



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

Dustin Kulaneck, Stephan M. Blank and Katja Kramp

Senckenberg Deutsches Entomologisches Institut, Müncheberg, Germany

## 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, a commercial high-throughput sequencing approach (HTS) was used for the very first identification of microsatellite motifs in the genome of *Xyela concava* and the design of primer pairs flanking those motifs. Sixteen of those primer pairs were selected and implemented successfully to answer questions on the phylogeography and population genetics of *X. 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 revealing 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 partial mitochondrial gene region 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 and population genetics of *X. concava*. Although a limited number of populations was analyzed, we nevertheless obtained new insights on the latter two topics. The microsatellites revealed a generally high gene flow between the populations, but also suggested a deep historical segregation into two genetic lineages. This deep genetic segregation was confirmed by *NaK*. While the high gene flow was unexpected, because of assumed restricted dispersal ability of *X. concava* and the discontinuous distribution of the host trees between the populations, the segregation of two lineages is comprehensible and could be explained by different refuge areas of the hosts during glacial times. The *COI* results showed a discordant strong genetic structure between all populations, which might be explained by the smaller effective population size of the mitochondrial genome. However, given the frequent evidence of a similar nature in recent studies on sawflies, we also consider and discuss mitochondrial introgression on population level as an alternative explanation.

Submitted 16 May 2019  
Accepted 8 October 2019  
Published 13 November 2019

Corresponding author  
Dustin Kulaneck,  
dustin.kulaneck@senckenberg.de

Academic editor  
Vladimir Uversky

Additional Information and  
Declarations can be found on  
page 14

DOI 10.7717/peerj.8010

© Copyright  
2019 Kulaneck et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

**Subjects** Entomology, Molecular Biology, Population Biology

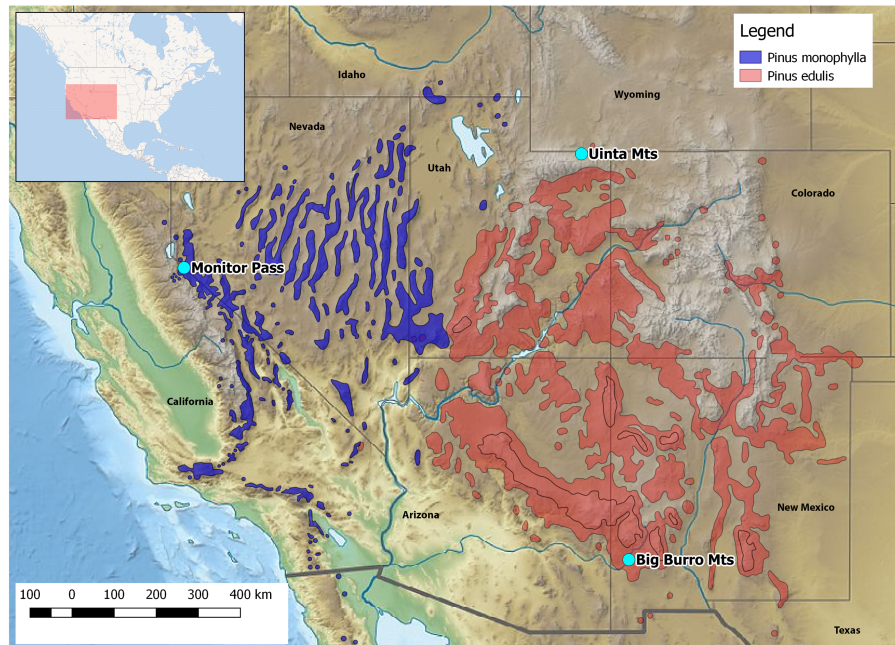
**Keywords** *Xyela concava*, Microsatellites, Population genetics, Phylogeography, Sawflies, Xyelidae, Mitochondrial introgression

## 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 (Ronquist *et al.*, 2012; Klopstein *et al.*, 2013; Malm & Nyman, 2015), which is traditionally divided into the paraphyletic “Symphyta” (missing the wasp waist) and the monophyletic Apocrita (sharing a wasp waist as a derived feature) (Malm & Nyman, 2015). The recent inconsistent phylogenetic placement of xyelids together with Pamphiliidae and Tenthredinidae as sister clade to all remaining Hymenoptera by Peters *et al.* (2017) might have been caused by an artificial grouping due to shared very slow mutation rates in those groups (Ronquist *et al.*, 2012).

The rich fossil record of Xyelidae includes the earliest fossil forms of Hymenoptera dating from the Middle–Upper Triassic (Kopylov, 2014). Proper knowledge of the phylogeography and population genetics of xyelids is therefore important in understanding the underlying evolutionary processes, which in turn will help to understand the evolution of other hymenopterans. Unfortunately, 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., Burdick, 1961; Blank, Shinohara & Byun, 2005; Blank, Shinohara & Altenhofer, 2013; Blank & Kramp, 2017; Blank, Kramp & Shinohara, 2017; Blank *et al.*, 2017). While a limited number of microsatellite studies has been conducted on sawflies (Hartel, Frederick & Shanower, 2003; Cook *et al.*, 2011; Caron *et al.*, 2013; Bittner *et al.*, 2017), none has focused on xyelids. Consequently, little is known about the population dynamics of this early-diverging group, including effects of ephemerality of imagines and their dispersal ability, host adaptation and host dependency, and reproduction mode.

Here, we report on the first developed and tested set of 16 polymorphic nuclear microsatellite markers for *Xyela concava* (Burdick, 1961), to shed light on the latter issues. *X. concava* is widely distributed in southwestern USA, where it is closely associated with the pinyon-juniper woodland vegetation type of higher elevation semideserts, i.e., with pine species of the subgenus *Strobus* subsection *Cembroides* (Farjon, 2010). Females oviposit into developing male cones of *Pinus cembroides* Zuccarini, 1830, *P. edulis* Engelman, 1848 and *P. monophylla* Torrey & Frémont, 1845 (Fig. 1), where the larvae feed on the sporophylls. After having ceased feeding, *Xyela* larvae dig into the soil below the host trees and may diapause up to several years before pupating (Blank, Shinohara & Altenhofer, 2013). Imagines of the next generation emerge during spring and often visit flowering plants with easily accessible anthers, such as mountain mahogany (*Cercocarpus* spp.) and cliff-rose (*Purshia* spp.), from which they gather pollen for nutrition with their adapted mouthparts (Burdick, 1961; SM Blank, pers. obs., 2013–2017). Flight behavior is described as erratic and slow (Burdick, 1961; Aron *et al.*, 2005). Therefore, given an assumed restricted dispersal



**Figure 1** Location of the collection areas and distribution of the host species. Credit *Pinus spec.* shape files: <https://data.usgs.gov/metadata>.

Full-size DOI: [10.7717/peerj.8010/fig-1](https://doi.org/10.7717/peerj.8010/fig-1)

ability and a close association with particular host species, it is intriguing to see how the variation within and among populations have been influenced by the distribution of the host trees during glacial and postglacial times. The high resolution power and therefore high capability of microsatellite markers to assess subtle and recent population genetic structures makes them well suited to this task. We used a commercial high-throughput 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 for 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*, sister taxon of *Xyela* (Smith, 1967), are illustrated.

## MATERIAL AND METHODS

### Sampling

*Xyela* larvae were extracted from staminate cones of pines as described by Blank, Shinohara & Altenhofer (2013) and stored in 100% ethanol at  $-20^{\circ}\text{C}$ . We included in the analysis larvae originating from three collection sites which are located 900–1,200 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 to species level morphologically, they were COI barcoded and identified by comparison with sequences from imagines identified as *X. concava* morphologically

(identification following [Burdick \(1961\)](#), reference sequences of imagines were published by [Blank, Kramp & Shinohara \(2017\)](#) 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 later 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 h incubation time at 55 °C (Thermomixer, without shaking) for cell lysis. The extracted DNA was stored at −20 °C until later use. The integuments were retained and stored in 70% ethanol.

### 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 commercial identification of microsatellite motifs and the design of motif flanking primer pairs. 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) and produced 3,043,190 paired-end reads. These paired-end reads were processed in Geneious 10.0.5 (Biomatters, Ltd.) using in-house developed scripts (property of AllGenetics & Biology) and overlapped into 1,521,595 sequences (trim error probability limit of 0.03). Primer design was carried out by AllGenetics & Biology in Primer 3 ([Koressaar & Remm, 2007](#); [Untergasser et al., 2012](#)) for 500 sequences containing microsatellite motifs. For a preliminary screening, 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 [Table S1](#)) 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 2× 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. Primer pairs that produced no amplification, multiple or unexpected size PCR products were discarded. Eighteen primer pairs, showing discernably 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  $\mu$ l containing 2.5  $\mu$ l DNA, 1.0  $\mu$ l 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  $\mu$ l of 2 $\times$ 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 amplified 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.*, 2004). 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  $\mu$ l containing 1.5–3.0  $\mu$ l DNA, 1.2–2.5  $\mu$ l of primers (5 pmol each) and 10.0–12.5  $\mu$ l of 2 $\times$ 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  $\mu$ l of the PCR product was visualized on a 1.4 % agarose gel. Primers and dNTPs were inactivated with FastAP and Exonuclease I (Thermo Fisher Scientific). 1.7–2.2 U of both enzymes were added to 17–22  $\mu$ l 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. 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

**Table 1** Sixteen polymorphic microsatellite loci and the corresponding flanking primer pairs identified in the pine catkin sawfly *Xyela con-cava*.

Locus	Size range (bp)	Motif	Ta in °C	Label	Primer sequence (5'—3')
AG_30887_445	75–93	AAG <sub>(11)</sub>	50	VIC	F: GTCTCGACTCCCTCCTACGA R: ACGGAAGTGCATCGGATCTTC
AG_30887_046	195–225	AGC <sub>(30)</sub>	50	PET	F: CCTTTCGTCTGGTTGACCA R: GATACGCCAGCCTATCCGTC
AG_30887_083	178–190	AAG <sub>(10)</sub>	50	6-Fam	F: TTCCAGTTTCTTGCAACGCG R: ATTTCGAAGCCTCTTCTGCA
AG_30887_188	179–188	AAT <sub>(9)</sub>	50	NED	F: GCGGCGGTATAATGAGTCGT R: GGAAAGTGACTGCTACCGGT
AG_30887_479	93–102	ACT <sub>(8)</sub>	50	PET	F: GCTGTTCACATGGCAGGTAG R: CCACCATCCCTACTACGGCT
AG_30887_193	110–134	AGC <sub>(17)</sub>	50	VIC	F: AGAGTGCCAACGTGGGAAAT R: TTAAGTTGCCATGCCATGC
AG_30887_234	376–424	AATGCG <sub>(8)</sub>	50	PET	F: AGTCTGATCCTTCCTGCGGA R: ATACGTGCCAGTTCGATCGT
AG_30887_282	239–263	AGC <sub>(10)</sub>	50	6-Fam	F: CTGTGCCTACGTCCCTTAGG R: CCCATCGTTTGGTCGGTAGA
AG_30887_286	103–121	AGC <sub>(8)</sub>	50	NED	F: GCGTCCGTCTGAAATCTTGG R: CATTTCGATTCGACGCACTC
AG_30887_179	111–126	AGC <sub>(9)</sub>	50	6-Fam	F: CCCGTTTCGTAATCGGTCCT R: GACGTGGAATCGGTGGACTC
AG_30887_460	90–116	AT <sub>(5)</sub>	50	PET	F: ACGTACTTATTGGGCGCGAA R: TTTACATGCTGTACACCGGGA
AG_30887_347	237–249	AAG <sub>(8)</sub>	50	PET	F: CCCGGACCTCGTGCTATTC R: GCGGACAATCCCACGTGATA
AG_30887_393	136–175	AAG <sub>(8)</sub>	50	6-Fam	F: CCATCACTGTGCCGCGATAT R: GCACCTCAGGGATCCTCAAT
AG_30887_414	122–179	AAG <sub>(8)</sub>	50	NED	F: TGATTTGTGCAACCGAGGGA R: CCCTTTATTCTCAGCAACCGC
AG_30887_012	130–148	AGG <sub>(9)</sub>	50	PET	F: TTCCGGACGACTTTGACCTG R: CCTCGATTCCGATTCCTCGTT
AG_30887_223	120–186	AAG <sub>(9)</sub>	50	6-Fam	F: TCAAAGCGGAGAAAGAGCGT R: TTAACCGCCATCGACCGTTC

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). Number of genotypes (NG)

**Table 2** 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	TTCAACWAATCATAAARAYATTGG	49	PCR, seq	<i>Prous et al. (2016)</i>
COI	symR1	TAAACTTCWGGRTGICCAAARAATC	49	PCR/ seq	<i>Prous et al. (2016)</i>
COI	symC1-J1751	GGAGCNCCTGATATAGCWTTYCC	49	seq	<i>Prous et al. (2016)</i>
NaK	NaK_263F	CTYAGCCAYGCRAARGCRAARGA	59	PCR/ seq	<i>Prous et al. (2017)</i>
NaK	NaK_907Ri	TGRATRAARTGRTGRATYTCYTTIGC	59	seq	<i>Prous et al. (2017)</i>
NaK	NaK_1250Fi	ATGTGGTTYGAYAAAYCARATYATIGA	59	seq	<i>Prous et al. (2017)</i>
NaK	NaK_1918R	GATTTGGCAATNGCTTTGGCAGTDAT	59	PCR/ seq	<i>Prous et al. (2017)</i>

in the dataset was counted with Excel. To test for isolation by distance, a Mantel test for the microsatellite data was performed (1,000 replicates) in ALLELES IN SPACE (*Miller, 2005*).

To assess the suitability of the microsatellite markers for assessing 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 (here without any comparison with a non-spatial assignment in STRUCTURE, since the model assumptions are likely to be violated for sequence data (*Falush, Stephens & Pritchard, 2003*)). In BAPS, a maximum number of  $K = 10$  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 identification of microsatellite motifs by using HTS yielded 500 potential markers of which 50 were picked for a preliminary screening. Sixteen were finally implemented. Alongside primer pairs that produced no amplification or were monomorphic, some also unexpectedly showed PCR products of multiple sizes and had to be discarded.

The microsatellite markers amplified 3–14 different alleles and 3–18 different genotypes per population and locus (*Table 3*). Observed heterozygosities ranged from 0.00 to 0.78 and were significantly lower than 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  $F_{IS}$

**Table 3** 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

**Notes.**

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

S.D., standard deviation.

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 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.

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 between 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 considerably 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). The Mantel test showed no isolation by distance ( $r^2 = 0.0173$ ,  $P < 0.001$ ). While spatial assignment tests for *NaK* and the microsatellites came up with the same pattern as the  $F_{ST}$  values—as indicated by the



**Table 4** 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

**Table 5** 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.

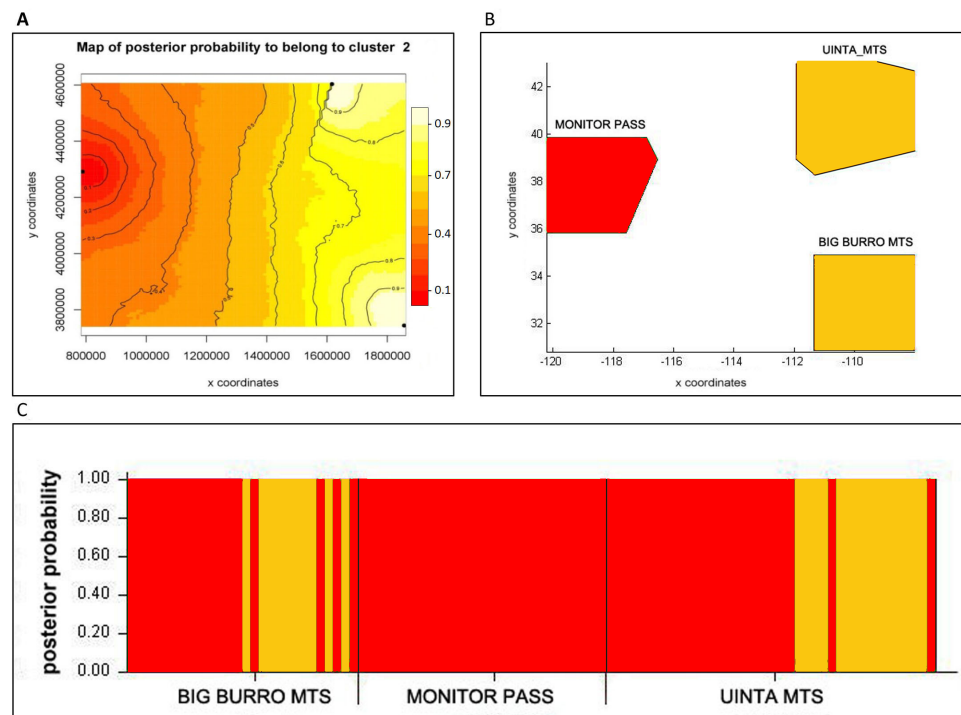
$F_{ST}$ uncorrected	Big Burro Mts	Monitor Pass	Uinta Mts
Big Burro Mts	*		
Monitor Pass	<b>0.09182</b>	*	
Uinta Mts	0.02254	<b>0.07705</b>	*
$F_{ST}$ ENA corrected	Big Burro Mts	Monitor Pass	Uinta Mts
Big Burro Mts	*		
Monitor Pass	0.083 [0.054, 0.115]	*	
Uinta Mts	0.015 [0.004, 0.028]	0.065 [0.041, 0.094]	*

assignment of two populations with high posterior probabilities to one genetic group or lineage (Figs. 2 and 3)—the non-spatial STRUCTURE analysis for microsatellites was slightly non-confirmative, with genotypes from the Big Burro Mountains and Uinta Mountains assigned to one separate lineage, yet with genotypes from all three populations assigned to one shared overlapping genetic lineage. The analysis of the *COI* data revealed that each population represented one distinct cluster ( $K = 3$ ) (Figs. 3B–3D).

All microsatellite markers were successfully tested for cross-species amplification. For the three additional species of *Xyela* and *Pleroneura*, four markers showed polymorphic

**Table 6** 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 ( $P < 0.005$ ).

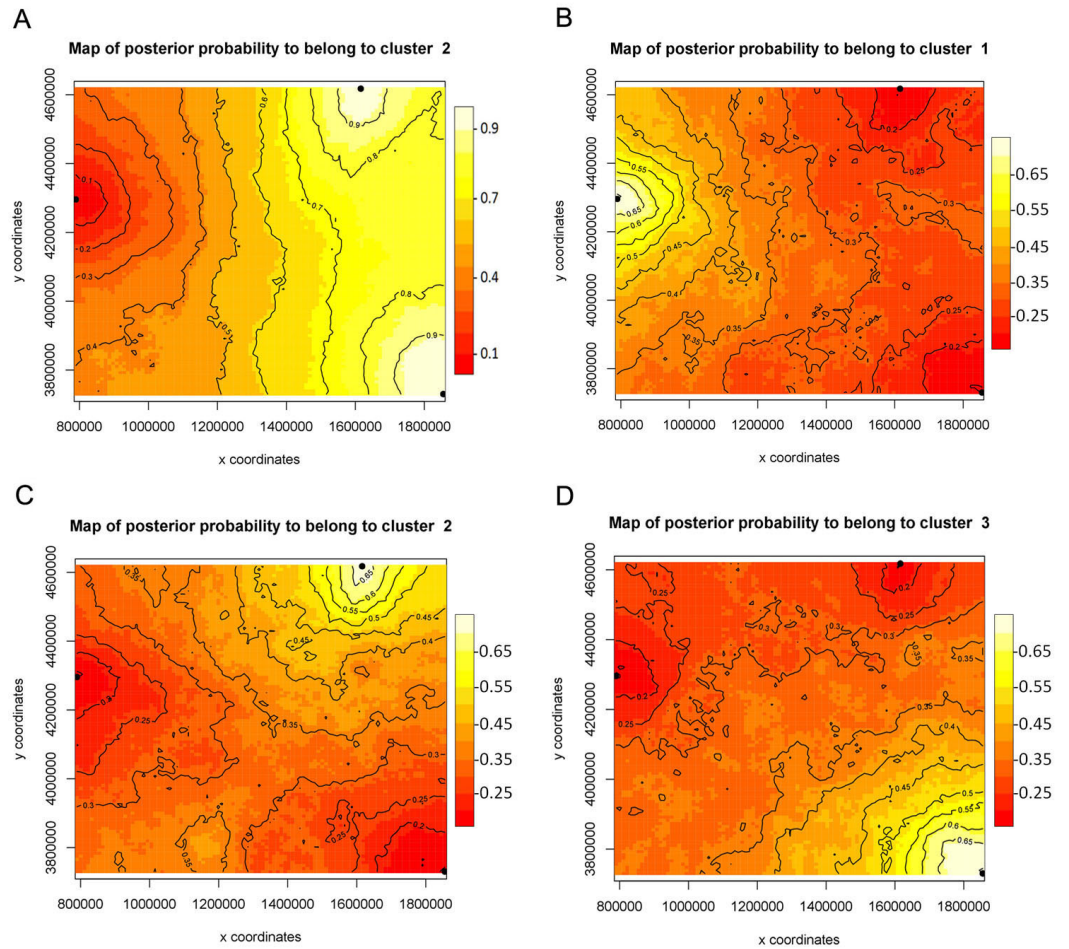
	Big Burro Mts	Monitor Mts	Uinta Mts
<b>NaK</b>			
Big Burro Mts	*		
Monitor Pass	<b>0.740</b>	*	
Uinta Mts	<b>0.215</b>	<b>0.680</b>	*
<b>COI</b>			
Big Burro Mts	*		
Monitor Pass	<b>0.699</b>	*	
Uinta Mts	<b>0.508</b>	<b>0.678</b>	*



**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.

Full-size [DOI: 10.7717/peerj.8010/fig-2](https://doi.org/10.7717/peerj.8010/fig-2)

products and five were apparently monomorphic 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).



**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.

Full-size [DOI: 10.7717/peerj.8010/fig-3](https://doi.org/10.7717/peerj.8010/fig-3)

## DISCUSSION

Of 50 initial markers, only 16 could finally be implemented. Such high drop-out rates due to large numbers of repetitive motifs throughout the genome causing nonspecific binding of primers are already known (Schoebel et al., 2013). Other recent studies on invertebrates, using the same commercial HTS approach for the identification of SSR motifs, resulted in 11 to 21 polymorphic microsatellite markers, which nonetheless could be applied successfully (Reineke et al., 2015; González-Castellano et al., 2018; Gomes et al., 2019).

The analyses demonstrated that the degree of variability of the new microsatellite marker set is adequate in that it reveals 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. Given the

**Table 7** Cross-species amplification.

Locus	<i>Xyela deserti</i>	<i>Xyela</i> spec. nov.	<i>Pleroneura</i> <i>koebelei</i>
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	+	++	–

**Notes.**

(–), no product; (+), monomorphic product; (++), polymorphic product.

ephemerality of the imagines and their fluctuating abundance due to extended diapausing, the major reason might have been a sampling bias, where only a fraction of each population was sampled (Wahlund effect). Furthermore, homozygote genotypes equally distributed across all populations indicated haploidy for altogether 26 specimens and may have had an impact on the discrepancy between the observed and expected heterozygosity. Thelytokous parthenogenesis—producing solely female offspring—which is known in xyelids ([Blank, Shinohara & Altenhofer, 2013](#)), also might have contributed to the deficiency. However, due to the observed genotypic variation across the data set, apomictic parthenogenesis seems unlikely ([Caron et al., 2013](#)).

The results based on the non-spatial model in STRUCTURE were not as confirmative as in the spatial-model based assignment tests. Since in STRUCTURE no spatial information and therefore fewer assumptions are incorporated, geographical barriers and distance as likely causes for differentiated populations might have been underestimated ([Coulon et al., 2006](#)). On the other hand, because it does not include spatial information, STRUCTURE may here indicate a subtler genetic structure with possible higher exchange rates of the nuclear genome among all populations. However, both model applications told a broadly concordant narrative for the microsatellite markers, which are also supported by the low but significant  $F_{ST}$  values. First, the recent, seemingly discontinuous distribution of the hosts, *Pinus edulis* and *P. monophylla*, at higher elevations in mountain ranges with up to 100 km between single patches, apparently does not represent a barrier for recent and present gene flow. This is also supported by the Mantel test, which indicated no isolation by distance. *X. concava* is assumed to be relatively stationary due to the observed slow

and erratic flight behavior (*Burdick, 1961*). Therefore, other explanations for the ability to disperse over long distances should be considered, such as passive dispersal by wind.

Second, the proposed geographically remote and restricted refugia of the host species during glacial times (*Bentancourt et al., 1991; Grayson, 2011; Duran, Pardo & Mitton, 2012*), and the considerably long distances between them, might have been sufficiently great to cause restricted gene flow and genetic segregation into two lineages. This assumption was also supported by the high and significant  $F_{ST}$  values and the genetic clustering of the *NaK* coding region, which due to the slower mutation rates presumably better reflects events in the past. In the  $F_{ST}$  statistics and assignment tests of the microsatellite data (displaying presumably more recent events) the segregation could still be detected, but also a recent state of admixture was indicated. To test this hypothesis and a possibly ongoing admixture of the segregated lineages, populations of *X. concava* in hybrid zones and overlapping distribution areas of the host species should be included in future studies.

Compared to the results of the nuclear microsatellites and *NaK*,  $F_{ST}$  values and Bayesian statistics for the mitochondrial *COI* region showed a clear non-congruent pattern with a strong genetic structure among all three populations. One explanation could be the different effective population sizes ( $N_e$ ) of the mitochondrial and nuclear genome, which will cause differences in the diversity of the genetic structure of populations over time (incomplete lineage sorting; *Harrison, 1989; Funk & Omland, 2003*). However, this non-congruent pattern also might have been caused by biased mitochondrial introgression as often found in “Symphyta” as recently discussed by *Prous, Lee & Mutanen (2019, preprint)*. The authors assume that mitochondrial introgression in sawflies might be promoted by a combination of the haplodiploid reproduction system of Hymenoptera (*Aron et al., 2005*) and the low mitochondrial mutation rates in sawflies (*Tang et al., 2019*). The assumption is partly based on theoretical models of *Patten, Carioscia & Linnen (2015)* showing that haplodiploid species are especially prone to biased mitochondrial introgression. Furthermore, *Sloan, Havird & Sharbrough (2017)* recently suggested that species with low mitochondrial mutation rates might favor a specific beneficial (possibly locally adapted and/or novel) mitochondrial haplotype to compensate for deleterious mitonuclear mutation loads. The specific haplotype then selectively sweeps through a population (or species) and purges deleterious mitochondrial mutations (the alternative solution being compensatory co-evolutionary changes in the nuclear genome). Tendentially, this would lead to a strong mitochondrial population structure and a mitonuclear discordance, which might be reflected in the data set. Given the evidence for the very low evolutionary rates of molecular characters in xyelids (*Ronquist et al., 2012*), this might be especially true for them. Additionally, mitochondrial introgression might likely be the cause for mitonuclear discordance in cases where there is a general agreement among large numbers of nuclear loci but discordance with mitochondrial genealogies (*Sloan, Havird & Sharbrough, 2017*). Therefore, this new set of microsatellites may also be an attractive tool to indicate mitochondrial introgression at the population level of *X. concava* and other closely related xyelids.

## 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.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

The authors received no funding for this work.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Dustin Kulanek conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Stephan M. Blank conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Katja Kramp conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

## DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The COI and NaK sequences are available at GenBank: [MK264919](#) to [MK265114](#).

## Data Availability

The following information was supplied regarding data availability:

The raw data are available in the [Supplemental Files](#) and at GenBank.

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.8010#supplemental-information>.

## REFERENCES

- Aron S, De Menten L, Van Bockstaele DR, Blank SM, Roisin Y. 2005. When hymenopteran males reinvented diploidy. *Current Biology* 15:824–827 DOI [10.1016/j.cub.2005.03.017](#).
- Bentancourt JL, Schuster WS, Mitton JB, Anderson RS. 1991. Fossil and genetic history of a pinyon pine (*Pinus edulis*). *Ecology* 72:1685–1697 DOI [10.2307/1940968](#).
- 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 DOI [10.4289/0013-8797.119.SpecialIssue.703](#).
- 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](https://doi.org/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](https://doi.org/10.1093/molbev/msl191).
- Cook N, Aziz N, Hedley PE, Morris J, Milne L, Karley AJ, Hubbard SF, Russell JR. 2011.** Transcriptome sequencing of an ecologically important gramminivorous sawfly: a resource for marker development. *Conservation Genetics Resources* **3**:789–795  
DOI [10.1007/s12686-011-9459-7](https://doi.org/10.1007/s12686-011-9459-7).
- Corander J, Sirén J, Arjas E. 2008.** Bayesian spatial modeling of genetic population structure. *Computational Statistics* **23**:111–129 DOI [10.1007/s00180-007-0072-x](https://doi.org/10.1007/s00180-007-0072-x).
- Corander J, Waldmann P, Sillanpää MJ. 2003.** Bayesian analysis of genetic differentiation between populations. *Genetics* **163**:367–374 DOI [10.1093/bioinformatics/bth250](https://doi.org/10.1093/bioinformatics/bth250).
- Coulon A, Guillot G, Cosson J-F, Angibault JMA, Aulagnier S, Cargnelutti B, Galan M, Hewison AJM. 2006.** Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology* **15**:1669–1679  
DOI [10.1111/j.1365-294X.2006.02861.x](https://doi.org/10.1111/j.1365-294X.2006.02861.x).
- Duran KL, Pardo A, Mitton JB. 2012.** From middens to molecules: phylogeography of the piñon pine, *Pinus edulis*. *Journal of Biogeography* **39**:1536–1544  
DOI [10.1111/j.1365-2699.2012.02704.x](https://doi.org/10.1111/j.1365-2699.2012.02704.x).
- Earl DA, vonHoldt BM. 2012.** STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**:359–361 DOI [10.1007/s12686-011-9548-7](https://doi.org/10.1007/s12686-011-9548-7).
- Excoffier L, Lischer HEL. 2010.** An integrated software package for population genetics data analysis. *Molecular Ecology Resources* **10**:564–567  
DOI [10.1111/j.1755-0998.2010.02847.x](https://doi.org/10.1111/j.1755-0998.2010.02847.x).
- Falush D, Stephens M, Pritchard JK. 2003.** Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**:1567–1587.
- Farjon A. 2010.** *A Handbook of the World's Conifers*. Leiden and Boston: Brill.
- Funk DJ, Omland KE. 2003.** Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics* **34**:397–423  
DOI [10.1146/annurev.ecolsys.34.011802.132421](https://doi.org/10.1146/annurev.ecolsys.34.011802.132421).
- Gomes SO, Souza IGB, Santos MF, Silva GR, Albrecht M, McKinley AS, Bentzen P, Diniz FM. 2019.** Discovery of novel NGS-mined microsatellite markers and an exploratory analysis of genetic differentiation between two Western Atlantic populations of *Cardisoma guanhumi* Latreille, 1825 (Decapoda: Brachyura: Gecarcinidae). *Journal of Crustacean Biology* **39**:181–185 DOI [10.1093/jcabi/ryy115](https://doi.org/10.1093/jcabi/ryy115).
- González-Castellano I, Perina A, González-Tizón AM, Torrecilla Z, Martínez-Lage A. 2018.** Isolation and characterization of 21 polymorphic microsatellite loci for the rockpool shrimp *Palaemon elegans* using Illumina MiSeq sequencing. *Scientific Reports* **8**:8–13 DOI [10.1038/s41598-018-35408-1](https://doi.org/10.1038/s41598-018-35408-1).



- Grayson DK. 2011.** *The great basin: a natural prehistory*. Berkeley and Los Angeles: University of California Press.
- Guillot G, Mortier F, Estoup A. 2005.** GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes* 5:712–715 DOI [10.1111/j.1471-8286.2005.01031.x](https://doi.org/10.1111/j.1471-8286.2005.01031.x).
- Guillot G, Renaud S, Ledevin R, Michaux J, Claude J. 2012.** A unifying model for the analysis of phenotypic, genetic, and geographic data. *Systematic Biology* 61:897–911 DOI [10.1093/sysbio/sys038](https://doi.org/10.1093/sysbio/sys038).
- Harrison RG. 1989.** Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology and Evolution* 4:6–11 DOI [10.1016/0169-5347\(89\)90006-2](https://doi.org/10.1016/0169-5347(89)90006-2).
- Hartel KD, Frederick BA, Shanower TG. 2003.** Isolation and characterization of microsatellite loci in wheat stem sawfly *Cephus cinctus* and cross-species amplification in related species. *Molecular Ecology Notes* 3:85–87 DOI [10.1046/j.1471-8286](https://doi.org/10.1046/j.1471-8286).
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004.** Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* 101:14812–14817 DOI [10.1073/pnas.0406166101](https://doi.org/10.1073/pnas.0406166101).
- Klopfstein S, Vilhelmsen L, Heraty JM, Sharkey M, Ronquist F. 2013.** The hymenopteran tree of life: evidence from protein-coding genes and objectively aligned ribosomal data. *PLOS ONE* 8(8):e69344 DOI [10.1371/journal.pone.0069344](https://doi.org/10.1371/journal.pone.0069344).
- Kopylov DS. 2014.** New sawflies of the subfamily Madygellinae (Hymenoptera, Xyelidae) from the Middle-Upper Triassic of Kyrgyzstan. *Paleontological Journal* 48:610–620 DOI [10.1134/s0031030114060070](https://doi.org/10.1134/s0031030114060070).
- Koressaar T, Remm M. 2007.** Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23:1289–1291 DOI [10.1093/bioinformatics/btm091](https://doi.org/10.1093/bioinformatics/btm091).
- Malm T, Nyman T. 2015.** Phylogeny of the symphytan grade of Hymenoptera: new pieces into the old jigsaw(fly) puzzle. *Cladistics* 31:1–17 DOI [10.1111/cla.12069](https://doi.org/10.1111/cla.12069).
- Miller MP. 2005.** Alleles In Space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity* 96:722–724 DOI [10.1093/jhered/esi119](https://doi.org/10.1093/jhered/esi119).
- Patten MM, Carioscia SA, Linnen CR. 2015.** Biased introgression of mitochondrial and nuclear genes: a comparison of diploid and haplodiploid systems. *Molecular Ecology* 24:5200–5210 DOI [10.1111/mec.13318](https://doi.org/10.1111/mec.13318).
- Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, Diez PA, Heraty J, Kjer KM, Klopfstein S, Meier R, Polidori C, Schmitt T, Liu S, Zhou X, Wappler T, Rust J, Misof B, Niehuis O. 2017.** Evolutionary History of the Hymenoptera. *Current Biology* 27:1013–1018 DOI [10.1016/j.cub.2017.01.027](https://doi.org/10.1016/j.cub.2017.01.027).
- 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](https://doi.org/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, Lee KM, Mutanen M. 2019.** Detection of cross-contamination and strong mitonuclear discordance in two species groups of sawfly genus *Empria*. *bioRxiv* DOI 10.1101/525626.
- 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.
- Reineke A, Assaf HA, Kulaneck D, Mori N, Pozzebon A, Duso C. 2015.** A novel set of microsatellite markers for the European Grapevine Moth *Lobesia botrana* isolated using next-generation sequencing and their utility for genetic characterization of populations from Europe and the Middle East. *Bulletin of Entomological Research* 105:408–416 DOI 10.1017/S0007485315000267.
- 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.
- Schoebel CN, Brodbeck S, Buehler D, Cornejo C, Gajurel J, Hartikainen H, Keller D, Leys M, Řičanová Š, Segelbacher G, Werth S, Csencsics D. 2013.** Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequencing. *Journal of Evolutionary Biology* 26:600–611 DOI 10.1111/jeb.12077.
- Sloan DB, Havird JC, Sharbrough J. 2017.** The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology* 26:2212–2236 DOI 10.1111/mec.13959.
- Smith DR. 1967.** A review of the larvae of Xyelidae, with notes on the family classification (Hymenoptera). *Annals of the Entomological Society of America* 60:376–384 DOI 10.1093/aesa/60.2.376.
- Tang P, Zhu J, Zheng B, Wei S, Sharkey M, Chen X, Vogler AP. 2019.** Mitochondrial phylogenomics of the Hymenoptera. *Molecular Phylogenetics and Evolution* 131:8–18 DOI 10.1016/j.ympev.2018.10.040.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012.** Primer3—new capabilities and interfaces. *Nucleic Acids Research* 40:1–12 DOI 10.1093/nar/gks596.