Universiteit
Leiden
The Netherlands

Towards understanding the architecture of the Bicyclus anynana genome

Hof, Arjèn Emiel van 't

Citation
Hof, A. E. van 't. (2011, June 23). *Towards understanding the architecture of the Bicyclus anynana genome*. Faculty of Science, Leiden University. Retrieved from https://hdl.handle.net/1887/17726

Version: Not Applicable (or Unknown)
License:
Downloaded from: https://hdl.handle.net/1887/17726

Note: To cite this publication please use the final published version (if applicable).

Towards understanding the architecture of the *Bicyclus anynana* genome

Arjèn van 't Hof

Towards understanding the architecture of the *Bicyclus anynana* genome

PROEFSCHRIFT

ter verkrijging van de graad van Doctor aan de Universiteit Leiden,

op gezag van de Rector Magnificus Prof. Mr. P.F. van der Heijden,

volgens besluit van het College voor Promoties

te verdedigen op donderdag 23 juni 2011

klokke 13.45 uur

door

Arjèn Emiel van 't Hof
geboren te Numansdorp in 1973

PROMOTIECOMMISSIE:

Promotors:	Prof. Dr. Paul Brakefield Prof. Dr. Bas Zwaan
Overige leden:	Prof. Dr. Carel J. ten Cate Dr. Patrícia Beldade Dr. Chris D. Jiggins Prof. Dr. František Marec Dr. Ilik J. Saccheri Prof. Dr. Peter G.L. Klinkhamer

TABLE OF CONTENTS

General introduction 1

Characterization of 28 microsatellite loci for the butterfly *Bicyclus anynana* 15

Evolutionary dynamics of multi-locus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera 24

Cytogenetic characterization of the butterfly *Bicyclus anynana* 61

An AFLP-based genetic linkage map for the butterfly *Bicyclus anynana*, covering all 28 chromosomes 72

Quantitative Trait Loci affecting the size of wing eyespots in a seasonally polyphenic butterfly 105

Dutch summary / Nederlandse samenvatting 135

Curriculum vitae 142

Acknowledgements 144

Chapter 1

General introduction

Evolutionary research: beyond case studies

Raising our understanding of processes involved in evolution to new levels is the primary goal in evolutionary biology research. The main focus should therefore be on how traits evolve within or between species rather than merely unravelling species-specific characteristics that would only appeal to a niche audience. From a practical point of view however, there is often no escape in examining a species into great detail prior to addressing more universal questions. Comprehensive species-specific knowledge is essential for robust experimental design, development of research tools, analysis and interpretation of results, and comparison with other species. The importance of building a foundation to facilitate research is illustrated by the excitement about the completion of the human genome sequence (MACILWAIN 2000). This excitement was not at all caused by the result itself (a dull near-endless stretch composed of four different nucleotides), but rather by what it could potentially be used for. A substantial part of this thesis deals with tool development and description of the genomic architecture of the afrotropical butterfly *Bicyclus anynana* to provide the basis to address specific current questions, but also to aid future research involving this emerging model organism. Treating the work presented here as merely a species-specific characterization would not do justice to the achievements though. Many of the findings fit into a broader context and the chapters that have been published so far (chapters 2-5) have been cited frequently, not restricted to insect-based publications, which implies a degree of more universal significance. Expanding the knowledge of a species also allows comparative analysis which can be used to test generalised rules. This can either strengthen these rules by confirmation, weaken them by demonstrating an exception to a rule or it can reveal that commonly accepted principles should perhaps be reconsidered. Examples of these three options found in *B. anynana* (this thesis) are the confirmation that meiotic recombination is absent in females, the unusually small W-chromosome, and the contribution to a growing body of evidence that microsatellites often do not match their defined characteristics.

Seasonal polyphenism in *Bicyclus anynana*: the rise of a model organism

B. anynana initially attracted scientific attention because it displays striking seasonal polyphenism (BRAKEFIELD *et al.* 1996), with distinct morphological features during different seasons (SHAPIRO 1976). It has developed into one of the most intensively investigated species in this field, partly because it can be maintained and studied in a controlled laboratory environment and has a relatively short generation time. The dry season adult form of *B. anynana* is cryptic against dry leaves with very small eyespots to avoid visual predation. The wet season form has a far more conspicuous appearance with a row of eyespots along the ventral wing margins; they usually rest with their wings closed above the body displaying the ventral wing surfaces. Naïve birds sometimes misdirect their attacks towards these conspicuous eyespot 'targets'. This gives the butterfly an opportunity to escape although perhaps

with the sacrifice of a small section of the marginal wing tissue should the wings be grasped by the predator (LYYTINEN *et al.* 2004). The morphological differences are triggered by ambient temperature during the last larval instar (Fig. 1.1), with temperatures around 17°C resulting in the dry season form and temperatures

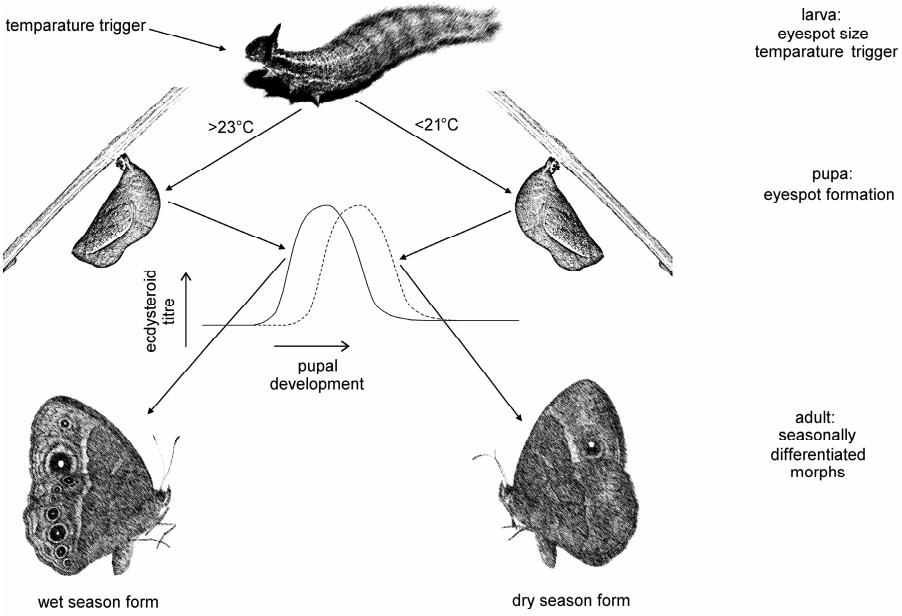


Figure 1.1 Key events that lead to the differentiation in ventral eyespot sizes.
Ambient temperature during the late 5th larval instar triggers the eyespot size (image shows a 4th instar larva). A temperature below 21°C causes a late ecdysteroid peak in the pupa, which leads to a dry season form adult with small eyespots (right) and a temperature above 23°C results in an early ecdysteroid peak, producing a wet season form adult (left). The ecdysteroid timing curve is a schematic based on Fig. 4 in OOSTRA *et al.* (2010).

above 23°C producing the wet season form (BRAKEFIELD *et al.* 1996). Intermediate temperatures produce intermediate eyespot sizes following a reaction norm (WIJNGAARDEN and BRAKEFIELD 2001), which is strongly associated with the timing of ecdysteroid peaks in the early pupa (OOSTRA *et al.* 2010).
The developmental pathway of the eyespots and their different components have been thoroughly studied on a molecular level, however much of this attention has been directed towards eyespot formation rather than eyespot size variation. Exceptions are a study that revealed an effect of *Distal-less* (or a tightly linked unknown gene) on dorsal forewing eyespot size (BELDADE *et al.* 2002) and a number of experiments that linked the effect of ecdysteroid concentrations and timing thereof to the observed size variation (KOCH *et al.* 1996; OOSTRA *et al.* 2010; ZIJLSTRA *et al.* 2004). A comprehensive description of how eyespot sizes are determined remains to be uncovered and will require additional investigations.

Bicyclus anyana beyond seasonal polyphenism: the establishment as a model organism

Research using *B. anynana* as study organism has expanded beyond seasonal polyphenism into fields such as evolution and development (evo-devo) life history evolution, ageing, functional genomics, population genetics, developmental modularity, pheromone signalling and comparative genomics (BELDADE and BRAKEFIELD 2002; BELDADE *et al.* 2005; BELDADE *et al.* 2009a; BRAKEFIELD *et al.* 2009; FISCHER *et al.* 2003; MARCUS *et al.* 2004; MONTEIRO *et al.* 2007; NIEBERDING *et al.* 2008; PIJPE *et al.* 2008; SACCHERI and BRAKEFIELD 2002; ZIJLSTRA *et al.* 2003). This thesis builds further on ongoing research involving seasonal polyphenism, but also extends beyond the current fields of investigation by addressing chromosome and genome evolution and architecture. Life history evolution has also been addressed as part of the PhD research by examining crosses between lines with fast and slow developing larvae. These experiments were set up on such an enthusiastic scale that the caterpillars depleted all available maize plants before pupation and were supplied with intermittent batches of food subsequently. A continuous food supply is essential to produce reliable developmental time data and thus it came as no surprise that the results did not show a sensible pattern. Therefore the fast-slow experiments were not transformed into a scientific publication or thesis chapter.

The special characteristics of Lepidopteran genomes

Lepidoptera (butterflies and moths) possess a number of characteristics that are different from what is generally observed in other organisms. Most of these uncommon features are not unique to Lepidoptera, it is however unusual that they are concentrated within one taxon. Lepidoptera have an uncommon sex determination system, with females being the heterogametic sex. This is only known in a few groups, such as in sister taxon Trichoptera (caddis flies) (TRAUT *et al.* 2007) and the birds. Instead of XX females and XY males, Lepidoptera have ZZ males, and WZ (or variations thereof) females (TRAUT *et al.* 2007). Lepidoptera chromosomes lack a distinct primary constriction (the centromere), but have a holokinetic arrangement that results in full-length alignment of sister chromatids during mitotic metaphase followed by their parallel disjunction. Besides Lepidoptera and their sister group Trichoptera, this is only known in a limited number of taxa, including hemipteran insects, some spiders, mites and nematodes. (RODRÍGUEZ GIL *et al.* 2002; WRENSCH *et al.* 1994) and nematodes (MADDOX *et al.* 2004). Centromere positions are important landmarks for identification of chromosomes, and their absence in Lepidoptera makes the chromosomes less distinct. Moreover, diagnostic banding patterns or discrete size distributions often do not occur in Lepidoptera, which usually prevents visual identification of chromosomes. The chromosome number in Lepidoptera has the highest variability and the highest total number encountered in animals (WHITE 1973). Observed chromosome numbers vary between n = 7 in *Erebia aethiopellus* and n = 217-223 in *Plebicula atlanticus* (WHITE 1973) and the rate in which these differences are generated are sometimes exceptionally high, with large differences between closely related species or even between different populations of the same species (DE LESSE 1963; KANDUL *et al.* 2007; SUOMALAINEN and BROWN 1984). These examples

illustrate extreme cases that are restricted to a few lepidopteran families though. The general chromosome counts are concentrated around 29-31 (WHITE 1973), with 31 considered the ancestral number (ROBINSON 1971).

Orthologous genomic regions in Lepidoptera are generally highly conserved in gene composition and order (SAHARA *et al.* 2007; YASUKOCHI *et al.* 2009), even between the relatively distantly related butterflies and moths (BELDADE *et al.* 2009a; PAPA *et al.* 2008; PRINGLE *et al.* 2007; YASUKOCHI *et al.* 2006). An undisturbed gene arrangement relative to a common ancestor is referred to as 'shared synteny' (also named 'conserved synteny') and is usually divided into two categories based on chromosomal scale with macrosynteny describing large chromosomal regions and microsynteny dealing with clusters of neighbouring genes. A multi-species macrosynteny comparison is currently difficult since there is little overlap between the annotated positions in Lepidopteran linkage maps. Therefore, the genome assembly of the silkworm (*Bombyx mori*) (XIA *et al.* 2004) is central to most shared synteny studies in Lepidoptera to date. An exception where two species other than *B. mori* are compared is a study where fluorescenty labelled Bacterial Artificial Chromosomes (BACs) derived from *Manduca sexta* are hybridised (BAC-FISH) to both *Manduca sexta* itself and *Agrius convolvuli*, which also showed a high degree of conserved macrosynteny (YASUKOCHI *et al.* 2009). Microsynteny is usually based on the comparison of BAC sequences with the *B. mori* genome assembly and this revealed high levels of conservation between *Heliconius* and *B. mori* (PAPA *et al.* 2008), however, two Noctuid moths revealed very small conserved synteny blocks relative to the *B. mori* reference genome (D'ALENÇON *et al.* 2010).

Meiotic recombination does not occur in those Lepidoptera females studied to date, with the exception of *Colias* butterflies (CARTER and WATT 1988). Complete absence of recombination in one of the sexes is very rare in other groups, with *Drosophila* males being one of the few examples outside Lepidoptera (MORGAN 1912). Another characteristic is the high proportion of transposable elements in Lepidoptera genomes (EICKBUSH 1995; INTERNATIONAL SILKWORM GENOME CONSORTIUM 2008) with extreme concentrations on the W-chromosome reported in *B. mori* (ABE *et al.* 2005). These characteristics have implications for experimental design, analysis and interpretation, which makes Lepidoptera research a specialist subject. The combined experiments described in this thesis were confronted with each of the unusual Lepidoptera features listed above.

Outline of this thesis

The work presented in this thesis uses a number of molecular-genetic and cytogenetic techniques to describe the general genomic architecture of *B. anynana*. This description consists of a cytological characterisation, a survey of the origin and replication methods of repetitive sequences, and the construction of a linkage map. The acquired knowledge can be used by the *Bicyclus* and Leptidopera communities to develop tools and experimental procedures as is illustrated by the analysis of quantitative genetic analysis variation for eyespot size. The sections that follow will deal with the different experiments in more detail and explain how these fit into a broader context.

Obtaining polymorphic microsatellites for *Bicyclus anynana*

The second chapter describes the development of microsatellite markers for *B. anynana*. Microsatellites are markers that are commonly used in population genetic, parentage, and mapping studies. They are often favoured over alternative genetic markers because they are presented as codominant single copy markers with conserved flanking regions that can potentially produce a large number of alleles per locus, and can be genotyped at relatively low cost when multiplexed. They were, therefore, considered the most suitable markers to use as annotated anchors for subsequent mapping studies in *B. anynana*.

The construction of microsatellite enriched libraries was handed over to a specialized company, following rumours of difficulties involved with microsatellite development in Lepidoptera. It became clear however that library construction is not the bottleneck for producing robust polymorphic microsatellite markers. The library suffered from extremely abundant common sequence fragments which made it impossible to produce distinct single-locus microsatellites. A substantial amount of microsatellites were required to cover the genome sufficiently, given the relatively large number of chromosomes in Lepidoptera. This warranted additional efforts to produce a decent panel of microsatellites, and a special screening technique was designed to avoid useless sequences. This resulted in a reasonable number of microsatellites with unique flanks. Some of these refused to amplify with polymerase chain reaction (PCR) in spite of exhaustive optimisation attempts including wide ranges of $MgCl_2$ concentrations and annealing temperatures. A blast search revealed that these sequences were identical to those of the Central American tungara frog (*Physalaemus pustulosus*) for which an enriched microsatellite library was constructed at the same time by the same company as our *B. anynana* library (Prof. Heike Pröhl, pers comm[1]). Therefore, I have established that tungara frog microsatellites do not cross-amplify in *B. anynana* (unpublished results). The remaining microsatellites produced a for lepidoptera exceptional number of 28 polymorphic loci. Unfortunately, many of the microsatellites suffered from heterozygote deficiency caused by PCR failure of certain alleles. Such null-alleles (a.k.a. allele dropout) are extremely common in Lepidoptera microsatellites (MEGLECZ *et al.* 2004), which makes them hardly useful for population genetic analysis. However, they can still be used when pedigrees are known (e.g. in a linkage mapping family) if the null alleles follow Mendelian segregation. The low proportion of unique microsatellites and the high numbers of null-alleles make these markers far less attractive in Lepidoptera than they are generally presented.

The 28 markers described in this thesis were later supplemented with 40 additional polymorphic microsatellite loci obtained from an expressed sequence tag (EST) library (BELDADE *et al.* 2009b). These combined loci provide one of the highest total numbers of microsatellites within Lepidoptera for *B. anynana*.

[1] Prof. Heike Pröhl used the same microsatellite enrichment company as I did and is not in any way responsible for contaminating my samples.

Repetitive sequences associated with microsatellites in *Bicyclus anynana* and Lepidoptera in general.

The third chapter elaborates on those microsatellites that have flanks which are present in multiple copies throughout the genome. This phenomenon has been frequently observed in Lepidoptera, but was mainly ignored as an interesting genomic trait (MEGLECZ *et al.* 2004; ZHANG 2004). One of the distinct difficulties in describing and interpreting these observations is that they are the result of dynamics over a (potentially) very long period of time. Sequences have been duplicated many times and diverged independently and all that can be analysed is the end result rather than the separate events. The sequences do still hold information though, which allows a number of questions to be addressed. The main issue investigated is the duplication mechanism that generated the multiple copies, which reveals that a combination of unequal recombination and replication of mobile elements describe the data most convincingly. The association between microsatellites and repetitive DNA is also explored. Transposable elements that generate microsatellites are known in other taxa and were recently also found in Lepidoptera (ARCOT *et al.* 1995; TAY *et al.* 2010; WILDER and HOLLOCHER 2001), but they do not account for the microsatellites described in *B. anynana*. Instead, the microsatellites are duplicated together with the flanking regions in *B. anynana*. Some of the over-represented sequences are not unique for *B. anynana*, but are also found in many other Lepidoptera species. One of these multi-copy sequences that are common in Lepidoptera corresponds with a bracovirus that is hosted as a mutualistic endosymbiont by parasitic wasps which lay their eggs in lepidopteran larvae. The bracovirus is incorporated in the parasitic wasp genome (BELLE *et al.* 2002; DREZEN *et al.* 2003) and released to inhibit the immune system of the parasitized caterpillar to secure the development of the wasp larvae (LAVINE and BECKAGE 1995) Finding this sequence in the genome of a butterfly indicates horizontal gene transfer from parasitic wasp to butterfly through a viral intermediate. A plausible explanation would be that some parasitized caterpillars managed to survive the invasion and developed into adults. The transferred sequence must have settled in the germ line to be maintained within the genome over multiple generations.

Cytogenetic characterisation of *Bicyclus anyana* chromosomes

Chapter four presents a cytogenetic characterization of *B. anynana*, which explores the sex chromosome makeup, the chromosome number and additional general karyotype features. Apart from increasing our general knowledge about *B. anynana*, this information is important because it can act as reference to determine the achieved coverage of linkage mapping (chapter 5). A publication specifying a haploid chromosome number of n = 28 in *B. anynana* (DE LESSE 1968) was initially overlooked because it did not occur in any electronic database at the time and was written in French, which makes it less accessible to the English speaking scientific community (chromosome numbers of n = 26 and n = 13 have been assumed previously (BELDADE *et al.* 2002; WIJNGAARDEN and BRAKEFIELD 2000)). Establishing the chromosome number in our stock population was still necessary though given the frequently observed geographical within-species karyotype variation in Lepidoptera (DE LESSE 1963; KANDUL *et al.* 2007; SUOMALAINEN and BROWN 1984). The n = 28 chromosome number was confirmed by our experiments based on

different meiotic stages in testes and ovaries. The sex chromosome composition of *B. anynana* was consistent with the most commonly observed lepidopteran system with WZ females and ZZ males. However, the W chromosome is extremely small in *B. anynana*, which results in an unusual synaptonemal complex during the pachytene stage with the Z-chromosome forming a circular or horseshoe shaped structure enclosing the W-chromosome. The small W-chromosome size was also observed in polyploid interphase nuclei of Malpighian tubules. This observation of an unusually small W-chromosome is important in the context of sex chromosome evolution because some Lepidoptera have a derived Z-♀, ZZ-♂ sex determination system (TRAUT *et al.* 2007). The observation in *B. anynana* therefore suggests that gradual erosion rather than an abrupt event may be responsible for losing the W-chromosome. The chromosomes of *B. anynana* lack a centromere during cell division and chromatids are paired over their entire length instead. Such a holokinetic (a.k.a. holocentric) association is common to Lepidoptera. Two active nucleolar organising regions (NORs) were observed in *B. anynana*, which provides a form of visual differentiation in the absence of other distinct features. Apart from adding visual characteristics to chromosomes within a species, NORs can also serve as landmarks in chromosome evolution, and possibly act as ectopic recombination sites that mediate chromosome restructuring (NGUYEN *et al.* 2010). One of the nucleoli is terminally positioned on an anonymous autosome and has two small but conspicuous heterochromatin sections associated with the NOR, which is uncommon in Lepidoptera. The second nucleolus has an unknown chromosomal position and is associated with one of the sex chromosomes, presumably the Z-chromosome given the transcriptional activity of the NOR.

An AFLP-based genetic linkage map for *B. anynana*

The construction of a genetic linkage map for *B. anynana* is described in Chapter 5. A Linkage map is a representation of the genome by genetic or morphological markers that ideally cover every chromosome with regular intervals along their entire length. The two essential components to produce a linkage map are meiotic recombination and polymorphisms. Polymorphisms reveal the segregation of loci from parents to offspring, and those loci that show a similar pattern of segregation are considered linked to each other (i.e. on the same chromosome). Recombination disrupts complete co-segregation of markers on the same chromosome and the degree in which this co-segregation differs between two loci is a measure for the distance that separates them, which is specified in centimorgans or map units. The total number of recombination events in a mapping family is proportional to the number of offspring, thus mapping resolution is strongly dependent on the sample size used. Linkage maps are important, if not essential, to localize genes that underpin specific phenotypic traits and their variation. A genetically controlled polymorphic phenotype will co-segregate with markers on the same chromosome and its inheritance pattern will be most similar to the markers nearest to it. This often provides the first step towards identifying the gene responsible for a certain phenotype. Identification of genes is not the only use of linkage maps though, especially when more than one locus is involved in a trait. Firstly, combinations of segregating loci that affect a phenotype can reveal the effects of gene interactions such as epistasis. Secondly, linkage maps can also reveal whether independent Mendelian segregation can be expected for two (or more) genes, depending on their degree of linkage. This information is important because

closely-linked genes with antagonistic effects respond very differently from unlinked genes when under selection. Information about the relative position of genes can also be extremely relevant because functionally-related genes are sometimes clustered as 'supergenes'. Such a supergene was found to regulate mimetic colour polymorphisms in *Heliconius numata* (JORON *et al.* 2006). Thus linkage maps can help distinguish pleiotropy from close genetic linkage.

The absence of recombination in female Lepidoptera has severe implications for linkage mapping. The segregation patterns of paternal and maternal polymorphisms cannot be analysed together and the maternally inherited portion does not hold any information on the position of the markers on the chromosome. This requires a Lepidoptera-specific mapping approach where genotype data needs to be dissected within each individual for each marker based on parental origin and then analysed separately. Apart from these difficulties, there are also a number of useful consequences resulting from the absence of recombination in females. The maternally inherited alleles possess a strong chromosome-specific signature because they fully co-segregate in the absence of recombination. This maternally derived segregation pattern is called the chromosome print (YASUKOCHI 1998), which can be used to define chromosomes, verify the quality of certain markers, and exclude the maternally inherited component before positional linkage mapping. This leaves a fully positionally informative paternal allele segregation matrix.

The linkage map presented in Chapter 5 is mainly composed of AFLP markers, which are generated by a sequence of procedures consisting of: Restriction – adapter ligation – two subsequent PCRs – and, fragment analysis. These markers have the advantage that they can be easily produced without any previously available sequence data and they usually provide a large number of polymorphic markers in a single reaction. However, there are three downsides to this technique: (*i*) AFLPs are dominant markers with heterozygotes usually undistinguishable from dominant homozygotes, (*ii*) different markers can produce PCR products of the same size, causing unreliable genotype scores due to overlapping peaks (*iii*) the markers are anonymous in the sense that no sequence data is available for them. The use of dominant AFLP markers requires a large proportion of data to be excluded from analysis because they often do not reveal which allele was obtained from the father. The exact proportion of data to be excluded depends on the ratio of markers for which both parents were heterozygous, which is particularly high in the current cross. This exclusion of uninformative data is named censoring and it results in two initially incompatible linkage groups per chromosome. These two separate linkage groups can be combined into one when a sufficient number of fully informative paternally segregating marker alleles (named anchoring markers) are available. This integration of linkage groups is commonly performed by using Mapmaker software (HECKEL *et al.* 1999; JIGGINS *et al.* 2005; KAPAN *et al.* 2006; MIAO *et al.* 2005; WANG and PORTER 2004; YAMAMOTO *et al.* 2008). Unfortunately, this method was considered unreliable for the *B. anynana* AFLP data because the proportion of incompatible markers that are included within a sliding window analysis is unacceptably high. Instead, where possible, a novel linkage group integration approach using Joinmap software was used, and the remaining incompatible linkage groups (with too few anchoring markers) were produced separately.

Another level of complexity was caused by the full-sib cross design. The reason to choose this setup was to generate the maximum phenotypic range for eyespot size (explained below) within a single cross, but the drawbacks associated with this approach were not fully anticipated when these experiments were initiated. A tailored

approach was used to expose the information that was unobtainable with standard analysis. The essence of the full-sib design effects is that the least useful markers, for which both parents are heterozygous, are overrepresented and that backcross markers are less common and even absent for 25% of the chromosomes. This results in dramatically lower frequencies of anchoring markers and an absence of chromosome prints for 25% of the chromosomes. The linkage map consists of 28 linkage groups, corresponding with the 28 chromosomes described in Chapter 4. It contains a few microsatellite markers and has the nuclear gene *tpi* on the Z-chromosome, which all act as annotated markers with sequences deposited in GenBank, and the majority of markers are AFLPs. A *B. anynana* linkage map using a very extensive number of single-copy nuclear genes was published after the AFLP-based linkage map (BELDADE *et al.* 2009a), setting a new standard for *B. anynana* and for Lepidoptera in general. The AFLP-based linkage map serves its own purpose though, since it provides a segregation reference to detect quantitative trait loci involved in eyespot size determination (Chapter 6), and it will hopefully deter lepidopterists from using full-sib crosses for mapping studies.

Quantitative trait loci affecting eyespot size

The final chapter explores the genetic components involved in ventral wing eyespot size determination, using the tools and resources developed in earlier chapters. Unravelling the control of eyespot formation in *B. anynana* is of great importance because eyespot morphology forms the most striking component (at least from human perception) of the two different seasonal forms in this species, and has ecological relevance in the context of crypsis and predator escape. It is a paradox to look for genetic polymorphisms to explain seasonal polyphenism, since the size of eyespots on the ventral wing surfaces for the seasonal forms of *B. anynana* is usually described in the context of phenotypic plasticity determined by ambient temperature rather than by genetic variation. However, genetic variation is still needed to account for evolution of the polyphenic trait. There is a strong genetic regulation of eyespot size, which is not obvious in a natural situation because there are many genetic components involved which normally rule each other out due to a stochastic mix of alleles with antagonistic effects. Local adaptation of the wing pattern plasticity response was found between 2 populations from distant localities with different climates (DE JONG *et al.* 2010), which indicates a degree of genetic control. Artificial selection revealed a high degree of heritable genetic variation for eyespot size (BELDADE *et al.* 2002; BRAKEFIELD *et al.* 1996; WIJNGAARDEN and BRAKEFIELD 2000; WIJNGAARDEN and BRAKEFIELD 2001) and truncated selection lines were generated for small and large eyespots (BRAKEFIELD *et al.* 1996), named Low-line and High-line respectively. A cross between these lines formed the basis of a quantitative trait loci (QTL) analysis. This technique reveals the genomic positions, effects and interactions of multiple genes involved in a phenotypic trait by analysing co-segregation between marker alleles and phenotypes. The great advantage of this approach is that, apart from the availability of a linkage map, prior genetic or biochemical knowledge is not needed to detect the separate components involved in producing different phenotypes.

A full-sib F_2 cross based on a grandmother from the High line and a grandfather from the Low line was used for the QTL analysis. This design was chosen to generate a maximal phenotypic range in the F_2 while keeping genetic background noise to a

minimum. This approach increases the proportion of alleles within the cross that are unique to one of the lines relative to a setup with four grandparents and should therefore give a strong allele - phenotype association. This full-sib design combined with dominant markers and absence of recombination in females had severe analytical consequences. Most of the difficulties encountered were similar to those that obstructed the linkage map construction. In hindsight, having a strong allele-phenotype association is useful in general, but needless in Lepidoptera QTL mapping because the chromosome print also reveals the origin (i.e. High or Low line) of the inherited alleles, even if the allele is not unique to one of the lines. Therefore it must be accepted that the full-sib design has many negative consequences and no obvious benefits. There is nevertheless still useful information within the QTL data. Firstly, a chromosome-by-chromosome QTL analysis that revealed linkage groups which affect mimetic colour patterns in *Heliconius melpomene* subspecies (BAXTER *et al.* 2008) could also be used for the *B. anynana* eyespot size data. Secondly, single marker analysis can reveal co-segregation between phenotype and genotype for each individual marker locus, albeit that the results are only reliable for markers that are closely linked to a QTL. Thirdly, interval mapping is possible when only the segregation of paternally inherited QTL alleles are compared with paternally inherited marker alleles. This requires a custom-designed approach once more with each chromosome divided in two linkage phases and analysed separately.

A moderate number of six QTL signals was found, with a striking difference of genotypic effects between the sexes consistent with earlier findings based on similar line crosses (WIJNGAARDEN and BRAKEFIELD 2000). The QTLs could be linked to chromosomes or sometimes chromosomal regions of the gene based *B. anynana* linkage map (BELDADE *et al.* 2009a) and the high degree of synteny in Lepidoptera allowed further predictions of potential candidate genes based on the *B. mori* genome assembly (XIA *et al.* 2004). This made it possible to compare the QTL positions with those of genes already known to be involved in eyespot formation and with genes that are potentially involved (candidate genes). One QTL is in a region that contains *ecdysone receptor* and *Phantom*, which are both related to ecdysteroid signaling (IGA and SMAGGHE 2010; ZHENG *et al.* 2010). Four other QTLs, each on separate chromosomes, are not positioned accurately enough to confirm or reject coincidence with eyespot candidates. Two QTLs mapped to regions without any known candidate genes, thus providing an exciting opportunity to expand future research on eyespot size determination. Surprisingly, no QTL association was found for *distal-less*, which co-segregated with eyespot size in a cross between lines that were selected for the size of non-plastic dorsal forewing, rather than the ventral hindwing, eyespots (BELDADE *et al.* 2002).

The QTL analysis gives an impression of the degree to which our current knowledge covers the processes involved in eyespot size determination, and it opens up new directions for further investigations in this field. A more detailed knowledge about the determination of eyespot size in the future will contribute to a better understanding of the overall processes involved in seasonal polyphenism, from ecological predator-prey interactions through signal transduction of the temperature trigger and timing of the ecdysteroid peaks to the deposition of pigments in the wing scales.

REFERENCES

ABE, H., K. MITA, Y. YASUKOCHI, T. OSHIKI and T. SHIMADA, 2005 Retrotransposable elements on the W chromosome of the silkworm, *Bombyx mori*. Cytogenet Genome Res **110:** 144-151.

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.

BAXTER, S. W., S. E. JOHNSTON and C. D. JIGGINS, 2008 Butterfly speciation and the distribution of gene effect sizes fixed during adaptation. Heredity **102:** 57-65.

BELDADE, P., and P. BRAKEFIELD, 2002 The genetics and evo-devo of butterfly wing patterns. Nat Rev Genet **3:** 442 - 452.

BELDADE, P., P. BRAKEFIELD and A. LONG, 2002 Contribution of *Distal-less* to quantitative variation in butterfly eyespots. Nature **415:** 315 - 318.

BELDADE, P., P. M. BRAKEFIELD and A. D. LONG, 2005 Generating phenotypic variation: prospects from "evo-devo" research on *Bicyclus anynana* wing patterns. Evolution & Development **7:** 101-107.

BELDADE, P., S. V. SAENKO, N. PUL and A. D. LONG, 2009a A Gene-Based Linkage Map for *Bicyclus anynana* Butterflies Allows for a Comprehensive Analysis of Synteny with the Lepidopteran Reference Genome. PLoS Genet **5:** e1000366.

BELDADE, P., A. E. VAN'T HOF, M. A. JERÓNIMO and A. D. LONG, 2009b Microsatellite markers associated with genes expressed in developing wings of *Bicyclus anynana* butterflies. Molecular Ecology Resources **9:** 1487-1492.

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.

BRAKEFIELD, P. M., P. BELDADE and B. J. ZWAAN, 2009 The African Butterfly *Bicyclus anynana*: A Model for Evolutionary Genetics and Evolutionary Developmental Biology. Cold Spring Harb Protoc **doi:** 10.1101/pdb.emo1122.

BRAKEFIELD, P. M., J. GATES, D. KEYS, F. KESBEKE, P. J. WIJNGAARDEN *et al.*, 1996 Development, plasticity and evolution of butterfly eyespot patterns. Nature **384:** 236-242.

CARTER, P. A., and W. B. WATT, 1988 Adaptation at specific loci v. metabolically adjacent enzyme loci may have very distinct experiences of selective pressures. Genetics **119:** 913-924.

D'ALENÇON, E., H. SEZUTSU, F. LEGEAI, E. PERMAL, S. BERNARD-SAMAIN *et al.*, 2010 Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements PNAS **107:** 7680-7685.

DE JONG, M. A., F. M. N. H. KESBEKE, P. M. BRAKEFIELD and B. J. ZWAAN, 2010 Geographic variation in thermal plasticity of life history and wing pattern in *Bicyclus anynana*. Climate Research **43:** 91-102.

DE LESSE, H., 1963 Variation chromosomique chez les *Agrodiaetus* (Lep. Lycaenidae). Rev Fr Entomol **30:** 182-189.

DE LESSE, H., 1968 Formules chromosomiques de Lépidoptères Rhopalocères d'Uganda et du Kenya. Ann. Soc. Ent. Fr. (N.S.) **4:** 581-599.

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.

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.

FISCHER, K., E. EENHOORN, A. N. M. BOT, P. M. BRAKEFIELD and B. J. ZWAAN, 2003 Cooler butterflies lay larger eggs: developmental plasticity versus acclimation. Proc R. Soc. B. **270:** 2051-2056.

HECKEL, D. G., L. J. GAHAN, Y.-B. LIU and B. E. TABASHNIK, 1999 Genetic mapping of resistance to *Bacillus thuringiensis* toxins in diamondback moth using biphasic linkage analysis. PNAS **96:** 8373-8377.

IGA, M., and G. SMAGGHE, 2010 Identification and expression profile of Halloween genes involved in ecdysteroid biosynthesis in *Spodoptera littoralis*. Peptides **31:** 456-467.

INTERNATIONAL SILKWORM GENOME CONSORTIUM, 2008 The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. Insect Biochem Mol Biol. **38:** 1036-1045.

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

JORON, M., R. PAPA, M. BELTRÁN, N. CHAMBERLAIN, J. MAVÁREZ *et al.*, 2006 A Conserved Supergene Locus Controls Colour Pattern Diversity in *Heliconius* Butterflies. PLoS Biol **4:** e303.

KANDUL, N. P., V. A. LUKHTANOV and N. E. PIERCE, 2007 Karyotypic diversity and speciation in *Agrodiaetus* butterflies. Evolution **61:** 546-559.

KAPAN, D. D., N. S. FLANAGAN, A. TOBLER, R. PAPA, R. D. REED *et al.*, 2006 Localization of Müllerian mimicry genes on a dense linkage map of *Heliconius erato*. Genetics **173:** 735-757.

KOCH, P. B., P. M. BRAKEFIELD and F. KESBEKE, 1996 Ecdysteroids control eyespot size and wing color pattern in the polyphenic butterfly *Bicyclus anynana* (Lepidoptera: Satyridae). J. Insect Physiol **42:** 223–230.

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

LYYTINEN, A., P. M. BRAKEFIELD, L. LINDSTRÖM and J. MAPPES, 2004 Does predation maintain eyespot plasticity in *Bicyclus anynana*? Proc. R. Soc. Lond. B **271:** 279-283.

MACILWAIN, C., 2000 World leaders heap praise on human genome landmark. Nature **405:** 983-984.

MADDOX, P. S., K. OEGEMA, A. DESAI and I. M. CHEESEMAN, 2004 "Holo"er than thou: Chromosome segregation and kinetochore function in *C. elegans* Chromosome Research **12:** 641-653.

MARCUS, J. M., D. M. RAMOS and A. MONTEIRO, 2004 Germline transformation of the butterfly *Bicyclus anynana*. Proc. R. Soc. Lond. B **271 Suppl 5:** S263-S265.

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.

MIAO, X.-X., S.-J. XUB, M.-H. LI, M.-W. LI, J.-H. HUANG *et al.*, 2005 Simple sequence repeat-based consensus linkage map of *Bombyx mori*. PNAS **102:** 16303-16308.

MONTEIRO, A., B. CHEN, L. SCOTT, L. VEDDER, J. H. PRIJS *et al.*, 2007 The combined effect of two mutations that alter serially homologous color pattern elements on the fore and hindwings of a butterfly. BMC Genetics **8:** 22.

MORGAN, T. H., 1912 Complete linkage in the second chromosome of the male of *Drosophila*. Science **36:** 719-720.

NGUYEN, P., K. SAHARA, A. YOSHIDO and F. MAREC, 2010 Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). Genetica **138:** 343-354.

NIEBERDING, C. M., H. DE VOS, M. V. SCHNEIDER, J.-M. LASSANCE, N. ESTRAMIL *et al.*, 2008 The Male Sex Pheromone of the Butterfly *Bicyclus anynana*: Towards an Evolutionary Analysis. PLoS ONE **3:** e2751.

OOSTRA, V., M. A. DE JONG, B. M. INVERGO, F. M. H. N. KESBEKE, F. WENDE *et al.*, 2010 Translating environmental gradients into discontinuous reaction norms via hormone signalling in a polyphenic butterfly. Proc R Soc B **doi:** 10.1098/rspb.2010.1560.

PAPA, R., C. MORRISON, J. WALTERS, B. COUNTERMAN, R. CHEN *et al.*, 2008 Highly conserved gene order and numerous novel repetitive elements in genomic regions linked to wing pattern variation in *Heliconius* butterflies. BMC Genomics **9:** 345.

PIJPE, J., P. M. BRAKEFIELD and B. J. ZWAAN, 2008 Increased life span in a polyphenic butterfly artificially selected for starvation resistance. Am Nat. **171:** 81-90.

PRINGLE, E. G., S. W. BAXTER, C. L. WEBSTER, A. PAPANICOLAOU, S. F. LEE *et al.*, 2007 Synteny and Chromosome Evolution in the Lepidoptera: Evidence From Mapping in *Heliconius melpomene*. Genetics **177:** 417-426.

ROBINSON, R., 1971 *Lepidoptera Genetics.* Pergamon Press, Oxford, UK.

RODRÍGUEZ GIL, S. G., L. M. MOLA, A. G. PAPESCHI and C. L. SCIOSCIA, 2002 Cytogenetic heterogeneity in common haplogyne spiders from Argentina (arachnida, araneae). The Journal of Arachnology **30:** 47-56.

SACCHERI, I. J., and P. M. BRAKEFIELD, 2002 Rapid spread of immigrant genomes into inbred populations. Proc Biol Sci. B. **269:** 1073-1078.

SAHARA, K., A. YOSHIDO, F. MAREC, I. FUKOVÁ, H.-B. ZHANG *et al.*, 2007 Conserved synteny of genes between chromosome 15 of *Bombyx mori* and a chromosome of *Manduca sexta* shown by five-color BAC-FISH. Genome **50:** 1061-1065.

SHAPIRO, A. M., 1976 Seasonal polyphenism. Evol. Biol. **9:** 259-333.

SUOMALAINEN, E., and K. S. J. BROWN, 1984 Chromosome number variation within *Philaethria* butterflies (Lepidoptera: Nymphalidae, Heliconiini). Chromosoma **90:** 170-176.

TAY, W. T., G. T. BEHERE, P. BATTERHAM and D. G. HECKEL, 2010 Generation of microsatellite repeat families by RTE retrotransposons in lepidopteran genomes. BMC Evolutionary Biology **10:** 144.

TRAUT, W., K. SAHARA and F. MAREC, 2007 Sex Chromosomes and Sex Determination in Lepidoptera. Sexual Development **1:** 332–346.

WANG, B., and A. H. PORTER, 2004 An AFLP-based interspecific linkage map of sympatric, hybridizing *Colias* butterflies. Genetics **168:** 215-225.

WHITE, M. J. D., 1973 *Animal Cytology and Evolution*. Cambridge Univ. Press, Cambridge.

WIJNGAARDEN, P. J., and P. M. BRAKEFIELD, 2000 The genetic basis of eyespot size in the butterfly *Bicyclus anynana*: an analysis of line crosses. Heredity **85:** 471-479.

WIJNGAARDEN, P. J., and P. M. BRAKEFIELD, 2001 Lack of response to artificial selection on the slope of reaction norms for seasonal polyphenism in the butterfly *Bicyclus anynana*. Heredity **87:** 410-420.

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

WRENSCH, D. L., J. B. KETHLEY and R. A. NORTON, 1994 Cytogenetics of Holokinetic Chromosomes and Inverted Meiosis: Keys to the Evolutionary Success of Mites, with Generalizations on Eukaryotes, pp. 282 – 342 in *Mites: Ecological and Evolutionary Analyses of Life-History Patterns*, edited by M. HOUCK. Chapman & Hall, New York.

XIA, Q., Z. ZHOU, C. LU, D. CHENG, F. DAI *et al.*, 2004 A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). Science **306:** 1937 - 1940.

YAMAMOTO, K., J. NOHATA, K. KADONO-OKUDA, J. NARUKAWA, M. SASANUMA *et al.*, 2008 A BAC-based integrated linkage map of the silkworm, *Bombyx mori* Genome Biology **9:** R21.

YASUKOCHI, Y., 1998 A dense genetic map of the silkworm, *Bombyx mori*, covering all chromosomes based on 1018 molecular markers. Genetics **150:** 1513-1525.

YASUKOCHI, Y., L. A. ASHAKUMARY, K. BABA, A. YOSHIDO and K. SAHARA, 2006 A second-generation integrated map of the silkworm reveals synteny and conserved gene order between lepidopteran insects. Genetics **173:** 1319-1328.

YASUKOCHI, Y., M. TANAKA-OKUYAMA, F. SHIBATA, A. YOSHIDO, F. MAREC *et al.*, 2009 Extensive Conserved Synteny of Genes between the Karyotypes of *Manduca sexta* and *Bombyx mori* Revealed by BAC-FISH Mapping. PLoS ONE **4:** e7465.

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

ZHENG, W.-W., D.-T. YANG, J.-X. WANG, Q.-S. SONG, L. I. GILBERT *et al.*, 2010 *Hsc70* binds to *ultraspiracle* resulting in the upregulation of 20-hydroxyecdsone-responsive genes in *Helicoverpa armigera*. Molecular and Cellular Endocrinology **315:** 282-291.

ZIJLSTRA, W. G., M. J. STEIGENGA, P. M. BRAKEFIELD and B. J. ZWAAN, 2003 Simultaneous selection on two fitness-related traits in the butterfly *Bicyclus anynana*. Evolution **57:** 1852–1862.

ZIJLSTRA, W. G., M. J. STEIGENGA, P. B. KOCH, B. J. ZWAAN and P. M. BRAKEFIELD, 2004 Butterfly Selected Lines Explore the Hormonal Basis of Interactions between Life Histories and Morphology. American Naturalist **163:** E76-E87.

Chapter 2

Characterization of 28 microsatellite loci for the butterfly *Bicyclus anynana*[1]

Arjèn E. van't Hof
Bas J. Zwaan
Ilik J. Saccheri
Derek Daly
Jeanette N.M. Bot
Paul M. Brakefield

ABSTRACT

We present 28 polymorphic microsatellite loci, including a sex-linked W-chromosome marker, for the afrotropical butterfly *Bicyclus anynana*. Our primary motivation to develop these markers was to apply them in QTL mapping studies. A technique is also proposed that may be useful in avoiding redundant sequences which are common in Lepidopteran enriched libraries. Pedigree analysis was performed to test Mendelian segregation of the markers and to address the issue of null alleles.

[1] This chapter has been published in Molecular Ecology Notes 2005 volume **5**, pp 169-172. The following text includes some detail that was removed in the published version to comply with the strict word number limitation of MEN.

INTRODUCTION

Obtaining polymorphic microsatellites in Lepidoptera is a more challenging task than for most other organisms. The yields are very low, but not because of low levels of polymorphism. Low microsatellite densities, PCR amplification problems and unusually high redundancy of sequences are some of the reported drawbacks that have prevented large numbers of these markers being found (MEGLECZ *et al.* 2004; ZHANG 2004). As a result, the highest number of applicable loci published in any single Lepidoptera species to date is 15 (FLANAGAN *et al.* 2002; REDDY *et al.* 1999) with an average yield of less than 8, (based on the 20 species examined to date) (AMSELLEM *et al.* 2003; ANTHONY *et al.* 2001; BEZZERIDES *et al.* 2004; BOGDANOWICZ *et al.* 1997; CALDAS *et al.* 2002; CASSEL 2002; COATES and HELLMICH 2003; DALY *et al.* 2004; FLANAGAN *et al.* 2002; HARPER *et al.* 2000; JI *et al.* 2003; KEYGHOBADI *et al.* 1999; KEYGHOBADI *et al.* 2002; KLÜTSCH *et al.* 2003; KOSHIO *et al.* 2002; MEGLÉCZ and SOLIGNAC 1998; PALO *et al.* 1995; REDDY *et al.* 1999; ROUSSELET *et al.* 2004; SCOTT *et al.* 2004; TAN *et al.* 2001; WARDILL *et al.* 2004; WILLIAMS *et al.* 2002). Moreover, polymorphic microsatellite loci in Lepidoptera often suffer from substantial proportions of null-alleles (AMSELLEM *et al.* 2003; CASSEL 2002; COATES and HELLMICH 2003; DALY *et al.* 2004; FLANAGAN *et al.* 2002; JI *et al.* 2003; KEYGHOBADI *et al.* 1999; ROUSSELET *et al.* 2004; WARDILL *et al.* 2004).

METHODS

Genomic libraries enriched for CA, GA, AAT, ATG, GAA, TACA and TAGA repeats were constructed by Genetic Identification Services (GIS, http://www.genetic-id-services.com; Chatsworth, CA, USA). The used material came from our laboratory stock population, which was founded and reared under conditions that sustain high

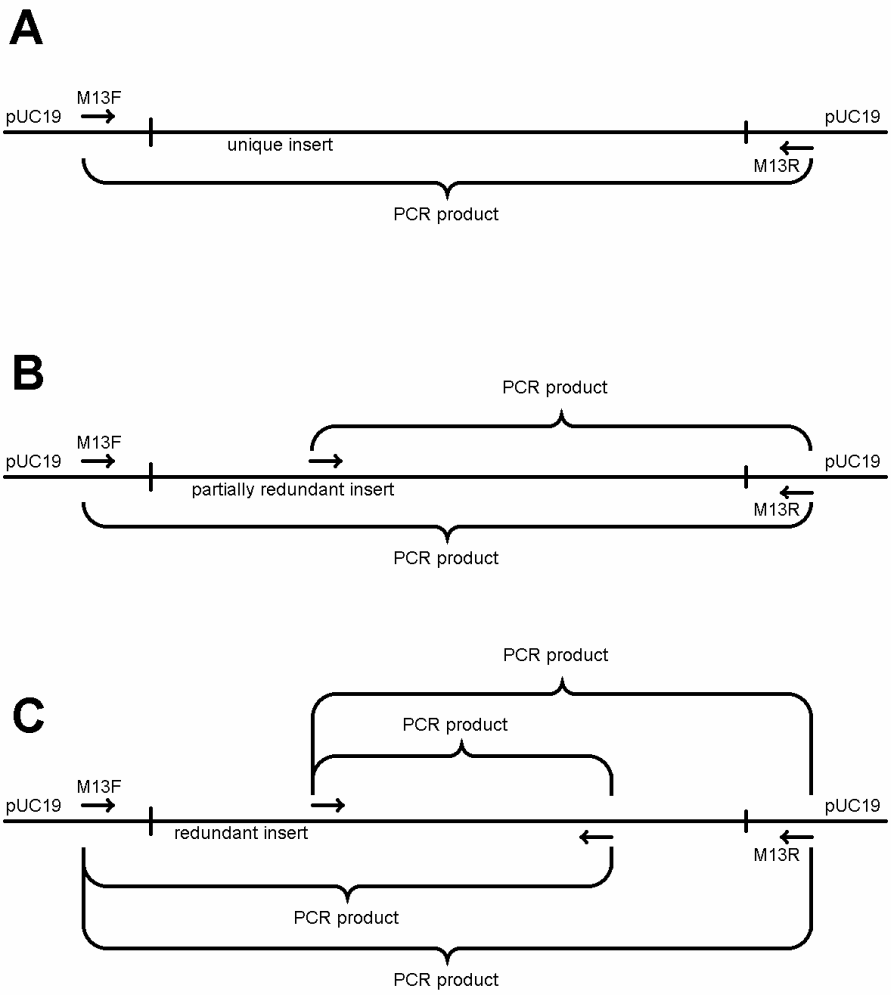


Figure 2.1 PCR strategy to detect and avoid redundant inserts.
The horizontal lines represent the pUC19 cloning vector with insert as PCR template that is exposed to a mix of M13 primers and oligos that match the most commonly found redundant sequences; (A) A unique insert gives a single PCR product generated with the M13 primers; (B) a partially redundant sequence produces two PCR products; (C) a fully redundant insert results in multiple PCR products. Only clones with single PCR products were used to develop microsatellite markers.

levels of heterozygosity (SACCHERI and BRUFORD 1993). DNA of a single female was extracted as described in Saccheri & Bruford (1993). The recombinant plasmids were transformed into *Escherichia coli* JM109 Competent Cells (Promega, Madison, WI, USA) and identified by blue/white screening following the manufacturer's instructions. Randomly picked white clones were grown and purified using the Qiaprep spin miniprep kit (Qiagen, Hilden, Germany). Sequencing was performed by different commercial facilities.

After detecting a substantial number of similar (but not identical) sequences, we developed a technique to avoid obtaining more redundant sequences by screening the plasmids first by multiplex PCRs, amplifying the 3 most common sequences making up more than half of the CA library. Universal M13 primers were used as positive control, thereby giving a single PCR fragment for unique clones and multiple bands for redundant sequences (Fig. 2.1 and 2.2). This technique proved to be 100% discriminative with miniprepped plasmids as template and significantly reduced the number of unusable sequences with colony PCRs. A total of 960 colonies were screened with this method. The PCR conditions for this test were 3 min. 94ºC, 20 cycles of 30 sec. 94ºC, 30 sec. 55ºC, 30 sec 72ºC and a 5 minute extension at 72ºC in 15 µl containing 0.025 U/µl taq polymerase (Qiagen), 1×PCR buffer, 0.2 mM of each dNTP and 0.67 µM of each primer and toothpicked colonies heated for 10 minutes at 95ºC in 7.5 µl H_2O.

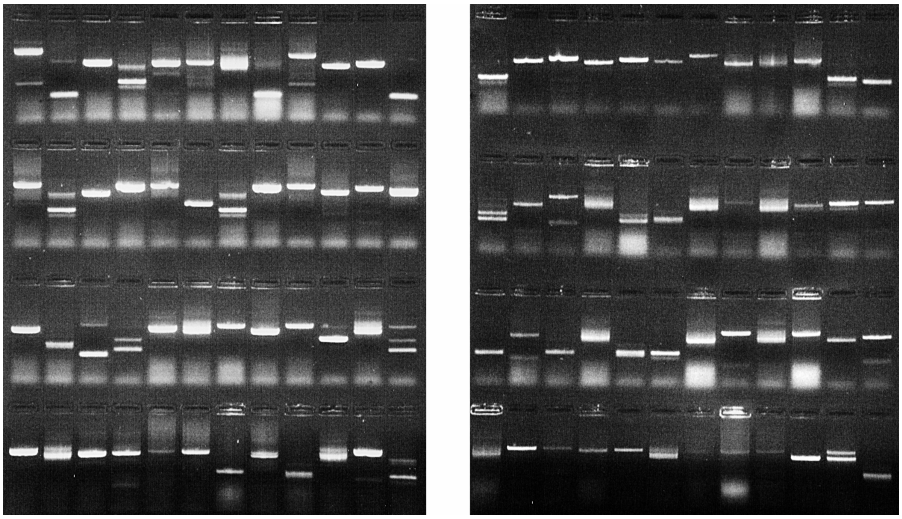


Figure 2.2 Colony PCR of 96 randomly picked clones (representing 10% of the total 960 clones) using a mix of four primer pairs per reaction, consisting of M13 forward and reverse and primers that match the three most frequently encountered redundant sequences. Clones that produced more than one band were excluded from further marker development. The smear beneath the PCR products is presumably degraded RNA from the *E. coli* plasmid host. A size ladder was not used.

From a total of 298 sequences, primers were designed for 82 sequences with OLIGO version 6 (RYCHLIK 2000) of which 51 showed successful amplification. Template for the optimization and subsequent steps described below was extracted from half a thorax with the DNeasy tissue kit (Qiagen). The amount of DNA was

Table 2.1 PCR conditions.
T_a = optimal annealing temperatue, $MgCl_2$ = final reaction concentration.

Locus	T_a (ºC)	$MgCl_2$ (mM)	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
BA-AAT1	55	1.5	CCGGACCGAGTTCCAACCT	GATTGCCACGACCCCAAAAT
BA-AAT2	55	3.0	TTGATCCCGACCGTGTGA	ACCGACAACAATGCGACAGC
BA-AAT3	55	1.5	GCGGCTGGCAACTTTATAATAACT	GTCACGGCAACGAGGATACCAAAC
BA-AAT4	50	1.5	CACACAGTATAAATGCGTGTAAGT	AAAGTTCAGTAAATAAAGGCTT
BA-ATG1	50	1.5	CTGCAGTGGACGTCCATCGG	CACAGACTACCTCGCGACAG
BA-ATG2	50	1.5	ACCGTATAAAGATGATGACGTT	CTCCCTGACACCATGCAAG
BA-ATG3	55	1.5	GCAGCAAGCGACGACAAGGT	CTGCAGTGGACGTCCATCGG
BA-CA1	55	3.0	TTGTCGTTTGTCGCAGATT	TAGCGACAGCGAGGACTAGA
BA-CA2	53	1.5	AAATCAACAGCGTTACCAAG	GCGACTAGCGGAAACTA
BA-CA3	55	1.5	GCGCACATTTTAATGTCT	GCACTGGGCAATATACTTAC
BA-CA4	50	3.2	TTTGTCCAAATCGCTTCAG	TGGAGGGAAAGTTTGTGGTAA
BA-CA5	55	1.5	CGCAAGTCCTCTCGTCATGG	CCGCAGTCAAGTCGTAGCTT
BA-CA6	55	1.5	GGAATGAAAAGTAGCCTATG	TTGGCTGAATCACACTATCA
BA-CA7	55	1.5	TCCGCGTCTGTACCCGTAGA	TCAGTAGCCGCAGCGAAAAG
BA-CA8	48	1.5	CATGCAAAATATGAAATAAGGA	ACTGGATATTACTGGATGCATT
BA-CA9	48	1.5	ACGTGGATAAACAGTAATA	TGGCACAGAGATAGTACAT
BA-CA10	55	1.5	CCGCAGTTGGAGTTTATCGT	(GTTTC)AACCTTGGGCTGTGGA
BA-CA11	53	1.5	GGCGCAAAAGAATGACCAAC	(GTTTA)TGGGGTGGATTGAGTGTA
BA-CA12	55	1.5	CTCGCCAGGACCGGTTCTAC	CACAGAGCCGACGTGTTCCA
BA-CA13	55	1.5	CAAATTCCAGCCAAATCGGT	GCTTCCATCGCCAGTAAAC
BA-CA14	53	1.5	GCTCTTCCCTGCTTAGATG	AACAGAGTTTGCAAATCGTC
BA-CA15	50	1.5	GCGCGGTGGTTTAAGTTACT	GTTCAATGGATGCGGTCTGG
BA-CA16	47	1.5	TTACGTCGTCAGAGTTATT	TGGGTATAACTAAAACTAAAGA
BA-GA1	50	1.5	ATGCCGGATCTTAGACTA	TGAGCTCGGACGAAGTGCAA
BA-GA2	47	1.5	TCACAGTGGAAATTCGGATAA	TGGGTGGAAGGTGTACCGAC
BA-GAA1	50	1.5	CTCAAAGGAGGAACAAACATAC	CCATTAGAAAAGCTGAGGATC
BA-GAA2	48	1.5	CAAATTAGAAATTAGGGTAT	CTCCACTTAGGGCATTACAC
BA-GAA3	55	1.5	ACTGCATATTCTCCGTGTTTC	AATCTAGTCAATGGCGATCAC

optimized as added volume instead of concentration. PCR was performed in 10 µl containing 5µl 2×Reddymix 1.5 (Abgene, Portsmouth, NH, USA) 0.33 µM of each primer, 1µl template and additional $MgCl_2$ where required. PCR conditions are: 3min. 95ºC, 30 cycles of 30sec. 94ºC; 30sec. T_a; 45sec. 72ºC, and a 30 min. 72ºC. $MgCl_2$ concentration and annealing temperature (T_a) are given in Table 2.1.

Polymorphism levels are based on characterization of butterflies from the stock population. Two of the polymorphic loci (BA-AAT1 and BA-CA1) were detected with ^{32}P-labelled primers on a polyacrylamide sequence gel in absence of a size marker (hence the size ranges for these two loci are not specified in Table 2.2). Banding patterns were visualized with phosphor imaging plates and were manually scored. Characteristics for these two loci are based on 16 individuals. The remaining loci were tested with fluorescently labeled primers (JOE, TAMRA, 5-FAM) with ROX-500 size standard, based on 29 individuals using an ABI-377 automatic sequencer (Applied Biosystems, Foster City, CA, USA) and analyzed with GENESCAN 3.12. Characteristics of the polymorphic loci were calculated by GENEPOP (RAYMOND and ROUSSET 1995) and are included in Table 2.2.

Table 2.2 Characteristics of the 28 polymorphic microsatellite loci.
n = sample size, N_a = number of alleles, H_O = Observed Heterozygosity, H_E = Expected Heterozygosity, HW eq. = Compliance with Hardy-Weinberg equilibrium, Mend. segr. = Mendelian segregation.

Locus	Core Sequence	n	N_a	H_O H_E	HW. eq.	Exp. Size	Mend. segr.	Size range	GENBANK acc. #
BA-AAT1	$(AAT)_{11}$	16	9	0.30 0.90	No	342	Yes	N.A.	AY785058
BA-AAT2	$(AAT)_{10}$	29	3	0.07 0.07	Yes	119	Yes	108-120	AY785059
BA-AAT3	$(AAT)_{11}$	29	2	0.36 0.30	Yes	238	Yes	144-238	AY785060
BA-AAT4	$(AAT)_{13}$	29	4	0.59 0.52	Yes	154	No	129-154	AY785061
BA-ATG1	$(ATG)_{11}$	29	8	0.83 0.76	Yes	139	Yes	121-157	AY785062
BA-ATG2	$(GAT)_4GAC\text{-}(GAT)_7$	29	4	0.34 0.60	No	101	No	73-100	AY785063
BA-ATG3	$(ATG)_{12}$	29	6	0.55 0.70	Yes	259	Yes	208-262	AY785064
BA-CA1	$(CA)_{13}$	16	2	0.31 0.51	Yes	170	Yes	N.A.	AY785065
BA-CA2	$(TG)_{10}AAGAC\text{-}(GT)_{3½}$	29	5	0.59 0.70	Yes	268	No	212-288	AY785066
BA-CA3	$(CA)_{13}$	29	4	0.36 0.56	Yes	156	Yes	148-158	AY785067
BA-CA4	$(CA)_{26}$	29	3	0.38 0.50	Yes	207	Yes	181-207	AY785068
BA-CA5	$(CA)_{32}AT(CA)_{6½}$	29	2	0.14 0.51	No	200	Yes	142-202	AY785069
BA-CA6	$(CA)_7$	29	4	0.28 0.43	No	130	Yes	129-195	AY785070
BA-CA7	$(GT)_{8½}ATCA(GT)_3AT(GT)_7$ $AT(GT)_8ATCA(GT)_3$	29	7	0.14 0.76	No	140	No	123-239	AY785071
BA-CA8	$(TG)_3C(GT)_3\text{-}TCGC(GT)_{8½}$	29	5	0.59 0.72	Yes	115	Yes	108-122	AY785072
BA-CA9	$(CA)_{4½}TA(CA)_4TA(CA)_3C(CA)_{11½}$	29	5	0.33 0.66	No	124	Yes	82-124	AY785073
BA-CA10	$(CA)_7CT(CA)_7TA\text{-}(CA)_3$	29	3	0.55 0.51	Yes	223	Yes	211-219	AY785074
BA-CA11	$(AC)_6GC(AC)_6GCACGC(AC)_{10}$	29	4	0.41 0.69	No	225	Yes	168-250	AY785075
BA-CA12	$(CA)_{18}$	29	6	0.59 0.65	Yes	215	Yes	181-259	AY785076
BA-CA13	$(CA)_{11}$	29	4	0.32 0.66	No	191	Yes	186-196	AY785077
BA-CA14	$(GT)_3ATGC(GT)_3GC(GT)_2GC$ $(GT)_{10}AT(GT)_5$	29	5	0.32 0.74	No	171	Yes	165-175	AY785078
BA-CA15	$(GT)_{12}GA(GT)_{13}AT(GT)_2$	29	7	0.68 0.76	Yes	181	Yes	130-240	AY785079
BA-CA16	$(CT)_{11}(CA)_{20}(TA)_3$	29	2	N.A.	N.A.	130	Yes	127-137	AY785080
BA-GA1	$(CT)_8CAT(CT)_3$	29	4	0.83 0.61	Yes	148	No	138-150	AY785081
BA-GA2	$(GA)_{24}$	29	4	0.37 0.40	Yes	180	No	169-179	AY785082
BA-GAA1	$(GAA)_8$	29	4	0.21 0.22	No	160	Yes	157-166	AY785083
BA-GAA2	CTT-GTT-$(CTT)_9$	29	2	0.14 0.27	Yes	274	Yes	273-281	AY785084
BA-GAA3	$(GAA)_8$	29	3	0.48 0.55	Yes	246	Yes	235-247	AY785085

RESULTS and DISCUSSION

An unusually high total of 28 polymorphic loci were found, a further 13 were monomorphic and 10 had uninterpretable banding patterns. Allele numbers vary between 2 and 9 with gene diversities ranging from 0.07 to 0.83. Departure from Hardy-Weinberg equilibrium is in 8 out of 10 cases caused by heterozygote deficit. Lepidoptera usually have WZ/ZZ sex chromosomes, in which the females possess the heterogametic WZ pair. The BA-CA16 locus only produces PCR products in females, and must therefore be on the W-chromosome. Such a marker can be useful to determine sex by means of PCR in life stages that do not display obvious sexual characteristics.

Pedigree analyses confirmed Mendelian inheritance for 22 loci and indicated that six loci deviated due to fully transmitted null-alleles or multiple bands (Table 2.2). Hence, many of the loci isolated here will be useful for population genetic studies, and all of them will be suitable as anchoring points in segregating crosses.

ACKNOWLEDGEMENTS

An NWO Short study visit grant provided the means to use the Lab facilities and the inspiring advice of Professor Steve Kemp of the Animal Genomics group of the University of Liverpool. The work was further supported by NWO grants 811-34.005 and Puls-805-48-002, and by the IOP grant IGE01014

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.

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.

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.

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.

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.

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.

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.

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.

RAYMOND, M., and F. ROUSSET, 1995 GENEPOP version 1.2.: Population genetics software for exact tests and ecumenicism. Journal of Heredity **86:** 248-249.

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.

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

SACCHERI, I. J., and M. W. BRUFORD, 1993 DNA fingerprinting in a butterfly, *Bicyclus anynana* (Satyridae). Journal of Heredity **84:** 195-200.

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.

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.

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

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.

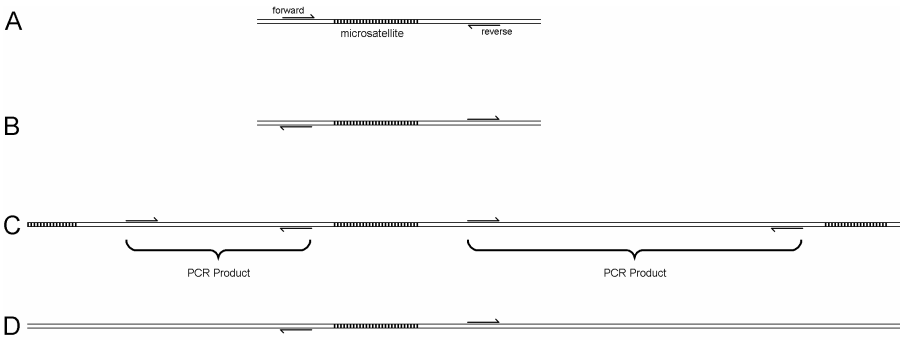


Figure 3.1 Response of PCR amplification to different microsatellite-flank arrangements. A: Example of normal microsatellite primer design with a forward and a reverse primer on either side of the repeat, initiating polymerization directed towards each other. B: Primer design for this particular experiment 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.

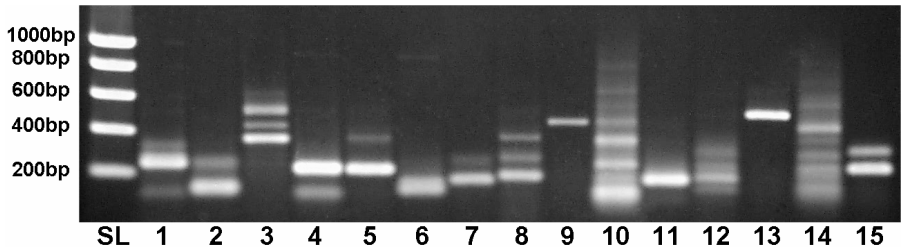


Figure 3.2 PCR product from 15 different primer pair combinations designed to test sequence associations found in the ATG library. Lane numbers correspond to the following primer combinations (see Fig. 3.3 for primer locations): SL = Eurogentec Smartladder; **1** = BA-ATG1 subgroup 2A upstream A and 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)

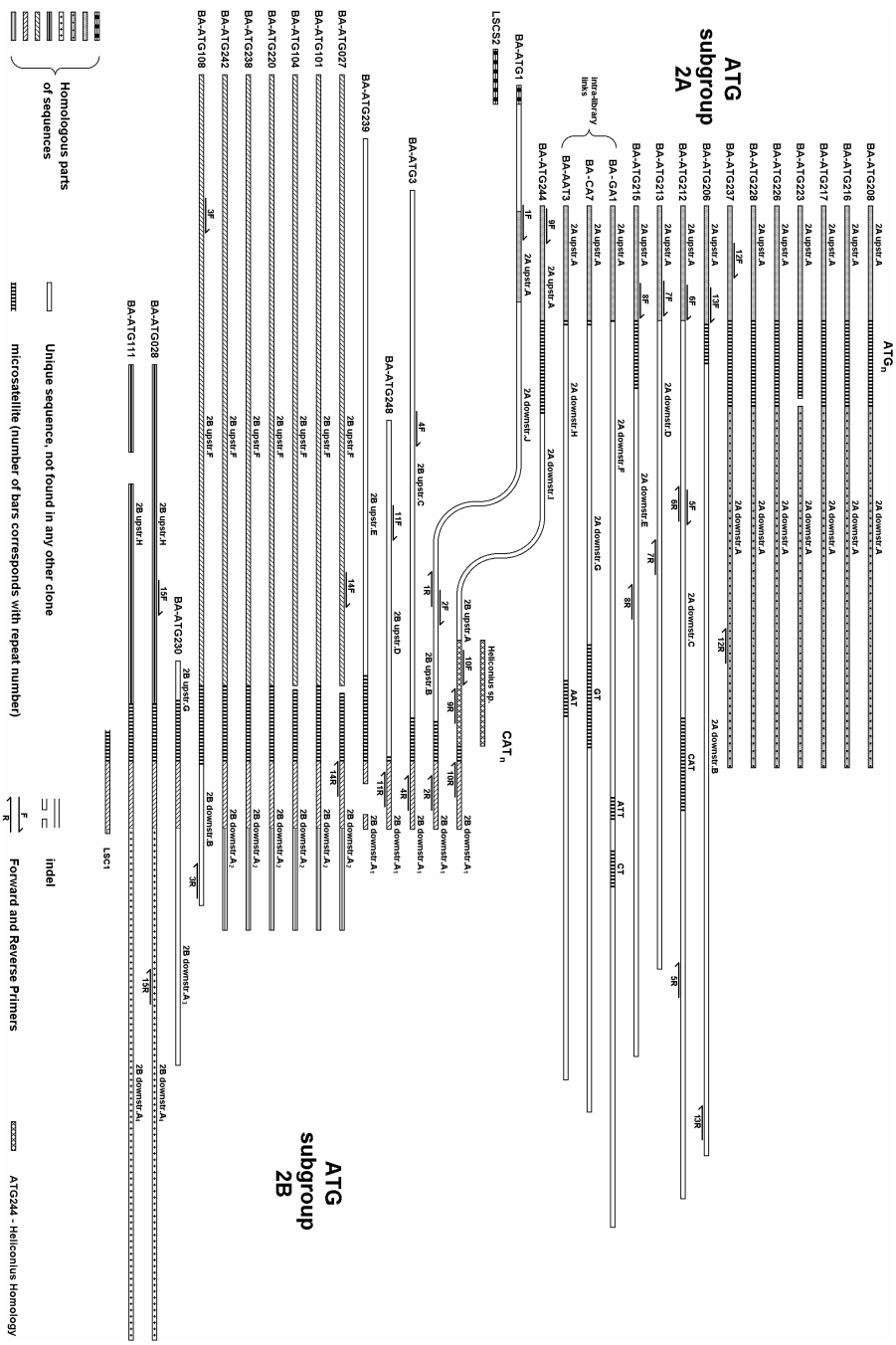


Figure 3.3 Schematic representation of the alignment of ATG group 2 sequences, showing two subgroups (2A & 2B) linked together by two sequences possessing characteristics of both. The majority are intra-library links, grouped to symmetrical sequence families with variable microsatellites, and asymmetrical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

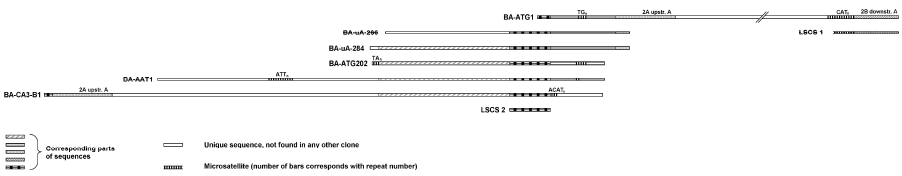


Figure 3.4 Schematic alignment of sequences upstream of subgroup ATG2A, showing the full LSCS2 alignment. There is partial overlap with Fig. 3.3, which uses the same patterns for homologous sections.

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).

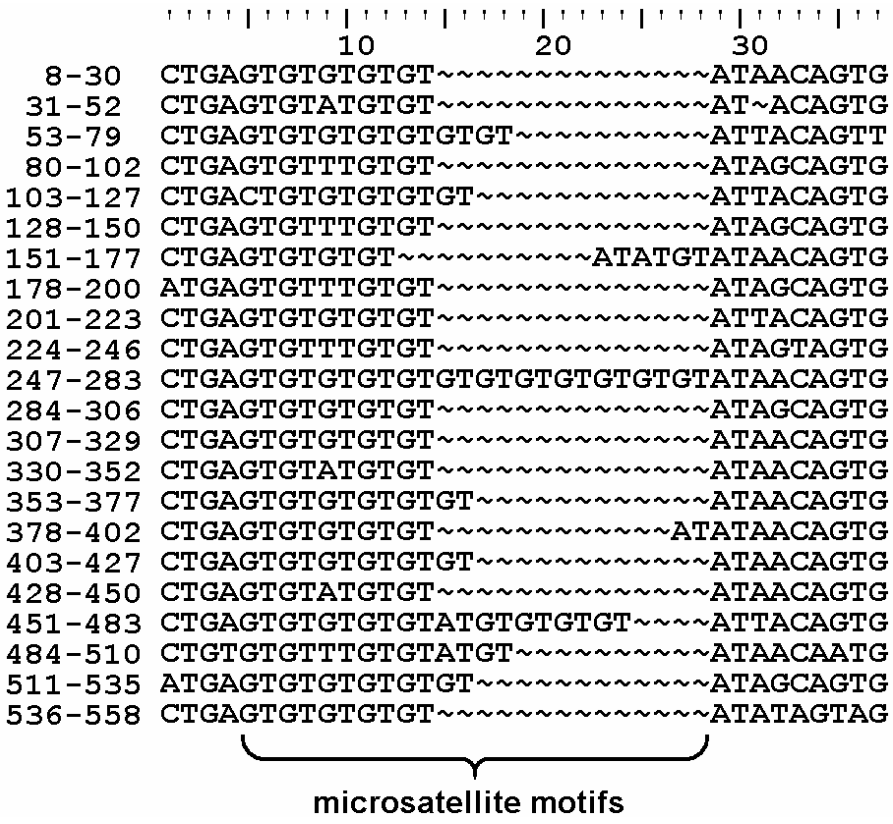


Figure 3.5 Internal alignment of a 551 bp stretch of BA-CA1-G4, showing 22 minisatellite units with incorporated variable microsatellites (GT_4-GT_{12}).

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

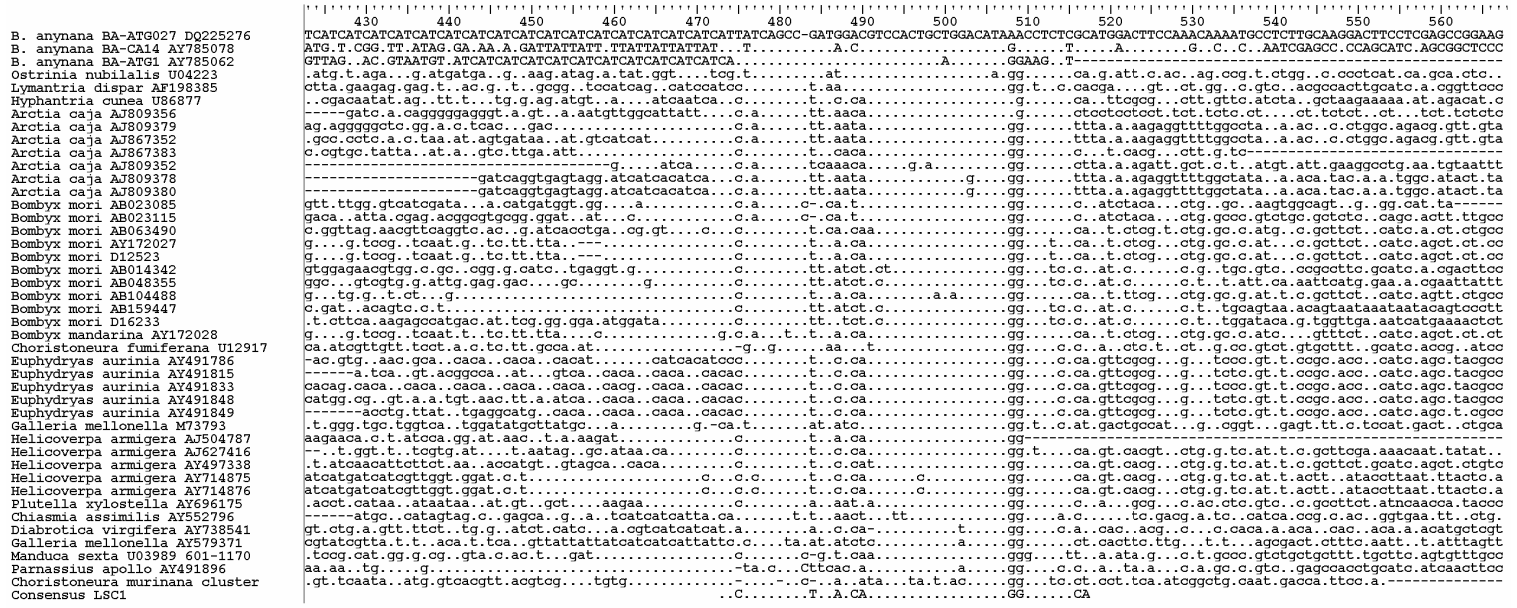
AAAATTTAGCCTATGTTACTCGGGAATAGTGTAGCTTTCCAACAGTGAAA
GAATTTTTCAAATCGGTTCAGTAGTTTCTGAGCCTATTCATTA-$CAAA_n$-
TCTTTCCTCTTTATAATATTAGTATAGAT

LSCS4

TTATTGGAACGAAGTTCCTTATCGCGCGTTGCGAAAGGGGGCTAGACGGA
AAAAATTAAGACCAAAAGTTGTCACGACACTTTTT

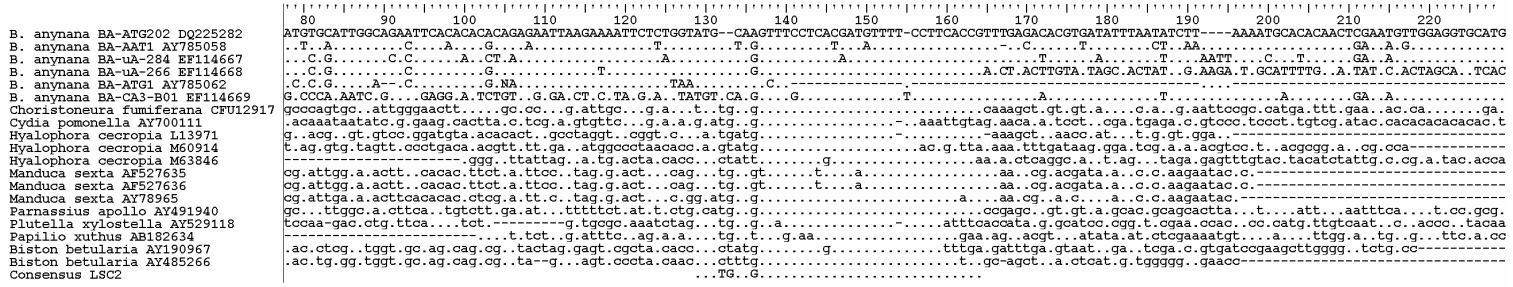
Appendix 3.6

LSCS1 alignment



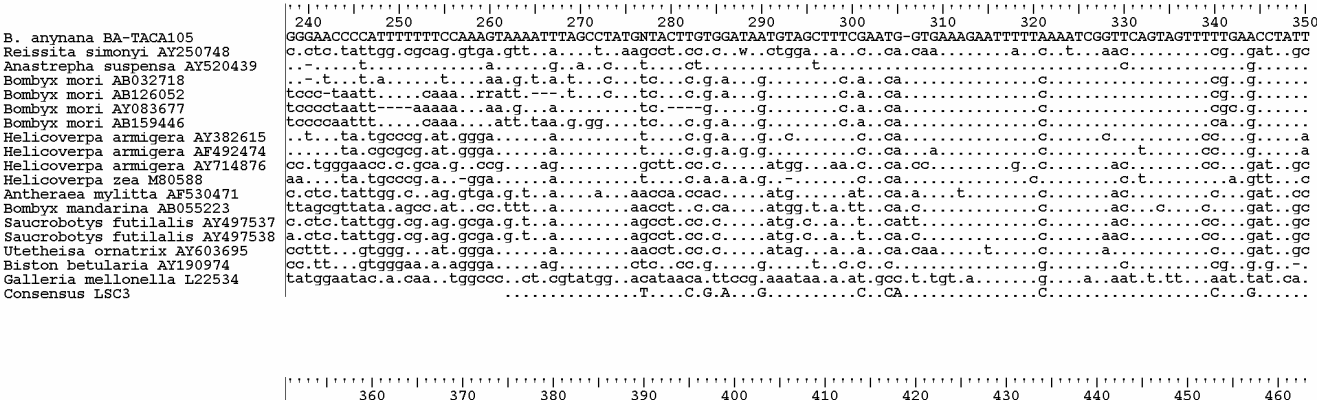
Appendix 3.7

LSCS2 alignment



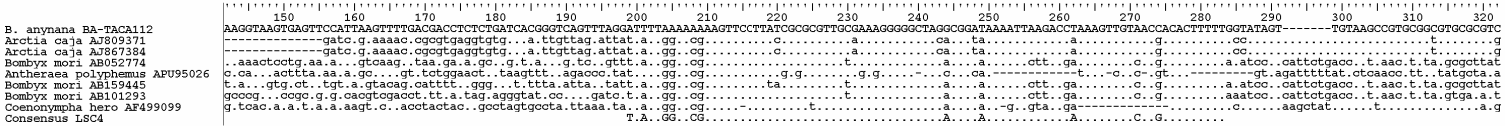
Appendix 3.8

LSCS3 alignment



Appendix 3.9

LSCS4 alignment

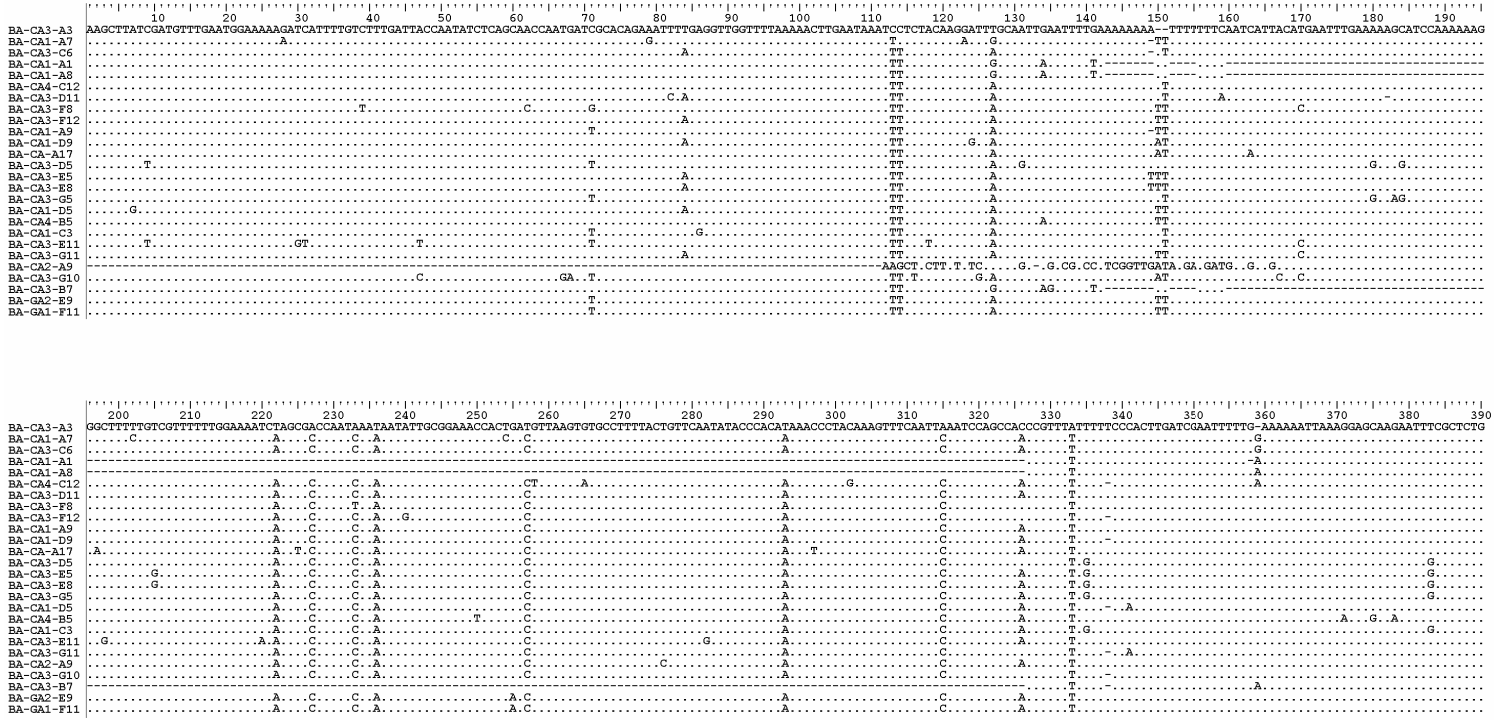


APPENDIX 3.6 – 3.9 DETAILS

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.

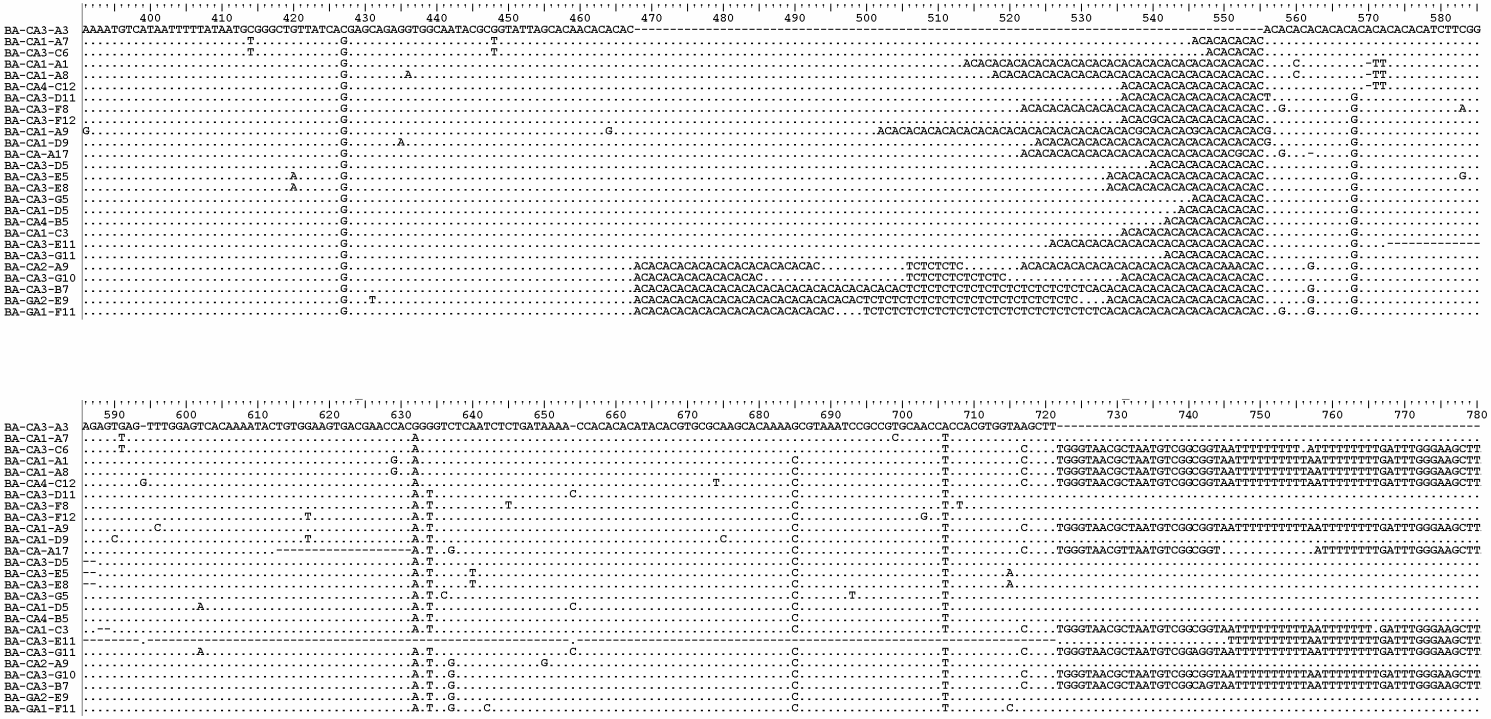
Appendix 3.10A

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  TCGGTTCAGTAGTTTCGGAGCCTATTCAATGCAAACAAACAAACAAACAAACAATC  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.

Chapter 4

Cytogenetic characterization of the butterfly *Bicyclus anynana*[1]

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

ABSTRACT

The chromosome characteristics of the butterfly *Bicyclus anynana*, have received little attention, despite the scientific importance of this species. This study presents the characterization of chromosomes in this species by means of cytogenetic analysis. Lepidoptera possess a female heterogametic W-Z sex chromosome system. The WZ-bivalent in pachytene oocytes of *B. anynana* consists of an abnormally small, heterochromatic W-chromosome with the Z-chromosome wrapped around it. Accordingly, the W-body in interphase nuclei is much smaller than usual in Lepidoptera. This suggests an intermediate stage in the process of secondary loss of the W-chromosome to a ZZ/Z sex determination system. Two nucleoli are present in the pachytene stage associated with an autosome and the WZ-bivalent respectively. Chromosome counts confirmed a haploid number of n=28. This study adds to the knowledge of chromosome structure and evolution of an intensively studied organism. On a broader scale it provides an insight in Lepidoptera sex chromosome evolution.

[1] Chapter 4 and 5 were published combined into one article in PLoS ONE 2008 **3**(12): e3882

INTRODUCTION

The butterfly *Bicyclus anynana* (Nymphalidae, Satyrinae) is among the most extensively studied Lepidoptera species. It has been established as an emerging model organism to address many evolutionary questions with a particular focus on genetic and environmental effects on wing pattern formation (BELDADE and BRAKEFIELD 2002; BELDADE *et al.* 2005; WIJNGAARDEN and BRAKEFIELD 2000), and on life history evolution and ageing (FISCHER *et al.* 2003; MARCUS *et al.* 2004; PIJPE *et al.* 2008; ZIJLSTRA *et al.* 2003). Although this species has received much scientific attention, the physical features of its genome have yet to be described.

Lepidoptera chromosome numbers are usually between 28 to 32 pairs (ROBINSON 1971; SUOMALAINEN 1969), but can vary widely probably as a result of their holokinetic chromosome arrangement (KANDUL *et al.* 2007; TRAUT *et al.* 2007). The most striking examples at the genus level are found in *Agrodiaetus*, with haploid chromosome numbers that vary between 10 and 134 (KANDUL *et al.* 2007). However, geographical intra-specific variability is also commonly observed in Lepidoptera (ROBINSON 1971; WHITE 1973). Geographical subspecies of the silk moth *Samia cynthia* show, besides different chromosome numbers, a high polymorphism of sex chromosomes (YOSHIDO *et al.* 2005), which may play a role in population and species divergence (CHARLESWORTH *et al.* 1987; PROWELL 1998). An extraordinary variation in chromosome numbers, ranging from n=12 to n=88, was reported between populations of a *Philaethria dido* species complex, which is no longer regarded as single species, since no evidence of hybrids between individuals of sympatric populations with different chromosome numbers was found (BROWN *et al.* 1992; SUOMALAINEN and BROWN 1984). The karyotype variation within the genus *Bicyclus* is less spectacular (Appendix 4.1). With the exception of *B. auricrudus* that has a reported haploid chromosome number of 14, all karyotyped species have between 26 and 29 pairs, with n=28 being the predominant count (DE LESSE 1968; DE LESSE and CONDAMIN 1962; DE LESSE and CONDAMIN 1965). However, geographical within-species variation has been observed in *B. funebris* with n=28 in Uganda and n=29 in Senegal (DE LESSE 1968; DE LESSE and CONDAMIN 1962). A haploid chromosome number of 28 was reported in *B. anynana* from Entebbe, Uganda (DE LESSE 1968), but given the geographical variability in Lepidoptera there is need for confirmation since the material used in the present study originates from Nkhata Bay in Malawi, about 1300 km to the south.

Identification of individual chromosomes based on size and banding patterns is difficult in Lepidoptera because of the large number of small and equally sized chromosomes that are not susceptible to banding techniques during mitosis. Much longer meiotic chromosomes in the pachytene stage provide better resolution, but their chromomere patterns are usually not fully distinctive (TANAKA *et al.* 2000; TRAUT *et al.* 2007). In addition, lepidopteran chromosomes are holokinetic, i.e. they lack a distinct primary constriction (the centromere) and spindle microtubules are attached to a large kinetochore plate, which covers significant part of the chromosome surface (WOLF 1996). Thus, the chromosomes cannot be distinguished or characterized by centromere position. The most useful visual characteristics to distinguish lepidopteran chromosomes are the presence of nucleolar organising regions (NORs) associated with nucleoli and heterochromatin of the W chromosome in the sex-chromosome (WZ) pachytene bivalents of females. However, this accounts only for a small fraction of the chromosomes (FUKOVÁ *et al.* 2005; MAREC *et al.* 2001).

METHODS

Spread preparations of pachytene oocytes were obtained following the protocol in (TRAUT 1976) for pachytene mapping. Ovaries of 5th instar larvae were dissected in physiological solution, then fixed for 20 min in Carnoy`s fixative (6 : 3 : 1 ethanol - chloroform - acetic acid), macerated in 60% acetic acid, spread on a slide at 45°C, dehydrated by three washes in increasing concentrations of ethanol (70%, 80%, and 96%, 30s each), and dried at room temperature, leaving the preparations suitable for different types of staining. Some preparations were stained for 5 min and mounted in 2.5% lactic acetic orcein. Others were stained with YOYO-1 fluorescent dye (Molecular Probes Inc., Eugene, OR, USA) under the following conditions: the dry preparations were first soaked for 5 min in PBS (phosphate buffered saline), then stained with 50 µl of 100 nM YOYO-1 in PBS for 20 min, briefly washed in tap water, air-dried and mounted in 20 µl of antifade based on DABCO (1,4-diazabicyclo(2.2.2)-octane; Sigma-Aldrich, St. Louis, MO, USA) (for details, see (MEDIOUNI *et al.* 2004)).

Male metaphase I and II chromosomes were obtained from testes of the 5th instar larvae. The testes were dissected in physiological solution, pretreated in hypotonic solution (0.075M KCl) for 15 min, and then fixed in Carnoy's fixative for 15 minutes. The testes were subsequently squashed in 20 µl of 50% acetic acid using a siliconised cover slip, followed by dehydration in an alcohol series as described above. Staining involved a 5 min incubation in PBS/1% Triton-X, followed by 15 min in PBS/1% Triton-X with 0.25 µg/ml DAPI (4′,6-diamino-2-phenylindole; Sigma-Aldrich). The slides were then rinsed for 5 min in PBS/1% Triton-X with 1% Kodak PHOTO-FLO, followed by 10s rinsing in H_2O containing 1% Kodak PHOTO-FLO. Finally, the preparations were mounted in 20 µl of antifade.

To determine the sex chromatin status (see (TRAUT and MAREC 1996)), preparations of polyploid nuclei were made from Malpighian tubules of 5th instar larvae. The tubules were dissected in physiological solution, fixed in Carnoy's fixative for 2 min, and then stained with 1.5% lactic acetic orcein for 4 min.

RESULTS

Chromosome number: The analysis of metaphase I bivalents and male metaphase II chromosomes in male meiosis, and pachytene bivalents in female meiotic prophase I showed a haploid chromosome number of 28 for *B. anynana* in our stock from Malawi (Fig. 4.1A-C). This is consistent with the findings of (DE LESSE 1968) for *B. anynana* from Uganda and thus, there is no evidence for geographical variation in chromosome numbers in this species. Orcein staining of pachytene bivalents provided the characteristic chromomere pattern that differentiated the chromosomes to a certain level (Fig. 4.1C). However, we did not assign chromosome numbers based on these patterns since it is not clear with which linkage groups they correspond.

Sex chromosomes: Male pachytene spreads displayed 28 bivalents per nucleus that were aligned over their full length. Female pachytene oocytes showed 27 fully-paired bivalents and a pair of sex chromosomes, consisting of a small heterochromatic W chromosome that has a circular arrangement and a Z chromosome that was wrapped around it in the majority of nuclei (Fig 4.1D); in some nuclei, the W chromosome was associated with a terminal segment of the Z chromosome (Fig. 4.1C) or less often with a central part of the Z chromosome (Fig. 4.1E) and formed a short thick rod or a body-like structure. A comparison of the male and female chromosome complements shows that *B. anynana* has a WZ/ZZ (female/male) sex chromosome system, typical for the majority of advanced Lepidoptera (reviewed in (TRAUT *et al.* 2007)).

Large, highly polyploid interphase nuclei of the Malpighian tubules do not form lobes as is seen in some Lepidoptera (cf. (MAREC and TRAUT 1994)), but have oval shapes. In females, each nucleus showed a small heterochromatin W-body (i.e. sex chromatin) that was absent in males (Fig. 4.1F, G). The small size of the W-body was consistent with the tiny W chromosome observed in pachytene oocytes.

Nucleolar organising regions: Two distinct nucleoli were regularly observed in YOYO-1-stained pachytene spreads. One was associated with an autosome bivalent, the other with the WZ bivalent (Fig. 4.1D, H, I). The association with the WZ bivalent is not apparent in Fig. 4.1D and 4.1H since the nucleolus also borders autosomal bivalents, but it was consistent in all examined nuclei. At the end of the autosome bivalent, a pair of YOYO-1-positive dots was immersed into the nucleolus mass. The dots probably composed of heterochromatin were often separated from the main chromosome bodies by a constriction, obviously representing the nucleolus organizing region (NOR) (Fig 4.1H). In orcein-stained pachytenes, two conspicuous chromomeres were seen at the end of this NOR-bivalent (Fig 4.1I). These chromomeres most likely correspond with the two heterochromatic dots highlighted with YOYO-1 (Fig. 4.1D, H).

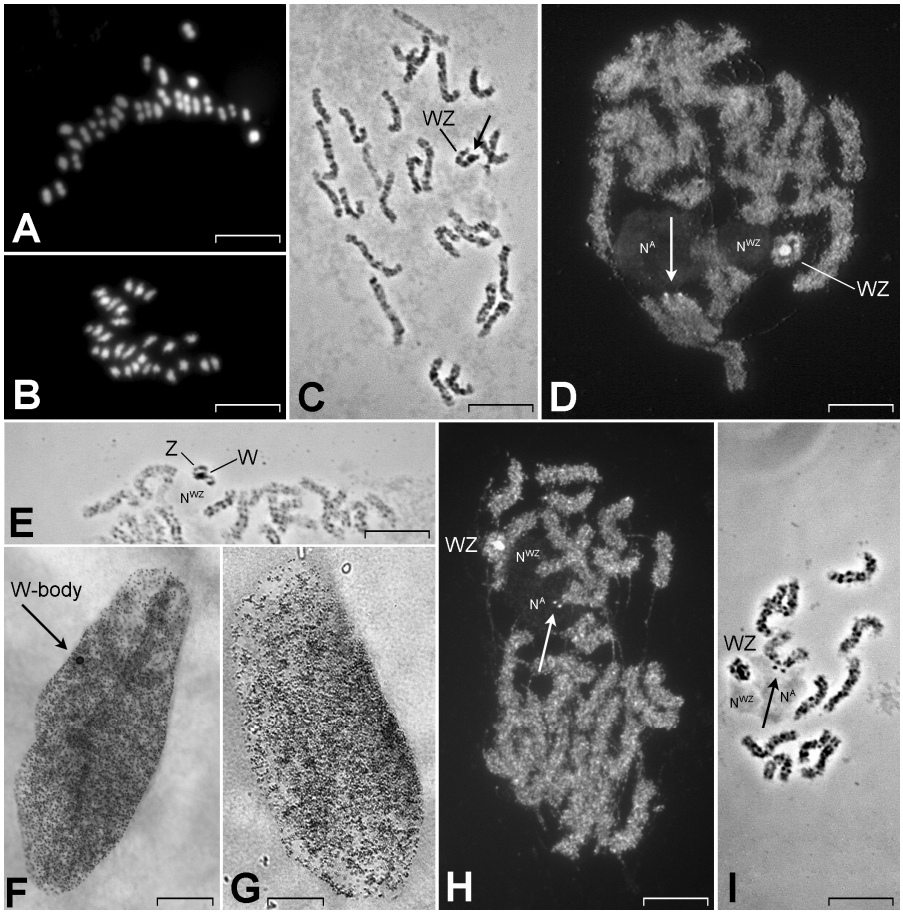


Figure 4.1 Preparations of meiotic cells and somatic interphase nuclei in *Bicyclus anynana*. (A) Squashed DAPI-stained male metaphase I bivalents; (B) squashed DAPI-stained male metaphase II chromosomes; (C) spread orcein-stained female pachytene complement showing chromomere patterns; note the small heterochromatic W chromosome associated with the terminal segment of the Z chromosome (arrow); (D) spread YOYO-1-stained female postpachytene complement showing a curious WZ bivalent, in which the Z chromosome strand is wrapped around the body-like W chromosome, and two nucleoli, one associated with an autosome bivalent (N^A) and the other with the WZ bivalent (N^{WZ}); note small heterochromatin dots (arrow) highlighted with YOYO-1 at the end of each chromosome of the NOR-autosome bivalent; (E) orcein-stained female pachytene spread, showing a WZ-bivalent where the W chromosome is associated with the central part of the Z chromosome; (F) a polyploid nucleus of the female Malpighian tubule cell showing a small sex-chromatin body (arrow), representing multiple copies of the tiny W chromosome; (G) a polyploid nucleus of the male Malpighian tubule cell without sex chromatin; (H) YOYO-1 stained female pachytene spread showing the NOR as stalked dots (arrow) in the nucleolus; (I) orcein-stained female pachytene spread with two conspicuous chromomeres (arrow) within the nucleolus. Scale bars indicate 10 μm in (A-D and H,I) and 50 μm in (F, G).

DISCUSSION

The cytogenetic characteristics of *B. anynana* correspond to those generally found in Lepidoptera. Female heterogamety is confirmed by the presence of a WZ bivalent in pachytene oocytes and the presence of a heterochromatic W-body in female somatic interphase nuclei, which are absent in males. The chromosomes are indistinguishable in different stages of both mitotic and meiotic divisions, except for orcein stained pachytene, where different bivalents can be differentiated to a certain degree. We regularly identified two distinctive bivalents that were associated with two different nucleoli in female pachytene spreads. One of these nucleoli is associated with the WZ bivalent and the other with an autosome bivalent. The autosome bivalent carried a terminally located NOR that was associated with small but clear heterochromatin. The presence of heterochromatin at the NORs is common in animals (e.g. (HIRAI *et al.* 1994; KING *et al.* 1990)) but in Lepidoptera has been reported only in the silkworm *B. mori* (SAHARA *et al.* 2003).

It remained unclear whether the sex-linked NOR of *B. anynana* was located on the W- or on the Z chromosome or on both sex chromosomes since we did not examine pachytene spermatocytes for a comparison. Due to the circular form of the WZ bivalent it was not possible to determine whether the sex-linked NOR is terminal or interstitial. Nevertheless, we favor location of the sex-linked NOR on the Z chromosome as the W chromosome appears composed entirely of heterochromatin, which would inhibit a high transcriptional activity of the active NOR.

The pachytene WZ bivalent of *B. anynana* is exceptional due to the tiny W chromosome. The W chromosome of the oriental tussock moth, *Artaxa subflava* is about half the size of the Z chromosome (YOSHIDO *et al.* 2006) and in the other lepidopteran species examined so far, the W chromosome was either only slightly smaller or comparable in size to the Z chromosome (e.g. (FUKOVÁ *et al.* 2005; TRAUT *et al.* 1999; VÍTKOVÁ *et al.* 2007). Compatible lengths in the pachytene stage of such relatively similar sized W and Z chromosomes undoubtedly facilitate their complete pairing. A regular synaptonemal complex can be formed in spite of their obvious non-homology by means of twisting and synaptic adjustment (MAREC and TRAUT 1994; WEITH and TRAUT 1986). However, the size difference of W and Z is too large in *B. anynana* to form a regular bivalent. Instead, the much longer Z chromosome often forms a circle or horseshoe structure with the W chromosome closed inside. This arrangement could be considered an extreme case of synaptic adjustment as it allows the sex chromosomes to pair along their entire length. A similar mode of pairing was observed in mutants of the flour moth (*Ephestia kuehniella*), in which the W chromosome was shortened by irradiation (TRAUT *et al.* 1986), and also in *A. subflava*, in which the W chromosome comprises about half of the Z chromosome but shows still a conspicuous heterochromatic mass (see Fig. 3 in (YOSHIDO *et al.* 2006)). On the other hand, we cannot exclude that the W and Z chromosomes pair by means of some sequence homology, for example, in telomeric regions or via rDNA in the case of shared NORs. The *B. anynana* W chromosome is composed of constitutive heterochromatin as in many other Lepidoptera. This observation, combined with recent findings on the composition of W chromosomes in *B. mori*, *C. pomonella*, and several pyralids (ABE *et al.* 2005; FUKOVÁ *et al.* 2007; SAHARA *et al.* 2003; VÍTKOVÁ *et al.* 2007), suggests that the *B. anynana* W chromosome is probably gene-poor and rich in interspersed repetitive sequences, such as transposable elements, which are known to be abundant in *B. anynana* in general (VAN'T HOF *et al.* 2007). The small size of the W chromosome is also reflected by a small heterochromatin body in

Malpighian tubule nuclei of females. The size could indicate an intermediate stage in the process of secondary loss of the W chromosome as is the case in Lepidoptera that have adopted a ZZ/Z sex determination system after loss of the W chromosome (TRAUT *et al.* 2007).

REFERENCES

ABE, H., K. MITA, Y. YASUKOCHI, T. OSHIKI and T. SHIMADA, 2005 Retrotransposable elements on the W chromosome of the silkworm, *Bombyx mori*. Cytogenet Genome Res **110:** 144-151.

BELDADE, P., and P. M. BRAKEFIELD, 2002 The genetics and evo-devo of butterfly wing patterns. Nature Reviews Genetics **3:** 442-452.

BELDADE, P., P. M. BRAKEFIELD and A. D. LONG, 2005 Generating phenotypic variation: prospects from "evo-devo" research on *Bicyclus anynana* wing patterns. Evol. Dev. **7:** 101-107.

BROWN, K. S., T. EMMEL, P. J. ELIAZAR and E. SUOMALAINEN, 1992 Evolutionary patterns in chromosome numbers in neotropical Lepidoptera. I. Chromosomes of the Heliconiini (family Nymphalidae: subfamily Nymphalinae). Hereditas **117:** 109-125.

CHARLESWORTH, B., J. A. COYNE and N. H. BARTON, 1987 The Relative Rates of Evolution of Sex Chromosomes and Autosomes. The American Naturalist **130:** 113.

DE LESSE, H., 1968 Formules chromosomiques de Lépidoptères Rhopalocères d'Uganda et du Kenya. Ann. Soc. Ent. Fr. (N.S.) **4:** 581-599.

DE LESSE, H., and M. CONDAMIN, 1962 Formules chromosomiques de quelques Lépidoptères Rhopalocères du Sénegal. Bulletin de l'I.F.A.N. **24:** 464-473.

DE LESSE, H., and M. CONDAMIN, 1965 Formules chromosomiques de quelques Lépidoptères Rhopalocères du Sénegal et de Côte d'Ivoire. Bulletin de l'I.F.A.N. **27:** 1089-1094.

FISCHER, K., E. EENHOORN, A. N. M. BOT, P. M. BRAKEFIELD and B. J. ZWAAN, 2003 Cooler butterflies lay larger eggs: developmental plasticity versus acclimation. Proc R. Soc. B. **270:** 2051-2056.

FUKOVÁ, I., P. NGUYEN and F. MAREC, 2005 Codling moth cytogenetics: karyotype, chromosomal location of rDNA, and molecular differentiation of sex chromosomes. Genome **48:** 1083-1092.

FUKOVÁ, I., W. TRAUT, M. VÍTKOVÁ, P. NGUYEN, S. KUBÍČKOVÁ *et al.*, 2007 Probing the W chromosome of the codling moth, *Cydia pomonella*, with sequences from microdissected sex chromatin. Chromosoma **116:** 135-145.

HIRAI, H., M. T. YAMAMOTO, K. OGURA, Y. SATTA, M. YAMADA *et al.*, 1994 Multiplication of 28S rDNA and NOR activity in chromosome evolution among ants of the *Myrmecia pilosula* species complex. Chromosoma **103:** 171-178.

KANDUL, N. P., V. A. LUKHTANOV and N. E. PIERCE, 2007 Karyotypic diversity and speciation in *Agrodiaetus* butterflies. Evolution **61:** 546-559.

KING, M., N. CONTRERAS and R. L. HONEYCUTT, 1990 Variation within and between nucleolar organizer regions in Australian hylid frogs (Anura) shown by 18S + 28S in-situ hybridization. Genetical Research **80:** 17-29.

MARCUS, J. M., D. M. RAMOS and A. MONTEIRO, 2004 Germline transformation of the butterfly *Bicyclus anynana*. Proc. R. Soc. Lond. B **271 Suppl 5:** S263-S265.

MAREC, F., A. TOTHOVÁ, K. SAHARA and W. TRAUT, 2001 Meiotic pairing of sex chromosome, fragments and its relation to atypical transmission of a sex-linked marker in *Ephestia kuehniella* (Insecta : Lepidoptera). Heredity **87:** 659-671.

MAREC, F., and W. TRAUT, 1994 Sex chromosome pairing and sex chromatin bodies in W-Z translocation strains of *Ephestia kuehniella* (Lepidoptera). Genome **37:** 426-435.

MEDIOUNI, J., I. FUKOVÁ, R. FRYDRYCHOVÁ, M. H. DHOUIBI and F. MAREC, 2004 Karyotype, sex chromatin and sex chromosome differentiation in the carob moth, *Ectomyelois ceratoniae* (Lepidoptera: Pyralidae). Caryologia **57:** 184-194.

PIJPE, J., P. M. BRAKEFIELD and B. J. ZWAAN, 2008 Increased life span in a polyphenic butterfly artificially selected for starvation resistance. Am Nat. **171:** 81-90.

PROWELL, D. P., 1998 Sex linkage and speciation in Lepidoptera in *Endless forms: species and speciation*, edited by D. J. HOWARD and S. H. BERLOCHER. Oxford University Press, New York.

ROBINSON, R., 1971 *Lepidoptera Genetics.* Pergamon Press, Oxford, UK.

SAHARA, K., A. YOSHIDO, N. KAWAMURA, A. OHNUMA, H. ABE *et al.*, 2003 W-derived BAC probes as a new tool for identification of the W chromosome and its aberrations in *Bombyx mori*. Chromosoma **112:** 48–55.

SUOMALAINEN, E., 1969 Chromosome evolution in the Lepidoptera. Chromosomes Today **2:** 132-138.

SUOMALAINEN, E., and K. S. J. BROWN, 1984 Chromosome number variation within *Philaethria* butterflies (Lepidoptera: Nymphalidae, Heliconiini). Chromosoma **90:** 170-176.

TANAKA, N., T. YOKOYAMA, H. ABE, K. TSUCHIDA, O. NINAGI *et al.*, 2000 Cytogenetic analysis shows that the unusually large chromosome in the sex-limited pB silkworm (*Bombyx mori*) strain consists of three chromosomes. Hereditas **133:** 95-103.

TRAUT, W., 1976 Pachytene mapping in the female silkworm, *Bombyx mori* L. (Lepidoptera). Chromosoma **58:** 275-284.

TRAUT, W., and F. MAREC, 1996 Sex chromatin in Lepidoptera. Q. Rev. Biol. **71:** 239-256.

TRAUT, W., K. SAHARA and F. MAREC, 2007 Sex Chromosomes and Sex Determination in Lepidoptera. Sexual Development **1:** 332–346.

TRAUT, W., K. SAHARA, T. D. OTTO and F. MAREC, 1999 Molecular differentiation of sex chromosomes probed by comparative genomic hybridization. Chromosoma **108:** 173-180.

TRAUT, W., A. WEITH and G. TRAUT, 1986 Structural mutants of the W chromosome in *Ephestia* (Insecta, Lepidoptera). Genetica **70:** 69-79.

VAN'T HOF, A. E., P. M. BRAKEFIELD, I. J. SACCHERI and B. J. ZWAAN, 2007 Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera. Heredity **98:** 320-328.

VÍTKOVÁ, M., I. FUKOVÁ, S. KUBÍČKOVÁ and F. MAREC, 2007 Molecular divergence of the W chromosomes in pyralid moths (Lepidoptera). Chromosome Res. **15:** 917-930.

WEITH, A., and W. TRAUT, 1986 Synaptic adjustment, non-homologous pairing, and non-pairing of homologous segments in sex chromosome mutants of *Ephestia kuehniella* (Insecta, Lepidoptera). Chromosoma **94:** 125-131.

WHITE, M. J. D., 1973 *The chromosomes*. Chapman & Hall, London.

WIJNGAARDEN, P. J., and P. M. BRAKEFIELD, 2000 The genetic basis of eyespot size in the butterfly *Bicyclus anynana*: an analysis of line crosses. Heredity **85:** 471-479.

WOLF, K. W., 1996 The structure of condensed chromosomes in mitosis and meiosis of insects. Int J Insect Morphol Embryol **25:** 37-62.

YOSHIDO, A., F. MAREC and K. SAHARA, 2005 Resolution of sex chromosome constitution by GISH and telomere-FISH in some species of Lepidoptera. Chromosoma **114:** 193-202.

YOSHIDO, A., Y. YAMADA and K. SAHARA, 2006 The W chromosome detection in several Lepidopteran species by genomic *in situ* hybridization (GISH). Journal of insect biotechnology and sericology **75:** 147-151.

ZIJLSTRA, W. G., M. J. STEIGENGA, P. M. BRAKEFIELD and B. J. ZWAAN, 2003 Simultaneous selection on two fitness-related traits in the butterfly *Bicyclus anynana*. Evolution **57:** 1852–1862.

APPENDIX 4.1

Haploid chromosome numbers of different *Bicyclus* species and their geographical origin

species	n	Origin	Reference
B. auricrudus parvocellata	*ca* 14[1]	Entebbe (Uganda)	DE LESSE 1968
B. jefferyi	28	Entebbe (Uganda)	DE LESSE 1968
B. sophrosyne	28	Entebbe (Uganda)	DE LESSE 1968
B. mollita	28	Entebbe (Uganda)	DE LESSE 1968
B. safitza	28	Entebbe (Uganda)	DE LESSE 1968
B. dentatus	29	Fort Portal (Uganda); Katamayo Forest (Kenya)	DE LESSE 1968
B. anynana	28	Entebbe (Uganda)	DE LESSE 1968
B. funebris agraphis	28	Entebbe (Uganda)	DE LESSE 1968
B. funebris funebris	29	Forêt de Tobor (Senegal)	DE LESSE and CONDAMIN 1962
B. saussurei	28	Entebbe (Uganda)	DE LESSE 1968
B. zinebi	26	Forêt de Santiaba-Mandjak (Senegal)	DE LESSE and CONDAMIN 1965
B. vulgaris	28	Lamto (Ivory Coast); Forêt de Tobor (Senegal)	DE LESSE and CONDAMIN 1965 DE LESSE and CONDAMIN 1962
B. sandace	28	Sangalkam (Senegal)	DE LESSE and CONDAMIN 1962

[1] DE LESSE 1968: "Quant à *B. auricrudus*, à n = *ca* 14, il représente sans doute, comme ailleurs (cf. *Eurema brigitta*, p 592), un cas de réunion de chromosomes." This translates to: As for *B. auricrudus*, with n = *ca* 14, it represents without a doubt, as in others (cf. *Eurema brigitta*, p 592), a case of chromosome fusion.

Chapter 5

An AFLP-based genetic linkage map for the butterfly *Bicyclus anynana*, covering all 28 chromosomes[1]

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

ABSTRACT

The butterfly *Bicyclus anynana* is emerging as a model organism in the field of evo-devo, life history evolution and ageing. However, there are limitations to the detail in which these traits can be unraveled without basic knowledge about genome architecture. A genetic linkage map allows separate components of traits to be identified and linked to distinct chromosomal regions and to study their effect in more detail in a context of functionality and interaction. Therefore, a linkage map was constructed for *B. ananynana* to serve as a resource to facilitate future investigations. Linkage mapping had to take account of absence of crossing-over in female Lepidoptera, and of our use of a full-sib crossing design. We developed a new method to determine and exclude the non-recombinant uninformative female inherited component in offspring. The linkage map was constructed using a novel approach that uses exclusively JOINMAP-software for Lepidoptera linkage mapping. This approach simplifies the mapping procedure, avoids over-estimation of mapping distance and increases the reliability of relative marker positions. A total of 347 AFLP markers, 9 microsatellites and one single-copy nuclear gene covered all 28 chromosomes, with a mapping distance of 1354 cM. Conserved synteny of *Tpi* on the Z-chromosome in Lepidoptera was confirmed for *B. anynana*. The results are discussed in relation to other mapping studies in Lepidoptera. This study contributes to the knowledge of chromosome structure and evolution of an intensively studied organism. On a broader scale it provides an insight in Lepidoptera sex chromosome evolution and it proposes a simpler and more reliable method of linkage mapping than used for Lepidoptera to date.

[1] Chapter 4 and 5 were published combined into one article in PLoS ONE 2008 **3**(12): e3882.

INTRODUCTION

Understanding genetic traits greatly benefits from a general knowledge about genome structure. This is commonly achieved by the construction of linkage maps with genetic or morphological markers that co-segregate per chromosome (i.e. linkage group) and are spaced thereon based on recombination events. Linkage maps can form the basis to reveal fundamental characteristics of a gene such as where it is, what it is, what it does, and how it interacts with other genes. There are two important benefits of linkage maps: Firstly, it provides a position for loci relative to other markers and it assigns them to a specific chromosome. Secondly, it allows exploring the segregation of markers, or combinations thereof in relation to phenotypic traits in greater detail. This is particularly useful when little is known about the genetic mechanism of a trait because linked neutral markers can identify the region that harbors an unknown gene with a specific phenotypic effect. The absence of a linkage map for the butterfly *Bicyclus anynana* has biased the analysis of traits towards genes with known function in other insects. A linkage map for *B. anynana* would provide the means to investigate all components that are relevant in this species rather than a preselected subset and would thus open up new directions in ongoing and in novel research.

Despite the abundance of lepidopteran species and their economical relevance, linkage maps are currently available for only six species. One reason for this is that the generally large number of chromosomes in this taxon requires a relatively large number of markers to cover all chromosomes with sufficient density. Additionally, a substantial part of the polymorphisms in the offspring cannot be used for positional mapping since the maternally transmitted markers are non-recombinant in Lepidoptera. The maternally transmitted markers obscure a large part of the paternally transmitted genotypes when using dominant markers, resulting in an even greater loss of information (JIGGINS *et al.* 2005; KAPAN *et al.* 2006). The most detailed linkage information in Lepidoptera comes from the domesticated silkworm *Bombyx mori*, for which a number of linkage maps have been constructed based on RAPD (PROMBOON *et al.* 1995; YASUKOCHI 1998), RFLP (SHI *et al.* 1995), AFLP (TAN *et al.* 2001), microsatellites (MIAO *et al.* 2005), and BAC sequences (YAMAMOTO *et al.* 2006; YAMAMOTO *et al.* 2008; YASUKOCHI *et al.* 2006). In addition, all genetic linkage groups (LGs) were successfully assigned to individual chromosomes in this species (YOSHIDO *et al.* 2005). The other lepidopteran linkage maps have been constructed for *Heliconius melpomene* (JIGGINS *et al.* 2005), *H. erato* (KAPAN *et al.* 2006; TOBLER *et al.* 2005), *Colias eurytheme-C. philodice* hybrid (WANG and PORTER 2004), *Ostrinia nubilalis* (DOPMAN *et al.* 2004) and *Plutella xylostella* (HECKEL *et al.* 1999) based on RFLP, AFLP, microsatellites, allozymes and single copy nuclear genes.

When using a cross with dominant markers such as AFLP's, the general approach in Lepidoptera mapping procedures is to divide the offspring marker data into three groups based on the F_1 marker genotypes. Markers that are heterozygous in both F_1 parents segregate in the F_2 with a 3:1 Mendelian ratio. Markers that are recessive homozygous in the F_1 male and heterozygous in the F_1 female have a 1:1 ratio in the F_2 offspring. These markers are used for LG assignment and for identification and exclusion of the uninformative female-inherited component in the 3:1 markers. The markers that are recessive homozygous in the F_1 female and heterozygous in the F_1 male also have a 1:1 ratio in the offspring. These markers, combined with the male-inherited component of the 3:1 marker genotypes, are used for constructing the final linkage map (JIGGINS *et al.* 2005; KAPAN *et al.* 2006; TOBLER *et al.* 2005).

When using only the 3:1 markers, the outcome is a linkage map with two LGs per chromosome (2n LGs). The two sets of homologous LGs are incompatible and can only be combined with anchoring markers. Male informative markers, allelic AFLPs and microsatellites can act as such anchors and there are various approaches to integrate the two sets of dominant markers. For example, Lepidoptera specific software was designed to create a linkage map for *B. mori* because it was argued that MAPMAKER 3.0 (LANDER *et al.* 1987) is unsuitable for this purpose (SHI *et al.* 1995). In other studies, the final step is performed with MAPMAKER 3.0, allthough in some cases the preceding steps were done in JOINMAP 3.0 or specifically designed programs (HECKEL *et al.* 1999; JIGGINS *et al.* 2005; KAPAN *et al.* 2006; TOBLER *et al.* 2005). Alternatively, the LGs in repulsion were presented as two different sets (PROMBOON *et al.* 1995; YASUKOCHI 1998), or one integrated set that was based on the average distances of anchoring markers (YASUKOCHI *et al.* 2006).

Here we report on a novel approach for the final step in Lepidoptera linkage maping by using the option in JOINMAP to join maps, i.e. to present the two opposite phased homologous maps as different mapping populations and use the software to integrate them based on the anchoring markers. The advantages are that the female-derived component can be removed instead of presented as missing data, and the same software combines the two phases automatically. To compare our mapping distance with that of other species of butterfly, we also performed a MAPMAKER analysis because Mapping distances generated by the two programs can differ substantially. In general, these differences are caused by the different algorithms that are used. MAPMAKER determines the mapping distance based on maximum likelihood multipoint estimates, while JOINMAP uses linear regression of pairwise distances. Additionally, when using dominant markers in species with only one recombining sex, the manner in which the uninformative part of the data are treated also has an effect on mapping distance.

METHODS

Linkage analysis and map construction

Cross design: The linkage analysis was based on a cross between individuals from divergent selection lines for eyespot size on the ventral hindwing, designated High (H) and Low (L) for large and small eyespots respectively (BRAKEFIELD *et al.* 1996; WIJNGAARDEN and BRAKEFIELD 2000). An H-female was mated to an L-male (P generation), and subsequently, 15 full-sib F_1 crosses were set up by combining random brothers and sisters to produce segregating F_2 offspring. The larvae were raised on maize plants and the adults were fed with banana. They were reared at 23ºC to minimize the effect of temperature on eyespot size, since this temperature is an intermediate between the temperature that would produce small (20°C) and large (27°C) eyespots as a result of phenotypic plasticity. The cross that produced the largest amount of F_2 adults was selected to produce the linkage map. All procedures have been performed following our institutional animal husbandry guidelines. From a total offspring of 71 males and 113 females, 23 individuals from both ends of the phenotypic extremes of the F_2 generation were genotyped in each sex (i.e. 92 F_2 individuals in total). DNA was extracted from half a thorax using DNeasy tissue spin columns (Qiagen GmbH, Hilden, Germany).

AFLP: We followed a modified procedure of the AFLP technique (VOS *et al.* 1995). Digestion and ligation were performed simultaneously for two hours at 37°C in 25 µl 1× T4 ligase buffer containing 1.2 units of both *Mse*I and *Eco*RI (NEB, Ipswich, MA, USA), 0.612 µM Mse-adapter (5'-GACGATGAGTCCTGAG-3' + 5'-TACTCAGGACTCAT-3'), 0.068 µM Eco-adapter (5'-CTCGTAGACTGCGTACC-3' + 5-AATTGGTACGCAGTCTAC-3'), 0.6 Weiss units T4 Ligase, 2.5 µg BSA and 5 µl DNA extract from the 2nd Qiagen DNeasy tissue kit elution (corresponding to approximately 125 ng DNA).

Preamplification was performed in 15 µl 1 × AFLP Amplification Core Mix Module (Applied Biosystems, Foster City, CA, USA) supplied with 0.12 µM Eco+A primer (5'-GACTGCGTACCAATTCA-3'), 0.92 µM Mse+C primer (5'-GATGAGTCCTGAGTAAC-3'), and 2 µl undiluted restriction-ligation product as template. Preamplification PCR cycle was 120s 72°C, 120s 94°C, followed by 20 cycles of 10s 94°C, 30s 56°C, 120s 72°C.

Selective amplifications with 33 different primer combinations were processed in 10µl 1 × Core Mix with 0.05 µM fluorescently labeled Eco+ANN primer, 0.25 µM Mse+CNN primer and 1 µl 10 × diluted preamplified product as template. For sequence and fluorescent labels of the primers see Table 5.1. Amplification was performed with 60s 94°C, then 9 cycles of 10s 94°C, 30s T_a (annealing temperature), 120s 72°C, with T_a decreasing 1°C per cycle from 65°C down to 57°C. Then 25 cycles of 10s 94°C, 30s 56°C, 120s 72°C, and a final extension of 30 min at 72°C. Twelve of the combinations were genotyped on an ABI 377 automated sequencer with 3 different dyes and ROX500 size standard, and an additional 21 on an ABI 3100 with 4 dyes and LIZ500 size standard. The ABI377 data output was analyzed with GENOGRAPHER 1.6.0 (BENHAM *et al.* 1999) and the ABI3100 generated data with GENOTYPER 3.6. (Applied Biosystems). We use the term "peakpresent" to indicate an AFLP amplicon that shows up as a peak on capillary fragment analysis systems and which is either homozygous or heterozygous and "peakabsent" for the recessive homozygote.

Microsatellites: The microsatellite markers available for this species were processed under the conditions described in (VAN'T HOF *et al.* 2005), except in this case they

were amplified with NED, PET, 6-FAM or HEX modified fluorescent primers, run with LIZ-500 size standard on an ABI 3700 fragment analysis instrument and analysed with Genotyper 3.6 (primers, size standard, software and ABI 3700 from Applied Biosystems).

***Tpi* genotyping:** RNA was extracted from ground thorax with TRIZOL (Invitrogen, Carlsbad, CA, USA) following the methods suggested by the manufacturer. cDNA was synthesized with SUPERSCRIPT III (Invitrogen) with 50 ng template and a T_{17} primer under standard conditions. A section of the *Tpi* (Triose-phosphate isomerase) gene was amplified with arthropod-specific degenerate primers *197fin*1F and *197fin*2R (REGIER 2005). PCR was performed in 1 × Amplitaq PCR buffer I, 0.6 units Amplitaq Gold polymerase (buffer and polymerase supplied by Applied Biosystems) 0.4 µM of each primer, 0.2 µM dNTP and 1 µl of the cDNA in a final volume of 20 µl. The PCR cycle was 9 min 94°C, then 35 cycles of 30s 94°C, 30s 50°C, 45s 72°C. The PCR product was purified with EXOSAP-IT (Amersham plc, Little Chalfont, UK), sequenced with the BigDye 3.1 kit (Applied Biosystems), and analyzed on an ABI 3100 sequencer (Applied Biosystems). Gene-specific primers Ba_TPI_207U (TTCGGCTGAGATGATAAAGG) and Ba_TPI_473L (AGTACCAATGGCCCACACTG) were designed within the *Tpi* sequence to amplify an intronic region, using the same genomic template as for the AFLP reactions. PCR conditions were as described above, except for using T_a 52°C instead of 50°C. The F_1 parents were screened for SNPs (single nucleotide polymorphisms) by means of sequencing the intron. Genotyping the F_2 offspring was based on PCR amplification (as above), the amplicons were subsequently treated with 1 unit of *Alu*I restriction enzyme (NEB) for 2h at 37°C, which either cuts a 230bp fragment into 30bp and 200bp or leaves it intact, depending on the genotype. The restriction pattern was screened on a 3% agarose gel. The *Tpi* partial cds and intron sequence are submitted to GENBANK under accession numbers EU675861 and EU675862.

Data sorting into FI, MI, BI, and sex-linked markers: The AFLP markers were divided into different groups, depending on the F_1 genotypes. Female informative (FI) markers are present in the F_1 female and absent in the F_1 male and segregate 1:1, male informative (MI) markers are present in the F_1 male and absent in the F_1 female and segregate 1:1 as well. BI (both informative) markers segregate with a 3:1 ratio, resulting from F_1 male and female that are both heterozygous peakpresent. Z-linked markers were identified by a peakpresent in all male offspring and a 1:1 ratio in the female offspring (representing an F_1 WZ^+ (♀) × Z^+Z^- (♂) F_1 cross, with "+" = peakpresent allele and "–" = peakabsent allele). All F_2 female MI-markers were compared with this Z-specific 1:1 pattern in JOINMAP to reveal the WZ^- × Z^+Z^- crosses in which both male and female offspring have a 1:1 ratio.

Identification of chromosome prints: Due to the absence of meiotic recombination in females, syntenic FI-markers are transmitted to the offspring in complete association, independent of their relative position. Consequently, they cannot be positioned within LGs. A cluster of syntenic FI-markers displays a chromosome-specific pattern of F_2 genotypes, which is identical for all loci on the same chromosome and which displays the exact opposite pattern in all markers in repulsion. This fixed set of genotypes has been named the "chromosome print" (YASUKOCHI 1998). The number of chromosome prints per individual equals the haploid autosomal chromosome number. Their identification in *B. anynana* was carried out as described in (JIGGINS *et al.* 2005), by grouping the FI-markers together with JOINMAP under strict conditions (LOD >8), allowing just a small number of genotyping errors. The linkage phase describes on which parental (F_1) chromosome the peakpresent of a

marker lies and from which grandparent (P-generation) it came. If the marker is present in the grandmother and absent in the grandfather, the linkage phase is "0", the reverse gives linkage phase "1". When the marker is present in both grandparents, the linkage phase is determined by the software based on co-segregation in the F_2 with markers for which linkage phases are known. Linkage phases consist of a maternal and a paternal component, indicating marker orientation (and P-origin) in the F_1 mother and the F_1 father.

The chromosome prints were numbered based on the output order of the software. It is important to reduce the number of errors in chromosome prints to a minimum because they are subsequently used for error detection and identification and removal of uninformative markers. With multiple FI-markers defining a chromosome print, inconsistencies were rescored and when persistent, the chromosome print was based on the most common genotype in the inconsistent individual.

Chromosome prints for chromosomes without FI-markers were reconstructed based on BI and MI-markers as described in Appendix 5.1. This was done after the LG assignment described below. In addition, ten LGs with available (FI-based) chromosome prints were also reconstructed in this way to validate the reconstruction technique.

BI and microsatellite linkage group assignment: BI and microsatellite markers were grouped by screening them against the 21 chromosome print patterns in JOINMAP with a LOD threshold of 3. This mapping step also established the linkage phase of the markers. Markers in the six LGs for which chromosome prints were initially not available were assigned to LG22-LG27.

The markers were subsequently screened by a "forbidden genotype" analysis to confirm or reject correct LG assignment and to detect scoring errors (HECKEL *et al.* 1999; SHI *et al.* 1995). This procedure is based on the fact that certain marker combinations within an individual cannot occur because it would involve recombination in females. This screening procedure is explained in more detail in Appendix 5.2. The threshold to exclude markers from further analysis was set to three or more forbidden genotypes.

Identification of allelic (codominant) AFLPs: Part of the observed variation in AFLP data is caused by indels (insertions or deletions) between the two restriction sites at a single locus, resulting in amplicons of different sizes. To determine whether two BI loci are in fact different alleles of the same locus, we applied the following criteria: (1) they must have the same primer combination; (2) they must group together in the same LG when presented as independent loci in the initial uncensored BI screening; and (3) linkage phases of markers with 3:1 ratio must be opposite for both the maternal and the paternal component. Either one or both peaks present in an individual would be a prerequisite for codominance in species with recombination in both sexes, but with non-recombining females, that same condition is already covered by forbidden genotype restrictions.

MI alleles were detected as well, but they do not provide more analytical power when combined together into one codominant marker as is the case in the BI-markers. Their opposite paternal linkage phases produce fully complementary peak patterns that hold the same mapping information.

Censoring of female-derived BI-markers in the F_2: The BI-markers (with a 3:1 ratio in the offspring) obtain half their peakpresent alleles from the F_1 mother and the other half from the F_1 father. A female derived peakpresent obscures the male-derived allele in dominant markers, so that it is impossible to distinguish between F_2 homozygotes and heterozygotes. This is not an issue when mapping species with

recombination in both sexes, because mapping software can treat these unknown allele combinations as "either heterozygous or homozygous". However, without recombination in the females, genotype scores that have a positive F_1 female signal have to be excluded from analysis, which means that part of the paternal information is also lost. What remains are scores for those individuals that obtained a peakabsent from the female and either a present or absent from the male in a 1:1 ratio. The criteria for filtering out the female component is straightforward because the female BI peakpresent is always fully linked to either a positive or negative chromosome print value, depending on their relative maternal linkage phases (Appendix 5.3). Markers from individuals with a positive chromosome print value must be removed when they have the same maternal linkage phase as the chromosome print, and markers in repulsion with the chromosome print must be removed in the remainder of the individuals.

Assignment of linkage groups for MI-markers: The censored BI genotypes are initially replaced with "missing data". The BI and microsatellite markers with their LG designations are then analyzed together with the MI and microsatellite markers in JOINMAP to establish to which LGs they belong.

Final map construction: Microsatellites were translated to their male informative component as described in Appendix 5.4, resulting in MI-markers with a 1:1 ratio. These were then combined with the MI- and censored BI-markers for each separate chromosome. Each chromosome set was then divided in two subsets, based on their chromosome print values (Appendix 5.3). The BI markers in these two subsets are of opposite maternal linkage phase as a result of the exclusion of the censored BI genotypes. All the subsets were individually presented to JOINMAP for linkage map construction. Subsequently, the sets of linkage maps representing the same chromosomes with suitable anchoring markers are combined with the "Combine groups for map integration" command in JOINMAP. The remaining sets (without anchoring markers) remain as separate LGs. The integration of the two subsets is represented schematically in Appendix 5.3.

The Z chromosome markers were divided into male- and female F_2 offspring. The female F_2 offspring have a 1:1 ratio for all markers, while the F_2 males have 100% peakpresent when the F_1 female is also peakpresent. These 100% male scores were excluded from analysis and all the female markers and the remaining male markers were separately mapped and then joined as described above.

Comparison between JOINMAP and MAPMAKER: Besides the linkage map construction with JOINMAP, we followed the procedures described in (KAPAN *et al.* 2006) for constructing a linkage map with MAPMAKER 3.0.

All steps except the "Final map construction" were identical to the procedures described above, since (KAPAN *et al.* 2006) used JOINMAP for that part of the analysis. The main difference from the JOINMAP approach in this final step is that the censored BI-markers were replaced by "missing data" rather than excluded, and that the markers belonging to the same LGs were analysed together instead of in two separate groups. For LGs without sufficient anchoring markers, the subgroups with the largest mapping distance were compared.

Table 5.1 AFLP primer combinations and fluorescent dyes.
The first column contains the different *Mse*I-based primers used. The next four columns contain the fluorescently labeled *Eco*RI-based primers that were used in combination with the *Mse*I-based primer within the same row. The primers are 19bp in length and consist of a 16bp core sequence and a 3bp extension. "m" is short for a GATGAGTCCTGAGTAA core sequence and "e" stands for a GACTGCGTACCAATTC core sequence. "m" and "e" are followed by the three base extensions that differentiate them. The colors of the fluorescent labels of the *Eco*RI-based primers are presented in the column headers, and the fluorescent 5' modifications in the cells below them (5-FAM, 6-FAM, JOE, VIC, NED and PET). Individual AFLP markers in Fig. 5.1 & 5.2 are characterized by the eNNN-mNNN combinations shown in this table and the PCR product size. The final column describes which fragment analysis instrument was used.

*Mse*I-based primer	*Eco*RI-based primer blue	*Eco*RI-based primer green	*Eco*RI-based primer yellow	*Eco*RI-based primer red	Instrument
mCAA	eACA 5-FAM	eACC JOE	eAAC NED	not used	ABI 377
mCAC	eACA 5-FAM	eACC JOE	eAAC NED	not used	ABI 377
mCAT	eACA 5-FAM	eACC JOE	eAAC NED	not used	ABI 377
mCGC	eACA 5-FAM	eACC JOE	eAAC NED	not used	ABI 377
mCAG	eACA 6-FAM	eAAC VIC	eACC NED	not used	ABI 3100
mCGA	eACA 6-FAM	eAAC VIC	eACC NED	not used	ABI 3100
mCGG	eACA 6-FAM	eAAC VIC	eACC NED	eACG PET	ABI 3100
mCGT	eACA 6-FAM	eAAC VIC	not used	eACG PET	ABI 3100
mCTC	eACA 6-FAM	eAAC VIC	eACC NED	eACG PET	ABI 3100
mCTG	eACA 6-FAM	eAAC VIC	eACC NED	eACG PET	ABI 3100

RESULTS

Genetic markers: A total number of 458 polymorphic segregating markers was generated with AFLPs. The effective number was smaller because the female informative markers do not contribute to mapping, a small number of markers failed the forbidden genotype screening, and 52 markers that behaved as alleles were merged to form 26 single locus codominant markers. This resulted in 347 AFLP loci that could be used for the construction of the linkage map. The markers cover all chromosomes except for the W chromosome, which cannot be mapped even if markers were available because this chromosome not involved in recombination. Additionally, there were seven polymorphic microsatellites that could be positioned on the map and another two that could only be assigned to specific LG's by their female informative component because they were homozygous in the F_1 male. This number is far lower than the number of microsatellite loci available for *B. anynana* because many were not informative in the P-generation to start with, and other loci inherited an uninformative set of alleles from the P-generation to the F_1 due to the bottleneck conditions of the full-sib cross design. The *Alu*I digestion of the genomic *Tpi* amplicons gave a restriction pattern in male F_2 offspring of either a 230 bp fragment, a 200 bp (and a 30 bp) fragment, or both of them within the same individual, thus representing both homozygotes and the heterozygote. Female F_2 offspring had either the 230 bp or the 200 bp fragment (but not both) per individual, thereby showing a hemizygous (Z-linked) pattern.

Chromosome prints based on FI-markers were available for 21 of the 27 autosomes, another three were reconstructed from BI and MI-markers (LGs 22, 25, 27) and the remaining three were based on BI-markers alone (LGs 23, 24, 26), with random 1:1 designation for the unassigned values as described in Appendix 5.1. The empirical verification of the BI+MI based reconstruction for LGs with chromosome prints already available gave an exact match between "chromosome print" and "reconstructed chromosome print" in eight out of 10 cases, one with a single error and one with three, giving a total of only four inconsistent values out of 920. The verification of difference between the BI-only reconstructed maps and the actual maps (performed on the same 10 control LGs) showed a deviation of 2 cM at most for the entire mapping distance, and the markers always remained in the same order.

The final linkage map is shown in Fig. 5.1 and 5.2. Twenty chromosomes had sufficient anchoring markers to create integrated LGs following the procedures outlined in Appendix 5.3. Eight chromosomes had either one or no anchoring markers (chromosomes 11, 12, 14, 17, 20, 22, 23, 24), which prevented integration. These are represented in Fig. 5.1 and 5.2 as separate linkage groups per chromosome with unknown position and orientation relative to each other. These two subsets represent markers available from the high and low eyespot selection lines respectively. The Z chromosome contains 18 evenly dispersed markers and the *Tpi* gene. The mapping lengths of the chromosomes range from 8 to 84 cM, but we assume that the smaller linkage groups have insufficient coverage rather than representing chromosomes that are relatively small. Therefore, the estimated map length does not necessarily reflect the actual chromosome length.

Comparing mapping procedures; JOINMAP with separate phase analysis vs. MAPMAKER with missing data censoring:

The mapping order in MAPMAKER was similar to the JOINMAP output for most chromosomes. However, in some LG's with low proportions of anchoring markers vs. BI markers, or unevenly distributed anchoring markers, large rearrangements sometimes occurred. This is caused by the fact that MAPMAKER compares small subsets of markers rather than all representatives of an LG at the same time. MAPMAKER initially uses a maximum of eight markers, and subsequently positions additional markers within the initial (eight marker) map. Finally, the mapping order is fine-tuned by using a sliding window of five markers (ripple command). The use of a subset of markers (i.e. eight initial markers or five ripple markers) that is made up of BI markers of both maternal linkage phases and less than two anchoring markers, results in an unreliable suggested marker order. The reliability of the initial (eight marker) map can be improved by including all available MI and codominant markers, but with the ripple command the representative markers cannot be hand-picked because their grouping depends on the provisional marker order suggested by MAPMAKER. Similar to the ripple command that is used to determine marker order, a sliding window analysis also reveals the reliability of the marker order, by comparing the likelihood of the most likely marker order with alternative orders (flips test). This test is confronted with the same bi-phasic incompatibility problems and cannot be used on a censored data set with missing data. The consequences of comparing only subsets of markers within a linkage group (i.e. sliding window) are illustrated with an example based on LG21, which is characterized by codominant anchoring markers close to both ends and ten dominant markers of both phases in between them (Appendix 5.5). JOINMAP also performs a ripple test, which is based on a sliding window of only three markers. With "missing data" input, this results in even more problems than in MAPMAKER because the chance that two anchoring markers are included in a subset of just three markers is far smaller than in a subset of five. This is presumably the reason why for the final mapping step in some butterfly linkage maps JOINMAP has been replaced by MAPMAKER.

The marker order suggested by JOINMAP (following the procedures used for the present linkage map) is far more reliable than the MAPMAKER approach because it does not attempt to map incompatible BI markers relative to each other directly. The ripple test, which can cause serious problems with missing data analysis, strongly increases the reliability of the marker order when analyzing markers of each maternal linkage phase separately in JOINMAP. Instead of reporting a flips test value, JOINMAP simply excludes markers that do not meet the criteria for reliable neighboring markers (recombination frequency smaller than 0.4 and LOD larger than 1.0). MAPMAKER on the other hand always suggests a mapping order and will always produce a linkage map that includes all presented markers.

The mapping distances given by MAPMAKER were larger than those produced by JOINMAP under all circumstances. The mapping distances decreased substantially with error detection activated in MAPMAKER, but were on average still 38% larger than in JOINMAP, ranging from 1.02 to 2.14 times in size for the different LGs (Appendix 5.6). The total mapping distances are 1873 vs. 1354 cM for MAPMAKER and JOINMAP respectively. The data are presented in different ways to each program, with the censored BI-markers as missing data in MAPMAKER and excluded in JOINMAP. Since JOINMAP has difficulties with high proportions of non-overlapping missing data, a comparison with identical data input was not possible for the MI-markers combined with censored BI-markers. Therefore, the software was

also compared based only on MI-markers, thus avoiding censoring of markers. Fourteen LGs had sufficient MI-markers to construct linkage maps with MAPMAKER again giving higher values than JOINMAP, but now with only 17% difference. The genome size of *B. anynana* is 0.49 pg (GREGORY and HEBERT 2003), which corresponds with approximately 480 Mb (DOLEŽEL *et al.* 2003). This implies that the JOINMAP based linkage map is 355 Kb/cM and the MAPMAKER based map 256 Mb/cM.

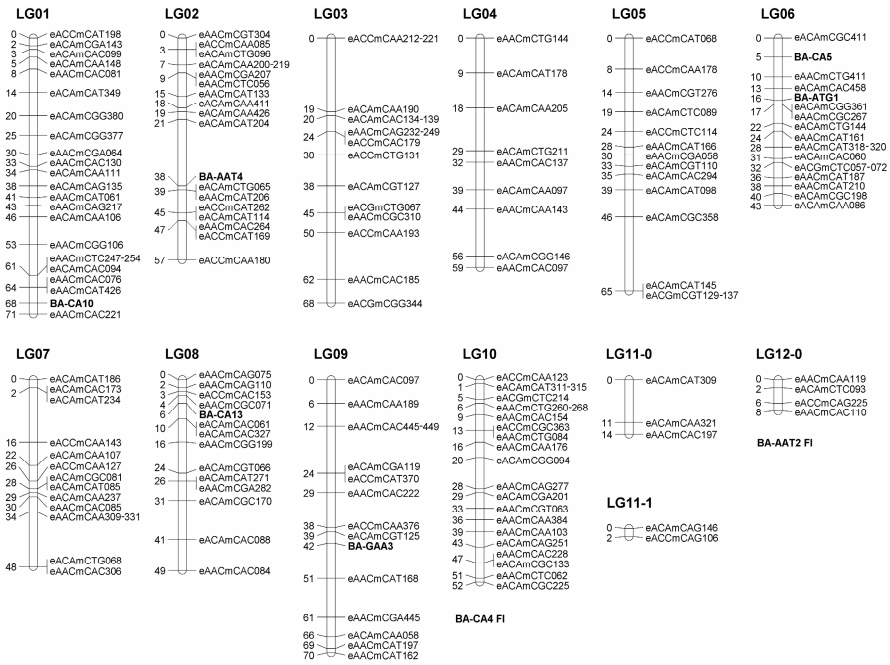


Figure 5.1 Linkage map of LG1-12. Vertical bars represent chromosomes and show the mapping distance in centimorgan (cM) on the left and the corresponding markers on the right. Microsatellites are displayed in bold and start with "BA", the two microsatellites with only FI polymorphism are placed underneath the LG's they belong to. AFLPs are named according to their selective primer extension and amplicon size. The "e" stands for the fluorescent *Eco*RI-based primer and the "m" stands for the non-fluorescent *Mse*I-based primer. AFLPs with two amplicon sizes per primer combination (e.g. eACCmCAA212-221 in LG03) are codominant. A vertical line indicates that markers are less than 1 cM apart (e.g. eACAmCGA119 and eAACmCAT370 in LG09)

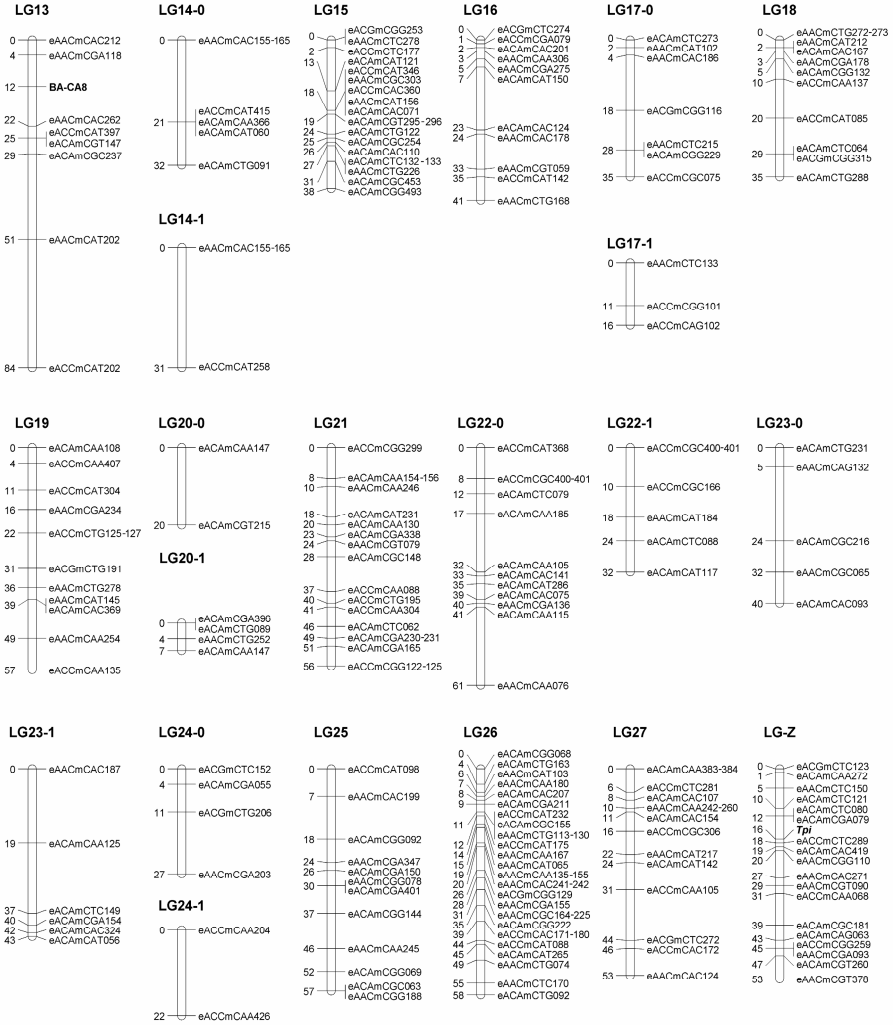


Figure 5.2 Linkage map of LG13-27 and Z.

DISCUSSION

How to get the most out of an F_2 design: The full-sib F_2 cross design was chosen for the purpose of mapping QTL for ventral eyespot size. It generates a maximum phenotypic range in the offspring while keeping random genetic variation to a minimum. As a downside, this design is not ideally suited for linkage mapping with dominant markers.

One effect of having just one set of grandparents is that BI markers carry information in only one of both paternal linkage phases for most LGs (Appendix 5.7). Another effect is that it creates a strong bottleneck, that results in a lower proportion of FI and MI markers relative to BI markers than in an outbred cross (Appendix 5.7). This is most striking when the F_1 male and female inherit the same set of P chromosomes, where 1:1 segregating markers can only arise as a result of recombination in the P-male. This unfavorable F_1 gamete combination occurs in 25% of the chromosomes, and is reflected by the complete absence of FI-markers in six LGs. Without recombination in the P-male for such LGs, generating more AFLP markers will not produce FI-markers because they do not exist for these linkage groups. Therefore, the chromosome prints for these six FI-devoid autosomes had to be obtained from BI and MI-markers instead. This reconstruction is based on the forbidden genotype restrictions, and the assumptions that either the unassigned individuals received a MI-marker that was fully associated with a non-recombinant BI-marker region (BI + MI reconstruction), or that the female BI component segregation is 1:1 (BI only reconstruction). Empirical tests based on LGs with available chromosome prints showed that this approach creates chromosome prints that are identical or nearly identical to the available ones, and linkage maps that are very similar to those based on conventionally censored datasets. The stochastic deviations from the 1:1 segregation have a negligible effect on the mapping distance and no effect on the mapping order. This validates the BI censoring approach for LGs without FI-markers.

The selective genotyping approach was chosen to avoid genotyping intermediate eyespot phenotypes in the offspring, since they provide hardly any additional information in QTL mapping compared to that of the extreme phenotypes (MURANTY *et al.* 1997). As a result of this, the linkage map itself is based on a non-random set of offspring. The effect of this on the reliability of the linkage map is negligible because it does not affect the three main characteristics in linkage mapping: namely, marker grouping, marker order and marker distance. There could, however be an effect of selection on the ratios of segregating markers, since dominance promotes extreme phenotypes in recessive homozygotes and additive alleles produce extreme phenotypes in both types of homozygotes. Markers that are linked to genes which are involved with eyespot formation may therefore deviate from 3:1 or 1:1 ratios due to hitch-hiking.

Effects of data censoring: Using MAPMAKER with censored BI-markers as missing data resulted in a map that was 38% larger in size than the one produced from two subsets per chromosome with JOINMAP. This size difference is caused by two factors. Firstly, there is a software effect (i.e. algorithms used) that is revealed by analyzing only the (uncensored) MI-markers, that accounts for 17% of the difference in this study. The rest of the difference is caused by the treatment of the incompatible bi-phasic censored BI-markers. The main purpose of the MAPMAKER analysis was to allow comparison of mapping distance in *B. anynana* with other Lepidoptera linkage maps, since this is the first species in this taxon for which the final mapping

step was performed in JOINMAP. This software has not been used before for Lepidoptera linkage maps, presumably because it is less able to deal with a substantial portion of non-overlapping genotypes than MAPMAKER. Our approach avoided this problem by adapting that of (YASUKOCHI *et al.* 2006) which involves splitting up the dataset based on chromosome print value and omitting female-derived information rather than treating it as missing data. This results in two linkage maps per chromosome that are then juxtaposed and integrated based on common MI and codominant markers and their average distances. Rather than just using the average distance between the anchoring markers to combine the two phases, JOINMAP also takes the number of individuals representing both subsets into account (STAM 1993).

Linkage groups and chromosome number: The number of LGs matches the karyotype, thus markers are available for all 27 autosomes and the Z chromosome. There are no markers available for the W chromosome, probably due to its small size in *B. anynana* (VAN'T HOF *et al.* 2008). The marker densities and distances vary substantially between the different chromosomes, but given the uniform lengths of the pachytene bivalents, we interpret this as incomplete marker coverage rather than a difference in chromosome size. We aimed to present an integrated linkage map, with relative marker positions and distances based on both sets of incompatible BI-markers linked together with MI, codominant AFLP and microsatellite markers. We succeeded for 20 LGs, and mapped the remaining eight separately because they lacked sufficient anchoring markers. The presence of the *Tpi* gene of *B. anynana* is consistent with all (distantly related) Lepidoptera species for which this gene has been mapped to date (summarized in (TRAUT *et al.* 2007)). This strengthens the hypothesis of taxon-wide conserved synteny for at least part of the Lepidoptera Z chromosome.

Linkage and physical maps in Lepidoptera: The present linkage map provides the basis for the assignment of the number, position, effect and interactions of QTLs involved with the development of wingspot size. We will further anchor the map using SNP markers (BELDADE *et al.* 2006), with a main focus on genes that are involved in eyespot formation in *B. anynana* and eyespot and wing pattern formation in Lepidoptera in general. Additionally, physical anchoring of linkage groups to specific chromosomes by means of BAC-FISH, as has been performed in *B. mori* (YOSHIDO *et al.* 2005), will provide a solid framework for future mapping studies.

The MAPMAKER mapping distance of 1873 cM in *B. anynana* is within the 1430-2542 cM range reported for other butterfly species (JIGGINS *et al.* 2005; KAPAN *et al.* 2006; WANG and PORTER 2004). The accuracy of these mapping distances may however be limited, since mapping distances of both 1430 cM and 2400 cM were reported in *Heliconius erato* (KAPAN *et al.* 2006; TOBLER *et al.* 2005) and distances ranging from 1305 cM to 6512 cM in *Bombyx mori* (TAN *et al.* 2001; YAMAMOTO *et al.* 2006) when using MAPMAKER software. One mapping software package that does support sex-specific map construction is CRI-MAP (LANDER and GREEN 1987), which has been used to build many mammalian genetic maps. To our knowledge, CRI-MAP has never been used to compute a Lepidoptera map based on dominant markers. CRI-MAP shares some of its origins with MAPMAKER and suffers from the same deficiencies of MAPMAKER we have explained above. Notably, CRI-MAP includes (1) no robust method to choose an initial order of markers and (2) no systematic method to decide whether a marker should be excluded from the map because it cannot be reliably ordered. Our proposed mapping strategy avoids Lepidoptera specific issues that have an effect on mapping distance and order, but it still requires a large number of analysis steps. Therefore, we would welcome the implementation of sex-specific recombination in the analysis parameters of

JOINMAP. This would not just be an asset to linkage mapping in Lepidoptera, but for all organisms in which sex-specific recombination rates have been reported.

ACKNOWLEDGEMENTS

We thank Dr Durrell Kapan and Dr. Ir. Johan W. van Ooijen for valuable comments on linkage mapping in Lepidoptera and its implications for JOINMAP software.

REFERENCES

BELDADE, P., S. RUDD, J. D. GRUBER and A. D. LONG, 2006 A wing expressed sequence tag resource for *Bicyclus anynana* butterflies, an evo-devo model. BMC Genomics **7:** 130.

BENHAM, J., J.-U. JEUNG, M. JASIENIUK, V. KANAZIN and T. BLAKE., 1999 Genographer: A graphical tool for automated AFLP and microsatellite analysis. J. Agric. Genomics **4**.

BRAKEFIELD, P. M., J. GATES, D. KEYS, F. KESBEKE, P. J. WIJNGAARDEN *et al.*, 1996 Development, plasticity and evolution of butterfly eyespot patterns. Nature **384:** 236-242.

DOLEŽEL, J., J. BARTOŠ, H. VOGLMAYR and J. GREILHUBER, 2003 Nuclear DNA content and genome size of trout and human. Cytometry **51A:** 127-128.

DOPMAN, E. B., S. M. BOGDANOWICZ and R. G. HARRISON, 2004 Genetic mapping of sexual isolation between E and Z pheromone strains of the European corn borer (*Ostrinia nubilalis*). Genetics **167:** 301-309.

GREGORY, T. R., and P. D. N. HEBERT, 2003 Genome size variation in lepidopteran insects. Canadian Journal of Zoology **81:** 1399-1405.

HECKEL, D. G., L. J. GAHAN, Y.-B. LIU and B. E. TABASHNIK, 1999 Genetic mapping of resistance to *Bacillus thuringiensis* toxins in diamondback moth using biphasic linkage analysis. PNAS **96:** 8373-8377.

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

KAPAN, D. D., N. S. FLANAGAN, A. TOBLER, R. PAPA, R. D. REED *et al.*, 2006 Localization of Müllerian mimicry genes on a dense linkage map of *Heliconius erato*. Genetics **173:** 735-757.

LANDER, E. S., and P. GREEN, 1987 Construction of multilocus genetic linkage maps in humans. Proc. Natl. Acad. Sci. USA **84:** 2363-2367.

LANDER, E. S., P. GREEN, J. ABRAHAMSON, A. BARLOW, M. J. DALY *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181.

MIAO, X.-X., S.-J. XUB, M.-H. LI, M.-W. LI, J.-H. HUANG *et al.*, 2005 Simple sequence repeat-based consensus linkage map of *Bombyx mori*. PNAS **102:** 16303-16308.

MURANTY, H., B. GOFFINET and F. SANTI, 1997 Multitrait and multipopulation QTL search using selective genotyping. Genetical Research **70:** 259-265.

PROMBOON, A., T. SHIMADA, H. FUJIWARA and M. KOBAYASHI, 1995 Linkage map of Random Amplified Polymorphic DNAs (RAPDs) in the silkworm, *Bombyx mori*. Genetical Research **66:** 1-7.

REGIER, J. C., 2005 Protocols, Concepts, and Reagents for preparing DNA sequencing templates. Version 3/14/06. www.umbi.umd.edu/users/jcrlab/PCR_primers.pdf pp.

SHI, J., D. G. HECKEL and M. R. GOLDSMITH, 1995 A genetic linkage map for the domesticated silkworm, *Bombyx mori*, based on restriction fragment length polymorphisms. Genetical Research **66:** 109-126.

STAM, P., 1993 Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J **3:** 739-744.

TAN, Y. D., C. WAN, Y. ZHU, C. LU, Z. XIANG *et al.*, 2001 An amplified fragment length polymorphism map of the silkworm. Genetics **157:** 1277-1284.

TOBLER, A., D. KAPAN, N. FLANAGAN, C. GONZALEZ, E. PETERSON *et al.*, 2005 First-generation linkage map of the warningly colored butterfly *Heliconius erato*. Heredity **94:** 408–417.

TRAUT, W., K. SAHARA and F. MAREC, 2007 Sex Chromosomes and Sex Determination in Lepidoptera. Sexual Development **1:** 332–346.

VAN'T HOF, A. E., F. MAREC, I. J. SACCHERI, P. M. BRAKEFIELD and B. J. ZWAAN, 2008 Cytogenetic Characterization and AFLP-Based Genetic Linkage Mapping for the Butterfly *Bicyclus anynana*, Covering All 28 Karyotyped Chromosomes. PLoS ONE **3:** e3882.

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*. Molecular Ecology Notes **5:** 169-172.

VOS, P., R. HOGERS, M. BLEEKER, M. REIJANS, T. V. D. LEE *et al.*, 1995 AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research **23:** 4407-4414.

WANG, B., and A. H. PORTER, 2004 An AFLP-based interspecific linkage map of sympatric, hybridizing *Colias* butterflies. Genetics **168:** 215-225.

WIJNGAARDEN, P. J., and P. M. BRAKEFIELD, 2000 The genetic basis of eyespot size in the butterfly *Bicyclus anynana*: an analysis of line crosses. Heredity **85:** 471-479.

YAMAMOTO, K., J. NARUKAWA, K. KADONO-OKUDA, J. NOHATA, M. SASANUMA *et al.*, 2006 Construction of a single nucleotide polymorphism linkage map for the silkworm, *Bombyx mori*, based on bacterial artificial chromosome end sequences. Genetics **173:** 151-161.

YAMAMOTO, K., J. NOHATA, K. KADONO-OKUDA, J. NARUKAWA, M. SASANUMA *et al.*, 2008 A BAC-based integrated linkage map of the silkworm, *Bombyx mori* Genome Biology **9:** R21.

YASUKOCHI, Y., 1998 A dense genetic map of the silkworm, *Bombyx mori*, covering all chromosomes based on 1018 molecular markers. Genetics **150:** 1513-1525.

YASUKOCHI, Y., L. A. ASHAKUMARY, K. BABA, A. YOSHIDO and K. SAHARA, 2006 A second-generation integrated map of the silkworm reveals synteny and conserved gene order between lepidopteran insects. Genetics **173:** 1319-1328.

YOSHIDO, A., H. BANDO, Y. YASUKOCHI and K. SAHARA, 2005 The *Bombyx mori* karyotype and the assignment of linkage groups. Genetics **170:** 675-685.

APPENDIX 5.1A

Reconstruction of the chromosome print based on forbidden genotype restrictions

As a result of the full-sib cross design, one out of four linkage groups has a very small proportion of 1:1 informative markers, which makes it difficult to generate chromosome prints for these linkage groups. Without a chromosome print available, BI markers can still be mapped, but without censoring, the software will underestimate the mapping distance about twofold because absence of recombination from the female-derived genotypes is interpreted as close linkage. Therefore, the markers on these chromosomes need to be censored, but in an alternative way from the FI marker based censoring.

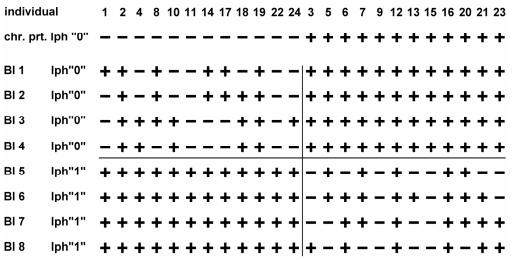


Figure 5.3 Example of the relation between chromosome print and dominant marker phenotype of BI markers in the two different maternal linkage phases. Chromosome print values and the peakscores of eight BI markers of 24 F_2 individuals (+ = peakpresent, – = peakabsent) are shown. All markers (and the chromosome print) belong to the same linkage group and both maternal linkage phases (lph"0" and lph"1") are represented by the BI markers (four of each). All markers have the same paternal linkage phase. The individuals have been grouped based on their chromosome print value and maternal linkage phase. The grouping reveals two clusters with exclusively peakpresents and two clusters with a mixture of peakpresents and peakabsents. This figure shows the pattern on which the reconstruction is based and not the actual reconstruction itself. Thus individuals 2, 12, 16 and 19 stay unassigned when using this pattern for reconstruction without chromosome prints and without additional (e.g. MI) information.

Given the forbidden genotype restrictions, an individual that has one or more peakabsents in one maternal linkage phase of BI markers must have exclusively peakpresents in the alternative maternal linkage phase. This peakabsent-linkage phase combination gives a pattern from which a substantial part of the chromosome print can be reconstructed. This pattern is shown in Fig. 5.3 where markers with the same linkage phases are grouped vertically and the individuals with the same chromosome print values horizontally. The pattern is a direct consequence of the absence of recombination in the females. A "–" chromosome print value combined with BI marker in linkage phase "1" produces exclusively peakpresents (bottom left). A chromosome print "+" together with BI linkage phase "0" also gives exclusively peakpresents (top right). These two "exclusively peakpresent" clusters consist of a mixture of heterozygotes and homozygote peakpresents that cannot be told apart. The remaining genotypes (top left and bottom right) are either homozygous peakabsent or heterozygotes that inherited a paternal peakpresent (and a maternal peakabsent). The

chromosome print values can be reconstructed directly from this pattern for individuals that have at least one peakabsent and known linkage phases. The remaining individuals have peakpresent values for all markers in both maternal linkage phases (e.g. individuals 2, 12, 16 and 19 in Fig. 5.3), so that distinction between F_2 male- and female component is not possible and the chromosome print values for these individuals stay "unassigned". The proportion of individuals with "all-present" in both linkage phases is substantial in a full-sib cross as explained in Appendix 5.7. They reflect non-recombinant regions on the F_1 male chromosome. In contrast, all-presents are far less abundant in outbred crosses, which makes the reconstruction of a chromosome print easier, but it will generally not be necessary to reconstruct them in the first place because FI markers are more common in outbred crosses.

APPENDIX 5.1B

Reconstruction of the chromosome print for unassigned individuals with only BI markers available

The unasigned individuals are all-peakpresent in both linkage phases of BI markers. Therefore, it remains unclear what linkage phase should be censored. However, it does not have any consequences for the linkage map whether the actual female component is censored, or whether an equal number of random "all-peakpresents" is excluded. This means that for half these unassigned individuals, the chromosome print can be set to peakabsent and for the other half to peakpresent, based on the assumption that the female marker inheritance is 1:1. In case of the example given in Fig. 5.3, individuals 2 and 12 would be assigned chromosome print value "–" and 16 and 19 assigned value "+". This means that 12 and 19 are grouped in the wrong cluster, but this has no consequences for marker distribution, and therefore no effect on the linkage map. In addition, because stochastic departures from the 1:1 ratio have an effect on the mapping distance, chromosome prints were reconstructed based on the extremes of the binomial 95% confidence interval for 1:1 segregation in order to determine the error margins of this approach. Shifting the censoring to the boundaries of the 95% binomial confidence level changed the total mapping distance 2 cM or less, and kept the marker order intact. We applied the same reconstruction method to the first ten linkage groups (with chromosome prints available) to determine the consistency of mapping order and distance for the BI-only linkage groups when using this technique.

APPENDIX 5.1C

Reconstruction of the chromosome print for unassigned individuals with BI and MI markers available

The random 1:1 reconstruction described in Appendix 5.1b is only suitable for linkage groups with exclusively BI markers. It is not appropriate for a combination of BI and MI markers, because there is an effect on the linkage map when MI markers end up in the wrong censoring group due to random chromosome print value assignment. However, with MI markers available, the chromosome print values of the unassigned individuals can also be determined. For the following reconstruction, we take only BI markers with one of the two possible paternal linkage phases into

account. We choose the paternal phase with the most representatives available, but with a full-sib design, this usually means that all markers are included because the full-sib cross causes most BI markers to be in the same paternal phase (Appendix 5.7). Specifically, this means that we only look at markers that are peakpresent on the same F_1 male chromosome and leave out the markers that are peakpresent on the other. An individual that inherited BI markers that are all in a non-recombinant region of the F_1 male chromosome will be either exclusively peakpresent or exclusively peakabsent within a maternal linkage phase. In Fig. 5.3, the non-recombinant "all peakpresents" are represented by individuals 2, 12, 16 and 19 and the non-recombinants with exclusively peakabsents in one maternal phase are represented by individuals 9, 11, 15 and 22. A MI marker that lies within such a homogenous (non-recombinant) region will share the same pattern, with exclusively MI peakpresents associated with BI all-peakabsents in one phase, and MI peakabsents with BI all-peakabsents in the opposite phase. Thus by using the pattern of MI loci that meet these criteria, the male and female component in the unassigned (BI all-present) individuals can be determined, which allows reconstruction of the chromosome print for all unassigned individuals. This approach to identify the male component of F_2 markers is similar to the identification of the female component based on FI markers (and chromosome prints). The main difference is that FI markers are always fully linked to BI markers because all female inherited chromosomes are completely non-recombinant while identification of the male component is based on non-recombinant chromosome regions that need to be identified first.

In order to test and validate the BI+MI-based chromosome print reconstruction empirically, we used the first ten linkage groups (with chromosome prints available) to determine the difference between the actual chromosome print and reconstructed chromosome print for BI+MI linkage groups.

APPENDIX 5.2

Forbidden genotype screening

The manifestation of a forbidden genotype combination of a BI marker and the chromosome print depends on their relative linkage phases in the F_1 female. When the chromosome print positive signal and the BI marker positive signal are on the alternate chromosomes in the F_1 female, the offspring must always be positive peakpresent either in the chromosome print, screened marker, or in both. Both absent would mean that the marker negative signals formed a novel combination (i.e. both on the same chromosome) in the female, and thus resulted from forbidden recombination (Fig. 5.4a). Similarly, both positive signals on the same chromosome would also indicate forbidden recombination. However, such a haplotype does not give a unique detectable combination with dominant markers. In fact, only one out of four forbidden genotype combinations can be detected when using dominant markers (Fig. 5.4a). The alternative F_1 female marker combination has the positive signals on the same chromosome, and both negatives on the other (Fig. 5.4b). In this case, absence of the BI marker in combination with presence of the chromosome print marker is not allowed. Again, only one out of four forbidden genotypes is detected. Therefore we excluded loci with more than two inconsistencies from further analysis. Forbidden genotype screening in microsatellites is similar, but the proportion of detectable forbidden genotypes is higher (50% or 100% depending on the number of alleles involved).

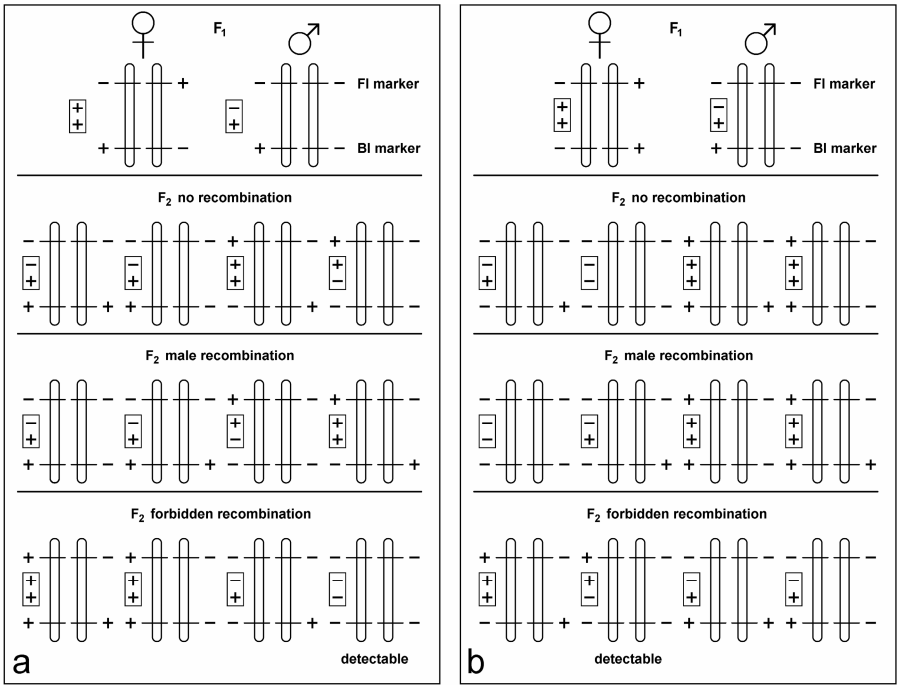


Figure 5.4 Detection of forbidden genotypes. (a) Cross between an F_1 female with positive signals in repulsion and an F_1 male with one positive BI signal. Vertical bars represent chromosomes with marker loci on both ends. The chromosome print is here represented by an FI marker. The characters in the box represent the dominant marker output (i.e. + = peakpresent, – = peakabsent) for both loci. Male recombination does not result in new marker combinations and is only included for completeness. Of the four possibilities, only the double negative marker combination is distinguishable as forbidden genotype; (b) Cross with the positive signals linked in the F_1 female. In this case, an FI peakpresent combined with a BI peakabsent is detectable as a forbidden recombination.

APPENDIX 5.3

Censoring of BI markers and map integration with anchoring markers

BI markers that have F_1 female inherited peakpresents must be excluded from analysis because the absence of recombination in females makes them uninformative. The chromosome print defines which individual-linkage phase combinations have such an allele. The pattern shown in Fig. 5.3 that was used for chromosome print reconstruction, also groups the data that is to be excluded from analysis. The "exclusively peakpresent" clusters with a female inherited component that need to be censored are shown in Fig. 5.5. The issue of two incompatible linkage groups per

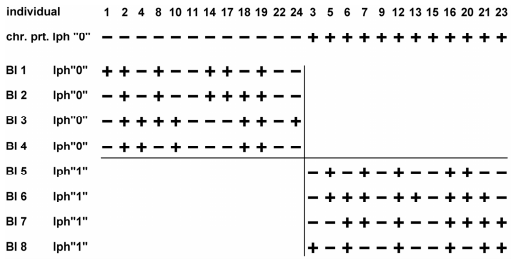


Figure 5.5 Censored BI markers. The same diagram as in Fig. 5.3, but now with the uninformative clusters removed.

chromosome is also clearly illustrated in Fig. 5.5. Excluding the two peakpresent clusters from analysis leaves two subsets of data (top left and bottom right) that cannot be linked to each other without anchoring markers because they do not hold information within the same individuals. Fig. 5.6 shows two MI markers in addition to the eight BI markers in Fig. 5.5. They provide the information that allows integration of the two clusters because their genotypes are informative in all F_2 individuals. In our approach, we first construct two separate linkage maps. One for individuals with chromosome print value "–" and the other for individuals with chromosome print value "+".With at least two anchoring markers that are not closely linked available, they can be integrated by JOINMAP as demonstrated in Fig. 5.7.

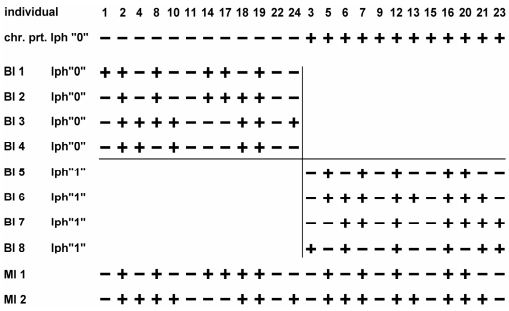


Figure 5.6 Censored BI markers with MI anchoring markers. With anchoring markers (here represented by MI markers), both BI clusters can be linked together and integrated into a single linkage group because the MI markers are informative in all individuals. The linkage phase of the MI markers is not shown here because MI markers do not have a maternal linkage phase and the paternal linkage phase is not relevant in this example.

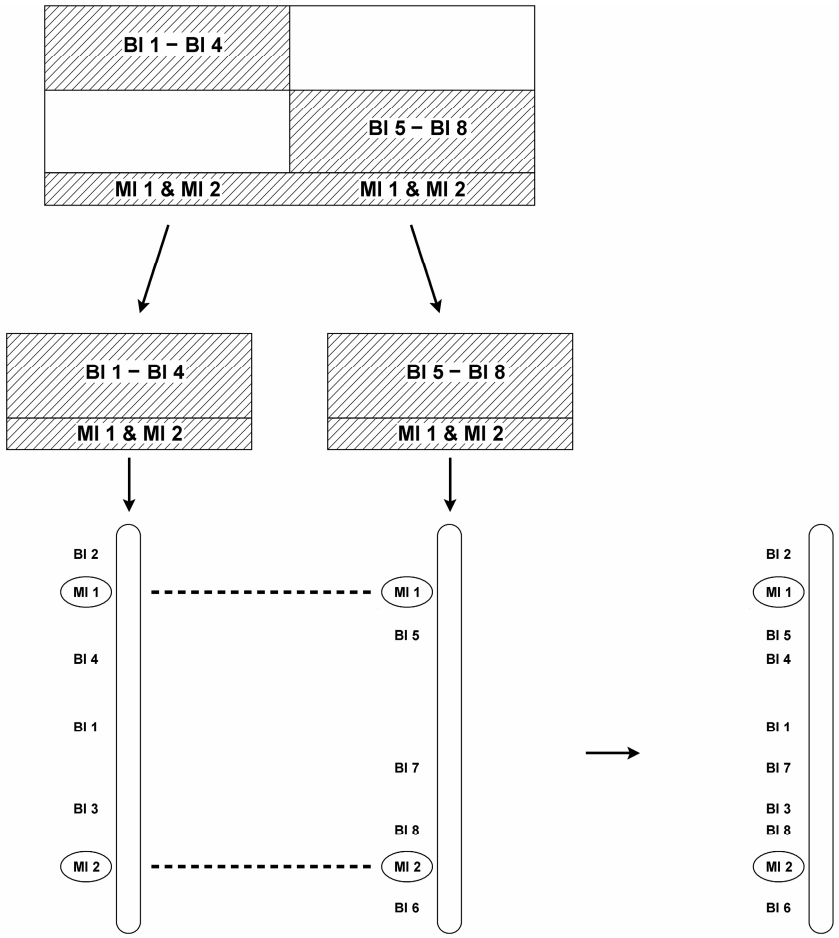


Figure 5.7 Mapping and Integration of two incompatible BI linkage groups. This is a schematic representation of the censored data as shown in Fig. 5.6, followed by the final mapping steps. Individuals with different chromosome print values are mapped separately. The two anchoring (MI) markers facilitate integration into a single linkage map.

APPENDIX 5.4

Microsatellite censoring

Microsatellite genotypes need to be converted so that only the paternal component in the F_2 is used for mapping. F_1-male-specific alleles can be translated directly into a present-absent pattern for that allele, and then used as a MI marker.

When the F_1 male and female are both heterozygous with the same sets of alleles, it is not immediately clear which allele came from which parent in the F_2 heterozygotes. However, the chromosome print can reveal the origin of these heterozygous alleles. The two alternative F_2 heterozygotes are fully associated with the opposite chromosome print values. Distinction between the male and female component is illustrated by an example below.

We consider a microsatellite locus with a 10 repeat allele and a 12 repeat allele in both F_1 male and F_1 female.

F_1 female $10^a/12^a$
F_1 male $10^b/12^b$

F_2 (offspring)	$10^a/10^b$	$10^a/12^b$	$12^a/10^b$	$12^a/12^b$
Chromosome print	–	–	+	+
Translated to MI:	10	12	10	12

The 10/10 F_2 individuals should all have the same chromosome print values (e.g. "–") and the 12/12 F_2 individuals should all have the opposite chromosome print values (i.e. "+"). This homozygote-chromosome print combination reveals which chromosome print value is associated with the maternal "10^a" allele and which with the maternal "12^a" allele. The chromosome print-allele association is the same in the F_2 heterozygotes and thus reveals the female component that is to be excluded from analysis. This codominant censoring does not result in the exclusion of half the individuals for such a locus, as is the case in BI marker censoring. Instead, these microsatellites can be treated as MI markers representing all F_2 individuals, and serve as anchoring markers.

APPENDIX 5.5

Implications of sliding window analysis of a "missing data"-censored dataset based on an example.

Linkage group 21 is used to illustrate the consequences of sliding window analysis on a linkage group that has no evenly spaced anchoring markers. Fig. 5.8 shows the linkage map of LG21 with anchoring markers and BI markers of both phases in different colors. A sliding window of five neighboring markers usually results in incompatible combinations. Table 5.2 shows the censored genotype score of all LG21 markers in 14 F_2 individuals. Below that are three examples of marker subsets that are compared with different sliding window positions (Table 5.3a-5.3c). The first two examples allow reliable marker ordering, but the third example shows why a missing data approach can have severe consequences for the reliability of the marker order. The mapping approach described in Supplement 5.3 avoids these compatibility issues because it only compares markers that have genotype scores available within the same individuals.

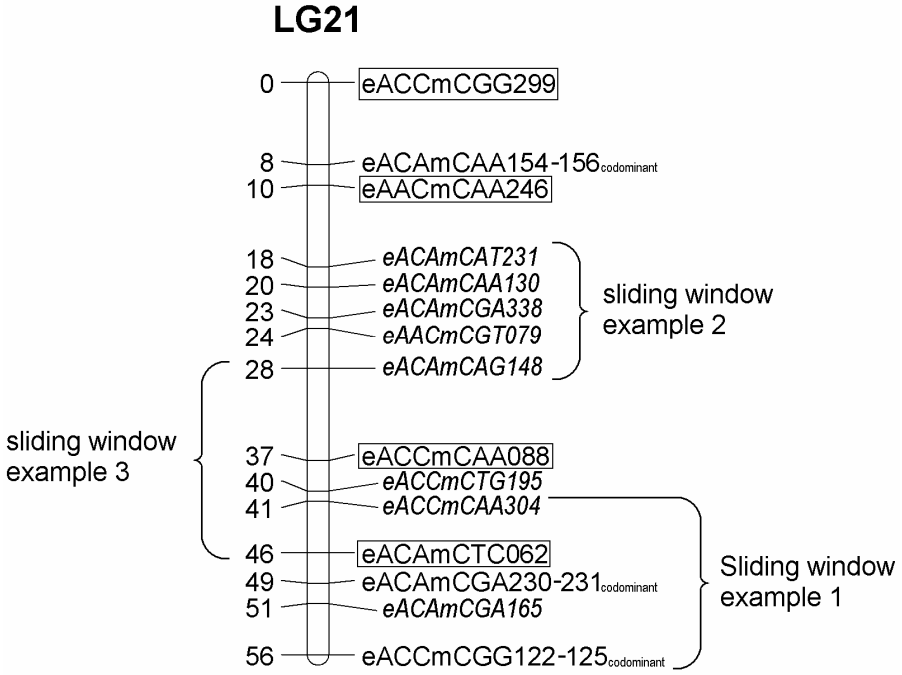


Figure 5.8 Mapping order, mapping distance and maternal linkage phase of BI markers in linkage group 21. Mapping order and distance are based on separate phase mapping followed by integration in JOINMAP. Markers in rectangles have maternal linkage phase "0", markers positioned slightly to the right and italicized have maternal linkage phase "1". The remaining three markers are codominant. The linkage phase of the codominant markers is not relevant in this context. The three subsets of five markers that are dealt with in the sliding window examples below are indicated with braces.

Table 5.2 Censored peakscore of LG21.
Markers in grey cells have maternal linkage phase "0", markers aligned right and italicized have maternal linkage phase "1". The remaining three markers are codominant. The censored values are black cells. The codominant markers have been merged to show only their male informative component. They have a genotype score in all individuals. Only 14 of the 92 individuals are shown. Missing data (md) represents the censored genotypes. The censored genotypes correspond with the chromosome print values (bottom row) with combinations [phase 0/print 0] and [phase 1/print 1] defined as md.

F_2 Individual	1	2	3	4	5	6	7	8	9	10	11	12	13	14
eACCmCGG299	md	md	md	0	0	md	0	md	md	md	1	0	md	0
eACAmCAA154-156	0	1	0	0	0	1	1	0	1	0	1	0	1	1
eAACmCAA246	md	md	md	0	0	md	0	md	md	md	1	0	md	1
eACAmCAT231	1	0	1	md	md	0	md	1	0	0	md	md	1	md
eACAmCAA130	1	0	1	md	md	0	md	0	0	0	md	md	1	md
eACAmCGA338	1	0	1	md	md	0	md	1	0	0	md	md	1	md
eAACmCGT079	1	1	1	md	md	0	md	1	0	0	md	md	0	md
eACAmCGC148	1	0	1	md	md	0	md	1	1	0	md	md	1	md
eACCmCAA088	md	md	md	0	0	md	1	md	md	md	1	0	md	1
eACCmCTG195	0	0	1	md	md	1	md	1	0	0	md	md	1	md
eACCmCAA304	1	1	1	md	md	0	md	1	0	1	md	md	0	md
eACAmCTC062	md	md	md	0	0	md	1	md	md	md	1	0	md	1
eACAmCGA230-231	0	0	1	0	0	1	1	0	1	0	1	0	1	1
eACAmCGA165	1	1	0	md	md	0	md	1	0	1	md	md	0	md
eACCmCGG122-125	0	0	1	0	0	1	1	0	1	0	1	0	1	0
Chromosome print 21	0	0	0	1	1	0	1	0	0	0	1	1	0	1

Table 5.3a Sliding window example 1
These are the five markers at the bottom of the linkage group. This includes two anchoring markers, two BI markers with maternal linkage phase "1" and one with maternal linkage phase "0". The mapping order can be determined for all markers due to the presence of two anchoring markers.

F_2 Individual	1	2	3	4	5	6	7	8	9	10	11	12	13	14
eACCmCAA304	1	1	1	md	md	0	md	1	0	1	md	md	0	md
eACAmCTC062	md	md	md	0	0	md	1	md	md	md	1	0	md	1
eACAmCGA230-231	0	0	1	0	0	1	1	0	1	0	1	0	1	1
eACAmCGA165	1	1	0	md	md	0	md	1	0	1	md	md	0	md
eACCmCGG122-125	0	0	1	0	0	1	1	0	1	0	1	0	1	0

Table 5.3b Sliding window example 2
Five neighboring BI markers with the same maternal linkage phase. Marker order and integrity can be determined for all five markers in this subset because the informative genotype scores are available for all individuals.

F_2 Individual	1	2	3	4	5	6	7	8	9	10	11	12	13	14
eACAmCAT231	1	0	1	md	md	0	md	1	0	0	md	md	1	md
eACAmCAA130	1	0	1	md	md	0	md	0	0	0	md	md	1	md
eACAmCGA338	1	0	1	md	md	0	md	1	0	0	md	md	1	md
eAACmCGT079	1	1	1	md	md	0	md	1	0	0	md	md	0	md
eACAmCGC148	1	0	1	md	md	0	md	1	1	0	md	md	1	md

Table 5.3c Sliding window example 3
This subset contains markers of both maternal linkage phases but no anchoring markers. It is not possible to establish marker order or marker order integrity for all five markers because genotypes in markers of maternal phase "1" are always compared with missing data in markers with linkage phase "0" and vice versa. In such a case, JOINMAP refuses to suggest a marker order, but MAPMAKER will give a marker order even though it may not correspond with the actual order of the markers.

F_2 Individual	1	2	3	4	5	6	7	8	9	10	11	12	13	14
eACAmCGC148	1	0	1	md	md	0	md	1	1	0	md	md	1	md
eACCmCAA088	md	md	md	0	0	md	1	md	md	md	1	0	md	1
eACCmCTG195	0	0	1	md	md	1	md	1	0	0	md	md	1	md
eACCmCAA304	1	1	1	md	md	0	md	1	0	1	md	md	0	md
eACAmCTC062	md	md	md	0	0	md	1	md	md	md	1	0	md	1

APPENDIX 5.6

Linkage group sizes of *Bicyclus anynana* produced with MAPMAKER and with JOINMAP

Mapping with MAPMAKER instead of JOINMAP resulted in larger mapping distances. With two incompatible linkage groups per chromosome, only the larger one was mapped. The "MM" "MM err. det." and "JM" columns give the linkage group sizes based on all informative markers. The three columns on the right report the male-informative mapping distances. The "err. det." columns specify the MAPMAKER output with error detection activated. The MI-based linkage groups are generally smaller size, partly because the MI data is more reliable, but mainly because they are physically closer to each other than the BI markers at the extremes of the linkage groups.

	MM	MM err. det.	JM	MM MI	MM MI err. det.	JM MI
LG01	138.8 cM	97.1	71 cM	88.5 cM	66.1 cM	54 cM
LG02	73.1 cM	66.6	57 cM	46.2 cM	44.1 cM	41 cM
LG03	95.2 cM	77.6	68 cM			
LG04	74.5 cM	65.9	59 cM	51.9 cM	50.4 cM	42 cM
LG05	78.7 cM	74.5	65 cM			
LG06	75.6 cM	51.0	43 cM	52.6 cM	50.8 cM	47 cM
LG07	74.8 cM	60.9	48 cM	63.9 cM	60.8 cM	51 cM
LG08	81.8 cM	60	49 cM	59.3 cM	51.5 cM	42 cM
LG09	99.3 cM	77.6	70 cM	71.3 cM	64.6 cM	52 cM
LG10	88.8 cM	66.6	52 cM	25.4 cM	24.1 cM	21 cM
LG11	18.0 cM	14.3	14 cM			
LG12	8.4 cM	8.4	8 cM			
LG13	104.2 cM	102.4	84 cM	31.8 cM	31.8 cM	29 cM
LG14	37.6 cM	37.6	32 cM			
LG15	63.4 cM	46.7	38 cM	11.0 cM	11 cM	10 cM
LG16	52.8 cM	51.6	41 cM	35.9 cM	36.1 cM	32 cM
LG17	73.5 cM	68.6	35 cM			
LG18	50.8 cM	47.6	35 cM			
LG19	82.7 cM	94.5	57 cM	12.3 cM	12.3 cM	11 cM
LG20	23.3 cM	23.3	20 cM			
LG21	108.5 cM	83.9	56 cM			
LG22	128.9 cM	116.5	61 cM			
LG23	62.6 cM	54.8	43 cM			
LG24	58.1 cM	29.6	27 cM			
LG25	111.5 cM	122.2	57 cM	31.0 cM	31 cM	25 cM
LG26	153.6 cM	119.8	58 cM			
LG27	107.0 cM	86.1	53 cM	15.1 cM	15.1 cM	13 cM
LGZ	83.0 cM	67.0	53 cM			
Total	2208.5 cM	1872.7	1354 cM	596.2 cM	549.7 cM	470 cM

APPENDIX 5.7

Implications of a full-sib design

The use of a full-sib design has implications for the availability and ratios of the different marker types (FI, MI, BI), and for the linkage phases of the informative markers. In an outbred full-sib design, there are four different (autosomal) chromosome sets present in the P-generation (Fig. 5.9). Consequently, there are sixteen possible non-recombinant chromosome combinations in the F_1. Fig. 5.9 and 5.10 show only four of these combinations, because the other twelve (three sets of four) have the same implications. Each set contains: *i* F_1 male and female with the same chromosome combination, *ii* F_1 male and female with the same maternal chromosome and different paternal chromosomes, *iii* F_1 male and female with the same paternal chromosome and different maternal chromosomes, *iv* F_1 male and female with four different chromosomes. These four types of F_1 combinations are shown in Fig. 5.9 and 5.10, which illustrate the effects of a full-sib cross on the 1:1 and 3:1 segregating markers respectively.

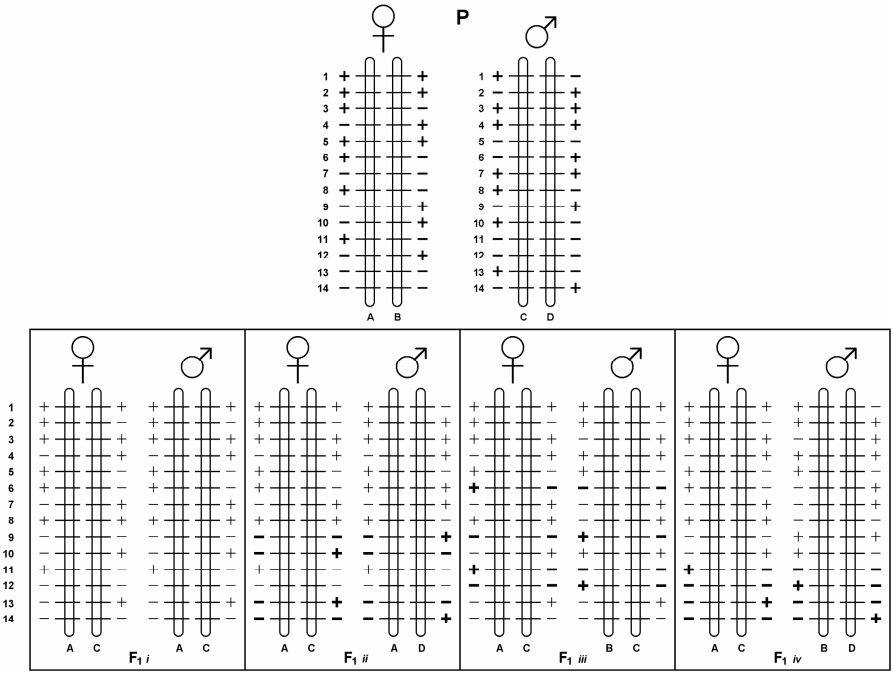


Figure 5.9 Marker associations in P- and F_1 generation of an outbred full-sib cross. The P-generation has four different chromosomes (A, B, C, D). The fourteen loci on these chromosomes cover all the genotype combinations that are potentially informative in the offspring ("+" = dominant allele, "–" = recessive). FI and MI markers are in bold, (the unmarked loci are uninformative allele combinations and BI markers). F_1 *i* has no FI or MI markers, while the other combinations have two of each. Recombination is not included in this scheme.

When the male and female have the same chromosome combination, there are no MI or FI markers available (Fig. 5.9 F_1 *i*). However, MI and FI markers can (but not necessarily do) occur if the P-male chromosomes recombine within the region in which the markers are located before they are inherited by the F_1 male. Thus a consequence of this full-sib design is that there is a deficiency of FI and MI markers in 25% of the linkage groups, and chromosome prints cannot be determined directly for them without FI marker information available. The chromosome prints for these groups must therefore be reconstructed in an alternative way as described in Appendix 5.1. With one chromosome in common (i.e type *ii* and *iii*), FI and MI markers do occur, but their linkage phases are all identical as can be seen in Fig. 5.9, where the "+" signals are restricted to one of the chromosomes in both individuals.

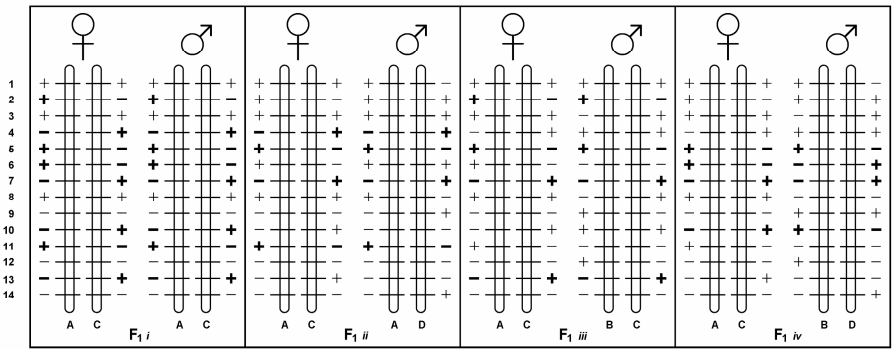


Figure 5.10 Same F_1 combinations as in Fig. 5.9, but now with only BI markers highlighted. Combinations *i*, *ii* and *iii* have marker combinations in fixed phases (i.e. "+–/+–" or "–+/–+"). F_1 *iv* has independent marker associations (i.e. "+–/+–"; "–+/–+";"+–/–+";"–+/+–").

The presence of chromosomes that are identical by descent in both F_1 male and female (i.e. type *i*, *ii*, *iii*) results in informative[1] BI markers with exclusively peakpresents or exclusively peakabsents within an F_2 individual if the chromosome that was inherited from the F_1 male is non-recombinant[2]. This is demonstrated in Fig. 5.10, where maternal and paternal syntenic BI marker values (i.e. + or –) are either exclusively identical or exclusively opposite (in *i*, *ii*, *iii*). E.g., If an F_2 individual inherits the "A" chromosome of the F_1 female from the type *ii* combination in Fig. 5.10, markers 4 and 7 are informative because they are peakabsent on that chromosome. The same F_2 individual inherits either the "A" or the "D" chromosome from the F_1 male. In case of the "A" chromosome, all markers (4 and 7) are peakabsent, whereas the "D" chromosome gives exclusively peakpresents for these loci. This example is representative for all chromosome combinations in the F_2 offspring (of type *i*, *ii* and *iii* crosses).

Recombination (in the F_1 male) results in a mix of both peakpresent and peakabsent BI markers within an F_2 individual. The linkage groups with four different chromosomes in the F_1 (type *iv*) have no restrictions on BI marker combinations.

[1] "Informative" BI markers are those that remain after censoring, i.e. those that inherited a peakabsent from the F_1 female and thus have only a male-informative component.

[2] Or a F_1 male inherited chromosome that is at least non-recombinant within the region for which markers are available. The female chromosome is obligatory non-recombinant.

Chapter 6

Quantitative Trait Loci affecting the size of wing eyespots
in a seasonally polyphenic butterfly[1]

Arjèn E. van't Hof
Patrícia Beldade
Fanja Kesbeke
Helene de Vos
Paul M. Brakefield
Bas J. Zwaan

ABSTRACT

The afrotropical butterfly *Bicyclus anynana* has two strikingly different seasonal forms, with rows of conspicuous large eyespots along the wing margins in the wet season, and an overall dull brown appearance with small eyespots in the dry season. The wet season form provides a functional strategy to escape from bird attacks, while the dry season form avoids attacks from visual predators by crypsis against dry vegetation. The question of how such extreme phenotypic variation can emerge from a single genome has sparked a broad range of evo-devo research to unravel the biochemical and genetic mechanisms underlying eyespot formation. A substantial body of knowledge about the genetic and hormonal regulation of eyespot pigmentation and size determination has accumulated, most notably the strong correlation between the dynamics of ecdysteroid titres and eyespot size. These achievements often relied on *a priori* assumptions that were based on the function of genes and the effect of their alleles in other insects, which possibly prevented full disclosure of the components involved. A QTL analysis was performed to open new directions in eyespot formation research, since prior genetic or biochemical knowledge is not required to detect chromosomal regions harboring genes that affect variation within a trait with this approach. A cross between truncated selection lines for large and small eyespots was used for the analysis, which resulted in six chromosomes with QTLs above threshold. One QTL was close enough to potentially coincide with *phantom* or *ecdysone receptor*, which are involved in ecdysteroid synthesis and recognition respectively, some are in poorly-defined regions that may contain known candidate genes and two QTLs are in chromosomal regions without eyespot candidates. Closer examination of these newly found regions could pinpoint the genes responsible for the observed variation and provide new insights in butterfly eyespot size determination and possibly in mechanisms behind polyphenism.

[1] This chapter will be submitted to a scientific journal after the most relevant candidate genes have been mapped relative to the QTL peaks.

INTRODUCTION

The afrotropical butterfly *Bicyclus anynana* displays extreme seasonal polyphenism, which gives it the ability to match the adult phenotype with the ecological challenges of different seasons (BRAKEFIELD and REITSMA 1991). The morphological seasonal differences in *B. anynana* are most apparent in the eyespots on the ventral wing surfaces. During the colder dry season, the wings display small eyespots enhancing crypsis against the dried out vegetation. In the warmer wet season the wings are conspicuous with large eyespots arranged along the wing margins (WINDIG *et al.* 1994). Such a pattern is likely to function in deflecting or mis-directing some attacks of birds away from body parts that are more vital and would allow a firmer grasp. The attack is directed towards the eyespot 'targets' on the wing edges which can then tear away readily if the insect is grasped so that the prey escapes, albeit having lost a little wing tissue (LYYTINEN *et al.* 2004; OLOFSSON *et al.* 2010).

The translation of ambient temperature into size variation of eyespots has two key characteristics. The difference in eyespot size between the seasons is triggered by temperature during the late larval stage (BRAKEFIELD *et al.* 1996) and there is a reaction norm with a gradual response to a temperature gradient. The eyespots are formed around a focus in which morphogens are released, that results in concentric concentration gradients on the developing wing surface (BRAKEFIELD and FRENCH 1995). A threshold response of the target tissue to the spatially variable morphogen titre determines the pigmentation of the wing scales in the adult. Fully developed eyespots are white in the centre, surrounded by a larger black area and a gold ring that presumably correspond with high, intermediate and low morphogen concentrations respectively.

Considerable progress has been made in identifying separate components of butterfly eyespot formation. Ecdysteroids induce larger eyespots at an early pupal stage (KOCH *et al.* 1996; ZIJLSTRA *et al.* 2004) and the expression of *Distal-less* (*Dll*) (BELDADE *et al.* 2002a; CARROLL *et al.* 1994; KOCH *et al.* 2003; MCMILLAN *et al.* 2002; REED and SERFAS 2004), *Notch* (REED and SERFAS 2004), *hedgehog, patched*, *cubitus interruptus, engrailed/invected* (KEYS *et al.* 1999) and *ecdysone receptor* (*EcR*) (KOCH *et al.* 2003) spatially coincide with the location of eyespot foci during different stages of wing development. The golden ring area of the *B. anynana* eyespots is fully associated with the *engrailed* expression pattern (BRUNETTI *et al.* 2001; SAENKO *et al.* 2008), and *Dll* and *spalt* play a role in the formation of black wing scales (BRUNETTI *et al.* 2001). Other genes reported to be involved (or potentially involved) in butterfly eyespot formation or butterfly wing-pattern formation in general are *vermilion*, *white*, *cinnabar* (REED and NAGY 2005), *ruby, vermilion* (BELDADE *et al.* 2005), *decapentaplegic, scalloped* (CARROLL *et al.* 1994; MCMILLAN *et al.* 2002) *Ultrabithorax* (WEATHERBEE *et al.* 1999), *wingless* (MCMILLAN *et al.* 2002; MONTEIRO *et al.* 2006; SAENKO *et al.* 2010), *Dopa decarboxylase* (KOCH *et al.* 1998), *Smad* homologs (*Smad on X*, *Mothers against dpp*, *Medea*) (MONTEIRO *et al.* 2006), *ultraspiracle* (*USP*) (KOCH *et al.* 2003), *APC-like*, *groucho*, *naked cuticle*, *cinnamon*, *echinus*, *Catalase*, *Heat-shock protein 70* (*Hsp70*), *Henna*, *split ends*, *scabrous* (BELDADE *et al.* 2009) and the *achaete-scute complex* (*AS-C*) gene family (GALANT *et al.* 1998), which is composed of *ASH1-3* and *asense* in Lepidoptera (ZHOU *et al.* 2008). A number of additional genes have not been proposed as Lepidoptera pigmentation genes as such, but may be involved because in other insects, they are associated with orthologs of genes involved in butterfly eyespot formation. The *Notch* receptor is inhibited by *numb* in *Drosophila melanogaster*

(FRISE 1996), *Heat-shock cognate 70* (*Hsc70*) upregulates the response to 20-hydroxyecdsone in *Helicoverpa armigera* (ZHENG *et al.* 2010), *Hsc70* and *Heat-shock protein 90* (*Hsp90*) activate the EcR/USP heterodimer in *D. melanogaster* (ARBEITMAN and HOGNESS 2000), *immunophilin* is part of the ecdysone receptor complex in insects (SONG *et al.* 1997), and the Halloween genes *spook*, *phantom*, *disembodied*, *shadow* and *shade* are involved in ecdysteroid biosynthesis in *Spodoptera littoralis* and *Bombyx mori* (IGA and SMAGGHE 2010; MAEDA *et al.* 2008; NIWA *et al.* 2004).

The temperature response is not the only factor that determines the eyespot size because there is also heritable variation in eyespots within the dry- and wet season forms. Selection lines for large and small eyespots under standardized temperatures (i.e. minimal variation in environmental effect) responded rapidly to artificial selection on dorsal eyespots (with correlated responses ventrally) (BELDADE *et al.* 2002b; BELDADE *et al.* 2002c; MONTEIRO *et al.* 1997) and on ventral eyespots (BRAKEFIELD *et al.* 1996). The relatively quick leveling of the selection response (BRAKEFIELD *et al.* 1996; WIJNGAARDEN and BRAKEFIELD 2000) suggested the contribution of a limited number of quantitative trait loci (QTLs), as was confirmed by quantitative genetic analyses (BRAKEFIELD *et al.* 1996; WIJNGAARDEN and BRAKEFIELD 2000). The genotypically (and phenotypically) diverged ventral eyespot lines provide the basis for a QTL analysis to explore the genes that underpin natural genetic variation for this trait, which could bring us closer to understanding how a single genome can form the basis of different functional adult forms.

QTL analysis of morphological traits reveals an accumulation of gene expression events caused by underlying allelic variation rather than describing developmental processes probed at a fixed moment in time, effects of a limited number of genes, or processes taking place only in the examined target tissue. This can in some cases give QTL mapping an advantage over gene expression analysis techniques such as quantitative PCR, microarrays, RNAi silencing, immune response, or EST expression profile analysis when exploring the underlying causes of variation. Besides, prior genetic or biochemical knowledge is not essential for QTL analysis. QTL mapping is not superior to the gene expression techniques though, but it provides a top-down approach that starts from variation and attempts to narrow down to genes or regulatory elements rather than exploring the effects of a selection of genes. A major drawback of QTL mapping is that QTL phenotypes typically do not pinpoint an exact genomic location, but rather identify a region on a chromosome which subsequently needs to be screened for candidates. The involvement of these candidates then needs to be verified with gene expression assays. QTL analysis can still be extremely useful even when many factors influencing a trait have already been identified. One of the key advantages of a QTL approach is the direct link of the identified gene region to a phenotypic effect. The gene-based genetic linkage map of *B. anynana* (BELDADE *et al.* 2009) includes a number of visual eyespot mutants and eyespot candidate genes and the degree in which QTL signals for eyespot size coincide with these mapped genes could give an indication of the level in which the currently available set of genes is determining (part of) the standing genetic variation of the eyespot size trait. A full match between so far identified eyespot genes and QTLs is not expected though because genes involved in eyespot formation do not necessarily have an effect on eyespot size. Neither should the list of candidate genes be limited to those that are involved in pigment formation, since temporal shifts of hormone titers caused by heterochronic polymorphisms can also change eyespot morphology (KOCH *et al.* 2000; NIJHOUT 1999; ZIJLSTRA *et al.* 2004).

So far, the transcription factor *Dll* is the only isolated gene that has been directly associated with eyespot size variation in *B. anynana*. A QTL analysis revealed significant co-segregation between different alleles of this gene and dorsal forewing eyespot sizes (BELDADE *et al.* 2002a). It is still premature to label *Dll* as an eyespot size gene rather than an eyespot size candidate gene though, because a closely linked polymorphism may be responsible for the observed phenotypic variation instead (BELDADE *et al.* 2002a). Since ecdysteroid titers are strongly correlated with the reaction norm of eyespot size (OOSTRA *et al.* 2010), genes involved in producing these hormones and those that are triggered thereby rank highest as additional candidates. The *B. anynana* genes in the ecdysteroid pathway that are currently identified are *EcR, Hsp70*, *Hsp90, Hsc70* and *immunophilin,* but *USP* and the Halloween genes may also be involved. Genes that are not directly related to ecdysteroid signal transduction can however also play a role since the association between ecdysteroid titre and eyespot size can diminish under artificial selection (ZIJLSTRA *et al.* 2004). The most likely visible mutant to coincide with an eyespot size QTL is *Bigeye* because it enlarges all ventral eyespots (BRAKEFIELD *et al.* 1996), while *067*, although also affecting eyespot size (BRAKEFIELD *et al.* 2008) is less likely to emerge because it does not affect the eyespot that was used for truncating selection. Juvenile hormones have been suggested to potentially affect eyespot size because they are involved in polyphenic traits in insects (NIJHOUT 1999; WIJNGAARDEN *et al.* 2002). These hormones and their associated genes are no longer considered eyespot candidates though since juvenile hormone titres are not influenced by temperature during the critical period for eyespot size determination (OOSTRA *et al.* 2010).

Seasonal polyphenism and the expected effect size of QTLs. There has been some debate and a shifting insight both on the amount of QTLs and the distribution of their effects involved in quantitative traits (ORR 2005). It has been argued that QTLs of large effects may be uncommon (or short-lived) because they are likely to become fixed, while QTLs of small effect disappear through drift (reviewed in Orr 2005). Following this line of thought, the majority of QTLs found in natural populations must be of intermediate effect. This could apply under many circumstances, but may be too generalized for a number of systems. The *B. anynana* eyespot quantitative trait is a response to seasonally different challenges. Evolutionary forces have maintained, and presumably even increased the distance between the extreme phenotypes, which could suggest that balancing selection preserved genetic variation. However, the narrow sense h^2 for eyespot size is high (around 0.5; (BELDADE *et al.* 2002a; MONTEIRO *et al.* 1997; WIJNGAARDEN and BRAKEFIELD 2000; WIJNGAARDEN and BRAKEFIELD 2001)), thus arguing against balancing selection which would erode additive genetic variation (ROFF 1997). A more likely scenario is that different genes are expressed during the dry and the wet season. This would prevent selection against alleles that are beneficial to adults eclosing in the following season.

The selection for eyespot size affected the elevation of the reaction norm rather than the slope. Genes affecting the slope of the reaction norm are expected to be more closely associated with the temperature-induced seasonal polyphenism though. However, the slope is far less susceptible to artificial selection than the elevation and it has an asymmetric response. A steeper slope was established after three selection attempts ((WIJNGAARDEN and BRAKEFIELD 2001; WIJNGAARDEN *et al.* 2002) P.M. Brakefield unpub. data), but selection failed to reduce the temperature-phenotype association. This apparent immunity to selection aimed at leveling the slope suggests a high degree of fixation for polyphenism genes, which is in stark contrast with the

'elevation' genes. This difference between slope and elevation response exposes the complexity of eyespot gene interactions and exploring these quantitative dynamics may serve to broaden our understanding about QTL evolution and maintenance of genetic variation in a more general perspective than in wing-pattern formation alone.

The aims of this study are to: (*i*) describe the number and contributions of QTLs to variation in wing pattern traits, (*ii*) link the QTL positions to candidate genes involved in eyespot formation and phenotypic plasticity, and (*iii*) relate this to the evolution of plasticity and selection in seasonal environments.

METHODS

Selection lines, Cross design and trait measurements. Artificial selection for eyespot size resulted in lines that produce only one of the two alternative seasonal forms across all rearing temperatures (BRAKEFIELD *et al.* 1996). The High (H) line produces the wet season form (wsf) and the Low (L) line the dry season form (dsf). Only the ventral eyespots express the plasticity, effectively behaving as a module, producing all larger or all smaller eyespots (with partial exception of posterior portion of the white area of the 5th ventral forewing eyespot).

The design and rearing conditions of the cross that was used for the QTL analysis is described in detail in (VAN'T HOF *et al.* 2008). In short: The grandparents (P-generation) of the cross were representatives of two lines that resulted from truncated selection on the fifth ventral hindwing eyespot. The grandmother came from the H line with large eyespots and the grandfather from the L line with small eyespots. The F_2 was obtained from a full-sib (brother-sister) F_1 cross. The effect of temperature-induced phenotypic plasticity was reduced to a minimum by rearing the crosses at 23ºC, which is an intermediate between the temperatures that produce large and small eyespots. The genotyped individuals were selected from the extremes of the phenotypic distribution (selective genotyping) because the intermediate phenotypic values only give weak marker-phenotype associations.

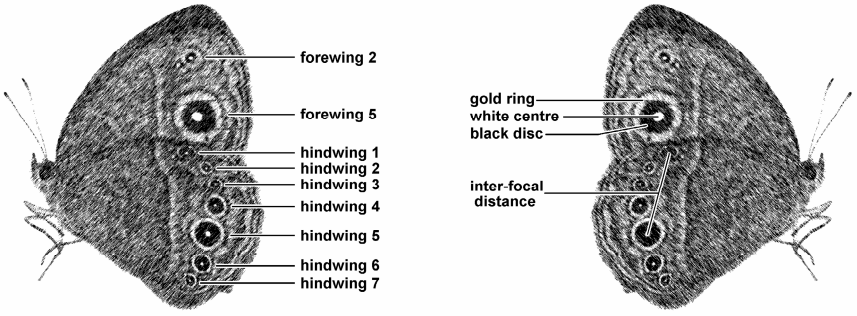


Figure 6.1 Eyespot numbering and eyespot pigment areas.

The wings were photographed with a Leica DC 2000 attached to a microscope capturing an image of 3.3×2.6 cm in 1.3 megapixels 8bit-RGB. The surface of the

white and black areas of forewing eyespots 2 and 5, and hindwing eyespots 2, 4, and 5, and the inter-focal distance (ID), which is the distance between the centre of the first and fifth hindwing spots (Fig. 6.1) were measured with Scion Image version 3b (Scion corporation, MD, USA).

Statistic analysis of the phenotypic variation. All Statistic analyses were performed separately for males and females using the phenotypic values of the ventral right fore- and hindwing. This general statistical description of trait values was based on all offspring rather than the subset of most informative individuals that was used for QTL analysis (explained below). The main trait of interest was the black area of the fifth hindwing eyespot (Fig. 6.1) which is abbreviated hrb5 onwards ('h' for hindwing, 'r' for right, 'b' for black, and '5' for 5^{th}). Additionally all other traits measured in the right wings were also included in the statistical analysis. The phenotypic distributions were tested for normality with a Kolmogorov-Smirnov test to indicate whether parametric or non-parametric test should be used for the subsequent analysis of correlation between the different eyespot components. The pigmented areas were statistically analysed as surface as well as diameter (using the square root of the surface area, √surface, as a proportional measure of diameter).

The effect of wing size on eyespot size was determined with a linear regression on the diameter (√surface) of the traits relative to inter-focal distance (ID). The correlation between the different traits was calculated with a non-parametric Spearman-rank test.

The grandparents (P-generation) were raised at different temperatures to synchronise the adult stages (to set up crosses). Therefore, the trait values of the P-generation cannot be compared with the F_1 and F_2, which were reared at 23°C. The average hindwing 5^{th} eyespot surfaces (mm^2) of the High and the Low line reared at 23°C by Wijngaarden & Brakefield (2000) are calculated from the diameter of the black area ($\pi/4 \times d^2$), which also includes the white spot, and these values are therefore slightly larger than for the black area alone.

QTL analysis. The current QTL analysis is confronted with a combination of (*i*) a full-sib cross (*ii*) dominant AFLP markers and (*iii*) absence of recombination in females. This prevents standard QTL mapping procedures such as composite interval mapping from being used and demands a careful analysis approach and interpretation of the results (Appendix 6.1).

The absence of recombination in the mother results in a fixed maternal segregation pattern. This pattern is named the chromosome print (YASUKOCHI 1998), which defines which chromosomes an individual received from its mother. This information can be used to divide offspring into two groups according to the maternal chromosome they possess. This grouping needs to be performed for each chromosome pair separately (i.e. 28 times) because the sets of individuals that make up the groups are completely different per chromosome. Because these maternal chromosomes are non-recombinant, all grouped offspring will share the same maternal QTL allele (if a QTL is present on that chromosome). The analysis can then proceed in three different ways.

Firstly, a significant difference in phenotype values could emerge when offspring with maternal chromosomes carrying a QTL are separated into two chromosome print groups. This method has been introduced by (BAXTER *et al.* 2008) to explore QTLs associated with mimetic wing patterns in Heliconius butterflies. There is a great deal

of background noise inherent to this method though because the paternal QTL alleles also contribute to variation, but this variation is not grouped and additionally QTLs on other chromosomes also generate noise. The noise generated by QTLs on other chromosomes is not unique to this analysis, and is in fact inherent to QTL dynamics, but in this case it generates an accumulation of background noise together with the paternal phenotypic contribution on the examined chromosome. The comparison between the phenotype distributions of individuals that were grouped based on the chromosome print is performed with a non-parametric Mann–Whitney Rank Sum Test because the phenotypes of the genotyped samples are not expected to have a Gaussian distribution after selective genotyping. The analysis was performed as a one-tailed test because the subsets are characterized by having either the maternal High-line or maternal Low-line chromosome, which gives a directional expectation rather than a difference in either direction.

Secondly, with all maternal QTL alleles being equal per offspring subset (as divided by chromosome print), the variation within such a group is caused by the paternal allele, and thus such a subset behaves like a normal male informative backcross. Since the censored AFLPs also behave as male informative backcross, the phenotype and genotype are fully compatible, albeit that QTLs on other chromosomes still introduce the inevitable background noise (not unique to this approach). This backcross genotype-phenotype association can be used for interval mapping (IM), which is the only reliable method for positional QTL analysis given the current (suboptimal) experimental setup. There are a number of drawbacks though, which could make the positive results less reliable because the number of replicate analyses is increased, and it also increases the possibility that a number of QTLs remain undetected. The success of QTL analysis partly depends on sample size, and 92 offspring would provide a suitable number, but here it needs to be divided in two (chromosome print) subsets and analysed separately, which halves the sample size per analysis. Moreover, males and females cannot be analysed together because the eyespots are sexually dimorphic and thus the subsets are split in two once more. Therefore, IM analysis needs to be performed in four subsets per autosome and in two for the Z chromosome, giving a total of 110 ($27\times4 + 2$)[1] analyses using only 23 individuals on average for the autosomes. The number of individuals that could be analysed was even lower for the chromosomes with chromosome prints that could not be fully reconstructed (VAN'T HOF *et al.* 2008). Using artificial values in chromosome prints has a negligible negative effect combined with a large positive effect on linkage mapping quality, but randomly assigned values cannot be used for QTL analysis and the corresponding individuals (with random chromosome print assignment) must be excluded from analysis.

The effect of dominance also needs to be taken into account when interpreting the results of these procedures. All individuals per group have the same maternal QTL allele, and if this allele happens to be completely dominant, the paternally inherited variation at this locus is completely lost and no QTL signal will be detected. The other group will receive only the recessive maternal QTL allele and in this case, all paternal QTL alleles contribute to the phenotype. Therefore, it is possible that one subset of individuals shows that there is a strong QTL signal on a chromosome, while the other subset representing the same chromosome reveals nothing whatsoever. Although

[1] 27 is the haploid autosome number, 4 represents two maternal linkage phases and two sexes, and 2 covers males and females for the Z chromosome

incomplete dominance would not completely eliminate the paternal phenotypic effect, the strength of the QTL signal will be diminished and is less likely to exceed the threshold value. These implications of dominance are not a unique feature of the current data treatment, they also occur in normal backcrosses. It is however more paradoxical if such seemingly inconsistent patterns emerge from within a single cross.

A third approach to detect QTL signals makes use of co-segregation of QTL and single markers with markers being nearest to the QTL having a higher genotype-phenotype association than more distant markers. This single marker analysis (SMA) can be an acceptable method to detect QTL positions in spite of the non-recombinant mother if the results are interpreted with care. Markers that are distant from the QTL would give overestimated values caused by the maternally inherited chromosome, and should be interpreted accordingly. However, the marker with the highest value (i.e. closely linked to the QTL) should give a good representation of the actual QTL effect at that position because it is made up of the largest paternal phenotypic effect on a chromosome and the maternal phenotypic baseline. This should only work when using co-dominant BI markers though, because the phenotypes also originate from BI QTL polymorphisms (i.e. HL×HL cross). Most of the AFLP-based markers are dominant BI and can therefore not be used for SMA. In contrast, the microsatellites and the co-dominant BI AFLPs do give proper 1:2:1 segregation and can be used for this purpose. Chromosome prints combined with MI markers easily translate into co-dominant BI markers and can also be used. SMA was not performed for all chromosomes, but rather as confirmation of the QTLs found with IM because it is only reliable with a single QTL per chromosome. The phenotypic values were also compared with the three co-dominant genotypes to verify whether they matched the expected distribution (i.e. LL = small, HL = intermediate, HH = large). The SMA and IM were performed with WinQtlCart version 2.5 (WANG *et al.* 2010) using SF2 cross (selfing cross with two generations) for SMA and B1 (backcross) for IM. Calculated LOD thresholds for IM were based on 1000 permutations with default significance level 0.05. The LOD thresholds corresponding with lower significance levels were calculated for QTLs above the default threshold (also based on 1000 permutations). The primary trait was the size of the hrb5, but additionally, the black and white surfaces of all other eyespots were also analysed with IM.

Linking B. anynana linkage maps. Besides the AFLP based linkage map for *B. anynana* presented in (VAN'T HOF *et al.* 2008), a gene-based linkage map has also become available (BELDADE *et al.* 2009). The gene-based linkage map was constructed from twelve families each with 22 offspring genotyped. Family 12 of the gene-based map is the same family as used for the AFLP map and therefore, the two maps can be linked to reveal which LGs correspond with each other. The majority of the chromosomes could be linked using the chromosome prints (i.e. identical maternally inherited segregation pattern). Family 12 of the gene-based linkage map showed the same shortcoming as the AFLP markers, namely absence of female informative (FI) markers for 12 of the 28 LGs. Therefore, the chromosome prints for these gene-based LGs were reconstructed as described in supplement S2 of (VAN'T HOF *et al.* 2008), so that they could be matched to the reconstructed chromosome prints of the AFLP-based LGs. Three chromosomes of the gene-based linkage map (9, 27 and 28) did not have markers that were suitable for chromosome print reconstruction in family 12 and the links for these three chromosomes were

established based on identical or nearly identical MI segregating markers between the gene-based- and the AFLP linkage maps.

A full spatial integration of the genes and AFLPs was not possible because most AFLPs are dominant BI, which leaves only 11 individuals (out of 22 genotyped offspring) on average for comparison after censoring, which is too few for reliable map integration. The MI segregation patterns of markers in the AFLP map close to QTL peaks were compared with the gene-based map to get a rough estimate of the chromosomal region that contains the QTL.

Predicting the position of candidate genes based on conserved synteny. There is a high degree of shared synteny between *B. anynana* and the domesticated Silkworm *Bombyx mori*, demonstrated by high-density parallel links between the *B. anynana* gene-based linkage map (BELDADE *et al.* 2009) and the SilkDB *B. mori* genome assembly (WANG *et al.* 2005). This gives the opportunity to predict the position of the candidate genes that have not yet been mapped in *B. anynana*. Genes in *B. mori* present in regions that are undisturbed between the two species (i.e. conserved synteny blocks) are expected to be in the same chromosomal region in *B. anynana*. The sequences of the unmapped candidate genes were obtained from NCBI, if possible for *B. mori*, otherwise from *D. melanogaster*. These genes were then identified within the SilkDB annotated genes and genome assembly with blastx and tblastx respectively (ALTSCHUL *et al.* 1990). The tblastx search was necessary because some genes are not annotated in SilkDB version 2.0 (WANG *et al.* 2005). E-value thresholds of e-30 and e-50 were used for *D. melanogaster* vs SilkDB using tblastx and blastx respectively. The tblastx threshold was set less stringent because the values of the different exons are not combined in the tblastx report, thus giving a large under-representation of the actual similarity. Higher e-value thresholds of e-70 (tblastx) & e-100 (blastx) were used for *B. mori* vs SilkDB since there should be a (near) perfect match when performing within-species comparison. A within-species match that is not perfect does not necessarily indicate a paralog hit though. Differences between query and annotated target can be caused by (*i*) alternative splicing, (*ii*) genetic variation, (*iii*) gene-prediction errors. The latter of these three is the case for a substantial proportion of the SilkDB annotated genes (version 2.0), and is usually revealed by means of tblastx. The positions of the candidate genes in SilkDB were compared with the *B. anynana* gene-based linkage map to confirm that they are in interlinked regions that have not been re-arranged between these two Lepidoptera species.

RESULTS

Statistics of eyespot size components. The phenotypic distribution of the F_2 covered the entire expected range from nearly absent to very large eyespots (Fig. 6.2). The basic statistics for hrb5 surfaces of the F_1 and F_2 are presented in Table 6.1. As expected, the means are similar in the F_1 and F_2 with a narrower size distribution in

Figure 6.2 Extreme phenotypes in the F_2 offspring.
Hindwings of two sisters reared under identical conditions representing both ends of the eyespot size range.

the F_1. The F_1 is expected to be genetically uniform (heterozygous High/Low for all QTLs), yet for the males it has nearly the same phenotypic diversity as the F_2, and a t-test found no significant difference between the two generations. This suggests that much of the phenotypic variation in the F_2 males may not have a genetic background. The differences between the F_1 and F_2 in females are much more obvious and they are statistically confirmed with a t-test.

Table 6.1 Basic statistics of hrb5 surfaces (mm^2) in the F_1 and F_2 generation for males and females reared at 23°C. The calculated eyespot areas of the High and Low line reared at 23°C (WIJNGAARDEN and BRAKEFIELD 2000) are included for comparison, albeit that these also include the small white area.

	N	min	max	mean	Std Dev
High males hrb5+hrw5[1]	105	N/A	N/A	2.516	0.209
Low males hrb5+hrw5[1]	123	N/A	N/A	0.181	0.117
F1 males hrb5	31	0.320	2.900	1.482	0.697
F2 males hrb5	71	0.186	3.164	1.491	0.832
High females hrb5+hrw5[1]	137	N/A	N/A	3.237	0.251
Low females hrb5+hrw5[1]	113	N/A	N/A	0.108	0.123
F1 females hrb5	34	0.860	3.820	2.370	0.710
F2 females hrb5	112	0.292	5.275	2.854	1.040

[1]from Wijngaarden and Brakefield (2000)

The distribution of the phenotypes did not fit a normal distribution for each trait. Table 6.2 summarizes the results of a Kolmogorov-Smirnov normality test for all offspring (not just the selectively genotyped offspring). Males and females were analysed separately and normality was tested for eyespot surface and for the square-root thereof as a measure that is proportional to the diameter of the eyespot. Surprisingly, the female surface distributions are generally consistent with normality, while in the male, the values representing diameter (√surface) have a far better fit.

Given this inconsistency between males and females, correlation between the different traits is based on a non-parametric spearman-rank test.

Table 6.2 Kolmogorov-Smirnov test for normality of traits in the right wings.

	♀ surface P-value		♀ √surface P-value		♂ surface P-value		♂ √surface P-value	
Forewing white 2	0.098	Passed	0.431	Passed	0.111	Failed	0.027	Failed
Forewing white 5	0.128	Passed	0.432	Passed	0.042	Failed	0.066	Passed
Forewing black 2	0.002	Failed	<0.001	Failed	0.003	Failed	0.189	Passed
Forewing black 5	0.479	Passed	0.537	Passed	0.141	Passed	0.252	Passed
Hindwing white 1	0.295	Passed	0.862	Passed	0.205	Passed	0.533	Passed
Hindwing white 4	0.352	Passed	0.111	Passed	0.011	Failed	0.323	Passed
Hindwing white 5	0.156	Passed	0.163	Passed	0.401	Passed	0.466	Passed
Hindwing black 1	0.061	Passed	<0.001	Failed	0.001	Failed	0.069	Passed
Hindwing black 4	0.532	Passed	0.020	Failed	<0.001	Failed	0.002	Failed
Hindwing black 5	0.202	Passed	<0.001	Failed	0.007	Failed	0.101	Passed

Relation between wing size and eyespot size. The distance between the centre of the first and fifth hindwing eyespots (inter-focal distance (ID)) was used as a measure of wing size. This value was also used as a standard to compare forewing traits against, since there is a strong correlation between forewing and hindwing size (FRANKINO *et al.* 2005). Linear regression with ID as independent- and the square root of trait values as dependent measure revealed that only a small part of the eyespot variation was a result of wing size variation. The adjusted R^2 values, which represent the proportional effect of wing size on the traits are given in Table 6.3. Again, there are substantial differences between females and males. In males, the wing size has a negligible or even immeasurable effect on the amount of black in the eyespot, but the variation in the female is to some extent influenced by wing size (all regression coefficients are positive). The white areas are more dependent on wing size than the black areas, with the strongest effect again in females. The primary trait (hrb5) has an adjusted R^2 value of 0.082 in females, which means that nearly 92% of the variation in this trait is independent of the wing size. Therefore, the hrb5 size was not adjusted for wing size in the QTL analysis. The other traits were also left unadjusted.

Table 6.3 Relation between wing size and eyespot phenotypes. Linear regression values of the square root of right-wing trait values as dependent measure of the inter-focal distance. R^2 and adjusted R^2 are given for all traits in both sexes.

	females		males	
trait	R^2	Adjusted R^2	R^2	Adjusted R^2
Forewing white 2	0.0115	0.00251	0.00349	0.000
Forewing white 5	0.332	0.326	0.239	0.228
Forewing black 2	0.0170	0.00776	0.00163	0.000
Forewing black 5	0.203	0.195	0.00122	0.000
Hindwing white 1	0.0795	0.0711	0.0456	0.0314
Hindwing white 4	0.111	0.103	0.0637	0.0497
Hindwing white 5	0.157	0.149	0.120	0.107
Hindwing black 1	0.0925	0.0841	0.0226	0.00713
Hindwing black 4	0.0845	0.0760	0.000428	0.000
Hindwing black 5	0.0899	0.0816	0.000212	0.000

A Spearman-rank test explored the correlation between the different traits (Table 6.4). As before, a clear difference between males and females became apparent. The female traits were correlated in all combinations except for frw5-frb2, while the white and black phenotypes were mostly independent in the males. This white and black independence in the male offspring was especially the case when comparing all traits

against each other. The within-eyespot correlation was above the significance threshold for all hindwing eyespots and the forewing eyespots in females, but not in the forewing eyespots in males (Table 6.4 and Fig. 6.3).

Table 6.4 correlation between eyespot traits. Spearman-rank correlation between different eyespot traits based on all offspring. Female values are above the diagonal of black cells, male values below. Cells contain correlation coefficient (top) and p-value (bottom). Cells with p-values greater than 0.005 are in grey, the stringent p-threshold takes multiple comparisons into account. The within-eyespot values are in bold. Column and row headers: fr = right forewing, hr = right hindwing, w# = white central area of eyespot with eyespot number, b# is black area with eyespot number.

		F_2 females									
		frw2	frw5	frb2	frb5	hrw1	hrw4	hrw5	hrb1	hrb4	hrb5
F_2 males	frw2		0.436 0	**0.443** **0**	0.435 0	0.278 0.00308	0.497 0	0.413 0	0.265 0	0.324 0	0.258 0.00612
	frw5	0.257 0.0317		0.180 0.0606	**0.620** **0**	0.510 0	0.470 0	0.532 0	0.314 0	0.290 0.00212	0.298 0.00157
	frb2	**-0.00908** **0.943**	-0.201 0.111		0.602 0	0.317 0	0.556 0	0.540 0	0.660 0	0.776 0	0.696 0
	frb5	-0.127 0.297	**0.00859** **0.943**	0.747 0		0.477 0	0.593 0	0.711 0	0.739 0	0.751 0	0.759 0
	hrw1	0.110 0.362	0.263 0.0268	0.237 0.0591	0.063 0.604		0.585 0	0.627 0	**0.669** **0**	0.499 0	0.523 0
	hrw4	0.069 0.569	0.240 0.0441	0.282 0.0242	0.184 0.126	0.544 0		0.675 0	0.604 0	**0.707** **0**	0.582 0
	hrw5	0.0301 0.804	0.234 0.0492	0.484 0	0.560 0	0.471 0	0.661 0		0.776 0	0.750 0	**0.834** **0**
	hrb1	-0.227 0.0584	-0.239 0.0446	0.766 0	0.775 0	**0.254** **0.0325**	0.219 0.0664	0.548 0		0.879 0	0.906 0
	hrb4	-0.169 0.162	-0.194 0.104	0.768 0	0.877 0	0.162 0.176	**0.312** **0.00828**	0.662 0	0.853 0		0.926 0
	hrb5	-0.144 0.235	-0.191 0.111	0.761 0	0.881 0	0.151 0.207	0.253 0.0334	**0.685** **0**	0.847 0	0.971 0	

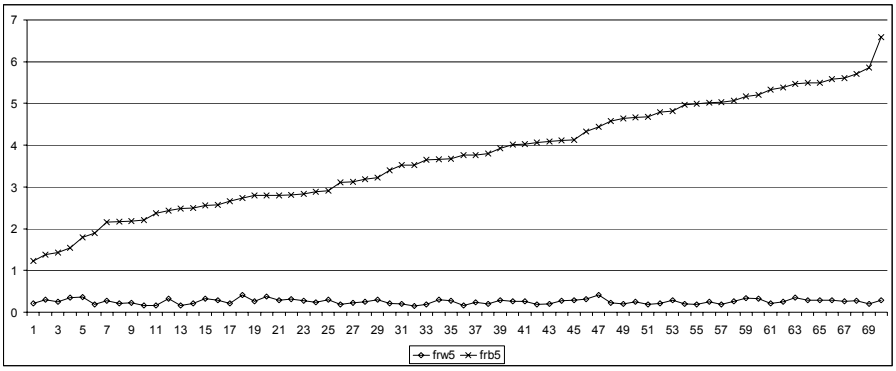


Figure 6.3 White and black area in male right forewing eyespot 5.
Male offspring are sorted by increasing size of the black area horizontally. The vertical scale represents the pigmented surface of the black and white areas (mm^2). The two traits are clearly independent.

QTL analysis. The marker-QTL association was analysed with three different approaches. (*i*) QTLs on maternally inherited non-recombining chromosomes were identified by analyzing the phenotypic effects associated with the different maternally inherited chromosomes. These differences were detected by means of a Mann–Whitney Rank Sum Test and the p-values of this test are given in Table 6.5 if a significant difference was found for hrb5. LG1, LG5 and LG24 possess clear QTL-chromosome print associations, but the effect was more obvious in the sons than in the daughters. This method identified chromosomes that contain a QTL, but it does not reveal the position of these QTLs because there is no recombination involved. (*ii*) Interval Mapping (IM) produced QTL peaks for hrb5 above the calculated threshold (1000 permutations, significance level 0.05) in six different LGs. These six LGs include the three that were identified with chromosome print–phenotype associations and additionally LG-Z, LG7 and LG13 (Fig. 6.4a-d & Fig. 6.5 a, c, d). The LOD thresholds were also calculated with more stringent significance levels to assess the strength and reliablility of the QTLs (Table 6.6).

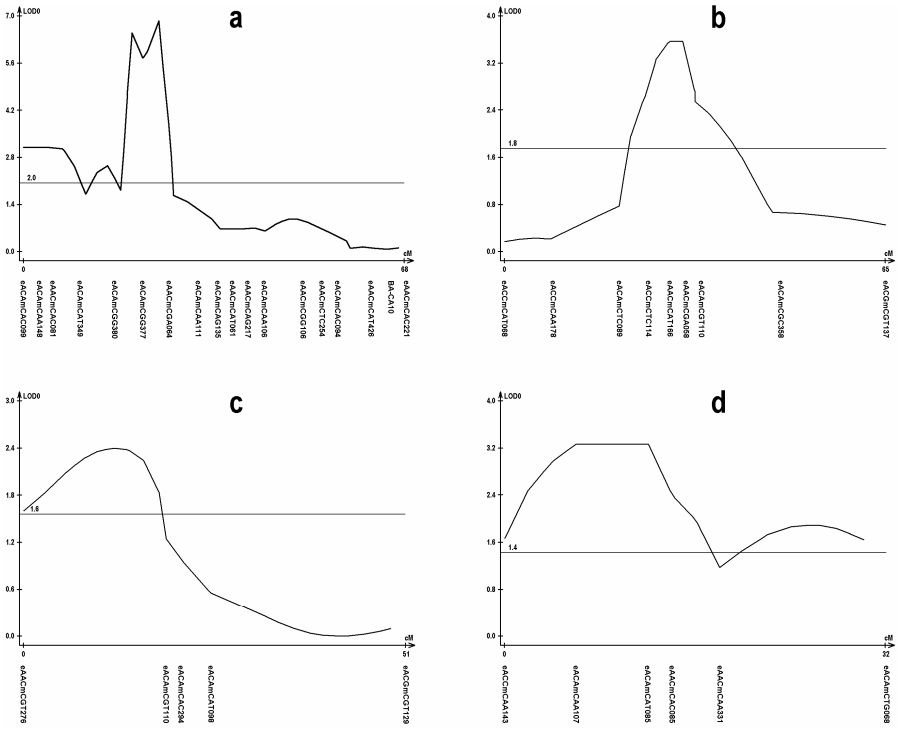


Figure 6.4 Interval mapping QTL traces of hrb5 for 4a: LG1 daughters phase 1, 4b: LG5 daughters phase 0, 4c: LG5 sons phase 1, 4d: LG7 daughters phase 0. The horizontal line is the LOD threshold calculated with 1000 permutations and significance level 0.05.

The QTL signals were usually not present in both linkage phases, which is presumably caused by dominance effects, since every analysed subset had a pool of identical maternal alleles (as explained in the Methods). Differences may also be caused by the small sample sizes per analysed subset, which are 23 on average, but may be far less if the Mendelian segregation of the maternal chromosome was skewed. The QTLs are usually not significant above the threshold in both sons and daughters per chromosome (Table 6.5). (*iii*) Single marker analysis (SMA) was used to validate the IM QTL peaks and it confirmed marker-phenotype associations when

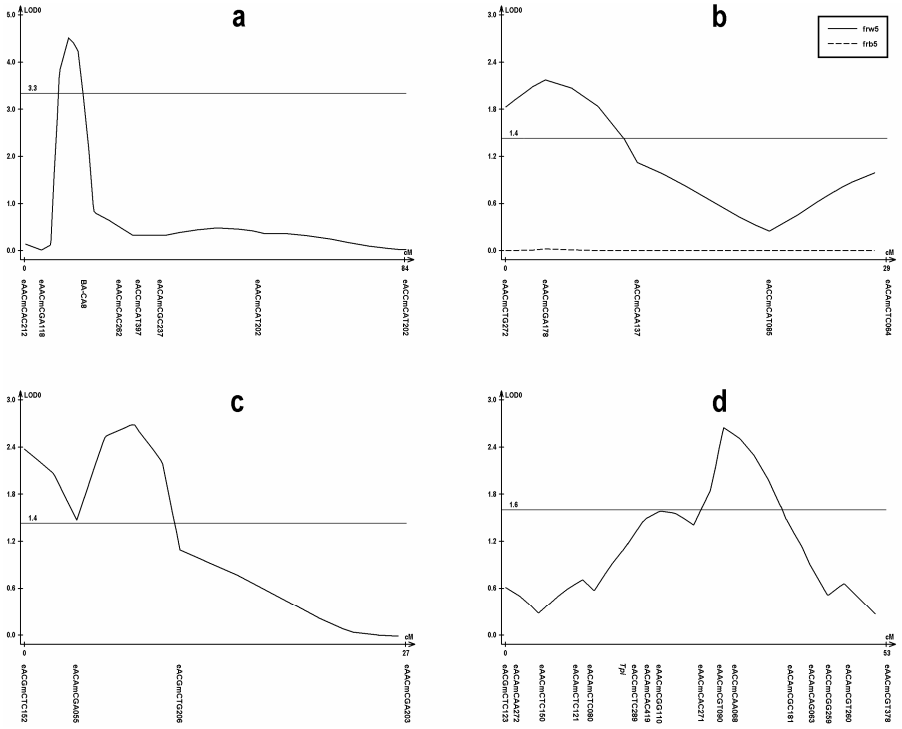


Figure 6.5 Interval mapping QTL traces of hrb5 (6.5a, 6.5c, 6.5d) and both frw5 and frb5 plotted with solid and interrupted lines respectively (5b). The graphs represent; 5a: LG13 sons phase 1, 5b: LG18 sons phase 0, 5c: LG24 sons phase 0, 5d: LG-Z daughters. The horizontal line is the LOD threshold calculated with 1000 permutations and significance level 0.05, Fig. 6.5b contains two traits, but the calculated threshold is identical for both.

possible for all examined QTL peaks, but again this was not fully consistent between sons and daughters (Tables 6.5 & 6.7). The average phenotypic values per co-dominant genotype listed in Table 6.7 show that the phenotypes are usually as expected with LL giving smaller eyespots, HH larger, and HL intermediate, but the distribution of the daughters did not match this expectation on LG1 and LG13. This can be caused by overdominance and underdominance for LG1 and Lg13 respectively, but it could also be the result of selective genotyping where the intermediate phenotypes are under-represented. This effect is expected more strongly

in the females than in the males because there were more female offspring and thus more daughters with intermediate phenotypes were excluded (both absolute and proportionally).

The additional traits examined behaved similar to the hrb5 area, albeit that the QTL signal was usually less intense. This in line with expectations because the strongest QTL signal is found in the trait that was used for truncating selection, and the correlation values in Table 6.4 confirm that most traits are correlated but not fully dependent (i.e. correlation coefficients <1). The forewing white areas in the male did not give QTL peaks that coincided with hrb5, which is unsurprising given the absence of correlation (Table 6.4, Fig. 6.3). A peak for right forewing fifth eyespot white

Table 6.5 QTL signals using different analysis methods. Only the linkage groups with QTL signals are included. Interval mapping (IM) is divided into sons and daughters and for the autosomes also into maternal linkage phase. A '+' in the IM columns indicates a QTL peak that is above threshold for hrb5. The chromosome print columns give the p-value if it is less than 0.05, which indicates that different phenotypic values for hrb5 are associated with the different maternal chromosomes. The single marker analysis (SMA) column shows a '+' when a co-dominant BI marker had a significant association with the phenotype. The SMA characteristics are presented in more details in Table 6.7. LG18 has only a frw5 QTL present in the sons. LG24 has a hrb5 QTL, but there are no MI or co-dominant BI markers available and therefore co-dominant BI SMA is not possible. LG-Z also has a hrb5 QTL, but cannot be divided into two chromosome print groups and it does not need to be analysed with SMA because the IM QTL analysis for this LG did not need to be performed in four separate groups.

LG	IM				Chr. print		SMA	
	sons		daughters		sons	daughters	sons	daughters
	ph0	ph1	ph0	ph1				
LG-Z			+		N/A	N/A	N/A	N/A
LG1				+	0.001495		+	+
LG5		+	+		0.000642		+	
LG7			+				+	+
LG13		+					+	
LG18	frw5						frw5	
LG24	+				0.004815	0.020425	N/A	N/A

(frw5) emerged most convincingly on LG18 and is clearly above the threshold, while the black area of the same eyespot produces a flat line (Fig 6.5b). The QTL is below the LOD threshold when the significance level is slightly increased though, which may indicate that the QTL effect is not very strong, or that it is in fact a ghost QTL. However, the SMA gives a significant value (Table 6.7), but the average phenotypes per marker genotype are not robustly different (Table 6.7). It should be taken into account though that eAACmCTG272 is not positioned exactly underneath the QTL peak, but approximately 3 cM to the left (Fig 6.5b), which reduces the marker-QTL association. The eAACmCGA178 AFLP is better positioned, but it is not a co-dominant marker and cannot be used for SMA.

Linking *B. anynana* linkage maps and predicting additional candidate gene locations. Links could be established unambiguously between all 28 LGs of the gene-based and AFLP-based linkage maps (Table 6.8). Twenty-five sets had identical FI segregation and the remaining three links were obtained from co-segregating MI markers. This also links the AFLP linkage map and the QTL signals thereon to a number of mapped eyespot candidates and to the lepidopteran reference genome of *B.*

mori (WANG *et al.* 2005). The LG numbers of the AFLP map have not been adjusted to match the gene-based map of *B. anynana* since it has already been published with the original numbers (VAN'T HOF *et al.* 2008). To discriminate between the AFLP-based and the gene-based LG numbers, the AFLP-based LGs will be named 'LG#', while the gene-based LGs are named 'Bany#' (they are originally named 'Bany LG#' in (BELDADE *et al.* 2009)).

Table 6.6 LOD thresholds associated with different significance levels.
Calculated LOD thresholds associated with different significance levels (1000 permutations) for all QTL signals that were above the LOD threshold with default significance level 0.05. The sign between brackets indicates whether the LOD threshold is below (<) or above (>) the LOD peak.

	LOD peak height	0.05	0.01	0.005	0.001
LG-Z daughters hrb5	2.64	1.6 (<)	2.3 (<)	3.2 (>)	(>)
LG1 daughters phase 1hrb5	6.88	2.0 (<)	2.9 (<)	3.5 (<)	6.9 (>)
LG5 daughters phase 0 hrb5	3.57	1.8 (<)	3.0 (<)	3.6 (>)	9.6 (>)
LG5 sons phase 1 hrb5	4.40	1.6 (<)	4.9 (>)	(>)	(>)
LG7 daughters phase 0 hrb5	3.27	1.4 (<)	2.5 (<)	3.1 (<)	4.3 (>)
LG13 sons phase 1 hrb5	4.10	3.3 (<)	6.5 (>)	(>)	(>)
LG18 sons phase 0 frw5	2.18	1.4 (<)	2.4 (>)	(>)	(>)
LG24 sons phase 0 hrb5	2.38	1.4 (<)	2.5 (>)	(>)	(>)

Table 6.7 Single marker analysis with co-dominant BI markers.
The average phenotypic values (mm^2) of the hrb5 trait associated with co-dominant BI genotypes are presented in columns LL, HL and HH in rows LG1-LG13. The values for LG18 represent the frw5 trait. The pr(F) value for SMA plus their significance level as calculated by QtlCart 2.5 are in the 'pr(F)' column, with *, **, *** and ****, representing significance levels of 5%, 1%, 0.1% and 0.01% respectively.

LG	marker	sons/daughters	LL	HL	HH	pr(F)
LG1	eAACmCGA064	daughters	1.867	3.305	2.625	0.009**
LG1	eACAmCGG377	daughters	1.504	3.594	2.625	0.000****
LG1	eAACmCGA064	sons	0.942	1.547	2.177	0.009**
LG1	eACAmCGG377	sons	0.962	1.522	2.177	0.013*
LG5	eACAmCGT110	daughters	2.572	2.272	4.068	0.221
LG5	eACAmCGT110	sons	0.711	1.768	2.134	0.000***
LG7	eACAmCAT085	daughters	1.417	3.215	3.071	0.009**
LG7	eACAmCAT085	sons	0.918	1.711	1.865	0.023*
LG13	BA-CA8	daughters	2.353	1.861	3.151	0.223
LG13	BA-CA8	sons	0.957	1.652	1.985	0.009**
LG18	eAACmCTG272	sons	0.297	0.242	0.240	0.022 *

The gene order between *B. anynana* and *B. mori* is highly conserved in general (BELDADE *et al.* 2009), which allows prediction of the positions of additional candidate genes. Twenty-nine unmapped candidate genes are in regions that are highly similar in *B. anynana* and *B. mori*, and their predicted chromosome assignment can be considered reliable. The prediction for three other candidates is less convincing because *ruby* is in a region that is not covered in *B. anynana* (SilkDB nscaf2868), and *Ultrabithorax* and *shade* are in regions that experienced some rearrangements between the two species, but nevertheless the predicted chromosomes are the most likely candidates for these three genes. The positions of *spook* and *Hsp90* are uncertain, because there are only relocated genes within range, their predicted positions can however still be limited to two sets of most likely chromosomes. *Spook* is presumably on Bany6 or Bany11, which do not contain QTLs. *Hsp90* is expected to

be on Bany 9 (with QTL) or on Bany 27 (without QTL). The unmapped candidate genes and the chromosome that they are assumed to be on are included together with the mapped candidate genes and visible mutants in Table 6.8. The accession numbers of the sequences used to determine the SilkDB positions and the corresponding e-values are listed in Appendix 6.2.

Although the LGs of the gene-based map and the AFLP based map could be unambiguously linked, positional integration of the markers was not possible. This is mainly a result of the dominant characteristics of the AFLPs combined with the small number of overlapping offspring (22 individuals). Additionally, a number of genes were not polymorphic in the mapping family, and some genes that did segregate are not positionally mapped in the gene-based linkage map.

Comparison of QTL positions with gene positions. MI markers that are close to QTL peaks were compared with the segregation pattern of the gene-based linkage map to provide rough estimates of the regions where the QTLs are expected to reside. This is important for further investigations of the QTLs because the AFLP markers are anonymous, while the genes provide annotated anchors. With only 22 offspring included, there is only a limited number of recombinations available to define the position of markers. A single recombination under these conditions is responsible for a mapping distance of approximately 4.5 centimorgan (100/22), which gives a rather crude resolution. It may nevertheless place a QTL signal within a limited section of the chromosome, which can be subsequently targeted by fine-mapping.

The MI AFLP eACAmCGG377 on LG1/Bany13 (Fig. 6.4a) is separated from *C5123* by one recombination, but this does not reveal an approximate position because the coverage of this chromosome is not very detailed in the gene-based linkage map. This linkage group has the strongest QTL signal for eyespot size and the candidate gene cluster *AS-C* and *scalloped* are also expected to be on this chromosome. Since the *AS-C* is 4.4 Mb apart from *C5123* in *B. mori*, it is probably not responsible for the observed phenotypic differences. It should however not be completely rejected as candidate for the QTL because *C5123* itself does not fully coincide with the QTL peak either, thus *AS-C* may be 4.4 Mb from *C5123*, but could be much closer to the actual QTL. *Scalloped* is even further apart from *C5123* in *B. mori* (~6 Mb) and therefore unlikely the source of the QTL signal.

The nearest MI marker to the LG5/Bany16 QTL peak (eACAmCGT110, Fig. 6.4b and 6.4c) is one recombination apart from *C5636* and *C1223*, which positions the QTL peak near the centre of Bany16. There are no known candidates expected on this chromosome.

On LG7/Bany10, *C5726* is between eACAmCAT085 and eACCmCAA143 with one recombination relative to eACAmCAT085 and two in the opposite direction relative to eACCmCAA143 (Fig 6.4d). The QTL signal is also positioned between these two AFLPs and closer to eACAmCAT085 than to eACCmCAA143, which suggests that *C5726* may be relatively close to the QTL. The eACAmCAA107 AFLP that coincides with the QTL (Fig 6.4d) is dominant BI which prevents reliable segregation comparison. Bany10 (corresponding with LG7) has *EcR* as mapped candidate gene, one mapped visible mutant and the potential candidate genes *phantom*, *white* and *cinnabar* that have not been mapped in *B. anynana*. Bany10 and *B. mori* chromosome 10 have diverged very little in terms of gene order and composition, which strengthens the confidence of gene position predictions in *B. anynana*. The QTL peak on LG7/Bany10 is surrounded by MI markers eACCmCAA143 and eACAmCAT085, which are each one recombination apart from

EcR and from *C5726* respectively. The unmapped ecdysteroid biosynthesis related gene *phantom* (IGA and SMAGGHE 2010) is only 2Mb 'below' the *C5726* ortholog in *B. mori* (BGIBMGA006848) which could still be considered within range of the QTL given the crudeness of the 22-individual based resolution. The positions of unmapped genes *cinnabar* and *white* are ~4.5Mb and ~6.25Mb 'below' BGIBMGA006848 in *B. mori* respectively, which is probably too distant to account for the QTL signal. There is an additional QTL region above the threshold line on the right in Fig. 6.4d, which could possibly contain these candidate genes, however this peak does not coincide with markers and may not represent a real QTL. The visible mutant *Spotty* is very distant from the QTL peak and the corresponding region is not represented in Fig. 6.4d because the chromosome is not fully covered with AFLP markers. It may be that *Spotty* is in fact in the bottom part of the LG (BELDADE *et al.* 2009), which would also position it far from the QTL peak.

The BA-CA8 microsatellite that coincides with the QTL peak of LG13/Bany8 (Fig. 6.5a) fully co-segregates with *C3090* (all 22 offspring are identical), but the gene is not positionally mapped on the gene-based linkage map, which also leaves the QTL position unspecified. This chromosome is expected to include *shadow*, which is related to ecdysteroid biosynthesis.

LG24/Bany9 (Fig. 6.5c) has no MI markers, and the position of the QTL peak can therefore not be linked to the gene-based linkage map. Bany9 includes the confirmed and expected candidate genes *Hsp70*, *echinus*, and *hedgehog*, but their relation to the QTL peak remains unclear.

The Z-chromosome QTL (Fig. 6.5d, LG-Z/Bany1) cannot be positioned because there are too few polymorphic genes available in the family of the AFLP-based map. AFLP eACAmCGT260 has a segregation pattern identical to *C1211*, but that gene is not positionally mapped. The only link between markers that are positioned in both linkage maps is *C2115*-eACAmCAA272, which have identical segregation patterns, but these markers are far from the QTL peak. Therefore it remains unclear whether the candidate genes *Henna*, *Catalase* and *scabrous* are close to the QTL.

The QTL peak on LG18/Bany25 (Fig. 6.5b) that is associated with white in the forewing eyespot 5 is near AFLP eAACmCTG272, which fully co-segregates with *C5686* and *C5713*, but these genes are not positioned. The nearest mapped genes are *C1852* and *C4388*, which both have 3 recombinations relative to eAACmCTG272 and are identical to each other in the 22 offspring. No known candidate genes are expected on this chromosome.

Table 6.8 (next page) Links between AFLP- and gene based map and distribution of candidate genes.
The first column includes the LG numbering of the gene-based linkage map (BELDADE *et al.* 2009), the second column the LG numbering of the AFLP map. The third column shows mapped candidate genes in bold, mapped visible mutants bold-underlined (mapped in (BELDADE *et al.* 2009)), and candidate genes with predicted chromosome assignments in normal font. SilkDB position gives the position of the gene orthologs in the genome assembly of *B. mori*. SilkDB gene ID lists the gene identifiers of the *B. mori* orthologs. The final column indicates whether a hrb5 QTL is detected (+) or in the case of Bany25/LG18 a rfw5 QTL. *Hsp90* (SilkDB BGIBMGA004612) and *spook* (SilkDB BGIBMGA001753) are not included because their positions are uncertain.

Gene map (Bany)	AFLP map (LG)	**mapped genes** **mapped visible mutants** predicted genes	SilkDB position	SilkDB gene ID	QTL signal
1	Z	***Henna***	nscaf2734:549603-556474	BGIBMGA003866	
		Catalase	nscaf1690:5265439-5267758	BGIBMGA000701	+
		scabrous	nscaf1690:7881371-7884088	BGIBMGA000480	
2	9	***Engrailed***	nscaf2964:4110461-4111273	BGIBMGA009797	
		Smad on X	nscaf2964:3722236-3736526	BGIBMGA009650	
		Distal-less	nscaf2964:2064122-2064817	BGIBMGA009672	
		Invected	nscaf2964:4269535-4271098	BGIBMGA009643	
3	25	*split ends*	nscaf2930:3446818-3468207	BGIBMGA009003	
4	17	***Band***	N/A	N/A	
		Dopa decarboxylase	nscaf2589:5804843-5821387	BGIBMGA003199	
		spalt related	nscaf2589:6654523-6680893	BGIBMGA002946	
		wingless	nscaf2847:4290086-4300264	BGIBMGA006146	
		ultraspiracle	nscaf2847:6419185-6441959	BGIBMGA006183	
5	16	***naked cuticle***	nscaf2674:5380977-5383479	BGIBMGA003495	
		patched	nscaf2674:2648112-2656594	not annotated	
6	27	***APC-like***	nscaf3079:978040-998734	BGIBMGA013580	
		Ultrabithorax	nscaf2853:3215464-3215931	BGIBMGA006389	
7	19	*Medea*	nscaf2986:4943147-4952874	BGIBMGA010110	
		shade	nscaf2986:2935439-2939078	BGIBMGA010239	
8	13	*shadow*	nscaf2828:4597141-4613087	BGIBMGA005496	+
9	24	***Heat-shock protein 70***	nscaf2801:598000-599981	BGIBMGA004614	
		echinus	nscaf2511:1863890-1877092	BGIBMGA002483	+
		hedgehog	nscaf3048:978705-989746	BGIBMGA012535	
10	7	***ecdysone receptor***	nscaf2855:5976936-5989189	BGIBMGA006767	
		Spotty	N/A	N/A	
		white	nscaf2575:3237764-3274876	BGIBMGA002922	+
		phantom	nscaf2860:608241-614972	BGIBMGA006936	
		cinnabar	nscaf2860:3025665-3028983	BGIBMGA006968	
11	15				
12	23	*numb*	nscaf2993:7375661-7384165	BGIBMGA010576	
		Mothers against dpp	nscaf2993:355075-361490	BGIBMGA010481	
		decapentaplegic	nscaf2993:6692859-6696343	BGIBMGA010384	
13	1	*scalloped*	nscaf1898:4585777-4618439	BGIBMGA001129	
		achaete-scute homolog 1	nscaf1898:6056628-6057209	BGIBMGA001001	
		achaete-scute homolog 2	nscaf1898:6175916-6176635	BGIBMGA001148	+
		achaete-scute homolog 3	nscaf1898:6222871-6223434	BGIBMGA001000	
		asense	nscaf1898:5940257-5941471	BGIBMGA001002	
14	11	*vermilion*	nscaf2943:1741136-1757256	BGIBMGA009276	
15	26	***Heat-shock cognate 70***	nscaf2888:8342475-8344810	BGIBMGA007950	
		Notch	nscaf2888:7526277-7586024	BGIBMGA007929	
16	5				+
17	10	***Bigeye***	N/A	N/A	
		067	N/A	N/A	
18	21	***cinnamon***	N/A	N/A	
19	8				
20	3				
21	6	*groucho*	nscaf2136:5066518-5124531	not annotated	
		immunophilin	nscaf2136:114877-119763	BGIBMGA001490	
		ruby	nscaf2868:530839-555239	BGIBMGA007196	
22	2	*disembodied*	nscaf1681:277196-285831	BGIBMGA000368	
23	4	***Cyclops***	N/A	N/A	
24	20	***cubitus interruptus***	nscaf2800:2013357-2027060	BGIBMGA004545	
25	18				rfw5
26	22				
27	14				
28	12	***Goldeneye***	N/A	N/A	

DISCUSSION

The QTL analysis succeeded in identifying a number of genomic regions that are statistically linked to butterfly wing eyespot size differences involving the seasonal polyphenism. Additional aims were to unravel the phenotypic, additive, dominance, pleiotropic and epistatic effects of these genes. This proved to be an over-ambitious goal when using a combination of a full-sib cross design, mostly dominant genetic markers, and a species that has no female recombination. The results nevertheless provide relevant contributions in piecing together the components of eyespot formation and from a broader point of view understanding genetic variation underlying phenotypic differences that is maintained despite natural selection.

QTL peaks above the LOD threshold were found on six different linkage groups for the trait that was used for truncation selection (hrb5). Integration between the AFLP based linkage map used for the QTL analysis and the gene-based linkage map constructed by (BELDADE *et al.* 2009) made it possible to link the QTL peaks to annotated LGs and in some cases determine their approximate position. The only genes that are found to be close to a QTL peak are *EcR* and *phantom* on LG7/Bany10, but this is no guarantee that they are in fact responsible for the QTL. Three other genes involved in ecdysteroid biosynthesis that occur on chromosomes with QTLs are *shadow, Hsp70 and Hsp90*, albeit that there is currently no positional information available. The Z-chromosome contains a clear QTL peak, which consistent with a previous analysis of line crosses that predicted at least one Z-linked locus responsible for 9.3% of the difference between the High and Low lines (WIJNGAARDEN and BRAKEFIELD 2000). The Z-chromosome harbors candidate genes *Henna*, *Catalase* and potentially *scabrous*, but further effort is needed to determine whether one of these coincides with the QTL. The QTL signal on LG1/Bany13 may originate from *achaete-scute complex*, but is more likely caused by a so far unidentified gene. One QTL peak is on a chromosome that does not possess known or expected eyespot candidate genes, and other QTL peaks are in chromosomal regions that were not previously associated with eyespot formation, which demonstrates that our knowledge of eyespot formation is not yet complete. It is notable that the QTL peaks do not coincide with the expected positions of *Notch, engrailed, patched*, *cubitus interruptus* and *spalt*, which are key genes in eyespot formation. The genes involved in the production of eyespots are not necessarily the genes that determine their sizes though and there has been no indication so far that these genes are in fact involved in eyespot size determination. The absence of a QTL signal on LG9/Bany2 is surprising since *Dll* convincingly co-segregated with eyepot size in other selection lines (BELDADE *et al.* 2002a). These lines were selected for eyespot sizes in the dorsal (non-plastic) forewing though, while the lines used for the current QTL analysis were selected for hindwing eyespot size, which could possibly explain the absence of a QTL peak. Ecdysone titers have been shown to influence eyespot size and the QTL peak that appears to coincide with *EcR* and possibly with *phantom* is therefore more in line with expectations. No QTL was found on LG17, which is unexpected given the strong eyespot size effect of *Bigeye* and the presence of two other mutants at the same locus that alter eyespot morphology (SAENKO *et al.* 2010). An additional reason to expect a QTL on this LG is the presence of regions that are involved in pigment patterning on the orthologous chromosomes in other Lepidoptera (GOLDSMITH and WILKINS 1995; JORON *et al.* 2006).

The majority of different examined traits was correlated within individuals, and also produced overlapping QTL peaks. Butterfly eyespots can be regulated independently (BELDADE *et al.* 2002b; MONTEIRO *et al.* 2007; REED and SERFAS 2004) but this does not seem to be the case in the High/Low lines because the correlation between black eyespot surfaces is strong and there are no QTL peaks that emerged for particular eyespots, while being absent for others. The black and white pigment areas were however often independent in the male offspring. This independence was the strongest in the forewing eyespots, and this may be explained by a QTL that was associated with the white area of the 5th forewing eyespot, but not with the black area of the same eyespot.

The confidence levels for the different QTLs is variable, with the LG1 and LG7 peaks still above threshold with a confidence level of 0.005, while other peaks are already below threshold with a confidence level of 0.01. One reason for this is the low number of individuals that could be effectively analysed in one batch with the current setup, but it could also reflect the difference in phenotypic effects of the different QTLs. Alternatively, some of the signals may be ghost QTLs (i.e. type I statistical errors), which is not unlikely with the current analysis. The data required partitioning and independent analysis in 110 subsets, which substantially increases the chance of such type I errors. The QTL on LG5 in males and the one on LG24 have to be considered with care in this respect since the QTL peaks do not coincide with genetic markers (Fig 6.4c and 6.5c). Bonferroni correction would have disqualified all QTLs, which would most likely have generated many type II errors. The most practical way to deal with this is to treat the QTL signals as good starting points for further investigation rather than as positions that harbor an eyespot gene with absolute certainty. It is reassuring though that the alternative methods (in addition to interval mapping) that assessed marker-phenotype relationships often confirm the QTL peaks.

The differences between males and females in regulating eyespot size are remarkable, but more or less in line with the findings of Wijngaarden and Brakefield (2000) who estimated different ranges for the number of loci affecting eyespot size in males and females at 23°C. Additionally, empirical evidence demonstrated that ecdysone titers peak earlier during pupal development in females than in males (ZIJLSTRA *et al.* 2004), and *Dll* has a sex-specific allelic effect on eyespot size (BELDADE *et al.* 2002a). The following explanations for these differences are purely hypothetical at this point, but there are a number of notable sex-specific differences. The QTL signal on the Z-chromosome is only present in the daughters and is absent in the sons. As a result of the WZ sex determination system, with females being the heterogametic sex, the daughters receive either one of the father's Z-chromosomes, but none of the mother's. Since the father is heterozygous High/Low, the phenotypic effect in the daughters is caused by either the High allele or the Low allele, but not a combination of alleles (the W chromosome does not contribute), which prevents heterozygous intermediates (i.e. hemizygous inheritance). Absence of dosage compensation in Lepidoptera (MANK 2009; TRAUT *et al.* 2007) could cause variation between the sexes, but it is not expected to explain the difference in this case because the QTL signal on the Z-chromosome is strong in females and absent in males, while absence of dosage compensation results in decreased gene expression in the heterogametic sex. It is possible though that these genes act as inhibitors to other eyespot genes, which would result in an increased phenotypic effect associated with the lower expression levels. The QTL on the Z-chromosome only explains part of the phenotypic variation and autosomal sex-specific QTL effects are also present in the

current cross. Such sex-specific differences between QTL effects have been frequently observed in *D. melanogaster* traits (LONG *et al.* 1995; NUZHDIN *et al.* 1997; WAYNE *et al.* 2001). QTLs for abdominal bristle number in *D. melanogaster* are highly specific to sex, with less than half detected both in males and in females and those that were present in both had opposite phenotypic effects (DILDA and MACKAY 2002).

Another difference between males and females in the F_2 of the QTL family is caused by selective genotyping, to amplify the QTL effect. The complete F_2 consists of 72 males and 112 females while only 46 of each have been included in the QTL analysis (23 from both tails of the phenotype distribution), and therefore the phenotypic differences in the analysed females were much more extreme. Additionally, the phenotypic variation in males (at least in the current cross) appears to be less dependent on the genotype than in the females, which makes it more difficult to detect marker-phenotype associations.

An F_2 cross design was used to produce the maximum phenotypic range within a single cross. This setup had a number of methodological consequences, but may also have affected the phenotypic range itself as it could potentially be reduced as a result of biased inbreeding depression. Homozygous QTLs contribute towards extreme phenotypes, but within a full-sib cross from divergent selection lines, homozygotes are identical by descent in 50% of the cases (cf. (VAN'T HOF *et al.* 2008), supplement S8). Adults with identical-descent homozygous QTLs (and associated extreme phenotypes), will not exist in the offspring if lethals are linked to these QTLs. It is unclear whether this actually occurred in this cross, but it could be a possible explanation for the relatively small difference in phenotypic range between F_1 and F_2 in males. Sex-specific inbreeding effects have been reported in *B. anynany* (SACCHERI *et al.* 2005) and may also account for with the skewed sex ratio in the offspring.

The current study suggests that the size of the eyespots is influenced by a small number of QTLs with large effect, although it is possible that QTLs with small effect have been overlooked due to the fragmented analysis (i.e. large effect QTLs are simply easier to detect than small effect QTLs). A small number of large effect QTLs does fit the expectation posted earlier, that fixation of genes causing variation is unlikely when this variation is essential in an ecological context. The appealing feature about seasonal polyphenism is the fact that different adult forms can originate from a single genome. Such differences in phenotypes are likely to be regulated by differential gene expression and since natural selection only acts on genes that are actually expressed, the dry season could create a genomic refugium for wet season genes and the wet season could do so for dry season genes and thus genetic diversity can be maintained. This hypothesis is in line with high heritabilities for the selected traits.

This QTL analysis focused on the genetic variation underlying variation in eyespot size. In a natural situation, the main determinant of eyespot sizes is ambient temperature rather than genetic variation and the current analysis does not necessarily reveal the genes that are temperature induced (although the case is the strongest for the involvement of the *EcR* and *phantom*). However, the QTL analysis has opened up new directions to explore eyespot formation on butterfly wings, which may lead towards a better understanding of the molecular dynamics of seasonal polyphenism. The wealth of genetic markers developed and positioned on the gene-based linkage map by Beldade et al. (2009) and additional genes from an EST library (BELDADE *et*

al. 2006) can be used to construct a finer-scaled and co-dominant profile of the regions where the QTLs were found. The identification of QTL candidates can be refined further by using orthologous regions on the Silkworm reference genome to develop additional genetic markers and to explore the function of genes that are expected in the region based on conserved synteny. Ultimately, a select set of candidate genes should emerge for each QTL, that can be tested for correlated gene expression.

REFERENCES

ALTSCHUL, S., W. GISH, W. MILLER, E. MYERS and D. LIPMAN, 1990 Basic local alignment search tool. J Mol Biol **215:** 403 - 410.

ARBEITMAN, M. N., and D. S. HOGNESS, 2000 Molecular Chaperones Activate the *Drosophila* Ecdysone Receptor, an RXR Heterodimer Cell **101:** 67-77.

BAXTER, S. W., S. E. JOHNSTON and C. D. JIGGINS, 2008 Butterfly speciation and the distribution of gene effect sizes fixed during adaptation. Heredity **102:** 57-65.

BELDADE, P., P. BRAKEFIELD and A. LONG, 2002a Contribution of *Distal-less* to quantitative variation in butterfly eyespots. Nature **415:** 315 - 318.

BELDADE, P., P. M. BRAKEFIELD and A. D. LONG, 2005 Generating phenotypic variation: prospects from "evo-devo" research on *Bicyclus anynana* wing patterns. Evolution & Development **7:** 101-107.

BELDADE, P., K. KOOPS and P. BRAKEFIELD, 2002b Developmental constraints versus flexibility in morphological evolution. Nature **416:** 844 - 847.

BELDADE, P., K. KOOPS and P. M. BRAKEFIELD, 2002c Modularity, individuality, and evo-devo in butterfly wings. PNAS **99:** 14262-14267.

BELDADE, P., S. RUDD, J. GRUBER and A. LONG, 2006 A wing expressed sequence tag resource for *Bicyclus anynana* butterflies, an evo-devo model. BMC Genomics **7:** 130.

BELDADE, P., S. V. SAENKO, N. PUL and A. D. LONG, 2009 A Gene-Based Linkage Map for *Bicyclus anynana* Butterflies Allows for a Comprehensive Analysis of Synteny with the Lepidopteran Reference Genome. PLoS Genet **5:** e1000366.

BRAKEFIELD, P. M., P. BELDADE and B. J. ZWAAN, 2008 The African butterfly *Bicyclus anynana*: evolutionary genetics and evo-devo. in *Emerging Model Organisms: A Laboratory Manual*, edited by R. R. BEHRINGER, A. D. JOHNSON and R. E. KRUMLAUF. Laboratory Press, Cold Spring Harbor.

BRAKEFIELD, P. M., and V. FRENCH, 1995 Eyespot Development on Butterfly Wings: The Epidermal Response to Damage. Developmental Biology **168:** 98-111.

BRAKEFIELD, P. M., J. GATES, D. KEYS, F. KESBEKE, P. J. WIJNGAARDEN *et al.*, 1996 Development, plasticity and evolution of butterfly eyespot patterns. Nature **384:** 236-242.

BRAKEFIELD, P. M., and N. REITSMA, 1991 Phenotypic plasticity, seasonal climate and the population biology of *Bicyclus* butterflies (Satyridae) in Malawi. Ecol Entomol **16:** 291–303.

BRUNETTI, C. R., J. E. SELEGUE, A. MONTEIRO, V. FRENCH, P. M. BRAKEFIELD *et al.*, 2001 The generation and diversification of butterfly eyespot color patterns. Current Biology **11:** 1578-1585.

CARROLL, S., J. GATES, D. KEYS, S. PADDOCK, G. PANGANIBAN *et al.*, 1994 Pattern formation and eyespot determination in butterfly wings. Science **265:** 109-114.

DILDA, C. L., and T. F. C. MACKAY, 2002 The Genetic Architecture of *Drosophila* Sensory Bristle Number. Genetics **162:** 1655-1674.

FRANKINO, W. A., B. J. ZWAAN, D. L. STERN and P. M. BRAKEFIELD, 2005 Natural Selection and Developmental Constraints in the Evolution of Allometries. Science **307:** 718-720.

FRISE, E., 1996 The *Drosophila* Numb protein inhibits signaling of the Notch receptor during cell-cell interaction in sensory organ lineage. Proc Natl Acad Sci USA **93:** 11925-11932.

GALANT, R., J. B. SKEATH, S. PADDOCK, D. L. LEWIS and S. B. CARROLL, 1998 Expression pattern of a butterfly *achaete-scute* homolog reveals the homology of butterfly wing scales and insect sensory bristles. Current Biology **8:** 807-813.

GOLDSMITH, M. R., and A. S. WILKINS, 1995 *Molecular Model Systems in the Lepidoptera*. Cambridge University Press, Cambridge.

IGA, M., and G. SMAGGHE, 2010 Identification and expression profile of Halloween genes involved in ecdysteroid biosynthesis in *Spodoptera littoralis*. Peptides **31:** 456-467.

JORON, M., R. PAPA, M. BELTRÁN, N. CHAMBERLAIN, J. MAVÁREZ *et al.*, 2006 A Conserved Supergene Locus Controls Colour Pattern Diversity in *Heliconius* Butterflies. PLoS Biol **4:** e303.

KEYS, D. N., D. L. LEWIS, J. E. SELEGUE, B. J. PEARSON, L. V. GOODRICH *et al.*, 1999 Recruitment of a *hedgehog* regulatory circuit in Butterfly eyespot evolution. Science **283:** 532-534.

KOCH, P. B., P. M. BRAKEFIELD and F. KESBEKE, 1996 Ecdysteroids control eyespot size and wing color pattern in the polyphenic butterfly *Bicyclus anynana* (Lepidoptera: Satyridae). J. Insect Physiol **42:** 223–230.

KOCH, P. B., D. N. KEYS, T. ROCHELEAU, K. ARONSTEIN, M. BLACKBURN *et al.*, 1998 Regulation of dopa decarboxylase expression during color pattern formation in wild-type and melanic tiger swallowtail butterflies. Development **125:** 2303–2313.

KOCH, P. B., U. LORENZ, P. M. BRAKEFIELD and R. H. FFRENCH-CONSTANT, 2000 Butterfly wing pattern mutants: developmental heterochrony and co-ordinately regulated phenotypes. Development Genes and Evolution **210:** 536-544.

KOCH, P. B., R. MERK, R. REINHARDT and P. WEBER, 2003 Localization of ecdysone receptor protein during colour pattern formation in wings of the butterfly *Precis coenia* (Lepidoptera: Nymphalidae) and co-expression with Distal-less protein. Development Genes and Evolution **212:** 571-584.

LONG, A. D., S. L. MULLANEY, L. A. REID, J. D. FRY, C. H. LANGLEY *et al.*, 1995 High Resolution Mapping of Genetic Factors Affecting Abdominal Bristle Number in *Drosophila melanogaster*. Genetics **139:** 1273-1291.

LYYTINEN, A., P. M. BRAKEFIELD, L. LINDSTRÖM and J. MAPPES, 2004 Does predation maintain eyespot plasticity in *Bicyclus anynana*? Proc. R. Soc. Lond. B **271:** 279-283.

MAEDA, S., A. NAKASHIMA, R. YAMADA, N. HARA, Y. FUJIMOTO *et al.*, 2008 Molecular cloning of ecdysone 20-hydroxylase and expression pattern of the enzyme during embryonic development of silkworm *Bombyx mori*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **149:** 507-516.

MANK, J. E., 2009 The W, X, Y and Z of sex-chromosome dosage compensation Trends in Genetics **25:** 226-233.

MCMILLAN, W. O., A. MONTEIRO and D. D. KAPAN, 2002 Development and evolution on the wing. TREE **17**.

MONTEIRO, A., P. BRAKEFIELD and V. FRENCH, 1997 The genetics and development of an eyespot pattern in the butterfly *Bicyclus anynana*: response to selection for eyespot shape. Genetics **146:** 287 - 294.

MONTEIRO, A., B. CHEN, L. C. SCOTT, L. VEDDER, H. J. PRIJS *et al.*, 2007 The combined effect of two mutations that alter serially homologous color pattern elements on the fore and hindwings of a butterfly. BMC Genetics **8:** 22.

MONTEIRO, A., G. GLASER, S. STOCKSLAGER, N. GLANSDORP and D. RAMOS, 2006 Comparative insights into questions of lepidopteran wing pattern homology. BMC Developmental Biology **6:** 52.

NIJHOUT, H. F., 1999 Control mechanisms of polyphenic development in insects. BioScience **49:** 181-192.

NIWA, R., T. MATSUDA, T. YOSHIYAMA, T. NAMIKI, K. MITA *et al.*, 2004 CYP306A1, a cytochrome P450 enzyme, is essential for ecdysteroid biosynthesis in the prothoracic glands of *Bombyx* and *Drosophila*. J. Biol. Chem. **279:** 35942-35949.

NUZHDIN, S. V., E. G. PASYUKOVA, C. L. DILDA, Z.-B. ZENG and T. F. C. MACKAY, 1997 Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster* PNAS **94:** 9734-9739.

OLOFSSON, M., A. VALLIN, S. JAKOBSSON and C. WIKLUND, 2010 Marginal Eyespots on Butterfly Wings Deflect Bird Attacks Under Low Light Intensities with UV Wavelengths. PLoS ONE **5:** e10798.

OOSTRA, V., M. A. DE JONG, B. M. INVERGO, F. M. H. N. KESBEKE, F. WENDE *et al.*, 2010 Translating environmental gradients into discontinuous reaction norms via hormone signalling in a polyphenic butterfly. Proc R Soc B **doi:** 10.1098/rspb.2010.1560.

ORR, H. A., 2005 The genetic theory of adaptation: a brief history. Nat Rev Genet **6:** 119-127.

REED, R. D., and L. M. NAGY, 2005 Evolutionary redeployment of a biosynthetic module: expression of eye pigment genes *vermilion*, *cinnabar*, and *white* in butterfly wing development. Evolution & Development **7:** 301-311.

REED, R. D., and M. S. SERFAS, 2004 Butterfly Wing Pattern Evolution Is Associated with Changes in a Notch/Distal-less Temporal Pattern Formation Process. Current Biology **14:** 1159-1166.

ROFF, D. A., 1997 *Evolutionary quantitative genetics.* Chapman and Hall, New York.

SACCHERI, I. J., H. D. LLOYD, S. J. HELYAR and P. M. BRAKEFIELD, 2005 Inbreeding uncovers fundamental differences in the genetic load affecting male and female fertility in a butterfly. Proc. R. Soc. B **272:** 39-46.

SAENKO, S., P. BRAKEFIELD and P. BELDADE, 2010 Single locus affects embryonic segment polarity and multiple aspects of an adult evolutionary novelty. BMC Biology **8:** 111.

SAENKO, S. V., V. FRENCH, P. M. BRAKEFIELD and P. BELDADE, 2008 Conserved developmental processes and the formation of evolutionary novelties: examples from butterfly wings. Phil. Trans. R. Soc. B **363:** 1549-1556.

SONG, Q., E. S. ALNEMRI, G. LITWACK and L. I. GILBERT, 1997 An Immunophilin is a Component of the Insect Ecdysone Receptor (EcR) Complex. Insect Biochemistry and Molecular Biology **27:** 973-982.

TRAUT, W., K. SAHARA and F. MAREC, 2007 Sex Chromosomes and Sex Determination in Lepidoptera. Sexual Development **1:** 332–346.

VAN'T HOF, A. E., F. MAREC, I. J. SACCHERI, P. M. BRAKEFIELD and B. J. ZWAAN, 2008 Cytogenetic Characterization and AFLP-Based Genetic Linkage Mapping for the Butterfly *Bicyclus anynana*, Covering All 28 Karyotyped Chromosomes. PLoS ONE **3:** e3882.

WANG, J., Q. XIA, X. HE, M. DAI, J. RUAN *et al.*, 2005 SilkDB: a knowledgebase for silkworm biology and genomics. Nucl. Acids Res. **33:** D399-402.

WANG, S., C. J. BASTEN and Z.-B. ZENG, 2010 Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (htto://statgen.ncsu.edu/qtlcart/WQTLCart.htm).

WAYNE, M. L., J. B. HACKETT, C. L. DILDA, S. V. NUZHDIN, E. G. PASYUKOVA *et al.*, 2001 Quantitative trait locus mapping of fitness-related traits in *Drosophila melanogaster*. Genetical Research **77:** 107-116.

WEATHERBEE, S. D., H. F. NIJHOUT, L. W. GRUNERT, G. HALDER, R. GALANT *et al.*, 1999 *Ultrabithorax* function in butterfly wings and the evolution of insect wing patterns. Current Biology **9:** 109-115.

WIJNGAARDEN, P. J., and P. M. BRAKEFIELD, 2000 The genetic basis of eyespot size in the butterfly *Bicyclus anynana*: an analysis of line crosses. Heredity **85:** 471-479.

WIJNGAARDEN, P. J., and P. M. BRAKEFIELD, 2001 Lack of response to artificial selection on the slope of reaction norms for seasonal polyphenism in the butterfly *Bicyclus anynana*. Heredity **87:** 410-420.

WIJNGAARDEN, P. J., P. B. KOCH and P. M. BRAKEFIELD, 2002 Artificial selection on the shape of reaction norms for eyespot size in the butterfly *Bicyclus anynana*: direct and correlated responses. Journal of Evolutionary Biology **15:** 290-300.

WINDIG, J. J., P. M. BRAKEFIELD, N. REITSMA and J. G. M. WILSON, 1994 Seasonal polyphenism in the wild: survey of wing patterns in five species of *Bicyclus* butterflies in Malawi. Ecological Entomology **19:** 285-298.

YASUKOCHI, Y., 1998 A dense genetic map of the silkworm, *Bombyx mori*, covering all chromosomes based on 1018 molecular markers. Genetics **150:** 1513-1525.

ZHENG, W.-W., D.-T. YANG, J.-X. WANG, Q.-S. SONG, L. I. GILBERT *et al.*, 2010 *Hsc70* binds to ultraspiracle resulting in the upregulation of 20-hydroxyecdsone-responsive genes in *Helicoverpa armigera*. Molecular and Cellular Endocrinology **315:** 282-291.

ZHOU, Q., T. ZHANG, W. XU, L. YU, Y. YI *et al.*, 2008 Analysis of four achaete-scute homologs in *Bombyx mori* reveals new viewpoints of the evolution and functions of this gene family. BMC Genetics **9:** 24.

ZIJLSTRA, W. G., M. J. STEIGENGA, P. B. KOCH, B. J. ZWAAN and P. M. BRAKEFIELD, 2004 Butterfly Selected Lines Explore the Hormonal Basis of Interactions between Life Histories and Morphology. American Naturalist **163:** E76-E87.

APPENDIX 6.1

QTL mapping with a full-sib cross, dominant markers and non-recombining females.

QTL mapping is based on the same underlying principles as linkage mapping, namely the disruption of co-segregation by recombination between different loci on the same linkage group (LG). The proportion of offspring that recombined between loci is a measure of the mapping distance between these loci. There are also major differences between linkage mapping and QTL mapping though, especially in Lepidoptera where the females do not recombine during meiosis. For linkage mapping, loci involved are discrete genetic markers while QTL mapping is based on a combination of discrete genetic markers and partial phenotypic contributions of linked QTLs, whose effects can be counteracted by other (independent) loci. This gives far weaker associations of marker-QTL co-segregation than for genetic markers alone. With linkage mapping in Lepidoptera, the (non-recombinant) maternally inherited alleles must be removed from analysis because they prevent reliable mapping (JIGGINS *et al.* 2005; KAPAN *et al.* 2006). This introduction of errors by maternally inherited alleles is also true for QTL mapping, but linked QTLs on the maternal chromosomes do in fact contribute to the phenotype. Removing a marker allele, but leaving its (fully linked because non-recombinant) phenotypic effect intact disrupts the marker-phenotype association.

The full-sib cross design was chosen to generate the maximal phenotypic range in the offspring, and to obtain the highest possible association between genetic markers and QTL. The choice for such a cross has downstream consequences though. In the following explanation, 'H' is used to indicate marker alleles from the High line and 'L' to represent those from the Low line, and 'Q' and 'q' are used for QTL alleles causing an increase and decrease in eyespot size respectively. In a HL×HL/Qq×Qq F_1 cross, the marker alleles are in full association with the QTL alleles and a marker that is fully linked to a QTL will give HH/QQ, HL/Qq, LH/qQ, and LL/qq offspring. The analysis of such combinations is very straightforward in general, especially when the QTL is additive (i.e. QQ = large, Qq intermediate, qq = small). It is however not possible to make full use of this association with the current experimental setup due to the dominant nature of the markers used and the absence of recombination in female Lepidoptera. When using dominant markers (such as AFLPs), the marker-QTL association in the offspring becomes H/QQ, H/Qq, H/qQ, L/qq if the H-marker allele is dominant, which has a far lower resolution than with co-dominant markers. The L allele could also be dominant and QTL themselves can potentially be dominant in either direction, which can result in combinations that are hardly informative.

Introducing the non-recombinant characteristic of females adds another level of complexity. The ideal markers for Lepidoptera are markers that have only a paternally inherited segregation pattern. Such markers are male informative (MI) markers and censored co-dominant both informative (BI) markers from which the maternal component has been removed. This removal of the maternal component is based on the fixed segregation patterns of the non-recombinant maternal chromosomes which facilitates unambiguous prediction of the alleles that the mother transferred to each individual offspring. This segregation matrix is named the chromosome print (YASUKOCHI 1998) and it simply defines which one of the two maternal chromosomes each offspring received for each chromosome pair. The phenotypic range represents an F_2 cross though and the phenotypic maternal component cannot be censored in a

similar way as for the genetic markers. This has a two-fold undesirable effect: (*i*) the QTL on the maternal chromosome cannot be pinpointed because its phenotypic effect is fully linked over the entire length of the chromosome and (*ii*) a paternal marker backcross pattern with 1:1 segregation, representing only half the phenotypic range could be wrongfully compared with a fully diverse F_2 eyespot size range with 1:2:1 or 3:1 segregation. This means that when using Lepidoptera, the best strategy would be to start off with two backcrosses (in both phenotypic directions) using heterozygous males and homozygous females to cover the full phenotypic range for QTL analysis. Since the current data are not based on backcrosses, they need a specific approach to extract QTL mapping information.

There are further restrictions to QTL mapping as a result of the current experimental setup. Composite interval mapping and multiple interval mapping that are used to detect genetic interactions between the QTLs cannot be performed because they need to compare the genotypic and phenotypic values from all chromosomes at once. Each chromosome has a unique set of ~50% individuals that are excluded by the censoring of dominant BI markers. This means that the number of individuals that can be compared decrease exponentially per chromosome. (e.g. two chromosomes have only 25% overlap in genotyped individuals, and 28 chromosomes have only 2^{-28} (~ 0) fully overlapping individuals).

APPENDIX 6.1 REFERENCES

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

KAPAN, D. D., N. S. FLANAGAN, A. TOBLER, R. PAPA, R. D. REED *et al.*, 2006 Localization of Müllerian mimicry genes on a dense linkage map of *Heliconius erato*. Genetics **173:** 735-757.

YASUKOCHI, Y., 1998 A dense genetic map of the silkworm, *Bombyx mori*, covering all chromosomes based on 1018 molecular markers. Genetics **150:** 1513-1525.

APPENDIX 6.2

Genbank accession numbers of sequences used to locate candidate genes that are unmapped in *Bicyclus anynana* in the SilkDB silkworm genome assembly by means of a blast search. *Bombyx mori* coding sequences were used as query input when available and otherwise the *Drosophila melanogaster* (*Dmel* in table) orthologs were used. Blastx values are given for genes with proper coding sequence predictions, tblastx results are used for genes with prediction errors.

Gene name	Gene symbol	blast	sequence origin	genbank acc #	SilkDB scaffold	e-value
achaete-scute homolog 1	*ASH1*	blastx	*B. mori*	DQ350890	nscaf1898:6056628-6057209	e-109
achaete-scute homolog 2	*ASH2*	blastx	*B. mori*	NM_001105222	nscaf1898:6175916-6176635	e-135
achaete-scute homolog 3	*ASH3*	blastx	*B. mori*	NM_001105224	nscaf1898:6222871-6223434	e-104
asense	*ase*	blastx	*B. mori*	NM_001105226	nscaf1898:5940257-5941471	0
cinnabar	*cn*	blastx	*B. mori*	EF210328	nscaf2860:3025665-3028983	e-143
decapentaplegic	*dpp*	blastx	*B. mori*	NM_001145329	nscaf2993:6692859-6696343	0
disembodied	*dib*	blastx	*B. mori*	NM_001043488	nscaf1681:277196-285831	0
Distal-less	*Dll*	tblastx	*Dmel*	NM_079133	nscaf2964:2055452-2184269	7e-32
groucho	*gro*	tblastx	*B. mori*	NM_001134889	nscaf2136:5066518-5124531	0
Heat-shock protein 90	*Hsp90*	blastx	*B. mori*	NM_001043411	nscaf2801:616538-618688	0
hedgehog	*hh*	blastx	*Dmel*	L05404	nscaf3048:978705-989746	1e-63
Medea	*Med*	blastx	*Dmel*	NM_079871	nscaf2986:4943147-4952874	e-107
Mothers against dpp	*Mad*	blastx	*Dmel*	U10328	nscaf2993:355075-361490	0
Notch	*N*	blastx	*B. mori*	NM_001163898	nscaf2888:7526277-7586024	0
numb	*numb*	blastx	*Dmel*	NM_078799	nscaf2993:7375661-7384165	e-101
patched	*ptc*	tblastx	*Dmel*	NM_078937	nscaf2674:2648112-2656594	0
phantom	*phm*	blastx	*B. mori*	AB124840	nscaf2860:608241-614972	0
ruby	*rb*	blastx	*Dmel*	AF247194	nscaf2868:530839-555239	e-126
scabrous	*sca*	tblastx	*Dmel*	M60065	nscaf1690:7881371-7884088	1e-72
scalloped	*sd*	blastx	*Dmel*	NM_001169291	nscaf1898:4585777-4618439	3e-59
shade	*shd*	blastx	*B. mori*	AB236417	nscaf2986:2935439-2939078	e-101
shadow	*sad*	blastx	*B. mori*	AB167737	nscaf2828:4597141-4613087	e-179
Smad on X	*Smox*	blastx	*Dmel*	NM_078524	nscaf2964:3722236-3736526	0
spalt-related	*salr*	blastx	*Dmel*	NM_078824	nscaf2589:6654523-6680893	4e-94
split ends	*spen*	blastx	*Dmel*	NM_164374	nscaf2930:3446818-3468207	7e-85
spook	*spo*	blastx	*B. mori*	AB124841	nscaf2176:1260891-1266149	0
Ultrabithorax	*Ubx*	tblastx	*B. mori*	AB505052	nscaf2853:3215464-3215931	1e-74
ultraspiracle	*usp*	tblastx	*B. mori*	AB182582	nscaf2847:6419185-6441959	e-132
vermilion	*v*	blastx	*Dmel*	NM_078558	nscaf2943:1741136-1757256	e-142
white	*w*	blastx	*B. mori*	NM_001043569	nscaf2575:3237764-3274876	0
wingless	*wg*	blastx	*B. mori*	NM_001043850	nscaf2847:4290086-4300264	0

Nederlandse samenvatting

(Dutch summary)

Een bijdrage aan de kennis over de genoom architectuur van *Bicyclus anynana*

Het promotieonderzoek dat in dit proefschrift beschreven wordt heeft als belangrijkste doel het ontrafelen van de genetische factoren die van invloed zijn op het verschil in oogvlekgrootte op de ondervleugels van de Afrikaanse vlindersoort *Bicyclus anynana*. Evolutionair biologen zijn geïnteresseerd in deze eigenschap omdat het centraal staat in een systeem waar ecologie, evolutie en ontwikkelingsgenetica nauw met elkaar verbonden zijn. Daardoor kunnen de afzonderlijke componenten die bijdragen aan de verschillende fenotypen in een bredere context geplaatst worden, wat leidt tot een beter begrip van complexe evolutionaire processen.

Volledig ontwikkelde oogvlekken op de vleugels van *B. anynana* bestaan uit een witte kern omringd door een zwarte en een gouden cirkel. De grootte en samenstelling van de oogvlekken verschillen per jaargetijde en zijn klein en onopvallend in het droge seizoen en groot en opvallend in het natte seizoen. Naast de oogvlekken hebben de rest van de vleugels en het gedrag van de vlinders ook een seizoensgebonden patroon. De vlinders zijn passief en vrij egaal bruin in het droge seizoen waardoor ze nauwelijks opvallen tussen dorre bladeren. Dit biedt ze bescherming tegen visueel ingestelde predatoren. In het natte seizoen zijn ze actief en hebben ze een reeks opvallende oogvlekken langs de rand van de ondervleugels. Het begrip 'ondervleugels' is een anatomische benaming, en betekent niet automatisch dat dit vleugeloppervlak aan het zicht onttrokken wordt (zoals bij een dagpauwoog het geval zou zijn). Bij *B. anynana* zijn de ondervleugels juist de meest zichtbare zijde omdat deze soort met de vleugels gevouwen boven het lichaam rust. Vogels die op zicht foerageren hebben de neiging om in plaats van het lichaam de oogvlekken op de vleugels met hun snavel te grijpen. Daardoor kunnen de vlinders ontsnappen met als enig gevolg dat de vleugel licht beschadigd raakt. Kort samengevat geeft het seizoensgebonden specifieke uiterlijk van de vlinders dus een schutkleur tijdens het droge seizoen en een ontsnappingsstrategie tijdens het natte seizoen.

Eén van de fascinerende aspecten van deze seizoensgebonden verschillen in uiterlijk is dat het in een natuurlijke situatie niet primair door genetische variatie wordt bepaald, maar door fenotypische plasticiteit. In dit geval wordt doorgaans de term 'seizoens-polyfenisme' gebruikt vanwege de twee afzonderlijke fenotypen gerelateerd aan de twee verschillende seizoenen. Het uiterlijk van de adult wordt beïnvloed door omgevingstemperatuur tijdens het laatste larvale (rups) stadium. Een lage temperatuur leidt tot het droge seizoen fenotype met kleine oogvlekken en een hoge temperatuur leidt tot het natte seizoen fenotype met grote oogvlekken. Laboratorium experimenten hebben aangetoond dat intermediaire temperaturen oogvlekken met tussenliggende grootte veroorzaken, maar in de natuurlijke omgeving (met discrete temperatuurverschillen tussen de seizoenen) komen vooral de twee extremen voor. Het is tot dusver onbekend hoe de temperatuur direct de biochemische en/of genetische reacties beïnvloedt, maar iets later in het proces van oogvlekontwikkeling is er een zeer sterke correlatie gevonden tussen het moment waarop de concentratie van het ecdysone hormoon zijn hoogste waarde bereikt en de oogvlekgrootte.

Genetische variatie vormt vaak de basis voor het opsporen van genen die betrokken zijn bij de vorming van verschillen in fenotypen. Hierboven is echter uitgelegd dat temperatuur in plaats van genetische variatie de seizoensverschillen bepaalt. In werkelijkheid is er wel degelijk een effect van genetische variatie op oogvlekgrootte, maar dat is onder natuurlijke omstandigheden moeilijk waar te nemen. Een natuurlijke populatie heeft namelijk een min of meer willekeurige genetische samenstelling, waarin antagonistische allelen van verschillende genen elkaar's effect afzwakken met als resultaat dat slechts het temperatuureffect aantoonbaar verantwoordelijk is voor het fenotype. Een dergelijk samenspel van verschillende genen met verschillend effect en intensiteit en de omgeving resulteert in kwantitatieve eigenschappen met een spectrum van fenotypische variatie (in tegenstelling tot discrete eigenschappen zoals geslacht of oogkleur in de mens). Deze kwantitatieve eigenschappen worden in het Engels 'quantitative traits' genoemd en de afzonderlijke betrokken genen (gebaseerd op hun positie/locus in het genoom) worden aangeduid met 'quantititive trait loci' of QTLs.

De genetische effecten van QTLs kunnen aan het licht gebracht worden door kunstmatige selectie toe te passen. Hierdoor accumuleren allelen van verschillende genen met een zelfde effect over een aantal generaties in kunstmatige populaties. Voor *B. anynana* zijn zulke selectielijnen geproduceerd voor grote en kleine oogvlekken met als resultaat dat de genetische verschillen die zijn opgebouwd tussen de lijnen verschillende fenotypen produceren bij dezelfde omgevingstemperatuur. De selectielijnen vormen de basis om te onderzoeken welke genen verantwoordelijk zijn voor de genetisch gestuurde fenotypische variatie.

Het is echter niet mogelijk om simpelweg de genetische verschillen tussen de twee lijnen in kaart te brengen en te concluderen dat deze ook de fenotypische verschillen veroorzaken. Naast de effectieve (oogvlek-fenotype sturende) genetische verschillen is er namelijk ook een veelvoud aan genetische variatie die niet gerelateerd is aan het fenotype. Daarnaast is het met de huidige stand van techniek bijzonder moeilijk om alle genetische verschillen goed te karakteriseren. De genetische variatie die verantwoordelijk is voor de verschillen in oogvlekgrootte kunnen opgespoord worden door de selectielijnen te kruisen over twee generaties en de genetische variatie met intervallen op de chromosomen te bemonsteren. Wanneer een bepaalde sectie van een chromosoom een vergelijkbaar overervingpatroon vertoont als het fenotype dan is dat een goede indicatie dat deze chromosoom regio een oogvlekgrootte beïnvloedend gen (QTL) bevat. Deze identificatie van een chromosomale regio die gerelateerd is aan een fenotype is een eerste stap naar het vinden van het verantwoordelijke gen zelf.

Het isoleren van een gen en het bevestigen van de betrokkenheid bij de totstandkoming van een bepaald fenotype vereist vervolgonderzoek, maar desondanks kan deze ruwe positionering van een gen in combinatie met overervingpatronen van chromosoom secties en fenotypen zeer belangrijke informatie onthullen. Aan de hand van een chromosomale regio die een oogvlekgrootte regulerend gen bevat kan het effect en intensiteit van de allelen van een gen bepaald worden en de interacties met andere genen. Dit betekent dat, in een ideaal geval, veel genetische processen al gekarakteriseerd kunnen worden voordat de afzonderlijke genen geïdentificeerd zijn. Door intensief onderzoek aan *B. anynana* zijn de genetische processen die betrokken zijn bij de vorming van oogvlekken voor een deel bekend. Voor een aantal genen is overtuigend aangetoond dat ze betrokken zijn bij oogvlek ontwikkeling. Dat betekent echter niet automatisch dat deze genen ook betrokken zijn bij de variatie in oogvlekgrootte. In hoofdstuk 6 wordt onderzocht of deze bekende oogvlek genen ook voor variatie zorgen en of er nog andere genen verantwoordelijk zijn voor de

verschillende seizoensvormen. De hierboven genoemde vragen konden echter niet behandeld worden voordat de genoom architectuur van *B. anynana* tot op zekere hoogte in kaart was gebracht. De hoofdstukken voorafgaand aan hoofdstuk 6 beschrijven elk een onderdeel van de karakterisering van het genoom op verschillende niveaus.

Microsatellieten

Hoofdstuk 2 beschrijft de ontwikkeling van microsatellieten voor *B. anynana* als genetische merkers. Genetische merkers zijn DNA fragmenten waarvan verschillende varianten (polymorfismen) bestaan. Genetische merkers worden voor vele doeleinden gebruikt in verschillende wetenschappelijke disciplines, maar in dit geval worden ze gebruikt om overervingpatronen te traceren. Microsatellieten zijn korte stukjes DNA met een repeterend motief van hooguit zes basenparen per éénheid, zoals een di-nucleotide herhalingsstructuur met elf éénheden: CACACACACACACACACACACA (= CA_{11}). Er zijn een aantal redenen waarom microsatellieten veel gebruikt worden in biologische studies. Ten eerste bevatten ze veel variatie in het aantal herhalingséénheden (genetische variatie is één van de belangrijkste eigenschappen, zoniet dé belangrijkste eigenschap in biologisch onderzoek). Een individu kan bijvoorbeeld op het ene chromosoom een CA_{11} hebben terwijl zich op het andere chromosoom op dezelfde plek een CA_{15} bevindt. Een tweede groot voordeel is dat microsatellieten co-dominant zijn, wat betekent dat heterozygoten van homozygoten onderscheiden kunnen worden. Verder worden vaak als voordelen genoemd dat hun flankerende sequenties sterk geconserveerd zijn en dat deze flankerende sequenties slechts eenmaal per (haploïd) genoom voorkomen. Deze laatste twee eigenschappen zijn echter in geen geval op alle microsatellieten van toepassing, wat later uitgelegd zal worden.

Het ontwikkelen van microsatellieten kan een zeer moeilijke onderneming zijn, vooral wanneer een groot aantal benodigd is. Om die reden werd het eerste deel van de procedure voor *B. anynana* uitbesteed aan een gespecialiseerd biotechnologisch bedrijf. De vervolgstappen omvatten doorgaans beproefde procedures om bruikbare microsatellieten te oogsten. Dat was hier echter niet het geval omdat de flankerende sequenties van de microsatellieten vele malen voorkwamen in verschillende combinaties. Dit betekent dat er vele kopieën van deze microsatellieten in het genoom aanwezig zijn, wat ze onbruikbaar maakt om een unieke positie in het genoom te karakteriseren. Dit verschijnsel is in strijd met de bewering dat flankerende sequenties slechts eenmaal per genoom voorkomen. De microsatellieten die wel uniek waren hadden vaak grote variatie in hun flankerende sequenties, wat niet in overeenstemming is met de bewering dat flankerende sequenties sterk geconserveerd zijn. Door soortgelijk onderzoek in andere soorten is het duidelijk geworden dat deze ongewenste eigenschappen van microsatellieten eerder regel dan uitzondering vormen voor Lepidoptera (de taxonomische groep vlinders en nachtvlinders). Om toch een acceptabel aantal bruikbare microsatelliet merkers te bemachtigen is een speciale techniek ontwikkeld om de ongewenste sequenties te vermijden. Dit leverde een, voor vlinders, ongekend hooge opbrengst van 28 polymorfe microsatelliet merkers op, wat lange tijd met afstand het grootse aantal in vlinders is geweest.

Replicatie van genoom fragmenten

Hoofdstuk 3 beschrijft hoe een analyse van de microsatellieten en hun flankerende sequenties gebruikt kon worden om de dynamiek van het genoom beter te begrijpen. Genomen zijn constant aan veranderingen onderhevig en een deel van deze veranderingen hebben duidelijke gevolgen, zoals mutaties die de aminozuursamenstelling van een eiwit veranderen. Andere veranderingen, zoals duplicatie van genoom fragmenten, en de consequenties daarvan worden minder goed begrepen en karakterisering van deze dynamiek draagt daardoor bij aan de algemene kennis van genoom evolutie. De microsatellieten van *B. anynana* dragen sporen van de vermenigvuldigingen die tot meerdere kopieën hebben geleidt. Hierdoor kunnen deze gebeurtenissen tot op zekere hoogte gereconstrueerd worden. Er zijn twee algemene processen die tot multiplicatie van genoom segmenten kunnen leiden. Het eerste proces is ongelijke recombinatie, waarbij niet-overeenkomstige delen van chromosomen hergroeperen, wat in sommige gevallen tot duplicatie leidt. De tweede categorie is een verzameling van processen waar stukken DNA, aangeduid als 'transposable elements' verplaatst worden. Bij transpositie worden de sequenties vaak eerst gedupliceerd voordat ze verplaatst worden waardoor er meerdere kopieën ontstaan (replicatieve transpositie). De herkomst van transposable elements is divers, soms komen ze uit het genoom zelf, maar ze kunnen ook een andere oorsprong hebben, zoals een virus.

Ongelijke recombinatie resulteert doorgaans in kopieën die dicht bij elkaar op het chromosoom liggen. Dit wordt inderdaad waargenomen, maar overervingpatronen geven ook aan dat een deel van de gedupliceerde fragmenten op verschillende chromosomen liggen, wat wijst op transposable elements. De sequenties in B. anynana passen daarom het best in een model waar zowel ongelijke recombinatie en transpositie verantwoordelijk zijn voor duplicatie.

De gemultipliceerde fragmenten in *B. anynana* vertonen sterke gelijkenis met sequenties in andere Lepidoptera soorten. Bovendien zijn deze overeenkomstige sequenties in de andere soorten eveneens bijna altijd direct gekoppeld aan een microsatelliet. Voor de meeste van deze Lepidoptera specifieke sequenties is het niet duidelijk wat hun gemeenschappelijke herkomst en functie is en waarom ze met microsatellieten geassocieerd zijn. Van één van de sequenties is de herkomst echter wel duidelijk. Het is afkomstig van een symbiotisch virus dat zich in een sluipwesp genoom heeft ingebouwd. De sluipwesp legt haar eieren in rupsen van vlinders en het virus onderdrukt het immuunsysteem van de rupsen, waardoor de sluipwesplarven zich kunnen ontwikkelen ten koste van de rups. Het virus DNA is waarschijnlijk opgenomen in het genoom van een rups die heeft weten te overleven, waarna het is doorgegeven aan volgende generaties.

Cytogenetica

Hoofdstuk 4 omvat een cytogenetische karakterisering van het *Bicyclus anynana* genoom. Het primaire doel was het aantal chromosomen per haploïde cel (n) in onze laboratoriumpopulatie van *Bicyclus anynana* te bepalen. Dit was nodig omdat het chromosoom aantal in Lepidoptera extreem variabel is. De totale spreiding tussen soorten varieert van n = 7 tot meer dan 200 en zelfs variatie in chromosoom aantallen tussen verschillende populaties van dezelfde soort is niet ongebruikelijk. Een chromosoom aantal van n = 28 werd geteld in onze laboratoriumpopulatie, wat in de buurt ligt van het meest voorkomende aantal in Lepidoptera (n = 31).

Naast het aantal werden enkele algemene chromosoom eigenschappen onderzocht. De chromosomen van Lepidoptera hebben een aantal eigenaardige kenmerken. Vrouwtjes zijn in Lepidoptera het heterogametische geslacht, wat betekent dat zij twee verschillende geslachtschromosomen hebben (vergelijkbaar met XY in mannen bij de mens). Omdat het een afwijkend systeem is worden deze chromosomen niet aangeduid met XY, maar met WZ (Lepidoptera mannetjes zijn ZZ). Meest bekend bij een breed publiek zijn chromosomen die tijdens de celdeling in langwerpige paren van chromatiden bijeen gehouden worden op een bepaald punt, genaamd de centromeer. Daardoor ontstaan X-vormige chromosoom paren in het geval dat de centromeer nabij het midden ligt. De positie van de centromeer is echter variabel, wat gebruikt kan worden om de afzonderlijke chromosomen te karakteriseren. Bovendien zijn er kleuringtechnieken die vaak een chromosoomspecifiek banderingspatroon opleveren. Deze determinatie kenmerken kunnen niet toegepast worden in Lepidoptera omdat ze een holokinetische binding tussen de chromatiden hebben, waardoor ze over hun gehele lengte met elkaar gepaard zijn (er is dus geen centromeer aanwezig). Tijdens de mitose zijn de chromosomen niet langgerekt, maar sterk gecomprimeerd, waardoor geen banderingspatronen gevisualiseerd kunnen worden. Het enige stadium waarin Lepidoptera chromosomen tot op een zeker niveau gekarakteriseerd kunnen worden is tijdens de pachyteen fase (onderdeel van meiose 1, in de profase 1). De chromosomen zijn dan enigszins langgerekt, maar zijn nog steeds niet afzonderlijk herkenbaar en kunnen hooguit gegroepeerd worden in klein, gemiddeld en groot. Daarom worden in Lepidoptera vaak fluorescerende probes gebruikt om bepaalde chromosomen te identificeren (in dit proefschrift niet gebruikt). De enige kenmerken die hier gebruikt konden worden voor identificatie waren (*i*) kernlichaampjes (nucleoli) die zichtbaar zijn als uitstulpingen veroorzaakt door genexpressie van ribosoomaal RNA, en (*ii*) de geslachtschromosoom bivalent in vrouwtjes die zich van de rest onderscheid doordat de gepaarde chromosomen niet identiek zijn. Er waren twee kernlichaampjes aanwezig, die geassocieerd waren met een autosoom en één van de geslachtschromosomen. Het W-chromosoom bleek uitermate klein te zijn, waardoor de geslachtschromosomen in vrouwtjes op een opmerkelijke manier paarden (zie figuur 4.1). Volledig afwezige W-chromosomen zijn bekend in sommige soorten Lepidoptera (mannetjes ZZ, vrouwtjes Z), maar een dergelijk sterk gereduceerd W-chromosoom was nog niet eerder waargenomen. Het geeft dus aan dat de afwezigheid van W-chromosomen in sommige Lepidoptera waarschijnlijk het resultaat is van een geleidelijke afbraak in plaats van een plotseling verlies van het W-chromosoom.

Linkage map

Hoofdstuk 5 behandelt de constructie van een genetische linkage map. Een genetische linkage map is een weergave of kaart van een genoom gebaseerd op de overerving van genetische merkers. Genetische merkers die niet op het zelfde chromosoom liggen worden tijdens de meiose willekeurig verdeeld en kunnen dus in alle mogelijke combinaties voorkomen in de nakomelingen. Genetische merkers die op het zelfde chromosoom liggen zijn echter niet onafhankelijk van elkaar en worden gezamenlijk doorgegeven aan de nakomelingen. Hierdoor kunnen merkers gegroepeerd worden per chromosoom. Deze koppeling van merkers op een zelfde chromosoom kan echter verbroken worden door recombinatie en de mate waarin dat gebeurt is evenredig met de afstand tussen de merkers. Aan de hand van overerving en recombinatie kan dus bepaald worden (*i*) op welk chromosoom een merker ligt, (*ii*)

wat de onderlinge afstand tussen de merkers op het chromosoom is en daaruit voortkomend (*iii*) wat de volgorde van de merkers is. Het resultaat is een kaart met voor elk chromosoom een lijn met daarop de positie van de merkers die zich op dat chromosoom bevinden.

Een belangrijke toepassing van een linkage map is het positioneren van genetische polymorfismen die fenotypische variatie veroorzaken. De fenotypen zullen, net als de genetische merkers, een overervingpatroon hebben die afhankelijk is van de chromosoom positie waarop het zich bevindt. Daardoor kan de genetische basis voor een bepaald fenotype gepositioneerd worden ten opzichte van de genetische merkers. Naast de eerder beschreven microsatellieten zijn ook AFLP's als genetische merkers gebruikt. Deze merkers hebben als voordeel dat ze met een universele techniek gegenereerd kunnen worden en meerdere merkers produceren per experiment. Het is dus een uitstekende manier om snel veel merkers beschikbaar te krijgen. De nadelen ten opzichte van microsatellieten zijn dat ze dominant zijn waardoor heterozygoten niet onderscheiden kunnen worden van één van de twee homozygoten en dat de sequenties van de stukjes DNA die de AFLP's vertegenwoordigen onbekend blijven waardoor ze in tegenstelling tot microsatellieten 'anoniem' zijn.

De kruising die gebruikt werd om de linkage map te genereren werd opgestart met een P-generatie bestaande uit een grootmoeder uit de selectielijn voor grote oogvlekken en een grootvader uit de selectielijn voor kleine oogvlekken. Deze produceerde hybride nakomelingen waarmee een broer-zus F1 kruising werd opgezet, die vervolgens een F2 generatie produceerde. Een dergelijk kruisingsschema heeft als voordeel dat de F2 generatie een maximale fenotypische spreiding heeft, terwijl de genetische achtergrondruis drastisch verminderd word. Dominante merkers (AFLP's) bleken echter bij nader inzien slechts deels geschikt om een linkage map van een dergelijke kruising te construeren omdat bepaalde overervingpatronen niet meer te detecteren zijn. Dit probleem werd bovendien versterkt door de afwezigheid van recombinatie in vrouwelijke Lepidoptera. Ondanks deze obstakels kon er met een speciaal voor deze kruising ontwikkelde methode een linkage map geconstrueerd worden die alle 28 chromosomen omvat.

QTL analyse

De analyse gepresenteerd in **Hoofdstuk 6** identificeert een aantal chromosoom regionen die waarschijnlijk genen bevatten die oogvlekgrootte beïnvloeden. Om de genetische processen die oogvlekgrootte bepalen te begrijpen is het belangrijk om te weten welke genen er bij betrokken zijn, waar deze genen zich bevinden, wat hun effect is (toename of afname in grootte) en wat hun interactie is met, of hun effect op, andere genen. Er is al veel bekend over genetische processen die betrokken zijn bij de vorming van oogvlekken en ook op het gebied van variatie is vooruitgang geboekt. Deze bevindingen zijn echter vooral gebaseerd op kennis die vergaard is in de fruitvlieg (*Drosophila melanogaster*). Wanneer een bepaald gen in de fruitvlieg een functie heeft die vergelijkbaar is met (of gerelateerd aan) het proces in een ander insect bestaat er een redelijke kans dat hetzelfde gen verantwoordelijk is. Daarom wordt een dergelijk gen een kandidaat-gen genoemd. Het onderzoek naar de functie van deze kandidaat-genen in *B. anynana* heeft aangetoond dat een aantal inderdaad betrokken zijn bij oogvlek vorming. Deze aanpak maakt het ontrafelen van oogvlekvorming en de variatie echter sterk afhankelijk van wat bekend is in de fruitvlieg en van de aanname dat fruitvlieg genen dezelfde functie hebben in andere insect soorten.

Een QTL analyse kan onafhankelijk van voorkennis of aannames regionen identificeren waarin genen liggen die verantwoordelijk zijn voor bepaalde fenotypen, waardoor een completer beeld van alle betrokken processen verkregen wordt. Vervolgens kan onderzocht worden of er zich inderdaad kandidaat-genen in deze regio's bevinden of dat er onbekende, of onverwachte genen betrokken zijn. Omdat is aangetoond dat de concentratie van het hormoon ecdysone sterke invloed heeft op oogvlekgrootte zijn de genen die betrokken zijn bij de productie van ecdysone en genen die er door geactiveerd of aangestuurd worden de belangrijkste kandidaat-genen.

In de voorgaande sectie werd uitgelegd hoe overeenkomsten tussen overerving van merkers en fenotypen de positie van een gen verantwoordelijk voor dat fenotype kunnen lokaliseren. Deze positionering is het eenvoudigst wanneer een fenotype door een enkel gen wordt bepaald, maar ook bij eigenschappen, waar meerdere genen betrokken zijn (QTLs) kunnen statistische associaties tussen overerving van merkers en fenotypen de locaties van deze genen aan het licht brengen. De associatie is alleen minder sterk omdat er per individu een mix van genen zijn die elkaar versterken en afzwakken. Bovendien hebben de betrokken genen verschillende intensiteit van effect, verschillende mate van dominantie en kunnen sommige genen het effect van anderen zelfs teniet doen. Daardoor zijn de intervallen waarin een gen op het chromosoom geplaatst wordt groter dan wanneer een enkel gen een fenotype veroorzaakt (de resolutie van een QTL is dus lager). Bovendien wordt de aanwezigheid van een 'bijdragend' gen (QTL) gedefinieerd met een statistische waarschijnlijkheid, waardoor een dergelijk gen in sommige gevallen over het hoofd gezien kan worden.

De generaties in de selectielijn-kruising produceerden de verwachte fenotypische verdelingen met intermediaire oogvlekken in de F1 en een maximale spreiding van klein tot groot in de F2 (fig. 6.2). Mannetjes en vrouwtjes hebben beiden variabele oogvlekken, maar er is desondanks sterke seksuele dimorfie voor dit kenmerk. Daarom werden beide geslachten afzonderlijk geanalyseerd. Er waren sterke verschillen in QTL-spreiding tussen de geslachten, hetgeen niet ongebruikelijk is. Bij de vrouwtjes waren de signalen het sterkst. Er bevindt zich een QTL op het Z (geslacht) chromosoom die mogelijk overeen komt met één van de kandidaat genen *Henna*, *Scabrous*, of *catalase*. Een ander QTL valt mogelijk samen met *ecdysone receptor*, wat relevant is omdat het gerelateerd is aan het hormoon wat oogvlekgrootte beïnvloedt. Verder zijn er nog QTLs waarbij het onduidelijk is of ze overeen komen met de positie van een kandidaat-gen, en bovendien zijn er QTLs die op chromosomen liggen waarop zich geen kandidaat-genen bevinden. Dit geeft aan dat er hooguit een gedeeltelijke overlap is tussen wat verwacht werd en wat experimenteel is aangetoond. Dat betekent dus ook dat wat we momenteel weten over oogvlekvorming, en met name de variatie in grootte, slechts voor een deel ontrafeld is. De QTL analyse die hier beschreven wordt geeft geen definitief antwoord op de openstaande vragen betreffende oogvlekgrootte variatie, maar het heeft wel een aantal chromosoom regionen geïdentificeerd die vervolgens tot in detail onderzocht kunnen worden op hun betrokkenheid bij het reguleren van seizoens-polyfenisme. Daardoor brengt dit onderzoek ons een stap dichter bij het begrijpen van de processen die oogvlekgroottes en hun evolutie bepalen.

CURRICULUM VITAE

Arjèn Emiel van 't Hof 11 April 1973, in Numansdorp, The Netherlands

PRE-UNIVERSITY EDUCATION:
1985-1990 HAVO Lambert Franckens College, Elburg, The Netherlands
1990-1992 VWO Lambert Franckens College, Elburg, The Netherlands

ACADEMIC EDUCATION
1992-2000 MSc in Biology, primary subject: Ecology
University of Groningen, Department of Biology, The Netherlands

Masters thesis projects:

'Gene Flow in *Drosophila melanogaster* metapopulations'
University of Groningen, Evolutionary Genetics Group, The Netherlands

'Hatching strategies of the neotropical Giant Damselfly *Megaloprepus coerulatus*'
Smithsonian Tropical Research Institute, Barra Colorado Island, Panama

PHD RESEARCH
2000-2005[1] 'Towards understanding the architecture of the *Bicyclus anynana* genome'
University of Leiden, Institute of Biology Leiden, Evolutionary Biology, Leiden, The Netherlands

POST-PHD RESEARCH
2006-2009 Post-doctoral research assistant 'Adaptation to anthropogenic environmental change: the molecular genetic basis of industrial melanism in peppered moths'
University of Liverpool, Institute of Integrative Biology, Evolution, Ecology and Behaviour, Liverpool, UK

2010 Post-doctoral research assistant 'GRAPPLE – Iterative modeling of gene regulatory interactions underlying stress, disease and ageing in *C. elegans*'
University of Liverpool, Institute of Integrative Biology, Functional and Comparative Genomics, Liverpool, UK

2010-2013 Post-doctoral research assistant 'Effect of rapid environmental change on genetic diversity through space and time: selective sweeps and industrial melanism in peppered moths'
University of Liverpool, Institute of Integrative Biology, Evolution, Ecology and Behaviour, Liverpool, UK

Current work address:
University of Liverpool, Institute of Integrative Biology, Evolution, Ecology and Behaviour, Biosciences Building, Crown Street, Liverpool L69 7ZB, United Kingdom

[1] not including final analysis and thesis preparation

PUBLICATIONS

VAN'T HOF, A. E., P. BELDADE, F. KESBEKE, H. DE VOS, P. M. BRAKEFIELD and B. J. ZWAAN, 2011 Quantitative Trait Loci affecting the size of wing eyespots in a seasonally polyphenic butterfly. (in prep)

VAN'T HOF, A. E., N. EDMONDS, F. MAREC, M. DALÍKOVÁ and I. J. SACCHERI, 2011 Industrial Melanism in British Peppered Moths has a Singular and Recent Mutational Origin. Science (conditionally accepted)

VAN'T HOF, A. E., N. EDMONDS, F. MAREC, M. DALÍKOVÁ, P. NGUYEN and I. J. SACCHERI, 2011 Linkage mapping in the peppered moth positions melanisation genes, reconstructs chromosome evolution events, and reveals shared synteny of sex-linked genes. (in prep)

VAN'T HOF, A. E. and I. J. SACCHERI, 2010 Industrial Melanism in the Peppered Moth Is Not Associated with Genetic Variation in Canonical Melanisation Gene Candidates. PLoS ONE **5**: e10889.

BELDADE, P., A. E. VAN'T HOF, M. A. JERÓNIMO and A. D. LONG, 2009 Microsatellite markers associated with genes expressed in developing wings of *Bicyclus anynana* butterflies. Molecular Ecology Resources **9:** 1487-1492.

VAN'T HOF, A. E., F. MAREC, I. J. SACCHERI, P. M. BRAKEFIELD and B. J. ZWAAN, 2008 Cytogenetic Characterization and AFLP-Based Genetic Linkage Mapping for the Butterfly *Bicyclus anynana*, Covering All 28 Karyotyped Chromosomes. PLoS ONE **3:** e3882.

VAN'T HOF, A. E., P. M. BRAKEFIELD, I. J. SACCHERI and B. J. ZWAAN, 2007 Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera. Heredity **98:** 320-328.

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

ACKNOWLEDGEMENTS

This is where I say ‘thank you’. I am aware that this is the only part of the thesis that most people will ever read and therefore I should pay extra attention to this section. However, the university guidelines dictate me to acknowledge “with the restraint customary for acknowledgements in international scientific literature”, so please don’t get upset when it seems a bit cold and impersonal.

First of all my deepest thanks to Bas Zwaan and Paul Brakefield, it was a great privilege to be supervised by two scientific heavyweights, but equally important, by two great human beings. Besides my supervisors, there were a few persons who guided this work substantially. Ilik Saccheri has supported me at crucial moments and helped me improve the text of many chapters. František Marec had an enormous impact on shaping my understanding of Lepidoptera genetics. Patrícia Beldade has done me an unforgettable favour by being open about her work in progress and has been a great help in improving chapter 6. Fanja Kesbeke and Helene de Vos have significantly contributed to this work by helping out with lab work and analysis and were very pleasant people to work with. A steady supply of maize plants was provided by Els Schlatmann, Niels Wurzer, Mariëlle Lavrijsen and Bert de Winter. This thesis was assessed and approved by a reading committee consisting of Prof. Dr. Anthony D. Long, Dr. Chris Jiggins, Dr. Ilik J. Saccheri, Prof. Dr. František Marec, Dr. Patrícia Beldade, and Prof. Dr. Carel J. ten Cate. I would like to thank them for taking the time to examine my work, and for their positive feedback. Hideki Uneno was a visiting scientist from Japan who worked in our group on Ladybirds. He was also a wonderful person and friend, which makes it all the more difficult that he passed away.

As for moral support: I have a number of close friends, most of which are unfortunately not very close in terms of travelling distance. I will put you in alphabetical order because I do not want to rank friends. Vidya Athreya, Derek Daly, Nicola Edmonds, René Erkens, Azamit Haile, Dries Kuijper, Shomita Mukherjee, Darlina Naim, Jeroen Reneerkens, Thor Veen. You all made a difference in my darkest and my brightest moments and also in between. Thor Veen not only gave moral support, but also lent me his macbook to analyse the AFLP data and taught me statistics for chapter 6. I received a warm welcome from woongroep Bonobo and a badly needed quiet place of retreat thanks to Lenze Hofstee. This PhD project started off in Leiden, but final experiments, data analysis and manuscript preparation was wrapped up in Liverpool. My move there has placed me in a work environment with highly skilled molecular biologists, who are outstanding people as well. I would like to thank Chris Lowe, Suzanne Kay, Lisa Olohan, Jamie Wilson, Daryl Williams, Michelle Habets, Phill Watts and many others who dwell or dwelled in lab E and beyond, most of whom I would actually rather call friends than colleagues.

My family has always unconditionally supported me in many ways. One person whom I would like to mention is my grandfather who has been the most influential person in my life. He showed me the inspiring world of wildlife, natural history and photography; things that went down quite well for me and made me decide to become a biologist. I will close this thesis with my deepest gratitude to those who are, or have been immeasurably special to me: Sanna, Pixie, and the AX.