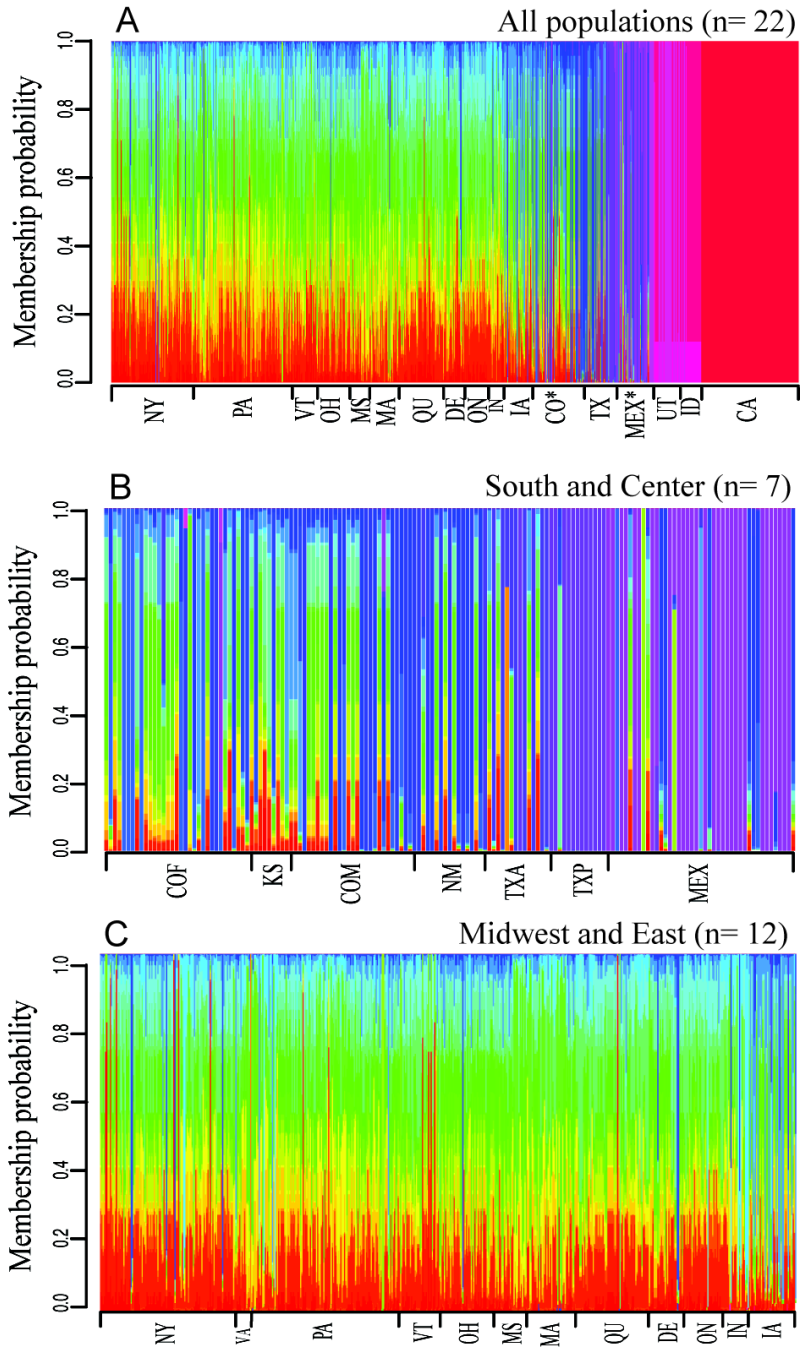


Figure 2.8. Clustering analysis based on the discriminant analysis of principal components (DAPC). Each vertical bar represents an individual that is grouped by the population of origin (x-axis). Colors represent the probability of belonging to a genetic cluster quantified by the membership probability (y-axis). Results from all populations (A), populations from the south an center (B), and the Midwest and East (C) are shown separately.

cluster group includes the individuals from Idaho and Utah, and the third group includes the rest of the individuals (Fig. 2.8). Overall, individuals from the west of the US, center of the US, and Mexico are likely to belong to the same deme (Fig. 2.8 A-B). Membership probabilities were 1 for all individuals collected in California. The membership probabilities for the populations of Idaho, Utah and Arizona were also high, making these individuals genetically differentiated from individuals from other populations (Fig. 2.8A). Most individuals from Mexico, Texas El Paso (TXP) and Texas Austin (TXA) have high membershipprobabilities from their populations of origin but some show admixed origin (Fig. 2.8B). The rest of the individuals (north of Texas and east of the Rocky Mountains) show high levels of admixture, making them one homogenous genetic unit (Fig. 2.8C). Individuals from Colorado, Kansas and New Mexico (center) have admixed individuals from the east coast and individuals that genetically belong to Mexico-Texas cluster (Fig. 2.8B).

Demographic parameters and colonization history

We investigated signatures of population range expansion in *P. pruinosa* by estimating the relative difference in effective population size between the current (*N0*) and ancestral populations (*N1*). All coalescent simulations included in our parameter estimation converged, as indicated by the Gelman-Rubin convergence statistic (<1.1). Furthermore, all independent chains provided consistent marginal posterior probability distribution (Fig. 2.9). Comparisons between current and ancestral population sizes consistency showed a dramatic reduction in population size that varied between five and ten-fold in the log scale. Estimates of time since the population size change event were highly dependent on priors, thus are not reported.

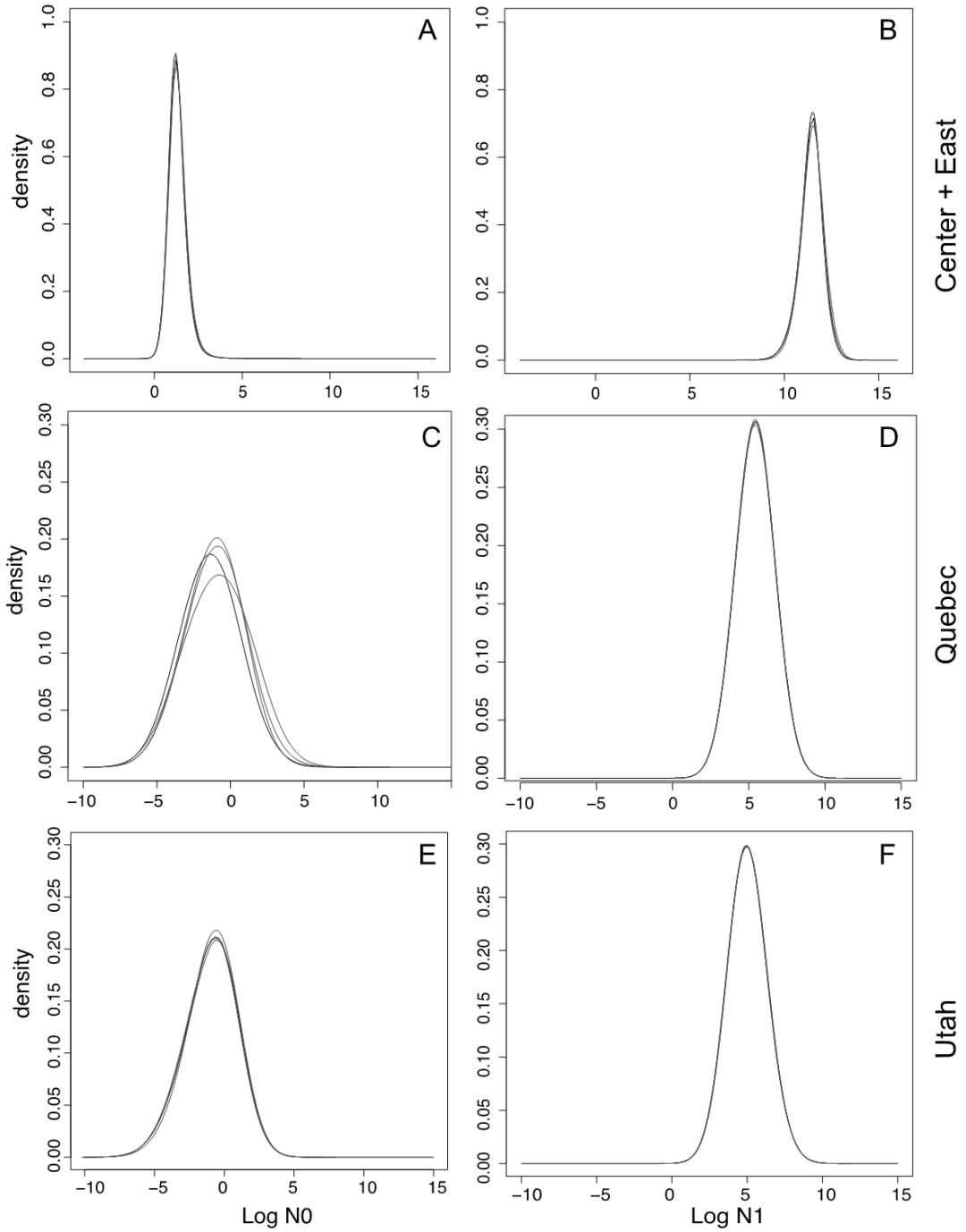


Figure 2.9. Demographic parameter estimation of *Peponapis pruinosa.* Marginal posterior probability density of ancestral (N0, left panel) and current (N1, right panel) population size in the log-scale for Center+East, Quebec and Utah populations. Black and grey lines show density estimation for all and independent runs, respectively.

The distance NJ tree revealed the distance-based relationships between populations providing information about a possible route of colonization of *P. pruinosa* into NA. The topology of the tree supports the presence of one western lineage grouping the populations from Texas 'El Paso', Arizona, Utha, Idaho and California (orange, Fig. 2.10). The lineage that groups the remaining populations places populations from the center at the base (purple, Fig. 2.10), and it separates three different groups: one clustering populations from the Midwest, and two separate groups in the northeast (green, Fig. 2.10). This tree topology provides partial support to the Midwest hypothesis given that populations in the northeast are genetically closer to the populations in the Midwest. However, the interpretation of this tree topology is limited by the fact that no bootstrap support was found for any of the branches.

Comparisons between the three colonization hypotheses to western North America indicated that the scenario of two independent colonization events from Mexico (Fig. 2.3 A; SI.1) received the lowest support (Table 2.5). Therefore, only one lineage of *P. pruinosa* colonized the west of US and it is the ancestral population of the two well differentiated lineages found in California and Utah+Idaho. However, distinguishing the ancestral population between these two lineages (Fig, 2.3 B-C; SI.2 vs. SI.3) was not possible and both scenarios showed equal posterior probabilities (Table 2.5). The ABC analysis for the two possible demographic scenarios of the colonization history of *P. pruinosa* to the northeast of NA highly supported scenario SII.2 (Fig. 2.3 E). The posterior probability of the demographic scenario of colonization of the northeast of NA after the second domestication event of *C. pepo* (Fig. 2.1) is highly supported by the data (90%, Table 3). Evaluation of the performance of

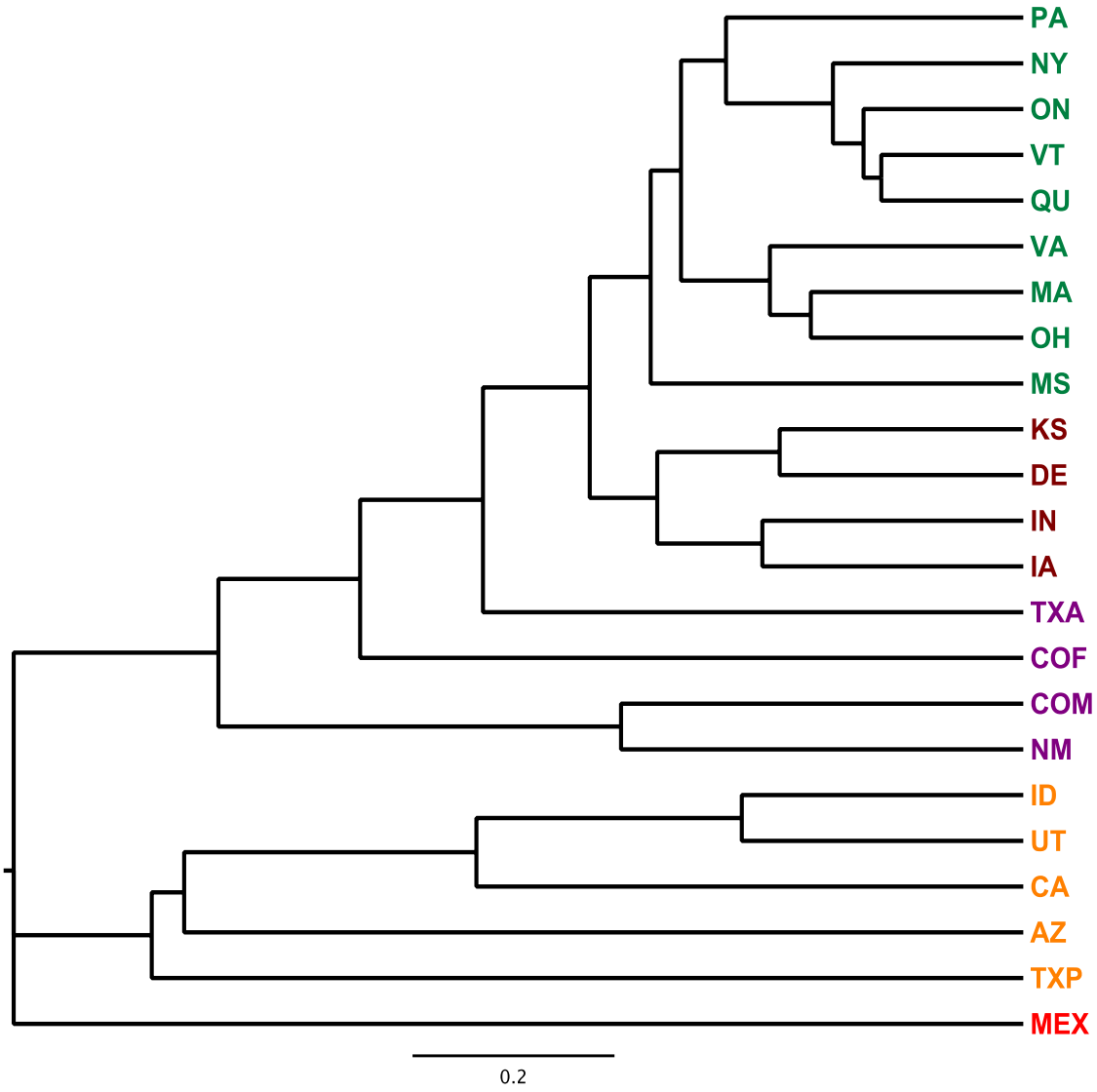


Figure 2.10. Neighbor-joining tree based on the Cavalli-Sforza chord distance. Tree is rooted with the population from Mexico (MX). In orange, populations from the west of NA [Texas 'El Paso' (TXP), Arizona (AZ), California (CA), Utah (UT), Idaho (ID)]. In purple, populations from the center of the United States [New Mexico (NM), Colorado 'Montrose' (COM), Colorado 'Fort Collins' (COF), Texas 'Austin' (TXA)]. In burgundy, populations from the Midwest [Iowa (IA), Kansas (KA), Indiana (IN)] + Delaware (DE). In green, populations from the east and northeast of NA [Mississippi (MS), Ohio (OH), Massachusetts (MA), Virginia (VA), Quebec (QU), Vermont (VT), Ontario (ON), New York (NY), Pennsylvania (PA)].

Table 2.5. Model choice for colonization scenarios based on the Approximate Bayesian Computation (ABC) analysis.

	SI.1	SI.2	SI.3	SII.1	SII.2
Posterior Probability	23.2%	38.2%	38.6%	9.6%	90.4%
	[0.0-60.20]	[0.0-80.79]	[0.0-81.27]	[0.0-35.42]	[64.58-100.0]
Model performance of best scenario					
Type I error	90%			19.2%	
Type II error		43%	32.8%		20%
Number of summary statistics significantly different than the observed data					
P < 0.05	3	1	2	4	1
P < 0.01	0	0	0	1	0
P < 0.001	0	0	0	0	0

each model agrees with these results. For the first set of demographic models, 90% of the pseudo-replicates of the simulated dataset identified SI.1 as the most likely scenario. This high type I error indicates that the data is uninformative to differentiate between the hypothesized scenarios. For hypotheses SII.1 and SII.2, the statistical power to differentiate between models was higher. Of the 500 pseudo-replicates, 20% were assigned to the wrong model (type I error), while the statistical power was 80% (1- β [type II error]). Parameter estimates of the best-supported demographic model are provided in supplementary material (Table 2.6).

Discussion

Our results strongly support the evidence that the current distribution of the squash bee *P. pruinosa* is the result of a spatial range expansion from tropical to temperate areas of NA. The presence of the highest genetic diversity in the population of Mexico and the significant reduction of genetic diversity with increasing distance from Mexico, indicate that the center of origin of *P. pruinosa* is in the southern range of its current distribution.

Understanding the demography, routes of invasion and life-history traits of biological invaders is key to generate and test hypotheses that predict which species will likely be successful colonizers of new areas. We reconstructed the demographic history of the range expansion of *P. pruinosa* by testing several scenarios of colonization from Mexico into NA. For the colonization of the west, the ABC analysis supported the scenario of a single ancestral lineage diverging into the populations of California and Utah+Idaho, instead of two independent colonization events (Fig. 2.3 A-C; Table 2.5). On the other hand,

Table 2.6. Demographic point estimates of the best-supported models for sets of demographic scenarios I and II based on the marginal posterior distribution of each parameter.

Parameter	Mode	2.5 quantile	97.5 quantile
Priors for demographic parameters – Scenarios I			
N1	3200	945	8820
N2	102	38	673
N3	193	72	1530
t1	2130	389	7330
t2	8430	4670	9720
Priors for demographic parameters – Scenarios II			
N1	19100	4430	36800
N2	2670	885	17000
N3	2610	841	17600
t1	219	125	2920
t2	2900	1280	9440

the colonization of the northeast from populations in the Midwest was highly supported by the data with a high statistical power to differentiate between the two proposed scenarios (Table 2.5). Therefore, our results corroborate previous studies based on wing morphometrics that suggested that the northeast of NA was colonized from the west and not from Mexico after the first domestication event 10000 y.a. (Bischoff *et al.* 2009).

All the tested demographic scenarios of range expansion require a host-plant shift from *C. foetidissima* to *C. pepo* outside the range of the native host-plant. Therefore, domestication and human-mediated range shifts of *C. pepo* are the main driver of range shifts in this species. Some species of the genus *Peponapis* have narrow host-plant associations limited to collecting pollen from one or two wild *Cucurbita* spp. (e.g. *P. atrata*, *P. citrullina*, *P. melonis*, and *P. timberlakei*) and they tend to show limited geographic distribution (Hurd *et al.* 1971). Unlike the other species in the genus, *P. pruinosa* has experienced multiple ecological shift between wild and cultivated *Cucurbita* allowing this bee species to become widespread outside the range of its wild host. This ecological plasticity is probably one of the most important life-history traits of this species responsible for the successful invasion from tropical to temperate ecosystems.

The DAPC clearly identified three genetic clusters within the sampled individuals of *P. pruinosa* (Fig. 2.6). The individuals from California and Utah+Idaho are genetically differentiated from the rest with high assignment probability to their populations of origin (Fig. 2.7). Even though estimates of the time since divergence were not possible due to the low levels of genetic diversity (uninformative data), the current distribution of the populations from California and Utah+Idaho are outside the range of the native *C. foetidissima.*

This distribution pattern implies a recent colonization of California and Utah+Idaho probably after European settlers started annually cultivating *C. pepo*. However, unlike other recently colonized areas in the northeast, the range expansion to California and Utah+Idaho required dispersal thought geographic barriers due to the topographic complexity of western NA. As a result of severe bottlenecks and lack of gene flow with the source, California and Utah+Idaho populations are under strong influence of drift that can explain the quick divergence from the source population. Furthermore, the combination of genetic isolation and differences in climatic conditions may facilitate rapid divergence through adaptation to local environmental conditions. Morphological differences in vesture coloration gave a different taxonomic name to the *P. pruinosa* population from California (*P. angelica* Cockerell 1902) (Hurd & Linsley 1966). Our genetic data suggest the taxonomic validity of this species name should be re-evaluated. In contrast to the striking differentiation of the two western lineages, all the populations from the Midwest and east of NA showed low genetic differentiation (Fig. 2.7) despite a significant pattern of isolation-by-distance (Fig. 2.5). This significant pattern of isolation by distance suggests that there is equilibrium between gene flow and genetic drift in the populations from Midwest and East of NA (Hutchison & Templeton 1999). Therefore, we conclude that in the absence of topographic complexity populations of this species can maintain drift-migration equilibrium at large geographic scales.

In spite of the clear pattern of spatial range expansion, the coalescent-based estimation of demographic parameters detected severe reductions in effective size in all populations (Fig. 2.9). Therefore, the high census numbers of this bee observed in *Cucurbita* fields throughout NA does not correlate with

the effective sizes of *P. pruinosa* populations. The extremely low levels of genetic variability at the periphery of its range and the unbalanced allele frequency spectrum in all loci indicate severe bottlenecks across the whole distribution of the species (Fig. 2.9). This pattern of depletion of genetic diversity is expected after consecutive bottlenecks if the number of founder individuals is small, gene flow between populations is low, and the reproductive rate of the species is low (Excoffier *et al.* 2009). However, we also detected signals of population decline in the center of origin of the species suggesting that other extrinsic factors besides the spatial range expansion are driving these demographic changes. Because *P. pruinosa* is a host-plant specialist that relies on domesticated species of *Cucurbita* in most of its current distribution, populations are under constant disturbance as a consequence of tillage for crop rotation (Shuler *et al.* 2005). We hypothesize that the reduced genetic variation and evidence of severe bottlenecks in *P. pruinosa* is the result of a combination of the historical spatial range expansion and a meta-population dynamics driven by agricultural practices.

Severe reductions in genetic variability have detrimental effects on populations of haplodiploid insects that have a single locus complementary sex determination (csd) system (Zayed & Packer 2005). In these species, low genetic variability and small effective population sizes increase the frequency of homozygosity at the csd locus leading to increasing production of diploid (sterile) males in the population (Hedrick *et al.* 2006). Based on these conditions, short-term population viability is expected in bee populations that have depleted levels of genetic variability due to severe demographic bottlenecks. Despite the low levels of genetic variability, low effective population size and evidence of strong bottlenecks, *P. pruinosa* is an

abundant species throughout NA that continues expanding its geographic range (Payette & Payette 2003) and provides significant pollination services to crop species in the genus *Cucurbita* (Cane *et al.* 2011; Hurd *et al.* 1974). Given the low frequency of diploid males found in this study (2 out of 438), we infer that *P. pruinosa* possesses mechanisms to avoid or reduce high frequencies of diploid males despite low levels of genetic variability. One possible mechanism is the use of chemical recognition to evade mating with individuals that have the same alleles at the csd locus. Even though this mechanism has not been investigated, the existence of male mate choice through chemical kin recognition has long been known in bees (Smith 1983; Wcislo 1992). Alternatively, sex determination could have a multi-locus genetic basis, making diploid male production possible only in eggs with homozygosity at all loci (Crozier 1971). In any case, this is the second reported case of a successful biological invasion by a solitary bee despite greatly reduced genetic variability (Zayed *et al.* 2007). These studies suggest solitary bees can be effective colonizers of new areas despite severe founder events. However, levels of genetic variability in other successful invasive bee species (e.g. *Osmia cornifrons, Anthidium manicatum*) have not been investigated, thus we cannot assert the ubiquity of this pattern.

Hypothesis testing of different demographic scenarios is necessary to identify the colonization routes of invading organisms. In this study, we used genetic data in combination with historical records on the domestication and range expansion of *Cucurbita* spp. to unravel the demographic history of a host-plant specialist bee. To our knowledge, this is among the first studies to specifically investigate the consequences of plant domestication and range shifts by early human societies in the demographic history of a mutualistic

pollinating species. Our results have important implications for the conservation and management of wild bees for crop pollination. We show that some species can be resilient to the negative effects of low genetic variability. Although, *P. pruinosa* is a successful invader that keeps expanding its geographic range in spite of severe and repeated bottleneck, we do not have evidence of the ubiquity of this pattern in bees. Therefore, studies investigating the precise mechanism of inbreeding avoidance in bees are necessary to better provide management recommendations for the reintroduction of bees in regions where habitat loss and intense agricultural systems has extinguish the native bee pollinator community.

Acknowledgements

We thank all the people involved in the sampling of *Peponapis pruinosa* across Mexico, the United States and Canada: Derek Artz, Jessica Petersen, Ta'i Roulston, Sam Droege, Karen Stickler, Beatriz Moisset, Leif Richardson, Carlos Vergara, Jason Gibbs, John Wenzel, Blair Simpson, Elizabeth Evans, Michael Veit, Andre Payette, Heather Harmon, Katharina Ullmann, Karen Goodell, John Purdy, Sheena Shidu, Christopher Mayack, Jack Neff, Liz Maynard, Jennifer Thomas, and Victor Gonzalez. We also thank Steve Bogdanowicz for help with microsatellite development and Christine Santiago for assistance with microsatellite genotyping. Funding for this study was provided by grants from the Grace Griswold Endowment (Cornell University to MMLU), and by awards from the National Science Foundation (DEB-0814544 and DEB-0742998 to B.N.D.).

REFERENCES

Austerlitz F, Jung-Muller B, Godelle B, Gouyon P-H (1997) Evolution of coalescence times, genetic diversity and structure during colonization. Theoretical Population Biology 51, 148-164.

Bischoff I (2003) Population dynamics of the solitary digger bee *Andrena vaga* Panzer (Hymenoptera, Andrenidae) studied using mark-recapture and nest counts. Population Ecology 45, 197-204.

Bischoff I, Schröder S, Misof B (2009) Differentiation and range expansion of North American squash bee, *Peponapis pruinosa* (Apidae: Apiformes) populations assessed by geometric wing morphometry. Annals Of The Entomological Society Of America 102, 60-69.

Brooks, S.P. and Gelman, A. (1998). "General Methods for Monitoring Convergence of Iterative Simulations". Journal of Computational and Graphical Statistics, 7, p. 434--455.

Cane JH, Sampson BJ, Miller SA (2011) Pollination Value of Male Bees: The Specialist Bee *Peponapis pruinosa* (Apidae) at Summer Squash (*Cucurbita pepo*). Environmental Entomology 40, 614-620.

Cornuet J-M, Santos F, Beaumont MA, et al. (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics 24, 2713-2719.

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

Crozier RH (1971) Heterozygosity and sex determination in haplo-diploidy. The American Naturalist 105, 399-412.

Decker DS, Wilson HD (1987) Allozyme variation in the *Cucurbita pepo* complex: *C. pepo* var. *ovifera* vs. *C. texana*. Systematic Botany 12, 263-273.

Dieringer D, Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3, 167-169.

Estoup A, Jarne P, Cornuet J (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular Ecology 11, 1591–1604.

Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. Annual Review of Ecology, Evolution, and Systematics 40, 481-501.

Fontaine MC, Austerlitz F, Giraud T, Labbé F (2013) Genetic signature of a range expansion and leap-frog event after the recent invasion of Europe by the grapevine downy mildew pathogen *Plasmopara viticola*. Molecular Ecology 22, 2771–2786.

Furrer R, Nychka D, Sain S (2011) Fields: Tools for spatial data R package version 6.6.

Goudet J (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics J Hered 86: 485-486.

Hedrick PW, Gadau J, Page RE (2006) Genetic sex determination and extinction. Trends Ecol Evol 21, 55-57.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the. Phil. Trans. R. Soc. Lond. B 359, 183-195.

Hurd PD, Linsley EG (1964) The squash and gourd bees genera *Peponapis* Robertson and *Xenoglossa* Smith inhabiting America North of Mexico (Hymenoptera: Apoidea). Hilgardia 35: 375-477.

Hurd PD, Linsley EG (1966) The Mexican squash and gourd bees of the genus *Peponapis* (Hymenoptera: Apoidea). Annals Of The Entomological Society Of America 59, 835-851.

Hurd PD, Linsley EG, Michelbacher AE (1974) Ecology of the squash and gourd bee, *Peponapis pruinosa*, on cultivated cucurbits in California (Hymenoptera: Apoidea).

Hurd PD, Linsley EG, Whitaker TW (1971) Squash and gourd bees (*Peponapis*, *Xenoglossa*) and the origin of the cultivated *Cucurbita*. Evolution, 218-234.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measure: inferring the relative influences of gene flow and drift on distribution of genetic variability. Evolution 53, 1898-1914.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. BMC Genetics 6, 13.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11, 94.

Kirkpatrick KJ, Wilson HD (1988) Interspecific gene flow in *Cucurbita*: *C. texana* vs. *C. pepo*. American Journal of Botany 75, 519-527.

López-Uribe MM, Santiago CK, Bogdanowicz SM, Danforth BN (2013) Discovery and characterization of microsatellites for the solitary bee

Colletes inaequalis using Sanger and 454 pyrosequencing. Apidologie 44, 163-172.

Meirmans PG, van Tienderen PH (2004) Genotype and Genodive: Two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4, 792-794.

Miller-Rushing AJ, Høye TT, Inouye DW, Post E (2010) The effects of phenological mismatches on demography. Phil. Trans. R. Soc. B 365, 3177–3186.

Oreskes N (2004) Beyond the ivory tower: The scientific consensus on climate change. Science 306, 1686-1686.

Paradis E., Claude J. & Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289-290.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421, 37-42.

Payette A, Payette M (2003) Première mention de l'abeille *Peponapis pruinosa* (Say) (Hymenoptera: Apidae) pour le Québec. Fabreries 28, 37-47.

Petersen JD, Reiners S, Nault BA (2013) Pollination services provided by bees in pumpkin fields supplemented with either *Apis mellifera* or *Bombus impatiens* or not supplemented. PLoS ONE 8, e69819.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Ran J-H, Wei X-X, Wang X-Q (2006) Molecular phylogeny and biogeography of *Picea* (Pinaceae): Implications for phylogeographical studies using cytoplasmic haplotypes. Molecular Phylogenetics and Evolution 41, 405-419.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. J Heredity 86, 248–249.

Rosendahl S, Mcgee P, Morton JB (2009) Lack of global population genetic differentiation in the arbuscular mycorrhizal fungus *Glomus mosseae* suggests a recent range expansion which may have coincided with the spread of agriculture. Molecular Ecology 18, 4316-4329.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145, 1219-1228.

Sanjur OI, Piperno DR, Andres TC, Wessel-Beaver L (2002) Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. Proceedings of the National Academy of Science of the United States of America 99, 535–540.

Shuler RE, Roulston TH, Farris GE (2005) Farming practices influence wild pollinator populations on squash and pumpkin. Journal of Economic Entomology 98, 790-795.

Smit, A.F.A., Hubley, R., Green, P. (2010) Repeat-Masker Open-3.0. Available from: http://www. repeatmasker.org

Smith, BD (1997) The initial domestication of Cucurbita pepo in the Americas 10,000 years ago. Science 276, 932-934.

Smith BD (2006) Eastern North America as an independent center of plant domestication. PNAS 103, 12223-12228.

Smith BH (1983) Recognition of female kin by male bees through olfactory signals. PNAS 80, 4551-4553.

Suarez AV, Tsutsui ND (2008) The evolutionary consequences of biological invasions. Molecular Ecology 17, 351-360.

Vavilov, N. I. (1992) “The Origin and Geography of Cultivated Plants” (Cambridge Univ. Press, Cam- bridge, U.K.)

Wcislo WT (1992) Attraction and learning in mate-finding by solitary bees, *Lasioglossum* (*Dialictus*) *figueresi* Wcislo and *Nomia triangulifera* Vachal (Hymenoptera, Halictidae). Behavioral Ecology And Sociobiology 31, 139-148.

Whitaker T, Bemis W (1964) Evolution in the genus *Cucurbita*. Evolution 18, 553-559.

Whitaker TW (1981) Archeological cucurbits. Economic Botany 35, 460-466.

Wilson, HD (1990). Gene flow in squash species. BioScience 40, 449-455.

Wilson HD, Doebley J, Duvall M (1992) Chloroplast DNA diversity among wild and cultivated members of *Cucurbita* (Cucurbitaceae). Theoretical and Applied Genetics 84, 859-865.

Wilson JR, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM (2009) Something in the way you move: dispersal pathways affect invasion success. Trends In Ecology & Evolution 24, 136-144.

Xiang Q, Manchester SR, Thomas DT, Zhang W (2005) Phylogeny, biogeography, and molecular dating of cornelian cherries (*Cornus, cornaceae*): Tracking tertiary plant migration. Evolution 59, 1685–1700.

Zayed A, Constantin Ş, Packer L (2007) Successful biological invasion despite a severe genetic load. PLoS ONE 2, e868.

Zayed A, Packer L (2005) Complementary sex determination substantially increases extinction proneness of haplodiploid populations. PNAS 102, 10742-10746.

Zeder M, Emshwiller E, Smith B, Bradley D (2006) Documenting domestication: the intersection of genetics and archaeology. Trends In Genetics 22, 139-155.

CHAPTER 3

Mating locally despite male-biased dispersal: Fine-scale population structure in solitary ground-nesting bees

Margarita M. López-Uribe[1], Stephen J. Morreale[2,] Christine K. Santiago[1] and Bryan N. Danforth[1]

[1] Department of Entomology, Cornell University, Ithaca, NY 14853

[2] Department of Natural Resources, Cornell University, Ithaca, NY 14853

Abstract

Fidelity to natal sites is a common behavior in organisms that rely on resources that are patchily distributed. Unless mechanisms of inbreeding avoidance evolve, philopatric behavior can have detrimental effects on population viability due to mating with close kin over long time periods. Bee species can show strong philopatric behavior as a result of limited nesting resources and are prone to inbreeding depression due to increased diploid male production when genetic variability is reduced at the complementary sex determination locus. However, mechanisms of inbreeding avoidance in bees are still poorly understood. In this study, we investigate the fine-scale population structure of the solitary bee *Colletes inaequalis* by sampling nine nest aggregations within a small geographic scale (< 4km), and two nest aggregations at a larger geographic scale (45-90km). Relatedness estimates and genetic spatial autocorrelation confirmed greater genetic 'similarity' among individuals within a nest aggregation than among randomly chosen individuals.

We found weak significant genetic differentiation among nest aggregations at both small (Gst= 0.011) and large geographic scales (Gst= 0.017). Reconstruction of parental genotypes revealed greater genetic relatedness among females than among males within each nest aggregation suggesting male-mediated dispersal as a possible mechanism of population connectivity and inbreeding avoidance. Nesting area was positively correlated with effective population size (r=0.71, p=0.03) but not with other estimators of genetic diversity. Geographic distance among nests was the strongest predictor of genetic differentiation between nest aggregations after removing one isolated nest aggregation from the dataset (r= 0.579; p= 0.009). These findings emphasize the importance of nest density and distribution for enhancing bee population abundance and connectivity. Our results highlight the importance of incorporating landscape genetic studies at fine-geographic scales for the understanding of bee mating behavior and population connectivity into studies for the conservation and management of native pollinators.

Introduction

The classical model of genetic structure assumes population subdivision as discrete islands at equilibrium between migration and drift (Wright 1978). However, increasing evidence shows that structure within subpopulations is widespread in many organisms generating hierarchical patterns in some cases (Dionne *et al.* 2008). The minimal spatial scale of population structure depends on the dispersal ability, mating system, and habitat specialization of the species (Shafer *et al.* 2012). For example, fine-scale genetic structure is expected in taxa with low dispersal abilities (Johansson *et al.* 2008) and

advanced social interactions (Bilde *et al.* 2005). However, organisms that have capacity for long-distance dispersal but exhibit philopatric behavior and rely on habitats with patchy distribution are also prone to genetic differentiation at small geographic scales (Buchalski *et al.* 2013; Pierson *et al.* 2013). Because population structure constrains the amount of genetic variability upon which evolutionary forces can act, identifying hierarchical population structure is key to understanding the scale at which different evolutionary processes occur.

Fine-scale population structure can provide evolutionary benefits under certain conditions. For example, to maintain local adaptation to environmental conditions (Stelkens *et al.* 2012), decrease high cost of dispersal (Bonte *et al.* 2012), increase probability of mate encounter when population densities are low (Matthysen 2005), and enhance cooperation among members of a social group (Bilde *et al.* 2005). However, fine-scale structures increase the risk of extreme coancestry within populations and can lead to inbreeding depression, a reduction in individual fitness due to increasing levels of genetic homozygosity and expression of partially deleterious alleles (Charlesworth & Charlesworth 1999). The negative effects of inbreeding depression on fitness have been demonstrated in many species of mammals (Ralls et al 1988), birds (Van Noordwijk & Scharloo 1981), plants (Richards 2000) and in threatened populations with low effective population sizes (Frankham et al 2010). However, most wild populations of animals and plants are not inbred. Due to its detrimental effects to population long-term viability, different strategies have evolved across taxa to avoid high levels of relatedness between individuals in a population. For example, self-incompatibility and female mate choice based on relatedness are widespread mechanisms of inbreeding avoidance in plants and animals, respectively (Castric & Vekemans 2004; Lehmann & Perrin

2002). Many animal species also exhibit natal dispersal (Greenwood 1980; Pusey & Wolf 1996), while others disperse temporarily to mate with individuals from other reproductive groups but return to their natal sites (Buchalski *et al.* 2013; Clark *et al.* 2007).

Alterations in the distribution of species resources and the presence of impervious habitat by-product of human actions influence individual movements and alter connectivity, gene flow and species population structure. Thus, landscape changes can significantly impact the genetic diversity and the evolutionary potential of populations (Miller *et al.* 2012). In continuous habitat, distance is expected to be the best predictor of population structure generating a pattern of isolation-by-distance (Slatkin 1993). On the other hand, when the distribution of resources constraints individual movement in heterogeneous habitats, landscape composition and configuration determine individual movement and population structure (Cushman *et al.* 2006). Investigating how landscape features shape evolution in human dominated habitats is key to understanding adaptation and species responses to anthropogenic change (Goldberg & Waits 2010). Furthermore, understanding how populations are connected through gene flow in highly modified landscapes is necessary to inform management decisions that enhance population connectivity for the conservation of species that are in decline (Palsboll *et al.* 2007).

Bees are the main pollinators of flowering plants in natural and agricultural ecosystems (Kearns *et al.* 1998; Klein *et al.* 2007). However, worldwide trends of bee declines are raising concerns about the economic and environmental consequences associated with the loss of the pollination services they provide (Gallai et al. 2008; Kremen et al. 2007). Multiple studies over the past 10 years have identified three major drivers of decline in bee

population abundance: (1) environmental stressors, including habitat loss and pesticide exposure; (2) increasing pests and pathogens; and (3) loss of genetic diversity (Potts *et al.* 2010). Despite its importance, levels of genetic diversity, population structure and connectivity remain unexplored in wild bee species that need to be enhanced or can be potentially managed for crop pollination. Due to their high mobility and potential for long distance dispersal, genetic structure at fine geographic scales is unexpected in most bees. Even though species in highly fragmented habitats show moderate to high levels of genetic differentiation (Darvill *et al.* 2006; Davis *et al.* 2010), studies of population structure in bees generally indicate low population structure at regional and even continental geographic scales (Beveridge & Simmons 2006; Exeler *et al.* 2008; López-Uribe *et al.* 2014; Zayed & Packer 2002). In spite of that, bee populations can show high levels of inbreeding due to strong philopatric behavior and intranidal nest mating behavior that could generate population structure at fine geographic scales (Danforth *et al.* 2003; Paxton *et al.* 1996). Nevertheless, the evolutionary processes that could maintain weak genetic structure patterns at large scales but presence of genetic structure at small geographic scales are not well understood and remain understudied.

Due to their sex determination system, bees are particularly prone to inbreeding depression. Bees are haplodiploid insects that have a single complementary sex determination (csd) locus that, when homozygous, causes production of sterile diploid males instead of females (Beye *et al.* 2003a). Long-term diploid male production reduces effective population size and increases population genetic load (Hedrick *et al.* 2006) that can lead populations into an “extinction vortex” (Zayed & Packer 2005). Therefore, selection for the evolution of mechanisms of inbreeding avoidance is expected

to be strongly favored in bees. Pre-mating kin recognition through cuticular carbohydrates has been demonstrated as a mechanism of inbreeding avoidance in some bee species (Smith 1983; Wcislo 1992). However, unlike most other species with intense male competition, males are the more discriminating sex in bees (Wcislo 1987). In many solitary bees that nest gregariously, mate location and recognition is based on male attraction by pheromones released from the female body. In this mating system, males actively prefer female odors from different populations than from their own (Vereecken *et al.* 2007). Given the strong female philopatry and male preference for “exotic” female odors, male-mediated dispersal is an expected mechanism to avoid the negative effects of inbreeding (Smith 1983).

One species likely to exhibit the conflicting evolutionary forces between mating with close relatives and inbreeding avoidance, is the ground-nesting bee *Colletes inaequalis.* This bee is solitary, meaning each female builds her own nest and provisions her own offspring, but nests in dense aggregations of up to 100 nests/m^2 (Fig. 3.1A-B). *C. inaequalis* is abundant in eastern North America and it is one of the earliest spring bees in this region (Bartomeus *et al.* 2013). Due to its phenology, *C. inaequalis* is a wild pollinator of early blooming trees such as red maple and willow trees, and it is potentially important for blueberry and apple pollination (Batra 1997). Males emerge a few days before females and patrol trees and nesting sites close to their emergence areas. Mating typically takes place at nest sites where males pounce on females emerging from the ground or while entering and leaving their nest. Unlike males that die after two weeks, females are active for about eight weeks and visit a wider range of plants that flower during their activity period. After emergence, females initiate the construction of their nests and

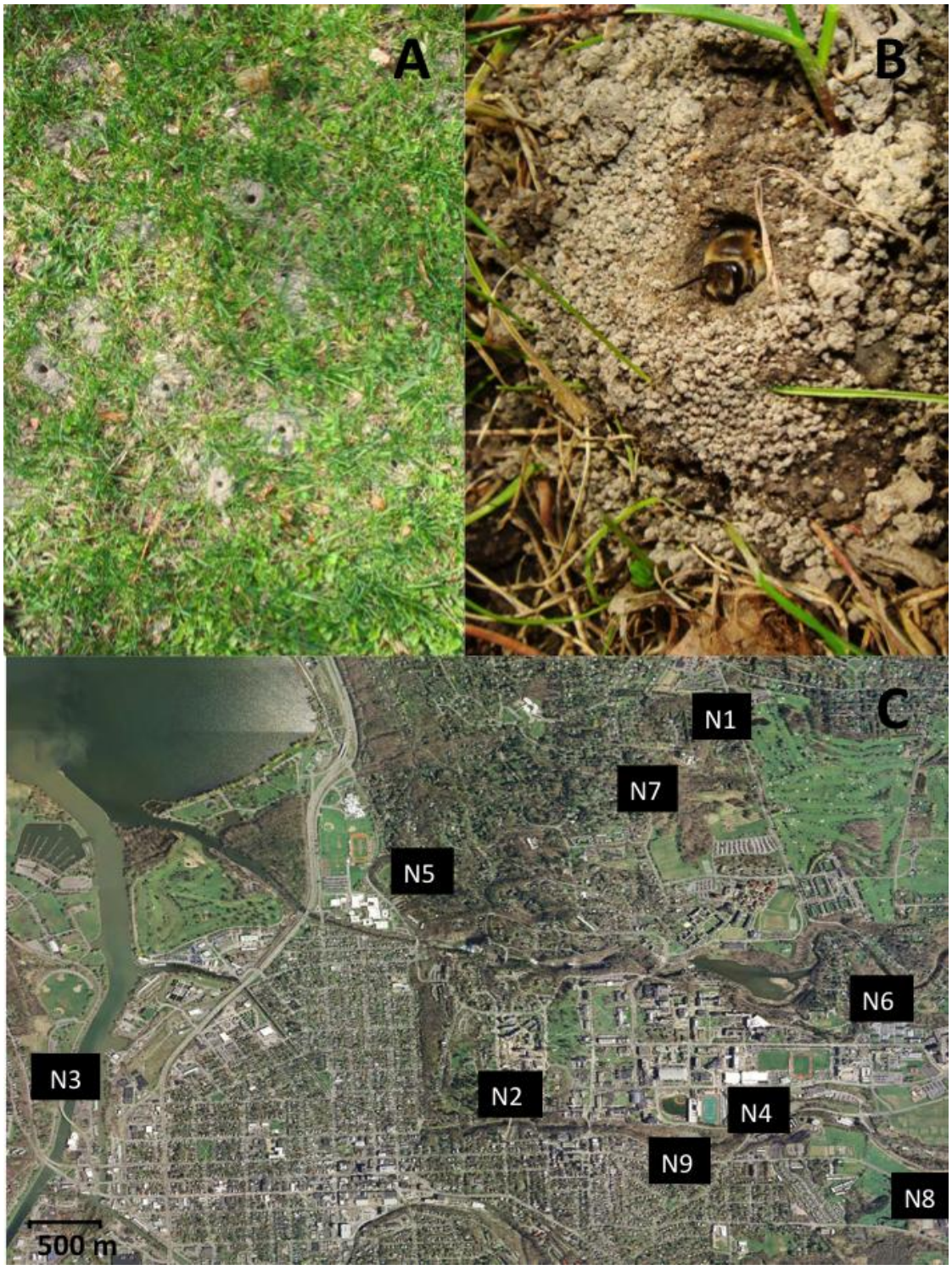


Figure 3.1. (A) Nest aggregations of the ground nesting bee *Colletes inaequalis.* (B) Solitary female at the entrance of her nest. (C) Sampled nest aggregation sites in the city of Ithaca, NY.

cell provisioning close to their emergence points. *C. inaequalis* is a proterandrous species suggesting strong male competition for mates. However, females are receptive for several weeks implying that they can mate more than once, even though males prefer newly emerged females (Batra 1980). The distribution of *C. inaequalis* nesting sites at the landscape level is patchy due to female preference for sandy soils and south-facing slopes on sparse grassy areas (Batra 1980; Cane 1991). Thus, female natal philopatry and the patchy distribution of their nesting habitats suggest that *C. inaqualis* could naturally experience high levels of inbreeding and may be prone to genetic differentiation at fine geographic scales. On the other hand, male preference for “exotic” female odors may favor male dispersal or copulation with females from other nest aggregations.

In this study, we investigate the fine scale genetic structure (<4 km) of the solitary bee *C. inaequalis* in an urban/suburban habitat in central New York, USA. We aimed to disentangle the effect of female natal philopatry and spatial distribution of available nesting sites and floral resources on the population genetic structure and connectivity of *C. inaequalis*. Our approach combines information from highly polymorphic markers and landscape conductance models to test whether (i) individuals within a nest aggregation are inbred and more genetically related to each other than would be expected at random, (ii) females within a nest aggregation are more genetically related than males as expected from female philopatric behavior, and (iii) distribution and availability of nesting sites and floral resources mediate connectivity between nest aggregations. We hypothesize that *C. inaequalis* would exhibit deviations from random mating at very fine geographic scales (< 1 km) due to natal philopatric behavior. Thus, we predict significant differences in the

partition of genetic variability (genetic structure), levels of relatedness and spatial correlation among individuals within and between nest aggregations. Our second hypothesis is that connectivity between nest aggregations is male-mediated. Therefore, we predict higher genetic relatedness between females than males at each nest aggregation. Lastly, we aimed to differentiate the role of the distribution of nesting sites and floral resources as facilitators of population connectivity. Given that the highest density of females is found in nest aggregations and most copulations occur at the nesting sites, we predict nest distribution will be the best predictors of the degree of connectivity between aggregations. Results from this study are relevant for bee conservation because (i) we provide evidence of a possible mechanism of inbreeding avoidance in solitary bees and (ii) we identify landscape features necessary to maintain and potentially enhance bee population connectivity and minimize the risk of inbreeding depression in urban/suburban populations. Our management recommendations also have implications for other native species that share similar phenology and nest habitat specialization.

Materials and Methods

Study area

We sampled *C. inaequalis* nest aggregations in urban/suburban areas of Ithaca, Geneva, and Rochester (New York, USA). Ithaca is a small city (14.1 km^2, *ca.* 100,000 people) surrounded by rural areas consisting of agricultural fields, extensive forests, and scattered residences. The urban area and perimeter contain a large number of gardens and forested areas (Smith *et al.* 1993) that provide a wide variety of floral resources to bees during the spring and summer. Geneva is also a small urban area (11 km^2, *ca.* 13,000 people)

but, unlike Ithaca, it is surrounded by agricultural land dedicated mainly to apple, grape and pumpkin production. Rochester is a heavily urbanized area (93 km^2, *ca.* 210,000 people) with a dense city park system within the perimeter of the city. The climate in Central New York is characterized by large variations in temperature throughout the year with cold winters with temperatures reaching -18ºC, and warm and humid summers that reach 32ºC.

Microsatellite genotyping

DNA was extracted from individual antennae using 150 µl of a 10% Chelex® 100 solution, 5 µl of Proteinase K (Oi *et al.* 2013) and incubated at 55 ºC for 1 h and at 99 ºC for 30 min. Supernatant was used for the amplification of 18 microsatellite loci previously developed for *C. inaequalis* (CI010, CI12, CI15, CI23, CI27, CI028, CI35, CI62, CI66, CI73, CI075, CI87, CI98, CI099, CI102, CI106, CI131, CI179) that were multiplexed according to López-Uribe *et al.* (2013). PCR products were diluted and mixed with Hi-Di Formamide and GeneScan-500 LIZ for genotyping on an Applied Biosystems solution 3730xl DNA Analyzer (Applied Biosystems).

We estimated the power of the microsatellite markers to infer family relatedness using the software KinInfor v.1 (Wang 2006). We calculated the power for relationship inference (PWR) by the simulation procedure that is based on estimated allele frequencies from the microsatellite markers. Our primary hypothesis was the detection of fullsib ($\Delta_1 = 0.5$, $\Delta_2 = 0.5$), paternal halfsibs ($\Delta_1 = 1$, $\Delta_2 = 0$), maternal halfsibs ($\Delta_1 = 0.5$, $\Delta_2 = 0$) and unrelated ($\Delta_1 = 0$, $\Delta_2 = 0$) individuals. We ran 10^6 simulated pairs of genotypes and set the confidence level at 0.05. Genotyping error varied from 0 to 0.02 depending on the locus. We used the reciprocal of the mean squared deviations (RMSD) of different relatedness estimates [Ritland *et al.* (1996) Li *et al.* (1993), Goodnight

and Queller (1999), Lynch and Ritland (1999), and Wang (2002)] to identify the estimator for which our dataset contains the most information and greatest power of relatedness detection.

Genetic diversity and relatedness

The presence of null alleles and large allele dropouts were examined using MicroChecker v.2.2.3 (van Oosterhout *et al.* 2004). We tested for Hardy-Weinberg and linkage disequilibrium using the exact test incorporated in GENEPOP (Rousset 2008). We characterized the genetic diversity of each nest aggregation after removing non-neutral loci by calculating the mean number of alleles (*A*), and observed and expected heterozygosity (*Ho* and *Hs*), using the software GenoDive v.2.0b25 (Meirmans & van Tienderen 2004). Allele richness corrected for sample size and inbreeding coefficients (*Fis*) were estimated in FSTAT v.2.9.3.2 (Goudet 1995). We estimated sibship among females within nest aggregation using the software COLONYv.2.0.4.4 (Jones & Wang 2010) by running two chains using the very high precision method, assuming a polygamous mating system and presence of inbreeding. We removed full sibs from our dataset for all the downstream analysis to have unbiased estimators of population structure from unrelated individuals.

Non-random mating

We used three different methods to test the hypothesis of non-random mating between individuals within and between nest aggregations. First, we tested for population genetic structure using three G-statistics, which use the proportion of genetic diversity found within a population compared to all populations. We compared estimates from Nei's Gst, Hedrick's Gst and Jost's D calculated in GenoDive 2.0b25 (Meirmans & van Tienderen 2004). We estimated pairwise Fst using the *fstat* function in the R package "adegenet" (Jombart *et al.* 2008),

which estimates genetic differentiation based on information from allele frequencies and heterozygosity. Significance was assessed by a permutation test with 10000 repetitions. In the second approach, we estimated pairwise relatedness between individuals to compare mean relatedness in individuals within and among nest aggregations. We used Queller & Goodnight's relatedness estimators because of its robustness to differences in the number of alleles per locus, the allele frequency distribution, sample size, and the presence of inbreeding (Wang 2002). Furthermore, the KinInfo analysis identified this estimator as the one for which our dataset contains the greatest power of relatedness detection (see results). We compared the average pairwise relatedness among individuals within nest aggregations against the null hypothesis of panmixia. We compared the average pairwise relatedness among individuals within nest aggregations to expectations under panmixia was assessed through the bootstrapping method implemented in COANCESTRY v.1.0.1.1 (Wang 2010). Relatedness "within" nest aggregations greater than expected by random mating indicate non-panmictic conditions. Last, we investigated the spatial autocorrelation between individuals every 500 m, using the software SPAGeDi (Hardy & Vekemans 2002). Standard errors were estimated by jackknifing over loci. Confidence intervals around the null expectation of no genetic differentiation were assessed by permutating 1000 multi-locus genotypes and spatial coordinates.

Effective population size and dispersal

Contemporary effective population sizes were calculated using the sibs method implemented in COLONYv.2.0.4.4 (Jones & Wang 2010). This method estimates effective population sizes based on the frequency of sibs within subpopulation with respect to the whole population, while correcting for

deviations from non random mating (Wang 2009). To test the hypothesis of male mediated dispersal, we reconstructed parental genotypes from female offspring genotypes using the full maximum likelihood method in the software COLONYv.2.0.4.4 (Jones & Wang 2010). The algorithm used in COLONY generates posterior probabilities for the best 5 hypothesized parental genotypes. We only included in our analyses genotypes with a posterior probability >0.75 and removed individuals with less than 5 accurately inferred genotypes. Relatedness between female and male individuals in each nest aggregations were quantified using the Queller & Goodnight's (*QGt)* point estimator (Queller & Goodnight 1989). Significant differences in relatedness between females and males across nest aggregations were tested using a one-way ANOVA, assuming unequal variance among samples.

We estimated migration rates between nest aggregations using the Bayesian method implemented in BayesAss v.3.0 (Wilson & Rannala 2003). Following a burnin of 10^6, we performed 4 independent runs of 10^7 iterations with a sampling frequency of 1000. All mixing parameters were set to 1. Convergence was assessed by looking at the repeatability of the four different chains that started with different random seed numbers. We also detected first generation migrants in each nest aggregation using the Rannala and Mountain (1997) Bayesian criterion for likelihood estimation, implemented in GENECLASS2 v.2 (Piry *et al.* 2004). We identified individuals that had a probability <0.01 of belonging to a population different than the one where the individuals were sampled, using a Monte Carlo resampling method (Paetkau *et al.* 2004). Migrants were identified based on the ratio between the likelihood of the individual genotype belonging home and the highest value available between all nest aggregations (L_home/L_max).

Landscape analysis

We visualized the spatial distribution of genetic diversity by calculated the inversed weighted distance (IWD) between nest aggregations using Eucledian distance, inbreeding coefficient (Fis) and spatial principal components (sPCA). We hypothesize that the distribution of genetic variability at this fine geographic scale (~ 14km^2) is dependent on landscape features. Thus, we quantified three landscape features to serve as predictors of the distribution of genetic diversity in the landscape. First, we calculated the area of each nest aggregation as a predictor of 'within nest aggregation' genetic diversity (effective population size, number of alleles and expected heterozygosity). Second, we built a cost distance layer coding early blooming trees as high conductance habitat to explain the distribution of genetic diversity among nest aggregations (Table 5). We used three different conductance values to test the sensitivity of these values to our analysis using layers at a resolution of 0.0001 degrees. The flower distribution cost layer (model 1) and resistance (model 2) laers were compared to Euclidean distance between nest aggregations (null model) (McRae 2006). All distances were correlated with estimates of genetic diversity (pairwise Fst) using Mantel tests as implemented in program zt (Bonnet & Van de Peer 2002). All geospatial operations were performed in ArcGIS v.10.2, and all statistical analyses were performed in R.

Results

Sampling

We collected a total of 548 female individuals of *C. inaequalis* from nest aggregations within the suburban/urban areas of Ithaca (n=9), Geneva (n=1) and Rochester (n=1) (New York, USA) (Fig. 3.1). We used Ithaca populations

to examine the fine-scale genetic structure between nest aggregations, whereas Geneva and Rochester served as comparisons to investigate genetic structures at the broader regional scale. An average of 50 females (±10.7) were collected from each nest aggregation (Table 3.1) in April 2011 to ensure that all individuals belonged to the same generation. Sampling individuals from the same non-overlapping generation allowed us to estimate migration rates and first generation migrants. Linear distances between sampled nest aggregations in Ithaca ranged from 470 m to 4 km. Nest aggregations from Geneva and Rochester were located at 45 and 90 km away from Ithaca, respectively.

Genetic diversity

Significant evidence of null alleles was detected for loci CI35 and CI87 across all nest aggregations; therefore, these loci were excluded from all analyses. All loci from the nest aggregation in Rochester exhibited Hardy-Weinberg disequilibrium (deficit of heterozygotes). Excluding this sample, nine loci (CI27, CI12, CI66, CI62, CI099, CI075, CI15, CI23, CI179) exhibited no significant deviations from Hardy-Weinberg across all nest aggregations and seven (CI028, CI131, CI73, CI106, CI102, CI98) loci exhibited significant heterozygote deficiency in no more than three nest aggregations (Table 3.2). The remaining 16 loci were at linkage equilibrium except for locus CI102, which was excluded from all analyses. Based on the 15 neutral loci, the estimated power for relationship inference (PWR) was 0.981.

The microsatellite loci showed an average number of alleles (*Na*) of 7.73 (±0.443), allele richness (*Ar*) corrected for sample size of 6.38 (±0.191), and observed and expected heterozygosity of 0.568 (±0.055) and 0.615 (±0.061), respectively (Table 3.1). After Bonferroni correction, significant

Table 3.1. Geographic location and genetic diversity indices for each nest aggregations. Number of individuals sampled (*N*), mean number of alleles (*Na*), observed heterozygosity (*Ho*), heterozygosity within subpopulation (*Hs*), inbreeding coefficient (*Gis*), inbreeding coefficient (*Fis*) and significance adjusted for Bonferroni correction (5% level is 0.0003).

Nest aggregation	***Code***	***Coordinates***	***N***	***Na***	***Ar***	***Ho***	***Hs***	***Fi***	***p-value***
Pleasant Grove	N1	42° 27' 38.14" -76° 28' 39.36"	57	8.00	6.575	0.621	0.617	-0.006	0.5742
Ithaca Cemetery	N2	42° 26' 40" -76° 29' 26.3"	55	7.73	6.312	0.595	0.616	0.034	0.0788
Cass Park	N3	42° 26' 45.00" -76° 30' 53.30"	50	7.87	6.508	0.596	0.602	0.01	0.3418
Football field	N4	42° 26' 38.2" -76° 28' 36.8"	53	7.60	6.228	0.597	0.62	0.037	0.0288
Cayuga heights	N5	42° 27' 24.7" -76° 29' 43.2"	58	7.80	6.438	0.567	0.606	0.065	0.0015
Tunnel	N6	42° 26' 58.50" -76° 28' 12.50"	63	7.67	6.081	0.605	0.606	0.001	0.3967
Jim’s House	N7	42° 27' 41.19" -76° 28' 59.32"	52	7.67	6.317	0.603	0.607	0.007	0.3833
East Hill	N8	42°26' 20.60" -76° 28' 4.10"	49	7.20	6.153	0.624	0.618	-0.01	0.6491
Ctown	N9	42°26' 28" -76° 28' 52.5"	48	8.07	6.670	0.607	0.625	0.028	0.117
Geneva	N10	42°46' 3.7" -76° 59' 8.4"	24	6.87	6.597	0.616	0.625	0.014	0.347
Rochester	N11	43°8' 22.87" -77° 34' 7.38"	39	6.93	6.314	0.215	0.624	0.655	0.0003

Table 3.2. Descriptive summary of PCR primers and conditions, and genetic diversity indices of the 18 microsatellite loci. Genetic diversity indices show: Number of individuals sampled (*N*), number of alleles (*Na*), observed heterozygosity (*Ho*), heterozygosity within subpopulation (*Hs*), and inbreeding coefficient (*Gis*).

Locus	Motif	T_a	Allele sizes	N_a	H_O	H_S	Gis
CI028	$(CA)_7$	59	142 - 164	5	0.474	0.475	0.002
CI27	$(TG)_7GG(TG)_6$	55	193 - 223	11	0.356	0.368	0.032
CI12	$(CT)_{10}$	58	207 - 132	12	0.710	0.788	0.100
CI131	$(GTT)_3(GCT)_2(GTT)_3$	59	248 - 270	3	0.450	0.465	0.031
CI66	$(AAG)_{15}$	54	350 - 410	16	0.814	0.853	0.045
CI73	$(TAGA)_{11}$	55	181 - 205	6	0.056	0.063	0.116
CI35	$(TG)_{23}$	54	238 - 314	13	0.207	0.353	0.414
CI106	$(GT)_{11}$	53	119 - 164	25	0.699	0.790	0.116
CI62	$(CAA)_8$	55	119 - 164	15	0.716	0.763	0.062
CI102	$(TTGT)_8$	54	181 - 205	12	0.735	0.803	0.085
CI099	$(AC)_9$	54	367 - 395	6	0.642	0.683	0.060
CI87	$(CA)_{16}$	54	304 - 356	30	0.769	0.904	0.150
CI075	$(TG)_{15}CG(TG)_2TA(TG)_6$	59	367 - 395	10	0.305	0.332	0.082
CI15	$(TG)_{11}$	52	114 - 132	8	0.509	0.591	0.139
CI010	$(CA)_{15}$	53	160 - 194	18	0.817	0.910	0.102
CI98	$(CA)_{15}$	55	221 - 249	19	0.687	0.756	0.092
CI23	$(TG)_7AC(TC)_7$	55	248 - 270	13	0.756	0.829	0.088
CI179	$(CT)_9$	59	299 - 305	4	0.527	0.557	0.054

heterozygote deficiencies were only found in the nest aggregation from Rochester (Table 3.1). Therefore, the use of the Queller & Goodnight's estimator gave an unbiased estimation of relatedness in our analyses (Wang2014). The COLONY analysis detected five full sibs in our dataset (nest aggregations N1, N5, N6, N9 and N10). One individual of these dyads was randomly removed for all downstream analyses.

Non-random mating

All G-statistic estimators indicated significant ($p<0.005$) but weak genetic structure (Nei Gst= 0.011; Hedrick Gst= 0.0317; Jost D= 0.0195; φst= 0.0538) suggesting deviations from random mating at the local scale in the nest aggregations from Ithaca (Table 3.3). For pairwise Fst, ten out of the 36 pairwise comparisons were significantly different from zero (Table 3.4). When nest aggregations from Geneva and Rochester were included, higher but still weak genetic differentiation was detected (Nei Gst= 0.017; Hedrick Gst= 0.0474; Jost D= 0.0293**;** φst**=** 0.071). Comparatively, Nei's Gst estimator gave the lowest mean values and variances of differentiation but it behaved non-linearly when compared to corrected measures (Fig. 3.2).

Mean levels of relatedness within each nest aggregations were low and exhibited large variances (*QGt*= 0.003± 0.0134). Overall, average relatedness was higher among individuals within nest aggregations than among individuals from different aggregations (QGt_{wihin}-$QGt_{between}$= 0.0052, p-value<0.01). However, this difference in relatedness between groups was variable when each nest was assessed independently (Fig. 3.3). The spatial autocorrelation analysis further supported evidence of non-random mating among individuals from different nest aggregations. The autocorrelogram shows that individuals within nest aggregations significantly deviate from panmixia (Fig. 3.4).

Table 3.3. Average global G-statistics for the nest aggregations in Ithaca and all nest aggregations (including Geneva and Rochester). Nei's Gst is corrected for average heterozygosity. Global estimates and confidence intervals were estimated by bootstrapping over loci and significance was assessed after 10000 permutations.

Estimator	**Ithaca**			**All**		
	Mean	***CI***		***Mean***	***CI***	
Nei's Gst	0.011	0.009-0.0135	0.003	0.0172	0.0145-0.0202	0.000
Hedrick's Gst	0.0321	0.0257-0.0387	0.003	0.048	0.0407-0.0564	0.000
Jost's D	0.0198	0.0158-0.0238	0.003	0.0298	0.0251-0.035	0.000
Phi st	0.0538	0.0444-0.0645	0.003	0.071	0.0612-0.0819	0.000

Table 3.4. Pairwise Fst values (below) and p-values (above) of the genetic differentiation among nest aggregations in Ithaca based on 15 microsatellite markers. Numbers in bold denote significant differentiation at the alpha-level 0.05.

	Plea_Grove (N1)	Itha_Ceme (N2)	Cass_Park (N3)	Foot_Field (N4)	Cayu_Heig (N5)	Tunnel (N6)	Jim_House (N7)	East_Hill (N8)	Ctown (N9)
(N1)	-	0.822	0.194	0.956	0.597	0.496	0.053	0.121	0.467
(N2)	0.0041	-	0.668	0.487	0.68	0.509	0.386	0.4	0.139
(N3)	0.0071	0.0055	-	**0.049**	0.223	**0.014**	**0.049**	**0.046**	**0.047**
(N4)	0.0033	0.0052	**0.0083**	-	0.573	0.419	0.203	0.08	0.178
(N5)	0.0045	0.0036	0.0059	0.0051	-	0.093	0.077	0.112	**0.004**
(N6)	0.0043	0.0045	**0.0082**	0.0049	0.0039	-	0.024	**0.02**	**0.016**
(N7)	0.0062	0.0053	**0.0080**	0.0058	0.0054	0.0068	-	0.018	**0.002**
(N8)	0.0060	0.0054	**0.0088**	0.0071	0.0063	**0.0072**	0.0079	-	**0.011**
(N9)	0.0054	0.0068	**0.0100**	0.0066	**0.0088**	**0.0081**	**0.0106**	**0.0090**	

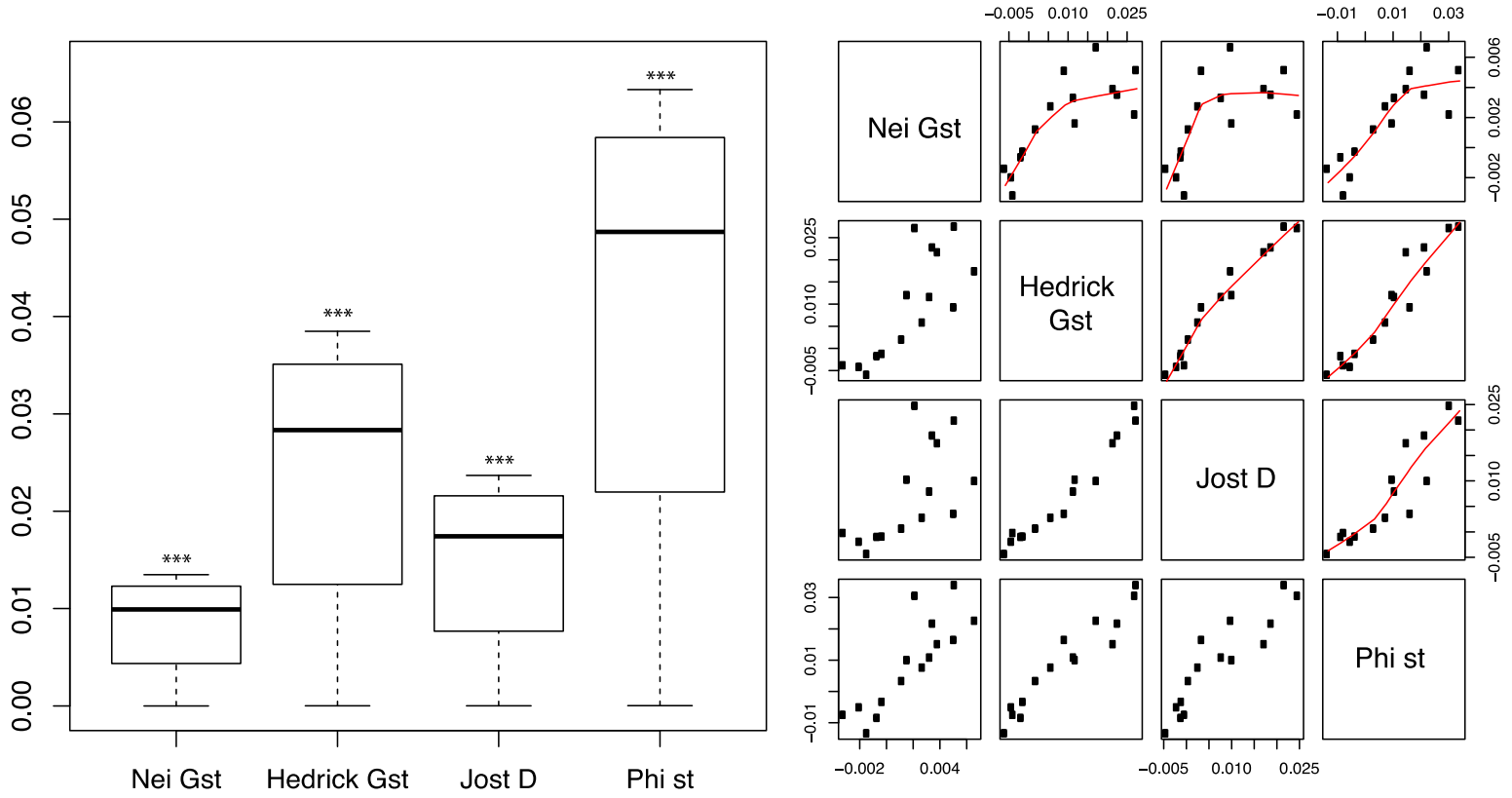


Figure 3.2. Global estimators of population structure for the nine nest aggregations in Ithaca, NY and the relationship between them. Mean values and 95% confident intervals for Nei's Gst, Hedrick's Gst, Jost's D and Phi st. All estimators were significant with p-value<0.005.

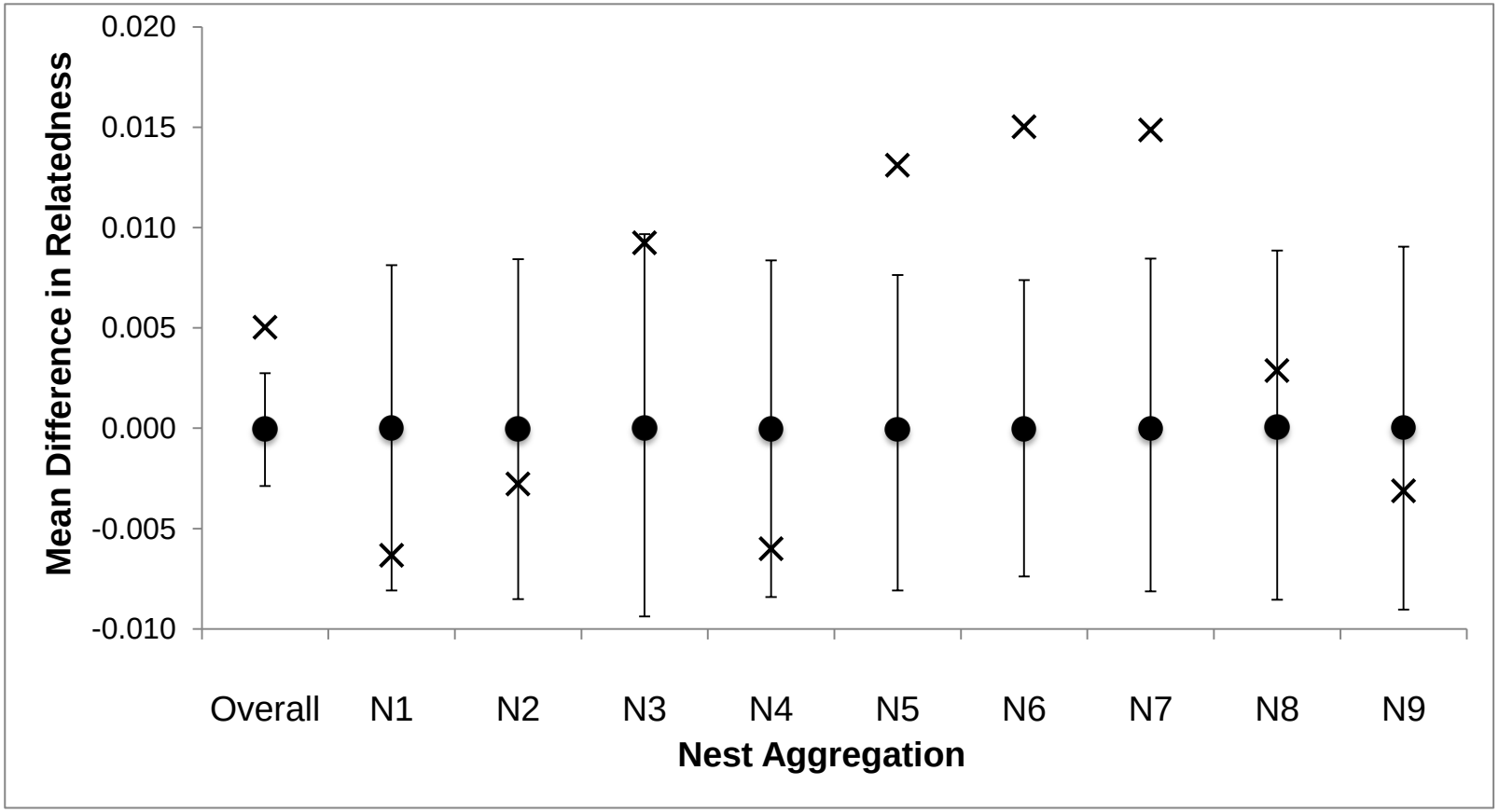


Figure 3.3. Mean relatedness values (Queller and Goodnight's) between individuals within nest aggregations (x) compared to the expectations under panmixia. Black dots represent the mean and bars represent the 95% confidence interval expected under random mating after 10000 bootstrapping repetitions.

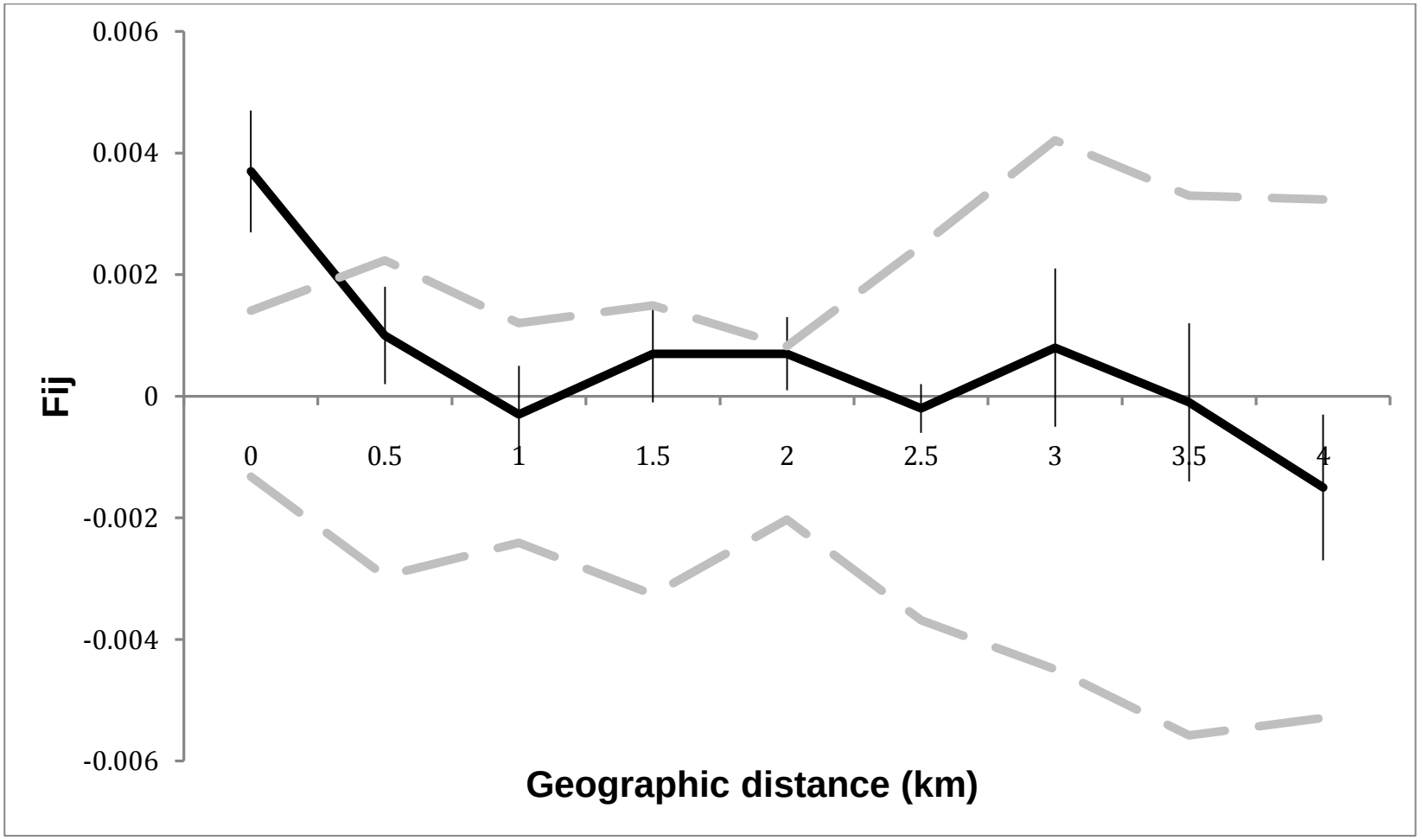


Figure 3.4. Spatial autocorrelation diagram based on kinship coefficient *Fij* at different distance intervals (black line). Gray lines depict the 95% confident intervals showing the expected range of *Fij* if there is no correlation between kinship and distance.

Effective population size and dispersal

Effective population sizes ranged from 55 to 71 across all nest aggregations (mean= 60). Except for the nest aggregation in Rochester, the presence of inbreeding within nest aggregations did not significantly affect the estimation of effective population size. We reconstructed parental genotypes of 221 females and 490 males using the conservative criteria described above. We found a significantly higher relatedness among females than males within each nest aggregation ($p<0.0001$) (Fig. 3.5). The BayesAss runs reached convergence into two different estimates of migration. Migration rates were high (~30%) for N4 and N8, respectively, while the rates of migration between the remaining nest aggregation were ~0.5%. We detected 7 first generation migrants from N1, N2, N3, N4, N6, and N8 using the maximum likelihood method (Table 3.5).

Landscape analysis

The IDW Euclidian distance map clearly shows that N9 is the nest aggregation most geographically distant (Fig. 3.6A). For the levels of inbreeding, N6 shows the highest levels (more positive values), while N5 and N7 show the lowest levels (Fig. 3.6B). The sPCA values represent relative genetic similarity weighted by spatial distribution. The sPCA map shows that that N6 is the most distinct nest aggregation in a spatially explicit context (Fig. 3.6C). We found a significant correlation between nesting area and effective population size ($r= 0.71$, $p=0.03$) but not with number of alleles, expected heterozygosity or inbreeding (Fig. 3.7). Euclidean distance between nest aggregations, least-cost and resistance layers through flower resources significantly explained the genetic differentiation between sites but only after removing N9 from the dataset (Table 3.6). The genetic differentiation of this nest aggregation was higher than expected by distance (Fig. 3.8). Distance between nests was the

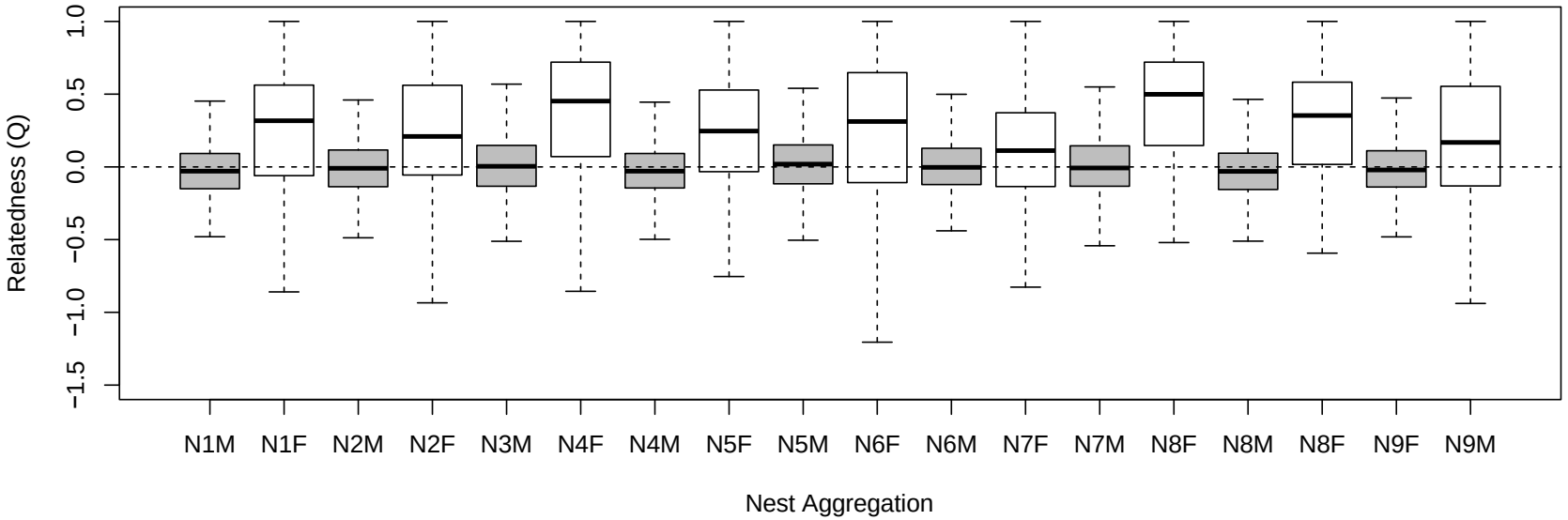


Figure 3.5. Comparison of relatedness between females and males in each nest aggregation. Relatedness calculated based on Queller & Goodnight's index (Q). Grey and white boxplots show distribution of relatedness between males and females of the same nest aggregation respectively. Differences between mean relatedness of males and females was highly significant (F=1244.16, p-value<0.000).

Table 3.5. First generation migrants identified in the nest aggregations from Ithaca, NY. Sampling and hypothesizes origin of each detected migrant, and significance of the inference based on a Monte Carlo resampling method

ID	***Nest sampled***	***Origin***	***p-value***
Cin064	N1	N2	0.003
Cin198	N4	N8	0.002
Cin355	N7	N4	0.001
Cin432	N8	N6	0.01
Cin467	N9	N1	0.002

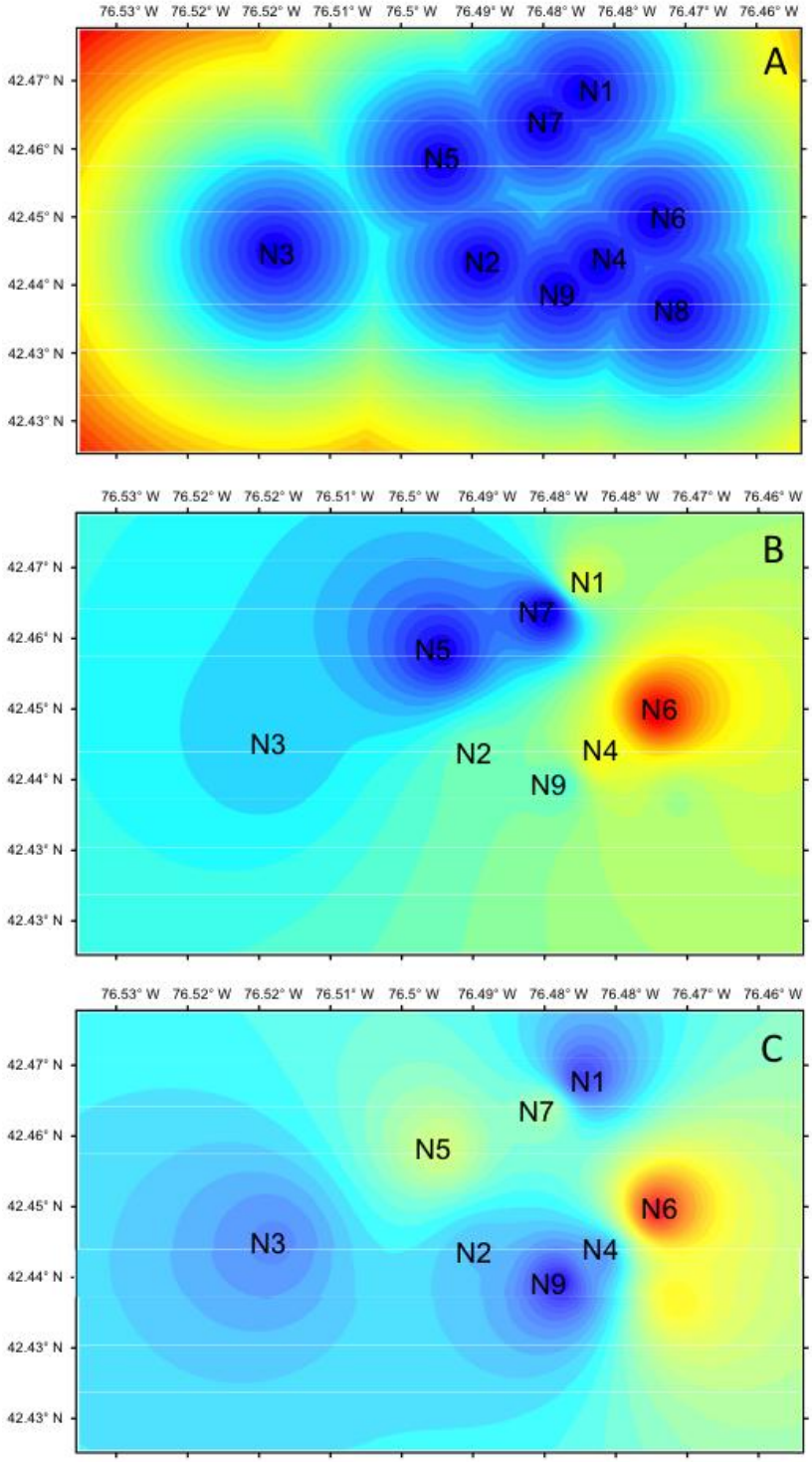


Figure 3.6. Inverted weighted distance of (A) Eucledian distance, (B) inbreeding coefficient (Fis), and (C) spatial principal components between nest aggregations of *Colletes inaequalis* in Ithaca, NY.

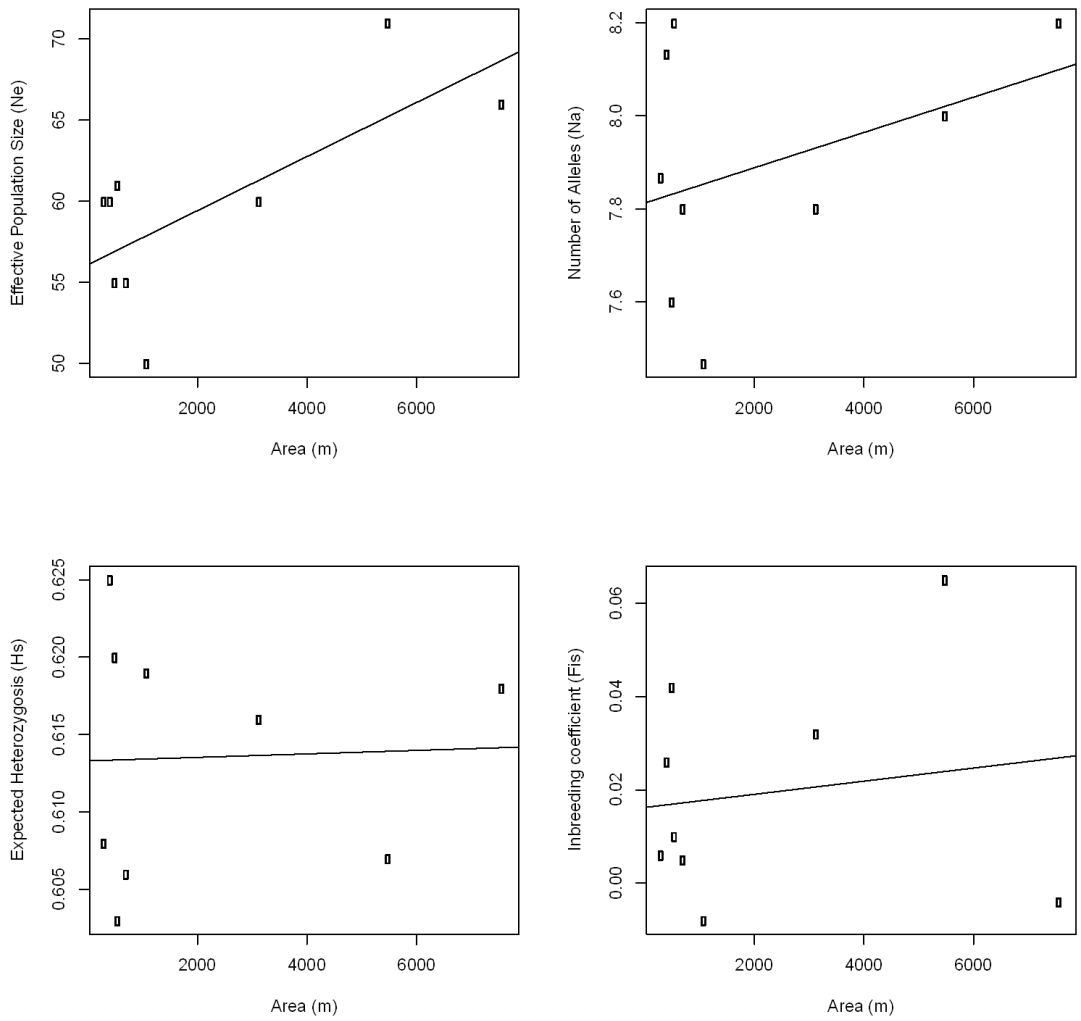


Figure 3.7. Correlation between nesting area (meters) and (A) effective population size (r=0.71, p=0.03), (B) number of alleles (r=0.39, p=0.31), (C) expected heterozygosity (r=0.04, p=0.92), and (D) inbreeding coefficient (r=0.16, p=0.69) all nest aggregation in Ithaca.

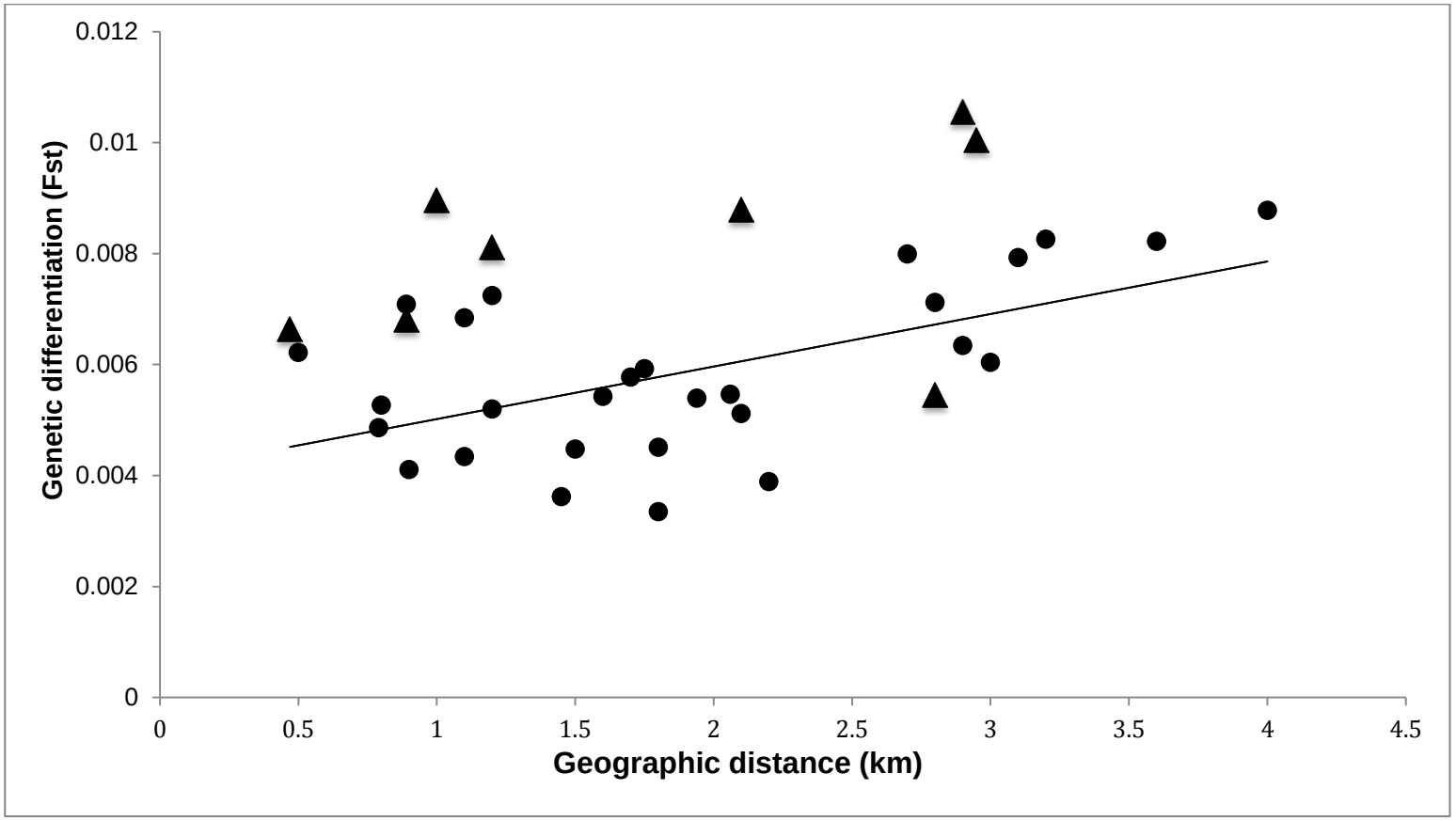


Figure 3.8. Relationship between pairwise genetic differentiation, measured as Fst corrected for heterozygosity, and geographic distance (km) for the nine nest aggregations in Ithaca, NY. Mantel test was not significant for the 9 nest aggregations (R^2= 0.179, r= 0.37, p= 0.096) but highly significant after removing nest aggregation N9 (R^2= 0.338, r= 0.59, p= 0.006). Triangles represent pairwise comparisons of N9 and circles represent all the rest of the other aggregation.

Table 3.6. Mantel test results for the Euclidean distance (null model), least-cost paths (model 1) and resistance (model 2) and their relationship with genetic differentiation measured as Gst corrected for heterozygosity. Results for a partial Mantel test after controlling for Eucledian distance are also shown. A sensitivity analysis was performed using 3 different cost values for each layer but correlation values did not change.

Model	Cost Value			Cost Layer	Simple	Partial
	High	Medium	Low			
Euclidian	-	-	-	Euclidian	r= 0.37; p=0.098 (*r= 0.579; p= 0.009*)[a]	-
Least coast	250	50	5	Trees	r= 0.322; p= 0.131 (*r= 0.497; p=0.022*)[a]	r= -0.223; p=0.207
Conductance	250	50	5	Trees	r=0.257; p=0.189 (*r= 0.312; p=0.018*)[a]	r= -0.391; p=0.007

[a] Spearman correlation coefficients after removing N9 from the dataset.

strongest landscape predictor of genetic differentiation (r= 0.579; p= 0.009). After controlling for Euclidian distance, the flower cost layer did not have any predictive power for explaining the distribution of genetic patterns (Table 3.6).

Discussion

Fine-scale population structure

Our results from G-statistics, relatedness and spatial autocorrelation support the hypothesis that *C. inaequalis* individuals mate with individuals from the same nest aggregation more frequently than at random. However, in some nest aggregations levels of relatedness were lower than expected at random. We detected significant genetic differentiation between nest aggregations sampled from an area of ~14 km^2, which demonstrates that solitary bee populations can show significant genetic structure at small geographic scales despite their ability to fly. Other studies in solitary bees with similar body size have found weak or no detectable genetic structure at broader geographic scales (Cerantola *et al.* 2010; Exeler *et al.* 2008; Suni *et al.* 2014). Even though we only sampled two nest aggregations outside the Ithaca area, levels of genetic differentiation at the broader geographic scale were also weak (Gst= 0.02). This pattern of high genetic relatedness at fine-scales but low genetic differentiation across their geographic range is also found in other species with high mobility and strong philopatric behavior, such as birds and bats (Buchalski *et al.* 2013; Pierson *et al.* 2013) . Our results suggest that solitary bees can show local genetic differentiation in the face of high levels dispersal that homogenizes genetic variation at larger geographic scales. Differences between local and regional genetic differentiation have also been found in bumble bees that have large home ranges and the capacity for long-

distance dispersal (Jha & Kremen 2013). Thus, this study emphasizes the importance of fine scale processes for the distribution of genetic diversity in bee populations. Although we detected deviations from random mating at our sampling scale, the group of nest aggregations within the city of Ithaca should be considered a single population due to the low levels of genetic differentiation (Gst= 0.01) and moderate number of first generation migrants detected (1%).

The significant genetic structure revealed in this study is not surprising given the high degree of philopatry observed in this species (Batra 1980). Even though, we expected higher levels of inbreeding due to mating among closely related individuals over many generations, the nest aggregation from Rochester was the only one that showed a significant evidence of deficit of heterozygotes (Fis= 0.664). Absence of inbreeding despite non-random mating at this small geographic scale may be the result of one or a combination of the following: (1) large effective population sizes in nest aggregations that allow for low probability of mating between sibs (negative effects of genetic drift in small populations), (2) high levels of gene flow among nest aggregations that lowers levels of coancestry within each nest aggregation, or (3) negative assortative mating based on kin recognition that prevents mating between relatives. Our data show that the average effective population size of these nest aggregations is small (~60 individuals) but migration rates between nest aggregations are moderate (~1%). Thus, we speculate that both gene flow and kin recognition may be important mechanisms of inbreeding avoidance in this gregariously philopatric bee.

Male-biased dispersal

We provide evidence that females are consistently more genetically related than males across all sampled nest aggregations suggesting that males disperse more than females in *C. inaequalis.* Sex-biased dispersal is a common behavior in mammals, birds and reptiles (Greenwood 1980; Lane & Shine 2011) and it has recently been demonstrated in bees (Ulrich *et al.* 2009). Mating system in bees are generally thought to be monandrous in females and polygamous in males, even though genetic data supporting this premise is lacking (Paxton 2005). In monandrous systems, receptive females are considered a limited resource, thus male dispersal is expected as a strategy to maximize their reproductive success. However, *C. inaequalis* females are receptive for most of their lives and can probably mate more than once (Batra 1980). Because female philopatry in *C. inaequalis* may dramatically increase coancentry of individuals at natal sites and males have the capacity for kin discrimination (Vereecken *et al.* 2007), we hypothesize that male-biased dispersal may be favored to avoid kin competition and long-term genetic load (Perrin & Mazalov 2000).

Evidence for intranidal mating and high levels of inbreeding reported in several solitary bees with gregarious nesting (Paxton *et al.* 1999; Paxton *et al.* 1996) suggest that inbreeding may be widespread phenomenon among bees due to their high fidelity to natal sites and patchy distribution of resources (Paxton 2005). However, mating systems and behaviors that reduce inbreeding are expected to evolve in bees due to potential production of high frequencies of diploid males that can increase genetic load in populations. In this study, we provide evidence suggesting that male-biased dispersal is a possible mechanism for inbreeding avoidance in bees in spite of non-random mating at natal sites. On the other hand, our results also indicate that female

C. inaqualis may have limited colonization capacity due to the strong philopatric behavior and the high cost associated with dispersing to search for nesting sites with low density and patchy distributions in the landscape.

Landscape mediated responses

Our data emphasize the importance of the density and distribution of nesting areas for the enhancement of solitary bee populations (Lonsdorf *et al.* 2009). We found that geographic distance among nest aggregations significantly correlates with genetic structure and that the size of the nesting area positively correlates with the effective population sizes of the nest aggregations (Fig. 4). Thus, we provide evidence that increasing the density and amount of suitable nesting habitat in the landscape may increase bee population abundance and connectivity. Management strategies for bee conservation generally focus on improving habitat quality for females because they provide the majority of pollination services while males are short-lived. Our results highlight the important role of males for bee population dynamics and long-term persistence, as they are the mediators of population connectivity.

Nest aggregations of *C. inaequalis* sampled in Ithaca and Geneva showed low levels of inbreeding and high levels of genetic diversity suggesting these nest aggregations are ‘healthy’ in a genetic sense. However, the nest aggregation from the most heavily urbanized area (Rochester, NY) showed very high levels of inbreeding. Studies of urban ecology have demonstrated that species diversity can peak in transitional regions between urban and rural areas (Magura *et al.* 2010). Human activity can increase landscape heterogeneity to a certain degree, but high degrees of urbanization have a strong homogenizing effect that dramatically lowers species diversity (Cam *et al.* 2000). In the case of bees, urban areas can provide suitable habitat for

pollen generalist due to the wide variety of floral resources that these habitats provide throughout the year (Hernandez & Frankie 2009). In regards to nesting habitats, cavity-nesters are the species that show the strongest positive effect to urbanization because of increased nest suitable areas in human-made structures (Cane *et al.* 2006a). However, nesting sites are the limiting resource for ground nesting bees in urban areas where the soil is heavily modified and the proportion of bare soil is dramatically reduced. Even though we did not directly examine nest aggregations in a gradient of urbanization, our results suggest that areas with intermediate levels of urbanization and high density distribution of nesting habitat and floral resources may provide high quality habitat for solitary wild bees.

Landscape management recommendations

The key finding of this study from a management perspective is that the distribution and density of suitable nesting areas in urban/suburban areas strongly drive population size and connectivity between solitary ground nesting bees that have narrow preferences for soil characteristics. Specifically for *C. inaequalis*, increasing the amount of bare undisturbed sandy soil is desirable to enhance population sizes. However, we predict the colonization rate of solitary bees may be low due to the strong tendency for philopatric behavior in bees. Therefore, if enhanced populations of solitary bees are required for pollination, seeding suitable nesting sites with overwintering adults may be an effective approach.

Acknowledgements

We thank L. Duque, K. Sullivan and students from the Applied Conservation course at Cornell for their assistance in the field; J. Greenberg for providing

specimens from Rochester; S. DeGloria and K. Jenkins for their assistance with GIS analyses. We also thank members of the Zamudio lab and xx reviewers for feedback on earlier versions of the manuscript. Funding was provided by grants from the Andrew W. Mellon Foundation at Cornell University (MMLU), the Sarah Bradley Fellowship (MMLU), the Cornell Biology Research Fellowship Program (CKS), and by awards from the National Science Foundation (DEB-0814544 and DEB-0742998 to BND).

REFERENCES

Bartomeus I, Ascher JS, Gibbs J*, et al.* (2013) Historical changes in northeastern US bee pollinators related to shared ecological traits. *PNAS*.

Batra S (1997) Solitary bees for *Vaccinium* pollination. *Acta Horticulturae* **476**, 71-76.

Batra SWT (1980) Ecology, behavior, pheromones, parasites and management of the sympatric vernal bees *Colletes inaequalis, C. thoracicus* and *C. validus*. *Journal of the Kansas Entomological Society* **53**, 509-538.

Beveridge M, Simmons L (2006) Panmixia: an example from Dawson's burrowing bee (*Amegilla dawsoni*)(Hymenoptera: Anthophorini). *Molecular Ecology* **15**, 951-957.

Beye M, Hasselmann M, Fondrk MK, Page RE, Omholt SW (2003) The gene csd is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell* **114**, 419-429.

Bilde T, Lubin Y, Smith D, Schneider JM, Maklakov AA (2005) The transition to social inbred mating systems in spiders: Role of inbreeding tolerance in a subsocial predecessor. *Evolution* **59**, 160-174.

Bonnet E, Van de Peer Y (2002) zt: a software tool for simple and partial Mantel tests. *Journal of Statistical Software* **7**, 1-12.

Bonte D, Van Dyck H, Bullock JM*, et al.* (2012) Costs of dispersal. *Biological Reviews* **87**, 290-312.

Buchalski MR, Chaverri G, Vonhof MJ (2013) When genes move farther than offspring: gene flow by male gamete dispersal in the highly philopatric bat species *Thyroptera tricolor*. *Molecular Ecology* **23**, 464-480.

Cam E, Nichols JD, Sauer JR, Hines JE, Flather CH (2000) Relative species richness and community completeness: birds and urbanization in the mid-Atlantic states. *Ecological Applications* **10**, 1196-1210.

Cane J (1991) Soils of ground-nesting bees (hymenoptera, apoidea) - texture, moisture, cell depth and climate. *Journal of the Kansas Entomological Society* **64**, 406-413.

Cane JH, Minckley RL, Kervin LJ (2006) Complex responses within a desert bee guild (Hymenoptera: Apiformes) to urban habitat fragmentation. *Ecological Applications* **16**, 632-644.

Castric V, Vekemans X (2004) Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. *Molecular Ecology* **13**, 2873-2889.

Cerantola NdCM, Oi CA, Cervini M, Del Lama MA (2010) Genetic differentiation of urban populations of Euglossa cordata from the state of São Paulo, Brazil. *Apidologie*, 1-9.

Charlesworth B, Charlesworth D (1999) The genetic basis of inbreeding depression. *Genetical Research* **74***,* 329-340.

Clark RW, Brown WS, Stechert R, Zamudio KR (2007) Integrating individual behaviour and landscape genetics: the population structure of timber rattlesnake hibernacula. *Molecular Ecology* **17**, 719-730.

Cushman SA, Mckelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist* **168**, 486-499.

Danforth B, Ji S, Ballard L (2003) Gene flow and population structure in an oligolectic desert bee, *Macrotera* (*Macroteropsis*) *portalis* (Hymenoptera: Andrenidae). *Journal of the Kansas Entomological Society* **76**, 221-235.

Darvill B, Ellis J, Lye G, Goulson D (2006) Population structure and inbreeding in a rare and declining bumblebee, *Bombus muscorum* (Hymenoptera: Apidae). *Molecular Ecology* **15**, 601-611.

Davis ES, Murray TE, Fitzpatrick U, Brown MJF, Paxton RJ (2010) Landscape effects on extremely fragmented populations of a rare solitary bee, *Colletes floralis*. *Molecular Ecology*, 1-14.

Dionne M, Caron F, Dodson JJ, Bernatchez L (2008) Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. *Molecular Ecology* **17**, 2382-2396.

Exeler N, Kratochwil A, Hochkirch A (2008) Strong genetic exchange among populations of a specialist bee, *Andrena vaga* (Hymenoptera: Andrenidae). *Conservation Genetics* **9**, 1233-1241.

Goldberg CS, Waits LP (2010) Comparative landscape genetics of two pond-breeding amphibian species in a highly modified agricultural landscape. *Molecular Ecology* **19**, 3650-3663.

Goodnight K, Queller D (1999) Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Molecular Ecology* **8**, 1231-1234.

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.

Greenwood PJ (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour* **28***,* 1140-1162.

Hardy OJ, Vekemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620.

Hedrick PW, Gadau Jr, Jr REP (2006) Genetic sex determination and extinction. *Trends in Ecology and Evolution* **21**, 55-57.

Hernandez JL, Frankie GW (2009) Ecology of urban bees: a review of current knowledge and directions for future study. *Cities and the Environment* **2**, 1-15.

Jha S, Kremen C (2013) Urban land use limits regional bumble bee gene flow. *Molecular Ecology* **22**, 2483-2495.

Johansson ML, Banks MA, Glunt KD, Hassel-Finnegan HM, Buonaccorsi VP (2008) Influence of habitat discontinuity, geographical distance, and oceanography on fine-scale population genetic structure of copper rockfish (*Sebastes caurinus*). *Molecular Ecology* **17**, 3051-3061.

Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* **101**, 92-103.

Jones OR, Wang J (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* **10**, 551-555.

Kearns CA, Inouye DW, Waser NM (1998) Endangered mutualisms: the conservation of plant-pollinator interactions. *Annual Review of Ecology and Systematics* **29**, 83-112.

Klein A-M, Vaissière BE, Cane JH*, et al.* (2007) Importance of pollinators in changing landscapes for world crops. *Proceedings Of The Royal Society B-Biological Sciences* **274**, 303-313.

Lane A, Shine R (2011) Intraspecific variation in the direction and degree of sex-biased dispersal among sea-snake populations. *Molecular Ecology* **20**, 1870-1876.

Lehmann L, Perrin N (2002) Altruism, dispersal, and phenotype-matching kin recognition. *The American Naturalist* **159**, 451-468.

Li C, Weeks D, Chakravarti A (1993) Similarity of DNA fingerprints due to chance and relatedness. *Human heredity* **43**, 45-52.

Lonsdorf E, Kremen C, Ricketts T*, et al.* (2009) Modelling pollination services across agricultural landscapes. *Annals of Botany* **103**, 1589-1600.

López-Uribe MM, Santiago CK, Bogdanowicz SM, Danforth BN (2013) Discovery and characterization of microsatellites for the solitary bee *Colletes inaequalis* using Sanger and 454 pyrosequencing. *Apidologie* **44**, 163-172.

López-Uribe MM, Zamudio KR, Cardoso CF, Danforth BN (2014) Climate, physiological tolerance and sex-biased dispersal shape genetic structure of Neotropical orchid bees. *Molecular Ecology* **23**, 1874–1890.

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics* **152**, 1753-1766.

Magura T, Horváth R, Tóthmérész B (2010) Effects of urbanization on ground-dwelling spiders in forest patches, in Hungary. *Landscape Ecology* **25**, 621-629.

Matthysen E (2005) Density-dependent dispersal in birds and mammals. *Ecography* **28**, 403-416.

McRae BH (2006) Isolation by resistance. *Evolution* **60**, 1551-1561.

Meirmans PG, van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792-794.

Miller MP, Bianchi CA, Mullins TD, Haig SM (2012) Associations between forest fragmentation patterns and genetic structure in Pfrimer's Parakeet (*Pyrrhura pfrimeri*), an endangered endemic to central Brazil's dry forests - Online First - Springer. *Conservation Genetics* **14**, 333-343.

Oi CA, López-Uribe MM, Cervini M, Del Lama MA (2013) Non-lethal method of DNA sampling in euglossine bees supported by mark–recapture experiments and microsatellite genotyping. *Journal of Insect Conservation* **17**, 1071-1079.

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.

Palsboll PJ, Berube M, Allendorf FW (2007) Identification of management units using population genetic data. *Trends In Ecology & Evolution* **22**, 11-16.

Paxton R, Giovanetti M, Andrietti F, Scamoni E, Scanni B (1999) Mating in a communal bee, *Andrena agilissima* (Hymenoptera Andrenidae). *Ethology Ecology and Evolution* **11**, 371-382.

Paxton R, Thorén P, Tengo J, Estoup A, Pamilo P (1996) Mating structure and nestmate relatedness in a communal bee, *Andrena jacobi* (Hymenoptera, Andrenidae), using microsatellites. *Molecular Ecology* **5**, 511-519.

Paxton RJ (2005) Male mating behaviour and mating systems of bees: an overview. *Apidologie* **36**, 145-156.

Perrin N, Mazalov V (2000) Local competition, inbreeding, and the evolution of sex-biased dispersal. *The American Naturalist* **155**, 116-127.

Pierson JC, Allendorf FW, Drapeau P, Schwartz MK (2013) Breed locally, disperse globally: fine-scale genetic structure despite landscape-scale panmixia in a fire-specialist. *PLoS ONE* **8**, e67248.

Piry S, Alapetite A, Cornuet J, Paetkau D (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95***,* 536-539*.*

Potts SG, Biesmeijer JC, Kremen C, Kremen C, Neumann P, Schweiger O, Kunin WE (2010) Global pollinator declines: trends, impacts and drivers. *Trends in Ecology and Evolution* **25**, 345-353.

Pusey A, Wolf M (1996) Inbreeding avoidance in animals. *Trends in Ecology and Evolution* **11**, 201-206.

Queller DC, Goodnight KF (1989) Estimating relatedness using molecular markers. *Evolution* **43,** 258-275.

Richards CM (2000) Inbreeding depression and genetic rescue in a plant metapopulation. *The American Naturalist* **155**, 383-394.

Ritland K, (1996) A marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* **152**, 1062-1073.

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

Shafer ABA, Northrup JM, White KS Boyce MS, Côté SD, Coltman DW (2012) Habitat selection predicts genetic relatedness in an alpine ungulate. *Evolution* **93**, 1317–1329.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264-279.

Smith BE, Marks P, Gardescu S (1993) Two hundred years of forest cover changes in Tompkins County, New York. *Bulletin of the Torrey Botanical Club* **120**, 229-247.

Smith BH (1983) Recognition of female kin by male bees through olfactory signals. *PNAS* **80**, 4551-4553.

Stelkens RB, Jaffuel G, Escher M, Wedekind C (2012) Genetic and phenotypic population divergence on a microgeographic scale in brown trout. *Molecular Ecology* **21**, 2896-2915.

Suni SS, Bronstein JL, Brosi BJ (2014) Spatio-temporal genetic structure of a tropical bee species suggests high dispersal over a fragmented landscape. *Biotropica* **46**, 202-209.

Ulrich Y, Perrin N, Chapuisat M (2009) Flexible social organization and high incidence of drifting in the sweat bee, *Halictus scabiosae*. *Molecular Ecology* **18**, 1791-1800.

Van Noordwijk AJ, Scharloo W (1981) Inbreeding in an island population of the great tit. *Evolution* **35**, 674-688.

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

Wang J (2002) An estimator for pairwise relatedness using molecular markers. *Genetics* **160**, 1203-1215.

Wang J (2006) Informativeness of genetic markers for pairwise relationship and relatedness inference. *Theoretical Population Biology* **70**, 300-321.

Wang J (2009) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions B*.

Wang J (2010) COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources* **11**, 141-145.

Wang J (2014). Marker- based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *Journal of evolutionary biology* **27**, 518-530.

Wcislo WT (1987) The role of learning in the mating biology of a sweat bee *Lasioglossum zephyrum* (Hymenoptera, Halictidae). *Behavioral Ecology and Sociobiology* **20**, 179-185.

Wcislo WT (1992) Attraction and learning in mate-finding by solitary bees, *Lasioglossum* (*Dialictus*) *figueresi* Wcislo and *Nomia triangulifera* Vachal (Hymenoptera, Halictidae). *Behavioral Ecology and Sociobiology* **31**, 139-148.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177–1191.

Wright S (1978) Vol. 4: Variability within and among natural populations *University of Chicago Press.*

Zayed A, Packer L (2002) Genetic differentiation across a behavioural boundary in a primitively eusocial bee, *Halictus poeyi* Lepeletier (Hymenoptera, Halictidae). *Insectes Sociaux* **49**, 282-288.

Zayed A, Packer L (2005) Complementary sex determination substantially increases extinction proneness of haplodiploid populations. *PNAS* **102**, 10742-10746.