1-2023

# A Bioinformatics Approach for Evaluating Evolutionary Convergence of Gene Family Size in Hematophagous Insects

Mbemba Ceesay

**Abstract**

The act of blood-feeding can be nutritionally rewarding for blood-feeding arthropods. However, blood digestion can release pro-oxidant molecules such as heme and iron at potentially harmful levels. If left uncontrolled, this heme/iron can cause oxidative damage and eventually cell death. This has led to the evolution of various adaptations that protect blood-feeding arthropods against iron- and heme-associated damage. Here I postulate that the signature of this adaptation can be observed in patterns of gene family size. To test this hypothesis, I explore convergent evolutionary expansions and contractions of gene families in distinct lineages of hematophagous insects. Specifically, I compare the gene content present in available genomes from blood-feeding and non-blood feeding arthropods (including outgroup taxa in the Lepidoptera [moths & butterflies]), to identify possible changes in gene family size in the blood-feeding taxa. Of the 206 heme/iron-associated genes identified from the model insect, *Drosophila melanogaster*, five were overrepresented (potentially duplicated) in the blood-feeding taxa: spook (cyp307A1), spookier (cyp307A2), cytochrome P450 12e1 (cyp12e1), hormone receptor and 51 (Hr51), NADH dehydrogenase (ubiquinone) B16.6 subunit (ND-B16.6), and seven were underrepresented (potentially lost). However, when only Dipteran (fly) and Siphonaptera (flea) genomes were included in the analysis, just one iron gene and one heme gene (NADH dehydrogenase (ubiquinone) PDSW subunit (ND-PDSW) were overrepresented in the blood-feeding taxa. Interestingly, the expanded cytochrome genes are known detoxifiers of many compounds, including heme and iron. More broadly, the analytical approach I employee here could be used to evaluate functional convergence for other phenotypic traits, conditional on the availability of annotated genomic data

*Keywords:* evolutionary convergence, hematophagy, Diptera, Lepidoptera, Siphonaptera, gene ontology, gene family expansions, cytochrome P450**,** NADH dehydrogenase (ubiquinone) PDSW subunit (ND-PDSW)

MONTCLAIR STATE UNIVERSITY

# A Bioinformatics Approach for Evaluating Evolutionary Convergence of Gene Family size Expansions in Hematophagous Insects

by

Mbemba Ceesay

A Master's Thesis Submitted to the Faculty of

Montclair State University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

January 2023

College of Science and Mathematics

Department of Biology

Thesis Committee:

Dr. Matthew Aardema

[redacted]

Dr. Julian Keenan

[redacted]ber

Dr. Kevin Bilyk

Committee Member

[redacted]

# A BIOINFORMATICS APPROACH FOR EVALUATING EVOLUTIONARY CONVERGENCE OF GENE FAMILY SIZE IN HEMATOPHAGOUS INSECTS

A THESIS

Submitted in partial fulfillment of the requirements

For the degree of Master of Science

by

Mbemba Ceesay

Montclair State University

Montclair, NJ

2023

Acknowledgements

I begin by offering my profound thanks to my thesis supervisor Dr. Matthew Aardema. His guidance, availability, and expertise made this research project possible. I am immensely grateful to have learned the contents of this research project from his deep insights into the research subject. I am deeply thankful to my committee members, Dr. Julian Keenan and Dr. Kevin Bilyk, for their valuable contributions, which improved this research project. Finally, my immense appreciation goes to my family for their constant support.

## Contents

## List of Tables

### 2 Materials & Methods



### 3 Results

# List of Figures

## 2 Materials & Methods

# 1. Introduction

## 1.1 Evolutionary Convergence

Evolutionary convergence is the independent evolution of related traits in distinct taxa. Organisms with varied phylogenetic evolutionary histories may nonetheless exhibit similar phenotypes due to comparable selective pressures (Losos 2011; Wheeler and Carstens 2018). These selective pressures lead to the evolution of the same adaptive syndromes that allow the performance of a particular task or the adaptation to a new lifestyle (Reich et al. 2003). The independent evolution of eyes in octopi, vertebrates, and spiders is a classic example of convergent evolution. Additionally, flight in insects, birds, and bats also attests to the same phenomenon (Xu et al. 2020; Losos 2011).

Convergence observations may support a hypothesis that natural selection was a driving mechanism of specific traits (Stern 2013, Wheeler and Carstens 2018). However, evolutionary convergence is not limited to adaptation due to similar selective conditions related to natural selection (Losos 2011). Species can also evolve similar phenotypes due to coincidences unrelated to natural selection, thus excluding the occurrence of a particular trait for selective environmental reasons. Secondly, apparent convergence could be due to exaptation. In this case, a change in the function of a trait during evolution may occur if a species adopts a beneficial feature to a particular environment for a different reason, resulting in species sharing similar phenotypes in comparable selective environments but which evolved as an adaptation to different selective pressures (Losos 2011). Finally, constraints that bias the number of possible changes could push distantly related species to evolve convergent traits through nonadaptive processes (Losos 2011; Sackton and Clark 2019). For instance, the evolution of small size might be due to

a particular environment through paedomorphosis and consequently, give rise to additional paedomorphic features such as fewer digit numbers or webbed feet (Losos 2011). Salamanders are an example of this phenomenon. The number of cells in the salamander's limb bud partially determines the number of digits on the limb. Accordingly, evolutionary changes corresponding to a fewer number of cells can lead to decreased body size resulting in digit numbers reducing from five to four (Losos 2011; Alberch and Gale 1985).

**1.2 Measuring Convergent Evolution**

Convergence at the genetic level frequently occurs due to similar genetic modifications. It does so through parallel and collateral genetic evolution. Parallel genetic evolution is the independent evolutionary occurrence of mutations in different species or populations. In contrast, collateral genetic evolution occurs firstly through the evolution of polymorphic alleles within populations with shared ancestry and secondly through introgression, a process in which an allele from one population is introduced into another population through hybridization (Stern 2013). Genetic approaches for examining convergent evolution tend to focus on convergent patterns in biological pathways, gene expression, gene copy number, transposable element content (TE), amino acid (AA) substitutions in specific genes, genome-wide AA substitutions, and/or the percentage of guanines and cytosines (GC content) (Xu et al. 2017).

An example of parallel evolution between populations within the same species: The mouse-ear cress (*Arabidopsis thaliana*), which, during the spring, goes through cold temperature exposure, a requirement to induce flowering in a process called vernalization. However, many populations amongst this species can flower without exposure to cold temperatures and thus forego vernalization. This subpopulation of *Arabidopsis thaliana* has 20 instances of independent

mutations that deactivate the gene responsible for vernalization, namely the FRIGIDA gene. Consequently, 70% of the noticed differences in flowering times are attributable to changes in the FRIGIDA gene (Shindo et al. 2005; Stern 2013).

Mullerian mimicry complexes in *Heliconius* butterflies exemplify collateral evolution through hybridization. In Mullerian mimicry, *Heliconius* butterflies have evolved vibrant wing color patterns of red, orange, black, and white. These color combinations warn predators of the association between the manifested color combinations and unpalatability. (Heliconius Genome Consortium 2012; Stern 2013). Evidence suggests that gene duplication and reductions could arise convergently with far-reaching consequences. Pathways of genes essential to frigid responses and shade avoidance convergently retain duplicated gene copy numbers in various plant lineages. Possessing additional copies of these genes enhances plant survival in cold environments devoid of access to light (Wu et al. 2019). Conversely, convergent reductions in gene copy numbers could attenuate certain features, as is the case for seagrasses that have independently returned to an aquatic environment, consequently losing features such as complex floral structures and stomata (Kuo and Hartog 2007; Xu et al. 2020) Gene duplication and variations in gene copy numbers actively participate in adaptations to environmental challenges (Xu et al., 2017).

**1.3 Blood Feeders**

Estimates suggests there may be more than 1 million contemporary insect and arachnid species on Earth. Of these, ~14,000 have the capability to feed on vertebrate blood (Graça-Souza et al. 2006). For instance, blood-feeding ticks can ingest a volume of blood that is 100 times their pre-meal weight. (Reuben Kaufman 2007). An ability to feed on blood, generally known as

hematophagy, has occurred independently several times during insect evolution, (Graça-Souza et al. 2006), especially among dipterans (flies) which make up the most abundant and diverse order of blood feeders (Dashti et al. 2016). The evolution of blood-feeding may have arisen from an ability to feed on the skin or wastes of dead animals, and/or through pre-adapted piercing mouthparts that then allowed these species to penetrate animal skin to access blood (Graça-Souza et al. 2006).

Hematophagous insects rely on blood-feeding to acquire vital nutrients for development and reproduction (Dalton et al. 2004). While rewarding, the blood-feeding process can be dangerous for hematophagous invertebrates as vertebrates have defensive tendencies against blood-feeders. Accordingly, limiting contact with vertebrates is an essential survival trait for blood-feeders and forces them to ingest large amounts of blood in a single meal (Sterkel et al. 2017). This may expose the blood feeder to high concentrations of specific blood constituents. The chemical composition of dry vertebrate blood is almost 90% proteins with hemoglobin being the most abundant protein (~150mg/ml), then albumin (~50mg/ml) (Sterkel et al. 2017). Blood-feeding insects generally ingest 2 to 100 times their body weight during a single blood meal, and the rich protein meal digestion releases amino acids essential for egg protein, carbohydrate, and lipid synthesis (Whiten et al. 2017). Compared to proteins, carbohydrates, for instance, only account for 0.4% of dry vertebrate blood composition (Sterkel et al. 2017). Although nutritionally beneficial, the blood-feeding process releases heme and iron at potentially harmful levels, requiring hematophagous arthropods to evolve means to limit oxidative damage during digestion (Whiten et al. 2017).

Interestingly, the lamprey demonstrates convergent evolution due to possessing five unique globin genes (Adgb, Gbx, Cygb, Hbs, and Mbs) that convergently evolved from an

identical globin predecessor (Schwarze 2014). Vertebrate hemoglobin comprises two identical alpha subunits (141 residues) and two identical beta subunits (146 residues), each possessing a single heme group (Perutz 1970). The interactive hetero-tetrameric hemoglobin, the most abundant protein in the blood, promotes the transport of $O_2$ in the circulatory system (Brunori 1999; Rohlfing et.al 2016). During the formation of the heme-protein (hemoglobin), its cytotoxic and unstable constituents, free alpha and beta subunits, free-heme, and uncomplexed iron ions, promote free radical aggregation, thus requiring coordination to reduce harmful effects (Baker et al. 2003). In addition to oxygen transport, the globin proteins, including hemoglobin (Hgb), myoglobin (Mgb), neuroglobin (Ngb), cytoglobin (Cygb), globin's E, X, Y, and androglobin (Adgb), are known detoxifiers of reactive oxygen and nitrogen species. They do so through their conserved heme prosthetic group fold that binds reversibly to $O_2$ (Rohlfing et al. 2016).

Heme is composed of an iron atom bonded to four nitrogen atoms belonging to the highly reactive pyrrole ring of protoporphyrin IX (Toh et al. 2010). Metabolic heme and iron are sourced from dietary heme in some hematophagous parasites. Heme binds directly to important molecules by transitory interactions, which generate free radicals. Due to its destabilizing effects, heme promotes cell lysis as it binds to macromolecules. By binding to enzymes and transcriptional factors, heme modulates the transcription and translation of protein synthesis, affecting the regulation of cell development, apoptosis, and signal transduction (Hou et al. 2007; Toh et al. 2010). Heme generates hydroxyl radicals and reactive oxygen species due to the presence of iron via the Fenton reaction or the heme-$H_2O_2$ reaction. The presence of ROS degrades and destroys proteins, lipids, and DNA molecules. Yet, low amounts of ROS are vital for critical biological functions such as the up-regulation of heme synthesis, signal transduction, and enzyme activation in cells (Tsiftsoglou et al. 2016; Toh et al. 2010).

The breakdown of hemoglobin releases large amounts of free heme and free iron. Under favorable conditions, these compounds enable damaging reactions (Everse and Hsia 1997). Although iron is essential for biological redox reactions, the easy conversion between Fe(II) and Fe(III) enhances iron toxicity by producing oxygen-derived radicals and other damaging molecules (Pierre et al. 2002). Knowledge of iron metabolism in humans is extensive compared to insects and less still for blood-feeding insects. Iron metabolism in humans requires multiple functional genes such as iron regulatory proteins, transferrin receptors, transferrin, ferritin, and many more genes, with the majority absent of human homologs in insects (Dashti et al. 2016). Understanding iron metabolism can explain why hematophagous arthropods survive and flourish, given their potentially harmful lifestyle (Galay et al. 2015).

Hematophagous arthropods, over time, have developed immune antioxidant responses to tackle the customary oxidative stress post-blood-feeding. As the first line of defense against heme toxicity after a blood meal, hematophagous insects generate insoluble heme products for excretion with wastes, thus quantitatively expelling most of the heme generated from the blood meal (Sterkel et al. 2017). Additional responses are protein, lipid, and nucleic acid oxidation. Intracellular antioxidants such as Cu, Zn, and Mn superoxide dismutase (SOD) and extracellular antioxidants such as the heme and iron chelator xanthurenic acid relieve oxidative stress after a blood meal (Lima et al. 2012; Whiten et al. 2017) Blood-feeding parasites, for instance, utilize enhanced hemolytic and proteolytic enzymes to lyse and catabolize red blood cells, blood's most significant cellular component (Brindley 1998). They have also developed defenses in the gut lumen, cytoplasm, and midgut epithelial cells (Whiten et al. 2017). Most of the iron absorbed from the blood meal is sourced from heme iron. However, mosquitoes, for instance, prevent iron overload by excreting excess iron, blocking iron absorption, storing absorbed iron in ferritin, and

circulating iron to eggs and other tissues (Zhou et al. 2007). Parasite survival, symbiotic fitness, and successful reproduction all require rigid control of iron metabolism (Dashti et al. 2016).

### 1.4 Genome Evolution and Gene Ontology (GO) in Blood-Feeders

Examining genome sequences of blood-feeding insects may allow a better understanding of the physiological adaptations undergone by hematophagous insects. In the case of disease vector insects, genomic analyses may also allow the identification of novel targets in disease transmission. Additionally, comparing genomic data of hematophagous insects and their non-blood-feeding relatives can offer convincing evidence regarding the physiological consequences of blood-feeding (Dashti et al. 2016).

Using genomic data, it can be challenging to infer a common molecular role based on nucleotide similarity, as orthologous proteins may have very different functions, while evolutionary divergent proteins may be important in the production of similar phenotypic effects (Doolittle 1994). If defined based on function rather than nucleotide sequences, evolutionary syndromes could be a possible solution to the mentioned challenge because they allow genetic convergence to be defined by function. With this approach, convergence can now be determined by functional similarity (Wheeler and Carstens 2018).

The use of gene ontology (GO) coding can effectively measure functional convergence because it allows the pairing of distinct genes to a specific numeric code relying on available experimental data that has determined sequence similarity to gene function (Wheeler and Carstens 2018). GO terms are assigned numerical codes to potential biological activities and ranked based on specificity. Consequently, search terms that are synonymous with a specific function can be used to measure evolutionary convergence between different species.

Furthermore, GO codes allow the selection of relevant functions before comparing genes between species with similar convergent functions (Wheeler and Carstens 2018).

Gene family expansions in insects may be linked to hematophagy, as the expression of additional copies of genes associated with blood-feeding could assist in blood digestion (Freitas and Nery 2020). The evidence gathered from the literature suggests that the blood feast generates toxic amounts of heme and iron, which require proper and timely disposal to avoid harming host cells, proteins, and DNA. This study aims to identify evolutionary convergence within potentially associated gene families that may have expanded during the evolution of the blood-feeding process. Therefore, in this study, I predict that gene families related to the detoxification of heme/iron and implicated in the blood-feeding process may have expanded.

## 2. Materials & Methods

Arthropods that feed on vertebrate blood poses the need to detoxify the excess heme and iron present in the excessive blood feast. The need to detoxify heme and iron served as the basis for this study.

### 2.1 Species list

I generated a varied species list (Table 1) consisting of blood-feeding Diptera (flies) and Siphonaptera (fleas) representing six independent evolutionary lineages of hematophagy and a comparable number of non-blood-feeding Diptera (Figure 1). I also included several species of non-blood-feeding Lepidoptera (an outgroup sister Order to the Diptera/Siphonaptera clade) as a reference group to partially control for evolutionary relatedness within the Diptera and Siphonaptera. Annotated genomes of the dipterans were downloaded

from https://www.ncbi.nlm.nih.gov/assembly/?term=diptera, while the lepidopteran genomes were downloaded from https://www.ncbi.nlm.nih.gov/assembly/?term=lepidoptera. The non-blood-feeding Diptera and Lepidoptera species served as additional controls to account for phylogenetic effects, as closely related species may resemble each other compared to species selected randomly.

**2.2 Step 1: Get Genes Associated with Iron and Heme**

For my study, the model organism *Drosophila melanogaster* served as the reference. This was done because gene functions are well characterized in this species. I used the custom Perl script "Select_word-associated_GO_terms.pl" to locate any GO term that contains the words' heme' or 'iron' from the *D. melanogaster* go-basic.obo database downloaded from http://purl.obolibrary.org/obo/go/go-basic.obo to obtain the *D. melanogaster* associated heme/iron genes. These terms were than manually examined to assess their potential association with heme and/or iron molecular processes in insects.

**2.3 Step 2: Make Query File Once Per Search Term**

Once extracted, I created a query file (a searchable sequence against the blast database) for all the *D. melanogaster*-related heme and iron genes using the heme and iron GO terms generated from the previous step (go_terms_heme.txt and go_terms_iron.txt). The Perl script, “perl get_drosophila_exons_with_GO_term_list.pl” executed with the command line using the command: (perl get_drosophila_exons_with_GO_term_list.pl iron_go_terms.txt gene_association_v2.1.fb all_dmel_exons.fa > dros_iron_query_seqs.fa) and (perl get_drosophila_exons_with_GO_term_list.pl heme_go_terms.txt gene_association_v2.1.fb

all_dmel_exons.fa > dros_heme_query_seqs.fa) created the query sequences file, once for each search term. This step associated the heme and iron genes from the previous step with their reference *D. melanogaster* genes. I used the program samtools-1.14 https://www.htslib.org/download/) to process the sequence alignments between the heme and iron genes and the reference *D. melanogaster* genes within the Perl script. This step gathered all of the iron and heme genes that match their orthologs in the *D. melanogaster* exons as query files (dros_heme_query_seqs.fa, dros_iron_query_seqs.fa), for subsequent use with the Basic Local Alignment Search Tool (‘BLAST).

**2.4 Step 3a: Make Blast Database (Once per Species)**

BLAST aligns and compares a query DNA sequence with a sequence database. The matches generated from BLAST are ranked based on statistical significance (Altschul et al. 1990). Because BLAST requires a query sequence and a database to search against, the query sequences obtained from step 2 were searched against the databases generated in step 3a. Blast version: ncbi-blast-2.12.0+ was downloaded from https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download. Once downloaded, ncbi-blast-2.12.0+ created databases for each species with the command line using, the command: ~/bin/ncbi-blast-2.12.0+/bin/makeblastdb -in GCA_020740445.1_IBAB_Cspp_1.0_genomic.fna -dbtype nucl. Each database created identifies with the appropriate species accession number. Step 3a generated databases for each species’ genome, the second component required to conduct a blast search.

**2.5 Step 3b: TBLASTN against each species' genome sequence, once per species**

The BLAST tool ‘tblastn’ aligns protein sequences (query) to a translated nucleotide database (Gertz et al. 2006). Using tblastn, I searched amino acid sequences from *D. melanogaster* against the sequences of each taxon. The availability of species databases and search term query sequences from the previous steps made the blast searches possible. I searched *D. melanogaster* genes related to heme and iron query sequences against the genomic database of each species with the command line using the command: ~/bin/ncbi-blast-2.12.0+/bin/tblastn -query dros_iron_query_seqs.fa -db GCA_020740445.1_IBAB_Cspp_1.0_genomic.fna -outfmt 6 -evalue 1e-10> blast_out_iron_culicoides-tainanus.txt. And for heme, the command: ~/bin/ncbi-blast-2.12.0+/bin/tblastn -query dros_heme_query_seqs.fa -db GCA_020740445.1_IBAB_Cspp_1.0_genomic.fna -outfmt 6 -evalue 1e-10 > blast_out_heme_culicoides-tainanus.txt. The parameters are set to format tblastn outputs in a tabular format (-outfmt 6) and to filter out statistically significant tblastn matches (<1e-10). Individual blast outputs for each species (Diptera, Siphonaptera, and Lepidoptera) were generated from the above commands, highlighting the total number of *D. melanogaster*-related heme and iron genes and their total gene copy numbers. I assembled each species' blast results in a single tab-delimited excel file for further processing.

**2.6 Step 4: Calculate the average number of blast hits per gene, once per species**

The BLAST output with the *D. melanogaster-related* iron and heme genes and their copy numbers was further processed to obtain the average number of blast hits per gene (average gene copy numbers) using the Perl script "perl calculate_average_gene_blast_hit.pl". The Perl script, executed with the command line using the command: perl calculate_average_gene_blast_hit.pl blast_out_culicoides-tainanus.txt. 50 40 gene_list.txt > average_irongene_culicoides-

tainanus.txt. And perl calculate_average_gene_blast_hit.pl blast_out_culicoides-tainanus.txt. 50 40 gene_list.txt > average_hemegene_culicoides-tainanus.txt. I defined the blast search criteria as 50% minimum sequence similarity and 40 base pairs for sequence length to capture *D. melanogaster-related* heme and iron genes with 50% sequence similarity and at leaset 40 base pairs in length. This step gathered all the iron and heme genes within the study taxa genomes that satisfied these criteria.

**2.7 Statistical Analysis**

I assessed statically significant differences in gene family expansions or contractions between the blood-feeders Dipterans, non-blood-feeding dipterans, and the non-blood-feeding Lepidopterans. T-tests with the Perl script "calculate_t_test.pl" were conducted relying on the differences in "phenotypes," namely "Blood" and "No," to distinguish the blood-feeders from the non-blood-feeders. P-values below 0.05 were considered statically significant and representative of gene family expansions or contractions within the Diptera Siphonaptera and Lepidoptera taxa.

Additional analysis excluded the outgroup Lepidopterans and assessed gene family convergent expansions and contractions in the Dipteran blood and non-blood-feeders and the Siphonaptera blood-feeder. This analysis generated results that can be compared to the previous analysis that accounted for both the ingroups (Diptera & Siphonaptera) and the outgroup (Lepidoptera) to assess convergent gene family expansions and contractions within the ingroup. T-tests with the Perl script "calculate_t_test.pl" were conducted relying on the differences in "phenotypes," namely "Blood" and "No," to distinguish the blood-feeders from the non-blood-feeders. P-values below 0.05 were considered statically significant and representative of gene family expansions or contractions within the Diptera taxa.

I examined convergence using the average gene family copies of the blood-feeders and the non-blood feeders with p-values below 0.05. The selection of gene families based on appropriate p-value scores allows comparisons between average gene family copy numbers of the blood-feeding and non-blood-feeding taxa. The group with the higher average gene copy number is considered to have expanded or contracted convergently due to the blood-feeding lifestyle or the absence of blood-feeding.

## 3. Results

This study identified four hundred sixty-seven iron-associated GO terms and eighty-five heme-associated GO terms. An assessment to identify all heme and iron genes in all species and their copy numbers yielded 206 annotated genes in *D. melanogaster* that have possible relations to iron and heme. The gene copy numbers varied amongst the species and ranged from 0 to 218 copies.

### 3.1 Iron gene family Expansions and Contractions in the Dipteran, Siphonapteran and Lepidopteran Taxa

Regarding gene family expansions and contractions in the Dipteran, Siphonaptera and Lepidopteran taxa, a t-test (Table 2) comparison of hematophagous Diptera vs. non-hematophagous (Diptera & Lepidoptera) taxa revealed five iron gene families ($p < 0.05$) that were expanded in the hematophagous taxa and seven ($p > 0.05$) that contracted in the hematophagous taxa. The iron gene families with statically significant p-values favoring the hematophagous taxa and indicating gene family expansions (Table 3) are spook (Spo) ($p = 0.01233$), Cytochrome P450 12e1 (Cyp12e1) ($q = 0.01364$), hormone receptor 51 (Hr51) ($p =$

0.0205), spookier (Spok) (p = 0.02546), and the NADH dehydrogenase (ubiquinone) B16.6 subunit (ND-B16.6) (p = 0.03154).

The same t-test (Table 3) highlighted cardinal (Cd) (p = 0.003787), Curly su (Cysu) (p = 0.008056), Heavy metal tolerance factor 1(Hmt-1) (p = 0.01853), Peroxinectin-like (Pxt) (p = 0.03137), NADH dehydrogenase (ubiquinone) 13 kDa A subunit (ND-13A) (p = 0.03277), Nitric oxide synthase (Nos) (p = 0.04328), and Cytochrome P450 4d1 (Cyp 4d1) (p = 0.04479) as gene families that favored the non-hematophagous taxa thus indicative of contraction.

### 3.2 Heme gene family Expansions and Contractions in the Dipteran, Siphonapteran and Lepidopteran Taxa

The t-test analysis for heme gene family expansions and contractions in blood-feeding (Diptera and Siphonaptera) and non-blood-feeding (Diptera and Lepidoptera) yielded (table 4) spook (p=0.01233), Cytochrome P450 12e1 (p=0.01364), hormone receptor 51 (p= 0.0205), spookier(p=0.02546), and the NADH dehydrogenase (ubiquinone) B16.6 subunit(p=0.03154) as expanded (Table 5).

The same t-test found cardinal (p=0.003787), Curly su (p=0.008056), Heavy metal tolerance factor 1(q=0.01853), Peroxinectin-like (q=0.03137), NADH dehydrogenase (ubiquinone) 13 kDa A subunit(q=0.03277), Nitric oxide synthase(q=0.04328), and Cytochrome P450 4d1 (q=0.04479) as heme gene families that contracted (Table 5).

### 3.3 Iron and Heme Gene Family Expansions in Diptera and Siphonaptera

To further investigate the iron related gene family contractions and expansions in Dipteran and Siphonaptera hematophagous, and non-hematophagous Dipteran taxa, a t-test was conducted,

(Table 6) and it yielded ND-PDSW (q = 0.04776) as the only gene family that expanded (Table 7). The remaining six gene families namely, Cyp4d1(q = 0.008542), Cyp6a16 (q=0.01446), Cyp6a17 (q = 0.03699), Cyp6a23 (q = 0.03714), ND-B14.5A (q = 0.04678), and Cyp6t3 (q = 0.04681) contracted.

Averages copies of each gene family were noted for the hematophagous and non-hematophagous Diptera taxa (table 7), and these averages were used to establish gene family contractions and expansions.

In this analysis (Table 8), a t-test comparison of hematophagous Diptera and Siphonaptera and non-hematophagous Diptera taxa as a group, ND-PDSW (q = 0.04776), was noted as the only heme gene family that expanded (Table 9). The remaining six gene families namely, Cyp4d1(q = 0.008561), Cyp6a16 (q = 0.01446), Cyp6a17 (q = 0.03699), Cyp6a23 (q = 0.03714), ND-B14.5A (q = 0.04678), and Cyp6t3 (q = 0.04681) contracted (Table 9).

## 4. Discussion

This study was designed to identify convergent evolutionary expansions within gene families that may have occurred during the evolution of hematophagy in dipterans. Accordingly, it was hypothesized that genes implicated in heme/iron detoxification during blood digestion may have expanded to limit oxidative damage during digestion, and additional copies of genes associated with hematophagy could assist in blood digestion.

Notably, in the analysis of heme and iron-related gene family expansions in the Diptera, Siphonaptera and Lepidopteran taxa, the results identified potential expansions in the following protein-coding genes: spook (cyp307A1), spookier (cyp307A2), cytochrome P450 12e1 (cyp12e1), hormone receptor and 51 (Hr51), NADH dehydrogenase (ubiquinone) B16.6 subunit

(ND-B16.6). When assessing iron and heme gene family expansions in the Diptera and Siphonaptera taxa alone, NADH dehydrogenase (ubiquinone) PDSW subunit (ND-PDSW) expanded.

The spook and spookier genes code for cytochrome P450 enzymes and are part of the more prominent cytochrome P450 gene family. Spook precisely retains distinct sequences found in P450 proteins, including the heme-binding domain (PFxxGXRxCxG), and it also plays a role in insect molting (Ono et al. 2006). The presence of the conserved heme-binding domain suggests a possible role of spook and other cytochrome P450s in detoxifying heme. Additionally, cytochrome enzymes are known detoxifiers of endogenous and exogenous compounds through oxidation reactions (Li et al. 2007). Experimental data are needed to verify the role of cytochrome P450 proteins in detoxifying excess heme after a blood meal in the taxa represented in this study. However, it is reasonable to suppose a detoxification role since these proteins are known to carry out a diverse and often surprising number of oxidation reactions to expunge many compounds (Werck-Reichhart and Feyereisen 2000). Furthermore, due to the presence of the conserved heme-binding domain, it possibly expels excess heme after a blood meal in the hematophagous taxa.

The expansion and diversity of detoxification genes probably enhanced the adaptation of insects to their environments and, in recent evolutionary times, has helped them endure numerous artificial chemicals (Strode et al. 2007). This study found the NADH dehydrogenase (ubiquinone) B16.6 subunit (ND-B16.6) gene family to have potentially expanded. These proteins comprise the biggest enzyme complex in the mitochondrial electron transfer chain and facilitate electron transfers from NADH to ubiquinone. Its two functional arms, NADH and ubiquinone, are involved in oxidation and reduction reactions (Agip et al. 2019). There seems to

be a link between cytochrome P450s and NADH dehydrogenase enzymes in targeted defenses against oxidative agents. Cytochrome P450 proteins are known to utilize NADPH-linked electrons to produce molecular oxygen, which activates specific oxidative attacks against many substrates (Werck-Reichhart and Feyereisen 2000). A negative consequence of the blood meal is the generation of reactive oxygen species and general oxidative stress upon the host, thus requiring the elimination of these substances. Interactions between cytochrome P450s and NADPH dehydrogenases could offer a possible means of eliminating deleterious oxidative stress post blood meals in hematophagous arthropods.

Further evidence suggests the interactions between heme iron, cytochrome P450s, and NADPH in lowering oxidative stress. The convergence of cytochrome P450 genes and NADPH dehydrogenase genes amongst the hematophagous taxa indicates the functional interdependence of these genes to oppose oxidative stress resulting from the blood meal. In the presence of cytochrome P450, methemoglobin can be reduced by NADPH (Faassen et al. 1988). Methemoglobin is the oxidized form of hemoglobin through a spontaneous oxidation reaction that produces reactive oxygen species such as superoxides and hydrogen peroxide (Everse and Hsia 1997). Hematophagous arthropods may have retained these genes collectively due to the increased need to mitigate the harm resulting from the blood meal, given the large amounts of hemoglobin ingested. Consistent with the findings of this study, heme, although known for promoting free radicals, is an integral feature of proteins such as cytochromes, electron transport enzymes, cellular respiration, signal transduction, and detoxification (Hamza and Dailey 2012). The expanded Hr51 gene is a heme-binding protein (de Rosny et al. 2008).

This study found three proteins within the larger cytochrome P450 family that expanded. However, in the Diptera, Siphonaptera, and Lepidopteran taxa analysis, cytochrome P450 4d1

(Cyp4d1) contracted and had higher gene copies in the non-blood-feeding taxa than in blood-feeders. Gene contractions and expansions may suggest adapting to a new lifestyle (Feritas & Nery 2020). Gene contractions were noted in cardinal, curly su (Cysu), heavy metal tolerance factor 1 (Hmt-1), peroxinectin-like (Pxt), NADH dehydrogenase (ubiquinone) 13 kDa A subunit (ND-13A), nitric oxide synthase (Nos), and Cy4d1.

In the analysis of the hematophagous and non-hematophagous (Diptera) taxa and the hematophagous (Siphonaptera) taxa, five genes contracted (Cytochrome P450 4d1, Cytochrome P450 6a16, Cytochrome P450 6a17, Cytochrome P450 6a23, NADH dehydrogenase (ubiquinone) B14.5 A subunit and Cytochrome P450 6t3). These gene contractions may be due to adaptations to new lifestyles that find these genes energetically costly to retain, and therefore mutations that inhibit gene transcription may be advantageous, resulting in gene losses.

The methods of this paper are not limited to the represented study, extending to other comparable analyses, given the availability of genomic data and a " functional syndrome" that differs between closely related species. Such an approach would yield results that segregate and quantify genes implicated in the determined "functional syndrome."

The current use of pesticides to target disease vector insects threatens insects such as bees and other pollinators essential for global food production. Even at sub-lethal doses, pesticide exposure could make bees susceptible to pathogens and parasites and possibly a significant contributor to the global mortality of honeybee colonies (Pettis et al. 2012). Future conservation strategies could determine the genes involved in detoxifying the blood meal, such as cytochrome P450s, as these genes are known detoxifiers of many substances, including pesticides.

Interestingly the *A. mellifera* (honeybee) has a ten-fold decrease in xenobiotic detoxification genes such as cytochrome P450s compared to the hematophagous *A. gambiae*

(malaria mosquito). The significant decrease in these detoxification genes amongst bees could explain their sensitivity to pesticides (Claudianos et al. 2006) Approaches that block transcription of the expanded genes in this study offer a viable and environmentally friendly alternative to fight disease vector insects. A thorough study of these genes could help develop a class of pesticides specific to disease-causing insects, thus limiting the indiscriminate attack posed by pesticides on beneficial insects while also fighting against pesticide resistance.

Additionally, selective inhibition of heme and iron degradation pathways would neutralize disease-transmitting insects, allowing for the invention of novel compounds capable of targeted neutralization of disease vector insects. Indeed (Sterkel, Perdomo, et al. 2016) demonstrated that inhibition of the first and second enzymes in the phenylalanine/tyrosine degradation pathway, namely 4-hydroxyphenylpyruvate dioxygenase (HPPD) and tyrosine aminotransferase (TAT) with mesotrione an HPPD inhibitor caused the death of hematophagous insets after a blood meal, however, the non-hematophagous arthropods remained unaffected. Compared to the neurotoxic insecticides in current use, this less-toxic approach could bring about diseases vector insect controls that are less toxic to humans and the environment.

Predictably as taxa converge on similar phenotypes, functional genomic convergence transpires, as the results of this study indicate. This study discovered genes that expanded due to the blood-feeding lifestyle amongst dipterans. However, not all of the statistically significant genes expanded. There were also observed contractions in some gene families.

This study was limited by the absence of available annotated GO terms for hematophagous arthropods. *D. melanogaster* served as the reference species. However, a practical model organism*, D. melanogaster*, does not feed on blood, stressing the need for experimental assembly of GO terms for hematophagous arthropods. As GO terms of

hematophagous arthropods become available, greater accuracy in assessing the genes implicated in the blood-feeding process would result in studies that follow this methodology. The expanded gene sequences in this study could be compared amongst the hematophagous and non-hematophagous taxa to observe possible convergent mutation patterns. Future studies evaluating the role of iron and heme in hematophagous arthropods could generate knockouts of genes such as cytochrome P450s in hematophagous arthropods and compare their ability to detoxify excess heme and iron to their wild-type counterparts.

## 5. Conclusion

In conclusion, this study underlines the importance of genomic data in assessing evolutionary convergence within gene families that may have expanded resulting from the blood-feeding lifestyle amongst hematophagous insects. Notably, in the analysis of the Dipteran, Siphonapteran, and Lepidopteran taxa, spook (cyp307A1), spookier (cyp307A2), cytochrome P450 12e1 (cyp12e1), hormone receptor and 51 (Hr51), NADH dehydrogenase (ubiquinone) B16.6 subunit (ND-B16.6) possibly expanded due to hematophagy. NADH dehydrogenase (ubiquinone) PDSW subunit (ND-PDSW) is the only potential gene expansion in the Diptera and Siphonaptera taxa analysis. The approach presented in this study represents a promising approach for genomic studies aimed at discovering convergent gene family expansions in taxa with an established functional syndrome, available genomic data, and appropriate control taxa.

**Table 1**: Species list with unique sequence identifier (Accession Number).

| Scientific Name | Order | Family | Phenotype | Accession Number |
|---|---|---|---|---|
| *Aedes aegypti* | Diptera | Culicidae | Blood | GCA_002204515.1 |
| *Anopheles gambiae* | Diptera | Culicidae | Blood | GCA_000005575.1 |
| *Culex quinquefaciatus* | Diptera | Culicidae | Blood | GCA_015732765.1 |
| *Culicoides tainanus* | Diptera | Ceratopogonida | Blood | GCA_020740445.1 |
| *Glossina fuscipes* | Diptera | Glossinidae | Blood | GCA_014805625.1 |
| *Lutzomyia longipalpis* | Diptera | Psychodidae | Blood | GCA_000265325.1 |
| *Phlebotomus papatasi* | Diptera | Psychodidae | Blood | GCA_000262795.1 |
| *Sarcophaga bullata* | Diptera | Sacrophagidae | Blood | GCA_005959815.1 |
| *Stomoxys calcitrans* | Diptera | Muscidae | Blood | GCA_001015335.1 |
| *Bactrocera tryoni* | Diptera | Tephritidae | No | GCA_016617805.2 |
| *Belgica antarctica* | Diptera | Chironomidae | No | GCA_000775305.1 |
| *Bradysia coprophila* | Diptera | Sciaridae | No | GCA_014529535.1 |
| *Ceratitis capitata* | Diptera | Tephritidae | No | GCA_000347755.4 |
| *Clanio marinus* | Diptera | Chironomidae | No | GCA_900005825.1 |
| *Clogmia albipunctata* | Diptera | Psychodidae | No | GCA_001014945.1 |
| *Contarinia nasturtii* | Diptera | Cecidomyiidae | No | GCA_009176525.2 |
| *Hermeta illucens* | Diptera | Stratiomyidae | No | GCA_905115235.1 |
| *Lucilia sericata* | Diptera | Calliphoridae | No | GCA_015586225.1 |
| *Musca domestica* | Diptera | Muscidae | No | GCA_000371365.1 |
| *Polypedilum vanderplanki* | Diptera | Chironomidae | No | GCA_018290095.1 |
| *Polypedilum pembai* | Diptera | Chironomidae | No | GCA_014622435.1 |
| *Propsilocerus akamusi* | Diptera | Chironomidae | No | GCA_018397935.1 |
| *Rhagoletis pomonella* | Diptera | Tephritidae | No | GCA_013731165.1 |
| *Rhagoletis zephyria* | Diptera | Tephritidae | No | GCA_001687245.1 |
| *Teleopsis dalmanni* | Diptera | Diopsidae | No | GCA_002237135.2 |
| *Zeugodacus cucurbitae* | Diptera | Tephritidae | No | GCA_000806345.1 |
| *Ctenocephalides felis* | Siphonaptera | Pulicidae | Blood | GCA_003426905.1 |
| Aricia agestis | Lepidoptera | Lycaenidae | No | GCA_905147365.1 |
| Bombyx mori | Lepidoptera | Bombycidae | No | GCA_014905235.2 |
| Danaus plexippus | Lepidoptera | Nymphalidae | No | GCA_009731565.1 |
| Helicoverpa armigera | Lepidoptera | Noctuidae | No | GCA_002156985.1 |
| Manduca sexta | Lepidoptera | Sphingidae | No | GCA_014839805.1 |
| Parage aegeria | Lepidoptera | Nymphalidae | No | GCA_905163445.1 |
| Plutella xylostella | Lepidoptera | Plutellidae | No | GCA_905116875.3 |
| Spodoptera frugiperda | Lepidoptera | Noctuidae | No | GCA_011064685.1 |
| Trichoplusia ni | Lepidoptera | Noctuidae | No | GCA_003590095.1 |
| Zerene cesonia | Lepidoptera | Pieridae | No | GCA_012273895.2 |

**Table 2:** TBLASTN results of average *D.melanogaster* -related iron gene copy numbers for genes with $p < 0.05$ in comparing hematophagous Diptera and Siphonaptera and non-hematophagous Diptera and Lepidoptera taxa. Each gene is represented with the corresponding p-value score.

| Phenotype | Species | Order | Spo (P =0.0123) | Cyp4d1 (P = 0.0448) | Nos (P =0.0433) | ND-B16.6 (P = 0.032) | ND-13A (P = 0.033) | Hr51 (P = 0.0205) | Cyp 12e1 (P = 0.0136) | Hmt-1 (P = 0.019) | Cysu (P = 0.0081) | Spok (P = 0.0255) | Pxt (P = 0.031) | Cd (P = 0.0038) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Blood | *A. aegypti* | Diptera | 6 | 4.67 | 3.33 | 4 | 2 | 1 | 6 | 2.5 | 1.5 | 3 | 2.75 | 1.5 |
| Blood | *A. gambiae* | Diptera | 6 | 4.67 | 5.67 | 4 | 2 | 2 | 4 | 2.67 | 1.5 | 3 | 2.6 | 1.5 |
| Blood | *C. quinquefaciatus* | Diptera | 6 | 5.33 | 4.33 | 2 | 2 | 1 | 10 | 3.2 | 1 | 3 | 1.8 | 1.5 |
| Blood | *C. tainanus* | Diptera | 6 | 3.33 | 3.33 | 4 | 2 | 1 | 0 | 4 | 1.5 | 3 | 2 | 1 |
| Blood | *G. fuscipes* | Diptera | 9 | 4 | 3.27 | 4 | 2 | 1 | 2 | 2.8 | 1.4 | 6 | 2 | 2.33 |
| Blood | *L. longipalpis* | Diptera | 9 | 3.5 | 5 | 6 | 0 | 1 | 12 | 7 | 1.5 | 3 | 1.83 | 1.5 |
| Blood | *P. paptasi* | Diptera | 9 | 3.33 | 3.3 | 2 | 2 | 1 | 12 | 8.67 | 1.75 | 3 | 1.5 | 1.5 |
| Blood | *S. bullata* | Diptera | 9 | 4.5 | 3 | 4 | 2 | 1 | 10 | 3.2 | 1.2 | 4.5 | 1.67 | 1.33 |
| Blood | *S. calcitrans* | Diptera | 9 | 7.5 | 3.27 | 2 | 2 | 1 | 6 | 2.8 | 1.4 | 6 | 1.83 | 2 |
| Blood | *C. felis* | Siphonaptera | 6 | 0 | 7.5 | 8 | 2 | 1.33 | 10 | 7 | 3 | 3 | 2.67 | 1 |
| No | *B. tryoni* | Diptera | 3 | 4 | 6.75 | 4 | 4 | 1 | 14 | 3.6 | 1.4 | 3 | 1.83 | 1.5 |
| No | *B. antarctica* | Diptera | 6 | 8 | 5.4 | 2 | 2 | 1 | 0 | 2.5 | 1.6 | 0 | 1.67 | 1 |
| No | *B. coprophila* | Diptera | 6 | 9.5 | 7.67 | 4 | 2 | 1 | 2 | 5.5 | 1.75 | 3 | 2 | 1.5 |
| No | *C. capitata* | Diptera | 6 | 3.5 | 5.75 | 4 | 4 | 1.5 | 10 | 4 | 1.6 | 3 | 1.67 | 1.5 |
| No | *C. marinus* | Diptera | 3 | 9.33 | 3.43 | 2 | 2 | 1 | 0 | 3 | 1.75 | 3 | 1.8 | 1 |
| No | *C. albipunctata* | Diptera | 3 | 8.67 | 3 | 2 | 0 | 1 | 0 | 5.5 | 1.75 | 3 | 1.8 | 1 |
| No | *C. nasturtil* | Diptera | 6 | 5 | 3.82 | 2 | 2 | 2 | 0 | 7 | 2.25 | 0 | 1.83 | 2 |
| No | *C. sericata* | Diptera | 12 | 8 | 3.27 | 4 | 0 | 1.2 | 2 | 3.6 | 1.2 | 6 | 3.17 | 2 |
| No | *H. illucens* | Diptera | 12 | 10 | 3.3 | 2 | 2 | 1 | 0 | 7 | 1.25 | 6 | 2 | 1.5 |
| No | *M. domestica* | Diptera | 9 | 6 | 4.09 | 4 | 2 | 1 | 2 | 2.8 | 1.2 | 4.5 | 1.83 | 2 |
| No | *P. pembai* | Diptera | 3 | 3.5 | 3.6 | 0 | 2 | 1 | 0 | 4 | 1 | 0 | 2.5 | 1 |
| No | *P. vanderplanki* | Diptera | 0 | 5.5 | 3.6 | 0 | 2 | 1 | 0 | 5.5 | 1.25 | 0 | 3 | 1 |
| No | *P. akamusi* | Diptera | 9 | 2 | 3 | 2 | 2 | 1 | 0 | 2 | 1.75 | 3 | 2 | 1 |
| No | *R. pomonella* | Diptera | 6 | 6 | 11.5 | 4 | 6 | 1 | 6 | 3.2 | 1.2 | 3 | 1.33 | 1.5 |
| No | *R. zephyria* | Diptera | 6 | 4.67 | 10.5 | 4 | 4 | 1 | 6 | 6.8 | 3.4 | 3 | 2.33 | 2.5 |
| No | *T. dalmanni* | Diptera | 6 | 11.5 | 3.27 | 4 | 2 | 1.75 | 18 | 8 | 2.6 | 3 | 2.5 | 5 |
| No | *Z. cucurbitae* | Diptera | 6 | 7.33 | 6 | 2 | 4 | 1 | 10 | 3.6 | 1.2 | 3 | 1.5 | 1.33 |
| No | *A. agestis* | Lepidoptera | 3 | 2 | 6 | 2 | 2 | 0 | 0 | 9.33 | 3.5 | 3 | 4 | 5 |
| No | *B. mori* | Lepidoptera | 3 | 2 | 6.38 | 2 | 2 | 0 | 0 | 7.5 | 3.5 | 0 | 3.33 | 4 |
| No | *D. plexippus* | Lepidoptera | 3 | 2 | 6.75 | 2 | 2 | 0 | 0 | 14.5 | 3.25 | 3 | 4 | 2.5 |
| No | *H. armigera* | Lepidoptera | 6 | 6 | 6 | 2 | 4 | 0 | 0 | 11.33 | 3 | 3 | 3.33 | 5 |
| No | *M. sexta* | Lepidoptera | 6 | 6 | 5.62 | 2 | 2 | 1 | 0 | 10.67 | 5 | 0 | 3.67 | 2.5 |
| No | *P. aegeria* | Lepidoptera | 0 | 6 | 5.62 | 2 | 2 | 0 | 0 | 10 | 3.25 | 3 | 3 | 3.5 |
| No | *P. xylostella* | Lepidoptera | 3 | 10 | 5.67 | 2 | 2 | 0 | 0 | 13.33 | 3.25 | 3 | 2.33 | 3.5 |
| No | *S. frugiperda* | Lepidoptera | 12 | 4 | 5.62 | 2 | 4 | 0 | 0 | 12.67 | 4.5 | 0 | 3.67 | 3.5 |
| No | *T. ni* | Lepidoptera | 6 | 2 | 7.88 | 2 | 2 | 0 | 4 | 10 | 3 | 3 | 3.67 | 3.5 |
| No | *Z. cesonia* | Lepidoptera | 0 | 4 | 5.62 | 2 | 4 | 0 | 0 | 9.5 | 3 | 3 | 3 | 6 |

**Table 3:** Average Iron gene copy numbers for contracted and expanded genes comparing the hematophagous and non-hematophagous taxa. Spook (Spo), Cytochrome P450 12e1 (Cyp12e1), Hormone receptor 51(Hr51), Spookier (Spok) and NADH dehydrogenase (ubiquinone) B16.6 subunit (ND-B16.6) potentially expanded. Cardinal (Cd), Curly Su (Cysu), Heavy metal tolerance factor 1 (Hmt-1), Peroxinectin-like (Pxt), NADH dehydrogenase (ubiquinone) 13 kDa A subunit (ND-13A), Nitric oxide synthase (Nos), and Cytochrome P450 4d1 (Cyp4d1) potentially contracted. "Sdv"= standard deviation results.

| **Gene Family Name** | **Hematophagous average gene copy number** | **Non-hematophagous average gene copy number** |
|---|---|---|
| Spo | 7.5 (Sdv = 1.581) | 5.3 ((Sdv = 3.363) |
| Cyp12e1 | 7.2 (Sdv = 4.237) | 2.7 (Sdv = 4.872) |
| Hr51 | 1.1 (Sdv = 0.322) | 0.8 (Sdv = 0.597) |
| Spok | 3.8 (Sdv = 1.275) | 2.5 (Sdv = 1.715) |
| ND-B16.6 | 4 (Sdv = 1.886) | 2.4 (Sdv = 1.155) |
| Cd | 1.5 (Sdv = 0.403) | 2.5 (Sdv = 1.492) |
| Cysu | 1.6 (Sdv = 0.539) | 2.3 (Sdv = 1.112) |
| Hmt-1 | 4.4 (Sdv = 2.273) | 6.9 (Sdv = 3.638) |
| Pxt | 2.1 (Sdv = 0.445) | 2.5 (Sdv = 0.832) |
| ND-13A | 1.8 (Sdv = 0.632) | 2.5 (Sdv = 1.312) |
| Nos | 4.2 (Sdv = 1.461) | 5.5 (Sdv = 2.146) |
| Cyp4d1 | 4.1 (Sdv = 1.893) | 5.8 (Sdv = 2.834) |

**Table 4:** TBLASTN results of average heme gene copy numbers for genes with $p < 0.05$ in comparing hematophagous Diptera and Siphonaptera and non-hematophagous Diptera and Lepidoptera taxa. Each gene represented has their corresponding p-value scores.

| Phenotype | Species | Order | Spo (P=0.012) | Cyp4d1 (P=0.045) | Nos (P=0.043) | ND-B16.6 (P=0.032) | ND-13A (P=0.033) | Hr51 (P=0.021) | Cyp 12e1 (P=0.014) | Hmt-1 (P=0.002) | Cysu (P=0.008) | Spok (P=0.025) | Pxt (P=0.031) | Cd (P=0.004) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Blood | *A. aegypti* | Diptera | 4 | 2.33 | 2.22 | 4 | 2 | 1 | 3 | 3.75 | 1.5 | 2 | 2.75 | 1.5 |
| Blood | *A. gambiae* | Diptera | 4 | 2.33 | 3.78 | 4 | 2 | 2 | 2 | 4 | 1.5 | 2 | 2.6 | 1.5 |
| Blood | *quinquefaciat* | Diptera | 4 | 2.67 | 2.89 | 2 | 2 | 1 | 5 | 4.8 | 1 | 2 | 1.8 | 1.5 |
| Blood | *C. tainanus* | Diptera | 4 | 1.67 | 2.22 | 4 | 2 | 1 | 0 | 6 | 1.5 | 2 | 2 | 1 |
| Blood | *G. fuscipes* | Diptera | 6 | 2 | 2.18 | 4 | 2 | 1 | 1 | 4.2 | 1.4 | 4 | 2 | 2.33 |
| Blood | *L. longipalpis* | Diptera | 6 | 1.75 | 3.33 | 6 | 0 | 1 | 6 | 10.5 | 1.5 | 2 | 1.83 | 1.5 |
| Blood | *P. paptasi* | Diptera | 6 | 1.67 | 2.2 | 2 | 2 | 1 | 6 | 13 | 1.75 | 2 | 1.5 | 1.5 |
| Blood | *S. bullata* | Diptera | 6 | 2.25 | 2 | 4 | 2 | 1 | 5 | 4.8 | 1.2 | 3 | 1.67 | 1.33 |
| Blood | *S. calcitrans* | Diptera | 6 | 3.75 | 2.18 | 2 | 2 | 1 | 3 | 4.2 | 1.4 | 4 | 1.83 | 2 |
| Blood | *C. felis* | Siphonaptera | 4 | 0 | 5 | 8 | 2 | 1.33 | 5 | 10.5 | 3 | 2 | 2.67 | 1 |
| No | *B. tryoni* | Diptera | 2 | 2 | 4.5 | 4 | 4 | 1 | 7 | 5.4 | 1.4 | 2 | 1.83 | 1.5 |
| No | *B. antarctica* | Diptera | 4 | 4 | 3.6 | 2 | 2 | 1 | 0 | 3.75 | 1.6 | 0 | 1.67 | 1 |
| No | *B. coprophila* | Diptera | 4 | 4.75 | 5.11 | 4 | 2 | 1 | 1 | 8.25 | 1.75 | 2 | 2 | 1.5 |
| No | *C. capitata* | Diptera | 4 | 1.75 | 3.83 | 4 | 4 | 1.5 | 5 | 6 | 1.6 | 2 | 1.67 | 1.5 |
| No | *C. marinus* | Diptera | 2 | 4.67 | 2.29 | 2 | 2 | 1 | 0 | 4.5 | 1.75 | 2 | 1.8 | 1 |
| No | *C. albipunctat* | Diptera | 2 | 4.33 | 2 | 2 | 0 | 1 | 0 | 8.25 | 1.75 | 2 | 1.8 | 1 |
| No | *C. nasturtil* | Diptera | 4 | 2.5 | 2.55 | 2 | 2 | 2 | 0 | 10.5 | 2.25 | 0 | 1.83 | 2 |
| No | *C. sericata* | Diptera | 8 | 4 | 2.18 | 4 | 0 | 1.2 | 1 | 5.4 | 1.2 | 4 | 3.17 | 2 |
| No | *H. illucens* | Diptera | 8 | 5 | 2.2 | 2 | 2 | 1 | 0 | 10.5 | 1.25 | 4 | 2 | 1.5 |
| No | *M. domestica* | Diptera | 6 | 3 | 2.73 | 4 | 2 | 1 | 1 | 4.2 | 1.2 | 3 | 1.83 | 2 |
| No | *P. pembai* | Diptera | 2 | 1.75 | 2.4 | 0 | 2 | 1 | 0 | 6 | 1 | 0 | 2.5 | 1 |
| No | *P. vanderplank* | Diptera | 0 | 2.75 | 2.4 | 0 | 2 | 1 | 0 | 8.25 | 1.25 | 0 | 3 | 1 |
| No | *P. akamusi* | Diptera | 6 | 1 | 2 | 2 | 2 | 1 | 0 | 3 | 1.75 | 2 | 2 | 1 |
| No | *R. pomonella* | Diptera | 4 | 3 | 7.67 | 4 | 6 | 1 | 3 | 4.8 | 1.2 | 2 | 1.33 | 1.5 |
| No | *R. zephyria* | Diptera | 4 | 2.33 | 7 | 4 | 4 | 1 | 3 | 10.2 | 3.4 | 2 | 2.33 | 2.5 |
| No | *T. dalmanni* | Diptera | 4 | 5.75 | 2.18 | 4 | 2 | 1.75 | 9 | 12 | 2.6 | 2 | 2.5 | 5 |
| No | *Z. cucurbitae* | Diptera | 4 | 3.67 | 4 | 2 | 4 | 1 | 5 | 5.4 | 1.2 | 2 | 1.5 | 1.33 |
| No | *A. agestis* | Lepidoptera | 2 | 1 | 4 | 2 | 2 | 0 | 0 | 14 | 3.5 | 2 | 4 | 5 |
| No | *B. mori* | Lepidoptera | 2 | 1 | 4.25 | 2 | 2 | 0 | 0 | 11.25 | 3.5 | 0 | 3.33 | 4 |
| No | *D. plexippus* | Lepidoptera | 2 | 1 | 4.5 | 2 | 2 | 0 | 0 | 21.75 | 3.25 | 2 | 4 | 2.5 |
| No | *H. armigera* | Lepidoptera | 4 | 3 | 4 | 2 | 4 | 0 | 0 | 17 | 3 | 2 | 3.33 | 5 |
| No | *M. sexta* | Lepidoptera | 4 | 3 | 3.75 | 2 | 2 | 1 | 0 | 16 | 5 | 0 | 3.67 | 2.5 |
| No | *P. aegeria* | Lepidoptera | 0 | 3 | 3.75 | 2 | 2 | 0 | 0 | 15 | 3.25 | 2 | 3 | 3.5 |
| No | *P. xylostella* | Lepidoptera | 2 | 5 | 3.78 | 2 | 2 | 0 | 0 | 20 | 3.25 | 2 | 2.33 | 3.5 |
| No | *S. frugiperda* | Lepidoptera | 8 | 2 | 3.75 | 2 | 4 | 0 | 0 | 19 | 4.5 | 0 | 3.67 | 3.5 |
| No | *T. ni* | Lepidoptera | 4 | 1 | 5.25 | 2 | 2 | 0 | 2 | 15 | 3 | 2 | 3.67 | 3.5 |
| No | *Z. cesonia* | Lepidoptera | 0 | 2 | 3.75 | 2 | 4 | 0 | 0 | 14.25 | 3 | 2 | 3 | 6 |

**Table 5:** Average Heme gene copy numbers for contracted and expanded genes comparing the hematophagous and non-hematophagous taxa. Spook (Spo), Cytochrome P450 12e1 (Cyp12e1), Hormone receptor 51(Hr51), Spookier (Spok) and NADH dehydrogenase (ubiquinone) B16.6 subunit (ND-B16.6) expanded. Cardinal (Cd), Curly Su (Cysu), Heavy metal tolerance factor 1 (Hmt-1), Peroxinectin-like (Pxt), NADH dehydrogenase (ubiquinone) 13 kDa A subunit (ND-13A), Nitric oxide synthase (Nos), and Cytochrome P450 4d1 (Cyp4d1) contracted. "Sdv" = standard deviation scores.

| **Gene Family Name** | **Hematophagous average gene copy number** | **Non-hematophagous average gene copy number** |
|---|---|---|
| Spo | 5 (Sdv = 1.0540 | 3.5 (Sdv =2.284) |
| Cyp12e1 | 3.6 (Sdv = 2.119 | 1.3 (Sdv = 2.462) |
| Hr51 | 1.1 (Sdv = 0.322) | 0.7 (Sdv = 0.607) |
| Spok | 2.5 (Sdv = 0.850) | 1.7 (Sdv = 1.164) |
| ND-B16.6 | 4 (Sdv =1.886) | 2.4 (Sdv = 1.134) |
| Cyp4d1 | 2 (Sdv =0.946) | 2.9 (Sdv = 1.445) |
| Nos | 2.8 (Sdv =0.974) | 3.5 (Sdv =1.212) |
| ND-13A | 1.8 (Sdv = 0.632) | 2.4 (Sdv =1.134) |
| Hmt-1 | 6.6 (Sdv = 3.409) | 10.6 (Sdv = 5.449) |
| Cysu | 1.6 (Sdv =0.539) | 2.4 (Sdv = 1.110) |
| Pxt | 2.1 (Sdv = 0.445) | 2.6 (Sdv = 0.812) |
| Cd | 1.5 (Sdv =0.403) | 2.5 (Sdv = 1.508) |

**Table 6:** TBLASTN results of average iron gene copy numbers for genes with $p < 0.05$ in comparing hematophagous Diptera and Siphonaptera and non-hematophagous Diptera taxa. Each gene represented has their corresponding p-value scores.

| Phenotype | Species | Order | Cyp4da1 (P = 0.009) | Cyp6a16 (P = 0.014) | Cyp6a17 (P = 0.037) | Cyp6a23 (P = 0.037) | ND-B14.5A (P = 0.047) | Cyp6t3 (P = 0.047) | ND-PDSW (P = 0.048) |
|---|---|---|---|---|---|---|---|---|---|
| Blood | *A. aegypti* | Diptera | 4.67 | 2 | 2 | 2 | 0 | 0 | 4 |
| Blood | *A. gambiae* | Diptera | 4.67 | 0 | 2 | 2 | 0 | 2 | 4 |
| Blood | *C. quinquefaciatus* | Diptera | 5.33 | 2 | 2 | 2 | 0 | 0 | 4 |
| Blood | *C. tainanus* | Diptera | 3.33 | 0 | 2 | 3 | 0 | 0 | 2 |
| Blood | *G. fuscipes* | Diptera | 4 | 0 | 0 | 2 | 2 | 4 | 2 |
| Blood | *L. longipalpis* | Diptera | 3.5 | 2 | 10 | 8 | 0 | 0 | 2 |
| Blood | *P. paptasi* | Diptera | 3.33 | 3 | 12 | 15 | 0 | 0 | 2 |
| Blood | *S. bullata* | Diptera | 4.5 | 4 | 9 | 9 | 2 | 4 | 2 |
| Blood | *S. calcitrans* | Diptera | 7.5 | 2 | 17 | 17 | 2 | 8 | 2 |
| Blood | *C. felis* | Siphonaptera | 0 | 0 | 16 | 18 | 0 | 0 | 4 |
| No | *B. tryoni* | Diptera | 4 | 7 | 14 | 15 | 2 | 16 | 2 |
| No | *B. antarctica* | Diptera | 8 | 4 | 3 | 7 | 2 | 0 | 2 |
| No | *B. coprophila* | Diptera | 9.5 | 0 | 9 | 9 | 0 | 0 | 2 |
| No | *C. capitata* | Diptera | 3.5 | 4 | 14 | 13 | 2 | 18 | 2 |
| No | *C. marinus* | Diptera | 9.33 | 6 | 10 | 9 | 2 | 4 | 2 |
| No | *C. albipunctata* | Diptera | 8.67 | 8 | 34 | 45 | 0 | 0 | 2 |
| No | *C. nasturtil* | Diptera | 5 | 0 | 3 | 3 | 0 | 0 | 2 |
| No | *C. sericata* | Diptera | 8 | 6 | 17 | 14 | 2 | 2 | 2 |
| No | *H. illucens* | Diptera | 10 | 5 | 9 | 18 | 2 | 4 | 0 |
| No | *M. domestica* | Diptera | 6 | 3 | 14 | 14 | 2 | 6 | 2 |
| No | *P. pembai* | Diptera | 3.5 | 0 | 16 | 16 | 0 | 4 | 2 |
| No | *P. vanderplanki* | Diptera | 5.5 | 0 | 13 | 12 | 0 | 0 | 2 |
| No | *P. akamusi* | Diptera | 2 | 0 | 3 | 3 | 2 | 0 | 2 |
| No | *R. pomonella* | Diptera | 6 | 10 | 12 | 13 | 2 | 5 | 4 |
| No | *R. zephyria* | Diptera | 4.67 | 8 | 14 | 18 | 2 | 12 | 2 |
| No | *T. dalmanni* | Diptera | 11.5 | 0 | 23 | 20 | 2 | 24 | 2 |
| No | *Z. cucurbitae* | Diptera | 7.33 | 12 | 18 | 20 | 2 | 6 | 2 |

**Table 7:** Average Iron gene copies for hematophagous Diptera & Siphonaptera and non-hematophagous Diptera. NADH dehydrogenase (ubiquinone) PDSW subunit (ND-PDSW) expanded. Cytochrome P450 4d1 (Cyp4d1), Cytochrome P450 6a16 (Cyp6a16), Cytochrome P4 50 6a17 (Cyp6a17), Cytochrome P450 6a23 (Cyp6a23), NADH dehydrogenase (ubiquinone) B14.5 A subunit (ND-B15.5A) and Cytochrome P450 6t3 (Cyp6t3) contracted. "Sdv" = standard deviation scores.

| Gene Family Name | Hematophagous Average Gene Copy Number | Non-hematophagous Average Gene Copy Number |
|---|---|---|
| ND-PDSW | 2.8 (Sdv = 1.033) | 2 (Sdv = 0.707) |
| Cyp4d1 | 4.1 (Sdv = 1.893) | 6.6 (Sdv = 2.687) |
| Cyp6a16 | 1.5 (Sdv = 1.434) | 4.3 (Sdv = 3.917) |
| Cyp6a17 | 7.2 (Sdv = 6.391) | 13.3 (Sdv = 7.647) |
| Cyp6a23 | 7.8 (Sdv = 6.663) | 14.6 (Sdv = 9.387) |
| ND-B15.5A | 0.6 (Sdv = 0.966) | 1.4 (Sdv = 0.939) |
| Cyp6t3 | 1.8 (Sdv = 2.741) | 5.9 (Sdv = 7.284) |

**Table 8:** TBLASTN results of average Heme gene copy numbers for genes with $p < 0.05$ in comparing hematophagous Diptera and Siphonaptera and non-hematophagous Diptera taxa. Each gene represented has their corresponding p-value scores.

| Phenotype | Species | Order | Cyp4d1 (P = 0.009) | Cyp6a16 (P = 0.014) | Cyp6a17 (P = 0.037) | Cyp6a23 (P = 0.037) | ND-B14.5A (P = 0.047) | Cyp6t3 (P = 0.047) | ND-PDSW (P = 0.048) |
|---|---|---|---|---|---|---|---|---|---|
| Blood | *A. aegypti* | Diptera | 2.33 | 1 | 1 | 1 | 0 | 0 | 4 |
| Blood | *A. gambiae* | Diptera | 2.33 | 0 | 1 | 1 | 0 | 1 | 4 |
| Blood | *C. quinquefaciatus* | Diptera | 2.67 | 1 | 1 | 1 | 0 | 0 | 4 |
| Blood | *C. tainanus* | Diptera | 1.67 | 0 | 1 | 1.5 | 0 | 0 | 2 |
| Blood | *G. fuscipes* | Diptera | 2 | 0 | 0 | 1 | 2 | 2 | 2 |
| Blood | *L. longipalpis* | Diptera | 1.75 | 1 | 5 | 4 | 0 | 0 | 2 |
| Blood | *P. paptasi* | Diptera | 1.67 | 1.5 | 6 | 7.5 | 0 | 0 | 2 |
| Blood | *S. bullata* | Diptera | 2.25 | 2 | 4.5 | 4.5 | 2 | 2 | 2 |
| Blood | *S. calcitrans* | Diptera | 3.75 | 1 | 8.5 | 8.5 | 2 | 4 | 2 |
| Blood | *C. felis* | Siphonaptera | 0 | 0 | 8 | 9 | 0 | 0 | 4 |
| No | *B. tryoni* | Diptera | 2 | 3.5 | 7 | 7.5 | 2 | 8 | 2 |
| No | *B. antarctica* | Diptera | 4 | 2 | 1.5 | 3.5 | 2 | 0 | 2 |
| No | *B. coprophila* | Diptera | 4.75 | 0 | 4.5 | 4.5 | 0 | 0 | 2 |
| No | *C. capitata* | Diptera | 1.75 | 2 | 7 | 6.5 | 2 | 9 | 2 |
| No | *C. marinus* | Diptera | 4.67 | 3 | 5 | 4.5 | 2 | 2 | 2 |
| No | *C. albipunctata* | Diptera | 4.33 | 4 | 17 | 22.5 | 0 | 0 | 2 |
| No | *C. nasturtil* | Diptera | 2.5 | 0 | 1.5 | 1.5 | 0 | 0 | 2 |
| No | *C. sericata* | Diptera | 4 | 3 | 8.5 | 7 | 2 | 1 | 2 |
| No | *H. illucens* | Diptera | 5 | 2.5 | 4.5 | 9 | 2 | 2 | 0 |
| No | *M. domestica* | Diptera | 3 | 1.5 | 7 | 7 | 2 | 3 | 2 |
| No | *P. pembai* | Diptera | 1.75 | 0 | 8 | 8 | 0 | 2 | 2 |
| No | *P. vanderplanki* | Diptera | 2.75 | 0 | 6.5 | 6 | 0 | 0 | 2 |
| No | *P. akamusi* | Diptera | 1 | 0 | 1.5 | 1.5 | 2 | 0 | 2 |
| No | *R. pomonella* | Diptera | 3 | 5 | 6 | 6.5 | 2 | 2.5 | 4 |
| No | *R. zephyria* | Diptera | 2.33 | 4 | 7 | 9 | 2 | 6 | 2 |
| No | *T. dalmanni* | Diptera | 5.75 | 0 | 11.5 | 10 | 2 | 12 | 2 |
| No | *Z. cucurbitae* | Diptera | 3.67 | 6 | 9 | 10 | 2 | 3 | 2 |

**Table 9:** Average heme gene copies for hematophagous (Diptera & Siphonaptera) and non-hematophagous taxa (Diptera). NADH dehydrogenase (ubiquinone) PDSW subunit (ND-PDSW) expanded. Cytochrome P450 4d1 (Cyp4d1), Cytochrome P450 6a16 (Cyp6a16), Cytochrome P450 6a17 (Cyp6a17), Cytochrome P450 6a23 (Cyp6a23), NADH dehydrogenase (ubiquinone) B14.5 A subunit (ND-B15.5A) and Cytochrome P450 6t3 (Cyp6t3) contracted. "Sdv" = standard deviation scores.

| **Gene Family Name** | **Hematophagous Average Gene Copy Number** | **Non-hematophagous Average Gene Copy Number** |
|---|---|---|
| ND-PDSW | 2.8 (Sdv =1.033) | 2 (Sdv = 0.707) |
| Cyp4d1 | 2 (Sdv = 0.946) | 3.3 (Sdv = 1.344) |
| Cyp6a16 | 0.8 (Sdv = 0.717) | 2.1 (Sdv = 1.959) |
| Cyp6a17 | 3.6 (Sdv = 3.195) | 6.6 (Sdv = 3.823) |
| Cyp6a23 | 3.9 (Sdv = 3.332) | 7.3 (Sdv = 4.694) |
| ND-B15.5A | 0.6 (Sdv = 0.966) | 1.4 (Sdv = 0.939) |
| Cyp6t3 | 0.9 (Sdv = 1.370) | 3 (Sdv = 3.642) |

**Table10a:** *D. melanogaster* associated genes with corresponding FlyBase ID of each gene. Blue highlights potential gene expansions in the analysis of hematophagous Diptera, Siphonaptera, and non-hematophagous Diptera and Lepidoptera.

| ***D.melanogaster* associated genes** | **FlyBase ID** |
|---|---|
| Spook (Spo) | FBgn0003486 |
| Cytochrome P450 12e1 (Cyp12e1) | FBgn0037817 |
| Hormone receptor 51 (Hr51) | FBgn0034012 |
| Spookier (Spok) | FBgn0086917 |
| NADH dehydrogenase (ubiquinone) B16.6 subunit (ND-B16.6) | FBgn0029868 |
| Cardinal (Cd) | FBgn0263986 |
| Curly su (Cysu) | FBgn0038511 |
| Heavy metal tolerance factor 1(Hmt-1) | FBgn0038376 |
| Peroxinectin-like (Pxt) | FBgn0261987 |
| NADH dehydrogenase (ubiquinone) 13 kDa A subunit (ND-13A) | FBgn0031684 |
| Nitric oxide synthase (Nos) | FBgn0011676 |
| Cytochrome P450 4d1 (Cyp4d1) | FBgn0005670 |

**Table10b:** *D. melanogaster* associated genes with corresponding FlyBase ID of each gene. Blue highlights potential gene expansions in the analysis of hematophagous Diptera, Siphonaptera, and non-hematophagous Diptera.

| ***D.melanogaster* associated genes** | **FlyBase ID** |
|---|---|
| NADH dehydrogenase (ubiquinone) PDSW subunit (ND-PDSW) | FBgn0021967 |
| Cytochrome P450 4d1 (Cyp4d1) | FBgn0005670 |
| Cytochrome P450 6a16 (Cyp6a16) | FBgn0031726 |
| Cytochrome P450 6a17 (Cyp6a17) | FBgn0015714 |
| Cytochrome P450 6a23 (Cyp6a23) | FBgn0033978 |
| Cytochrome P450 6t3 (Cyp6t3) | FBgn0033697 |
| NADH dehydrogenase (ubiquinone) B14.5 A subunit (ND-B14.5A) | FBgn0033978 |

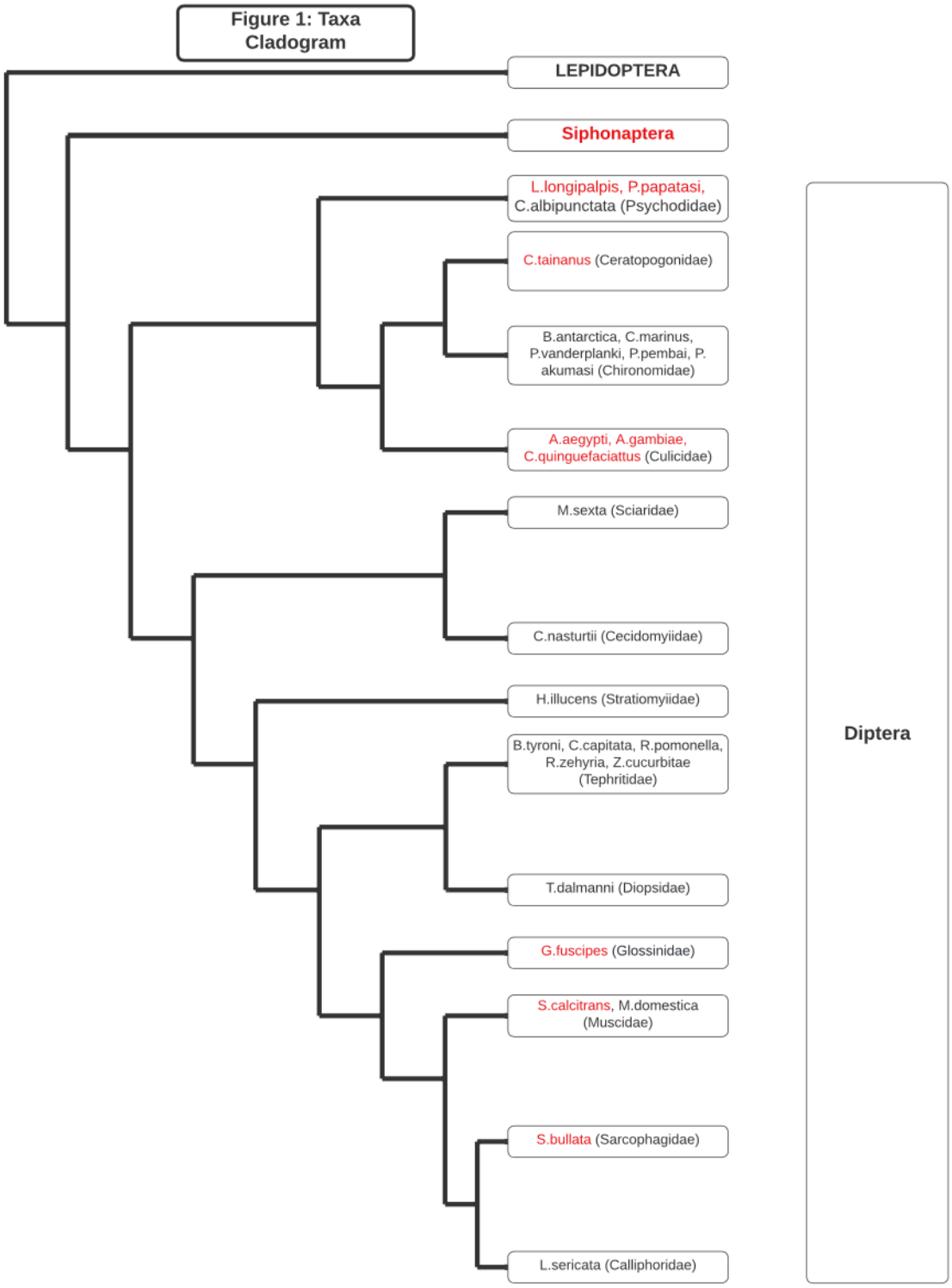
Figure 1: Taxa Cladogram
LEPIDOPTERA
Siphonaptera
L.longipalpis, P.papatasi, C.albipunctata (Psychodidae)
C.tainanus (Ceratopogonidae)
B.antarctica, C.marinus, P.vanderplanki, P.pembai, P. akumasi (Chironomidae)
A.aegypti, A.gambiae, C.quinguefaciattus (Culicidae)
M.sexta (Sciaridae)
C.nasturtii (Cecidomyiidae)
H.illucens (Stratiomyiidae)
B.tyroni, C.capitata, R.pomonella, R.zehyria, Z.cucurbitae (Tephritidae)
T.dalmanni (Diopsidae)
G.fuscipes (Glossinidae)
S.calcitrans, M.domestica (Muscidae)
S.bullata (Sarcophagidae)
L.sericata (Calliphoridae)
Diptera

## 6. Works Cited

Agip, Ahmed-Noor A., et al. "Mammalian Respiratory Complex I through the Lens of Cryo-EM." *Annual Review of Biophysics*, vol. 48, no. 1, 2019, pp. 165–184, doi:10.1146/annurev-biophys-052118-115704.

Alberch, Pere, and Emily A. Gale. "A Developmental Analysis of an Evolutionary Trend: Digital Reduction in Amphibians." *Evolution; International Journal of Organic Evolution*, vol. 39, no. 1, 1985, p. 8, doi:10.2307/2408513.

Altschul, S. F., et al. "Basic Local Alignment Search Tool." *Journal of Molecular Biology*, vol. 215, no. 3, 1990, pp. 403–410, doi:10.1016/S0022-2836(05)80360-2.

Baker, Heather M., et al. "Dealing with Iron: Common Structural Principles in Proteins That Transport Iron and Heme." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 7, 2003, pp. 3579–3583, doi:10.1073/pnas.0637295100.

Brindley, Paul J. "Proteolytic Degradation of Host Hemoglobin by Schistosomes." *Parasitology International*, vol. 47, 1998, p. 60, doi:10.1016/s1383-5769(98)80112-8.

Brunori, M. "Hemoglobin Is an Honorary Enzyme." *Trends in Biochemical Sciences*, vol. 24, no. 4, 1999, pp. 158–161, doi:10.1016/s0968-0004(99)01380-8.

Claudianos, C., et al. "A Deficit of Detoxification Enzymes: Pesticide Sensitivity and Environmental Response in the Honeybee." *Insect Molecular Biology*, vol. 15, no. 5, 2006, pp. 615–636, doi:10.1111/j.1365-2583.2006.00672.x.

Dalton, John P., et al. "Role of the Tegument and Gut in Nutrient Uptake by Parasitic Platyhelminths." *Canadian Journal of Zoology*, vol. 82, no. 2, 2004, pp. 211–232, doi:10.1139/z03-213.

Dashti, Zahra Jalali Sefid, et al. “Computational Characterization of Iron Metabolism in the Tsetse Disease Vector, Glossina Morsitans: IRE Stem-Loops.” *BMC Genomics*, vol. 17, no. 1, 2016, p. 561, doi:10.1186/s12864-016-2932-7.

de Rosny, Eve, et al. “DHR51, the Drosophila Melanogaster Homologue of the Human Photoreceptor Cell-Specific Nuclear Receptor, Is a Thiolate Heme-Binding Protein.” *Biochemistry*, vol. 47, no. 50, 2008, pp. 13252–13260, doi:10.1021/bi801691b.

Doolittle, R. F. “Convergent Evolution: The Need to Be Explicit.” *Trends in Biochemical Sciences*, vol. 19, no. 1, 1994, pp. 15–18, doi:10.1016/0968-0004(94)90167-8.

Everse, J., and N. Hsia. “The Toxicities of Native and Modified Hemoglobins.” *Free Radical Biology & Medicine*, vol. 22, no. 6, 1997, pp. 1075–1099, doi:10.1016/s0891-5849(96)00499-6.

Faassen, A. E., et al. “Hemoglobin: A Lifesaver and an Oxidant. How to Tip the Balance.” *Biomaterials, Artificial Cells, and Artificial Organs*, vol. 16, no. 1–3, 1988, pp. 93–104, doi:10.3109/10731198809132559.

Freitas, Lucas, and Mariana F. Nery. “Expansions and Contractions in Gene Families of Independently-Evolved Blood-Feeding Insects.” *BMC Evolutionary Biology*, vol. 20, no. 1, 2020, p. 87, doi:10.1186/s12862-020-01650-3.

Galay, Remil Linggatong, et al. “Iron Metabolism in Hard Ticks (Acari: Ixodidae): The Antidote to Their Toxic Diet.” *Parasitology International*, vol. 64, no. 2, 2015, pp. 182–189, doi:10.1016/j.parint.2014.12.005.

Gertz, E. Michael, et al. “Composition-Based Statistics and Translated Nucleotide Searches: Improving the TBLASTN Module of BLAST.” *BMC Biology*, vol. 4, no. 1, 2006, p. 41, doi:10.1186/1741-7007-4-41.

Graça-Souza, Aurélio V., et al. “Adaptations against Heme Toxicity in Blood-Feeding Arthropods.” *Insect Biochemistry and Molecular Biology*, vol. 36, no. 4, 2006, pp. 322–335, doi:10.1016/j.ibmb.2006.01.009.

Hamza, Iqbal, and Harry A. Dailey. “One Ring to Rule Them All: Trafficking of Heme and Heme Synthesis Intermediates in the Metazoans.” *Biochimica et Biophysica Acta. Molecular Cell Research*, vol. 1823, no. 9, 2012, pp. 1617–1632, doi:10.1016/j.bbamcr.2012.04.009.

Heliconius Genome Consortium. “Butterfly Genome Reveals Promiscuous Exchange of Mimicry Adaptations among Species.” *Nature*, vol. 487, no. 7405, 2012, pp. 94–98, doi:10.1038/nature11041.

Hou, Shangwei, et al. “Reversible Binding of Heme to Proteins in Cellular Signal Transduction.” *ChemInform*, vol. 38, no. 16, 2007, doi:10.1002/chin.200716271.

Kuo, J., and C. den Hartog. “Seagrass Morphology, Anatomy, and Ultrastructure.” *SEAGRASSES: BIOLOGY, ECOLOGYAND CONSERVATION*, Springer Netherlands, 2007, pp. 51–87.

Li, Xianchun, et al. “Molecular Mechanisms of Metabolic Resistance to Synthetic and Natural Xenobiotics.” *Annual Review of Entomology*, vol. 52, no. 1, 2007, pp. 231–253, doi:10.1146/annurev.ento.51.110104.151104.

Lima, Vitor L. A., et al. “The Antioxidant Role of Xanthurenic Acid in the Aedes Aegypti Midgut during Digestion of a Blood Meal.” *PloS One*, vol. 7, no. 6, 2012, p. e38349, doi:10.1371/journal.pone.0038349.

Losos, Jonathan B. “Convergence, Adaptation, and Constraint.” *Evolution; International Journal of Organic Evolution*, vol. 65, no. 7, 2011, pp. 1827–1840, doi:10.1111/j.1558-5646.2011.01289.x.

Ono, Hajime, et al. “Spook and Spookier Code for Stage-Specific Components of the Ecdysone Biosynthetic Pathway in Diptera.” *Developmental Biology*, vol. 298, no. 2, 2006, pp. 555–570, doi:10.1016/j.ydbio.2006.07.023.

Pettis, Jeffery S., et al. “Pesticide Exposure in Honey Bees Results in Increased Levels of the Gut Pathogen Nosema.” *The Science of Nature*, vol. 99, no. 2, 2012, pp. 153–158, doi:10.1007/s00114-011-0881-1.

Pierre, J. L., et al. *Biometals: An International Journal on the Role of Metal Ions in Biology, Biochemistry, and Medicine*, vol. 15, no. 4, 2002, pp. 341–346, doi:10.1023/a:1020259021641.

Reich, P. B., et al. “The Evolution of Plant Functional Variation: Traits, Spectra, and Strategies.” *International Journal of Plant Sciences*, vol. 164, no. S3, 2003, pp. S143–S164, doi:10.1086/374368.

Reuben Kaufman, W. “Gluttony and Sex in Female Ixodid Ticks: How Do They Compare to Other Blood-Sucking Arthropods?” *Journal of Insect Physiology*, vol. 53, no. 3, 2007, pp. 264–273, doi:10.1016/j.jinsphys.2006.10.004.

Rohlfing, Kim, et al. “Convergent Evolution of Hemoglobin Switching in Jawed and Jawless Vertebrates.” *BMC Evolutionary Biology*, vol. 16, no. 1, 2016, p. 30, doi:10.1186/s12862-016-0597-0.

Sackton, Timothy B., and Nathan Clark. “Convergent Evolution in the Genomics Era: New Insights and Directions.” *Philosophical Transactions of the Royal Society of London.*

*Series B, Biological Sciences*, vol. 374, no. 1777, 2019, p. 20190102, doi:10.1098/rstb.2019.0102.

Schwarze, Kim, et al. "The Globin Gene Repertoire of Lampreys: Convergent Evolution of Hemoglobin and Myoglobin in Jawed and Jawless Vertebrates." *Molecular Biology and Evolution*, vol. 31, no. 10, 2014, pp. 2708–2721, doi:10.1093/molbev/msu216.

Shindo, Chikako, et al. "Role of FRIGIDA and FLOWERING LOCUS C in Determining Variation in Flowering Time of Arabidopsis." *Plant Physiology*, vol. 138, no. 2, 2005, pp. 1163–1173, doi:10.1104/pp.105.061309.

Sterkel, Marcos, José Henrique M. Oliveira, et al. "The Dose Makes the Poison: Nutritional Overload Determines the Life Traits of Blood-Feeding Arthropods." *Trends in Parasitology*, vol. 33, no. 8, 2017, pp. 633–644, doi:10.1016/j.pt.2017.04.008.

Sterkel, Marcos, Hugo D. Perdomo, et al. "Tyrosine Detoxification Is an Essential Trait in the Life History of Blood-Feeding Arthropods." *Current Biology: CB*, vol. 26, no. 16, 2016, pp. 2188–2193, doi:10.1016/j.cub.2016.06.025.

Stern, David L. "The Genetic Causes of Convergent Evolution." *Nature Reviews. Genetics*, vol. 14, no. 11, 2013, pp. 751–764, doi:10.1038/nrg3483.

Strode, Clare, et al. "Genomic Analysis of Detoxification Genes in the Mosquito Aedes Aegypti." *Insect Biochemistry and Molecular Biology*, vol. 38, no. 1, 2008, pp. 113–123, doi:10.1016/j.ibmb.2007.09.007.

Toh, Shu Qin, et al. "Heme and Blood-Feeding Parasites: Friends or Foes?" *Parasites & Vectors*, vol. 3, no. 1, 2010, p. 108, doi:10.1186/1756-3305-3-108.

Werck-Reichhart, D., and R. Feyereisen. "Cytochromes P450: A Success Story." *Genome Biology*, vol. 1, no. 6, 2000, p. REVIEWS3003, doi:10.1186/gb-2000-1-6-reviews3003.

Wheeler, Gregory L., and Bryan C. Carstens. “Evaluating the Adaptive Evolutionary Convergence of Carnivorous Plant Taxa through Functional Genomics.” *PeerJ*, vol. 6, 2018, p. e4322, doi:10.7717/peerj.4322.

Whiten, Shavonn R., et al. “Ironing out the Details: Exploring the Role of Iron and Heme in Blood-Sucking Arthropods.” *Frontiers in Physiology*, vol. 8, 2017, p. 1134, doi:10.3389/fphys.2017.01134.

Wu, Shengdan, et al. “Genetic Contribution of Paleopolyploidy to Adaptive Evolution in Angiosperms.” *Molecular Plant*, vol. 13, no. 1, 2020, pp. 59–71, doi:10.1016/j.molp.2019.10.012.

Xu, Shaohua, et al. “Genomic Convergence in the Adaptation to Extreme Environments.” *Plant Communications*, vol. 1, no. 6, 2020, p. 100117, doi:10.1016/j.xplc.2020.100117.

Zhou, Guoli, et al. “Fate of Blood Meal Iron in Mosquitoes.” *Journal of Insect Physiology*, vol. 53, no. 11, 2007, pp. 1169–1178, doi:10.1016/j.jinsphys.2007.06.009.