

## SHORT REVIEW

## Butterfly genomics eclosing

P Beldade<sup>1</sup>, WO McMillan<sup>2</sup> and A Papanicolaou<sup>3</sup><sup>1</sup>Section of Evolutionary Biology, Institute of Biology, Leiden University, Leiden, The Netherlands; <sup>2</sup>Department of Biology, University of Puerto Rico, San Juan, PR, Puerto Rico and <sup>3</sup>Department of Entomology, Max Planck Institute for Chemical Ecology, Jena, Germany

Technological and conceptual advances of the last decade have led to an explosion of genomic data and the emergence of new research avenues. Evolutionary and ecological functional genomics, with its focus on the genes that affect ecological success and adaptation in natural populations, benefits immensely from a phylogenetically widespread sampling of biological patterns and processes. Among those organisms outside established model systems, butterflies offer exceptional opportunities for multidisciplinary research on the processes generating and maintaining variation in ecologically relevant traits. Here we highlight research on wing color pattern variation in two groups of Nymphalid butterflies, the African species *Bicyclus anynana* (subfamily Satyrinae) and species of the South American genus *Heliconius* (subfamily Heliconiinae), which are emerging as

important systems for studying the nature and origins of functional diversity. Growing genomic resources including genomic and cDNA libraries, dense genetic maps, high-density gene arrays, and genetic transformation techniques are extending current gene mapping and expression profiling analysis and enabling the next generation of research questions linking genes, development, form, and fitness. Efforts to develop such resources in *Bicyclus* and *Heliconius* underscore the general challenges facing the larger research community and highlight the need for a community-wide effort to extend ongoing functional genomic research on butterflies.

*Heredity* (2008) **100**, 150–157; doi:10.1038/sj.hdy.6800934; published online 7 February 2007

**Keywords:** evolutionary and ecological functional genomics; butterfly wing patterns; *Bicyclus*; *Heliconius*; EST; linkage maps

## Introduction

## Genomics outside established model organisms

The initial lament that genomics ‘would accelerate the migration of biologists to the ‘superb six’: humans, mice, fruitflies, worms, yeast, and *Arabidopsis*’ (Murray, 2000) has failed to materialize (Crawford, 2001). Less than 5 years after the first draft of the human genome was published, nearly 600 eukaryotic genome-sequencing projects are completed or underway (cf. <http://www.genomesonline.org/>). The advantages of phylogenetically broad genome coverage are clear, and comparative analysis of diverse genomes will certainly continue to yield important insights into genome evolution and the relationships among branches of the tree of life. However, more than accumulating sequence data for comparative analysis, genomic research offers a unique opportunity to pursue a complete understanding of how genetic information is translated to produce an organism, and how modifications in genomic composition and organization give rise to biological diversity. In this quest, research on a new class of ‘emerging’ model organisms is an essential complement to the in-depth and finely detailed analysis of traditional genetic model organisms.

## Evolutionary and ecological functional genomics

The relatively new field of ‘evolutionary and ecological functional genomics’ (EEFG), and its goal of finding ‘the genes that affect ecological success and evolutionary fitness in natural environments and populations’ (Feder and Mitchell-Olds, 2003), requires an expansion outside classical model organisms. Model organisms for EEFG must combine broad genetic and ecological tractability with naturally occurring, functional variation (Feder and Mitchell-Olds, 2003).

Lepidoptera in general, and butterflies in particular, offer outstanding opportunities for integrative research at the interface between genomes and biological complexity. In spite of their immense biological (very species rich), economical (pests, pollinators and silk production), and societal (education and public understanding of science) value, available genomic resources in Lepidoptera have been limited. This situation is finally changing and independent efforts to develop core resources are underway for several species of butterflies and moths. Here we provide an overview of the strengths of butterflies as models in EEFG and summarize current efforts to develop resources in two groups, *Bicyclus* and *Heliconius*, which offer unique and complementary opportunities to study the links between genomic, developmental, and phenotypic diversity. Within this context, we discuss the general challenges facing the research community and highlight the need for a community-wide effort to consolidate and extend ongoing research.

Correspondence: Dr P Beldade, Institute of Biology, University of Leiden, Kaiserstraat 63, 2311 GP Leiden, The Netherlands.

E-mail: [pbeldade@biology.leidenuniv.nl](mailto:pbeldade@biology.leidenuniv.nl)

Received 13 April 2006; revised 16 October 2006; accepted 27 November 2006; published online 7 February 2007

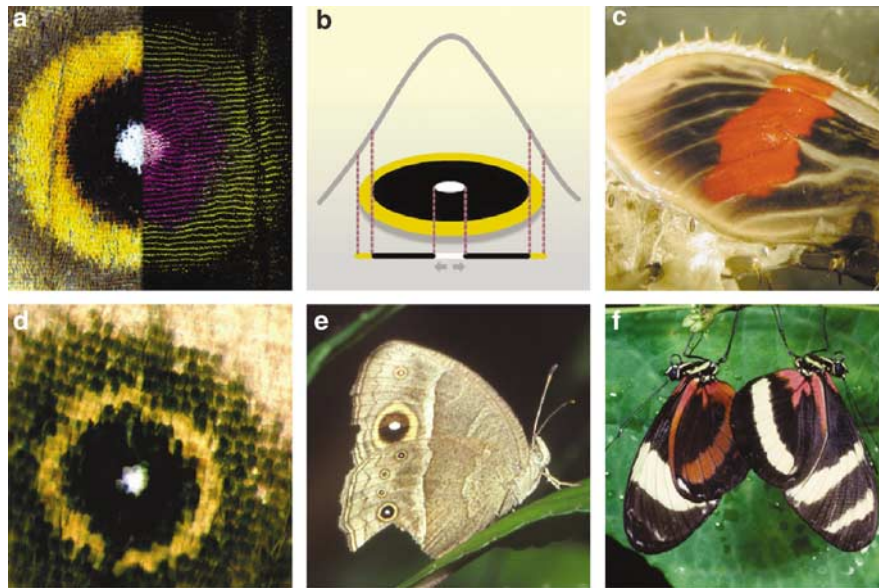
## Butterflies as emerging model organisms in genomics

The strength of butterflies as research targets derives from their extraordinary diversity, coupled with the exceptional opportunities to study the origins and maintenance of variation at nearly every biological level. The historical roots of butterfly research are deep, and the current research community is very active in a variety of areas of ecology and evolution (Boggs *et al.*, 2003) ranging from the molecular details of insect color-vision (Briscoe and Chittka, 2001; Stavenga, 2002) to the analysis of human impact on biodiversity (Kotiaho *et al.*, 2005; Mulder *et al.*, 2005). Different species have provided some of the most important case studies on diverse topics in ecology and evolution. These include (1) population genetics and metapopulation dynamics focusing on the Glanville fritillary, *Melitaea cinxia* (Hanski, 2005), (2) long distance migration of the monarch butterfly, *Danaus plexippus* (Brower, 1996; Wassenaar and Hobson, 1998; Froy *et al.*, 2003), (3) studies of Batesian mimicry, host plant detoxification, and pigment production in *Papilio* swallowtails (Li *et al.*, 2003; Nijhout, 2003), and (4) evolution and development of wing patterns in the buckeye, *Junonia coenia* and a number of other species, including *Bicyclus anynana* and *Heliconius*

(Beldade and Brakefield, 2002; McMillan *et al.*, 2002; Marcus, 2005).

## Evolution and development of butterfly wing patterns

Research on wing pattern formation is perhaps the most visually appealing example of the contribution butterflies can make to the understanding of the origins, maintenance, and modification of diversity. Virtually all of the more than 17 000 species of butterflies can be identified on the basis of the color patterns on their wings, and these highly diverse traits are emerging as invaluable systems for linking genes, gene networks, development, form, and function (Figure 1) (Nijhout, 1991; Beldade and Brakefield, 2002; McMillan *et al.*, 2002; Brakefield *et al.*, 2003; Evans and Marcus, 2006). Wings covered with colored scales (Figure 1d) are a morphological innovation of Lepidoptera and there is enormous pattern variation both within and across species. This variation is generally ecologically relevant and its adaptive value in natural populations has been extensively documented in relation to both biotic and abiotic factors (examples in Nijhout, 1991; Beldade and Brakefield, 2002; McMillan *et al.*, 2002). Furthermore, the production and maintenance of this variation can be studied across a multitude of levels of biological organization (reviewed in Beldade and Brakefield, 2002;



**Figure 1** Multidisciplinary research in two colorful dimensions. Panels illustrate the different levels at which the mechanisms governing the production and modification of butterfly wing patterns can be studied. (a) A number of developmental candidate genes have been implicated in the formation of particular pattern elements such as eyespots. The genes *spalt* (pink) and *engrailed* (green) are expressed in pupal wings (right) in the center and in the different color rings of the future adult eyespot (left) (Brunetti *et al.*, 2001). (b) Wing color pattern has also been studied in terms of the cellular interactions that underlie pattern formation and which are best understood for eyespots. In early pupal wings, the cells at the center of the presumptive eyespot produce a ‘morphogen’, which diffuses away from the center (arrows) to create a concentration gradient (gray curve). Neighboring cells then become fated to synthesize a particular color pigment depending on the morphogen concentrations they experience (where the vertical lines intersect the gray curve). (c) In *Heliconius*, the omochrome and melanin pathways (Nijhout, 1991) synthesize the pigment molecules that color the monochromatic scales. The deposition of different color pigments in different wing areas occurs late in pupal wing development (shown here for a pupa whose cuticle has been removed approximately 1 day before eclosion to expose the dorsal surface of the developing forewing). (d) The spatial arrangement of these scales in a single layer of parallel and overlapping rows produces the different pattern elements on the adult color phenotype (e.g. the eyespot on the photo). (e) Butterfly wing patterns play an important role in minimizing predation. The eyespots in *B. anynana*, for example, are thought to deflect predator’s attention away from the fragile body as seen in this specimen photographed in the wild. (f) In addition, wing patterns have also been shown to play a role in mate selection and speciation. For example, the wing patterns in *Heliconius* provide both a source of ecological post-mating isolation and mating cues important in the incipient stages of speciation. See acknowledgements regarding source of photographs.

McMillan *et al.*, 2002), ranging from the molecular details of pattern formation to the ecological relevance of pattern variation in natural populations (Figure 1).

During the last 15 years, explicit efforts to integrate methods and concepts from evolutionary and developmental biology have brought increased attention to research on butterfly wing patterns (Beldade and Brakefield, 2002; McMillan *et al.*, 2002; Beldade *et al.*, 2005; Joron *et al.*, 2006a). This research has illustrated such exciting findings as the co-option of conserved pathways to produce evolutionary novelties (Brakefield *et al.*, 1996; Brunetti *et al.*, 2001; Reed and Serfas, 2004), the contribution of key development candidate genes to phenotypic variation (Beldade *et al.*, 2002a; Kronforst *et al.*, 2006), the mapping to the same genomic location of color pattern switch genes from different species (Joron *et al.*, 2006b), and experimental tests of evolutionary constraints in morphological change (Beldade *et al.*, 2002b; Frankino *et al.*, 2005).

### Two complementary systems

The African bush-brown *B. anynana* (Nymphalidae, Satyriinae) and species within the South-American genus *Heliconius* (Nymphalidae, Heliconiinae) have emerged as important players in research on how the reciprocal interactions between development and selection shape functional diversity (Figure 1).

The wings of these two Nymphalid clades are very different in shape and pattern (Figure 2). Furthermore, the striking phenotypic differences are accompanied by clear differences in ecological function and in the

underlying genetic and developmental basis. Both groups are well suited for analysis at the molecular, organismal, and population levels and are textbook examples of natural polymorphisms. *Heliconius* is characterized by amazing geographic pattern divergence within species and pattern convergence between distantly related species (reviewed in Joron *et al.*, 2006a), and *B. anynana* by striking seasonal variation and adaptive phenotypic plasticity (Brakefield and French, 1999). In both groups, wing patterns play a role in avoiding predation (Figure 1f) (Benson, 1972; Mallet and Barton, 1989; Kapan, 2001; Langham, 2004; Lyytinen *et al.*, 2004; Brakefield and Frankino, 2006) and in mate selection (Figure 1e) (McMillan *et al.*, 1997; Jiggins *et al.*, 2001; Breuker and Brakefield, 2002; Robertson and Monteiro, 2005), but they seem to function in different manners. While the bright colors in *Heliconius* warn potential predators of the butterflies' distastefulness (Langham, 2004), those on *B. anynana* are associated to different seasonal strategies to avoid predation (camouflaging the dull-brown butterfly against a background of dry leaves, or attracting predators' attention away from the body against a green background; Figure 2a) (Brakefield and Frankino, 2006). These different ecological pressures lead to quite distinct modes of selection in natural populations: strong directional selection in *Heliconius* and divergent selection for opposite extreme phenotypes in the two seasonal environments experienced by *B. anynana* populations.

The genetic and developmental basis of wing pattern(s) formation also seems distinct in the two target groups. Study of laboratory populations of *B. anynana*



**Figure 2** Extensive morphological variation in the wing patterns of *B. anynana* and *Heliconius* provide exciting opportunities for comparative work into the interplay between genes, development, and ecology. (a) Variation in *Bicyclus* wing patterns is extensive within and across species, and laboratory *B. anynana* provides the opportunity to study different types of variation (e.g. due to plasticity, to many alleles of small effect, or to single alleles of large effect) in detail. The *B. anynana* Stock Center in Leiden maintains over 20 lines with divergent phenotypes generated by artificial selection and over 30 mutant stocks carrying spontaneous mutations of large effect. The panel shows the ventral surface of both fore- and hindwing in different stocks of *B. anynana*. The first two photos on the left correspond to the 'wild-type' outbred stock and illustrate the seasonal polyphenism that results from plasticity in relation to temperature and humidity during development (Brakefield and Frankino, 2006) (on the left, a butterfly with conspicuous eyespots typical of the 'wet season', and to the right a dull-colored butterfly more typical of the 'dry season'). The remaining photos correspond to different mutant stocks with altered eyespot patterns (from left to right: *Bigeye* with enlarged eyespots, *spotty* with extra eyespots on the forewing, *Goldeneye* with the typically black ring replaced with golden scales, and *Missing* with two eyespots absent from the hindwing). (b) The radiation in *Heliconius* color patterns couples both divergent evolution and multiple independent cases of convergent evolution. Different *Heliconius* species can be easily maintained in captivity and different populations or closely related species can be crossed to study naturally occurring variation. The panel shows geographic variation in the mimetic species, *H. erato* (top row) and *H. melpomene* (second row). The two species fall on divergent lineages in the genus, yet share identical wing patterns across their sympatric ranges and have undergone a parallel radiation into over 30 different geographic forms (Sheppard *et al.*, 1985). Color pattern variation in these species is largely explained by changes at 4–5 loci or complex of tightly linked loci of large effect. For example, allelic changes at the *Cr* locus in *H. erato* and in a complex of at least three tightly linked loci (*N*, *Yb*, *Sb*) in *H. melpomene* control most of the variation in yellow and white pattern elements among five geographic races shown. See acknowledgements regarding source of photographs.

have revealed both the presence of large amounts of segregating quantitative variation contributing to gradual response to artificial selection (Monteiro *et al.*, 1994; Monteiro *et al.*, 1997; Beldade *et al.*, 2002b), and a number of spontaneous mutant alleles with a dramatic effect on phenotype (Beldade and Brakefield, 2002; Beldade *et al.*, 2005). In *Heliconius*, in contrast, pattern variation is primarily attributable to a few genes of large effect with some minor effect modifiers (reviewed in Joron *et al.*, 2006a). Differences in overall genetic architecture are emphasized by a more detailed analysis of specific candidate genes and pathways. The formation of butterfly eyespots, including those in *B. anynana*, involves expression of genes from classical wing development pathways (Brakefield *et al.*, 1996; Brunetti *et al.*, 2001; Reed and Serfas, 2004) in and around the area of the centers (foci) of presumptive eyespots with described organizing properties (French and Brakefield, 1995; Figure 1b). However, with the notable exception of tight linkage between *wingless* and the white/yellow color switch locus K in *H. cydno* (Kronforst *et al.*, 2006), the bands and patches of color in *Heliconius* wings have so far shown no evidence for the involvement of the same developmental pathways (Reed and Gilbert, 2004; Jiggins *et al.*, 2005; Tobler *et al.*, 2005; Kapan *et al.*, 2006; Joron *et al.*, 2006a) or any type of patterning foci. Instead, genetic crosses and developmental mutants suggest that *Heliconius* patterns develop in a whole-wing proximodistal manner, independently of wing veins (Reed and Gilbert, 2004). These two seemingly distinct patterning systems within Nymphalid butterflies offer an excellent opportunity for a broad understanding of pattern formation and of the ecological consequences of variation in phenotype.

## Genomic resources in butterflies

Advances in available genomic resources are fueling genome-wide research in *B. anynana* and *Heliconius*. The functional analysis of genotypic and phenotypic variants can be pursued both at the level of the molecular details of gene function during wing development (e.g. using spontaneous mutations and genetic transformation techniques (Lewis *et al.*, 1999; Weatherbee *et al.*, 1999; Marcus *et al.*, 2004; Lewis and Brunetti, 2006)) and, at the other end of the spectrum, at the level of the ecological analysis of the adaptive value of variant phenotypes (Benson, 1972; Kapan, 2001; Langham, 2004; Mallet and Barton, 1989). Ultimately, this research promises to identify the genes and gene regions that underlie adaptive variation, link these to the genetic and biochemical networks responsible for pattern formation, and generate a fuller understanding of the interplay between genomic, developmental, and evolutionary processes.

### Genetic sequence information

Construction and analysis of both cDNA and gDNA libraries is expanding the amount of sequence information available in butterflies. In the last couple of years, moderate-scale sequencing of expression sequence tags (ESTs) has catapulted gene discovery in *B. anynana* and *Heliconius erato* and *H. melpomene*. ESTs derived from developing wings (Papanicolaou *et al.*, 2005; Beldade *et al.*, 2006) have been independently assembled resulting in the identification of thousands of putative gene objects

**Table 1** Resources in *B. anynana* and *Heliconius*

Species	Genome size	Linkage map <sup>a</sup>	UniGenes <sup>b</sup>
<i>B. anynana</i>	28 chromosomes 490 Mb 1361 cM <sup>a</sup>	352 AFLPs 8 msats	5721 (4251 <sup>c</sup> )
<i>H. erato</i>	21 chromosomes 395 Mb 1428 cM <sup>a</sup>	380 AFLPs 15 msats 16 SCNLs 9 isozymes	2981
<i>H. melpomene</i>	21 chromosomes 292 Mb 1616 cM <sup>a</sup>	219 AFLPs 23 msats 19 SCNLs	651

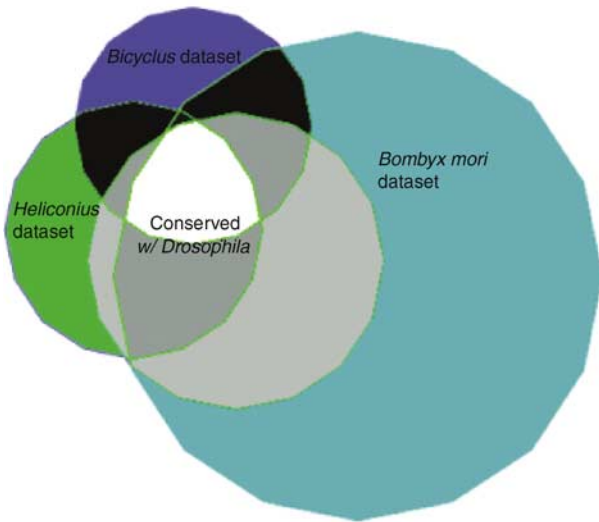
<sup>a</sup>Linkage maps for *B. anynana* (van't Hof *et al.*, 2007), *H. erato* (Tobler *et al.*, 2005; Kapan *et al.*, 2006), and *H. melpomene* (Jiggins *et al.*, 2005) include different types of markers: amplified fragment length polymorphisms (AFLPs), microsatellites (msats), single copy nuclear loci (SCNLs), and isozymes.

<sup>b</sup>Recent expression sequence tags (EST) projects have identified genes expressed in different tissues and developmental stages: *B. anynana* ESTs from five cDNA libraries made from developing wings at different stages, *H. erato* ESTs from a pooled cDNA library made from wing disc tissue collected from different geographic races at different developmental stages, and *H. melpomene* ESTs from three cDNA libraries made from whole-body pupae, late instar wing discs. All ESTs have been assembled and are available in *ButterflyBase*, [www.butterflybase.org](http://www.butterflybase.org).

<sup>c</sup>The use of different assembly algorithms in *ButterflyBase* (Papanicolaou *et al.*, 2005) and *openSputnik* (Beldade *et al.*, 2006) explains the difference in total number of gene objects (*openSputnik* assembly in brackets).

(Table 1; Figure 3). These, together with publicly available sequences from other Lepidopteran species, have been assembled in a dedicated and web-accessible database, *ButterflyBase* (via <http://www.butterflybase.org>), designed to optimize the retrieval of individual ESTs or assembled gene objects annotated based on sequence similarity and protein prediction algorithms (Papanicolaou *et al.*, 2005).

Gene discovery projects in *Heliconius* and *Bicyclus* have generated much sequence information, providing the first step towards enabling the study of genome evolution in butterflies. Many of the gene objects identified in initial EST scans showed similarity to genes in publicly available collections. This analysis has enabled the identification of genes from different functional categories, including genes known to be involved in insect wing development (candidate genes for wing pattern variation) and common 'house keeping genes' (valuable in comparative mapping studies, see below) (Papanicolaou *et al.*, 2005; Beldade *et al.*, 2006). However, there is a fairly large subset of coding regions that do not show clear homology to genes in publicly available collections (Beldade *et al.*, 2006 and *ButterflyBase*), including those of the insect model *Drosophila melanogaster* and the Lepidopteran model *Bombyx mori* (with recently published genome (Mita *et al.*, 2004; Xia *et al.*, 2004) and large-scale EST projects (Mita *et al.*, 2003; Cheng *et al.*, 2004)) (Figure 3). Particularly exciting are a few hundred fairly large predicted peptides that may be new or highly diverged genes in butterflies (Papanicolaou *et al.*, 2005; Beldade *et al.*, 2006). A functional analysis of these genes (e.g. with analysis of patterns of gene expression) and the expansion of gene collections within butterflies will help to better characterize these emerging patterns. In this respect, the planned addition



**Figure 3** Overlap in EST-derived gene collections of *Heliconius*, *B. anynana*, and *B. mori*. The scaled Venn diagram (created using Vennmaster 0.17a; <http://www.informatik.uni-ulm.de/ni/staff/HKestler/vennm/>) shows the overlap between the collections of three Lepidoptera EST-cluster data sets from *ButterflyBase* (5721 clusters for *B. anynana*, 28 036 for *B. mori*, and 3632 for the pooled *H. erato* and *H. melpomene* collections) and the *Drosophila* proteins from *FlyBase* (69 920 peptide sequences from the Genome Annotation, Release 3). Each Lepidoptera collection was compared to the known *Drosophila* proteins using BLASTX similarity analysis, and to each other lepidopteran data set using BLASTN analysis. Lepidopteran gene clusters were assigned to the different areas of the Venn diagram based on a bit-score cutoff point of 70 bits. A total of 11 541 Lepidoptera clusters were significantly similar to proteins from the insect model *Drosophila*, and a subset of 1769 (white area) were conserved between all data sets. A total of 2161 gene clusters are shared across at least two Lepidoptera, but show no similarity with *Drosophila* peptides (black areas). The areas with clusters having no significant similarity to the other collections (in color) will likely decrease as the publicly available EST collections in lepidopterans increase, since a large proportion of these clusters likely reflect limitations of sampling cDNA (relatively few ESTs are available for butterflies) and sequencing (short reads make it harder to detect sequence homology). The use of a rather conservative estimate BLAST cutoff significance level (minimum 70 bits score corresponding to  $e$ -values lower than  $E-12$ ) ensures lower rates of false positives (problematic when using gene collections that are not full sequence) but results in a potentially high number of false negatives (i.e. gene objects that do correspond to Lepidopteran homologs of annotated *Drosophila* peptides but which were not found significant here). Expansion of EST data sets for butterflies will enhance the estimates not only for large proteins but also of rapidly evolving genes or Lepidoptera- and Butterfly-specific genes.

of tens of thousands ESTs for *B. anynana* by the Joint Genome Institute (<http://www.jgi.doe.gov>) will provide an exciting data set of the genes expressed in different tissues and developmental stages in butterflies, and a powerful basis for comparative studies of Lepidopterans.

#### From identified genes to the genetic dissection of variation

The accumulation of sequence information is accelerating the development of the next generation of genomic resources in *Bicyclus* and *Heliconius* and expanding ongoing genetic mapping and expression profiling efforts.

High-density linkage maps predominately composed of amplified fragment length polymorphisms (AFLPs) and microsatellite markers are available for *B. anynana* and several *Heliconius* species (Table 1). These maps have been used to identify genomic regions that contribute to different types of phenotypic variation in the target Nymphalids (Jiggins *et al.*, 2005; Tobler *et al.*, 2005; Kapan *et al.*, 2006; Joron *et al.*, 2006b; van't Hof *et al.*, 2007). Finer resolution mapping is being pursued by (1) adding gene-based markers throughout the genome (see below) and (2) by using linked AFLP markers and *bacterial artificial chromosome* (BAC) libraries now available in *H. erato*, *H. melpomene*, *H. numata*, and *B. anynana* to develop markers in genomic regions of interest. The latter strategy has been used successfully in *Heliconius* to show that the *NYbSb* gene complex in *H. melpomene*, the *P* locus in *H. numata*, and the *Cr* locus in *H. erato* all map to homologous regions of the genome (Joron *et al.*, 2006b). This finding has been interpreted to suggest that a conserved, yet relatively unconstrained, mechanism affects pattern variation in *Heliconius*, and to imply that both convergent and divergent change can occur by the recruitment of homologous genomic regions. Positional cloning of these regions, now ongoing in all three species, will allow deeper insights into architecture, identity, and mode of action of this 'developmental hotspot' (cf. Richardson and Brakefield, 2003).

Current mapping efforts in both *Bicyclus* and *Heliconius* are concentrating on generating high-resolution gene-based maps. In this respect, ongoing EST projects are invaluable for the development of more markers for mapping and linkage analysis (Papanicolaou *et al.*, 2005; Beldade *et al.*, 2006). Sequence tags can be used to identify sequence polymorphisms in particular genes of interest, or, with targeted design, EST projects can directly combine gene and polymorphism discovery. In *B. anynana* and *Heliconius*, such a strategy has identified single-nucleotide polymorphisms and microsatellite repeats in thousands of gene objects (Beldade *et al.*, 2006 and assembled ESTs in *ButterflyBase*). These types of markers are being added to existing linkage maps and will be a very powerful tool in moving from mapped regions to the identification of the actual genes that contribute to phenotypic variation. Particularly relevant are genes whose described role in wing development makes good candidates for wing pattern variation (cf. Beldade *et al.*, 2002a). In addition, 'housekeeping' genes recurrent in EST projects of all species provide a common suite of reference markers for gene-based maps. Ribosomal protein genes, in particular, are ubiquitous in even moderate-scale EST scans and are excellent anchors for comparative linkage analysis (Yasukochi *et al.*, 2006) (Table 2). Initial analysis based on ~30 orthologous markers mapped in *H. erato*, *H. melpomene*, and *B. mori* shows surprising levels of synteny (Jiggins *et al.*, 2005; Kapan *et al.*, 2006; Yasukochi *et al.*, 2006). It will be exciting to confirm this observation for more markers in more species as the conservation of gene order would be a powerful tool to eventually identify mapped loci by comparison of maps from different species.

Gene mapping studies attempting to identify genes and gene regions contributing to variation in phenotype will be complemented with a detailed analysis of the changes in the levels of gene expression that accompany such variation. First generation high-density arrays

**Table 2** Ribosomal proteins as candidate anchor loci for comparative mapping

Organism	Taxon <sup>a</sup>	EST # <sup>b</sup>	RP # <sup>c</sup>	% <i>B. mori</i> <sup>d</sup>
<i>Heliconius melpomene</i>	Papilionoidea, Nymphalidae	1258	91	88.3
<i>H. erato</i>	Papilionoidea, Nymphalidae	8129	73	70.8
<i>Bicyclus anynana</i>	Papilionoidea, Nymphalidae	9205	96	93.2
<i>Papilio dardanus</i>	Papilionoidea, Papilionidae	698	59	57.2
<i>Euclidea glyphica</i>	Noctuoidea, Noctuidae	570	16	15.5
<i>Manduca sexta</i>	Sphingoidea, Sphingidae	1991	83	80.5
<i>Bombyx mori</i> ( <i>B. mori</i> )	Bombycoidea, Bombycidae	115 103	103	100.0
<i>Plutella xylostella</i>	Yponomeutoidea, Plutellidae	1129	91	88.3

<sup>a</sup>Classification within the order Lepidoptera (cf. National Center for Biotechnology Information (NCBI)).

<sup>b</sup>Total number of expression sequence tags (ESTs) in *ButterflyBase*.

<sup>c</sup>Number of ribosomal protein genes (RPs) identified in the different lepidopteran species using a BLASTN similarity search against the full coding sequences of *B. mori* RPs (GenBank AJ490511, AY578154, AY578155, AY583363, AY705974, AY706955-AY769343, DQ311196, DQ311216, DQ311285-DQ311290, DQ311379, DQ311405) with a cutoff threshold of 50 bits (cf. Parkinson et al., 2004).

<sup>d</sup>Number of ribosomal protein loci as a percentage of the total number of sequences searched against.

composed of genes expressed during wing development are being tested in both *Bicyclus* and *Heliconius* (Reed et al., 2007). These resources will allow expression profiling of different parts of the developing wing and different variants of the same species. Furthermore, the availability of BAC libraries will allow the characterization of regulatory regions around those genes whose map location or expression changes are associated with variation in phenotype. As the community continues to identify genes and genetic regulatory regions associated with pattern formation and pattern variation, the tools to test the functional importance of these loci are being perfected. Germline transformation technology has been developed in *B. anynana* (Marcus et al., 2004), and will be the basis for the next generation of functional experiments such as gene-targeted expression or knockouts.

## Extending genomic research in butterflies

Core resources for genomic research in butterflies have expanded substantially over the last few years. However, for butterflies to fully emerge as ecological and evolutionary genomic models, commitment of the whole research community is required. A concerted effort is crucial to stimulate the development of shared resources and strategies are required to turn butterflies into competitive players in the genomics era and to enable a more complete analysis of the questions that have made this group such powerful biological models over the last couple hundred years.

### Linking genomic, phenotypic, and ecological data

There is a rich history of collaborative multidisciplinary research in the butterfly community and the time has come to develop a common database containing both emerging genetic and genomic information and the vast amount of non-genomic data available for butterflies. Such database would link genomic/genetic diversity data (physical/linkage maps, expression data, ESTs, sequence polymorphisms, and genomic sequences) and phenotypic diversity data (quantitative and qualitative descriptions of phenotypes, images, and pedigrees) within the context of clear spatial (e.g. habitats and sampling sites) and temporal scales. Equally important is the development of common tools to utilize such a database and permit detailed queries across species

collections. These are challenging issues that require broad community participation. Fortunately, we are not alone and the challenges faced by the butterfly community are identical to those faced by other emerging model groups including *Mimulus*, Cichlids, Sticklebacks, *Daphnia*, and *Dictyostelium*. A number of bioinformatics solutions to these challenges are available including, for example, GMOD (<http://www.gmod.org>), a generalized open-source resource fully equipped with standard ontologies, file formats, web site and database options, and tools for organizing genomic data.

### Prioritize genome sequencing

Very importantly, the community must push forward efforts to get at least one butterfly genome sequenced. Genome sequence information will provide an invaluable anchor for all genetic and genomic research in this group. Genome projects in Lepidoptera are so far restricted to moths, with *B. mori* being the only published effort (Mita et al., 2004; Xia et al., 2004). It is hoped that newly available physical maps (Yamamoto et al., 2006; Yasukochi et al., 2006) will accelerate assembly and annotation of the silkworm genome, but it is still unclear how far this resource can be used in a detailed genetic analysis of butterflies. Butterflies and the Lepidopteran lineage containing *B. mori* have probably diverged more than hundred Mya (Vane-Wright, 2004), and have quite distinct biological properties related to the contrast between the diurnal (in butterflies) and the nocturnal (moths) lifestyles. Unfortunately, the same diversity that makes butterflies such attractive models has so far made community cohesion challenging. While genome projects continue to be a major financial and technical undertaking, the community will need to rally behind one or perhaps two species to be able to make the strongest possible argument for sequencing a butterfly genome. The creation of a 'Butterfly Consortium', similar to what has been put together for other organisms, is necessary to fuel discussions and overcome these types of challenges. As new technology reduces the cost of sequencing and enables the addition of new genomes (see Bonetta, 2006), the community will be well positioned to capitalize on the strength of lepidopteran diversity to study a wide array of biological processes.

## Butterfly genomics eclosing

These are exciting times, as we witness the metamorphosing of butterflies from classical organisms in ecological and evolutionary analysis to players in the genomics era. Indeed, research on *B. anynana* and *Heliconius* highlights the utility of butterflies as models for evolutionary and ecological genomic research, both satisfying essential EEFG criteria (Feder and Mitchell-Olds, 2003). With expanding genomic resources, EEFG on butterflies promises to provide important insights into the links between developmental diversity, phenotypic variation, and macroevolution. Ultimately, the combination of new tools, extraordinary diversity, and a rich history of research in ecology and evolution will ensure that butterflies can fully realize the long promised potential illustrated by the words of the nineteenth century naturalist H.W. Bates, 'the study of butterflies – creatures selected as the types of airiness and frivolity... , will some day be valued as one of the most important branches of the Biological Sciences' (Henry Walter Bates, *The Naturalist on the River Amazons*, 1864).

## Acknowledgements

We thank our colleagues who shared images included in our figures: Craig Brunetti (panel 1a), Marcel Dix (1b), Bob Reed (1c), Paul Brakefield (1e), Larry Gilbert (1f), Antónia Monteiro (2a), and Jim Mallet (2b). PB is supported by the Dutch Research Organization NWO, WOM by the American NSF and the National Evolutionary Synthesis Center, and AP by the Max-Planck-Gesellschaft.

## References

- Beldade P, Brakefield PM (2002). The genetics and evo-devo of butterfly wing patterns. *Nat Rev Genet* 3: 442–452.
- Beldade P, Brakefield PM, Long AD (2002a). Contribution of *Distal-less* to quantitative variation in butterfly eyespots. *Nature* 415: 315–318.
- Beldade P, Brakefield PM, Long AD (2005). Generating phenotypic variation: prospects from 'evo-devo' research on *Bicyclus anynana* wing patterns. *Evol Dev* 7: 101–107.
- Beldade P, Koops K, Brakefield PM (2002b). Developmental constraints versus flexibility in morphological evolution. *Nature* 416: 844–847.
- Beldade P, Rudd S, Gruber JD, Long AD (2006). A wing expressed sequence tag resource for *Bicyclus anynana* butterflies, an evo-devo model. *BMC Genomics* 7: 130.
- Benson WW (1972). Natural selection for Müllerian mimicry in *Heliconius erato* in Costa Rica. *Science* 176: 936–939.
- Boggs CL, Watt WB, Ehrlich PR (eds) (2003). *Butterflies: Ecology and Evolution Taking Flight*. The University of Chicago Press. p 756.
- Bonetta L (2006). Genome sequencing in the fast lane. *Nat Methods* 3: 141–147.
- Brakefield PM, Frankino WA (2006). Polyphenisms in Lepidoptera: multidisciplinary approaches to studies of evolution. In: Ananthakrishnan TN, Whitman DW (eds). *Phenotypic Plasticity in Insects*. Oxford University Press: Oxford.
- Brakefield PM, French V (1999). Butterfly wings: the evolution of development of colour patterns. *BioEssays* 21: 391–401.
- Brakefield PM, French V, Zwaan BJ (2003). Development and the genetics of evolutionary change within insect species. *Ann Rev Ecol Evol System* 34: 633–660.
- Brakefield PM, Gates J, Keys D, Kesbeke F, Wijngaarden PJ, Monteiro A et al. (1996). Development, plasticity and evolution of butterfly wing patterns. *Nature* 384: 236–242.
- Breuker CJ, Brakefield PM (2002). Female choice depends on size but not symmetry of dorsal eyespots in the butterfly *Bicyclus anynana*. *Proc R Soc Lond Ser B-Biol Sci* 269: 1233–1239.
- Briscoe AD, Chittka L (2001). The evolution of color vision in insects. *Annu Rev Entomol* 46: 471–510.
- Brower LP (1996). Monarch butterfly orientation: missing pieces of a magnificent puzzle. *J Exp Biol* 199: 93–103.
- Brunetti CR, Selegue JE, Monteiro A, French V, Brakefield PM, Carroll SB (2001). The generation and diversification of butterfly eyespot color patterns. *Curr Biol* 11: 1578–1585.
- Cheng TC, Xia QY, Qian JF, Liu C, Lin Y, Zha XF et al. (2004). Mining single nucleotide polymorphisms from EST data of silkworm, *Bombyx mori*, inbred strain Dazao. *Insect Biochem Mol Biol* 34: 523–530.
- Crawford DL (2001). Functional genomics does not have to be limited to a few select organisms. *Genome Biol* 2: INTERACTIONS1001.
- Evans TM, Marcus JM (2006). A simulation study of the genetic regulatory hierarchy for butterfly eyespot focus determination. *Evol Dev* 8: 273–283.
- Feder ME, Mitchell-Olds T (2003). Evolutionary and ecological functional genomics. *Nat Rev Genet* 4: 649–655.
- Frankino WA, Zwaan BJ, Stern DL, Brakefield PM (2005). Natural selection and developmental constraints in the evolution of allometries. *Science* 307: 718–720.
- French V, Brakefield PM (1995). Eyespot development on butterfly wings: the focal signal. *Dev Biol* 168: 112–123.
- Froy O, Gotter AL, Casselman AL, Reppert SM (2003). Illuminating the circadian clock in monarch butterfly migration. *Science* 300: 1303–1305.
- Hanski I (2005). Landscape fragmentation, biodiversity loss and the societal response. The long term consequences of our use of natural resources may be surprising and unpleasant. *EMBO Rep* 6: 388–392.
- Jiggins CD, Mavarez J, Beltran M, McMillan WO, Johnston JS, Bermingham E (2005). A genetic linkage map of the mimetic butterfly *Heliconius melpomene*. *Genetics* 171: 557–570.
- Jiggins CD, Naisbit RE, Coe RL, Mallet J (2001). Reproductive isolation caused by colour pattern mimicry. *Nature* 411: 302–305.
- Joron M, Jiggins CD, Papanicolaou A, McMillan WO (2006a). *Heliconius* wing patterns: an evo-devo model for understanding phenotypic diversity. *Heredity* 97: 157–167.
- Joron M, Papa R, Beltran M, Chamberlain N, Mavarez J, Baxter S et al. (2006b). A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biol* 4: e303.
- Kapan DD (2001). Three-butterfly system provides a field test of mullerian mimicry. *Nature* 409: 338–340.
- Kapan DD, Flanagan NS, Tobler A, Papa R, Reed RD, Gonzalez JA et al. (2006). Localization of Müllerian mimicry genes on a dense linkage map of *Heliconius erato*. *Genetics* 173: 735–757.
- Kotiaho JS, Kaitala V, Kolmonen A, Paivinen J (2005). Predicting the risk of extinction from shared ecological characteristics. *Proc Natl Acad Sci USA* 102: 1963–1967.
- Kronforst MR, Young LG, Kapan DD, McNeely C, O'Neill RJ, Gilbert LE (2006). Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proc Natl Acad Sci USA* 103: 6575–6580.
- Langham GM (2004). Specialized avian predators repeatedly attack novel color morphs of *Heliconius* butterflies. *Evolution* 58: 2783–2787.
- Lewis DL, Brunetti CR (2006). Ectopic transgene expression in butterfly imaginal wing discs using vaccinia virus. *Bio-techniques* 40: 48–52.
- Lewis DL, DeCamillis MA, Brunetti CR, Halder G, Kassner VA, Selegue JE et al. (1999). Ectopic gene expression and homeotic transformations in arthropods using recombinant Sindbis viruses. *Curr Biol* 9: 1279–1287.
- Li W, Schuler MA, Berenbaum MR (2003). Diversification of furanocoumarin-metabolizing cytochrome P450 monooxy-

- genes in two papilionids: specificity and substrate encounter rate. *Proc Natl Acad Sci USA* **100**: 14593–14598.
- Lyytinen A, Brakefield PM, Lindstrom L, Mappes J (2004). Does predation maintain eyespot plasticity in *Bicyclus anynana*? *Proc R Soc Lond Ser B-Biol Sci* **271**: 279–283.
- Mallet J, Barton NH (1989). Strong natural selection in a warning-color hybrid zone. *Evolution* **43**: 421–431.
- Marcus JM (2005). Jumping genes and AFLP maps: transforming lepidopteran color pattern genetics. *Evol Dev* **7**: 108–114.
- Marcus JM, Ramos DM, Monteiro A (2004). Germline transformation of the butterfly *Bicyclus anynana*. *Proc R Soc Lond Ser B-Biol Sci* **271**: S263–S265.
- McMillan WO, Jiggins CD, Mallet J (1997). What initiates speciation in passion-vine butterflies? *Proc Natl Acad Sci USA* **94**: 8628–8633.
- McMillan WO, Monteiro A, Kapan DD (2002). Development and evolution on the wing. *Trends Ecol Evol* **17**: 125–133.
- Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H *et al.* (2004). The genome sequence of silkworm, *Bombyx mori*. *DNA Res* **11**: 27–35.
- Mita K, Morimyo M, Okano K, Koike Y, Nohata J, Kawasaki H *et al.* (2003). The construction of an EST database for *Bombyx mori* and its application. *Proc Natl Acad Sci USA* **100**: 14121–14126.
- Monteiro A, Brakefield PM, French V (1997). Butterfly eyespots: the genetics and development of the color rings. *Evolution* **51**: 1207–1216.
- Monteiro AF, Brakefield PM, French V (1994). The evolutionary genetics and developmental basis of wing pattern variation in the butterfly *Bicyclus anynana*. *Evolution* **48**: 1147–1157.
- Mulder C, Aldenberg T, de Zwart D, van Wijnen HJ, Breure AM (2005). Evaluating the impact of pollution on plant–Lepidoptera relationships. *Environmetrics* **16**: 357–373.
- Murray AW (2000). Whither genomics? *Genome Biol* **1**: COMMENT003.
- Nijhout HF (1991). *The Development and Evolution of Butterfly Wing Patterns*. Smithsonian Inst. Press: Washington.
- Nijhout HF (2003). Polymorphic mimicry in *Papilio dardanus*: mosaic dominance, big effects, and origins. *Evol Dev* **5**: 579–592.
- Papanicolaou A, Joron M, Mcmillan WO, Blaxter ML, Jiggins CD (2005). Genomic tools and cDNA derived markers for butterflies. *Mol Ecol* **14**: 2883–2897.
- Parkinson J, Mitreva M, Whitton C, Thomson M, Daub J, Martin J *et al.* (2004). A transcriptomic analysis of the phylum Nematoda. *Nat Genet* **36**: 1259–1267.
- Reed RD, Gilbert LE (2004). Wing venation and distal-less expression in *Heliconius* butterfly wing pattern development. *Dev Genes Evol* **214**: 628–634.
- Reed RD, McMillan WO, Nagy LM (2007). Gene regulation underlying adaptive variation: cinnabar and vermilion in the development and polymorphism of *Heliconius* butterfly wing patterns (in preparation).
- Reed RD, Serfas MS (2004). Butterfly wing pattern evolution is associated with changes in a notch/distal-less temporal pattern formation process. *Curr Biol* **14**: 1159–1166.
- Richardson MK, Brakefield PM (2003). Developmental biology – hotspots for evolution. *Nature* **424**: 894–895.
- Robertson KA, Monteiro A (2005). Female *Bicyclus anynana* butterflies choose males on the basis of their dorsal UV-reflective eyespot pupils. *P Roy Soc B-Biol Sci* **272**: 1541–1546.
- Sheppard PM, Turner JRG, Brown KS, Benson WW, Singer MC (1985). Genetics and the evolution of Müllerian mimicry in *Heliconius* butterflies. *Phil Trans Roy Soc B* **308**: 433–613.
- Stavenga DG (2002). Reflections on colourful ommatidia of butterfly eyes. *J Exp Biol* **205**: 1077–1085.
- Tobler A, Kapan D, Flanagan NS, Gonzalez C, Peterson E, Jiggins CD *et al.* (2005). First-generation linkage map of the warningly colored butterfly *Heliconius erato*. *Heredity* **94**: 408–417.
- Vane-Wright D (2004). Entomology – butterflies at that awkward age. *Nature* **428**: 477–480.
- van't Hof AE, Saccheri IJ, Marec F, Brakefield PM, Zwaan B (2007). A high density AFLP-based genetic linkage map for the butterfly *Bicyclus anynana*, covering all 28 karyotyped chromosomes (in preparation).
- Wassenaar LI, Hobson KA (1998). Natal origins of migratory monarch butterflies at wintering colonies in Mexico: New isotopic evidence. *Proc Natl Acad Sci USA* **95**: 15436–15439.
- Weatherbee SD, Nijhout HF, Grunert LW, Halder G, Galant R, Selegue J *et al.* (1999). *Ultrabithorax* function in butterfly wings and the evolution of insect wing patterns. *Curr Biol* **9**: 109–115.
- Xia QY, Zhou ZY, Lu C, Cheng DJ, Dai FY, Li B *et al.* (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* **306**: 1937–1940.
- Yamamoto K, Narukawa J, Kadono-Okuda K, Nohata J, Sasanuma M, Suetsugu Y *et al.* (2006). Construction of a single nucleotide polymorphism linkage map for the silkworm, *Bombyx mori*, based on BAC end-sequences. *Genetics* **173**: 151–161.
- Yasukochi Y, Ashakumary LA, Baba K, Yoshido A, Sahara K (2006). A second generation integrated map of the silkworm reveals synteny and conserved gene order between lepidopteran insects. *Genetics* **173**: 1319–1328.