# First microsatellite markers for the pine catkin sawfly *Xyela concava* (Hymenoptera, Xyelidae) and their application in phylogeography and population genetics

Dustin Kulanek[1], Stephan M. Blank[1], Katja Kramp[1]

[1]Senckenberg Deutsches Entomologisches Institut, Müncheberg, Germany

Corresponding Author:
Dustin Kulanek[1]
Eberswalder Str. 90, 15374 Müncheberg, Germany
Email address: Dustin.Kulanek@senckenberg.de

## Abstract

16  Microsatellites are widely used as powerful markers in population genetics because of their
17  ability to access recent genetic variation and to resolve subtle population genetic structures.
18  However, their development, especially for non-model organisms with no available genome-
19  wide sequence data, has been difficult and time-consuming. Here, a commercial high-throughput
20  sequencing approach (HTS) was used for the very first identification of microsatellite motifs in
21  the genome of *Xyela concava* and the design of primer pairs flanking those motifs. Sixteen of
22  those primer pairs were selected and implemented successfully to answer questions on the
23  phylogeography and population genetics of *X. concava*. The markers were characterized in three
24  geographically distinct populations of *X. concava* and tested for cross-species amplification in
25  two additional *Xyela* and one *Pleroneura* species (Xyelidae). All markers showed substantial
26  polymorphism as well as revealing subtle genetic structures among the three genotyped
27  populations. We also analyzed a fragment of the nuclear gene region of sodium/potassium-
28  transporting ATPase subunit alpha (*NaK*) and a mitochondrial gene region partly coding for
29  cytochrome oxidase subunit I (*COI*) to demonstrate different genetic resolutions and sex-biased
30  patterns of these markers, and their potential for combined use in future studies on the
31  phylogeography and population genetics of *X. concava*. Although a limited number of
32  populations was analyzed, we nevertheless obtained new insights on the latter two topics. The
33  microsatellites revealed a generally high gene flow between the populations, but also suggested a
34  deep historical segregation into two genetic lineages. This deep genetic segregation was
35  confirmed by *NaK*. While the high gene flow was unexpected, because of assumed restricted
36  dispersal ability of *X. concava* and the discontinuous distribution of the host trees between the
37  populations, the segregation of two lineages is comprehensible and could be explained by
38  different refugial areas of the hosts during glacial times. The *COI* results showed a discordant
39  strong genetic structure between all populations, which might be explained by the smaller
40  effective population size of the mitochondrial genome. However, given the frequent evidence of
41  a similar nature in recent studies on sawflies, we also consider and discuss mitochondrial
42  introgression on population level as an alternative explanation.

43

## Introduction

44

45  Xyelidae have always attracted the attention of taxonomists and systematists. They represent the
46  sister group of the rest of the megadiverse insect order Hymenoptera (Ronquist et al., 2012;
47  Klopfstein et al., 2013; Malm & Nyman, 2015), which is traditionally divided into the
48  paraphyletic "Symphyta" (missing the wasp waist) and the monophyletic Apocrita (sharing a
49  wasp waist as a derived feature ) (Malm & Nyman, 2015). The recent inconsistent phylogenetic
50  placement of xyelids together with Pamphiliidae and Tenthredinidae as sister clade to all
51  remaining Hymenoptera by Peters et al. (2017) might have been caused by an artificial grouping
52  due to shared very slow mutation rates in those groups (Ronquist et al., 2012).
53  The rich fossil record of Xyelidae includes the earliest fossil forms of Hymenoptera dating from
54  the Middle–Upper Triassic (Kopylov, 2014). Proper knowledge of the phylogeography and

55  population genetics of xyelids is therefore important in understanding the underlying
56  evolutionary processes, which in turn will help to understand the evolution of other
57  hymenopterans. Unfortunately, such data are scarce for xyelids due to the rarity of many species,
58  ephemerality of the imagines, and considerable problems in identifying species morphologically
59  as well as genetically (e.g., Burdick 1961; Blank, Shinohara & Byun, 2005; Blank, Shinohara &
60  Altenhofer, 2013; Blank et al., 2017; Blank & Kramp 2017). While a limited number of
61  microsatellite studies has been conducted on sawflies (Hartel, Frederick & Shanower, 2003;
62  Cook et al., 2011; Caron et al., 2013; Bittner et al., 2017), none has focused on xyelids.
63  Consequently, little is known about the population dynamics of this early-diverging group,
64  including effects of ephemerality of imagines and their dispersal ability, host adaption and host
65  dependency, and reproduction mode.
66  Here, we report on the first developed and tested set of 16 polymorphic nuclear microsatellite
67  markers for *Xyela concava* Burdick, 1961, to shed light on the latter issues. *X. concava* is widely
68  distributed in southwestern USA, where it is closely associated with the pinyon-juniper
69  woodland vegetation type of higher elevation semideserts, i.e., with pine species of the subgenus
70  *Strobus* subsection *Cembroides* (Farjon 2010). Females oviposit into developing male cones of
71  *Pinus cembroides* Zuccarini, 1830, *P. edulis* Engelmann, 1848 and *P. monophylla* Torrey &
72  Frémont, 1845 (Fig. 1), where the larvae feed on the sporophylls. After having ceased feeding,
73  *Xyela* larvae dig into the soil below the host trees and may diapause up to several years before
74  pupating (Blank, Shinohara & Altenhofer, 2013). Imagines of the next generation emerge during
75  spring and often visit flowering plants with easily accessible anthers, such as mountain
76  mahogany (*Cercocarpus* spp.) and cliff-rose (*Purshia* spp.), from which they gather pollen for
77  nutrition with their adapted mouthparts (Burdick 1961; Blank, personal observation). Flight
78  behavior is described as erratic and slow (Burdick 1961). Therefore, given an assumed restricted
79  dispersal ability and a close association with particular host species, it is intriguing to see how
80  the variation within and among populations have been influenced by the distribution of the host
81  trees during glacial and postglacial times. The high resolution power and therefore high
82  capability of microsatellite markers to assess subtle and recent population genetic structures
83  makes them well suited to this task. We used a commercial high-throughput sequencing
84  approach for the development of the microsatellite markers and applied them to describe genetic
85  structures and variation among and within three geographically distinct populations of *X.*
86  *concava*. Furthermore, we compared the resolution of genetic variation of these markers with
87  compiled data for one nuclear and one mitochondrial gene coding region and discuss their
88  possible combined suitability for identifying genealogical lineages and answering
89  phylogeographical questions. Finally, cross-amplification patterns for two species of *Xyela* and
90  one of *Pleroneura*, sister taxon of *Xyela* (Smith 1967), are illustrated.

91

## Material and methods

92

93

### Sampling

94

95   *Xyela* larvae were extracted from staminate cones of pines as described by Blank, Shinohara &
96   Altenhofer (2013) and stored in 100 % ethanol at -20 °C. We included in the analysis larvae
97   originating from three collection sites which are located 900–1,200 km from each other (see
98   Table S1). The specimens are preserved in the Senckenberg Deutsches Entomologisches Institut,
99   Müncheberg, Germany. Since it is impossible to identify *Xyela* larvae to species level
100  morphologically, they were *COI* barcoded and identified by comparison with sequences from
101  imagines identified as *X. concava* morphologically (identification following Burdick (1961),
102  reference sequences of imagines were published by Blank, Kramp & Shinohara (2017) and are
103  deposited in the GenBank (NCBI) database, accession numbers KY198313 and KY198314).
104  Finally, 98 larvae of *X. concava* were selected for the analysis (for detailed data see Table S1).

105

### DNA extraction

106

107  Whole larvae were used for DNA extraction. The integument was slightly cut with a scalpel, so
108  that the exterior stayed intact for later morphological inspection. DNA was extracted and purified
109  with E.Z.N.A. Tissue DNA Kit (Omega Bio-Tek) according to the manufacturer's protocol, but
110  with an extended 2 hour incubation time at 55 °C (Thermomixer, without shaking) for cell lysis.
111  The extracted DNA was stored at -20 °C until later use. The integuments were retained and
112  stored in 70 % ethanol.

113

### Microsatellite marker development and screening

114

115  Total genomic DNA of a single female of *X. concava* (specimen ID: DEI-GISHym 30887, see
116  Table S1) was extracted following the protocol described above. 10 ng/µl DNA in a total volume
117  of 20 µl was sent to AllGenetics & Biology (Coruña, Spain) for the commercial identification of
118  microsatellite motifs and the design of motif flanking primer pairs. A library was prepared for
119  the DNA sample using the Nextera XTDNA kit (Illumina), following the manufacturer's
120  instructions. The library was enriched with the following microsatellite motifs: AC, AG, ACG,
121  and ATCT. Enriched DNA was sequenced in the Illumina MiSeq platform (PE300) and
122  produced 3,043,190 paired-end reads. These paired-end reads were processed in Geneious 10.0.5
123  (Biomatters, Ltd.) using in-house developed scripts (property of AllGenetics & Biology) and
124  overlapped into 1,521,595 sequences (trim error probability limit of 0.03). Primer design was
125  carried out by AllGenetics & Biology in Primer 3 (Koressaar & Remm 2007; Untergrasser et al.,
126  2012) for 500 sequences containing microsatellite motifs. For a preliminary screening, fifty
127  primer pairs were picked and four *X. concava* larvae (DEI-GISHym 32824–32827) were used for
128  tests of polymorphism. Furthermore, 12 specimens of *X. deserti* Burdick, 1961, 12 specimens of
129  an undescribed *Xyela* species, possibly a member of the *X. alpigena* group (Blank & Kramp
130  2017), and six specimens of *Pleroneura koebelei* Rohwer, 1910 (see Table S1) were tested for
131  cross-species amplification to check the marker system for potential use on two closely and one

132   more distantly related xyelid species. The PCR analysis included a temperature gradient in the
133   primer annealing step to find the best conditions for each primer pair. PCR was carried out in a
134   total volume of 5 μl containing 0.5 μl DNA, 0.1 μl of primers (10 pmol each) and 2.5 μl of 2x
135   Multiplex PCR Plus Master mix (QIAGEN). The PCR protocol consisted of an initial DNA
136   polymerase (HotStar Taq) activation step at 95 °C for 5 min, followed by 35 cycles of 30 s of
137   95 °C (DNA denaturation step), 90 s at 50 °C, 52 °C, 54 °C and 56 °C (primer annealing step,
138   temperature ramp), and 30 s at 72 °C (elongation step); the last cycle was followed by a final
139   10 min extension step at 68 °C. 5 μl of the PCR product was visualized on a 2 % agarose gel.
140   Primer pairs that produced no amplification, multiple or unexpected size PCR products were
141   discarded. Eighteen primer pairs, showing discernably strong and specific signals, were picked
142   for further analysis. 5'-end fluorescently labelled reverse primers (6-Fam (Biomers) and NED,
143   VIC, PET (Thermo Fisher Scientific)) for the selected primer pairs were synthesized for
144   multiplexing and capillary electrophoresis. PCR was carried out in four multiplex reactions for
145   four *X. concava* DNA samples in a total volume of 10 μl containing 2.5 μl DNA, 1.0 μl of
146   fluorescently labelled primer pair mix (0.5 pmol each, containing up to five primer pairs,
147   depending on compatible annealing temperature, dye and expected fragment size range) and
148   5.0 μl of 2x Multiplex PCR Plus Master mix (QIAGEN). PCR reaction conditions were as
149   described above with the respective optimal annealing temperature for each primer pair mix.
150   Reactions were diluted 1:2 and sent to Macrogen Europe (Amsterdam, the Netherlands) for
151   fragment analysis.
152   Allele sizes were scored using GeneMapper 5.0 (Applied Biosystems). No marker showed strong
153   stutter peaks or intensive background signal. Two primer pairs appeared to be monomorphic and
154   were excluded from further analyses. Sixteen primer pairs showed apparent polymorphism for
155   the four tested samples and were finally selected (Table 1).
156
157   **COI and NaK polymerase chain reaction analysis**
158   Primers used for amplification and sequencing are listed in Table 2. The mitochondrial region
159   amplified is a 1,078 bp long fragment of cytochrome oxidase subunit I gene (*COI*). The first
160   658 bp of this fragment (from the 5' end) correspond to the standard barcode region of the animal
161   kingdom (Hebert et al. 2004). Additionally, a 1,654 bp long fragment of the nuclear gene region
162   of sodium/potassium-transporting ATPase subunit alpha (*NaK*) was amplified.
163   PCR reactions were carried out in a total volume of 20–25 μl containing 1.5–3.0 μl DNA, 1.2–
164   2.5 μl of primers (5 pmol each) and 10.0–12.5 μl of 2x Multiplex PCR Plus Master mix
165   (QIAGEN). The PCR protocol consisted of an initial DNA polymerase (HotStar Taq) activation
166   step at 95 °C for 5 min, followed by 38–40 cycles of 30 s at 95 °C, 90 s at 49–59 °C depending
167   on the primer set used, and 50–120 s (depending on the amplicon size) at 72 °C; the last cycle
168   was followed by a final 30 min extension step at 68 °C. 3 μl of the PCR product was visualized
169   on a 1.4 % agarose gel. Primers and dNTPs were inactivated with FastAP and Exonuclease I
170   (Thermo Fisher Scientific). 1.7–2.2 U of both enzymes were added to 17–22 μl of PCR solution
171   and incubated for 15 min at 37 °C, followed by 15 min at 85 °C. Purified PCR products were
172   sent to Macrogen Europe (Amsterdam, the Netherlands) for sequencing. To obtain unequivocal

173 sequences, both sense and antisense strands were sequenced. Sequences were aligned manually
174 with Geneious 11.0.5. Ambiguous positions (i.e., double peaks in chromatograms of both
175 strands) due to heterozygosity or heteroplasmy were coded using IUPAC symbols. Sequences
176 have been deposited in the GenBank (NCBI) database (accession numbers MK265017–
177 MK265114 and MK264919–MK265016, for detailed data see Table S1).
178

179 **Genetic data analysis**
180 Estimations of genetic variation were obtained by calculating average number of alleles ($N_A$),
181 observed ($H_O$) and expected heterozygosity ($H_E$) as well as deviations from Hardy-Weinberg
182 equilibrium (HWE) for each locus for all *X. concava* populations using ARLEQUIN 3.5.2.2
183 (Excoffier & Lischer, 2010) and 1,000 permutations. The same program was used to assess the
184 suitability of resolving population differentiation by estimating population pairwise measures of
185 $F_{ST}$ (1,000 permutations). The program GENEPOP 4.7.0 (Rousset, 2008) was used to estimate
186 the inbreeding coefficient $F_{IS}$ (1,000 permutations). GENEPOP was also used in combination
187 with the ENA correction implemented in the program FreeNA (Chapuis & Estoup, 2007) to test
188 for the presence and frequency of null alleles in the populations and to correct for the potential
189 overestimation of $F_{ST}$ values induced by the occurrence of null alleles (1,000 permutations).
190 Number of genotypes (NG) in the dataset was counted with Excel. To test for isolation by
191 distance, a Mantel test for the microsatellite data was performed (1,000 replicates) in ALLELES
192 IN SPACE (Miller, 2005).
193 To assess the suitability of the microsatellite markers for assessing genetic population structures,
194 three independent Bayesian assignment tests were carried out, one non-spatial using
195 STRUCTURE 2.3.4 (Pritchard, Stephens & Donnelly, 2000) and two spatial model based using
196 BAPS 6.0 (Corander, Waldmann & Sillanpää, 2003; Corander, Sirén & Arjas, 2008) and
197 GENELAND 4.0.8 (Guillot, Mortier & Estoup, 2005). GENELAND assignment results for the
198 microsatellite markers were also compared with results in GENELAND for the mitochondrial
199 and nuclear gene coding markers (here without any comparison with a non-spatial assignment in
200 STRUCTURE, since the model assumptions are likely to be violated for sequence data (Falush,
201 Stephens & Pritchard, 2003)). In BAPS, a maximum number of 10 $K$ was given as a prior. In
202 STRUCTURE, ten replicates for each $K$ from 1 to 10 were carried out with 50,000 burn-in steps
203 followed by 100,000 MCMC. The online program STRUCTURE HARVESTER (Earl &
204 vonHoldt, 2012) was used to infer the most likely value of $K$. GENELAND was carried out with
205 an uncertainty on coordinates of 25 km, 100,000 iterations, a thinning to every 100 replicate and
206 10 independent runs. In STRUCTURE and GENELAND, a no admixture model and
207 independency of allele frequency (uncorrelated model) was assumed, since correlated frequency
208 models, though more powerful in detecting subtle differentiations, are more sensitive to
209 departure from model assumptions (Guillot et al., 2012).
210

## Results

211  **Results**

212  The identification of microsatellite motifs by using HTS yielded 500 potential markers of which

213  50 were picked for a preliminary screening. Sixteen were finally implemented. Alongside primer

214  pairs that produced no amplification or were monomorphic, some also unexpectedly showed

215  PCR products of multiple sizes and had to be discarded.

216  The microsatellite markers amplified 3–14 different alleles and 3–18 different genotypes per

217  population and locus (Table 3). Observed heterozygosities ranged from 0.00 to 0.78 and were

218  significantly lower than those expected under Hardy-Weinberg equilibrium except for one locus,

219  indicating a deficiency of heterozygotes in the analyzed *Xyela concava* populations and/or the

220  presence of null alleles. This deficit is also confirmed by positive $F_{IS}$ values obtained for all but

221  three loci in one population. Estimated frequencies of null alleles were variable depending on the

222  respective microsatellite locus and *X. concava* population and varied between 0 and 39 % (Table

223  4).

224  The estimation of the frequency of null alleles, though highly variable depending on the locus-

225  population combination, did not introduce any bias to our dataset and thus did not cause an

226  overestimation of pairwise $F_{ST}$ values.

227  The $F_{ST}$ values uncorrected and corrected for the presence of null alleles showed higher values

228  between the populations of Monitor Pass and Uinta Mountains as well as between the

229  populations of Monitor Pass and Big Burro Mountains than the values between the populations

230  of Uinta Mountains and Big Burro Mountains (Table 5). In general, all $F_{ST}$ values were

231  comparatively low (0.028–0.113) but either had a considerably narrow confidence interval or

232  were significant or approaching the level of significance ($P = 0.055$). The $F_{ST}$ values for *NaK* and

233  *COI* were, in comparison, higher (0.215–0.740). While the values for *NaK* showed the same

234  pattern as the microsatellite markers in respect of genetic relationship of the populations, the $F_{ST}$

235  values for *COI* indicated relatively high differences between all populations (Table 6). The

236  Mantel test showed no isolation by distance ($r^2 = 0.0173$, $P < 0.001$). While spatial assignment

237  tests for *NaK* and the mircosatellites came up with the same pattern as the $F_{ST}$ values – as

238  indicated by the assignment of two populations with high posterior probabilities to one genetic

239  group or lineage (Figs. 2 and 3) – the non-spatial STRUCTURE analysis for microsatellites was

240  slightly non-confirmative, with genotypes from the Big Burro Mountains and Uinta Mountains

241  assigned to one separate lineage, yet with genotypes from all three populations assigned to one

242  shared overlapping genetic lineage. The analysis of the *COI* data revealed that each population

243  represented one distinct cluster ($K = 3$) (Fig. 3B, C, D).

244  All microsatellite markers were successfully tested for cross-species amplification. For the three

245  additional species of *Xyela* and *Pleroneura*, four markers showed polymorphic products and five

246  were apparently monomorphic for *X. deserti*. Eight markers showed polymorphic products for

247  the new *Xyela* species of the *alpigena* group, while no or unspecific fragments were amplified

248  for *Pleroneura koebelei* (Table 7).

249

## Discussion

250
251 Of 50 initial markers, only 16 could finally be implemented. Such high drop-out rates due to
252 large numbers of repetitive motifs throughout the genome causing nonspecific binding of primers
253 are already known (Schoebel et al., 2013). Other recent studies on invertebrates, using the same
254 commercial HTS approach for the identification of SSR motifs, resulted in 11 to 21 polymorphic
255 microsatellite markers, which nonetheless could be applied successfully (Reineke et al., 2015;
256 González-Castellano et al., 2018; Gomes et al., 2019).
257 The analyses demonstrated that the degree of variability of the new microsatellite marker set is
258 adequate in that it reveals polymorphic alleles within and across populations. The low significant
259 deviations from Hardy-Weinberg equilibrium as well as positive $F_{IS}$ values for almost all loci in
260 all populations could, however, have several causes. Given the ephemerality of the imagines and
261 their fluctuating abundance due to extended diapausing, the major reason might have been a
262 sampling bias, where only a fraction of each population was sampled (Wahlund effect).
263 Furthermore, homozygote genotypes equally distributed across all populations indicated
264 haploidy for altogether 26 specimens and may have had an impact on the discrepancy between
265 the observed and expected heterozygosity. Thelytokous parthenogenesis – producing solely
266 female offspring – which is known in xyelids (Blank, Shinohara & Altenhofer, 2013), also might
267 have contributed to the deficiency. However, due to the observed genotypic variation across the
268 data set, apomictic parthenogenesis seems unlikely (Caron et al., 2013).
269 The results based on the non-spatial model in STRUCTURE were not as confirmative as in the
270 spatial-model based assignment tests. Since in STRUCTURE no spatial information and
271 therefore fewer assumptions are incorporated, geographical barriers and distance as likely causes
272 for differentiated populations might have been underestimated (Coulon et al., 2006). On the other
273 hand, because it does not include spatial information, STRUCTURE may here indicate a subtler
274 genetic structure with possible higher exchange rates of the nuclear genome among all
275 populations. However, both model applications told a broadly concordant narrative for the
276 microsatellite markers, which are also supported by the low but significant $F_{ST}$ values. First, the
277 recent, seemingly discontinuous distribution of the hosts, *P. edulis* and *P. monophylla*, at higher
278 elevations in mountain ranges with up to 100 km between single patches, apparently does not
279 represent a barrier for recent and present gene flow. This is also supported by the Mantel test,
280 which indicated no isolation by distance. *X. concava* is assumed to be relatively stationary due to
281 the observed slow and erratic flight behavior (Burdick, 1961). Therefore, other explanations for
282 the ability to disperse over long distances should be considered, such as passive dispersal by
283 wind.
284 Second, the proposed geographically remote and restricted refugia of the host species during
285 glacial times (Bentancourt et al., 1991; Grayson 2011; Duran, Pardo & Mitton, 2012), and the
286 considerably long distances between them, might have been sufficiently great to cause restricted
287 gene flow and genetic segregation into two lineages. This assumption was also supported by the
288 high and significant $F_{ST}$ values and the genetic clustering of the *NaK* coding region, which due to
289 the slower mutation rates presumably better reflects events in the past. In the $F_{ST}$ statistics and

290    assignment tests of the microsatellite data (displaying presumably more recent events) the
291    segregation could still be detected, but also a recent state of admixture was indicated. To test this
292    hypothesis and a possibly ongoing admixture of the segregated lineages, populations of *X.*
293    *concava* in hybrid zones and overlapping distribution areas of the host species should be
294    included in future studies.
295    Compared to the results of the nuclear microsatellites and *NaK,* $F_{ST}$ values and Bayesian
296    statistics for the mitochondrial *COI* region showed a clear non-congruent pattern with a strong
297    genetic structure among all three populations. One explanation could be the small effective
298    population size ($N_e$) of the mitochondrial genome due to uniparental inheritance, which increases
299    the rate at which populations will become genetically more structured (lineage sorting; Harrison,
300    1989). However, this non-congruent pattern also might have been caused by biased
301    mitochondrial introgression as often found in "Symphyta" as recently discussed by Prous, Lee &
302    Mutanen (2019, preprint). The authors assume that mitochondrial introgression in sawflies might
303    be promoted by a combination of the haplodiploid reproduction system of Hymenoptera and the
304    low mitochondrial mutation rates in sawflies. The assumption is partly based on theoretical
305    models of Patten, Carioscia & Linnen (2015) showing that haplodiploid species are especially
306    prone to biased mitochondrial introgression. Furthermore, Sloan, Havird & Sharbrough (2017)
307    recently suggested that species with low mitochondrial mutation rates might favor a specific
308    beneficial (possibly locally adapted and/ or novel) mitochondrial haplotype to compensate for
309    deleterious mitonuclear mutation loads. The specific haplotype then selectively sweeps through a
310    population (or species) and purges deleterious mitochondrial mutations (the alternative solution
311    being compensatory co-evolutionary changes in the nuclear genome). Tendentially, this would
312    lead to a strong mitochondrial population structure and a mitonuclear discordance, which might
313    be reflected in the data set. Given the evidence for the very low evolutionary rates of molecular
314    characters in xyelids (Ronquist et al., 2012), this might be especially true for them. Additionally,
315    mitochondrial introgression might likely be the cause for mitonuclear discordance in cases where
316    there is a general agreement among large numbers of nuclear loci but discordance with
317    mitochondrial genealogies (Sloan, Havird & Sharbrough, 2017). Therefore, this new set of
318    microsatellites may also be an attractive tool to indicate mitochondrial introgression at the
319    population level of *X. concava* and other closely related xyelids.
320

## Conclusions

The implemented new set of microsatellite markers will be valuable for future analyses of additional and less distantly located populations while unraveling the population structure of *Xyela concava*. Together with other nuclear gene coding markers it can be used to elucidate both old and recent divisions in the gene pool to reveal more details of the phylogeography of this species. Furthermore, especially because of different underlying evolutionary processes affecting the nuclear and mitochondrial genome, this new set of microsatellites can potentially be used to reveal processes such as mitochondrial introgression at population level.

Even from this small data set, some tentative phylogeographic trends can be stated for *X. concava*. This study covers only three populations but nevertheless indicates a segregation of two genetic lineages and a recent state of admixture, which might have been caused by glacial retreat events. This would agree with proposed geographically separate glacial refugia of the host species. However, more populations covering the complete distribution of *X. concava,* especially populations from overlapping distribution areas of the hosts, need to be analyzed to test this hypothesis.

### Acknowledgements

We are grateful to C. Kutzscher (SDEI Müncheberg) for joining S.M. Blank during field work and for his support in the genetic lab. We thank A. Liston (SDEI Müncheberg) for a linguistic check of an earlier draft of the manuscript. We acknowledge the improvement of the manuscript by S.K. Monckton (Toronto) and an anonymous referee.

### References

Aron S, de Menten L, Van Bockstaele DR, Blank SM, Roisin Y. 2005. When Hymenopteran Males Reinvented Diploidy. *Current Biology* 15:824–827.

Bittner TD, Hajek AE, Haavik L, Allison J, Nahrung H. 2017. Multiple introductions of *Sirex noctilio* (Hymenoptera: Siricidae) in northeastern North America based on microsatellite genotypes, and implications for biological control. *Biological Invasions* 19:1431–1447 DOI: 10.1007/s10530-016-1365-1.

Bentancourt JL, Schuster WS, Mitton JB, Anderson RS. 1991. Fossil and Genetic History of a Pinyon Pine (*Pinus edulis*). *Ecology* 72:1685–1697.

Blank SM, Kramp K. 2017. *Xyela davidsmithi* (Hymenoptera, Xyelidae), a new pine catkin sawfly with an unusual host association from the Sierra Nevada. *Proceedings of the Washington Entomological Society* 119:703–717.

Blank SM, Kramp K, Shinohara A. 2017. *Xyela fusca* spec. nov. from Japan elucidates East Asian–North American relationships of *Xyela* (Hymenoptera, Xyelidae). *Zootaxa* 4303:103–121 DOI: 10.11646/zootaxa.4303.1.6

Blank SM, Kramp K, Smith DR, Sundukov YN, Wei M, Shinohara A. 2017. Big and beautiful: The *Megaxyela* species (Hymenoptera, Xyelidae) of East Asia and North America. *European Journal of Taxonomy* 348:1–46 DOI: 10.5852/ejt.2017.348.

Blank SM, Shinohara A, Altenhofer E. 2013. The Eurasian species of *Xyela* (Hymenoptera, Xyelidae): Taxonomy, host plants and distribution. *Zootaxa* 3629:1–106 DOI: 10.11646/zootaxa.3629.1.1.

Blank SM, Shinohara A, Byun B-K. 2005. The East Asian *Xyela* species (Hymenoptera: Xyelidae) associated with Japanese Red Pine (*Pinus densiflora*; Pinaceae) and their distribution history. *Insect Systematics & Evolution* 36:259–278 DOI: 10.1163/187631205788838393.

Burdick DJ. 1961. A Taxonomic and Biological Study of the Genus *Xyela* Dalman in North America. *University of California Publications in Entomology* 17:285–356.

Caron V, Norgate M, Ede FJ, Nyman T, Sunnucks P. 2013. Novel microsatellite DNA markers indicate strict parthenogenesis and few genotypes in the invasive willow sawfly *Nematus oligospilus*. *Bulletin of Entomological Research* 103:74–88 DOI: 10.1017/S0007485312000429.

Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24:621–631 DOI: 10.1093/molbev/msl191.

377   Cook N, Aziz N, Hedley PE, Morris J, Milne L, Karley AJ, Hubbard SF, Russell JR. 2011.
378       Transcriptome sequencing of an ecologically important graminivorous sawfly: a resource
379       for marker development. *Conservation Genetics Resources* 3:789–795 DOI:
380       10.1007/s12686-011-9459-7.
381   Corander J, Sirén J, Arjas E. 2008. Bayesian spatial modeling of genetic population structure.
382       *Computational Statistics* 23:111–129 DOI: 10.1007/s00180-007-0072-x.
383   Corander J, Waldmann P, Sillanpää MJ. 2003. Bayesian analysis of genetic differentiation
384       between populations. *Genetics* 163:367–374 DOI: 10.1093/bioinformatics/bth250.
385   Coulon A, Guillot G, Cosson J-F, Angibault JMA, Aulagnier S, Cargnelutti B, Galan M,
386       Hewison AJM. 2006. Genetic structure is influenced by landscape features: empirical
387       evidence from a roe deer population. *Molecular Ecology* 15:1669–1679 DOI:
388       10.1111/j.1365-294X.2006.02861.x.
389   Duran KL, Pardo A, Mitton JB. 2012. From middens to molecules: Phylogeography of the piñon
390       pine, *Pinus edulis*. *Journal of Biogeography* 39:1536–1544 DOI: 10.1111/j.1365-
391       2699.2012.02704.x.
392   Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for
393       visualizing STRUCTURE output and implementing the Evanno method. *Conservation*
394       *Genetics Resources* 4:359–361 DOI: 10.1007/s12686-011-9548-7.
395   Excoffier L, Lischer HEL. 2010. An Integrated Software Package for Population Genetics Data
396       Analysis. *Molecular Ecology Resources* 10:564–567 DOI: 10.1111/j.1755-
397       0998.2010.02847.x.
398   Funk DJ, Omland KE. 2003. Species-Level Paraphyly and Polyphyly: Frequency, Causes, and
399       Consequences, with Insights from Animal Mitochondrial DNA. *Annual Review of Ecology,*
400       *Evolution, and Systematics* 34:397–423. DOI: 10.1146/annurev.ecolsys.34.011802.132421.
401   Gomes SO, Souza IGB, Santos MF, Silva GR, Albrecht M, McKinley AS, Bentzen P, Diniz FM.
402       2019. Discovery of novel NGS-mined microsatellite markers and an exploratory analysis of
403       genetic differentiation between two Western Atlantic populations of *Cardisoma guanhumi*
404       Latreille, 1825 (Decapoda: Brachyura: Gecarcinidae). *Journal of Crustacean Biology*
405       39:181–185. DOI: 10.1093/jcbiol/ruy115.
406   González-Castellano I, Perina A, González-Tizón AM, Torrecilla Z, Martínez-Lage A. 2018.
407       Isolation and characterization of 21 polymorphic microsatellite loci for the rockpool shrimp
408       *Palaemon elegans* using Illumina MiSeq sequencing. *Scientific Reports* 8:8–13. DOI:
409       10.1038/s41598-018-35408-1.
410   Grayson DK. 2011. *The Great Basin: A Natural Prehistory*. Berkley and Los Angeles:
411       University of California Press.
412   Guillot G, Mortier F, Estoup A. 2005. GENELAND: A computer package for landscape
413       genetics. *Molecular Ecology Notes* 5:712–715 DOI: 10.1111/j.1471-8286.2005.01031.x.
414   Guillot G, Renaud S, Ledevin R, Michaux J, Claude J. 2012. A Unifying Model for the Analysis
415       of Phenotypic, Genetic, and Geographic Data. *Systematic Biology* 61:897–911 DOI:
416       10.1093/sysbio/sys038.

417  Harrison RG. 1989. Animal mitochondrial DNA as a genetic marker in population and
418      evolutionary biology. *Trends in Ecology and Evolution* 4:6–11. DOI: 10.1016/0169-
419      5347(89)90006-2.
420  Hartel KD, Frederick BA, Shanower TG. 2003. Isolation and characterization of microsatellite
421      loci in wheat stem sawfly *Cephus cinctus* and cross-species amplification in related species.
422      *Molecular Ecology Notes* 3:85–87 DOI: 10.1046/j.1471-8286.
423  Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA
424      barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*.
425      *Proceedings of the National Academy of Sciences of the United States of America*
426      101:14812–14817. DOI: 10.1073/pnas.0406166101.
427  Klopfstein S, Vilhelmsen L, Heraty JM, Sharkey M, Ronquist F. 2013. The Hymenopteran Tree
428      of Life: Evidence from Protein-Coding Genes and Objectively Aligned Ribosomal Data.
429      *PLoS ONE* 8. DOI: 10.1371/journal.pone.0069344.
430  Kopylov DS. 2014. New sawflies of the subfamily Madygellinae (Hymenoptera, Xyelidae) from
431      the Middle-Upper Triassic of Kyrgyzstan. *Paleontological Journal* 48:610–620 DOI:
432      10.1134/s0031030114060070.
433  Koressaar T, Remm M. 2007. Enhancements and modifications of primer design program
434      Primer3. *Bioinformatics* 23:1289–1291. DOI: 10.1093/bioinformatics/btm091.
435  Malm T, Nyman T. 2015. Phylogeny of the symphytan grade of Hymenoptera: new pieces into
436      the old jigsaw(fly) puzzle. *Cladistics* 31:1–17. DOI: 10.1111/cla.12069.
437  Miller MP. 2005. Alleles In Space (AIS): Computer software for the joint analysis of
438      interindividual spatial and genetic information. *Journal of Heredity* 96:722–724. DOI:
439      10.1093/jhered/esi119.
440  Patten MM, Carioscia SA, Linnen CR. 2015. Biased introgression of mitochondrial and nuclear
441      genes: A comparison of diploid and haplodiploid systems. *Molecular Ecology* 24:5200–
442      5210. DOI: 10.1111/mec.13318.
443  Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A,
444      Podsiadlowski L, Petersen M, Lanfear R, Diez PA, Heraty J, Kjer KM, Klopfstein S, Meier
445      R, Polidori C, Schmitt T, Liu S, Zhou X, Wappler T, Rust J, Misof B, Niehuis O. 2017.
446      Evolutionary History of the Hymenoptera. *Current Biology* 27:1013–1018. DOI:
447      10.1016/j.cub.2017.01.027.
448  Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus
449      genotype data. *Genetics* 155:945–959 DOI: 10.1111/j.1471-8286.2007.01758.x.
450  Prous M, Kramp K, Vikberg V, Liston A. 2017. North-Western Palaearctic species of
451      *Pristiphora* (Hymenoptera, Tenthredinidae). *Journal of Hymenoptera Research* 59:1–190
452      DOI: 10.3897/jhr.59.12656.
453  Prous M, Lee KM, Mutanen M. 2019. Detection of cross-contamination and strong mitonuclear
454      discordance in two species groups of sawfly genus *Empria*. *bioRxiv*. DOI:
455      http://dx.doi.org/10.1101/525626.
456  Prous M, Vikberg V, Liston A, Kramp K. 2016. North-Western Palaearctic species of the

457     *Pristiphora ruficornis* group (Hymenoptera, Tenthredinidae). *Journal of Hymenoptera*
458     *Research* 51:1–54 DOI: 10.3897/jhr.51.9162.

459 Reineke A, Assaf HA, Kulanek D, Mori N, Pozzebon A, Duso C. 2015. A novel set of
460     microsatellite markers for the European Grapevine Moth *Lobesia botrana* isolated using
461     next-generation sequencing and their utility for genetic characterization of populations from
462     Europe and the Middle East. *Bulletin of Entomological Research* 105:408–416. DOI:
463     10.1017/S0007485315000267.

464 Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. 2012. A
465     Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the
466     Hymenoptera. *Systematic Biology* 61:973–999 DOI: 10.1093/sysbio/sys058.

467 Rousset F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for
468     Windows and Linux. *Molecular Ecology Resources* 8:103–106 DOI: 10.1111/j.1471-
469     8286.2007.01931.x.

470 Schoebel CN, Brodbeck S, Buehler D, Cornejo C, Gajurel J, Hartikainen H, Keller D, Leys M,
471     Říčanová Š, Segelbacher G, Werth S, Csencsics D. 2013. Lessons learned from
472     microsatellite development for nonmodel organisms using 454 pyrosequencing. *Journal of*
473     *Evolutionary Biology* 26:600–611. DOI: 10.1111/jeb.12077.

474 Sloan DB, Havird JC, Sharbrough J. 2017. The on-again, off-again relationship between
475     mitochondrial genomes and species boundaries. *Molecular Ecology* 26:2212–2236. DOI:
476     10.1111/mec.13959.

477 Smith DR. 1967. A review of the larvae of Xyelidae, with notes on the family classification
478     (Hymenoptera). *Annals of the Entomological Society of America* 60:376–384. DOI:
479     10.1093/aesa/60.2.376.

480 Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012.
481     Primer3 – new capabilities and interfaces. *Nucleic Acids Research* 40:1–12. DOI:
482     10.1093/nar/gks596.

483

Sixteen polymorphic microsatellite loci and the corresponding flanking primer pairs identified in the pine catkin sawfly *Xyela concava*

| Locus | Size range (bp) | Motif | Ta in °C | label | Primer sequence (5'─3') |
|---|---|---|---|---|---|
| AG_30887_445 | 75–93 | $AAG_{(11)}$ | 50 | VIC | F: GTCTCGACTCCCTCCTACGA R: ACGGAAGTGCATCGGATCTTC |
| AG_30887_046 | 195–225 | $AGC_{(30)}$ | 50 | PET | F: CCTTTCGTCCTGGTTGACCA R: GATACGCCAGCCTATCCGTC |
| AG_30887_083 | 178–190 | $AAG_{(10)}$ | 50 | 6-Fam | F: TTCCAGTTTCTTGCAACGCG R: ATTCGCAAGCCTCTTCTGCA |
| AG_30887_188 | 179–188 | $AAT_{(9)}$ | 50 | NED | F: GCGGCGGTATAATGAGTCGT R: GGAAAGTGACTGCTACCGGT |
| AG_30887_479 | 93–102 | $ACT_{(8)}$ | 50 | PET | F: GCTGTTCACATGGCAGGTAG R: CCACCATCCCTACTACGGCT |
| AG_30887_193 | 110–134 | $AGC_{(17)}$ | 50 | VIC | F: AGAGTGCCAACGTGGGAAAT R: TTACTTTGCCCATGCCATGC |
| AG_30887_234 | 376–424 | $AATGCG_{(8)}$ | 50 | PET | F: AGTCTGATCCTTCCTGCGGA R: ATACGTGCCAGTTCGATCGT |
| AG_30887_282 | 239–263 | $AGC_{(10)}$ | 50 | 6-Fam | F: CTGTGCCTACGTCCCTTAGG R: CCCATCGTTTGGTCGGTAGA |
| AG_30887_286 | 103–121 | $AGC_{(8)}$ | 50 | NED | F: GCGTCCGTCTGAAATCTTGG R: CATTCGCATTCGACGCACTC |
| AG_30887_179 | 111–126 | $AGC_{(9)}$ | 50 | 6-Fam | F: CCCGTTCGTAAATCGGTCCT R: GACGTGGAATCGGTGGACTC |
| AG_30887_460 | 90–116 | $AT_{(5)}$ | 50 | PET | F: ACGTACTTATTGGGCGCGAA R: TTTACATGCTGTACACCGGGA |
| AG_30887_347 | 237–249 | $AAG_{(8)}$ | 50 | PET | F: CCCGGACCTCGTGCTATTC R: GGCGACAATCCCACGTGATA |
| AG_30887_393 | 136–175 | $AAG_{(8)}$ | 50 | 6-Fam | F: CCATCACTGTGCCGCGATAT R: GCACCTCAGGGATCCTCAAT |
| AG_30887_414 | 122–179 | $AAG_{(8)}$ | 50 | NED | F: TGATTTGTGCAACCGAGGGA R: CCCTTTATTCTCAGCAACCGC |
| AG_30887_012 | 130–148 | $AGG_{(9)}$ | 50 | PET | F: TTCCGGACGACTTTGACCTG R: CCTCGATTCCGATTCCCGTT |
| AG_30887_223 | 120–186 | $AAG_{(9)}$ | 50 | 6-Fam | F: TCAAAGCGGAGAAAGAGCGT R: TTAACCGCCATCGACCGTTC |

Nuclear *NaK* and mitochondrial *COI* primers used for amplification (PCR) and sequencing (seq)

| Gene Region | Primer name | Primer sequence (5'-3') | Ta in °C | PCR/ Sequencing | Reference |
|---|---|---|---|---|---|
| COI | symF1 | TTTCAACWAATCATAAARAYATTGG | 49 | PCR, seq | Prous et al. 2016 |
| COI | symR1 | TAAACTTCWGGRTGICCAAARAATC | 49 | PCR/ seq | Prous et al. 2016 |
| COI | symC1-J1751 | GGAGCNCCTGATATAGCWTTYCC | 49 | seq | Prous et al. 2016 |
| NaK | NaK263F | CTYAGCCAYGCRAARGCRAARGA | 59 | PCR/ seq | Prous et al. 2017 |
| NaK | NaK907Ri | TGRATRAARTGRTGRATYTCYTTIGC | 59 | seq | Prous et al. 2017 |
| NaK | NaK1250Fi | ATGTGGTTYGAYAAYCARATYATIGA | 59 | seq | Prous et al. 2017 |
| NaK | NaK1918R | GATTTGGCAATNGCTTTGGCAGTDAT | 59 | PCR/ seq | Prous et al. 2017 |

Comparative genetic diversity values for the three *Xyela concava* populations

Analyzed for each of the 16 microsatellite loci and on average over all loci including number of alleles ($N_A$), Number of genotypes (NG), observed ($H_O$) and expected ($H_E$) heterozygosity and estimates of $F_{IS}$

| Locus | Big Burro Mountains | | | | Monitor Pass | | | | Uinta Mountains | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $N_A$/NG | $F_{IS}$ | $H_O$ | $H_E$ | $N_A$/NG | $F_{IS}$ | $H_O$ | $H_E$ | $N_A$/NG | $F_{IS}$ | $H_O$ | $H_E$ |
| AG_30887_445 | 6/7 | 0.91 | 0.07 | 0.78* | 6/11 | 0.50 | 0.40 | 0.79* | 7/12 | 0.43 | 0.40 | 0.69* |
| AG_30887_046 | 10/11 | 0.33 | 0.61 | 0.85* | 9/14 | 0.44 | 0.47 | 0.82* | 7/11 | 0.04 | 0.78 | 0.80* |
| AG_30887_083 | 5/7 | 0.39 | 0.43 | 0.69* | 3/5 | 0.63 | 0.20 | 0.53* | 4/5 | -0.18 | 0.63 | 0.53* |
| AG_30887_188 | 4/6 | 0.84 | 0.11 | 0.66* | 3/5 | 0.31 | 0.37 | 0.53* | 3/5 | 0.63 | 0.25 | 0.66* |
| AG_30887_479 | 3/4 | 0.22 | 0.43 | 0.54* | 4/5 | 0.79 | 0.10 | 0.47* | 4/6 | -0.06 | 0.53 | 0.49* |
| AG_30887_193 | 6/8 | 0.31 | 0.50 | 0.72* | 5/12 | 0.47 | 0.40 | 0.74* | 7/13 | 0.04 | 0.78 | 0.80* |
| AG_30887_234 | 6/9 | 0.34 | 0.50 | 0.75* | 6/8 | 0.42 | 0.40 | 0.68* | 6/9 | 0.21 | 0.63 | 0.78* |
| AG_30887_282 | 8/8 | 0.40 | 0.46 | 0.77* | 6/9 | 0.61 | 0.30 | 0.75* | 6/9 | -0.03 | 0.73 | 0.70* |
| AG_30887_286 | 6/8 | 0.76 | 0.18 | 0.74* | 5/9 | 0.24 | 0.47 | 0.61 | 7/11 | 0.53 | 0.35 | 0.74* |
| AG_30887_179 | 3/3 | 1.00 | 0.00 | 0.62* | 5/7 | 0.55 | 0.20 | 0.43* | 5/6 | 0.76 | 0.15 | 0.61* |
| AG_30887_460 | 6/6 | 0.75 | 0.14 | 0.55* | 4/4 | 0.30 | 0.13 | 0.18* | 6/6 | 0.74 | 0.15 | 0.56* |
| AG_30887_347 | 4/5 | 0.34 | 0.43 | 0.64* | 3/6 | 0.51 | 0.33 | 0.67* | 4/7 | 0.06 | 0.63 | 0.66* |
| AG_30887_393 | 7/7 | 0.82 | 0.11 | 0.59* | 6/10 | 0.44 | 0.40 | 0.71* | 5/9 | 0.67 | 0.20 | 0.59* |
| AG_30887_414 | 12/12 | 0.35 | 0.54 | 0.82* | 10/18 | 0.54 | 0.40 | 0.86* | 9/13 | 0.13 | 0.68 | 0.77* |
| AG_30887_012 | 5/7 | 0.90 | 0.07 | 0.73* | 3/4 | 0.51 | 0.27 | 0.54* | 3/4 | 0.67 | 0.23 | 0.67* |
| AG_30887_223 | 9/11 | 0.76 | 0.14 | 0.80* | 14/18 | 0.36 | 0.47 | 0.89* | 13/15 | 0.72 | 0.23 | 0.82* |
| Mean | | 0.59 | 0.29 | 0.71 | | 0.48 | 0.33 | 0.64 | | 0.33 | 0.46 | 0.68 |
| S.D. | | 0.26 | 0.21 | 0.09 | | 0.13 | 0.12 | 0.18 | | 0.33 | 0.24 | 0.10 |

2 * significant departure from H-W equilibrium (P < 0.05)

3 S.D. = standard deviation

Estimated null allele frequencies for each of the 16 polymorphic microsatellite loci and each population including the average null allele frequency

Estimated null allele frequency

| Locus | Big Burro Mts | Monitor Pass | Uinta Mts |
|---|---|---|---|
| AG_30887_445 | 0.395 | 0.221 | 0.167 |
| AG_30887_046 | 0.165 | 0.191 | 0.028 |
| AG_30887_083 | 0.175 | 0.229 | 0.041 |
| AG_30887_188 | 0.334 | 0.116 | 0.247 |
| AG_30887_479 | 0.095 | 0.267 | 0.040 |
| AG_30887_193 | 0.130 | 0.194 | 0.037 |
| AG_30887_234 | 0.161 | 0.184 | 0.087 |
| AG_30887_282 | 0.194 | 0.260 | 0.036 |
| AG_30887_286 | 0.314 | 0.073 | 0.208 |
| AG_30887_179 | 0.381 | 0.190 | 0.282 |
| AG_30887_460 | 0.259 | 0.000 | 0.257 |
| AG_30887_347 | 0.148 | 0.200 | 0.048 |
| AG_30887_393 | 0.309 | 0.163 | 0.247 |
| AG_30887_414 | 0.196 | 0.245 | 0.053 |
| AG_30887_012 | 0.378 | 0.183 | 0.264 |
| AG_30887_223 | 0.319 | 0.162 | 0.314 |
| Mean | 0.247 | 0.180 | 0.147 |

Pairwise $F_{ST}$ estimates between populations of *Xyela concava* for the 16 microsatellite loci including corresponding *P* values and confidence intervals

Estimates are given both uncorrected and corrected for the presence of null alleles. Bold typeface denotes pairwise $F_{ST}$ estimates that are significantly different from zero ($P < 0.005$). Values in square brackets indicate 95 % confidence intervals for pairwise corrected $F_{ST}$ estimates

| $F_{ST}$ uncorrected | Big Burro Mts | Monitor Pass | Uinta Mts |
|---|:---:|:---:|:---:|
| Big Burro Mts | * | | |
| Monitor Pass | **0.09182** | * | |
| Uinta Mts | 0.02254 | **0.07705** | * |

| $F_{ST}$ ENA corrected | Big Burro Mts | Monitor Pass | Uinta Mts |
|---|:---:|:---:|:---:|
| Big Burro Mts | * | | |
| Monitor Pass | 0.083 [0.054, 0.115] | * | |
| Uinta Mts | 0.015 [0.004, 0.028] | 0.065 [0.041, 0.094] | * |

Pairwise $F_{ST}$ estimates between populations of *Xyela concava* for *NaK* and *COI* including corresponding *P* values

Bold typeface denotes pairwise $F_{ST}$ estimates that are significantly different from zero ($P < 0.005$)

| *NaK* | BB Mts | Mon Pass | Uinta Mts |
|---|---|---|---|
| Big Burro Mts | * | | |
| Monitor Pass | 0.740 | * | |
| Uinta Mts | 0.215 | 0.680 | * |
| *COI* | BB Mts | Mon Pass | Uinta Mts |
| Big Burro Mts | * | | |
| Monitor Pass | 0.699 | * | |
| Uinta Mts | 0.508 | 0.678 | * |

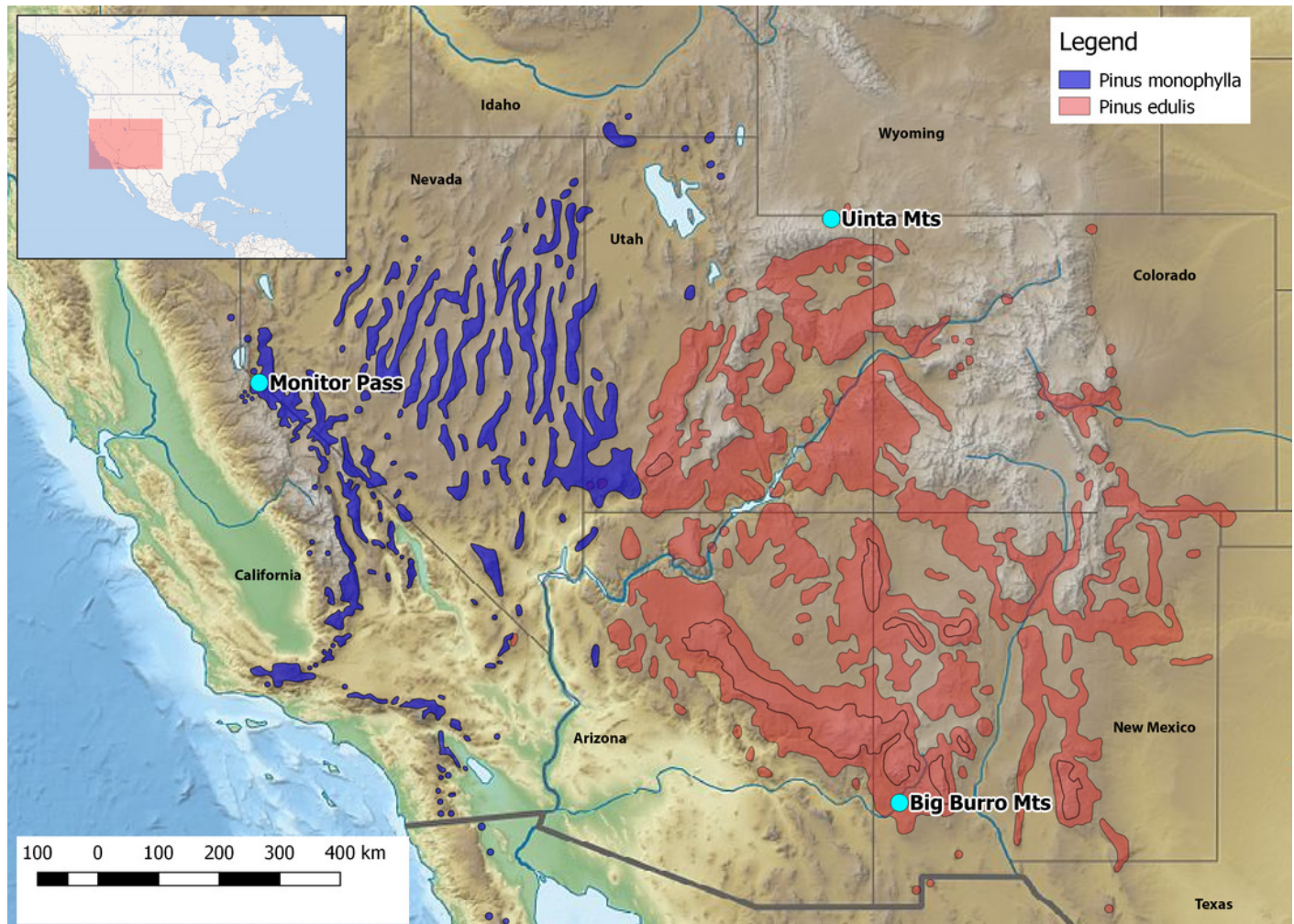
Cross-species amplification

(-) no product, (+) monomorphic product, (++) polymorphic product

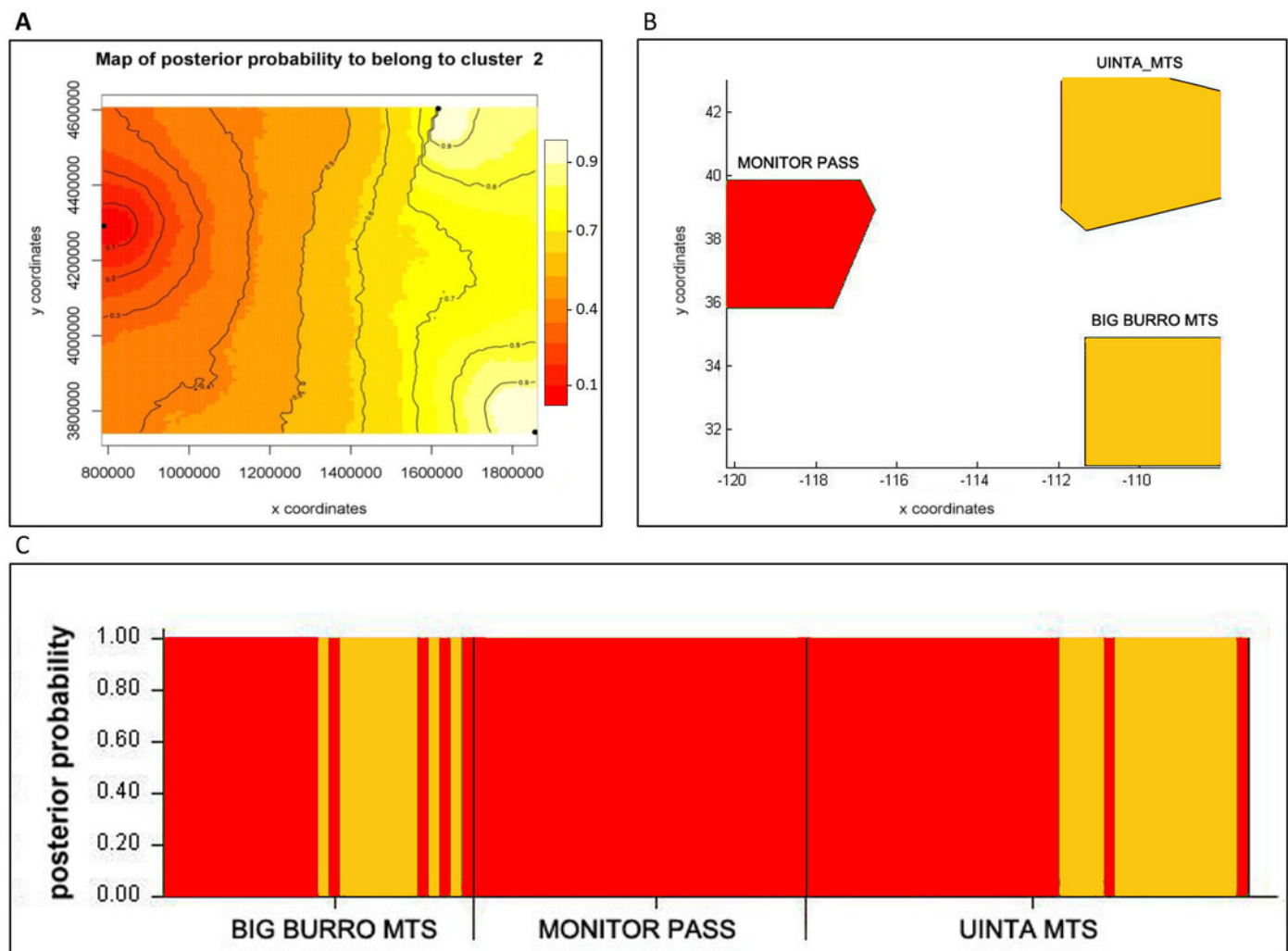
| Locus | *Xyela deserti* | *Xyela* spec. nov. | *Pleroneura koebelei* |
|---|---|---|---|
| AG_30887_445 | - | - | - |
| AG_30887_046 | + | ++ | - |
| AG_30887_083 | - | - | - |
| AG_30887_188 | - | - | - |
| AG_30887_479 | + | ++ | - |
| AG_30887_193 | - | ++ | - |
| AG_30887_234 | + | ++ | - |
| AG_30887_282 | ++ | ++ | - |
| AG_30887_286 | ++ | ++ | - |
| AG_30887_179 | - | - | - |
| AG_30887_460 | - | - | - |
| AG_30887_347 | ++ | ++ | - |
| AG_30887_393 | + | - | - |
| AG_30887_414 | ++ | - | - |
| AG_30887_012 | - | - | - |
| AG_30887_223 | + | ++ | - |

# Figure 1

Location of the collection areas and distribution of the host species

[p]Credit *Pinus spec*. shape files: https://data.usgs.gov/metadata[p]

# Figure 2

Bayesian assignment of *Xyela concava* populations to each of the identified clusters (*K* = 2) for the microsatellite markers

(A) GENELAND (Posterior probabilities are indicated in the scale bar. The contour lines in the maps indicate the spatial positions of genetic discontinuities. Lighter shading indicates a higher probability of belonging to the genetic population), (B) BAPS (the area of each population is proportional to the number of specimens used) and (C) STRUCTURE

# Figure 3

Bayesian spatial assignment (GENELAND) of *Xyela concava* populations to each of the identified clusters for (A) *NaK* ($K = 2$) and (B), (C), (D) *COI* ($K = 3$).

The different colors represent the estimated posterior probabilities of the membership to each cluster. Posterior probabilities are indicated in the scale bar. The contour lines in the maps indicate the spatial positions of genetic discontinuities. Lighter shading indicates a higher probability of belonging to the genetic population.

A

**Map of posterior probability to belong to cluster 2**

B

**Map of posterior probability to belong to cluster 1**

C

**Map of posterior probability to belong to cluster 2**

D

**Map of posterior probability to belong to cluster 3**