

# Microsatellites for the Marsh Fritillary Butterfly: *De Novo* Transcriptome Sequencing, and a Comparison with Amplified Fragment Length Polymorphism (AFLP) Markers

Melanie R. Smee<sup>1,2\*</sup>, Yannick Pauchet<sup>2,3</sup>, Paul Wilkinson<sup>2,4</sup>, Brian Wee<sup>5</sup>, Michael C. Singer<sup>6</sup>, Richard H. ffrench-Constant<sup>2</sup>, David J. Hodgson<sup>2</sup>, Alexander S. Mikheyev<sup>7</sup>

**1** Department of Biology, University of York, York, United Kingdom, **2** Centre for Ecology and Conservation, School of Biosciences, University of Exeter, Penryn, Cornwall, United Kingdom, **3** Max Planck Institute for Chemical Ecology, Jena, Germany, **4** School of Biological Sciences, University of Bristol, Bristol, United Kingdom, **5** National Ecological Observatory Network Incorporated, Washington, D. C., United States of America, **6** Section of Integrative Biology, University of Texas at Austin, Austin, Texas, United States of America, **7** Okinawa Institute for Science and Technology, Onna-son, Kunigami-gun, Okinawa, Japan

## Abstract

**Background:** Until recently the isolation of microsatellite markers from Lepidoptera has proved troublesome, expensive and time-consuming. Following on from a previous study of Edith's checkerspot butterfly, *Euphydryas editha*, we developed novel microsatellite markers for the vulnerable marsh fritillary butterfly, *E. aurinia*. Our goal was to optimize the process in order to reduce both time and cost relative to prevailing techniques. This was accomplished by using a combination of previously developed techniques: *in silico* mining of a *de novo* assembled transcriptome sequence, and genotyping the microsatellites found there using an economic method of fluorescently labelling primers.

**Principal Findings:** In total, we screened nine polymorphic microsatellite markers, two of which were previously published, and seven that were isolated *de novo*. These markers were able to amplify across geographically isolated populations throughout Continental Europe and the UK. Significant deviations from Hardy-Weinberg equilibrium were evident in some populations, most likely due to the presence of null alleles. However, we used an  $F_{st}$  outlier approach to show that these markers are likely selectively neutral. Furthermore, using a set of 128 individuals from 11 populations, we demonstrate consistency in population differentiation estimates with previously developed amplified fragment length polymorphism (AFLP) markers ( $r = 0.68$ ,  $p < 0.001$ ).

**Significance:** Rapid development of microsatellite markers for difficult taxa such as Lepidoptera, and concordant results with other putatively neutral molecular markers, demonstrate the potential of *de novo* transcriptional sequencing for future studies of population structure and gene flow that are desperately needed for declining species across fragmented landscapes.

**Citation:** Smee MR, Pauchet Y, Wilkinson P, Wee B, Singer MC, et al. (2013) Microsatellites for the Marsh Fritillary Butterfly: *De Novo* Transcriptome Sequencing, and a Comparison with Amplified Fragment Length Polymorphism (AFLP) Markers. PLoS ONE 8(1): e54721. doi:10.1371/journal.pone.0054721

**Editor:** Casper Breuker, Oxford Brookes University, United Kingdom

**Received:** September 27, 2012; **Accepted:** December 14, 2012; **Published:** January 21, 2013

**Copyright:** © 2013 Smee et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was funded by a BBSRC CASE studentship awarded to RHffC. ASM is supported by funding from the Okinawa Institute for Science and Technology (OIST). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: melaniesmee@gmail.com

## Introduction

Microsatellite markers are an important tool in population genetics and conservation biology. They typically have high levels of polymorphism and generate reliable data on genetic diversity, population structure and gene flow in highly protected species [1,2]. At the same time, microsatellites can be amplified from trace amounts of DNA, and permit non-lethal and non-invasive sampling. Although many new genotyping techniques have become available, microsatellites will likely remain relevant for the foreseeable future, given their low input DNA requirements and power for producing reliable genotypes suitable for calculation of familial relationships and for population genetic studies. For

many species the usual route to isolating and characterizing microsatellite markers involves a time-consuming process of enriching a genomic DNA library for a variety of di-, tri- or tetra-nucleotide repeat motifs, followed by a process of cloning and sequencing, to finally design species-specific primers [3–5]. The design and thorough testing of optimal multiplexes then requires the purchase, for each locus being genotyped, of an expensive primer carrying a fluorescent dye label. Despite this cost, microsatellites remain one of the most popular and useful molecular markers for studies of population structure and conservation decision-making. Capillary DNA sequencers automate microsatellite processing, permitting up to 96 samples to be

**Table 1.** Characteristics of 14 microsatellite loci in *Euphydryas aurinia* (Lepidoptera: Nymphalidae).

| PCR # | Locus      | Primer sequences (5' to 3')                                 | M13 Label | Repeat motif  | Range (bp) | N <sub>A</sub> | H <sub>O</sub> | H <sub>E</sub> | % missing | Average % null alleles | GenBank Acc. no. |
|-------|------------|---|-----------|---|------------|----------------|----------------|----------------|-----------|------------------------|------------------|
| 1     | Aurinia_01 | F:CGAGCGTATTTGTCAAAGAAAG<br>R:AGCGAATTAGGGTTGTACATT         | HEX       | (CAT) <sub>6</sub>  | 252–270    | 6              | 0.396          | 0.503          | 8.1%      | 10.4                   | JN116271         |
| 2     | Aurinia_13 | F:AACGTTAACTAGGGGTCTCA<br>R:TATGATATAGTGTACGCGGTTTT         | TET       | (ATT) <sub>6</sub>  | 228–237    | 4              | 0.332          | 0.294          | 0.85%     | 0.9                    | JN116273         |
| 2     | Aurinia_16 | F:CCCCTATGATCCATGTTTTA<br>R:AAATTCATTTGTACTTTCCGGTACAT      | TET       | (TTA) <sub>6</sub>  | 173–204    | 6              | 0.343          | 0.367          | 0.2%      | 4.6                    | JN116275         |
| 4     | Aurinia_18 | F:AAAAGCGCTGAAAGAAGAAAAA<br>R:CAGTCTCAAAGATTTTCGCATAAAA     | TET       | (TAT) <sub>5</sub>  | 189–193    | 4              | 0.327          | 0.384          | 1.1%      | 4.0                    | JN116276         |
| 1     | Aurinia_45 | F:GGTGAAATTCGGAATGAGT<br>R:TCCCCGCTACAGATGAAATC             | HEX       | (GTT) <sub>6</sub>  | 193–213    | 8              | 0.452          | 0.566          | 3.4%      | 8.9                    | JN116280         |
| 3     | Aurinia_64 | F:CAACCTGTAGCCGAAAAGA<br>R:GCTTTTCTGTTGCCATCGTT             | TAM       | (TAC) <sub>8</sub>  | 201–210    | 4              | 0.188          | 0.203          | 4.0%      | 3.5                    | JN116282         |
| 1     | Aurinia_70 | F:CAACTTCAGTATGATCTCATTGCTTT<br>R:TCACAATTTGCAGTGGCTCTAT    | HEX       | (GA) <sub>10</sub>  | 130–144    | 8              | 0.431          | 0.370          | 1.5%      | 1.1                    | JN116283         |
| 1     | EA26*      | F:CCGAGACTACACCTACAAG<br>R:CAGTGTATTTCCGGAACACAG            | HEX       | (TG) <sub>5</sub><br>TA(TG) <sub>11</sub><br>(AGCG) <sub>4</sub><br>(TG) <sub>3</sub> | 163–180    | 13             | 0.705          | 0.730          | 5.7%      | 4.2                    | AY491806         |
| 5     | EA51*      | F:TGACGACAGATGGGTGTTTC<br>R:TGTAAGCGACTCAGTCTCATTTT         | HEX       | (TGTA) <sub>9</sub><br>(TGAT) <sub>2</sub>  | 129–142    | 4              | 0.483          | 0.622          | 5.1%      | 11.9                   | AY491828         |
| -     | Aurinia_14 | F:TTTGATGGGGAGAATTTATTGTTT<br>R:TTTCTTTAATCACAGATAACCTTTTT  | -         | (AT) <sub>8</sub>   | 226–232    | 4              | -              | -              | -         | -                      | JN116274         |
| 3     | Aurinia_23 | F:TACAAGCGCTTACCGAAGAAAC<br>R:TCTGTCTGTTTCATCGCTCTCA        | TAM       | (GAT) <sub>7</sub>  | 147–162    | 5              | -              | -              | -         | -                      | JN116277         |
| 2     | Aurinia_31 | F:CAATTTAGGCGCAAATTAAGA<br>R:CCCAGCTAAGCGACTACACTA          | TET       | (AAC) <sub>5</sub>  | 252–258    | 2              | -              | -              | -         | -                      | JN116278         |
| 2     | Aurinia_35 | F:CAAGAACATGAATTTAGGTAAGCA<br>R:CATAAGATTATGCCGGTATATAAAGTT | TET       | (TAT) <sub>5</sub>  | 122–130    | 4              | -              | -              | -         | -                      | JN116279         |
| 1     | Aurinia_62 | F:TGTAGGCGACGTTCTTGCTA<br>R:AATGCATTTTCGCATTCTAGG           | HEX       | (AAT) <sub>5</sub>  | 240–243    | 2              | -              | -              | -         | -                      | JN116281         |

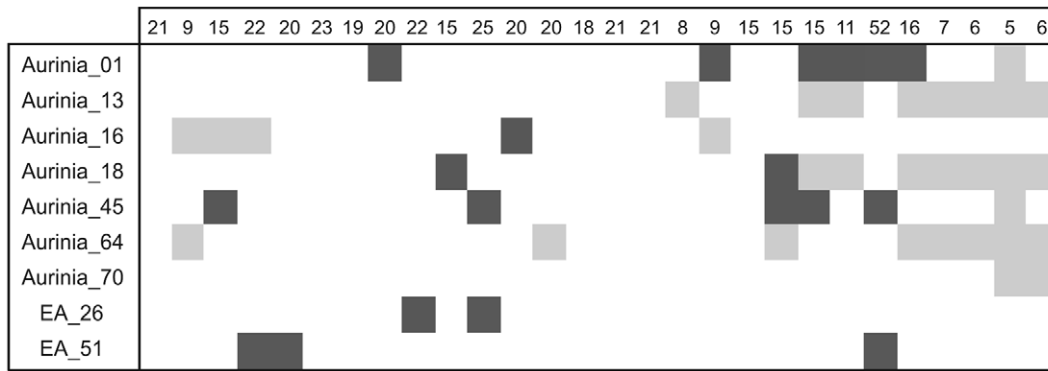
Summary statistics based on a survey of 468 individuals from 28 populations. N<sub>A</sub>, number of observed alleles; H<sub>O</sub> observed and H<sub>E</sub> expected heterozygosity per locus. All Forward primers had M13 tail (TGTAACGACGCCAGT) added to 5' end, as described in text. Loci below the line were lost due to a lack of amplification across all samples, or because of null alleles, but may prove of use in other populations or closely-related species. PCR reactions 1, 2 and 3 were pooled and analyzed together in the same sequencer run, as were reactions 4 and 5— see text for description. \*Primers were originally published by Petenian *et al.* [16]. doi:10.1371/journal.pone.0054721.t001

run simultaneously. Scoring of alleles can likewise be automated, although human verification remains necessary.

Microsatellites often need to be isolated *de novo* when a sequenced genome is unavailable. Although a wide variety of approaches for isolating microsatellites exist [6] they fail frequently with lepidopteran genomes [7–10]. This inherent difficulty may be associated with a lack of polymorphism, similarity in flanking regions of different microsatellite loci, presence of null alleles or possible associations with mobile elements [7–9,11,12]. However, there are still a large number of studies detailing the isolation of microsatellite markers from Lepidoptera [13–18]. Inevitably it seems, these studies usually show a low number of loci, low levels of polymorphism, high occurrence of null alleles, or strong deviations from Hardy-Weinberg equilibrium (HWE) [12–14,16,19]. Often, it is also the case that the developed markers prove not to be transferable between geographically distant

populations of the same species. For example, in a recent study on the marsh fritillary butterfly, *Euphydryas aurinia*, three out of an initial 96 loci showed evidence of transferability among geographically distant populations and sub-species [20]. In studies where investigations into geographical patterns of population differentiation are the main aim, such markers cannot be used.

The recent advent of next generation sequencing and the resulting increase in publicly available expressed sequence tag (EST) data has encouraged quick and easy isolation of microsatellite markers [15,20–22]. With the future cost of DNA sequencing decreasing at a rate exceeding Moore's law, the methods used here are becoming much more accessible to a wider research community. However, microsatellites associated with expressed regions of the genome may be under selection, even if they represent untranslated regions (UTRs) of genes, and hence often show strong deviations from HWE [15,21]. Fortunately, studies



**Figure 1. Hardy-Weinberg equilibrium (HWE) statistics.** ‘Exact test’ statistics for each locus, and for each of the 28 populations included in the study (each column represents a population with sample size in the top row, with the populations occurring in the same order, left to right, as numbered in Figure S1 and Table S1), after a sequential Bonferroni correction for multiple tests [43]. Dark shading indicates significant deviation from HWE (chi-squared test,  $p < 0.05$ ); light grey shading indicates monomorphic loci, or only a single copy of a second allele if present. For loci with less than five alleles the complete enumeration method was used [51] but for all others a Markov chain algorithm was implemented [52] in GENEPOP v.4.1 [37].

doi:10.1371/journal.pone.0054721.g001

that have compared EST-derived microsatellites with other genotyping methods have generally found similar results in studies of population genetics [15,23,24].

*E. aurinia* populations have severely declined in both the UK and continental Europe [25–27]. Therefore, a current need exists for tools to illuminate population genetic structure throughout this species, in order to aid the setting of conservation priorities. To this end, we report the characterization of seven novel microsatellite loci using next-generation 454-pyrosequencing methods and confirmation of two previously published loci for *E. aurinia*. These markers give concordant results for estimates of population differentiation with amplified fragment length polymorphism (AFLP) markers on the same samples [28]. Their ability to amplify across large geographical regions is also demonstrated. We therefore confirm that these new markers are suitable for use in a conservation genetics study with the aim of informing priorities and targets by determining population structure and gene flow across a fragmented landscape.

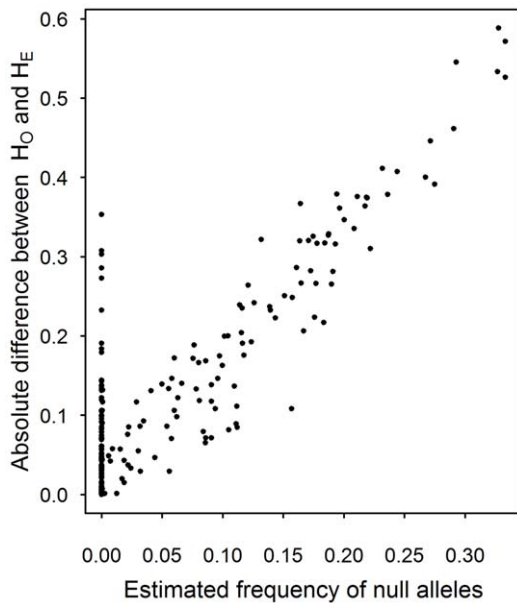
**Materials and Methods**

RNA was extracted from a pool of 64 *E. aurinia* larvae collected from 16 sites in south-west UK (under Natural England licence number 20081071), using TRIzol (Invitrogen). Genomic DNA contamination was removed by DNase treatment (TURBO DNase, Ambion) for 30 min at 37°C. RNA was further purified by using the RNeasy MinElute Clean up Kit (Qiagen) following the manufacturer’s protocol. Full-length, enriched, cDNAs were generated from 2 µg of total RNA using the SMART PCR cDNA synthesis kit (BD Clontech) following the manufacturer’s protocol. Reverse transcription was performed with the PrimeScript reverse transcriptase (Takara) for 60 min at 42°C and 90 min at 50°C. In order to prevent over-representation of the most common transcripts, the resulting double-stranded cDNAs were normalized using the Kamchatka crab duplex-specific nuclease method (Trimmer cDNA normalization kit, Evrogen) [29]. The resulting normalized cDNA library was sent to the Advanced Genomics facility at the University of Liverpool (<http://www.liv.ac.uk/agf>) for sequencing on the Roche 454 GS-FLX pyrosequencing platform. The assembly of the obtained reads was achieved using MIRA v2.9.26x3 ([http://sourceforge.net/apps/mediawiki/mira-assembler/index.php?title=Main\\_Page](http://sourceforge.net/apps/mediawiki/mira-assembler/index.php?title=Main_Page)). The mira assembler was

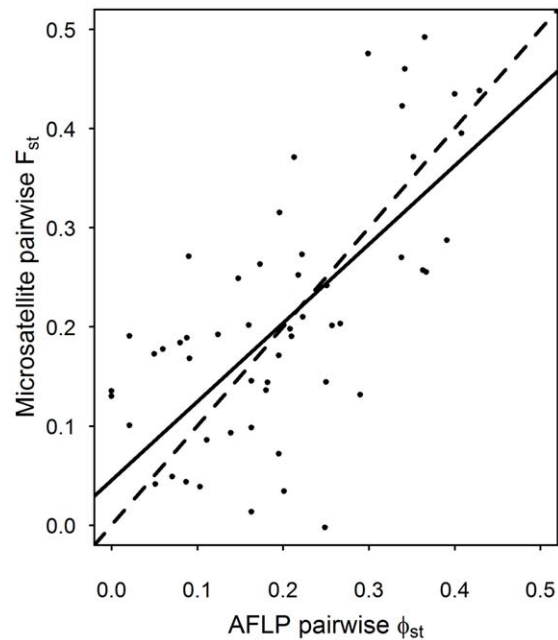
run using a perl wrapper script (EST2assembly [30]) which parameterizes and selects the optimal assembly, based on coverage, redundancy and the proportion of coverage to a reference transcriptome. The mira command line for the optimal assembly was: `mira-job = denovo, est, accurate,454-fasta-OUT: rrol = 1:rdl = 1:orc = 1:org = 0:ora = 0:ors = 0:otf = 0:otc = 0-GE: not = 1-CO:asir = 1-LR:mxti = 1-AS:sd = 0:uess = 0:urd = 0:ard = 1-SKIM:mmhr = 2:mnr = yes 454_SETTINGS-DP:ure = 0-CO:mrpg = 10-CL:pvlc = 0:cpat = 0:mbc = 1:mbcgs = 30:mbcmfg = 30:mbcmeg = 30:qc = 0-ED:ace = 1-AL:egp = no-ALIGN:bip = 20:bmax = 120:mo = 10:mrs = 80.`

The assembled data were queried for ≥5 perfect repeats of di-, tri- and tetra-nucleotide repeats using Msafinder version 2.0.9 [31]. Primers were then designed using Primer3 for 74 contigs containing microsatellites [32]. Forward primers were designed with an addition of a M13 tail at the 5’ end (M13: 5’-TGTAACGACGGCCAGT-3’) following the methods of Schuelke [33]. This allows the use of fewer expensive fluorescently labelled primers, since only one ‘universal’ fluorescently labelled M13 primer is required per multiplex. PCR conditions are such that a small quantity of unlabelled sequence-specific forward primer with a M13 tail is incorporated first, in partnership with a sequence-specific reverse primer. Once the forward primer is depleted, the annealing temperature is lowered so that the ‘universal’ fluorescently labelled M13 forward primer anneals instead (its annealing temperature is a few degrees lower). The fluorescent dye is thus incorporated into the PCR product [33]. A differently labelled M13 primer can be used for each multiplex and then all multiplexes can be combined when sequencing. We also synthesized previously published primers of *E. aurinia* microsatellite markers [16,34] with the addition of a M13 tail and tested these on our samples.

*E. aurinia* (Lepidoptera; Nymphalidae) is a univoltine species, with gregarious larvae [35]. A single 3rd/4th instar larva was collected from each of 281 communal webs found in September 2008 and 2009 (Natural England licence numbers 20082745 and 20091722; Countryside Council for Wales licence OTH:SPCA:09:2008; and Scottish Natural Heritage licence 9154) from a total of 16 sites spanning the butterfly’s mainland UK range (Figure S1). Each larva collected from south-west UK, Scotland and Wales was starved for 24 hours to remove plant material from the digestive tract and then dissected to remove the



**Figure 2. Deviation from HWE and estimated frequency of null alleles.** The absolute difference between expected and observed heterozygosity in a population, as a measure of deviation from HWE, plotted against the estimated frequency of null alleles per marker per population for all nine loci to be used in future population genetic studies. Each data point represents one locus in a specific population. doi:10.1371/journal.pone.0054721.g002



**Figure 3. A comparison of microsatellite and AFLP markers on the same samples.** A comparison of pairwise population differentiation (uncorrected  $F_{st}$  and  $\Phi_{st}$ ) estimated using the seven novel microsatellites developed in the current study and AFLP markers developed in a previous study by Wee [28] on the same set of 11 European populations (solid line). A dashed line is also included to show where a perfect correspondence between the two would lie. doi:10.1371/journal.pone.0054721.g003

midgut and potentially any parasitoids, before then being snap frozen. Following the methods of Martinez-Torres *et al.* [36] with slight modification (38  $\mu$ l of 3 M potassium acetate, pH 5.2, was used and the samples then incubated at  $-80^{\circ}\text{C}$  for 15 mins), genomic DNA was extracted, diluted to a working concentration ( $\sim 15$  ng/ $\mu$ l), and stored at  $-20^{\circ}\text{C}$  until required. DNA from samples collected from 12 sites in France and Catalonia (Figure S1) was extracted from the heads of larvae using DNeasy Tissue kits (Qiagen) [28]. All 74 designed primer pairs were initially tested for amplification on four samples from geographically widespread populations in the UK (site numbers 1, 6, 8 and 11: see Table S1 and Figure S1 for details and geographical distribution of all populations used in the study). Each 10  $\mu$ l PCR mix contained approximately 1.5 ng genomic DNA, 1  $\mu$ l 10x EX Buffer (Takara), 2  $\mu$ M dNTPs, 0.25 U EX Taq HS (Takara), 1  $\mu$ l 10 mg/ml BSA, 10 pmol fluorescently labelled M13 forward primer, 10 pmol reverse primer and 2.5 pmol of the sequence specific forward primer. Using a PTC-200 Thermo Cycler (MJ Research) the PCR conditions were as follows:  $95^{\circ}\text{C}$  for 5 mins; followed by 25 cycles of  $95^{\circ}\text{C}$  for 5 s,  $60^{\circ}\text{C}$  for 30 s,  $68^{\circ}\text{C}$  for 1 min; followed by 8 cycles of  $95^{\circ}\text{C}$  for 5 s,  $53^{\circ}\text{C}$  for 30 s,  $68^{\circ}\text{C}$  for 1 min; and ended with a final extension at  $72^{\circ}\text{C}$  for 30 mins. A maximum of 2  $\mu$ l of PCR product was then added to 10  $\mu$ l HiDi formamide (Applied Biosystems) and 0.1  $\mu$ l GeneScan 500 ROX standard (Applied Biosystems) and run on an ABI 3130xl Genetic Analyzer (Applied Biosystems). Allele presence and sizes were assigned using GeneMarker 1.7 (SoftGenetics, LLC).

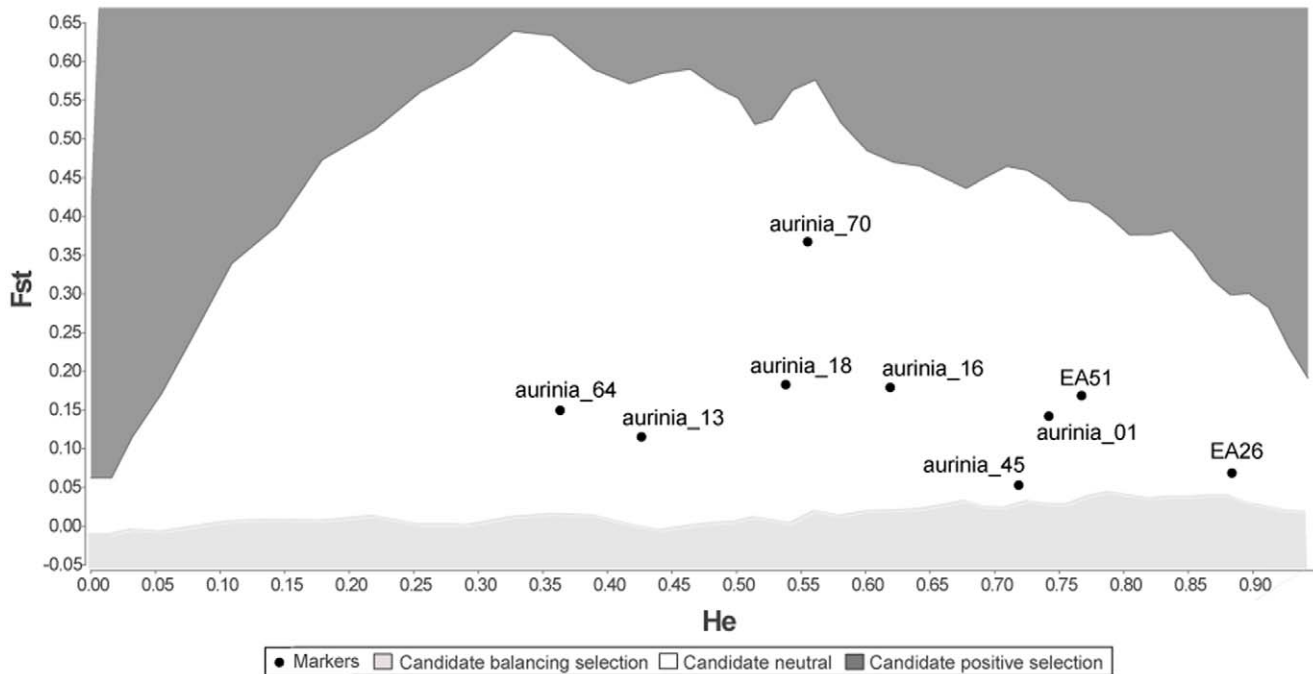
Potentially polymorphic loci were then tested on a further four samples from different populations in the UK (site numbers 9, 12, 13 and 14: see Table S1 and Figure S1). For loci that amplified across all eight samples and appeared polymorphic, we designed multiplexes using PCR products with non-overlapping size ranges. Multiplexes using differently labelled M13 ‘universal’ forward primers were then mixed together and analyzed on an ABI 3100

capillary sequencer (on average 0.5  $\mu$ l of each multiplex was added to the ROX and HiDi formamide mix, but this was adapted according to the intensity of the signal). All microsatellite sequences were submitted to GenBank under accession numbers JN116271 and JN116273 to JN116283.

GENEPOP v4.1 was used to assess deviations from HWE [37]. Estimates of null allele frequencies were achieved following the Expectation Maximization (EM) algorithm of Dempster *et al.* [38] in the online programme FreeNA [39]. The programme LOSITAN was then used to identify putative loci under selection across our samples, using a method of comparing  $F_{ST}$  values against expected heterozygosity ( $H_E$ ) to detect outliers [40,41]. As a subset of samples ( $N = 128$ ) had been previously genotyped using AFLP markers [28], the concordance between this previous analysis and the current study was assessed by comparing the  $\Phi_{st}$  and  $F_{st}$  matrixes generated by the two studies, respectively, using a Mantel test with 10,000 bootstrap replicates. Both  $\Phi_{st}$  and  $F_{st}$  matrixes were computed using Arlequin v3.5 [42].

## Results

The 454 run generated 186,835 reads, and a total of 40,865,774 bases, after quality filtering. These were assembled into 22,032 contigs whereby 50% were equal to or larger than a size of 932 bases ( $N_{50}$  size) resulting in an average 3.15-fold coverage. Primers were designed for 74 microsatellite loci out of the 97 found in the 454 transcriptome, as some were too close to the edge of the contig for primer placement. Across an initial sample of four individuals, 36 of the novel loci did not amplify at all, 24 appeared monomorphic and 14 were potentially polymorphic. Out of the eleven previously published loci, only two appeared polymorphic in our samples. Testing across another four samples resulted in



**Figure 4. LOSITAN results for all nine loci across all samples.** This workbench uses an  $F_{st}$  outlier method to detect potential loci under selection and therefore unsuitable for use in studies reliant on neutral markers, such as population genetics. Under a stepwise mutation model (SMM), all loci were found to be candidate neutral markers – located in the white region of the figure. The dark grey region above indicates where candidate positive selection markers would be found, and the lighter region below indicates where candidate balancing selection markers would be found. Under an infinite alleles model (IAM) of mutation, the result is the same. doi:10.1371/journal.pone.0054721.g004

discarding of two of the novel primers due to a lack of amplification across all samples, and possible null alleles (whereby a mutation in the primer annealing site prevents amplification). The remaining loci (see Table 1) were developed into multiplex sets to enable all samples to be genotyped in just two sequencing mixes. The locus Aurinia\_14 was discarded during multiplex design as it did not successfully amplify with other primer combinations, or even reliably amplify alone. Loci Aurinia\_23B, Aurinia\_31B, Aurinia\_62 and Aurinia\_35 either did not amplify well or gave large amounts of missing data, and thus were excluded from further analyses leaving nine loci for large-scale genotyping. In many of the sampled populations there was significant deviation from HWE even after sequential Bonferroni correction for multiple tests [43] (Figure 1).

These deviations from HWE are likely due to the presence of null alleles in many cases. Analyses using FreeNA [39] gave estimated null allele frequencies ranging from 0 to 0.33 per marker per population, which were highly correlated with the absolute difference between expected and observed heterozygosities – a measure of deviation from HWE (Figure 2).

The extent to which deviations from HWE may affect typical analyses using EST-derived markers is unclear. We were able to validate our markers using AFLP-derived distances computed for the same samples in an earlier study, despite the presence of null alleles [28]. Hence, we show very similar results for estimates of pairwise population differentiation using only the seven novel microsatellite loci developed in this study and AFLP markers developed previously [28] on a subset of the populations analysed here (Mantel test:  $r = 0.68$ ,  $n = 11$ ,  $p < 0.001$ , Figure 3). Adding the two previously developed microsatellite loci, EA26 and EA51 (Table 1), further strengthens this relationship (Mantel test:  $r = 0.72$ ,  $p < 0.0005$ ). The occurrence of null alleles can be

corrected for when calculating  $F_{st}$  and genetic distances in studies of population structure [39]. However, this correction does not explain any significant geographic variation in population structure, relative to the uncorrected matrix (partial Mantel test of geographic distance and the corrected  $F_{st}$  matrix, using the uncorrected matrix as the covariate;  $p = 0.23$ ), meaning our results are robust and do not change significantly after correction.

We were also able to demonstrate that all loci fell into the candidate neutral category after using the programme LOSITAN [40,41] to identify putative loci under selection across all our samples (Figure 4). Four pairs of loci showed evidence of linkage disequilibrium (LD) across all populations (EA26-Aurinia\_45, EA26-Aurinia\_01, EA51-EA26, EA51-Aurinia\_45) however, following a sequential Bonferroni correction only one remained marginally significant (EA51-Aurinia\_45).

## Discussion

This study has successfully isolated and characterized microsatellites for a threatened lepidopteran species using a rapid method combining data mining for perfect repeats with inexpensive primer labelling. Schuelke's [33] M13 primer protocol has particular advantages for small research groups performing low-throughput genetic analyses, by reducing costs of fluorescently labelled primers, particularly useful for initial primer screening, where primers may be used possibly only once.

Effective and reliable DNA markers are desirable for taxa with important roles in agriculture, ecology and horticulture, such as the Lepidoptera. The markers developed in this study originated from a 454-pyrosequencing run using first-generation chemistry, and hence likely do not represent the entire breadth of markers to be found. The low coverage of the transcriptome and the potential

for assembly errors, especially where there are homopolymer runs, may explain why so many of the initially designed primers failed to amplify. There were also time constraints, which highlights the efficiency of our approach, taking a full time researcher only two months from receiving primers to analysis.

This study also confirmed the successful use of two established microsatellite loci for the study species, although they both showed considerable deviations from HWE across the sampled populations [16]. Such deviations in HWE, aside from being the result of null alleles, can be expected in *E. aurinia* and other Lepidoptera with similar life histories, for several biologically plausible reasons. Butterflies in the subfamily Melitaeinae, to which the genus *Euphydryas* belongs, typically suffer frequent population extinctions and their persistence in a landscape can depend entirely on metapopulation dynamics [44]. Although population structure is often discrete [45] and most individuals travel very short distances in their lives, occasional longer distance dispersal events occur [46,47]. In *E. aurinia*, mean clutch size is several hundred eggs [48], such that each female is unlikely to lay more than two clutches in her lifetime. Effective population size is reduced by the correlated survival of siblings that results from this large clutch size and by extreme population fluctuations [49]. Also affecting HWE expectations are the high levels of inbreeding found in small subpopulations within metapopulations [50], which is an unavoidable result of habitat fragmentation and stochastic patch dynamics. Our final analysis using LOSITAN [40,41] confirms that all loci are putatively neutral, and so deviations are likely to be due to actual population structure and dynamics, but also, significantly, due to the presence of null alleles across the markers themselves. Studies using these markers to compute Hardy-Weinberg equilibrium should make corrections for the presence of null alleles, although  $F_{st}$ -based measures of population differentiation appear relatively robust (see also Mikheyev *et al* [15]).

The markers developed here from a 454 transcriptome fulfil a large number of desired criteria for developing 'robust' microsatellite markers for *E. aurinia* [20]. This includes perfect repeat motifs (neither compound microsatellites nor with any interruptions), being PCR 'multiplexable', and being transferable between geographically distant populations [20]. The one caveat, or unfulfilled criterion, is that some loci in some populations may have substantial numbers of null alleles. Sinama *et al* [20] recently developed microsatellite markers for *E. aurinia* using a combination of a biotin-enrichment protocol and next generation pyrosequencing on genomic DNA, resulting in three markers without the presence of null alleles. These three markers also showed transferability between populations geographically further apart

than those used in this current study, and between sub-species of *E. aurinia* [20]. Although null alleles can overestimate the genetic distance in significantly differentiated populations [39], the extent to which they affect the measurement of population differentiation in *E. aurinia* is unknown *a priori*. Furthermore, there is no guarantee that some population not sampled during marker development will not have null alleles. Aware of this problem in our markers, we validated their use for estimating genetic distances between populations with data from AFLP markers, with and without correction for the presence of null alleles [39]. Sinama *et al* [20] did not have AFLP markers as a comparison. However that study rejected four microsatellite markers because of deviations from HWE in only one out of the three populations tested. Such markers might actually prove usable if they could be validated. Approaches such as those used in both this study and by Sinama *et al* [20], may both be useful for efficient and cost effective development of robust markers for any endangered or non-model lepidopteran species, and should be the methods of choice for future studies. Marker validation through an alternative technique remains a plus.

## Supporting Information

**Figure S1 Map showing location of all 28 populations sampled in the present study.** Populations 18 to 28 are those also included in a previous study using AFLP markers [28]. (TIF)

**Table S1 Sites across the UK and Catalonia region of Europe from which *E. aurinia* was sampled.** (DOCX)

## Acknowledgments

The authors would like to thank Butterfly Conservation, the Cornwall Wildlife Trust, Natural England, Scottish Natural History and the Countryside Council for Wales, along with private landowners and volunteer collectors, for their assistance and enthusiasm for this project. We would like to thank Constanti Stefanescu for his expertise with southern populations of *E. aurinia* and Tanya Vo for laboratory assistance, and for sharing her considerable experience with microsatellite development and analysis.

## Author Contributions

Conceived and designed the experiments: MRS ASM DJH RHfC MCS. Performed the experiments: MRS YP PW. Analyzed the data: MRS ASM PW. Contributed reagents/materials/analysis tools: BW MCS RHfC ASM. Wrote the paper: MRS.

## References

1. Beaumont MA, Bruford MW (1999) Microsatellites in conservation genetics. In: Goldstein DB, Schlötterer C, editors. *Microsatellites: Evolution and Applications*. Oxford: Oxford University Press. pp.165–182.
2. Behura S (2006) Molecular marker systems in insects: current trends and future avenues. *Molecular Ecology* 15: 3087–3113.
3. Kandpal RP, Kandpal G, Weissman SM (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. *Proceedings of the National Academy of Sciences of the United States of America* 91: 88–92.
4. Glenn TC, Schable NA (2005) Isolating microsatellite DNA loci. *Methods in Enzymology* 395: 202–222.
5. Ostrander EA, Jong PM, Rine J, Duyk G (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proceedings of the National Academy of Sciences* 89: 3419–3423.
6. Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology* 11: 1–16.
7. Ji YJ, Zhang DX (2004) Characteristics of microsatellite DNA in lepidopteran genomes and implications for their isolation. *Acta Zoologica Sinica* 50: 608–614.
8. Zhang DX (2004) Lepidopteran microsatellite DNA: redundant but promising. *Trends in Ecology & Evolution* 19: 507–509.
9. Meglec E, Petician F, Danchin E, D'Acier AC, Rasplus JY, et al. (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.
10. Nève G, Meglec E (2000) Microsatellite frequencies in different taxa. *Trends in Ecology & Evolution* 15: 376–377.
11. 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.
12. Prasad MD, Muthulakshmi M, Madhu M, Archak S, Mita K, et al. (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.
13. Ji YJ, Zhang DX, Hewitt GM, Kang L, Li DM (2003) Polymorphic microsatellite loci for the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and some remarks on their isolation. *Molecular Ecology Notes* 3: 102–104.
14. Meglec E, Solignac M (1998) Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). *Hereditas* 128: 179–180.



15. Mikheyev AS, Vo T, Wee B, Singer MC, 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.
16. Petenian F, Meglec E, Genson G, Rasplus JY, Faure E (2005) Isolation and characterization of polymorphic microsatellites in *Parnassius apollo* and *Euphydryas aurinia* (Lepidoptera). *Molecular Ecology Notes* 5: 243–245.
17. Van't Hof AE, Zwaan BJ, Saccheri JJ, Daly D, Bot ANM, et al. (2005) Characterization of 28 microsatellite loci for the butterfly *Bicyclus anynana*. *Molecular Ecology Notes* 5: 169–172.
18. Vila M, Latasa T, Pino JJ, Verhulst G (2009) Characterization of ten polymorphic microsatellite markers for the endemic Chapman's ringlet, *Erebia palarica* (Lepidoptera: Nymphalidae). *European Journal of Entomology* 106: 485–490.
19. Keyghobadi N, Roland J, Strobeck C (1999) Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology* 8: 1481–1495.
20. Sinama M, Dubut V, Costedoat C, Gilles A, Junker M, et al. (2011) Challenges of microsatellite development in Lepidoptera: *Euphydryas aurinia* (Nymphalidae) as a case study. *European Journal of Entomology* 108: 261–266.
21. Kong L, Li Q (2008) Development of expressed sequence tag-derived microsatellite markers for the sea urchin *Hemicentrotus pulcherrimus*. *Molecular Ecology Resources* 8: 152–154.
22. Metzger C, Ferchaud A-L, Geiser C, Ursenbacher S (2011) New polymorphic microsatellite markers of the endangered meadow viper (*Vipera ursinii*) identified by 454 high-throughput sequencing: when innovation meets conservation. *Conservation Genetics Resources*.
23. Varshney RK, Chabane K, Hendre PS, Aggarwal RK, Graner A (2007) Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. *Plant Science* 173: 638–649.
24. Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Mackenzie K, et al. (2005) Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. *Molecular Ecology* 14: 1681–1695.
25. Asher J, Warren M, Fox R, Harding P, Jeffcoate G, et al. (2001) The millennium atlas of butterflies in Britain and Ireland. Oxford: Oxford University Press.
26. Fox R, Warren MS, Brereton TM, Roy DB, Robinson A (2010) A new Red List of British butterflies. *Insect Conservation and Diversity*.
27. Schtickzelle N, Chouff J, Goffart P, Fichet V, Bagnette M (2005) Metapopulation dynamics and conservation of the marsh fritillary butterfly: Population viability analysis and management options for a critically endangered species in Western Europe. *Biological Conservation* 126: 569–581.
28. Wee P-S (2004) Effects of Geographic Distance, Landscape Features and Host Association on Genetic Differentiation of Checkerspot Butterflies. PhD Thesis [PhD Thesis]. Austin: The University of Texas.
29. Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, et al. (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res* 32: e37.
30. Papanicolaou A, Stierli R, French-Constant RH, Heckel DG (2009) Next generation transcriptomes for next generation genomes using est2assembly. *BMC Bioinformatics* 10: 447.
31. Thurston MI, Field D (2005) Msatfinder: detection and characterisation of microsatellites. Distributed by the authors at <http://www.genomics.ceh.ac.uk/msatfinder/>. CEH Oxford, Mansfield Road, Oxford OX1 3SR. Accessed 24 May 2012.
32. Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for General Users and for Biologist Programmers. In: Krawetz S, Misener S, editors. *Bioinformatics Methods and Protocols: Methods in Molecular Biology: Humana Press, Totowa, NJ*. pp.365–386.
33. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.
34. Sigaard P, Pertoldi C, Madsen AB, Sogaard B, Loeschke V (2008) Patterns of genetic variation in isolated Danish populations of the endangered butterfly *Euphydryas aurinia*. *Biological Journal of the Linnean Society* 95: 677–687.
35. Porter K (1982) Basking behaviour in larvae of the butterfly *Euphydryas aurinia*. *Oikos* 38: 308–312.
36. Martinez-Torres D, Chandre F, Williamson MS, Darriet F, Berge JB, et al. (1998) Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector *Anopheles gambiae* s.s. *Insect Molecular Biology* 7: 179–184.
37. Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.
38. Dempster AP, Laird NM, Rubin DB (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society Series B (Methodological)* 39: 1–38.
39. Chapuis M-P, Estoup A (2007) Microsatellite Null Alleles and Estimation of Population Differentiation. *Molecular Biology and Evolution* 24: 621–631.
40. Beaumont M, Nichols R (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B* 263: 1619–1626.
41. Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* 9: 323.
42. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.
43. Rice WR (1989) Analyzing Tables of Statistical Tests. *Evolution* 43: 223–225.
44. Hanski I (1999) *Metapopulation Ecology*. Oxford: Oxford University Press.
45. Ehrlich PR (1965) The Population Biology of the Butterfly, *Euphydryas editha*. II. The Structure of the Jasper Ridge Colony. *Evolution* 19: 327–336.
46. Harrison S (1989) Long-Distance Dispersal and Colonization in the Bay Checkerspot Butterfly, *Euphydryas Editha Bayensis*. *Ecology* 70: 1236–1243.
47. Zimmermann K, Fric Z, Jiskra P, Kopeckova M, Vlasanek P, et al. (2011) Mark-recapture on large spatial scale reveals long distance dispersal in the Marsh Fritillary, *Euphydryas aurinia*. *Ecological Entomology* 36: 499–510.
48. Porter K (1981) The population dynamics of small colonies of the butterfly *Euphydryas aurinia*. PhD Thesis. Oxford: Oxford University.
49. Ford HD, Ford EB (1930) Fluctuation in numbers, and its influence on variation, in *Melitaea aurinia*, Rott. (Lepidoptera). *Transactions of the Royal Entomological Society of London* 78: 345–352.
50. Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, et al. (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature* 392: 491–494.
51. Louis EJ, Dempster ER (1987) An Exact Test for Hardy-Weinberg and Multiple Alleles. *Biometrics* 43: 805–811.
52. Guo SW, Thompson EA (1992) Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics* 48: 361–372.