Aliens in the CYPome of the black fungus gnat, *Bradysia coprophila*

René Feyereisen[a], John M. Urban[b] and David R. Nelson[c]

[a]*Laboratory of Agrozoology, Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000, Ghent, Belgium*

[b]*Carnegie Institution for Science, Department of Embryology, Howard Hughes Medical 8 Institute Research Laboratories, 3520 San Martin Drive, Baltimore, MD, 21218, USA*

[c]*Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, Memphis, TN 38163, USA*

Abstract

The diverse cytochrome P450 enzymes of insects play essential physiological roles and also play important roles in the metabolism of environmental chemicals such as insecticides. We manually curated the complement of P450 (CYP) genes, or CYPome, of the black fungus gnat, *Bradysia (Sciara) coprophila* (Diptera, Sciaroidea), a species with a variable number of chromosomes. This CYPome carries two types of "alien" P450 genes. The first type of alien P450s was found among the 163 CYP genes of the core genome (autosomes and X). They consist of 28 sequences resulting from horizontal gene transfer, with closest sequences not found in insects, but in other arthropods, often Collembola. These genes are not contaminants, because they are expressed genes with introns, found in synteny with regular dipteran genes, also found in *B. odoriphaga* and *B. hygida*. Two such "alien" genes are representatives of CYP clans not otherwise found in insects, a CYP53 sequence related to fungal CYP53 genes, and a CYP19-like sequence similar to some collembolan sequences but of unclear origin. The second type of alien P450s are represented by 99 sequences from germline-restricted chromosomes (GRC). While most are P450 pseudogenes, 33 are apparently intact, with half being more closely related to P450s from Cecidomyiidae than from Sciaridae, thus supporting the hypothesis of a cross-family hybridization origin of the GRC.

Keywords: Cytochrome P450, Horizontal gene transfer, dispensable chromosomes, pseudogenes

1. Introduction

With over 160,000 described species, the order Diptera is one of the most speciose groups of insects. The model organism *Drosophila melanogaster* and the malaria vector *Anopheles gambiae* were the first two insects for which a complete genome was published, and for which the full complement of cytochrome P450 genes, or CYPome, was obtained (Tijet et al., 2001, Ranson et al., 2002). With 87 and 100 genes respectively distributed in four clans (the CYP2, 3 , 4 and mitochondrial clans, Feyereisen, 2006), these two CYPomes have long served as references for the diversity of P450 sequences in insects. However, great variation in CYPome size and the occasional presence of P450s from other clans is now documented in arthropods (Dermauw et al., 2020). The Arthropoda assembly assessment catalogue (https://evofun.dscr.unil.ch/table.html), version v.2023-05-01 lists 4,149 genome assembly entries, with 886 Diptera, of which 430 Drosophila and 112 Culicidae. With 170,192 estimated species of Diptera (www.catalogueoflife.org: accessed: 2023-05-05), and 1,620 species of Drosophila or 3,680 Culicidae, it would seem that, as of this writing, genome assemblies are not evenly distributed across the phylogeny of Diptera, or indeed arthropods. As noted before (Feron and Waterhouse, 2022) such a survey "highlights the sparsity and taxonomic imbalance of current species sampling". Here, we take advantage of the recently published genome assemblies of the black fungus gnat *Bradysia (Sciara) coprophila* (Urban et al., 2021, Hodson et al., 2022) to analyze the CYPome of this representative of the Bibionomorpha, an infraorder of "lower" Diptera (Nematocera) rich in species and lodged between the Culicomorpha and the Schizophora (Wiegmann et al., 2011)(Fig.1). *B. coprophila* is of great interest for its ecology and chromosomal biology. As its name suggests, the larvae are soil-dwelling, feeding on decaying plant material and fungi as well as on plant roots, bulbs and seedlings. They are considered pests in crops such as garlic, onions and particularly Chinese chives. Adults have a short life span and do not feed. Insecticide treatments to control fungus gnats have led to P450-mediated resistance in *Bradysia* species (Chen et al., 2017, 2019, 2022; Zhang et al., 2022, 2023), thus justifying the annotation of their CYPome.

Quite unlike other Diptera as Drosophila, or insects in general, *B. coprophila* changes chromosome numbers during development. The initial germline genome consists of three pairs of autosomes, a pair of sex chromosomes (XX or XX'), and dispensable L chromosomes. The latter germline-restricted chromosomes (or GRC) are not found in the somatic genome, as they are eliminated from the soma in early embryogenesis. Furthemore, sex determination proceeds with males being X/0 and females X/X (having only sons) or X/X' (having only daughters). The complex developmental variation in genome DNA constancy (Urban et al., 2021; Hodson et al., 2022) meant a great effort was devoted to assembling a

high contiguity genome (Urban et al., 2021) with sequence anchored to chromosomes (Urban et al. 2022). Furthermore, contigs of the GRC genome were obtained independently (Hodson et al., 2022). Here we report the annotation and analysis of the CYPome of this species which provided several unexpected features and contrasts markedly from the mosquito or Drosophila CYPomes.

2. Methods

The black fungus gnat P450 sequences were annotated from genome assembly GCA_014529535.1, VSDI01, that was obtained from male embryos and male pupae thus covering the autosomes and X chromosome, but not X' chromosome, or GRCs. The RefSeq gene models (annotation release 100) were manually curated against the transcriptome shotgun assemblies (TSA) GJHU01 and GJHV01, and against a collection of 3,000 arthropod P450 sequences (Dermauw et al., 2020) to obtain full length P450 sequences. The depth of the transcriptome contributed to a high quality RefSeq set, with only 10 sequences requiring correction and four fused gene models split into their individual P450 sequences. One gene, confirmed by TSA, and a full-length pseudogene had no gene model in RefSeq. The P450 sequences from the germline-restricted chromosomes (GRC or L chromosomes) were annotated from genome contigs from the (male) somatic (head) and germline (testes with sperm) tissue. These were kindly provided by Dr. Hodson. P450 sequences of the Hodson et al. contigs that were absent from the Urban et al. assembly were by default assigned to the GRC, and these contigs were indeed identified as belonging to the GRC in Hodson et al. (2022; Suppl. Table S3). Chromosome anchoring was made possible with genome assembly GCA_014529535.2, VSDI02. The sequences were aligned by MAFFT (Katoh et al., 2019), and a maximum likelihood tree was built by IqTree with the Best-fit model: LG+F+R7 (Nguyen et al., 2015; Minh et al., 2020). The tree was drawn by FigTree (http://tree.bio.ed.ac.uk/).

3. Results and discussion

3.1 The CYPome: number and diversity of genes of the core genome

We first annotated the P450 genes from the three autosomes and X chromosome (see section 4 below for the GRC genes). Manual curation of the CYPome, matching transcripts and genomic sequences, yielded 173 full length P450 sequences including 8 pseudogenes (Suppl. data 1). With two alternatively spliced P450 transcripts (CYP4ZL1 and

CYP4ZL3, each resulting in two isoforms), this represents a total of 163 active P450 genes. There are two distinct CYP6AAK1 genes identical in sequence. Seven shorter sequence fragments and degenerate pseudogenes were not fully annotated. The genes were distributed in six CYP clans, the four clans typically found in insects (19 genes in the CYP2 clan, 85 in CYP3, 44 in CYP4, and 13 in the MITO clan) as well as one sequence in each of the CYP19 and CYP53 clans (see below)(Fig. 2). With 46 CYP families in *B. coprophila*, versus just 24 in *Drosophila*, and with 14 CYP families and 44 subfamilies not previously named from full CYPomes of other organisms, the *B. coprophila* CYPome is extremely diverse, and suggests that in large insect orders (Diptera, estimated 161,000 species) further P450 diversity remains to be discovered. Li et al. (2022) reported only 96 P450 genes in *Bradysia odoriphaga* (*cellarum*), a closely related species. This is significantly fewer, perhaps due in part to the less accurate gene prediction compared to manual curation, despite a reasonably good BUSCO (Benchmarking Universal Single-Copy Orthologue; Manni et al., 2021) score. However, the number of predicted gene models in *B. odoriphaga* (16,231) is also lower than that of *B. coprophila* (23,117, Urban et al., 2021; 22,029 Hodson et al., 2022), in a roughly similar proportion to the difference in CYPome size.

All CYP genes were assigned to one of the four chromosomes, with 30 on the X chromosome, 45 on chromosome II, 46 on III and 42 on IV (Suppl. data 2). The CYP genes of the germline-restricted chromosomes are discussed in section 4 below. With respect to chromosome size (Urban et al., 2022), the genes are not evenly distributed, but enriched on chromosomes II and III.

3.2. Aliens in the CYPome: extensive horizontal gene transfer (HGT)

Twenty-eight sequences, including one full length pseudogene, were clear outliers as their best BlastP matches were not sequences from Diptera, or even from insects. These sequences belong to CYP clans 2, 3 and 4, and one sequence each to CYP19 and CYP53 clans. Several lines of evidence indicate that these are real *B. coprophila* genomic sequences and not contaminations of some kind. Table 1 shows that all but two genes were represented by transcripts. All these “alien” P450s (except the pseudogene CYP4717A1P) had direct 1:1 (or 2:1 for CYP4721A1-2) orthologs in the *B. odoriphaga* genome (sequence identity 88-98%), and 13 had orthologs in *B. (Pseudolycoriella) hygida*. The geographical origin of the sequences for these three sciarids (USA, China, and Brazil) excludes a concerted or accidental contamination. Furthermore, only one genomic sequence was intronless (see below) with all 27 others being comprised of several exons with classical gt/ag

intron boundaries. These “alien” genes, marked in magenta on Fig. 2, are now discussed by clan.

3.2.1. CYP2 clan aliens

We found 19 CYP2 clan P450s, when most insects have ten or less (Dermauw et al., 2020). Ten CYP2 clan P450s are apparently of HGT origin. Their best matches in the NCBI NR database are all from Collembola, with overall identities between 40 and 60%. The maximum likelihood tree (Fig. 2) supports two clades of five alien genes each. The first supported clade (SH-aLRT 98.5 / UFboot 99) comprises the CYP3709 and 3710 families, with an average of 49% sequence identity to collembolan sequences. The second (99.1/100) comprises the CYP4721 and 4722 families, with an average of 41% sequence identity to collembolan sequences. This would suggest at least two independent HGT events, one somewhat more recent than the other, and each followed by gene duplications in the *Bradysia* ancestor.

3.2.2 CYP3 clan aliens

Fourteen CYP3 clan sequences are apparently of HGT origin, distributed in several clades with low (35%) identity to mite sequences or higher (65%) identity to collembolan sequences. Among the latter clades is one pseudogene, CYP4717A1P. This sequence is represented by two types of transcripts, covering 4 or 5 exons but both lacking the conserved Cys region essential to heme binding. It has no ortholog in *B. odoriphaga* or *B. hygida*, and its significance in the *B. coprophila* is unclear hence its pseudogene designation. One sequence, CYP4719A1, is 47% identical to oribatid mite sequences, and one, CYP3756A1, was previously identified as a probable P450 of HGT origin in *B. odoriphaga* (Dermauw et al., 2020). The different levels of sequence identity and the position of the sequences in the tree suggest three to six independent HGT events, of which one, the CYP4716 family, was followed by several gene duplication events in the ancestor of *B. coprophila* and *odoriphaga* (Fig. 3).

3.2.3 CYP4 clan aliens

Only one HGT event seem to have occurred in the CYP4 clan, bringing the ancestor of the two CYP4714A genes from a mite donor. This event is relatively ancient with percentage identities of 37-40% to the mite sequences, and the divergence from the *B. odoriphaga* (80-91%) and *hygida* (59-65%) paralogs is significant. Furthermore, despite their apparent origin from a duplication, the two genes (57% identity) are not tandemly clustered, with one on chromosome II and one on the X.

3.2.4. CYP53-like sequence from fungi

CYP53A57 represents the first insect CYP53 clan P450. The CYP53A57 gene of four exons on chromosome 4 is highly similar (about 74%) to CYP53 clan sequences from fungi, specifically Dothideomycetes (Ascomycota) (Fig. 4). This group of fungi is relatively poorly studied, but they comprise animal and plant pathogens as well as saprobes. The position and phase of the three phase 2 introns in *B. coprophila* are conserved in *Coniosporium apollinis* (72.7% id). The gene has an ortholog in *B. odoriphaga* genome KAG4071074 (94.4% id), and in *B. hygida* (KAG6633705.1; 78.8% id). The three introns are conserved in *B. odoriphaga*, but one has been lost in *B. hygida*. CYP53A57 is flanked by “dynein heavy-chain 7 axonemal” upstream and “class E bHLH protein 22” downstream in *B. coprophila*, a locus in conserved synteny in *B. odoriphaga* (JAFDOW010001058.1) and in other Diptera (*Contarinia nasturtii*, *Hermetia illucens* - but without the P450) (Fig. 5). Interestingly, a cyclopropane fatty acyl phospholipid synthase - like (PRK11705, cl30044) sequence is found at that locus as well. There are six such sequences in *B. coprophila*, three in *B. odoriphaga* and five in *B. hygida*. These are not found in other insects, but there are several in collembola, while their origin is clearly bacterial. This chromosomal locus therefore appears to be a “hotspot” for insertion of horizontally transferred genes.

The CYP53 clan is not universally present in fungi, generally as one CYP53A gene in Ascomycota, while Basidiomycota carry more diverse CYP53 genes (Jawallapersand et al., 2014). The CYP53A enzymes of ascomycetes are known to hydroxylate benzoic acid (Lah et al., 2011) and the CYP53D of basidiomycetes can dealkylate methoxystilbenes (Ide et al., 2012). This suggests that the *Bradysia* CYP53 may be involved in the metabolism of cyclic aromatic compounds.

3.2.5. CYP19-like sequence in *Bradysia* and collembolans

CYP4724A1 was a most unusual sequence, with best matches to sequences of the CYP19 clan. The top hit by BlastP (NR) was the collembolan *Folsomia candida* sequence XP_021967663.2 derived from genome assembly GCA_002217175.1 (LNIX01000067.1). A full-length sequence is found on another *F. candida* genome assembly GCA_020920555.1 (JAEMPG010000007.1; 140541-138549). The collembolan sequence, confirmed by transcripts, has a single phase 2 intron and codes for a 609 aa long protein about 47% identical to the fungus gnat sequence. The latter, however, was intronless in both *B. coprophila* and *B. odoriphaga* (JAFDOW010001459.1 plus 248531-250381) (92.4% identity). The loci of

CYP4724A1 on chromosome III and that of the orthologous gene in *B. odoriphaga* are not part of a syntenic block.

Similar CYP19-like sequences are found in the Collembola *Thalassaphorura encarpata*, *Pseudachorutes palmiensis* and *Neanura muscorum,* but not in *Sinella curviseta* or *Orchesella cincta.* Also, the Shanghai sexual strain of *Folsomia candida* (Luan et al., 2021), thought to be a different species from the parthogenetic strain sequenced before (Faddeeva-Vakhrusheva et al., 2017) lacks this CYP19-like sequence. This gene is therefore not present in all Collembola.

Until recently, the CYP19 clan represented sequences found in vertebrates, with CYP19A1 encoding the steroid A-ring aromatase enzyme. CYP19 was thought to have evolved late in the bilaterian lineage (Goldstone et al., 2016; Di Nardo et al., 2021). However, sequences similar to CYP19 and considered members of the CYP19 clan now appear to be more widely distributed. Of the 15 residues comprising the catalytic cleft in the human CYP19A1 structure (Ghosh et al., 2009) seven are identical in the *Bradysia* and *Folsomia* sequences and six are conservative substitutions, despite just an overall 26% identity of these arthropod sequences to the human one (Fig. 6). It is therefore expected that the substrate binding cavity of the fungus gnat CYP4724A1 would be small and accommodating a flat, perhaps cyclic aromatic substrate.

The distribution of P450 in clans in the collembolan *Sinella curviseta* is skewed toward the CYP2 clan (142 sequences versus 57 in the CYP3 clan and 59 in the CYP4 clan, with just 4 in the mito clan) (Dermauw et al., 2020). If this distribution is similar in most Collembola, then with two probable HGT events from the CYP2 clan and three or more from the CYP3 clan, it would seem that the probability of transfer is not random.

3.3 Paucity of gene clusters

In contrast with other Diptera (and more generally, arthropods, Dermauw et al., 2020), very few P450s are clustered on *B. coprophila* chromosomes. Lineage-specific expansions of CYP subfamilies or blooms (Feyereisen, 2011) are usually observed as gene clusters, but this is not the case in *B. coprophila*. For each of the seven CYP12BA, CYP4ZL or CYP6FV genes, only a pair are found as tandem duplicates. Even the CYP4716A “alien” genes and pseudogenes are dispersed at distinct loci in both *B. coprophila* and *B.odoriphaga*, even though they appear to result from a single HGT event followed by duplications in the host species (Fig. 3). Overall, only 24 genes of 165 were clustered, mostly as tandem duplicates, and the large majority of P450s were spread out in the genome as singletons.

This distribution of genes differs markedly from that seen in other insects with high quality genome assemblies, and is not constrained by the phylogeny (Fig. 7). In particular, highly conserved genes that are most commonly clustered in insects such as CYP18A1 and CYP306A1, or CYP15 and CYP305, are found on different chromosomes in *B. coprophila,* while CYP307A2 and the Rieske-domain protein neverland, are 17.7 Mb distant on the X chromosome.

There are seven CYP4G genes in *B. coprophila*, a singleton CYP4G15-type gene (CYP4G313) on the X and six CYP4G1-type genes, a distant tandem duplicate on the X and a cluster of three genes and a pseudogene on chromosome II. All Neoptera have at least one CYP4G gene involved in cuticular hydrocarbon (CHC) biosynthesis (Feyereisen, 2020). Mosquitoes have two, both contributing to the CHC blend (Kefi et al., 2019). In contrast, house flies have six and a pseudogene, with only CYP4G2 documented as CHC biosynthetic enzyme (Qiu et al., 2012) while the others are currently not functionally characterized.

Interestingly, one of the large clusters of P450 genes comprises four genes of the CYP3 clan, two CYP6SX genes, CYP6AAS1 and CYP3978B1. This is an unusual clustering of P450 genes from different (sub)families, when most clusters in other species are often long series of tandem duplicated genes from the same subfamily. The genomic organization was therefore analyzed in greater detail, revealing that this cluster was located in intron 3 of the activating signal cointegrator 1 complex subunit 2 gene (XP_037026646.1) on the X chromosome. This conserved, large gene (33.6 kb) has a long intron 3 of 31kb. Clearly, the large size of the intron is an opportunity allowing nesting. Four orthologous P450s were also nested in the same gene in *B. odoriphaga* and *B. hygida* (Fig. 8). The “host” gene in other Sciaroidea, *Bolitophila cinerea* (Mycetophilidae), *Contarinia nasturtii* and *Sitodiplosis mosellana* (Cecidomyiidae), also had a nested CYP6 gene. These two Cecidomyiidae genomes had additional CYP6 genes just downstream of the host gene (Fig. 8). Even in the mosquito *Aedes albopictus*, three CYP6 genes were nested in the third intron of the orthologous gene (XP_029710533.1). In Lepidoptera, this gene covers 12-15 exons, with no nested P450 gene.

Nesting genes in an intron seems to place a limit to recombination, and the resulting cluster in an intron is therefore stable over evolutionary time. Here, the age of nesting is at least 230MY given the divergence time of the host species (Fig. 3). It would be interesting to determine if the neighboring P450 genes in Cecidomyiidae are on their way in, or if they have actually escaped nesting. Other examples of nested P450 genes seem to be of more recent origin. The human CYP51P3 pseudogene is nested in the SASH1 gene and as such it is protected. The same pseudogene can be seen over 66 MY in eutherian mammals and

it has not been lost. The CYP2C64P pseudogene is found inside the MTMR1 gene and has representatives in grey mouse-lemur with that ancestral split dating to 58 million years.

3.4 Alternative splicing

In *B. coprophila*, the deep TSA coverage has confirmed two cases of alternative splicing in CYP4 family genes. These are named according to the rules (Nelson et al., 2004) describing the variable exons. CYP4ZL1_v1a2a uses distal exons 1a and 2a, while CYP4ZL1_v1b2b uses 2 proximal exons 1b and 2b. Similarly, CYP4ZL3_v1a2a has distal exons 1a and 2a, and CYP4ZL3_v1b1'b2b has proximal exons,1b-1'b (the homologous first exon is split by an additional intron) and 2b. In both cases the isoforms use common last 5 exons. The genome further shows two untranscribed exons (equivalent to exons 1 and 2) upstream of CYP4ZL1v1. Although alternative splicing is common for vertebrate P450s (Annalora et al., 2017), examples are far fewer in insects. Thus in *D. melanogaster*, there are alternative splice forms of CYP4D1, CYP4D1_v1a and CYP4D1_v1b (Tijet et al., 2001; Chung et al., 2009; Good et al., 2014). Similarly, the CYP4CE1 gene in *Nilaparvata lugens* encodes two alternative splice forms differing in the use of the first exon (Liu et al., 2021). It is likely that a thorough analysis of RNAseq data will reveal more examples of alternative splicing of P450 transcripts in insects.

4. P450s on the germline-restricted chromosomes

Genomic contigs covering the hologenome (Hodson et al., 2022) were mined for P450 sequences. All 163 genes from the X and autosomes were found, and a number of additional sequences were by default assigned to the L chromosomes, or genome-restricted chromosomes. This way, 99 sequences were found on the GRC scaffolds. We found 15 CYP2 clan sequences of which 10 are pseudogenes, 56 CYP3 clan sequences of which 36 are pseudogenes, 24 CYP4 clan sequences of which 16 are pseudogenes and 4 Mito clan sequences of which 2 are pseudogenes. There was also a CYP53A57 homolog pseudogene. Thus only 33 GRC sequences were apparently intact P450s, while 66 were clearly pseudogenes or gene fragments. None of the 33 intact P450 sequences were represented in the publicly available transcriptome shotgun assemblies (GJHU01 and GJHV01). The 66 pseudogenes were clearly recognizable P450 sequences interrupted by one or more stop codons, frameshifts or indels of various sizes. We counted 17 stop codons, 30 frameshifts

and 29 indels distributed among those pseudogenes, with 13 such defects impacting the nearly universally conserved cysteine that provides the thiolate ligand to the heme.

Hodson et al. (2022) proposed that the GRC (L chromosomes) originated from an ancient interspecific hybridization event between a *Bradysia* ancestor (Sciaridae) and an ancestor in the Cecidomyiidae. We therefore chose the genome of *Contarinia nasturtii* as an extant Cecidomyiid, and compared the GRC sequences to their closest match in *C. nasturtii* and in *B. coprophila* (A/X core genome). Of the 33 genes, 16 were on average 79.6% identical to the closest *B. coprophila* sequence and only 43.5% identical to the closest *C. nasturtii* sequence while 17 are more Cecidomyiid-like (68.2% vs 54.9%). The GRC sequences therefore appear to be of heterogeneous origin, yet quite distinct from the autosomal / X chromosome P450s. This is clearly seen from a plot of the respective % identities of the GRC sequences to their best *Bradysia* and *Contarinia* matches, which shows two distinct groups (Fig. 9). A striking example of a Cecidomyiid-like GRC sequence is CYP315A1, the autosomal copy has 8 exons, while the L copy has just 2. The L copy of CYP315A1 is 63.6% identical to the *C. nasturtii* CYP315A1, but just 46.1% identical to the autosomal copy. Moreover the *C. nasturtii* CYP315A1 has just one intron at the same position as in the L copy.

GRC sequences on the other hand also include eight “alien” P450s (see above), one intact gene (CYP3717A5) and five pseudogenes that all have an A/X homolog. In addition, there are two GRC pseudogenes that have no clear A/X homologs, but seem to be of collembolan origin, a CYP18A1 pseudogene (CYP2 clan) and a CYP3808A1 pseudogene (CYP4 clan). This suggests that these were aliens in a sciarid ancestor and were since lost in the extant three species for which we have WGS evidence.

5. Conclusion

The *B. coprophila* CYPome is quite unlike the CYPomes of other organisms as a result of the many genes resulting from HGT and of the unusual complement of genes and pseudogenes on germline-restricted chromosomes.

HGT with potential or demonstrated adaptive significance are increasingly being documented in arthropods for a variety of genes (Wybouw et al., 2014, 2016; Verster et al., 2021). Yet HGT in the P450 superfamily has so far been restricted to a few, albeit significant, cases. Ngcobo et al. (2023) recently proposed that all P450 genes in Archaea may originate from bacteria by plasmid-mediated transfer. Allene oxide synthase (CYP74), and related oxylipin-metablizing P450s are found in rhizobacteria, plants and cnidaria (Lee et al., 2008).

Fungal CYP55 (nitric oxide reductase P450nor) is a peculiar P450 that directly oxidizes NADH without a redox partner. It is thought to result from HGT from bacteria (Tomura et al., 1994, Higgins et al., 2018). A CYP55 gene is also found in chlorophytes (Hansen et al., 2021), suggesting another HGT event. The odd distribution of CYP19-like genes, that were initially thought to be a vertebrate innovation (Goldstone et al., 2016), is also suggestive of HGT. So is the similarity of *Bradysia* and collembolan CYP19 sequences, even if the origin of the collembolan CYP19 itself remains obscure. The HGT reported here, for CYP53 and for the CYP2, 3, and 4 clan genes are more "classical" P450s than the ones described above, and their transfer apparently quite recent, perhaps less than 100 MYA. Indeed, these alien P450s are found in three species of Sciaridae, but absent from other families of Sciaroideae. HGT of P450 genes appears to confer a selective advantage, as fungus gnat resistance to insecticides is linked to the expression of several of these genes, including CYP3356A1, mainly expressed in larval Malpighian tubules and midgut of *B. odoriphaga* (Chen et al., 2019) and CYP3828A1 (Zhang et al., 2023). The availability of a genetic transformation system for *B. coprophila* (Yamamoto and Gerbi, 2022) should allow testing the adaptive value of the HGT genes. One may expect to find other genes (i.e. beyond P450s), as horizontally transfered genes in *Bradysia*. Examples include bacterial tryptophan indole-lyase (Yuasa, 2022) and a few CAZY enzymes such as GH19 chitinases (Trinca et al., 2023). However, with most HGT events described here apparently originating within collembolans or mites, sorting differential gene loss from HGT may prove to be laborious given the relative phylogenetic proximity when compared to cross-kingdom HGT events.

With the P450 genes from the GRC consisting of about half Sciarid-related genes and half Cecidomyiid-related genes, the distant introgression of a haploid Cecidomyiid genome postulated by Hodson et al. (2022) was probably followed by interchromosomal exchanges. Such extensive genomic rearrangements could possibly have been propitious for HGT from organisms sharing the fungus gnats' ecological niche, such as collembola and fungi. As the alien P450s appear to be restricted to the *Bradysia* genus, and as some aliens are found in the GRC, it follows that the HGT are posterior to the cross-family introgression event. The divergence between Cecidomyiidae and Sciaridae dates to 145-120 MYA (Friedrich & Tautz, 1997; Wiegmann et al., 2011), which puts an upper limit to the age of the HGT events (Fig. 10).

Although the introgressed genome, now consisting of two GRCs, has seen translocation of host genes (Hodson et al., 2022 and Fig. 9), the initial event would amount to an allopolyploidization. This event therefore would impact pseudogenization rates to a different

degree depending on a gene's sensitivity to dosage imbalance. While most "Ohnologues" are rapidly pseudogenized, a significant number of often important genes in signaling and development are retained (van de Peer et al., 2017). Interestingly, Hodson et al. (2022) did not address pseudogenization of GRC genes, and the function of GRC in Sciaridae remains unclear (Hodson and Ross, 2021). With two thirds of the P450 genes from the GRC being pseudogenized, and some probably lost over evolutionary time, it is intriguing to see the bias in the genes that have been retained, apparently in an active state. Four of them are known to participate in juvenile hormone- (CYP15) and ecdysteroid biosynthesis (CYP315, CYP302 and CYP307). Interestingly, CYP307 and CYP315 are also retained Ohnologs in *Limulus polyphemus* (Dermauw et al., 2020). Proportionally, genes of the CYP2 clan and of the CYP9 family have mostly been retained as active genes in the GRC. Future work should explore the expression and function of these genes in the germline.

In *B. coprophila*, the extensive chromosome rearrangements which see the translocation of many genes from the core genome to the GRC, as well as the dispersion of duplicated genes throughout the genome, as opposed to their commonly observed clustering, may facilitate the HGT of genes from organisms which share its ecological niche.

Acknowledgements: We thank Dr. C. N. Hodson (University of Edinburgh) for access to the *Bradysia* hologenome contigs (Hodson et al., 2022).

Appendix A. Supplementary data

Supplementary data to this article can be found online at xxxxx

Suppl. data 1
FASTA file of the *Bradysia coprophila* CYPome. The sequences are given with the official CYP name. Full length pseudogenes are listed at the end of the file, and these sequences are not included in the tree of Fig. 2. For the pseudogene sequences, X represents a frameshift and * a stop codon. Sequences from the germline-restricted chromosomes are marked as GRC.

Suppl. data 2
Chromosomal assignment and position of the 173 full-length P450 sequences of the core genome.

Suppl. data 3
Phylogeny of the *Bradysia coprophila* CYPome (see Fig 2) shown with bootstrap support values (in % from 1000 bootstraps).

References

Annalora AJ, Marcus CB, Iversen PL. 2017. Alternative Splicing in the Cytochrome P450 Superfamily Expands Protein Diversity to Augment Gene Function and Redirect Human Drug Metabolism. Drug Metab Dispos. 45:375-389. doi: 10.1124/dmd.116.073254

Chen H, Lin L, Ali F, Xie M, Zhang G, Su W. 2017. Using Next-Generation Sequencing to Detect Differential Expression Genes in *Bradysia odoriphag*a after Exposure to Insecticides. Int J Mol Sci.18: 2445. doi: 10.3390/ijms18112445.

Chen C, Wang C, Liu Y, Shi X, Gao X. 2018. Transcriptome analysis and identification of P450 genes relevant to imidacloprid detoxification in *Bradysia odoriphaga*. Sci Rep. 8: 2564. doi: 10.1038/s41598-018-20981-2.

Chen C, Shan T, Liu Y, Shi X, Gao X. 2019. Identification of a novel cytochrome P450 CYP3356A1 linked with insecticide detoxification in *Bradysia odoriphaga*. Pest Management Science 75: 1006-1013

Chen C, Wang C, Liu Y, Shan T, Shi X, Gao X. 2022. Integration analysis of PacBio SMRT- and Illumina RNA-seq reveals P450 genes involved in thiamethoxam detoxification in *Bradysia odoriphaga*. Pestic Biochem Physiol. 186:105176. doi: 10.1016/j.pestbp.2022.105176

Chung H, Sztal T, Pasricha S, Sridhar M, Batterham P, Daborn PJ. 2009. Characterization of *Drosophila melanogaster* cytochrome P450 genes. Proc Natl Acad Sci U S A. 106:5731-6. doi: 10.1073/pnas.0812141106

Dermauw W, Van Leeuwen T, Feyereisen R. 2020. Diversity and evolution of the P450 family in arthropods. Insect Biochem Mol Biol. 127:103490. doi: 10.1016/j.ibmb.2020.103490

Di Nardo G, Zhang C, Marcelli AG, Gilardi G. 2021. Molecular and Structural Evolution of Cytochrome P450 Aromatase. Int J Mol Sci. 22:631. doi: 10.3390/ijms22020631

Faddeeva-Vakhrusheva A, Kraaijeveld K, Derks MFL, Anvar SY, Agamennone V, Suring W, Kampfraath AA, Ellers J, Le Ngoc G, van Gestel CAM, Mariën J, Smit S, van Straalen

NM, Roelofs D. 2017. Coping with living in the soil: the genome of the parthenogenetic springtail *Folsomia candida*. BMC Genomics. 18:493. doi: 10.1186/s12864-017-3852-x

Feron R, Waterhouse RM. 2022. Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. Gigascience. 11:giac006. doi: 10.1093/gigascience/giac006

Feyereisen R. 2006. Evolution of insect P450. Biochem Soc Trans. 34:1252-5. doi: 10.1042/BST0341252

Feyereisen R. 2011. Arthropod CYPomes illustrate the tempo and mode in P450 evolution. Biochim Biophys Acta. 1814:19-28. doi: 10.1016/j.bbapap.2010.06.012.

Feyereisen R. 2020. Origin and evolution of the CYP4G subfamily in insects, cytochrome P450 enzymes involved in cuticular hydrocarbon synthesis. Mol Phylogenet Evol. 143:106695. doi: 10.1016/j.ympev.2019.106695.

Friedrich M, Tautz D. 1997. Evolution and phylogeny of the Diptera: a molecular phylogenetic analysis using 28S rDNA sequences. Syst Biol. 46 :674-98. doi: 10.1093/sysbio/46.4.674.

Ghosh D, Griswold J, Erman M, Pangborn W. 2009. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. Nature. 457:219-23. doi: 10.1038/nature07614

Goldstone JV, Sundaramoorthy M, Zhao B, Waterman MR, Stegeman JJ, Lamb DC. 2016. Genetic and structural analyses of cytochrome P450 hydroxylases in sex hormone biosynthesis: Sequential origin and subsequent coevolution. Mol Phylogenet Evol. 94(Pt B):676-687. doi: 10.1016/j.ympev.2015.09.012

Good RT, Gramzow L, Battlay P, Sztal T, Batterham P, Robin C. 2014. The molecular evolution of cytochrome P450 genes within and between drosophila species. Genome Biol Evol. 6:1118-34. doi: 10.1093/gbe/evu083

Hansen CC, Nelson DR, Møller BL, Werck-Reichhart D. 2021. Plant cytochrome P450 plasticity and evolution. Mol Plant. 14:1244-1265. doi: 10.1016/j.molp.2021.06.028.
Higgins SA, Schadt CW, Matheny PB, Löffler FE. 2018. Phylogenomics Reveal the Dynamic Evolution of Fungal Nitric Oxide Reductases and Their Relationship to Secondary Metabolism. Genome Biol Evol. 10: 2474-2489. doi: 10.1093/gbe/evy187.

Hodson CN, Ross L. 2021. Evolutionary Perspectives on Germline-Restricted Chromosomes in Flies (Diptera). Genome Biol Evol. 13: evab072. doi: 10.1093/gbe/evab072.

Hodson CN, Jaron KS, Gerbi S, Ross L. 2022. Gene-rich germline-restricted chromosomes in black-winged fungus gnats evolved through hybridization. PLoS Biol. 20:e3001559. doi: 10.1371/journal.pbio.3001559

Ide M, Ichinose H, Wariishi H. 2012. Molecular identification and functional characterization of cytochrome P450 monooxygenases from the brown-rot basidiomycete *Postia placenta*. Arch Microbiol. 194:243-53. doi: 10.1007/s00203-011-0753-2.

Jawallapersand P, Mashele SS, Kovačič L, Stojan J, Komel R, Pakala SB, Kraševec N, Syed K. 2014. Cytochrome P450 monooxygenase CYP53 family in fungi: comparative structural and evolutionary analysis and its role as a common alternative anti-fungal drug target. PLoS One. 9:e107209. doi: 10.1371/journal.pone.0107209.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772-80. doi: 10.1093/molbev/mst010

Kefi M, Balabanidou V, Douris V, Lycett G, Feyereisen R, Vontas J. 2019. Two functionally distinct CYP4G genes of Anopheles gambiae contribute to cuticular hydrocarbon biosynthesis. Insect Biochem Mol Biol. 110:52-59. doi: 10.1016/j.ibmb.2019.04.018.

Lah L, Podobnik B, Novak M, Korošec B, Berne S, Vogelsang M, Kraševec N, Zupanec N, Stojan J, Bohlmann J, Komel R. 2011. The versatility of the fungal cytochrome P450 monooxygenase system is instrumental in xenobiotic detoxification. Mol Microbiol. 81:1374-89. doi: 10.1111/j.1365-2958.2011.07772.x

Lee DS, Nioche P, Hamberg M, Raman CS. 2008. Structural insights into the evolutionary paths of oxylipin biosynthetic enzymes. Nature. 455: 363-8. doi: 10.1038/nature07307.

Li M, Yang X, Fan F, Ge Y, Hong D, Wang Z, Lu C, Chen S, Wei G. 2022. De novo genome assembly of *Bradysia cellarum* (Diptera: Sciaridae), a notorious pest in traditional special vegetables in China. Insect Mol Biol. 31: 508-518. doi: 10.1111/imb.12776. Epub 2022 Apr 18.

Luan YX, Cui Y, Chen WJ, Jin JF, Liu AM, Huang CW, Potapov M, Bu Y, Zhan S, Zhang F, Li S. 2022. High-quality genomes reveal significant genetic divergence and cryptic speciation in the model organism *Folsomia candida* (collembola). Mol Ecol Resour. doi: 10.1111/1755-0998.13699

Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. 2021.BUSCO: Assessing genomic data quality and beyond. CurrentProtocols,1, e323. doi: 10.1002/cpz1.323

Minh, B.Q. , Schmidt, H.A. , Chernomor, O. , Schrempf, D. , Woodhams, M.D. , Von Haeseler, A. , Lanfear, R. , 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37, 1530–1534 .

Ngcobo PE, Nkosi BVZ, Chen W, Nelson DR, Syed K. 2023. Evolution of Cytochrome P450 Enzymes and Their Redox Partners in Archaea. Int J Mol Sci. 24:4161. doi: 10.3390/ijms24044161.

Nguyen L.-T., Schmidt H.A., von Haeseler A., and Minh B.Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol. Evol., 32:268-274.

Qiu Y, Tittiger C, Wicker-Thomas C, Le Goff G, Young S, Wajnberg E, Fricaux T, Taquet N, Blomquist GJ, Feyereisen R. 2012. An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. Proc Natl Acad Sci U S A. 109:14858-63. doi: 10.1073/pnas.1208650109

Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, Unger MF, Collins FH, Feyereisen R. 2002. Evolution of supergene families associated with insecticide resistance. Science. 298:179-81. doi: 10.1126/science.1076781

Ševčík J, Kaspřák D, Mantič M, Fitzgerald S, Ševčíková T, Tóthová A, Jaschhof M. 2016. Molecular phylogeny of the megadiverse insect infraorder Bibionomorpha sensu lato (Diptera). PeerJ. 4:e2563. doi: 10.7717/peerj.2563.

Sezutsu H, Le Goff G, Feyereisen R. 2013. Origins of P450 diversity. Philos Trans R Soc Lond B Biol Sci. 368:20120428. doi: 10.1098/rstb.2012.0428

Shiningavamwe A, Obiero G, Albertyn J, Nicaud JM, Smit M. 2006. Heterologous expression of the benzoate para-hydroxylase encoding gene (CYP53B1) from *Rhodotorula minuta* by *Yarrowia lipolytica*. Appl Microbiol Biotechnol. 72:323-9. doi: 10.1007/s00253-005-0264-7

Theron CW, Labuschagné M, Albertyn J, Smit MS. 2019. Heterologous coexpression of the benzoate-para-hydroxylase CYP53B1 with different cytochrome P450 reductases in various yeasts. Microb Biotechnol. 12:1126-1138. doi: 10.1111/1751-7915.13321.

Tijet N, Helvig C, Feyereisen R. 2001.The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny. Gene. 262:189-98. doi: 10.1016/s0378-1119(00)00533-3

Tomura D, Obika K, Fukamizu A, Shoun H. 1994. Nitric oxide reductase cytochrome P-450 gene, CYP 55, of the fungus *Fusarium oxysporum* containing a potential binding-site for FNR, the transcription factor involved in the regulation of anaerobic growth of *Escherichia coli*. J Biochem. 116:88-94. doi: 10.1093/oxfordjournals.jbchem.a124508.

Trinca V, Carli S, Uliana JVC, Garbelotti CV, Mendes da Silva M, Kunes V, Meleiro LP, Brancini GTP, Menzel F, Andrioli LPM, Torres TT, Ward RJ, Monesi N. 2023. Biocatalytic potential of Pseudolycoriella CAZymes (Sciaroidea, Diptera) in degrading plant and fungal cell wall polysaccharides. iScience. 26:106449. doi: 10.1016/j.isci.2023.106449

Urban JM, Foulk MS, Bliss JE, Coleman CM, Lu N, Mazloom R, Brown SJ, Spradling AC, Gerbi SA. 2021. High contiguity de novo genome assembly and DNA modification analyses for the fungus fly, *Sciara coprophila*, using single-molecule sequencing. BMC Genomics. 22 643. doi: 10.1186/s12864-021-07926-2

Urban JM, Gerbi SA, Spradling AC. 2022. Chromosome-scale scaffolding of the fungus gnat genome (Diptera: *Bradysia coprophila*). bioRxiv preprint doi.org/10.1101/2022.11.03.515061.

Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. Nat Rev Genet.18:411-424. doi: 10.1038/nrg.2017.26.

Verster KI, Tarnopol RL, Akalu SM, Whiteman NK. 2021. Horizontal Transfer of Microbial Toxin Genes to Gall Midge Genomes. Genome Biol Evol. 13: evab202. doi: 10.1093/gbe/evab202.

Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, Bertone MA, Cassel BK, Bayless KM, Heimberg AM, Wheeler BM, Peterson KJ, Pape T, Sinclair BJ, Skevington JH, Blagoderov V, Caravas J, Kutty SN, Schmidt-Ott U, Kampmeier GE, Thompson FC, Grimaldi DA, Beckenbach AT, Courtney GW, Friedrich M, Meier R, Yeates DK. 2011. Episodic radiations in the fly tree of life. Proc Natl Acad Sci U S A. 108:5690-5. doi: 10.1073/pnas.1012675108.

Wybouw N, Dermauw W, Tirry L, Stevens C, Grbić M, Feyereisen R, Van Leeuwen T. 2014. A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. Elife. 3: e02365. doi: 10.7554/eLife.02365.

Wybouw N, Pauchet Y, Heckel DG, Van Leeuwen T. 2016. Horizontal Gene Transfer Contributes to the Evolution of Arthropod Herbivory. Genome Biol Evol. 8:1785-801. doi: 10.1093/gbe/evw119.

Yamamoto Y, Gerbi SA. 2022. Development of Transformation for Genome Editing of an Emerging Model Organism. Genes (Basel).13:1108. doi: 10.3390/genes13071108.

Yuasa HJ. Metazoan tryptophan indole-lyase: Are they still active? 2022. Comp Biochem Physiol B Biochem Mol Biol. 263:110801. doi: 10.1016/j.cbpb.2022.110801.

Zhang C, Du S, Liu R, Dai W. 2022. Overexpression of Multiple Cytochrome P450 Genes Conferring Clothianidin Resistance in *Bradysia odoriphaga*. J Agric Food Chem. 70:7636-7643. doi: 10.1021/acs.jafc.2c01315

Zhang C, Zhou T, Li Y, Dai W, Du S. 2023. Activation of CncC pathway is involved in the regulation of P450 genes responsible for clothianidin resistance in *Bradysia odoriphag*a. Pest Manag Sci. doi: 10.1002/ps.7482.

Table 1 : P450 genes resulting from Horizontal Gene Transfer in *B. coprophila.* (CYP name, CYP clan, accession number, orthologs in *B. odoriphaga* and *B. hygida*, best blastp match).

Legends to Figures

Fig. 1. Schematic phylogeny of Diptera with approximate divergence times, showing the position of *B. coprophila*. There are many genomes available for mosquitoes (Culicomorpha) and Drosophilidae (Ephydroidea), some for Calyptratae and Tephrididoidea, but relatively fewer in between. Representative species of eight families from 251 are shown. For a more detailed phylogeny of Diptera, see Wiegmann et al. (2011).

Fig. 2 Phylogeny of the *B. coprophila* CYPome. The four major insect CYP clans are colored blue (CYP2 clan), red (mito clan), green (CYP3 clan) and brown (CYP4 clan). Alien P450 sequences are marked in magenta. P450 sequences from germline-restricted chromosomes are marked in blue. The basal CYP4724A1 sequence belongs to the CYP19 clan. Bootstrap values are shown in Suppl. data 3.

Fig. 3. Phylogenetic relationships between the CYP4716A genes from three *Bradysia* species. The tree suggests a single HGT event from mites, followed by duplications in *B. coprophila* and *B. odoriphaga*. These genes are all found at different loci on the chromosomes and not as tandem duplicates or clusters. Numbers at nodes are the % bootstrap support from 1000 bootstraps.

Fig. 4. Phylogeny of CYP53A genes of Ascomycota, suggesting the origin of the horizontal gene transfer to a *Bradysia* ancestor. Numbers at nodes are the % bootstrap support from 1000 bootstraps.

Fig. 5. The CYP53A57 locus in Diptera. Synteny is conserved between *B. coprophila* and *B. odoriphaga*. The locus is conserved in more distantly related Diptera, but lacks the horizontally transfered genes.

Fig. 6. Partial alignment of the CYP19 clan P450s from *B. coprophila*, *Folsomia candida* (Collembola) and human CYP19A1. blue: catalytic cleft as defined by Ghosh et al. (2009);

w on top of the alignment: van der Waals contact with substrate; magenta: conserved P450 residues; orange: conserved residues in Substrate Recognition Sites of aromatases (Di Nardo et al., 2021).

Fig. 7. Percentage of P450 genes found as singletons, tandem duplicates and clusters in various insect genomes. The distribution in *B. coprophila* is independent of its phylogenetic position.

Fig. 8. CYP3 clan genes nested in the activating signal cointegrator 1 complex subunit 2 gene (green) in *Bradysia* and other dipteran genomes.

Fig. 9. Percentage identity of P450 sequences from the GRC to their closest homolog in the *B. coprophila* (Sciaridae) core genome and in *Contarinia nasturtii* (Cecidomyiidae).

Fig. 10. Summary diagram showing the origin of the current *Bradysia* CYPome. Following the proposal of Hodson et al. (2022), hybridization of sciarid and cecidomyiid ancestors led to an allopolyploid ancestor. Massive pseudogenization of the newly acquired genes (*) was accompanied by substantial chromosome rearrangements (double arrow). The molecular processes leading to such rearrangements may have been propitious to the integration of genes from the environment (blue arrows) that have occured several times. These genes are now recognized as “alien” genes.

Figure1

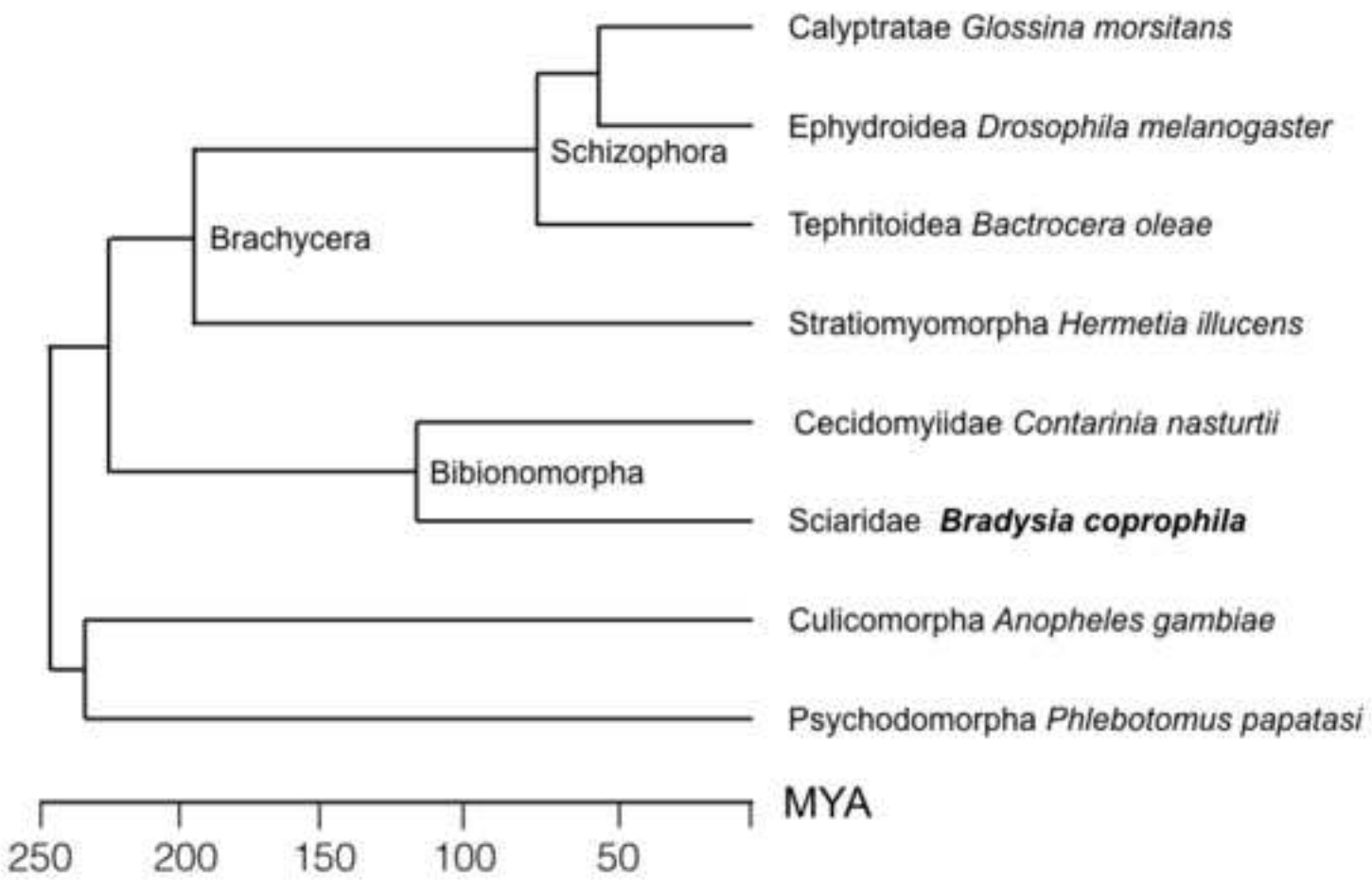

Figure2

Click here to access/download;Figure;Fig2.pdf

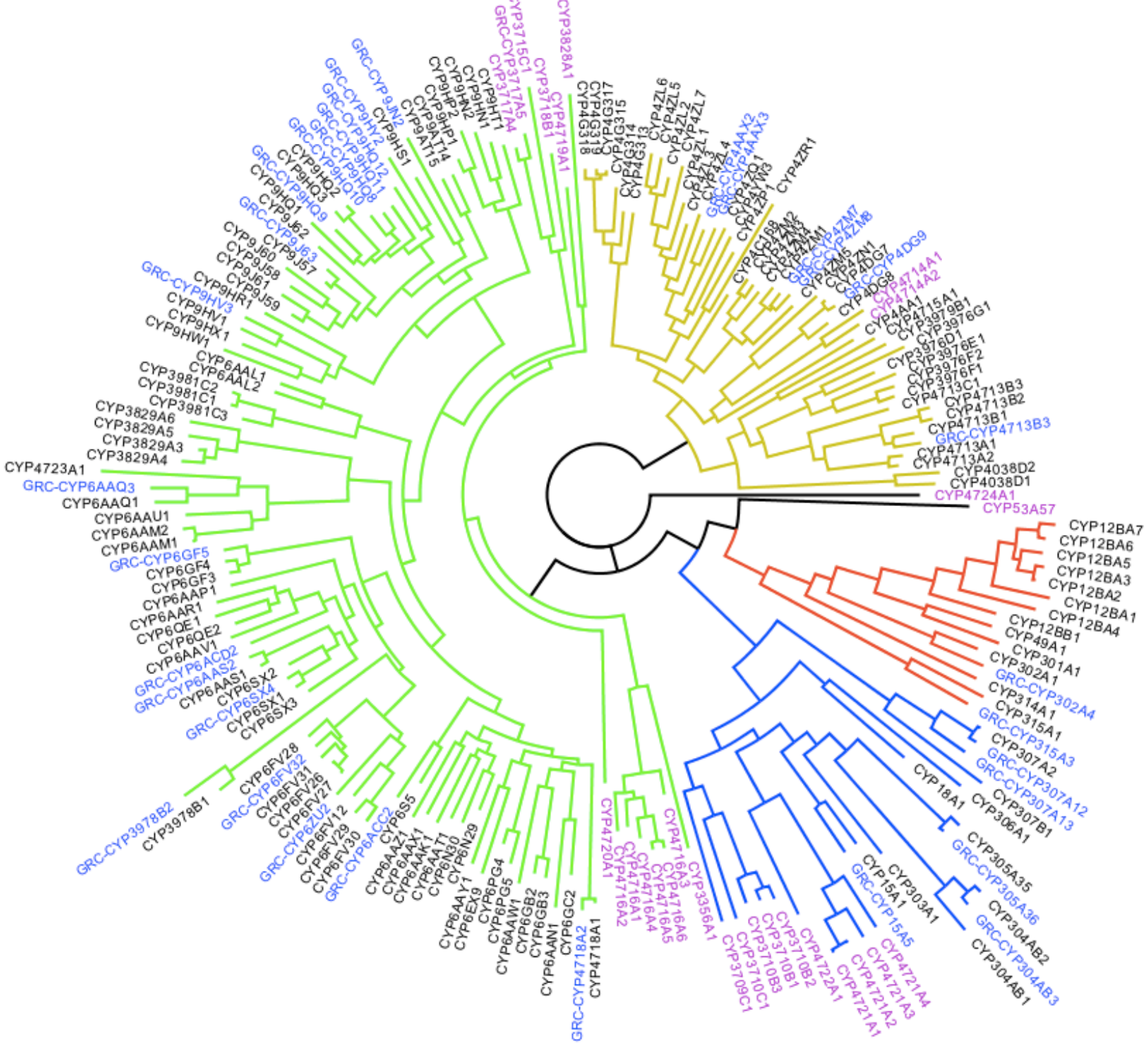

Figure 3

Click here to access/download;Figure;Fig3.001.tiff

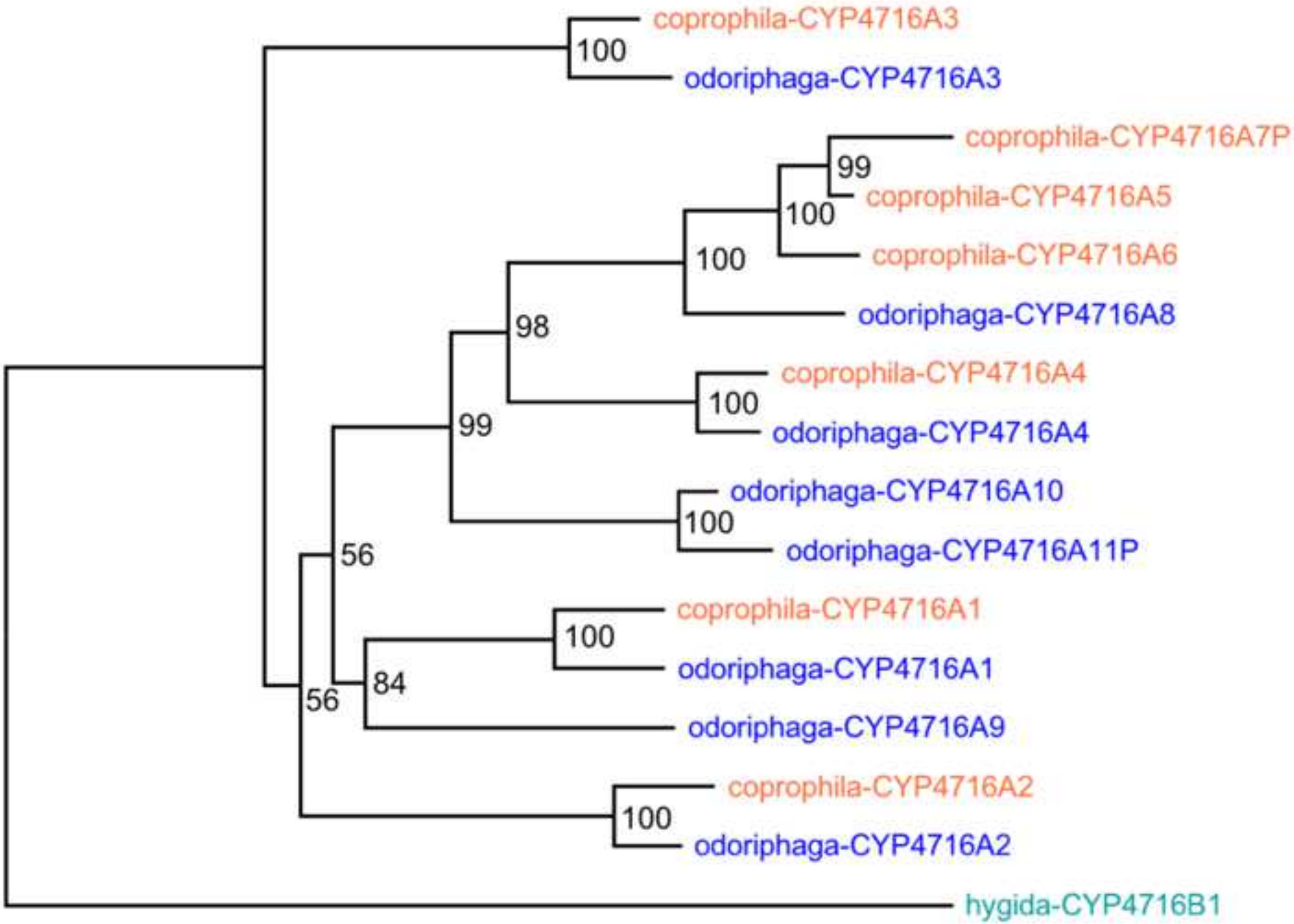

Figure 4

Click here to access/download;Figure;Fig4.001.tiff

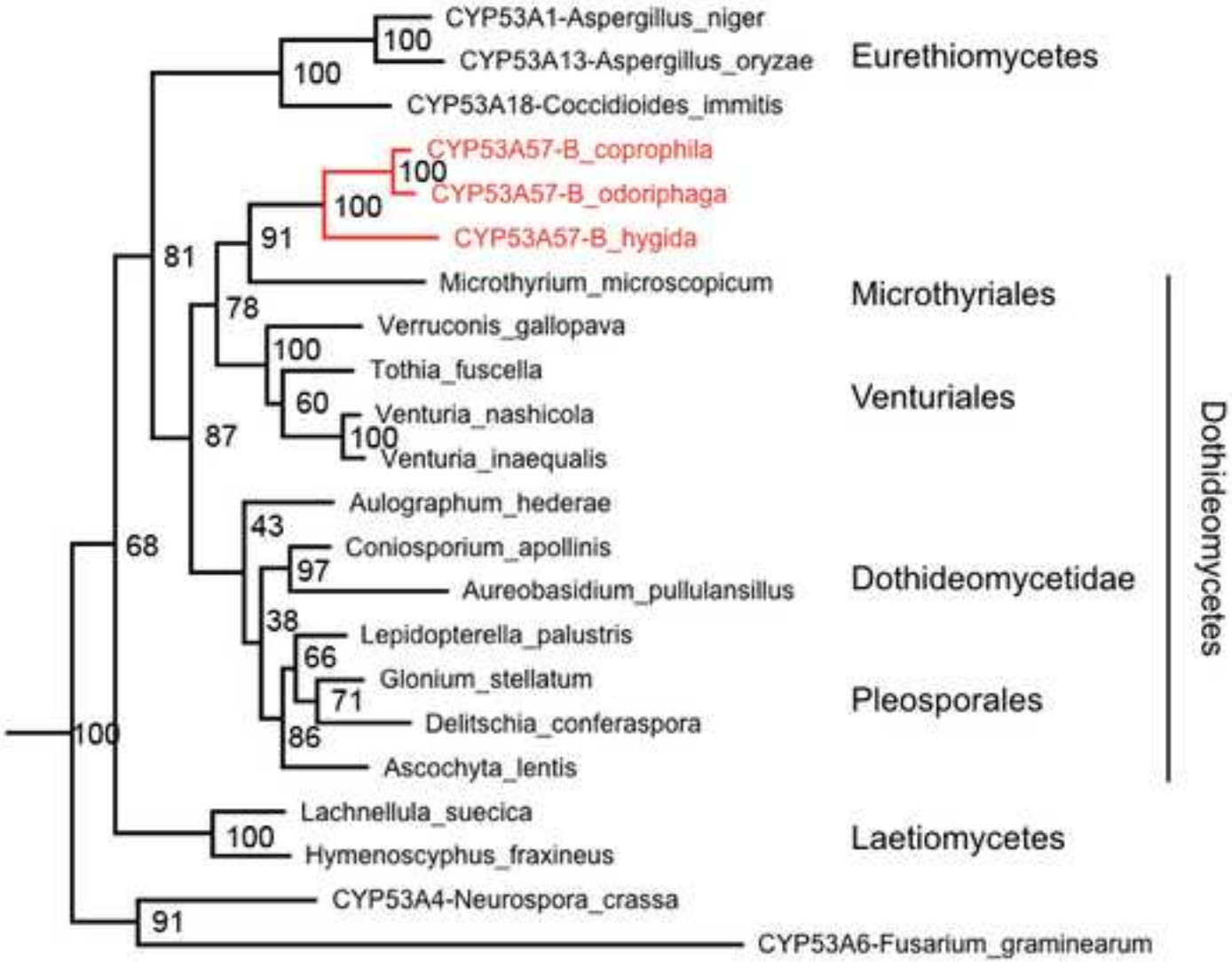

Figure5

Click here to access/download;Figure;Fig5.tiff

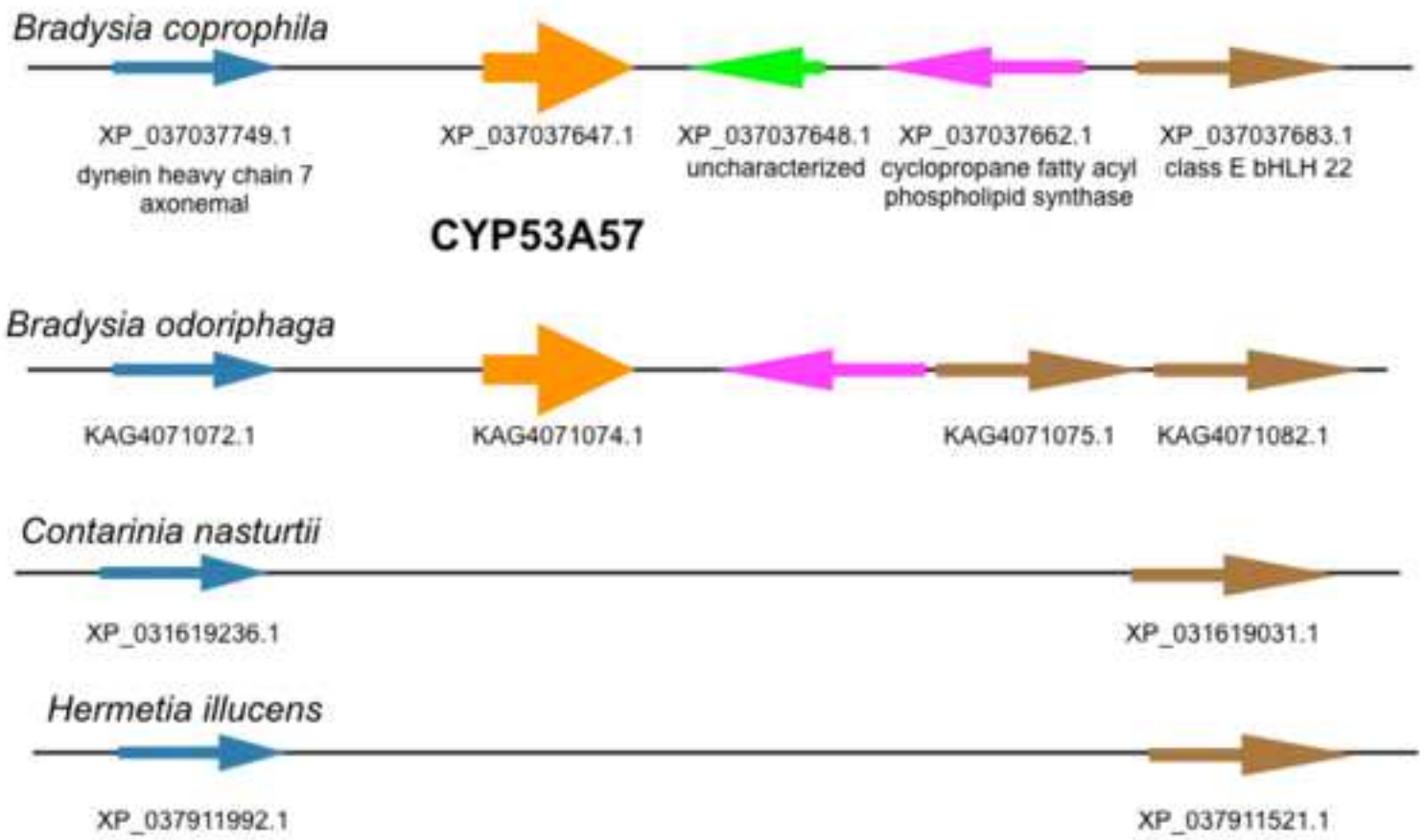

Figure6

Click here to access/download;Figure;Fig.6.docx

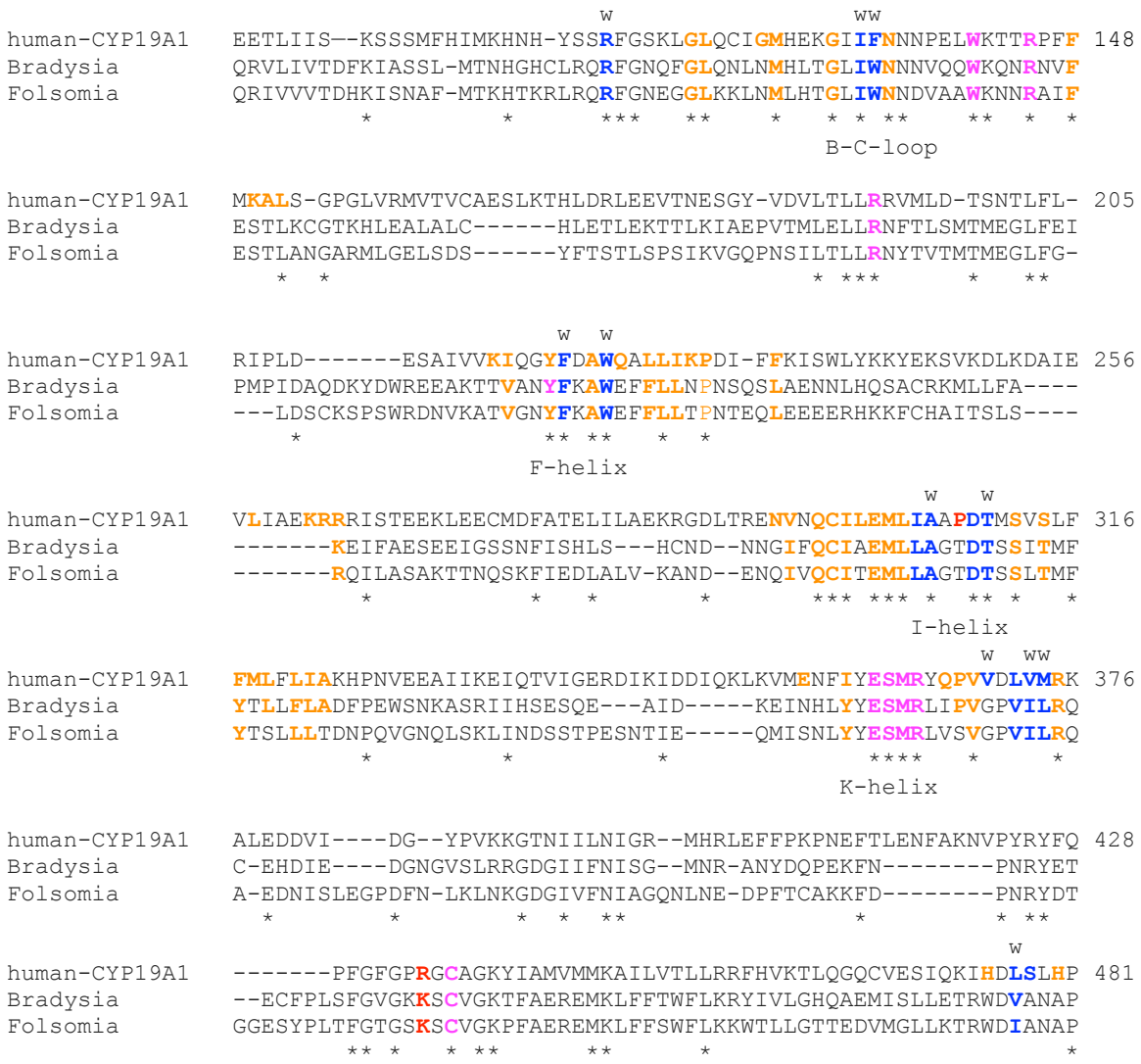

Figure7

Click here to access/download;Figure;Fig7.tiff

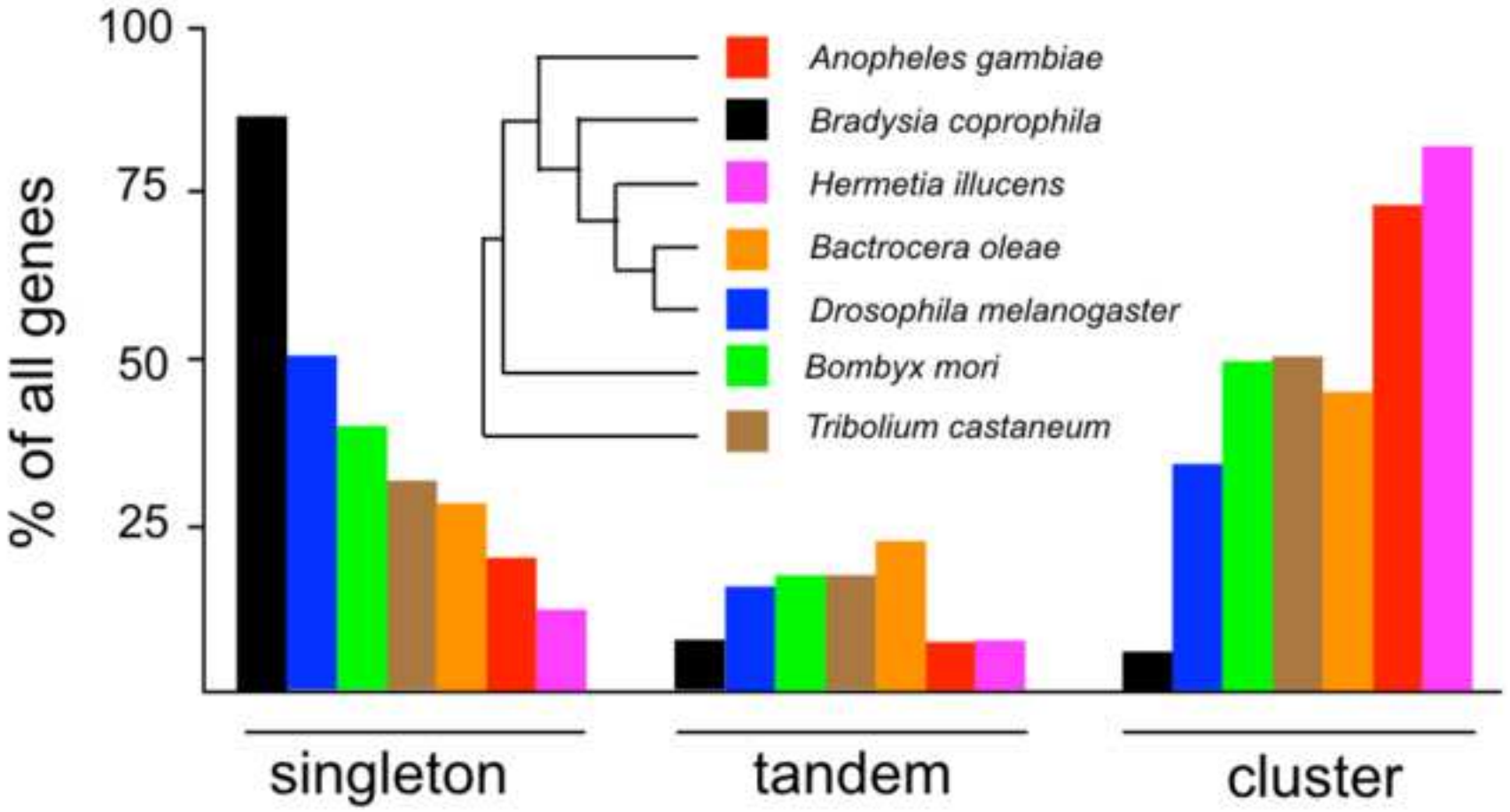

Figure8

Click here to access/download;Figure;Fig8.tiff

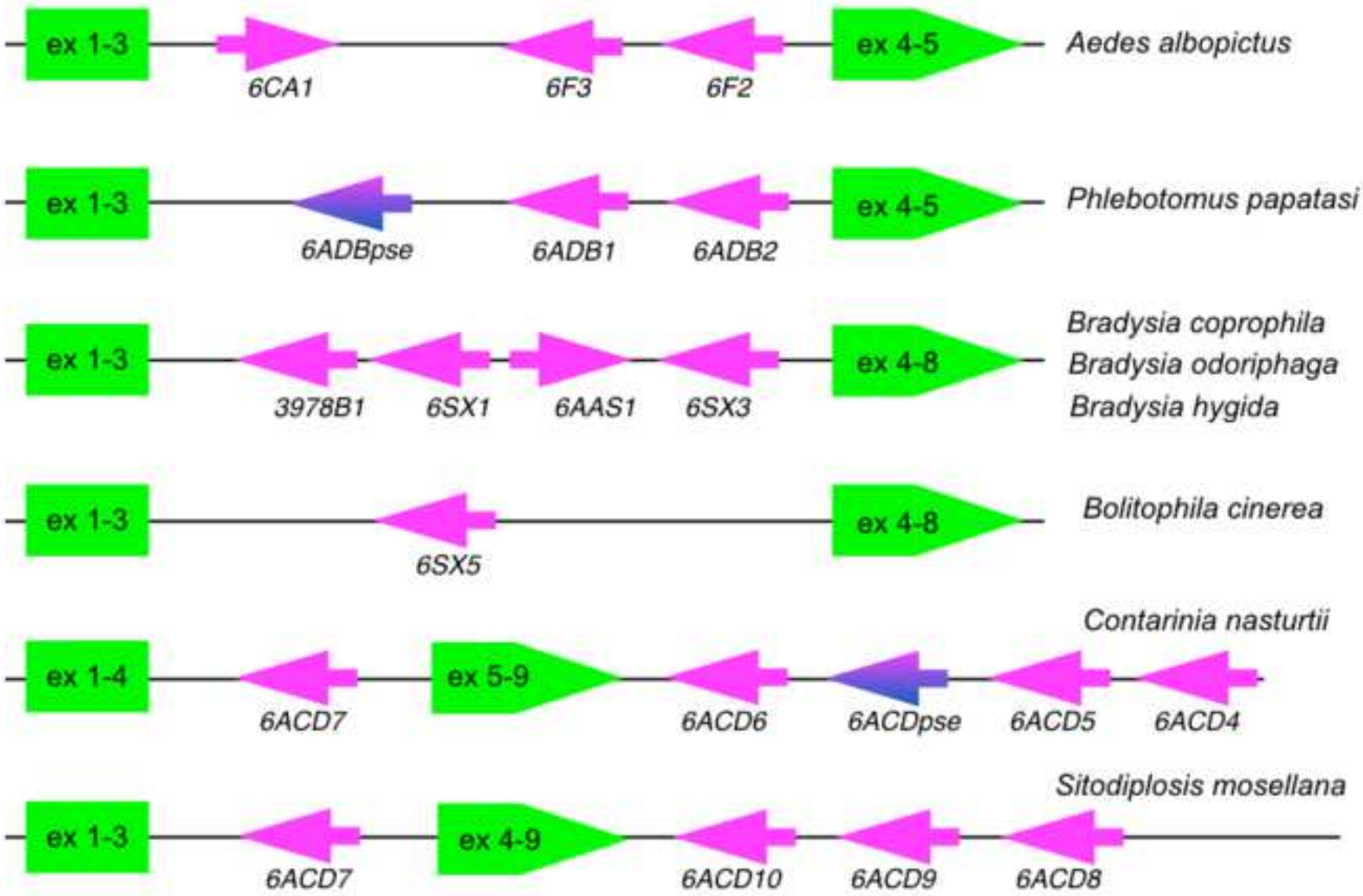

Figure9

Click here to access/download;Figure;Fig9.tiff

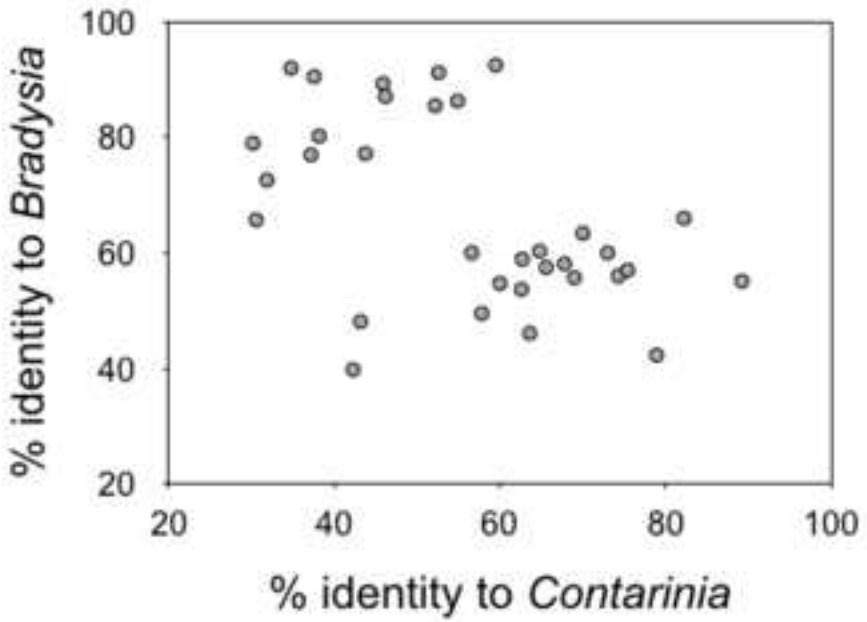

Figure 10

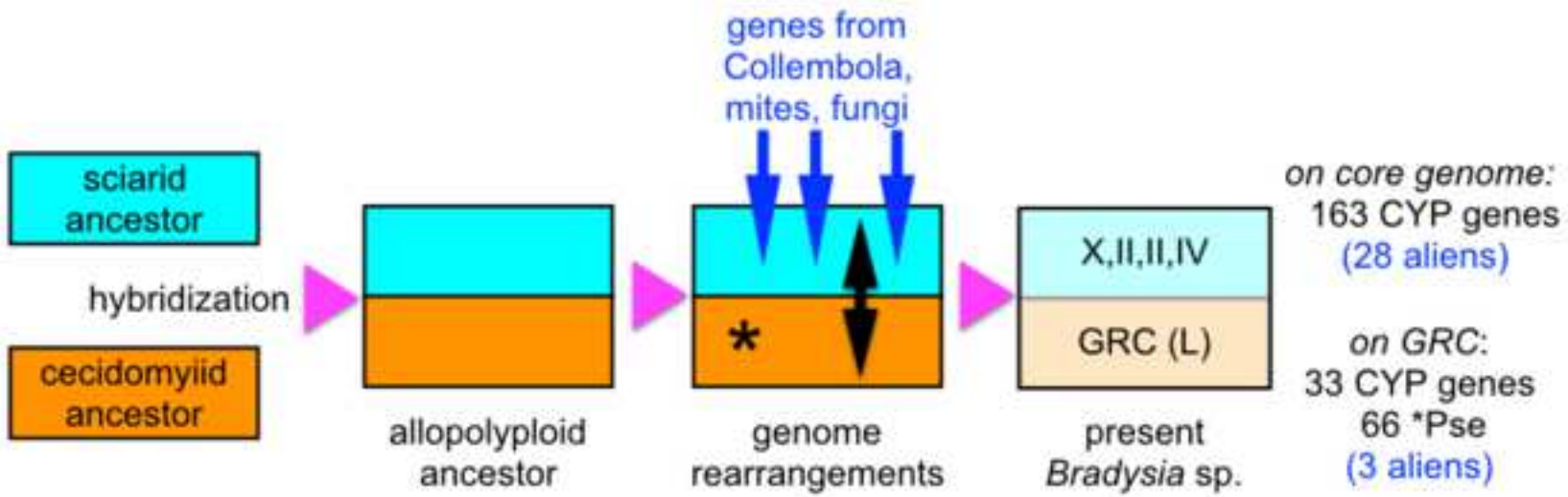

Table 1

Table 1						
CYP name	**CYP clan**	**accession number**	**Transcript in *Bradysia coprophila***	**Ortholog in *B. odoriphaga* (% identity)**	**Ortholog in *B.hygida* (% identity)**	**NCBI blastp NR best hit (% identity)**
CYP53A57	CYP53 clan	XP_037037647.1	yes	KAG4071074.1 (94.4%)	KAJ6633705.1 (partial gene model), but full gene on WJQU01001785.1 plus; transcript GICH01027002.1	*Tothia fuscella* KAF2436574.1(Ascomycota-Pezizomycotina-Dothideomycetes-Pleosporomycetidae-
CYP4724A1	CYP19 clan	XP_037047887.1	yes	KAG4066294.1 (partial gene model); but full gene on JAFDOW010001459.1 plus (92.4%)	not found in genome, no TSA match	*Folsomia candida* (Collembola) XP_021967663.2 (47%)
CYP3709C1	CYP2 clan	XP_037048891.1	yes	not found in genome	not found in genome, no TSA match	*Allacma fusca* (Collembola) CAG7833700.1 (60%)
CYP3710B1	CYP2 clan	XP_037051484.1	yes	KAG4069881.1 (97.3%)	KAJ6649701.1 (84.6%), on chr. A minus	*Folsomia candida* (Collembola) XP_021963340.1 (47%)
CYP3710B2	CYP2 clan	XP_037049572.1	yes	KAG4066248.1 (98.0%)	*paralog* KAJ6649701.1	*Folsomia candida* (Collembola) XP_021963340.1 (47%)
CYP3710B3	CYP2 clan	XP_037025072.1	yes	KAG4069821.1 (94.2%)	*paralog* KAJ6649701.1	*Folsomia candida* (Collembola) XP_021963340.1 (46%)
CYP3710C1	CYP2 clan	XP_037036945.1	yes	no gene model, but full gene on JAFDOW010000093.1 minus (95.1%)	KAJ6636179.1 (partial gene model), but full gene on chr. C minus; transcript GICH01072858.1	*Folsomia candida* (Collembola) XP_021963340.1 (46%)
CYP4721A1	CYP2 clan	XP_037042570.1	no	*paralog* KAG4079466.1	KAJ6646619.1, on chr. A minus; transcript GICH01007255.1 (61.1%)	*Folsomia candida* (Collembola) XP_021950528.2 (40%)
CYP4721A2	CYP2 clan	XP_037042572.1	yes	KAG4079466.1 (90.3%)	*paralog* KAJ6646619.1	*Folsomia candida* (Collembola) XP_021950528.2 (41%)
CYP4721A3	CYP2 clan	XP_037052143.1	yes	KAG4077849.1 (93.8%)	*paralog* KAJ6646619.1	*Folsomia candida* (Collembola) XP_021950528.2 (42%)
CYP4721A4	CYP2 clan	XP_037034543.1	yes	KAG4075367 (partial gene model); but full gene on JAFDOW010000603.1 minus 1719254-1716914 (91.9%)	KAJ6644313.1 (partial gene model); but full gene on chr. B plus. transcript GICH01052778.1	*Folsomia candida* (Collembola) XP_021950528.2 (39%)
CYP4722A1	CYP2 clan	XP_037044115.1	yes	KAG4072777.1 (93.6%)	KAJ6641655.1; chr. B plus; transcript GICH01117116.1	*Folsomia candida* (Collembola) XP_021966906.1 (41%)
CYP3356A1	CYP3 clan	XP_037050080.1	yes	KAG4071301.1 (93.7%)	*paralog* KAJ6635278.1 (59.0%); transcript GICH01030318.1	*Folsomia candida* (Collembola) XP_021962537.1 (45%)
CYP3715C1	CYP3 clan	XP_037045845.1	yes	KAG4076978.1 (95.5%)	KAJ6642747.1; chr. B minus; transcript GICH01005589.1 (76.2%)	*Orchesella cincta* (Collembola) ODM98968.1 (65%)
CYP3717A4	CYP3 clan	XP_037025508.1	yes	KAG4065341m (94.1%)	KAJ6627057.1 (partial gene model); but full gene on WJQU01003173.1 minus; transcript	*Folsomia candida* (Collembola) XP_021965563.1 (55%)
CYP3718B1	CYP3 clan	XP_037036450.1	yes	KAG4076446.1 (97.1%)	not found in genome, no TSA match	*Orchesella cincta* (Collembola) ODM98968.1 (65%)
CYP3828A1	CYP3 clan	XP_037049686.1	yes	KAG4077367.1 (89.8%)	no gene model; but full gene on chr. B plus; transcript GICH01007030.1 (56.8%)	*Allacma fusca* (Collembola) CAG7702444.1 (35%)
CYP4716A1	CYP3 clan	XP_037049098.1	yes	KAG4076059.1 (88.1%)	*paralog* KAJ6641019.1, no TSA match	*Halotydeus destructor* (Acari -Trombidiformes - Prostigmata) KAI1285102.1 (37%)
CYP4716A2	CYP3 clan	XP_037029925.1	yes	KAG4079242.1 (91.1%)	*paralog* KAJ6641019.1, no TSA match	*Halotydeus destructor* (Acari -Trombidiformes - Prostigmata) KAI1298148.1 (35%)
CYP4716A3	CYP3 clan	XP_037049483.1	yes	KAG4073301.1 (90.6%)	*paralog* KAJ6641019.1, no TSA match	*Halotydeus destructor* (Acari -Trombidiformes - Prostigmata) KAI1298148.1 (35%)
CYP4716A4	CYP3 clan	XP_037025913.1	yes	KAG4075516.1 (92.5%)	*paralog* KAJ6641019.1, no TSA match	*Halotydeus destructor* (Acari -Trombidiformes - Prostigmata) KAI1298148.1 (35%)
CYP4716A5	CYP3 clan	XP_037037225.1	no	KAG4075497.1 (partial gene model); but full gene on JAFDOW010000598.1 plus 1252564-12514200	*paralog* KAJ6641019.1, no TSA match	*Halotydeus destructor* (Acari -Trombidiformes - Prostigmata) KAI1298148.1 (35%)
CYP4716A6	CYP3 clan	XP_037035924.1	yes	KAG4075497.1 (partial gene model); but full gene on JAFDOW010000598.1 plus 12512519-12514203 (82.8%)	*paralog* KAJ6641019.1, no TSA match	*Halotydeus destructor* (Acari -Trombidiformes - Prostigmata) KAI1298148.1 (35%)
CYP4717A1P	CYP3 clan	XP_037043806.1	yes, pseudogene	-	-	*Allacma fusca* (Collembola) CAG7784925.1 (60%)
CYP4719A1	CYP3 clan	XP_037047695.1	yes	KAG4066506.1 (93.9%)	*paralog* KAJ6647417.1 (55.3%) transcript GICH01015780.1	*Oppiella nova* (Acari - Sarcoptiformes - Oribatida) CAD7644914.1(47%)
CYP4720A1	CYP3 clan	XP_037052058.1	yes	KAG4074179.1 (partial gene model); but full gene on JAFDOW010000796.1 minus, 1661015-154911 (95.6%)	KAJ6638610.1 (partial gene model); but full gene on chr.X minus; transcript GICH01039913.1	*Tetranychus kanzawai* (Acari - Trombidiformes- Prostigmata) QEV89135.1 (35%)
CYP4714A1	CYP4 clan	XP_037032840.1	yes	no gene model, but full gene on JAFDOW010001076.1 plus 2593699-2595338	KAJ6644108.1 (62.8%) on chr. B; also *paralogs* KAJ6644110.1 and KAJ6644109.1	*Halotydeus destructor* (Acari - Trombidiformes - Prostigmata) KAI1285563.1 (40%)
CYP4714A2	CYP4 clan	XP_037029512.1	yes	no gene model, but full gene on JAFDOW010000598.1 minus (91.5%)	no gene model; but full gene on chr. B plus; transcript GICH01052915.1	*Halotydeus destructor* (Acari - Trombidiformes - Prostigmata) KAI1285563.1 (37%)
on GRC			best match to core genome			
CYP3717A5	CYP3 clan	NODE_2245 plus 112-1853	79.0% match to CYP3717A4	not found in core genome	not found in core genome	*Folsomia candida* (Collembola) XP_021965563.1 (56.1 %)
CYP18A1-like pseudogene	CYP2 clan	NODE_14770 plus 3378-6663	50.4% match to CYP18A1	not found in core genome	not found in core genome	*Orchesella cincta* (Collembola) ODN04952.1 (74%)
CYP3808-like pseudogene	CYP4 clan	NODE_574 plus 7913-10097	31.9% march to CYP4ZM4	not found in core genome	not found in core genome	*Folsomia candida* (Collembola) XP_021962907.1 (44.6 %)