High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera

Emese Meglécz, Frederic Petenian, Etienne Danchin, Armelle Coeur d'Acier, Jean-Yves Rasplus, Eric Faure

▶ To cite this version:

Emese Meglécz, Frederic Petenian, Etienne Danchin, Armelle Coeur d'Acier, Jean-Yves Rasplus, et al.. High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: Parnassius apollo and Euphydryas aurinia. Molecular Ecology, 2004, 13 (6), pp.1693-1700. 10.1111/j.1365-294X.2004.02163.x . hal-02678374

HAL Id: hal-02678374
https://hal.inrae.fr/hal-02678374

Submitted on 22 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Molecular Ecology (2004) **13**, 1693–1700 doi: 10.1111/j.1365-294X.2004.02163.x

High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*

EMESE MEGLECZ,* FREDERIC PETENIAN,* ETIENNE DANCHIN,† ARMELLE COEUR D'ACIER,‡ JEAN-YVES RASPLUS‡ and ERIC FAURE*

**Evolutionary Systematics Laboratory, CASE 5, UPRES Biodiversité 2202, Université de Provence, 3 place Victor Hugo, F-13331, Marseille Cedex 3, France, †Phylogenomics Laboratory, UPRES Biodiversité 2202, Université de Provence, 3 place Victor Hugo, F-13331, Marseille Cedex 3, France, ‡Centre de Biologie et Gestion des Populations (CBGP), Campus International de Baillarguet, CS 30 016, 34 988 Montferrier, France*

Abstract

Microsatellite flanking regions have been compared in two butterfly species. Several microsatellite flanking regions showed high similarity to one another among different microsatellites within a same species, but very few similarities were found between species. This can be the consequence of either duplication/multiplication events involving large regions containing microsatellites or of microsatellites imbedded in minisatellite regions. The multiplication of microsatellites might also be linked to mobile elements. Furthermore, crossing over between nonhomologous microsatellites can lead to the exchange of the flanking regions between microsatellites. The same phenomenon was observed in both studied butterfly species but not in *Aphis fabae* (Hemiptera), which was screened at the same time using the same protocol. These findings might explain, at least partially, why microsatellite isolation in Lepidoptera has been relatively unsuccessful so far.

Keywords: crossing over, duplication, flanking region, Lepidoptera, microsatellites, mobile elements

Received 10 December 2003; revision received 15 January 2004; accepted 29 January 2004

Introduction

Microsatellites are excellent markers to study population structure because they are highly variable, more likely to be neutral than other genetic markers and the results are reproducible (Jarne & Lagoda 1996). These advantages generally counterbalance the relatively long and expensive isolation process and the establishment of the appropriate polymerase chain reaction (PCR) conditions. In some cases, however, obtaining well-behaved microsatellites requires considerable time and effort and some of the microsatellites might still have null alleles or single primer pair amplifies more than one locus (Ohnishi *et al*. 1998; Sunnucks & Wilson 1999; Wu *et al*. 1999; Wilson *et al*. 2002; Li *et al*. 2003). Because complete or nearly complete failures of microsatellite isolations are hardly publishable we do not have exact details on the proportion of the trials that do not yield sufficient results. Therefore we shall restrict our discussion to Lepidoptera, as we have more detailed information in this order. In this group, difficulties arise systematically during the isolation and characterization of microsatellites (Meglécz & Solignac 1998), leading to relatively few well-behaving loci compared to other taxa (Nève & Meglécz 2000). The cause of this phenomenon has been questioned (Sunnucks 2000), but so far only guesses have been suggested orally by numerous biologists. Lepidopterists encounter most of the problems during the design of appropriate primers and in setting up PCR conditions (I. Saccheri, N. Keygobadi, A. Cassel, N. Anthony, pers.comm.). In almost all the studies reporting microsatellite markers in Lepidoptera, a significant departure from Hardy–Weinberg equilibrium was observed at many of the loci due to the deficit of heterozygotes (Palo *et al*. 1995; Bogdanovicz *et al*. 1997; Meglécz & Solignac 1998; Keyghobadi *et al*. 1999, 2002; Harper *et al*. 2000; Anthony *et al*. 2001; Caldas *et al*. 2002;

Correspondence: Emese Meglécz, Unité Evolution, Génome, Environnement, CASE 36, Université de Provence, 3 place Victor Hugo, F-13331 cedex 3, Marseille, France. Fax: + 33 4 91 10 63 48; E-mail: meglecz@up.univ-mrs.fr

© 2004 Blackwell Publishing Ltd

Cassel 2002; Flanagan *et al*. 2002; Amsellem *et al*. 2003). These results suggest the presence of null alleles. As a consequence, even the loci with apparently clearly interpretable and reproducible patterns might still yield biased or erroneous results. The presence of null alleles and varying amplification intensities between individuals suggest that the flanking regions of microsatellites in this order are variable.

Another frequently observed problem is the amplification of more than two different bands with a single pair of primers (I. Saccheri, A. Estoup, D. Bourguet, pers. comm.; Emese Meglécz, unpubl. results). Two possible causes of this pattern are (i) the duplication or multiplication of microsatellite-containing regions or (ii) that microsatellites lie within a minisatellite repeat unit and have microsatellite length variations between minisatellite repeat units. In this study we have compared microsatellite-containing sequences for two butterfly species, *Parnassius apollo* (Papilionidae) and *Euphydryas aurinia* (Nymphalidae), in order to detect similarities among flanking sequences of different microsatellite loci. We have also analysed microsatellite flanking sequences of *Aphis fabae* (Aphididae) as a control.

Materials and methods

Screening and sequencing

Genomic DNA was isolated from individuals (six, eight and 400 for *P. apollo*, *E. aurina* and *A. fabae*, respectively) kept at −80 °C following standard procedures (Sambrook *et al*. 1989). The screening was carried out according to Streiff *et al*. (2002) and Dalecky *et al*. (2002). Total genomic DNA (3–4 μg) was digested to completion with *Rsa*I restriction enzyme. The extent of digestions was checked on agarose gel. The 300–900 base pairs (bp) fraction of the digested DNA was selected on agarose gel, purified and ligated to *Rsa* linkers.

The enrichment procedure followed the protocol from Kijas *et al*. (1994) based on streptavidin-coated magnetic particles (Magnesphere, Promega), with slight modifications. 5′-biotinylated $(CT)_{10}$, $(GT)_{10}$, $(AAG)_{10}$, $(AAT)_{10}$, $(ATCT)_{10}$ and $(TGTA)_{10}$ oligonucleotides were used as probes. The enriched single-stranded DNA was amplified using one of the *Rsa* linkers as primer to obtain a more enriched product and to recover double-stranded DNA. PCR products were purified and ligated into pGEM-T Easy Vector (Promega), and the plasmid transformed into *Escherichia coli* supercompetent cells (XL1 blue, Stratagene). Recombinant clones were transferred to Hybond-N nylon membranes (Amersham). Colonies were hybridized at 50 °C with the mixture of oligonucleotide probes $(CT)_{10}$, $(TG)_{10}$, $(AGA)_5AG$, $(ATA)_5AT$, $(CTAT)_6CT$ and $(TGTA)_6TG$ labelled with the DIG oligonucleotide tailing kit (Boehringer). Positive clones were identified visually and plasmid DNA was extracted from them. Positive clones having the strongest signal were sequenced on one strand using an automated sequencer (Genome Express SA). Sequences were deposited in GenBank. Accession nos are: AY491782–AY491857 for *E. aurinia* and AY491858–AY491940 for *P. apollo*. The number of the clones for each motif separately is shown in Table 2. The whole screening procedure was performed at the same time with the same method for all three species.

Sequence analyses

We performed an all-against-all comparison of the microsatellite-containing sequences for all sequences together, using the BLAST (Altschul *et al*. 1990) program with BLASTN option. We used standard parameters for expected value settings and filtered sequences for low complexity using the DUST filter (Hancock & Armstrong 1994) to avoid wrong alignments due to composition biases in sequences (i.e. repetitive DNA regions have been systematically ignored). One bank was formatted with sequences from all three species and then each sequence was compared against all sequences of the bank. We selected only regions that showed more than 85% identity in the pairwise comparison, and excluded fragments shorter than 40 bp. Note, however, that regions showing similarity between different microsatellites were usually shorter than the whole sequences, thus identities between pairs of regions are higher than identities between pairs of the corresponding whole sequences. Sequences that showed more than 98% of identity along the whole length, including the repeat array, were considered as duplicates. Their occurrence can be attributed to the enrichment protocol. They can be either different alleles of the same locus, or the same alleles including a few mistakes in sequencing.

We also scanned all the sequences of the data set against databases of repetitive sequences in order to detect the presence of mobile elements using the program REPEATMASKER2 (A.F.A. Smit & P. Green, unpubl. data, http://repeatmasker.genome.washington.edu/RM/webrepeatmaskerhelp.html). We performed the analysis with two different models, the *Drosophila* parameters set and the Primates set. We used default sensitivity parameters for both analyses.

Results

Based on similarities between the flanking regions of different microsatellite loci, we have identified several groups of (related) sequences. In *A. fabae*, only a *c.* 60 bp fragment was shared by four different flanking regions (AF4, AF70, AF86, AF97). Very few similarities were found between species: 80 bp between sequence EA85 of *E. aurinia* and PA27 of *P. apollo*, and *c.* 60 bp between PA85 of *P. apollo* and AF4, AF70, AF86 and AF97 of *A. fabae*. On the contrary, we found many highly similar regions between the flanking sequences of the different microsatellites within each of the butterfly

Table 1 Summary of the groups of related sequences based on regions showing higher than 85% identity (in pairwise comparison)

	P. apollo	*E. aurinia*	*A. fabae*
Sequences with similarities in both sides of the MS	**GR 12:** PA31-PA45-PA82-PA95† **GR 13:** PA21-PA71 **GR 14A:** PA12-PA35-PA46-PA55-PA59‡	**GR 1:** EA2-EA3-EA47-EA58-EA59-EA80 **GR 2:** EA12-EA23 **GR 5:** EA5-EA35-EA57-EA76-EA78	
Sequences with similarities in one side of the MS	**GR 15:** PA8-PA11-PA20-PA22-PA32-PA44-PA56-PA68-PA69-PA78-PA90-PA93-PA102-PA103§	**GR 3:** EA9-EA71	**GR 18:** AF4-AF70-AF86-AF97¶
	GR 16B: PA15-PA37-PA54-PA60‡ **GR 16C:** PA16-PA19-PA25-PA65‡	**GR 4:** EA10-EA11-EA93*	
Sequences with similarities in the only side of the MS	**GR 14B:** PA12-PA70-PA94‡ **GR 16A:** PA5-PA17-PA19-PA27-PA41-PA50-PA60-PA63-PA87-PA88‡ **GR 17:** PA2-PA7-PA26-PA48-PA76-PA97	**GR 6:** EA7-EA20 **GR 7:** EA18-EA39	
		GR 8: EA31-EA62 **GR 9:** EA56-EA70 **GR 10:** EA65-EA69 **GR 11:** EA87-EA89-EA90	

*The flanking region is similar in both sides for EA10 and EA11.
†In Group 12 two or three regions are found in each sequences that are similar not only to the corresponding regions of the other sequences but also to the other regions of the same sequence (see also Fig. 1).
‡Groups 16A, 16B, 16C and Groups 14A, 14B are subgroups of two larger groups (Groups 14 and 16; see Fig. 1).
§The flanking region is similar in both sides for PA11 and PA32.
¶Sequences AF4-AF70-AF86 have CA microsatellite motif, AF97 has CATA.

species (Table 1; see details in http://www.up.univ-mrs.fr/evol/Supplement/Meglecz-MolEcol/). Many of the similar regions were long (more than 100 bp), and the identity was often higher than 95% among them. In most of the cases, however, we can exclude that we have isolated different alleles of the same locus, because generally similar regions did not stretch all along the whole corresponding sequences. In all the cases where the sequences matched in their whole length (groups 2, 7, 11, 12, 13, 17), the similarity was lower than 95% between them in a pairwise comparison.

In the butterflies, the microsatellite motifs are the same within groups (CA for groups 1, 2, 5, 6, 7, 9, 10, 16, 17, CATA for groups 4, 8, 11, 12, 13, 14, 15) but repeat numbers can vary. The only group in *A. fabae* is unusual, as sequences AF4, AF70 and AF86 have CA motifs but AF 97 is a CATA microsatellite.

Table 1 shows that concerning the position of similar regions in the sequences, all three possible scenarios have been found: (i) similarities between different flanking sequences but only on one side of the microsatellites; (ii) similarities between different flanking sequences at both side of the microsatellites; and (iii) similarities between different flanking sequences at the only side available (the microsatellite is situated at the end of the sequence). Sequences of the groups 16A, 16B and 16C; 14A and 14B can be clustered together into two larger groups (Fig. 1). The most peculiar sequences are found in group 12 of *P. apollo*. Here, we found a *c.* 90 bp fragment which is repeated twice or three times in the same sequence (with slight variations within repetitions) and it was also detected in four different sequences twice or three times in each (Fig. 1). It is also interesting to note that the limit between the microsatellite and the flanking region is frequently difficult to define because pure microsatellite motifs are often surrounded by stretches of cryptically simple sequences (i.e. microsatellites that accumulated many mutations; see Hancock 1999).

We also found 20, 13 and 34 sequences that were considered duplicates in *P. apollo*, *E. aurinia* and *A. fabae*, respectively (Table 2). These were eliminated from the analyses.

The distribution of the number of independent and related sequences (grouped together on the basis of similarities in their flanking regions) is shown in Table 2 for each microsatellite motif. In *P. apollo*, the number of related CATA sequences is unexpectedly high compared to the number of related CA sequences (Fisher's exact test $P < 0.005$). For the two other species, however, the P-value was nonsignificant ($P < 0.39$, $P < 0.17$ for *E. aurinia* and *A. fabae*, respectively).

Scanning our Lepidopteran sequences against a *Drosophila* and a primate repetitive sequence bank, gave positive results in three cases. Sequence PA69 (*P. apollo*) showed similarity with a retrotransposable element of *Drosophila* (*Gypsy*5-I, substitutions: 24.6%), and sequences PA5 and PA13 (*P. apollo*) to a human endogenous retrovirus (Prima4-int, substitutions: 15.9% and 9.4%, respectively).

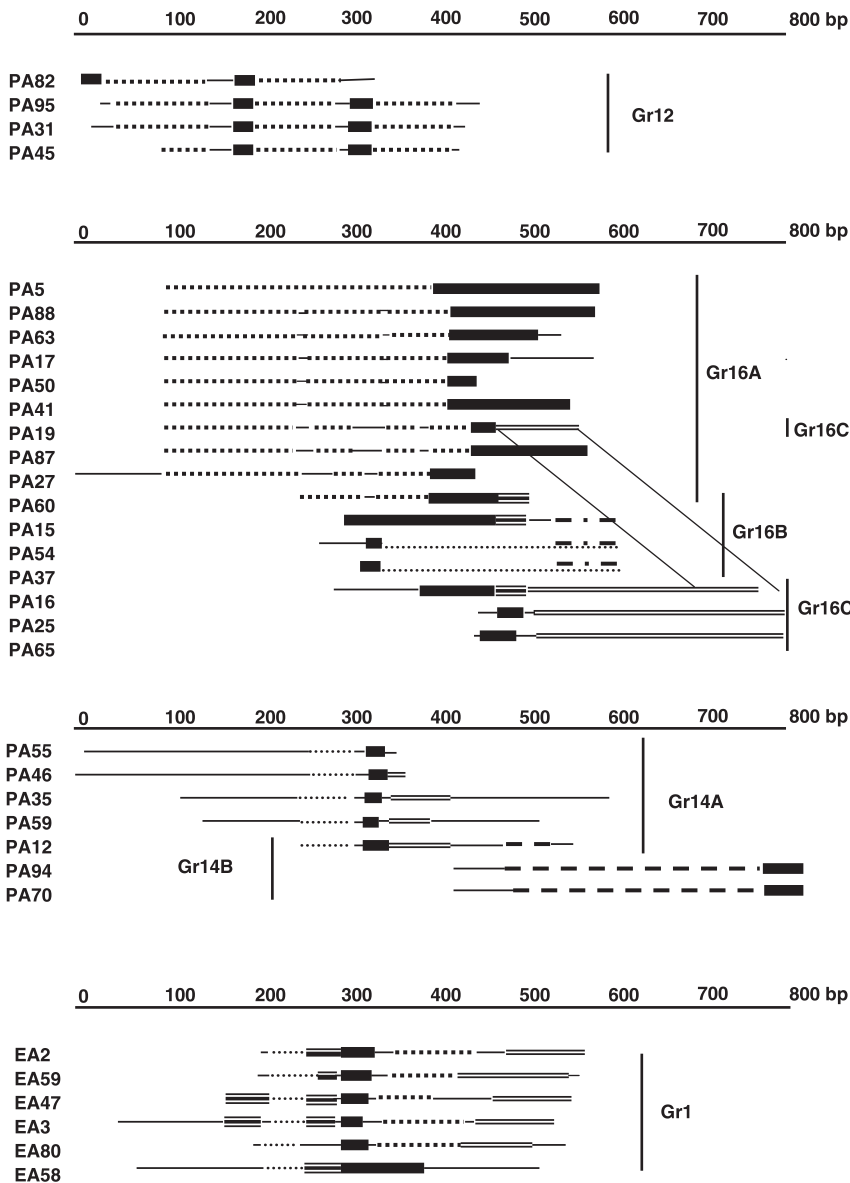

Fig. 1 Schematic representation of sequences of *P. apollo* and *E. aurinia* forming large groups of similarities. Wide black lines: microsatellites (all are CATA repeats in groups 12 and 14, CA in groups 1 and 16); simple lines: unique sequences (no apparent similarity to any of the other regions). All other types of lines are regions that are similar to other regions of the same group. Note that in group 12, regions represented by dashed lines are not only similar to corresponding regions in other sequences, but also to the other regions of the same sequence. In Group 16C the region represented by a double line in sequence PA19 is similiar to the region at *ca*. 650–750 bp of the other sequences of the same group.

Discussion

While there have been many studies on the mutation pattern of the microsatellites (Weber & Wong 1993; Michalakis & Veuille 1996; Chakraborty *et al*. 1997; Schlötterer 1998; Schug *et al*. 1998; Estoup & Cornuet 1999; Hancock 1999; Pupko & Graur 1999), little is known of their origin. It has been suggested that a primordial microsatellite of a few repeats is necessary for the expansion of the tandem repeats (Stallings *et al*. 1991), mainly through DNA replication slippage (Schlötterer & Tautz 1992; Schlötterer 2000). This indicates that the increase of the number of new microsatellites is slow. The presence of several similar but nonidentical flanking regions in our study suggests that there have been several duplication events involving both the microsatellites and their flanking regions, implying that duplication events might be an important factor for increasing the number of microsatellites per genome. Alternatively, it is possible that at least some of the microsatellites lie within a minisatellite repeat unit. This might be the case in sequences in group 12 of *P. apollo* (Fig. 1), because here a *c*. 90 bp DNA fragment is repeated twice or three times within the same sequences. This fragment, however, is not identical between repetitions (within the same sequence), but shows only 85–97% identity in pairwise comparisons. Furthermore, short unique DNA stretches (i.e. no apparent similarities to other parts of the same sequence or to the other sequences) can also be found between the putative minisatellite repeat motifs. For the other groups of sequences we did not find any evidence between association of microsatellite and minisatellite motifs.

Table 2 Distribution of microsatellites among different repetitive motifs. *Scr.*: Number of clones screened. *Pos.*: Number of positive clones detected visually. *Seq.*: Number of clones sequenced successfully. *Indep*: Number of sequences that did not show similarities in their flanking regions to others. *Rel.*: Number of sequences that were grouped together based on similarities in their flanking regions. *Red.*: Number of sequences that were nearly identical (identity greater than 98% along the whole sequence) to other sequences. *CSS*: Number of sequences that had long stretches of cryptically simple sequences, making alignment impossible. *False*: Number of sequences that did not contain microsatellite motif

	Scr.	Pos.	Seq.	Indep.	Rel.	Red.	CSS	False
P. apollo								
CA	663	65	55	18	22	7	7	1
ACAT	1105	80	47	4	27	13	2	1
TC	1326	3	2	1	0	0	1	0
AAG	1105	0	0	0	0	0	0	0
AAT	1547	0	0	0	0	0	0	0
ATCT	663	0	0	0	0	0	0	0
Total	6409	148	104	23	49	20	10	2
E. aurinia								
CA	1105	353	62	36	21	4	0	1
ACAT	884	87	20	8	8	6	0	2
ATCT	884	4	3	2	0	0	0	1
GAA	1105	51	8	3	2	3	0	0
AAT	0	0	0	0	0	0	0	0
CT	1105	0	0	0	0	0	0	0
Total	5083	495	93	45	31	13	0	4
A. fabae								
CA	454	207	86	60	3	24	0	1
ACAT	454	47	16	2	1	10	0	3
GAA	908	18	0	0	0	0	0	0
AAT	454	0	0	0	0	0	0	0
CT	1362	0	0	0	0	0	0	0
Total	3632	272	102	62	4	34	0	4

In some cases, the sequences matched in their whole length showing 85–95% similarities (groups 2, 7, 11, 12, 13, 17). Those sequences could be different alleles of the same locus or paralog sequences of multiplicated loci. The extent of similarities among flanking regions of alternative alleles of a microsatellite or among paralog loci is little studied. Microsatellite flanking regions are studied mainly by comparing homologue loci in different species (Choumane *et al*. 2000; Di Gaspero *et al*. 2000; Noor *et al*. 2001). The variability of flanking region depends strongly on the phylogenetic relation between the species. Few studies have information on the extent of similarities between microsatellite flanking regions of alleles of the same locus (Blankenship *et al*. 2002). One study on four species of mice provides details of both the flanking sequence variability among alleles and among homologue loci of different species (Makova *et al*. 2000). Here, differences between flanking sequences among alleles range from 0% to 2%, and among homologue sequences between 1.25 and 9.5%. To our knowledge, only one study on *Cyprinus caprio* (David *et al*. 2003) compares flanking sequences of alleles vs. paralogs. Here differences between alleles range from 0% to 4.7% and differences between paralogs from 2% to 12%. It is impossible to establish a threshold of similarity that would clearly separate these two scenarios. Based on the above results we have treated sequences with more than 5% difference as paralogs, keeping in mind that in a few cases it is possible that in reality we have sequenced very different alleles of the same loci. Those cases, however, are marginal compared to the number of paralog sequences.

PCR-mediated recombination can occur in the presence of two or more highly similar but nonidentical templates if prematurely terminated products re-anneal with nonidentical templates and then they are extended in the next cycle (Saiki *et al*. 1988; Myerhans *et al*. 1990). We have PCR amplified the enriched DNA. At this step, if different alleles or highly similar paralog sequences are present, artificial chimera formation is possible. This possibility is more probable if sequences are similar to each other in their whole length and thus can be the case essentially in the sequences discussed above (groups 2, 7, 11, 12, 13, 17). Those groups should thus be intrepreted with caution, as some of the sequences can be either chimeras or alleles and not paralogs. Judo *et al*. (1998) have shown that large numbers of cycles and short elongation times increase the probability of PCR recombination. By using a 2-min elongation time and 27 cycles we attempted to minimize the probability of this artefact.

An interesting result is the frequent asymmetry of flanking regions (i.e. similarities between the flanking region of one side but not the other). Our hypothesis is that crossing over between nonhomologous microsatellites might occur, leading to the exchange between the flanking regions of different microsatellites. Alternatively, PCR-mediated recombination during the enrichment procedure could also generate microsatellites matching on one side but not on the other. However, the probability of this event is slight because it can happen only if PCR elongation is interrupted at the microsatellite motif. In this case, the micosatellite at the end of the sequence can anneal with another nonrelated microsatellite and be elongated, forming a chimera with two nonrelated flanking sequences on the two sides of the microsatellite. In this case, we would expect a nearly complete identity (differences are due to rare mistakes during PCR) between the cross-over point (within the microsatellite motif) and the end of the sequence. This is rarely the case in our sequences. Only the following pairs or groups of sequences showed higher than 98% similarities in the above-mentioned regions (along the flanking sequence on one side of the microsatellite): EA35-EA76-EA78 of group 5, EA18-EA39 of group 7 and PA20-PA103 of group 15.

In recent years, numerous microsatellites have been found associated physically with various types of mobile elements

(Beckmann & Weber 1992; Nadir *et al*. 1996; Crouau-Roy & Clisson 2000; Souames *et al*. 2003). Mobile elements could play a role in the genesis of microsatellites in Diptera (Wilder & Hollocher 2001) and LINE and *Alu* insertions in the human genome (Kulski *et al*. 1997; Ovchinnikov *et al*. 2001). According to Nadir *et al*. (1996), the juxtaposition of microsatellites and retrotransposons suggests their coevolution, as the microsatellites could be generated by a 3′-extension of retrotranscripts, similar to mRNA polyadenylylation, and that they serve in turn as 'retroposition navigators', directing the retrotransposons via homology-driven integration into defined sites.

Because mutation rates are high in mobile elements and because they have been little studied in Lepidoptera, it is difficult to detect their presence in the relatively short flanking sequences we had in this study. In spite of this fact, scanning our Lepidopteran sequences against a *Drosophila* and a primate repetitive sequence bank still gave positive results in three cases. Microsatellite flanking regions present sequence similarities with a primate endogenous retrovirus and a putative invertebrate retrovirus (*Gypsy*). It is interesting to note that in some Lepidoptera genomes the *TED* retrotransposon has been found (Ozers & Friesen 1996), which belongs to the *Gypsy* group.

We have screened all three species for five to six different microsatellite motifs (Table 2). Strikingly, we found most of the positive clones with CA and TACA probes and very few with the other motifs. It is possible that the low stringency conditions used during enrichment selected repetitive DNA in general or AT-rich regions, and not only microsatellites. In this case, however, we would expect many positive clones for AAT and a high proportion of false positives. On the contrary, we did not find any positive clones for AAT and the proportion of false positives was low (Table 2). Furthermore, in a previous study on *P. mnemosyne* the proportion of CA and CATA clones was also very high (97%) compared to the other screened motifs, in spite of the fact that we used a nonenriched library for this species (Meglécz & Solignac 1998). It is also observed frequently in many species that some microsatellite motifs are more common than others. In human and mice, Beckmann & Weber (1992) showed by database analyses of tandem repeats in genomic sequences that CA microsatellites are the most abundant among dinucleotide repeats. CA repeats are the most common in *D. melanogaster* (Schug *et al*. 1998) and in a wasp, *Venturia canescence* (Butcher *et al*. 2000) as well. Therefore we think that the observed high frequency of CA and CATA microsatellites in the investigated species gives a fair approximation of their frequency in the genome. As similarities between flanking regions were observed for both predominant microsatellite motifs (CA, CATA) it is likely that problematic microsatellites are characteristic of the genome as a whole and not only to a subset of loci.

The fact that very few shared regions were found between sequences of *A. fabae* in spite of the identical screening procedure for all three species strongly reduces the possibility that the described phenomena observed in butterflies would be the results of laboratory artefacts. Frequent locus duplication, association between microsatellites and minisatellites or mobile elements suggest that in the two investigated Lepidoptera many of the microsatellites are situated in genetically disturbed region of the genome.

The above findings might have several technical consequences on the isolation of usable microsatellites for population genetics studies. The fact that several similar flanking regions are present in the genome of the two studied butterfly species can make clear PCR amplification difficult. If primers are placed into sufficiently conserved regions, the use of a single pair of primers leads to the amplification of more than one locus at the same time, making exact genotyping difficult or impossible. This phenomenon has already been observed (Nusha Keyghobadi pers. comm.; E. Meglécz & F. Péténian unpubl. results).

It is interesting to note that although the overall pattern in the two investigated butterfly species is similar, the extent of problems is different in the two cases. While in *P. apollo* many large and complicated similarity groups have been identified, in *E. aurinia* the extent of similarities between the sequences seems to be smaller. The proportion of independent sequences (i.e. no apparent similarities to other sequences of the same bank) is also quite different in the two species (Table 2). It would thus be unwise to generalize at this stage of the work. However, we know that the above-described phenomena have been found in other Lepidoptera as well (J.-F. Silvain, N. Faure, A. van't Hof, A. Estoup, D. Bourguet, pers. comm.). This fact and the frequent problems of finding suitable microsatellites in this order suggest that the above-mentioned problems are probably widespread in this group and possibly also in other taxa (i.e. Orthoptera). In order to support this hypothesis, all sequences coming from the screening process should be analysed systematically to search for similarities in the flanking regions rather than checking the presence of duplicate copies of the same locus. Such an analysis could also be a valuable aid to avoid choosing microsatellites that have similar flanking regions.

Acknowledgements

We thank Gwenaelle Genson for valuable technical help during the screening process, Henri Descimon and Gabriel Nève for constructive comments on this manuscript and Ilik Saccheri, Nusha Keygobadi, Anna Cassel, Nicola Anthony, Arnaud Estoup, Dennis Bourguet, Jean-François Silvain, Nathalie Faure and Arjen van't Hof for valuable information on their unpublished data. Robert Butcher, Paul Sunnucks and two anonymous referees provided valuable help in improving the manuscript during the peer-reviewing process.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Amsellem L, Risterucci AM, Benrey B (2003) Isolation and characterization of polymorphic microsatellite loci in *Lobesia botrana* Den. & Schiff. (Lepidoptera: Tortricidae). *Molecular Ecology Notes*, **3**, 117–119.

Anthony N, Gelembiuk G, Raterman G, Nice C, Ffrench-Constant R (2001) Isolation and characterization of microsatellite markers from the endangered Karner blue butterfly *Lycaeides melissa samuelis* (Lepidoptera). *Hereditas*, **134**, 271–273.

Beckmann JS, Weber JL (1992) Survey of human and rat microsatellites. *Genomics*, **12**, 627–631.

Blankenship SM, May B, Hedgecock D (2002) Evolution of a perfect simple sequence repeat locus in the context of its flanking sequence. *Molecular Biology and Evolution*, **19**, 1943–1951.

Bogdanovicz SM, Mastro VS, Prasher DC, Harrison RG (1997) Microsatellite variation among Asian and North American gypsy moths (Lepidoptera: Lymantriidae). *Annals of the Entomological Society of America*, **90**, 768–775.

Butcher RDJ, Hubbard SF, Whitfield WGF (2000) Microsatellite frequency and size variation in the parthenogenetic parasitic wasp *Venturia canescence* (Gravenhorst) (Hymenoptera: Ichneumonidae). *Insect Molecular Biology*, **9**, 375–384.

Caldas A, Hawthorne DJ, Barbosa P (2002) Isolation and characterisation of microsatellite markers from *Zale galbanata* (Lepidoptera: Noctuidae) and amplification in other members of the genus. *Molecular Ecology Notes*, **2**, 296–297.

Cassel A (2002) Characterization of microsatellite loci in *Coenonympha hero* (Lepidoptera: Nymphalidae). *Ecology Notes*, **2**, 566–568.

Chakraborty R, Kimmel M, Stivens ND, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri- and tetranucleotide microsatellite loci. *Proceedings of National Academy of Sciences USA*, **94**, 1041–1046.

Choumane W, Winter P, Weigand F, Kahl G (2000) Conservation and variability of sequence-tagged microsatellite sites (STMSs) from chickpea (*Cicer aerietinum* L.) within the genus *Cicer*. *Theoretical and Applied Genetics*, **101**, 269–278.

Crouau-Roy B, Clisson I (2000) Evolution of an *Alu* DNA element of the type Sx in the lineage of primates and the origin of an associated tetranucleotide microsatellite. *Genome*, **43**, 642–648.

Dalecky A, Debout G, Mondor G, Rasplus J-Y, Estoup A (2002) PCR primers for polymorphic microsatellite loci in the facultatively polygynous plant-ant *Petalomyrex phylax* (Formicidae). *Molecular Ecology Notes*, **2**, 404–407.

David L, Blum S, Feldman MW, Lavi U, Hillel J (2003) Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. *Molecular Biology and Evolution*, **20**, 1425–1434.

Di Gaspero G, Peterlunger E, Testolin R, Edwards KJ, Cipriani G (2000) Conservation of microsatellite loci within the genus *Vitis*. *Theoretical and Applied Genetics*, **101**, 301–308.

Estoup A, Cornuet J-M (1999) Microsatellite evolution: inferences from population data. In: *Microsatellites: Evolution and Applications* (eds Goldstein DB, Schlötterer C), pp. 49–65. Oxford University Press, Oxford.

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

Hancock JM (1999) Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: *Microsatellites: Evolution and Applications* (eds Goldstein DB, Schlötterer C), pp. 1–9. Oxford University Press.

Hancock JM, Armstrong JS (1994) SIMPLE34: an improved and enhanced implementation for VAX and Sun computers of the SIMPLE algorithm for analysis of clustered repetitive motifs in nucleotide sequences. *Computer Applications in the Biosciences*, **10**, 67–70.

Harper GL, Piyapattanakorn S, Goulson D, Maclean N (2000) Isolation of microsatellite markers from the adonis blue butterfly (*Lysandra bellargus*). *Molecuar Ecology*, **9**, 1948–1949.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, **11**, 424–429.

Judo MSB, Wedel AB, Wilson C (1998) Stimulation and suppression of PCR-mediated recombination. *Nucleic Acid Research*, **26**, 1819–1825.

Keyghobadi N, Roland J, Strobeck CE (1999) Influence of landscape on the population of the alpine butterfly *Parnassius smitheus* (Papilionidae). *Molecular Ecology*, **8**, 1481–1495.

Keyghobadi N, Roland J, Strobeck CE (2002) Isolation of novel microsatellite loci in the Rocky Mountain apollo butterfly, *Parnassius smitheus*. *Hereditas*, **136**, 247–250.

Kijas JMH, Fowler JCS, Garbett CA, Thomas MR (1994) Enrichment of microsatellites from citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particules. *Biotechniques*, **16**, 657–662.

Kulski JK, Gaudieri S, Bellgard M *et al.* (1997) The evolution of MHC diversity by segmental duplication and transposition of retroelements. *Journal of Molecular Evolution*, **45**, 599–609.

Li G, Hubert S, Bucklin K, Ribes V, Hedgecock D (2003) Characterization of 79 microsatellite DNA markers in the Pacific oyster *Crassostrea gigas*. *Molecular Ecology Notes*, **3**, 228–232.

Makova KD, Nekrutenko A, Baker RJ (2000) Evolution of microsatellite alleles in four species of mice (genus *Apodemus*). *Journal of Molecular Evolution*, **51**, 166–172.

Meglécz E, Solignac M (1998) Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). *Hereditas*, **128**, 179–180.

Michalakis Y, Veuille M (1996) Length variation of CAG/CAA trinucleotide repeats in natural population of *Drosophila melanogaster* and its relation to recombination rate. *Genetics*, **143**, 1713–1725.

Myerhans A, Vartanian JP, Wain-Hobson S (1990) DNA recombination during PCR. *Nucleic Acid Research*, **18**, 1687–1691.

Nadir E, Margalit H, Gallily T, Ben-Sasson SA (1996) Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. *Proceedings of the National Academy of Sciences USA*, **93**, 6470–6475.

Nève G, Meglécz E (2000) Microsatellite frequencies in different taxa. *Trends in Ecology and Evolution*, **15**, 376–377.

Noor MAF, Kliman RM, Machado CA (2001) Evolutionary history of microsatellites in the obscura group of *Drosophila*. *Molecular Biology and Evolution*, **18**, 551–556.

Ohnishi N, Ishibashi Y, Saitoh T, Abe S, Yoshida MC (1998) Polymorphic microsatellite DNA markers in the Japanese wood mouse *Apodemus argenteus*. *Molecular Ecology*, **7**, 1431–1432.

Ovchinnikov I, Troxel AB, Swergold GD (2001) Genomic characterization of recent human LINE-1 insertions: evidence supporting random insertion. *Genome Research*, **11**, 2050–2058.

Ozers MS, Friesen PD (1996) The Env-like open reading frame of the baculovirus-integrated retrotransposon TED encodes a retrovirus-like envelope protein. *Virology*, **226**, 252–259.

Palo J, Varvio SL, Hanski I, Väinölä R (1995) Developing microsatellite markers for insect population structure: complex variation in a checkerspot butterfly. *Hereditas*, **123**, 295–300.

Pupko T, Graur D (1999) Evolution of microsatellites in yeast *Saccharomyces cerevisiae*: role of length and number of repeated units. *Journal of Molecular Evolution*, **48**, 313–316.

Saiki RK, Gelfand DH, Stoffel S *et al*. (1988) Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase. *Science*, **239**, 487–491.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.

Schlötterer C (1998) Are microsatellites really simple sequences? *Current Biology*, **8**, R132–R134.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma*, **109**, 365–371.

Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acid Research*, **20**, 211–215.

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TF, Aquadro CF (1998) The mutation rates of di- tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **15**, 1751–1760.

Souames S, Bonnivard E, Bazin C, Higuet D (2003) High mutation rate of TPE repeats: a microsatellite in the putative transposase of the hobo element in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **20**, 1826–1832.

Stallings RL, Ford AF, Nelson D, Torney DC, Hildebrand CE, Moysis B (1991) Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics*, **10**, 807–815.

Streiff R, Mondor-Genson G, Audiot P, Rasplus J-Y (2002) Microsatellite DNA markers for a grasshopper: *Prionotropis hystrix rhodanica* (Orthoptera, Pamphagidae). *Molecular Ecology Notes*, **2**, 268–270.

Sunnucks P (2000) Reply from P. Sunnucks. *Trends in Ecology and Evolution*, **15**, 377.

Sunnucks P, Wilson ACC (1999) Microsatellite markers for the onychophoran *Euperipatoides rowelli*. *Molecular Ecology*, **8**, 899–900.

Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics*, **2**, 1123–1128.

Wilder J, Hollocher H (2001) Mobile elements and the genesis of microsatellites in dipterans. *Molecular Biology and Evolution*, **18**, 384–392.

Wilson ACC, Sunnucks P, Barker JSF (2002) Isolation and characterization of 20 polymorphic microsatellite loci for *Scaptodrosophila hibisci*. *Molecular Ecology Notes*, **2**, 242–244.

Wu L, Kaufman L, Fuerst PA (1999) Isolation of microsatellite markers in *Astatoreochromis alluaudi* and their cross-species amplifications in other African cichlids. *Molecular Ecology*, **8**, 895–897.