Entomological Science

Entomological Science (2015) **18,** 283–287 doi:10.1111/ens.12121

SHORT COMMUNICATION

Development of microsatellite loci from a reference genome for the Neotropical butterfly *Heliconius numata* and its close relatives

Mathieu CHOUTEAU[1], Annabel WHIBLEY[1], Bernard ANGERS[2] and Mathieu JORON[1]

[1]*Institut de Systématique, Evolution et Biodiversité, Muséum National d'Histoire Naturelle, Paris, France; and* [2]*Department of Biological Sciences, Université de Montréal, Montreal, Canada*

Abstract

The Neotropical butterfly *Heliconius numata* (Lepidoptera: Nymphalidae: Heliconiinae) is known for its striking diversity of wing color patterns driven by the Müllerian mimicry of multiple local models and controlled by a single supergene locus. Such fine-scale variation of traits under strong selection offers a unique opportunity for the study of the ecology and genetics of adaptation. However, little is still known of the population processes driving geographical variation in wing-pattern phenotypes. We report the characterization of 26 microsatellite markers for the butterfly *H. numata*, including six located inside the wing color-pattern supergene region. All markers are polymorphic, with allele numbers ranging from 2 to 21 per locus, an observed heterozygosity of 0.111 to 0.848 and an expected heterozygosity of 0.126 to 0.942. A subset of 18 of these markers was tested on five closely related sympatric *Heliconius* species with an amplification success ranging from 88% to 94%. The obtained set of microsatellite markers provides a new and useful set of tools to investigate patterns of differentiation and selection in populations of mimetic *Heliconius* butterflies. Moreover, markers developed within the color-pattern supergene will facilitate characterization of the association between the genetic architecture and the functional diversity of wing patterns. Finally, the cross-species amplification success of the described markers extends their utility to also encompass comparative population genetic studies of closely related species within a clade of rapidly diversifying species.

Key words: *Heliconius elevatus*, *Heliconius ethilla*, *Heliconius hecale*, *Heliconius pardalinus*, Lepidoptera, Nymphalidae, supergene.

The Neotropical butterflies in the genus *Heliconius* (Nymphalidae: Heliconiinae) are widely known for the remarkable diversity of wing color patterns displayed within species across their range and for the local convergence of phenotypes between distantly related species involved in mimetic assemblages (Brown & Benson 1974; Jiggins & McMillan 1997). Most *Heliconius* species adopt a single phenotype in a given geographic region that mimics one of the local warning color patterns adopted by other butterflies in the community, and polymorphism is typically restricted to hybrid zones. However, *Heliconius numata* (Cramer) displays widespread polymorphism in warning color patterns within populations, with two to seven distinct morphs co-existing in a single locality (Brown & Benson 1974). Variation is also found among populations, with adjacent regions harboring distinct combinations of morphs, following geographic variation in the diversity of warning signals in butterfly communities. Each *H. numata* morph is a perfect mimic of a distinct set of species in the ithomiine genera: *Melinaea*, *Athyrtis* and *Mechanitis* (Nymphalidae: Danainae), with which it is sympatric (Joron *et al.* 1999). Previous molecular work performed using allozyme markers did not resolve the

Correspondence: Mathieu Chouteau, Institut de Systématique, Evolution et Biodiversité, UMR 7205, Muséum National d'Histoire Naturelle, CP50, 45 rue Buffon, 75005 Paris, France.
Email: mathieu.chouteau@umontreal.ca

Received 28 March 2014; accepted 9 June 2014.

fine-scale population structure believed to characterize populations locally adapted to distinct mimetic assemblages, and little is known about demographics and adaptive evolution in those populations (Joron *et al.* 1999). Mimetic diversity in *H. numata* is controlled by a single genomic region, the supergene locus, *P*, mapped to an interval of approximately 400 kb, and which contain 18 genes (Joron *et al.* 2011). The functional variants participating in the supergene have yet to be identified. Here we present primers that amplify 26 microsatellite loci, including six located in the supergene *P*, for *H. numata*, in addition to the successful cross-amplification of 18 of these loci in populations of five closely related species found in sympatry in northern Peru.

Candidate microsatellite loci for *H. numata* were first identified from the publicly available reference genome of *Heliconius melpomene* (Linnaeus) (The *Heliconius* Genome Consortium 2012), and from sequences of bacterial artificial chromosome (BAC) clones of *H. numata* (Joron *et al.* 2011). The isolation of useful microsatellite markers in the Lepidoptera is complicated due to the presence of microsatellite DNA families comprising mobile elements containing microsatellite repeats with nearly identical flanking regions (Meglecz *et al.* 2004; Zhang 2004). To maximize the chances of isolating single-copy microsatellites, the *H. melpomene* genome was scanned for repetitive elements that were masked from the genome prior to candidate selection. Sequences with low complexity or homology to known lepidopteran repetitive elements were excluded using RepeatMasker v3.3.0 (http://repeatmasker.org) in conjunction with a custom library (The *Heliconius* Genome Consortium 2012; available from http://butterflygenome.org). Microsatellites were then detected using MSATCOMMANDER v1.0.8 (Faircloth 2008). From the original unmasked genome, which contained a total of 6974 perfect microsatellites with more than ten repetitions, 1087 were identified as not being part of the repetitive elements. Among these, 85 were selected for further investigation, with representatives on each of the 21 *Heliconius* chromosomes, including 12 within the wing-pattern supergene *P*. Sequences flanking the selected microsatellites were examined in alignments of whole genome re-sequence data from four *H. numata* individuals from natural populations of Peru. This step enabled us to highlight conserved regions for primer design, thus maximizing amplification rate and lowering the odds of designing primers with null alleles, which is a common problem with Lepidoptera (Zhang 2004). Briefly, for each wild-caught individual, standard short insert Illumina paired end libraries were prepared from gDNA extracted from specimens preserved in salt-saturated DMSO solution using the DNeasy kit (Qiagen, Hilden, Germany). Approximately 100 million 100 bp paired-end reads were obtained for each individual. Reads were aligned to the *H. melpomene* reference genome (v1.1) as described in Martin *et al.* (2013) using Stampy v1.0.17 (Lunter & Goodson 2011) and were visualized using the Integrative genomics viewer v2.3 (Robinson *et al.* 2011). Of the 85 microsatellites initially selected, 22 possessed conserved flanking regions suitable for primer design (Table 1). Primers were designed with the aid of Primer3 v0.4.0 (Untergasser *et al.* 2012).

DNA was extracted from thoracic tissue using a Qiagen DNeasy tissue kit. Polymerase chain reaction (PCR) conditions (10 μL volume) were: 1 μL of 10× *Taq* Buffer, 0.2 mM dNTPs, 0.2 U of *Taq* Polymerase (DreamTaq; Thermo Scientific, Waltham, MA, USA), 5–40 ng of genomic DNA, 0.1 μM M13-tailed forward primer, 0.4 μM dye-labeled (NED, VIC, PET and 6-FAM) M13 primer (5′CACGACGTTGTAAAACGAC3′) and 0.4 μM reverse primer. The PCR amplification temperature profile consisted of an initial denaturation at 92°C for 2 min, followed by 40 cycles at 92°C for 30 s, 53°C for 30 s and 72°C for 1 min and a final elongation at 72°C for 15 min. Reaction products were multiplexed (up to eight loci) and 1 μL of the resulting multiplexed solution was added to 8.8 μL Hi-Di formamide (Life Technologies, Carlsbad, CA, USA) and 0.2 μL size standard (Genescan 500 LIZ; Life Technologies). Products were resolved on an ABI 3130 Genetic Analyzer capillary DNA sequencer (Life Technologies) and scored and binned using SoftGenetics GeneMarker v.4.0 (LLC, State College, PA, USA). Amplification success and polymorphism of the reported microsatellites and of four microsatellites developed for *H. melpomene* (Flanagan *et al.* 2002) were tested on 50 *H. numata* individuals from the Tarapoto area of northern Peru. Moreover, a selection of 18 microsatellites were tested on 18–28 individuals of five closely related species/subspecies of Peruvian silvaniform *Heliconius* taxa found in sympatry with *H. numata* (Table 2): *H. pardalinus sergestus* Bates, *H. pardalinus butleri* Bates, *H. ethilla aerotome* Godart, *H. elevatus* Nöldner and *H. hecale felix* Fabricius.

Observed and expected heterozygosities and estimates of linkage disequilibrium between loci were calculated using GenePop v4.2 (Raymond & Rousset 1995). In *H. numata*, all loci were polymorphic with allele numbers ranging from 2 to 21. For loci outside of the wing color-pattern supergene *P*, observed heterozygosity

Table 1 Set of 22 microsatellite markers developed for *Heliconius numata*

Locus name	Chromosome	Repeat motif	Primer sequences (5′–3′)	Fragment length (bp)	GenBank accession No.
Num1-1	1	AT	F: TTTTTAAGGGGGAAAACAGA R: TTTGAAGCTSCGAGTAAAAAG	235–273	KJ786527
Num1-2	1	AT	F: GTGTTAACAAAAATAGAGAGCTTG R: GTTGAACTGAATGCTGCTGCTG	220–222	KJ786528
Num2-2	2	CGGAGT	F: GTGAAACAACATCGCGTCA R: GATTGTAAGTGCCTTCCGAGT	190–358	KJ786529
Num3-1	3	AT	F: CACKGTTATATCACTTTCAAAAAC R: GGAGTATAACCTCGCTMGAA	161–225	KJ786530
Num3-2	3	AC	F: GGTTCCCATTTCATTCTCCT R: TGCGATGGCGTTATTTATATC	178–210	KJ786531
Num3-3	3	AT	F: CTYCCAACAACTAACATATTTCAT R: AGGGGAGRATATCGATTTGT	180–204	KJ786532
Num4-1	4	AT	F: CAAATRCCAACGCCTACT R: TAAATCCRTCAACGTGCGC	142–190	KJ786533
Num5-1	5	AT	F: GTTCAGGTTGACTCGGATGC R: ACAGGAGAACGTTACGGTACTTAG	130–224	KJ786534
Num6-1	6	CT	F: CAAATTGTGGGGACGAAGTC R: ACCCTCGCACGAGAGCYA	191–220	KJ786535
Num11-1	11	AT	F: CGGCTCAAATCTTTGTAGGT R: GCGTGTGGACGAGATTAATAC	291–317	KJ786536
Num11-2	11	AT	F: CAATGGCTGTGAATTGTTAACTG R: TTTTTGTAAGCTTAGTTTTGATAAT	207–212	KJ786537
Num14-1	14	AT	F: CAAGCAATGATTTAACAACTAATCA R: TGCTGTTTCACCTTYTTGAATC	154–192	KJ786538
Num15-1	15	CAT	F: CAAGTCCGTAGCAACAGTCA R: GATGGCCTAGGATCATTGTG	163–175	KJ786539
Num17-1	17	AC	F: ATTTCGGCATACTGCGTGTA R: CCACGGTTAACGAGAATACG	168–196	KJ786540
Num17-2	17	AT	F: GCYACTATAGGGACACTTTTT R: GGGCAGWCTATTCGTGTCC	107–175	KJ786541
Num18-2	18	AT	F: TTGTTACATAACGCAATTCAAA R: TTGTCACTTGTTTATTACCACAGA	203–219	KJ786542
NumP1	15	AT	F: ACCCCGCCCAGAGACTAATAG R: TATTTCCTAAAATTCATAATGTAGAG	212–214	KJ786543
NumP2	15	ACAT	F: TGATGTTCAACTTTTATATTCTTTTG R: TGAGATGGCATGACTCACAT	159–186	KJ786544
NumP3	15	ATC	F: CAGCACTTTATCTAGAAATTATATAGA R: GCAHTGCAAAGGATGGTAATG	183–282	KJ786545
NumP10	15	TGG	F: TCGTAGGTATTCGGAGAACG R: CTCTGCGTTCCCATTAAGAA	317–325	KJ786546
NumP11	15	CA	F: ACCACATGGGGGTCTAAAGT R: CGAACTTCCGTTGCACTCT	125–145	KJ786547
NumP12	15	TTA	F: TCAAATTTTAGGTTGACAATGTT R: TGCCTACTCTGCTATGAGTGTG	198–229	KJ786548

(H_o) ranged from 0.111 to 0.848 and the expected heterozygosity (H_e) ranged from 0.126 to 0.942. For the six loci located inside the supergene *P* region (i.e. NumP1–NumP12), observed heterozygosity ranged from 0.176 to 0.666 and expected heterozygosity from 0.165 to 0.826. Seven loci outside of *P* (Num1-1, Num3-1, Num11-1, Num18-2, Hm02, Hm19, Hm22) and one loci within *P* (NumP12) displayed significant deviation from Hardy–Weinberg equilibrium ($P < 0.05$) caused by heterozygote deficiency and may indicate the presence of null alleles. Linkage disequilibrium was significant ($P < 0.01$) for only one pair of loci, NumP3–NumP10, lying within the supergene region. Among the microsatellite loci located within the supergene interval, NumP3 showed an association with different *H. numata* phenotypes. All *H. n. bicoloratus* individuals ($n = 28$) carried at least one 204-bp allele, all *H. n. tarapotensis* ($n = 20$) carried at least one 183-bp allele and never any 204-bp alleles, and both *H. n. silvana* tested ($n = 2$) were homozygous for the 197-bp allele. This finding is consistent with hierarchical dominance of the supergene alleles ($P^{sil} < P^{tar} < P^{bic}$, Joron *et al*. 2011). Furthermore,

Table 2 Characterization of microsatellite loci in *Heliconius numata*, *H. pardalinus sergestus*, *H. pardalinus butleri*, *H. ethilla aerotome*, *H. hecale felix* and *H. elevatus*

Locus name	*H. numata* ($n = 50$)			*H. pardalinus sergestus* ($n = 20$)			*H. pardalinus butleri* ($n = 18$)			*H. ethilla aerotome* ($n = 28$)			*H. hecale felix* ($n = 18$)			*H. elevates* ($n = 18$)		
	N_a	H_o	H_e	N_a	H_o	H_e	N_a	H_o	H_e	N_a	H_o	H_e	N_a	H_o	H_e	N_a	H_o	H_e
Num1-1	11	0.233*	0.376															
Num1-2	2	0.133	0.126															
Num2-2	21	0.740	0.829	18	0.947	0.946	16	0.666**	0.926	2	0.560	0.503	11	0.933	0.908	11	0.846	0.907
Num3-1	14	0.366**	0.568															
Num3-2	16	0.848	0.917	6	0.348	0.349	14	0.722	0.852	6	0.740	0.735	4	0.666	0.581	13	1.000	0.913
Num3-3	6	0.222	0.303	4	0.136**	0.544	4	0.384*	0.563	1	–	–	1	–	–	5	0.071**	0.627
Num4-1	10	0.743	0.753	5	0.636*	0.540	6	0.000**	0.751	1	–	–	2	0.133	0.128	6	0.55	0.744
Num5-1	10	0.400	0.455	5	0.087**	0.206	5	0.083*	0.735	1	–	–	4	0.422*	0.660	8	0.090**	0.869
Num6-1	15	0.592*	0.817	5	0.136	0.175	9	0.705	0.877		NA		7	0.334	0.445	7	0.714	0.814
Num11-1	17	0.312**	0.42															
Num11-2	6	0.629	0.707	4	0.363	0.546	4	0.332**	0.633	5	0.107**	0.234	5	0.400	0.358	6	0.615**	0.886
Num14-1	7	0.259	0.336															
Num15-1	5	0.481	0.544	1	–	–	4	0.277	0.339	2	0.107	0.165	3	0.400	0.397	3	0.333	0.298
Num17–1	19	0.87	0.94	6	0.412*	0.620	18	0.775	0.938	3	0.074	0.073	10	0.666*	0.862	16	0.833	0.945
Num17–2	6	0.687	0.556															
Num18–2	7	0.523*	0.685		NA		5	0.102**	0.763	5	0.222	0.210	1	–	–		NA	
Hm02†	13	0.296**	0.706	4	0.652*	0.561		NA		3	0.000**	0.356	5	0.533	0.712	11	0.727**	0.904
Hm06†	7	0.555	0.611	4	0.174	0.205	3	0.470	0.570	2	1.000	0.935	5	0.615	0.700	6	0.642	0.661
Hm19†	9	0.333**	0.627	3	0.066*	0.190		NA			NA			NA		2	0.000**	0.507
Hm22†	6	0.111**	0.605	5	0.588	0.639	9	0.437**	0.808	1	–	–	5	0.600	0.762	8	0.666	0.812
NumP1	2	0.18	0.165															
NumP2	11	0.666	0.728	5	0.705	0.684	10	0.626	0.802	5	0.357	0.490	2	0.077	0.077	7	0.502**	0.851
NumP3	6	0.578	0.676	3	0.266	0.245	3	0.5	0.469	3	0.166	0.163	4	0.289	0.391	3	0.598	0.659
NumP10	5	0.334	0.356	3	0.087	0.166	2	0.000**	0.225	2	0.043	0.043	5	0.348	0.409	2	0.000**	0.270
NumP11	5	0.480	0.602	2	0.000**	0.162	4	0.266*	0.562	2	0.043	0.043	2	0.071	0.071	7	0.000**	0.667
NumP12	9	0.238**	0.826	5	0.524	0.629	7	0.588	0.734	4	0.269*	0.377	3	0.400	0.384	4	0.143**	0.751

Loci for which observed heterozygosity deviates from expectations under Hardy–Weinberg equilibrium are marked with $*P < 0.05$ and $**P < 0.01$.

†Loci belong to Flanagan *et al.* (2002) microsatellite library. NA, absent or inconsistent amplification; blank cells, loci not tested for cross amplification; –, metrics not calculated due to allele fixation; N_a, number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity.

the 216-bp allele of the NumP12 locus was carried by all *H. n. bicoloratus* specimens, and by none of the other forms. Other loci developed within the supergene *P* do not show any clear association with phenotype, but more samples are needed to verify this. The 18 microsatellite loci tested in populations of closely related species showed amplification success varying from 94% for *H. pardalinus sergestus*, *H. hecale felix* and *H. elevatus* to 88% for *H. pardalinus butleri* and *H. ethilla aerotome*. Loci deviating from Hardy–Weinberg equilibrium ($P < 0.05$) ranged from two (*H. ethilla aerotome*) to nine (*H. pardalinus butleri*; Table 2).

The obtained set of microsatellites for *H. numata* consists of 20 loci located outside the wing color-pattern supergene *P* and of six loci located within the supergene *P*. The 20 loci located outside the supergene *P* will help assess the fine-scale population structure and demographic parameters, whereas the loci inside the wing color pattern determining interval may facilitate inferences regarding the geographic pattern of selection on wing color. Microsatellites located inside the supergene *P* also provide a useful tool for the genetic mapping of patterning elements and will facilitate the identification of individual genotypes (parental alleles) in controlled crosses, an indispensable tool to better understand the genetic architecture and dominance of this genomic region. Finally, the success of cross-amplification in other sympatric species of the *Heliconius* silvaniform clade is encouraging for comparative population genetic studies of this species complex. Interspecific hybridization and introgression of wing color-pattern variants have recently been shown to be of great importance in adaptive evolution and mimicry (Pardo-Diaz *et al*. 2012; The *Heliconius* Genome Consortium 2012). Important avenues for future research such as assessing rates of hybridization and introgression in the wild will be greatly facilitated by the use of these markers.

ACKNOWLEDGMENTS

We thank M McClure and F Prunier for help in the field and in the laboratory. This research was supported by NSERC to MC and by ERC starting grant MIMEVOL and ANR HybEvol to MJ.

REFERENCES

Brown KS Jr, Benson WW (1974) Adaptive polymorphism associated with multiple Müllerian mimicry in *Heliconius numata* (Lepid. Nymph). *Biotropica* **6,** 205–228.

Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* **8,** 92–94.

Flanagan NS, Blum MJ, Davison A *et al*. (2002) Characterization of microsatellite loci in neotropical *Heliconius* butterflies. *Molecular Ecology Notes* **2,** 398–401.

Jiggins CD, McMillan WO (1997) The genetic basis of an adaptive radiation: warning colour in two *Heliconius* species. *Proceedings of the Royal Society of London Series B, Biological Sciences* **264,** 1167–1175.

Joron M, Wynne IR, Lamas G, Mallet J (1999) Variable selection and the coexistence of multiple mimetic forms of the butterfly *Heliconius numata*. *Evolutionary Ecology* **13,** 721–754.

Joron M, Frezal L, Jones RT *et al*. (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477,** 203–206.

Lunter G, Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research* **21,** 936–939.

Martin SH, Dasmahapatra KK, Nadeau NJ *et al*. (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23,** 1817–1828.

Meglecz E, Petenian F, Danchin E, D'Acier AC, Rasplus J-Y FE (2004) High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Molecular Ecology* **13,** 1693–1700.

Pardo-Diaz C, Salazar C, Baxter SW *et al*. (2012) Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics* **8,** e1002752.

Raymond M, Rousset F (1995) GENEPOP (VERSION 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86,** 248–249.

Robinson JT, Thorvaldsdóttir H, Winckler W *et al*. (2011) Integrative genomics viewer. *Nature Biotechnology* **29,** 24–26.

The *Heliconius* Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487,** 94–98.

Untergasser A, Cutcutache I, Koressaar T *et al*. (2012) Primer3–new capabilities and interfaces. *Nucleic Acids Research* **40,** e115.

Zhang D-X (2004) Lepidopteran microsatellite DNA: redundant but promising. *Trends in Ecology & Evolution* **19,** 507–509.