Karyotypes versus genomes: the nymphalid butterflies *Melitaea cinxia, Danaus plexippus* and *D. chrysippus*

Walther Traut[1], Virpi Ahola[2], David A. S. Smith[3], Ian J. Gordon[4] and Richard H. ffrench-Constant[5]

[1)] Institut für Biologie, Zentrum für medizinische Struktur- und Zellbiologie, Universität zu Lübeck, Ratzeburger Allee 160, D-23538 Lübeck, Germany; traut@bio.uni-luebeck.de

[2)] Department of Biosciences, University of Helsinki, FI-00014 Helsinki, Finland.

[3)] Natural History Museum, Eton College, Windsor SL4 6DW, UK

[4)] BirdLife International, Africa Partnership Secretariat, Box 3502-00100, Nairobi, Kenya

[5)] Centre for Ecology and Conservation, University of Exeter, Penryn Campus, Penryn, Cornwall TR10 9FE, UK

Abstract

The number of sequenced Lepidopteran genomes is increasing rapidly. Rarely, however, do the corresponding assemblies represent whole chromosomes and generally they also lack the highly repetitive W sex chromosome. Knowledge of karyotypes can facilitate genome assembly and further our understanding of sex chromosome evolution in Lepidoptera. Here, we desribe the karyotypes of the Glanville fritillary *Melitaea cinxia* (n = 31), the Monarch *Danaus plexippus* (n = 30), and the African Queen *D. chrysippus* (2n = 60 or 59, depending on the source population). We show by FISH that the telomeres are of the $(TTAGG)_n$ type, as found in most insects. *M. cinxia* and *D. plexippus* have 'conventional' W chromosomes, which are heterochromatic in meiotic and somatic cells. In *D. chrysippus*, the W is inconspicuous. Neither telomeres nor W chromosomes are represented in the published genomes of *M. cinxia* and *D. plexippus*. Representation analysis in sequenced *D. chrysippus* female and male genomes detected an evolutionarily old autosome-Z chromosome fusion in *Danaus*. Conserved synteny of whole chromosomes, so called 'macro synteny', in Lepidoptera permitted us to identify the chromosomes involved in this fusion. An additional and more recent sex chromosome fusion was found in *D. chrysippus* by karyotype analysis and classical genetics. In a hybrid population between two subspecies, *D. c. chrysippus* and *D. c. dorippus*, the W chromosome was fused to an autosome that carries a wing colour locus. Thus, cytogenetics and and the present state of genomes complement one another to reveal the evolutionary history of the species.

Introduction

There is currently a rapid growth in both the number and quality of insect genomes published, owing to parallel improvements in sequencing technology. 'Next generation sequencing' has moved on to 3rd and 4th generation sequencing, facilitating high quality assemblies of even the most complex genomes (Heather and Chain, 2016). For such sequence data, knowledge of the karyotype is necessary to predict the behaviour of specific loci in transmission genetics and is indeed also a prerequisite for final genome reconstruction and linkage mapping. In vertebrates and flowering plants, mitotic karyotypes are often easy to obtain and are therefore well characterised and readily available as a basis for molecular research. In animal groups with lower DNA content and numerous chromosomes, however, like butterflies and moths (Lepidoptera), karyotyping is more difficult and data scarcer. For the majority of butterflies and moths, only chromosome counts from male metaphase I have been published. To facilitate the assembly of linkage groups and indeed chromosomes we describe and update descriptions of karyotypes from three model butterflies: the Glanville fritillary *Melitaea cinxia*, the Monarch *Danaus plexippus*, and the African Queen *Danaus chrysippus*.

We specifically examine the behaviour of sex chromosomes as they often pose problems in genome assembly. Like birds it is the female sex that is heterogametic in

Lepidoptera. The sex chromosome constitution in lepidopteran females is mostly WZ, but Z0 systems and systems with multiple sex chromosomes like W_1W_2Z, WZ_1Z_2, and up to $W_{1-3}Z_{1-6}$ do occur (Sahara et al., 2012; Šíchová et al., 2016). The constitution of the homogametic males is ZZ, or occasionally $Z_1Z_1Z_2Z_2$. The W chromosome is heterochromatic in somatic cells of most Lepidoptera (Traut and Marec, 1996). Like the Y chromosome in mammals it is rich in repetitive sequences (Traut et al., 2013). Genome projects, therefore, usually omit the W chromosomes by sequencing male genomes in Lepidoptera, as the repetitive W chromosome is too difficult to assemble. We show here that, at the present stage, combined efforts in karyotype and genome analysis reveal a sequence of sex chromosome evolution events in *Danaus* species, starting wih a hypothetical common ancestor which like *M. cinxia* has the basic lepidopteran chromosome number of n = 31.

Materials and Methods

Animals. *Melitaea cinxia* larvae and pupae were from broods kept at the Lammi biological station of the University of Helsinki (Lammi, Finland) (Ojanen et al., 2013). The broods were founded by larvae from the Åland Islands, Finland. *D. plexippus* pupae were provided by butterfly farms (Stratford-upon-Avon Butterfly Farm, UK, and VlindersParadijs, Havelte, Netherlands) and came originally from Costa Rica. *D. chrysippus* sources were broods from Southeast Asia, supplied by Michael Boppré (Freiburg, Germany), from Israel, supplied by Dubi Benyamini (Tel Aviv, Israel), and from various places in Africa bred by Steve Collins (Nairobi, Kenya), I.J.G., and D.A.S.S.

Cytogenetic preparations. Malpighian tubules were fixed for a few seconds in ME (methanol-acetic acid, 3 : 1), stained and mounted in lactic acetic orcein.

Male metaphase I chromosomes were prepared from spermatogenic cysts of last instar larvae and early pupae. Fixation with Carnoy`s fluid (ethanol : chloroform : acetic acid, 6 : 3 : 1) or ME (methanol : acetic acid, 3 : 1) was followed by transfer to slides, a short exposure (~10 sec) to 60% acetic acid, air drying, and staining with either lactic acetic orcein or DAPI (4'6-diamidino-2-phenylindole).

Pachytene preparations were done as described in Traut (1976) with a few modifications. In brief, ovary anlagen from larvae or tips of ovarioles dissected from pupae and adults were fixed in Carnoy, transferred to a drop of 60% acetic acid on a slide and minced with sharp tungsten needles. On a heating plate (50-60°), the drop was moved around by tilting the slide, and discarded after no more than 30 sec. Chromosomes were stained with DAPI and mounted in Antifade (0.233 g 1,4-diazabicyclo(2,2,2)octane, 1 ml 0.2M Tris–HCl, pH 8.0, and 9 ml glycerol).

Telomere FISH was performed with an insect-specific telomere probe, $(TTAGG)_n$, produced by template-free PCR (primers: TAGGTTAGGTTAGGTTAGGT and CTAACCTAACCTAACCTAAC) according to Sahara et al. (1999) and labelled with Orange-dUTP. Hybridisation was performed for 2 days at 37° in 12 µl/slide hybridisation mix containing ~10 ng probe DNA, 2.5 µg salmon sperm DNA, 6 µl deionised formamide, and 6 µl 20% dextran sulfate in 4xSSC. Stringent washes in 0.1xSSC/1% Triton X100 at 62° were followed by staining in DAPI and mounting in Antifade.

Representation analysis. Libraries of 250 bp Illumina reads from *D. chrysippus* were screened with protein queries (Supplementary Material Table s1) by TBLASTN for coverage of the respective genes. The libraries consisted of (1) 170,635,319 reads from a female of the Kitengela (Kenya) population, (2) 162,558,554 reads from a female of the Watamu (Kenya) population, and (3) 165,283,721 reads from a male of the Watamu (Kenya) population. A threshold of E = 1e-20 was set. Only data sets with >50 hits in all three libraries were

included in the analysis. Ratios of female/male hits were corrected for differences in library size.

Results

Melitaea cinxia

The chromosome number n = 31 was originally described by Federley (1938) from metaphase I of oocytes. We confirm this chromosome number from both male metaphase I, male metaphase II (Fig. 1a, b) and the female pachytene (Fig. 1c). It is the supposed basic chromosome number in Lepidoptera (Ahola et al., 2014; Suomalainen, 1969). The pachytene bivalents vary little in length distribution, the shortest and longest bivalents differing by no more than a factor of ~2. They display patterns of chromomeres and interchromomeres and terminate in $(TTAGG)_n$ telomeres which are the conventional type of telomeres in most insect groups (Sahara et al., 1999; Vitkova et al., 2005). The WZ bivalent is conspicuous among the bivalents. Only one of the partner chromosomes, the presumed Z, displays an autosome-like chromomere-interchromomere pattern while the other one, the presumed W, is heterochromatic with condensed chromatin, highly fluorescent in DAPI-stained preparations (Fig. 1c WZ). Depending on the stage, the W chromosome is stretched and either pairs quasi-normally with the Z chromosome (Fig. 2a) or is more or less condensed (Fig. 2b, d) or even collapsed to a ball with the Z chromosome coiled around the ball (Fig. 2d). In polyploid somatic cells such as those from Malpighian tubules, the W chromosome forms a sex-specific heterochromatin body, which is present in females but absent in males (Fig. 3).

The genome of *M. cinxia* has been successfully assembled to scaffolds that have been placed to 31 chromosomes (Ahola et al., 2014) and is available at EnsemblMetazoa (http://metazoa.ensembl.org/). It has been established from males. Hence, the W is not included. BLASTN searches with an insect telomere query, $(TTAGG)_{30}$, did not find any matching sequence of that size in the scaffolds. Only a short stretch of $(TTAGG)_5$ turned up in the search. This means that the scaffolds of the present version have not yet reached the chromosome ends, probably due to the high content of repeats in subtelomeric regions.

Danaus plexippus

The chromosome number n = 30 was originally published by Rao and Murty (1975), but in a preprint publication, Hamm (2017) maintains that the Indian species investigated by Rao and Murty (1975) is in fact *Danaus genutia*, whereas *D. plexippus* from Kansas, USA, has a different chromosome number, n = 28. In another preprint publication, Mongue et al. (2016) find a haploid chromosome number of n = 30 in *D. plexippus*. Our counts from male metaphase I and female pachytene clearly give n = 30 for *D. plexippus* from Costa Rica (Fig. 4). Racial differences in chromosome numbers are well-known but the data from Kansas should be reconfirmed.

The 29 autosomal bivalents are inconspicuous in the female pachytene and post-pachytene stages, while the WZ bivalent is quite conspicuous, having a heterochromatic W chromosome and an autosome-like Z chromosome. Telomeres are of the insect-type $(TTAGG)_n$ (Figs. 4, 5). Bivalent lengths in the pachytene come in a smoothly graded series; the longest and shortest bivalent differ by a factor of ~2. The W chromosome consists of two segments, a longer one with strong and a shorter one with inconspicuous DAPI fluorescence (Fig. 4b, Fig. 5a) or is condensed to an oblong mass in pachytene (Fig. 5b). In post-pachytene, the W chromosome appears to be resolved into a chain of condensed beads (Fig. 5c, d). Somatic cells display typical sex chromatin which forms a single large female-specific heterochromatin body in the highly polyploid cells of Malpighian tubules (Fig. 6).

The *D. plexippus* genome was published by Zhan et al. (2011) and is publicly available at MonarchBase (http://monarchbase.umassmed.edu/). In contrast to *M. cinxia*, females were used to produce the sequencing libraries. Therefore, the database should contain

also W chromosomal scaffolds. However, scaffolds have not yet been assigned to chromosomes. Like in *M. cinxia*, telomeres are not included in the scaffolds; only three very short stretches of $(TTAGG)_n$ have been detected by BLASTN searches somewhere within scaffolds.

Danaus chrysippus

Here we confirm that the chromosome number for *D. chrysippus* is also n = 30 (de Lesse and Condamin, 1962; Gupta, 1964; Smith et al., 2016) for males from Southeast Asia, Israel, Ghana, Watamu (Kenya) and Kitengela (Kenya) (examples shown in Fig. 7a and b). Both females and males of the subspecies *D. c. dorippus* from Watamu (Kenya) as well as of the subspecies *D. c. chrysippus* from Israel have 30 bivalents. Pachytene complements show all 30 bivalents with a chromomere-interchromomere pattern and insect-type telomeres (Fig. 7c). There are no abrupt length differences, the shortest and longest bivalents differ by a factor of ~2.5. The W chromosome is not visibly heterochromatic, the WZ bivalent could not be identified. Even comparative genomic hybridisation (CGH) with differently coloured female and male whole genomic probes did not reveal the W chromosome (not shown). Cell nuclei from Malpighian tubules contained small heterochromatic bodies in both sexes (Fig. 8). None could be safely associated with the female sex alone. Similar cases had been found in other Lepidoptera species (Traut and Marec, 1996). The heterochomatin bodies were probably formed by aggregation of small autosomal segments of heterochromatin in these highly polyploid nuclei.

Females from a population at Kitengela (near Nairobi in Kenya) form a special case. The population is part of a large hybrid zone between the subspecies *D. c. dorippus* and *D. c. chrysippus*. Females from this population display a sex chromosome trivalent, neo-W/Z_1Z_2, besides 28 bivalents in meiosis and are infected with a male-killing bacterium, *Spiroplasma ixodetis* (Smith et al., 2016). The neo-W is interpreted as the result of a fusion between the original W chromosome and an autosome, carrying the phenotypically defined wing colour gene C (Smith et al., 2010). We show here the trivalents formed in pachytene complements (Fig. 9). The two parts of the neo-W chromosome can be distinguished by their pairing with the shorter Z_1 and the longer Z_2 and by a different chromomere pattern. Z_2 is considered to be the former autosome; the chromomere patterns of Z_2 and the respective part of the neo-W are largely similar and, therefore, suggest a relatively short span of divergent evolution. The pattern of Z_1, however, is very different from that of the neo-W segment paired with it. They are probably the original Z and W chromosomes which have spent a long history of divergent evolution.

The genomes of two *D. chrysippus* females (Watamu and Kitengela populations) and a male (Watamu population) were sequenced. Scaffolds have not yet been assigned to chromosomes. Only few scaffolds of >1 kbp size contained >100 bp long stretches of telomere sequence at their ends, indicating true telomere regions (5 in Kitengela female, 6 in Watamu female, and 4 in Watamu male). The expectation was much higher: 60 in the Watamu male (both ends of Z and 29 autosomes) and 62 in the Watamu female (both ends of W, Z, and 29 autosomes) and the Kitengela female (both ends of neoW, Z_1, Z_2, and 28 autosomes).

We made use of the read libraries for a representation analysis of protein-coding genes. As probes we selected *B. mori* protein sequences which had formerly been used by Sahara et al. (2013) and Van't Hof et al. (2013) for demonstrating conserved synteny of chromosomes from *Helicoverpa armigera* and *Biston betularia*, repectively, with chromosomes of *B. mori*. Coverage of genes from all but one autosome was about the same in female and male libraries, the female/male ratio being equal to ~1.0. Female/male ratios of Z chromosomal genes and *B. mori* chromosome #16 genes, in contrast, were equal to ~0.5 (supplementary material Table s1). This shows that the chromosome #16 homologue in *D.*

chrysippus like the Z chromosome is present in only one copy in females but two copies in males. It is either a second Z chromosome or fused to the original Z chromosome (Fig. 10 'hypothetical intermediate' and 'alternative intermediate'). Both, female and male karyotypes in non-hybrid populations of *D. chrysippus* consist of 30 bivalents in meiosis (see Figure 7). Therefore, the presence of two Z chromosomes can be excluded. The *D. chrysippus* Z chromosome is apparently composed of two moieties, one with conserved synteny to the Z chromosome of other lepidopteran species and the other with conserved synteny to *M. cinxia* chromosome #21, *B. mori*, *H. armigera*, *B. betularia* chromosome #16, and *H. melpomene* chromosome #2 (Table 1, Fig. 10 '*D. plexippus*, *D. chrysippus* (Watamu)'). In a similar approach, representation of scaffolds in read libraries, with other species of *Danaus,* Walters and Mongue (2016) and Mongue et al. (2016) suggested a fusion event in an ancestor of the genus as the chromosome number of *Danaus* species is n = 30, one fewer than the supposed basic number in Lepidoptera, n = 31 (Ahola et al., 2014; Suomalainen, 1969).

The W chromosome presents a theoretical problem if the postulated Z-A fusion has indeed occurred in some ancestor of *Danaus*. In that case, we should have expected a W_1W_2 / neo-Z sex chromosome constitution as a result of the fusion (Fig. 10 'hypothetical intermediate'). We do not see that in *D. plexippus* or *D. chrysippus*. Hence, a fusion of the original W with the homologue of that autosome must have occurred in the evolutionary past of the genus, or else one of the hypothetical Ws was lost since (Fig. 10 'hypothetical intermediate' and 'alternative intermediate'). If the former autosome was not lost, it must have undergone considerable decay since, as it does not turn up in representation analysis.

A more recent fusion event, the W chromosome fused to an autosome, in the hybrid population of *D. chrysippus* at Kitengela, however, evidently produced a second Z chromosome (see Fig. 9, Fig. 10 '*D. chrysippus* (Kitengela)'). Our representation analysis did not identify the source of the second Z chromosome, the one carrying the otherwise autosomal C locus. Coverage ratios of genes from Kitengela and Watamu females were similar for all chromosomes (see supplementary material Table s1). Presumably, the W-autosome fusion was a phylogenetically rather recent event. Genes in the fused autosome, therefore, did not change or decay sufficiently yet to affect the representation analysis.

Discussion

Sequencing and electronically annotating a genome is no longer a difficult job, as one can see from the growing numbers of draft genomes e.g. in LepBase (http:// lepbase.org). Knowledge of the karyotype as the framework for these molecular data, however, is often incomplete or missing. Worse still, and even though butterflies are large and charismatic, targets have even been mis-identified leading to further confusion around cytogenetics and genomics. Without this precise inter relationship between taxonomy, cytogenetics and genomics we will end up with numerous draft genomes whose contigs and scaffolds bear no relationship to proper linkage maps or the chromosomes with which they are associated. This paper therefore attempts to tie down the karyotypes for three key nymphalid genomes which we hope will then act as models for further butterfly genomics.

Besides these general issues of taxonomy and cytogenetics, there are at least two problematic chromosomal regions for Lepidoptera genomics: chromosome ends and W sex chromosomes. As we show here, chromosome ends are rarely represented or are indeed totally missing in scaffold libraries, probably due to the repetive nature of subtelomeric regions and telomeres. These problems will, however, presumably be solved soon, when long reads as those from PacBio or Nanopore sequencing come into more general use. On the other hand, centromere regions, which are a problem e.g. in vertebrates, appear not to be critical for the assembly of Lepidoptera genomes, probably owing to the largely holocentric nature of Lepidoptera chromosomes (Wolf, 1996).

W chromosomes pose a substantial problem for DNA sequence assembly. They are usually heterochromatic and highly repetive in Lepidoptera (Sahara et al., 2012; Traut and Marec, 1996). To this end the W chromosomes of neither *M. cynthia* nor *B. mori* have been assembled to date (Ahola et al., 2014; Mita et al., 2004; Xia et al., 2004). In fact because of the expected problems with high W repeat content only male butterfly genomes are often sequenced. The W chromosome of the meal moth, *Ephestia kuehniella*, was isolated by microdissection and subjected to high-thoughput sequencing, but no anchor genes were found and the assembly produced only short contigs apparently not covering the whole W chromosome (Traut et al., 2013).

It is clear from our results that two of the species investigated here, *M. cinxia* and *D. plexippus*, have this type of W chromosome. It forms a sex-specific heterochromatin body, the W chromatin, in somatic cells. It is a highly conspicuous body in polyploid cells such as those of Malpighian tubules (see Figures 3 and 6). The W is even seen as a heterochromatic thread or 'blob' when synapsed to the Z chromosome in the pachytene stage of meiosis (see Figures 1c, 2, 4b, 5). The chances of proper reconstruction of the W from its likely highly repetitive DNA sequences remain low. However, the W chromosome of the third species, *D. chrysippu*s, in contrast, does not appear heterochromatic. It is inconspicuous in meiotic as well as somatic cells (see Figures 7-9) and has apparently lost most or all heterochromatic segments, supposed to have been present in the common ancestor of *D. chrysippus* and *D. plexippus*. It may in fact have retained some remnants from the autosome it was associated with in its evolutionary past (this paper: *Danaus chrysippus*).

Acknowledgements
We thank Dubi Benyamini (Tel Aviv, Israel), Michael Boppré (Freiburg, Germany), and Steve Collins (Nairobi, Kenia) for providing us with broods of *D. chrysippus*. The technical assistance of Conni Reuter, Ulrike Eickhoff (Lübeck, Germany), Suvi Ikonen, and Laura Häkkinen (Helsinki, Finland) is greatfully acknowledged.

References

Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, et al: The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nature Communications 5:4737 (2014).doi:10.1038/ncomms5737

de Lesse H, Condamin M: Formules chromosomiques de quelques lépidoptères rhopalocères du Sénégal. Bull Inst Fond Afr Noire 24:464–473 (1962).

Federley H: Chromosomenzahlen finnlandischer Lepidopteren. I. Rhopalocera. Hereditas 24:397–464 (1938).

Gupta Y: Chromosome studies of some Indian Lepidoptera. Chromosoma 15:540–561 (1964).

Hamm CA: Chromosome number of the monarch butterfly, *Danaus plexippus* (Linnaeus 1758) and the Danainae. bioRxiv preprint (2017). http://dx.doi.org/10.1101/107144.

Heather JM, Chain B: The sequence of sequencers: The history of sequencing DNA. Genomics 107:1-8 (2016).

Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, et al: The genome sequence of silkworm, *Bombyx mori*. DNA Res 11:27-35 (2004).

Mongue AJ, Nguyen P, Volenikova A, Walters JR: Neo-sex chromosomes in the Monarch butterfly, Danaus plexippus. bioRxiv preprint (2016). http://dx.doi.org/10.1101/036483.

Ojanen S, Nieminen M, Meyke E, Pöyry J, Hanski I: Long-term metapopulation study of the Glanville fritillary butterfly (*Melitaea cinxia*): Survey methods, data management, and long-term population trends. Ecology and Evolution:3713–3737 (2013).

Rao N, Murty AS: Chromosome number of *Danaus plexippus* (Lepidoptera: Danaidae). Current Science 44:894–895 (1975).
Sahara K, Marec F, Traut W: TTAGG telomeric repeats in chromosomes of some insects and other arthropods. Chromosome Research 7:449 - 460 (1999).
Sahara K, Yoshido A, Shibata F, Fujikawa-Kojima N, Okabe T, et al: Fish identification of Helicoverpa armigera and Mamestra brassicae chromosomes by BAC and fosmid probes. Insect Biochemistry and Molecular Biology 43:644-653 (2013).
Šíchová J, Ohno M, Dincă V, Watanabe M, Sahara K, Marec F: Fissions, fusions, and translocations shaped the karyotype and multiple sex chromosome constitution of the northeast-asian wood white butterfly, Leptidea amurensis. Biological Journal of the Linnean Society 118:457–471 (2016).
Sahara K, Yoshido A, Traut W: Sex chromosome evolution in moths and butterflies. Chromosome Research 20:83-94 (2012).
Smith DAS, Gordon IJ, Allen JA: Reinforcement in hybrids among once isolated semispecies of Danaus chrysippus (L.) and evidence for sex chromosome evolution. Ecological Entomology 35 (Suppl. 1), 77-89 (2010). DOI: 10.1111/j1365-2311.2009.01143.x.
Smith DAS, Gordon IJ, Traut W, Herren J, et al: A neo-W chromosome in a tropical butterfly links colour pattern, male-killing, and speciation. Proc R Soc B 283:20160821 (2016).
Suomalainen E: Chromosome evolution in the Lepidoptera. Chromosomes Today 2:132–138 (1969).
Traut W: Pachytene mapping in the female silkworm *Bombyx mori* l. (Lepidoptera). Chromosoma 58:275-284 (1976).
Traut W, Marec F: Sex chromatin in Lepidoptera. Quarterly Rev Biol 71:239 - 256 (1996).
Traut W, Vogel H, Glöckner G, Hartmann E, Heckel DG: High-throughput sequencing of a single chromosome: A moth W chromosome. Chromosome Research 21:491-505 (2013).
Van't Hof AE, Nguyen P, Dalíková M, Edmonds N, Marec F, et al: Linkage map of the peppered moth, Biston betularia (Lepidoptera, Geometridae): A model of industrial melanism. Heredity 110:283–295 (2013).
Walters JR, Mongue AJ: A neo-sex chromosome in the monarch butterfly, Danaus plexippus. bioRxiv preprint (2016). http://dx.doi.org/10.1101/036483
Wolf KW: The structure of condensed chromosomes in mitosis and meiosis of insects. International Journal of Insect Morphology and Embryology 25:37-62 (1996).
Vitkova M, Kral J, Traut W, Zrzavy J, Marec F: The evolutionary origin of insect telomeric repeats, (TTAGG)n. Chromosome research 13:145-156 (2005).
Xia Q, Zhou Z, Lu C, Cheng D, Dai F, et al: A draft sequence for the genome of the domesticated silkworm (Bombyx mori). Science 306:1937-1940 (2004).
Zhan S, Merlin C, Boore JL, Reppert SM: The monarch butterfly genome yields insights into long-distance migration. Cell 147:1171–1185 (2011).

Figures

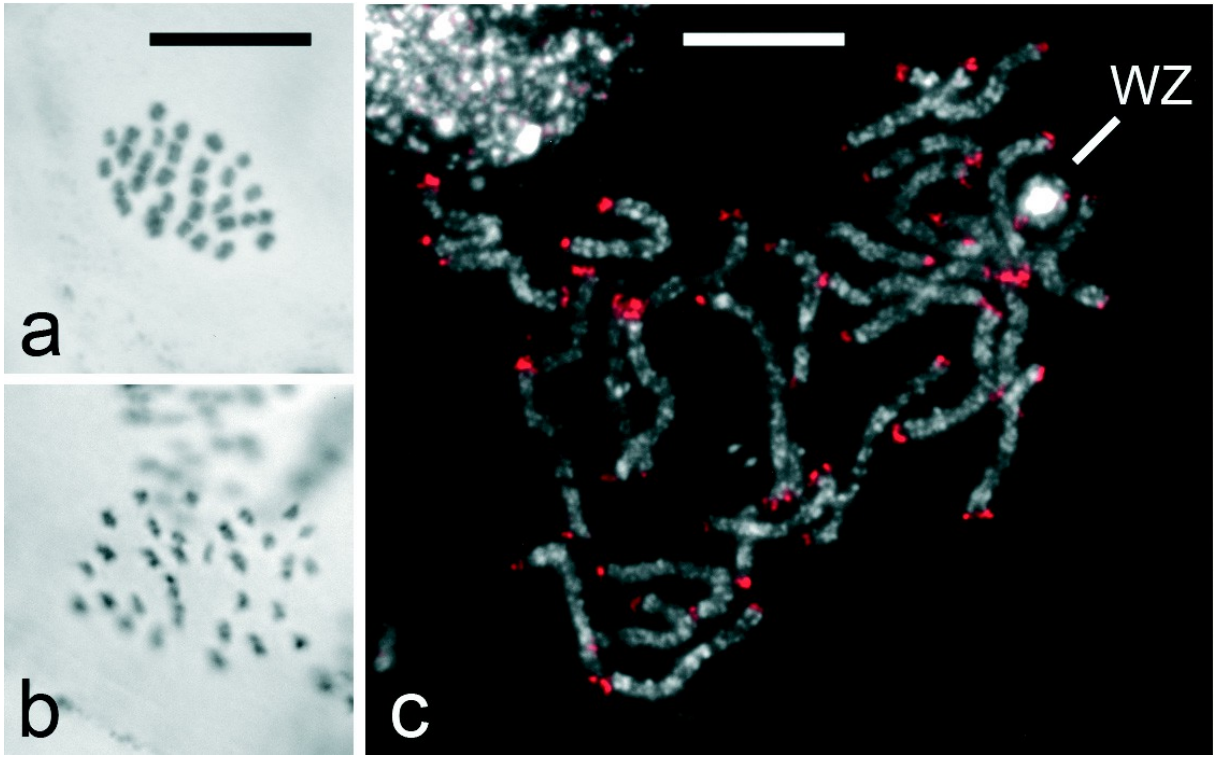

Fig. 1 *Melitaea cinxia*. Male metaphase I (a), male metaphase II (b) both orcein-stained, and female pachytene (c) DAPI-stained (white), telomere-FISH signals (red). WZ sex chromosome bivalent, bar 10 µm.

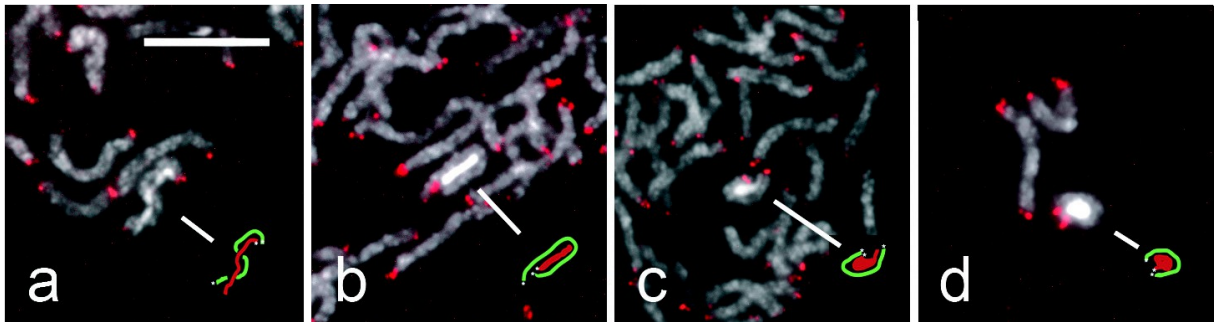

Fig. 2 *Melitaea cinxia*. WZ bivalents in the female pachytene stage, W chromosome compaction varies from extended (a) to more and more compacted (b, c, d). Note the missing telomere signal at one end of the W chromosome. DAPI staining (white), telomere-FISH (red). Diagrams interpreting the WZ pairing: red W, green Z chromosome, white asterisks telomeres. Bar 10 µm.

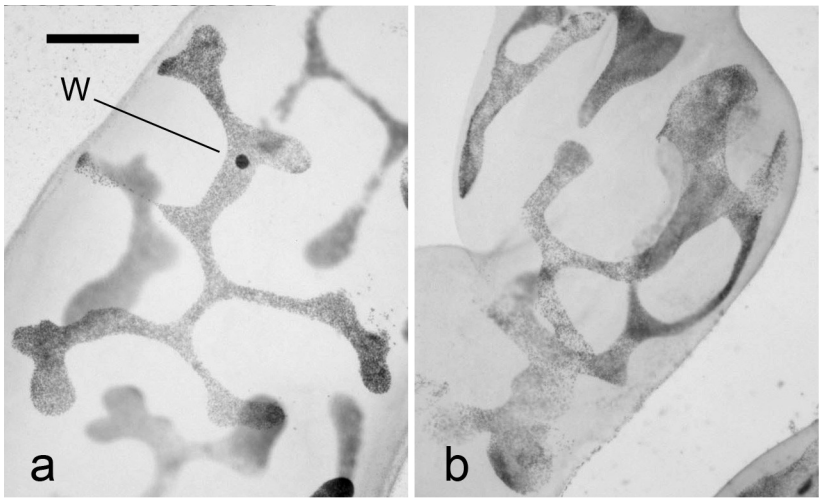

Fig. 3 *Melitaea cinxia*. Sex chromatin (W) in highly polyploid, lobed nuclei from Malpighian tubules. Sex chromatin is present in females (a) and absent in males (b). Orcein-stained, bar 50 μm.

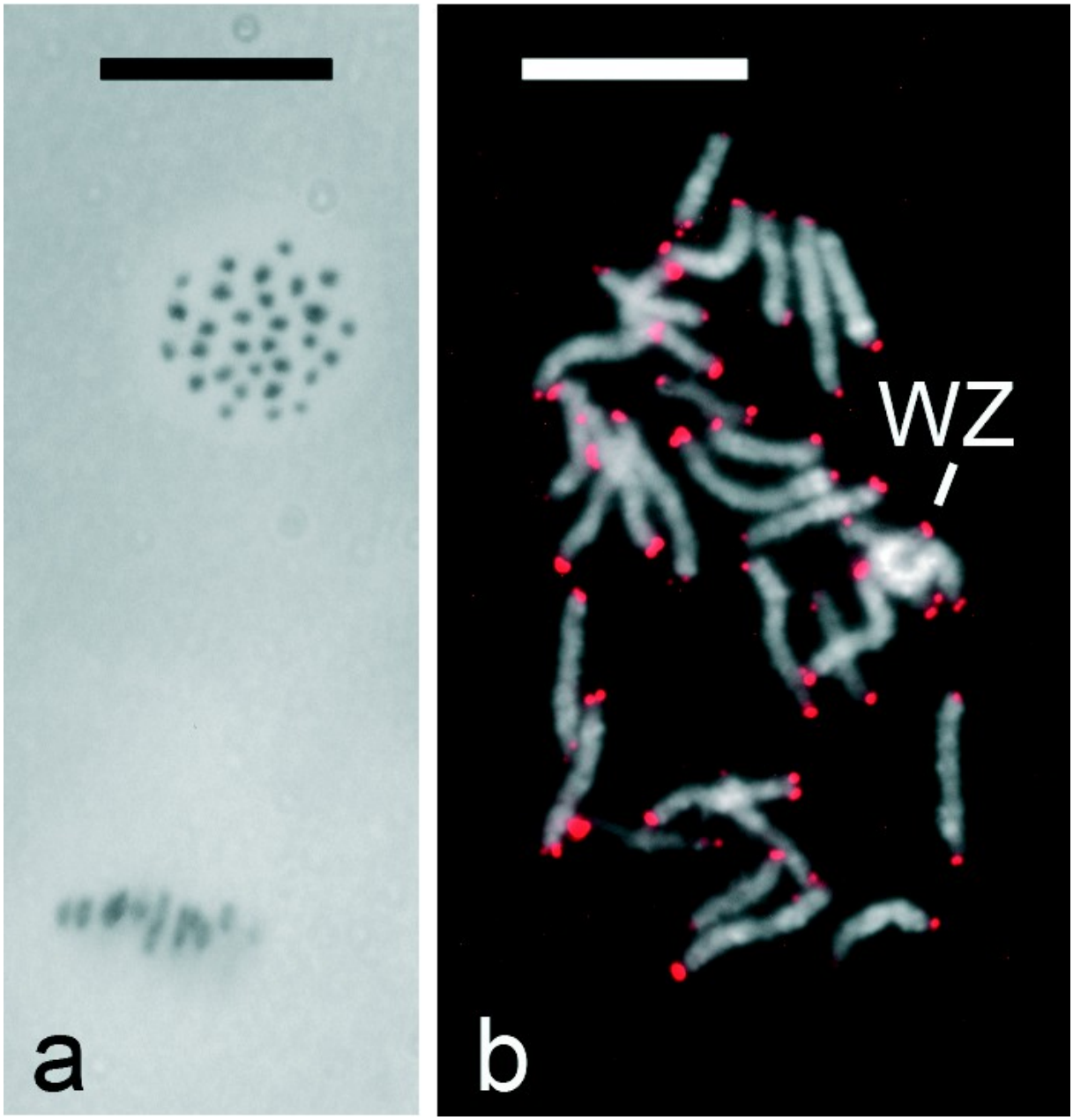

Fig. 4 *Danaus plexippus*. Male metaphase I (a) orcein-stained, and female pachytene (b). DAPI stained (white), telomere-FISH signals (red). WZ sex chromosome bivalent, bar 10 µm.

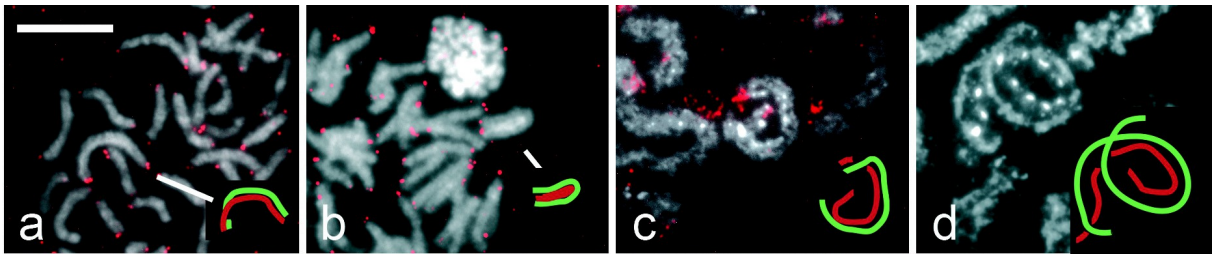

Fig. 5 *Danaus plexippus*. WZ bivalents in pachytene (a, b) and post-pachytene (c, d). Note the diffuse telomere signals in post-pachytene (c). Diagrams interpreting the WZ pairing: red W, green Z chromosome. Bar 10 µm.

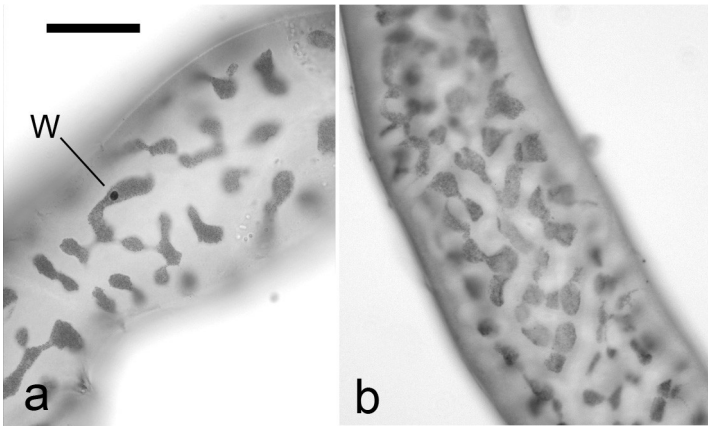

Fig. 6 *Danaus plexippus*. Sex chromatin (W) in highly polyploid, lobed nuclei from Malpighian tubules. Sex chromatin is present in females (a) and absent in males (b). Orcein-stained, bar 50 µm.

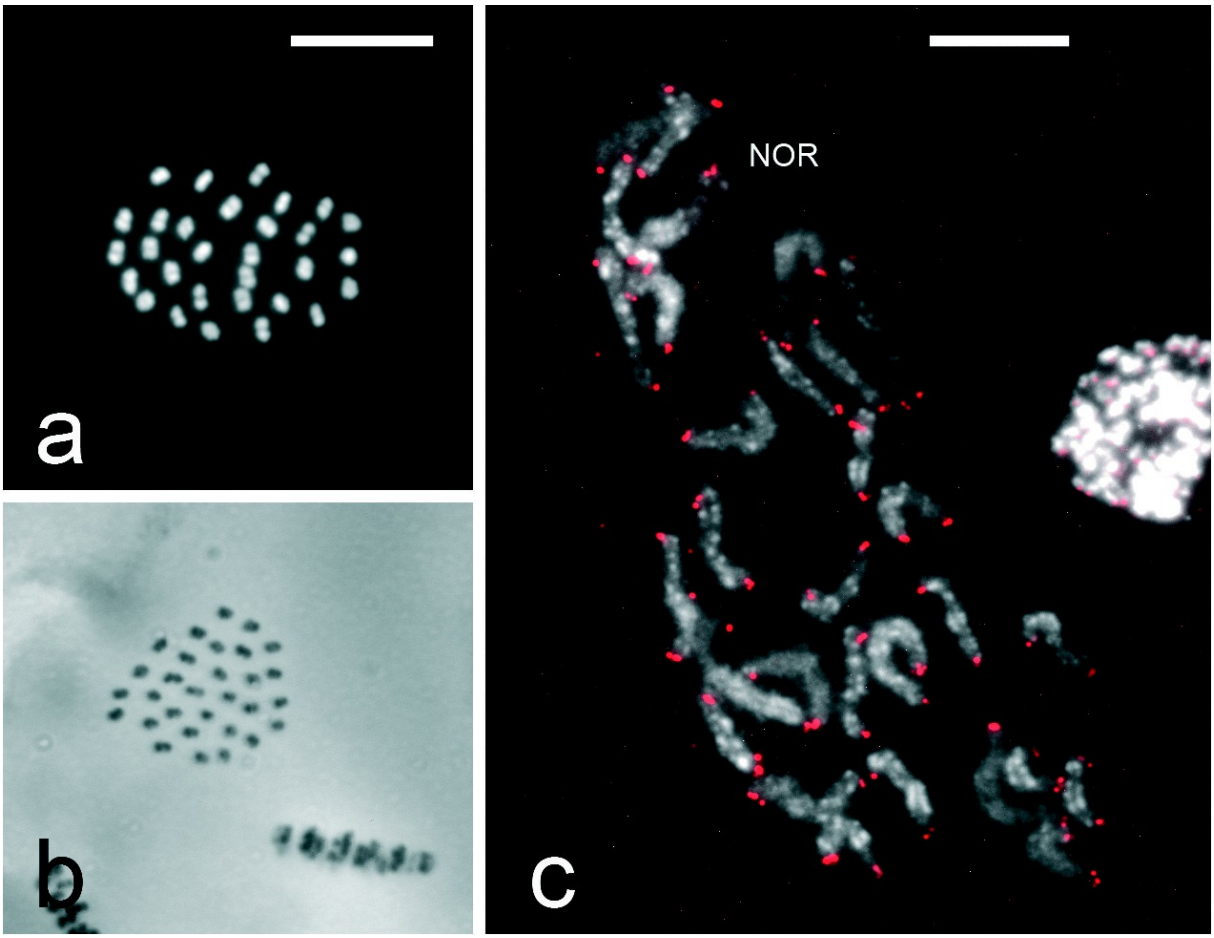

Fig. 7 *Danaus chrysippus*. Metaphase I of males from Watamu, Kenya (a) and Kitengela, Kenya (b), DAPI and orcein-stained, respectively. Female pachytene of a female from Watamu, Kenya (c) DAPI-stained, telomeres marked by FISH (red). NOR bivalent carrying the nucleolus-organiser region, bar 10 µm.

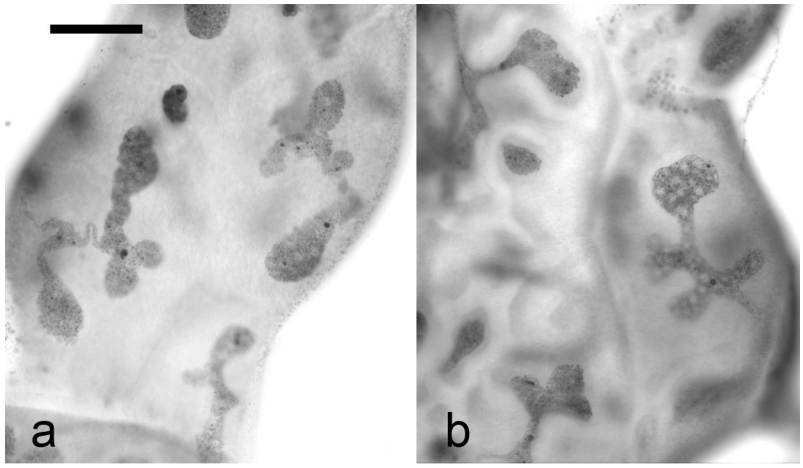

Fig. 8 *Danaus chrysippus*. Highly polyploid, lobed nuclei from Malpighian tubules with several small heterochromatin bodies present in both, females (a) and males (b). Source population: Watamu, Kenya. Orcein-stained, bar 50 µm.

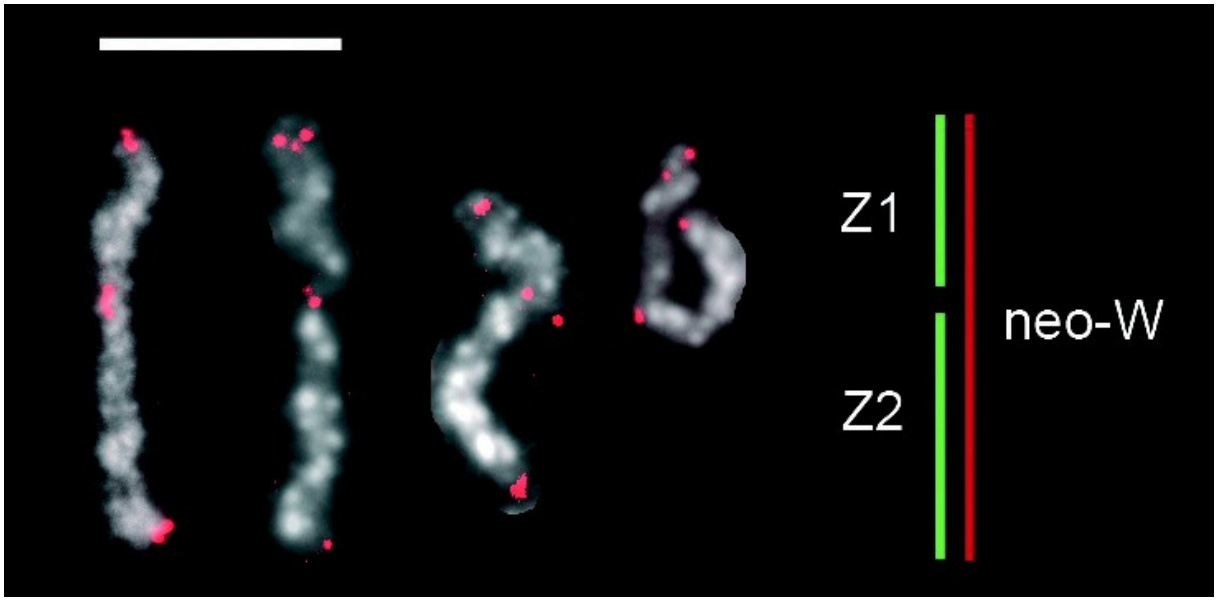

Fig. 9 *Danaus chrysippus*. Neo-W/Z_1Z_2 trivalents of pachytene complements from Kitengela, Kenya, DAPI-stained, telomeres marked by FISH (red). Diagramatic representation: red W, green Z chromosomes. Bar 10 µm.

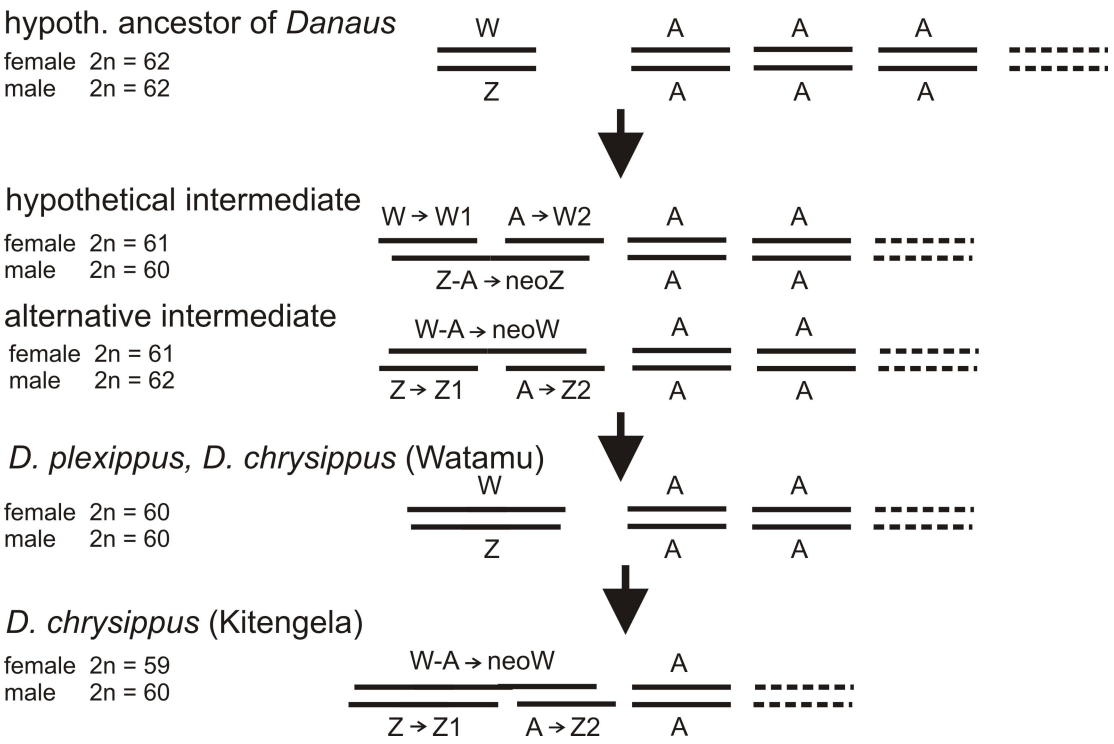

Fig. 10 Sex chromosome evolution by fusions in *Danaus*, from a hypothetical ancestor of the genus with the basic chromosome number of Lepidoptera, n = 62, via hypothetical intermediates to the actual situation of 2n = 60 in *D. plexippus* and *D. chrysippus* (Watamu), and on to *D. chrysippus* (Kitengela) with 2n = 59/60. Based on Mongue et al. (2016), Smith et al. (2016) and this paper.

Table 1 Conserved synteny among chromosomes of *M. cinxia* (Nymphalidae), *H. armigera* (Noctuidae), *B. betularia* (Geometridae), *B. mori* (Bombycidae) *H. melpomene* (Nymphalidae), and *D. chrysippus* (Nymphalidae). Data sources: *M. cinxia/ H. melpomene/ B. mori* (Ahola et al., 2014), *H. armigera/ B. mori* (Sahara et al., 2013), *B. betularia/ B. mori* (Van't Hof et al., 2013), *H. armigera* and *B. betularia* chromosomes #11, #23, #24, #29, #30, #31 versus *M. cinxia* (this paper, Tables s2 and s3), *D. chrysippus* (this paper).

Melitaea cinxia n = 31	*Helicoverpa armigera* n = 31	*Biston betularia* n = 31	*Bombyx mori* n = 28	*Heliconius melpomene* n= 21	*Danaus chrysippus* n = 30
#1(Z)	#1(Z)	#1(Z)	#1(Z)	#21(Z)	part of Z
#2	#4	#4	#4	part of #1	
#3	#15	#15	#15	#11	
#4	#12	#12	#12	part of #19	
#5	#6	#6	#6	#3	
#6	#5	#5	#5	part of #10	
#7	#18	#18	#18	#16	
#8	#17	#17	#17	#15	
#9	#10	#10	#10	part of #20	
#10	#9	#9	#9	part of #6	
#11	#22	#22	#22	part of #13	
#12	#29	#29	part of #11	part of #7	
#13	#8	#8	#8	part of #12	
#14	#23	#23	part of #23	part of #18	
#15	#13	#13	#13	part of #17	
#16	#19	#19	#19	#14	
#17	#3	#3	#3	#5	
#18	#25	#25	#25	#8	
#19	#21	#21	#21	#4	
#20	#7	#7	#7	#9	
#21	#16	#16	#16	#2	part of Z
#22	#28	#28	#28	part of #10	
#23	#26	#26	#26	part of #19	
#24	#27	#27	#27	part of #18	
#25	#20	#20	#20	part of #12	
#26	#14	#14	#14	part of #13	
#27	#31	#31	part of #24	part of #1	
#28	#2	#2	#2	part of #7	
#29	#24	#24	part of #24	part of #17	
#30	#30	#30	part of #23	part of #20	
#31	#11	#11	part of #11	part of #6	

Supplementary material

Table s1 Representation analysis. Coverage of protein-coding genes in read libraries from *D. chrysippus* females and a male analysed by TBLASTN. Queries are proteins from *B. mori*. The respective chromosome assignments are given together with the homologous chromosomes of *B. betularia*. Coverage data are presented as female/male ratios, corrected for library size.

Protein	Query acc. no.	Bombyx chr.#	Biston chr.#	WatamuFemale /WatamuMale ratio	KitengelaFema /WatamuMa rat
kettin	NP_001108348	1	1	0.51	0.4
lactate dehydrogenase	NP_001095933	1	1	0.47	0.4
paramyosin	ACF21977	1	1	0.44	0.4
ATP-binding cassette sub-family F member 2	NP_001040334	1	1	0.48	0.4
distal-less	XP_004933291	2	2	0.92	0.7
ARP6	NP_001040469	2	2	0.97	0.9
HMG-R_Bm	BAF62108	2	2	0.99	0.7
ribosomal protein L23	AAV34834	2	2	1.14	1.2
v-ATPase H subunit	ABF51492	3	3	1.09	0.9
Mn superoxide dismutase	BAD51413	3	3	0.94	1.2
LAMTOR3	XP_004924436	3	3	1.14	1.2
protein msta isoform B	XP_012546907	3	3	1.02	1.0
kinase 7-like	XP_004931740	3	3	1.30	1.1
RACK 1	NP_001041703	3	3	1.37	1.3
dopa decarboxylase	XP_012549895	4	4	1.07	0.9
Wnt-1	NP_001037315	4	4	0.97	0.8
DDC	AAK48988	4	4	1.06	0.9
kiser	AAS91007	4	4	1.12	0.9
EF-1a	D13338	5	5	0.97	0.7
EIF3S6	ABD36299	5	5	1.15	0.9
patched	XP_004930076	5	5	1.02	0.9
ultrabithorax	NP_001107632	6	6	0.97	0.8
vacuolar protein sorting 4	NP_001161188	6	6	0.91	0.8
antennapedia	BAA04087	6	6	1.01	0.9
cadherin-like membrane protein	BAA99404	6	6	1.19	1.1
BmChi-h	BAC67246	7	7	1.00	0.8
single-minded homolog 2	XP_012553046	7	7	1.22	1.2
ribonuclease L inhibitor	NP_001036911	7	7	1.15	1.1
enolase	NP_001091831	8	8	0.91	1.1
SEC62	XP_004924120	8	8	0.92	0.8
BmBR-C Z4-1	BAD24049	8	8	0.88	0.8
BmHSC70-4	BAB92074	9	9	1.01	0.9
aminopeptidase N	BAA32140	9	9	1.05	0.8
ecdysteroid-inducible angiotensin-converting	NP_001036859	9	9	1.02	0.9
c-Cbl-associated protein	NP_001166801	9	9	1.19	1.1
ribosomal protein S14	Q5UAM9	9	9	1.29	1.3
laccase 2A	BAG70891	10	10	1.05	0.9
ABC transporter	BAH03523	10	10	1.13	0.9
kynurenine 3-monooxygenase	ABE68382	10	10	0.96	0.9
iNOS-LP	BAB33296	10	10	0.89	1.2
aconitase	XP_004922163	11	11	0.83	0.8
orb2	XP_012548285	11	11	0.91	0.7
aprataxin	XP_004932950	11	29	1.21	1.0
eukaryotic transl. initiation factor 3 subunit L	NP_001037217	11	29	0.91	0.9
ribosomal protein L18	XP_004924182	11	29	0.92	0.6

signal sequence receptor	NP_001091760	11	29	1.20	0.
cysteine sulfinic acid decarboxylase	XP_004932908	11	29	1.06	0.
decapentaplegic	NP_001138801	12	12	1.00	0.
serine protease easter-like	XP_012548716	12	12	0.97	0.
XDH I	BAA07348	12	12	0.98	0.
ribosomal protein S2 (RpS2)	AAV34857	13	13	0.94	1.
cytochrome P450	ABF51415	13	13	1.23	1.
troponin T	NP_001040221	13	13	0.89	0.
ribosomal protein S3 (RpS3)	AAV34858	14	14	0.91	0.
enoyl-CoA hydratase	ABD36107	14	14	1.10	1.
eukaryotic translation initiation factor 3	NP_001037656	14	14	0.87	0.
M6-b	XP_004928477	15	15	0.95	0.
Notch	NP_001157370	15	15	1.02	0.
clathrin heavy chain	NP_001136443	16	16	0.53	0.
NIPSNAP	NP_001040139	16	16	0.48	0.
phenoloxidase subunit 1	NP_001037335	16	16	0.63	0.
Sex-lethal	BAE86938	16	16	0.66	0.
eIF3-S8	ABG54288	16	16	0.55	0.
cytoplasmic actin A4	AAC47432	17	17	0.97	0.
eIF3-S9	ABF55967	17	17	1.05	0.
snmp1	CAB65730	18	18	1.55	1.
gsk3	ADM32521	18	18	1.18	1.
microtubule-associated protein EB3	ABD36296	19	19	1.22	1.
eEF-2	ABF51485	19	19	1.03	0.
beta-tubulin	BAA19845	20	20	0.99	0.
cytochrome c oxidase	NP_001073120	20	20	0.89	1.
glucose-6-phosphate isomerase	NP_001091761	20	20	1.08	0.
ribosomal protein L7	NP_001037135	20	20	1.02	0.
MAP kinase-ERK kinase	NP_001036922	21	21	1.32	1.
ribosomal protein S6 (RpS6)	AAV34862	21	21	0.91	0.
RpS4	AAV34860	22	22	1.20	1.
calreticulin	BAC57964	22	22	1.07	0.
ribosomal protein L26	NP_001037233	23	23	1.03	0.
acyl-coenzyme A dehydrogenase	NP_001037672	23	23	1.19	1.
superoxide dismutase [Cu-Zn]	NP_001037084	23	23	1.11	1.
ribosomal protein S3Ae	NP_001037255	23	30	0.90	0.
SID1	XP_004930735	23	30	1.27	1.
deoxyhypusine synthase	NP_001129357	24	24	1.25	1.
PRP1	XP_004926349	24	24	0.93	0.
actin-depolymerizing factor 1	NP_001093278	24	31	1.06	1.
rho guanine nucleotide exchange factor 11	XP_012543780	24	31	1.49	0.
ryanodine receptor 44F	XP_012544761	24	31	1.10	1.
nephrin-like	XP_012552892	24	31	1.25	1.
dsx	BAB19780	25	25	0.96	0.
PPAE-3	AAL31707	25	25	1.02	0.
octopamine receptor	BAF33393	26	26	1.11	1.
ebony	XP_012551705	26	26	1.02	0.
attractin	NP_001138793	26	26	1.17	1.
heat shock protein 70	ABD36134	27	27	0.97	0.
lethal(2)essential for life-like	XP_004923510	27	27	1.05	0.
cubitus interruptus	XP_012553219	27	27	1.14	0.
serpin-4A	AAS68505	28	28	1.02	0.
H+ transporting ATP synthase gamma subunit	ABF51367	28	28	0.87	1.
6-phosphogluconolactonase	NP_001091839	28	28	1.16	1.
RRP41	XP_004933567	28	28	1.04	0.

Table s2 Conserved synteny. Mapping of protein-coding genes from *B. betularia* to *M. cinxia* scaffolds and chromosomes by TBLASTN.

Query	nt/aa	Protein	Biston chr #	Melitaea scaffold	Melitaea chr #	
AEP43804	aa	aconitase	11	scaffold2552	31	
ADF43214	aa	yellow2	11	scaffold1237	18	!
AEP43795	aa	orb2	11	scaffold4121	31	
AEP43782	aa	rpl4	11	scaffold1765	31	
				and scaffold3468	31	
AEP43785	aa	Cu-Zn superoxide dismutase	23	scaffold664	14	
AEP43800	aa	rpl26	23	scaffold4093	14	
ADO33057	aa	sex comb on midleg	23	scaffold2353	14	
AEP43787	aa	hexokinase	23	scaffold1715	14	
AEP43789	aa	rps30	23	scaffold3712	14	
				and scaffold11020	unassigned	
AEP43801	aa	acyl-CoA dehydrogenase	23	scaffold4319	14	
ADO32997	aa	ATP-dependent RNA helicase	24	scaffold663	29	
ADF43205	aa	ferritin heavy chain	24	scaffold3190	29	
ADF43204	aa	ferritin light chain	24	scaffold3190	29	
ADV76084	aa	AX11	24	scaffold1431	29	
ADO33016	aa	H+ tr ATP synthase subunit e	24	scaffold2396	29	
ADO33048	aa	rpl18	29	scaffold4822	12	
AEP43797	aa	radical sam	29	scaffold3566	12	
AEP43784	aa	zinc and ring finger 2	29	scaffold272	12	
ADO32992	aa	aprataxin	29	scaffold1663	12	
AEP43793	aa	black	29	scaffold795	12	
ADO33072	aa	wee	30	scaffold1228	30	
AEP43796	aa	pollux	30	scaffold601	30	
AEP43786	aa	fatty-acyl CoA reductase 3	30	scaffold1218	30	
AEP43788	aa	rps3a	30	scaffold1318	30	
ADO33043	aa	RhoGEF2	31	scaffold1271	27	
ADF43206	aa	GTPCHI	31	scaffold7454	27	
ADO33053	aa	ryanodine receptor	31	scaffold2911	27	
ADO33036	aa	puff	31	scaffold1406	27	
ADO33044	aa	ribonucleoprotein	31	scaffold1393	27	
ADO33068	aa	twinstar	31	scaffold1179	27	
				and scaffold494	27	
ADN84940	aa	sticks and stones	31	scaffold4925	27	

Table s3 Conserved synteny. Mapping of DNA segments and protein-coding genes from *H. armigera* to *M. cinxia* scaffolds and chromosomes by BLAST and TBLASTN.

Query	nt/aa	Protein	Helicoverpa chr #	Melitaea scaffold	Melitaea chr #	
BU038739	aa		11	scaffold1524	31	
KF982847	nt	18s rDNA	11	scaffold34886	31	
EY120734	nt		23	scaffold4395	14	
EY121788	nt		23	scaffold4319	14	
AAB14955	aa	methionine-rich storage protein	23	scaffold3593	14	
EY118312	nt		23	scaffold455	14	
EY117529	nt		23	scaffold2432	14	
BAN14651	aa	phosphatidylethanolamine binding	24	scaffold4488	29	
EE399561	nt	ribosomal protein L32	24	scaffold3837	29	
EY121429	nt		24	scaffold2437	29	
EY120156	nt		29	scaffold547	12	
EY121096	aa		29	scaffold5256	12	
EY120701	aa		29	scaffold2053	12	
EY119508	nt		29	scaffold6500	12	
EY117442	nt		29	scaffold1896	12	
ACC63240	aa	olfactory receptor 20	30	scaffold6628	5	!
GR968654	nt		30	scaffold2688	30	
EY119869	aa		31	scaffold494	27	
				and scaffold1179	27	
BU038461	nt		31	scaffold2911	27	