3
4
5 **Phylogeographic assessment of mtDNA paraphyly and the evolution of**
6 **unisexuality in *Calligrapha* (Coleoptera: Chrysomelidae)**

7
8 Running title: *Calligrapha* phylogeography and unisexuality

9
10
11 JESÚS GÓMEZ-ZURITA

12 &

13 ANABELA CARDOSO

14
15
16 *Animal Biodiversity and Evolution, Institute of Evolutionary Biology*
17 *(CSIC-Universitat Pompeu Fabra), 08003 Barcelona, Spain.*

18
19
20 **Correspondence:**

21 Jesús Gómez-Zurita

22 Institute of Evolutionary Biology (CSIC-Univ. Pompeu Fabra)

23 Pg. Marítim de la Barceloneta 37-49

24 08003 Barcelona, Spain

25 Phone: +34 93 2309643

26 E-mail: j.gomez-zurita@csic.es

27
28
29 *Keywords:* Cryptic species; glacial refugia; hybridization; parthenogenesis; range
30 expansions.

Abstract. The leaf beetle genus *Calligrapha* is one of the few examples of animals with several obligate unisexual, female-only species. Previous work showed that each one arose independently from interspecific hybridization events involving different species. However, all of them clustered in a single mtDNA clade together with some individuals of the parental bisexual species, which appeared as deeply polyphyletic in the mtDNA genealogy of the genus. The dating of these splits using a molecular clock placed them in the Quaternary and it was hypothesized that climatic change during this period may have favored range expansions and secondary contacts required for hybridization. In this work, we test this hypothesis and the origins of unisexuality in *Calligrapha* examining the diversity of mitochondrial (*cox1*) and nuclear (*wingless*, *Wg*) genes and the Bayesian continuous mtDNA phylogeography of a sample of more than 500 specimens of two bisexual species of *Calligrapha* at a continental scale and two unisexual species derived from them. Besides a major topological difference, whereby each bisexual species is monophyletic for *Wg* but paraphyletic for *cox1*, both gene datasets are consistent with a minimum of seven evolutionary lineages, coherent with geography and consistent with an ordered expansion to occupy their current ranges. The results also imply their survival in well-established glacial refuges during the Last Glacial Maximum (LGM). Thus, for bisexual *C. multipunctata* there are two main, southern and northern lineages. The southern lineage expanded its range in two evolutionary branches, to the Rocky Mountains and to the northern Mississippi and Ohio River basins, respectively. The northern lineage has one branch in the Upper Mississippi, and one that expanded west to the Pacific Northwest and east to the northeastern North Atlantic, finding refuge in both areas. These major lineages are parapatric in the Northern Great Plains, an area consistent with them having found refuge in the so-called Driftless region during the LGM. For bisexual *C. philadelphica*, one northern lineage expanded west from the northern Appalachians and one east and southwest along the axis of the Appalachians, and the timing of events is consistent with their persistence in glacial refugia at both sides of the main Great Lakes lobe of the Laurentide Ice Sheet. There is evidence that the northeastern North Atlantic lineages of both species hybridized at the edge of their ranges after the LGM. The additional, divergent mtDNA lineage of these species shows evidence of range expansions of two lineages, one for each species, in an area south of the Laurentide Ice Sheet, and giving origin to the unisexual species by way of hybridization with other species in the Alleghanian region after the LGM. Interestingly, the individuals of supposedly bisexual

65 species in this clade are all females. This means that unisexuality actually predates the
66 origin of unisexual taxa in this system and that some bisexual species in *Calligrapha*
67 may be species complexes instead, with cryptic species differing in their reproductive
68 mode.

69

70

Introduction

There is a relatively small, but taxonomically widespread number of emblematic examples of organisms among all animal phyla that reproduce by means other than the vastly predominant process of sexual reproduction (Suomalainen, Saura & Lokki, 1987; Avise, 2008). Understanding the evolutionary peculiarities of these reproductively eccentric animals has been recognized, for instance, as our best chance to fathom precisely the motives behind the unquestionable success of sexuality (e.g., Bell, 1982). For years, we have been interested in the study of the mechanisms originating unisexual reproduction in one of these examples, a group of beetles with several obligate unisexual, female-only species in North America (Robertson, 1966), the leaf beetle genus *Calligrapha* Chevrolat. This is a large genus of herbivore beetles distributed from Alaska to Argentina, with some 130 species, one quarter of them from a distinctive North American lineage particularly diverse in the Great Lakes region (Gómez-Zurita, 2005, 2015; Montelongo & Gómez-Zurita, 2014). An ecological and evolutionary transition perhaps favored by the humid and warmer conditions during the Pliocene, shifted the preference of this lineage from steppic mallows and composites to diverse trees associated to waterways (Montelongo & Gómez-Zurita, 2014). Very early in the phylogenetic studies of North American *Calligrapha*, it was unveiled an extended pattern of incongruence between mitochondrial markers and taxonomy, ecology and nuclear markers (Gómez-Zurita, Vogler & Funk, 2004; Gómez-Zurita, Funk & Vogler, 2006). This inconsistency, which was so rampant as to showing every species for which we had samples from several sources as polyphyletic for mtDNA, was taken as part of the evidence pointing to interspecific hybridization as a leading process in the origin of unisexual species in the genus (Gómez-Zurita et al., 2006). More recently, the analysis of allelic diversity of three single copy nuclear genes in five of the six unisexual *Calligrapha* species and a representative sample of bisexual species allowed validating this hypothesis. Indeed, individuals of unisexual species proved to be genetic mosaics sharing alleles with two and up to three bisexual species, extant representatives of the lineages involved in independent historical interspecific hybridization events (Montelongo & Gómez-Zurita, 2015). Interestingly, this study also revealed some unexpected and intriguing findings. Unisexual species were polyphyletic for all tested nuclear markers and the result of anastomosis of few identifiable parental lineages, different for each unisexual species. However, the mtDNA diversity present in all

unisexual species belonged to a single old evolutionary lineage of uncertain taxonomic attribution, which we named U-clade, and which was sister to a B-clade with most individuals of bisexual species. In fact, this mtDNA lineage, the U-clade, also included representatives of the bisexual species recognized as being involved in the interspecific crosses that led to each of the unisexual species, generalizing the idea of species mtDNA paraphyly in this group. The most parsimonious explanation for this complex pattern required a single ancestral species that introgressed its mtDNA some 2-3 Ma ago via hybridization into several lineages of *Calligrapha* and without fixation in the populations, so that they remained polymorphic for their mtDNA, as observed today (Montelongo & Gómez-Zurita, 2014, 2015). Later, already in the Pleistocene, species in these lineages were involved in new waves of hybridization events, and when these crosses involved females carrying the introgressed DNA, the resulting hybrid lineages would persist and eventually become unisexual.

The current distribution of most *Calligrapha* species in North America, particularly of hybrid unisexual species, around the Great Lakes area (Gómez-Zurita, 2005), as well as the estimated age of hybridization waves in these leaf beetles strongly suggest that climate change may have had a major role as the driver of range changes, secondary contacts and evolutionary processes in this group, as demonstrated for several other organisms, mainly in plants (e.g., Swenson & Howard, 2005; Beatty & Provan, 2011; Peterson & Graves, 2016). Indeed, this hypothesis is unquestionably orthodox, considering that current species and genetic pool ranges in North America are generally understood taking into account the oscillating climatic change during the Quaternary (Hewitt, 2000, 2004). In particular, this model features predominantly the change imposed by the Last Glacial Maximum, which covered with glaciers most of North America, with the massive Laurentide Ice Sheet reaching at its maximum extent as far south as latitude 39°N in today's southern Ohio and Indiana, some 21.8 ka BP (Margold, Stokes & Clark, 2018). It is assumed that these oscillating climatic conditions shifted the distribution of biomes along a latitudinal axis (Hewitt, 2000), with successive cycles of recolonization involving dispersal also in the east-west axis (e.g., Soltis et al., 2006; Beatty & Provan, 2010, 2011), a circumstance that makes it challenging to infer past species distributions based on their current ranges. Surviving populations of temperate species remained in multiple refugia during the height of glaciations (Beatty & Provan, 2010), in some cases relatively close to the ice sheet and colder tundra and boreal forest

biomes, as mounting evidence suggests (Rowe, Heske, Brown & Paige, 2004; Soltis et al., 2006; Beatty & Provan, 2010). One solid example of the latter is the so-called Driftless Region, east to the Northern Great Plains, between the Des Moines and Green Bay lobes of the Laurentide Ice Sheet in the Wisconsinan glaciation (Mickelson & Colgan, 2003; Rowe et al., 2004; Peterson & Graves, 2016). In every interglacial, these ice age survivors colonized newly available habitat after the retreating ice and the re-establishment of favorable conditions in previously glaciated areas relatively quickly, at least in the case of plants (Clark et al., 1998). These migrations took advantage of the physiographic configuration of North America, offering limited barriers to south-north dispersal (Soltis et al., 2006), but also to the mentioned fact that temperate species may have remained close to the newly available habitat as suggested by the analysis of dispersal rates and recent phylogeographic evidence in plants (e.g., Clark et al., 1998; Rowe et al., 2004; Snell & Cowling, 2015; Peterson & Graves, 2016). The cyclical range expansions throughout Ice Ages changed the distribution of species following a variety of, sometimes relatively congruent patterns (e.g., Soltis et al., 2006). Secondary encounters of previously isolated lineages resulted in the establishment of hybrid and suture zones (e.g., Swenson & Howard, 2005) and possibly triggered phylogeographic processes, such as the hypothesized leading-edge effect, which implies that early habitat colonization by a lineage constrains the range expansion of subsequent lineages (Hewitt, 1993, 2004).

Testing these ideas and scenarios for *Calligrapha*, including the origin of hybrid unisexual lineages due to range shifts favoring secondary encounters of formerly isolated evolutionary lineages, requires an understanding of the phylogeographic history of the group in a broader context. The advent of molecular methods and the formalization of phylogeography as a discipline (Avise, 2000) found almost immediately appealing subjects of study among species or species groups with alternative reproductive modes, both to understand the history and origin of unisexual lineages, but also to investigate how differences in reproductive biology might affect colonization potential (e.g., Weider, Hobaek, Hebert & Crease, 1999; Jensen, Enghoff, Frydenberg & Parker, 2002; Law & Crespi, 2002; Facon et al., 2003). Yet, phylogenetic studies trying to integrate the history of episodic interspecific hybridization events and the origin of unisexual species usually examine the problem through coarse-grained approaches, above the population level (e.g., Gómez-Zurita et al., 2006; Grismer et al.,

2014; Freitas et al., 2016). The resolution and explanatory power attained when similar genealogical approaches are applied at a phylogeographic or population level are obviously much better. They have the potential to capture not only local levels of genetic variation, but also to offer a better grasp on recent and contemporaneous processes consistent with the relatively recent origins, during the Pleistocene, of most of the unisexual species known to date (Kearney, 2005), including those in *Calligrapha* (Montelongo & Gómez-Zurita, 2014). Accordingly, this type of population-level analyses has offered the power and resolution to estimate single or multiple origins of unisexuality or to support, for example, the role of climate change, through glaciations and aridification, as a significant driving force behind the changes in the ranges of bisexual species, with colonization favoring the origin of unisexuality (e.g., Pongratz, Storhas, Carranza & Michiels, 2003; Strasburg, Kearney, Moritz & Templeton, 2007; Kearney & Blacket, 2008; Buckley, Marske & Attanayake, 2009; Freitas et al., 2016).

In this study, we will investigate the phylogeography of two bisexual species of *Calligrapha* with transcontinental distributions in North America, and two unisexual species evolutionarily derived from each of the former. This will be accomplished by characterizing the diversity, distribution and consistency of one mtDNA marker and one fast-evolving single-copy nuclear marker throughout the species ranges and across time and space, distinguishing between the evolutionary fates of the mtDNA B- and U-clade lineages and the origins of taxonomically recognizable unisexual lineages. We aim at linking the inferred patterns with the expected general processes triggered by Pleistocene glaciations both taking into account the time framework for phylogeographic events and the geographic distribution of evolutionary lineages.

Material and methods

Sampling, mtDNA and nuclear gene typing of Calligrapha

Four species of *Calligrapha* were sampled across their native ranges, including two bisexual and two unisexual species. Bisexual species included the dogwood specialist *C. philadelphica* (L.) and two subspecies of the willow specialist *C. multipunctata* (Say). In turn, unisexual species were *C. vicina* Schaeffer and *C. suturella* Schaeffer, closely related by descent to and sharing ecologies with each of the previous two, respectively

(Gómez-Zurita et al., 2006; Montelongo & Gómez-Zurita, 2015). In total, 326 specimens of *C. multipunctata bigsbyana* (Kirby), 24 of *C. multipunctata s. str.*, 151 of *C. philadelphia*, eight of *C. suturella* and seven of *C. vicina* were available for analysis (Appendix 1; Supporting Information Figure S1). Unisexual species have much smaller ranges, very low densities and typically coexist with more abundant bisexual relatives, which makes them considerably difficult to sample. Non-destructive, whole-body total DNA extractions were carried out using the DNeasy Blood & Tissue kit (Qiagen Iberia, Madrid). The exoskeleton of the beetle was retrieved after the extraction and mounted dry, together with prepared genitalia for future reference in the IBE-JGZ collection (Barcelona, Spain). Total DNA was used as template for amplification of a fragment of the cytochrome *c* oxidase subunit 1 (*coxI*) gene with the primers C1-J-2183 (5'-CAACATTTATTTTGATTTTTT) and TL2-N-3014 (5'-TCCAATGCACTAATCTGCCATATTA) of Simon et al. (1994) and of an intronless region of the wingless (*Wg*) gene with the primers Wg550F (5'-ATGCGTCAGGARTGYAARTGYCAYGGYATGTC) and WgAbRZ (5'-CACTTNACYTCRCARCACCARTG), designed by Wild & Maddison (2008) and using the same conditions that we have used successfully for *Calligrapha* over the years (e.g., Gómez-Zurita et al., 2006; Montelongo & Gómez-Zurita, 2014). PCR products of 829 (*coxI*) and 472 (*Wg*) bp were purified with 5M ammonium acetate and isopropanol and sequenced in both directions using the same PCR primers and Big Dye Terminator 3.1 technology on a capillary sequencing 3730xl DNA Analyzer (Applied Biosystems, Foster City CA). Sequences were edited and contigs assembled using Geneious R9 (Biomatters Ltd., Auckland). Newly obtained *coxI* sequences were deposited in the European Nucleotide Archive (EMBL-EBI) under sequence accession numbers LR030286-LR030371 (*coxI*) and LR030372-LR030440 (*Wg*). The multiple sequence alignments used for the study were deposited in the repository TreeBASE and retrievable in the URL: <http://purl.org/phylo/treebase/phyloids/study/TB2:S23718>.

Genetic markers diversity and gene genealogies

Length invariant *coxI* and *Wg* data were aligned manually with the help of their amino acid translations to generate matrices of 729 and 472 homologous nucleotide positions, respectively. In the case of *Wg* data, ambiguous positions of heterozygotes were resolved in most cases by reference to known alleles in the population as observed in

homozygotic individuals in the alignment or, in a second round of inference, considering the new alleles deduced in the previous step. Moreover, relevant in the case of unisexual species of known hybrid origin, these data were aligned and compared with the homologous sequences of putative parental *Calligrapha* species from Montelongo & Gómez-Zurita (2015). Information on sample diversity, including segregating sites, haplotype and nucleotide diversity, and theta, as a measure of divergence, as well as measures of heterozygosity in the case of *Wg*, were obtained using DnaSP 5.10.1 (Librado & Rozas, 2009). The same software was used to calculate the neutrality D statistic and its significance (Tajima, 1989). Polymorphism measures for *Wg* data were estimated from an alternative automatic phasing of alleles in DnaSP using PHASE (Stephens, Smith & Donnelly, 2001) and a model accepting recombination. The mtDNA and nuclear gene matrices with all available sequences were used to generate unrooted X-trees, i.e. allowing for observations in all vertices, based on the principle of statistical parsimony using TCS 1.21 (Clement, Posada & Crandall, 2000).

Congruence analysis of mtDNA and wingless data: association test

Given our previous knowledge and expectation of marked mtDNA species paraphyly in this system, assessing congruence between datasets is of major importance in this phylogeographic study. In order to make a coherent assessment of the phylogeography of the species (not of their mtDNA only), it was crucial to identify evolutionary lineages not contradicted by any of the data available. Given the characteristics of *Wg* data (see results), standard phylogenetic analyses for formal congruence analysis with the *coxI* phylogeny were not straightforward, and to test for potential associations between mtDNA and *Wg* diversity, we used permutation contingency tests of the latter on a *coxI* genealogy by means of nested clade analysis (Templeton & Sing, 1993). The observed diversity of *Wg* (alternatively coded as alleles or genotypes) was tested for associations with the statistical parsimony genealogy of *coxI* haplotypes upon transformation of the latter into a categorical variable by interpreting the *coxI* X-tree as a nested-clade design following the algorithm of Templeton, Boerwinkle & Sing (1987). Permutation Chi-square exact contingency tests of the variables defined by *Wg* alleles/genotypes on the structure defined by the *coxI* genealogy were carried out for all contingency matrices where both variables had two or more states. The tests and their significance were carried out using 9999 data permutations for each contingency matrix using the function

perm.ind.test of the R package wPerm 1.0.1 (Weiss, 2015). Permutation of the contingency matrices simulated the null hypothesis of no association between genetic differences of mtDNA and nuclear markers, and a significant result was interpreted as the existence of statistical differences between the nuclear genetic backgrounds of tested mtDNA groups. A graphical interpretation of this same association was also obtained by plotting mtDNA diversity on the statistical parsimony genealogy of *Wg* alleles using PopART (Leigh & Bryant, 2015).

Continuous phylogeography of Calligrapha

Simplified matrices of either B or U haplotypes (using the nomenclature for 'bisexual' and 'unisexual' haplotypes of Montelongo & Gómez-Zurita, 2015) by excluding sequences representing both redundant haplotype and location data were used for continuous phylogeography inference (Lemey, Rambaut, Welch & Suchard, 2010). This Bayesian method uses genetic georeferenced data from extant individuals to coestimate their genealogical relationships and other population parameters together with an inference of the ancestral population locations. To this end, continuous phylogeography treats geographic coordinates as inherited traits that co-vary with the phylogeny. Both matrices were assessed for their optimal evolutionary model among 88 alternatives using jModelTest 2.1.10 (Darriba, Taboada, Doallo & Posada, 2012) under ML optimization with NNI tree rearrangements. Bayesian phylogenetic inference in BEAST 1.8.4 (Drummond, Suchard, Xie & Rambaut, 2012) was used to generate the sample of trees to estimate the ancestral areas of all nodes in the genealogy by treating geographic coordinates of the tree terminals (e.g. individuals with a given haplotype found in a certain location) as continuous traits (Lemey et al., 2010). Two independent BEAST analyses were run for each dataset for 350 million generations, sampling one tree every 35000 trees. The nucleotide substitution model was set based on the results of jModelTest and the bivariate continuous trait model as a lognormal Repeated Random Walk, adding random jitter to tips (factor=0.001) to avoid having two samples with exactly the same coordinates. An uncorrelated lognormal relaxed clock (Drummond, Ho, Phillips & Rambaut, 2006) and a constant size coalescent tree prior (Kingman, 1982) were specified in both analyses. The results of the analyses were inspected for optimal effective sample size values in all parameters in Tracer 1.6.0 (Rambaut et al., 2014). A summary tree with all nodes annotated with relevant parameters, particularly inferred

ancestral location traits, was obtained with TreeAnnotator 1.8.4 (Drummond et al., 2012) and after discarding 15-20% of initial trees. The Maximum clade credibility trees annotated with continuous location traits were mapped onto the geography using SpreaD3 (Bielejec et al., 2016) to inspect the history of spatial distribution of lineages, including the timing of their spread based on previous analyses. *Wg* data was not subject to this type of analysis due to the relatively poorly structured diversity and high levels of heterozygosity, with a reduced number of alleles that were present in most individuals and throughout the whole range. Moreover, trials focusing on the dispersal of alleles—not individuals, since the method does not account for biallelic condition of a diploid system—proved useless because of persistent failure to homogenize parameters, mostly related to the coalescent process. In this case, the geographic patterns of alleles were investigated through georeferencing and their association with mtDNA lineages.

Results

MtDNA diversity of Calligrapha

The sample of 516 *Calligrapha* specimens yielded 504 *cox1* sequences making up to 86 haplotypes in two divergent groups (Table 1 and Appendix 1): B1-B73, only found in the three bisexual taxa included in the study, and U1-U13, found in *C. multipunctata bigsbyana* and *C. philadelphica*, and the two unisexual species. Haplotypes were nearly always exclusive of one of the species, except in the case of B1, shared between *C. multipunctata bigsbyana* and eight specimens of *C. philadelphica* from New Brunswick; B33, between the two subspecies of *C. multipunctata*; U3, shared between *C. philadelphica* and *C. vicina*; and U5, between *C. multipunctata bigsbyana* and *C. suturella*. Haplotype and nucleotide diversities were highest in *C. philadelphica* and lowest in the two unisexual species, and these two subsets of data were also the only ones producing positive values for Tajima's D, albeit non-significant in all cases (Table 1).

The average uncorrected nucleotide divergence between both groups of haplotypes was 0.04485, more than twice the average proportion of nucleotide differences in the most diverse of these groups, the one including all B haplotypes ($\theta=0.01526$). Haplotype

relationships within each of these two groups were best represented as unrooted X-trees instead of bifurcating phylogenetic trees, since nearly half of internal nodes were occupied by observed haplotypes in the case of the B group, and all internal nodes in the case of the U group (Figure 1). The haplotypes of *C. multipunctata* appeared in two paraphyletic groups, one by inclusion of at least two groups of *C. philadelphia* haplotypes in the B-clade and subdivided in three main subgroups, and one of *C. suturella* haplotypes in the U-clade. In turn, *C. philadelphia* appeared as polyphyletic in as many as four different haplotype groups, three in the B-clade, with one of them paraphyletic with *C. multipunctata bigsbyana*, and one in the U-clade, paraphyletic with *C. vicina*.

Wg diversity in Calligrapha

We obtained *Wg* sequences for 472 specimens, including 293 individuals of *C. multipunctata bigsbyana*, 22 of *C. multipunctata s. str.*, 143 of *C. philadelphia*, and seven each of *C. suturella* and *C. vicina*. Up to eight *Wg* alleles in *C. multipunctata* (m1-m8) and seven in *C. philadelphia* (p1-p7) were directly observed from homozygous individuals in the sample, and 46 extra alleles (43 in *C. multipunctata* and three in *C. philadelphia*) were deduced as resulting from combinations of one of these 15 alleles with another sequence variant, differing by a single point mutation and rarely changing the translated amino acid (only seven of the deduced mutations would imply an amino acidic change). Assuming that these alleles were present in the sample, it was possible to deduce eight extra alleles, six of them differing by 2-3 nucleotide substitutions from an observed allele. Automatic phasing of alleles yielded nearly identical allele inferences as the manual, sequential procedure based on individual examination of polymorphic sites in the multiple sequence alignment. In all, 58 *Wg* alleles were characterized for *C. multipunctata s. l.* and *C. suturella*, and eleven for *C. philadelphia* and *C. vicina*. The two unisexual species of hybrid origin, and as expected based on the results of a previous study (Montelongo & Gómez-Zurita, 2015), showed sequence data compatible with additionally including *Wg* alleles from *C. ignota* (two alleles) in the case of *C. suturella*, and from *C. rowena* (one allele) in the case of *C. vicina*. Figure 2a shows the frequencies of *Wg* alleles in the sample (considering their double representation in homozygous individuals), highlighting the clear predominance of the m1 allele in the populations of *C. multipunctata s. l.* (allelic frequency in this

species, $p_{m1} = 0.456$), followed by $m2$ ($p_{m2} = 0.129$), $m4$ ($p_{m4} = 0.052$) and $m6$ ($p_{m6} = 0.049$), and a more balanced contribution of alleles in the case of *C. philadelphica*, with $p1$ ($p_{p1} = 0.290$), $p2$ ($p_{p2} = 0.218$), $p6$ ($p_{p6} = 0.171$) and $p3$ ($p_{p3} = 0.130$) as the most prevalent in the populations. In *C. multipunctata s. l.* (and *C. suturella*), 39 alleles appeared as doubletons or singletons and represented only 7.55% of the global allelic diversity, while in the case of *C. philadelphica* (and *C. vicina*), these corresponded to four alleles or 2.05% of allelic diversity. Nucleotide diversity was lower in bisexual ($\pi = 0.00448$ - 0.00656) than in unisexual species ($\pi = 0.00913$ - 0.01227), the latter expectedly in the range of interspecific measures, since they are genetic mosaics with genomic contributions from several species as demonstrated previously (Montelongo & Gómez-Zurita 2015) (Table 2).

The genealogy of *Wg* alleles is shown in Figure 3, where those from *C. multipunctata* and *C. suturella* appear separated, albeit by a small number of mutations, from those of *C. philadelphica* and *C. vicina* (the highly divergent $p11$ allele is resolved with a somewhat intermediate position). The two species groups showed clearly contrasting patterns, with *C. multipunctata s. l.* presenting a typical star-tree pattern with a core of two highly frequent alleles and numerous, rare allelic variants directly derived from them usually by single mutations, and *C. philadelphica* with lower diversity, more structured and relatively balanced, also in terms of allele frequencies. Consistently with the observed differences in genealogical patterns, *C. multipunctata s. l.* yielded a significantly negative *D* value ($D = -2.0542$) in the Tajima's test of neutrality, while in the case of *C. philadelphica* it was slightly negative too ($D = -0.5201$), but without statistical significance (Table 2).

The combinations of alleles produced 113 genotypes as characterized in the sample, with 36.02% of individuals presenting the homozygotic condition for 15 alleles, whereby three alleles ($m1$, $p1$ and $p2$) accounted for nearly three quarters of these specimens (Figure 2b). Most individuals were heterozygous, displaying up to 98 *Wg* genotypes, of which six characterized 37.42% of heterozygotes: $m1/m2$, $m1/m6$, $m2/m10$ in *C. multipunctata*; $m1/i1$ in both *C. multipunctata bigsbyana* and *C. suturella*; and $p1/p6$ and $p2/p3$ in *C. philadelphica*. Thirteen genotypes were doubletons and 57 singletons, about half of them representing in most cases combinations of an inferred allele with one of the abundant $m1$, $m2$, $p1$ or $p2$ alleles (Figure 2c). Gene

diversity was relatively high in these species, ranging from 0.699 in *C. multipunctata* to 0.810 in *C. philadelphica* (with three alleles shared by all specimens of polyploid *C. vicina*, the obtained heterozygosity value, $H_d = 0.538$, is only a rough estimation).

Geographic distribution and genetic associations between cox1 and Wg

Some of the most abundant *Wg* alleles also had a widespread distribution throughout the range of the species, including m1, m2, m5, p2 and p5 (Supporting Information Figure S2). Other *Wg* alleles and *cox1* haplotypes showed restricted distributions in different parts of the species ranges. Considering the genetic variants present in two or more localities, these could be roughly defined as Pacific Northwest (B14), Pacific northwest to Northern Great Plains (m4, m6), Rocky Mountains (B33), Rocky Mountains and southern Mississippi (m12), eastern Alberta plains (B27, B28), Northern Great Plains (B9, B11, B16, B24, B50, B53, U9), Upper Mississippi and Ohio rivers (B49), Northern Great Plains to Appalachians (U1), Northern Great Plains and Northeast North America (B1, m7, m8, m10, m24, m26, m29, m35, m56, p1, p6, p10), Great Lakes to New England (U5), Lakes Erie and Ontario and Saint Lawrence river (B29, B42, B59, U11), Saint Lawrence river, northern Appalachians to Nova Scotia (B18, B63), Appalachians (U3, p3), northern Appalachians (B67, m3, m21, m22, m39, p4, p9), Gaspé peninsula and Nova Scotia (U8, m9, m27), northern Appalachians and southern Mississippi (m16, m38), and southern Mississippi (B46, B57, B58).

Apart from this intuitive geographic correspondence of some of the genetic diversity in the sample, the statistical tests of independence based on the full genealogical structure of *cox1* revealed significant differentiation for *Wg* composition, both in terms of genotypes and also for alleles, between the B- and the U-clades, but also among the three main groups as defined by the nested clade design, namely B4-1, B4-2 and B4-4 (Table 3). When the data were partitioned by species (not shown), given that the basic structure of the nested design was retained, the test also revealed statistically significant genotypic differences for the nuclear marker among the three main B-lineages of *C. multipunctata* ($N = 302$, chi-square = 226.3561, $P = 0.003$) and the three main B-lineages of *C. philadelphica* ($N = 105$, chi-square = 76.0232, $P < 0.001$). Within each of these lineages, at lower divergence levels, statistically significant associations were found mainly between and within nested clades of *C. philadelphica* (Table 3). The

evolutionary lineage carrying U-haplotypes showed significantly different *Wg* characteristics between U1-2 (*C. multipunctata* and *C. suturella*), carrying *C. ignota* and m-type *Wg* alleles, and the other nested clades (*C. philadelphia* and *C. vicina*), exclusively carrying *C. rowena* and p-type *Wg* alleles. In summary, major mtDNA lineages did not show statistically significant intra-lineage differentiation in their *Wg* diversity, but these differences were significant between the lineages.

Between the B- and U-clades of *C. multipunctata* s. l., despite their large difference in size, there were obvious genotypic differences for *Wg* owing to the presence in all but one individuals in the latter of a mosaic condition with *Wg* alleles from *C. ignota*, but also by the presence of an exclusive allele, m54, in 15% of the individuals. The three main B-lineages of *C. multipunctata* showed differences in their allelic and genotypic characteristics. The samples in clade B4-1, distributed at the highest sampled latitudes (north of parallel 42°) from the Pacific northwest to Nova Scotia, had several exclusive alleles present in three or more individuals, including m3, m9, m23 and m27, as well as up to 27 double- or singleton alleles, a significantly large proportion of individuals homozygotic for m1 and m1/m2 heterozygotes, and up to 48 exclusive genotypes. Clade B4-2, distributed in the sampled areas within the Rocky Mountains and in the southern Mississippi river, showed a predominance of m4, m5 and m6 alleles, as well as m12 in exclusivity, and a number of exclusive genotypes, generally involving some of these dominant alleles (m1/m12, m2/m5, m2/m12, m5/m5 and m5/m6). Finally, clade B4-3, mainly distributed in the northern central Plains and the Mississippi and Ohio rivers basins, showed proportionally low frequencies of m4 and m6, and up to nine exclusive alleles, although at low frequency, and twelve exclusive genotypes. In the case of *C. philadelphia*, there were also genetic and genotypic differences for *Wg* among the three main mtDNA groups: (1) individuals in B3-1, in the northeastern extreme (Maine and New Brunswick) of the sampled range, had a predominance of p2 and p6 alleles, present in homozygosis or as p2/p6 heterozygotes, the latter in exclusivity (the genotype p6/p10 was also exclusive of this group); (2) B3-2 *C. philadelphia*, in northern parts of the range in southern Alberta and eastern Ontario, had the allele p11 in exclusivity and two unique genotypes (p3/p5 and p5/p10); and (3) the specimens in B3-4, mainly an Appalachian lineage, had a predominance of p1 and p3 alleles, mainly in homozygosis), with p7 and p8 as exclusive alleles, and up to eight private genotypes (p1/p3, p1/p4, p1/p8, p2/p3, p2/p4, p5/p5, p5/p6, p7/p7).

MtDNA phylogeography of *Calligrapha*

The evolutionary models best accounting for the observed variation in the *cox1* matrices for U and B haplotypes were HKY and HKY+I models, respectively (the optimal evolutionary model for the full dataset was an HKY+G+I). The time-calibrated inferred Bayesian phylogeographic trees implementing these models were consistent with the statistical parsimony genealogy, with up to five lineages with high posterior probability, four in the B-clade and one corresponding to the U-clade (Figure 4). These different mtDNA lineages, otherwise homogeneous for *Wg*, constitute sound evolutionary units for phylogeographic assessment and will be referred to by explicit names hereafter: $m_{NL}+m_{CL}$, m_{ML} , p_{EL} and p_{WL} in the B-clade, and $m_{UL}+p_{UL}$ in the U-clade (Figure 4). The split of both divergent clades was estimated at 5.12 Ma with a large confidence interval reaching up to 9.46 Ma, but their diversification was dated much more recently, between 0.48-1.09 Ma in the case of the *C. multipunctata* and *C. philadelphica* B-clade and between 0.06-0.22 Ma in the case of the U-clade involving the same species and their unisexual counterparts. A loose taxonomic and phylogeographic pattern could be appreciated with two groups for every bisexual lineage of both species: one northern lineage across the continent and one lineage largely parapatric with the former in the Great Lakes, New England and eastern Canada, but reaching localities further south, as far south as Texas and the Mississippi mouth in the case of *C. multipunctata*, and the Blue Ridge mountains in the case of *C. philadelphica*. Moreover, data were consistent with very recent genetic exchange between both species in the NE edge of their common ranges. In turn, the U-clade branches represented by these species, of very recent origin, included only samples from the NE range of *C. multipunctata* and the eastern populations of *C. philadelphica*, including the southernmost localities in North Carolina and Georgia. The continuous phylogeographic analyses of these lineages suggested a history of range expansions of each lineage across the continent. Bisexual *C. philadelphica* would have had an origin around the St. Lawrence basin in NE North America, and expanded west toward Alberta, and eastward and southward along the Appalachians (Figure 5). Bisexual *C. multipunctata* would have had two centers of expansion, one from the southern Mississippi basin and one from the Great Lakes area and both expanding across North America, meeting in the northern central Plains and in the western slopes of the Appalachians (Figures 6 and 7). The lineages of these species

eventually involved in the origin of unisexual species originated in the Great Lakes area and expanded east and westward in the case of *C. multipunctata* and also southward in the case of *C. philadelphica* (Figure 8).

Discussion

Suitability of mtDNA to investigate the phylogeography of Calligrapha

This study describes the first phylogeographic analysis to date for any *Calligrapha* species. The results are consistent with elements, including the timing of events, that suggest that the conditions and processes affecting North America in the Pleistocene may have modeled the current species ranges, their diversification and the origin of hybrid lineages, including the unisexual species. Due to our phylogenetic approach to substitution rates, age estimates may be inflated (Ho & Larson, 2006), and thus our results are in fact conservative about the role of glaciations. It becomes apparent that, despite initial concerns that mtDNA might not be a good proxy for this type of study in these animals because of species polyphyly (Montelongo & Gómez-Zurita, 2015), there is in fact a close association of mtDNA diversity with geographic distribution and also with the underlying nuclear genetic background, at least for the main evolutionary lineages and for the nuclear marker of choice here. In fact, despite the lack of clear juxtaposed phylogenetic structure for these markers and the marked discordance in a major pattern, i.e. species paraphyly for mtDNA and reciprocal monophyly for *Wg* (the reader is referred to Montelongo & Gómez-Zurita [2015] for detailed explanations on this pattern), the hierarchical association study helped unveiling some interesting patterns both to understand the history of these *Calligrapha* species and the origins of unisexuality. Species mtDNA polyphyly and marked sympatry or parapatry of divergent mtDNA lineages in *Calligrapha* confound the geographic patterns required for meaningful phylogeographic assessment. However, these confused geographic patterns could be disentangled when each lineage was resolved, showing how intra- but also interspecific mtDNA lineages repeatedly colonized the same parts of the global range of the species, expanding their ranges from different areas of North America in a coherent manner (Figures 4-8).

Phylogeography of Calligrapha philadelphica and C. multipunctata in the B-clade

Bayesian phylogeographic analyses identified four main lineages (Posterior Probability, PP = 0.902-1.000) in the B-clade that shared a common ancestor some 0.75 Ma ago [0.48-1.09], in the transition from the Early to the Middle Pleistocene, in the later stages of the pre-Illinoian times (Figure 4). Two lineages, p_{WL} (i.e., B3-2) and p_{EL} (i.e., B3-4), are exclusively composed of *C. philadelphica* samples for which the founding event is placed in the Northern Appalachians during the Middle Pleistocene, at 0.46 Ma [0.26-0.68], in the period when the ice ages intensified their frequency and extreme alternation of conditions (Chalk et al. 2017). One of these lineages, p_{WL} , may have started to spread west along the southern Canadian Shield perhaps after the Illinoian glaciation (gray symbols in Figure 5b-d), during the interglacial period that lasted between 80-130 ka BP (Richmond & Fullerton, 1986). This lineage reached the northwestern interior plains in Alberta, east from the Rocky Mountains, before the Last Glacial Maximum (Figure 5d). This results suggests that in this period there must have been already at least two refugia at both sides of the Great Lakes area for this stock of *C. philadelphica*, since populations of these lineages originated before the LGM survived until the recent in both areas. The existence of an east-west genetic divide for organisms in North America is a general pattern and it has been usually linked to the disruptive effects of glaciations (e.g., references in Soltis et al., 2006; Colbeck et al., 2008). Starting some 0.27 Ma ago [0.11-0.44], the other *C. philadelphica* lineage (p_{EL}) diversified and expanded its range in the Northern Appalachian region reaching the Gaspé Peninsula to the northeast, southwestern Ontario to the west along the St. Lawrence lowlands, and the Blue Ridge Mts. along the Appalachians to the southwest (white symbols in Figure 5b-e). These range expansions toward coastal areas in the east (p_{EL1}) and particularly the latter toward the south (p_{EL2}) are coincident in time with the height of the Wisconsinian glaciation (12-35 ka BP) and consistent with the predictions of species range changes during the glaciations (Hewitt, 2000, 2004). However, in our phylogeographic reconstruction, they also require that descendant populations remained close to their ancestral ranges, since there is no inferred evidence for recolonization from the hypothetical eastern and southern refugia into the ancestral range (as proposed for other organisms; e.g., Peterson & Graves, 2016). Moreover, the p_{EL} lineage did not colonize, at least from a mtDNA perspective, the interior areas now occupied by descendants of the first group of *C. philadelphica*, indicating the possibility of some exclusive competition or demographic, leading-edge effects in this group (Hewitt, 2004).

580

581 The other two major lineages include most of the samples of *C. multipunctata*, with one
582 southern group, m_{ML} (i.e., B3-3, B3-5 and B3-6) associated to the Mississippi basin
583 (Figure 6), and a northern group, $m_{CL}+m_{NL}$ (i.e., B3-1) distributed from Eastern Canada
584 and New England to the Pacific Northwest (Figure 7). The signature from the nuclear
585 gene investigated here, namely the significant negative Tajima's D value and the star
586 shape of the allele genealogy, suggests a history of rapid range expansions for this
587 group after relatively recent bottlenecks. These two *C. multipunctata* lineages are
588 parapatric, sharing an area of sympatry in the Upper Mississippi and Missouri basins as
589 well as the southern plains of the central Canadian provinces, in their evolutionary and
590 geographic fronts of expansion. This area is of particular interest, since it roughly
591 corresponds to the area of expansion from the aforesaid Driftless region, a glacial refuge
592 of temperate fauna and flora (Rowe et al., 2004), which is revealed here as very
593 important for the evolution and maintenance of diversity of several lineages of *C.*
594 *multipunctata*. The oldest of the Mississippian branches (by virtue of its basal
595 paraphyly; Supporting Information Figure S3b), m_{ML1} , following perhaps the Missouri,
596 Snake and Columbia River basins and the intervening steppe environments of the
597 Wyoming Basin, colonized through the Rocky Mountains and across the continental
598 divide later than 0.19 Ma [0.05-0.34] and before 0.04 Ma [0.003-0.09]. This range
599 expansion may have happened in the context of northern migrations in interglacial
600 periods and without discarding the one after the Illinoian glaciation (gray symbols in
601 Figure 6c-e). This lineage established isolated populations that can be morphologically
602 recognized as *C. multipunctata s. str.*, and must have survived in a western refuge,
603 south of the area of confluence between the Laurentide and the Cordilleran Ice Sheets
604 through the last glacial cycle. Rivers, and particularly the Mississippi river in North
605 America, act as important barriers for terrestrial organisms (e.g., Soltis et al., 2006), but
606 it is important to note that rivers may actually favor the dispersal of this group of beetles
607 considering that their full life-cycle is associated to riparian trees, in this case of the
608 genus *Salix* (Brown, 1945). Another branch of this same lineage, m_{ML2} (i.e., B3-5, B3-6)
609 also spread northward upstream from the Mississippi Embayment to the Upper
610 Mississippi during the same period where it started to diversify some 0.07 Ma ago
611 [0.03-0.12], finally reaching the interior plains in southern Manitoba to the north, using
612 the Driftless region as a secondary dispersal center after the LGM less than 10 ka BP
613 (white symbols in Figure 6b-e). In the northward expansion from the Mississippi

Embayment, this lineage also colonized the Ohio River basin to the northeast at least in two independent colonization waves before the LGM, where descendants of both arrivals survived and dispersed.

The second, northern lineage of *C. multipunctata bigsbyana* (clade B3-1), currently distributed in previously glaciated regions during the LGM, may have had an origin in Eastern Canada, north of the St. Lawrence River and Gulf or perhaps in the region in the upper course of the Mississippi river some 0.29 Ma ago [0.15-0.46]. From this northern area, one evolutionary branch, m_{CL} (i.e., B2-2) colonized the central northern plains in the Late Pleistocene, some 0.10 Ma ago [0.03-0.19] (Figure 7a-e), most likely surviving the LGM again in the Driftless region. This lineage coexisted in this region up to the recent with descendants of other groups, including those discussed previously, reaching the area from the south and, as will be seen later, of an evolutionarily derived group of m_{NL} , coming from the east or the north. The recolonization of northernmost localities from the Driftless region right after the LGM could have followed the rapid advance of the pioneering Aspen Parkland biome, an ecological transition between grasslands and boreal forests and characterized by the presence of *Salix* (Strong & Hills, 2005). One of the evolutionary branches of the main m_{NL} lineage (Figure 7f-l), m_{NL1} (i.e., B2-1), in an early stage of its evolution, before 0.11 Ma [0.03-0.20], expanded its range west in several waves of colonization from the upper courses of the Mississippi and the Missouri rivers, crossing the Rocky Mountains to the west most likely through a northern route, north of the Cascade range, and reaching the Pacific coast of Washington and Oregon some 26 ka BP [4-55]. These populations were pushed to or survived in these coastal areas of the Pacific Northwest through the LGM, a well-established refuge for temperate species of plants and animals (e.g., Soltis, Gitzendanner, Streng & Soltis, 1997; Beatty & Provan, 2010; Roberts & Hamann, 2015). The same lineage also left descendants (m_{NL2}) in exactly the same areas associated to the Driftless region and with the same postglacial range expansion dynamics as the two groups of *C. multipunctata bigsbyana* discussed so far. Contemporarily, another evolutionary branch of this northern group (m_{NL3}) spread east across the St. Lawrence river basin and into the northern Appalachian region where it expanded its range north up to the Gaspé peninsula and Nova Scotia. This evolutionary lineage must have survived the LGM in one or several local refugia, since we found no evidence for recolonization of the area from southern populations, being southernmost

parts of the *C. multipunctata* range occupied by representatives of the southern, Mississippian lineage. Indeed, the existence of a northeastern coastal glacial refugium has been hypothesized for several organisms (e.g., references in Colbeck et al., 2008). In the final front of this expansion, in Maine and New Brunswick, this *C. multipunctata bigsbyana* lineage hybridized with *C. philadelphia*, leaving the signature of these events in several individuals of the latter with *cox1* haplotypes shared with or derived from *C. multipunctata* (gray symbols in Figure 7k). Judging from the diversity of mtDNA haplotypes involved, these interspecific hybridization events occurred several times and at different time-frameworks, the oldest as far back as 0.07 Ma [0.06-0.08]. However, most of them were fairly recent, consistent with secondary contacts in the range expansion after the LGM, coincident in the same region for both species involved. The *C. philadelphia* stock that was involved in these hybridization events was most likely the one currently distributed and expanding its range in the same geographic area, as also supported by the results of paired association tests of *Wg* genotypes for the three main mtDNA lineages of *C. philadelphia* (B3-1 vs B3-2: N = 53, chi-square = 23.3189, $P = 0.011$; B3-1 vs B3-3: N = 69, chi-square = 20.7078, $P = 0.446$).

Phylogeography of unisexual species

Knowing the phylogeographic origins of unisexual species requires understanding the phylogeography of the polytypic mtDNA U-clade, which contains the mtDNA-donor parental lineage of the interspecific hybridization event or events which eventually gave rise to these taxa. While the origin of the U-clade can be placed with confidence in the Pliocene (Montelongo & Gómez-Zurita, 2014), it was only captured by the species treated in this work in a much later evolutionary stage, at the turn from the Middle to the Late Pleistocene. Indeed, the representatives of the U-clade included in this study shared an ancestor about 0.13 Ma [0.06-0.22] and inferred on the eastern Great Lakes drainage basin (Figures 4, 8a and 8e). Thus, this particular lineage is much younger than the others discussed so far and in our study it includes two main, closely related sublineages, albeit with low posterior probabilities (PP = 0.50-0.78), one with *C. multipunctata bigsbyana* and *C. suturella*, m_{UL} (i.e., U1-2; Figure 8a-d), and one with *C. philadelphia* and *C. vicina*, p_{UL} (i.e., U1-1, U1-3, U1-4; Figure 8e-j). The timing and dynamics of their range expansions can be linked to the changes during the Late Pleistocene and the Wisconsinan glaciation.

From the interglacial Great Lakes distribution, the branch of *C. multipunctata* expanded its range westward, reaching southern Manitoba after the LGM, and eastward to the northern Appalachian region. From the latter region, it expanded to southern and eastern coastal areas in a period inferred around the LGM, some 14.2 ka BP [3.1-30.7] and reaching Nova Scotia some 4 ka BP [1-14]. The eastern or Northern Appalachian branch of this lineage of *C. multipunctata bigsbyana* hybridized with *C. ignota* (Montelongo & Gómez-Zurita, 2015) less than 13 ka BP [3-32] (oldest splitting event relating *C. multipunctata* and *C. suturella* haplotypes), giving rise to *C. suturella* (gray symbols in Figure 8c-d). The continuous phylogeographic inference, which also models geographic distribution besides genetic differences, identifies up to three different origins of *C. suturella*. While limited mtDNA diversity for these samples seems to be at odds with this interpretation, *Wg* data reveals up to three different genotypes for *C. suturella* and involving alleles and genotypes present or shared with different stocks of *C. multipunctata*, besides two different alleles related to *C. ignota*. Given the relatively high diversity of the nuclear background of otherwise clonal, thelytokous species, it is plausible to consider multiple origins for this species associated to population range changes at or after the end of the Wisconsinan glaciation.

In turn, the other sublineage, p_{UL} , starting some 0.07 Ma ago [0.03-0.12], in the interglacial before the Late Wisconsinan glaciation, expanded its range to the west reaching the upper course of the Upper Mississippi, to the east along the Saint Lawrence lowlands, and to the south along the Ohio river basin on the western slopes of the Appalachians. These three lineages survived the LGM and expanded their ranges locally after this event. The origin of *C. vicina* may be related to interspecific hybridization events of the latter with *C. rowena* (or a close relative; Montelongo & Gómez-Zurita, 2015) toward the end or after the LGM, some 13 ka BP [3-29] in the northern Alleghanian region. This unisexual species, for which we cannot rule out a single origin because of the genetic uniformity of all individuals studied, expanded its range northeast and southwest along the western slopes of the Appalachians after the end of the LGM (gray symbols in Figure 8h-j).

An unexpected finding for the model of evolution of unisexuality in Calligrapha

The last remarkable finding of this study, which had gone unnoticed so far, since we never achieved the sample size and geographic coverage used in this work, is that the 54 individuals of *C. multipunctata* and *C. philadelphica* with an mtDNA type belonging to the U-lineage turned out to be all females. This raises very clearly the possibility that these individuals, moreover with subtle traits that distinguish them morphologically from conspecific and sympatric bisexual populations with a different mtDNA type, are themselves thelytokous. If this were proven to be the case, these populations would carry already the factor for unisexuality that was transferred to the tetraploid species of demonstrated hybrid origin that are well known to reproduce unisexually (Robertson, 1966; Montelongo & Gómez-Zurita, 2015). This factor, which would still come from a single source by virtue of linkage with a monophyletic mtDNA group, may be nonetheless the key for the survival and establishment of hybrid lineages and yet an additional twist to understanding how this evolutionary transition occurred.

In our previous model for the evolution of thelytoky in *Calligrapha*, we expressed doubts about the sequence of events for the evolutionary transition (Montelongo & Gómez-Zurita, 2015). However, it appears that at least for the hybrid thelytokous species, some form of obligatory or facultative unisexuality would have predated the interspecific hybridization events that gave rise to these species. This still avoids the explanation of the origin of this altered reproductive strategy in *Calligrapha* or at least of highly distorted sex ratios in species such as *C. multipunctata bigsbyana* and *C. philadelphica* as shown in the current study. One possibility is that the shift to unisexuality linked to a particular mtDNA type would have been triggered by an intracellular symbiont able to exert reproductive manipulation, such as *Wolbachia*, *Rickettsia*, *Cardinium* or *Spiroplasma* (Engelstädter & Hurst, 2009). While the first is highly prevalent in *Calligrapha* and it may be a factor to be considered in the evolution of unisexuality in *Calligrapha* (Gómez-Zurita, submitted), a preliminary screening for other prokaryote genera in a taxonomically diverse sample of this genus, including the species represented in the current study proved unfruitful (data not shown). Another possibility is that thelytoky was achieved in *Calligrapha* through an intermediate facultative parthenogenetic stage as proposed for *C. scalaris* based on cytological data (Robertson, 1966). However, high genetic mtDNA differentiation, different nuclear backgrounds (as assessed based on *Wg*) and the existence of some morphological differences between reproductive types of each of those species despite their

coexistence (also found in *C. scalaris*; work in preparation) suggest that they are in fact entirely independent evolutionary lineages, each with their own reproductive biology. Of these independent lineages, only the ones already departing from sexual reproduction or with highly biased sex ratios (and with mtDNA of the U-type) would have been able to experience rare interspecific sex already in the Holocene, and establish the taxonomically recognizable unisexual *Calligrapha* (Montelongo & Gómez-Zurita, 2015).

Conclusions

The first mtDNA phylogenies for *Calligrapha* suggested that, owing to generalized species paraphyly of North American species for this genetic marker, it would be as interesting as challenging to use it for the analysis of recent evolutionary phenomena in this genus (Gómez-Zurita et al., 2004, 2006; Montelongo & Gómez-Zurita, 2015). This marker, and to some extent also a fast evolving single-copy nuclear protein encoding gene as shown in this study, exhibits high diversity and deep divergence. These characteristics, examined at large geographic scales and for lineages that have modified their ranges, coupled with episodes of genetic introgression, can potentially hinder the feasibility and reliability of phylogeographic research (Funk & Omland, 2003). However, we have shown here that when this diversity is partitioned in meaningful evolutionary units, e.g. based on monophyletic mtDNA assemblages with homogenous nuclear backgrounds, the patterns actually emerge and provide great detail to the understanding both of introgression, but also of spatial changes and the origins of sympatry and high local diversity. A partitioned analysis of transcontinental genetic data has shown that the distribution of selected species of *Calligrapha* has been altered during the Pleistocene and that it has been affected by climatic as well as physiographic changes in North America. The main engines for this process seem to include a dominant role of the Wisconsinan glaciation and possibly, to some extent, by the Illinoian glaciation too in terms of range changes, but also of rivers as postglacial highways for dispersal. In this regard, most species of *Calligrapha*, particularly the ones included in this study, are inextricably associated to plants living in moist habitats and waterways, which in turn have been shown to have had their postglacial distribution favored and also conditioned by river drainages (e.g., Carstens et al., 2013). We identified several genetic stocks, monophyletic mtDNA lineages of otherwise

paraphyletic *C. multipunctata* and *C. philadelphica*, and inferred range expansion dynamics, often bringing descendants of these, mostly allopatric lineages into sympatry and consistent with the survival of the species in several refugia during glacial maxima, very often not far from the southern edge of the Laurentide Ice Sheet. In fact, with the exception of the Mississippian lineage of *C. multipunctata*, the others examined here occupy ranges consistent with their distribution close to the glaciated areas and without evidence for recolonization from distant regions or exchange among inferred refugia. The main refugia identified, which were shared with a variety of temperate plants and animals as recognized in previous studies (Beatty & Provan, 2010), are the Pacific northwest coastal areas, maybe the Great Basin south of the northern Rocky Mountains, the Atlantic maritime regions of New England, Quebec and New Brunswick, and very predominantly the Driftless region in the confluence between current Wisconsin, Minnesota and Iowa. The latter would have offered refuge to a minimum of three *C. multipunctata bigsbyana* lineages, which all of them secondarily expanded their range in the area and in previously glaciated areas up north. Finally, it was possible to associate the hybrid origin of unisexual species in *Calligrapha*, which were known derived from populations carrying a particular mtDNA type, to a particular place, a particular time, a particular process and specific lineages of their parental species. The place would be an area west of the Northern Appalachians and east of the main Great Lakes lobe of the Laurentide Ice Sheet; the time would be in the Holocene, less than 12 ka BP and the end of the LGM; and the process would specifically require range expansions of local populations formerly in nearby refugia. Put together, this information offers unprecedented depth of phylogeographic understanding of the evolution of these species and solid starting hypotheses for specific studies focusing on the unisexual taxa. Among these, an important discovery of this study has been recognizing the hybrid unisexual species not only embedded in a genetically distinct lineage of their parental species, but the fact that this is a geographically coherent, female-only lineage, raising more questions than answers about the evolutionary causes of the reproductive transition in this system.

Acknowledgements. We are indebted to Tinguaro Montelongo (IBE, Barcelona), Dr. Daniel J. Funk (Vanderbilt University, Tennessee), and Dr. Daniel P. Duran (Drexel University, Philadelphia), as well as individuals who occasionally helped us over the years to put together part of the sampling that allowed this study. This research was

possible specifically thanks to funds made available by the Spanish Ministry of Economy (MINECO) through Grant No. CGL2014-52937-P, but it also benefited from samples and knowledge acquired thanks to previous projects funded by the same Organism, including Grant Nos. CGL2011-23820 and CGL2008-00007, the latter with contribution from the European Union Regional Development Fund (ERDF). Dr. Alexander Blanke and three anonymous referees kindly devoted some of their time to revise our work and help us to improve it in several respects, and we are grateful for that.

References

- Avise, J. C. (2000). *Phylogeography. The history and formation of species*. Cambridge, MA: Harvard University Press.
- Avise, J. C. (2008). *Clonality. The genetics, ecology, and evolution of sexual abstinence in vertebrate animals*. New York: Oxford University Press.
- Beatty, G. E., & Provan, J. (2010). Refugial persistence and postglacial recolonization of North America by the cold-tolerant herbaceous plant *Orthilia secunda*. *Molecular Ecology*, 19, 5009–5021.
- Beatty, G. E., & Provan, J. (2011). Phylogeographic analysis of North American populations of the parasitic herbaceous plant *Monotropa hypopitys* L. reveals a complex history of range expansion from multiple late glacial refugia. *Journal of Biogeography*, 38, 1585–1599.
- Bell, G. (1982). *The Masterpiece of Nature: The evolution and genetics of sexuality*. Berkeley, CA: University of California Press.
- Bielejec, F., Baele, G., Vrancken, B., Suchard, M. A., Rambaut, A., & Lemey, P. (2016). Spread3: interactive visualization of spatiotemporal history and trait evolutionary processes. *Molecular Biology and Evolution*, 33, 2167–2169.
- Brown, W. J. (1945). Food-plants and distribution of the species of *Calligrapha* in Canada, with descriptions of new species (Coleoptera, Chrysomelidae). *Canadian Entomologist*, 77, 117–133.
- Buckley, T. R., Marske, K. A., & Attanayake, D. (2009). Identifying glacial refugia in a geographic parthenogen using palaeoclimate modelling and phylogeography: the New Zealand stick insect *Argosarchus horridus* (White). *Molecular Ecology*, 18, 4650–4663.
- Carstens, B. C., Brennan, R. S., Chua, V., Duffie, C. V., Harvey, M. G., Koch, R. A., McMahan, C. D., Nelson, B. J., Newman, C. E., Satler, J. D., Seeholzer, G., Posbic, K., Tank, D. C., & Sullivan, J. (2013). Model selection as a tool for phylogeographic inference: an example from the willow *Salix melanopsis*. *Molecular Ecology*, 22, 4014–4028.
- Chalk, T. B., Hain, M. P., Foster, G. L., Rohling, E. J., Sexton, P. F., Badger, M. P. S., Cherry, S. G., Hasenfratz, A. P., Haug, G. H., Jaccard, S. L., Martínez-García, A., Pälike, H., Pancost, R. D., & Wilson, P. A. (2017). Causes of ice age intensification across the Mid-Pleistocene Transition. *Proceedings of the National Academy of*

861 *Sciences USA*, 114, 13114–13119.

862 Clark, J. S., Fastie, C., Hurtt, G., Jackson, S. T., Johnson, C., King, G. A., Lewis, M.,
863 Lynch, J., Pacala, S., Prentice, C., Schupp, E. W., Webb III, T., & Wyckoff, P.
864 (1998). Reid's paradox of rapid plant migration: dispersal theory and interpretation of
865 paleoecological records. *BioScience*, 48, 13–24.

866 Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: a computer program to
867 estimate gene genealogies. *Molecular Ecology*, 9, 1657–1660.

868 Colbeck, G. J., Gibbs, H. L., Marra, P. P., Hobson, K., & Webster, M. S. (2008).
869 Phylogeography of a widespread North American migratory songbird (*Setophaga*
870 *ruticilla*). *Journal of Heredity*, 99, 453–463.

871 Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more
872 models, new heuristics and parallel computing. *Nature Methods*, 9, 772.

873 Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed
874 phylogenetics and dating with confidence. *PLoS Biology*, 4, e88.

875 Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian
876 phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*,
877 29, 1969–1973.

878 Engelstädter, J., & Hurst, G. D. D. (2009). The ecology and evolution of microbes that
879 manipulate host reproduction. *Annual Review of Ecology, Evolution and Systematics*,
880 40, 127–149.

881 Facon, B., Pointier, J. P., Glaubrecht, M., Poux, C., Jarne, P., & David, P. (2003). A
882 molecular phylogeography approach to biological invasions of the New World by
883 parthenogenetic Thiarid snails. *Molecular Ecology*, 12, 3027–3039.

884 Freitas, S., Rocha, S., Campos, J., Ahmadzadeh, F., Corti, C., Sillero, N., Ilgaz, Ç,
885 Kumlutaş, Y., Arakelyan, M., Harris, D. James, & Carretero, M. A. (2016).
886 Parthenogenesis through the ice ages: a biogeographic analysis of Caucasian rock
887 lizards (genus *Darevskia*). *Molecular Phylogenetics and Evolution*, 102, 117–127.

888 Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: frequency,
889 causes, and consequences, with insights from animal mitochondrial DNA. *Annual*
890 *Review of Ecology, Evolution and Systematics*, 34, 397–423.

891 Gómez-Zurita, J. (2005). New distribution records and biogeography of *Calligrapha*
892 species (leaf beetles), in North America (Coleoptera: Chrysomelidae,
893 Chrysomelinae). *Canadian Field-Naturalist*, 119, 88–100.

894 Gómez-Zurita, J. (2015). Systematic revision of the genus *Calligrapha* Chevrolat

895 (Coleoptera: Chrysomelidae, Chrysomelinae) in Central America: The group of
896 *Calligrapha argus* Stål. *Zootaxa*, 3922, 1–71.

897 Gómez-Zurita, J., Vogler, A. P., & Funk, D. J. (2004). Diagnosing an overlooked North
898 American taxon: biological observations and mitochondrial insights on *Calligrapha*
899 *suturella* Schaeffer, 1933 new status (Coleoptera, Chrysomelidae). *Annals of the*
900 *Entomological Society of America*, 97, 28–36.

901 Gómez-Zurita, J., Funk, D. J., & Vogler, A. P. (2006). The evolution of unisexuality in
902 *Calligrapha* leaf beetles: Molecular and ecological insights on multiple origins via
903 interspecific hybridization. *Evolution*, 60, 328–347.

904 Grismer, J. L., Bauer, A. M., Grismer, L. L., Thirakhupt, K., Aowphol, A., Oaks, J. R.,
905 Wood Jr., P. L., Onn, C. K., Thy, N., Cota, M., & Jackman, T. (2014). Multiple
906 origins of parthenogenesis, and revised species phylogeny for the Southeast Asian
907 butterfly lizards, *Leiolepis*. *Biological Journal of the Linnean Society*, 113, 1080–
908 1093.

909 Hewitt, G. M. (1993). Postglacial distribution and species substructure: lessons from
910 pollen, insect and hybrid zones. In D. R. Lees and D. Edwards (Eds.), *Evolutionary*
911 *patterns and processes* (pp. 97–123). London: Linnean Society / Academic Press.

912 Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907–
913 913.

914 Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary.
915 *Philosophical Transactions of the Royal Society of London B*, 359, 183–195.

916 Ho, S. Y. W., & Larson, G. (2006). Molecular clocks: when times are a-changin'.
917 *Trends in Genetics*, 22, 79–83.

918 Jensen, L. H., Enghoff, H., Frydenberg, J., & Parker, E. D. (2002). Genetic diversity
919 and the phylogeography of parthenogenesis: comparing bisexual and thelytokous
920 populations of *Nemasoma varicorne* (Diplopoda: Nemasomatidae) in Denmark.
921 *Hereditas*, 136, 184–194.

922 Kearney, M. (2005). Hybridization, glaciation and geographical parthenogenesis.
923 *Trends in Ecology and Evolution*, 20, 495–502.

924 Kearney, M., & Blacket, M. J. (2008). The evolution of sexual and parthenogenetic
925 *Warramaba*: a window onto Plio-Pleistocene diversification processes in an arid
926 biome. *Molecular Ecology*, 17, 5257–5275.

927 Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and their Applications*,
928 13, 235–248.

929 Law, J. H., & Crespi, B. J. (2002). The evolution of geographic parthenogenesis in
930 *Timema* walking-sticks. *Molecular Ecology*, 11, 1471–1489.

931 Leigh, J. W., & D. Bryant. (2015). PopART: full-feature software for haplotype
932 network construction. *Methods in Ecology and Evolution*, 6, 1110–1116.

933 Lemey, P., Rambaut, A., Welch, J. J., & Suchard, M. A. (2010). Phylogeography takes
934 a relaxed random walk in continuous space and time. *Molecular Biology and*
935 *Evolution*, 27, 1877–1885.

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

938 Margold, M., Stokes, C. R., & Clark, C. D. (2018). Reconciling records of ice streaming
939 and ice margin retreat to produce a palaeogeographic reconstruction of the
940 deglaciation of the Laurentide Ice Sheet. *Quaternary Science Reviews*, 189, 1–30.

941 Mickelson, D. M., & Colgan, P. M. (2003). The southern Laurentide Ice Sheet.
942 *Developments in Quaternary Science*, 1, 1–16.

943 Montelongo, T., & J. Gómez-Zurita, J. (2014). Multilocus molecular systematics and
944 evolution in time and space of *Calligrapha* (Coleoptera: Chrysomelidae,
945 Chrysomelinae). *Zoologica Scripta*, 43, 605–628.

946 Montelongo, T., & Gómez-Zurita, J. (2015). Nonrandom patterns of genetic admixture
947 expose the complex historical hybrid origin of unisexual leaf beetle species in the
948 genus *Calligrapha*. *American Naturalist*, 185, 113–134.

949 Peterson, B. J., & Graves, W. R. (2016). Chloroplast phylogeography of *Dirca palustris*
950 L. indicates populations near the glacial boundary at the Last Glacial Maximum in
951 eastern North America. *Journal of Biogeography*, 43, 314–327.

952 Pongratz, N., Storhas, M., Carranza, S., & Michiels, N. K. (2003). Phylogeography of
953 competing sexual and parthenogenetic forms of a freshwater flatworm: patterns and
954 explanations. *BMC Evolutionary Biology*, 3, 23.

955 Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). *Tracer v1.6*.
956 Retrieved from <http://beast.bio.ed.ac.uk/Tracer>

957 Richmond, G. M., & Fullerton, D. S. (1986). Summation of Quaternary glaciations in
958 the United States of America. *Quaternary Science Reviews*, 5, 183–196.

959 Roberts, D. R., & Hamann, A. (2015). Glacial refugia and modern genetic diversity of
960 22 western North American tree species. *Proceedings of the Royal Society of London*
961 *B*, 282, 20142903.

962 Robertson, J. G. (1966). The chromosomes of bisexual and parthenogenetic species of

963 *Calligrapha* (Coleoptera: Chrysomelidae) with notes on sex ratio, abundance and
 964 egg number. *Canadian Journal of Genetics and Cytology*, 8, 695–732.

965 Rowe, K. C., Heske, E. J., Brown, P. W., & Paige, K. N. (2004). Surviving the ice:
 966 northern refugia and postglacial colonization. *Proceedings of the National Academy*
 967 *of Sciences USA*, 101, 10355–10359.

968 Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Flook, P. (1994). Evolution,
 969 weighting, and phylogenetic utility of mitochondrial gene sequences and a
 970 compilation of conserved polymerase chain reaction primers. *Annals of the*
 971 *Entomological Society of America*, 87, 651–701.

972 Snell, R. S., & Cowling, S. A. (2015). Consideration of dispersal processes and northern
 973 refugia can improve our understanding of past plant migration rates in North
 974 America. *Journal of Biogeography*, 42, 1677–1688.

975 Soltis, D. E., Gitzendanner, M. A., Streng, D. D., & Soltis, P. S. (1997). Chloroplast
 976 DNA intraspecific phylogeography of plants from the Pacific Northwest of North
 977 America. *Plant Systematics and Evolution*, 206, 353–373.

978 Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006).
 979 Comparative phylogeography of unglaciated eastern North America. *Molecular*
 980 *Ecology*, 15, 4261–4293.

981 Stephens, M., Smith, N., & Donnelly, P. (2001). A new statistical method for haplotype
 982 reconstruction from population data. *American Journal of Human Genetics*, 68, 978–
 983 989.

984 Strasburg, J. L., Kearney, M., Moritz, C., & Templeton, A. R. (2007). Combining
 985 phylogeography with distribution modeling: multiple Pleistocene range expansions
 986 in a parthenogenetic gecko from the Australian arid zone. *PLoS ONE*, 8, e760.

987 Strong, W. L., & Hills, L. V. (2005). Late-glacial and Holocene palaeovegetation zonal
 988 reconstruction for central and north-central North America. *Journal of Biogeography*,
 989 32, 1043–1062.

990 Suomalainen, E., Saura, A. & Lokki, J. (1987). *Cytology and evolution in*
 991 *parthenogenesis*. Boca Raton, FL: CRC Press.

992 Swenson, N. G., & Howard, D. J. (2005). Clustering of contact zones, hybrid zones, and
 993 phylogeographic breaks in North America. *American Naturalist*, 166, 581–591.

994 Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by
 995 DNA polymorphism. *Genetics*, 123, 585–595.

996 Templeton, A. R., & Sing, C. F. (1993). A cladistic analysis of phenotypic associations

997 with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses
 998 with cladogram uncertainty and recombination. *Genetics*, 134, 659–669.
 999 Templeton, A. R., Boerwinkle, E., & Sing, C. F. (1987). A cladistic analysis of
 1000 phenotypic associations with haplotypes inferred from restriction endonuclease
 1001 mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in
 1002 *Drosophila*. *Genetics*, 117, 343–351.
 1003 Weider, L. J., Hobaek, A., Hebert, P. D. N., & Crease, T. J. (1999). Holarctic
 1004 phylogeography of an asexual species complex. II. Allozymic variation and clonal
 1005 structure in Arctic *Daphnia*. *Molecular Ecology*, 8, 1–13.
 1006 Weiss, N. A. (2015). *wPerm: Permutation Tests*. R package version 1.0.1. Retrieved
 1007 from <https://CRAN.R-project.org/package=wPerm>
 1008 Wild, A. L., & Maddison, D. R. (2008). Evaluating nuclear protein-coding genes for
 1009 phylogenetic utility in beetles. *Molecular Phylogenetics and Evolution*, 48, 877–
 1010 891.
 1011
 1012

Figure captions

Figure 1. Statistical parsimony unrooted X-trees of *coxI* B (B1-B73) and U (U1-U13) haplotypes characterized from three species and two subspecies of *Calligrapha*, identified by shade intensity according to legend. Hypothetical placements of the root on both trees are signaled with an arrow, as the likely anchoring points of the branch connecting them. Haplotypes are grouped in increasingly larger groups until they encompass the whole tree according to the nesting algorithm of Templeton et al. (1987).

Figure 2. Frequencies of *Wg* alleles (a), homozygotes (b) and heterozygotes (c) in the studied sample of *Calligrapha*. The alleles found in homozygosis are highlighted with an arrowhead. Allele and heterozygotes observed twice (doubletons) or once (singletons) are binned into these respective categories, but identified in the accompanying legend. m-alleles: *C. multipunctata s. l.* and *C. suturella*; p-alleles: *C. philadelphica* and *C. vicina*.

Figure 3. Statistical parsimony network of *Wg* alleles in *Calligrapha*, where m-alleles are found in *C. multipunctata s. l.* and *C. suturella*, and p-alleles in *C. philadelphica* and *C. vicina*. The size of each node is proportional to the number of observations in the sample, and alleles highlighted in bold were found in homozygosis and all the others in heterozygotic individuals, usually with one of the former. The frequency of each *coxI* type, represented here at their first nesting level as shown in Figure 1, is shown for each *Wg* allele. Color-coded mtDNA groups are shown separately for each taxon and the order in which colors are used in the pie charts always start at 3 o'clock and progress counterclockwise in the same order as they appear in the legend.

Figure 4. Clock constrained Bayesian tree resulting from the continuous phylogeographic analysis and simplified to show the main lineages recovered and their posterior probabilities. For each lineage, coincident with hierarchical groups established based on the statistical parsimony X-tree of Figure 1, the encompassed species and the number of individuals from large administrative units (as an indication of geographical range) are given. The detailed topologies of every group as well as the phylogeographic inferences for their dispersal are shown in Supporting Information Figure S3 and Figures 5-8.

Figure 5. Continuous mtDNA phylogeography with inferred geographic dispersal of *Calligrapha philadelphia*. The maps show the history of lineages p_{WL} (gray circles) and p_{EL} (white circles) of *C. philadelphia*. The output of Spread3 also yields information on the proportion of lineages that occupy a given area, shown as a colored polygons proportional to the square root of the number of MCC branches in a given area. The approximate time period is given according to the dated tree topologies shown in Supporting Information Figure S3.

Figure 6. Continuous mtDNA phylogeography with inferred geographic dispersal of the Mississippian lineage of *Calligrapha multipunctata*. The maps show the history of the m_{ML} lineage (a-f) of *C. multipunctata*, including the origin of *C. multipunctata s. str.* (gray circles) in B3-3. Other details as in the caption to Figure 5.

Figure 7. Continuous mtDNA phylogeography with inferred geographic dispersal of the northern lineages of *Calligrapha multipunctata*. The maps show the history of lineage m_{CL} (a-e) and lineage m_{NL} (f-l) of *C. multipunctata bigsbyana*, including mtDNA introgression to *C. philadelphia* (gray circles) in the latter. Other details as in the caption to Figure 5.

Figure 8. Continuous mtDNA phylogeography with inferred geographic dispersal of the unisexual samples of *Calligrapha*. The maps show the history of lineage m_{UL} of *C. m. bigsbyana* (a-d) and lineage p_{UL} of *C. philadelphia* (e-j), including the origins of *C. suturella* and *C. vicina* (gray circles), respectively. Other details as in caption to Figure 5.

Supporting Information:

Supporting Information Figure S1. Geographic sources of *Calligrapha* samples. (a) Sampling localities of *C. multipunctata bigsbyana* (Kirby) (white circles) and of *C. multipunctata s. str.* (gray circles); (b) sampling localities of *C. philadelphia* (L.); and (c) sampling localities of *C. suturella* Schaeffer (white squares) and of *C. vicina* Schaeffer (black squares). Detailed data for these localities are shown in Appendix 1.

1080 **Supporting Information Figure S2.** Geographic distribution of inferred *Wg* alleles for
1081 the sample of *Calligrapha multipunctata* (a-l), *C. philadelphia* (m-r), *C. suturella* (a, d,
1082 j) and *C. vicina* (n, q).

1083

1084 **Supporting Information Figure S3.** Inferred trees used for continuous mtDNA
1085 phylogeographic analyses of *Calligrapha philadelphia* ($p_{EL}+p_{WL}$ lineages) and *C.*
1086 *multipunctata* (m_{ML} and $m_{CL}+m_{NL}$ lineages) in the B-clade, and of these species and *C.*
1087 *suturella* and *C. vicina* in the U-clade ($m_{UL}+p_{UL}$) as defined in Figure 4.

1088 Phylogeographic inferences for each of these clades are summarized in Figures 5-8.

1089

1089 **Table 1.** Genetic diversity in the *coxI* sample of *Calligrapha*. Taxonomic and
1090 phylogenetic subgroups of data are presented according to the number of specimens
1091 considered (N), number of segregating (S) and informative (Inf.) sites, number of
1092 haplotypes (h), haplotype (Hd) and nucleotide (Pi) diversities, the latter with an estimate
1093 of standard deviation (SD), divergence (Theta), and the results of neutrality test based
1094 on Tajima's D.
1095

Group	N	S	Inf.	h	Hd	Pi (SD)	Theta ^a	Tajima's D ^b
<i>C. multipunctata</i>	344	71	54	52	0.877	0.00898 ± 0.00077	0.01604	-1.28887 ^{ns}
<i>C. multipunctata bigsbyana</i>	322	68	51	48	0.860	0.00911 ± 0.00083	0.01558	-0.48375 ^{ns}
<i>C. multipunctata multipunctata</i>	22	4	3	5	0.699	0.00132 ± 0.00023	0.00147	-0.27418 ^{ns}
<i>C. philadelphica</i>	146	58	51	34	0.930	0.02224 ± 0.00103	0.01504	1.48430 ^{ns}
<i>C. suturella</i>	7	1	0	2	0.250	0.00034 ± 0.00025	0.00053	-1.05482 ^{ns}
<i>C. vicina</i>	7	0	0	1	0	0	0	NA
B-haplotypes	437	69	46	73	0.911	0.00782 ± 0.00024	0.01526	-1.40362 ^{ns}
U-haplotypes	67	11	9	13	0.835	0.00206 ± 0.00018	0.00314	-0.94677 ^{ns}
Bisexual species	490	92	71	85	0.927	0.01502 ± 0.00082	0.01986	-0.70731 ^{ns}
Unisexual species	14	3	2	3	0.600	0.00165 ± 0.00020	0.00127	0.89425 ^{ns}
All	504	92	71	86	0.931	0.01647 ± 0.00082	0.02034	-0.48375 ^{ns}

1096 ^aEstimated per site and based on the total number of mutations.

1097 ^bUsing total number of mutations.

1098

1099

1099 **Table 2.** Genetic diversity in the *Wg* sample of *Calligrapha*. Taxonomic and
1100 phylogenetic subgroups of data are presented according to the diploid number of gene
1101 copies considered (N), number of variable sites (S), number of alleles (h), gene (=
1102 heterozygosity; Hd) and nucleotide (Pi) diversities, the latter with an estimate of
1103 standard deviation (SD), divergence (Theta), and the results of neutrality test based on
1104 Tajima's D.
1105

Group	N	S	h ^a	Hd	Pi (SD)	Theta ^b	Tajima's D ^c
<i>C. multipunctata</i>	630	54	62 [58]	0.699	0.00448 ± 0.00029	0.01750	-2.05420*
<i>C. multipunctata bigsbyana</i>	586	54	62 [58]	0.679	0.00446 ± 0.00031	0.01768	-2.07444*
<i>C. multipunctata multipunctata</i>	44	4	6 [5]	0.808	0.00355 ± 0.00019	0.00195	1.92207 ^{ns}
<i>C. philadelphica</i>	286	23	12 [11]	0.810	0.00656 ± 0.00033	0.00816	-0.52013 ^{ns}
<i>C. suturella</i>	14	12	5 [5]	0.725	0.01227 ± 0.00117	0.00799	2.12960*
<i>C. vicina</i>	14	8	2 [3]	0.538	0.00913 ± 0.00089	0.00533	2.68630**
B-haplotypes	814	64	71 [69]	0.818	0.00970 ± 0.00028	0.02096	-1.48407 ^{ns}
U-haplotypes	130	24	14 [14]	0.897	0.01294 ± 0.00061	0.00935	1.10524 ^{ns}
Bisexual species	916	66	74 [70]	0.839	0.01082 ± 0.00025	0.02120	-1.34684 ^{ns}
Unisexual species	28	20	7 [8]	0.823	0.01582 ± 0.00093	0.01089	1.60031 ^{ns}
All	944	66	75 [72]	0.843	0.01110 ± 0.00024	0.02111	-1.30188 ^{ns}

1106 number of alleles inferred from manual assignment is given in square brackets.

1107 estimated per site and based on the total number of mutations.

1108 ng total number of mutations (ns: non-significant; *P<0.05; **: P<0.01).

1109

1110

Table 3. Results of permutation contingency tests between categorical variables defined by the *coxI* nested groups within each higher-level nesting category shown in Figure 1 and *Wg* diversity as defined by alleles (A) and genotypes (G). Only nesting categories with any of the tests producing significant results are shown.

Clade	N ^a	<i>Wg</i> genotypes	<i>Wg</i> alleles
B1-1	100	399.7283ns	518.3126**
B1-3	38	37.0194ns	56.8479***
B1-11	18	14.9625***	21.7684***
B1-12	18	9.5625ns	17.2446*
B1-20	9	12.6000ns	21.1500*
B2-1	196	593.6011**	654.5214***
B2-3	36	30.6667***	45.5652***
B2-7	37	49.3095**	40.3604***
B3-3	28	40.9748*	31.1860**
B3-4	52	17.7940ns	19.6888**
B4-1	244	192.5845***	341.4076***
B4-2	80	80.0000***	160.0000***
B5-1	408	352.8875***	422.4216***
U1-1	33	89.9583***	29.0042ns
U2-1	65	145.5421***	231.8475***
Total	472	361.4377***	396.3712***

^aIn the contrasts considering alleles, the corresponding sample size is doubled.

Significance: ns, non-significant; * $0.05 \geq P > 0.01$; ** $0.01 \geq P > 0.001$; *** $P \leq 0.001$.

B4-1

B4-3

B5-1

* Eight out of 105 specimens with haplotype B1 are *C. philadelphia* from New Brunswick
 ** One out of 13 specimens with haplotype B33 is a *C. multipunctata bigsbyana* from Louisiana

- *Calligrapha philadelphia*
- *Calligrapha multipunctata bigsbyana*
- *Calligrapha multipunctata multipunctata*
- ◐ *Calligrapha suturella*
- ◑ *Calligrapha vicina*

↖ hypothetical placement of the root

*Haplotype 31, originally in the B3-3 group, in this analysis joins all other haplotypes in the B3-1 group