Apidologie
© INRA, DIB and Springer-Verlag France, 2012
DOI: 10.1007/s13592-012-0168-3

Original article

Discovery and characterization of microsatellites for the solitary bee *Colletes inaequalis* using Sanger and 454 pyrosequencing

Margarita M. López-Uribe[1], Christine K. Santiago[1], Steve M. Bogdanowicz[2], Bryan N. Danforth[1]

[1]Department of Entomology, Cornell University, Ithaca, NY 14853, USA
[2]Evolutionary Genetics Core Facilities, Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

Received 5 June 2012 – Revised 8 August 2012 – Accepted 18 September 2012

Abstract – The recent implementation of next-generation sequencing for the discovery of microsatellite markers has made this technology the most effective method for generating genetic markers in non-model organisms. Here, we report the de novo discovery of microsatellite markers for the solitary bee *Colletes inaequalis* using cloning/Sanger sequencing and direct 454 pyrosequencing from microsatellite-enriched genomic libraries. We identified and successfully multiplexed 18 highly variable microsatellite markers in 585 individuals. The number of alleles per locus ranged from 3 to 23, and the expected heterozygosity ranged from 0.056 to 0.912. These genetic markers will allow for the investigation of levels of inbreeding and fine-scale population structure in *C. inaequalis*. Our results contribute to the literature demonstrating that 454 sequencing is more time- and cost-efficient than cloning/Sanger sequencing at identifying a large number of genomic regions with microsatellite repeat motifs.

SSRs / cloning / next-generation sequencing / Colletidae

1. INTRODUCTION

Simple sequence repeats (SSRs) or microsatellites are short DNA sequences consisting of tandem repeated motifs that vary in length, typically from 1 to 6 bp long. SSRs are most commonly present in non-coding regions and are characterized by high levels of repeat length polymorphism that are the result of two mutation mechanisms, replication slippage and unequal crossover (Schlötterer and Tautz 1992). Due to their high levels of polymorphism, these co-dominant multi-allelic markers have been widely used in genome mapping, parentage analysis, population structure, and phylogeographic studies, for organisms ranging from bacteria to humans (Blair and McCouch 1997; Weber et al. 1991).

In the past, one of the major limitations for the use of microsatellite markers in non-model organisms was the time and cost of developing these markers de novo (Zane et al. 2002). When microsatellite markers were first described (Tautz 1989), their isolation and characterization were time-consuming and expensive because protocols were based on a random screening of whole genome libraries. Currently, there are several enrichment protocols for microsatellite libraries (Zane et al. 2002), which

Corresponding author: M.M. López-Uribe, mml82@cornell.edu
Manuscript editor: Marina Meixner

Published online: 23 October 2012

can greatly accelerate the process of microsatellite development by making the screening process more targeted. Moreover, with the availability of next-generation sequencing technology, de novo isolation of microsatellite markers has become routine for non-model organisms (Mikheyev et al. 2010; Perry and Rowe 2011; Zalapa et al. 2012). A large number of microsatellite loci can be detected by direct pyrosequencing of enriched libraries followed by bioinformatic detection of simple repeated sequences in the assembled sequenced data (Andrés and Bogdanowicz 2011).

Colletes inaequalis is a solitary bee, meaning each nest is occupied by a single reproductively active female, native to eastern North America. This species is a wild pollinator of early spring wild plants and could become an important native pollinator for crops such as blueberries (Batra 1995) and apples (Gardner and Ascher 2006). *C. inaequalis* has a solitary life cycle characterized by adults emerging from their underground cocoons early in the spring. Both females and males have a strong tendency to stay close to their emergence sites. Males are the first to become active, and they remain patrolling the emergence area waiting for females to mate with. When females emerge, males actively search for them to copulate, while females immediately start building and provisioning their nests in the natal area (Batra 1980). Observations from behavioral studies indicate that nest aggregations of *C. inaequalis*, and other early spring *Colletes* spp., naturally experience philopatric behavior and high levels of inbreeding (Vereecken 2008). However, this question has not been investigated due to the lack of species-specific hypervariable genetic markers.

In this study, we describe and characterize 18 hypervariable microsatellite markers for the ground-nesting bee *C. inaequalis* (Colletidae). We enriched for microsatellite loci by hybridizing biotinylated nucleotide repeats to restriction enzyme digested genomic DNA. We then screened for microsatellites using two different methods: (1) the traditional cloning protocol that selects positive clones containing microsatellites by hybridization of radiolabeled oligonucleotides and (2) direct sequencing of enriched libraries for microsatellites using Roche 454 sequencing technology. We estimated the variability of the identified SSRs markers using four microsatellite multiplex sets in 585 *C. inaequalis* individuals.

2. MATERIAL AND METHODS

2.1. Enriched library preparation

Genomic DNA was extracted from two *C. inaequalis* males collected in Ithaca (NY, USA, 42°26′39.8″N 76°29′32.4″W) using the Qiagen DNeasy Blood and Tissue Kit. DNA quality and concentration were measured using a Nanodrop 1000 spectrophotometer. We digested the genomic DNA using two restriction enzymes, *BsaA I* and *Hinc II*, to minimize problems of unequal sampling of genomic regions due to genome composition biases. During digestion, double-stranded SNX linkers were ligated to the ends of the genomic fragments (Hamilton et al. 1999). An aliquot of the digestion was incubated with synthetic, single-stranded, biotinylated oligonucleotides representing 12 repeat motifs (GT, TC, TA, TTA, GTT, TTC, GCT, TTTA, TTTG, TTTC, GATA, and GTAT) that had streptavidin attached to the 3′ end. Genomic DNA/biotinylated complexes were captured with streptavidin-coated magnetic beads. A PCR reaction using a SNX forward primer was then used to transform the captured genomic fragments into double-stranded DNA (Andrés and Bogdanowicz 2011).

2.2. Microsatellite discovery by cloning and Sanger sequencing

PCR products from the enrichment protocol were digested with *Nhe* I and ligated into pUC 19 plasmids previously digested with *Xba* I and dephosphorylated. Plasmids were introduced into *Escherichia coli* cells through electroporation. Transformed bacteria were cultured on plates containing ampicillin and X-gal, then transferred onto nylon membranes and fixed by autoclaving. These membranes were screened for microsatellites through hybridization to 33P-radiolabeled oligonucleotides with the same repeat motifs used for the enrichment process. Plasmid

DNA from bacterial colonies that were found to be positive for the presence of microsatellite repeats were extracted and Sanger sequenced using M13 primers at the vector/insert boundary.

2.3. Microsatellite discovery by 454 sequencing

Post-enrichment PCR products were purified with a Qiagen PCR purification kit and ligated to 454 A and B adapters. The *C. inaequalis* library was pooled with seven other libraries generated from other taxa for sequencing in half 454 run. Each library was identified with a unique multiplex identifier (MID). Small fragments from the pooled libraries were removed with Ampure beads (Beckman Coulter). The final pooled sample was quantified, subjected to bead titration, and sequenced with Roche/454 FLX chemistry. Reads were sorted out by MID before assembly.

We used CodonCode Aligner 4.0.2 (CodonCode Co.) to trim adapters and MIDs from all reads and discard short and low-quality data. Cleaned reads were assembled into unique contigs using CodonCode Aligner with the following parameters: minimum percentage of identity, 55; minimum overlap length, 75; word length, 24; and maximum successive failures, 500. Contigs were analyzed in the RepeatMasker (Smit et al. 2010) software to scan all sequences for simple tandem repeats.

2.4. Primer design and testing

We designed primers for DNA sequences obtained from Sanger and 454 sequencing that contained a minimum of six perfect repeats using the software PrimerSelect (DNASTAR). We used the following criteria for primer design: maximum primer length of 30 bp, melting temperature from 54 to 60 °C, maximum dimer duplexing of 3 bp, and maximum hairpin duplexing of 3 bp. For primer testing, we genotyped ~30 male individuals using the universal tag three-primer method (Schuelke 2000) that includes: 5′ 6-FAM M13 primer, marker-specific forward primer plus a long tag at the 5′ end that is complementary to the M13 primer, and marker-specific reverse primer plus a six base "pigtail" (GTTTCT) attached to the 5′ end to reduce stutter peaks (Brownstein et al. 1996). Simplex three-primer PCR reactions contained 5× GoTaq buffer pH 8.5, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.1 μM forward primer, 0.2 μM FAM M13 primer, 0.2 μM reverse primer, 1 U GoTaq DNA polymerase (Promega), and ~20 ng DNA in a final volume of 20 μL. PCR cycles were performed in a Biometra TGradient thermal cycler with 1 cycle at 94 °C for 30 s, 35 cycles at 94 °C for 30 s, 45 s at the marker-specific annealing temperature, and 45 s at 72 °C, followed by one step of 7 min at 72 °C. PCR products were visualized in 2.5 % agarose gels and genotyped on an Applied Biosystems 3730xl DNA Analyzer using GeneScan-500 LIZ as the allele size standard. We used PeakScanner v.1.0 (Applied BioSystems) to genotype each microsatellite loci independently and determine their variability and consistent repeat pattern across samples. PCR conditions were adjusted for loci that presented stutter or multiple peaks by increasing the $MgCl_2$ concentration and/or increasing the annealing temperature.

Table I. Comparison of the Sanger and 454 sequencing methods for microsatellite discovery of *Colletes inaequalis.*

	Sanger	454
No. of fragments sequenced	~200	290
No. of reads with microsatellite motif present	26 (13 %)	290 (68 %)
No. of reads where primers could be designed	23 (88 %)	19 (7 %)
No. of variable loci detected	23 (100 %)	11 (58 %)
Price per locus	$3.90	$2.50
Total time	2 months	3 weeks

Table II. Descriptive summary and genetic diversity indices of the 18 microsatellite markers isolated from *Colletes inaequalis*.

Locus	Primer sequences	Motif	T_a	Allele sizes	N_a	H_E	H_O	Accession no.
CI010[a]	F: TTT TTC TCG CCT CCG GTT TTT CAG R: GCC CCG CCA ATC ATC ACA GTT TC	$(CA)_{15}$	59	160–194	18	0.911	0.865	JX117857
CI12	F: GTT TCG CAT TTG GAA GTA CCG ACA AT R: CGT CAA ATG GCG ACT CGT GAT AAT C	$(CT)_{10}$	55	207–239	11	0.792	0.760	JX117858
CI15	F: GTG TGG TTG GCT GCA TTG CTG TA R: TGG ACG CAT ATT GTG CCA TCC T	$(TG)_{11}$	58	114–132	5	0.581	0.534	JX117859
CI23	F: CCA AAC TGA CAT GTC ATT CGT TCG R: CCC GAG TGG CGA GGA GTT CA	$(TC)_7AC(TC)_7$	59	248–270	12	0.831	0.794	JX117860
CI27	F: ACA AAG GGG CCG ACA GAG TAA T R: AAC GGC ACG GAA AAA GTA AGT AAA	$(TG)_7GG(TG)_6$	54	193–223	10	0.379	0.385	JX117861
CI028[a]	F: CTT TGC GCC CTT ATC CTC TTA CTC R: TAT TTA TGT TCG GCT ATC GGT TGC	$(CA)_7CCA$	55	142–164	5	0.463	0.494	JX117862
CI35	F: TTT TTG CGT GCA GTT TTA CGA GAA R: ACG CGT GCA CGC TTT CCT T	$(TG)_{23}$	54	238–314	11	0.361	0.229	JX117863
CI62	F: GCG CGA CGC TGC TAA CAA C R: ATT CGT CGC ATT ATA CGC ACC TG	$(CAA)_8$	53	119–164	15	0.733	0.734	JX117864
CI66	F: TCA CGC CCG GGA TAG GAA G R: GTA ACG GGC CGC GAC AAT AG	$(AAG)_{15}$	55	350–410	15	0.853	0.851	JX117865
CI73	F: ACC AAG TGA GCG AAC CCT CCT TC R: GAC TCC GGC GTC GTT TCT TCC	$(TAGA)_{11}$	54	181–205	6	0.056	0.051	JX117866
CI075[a]	F: CAA CCA AAT TTT GTG TAC GCT GTG R: GTC CCG ATC GTC TGG TAT TTT GTA	$(TG)_6CG(TG)_2TA(TG)_6$	54	367–395	9	0.320	0.320	JX117867
CI87	F: AAT GCG AGA AGC GTT CAA GTC T R: ATT TTA TGC GCG TGC TAT CCA	$(CA)_{16}$	54	304–356	22	0.912	0.786	JX117868
CI98	F: GGG ACC GCG GGA CAG AGA A R: CCC CCA CCT CCG TCC TTT TC	$(CA)_{15}$	59	221–249	14	0.752	0.723	JX117869
CI099[a]	F: TCA GTG TCG GTG ATA TCA AGG AAG R: CGT AGC AGG GCT GTA AAA GTT TCT	$(AC)_9$	52	250–264	6	0.682	0.686	JX117870
CI102	F: CGC CTA ACG ATC CGA AAA TAA T R: GCG TAA TCG CGA TAA TCA GAG A	$(TTGT)_8$	53	169–197	8	0.799	0.789	JX117871
CI106	F: GTT CCA CCA AAT TCC CTC CAC T	$(GT)_{11}$	55	350–410	23	0.801	0.757	JX117872

Table II (continued)

Locus	Primer sequences	Motif	T_a	Allele sizes	N_a	H_E	H_O	Accession no.
	R: AGC TGC ACG ATA CGG TAA AAG AAT							
CI131[a]	F: TAT ATG GCC AGG AGG TAG AGT TGA R: CGA TGC CGA TAA GCG TCT C	$(GTT)_3(GCT)_2(GTT)_3$	55	240–261	3	0.472	0.477	JX117873
CI179[a]	F: GGA TAG GTC GGG GAT AGT GCT GTC R: GCG CCG TCG ATA GAA GGA AAA A	$(CT)_9$	59	299–305	4	0.575	0.566	JX117874

Primer sequences (5′ to 3′), repeat motif, annealing temperature (T_a) in Celsius degrees (°C), allele size range, number of alleles (N_a), expected heterozygosity (H_E), observed heterozygosity (H_O), and GenBank accession numbers

[a] Loci identified using the 454 sequencing protocol

2.5. PCR multiplex design and marker variability

We selected loci for multiplex PCRs based on the following criteria: (1) clear amplification pattern, (2) unambiguous genotypes, and (3) presence of at least two alleles. Up to five microsatellite loci that amplified at the same annealing temperature (±1 °C) but exhibited non-overlapping allele sizes were selected for multiplexing in the same microsatellite set. Forward primers in each set were fluorescently labeled with one of the following dyes: 6-FAM (blue), VIC (green), NED (black), and PET (red). We used the Qiagen PCR Multiplex Kit with the fluorescently labeled forward primer and the reverse primer plus the "pigtail" in a final volume of 10 μL using the following PCR cycles: 30 s denaturation step at 94 °C, 90 s annealing step at 54–59 °C, and 90 s extension step at 72 °C for 35 cycles and a final extension step at 72 °C for 10 min. Multiplexed PCR products were diluted 1:12, and 1 μL from this dilution was mixed with 0.2 μL of GeneScan-500 LIZ in a 20-μL final volume Hi-Di Formamide solution for genotyping on an Applied Biosystems 3730xl DNA Analyzer. Electropherograms were visualized and genotyped in the software PeakScanner v.1.0 (Applied BioSystems). Allele frequencies, genetic diversity indices, Hardy–Weinberg, and linkage equilibrium were estimated using the software Genetix 4.05 (Belkhir et al. 2004) on 80 males and 505 females of *C. inaequalis* collected from nine nest aggregations in the city of Ithaca (NY, USA, 42°26′36″N 76°30′0″W).

3. RESULTS

More than 500 positive clones were obtained from the cloning protocol, and ~200 of these were Sanger sequenced. We successfully designed primers for 26 (13 %) DNA sequences, and we tested 21 primer pairs. Of the tested primer pairs, 3 did not give any PCR product, 6 did not show clean amplification, and 12 showed a variable and repeatable pattern (Table I).

From the 454 sequencing, we obtained 3,164 sequences with an average length of 495 bp after discarding short and low quality reads. These reads assembled into 428 contigs of which 290 contained simple microsatellite repeat

motifs. As expected, the most common repeat motifs were dinucleotide (51 %), followed by trinucleotide (29 %) and tetranucleotide (20 %) repeats. We designed primers for 19 (7 %) contigs that contained simple repeat motifs and tested 14 of these primer pairs. Four primer sets did not amplify, four were not variable, and six were variable and showed a clean amplification pattern (Table I). Combining our results from the cloning/Sanger and 454 protocols, we successfully amplified 18 loci comprised of 13 dinucleotide, 3 trinucleotide, and 2 tetranucleotide motif repeat microsatellite markers (Table II). These loci were amplified and genotyped using four microsatellite multiplex sets based on the criteria defined in the methods section (Figure 1).

A first simplex genotype test of *C. inaequalis* males (n=80) revealed that these 18 primer pairs showed clear repeat patterns and absence of null alleles because all haploid males amplified for all loci. No diploid males were detected among the male individual analyzed. We used the female dataset (n=505) to assess marker variability using the microsatellites multiplex protocol. The average number of alleles per locus was 11 (3–23), expected heterozygosity ranged from 0.056 to 0.912, and observed heterozygosity ranged from 0.051 to 0.865 (Table II). Four loci significantly deviated from Hardy–Weinberg equilibrium showing significant heterozygote deficiency (CI010, P=0.0008; CI35, P<0.000; CI87, P< 0.000; and CI131, P=0.0017). The amplification of all male individuals for all loci in our dataset suggests that deviations from Hardy–Weinberg equilibrium are not the results of null alleles but of the demographic history and/or mating system of this species. This hypothesis is supported by the detection of significant heterozygote deficiency overall loci when the female dataset was analyzed using the nesting sites as discrete sampling units (P<0.000 for six out of the nine nesting sites). Two pairs of loci exhibited significant linkage disequilibrium after Bonferroni correction for multiple tests (CI62 and CI102, P<0.000; CI15 and CI66, P<0.000).

4. DISCUSSION

The development of microsatellite markers for *C. inaequalis* provides the first species-specific molecular markers for this bee and adds to the growing number of de novo microsatellite markers developed for solitary bee species (Table III). These genetic tools are essential for the study of mating systems, sociogenetic structure, population structure, and conservation genetics of solitary bees (Zayed 2009). Solitary bee species are currently the focus of numerous ecological and genetic studies because they may serve as alternative agricultural pollination agents to compensate for the loss of managed

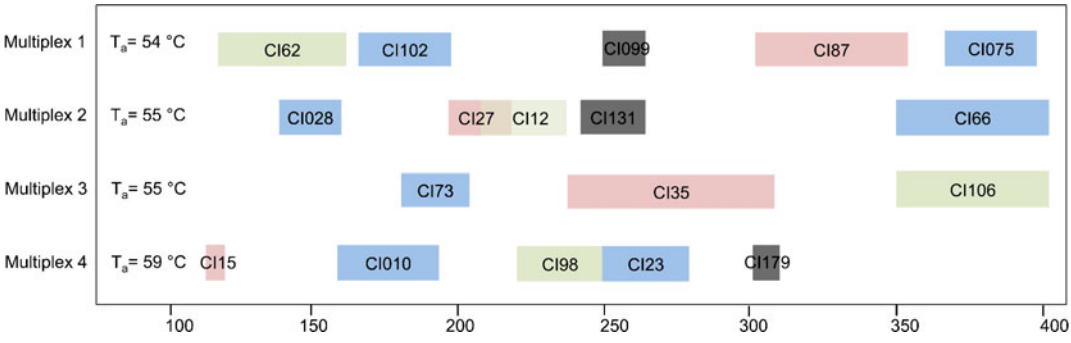

Figure 1. Diagram showing the annealing temperature and allele range for the microsatellite loci in each multiplex PCR set. *Colored boxes* represent the observed allele range for each locus. *Colors* represent the fluorescent dyes used to labeled forward primers: FAM (*blue*), NED (*black*), PET (*red*), and VIC (*green*).

Table III. Summary table of studies reporting microsatellite markers for solitary bees.

Species	Subfamily	Family	No. of loci	No. of alleles	H_E	Reference
Colletes inaequalis	Colletinae	Colletidae	18	10.9 (±6.0)	0.626 (±0.240)	This study
Xylocopa frontalis	Xylocopinae	Apidae	14	10.4 (±3.4)	0.799 (±0.073)	(Augusto et al. 2011)
Eulaema meriana	Apinae	Apidae	8	6.1 (±1.8)	0.685 (±0.153)	(López-Uribe et al. 2010)
Megalopta genalis	Halictinae	Halictidae	12	11 (±5.1)	0.825 (±0.093)	(Kapheim et al. 2009)
Halictus rubicundus	Halictinae	Halictidae	14	10.1 (±4.6)	0.704 (±0.163)	(Soro and Paxton 2009)
Euglossa annectans	Apinae	Apidae	9	5.3 (±2.2)	0.598 (±0.257)	(Paxton et al. 2009)
Eulaema nigrita	Apinae	Apidae	11	9.1 (±3.3)	0.794 (±0.086)	(Souza et al. 2007)
Euglossa cordata	Apinae	Apidae	9	16.4 (±9.3)	0.764 (±0.156)	(Souza et al. 2007)
Osmia rufa	Megachilinae	Megachilidae	6	4.2 (±0.7)	n.a.	(Neumann and Seidelmann 2006)
Lasioglossum leucozonium	Halictinae	Halictidae	10	2.4 (±0.5)[a]	0.440 (±0.183)[a]	(Zayed 2006)
			6	10.8 (±4.1)[b]	0.848 (±0.056)[b]	(Zayed et al. 2007)
Lasioglossum oenotherae	Halictinae	Halictidae	9	6.4 (±3.8)	0.616 (±0.306)	(Zayed 2006)
Ceratina flavipes	Xylocopinae	Apidae	8	n.a.	0.585 (±0.20)	(Azuma et al. 2005)
Amegilla dawsoni	Apinae	Apidae	12	6.2 (±6.4)	0.511 (±0.208)	(Beveridge and Simmons 2004)
Exoneura nigrescens	Xylocopinae	Apidae	12	18.6 (±16.7)	0.888 (±0.087)	(Langer et al. 2004)
Exoneura robusta	Xylocopinae	Apidae	13	20.9 (±18.7)	0.897 (±0.080)	(Langer et al. 2004)
Macrotera portalis	Panurginae	Andrenidae	7	6.1 (±1.8)	0.302 (±0.110)	(Danforth et al. 2003)
Lasioglossum malachurum	Halictinae	Halictidae	9	6.2 (±3)	0.744 (±0.287)	(Paxton et al. 2002)
Lasioglossum hemichalceum	Halictinae	Halictidae	10	14 (±4.4)	0.786 (±0.060)	(Kukuk et al. 2002)
Andrena vaga	Andreninae	Andrenidae	19	5.8 (±1.9)	0.747 (±0.202)	(Mohra et al. 2000)
Andrena jacobi	Andreninae	Andrenidae	4	7.3 (±2.7)	0.616 (±0.121)	(Paxton et al. 1996)

No. of alleles indicates the average number of alleles per species and the standard deviation. H_E indicates the average expected heterozygosity across loci

[a] Data derived from an introduced population to North America

[b] Data derived from a population from its native range in Europe

pollinators, such as the honeybee (Winfree et al. 2007).

To our knowledge, this is the first study that uses both cloning/Sanger and 454 sequencing protocols for the development of microsatellite markers. We found that the implementation of 454 sequencing for microsatellite discovery increases ~5-fold the number of identified sequences with simple microsatellite repeat motif, proving to be a more time effective method for de novo microsatellite discovery than the traditional Sanger sequencing method

(Table I). However, researchers should be aware of the methodological difficulties of this new approach. First, 454 sequencing has a higher sequencing error rate than Sanger sequencing particularly in homopolymeric regions (Huse et al. 2007). One of the main difficulties we faced when designing primer sequences for loci identified through pyrosequencing was the absence of high-quality flanking regions in many of the contigs. This difficulty can be overcome by (1) using newer pyrosequencing technology, such as the Roche Titanium chemistry, that generates longer average read lengths and/or (2) obtaining greater sequencing depth. Moreover, the assembly of sequence data with a high number of repeated regions can be challenging due to contigs built based on repeated regions that can generate chimeric artifacts (Trombetti et al. 2007).

Microsatellites are relatively common markers throughout the genome of most species (Li et al. 2002); thus, they can be efficiently isolated from genome sequencing of non-enriched (Abdelkrim et al. 2009) and enriched libraries (Santana et al. 2009). However, the implementation of library enrichment is more cost-effective due to the significantly smaller sequencing effort necessary to obtain large numbers of microsatellite markers (Andrés and Bogdanowicz 2011). Even though the enrichment approach is inherently biased towards a subset of marker motifs, when microsatellite markers are the tools and not the subject of the study (e.g., genome evolution), markers identified from enriched libraries are suitable and informative for genetic studies (Väli et al. 2008).

454 sequencing provides a time- and cost-effective approach for the identification of de novo microsatellites in non-model organisms. Time for library preparation, sequencing, and bioinformatics can be reduced to 3 weeks with the implementation of this new technology (Table I). Though effective, the cloning/Sanger sequencing protocol is more labor-intensive and requires weeks to months of laboratory work due to the need for individual sequencing of each positive clone detected. In terms of project costs, this study shows that the 454 protocol can decrease the cost per sequence to less than 1/3 of the cost of the cloning/Sanger sequencing protocol by using multiple enriched libraries pooled together in the same 454 run (Table I). Therefore, the increasing implementation of next-generation sequencing technology should accelerate the development of de novo markers for species that lack genomic resources.

ACKNOWLEDGMENTS

We would like to thank L. Duque, D. Dewey, J. Eitner, and J. Stitt for field assistance and A. Soro and two anonymous reviewers for comments on this manuscript. This work was supported by the Andrew W. Mellon Foundation at Cornell University (MMLU), the Sarah Bradley Fellowship (MMLU), and the Cornell Biology Research Fellowship Program (CKS). Additional funds were provided by National Science Foundation systematic grants (DEB-0814544 and DEB-0742998 to BND).

Découverte et caractérisation de microsatellites de l'abeille solitaire *Colletes inaequalis*, par utilisation de la méthode de Sanger et du pyroséquençage 454

SSR/ clonage/séquençage nouvelle génération/Colletidae

Entdeckung und Charakterisierung von Mikrosatelliten für die Solitärbiene *Colletes inaequalis* durch Sanger-Sequenzierung und 454 Pyrosequenzierung

SSR/Klonierung/next-generation Sequenzierung/Colletidae

REFERENCES

Abdelkrim, J., Robertson, B., Stanton, J.-A., Gemmell, N. (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. Biotechniques **46**, 185–192

Andrés, J.A., Bogdanowicz, S.M. (2011) Isolating microsatellite loci: looking back, looking ahead. In: Orgogozo, V., Rickman, M.V. (eds.) Molecular Methods for Evolutionary Genetics, pp. 211–232. Springer Protocols, Secaucus

Augusto, S.C., Gonçalves, P.H.P., Francisco, F.O., Santiago, L.R., Françoso, E.A., Suzuki, K.M., Sofia,

S.H., Simões, Z.L.P., Arias, M.C. (2011) Microsatellite loci for the carpenter bee *Xylocopa frontalis* (Apidae, Xylocopini). Conser. Genet. Resour. **4**, 315–317

Azuma, N., Takahashi, J., KIidoro, M., Higashi, S. (2005) Isolation and characterization of microsatellite loci in the bee *Ceratina flavipes*. Mol. Ecol. Notes **5**, 433–435

Batra, S.W.T. (1980) Ecology, behavior, pheromones, parasites and management of the sympatric vernal bees *Colletes inaequalis*, *C. thoracicus* and *C. validus*. J. Kans. Entomol. Soc. **53**, 509–538

Batra, S.W.T. (1995) Bees and pollination in our changing environment. Apidologie **26**, 361–370

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., Bonhomme, F. (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Montpellier (France): Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II

Beveridge, M., Simmons, L.W. (2004) Microsatellite loci for Dawson's burrowing bee (*Amegilla dawsoni*) and their cross-utility in other *Amegilla* species. Mol. Ecol. Notes. **4**, 379–381

Blair, M.W., McCouch, S.R. (1997) Microsatellite and sequence-tagged site markers diagnostic for the rice bacterial leaf blight resistance gene *xa*-5. Theor. Appl. Genet. **95**, 174–184

Brownstein, M.J., Carpten, J.D., Smith, J.R. (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. Biotechniques **20**, 1004–1010

Danforth, B.N., Ji, S., Ballard, L.J. (2003) Gene flow and population structure in an oligolectic desert bee, *Macrotera* (*Macroteropsis*) *portalis*. J. Kans. Entomol. Soc. **76**, 221–235

Gardner, K.E., Ascher, J.S. (2006) Notes on the native bee pollinators in New York apple orchard. J. N. Y. Entomol. Soc. **114**, 86–91

Hamilton, M.B., Pincus, E.L., Di Fiore, A., Fleischer, R.C. (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. Biotechniques **27**, 500–504

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., Welch, D.M. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol **8**, R143

Kapheim, K.M., Pollinger, J.P., Wcislo, W.T., Wayne, R.K. (2009) Characterization of 12 polymorphic microsatellite markers for a facultatively eusocial sweat bee (*Megalopta genalis*). Mol. Ecol. Resour. **9**, 1527–1529

Kukuk, P.F., Forbes, S.H., Zahorchack, R., Riddle, A., Pilgrim, K. (2002) Highly polymorphic microsatellite markers developed for the social halictine bee *Lasioglossum* (*Chilalictus*) *hemichalceum*. Mol. Ecol. Notes **2**, 529–530

Langer, P., Molbo, D., Keller, L. (2004) Polymorphic microsatellite loci in Allodapine bees for investigating the evolution of social behaviour. Mol. Ecol. Notes **4**, 303–305

Li, Y.C., Korol, A.B., Fahima, T., Beiles, A., Nevo, E. (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol. Ecol. **11**, 2453–2465

López-Uribe, M.M., Green, A.N., Ramírez, S.R., Bogdanowicz, S.M., Danforth, B.N. (2010) Isolation and cross-species characterization of polymorphic microsatellites for the orchid bee *Eulaema meriana* (Hymenoptera: Apidae: Euglossini). Conser. Genet. Resour. **1**, 21–23

Mikheyev, A.S., Vo, T., Wee, B., Singer, M.C., Parmesan, C. (2010) Rapid microsatellite isolation from a butterfly by *de novo* transcriptome sequencing: performance and a comparison with AFLP-derived distances. PLoS One **5**, e11212

Mohra, C., Fellendorf, M., Segelbacher, G., Paxton, R.J. (2000) Dinucleotide microsatellite loci for *Andrena vaga* and other andrenid bees from non-enriched and CT-enriched libraries. Mol. Ecol. **9**, 2189–2191

Neumann, K., Seidelmann, K. (2006) Microsatellites for the inference of population structures in the Red Mason bee *Osmia rufa* (Hymenoptera, Megachilidae). Apidologie **37**, 75–83

Paxton, R.J., Thorén, P.A., Tengö, J., Estoup, A., Pamilo, P. (1996) Mating structure and nestmate relatedness in a communal bee, *Andrena jacobi* (Hymenoptera, Andrenidae), using microsatellites. Mol. Ecol. **5**, 511–519

Paxton, R.J., Ayasse, M., Field, J., Soro, A. (2002) Complex sociogenetic organization and reproductive skew in a primitively eusocial sweat bee, *Lasioglossum malachurum*, as revealed by microsatellites. Mol. Ecol. **11**, 2405–2416

Paxton, R.J., Zobel, M.U., Steiner, J., Zillikens, A. (2009) Microsatellite loci for *Euglossa annectans* (Hymenoptera: Apidae) and their variability in other orchid bees. Mol. Ecol. Resour. **9**, 221–1223

Perry, J.C., Rowe, L. (2011) Rapid microsatellite development for water striders by next-generation sequencing. J. Hered. **102**, 125–129

Santana, Q., Coetzee, M., Steenkamp, E., Mlonyeni, O., Hammond, G., Wingfield, M., Wingfield, B. (2009) Microsatellite discovery by deep sequencing of enriched genomic libraries. Biotechniques **46**, 217–223

Schlötterer, C., Tautz, D. (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res. **20**, 211–215

Schuelke, M. (2000) An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. **18**, 233–234

Smit, A.F.A., Hubley, R., Green, P. (2010) *RepeatMasker Open-3.0*. Available from: http://www.repeatmasker.org

Soro, A., Paxton, R.J. (2009) Characterization of 14 polymorphic microsatellite loci for the facultatively eusocial sweat bee *Halictus rubicundus* (Hymenoptera, Halictidae) and their variability in related species. Mol. Ecol. Resour. **9**, 150–152

Souza, R.O., Cervini, M., Del Lama, M.A., Paxton, R.J. (2007) Microsatellite loci for euglossine bees (Hymenoptera: Apidae). Mol. Ecol. Notes **7**, 1352–1356

Tautz, D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. **17**, 6463–6471

Trombetti, G.A., Bonnal, R.J.P., Rizzi, E., De Bellis, G., Milanesi, L. (2007) Data handling strategies for high throughput pyrosequencers. BMC Bioinforma. **8**, S22

Väli, U., Einarsson, A., Waits, L., Ellegren, H. (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Mol. Ecol. **17**, 3808–3817

Vereecken, N.J. (2008) Pollinator-mediated selection, reproductive isolation and floral evolution in *Ophrys* orchids. Proc. Neth. Entomol. Soc. Meet. **19**, 9–21

Weber, J.L., Polymeropoulos, M.H., May, P.E., Kwitek, A.E., Xiao, H., McPherson, J.D., Wasmuth, J.J. (1991) Mapping of human chromosome 5 microsatellite DNA polymorphisms. Genomics **11**, 695–700

Winfree, R., Williams, N.M., Dushoff, J., Kremen, C. (2007) Native bees provide insurance against ongoing honey bee losses. Ecol. Lett. **10**, 1105–1113

Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R., Simon, P. (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am. J. Bot. **99**, 193–208

Zane, L., Bargelloni, L., Patarnello, T. (2002) Strategies for microsatellite isolation: a review. Mol. Ecol. **11**, 1–16

Zayed, A. (2006) Characterization of microsatellite loci from the solitary sweat bees *Lasioglossum leucozonium* and *Lasioglossum oenotherae* (Hymenoptera, Halictidae). Mol. Ecol. Notes **6**, 1154–1156

Zayed, A. (2009) Bee genetics and conservation. Apidologie **40**, 237–262

Zayed, A., Constantin, Ş.A., Packer, L. (2007) Successful biological invasion despite a severe genetic load. PLoS One **9**, e868