Published in final edited form as:

Nat Commun. ; 6: 8212. doi:10.1038/ncomms9212.

Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies

Xueyan Li[1,†], **Dingding Fan**[2,†], **Wei Zhang**[3,†], **Guichun Liu**[1,†], **Lu Zhang**[2,4,†], **Li Zhao**[1,5], **Xiaodong Fang**[2], **Lei Chen**[1,6], **Yang Dong**[1,‡], **Yuan Chen**[1,‡], **Yun Ding**[1,‡], **Ruoping Zhao**[1], **Mingji Feng**[2], **Yabing Zhu**[2], **Yue Feng**[2], **Xuanting Jiang**[2], **Deying Zhu**[2,7], **Hui Xiang**[1], **Xikan Feng**[2,4], **Shuaicheng Li**[4], **Jun Wang**[2,*], **Guojie Zhang**[2,8,*], **Marcus R. Kronforst**[3,*], and **Wen Wang**[1,*]

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China.

[2]BGI-Shenzhen, Shenzhen 518083, China.

[3]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA.

[4]City University of Hongkong, Hongkong, China.

[5]Department of Evolution and Ecology, University of California Davis, California 95616, USA.

[6]University of Chinese Academy of Sciences, Beijing 100049, China.

[7]School of Bioscience and Biotechnology, South China University of Technology, Guangzhou 510641, China.

Correspondence and requests for materials should be addressed to W.W (wwang@mail.kiz.ac.cn), M.R.K. (mkronforst@uchicago.edu), G.Z. (zhanggj@genomics.cn).

†These authors contributed equally to this work.

‡Current addresses: Dong Y., College of Life Science, Kunming University of Science and Technology, Kunming 650093, China; Chen Y., Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710, USA; Ding Y., Janelia Research Campus, Howard Hughes Medical Institute, Ashbum, VA 20147, USA.

*These authors jointly supervised this work.

Author contributions

W.W., X.L., Y. Dong and M.R.K. conceived the study and designed scientific objectives. W.W., X.L., and M.R.K. led the project and manuscript preparation. J.W., G.Z., X.Fang and M.F. supervised genome sequencing, assembly, and annotation and analysis. G.Z., D.F., D.Z. and L.Zhang developed the sequencing and assembling strategy. D.F., L.Zhang, Y.Z., Y.F., X. Feng and X.J. conducted genome and transcriptome assembly and evaluation, performed genome annotation. D.F., X.L., L.Zhao, and H.X. performed data analysis of genome and transcriptome. M.R.K. and W.Z. constructed butterfly cross, conducted RAD sequencing, constructed linkage map and performed analysis of genomic landscape of divergence. X.L., G.L., Y.Ding and Y.C. raised butterflies, provided materials, and prepared DNA and RNA samples for genomic and transcriptomic sequencing. G.L. and X.L. performed flow-cytometry analysis. X.L., G.L., L.C. and R.Z. performed CRISPR/Cas9 gene editing. X.L., W.W., M.R.K., L.Zhao, D.F., and W.Z. wrote the manuscript.

Additional information

Accession codes: The whole genome shotgun projects for the two swallowtail butterflies *Papilio xuthus* and *P. machaon* have been deposited in DDBJ/EMBL/GenBank under the accession codes LADI00000000 and LADJ00000000. The version described in this paper is the first version LADI01000000 and LADJ01000000. The genome assembly and annotation is available at GenBank as BioProject ID PRJNA270384 and PRJNA270386. The whole genome sequence reads have been deposited in the NCBI Sequence Read Archive (SRA) under the accession codes SRA230220 and SRA230221. The RAD reads have been deposited in the NCBI Sequence Read Archive (SRA) under the accession code SRA272689. The RNAseq data have been deposited in GenBank GEO under the accession codes GSE65280 and GSE65281, respectively.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications.

Competing financial interests: The authors declare no competing financial interests.

[8]Centre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, DK-2100 Copenhagen, Denmark.

Abstract

Butterflies are exceptionally diverse but their potential as an experimental system has been limited by the difficulty of deciphering heterozygous genomes and a lack of genetic manipulation technology. Here we use a hybrid assembly approach to construct high quality reference genomes for *Papilio xuthus* (contig and scaffold N50: 492 Kb, 3.4 Mb) and *P. machaon* (contig and scaffold N50: 81 Kb, 1.15 Mb), highly heterozygous species that differ in host plant affiliations and adult and larval color patterns. Integrating comparative genomics and analyses of gene expression yields multiple insights into butterfly evolution, including potential roles of specific genes in recent diversification. To functionally test gene function, we develop an efficient (up to 92.5%) CRISPR/Cas9 gene editing method that yields obvious phenotypes with three genes, *Abdominal-B, ebony*, and *frizzled*. Our results provide valuable genomic and technological resources for butterflies, and unlock their potential as a genetic model system.

Butterflies are famous for their extraordinarily diverse wing patterns, which differ not only among species, but also among populations, sexes, and even seasonal forms[1, 2]. Wing patterns are highly variable because they are multifunctional, involved in roles from crypsis to warning coloration, mimicry, thermoregulation, and mate selection[2, 3]. Beyond wing pattern, butterflies are also diverse in virtually all aspects of their biology, ranging from behavior and biogeography to cellular biology and biochemistry, with decades of study having placed much of this variation in a well-resolved ecological context[4]. These features make butterflies a promising system to explore the genetics, evolution, and development of morphological diversification and speciation[5, 6].

While butterflies have notable strengths in terms of natural variation and ecology, they also suffer critical shortcomings in the quest to characterize the genetic basis of organismal phenotypes, in particular difficulty in deciphering usually heterozygous genomes and a lack of functional genetics methodology. For instance, while there are estimated 18,000 butterfly species, there are currently only 6 butterfly genome sequences[7, 8, 9, 10, 11]. If additional, high quality genomes from closely related species were available, it would be feasible to comprehensively trace genetic changes responsible for phenotypic change over evolutionary time. However, the nature of high heterozygosity in most wild insects, including butterflies, has hampered efforts to obtain high quality reference genomes[7, 12, 13]. In addition, efficient and precise genetic manipulation technologies are indispensable for a model organism. However, general and efficient genetic manipulation of butterflies has not been reported, which, together with limited genomic resources, has greatly restricted their application as model organisms. The only case of genome editing in a butterfly is the use of zinc-finger nucleases in the monarch butterfly[14]. Recently, clustered regulatory interspaced short palindromic repeat (CRISPR) associated (Cas) based RNA-guided DNA endonucleases such as the *Streptococcus pyogenes* Cas9 (SpCas9) nuclease (CRISPR/Cas9) has emerged as an efficient tool for gene editing across a wide spectrum of organisms[15, 16], including insects such as fruitfly *Drosophila melanogaster*[17, 18] and silkworm *Bombyx mori*[19, 20].

Papilio xuthus and *Papilio machaon* are two closely related swallowtail butterfly species from the most basal lineage of butterflies, the family Papilionidae. Despite being closely related, *P. xuthus* and *P. machaon* differ in many aspects of their biology, including adult and larval color pattern, larval host plants, and geographic distribution, with *P. xuthus* mainly distributed in East Asia and *P. machaon* widely distributed across Asia, Europe and North America. Papilionidae is one of the most historically significant groups of butterflies; *Papilio machaon* was named as the type-species for all butterflies by Linnaeus[21] and since then the group has been a long-term focus in the study of mimicry, vision and learning, and pigmentation[3].

Here we present high quality reference genomes for the two highly heterozygous and closely related butterflies *P. xuthus* and *P. machaon*. Integrating comparative genomics and analyses of gene expression yields multiple insights into butterfly evolution. We develop an efficient and widely applicable CRISPR/Cas9 gene editing method that results in obvious phenotypes with three genes, *Abdominal-B*, *ebony*, and *frizzled*. Our results provide valuable genomic and technological resources for butterflies, and unlock their potential as a genetic model system.

Results

Genome sequencing and assembly

We collected *P. xuthus* and *P. machaon* samples from the suburb of Ya'an (Sichuan, China) (Supplementary Fig. 1). The heterozygosity of the two species was high, 1.008% for *P. xuthus* and 1.229% for *P. machaon*, and initial assembly using standard Illumina next generation sequencing data resulted in poor assemblies (Supplementary Note 1, Supplementary Table 1–3, Supplementary Fig. 2). Therefore we used 454FLXPlus long reads (≥ 700 base pair (bp)) in combination with a large quantity of Illumina short reads (100–150 bp, > 87 ×) to assemble contigs and we then used Illumina long-insert mate-pair sequencing data to generate scaffolds (Supplementary Note 1, Supplementary Table 1–11 and Supplementary Fig. 2–6). This hybrid assembly approach eliminated issues due to high heterozygosity, yielding a final assembly comprising 244 Megabases (Mb) with contig N50 of 492 Kilobases (Kb) and scaffold N50 of 3.4 Mb for *P. xuthus*, and a 281 Mb assembly with contig N50 of 81 Kb and scaffold N50 of 1.15 Mb for *P. machaon* (Table 1, Supplementary Tables 6–7). Three evaluation methods show that our assemblies are quite complete and reliable (Supplementary Note 1, Supplementary Tables 8–11). The assembled genome sizes are consistent with estimates by both *k-mer* analysis (Supplementary Table 3) and by flow-cytometry (Supplementary Table 5). Notably, the contig N50 sizes for both genomes are large compared to those of all published Lepidoptera genomes (Fig. 1a), and the contig N50 size of *P. xuthus* genome is the largest among all published animals genomes excluding such classical models as fruitfly *D. melanogaster*, human *Homo sapiens*, mouse *Mus musculus* and rat *Rattus norvegicus* (Supplementary Table 12).

We generated a linkage map using 74 offspring from a *P. xuthus* cross and used it to assemble 87% of the *P. xuthus* genome into 30 chromosomes (Supplementary Note 2, Supplementary Table 13 and Supplementary Fig. 7). Based on the *P. xuthus* chromosomal assembly, we compared syntenic relationships among *P. xuthus*, *P. machaon* and *B. mori*

(Fig. 1b), and found that chromosome 8 (chr8) of *P. xuthus* resulted from a fusion of ancestral lepidopteran chr8 and chr31. The Z chromosome (i.e, chr30) of *P. xuthus* is highly homologous to that of *B. mori* although a large fragment of *P. xuthus* chrZ is homologous to a region of *B. mori* chr5.

The assembled genomes of *P. xuthus* and *P. machaon* have similar composition of repetitive element (Supplementary Note 3, Supplementary Tables 14–16 and Supplementary Fig. 8) and are predicted to contain 15,322 and 15,499 protein-coding genes, respectively (Table 1, Supplementary Note 4, Supplementary Tables 17–18 and Supplementary Fig. 9–11). More than 80% of gene models are supported by evidence from at least two prediction methods (*ab initio*, homology, and RNA-seq) (Supplementary Fig. 9) and both genomes show similar gene features to those of other Lepidoptera (Supplementary Table 17 and Supplementary Fig. 10–11). About 50% of gene families are conserved in butterflies (Supplementary Table 18 and Supplementary Fig. 12). We inferred a phylogenetic tree among representative taxa spanning holometabolous insects including seven lepidopteran species (*P. xuthus*, *P. machaon*, *Heliconius melpomene*, *Danaus plexippus*, *Melitaea cinxia*, *B. mori*, *Plutella xylostella*) using 1,071 single-copy orthologs (Supplementary Note 5 and Supplementary Fig. 13). We analyzed lineage-specific genes and gene families and identified positively selected genes (Supplementary Note 5, Supplementary Tables 19–26 and Supplementary Fig. 13–14). We also obtained genome-wide gene expression profiles for *P. xuthus* and *P. machaon* at each of 10 development stages: egg, each of the five larval instars, male pupa, female pupa, male adult and female adult (Supplementary Note 6, Supplementary Tables 27–30 and Supplementary Fig. 15–19). By integrating these multiple layers of information, we inferred important genes and gene pathways in butterfly evolutionary history, and then we developed an efficient and widely-applicable genome editing method for butterfly functional genetics.

Assembly, gene annotations, gene family and developmental transcriptome can be obtained from ftp://ftp.genomics.org.cn/pub/papilio.

Comparative Genomics

Comparison to other insect genomes revealed interesting evolution patterns, in particular evolution in the juvenile hormone (JH) pathway in tandem with butterfly morphological diversification. Juvenile hormones are key hormones in insect development. There are six different JH forms in insects, and interestingly Lepidoptera have five of the six JH forms while other insects generally have only one JH form[22]. Short-chain isoprenyl diphosphate synthases (scIPPSs) are key enzymes in the JH synthesis pathway[23] (Supplementary Fig. 20). Based on comparisons across holometabolous insects, all butterflies experienced scIPPS gene expansion (Fig. 1c, Supplementary Note 7 and Supplementary Tables 31–32). Furthermore, scIPPSs can be classified into farnesyl diphosphate synthases (FPPSs) and geranylgeranyl pyrophosphate synthases (GGPPSs). In butterflies, FPPS genes expanded from one copy to 14 copies in *Papilio*, while GGPPS genes expanded from one copy to 7–8 copies in nymphalid butterflies (Fig. 1c and Supplementary Table 32). It has been suggested that in moths it is alternative splicing of scIPPS genes that contributes to diverse JH forms[24]. In addition, notable expansions of genes encoding JH epoxide hydrolase (JHEH) and JH diol kinase (JHDK) in *Papilio* (Supplementary Table 31 and Supplementary Fig. 21), combined with the prominent expansion of scIPPSs in butterfly JH biosynthesis, suggest significant diversification of JH metabolism in butterflies, which may be related to the evolution of their numerous morphological forms. Interestingly, scIPPS genes may have also played roles in the differentiation among *Papilio* species because we observed that while *P. machaon* and *P. xuthus* share the same 14 FPPS genes, these orthologs were differentially expressed between *P. machaon* and *P. xuthus* (Fig. 1e, f and Supplementary Table 33). Collectively the above observations suggest that scIPPSs might have played roles in both the deep evolutionary history of butterflies as well as very recent divergence between closely related species.

Comparative genomics also revealed evidence for genes critical in host plant adaptation. Although *P. machaon* and *P. xuthus* are congeneric species, their host plants are different. As larvae, *P. xuthus* feeds on plants in the family Rutaceae, such as citrus, while *P. machaon* feeds on plants in the family Apiaceae. Both groups of plants contain toxic furanocoumarins but Apiaceae have higher levels of furanocoumarins than Rutaceae[25]. By comparing *P. xuthus* and *P. machaon* genomes to those of *D. plexippus*, *H. melpomene*, *B. mori*, and *D. melanogaster*, we found CYP6, a cytochrome P450 monooxygenase (P450) gene family, expanded in swallowtail butterflies (Supplementary Note 8, Supplementary Table 34 and Supplementary Fig. 22). CYP6, together with CYP9 and alpha esterase families, are thought to contribute to xenobiotic detoxification in insects[26]. We observed that 50% of CYP6 members in the two *Papilio* genomes belong to the CYP6B subfamily, while there is no CYP6B in the monarch or passion-vine butterfly genomes, species whose host plants have no toxic furanocoumarins (Fig. 1d, Supplementary Table 35 and Supplementary Fig. 22). Higher expression of CYP6 genes in larvae of *P. machaon* and *P. xuthus*, compared to other developmental stages (Fig. 1g and h), suggests that CYP6 genes may play an important role in larval feeding. CYP6AB is another P450 subfamily found only in Lepidoptera. We found additional expansions of CYP6B and CYP6AB in *P. machaon* only (Fig.1d and Supplementary Fig. 22), which may account for this species' ability to feed on Apiaceae plants with a higher content of furanocoumarins.

Genomic divergence among swallowtail species

There are striking morphological differences between *P. xuthus* and *P. machaon*, especially at the fifth-instar larval ontogenetic stage (Supplementary Fig. 1). Similar to caterpillars of hundreds of other tropical butterfly species, the fifth instar larvae of *P. xuthus* display false eyespots on their metathorax which permit them to escape predation by mimicking venomous snakes[27, 28]. On the other hand, fifth instar larvae of *P. machaon* exhibit disruptive coloration characterized by highly contrasting patterns of black and green stripes with yellow spots, providing camouflage that conceals body shape at a distance while serving as warning coloration at close range[29, 30]. We found a number of genes related to morphological traits that were expanded, positively selected, or differentially expressed in *P. xuthus* and *P. machaon* (Supplementary Notes 9–11, Supplementary Tables 36–40 and Supplementary Fig. 22–36). Butterfly larvae, swallowtails' in particular, frequently have marked green pigmentation, which consists of blue bile binding to bilin-binding protein (BBP) and a yellow component binding to proteins encoded by yellow-related gene

(YRG)[31, 32]. Interestingly, we observed an expansion of the BBP gene family in butterflies, especially in swallowtails (Supplementary Table 36 and Supplementary Fig. 32), which may be related to the exceptionally green caterpillars in swallowtails. We also identified a *Papilio*-specific single-copy YRG gene, and found it was expressed at a much higher level in the fifth instar larvae of *P. machaon* compared to *P. xuthus* (Supplementary Table 36 and Supplementary Fig. 33–34), corresponding to the stronger yellow marking in *P. machaon* (Supplementary Fig. 1).

To better characterize patterns of genome evolution among closely related swallowtail species, we performed a genome-wide sliding window analysis of DNA sequence divergence among *P. xuthus*, *P. machaon*, and a third *Papilio* species, *P. polytes* (Fig. 2, Supplementary Note 12, Supplementary Tables 41–43 and Supplementary Fig. 37) with its resequencing data available[33]. A total of 915 genes showed strong divergence in all pairwise comparisons. By integrating this with results from our analyses of recent positive selection and differential gene expression, and looking for overlap among datasets, we found a small subset of genes that might have played an important role in recent diversification among these three closely related *Papilio* species (Supplementary Table 43). One particular gene was *frizzled (fz)*. We found that the genome window containing the 5′ untranslated region of *fz* emerged as a highly divergent segment in all pairwise comparisons among the three *Papilio* species but the coding region itself did not. This gene also showed strong differential expression between *P. xuthus* and *P. machaon* across all developmental stages (Supplementary Note 11, Supplementary Tables 39 and Supplementary Fig. 36), one of only 32 genes to do so (and one of only four highly divergent genes to do so) (Supplementary Fig. 17). *Frizzled* is well-known for its role polarizing epithelia during development (planar cell polarity, PCP), but it also plays a distinct role patterning the larval cuticle during embryogenesis[34]. In *Drosophila*, *fz* interacts with *wingless*, *armadillo* and other genes to determine the distribution of naked cuticle vs. denticle on larvae[35]. Given the highly divergent larval morphology among swallowtails, we hypothesized that *fz* may play a similar role in influencing larval cuticle development, in which case DNA sequence and expression divergence at this gene could be a result of natural selection on larval anatomy. Furthermore, another PCP pathway (*Fat/Dachsous*) that also influences larval denticle patterning in *Drosophila*[36] showed evidence of recent, positive selection in *P. machaon* (Supplementary Tables 38–40). Experimentally testing genes' potential role in morphological diversification among *Papilio* species required functional genetics methodology not available in butterflies.

CRISPR/Cas9 system

Given the recent, wide spread utility of CRISPR/Cas9[15, 16], we decided to implement a CRISPR/Cas9 genome editing approach in *P. xuthus* (Table 2, Supplementary Note 13, Supplementary Tables 44–50 and Supplementary Fig. 38–55). Prior to targeting candidate diversification genes, such as *fz*, we developed our pipeline (Supplementary Fig. 38) by focusing on the gene *Abdominal-B* (*Abd-B*) and we then verified it with *ebony* (*e*) (Supplementary Table 44 and Supplementary Fig. 39) because these two genes are expected to have predictable and visible morphological phenotypes. *Abd-B* is a HOX gene that plays a critical role in determining cell fate in the tail end of the organism. For instance, mutation or decreased expression of *Abd-B* causes development of extra prolegs on all segments posterior to the 6th abdominal segment (A6) in silkworm larvae[37]. We began our *Abd-B* gene editing experiment with low concentrations of single guide RNAs (sgRNAs) (150 ng per µl) and Cas9 mRNA (300 ng per µl), but failed to observe morphologically mutated individuals although sequencing yielded evidence of low frequency disruptions (Supplementary Table 45 and Supplementary Fig. 40). Increasing injection concentrations (sgRNAs:Cas9 (ng per µl), 200:300, 600:600) or the ratio (sgRNAs/Cas9mRNA (ng per µl): 100/150) gave rise to a low frequency (7.4%) of morphologically mutated larvae with varying abnormalities (Supplementary Tables 45–46 and Supplementary Fig. 41–44). Remarkably, further increasing injection concentration of sgRNAs (982 ng per µl) and Cas9 mRNA (1200 ng per µl), together with shortening the time period from egg laying to injection from 4 hours (hrs) to 2 hrs, resulted in both a high frequency (92.5%) of morphologically mutated individuals and a majority (90%) of mutated individuals with severe expected phenotypic abnormalities (Fig. 3a–d, Supplementary Tables 45–46 and Supplementary Fig. 43a). For example, all mutated individuals had a curled abdomen resulting from abnormal terga on segments A3 and after, and most individuals had prolegs on all or part of segments A7 - A10, segments that do not normally have prolegs (Supplementary Tables 45–46 and Supplementary Fig. 43). Sequencing of the targeted region showed a high frequency of gene disruption (Supplementary Table 46 and Supplementary Fig. 42). In addition, we also tested whether expression of the target gene (*Abd-B*) at RNA or protein level disappeared or was extremely reduced in mutants using both the quantitative reverse transcription-PCR (qRT-PCR) and western blotting. Both qRT-PCR and western blot results suggested that the level of Abd-B protein in mutants was extremely low compared to that of the wild type individuals (Supplementary Note 13, Supplementary Fig. 45a–b), further providing evidence of *Abd-B* gene disruption. Based on our experience with *Abd-B*, we conclude that the key factors for successful gene editing in butterflies include high concentrations of single guide (sgRNA) and Cas9 mRNA, an appropriate ratio, mixed injection of two or more sgRNAs with close targeting sites, and timing egg injection to target early embryogenesis (Supplementary Note 13). It is noteworthy that all the mutant test experiments used individuals that developed from injected eggs (G_0), and DNA, RNA as well as protein were extracted from G_0 whole body samples.

Having established the methodology, we further tested it on a second proof-of-principle gene, *ebony*. *Ebony* encodes the enzyme N-β-alanyl dopamine synthetase, a central component of melanin biosynthesis (Supplementary Fig. 25), and expresses not only in pupa and adult but also larvae of *P. xuthus* and *P. machaon*, including the green region of both larvae (Supplementary Fig. 26, 28c and 30). Swallowtail butterflies have a unique class of wing pigments, papiliochromes, that derive in part from melanin precursors such as tyrosine and β-alanine via the action of *ebony*[38]. *Ebony* is also known to influence behavior[39]. We obtained many *ebony* mutants of *P. xuthus* with the CRISPR/Cas9 system (Fig. 3e–j, Supplementary Note 13, Supplementary Tables 47–48 and Supplementary Fig. 46–51). Mutated 5th instar larvae displayed enhanced melanic pigmentation (Fig. 3e and Supplementary Fig. 46), consistent with *ebony* mutants in other insects[40, 41, 42], and they also showed an absence of orange color in the false eyespot (Fig. 3f and Supplementary Fig. 46). Interestingly, two mutated 5thinstar larvae also showed abnormal curticular structure in the location of eversible osmeterium, a swallowtail's specific defensive organ, and one or both of tubular arms of their osmeterium, unlike those of wild type individuals, could not evert when irritated by abrupt contact (Supplementary Fig. 47). Adult mutants showed brown pigmentation across the body and the regions of wings that were normally yellow (Supplementary Fig. 48). We also verified the decreased expression of *ebony* in the 5th instar larvae of mutants compared to that of wild type by qRT-PCR (*t* test, P=0.001) (Supplementary Note 13 and Supplementary Fig. 51), suggesting that *ebony* gene was largely disrupted in mutants.

Finally, we turned to the candidate diversification gene, *frizzled* (*fz*). Consistent with our hypothesis that *fz* plays a role in generating divergent larval morphology, we obtained mutants showing a variety of larval anatomy phenotypes, including asymmetric appendages and cuticular structures (Fig. 3k–n, Supplementary Tables 49–50 and Supplementary Fig. 52–55). For example, mutated larvae had smaller prolegs on one side of the body, smooth and colorless dorsal cuticle, or vestigial tubercles on the prothorax or metathorax. These results suggest that divergent evolution at the *frizzled* gene, as well as differential expression, may be related to the distinct larval anatomy between species.

In order to exclude the possibility that these morphological mutants were induced by off-target cleavage events during the genome editing process, we further identified all possible off-target sites in the genome using three methods (CasOT[43], Cas-OFFinder[44] and COSMID[45]) with the mismatches up to 5 bp. For the two target sites (Px_10703_e-T454 and -T6) of *ebony*, the three methods identified 197 possible off-target sites, but none of them is completely identical to the target sites (Supplementary Note 13 and Supplementary Table 51). Then we performed whole genome resequencing for an *ebony* 5th instar larva mutant (Supplementary Note 13). Our resequencing data confirmed disruption of the two target sites of *ebony* gene (Supplementary Fig. 56), but didn't disrupted any other functional gene in the mutant (Supplementary Table 52). Among the197 possible off-target sites of *ebony* T454 (143 sites) and T6 (54 sites), we observed that 33 sites have variation different from the reference genome. Among these 33 sites, 29 are due to SNVs, and 4 are due to indels. Three SNVs are found at the possible off-target sites of three genes, respectively, but they are all synonymous. Other 30 sites are all in non-coding regions. Therefore, the observed phenotypes are most possibly caused by the designed target genes' disruption rather than off-targeting. The 33 variation sites that didn't disrupt genes could result from polymorphisms given the high heterozygosity of *P. xuthus*, although the possibility of off-targeting could not completely excluded at this point. High specificity of Cas9 editing and low incidence of off-target mutations have recently been reported by whole-genome sequencing of human stem cell[46, 47], in butterflies future detailed analysis on offsprings from the same parents will allow to test whether some of the variations are from off-targeting or not. Nevertheless, our results show that Cas9 is an efficient and reliable tool in butterfly gene editing.

It is worth noting that we have only focused on *P. xuthus* at this point because we found difficult to raise *P. machaon* under laboratory conditions. However, our experience suggests that this Cas9 method will be widely applicable to any butterfly species as long as one can collect a sufficient number of eggs and then rear them to an appropriate developmental stage to observe mutants' phenotypes.

Discussion

The amazing diversity of butterflies provides a rich, natural experiment with which to explore the molecular mechanisms of morphological and species diversification, but doing so with butterflies has been hindered by complications associated with assembling genome sequences from highly heterozygous, outbred insects and an absence of widely-applicable gene editing technology. By focusing on a biologically significant group of butterflies, we have developed widely-applicable methods to overcome these challenges, thereby enabling the scientific community to fully harness the research potential of this remarkable insect radiation. Furthermore, our results lend unique insight into the evolutionary history of butterfly diversification. Much attention is often paid to adult butterflies, and their wing patterns in particular. However, like all holometabolous insects, the butterfly life cycle is a two-stage process consisting of larva and adult. Perhaps surprisingly, our genomic analyses point repeatedly to the caterpillar stage as a major driver of natural selection with many of the genes that emerge from our analyses connecting to the larval stage through effects on larval morphology or host plant use. This enriches our view of insect evolution and with these new genetic and genomic tools in hand, we can now test the generality of these findings.

Methods

Butterfly husbandry and sample collection

Pupae of *Papilio xuthus* Linnaeus (*Px*) and *P. machaon* Linnaeus (*Pm*) were mainly purchased from butterfly insectariums in the suburbs of Ya'an, Sichuan, China. A proportion of pupae and adults were also acquired by raising the larvae collected by Xueyan Li and Guichun Liu in the suburb of Kunming, Yunnan, China. Pupae were reared under the conditions of 26°C, 75% relative humidity and 18h/6h light/darkness. Emerged adults were crossed via hand-pairing[48, 49]. After mating, females were placed in net rooms with host plants for oviposition and then the mated adults were frozen at −80 °C for DNA extraction or kept as dried specimens. Eggs were collected to be used for experiments such as microinjection or allowed to develop on host plant at room temperature. *Px* larvae were fed with a horticulturally rutaceous plant (*Zanthxylum piperitum*) found by Xueyan Li, which appears to be a new host plant due to the species expansion and local adaptation. *Pm* larvae were fed with Umbelliferae plants fennel (*Foeniculum vulgare*) or carrot (*Daucus carota*).

In order to get high quality and high volume of DNA for library construction, head and thorax tissues (excluding wings) of six homogametric male (ZZ) adults of *P. xuthus* and *P. machaon* both from Sichuan, China, were dissected and used for DNA isolation using a Gentra Puregene Blood kit (Qiagen, Germany) following manual instructions. A single individual was used to construct short-insert libraries (150 bp, 250 bp, 500 bp); each of other five individuals was used to construct long-insert libraries (2 Kb, 5 Kb, 10 Kb, 20 Kb) and 454-library, respectively. RNA samples from different development stages of both species including egg, larvae from 1st instar (L1) to 5th instar (L5), pupa (female and male) and adult (female and male) were collected for transcriptome sequencing. Due to small body of egg and early development larvae, 10 eggs, 10 L1 and 5 L2 were used for constructing single RNA library, respectively; for other development stages, single individual was used to construct RNA library. For linkage mapping, *P. xuthus* pupae were obtained from a butterfly breeder, and one three-day old male and one newly emerged female were used to set up a cross. The two parents (adults) and their 94 F1 offspring (2^{nd}-4^{th} instar larvae) of *P. xuthus* were collected for DNA isolation and RAD sequencing. The eggs of *P. xuthus* for microinjection of Cas9 were collected in our greenhouse. The related protocols involved with butterfly in this study have been reviewed and approved by the internal review board of Kunming Institute of Zoology, Chinese Academy of Sciences.

Genome sequencing and assembly—All short reads from short- and long-insert libraries for genome assembly were produced with the Illumina His eq2000 platform at BGI (Shenzhen, China) (Supplementary Note 1). Long reads were produced with the Roche 454FLXPlus Titanium platform at Duke University (USA) using the 454 Life Science/ Roche protocol on the GS-FLX+ System. We carried out several rounds of assembly using high quality data (Supplementary Note 1) from different sequencing platforms, i.e, Illumina short reads, 454 long reads, or both combined (Supplementary Table 1). Considering that there exist a larger number of short contigs in the assemblies only by Illumina short reads (Supplementary Table 1), we took a hybrid assembly strategy (Supplementary Fig. 5) of combining Illumina short reads with 454 long reads for the final assembly. Firstly, about 30 × Illumina reads from short insert libraries were used to correct amplicon pyrosequencing errors in 454 reads. To do the correction, Coral 1.4 (http://www.cs.helsinki.fi/u/lmsalmel/coral/)[50] was used to correct sequencing errors by forming multiple alignments. We employed 454 model with options "–f –k 21 –e 0.07 –p 20" during correction. For *P. xuthus* and *P. machaon*, 1.60% and 1.10% errors in 454 read bases were corrected respectively. Secondly, corrected 454 reads were then fed into Newbler 2.6[51] in DataAnalysis_2.6 toolkit with options "-mi 90 -ml 40 -nrm -het -m -cpu 20 -l 500" to build contigs. Thirdly, paired-end sequencing Illumina short reads were employed to combine contigs by SSPACE-1.2 basic[52] (http://www.baseclear.com/landingpages/sspacev12/). The reads from short insert size libraries (<2 Kb) were first used to construct scaffolds (with parameters -a 0.6 -x 0 -g 1 –k 5). Then reads from long insert size libraries (≧2 Kb) were added in with parameters "-a 0.6 -x 0 -g 1 –k 3". Finally, we used short paired-end (150 bp, 250 bp, 500 bp libraries) information to retrieve gap-crossed read pairs with one end mapped to a unique contig and the other located in a gap and performed a local assembly for these collected reads to fill the gaps by GapCloser v1.12[53] (http://soap.genomics.org.cn/soapdenovo.html). We further estimated genome size using flow-cytometry (Supplementary Note 1) and then made haplotype separation and excluded microsporidia genome reads (Supplementary Note 1). Assembly quality was evaluated by three methods, i.e., aligning short reads to assemblies, aligning transcripts to evaluate completeness of assemblies, and evaluating completeness by CEGMA (Core Eukaryotic Genes Mapping Approach) (Supplementary Note 1).

RAD-based linkage mapping of scaffolds—RAD library preparation was performed mainly according to Etter et al.[54] with some modifications and sequencing was performed with an Illumina Hiseq 2000, and *Eco*RI site was used for RAD tag extraction

(Supplementary Note 2). Linkage maps were constructed using both regression (Kosambi mapping function) and maximum likelihood algorithms (Supplementary Note 2). Syntenic relationship of chromosomal linkage between *P. xuthus* and *P. machaon* was analyzed based on their orthologs (Supplementary Note 2).

Genome annotation—Repetitive sequences and transposable elements in *P. xuthus*, *P.machaon*, and two nymphalid butterflies (*Danaus plexippus* and *Heliconius melpomene*) were annotated using a combination of homology to Repbase sequences, *de novo* prediction approaches and Tandem Repeats Finder (Supplementary Note 4). Protein coding genes in *Papilio* genomes were predicted using homology-based methods, *ab initio* gene prediction and RNA-seq data (Supplementary Note 5). Gene function information, motifs and domains of their proteins were assigned by comparing with public databases including SwissProt, TrEMBL, KEGG, InterPro, Gene Ontology (GO) (Supplementary Note 5). In order to avoid biased comparisons of gene sets among butterflies species used in this study and two former studies (*D. plexippus* and *H. melpomene*), we conducted all these analyses for the four butterfly genomes.

Gene evolution—In order to explore gene evolution patterns among butterflies, gene cluster was analyzed including the genomes of five butterflies (*P. xuthus*, *P. machaon*, *D. plexippus*, *H. melpomene, Melitaea cinxia*), moths (*Bombyx mori*, *Plutella xylostella*), mosquito (*Anopheles gambiae*), fruit fly (*Drosophila melanogaster*), beetle (*Tribolium castaneum*), bee (*Apis mellifera*), which cover all orders of holometabolous insects (Supplementary Note 5). 1071 single-copy genes were used for constructing phylogenetic tree of 11 species (Supplementary Note 5). Gene orthologous relationship among Lepidopteran insects (*P. xuthus, P. machaon*, *D. plexippus*, *H. melpomene* and *B. mori*) was determined (Supplementary Note 5), and positively selected genes were identified in *P. xuthus, P. machaon* and both of them (Supplementary Note 5).

Transcriptome sequencing and analysis

Transcriptome sequencing for different developmental stages was performed with Illumina RNA-seq protocols, and two methods, i.e., *de novo* assembly of clean RNA reads and mapping them back to the assembled genomes, were carried out for transcriptome assembly (Supplementary Note 6). RPKM (Reads Per Kb per Million reads) was used to measure gene expression abundance. We analyzed gene expression dynamics in development (Supplementary Note 6), differentially expressed genes (Supplementary Note 6), expression patterns of positively selected genes, and lineage-specific genes (Supplementary Note 6).

Specific gene families and pathways—In order to explore possible significance of expansion of farnesyl pyrophosphate synthase (FPPS) gene family in *Papilio* butterflies, we investigated all genes encoding short-chain isoprenyl diphosphate synthases (scIPPS) and other enzymes in the pathways of juvenile hormone (JH) biosynthesis and degradation as well as protein prenyltransferasesin ten insect species (*P. xuthus, P. machaon*, *D. plexippus*, *H. melpomene, B. mori*, *P. xylostella, A. gambiae, D. melanogaster, T. castaneum*, *A. mellifera*), constructed their gene trees and analyzed their expression patterns (Supplementary Note 7). We also identified gene families involved in detoxification of various xenobiotics, including cytochrome P450 monooxygenases (P450s), glutathione S-transferase (GST) and carboxylesterases (COE) (Supplementary Notes 8–9). In addition, we also analyzed the genes related to body color (Supplementary Note 10) and planar cell polarity (PCP) genes (Supplementary Note 11) playing a fundamental role in morphogenesis of vertebrates and invertebrates. All phylogenetic trees were constructed using Maximum likelihood in PAML[55].

Analysis of genomic divergence—The divergence regions was analyzed based on the assembled genomes of *P. xuthus*, *P. machaon* and Illumina genome resequencing data of one individual of a third *Papilio* species (*P. polytes*) with its data available. We conducted pairwise local alignments using the *P. xuthus* genome as a reference and aligned paired-end reads of one male *P. polytes* individual and simulated paired-end reads of the *P. machaon* genome. Then we calculated fixed different SNPs per bp for each 50 kb window across the reference genome for three comparisons (*P. xuthus*/*P. machaon*, *P. xuthus*/*P. polytes* and *P. machaon*/*P. polytes*) and the highly divergent regions were characterized using 95% smoothed empirical likelihood quantiles among all three comparisons as a cutoff (Supplementary Note 12).

Genome editing in butterfly using CRISPR/Cas9 system

We designed sgRNA targ et sites by seeking sequences corresponding to N_{20}NGG on exon regions of the sense or antisense strand of the DNA by ZiFit Targeter program[56]. Then we BLAST (using Basic Local Alignment Search Tool) these candidate target sequence against the *P. xuthus* genome to eliminate those with off-target sites using strict criteria, where the candidate editable site is defined only when the seed region (12 nucleotides (nt) to protospacer adjacent motif (PAM) NGG) is unique[15]. From candidate editable sites, we selected those with the first two bases of GG, GA or AG for sgRNA synthesis. SgRNA can be synthesized by plasmid-based or PCR-based strategies and we used a PCR-based method to synthesize sgRNAs. Briefly, a unique oligonucleotide encoding T7 polymerase binding site and the sgRNA target sequence N_{20}) was designed as forward primer (CRISPRF = GAAATTAATACGACTCACTATA$\underline{N}_{20}$GTTTTAGAGCTAGAAATAGC), and a common oligonucleotide encoding the remaining sgRNA sequence (sgRNAR = AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCT TATTTTAACTTGCTATTTCTAGCTCTAAAAC) was designed as a reverse primer[18]. All the oligonucleotides (CRISPRF, sgRNAR) were synthesized by GENEray Company (Shanghai, China). The information of target sites and the primers for analyzing target sites of three genes (*Px_03961_Abd-B, Px_01073_e, Px_15230_fz*) are shown in Supplementary Table 44 and Supplementary Fig. 39. PCR of primer-self amplification was performed with Q5 high-quality DNA polymerase (BioLabs) in 100 μl reaction volumes. 300 ng purified PCR-product was used as DNA template to perform *in vitro* transcription with the MAXIscript T7 kit (Life technology, USA) for 4 h at 37 °C. The Cas9 mRNA was transcribed using *NotI*-digested Cas9 expression vector **pTD1-T7-Cas9**[19] and the mMESSAGE mMACHINE T7 ULTRA kit (Life Technologies). After the transcription reaction, the poly (A) tailing reaction and *DNase I* treatment were performed according to the manufacturer’s instructions. Both the sgRNA and the Cas9-encoding mRNA were then purified by LiCl precipitation and redissolved in RNase-free water.

Fresh eggs were collected from the host plant leaves, dipped into clear water and then aligned on the microscope slide with a soft paintbrush and fixed with glue. 2 nl mix of sgRNA(s) and Cas9-encoding mRNA, at varying concentrations (Supplementary Tables 45, 47 and 49) were injected through the chorion into each egg under a dissecting microscope (Nikon SMZ800), using a TransferMan NK2 and FemtoJet microinjection system (Eppendorf, Germany). After injection, eggs were put in a petri dish and then placed in an incubator at 25°C and 70% relative humidity. When embryos hatched, host plant leaves were placed into the dishes for newly hatched larva feeding. The leaves with larvae were carefully transferred into large dishes. The phenotype of G_0 generation larvae was carefully checked using a dissecting microscope and photographed using a digital camera. Pupae were transferred into plastic baskets to eclose. We observed morphologic changes in different developmental stages, using microscope and by eye. For *Abd-B*, we mainly observed variation of the larval abdomen; for *ebony* (*e*), we primarily noted color variation of larvae and adults. For *frizzled* (*fz*), we mainly observed cuticular structures.

For *Abd-B* mutants of *P. xuthus*, genomic DNA extraction for the whole body of unhatched G_0 larvae dissected from developed eggs and subsequent PCR were carried out using an Animal Tissue Direct PCR kit (FOREGENE, China) following manual instruction; and genomic DNA of whole body of G_0 hatched larvae were extracted using Gentra Puregene Blood Kit (Qiagen, Germany), and exTaq polymerase was used in PCR amplification. The primer pairs included F3/R6, F9/R11, and F10/R11 (Supplementary Fig. 39a). For *ebony* mutants, whole body of larvae, prepupae and adult were used to extract genomic DNA, using Gentra Puregene Blood Kit (Qiagen, Germany), and exTaq polymerase was used in PCR amplification. The primer pairs were F2/R2 for amplifying target sites of T2-T303 (*Px_01073_e*-I, *Px_01073_e*-II) or F7/R4 for amplifying T454-T6 (*Px_01073_e*-III) (Supplementary Fig. 39b). For *frizzled* mutants, whole body of larvae were used to extract genomic DNA, using Gentra Puregene Blood Kit (Qiagen, Germany), and exTaq polymerase was used in PCR amplification. The primer pair for amplifying target sites is F2/R2 (Supplementary Fig. 39c). PCR products of target sites were detected with T7 endonuclease I (T7EI) as previously described[57]. T7EI positive and morphologically mutated individuals were further confirmed by Sanger sequencing 12 TA clones. We also performed quantitative reverse transcription-PCR (qRT-PCR) and western blot analysis to check effect of disruption of *Abd-B* gene on its expression at RNA or protein levels (Supplementary Note 13). In addition, we also performed qRT-PCR to check effect of *ebony* gene on its expression at RNA level (Supplementary Note 13).

For those targets resulted in morphological mutants, in order to exclude the possibility that these morphological mutants were induced by off-target cleavage events during the genome editing process, we further analyzed possible off-target sites in the genome using three methods CasOT[43], Cas-OFFinder[44] and COSMID[45] with the least stringent parameters, and carried out whole genome next generation sequencing validation for the *ebony* gene mutant (Supplementary Note 13). The resequencing reads of mutant have been uploaded to the NCBI Sequence Read Archive (SRA) and are available via the accession number SRA272356.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This project was funded by grants from the National Natural Science Foundation of China (No. 31321002), Chinese Academy of Sciences (XDB13000000) and from Yunnan Provincial Science and Technology Department (No. 2013HA004) to W.W. M.R.K and W.Z were supported by National Institutes of Health Grant GM108626-02. The authors thank D. Yang and Y. Peng for kindly providing their greenhouse, Y. Wu and C. Zhou, Z. Dong and M. Li for help or assistance in butterfly husbandry, R. Huang for kindly gift of Cas9 expression vector, A. Tan and X. Huang for helpful suggestions in Cas9 experiments, J. Lyu for fruitful discussion in data analysis, and Q. Li for helping in off-target identification.

References

1. Joron M, Mallet JLB. Diversity in mimicry: paradox or paradigm? Trends in ecology & evolution. 1998; 13:461–466. [PubMed: 21238394]
2. Brakefield PMFV. Butterfly wings: the evolution of development of colour patterns. BioEssays : news and reviews in molecular, cellular and developmental biology. 1999; 27:11.
3. Nijhout, HF. The Development and Evolution of Butterfly Wing Patterns. Smithsonian Institution Scholarly Press; 1991.
4. Boogs, CL.; Ehrlich, PR.; Watt, WB. Butterflies: Ecology and Evolution Taking Flight. University of Chicago Press; 2003.
5. Beldade P, Brakefield PM. The genetics and evo-devo of butterfly wing patterns. Nature Reviews Genetics. 2002; 3:442–452.
6. McMillan WO, Monteiro A, Kapan DD. Development and evolution on the wing. Trends in ecology & evolution. 2002; 17:125–133.
7. Zhan S, Merlin C, Boore JL, Reppert SM. The Monarch Butterfly Genome Yields Insights into Long-Distance Migration. Cell. 2011; 147:1171–1185. [PubMed: 22118469]
8. Dasmahapatra KK, et al. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature. 2012; 487:94–98. [PubMed: 22722851]
9. Ahola V, et al. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nature communications. 2014; 5:4737.
10. Cong Q, Borek D, Otwinowski Z, Grishin NV. Tiger Swallowtail Genome Reveals Mechanisms for Speciation and Caterpillar Chemical Defense. Cell reports. 2015; 10:910–919.
11. Nishikawa H, et al. A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. Nature genetics. 2015
12. Langley CH, Crepeau M, Cardeno C, Corbett-Detig R, Stevens K. Circumventing heterozygosity: sequencing the amplified genome of a single haploid Drosophila melanogaster embryo. Genetics. 2011; 188:239–246. [PubMed: 21441209]
13. You MS, et al. A heterozygous moth genome provides insights into herbivory and detoxification. Nature genetics. 2013; 45:220–225. [PubMed: 23313953]
14. Merlin C, Beaver LE, Taylor OR, Wolfe SA, Reppert SM. Efficient targeted mutagenesis in the monarch butterfly using zinc-finger nucleases. Genome research. 2013; 23:159–168. [PubMed: 23009861]
15. Cong L, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 2013; 339:819–823. [PubMed: 23287718]
16. Gaj T, Gersbach CA, Barbas CF 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in biotechnology. 2013; 31:397–405. [PubMed: 23664777]
17. Gratz SJ, et al. Genome Engineering of Drosophila with the CRISPR RNA-Guided Cas9 Nuclease. Genetics. 2013; 194 1029-+
18. Bassett AR, Tibbit C, Ponting CP, Liu JL. Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System. Cell reports. 2013; 4:220–228. [PubMed: 23827738]

19. Wang YQ, et al. The CRISPR/Cas System mediates efficient genome engineering in Bombyx mori. Cell research. 2013; 23:1414–1416. [PubMed: 24165890]

20. Ma SY, et al. CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of BmKu70 in Bombyx mori. Scientific reports. 2014; 4

21. Linnaeus C. Systema naturae per regna tria naturae :secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Editio decima, reformata. 1758

22. Cusson M, et al. Characterization and tissue-specific expression of two lepidopteran farnesyl diphosphate synthase homologs: implications for the biosynthesis of ethyl-substituted juvenile hormones. Proteins. 2006; 65:742–758. [PubMed: 16972283]

23. Vandermoten S, Haubruge E, Cusson M. New insights into short-chain prenyltransferases: structural features, evolutionary history and potential for selective inhibition. Cellular and molecular life sciences : CMLS. 2009; 66:3685–3695. [PubMed: 19633972]

24. Kinjoh T, Kaneko Y, Itoyama K, Mita K, Hiruma K, Shinoda T. Control of juvenile hormone biosynthesis in Bombyx mori: cloning of the enzymes in the mevalonate pathway and assessment of their developmental expression in the corpora allata. Insect biochemistry and molecular biology. 2007; 37:808–818. [PubMed: 17628279]

25. Peroutka R, Schulzova V, Botek P, Hajslova J. Analysis of furanocoumarins in vegeTables (Apiaceae) anal citrus fruits (Rutaceae). Journal of the science of food and agriculture. 2007; 87:2152–2163.

26. Danielson PB, MacIntyre RJ, Fogleman JC. Molecular cloning of a family of xenobiotic-inducible drosophilid cytochrome p450s: evidence for involvement in host-plant allelochemical resistance. Proceedings of the National Academy of Sciences of the United States of America. 1997; 94:10797–10802. [PubMed: 9380713]

27. Janzen DH, Hallwachs W, Burns JM. A tropical horde of counterfeit predator eyes. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:11659–11665. [PubMed: 20547863]

28. Kodandaramaiah U, Vallin A, Wiklund C. Fixed eyespot display in a butterfly thwarts attacking birds. Animal behaviour. 2009; 77:1415–1419.

29. Schaefer HM, Stobbe N. Disruptive coloration provides camouflage independent of background matching. Proceedings of the Royal Society B-Biological Sciences. 2006; 273:2427–2432.

30. Tullberg BS, Merilaita S, Wiklund C. Aposematism and crypsis combined as a result of distance dependence: functional versatility of the colour pattern in the swallowtail butterfly larva. Proceedings of the Royal Society B-Biological Sciences. 2005; 272:1315–1321.

31. kawooya JK, Keim PAS, Law JH, Riley CT, Ryan RO, Shapiro JP. Why are green caterpillars green? ACS Symposium Series. 1985; 276:11.

32. Shirataki H, Futahashi R, Fujiwara H. Species-specific coordinated gene expression and trans-regulation of larval color pattern in three swallowtail butterflies. Evolution & development. 2010; 12:305–314. [PubMed: 20565541]

33. Kunte K, et al. doublesex is a mimicry supergene. Nature. 2014; 507 229-+

34. Goodrich LV, Strutt D. Principles of planar polarity in animal development. Development. 2011; 138:1877–1892. [PubMed: 21521735]

35. Povelones M, Howes R, Fish M, Nusse R. Genetic evidence that drosophila frizzled controls planar cell polarity and armadillo signaling by a common mechanism. Genetics. 2005; 171:1643–1654. [PubMed: 16085697]

36. Repiso A, Saavedra P, Casal J, Lawrence PA. Planar cell polarity: the orientation of larval denticles in Drosophila appears to depend on gradients of Dachsous and Fat. Development. 2010; 137:3411–3415. [PubMed: 20826534]

37. Tomita S, Kikuchi A. Abd-B suppresses lepidopteran proleg development in posterior abdomen. Developmental biology. 2009; 328:403–409. [PubMed: 19389347]

38. ffrench-Constant, R.; Koch, PB. Mimicry and melanism in swallowtail butterflies: toward a molecular understanding. In: Boggs, CL.; Watt, WB.; Ehrlich, PR., editors. Butterflies: Ecology And Evolution Taking Flight. Chicago University Press; 2003. (ed^

39. Takahashi A. Pigmentation and behavior: potential association through pleiotropic genes in Drosophila. Genes Genet Syst. 2013; 88:165–174. [PubMed: 24025245]

40. Futahashi R, et al. yellow and ebony Are the Responsible Genes for the Larval Color Mutants of the Silkworm Bombyx Mori. Genetics. 2008; 180:1995–2005. [PubMed: 18854583]
41. Wittkopp PJ, True JR, Carroll SB. Reciprocal functions of the Drosophila Yellow and Ebony proteins in the development and evolution of pigment patterns. Development. 2002; 129:1849–1858. [PubMed: 11934851]
42. Sakai RK, Baker RH, Iqbal MP. Genetics of Ebony, a Nonlethal Recessive Melanotic Mutant in a Mosquito. Journal of Heredity. 1972; 63 275-&.
43. Xiao A, et al. CasOT: a genome-wide Cas9/gRNA off-target searching tool. Bioinformatics. 2014; 30:1180–1182.
44. Bae S, Park J, Kim JS. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics. 2014; 30:1473–1475. [PubMed: 24463181]
45. Cradick TJ, Qiu P, Lee CM, Fine EJ, Bao G. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. Mol Ther-Nucl Acids. 2014; 3
46. Veres A, et al. Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing (vol 15, pg 27, 2014). Cell Stem Cell. 2014; 15:254–254.
47. Yang LH, et al. Targeted and genome-wide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells. Nature communications. 2014; 5
48. Arikawa KSD, Fujii T. Hindsight by genitalia: photo-guided copulation in butterflies. J comp Physiol A. 1997; 180:5.
49. Lederhouse RCAMP, Scriber JM. Evaluation of spermatorphore counts in studying mating systems of Lepidoptera. Journal of the Lepidopterists' Society. 1989; 43:9.
50. Salmela L, Schroder J. Correcting errors in short reads by multiple alignments. Bioinformatics. 2011; 27:1455–1461. [PubMed: 21471014]
51. Margulies M, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437:376–380. [PubMed: 16056220]
52. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011; 27:578–579. [PubMed: 21149342]
53. Li R, et al. De novo assembly of human genomes with massively parallel short read sequencing. Genome research. 2010; 20:265–272. [PubMed: 20019144]
54. Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol Biol. 2011; 772:157–178. [PubMed: 22065437]
55. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution. 2007; 24:1586–1591. [PubMed: 17483113]
56. Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, Dobbs D. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic acids research. 2010; 38:W462–W468. [PubMed: 20435679]
57. Guschin, DY.; Waite, AJ.; Katibah, GE.; Miller, JC.; Holmes, MC.; Rebar, EJ. A Rapid and General Assay for Monitoring Endogenous Gene Modification. In: Mackay, JP.; Segal, DJ., editors. Engineered Zinc Finger Proteins, Methods in Molecular Biology. Vol. 9. 2010.
58. Heikkila M, Kaila L, Mutanen M, Pena C, Wahlberg N. Cretaceous origin and repeated tertiary diversification of the redefined butterflies. Proceedings Biological sciences / The Royal Society. 2012; 279:1093–1099. [PubMed: 21920981]
59. Maeki K. A Use of Chromosome Numbers in the Study of Taxonomy of the Lepidoptera and Notes on the Internal Reproductive Anatomy (With 6 Text-figures). JOURNAL OF THE FACULTY OF SCIENCE HOKKAIDO UNIVERSITY Series VI ZOOLOGY. 1957; 13:4.
60. Zhan S, Reppert SM. MonarchBase: the monarch butterfly genome database. Nucleic acids research. 2013; 41:D758–D763. [PubMed: 23143105]
61. Xia QY, et al. A draft sequence for the genome of the domesticated silkworm (Bombyx mori). Science. 2004; 306:1937–1940. [PubMed: 15591204]

62. Gregory TR, et al. Eukaryotic genome size databases. Nucleic Acids Res. 2007; 35:D332–D338. [PubMed: 17090588]

Author Manuscript

Author Manuscript

Author Manuscript

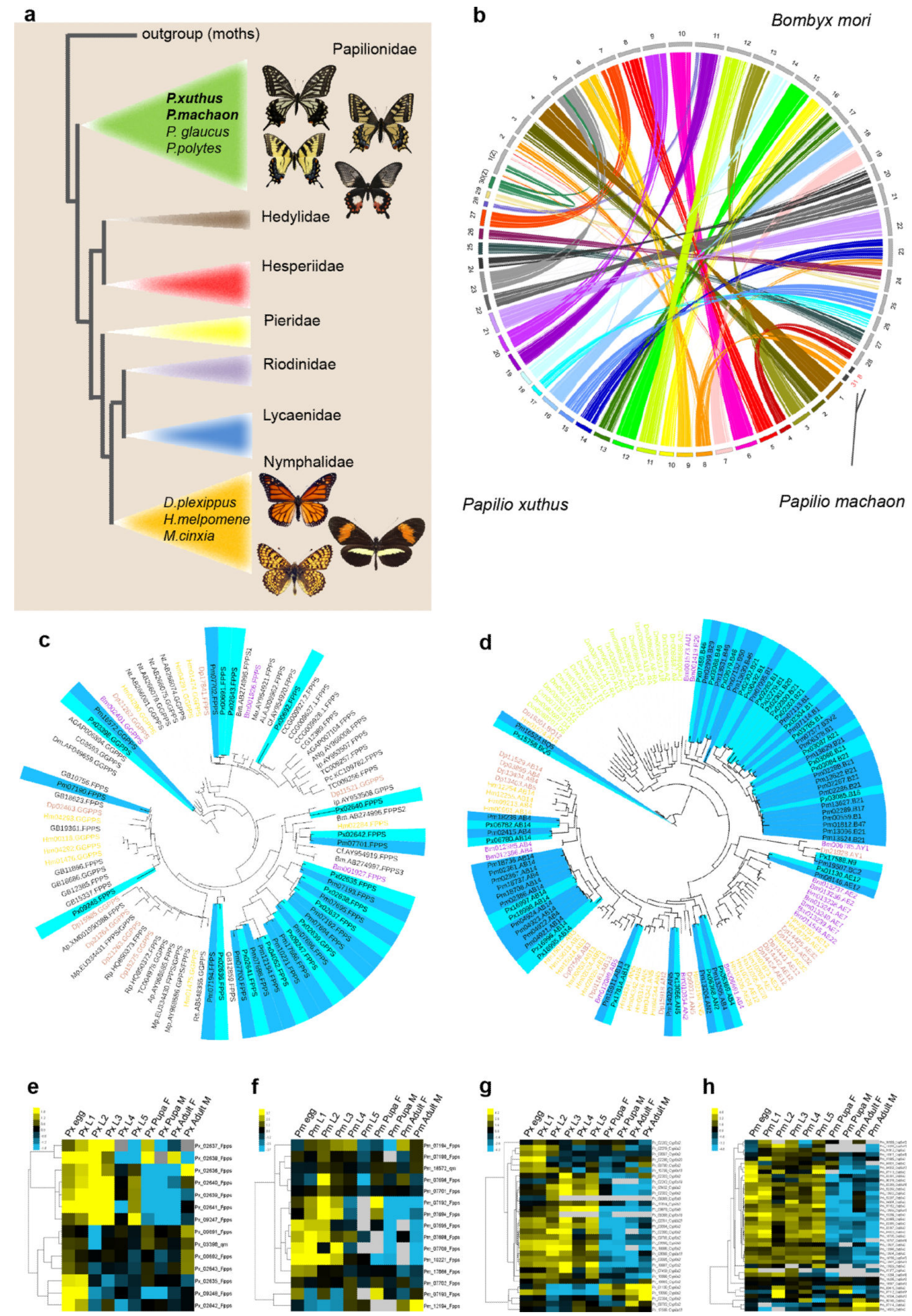

Figure 1. Butterfly comparative genomics

a, Phylogeny of butterfly families[58] showing relationship of *Papilio xuthus* and *P. machaon* to *Danaus plexippusHeliconius melpomeneMelitaea cinxiaP. glaucusP. polytes***b**, Chromosome mapping of *P. xuthus* (n=30) to *Bombyx mori* (n=28). For *P. machaon* (n=31) we only plot chromosome 8 (chr8) and chr31 which were fused in *P. xuthus***c**, Maximum-likelihood tree showing strong expansion of the short-chain isoprenyl diphosphate synthase (scIPPS) genes in the genomes of swallowtail butterflies. Included are all scIPPS genes identified in the genomes of ten holometabolous insects (*P. xuthus*, *P. machaonD.*

plexippusH. melpomene, B. moriPlutella xylostellaAnopheles gambiaeDrosophila melanogasterTribolium castaneumApis mellifera). The clades of *P. xuthus* and *P. machaon* are highlighted by light blue and deep blue, respectively. **d**, Maximum-likelihood tree showing expansions of the cytochrome P450 6B subfamily (CYP6B) genes in the genomes of *P. xuthus*and *P. machaon* and of the CYP6AB genes in the genome of *P. machaon*, as compared to CYP6 genes of *D. plexippus H. melpomene B. mori*, and *D. melanogaster*. In **c** and **d**, the clades of *P. xuthus* and *P. machaon* are highlighted by light blue and deep blue, respectively. **e–f**, Expression profiles of farnesyl diphosphate synthase (FPPS) and geranylgeranyl diphosphate synthase (GGPPS) genes at all development stage of *P. xuthus* (**e**) and *P.machaon* (**f**). **g–h**, Expression profiles of CYP6 genes at all development stage of *P. xuthus* (**g**) and *P machaon* (**h**). Expression measured in reads per kilobase of transcript per million reads mapped (RPKM).

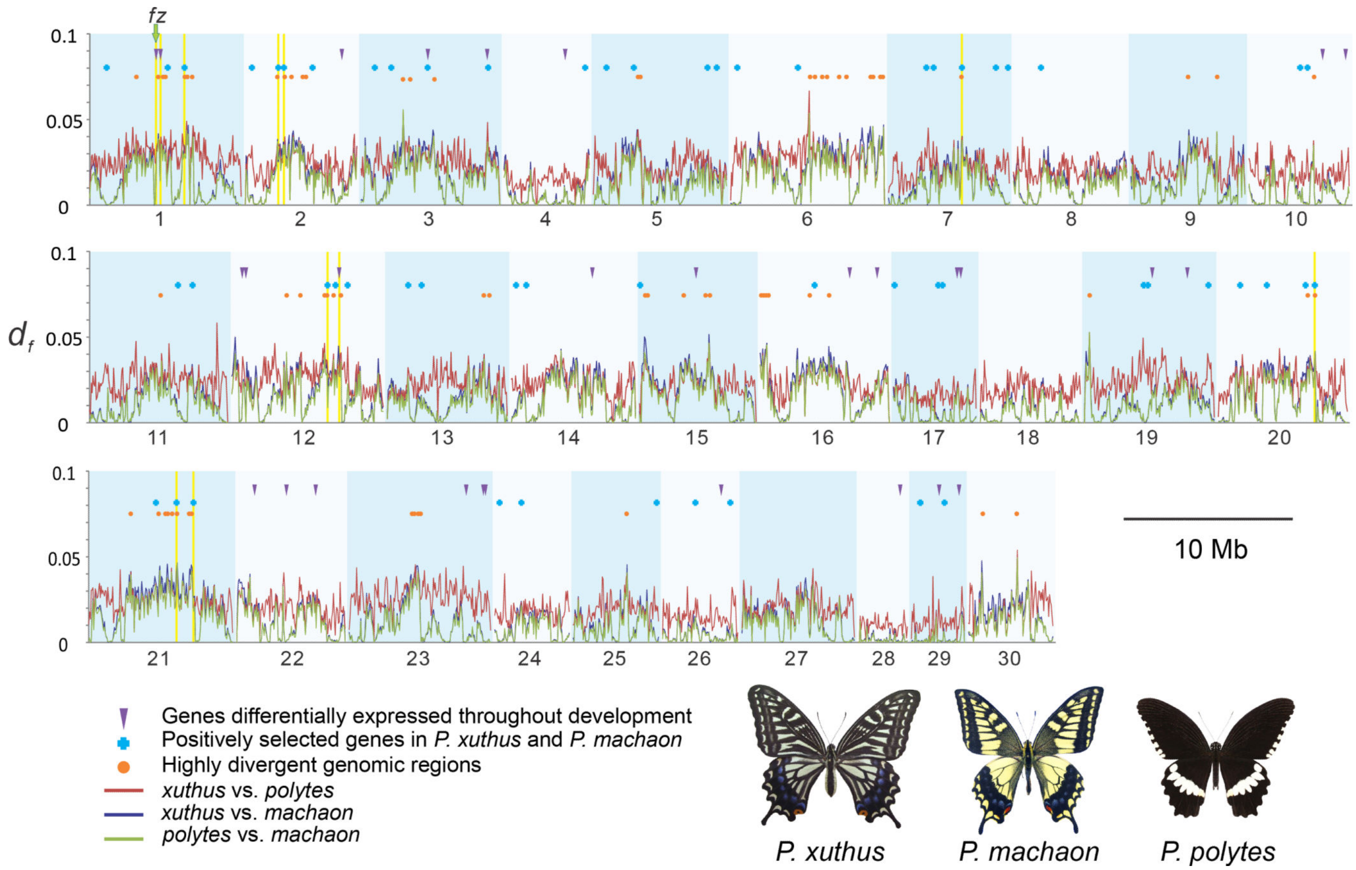

Figure 2. Genome-wide divergence among *Papilio* species

Pairwise genetic divergence among *P. xuthus P. machaon*, and *P. polytes* was calculated in 50-kb windows. This yielded 70 highly divergent genomic regions, windows in the upper 95th percentile in all three comparisons, encompassing 915 genes. When overlaid with signatures of positive selection and differential expression throughout development, 11 genes (highlighted in yellow) emerged as strong candidates for a role in recent diversification, including *frizzled*. Degree of freedom (df) is the density of different SNPs per bp of each 50-kb window.

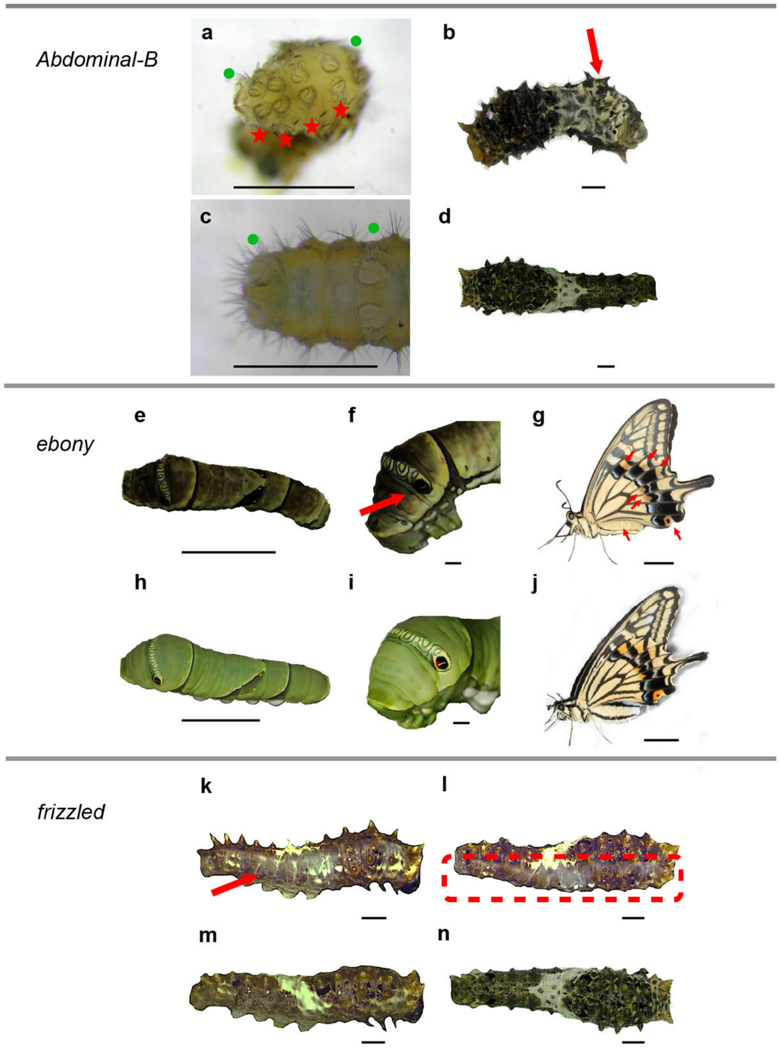

Figure 3. CRISPR/Cas9-induced morphological mutants of *Abdominal-B, ebony* and *frizzled* in *Papilio xuthus*

a–d, Mutations induced by the injection of Cas9 mRNA and *Abdominal-B* sgRNA with prolegs on segment A7 – A10 (**a**), segments that do not normally have prolegs (**c**), or with a curled abdomen resulting from abnormal terga on segments A3 and after (**b**), compared with wild type (**d**). Green dots in (**a)** and (**c)** show normal prolegs of A6 and A10, while red stars in (**a)** denote the redundant prolegs on A7-A10. **e–j**, Mutations induced by the injection of Cas9 mRNA and *ebony* sgRNA with enhanced melanic pigmentation (**e**) and an absence of orange color in the false eyespot (**f**) in fifth instar larvae, compared with wild type (**h, i**); and with brown pigmentation (**g**) across the body in regions that were normally yellow and in wing patches that were normally orange in adult (**j**). **k–n**, A mutant induced by the injection of Cas9 mRNA and *frizzled* sgRNA with smooth and colorless dorsal cuticle in right side (**k**) while normal in its left side (**m**), and its dorsal view (**l**) compared to wild type (**n**). Scale bars: 1mm in panels **a–d, f, i**, and **k-n**; 10 mm in panels **e, g, h** and **j**.

Table 1

Comparison of reference genomes among Lepidopterans.

	Pm	Px	Hm	Dp	Bm	PLX
Chromosome Number (n)	31[a]	30[a]	21[b]	29–30[c]	28[e]	31[f]
Genome size (Mb) based on the bases in contigs and in scaffolds	265/281	231/244	274/269[b]	242[c]/249[d]	432/481[e]	385/394[f]
Genome size (Mb) measured by C-value	234~256	218~238	300[g]	290[g]	520[g]	NA
N50 of contig (Kb)/scaffold (Mb)	**81/1.15**	**492/3.4**	51/0.2[b]	51[c]/0.7[d]	13/3.7[e]	49/0.72[f]
TE (% genome)	21.10	20.26	28.47[b]	12.21[c]	21.1[e]	33.97[f]
Number of protein-coding genes	15,499	15,322	15,984[h]	16,866[c]	14,623[e]	18,071[f]

Abbreviation: *Papilio machaon* (*Pm*), *P. xuthus* (*Px*), *Heliconius melponene* (*Hm*), *Danaus plexippus* (*Dp*), *Bombyx mori* (*Bm*), *Plutella xylostella* (*PLX*).

[a] From (Maeki, 1957)[59];

[b] From (Dasmahapatra et al., 2012)[8];

[c] From (Zhan et al., 2011)[7];

[d] From (Zhang et al., 2013)[60];

[e] From (Xia et al., 2004)[61];

[f] From (You et al., 2013)[13];

[g] From (Gregory et al., 2007)[62];

[h] Predicted in this study using the same pipeline as in *Pm* and *Px*.

Table 2

Mutagenesis efficiency of knocking out genes *Abdominal-B*, *ebony*, and *frizzled* in *P. xuthus*.

Gene	Target sites*	RNA con. (ng per μl)**	Phynotype rate (%)***	Mutation rate (%)	Mutation type****
Px_03961_Abd-B	T42	200; 300	7.95 (7/88)	35.38	D: 3–51; I: 4–18
		600; 600	8.26 (10/121)	25	D: 3–66; I; 4–5; M: 40
	T42/T95	50/50; 100	6 (6/100)	18.33	D: 5–59; I: 3–19
		566/416; 1200	91.79 (123/134)	90.85	D: 3–73; I: 3–64; M: 3–29
Px_01073_e	T2/T303	158/159; 1200	31.25 (5/16)	66.67	D: 5–33; I; 3–10; M; 3
	T454/T6	200/150; 1200	88 (22/25)	29.92	D: 6–62; I; 3–15; M; 3–8
Px_15230_fz	T268/T283	500/500; 1000	4.12 (4/96)	45.16	D: 3–64; I; 3–15; M; 4–5

* "/" denoting injection of mixed targets;

** the number before and after semicolon denotes injected concentration of sgRNA and Cas9 mRNA, respectively;

*** the number in the bracket denotes mutant/observed injected individuals;

**** D: delete fragment (bp); I: insert fragment (bp); M: mutated fragment (bp).