

RESEARCH ARTICLE

Open Access



# Expansions and contractions in gene families of independently-evolved blood-feeding insects

Lucas Freitas<sup>1,2\*</sup>  and Mariana F. Nery<sup>1</sup>

## Abstract

**Background:** The blood-feeding behavior evolved multiple times in Insecta lineages and it represents an excellent opportunity to study patterns of convergent molecular evolution regarding this habit. In insects the expansion of some gene families is linked with blood-feeding behavior, but a wide study comparing the evolution of these gene families among different lineages is still missing. Here we gathered genomic data from six independently-evolved hematophagous lineages, aiming to identify convergent expansions and/or contractions of gene families in hematophagous lineages of insects.

**Results:** We found four rapidly evolving gene families shared by at least two hematophagous independently-evolved lineages, including a heat-shock and a chemosensory protein. On the expression of these four rapidly evolving gene families we found more genes expressed in mated individuals compared with virgin individuals in rapidly-expanded families and more genes expressed in non-blood-feeding individuals compared with blood-feeding individuals in rapidly-contracted families.

**Conclusion:** Our results reveal a new set of candidate genes to be explored in further analysis to help the development of new strategies to deal with blood-feeding vectors and also presents a new perspective to study the evolution of hematophagy identifying convergent molecular patterns.

**Keywords:** Comparative genomics, Hematophagy, Adaptation, Evolution, Heat-shock protein, Chemosensory protein

## Background

Hematophagy, or the habit of blood-feeding, is a widespread feeding behavior, being found in several groups like insects, fishes, mammals, and birds [1]. Several of these hematophagous species are able to transmit viruses, bacteria, and parasites during the blood-feeding behavior, which are harmful to the host, causing huge economic loss and health impacts on humans and

animals. For instance, only on the American continent Dengue fever has an estimated cost of \$2.1 billion per year, in 2010 US dollars [2]. Further, more than 17% of all infectious diseases are carried by vectors, causing 700,000 deaths annually [3].

The main group responsible for vector-borne diseases is the class Insecta, where hematophagy independently evolved multiple times [4, 5]. To better understand the basis of insects' hematophagy, different approaches have been employed in many fields of science, e.g. animal behavior [6], ecology [7], biochemistry [8] and genomics [9, 10].

Comparative functional genomics studies in hematophagous insects usually target salivary glands tissues due to their specific adaptations to hematophagy, as reviewed by Arcà & Ribeiro (2018) [11]. For example,

\* Correspondence: [freitas.la@protonmail.ch](mailto:freitas.la@protonmail.ch)

<sup>1</sup>Laboratório de Genômica Evolutiva, Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, Rua Bertrand Russell, S/N, Cidade Universitária, IB, Bloco H, Campinas, SP, Brazil

<sup>2</sup>School of Biological and Chemical Sciences, Queen Mary University of London, London, UK



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

proteins found in salivary glands of hematophagous insects have anti-clotting and anti-itching properties allowing them to pierce other animals and feed on blood without being detected [12].

Since hematophagy evolved independently among insects, it is expected that distinct proteins would be found in salivary glands of independently-evolved lineages, which was reviewed by Ribeiro et al. (2010) [12]. However, some of these distinct proteins have the same function: lipocalins and D7 proteins act as a biogenic amines scavenger in ticks, kissing bugs and mosquitoes [13], revealing a functional convergence related with hematophagy.

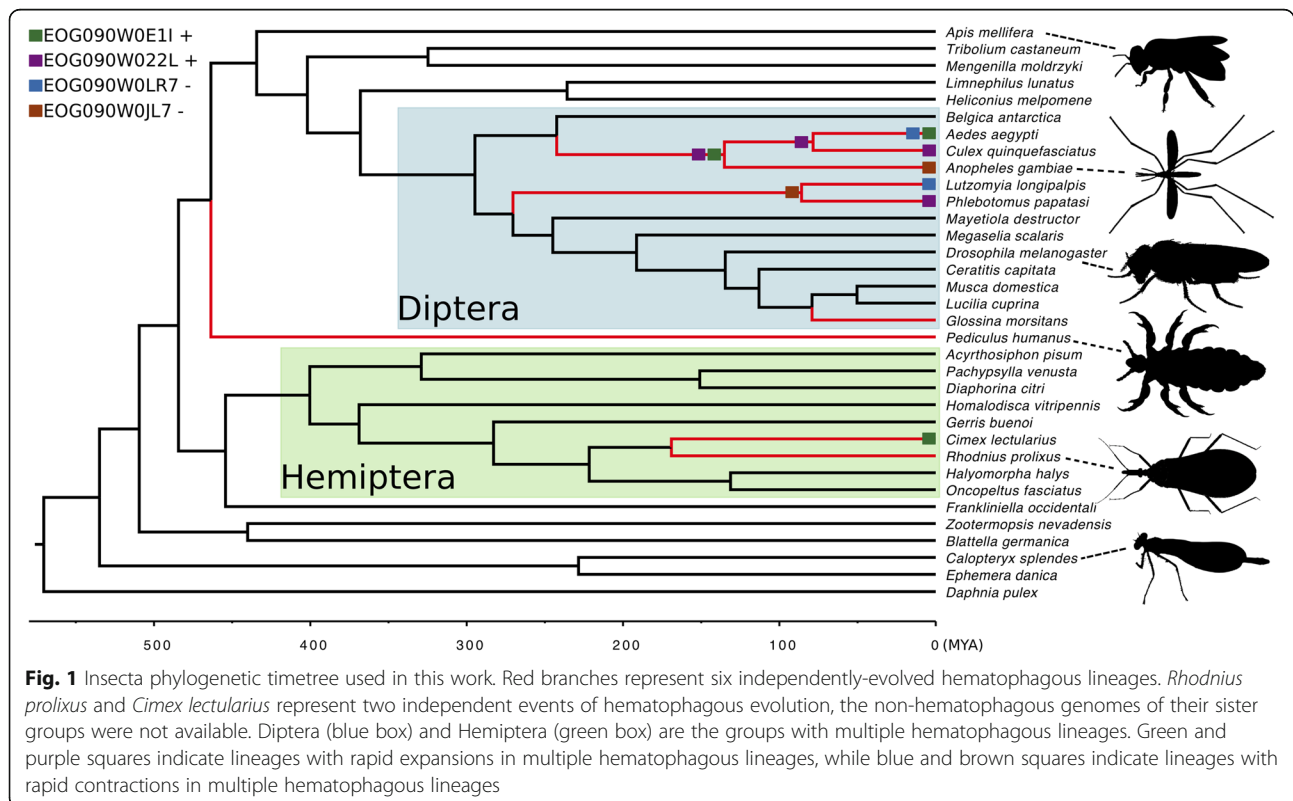
Still, other kinds of convergent evolution are found at molecular level. Convergent amino acid substitutions [14], convergent shift in evolutionary rates [15] and convergence in gene families copy number variation [16] are examples of convergent molecular evolution that are reported in the literature. In this context, the independent evolution of hematophagy in insects represents an excellent opportunity to study patterns of convergent molecular evolution. Although several strategies to infer convergent molecular evolution are being developed and used [17], few works aimed to study the convergence of gain and loss in gene copies number across distinct lineages [16]. For example, in insects, the expansion of some gene families has been linked to blood-feeding behavior,

since the expression of extra copies of genes would help in blood digestion [18–21].

Despite the importance of gene family turnovers, a wide study comparing the evolution of gene families among different blood-feeding lineages of insects is still missing. Accordingly, here we gathered genomic data from six independently-evolved hematophagous lineages, Culicidae, Psychodidae, Glossinidae (Diptera order), Reduviidae, Cimicidae (Hemiptera order) and Phthiraptera (Psocodea order), aiming to identify convergent expansions and/or contractions in gene families in hematophagous lineages of insects and their role in the evolution of the blood-feeding behavior using gene expression analysis.

## Results

Regarding the inference of gene families evolution, CAFE estimated an error distribution ( $e$ ) of 0.075 which means that less than 10% of gene families in our dataset have the wrong number of gene copies. Using this error distribution, we tested two models using the timetree inferred with MCMCtree: the global model, where the whole tree has the same  $\lambda$  rate, and the hemato model, where hematophagous branches have a distinct  $\lambda$  rate compared with non-hematophagous branches (Fig. 1). The  $\lambda$  rate for the global model was  $1.7 \times 10^{-3}$ , while this same  $\lambda$  rate was  $1.2 \times 10^{-3}$  for hematophagous branches and  $1.7 \times 10^{-3}$  for non-hematophagous



branches in the hemato model. The likelihood-ratio test was significant ( $p = 0$ ), showing that the hemato model fits better the data compared with the global model. This result shows that hematophagous lineages have a slower rate of gain and losses of genes compared with non-hematophagous lineages. Although hematophagous lineages had an overall slower rate of gain and losses, they have approximately the same mean number of gene families with rapid expansions ( $14.92 \pm 14.56$ ) compared to non-hematophagous lineages ( $14.6 \pm 21.15$ ), per lineage. On the other hand, hematophagous lineages had a significantly higher number of rapid contractions ( $8.83 \pm 8.25$ ) compared to non-hematophagous lineages ( $4.92 \pm 6.75$ ), per lineage (Wilcoxon rank sum test  $p$ -value = 0.015).

Since we aimed to identify gains/losses in independently-evolved hematophagous lineages, we focused on rapidly evolving families that were found in at least two of these lineages. We found four of such gene families: rapid expansions in a small heat shock protein HSP20 (EOG090W0E11) and in a carboxylesterase (EOG090W022L), and rapid contractions in a transmembrane protein (EOG090W0LR7) and in an odorant-binding protein (EOG090W0JL7) (Table 1), but only the HSP20 showed a convergent pattern of expansion considering a non-hematophagous ancestor. As control of this test, we applied the same procedure in nearby non-hematophagous lineages; in this case, only two genes, a glucose-methanol-choline oxidoreductase (EOG090W08UU) and an uncharacterized protein (EOG090W00HB), had rapid expansions in distinct orders and no multiple contractions were found. Another interest gene family was the EOG090W0B2X family, a carbohydrate kinase, which was present in 44 branches but missing in all hematophagous lineages.

To deep explore the role of these four candidate gene families in the hematophagous behavior in Insecta, we specifically searched for genes differentially expressed in the transcriptomes available at VectorBase. From 315 genes in all hematophagous species in these four gene families, we were able to identify 95 genes differentially expressed in VectorBase database (Additional file 1: Tables S1-S4). From these 95 genes, 78 were found in the sex set, 17 in the intercourse set, 52 were found in

the feeding set, 22 genes were shared between sex and feeding and 15 genes were shared by all categories (Fig. 2).

On gene families with rapid expansions, the HSP20 family contains 18 genes differentially expressed (Additional file 1: Table S1), among which five are shared by all sets, nine are shared between sex and feeding, two are exclusively found in the feeding set and other two are exclusively found in the sex set (Fig. 3A). The carboxylesterase family contains 33 genes differentially expressed (Additional file 1: Table S2), including six shared by all sets, nine shared by sex and feeding sets, 12 exclusively found in the sex set, seven in the feeding set and one found in the intercourse set (Fig. 3B). Functions of these gene families were assigned as metal ion binding and carboxylic ester hydrolase activity for HSP20 and carboxylesterase families, respectively.

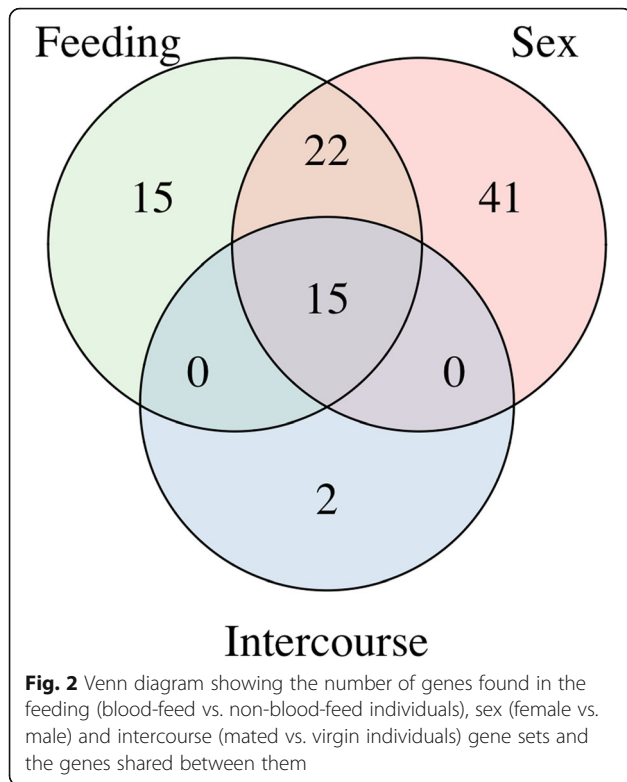
Among gene families with rapid contractions, the transmembrane protein EOG090W0LR7 contains only one gene differentially expressed, AGAP009139, found in the intercourse set, and present both in mated and virgin individuals. The odorant-binding EOG090W0JL7 contains 41 genes differentially expressed (Additional file 1: Table S4), but only four are shared between all sets, while the other four genes are shared between sex and feeding sets. In the sex set, we found 27 exclusive genes, while the feeding set contains six exclusive genes (Fig. 3C). Regarding the function of these families, both the transmembrane protein EOG090W0LR7 and the odorant-binding EOG090W0JL7 do not have functional annotation in VectorBase, thus we blasted them against InterPro [22] and Pfam [23] databases to assign their functions. Both families remained without a functional annotation, however, the odorant-binding EOG090W0JL7 was assigned as a member of the chemosensory protein family (CSP) based on OrthoDB annotation.

## Discussion

Shifts in the numbers of copies from a specific gene family represent a raw source of evolutionary innovation. More copies of a specific gene could increase the content of protein, allowing the organism to better deal with some tasks, like digestion [24] or detoxification of some compound [25]. On the other hand, some gene losses

**Table 1** Number of gene gain (+) and loss (–) (in parenthesis) and functional annotation for all rapidly-evolving gene families. In bold, rapid/expansions in independently-evolved hematophagous lineages

OrthoDB ID	Lineage (gene change)	Functional annotation
EOG090W0E11	<b><i>Aedes aegypti</i> (+ 8), <i>Culicidae</i> (+ 2), <i>Cimex lecturianus</i> (+ 2), <i>Culex quinquefasciatus</i> (– 7)</b>	Metal ion binding
EOG090W022L	<b><i>Culicinae</i> (+ 4), <i>Culex quinquefasciatus</i> (+ 3), <i>Phlebotomus papatasi</i> (+ 4), <i>Culicidae</i> (+ 4), <i>Lutzomyia longipalpis</i> (– 12)</b>	Carboxylic ester hydrolase activity
EOG090W0LR7	<b><i>Aedes aegypti</i> (– 3), <i>Lutzomyia longipalpis</i> (– 4), <i>Homoladisca vitripennis</i> (+ 3), <i>Lucilia cuprina</i> (+ 3), <i>Culex quinquefasciatus</i> (+ 2), <i>Zootermopsis nevadensis</i> (+ 2)</b>	–
EOG090W0JL7	<b><i>Anopheles gambiae</i> (– 12), <i>Psychodidae</i> (– 8), <i>Musca domestica</i> (+ 6), <i>Aedes aegypti</i> (+ 11), <i>Culicinae</i> (+ 7), <i>Culicidae</i> (+ 4), <i>Heliconius melpomene</i> (+ 3)</b>	Chemosensory protein



leading to phenotypic and metabolic changes are also thought to be adaptive [26]. In common, both expansions and contractions might indicate an adaptation to a new lifestyle.

The transition to hematophagy represents a new lifestyle, where insects independently developed new physiological adaptations to seek blood sources and to feed on blood [27, 28]. Here we explored the evolution of gene families in Insecta to identify significant convergent patterns of expansions and contractions in gene families of six independently-evolved blood-feeding lineages and their relationship with hematophagy.

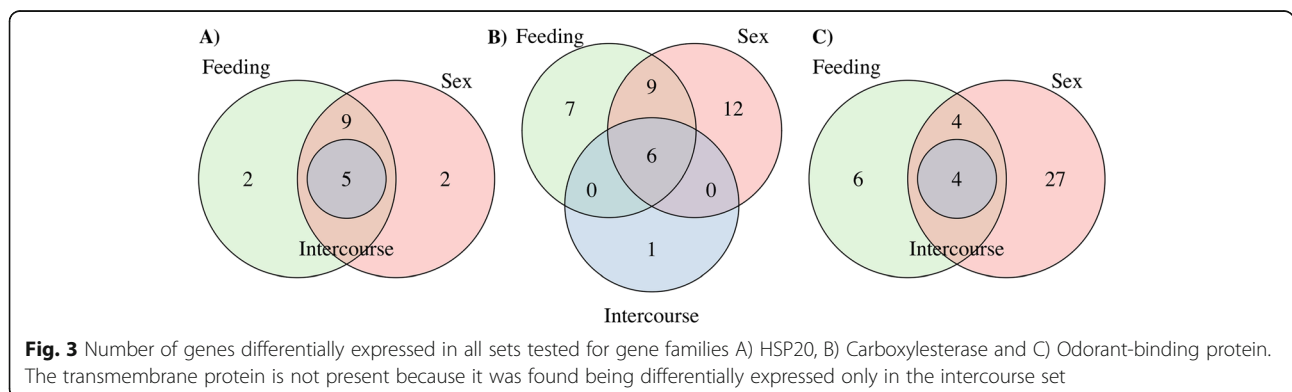
We found four rapidly evolving families in at least two hematophagous independently-evolved lineages. Gene

family EOG090W0E11 rapidly expanded independently in Culicidae and Cimicidae orders and it was the only true example of convergent evolution, where the expansion occurred after the transition to hematophagy. According to UniProt annotation, it belongs to the HSP20 family. Previous works showed high expression levels of other HSP, HSP70, after blood-feeding in mosquitoes, kissing bugs and bed bugs, suggesting that HSP protects their midguts from the stress caused by the warm blood ingested [29, 30].

Gene family EOG090W022L also rapidly expanded independently in Culicidae and Psychodidae orders and, according to UniProt annotation, it is a carboxylesterase. Carboxylesterase is a detoxification gene previously associated with resistance to insecticide in mosquitoes [31, 32] and is hypothesized that this expansion may help these groups to detoxify the huge amount of amino acids and heme following a blood-feeding [33].

Notably, our results show that hematophagous species have a slower  $\lambda$  rate than non-hematophagous species, contrary to the assumption that a transition to a new lifestyle would favor expansions and/or contractions. However, it is likely that many non-hematophagous species present a great number of expansions and contractions due several reasons. For example, the milkweed bug genome [34] shows expansions in the number of gene copies related to CSPs and retention of enzymes involved in the amino acid metabolism.

On the expression patterns of rapidly-evolved genes found in VectorBase and their role in the evolution of hematophagy, our results are unexpected. VectorBase data rely on Culicidae expression studies, thus we would expect more genes overexpressed in females, mated and blood-feeding insects [35]. However, from all sets, only intercourse set showed more expressed genes in the expected category, mated individuals. The other two sets, feeding, and sex, showed more expressed genes in the unexpected categories of male and non-blood-feeding individuals, respectively (Table 2).



**Table 2** Number of overexpressed genes found in VectorBase according to categories and sets

	Sex			Intercourse			Feeding		
	Female	Both sets	Male	Mated	Both sets	Virgin	Blood	Both sets	Non-blood
Rapid evolving families	13	4	26	9	2	1	9	11	18
Slow evolving families	10	16	9	2	2	1	2	1	11

Although expansions in salivary gland proteins in hematophagous insects are well reported in literature [9, 10, 36], none of the two expanded genes families found in this work are a classic salivary gland protein. A possible explanation is that we focused our searches on rapid expansions and contractions shared by two or more independently-evolved blood-feeding lineages, and salivary gland genes expansions are usually specific to each lineage [11, 12]. Besides it, the orthology assessment from OrthoDB may differ from the orthology assessment of salivary gland proteins, which is a well-studied group of proteins regarding hematophagous species [11] and where orthology is usually manually reviewed [37]. Indeed, from the 53 salivary proteins distributed in 24 gene families found in 19 *Anopheles* genomes by Àrca et al. (2017) [37], OrthoDB distributes them in 38 orthologous groups (gene families), which probably decreased turnovers rates inferred by CAFE for salivary proteins in our study. Only two salivary proteins from Àrca et al. (2017) [37] are not present in OrthoDB: a cE5-Anophelin and a hyp8.2.

Regarding the two gene families with rapid contractions, their functions are not well-annotated as gene families with rapid expansions. The only annotation for EOG090W0LR7 classified it as an integral component of membrane, while EOG090W0JL7 is annotated as a CSP but none function is associated with it. Assuming its role as a CSP, we cannot state that the contraction inferred in EOG090W0JL7 represents an adaptation. For example, chemosensory genes have a wide variation in the number of copies in Insecta, ranging from 1 to 265, where expansions and contractions both happened several times [38]. Even the number of CSPs copies found within Culicidae have a large variation:  $6.8 \pm 1.1$  CSPs were found in 19 species of the Anophelinae sub-family, while  $51 \pm 28.8$  CSPs were found in three species in the Culicinae sub-family [39]. On the expression patterns of slowly-evolved genes found in VectorBase and their role in the evolution of hematophagy, Table 2 does not show a clear pattern. Only the feeding set displays differences between distinct categories, showing more overexpressed genes in non-blood-feeding individuals compared with blood-feeding individuals.

Previously, several works explored the relationship between gene families expansions and hematophagy [10,

20, 37, 40], but they usually focused only in one species/group or few gene families and they did not employ a statistical approach to model the evolution of gene families. Two exceptions are the works of Neafsey et al. (2015) [9] and Thomas et al. (2020) [41]. The first described the evolution of gene families within Anophelinae and found a  $\lambda$  rate of  $3.2 \times 10^{-3}$  along the genome, estimating a relatively conserved number of CSP across this group.

Compared with Neafsey et al. (2015) [9], our  $\lambda$  rate for hematophagous lineages is approximately three times lower and we did find a contraction in CSP for *Anopheles gambiae*. However, our dataset spans a much deeper time-scale and the contraction in *A. gambiae* represents the loss from the ancestor of Anophelinae and Culicinae, which is not the case of Neafsey et al. (2015) [9]. Thomas et al. (2020) [41], who also explored a deep-time scale of insects, inferred the  $\lambda$  rate for several families (Table S13 on their work), where the mean found was  $1.18 \times 10^{-3}$ , similar to our estimates ( $1.12 \times 10^{-3}$  and  $1.17 \times 10^{-3}$  for hematophagous and non-hematophagous lineages, respectively).

## Conclusions

In this work, we investigated the evolution of gene families as a first approach to identify convergent evolution in blood-feeding insects. The significant turnover in gene numbers found in four gene families in more than one hematophagous lineage indicates evidence of convergent molecular evolution related to blood-feeding. Following this line, we also would expect these genes to be differentially expressed after blood-meal, intercourse or in females. Here we found more genes expressed in mated individuals compared with virgin individuals in rapidly-expanded families and more genes expressed in non-blood-feeding individuals compared with blood-feeding individuals in rapidly-contracted families.

Beyond revealing a new set of candidate genes to be explored in further analysis to help the development of new strategies to deal with blood-feeding vectors, our findings suggest that the same strategy could be efficiently used in more than one vector, due to convergent evolution, and also present a new perspective to study the evolution of hematophagy identifying convergent molecular patterns.

## Methods

### Genomic data set

All data used in this study was obtained through OrthoDB v. 9.1 [42]. We generated single-copy-only gene families (one gene copy per species per gene family) and a multiple-copies gene family matrix (one or more gene copies per species per gene family) for Metazoa. The first was used to estimate the divergence time, while the latter was used to estimate changes (gain and loss) in the evolution of gene families. We selected 33 Arthropoda species (Fig. 1) plus *Daphnia pulex* as an outgroup. We did not include *Bactrocera* and *Polypedilum* genus because other species within their families were already present.

### Divergence time estimation

We used Misof et al. (2014) [4] phylogenetic tree as input to our order-relationships. Within Diptera and Hemiptera orders, the relationships were based on Song et al. (2016) and Weirauch et al. (2019) [43, 44]. From the 38,842 orthologous groups (OGs) available at OrthoDB we filtered out species with multiple copies and aligned each OG with MAFFT v7.407 (FFT-NS-1 algorithm) [45]. Then we removed gap-rich regions with Gblocks v. 0.91b (minimum length of block = 5 and up to 50% gap positions were allowed) [46]. We only selected genes where more than 50% of amino acid sites were kept by GBlocks and more than 17 species were present in the alignment, resulting in a concatenated matrix of 683 genes and 288,500 amino acid sites.

Using this assembled tree and alignment we used MCMCtree v. 4.9 h [47] to perform a Bayesian divergence time estimation with the approximate likelihood calculation described in dos Reis & Yang (2011) [48]. Analyses were performed twice to check MCMC convergence, and fossil calibration priors followed Misof et al. (2014) [4].

### Gene families evolution

We used CAFE v. 4.1 [49] to identify gene families with rapid expansions and/or contractions in copy numbers along hematophagous-only lineages for the 38,842 OGs available at OrthoDB. Using the timetree estimated earlier and the number of gene copies per family, CAFE first estimates the distribution of error ( $\epsilon$ ) in the number of gene copies present in the dataset. Then, accounting to  $\epsilon$ , CAFE infers the birth and death ( $\lambda$ ) rate, which represents the average rate of genomic turnover (gains and losses) per gene per million years [50].

In this work, we compared a global model, with one  $\lambda$  rate for all branches, and the other with two  $\lambda$  rates, split between hematophagous and non-hematophagous branches. To compare both models and infer which one fits better the data, CAFE employs a set of simulations

for the global model. For each simulation, it performs a likelihood-ratio test (LRT) that will be used to draw a null distribution where the real LRT will be compared. Finally, we annotated the function of these genes using UniProt [51] and VectorBase [52] databases. We also used VectorBase to identify which of these genes were statistically differentially expressed in transcriptomes according to three sets: feeding (blood-feed vs. non-blood-feed individuals), intercourse (mated vs. virgin individuals) and sex (female vs. male) patterns, all of them related with the blood-feeding behavior [35].  $p$ -values < 0.05 according each study present in VectorBase were considered significant.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12862-020-01650-3>.

### Additional file 1.

### Abbreviations

HSP20: Small heat shock protein HSP20; CSP: Chemosensory protein family; OG: Orthologous group

### Acknowledgements

We would like to thank the reviewers and editors for improving the manuscript.

### Authors' contributions

LF and MFN designed the study, collected and analyzed the data, and wrote the final draft. All authors have read and approved the manuscript.

### Funding

This work was supported by the São Paulo Research Foundation (FAPESP) [Grant numbers 2017/25058–2, 2018/19682–8 and 2015/18269–1]. The funding body had no role in the design of the study and collection, analysis, and interpretation of data and in the writing of the manuscript.

### Availability of data and materials

The datasets generated and/or analysed during the current study are available in <https://github.com/freitas-lucas/GenFamBlood>.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 13 March 2020 Accepted: 6 July 2020

Published online: 17 July 2020

## References

1. Boulton RA. Hematophagy. In: Encyclopedia of Animal Cognition and Behavior. Springer, Cham; 2018. p. 1–5.
2. Shepard DS, Coudeville L, Halasa YA, Zambrano B, Dayan GH. Economic impact of dengue illness in the Americas. *Am J Trop Med Hyg.* 2011;84:200–7.
3. WHO. Vector-borne diseases. World Health Organization. 2017. <https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases>. Accessed 10 Oct 2019.

4. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science*. 2014;346:763–7.
5. Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim J-W, Lambkin C, et al. Episodic radiations in the fly tree of life. *Proc Natl Acad Sci U S A*. 2011;108:5690–5.
6. Tedrow RE, Rakotomanga T, Nepomichene T, Howes RE, Ratovonjato J, Ratsimbaoa AC, et al. *Anopheles* mosquito surveillance in Madagascar reveals multiple blood feeding behavior and Plasmodium infection. *PLoS Negl Trop Dis*. 2019;13:e0007176.
7. Alencar J, Dégallier N, Hannart A, Silva J dos S, Pacheco JB, Guimarães AE. Circadian and seasonal preferences for hematophagy among *Haemagogus capricornii*, *Hg. janthinomys*, and *Hg. leucocelaenus* (Diptera: Culicidae) in different regions of Brazil. *J Vector Ecol*. 2008;33:389–92.
8. Oliveira MF, Silva JR, Dansa-Petretski M, de Souza W, Lins U, Braga CMS, et al. Haem detoxification by an insect. *Nature*. 1999;400:517–8.
9. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science*. 2015;347:1258522.
10. Mesquita RD, Vionette-Amaral RJ, Lowenberger C, Rivera-Pomar R, Monteiro FA, Minx P, et al. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc Natl Acad Sci U S A*. 2015;112:14936–41.
11. Arcà B, Ribeiro JM. Saliva of hematophagous insects: a multifaceted toolkit. *Curr Opin Insect Sci*. 2018;29:102–9.
12. Ribeiro JMC, Mans BJ, Arcà B. An insight into the sialome of blood-feeding Nematocera. *Insect Biochem Mol Biol*. 2010;40:767–84.
13. Calvo E, Mans BJ, Andersen JF, Ribeiro JMC. Function and evolution of a mosquito salivary protein family. *J Biol Chem*. 2006;281:1935–42.
14. Foote AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, et al. Convergent evolution of the genomes of marine mammals. *Nat Genet*. 2015;47:272–5.
15. Partha R, Kowalczyk A, Clark NL, Chikina M. Robust method for detecting convergent shifts in evolutionary rates. *Molecular Biology and Evolution*. 2019;36:1817–30.
16. Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, et al. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science*. 2018;361.
17. Sackton TB, Clark N. Convergent evolution in the genomics era: new insights and directions. *Philos Trans R Soc Lond B Biol Sci*. 2019;374:20190102.
18. Nyanjom SG, Tare C, Wamunyokoli F, Obiero GF. Expression Levels of Odorant Receptor Genes in the Savanna Tsetse Fly, *Glossina morsitans morsitans*. *J Med Entomol*. 2018;55:855–61.
19. Pelletier J, Leal WS. Characterization of olfactory genes in the antennae of the Southern house mosquito, *Culex quinquefasciatus*. *J Insect Physiol*. 2011;57:915–29.
20. Martínez-Barnette J, Lavore A, Beliera M, Téllez-Sosa J, Zumaya-Estrada FA, Palacio V, et al. Adaptations in energy metabolism and gene family expansions revealed by comparative transcriptomics of three Chagas disease triatomine vectors. *BMC Genomics*. 2018;19:296.
21. Henriques BS, Gomes B, da Costa SG, da Silva MC, Mesquita RD, Dillon VM, et al. Genome wide mapping of peptidases in *Rhodnius prolixus*: identification of protease gene duplications, horizontally transferred proteases and analysis of peptidase A1 structures, with considerations on their role in the evolution of hematophagy in Triatominae. *Frontiers in Physiology*. 2017;8.
22. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*. 2019;47:D351–60.
23. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47:D427–32.
24. Ollivier M, Tresselt A, Bastian F, Lagoutte L, Axelsson E, Arendt M-L, et al. Amy2B copy number variation reveals starch diet adaptations in ancient European dogs. *Royal Society Open Science*. 2016;3:160449. <https://doi.org/10.1098/rsos.160449>.
25. Cheng T, Wu J, Wu J, Chilukuri RV, Huang L, Yamamoto K, et al. Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nature Ecology & Evolution*. 2017;1:1747–56. <https://doi.org/10.1038/s41559-017-0314-4>.
26. Sharma V, Hecker N, Roscito JG, Foerster L, Langer BE, Hiller M. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nat Commun*. 2018;9:1215.
27. Graça-Souza AV, Maya-Monteiro C, Paiva-Silva GO, Braz GRC, Paes MC, Sorgine MHF, et al. Adaptations against heme toxicity in blood-feeding arthropods. *Insect Biochem Mol Biol*. 2006;36:322–35.
28. Mans BJ. Evolution of vertebrate hemostatic and inflammatory control mechanisms in blood-feeding arthropods. *J Innate Immun*. 2011;3:41–51.
29. Benoit JB, Lopez-Martinez G, Patrick KR, Phillips ZP, Krause TB, Denlinger DL. Drinking a hot blood meal elicits a protective heat shock response in mosquitoes. *Proc Natl Acad Sci U S A*. 2011;108:8026–9.
30. Paim RMM, Araujo RN, Leis M, Sant'anna MRV, Gontijo NF, Lazzari CR, et al. Functional evaluation of Heat Shock Proteins 70 (HSP70/HSC70) on *Rhodnius prolixus* (Hemiptera, Reduviidae) physiological responses associated with feeding and starvation. *Insect Biochem Mol Biol*. 2016;77:10–20.
31. Wu X-M, Xu B-Y, Si F-L, Li J, Yan Z-T, Yan Z-W, et al. Identification of carboxylesterase genes associated with pyrethroid resistance in the malaria vector *Anopheles sinensis* (Diptera: Culicidae). *Pest Manag Sci*. 2018;74:159–69.
32. Main BJ, Everitt A, Cornel AJ, Hormozdiari F, Lanzaro GC. Genetic variation associated with increased insecticide resistance in the malaria mosquito, *Anopheles coluzzii*. *Parasit Vectors*. 2018;11:225.
33. Sterkel M, Oliveira JHM, Bottino-Rojas V, Paiva-Silva GO, Oliveira PL. The dose makes the poison: nutritional overload determines the life traits of blood-feeding arthropods. *Trends Parasitol*. 2017;33:633–44.
34. Panfilio KA, Vargas Jentszsch IM, Benoit JB, Erezylmaz D, Suzuki Y, Colella S, et al. Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Genome Biol*. 2019;20:64.
35. Roy DN. On the role of blood in ovulation in *Aedes aegypti*. *Linn. Bulletin of Entomological Research*. 1936;27:423–9.
36. Ribeiro JMC, Martin-Martin I, Arcà B, Calvo E. A Deep insight into the sialome of male and female *Aedes aegypti* mosquitoes. *PLoS One*. 2016;11:e0151400.
37. Arcà B, Lombardo F, Struchiner CJ, Ribeiro JMC. Anopheline salivary protein genes and gene families: an evolutionary overview after the whole genome sequence of sixteen *Anopheles* species. *BMC Genomics*. 2017;18:153.
38. Sánchez-Gracia A, Vieira FG, Rozas J. Comparative genomics of the major chemosensory gene families in arthropods. *Encyclopedia of Life Sciences: Almeida FC*; 2011.
39. Mei T, Fu W-B, Li B, He Z-B, Chen B. Comparative genomics of chemosensory protein genes (CSPs) in twenty-two mosquito species (Diptera: Culicidae): Identification, characterization, and evolution. *PLoS One*. 2018;13:e0190412.
40. Wu D-D, Wang G-D, Irwin DM, Zhang Y-P. A profound role for the expansion of trypsin-like serine protease family in the evolution of hematophagy in mosquito. *Mol Biol Evol*. 2009;26:2333–41.
41. Thomas GWC, Dohmen E, Hughes DST, Murali SC, Poelchau M, Glastad K, et al. Gene content evolution in the arthropods. *Genome Biol*. 2020;21:15.
42. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, et al. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res*. 2017;45:D744–9.
43. Song N, An S, Yin X, Cai W, Li H. Application of RNA-seq for mitogenome reconstruction, and reconsideration of long-branch artifacts in Hemiptera phylogeny. *Sci Rep*. 2016;6:33465.
44. Weirauch C, Schuh RT, Cassis G, Wheeler WC. Revisiting habitat and lifestyle transitions in Heteroptera (Insecta): insights from a combined morphological and molecular phylogeny. *Cladistics*. 2019;35:67–105.
45. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
46. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000;17:540–52.
47. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24:1586–91.
48. Reis MD, Yang Z. Approximate likelihood calculation on a phylogeny for Bayesian Estimation of Divergence Times. *Mol Biol Evol*. 2011;28:2161–72.
49. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol*. 2013;30:1987–97.
50. Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res*. 2005;15:1153–60.

51. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019;47:D506–15.
52. Giraldo-Calderón GI, Emrich SJ, MacCallum RM, Maslen G, Dialynas E, Topalis P, et al. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res.* 2015;43(Database issue):D707–13.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

