Molecular Phylogenetics and Evolution 143 (2020) 106695

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Origin and evolution of the CYP4G subfamily in insects, cytochrome P450 enzymes involved in cuticular hydrocarbon synthesis

René Feyereisen*

Department of Plant and Environmental Sciences, University of Copenhagen, Copenhagen, Denmark
Department of Plants and Crops, Ghent University, Ghent, Belgium

ARTICLE INFO

Keywords:
Cytochrome P450
Insect phylogeny
Gene birth and death
Cuticular hydrocarbons
Chemical communication

ABSTRACT

The large and diverse P450 (CYP) superfamily encodes enzymes with a wide spectrum of monooxygenase and related activities. Insect P450 enzymes of the CYP4G subfamily are known to catalyze the synthesis of cuticular hydrocarbons that serve multiple functions from desiccation resistance to chemical communication. These functions are essential for survival. In order to understand the evolution of insect CYP4G genes, 368 sequences from 24 insect orders and 167 species were mined and analyzed. The genomes of most species of Neoptera carry at least two CYP4G genes that are paralogs of the two *Drosophila* CYP4G genes. The duplication of the original CYP4G is basal to Neoptera and no CYP4G is found in Paleoptera, or beyond the class Insecta. The sequences of CYP4G and particularly their active site have been highly conserved over 400 MY, but all CYP4G sequences are characterized by a +44 residue insertion between the G and H helices, which protrudes from the globular structure of the enzyme distally from the membrane anchor. Although it is generally considered that genes with highly conserved sequence and function are evolutionarily "stable", the evidence from the CYP4G subfamily shows that since their initial duplication over 400 MYA, these genes have experienced many gene births and deaths. The CYP4G1 homolog has been lost several times, and is missing in five orders of insects. These losses are both ancient, as in all Hemiptera and Thysanoptera, and more recent as in honey bees. Serial duplications leading to CYP4G gene clusters have also been observed, as in house flies and in fireflies. The detailed evolutionary history of CYP4G genes does not support the "stability" of these essential genes, but rather a "revolving door" pattern where their essential function is maintained despite an apparently random birth and death process. The dual function of cuticular hydrocarbons, in desiccation resistance achieved mainly by the quantity of hydrocarbons produced and in chemical communication, achieved by the blend of hydrocarbons produced, may explain the apparently paradoxical evolution of CYP4G genes.

1. Introduction

The genes encoding P450 enzymes are part of one of the largest gene families in plants and animals, the CYP genes (Nelson, 2018). Studies on CYP genes in insects have contributed to our understanding of the evolution of such gene families (Feyereisen, 2011; Sezutsu et al., 2013, Good et al., 2014). One particular aspect of the CYP gene family in animals is their apparent dichotomy in function. On one hand, the P450 enzymes involved in the metabolism of xenobiotics (such as pesticides and drugs) and of plant and fungal compounds. On the other hand the P450 enzymes involved in the synthesis and degradation of signalling molecules, hormones and lipids. This apparent dichotomy is not related to phylogeny, as these two functional types are found throughout the P450 gene family and are not tightly clustered in two phylogenetically distinct gene lineages. In humans about 56% of the CYP genes play a role in physiological processes while 26% are involved in detoxification, the remainder being of "orphan" function (Guengerich and Cheng, 2011). Much less is known of the function of insect CYP genes (Feyereisen, 2015) and a physiological function is often found or suspected for single copy genes that are highly conserved within the class Insecta (Feyereisen, 2012).

The CYP4G subfamily is such a case. A preliminary survey of CYP4G sequences in insects (Qiu et al., 2012) revealed that there is an average of two CYP4G genes in most insect orders, a minimum of one in the honey bee and the pea aphid, and a handful of genes in Lepidoptera. The CYP4G subfamily is also remarkable because genes of this subfamily are very highly expressed. In the first global P450 transcriptome analysis of *Drosophila melanogaster* (Daborn et al., 2002), the CYP4G1

* Address: Department of Plant and Environmental Sciences, University of Copenhagen, Copenhagen, Denmark.
E-mail address: rene.feyereisen@gmail.com.

https://doi.org/10.1016/j.ympev.2019.106695
Received 5 October 2019; Received in revised form 23 November 2019; Accepted 26 November 2019
Available online 02 December 2019
1055-7903/ © 2019 Elsevier Inc. All rights reserved.

gene stood out as, by far, the most highly expressed P450 genes of all. Similarly, CYP4G19 is among the twenty most abundant assembled unigenes from the transcriptome of the cockroach *Blattella germanica* (Zhou et al., 2014). *Drosophila* CYP4G1 is highly expressed in the oenocytes, with null mutants or RNAi suppression causing mortality at the late pupal - adult emergence stage (Gutierrez et al., 2007; Chung et al., 2009).

The function of the CYP4G genes was first elucidated by Qiu et al. (2012) who showed that survivors of CYP4G1 RNAi are deficient in cuticular hydrocarbons (CHC), highly sensitive to desiccation stress, and impaired in their pheromone mediated courtship behavior. Furthermore, the recombinant *Musca domestica* CYP4G2-P450 reductase fusion protein was shown to catalyze the last step in CHC biosynthesis, the oxidative decarbonylation of long chain fatty aldehydes (Qiu et al., 2012). The role of CYP4G enzymes in the synthesis of alkanes and alkenes which serve as waterproofing agents on the insect epicuticle and in many pheromonal functions as well (Howard and Blomquist, 2005; Ferveur, 2005) has been confirmed in subsequent studies. Balabanidou et al. (2016) showed that both *Anopheles gambiae* CYP4G genes are highly expressed in the oenocytes and that CYP4G16 is a functional oxidative decarbonylase. Kefi et al. (2019) further demonstrated that both mosquito CYP4G genes, alone or in combination, could rescue the lethal phenotype CYP4G1 knockdown in transgenic *Drosophila*. RNAi of CYP4G102 in *Locusta migratoria* (Yu et al., 2016) and of the single CYP4G gene, CYP4G51, in the pea aphid *Myzus persicae* (Chen et al., 2016) causes increased mortality and decreases CHC content and desiccation tolerance. The single CYP4G of the honey bee, CYP4G11, catalyzes the oxidative decarbonylation of the respective fatty aldehydes to tridecane (C13) and heptadecane (C17) (Calla et al., 2018). In the mountain pine beetle, *Dendroctonus ponderosae*, both CYP4G enzymes are active in the biosynthesis of CHC and of (*Z*)-3-nonene, an intermediate in the pathway to the pheromone *exo*-brevicomin (MacLean et al., 2018). In another coleopteran, *Tenebrio molitor*, RNAi of the two CYP4G genes causes a decrease in cuticular hydrocarbon content and in desiccation tolerance (Wang et al., 2019a). Similar results were obtained for the two CYP4G genes of the brown planthopper, *Nilaparvata lugens* (Wang et al., 2019b). Thus, direct and indirect evidence from phylogenetically distant species strongly suggest that CYP4G enzymes share a common biochemical function as oxidative decarbonylases essential in hydrocarbon biosynthesis, whether these have a structural function (as CHC) or a signalling one (as pheromones).

Constitutive overexpression of CYP4G genes in insecticide-resistant strains also provides indirect evidence for a role of CYP4G genes in resistance. Initial observations were only correlative, i.e. high CYP4G expression and resistance (Pittendrigh et al., 1997; Pridgeon et al., 2003; Müller et al., 2008; Jones et al., 2013). RNAi of CYP4G19 in *Blattella germanica* and of CYP4G14 in *Tribolium castaneum* increases toxicity of pyrethroids (Guo et al., 2010; Chen et al., 2019; Kalsi and Palli, 2017). It appears that the control of CHC production by CYP4G enzymes affects insecticide penetration, and hence contributes to resistance (Balabanidou et al., 2016, 2018, 2019; Wang et al., 2019b). In the laboratory-selected resistant strain of *D. melanogaster* called 91-R, CYP4G1 is one of several constitutively overexpressed genes, leading to an increase in CHC content. RNAi experiments suggest that this contributes to DDT resistance in this strain (Kim et al., 2018). The CYP4G enzymes are therefore not only essential for the synthesis of CHC and their multiple structural and communication roles, but they can also play a role in a poorly understood toxicokinetic aspect of insecticide resistance, i.e. resistance to penetration through the cuticle.

It was therefore felt that a phylogenomic study of the CYP4G subfamily was needed to understand the origin and evolution of these enzymes which, among P450 enzymes, have a unique catalytic mechanism. Indeed no P450 outside insects is known to cleave a carbon-carbon bond to yield a terminal, saturated n-alkane moiety, a highly unusual P450 reaction (Ortiz de Montellano, 2005; Guengerich and Yoshimoto, 2018). In particular, the observation that most insect species have two CYP4G genes was followed, and instances where this is not the case were analyzed in greater detail. A global phylogeny of the CYP4G sequences is presented. The analysis covers 368 sequences from 167 insect species (40 species of Polyneoptera, 20 species of Hemipteroids (Hemiptera, Thysanoptera, Psocodea) and 107 species of Holometabola). A commonly held assumption about highly conserved "essential" genes is that they are evolutionarily "stable", less subject to the birth and death processes seen in large gene families (Krylov et al., 2003; Waterhouse et al., 2010). Here the CYP4G genes are shown to have a highly conserved sequence yet to be "unstable" with multiple instances of both ancient and recent gene births and deaths.

2. Material and methods

Sequence databases were searched by blastp and tblastn in a recursive manner, by using known CYP4G sequences from different insect orders. The sequences were curated manually to verify and correct length and intron–exon junctions, and this was facilitated by WebScipio (www.webscipio.org; Hatje et al., 2011). Frameshifts were verified when possible by transcriptome data. The richness and completeness of the transcriptomes derived from the "1-KITE" project in particular (Misof et al., 2014; Peters et al., 2017; Johnson et al., 2018; Wipfler et al., 2019) is gratefully acknowledged. These data were remarkably concordant with genomic data. Many cases of gene presence or absence could be assigned with a high degree of confidence, either when the genomic data were incomplete, or when some phylogenetic lineages were devoid of genomic data. The (current) genomic "deserts" are often covered by transcriptome sequence assemblies (hereafter TSA), which are well suited for CYP4G mining. Indeed, as essential gene(s) in cuticular hydrocarbon biosynthesis, these genes are among the most highly expressed P450 genes in insects. The CYP4G genes are so abundantly expressed that they are occasionally found as full length transcripts from parasitoids in their lepidopteran or neuropteran hosts e.g. in *Cnaphalocrocis medinalis* (KP001141) and *Pseudomallada prasinus* (GAVV02039744). Sequences from TSA data were converted to protein sequence with orf finder (www.ncbi.nlm.nih.gov/orffinder/). Only full length sequences were retained except for five TSA sequences from Polyneoptera averaging 525 aa, and for which no alternative and equally representative full length sequence was available. It was attempted to obtain full-length sequences from at least two species in each insect order, with diverse and species rich orders sampled more thoroughly. Sequences were aligned by MAFFT (Katoh and Standley, 2013) and trees were built by RAxML (Stamatakis, 2014) on the CIPRES Science gateway V.3.3 (www.phylo.org). Consensus sequences were illustrated by WebLogo (weblogo.berkeley.edu; Crooks et al., 2004). CYP4G16 of *Anopheles gambiae* was modeled by I-TASSER (Yang et al., 2015) and the resulting pdb file visualized by iCn3D (Wang et al., 2019c) along with its template 5T6Q, the rabbit CYP4B1 with bound octane (Hsu et al., 2017). Species phylogenies followed Misof et al. (2014) at the order level, and followed more detailed phylogenies as noted in the results. Ranking within lineages used the taxonomy browser at NCBI below the order level unless otherwise noted.

3. Results and discussion

3.1. Two main classes of CYP4G genes in insects

The CYP4G protein sequences are characterized by a substantial insertion between the G and H helices (Qiu et al. 2012). This insertion of ± 44 residues facilitates the identification of CYP4G sequences. The CYP4G genes were followed along the different branches of insect phylogeny, to highlight the evolutionary dynamics of these genes over time, and to discern their origin. The list of species and their CYP4G sequences obtained with their characteristics are listed in Table S1. A FASTA file of 368 CYP4G sequences is provided as Supplementary Material S1. These sequences were aligned along with highest blastp

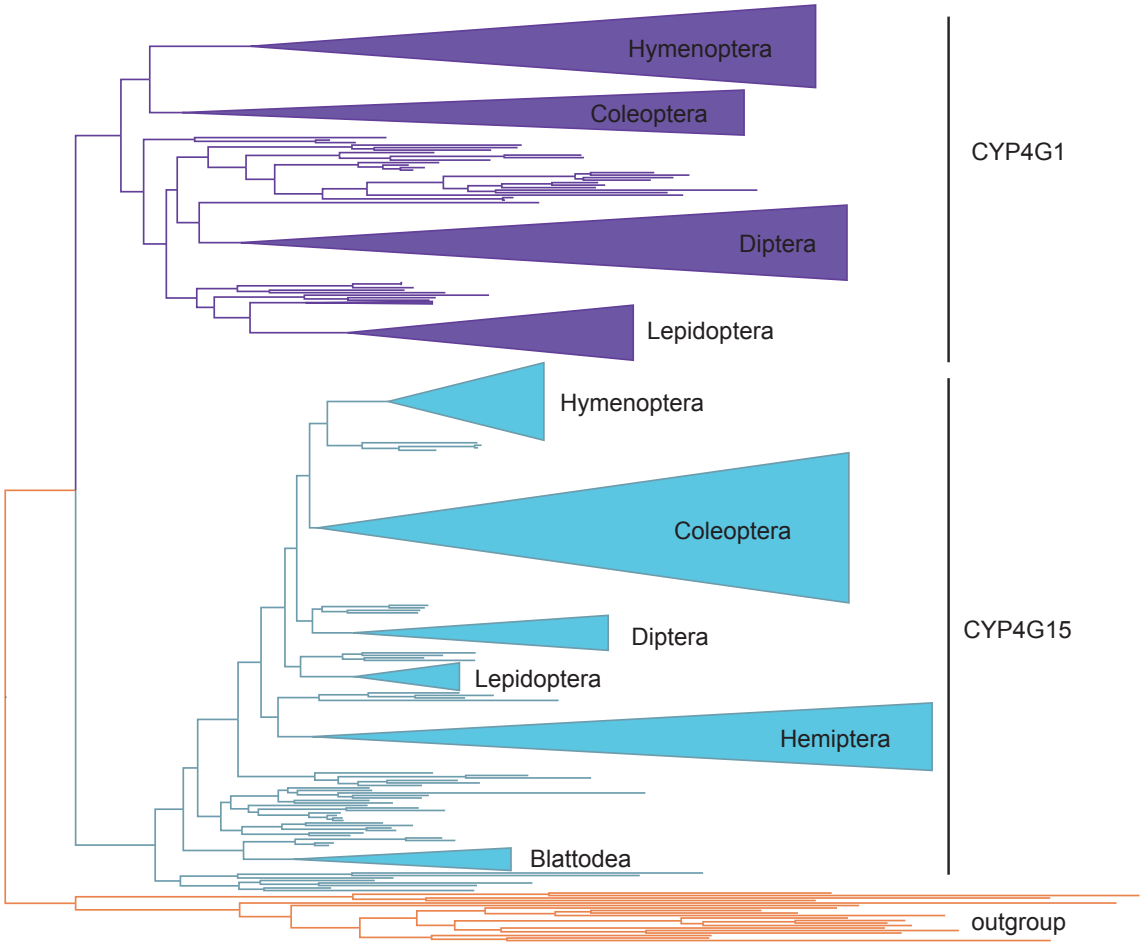

Fig. 1. Phylogeny of the CYP4G genes. Sequences for CYP4G and closely related CYP4 clan P450s (outgroup) were aligned by MAFFT and the tree shown built by RAxML. Sequences in major insect orders were collapsed for clarity. The full tree is provided as Supplementary Fig. 1. Keys to the species names and sequences are in Supplementary Table 1.

scoring P450 sequences from various orders including Archaeognatha, Diplura, Collembola and Protura, that did not contain this insertion. These 20 sequences were thus considered to be non-CYP4G sequences and served as outgroup. The resulting tree of 388 sequences is shown on Fig. 1 and has three clearly distinguished, bootstrap supported clades (see Supplementary Fig. S1 for the full tree with no collapsed branches). All the outgroup sequences cluster together as one clade, while the CYP4G sequences form the two other clades, which are designated as the CYP4G1 and CYP4G15 clades, as they contain the two eponymous sequences from *Drosophila*. To facilitate reading, CYP4G1 and CYP4G15 are hereafter meant *sensu lato*, as the genes belonging to the two orthologous groups (Gabaldón and Koonin, 2013) defined in the tree of Fig. 1, and thus do not imply strict orthology except in the genus *Drosophila*. This is preferred to arbitrary "A" and "B" genes or something equivalent or cumbersome, and it does not imply any special status of *Drosophila*, except and solely with reference to the first complete set of insect P450 genes (or CYPome), that of the fruit fly (Tijet et al., 2001). The bootstrap values within each clade are often low, but each of the CYP4G1 and CYP4G15 sequences representing each insect order cluster together as expected from their high degree of identity (see Section 3.13). Alignments of 148 CYP4G1 sequences and 210 CYP4G15 sequences are provided in Supplementary Materials (S2, S3). The following sections analyze detailed features of the insect CYP4G sequences in various orders.

3.2. Diptera and the origin of the intronless CYP4G1 gene in Drosophila melanogaster

The CYP4G1 gene of *Drosophila melanogaster* is the most highly expressed P450 gene in the fruit fly (Daborn et al., 2002), a remarkable observation given that its expression is almost exclusively limited to the oenocytes (Chung et al., 2009; Qiu et al., 2012). This gene is one of the few intronless P450 genes of *Drosophila*, and this feature was traced in the *Drosophila* phylogeny.

A single intronless CYP4G1 gene was found in 28 genomes of the Ephydroidae superfamily, including all members of the Drosophilinae subfamily within Drosophilidae, with members of the Drosophilini tribe, as well as *Scaptodrosophila lebanonensis* of the Colocasiomyini tribe. Thus, all *Drosophila* relatives have an intronless CYP4G1 gene. However, in the next closest groups (Wiegmann et al., 2011), *Phortica variegata* (family Drosophilidae, subfamily Steganinae) and the shore flies *Ephydra gracilis* and *Cirrula hians* (family Ephydridae) have one or more introns interrupting the coding sequence of CYP4G1 (Fig. 2). In those genomes, the single common intron in CYP4G1 is a phase 0 intron in the sequence coding for the I helix and is an "ancient" P450 intron conserved in many P450 genes at least through Cnidaria. Introns are generally found in other Dipteran CYP4G1 genes, so that the complete loss of introns is a recent event (less than 50 MY).

Although genes can become intronless by gradual loss of introns over time, they can also lose all introns simultaneously when the gene arises as a retrogene (retrotransposed transcript). FlyBase reports that "the CYP4G1 gene may have been derived from the CYP4AC1 gene by

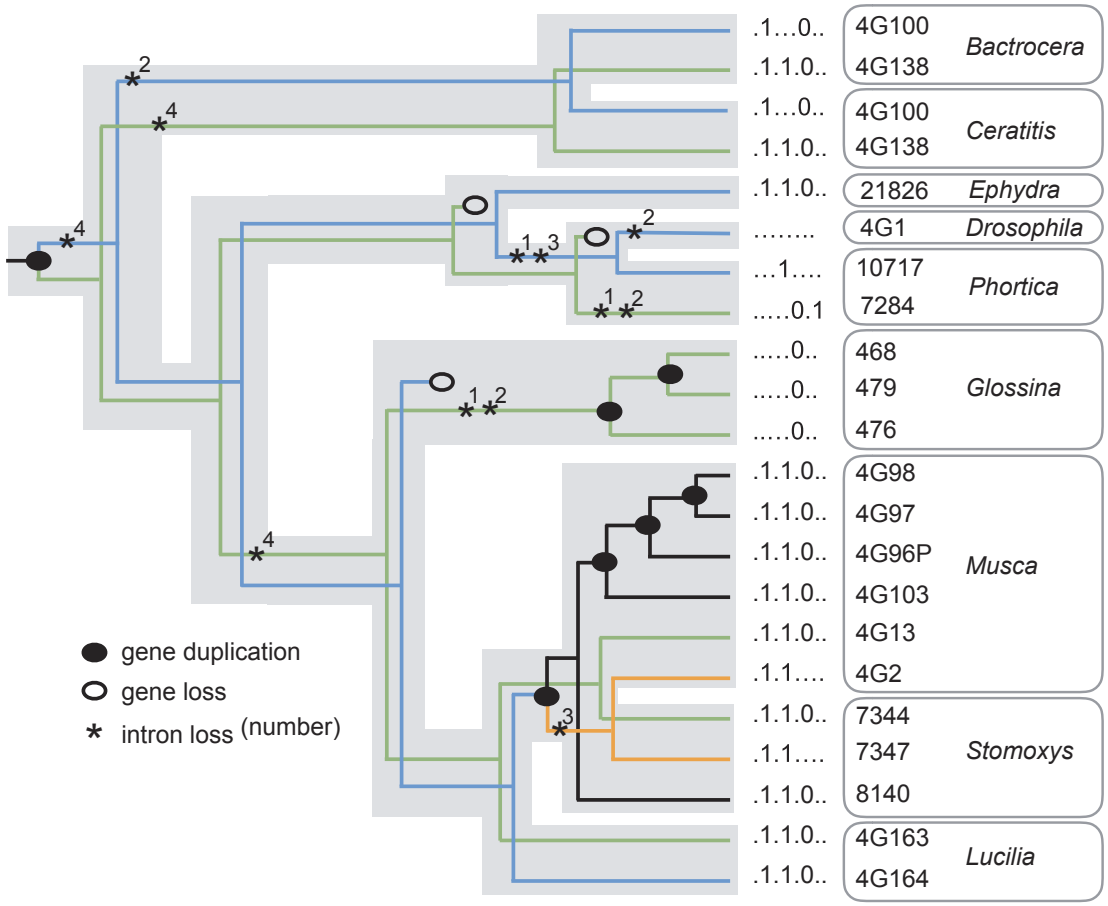

Fig. 2. Phylogeny and intron-exon organization of the CYP4G1 genes of Schizophora. The species tree follows the Dipteran phylogeny of Wiegmann et al. (2011). The gene tree has been superposed onto the species tree (in light grey). Branch lengths are not to scale. Gene duplication is indicated by a full circle, and gene death by an open circle. Position and phase of the introns are indicated graphically next to the gene names. Asterisks with numbers indicate the loss of an intron and the position of the intron. The fourth intron, seen here only in one of the two *Phortica variegata* genes, is an ancient conserved intron (phase 1, one codon after the conserved Cys codon), and the Schizophoran ancestor had four introns. A CYP4G1 duplication occurred before the radiation of Schizophora, so that many species have two CYP4G1 paralogs, while mosquitoes have only one CYP4G1 gene. Several gene duplication and gene death events then followed in various branches (see text for details). For clarity, *Ephydra gracilis* is shown as sister clade to *Drosophila / Phortica*, but a nearly identical gene sequence and structure is found in *Cirrula hians*.

retroposition" (Pan and Zhang, 2009), and is listed in the summary of retrogenes (http://flybase.org/reports/FBrf0225799.html). Yet the closest non-CYP4G sequence in *Drosophila* is CYP4C3 (35% identity) (see also below, origin of CYP4G), and not CYP4AC1 which has only 27% amino acid identity with CYP4G1. As retrotransposition usually inserts the new gene away from its original position in the genome, the synteny of Dipteran CYP4G1 genes with or without introns was compared. Despite the more fragmented quality of some genome assemblies (limited length of contigs), the identity of neighbouring genes was determined in a number of species. Two adjacent genes to CYP4G1 in *Drosophila melanogaster* were searched, the *ebo* (ellipsoid body open, pfam03810, pfam08389) gene, CG3923, is upstream of CYP4G1, and *ase* (asense, or achaete scute complex protein T8, a bHLH protein), CG3258, is downstream of CYP4G1 on the opposite strand. Homologous sequences were found in *Drosophila* species and beyond (Fig. 3).

The arrangement [> *ebo* > *Cyp4g1* < *ase*] was also found in *D. ananassae*, *D. pseudoobscura*, *D. virilis*, *D. mojavensis*, *D. miranda*, *D. willistoni*, although there was an inversion in *D. grimshawi* [> *ase* > *Cyp4g1* < *ebo*]. The coding sequence of all these CYP4G1 genes is intronless. Genome rearrangements obscured syntenic relationships in *Phortica variegata* and *Ephydra gracilis*. However, the head to head arrangement of the CYP4G1 and *ase* orthologs was also found in *Glossina pallipides*, *Musca domestica*, *Stomoxys calcitrans*, *Bactrocera dorsalis*, *Ceratitis capitata* and *Anopheles gambiae*. Beyond Diptera, the [> *CYP4G1* < *ase*] synteny was also observed in some Lepidoptera, Coleoptera, and in *Habropoda laboriosa, Polistes dominula* and *Cephus cinctus* (Hymenoptera). Synteny with *ase* and members of the achaete scute complex genes in such distant species and different orders (Fig. 3), and the presence of one or more introns in CYP4G1 proves that *Drosophila* CYP4G1 is not a retrogene, and that the last intron was lost in the common ancestor of the subfamily Drosophilinae (Fig. 2). Interestingly, in the midge, *Chironomus tentans*, one of two tandem arrayed CYP4G1 genes is also intronless.

3.3. Birth and death of CYP4G1 genes in Diptera

The CYP4G1 gene was found duplicated in several species of Brachycera and Nematocera, but no duplication of the CYP4G15 gene was observed in Diptera. Several duplications of CYP4G1 in Schizophora as well as gene losses make orthology assignments difficult (Fig. 2). Successive duplications have led to a cluster of five CYP4G1 genes and one pseudogene in the house fly (Fig. 3). Although the cactophilic, desert-dwelling *Drosophila* species such as *D. mojavensis* have significantly more cuticular hydrocarbons than *D. melanogaster* (Etges and Jackson, 2001), they have just one CYP4G1 and one CYP4G15 gene as in other species of this family. The midge *Polypedilum vanderplanki* is known for its extreme desiccation tolerance (Gusev et al. 2014). Comparison of its complement of CYP4G genes with that of the related species *P. nubifer* showed that both species carry three CYP4G genes, so that in this case at least, desiccation tolerance cannot be related to an

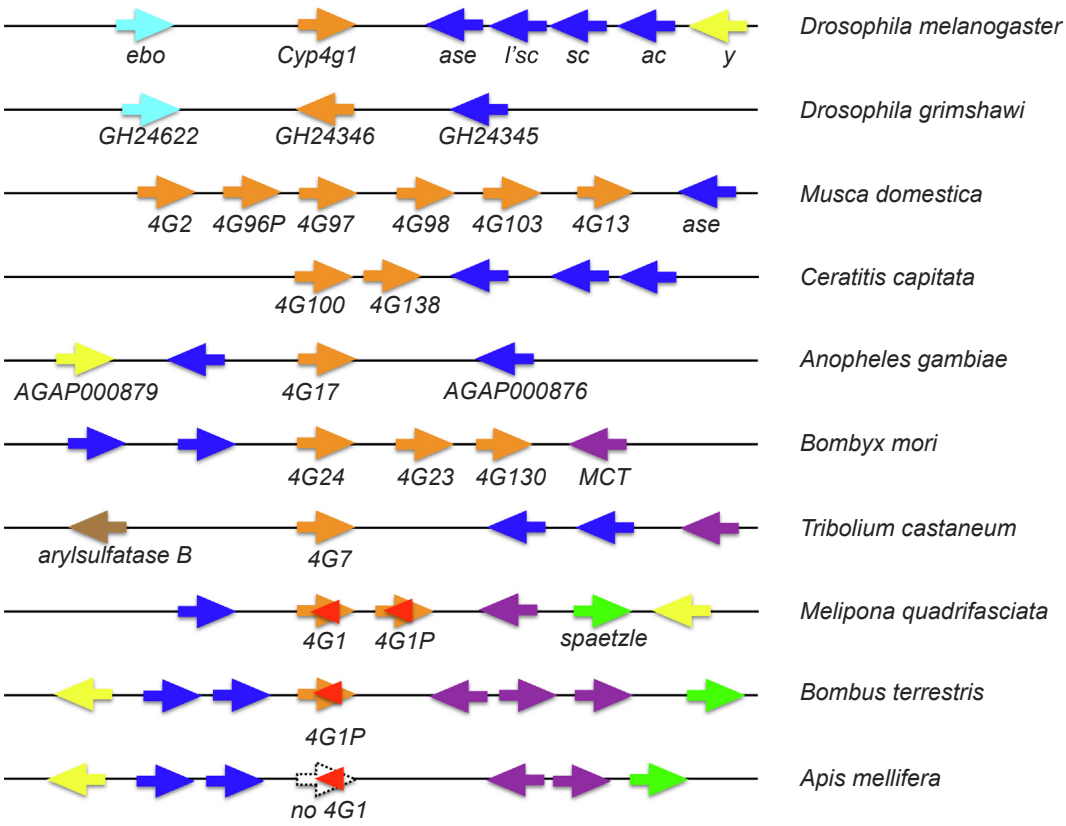

Fig. 3. Synteny at the CYP4G1 locus. The intronless CYP4G1 gene of *Drosophila* is not a retrogene as shown by the conserved synteny from Diptera to at least Hymenoptera. Genes of the same color are orthologs or close paralogs: Orange: CYP4G1; cyan: *ebo*; blue: *ase*; yellow: *yellow*; purple: MCT, monocarboxylate transporter; brown: arylsulfatase B; green: spaetzle. In Hymenoptera, a "cytochrome *c* testis-specific like" gene (red) is inserted on the opposite strand of an intron of the CYP4G1 gene. *Melipona* has one gene and one pseudogene. In *Bombus*, only the pseudogene is remaining, and in *Apis mellifera* this pseudogene has left no trace (stippled arrow), but the cyt c gene is still found, as well as the neighbouring genes, thus identifying the former CYP4G1 locus. Schematic representation, not drawn to scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

increase in CYP4G copy number in the genome.

3.4. The alternative C-termini of CYP4G16 (CYPG15) in A. gambiae

Balabanidou et al. (2016) reported an alternative C-terminus for CYP4G16 in *A. gambiae*. The annotation of CYP4G16 by Vectorbase (AGAP001076) has four transcripts, of which two forms PA and PD differ at the C-terminal protein sequence. The PD form was not found experimentally by Balabanidou et al. However, not predicted by Vectorbase but found by Balabanidou et al. at a low level was a form they called CYP4G16-PD1, in which the last intron is not spliced out (Fig. 4). This form is 160x less expressed, which, taking into account the very high basal expression, is still significant. This in-frame run-through translation cannot occur in *A. aegypti*, nor is there a possibility for an alternate 5th exon. The PD version of CYP4G16 in Vectorbase is therefore probably erroneous, but the experimentally discovered PD1 version suggests that variation can occur at the C-termini of CYP4G genes.

3.5. CYP4G in Mecoptera and Siphonaptera

There is no genome available to date for Mecoptera. TSA evidence shows that *Boreus hyemalis* and *Nannochorista philpotti* express both CYP4G1 and 4G15 genes. However, *Bittacus pilicornis* and *Panorpa vulgaris* express only one each, CYP4G15 and CYP4G1 respectively. Given the high quality TSA data this would suggest either reciprocal gene losses in these two families or alternative expression. In Siphonaptera, TSA shows expression of both CYP4G genes in *Oropsylla silantiewi*, *Ceratophyllus gallinae* (not shown) and *Ctenocephalides felis*. The cat flea genome reveals a more complex background, as it has a tandem duplication of the CYP4G15 gene (with identical open reading frame), head to head with a CYP4G15 pseudogene (> Pse (5 last exons) < XP_026468773 < XP_026468774). Interestingly the cat flea also has two duplicated, head to head CYP4G1 genes, differing by just one amino acid, with a CYP4G1 pseudogene in between them.

3.6. CYP4G in Lepidoptera and Trichoptera

Lepidopteran genomes have both CYP4G genes, with one copy each

```
AGAP001076-PA  503 RSCV g(110 nt)ga GRKYAMLKLKIILSTILRNFRVYSDLKEEEFKLQADIILKREEGFQIRLEPRQRKSKTL 565*

AGAP001076-PD  503 RSCV g( 72 nt )tg                              VKDTNIYIITSFFQDVNMLC 525*

Experiment PD1 503 RSCV (read through) GKRKIIKKVKYKFKKIKFLCKQFAVKDTNIYIITSFFQDVNMLC 549*
```

Fig. 4. Alternative C-terminal sequences of CYP4G16 in *A. gambiae*. PA is the major form, PD1 a very minor form experimentally found (Balabanidou et al., 2016) in which the last intron is unspliced and read through and PD is probably an erroneous annotation in Vectorbase, unsupported by any *Anopheles* sp. transcript (TSA or EST). Modified from Balabanidou et al. (2016).

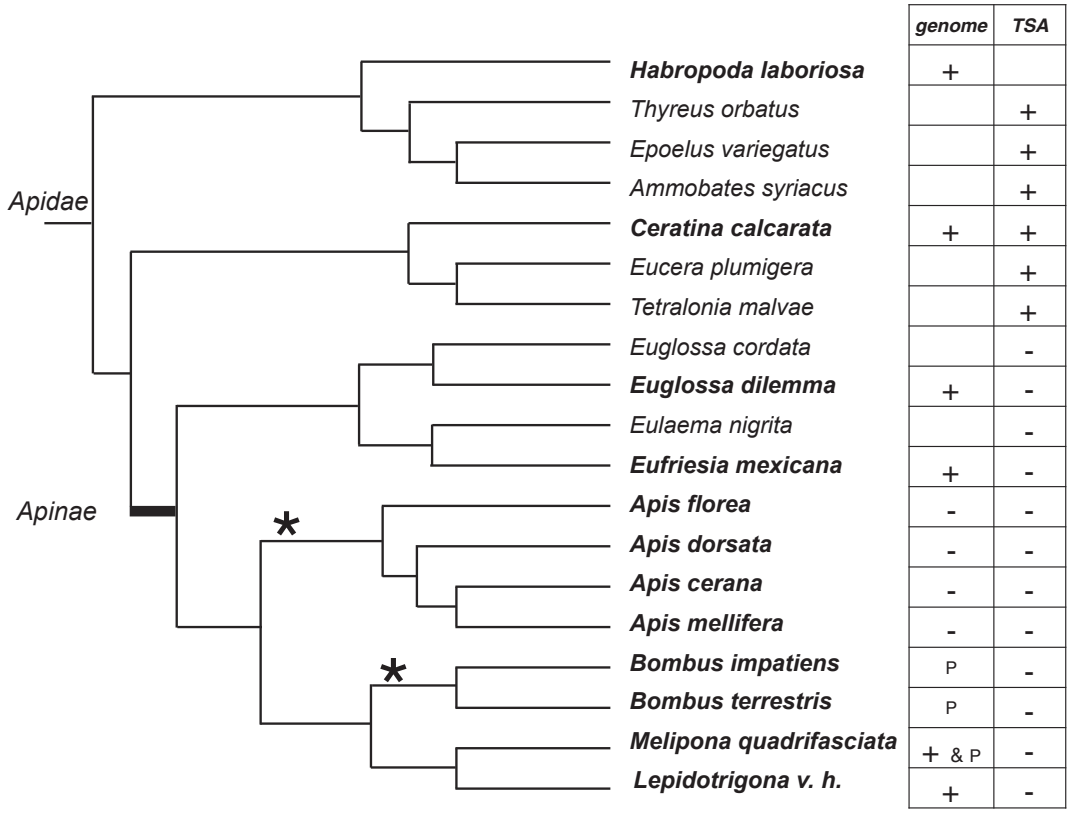

Fig. 5. Loss of the CYP4G1 gene in Apidae. Phylogeny of the Apidae adapted from Bossert et al. (2019) with branch lengths not drawn to time scale. The root of the Apidae shown here dates to approx.100 MYA (Peters et al., 2017; Sann et al., 2018). Presence or absence of the CYP4G1 gene in the genome or TSA data are shown in the table. Species for which a genome is currently available are marked in bold (+/-: gene present absent; P: pseudogene). Bold line for the Apinae clade suggests loss of CYP4G1 expression. Asterisks show independent losses of the CYP4G1 gene.

in the "basal" *Plutella xylostella* and in Papilionoidea (nine species in five families) (Table S1). However the CYP4G1 gene has been duplicated several times independently, in Tortricidae and Gelechioidea as well as in eight species of moths. Several tandem duplications are noted, forming small clusters. In *Helicoverpa armigera,* the CYP4G1 cluster is arranged [< CYP4G9 < CYP4G26 > CYP4G10 < CYP4G87 < > x < ase] thus maintaining the synteny with ase (see above). Similarly, there is a cluster of three CYP4G1 in *Spodoptera litura*, *Manduca sexta* and *Bombyx mori* (Fig. 3; Table S1). Initiation of diapause in the silkmoth *Antheraea yamamai* is associated with high CYP4G25 (BAD81026) expression (Yang et al., 2008). In caddisflies (Trichoptera), transcriptomes of several species as well as the genome of *Stenopsyche tienmushanensis* indicate the presence of both CYP4G1 and CYP4G15, and in *Annulipalpia* sp. each has been duplicated at least once.

3.7. CYP4G in Coleoptera and in the sister groups of Coleoptera

Coleoptera are the most diverse order of insects in number of species, and aspects of their phylogeny remain unresolved. The phylogenies of McKenna et al. (2015) and Zhang et al. (2018) were used to guide a representative survey of genomes and TSA from 28 species in 17 superfamilies. Most carried both CYP4G1 and CYP4G15 genes, but there were no clear patterns, perhaps owing to the long evolutionary history of the order, with the majority of extant families dating to the Cretaceous. Moreover, the exact number of CYP4G genes was often obscured by fragmented and incomplete genome assemblies. In Myxophaga, *Lepicerus* sp. expresses no CYP4G1, but at least three CYP4G15 genes, while in Arcostemata, fragmentary evidence from *Priacma serrata* TSA suggests the expression of two CYP4G15 genes and one CYP4G1 gene.

Adephaga were represented in Caraboidea by the genome of *Pogonus chalceus* which carries two CYP4G15 genes, both are expressed, but no CYP4G1 was found. The TSA of *Carabus granulatus* also shows expression of just two CYP4G15 genes. In Dytiscoidea and Gyrinoidea, *Cybister lateramarginalis*, *Liopterus haemorrhoidalis* and *Gyrinus marinus* (TSA) express a single copy of each CYP4G1 and CYP4G15 genes, but *Amphizoa insolens* also expresses a second CYP4G15 gene.

The evidence from Polyphaga is similarly complex with multiple gene birth and losses in three of the major branches, the Staphyliniformia and Scarabaeoidea, the Elateriformia and the Cucujiformia. In Staphylinoidea both CYP4G1 and CYP4G15 genes are found in the genome of *Nicrophorus vespilloides* (Silphidae), and both are expressed in *N. vespilloides* and *N. orbicollis*. In contrast, *Aleochara curtula* (Staphylinidae), has no CYP4G15 transcript but two CYP4G1 transcripts. Transcripts of CYP4G1 and CYP4G15 are found in Histeroidea, but there is only a CYP4G15 transcript in Hydrophiloidea.

In Scarabaeoidea, the genome of *Ontophagus taurus* has only a CYP4G1 gene, and in *O. nigriventris*, *Popillia japonica* and *Oxysternon conspillatum*, only a CYP4G1 transcript is found. However, in *Trypoxylus dichotomus* and *Heterochelus* sp. transcripts of both CYP4G1 and CYP4G15 are found. The fragmented genome assemblies of *Oryctes borbonicus* and *Protaetia brevitarsis* appear to carry two or more CYP4G genes.

Elateriformia show multiple instances of CYP4G15 duplications. In Dascilloidea, *Anorus arizonicus* expresses four distinct CYP4G15 genes, three of them found as full length transcripts with no trace of a CYP4G1 transcript. The genome of *Agrilus planipennis* (Buprestoidea) carries one copy each of CYP4G1 and CYP4G15, while only two CYP4G15 transcripts are found in *Dryops* sp. (Byrrhoidea). Major duplication events are seen in Elateroidea: In fireflies and click beetles (Lampyridae and Elateridae) the CYP4G1 gene was lost, but the CYP4G15 gene was

repeatedly duplicated to three clustered genes in *Aquatica lateralis*, five in *Ignelater luminosus* (three of which are incomplete or pseudogenes, the genome assembly is fragmented), six in *Photinus pyralis* and ten genes in *Pyrocoelis pectoralis*, of which eight are clustered on a single scaffold. In each case, the cluster of CYP4G15 genes is flanked by a CYP4C-like gene. TSA data show that all six *Photinus* genes are expressed. Duplications of the CYP4G15 gene started to occur over 125MYA. If bioluminescence evolved independently in click beetles and in fireflies (Fallon et al., 2018), then the expansion of CYP4G15 genes predated the emergence of bioluminescence. Whether it is related at all to bioluminescence remains conjectural, but several observations would justify a closer examination. The firefly lanterns of *P. pyralis* and *A. lateralis* highly and differentially express the same CYP4G gene (Fig. 5 of Fallon et al., 2018) one of only 26 genes in this class. Photogenic tissues of *Phrixothrix hirtus* and *Aspisoma lineatum* also highly express CYP4G genes (Amaral et al., 2017a,b).

Significantly, luciferase has evolved from fatty acid acyl-CoA synthase, an enzyme highly enriched in oenocytes (e.g. Gutierrez et al., 2007), as are the CYP4G enzymes. Although the origin of photocytes and lanterns is often ascribed to fat body cells, it is possible that the fat body cells located close to abdominal epidermis were misidentified, being in fact oenocytes. At least in *Phengodes* and *Phrixothrix* lanterns are reported to be of oenocyte origin (Buck, 1948; Viviani, 2002). The high oxygen consumption of luciferase is accomodated by the unusually rich tracheolization of light organs (Buck, 1948; Smith, 1963). One attractive hypothesis would be that the CYP4G enzymes are providing the hydrocarbons needed to coat the inner epicuticle of tracheal tubes (Moussian, 2010), and thus play an important role in oxygen supply for light flashing (Tsai et al., 2014). Two other and unrelated close associations between CYP4G-derived CHC and specialized trachae are found around the subgenual organ in Gasteruptiid wasps (Mikó, 2019), and possibly close to the tympanal membranes of tree weta “ears” (Lomas et al., 2012) as well.

The very diverse Cucujiformia also show diverse patterns of CYP4G gains and losses. In the genomes and TSA of Tenebrionoidea (*Tribolium castaneum* and *Tenebrio molitor*), Cleroidea (*Thanasimus formicarius*) and Chrysomeloidea (*Leptinotarsa decemlineata* and *Anoplophora glabripennis*), the 1:1 pattern CYP4G1:CYP4G15 is found. In Curculinoidea, the CYP4G15 gene was duplicated in *Hypothenemus hampei*, but not in *Dendroctonus ponderosae*. However, there is no CYP4G1 and only a single CYP4G15 in the genome of Coccinelloidea. In *Harmonia axydiris*, four available genomes were searched but all had a only a single copy of the CYP4G15 gene. This was also the case for the genome of *Coccinella septempunctata* and the TSA of *Serangium japonicum* (GGMU01045764), as well as the TSA of *Xylobiops basilaris* (Bostrichoidea). In *Aethina tumida* (Cucujoidea) no CYP4G1 was found whereas three CYP4G15 genes were present, two of them in close proximity to a CYP4C-like gene. These data suggest that the CYP4G1 gene was lost independently in Coccinelloidea, Bostrichoidea and Cucujoidea.

In conclusion, at least eight independent losses of CYP4G1 and at least six duplications of CYP4G15 can be counted in Coleoptera, with, in addition, the CYP4G15 bloom(s) in Elateriformia. The earlier diversification of Coleoptera, when compared to Lepidoptera, may partially explain the greater “instability” of CYP4G genes in Coleoptera because there was nearly double the evolutionary time for changes to occur.

Evidence for the presence in the genome and expression of both CYP4G genes in *Mengenilla moldrzyki* (Strepsiptera) was obtained, but the sources are fragmented and only the CYP4G15 gene could be fully reconstructed. Both CYP4G1 and CYP4G15 genes are found in Neuroptera. In *Pseudomallada prasinus* the TSA sequences are apparently contaminated with CYP4G sequences of hymenopteran origin (GAVV02039744 and GAVV02048646, not shown on tree). The evidence from Megaloptera is more tenuous, lacking full length transcripts of the CYP4G genes. Alignment of the fragments with the Neuropteran genes shows that *Corydalus cornutus* expresses both types of CYP4G genes. There is no transcript evidence for a CYP4G1 gene in Raphidioptera, but the CYP4G15 gene has been duplicated with two copies in *Inocellia crassicornis* and four copies in *Xanthostigma xanthostigma*.

3.8. Origin of the single CYP4G gene in the honey bee

When the sequence of the honey bee genome became available, it stood out as an outlier within insects. It carries relatively fewer P450 genes than other insects (Claudianos et al., 2006) and has just four members of the CYP4 clan. Among these is a single CYP4G gene, CYP4G11, which is a CYP4G15 gene (Fig. 1). The second genome of the order Hymenoptera, that of the jewel wasp *Nasonia vitripennis*, has a more typical number of P450 genes (about 91) and also the more typical pair of CYP4G1 and 4G15 genes. The increasing number of hymenopteran genomes now available and the improved phylogeny of Hymenoptera (Peters et al. 2017) in general and Apoidae in particular (Sann et al., 2018; Bossert et al., 2019) made it possible to determine when, during the > 180 MY of evolution of bees and apoid wasps, the CYP4G1 homolog was lost.

No CYP4G1 is found in the genomes of Apini (*Apis mellifera*, *A. florea*, *A. cerana*, *A. dorsata*). Their sister groups the Bombini and Meliponini, however, show either pseudogenized or complete CYP4G1 genes. In Bombini, *Bombus terrestris* and *B. impatiens* CYP4G1 genes are truncated, having lost the first three exons, and the remaining five exons contain several frameshift causing indels. There is no transcript evidence for a CYP4G1 gene in any *Bombus* sp., whereas the CYP4G15 genes in *B. terrestris* and *B. rupestris* have full transcript support. In Meliponini, *Melipona quadrifasciata* CYP4G1 is represented by two tandemly arrayed genes, of which one is a recent pseudogene (one nt indel frameshift) (Fig. 3). A CYP4G1 gene is also found in *Lepidotrigona ventralis hoosana*. Synteny relationships (Fig. 3) show that the CYP4G1 locus is conserved, and that the loss of CYP4G1 in Bombini and Apini was not due to a single segmental deletion, but rather to a more gradual process of pseudogenization and obliteration. Indeed, the *Bombus terrestris* CYP4G1 pseudogene is flanked by *yellow* and *ase* homologs, so it has retained the syntenic relationships described above. In the honey bee, this genomic region is also found, but without a remaining trace of the CYP4G1 gene. The “cyt c testis-specific-like” gene that is found in an intron of other hymenopteran CYP4G1 genes is still present in the honey bee (Fig. 3).

The relative position of honeybees, stingless bees and orchid bees has recently been resolved (Bossert et al., 2019). As *Melipona* and *Bombus* are in the same clade, it implies that the CYP4G1 gene was lost at least twice, once in the *Apis* lineage and once in the *Bombus* lineage. Indeed, the closest clade to honeybees and stingless bees are the orchid bees represented by the genomes of *Euglossa dilemma* and *Eufriesea mexicana*. Each have both CYP4G1 and 4G15 genes (Fig. 5), although transcripts are only seen for the 4G15 gene in the *Euglossa* genus (*E. dilemma* and *E. cordata*) and in *Eulaema nigrita*. Similarly, there is no transcript evidence in TSA for the CYP4G1 genes of *Melipona* or *Lepidotrigona*. Although tentative, this would suggest that the losses of the CYP4G1 genes in Apinae may have been preceded by their transcriptional silencing (Fig. 5).

The sister clade to Apinae is the Eucerinae, for which there is currently no genomic evidence. However CYP4G1 genes are expressed in *Eucera plumigera* and *Tetralonia malvae*, while a CYP4G15 gene is found in the transcriptomes of *Tetraloniella nigriceps* and *E. syriaca* females (but not *E. plumigera male*). The adult specimens of the two *Eucera* species were sampled at the same time and place, and differ only by sex, suggesting a sex-specific expression of the CYP4G genes in this family. In Xylocopinae (carpenter bees), the two genes are found, *Ceratina calcarata* (genome) and *C. chalybea* (transcriptome, GBPU01009783 and GBPU01010525). In Nomadinae (cuckoo bees), there is currently no genome, but the two genes are found in the TSA of four species. In Anthophorinae (digger bees), The *Habropoda laboriosa* genome appears

to carry both CYP4G genes, although the CYP4G1 gene is fragmented by gaps. *Anthophora plumipes* expresses both genes, but the CYP4G1 transcript is partial and unspliced. Evidence for a functional CYP4G1 gene in digger bees thus remains somewhat conjectural.

Beyond Apidae, species representing ten superfamilies (Apoidea, Pompiliodea, Vespoidea, Cynipoidea, Chalcidoidea, Ichneumonoidea, Orussoidea, Cephoidea and Tenthrinidoidea) carry one copy each of the CYP4G1 and CYP4G15 genes. In ants, (Formicoidea) however, the CYP4G1 gene is duplicated in several species, particularly in Ponerine ants, with up to three genes and one pseudogene in *Harpegnathos saltator* and three genes and seven fragments/pseudogenes in *Dinoponera quadriceps*. There are also two CYP4G1 pseudogenes in *Linepithema humile*, when overall CYP4G pseudogenes are not frequent in this insect-wide survey, suggesting an unstable expansion of the CYP4G1 gene in ants. CHC play particularly important and diverse chemical communication roles in ants, and genes for odorant receptors tuned to CHC are expanded in e.g. *H. saltator* (Pask et al., 2017). Evidence that CYP4G enzymes help determine the complexity of the CHC blend (Kefi et al., 2019) suggests therefore that there may be a reciprocal evolution of the biosynthesis of CHC and of their receptors in ants. Ants also show fine tuning of CYP4G expression during the switch from nursing to foraging. In *Camponotus fellah*, one of the CYP4G1 genes is upregulated when workers leave the nest to forage, in response to signalling by the peptide inotocin whose receptor is expressed in oenocytes (Koto et al., 2019).

3.9. CYP4G in Psocodea

The phylogeny of Psocodea was recently clarified by Johnson et al. (2018) and the only representative to date with a full genome is the human body louse, *Pediculus humanus*, while the other groups of parasitic lice, bark lice and book lice are only covered by TSAs. Within Phthiraptera, *Pediculus* has two CYP4G genes, CYP4G39 (4G1) and CYP4G38 (4G15). The CYP4G39 gene is incorrectly annotated in XP_002423839 and in Vectorbase, missing an exon. The TSA of *Menopon gallinae* shows several CYP4G transcripts, with at least full length representatives of CYP4G1 and CYP4G15 (not shown). In book lice, *Liposcelis bostrychophila* and *Badonnelia titei* (incomplete sequences, not shown) also express the two CYP4G genes. In Psocomorpha, however, only CYP4G15 transcripts were found, one in *Aaroniella sp*. (GDEY01001503) and two in *Bertkauia sp*. No CYP4G1 was found in the TSA of Trogiomorpha (*Cerobasis guestfalica*). This suggests two independent losses of CYP4G1 in Psocodea, although only genomic sequences can verify this.

3.10. Hemipteroid insects and the origin of the single CYP4G gene in the pea aphid

A similar question to the presence of a single CYP4G gene in honeybees was raised by the presence of a single CYP4G gene in the pea aphid *Acyrthosiphon pisum*. In this species, extensive duplications were reported in many gene families (Consortium, 2010), yet CYP4G51 is the single CYP4G member found in aphids. The function of this CYP4G15 gene in cuticular hydrocarbon biosynthesis was demonstrated (Chen et al., 2016).

A survey from the pea aphid to beyond Aphididae revealed that the CYP4G1 gene was absent not just in aphids, but apparently in the whole order Hemiptera. The genome of a number of species of Hemiptera have been sequenced, and the loss of the CYP4G1 gene seems to have been accompanied by duplications of CYP4G15 in Hemiptera, with aphids the apparent exception with a single CYP4G15 gene. Phylogenetic analysis of the hemipteran CYP4G15 genes in the three major groups, Sternorrhyncha, Auchenorrhycha and Heteroptera or true bugs, as well as the sequences from thrips, the sister group to Hemiptera, suggests that at least one duplication of CYP4G15 occurred in their common ancestor, further implying that the Aphidoidea lineage actually lost one of two CYP4G15 gene recently.

The whitefly *Bemisia tabaci* has four CYP4G15 genes of 12 coding exons each. Three of the genes, CYP4G68, 128 and 129 are clustered on the chromosome. The transcriptome of the closely related *Trialeurodes vaporarium* shows three CYP4G15 genes. The brown planthopper *Nilaparvata lugens* has two CYP4G15 genes. Among psyllids, *Pachypsylla venusta* has a single intronless CYP4G15 gene, whereas *Diaphorina citri* appears to have five genes including three pseudogenes (four genes in tandem array with two partial genes). The bed bug *Cimex lectularius* has two sets of tandemly duplicated CYP4G15 genes. The brown marmorated stink bug *Halyomorpha halys* has two clusters of CYP4G15 genes, one with six tandem duplicated genes and another with two. Both clusters are flanked by paralogs of an arylsulfatase B-like gene, indicating an ancient segmental duplication encompassing more than the clustered P450 genes. Two P450s of the six gene cluster have been named into new subfamilies (CYP4GW1 and CYP4GX1) owing to their divergence from the original CYP4G sequence. However, they have maintained the 12 exons structure of their close paralogs. Three CYP4G15 genes are expressed in the harlequin bug, *Murgantia histrionica*. It is tempting to suggest that the loss of the CYP4G1 gene in the ancestral hemipteran lifted some constraint on the evolution of the CYP4G15 gene, leading not just to several gene gains and losses, but also to greater sequence divergence as seen in the stinkbug *H. halys*. This idea would then lead to the further suggestion that following the initial duplication of the CYP4G in a basal neopteran insect, the presence of two CYP4G genes as seen in so many insect lineages had a mutually stabilizing influence, the nature of which remains enigmatic.

Frankliniella occidentalis (Thysanoptera) has no CYP4G1 gene but two 4G15 genes. Each of the two genes appear duplicated in *Gynaikothrips ficorum*. Thysanoptera are accepted as sister group of Hemiptera, and the lack of CYP4G1 genes in this group suggests a single loss after divergence from the [Psocodea: Holometabola] clade between 400 MYA (Johnson et al., 2018) and 360 MYA (nodes 37–105 of Misof et al., 2014). Although the precise sequence of events is unknown, it is tempting to suggest that a CYP4G15 duplication was concomitant to / compensated for the loss of CYP4G1 in the common ancestor of Thysanoptera and Hemiptera about 400 MYA.

3.11. CYP4G genes in Polyneoptera

In Polyneoptera (Wipfler et al., 2019) the number of available genomes is more restricted, and the available transcriptomes can prove the presence, but not the absence of a gene. No CYP4G1 gene can be found in the genomes of Blattodea or in the transcriptomes of Mantodea. The CYP4G15 gene is found in one copy in *Blattella germanica* but is duplicated in *Periplaneta americana,* where each copy is expressed. In termites (Blattodea-Isoptera), the genomes of *Zootermopsis nevadensis* and *Cryptotermes secundus* have just one CYP4G15 gene, while that of *Macrotermes natalensis* has three, as noted by Harrison et al., 2018 (who mistakenly called these genes CYP4G1). Two of these genes were missing one the 12 exons and could not be fully reconstructed because of the many small gaps in this large 1.3 Gb genome. Two CYP4G15 genes are expressed in *Coptotermes gestroi* and at least five species of Mantodea also express two different CYP4G15 genes. As Blattodea/Isoptera and Mantodea form a monophyletic clade, the loss of the CYP4G1 gene probably occurred once, before node 117 of Misof et al. (2014) over 200 MYA.

The genomes of Phasmatodea (*Timema cristinae* and *Dryocelus australis*) carry both CYP4G1 and 4G15 genes. The 4G1 gene is duplicated in *T. cristinae* and evidence for duplication of CYP4G1 also comes from the TSA of *Aretaon asperrimus* (GAZQ02040172, not full length) and *Peruphasma schultei*. The latter does not express (or lacks) the CYP4G15 gene. In Embioptera (web spinners), both CYP4G genes are expressed, and the CYP4G1 gene is duplicated at least in *Aposthonia japonica* (partial transcript GAWU02032889) and in *Ptilocerembia catherinae*. Only two species of Grylloblattodea have data in the TSA database, and

each appears to express only a CYP4G15 gene, with no trace of a CYP4G1 transcript. Similarly, only two species of Mantophasmatodea have been studied to date, both express a CYP4G1 gene, and incomplete transcripts suggest the presence of a CYP4G15 gene as well. A definitive answer will require mining of a genome from this order.

In Orthoptera, the transcriptomes of *Stenobothrus lineatus*, *Aularches miliaris* and *Tetrix sp*. (Caelifera, Acrididea) showed the expression of at least one copy each of the CYP4G1 and CYP4G15 genes. The genomes of *Locusta migratoria* and of the closely related acridid *Xenocatantops brachycerus* have indeed one copy of both genes. In Ensifera as well, TSA of *Ceuthophilus* sp. (Rhaphidophoroidea) and *Nippancistroger testaceus* (Tettigoniidea) revealed both CYP4G1 and CYP4G15 transcripts. In contrast to other Orthoptera, no CYP4G15 gene was detected in the genome of the cricket *Laupala kohalensis* (Ensifera, Gryllidae), and the TSA of other crickets (*Laupala cerasina*, *Acheta domesticus*, *Gryllus firmus*) also indicated a single CYP4G1 gene.

No CYP4G1 gene was found in genomes or transcriptomes of Plecoptera (stoneflies), but the CYP4G15 gene appears to be duplicated at least once in the genome of *Lednia tumana* (genes incomplete, short contigs). There are two transcripts from *Apteroperla tikumana* and from *Perla marginata* (one short at N-term, not shown) and three from *Leuctra sp*. After the divergence of Dermaptera, Plecoptera are considered a sister group to all other Polyneoptera (Misof et al., 2014; Wipfler et al., 2019).

In Dermaptera (earwigs), there is only evidence for the expression of CYP4G15 in six species including *Forficula auricularia*, *Apachyus charteceus* and *Diplatys* sp., so the absence of CYP4G1 may be either the loss of the gene or a very low expression level. In *Diplatys* sp. three CYP15 genes are expressed. Evidence from Zoraptera (ground lice) is only fragmentary, but *Zorotypus caudelli* expresses at least one CYP4G15 gene, as does *Z. gurneyi* (partial transcripts GABA01002401 and GABA01008920).

3.12. The original insect CYP4G duplication and origin of the CYP4G gene in the CYP4 family

As four orders of Polyneoptera (Orthoptera, Mantophasmatodea, Embiptera and Phasmatodea) carry a CYP4G1 gene that is monophyletic with all other CYP4G1 sequences (Fig. 1), the initial CYP4G1/4G15 duplication must have occurred about 400 MYA (between nodes 133 and 134 of Misof et al., 2014). CYP4G1 was then lost four independent times in Polyneopteran lineages (once in Blattodea / Mantodea, once in Grylloblattodea, once in Plecoptera, and once in Zoraptera/Dermaptera) (Fig. 6). Furthermore, the presence of the two types of genes in Holometabola and Polyneoptera (e.g. *Locusta*), but just of the CYP4G15 gene in Condylognatha (aka Acercaria, i.e. Hemiptera and Thysanoptera) would suggest an additional loss of the CYP4G1 gene around node 37 about 350 MYA.

The genomes and transcriptomes as available for this survey provided a clue to the timing of appearance of the CYP4G genes. No CYP4G sequence was found in Paleoptera (Ephemeroptera and Odonata), in Archaeognatha (wingless bristletails) and Zygentoma (silverfish) or indeed outside the class Insecta, in "basal" Hexapoda such as Protura, Collembola and Diplura. No CYP4G sequence was found beyond Hexapoda, in Crustacea, Myriapoda or Chelicerata. The absence of CYP4G gene in Paleoptera and beyond indicates that the CYP4G genes are not common to all insects, and that the proposal of Qiu et al. (2012) that CYP4G were a key innovation contributing to the success of land colonization by a hexapod ancestor needs to be revised.

Although there is a wealth of information on CHC in a wide variety of species, there is little information on CHC in Paleoptera (Odonata and Ephemeroptera) in which no CYP4G gene (in *Calopteryx splendens* and *Ladona fulva*) or CYP4G transcript (in the TSA of six species) can be found. This raises the question of how these insects protect themselves from desiccation. The elegant work of Futahashi et al. (2019) provides an answer. They showed that dragonfly waxes which are also responsible for color changes and UV light reflection are unusual (for insects), consisting mainly of very long chain fatty aldehydes (C24 to C30) and 2-methyl ketones (C25 to C29), and not the usual (for insects) alkanes or alkenes. Apparently then, CYP4G substrates (long chain fatty aldehydes) and products of modified fatty acyl-CoA oxidation (methyl ketones) are sufficiently hydrophobic to be functional equivalents of hydrocarbons, and indeed the cuticular surface shows a high degree of water repellency (Futahashi et al., 2019). However, the older work of Jacob and Hanssen (1979) noted the presence of hydrocarbons as a small proportion of total cuticular lipids in two *Aeshna* (up to 5%) and two *Sympetrum* species (up to 29%), although these authors did not report absolute quantities. These data are difficult to reconcile with those of Futahashi et al. who did not detect any CHC on a different *Sympetrum* species and on two *Orthetrum* species. There are two possibilities. Either the CYP4G genes evolved before the radiation of Paleoptera, and were subsequently lost in some lineages, or the CHC that are found in some Odonata are not synthesized by a CYP4G enzyme, but rather by a CYP4C enzyme. The second hypothesis would be preferred. Since CYP4G evolved from a CYP4C (see below), some oxidative decarbonylase activity may have been present in an ancestral, progenitor CYP4C enzyme, and it was the optimization of this activity ("neofunctionalization") that makes a CYP4G distinctive. Some Paleoptera may therefore make small amounts of CHC with a CYP4C enzyme but depend mostly on other lipids to coat their epicuticle, whereas the Neoptera see the predominance of CHC.

Beyond insects, and in the absence of CYP4G genes, the question of the origin of alkanes and alkenes in some arthropods is raised. A confounding factor can be the acquisition of CHC from insect prey as in scorpions or spiders, so that clear evidence of synthesis is needed. There is convincing evidence that astigmatic mites can synthesize alkane and alkenes de novo from acetate, with a terminal aldehyde as probable intermediate, as in insects (Shimizu et al., 2017). However, the decarbonylation enzyme has not been identified to date.

The origin of CYP4G genes is therefore in the Devonian around 400 MYA, between nodes 133 and 134 of Misof et al. (2014), shortly before the duplication into CYP4G1 and 4G15 discussed above (Fig. 6). It can only be speculated whether the considerable investment in structural CHC represented by the evolution of the O_2-dependent CYP4G enzymes was facilitated by the increasing oxygen levels at that time or was a less toxic alternative to the use of aldehydes and 2-methylketones as waterproofing agents of the epicuticle. The second hypothesis is more plausible, as P450 enzyme function (today) is not limited, even by oxygen concentrations far below that of the Cambrian.

As the closest P450 sequences to the CYP4G are members of the CYP4C subfamily, it was hypothesized that the original CYP4C-CYP4G (tandem) duplication may have left a trace such as a close synteny between a CYP4G and a CYP4C gene. Genomes were therefore searched for close association between CYP4G genes and other CYP4 genes. Unfortunately, too many key genomes have very fragmented assemblies, so synteny can be difficult to assess. One of the *Halyomorpha halys* CYP4G15 clusters (see above) consists of four CYP4G15 paralogs and two closely related sequences, CYP4GW1 and CYP4GX1. However, these two are clearly recent genes that have diverged from a CYP4G15 ancestor. In contrast, several examples of CYP4 synteny with CYP4G were found in Coleoptera and in Isoptera (Blattodea), along with highly conserved non-P450 genes. Comparative analysis suggests that the CYP4G genesis locus (to adopt the formulation of Nelson et al., 2013) contained the originator CYP4C gene, the newly duplicated CYP4G gene, a perilipin gene, a TWIK potassium channel gene and an arylsulfatase B gene. Fig. 7 schematically shows the conservation of this locus throughout the phylogeny. In Hemiptera and in Hymenoptera, rearrangements have split the original locus, but some members have retained their syntenic relationship. In both termites and Coleoptera, it is a CYP4G15 (or a cluster of CYP4G15 genes in fireflies) that was found close to a CYP4C gene. In the stick insect *T. cristinae*, the CYP4G1 and 4G15 genes are present as a tandem duplicate. This analysis therefore

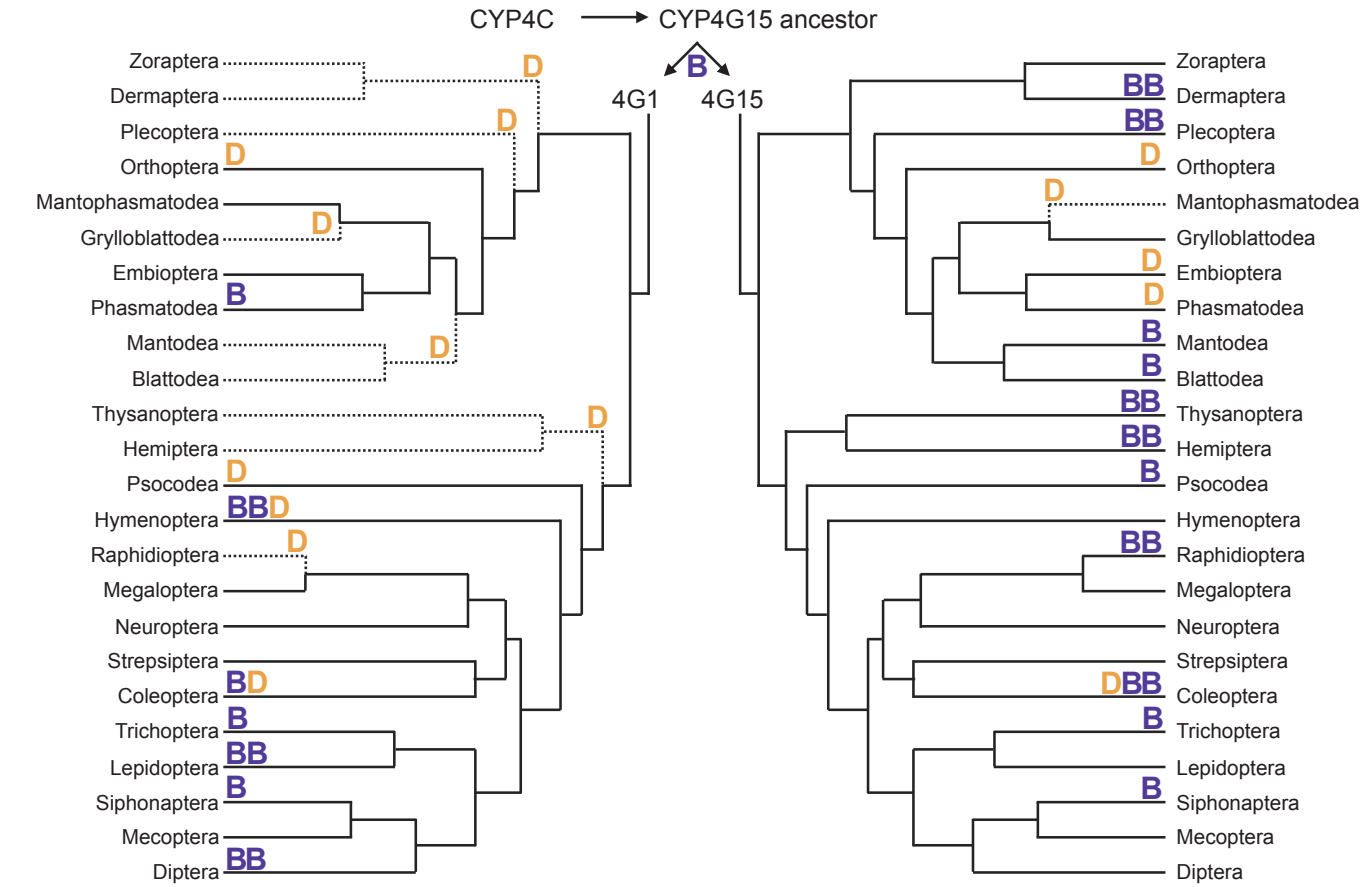

Fig. 6. Phylogeny of neopteran orders and fate of the CYP4G1 and 4G15 paralogs. The insect phylogeny is adapted from Misof et al. (2014). An ancestral CYP4G resulted from the neofunctionalization of a duplicated CYP4C gene. This ancestral CYP4G then duplicated resulting in a family of CYP4G1 and CYP4G15 paralogs found in 24 extant insect orders. D indicates gene loss within the lineage, not excluding independent losses. B indicates gene duplication and BB indicates more than a single duplication within the lineage. Lineages as stippled lines indicate gene loss in the entire lineage.

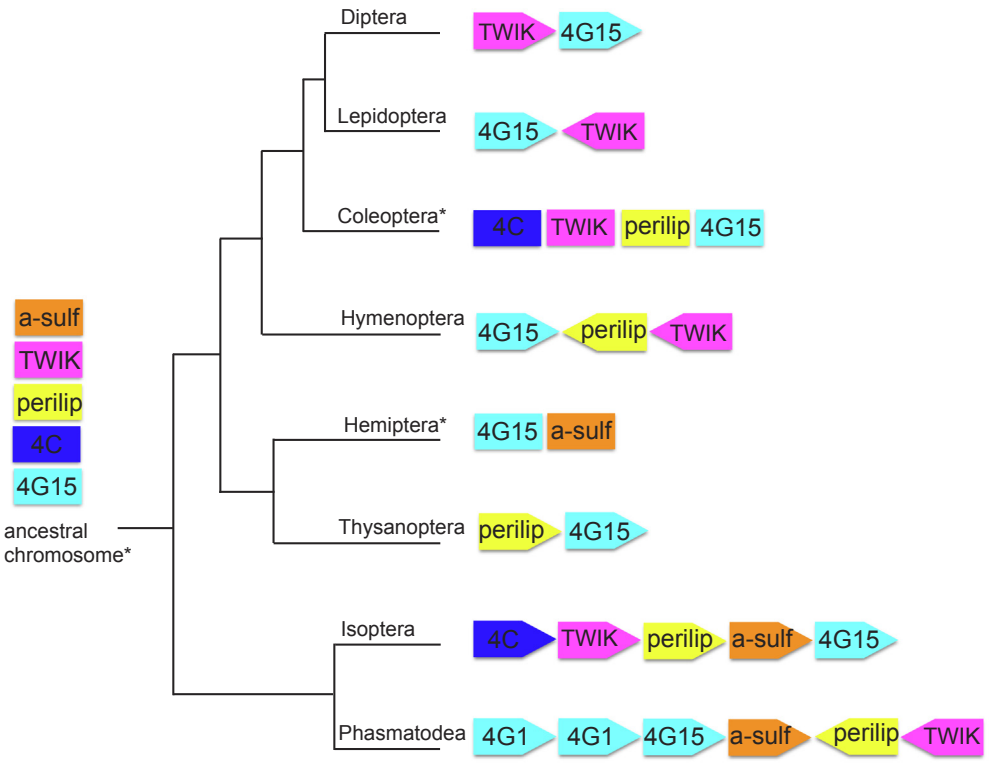

Fig. 7. The CYP4G genesis locus. The tree topology follows the insect phylogeny of Misof et al. (2014). The species representing the various orders are as follows: Phasmatodea: *Timema cristinae*; Isoptera: *Cryptotermes secundus*, *Zootermopsis nevadensis*; Thysanoptera: *Frankliniella occidentalis*; Hemiptera: *Halyomorpha halys*; Hymenoptera: *Acromyrmex echiniator;* Coleoptera: *Aethina tumida*, *Anoplophora glabripennis*, *Leptinotarsa decemlineata*, *Nicrophorus vespilloides*, *Tribolium castaneum*; Lepidoptera: *Amyeolis transitella*, *Plutella xylostella*; Diptera: *Aedes aegypti*, *Ceratitis capitata*. An asterisk * indicates that the genes are listed without direction or order on the chromosome, as several configurations are found in different species of that insect order. The CYP4G genesis locus on the ancestral chromosome is thought to have comprised a CYP4C gene, recently duplicated and diverged into a new CYP4G15 gene. The neighbouring genes included a perilipin (perilip), a potassium channel (TWIK) and an arylsulfatase B type gene (a-sulf). (see text for details).

suggests that the CYP4G1 ancestor is derived from a CYP4G15 ancestor, in a duplication that occurred shortly after the initial CYP4C-CYP4G duplication. Significantly, in *Tribolium*, the CYP4G1 gene (CYP4G7) is very close to the arylsulfatase B and *ase* genes (Fig. 3), whereas the CYP4G15 gene (CYP4G) has remained close to the TWIK and perilipin genes (Fig. 7). The CYP4C subfamily, as probable progenitor of the CYP4G subfamily, is more ancient with representatives found in insects and crustaceans. To understand the relationship between CYP4G and CYP4C, it may be instructive to review the known functions of CYP4C genes. CYP4C7 in the viviparous cockroach *Diploptera punctata* is expressed specifically in the corpora allata. In adult females, it is expressed when vitellogenesis is completed (Sutherland et al., 1998). The enzyme hydroxylates a number of sesquiterpenoids in omega, and this gene is under strict transcriptional control by the ovaries. The role of CYP4C7 is thus to clear the corpora allata from JH precursors when the glands receive a "stop JH production" signal from the ovaries (Sutherland et al., 2000). Closely related CYP4C genes are expressed in the fat body of this cockroach and are of unknown function (Sutherland et al., 1998). The first named member of the CYP4C subfamily in insects, CYP4C1 is expressed in the fat body of another cockroach, *Blaberus discoidalis* (Bradfield et al., 1991) under the control of JH and of the peptidic hypertrehalosemic hormone (HTH), suggesting a regulatory role in energy/lipid mobilization (Lu et al., 1999). CYP4C15 is found selectively in the ecdysteroidogenic Y-organs of crayfish where its function is unknown (Dauphin-Villemant et al., 1999; Aragon et al., 2002). Neofem4 is a CYP4C gene in neotenic termites. Interestingly, it controls production of queen-specific CHC (Weil et al., 2007; Hoffmann et al., 2014). This survey shows that although CYP4C genes appear to have lipid metabolic functions under tight physiological control, too little is currently known about CYP4C functions to shed light on the transition to CYP4G function.

3.13. Structure of CYP4G proteins

Comparison of the alignments of 148 CYP4G1 and 210 CYP4G15 sequences do not reveal consistent and obvious differences (S2, S3). Residues that are conserved are mostly conserved in both types of sequences. 67 and 68 residues are 100% conserved in CYP4G1 and CYP4G15 sequences, respectively. Of those, 51 are 100% conserved in all CYP4G sequences analyzed here. Over a hundred residues are conservative substitutions. Although the insertion between helices G and H in CYP4G15 is longer than in CYP4G1, this is particular for *Drosophila*, and the mean length of the insertion is the same (44 and 43 amino acids respectively) in the two types of CYP4G sequences. The amino acid composition of the insertion showed an enrichment in acidic residues (Asp 15.6%, Glu 9.4%, average from 204 sequences). This is about double the natural abundance of acidic residues in proteins (Carugo, 2008; Hormoz, 2013).

Because the insertion aligns poorly and is of variable length, several CYP4G1 and 4G15 sequences were modeled by I-Tasser. As expected, the crystal structure of the rabbit CYP4B1 (Hsu et al., 2017), a short chain fatty acid and alkane omega-hydroxylase, served as the best template, simply because it is the only CYP4 structure currently public (about 30% identity). Fig. 8 shows a comparison of the CYP4B1 structure with the model for CYP4G16. The insertion was protruding from the globular structure of the P450, on the cytoplasmic side distal from the membrane surface in which the N-terminal is anchored and the loop between helices F and G is dipping. The variable sequences of the CYP4G insertions resulted in varying predictions of secondary structure, from random coil over the entire length, to a short helix between helices G and H, to a lengthened helix G and random coil. It appears then that the acidic nature of the insertion is the only common feature. The insertion does not interfere with interaction site of the P450 with NADPH cytochrome P450 reductase (CPR), the obligatory electron donor that is highly enriched in oenocytes along with CYP4G enzymes (Lycett et al., 2006; Qiu et al., 2012). This interaction site of the FMN domain of CPR with P450s is thought to be located near residues of the N-terminus of the I helix and of the B' helix (Estrada et al., 2016), i.e. in front of the view shown in Fig. 8. However, the CPR is a larger multi-domain enzyme which undergoes conformational changes during catalysis (Laursen et al., 2011), and it is possible that the CYP4G insertion could have ionic interactions that stabilize the P450-CPR complex. Alternatively, the insertion might serve to tether yet unidentified proteins to form a metabolon with CYP4G and CPR. Candidates would be fatty acyl-CoA reductases (FAR), the enzymes providing CYP4G with their substrates.

The consensus from 358 CYP4G sequences was illustrated as "web logo" and portions of this CYP4G consensus were aligned with rabbit CYP4B1 (Hsu et al., 2017). Fig. 9 shows substrate recognition sites (SRS) as well as the region surrounding the Cys axial ligand to the heme. The figure shows that these regions close to the active site of the CYP4G enzymes have been highly conserved for 400 MY. Furthermore, of the 17 amino acid residues lining the active site of CYP4B1 (Hsu et al., 2017) eight are identical and four are conserved substitutions. This suggests that the ancestor CYP4C may have been a fatty acid omega hydroxylase and that there are structural constraints to maintain the regioselectivity towards the omega position. The broader phylogeny of CYP4 genes (Kirischian and Wilson, 2012) indicates that insect CYP4 sequences are monophyletic with the vertebrate CYP4V subfamily, known to encode omega hydroxylases.

In CYP4 omega hydroxylases of vertebrates, a conserved Glu residue of the I helix forms a covalent (ester) bond to a hydroxyl group of the heme 5-methyl group (LeBrun et al., 2002) and this favors omega position. This Glu310 in the I helix of rabbit CYP4B1 is totally conserved in all insect CYP4G sequences, except in one of the four *B. tabaci* CYP4G15 (CYP4G69, Ala) and in one of the three Dinoponera CYP4G1 (Val) with a divergent I helix. However, no evidence for a covalent bond between this Glu and the heme porphyrin could be demonstrated in *Drosophila melanogaster* CYP4G1 by TOF-MS and peptide mapping.

4. General discussion and conclusion

It is commonly held that single copy genes with a conserved function are "stable", being less subject to duplications and losses (Krylov et al., 2003; Waterhouse et al., 2010). This statement has also been made for P450 genes, specifically that biosynthetic-type P450s are less subject to gene duplication and loss than detoxification-type P450s (Thomas, 2007; Feyereisen, 2011; Good et al., 2014; Kawashima and Satta, 2014). From a broad perspective then, the notion of conserved, "stable" genes has some merit. Yet there are examples of "unstable" P450 genes with highly conserved function (Sezutsu et al., 2013). These include for instance CYP307, the first P450 in the ecdysteroid synthesis pathway with multiple instances of birth and death. CYP51, a highly conserved P450 involved in 14-demethylation of membrane sterol precursors, has been duplicated to a series of paralogs in grasses that are involved in the synthesis of defense compounds, and lost in lineages such as nematodes and arthropods.

An initial survey of just a few insect genomes as available a few years ago would have suggested "stability" of the CYP4G genes. Indeed, when compared to e.g. CYP6 genes which are highly variable in number and sequence, the initially sequenced insect genomes revealed just two CYP4G genes, with some exceptions. Bees have just one CYP4G15 gene, but they have overall very few P450 genes, and just four CYP4 family genes (Claudianos et al., 2006). The two CYP4G genes of *Drosophila* were also considered "stable" in an elegant survey of twelve species (Good et al., 2014). Here, a closer look at 167 species from 25 orders gives a different picture. The CYP4G genes of Neoptera are shown to be unstable during their 400 MY of evolution (Fig. 6). This instability is observed even though CYP4G genes play an essential role in the prevention of water loss, a key physiological problem to insects because of their large surface to volume ratio. Loss of the CYP4G genes involved in CHC production is lethal, and this CYP4G function is conserved for the

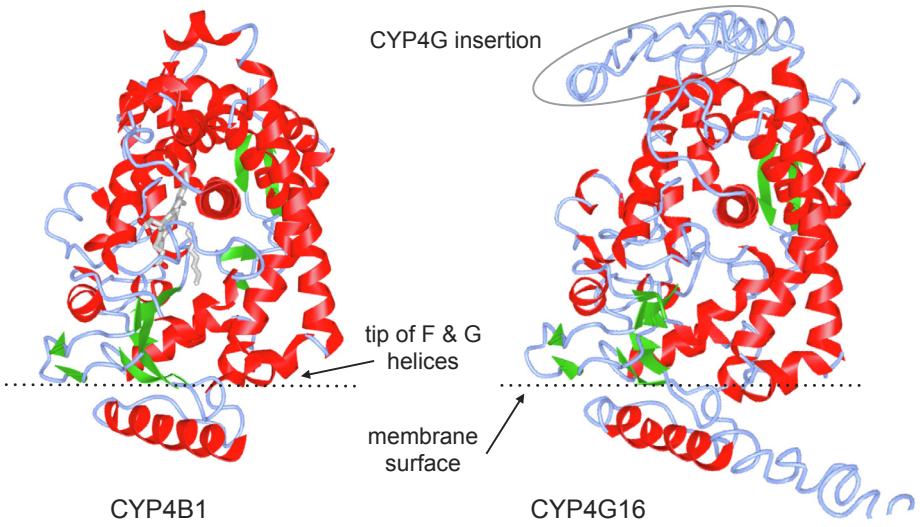

Fig. 8. Comparison of the model of Anopheles CYP4G16 and the structure of rabbit CYP4B1. The CYP4G16 model obtained by I-TASSER on the right panel and CYP4B1 with bound heme and octane substrate on the left panel (PDB 5t6q). Helices are in red and sheets in green. The view is through the I helix, with the N-terminal transmembrane helix and the tips of the F and G helices at the bottom. The CYP4G insertion is clearly visible on top of the model, after the G helix. The approximate position of the membrane surface is shown as stippled line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

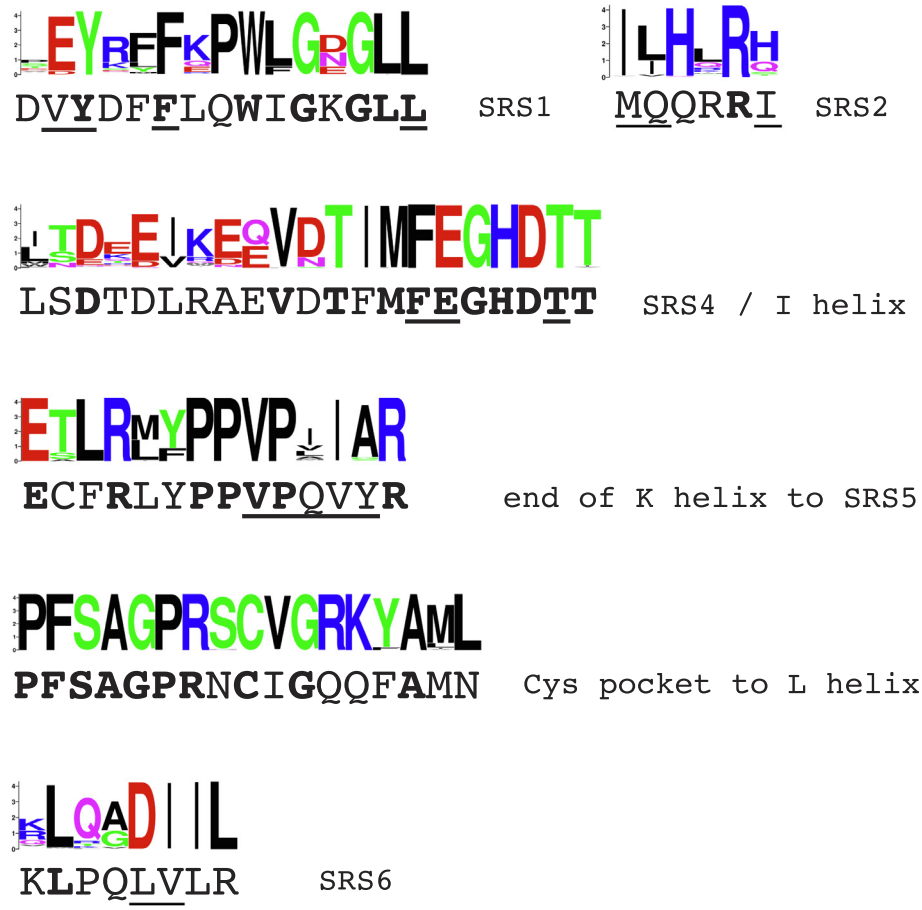

Fig. 9. Structural features of the CYP4G active site. Consensus from 358 CYP4G sequences shown as color WebLogos (Crooks et al., 2004) were aligned on top of the sequence of rabbit CYP4B1 (Hsu et al., 2017). Active site residues in the structure of CYP4B1 are underlined, residues conserved between CYP4G and CYP4B1 are in bold.

single CYP4G genes in aphids (Chen et al., 2016) or bees (Calla et al., 2018), and for the two genes in mosquitoes (Kefi et al., 2019).

Why are there generally two CYP4G genes for over 400 MY, when a single gene would be sufficient for survival? The two genes are not fully redundant as the mosquito genes, which are both expressed in oenocytes, have a different efficacy in replacing the CYP4G1 gene of *Drosophila* in transgenic flies (Kefi et al., 2019), and indeed appear to have a slightly different substrate specificity. The two CYP4G enzymes of the mountain pine beetle may have slightly different velocities (MacLean et al., 2018). The two CYP4G genes of the brown planthopper both contribute to CHC synthesis and desiccation resistance, yet slightly differ in their effect on insecticide penetration (Wang et al., 2019b). In *Drosophila* itself, CYP4G1 plays a key role in CHC production, while the CYP4G15 gene is not expressed in oenocytes but in the brain (Maibeche-Coisne et al., 2000), where its function is unknown and CYP4G15-RNAi flies are viable (Chung et al., 2009). Expression of CYP4G genes outside the oenocytes, in tissues not known to produce hydrocarbons such as prothoracic glands in *M. sexta* (AY635178.1), *B. mori* and *D. melanogaster* (Niwa et al., 2011) and a variety of other tissues (Maibeche-Coisne et al., 2005; Pondeville et al., 2013) is therefore enigmatic. Much remain to be discovered about the function of CYP4G enzymes. Loss of CYP4G1 in *Drosophila* is lethal (Gutierrez

et al., 2007) because ectopic expression of CYP4G15 (i.e. outside the oenocytes) leaves CYP4G1 as the only CHC producer, a situation analogous to bees (see 3.8) or aphids (see 3.10). Lethality of the *Drosophila* CYP4G1 gene does therefore not mean that it *gained* a new essential function in the *Drosophila* lineage. In fact, in those species with two CYP4G genes, both can be considered ecologically "lethal" genes, as indeed RNAi of either one or of both leads to severely compromised insects that would not be expected to survive in the wild (Wang et al., 2019a,b).

In a comparison of human, chimpanzee and macaque genomes, Chen et al. (2010) noted that duplicate genes with family size conservation evolved more slowly than those without family size conservation, an observation which seem to apply to the CYP4G genes, where blooming behavior, as seen in other CYP families (Feyereisen, 2011; Sezutsu et al, 2013), is rare. Blooming of the CYP4G genes as in the brown marmorated stink bug *H. halys* leads to the generation of new and divergent P450s of unknown function. To understand the reasons for the maintenance of an old (about 400MY) duplication, and the subsequent evolutionary "instability" of the duplicate pair (Fig. 6), a closer look at CYP4G gene phenotypes is in order. Oxidative decarbonylation of long chain fatty aldehydes (Qiu et al., 2012) or alcohols (MacLean et al., 2018) is the only biochemical reaction known to be catalyzed by CYP4G enzymes. However, their hydrocarbon products serve many functions in insect physiology, from desiccation resistance to chemical communication in a myriad ways (Howard and Blomquist, 2005). The hydrocarbons produced, several dozen, differ in their chain length, degree and position of desaturation and of (methyl) branching. Many fatty acid synthases (FAS), elongases, desaturases and fatty acyl-CoA reductases (FAR) contribute to generate the blend of substrates for the "terminal" CYP4G enzymes. In addition, largely unknown processes contribute to the deposition of the CYP4G enzyme products on the epicuticle. It is unlikely, indeed quite impossible that the kinetic properties of CYP4G enzymes (k_{cat}, K_M) would be exactly the same for all their substrates so that they, too, contribute to hydrocarbon diversity. Experimental evidence for this proposition was obtained by Kefi et al. (2019) who modified *Drosophila* CHC profiles by transgenic expression of two mosquito CYP4G genes. Furthermore, current biochemical evidence indicates that the CYP4G enzymes are rather sluggish, perhaps as a result of their complex and unusual catalytic activity. This would explain the high concentration of CYP4G enzymes in CHC synthesizing cells (Qiu et al., 2012), i.e. high [enzyme] would compensate for low k_{cat}. This, in turn, implies that the level of expression of CYP4G genes also affects CHC diversity. Perhaps the initial CYP4G15/CYP4G1 duplication was favored by its effect on gene dosage. Moreover, the phenotypic plasticity of CHC in insects, both quantitative and qualitative, both stepwise and gradual (Otte et al., 2018), would be facilitated by the presence of two, ever so slightly different CYP4G genes. This is a difficult hypothesis to test, but is indirectly validated by the elegant work of Chung et al. (2014) who showed that the lack of expression of just one of the FAS genes in *Drosophila birchii* is sufficient to decrease the amounts of methyl branched CHC and desiccation resistance, and may be responsible for reproductive isolation from its sister species *D. serrata*. Such a dual trait (desiccation resistance and mate choice by chemical communication), amounts to a pleiotropic function of a gene with a single biochemical function (FAS). The same should hold for CYP4G genes, and duplication can be a way to mitigate the risk of inactivation of a single gene with pleiotropic functions. The "instability" of CYP4G genes may therefore be related to the ever changing needs of CHC complexity that accompany speciation events. The fatty acyl-CoA reductase (FAR) genes provide a startling clear confirmation of this idea. In a survey of 200 FAR genes in twelve *Drosophila* species, it was shown that "stable" FAR genes were essentially involved in intermediary lipid metabolism, whereas the "unstable" FAR genes were chiefly expressed in oenocytes and participating in determining the CHC profile of flies, and, by implication, also playing a role in speciation (Finet et al., 2019). Pre-mating reproductive isolation between the recently diverged but morphologically indistinguishable malaria vectors *Anopheles gambiae* and *A. coluzzii* is associated with a "genomic island of speciation" on the X chromosome (Aboagye-Antwi et al., 2015), of which the most striking molecular marker is a variant of the CYP4G16 gene (Caputo et al., 2016). It is tempting to suggest that slight variations in CHC may be the close range mating recognition cues differentiating the two species. In conclusion, maintaining the essential desiccation resistance role of CYP4G enzymes while allowing variation of their chemical communication role seems to be achieved by a "revolving door" (Demuth and Hahn, 2009) of CYP4G gene birth and death over 400 MY.

Acknowledgement

The comments and advice of Dr. W. Dermauw and valuable information provided by an anonymous reviewer are gratefully acknowledged.

Conflicts of Interest

None

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2019.106695.

References

Aboagye-Antwi, F., Alhafez, N., Weedall, G.D., Brothwood, J., Kandola, S., et al., 2015. Experimental swap of *Anopheles gambiae*'s assortative mating preferences demonstrates key role of X-chromosome divergence island in incipient sympatric speciation. PLoS Genet. 11, e1005141.

Amaral, D.T., Silva, J.R., Viviani, V.R., 2017a. Transcriptomes from the photogenic and non-photogenetic tissues and life stages of the *Aspisoma lineatum* firefly (Coleoptera, Lampyridae) Implications for the evolutionary origins of bioluminescence and its associated light organs. Gene Reports 8, 150–159.

Amaral, D.T., Silva, J.R., Viviani, V.R., 2017b. Transcriptional comparison of the photogenic and non-photogenic tissues of *Phrixothrix hirtus* (Coleoptera: Phengodidae) and non-luminescent *Chauliognathus flavipes* (Coleoptera: Cantharidae) give insights on the origin of lanterns in railroad worms. Gene Reports 78–86.

Aragon, S., Claudinot, S., Blais, C., Maibeche, M., Dauphin-Villemant, C., 2002. Molting cycle-dependent expression of CYP4C15, a cytochrome P450 enzyme putatively involved in ecdysteroidogenesis in the crayfish, *Orconectes limosus*. Insect Biochem. Mol. Biol. 32, 153–159.

Balabanidou, V., Grigoraki, L., Vontas, J., 2018. Insect cuticle: a critical determinant of insecticide resistance. Curr. Opin. Insect Sci. 27, 68–74.

Balabanidou, V., Kampouraki, A., MacLean, M., Blomquist, G.J., Tittiger, C., et al., 2016. Cytochrome P450 associated with insecticide resistance catalyzes cuticular hydrocarbon production in *Anopheles gambiae*. Proc. Natl. Acad. Sci. USA 113, 9268–9273.

Balabanidou, V., Kefi, M., Aivaliotis, M., Koidou, V., Girotti, J.R., et al., 2019. Mosquitoes cloak their legs to resist insecticides. Proc. Roy. Soc. B: Biol. Sci. 286, 20191091.

Bossert, S., Murray, E.A., Almeida, E.A.B., Brady, S.G., Blaimer, B.B., et al., 2019. Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. Mol. Phylogenet. Evol. 130, 121–131.

Bradfield, J., Lee, Y., Keeley, L., 1991. Cytochrome P450 family 4 in a cockroach: molecular cloning and regulation by regulation by hypertrehalosemic hormone. Proc. Natl. Acad. Sci. U S A 88, 4558–4562.

Buck, J.B., 1948. The anatomy and physiology of the light organ in fireflies. Ann. N. Y. Acad. Sci. 49, 397–483.

Calla, B., MacLean, M., Liao, L.H., Dhanjal, I., Tittiger, C., Blomquist, G.J., Berenbaum, M.R., 2018. Functional characterization of CYP4G11-a highly conserved enzyme in the western honey bee *Apis mellifera*. Insect Mol. Biol. 27, 661–674.

Caputo, B., Pichler, V., Mancini, E., Pombi, M., Vicente, J.L., et al., 2016. The last bastion? X chromosome genotyping of *Anopheles gambiae* species pair males from a hybrid zone reveals complex recombination within the major candidate 'genomic island of speciation'. Mol. Ecol. 25, 5719–5731.

Carugo, O., 2008. Amino acid composition and protein dimension. Protein Sci. 17, 2187–2191.

Chen, F.-C., Chen, C.-J., Li, W.-H., Chuang, T.-J., 2010. Gene family size conservation is a good indicator of evolutionary rates. Mol. Biol. Evol. 27, 1750–1758.

Chen, N., Fan, Y.-L., Bai, Y., Li, X.-d., Zhang, Z.-F., et al., 2016. Cytochrome P450 gene, CYP4G51, modulates hydrocarbon production in the pea aphid, *Acyrthosiphon pisum*. Insect Biochem. Mol. Biol. 76, 84–94.

Chen, N., Pei, X.-J., Li, S., Fan, Y.-L., Liu, T.-X., 2019. Involvement of integument-rich CYP4G19 in hydrocarbon biosynthesis and cuticular penetration resistance in *Blattella germanica* (L.). Pest Manage. Sci. https://doi.org/10.1002/ps.5499.

Chung, H., Sztal, T., Pasricha, S., Sridhar, M., Batterham, P., Daborn, P.J., 2009. Characterization of *Drosophila melanogaster* cytochrome P450 genes. Proc. Natl. Acad. Sci. USA 106, 5731–5736.
Chung, H., Loehlin, D.W., Dufour, H.D., Vaccarro, K., Millar, J.G., et al., 2014. A single gene affects both ecological divergence and mate choice in *Drosophila*. Science 343, 1148–1151.
Claudianos, C., Ranson, H., Johnson, R.M., Biswas, S., Schuler, M.A., et al., 2006. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. Insect Mol. Biol. 15, 615–636.
Consortium, I., 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol. 8, e1000313.
Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: A sequence logo generator. Genome Res. 14, 1188–1190.
Daborn, P., Yen, J., Bogwitz, M., Le Goff, G., Feil, E., et al., 2002. A single P450 allele associated with insecticide resistance in *Drosophila*. Science 297, 2253–2256.
Dauphin-Villemant, C., Bocking, D., Tom, M., Maibeche, M., Lafont, R., 1999. Cloning of a novel cytochrome P450 (CYP4C15) differentially expressed in the steroidogenic glands of an arthropod. Biochem. Biophys. Res. Commun. 264, 413–418.
Demuth, J., Hahn, M., 2009. The life and death of gene families. BioEssays 31, 29–39.
Estrada, D.F., Laurence, J.S., Scott, E.E., 2016. Cytochrome P450 17A1 Interactions with the FMN Domain of Its Reductase as Characterized by NMR. J. Biol. Chem. 291, 3990–4003.
Etges, W.J., Jackson, L.L., 2001. Epicuticular hydrocarbon variation in *Drosophila mojavensis* cluster species. J. Chem. Ecol. 27, 2125–2149.
Fallon, T.R., Lower, S.E., Chang, C.-H., Bessho-Uehara, M., Martin, G.J., et al., 2018. Firefly genomes illuminate parallel origins of bioluminescence in beetles. Elife 7. https://doi.org/10.7554/eLife. 36495.
Ferveur, J., 2005. Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. Behav. Genet. 35, 279–295.
Feyereisen, R., 2011. Arthropod CYPomes illustrate the tempo and mode in P450 evolution. BBA 1814, 19–28.
Feyereisen, R., 2012. Insect CYP genes and P450 enzymes. In: Gilbert, L.I. (Ed.), Insect Molecular Biology and Biochemistry. Elsevier B.V, London, pp. 236–316.
Feyereisen, 2015. Insect P450 inhibitors and insecticides: challenges and opportunities. Pest Manage. Sci. 71, 793–800.
Finet, C., Slavik, K., Pu, J., Carroll, S.B., Chung, H., 2019. Birth-and-Death Evolution of the Fatty Acyl-CoA Reductase (FAR) Gene Family and Diversification of Cuticular Hydrocarbon Synthesis in *Drosophila*. Genome Biol. Evolut. 11, 1541–1551.
Futahashi, R., Yamahama, Y., Kawaguchi, M., Mori, N., Ishii, D., et al., 2019. Molecular basis of wax-based color change and UV reflection in dragonflies. Elife 8. https://doi.org/10.7554/eLife. 43045.
Gabaldón, T., Koonin, E.V., 2013. Functional and evolutionary implications of gene orthology. Nat. Rev. Genet. 14, 360–366.
Good, R.T., Gramzow, L., Battlay, P., Sztal, T., Batterham, P., et al., 2014. The molecular evolution of cytochrome P450 genes within and between *Drosophila* species. Genome Biol. Evolut. 6, 1118–1134.
Guengerich, F.P., Cheng, Q., 2011. Orphans in the human cytochrome P450 superfamily: approaches to discovering functions and relevance in pharmacology. Pharmacol. Rev. 63, 684–699.
Guengerich, F.P., Yoshimoto, F.K., 2018. Formation and cleavage of C-C bonds by enzymatic oxidation-reduction reactions. Chem. Rev. 118, 6573–6655.
Guo, G.-Z., Geng, Y.-J., Huang, D.-N., Xue, C.-F., Zhang, R.-L., 2010. Level of CYP4G19 expression is associated with pyrethroid resistance in *Blattella germanica*. J. Parasitol. Res. https://doi.org/10.1155/2010/517534.
Gusev, O., Suetsugu, Y., Cornette, R., Kawashima, T., Logacheva, M.D., et al., 2014. Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. Nat. Commun. 5, 4784.
Gutierrez, E., Wiggins, D., Fielding, B., Gould, A.P., 2007. Specialized hepatocyte-like cells regulate *Drosophila* lipid metabolism. Nature 445, 275–280.
Harrison, M.C., Jongepier, E., Robertson, H.M., Arning, N., Bitard-Feildel, T., et al., 2018. Hemimetabolous genomes reveal molecular basis of termite eusociality. Nat. Ecol. Evol. 2, 557–566.
Hatje, K., Keller, O., Hammesfahr, B., Pillmann, H., Waack, S., Kollmar, M., 2011. Cross-species protein sequence and gene structure prediction with fine-tuned Webscipio 2.0 and Scipio. BMC Research Notes 4: 265.
Hoffmann, K., Gowin, J., Hartfelder, K., Korb, J., 2014. The scent of royalty: a P450 gene signals reproductive status in a social insect. Mol. Biol. Evol. 31, 2689–2696.
Hormoz, S., 2013. Amino acid composition of proteins reduces deleterious impact of mutations. Sci. Rep. 3, 2919.
Howard, R.W., Blomquist, G.J., 2005. Ecological, behavioral, and biochemical aspects of insect hydrocarbons. Ann. Rev. Entomol. 50, 371–393.
Hsu, M.-H., Baer, B.R., Rettie, A.E., Johnson, E.F., 2017. The Crystal Structure of Cytochrome P450 4B1 (CYP4B1) monooxygenase complexed with octane discloses several structural adaptations for ω-Hydroxylation. J. Biol. Chem. 293, 5610–5621.
Jacob, J., Hanssen, H.P., 1979. The chemical composition of cuticular lipids from dragonflies (Odonata). Z. Naturforsch. 34c, 498–502.
Johnson, K.P., Dietrich, C.H., Friedrich, F., Beutel, R.G., Wipfler, B., et al., 2018. Phylogenomics and the evolution of hemipteroid insects. Proc. Natl. Acad. Sci. USA 115, 12775–12780.
Jones, C.M., Haji, K.A., Khatib, B.O., Bagi, J., Mcha, J., et al., 2013. The dynamics of pyrethroid resistance in *Anopheles arabiensis* from Zanzibar and an assessment of the underlying genetic basis. Parasit Vectors. 6, 343.
Kalsi, M., Palli, S.R., 2017. Cap n' collar transcription factor regulates multiple genes coding for proteins involved in insecticide detoxification in the red flour beetle, *Tribolium castaneum*. Insect Biochem. Mol. Biol. 90, 43–52.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kawashima, A., Satta, Y., 2014. Substrate-dependent evolution of cytochrome P450: rapid turnover of the detoxification-type and conservation of the biosynthesis-type. PLoS ONE 9, e100059.
Kefi, M., Balabanidou, V., Douris, V., Lycett, G., Feyereisen, R., et al., 2019. Two functionally distinct CYP4G genes of *Anopheles gambiae* contribute to cuticular hydrocarbon biosynthesis. Insect Biochem. Mol. Biol. 110, 52–59.
Kim, J.H., Moreau, J.A., Zina, J.M., Mazgaeen, L., Yoon, K.S., et al., 2018. Identification and interaction of multiple genes resulting in DDT resistance in the 91-R strain of *Drosophila melanogaster* by RNAi approaches. Pestic. Biochem. Physiol. 151, 90–99.
Kirischian, N.L., Wilson, J.Y., 2012. Phylogenetic and functional analyses of the cytochrome P450 family 4. Mol. Phylogenet. Evol. 62, 458–471.
Koto, A., Motoyama, N., Tahara, H., McGregor, S., Moriyama, M., et al., 2019. Oxytocin/vasopressin-like peptide inotocin regulates cuticular hydrocarbon synthesis and water balancing in ants. Proc. Natl. Acad. Sci. USA 116, 5597–5606.
Krylov, D., Wolf, Y., Rogozin, I., Koonin, E., 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res. 13, 2229–2235.
Laursen, T., Jensen, K., Møller, B.L., 2011. Conformational changes of the NADPH-dependent cytochrome P450 reductase in the course of electron transfer to cytochromes P450. Biochim. Biophys. Acta 1814, 132–138.
LeBrun, L., Hoch, U., Ortiz de Montellano, P., 2002. Autocatalytic mechanism and consequences of covalent heme attachment in the cytochrome P4504A family. J. Biol. Chem. 277, 12755–12761.
Lomas, K.F., Greenwood, D.R., Windmill, J.F.C., Jackson, J.C., Corfield, J., et al., 2012. Discovery of a lipid synthesising organ in the auditory system of an insect. PLoS ONE 7, e51486.
Lu, K.H., Bradfield, J.Y., Keeley, L.L., 1999. Juvenile hormone inhibition of gene expression for cytochrome P4504C1 in adult females of the cockroach, *Blaberus discoidalis*. Insect Biochem. Mol. Biol. 29, 667–673.
Lycett, G.J., McLaughlin, L.A., Ranson, H., Hemingway, J., Kafatos, F.C., et al., 2006. *Anopheles gambiae* P450 reductase is highly expressed in oenocytes and in vivo knockdown increases permethrin susceptibility. Insect Mol. Biol. 15, 321–327.
MacLean, M., Nadeau, J., Gurnea, T., Tittiger, C., Blomquist, G.J., 2018. Mountain pine beetle (*Dendroctonus ponderosae*) CYP4Gs convert long and short chain alcohols and aldehydes to hydrocarbons. Insect Biochem. Mol. Biol. 102, 11–20.
Maibeche-Coisne, M., Merlin, C., Francois, M., Porcheron, P., Jacquin-Joly, E., 2005. P450 and P450 reductase cDNAs from the moth *Mamestra brassicae*: cloning and expression patterns in male antennae. Gene 346, 195–203.
Maibeche-Coisne, M., Monti-Dedieu, L., Aragon, S., Dauphin-Villemant, C., 2000. A new cytochrome P450 from *Drosophila melanogaster*, CYP4G15, expressed in the nervous system. Biochem. Biophys. Res. Commun. 273, 1132–1137.
McKenna, D.D., Wild, A.L., Kanda, K., Bellamy, C.L., Beutel, R.G., et al., 2015. The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the Cretaceous terrestrial revolution. Syst. Entomol. 40, 835–880.
Mikó, I., 2019. Fat in the Leg: Function of the Expanded Hind Leg in Gasteruptiid Wasps (Hymenoptera: Gasteruptiidae). Insect System. Div. 3, 1–16.
Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., et al., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767.
Moussian, B., 2010. Recent advances in understanding mechanisms of insect cuticle differentiation. Insect Biochem. Mol. Biol. 40, 363–375.
Müller, P., Chouaibou, M., Pignatelli, P., Etang, J., Walker, E., et al., 2008. Pyrethroid tolerance is associated with elevated expression of antioxidants and agricultural practice in *Anopheles arabiensis* sampled from an area of cotton fields in Northern Cameroon. Mol. Ecol. 17, 1145–1155.
Nelson, D.R., 2018. Cytochrome P450 diversity in the tree of life. BBA - Proteins Proteomics 1866, 141–154.
Nelson, D.R., Goldstone, J.V., Stegeman, J.J., 2013. The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. Philos. Trans. Roy. Soc. B: Biol. Sci. 368, 20120474.
Niwa, R., Sakudoh, T., Matsuya, T., Namiki, T., Kasai, S., Tomita, T., Kataoka, H., 2011. Expressions of the cytochrome P450 monooxygenase gene CYP4G1 and its homolog. Appl. Entomol. Zool. 46, 533–543.
Ortiz de Montellano, P.R., editor, 2005. Cytochrome P450, Structure, Mechanism, and Biochemistry. 689 p.
Otte, T., Hilker, M., Geiselhardt, S., 2018. Phenotypic plasticity of cuticular hydrocarbon profiles in insects. J. Chem. Ecol. 44, 235–247.
Pan, D., Zhang, L., 2009. Burst of young retrogenes and independent retrogene formation in mammals. PLoS ONE 4, e5040.
Pask, G.M., Slone, J.D., Millar, J.G., Das, P., Moreira, J.A., et al., 2017. Specialized odorant receptors in social insects that detect cuticular hydrocarbon cues and candidate pheromones. Nat. Commun. 8, 297.
Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., et al., 2017. Evolutionary history of the hymenoptera. Curr. Biol. 27, 1013–1018.
Pittendrigh, B., Aronstein, K., Zinkovsky, E., Andreev, O., Campbell, B., et al., 1997. Cytochrome P450 genes from *Helicoverpa armigera*: expression in a pyrethroid-susceptible and -resistant strain. Insect Biochem. Mol. Biol. 27, 507–512.
Pondeville, E., David, J.-P., Guittard, E., Maria, A., Jacques, J.-C., et al., 2013. Microarray and RNAi analysis of P450s in *Anopheles gambiae* male and female steroidogenic tissues: CYP307A1 Is required for ecdysteroid synthesis. PLoS ONE 8, e79861.
Pridgeon, J.W., Zhang, L., Liu, N., 2003. Overexpression of CYP4G19 associated with a pyrethroid-resistant strain of the German cockroach, *Blattella germanica* (L.). Gene 314, 157–163.
Qiu, Y., Tittiger, C., Wicker-Thomas, C., Le Goff, G., Young, S., et al., 2012. An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. Proc. Natl. Acad. Sci. USA 109, 14858–14863.

Sann, M., Niehuis, O., Peters, R.S., Mayer, C., Kozlov, A., et al., 2018. Phylogenomic analysis of Apoidea sheds new light on the sister group of bees. BMC Evol. Biol. 18, 71.
Sezutsu, H., Le Goff, G., Feyereisen, R., 2013. Origins of P450 diversity. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368, 20120428.
Shimizu, N., Sakata, D., Schmelz, E.A., Mori, N., Kuwahara, Y., 2017. Biosynthetic pathway of aliphatic formates via a Baeyer-Villiger oxidation in mechanism present in astigmatid mites. Proc. Natl. Acad. Sci. USA 114, 2616–2621.
Smith, D.S., 1963. The organization and innervation of the luminescent organ in a firefly, *Photuris pennsylvanica* (Coleoptera). J. Cell Biol. 16, 323–359.
Stamatakis, A., 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Sutherland, T., Unnithan, G., Andersen, J., Evans, P., Murataliev, M., et al., 1998. A cytochrome P450 terpenoid hydroxylase linked to the suppression of insect juvenile hormone synthesis. Proc. Natl. Acad. Sci. USA 95, 12884–12889.
Sutherland, T., Unnithan, G., Feyereisen, R., 2000. Terpenoid omega-hydroxylase (CYP4C7) messenger RNA levels in the corpora allata: a marker for ovarian control of juvenile hormone synthesis in *Diploptera punctata*. J. Insect Physiol. 46, 1219–1227.
Thomas, J., 2007. Rapid birth-death evolution specific to xenobiotic cytochrome P450 genes in vertebrates. PLoS Genet. 3, e67.
Tijet, N., Helvig, C., Feyereisen, R., 2001. The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny. Gene 262, 189–198.
Tsai, Y.-L., Li, C.-W., Hong, T.-M., Ho, J.-Z., Yang, E.-C., et al., 2014. Firefly light flashing: oxygen supply mechanism. Phys. Rev. Lett. 113, 258103.
Viviani, V.R., 2002. The origin, diversity, and structure function relationships of insect luciferases. Cell. Mol. Life Sci. 59, 1833–1850.
Wang, J., Youkharibache, P., Zhang, D., Lanczycki, C.J., Geer, R.C., Madej, T., Phan, L., Ward, M., Lu, S., Marchler, G.H., Wang, Y., Bryant, S.H., Geer, L.Y., Marchler-Bauer, A., 2019c. iCn3D, a Web-based 3D Viewer for Sharing 1D/2D/3D Representations of Biomolecular Structures. Bioinformatics. https://doi.org/10.1093/bioinformatics/btz502.
Wang, S.Y., Hackney Price, J., Zhang, D., 2019a. Hydrocarbons catalysed by TmCYP4G122 and TmCYP4G123 in *Tenebrio molitor* modulate the olfactory response of the parasitoid *Scleroderma guani*. Insect Mol. Biol. https://doi.org/10.3389/fphys.2019.00913.
Wang, S., Li, B., Zhang, D., 2019b. NlCYP4G76 and NlCYP4G115 Modulate Susceptibility to Desiccation and Insecticide Penetration Through Affecting Cuticular Hydrocarbon Biosynthesis in *Nilaparvata lugens* (Hemiptera: Delphacidae). Front. Physiol. 10, 913.
Waterhouse, R., Zdobnov, E., Kriventseva, E., 2010. Correlating traits of gene retention, sequence divergence, duplicability and essentiality in vertebrates, arthropods, and fungi. Genome Biol. Evol. 3, 75–86.
Weil, T., Rehli, M., Korb, J., 2007. Molecular basis for the reproductive division of labour in a lower termite. BMC Genom. 8, 198.
Wiegmann, B.M., Trautwein, M.D., Winkler, I.S., Barr, N.B., Kim, J.-W., et al., 2011. Episodic radiations in the fly tree of life. Proc. Natl. Acad. Sci. USA 108, 5690–5695.
Wipfler, B., Letsch, H., Frandsen, P.B., Kapli, P., Mayer, C., et al., 2019. Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects. Proc. Natl. Acad. Sci. USA 116, 3024–3029.
Yang, P., Tanaka, H., Kuwano, E., Suzuki, K., 2008. A novel cytochrome P450 gene (CYP4G25) of the silkmoth *Antheraea yamamai*: cloning and expression pattern in pharate first instar larvae in relation to diapause. J. Insect Physiol. 54, 636–643.
Yang, J., Yan, R., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: Protein structure and function prediction. Nature Methods 12, 7–8.
Yu, Z., Zhang, X., Wang, Y., Moussian, B., Zhu, K.Y., et al., 2016. LmCYP4G102: An oenocyte-specific cytochrome P450 gene required for cuticular waterproofing in the migratory locust, *Locusta migratoria*. Sci. Rep. 6, 29980.
Zhang, S.-Q., Che, L.-H., Li, Y., Liang, D., Pang, H., et al., 2018. Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. Nat. Commun. 9, 205.
Zhou, X., Qian, K., Tong, Y., Zhu, J.J., Qiu, X., et al., 2014. De novo transcriptome of the Hemimetabolous German cockroach (*Blattella germanica*). PLoS ONE 9, e106932.