Development of Multiple Polymorphic Microsatellite Markers for Ceratina calcarata (Hymenoptera: Apidae) Using Genome-Wide Analysis

Wyatt A. Shell and Sandra M. Rehan

Abstract

The small carpenter bee, Ceratina calcarata (Robertson), is a widespread native pollinator across eastern North America. The behavioral ecology and nesting biology of C. calcarata has been relatively well-studied and the species is emerging as a model organism for both native pollinator and social evolution research. C. calcarata is subsocial: reproductively mature females provide extended maternal care to their brood. As such, studies of C. calcarata may also reveal patterns of relatedness and demography unique to primitively social Hymenoptera. Here, we present 21 microsatellite loci, isolated from the recently completed C. calcarata genome. Screening in 39 individuals across their distribution revealed that no loci were in linkage disequilibrium, nor did any deviate significantly from Hardy-Weinberg following sequential Bonferroni correction. Allele count ranged from 2 to 14, and observed and expected heterozygosities ranged from 0.08 to 0.82 (mean 0.47) and 0.26 to 0.88 (mean 0.56), respectively. These markers will enable studies of population-wide genetic structuring across C. calcarata’s distribution. Such tools will also allow for exploration of between and within-colony relatedness in this subsocial native pollinator.

Key words: Ceratina calcarata, microsatellite marker, Illumina

Microsatellite markers are popular and frequently employed for studies of relatedness and population genetics. Owing to their high mutation rate and population variability, microsatellite loci can be targeted to reveal subtle changes in population structure and composition, kinship, patterns of paternity, and heritability (reviewed in Powell et al. 1996; Sunnucks 2000). The design and optimization of these powerful molecular tools has been improved and greatly expedited by next-generation sequencing (Grover and Sharma 2016). Given their multiscalar capacity to reveal patterns of both population and family structure, the development of microsatellite markers remains an informative and important endeavor.

Bees are represented by over 20,000 described species and occur on all continents except Antarctica (Michener 2007). Highly efficient pollinators, bees make a significant contribution to the productivity of both agricultural and natural systems (Kremen et al. 2002; Klein et al. 2007; Brittain et al. 2012). Fine scale population research has revealed a great deal about how bees affect and are affected by ecological conditions (Cameron et al. 2011; Bartomeus et al. 2013). Such studies have helped distinguish subspecies (e.g. Melipona spp. Tavares et al. 2013), population distribution (Bombus spp. Geib et al. 2015), effects of land use practices (Dreier et al. 2014), as well as conflicts between managed and wild bee populations (Moreira et al. 2015). The development of microsatellite markers thus allows for the comprehensive study of bee biology and demography at a macroscopic scale, and informs our ability to implement biologically meaningful pollinator conservation practices.

The small carpenter bee, Ceratina calcarata (Robertson), is one of five very recently diverged and largely sympatric species of Ceratina found across eastern North America (Rehan and Sheffield 2011; Shell and Rehan 2016). C. calcarata is a generalist pollinator (McFrederick and Rehan 2016) and, given its broad range and high abundance, contributes to the productivity of a large number of ecological and agricultural systems. C. calcarata is also subsocial: females provide extended maternal care to their brood, and defend and clean their offspring into adulthood (Rehan and Richards 2010). Subsocial behavior is considered foundational to the evolution of more complex social forms (reviewed in Rehan and Toth 2013); as such, C. calcarata is also emerging as a model organism for studies of social evolution (Rehan et al. 2014).

Significant molecular resources are available for C. calcarata in the form of an annotated transcriptome (Rehan et al. 2014), methylome and genome (Rehan et al. 2016). Among many powerful and
practical applications, such data avails the rapid and reliable development of molecular markers. Here, we isolated microsatellite loci from the *C. calcarata* genome, and optimized a suite of 21 polymorphic markers in 39 individuals from across the species’ range. These primers make available multiscalar population genetics studies, and will allow researchers to investigate relatedness and patterns of parentage in this unique subsocial pollinator.

**Materials and Methods**

Microsatellite loci were isolated from the recently published *C. calcarata* genome (Rehan et al. 2016) using the Microsatellite Identification Tool (MISA; Thiel et al. 2003) interfaced with an executable version of Primer 3 (Koressaar and Remm 2007; Untergasser et al. 2012). MISA was used to trim genomic reads to lengths of 20 kb to facilitate scans. To ensure microsatellite quality, MISA was configured to select loci based on strict minimum motif repeat requirements (mononucleotides = 15; dinucleotides = 7; tri-through hexanucleotides = 5). The resulting 2,010 putative microsatellite loci and flanking regions were sorted in descending order of motif length and motif repeat count. Flanking regions were then visually inspected for self-complementarity to prevent hairpin or excessive primer dimer formation. Putative primer pairs were also screened to ensure $T_a < 1/C_{14}$ difference in melting temperature between forward and reverse oligos.

A top fifty candidate primers were then assessed for amplification performance and polymorphism in 16 females from across *C. calcarata*’s range (*C. calcarata* is haplodiploid, thus primers must be screened in females to accurately assess heterozygosity). We followed the methodology of Schuelke (2000) and designed forward primers modified with a partial M13 tail. This M13 oligo extension

| Table 1. Primer sequences and locus characteristics of twenty-one microsatellite loci developed for *C. calcarata* |
|---------------------------------------------------------------|
| Locus | Primer Sequence 5’-3’ | GenBank | Repeat Motif | k | Allele size range (T_a) | N | H_o | H_e | HWE P |
|-------|-----------------------|---------|--------------|---|------------------------|---|------|------|-------|
| Ccal01 | F: ACAAAACAAAGCGGGCACA KU945359 GGTGAC 8 | 263–311 | 69 | 0.658 | 0.752 | 0.011 |
|       | R: GGATTGTGTCGAGGCGGGAG | | | | | |
| Ccal02 | F: AAATCAACCTACGCCCCAGC KU945360 CAGCTC 5 | 226–280 | 64 | 0.308 | 0.409 | 0.259 |
|       | R: TACACACAGGTGTCAGTG | | | | | |
| Ccal03 | F: AATAGGCCGAGAGCAGCCG R: TTGTTCATCTTGCCAGCCG | KU945361 AGGCAG 7 | 152–188 | 64 | 0.526 | 0.664 | 0.220 |
|       | | | | | | |
| Ccal04 | F: GGAGAAACGAGATACCCAGG R: TCCCACTTTTAACGGCTCCC | KU945362 ACCGA 2 | 91–96 | 67 | 0.083 | 0.579 | 0.367 |
|       | | | | | | |
| Ccal08 | F: TCGATTCAGCAGAAGCTGAC R: GAGATTCGCACCGCTAAT | KU945363 CTGA 11 | 235–287 | 68 | 0.579 | 0.854 | 0.004 |
|       | | | | | | |
| Ccal11 | F: ATAGGAGAGAGCCGCTTTCG R: TGTGCCAGACCATACAAAT | KU945364 AGGTT 6 | 243–278 | 65 | 0.633 | 0.698 | 0.004 |
|       | | | | | | |
| Ccal14 | F: GGCGTGTGAGAAACACGAAGC R: AGGCTGTGCGTCTGAA | KU945365 AACCT 3 | 164–174 | 66 | 0.308 | 0.406 | 0.005 |
|       | | | | | | |
| Ccal16 | F: CAGGGAGCCGCTTACCTTTT R: GGCAGGAGGATACTCGGC | KU945366 AGGTT 3 | 252–262 | 67 | 0.648 | 0.502 | 0.072 |
|       | | | | | | |
| Ccal17 | F: GTGCCGCAGTGAGAACCAAGC R: AGGCTGTGCGTCTGAA | KU945367 GGCAG 5 | 195–215 | 67 | 0.179 | 0.295 | 0.017 |
|       | | | | | | |
| Ccal18 | F: GTCATTTCGTCGTCGACG | KU945368 GTTCT 2 | 235–240 | 66 | 0.256 | 0.26 | 1.000 |
|       | R: CTGAGCCGCTATCGCATA | | | | | |
| Ccal19 | F: TCATTATCTCTAGGGGGGTCG R: CTGCCATCTCTGTCCTCGT | KU945369 GAACA 6 | 261–271 | 67 | 0.316 | 0.508 | 0.020 |
|       | | | | | | |
| Ccal23 | F: AATTCGGCCAAGCTTCTGACA R: GGAACTGTGCTGTGGGCCC | KU945370 GTGCCG 6 | 163–188 | 70 | 0.795 | 0.585 | 0.157 |
|       | | | | | | |
| Ccal25 | F: ACGGCGGAGTTACAAAAACG R: ACTTGTAGGTTGCTCAAAT | KU945371 CCGCA 7 | 188–218 | 69 | 0.436 | 0.645 | 0.003 |
|       | | | | | | |
| Ccal29 | F: AGCTTGTGAGACACATGACA R: CGGGTGCTCCCTAAATCAC | KU945372 AACCT 3 | 270–280 | 64 | 0.359 | 0.331 | 1.000 |
|       | | | | | | |
| Ccal30 | F: TCTATGAGTAGTTGCCGCGG R: GACGACTGTTGGCCAGATAC | KU945373 ATCAT 5 | 267–302 | 63 | 0.436 | 0.656 | 0.015 |
|       | | | | | | |
| Ccal37 | F: CGTCTCCTACGAACGATAC R: AGGAACTGGTGGCCCGTCTAC | KU945374 AGAA 12 | 148–196 | 64 | 0.769 | 0.868 | 0.056 |
|       | | | | | | |
| Ccal39 | F: CAGAAATGTGGCCTCGGACR | KU945375 TTAT | 253–277 | 65 | 0.553 | 0.765 | 0.010 |
|       | | | | | | |
| Ccal44 | F: TTCACCGTGATGCTGTCAC R: AGGTCTGAGATGCTGTCAC | KU945376 GTCT 6 | 201–297 | 67 | 0.289 | 0.319 | 0.087 |
|       | | | | | | |
| Ccal48 | F: CGATCGTGTTGAGGAACGCAAG R: CTGCCTTCTCTCTCTCTCT | KU945377 GAGAA 6 | 106–126 | 65 | 0.513 | 0.584 | 0.049 |
|       | | | | | | |
| Ccal49 | F: CTGCCCTTCTCTCTCTCTCTCT | KU945378 GCAC 8 | 234–262 | 65 | 0.744 | 0.787 | 0.571 |
|       | R: GAGAGCGGCGGCTAATAAG | | | | | |
| Ccal50 | F: CCACACTCTCTCTCTCAGG R: TCTCTGTCTCTCTCTCTCAGG | KU945379 TGTA 14 | 227–283 | 69 | 0.821 | 0.879 | 0.028 |

Information presented includes primer sequence; GenBank Accession Number; repeat motif; allele count (k); allele size range; annealing temperature in °C ($T_a$); number of individuals successfully screened (N); observed (H_o) and expected (H_e) heterozygosities; and Hardy-Weinberg exact test P-values (HWE P).
allowed fluorescently dyed M13 probes to be incorporated during PCR. A set of universal primers was labeled with three dyes from the D5-33 set (FAM, PET, and VIC) to allow for PCR product multiplexing downstream. PCR reactions were executed in an 11 µl volume (5.45 µl dH2O; 2.0 µl 5× HF Buffer; 0.2 µl [10 mM] dNTPs; 0.1 µl Phusion HF Taq Polymerase (Thermo Scientific, Waltham, MA); 0.25 µl [10 mM] forward primer; 0.5 µl Fluorescent M13 oligo [10 mM]; 0.5 µl [10 mM] reverse primer; 2.0 µl DNA template) using an Eppendorf Mastercycler gradient thermocycler (Eppendorf North America, Hauppauge, NY). PCR reactions involved five stages: 1) initial denaturing at 98 °C for 40 s, 2) a touchdown series of 10 cycles at 98 °C for 10 s, 72 °C for 15 s (cooling incrementally to primer-specific Tm), and 72 °C for 15 s, 3) 20 cycles at 98 °C for 10 s, primer Tm for 15 s, and 72 °C for 15 s, 4) 8 cycles at 98 °C for 10 s, 62 °C for 15 s, and 72 °C for 15 s, 5) final extension at 72 °C for 10 min. PCR products were mixed with HiDi Formamide (Applied Biosystems, Foster City, CA, USA) before being sent to the DNA Analysis Facility at Yale University for fragment analysis on a 3730xl Analyzer. Alleles were manually scored using Peak Scanner 2 (Applied Biosystems, Foster City, CA, USA).

After initial PCR, 21 loci were further screened for polymorphism and performance in an additional 23 females (total screening panel N = 39). Loci were tested for Hardy-Weinberg and linkage disequilibrium using GenePop 4.2 (Raymond and Rousset 1995; Rousset 2008). Each locus was then assessed for expected and observed heterozygosities ranging from 0.08 for the genus using a reference genome. No loci were found to be in linkage disequilibrium using an Eppendorf Mastercycler gradient thermocycler (Eppendorf North America, Hauppauge, NY). PCR reactions involved five stages: 1) initial denaturing at 98 °C for 40 s, 2) a touchdown series of 10 cycles at 98 °C for 10 s, 72 °C for 15 s (cooling incrementally to primer-specific Tm), and 72 °C for 15 s, 3) 20 cycles at 98 °C for 10 s, primer Tm for 15 s, and 72 °C for 15 s, 4) 8 cycles at 98 °C for 10 s, 62 °C for 15 s, and 72 °C for 15 s, 5) final extension at 72 °C for 10 min. PCR products were mixed with HiDi Formamide (Applied Biosystems, Foster City, CA, USA) before being sent to the DNA Analysis Facility at Yale University for fragment analysis on a 3730xl Analyzer. Alleles were manually scored using Peak Scanner 2 (Applied Biosystems, Foster City, CA, USA).

When microsatellite markers are affordable and adaptable tools, suited to a wide range of research.

Results and Discussion

This suite of 21 microsatellite loci are the first primers developed for a member of the New World Ceratina, and the first set developed for the genus using a reference genome. No loci were found to be in linkage disequilibrium, nor did they diverge significantly from Hardy-Weinberg equilibrium following sequential Bonferroni correction. Observed and expected heterozygosities ranged from 0.08 to 0.82 (mean 0.47) and 0.26 to 0.88 (mean 0.56), respectively (Table 1). Our results are thus very similar to those of other recent hymenopteran microsatellite development projects (Rabeling et al. 2014; Chen et al. 2015; Vickruck 2015). These loci will be informative in exploring the population structure, patterns of parentage, and kinship dynamics in C. calcarata. As C. calcarata is closely related to four other native Ceratina species (Rehan and Sheffield 2011; Shell and Rehan 2016), these loci may also cross-amplify (as in Ceratina flavipes, Azuma et al. 2005; and Halictus rubicundus (Christ), Soro and Paxton 2009) allowing for genus-wide research and conservation.

Additionally, these microsatellite loci can be powerful markers for understanding the evolution of social structure. Microsatellite markers have been used to reveal variation in queen-worker dynamics by geography (Richards et al. 2005), and deviations from expected worker kinship in a eusocial sweat bee (Lasio glossum malacburum (Kirby), Soro et al. 2009). Such markers have also been used to reveal mating structure in a communal bee (Andrena jacobi (Perkins), Paxton et al. 1996) and polymorphism in reproductive strategy and sociality among populations of Halictus scabiosae (Rossi) (Ulrich et al. 2009). As C. calcarata is subsocial (Rehan and Richards 2010) it likely represents an early stage in the evolution of eusociality (Rehan and Toth 2015). By employing even a subset of our 21 markers in a study of C. calcarata inter- and intracolony relatedness, we may be able to uncover similarly cryptic social dynamics in a bee putatively on the cusp of incipiently social reproductive behavior (Rehan et al. 2014).

Protocols for the discovery and optimization of microsatellite loci are numerous and have evolved over decades of population genetics studies (e.g. Glenn and Schable 2005; reviewed in Zane et al. 2002). To secure even one useful microsatellite was originally a laborious procedure with few guarantees: isolated loci were random, or limited to sites complimentary to specially designed probes (Ostrander et al. 1992; Queller et al. 1993; Kijas et al. 1994). Whole genome sequencing technologies have made great advancements in quality and accessibility of genetic resources over the past decade (Hudson 2008; Ekblom and Galindo 2011; vanDijk et al. 2014). Improved availability of such powerful resources has greatly expedited and improved the generation of microsatellite primers for many novel species (e.g. copperhead snakes, Castoe et al. 2010; water striders, Perry and Rowe 2011). The generation and analysis of whole genomes is still an expensive and bioinformatically exacting endeavor (reviewed in Lemmon and Lemmon 2013); by contrast, microsatellite markers are affordable and adaptable tools, suited to a wide range of research.

Acknowledgments

We would like to thank Krista Ciaccio, Dusty Durant, Sean Lombard, Nicholas Piza, and Jacob Withie for assistance with field collections. We also thank Michael Steffen and Sarah Lawson for their helpful comments in preparing this manuscript. Funding from the University of New Hampshire and the National Science Foundation (Behavioral Systems Award no. IOS-1456296) supported this work. Partial funding was provided by the New Hampshire Agricultural Experiment Station. This is Scientific Contribution Number 2674. This work was supported by the USDA National Institute of Food and Agriculture Hatch Project 1004515.

References Cited

Azuma, N., J. Tagkahashi, M. Kidokoro, and S. Higashi. 2005. Isolation and characterization of microsatellite loci in the bee Ceratina flavipes. Mol Ecol Notes 5: 433–435.

Bartomeus, I., J. S. Ascher, J. Gibbs, B. N. Danforth, D. L. Wagner, S. M. Hedrle, and R. Winfree. 2013. Historical changes in northeastern US bee pollinators related to shared ecological traits. Proc Natl Acad Sci USA 110: 4656–4660.

Brittain, C., N. Williams, C. Kremen, and A. M. Klein. 2012. Synergistic effects of non-Apis bees and honey bees for pollination services. P R Soc B 280: 20122767.

Cameron, S. A., J. D. Lozier, J. P. Strange, J. B. Koch, N. Cordes, L. F. Soltner, and T. L. Griswold. 2011. Patterns of widespread decline in North American bumble bees. Proc Natl Acad Sci USA 108: 662–667.

Castoe, T. A., A. W. Poole, W. Gu, A. P. J. De Koning, J. M. Daza, E. N. Smith, and D. D. Pollock. 2010. Rapid identification of thousands of copperhead snake (Agkistrodon contortrix) microsatellite loci from modest amounts of 454 shotgun genome sequence. Mol Ecol Resour 10: 341–347.

Chen, W., L. Fang, L. Liu, Z. He, and H. Y. Hu. 2015. Isolation and characterization of polymorphic microsatellite loci for Pachyrhizodus vindemmiae (Rondani) (Hymenoptera: Pteromalidae). Mol Gen Mol Biol 41: 1798–1801.

Dreier, S., J. W. Redhead, I. A. Warren, A. F. G. Bourke, M. S. Heard, W. C. Bartomeus, I., J. S. Ascher, J. Gibbs, B. N. Danforth, D. L. Wagner, S. M. Hedrle, and R. Winfree. 2013. Historical changes in northeastern US bee pollinators related to shared ecological traits. Proc Natl Acad Sci USA 110: 4656–4660.

Ekblom, R., and J. Galindo. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107: 1–15.
