Towards understanding the architecture of the Bicyclus anynana genome

Hof, Arjèn Emiel van 't

Chapter 3

Evolutionary dynamics of multi-locus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera.[1]

Arjèn E. van't Hof
Paul M. Brakefield
Ilik J. Saccheri
Bas J. Zwaan

ABSTRACT

The sequences flanking microsatellites isolated from the butterfly *Bicyclus anynana* display high levels of similarity amongst different loci. We examined sequence data for evidence of the two mechanisms most likely to generate these similarities, namely recombination mediated events, such as unequal crossing over or gene conversion, and through transposition of Mobile Elements (ME). Many sequences contained tandemly arranged microsatellites, lending support to recombination as the multiplication mechanism. There is, however, also support for ME-mediated multiplication of microsatellites and their flanking sequences. Homology with a known Lepidopteran ME was found in *B. anynana* microsatellite regions, and polymorphic microsatellite markers with partial similarities in their flanking sequences were passed on to the next generation independently, indicating that they are not linked. Therefore, the rise of these similarities appears to be mediated through both processes, either as an interaction between the two, or by each being responsible for part of the observations. A large proportion of microsatellites embedded in repetitive DNA is representative for most studied butterflies and moths, and a BLAST survey of the *B. anynana* sequences revealed four short microsatellite-associated sequences that were present in many species of Lepidoptera. The similarities usually start to deviate beyond these sequences, which suggests that they define the extremes of a repeated unit. Further study of these conserved sequences may help to understand the mechanism underlying the multiplication events, and answer the question of why these redundancies are predominantly found in this insect group.

[1] This chapter was published in Heredity 2007 volume **98**, pp 320-328. Appendix 3.11 contains post-publication information which convincingly demonstrates that one of the multi-locus microsatellites has a viral origin that was presumably horizontally transferred from parasitic wasps.

INTRODUCTION

Microsatellites, consisting of tandemly repeated units of 2-6 bp, have proved to be one of the most versatile molecular markers available due to their high level of repeat number variation and widespread distribution in eukaryotic genomes. The classical model for their evolution proposes that the initial repeated motif seed arises through random base substitution, followed by stepwise mutation through replication slippage (reviewed in (ELLEGREN 2004). However, the expanding microsatellite database, particularly from Lepidoptera, suggests that other mechanisms play an important role in the genesis of microsatellites.

In Lepidoptera, microsatellites and their flanking sequences often possess features which are uncommon in most other taxa. These features have impeded development of microsatellite markers, as illustrated by the relative paucity of lepidopteran microsatellites described in the literature (NÈVE and MEGLÉCZ 2000); Appendix 3.1).

Only recently has the collective set of observations been recognized as a genetic phenomenon in itself rather than being treated as a methodological nuisance for obtaining an acceptable number of markers (MEGLECZ *et al.* 2004; ZHANG 2004). The four major features of Lepidopteran microsatellites that have been suggested as possible causes of these low yields of markers are: (*i*) low genomic frequency of microsatellites, (*ii*) low proportions of polymorphic vs. monomorphic markers, (*iii*) unstable flanking sequences interfering with PCR amplification, and (*iv*) multiple occurrences of similar flanking sequences. The following sections treat these reported features in turn.

(*i*) Frequencies of microsatellites in Lepidoptera have been described in *Parnassius mnemosyne* and *Bombyx mori* (MEGLÉCZ and SOLIGNAC 1998; PRASAD *et al.* 2005; REDDY *et al.* 1999). These show an average CA-repeat occurrence every 97kb in *P. mnemosyne* and 40kb in *B. mori*, which is larger than the interval found in most other taxa, but not unusual, and not nearly enough to explain the differences in yields with other (insect) groups (NÈVE and MEGLÉCZ 2000). Moreover, enrichment techniques used in the more recent studies did not substantially improve genetic marker yields, implying that the relative scarcity of microsatellites is not the primary cause for the poor results.

(*ii***)** Where specified, the proportion of monomorphic loci is usually low in Lepidoptera, and never high enough to explain the low number of discriminating markers as can be seen in Appendix 3.1.

(*iii***)** Heterozygote deficiency has been reported in a large proportion of markers in most Lepidoptera studies (Appendix 3.1). This is primarily caused by the frequent occurrence of null alleles (CASSEL 2002; JIGGINS *et al.* 2005; VAN'T HOF *et al.* 2005). There is substantial evidence that many null alleles in Lepidoptera are caused either by mutations in primer binding sites resulting in unsuccessful PCR, or by indels that produce alleles with PCR fragment sizes which fall outside the standard detection range (FLANAGAN *et al.* 2002; JIGGINS *et al.* 2005; KEYGHOBADI *et al.* 1999; PALO *et al.* 1995; REDDY *et al.* 1999). Therefore, this relatively high flanking sequence variability, that manifests itself as null alleles, is in part responsible for the low yields.

(*iv***)** The primary cause of the difficulties in obtaining markers, however, is not that flanking sequences differ too much for successful amplification as described above, but rather that these sequences at more than one locus are too much alike. This usually results in more than two different distinguishable PCR products, causing uninterpretable banding patterns (ANTHONY *et al.* 2001; BOGDANOWICZ *et al.* 1997; JI *et al.* 2003; PALO *et al.* 1995; WILLIAMS *et al.* 2002). Our own data, based on several

microsatellite-enriched libraries of the Afrotropical butterfly, *Bicyclus anynana* (Satyridae), are consistent with such unusual microsatellite characteristics. Thus, we found that most sequences surrounding microsatellites show similarities. Of these, we found those with similar sequences on both sides of the microsatellite and those where only one flank matches other sequences. These two categories of flanking sequence similarity have been named symmetrical and asymmetrical respectively by (MEGLECZ *et al.* 2004) after finding analogous structures in two other butterfly species.

The present study focuses on the origins of the multiplications that have led to these multi-copy sequences, and on why this process is so widespread in Lepidoptera. We first consider the possibility that asymmetrical sequences might in fact be artifacts, representing chimeric PCR products formed during the enrichment PCR step (PÄÄBO *et al.* 1990).

Secondly, we focus on the mechanisms through which multi-copy DNA arises and how they are involved in *B. anynana* microsatellites. The two main pathways are by means of transposition of Mobile Elements (ME's) and by recombination. We surveyed the dataset for tandemly repeated patterns as would be the case after unequal crossing over or gene conversion, and also screened it for ME characteristics such as direct- or inverted repeats and for similarities with sequence data for known ME's. Furthermore, we examined whether the microsatellites co-migrate within their surrounding sequences or whether they were formed from proto-microsatellites after the multiplication event, as is the case in mini-me's in *Drosophila* (WILDER and HOLLOCHER 2001), primate Alu elements (ARCOT *et al.* 1995), and in introns of human and desert locust (*Schistocerca gregaria*) FABP genes (WU *et al.* 2001).

Finally, we consider our data in a broader perspective by making comparisons to other species with a particular emphasis on the Lepidoptera. We thus aim to find clues about a unitary mechanism, and to find out why these phenomena are mainly reported from butterflies and moths.

MATERIALS AND METHODS

DNA extraction, library construction and sequencing

The source material for all analyzed sequences is DNA extracted from thorax and head of a single butterfly using a standard Phenol-CIA protocol as described in (VAN'T HOF *et al.* 2005). A female was used to incorporate both the W and Z chromosomes. Enrichment for CA, GA, AAT, ATG, GAA and TACA motifs was performed by Genetic Identification Services (GIS, http://www.genetic-id-services.com; Chatsworth, CA, USA) using Hind III restriction and adapters, and a single round of enrichment with biotinylated microsatellite sequences as capture molecules. Positive DNA fragments of 350-700 were cloned in pUC19. The libraries were transformed into JM109 (Promega, Madison, WI, USA), followed by blue-white screening. Positive clones were grown in 200 µl LB with 100 µg/ml ampicillin and miniprepped using the Qiaprep spin miniprep kit (Qiagen, Hilden, Germany). Sequencing was outsourced to commercial facilities. The numbers of sequenced clones per library are given in Table 3.1.

Detection of intra-specific similarities

Similarities within this dataset were detected by comparing the sequences from all libraries with each other by means of 'all against all' standalone Nucleotide-nucleotide BLAST (BLASTN) (ALTSCHUL *et al.* 1997) and then manually fine-aligning where needed using BIOEDIT (HALL 1999). The length threshold for

considering sequence homologues was set to 40 bp. Shorter homologies with adjoining microsatellites that were omitted by BLASTN due to their repetitive nature were included.

Detection of inter-specific homologies

Homologies between our data and sequences submitted to GenBank were surveyed with online BLASTN using default settings. Distinction between hits that occur by chance and true 'common origin' data is not fully represented in the 'Blast Score' since it does not compensate for the differences in available sequences per species. Therefore, we used a threshold of 50 to include hits from large scale genome surveys, and a threshold of 40 for species with under-represented sequence data resources. The hits matching these criteria were then manually realigned with BIOEDIT for two reasons. First of all, repeat structures are not included in the BLASTN output whereas the detected match often continued into a shared microsatellite or even beyond, and secondly, many obvious homologies surrounding the returned sequence match were not reported by BLASTN.

Sequence regions that were reported from multiple species were aligned with BIOEDIT to construct a consensus sequence. Subsequently, this sequence was re-analyzed with online BLASTN, followed by an update of the consensus based on the additional hits. This process was repeated until no more new hits occurred.

Experiment I: Confirming the presence of specific sequences in genomic DNA

To test whether the different combinations of flanking sequences were an artifact caused by enrichment procedures, or in fact occur in the observed association in the butterfly genome, we designed primers with OLIGO version 6 (RYCHLIK 2000) to amplify 15 different combinations of symmetrical and asymmetrical sequence clusters in the ATG library. Product was detected with ethidium bromide-stained 1% agarose gel. PCR was performed in 10 µl, containing 5µl 2×Reddymix 1.5 (Abgene, Portsmouth, NH, USA), 0.33 µM of each primer, with 1µl 2nd elution DNeasy-tissue (Qiagen) extracted thorax as template. Thermal cycle was: 3 min. @ 95ºC; 30 cycles of 30 sec. 94ºC, 30 sec. T_a, 45 sec. 72ºC; followed by 30 min. @ 72ºC. T_a was 50°C for all but primer-pair 9 (BA-ATG244), where T_a = 47°C. The primer sequences are listed in Appendix 3.2.

Experiment II: Exploration of the spatial organization of common sequences

PCR primers were designed with an outward orientation instead of inward on both ends of the cloned insert (i.e. primers amplifying away from the microsatellite instead of towards it, as in inverse PCR). They were based on the consensus sequences of six symmetrical (microsatellite flanking sequence) groups (AAT group 1, ATG group 2A upstr.A-dstr.A, ATG group 2B upstr.F-dstr.A, CA groups 1, 2 & 3). This arrangement of primers will only result in amplification if the complementary primer is within range (see Fig. 3.1). PCR was performed as in experiment I, but with a 55ºC T_a for ATG group 2A upstr.A-dstr.A, which is 5 ºC above the advised T_a to reduce the chance of non-specific priming. Amplification products were detected with ethidium bromide-stained 1% agarose gel. The primer sequences are given in Appendix 3.3.

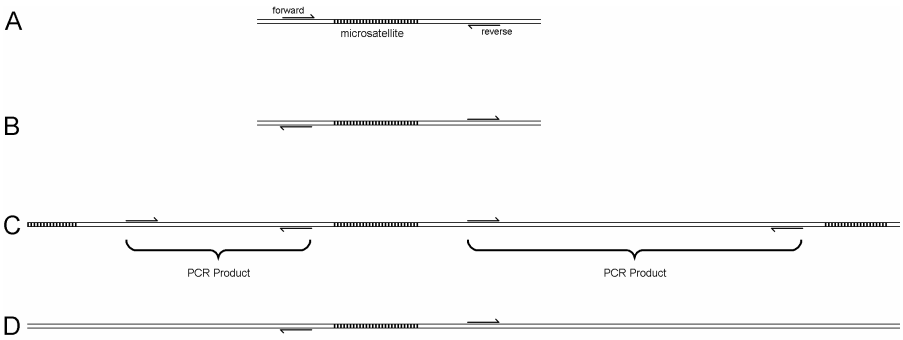


Figure 3.1 Response of PCR amplification to different microsatellite-flank arrangements. A: Example of normal microsatellite primer design with a forward and a reverse primer on either side of the repeat, initiating polymerization directed towards each other. B: Primer design for this particular experiment with primers oriented away from the microsatellite and more importantly, away from each other. C & D: The two possible scenarios; C: Tandem arrangement with a relatively short distance between the units, resulting in exponential amplification, or D: No tandem arrangement, or large repeat units with too distant primer recognition sites for successful amplification.

A band of approximately 275bp that consisted of two merged amplicons produced with the ATG group 2A upstr.A-dstr.A primer combination was excised and purified with Qiaquick gel extraction kit and inserted into a cloning vector using the pGEM-T system (Promega). Transformation, cloning and sequencing was performed as described above for the enriched library.

RESULTS

Sequence similarities within the *B. anynana* libraries

Most sequences from the *B. anynana* enriched libraries showed typical Lepidopteran microsatellite characteristics, such as symmetrical and asymmetrical flanking regions surrounding the repeat structure. These multi-copy sequences were found in all of the six libraries and their details are summarized in Table 3.1. The standalone 'all against all' BLASTN revealed that sequences are not only associated within the different enriched libraries, but also frequently between them. Compound microsatellites selected by multiple enrichment probes make up just a small fraction of these intra-library links. The proportion of clones that show no similarity is 80 out of 289, which is an overestimate, since large numbers of redundant clones were filtered out before sequencing (VAN'T HOF *et al.* 2005). Sequence data have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY785060, AY785062, AY785064, AY785071, AY785080, AY785081, DQ225274-DQ225304, EF114667-EF114669.

Table 3.1 Properties of the sequences extracted from the six enriched libraries.

'S.C.' stands for Single Copy, 'M.C.' for Multi Copy. For the three different M.C. classes, the numbers of homologous groups are given for the intra-library homologies '(# of groups)'. The category 'M.C. microsat' is composed of symmetrical, asymmetrical and partial homologies. 'No tandem repeats' consists of single- and multi-copy clones without microsatellite or minisatellite structures.

Redundancy and repeat type categories	Libraries						Total characteristics [shared characteristics]
	CA	GA	AAT	ATG	GAA	TACA	
S.C. microsat	24	5[1]	3	6	0	3	41[1] [1]
M.C. microsat (# of groups)	117 (12)[1]	3 (1)	13 (4)	34 (4)	8 (3)	9 (2)	184[1] [4]
S.C. minisat	7	1[1]	4	2	2	1	17[1] [1]
S.C. minisat with microsat.	9	2	0	1	0	1	13
M.C. minisat (# of groups)	12 (4)[1]	0	2 (1)	0	0	0	14[1] [4]
M.C. minisat with microsat (# of groups)	2 (1)	0	0	2 (1)	1	1	6
No tandem repeats (of which M.C.)	6 (1)	0	1 (1)	8 (5)	3 (2)	1 (0)	19 (9)
Total characteristics (total clones)	177 (173) [1]	11 (10) [1]	23	53	14	16	294 (289) [1]

Confirmation of the presence of cloned sequences in genomic DNA

The PCR amplification of different asymmetrical combinations gave robust amplification products in each of the 15 different upstream-downstream primer combinations (Fig. 3.2). This showed that the observed data is not an enrichment artifact, but that these asymmetrical structures actually occur as contiguous sequences in the *B. anynana* genome. Most of the PCR products showed more than one distinct band, indicative of multiple copies with a variable distance between the primer binding sites.

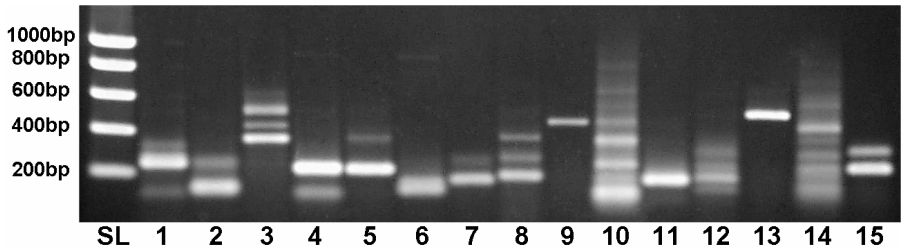


Figure 3.2 PCR product from 15 different primer pair combinations designed to test sequence associations found in the ATG library. Lane numbers correspond to the following primer combinations (see Fig. 3.3 for primer locations): SL = Eurogentec Smartladder; **1** = BA-ATG1 subgroup 2A upstream A and downstream J (2A-uA-dJ); **2** = BA-ATG1/2B-uB-dA$_1$; **3** = BA-ATG108/2B-uF-dB; **4** = BA-ATG3/2B-uC-dA$_1$; **5** = BA-ATG212/ single copy microsatellite region; **6** = BA-ATG212/2A-uA-dC; **7** = BA-ATG213/2A-uA-dD; **8** = BA-ATG215/2A-uA-dE; **9** = BA-ATG244/2A-uA-dI; **10** = BA-ATG244/2B-uA-dA$_1$; **11** = BA-ATG248/2B-uD-dA$_1$; **12** = consensus 2A-uA-dA; **13** = consensus 2A-uA-dB; **14** = consensus 2B-uF-dA$_3$; **15** = consensus 2B-uH-dA$_5$.

1 Some sequences contain both a microsatellite and a minisatellite (not to be confused with a microsatellite inside a minisatellite) and are, therefore, included twice in the statistics. For that reason, the table states both "total characteristics" and "total clones". The sum of '[shared characteristics]' divided by two (10/2 =5), subtracted from 'Total characteristics' provides the total number of clones (294-5=289)

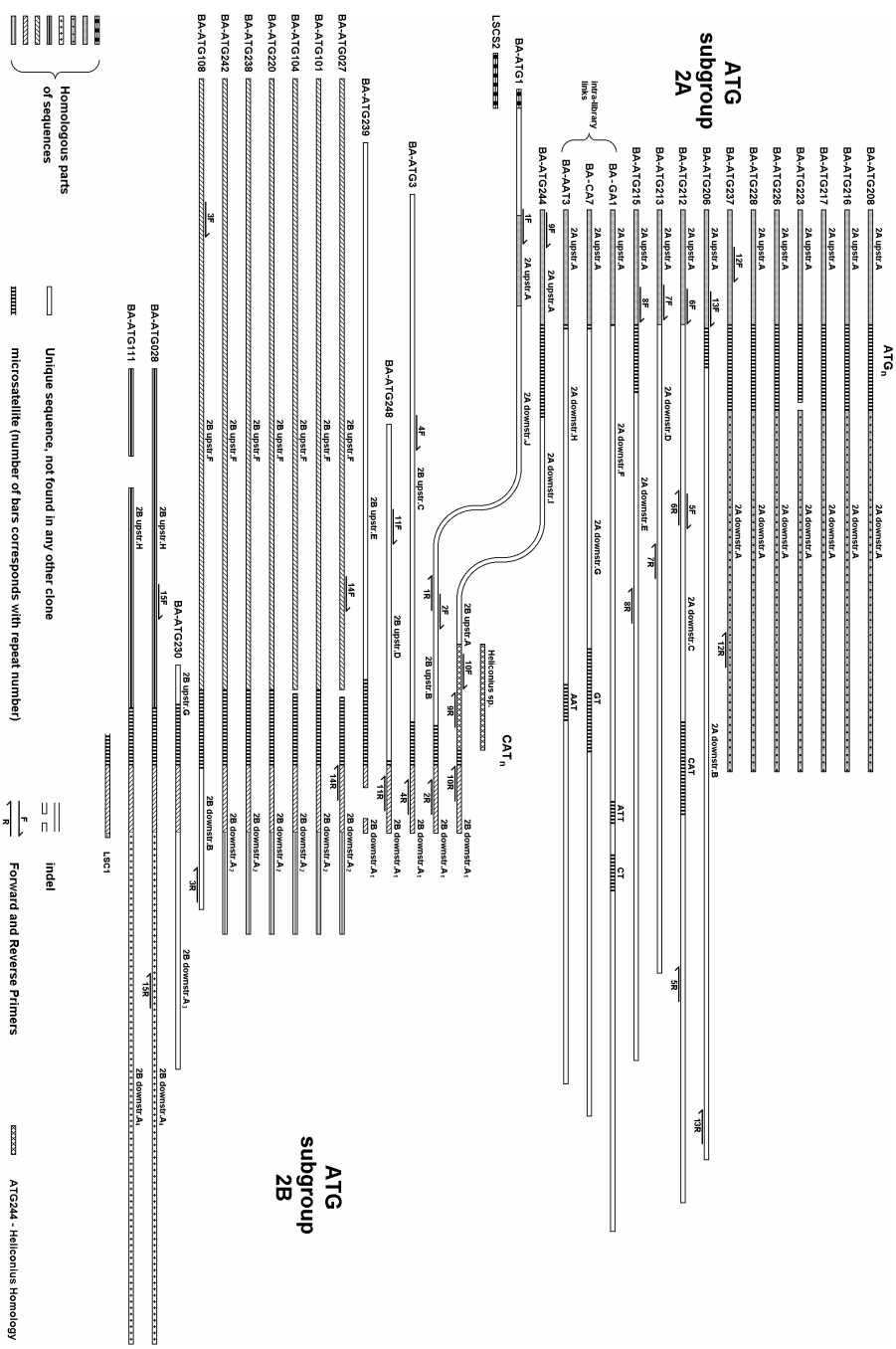


Figure 3.3 Schematic representation of the alignment of ATG group 2 sequences, showing two subgroups (2A & 2B) linked together by two sequences possessing characteristics of both. The majority are intra-library links, grouped to symmetrical sequence families with variable microsatellites, and asymmetrical alignments. Inter-library links are shown in subgroup ATG 2A and inter-specific hits are represented in

subgroup ATG 2B by the Lepidoptera Specific Core Sequence LSCS1, a small section of LSCS2 in BA-ATG1 (see inter-specific comparison section in Results) and a *Heliconius* sequence. The arrows represent the location of the primer primers used in the control experiment to verify the existence of several upstream-downstream combinations, forward (F) and reverse (R) primer numbers corresponding to the lane numbers in Fig. 3.2.

Relative orientation of common sequences

A sequence family from the ATG library is represented as a schematic alignment in Fig. 3.3 to provide an example of the similarity patterns. The ATG2 sequence family consists of two subgroups that are linked together by sequences that possess characteristics of both clusters (BA-ATG244 and BA-ATG1). Subgroup 2A is defined by a 60bp sequence directly adjacent (upstream) to the ATG_n repeat (2A upstr. A), and subgroup 2B is characterized by a 31bp sequence immediately beyond a common CAT_n repeat (2B downstr. A). The relative positions of the different sequence regions are designated by (*i*): the alignment subgroup (2A or 2B), (*ii*): their position upstream/downstream (u/d) relative to the aligned microsatellites, and (*iii*): by their class of similarity within each subgroup (A-J). Two clusters, 2A-uA-dA and 2B-uF-dB are typical examples of symmetrical associations, possessing similarities on both sides of the microsatellites. Both subgroups also have many asymmetrical associations with some flanks overrepresented, rather than a random mixture of upstream-downstream combinations (e.g. BA-ATG206, 212, 213, 215 etc.). The prevalence of one type of flank on one side and variation on the other side of the microsatellite is a characteristic of most other asymmetrical groups that were found in *B. anynana*. Asymmetrical inter-library alignments are represented in Fig. 3.3 by BA-GA1, BA-CA7 & BA-AAT3. They match up with 2A-uA, followed by an ATG_1 or ATG_2 in line with the ATG_n site.

The two main aligned microsatellites in Fig. 3.3 both differ markedly in repeat numbers with zero to 26 repeats in 2A and three to 29 in 2B. Additional microsatellites present in some "2A" sequences appear to be unrelated to the aligned ATG_n, and consist of different repeat types. These sequences often align partially or asymmetrically to other sequences or groups of sequences either within or between libraries (not shown in Fig. 3.3).

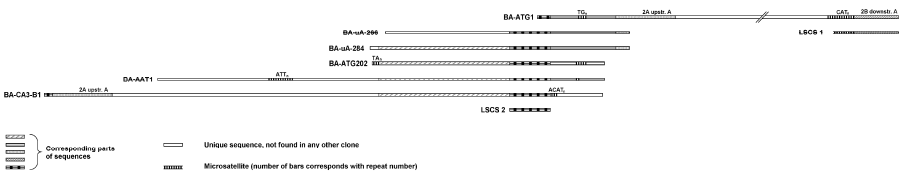


Figure 3.4 Schematic alignment of sequences upstream of subgroup ATG2A, showing the full LSCS2 alignment. There is partial overlap with Fig. 3.3, which uses the same patterns for homologous sections.

The experiment to explore sequences surrounding some of the sequence families gave a positive amplification result in three of the six combinations (ATG-2A-uA-dA, ATG-2B-uF-dA, CA group 3). This implies that some common sequences are repeated relatively closely beyond the known sequence.

The ATG-2A-uA-dA band that was sequenced from these PCR products consists of a 266bp and a 284bp fragment (BA-uA-266 and BA-uA-284). They both match with the upstream-A flank, including the BA-ATG1 extension (Fig. 3.3), but shared little more than the primer sequence with the ATG-2A-dA region. The BA-uA-266 and BA-uA-284 sequences form a link between a sequence cluster consisting of BA-ATG202, BA-AAT1 and BA-CA3-B1 plus the upstream part of the ATG2A subgroup. The schematic alignment of these sequences is presented in Fig. 3.4, which has partial overlap with Fig. 3.3. The BA-uA-266 and BA-uA-284 sequences are nearly identical for about half their length, but loose their similarity immediately after a 35bp non-random sequence that is associated with multi-copy microsatellites in many Lepidoptera species. This sequence, designated LSCS2, will be discussed in detail below.

The ATG-2A-uA sequence that characterizes subgroup ATG 2A recurs further upstream in the BA-CA3-B1 sequence (Fig. 3.4). Furthermore, this group of sequences incorporates a microsatellite that is variable in repeat number, but whose variability does not alter the overall length of the sequence (i.e. caused by base substitutions rather than by means of DNA replication slippage). This could either represent the different stages of a developing proto-microsatellite, or a microsatellite in decay.

Inter-specific comparison with *B. anynana* microsatellite sequences

The online BLASTN comparison of the *B. anynana* sequences resulted in hits with nine butterflies, 23 moths, one Coleoptera, two Diptera and two Hymenoptera (the species list is available in Appendix 3.4). Four distinct Lepidoptera Specific Core Sequences (LSCS), nearly exclusively matching a wide range of Lepidoptera species, were identified from these BLAST hits. They are generally situated next to a microsatellite, and usually define the position where similar regions start to differ in sequence.

LSCS1 is a 38bp sequence that corresponds with the ATG2B-dA sequence that is aligned in Fig. 3.3 and in Fig. 3.4. A BLAST search of this core sequence results in over 40 hits within 15 Lepidoptera species and one Coleoptera species (*Diabrotica virgifera*). With one exception, they all have a microsatellite in the same position as the CAT_n region in *B. anynana*. In addition to the predominant CAT_n repeats in these BLAST hits, several of these sequences also contain ATT_n, $CCAT_n$, $CAAT_n$ or CA_n arrays. The LSCS1 in *D. virgifera* is tightly between two microsatellites (CAT_n and CA_n). The 35bp LSCS2 matches the common sequence in the aligned cluster shown in Fig. 3.4, and also aligns with the extreme end of BA-ATG1 (Fig. 3.3). This core sequence is present in 13 deposited sequences from eight Lepidoptera species. In contrast to the other three LSCS, this sequence is not typically bordered by a microsatellite, although there is a small microsatellite immediately beyond it in BA-CA3-B1. The 150bp LSCS3 was detected in 11 Lepidoptera species, based on the BA-TACA105 BLAST hits. It spans both flanks of a common $CAAA_n$ microsatellite and is associated with retrotransposons in *Bombyx mandarina* (GenBank acc# AB055223), *B. mori* (GenBank acc# AB032718) and *Antheraea mylitta* (GenBank acc# AF530471). The LSCS4, identified from BA-TACA112, consists of a 85bp sequence and was found in six Lepidoptera species, usually bordered by a microsatellite. The sequences of the four LSCS and the alignments with their BLAST hits are presented in appendices 3.5-3.9.

Besides the four core sequences that were present in many Lepidoptera species, there were also a number of more solitary hits, but still predominantly from

Lepidoptera and often associated with microsatellites. One of these inter-specific links is represented in Fig. 3.3 by *Heliconius cydno* and *H. melpomene* microsatellite flanks corresponding with part of the BA-ATG244 sequence.

Minisatellite structures

In addition to the microsatellites, 15% of the clones contained minisatellites with repeat units ranging from 14 to 55 bp, either with or without a microsatellite incorporated within each unit. Most of the microsatellites embedded in minisatellite units showed repeat number variation, which is possibly (but not necessarily) caused by slipped strand mispairing (relatively frequently occurring mutations adding or removing a repeat unit) as is the case in solitary microsatellites (Fig 3.5).

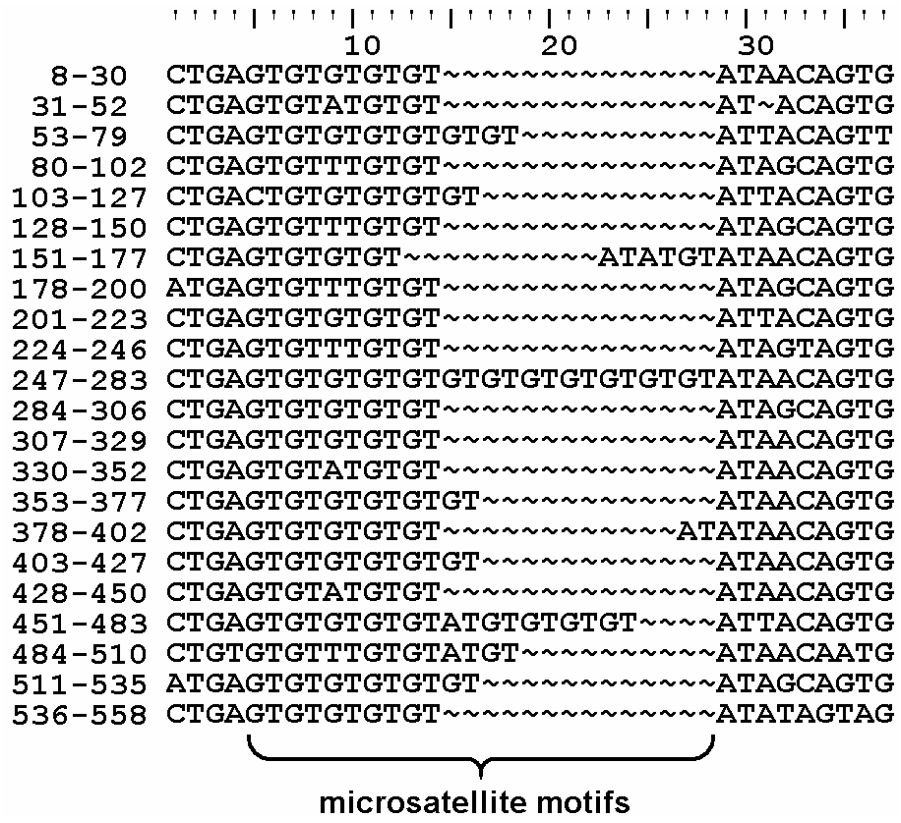


Figure 3.5 Internal alignment of a 551 bp stretch of BA-CA1-G4, showing 22 minisatellite units with incorporated variable microsatellites (GT_4-GT_{12}).

Many of the minisatellites could be grouped together in gene families in the same way as described above for the multi-copy microsatellites. The different representatives of each family show variation in number of repeat units, composition

of the units and of their flanking sequences. An overview of the numbers of clones containing the different minisatellite characteristics can be found in Table 3.1. The 10bp Jeffreys core sequence (GGGCAGGANG) (JEFFREYS *et al.* 1985) was found as a 9/10 base match and a 100% match in the repeat units of BA-CA4-C1 and BA-AAT2-B11, respectively.

DISCUSSION

In (GOLDSTEIN and SCHLÖTTERER 1999), the flanking region is described as "The single-copy DNA sequence immediately upstream and downstream of a microsatellite locus that allows the design of specific primers that preferentially amplify the target microsatellite". The *B. anynana* dataset presented here suggests that this definition cannot be universally applied, because most microsatellites in this species are located within repetitive DNA. This appears to be a general characteristic of Lepidoptera (Appendix 3.1), and has also been found in some other insects, such as Coleoptera ((LIEWLAKSANEEYANAWIN *et al.* 2001); N. Margraf pers. comm) and Diptera (FAGERBERG *et al.* 2001; WILDER and HOLLOCHER 2001). Apart from observations in insects, microsatellites associated with repetitive DNA have also been reported in vertebrates (ALEXANDER *et al.* 1995; ARCOT *et al.* 1995; BAND and RON 1996) and in plants (RAMSAY *et al.* 1999; TEMNYKH *et al.* 2001; TERO *et al.* 2006). The possibility that multiple variants of a certain locus were incorporated in the genomic library by means of chimeric reassociation during the PCR-based enrichment (PÄÄBO *et al.* 1990) was dismissed by the successful genomic PCR amplification of 15 different repetitive DNA sequences. We usually found amplicons of different sizes per amplification, indicating that they originate from multiple loci (Fig 3.2). A similar experiment was performed by (TERO *et al.* 2006), who found that 82.1% of the tested primer combinations confirmed that the sequences derived from their genomic library were contiguous in *Silene tatarica*, and sequencing of amplification products of different sizes revealed that they represent heterogeneous loci.

Another indication that the sequences obtained from the *B. anynana* library are contiguous is given by the fact that a number of sequences with similar regions can be amplified uniquely and serve as polymorphic microsatellite markers as long as the primers target unique parts of these sequences (BA-GA1, BA-CA7, BA-AAT3, BA-ATG1 and BA-ATG3, all represented in Fig. 3.3).

Enrichment procedures may however have a bias towards certain sequences other than the repeat itself. For instance, the BA-ATG213 sequence that belongs to the ATG2 family was included in the library in spite of not containing a microsatellite.

Repeat unit definition

The two main mechanisms for multiplication of DNA sequences are by means of transposition of mobile elements (ME's) that have the ability to incorporate copies of themselves elsewhere in the genome, or even in other individuals, and recombination related events, such as unequal crossing over (UCO) and gene conversion that result in tandemly arranged homologues.

One limitation of the material studied here to distinguish between these two possibilities is that it is not always clear what defines the extremes of a repeated unit. Inserts were selected in the 350-700bp range, while many ME's and recombination-products are larger. There are however two common structures where similar sequences start to differ. First of all, asymmetrical sequences are by definition

identical on one side of the microsatellite and different on the other side. Secondly, there are the LSCS structures that usually define the start of sequence divergence.

Mobile elements

There is support in our data for the hypothesis that ME's are responsible for the abundance of similarity surrounding microsatellites. The BA-TACA105 derived LSCS3 fully matched to Lepidopteran retrotransposons of *Bombyx mori*, *B. mandarina* and *Antheraea mylitta*. It is possible that the other three LSCS are structural units of mobile elements as well. The fact that LSCS1 and LSCS2 are present in a single sequence (BA-ATG1) would indicate that they define different parts of the same mobile element.

Mobile elements usually have specific characteristics such as inverted or direct repeats at their extremes, or poly-A tracts (for an overview see (BERG and HOWE 1989)). A small number of short direct and inverted repeats were found in *B. anynana*, and 25 clones contained a poly-A homopolymer of 10 or more base pairs.

Another observation in *B. anynana* that supports ME's rather than recombination is the independent inheritance of asymmetrical loci in an F2 cross, indicating that the microsatellites in question are not closely linked (VAN'T HOF *et al.* 2005).

Examples of ME associated microsatellites in other Lepidoptera species are those in the very common *Bombyx mori* BM1 elements, which are "surrounded by short direct repeats (2-6bp)" (EICKBUSH 1995) and the similarities between *Parnassius* microsatellite clones and a *Drosophila* retrotransposable element and a human retrovirus (MEGLECZ *et al.* 2004).

At odds with the involvement of proto-microsatellite containing ME's (WILDER and HOLLOCHER 2001) are some very distinct polymorphisms that interrupt the microsatellites in *B. anynana.* They manifest themselves in different loci or repeat units (e.g. CA group 2, Appendix 3.10), indicating that the microsatellites must have been present before the multiplication event, and hitchhiked in conjunction with the flanking sequences.

Recombination as cause for repetitive sequences

There is also support for the involvement of recombination as a mechanism for part of the observed repetitive sequences from the present dataset. Minisatellites are generated trough recombination, and each minisatellite unit of a microsatellite-containing minisatellite can be described as a microsatellite with flanking regions, just as in a solitary microsatellite, only with much shorter flanks. On a larger scale, the BA-CA3-E3 sequence shows two tandemly arranged units of approximately 100bp each, that both include a CA_n repeat (CA_9 and CA_{13} respectively), which can also be defined as microsatellites with flanking sequences. When similar microsatellite-containing repeated units become much larger (i.e. larger than the cloned insert) it is impossible to detect their higher order repetitive nature within the currently available sequences. It is therefore possible that part of the repetitiveness is comparable to the microsatellite-containing minisatellites, but with a much larger unit size. The BA-CA3-B1 sequence indirectly positions the ATG2A-uA sequence upstream of the main ATG2A-uA alignment (Fig. 3.4), which may represent tandem arrangement. The fact that the alignment ends after the $ACAT_n$ microsatellite in this sequence could be due to an indel as described below, thus it is not unlikely that the ATG2A-uA sequence actually recurs downstream of this sequence.

The asymmetrical sequence arrangements fit perfectly within the description of UCO (i.e., where a chiasma occurs at two imperfectly aligned microsatellites with

shared repeat units, leaving two new upstream-downstream combinations) (MEGLECZ *et al.* 2004). There are, however, some discrepancies. One of the features in Lepidoptera microsatellites is that they often possess indels of various sizes directly adjacent to the microsatellite (FLANAGAN *et al.* 2002; REDDY *et al.* 1999). If such an indel is too large to find a match within a sequence family, it may be misinterpreted as a completely different flank. For example, the BA-CA3-E11 clone, belonging to CA group 2, contains a 173bp deletion immediately after the microsatellite, and rejoins at the end of the main alignment with a perfect match of 35bp (Appendix 3.10). Had the deletion been 35 or more bases larger no matching sequence would have been found and it might have been wrongly attributed to misaligned-microsatellite UCO. The fact that there are instances where indels form an alterative explanation for the observed asymmetries does not however rule out recombination as a contributory mechanism for repetitiveness altogether.

Lepidoptera specific homologues

The comparison of *B. anynana* clones with GenBank resulted in a large number of hits that were very strongly biased towards butterflies and moths. One could argue that it is not surprising to BLAST Lepidoptera sequences and get Lepidoptera hits in return. The issue here however, is that some regions seem to be very widely conserved in Lepidoptera, and more importantly, they are associated with the very phenomena we are exploring, namely multi-copy microsatellite flanking regions in Lepidoptera. It seems therefore, that there is a shared mechanism involved in the Lepidoptera that is reflected in the conservation of certain sequences. In particular, the four LSCS seem to be so frequent and widely distributed in this group that they may be key sequences for further investigation of these issues.

Sister chromatid association in Lepidoptera

The impression that the patterns described are peculiar to Lepidoptera raises the question of what might distinguish them from other groups. One uncommon feature present in all Lepidoptera is their holocentric chromosome arrangement, where chromatids assemble over their entire length instead of being joined at a centromere. Depletion of KLP-19, an essential microtubule motor, caused misalignment of holocentric kinetochores in the cabbage moth, *Mamestra brassicae* (MANDRIOLI *et al.* 2003). This suggests a direct link between holocentric chromosomes in Lepidoptera and UCO. However, a survey of other species with holocentric chromosomes, including *Caenorhabditis elegans*, species of Hemiptera and certain plants did not reveal similar microsatellite flank redundancies, while other species that did possess them, such as some Coleoptera and Diptera, have centromere associated chromosomes.

Over-representation of multi-copy microsatellites vs. under-representation of unique microsatellites

The low ratio of single- to multi-copy microsatellites from various studies on Lepidoptera has generally been interpreted as indicating high frequencies of the latter, relative to other taxa. An alternative, or complementary, interpretation is that single-copy microsatellites are scarce in Lepidoptera. This may also be reflected in the large number of null-alleles reported in Lepidoptera, since if there are too few alternatives to these suboptimal microsatellite loci, they are more likely to be utilized and published. The Introduction cited data from (PRASAD *et al.* 2005), interpreted as indicating that microsatellite densities are not unusually low in *Bombyx mori*;

however, in this study the microsatellite densities obtained from more than 4400 in silico detected loci (total density of one locus per 6.4kb) are not separately specified as single- and multi-copy loci, which makes it difficult to determine whether multi-copy microsatellites are unusually abundant or unique microsatellites scarce.

Conclusion

Our exploration of different hypotheses that may explain these unusual observations provided no clear-cut mechanism, since there is support for both recombination and ME's being implicated in the multiplication events. Therefore, a combination of both explains our observations best. The question remains as to whether we are dealing with two separate processes, that both lead to redundancy, or if it is an integrated mechanism.

Analysis of the repetitive microsatellite characteristics in *B. anynana* and other Lepidoptera species revealed a number of Lepidoptera specific patterns that provides a basis for further research on this subject. The four core sequences appear to hold valuable information and may serve as a starting point for further investigations (e.g. *in situ* hybridization), leading to a better understanding of the mechanisms involved, and possibly in defining a new type of Lepidopteran Mobile Element. These findings may not only lead to a more complete knowledge of micro- and minisatellites in Lepidoptera, but may have general implications for understanding VNTR dynamics.

ACKNOWLEDGEMENTS

We thank two anonymous referees and the subject editor for comments on a previous version of this paper.

REFERENCES

ALEXANDER, L. G., G. ROHRER and C. BEATTIE, 1995 Porcine SINE-associated microsatellite markers: evidence for new artiodactyl SINEs. Mamm. Genome **6:** 464-468.

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHÄFFER, J. ZHANG, Z. ZHANG *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389-3402.

ANTHONY, N., G. GELEMBIUK, D. RATERMAN, C. NICE and R. FFRENCH-CONSTANT, 2001 Isolation and characterization of microsatellite markers from the endangered Karner blue butterfly *Lycaeides melissa* samuelis (Lepidoptera). Hereditas **134:** 271-273.

ARCOT, S., Z. WANG, J. WEBER, P. DEININGER and M. BATZER, 1995 Alu repeats-a source for the genesis of primate microsatellites. Genomics **29:** 136-144.

BAND, M., and M. RON, 1996 Creation of a SINE enriched library for the isolation of polymorphic (AGC)n microsatellite markers in the bovine genome. Animal Genetics **27:** 243-248.

BERG, D. E., and M. M. HOWE, 1989 *Mobile DNA*. American Society for Microbiology, Washington, DC.

BOGDANOWICZ, S. M., V. C. MASTRO, D. C. PRASHER and R. G. HARRISON, 1997 Microsatellite DNA variation among Asian and North American gypsy moths (Lepidoptera : Lymantriidae). Ann. Ent. Soc. Am. **90:** 768-775.

CASSEL, A., 2002 Characterization of microsatellite loci in *Coenonympha hero* (Lepidoptera : Nymphalidae). Mol Ecol. Notes **2:** 566-568.

EICKBUSH, T. H., 1995 Mobile elements of lepidopteran genomes, pp. 77-105 in *Molecular model systems in the Lepidoptera*, edited by M. R. GOLDSMITH and A. S. WILKINS. Cambridge University Press, Cambridge.

ELLEGREN, H., 2004 Microsatellites: simple sequences with complex evolution. Nat Rev Genet **5:** 435-445.

FAGERBERG, A. J., R. E. FULTON and W. C. BLACK, 2001 Microsatellite loci are not abundant in all arthropod genomes: analyses in the hard tick, *Ixodes scapularis* and the yellow fever mosquito, *Aedes aegypti*. Insect Mol. Biol. **10:** 225-236.

FLANAGAN, N. S., M. J. BLUM, A. DAVISON, M. ALAMO, R. ALBARRAN *et al.*, 2002 Characterization of microsatellite loci in neotropical *Heliconius* butterflies. Molecular ecology Notes **2:** 398-401.

GOLDSTEIN, D. B., and C. SCHLÖTTERER, 1999 *Microsatellites; Evolution and Applications*. Oxford University Press, New York.

HALL, T., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids. Symp. Ser. **41:** 95–98.

JEFFREYS, A. J., V. WILSON and S. W. THEIN, 1985 Hypervariable 'minisatellite' regions in human DNA. Nature **314:** 67-73.

JI, Y.-J., D.-X. ZHANG, G. M. HEWITT, L. KANG and D.-M. LI, 2003 Polymorphic microsatellite loci for the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and some remarks on their isolation. Mol. Ecol. Notes **3:** 102-104.

JIGGINS, C. D., J. MAVAREZ, M. BELTRÁN, W. O. MCMILLAN, J. S. JOHNSTON *et al.*, 2005 A genetic linkage map of the mimetic butterfly, *Heliconius melpomone*. Genetics **171:** 557-570.

KEYGHOBADI, N., J. ROLAND and C. STROBECK, 1999 Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilonidae). Molecular ecology **8:** 1481-1495.

LIEWLAKSANEEYANAWIN, C., C. E. RITLAND, C. H. NEWTON and Y. A. EL-KASSABY, 2001 Characterization of microsatellite loci in white pine weevil (*Pissodes strobi*). Mol. Ecol. Notes **1:** 248–249.

MANDRIOLI, M., G. C. MANICARDI and F. MAREC, 2003 Cytogenetic and molecular characterization of the MBSAT1 satellite DNA in holokinetic chromosomes of the cabbage moth, *Mamestra brassicae* (Lepidoptera). Chromosome Research **11:** 51-56.

MEGLECZ, E., F. PETENIAN, E. DANCHIN, A. C. D'ACIER, J. Y. RASPLUS *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.

MEGLÉCZ, E., and M. SOLIGNAC, 1998 Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). Hereditas **128:** 179-180.

NÈVE, G., and E. MEGLÉCZ, 2000 Microsatellite frequencies in different taxa. TREE **15:** 376-377.

PÄÄBO, S., D. M. IRWIN and A. C. WILSON, 1990 DNA damage promotes jumping between templates during enzymatic amplification. J. Biol. Chem. **265:** 4718-4712.

PALO, J., S. VARVIO, I. HANSKI and R. VÄINÖLÄ, 1995 Developing microsatellite markers for insect population structure: complex variation in a checkerspot butterfly. Hereditas **123:** 295-300.

PRASAD, M. D., M. MUTHULAKSHMI, M. MADHU, S. ARCHAK, K. MITA *et al.*, 2005 Survey and Analysis of Microsatellites in the Silkworm, *Bombyx mori*. Genetics **169:** 197-214.

RAMSAY, L., M. MACCAULAY, L. CARDLE, M. MORGANTE, S. D. IVANISSEVICH *et al.*, 1999 Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J **17:** 415-425.

REDDY, K. D., E. G. ABRAHAM and J. NAGARAJU, 1999 Microsatellite in the silkworm, *Bombyx mori*: abundance, polymorphism, and strain characterization. Genome **42:** 1057-1065.

RYCHLIK, W., 2000 OLIGO: primer analysis software, pp. in *Molecular Biology Insights*, West Cascade, CO, USA.

TEMNYKH, S., G. DECLERCK, A. LUKASHOVA, L. LIPOVICH, S. CARTINHOUR *et al.*, 2001 Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. **11:** 1441-1452.

TERO, N., R. NEUMEIER, R. GUDAVALLI and C. SCHLÖTTERER, 2006 *Silene tatarica* microsatellites are frequently located in repetitive DNA. J. Evol. Biol. **19:** 1612-1619.

VAN'T HOF, A. E., B. J. ZWAAN, I. J. SACCHERI, D. DALY, A. N. M. BOT *et al.*, 2005 Characterization of 28 microsatellite loci for the butterfly *Bicyclus anynana*. Mol. Ecol. Notes **5:** 169-172.

WILDER, J., and H. HOLLOCHER, 2001 Mobile elements and the genesis of microsatellites in Dipterans. Mol. Biol. Evol. **18:** 384-392.

WILLIAMS, B. L., J. D. BRAWN and K. N. PAIGE, 2002 Highly polymorphic microsatellite loci for *Speyeria idalia* (Lepidoptera : Nymphalidae). Mol. Ecol. Notes **2:** 87-88.

WU, Q., P. ANDOLFATTO and N. H. HAUNERLAND, 2001 Cloning and sequence of the gene encoding the muscle fatty acid binding protein from the desert locust, *Schistocerca gregaria*. Insect Biochemistry and Molecular Biology **31:** 553-562.

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

APPENDIX 3.1

Summary of Lepidoptera microsatellite marker development publications.

Reference	Species	Poly-morphic	Mono-morphic	Hetz. Deff.[1]	Multi locus flanks[2]
AMSELLEM *et al.* 2003	*Lobesia botrana*	7	0	YES	n.s.
ANTHONY *et al.* 2001	*Lycaeides melissa*	4	0	YES	n.s.
PALO *et al.* 1995	*Melitea cinxia*	2	0	YES	YES
REDDY *et al.* 1999	*Bombyx mori*	15	0	YES	YES
PRASAD *et al.* 2005[3]	*Bombyx mori*	36	n.s.	YES	n.s.
WILLIAMS *et al.* 2002	*Speyeria idalia*	4	1	YES	YES
ROUSSELET *et al.* 2004	*Thaumetopoea pityocampa*	5	0	NO	n.s.
DALY *et al.* 2004	*Biston betularia*	14	n.s.	YES	YES
KLÜTSCH *et al.* 2003	*Reissita simonyi*	14	n.s.	YES	n.s.
CALDAS *et al.* 2002	*Zale galbanata*	5	n.s.	YES	n.s.
CASSEL 2002	*Coenonympha hero*	7	n.s.	YES	n.s.
COATES and HELLMICH 2003	*Ostrinia nubilalis*	1[4]	n.s.	YES	YES[5]
COATES *et al.* 2005	*Ostrinia nubilalis*	10	n.s.	YES	YES
FLANAGAN *et al.* 2002	*Heliconius erato*	15	n.s.	YES	YES[6]
FLANAGAN *et al.* 2002	*Heliconius melpomene*	8	n.s.	YES	YES
JIGGINS *et al.* 2005[7]	*Heliconius melpomene*	18[8]	3	YES	YES
HARPER *et al.* 2000	*Lysandra bellargus*	5	1	YES	n.s.
WARDILL *et al.* 2004	*Chiasmia assimilis*	12	n.s.	YES	n.s.
BEZZERIDES *et al.* 2004	*Utetheisa ornatrix*	5	3	YES	YES
IBRAHIM *et al.* 2004	*Spodoptera exempta*	8	n.s.	YES	n.s.
SCOTT *et al.* 2004	*Helicoverpa armigera*	5	n.s	YES	n.s.
TAN *et al.* 2001	*Helicoverpa armigera*	5	0	YES	n.s.
JI *et al.* 2003	*Helicoverpa armigera*	5	n.s.	YES	YES
KEYGHOBADI *et al.* 1999	*Parnassius smintheus*	4	n.s.	YES	n.s.
KEYGHOBADI *et al.* 2002	*Parnassius smintheus*	4	0	YES	n.s.[9]
MEGLÉCZ and SOLIGNAC 1998	*Parnassius mnemosyne*	3	n.s.	YES	n.s.
PETENIAN *et al.* 2005	*Parnassius apollo*	6	n.s.	YES	YES
PETENIAN *et al.* 2005	*Euphydryas aurinia*	5	n.s.	YES	YES
BOGDANOWICZ *et al.* 1997	*Lymantria dispar*	4	n.s.	YES	n.s.
KOSHIO *et al.* 2002	*Lymantria dispar*	3	0	n.s.	n.s.
ENDERSBY *et al.* 2005	*Plutella xylostella*	6	n.s.	YES	YES
FAURE and SILVAIN 2005	*Busseola fusca*	8	n.s.	n.s.	n.s.
FAUVELOT 2005	*Drupadia theda*	5	n.s.	YES	YES
FAUVELOT 2005	*Arhopala epimuta*	5	n.s.	YES	YES
ZHOU *et al.* 2005	*Cydia pomonella*	11	6	NO	YES
FRANCK *et al.* 2005	*Cydia pomonella*	22	2	YES	YES
DELPORT *et al.* 2005	*Gonometa postica*	6	n.s.	YES[10]	n.s.
VAN'T HOF *et al.* 2005	*Bicyclus anynana*	28	13	YES	YES
GRACE *et al.* 2005	*Plodia interpunctella*	9	6	YES	n.s.
ZEISSET *et al.* 2005	*Maculinea nausithosus*	11	n.s.	YES	n.s.
ZEISSET *et al.* 2005	*Maculinea alcon*	1	5	YES	n.s.
JI *et al.* 2005	*Dendrolimus punctatus*	10	0	YES	YES

Appendix 3.1 presents 38 publications covering microsatellite marker development of 34 Lepidoptera species. Some publications cover multiple species, therefore, the total "marker development efforts" is 42, resulting in 361 polymorphic markers, giving an average of 8.6 markers per "marker development effort".
n.s. = "not specified"

1 Hetz. deff. stands for heterozygote deficiency mentioned in the text as: Heterozygote deficiency, Null alleles, unexpected large allele size differences, flank indels, flank mutations.

2 Mentioned in the text either explicitly or cryptically as: Multi-copy, multi-locus, duplicated locus, non-specific amplicons, nonspecific amplification, multiband patterns, "Too many genotypes, at one locus or the other, to be accounted for by a single mating", "very similar sequences in the regions flanking microsatellite repeats", redundancy, etc.

3 *In silico* developed Microsatellites derived from more than 28 Mb consisting of random sequences, Z-chromosomal BAC sequences, and non-redundant EST's. Mononucleotide tracts and relatively short microsatellites (from 5 repeats) are included in this study. This resulted in 198 microsatellite loci of which 36 were polymorphic. Whether or not the remaining 162 are monomorphic, or not considered polymorphic for other reasons, is not specified.

4 The title suggests two microsatellites, but in fact it is one microsatellite and one minisatellite.

5 Redundancy in GenBank deposited sequences

6 “Primers were designed for 31, 18 and 11 unique repeat sequences”; “we identified those loci that amplified a single or double band”

7 JIGGINS *et al.* 2005 refer to MAVAREZ & GONZALES 2004 as the reference wherein the microsatellite development is described.

8 This paper describes 23 loci, of which five have been previously published.

9 An identical sequence is mentioned

10 Authors claim that observed heterozygote deficiency is not due to null-alleles, but due to population size fluctuations.

APPENDIX 3.1 REFERENCES

AMSELLEM, L., A. M. RISTERUCCI and B. BENREY, 2003 Isolation and characterization of polymorphic microsatellite loci in *Lobesia botrana* Den. & Schiff. (Lepidoptera: Tortricidae). Molecular Ecology Notes **3:** 117-119.

ANTHONY, N., G. GELEMBIUK, D. RATERMAN, C. NICE and R. FFRENCH-CONSTANT, 2001 Isolation and characterization of microsatellite markers from the endangered Karner blue butterfly *Lycaeides melissa* samuelis (Lepidoptera). Hereditas **134:** 271-273.

BEZZERIDES, A., J. BEZZERIDES and T. EISNER, 2004 Isolation of five polymorphic microsatellite markers in *Utetheisa ornatrix* (Lepidoptera: Arctiidae). Molecular Ecology Notes **4:** 566-567.

BOGDANOWICZ, S. M., V. C. MASTRO, D. C. PRASHER and R. G. HARRISON, 1997 Microsatellite DNA variation among Asian and North American gypsy moths (Lepidoptera : Lymantriidae). Ann. Ent. Soc. Am. **90:** 768-775.

CALDAS, A., D. J. HAWTHORNE and P. BARBOSA, 2002 Isolation and characterization 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). Mol Ecol. Notes **2:** 566-568.

COATES, B. S., and R. L. HELLMICH, 2003 Two sex-chromosome-linked microsatellite loci show geographic variance among North American *Ostrinia nubilalis*. Journal of Insect Science **3:** 6pp.

COATES, B. S., R. L. HELLMICH and L. C. LEWIS, 2005 Polymorphic CA/GT and GA/CT microsatellite loci for *Ostrinia nubilalis* (Lepidoptera: Crambidae). Mol Ecol. Notes **5:** 10-12.

DALY, D., K. WALTHAM, J. MULLEY, P. C. WATTS, A. ROSIN *et al.*, 2004 Trinucleotide microsatellite loci for the peppered moth (*Biston betularia*). Molecular Ecology Notes **4:** 179-181.

DELPORT, W., J. W. FERGUSON and P. BLOOMER, 2005 Characterization of six microsatellite loci in the African wild silk moth (*Gonometa postica*, Lasiocampidae). Mol. Ecol. Notes **5:** 860-862.

ENDERSBY, N. M., S. W. MCKECHNIE, H. VOGEL, L. J. GAHAN, S. W. BAXTER *et al.*, 2005 Microsatellites isolated from diamondback moth, *Plutella xylostella* (L.), for studies of dispersal in Australian populations. Mol. Ecol. Notes **5:** 51-53.

FAURE, N., and J. F. SILVAIN, 2005 Characterization of eight microsatellite loci in the maize stalk borer *Busseola fusca* Fuller, 1901 (Lepidoptera: Noctuidae). Mol. Ecol. Notes **5:** 846-848.

FAUVELOT, C., 2005 Isolation and characterization of microsatellites in two tropical butterflies, *Drupadia theda* and *Arhopala epimuta* (Lepidoptera: Lycaenidae). Mol. Ecol. Notes **5:** 724-726.

FLANAGAN, N. S., M. J. BLUM, A. DAVISON, M. ALAMO, R. ALBARRAN *et al.*, 2002 Characterization of microsatellite loci in neotropical *Heliconius* butterflies. Molecular ecology Notes **2:** 398-401.

FRANCK, P., F. GUÉRIN, A. LOISEAU and B. SAUPHANOR, 2005 Isolation and characterization of microsatellite loci in the codling moth *Cydia pomonella* L. (Lepidoptera, Tortricidae). Mol. Ecol. Notes **5:** 99-102.

GRACE, T., S. KAMBHAMPATI and B. SUBRAMANYAM, 2005 Characterization of polymorphic microsatellite markers in the Indian meal moth, *Plodia interpunctella* (Hübner). Mol. Ecol. Notes **5:** 321-322.

HARPER, G. L., S. PIYAPATTANAKORN, D. GOULSON and N. MACLEAN, 2000 Isolation of microsatellite markers from the Adonis blue butterfly (*Lysandra bellargus*). Molecular Ecology Notes **9:** 1919-1952.

IBRAHIM, K. M., Y. YASSIN and A. ELGUZOULI, 2004 Polymerase chain reaction primers for polymorphic microsatellite loci in the African armyworm, *Spodoptera exempta* (Lepidoptera: Noctuidae). Mol. Ecol. Notes **4:** 653-655.

JI, Y. J., Y. P. HUA, Y. D. LIU, H. T. LIU, C. M. SHI *et al.*, 2005 Ten polymorphic microsatellite markers developed in the masson pine moth *Dendrolimus punctatus* Walker (Lepidoptera: Lasiocampidae). Mol. Ecol. Notes **5:** 911-913.

JI, Y. J., D. X. ZHANG, G. M. HEWITT, L. KANG and D. M. LI, 2003 Polymorphic microsatellite loci for the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and some remarks on their isolation. Mol. Ecol. Notes **3:** 102-104.

JIGGINS, C. D., J. MAVAREZ, M. BELTRÁN, W. O. MCMILLAN, J. S. JOHNSTON *et al.*, 2005 A genetic linkage map of the mimetic butterfly, *Heliconius melpomone*. Genetics **171:** 557-570.

KEYGHOBADI, N., J. ROLAND and C. STROBECK, 1999 Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilonidae). Molecular ecology **8:** 1481-1495.

KEYGHOBADI, N., J. ROLAND and C. STROBECK, 2002 Isolation of novel microsatellite loci in the Rocky Mountain apollo butterfly, *Parnassius smintheus*. Hereditas **136:** 247-250.

KLÜTSCH, C. F. C., B. MISOF and C. M. NAUMANN, 2003 Characterization of microsatellite loci for *Reissita simonyi* (Rebel, 1899) (Lepidoptera, Zygaenidae). Molecular Ecology Notes **3:** 528-531.

KOSHIO, C., M. TOMISHIMA, K. SHIMIZU, H. S. KIM and O. TAKENAKA, 2002 Microsatellites in the gypsy moth, *Lymantria dispar* L. (Lepidoptera : Lymantriidae). Applied Entomology and Zoology. **37:** 309-312.

MEGLÉCZ, E., and M. SOLIGNAC, 1998 Microsatellite loci for *Parnassius mnemosyne* (Lepidoptera). Hereditas **128:** 179-180.

PALO, J., S. VARVIO, I. HANSKI and R. VÄINÖLÄ, 1995 Developing microsatellite markers for insect population structure: complex variation in a checkerspot butterfly. Hereditas **123:** 295-300.

PETENIAN, F., E. MEGLÉCZ, G. GENSON, J. Y. RASPLUS and E. FAURE, 2005 Isolation and characterization of polymorphic microsatellites in *Parnassius apollo* and *Euphydryas aurinia* (Lepidoptera). Mol. Ecol. Notes **5:** 243-245.

PRASAD, M. D., M. MUTHULAKSHMI, M. MADHU, S. ARCHAK, K. MITA *et al.*, 2005 Survey and Analysis of Microsatellites in the Silkworm, *Bombyx mori*. Genetics **169:** 197-214.

REDDY, K. D., E. G. ABRAHAM and J. NAGARAJU, 1999 Microsatellite in the silkworm, *Bombyx mori*: abundance, polymorphism, and strain characterization. Genome **42:** 1057-1065.

ROUSSELET, J., E. MAGNOUX and C. KERDELHUÉ, 2004 Characterization of five microsatellite loci in the Pine Processionary Moth *Thaumetopoea pityocampa* (Lepidoptera Notodontidae Thaumetopoeinae). Molecular Ecology Notes **4:** 213-214.

SCOTT, K. D., C. L. LANGE, L. J. SCOTT and G. C. GRAHAM, 2004 Isolation and characterization of microsatellite loci from *Helicoverpa armigera* Hübner (Lepidoptera: Noctuidae). Molecular Ecology Notes **4:** 204-205.

TAN, S., X. CHEN, A. ZHANG and D. LI, 2001 Isolation and characterization of DNA microsatellite from cotton bollworm (*Helicoverpa armigera*, Hübner). Molecular Ecology Notes **1:** 243-244.

VAN'T HOF, A. E., B. J. ZWAAN, I. J. SACCHERI, D. DALY, A. N. M. BOT *et al.*, 2005 Characterization of 28 microsatellite loci for the butterfly *Bicyclus anynana*. Mol. Ecol. Notes **5:** 169-172.

WARDILL, T. J., K. D. SCOTT, G. C. GRAHAM and M. P. ZALUCKI, 2004 Isolation and characterization of microsatellite loci from *Chiasmia assimilis* (Warren, 1899) (Lepidoptera: Geometridae). Molecular Ecology Notes **4:** 358-360.

WILLIAMS, B. L., J. D. BRAWN and K. N. PAIGE, 2002 Highly polymorphic microsatellite loci for *Speyeria idalia* (Lepidoptera : Nymphalidae). Mol. Ecol. Notes **2:** 87-88.

ZEISSET, I., T. D. ALS, J. SETTELE and J. J. BOOMSMA, 2005 Microsatellite markers for the large blue butterflies *Maculinea nausithous* and *Maculinea alcon* (Lepidoptera: Lycaenidae) and their amplification in other *Maculinea* species. Mol. Ecol. Notes **5:** 165-168.

ZHOU, Y., H. GU and S. DORN, 2005 Isolation of microsatellite loci in the codling moth, *Cydia pomonella* (Lepidoptera: Tortricidae). Mol. Ecol. Notes **5:** 226-227.

APPENDIX 3.2

Primers used in the experiment to confirm the presence of specific sequences in genomic DNA (Experiment I)

Clone	Primer designation in Fig. 2 & 3	Primer sequence (5'-3')	GenBank acc. #
BA-ATG1	1F	TTGGCCTAACCCCTCTCATTCTGAGC	AY785062
	1R	CGCGAGGTAGTCTGTGTGTTCCTAGC	
BA-ATG1	2F	CACAGACTACCTCGCGACAG	AY785062
	2R	CTGCAGTGGACGTCCATCGG	
BA-ATG108	3F	TGCTACGTGTGTTCGGTGCAT	DQ225280
	3R	CCGTCAAAAACGTCTATTGGC	
BA-ATG3	4F	GCAGCAAGCGACGACAAGGT	AY785064
	4R	CTGCAGTGGACGTCCATCGG	
BA-ATG212	5F	CATTAGCTTTGTGGCAACCTT	DQ225285
	5R	TGGCTCAGGATCGTGACGTTT	
BA-ATG212	6F	GCCGTATATGGGTTGATAAT	DQ225285
	6R	GGTTGCCACAAAGCTAATGA	
BA-ATG213	7F	TCAGCAGTGAGCCGAATATG	DQ225286
	7R	CCAATACTTTCCGGACTGTT	
BA-ATG215	8F	AGTGAGCCGTATATGGGTT	DQ225287
	8R	ATTTAGGTATTTGCGTACTCGT	
BA-ATG244	9F	GCTTCCTAACCCCAATCATT	DQ225299
	9R	TTGAGTTTCTTATCGGCTCT	
BA-ATG244	10F	CGATTCGGAAGGCAGGTCCT	DQ225299
	10R	GGACGTCCATCGGCTGATAT	
BA-ATG248	11F	GGGAATTCACAGCGCTTGAT	DQ225275
	11R	CTGCAGTGGACGTCCATCGG	
BA-ATG237	12F	TTCTAAGAGGAGACTCGAGC	DQ225295
	12R	CGCCTAGTTGGGACTACTT	
BA-ATG206	13F	GTGAGCCGAATATAGGTGA	DQ225283
	13R	ATCTTCCACGACTCGCTTCA	
BA-ATG027	14F	CAGGTGTAGTCAAGGGCTAA	DQ225276
	14R	GTCCATCGGCTGATAATG	
BA-ATG028	15F	CCTTGTGTTGACTTGCGGAAAC	DQ225277
	15R	GGCGTGGCAAAGGGTGTC	

APPENDIX 3.3

Primers used in the experiment to examine the proximity of homologues (experiment II).

Sequence family	Primer designation	Primer sequence (5'-3')
AAT group 1	AAT-gr1-upstr	TCAAATGGACACGCAACTTTACC
	AAT-gr1-downstr	AGCGCTTGACGAAGATAGTGTTAG
ATG group 2A upstr.A-dstr.A	ATG-gr2A-upstr	TCACTGTTGAGCTCGAGTCTCC
	ATG-gr2A-downstr	CTAAAGTAGTCCCAACTAGGC
ATG group 2B upstr.F-dstr.A	ATG-gr2B-upstr	CCGGTAGGGTGGTAACTAGCC
	ATG-gr2B-downstr	TCTTGCAAGGACTTCCTCGAGC
CA group 1	CA-gr1-upstr	TGCACTGCGGCTACTGA
	CA-gr1-downstr	GATAGCCCAGTGGATACGGA
CA group 2	CA-gr2-upstr	CGTGATAACAGCCCGCATTA
	CA-gr2-downstr	ATCCGCCGTGCAACCAC
CA group 3	CA-gr3-upstr	CTGGATTAACACATAGGCT
	CA-gr3-downstr	CAAAACCTTAAAATATACAGGT

APPENDIX 3.4

List of species with relevant Blast hits matching sequences from the *Bicyclus anynana* enriched libraries.

Species	Insect group	GenBank accession #
Antheraea mylitta	Lepidoptera / Moth	AF530471
Arctia caja	Lepidoptera / Moth	AJ809352; AJ809356; AJ809371; AJ809378; AJ809379; AJ809380; AJ867352; AJ867362; AJ867383; AJ867384
Bicyclus anynana[1]	Lepidoptera / Butterfly	AY766157
Biston betularia	Lepidoptera / Moth	AY190966; AY190967; AY190974; AY485266
Bombyx mori	Lepidoptera / Moth	AF226688; AB014342; AB023085; AB023115; AB032718; AB035269; AB048355; AB052774; AB052773; AB063490; AB080675; AB090307; AB090308; AB101293; AB104488; AB126052; AB159445; AB159446; AB159447; AF541967; AY083677; AY172027; D10742; D12523; D16230; D16233; D66906; D78138; D86623; M24370; J04829; M76430; X04226; X02223; Z14101; Z14101; Z15048; Z15048
Bombyx mandarina	Lepidoptera / Moth	AY172028; AY172028
Chiasmia assimilis	Lepidoptera / Moth	AY552796
Choristoneura fumiferana	Lepidoptera / Moth	CFU12917
Choristoneura murinana	Lepidoptera / Moth	AF177645; AF177646; AF177647; AF177662
Coenonympha hero	Lepidoptera / Butterfly	AF499099
Cydia pomonella	Lepidoptera / Moth	AY700111
Euphydryas aurinia	Lepidoptera / Butterfly	AY491786; AY491815; AY491833; AY491848; AY491849
Galleria mellonella	Lepidoptera / Moth	M73793; L22534; M73793
Heliconius cydno	Lepidoptera / Butterfly	AY429264
Heliconius melpomene	Lepidoptera / Butterfly	AY429262; AY429263
Helicoverpa armigera	Lepidoptera / Moth	AF271059; AF492474; AJ504787; AJ627416; AY382615; AY497338; AY714875; AY714876
Helicoverpa zea	Lepidoptera / Moth	M80588
Heliothis virescens	Lepidoptera / Moth	AF072458
Hyalophora cecropia	Lepidoptera / Moth	L13971; M60914; M63846
Hyphantria cunea	Lepidoptera / Moth	U86877
Lymantria dispar	Lepidoptera / Moth	AF004228; AF198385
Manduca sexta	Lepidoptera / Moth	AF527635; AF527636; AY789465; U03989
Ostrinia nubilalis	Lepidoptera / Moth	U04223
Papilio helenus	Lepidoptera / Butterfly	AB013152
Papilio xuthus	Lepidoptera / Butterfly	AB182634
Pararge xiphia	Lepidoptera / Butterfly	AF214612
Parnassius apollo	Lepidoptera / Butterfly	AY491896; AY491940
Plutella xylostella	Lepidoptera / Moth	AY696174; AY696175
Reissita simonyi	Lepidoptera / Moth	AY250748
Saucrobotys futilalis	Lepidoptera / Moth	AY497537; AY497538
Utetheisa ornatrix	Lepidoptera / Moth	AY603695
Zale galbanata	Lepidoptera / Moth	AF484812
Anastrepha suspensa	Diptera / Caribbean fruit fly	AY520439
Anopheles gambiae	Diptera / Malaria mosquito	XM_308573
Diabrotica virgifera	Coleoptera / Corn rootworm	AY738541
Apis mellifera	Hymenoptera / Honey bee	XM_397014
Pheidole pallidula	Hymenoptera / Mediterranean ant	AF426753

[1]Inter- or intra library hits of *Bicyclus anynana* are not included in this table. *Bicyclus anynana* AY766157 is an independently deposited sequence.

APPENDIX 3.5

Lepidoptera Specific Core Sequences (LSCS)

LSCS1

ATCATCAGCCTATAGCAGTCCACTGCTGGACATAGGCCTCTCCA

LSCS2

ATGTGCAGGTTTCCTCACGATGTTTTCCTTCACCG

LSCS3

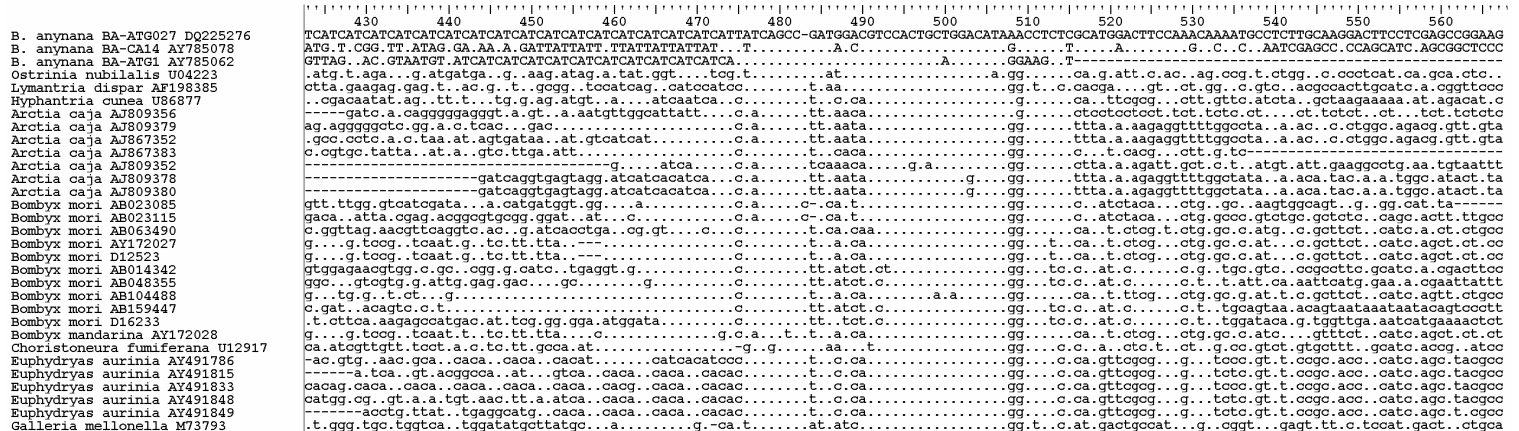
AAAATTTAGCCTATGTTACTCGGGAATAGTGTAGCTTTCCAACAGTGAAA
GAATTTTTCAAATCGGTTCAGTAGTTTCTGAGCCTATTCATTA-$CAAA_n$-
TCTTTCCTCTTTATAATATTAGTATAGAT

LSCS4

TTATTGGAACGAAGTTCCTTATCGCGCGTTGCGAAAGGGGGCTAGACGGA
AAAAATTAAGACCAAAAGTTGTCACGACACTTTTT

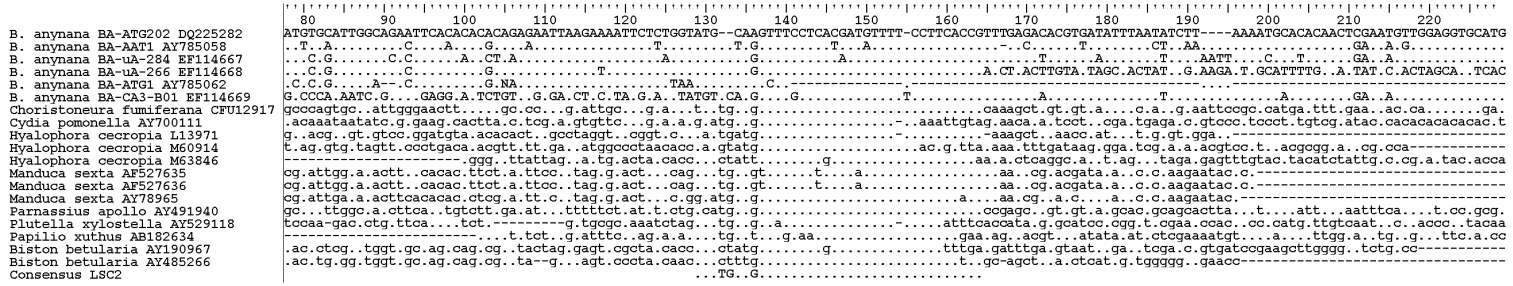
Appendix 3.6

LSCS1 alignment



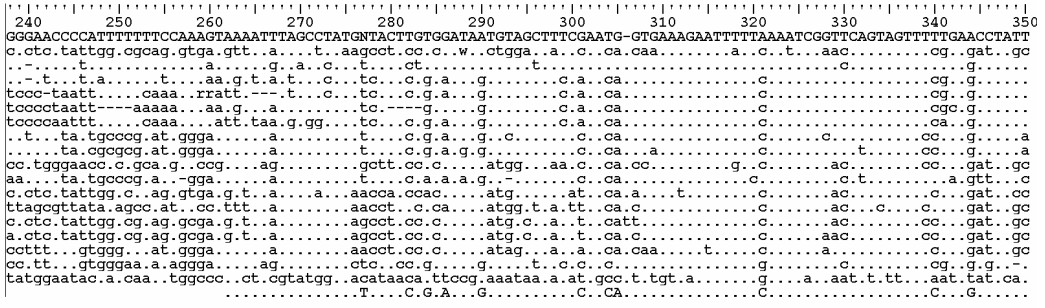
Appendix 3.7

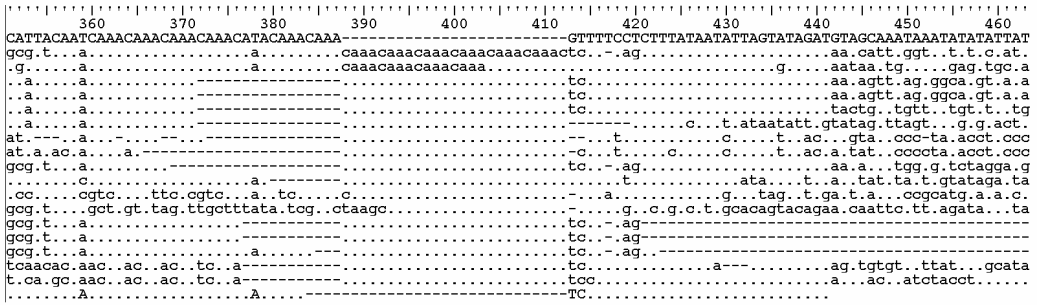
LSCS2 alignment



Appendix 3.8

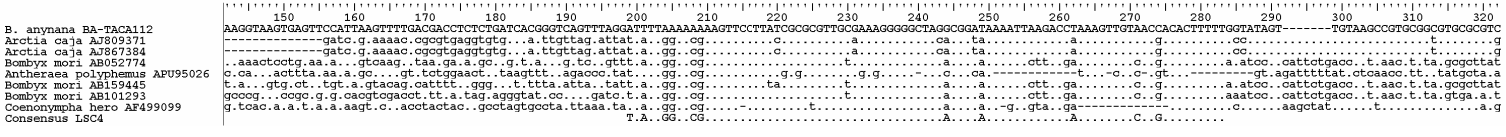
LSCS3 alignment





Appendix 3.9

LSCS4 alignment

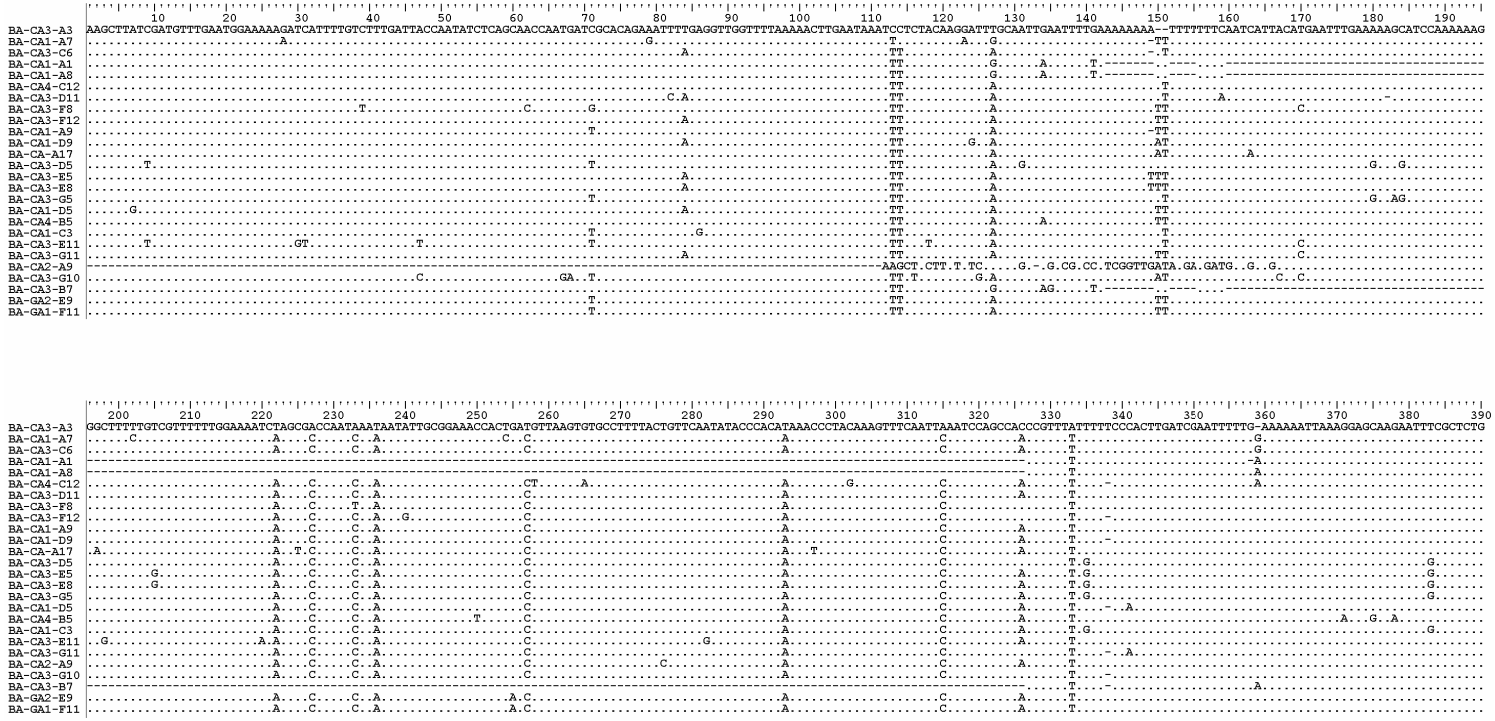


APPENDIX 3.6 – 3.9 DETAILS

Lepidoptera specific core sequence (LSCS) 1-4 alignments based on iterative blastn hits. The top row is the *Bicyclus anynana* sequence that was used for the initial blast search against the NCBI database. Below are the blast hits aligned against the *B. anynana* sequence with species and GenBank accession number specified. A dot (.) indicates a nucleotide that is identical to the top row, a dash (-) indicates a gap (indel). The bottom rows contain the aligned LSCS. Only the middle section of the alignment that includes the conserved region is shown.

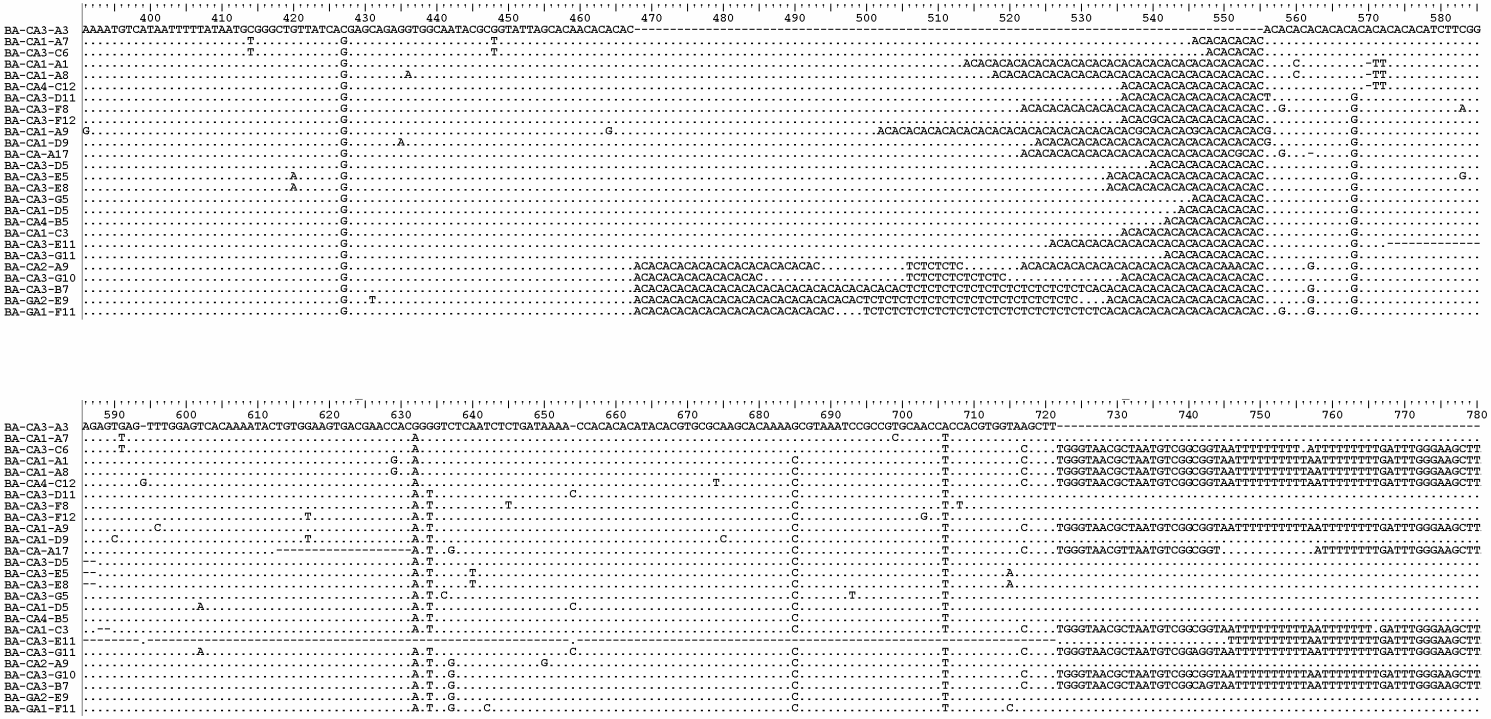
Appendix 3.10A

CA group 2 alignment first half



Appendix 3.10B

CA group 2 alignment second half



APPENDIX 3.10 DETAILS

Alignment of *Bicyclus anynana* sequences showing various degrees of similarity in flanks and microsatellites in different loci. The microsatellites vary both in repeat number and in composition, with some being compound CA+CT while others have exclusively CA repeats. The bottom two sequences originate from the GA library, wile the rest were in the CA library. Dots and dashes as in the previous alignments.

APPENDIX 3.11

LSCS3 is potentially incorporated in *B. anynana* by horizontal gene transfer.

After the contents of this chapter were published in Heredity, a number of bracovirus sequences have become available in GenBank (e.g. acc # EF710635.1 & EF710642.1) that shed new light on one of the Lepidoptera specific core sequences (LSCS). These bracovirus sequences have a nearly-perfect match with LSCS3 (blast results in Fig 3.6). The homologous region includes a $CAAA_6$ repeat that fully coincides with the $CAAA_6$ in LSCS3.

Bracoviruses are double-stranded DNA viruses incorporated in the genomes of parasitic wasps as pro-viral sequences (DREZEN *et al.* 2003) that act as mutualistic endosymbionts. They enhance parasitoid survival in caterpillars, by suppressing the host's immune system (DESJARDINS *et al.* 2008; DESJARDINS *et al.* 2007; FLEMING 1991; LAVINE and BECKAGE 1995; WYDER *et al.* 2003).

The sequence found in *B. anynana* cannot be a contaminant of the enriched library by a parasite because (*i*) bracoviruses do not replicate in host tissue (*ii*) adult material was used for library construction and (*iii*) the adult came from a laboratory stock that has been maintained in a controlled, parasitic wasp free environment for many years. Therefore it is most likely part of the *B. anynana* genome, which would indicate horizontal gene transfer from parasitic wasp to the butterfly. In-situ hybridization on chromosomes confirmed the incorporation of these viruses in the wasp genomes (BELLE *et al.* 2002), and this technique could also be used to demonstrate genomic incorporation in Lepidoptera.

```
>gb|EF710635.1| Cotesia sesamiae Kitale bracovirus clone BAC 2004, complete sequence
Length=113692

 Score =  211 bits (114),  Expect = 8e-52
 Identities = 140/152 (92%), Gaps = 4/152 (2%)
 Strand=Plus/Plus

Query  3       AATTTAGCCTATGTTACTCGGGAATAGTGTAGCTTTCCAACAGTGAAAGAATTTTTCAAA  62
               ||| |||||||||| ||||||| ||||||||| ||| ||||||||||||||||||||||
Sbjct  103395  AATATAGCCTATGTCACTCGGGGATAGTGTAGGTTTTCAACAGTGAAAGAATTTTTCAAA  103454

Query  63      TCGGTTCAGTAGTTTCTGAGCCTATTC-AT---TACAAACAAACAAACAAACAAACAAAC  118
               ||||||||||||||| |||||||||| ||    |||||||||||||||||||||||| |
Sbjct  103455  TCGGTTCAGTAGTTTCGGAGCCTATTCAATGCAAACAAACAAACAAACAAACAAATC  103514

Query  119     AAATCTTTCCTCTTTATAATATTAGTATAGAT  150
               ||||||||||||||||||||||||||||||||
Sbjct  103515  AAATCTTTCCTCTTTATAATATTAGTATAGAT  103546


>gb|EF710642.1| Cotesia sesamiae Mombasa bracovirus clone BAC 14G12, complete
sequence
Length=104039

 Score =  196 bits (106),  Expect = 2e-47
 Identities = 135/148 (91%), Gaps = 5/148 (3%)
 Strand=Plus/Minus

Query  3      AATTTAGCCTATGTTACTCGGGAATAGTGTAGCTTTCCAACAGTGAAAGAATTTTTCAAA  62
              ||| |||||||||| ||||||| ||||||||| ||| |||||||||||||||||||||||
Sbjct  88348  AATATAGCCTATGTCACTCGGGGATAGTGTAGGTTTTCAACAGTGAAAGAATTTTTCAAA  88289

Query  63     TCGGTTCAGTAGTTTCTGAGCCTATTCATTACAAACAAACAAACAAACAAACAAA--  120
              |||||||||||||||| |||||||||| | |||||||||||||||||||||||||||
Sbjct  88288  TCGGTTCAGTAGTTTCGGAGCCTATTCAATGCAAACAAACAAACAAACAAACAATC  88229

Query  121    A--TCTTTCCTCTTTATAATATTAGTAT  146
              |  ||||||||||||||||||| |||
Sbjct  88228  AAATCTTTCCTCTTTATAATATTA-TAT  88202


>gb|AC239123.1| Bicyclus anynana clone BA_Ba84B11, complete sequence
Length=129178

 Score =  189 bits (102),  Expect = 4e-45
 Identities = 136/152 (89%), Gaps = 4/152 (2%)
 Strand=Plus/Plus

Query  1       AAAATTTAGCCTATGTTACTCG-GGAATAGTGTAGCTTTCCAACAGTGAAAGAATTTTTC  59
               ||||||||||||||||||||| | || ||| |||||||||| ||  |||||||| |||||
Sbjct  111795  AAAATTTAGCCTATGTTACTTGTGG-ATAATGTAGCTTTCGAATGGTGAAAGAGTTTTTA  111853

Query  60      AAATCGGTTCAGTAGTTTCTGAGCCTATTCATTACAAACAAACAAACAAACAAACAAACA  119
               |||||||| ||| |||||| ||||||||||||||||||||||||||||||||||| |||||||
Sbjct  111854  AAATCGGTCCAGCAGTTTTTGAGCCTATTCATTACAAACAAACAAACAAACATACAAACA  111913

Query  120     AA-TCTTTCCTCTTTATAATATTAGTATAGAT  150
               || | |||| ||||||||||||||||||||||
Sbjct  111914  AAGTTTTTC-TCTTTATAATATTAGTATAGAT  111944


>emb|FP340425.1| 64G08_HaBAC_fin, Helicoverpa armigera BAC, pupae DNA
Length=103563

 Score =  182 bits (98),  Expect = 6e-43
 Identities = 138/156 (88%), Gaps = 8/156 (5%)
 Strand=Plus/Plus

Query  1     AAAATTTAGCCTATGTTACTCG-GGAATAGTGTAGCTTTCCAACAGTGAAAGAATTTTTC  59
             ||||| |||||||||||||||| || ||| |||||||||| ||  |||||||||| |||
Sbjct  4767  AAAATATAGCCTATGTTACTCGTGG-ATAATGTAGCTTTCGAATGGTGAAAGAATGTTTA  4825

Query  60    AAATCGGTTCAGTAGTTTCTGAGCCTATTCATT---A-CAAACAAACAAACAAACAAACA  115
             ||| |||| |||||||| ||||||||||||||   | ||||||||||||||||||||||
Sbjct  4826  AAAACGGTCCAGTAGTTTTTGAGCCTATTCATTACAACCAAACAAACAAACAAACAAACA  4885

Query  116   AACAAA-TCTTTCCTCTTTATAATATTAGTATAGAT  150
             |||||| | |||||||||||||||||||||||||||
Sbjct  4886  AACAAAGT-TTTCCTCTTTATAATATTAGTATAGAT  4920
```

Figure 3.6 The first four NCBI blastn hits returned for LSCS3. Low complexity filtering was disabled to include the CAAA repeat in the alignment. The two bracoviruses have higher e-values than the two Lepidoptera hits that follow, even though LSCS3 was based on lepidopteran sequences. The full list of blast results consists exclusively of butterfly, moth and bracovirus sequences.

APPENDIX 3.11 REFERENCES

BELLE, E., N. BECKAGE, J. ROUSSELET, M. POIRIE, F. LEMEUNIER *et al.*, 2002 Visualization of polydnavirus sequences in a parasitoid wasp chromosome. J Virol **76:** 5793 - 5796.

DESJARDINS, C., D. GUNDERSEN-RINDAL, J. HOSTETLER, L. TALLON, D. FADROSH *et al.*, 2008 Comparative genomics of mutualistic viruses of Glyptapanteles parasitic wasps. Genome Biology **9:** R183.

DESJARDINS, C., D. GUNDERSEN-RINDAL, J. HOSTETLER, L. TALLON, R. FUESTER *et al.*, 2007 Structure and evolution of a proviral locus of Glyptapanteles indiensis bracovirus. BMC Microbiol **7:** 61.

DREZEN, J. M., B. PROVOST, E. ESPAGNE, L. CATTOLICO, C. DUPUY *et al.*, 2003 Polydnavirus genome: integrated vs. free virus. Journal of Insect Physiology **49:** 407-417.

FLEMING, J., 1991 The integration of polydnavirus genomes in parasitoid genomes: implications for biocontrol and genetic analyses of parasitoid wasps. Biol Control **1:** 127 - 135.

LAVINE, M., and N. BECKAGE, 1995 Polydnaviruses: potent mediators of host insect immune dysfunction. Parasitol Today **11:** 368 - 378.

WYDER, S., F. BLANK and B. LANZREIN, 2003 Fate of polydnavirus DNA of the egg-larval parasitoid Chelonus inanitus in the host Spodoptera littoralis. J Insect Physiol **49:** 491 - 500.