Genomics 116 (2024) 110956

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Genomic evolution and patterns of horizontal gene transfer in *Papilio*

Jiajia Wang, Yunfei Wu, Linxin Zhu, Kaixin Guo, Shichen Gao, Yan Dong *

College of Biology and Food Engineering, Chuzhou University, Chuzhou 239000, China

ARTICLE INFO

Keywords:
Genomic evolution
Horizontal gene transfer
Papilio
ABC gene family

ABSTRACT

The Papilio genus, known for its ecological and phenotypic diversity, is a valuable model for evolutionary studies. This study conducted a comparative genomic analysis of 11 *Papilio* species, revealing species-specific gene family expansions, including the UDP-glucosyltransferase 2 gene associated with insect detoxification, particularly expanding in *Papilio polyxenes*. Our analysis also revealed 199 horizontal gene transfer (HGT) acquired genes from 76 microbial species, with *Pseudomonadota* and *Bacillota* as common HGT donors across these genomes. Furthermore, we examined the evolutionary patterns of nine ABC transporter subfamilies, uncovering potential links between gene family evolution and environmental adaptation. This study provides new insights into evolutionary relationships and genomic adaptations within the *Papilio* genus, contributing to broader butterfly evolutionary research.

1. Introduction

Butterflies, with their rich diversity in appearance and variable behaviors, hold significant value for both observation and scientific research [1,2]. Since the time of Darwin, they have been considered a key species for studying adaptive evolution [3]. In recent years, advances in high-throughput sequencing and genomic analysis have established butterflies as ideal models for studying morphogenesis, evolution, and development. Numerous studies have investigated the associations between butterfly genomes and their phenotypic traits [4,5], driving progress in fields such as evolutionary biology, population genetics, and ecology.

The *Papilio* genus holds significant research value, and as one of the most well-known and extensively studied groups of insects, it serves as a model for evolutionary biology, ecology, genomics, and conservation biology [6]. Numerous studies have explored various *Papilio* species through genome assembly and comparative genomics. For instance, a genome study on *Papilio bianor*, widely distributed across east Asia, south Asia, and southeast Asia, has provided insights into body color evolution and species differentiation [7]. This research assembled a chromosome-level genome for *P. bianor*, which includes 29 autosomes and one sex chromosome. Furthermore, it was assembled using high-throughput chromosome conformation capture techniques, serves as a vital reference for genomic studies of the *Papilio* genus. Martijn et al. conducted a comparative genomic analysis of *Papilio dardanus*, revealing that a single gene is responsible for variations in wing patterns [8]. However, their analysis was based on a population study of a single species and did not explore the evolutionary characteristics among different species within the *Papilio* genus. Similarly, Pan et al. assembled the genome of *Papilio elwesi Leech* and performed a comparative genomic analysis [9], but their study only examined gene family expansions and contractions in six *Papilio* species without delving into further in-depth analysis.

Horizontal Gene Transfer (HGT) refers to the direct transfer of genes between different species, distinct from traditional evolutionary processes that involve the vertical inheritance of genes. While this phenomenon is well-documented in prokaryotes, recent studies indicate that HGT also frequently occurs in insects. HGT can substantially influence genome composition, offering new evolutionary pathways for insects to adapt to various environments and ecological niches. For example, insects have acquired plant genes through HGT to metabolize harmful phenolic glycosides, enhancing their adaptability to a wide range of host plants [10]. Through these transferred genes, insects gain complex biological capabilities essential for survival across diverse ecosystems, thereby increasing their resilience to environmental changes. Advances in genome sequencing have revealed numerous potential HGT events in insect genomes. For instance, a study combining 218 high-quality insect genomes identified 1410 HGT-acquired genes, one of which has been experimentally shown to assist in male courtship behavior [11].

In this study, we used the *Papilio bianor* genome as a reference to improve the genomes of 9 other Papilio species and re-annotated the

* Corresponding author.
E-mail address: dongyan@chzu.edu.cn (Y. Dong).

https://doi.org/10.1016/j.ygeno.2024.110956
Received 4 July 2024; Received in revised form 14 October 2024; Accepted 2 November 2024
Available online 12 November 2024

improved genomes. Through comparative genomic analysis, this study estimated the divergence times among various *Papilio* species and identified both shared and unique genes, as well as gene family expansions and contractions across the eleven species. Additionally, 199 HGT-acquired genes were identified within these upgraded *Papilio* genomes, and functional enrichment analysis suggested potential roles for these genes. With the *ABC* gene family as an example, we explored its evolutionary patterns across the 11 *Papilio* species. These findings provide valuable insights into the genomic evolution of *Papilio* species and the variation of functional genes, serving as a reference for further research.

2. Methods

2.1. Data acquisition

The genomes of 11 different species within Papilio were obtained from the NCBI Assembly database, including *P. bianor* (PRJNA530186), *P. machaon* (PRJEB47027), *P. aristodemus* (PRJNA434126), *P. dardanus tibullus* (PRJNA451133), *P. demoleus* (PRJNA714807), *P. glaucus* (PRJNA270125), *P. helenus* (PRJNA609337), *P. memnon* (PRJNA345222), *P. polyxenes* (PRJNA345143), *P. protenor* (PRJNA714807), and *P. xuthus* (PRJNA270384). Among these, the genome of the *P. bianor* was assembled to chromosome level using long-read single-molecule sequencing and Hi-C based chromatin interaction maps [7].

Protein sequences and coding sequences (cds) of *B. mori* were sourced from SilkDB 3.0 (https://silkdb.bioinfotoolkits.net/main/species-info/-1) [12]. The protein sequences and cds of the *H. armigera* were obtained from the study by Pearce et al. [13].

2.2. Genome scaffolding and improving

In the presence of high-quality reference genomes from closely related species (preferably within the same genus), RagTag (v2.1.0) can be employed to enhance the quality and completeness of other genomes of low-quality using a high-quality genome as a reference [14, 15]. Given that the genome of the *P. bianor* and *P. machaon* are at the chromosomal level, and only *P. bianor* has a chromosome-level genome assembled using a combination of Hi-C and third-generation sequencing strategies. This study utilized the genome of the *P. bianor* as a reference and employed RagTag with ragtag correct and scaffold model to perform scaffolding and improvement on the genomes of the other ten species within *Papilio*. The completeness of the assemblies was evaluated by BUSCO v5.4.6 [16] with the insecta_odb9 BUSCO set.

2.3. Genome annotation

We collected all available transcriptome datasets for *Papilio* species currently in the database, obtaining a total of 134 RNA-seq datasets from 128 samples across seven *Papilio* species from NCBI (Supplementary Table S1, Supplementary Table S2), and used fastp to remove low-quality sequences followed by de novo assembly with Trinity [17,18]. The assembled transcripts were merged, then redundancy was eliminated using cd-hit-est [18]. Repeat sequences in ten upgraded *Papilio* genomes were annotated using RepeatMasker (v4.1.2) (http://repeatmasker.org) with the parameters "-nolow -no_is -norna -engine ncbi" based on RepBase (version 17.01, http://www.girinst.org/repbase). Concurrently, RepeatModeler (v1.0.8) was employed with the parameters "-engine ncbi" to build a de novo library of repeat sequences for each species [19], which was subsequently annotated again with RepeatMasker (−nolow -no_is -norna -engine ncbi). In addition, Tandem Repeats in each genome were annotated using Tandem Repeats Finder (v4.09) with parameters "2 7 7 80 10 50 2000 -d -h" [20]. The model was trained using Augustus with the genome and structural annotation information of *P. bianor* [21]. Finally, the gene structure annotation for each genome was conducted using the MAKER2 (v3.01.03) annotation pipeline [22].

2.4. Calculation of Ks for homologous genes

Paralogous genes were identified by conducting BLASTP reciprocal best hit (RBH) pairwise comparisons within the species genome using BLAST v2.14.0. [23]. The synonymous substitution rates (Ks values) for both orthologous and paralogous gene pairs were calculated using ParaAT [24] and KaKs Calculator v2.0 [25]. Density plots were drawn using the R ggridges package [26].

2.5. Differentiation time and analysis of contraction and expansion of gene families

The *B. mori* and the *H. armigera* were selected as outgroups, and species-specific single-copy genes were identified using OrthoFinder (v2.5.5) [27], and the longest protein was selected to represent loci with multiple transcripts. A phylogenetic tree was constructed using Easy-SpeciesTree v1.0. The concatenation phylogenetic tree obtained in the previous step serves as the input file for r8s [28] to calculate a timetree. The divergence time between the *B. mori* and the *H. armigera* was queried on Timetree (http://timetree.org/) [29], and the node for *B. mori - H. armigera* in the tree was calibrated to 95 Ma.

The ultrametric tree obtained from r8s was incorporated into CAFE v4.2 [30] with parameters "-k 3 -p –pvalue 0.05". This model employs a stochastic birth and death model to estimate the size of each family at each ancestral node and obtains family-wise *p*-values (based on a Monte–Carlo re-sampling procedure) to indicate whether each gene family across species has undergone significant expansion or contraction.

2.6. HGT-acquired genes identification

The identification of HGT-acquired genes was performed according to the method described by Li et al. [31]. In brief, protein sequences of 11 *Papilio* species were separately carried out a BLASTP in DIAMOND v0.9.19.120 [32] with the Refseq database (version 23–5-5) and 215 insect species protein sequences, including the species of *Papilio*. (Supplementary file 1), with an e-value threshold set to 1e-5 and maximum target sequences set to 1000. Subsequently, the Alien Index (AI) and the out_pct (the percentage of species from the OUTGROUP lineage) were calculated for each gene using HGTfinder v1.0 [33].

Genes with out_pct $\geq$ 65 and AI >0 were assumed as HGT- acquired genes candidates, and phylogenetic trees were constructed for the HGT-acquired genes candidates to confirm the candidates (There was substantial topological disagreement between the gene tree and its associated species tree and the *Papilio* gene sequence was robustly nested within the donor lineage in the gene tree). Multiple sequence alignment was performed using MAFFT v7.508 [34] with the default settings, the alignment results were trimmed using trimAL v1.4.rev15 [35] with default settings, and phylogenetic trees were constructed using IQ-TREE v2.2.0.3 [36]. The resulting trees were visualized using the R package ggtree [37].

The final HGT-acquired genes were filtered according to the following criteria: (1) There was substantial topological disagreement between the gene tree and its associated species tree and the *Papilio* gene sequence was robustly nested within the donor lineage in the gene tree. (2) Genes located on contigs with a length greater than 100 kb; (3) Similarity with non-metazoan organisms proteins less than 90 %; (4) The proportions of HGT-acquired genes in the contigs that harbor the HGT-acquired genes less than 5 %.

2.7. Functional annotation and enrichment of genes

KEGG pathway analysis was conducted using kobas 2.0 [38], and

protein domain identification was performed using HMMER 3.3.2 [39] with parameters "–noali -E 1e-5" to search the InterProScan database [40] (https://www.ebi.ac.uk/interpro/) with Pfam-A HMMs. Gene annotation for GO terms was carried out using Eggnog 2.0.1 [41].

2.8. Identification of the ABC gene family

We identified the ABC gene family protein sequences of 11 *Papilio* species using ABC scan [42] and obtained their subfamilies. Multiple sequence alignments of ABC transporter protein sequences were performed using MAFFT v7.508 [34] with the default settings, and ambiguously aligned regions removed using trim-AI [35] with default settings. The best model, VT, was selected with prottest v3.4.2 [43], followed by phylogenetic analysis using RAxML [44] with parameters '-m PROTGAMMAVT -p 12345'based on the VT model. The tree were visualized using iTOL v6 [45].

Gene co-linearity relationships among different species were identified using OrthoFinder (v2.5.5) [27]. For the orthologs ABC transporters, Ka/Ks analysis were done using KaKs Calculator v2.0 version [46] with NG method [47].

3. Results

3.1. Genome assembly and evolutionary analysis of Papilio

3.1.1. Genome assembly

This study obtained 11 *Papilio* genomes from the NCBI Assembly database and upgraded the genomes of 10 species using the *P. bianor* genome as a reference (Supplementary Table S3). Among the 11 genomes, *P. bianor* has the largest genome size at 421 Mb, while *P. dardanus tibullus* has the smallest at 231 Mb (Fig. 1A). A simple linear regression analysis of genome size revealed a significant correlation between the amount of repeat sequence and genome size (Supplementary Fig. S1), suggesting that the content of repeat sequence may be one of the main drivers of genome size variation within *Papilio* species. Additionally, using 128 *Papilio* RNA-seq datasets, we annotated the genomes and found that *P. polyxenes* has the highest gene count (16,570 genes), while *P. machaon* has the lowest (13,379 genes). These results indicate substantial variations among the 11 *Papilio* genomes and suggest that changes in genome size do not necessarily correlate with changes in gene number. This implies that different evolutionary patterns may exist among protein-coding genes and other genomic elements across these species.

3.1.2. Evolutionary analysis

In this study, we identified a total of 2398 single-copy genes in the genomes of the *Bombyx mori*, *Helicoverpa armigera*, and 11 *Papilio* species. Additionally, using the *Bombyx mori* and *Helicoverpa armigera* as outgroups, we constructed a phylogenetic tree based on these single-copy genes and estimated their divergence times (Fig. 1B). The results are consistent with previous studies. For instance, we determined the divergence time between *Papilio protenor* and *Papilio memnon* to be 7.8 Mya, which is consistent with the results reported by Condamine et al. [6], supporting the reliability of our analysis. Furthermore, *P. glaucus* and *P. aristodemus* diverged from the common ancestor of *Papilio* earlier, around 21.4 million years ago. *P. machaon* and *P. polyxenes* belong to a branch where their most recent common ancestor is closer than that of other *Papilio* species. Their speciation occurred approximately 4.6 million years ago, making it the most recent speciation event among the 11 *Papilio* species. The distribution of Ks values for the paralogous genes among the 11 *Papilio* genomes indicates that they have undergone two whole-genome duplication (WGD) events. The peaks in the distribution suggest varying gene loss rates among the 11 species, potentially due to different selective pressures they have encountered (Fig. 1C). The Ks values of syntenic genes between two species can reflect the rate of genomic evolution. Therefore, we calculated the Ks values for syntenic genes between pairs of *Papilio* species (Fig. 1D and Supplementary Fig. S2). Most *Papilio* species have Ks value densities ranging from 0 to 1. However, the Ks values for *P. protenor* and *P. aristodemus* peak at the far right, indicating they have more distant evolutionary relationship with *P. xuthus* (Fig. 1D).

This study used CAFE to analyze the expansion and contraction of gene families in 11 *Papilio* species and two outgroups (*H. armigera* and *B. mori*) (Fig. 2A). We found that *P. polyxenes* exhibited the most extensive gene family expansion, with a total of 1348 gene families. Notably, KEGG enrichment analysis of these expanded gene families indicated that *P. polyxenes* experienced significant expansion of genes related to detoxification and metabolism of xenobiotic substances during its evolution. These genes are associated with pathways such as metabolism of xenobiotics by cytochrome P450 (13 genes), drug metabolism − other enzymes (6 genes), and drug metabolism − cytochrome P450 (13 genes) (Fig. 2B). In addition to *P. polyxenes*, several other *Papilio* species, particularly *P. aristodemus* and *P. memnon*, also showed enrichment in these KEGG pathways, although they have fewer expanded gene families compared to *P. polyxenes*. Specifically, *P. aristodemus* and *P. memnon* have 573 and 736 expanded gene families, respectively. However, the number of annotated genes involved in the aforementioned functions is greater in these species. For instance, in the *P. memnon* genome, there are 27, 20, and 27 genes involved in the three respective pathways. Furthermore, the expanded genes in different species are also enriched in pathways related to energy metabolism, visual perception, and other functions (Supplementary Table S4). These processes are closely linked to insect growth, development, and survival. While several gene families have been expanded across various critical functions, *P. aristodemus* exhibited the most contraction, with 1179 gene families undergoing reduction.

In this study, we identified a total of 23,528 gene families across the 11 *Papilio* genomes. Among them, *P. polyxenes* had the highest number of gene families (12,290) (Fig. 2C), while *P. machaon* had the fewest (10,691). There are 6549 gene families that are shared across all 11 *Papilio* genomes, while 480 gene families are unique to individual species (Fig. 2D). *P. bianor* and *P. memnon* exhibit the most unique genes. Enrichment analysis of these unique genes revealed their associations with physiological processes related to lifespan, defense mechanisms, and stress responses (Supplementary Table S5).

3.2. HGT-acquired genes in the genomes of Papilio

3.2.1. Identification of the HGT-acquired genes

HGT is widely present in both prokaryotic and eukaryotic genomes, and it can enhance the capabilities of organisms in environmental adaptation and reproduction. In order to gain further insights into the genomic evolution of *Papilio*, this study identified HGT-acquired genes in 11 species within *Papilio*. Contigs with a length of less than 100 kb were excluded for increased accuracy (Supplementary Table S6). The overall distribution of contig lengths in each genome showed that 142 contigs containing 199 HGT-acquired genes were longer than 226 contigs without HGT-acquired genes (Fig. 3A). Additionally, none of the 142 contigs contained HGT-acquired genes with a frequency greater than 0.2 % (Fig. 3B). The protein sequence similarity between HGT-acquired genes in *Papilio* receptors and the closest homologous genes in non-metazoan donors ranged from 26.5 % to 85 %, with an average of 53.3 % (Fig. 3C).

3.2.2. The origin of HGT-acquired genes

In *Papilio*, a total of 199 HGT-acquired genes from 76 microbial species were identified. These HGT-acquired genes primarily originated from bacteria, with only five coming from archaea. Within bacteria, *Pseudomonadota* and *Bacillota* were the primary donor species for the 11 HGT-acquired genes in *Papilio* (Fig. 3D). However, there were also significant differences in HGTs between different species within *Papilio*.

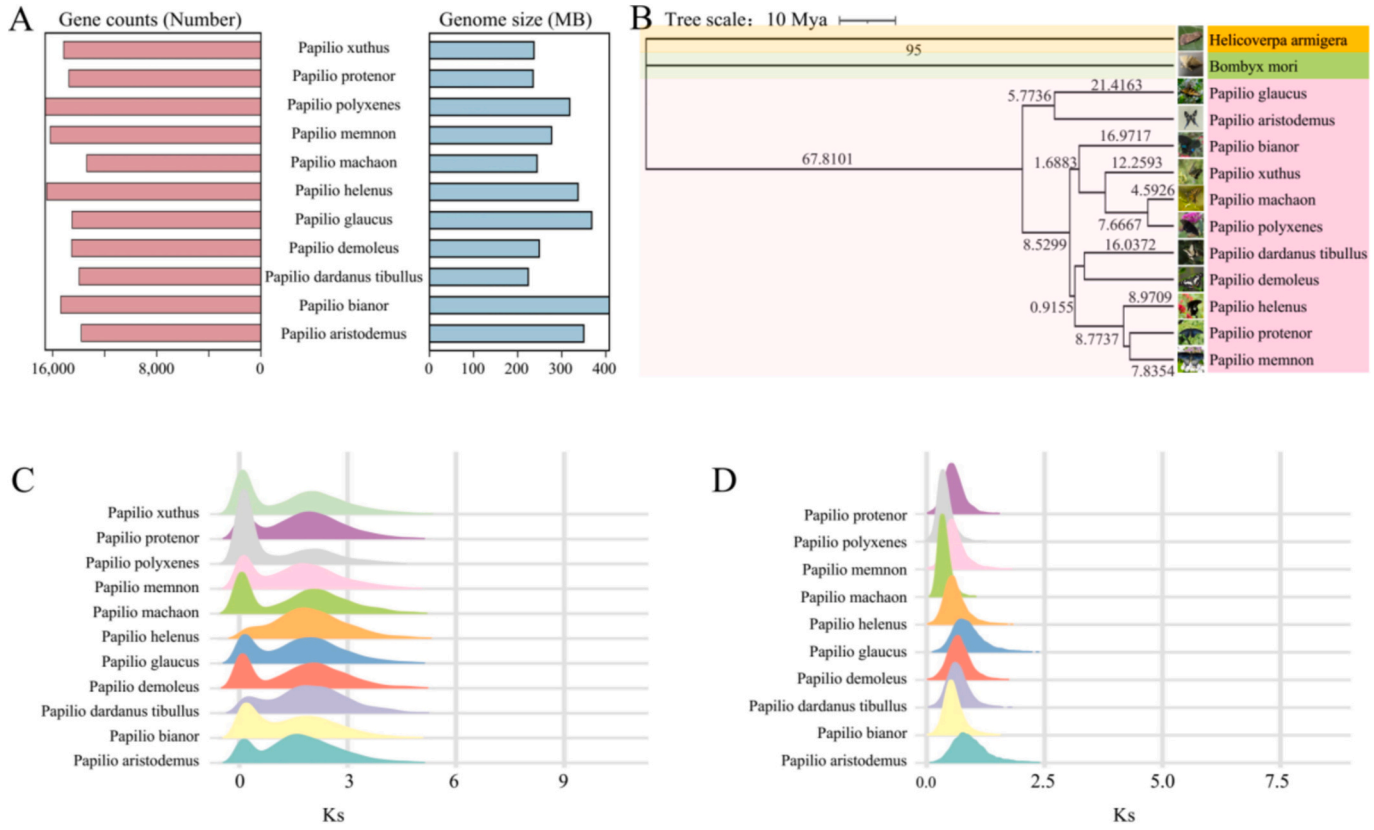

Fig. 1. Comparative genomic analysis of *Papilio*. (A) Genome size and gene counts of 11 species of *Papilio*. The bar chart on the left shows the gene count for each genome, while the bar chart on the right shows the genome size. (B) Phylogenetic tree with 2397 single-copy orthologs from 13 species identified by Orthofinder to show divergence times. The numbers on the tree indicate divergence times. (C) Ks value distribution of paralogous genes in 11 *Papilio*. The x-axis represents Ks values, and the y-axis represents species. (D) Ks value distribution of orthologous between the *P. xuthus* and other *Papilio*. The x-axis represents Ks values, and the y-axis represents species.

Among the donor species from different species, HGT-acquired genes from *Pseudomonadota* were the most abundant (50 %). The *P. glaucus* had the most uniform source of HGT-acquired genes, only obtaining genes from *Pseudomonadota* and *Bacillota*. *Candidatus Lokiarchaeota* specifically supplied genes to the *P. helenus* (Fig. 3E). This further suggests that in addition to HGT-acquired genes inherited from ancestral genomes, independent gene transfer events at the gene level may have occurred in different *Papilio* species due to differences in their respective environmental adaptations after species differentiation.

3.3. Functional analysis of HGT-acquired genes

3.3.1. Enrichment analysis of HGT-acquired genes

To further explore the reasons behind these HGT occurrences and their potential functional implications, this study conducted an enrichment analysis of HGT-acquired genes (Fig. 4A-C). With the background of all genes in 11 *Papilio* genomes, the HGT-acquired genes were found to be enriched in cysteine and methionine metabolism pathways. Furthermore, KEGG enrichment analysis conducted separately for HGT-acquired genes from different species revealed that all 11 species are enriched in the same pathway, named as the sulfur metabolism pathway. Additionally, the donor genes of the HGT-acquired genes enriched in this pathway are all from *Methylobacterium*, with the protein accession number WP_012320945.1, which encodes a cysteine synthase family protein. It has been reported that the *Tetranychus evansi* has acquired a cysteine synthase from *Methylobacterium*, which facilitates the detoxification of cyanides produced by plants [48].

Furthermore, GO enrichment analysis were conducted on HGT-acquired genes across 11 *Papilio*(Supplementary Table S7 and Supplementary Fig. S3). Among the five species with significant GO enrichment results, most exhibited enrichment in amide metabolic processes and cytoplasmic translation, likely related to the metabolism of basic substances such as proteins, hormones, amino acids, and carbohydrates. In the *P. dardanus tibullus*, HGT-acquired genes are primarily involved in reproductive biological processes (maintenance of female germ-line stem cell populations). GO analysis results of the HGT-acquired genes in *P. helenus* reveals some HGT-acquired genes participate in responses to viruses (cellular response to virus), symbiotic interactions (regulation of biological processes involved in symbiotic interaction), and oxidative stress responses (regulation of response to oxidative stress), among others. Additionally, the *P. helenus* showed enrichment in biological processes related to sulfur compound metabolism (sulfur amino acid biosynthetic process, sulfur compound metabolic process).

3.3.2. The presence and absence of HGT-acquired genes in the Papilio

The research findings mentioned above indicate that, following speciation, independent HGT events may have occurred in the genomes of various species within *Papilio*. Accordingly, the result of Orthofinder was used to display the homologous relationships of HGT-acquired genes in 11 *Papilio* species and reveal the presence and absence of these genes across 11 *Papilio* species(Fig. 4D). Some HGT-acquired genes were present in all 11 *Papilio* species, which might have been acquired before the diversification of these 11 *Papilio* species, or could be due to the existence of the same donor species in the environment after their diversification. These genes may be beneficial for their survival and thus were retained. However, certain HGT-acquired genes were specific to particular species, such as OG0000032 (Thioredoxin), OG0000033

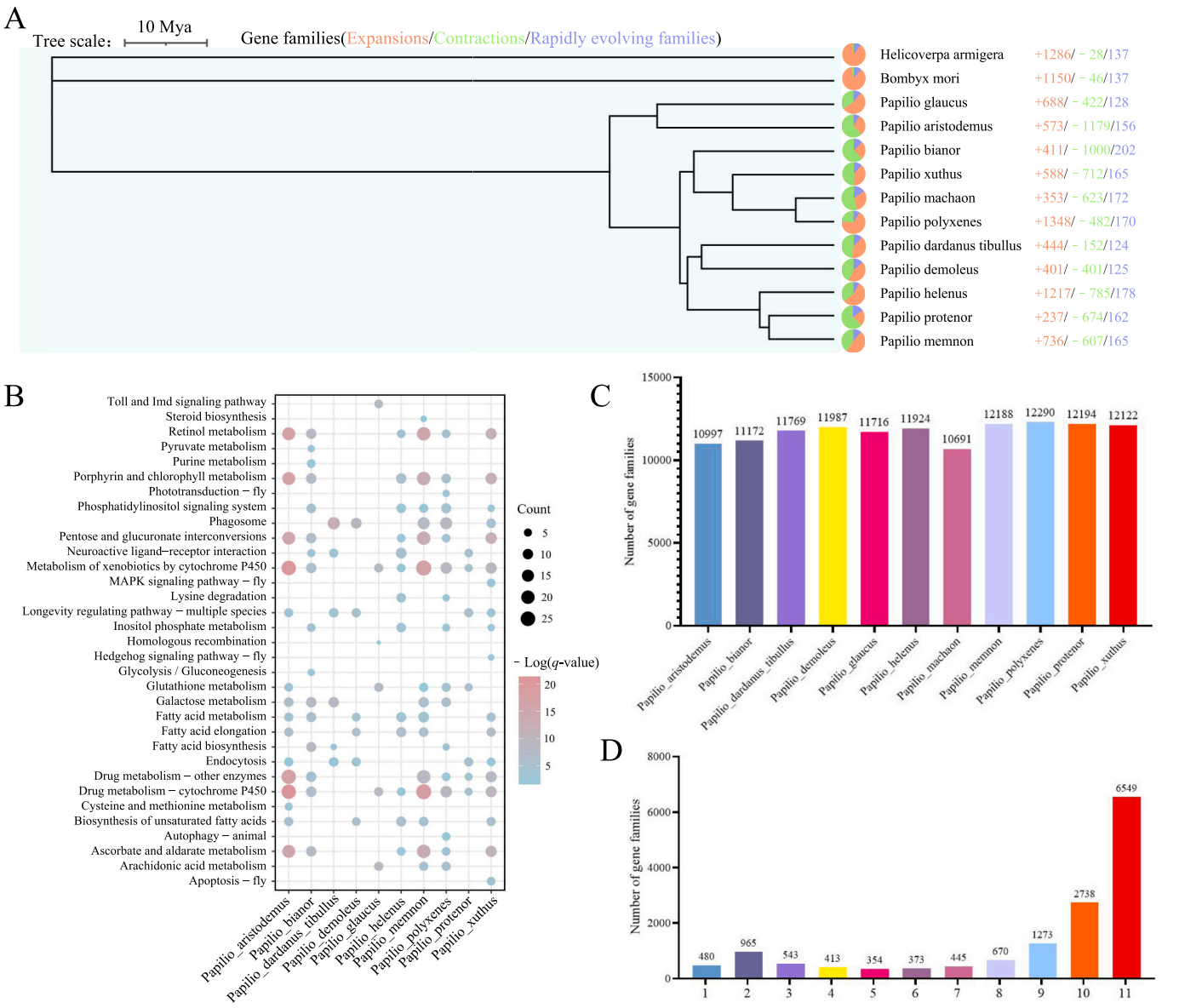

Fig. 2. Gene family expansion and contraction. (A). Estimation of gene family expansion and contraction on each evolutionary branch. The pie chart on the right shows the proportion of expanded, contracted, and significantly changed gene families. The numbers in the right of the branches represent the expansion and contraction of gene families. Red indicates the number of expansions, turquoise represents the number of contractions, and blue represents the number of significantly expanded and contracted gene families (with *p*-value ≤0.01). (B). KEGG enrichment analysis of significantly expanded gene families. The x-axis represents the species, and the y-axis represents KEGG pathways. The size and color of each dot represent the number of genes enriched in the pathway and -log(q-value), respectively. (C). The number of gene families identified in 11 *Papilio* genomes. The numbers on top of each bar indicate the number of gene families in each species. (D). The number of gene families shared by 1 to 11 species. The number of each bar represents the number of gene families shared by different numbers of species. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Malformin synthetase mlfA), OG0000034 (Electron transfer flavoprotein beta subunit lysine methyltransferase), and OG0000035 (Cystathionine gamma-lyase) in the *P. helenus*, potentially related to unique traits of the *P. helenus*. These findings provide crucial information for integrating ecological research into the evolutionary patterns of *Papilio*.

3.4. Evolution of the ABC gene family in Papilio

3.4.1. Identification of ABC gene family in Papilio

The process of genome evolution is often accompanied by the evolution of gene families with critical biological functions. ABC transporters, for example, are a major transmembrane protein family involved not only in the development of resistance to xenobiotics but also in the transport of pigment precursors and related metabolites essential for pigmentation, which plays a role in sexual selection, predator evasion, and thermoregulation in Papilio [49]. Moreover, this gene family enables insects to resist exogenous toxins, such as plant secondary metabolites and pesticides [50,51]. To investigate the evolutionary history of ABC transporter genes in Papilio, we used a standardized pipeline [42] to identify 555 ABC transporter family members in 11 *Papilio* species. *P. polyxenes* has the highest number of ABC transporters among the 11 *Papilio* species, with a total of 58, while *P. aristodemus* has the fewest, with only 32 ABC transporters. The phylogenetic tree of ABC transporters across the 11 *Papilio* species divides the ABC gene family into eight subfamilies (Fig. 5).

3.4.2. The subfamily of ABC transporters in the Papilio

The conservation and functional diversification of each ABC

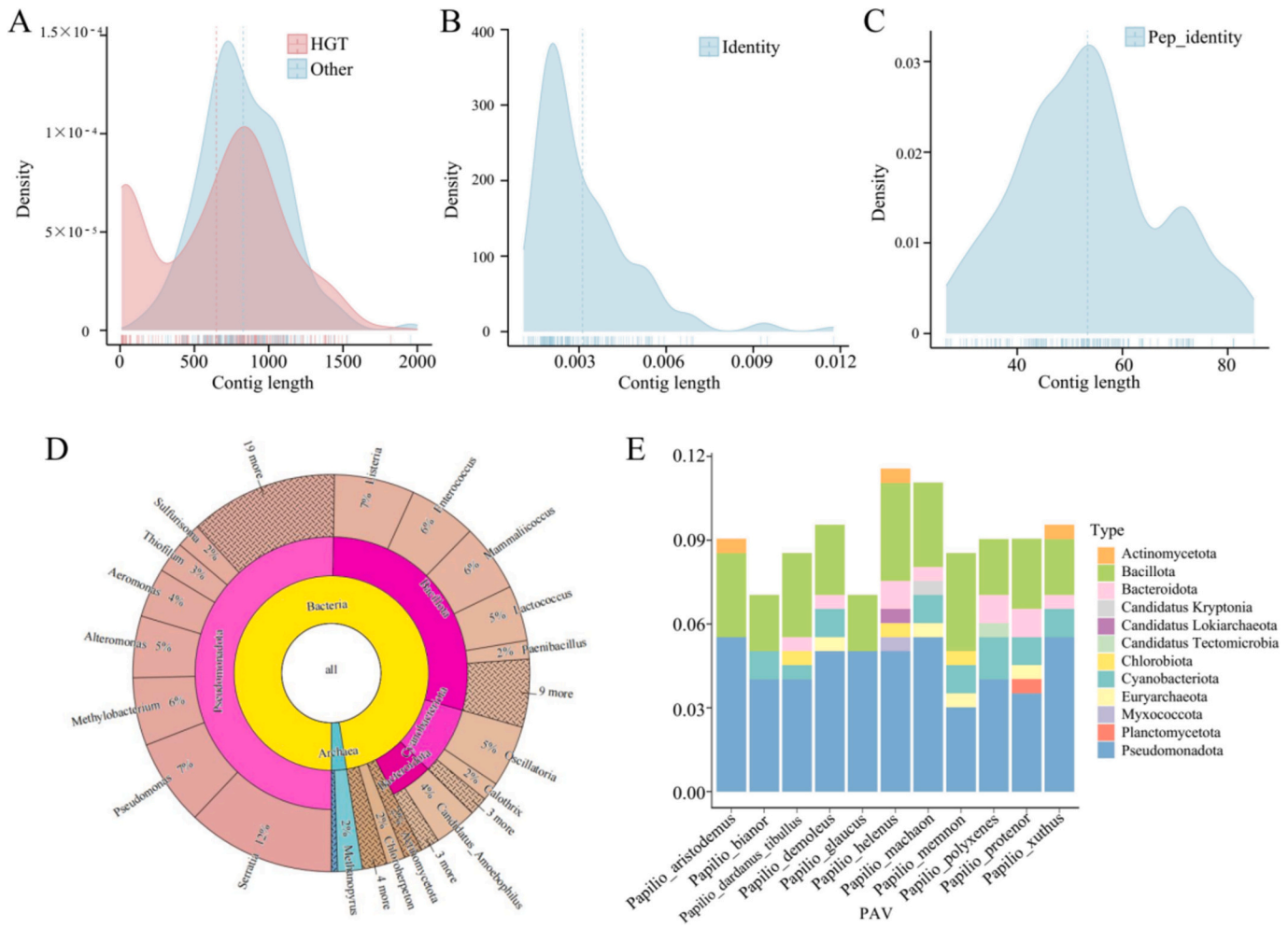

Fig. 3. HGT-acquired genes in the Genomes of *Papilio*. (A) Distributions of sequence lengths of genomic contigs with and without HGT-acquired genes. (B) Distribution of proportions of HGT-acquired genes in each of the contigs that harbors the 199 inferred HGT-acquired genes. (C) Distribution of the protein sequence similarity between the sequence in the insect recipient genome and its closest hit in a non-metazoan donor genome for all 199 HGT-acquired genes. (D) The distribution of 76 putative donor species. (E) The ratio of HGT-acquired genes from the different HGT donor resources.

transporter subfamily exhibit a rich diversity. It is evident from the distribution of the number of genes across subfamilies that ABC-D, ABC-E, and ABC-H subfamilies contain fewer genes (Fig. 6A). The ABC-D subfamily is reported to have a highly conserved function [51], and the gene duplication frequency in this subfamily is low, with only 14 ABC-D members identified across 11 species of *Papilio*. Many proteins in the ABC-G subfamily have been proven to be multidrug resistance proteins [52]; however, the number of ABC-G varies significantly across species and is the highest among all subfamilies (161 members), possibly related to the adaptation of these *Papilio* species to detoxify various toxic substances during their evolutionary process. No genes of the ABC-H subfamily were detected in the *P. helenus*, but it has a higher number of ABC-E subfamily members compared to others. ABC-B transporters are often referred to as Multidrug Resistance Proteins (MDRs) or Permeability Glycoproteins (P-glycoproteins; P-gps) due to their ability to secrete exogenous compounds. This subfamily is divided into smaller subfamilies, ABC-BF and ABC-BH. Although compared with other *Papilio* species, the *P. aristodemus* has the smallest number of ABC transporter gene family members, it has a larger proportion of ABC-BF (Fig. 6B).

3.4.3. The evolution of ABC transporters in the Papilio

The result of Orthofinder was used to display the orthologous relationships of ABC transporter genes among the genomes of 11 *Papilio* species. Based on these orthologous relationships, the study analyzed the distribution of ka/ks values among orthologous groups (Fig. 6C). The peak ka/ks values for all ABC transporters are far below 1, indicating purifying selection on ABC transporters across 11 *Papilio* species. Further analysis revealed a rich presence-absence variation of different ABC transporter gene family members across the 11 species (Fig. 6D). This phenomenon also led to some genes being present only in certain species, and hence, some orthologous groups lacked sufficient genes for analysis during the ka/ks value analysis process. Most members of the ABC transporter gene family in the *P. machaon* have no orthologous ABC transporters with the other 10 species, including with the *P. polyxenes*, which is closely related evolutionarily. This reveals that the evolution of gene families does not always coincide with the evolutionary relationships of species divergence times, suggesting that the *P. machaon* might have undergone unique evolutionary events in the ABC transporter gene family after species differentiation.

4. Discussion

The *Papilio* genus, with its rich phenotypic and ecological diversity, provides invaluable material for evolutionary studies. However, apart from a few genomic and epigenetic studies [7,53], there has been a lack of systematic research on *Papilio* species using high-quality genomes. The limited quality of existing genomes has hindered more in-depth comparative analyses. Thanks to advances in genome sequencing and

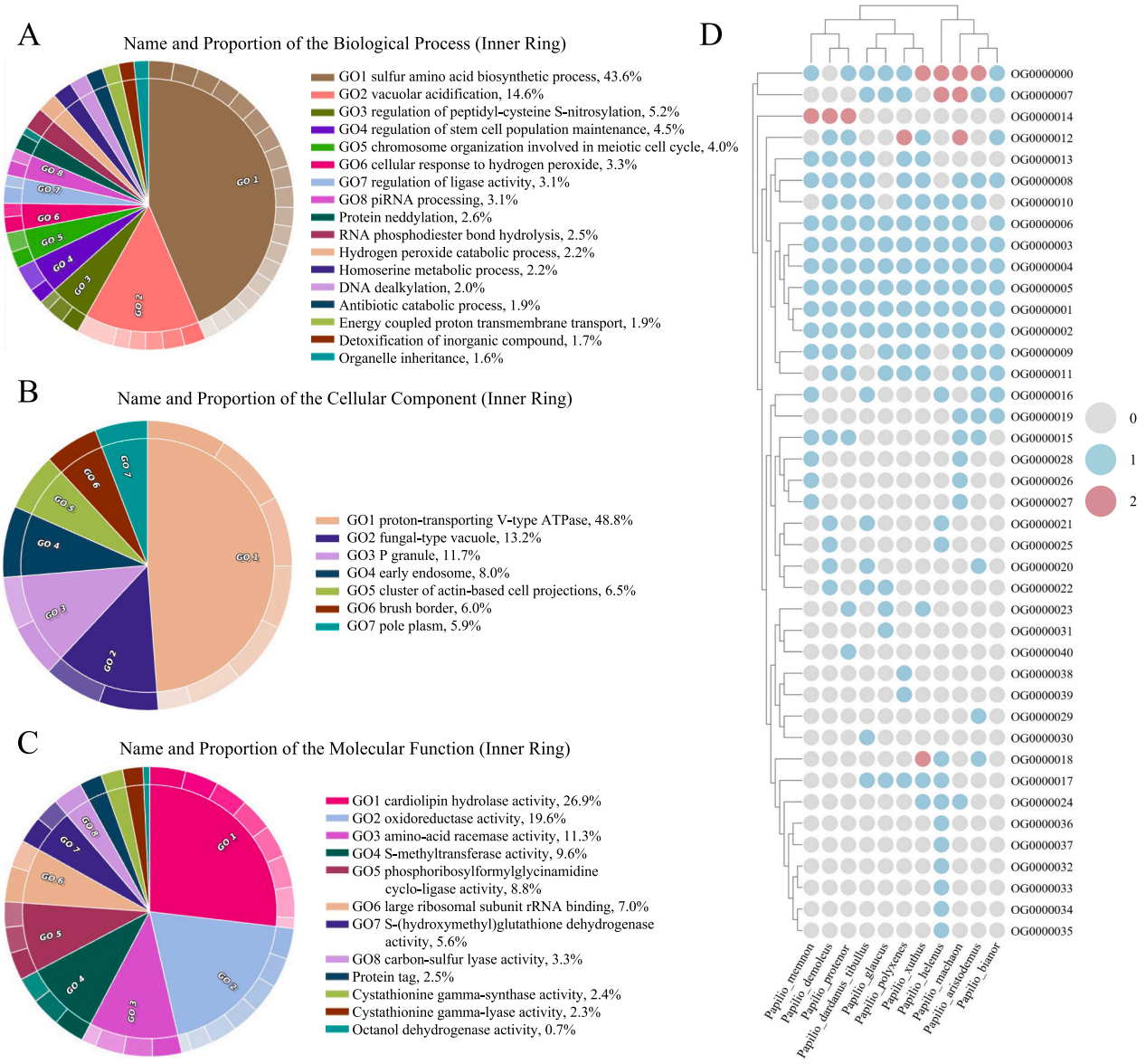

Fig. 4. GO enrichment of HGT-acquired genes in terms of (A) biological processes, (B) molecular functions, and (C) cellular components (D) presence and absence of HGT-acquired genes in 11 species.

assembly technologies, we have obtained a high-quality genome for *P. bianor*, assembled with high-throughput chromosome conformation capture techniques. We performed scaffolding and patching on 9 lower-quality genomes with *P. bianor* as reference [14], leading to improvements in N50, N90, and BUSCO scores for these genomes (Supplementary Table S3). Furthermore, we re-annotated these upgraded genomes, laying a foundation for subsequent analyses. These efforts have enabled us to conduct a systematic comparative genomic analysis across 11 *Papilio* species.

The larvae of *Papilio* feed on plants, establishing a long-term co-evolutionary relationship between plant defense and larval attack [54]. Plants have evolved a vast array of "metabolic weapons" [55] through specific metabolic pathways, posing great challenges to the growth and development of *Papilio* larvae and driving their evolution through natural selection. For instance, the *P. polyxenes* is often used by plant protection experts [56,57] for detoxification experiments against toxic plant metabolites, with members of the P450 gene family found to be involved in the detoxification of plant secondary metabolites. Interestingly, this study identified several detoxification-related genes, such as UDP-glucosyltransferase 2, UDP-glycosyltransferase UGT5, and UDP-glucuronosyltransferase 2B17, significantly expanded in the *P. polyxenes*. These genes are enriched in pathways like metabolism of xenobiotics by cytochrome P450 and drug metabolism - cytochrome P450, revealing diversification in evolutionary directions among different species within *Papilio*. These results provide further genetic resources for studying the evolution of detoxification-based environmental adaptability among *Papilio* species, using the *P. polyxenes* as a reference.

Thanks to the improved quality of the genomes, this study identified HGT-acquired genes in 11 species of the *Papilio* genus. We assessed the reliability of the HGT-acquired genes based on several factors: the length of contigs containing these genes, the proportion of HGT-acquired genes on each contig, and the protein similarity between the HGT-acquired genes and their closest homologs in non-metazoan donors. Some of these metrics are consistent with the study of Li et al. [31], further confirming the reliability of the 199 HGT-acquired genes identified in this study. This provides a comprehensive understanding of gene-level HGT in the genomes of 11 species within *Papilio*. Various roles of

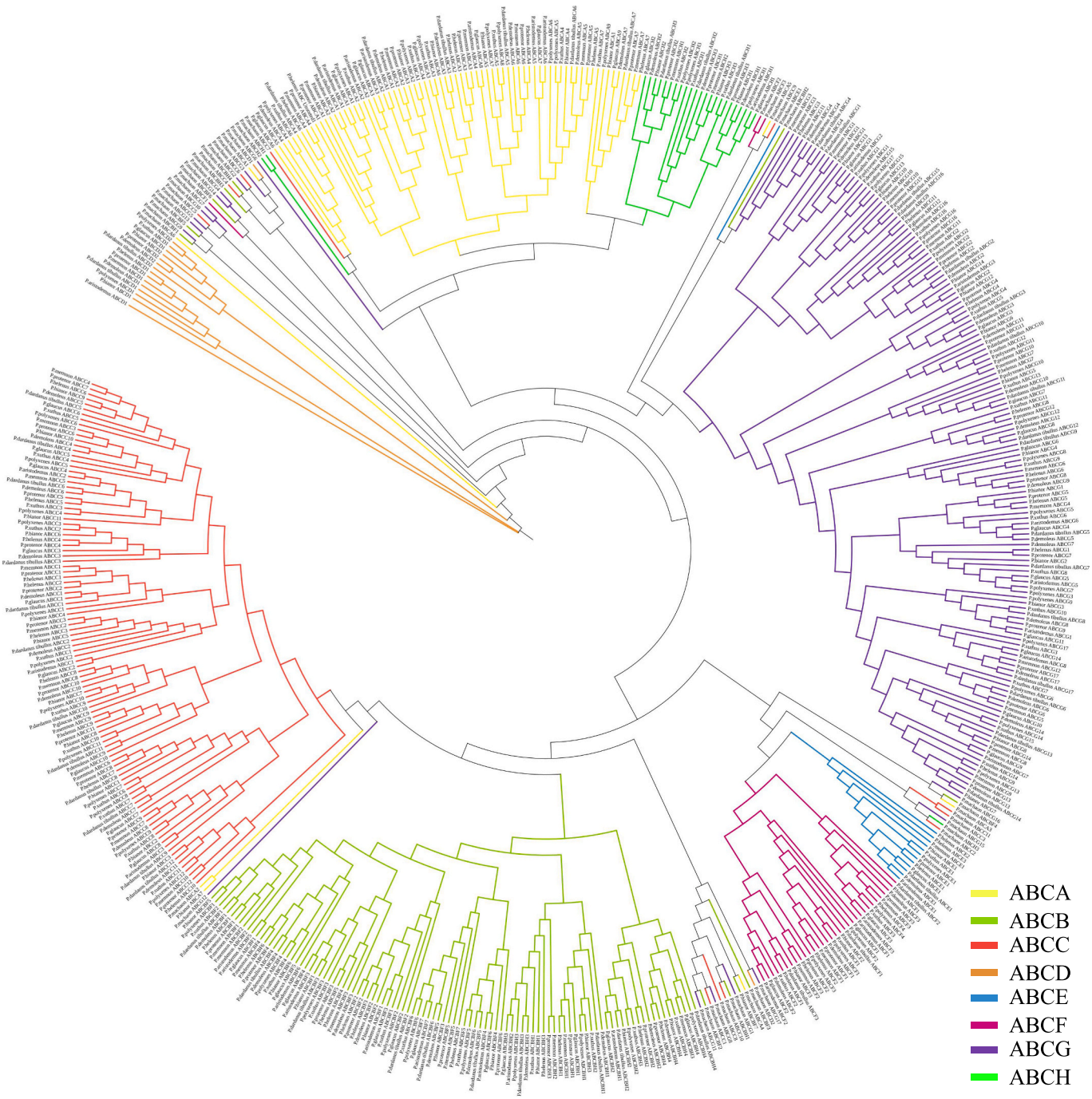

Fig. 5. Phylogenetic evolutionary tree of ABC transporters among 11 *Papilio* species.

HGT-acquired genes in the reproduction and growth of insects have been systematically revealed [31]. *Pseudomonadota* and *Bacillota* are important bacterial phyla in the gut microbiomes of various insects, with their composition at lower taxonomic levels varying with the host's physiological and ecological environment [58]. HGT-acquired genes identified in the genomes of *Papilio* species mainly originate from *Pseudomonadota* and *Bacillota*, suggesting that HGT-acquired genes in *Papilio* may primarily be related to food digestion. Additionally, interesting HGT events were discovered, such as HGT-acquired genes related to sulfide metabolism in the *P. helenus*, potentially associated with the larvae's ability to secrete a foul-smelling liquid to deter predators and parasites upon irritation. These results revealed a rich phenomenon of gene transfer between Papilio species and microorganisms, providing new insights into the formation of environmental adaptability and corresponding evolutionary processes in different species within *Papilio*.

As one of the important detoxification gene families in insects, the analysis of the evolutionary patterns of ABC transporters in *Papilio* reveals significant differences between species. Given that these species of *Papilio* theoretically share a common ancestor, it is challenging to infer the evolutionary process of the gene family from monophyletic groups within the ABC transporter gene family phylogenetic tree. However, analysis of the phylogenetic tree still reveals unique phenomena, such as the ABC-E gene family having only one member in the genomes of 10 species, except for the *P. helenus*. In the *P. helenus* genome, a gene duplication event occurred, resulting in two ABC-E subfamily genes. The functional analysis of each subfamily shows that different environmental selections have a significant impact on functional genes. Although the absence of ABC-H transporters in the *P. helenus* genome

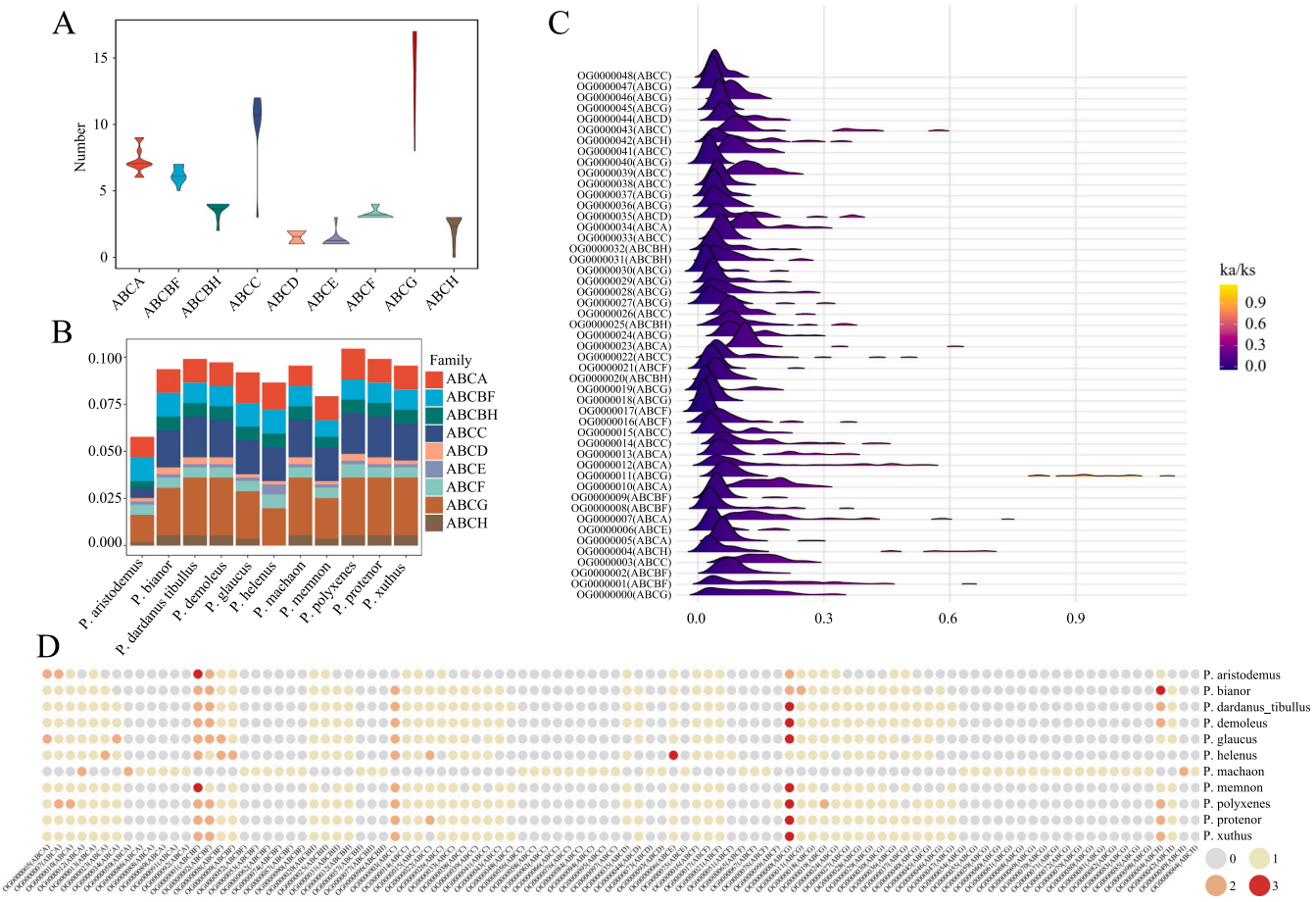

Fig. 6. Analysis of the ABC transporter gene families in 11 species. (A) The distribution of the number of 9 ABC transporter subfamilies. (B) The proportion of the number of genes in each ABC transporter subfamily within the genomes of the 11 species. The total of the ABC transporter gene families across the 11 species is set to 1. (C) The Ka/Ks distribution of the orthologous ABC transporters. (D) Presence and absence of ABC transporters in the 11 species. The numbers 0, 1, 2, and 3 indicate the number of genes contained in each homologous group across the species.

could be due to reasons related to genome assembly and structural annotation, this is a remarkable event given the known significant functions of ABC-H. Structurally, ABC-H proteins are half-transporters and exhibit an inverse domain structure similar to the ABC-G family. Functional characterization of these proteins has been conducted in fruit flies, where the ABC-H transporter Snustorr (Snu) is localized in the cuticle and shown to transport cuticular hydrocarbons [59]. RNAi knockdown of *Snu* homologs in other arthropods leading to death from desiccation suggests a similar function of ABC-H transporter [60,61]. The survival mechanism in the absence of ABC-H presents an interesting scientific question. Although this study lacks the *P. helenus* samples to verify the results, our analysis broadly reveals the variation patterns of ABC transporters across the genomes of *Papilio* species, providing a reference for studying the evolution of *Papilio* with respect to ABC transporters. For example, the poor reproductive capacity of the *P. aristodemus*, a significant factor in its endangerment, correlates with its minimal numbers of the resistant protein subfamilies ABC-G and ABC-C, potentially a crucial reason for the high endangerment risk of the *P. aristodemus*.

In summary, this study utilized comparative genomics, HGT-acquired genes identification, and an in-depth analysis of ABC transporters as a case study to shed light on gene evolution within the Papilio genus. These approaches provided new insights into the evolutionary processes and genomic adaptations in *Papilio* species, offering a perspective on the genetic mechanisms underlying their ecological and phenotypic diversity.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2024.110956.

Funding

This work was supported by Natural Science Foundation of the Higher Education Institutions of Anhui Province (2022AH040152); Outstanding Youth Science Foundation of the Higher Education Institutions of Anhui Province (2022AH030110); Research Initiation Fund Project of Chuzhou University (2022qd015).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

CRediT authorship contribution statement

Jiajia Wang: Writing – original draft, Methodology, Investigation. **Yunfei Wu:** Investigation. **Linxin Zhu:** Conceptualization. **Kaixin Guo:** Investigation. **Shichen Gao:** Visualization. **Yan Dong:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Files can be download at Figshare (10.6084/m9.figshare.27215379、10.6084/m9.figshare.25272262、10.6084/m9.figshare.25333684.v1).

References

[1] A. Sucháčková Bartoňová, D. Linke, I. Klečková, G. de Ribeiro, P. Matos-Maraví, Incorporating genomics into insect conservation: butterflies as a model group, Insect. Conserv. Divers. 16 (2023) 427–440.
[2] A.R.P. Martins, N.B. Warren, W.O. McMillan, R.D.H. Barrett, Spatiotemporal dynamics in butterfly hybrid zones, Insect Sci. 31 (2024) 328–353.
[3] Fleeming Jenkin, The origin of species, North Br. Rev. (1867) 277–318.
[4] S. Wang, D. Teng, X. Li, P. Yang, W. Da, Y. Zhang, et al., The evolution and diversification of oakleaf butterflies, Cell 185 (2022) 3138–3152.e20.
[5] G. Huang, L. Song, X. Du, X. Huang, F. Wei, Evolutionary genomics of camouflage innovation in the orchid mantis, Nat. Commun. 14 (2023) 4821.
[6] F.L. Condamine, R. Allio, E.L. Reboud, J.R. Dupuis, E.F.A. Toussaint, N. Mazet, et al., A comprehensive phylogeny and revised taxonomy illuminate the origin and diversification of the global radiation of Papilio (Lepidoptera: Papilionidae), Mol. Phylogenet. Evol. 183 (2023) 107758.
[7] S. Lu, J. Yang, X. Dai, F. Xie, J. He, Z. Dong, et al., Chromosomal-level reference genome of Chinese peacock butterfly (Papilio bianor) based on third-generation DNA sequencing and hi-C analysis, Gigascience 8 (2019) giz128.
[8] M.J.T.N. Timmermans, S.W. Baxter, R. Clark, D.G. Heckel, H. Vogel, S. Collins, et al., Comparative genomics of the mimicry switch in Papilio dardanus, Proc. R. Soc. B Biol. Sci. 281 (2014) 20140465.
[9] Z. Pan, Y. Ding, S. Zhang, L. Li, F. Ma, Chromosome-level genome assembly of papilio elwesi Leech, 1889 (Lepidoptera: Papilionidae), Insects (2023) 14.
[10] J. Xia, Z. Guo, Z. Yang, H. Han, S. Wang, H. Xu, et al., Whitefly hijacks a plant detoxification gene that neutralizes plant toxins, Cell 184 (2021) 1693–1705.e17.
[11] Y. Li, Z. Liu, C. Liu, Z. Shi, L. Pang, C. Chen, et al., HGT is widespread in insects and contributes to male courtship in lepidopterans, Cell 185 (2022) 2975–2987.e10.
[12] F. Lu, Z. Wei, Y. Luo, H. Guo, G. Zhang, Q. Xia, et al., SilkDB 3.0: visualizing and exploring multiple levels of data for silkworm, Nucleic Acids Res. 48 (2020) D749–D755.
[13] S.L. Pearce, D.F. Clarke, P.D. East, S. Elfekih, K.H.J. Gordon, L.S. Jermiin, et al., Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive Helicoverpa pest species, BMC Biol. 15 (2017) 63.
[14] M. Alonge, L. Lebeigle, M. Kirsche, K. Jenike, S. Ou, S. Aganezov, et al., Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing, Genome Biol. 23 (2022) 258.
[15] T. Zhou, L. Lu, C. Li, Optimization of the “in-silico” mate-pair method improves contiguity and accuracy of genome assembly, Ecol. Evol. 13 (2023) e9745.
[16] M. Manni, M.R. Berkeley, M. Seppey, F.A. Simão, E.M. Zdobnov, BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes, Mol. Biol. Evol. 38 (2021) 4647–4654.
[17] S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (2018) i884–i890.
[18] M. Grabherr, B. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, Trinity: Reconstructing a Full-length Transcriptome Without a Genome from RNA-Seq data 29, 2011, pp. 644–652.
[19] J.M. Flynn, R. Hubley, C. Goubert, J. Rosen, A.G. Clark, C. Feschotte, et al., RepeatModeler2 for automated genomic discovery of transposable element families, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 9451–9457.
[20] G. Benson, Tandem repeats finder: a program to analyze DNA sequences, Nucleic Acids Res. 27 (1999) 573–580.
[21] M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: ab initio prediction of alternative transcripts, Nucleic Acids Res. (2006) W435–9 (34 Web Server issue).
[22] C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects, BMC Bioinformat. 12 (2011) 1–14.
[23] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, et al., BLAST+: architecture and applications, BMC Bioinformat. (2009).
[24] Z. Zhang, J. Xiao, J. Wu, H. Zhang, G. Liu, X. Wang, et al., ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments, Biochem. Biophys. Res. Commun. 419 (2012) 779–781.
[25] D. Wang, Y. Zhang, Z. Zhang, J. Zhu, J. Yu, KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies, Genom. Proteom. Bioinform. 8 (2010) 77–80.
[26] S. Aldahmani, T. Zoubeidi, Graphical group ridge, J. Stat. Comput. Simul. 90 (2020) 3422–3432.
[27] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, Genome Biol. 20 (2019) 238.
[28] M.J. Sanderson, r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock, Bioinformatics 19 (2003) 301–302.
[29] S. Kumar, M. Suleski, J.M. Craig, A.E. Kasprowicz, M. Sanderford, M. Li, et al., TimeTree 5: an expanded resource for species divergence times, Mol. Biol. Evol. 39 (2022) msac174.
[30] T. De Bie, N. Cristianini, J.P. Demuth, M.W. Hahn, CAFE: a computational tool for the study of gene family evolution, Bioinformatics 22 (2006) 1269–1271.
[31] Y. Li, Z. Liu, C. Liu, Z. Shi, L. Pang, C. Chen, et al., HGT is widespread in insects and contributes to male courtship in lepidopterans, Cell 185 (2022) 2975–2987.e10.
[32] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (2015) 59–60.
[33] X.X. Shen, D.A. Opulente, J. Kominek, X. Zhou, J.L. Steenwyk, K.V. Buh, et al., Tempo and mode of genome evolution in the budding yeast subphylum, Cell 175 (2018) 1533–1545.e20.
[34] S. Kuraku, C.M. Zmasek, O. Nishimura, K. Katoh, aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity, Nucleic Acids Res. 41 (2013) W22–W28.
[35] S. Capella-Gutiérrez, J.M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, Bioinformatics 25 (2009) 1972–1973.
[36] L.T. Nguyen, H.A. Schmidt, A. Von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2015) 268–274.
[37] G. Yu, D.K. Smith, H. Zhu, Y. Guan, T.T.-Y. Lam, Ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data, Methods Ecol. Evol. 8 (2017) 28–36.
[38] C. Xie, X. Mao, J. Huang, Y. Ding, J. Wu, S. Dong, et al., KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases, Nucleic Acids Res. 39 suppl_2 (2011) W316–W322.
[39] S.R. Eddy, Accelerated profile HMM searches, PLoS Comput. Biol. 7 (2011) e1002195.
[40] T. Paysan-Lafosse, M. Blum, S. Chuguransky, T. Grego, B.L. Pinto, G.A. Salazar, et al., InterPro in 2022, Nucleic Acids Res. 51 (2023) D418–D427.
[41] C.P. Cantalapiedra, A. Hernández-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas, eggNOG-mapper v2: functional annotation, Orthology assignments, and domain prediction at the metagenomic scale, Mol. Biol. Evol. 38 (2021) 5825–5829.
[42] S. Denecke, I. Rankić, O. Driva, M. Kalsi, N.B.H. Luong, B. Buer, et al., Comparative and functional genomics of the ABC transporter superfamily across arthropods, BMC Genom. 22 (2021) 553.
[43] F. Abascal, R. Zardoya, D. Posada, ProtTest: selection of best-fit models of protein evolution, Bioinformatics 21 (2005) 2104–2105.
[44] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (2014) 1312–1313.
[45] I. Letunic, P. Bork, Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation, Nucleic Acids Res. 49 (2021) W293–W296.
[46] D.-P. Wang, H.-L. Wan, S. Zhang, J. Yu, γ-MYN: a new algorithm for estimating Ka and Ks with consideration of variable substitution rates, Biol. Direct 4 (2009) 20.
[47] M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, Mol. Biol. Evol. 3 (1986) 418–426.
[48] N. Wybouw, W. Dermauw, L. Tirry, C. Stevens, M. Grbić, R. Feyereisen, et al., A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning, Elife 3 (2014) e02365.
[49] I. Medina, R. Vega-Trejo, T. Wallenius, M.R.E. Symonds, D. Stuart-Fox, From cryptic to colorful: evolutionary decoupling of larval and adult color in butterflies, Evol. Lett. 4 (2020) 34–43.
[50] C. Wu, S. Chakrabarty, M. Jin, K. Liu, Y. Xiao, Insect ATP-binding cassette (ABC) transporters: roles in xenobiotic detoxification and Bt insecticidal activity, Int. J. Mol. Sci. 20 (2019) 2829.
[51] W. Dermauw, T. Van Leeuwen, The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance, Insect Biochem. Mol. Biol. 45 (2014) 89–110.
[52] A. Kovalchuk, A.J.M. Driessen, Phylogenetic analysis of fungal ABC transporters, BMC Genom. 11 (2010) 177.
[53] W. Wan, P. Hu, Z. Chang, Y. Ren, Z. Dong, J. Yang, et al., Genome-wide survey of open chromatin regions in two swallowtail butterflies Papilio machaon and P. bianor, Arch. Insect Biochem. Physiol. 111 (2022) e21952.
[54] G. Sollai, I. Tomassini Barbarossa, C. Masala, P. Solari, R. Crnjar, Gustatory sensitivity and food acceptance in two phylogenetically closely related Papilionid species: Papilio hospiton and Papilio machaon, PLoS One 9 (2014) e100675.
[55] M. Erb, C.A.M. Robert, Sequestration of plant secondary metabolites by insect herbivores: molecular mechanisms and ecological consequences, Curr. Opin. Insect. Sci. 14 (2016) 8–11.
[56] Z. Wen, M.R. Berenbaum, M.A. Schuler, Inhibition of CYP6B1-mediated detoxification of Xanthotoxin by plant Allelochemicals in the black swallowtail (Papilio polyxenes), J. Chem. Ecol. 32 (2006) 507–522.
[57] J.G. Zumwalt, J.J. Neal, Cytochromes P450 from Papilio polyxenes: adaptations to host plant allelochemicals, Comp. Biochem. Physiol. Part C Pharmacol. Toxicol. Endocrinol. 106 (1993) 111–118.
[58] S. Suenami, A. Koto, R. Miyazaki, Basic structures of gut bacterial communities in eusocial insects, Insects 14 (2023) 444.
[59] R. Zuber, M. Norum, Y. Wang, K. Oehl, N. Gehring, D. Accardi, et al., The ABC transporter Snu and the extracellular protein Snsl cooperate in the formation of the

lipid-based inward and outward barrier in the skin of Drosophila, Eur. J. Cell Biol. 97 (2018) 90–101.
[60] Z. Guo, S. Kang, X. Zhu, J. Xia, Q. Wu, S. Wang, et al., The novel ABC transporter ABCH1 is a potential target for RNAi-based insect pest control and resistance management, Sci. Rep. 5 (2015) 13728.
[61] Z. Yu, Y. Wang, X. Zhao, X. Liu, E. Ma, B. Moussian, et al., The ABC transporter ABCH-9C is needed for cuticle barrier construction in Locusta migratoria, Insect Biochem. Mol. Biol. 87 (2017) 90–99.