Heredity

npg

ORIGINAL ARTICLE

Linkage map of the peppered moth, *Biston betularia* (Lepidoptera, Geometridae): a model of industrial melanism

AE Van't Hof[1], P Nguyen[2], M Dalíková[2], N Edmonds[1], F Marec[2] and IJ Saccheri[1]

We have constructed a linkage map for the peppered moth (*Biston betularia*), the classical ecological genetics model of industrial melanism, aimed both at localizing the network of loci controlling melanism and making inferences about chromosome dynamics. The linkage map, which is based primarily on amplified fragment length polymorphisms (AFLPs) and genes, consists of 31 linkage groups (LGs; consistent with the karyotype). Comparison with the evolutionarily distant *Bombyx mori* suggests that the gene content of chromosomes is highly conserved. Gene order is conserved on the autosomes, but noticeably less so on the Z chromosome, as confirmed by physical mapping using bacterial artificial chromosome fluorescence *in situ* hybridization (BAC-FISH). Synteny mapping identified three pairs of *B. betularia* LGs (11/29, 23/30 and 24/31) as being orthologous to three *B. mori* chromosomes (11, 23 and 24, respectively). A similar finding in an outgroup moth (*Plutella xylostella*) indicates that the *B. mori* karyotype ($n = 28$) is a phylogenetically derived state resulting from three chromosome fusions. As with other Lepidoptera, the *B. betularia* W chromosome consists largely of repetitive sequence, but exceptionally we found a W homolog of a Z-linked gene (*laminin A*), possibly resulting from ectopic recombination between the sex chromosomes. The *B. betularia* linkage map, featuring the network of known melanization genes, serves as a resource for melanism research in Lepidoptera. Moreover, its close resemblance to the ancestral lepidopteran karyotype ($n = 31$) makes it a useful reference point for reconstructing chromosome dynamic events and ancestral genome architectures. Our study highlights the unusual evolutionary stability of lepidopteran autosomes; in contrast, higher rates of intrachromosomal rearrangements support a special role of the Z chromosome in adaptive evolution and speciation.

Heredity (2013) **110**, 283–295; doi:10.1038/hdy.2012.84; published online 5 December 2012

INTRODUCTION

Linkage maps have been most frequently constructed for the primary purpose of identifying loci controlling specific phenotypic traits (Sturtevant, 1913); however, genome-wide linkage maps are also an efficient starting point for exploring genome architecture and chromosome dynamics (Beldade *et al.*, 2009), complementing the current revolution in next-generation whole-genome sequencing (Wang *et al.*, 2005). Lepidoptera are an interesting group in this respect, characterized by wide variation in chromosome number (Suomalainen and Brown, 1984; Kandul *et al.*, 2007) around a strong, putatively ancestral mode (haploid $n = 31$), encountered in approximately one-third of the species (White, 1973). The dynamism implied by high rates of chromosome fission/fusion contrasts with a highly conserved synteny of genes between chromosomes across distantly related taxa (Pringle *et al.*, 2007; Sahara *et al.*, 2007; Beldade *et al.*, 2009; Yasukochi *et al.*, 2009; d'Alençon *et al.*, 2010; Baxter *et al.*, 2011). This latter feature allows unambiguous identification of orthologous chromosomal regions in different species to reconstruct the events that produced the current pattern of structural diversity.

Lepidopteran linkage maps have been published for 8 species of Lepidoptera, within 4 of the 43 (Van Nieukerken *et al.*, 2011) currently recognized superfamilies. This sample is concentrated on species within the Papilionoidea (*Heliconius* spp. (Jiggins *et al.*, 2005;

Kapan *et al.*, 2006; The *Heliconius* Genome Consortium, 2012), *Bicyclus anynana* (Van't Hof *et al.*, 2008; Beldade *et al.*, 2009), *Colias* hybrid (Wang and Porter, 2004), *Papilio* hybrid (Winter and Porter, 2010)) and one other macrolepidopteran, *Bombyx mori* (Bombycoidea) (Yasukochi *et al.*, 2006; Yamamoto *et al.*, 2008). Linkage maps are also available for two other distantly related superfamilies, Pyraloidea (*Ostrinia nubilalis*; Dopman *et al.*, 2004) and Yponomeutoidea (*Plutella xylostella* (Heckel *et al.*, 1999; Baxter *et al.*, 2011)). *Biston betularia*, the peppered moth of industrial melanism fame (Majerus, 1998; Cook, 2003), is a significant addition to this list, representing the first linkage map for the Geometroidea, the second largest superfamily of Lepidoptera (after the Noctuoidea), which includes several major crop pests and ecological models.

A comprehensive reconstruction of lepidopteran chromosome evolution is currently lacking as a result of the wide use of anonymous markers, nonoverlapping panels of annotated markers, taxonomic skew and low number of mapped genomes. Linkage map resolution and genome coverage are also important limiting factors in comparative analyses. High marker number allows accurate comparison of chromosome content, whereas high resolution is required for fine-scale reconstruction of intrachromosomal rearrangements. In practice, linkage map construction involves a tradeoff between marker number and resolution, the latter being fundamentally determined by the

[1]Institute of Integrative Biology, University of Liverpool, Liverpool, UK and [2]Biology Centre ASCR, Institute of Entomology, and Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic
Correspondence: Dr I Saccheri, Institute of Integrative Biology, University of Liverpool, The Biosciences Building, Crown Street, Liverpool L69 7ZB, UK.
E-mail: saccheri@liv.ac.uk

number of recombinant offspring. A further difficulty arises when attempting to reconstruct the sequence of chromosomal rearrangement events on the basis of a minimal sample of independently derived states, such as *B. anynana* ($n = 28$) and *Heliconius melpomene* ($n = 21$) versus *B. mori* ($n = 28$) (Pringle et al., 2007; Beldade et al., 2009). The addition of linkage maps for species with the ancestral $n = 31$, of which there is currently only one (Heckel et al., 1999; Baxter et al., 2011), should eliminate much of this uncertainty.

Lepidopteran sex chromosomes and sex-linked genes have historically been investigated more thoroughly than the autosomes, partly because they are easier to identify. Females are the heterogametic sex in Lepidoptera, with most commonly a W/Z female, Z/Z male system (for deviations see Traut et al., 2007). The sex chromosomes stand out in female pachytene spreads by their heterochromatic appearance and produce hemizygous inheritance patterns. In contrast, lepidopteran autosomes are far less distinct because they are cytogenetically uniform as a result of their holokinetic arrangement (that is, lacking a primary constriction, the centromere; Murakami and Imai, 1974; Carpenter et al., 2005), large number, small size and insusceptibility to banding techniques. The gene composition of the Z chromosome is conserved in Lepidoptera, although not as strongly as some of the autosomes (Beldade et al., 2009; Baxter et al., 2011; Kroemer et al., 2011). A number of genes, such as *kettin*, *tpi*, *6-PGD*, *period*, *apterous* and *Ldh*, have been assigned to the Z chromosomes of various Lepidoptera (Johnson and Turner, 1979; Gotter et al., 1999; Suzuki et al., 1999; Dopman et al., 2004; Narita et al., 2006; Pringle et al., 2007; Putnam et al., 2007; Van't Hof et al., 2008; Kroemer et al., 2011; Yoshido et al., 2011) and are now established lepidopteran Z-genes. The silkworm (*Bombyx mori*) whole-genome sequence assembly (The International Silkworm Genome Consortium, 2008) includes a large number of additional Z-linked genes, many of which have since also been found on the Z chromosome in other Lepidoptera (Beldade et al., 2009; Baxter et al., 2011; Kroemer et al., 2011).

There are very few active genes described on lepidopteran W chromosomes (reviewed in Marec et al., 2010), which is consistent with undetectable transcription activity levels of W chromatin in interphase nuclei of somatic cells (Traut and Scholz, 1978). Not a single gene has been found on the W chromosome in *Heliconius* spp., *B. anynana*, or *P. xylostella*, in spite of comprehensive gene-based surveys (Pringle et al., 2007; Beldade et al., 2009; Baxter et al., 2011). An exceptionally small W chromosome may partly explain this paucity of genes in *B. anynana* (Van't Hof et al., 2008), but there is no such association for the *B. mori* W chromosome that is almost entirely composed of repetitive sequences originating from retrotransposons (Abe et al., 2005). Some records of W-linked genes are disputed, such as the dark morph locus in *Papilio glaucus* (Andolfatto et al., 2003) and an egg size determining gene, *Esd*, in the silkworm (Kawamura, 1988; but see Fujii et al., 2010). A genuine example is the sex determining gene *Fem* in the silkworm, mapped to a narrow region of the W chromosome (Abe et al., 2008) that contains two genes encoding a putative zinc-finger protein; however, its role in sex determination remains to be elucidated (Suzuki, 2010).

Industrial melanism in *B. betularia* is perhaps the best-known example of a rapid adaptive response to environmental change (Cook, 2003), although many other moths, especially geometrids, responded similarly (Lees, 1981). The initial motivation for this linkage map was therefore to find the locus determining the difference between the wild-type (typical) and the black (*carbonaria*) forms. The *carbonaria* locus has been localized to a 200-kb region on the ortholog of *B. mori* chromosome 17 but remains to be pinpointed (Van't Hof et al., 2011). The linkage map presented here, for all the chromosomes,

featuring the major genes involved in melanin biosynthesis as well as several other important patterning genes, serves as a frame of reference for melanism research in Lepidoptera, providing an essential basis for unraveling the regulatory cascade emanating from the melanism 'switch' locus. The *B. betularia* map also sheds new insight into the likely chromosomal architecture of the ancestral lepidopteran genome, and into interaction and shared gene content of the Z and W chromosomes.

MATERIALS AND METHODS
Linkage mapping
The linkage map is based on family no. 16, a cross between a typical female and a heterozygous *carbonaria* male (the *carbonaria* allele is dominant) and 73 offspring that was used previously to screen candidate genes (Van't Hof and Saccheri, 2010) and to construct the LG17 map that includes the *carbonaria* locus (Van't Hof et al., 2011). The previously published LG17 map was also based on family no. 8, whereas the additional markers presented here are based on family no. 16 alone. Anonymous amplified fragment length polymorphisms (AFLPs) were used to provide a genome-wide recombination-segregation matrix that allowed the positioning of genes, bacterial artificial chromosomes (BACs), microsatellites and the *carbonaria* locus as a morphological marker. A number of genes in the linkage map were specifically included because of their role in insect melanization or pigment pattern formation, and others to provide annotated anchors, improve coverage in the absence of other markers or target regions of special interest (such as the *carbonaria* chromosome, the Z chromosome and putative *B. mori* fusion chromosomes). The linkage map was generated with Joinmap 4 (Kyazma, Wageningen, The Netherlands) using regression mapping with independence LOD (logarithm of the odds) test to group markers and Haldane's mapping algorithm to position markers using Joinmap's default parameters.

The segregation pattern of individual AFLPs determines their utility as mapping markers. Paternally segregating markers (often named male informative (MI) markers), inherited from a heterozygous father and a homozygous mother, have a 1:1 offspring ratio and can be used for positional mapping based on recombination frequencies between markers. Maternally segregating markers (female informative (FI); mother heterozygous, father homozygous) also have a 1:1 ratio and can be assigned to chromosomes, but cannot be positioned because Lepidoptera females do not recombine (see, for example, Traut, 1977; Nokkala, 1987). Interpretation of 3:1 segregating both informative (BI) dominant AFLP markers (mother and father heterozygous) is problematic in Lepidoptera because heterozygotes cannot be distinguished from the dominant homozygotes in the offspring and, moreover, they include the nonrecombinant maternal component. A number of approaches have been used to deal with these dominant BI markers in Lepidoptera (Jiggins et al., 2005; Kapan et al., 2006; Yasukochi et al., 2006; Van't Hof et al., 2008), each predicted to give a reasonable approximation of the actual marker order. We opted not to compromise the reliability of the linkage map and use only markers that are fully informative in all offspring. Dominant BI markers in marker-poor regions were sequenced and used to isolate BAC clones to obtain codominant polymorphisms for these positions (replacing the original dominant BI AFLP). Dominant BI markers that would only increase marker number but not marker coverage were excluded from the final linkage map.

A small proportion of BI AFLPs were codominant owing to an indel within the amplicons and thus segregated with a 1:2:1 ratio (see, for example, Supplementary Figure S1). These markers still contain the nonrecombinant maternal component, which can be easily identified because the maternal alleles are inherited in fixed chromosome-specific patterns named chromosome prints (Yasukochi, 1998). This chromosome print can be used to remove the maternal component from codominant BI markers, leaving only the fully informative paternal segregation pattern. The same BI 1:2:1 → MI 1:1 procedure was used to 'clean up' codominant BI genes and BAC segregation patterns (see, for example, following explanations in Jiggins et al., 2005; Kapan et al., 2006).

The Z-chromosome markers were identified from a segregation pattern defined by AFLP peak presence in all sons and half the daughters. The segregation pattern in these daughters was used to identify Z-linked MI

markers that segregate in both sexes. FI markers, which were exclusively present in daughters, are on the W chromosome, and those that were exclusively present in sons were assigned to the Z chromosome.

Marker development

AFLP markers were generated using a protocol modified from Vos *et al.* (1995). Phenol/chloroform extracted genomic DNA from half a thorax was digested with restriction enzyme combinations *Eco*RI/*Mse*I and *Eco*RI/*Hha*I and ligated to standard *Eco*RI and *Mse*I adapters (Vos *et al.*, 1995) and custom-designed *Hha*I adapters (5′-GACGATGAGTCCTGAACG-3′ + 5′-TTCAGG ACTCAT-3′) in a single reaction. The restriction–ligation reactions were performed for 6 h at 37 °C in 40 μl containing 150 ng genomic DNA, 1.5 units of each restriction enzyme, 1 × NEBuffer (NEB2 for *Mse*I/*Eco*RI and NEB4 for *Hha*I/*Eco*RI), 4 μg bovine serum albumin, 64 units T4 ligase, 1 mM ATP, 10 mM dithiothreitol, 68 nM *Eco* adapter and 612 nM *Mse* adapter or *Hha* adapter. Restriction enzymes, ligase, bovine serum albumin, NEBuffers and dithiothreitol were sourced from New England Biolabs (Ipswich, MA, USA). A PCR pre-amplification (preamp) was performed in a total volume of 15 μl

containing 2 μl restriction–ligation product, 200 μM of each dNTP, 0.9 units *AmpliTaq* DNA polymerase (Applied Biosystems, Foster City, CA, USA), 1 × GeneAmp buffer I, 200 nM *Eco* + A primer, 1.5 μM *Mse* + 1 or *Hha* + 1 primer. The primers have a fixed core and an extension that varies in length and nucleotide composition. The core sequences of the *Mse* and *Eco* primers are as in Vos *et al.* (1995), the *Hha* core sequence is 5′-GATGAGTCCTGAACGC-3′ and ' + 1' represents single-nucleotide extensions, with C and G used for *Hha* primers and all four nucleotides used for *Mse* primers. The subsequent selective PCR amplification (selamp) used the same reaction mix as preamp except that 1.5 μl diluted preamp product replaced restriction–ligation as the template and the primers used were 100 nM fluorescent *Eco* + 3 and 300 nM *Mse* + 2 or *Hha* + 2. Preamp and selamp thermal cycling profiles are described in Van't Hof *et al.* (2008). The codes of the AFLP markers in Figures 1 and 2 refer to the primers and fragment length, with lowercase m, h and e for *Mse*, *Hha* and *Eco* primer cores, respectively, followed by the extensions in uppercase and the fragment size (for example, mAGeACG178). The AFLP genotypes were separated with an ABI3130XL fragment analyzer and scored in Genotyper 3.6 (Applied Biosystems).
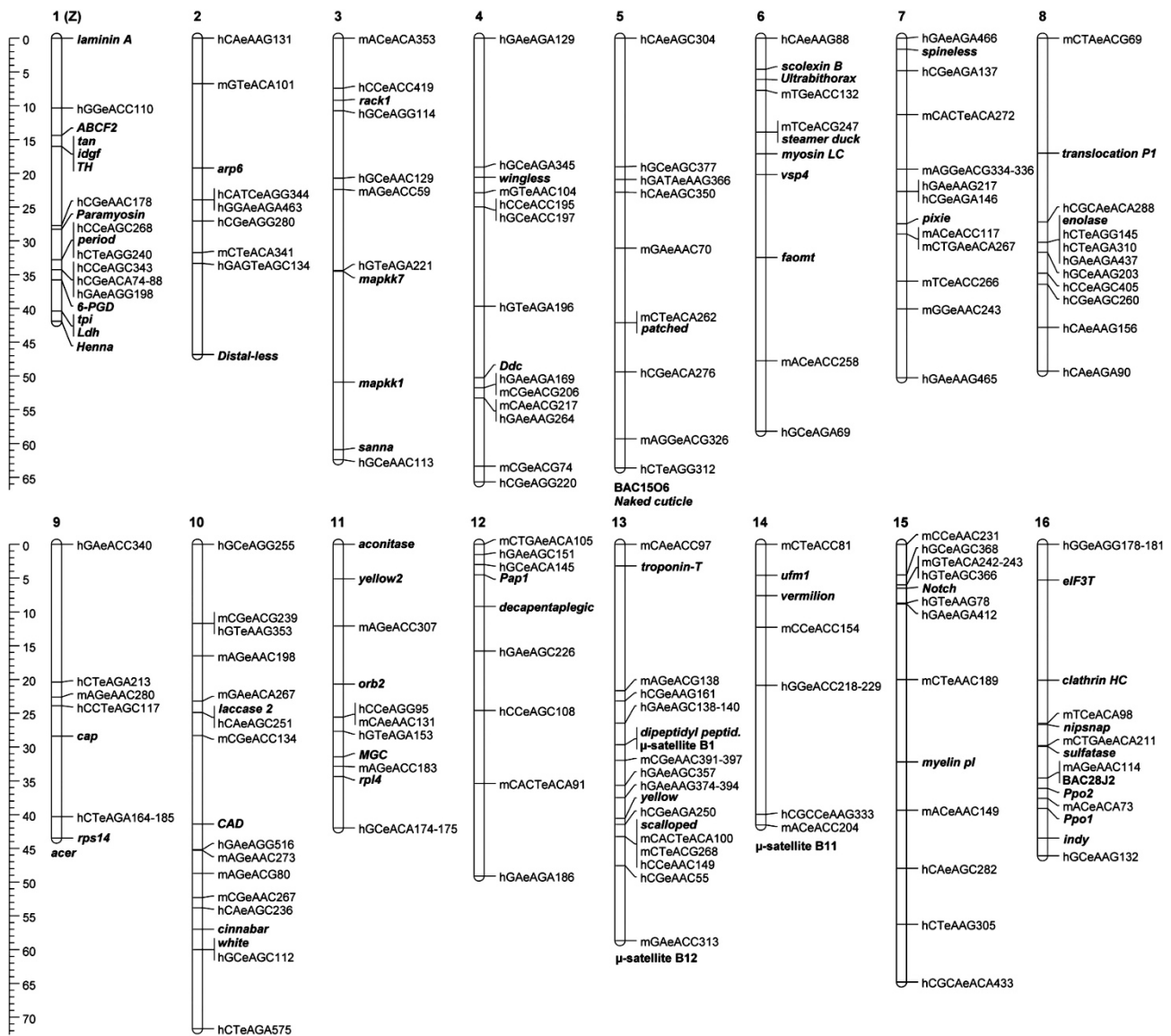

Figure 1 Linkage map of *Biston betularia* chromosomes 1–16. The linkage map is composed of AFLPs, nuclear genes, BACs and microsatellites. AFLP markers are defined by their initial restriction enzymes and selective primer combinations. Nuclear genes are in bold (full gene names are listed in Supplementary Table S1). BACs and microsatellites are also in bold, with BACs named according to library plate position. Assigned but unpositioned (FI) markers are indicated at the bases of the LGs.
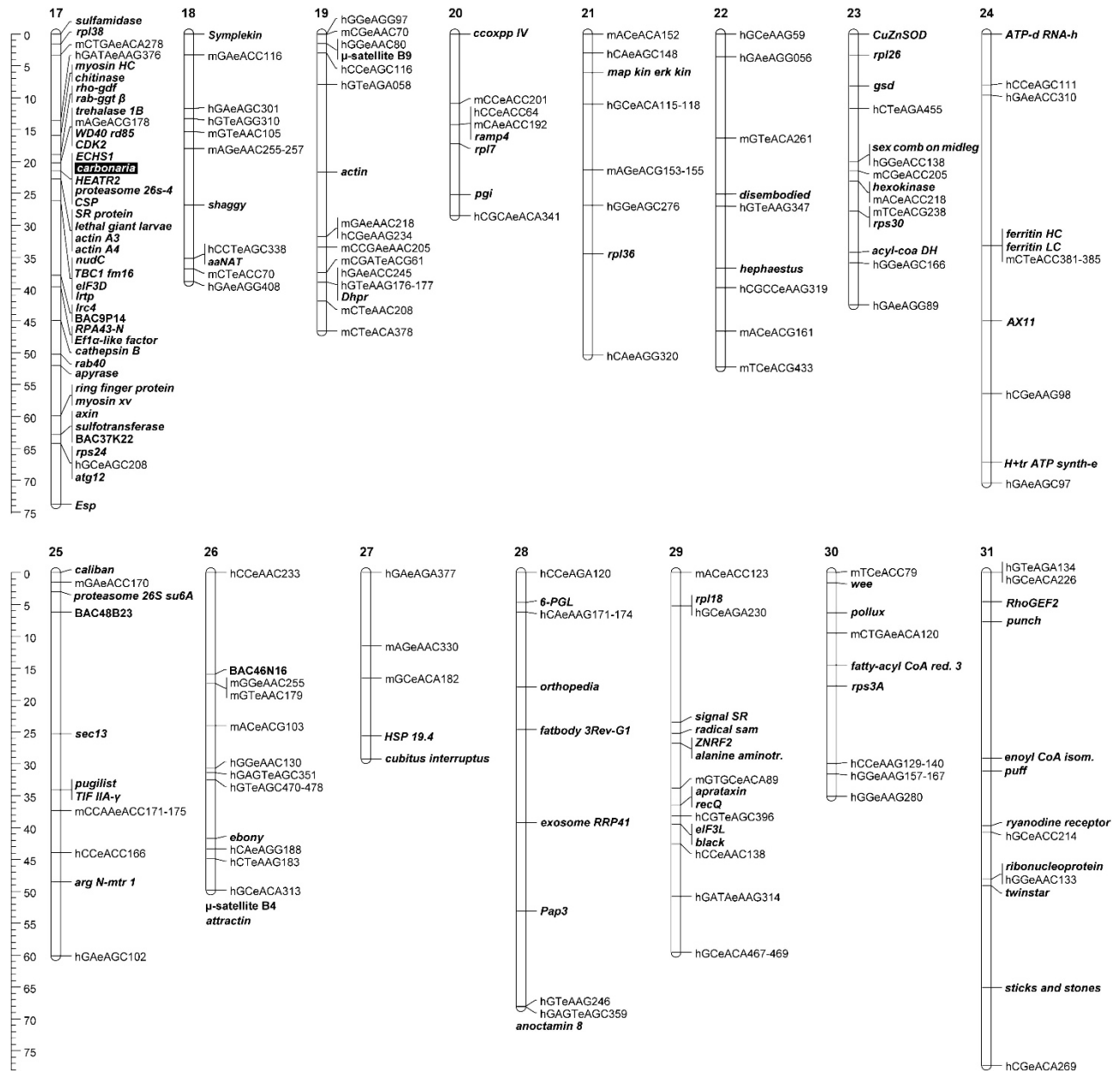
Figure 2 Linkage map of *Biston betularia* chromosomes 17–31. The linkage map has been continued from Figure 1, including the morphological *carbonaria* locus on LG17.

Selective amplification used primers with two selective nucleotides on the '4-cutter' end of the fragment (matching the *Hha* and *Mse* adapters) as the more usual three selective bases generated too low density of fragments in *B. betularia*. A drawback of this strategy was greater variability in marker quality, including weak, inconsistent or overlapping peaks. Such low-quality markers were not included in the final linkage map, but they were screened in a preliminary draft map for their potential to improve linkage map coverage. Low-quality AFLPs that appeared to map to regions with low coverage were made more reliable by extending primers with additional selective bases, which increases peak strength, decreases peak density and differentiates similar sized fragments (Supplementary Figures S1 and S2). Genes were obtained primarily as mRNA-coding sequences either from a transcriptome library (Van't Hof and Saccheri, 2010) or with gene-specific degenerate primers using complementary DNA or genomic DNA as template following the methods described in Van't Hof and Saccheri (2010). All primers were designed specifically for this project,

with the exception of *Ddc* and *Pgi* degenerate primers (Regier, 2007) and *Ppo*I_F (Hartzer *et al.*, 2005); the primer sequences are included in Supplementary Table S2. Several gene sequences were characterized from BACs that were isolated with PCR probes derived from mapped and sequenced AFLPs. The segregation patterns of polymorphisms within these BACs were compared against the AFLP segregation to exclude false-positive BACs. Three genes (*enolase*, *CAD* and *period*) were available for *B. betularia* in GenBank (accession numbers in Supplementary Table S2). The polymorphisms used for genotyping genes in the mapping family were usually found within introns, exons or untranslated regions, but occasionally, when such polymorphisms were not available, their segregation patterns were inferred from polymorphisms in end-sequences of BACs that included the gene. The polymorphisms were genotyped either by PCR/restriction fragment length polymorphism (single-nucleotide polymorphisms (SNPs)), separation on agarose gel (large insertion/deletions (indels)) or sequencing (SNPs and small indels).

Rpl36 was also mapped in the butterfly *B. anynana* to investigate the chromosomal origin of this gene. The coding sequence was obtained from expressed sequence tag (EST) sequences deposited in GenBank (Beldade et al., 2006) and an intron (GenBank accession number JF811431) containing a polymorphism was mapped in family 12 of the gene-based linkage map of *B. anynana* (Beldade et al., 2009). The gene was amplified with primer combination: forward 5′-GCGAAAAGGCCACAAAACAACCAAA-3′ and reverse 5′-CCTCGCGGACCAGATCACGGACA-3′, and genotyped by Sanger sequencing.

Gene identity was verified with blastx against GenBank using an e-value threshold of 1e−20, with the higher e-values usually representing relatively short available coding sequence rather than highly diverged orthologous proteins (e-value is not only dependent on similarity, but also on length). To explore shared synteny with *B. mori*, a blastx and tblastx (to cover genes with annotation errors) against SilkDB were used to identify orthologs. Only unambiguous single-copy nuclear genes without known highly or moderately similar paralogs in *B. mori* were used as anchoring genes. Some of the melanization genes are members of gene-families though, and the default best blast hits were tested for being the true ortholog (rather than a paralog) using the stringent criteria described in Van't Hof and Saccheri (2010). The sequences of the genes and BACs are deposited in GenBank with accession numbers listed in Supplementary Table S2. The *B. betularia* microsatellites described in Daly et al. (2004) were amplified and sequenced in the parents of the mapping family and screened for polymorphisms. The informative loci were genotyped in the offspring by separation on 2% agarose for μsat-B01 and B09 that have 23-bp and 18-bp indels, respectively, and by Sanger sequencing for the rest.

Cytogenetic procedures

Mitotic and meiotic chromosomes were obtained from testes of mid-late instar male larvae from a cross between a typical Scottish mother and a melanic Austrian father, and from both testes and ovaries of diapausing pupae originating from a cross between a typical female (Scotland) and a typical male (Devon). Spread chromosome preparations were made essentially following the procedure described in Sahara et al. (1999). Briefly, gonads were dissected in physiological solution, swollen for 10 min in hypotonic solution (83 mM KCl and 17 mM NaCl), fixed for 15–30 min in Carnoy's fixative (ethanol/chloroform/acetic acid 6:3:1), dissociated in 60% acetic acid and spread on the slide using a heating plate at 45 °C. The preparations were then passed through an ethanol series (70, 80, and 100%; 30 s each) and stored at −80 °C until further use. Shortly before use, selected preparations were removed from the freezer, dehydrated in the graded ethanol series and air-dried. For basic karyotype analysis, the dry preparations were stained with fluorescent dyes, either 0.5 μg ml^{-1} DAPI (4,6-diamino-2-phenylindole; Sigma-Aldrich, St Louis, MO, USA) or 100 nM YOYO-1 (Molecular Probes Inc., Eugene, OR, USA), and mounted in antifade based on DABCO (1,4-diazabicyclo(2.2.2)-octane; Sigma-Aldrich) as described in Van't Hof et al. (2008).

BAC probes for fluorescence in situ hybridization (FISH) were prepared as follows. BAC-DNA was extracted with the BACMAX DNA Purification Kit (Epicentre Biotechnologies, Madison, WI, USA), amplified using illustra GenomiPhi HY DNA Amplification Kit (GE Healthcare Life Sciences, Milwaukee, WI, USA), and labeled with a fluorochrome, either Amersham Cy3-dUTP (GE Healthcare UK Ltd, Buckinghamshire, UK), Alexa Fluor 488-dUTP (Molecular Probes Inc., Eugene, OR, USA) or Green-dUTP (Abbott Molecular Inc., Des Plaines, IL, USA), using a Nick Translation Mix (Roche Diagnostics GmbH, Mannheim, Germany). Two-color BAC-FISH was carried out following the procedure described in Sahara et al. (2003) with some modifications. The probe cocktail for one slide (10 μl; 50% deionised formamide and 10% dextran sulphate in 2× SSC buffer) contained 160–450 ng of each labeled BAC probe, 3–5 μg of unlabeled sonicated genomic DNA (extracted from *B. betularia* male by standard phenol/chloroform procedure) used as a species-specific competitor and 25 μg of sonicated salmon sperm DNA (Sigma-Aldrich). For BAC-FISH mapping of 10 genes on the Z chromosome, we used a protocol for reprobing described in (Shibata et al., 2009). In this case, three different BAC probes were mixed in each cocktail as

follows: (1) 160 ng Cy3-labeled BAC, (2) 400 ng Green-labeled BAC and (3) 600 ng two-color labeled BAC (200 ng Cy3 labeled and 400 ng Green labeled). Three BAC probes were hybridized to chromosomes in the first FISH round, and then the preparation was twice reprobed always with three other BAC probes and once reprobed with the last BAC probe. In each BAC-FISH experiment, the probe cocktail was denatured at 95 °C for 5 min. Denaturation of chromosomes was done at 68 °C for 3.5 min in 70% formamide in 2× SSC. The preparations were hybridized with the probe cocktail for 3 days at 37 °C. The slides were then washed for 5 min in 0.1× SSC containing 1% Triton X-100 at 62 °C and for 5 min in phosphate-buffered saline containing 1% Triton X-100 at room temperature. The chromosomes were counterstained with 0.5 μg ml^{-1} of DAPI and mounted in antifade as described in Van't Hof et al. (2008).

Chromosome preparations were observed in a Zeiss Axioplan 2 microscope (Carl Zeiss, Jena, Germany) equipped with appropriate fluorescence filter sets. Black-and-white images were recorded with a cooled F-View CCD camera and captured separately for each fluorescent dye with AnalySIS software, version 3.2 (Soft Imaging System GmbH, Münster, Germany). In BAC-FISH preparations, the images were pseudocolored (light blue for DAPI, green for Alexa Fluor 488-dUTP or Green-dUTP and red for Cy3-dUTP) and superimposed with Adobe Photoshop CS4, version 11.0 (Adobe Systems Inc., San Jose, CA, USA).

To determine the sex chromatin status and thus infer the sex chromosome system in *B. betularia* (see Traut and Marec, 1996), preparations of polyploid nuclei were made from Malpighian tubules of the last instar female and male larvae. The tubules were dissected in physiological solution, briefly fixed in Carnoy's fixative and then stained for 3–5 min with 1.5% lactic acetic orcein. After mounting in the staining solution the preparations were inspected in a light microscope.

RESULTS

Cytogenetic characteristics

The chromosome number was determined by analysis of several tens of mitotic and meiotic chromosome complements on spread preparations from testes. Mitotic spermatogonia showed 62 chromosomes of a holokinetic type (Supplementary Figure S3a). The chromosomes were mostly rod shaped, ∼1–2 μm large, and lacked any morphological signs including primary constrictions (the centromeres) that would allow their identification. In metaphase I spermatocytes, 31 meiotic bivalents were observed (Supplementary Figure S3b). This number was confirmed by counting pachytene bivalents in both sexes (Supplementary Figures S3c and d). Thus, our investigation established that the karyotype of *Biston betularia* consists of $2n = 62$ chromosomes and that the expected number of linkage groups (LGs; that is, the haploid chromosome number) is 31.

Preparations of the Malpighian tubules from larvae of both sexes showed large lobed nuclei. In females, each nucleus displayed a conspicuous spherical heterochromatin body (Supplementary Figure S3e), representing the so-called sex chromatin (W chromatin), which is a characteristic trait in females of the majority of advanced Lepidoptera including Geometridae (Traut and Marec, 1996; Traut et al., 2007). In contrast, somatic polyploid nuclei of males displayed a uniform texture of chromatin grains without any heterochromatin (Supplementary Figure S3f). The presence of sex chromatin in females and its absence in males indicated that *B. betularia* has a WZ/ZZ (female/male) sex chromosome system. In keeping with the sex chromatin status, preparations of pachytene oocytes without a hypotonic treatment showed a WZ bivalent that was discernible from the DAPI-highlighted thread of the W chromosome (Supplementary Figure S3d). However, autosomal bivalents each displayed a regular chromomere pattern that is typical for the synapsis of homologous chromosomes (Traut et al., 2007). In most pachytene oocyte complements seen, the WZ bivalent was one of the longest bivalents. The DAPI-positive staining of the whole W thread indicated that this

chromosome is largely composed of A/T-rich heterochromatin as found in several other lepidopteran species (Fuková et al., 2005; Marec et al., 2010). It should be noted that B. betularia is the first geometrid species in which the sex chromosomes have been identified (cf. Table 1 in Traut et al., 2007).

The linkage map
The map (Figures 1 and 2) is defined by 227 AFLPs (17 of which are 1:2:1 segregating codominant markers, the remainder being paternally 1:1 segregating backcross markers), 156 genes, 6 BACs, 5 microsatellites and the carbonaria phenotype, included as a morphological backcross marker. The ortholog of B. mori kettin was positioned on the Z chromosome by means of BAC-FISH only, as no polymorphisms were available in the mapping family (making a total of 157 genes).

The carbonaria morph maps to chromosome 17 (Van't Hof et al., 2011). Interestingly, none of the major genes known to be involved in insect melanization (aaNat, black, Ddc, Dhpr, ebony, ferritinLC, ferritinHC, henna, laccase2, punch, Pap1, Pap3, Ppo1, Ppo2, tan, TH, yellow, yellow2; see Supplementary Table S1 for full names) map to this chromosome. This confirms earlier findings, based on cosegregation, of a lack of association between genetic variation in these genes and the carbonaria morph (Van't Hof and Saccheri, 2010). In addition to the melanization candidates, vermilion, cinnabar, white, cubitus interruptus, decapentaplegic, patched, Ultrabithorax, wingless, Distal-less, naked cuticle, Notch and scalloped were specifically mapped because of their role in lepidopteran pigment patterning (Carroll et al., 1994; Keys et al., 1999; Weatherbee et al., 1999; Beldade et al., 2002; McMillan et al., 2002; Reed and Serfas, 2004; Reed and Nagy, 2005), but none of these candidates maps to chromosome 17 either.

Consistent with the karyotype, the maternal backcross markers define 31 LGs with the reliability of marker assignment supported by the maximum possible independence test LOD of 10, and (independently of FI markers) paternally segregating markers also separate into 31 LGs with LOD $\geqslant 7$ (average LOD = 9.35). These LOD values are high because (1) low-density regions were specifically targeted with additional genes predicted to fill gaps based on shared synteny with B. mori; (2) 3:1 dominant markers that can diminish the reliability of marker order were excluded; and (3) for AFLP, which typically generate markers with a wide quality range, only those that produced unambiguous genotypes were included. Manual inspection of the recombinants per chromosome ('graphical genotypes' in JoinMap 4, Supplementary Table S2) show marker orders that are consistent with the predictions of the mapping algorithm, most positions being defined by single recombinations per chromosome and a small fraction of double recombinations. The effect of missing values on the linkage map is negligible because only 221 genotypes out of ~28 000 (<1%) remained unscored. A few genotypes suggesting short-range double recombinations (12 out of ~28 000) were presumed to be genotyping errors because nearby crossovers are inconsistent with rules of interference (Muller, 1916). The low number of inconsistencies reflects the stringent marker quality threshold that was used.

The total map length is 1638 cM, with LG size ranging from 28.5 cM (LG20) to 77.2 cM (LG31) and averaging 52.9 cM. Variation in recombinational length estimates among LGs reflects differences in the physical size of the chromosomes, but is also influenced by chromosome-specific recombination frequencies, and by the degree of coverage. The pachytene spreads (Supplementary Figures S3c and d) suggest that there is less variation in chromosome size than implied by LGs, reflecting incomplete coverage for several chromosomes (that

is, some chromosomes may not have markers near one or both telomeric regions). The total recombinational length estimate should therefore be taken as an underestimate, but implies that, on average, a cM is equivalent to ~250–300 kb, assuming a genome size of 500 Mb, the average for geometrid moths (Gregory and Hebert, 2003).

The previously published map of LG17, which includes the carbonaria locus (Van't Hof et al., 2011), has been supplemented with three additional genes, including epidermal stripes and patches (Esp), which extends the LG at the 'lower' end. This is the gene closest to the 'bottom' end of B. mori chromosome 17 and, based on shared synteny, Esp is assumed to also mark the lower end in B. betularia. With sulfamidase close to the other end in B. betularia (demonstrated with BAC-FISH in Van't Hof et al., 2011), LG17 is expected to be fully covered.

Conserved synteny (autosomes)
Every chromosome in the linkage map is anchored with two or more single-copy nuclear genes, allowing presumptive orthologous links to be made to each of the 28 chromosomes in Bombyx mori, the least taxonomically distant reference genome currently available (Xia et al., 2004) (Supplementary Table S1). In all, 25 B. betularia LGs were numbered according to B. mori single-chromosome orthologs. The six B. betularia LGs with orthologous links to the three remaining B. mori chromosomes follow the numbering scheme used for P. xylostella (Baxter et al., 2011), which also possesses 31 chromosomes. Comparison of annotated anchoring markers with the positions of their orthologs in B. mori demonstrates that genes are occasionally rearranged within a chromosome, but rarely translocated to other chromosomes. Although the density of genes across the linkage map is in general too sparse to allow extensive analysis of synteny, chromosomes with two mapped genes can be used to verify co-occurrence and those with three or more can reveal gene order.

Genes that are on the same LG in B. betularia are always grouped together on the same chromosome in B. mori except for rpl36 (B. betularia LG21) and recQ (B. betularia LG29) that are located on B. mori chromosomes 1 (Z) and 17, respectively. Rpl36 also mapped to the LG21 ortholog of B. anynana (Bany LG21), which implies that this gene translocated to the Z chromosome in B. mori, rather than away from it in both B. anynana and B. betularia. Conversely, the location of recQ is similar in B. mori chromosome 17 and its H. melpomene ortholog (Ferguson et al., 2010), suggesting that recQ is relocated in B. betularia relative to its ancestral position. The position of the melanization gene punch is ambiguous in the silkworm genome assembly with the 5′-untranslated region and part of the coding sequence of the GenBank deposited sequence on chromosome 24 and the remainder on chromosome 13. An attempt to map punch in B. mori was unsuccessful (Futahashi et al., 2008), but we nevertheless assume the sequence on chromosome 24 to be in the correct position, implying that B. mori and B. betularia share synteny for this gene. In total, only 2 out of 157 mapped genes reside on different chromosome orthologs between these two species (Supplementary Table S1).

Of the 17 LGs with three or more genes mapped, 12 have gene orders that are identical to B. mori. This includes the more densely mapped 'carbonaria chromosome', LG17, as described in Van't Hof et al. (2011). The exceptions are LG1 (Z), LG11, LG16, LG28 and LG29, in which several genes are rearranged by means of inversions such as (PpoI, PpoII, indy) and (eIF3T, nipsnap, sulfatase) on LG16 (Supplementary Figure S4) and the upper half of LG29 (Figure 3). Other intrachromosomal rearrangements are caused by transpositions, as demonstrated by clathrin HC on LG16 (Supplementary Figure S4) and orb2 on LG11 (Figure 3).
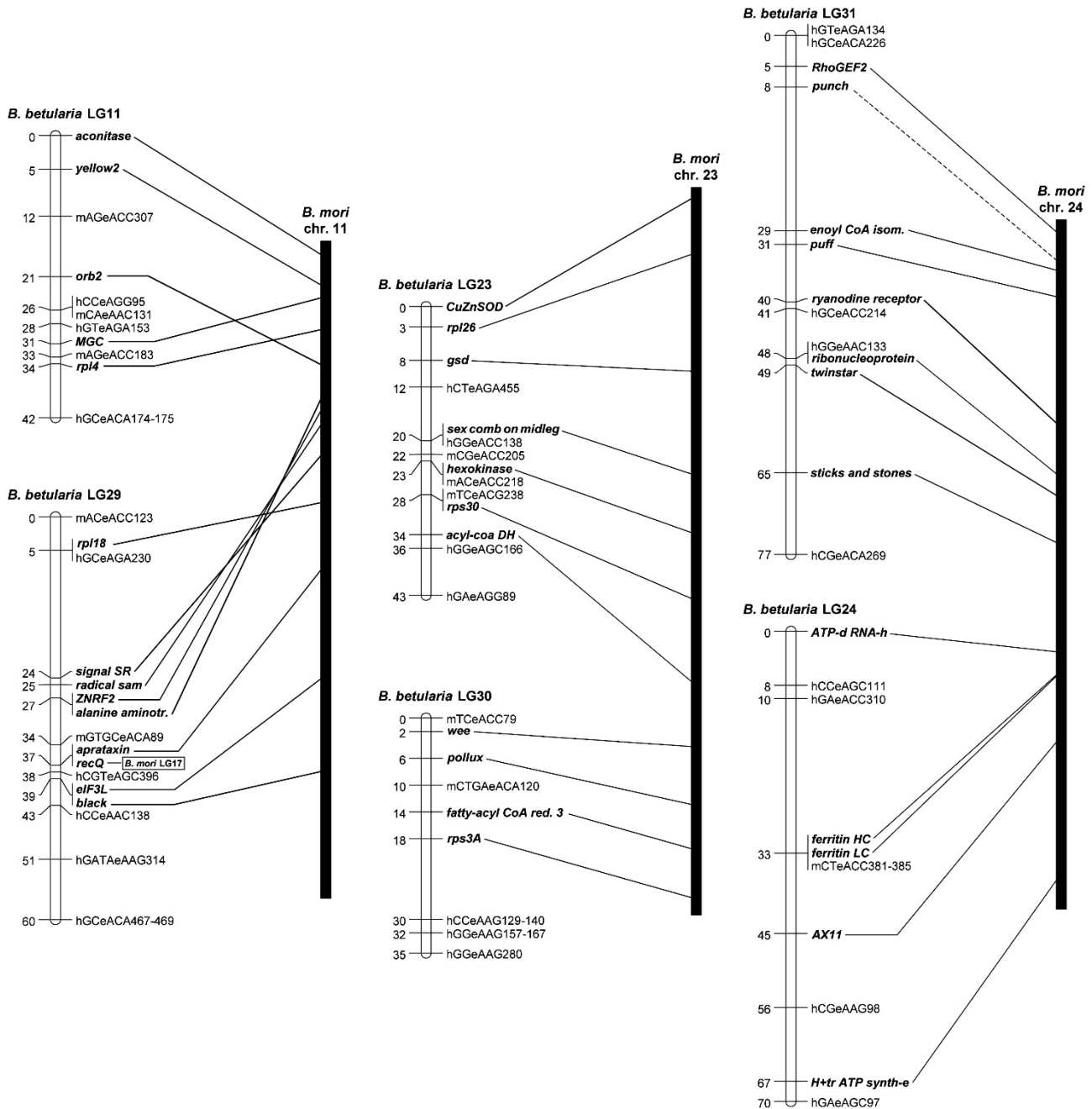
Figure 3 Putatively fused chromosomes in *B. mori*. Six *B. betularia* LGs share gene content with distinct sections on three *B. mori* chromosomes. This suggests three fusion events in *B. mori* that would account for the haploid chromosome number reduction from the presumed ancestral *n* = 31 to *n* = 28. The gene order between *B. betularia* LG11 and its orthologous region in *B. mori* is rearranged such that the upper half of *B. betularia* LG29 is inversed relative to *B. mori*. The remaining LGs (23, 24, 30 and 31) have gene orders identical to *B. mori*. *Punch* is linked with an interrupted line because part of the gene is assigned to a different chromosome in the *B. mori* genome sequence, presumably because of a sequence assembly error.

B. mori chromosome fusions

As *B. mori* has three fewer autosomes than *B. betularia* (28 vs 31) we expected to find genes that are linked in *B. mori* to occur on different LGs in *B. betularia*. Moreover, as 31 is considered to be the ancestral chromosome number for Lepidoptera (Suomalainen, 1969; Lukhtanov, 2000), these patterns suggest fusions between specific ancestral chromosomes. *B. betularia* LGs 11/29, 23/30 and 24/31 had links to *B. mori* chromosomes 11, 23 and 24, respectively. These six LGs were targeted with additional genes to test the initial evidence for three

chromosome fusions in the ancestors of *B. mori*, and to determine the orientation and approximate position of the fusion points (Figure 3). The links between orthologous genes clearly demonstrate how orthologous regions of the two species match, although *B. betularia* LG11 and LG29 experienced a transposition and an inversion, respectively.

Sex chromosomes

The Z-chromosome map (LG1) is defined by 11 anchoring orthologs of *B. mori* Z-linked genes, which include 3 melanization genes
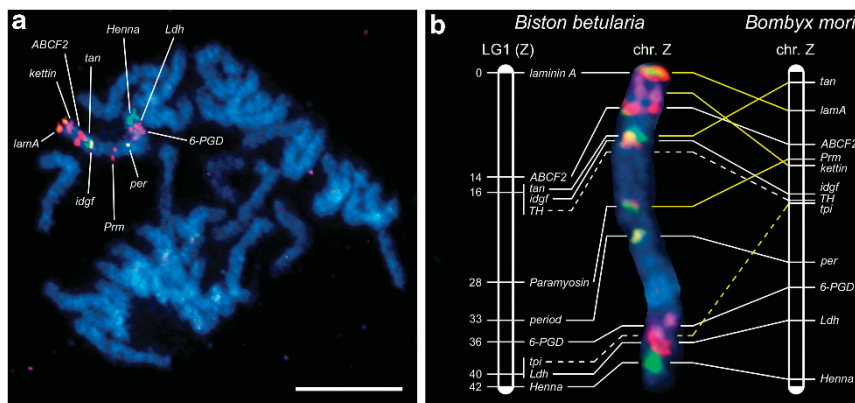
Figure 4 BAC-FISH and linkage mapping of the Z chromosome in *B. betularia*. Meiotic chromosomes from males were counterstained with DAPI (light blue). The BAC probes were labeled with Cy3-dUTP, Green-dUTP and with a combination of both Cy3-dUTP and Green-dUTP, respectively. Hybridization signals indicate the physical positions of the 10 loci: *laminin A* (orange), *kettin* (violet), *ABCF2* (red), *tan* (green), *idgf* (yellowish to orange), *Paramyosin* (red and green), *period* (yellow), *6-PGD* (violet), *Ldh* (red) and *Henna* (green). (**a**) Pachytene spermatocyte complement with the ZZ bivalent identified by hybridization signals of 10 Z-linked gene-containing BAC probes. Scale bar = 10 μm. (**b**) Chromosomal localization and synteny of 10 Z-linked loci between *B. betularia* and *B. mori* Z chromosomes. The recombinational positions of genes mapped to *B. betularia* linkage group 1 (LG1; left-hand vertical bar) are related to their physical positions in the BAC-FISH-painted Z chromosome of *B. betularia* (middle image) and in a scheme of the *B. mori* Z chromosome (right-hand vertical bar) by interconnecting lines; white lines indicate preserved gene order and yellow lines rearranged gene order between the two species.

(*Henna*, *tan* and *TH*), 3 other enzyme-coding genes (*6-PGD*, *Ldh* and *tpi*) and 5 protein-coding genes without enzymatic function (*ABCF2*, *idgf*, *laminin A*, *Paramyosin* and *period*). The gene *kettin*, encoding insect muscle protein, was screened for nearly 10 kb of genomic sequence, but did not contain polymorphisms in the mapping family and was positioned by means of BAC-FISH instead (Figure 4).

Out of 13 Z-linked *B. mori* gene orthologs, 12 mapped to the Z chromosome of *B. betularia* (Figures 4a and b), with *rpl36* being autosomal in *B. betularia*. However, although Z linkage of most genes is conserved between these species, gene order within the Z chromosome differs in some respect. Comparison of the Z-chromosome map of *B. betularia*, constructed by linkage and BAC-FISH mapping, with the physical map of *B. mori* Z chromosome (Figure 4b) revealed conserved gene order in a Z-chromosome segment between the *period* and *Henna* loci (that is, *period*, *6-PGD*, *Ldh* and *Henna*), except the *tpi* locus that is positioned near *Ldh* in a subterminal region of the *B. betularia* Z, but located in the middle part of the *B. mori* Z. In contrast, the upper half of the Z chromosome showed altered order of four genes, namely *laminin A*, *kettin*, *tan* and *Paramyosin*. The observed pattern of divergence can be hypothetically accounted for by two inversions in the upper part of the chromosome and two transpositions (Supplementary Figure S5).

Distances between some genes differ considerably in the linkage and physical (BAC-FISH) maps of the *B. betularia* Z chromosome (Figure 4b). In particular, the *ABCF2* locus is separated from the terminal *laminin A* locus by a large gap covering one-third of the LG, whereas BAC-FISH positions these loci relatively close to each other. In contrast, the physical distance between the *period* and *6-PGD* loci is much larger than the cM distance. The discrepancies between physical and cM distances suggest a recombination hot spot near the upper end of the Z chromosome and a reduced recombination region at the lower end.

Exceptionally for a Z-linked gene, *laminin A* also occurs on the W chromosome, as revealed by the observation that an exonic SNP within the gene was heterozygous A/C (not hemizygous) in both the mother and father. The same heterozygous genotype was shared by approximately half of the sons, the other half being homozygous for

Table 1 Genotype frequencies of the *laminin A* WZ single-nucleotide polymorphism (SNP) in sons and daughters inherited from mother A_ZC_W × father A_ZC_Z

Genotype	Sons (ZZ)	Daughters (WZ)
A/A	17 (A_ZA_Z)	0
A/C	17 (A_ZC_Z)	17 (A_ZC_W)
C/C	0	22 (C_ZC_W)

their mother's Z allele, whereas (WZ) daughters were either heterozygous or homozygous for the alternate allele (Table 1). Additionally, the *laminin A* alleles cosegregated with the Z-linked markers at the 'upper' end of the chromosome. This pattern of inheritance is only possible for a Z-linked/W-linked locus. Full-length gene identity and functionality of the W representative remain to be confirmed but the available sequence lacks characteristics of a pseudogene; in particular, the W and Z copies are identical over the full length of a 566-bp PCR fragment apart from the synonymous SNP (that is, identical amino-acid sequence) and there are no frameshifts or premature stop codons relative to the *B. mori* sequence. Moreover, the observation that the W-linked SNP allele also occurs on the Z chromosome highlights that there is no unique feature in the W sequence.

Four BAC clones containing *laminin A* sequence were identified and used as probes for BAC-FISH to verify the presence of the gene on the W chromosome. Each BAC clone provided a different hybridization pattern on spread chromosomes of pachytene oocytes (Figures 5a–f). The clone 2I7 hybridized to the ends of both the W and Z chromosomes (Figures 5a and b), indicating remnants of sequence homology at one end of both sex chromosomes. Two clones provided a discrete hybridization signal at the end of one sex chromosome only, either Z (clone 1A21) or W (clone 8M12) (Figures 5e and f). This suggests that clone 1A21 is derived from the Z chromosome, whereas clone 8M12 originates from the W chromosome. It should be noted that in the two BAC-FISH experiments, pairing orientation of the sex chromosomes in the
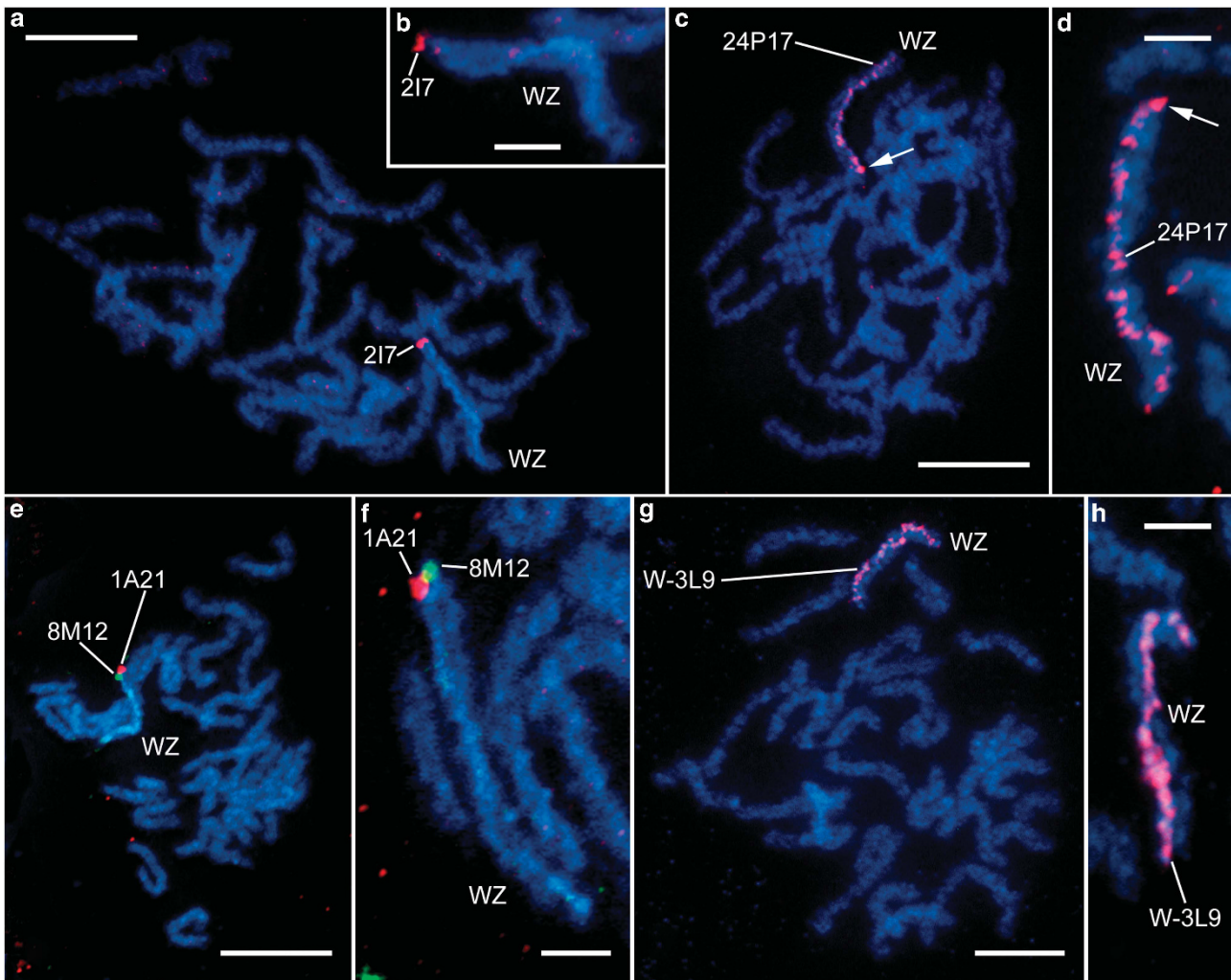
Figure 5 BAC-FISH mapping on both sex chromosomes (W and Z) in female *B. betularia.* Hybridization patterns of four laminin *A*-containing clones and one W-linked AFLP marker-derived clone on W and Z chromosomes. (**a**, **c**, **e**, **g**) Pachytene oocyte complements. (**b**, **d**, **f**, **h**) Details of WZ bivalents. Chromosomes were counterstained with DAPI (light blue); BAC probes were labeled with Cy3-dUTP (red) or Alexa Fluor 488-dUTP (green). The sex chromosome bivalent (WZ) is discernible from a DAPI-positive heterochromatic thread of the W chromosome and also by hybridization signals of the BAC probes. (**a**, **b**) *laminin A* BAC probe 2I7 (red) shows hybridization signals at the end of both the W and Z chromosomes. (**c**, **d**) *laminin A* BAC probe 24P17 (red) shows scattered hybridization signals along the whole W-chromosome thread; a stronger signal at one end of the W chromosome (arrow) possibly indicates a physical position of the BAC clone. (**e**, **f**) *laminin A* BAC probes 1A21 (red) and 8M12 (green) indicate physical positions of *laminin A* at the ends of the Z and W chromosomes, respectively. (**g**, **h**) AFLP-derived BAC probe W-39L (red) paints the entire W-chromosome thread. Scale bars = 10 μm (**a**, **c**, **e**, **g**) or 3 μm (**b**, **d**, **f**, **h**).

WZ bivalent was identical in all pachytene complements examined, with either two 2I7 signals or 1A21 plus 8M12 signals always at one end of the bivalent, which suggests specific pairing. Unlike the other *laminin A*-containing clones, which showed discrete signals, the clone 24P17 painted the whole W chromosome in a scattered manner (Figures 5c and d). This suggests that this particular clone is rich in repetitive sequences spread throughout the W chromosome. For a comparison, we also mapped two W-derived BAC clones, W-3L9 and 38L11, identified with W-specific AFLP markers. The clone W-3L9 showed uniform, almost continuous painting of the whole W chromosome (Figures 5g and h), whereas clone 38L11 showed scattered signals on the whole W chromosome (not shown), similar to the *laminin A*-containing clone 24P17.

One end sequence of BAC 8M12 contains two additional *laminin A* exons (and four introns) upstream of the exon used to map family no. 16. The mother of family no. 67, whose daughters were used for BAC construction, has a FI SNP in one of these introns. One of the alleles is inherited by family no. 67 daughters exclusively (11 daughters, 16 sons), allowing unambiguous assignment of BACs 8M12 and 24P17 as W derived and 1A21 and 2I7 as Z derived, respectively, confirming the BAC-FISH results. The presence of additional *laminin A* exons on a W-BAC (8M12) suggests that the W chromosome contains a substantial section of the *laminin A* gene. Obtaining full sequence from W-BACs would reveal whether the W representative of *laminin A* is complete and intact, but is not yet available owing to the large genomic size of the gene (66 and 21 kb in *B. mori* and *Danaus plexippus*, respectively).

DISCUSSION

In the age of whole-genome sequence assembly, linkage mapping remains an important tool for analyzing the genesis of genome architecture. The *Biston betularia* linkage map, the first for a

geometrid moth, illustrates the value of broadening the taxonomic coverage to reconstructing chromosome dynamics (inversion, transposition, fission and fusion). Such a phylogenetic perspective is necessary to distinguish ancestral from derived states, but has until recently been limited by the availability of a single lepidopteran (*B. mori*) whole-genome reference sequence or comprehensive linkage map against which to make comparisons. (Two additional lepidopteran whole genomes are now available for *H. melpomene* (The *Heliconius* Genome Consortium, 2012) and *D. plexippus* (Zhan *et al.*, 2011), although the latter's scaffolds are not assigned to chromosomes). A case in point is the reconstruction of the direction of a translocation event that led to *rpl36* being Z linked in *B. mori* but autosomal in *Ostrinia nubilalis*. On the basis of pairwise comparison, Kroemer *et al.* (2011) suggested that Z linkage is the ancestral state. However, a three-way comparison of the position of *rpl36* between *B. mori* (Z), *B. betularia* (LG21) and *B. anynana* (mapped for this study to the ortholog of *B. mori* chromosome 21) provides stronger support for the alternative hypothesis. In contrast, the same approach, using *B. mori*, *B. betularia* and *Heliconius melpomene* (Ferguson *et al.*, 2010), applied to *recQ*, the only other gene (out of 157 mapped genes) whose LG differed between *B. betularia* and *B. mori*, implies that the position in *B. mori* is the ancestral state.

Phylogenetically informed reconstruction is also essential for properly accounting for differences in chromosome number among species by means of fissions or fusions. In this respect, *B. mori* is not an ideal reference for Lepidoptera, having three fewer chromosomes than the accepted ancestral 31 (Suomalainen, 1969; Lukhtanov, 2000). The problem of reconstructing the sequence of fission/fusion events is compounded in comparisons between independently derived chromosome complements, such as between *B. mori* and *B. anynana* (Beldade *et al.*, 2009) or *B. mori* and *H. melpomene* (Pringle *et al.*, 2007). We have established that *B. betularia* has 31 chromosomes, and that three pairs of these chromosomes are orthologous to three *B. mori* chromosomes (*Bb*LG11/29-*Bm*11; *Bb*LG23/30-*Bm*23; and *Bb*LG24/31-*Bm*24). This pattern is entirely consistent with the RAD (restriction site associated DNA)-based, *n* = 31, linkage map for *P. xylostella* (Baxter *et al.*, 2011), but is additionally based on gene order and not just content. Moreover, the large phylogenetic distance between *B. betularia* and *P. xylostella* (*P. xylostella* is a distant outgroup to *B. betularia* and *B. mori*) effectively establishes that the difference in chromosome number is the result of three fusions in the *B. mori* lineage, rather than the same three chromosome fissions, independently, in the *B. betularia* and *P. xylostella* lineages. A recent comparison between *H. melpomene* and *B. mori* genomes arrived at the same conclusion (The *Heliconius* Genome Consortium, 2012).

The *B. betularia* linkage map provides further evidence for deep conservation of synteny and gene order within the Lepidoptera. The degree of shared synteny among representatives of the Lepidoptera, currently all within the infraorder Heteroneura (diverged ~160 Myr ago; Labandeira *et al.*, 1994), is comparable to that found within the single, 65-Myr-old genus *Drosophila* (Singh *et al.*, 2009). By comparison, the genomes of two distantly related dipterans, *Anopheles gambiae* and *Drosophila melanogaster*, have been extensively reshuffled during 250 Myr of divergence (Bolshakov *et al.*, 2002). The chromosome scale patterns in Lepidoptera may mask higher rates of gene order rearrangement at the finest scale (that is, neighboring genes), as revealed by interspecific comparisons of contiguous sequence (d'Alençon *et al.*, 2010; Conceição *et al.*, 2011), possibly associated with holocentrism and transposable elements.

According to the minimum-interaction hypothesis, karyotypes consisting of high number of small elements, as is typical for

Lepidoptera, suffer fewer deleterious rearrangements such as translocations because the likelihood of interaction among chromosomes during meiosis is reduced (Imai *et al.*, 1986; Hirai *et al.*, 1994). The centromere-drive hypothesis (nonrandom segregation based on centromere activity differences in asymmetric female meiosis) was formulated to explain karyotype evolution in mammals (Pardo-Manuel de Villena and Sapienza, 2001). It has been speculated that holokinetic chromosomes evolved as suppressors of meiotic drive (Talbert *et al.*, 2009; Zedek and Bureš, 2012). Thus, the holokinetic structure of lepidopteran chromosomes might contribute to their evolutionary stability. This is a paradox because holokinetic chromosomes are expected to facilitate karyotype evolution via fusion and fission by reducing the risk of the formation of dicentric and/or acentric chromosomes (Wrensch *et al.*, 1994; Carpenter *et al.*, 2005). These hypotheses suggest that small and numerous holokinetic chromosomes predispose lepidopteran genomes to structural stasis.

The *B. betularia* Z map extends previous efforts of mapping genes on non-silkworm lepidopteran Z chromosomes (Beldade *et al.*, 2009; Kroemer *et al.*, 2011) and suggests that the synteny conserving mechanism is weaker for the Z chromosome than the autosomes, with five gene order differences and one gene content difference (*rpl36*) compared with *B. mori*. Taking into account the available data on Z-linked genes in other Lepidoptera (Supplementary Table S3), the emerging picture is that although its gene content is highly conserved, the Z may be characterized by higher rates of intrachromosomal rearrangement. This may reflect its disproportionate role in adaptive evolution and speciation, the so-called Large-Z effect (Prowell, 1998; Naisbit *et al.*, 2002; Presgraves, 2002). The rearranged order of Z-linked genes between *B. mori* and *B. betularia* suggests two inversions and two transpositions during divergence of their Z chromosomes from a common ancestor (Supplementary Figure S5). Inversions in particular can play an important role in speciation by restricting meiotic recombination in the whole inverted region, protecting favorable haplotypes from being broken down, and allowing the accumulation of genetic incompatibilities in the loci involved (Faria and Navarro, 2010). The loss of Z-chromosome colinearity in comparison with the autosomes, recently found between scaffolds of *B. mori* and *D. plexippus*, supports this hypothesis (Zhan *et al.*, 2011).

A special feature of lepidopteran genomes is the W chromosome. Being inherited via females, the complete absence of recombination has resulted in its independent evolution accompanied with extensive genetic erosion and accumulation of repetitive DNA sequences (Marec *et al.*, 2010). So far, the repetitive nature of the W chromosome, together with a lack of functional genes, have obstructed detailed structural analysis, even in *B. mori* (Abe *et al.*, 2010). The relatively high number of W-linked AFLPs (18) and one gene (*laminin A*) in *B. betularia* allowed the identification of W-linked BACs as a tool to examine the characteristics of a lepidopteran W chromosome further. Three W-linked BAC clones (two AFLP derived and one *laminin A* derived) did not localize discrete sites with BAC-FISH but painted the whole W chromosome (Figures 5c, d, g and h), similar to W-derived BAC clones in *B. mori* (Sahara *et al.*, 2003). Although unsuitable for positional mapping, these hybridization patterns do add further evidence of the repetitive nature and uniformity of the lepidopteran W chromosome (Marec *et al.*, 2010). The chromosome-wide hybridization signals, especially of 3L9, suggest the presence of repetitive elements that exist (almost) exclusively on the W chromosome.

The identification of a W homolog of the *laminin A* gene, revealed by nonhemizygous segregation (Table 1), and its localization to the

end of the *B. betularia* W chromosome (Figures 5a, b, e and f) is a significant addition to the very small number of confirmed protein-coding sequences on lepidopteran W chromosomes (although transcription of the W representative remains to be confirmed). As the Z-linked *laminin A* is also located at the end of the Z chromosome, the W-linked copy of this gene could have arisen and be maintained by ectopic recombination between subtelomeric (terminal) regions resulting in gene conversion (Schlecht *et al.*, 2004; Linardopoulou *et al.*, 2005). Besides the putative zinc-finger protein (*Fem*) in the silkworm (Suzuki, 2010), the only W-linked genes with known sequence are in *Antheraea pernyi*, which has two variants of the *period* gene on the W chromosome, one producing a truncated protein and the other antisense RNA transcript (Gotter *et al.*, 1999). These W-linked *period* variants might originate from the functional *period* gene, which is also Z linked in *A. pernyi*, as in *B. mori* and *B. betularia* (Figure 4b).

Valid synteny comparisons are dependent on linkage map quality, which in turn depends on the choices made for experimental design and analysis. The linkage mapping strategies for *P. xylostella* (Baxter *et al.*, 2011) and *B. anynana* (Beldade *et al.*, 2009), based on RAD and EST polymorphisms, respectively, used single or multiple small families (20 and 22 offspring), thus prioritizing the number of markers over resolution. In the *P. xylostella* map, marker order is reliable but markers are clustered in ∼5 cM intervals. In the *B. anynana* map, the reliability of gene order may be partially compromised in LGs lacking polymorphic anchoring markers to join the maps from each of the 12 mapping families. RAD mapping (using a single family of 43 offspring) and *de novo* genome assembly were powerfully combined to order sequence scaffolds onto chromosomes in *H. melpomene* (The *Heliconius* Genome Consortium, 2012). For the *B. betularia* map, we opted for a single relatively large family (73 offspring) at modest marker density (a second family was additionally used for LG17). Marker number was supplemented to fill specific gaps and for LGs of particular interest (Z, *carbonaria* locus, *B. mori* fusions) using genes from orthologous *B. mori* chromosomes. The resulting chromosome maps are highly reliable in terms of marker order and, conditional on variable marker density, are relatively well resolved. Parallel patterns of orthology for *B. betularia* and *P. xylostella* against *B. mori*, across all of the 31 LGs, strongly suggest that the genome structure of these two species closely resembles the ancestral chromosome print for the Lepidoptera (direct comparison between *B. betularia* and *P. xylostella* cannot be made because only uniref90 blastx positives are available for *P. xylostella*). Furthermore, this implies that the strong mode of $n = 31$ chromosomes in Lepidoptera is maintained through stasis rather than a dynamic fusion/fission equilibrium.

As the number of lepidopteran linkage maps and mapped genomes expands, it will be possible to reconstruct the architecture of ancestral lepidopteran genomes at increasingly fine taxonomic and genetic scales. The significance of this exercise lies not only in a better understanding of chromosome dynamics *per se*, but is highly relevant to analyzing the role of genetic rearrangements and LG conservation in phenotypic evolution and reproductive isolation. An elegant example of this is the inversion polymorphism defining the wing-pattern supergene (*P*) in *Heliconius numata*, which acts to suppress low-fitness (nonmimic) recombinant phenotypes (Joron *et al.*, 2011). Through an unlikely coincidence, the *carbonaria* locus controlling melanism in *B. betularia* is orthologous to *P*, suggesting the existence of a deeply conserved switch of pattern regulation in this region. The positional information provided by the *B. betularia* linkage map for the majority of genes known to be directly involved in melanin

biosynthesis (and some additional patterning genes) sets the genomic context for unraveling the cascade of interactions underlying this widespread polymorphism.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Abe H, Fujii T, Tanaka N, Yokoyama T, Kakehashi H, Ajimura M *et al.* (2008). Identification of the female-determining region of the W chromosome in *Bombyx mori*. *Genetica* **133**: 269–282.

Abe H, Mita K, Yasukochi Y, Oshiki T, Shimada T (2005). Retrotransposable elements on the W chromosome of the silkworm, *Bombyx mori*. *Cytogenet Genome Res* **110**: 144–151.

Abe H, Fujii T, Shimada T (2010). Sex chromosomes and sex determination in *Bombyx mori*. In: Goldsmith MR, Marec F (eds). *Molecular Biology and Genetics of the Lepidoptera*. CRC Press: Boca Raton, FL, USA. pp 65–87.

Andolfatto P, Scriber JM, Charlesworth B (2003). No association between mitochondrial DNA haplotypes and a female-limited mimicry phenotype in *Papilio glaucus*. *Evolution* **57**: 305–316.

Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD *et al.* (2011). Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS ONE* **6**: e19315.

Beldade P, Brakefield PM, Long AD (2002). Contribution of *Distal-less* to quantitative variation in butterfly eyespots. *Nature* **415**: 315–318.

Beldade P, Rudd S, Gruber J, Long A (2006). A wing expressed sequence tag resource for *Bicyclus anynana* butterflies, an evo-devo model. *BMC Genomics* **7**: 130.

Beldade P, Saenko SV, Pul N, Long AD (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.

Bolshakov VN, Topalis P, Blass C, Kokoza E, Della TA, Kafatos FC *et al.* (2002) A comparative genomic analysis of two distant Diptera, the fruit fly, *Drosophila melanogaster*, and the malaria mosquito, *Anopheles gambiae*. *Genome Res* **12**: 57–66.

Carpenter JE, Bloem S, Marec F (2005). Inherited sterility in insects. In: Dyck VA, Hendrichs J, Robinson AS (eds). *Sterile Insect Technique. Principles and Practice in Area-Wide Integrated Pest Management*. Springer: Dordrecht. pp 115–146.

Carroll S, Gates J, Keys D, Paddock S, Panganiban G, Selegue J *et al.* (1994). Pattern formation and eyespot determination in butterfly wings. *Science* **265**: 109–114.

Conceição IC, Long AD, Gruber JD, Beldade P (2011). Genomic sequence around butterfly wing development genes: annotation and comparative analysis. *PLoS ONE* **6**: e23778.

Cook LM (2003). The rise and fall of the *carbonaria* form of the peppered moth. *Q Rev Biol* **78**: 399–417.

d'Alençon E, Sezutsu H, Legeai F, Permal E, Bernard-Samain S, Gimenez S *et al.* (2010). Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements. *Proc Natl Acad Sci USA* **107**: 7680–7685.

Daly D, Waltham K, Mulley J, Watts PC, Rosin A, Kemp SJ *et al.* (2004). Trinucleotide microsatellite loci for the peppered moth (*Biston betularia*). *Mol Ecol Notes* **4**: 179–181.

Dopman EB, Bogdanowicz SM, Harrison RG (2004). Genetic mapping of sexual isolation between E and Z pheromone strains of the European corn borer (*Ostrinia nubilalis*). *Genetics* **167**: 301–309.

Faria R, Navarro A (2010). Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol Evol* **25**: 660–669.

Ferguson L, Lee SF, Chamberlain N, Nadeau N, Joron M, Baxter S *et al.* (2010). Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the *HmYb/Sb* locus. *Mol Ecol* **19**: 240–254.

Fujii T, Abe H, Shimada T (2010). Molecular analysis of sex chromosome-linked mutants in the silkworm *Bombyx mori*. *J Genet* **89**: 365–374.

Fuková I, Nguyen P, Marec F (2005). Codling moth cytogenetics: karyotype, chromosomal location of rDNA, and molecular differentiation of sex chromosomes. *Genome* **48**: 1083–1092.

Futahashi R, Sato J, Meng Y, Okamoto S, Daimon T, Yamamoto K et al. (2008). *yellow* and *ebony* are the responsible genes for the larval color mutants of the silkworm *Bombyx mori*. *Genetics* **180**: 1995–2005.

Gotter AL, Levine JD, Reppert SM (1999). Sex-linked *period* genes in the silkmoth, *Antheraea pernyi*: implications for circadian clock regulation and the evolution of sex chromosomes. *Neuron* **24**: 953–965.

Gregory TR, Hebert PDN (2003). Genome size variation in lepidopteran insects. *Can J Zool* **81**: 1399–1405.

Hartzer KL, Zhu KY, Baker JE (2005). Phenoloxidase in larvae of *Plodia interpunctella* (Lepidoptera: Pyralidae): molecular cloning of the proenzyme cDNA and enzyme activity in larvae paralyzed and parasitized by *Habrobracon hebetor* (Hymenoptera: Braconidae). *Arch Insect Biochem Physiol* **59**: 67–79.

Heckel DG, Gahan LJ, Liu Y-B, Tabashnik BE (1999). Genetic mapping of resistance to *Bacillus thuringiensis* toxins in diamondback moth using biphasic linkage analysis. *Proc Natl Acad Sci USA* **96**: 8373–8377.

Hirai H, Yamamoto MT, Ogura K, Satta Y, Yamada M, Taylor RW 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.

Imai HT, Maruyama T, Gojobori T, Inoue Y, Crozier RH (1986). Theoretical bases for karyotype evolution. 1. The minimum-interaction hypothesis. *Am Nat* **128**: 900–920.

Jiggins CD, Mavarez J, Beltrán M, McMillan WO, Johnston JS, Bermingham E (2005) A genetic linkage map of the mimetic butterfly *Heliconius melpomene*. *Genetics* **171**: 557–570.

Johnson MS, Turner JRG (1979). Absence of dosage compensation for a sex-linked enzyme in butterflies (*Heliconius*). *Heredity* **43**: 71–77.

Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR et al. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**: 203–206.

Kandul NP, Lukhtanov VA, Pierce NE (2007). Karyotypic diversity and speciation in *Agrodiaetus* butterflies. *Evolution* **61**: 546–559.

Kapan DD, Flanagan NS, Tobler A, Papa R, Reed RD, Acevedo Gonzalez J et al. (2006). Localization of Müllerian mimicry genes on a dense linkage map of *Heliconius erato*. *Genetics* **173**: 735–757.

Kawamura N (1988). The egg size determining gene, *Esd*, is a unique morphological marker on the W chromosome of *Bombyx mori*. *Genetica* **76**: 195–201.

Keys DN, Lewis DL, Selegue JE, Pearson BJ, Goodrich LV, Johnson RL et al. (1999). Recruitment of a *hedgehog* regulatory circuit in butterfly eyespot evolution. *Science* **283**: 532–534.

Kroemer J, Coates B, Nusawardani T, Rider S, Fraser L, Hellmich R (2011) A rearrangement of the Z chromosome topology influences the sex-linked gene display in the European corn borer, *Ostrinia nubilalis*. *Mol Genet Genomics* **286**: 37–56.

Labandeira CC, Dilcher DL, Davis DR, Wagner DL (1994). Ninety-seven million years of angiosperm-insect association: paleobiological insights into the meaning of coevolution. *Proc Natl Acad Sci USA* **91**: 12278–12282.

Lees DR (1981). Industrial melanism: genetic adaptation of animals to air pollution. In: Bishop JA, Cook LM (eds). *Genetic Consequences of Man Made Change*. Academic Press: London. pp 129–176.

Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ (2005). Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437**: 94–100.

Lukhtanov VA (2000). Sex chromatin and sex chromosome systems in nonditrysian Lepidoptera (Insecta). *J Zool Syst Evol Res* **38**: 73–79.

Majerus MEN (1998). *Melanism: Evolution in Action*. Oxford University Press: Oxford.

Marec F, Sahara K, Traut W (2010). Rise and fall of the W chromosome in Lepidoptera. In: Goldsmith MR, Marec F (eds). *Molecular Biology and Genetics of the Lepidoptera*. CRC Press: Boca Raton, FL, USA. pp 49–63.

McMillan WO, Monteiro A, Kapan DD (2002). Development and evolution on the wing. *Trends Ecol Evol* **17**: 125–133.

Muller HJ (1916). The mechanism of crossing-over II. *Am Nat* **50**: 284–305.

Murakami A, Imai HT (1974). Cytological evidence for holocentric chromosomes of the silkworms, *Bombyx mori* and *B. mandarina* (Bombycidae, Lepidoptera). *Chromosoma* **47**: 167–178.

Naisbit RE, Jiggins CD, Linares M, Salazar C, Mallet J (2002). Hybrid sterility, Haldane's rule and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics* **161**: 1517–1526.

Narita S, Nomura M, Kato Y, Fukatsu T (2006). Genetic structure of sibling butterfly species affected by *Wolbachia* infection sweep: evolutionary and biogeographical implications. *Mol Ecol* **15**: 1095–1108.

Nokkala S (1987). Cytological characteristics of chromosome behaviour during female meiosis in *Sphinx ligustri* L. (Sphingidae, Lepidoptera). *Hereditas* **106**: 169–179.

Pardo-Manuel de Villena F, Sapienza C (2001). Female meiosis drives karyotypic evolution in mammals. *Genetics* **159**: 1179–1189.

Presgraves DC (2002). Patterns of postzygotic isolation in Lepidoptera. *Evolution* **56**: 1168–1183.

Pringle EG, Baxter SW, Webster CL, Papanicolaou A, Lee SF, Jiggins CD (2007). Synteny and chromosome evolution in the Lepidoptera: evidence from mapping in *Heliconius melpomene*. *Genetics* **177**: 417–426.

Prowell DP (1998). Sex linkage and speciation in Lepidoptera. In: Howard DJ, Berlocher SH (eds). *Endless Forms: Species and Speciation*. Oxford University Press: Oxford. pp 309–319.

Putnam AS, Scriber JM, Andolfatto P (2007). Discordant divergence times among Z-chromosome regions between two ecologically distinct swallowtail butterfly species. *Evolution* **61**: 912–927.

Reed RD, Nagy LM (2005). Evolutionary redeployment of a biosynthetic module: expression of eye pigment genes *vermilion*, *cinnabar*, and *white* in butterfly wing development. *Evol Dev* **7**: 301–311.

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.

Regier JC (2007). Protocols, concepts, and reagents for preparing DNA sequencing templates. *wwwumbiumdedu/users/jcrlab/PCR_primerspdf*.

Sahara K, Marec F, Traut W (1999). TTAGG telomeric repeats in chromosomes of some insects and other arthropods. *Chromosome Res* **7**: 449–460.

Sahara K, Yoshido A, Kawamura N, Ohnuma A, Abe H, Mita K 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.

Sahara K, Yoshido A, Marec F, Fuková I, Zhang H-B, Wu C-C 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.

Schlecht HB, Lichten M, Goldman ASH (2004). Compartmentalization of the yeast meiotic nucleus revealed by analysis of ectopic recombination. *Genetics* **168**: 1189–1203.

Shibata F, Sahara K, Naito Y, Yasukochi Y (2009). Reprobing multicolor FISH preparations in lepidopteran chromosome. *Zool Sci* **26**: 187–190.

Singh ND, Larracuente AM, Sackton TB, Clark AG (2009). Comparative genomics on the *Drosophila* phylogenetic tree. *Annu Rev Ecol Evol Syst* **40**: 459–480.

Sturtevant AH (1913). The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J Exp Zool* **14**: 43–59.

Suomalainen E (1969). Chromosome evolution in the Lepidoptera. *Chromosomes Today* **2**: 132–138.

Suomalainen E, Brown KSJ (1984). Chromosome number variation within *Philaethria* butterflies (Lepidoptera: Nymphalidae, Heliconiini). *Chromosoma* **90**: 170–176.

Suzuki MG (2010). Sex determination: insights from the silkworm. *J Genet* **89**: 357–363.

Suzuki MG, Shimada T, Kobayashi M (1999). *Bm kettin*, homologue of the *Drosophila kettin* gene, is located on the Z chromosome in *Bombyx mori* and is not dosage compensated. *Heredity* **82**: 170–179.

Talbert PB, Bayes JJ, Henikoff S (2009). Evolution of centromeres and kinetochores: A two-part fugue. In: De Wulf P, Earnshaw WC (eds). *The Kinetochore: From Molecular Discoveries to Cancer Therapy*. Springer: New York. pp 193–229.

The *Heliconius* Genome Consortium (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**: 94–98.

The International Silkworm Genome Consortium (2008). The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem Mol Biol* **38**: 1036–1045.

Traut W (1977). A study of recombination, formation of chiasmata and synaptonemal complexes in female and male meiosis of *Ephestia kuehniella* (Lepidoptera). *Genetica* **47**: 135–142.

Traut W, Marec F (1996). Sex chromatin in Lepidoptera. *Q Rev Biol* **71**: 239–256.

Traut W, Sahara K, Marec F (2007). Sex chromosomes and sex determination in Lepidoptera. *Sex Dev* **1**: 332–346.

Traut W, Scholz D (1978). Structure, replication and transcriptional activity of the sex-specific heterochromatin in a moth. *Exp Cell Res* **113**: 85–94.

Van Nieukerken EJ, Kaila L, Kitching IJ, Kristensen NP, Lees DC, Minet J et al. (2011). Order Lepidoptera Linnaeus, 1758. *Zootaxa* **3148**: 212–221.

Van't Hof AE, Edmonds E, Dalíková M, Marec F, Saccheri IJ (2011). Industrial melanism in British peppered moths has a singular and recent mutational origin. *Science* **332**: 958–960.

Van't Hof AE, Marec F, Saccheri IJ, Brakefield PM, Zwaan BJ (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 AE, Saccheri IJ (2010). Industrial melanism in the peppered moth is not associated with genetic variation in canonical melanisation gene candidates. *PLoS ONE* **5**: e10889.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23**: 4407–4414.

Wang B, Porter AH (2004). An AFLP-based interspecific linkage map of sympatric, hybridizing *Colias* butterflies. *Genetics* **168**: 215–225.

Wang J, Xia Q, He X, Dai M, Ruan J, Chen J et al. (2005). SilkDB: a knowledgebase for silkworm biology and genomics. *Nucleic Acids Res* **33**: D399–D402.

Weatherbee SD, Nijhout HF, Grunert LW, Halder G, Galant R, Selgue J et al. (1999). *Ultrabithorax* function in butterfly wings and the evolution of insect wing patterns. *Curr Biol* **9**: 109–115.

White MJD (1973). *Animal Cytology and Evolution*. Cambridge University Press: Cambridge.

Winter CB, Porter AH (2010). AFLP linkage map of hybridizing swallowtail butterflies, *Papilio glaucus* and *Papilio canadensis*. *J Hered* **101**: 83–90.

Wrensch DL, Kethley JB, Norton RA (1994). Cytogenetics of holokinetic chromosomes and inverted meiosis: keys to the evolutionary success of mites, with generalizations on eukaryotes. In: Houck M (ed.). *Mites: Ecological and Evolutionary Analyses of Life-History Patterns*. Chapman & Hall: New York. pp 282–342.

Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B et al. (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* **306**: 1937–1940.

Yamamoto K, Nohata J, Kadono-Okuda K, Narukawa J, Sasanuma M, Sasanuma S *et al.* (2008). A BAC-based integrated linkage map of the silkworm, *Bombyx mori Genome Biol* **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, 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.

Yasukochi Y, Tanaka-Okuyama M, Shibata F, Yoshido A, Marec F, Wu C *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.

Yoshido A, Sahara K, Marec F, Matsuda Y (2011). Step-by-step evolution of neo-sex chromosomes in geographical populations of wild silkmoths, *Samia cynthia* ssp. *Heredity* **106**: 614–624.

Zedek F, Bureš P (2012). Evidence for centromere drive in the holocentric chromosomes of *Caenorhabditis*. *PLoS ONE* **7**: e30496.

Zhan S, Merlin C, Boore JL, Reppert SM (2011). The monarch butterfly genome yields insights into long-distance migration. *Cell* **147**: 1171–1185.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)