Evolution of genes involved in feeding preference in Calliphoridae (Diptera: Calyptratae)

Gisele Antoniazzi Cardoso [1], **Marco Antonio Tonus Marinho** [2], **Raquel Dietsche Monfardini** [1], **Ana Lima de Azeredo Espin** [3], **Tatiana Teixeira Torres** Corresp. [1]

[1] Department of Genetics and Evolutionary Biology, University of Sao Paulo, Sao Paulo, Sao Paulo, Brazil

[2] Department of Biology, Faculty of Philosophy, Science, and Letters, University of Sao Paulo, Ribeirao Preto, Sao Paulo, Brazil

[3] Department of Genetics, Evolution, and Bioagents, State University of Campinas, Campinas, Sao Paulo, Brazil

Corresponding Author: Tatiana Teixeira Torres
Email address: tttorres@usp.br

Background. The genotype-phenotype interactions among traits governing feeding preference are of fundamental importance to behavioral genetics and evolutionary biology. The genetic basis of behavioral traits has been explored in different taxa using different approaches. However, the complex nature of the genetic mechanisms undergirding behavior is poorly understood. Here, we present an evolutionary study of candidate genes related to parasitism in Calliphoridae (Diptera: Calyptratae). Closely related species in this family exhibit distinct larval feeding habits, most notably necro-saprophagy and obligate parasitism. **Methods**. To understand the genetic and molecular bases underlying these habits, expression levels of eight candidate genes for feeding behavior—*Cyp6g2, foraging, glutamate dehydrogenase, Jonah65aiv, Malvolio, PGRP-SC2*, *RPS6-p70-protein kinase,* and *smooth*—were measured in four species using qPCR. Moreover we used expression values and sequence information to reconstruct the relationship among species and the d_N/d_S rate to infer possible sites under selection. **Results.** For most candidate genes, no statistically significant differences were observed, indicating a high degree of conservation in expression. However, *Malvolio* was differentially expressed between habits. Evolutionary analyses based on transcript levels and nucleotide sequences of *Malvolio* coding region suggest that transcript levels were correlated to feeding habit preferences among species, although deviations under a strictly neutral model were also observed in statistical tests. **Discussion.** *Malvolio* was the only gene demonstrating a possible connection to feeding habit. Differences in gene expression may be involved in (or be a result of) the genetic regulation of Calliphoridae feeding habit. Our results are the first steps towards understanding the genetic basis and evolution of feeding behavior in Calliphoridae using a functional approach.

Evolution Of Genes Involved In Feeding Preference In Calliphoridae (Diptera: Calyptratae)

Gisele Antoniazzi Cardoso[*], Marco Antonio Tonus Marinho[†], Raquel Dietsche Monfardini[*], Ana Maria Lima Azeredo Espin[‡], Tatiana Teixeira Torres[*,§]

[*]Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo (USP), São Paulo, SP, Brazil, 05508-090.

[†]Department of Biology, Faculty of Philosophy, Science, and Letters, University of São Paulo, Ribeirão Preto, SP, Brazil, 14040-901.

[‡]Department of Genetics, Evolution, and Bioagents, State University of Campinas (UNICAMP), Campinas, SP, Brazil, 13083-875.

[§]Research Center on Biodiversity and Computing (BioComp-USP), University of São Paulo, São Paulo, SP, Brazil,05508-900.

Corresponding author: Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo (USP), Rua do Matão, 277, 05508-090, São Paulo, SP, Brazil. + 55-11-3091-8759. E-mail: tttorres@ib.usp.br

23

24

25 ABSTRACT

26 **Background.** The genotype-phenotype interactions among traits governing feeding preference
27 are of fundamental importance to behavioral genetics and evolutionary biology. The genetic
28 basis of behavioral traits has been explored in different taxa using different approaches.
29 However, the complex nature of the genetic mechanisms undergirding behavior is poorly
30 understood. Here, we present an evolutionary study of candidate genes related to parasitism in
31 Calliphoridae (Diptera: Calyptratae). Closely related species in this family exhibit distinct larval
32 feeding habits, most notably necro-saprophagy and obligate parasitism.
33 **Methods**. To understand the genetic and molecular bases underlying these habits, expression
34 levels of eight candidate genes for feeding behavior—*Cyp6g2, foraging, glutamate*
35 *dehydrogenase, Jonah65aiv, Malvolio, PGRP-SC2*, *RPS6-p70-protein kinase,* and *smooth*—were
36 measured in four species using qPCR. Moreover we used expression values and sequence
37 information to reconstruct the relationship among species and the d_N/d_S rate to infer possible
38 sites under selection.
39 **Results.** For most candidate genes, no statistically significant differences were observed,
40 indicating a high degree of conservation in expression. However, *Malvolio* was differentially
41 expressed between habits. Evolutionary analyses based on transcript levels and nucleotide
42 sequences of *Malvolio* coding region suggest that transcript levels were correlated to feeding
43 habit preferences among species, although deviations under a strictly neutral model were also
44 observed in statistical tests.

Discussion. *Malvolio* was the only gene demonstrating a possible connection to feeding habit. Differences in gene expression may be involved in (or be a result of) the genetic regulation of Calliphoridae feeding habit. Our results are the first steps towards understanding the genetic basis and evolution of feeding behavior in Calliphoridae using a functional approach.

INTRODUCTION

One of the most interesting and complex phenotypic traits that challenges molecular biologists is behavior. Behavioral traits are very plastic and can be influenced by multiple genes, organism needs, and environmental factors. Despite the complexity of behavioral genetics, previous studies were able to successfully identify genes underlying specific behaviors, especially in model species as *Drosophila melanogaster* (Sokolowski, 2001). A large number of genes have been described to play a role in behaviors including aggressiveness (*fruitless*), courtship dynamics (*courtless*), and circadian rhythms (*period*) (Greenspan and Ferveur, 2000; Ryner et al., 1996; Waddell and Quinn, 2001). Understanding the genetic complexity and evolution of these traits is of fundamental importance to elucidate the behavioral influences on ecological diversification, adaptation, and speciation. Variations in gene expression patterns are responsible for a large proportion of phenotypic differences among species (Enard et al., 2002; King and Wilson, 1975). These differences may be due to mutations in coding or regulatory regions or changes in the chromatin structure that affect accessibility of transcriptional

machinery to the genetic template (Enard et al., 2002; Yang et al., 2006). Traditionally, researchers have focused on coding regions, but there is now a growing number of studies investigating evolution through differentiation of gene expression patterns. Many of these studies have attempted to address the question of how phenotypic differences among closely related species are connected with differences in gene expression. In Darwin finches, for example, differences in expression patterns of the *bone morphogenetic protein 4* (*Bmp4*) and *calmodulin* (*CaM*) genes have been shown to be involved in beak shape variation among closely related species, a trait that directly influences diet and exploitation of food resources (Abzhanov et al., 2006; Abzhanov et al., 2004). In addition, regulatory differences in the *Pituitary homeobox 1* gene (*Pitx1*) between marine and fresh water populations of *Gasterosteus aculeatus* have been shown to be associated with the progressive reduction of pelvic spines, which have an adaptive function in defense against predation (Shapiro et al., 2004). These differences in regulatory regions are to a large extent responsible for the morphological variation on which natural selection can act.

In this context, the Calliphoridae family (Diptera: Calyptratae) constitutes an ideal system to investigate differential patterns of gene expression related to the evolution of feeding behavior. Most species in this family exhibit two main feeding habits in their immature stage: (1) necro-saprophagy, in which larvae feed on animal decaying organic matter; and (2) parasitic, in which larvae can in an obligate or facultative manner feed on living tissues of warm-blooded vertebrates, causing infestations called myiasis (F, 1965; Hall and Wall, 1995).

The necro-saprophagous habit of blowflies seems to comprise the ground-plan for the group and, according to the current phylogenetic hypothesis for this family, obligate parasitism appears to have evolved independently at least three times (Stevens and Wallman, 2006). In the

most parsimonious scenario, the primitive necro-saprophagous habit evolved to an obligate parasitic habit through the appearance of opportunistic infestations of living animals, which were weak or with partly necrotic wounds (F, 1965; YZ, 1989). Two Calliphoridae genera in the subfamily Chrysomyinae, *Chrysomya* and *Cochliomyia*, contain both obligate parasites and necro-saprophagous species, allowing the comparison of different feeding habits among closely related species.

Here, we use a candidate gene approach as a first step to investigate the evolution of feeding habits in Calliphoridae. The qPCR analyses of transcripts from eight candidate genes in four closely related species, *Chrysomya albiceps*, *Chrysomya megacephala*, *Cochliomyia macellaria* (all with necro-saprophagous larvae), and *Cochliomyia hominivorax* (whose larvae are obligate parasites of mammals) were used to determine genetic associations with different feeding habits and the types of selection potentially influencing gene expression. Furthermore, the results were corroborated with molecular evolution and sequence analyses of the candidate genes.

MATERIALS AND METHODS

Candidate gene selection and primer design: Candidate genes were selected from previous studies in Diptera (Table S1). To expand our search, the Gene Ontology (GO) database was also used (available at http://www.geneontology.org). Sequences of the selected genes were obtained from Flybase (http://flybase.org/). The sequences of the twelve *Drosophila* species with whole genome sequenced (Clark et al., 2007) were aligned against all *C. hominivorax* transcriptome contigs (Carvalho et al., 2010) to identify corresponding orthologs. Identified contigs were aligned against the BLAST database repository using blastn (Altschul et al., 1990)

to search for orthologs in closely related species. Primers for each gene were then designed in conserved regions among the selected orthologs with Primer3 (Rozen and Skaletsky, 2000).

RNA isolation and cDNA synthesis: Live flies were obtained from the rearing facilities at the Laboratory of Animal Genetics and Evolution (CBMEG/UNICAMP). Rearing conditions were as described in a previous study (Cardoso et al., 2014). RNA extractions were performed on two separate generations of larvae, and adult males and females of each species (*C. albiceps*, *C. megacephala*, *C. hominivorax*, and *C. macellaria*). Total RNA was extracted from 10 individuals from each generation using TRIzol® reagent (Invitrogen, USA) according to the manufacturer's protocol. For subsequent analyses, individuals of the same generation were combined into two RNA samples, each containing five individuals, for a total of four biological replicates.

All RNA extractions were treated with TURBODNase (Ambion® Life Technologies, ThermoFisher, MA USA) to degrade any remaining genomic DNA. DNase was inactivated by heating samples at 75° for 10 min followed by the addition of EDTA to a final concentration of 2.5 mM. After DNAse treatment, a PCR was performed using *Rp49* primers (as described in the Amplification section) to test the treated samples.

Total RNA was quantified with a Qubit® fluorometer (Invitrogen™, ThermoFisher, MA USA) using the Qubit RNA® assay kit. Synthesis of cDNA was performed using the First Strand cDNA Synthesis Kit (ThermoScientific™, ThermoFisher, MA USA), following the manufacturer's protocol and using 0.5 µg of total RNA as template.

PCR amplification of selected genes: To test the candidate gene primers (Table S2), a PCR was performed using the cDNA of the four Calliphoridae species as template. PCR reactions were conducted for a final volume of 20 µL containing 2.0 mM $MgCl_2$, 0.6 µM of each primer, 200 µM dNTPs, and 1 U *Taq* DNA polymerase (Thermo Scientific, MA, USA). Reaction

conditions consisted of an initial denaturing step of 3 min at 94 C °, followed by 35 cycles of 50s at 94 ºC, 30s at 60 ºC, and 30s at 72 ºC, and a final extension step at 72 ºC for 5 min.

Quantitative PCR: qPCR reactions were performed in a final volume of 12.5 µL using the SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA) according to manufacturer's instructions. In each reaction, 1 µL of cDNA was used and primers were added to a final concentration of 200 nM. The qPCRs were run in two technical replicates to assess intra-assay variation on an ABI 7500 Fast Real Time PCR System (Applied Biosystems, Foster City, CA) using the following cycling conditions: 2 min at 50 ºC, 10 min at 95 ºC, and 40 cycles consisting of 15 min at 95 ºC and 60s at 60 ºC. After the 40th cycle, the temperature was gradually increased from 60 ºC to 95 ºC (1 °C per minute for 35 min) to obtain a dissociation curve. This step was performed to confirm the formation of a single product during the reaction. A five point standard curve was used to access the efficiency of each primer. Efficiency values were calculated according to the equation $E= 10^{1/slope}-1$ (Higuchi et al., 1993).

Relative expression levels were calculated using the Mean Normalized Expression (MNE) formula proposed by Simon (2003). MNE is the ratio of the mean of the reference and target genes Cycle thresholds (*Ct*s), including efficiency values: $MNE = (E_{reference})^{Ct\ reference}/(E_{target})^{Ct\ target}$. All expression data were normalized using the mean of the reference genes *Gapdh* and *Rp49*, both validated in a previous study (Cardoso et al., 2014).

Multivariate analyses: Statistically significant differences ($\alpha = 0.01$) in gene expression levels between the studied species and different feeding habits were determined using two-way Analyses of Variance (ANOVA) (species × feeding habit). The MNE values were used as input data for all statistical analyses. All tests were performed in the statistical package R (Team, 2008). Genes differently expressed among species were candidates to be under neutral evolution

160 or positive selection and genes with no significant difference were candidates to be under
161 purifying selection.

162 **Gene expression analyses**: To test the hypothesis that species divergence in gene
163 expression evolves neutrally, pair-wise differences in mRNA transcript levels were compared
164 with a matrix of pair-wise genetic distances using linear regression analyses and the Pearson's
165 correlation index.

166 Differences in mRNA levels were calculated using the square difference between each
167 species pair. Pair-wise genetic distances among given species were based on a published
168 concatenated dataset of COI, 16S, 28S, and ITS2 sequences (Marinho et al., 2012), and estimated
169 using the Maximum Composite Likelihood method (with the TN93+G substitution model) in
170 MEGA 5 (Tamura et al., 2011).

171 **Expression trees (clustering analyses)**: Pair-wise differences (absolute values) in
172 expression levels among species, sexes, and developmental stages for each candidate gene were
173 used to generate distance matrices for clustering analyses. These matrices were then analyzed
174 using FastME software (Desper and Gascuel, 2002), which computes distance trees using the
175 Minimum Evolution principle. Cluster diagrams (expression trees) were estimated separately for
176 sex and developmental stage and with combined data. The resulting trees were then compared
177 with the known phylogenetic relationships among these species, used as a null hypothesis for the
178 observed pattern of gene expression data.

179 **Sanger sequencing of candidate genes**: Five of the candidate genes (*for*, *Gdh*, *Jon65aiv*,
180 *Mvl,*and *S6k*) showing the most promising results in the expression analyses were selected for
181 further investigation using coding sequence data. Primers for PCR-amplification and sequencing
182 of these gene regions were designed (Table S3), using the previously synthesized cDNA as a

183 template.

184 PCR reactions of coding regions were performed in a 25 μL final volume, containing
185 0.45 μM of each primer, 200 μM dNTPs, and 1.25 U of DreamTaq DNA polymerase (Thermo
186 Scientific, MA USA). Reaction conditions comprised an initial denaturing step of 3 min at 92 °,
187 followed by 30 cycles of 30 s at 95 º, 30 s at specific primers' annealing temperatures (Table S3)
188 and 30 s at 72 º, witha final step of extension at 72 º for 5 min.

189 PCR products were purified with Exonuclease I and FastAP Alkaline Phosphatase
190 (ExoFap, Thermo Scientific) and sequenced in an ABI 3730 DNA Analyzer Sanger platform
191 (Applied Biosystems, Foster City, CA) using the BigDye® Terminator v3.1 Cycle Sequencing
192 Kit (Life Technologies, CA USA).

193 **Phylogenetic analyses (sequence dendrograms)**: Sequences of the five genes obtained
194 in the previous section were used to infer phylogenetic relationships among the species under
195 study. Sequence alignments were generated using MAFFT v 7.149 software (Katoh et al., 2002;
196 Katoh and Standley, 2013) and resultant alignments were checked for out-of-frame indels and
197 premature stop codons using the "Alignment Explorer" tool available in MEGA 6 software
198 (Tamura et al., 2013). Maximum-likelihood (ML) and Bayesian Inference (BI) trees were
199 inferred using GARLI v2.0 (DJ, 2006) and MrBayes v3.2.3 (Ronquist et al., 2012), respectively.
200 In both analyses, the substitution model was estimated using jModelTest v2.1.5 (Darriba et al.,
201 2012), and the best-fit model was selected for analyses. Resulting phylogenies were compared
202 with the known phylogenetic relationships among the species, the latter of which were used as
203 the null hypothesis for primary sequence evolution (i.e., neutral if coding sequences have
204 evolved following lineage diversification; selection if there are forces constraining evolution
205 based on larval alimentary habit).

206 Moreover, a maximum-likelihood (ML) tree was constructed for the *Mvl* gene (as
207 described above) using the sequences of eight flies with different feeding habits: necro-
208 saprophagous (*C. albiceps*, *C. megacephala*, *C. macellaria,* and *Lucilia sericata*), obligate
209 myiasis-causing parasites (*C. hominivorax*, *Dermatobia hominis,* and *Oestrus ovis*), facultative
210 myiasis-causing parasites (*Lucilia cuprina*), obligate hematophagous parasites (*Glossina*
211 *morsitans*), and generalists (*Musca domestica*).

212 **Coding sequence evolution**: To perform a quantitative evaluation of the neutral
213 hypothesis of sequence evolution, the Tajima's Relative Rates Test (Tajima, 1993) was
214 conducted as implemented in MEGA v.6 (Tamura et al., 2013). This test provides a statistical
215 framework to compare the evolutionary rates of two ingroup sequences in relation to a third
216 (outgroup) sequence, and checks for deviations from the molecular clock hypothesis. For these
217 analyses, the two *Cochliomyia* species (one necro-saprophagous and one obligate parasite) were
218 defined as the ingroup taxa, and both *Chrysomya* species were tested as the outgroup.

219 In addition, a maximum-likelihood method was used to estimate the ratio of the rate of
220 non-synonymous (non-silent) substitutions per site to the rate of synonymous (silent)
221 substitutions per site (d_N/d_S, *Ka/Ks*, or ω) for *Mvl* and to obtain the likelihoods ratio, using as
222 input a neighbor joining (NJ) phylogeny. The codeml program from the PAML v.4 package
223 (Yang, 2007) was used to test five models: a) the null expectation from a neutral model with a
224 fixed ω = 1 (neutral model); b) average ω for the whole tree (model 1); c) specific ω for *C.*
225 *hominivorax* and an average ω for all other species (model 2); d) an average ω for all parasites
226 (*C. hominivorax*, *D. hominis*, *G. morsitans*, *L. cuprina,* and *O. ovis*) and an average ω for the
227 remaining species (model 3); and e) average ω for parasites feeding on living tissues and blood
228 (*C. hominivorax*, *D. hominis,* and *G. morsitans*) and an average ω for all other species (model 4).

The likelihood estimate for the neutral model was compared to the likelihoods of each model in test (model1-model4) using a χ^2 statistics. The possible evolutionary processes involved in sequence evolution (neutral: $\omega = 1$, purifying selection $\omega < 1$ and positive selection $\omega > 1$) were evaluated for the significant comparisons ($\alpha = 0.05$).

Phylogenetic independent contrasts analysis: To evaluate the effects of autocorrelation in feeding habits due to phylogenetic relationships among the species under study, we performed a phylogenetic independent contrasts (PIC) analysis. For the PIC analysis, the average MNE for the four replicates per species and the phylogeny based on Marinho et al. (2012) were used as inputs in the ape package of the R environment (Paradis et al., 2004). Each species was assigned a number corresponding to its feeding preference (0: obligate parasite and 1: necro-saprophagous).

RESULTS

Candidate gene selection: From the 21 possible candidates, eight genes were amplified without ambiguities and with high qPCR efficiency: *Cyp6g2*, *foraging*, *Glutamate dehydrogenase*, *Jonah 65 aiv*, *Malvolio*, *PGRP-SC2*, *RPS6-p70-protein kinase,* and *smooth* (see Tables S1-S2 for details).

Expression analyses: We used the comparative *Ct* method with efficiency correction (Mean Normalized Expression or MNE; Simon 2003) to compare gene expression among Calliphoridae species with different feeding habits in four biological replicates. PCR efficiencies were obtained from the standard curve of each primer, ranging from 91 to 104% (Table S2).

252 To identify genes potentially correlated with distinct feeding habits or with species
253 diversification patterns, subsequent statistical inferences were conducted using a two-way
254 ANOVA. Statistical inferences and expression values are shown in Tables S5-S7.
255 Possible candidate genes could be classified into three groups according to the divergence
256 in expression patterns among the different species, by sex and developmental stage: (i) genes
257 with conserved expression levels among all species; (ii) genes with different expression levels in
258 at least one comparison, and (iii) genes with different expression levels in which the strongest
259 factor was feeding preference.
260 The comparisons within larvae demonstrated low variation in gene expression in all
261 genes, except *Mvl*. For them, no statistically significant differences were observed among species
262 that might be correlated with feeding habit *(P*> 0.01; Figure S1). This finding indicates a high
263 degree of conservation in expression of these genes in the larval stage. For the adult stage, no
264 delimited and clear pattern was observed, with two genes (*Cyp6g2* and *Gdh*) showing no
265 statistically significant differences in comparisons between sexes and among species. The
266 remaining genes showed significant differences in at least one comparison, but without any
267 particular correlation with sex or species.
268 Expression levels of *Pgrp-SC2* and *for* varied within species only in adult females *(P*=
269 0.0003 and 0.002; respectively) and were conserved in males. A similar result was observed for
270 *S6k,* but expression levels were conserved in adult males instead (*P*= 0.01). The difference
271 observed in *Jon65aiv* in females was significantly correlated with feeding habit *(P*= 0.009) and
272 in adult males there was a difference among species *(P*= 0.006). The gene *sm* was differentially
273 expressed among the different species in both sexes (adult males *P*= 0.001 and females *P*= 0.01).

Linear correlation analyses were then used to test if evolution in expression levels of the candidate genes were correlated with overall genetic distances among species (neutral sequence evolution represented the null hypothesis). Pair-wise nucleotide sequence distances were calculated and correlated with the pair-wise gene expression divergence. None of the genes showed a significant correlation after the application of multiple testing corrections.

Interestingly, among all our candidate genes, only *Malvolio* showed a significant difference in expression between feeding habits in larvae *(P*=0.00002) and both adult males and females *(P*=0.0007 and *P*= 0.00003, respectively). The *Mvl* gene is expressed at higher levels in species with parasitic habit, while the necro-saprophagous species share a common pattern and level of expression, revealing a habit-specific expression pattern (Figure 1).

These analyses, however, do not account for the evolutionary dependence among the species. Hence, we used PIC analysis to include the phylogeny of the chosen species and correlate feeding habit with gene expression. To this end, we corroborated our previous result that *Mvl* is somehow related to feeding habit (larvae: $r^2 = 0.9929$, *P*= 0.01). Surprisingly, *Gdh* had a similar significant result ($r^2 = 1$, $P < 0.001$ for all comparisons), indicating that it might be related to a feeding habit trait but is not specific to parasitism as shown previously.

Evolution of expression and coding sequence: Expression divergence (all genes) and coding sequence data (five of the candidate genes - see Methods section for details) were used to infer trees. Evolution of each gene was considered to be independent and tested against a standard neutral model using the phylogeny proposed by Marinho et al. (2012).

Dendrograms (trees) constructed with expression level differences for all developmental stages and sexes showed a complex pattern among species. Dendrograms were built for each sex and developmental stage separately. A clear pattern relating expression of the *Mvl* gene with

297 feeding habits was observed, with obligate parasite species separated from the necro-
298 saprophagous species (Figure 2). Indeed, feeding behavior was the major determinant of gene
299 expression divergence, supporting previous statistical analyses.
300 For five of the candidate genes, we also inferred phylogenetic trees using coding
301 sequence data. In all cases, neither the maximum likelihood (ML) or the Bayesian inference (BI)
302 methods showed deviations from the expectations under a neutral evolution model (i.e., grouping
303 congeneric species together). However, results of the Tajima's Relative Rates Test (Table S8)
304 showed a significant deviation from neutrality for at least two genes, *Mvl* (in both comparisons
305 using different outgroups) and *for* (only when *C. megacephala* was specified as the outgroup
306 taxon).
307 Further, we generated a phylogenetic tree adding six more Diptera species (parasites and
308 non-parasites). All obligate parasites except *D. hominis* have a longer branch (the branch length
309 represent the amount of genetic change) when compared to non-parasitic species (Figure 3). The
310 longer branch belongs to *G. morsitans*, the only species with a 69-bp indel in its sequence.
311 Another notable observation is the clustering of *D. hominis* with *Cochliomyia* species.
312 To expand our knowledge of the evolutionary processes involved in *Mvl* coding
313 sequence, we estimated the rate ratio of synonymous to non-synonymous substitutions per site
314 using the sequence and phylogeny of ten Diptera species (Table S9). First, we estimated the
315 average ω for the whole tree and compared its likelihood with the tree generated from
316 expectations under a neutral model (fixed ω= 1). The result was statistically significant, allowing
317 us to reject the null hypothesis in favor of a model in which this gene is evolving under a strong
318 purifying selection (ω= 0.027, P<000.1). To test for differences in the selective processes
319 constraining the tree, we performed the same test considering specific branches (see Methods).

We found an elevated ω rate ratio in the *C. hominivorax* branch compared to all other species (0.111 and 0.023, respectively), indicating relaxation of constraint at this branch.

Together, these results indicate that *Mvl* influences feeding preferences among Calliphoridae species, largely due to differences in expression levels and also with some influence of coding sequence variation. The molecular mechanisms of these sequence variations are probably located in specific sites, thus accounting for the discordance between the coding sequence tree and expression data.

DISCUSSION

A candidate gene approach was used to study the evolution of feeding habit in Calliphoridae. Messenger RNA (mRNA) levels of eight candidate genes were measured in larvae, adult males, and adult females of three necro-saprophagous species and one obligate parasite using quantitative PCR (qPCR).

A high degree of conservation in gene expression (transcript) levels was found in larvae of the four species. This suggests that regulatory variation in these genes is constrained by purifying selection and that these genes have an important function in development. The pervasive role of purifying selection has been described previously in genome-wide reports of various organisms, including mammals and flies (Brawand et al., 2011; Nuzhdin et al., 2004).

Some genes appear to be under different evolutionary selective pressures during different developmental stages. One of the most promising candidate genes, *For*, encodes a cGMP-dependent protein kinase (PKG), which participates in the modulation of the division of labor in social insect species and foraging behavior in *Drosophila melanogaster* (Ben-Shahar, 2005; Fitzpatrick and Sokolowski, 2004). This gene was differentially expressed among species in

adult females, but neither in males nor larvae. This difference between sexes and stages requires the presence of different regulatory modules and the action of different selection regimes in each module. There was a difference in gene expression among species in one sex but not in the other for *S6K*, *Jon65aiv,* and *Pgrp-sc2* as well. Different phenotypes mediated by the presence of different regulatory factors has already been observed in stickleback fish (Chan et al., 2010). Chan and colleagues documented that the deletion of an enhancer in the gene *Pitx1* causes a contrasting variation of the pelvic skeleton between stickleback populations by reducing transcript levels. The result of this mutation is the presence of a population with reduced pelvises; however, the expression of this gene in other tissues essential for organism viability (non-candidate housekeeping genes) is similar between populations. These findings suggest that different regulatory factors are involved in the regulation of specific and essential structures by *Pitx1* during the fish development (Shapiro et al., 2004).

To test the hypothesis that the evolution of expression of these seven genes (in each stage or sex) was predominantly neutral (i.e., that regulatory variation accumulates in the absence of selection), we tested measures of expression and sequence divergence. In a neutral scenario, divergences in expression should accumulate linearly with respect to time. Pair-wise nucleotide sequence distances between species were calculated using four genetic markers. No significant evidence of gene expression evolving under neutrality was found. Together, our observations suggest that candidate gene expression levels are not strictly evolving under a standard neutral model of evolution. These findings are consistent with several studies that have demonstrated that gene expression is, in large part, subject to selection (Nuzhdin et al., 2004; Ometto et al., 2011; Rifkin et al., 2003).

365 In this study, *Malvolio* was found to deviate from neutral expectations with respect to
366 gene expression levels between habits in all comparisons. *Mvl* is involved in the gustatory signal
367 transduction via metal-ion transport into neurons regulating the sensory perception of sweet taste
368 (Ben-Shahar et al., 2004; Orgad et al., 1998). In *D. melanogaster*, behavioral studies have
369 revealed the association of this gene with food preference (Orgad et al., 1998). In contrast, in the
370 honeybee *Apis mellifera*, *Mvl* can define specialized tasks in the hive. Forager bees have higher
371 *Mvl* mRNA transcript levels and increased responsiveness to sugar, while nurse bees exhibit
372 lower expression levels and decreased responsiveness to sugar (Ben-Shahar et al., 2004). Indeed,
373 *Mvl* plays an important role in foraging and food preference. According to our expression data,
374 the obligate parasite, *C. hominivorax*, has higher *Mvl* mRNA transcript levels compared to the
375 necro-saprophagous species. In addition, taking into consideration the phylogenetic dependence
376 among species, this study demonstrates a remarkable correlation between feeding habit and *Mvl*
377 expression level.
378 Analyses under a phylogenetic framework, constructing dendrograms based on
379 expression level differences and phylogenies based on the nucleotide sequence of coding
380 regions, provided additional evidence for a possible role of *Mvl* in feeding habit differences
381 among species. In the dendrogram generated for this gene, almost all parasites had a longer
382 branch than the necro-saprophagous species. Additionally, the parasite *D. hominis* clustered with
383 the genus *Cochliomyia*. However, it was previously observed that the affinities of Oestridae
384 species with the remaining Oestroidea lineages are more difficult to recover with confidence and
385 the group is usually recovered with variable placements in the superfamily's phylogeny (Marinho
386 et al., 2012).

387 The correlation of *Mvl* with feeding behavior may be attributable to differences in
388 transcript levels between necro-saprophagous and obligate parasitic species, although there may
389 be some influence by signatures of sites under selection or similar molecular evolutionary
390 processes (as suggested by the Tajima's test statistics). We confirmed with ω analysis that *Mvl* is
391 constrained by purifying selection, but it is important to note that this is just a part of *Mvl* coding
392 sequence. We might find molecular evolutionary signatures in other portions of the coding
393 region or even in regulatory regions of the gene as well. The *for* gene might also be under
394 influence of these differences, but these are probably due to specific sites under selection in the
395 coding region. No evident pattern or significant difference in expression level was observed for
396 the *for* gene.
397
398 CONCLUSION
399 In this study, we tested the hypothesis that variation in gene expression is directly involved in
400 feeding behavior. To date, most previous studies in this subject have been based on phylogenetic
401 information alone (Stevens and Wall, 1997; Stevens, 2003). The results of the present study
402 represent the first and expanded effort to understand the evolution of parasitism in Calliphoridae
403 combining functional and molecular evolutionary approaches. Our findings indicate that
404 *Malvolio* may be involved in the evolutionary genetic regulation of Calliphoridae feeding
405 behavior, and additional studies into the molecular mechanisms undergirding the adaptive
406 evolution of this candidate gene are needed and underway.
407
408
409 ADDITIONAL INFORMATION

Accession codes: Public *C. hominivorax* sequences (SRA accession number SRR060179) were used to design the primers for the qPCR experiments (details provided in Table S1). Sequences generated by Sanger sequencing for *C. abiceps*, *C. megacephala*, *C. hominivorax* and *C. macellaria* were deposited in GenBank under the accession numbers KU234347 to KU234388 (Tables S4). *Mvl* sequences of *D. hominis* and *O. ovis* were obtained from non-public RNA-seq data and were deposited in GenBank under the accession numbers, KU234379 and KU234387, respectively. *Mvl* orthologs in *L. cuprina* and *L. sericata* were recovered from public datasets (accession numbers JRES01000000 and SRR350015, respectively).

ACKNOWLEDGEMENTS

The authors thank Rosangela A. Rodrigues for technical assistance and Maria Salete Couto for maintaining fly stocks. In addition, the authors thank Professor Cláudio José Von Zubem, PhD (UNESP, Rio Claro campus) for providing *C. megacephala* stocks and Professor Ana Cláudia Lessinger, PhD (UFSCar, Sorocaba campus) for providing *C. macellaria* stocks. The authors are grateful for statistical support provided by Bárbara Bitarello for PAML analyses.

REFERENCES

Abzhanov, A., Kuo, W.P., Hartmann, C., Grant, B.R., Grant, P.R., Tabin, C.J., 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. Nature 442, 563-567.

Abzhanov, A., Protas, M., Grant, B.R., Grant, P.R., Tabin, C.J., 2004. Bmp4 and morphological variation of beaks in Darwin's finches. Science 305, 1462-1465.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology 215, 403-410.

435 Ben-Shahar, Y., 2005. The foraging gene, behavioral plasticity, and honeybee division of labor.
436 Journal of comparative physiology. A, Neuroethology, sensory, neural, and behavioral
437 physiology 191, 987-994.
438 Ben-Shahar, Y., Dudek, N.L., Robinson, G.E., 2004. Phenotypic deconstruction reveals
439 involvement of manganese transporter malvolio in honey bee division of labor. The Journal of
440 experimental biology 207, 3281-3288.
441 Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M.,
442 Liechti, A., Aximu-Petri, A., Kircher, M., Albert, F.W., Zeller, U., Khaitovich, P., Grutzner, F.,
443 Bergmann, S., Nielsen, R., Paabo, S., Kaessmann, H., 2011. The evolution of gene expression
444 levels in mammalian organs. Nature 478, 343-348.
445 Cardoso, G.A., Matiolli, C.C., de Azeredo-Espin, A.M., Torres, T.T., 2014. Selection and
446 validation of reference genes for functional studies in the Calliphoridae family. J Insect Sci 14, 2.
447 Carvalho, R.A., Azeredo-Espin, A.M., Torres, T.T., 2010. Deep sequencing of New World
448 screw-worm transcripts to discover genes involved in insecticide resistance. BMC genomics 11,
449 695.
450 Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Jr., Shapiro, M.D., Brady, S.D., Southwick,
451 A.M., Absher, D.M., Grimwood, J., Schmutz, J., Myers, R.M., Petrov, D., Jonsson, B., Schluter,
452 D., Bell, M.A., Kingsley, D.M., 2010. Adaptive evolution of pelvic reduction in sticklebacks by
453 recurrent deletion of a Pitx1 enhancer. Science 327, 302-305.
454 Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman,
455 T.C., Kellis, M., Gelbart, W., Iyer, V.N., Pollard, D.A., Sackton, T.B., Larracuente, A.M., Singh,
456 N.D., Abad, J.P., Abt, D.N., Adryan, B., Aguade, M., Akashi, H., Anderson, W.W., Aquadro,
457 C.F., Ardell, D.H., Arguello, R., Artieri, C.G., Barbash, D.A., Barker, D., Barsanti, P.,
458 Batterham, P., Batzoglou, S., Begun, D., Bhutkar, A., Blanco, E., Bosak, S.A., Bradley, R.K.,
459 Brand, A.D., Brent, M.R., Brooks, A.N., Brown, R.H., Butlin, R.K., Caggese, C., Calvi, B.R.,
460 Bernardo de Carvalho, A., Caspi, A., Castrezana, S., Celniker, S.E., Chang, J.L., Chapple, C.,
461 Chatterji, S., Chinwalla, A., Civetta, A., Clifton, S.W., Comeron, J.M., Costello, J.C., Coyne,
462 J.A., Daub, J., David, R.G., Delcher, A.L., Delehaunty, K., Do, C.B., Ebling, H., Edwards, K.,
463 Eickbush, T., Evans, J.D., Filipski, A., Findeiss, S., Freyhult, E., Fulton, L., Fulton, R., Garcia,
464 A.C., Gardiner, A., Garfield, D.A., Garvin, B.E., Gibson, G., Gilbert, D., Gnerre, S., Godfrey, J.,
465 Good, R., Gotea, V., Gravely, B., Greenberg, A.J., Griffiths-Jones, S., Gross, S., Guigo, R.,
466 Gustafson, E.A., Haerty, W., Hahn, M.W., Halligan, D.L., Halpern, A.L., Halter, G.M., Han,
467 M.V., Heger, A., Hillier, L., Hinrichs, A.S., Holmes, I., Hoskins, R.A., Hubisz, M.J., Hultmark,
468 D., Huntley, M.A., Jaffe, D.B., Jagadeeshan, S., Jeck, W.R., Johnson, J., Jones, C.D., Jordan,
469 W.C., Karpen, G.H., Kataoka, E., Keightley, P.D., Kheradpour, P., Kirkness, E.F., Koerich, L.B.,
470 Kristiansen, K., Kudrna, D., Kulathinal, R.J., Kumar, S., Kwok, R., Lander, E., Langley, C.H.,
471 Lapoint, R., Lazzaro, B.P., Lee, S.J., Levesque, L., Li, R., Lin, C.F., Lin, M.F., Lindblad-Toh,
472 K., Llopart, A., Long, M., Low, L., Lozovsky, E., Lu, J., Luo, M., Machado, C.A., Makalowski,
473 W., Marzo, M., Matsuda, M., Matzkin, L., McAllister, B., McBride, C.S., McKernan, B.,
474 McKernan, K., Mendez-Lago, M., Minx, P., Mollenhauer, M.U., Montooth, K., Mount, S.M.,
475 Mu, X., Myers, E., Negre, B., Newfeld, S., Nielsen, R., Noor, M.A., O'Grady, P., Pachter, L.,
476 Papaceit, M., Parisi, M.J., Parisi, M., Parts, L., Pedersen, J.S., Pesole, G., Phillippy, A.M.,
477 Ponting, C.P., Pop, M., Porcelli, D., Powell, J.R., Prohaska, S., Pruitt, K., Puig, M., Quesneville,
478 H., Ram, K.R., Rand, D., Rasmussen, M.D., Reed, L.K., Reenan, R., Reily, A., Remington,
479 K.A., Rieger, T.T., Ritchie, M.G., Robin, C., Rogers, Y.H., Rohde, C., Rozas, J., Rubenfield,
480 M.J., Ruiz, A., Russo, S., Salzberg, S.L., Sanchez-Gracia, A., Saranga, D.J., Sato, H., Schaeffer,

S.W., Schatz, M.C., Schlenke, T., Schwartz, R., Segarra, C., Singh, R.S., Sirot, L., Sirota, M., Sisneros, N.B., Smith, C.D., Smith, T.F., Spieth, J., Stage, D.E., Stark, A., Stephan, W., Strausberg, R.L., Strempel, S., Sturgill, D., Sutton, G., Sutton, G.G., Tao, W., Teichmann, S., Tobari, Y.N., Tomimura, Y., Tsolas, J.M., Valente, V.L., Venter, E., Venter, J.C., Vicario, S., Vieira, F.G., Vilella, A.J., Villasante, A., Walenz, B., Wang, J., Wasserman, M., Watts, T., Wilson, D., Wilson, R.K., Wing, R.A., Wolfner, M.F., Wong, A., Wong, G.K., Wu, C.I., Wu, G., Yamamoto, D., Yang, H.P., Yang, S.P., Yorke, J.A., Yoshida, K., Zdobnov, E., Zhang, P., Zhang, Y., Zimin, A.V., Baldwin, J., Abdouelleil, A., Abdulkadir, J., Abebe, A., Abera, B., Abreu, J., Acer, S.C., Aftuck, L., Alexander, A., An, P., Anderson, E., Anderson, S., Arachi, H., Azer, M., Bachantsang, P., Barry, A., Bayul, T., Berlin, A., Bessette, D., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Bourzgui, I., Brown, A., Cahill, P., Channer, S., Cheshatsang, Y., Chuda, L., Citroen, M., Collymore, A., Cooke, P., Costello, M., D'Aco, K., Daza, R., De Haan, G., DeGray, S., DeMaso, C., Dhargay, N., Dooley, K., Dooley, E., Doricent, M., Dorje, P., Dorjee, K., Dupes, A., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Fisher, S., Foley, C.D., Franke, A., Friedrich, D., Gadbois, L., Gearin, G., Gearin, C.R., Giannoukos, G., Goode, T., Graham, J., Grandbois, E., Grewal, S., Gyaltsen, K., Hafez, N., Hagos, B., Hall, J., Henson, C., Hollinger, A., Honan, T., Huard, M.D., Hughes, L., Hurhula, B., Husby, M.E., Kamat, A., Kanga, B., Kashin, S., Khazanovich, D., Kisner, P., Lance, K., Lara, M., Lee, W., Lennon, N., Letendre, F., LeVine, R., Lipovsky, A., Liu, X., Liu, J., Liu, S., Lokyitsang, T., Lokyitsang, Y., Lubonja, R., Lui, A., MacDonald, P., Magnisalis, V., Maru, K., Matthews, C., McCusker, W., McDonough, S., Mehta, T., Meldrim, J., Meneus, L., Mihai, O., Mihalev, A., Mihova, T., Mittelman, R., Mlenga, V., Montmayeur, A., Mulrain, L., Navidi, A., Naylor, J., Negash, T., Nguyen, T., Nguyen, N., Nicol, R., Norbu, C., Norbu, N., Novod, N., O'Neill, B., Osman, S., Markiewicz, E., Oyono, O.L., Patti, C., Phunkhang, P., Pierre, F., Priest, M., Raghuraman, S., Rege, F., Reyes, R., Rise, C., Rogov, P., Ross, K., Ryan, E., Settipalli, S., Shea, T., Sherpa, N., Shi, L., Shih, D., Sparrow, T., Spaulding, J., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Strader, C., Tesfaye, S., Thomson, T., Thoulutsang, Y., Thoulutsang, D., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Vo, A., Wangchuk, T., Wangdi, T., Weiand, M., Wilkinson, J., Wilson, A., Yadav, S., Young, G., Yu, Q., Zembek, L., Zhong, D., Zimmer, A., Zwirko, Z., Alvarez, P., Brockman, W., Butler, J., Chin, C., Grabherr, M., Kleber, M., Mauceli, E., MacCallum, I., 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450, 203-218.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9, 772.

Desper, R., Gascuel, O., 2002. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. J Comput Biol 9, 687-705.

DJ, Z., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. University of Texas at Austin.

Enard, W., Khaitovich, P., Klose, J., Zollner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., Doxiadis, G.M., Bontrop, R.E., Paabo, S., 2002. Intra- and interspecific variation in primate gene expression patterns. Science 296, 340-343.

F, Z., 1965. Myiasis in man and animals in the Old World, London: Butterworths.

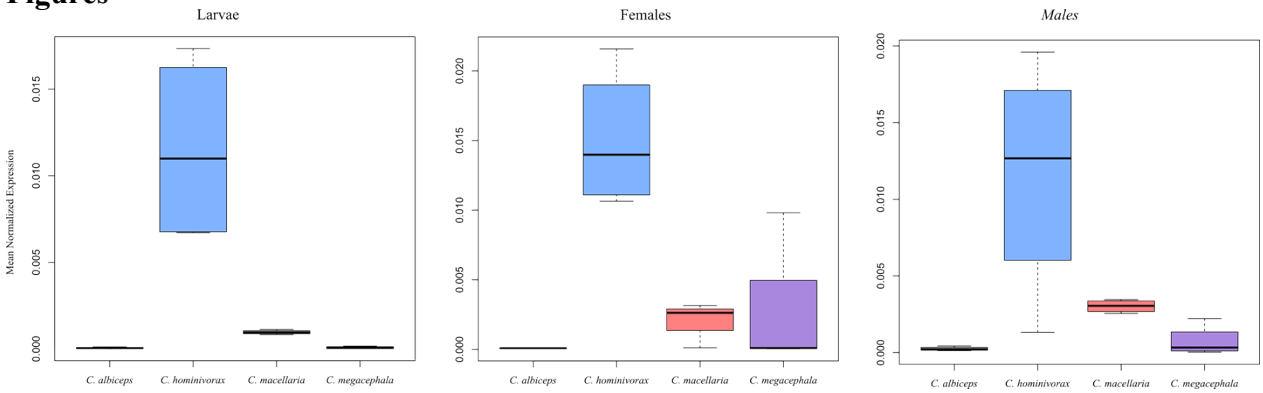
Fitzpatrick, M.J., Sokolowski, M.B., 2004. In Search of Food: Exploring the Evolutionary Link Between cGMP-Dependent Protein Kinase (PKG) and Behaviour. Integrative and comparative biology 44, 28-36.

Greenspan, R.J., Ferveur, J.F., 2000. Courtship in Drosophila. Annu Rev Genet 34, 205-232.

527 Hall, M., Wall, R., 1995. Myiasis of humans and domestic animals. Adv Parasitol 35, 257-334.
528 Higuchi, R., Fockler, C., Dollinger, G., Watson, R., 1993. Kinetic PCR analysis: real-time
529 monitoring of DNA amplification reactions. Biotechnology (N Y) 11, 1026-1030.
530 Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple
531 sequence alignment based on fast Fourier transform. Nucleic Acids Res 30, 3059-3066.
532 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7:
533 improvements in performance and usability. Mol Biol Evol 30, 772-780.
534 King, M.C., Wilson, A.C., 1975. Evolution at two levels in humans and chimpanzees. Science
535 188, 107-116.
536 Marinho, M.A., Junqueira, A.C., Paulo, D.F., Esposito, M.C., Villet, M.H., Azeredo-Espin,
537 A.M., 2012. Molecular phylogenetics of Oestroidea (Diptera: Calyptratae) with emphasis on
538 Calliphoridae: insights into the inter-familial relationships and additional evidence for paraphyly
539 among blowflies. Mol Phylogenet Evol 65, 840-854.
540 Nuzhdin, S.V., Wayne, M.L., Harmon, K.L., McIntyre, L.M., 2004. Common pattern of
541 evolution of gene expression level and protein sequence in Drosophila. Mol Biol Evol 21, 1308-
542 1317.
543 Ometto, L., Shoemaker, D., Ross, K.G., Keller, L., 2011. Evolution of gene expression in fire
544 ants: the effects of developmental stage, caste, and species. Mol Biol Evol 28, 1381-1392.
545 Orgad, S., Nelson, H., Segal, D., Nelson, N., 1998. Metal ions suppress the abnormal taste
546 behavior of the Drosophila mutant malvolio. The Journal of experimental biology 201, 115-120.
547 Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R
548 language. Bioinformatics 20, 289-290.
549 Rifkin, S.A., Kim, J., White, K.P., 2003. Evolution of gene expression in the Drosophila
550 melanogaster subgroup. Nature genetics 33, 138-144.
551 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B.,
552 Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic
553 inference and model choice across a large model space. Syst Biol 61, 539-542.
554 Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist
555 programmers. Methods in molecular biology 132, 365-386.
556 Ryner, L.C., Goodwin, S.F., Castrillon, D.H., Anand, A., Villella, A., Baker, B.S., Hall, J.C.,
557 Taylor, B.J., Wasserman, S.A., 1996. Control of male sexual behavior and sexual orientation in
558 Drosophila by the fruitless gene. Cell 87, 1079-1089.
559 Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jonsson, B., Schluter,
560 D., Kingsley, D.M., 2004. Genetic and developmental basis of evolutionary pelvic reduction in
561 threespine sticklebacks. Nature 428, 717-723.
562 Simon, P., 2003. Q-Gene: processing quantitative real-time RT-PCR data. Bioinformatics 19,
563 1439-1440.
564 Sokolowski, M.B., 2001. Drosophila: genetics meets behaviour. Nature reviews. Genetics 2,
565 879-890.
566 Stevens, J., Wall, R., 1997. The evolution of ectoparasitism in the genus Lucilia
567 (Diptera:Calliphoridae). Int J Parasitol 27, 51-59.
568 Stevens, J.R., 2003. The evolution of myiasis in blowflies (Calliphoridae). Int J Parasitol 33,
569 1105-1113.
570 Stevens, J.R., Wallman, J.F., 2006. The evolution of myiasis in humans and other animals in the
571 Old and New Worlds (part I): phylogenetic analyses. Trends in parasitology 22, 129-136.

572 Tajima, F., 1993. Simple methods for testing the molecular evolutionary clock hypothesis.
573 Genetics 135, 599-607.
574 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5:
575 molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and
576 maximum parsimony methods. Mol Biol Evol 28, 2731-2739.
577 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular
578 Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30, 2725-2729.
579 Team, R.C., 2008. R: A language and environment for statistical computing, in: R Foundation
580 for Statistical Computing, V., Austria (Ed.).
581 Waddell, S., Quinn, W.G., 2001. Flies, genes, and learning. Annu Rev Neurosci 24, 1283-1309.
582 Yang, X., Schadt, E.E., Wang, S., Wang, H., Arnold, A.P., Ingram-Drake, L., Drake, T.A., Lusis,
583 A.J., 2006. Tissue-specific expression and regulation of sexually dimorphic genes in mice.
584 Genome research 16, 995-1004.
585 Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24,
586 1586-1591.
587 YZ, E., 1989. The origin of parasitism in blowflies. British Journal of Entomology and Natural
588 History 2, 125-127.
589
590

591 **Figures**



592
593

594 **Figure 1. mRNA levels of *Malvolio* in adult females, adult males, and larvae**. The three

595 boxplots show that the obligate parasite, *C. hominivorax*, has a specific pattern of expression

596 while necro-saprophagous species have a similar pattern among them. The highest expression

597 difference observed was between *C. albiceps* and *C. hominivorax* larvae. This gene was

598 approximately 210 times more expressed in the obligate parasite. The smallest difference was

599 among necro-saprophagous species. *C. megacephala* had 1.2 times higher expression than *C.*

600 *macellaria*.

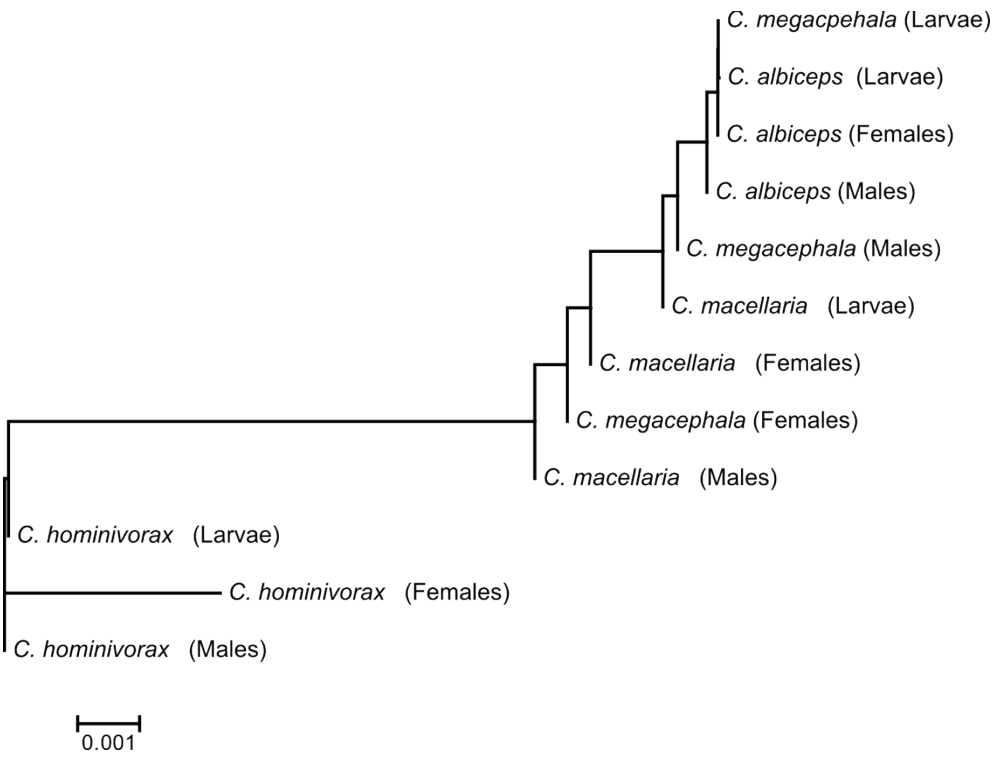
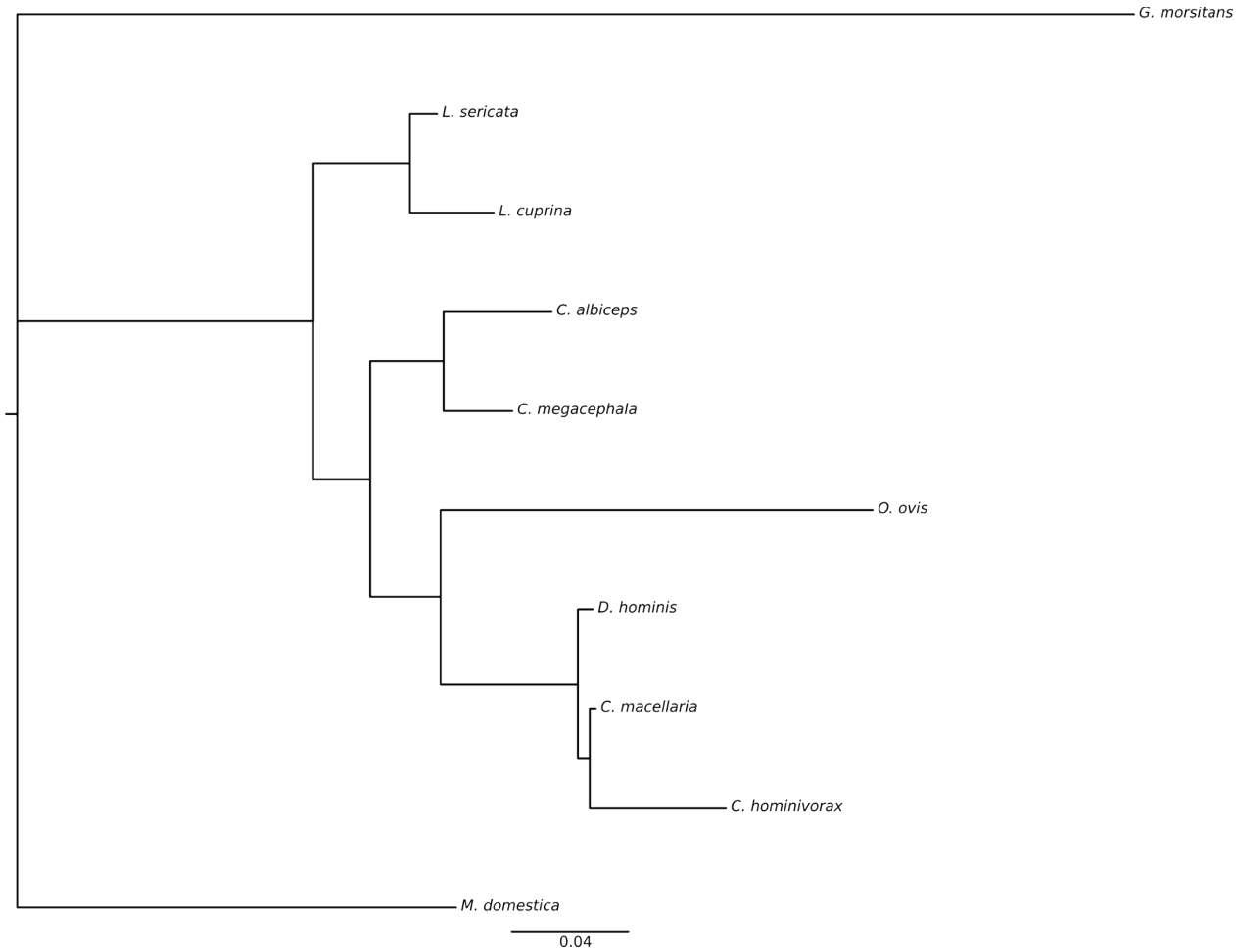


Figure 2. Dendrogram of *Mvl* expression using qPCR data. The diagram shows the separation of *C. hominivorax* from the necro-saprophagous species. This finding corroborates the idea that divergence in expression of *Mvl* is associate with feeding behavior. The scale bar represents the branch length estimated by expression divergence.



606

607 **Figure 3. Dendrogram of the *Mvl* gene including a broader range of species**. Species include
608 obligate parasites *C. hominivorax*, *D. hominis*, *L. cuprina*, *G. morsitans,*and *O. ovis*, and non-
609 parasites *C. albiceps*, *C. megacephala*, *C. macellaria*, *L. sericata,* and *M. domestica*. The
610 majority of parasitic blowflies have longer branches compared to non-parasite species. The scale
611 bar represents 0.04 substitutions per site.