Journal of Genetics (2021)100:67
https://doi.org/10.1007/s12041-021-01318-x

© Indian Academy of Sciences

RESEARCH ARTICLE

Development and preliminary application of novel genomewide SSR markers for genetic diversity analysis of an economically important bio-control agent *Platygaster robiniae* (Hymenoptera: Platygastridae)

J. YANG[1,2], L. HUANG[1], Z. R. LI[1], H. Q. SUN[1], W. X. ZHAO[1] and Y. X. YAO[1]*

[1]*Key Laboratory of Forest Protection of National Forestry and Grassland Administration, Research Institute of Forest Ecology, Environment and Protection, Chinese Academy of Forestry, Beijing 100091, People's Republic of China*
[2]*United Pesticide Industry Cropscience Co. Ltd., Jinan 300384, People's Republic of China*
*For correspondence. E-mail: yaoyx@caf.ac.cn.

Received 8 February 2021; revised 6 April 2021; accepted 27 May 2021

Abstract. *Platygaster robiniae* Buhl and Duso (Hymenoptera: Platygastridae) is an egg-larvae parasitoid of the black locust gall midge (*Obolodiplosis robiniae*) (Haldeman) (Diptera: Cecidomyiidae) which is a serious invasive pest in China, where it attacks an important hardwood species, the black locust tree, *Robini pseudoacacia* L. (Fabales: Fabaceae). Despite the use of *P. robiniae* as an effective bio-control agent, the absence of sequence data and other molecular markers have limited its genetic applications for pest management in forests. Simple-sequence repeats (SSRs) are valuable molecular markers for population genetic structure studies. In the present study, we identified 14,123 SSRs, of which 7799 SSR primer pairs were successfully designed. Subsequently, 240 SSR were chosen and tested with 48 *P. robiniae* accessions from two geographically separated populations in north and south China. Of these, 34 were polymorphic, with an average of three alleles (N_a) and four genotypes (N_G) each. The average values of observed heterozygosity (H_o) was 0.3514, expected heterozygosity (H_e) 0.4167, Shannon's information index (I) 0.7143, and polymorphism information content (PIC) 0.3558, respectively. Neighbour joining analysis (bootstrap 1000) revealed that Chengdu (CD) and Dangdong (DD) popluations clustered into two main divisions, and some individuals from two popluations clustered together as the third devision, which indicated the gene flow and genetic differentiation were present between two populations. Our finding indicates that these SSR markers will be useful for further studies on the genotype identification and genetic mapping of the genus *Platygaster*.

Keywords. simple-sequence repeat markers; genome sequences; resource for population genetic; *Platygaster robiniae*.

Introduction

The black locust gall midge, *Obolodiplosis robiniae* (Haldeman) (Diptera: Cecidomyiidae) is a North American species that has recently been extensively introduced throughout Asia and Europe (Yao *et al.* 2020), and has been considered a highly invasive pest insect in the new regions in the last decades (Yao *et al.* 2015). Its main host is the black locust tree *(Robinia pseudoacacia*) although it is occasionally found on *R. pseudoacacia* cv. 'Frisia' (Badmin 2016). The black locust tree which originated from the eastern United States was introduced into Europe and Asia (including China) for ornamental purposes, and now it has become economically and ecologically important type of afforestation tree due to its good characteristics, such as fast growth, strong adaptability, tolerate drought, salty and poor soil, etc (Yao *et al.* 2020). The gall midge causes leaf rolling and premature leaf shedding, resulting in the deterioration of host and increased susceptibility to other pests including wood borers such as longhorn beetles.

Platygaster robiniae Buhl and Duso (Hymenoptera: Platygastridae), a critical natural enemy of *O. robiniae* was

J. Yang and L. Huang contributed equally to this work.

Supplementary Information: The online version contains supplementary material available at https://doi.org/10.1007/s12041-021-01318-x.

Published online: 18 September 2021

first reported and named by Buhl and Duso in 2008 based on specimen materials from Italy, Japan, and the Czech Republic, and the hosts of those materials were all *O. robiniae* on *R. pseudoacacia* (Leguminosae) (Buhl and Duso 2008). Recently, the wasp has been found in other European and Asian countries, including Switzerland, Serbia, Croatia, Denmark, Slovakia, Roma, Ukraine, Bulgaria, and South Korea (Ljubodrag *et al.* 2008; Jørgensen 2009; Park *et al.* 2009; Pernek and Matošević 2009; Sviridov and Bazhenova 2009; Bálint *et al.* 2010; Tóth *et al.* 2011; Tomov 2014). In China, it was first reported in Qinhuangdao city, Hebei Province in 2010 (Lu *et al.* 2010), and later was found in 17 other provinces at 29 sites (Yang *et al.* 2019).

The parasitoid is a gregarious, koinobiont endoparasitoid; lays egg in *O. robiniae* egg and both the larvae develop to maturity simultaneously (Kim *et al.* 2011). When the host larvae mature, the wasp kills the host, consumes all its organs and nutrients, and pupates in it. At this time, the pupae of the wasp, arranged in a tight spiral, can be seen through the host epidermis (Buhl and Duso, 2008; Yang *et al.* 2019). Although the percentage of parasitism on *O. robiniae* differs in each new region (Pernek and Matošević, 2009; Lu *et al.* 2010; Tóth *et al.* 2011), *P. robiniae* is always the predominant natural enemy of *O. robiniae* (Duso *et al.* 2005; Park *et al.* 2009; Pernek and Matošević, 2009; Lu *et al.* 2010). For the parasitism frequency increasing along with generation development of *O. robiniae*, the differences across regions probably were due to the generation difference when they were investigated. Since there is no data on wasp, although *P. robiniae* was found in European and Asian countries and regions where the host *O. robiniae* was introduced (Yang *et al.* 2019), we are not sure whether the parasitoid was introduced from North America along with its host and spread gradually.

Knowledge of the population genetic diversity and structure of insects would be of great assistance to understand the history of occurrence in the areas recently colonized by their hosts, explore the mechanism of population colonization, and assess the successful invasion of an alien organism (Sakai *et al.* 2001; Lee *et al.* 2017; Zalewski *et al.* 2010; Amouroux *et al.* 2013; Kirk *et al.* 2013; Terhorst and Lau 2015). SSR markers and repeated nucleotide sequence motifs have well known characteristics of codominant inheritance, high polymorphism, variability, and suitability for cross-species transfer ability and are used in population genetics and ecological studies of many groups (Cavagnaro *et al.* 2010; Terhorst and Lau 2015; Linløkken 2018; Simonato *et al.* 2019). Thus, they are valuable markers to study the intraspecific or interspecific variation and genetic structure of a biological population (Yao *et al.* 2020), which helped in tracing back to the origin, especially for the introduction of natural enemies of the invasive pest insect. Next-generation sequencing (NGS) enables the creation of massive genomic sequences that provide opportunities to develop SSR markers (Li *et al.* 2010; Bentur *et al.* 2011; Guichoux *et al.* 2011; Cerna and Straka 2013; Gil *et al.* 2017; Lee *et al.* 2017).

Because of the scarcity of molecular markers, there is little information on the genetic relationships and diversity of the *P. robiniae* population. We have developed 34 polymorphic SSR markers from several *P. robiniae* accessions by NGS and bioinformatics approaches. The findings from such a study should provide useful information for studies of population genetic diversity and structural exploration studies of the parasitoid. It also provides information support for further prevention and control of *O. robiniae*.

Materials and methods

Sample collection and DNA extraction

As *O. robiniae* is parasitized by *P. robiniae* at a very high rate in the late summer, the rolled leaves of host tree *R. pseudoacacia* provide a good source for the parasitoid. The rolled leaves of *R. pseudoacacia* were collected from three sites (Beijing, 40°21.768′N, 116°34.806′E; Chengdu, 30°48.996′N, 103°51.816′E; Dandong, 40°07.794′N, 124°22.542′E) (table 1 in electronic supplementary material at http://www.ias.ac.in/jgenet/) in the late summer and placed in large jars (L × H; 30 × 14 cm), sealed with a piece of gauze, and bound with a rubber band. The jars were maintained at 25°C and 70% humidity. When the adults of *P. robiniae* emerged, they were collected carefully with a brush wetted with ethanol and transferred to a vial with 100% ethanol. DNA extraction was conducted with the TIAN amp Genomic DNA kit (Tiangen Col., Beijing) on 15 individuals from Beijing site. For other two sites, Dandong (Liaoning province) is probably one of the earliest introduction sites for *O. robiniae*, according to the result of our former studies of the host, and represented north Chinese population, while Chengdu (Sichuan province) is relatively far away from the site and represented south Chinese population.

Genomic sequencing and sequence assembly

Sequencing libraries for DNA samples were prepared with a TruSeq Nano DNA Sample Prep kit, according to the manufacturer's instructions (Illumina, San Diego, USA). The libraries were subjected to paired-end sequencing with a 150 bp read length on the Illumina HiSeq 2500 platform (Illumina). For each library, 4 Gb of sequence data were obtained after removing reads that contained adaptor contamination poly-Ns and low-quality bases (quality score < 20, more than 10% uncalled based ('N' characters), and the duplicated sequences) with the software Adapter Removal v. 2.2 (Schubert *et al.* 2016). The clean short reads were assembled with SOAP *de novo* v. 2 (Luo *et al.* 2012) and an adjusted k-mer size of 41. Gap filling and base

Table 1. Characteristics of 34 SSR markers for *P. robiniae.*

Locus	GenBank number	Motif	Primer sequence (5′–3′)	T_a (°C)	Allele size
P-LH550	MT407333	(TA)12	F:GAAATTGATCACTCGGAGCC R:AAGTTGGGCCATTGTAGGTG	55	338–346
P-LH620	MT407334	(TC)19	F:CCACAATCATGCTGAAAACG R:GAAAGCGCAGGTCGATACTC	55	287–385
P-LH1099	MT407335	(AAT)12	F:ATTTGGCCATTTTCGTGTGT R:GGATGCTAAAATCGAGCCTG	55	295–328
P-LH1431	MT407337	(CTT)12	F:TCTCGGTCACCTGCTTCTCT R:CAACGGAGGTCAAGATGGTT	55	176–239
P-LH6835	MT407338	(CCGTTG)8	F:CATTTTCCCTAGTCGTTGCC R:TCCCGCGCTTTTCTATAATG	55	114–250
P-LH8765	MT407339	(CTA)12	F:TACCCGCAGTCAACACAGAG R:GCCTTGGGAAATTGGAAAAT	55	204–217
P-LH17682	MT407340	(GAG)13	F:TTGAAGCACACAAGGTGGAG R:GCGAGTTGGTCAGCTCTTTC	55	285–288
P-LH18283	MT407341	(ATC)12	F:CAACCCTGGCAAACATTCTT R:ATACACCGCCCCTCTTTCTT	55	378–381
P-LH20524	MT407342	(ATG)13	F:ATCGTCGAAATGGATTCAGC R:TGCAACAATCATTCAGCCAT	55	419–442
P-LH23400	MT407343	(AT)12	F:TCGATCCAACATTCTCCACTC R:ACTCTAGTCAGGCCCTTGCC	55	224–232
P-LH28311	MT407344	(GCT)13	F:CGAGGAGTTGAACGAGGAAG R:CCGCGGAATATTGTACCACT	54	244–271
P-LH34007	MT407345	(TTC)12	F:ACCGTATCAGAATGGCTTGC R:GACGTTTGATCTACCTGGGG	55	263–279
P-LH44376	MT407346	(AT)12	F:GAAGACGTCGAAGAAGCCAC R:AAAATCACGCAGAAAAAACGG	55	191–205
P-LH45531	MT407336	(AGC)13	F:TCGAGACGATTCTGGAGGAC R:GCACAGCGTTTCAGTGTCAT	55	219–231
P-LH50591	MT407347	(TC)19	F:TCTTTCATCCACCCGATCTC R:TGTCACAGCAGTACGCCTTC	55	345–347
P-LH54936	MT407348	(AAG)13	F:CGACGAGTCGGAGAGTTTTC R:TACTTTTGTCCTTGGTCGGG	55	148–297
P-LH61231	MT407349	(GGA)13	F:ACGACGATAATGGGCATGTT R:AACGTCTATGACACGAGGGG	55	317–332
P-LH63418	MT407350	(TTG)13	F:GGAATTTCGGGCTTTTCTTC R:TTGCTTCAGCCACAACAGTC	55	276–282
P-LH64210	MT407351	(AC)18	F:TCACCATGTTCCTGGGAAGT R:TGACCTTTTGTGATTTCGTGTC	55	167–243
P-LH74065	MT407352	(GTT)13	F:TTTTGCACGAAAATGCTCAG R:ACAAGCCCAACAACCTATGC	55	208–224
P-LH82949	MT407353	(TAT)12	F:TCGGCAGAATTTCGTGATTT R:CGAATTTGGTGTGTCAATGC	55	242–383
P-LH87349	MT407354	(TC)19	F:AACGCCGCACCACACTATAC R:CCCAAAGCGACAACGTTTAT	55	147–308
P-LH123303	MT407355	(TTA)12	F:TGGTCATTTTCCCTCGTTTC R:CGAAGCATTCCATTGTTGTG	55	147–390
P-LH129182	MT407356	(AAT)13	F:TTTTGCAGCGTGCTTGTAAC R:CATGACGCAGGGTATTTGTG	55	191–204
P-LH133174	MT407357	(AC)17	F:CGAGAAATACGTTTCACGCC R:CTATGGCCCATTGAAAAACG	55	208–214
P-LH150918	MT407358	(CGT)12	F:CGGGTACAACTGCTGGATTT R:AGGGGTAACTTTTCGACGCT	55	228–231
P-LH158987	MT407359	(AAG)13	F:ATAGCGGCATCATTTTCACC R:AAAGGGCGCTTCTTCTTTTC	55	160–238
P-LH165567	MT407360	(ACG)13	F:CCACCGTTCTTGCTCAATG R:ACGTAATTTCAGCCACGTCC	55	170–202
P-LH185124	MT407361	(TTA)12	F:GCCATGATTGTTGTTATTTTCG R:GAGGTGGAGGTAGAGAGGGG	55	358–408
P-LH206646	MT407362	(ATT)12	F:TGCTTCTTCCAAATTCCAGG R:GCCATGGGCAAATATACGAT	55	314–330
P-LH246376	MT407363	(TC)19	F:CGAGGAAACAGCAGGTAAGG R:ACGTCATGCAATGCTGAAAG	54	248–250

Table 1 (*contd*)

Locus	GenBank number	Motif	Primer sequence (5′–3′)	T_a (°C)	Allele size
P-LH269819	MT407364	(AAG)13	F:CTTCCGGCGTACTCGTTTC R:GGCTAGAGCCAAGCTGTCTG	55	242–247
P-LH307411	MT407365	(GGTTC)6	F:TTAAACGCTGCCGAAAGACT R:GAACCGAAACGAACTGAACC	55	172–211
P-LH375606	MT407366	(AC)18	F:AATCCCTATGTCACGCGAAG R:TCGCCTAAATAACTGCCACC	55	203–240

correction were screened in the assembled contigs using GapCloser v. 1.12.

SSR findings and primer designs

SSRs were identified using the microsatellite identification tool (Beier *et al.* 2017) with the following parameters: (i) microsatellites consisting of tandem repeats of 2–6 bp with a minimum of 5; (ii) a minimum space between two SSRs of 100 bp; and (iii) no variation (mutation) in repeat motifs. Primers were designed from flanking sequences of SSR microsatellite loci with Primer3 v. 2.3.7 (Rozen and Skaletsky 1999) with the following parameters: primer length = 19–22 bp with 20 bp as the optimum; GC % = 40–60% with 52% as the optimum; annealing temperature (T_a) = 55–65°C with optimum at 55°C; and product size range = 150–500 bp. Forward primers were labelled with a virtual dye TAMRA, ROX, HEX, and FAM (Applied Biosystems).

PCR amplification and scanning

To each marker we added an M13 tail (5′-TGTAAAAC-GACGGCCAGT-3′) sequence at the 5′-end of both the forward or reverse primers. PCR reactions were performed with 0.5 ng genomic DNA in each 10 μL reaction. In addition to template DNA, PCR reaction mixtures contained 5 μL of 2×Taq Mix, 0.05 μL of 10 μM forward primer, 0.15 μL of revised primer, 0.1 μL of TP-M13 (5 mol/L) primer joint, 0.5 μL of DNA template, and 4.2 μL of double distilled H_2O. The thermal cycling conditions were as follows: a 5 min initial denaturation at 95°C; 35 cycles of 30 s denaturation at 95°C, 30 s of annealing at 55°C, a 40 s extension at 72°C; and a 10 s final extension at 72°C. The PCR products were separated according to different fragment sizes with specific sequencing glue and scanned on the 3730xl DNA analyzer gene analyser (ABI), then detected using the standard, GeneScan-500 Orange. Subsequently, the data were analysed with GeneMarker 2.2.0 (SoftGenetics, LLC, USA) and GenePop software v. 4.3 (François 2008). We used GenAlEx6.502 (Peakall and Smouse 2010) and PowerMarker v. 3.25 software (Liu and Muse 2005) to construct the NJ dendrogram.

Results

Characteristics of genomic SSR in the NGS assemblies of P. robiniae

The libraries were constructed from 20 *P. robiniae* individuals by means of paired-end sequencing with a 150 bp read length on the Illumina HiSeq 2500 platform (Illumina). The SOAPdenovo2 assembler (v. 2.04) was used to assemble the

Table 2. The major frequent alleles of the polymorphic SSR markers amplified in two populations of *P. robiniae*.

Locus	Allele	CD	DD
P-LH550	356	0.8958	0.9583
P-LH620	385	0.4583	0.3958
P-LH1099	313	0.8333	0.9583
P-LH1431	194	0.5000	0.6042
P-LH6835	157	0.5417	0.5000
P-LH8765	232	1.0000	0.9583
P-LH17682	306	0.6458	0.6250
P-LH18283	399	0.7500	0.7292
P-LH20524	439	0.8333	0.6457 (460)*
P-LH23400	242	0.5000 (248)	0.6042
P-LH28311	262	0.813	0.771
P-LH34007	297	0.9583	1.0000
P-LH44376	209	0.9583	0.4167
P-LH45531	240	0.9167	1.0000
P-LH50591	365	0.6250	0.6250 (363)*
P-LH54936	166	0.5417	0.5417 (315)*
P-LH61231	350	0.9167	0.9583
P-LH63418	300	0.9375	0.9792
P-LH64210	236	0.4375	0.5208 (195)*
P-LH74065	229	0.7917	0.7292
P-LH82949	260	0.7292	0.8750
P-LH87349	329	0.5208	0.6875 (165)*
P-LH123303	384	0.3542	0.3542
P-LH129182	209	0.4583	0.9583 (222)*
P-LH133174	228	0.6250	0.6250
P-LH150918	249	0.9583	0.9167
P-LH158987	178	0.4375	0.6875
P-LH165567	188	0.5000	0.4165 (217)*
P-LH185124	423	0.8542	0.4167
P-LH206646	348	0.8125	0.7500
P-LH246376	268	0.9167	0.9583
P-LH269819	263	0.9792	0.9375
P-LH307411	190	0.5417	0.6250
P-LH375606	225	0.4167	0.4792

*Different allele in DD population.

P. robiniae genome *de novo*, generating 880,792 contigs representing 340 Mb in length from the wasp genome. The [G+C] content was 32.96%; while N content was 0.802%. We identified 14,123 SSRs from the *de novo* assembly of the genome, which included 7881 dinucleotide SSRs constituted 55.8%, 5168 trinucleotide SSRs constituted 36.6%, 882 tetranucleotide SSRs constituted 6.2%, 134 pentanucleotide SSRs constituted 0.9% and 58 hexanucleotide SSRs constituted 0.4% (table 2 in electronic supplementary material).

SSR marker development and characterization

We successfully designed 7799 SSR primer pairs of 14,123 SSRs from a unique genomic region of the wasp genome assembly. The majority of the designed SSR primer sets were dinucleotides (4579, in SSR, 58.71%), trinucleotide (2711, in SSR, 34.76%), tetranucleotide (422, in SSR, 5.41%), pentanucleotide (61, in SSR, 0.78%), and hexanucleotide (26, in SSR, 0.33%) motifs.

Two hundred and forty SSR markers were used to analyse the polymorphism of each primer set. Seventy-six SSR primer sets yielded balanced, intact, reproducible, and polymorphic amplicons in 4% agarose gels. The other SSR primer sets when amplified showed unexpected amplicon sizes. Subsequently, 34 of 76 polymorphism SSR markers contained dinucleotides (repeats motifs over 10) and equal/more trinucleotides were scored with 48 *P. robiniae* accessions (table 1). As a result, the frequency of the predominant allele for each population varied from 0.3542 to 1.00 (table 2), and 31 loci showed fixed alleles or alleles close to fixation in those two. Also, major frequency of 26 loci resulted from completely same allele in Chengdu (CD) and Dandong (DD), while other eight monomorphic loci was present respectively in two populations, one in CD and seven in DD, however, one of the two alleles in those eight loci was also shared by two populations (table 2). Further, with the allele frequency for each locus across populations, 13 rare alleles were found only (figure 1) in CD. Estimates of polymorphism indicated that the number of alleles (N_a) per locus varied widely among the markers, with an average of 3.405 alleles (range 2–8). The number of genotypes (N_g) had an average value of 4 (range 210). Effective number of alleles (N_e) had an average value of 1.9444 (range 1.0425–4.1967). In addition, the observed heterozygosity (H_o) had an average value of 0.3514 (range 0–1); the excepted heterozygosity (H_e) had an average value of 0.4124 (range 0.0414–0.7697). Finally, Shannon's information

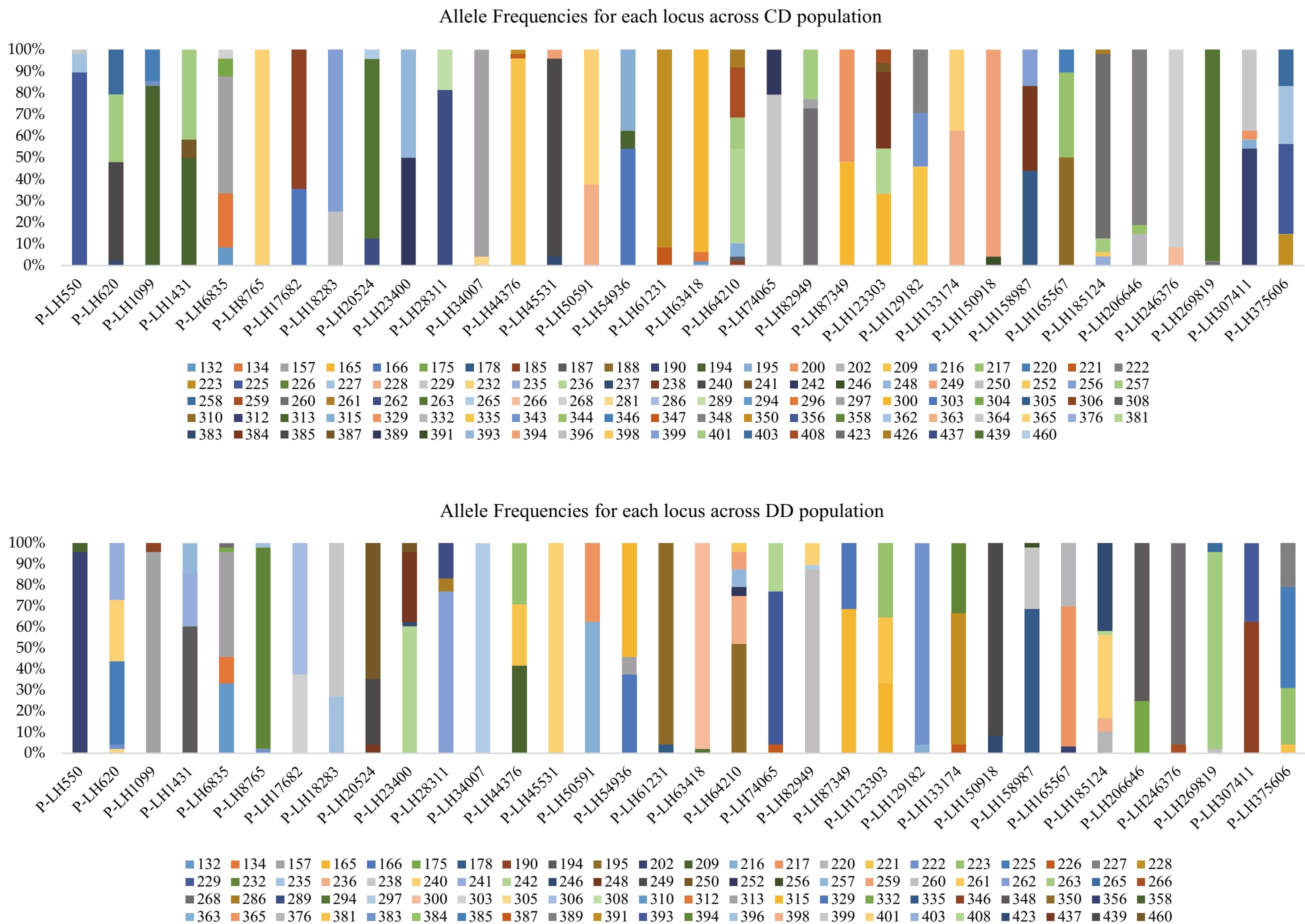

Figure 1. Allele frequencies of each locus across two populations.

Table 3. Diversity statistics from each primer used for screening 48 *P. robiniae* individuals.

Locus	N_a	N_g	N_e	H_o	H_e	I	PIC
P-LH550	4	4	1.1604	0.1250	0.1397	0.3308	0.1343
P-LH620	5	6	3.0157	0.6667	0.6754	1.1955	0.6033
P-LH1099	3	4	1.2324	0.0625	0.1906	0.3680	0.1743
P-LH1431	3	3	2.4291	0.8958	0.5945	0.9834	0.5188
P-LH6835	5	6	2.8287	0.9583	0.6533	1.2426	0.5982
P-LH8765	3	3	1.0428	0.0417	0.0414	0.1157	0.0405
P-LH17682	2	3	1.8633	0.1458	0.4682	0.6560	0.3560
P-LH18283	2	3	1.6265	0.1875	0.3893	0.5735	0.3110
P-LH20524	3	4	2.2058	0.0625	0.5524	0.8933	0.4629
P-LH23400	4	5	2.0879	0.4375	0.5265	0.8209	0.4146
P-LH28311	3	4	1.5173	0.2917	0.3445	0.5998	0.3003
P-LH34007	2	2	1.0425	0.0417	0.0412	0.1013	0.0399
P-LH44376	3	2	1.9176	0.3125	0.4836	0.8377	0.4311
P-LH45531	3	3	1.0878	0.0833	0.0816	0.2021	0.0790
P-LH50591	2	2	2.0000	0.0000	0.5053	0.6931	0.3750
P-LH54936	3	3	2.3415	0.0000	0.5789	0.9222	0.4788
P-LH61231	3	3	1.1350	0.0000	0.1202	0.2736	0.1151
P-LH63418	3	3	1.0878	0.0833	0.0816	0.2021	0.0790
P-LH64210	8	10	4.1967	0.7083	0.7697	1.6129	0.7252
P-LH74065	3	4	1.5961	0.2708	0.3774	0.6214	0.3176
P-LH82949	3	4	1.4879	0.3125	0.3314	0.5838	0.2908
P-LH87349	2	3	1.9459	0.5833	0.4912	0.6792	0.3680
P-LH123303	5	8	3.2704	0.5000	0.7015	1.2732	0.6333
P-LH129182	3	5	2.1533	0.1875	0.5412	0.9122	0.4757
P-LH133174	3	4	1.9361	0.1250	0.4886	0.7420	0.3850
P-LH150918	2	2	1.1327	0.0000	0.1184	0.2338	0.1103
P-LH158987	3	5	2.2555	0.3333	0.5625	0.9126	0.4743
P-LH165567	4	6	2.6468	0.7708	0.6287	1.0559	0.5437
P-LH185124	6	10	2.1573	0.2708	0.5421	1.0442	0.4874
P-LH206646	3	4	1.5386	0.0625	0.3537	0.5941	0.3016
P-LH246376	2	2	1.1327	0.0000	0.1184	0.2338	0.1103
P-LH269819	3	3	1.0878	0.0417	0.0816	0.2021	0.0790
P-LH307411	4	4	2.0757	0.7500	0.5237	0.8435	0.4216
P-LH375606	5	4	3.2382	0.8958	0.6985	1.3166	0.6412
Mean	3.405	4	1.9444	0.3514	0.4167	0.7143	0.3558

N_a, number of alleles; N_g, genotype number; N_e, effective number of alleles; I, Shannon‘s information index; H_o, observed heterozygosity; H_e, expected heterozygosity; PIC, polymorphism information content.

Table 4. The Ewens–Watterson test for neutrality in 34 microsatellite loci of *P. robiniae*.

Locus	Obs.F	SE	L95	U95
P-LH550	0.8618	0.0290	0.3153	0.8997
P-LH620	0.3316	0.0240	0.2624	0.8427
P-LH1099	0.8114	0.0315	0.3726	0.9590
P-LH1431	0.4117	0.0333	0.3704	0.9590
P-LH6835	0.3535	0.0247	0.2617	0.8427
P-LH8765	0.9590	0.0314	0.3752	0.9590
P-LH17682	0.5367	0.0271	0.5020	0.9794
P-LH18283	0.6148	0.0280	0.5035	0.9794
P-LH20524	0.4533	0.0318	0.3804	0.9590
P-LH23400	0.4789	0.0310	0.3079	0.9191
P-LH28311	0.6591	0.0324	0.3674	0.9590
P-LH34007	0.9592	0.0280	0.5035	0.9794
P-LH44376	0.5215	0.0316	0.3713	0.9590
P-LH45531	0.9193	0.0316	0.3743	0.9590
P-LH50591	0.5000*	0.0281	0.5035	0.9794
P-LH54936	0.4271	0.0336	0.3609	0.9590
P-LH61231	0.8811	0.0330	0.3655	0.9590
P-LH63418	0.9193	0.0328	0.3743	0.9590
P-LH64210	0.2383	0.0129	0.1866	0.6222
P-LH74065	0.6265	0.0341	0.3674	0.9590
P-LH82949	0.6721	0.0320	0.3602	0.9590
P-LH87349	0.5139	0.0279	0.5020	0.9794
P-LH123303	0.3058	0.0248	0.2671	0.8427
P-LH129182	0.4644	0.0329	0.3804	0.9590
P-LH133174	0.5165	0.0324	0.3730	0.9590
P-LH150918	0.8828	0.0276	0.5020	0.9794
P-LH158987	0.4434	0.0323	0.3700	0.9590
P-LH165567	0.3778	0.0298	0.3097	0.8997
P-LH185124	0.4635	0.0203	0.2285	0.7552
P-LH206646	0.6500	0.0311	0.3750	0.9590
P-LH246376	0.8828	0.0278	0.5035	0.9794
P-LH269819	0.9193	0.0324	0.3652	0.9590
P-LH307411	0.4818	0.0289	0.3030	0.8997
P-LH375606	0.3088	0.0246	0.2630	0.8247

Obs. F, observed sum of the squared of allelic frequency; L95 and U95, 95% confidence interval upper and lower limit; SE, standard error for observed F were calculated using 1000 simulated sample.
*F value outside the limit (lower and upper) of 95% confidence region.

index (*I*) had an average value of 0.7143 (range 0.2021–1.6129); the polymorphic information content (PIC) had an average value of 0.3558 (range 0.0405–0.7252) (table 3). *P*-value deviations from the expected Hardy–Weinberg equilibrium (HWE) were evaluated with the Markov chain algorithm (10,000 steps) for each locus. Significant deviation from HWE was observed for seven markers in CD and 15 markers in DD, genotypic LD was tested between all pairs of alleles for loci across populations based on a permutation procedure ($P<0.05$) in the same program (10,000 permutations). Significant LD was detected between 48 pairs of alleles for loci, 24 in each population.

F-statistics analysis showed that the F_{is} had average value of 0.1022 per locus (range −0.9233–1), and 15 of 34 loci with negative F_{is} values, indicating heterozygote excess (outbreeding) in each locus; the F_{st} had an average value of 0.0539 (range 0.0050–0.3258). Gene flow (N_m) averaged 4.3849, and ranged from 0.5174 at P-LH50591 to 533.5000 at P-LH17682 (table 3 in electronic supplementary material). The Ewens–Watterson test for neutrality of microsatellite markers was conducted with the infinite-alleles model in Arlequin software v. 3.5 (Excoffier and Lischer 2010). The result showed that *F*-value (sum of square of allelic frequency) lay outside the lower and upper limits of 95% confidence region, except at one locus (P-LH50591) (table 4), which indicates that most of markers are neutral. NJ analysis (bootstrap 1000) showed that CD and DD populations clustered into two main divisions, however some individuals from two populations clustered together as the third division (figure 1 in electronic supplementary material), which indicates the gene flow between two populations and the within-individual differentiation were present.

Discussion

In this study, NGS methods was applied to the genome sequence of *P. robiniae*. From the unique genomic regions of the wasp 14,123 SSRs was identified, ranging from

dinucleotide to hexanucleotide motifs. From which 7799 SSR primer pairs were designed from the putative SSR regions identified. Subsequently, 34 SSR marker of 76 polymorphism SSR markers, with over trinucleotides or with dinucleotides but more repeats motifs (over 10) which were considered stable and can be used for population genetics were scored with 48 *P. robiniae* accessions.

For the 34 SSR loci, average of observed heterozygosity, expected heterozygosity, Shannon‘s information index and polymorphism information content (0.3514, 0.4167, 0.7143 and 0.3558, respectively) were moderate. Among them, H_o =0.0 for five loci (table 3), which probably reduced the mean values across all 34 loci. Nevertheless, it is higher than previous mean levels found in other hymenopterans (Anjos *et al.* 2016; Simonato *et al.* 2019). Allele numbers per loci were in the range 2–6. In the absence of comparable data from conspecific populations or congeneric species, we do not know whether the low number of alleles is a natural feature of *P. robiniae*, or is due to the application of methodologies. In addition, the low number of alleles observed at each locus, the high frequency of fixed alleles, together with rare alleles and alleles at low frequencies, suggests that the exchange of wasps between CD and DD is insufficient to maintain the two populations as distinct genetic entities.

According to the result of the Ewens–Watterson test for neutrality of microsatellite markers, two loci were not neutral, which may be related to some selection traits. This is likely associated with genetic draft, which is well known as a potent factor in changing allelic frequency and heterozygosity (Kumar *et al.* 2009). The possible association with selection near these loci should be further investigated.

The genus *Platygaster* contains many species, and the biology of the most species is unknown. Although some species may parasitize different kinds of gall midges (Tóth *et al.* 2011; Buhl and Eiseman 2016), *P. robiniae* has only been found to parasitize *O. robiniae*, hence, it is considered to be a specific parasitoid of the black locust leaf gall midge, and may have been introduced into Europe and Asia from North America along with its host and has spread gradually. We have shown that Chinese *O. robiniae* populations cluster into two groups independent of geographical distance, which initially shaped the genetic structure of the Chinese *O. robiniae* populations (Yao *et al.* 2020). By means of the novel markers developed here, we can now analyse the population genetics of Chinese populations of *P. robiniae,* and determine whether the synchronicity of colonization is present with the host in the future.

Acknowledgments

We thank Mr Xiang-Chen Chen (General Station of Forest and Grassland Pest Management, National Forestry and Grassland Administration, Shenyang) for the help in samplings. We also sincerely thank the other members of the Plant Quarantine Lab of the Research Institute of Forest Ecology, Environment and Protection, Chinese Academy of Forestry. This research was funded by the National Key Research and Development Program of China (2016YFC1201200).

Author contribution

Conceptualization, YX and WX; methodology, YJ and YX; software, YJ and HL; validation, YJ, HL, ZR, and YX; formal analysis, HL, ZR; investigation, YJ, HL and HQ; resources, YJ, HL and HQ; data curation, YJ, HL and ZR; writing - original draft preparation, YJ, HL, ZR, HQ, WX and YX; writing - review and editing, YJ, HL, ZR, HQ, WX and YX; visualization, YJ, HL; supervision, WX and YX; project administration, YX. funding acquisition, YX.

References

Amouroux P., Normand F., Nibouche S. and Delatte H. 2013 Invasive mango blossom gall midge, *Procontarinia mangiferae* (Felt) (Diptera: Cecidomyiidae) in Reunion Island: ecological plasticity, permanent and structured populations. *Biol. Invasions* **15**, 1677–1693.

Anjos L. O. S., Peixoto R. F., Chanquinie D. M. S., Pinto L. R., Creste S. A., Dinardo-Miranda L. L. *et al*. 2016 Microsatellite loci and genetic structure of artificial populations of *Cotesia flavipes* (Hymenoptera, Braconidae). *Genet. Mol. Res.* **15**, 1–15.

Badmin J. 2016 *Obolodiplosis robiniae* (Haldeman) (Diptera: Cecidomyiidae) on False–acacia cv Frisia. *Brit. J. Entomol. Nat. Hist.* **29**, 245.

Bálint J., Neacşu P., Balog A., Fail J. and Vétek G. 2010 First record of the black locust gall midge, *Obolodiplosis robiniae* (Haldeman) (Diptera: Cecidomyiidae), in Romania. *North-West. J. Zool.* **6**, 319–322.

Beier S., Thiel T., Munch T., Scholz U. and Mascher M. 2017 MISA-web: a web server for microsatellite prediction. *Bioinformatics (oxford, England).* **33**, 2583–2585.

Bentur J. S., Sinha D. K., Padmavathy C., Revathy C., Muthulakshmi M. and Nagaraju J. 2011 Isolation and characterization of microsatellite loci in the Asian rice gall midge (*Orseolia oryzae*) (Diptera: Cecidomyiidae). *Int. J. Mol. Sci.* **12**, 755–772.

Buhl P. N. and Duso C. 2008 *Platygaster robiniae* n sp (Hymenoptera Platygastridae) Parasitoid of *Obolodiplosis robiniae*(Diptera Cecidomyiidae) in Europe.pdf. *Ann. Entomol.* **101**, 297–300.

Buhl P. N. and Eiseman C. S. 2016 Three new reared species of *Platygaster* (Hymenoptera, Platygastroidea: Platygastridae) from the United States. *Int. J. Environ. Stu.* **73**, 1006–1015.

Cavagnaro P. F., Senalik D. A., Yang L., Simon P. W., Harkins T. T., Kodira C. D. *et al.* 2010 Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics* **11**, 569.

Cerna K. and Straka J. 2013 Identification of 37 microsatellite loci for *Anthophora plumipes* (Hymenoptera: Apidae) using next generation sequencing and their utility in related species. *Eur. J. Entomol.* **109**, 155–160.

Duso C., Fontana P. and Tirello P. 2005 Spread of the gall midge *Obolodiplosis robiniae* (Haldeman) injurious to black locust in Italy and Europe [*Robinia pseudoacacia* L.]. *Inform. Fitopatol.* **55**, 30–33.

Excoffier L. and Lischer H. 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567.

François R. 2008 genepop’007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* **1**, 103–106.

Gil J., Um Y., Kim S., Kim O. T., Koo S. C., Reddy C. S. *et al*. 2017 Development of genome-wide SSR markers from *Angelica gigas* Nakai using next generation sequencing. *Genes (Basel)* **8**, 238.

Guichoux E., Lagache L., Wagner S. *et al*. 2011 Current trends in microsatellite genotyping. *Mol. Ecol. Resour.* **20**, 591–611.

Jørgensen J. 2009 *Obolodiplosis robiniae* (Haldeman, 1847)(Cecidomyiidae) and its parasitoid *Platygaster robiniae* Buhl et Duso, 2007 (Platygastridae) two species new for Denmark. *Entomol. Meddelelser* **77**, 141–144.

Kim I.-K., Park J. D., Shin S. C. and Park I.-K. 2011 Prolonged embryonic stage and synchronized life-history of *Platygaster robiniae* (Hymenoptera: Platygastridae), a parasitoid of *Obolodiplosis robiniae* (Diptera: Cecidomyiidae). *Biol. Control* **57**, 24–30.

Kirk H., Dorn S. and Mazzi D. 2013 Molecular genetics and genomics generate new insights into invertebrate pest invasions. *Evol. Appl.* **6**, 842–856.

Kumar S., Dixit S. P., Verma N. K., Singh K. D. and Pande A. 2009 Genetic diversity analysis of the gohilwari breed of Indian goat (Capra hircus) using microsatellite markers. *Am. J. Anim. Vet. Sci.* **4**, 49–57.

Lee H. Y., Moon S., Shim D., Hong C. P., Lee Y., Koo C. D. *et al*. 2017 Development of 44 novel polymorphic SSR markers for determination of shiitake mushroom (*Lentinula edodes*) cultivars. *Genes (Basel)* **8**, 109.

Li R., Zhu H., Ruan J., Qian W., Fang X., Shi Z. *et al*. 2010 De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265.

Linløkken A. N. 2018 Genetic diversity in small populations. In Genetic diversity and disease susceptibility (ed. Y. M. Liu), pp. 43–55. IntechOpen, London.

Liu K. and Muse S. V. 2005 Powermarker: integrated analysis environment for genetic marker data. *BMC Bioinformatics* **21**, 2128–2129.

Ljubodrag M., Glavendeki M. M., Ivana J. and Saa M. 2008 *Obolodiplosis robiniae* (Haldeman) (Diptera: Cecidomyiidae): A new invasive insect pest on black locust in Serbia. *Glasnik Umarskog Fakulteta* **97**, 197–207.

Lu C. K., Buhl P. N., Duso C., Zhao C. M., Zhang J. S., Ji Z. X. *et al*. 2010 First discovery of *Platygaster robiniae* (Hymenoptera: Platygastridae) parasitizing the invasive *Obolodiplosis robiniae* (Diptera: Cecidomyiidae), a gall maker in China. *Acta. Entomol. Sinica* **53**, 233–237.

Luo R. B., Xie Y. L., Li Z. Y., Huang W. H., Yuan J. Y., He G. Z. *et al*. 2012 SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18.

Park J. D., Shin S. C., Kim C. S., Jeon M. J. and Park I.-K. 2009 Biological characteristic of *Obolodiplosis robiniae* and control effects of some insecticides. *Korean J. Appl. Entomol.* **48**, 327.

Peakall R. and Smouse P. E. 2010 GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288–295.

Pernek M. and Matošević D. 2009 Black locust gall midge (*Obolodiplosis robiniae*), new pest on black locust trees and first record of parasitoid *Platygaster robiniea* in Croatia. *Sumarski List.* **133**, 157–163.

Rozen S. and Skaletsky H. 1999 Primer3 on the WWW for general users and for biologist programmers. bioinformatics methods and protocols (ed. S. Misener and S. A. Krawetz). Humana Press, Totowa.

Sakai A. K., Allendorf F. W. and Holt J. S. 2001 The population biology of invasive species. *Ann. Rev. Ecol. Syst.* **32**, 305–332.

Schubert M., Lindgreen S. and Orlando L. 2016 AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88.

Simonato M., Pilati M., Magnoux E., Courtin C., Sauné L., Rousselet J. *et al.* 2019 A population genetic study of the egg parasitoid *Baryscapus servadeii* reveals large scale automictic parthenogenesis and almost fixed homozygosity. *Biol. Cont.* **139**, 104097..

Sviridov S. V. and Bazhenova T. N. 2009 The first record of *Platygaster robiniae* (Hymenoptera, Platygastridae), parasitoid of *Obolodiplosis robiniae* (Diptera, Cecidomyiidae) for Ukraine. *Vestnik Zoologii* **43**, 106.

Terhorst C. P. and Lau J. A. 2015 Genetic variation in invasive species response to direct and indirect species interactions. *Biol. Invasions* **17**, 651–659.

Tomov R. 2014 A review of pest status of recently recorded alien insects in Bulgaria [Conference poster]. 4th ESENIAS Workshop: International Workshop on IAS in Agricultural and Non-Agricultural Areas in ESENIAS Region, 16-17 December 2013 anakkale, Turkey.

Tóth P., Váňová M. and Luká J. 2011 Impact of natural enemies on *Obolodiplosis robiniae* invasion. *Biologia* **66**, 870–876.

Yang J., Zhao W. X., Ma L. and Yao Y. X. 2019 Geographic distribution of *Platygaster robiniae* (Hymenoptera: Platygastridae) parasitizing an invasive insect pest *Obolodiplosis robiniae* (Diptera: Cecidomyiidae) in China. *J. Environ. Entomol.* **41**, 167–172.

Yao Y. X., Zhao W. X. and Shang X. P. 2015 Development of polymorphic microsatellite markers of *Obolodiplosis robiniae*(Haldeman) (Diptera: Cecidomyiidae), a north American pest invading Asia. *J. Insect Sci.* (online) **15**, 127.

Yao Y. X., Shang X. P., Yang J., Lin R. Z., Huai W. X. and Zhao W. X. 2020 Genetic variation may have promoted the successful colonization of the invasive gall mMidge, *Obolodiplosis robiniae*, in China. *Front. Genet.* **11**, 387.

Zalewski A., Michalska-Parda A., Bartoszewicz M., Kozakiewicz M. and Brzeziński M. 2010 Multiple introductions determine the genetic structure of an invasive species population: American mink Neovison vison in Poland. *Biol. Conserv.* **143**, 1355–1363.

Corresponding editor: Steven M. Carr