Acknowledgements

This study was supported by key program of Science and Technology Department Foundation of Fujian Province (Grant No 2007N0050). We also thank Dr Mark Roberts for constructive comments on this manuscript.

References

Adams S, Mapstone BD, Russ GR, Davies CR (2000) Geographic variation in the sex ratio, sex-specific size and age structure of *Plectropomus leopardus* (Serranidae) between reefs open and closed to fishing on the Great Barrier Reef. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 1448–1458.

Armsworth PR (2001) Effects of fishing on a protogynous hermaphrodite. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 568–578.

Armsworth PR (2002) Recruitment limitation, population regulation and larval connectivity in reef fish metapopulations. *Ecology*, **83**, 1092–1104.

Frisch A, Anderson T (2005) Physiological stress responses of two species of coral trout (*Plectropomus leopardus* and *Plectropomus maculatus*). *Comparative Biochemistry and Physiology A—Molecular & Integrative Physiology*, **140**, 317–327.

Heemstra PC, Randall JE (1993) *FAO Fisheries Synopsis*. No. 125, vol. 16, pp. 292–293. FAO, Rome.

van Herwerden L, Choat JH, Dudgeon CL *et al*. (2006) Contrasting patterns of genetic structure in two species of the coral trout *Plectropomus* (Serranidae) from east and west Australia: Introgressive hybridisation or ancestral polymorphisms. *Molecular Phylogenetics and Evolution*, **41**, 420–435.

Munday PL, Jones GP, Pratchett MS, Williams AJ (2008) Climate change and the future for coral reef fishes. *Fish and Fisheries*, **9**, 261–285.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–3323.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Yeh FC, Yang RC, Boyle T (1999) *PopGene (version 1.32). Microsoft Window-base Software for Population Genetic Analysis: A Quick User's Guide*. University of Alberta, Center for International Forestry Research, Alberta, Canada.

doi: 10.1111/j.1755-0998.2009.02690.x

© 2009 Blackwell Publishing Ltd

Microsatellite markers associated with genes expressed in developing wings of *Bicyclus anynana* butterflies

PATRÍCIA BELDADE,*† ARJÈN E. VAN'T HOF,‡ MARIA ADELINA JERÓNIMO† and ANTHONY D. LONG§

**Institute of Biology Leiden University, PO Box 9516, 2300 RA, Leiden, The Netherlands, †Instituto Gulbenkian de Ciência, Apartado 14, P-2781-901 Oeiras, Portugal, ‡School of Biological Sciences, Biosciences Building, Crown Street, University of Liverpool, Liverpool L69 7ZB, UK, §Department of Ecology and Evolutionary Biology, University of California at Irvine, 5210 McGaugh Hall, Irvine, 92697-2525 CA, USA*

Abstract

Deriving useful microsatellite markers in lepidopterans has been challenging when relying on scans of genomic DNA libraries, presumably due to repetitiveness in their genomes. We assayed 96 of 320 microsatellites identified *in silico* from a collection of *Bicyclus anynana* ESTs, in 11 independent individuals from a laboratory population. From the 68 successful assays, we identified 40 polymorphic markers including 22 with BLAST-based annotation. Nine of 12 selected polymorphic markers tested in a panel of 24 wild-caught individuals converted to successful assays and were all polymorphic. We discuss how microsatellite discovery in ESTs is an efficient strategy with important attendant advantages.

Keywords: *Bicyclus anynana*, EST, genetic markers, Lepidoptera, microsatellite, polymorphism

Received 20 June 2008; revision accepted 26 February 2009

Correspondence: Patrícia Beldade, Fax: +31 71 5274900;
E-mail: p.beldade@biology.leidenuniv.nl

Microsatellites are widely used as neutral genetic markers for a variety of purposes; including analysis of relatedness, population structure, demographic and selective

processes, and genetic mapping (Ellegren 2004; Schlötterer 2004). Moreover, several studies have suggested that microsatellites can also play important roles in generating variation (Rockman & Wray 2002; Fondon & Garner 2004; Kashi & King 2006). Microsatellites have been explored in a variety of organisms, but have proven difficult to obtain for lepidopterans (butterflies and moths) using traditional methods that rely on genomic DNA libraries. This is potentially because lepidopteran microsatellites often appear within multi-copy genomic regions (Zhang 2004, Van't Hof *et al.* 2007), which results in PCR failure.

Lepidopterans have long been favourite targets of biological research (also because of their economic importance), and have some unusual genetic properties (such as heterogametic females and holocentric chromosomes). More recently, the derived and diverse colour patterns on butterfly wings have been the focus of efforts to develop genomic resources to link genotypes, to development, to phenotypes, to fitness (Beldade *et al.* 2008). Linkage maps are now available for a few species (Jiggins *et al.* 2005; Kapan *et al.* 2006; Van't Hof *et al.* 2008, Wang & Porter 2004), but are made mostly of anonymous markers whose value for across-species comparisons (e.g. Yasukochi *et al.* 2006; Pringle *et al.* 2007; Beldade *et al.* 2009) is limited. Having markers associated with expressed genes has advantages for downstream use of the maps and should improve marker conversion success. Growing EST collections for butterflies (Papanicolaou *et al.* 2008) represents a potential valuable source for microsatellites in this group. Here, we analyse microsatellites found in an EST collection from *Bicyclus anynana* butterflies, an emerging model for the evolutionary-developmental analysis of adaptive phenotypes (Beldade & Brakefield 2002).

Our *B. anynana* EST collection was derived from developing wings of a large number of outbred individuals to combine gene discovery with DNA sequence polymorphism identification (Beldade *et al.* 2006). A custom PERL script designed to search for microsatellite repeats detected 320 di-, tri-, tetra- and penta-nucleotide repeat microsatellites (perfect repeats with repeat number greater than 5 for dinucleotide microsatellites and greater than 3 for the remainder) in 4251 assembled unigenes (Beldade *et al.* 2006). Ten of the 73 microsatellites in non-singleton unigenes were polymorphic in our EST collection with up to seven alleles.

We selected 96 of the 320 EST-based microsatellites to test in a panel of 11 individuals (five females and six males) from a laboratory outbred stock (Brakefield *et al.* 2009). To maximize the likelihood of finding polymorphisms, we identified 151 microsatellites that were either polymorphic in the EST collection (Beldade *et al.* 2006) or that had relatively high repeat number (dinucleotides with greater than 5 repeats, trinucleotides with greater than 4 repeats, and tetra- and pentanucleotides with greater than 3 repeats). Primer3 software (Rozen & Skaletsky 2000; default settings unless noted) was used to design primers (optimal size = 22, and range = 20–25) to amplify a fragment of 100–150 bp around the microsatellite, avoiding priming low Phred and potential polymorphic bases (Beldade *et al.* 2006). Of the 151 original target microsatellites, this process led to automated design of 94 primer pairs (design failure was typically associated to microsatellites near the end of the available sequence). To increase the target group to 96, we manually selected two extra microsatellites from the 320. These were trinucleotides with 4 repeats in two unigenes annotated as Ribosomal Proteins (contigs 702 and 835 in Table S1). In total, primer pairs were designed to amplify 96 microsatellite loci in 94 unigenes, of which 60 were annotated based on sequence similarity with public databases (Beldade *et al.* 2006).

A 5′ M13-tag (TGTAAAACGACGGCCAGT) was added to each of the 96 forward primers to enable the use of a single fluorescently labelled primer (M13-FAM, *cf.* Schuelke 2000). 0.5 pmol forward and 2 pmol reverse

Table 1 Microsatellite markers in *Bicyclus anynana* derived from genomic and cDNA libraries

Source	Found	Assayed	PCR	Called	Mono morphic	Polymorphic
cDNA	320	96	76	68	28	40 [18] 2 alleles = 18 [6] 3 alleles = 15 [7] 4 alleles = 5 [3] >5 alleles = 2 [2]
gDNA	252	80	51	41	13	28 2 alleles = 6 3 alleles = 6 4 alleles = 10 >5 alleles = 6

Microsatellites identified in cDNA (Beldade *et al.* 2006) and gDNA (Van't Hof *et al.* 2005) libraries were typed in 11 and 28 outbred individuals, respectively. Counts are given for: (i) the total number of microsatellites identified (Found) following specific criteria (see Van't Hof *et al.* 2005; and Beldade *et al.* 2006), (ii) the subgroup for which primer pairs were designed and tested in genomic DNA (Assayed), (iii) primer pairs for which the PCR was successful (PCR), *i.e.* excluding those for which the microsatellite had either no bands on any of the test individuals or had bands for fewer than 5 of the 11 test individuals, (iv) microsatellites that we were actually able to type (Called), *i.e.* excluding failed PCR and inconclusive and (v) the number of Monomorphic and Polymorphic markers among the latter. Numbers in square brackets correspond to insertion-deletion polymorphisms or polymorphic microsatellites with imperfect repeats (see text).

primer, and 2 pmol M13-FAM (IDT) were used in a 12 µL PCR reaction with ca. 10 ng of genomic DNA (QIAGEN DNeasy), 0.3 U Taq polymerase (MasterTaq, Eppendorf) with buffer provided and enhancer additive (0.5× final concentration), and 10 nmol dNTPs. Denaturing at 95 °C during 3 min was followed by 10 cycles of

Table 2 EST-based microsatellite markers polymorphic in a laboratory population

ID	BLAST	Forward primer	Reverse primer	rep seq	rep nr	poly EST	alleles	size range	indiv	HW
120	1	GGTAAAACAATTCGATTATCAACA	TTGGACATCAATGATACTGAAGTTA	ATT	5	1	2	116–142	11	1
213	0	ACTTGAACCTAAAACACCAGCC	AGAAGTACGAACAACGAACACG	TGACT	5	0	3	104–119	10	0.543
248	0	TGACATAGAGAGCTCCTGGGAT	CACAGAGACATGAAGATGAGTGC	AC	6	1	2	118–120	11	0.542
349	1	AGATTTTGGGCGGAATCTACTT	AAAAGTACGACCCAACACAAGC	ACTA	4	0	2	146–148	8	**0.016**
438	1	GTTGTCGTTTTGTTGGGTGGTA	GTCAGAGACGCACCGATCTAAC	GAGT	8	1	2	100–104	8	0.200
455	1	AAATCGGACCAGTAGTTCCAGA	GAAATACGGAAAGAGGTTGACG	CAAA	5	1	3	123–131	11	0.143
572	0	TTGTGTTGCAATATGGTTAGGC	CAGCTGCAGTTTGTTACTACCAC	AATA	4	0	2	134–138	10	1
575	1	ACGTCCATCGCCTGTTCATTAT	TGCAAAATATCGTTTTATTTCAGA	ATG	5	0	3	142–146	10	1
698	1	GACATTATTTCACTTCAACAACAGG	AAATCCAATTCAAGGAATAACAAC	TA	6	0	2	127–129	11	0.108
780	1	CACGTAAAATGTTGCGTTTTATT	GAGAAGATGTCCAGATTCCTCG	AAT	5	0	3	147–156	11	1
818	1	TCGCATATTAAGTCTTCAAGCA	TAGACCCACTGAAATGAAAGGG	TTC	4	1	5	142–156	11	**0.005**
820	1	GCGTCGAGTGTGCTACAGTTAT	GCCTTACGAACAACACGAGTTTA	TTAT	4	0	4	117–128	10	1
902	1	TAACGGAGTTCGTTTATTTCGG	CAACACCGATTGTCCTCTACAA	AT	6	0	2	118–120	11	–
1006	1	TCAAATCGGACCAGTAGTTCCT	CCATTTTCAAGTTTGTCACTATTCT	CAAA	5	0	2	127–131	6	–
1369	0	TTGTTTTTGTTTTGTGCAATAAAGT	AGGTTGTCTCATGTAAGCGTTG	ATAA	6	0	2	127–135	11	0.232
1373	1	CACAGGCCAGTATTATGTAAGAGA	TTCATCATAGTGGGTACAACAAAAA	AC	7	0	5	153–160	11	1
1710	1	GCAATTTGCATTTTCAATACCA	CGTTGACTAACTCTTAGCTTTGACA	CA	6	0	3	117–123	11	0.050
1808	0	AAAATCCAACCCAAACAACAAC	TGTTGGTGATCACACTGAACAA	ATT	7	0	3	143–149	7	0.076
1876	1	TGAATCGCTTCATTCTTTCATA	AATTGAGGAAAATAACGCAGGT	GTAG	6	0	4	122–136	9	0.525
1899	0	ATTTGAATCGCGCATTTATTTT	TATGCAAATTTTATTTCGGGCT	ATA	6	0	2	156–158	10	1
2339	0	TGTTGAAAACAAAATTAACAAAACA	TCTGGATCCTGTAACAAACGTG	AT	6	0	4	131–137	11	**0.008**
2484	0	GGCGCGTACCAACTACATACTTA	TTGAAACAAAATAAACATTTTGGAA	ATT	5	0	2	123–130	10	1
2555	0	TTAAACTTCCTGGTCCACAGGT	TAAGTTTGACCCACCTTTTGCT	TTAA	4	0	2	140–144	11	–
2562	0	GTAATGTCTTCCCGGGTATTGA	AAAATATTGCGAGTAGGTTGCC	TAAA	7	0	3	135–143	8	0.513
2577	1	TTTTCTTCATTATGACCCCCTG	GATTTGCCAGAAGAACCAGAAA	TTC	5	0	2	152–163	11	0.255
2636	0	AGACGGGAGTGAGAAAAGAGC	GGCAGAACGAATATTTGAGTCC	TTCA	4	0	3	115–201	10	1
2669	0	TGTTGTTTAATTGGAAGTTTTCAGA	CGTGGGAAATTGGGAATATAGA	ACA	5	0	3	138–147	9	0.525
2904	1	AAATTAGCGTATCGATCAGTACAA	GAACTCAGTAACAATATGCTTGGAA	TAAT	4	0	2	136–140	11	1
3167	0	TGGAATGTTGTATTCTTGCAAAT	AGGCGTTGCTTAGATATTCTGG	ATT	5	0	2	156–159	11	0.279
3410	0	CCCATAATATTGGCTTAATGGT	GAGCGGTCCACTTAGTGAAGTT	TTTA	4	0	2	139–147	11	1
3457	1	CACACGGCTTTTTGTGTATTTT	CGATTTTGCTAAGCATGTTAGTTC	TTTA	5	0	4	155–160	11	0.509
3655	0	CCTAGTCATCATCCCGATTCTT	CATAGTCGGATAGTTCATCAATACA	TTA	5	0	2	127–130	10	1
3752	1	ACTTTCGAAACGTCCGGTAATA	CGATGACTAGATCACGGAAACA	ATAA	4	0	3	164–174	9	**0.021**
3833	1	AACCAGTATTCTTGCCACCATT	TGCCCTACAGTAAACATTCATCA	TCA	5	0	3	149–161	10	**0.011**
4122	1	ACCGCCTTTGCACATACTTATT	TCAATGCAAACAAACAAACAAA	AAAT	4	0	3	146–154	10	0.173
4122	1	TGTTATTGTTTTTGTTTTGTGCAAT	GCTTCCCAACAGTGGATTTTTA	GTTT	4	0	4	154–166	9	**0.020**
4190	0	ATTATGCACAGCACTCTGTCGT	TAGGATTGGAGAACGGAACGTA	CA	6	0	3	133–137	10	1
4316	0	TTCGCGAATATAACGTGAAAAA	TGTGGATATTTCACAGCAAAAA	GT	7	0	3	155–164	10	**0.001**
4431	1	TCCTACACTATGCGGTTCACTG	ACATCGACTACAACAACAACGC	TGT	5	0	2	101–122	11	1
4442	0	TGGGTTCGAAATAAACGCAT	CGGTACATTTCATCCTGTCAGA	TTTC	4	0	3	145–155	11	0.079

Data on the microsatellite markers identified *in silico* in a *B. anynana* EST collection (Beldade *et al*. 2006): *ID* is the contig number with underlining marking 'indels' and italics for markers tested also in a wild-caught panel (NCBI ESTdb accession numbers and contig consensus sequences are provided in Supplementary Table S1); *BLAST* is a binary code with 1 for unigenes annotated based on BLAST analysis against publicly available sequences, and 0 to those with no such annotation (details in Supplementary Table S1); *rep seq* is the microsatellite motif; *rep nr* is the number of repeats in the microsatellite in ESTs; and *poly EST* is a 0/1 code with 1 for the microsatellites found to be polymorphic in the EST collection. The remaining columns refer to the analysis in this paper: sequences of the Forward and reverse primers designed to type each microsatellite; *alleles* is the number of alleles found in the test panel; *size range* is the range for the length (bp) of the amplicons detected in the laboratory panel; *indiv* is the number of individuals from the panel genotyped for each microsatellite marker; *HW* is the p-value for the test for deviations from Hardy–Weinberg equilibrium [calculated with GENEPOP (Rousset 2008) based on the genotype data in Supplementary Table S1] with bold indicating significant deviations and - for alleles that did not meet the test criteria (see text). More details on these polymorphic markers, as well as details for the remaining markers tested (PCR failures and monomorphic microsatellites) are available in Supplementary Table S1.

94 °C (35 s), 58 °C (35 s) and 72 °C (45 s), then 30 cycles of 94 °C (35 s), 50 °C (35 s) and 72 °C (45 s), and finally 10 min at 72 °C (ABI 9700 thermal cycler). The amplicons thus generated were diluted 1:30 in water and paired so that two amplicons from the same individual but corresponding to microsatellites with expected PCR bands of different sizes (cf. Table S1) were further processed together. One μL of the plexed diluted amplicon mix was used with 9.5 μL of HiDi (ABI) and 0.5 μL of Genescan ROX ladder 50–350 bp (Gel Company) and run on the ABI 3100 Genetic Analyzer. Scoring of fluorescence peaks was performed using GeneMapper4.0 and confirmed manually.

Visual inspection of peak sizes in all test individuals led to the identification of different marker categories (Table 1). Of the 96 test microsatellites, nine had no detectable band in any of the individuals in the panel and 11 had bands in fewer than five of the 11 test individuals. Together, these 20 assays were scored as failed PCR reactions. PCR failures can result from unsuitable primers (e.g. due to sequencing errors in primed region; Long *et al.* 2007) and/or from the presence of introns in the genomic DNA (either leading to PCR failure or to amplicons of size greater than our maximum detection limit of 350 bp). Of the remaining primer pairs, eight produced peak patterns that were inconclusive and did not allow accurate genotyping of the test panel. These (8), together with the PCR failures (20), were excluded from the analysis. Twenty-eight of the remaining 68 loci were monomorphic in the test panel and 18 of the 40 polymorphic loci had alternative fragments differing in size by a number of nucleotides that was not a multiple of the repeat size. The latter can be insertion-deletion polymorphisms in the microsatellite-containing amplicon and/or non-perfect microsatellites, but remain potentially useful markers. Of the 40 polymorphic loci, 22 were associated with annotated unigenes (Table 2). Ten of the 40 markers might be of limited usefulness: seven showed heterozygote deficiency (commonly found in lepidopterans and usually caused by null alleles; Van't Hof *et al.* 2007) and three did not meet the criteria for Hardy–Weinberg analysis due to a low-frequency allele in a small sample (Table 2).

We selected 12 of the markers polymorphic in the laboratory population (italics in Table 2) to assay in a panel of 24 wild-caught individuals; 12 males and 12 females from South Africa. Three of them (contigs 213, 248, and 3655) had peak patterns that did not allow for clear genotyping and were discarded from further analysis. The remaining nine markers were polymorphic in the wild-caught panel. Three of them were not in Hardy–Weinberg equilibrium and two pairs were in linkage disequilibrium in the South African population (Table 3), but this was not significant when both test panels were taken into account ($P > 0.5$ for all pairwise tests involving the nine markers).

Compared with microsatellites in genomic libraries, those in ESTs generally have lower levels of polymorphism: fewer are polymorphic, they have fewer segregating alleles and the size difference between alleles is smaller (Table 1; Prasad *et al.* 2005). However, they have important advantages: there are typically fewer problems with PCR amplification when using primers designed

Table 3 EST-based microsatellite markers polymorphic in a wild-caught population

ID	rep size	alleles laboratory	alleles SA	indiv SA	HW SA	H_O SA	H_E SA	LD SA
455	4	3	3	19	**0.001**	0.158	0.496	
698	2	2	3	18	**0.013**	0.111	0.294	
780	3	3	3	23	0.292	0.261	0.308	a
902	2	2	3	22	1	0.409	0.394	
1369	4	2	4	18	**0**	0.167	0.567	
1710	2	3	2	19	0.080	0.053	0.149	
2555	4	2	3	23	0.362	0.273	0.369	b
2904	4	2	3	22	**0.038**	0.409	0.551	b
3410	4	2	3	23	0.260	0.565	0.456	a

ID is the contig number with underlining marking a possible 'indel' (*i.e.* alternative alleles differing by a number of nucleotides that is not a multiple of repeat size); *rep size* is the number of nucleotides in the microsatellite repeat; *alleles laboratory* is the number of alleles found in the laboratory panel of 11 butterflies; *alleles SA* is the number of alleles found in the panel of 24 test butterflies from a wild-caught South African population; *indiv SA* is the number of individuals from the South African panel that were genotyped for each microsatellite marker; *HW* is the p-value for the test for deviations from Hardy–Weinberg equilibrium in the South African panel with bold indicating significant deviations; *Ho* observed frequency of heterozygotes; *He* expected frequency of heterozygotes; *LD* corresponds to a letter code for groups of markers in linkage disequilibrium in the South African population: (a) $P = 0.0346$, (b) $P = 0.0352$, and empty cells having $P > 0.16$ for all pairwise LD tests. *HW* and *LD* were calculated with GENEPOP (Rousset 2008) based on the genotype data in the Supplementary Table S2. More details on these polymorphic markers, as well as details for the remaining markers tested (including three markers excluded from further analysis) are available in Supplementary Table S2.

against cDNA sequence and the markers are more likely to be transferable and comparable between species and are valuable in efforts to move from implicated genomic regions to implicated genes. Different recent studies have suggested that ESTs are a good source for microsatellite markers (Wren *et al*. 2000; Ellis & Burke 2007). Scanning EST collections for microsatellites can be especially valuable for groups where gDNA-based microsatellite development has proven challenging.

Acknowledgements

The authors wish to thank Stuart Macdonald and Steven Lockton for help with data collection and analysis, Maaike de Jong and Bas Zwaan for access to biological material from a wild population and Nicolien Pul for technical assistance with DNA preps. This work was supported by funds from the American National Science Foundation (NSF DEB-0235697 to ALD and PB), the Dutch Scientific Research Organization (NWO VENI 863.04.013 to PB) and the Portuguese Foundation for Science and Technology (FCT PTDC/BIA-BDE/65295/2006 to PB).

References

Beldade P, Brakefield PM (2002) The genetics and evo-devo of butterfly wing patterns. *Nature Reviews Genetics*, **3**, 442–452.

Beldade P, Rudd S, Gruber JD, Long AD (2006) A wing expressed sequence tag resource for *Bicyclus anynana* butterflies, an evo-devo model. *BMC Genomics*, **7**, 130.

Beldade P, Mcmillan WO, Papanicolaou A (2008) Butterfly genomics eclosing. *Heredity*, **100**, 150–157.

Beldade P, Saenko SV, Pul N, Long AD (2009) A gene-based linkage map for *Bicyclus anynana* butterflies allows for a comprehensive analysis of synteny with the lepidopteran reference genome. *PLoS Genetics*, **5**, e1000366.

Brakefield PM, Beldade P, Zwaan BJ (2009) The African butterfly *Bicyclus anynana*: evolutionary genetics and evo-devo. In: *Emerging Model Organisms: A Laboratory Manual, Volume 1*: (eds Behringer RR, Johnson AD & Krumlauf RE) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 291–330.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews. Genetics*, **5**, 435–445.

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity*, **99**, 125–132.

Fondon JW Iii, Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 18058–18063.

Jiggins CD, Mavarez J, Beltran M, Mcmillan WO, Johnston JS, Bermingham E (2005) A genetic linkage map of the mimetic butterfly *Heliconius melpomene*. *Genetics*, **171**, 557–570.

Kapan DD, Flanagan NS, Tobler A *et al*. (2006) Localization of Müllerian mimicry genes on a dense linkage map of *Heliconius erato*. *Genetics*, **173**, 735–757.

Kashi Y, King DG (2006) Simple sequence repeats as advantageous mutators in evolution. *Trends in Genetics*, **22**, 253–259.

Long AD, Beldade P, Macdonald SJ (2007) Estimation of population heterozygosity and library construction-induced mutation rate from expressed sequence tag collections. *Genetics*, **176**, 711–714.

Papanicolaou A, Gebauer-Jung S, Blaxter ML, Owen Mcmillan W, Jiggins CD (2008) ButterflyBase: a platform for lepidopteran genomics. *Nucleic Acids Research*, **36**, D582–D587.

Prasad MD, Muthulakshmi M, Madhu M, Archak S, Mita K, Nagaraju J (2005) Survey and analysis of microsatellites in the silkworm, *Bombyx mori*: frequency, distribution, mutations, marker potential and their conservation in heterologous species. *Genetics*, **169**, 197–214.

Pringle EG, Baxter SW, Webster CL, Papanicolaou A, Lee SF, Jiggins CD (2007) Synteny and chromosome evolution in the lepidoptera: evidence from mapping in *Heliconius melpomene*. *Genetics*, **177**, 417–426.

Rockman MV, Wray GA (2002) Abundant raw material for *cis*-regulatory evolution in humans. *Molecular Biology and Evolution*, **19**, 1991–2004.

Rousset F (2008) GENEPOP'007: a complete reimplementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103–106.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. (eds Krawetz S & Misener S), pp. 365–386. Humana Press, Totowa, NJ.

Schlötterer C (2004) The evolution of molecular markers – just a matter of fashion? *Nature Reviews Genetics*, **5**, 63–69.

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

Van't Hof AE, Zwaan BJ, Saccheri IJ, Daly D, Bot ANM, Brakefield PM (2005) Characterization of 28 microsatellite loci for the butterfly *Bicyclus anynana*. *Molecular Ecology Notes*, **5**, 169–172.

Van't Hof AE, Brakefield PM, Saccheri IJ, Zwaan BJ (2007) Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera. *Heredity*, **98**, 320–328.

Van't Hof AE, Marec F, Saccheri IJ, Brakefield PM, Zwaan BJ (2008) Cytogenetic characterization and AFLP-based genetic linkage mapping for the butterfly *Bicyclus anynana*, covering all 28 karyotyped chromosomes. *PLoS ONE*, **3**, e3882.

Wang B, Porter AH (2004) An AFLP-based interspecific linkage map of sympatric, hybridizing Colias butterflies. *Genetics*, **168**, 215–225.

Wren JD, Forgacs E, Fondon JW Iii *et al*. (2000) Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. *American Journal of Human Genetics*, **67**, 345–356.

Yasukochi Y, Ashakumary LA, Baba K, Yoshido A, Sahara K (2006) A second generation integrated map of the silkworm reveals synteny and conserved gene order between lepidopteran insects. *Genetics*, **173**, 1319–1328.

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

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Details for all 96 microsatellites tested in the laboratory panel

Table S2 Details for the 12 microsatellites tested in the wild-caught panel

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

doi: 10.1111/j.1755-0998.2009.02691.x

© 2009 Blackwell Publishing Ltd

Characterization of microsatellite loci isolated from the wasp, *Microstigmus nigrophthalmus* (Hymenoptera)

ERIC R. LUCAS,*† GAVIN J. HORSBURGH,* DEBORAH A. DAWSON* and JEREMY FIELD†

**NERC Molecular Genetics Facility, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK, †Department of Biology and Environmental Science, University of Sussex, Brighton BN1 9QG, UK*

Abstract

Fifty-two microsatellite loci were characterized in 22–31 unrelated females of the wasp (*Microstigmus nigrophthalmus*) collected from the Mata do Paraiso, Viçosa, M.G., Brazil. Fifty-one of these loci were developed from a microsatellite-enriched genomic library derived from *M. nigrophthalmus* and one was derived from the wasp, *Ormyrus nitidulus*. The genus *Microstigmus* represents an independent origin of social behaviour in the Hymenoptera and is thus of great potential in the study of social evolution.

Keywords: Hymenoptera, microsatellite, *Microstigmus nigrophthalmus*, social organization, wasp

Received 21 December 2008; revision accepted 24 March 2009

Microstigmus is a neotropical genus of apoid wasp characterized by the construction of nests using a combination of external organic material and silk produced by adult females. In contrast to most other apoid wasps, *Microstigmus* demonstrates evidence of social organization (Matthews 1968, 1991). The species *Microstigmus nigrophthalmus* (Melo 1992) lives in groups of one to five females (Melo & Campos 1993) and has great potential for use in the study of social evolution. We describe the identification of a set of polymorphic dinucleotide microsatellite loci for *M. nigrophthalmus*.

Genomic DNA was extracted from the thorax and abdomen of a single female *M. nigrophthalmus* pupa (046P1) captured from the Mata do Paraiso, Viçosa, M.G., Brazil. Genomic DNA was digested with *Mbo*I and size-selected (250–750 bp). The restriction fragments were enriched for $(CA)_n$ and $(GA)_n$ and their complements, as described by Armour *et al.* (1994) but without the pre-enrichment hybridization PCR-amplification step. Enriched fragments were ligated into *Bam*HI-digested, CIP-dephosphorylated pBluescript SK+ (Stratagene) and screened for $(CA)_n$ and $(GA)_n$ and their complements. Positive clones were sequenced in both directions, a consensus sequence created and primers designed using Primer 3 (Rozen & Skaletsky 2000). One-hundred and seven unique sequences were identified using BLASTN v.2.2.4 (Altschul *et al.* 1997, EMBL accession numbers FM957351–FM957457). Primer sets were designed for 86 loci with at least seven tandem repeat units.

As a result of the haplodiploid sex determination system found in the Hymenoptera, no part of the nuclear genome exists exclusively in one sex. Adult males and females of this species can be distinguished based on morphology. In addition, males are haploid, whereas females are diploid. All individuals genotyped for the characterization of our loci were female. Thirty-seven females were captured from the Mata do Paraiso (see above) and stored in 1 mL of absolute ethanol in screw-capped microfuge tubes at room temperature for several months, then in a −20 °C freezer. Genomic DNA was extracted from whole thoraces and each locus tested for amplification and polymorphism with between 22 and 31 unrelated individuals from the Mata do Paraiso population. Each 10-μL PCR contained approximately 50 ng of genomic DNA, 1.0 μM of each primer, 0.20 mM of each dNTP, 2.0 mM $MgCl_2$ and 0.25 units of *Taq* DNA polymerase (Biotaq; Bioline) in the manufacturer's buffer. PCR

Correspondence: Jeremy Field, Fax: +44 (0) 1273 678937; E-mail: j.field@sussex.ac.uk