

Research article

Open Access

# The cytochrome P450 (CYP) gene superfamily in *Daphnia pulex*

William S Baldwin\*<sup>1</sup>, Peter B Marko<sup>2</sup> and David R Nelson<sup>3</sup>

Address: <sup>1</sup>Clemson University, Biological Sciences, Institute of Environmental Toxicology, 509 Westinghouse Road, PO Box 709, Pendleton, SC, USA, <sup>2</sup>Clemson University, Biological Sciences, 132 Long Hall, Clemson, SC, USA and <sup>3</sup>University of Tennessee-Memphis, Department of Molecular Sciences, G01 Molecular Science Building, Memphis, TN, USA

Email: William S Baldwin\* - [baldwin@clermson.edu](mailto:baldwin@clermson.edu); Peter B Marko - [pmarko@clermson.edu](mailto:pmarko@clermson.edu); David R Nelson - [dnelson@utmem.edu](mailto:dnelson@utmem.edu)

\* Corresponding author

Published: 21 April 2009

Received: 13 February 2008

BMC Genomics 2009, 10:169 doi:10.1186/1471-2164-10-169

Accepted: 21 April 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/169>

© 2009 Baldwin et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** Cytochrome P450s (CYPs) in animals fall into two categories: those that synthesize or metabolize endogenous molecules and those that interact with exogenous chemicals from the diet or the environment. The latter form a critical component of detoxification systems.

**Results:** Data mining and manual curation of the *Daphnia pulex* genome identified 75 functional CYP genes, and three CYP pseudogenes. These CYPs belong to 4 clans, 13 families, and 19 subfamilies. The CYP 2, 3, 4, and mitochondrial clans are the same four clans found in other sequenced protostome genomes. Comparison of the CYPs from *D. pulex* to the CYPs from insects, vertebrates and sea anemone (*Nematostella vectensis*) show that the CYP2 clan, and to a lesser degree, the CYP4 clan has expanded in *Daphnia pulex*, whereas the CYP3 clan has expanded in insects. However, the expansion of the *Daphnia* CYP2 clan is not as great as the expansion observed in deuterostomes and the nematode *C. elegans*. Mapping of CYP tandem repeat regions demonstrated the unusual expansion of the CYP370 family of the CYP2 clan. The CYP370s are similar to the CYP15s and CYP303s that occur as solo genes in insects, but the CYP370s constitute ~20% of all the CYP genes in *Daphnia pulex*. Lastly, our phylogenetic comparisons provide new insights into the potential origins of otherwise mysterious CYPs such as CYP46 and CYP19 (aromatase).

**Conclusion:** Overall, the cladoceran, *D. pulex* has a wide range of CYPs with the same clans as insects and nematodes, but with distinct changes in the size and composition of each clan.

## Background

The cytochrome P450s (CYPs) have widespread and diverse functions in animals. CYPs in families 1–4 are critical and often inducible components of the phase I detoxification systems of vertebrates, invertebrates, and plants [1-4]. They are also important in lipid metabolism, including fatty acids, retinoids, eicosanoids, steroids, vitamin D and bile acids, and some are regulated in a sexually dimorphic fashion through endocrine mechanisms [5-8].

In addition, the CYP26 family contributes to maintenance of sharp retinoic acid boundaries between rhombomeres in the developing mammalian embryo and in the eye [9-11]. In insects CYPs are essential to the function of sensory organs such as antennae, where they may be involved in odorant clearance [12].

CYPs are important in the metabolism of and tolerance to anthropogenic chemicals [13] and plant allelochemicals



[14,15]. Inducibility, constitutive overexpression, and genomic studies investigating quantitative trait loci (QTL) mapping have demonstrated tolerance or resistance to environmental chemicals due to CYPs [13,16,17]. Furthermore, it has been hypothesized that pesticide sensitivity in honeybee is associated with reduced CYP numbers [18]. In addition, hormone structure is dependent on CYPs [19], and therefore small differences in the hormones utilized (i.e. juvenile hormones, methyl farnesoate) may be dependent on the CYPs available or the timing of expression.

*Daphnia pulex*, commonly known as the waterflea, is the first crustacean to have its genome sequenced. *D. pulex* has a genome of approximately 200 Mb with 31,000 genes <http://wfleabase.org>. Daphnid species are globally distributed zooplankton, in the order Branchiopoda, suborder Cladocera, and there are several daphnid species, including *Daphnia magna*, *Daphnia pulex*, and *Ceriodaphnia dubia*. Daphnids are commonly studied zooplankton because of their importance to aquatic ecosystems, ability to contend with environmental challenges, amenability to culture, short life-cycle, and parthenogenic reproduction. Furthermore, daphnids are commonly used in several toxicity tests for multiple applications. Studies on the *Daphnia pulex* genome and specifically the CYPs may provide important insight on how genome and gene expression alterations promote individual and population-level fitness following environmental change.

It has become clear in the genomic era that previous estimates for the number of CYPs in daphnids such as *D. magna* [20] was a gross underestimate, and that most metazoans have a large number of CYP genes. For example, humans have 57, mice have 102, *Strongylocentrotus purpuratus* (sea urchin) have 120, *Anopheles gambiae* (mosquito) have 106, and *Drosophila melanogaster* have 83 functional CYPs (Additional file 1). In continuation of this P450 research, we are working with the *Daphnia* Genomics Consortium to assemble, annotate, and compare the CYPs in the *Daphnia pulex* genome with several other species. We have also assigned names to all of the CYPs and CYP pseudogenes based on previously published rules using phylogenetic trees and amino acid sequence identity to determine clan, family and subfamily membership [21]. Clans represent the deepest diverging gene clades in the CYP nomenclature. There are ten P450 clans among animals, but only four are present in the protostomes: CYP2, CYP3, CYP4 and mitochondrial. The general web repository for P450 nomenclature and sequence data is <http://drnelson.utmem.edu/cytochromeP450.html>. Overall, this work provides a continuation of earlier projects to comprehensively annotate CYPs and determine the putative role of *Daphnia pulex* CYPs in sensory adaptation, sensitivity to toxicants, adap-

tation to environmental challenges, and as biomarkers of chemical or endocrine stimulation.

## Results and discussion

### Manual Annotation

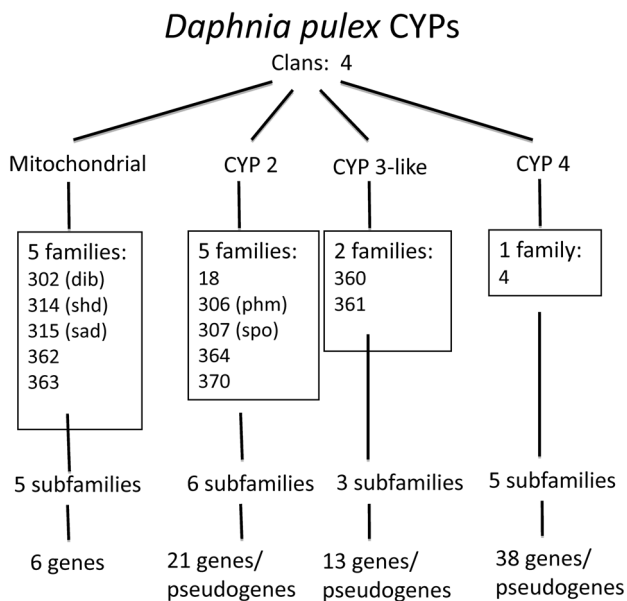
Manual annotation and curation of the CYPs in *Daphnia pulex* v1.1 draft genome sequence assembly (September, 2006) produced 73 full length P450s and three pseudogenes. In addition, there are two ESTs in the CYP4 clan that were not found in the *D. pulex* genome of the chosen parthenogenic individual, but partial sequences (Cyp4C32, Cyp4AN1) were observed previously in specific *D. pulex* ecotypes [15]. Overall, this number is similar to other CYP genomes (Additional file 1), but slightly lower than the typical invertebrate or insect with the exception of honeybee [22].

The *Daphnia pulex* full length CYPs are divided into four distinct clans; mitochondrial, CYP2, CYP3, and CYP4, the same clans that are found in insects [22]. The four clans can be further subdivided into 13 families, 19 subfamilies, and 75 putatively functional isoforms, and 3 pseudogenes (Fig. 1). The phylogenetic tree in combination with the previously described rules for naming CYPs (40% identity for family designation, 55% identity for subfamily designation) were used to name the different CYPs in *D. pulex* [21]. Six new CYP families were discovered; mitochondrial clan families CYP362 and 363, CYP2 clan families CYP364 and 370, and CYP3 clan families CYP360 and 361 (Fig. 1, 2).

### Phylogenetic tree

Bayesian inference and maximum parsimony yielded nearly identical topologies, so we have presented the tree in which branch lengths are proportional to the amount of change occurring in each lineage (Fig. 2). This tree is also available in a detailed, expandable, readable pdf document as supplementary material (Additional file 2). The Bayesian tree includes posterior probabilities at nodes which reflect the proportion of trees sampled during the search that included each particular node. The sea anemone genome has not been manually curated and therefore the data are subject to potential assembly errors until validated. However, the sea anemone provides an ancient anthozoan class cnidarian, and more importantly a diploblast, to the analysis as the only non-triploblastic species in our phylogenetic tree. The honeybee and fruitfly provide a protostome insect CYP genome for comparison, while the human CYPs provide a deuterostome for comparison to the Branchiopod crustacean, *D. pulex*.

In general, the tree shows six distinct monophyletic clades, all but one supported by posterior probabilities of 1.00. These include the mitochondrial, CYP2, CYP3, and CYP4 clans, and two deep branches that do not include



**Figure 1**  
***Daphnia pulex* CYPs are divided into four clans; mitochondrial, CYP2, CYP3, and CYP4.** The four clans can be further subdivided into 13 families, 19 subfamilies, and 78 isoforms, including pseudogenes. The numbers represented above include the pseudogenes, of which there is one pseudogene in each of the CYP2, 3, and 4 clans.

any arthropod CYPs. The non-arthropod lineages are nearly deuterostome exclusive and would be if not for the two anemone CYPs that show their closest relationships to CYP26. These anemone CYPs found nested within a vertebrate lineage indicate that a CYP26-like ancestor may have existed in cnidarians but was lost in protostomes. Similar inferences concerning CYP loss in protostomes have already been made about CYP51, CYP20 and possibly CYP7. The history of CYP20 in protostomes may be complex since an ortholog has been detected in the annelid leech *Haementeria depressa* (CN807321). In addition, CYP19 (aromatase) clusters with the CYP2 and mitochondrial clans, and CYP46 and related sea anemone CYPs are part of a sister clade to the CYP4 clan (Additional file 2).

Overall, the tree demonstrates that the four major clans found in insects (mitochondrial, CYP2, 3, 4) encompass all of the CYPs in *D. pulex*. An Excel table of each of the CYP genes and pseudogenes found in the *Daphnia pulex* genome, their nucleotide and amino acid sequences, and links to their scaffold position is available as supplementary material (Additional file 3).

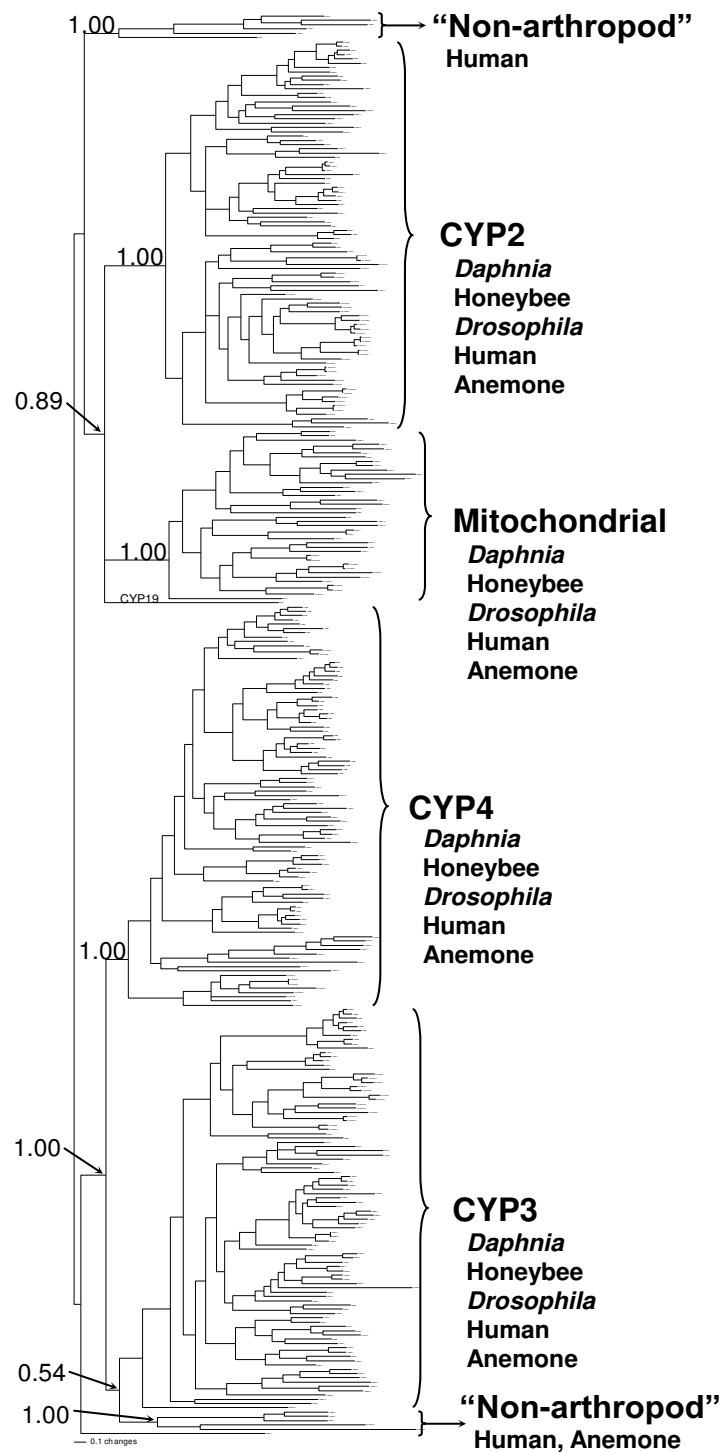
The mitochondrial clan of *D. pulex* contains six members in five families and five subfamilies. Three of the members

are highly conserved Halloween genes involved in ecdysone synthesis [19,23]. Specifically, *disembodied* (Cyp302A1; *dib*), *shade* (Cyp314A1, *shd*), and *shadow* (Cyp315A1; *sad*) are all mitochondrial CYPs involved in the last three steps of 20-hydroxyecdysone (20-HE) synthesis. Cyp314A1 is the CYP required for the conversion of ecdysone to its active form, 20-HE [19,23]. The other three CYPs are divided into two new families CYP362 (Cyp362A1, Cyp362A2) and Cyp363A1, indicating that these CYPs are not as well conserved and probably have taken on new roles. A tree of the mitochondrial CYP clan is available as supplementary material (Additional file 4).

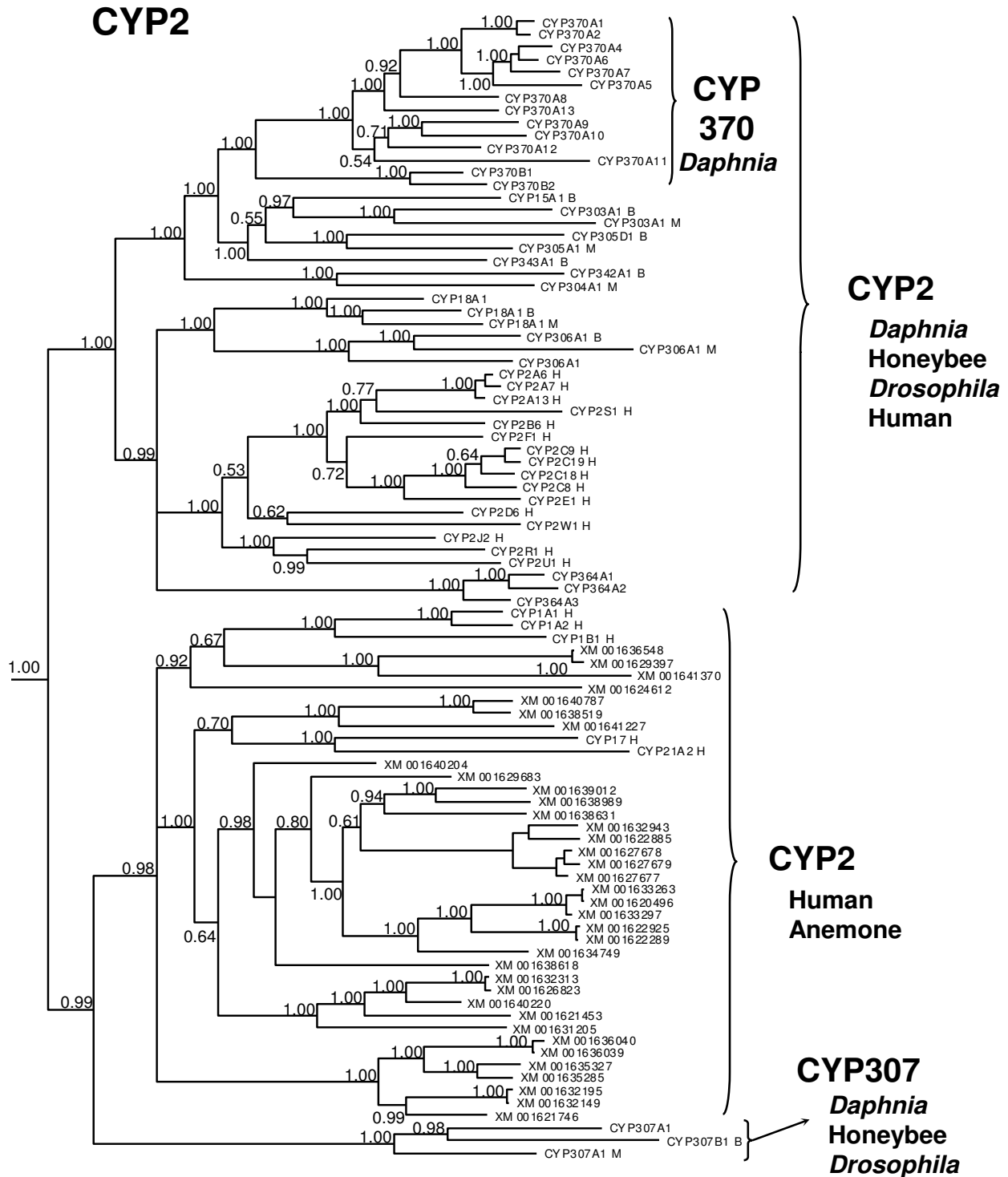
The CYP2 clan of *D. pulex* contains 21 members, including one pseudogene (not shown in Fig. 2, 3), separated into five distinct families (Cyp18, 306, 307, 364, 370) and six subfamilies. Two of the genes, *phantom* (Cyp306A1; *phm*) and *spook* (Cyp307A1; *spo*) are conserved Halloween genes involved in the early stages of ecdysone biosynthesis [19,24]. The exact role of Cyp307A1 in ecdysone biosynthesis is unknown, but appears to be part of the "Black box" of reactions involved in the earlier steps of ecdysone synthesis from cholesterol. *Daphnia pulex* only contains one Cyp307 gene similar to sequenced lepidopterans, but different than *Drosophila* that use two Cyp307 genes at different stages of their lifecycle [24,25]. Interestingly, the Cyp307 genes are a sister group to the non-arthropod CYP2 members (Fig. 3).

The other three CYP2-clan families in *D. pulex* (Cyp18, 364, 370) are divided into four subfamilies and contain 19 genes. Interestingly, a sister-group in the CYP2 clan contains no arthropod CYPs. This group primarily contains anemone CYPs, but also contains CYP17, CYP21, and the CYP1 family members inducible by chlorinated hydrocarbons in vertebrates [26] (Fig. 3; also available as Additional file 5). The Cyp370 family is the largest of the CYP2-clan families, containing 15 members with 13 members in the 370A subfamily and 2 members in the 370B subfamily (Cyp370B1,2). There is also one pseudogene in the Cyp370A subfamily, Cyp370A3P.

The *Daphnia* CYP370 family is greatly expanded relative to the single gene CYP15 and CYP303 families in insects it most closely resembles. Cyp15A1 is a regio- and stereo-specific epoxidase critical in the formation of juvenile hormone III (JH III) from methyl farnesoate in the corpora allata of the cockroach [27]. However, methyl farnesoate, a juvenile hormone precursor, is considered the major terpenoid hormone in crustaceans [28]; therefore, a methyl farnesoate epoxidase is unnecessary and it is unlikely that the CYP370A and CYP370B subfamily members specifically perform this function. The role of the Cyp370 family in *Daphnia* is currently unknown. Cyp18 and Cyp364 are both close relatives of Cyp306 (*phantom*), suggesting



**Figure 2**  
**Phylogenetic relationship of the different CYP clans.** Five CYP genomes were subjected to phylogenetic comparisons using MrBayes. An (H) after the CYP name denotes human sequences, a (B) denotes honeybee sequences, a (M) denotes *Drosophila melanogaster* sequences, CYP names lacking a letter are *D. pulex* sequences, and anemone sequences are noted with their GenBank protein accession numbers (start with XM). Numbers at nodes are posterior probabilities from the Bayesian analysis. Figure 2 is also available in an easily readable, expandable pdf format as a supplementary figure that allows recognition of individual CYP isoforms within each clade (Additional file 2).



**Figure 3**  
**Phylogenetic relationship of the different CYP2 clan members.** Three CYP2 clades are observed; two that contain arthropod CYP2 members and one that does not. In addition, the CYP2 clan is expanded in *Daphnia pulex* because of expansion of the Cyp370 family. An (H) after the CYP name denotes human sequences, a (B) denotes honeybee sequences, a (M) denotes fruitfly sequences, CYP names lacking a letter are *D. pulex* sequences, and anemone sequences are noted with their GenBank protein accession numbers (start with XM). Numbers at nodes are posterior probabilities.

potential involvement in ecdysone synthesis or catabolism (Fig. 3). The Cyp364 family is a new family that contains three genes (Cyp364A1,2,3). Cyp18, which is also found in insects, is induced by 20-HE in *Drosophila* [29].

The CYP3 clan consists of numerous CYPs involved in detoxification of xenobiotics and endobiotics [30-33]. Some CYP3 clan members are inducible by hormones such as progesterone [34] and ecdysone [17], and are responsible for the metabolism and elimination of steroid hormones in vertebrates [20,35,36]. Although the posterior probability at the base of the CYP3 clan is low (0.54), this only reflects the uncertainty of the position of the first two *Drosophila* lineages at the base of the CYP3 clan. The CYP3 clade is strongly supported when *Drosophila* is not included in the analysis (1.00). In *Daphnia pulex*, the CYP3 clan contains 12 genes and one pseudogene, arranged into two new families (Cyp360, Cyp361) and three subfamilies. Eleven of these thirteen genes are in the Cyp360A subfamily leaving just Cyp361A1 and Cyp361B1 outside this subfamily in the Cyp3 clan of *D. pulex*. The closest relatives of the Cyp360 subfamily in the tree are the Cyp6 and Cyp9 subfamily members of insects involved in endobiotic and xenobiotic metabolism and detoxification. Similarly, the closest relatives of the two Cyp361 subfamily members are the anemone CYP3-like group, and the human CYP3A and CYP5A subfamily members involved in detoxification and thromboxane A<sub>2</sub> biosynthesis. In general, the CYP3 clan in insect species has more CYPs than the *Daphnia pulex* CYP3 clan. Most CYP families in insects have only a few members, but the CYP6AS subfamily has 18 members, 37.5% of the honeybee P450s, and there are 35 members of the CYP3 clan distributed between seven families making up 42% of the *D. melanogaster* P450s. The CYP3 clan in insects, and in particular the CYP6AS subfamily in honeybee was recruited for major gene expansion as were the CYP360 and CYP370 families in *D. pulex*, and to a lesser degree the CYP4C subfamily. A tree of the CYP3 clan is available as an additional file (Additional file 6).

The CYP4 clan, which is the sister-group to a clade containing the CYP3 clan plus one of the two "non-arthropod" clans, consists of 38 members all in the same family (Cyp4) and arranged into five subfamilies (Cyp4C, Cyp4AN, Cyp4AP, Cyp4BX, Cyp4BY) with 4-10 members in each subfamily (Additional file 7). There is also a pseudogene in the Cyp4C subfamily. Two members of the Cyp4 family were not observed in the *Daphnia pulex* genome v1.1 draft genome sequence assembly (September, 2006), but were cloned by degenerative PCR in a previous study in which nine CYP4 members were partially cloned [15]. The two absent CYPs, Cyp4C32 (95% identical to 4C34v1; 89% to 4C34v2 from the *Daphnia pulex* genome) and Cyp4AN1 (92% to 4AN2v1; 96% identical

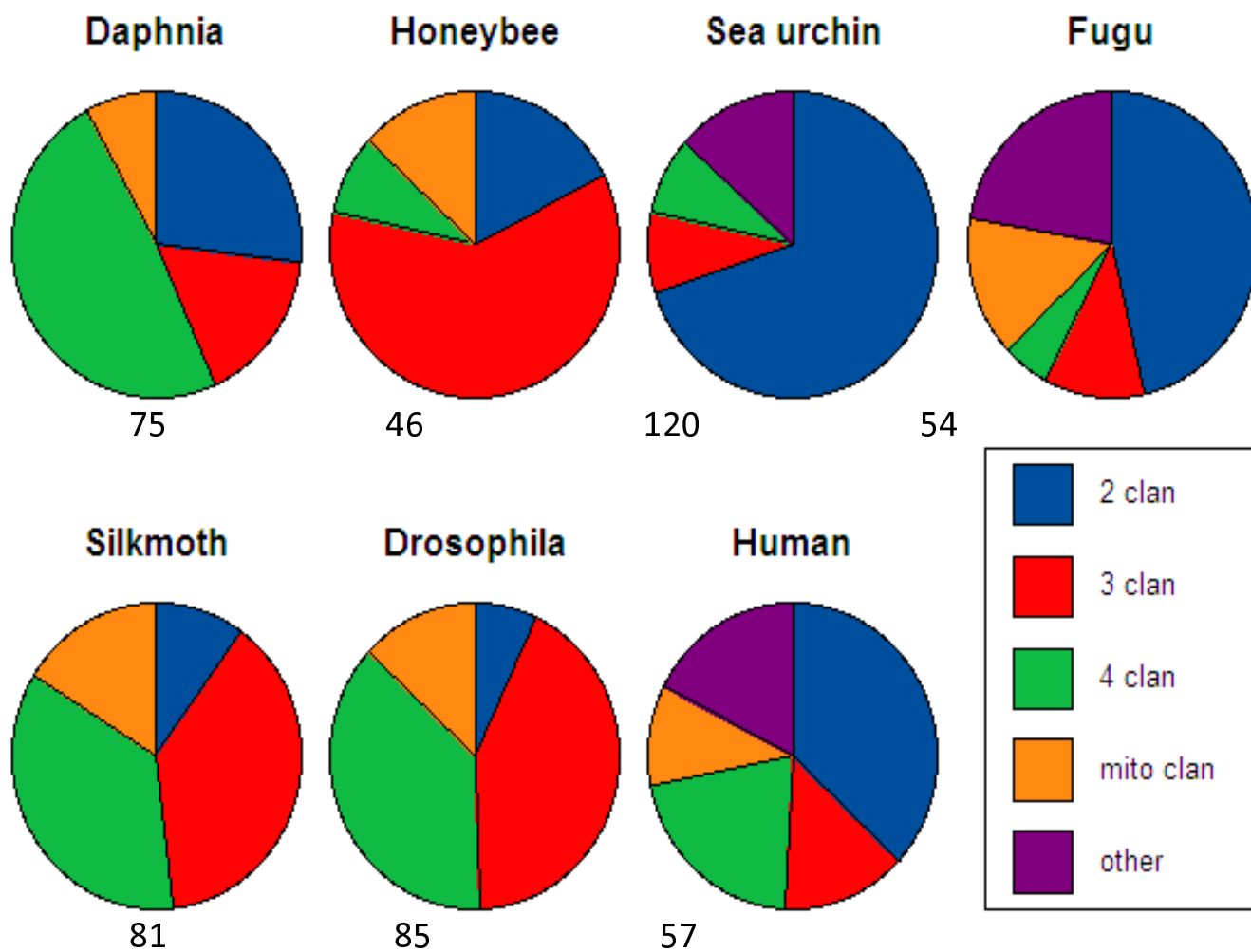
to CYP4AN2v2 from the *Daphnia pulex* genome), are available on GenBank (BQ703381 and BQ703379, respectively). The *D. pulex* genome sequence coverage was 8.7×, therefore our inability to find these two CYPs may be due to the known gaps in the genome assembly. It is also possible that these two genes were deleted from the *Daphnia* Genomics Consortium's chosen parthenogenic *D. pulex* or "chosen one" due to strain differences.

The Cyp4 members are considered the least studied of the CYP clans in insects [22] and are involved in fatty acid metabolism, including inflammatory arachidonic acid metabolites, and xenobiotic metabolism in mammals [37]. Some CYP4 members may be involved in sensory perception in insects as they are found in the antenna [38]. A Cyp4c member is also involved in the biosynthesis of juvenile hormone, and another is inducible by hypertrhalosemic hormone, a key hormone in arthropod carbohydrate metabolism [39]. Furthermore, some Cyp4 members are down-regulated by ecdysteroids [40], indicating that Cyp4 members may play a key role in sensory and hormonal functions in *D. pulex*.

Primarily CYP3, but some CYP4 and mitochondrial clan members have been associated with resistance to pesticides [17,30,41-44]. Several Cyp3 clan members, such as Cyp6g1 and Cyp6a5 are associated with DDT or pyrethroid resistance, respectively [41,44]. Cyp4D10 and other CYP4 members in *Drosophila* are inducible by plant alkaloids and may be important in plant host interactions [45]. In *D. pulex*, differential expression of two Cyp4 genes is associated with resistance to tannic acid and leaf litter [15]. Cyp4C32 expression is much higher in ecotype 1 and Cyp4AP1 shows much higher levels of transcription in ecotype 2, which is exposed to high amounts of leaf litter and polyphenols, and in turn resistant to toxic leaf litters [15]. Interestingly, the CYPs that show differential expression based on ecotype and leaf litter exposure, are the two CYPs that are not found within the *D. pulex* v1.1 genome sequence assembly. The complete sequence of these two CYPs is not available [15].

#### Expansion of the CYP 2 and 4 clans

A comparison of the number of genes in each of the CYP clans from *D. pulex*, silkworm (*Bombyx mori*), human (*Homo sapiens*), pufferfish (*Fugu rubripes*), honeybee (*Apis mellifera*), sea urchin (*Strongylocentrotus purpuratus*), and fruitfly (*Drosophila melanogaster*) indicates both subtle and demonstrative differences in CYP clan numbers between species (Fig. 4). In general, the protostomes (fruitfly, honeybee, *D. pulex*) have significantly fewer CYP2 clan members than deuterostomes, indicating an expansion of the CYP2 clan in deuterostomes (Fig. 3). Of the protostomes investigated, *D. pulex* has the greatest percentage of CYP2 clan members. The CYP2 clan encom-



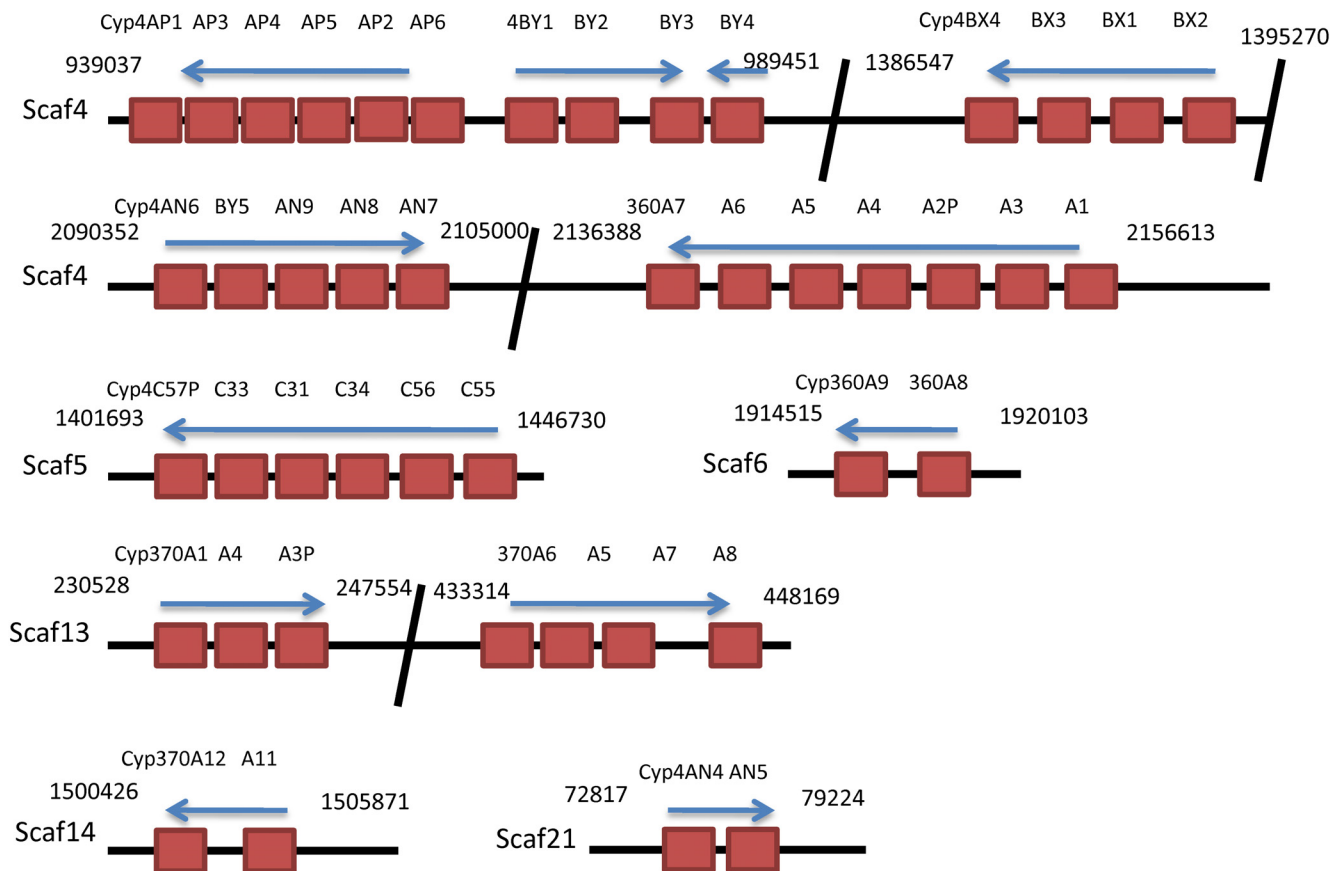
**Figure 4**  
**Comparison of the number of genes in each of the CYP clans from Daphnia, Honeybee, Sea Urchin, Fugu (pufferfish), Silkmoth, Drosophila, and Human genomes.** The data was derived from [22,46-49] and manual curation of the *Daphnia pulex* genome.

passes approximately 5.5–10% of the total CYPs in most insects with honeybee the exception at 17.4% of its CYPs in the CYP2 clan [22]. In contrast, *D. pulex* has 21 CYPs in the CYP2 clan, nearly double of any sequenced insect (11/132 in *Aedes aegypti*), and *D. pulex*'s CYP2 clan members encompass 26.9% of its CYPs, indicating a significant expansion of this clan relative to the insects. No other crustacean has been sequenced, so whether the expansion of the CYP2 clan is typical of crustaceans is unknown.

The CYP4 clan is also slightly expanded in *D. pulex* relative to the insects. There are 38 CYP4 members that encompass 49% of the CYPs in the genome. The CYP4 clan varies from 8.6–42% of the CYP genome in sequenced insects

[22]. Excluding the honeybee, which only has 4 CYP4 members, the rest of the insect's CYP4 members vary from 30.7–42% of the total CYPs. Relative increases in CYP2 and CYP4 members in *D. pulex* leaves a relative reduction in the CYP3 clan compared to insects. Only 16.7% (13/78) of the members of the *D. pulex* CYP genome are CYP3 clan members; whereas 38–61% (28–76) of the insect CYPs are CYP3 clan members.

There are several CYP2-clan members similar in structure to other ecdysone metabolizers (CYP18, CYP364 members). However, the formation of juvenile hormone III from methyl farnesoate by CYP15A1 is unnecessary in crustaceans and *D. pulex* in turn lacks the CYP15A1 gene.

**Figure 5**

**The Cytochrome P450s are often found in tandem repeats on the genome.** Forty-five of the seventy-seven discovered CYPs are found on tandem repeat regions, and an additional CYP (Cyp4C53; 1481785) is found adjacent to the Cyp4C-rich region on scaffold 5 (not shown). The scaffold numbers are presented on the left hand side of the figure and the positions of each tandem repeat on the scaffold are presented before and after each break in the tandem repeat region. Arrows indicate the direction of the repeat regions, and the names of each CYP are provided above each block that represents a single CYP gene.

In addition, *D. pulex* lacks the CYP303 subfamily members with unknown but putative external sensory development function [5]. Nevertheless, the CYP370 family, which is phylogenetically related to the CYP15 and CYP303 families, has expanded dramatically in *D. pulex*, and this family lacks a specific enzyme with a known function. Based on our current knowledge of CYPs, the expansion of the CYP370 family is probably necessary for responses to environmental stressors such as toxicants, and/or other growth or behavioral stimulators such as plant alkaloid toxins.

#### **Tandem repeat regions**

Tandem duplicates are genes that are within intron distances of each other, nearly identical (95+% identity), and may harbor some interesting biology. Gene expansion by tandem duplication is common in P450 evolution but the basis for recruitment of a founder gene is not understood. There are several CYP tandem repeat regions in the *D. pulex* genome (Fig. 5). Overall, 45 of the 77 (58%) CYP genes are located within tandem repeats. Twenty-eight of the 37 (76%) CYP4 family members, including Cyp4C53, are found in tandem repeat regions, which explain in part the expansion of the CYP4 clan in *D. pulex*. In contrast, the CYP3 clan has 69% (9/13) of its genes found in tandem



repeat regions, including 9/11 of the Cyp360 family members. However, the CYP3 clan is small relative to the insect genomes, thus presence in tandem repeats is not correlated to the size of the clan. Lastly, none of the mitochondrial CYPs, many of which have specific functions in ecdysone synthesis [19], are members of a tandem repeat region (Fig. 4). Expansions of conserved endogenous functions are rare and may be selected against. Conversely, tandem duplications may indicate diversity in exogenous xenobiotic substrates [3].

The CYP2 clan is highly expanded relative to the insects, but only 9 of the 21 CYP2 (43%) clan genes are members of a tandem repeat region. Initially, we thought this indicated that tandem repeat regions had little to do with the expansion of the CYP2 clan in *D. pulex*. However, all nine of the tandemly repeated CYP2 clan members are in the CYP370 family. It is interesting that the rest of the CYP2 clan members are not found in tandem repeat regions, and it is tempting to speculate that most *D. pulex* CYP2 clan members, just as many of the mitochondrial CYPs, have highly specific functions such as ecdysone biosynthesis. Needs for P450s are often met by expansion via tandem duplication leading to gene clusters. This suggests that the CYP370 family expanded while under selective pressure. Which genes expand may be independent of clan or family membership and depend on substrate specificity required to cope with a new xenobiotic stress. The *C. elegans* genome has almost half of its P450s in the CYP2 clan, yet it only has one mitochondrial clan member, CYP44. Insects have expanded the CYP3 clan into the large CYP6 and CYP9 families and several spinoff families. Deuterostomes have expanded CYP2 extensively, while *Trichoplax adhaeren* has only one CYP2 clan member (e\_gw1.8.275.1|Triad1 at JGI).

Scaffold 4 is especially rich in tandem repeat regions. It contains 26 of the 44 (59%) CYP genes located in tandem repeats, and 19 of these genes are CYP4 family members. Scaffold 4 contains tandem repeats for CYP4AN, 4AP, 4BX, and 4BY subfamilies. Interestingly, a Cyp4BY subfamily member (Cyp4BY5) is located in the middle of the Cyp4AN subfamily tandem repeat region. This is unusual as most of the genes adjacent to each other in a tandem repeat belong to the same subfamily. As the genes diversify a single gene cluster may contain multiple subfamilies as in the CYP2ABFGST and the CYP4ABXZ gene clusters in mammals [46]. An error could have been made in the annotation of this CYP; however, re-examination of Cyp4BY5 gene provided no evidence of this and this CYP firmly fits in the Cyp4BY subfamily based on identity and phylogenetic status (Fig. 2).

## Conclusion

The cladoceran, *D. pulex* has a wide range of CYPs with the same clans as insects, but with distinct changes in the population of each clan. Of note is the expansion of the newly discovered CYP370 family. Elucidation of the function of the different CYP families and subfamilies will no doubt provide important insight into the ability of *D. pulex* to respond to environmental changes, predator-prey relationships, hormonal changes, plant allelochemicals, and sensory stimuli. Ultimately, the understanding of the evolution of animals will require deciphering the history of the P450s and the *Daphnia* genome may contribute to unraveling molecular aspects of the ecology and physiology of this commonly used crustacean species.

## Methods

### Manual curation

The Joint Genome Institute (JGI) <http://www.jgi.doe.gov/Daphnia/> and wFleaBase <http://wfleabase.org> generated gene models for protein-coding genes in *Daphnia pulex* using multiple algorithms. The models were constructed and filtered in order to reduce redundancy based on homology and EST support, which in turn produced the Dappu v1.1 gene builds (July, 2007). The gene models were directly compared to other genomes such as human, mouse, *Drosophila*, zebrafish (*Danio rerio*), *Xenopus*, and bovine (*Bos taurus*) on the *Daphnia* genome portal. We searched for cytochrome P450 (CYP) gene models using KEGG and KOG pathways in addition to the Advanced Search Options on the JGI *Daphnia pulex* portal, and then manually curated each of the CYPs with the help of the gene models and genome comparisons detailed above. Our manual annotations included corrections to several models based on knowledge of intron-exon boundaries in related genes and BLAST searches as described previously [47], and later assigning gene names based on the homology of *Daphnia* CYPs to other species using defined nomenclature and naming rules for complete genes and pseudogenes [21,46].

### Gross Comparisons and Phylogeny

The different CYP clans and families from *Daphnia pulex* were compared to CYP genomes from human (*Homo sapiens*), honeybee (*Apis mellifera*), fruitfly (*Drosophila melanogaster*), silkworm (*Bombyx mori*), purple sea urchin (*Strongylocentrotus purpuratus*), and pufferfish (*Fugu rubripes*) [22,46-49]. In addition, phylogenetic comparisons of the different full length *Daphnia pulex* CYP genes were performed with full length human (*Homo sapiens*), fruitfly (*Drosophila melanogaster*), and honeybee (*Apis mellifera*), and select starlet sea anemone (*Nematostella vectensis*) CYPs. Honeybee, fruitfly, and human CYPs have been manually curated, but the anemone CYPs are only available through GenBank, <http://www.stellabase.org>, or the Nelson cytochrome P450 webpage [49], and have not

been curated or officially named yet. This genome was chosen as an ancient, diploblast for comparison to protostome (honeybee, fruitfly, *Daphnia*) and deuterostome (human) CYP genomes.

To construct phylogenetic trees, all of the *Daphnia*, human, fruitfly, honeybee, and anemone CYP amino acid sequences were first aligned using default parameters in ClustalX [50]. Trees were constructed using two methods. First, we used maximum parsimony as implemented in PAUP 4.0b10 [51], a method that minimizes the number of evolutionary events but does not use an explicit substitution model. The parsimony tree was based on a heuristic search with 10 random addition sequence replications and tree-bisection-reconnection branch swapping, with gaps treated as missing data. A 50% majority-rule tree was computed for all equally parsimonious trees.

Next, we constructed several trees using Bayesian inference, a probabilistic model-based method of phylogeny reconstruction that is similar to maximum likelihood but which has substantially reduced computation time. Bayesian trees were constructed with MrBayes version 3.1.2 [52] on a computing cluster provided by the Computational Biology Service Unit of Cornell University <http://cbsuapps.tc.cornell.edu/mrbayes.aspx>. We built trees using the "mixed-model" approach in which the Markov chain Monte Carlo sampler explores nine different fixed-rate amino acid substitution models implemented in MrBayes. We used 4 chains with runs of 2.5 million generations with chains sampled every 100 generations and with a burnin of 10,000 trees; the WAG [53] model was selected as the best fitting substitution model by MrBayes. Due to the difficulty in choosing an outgroup for such a diverse and ancient gene family, we elected to present phylogenies with a midpoint rooting in which the root is placed halfway between the two most divergent sequences. Midpoint rooting will accurately determine the root of a phylogeny provided that rates of substitution do not vary across the tree. To examine this assumption of a molecular "clock", we repeated the Bayesian analyses with the added constraint of constant rates of amino acid substitution and rooted the resulting trees with a midpoint rooting. The "clock" constrained trees yielded the same overall topology and root position as the unconstrained analyses, indicating that our application of a midpoint rooting is justified.

### Tandem repeat regions

Automated genome annotation software that relies on alignment with gapping, from BLAST and BLAT thru GeneWise, Exonerate and similar tools, may have problems identifying tandem genes and pseudogenes in areas with highly identical exons [54]. Software may skip over one exon and link to its nearly identical downstream gene

model in tandem repeat regions. The *Daphnia pulex* genome appears rich in tandem genes, and CYPs are often found in tandem repeat regions. Tandemgenes, or 'Tandy', software <http://eugenes.org/gmod/tandy/> was used to address this problem, and potential CYPs found in tandem repeat areas were provided by Don Gilbert, Biology, Indiana University [http://wflbase.org/genome-summaries/gene-duplicates/tdpages/Cytochrome\\_P450\\_15.html](http://wflbase.org/genome-summaries/gene-duplicates/tdpages/Cytochrome_P450_15.html). Areas potentially containing tandem repeats were carefully mined and manually curated using the newly generated models, and maps were made of each of the tandem repeat regions.

### Authors' contributions

WSB conceived and coordinated the study, annotated the CYPs, aligned the sequence data, and was the primary author drafting the manuscript. PBM carried out the phylogenetic analysis and helped draft the manuscript. DRN checked the CYP annotations, named the CYPs, provided valuable direction, insight, and expertise on CYP evolution, and helped draft the manuscript. All authors read and approved the final manuscript.

### Additional material

#### Additional file 1

*Comparison of the number of functional CYP genes in different genomes. A pdf table comparing the number of functional CYP genes from several different organisms.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-169-S1.pdf>]

#### Additional file 2

*Phylogenetic relationship of the different CYP clans. Five CYP genomes were subjected to phylogenetic comparisons using MrBayes. An (H) after the CYP name denotes human sequences, a (B) denotes honeybee sequences, a (M) denotes fruitfly sequences, CYP names lacking a letter are *D. pulex* sequences, and anemone sequences are noted with their GenBank protein accession numbers (start with XM).*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-169-S2.pdf>]

#### Additional file 3

*Detailed summary of each CYP sequence, its position on the *D. pulex* genome, and other potentially helpful information. An Excel workbook (xlsx) table of each of the CYP genes and pseudogenes found in the *Daphnia pulex* genome, their nucleotide and amino acid sequences, links to their scaffold position, manual protein identifying number, and comments about the annotation and/or presence of an expressed sequence tag (EST).*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-169-S3.zip>]

**Additional file 4**

*Phylogenetic tree of the mitochondrial CYP clan members. An (H) after the CYP name denotes human sequences, a (B) denotes honeybee sequences, a (M) denotes fruitfly sequences, CYP names lacking a letter are D. pulex sequences, and anemone sequences are noted with their GenBank protein accession numbers (start with XM).*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-169-S4.pdf>]

**Additional file 5**

*Phylogenetic tree of the CYP2 clan members. An (H) after the CYP name denotes human sequences, a (B) denotes honeybee sequences, a (M) denotes fruitfly sequences, CYP names lacking a letter are D. pulex sequences, and anemone sequences are noted with their GenBank protein accession numbers (start with XM).*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-169-S5.pdf>]

**Additional file 6**

*Phylogenetic tree of the CYP3 clan members. An (H) after the CYP name denotes human sequences, a (B) denotes honeybee sequences, a (M) denotes fruitfly sequences, CYP names lacking a letter are D. pulex sequences, and anemone sequences are noted with their GenBank protein accession numbers (start with XM).*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-169-S6.pdf>]

**Additional file 7**

*Phylogenetic tree of the CYP4 clan members. An (H) after the CYP name denotes human sequences, a (B) denotes honeybee sequences, a (M) denotes fruitfly sequences, CYP names lacking a letter are D. pulex sequences, and anemone sequences are noted with their GenBank protein accession numbers (start with XM).*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-169-S7.pdf>]

**Acknowledgements**

The sequencing and portions of the analyses were performed at the DOE Joint Genome Institute under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231, Los Alamos National Laboratory under Contract No. W-7405-ENG-36 and in collaboration with the Daphnia Genomics Consortium (DGC) <http://daphnia.cgb.indiana.edu>. Additional analyses were performed by wFleaBase, developed at the Genome Informatics Lab of Indiana University with support to Don Gilbert from the National Science Foundation and the National Institutes of Health. Coordination infrastructure for the DGC is provided by The Center for Genomics and Bioinformatics at Indiana University, which is supported in part by the METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc. Our work benefits from, and contributes to the Daphnia Genomics Consortium. This is Technical Contribution No. 5608 of the Clemson Univer-

sity Experiment Station, supported by the CSREES/USDA (project number SC-1700342).

**References**

- Wei P, Zhang J, Egan-Hafley M, Liang S, Moore DD: **The nuclear receptor CAR mediates specific xenobiotic induction of drug metabolism.** *Nature* 2000, **407**:920-923.
- Kang HS, Angers M, Beak JY, Wu X, Gimble JM, Wada T, Xie W, Collins JB, Grissom SF, Jetten AM: **Gene expression profiling reveals a regulatory role for ROR alpha and ROR gamma in phase I and phase II metabolism.** *Physiol Genomics* 2007, **31**:281-294.
- Li X, Schuler MA, Berenbaum MR: **Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics.** *Annu Rev Entomol* 2007, **52**:231-253.
- Estabrook RW: **A passion for P450s (remembrances of the early history of research on cytochrome P450).** *Drug Metab Dispos* 2003, **31**:1461-1473.
- Willingham AT, Keil T: **A tissue specific cytochrome P450 required for the structure and function of Drosophila sensory organs.** *Mech Dev* 2004, **121**:1289-1297.
- Waxman DJ: **Interactions of hepatic cytochromes P-450 with steroid hormones: Regioselectivity and stereoselectivity of steroid metabolism and hormonal regulation of rat P-450 enzyme expression.** *Biochem Pharmacol* 1988, **37**:71-84.
- Waxman DJ, Pampori NA, Ram PA, Agrawal AK, Shapiro BH: **Inter-pulse interval in circulating growth hormone patterns regulates sexually dimorphic expression of hepatic cytochrome P450.** *Proc Natl Acad Sci USA* 1991, **88**:6868-6872.
- Muerhoff AS, Griffin KJ, Johnson EF: **The peroxisome proliferator-activated receptor mediates the induction of CYP4A6, a cytochrome P450 fatty acid omega-hydroxylase, by clofibrate.** *J Biol Chem* 1994, **267**:19051-19053.
- Luo T, Sakai Y, Wagner E, Dräger UC: **Retinoids, eye development, and maturation of visual function.** *J Neurobiol* 2006, **66**:677-686.
- Sirbu IO, Gresh L, Barra J, Duester G: **Shifting boundaries of retinoic acid activity control hindbrain segmental gene expression.** *Development* 2005, **132**:2611-2622.
- Tahayato A, Dollé P, Petkovich M: **Cyp26C1 encodes a novel retinoic acid-metabolizing enzyme expressed in the hindbrain, inner ear, first branchial arch and tooth buds during murine development.** *Gene Expr Patterns* 2003, **3**:449-454.
- Maibèche-Coisne M, Merlin C, François MC, Porcheron P, Jacquinjoly E: **P450 and P450 reductase cDNAs from the moth Mamestra brassicae: cloning and expression patterns in male antennae.** *Gene* 2005, **346**:195-203.
- Wondji CS, Morgan J, Coetzee M, Hunt RH, Steen K, Black WCt, Hemingway J, Ranson H: **Mapping a quantitative trait locus (QTL) conferring pyrethroid resistance in the African malaria vector Anopheles funestus.** *BMC Genomics* 2007, **8**:34.
- Fogleman JC, Danielson PB, MacIntyre RJ: **The molecular basis of adaptation in Drosophila – The role of cytochrome P450s.** *Evol Biol* 1998, **30**:15-77.
- David P, Dauphin-Villemant D, Mesneau A, Meyran C: **Molecular approach to aquatic environmental bioreporting: differential response to environmental inducers of cytochrome P450 monooxygenase genes in the detritivorous subalpine planktonic Crustacea, Daphnia pulex.** *Mol Ecol* 2003, **12**:2473-2481.
- Yang Y, Chen S, Wu S, Yue L, Wu Y: **Constitutive overexpression of multiple cytochrome P450 genes associated with pyrethroid resistance in Helicoverpa armigera.** *J Econ Entomol* 2006, **99**:1784-1789.
- Le Goff G, Hilliou F, Siegfried BD, Boundy S, Wajnberg E, Sofer L, Audant P, French-Constant RH, Feyereisen R: **Xenobiotic response in Drosophila melanogaster: sex dependence of P450 and GST gene induction.** *Insect Biochem Mol Biol* 2006, **36**:674-682.
- Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, Berenbaum MR, Feyereisen R, Oakeshott JG: **A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee.** *Insect Mol Biol* 2006, **15**:615-636.
- Rewitz KF, Rybczynski R, Warren JT, Gilbert LI: **The Halloween genes code for cytochrome P450 enzymes mediating synthesis of the insect moulting hormone.** *Biochem Soc Trans* 2006, **34**:1256-1260.

20. Baldwin WS, LeBlanc GA: **Identification of multiple steroid hydroxylases in *Daphnia magna* and their modulation by xenobiotics.** *Environ Toxicol Chem* 1994, **13**:1013-1021.
21. Nelson DR: **Cytochrome P450 nomenclature, 2004.** *Methods Mol Biol* 2006, **320**:1-10.
22. Feyereisen R: **Evolution of insect P450.** *Biochem Soc Trans* 2006, **34**:1252-1255.
23. Rewitz KF, Gilbert LI: ***Daphnia* Halloween genes that encode cytochrome P450s mediating the synthesis of the arthropod molting hormone: evolutionary implications.** *BMC Evol Biol* 2008, **8**:60.
24. Rewitz KF, O'Connor MB, Gilbert LI: **Molecular evolution of the insect Halloween family of cytochrome P450s: phylogeny, gene organization and functional conservation.** *Insect Biochem Mol Biol* 2007, **37**:741-753.
25. Ono H, Rewitz KF, Shinoda T, Itoyama K, Petryk A, Rybczynski R, Jarcho M, Warren JT, Marques G, Shimell MJ, Gilbert LI, O'Connor MB: **Spook and Spookier code for stage-specific components of the ecdysone biosynthetic pathway in Diptera.** *Dev Biol* 2006, **298**:555-570.
26. Hahn ME: **Aryl hydrocarbon receptors: diversity and evolution.** *Chem Biol Interact* 2002, **141**:131-160.
27. Helvig C, Koener JF, Unnithan GC, Feyereisen R: **CYP15A1, the cytochrome P450 that catalyzes epoxidation of methyl farnesoate to juvenile hormone III in cockroach corpora allata.** *Proc Natl Acad Sci USA* 2004, **101**:4024-4029.
28. LeBlanc GA: **Crustacean endocrine toxicology: a review.** *Eco-toxicology* 2007, **16**:61-81.
29. Bassett MH, McCarthy JL, Waterman MR, Sliter TJ: **Sequence and developmental expression of Cyp18, a member of a new cytochrome P450 family from *Drosophila*.** *Mol Cell Endocrinol* 1997, **131**:39-49.
30. Strode C, Wondji CS, David JP, Hawkes NJ, Lumjuan N, Nelson DR, Drane DR, Karunaratne SH, Hemingway J, Black WC 4th, Ranson H: **Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*.** *Insect Biochem Mol Biol* 2008, **38**:113-123.
31. Kretschmer XC, Baldwin WS: **CAR and PXR: Xenosensors of Endocrine Disrupters?** *Chem-Biol Interact* 2005, **155**:111-128.
32. Waxman DJ: **P450 gene induction by structurally diverse xenobiotics; central role of nuclear receptors CAR, PXR, and PPAR.** *Arch Biochem Biophys* 1999, **369**:11-23.
33. Zhang J, Huang W, Qatanani M, Evans RM, Moore DD: **The constitutive androstane receptor and pregnane X receptor function coordinately to prevent bile acid-induced hepatotoxicity.** *J Biol Chem* 2004, **279**:49517-49522.
34. Masuyama H, Hiramatsu Y, Mizutani Y, Inoshita H, Kudo T: **The expression of pregnane X receptor and its target gene, cytochrome P450 3A1 in perinatal mouse.** *Mol Cell Endocrinol* 2001, **14**:47-56.
35. Waxman DJ, Ko A, Walsh C: **Regioselectivity and stereoselectivity of androgen hydroxylations catalyzed by cytochrome P-450 isozymes purified from phenobarbital-induced rat liver.** *J Biol Chem* 1983, **258**:11937-11947.
36. Acevedo R, Villanueva H, Parnell PG, Chapman LM, Gimenez T, Gray SL, Baldwin WS: **The contribution of hepatic steroid metabolism to serum estradiol and estrion concentrations in nonylphenol treated MMTVneu mice and its potential effects on breast cancer incidence and latency.** *J Appl Toxicol* 2005, **25**:339-353.
37. Simpson AECM: **The cytochrome P450 4 family.** *Gen Pharmacol* 1997, **28**:351-359.
38. Ono H, Ozaka K, Yoshikawa H: **Identification of cytochrome P450 and glutathione S-transferase genes preferentially expressed in chemosensory organs of the swallowtail butterfly, *Pailio xuthus* L.** *Insect Biochem Mol Biol* 2005, **35**:837-846.
39. Bradfield JY, Lee YH, Keeley LL: **Cytochrome P450 family 4 in a cockroach: molecular cloning and regulation by regulation by hypertrehalosemic hormone.** *Proc Natl Acad Sci USA* 1991, **88**:4558-4562.
40. Davies L, Williams DR, Aguiar-Santana IA, Pedersen J, Turner PC, Rees HH: **Expression and down-regulation of cytochrome P450 genes of the CYP4 family by ecdysteroid agonists in *Spodoptera littoralis* and *Drosophila melanogaster*.** *Insect Biochem Mol Biol* 2006, **36**:801-807.
41. Zhu F, Liu N: **Differential expression of CYP6A5 and CYP6A5v2 in pyrethroid-resistant house flies, *Musca domestica*.** *Arch Insect Biochem Physiol* 2007, **67**:107-119.
42. Joußen N, Heckel DG, Haas M, Schuphan I, Schmidt B: **Metabolism of imidacloprid and DDT by P450 CYP6G1 expressed in cell cultures of *Nicotiana tabacum* suggests detoxification of these insecticides in *Cyp6g1*-overexpressing strains of *Drosophila melanogaster*, leading to resistance.** *Pest Manag Sci* 2007, **64**:65-73.
43. Scharf ME, Parimi S, Meinke LJ, Chandler LD, Siegfried BD: **Expression and induction of three family 4 cytochrome P450 (CYP4)\* genes identified from insecticide-resistant and susceptible western corn rootworms, *Diabrotica virgifera virgifera*.** *Insect Mol Biol* 2001, **10**:139-146.
44. Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham P, Feyereisen R, Wilson TG, ffrench-Constant RH: **A single P450 allele associated with insecticide resistance in *Drosophila*.** *Science* 2002, **297**:2253-2256.
45. Danielson PB, Foster JLM, McNahill MM, Smith MK, Fogleman JC: **Induction by alkaloids and phenobarbital of family 4 P450s in *Drosophila*: evidence for host plant utilization.** *Mol Gen Genet* 1998, **259**:54-59.
46. Nelson DM, Zeldin DC, Hoffman SMG, Maltais LJ, Wain HM, Nebert DW: **Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative splice variants.** *Pharmacogenetics* 2004, **14**:1-18.
47. Nelson DR: **Comparison of P450s from human and fugu: 420 million years of vertebrate P450 evolution.** *Arch Biochem Biophys* 2003, **409**:18-24.
48. Goldstone JV, Hamdoun A, Cole BJ, Howard-Ashby M, Nebert DW, Scally M, Dean M, Epel D, Hahn ME, Stegeman J: **The chemical defense: Environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome.** *Dev Biol* 2006, **300**:366-384.
49. Nelson DR: **Cytochrome P450 Homepage.** [<http://drnelson.utm.edu/CytochromeP450.html>].
50. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The clustalx windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **24**:4876-4882.
51. Swofford DL: **PAUP\*: phylogenetic analysis using parsimony (\*and other methods), version 4.0b10.** Sunderland, MA: Sinauer Associates; 2001.
52. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
53. Whelan S, Goldman N: **A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach.** *Mol Biol Evol* 2001, **18**:691-699.
54. Nelson DR: **Frankenstein genes, or the Mad Magazine version of the human pseudogenome.** *Human Genomics* 2004, **1**:310-316.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

