RESEARCH

Isolation and Characterization of Microsatellite Markers Useful for Exploring Introgression Among Species in the Diverse New Zealand Cicada Genus *Kikihia*

Elizabeth J. Wade[1,2] and Chris Simon[1.3]

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, 75 North Eagleville Rd., CT 06269
[2]Corresponding author, e-mail: elizabeth.wade@uconn.edu
[3]School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand

ABSTRACT. The New Zealand cicada genus *Kikihia* Dugdale 1971 exhibits more than 20 contact zones between species pairs that vary widely in their divergence times (between 20,000 and 2 million years) in which some level of hybridization is evident. Mitochondrial phylogenies suggest some movement of genes across species boundaries. Biparentally inherited and quickly evolving molecular markers like microsatellites are useful for assessing gene flow levels. Here, we present six polymorphic microsatellite loci that amplify DNA from seven species across the genus *Kikihia*; *Kikihia* "northwestlandica," *Kikihia* "southwestlandica," *Kikihia muta, Kikihia angusta*, *Kikihia* "tuta," *Kikihia* "nelsonensis," and *Kikihia* "murihikua." The markers were developed using whole-genome shotgun sequencing on the 454 pyrosequencing platform. Moderate to high levels of polymorphisms were observed with 14–47 alleles for 213 individuals from 15 populations. Observed and expected heterozygosity range from 0 to 1 and 0.129 to 0.945, respectively. These new markers will be instrumental for the assessment of gene flow across multiple contact zones in *Kikihia*.

Key Words: Cicadidae, codominant genetic marker, population genetics

Kikihia (Dugdale 1971) is a monophyletic endemic New Zealand genus of cicadas (Buckley et al. 2002, Arensburger et al. 2004a). Currently, there are 14 described species and 16 yet-to-be-described species and subspecies (Dugdale 1971; Fleming 1984; Marshall et al. 2008, 2011; Larivière et al. 2010). Species of this cicada genus are found on the three main islands of New Zealand including North, South, and Stewart islands as well as smaller near-by islands and the more distantly located Kermadec and Norfolk islands. *Kikihia* is widespread with species found in most terrestrial habitats including grass, shrub, and forest ecosystems from lowland to subalpine regions of New Zealand.

Phylogenetic and phylogeographic studies using mitochondrial and nuclear gene sequences have provided insight into the evolutionary history of the genus (Arensburger et al. 2004a,b; Marshall et al. 2008, 2009, 2011). Extensive field study and molecular investigation into this genus identified 20 contact zones between different species that are possible sites of hybridization (Marshall et al. 2011). Individuals found in these proposed hybrid zones possess morphological and acoustic traits that are intermediate between parental species found away from the hybrid zones. Although species in the genus are estimated to have diverged more than 6 million years ago, hybridization seems confined to species that diverged 2 million years ago or less (Marshall et al. 2008).

Here, we present six newly discovered microsatellite markers to investigate the population dynamics, levels of gene flow, and impact of introgression on species boundaries throughout this diverse cicada genus. Because processes related to speciation are difficult to observe after the fact, the study of the interaction at contact zones between incipient species pairs or between recently diverged species provides insight into the speciation process (Barton and Hewitt 1989, Harrison 1993, Mallet 2007, Hewitt 2011). We have developed six novel highly polymorphic microsatellites that amplify in seven species of *Kikihia*; *Kikihia* "northwestlandica," *Kikihia* "southwestlandica," *Kikihia muta* (Fabricius, 1775), *Kikihia angusta* (Walker, 1850), *Kikihia* "tuta," *Kikihia* "nelsonensis," and *Kikihia* "murihikua." Data for pure populations (as far from hybrid zones as possible) are used to test the microsatellite markers' ability to differentiate these species.

Materials and Methods

Specimen Collection. Cicada specimens were identified in the field based on morphology and song traits. Taxonomic descriptions are pending for many *Kikihia* species as denoted by informal names in quotes. Populations located as far from putative hybrid zones as possible were chosen as species-typical or "parental" populations. Subsequently, 24–36 individuals per species from 1 to 5 largely species-typical populations were investigated to test these new microsatellite markers (Table 1). Whole body specimens were placed in 95% ethanol or three legs were removed and stored in 95% ethanol and the bodies were pinned. All ethanol specimens are stored at −20°C.

Microsatellite Development. Nine species of *Kikihia* from throughout the genus were used to develop microsatellites: *K.* "southwestlandica," *K. muta, K. angusta, K.* "nelsonensis," *Kikihia cutora cutora, Kikihia scutellaris, Kikihia* "peninsularis," *Kikihia horologium*, and *Kikihia* "aotea western." Genomic DNA was extracted from cicada legs using the Qiagen DNeasy Blood & Tissue kit (Qiagen, Boston, MA). RNA was removed with 20 µg RNAase A (New England Biolabs, Ipswich, MA) per sample during DNA purification. The 454 GS FLX Rapid Library Preparation kit (454 Life Sciences, Branford, CT) with Rapid Library multiplex identifiers (MIDs) were used for one specimen from each of the nine species. MIDs are unique sequences incorporated in the 454 primers for the purpose of identification. Sequencing was done unidirectionally. Sample preparation and sequencing were performed according to the manufacturer's instructions (454 Life Sciences, Branford, CT). One library (*K. muta*) was sequenced on 1/16th of a GS FLX Titanium PicoTiter Plate. The eight remaining libraries were pooled and sequenced on one region of a two-region GS FLX Titanium PicoTiter Plate. GS De Novo Assembler v2.3 was used to align reads with the heterogeneous and large genome options. This resulted in 3,683 contigs with an average size of 743 bp and a range of 100–5,909 bp. MsatCommander (Faircloth 2008, Abdelkrim et al. 2009), a program that searches for di- tri-, and tetranucleotide repeats with sufficient flanking sequence for primer design, was used to scan contigs. Primer3 (Untergrasser et al. 2012) was used to design primers.

Table 1. *Kikihia* species and populations used for microsatellite discovery

Species	N	Site Code	Lat	Long	E	Location description
NW	29	NN.KOB	−41.11	172.10	28	Kohaihai Bluff, 16.3 km north of Karamea
SW	33	WD.OKT	−43.23	170.16	6	Okarito, beach pulloff
Mur	18	OL.RSC	−44.50	168.78	378	3.4 km from Raspberry Flat/Mt. Aspiring Trailhead
	6	OL.NLW	−44.48	169.25		W side of Lake Hawea, SH6, north of Wanaka
Muta	10	MB.GRO	−41.48	173.96	15	2.5 km N. of Opawa River on SH1, S. of Grovestown
	4	MB.WCR	−41.42	173.88	43	Waikakaho Creek Rd 3.7 km north of Kaituna-Tuamanina Rd
	7	MB.LVR	−41.37	173.76	44	Long Valley Road at River (Kaituna)
	7	MB.LVB	−41.37	173.74	63	Long Valley Road at bridge
	5	MB.LVM	−41.36	173.73	64	Long Valley Road 3 km west of Kaituna River
Nel	20	NN.HRN	−41.00	172.89	681	2.6 km north of SH60 on Canaan Rd (Harwood hole road)
	16	NN.WGB	−41.45	173.08	144	10.8 km East of Wakefield on Pig Valley Rd
Tuta	16	NN.BES	−41.28	173.15		Rabbit Island, boat launch at Best's Rd
	17	SD.TAP	−41.29	173.67	25	Junction of SH6 and Tapps Rd at Canvastown
Ang	9	MB.BAV	−42.16	172.82	1263	North of Island Saddle, on the Rainbow/Hanmer Rd.
	16	SC.IDN	−44.78	170.87	141	Ritchies Rd, SOUTH of Douglas

Species NW, *K.* "northwestlandica"; SW, *K.* "southwestlandica"; Mur, *K.* "murihikua"; Muta, *K. muta, K.* "nelsonensis," *K.* "tuta," *K. angusta*; *N*, number of specimens per population; site code, the first two letters represent New Zealand district codes and the last three letters are unique collecting location codes; Lat, latitude; long, longitude; *E*, elevation in meters

In total, 30 primer pairs were screened. Six primer pairs were identified representing variable microsatellite loci that amplified in all species. These primer pairs were tested in multiple *Kikihia* populations.

Fragment Analysis. Polymerase chain reaction (PCR) primer pairs were tested with the forward primer tagged with a M13 sequence on the 5'-end and the reverse primers for each locus plus a 6-FAM-labeled M13 primer (Schuelke 2000). PCR amplifications were performed in 15 µl reactions containing 0.4 U *Ex Taq* DNA polymerase (TaKaRa Biomedical Inc., Otsu, Shiga, Japan), 1x PCR buffer, 0.3 mM dNTP mix, 3 pmol forward primer, 12 pmol reverse primer, 12 pmol M13 primer with a 5' 6-FAM fluorescent label reactions, and up to 100 ng template DNA. PCR conditions were as follows: an initial 94°C for 5 min followed by 30 cycles of 94°C for 30 s, 57°C for 45 s, 72°C for 45 s, followed by 8 cycles of 94°C for 30 s, 53°C for 45 s, 72°C for 45 s, followed by a final extension of 72°C for 10 min. PCR was diluted 1:9 in HiDi Formamide (Invitrogen Life Technologies, Grand Island, NY) and run on an ABI 3130xl DNA sequencer (Applied Biosystems, Grand Island, NY). Alleles were designated according to amplicon size relative to LIZ 600 size standard (Invitrogen Life Technologies, Grand Island, NY). PCR forward primers that amplified specimens from across the *Kikihia* phylogeny and revealed variable microsatellites were then fluorescently labeled with the G5 (Invitrogen Life Technologies, Grand Island, NY) labels (Table 2). Testing of *Kikihia* populations for the six described primer pairs were multiplex amplified using the Qiagen Type-it microsatellite PCR kit according to the manufacturer's instructions (Qiagen, Boston, MA), including an annealing temperature of 57°C for all multiplexed PCRs.

Microsatellite Analysis. GeneMarker v2.2.0 (SoftGenetics, LLC, Kalamazoo, MI) was used to visualize and score alleles.

Files were converted between different formats using CONVERT v1.31 (Glaubitz 2004). Arlequin v3.5.1.2 (Excoffier et al. 2005) was used to assess the number of alleles per locus, the allele frequencies, and expected and observed heterozygosity. Deviations from Hardy–Weinberg equilibrium (HWE) for each locus in each population were tested with Genepop v4.2.1 (Rousset 2008) using the Markov chain method. Microchecker v2.2.3 (van Oosterhout et al. 2004) was used to test for the presence of null alleles and large allele dropout. Microsatellite neutrality was tested using the Fst-outlier method implemented in LOSITAN (Beaumont and Nichols 1996, Antao et al. 2008). LOSITAN was run for 100 simulations using the neutral mean Fst and force mean Fst settings for a 0.99 confidence interval and infinite alleles model.

Results

Primer sequences, microsatellite repeat motifs, PCR amplicon sizes, and the number of alleles are provided in Table 2. Only one marker,

Table 2. Characterization of the microsatellite markers

Locus	Primer sequence (5'–3')	Label	Motif	S_A	MM	N_A
A553	F TGGAATTAGGCTATCTCACACTC R ACAAAACTCTGGCGGTTGG	6-FAM	(CT)10	76-110	1	23
M1010	F TGGCTCCTCAAGAACGACC R CGAAGTGCGCATCCAATTTC	VIC	(AC)7	167-233	1	47
M1422	F GTGTCCACACGAAGCTATTTG R AGTTTCATGCCAACGTCCC	PET	(GAT)8	192-267	1	27
A1267	F AAACCGCACGCACATTGAC R ACCAGACGGACTGACACAC	6-FAM	(AC)7	83-125	2	14
M2333	F TGTCTGCTTCTACTGGGATG R AAGCTATAGGGATCGTCGC	VIC	(AC)16	166-234	2	35
K1466	F CACTTGCACGGCCTCTTTC R AATGTGGCTTAACAACCGTC	NED	(AC)9	196-260	2	22

F, forward primer sequence; R, reverse primer sequence; label, fluorescent label on 5'-end of forward primer; *SA*, amplicon size range; MM, multiplex mix; *NA*, total number of alleles.

A553, showed stutter patterns that made scoring of some individuals difficult. The largest peak was used, and individuals that were ambiguous because of stutter were marked as missing data. The other five markers had minimal stuttering that did not influence peak calling. The number of alleles per locus varied from 14 to 47 in a total of 213 individuals from seven species. Table 3 summarizes the genetic diversity estimated for each locus in each population for seven species. Observed heterozygosity ranged from 0 to 1 and the expected heterozygosity ranged from 0.129 to 0.945. There was a total of 62 private alleles with all but one population, OL.NLW having at least one private allele. The HWE test, which indicates deviations between observed and expected heterozygosity showed significant deviations in three loci for one to two populations each after a Bonferroni correction ($\alpha = 0.00055$). Three loci were in HWE in all populations. Null alleles were detected in two of the six loci, A1267 and M2333, which had overlapping amplicon ranges and were multiplexed in the same reaction. Null alleles in these two markers are likely due to some interference between them in the electropherograms. Using the Fst-oulier method implemented in LOSITAN, one locus was identified as being a candidate of balancing selection, M2333. This may be due because of the presence of null alleles, which would likely reduce heterozygosity and make this locus. The A1267 locus was unusual because over 50% of the specimens tested were homozygous for one very common allele, whereas the other loci were much more heterogeneous. Some populations failed to amplify one of the multiplex reactions.

Table 3. Characterization of microsatellite markers for each population

Loci	Populations											
	NN.KOB				WD.OKT				OL.RSC			
	N_A	N_P	H_O	H_E	N_A	N_P	H_O	H_E	N_A	N_P	H_O	H_E
A553	6	1	0.461	0.617	7	1	0.786	0.733	4	0	0.688	0.528
M1010	9	1	0.500	0.753	6	0	0.697	0.778	7	0	0.765	0.799
M1422	7	3	0.567	0.801	4	0	0.606	0.672	2	0	0.133	0.129
A1267	5	3	0.233	0.219	5	1	0.060	0.174	4	1	0.333	0.602
M2333	8	0	0.055	0.849	10	1	0.625	0.835	7	3	0.444	0.740
K1466	8	1	0.533	0.837	10	3	0.818	0.794	5	1	0.000	0.589
	OL.NLW				MB.GRO				MB.WCR			
	N_A	N_P	H_O	H_E	N_A	N_P	H_O	H_E	N_A	N_P	H_O	H_E
A553	5	0	1.000	0.822	5	0	0.600	0.695	4	1	1.000	0.750
M1010	5	0	0.833	0.803	5	0	0.000	0.800	4	0	0.000	0.857
M1422	2	0	0.167	0.167	5	0	0.500	0.653	2	0	0.000	0.429
A1267	X	0	X	X	3	0	0.400	0.484	3	0	0.500	0.464
M2333	X	0	X	X	9	3	0.800	0.863	5	0	0.750	0.857
K1466	X	0	X	X	3	0	0.400	0.353	3	1	0.500	0.464
	MB.LVR				MB.LVB				MB.LVM			
	N_A	N_P	H_O	H_E	N_A	N_P	H_O	H_E	N_A	N_P	H_O	H_E
A553	5	2	0.714	0.769	4	0	0.714	0.780	3	0	0.400	0.378
M1010	5	0	0.429	0.780	8	1	0.286	0.923	6	2	0.600	0.889
M1422	4	0	0.333	0.636	7	3	0.429	0.879	6	2	0.800	0.844
A1267	2	0	0.500	0.429	2	0	0.400	0.356	3	1	0.500	0.833
M2333	4	0	0.667	0.636	8	1	0.667	0.924	5	0	1.000	0.867
K1466	X	X	X	X	3	0	0.500	0.439	X	X	X	X
	NN.HRN				NN.WGB				NN.BES			
	N_A	N_P	H_O	H_E	N_A	N_P	H_O	H_E	N_A	N_P	H_O	H_E
A553	10	1	0.842	0.838	5	0	0.688	0.784	7	0	0.938	0.766
M1010	16	3	0.789	0.871	8	2	0.375	0.603	15	6	0.875	0.921
M1422	7	0	0.500	0.800	7	1	0.533	0.795	5	0	0.286	0.725
A1267	4	0	0.667	0.608	3	0	0.267	0.343	2	0	0.357	0.452
M2333	9	0	0.556	0.882	16	3	0.733	0.945	7	0	0.667	0.826
K1466	6	2	0.222	0.765	3	0	0.143	0.670	3	0	0.357	0.320
	SD.TAP				MB.BAV				SC.IDN			
	N_A	N_P	H_O	H_E	N_A	N_P	H_O	H_E	N_A	N_P	H_O	H_E
A553	6	0	0.813	0.782	4	0	0.923	0.692	4	1	0.750	0.567
M1010	13	2	0.765	0.868	X	X	X	X	2	1	0.188	0.272
M1422	6	0	0.667	0.844	5	1	0.769	0.671	9	0	0.938	0.857
A1267	4	1	0.529	0.594	X	X	X	X	X	X	X	X
M2333	9	0	0.647	0.818	X	X	X	X	X	X	X	X
K1466	6	1	0.563	0.782	X	X	X	X	X	X	X	X

N_A, number of alleles; N_P, number of private alleles; H_O, observed heterozygosity; H_E, expected heterozygosity; X, missing data. H_E cells colored gray represent loci that violate HWE.

Discussion

Next-generation sequencing using 454 pyrosequencing was employed in the development of novel microsatellite markers for the New Zealand cicada genus *Kikihia*. This technology is becoming more commonly used in microsatellite discovery from a range of taxa (Abdelkrim et al. 2009, Ekblom and Galindo 2010, Gardner et al. 2011). This is the first published discovery of microsatellite markers for any New Zealand cicadas. Unlike most studies that are focused on intraspecific diversity, we are focused on intra- and interspecific diversity in a genus that is approximately 10–12 million years old with the earliest extant species split more than 6 million years ago (Marshall et al. 2008). Microsatellite markers that were variable and cross-amplified in all seven species of *Kikihia* are presented here.

Three of the six loci were found to violate the HWE in no more than 2 of the 15 populations tested. Heterozygote deficiency increases due to factors such as inbreeding, population stratification, null alleles, and genotyping errors. It is not surprising that species of *Kikihia* display high levels of inbreeding or population stratification since field observations suggest low dispersal rates. However, these were not explicitly tested here.

The large number of potential hybrid zones between multiple species of *Kikihia* makes this genus uniquely suited for the study of hybrid zones. Many of these potential hybrid zones are between nonsister species that vary in their divergence times from less than 1 million years to more than 3 million years (Marshall et al. 2008, 2011). The markers developed in this study will be useful for investigations of the evolutionary past and future of this interesting species radiation.

References Cited

Abdelkrim, J., B. C. Robertson, J. L. Stanton, and N. J. Gemmell. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. BioTechniques 46:185.

Antao, T., A. Lopes, R. J. Lopes, A. Beja-Pereira, and G. Luikart. 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics 9:323.

Arensburger, P., T. R. Buckley, C. Simon, M. Moulds, and K. E. Holsinger. 2004a. Biogeography and phylogeny of the New Zealand cicada genera (Hemiptera: Cicadidae) based on nuclear and mitochondrial DNA data. J. Biogeogr. 31:557–569.

Arensburger, P., C. Simon, and K. Holsinger. 2004b. Evolution and phylogeny of the New Zealand cicada genus *Kikihia* Dugdale (Homoptera: Auchenorrhyncha: Cicadidae) with special reference to the origin of the Kermadec and Norfolk Islands' species. J. Biogeogr. 31:1769–1783.

Barton, N., and G. M. Hewitt. 1989. Adaptation, speciation and hybrid zones. Nature 341:497–503.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. London B Biol. Sci. 263: 1619–1626.

Buckley, T. R., P. Arensburger, C. Simon, and G. K. Chambers. 2002. Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. Syst. Biol. 51:4–18.

Dugdale, J. S. 1971. Genera of New Zealand Cicadidae (Homoptera). N Z J. Sci. 14:856–882.

Ekblom, R., and J. Galindo. 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107:1-15.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. Online 1:47.

Fabricius, J. C. 1775. *Systema entomologiae, sistens insectorum classes, ordines, genera, species, adiectis synonymis, locis, descriptionibus, observationibus*. Officina Libraria Kortii, Flensburgi et Lipsiae xxxii:832 pp.

Faircloth, B. C. 2008. msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol. Ecol. Resour. 8:92–94.

Fleming, C. A. 1984. The cicada genus *Kikihia* Dugdale (Hemiptera, Homoptera): the New Zealand Green Foliage Cicadas. National Museum of New Zealand, 18:191–206.

Gardner, M. G., A. J. Fitch, T. Bertozzi, and A. J. Lowe. 2011. Rise of the machines—recommendations for ecologists when using next generation sequencing for microsatellite development. Mol. Ecol. Resour. 11:1093–1101.

Glaubitz, J. C. 2004. Convert: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Mol. Ecol. Notes 4:309–310.

Harrison, R. G. 1993. Hybrid zones and the evolutionary process. Oxford University Press, Oxford, UK.

Hewitt, G. M. 2011. Quaternary phylogeography: the roots of hybrid zones. Genetica 139:617–638.

Larivière, M., M. J. Fletcher, and A. Larochelle. 2010. Auchenorrhyncha (Insecta: Hemiptera): catalogue. Fauna N. Z. 63.

Mallet, J. 2007. Hybrid speciation. Nature 446:279–283.

Marshall, D. C., K. Slon, J. R. Cooley, K. B. Hill, and C. Simon. 2008. Steady Plio-Pleistocene diversification and a 2-million-year sympatry threshold in a New Zealand cicada radiation. Mol. Phylogenet. Evol. 48: 1054–1066.

Marshall, D. C., K. B. Hill, K. M. Fontaine, T. R. Buckley, and C. Simon. 2009. Glacial refugia in a maritime temperate climate: cicada (*Kikihia subalpina*) mtDNA phylogeography in New Zealand. Mol. Ecol. 18:1995–2009.

Marshall, D. C., K. B. Hill, J. R. Cooley, and C. Simon. 2011. Hybridization, mitochondrial DNA phylogeography, and prediction of the early stages of reproductive isolation: lessons from New Zealand cicadas (Genus Kikihia). Syst. Biol. 60:482–502.

Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8:103–106.

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

Untergrasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. 2012. Primer3 – new capabilities and interfaces. Nucleic Acids Research 40:e115.

van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4:535–538.

Walker, F. 1850. List of the specimens of Homopterous Insects in the Collection of the British Museum. Part 1, British Museum (Natural History), London. pp.1–260.