New approach to old wasps: First set of microsatellite markers for the pine catkin sawfly *Xyela concava* (Hymenoptera, Xyelidae) using next-generation sequencing (#37530)

First submission

Guidance from your Editor

Please submit by **10 Jun 2019** for the benefit of the authors (and your $200 publishing discount).

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Raw data check
Review the raw data. Download from the materials page.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files
Download and review all files from the materials page.

2 Figure file(s)
9 Table file(s)

Custom checks

DNA data checks

- Have you checked the authors data deposition statement?
- Can you access the deposited data?
- Has the data been deposited correctly?
- Is the deposition information noted in the manuscript?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

New approach to old wasps: First set of microsatellite markers for the pine catkin sawfly *Xyela concava* (Hymenoptera, Xyelidae) using next-generation sequencing

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

[1] Senckenberg Deutsches Entomologisches Institut, Müncheberg, Germany

Corresponding Author: Dustin Kulanek
Email address: dustin.kulanek@senckenberg.de

Microsatellites are widely used as powerful markers in population genetics because of their ability to access recent genetic variation and to resolve subtle population genetic structures. However, their development, especially for non-model organisms with no available genome-wide sequence data, has been difficult and time-consuming. Here, we used a next generation sequencing approach (NGS) for the development of a new set of microsatellite markers and implemented them successfully to answer questions on the population genetics and phylogeography of *Xyela concava*. The markers were characterized in three geographically distinct populations of *X. concava* and tested for cross-species amplification in two additional *Xyela* and one *Pleroneura* species (Xyelidae). All markers showed substantial polymorphism as well as revealed subtle genetic structures among the three genotyped populations. We also analyzed a fragment of the nuclear gene region of sodium/potassium-transporting ATPase subunit alpha (*NaK*) and a mitochondrial gene region partly coding for cytochrome oxidase subunit I (*COI*) to demonstrate different genetic resolutions and sex-biased patterns of these markers, and their potential for combined use in future studies on the phylogeography of *X. concava.*

New approach to old wasps: First set of microsatellite markers for the pine catkin sawfly *Xyela concava* (Hymenoptera, Xyelidae) using next-generation sequencing

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

Microsatellites are widely used as powerful markers in population genetics because of their ability to access recent genetic variation and to resolve subtle population genetic structures. However, their development, especially for non-model organisms with no available genome-wide sequence data, has been difficult and time-consuming. Here, we used a next generation sequencing approach (NGS) for the development of a new set of microsatellite markers and implemented them successfully to answer questions on the population genetics and phylogeography of *Xyela concava*. The markers were characterized in three geographically distinct populations of *X. concava* and tested for cross-species amplification in two additional *Xyela* and one *Pleroneura* species (Xyelidae). All markers showed substantial polymorphism as well as revealed subtle genetic structures among the three genotyped populations. We also analyzed a fragment of the nuclear gene region of sodium/potassium-transporting ATPase subunit alpha (*NaK*) and a mitochondrial gene region partly coding for cytochrome oxidase subunit I (*COI*) to demonstrate different genetic resolutions and sex-biased patterns of these markers, and their potential for combined use in future studies on the phylogeography of *X. concava.*

Introduction

Xyelidae have always attracted the attention of taxonomists and systematists. They represent the sister group of the rest of the megadiverse insect order Hymenoptera comprising, among others, bees, wasps and ants (Ronquist et al., 2012). The rich fossil records of Xyelidae include the earliest fossil forms of Hymenoptera dating from the Middle–Upper Triassic (Kopylov, 2014). Proper knowledge of its biology (e.g., larval and imaginal ecology, phylogeography, behavior) is crucial to reconstruct the ground plan of Hymenoptera and to understand the evolution of the non-xyelid hymenopteran lineages. However, such data are scarce for xyelids due to the rarity of many species, ephemerality of the imagines, and considerable problems in identifying species morphologically as well as genetically (e.g., Blank et al. 2013, 2017; Blank & Kramp 2017; Burdick 1961; Byun et al. 2005).

So far, only a limited number of microsatellite studies have been conducted on sawflies (Hartel, Frederick & Shanower, 2003; Cook et al., 2011; Caron et al., 2013; Bittner et al., 2017). Here, we report on the first developed and established set of 16 polymorphic nuclear microsatellite markers for *Xyela concava* Burdick, 1961, which will help to contribute to the understanding of the phylogeography and population dynamics of a member of the ancestral Xyelidae.

Xyela concava is widely distributed in southwestern USA, where it is closely associated with the pinyon-juniper woodland vegetation type of higher elevation semideserts, in particular with pine species of the subgenus *Strobus* subsection *Cembroides* (Farjon 2010). Females oviposit into developing male cones of *Pinus cembroides*, *P. edulis* and *P. monophylla*, where the larvae feed on the sporophylls. Imagines of the next generation emerge during the following spring and often visit flowering plants with easily accessible anthers, such as mountain mahogany (*Cercocarpus*

spp.) and bitterbrush or cliff-rose (*Purshia* spp.), from which they gather pollen for nutrition with their adapted mouthparts (Burdick 1961, Blank & Kramp, unpublished data).
We used a next generation sequencing approach for the development of the microsatellite markers and applied them to describe genetic structures and variation among and within three geographically distinct populations of *X. concava*. Furthermore, we compared the resolution of genetic variation of these markers with compiled data of one nuclear and one mitochondrial gene coding region and discuss their possible combined suitability for identifying genealogical lineages and answering phylogeographical questions. Finally, cross-amplification patterns for two species of *Xyela* and one of *Pleroneura* are illustrated.

Material and methods

Sampling

Xyela larvae were extracted from staminate cones of pines as described by Blank et al. (2013) and stored in 100 % ethanol at –20 °C. We included in the analysis larvae originating from three collection sites which are located 1,000–1,300 km from each other (see Table S1). The specimens are preserved in the Senckenberg Deutsches Entomologisches Institut, Müncheberg, Germany. Since it is impossible to identify *Xyela* larvae at species level morphologically, they were *COI* barcoded and identified by comparison with sequences from imagines already identified as *X. concava* morphologically (identification following Burdick 1961, reference sequences of imagines were published by Blank et al. 2017 and are deposited in the GenBank (NCBI) database, accession numbers KY198313 and KY198314). Finally, 98 larvae of *X. concava* were selected for the analysis (for detailed data see Table S1).

DNA extraction

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

Microsatellite marker development and screening

Total genomic DNA of a single female of *X. concava* (specimen ID: DEI-GISHym 30887, see Table S1) was extracted following the protocol described above. 10 ng/µl DNA in a total volume of 20 µl was sent to AllGenetics & Biology (Coruña, Spain) for the development of microsatellite markers. A library was prepared for the DNA sample using the Nextera XTDNA kit (Illumina), following the manufacturer's instructions. The library was enriched with the following microsatellite motifs: AC, AG, ACG, and ATCT. Enriched DNA was sequenced in the Illumina MiSeq platform (PE300). Microsatellite-containing reads were identified and selected. For 500 reads specific oligonucleotide primer pairs for the flanking regions of the microsatellite-sequences were designed and synthesized. Fifty primer pairs were picked and four *X. concava*

larvae (DEI-GISHym 32824–32827) were used for tests of polymorphism. Furthermore, 12 specimens of *X. deserti* Burdick, 1961, 12 specimens of an undescribed *Xyela* species, possibly a member of the *X. alpigena* group (Blank & Kramp 2017), and six specimens of *Pleroneura koebelei* Rohwer, 1910 (see appendix) were tested for cross-species amplification to check the marker system for potential use on two closely and one more distantly related xyelid species. The PCR analysis included a temperature gradient in the primer annealing step to find the best conditions for each primer pair. PCR was carried out in a total volume of 5 µl containing 0.5 µl DNA, 0.1 µl of primers (10 pmol each) and 2.5 µl of 2x Multiplex PCR Plus Master mix (QIAGEN). The PCR protocol consisted of an initial DNA polymerase (HotStar Taq) activation step at 95 °C for 5 min, followed by 35 cycles of 30 s of 95 °C (DNA denaturation step), 90 s at 50 °C, 52 °C, 54 °C and 56 °C (primer annealing step, temperature ramp), and 30 s at 72 °C (elongation step); the last cycle was followed by a final 10 min extension step at 68 °C. 5 µl of the PCR product was visualized on a 2 % agarose gel. Eighteen primer pairs, showing discernable polymorphic, strong and specific signals, were picked for further analysis. 5'-end fluorescently labelled reverse primers (6-Fam (Biomers) and NED, VIC, PET (Thermo Fisher Scientific)) for the selected primer pairs were synthesized for multiplexing and capillary electrophoresis. PCR was carried out in four multiplex reactions for four *X. concava* DNA samples in a total volume of 10 µl containing 2.5 µl DNA, 1.0 µl of fluorescently labelled primer pair mix (0.5 pmol each, containing up to five primer pairs, depending on compatible annealing temperature, dye and expected fragment size range) and 5.0 µl of 2x Multiplex PCR Plus Master mix (QIAGEN). PCR reaction conditions were as described above with the respective optimal annealing temperature for each primer pair mix. Reactions were diluted 1:2 and sent to Macrogen Europe (Amsterdam, the Netherlands) for fragment analysis.
Allele sizes were scored using GeneMapper 5.0 (Applied Biosystems). No marker showed strong stutter peaks or intensive background signal. Two primer pairs appeared to be monomorphic and were excluded from further analyses. Sixteen primer pairs showed apparent polymorphism for the four tested samples and were finally selected (Table 1).

COI and *NaK* Polymerase chain reaction analysis

Primers used for amplification and sequencing are listed in Table 2. The mitochondrial region used is a 1,078 bp long fragment of cytochrome oxidase subunit I gene (*COI*). The first 658 bp of this fragment (from the 5' end) correspond to the standard barcode region of the animal kingdom (Hebert et al. 2003). Additionally, a 1,654 bp long fragment of the nuclear gene region of sodium/potassium-transporting ATPase subunit alpha (*NaK*) was amplified.
PCR reactions were carried out in a total volume of 20–25 µl containing 1.5–3.0 µl DNA, 1.2–2.5 µl of primers (5 pmol each) and 10.0–12.5 µl of 2x Multiplex PCR Plus Master mix (QIAGEN). The PCR protocol consisted of an initial DNA polymerase (HotStar Taq) activation 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 on the primer set used, and 50–120 s (depending on the amplicon size) at 72 °C; the last cycle was followed by a final 30 min extension step at 68 °C. 3 µl of the PCR product was visualized on a 1.4 % agarose gel and then purified with FastAP and Exonuclease I (Thermo Fisher

Scientific). 1.7–2.2 U of both enzymes were added to 17–22 µl of PCR solution and incubated for 15 min at 37 °C, followed by 15 min at 85 °C. Purified PCR products were sent to Macrogen Europe (Amsterdam, the Netherlands) for sequencing. To obtain unequivocal sequences, both sense and antisense strands were sequenced. Sequences were aligned manually with Geneious 11.0.5 (www.geneious.com). Ambiguous positions (i.e., double peaks in chromatograms of both strands) due to heterozygosity or heteroplasmy were coded using IUPAC symbols. Sequences have been deposited in the GenBank (NCBI) database (accession numbers MK265017–MK265114 and MK264919–MK265016, for detailed data see Table S1).

Genetic data analysis

Estimations of genetic variation were obtained by calculating average number of alleles (N_A), observed (H_O) and expected heterozygosity (H_E) as well as deviations from Hardy-Weinberg equilibrium (HWE) for each locus for all *X. concava* populations using ARLEQUIN 3.5.2.2 (Excoffier & Lischer, 2010) and 1,000 permutations. The same program was used to assess the suitability of resolving population differentiation by estimating population pairwise measures of F_{ST} (1,000 permutations). The program GENEPOP 4.7.0 (Rousset, 2008) was used to estimate the inbreeding coefficient F_{IS} (1,000 permutations). GENEPOP was also used in combination with the ENA correction implemented in the program FreeNA (Chapuis & Estoup, 2007) to test for the presence and frequency of null alleles in the populations and to correct for the potential overestimation of F_{ST} values induced by the occurrence of null alleles (1,000 permutations). To assess the suitability of the microsatellite markers for ~~accessing~~ genetic population structures, three independent Bayesian assignment tests were carried out, one non spatial using STRUCTURE 2.3.4 (Pritchard, Stephens & Donnelly, 2000) and two spatial model based using BAPS 6.0 (Corander, Waldmann & Sillanpää, 2003; Corander, Sirén & Arjas, 2008) and GENELAND 4.0.8 (Guillot, Mortier & Estoup, 2005). GENELAND assignment results for the microsatellite markers were also compared with results in GENELAND for the mitochondrial and nuclear gene coding markers. In BAPS, a maximum number of 10 *K* was given as a prior. In STRUCTURE, ten replicates for each *K* from 1 to 10 were carried out with 50,000 burn-in steps followed by 100,000 MCMC. The online program STRUCTURE HARVESTER (Earl & vonHoldt, 2012) was used to infer the most likely value of *K*. GENELAND was carried out with an uncertainty on coordinates of 25 km, 100,000 iterations, a thinning to every 100 replicate and 10 independent runs. In STRUCTURE and GENELAND, a no admixture model and independency of allele frequency (uncorrelated model) was assumed, since correlated frequency models, though more powerful in detecting subtle differentiations, are more sensitive to departure from model assumptions (Guillot et al., 2012).

Results

The 16 microsatellite markers amplified 3–14 different alleles per population and locus (Table 3). Observed heterozygosities ranged from 0.00 to 0.78 and were significantly lower to those expected under Hardy-Weinberg equilibrium except for one locus, indicating a deficiency of

heterozygotes in the analyzed *Xyela concava* populations and/or the presence of null alleles. This deficit is also confirmed by positive F_{IS} values obtained for all but three loci in one population. Estimated frequencies of null alleles were variable depending on the respective microsatellite locus and *X. concava* population and varied between 0 and 39 % (Table 4).

The F_{ST} values uncorrected and corrected for the presence of null alleles showed higher values between the populations of Monitor Pass and Uinta Mountains as well as the populations of Monitor Pass and Big Burro Mountains than the values between the populations of Uinta Mountains and Big Burro Mountains (Table 5). In general, all F_{ST} values were comparatively low (0.028–0.113) but either had a considerable narrow confidence interval or were significant or approaching the level of significance ($P = 0.055$). The F_{ST} values for *NaK* and *COI* were, in comparison, higher (0.215–0.740). While the values for *NaK* showed the same pattern as the microsatellite markers in respect of genetic relationship of the populations, the F_{ST} values for *COI* indicated relatively high differences between all populations (Table 6). All assignment tests came up with the same pattern as the F_{ST} values indicated (Figs. 1 and 2). Based on data of the microsatellite and *NaK* markers, Monitor Pass represented one cluster and the remaining two populations were assigned to a second cluster ($K = 2$). Both clusters showed high posterior probabilities. The analysis of the *COI* data revealed that each population represented one distinct cluster ($K = 3$) (Fig. 2B, C, D).

All microsatellite markers were successfully tested for cross-species amplification. For the three additional species of *Xyela* and *Pleroneura*, five markers showed apparent monomorphic and four polymorphic bands for *X. deserti.* ~~Eight~~ markers showed polymorphic products for the new *Xyela* species of the *alpigena* group, while no or unspecific fragments were amplified for *Pleroneura koebelei* (Table 7).

Discussion

All analyses demonstrated that the degree of variability of the new microsatellite marker set appears substantial by showing polymorphic alleles within and across populations. The low significant deviations from Hardy-Weinberg equilibrium as well as positive F_{IS} values for almost all loci in all populations could, however, have several causes (e.g., sampling bias, wahlund effects). Nevertheless, the results may also reflect the true population structure. In this case, factors such as isolation, parthenogenesis, and/or founder effects may have been the causes of inbreeding or bottle-necking.

The estimation of the frequency of null alleles, though highly variable depending on the locus-population combination, did not introduce any bias to our dataset and thus did not cause an overestimation of pairwise F_{ST} values.

Since only larvae of unknown sex were analyzed in this study, the impact of haplodiploidy as the ancestral reproductive mode for all Hymenoptera (Aron et al., 2005) on the results remains to a certain degree unknown. However, negative *F*st values and Hardy-Weinberg equilibrium for some loci indicate no systematic deficiency in heterozygosity due to haplodiploidy.

The results in STRUCTURE were not as confirmative as in the spatial-model based assignment tests. Some individuals from the Big Burro Mountains and Uinta Mountains were assigned to a different cluster than the remaining individuals from those populations. Since in STRUCTURE no spatial information and therefore less assumptions are incorporated, geographical barriers and distance as most likely causes for differentiated populations are underestimated (Coulon et al., 2006). This demonstrates the necessity of the independent performance of two or more assignment methods to gain more confidence in the results (Pearse et al., 2006).
Compared to the results of the microsatellites and *NaK,* F_{ST} values and Bayesian statistics for the *COI* region showed a non-congruent pattern for the genetic structure. This could be expected because of higher mutation rates and smaller effective population size of mitochondrial genomes due to uniparental inheritance (Struck et al., 2018). Nevertheless, for future phylogeographical studies of *Xyela concava*, especially because of these different patterns, an analyzing system including markers for independent loci should be applied to gain a better understanding of the underlying evolutionary processes by additionally revealing events of sex-biased dispersal (Balloux et al., 2002).
In contrast, Bayesian posterior probabilities and F_{ST} values for the microsatellite loci and the nuclear *NaK* region showed a congruent pattern. The lower pairwise F_{ST} values in the microsatellite analysis can be interpreted as a higher resolution capacity and indicate subtle structures among populations which most likely emerged more recently (on a geological time scale).

Conclusions

The implemented new set of microsatellite markers will be supportive for future analyses of additional and less distantly located populations to unravel the population structure of *Xyela concava*. Furthermore, microsatellite markers and nuclear gene coding markers can be used to elucidate both old and recent divisions in the gene pool to reveal more details of the phylogeography of this species.
Even from this small data set, some tentative phylogeographic trends can be stated for *X. concava*. The results from this study cover only three populations but indicate a comparatively recent separation of two genetic lineages, which might have been caused by retreat events to different refugial areas during the last glacial maximum (LGM). This would agree with proposed geographically separate Pleistocene retreat areas of the host trees (Duran et al. 2012, Grayson 2011). However, more populations covering the complete distribution area of *X. concava* 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 the anonymous referees.

References

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

Balloux F, Lugon-Moulin N, Article R, Balloux F, Lugon-Moulin N. 2002. The estimation of population differentiation with microsatellite markers. *Molecular Ecology* 11:155–165 DOI: 10.1046/j.0962-1083.2001.01436.x.

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.

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.

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.

Byun B-K, Blank S, Shinohara A. 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.

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.

290 Cook N, Aziz N, Hedley PE, Morris J, Milne L, Karley AJ, Hubbard SF, Russell JR. 2011.
291 Transcriptome sequencing of an ecologically important graminivorous sawfly: a resource
292 for marker development. *Conservation Genetics Resources* 3:789–795 DOI:
293 10.1007/s12686-011-9459-7.
294 Corander J, Sirén J, Arjas E. 2008. Bayesian spatial modeling of genetic population structure.
295 *Computational Statistics* 23:111–129 DOI: 10.1007/s00180-007-0072-x.
296 Corander J, Waldmann P, Sillanpää MJ. 2003. Bayesian analysis of genetic differentiation
297 between populations. *Genetics* 163:367–374 DOI: 10.1093/bioinformatics/bth250.
298 Coulon A, Guillot G, Cosson J-F, Angibault JMA, Aulagnier S, Cargnelutti B, Galan M,
299 Hewison AJM. 2006. Genetic structure is influenced by landscape features: empirical
300 evidence from a roe deer population. *Molecular Ecology* 15:1669–1679 DOI:
301 10.1111/j.1365-294X.2006.02861.x.
302 Duran KL, Pardo A, Mitton JB. 2012. From middens to molecules: Phylogeography of the piñon
303 pine, *Pinus edulis*. *Journal of Biogeography* 39:1536–1544 DOI: 10.1111/j.1365-
304 2699.2012.02704.x.
305 Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for
306 visualizing STRUCTURE output and implementing the Evanno method. *Conservation*
307 *Genetics Resources* 4:359–361 DOI: 10.1007/s12686-011-9548-7.
308 Excoffier L, Lischer HEL. 2010. An Integrated Software Package for Population Genetics Data
309 Analysis. *Molecular Ecology Resources* 10:564–567 DOI: 10.1111/j.1755-
310 0998.2010.02847.x.
311 Grayson DK. 2011. *The Great Basin: A Natural Prehistory*. Berkley and Los Angeles:
312 University of California Press.
313 Guillot G, Mortier F, Estoup A. 2005. GENELAND: A computer package for landscape
314 genetics. *Molecular Ecology Notes* 5:712–715 DOI: 10.1111/j.1471-8286.2005.01031.x.
315 Guillot G, Renaud S, Ledevin R, Michaux J, Claude J. 2012. A Unifying Model for the Analysis
316 of Phenotypic, Genetic, and Geographic Data. *Systematic Biology* 61:897–911 DOI:
317 10.1093/sysbio/sys038.
318 Hartel KD, Frederick BA, Shanower TG. 2003. Isolation and characterization of microsatellite
319 loci in wheat stem sawfly *Cephus cinctus* and cross-species amplification in related species.
320 *Molecular Ecology Notes* 3:85–87 DOI: 10.1046/j.1471-8286.
321 Kopylov DS. 2014. New sawflies of the subfamily Madygellinae (Hymenoptera, Xyelidae) from
322 the Middle-Upper Triassic of Kyrgyzstan. *Paleontological Journal* 48:610–620 DOI:
323 10.1134/s0031030114060070.
324 Pearse DE, Arndt AD, Valenzuela N, Miller BA, Cantarelli V, Sites JW. 2006. Estimating
325 population structure under nonequilibrium conditions in a conservation context: Continent-
326 wide population genetics of the giant Amazon river turtle, *Podocnemis expansa* (Chelonia;
327 Podocnemididae). *Molecular Ecology* 15:985–1006 DOI: 10.1111/j.1365-
328 294X.2006.02869.x.
329 Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus

genotype data. *Genetics* 155:945–959 DOI: 10.1111/j.1471-8286.2007.01758.x.
Prous M, Kramp K, Vikberg V, Liston A. 2017. North-Western Palaearctic species of *Pristiphora* (Hymenoptera, Tenthredinidae). *Journal of Hymenoptera Research* 59:1–190 DOI: 10.3897/jhr.59.12656.
Prous M, Vikberg V, Liston A, Kramp K. 2016. North-Western Palaearctic species of the *Pristiphora ruficornis* group (Hymenoptera, Tenthredinidae). *Journal of Hymenoptera Research* 51:1–54 DOI: 10.3897/jhr.51.9162.
Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. 2012. A Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the Hymenoptera. *Systematic Biology* 61:973–999 DOI: 10.1093/sysbio/sys058.
Rousset F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8:103–106 DOI: 10.1111/j.1471-8286.2007.01931.x.
Struck TH, Feder JL, Bendiksby M, Birkeland S, Cerca J, Gusarov VI, Kistenich S, Larsson K-H, Liow LH, Nowak MD, Stedje B, Bachmann L, Dimitrov D. 2018. Finding Evolutionary Processes Hidden in Cryptic Species. *Trends in Ecology & Evolution* 33:153–163 DOI: 10.1016/j.tree.2017.11.007.

Table 1(on next page)

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

1

Table 2(on next page)

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

1

Table 3(on next page)

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 (NA), observed (H_O) and expected (H_E) heterozygosity and estimates of F_{IS}

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

1

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

3 S.D. = standard deviation

Table 4(on next page)

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

1

Table 5(on next page)

~~Uncorrected and for the presence of null alleles corrected~~ pairwise F_{ST} estimates between populations of *Xyela concava* for the 16 microsatellite loci including corresponding *P* values and confidence intervals

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]	*

1

Table 6(on next page)

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**	*

1

Table 7 (on next page)

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	+	++	-

1

Figure 1

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) , (B) BAPS and (C) STRUCTURE

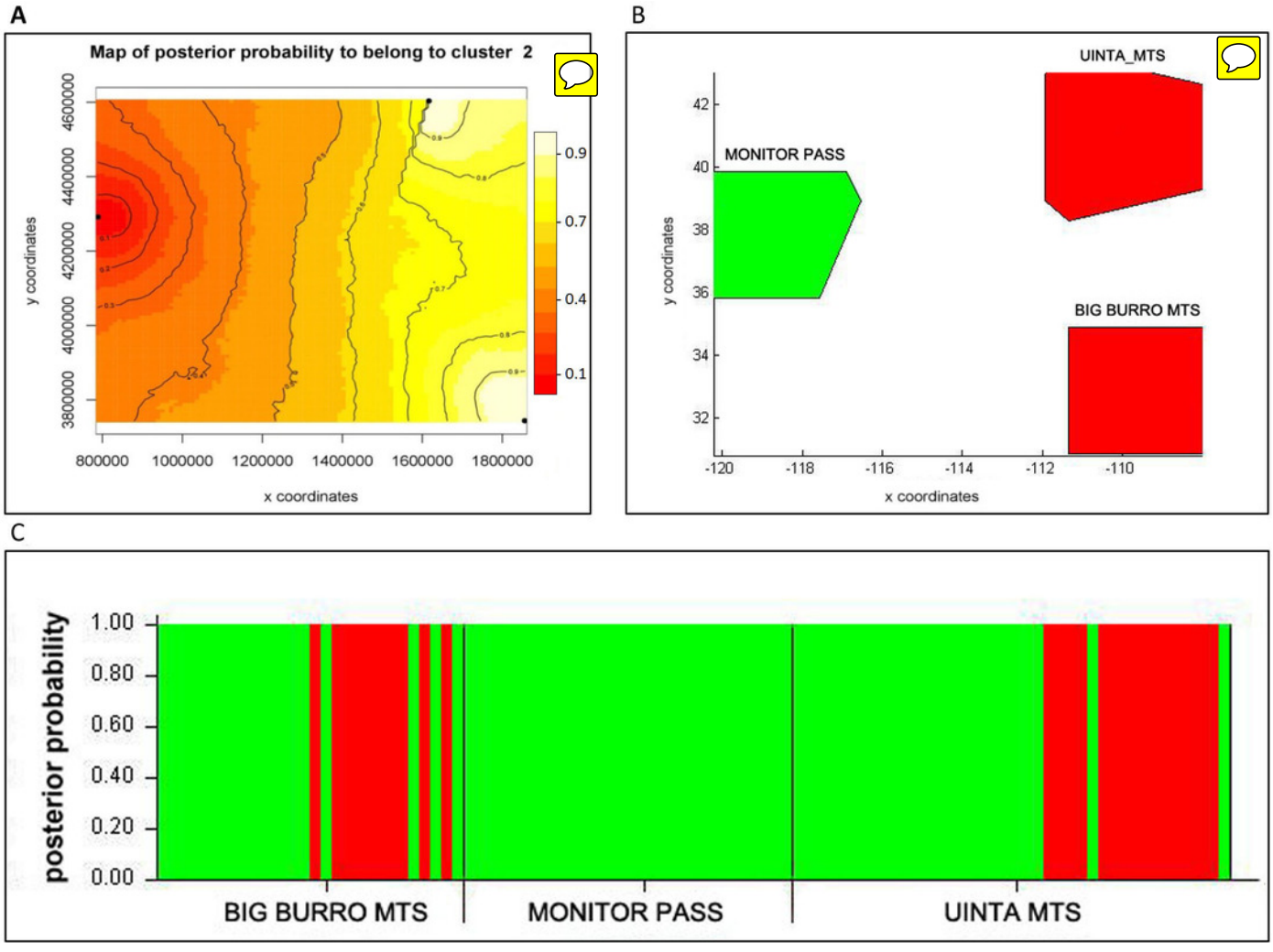

Figure 2

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