



OPEN

Development of microsatellite loci and optimization of a multiplex assay for *Latibulus argiolus* (Hymenoptera: Ichneumonidae), the specialized parasitoid of paper wasps

Agata Kostro-Ambroziak^{1✉}, Anna Siekiera², Magdalena Czajkowska³, Jan J. Pomorski² & Hanna Panagiotopoulou²

Microsatellite loci are commonly used markers in population genetic studies. In this study, we present 40 novel and polymorphic microsatellite loci elaborated for the ichneumonid parasitoid *Latibulus argiolus* (Rossi, 1790). Reaction condition optimisation procedures allowed 14 of these loci to be co-amplified in two PCRs and loaded in two multiplex panels onto a genetic analyser. The assay was tested on 197 individuals of *L. argiolus* originating from ten natural populations obtained from the host nests of paper wasps. The validated loci were polymorphic with high allele numbers ranging from eight to 27 (average 17.6 alleles per locus). Both observed and expected heterozygosity values were high, ranging between 0.75 and 0.92 for H_o (mean 0.83) and from 0.70 to 0.90 for H_e (mean 0.85). The optimized assay showed low genotyping error rate and negligible null allele frequency. The designed multiplex panels could be successfully applied in relatedness analyses and genetic variability studies of *L. argiolus* populations, which would be particularly interesting considering the coevolutionary context of this species with its social host.

The ichneumonid wasp *Latibulus argiolus* (Rossi, 1790) of the subfamily Cryptinae (Hymenoptera: Ichneumonidae) is a specialized ectoparasitoid of social paper wasps (*Polistes* Latreille, 1802) with respect to its larvae and pupae^{1,2}. Because parasitoids use only one host for their full larval development³, they are under strong natural selection that in turn leads to optimal exploitation of the host⁴. Therefore, the most important behaviour increasing the fitness of free-living female parasitoids is their oviposition decision, including the choice of the host and number of eggs laid on a specific host among other factors. A nest of paper wasps containing many potential hosts (various developmental stages of the wasp), can be attractive for the *L. argiolus* female allowing it to theoretically maximize its fitness. On the other hand, the strong defensive capabilities of *Polistes* colonies against intruders can influence the oviposition mode of *L. argiolus*, for example, by reducing the number of eggs laid in a single nest and in turn, by enlarging the number of host nests parasitized by this parasitoid. Furthermore, up to 34% to 78% of the host nests are destroyed by predators (such as birds and ants)⁵, that can lead to high *L. argiolus* mortality and lower its genetic variability. In turn, this process might cause a decrease in the fitness of this parasitoid and might function as an additional factor for females not to limit their reproduction to a single host nest. By linking the oviposition decision and genetic variability of *L. argiolus*, we can interpret the biology of this parasitoid in a broad ecological and evolutionary context. Investigations of the *L. argiolus* and paper wasp relationships may serve as a model for the coevolution of solitary parasitoid–social host interactions.

¹Laboratory of Insect Evolutionary Biology and Ecology, Faculty of Biology, University of Białystok, K. Ciołkowskiego 1J, 15-245 Białystok, Poland. ²Museum and Institute of Zoology Polish Academy of Sciences, Wilcza 64, 00-679 Warsaw, Poland. ³Division of Molecular Biology, Faculty of Biology, University of Białystok, K. Ciołkowskiego 1J, 15-245 Białystok, Poland. ✉email: ambro@uwb.edu.pl

The marker of choice for many areas of population research, such as mating systems, kinship structure, demography or conservation genetics, are microsatellite loci^{6–8}. The usefulness of these markers results from their codominance, high polymorphism, and abundance throughout the genome⁹. The high popularity of microsatellite loci, resulting in their frequent application in population genetics studies, is caused also by the beneficial proportion between the amount of information gained and the sum of financial costs and expended work efforts. Thus far, no such loci have been described for the *L. argiolus* species; therefore, the purpose of this study was to investigate these loci. Several methods for novel microsatellite marker development exist. One of them introduces third generation sequencing on a PacBio RSII platform. The described de novo loci must be verified first for reliable amplification and high polymorphisms. Microsatellites fulfilling the criteria of error free amplification, high genotyping quality, and polymorphisms may be applied for population genetic analysis. In studies applying microsatellites, at least over a dozen loci are required to obtain high statistical power of the calculations results^{10,11}. Therefore, concurrent amplification of markers in multiplex polymerase chain reactions (PCR) is desirable because this procedure will significantly shorten the time of laboratory work and reduce the cost of analysis. Such assays need to be first optimized and validated to justify the use of the loci in future research projects^{12–14}.

This study aimed to develop a set of novel microsatellite loci and then to optimize and verify the newly designed multiplexes that could be applied in population and evolutionary studies of the specialized ichneumonid parasitoid *L. argiolus* species.

Material and methods

Sampling and DNA extraction. The study material consisted of 213 *L. argiolus* specimens (126 diploid females and 87 haploid males). Specimens were obtained from 43 *Polistes* nests from ten wild populations in Poland: (1) Białystok-Lotnisko (53°05'59.5"N 23°09'37.6"E), (2) Hački (52°50'05.5"N 23°10'57.9"E), (3) Kalitnik (53°00'54.5"N 23°42'54.7"E), (4) Krzyżanowice (50°27'12.4"N 20°33'29.1"E), (5) Mietlica (52°32'28.0"N 018°22'57.3"E), (6) Przewóz (52°28'51"N 18°23'25"E), (7) Skupowo (52°49'51.0"N 23°42'55.9"E), (8) Suraż (52°55'19.3"N 22°57'49.9"E), (9) Stawiany (50°35'30.6"N 20°37'19.7"E), and (10) Zalesiany (53°04'10.4"N 23°02'32.9"E). The wasp nests collected in the field (in May/June) were transferred into plastic cages (24 × 17 × 16 cm) and kept in the laboratory at temperatures of 23 to 27 °C under natural light:dark (L:D) conditions. The individuals of *L. argiolus*, emerging from the host nest in June/July, were frozen and preserved at –20 °C. DNA was obtained from the hind legs using the DNeasy Blood & Tissue Kits (cat. no. 69506, Qiagen).

Pacific biosciences RS platform sequencing. The PacBio library was constructed using 9 µg of genomic DNA consisting of 16 (12.5 µl) equally mixed female DNA samples. DNA was fragmented using miniTube (cat. no. 520064) in the Covaris System to a size of 1.5 to 3 kb according to protocol with minor modifications in which the intensity was set to 0.2 while the sonication time was reduced to 40 s. The shared sample was purified applying 0.6 × volume ratio of AMPure PB Beads (cat. no. 100-265-900, Pacific Biosciences). Furthermore, the 2 kb library was prepared using 750 ng of DNA according to the Pacific Biosciences protocol available online¹⁵ and using the SMRTbell Template Prep Kit 1.0 (cat. no. 100-259-100, Pacific Biosciences). The procedure included DNA damage and end repair followed by blunt ligation of hairpin adaptors at both ends of the DNA fragments. Failed ligation products were removed with *ExoIII* and *ExoVII* enzymes. Purifications steps separating enzymatic reactions were performed by applying 0.6 × volume ratio of AMPure PB Beads. Size distribution of DNA fragments during the library preparation procedure and for the final library was checked by electrophoresis on 0.5% agarose gels. The resulting library was doubly purified and prepared for sequencing using DNA/Polymerase Binding Kit P6 v2 (cat. no. 100-372-700, Pacific Biosciences) and applying MagBeads Kit v2 (cat. no. 100-676-500, Pacific Biosciences) for loading onto the sequencer according to the protocol generated with Binding Calculator v.2.3.1.1 (available online: <https://github.com/PacificBiosciences/BindingCalculator>). Single molecule real-time testing (SMRT) was carried out on a PacBio RS II sequencer running two SMRT Cells 8Pac v3 (cat. no. 100-171-800, Pacific Biosciences) and a 360-min data collection mode.

Data gathered from sequencing were subject to the RS_ReadsoftInsert.1 protocol analysis with default settings to generate reads of insert (ROI). The ROI reads were subsequently subjected to microsatellite analysis and primer design using msatcommander v.1.08¹⁶ with a threshold of at least eight or ten repetitions for tri- and tetra-nucleotide repeats, excluding mononucleotide repeats. Primer design parameters including the msatcommander option “combine loci” consisted of several parameters: (1) size of primers ranging from 18 to 22 bp, (2) annealing temperature (T_m) ranging from 58 to 62 °C, (3) GC content ranging from 30 to 70%, and (4) amplicon product size range of 80 to 400 bp. Finally, only microsatellite sequences containing tri- or tetra-nucleotide motif with at least 10 repeats were used for marker selection and amplification test performance.

Microsatellite marker selection and testing. Loci with the highest number of tandem repeats were chosen from the identified microsatellite sequences set. The DNA sequences of the newly developed loci were aligned against the NCBI database of already described nucleotide sequences using BLAST algorithms (BLASTN 2.10.1 + and BLASTX 2.10.1 + programmes)^{17–19}. In the next step, loci were filtered for their optimal amplification performance by applying following criteria calculated in silico: (1) maximal PCR efficiency and no dimer formation, (2) difference in annealing temperature between primer pair below 2 °C, (3) minimal penalty score, and (4) sequence length of microsatellite ranging from 100 to 350 bp. For multiplex design purposes, the selected markers were further ordered according to the predicted PCR product size into three categories long (> 300 bp), medium (150–300 bp), and short (< 150 bp). The finally selected markers were amplified on DNA of 16 diploid females. Polymorphisms and the allele range of the chosen markers were tested using the universal primer labeling method^{20,21}. The forward primers were tailed at the 5' end with a CAC-sequence tag (5'-CACGACGTT

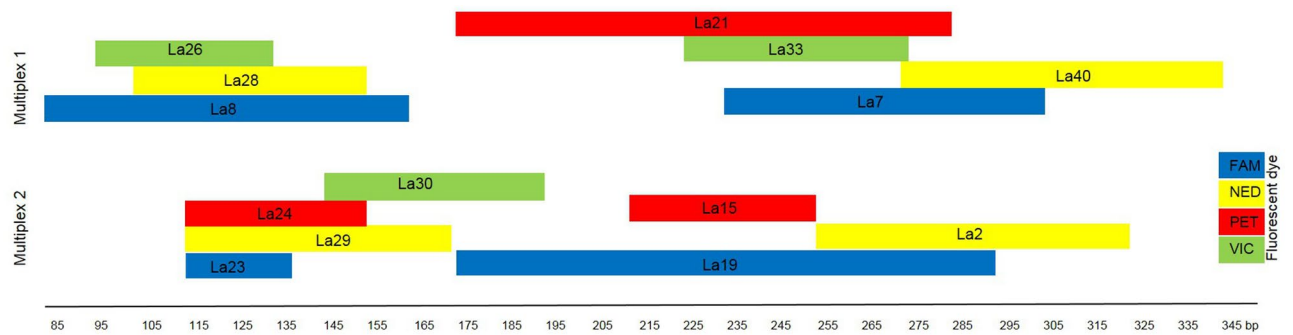


Figure 1. Two multiplex sets developed for simultaneous amplification of 14 microsatellite loci for *Latibulus argiolus*.

GTAAAACGAC). This tag sequence was identical to four fluorescent dye-labelled universal forward primers (labelled with the dyes, including 6-Fam, Hex, Tamra, and Rox). The PCR reactions were performed in a final volume of 15 μ l containing 9–227 ng of DNA (1 μ l), 1 \times MasterMix of DreamTaq Hot Start Green PCR Master Mix (cat. no. K9022, Thermo Fisher Scientific), 0.02 M tagged forward locus specific primer, 0.2 M locus specific reverse primer, and 0.2 M fluorescent labelled universal forward primers. The amplification conditions consisted of several steps: (1) initial denaturation 95 $^{\circ}$ C for 5 min, (2) 35 cycles of 30 s at 94 $^{\circ}$ C, (3) 90 s at 60 $^{\circ}$ C, (4) 40 s at 72 $^{\circ}$ C, and (5) a final elongation step of 10 min at 72 $^{\circ}$ C. Negative controls were included for each PCR reaction. Positive controls were not used as the DNA concentration and quality of all samples was satisfactory enough to permit successful PCR reactions. All loci were amplified separately, and the reaction efficiencies were visualized on 2% agarose gels. For loci in which no amplification products were observed, the amplifications were repeated by lowering the annealing temperature to 57 $^{\circ}$ C. PCR products were separated using an automated sequencer ABI 3500XL Genetic Analyzer applying GeneScan 600 LIZ as an internal lane size standard (cat. no. 4366589, Applied Biosystems). PCR products labelled with different fluorescent dyes were mixed for joint analysis. The resulting fragment sizes were read using GeneMapper v.4.1 (Applied Biosystems) software. Out of the tested set, only loci fulfilling following criteria were selected for multiplex panel optimization: (1) clear peak morphology, (2) no artefacts co-amplification, and (3) high number of alleles.

Multiplex panel optimization. Out of 41 de novo described loci, 14 microsatellite markers were selected and organized into two multiplex PCR sets (seven loci each) by maximizing the number of loci labelled with the same dye (avoiding though allele overlap between loci) as shown in Fig. 1. The gaps between the neighbouring loci were set to the size of 66 to 137 bp according to the genotype data set of 16 females obtained during single microsatellite marker testing. The forward primer of each marker was fluorescence-labelled (Table 1, Fig. 1). Multiplex PCRs were performed with Labcycler Gradient (SensoQuest GmbH) in 5 μ l reaction volume containing 2 μ l genomic DNA (~20 ng), 1.7 μ l Multiplex PCR 1 \times MasterMix (cat. no. 206143, Qiagen), 0.3 μ l mix of primers (cat. no. 450056, Life Technologies), and 1 μ l RNase-free water (cat. no. 206143, Qiagen). Table 1 shows the final concentration of each primer pair. For each PCR reaction, a negative control was used as a sample to which no DNA was added. Each multiplex PCR started with an initial activation step of 95 $^{\circ}$ C for 15 min followed by 22 cycles of denaturation at 94 $^{\circ}$ C for 30 s, annealing at 57 $^{\circ}$ C for 90 s, extension at 72 $^{\circ}$ C for 60 s, and ended with a final extension at 60 $^{\circ}$ C for 30 min. The PCR products were mixed with 10 μ l ultra-grade formamide (cat. no. 4311320, Applied Biosystems) and 0.2 μ l GeneScan 500 LIZ size standard (cat. no. 4322682, Applied Biosystems), denatured at 95 $^{\circ}$ C for 5 min, rapidly cooled, and then subjected to fragment length analysis by using a four-capillary ABI 3130 Genetic Analyzer (Applied Biosystems). The fragment lengths of microsatellite alleles were estimated automatically using the AutoBin feature in GeneMapper 4.0 software (Applied Biosystems), checked, and then optionally corrected manually to avoid false alleles calling of the software. In addition, approximately 10% of the samples (N = 20) was randomly chosen, re-amplified, and genotyped in order to calculate the genotyping error of the optimized multiplex.

Data analysis. The number of alleles (N_A) and their size ranges for each of the preliminarily tested locus were calculated with GenAlEx v.6.503 programme²². The software MICRO-CHECKER 2.2.1²³ was used for identifying possible genotyping errors (stuttering, large allele drop-out, false, and null allele frequencies) by performing 1,000 randomizations. We used CERVUS 3.0.3²⁴ to estimate the number of alleles per locus (N_A) and their size ranges. CERVUS 3.0.3 was also used to calculate the expected and observed heterozygosity (H_E and H_O , respectively), the mean polymorphic information content (PIC), and the null allele frequency values (F_{null}), and to estimate departures from Hardy–Weinberg equilibrium (HWE) with Bonferroni correction²⁴ for evaluating the significance of HWE deviations. Linkage disequilibrium between loci was tested using GENEPOP v4.0²⁵ with the Markov chain method (10,000 dememorization steps, 100 batches, 5,000 interactions) and Fisher’s exact test; a Bonferroni correction for multiple testing was then applied²⁶. The described calculations were performed exclusively for diploid females (N = 110) with exceptions for the N_A and size range values, which were conducted for the whole sample set (N = 197), including haploid males (N = 87). Because the analysed females came from populations, in which the number of females was on average only 7.2, calculations were performed for the pooled sample of 110 diploid individuals.

Locus/ GenBank	Repeat motif	Primer sequences	Multiplex (Mix)	Fluorescent dye	μ M	Size range	N_A	H_E	H_O	PIC	F_{null}	$F_{null\ micro}$	P_{HWE}
La7	ACG	(F) TTTGAT TCGCTCGTT GGGTG	LaI (1)	6-FAM	0.3	237–307	17	0.84	0.84	0.81	–0.003	–0.003	0.278
MN531314		(R) CGACGT AGCACAACC CTTC											
La8	ACG	(F) TCAATC GCGCGTACA AGAAC	LaI (1)	6-FAM	0.15	86–166	21	0.90	0.92	0.89	–0.014	–0.014	NE
MN531315		(R) GCGGTG ACTGTTCTT CCAC											
La21	ACG	(F) GCGCTC GGTAAGTAC GAAAC	LaI (1)	PET	0.375	176–280	27	0.84	0.79	0.83	0.034	0.033	0.084
MN531328		(R) ATCAGT TCGAAGCGT GTTGG					(176–234)	(25)					
La26	AAG	(F) GTTCCA CGGCGAAAG AAGAG	LaI (1)	VIC	0.05	95–134	14	0.83	0.82	0.81	0.004	0.004	0.860
MN531333		(R) AGATAT AACGGCGGG TCAGG						(13)					
La28	AGG	(F) GTACAA TCGGAAGCG GAACC	LaI (1)	NED	0.1	101–150	14	0.87	0.82	0.86	0.030	0.029	NE
MN531335		(R) ACTCTT CTCGTTCGT CCGTC					(110–150)	(12)					
La33	AAG	(F) GTTCGG AAAGCATGG GACTG	LaI (1)	VIC	0.275	220–279	18	0.89	0.87	0.87	0.007	0.007	NE
MN531340		(R) GGATTC GTTTCGCCA GTCC											
La40	ACG	(F) GAAGGA CAACAAAGG CGGAC	LaI (1)	NED	0.5	277–344	22	0.88	0.82	0.86	0.031	0.030	0.951
MN531347		(R) AGCGAG AGAGGTCA TTCGG											
La2	ACG	(F) CGTTGT GGATCGTGT TCTACC	LaII (2)	NED	0.8	260–324	18	0.87	0.85	0.85	0.013	0.013	0.930
MN531309		(R) AATTCT TGCTACCGA TCGCC											
La15	AGC	(F) TACTAC GAACGAGGT GAGGC	LaII (2)	PET	0.8	213–251	13	0.86	0.91	0.84	–0.033	–0.034	0.042
MN531322		(R) AGGAAG GAACGAGAG AAGGC						(12)					
La19	AAG	(F) GCGCGC GGAAATTC ATATC	LaII (2)	6-FAM	0.15	177–299	27	0.89	0.79	0.88	0.059	0.056	0.061
MN531326		(R) ACTCCC GACAGCTGA TTGAG						(23)					
La23	AAG	(F) CCCTTC GTCCGGATC TACAG	LaII (2)	6-FAM	0.1	114–136	8	0.70	0.76	0.65	–0.059	–0.064	0.124
MN531330		(R) ATATGG AGACTCTGC CGCTG											
La24	AAG	(F) ATCTTT ATCTCGGGA GGCCG	LaII (2)	PET	0.1	113–158	16	0.78	0.75	0.75	0.029	0.028	0.001*
MN531331		(R) TTAATG GGCAGAAGA CGGAG											

Continued

Locus/ GenBank	Repeat motif	Primer sequences	Multiplex (Mix)	Fluorescent dye	μ M	Size range	N_A	H_E	H_O	PIC	Fnull	Fnull micro	P_{HW}
La29	AAG	(F) GCAGGG CGTCTTTCT CAATC	LaII (2)	NED	0.1	118–177	17	0.90	0.84	0.89	0.036	0.035	NE
MN531336		(R) TCTTTG GGTACGAGA CGGAG											
La30	ACG	(F) CGGAAG TTCAAAGGT ACCGC	LaII (2)	VIC	0.1	146–193	15	0.80	0.78	0.77	0.004	0.004	0.593
MN531337		(R) CTTCCG ACAAATGCGC TGATC											
Overall							17.6 (16.9)	0.85	0.83	0.83			

Table 1. Characterization of 14 microsatellite loci developed for *Latibulus argiolus* and organized in two multiplex sets (1, 2). *GeneBank* GeneBank accession numbers; *Fluorescent dye* dye used for primers labeling; μ M optimized primer concentration; *Size range* observed size range in bp of the amplified alleles; N_A number of alleles; H_E expected heterozygosity; H_O observed heterozygosity; *PIC* mean polymorphic information content; *Fnull* null allele frequency (CERVUS 3.0.3); *Fnull micro* null allele frequency (MICRO-CHECKER 2.2.1; van Oosterhout values); P_{HW} Hardy–Weinberg probability test, *NE* not evaluated, *Statistical significance following sequential Bonferroni correction ($P < 0.001$). H_E , H_O , *PIC*, *Fnull*, P_{HW} , as well as *Size range* and N_A values in brackets were calculated exclusively for diploid females ($N = 110$).

Results

De novo sequencing of microsatellite loci. Library sequencing allowed us to obtain 1,133 Mb of raw data covering 65,556 reads for the jointly used SMRT Cells. The average sizes of the inserts were 2,241 and 2,464 (with quality of the inserts reads exceeding 0.92). The final output of reads length reached 151.4 Mb. Out of these sequences set, 48,515 ROIs were acquired and accumulated with 90,546,727 bp. The average ROI length equalled 1,874 bp, and the average number of passes was over 17 (17.74). The final sequence list consisted of 11,355 DNA sequences that were subjected to microsatellite loci searching. Sorting of microsatellites from this set allowed 415 sequences to be found that had an at least eight repetition threshold for which 401 primer pairs could be successfully designed. Altogether, 194 microsatellite loci covering tri- or tetra-nucleotide repeats at a number higher or equal to 10 were identified (Supplementary Table S1), for which 173 primers pairs could be designed (Supplementary Table S2). Only six tetra-nucleotide loci fulfilling the microsatellite selection criteria (out of these set) were detected (Supplementary Table S1). From this set, 41 tri-nucleotide loci, named LA1–LA41 were selected for PCR amplification and testing (Supplementary Table S3) and then deposited in GenBank under accession numbers MN531308–MN531348.

Out of 41 amplified markers tested in 16 diploid *L. argiolus* females, 40 (97.6%) were successfully amplified and were polymorphic (Supplementary Table S4). For four markers (LA1, 14, 16, and 38), the annealing temperature had to be lowered to 57 °C in order to obtain PCR products. Most of the markers (78%) amplified well and yielded clear and readable products. Twenty-four (60%) of the markers gave single peaks, while in the others 16 (40%), additional loci peaks impeding potential uncertainty in allele scoring were observed. These peaks consisted of stutter bands or after-peaks. In seven cases, two peaks differing by one bp in length were observed. The latter effect is caused in cases in which polymerase does not add adenine to the 3' end of the newly synthesised strand during replication (resulting in so called – A and + A bands occurrence)²⁷. BLASTX analysis showed no results for neither of sequence of the 41 tested loci. BLASTN results indicated that eight DNA sequences were similar to previously described DNA or cDNA sequences deposited in GenBank (Supplementary Table S5). However, in all results, the obtained query cover values were too low for convenient sequence annotation to the already described in GenBank genes or transcripts. This indicates that the selected loci are probably located in the non-coding regions of the *L. argiolus* genome. The exception could be locus La36 for which the query cover values exceeded 50% together with low E value scores and relatively high percentage identity of the aligned sequences for several insect species. In most of these cases BLAST search identified transcription factors sequences as the most probable results. The observed allele number in the tested loci ranged from 4 to 18 with 25 loci exhibiting at least 10 alleles. The best amplified, readable, and most polymorphic loci (Supplementary Table S4) were considered for multiplex development. LA11 was excluded from this set since it exhibited difficulties in raw read rounding, which could have been caused by its hypervariability ($N_A = 18$).

Development and characterization of the microsatellite multiplex assay. For multiplex optimization purposes, 14 loci were selected and organized into two multiplex panels (Table 1, Fig. 1). The combined microsatellite primers pairs in these assays successfully amplified the chosen loci to yield peaks that were clear and easy to score. For most of the loci, the initial primer concentration had to be adjusted to achieve balanced amplification of the markers (Table 1). We did not detect any evidence of genotypic errors using the MICRO-CHECKER 2.2.1 software, a finding that was also proven by the low frequencies of null alleles explored by CERVUS 3.0.3 (Table 1). The amplification success of the multiplexes was high, reaching 197 (100%) of the microsatellite profiles. Robustness of the optimized assay was further verified by obtaining 100% correctly genotyped samples for 20 duplicated samples (~10% of the studied samples set as recommend by Pompanon et al.²⁷). No

significant linkage disequilibrium among any pair of the loci under study after Bonferroni correction (adjusted value of $P < 0.004$) was found, indicating that the studied loci were most probably not linked. Deviations from Hardy–Weinberg equilibrium were significant ($P < 0.001$) for only one (La24) of the 14 loci tested after Bonferroni correction (Table 1).

The average number of alleles (N_A) per locus was 17.6, ranging from 8 in locus La23 to 27 in loci La19 and 21. The set of diploid female samples was characterized by the highly observed heterozygosity (mean $H_O = 0.83$) ranging from 0.75 in locus La24 to 0.92 in locus La8 in addition to the highly expected heterozygosity (mean $H_E = 0.85$) ranging from 0.70 in locus La23 to 0.90 in loci La8 and 29. The mean polymorphic information content (PIC) ranged from 0.65 in locus La23 to 0.89 in loci La8 and 29 (Table 1).

Discussion

In this study, we successfully describe new loci and optimization of a multiplex assay for *L. argiolus*, the highly specialized parasitoid of social paper wasps. PacBio RSII platform sequencing allowed for reliable discovery of microsatellite markers. This method has been proven to perform very well in de novo microsatellite description and was recently often applied^{28–30}. Application of this method is especially useful in non-model organisms in which no information about genomic data exists. In such cases, it also outperforms other methods by being more efficient and accurate^{30,31}. This method is also relatively cost and time efficient compared to the classical methods of microsatellite development^{9,28}. Out of 11,355 DNA sequences, approximately 1.7% contained tri- or tetra-nucleotide motifs exceeding ten repeats. Tri- and tetra-nucleotide loci were selected as they were proven to cause less difficulties in scoring and rounding³². This finding is important, especially since insects are considered problematic for microsatellite isolation and genotyping and are often biased with particularly high frequencies of null alleles^{9,33,34}. Tri- and tetra-nucleotide loci have been suggested to be less frequent than di-nucleotide in the genome; however, they are less prone to amplification errors, especially stuttering²⁹. For this reason, primers and in silico amplifications were very carefully chosen. Sequences with high repeat numbers were selected as it was proven that such loci have higher polymorphism due to higher mutation rate caused by polymerase slippage³⁵.

The amplification success of the designed markers was very high (97.6%), and most of the loci amplified well, which points to the high quality of the RSII sequencing and well-defined parameters for primers design^{28–31}. The relatively long reads (on average, exceeding 2 kb) facilitated the design of primers that enabled high PCR performance²⁸.

The approach of applying universal primer labelling²⁰ has been reported to be very efficient and cost effective in other studies^{21,32,36}. In our current study, it was also proven to be a very useful and suitable method for primer labelling. The selected markers amplified well, and in most cases, produced (78%) clear bands and in 60% produced single peaks that permitted confident reading. Almost all of the tested markers were polymorphic, which may have been the consequence of selecting sequences with more than ten repeats. The rate of polymorphic loci in conventional methods of de novo microsatellite isolation may be as low as 30%^{30,35}. This finding indicates that appropriate microsatellite sequences were selected for amplification tests in single specimens. This finding may also point to a high intrapopulation polymorphism of the studied species. However, a trade-off between high variability and possible artefacts may exist. The extremely polymorphic markers may be biased with high null allele frequency as a consequence of high a mutation rate in these loci^{9,27,33}. This bias could have occurred in the LA11 locus. For multiplex optimization, direct fluorescent labelling was applied to reduce primer-dimer formation and enhance the PCR efficiency³², which was 100% in the studied sample set.

Our data show that the 14 microsatellite loci selected for assay development are reliable markers for genetic analyses of *L. argiolus* individuals with a good quality of detection, high polymorphism, and low frequency of null alleles. It is true that one of the loci was not in HWE, but it should be emphasized that the HWE test was performed for pooled diploid individuals that came from different populations resulting most probably in Wahlund effect. We believe that the presented tool will be very useful in *L. argiolus* studies, especially in the context of its life history strategies after considering co-evolution of this haplodiploid parasitoid with its social host. Because oviposition decisions of female parasitoids primarily influence their fitness, we hypothesize that female of *L. argiolus* lays more eggs in larger nests of the paper wasp and exhibits adaptive sex ratio manipulation in response to the number of her offspring produced in a single nest of the host. The previously mentioned hypotheses could be now addressed by applying the developed marker set that would allow determination of kinship and could also be a very good tool for the rapid identification of sex of preimaginal stages of *L. argiolus*. We believe that application of microsatellite markers will yield many answers concerning the biology and genetic structure of this specialized parasitoid. Additionally, the deposited sequence data may be subject to further microsatellite searches and testing in other closely related species of this interesting group of parasitoids.

Data availability

All data generated during this study are included in this published article and its supplementary information files.

Received: 30 January 2020; Accepted: 9 September 2020

Published online: 30 September 2020

References

1. Makino, S. Biology of *Latibulus argiolus* (Hymenoptera, Ichneumonidae), a Parasitoid of the Paper Wasp *Polistes biglumis* (Hymenoptera, Vespidae). *Kontyû* **51**, 426–434 (1983).
2. Oh, S.-H., An, S.-L. & Lee, J.-W. Review of Korean *Latibulus* (Hymenoptera: Ichneumonidae: Cryptinae) and a key to the world species. *Can. Entomol.* **144**, 509–525 (2012).
3. Quicke, D. L. J. *The Braconid and Ichneumonid Parasitoid Wasps: Biology Systematics, Evolution and Ecology* (Wiley Blackwell, Amsterdam, 2015).

4. Paukku, S. & Kotiaho, J. S. Female oviposition decisions and their impact on progeny life-history traits. *J. Insect Behav.* **21**, 505–520 (2008).
5. Rusina, L. Y. The role of parasitoids in regulation of *Polistes* Wasp population (Hymenoptera, Vespidae: Polistinae). *Entomol. Rev.* **93**, 271–280 (2012).
6. Coelho, N. H. P. *et al.* Understanding genetic diversity, spatial genetic structure, and mating system through microsatellite markers for the conservation and sustainable use of *Acrocomia aculeata* (Jacq.) Lodd. Ex Mart. *Conserv. Genet.* **19**, 879–891 (2018).
7. Manlik, O. *et al.* Demography and genetics suggest reversal of dolphin source-sink dynamics, with implications for conservation. *Mar. Mammal Sci.* **35**, 732–759 (2019).
8. Nowicki, P. *et al.* What keeps “living dead” alive: demography of a small and isolated population of *Maculinea (=Phengaris) alcon*. *J. Insect Conserv.* **23**, 201–210 (2019).
9. Selkoe, K. A. & Toonen, R. J. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Lett.* **9**, 615–629 (2006).
10. Olafsson, K., Hjorleifsdottir, S., Pampoulie, C., Hreggvidsson, G. O. & Gudjonsson, S. Novel set of multiplex assays (SalPrint15) for efficient analysis of 15 microsatellite loci of contemporary samples of the Atlantic salmon (*Salmo salar*). *Mol. Ecol. Resour.* **10**, 533–537 (2010).
11. Randi, E. *et al.* Multilocus detection of wolf x dog hybridization in Italy, and guidelines for marker selection. *PLoS ONE* **9**, e86409 (2014).
12. Hung, C.-M., Yu, A.-Y., Lai, Y.-T. & Shaner, P.-J.L. Developing informative microsatellite markers for nonmodel species using reference mapping against a model species’ genome. *Sci. Rep.* **6**, 23087 (2016).
13. Marcus, T., Assmann, T., Durka, W. & Drees, C. A suite of multiplexed microsatellite loci for the ground beetle *Abax parallelepipedus* (Piller and Mitterpacher, 1783) (Coleoptera, Carabidae). *Conserv. Genet. Resour.* **5**, 1151–1156 (2013).
14. Panagiotopoulou, H., Baca, M., Baca, K., Stanković, A. & Žmihorski, M. Optimization and validation of a multiplex assay for microsatellite loci analysis in the field cricket, *Gryllus campestris* (Orthoptera: Gryllidae). *J. Asia-Pac. Entomol.* **18**, 421–424 (2015).
15. PacBio Pacific Biosciences, Procedure & Checklist: 2 kb Template Preparation and Sequencing. <https://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-2-kb-Template-Preparation-and-Sequencing.pdf>.
16. Faircloth, B. C. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol. Ecol. Resour.* **8**, 92–94 (2008).
17. Zhang, Z., Schwartz, S., Wagner, L. & Webb, M. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214 (2000).
18. Morgulis, A. *et al.* Database indexing for production MegaBLAST searches. *Bioinformatics* **24**, 1757–1764 (2008).
19. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
20. Schuelke, M. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* **18**, 233–234 (2000).
21. Austin, J. D. *et al.* Permanent genetic resources added to Molecular Ecology Resources Database 1 February 2011–31 March 2011. *Mol. Ecol. Resour.* **11**, 757–758 (2011).
22. Peakall, R. & Smouse, P. E. GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537–2539 (2012).
23. Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**, 535–538 (2004).
24. Kalinowski, S. T., Taper, M. L. & Marshall, T. C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* **16**, 1099–1106 (2007).
25. Rousset, F. GENEPOP’007: a complete reimplementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106 (2008).
26. Rice, W. R. Analyzing tables of statistical tests. *Evolution* **43**, 223–225 (1989).
27. Pompanon, F., Bonin, A., Bellemain, E. & Taberlet, P. Genotyping errors: causes, consequences and solutions. *Nat. Rev. Genet.* **6**, 846–847 (2005).
28. Grohme, M. A., Soler, R. F., Wink, M. & Fröhme, M. Microsatellite marker discovery using single molecule real-time circular consensus sequencing on the Pacific Biosciences RS. *Biotechniques* **55**, 253–256 (2013).
29. Liljegren, M. M., de Muinck, E. J. & Trosvik, P. Microsatellite length scoring by single molecule real time sequencing—effects of sequence structure and PCR regime. *PLoS ONE* **11**, e0159232 (2016).
30. Dutta, N. *et al.* Microsatellite marker set for genetic diversity assessment of primitive *Chitala chitala* (Hamilton, 1822) derived through SMRT sequencing technology. *Mol. Biol. Rep.* **46**, 41–49 (2018).
31. Wei, N., Bemmels, J. B. & Dick, C. W. The effects of read length, quality and quantity on microsatellite discovery and primer development: from Illumina to Pac Bio. *Mol. Ecol. Resour.* **14**, 953–965 (2014).
32. Corner, S., Yuzbasiyan-Gurkan, V., Agnew, D. & Venta, P. J. Development of a 12-plex of new microsatellite markers using a novel universal primer method to evaluate the genetic diversity of jaguars (*Panthera onca*) from North American zoological institutions. *Conservation Genet. Resour.* **11**, 487–497 (2019).
33. Chapuis, M.-P. & Estoup, A. Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* **24**, 621–631 (2007).
34. Blondin, L. *et al.* Characterization and comparison of microsatellite markers derived from genomic and expressed libraries for the desert locust. *J. Appl. Entomol.* **137**, 673–683 (2013).
35. Chistiakov, D. A., Hellems, B. & Volckaert, F. A. Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture* **255**, 1–29 (2006).
36. Cheng, L., Zhang, Y., Lu, C.-Y., Li, C. & Sun, X.-W. Development and characterization of four moderate multiplex microsatellite panels in crucian carp (*Carassius auratus*). *Conserv. Genet. Resour.* **5**, 821–823 (2013).

Acknowledgements

We kindly thank Aleksandra Kozłowska for her laboratory assistance. The study was carried out through financing agreement with the National Science Centre by Grant 2017/01/X/NZ8/01081.

Author contributions

A.K.-A. carried out the field work; A.K.-A., H.P., M.C. and A.S. conducted the laboratory work; J.P. performed bioinformatics analysis; A.K.-A., H.P. and M.C. drafted the article. All authors contributed critically to the drafts and gave final approval for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-72923-6>.

Correspondence and requests for materials should be addressed to A.K.-A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020