Molecular Evolution of the Ligands of the Insulin-Signaling Pathway: *dilp* Genes in the Genus *Drosophila*

Sara Guirao-Rico and Montserrat Aguadé*

1Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain
2Institut de Recerca de la Biodiversitat, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain

*Corresponding author: E-mail: maguade@ub.edu.

Associate editor: John H McDonald

Abstract

*Drosophila melanogaster*, unlike mammals, has seven insulin-like peptides (DILPs). In *Drosophila*, all seven genes (*dilp*1–7) are single copy in the 12 species studied, except for *D. grimshawi* with two tandem copies of *dilp2*. Our comparative analysis revealed that genes *dilp1–dilp7* exhibit differential functional constraint, which is indicative of some functional divergence. Species of the subgenera Sophophora and *Drosophila* differ in some traits likely affected by the insulin-signaling pathway, such as adult body size. It is in the branch connecting the two subgenera that we found the footprint left by positive selection driving nonsynonymous changes at some *dilp1* codons to fixation. Finally, the similar rate at which the two *dilp2* copies of *D. grimshawi* have evolved since their duplication and the presence of a putative regulatory region highly conserved between the two paralogs would suggest that both copies were preserved either because of subfunctionalization or dose dependency rather than by the neofunctionalization of one of the two copies.

Key words: natural selection, gene duplication, insulin, *Drosophila*.

In *Drosophila melanogaster*, the insulin-signaling pathway controls body size and some life history traits, such as fertility and lifespan (as reviewed in Zera et al. 2007). Because these characters generally reflect adaptive responses to environmental pressures, both positive and negative selection might have played an important role in the molecular evolution of the underlying genes. Here, we have focused on the seven genes (*dilp*1–7; fig. 1) encoding the *Drosophila* insulin-like peptides (DILPs), which are the ligands that trigger the insulin-signaling cascade. Despite the high divergence between members of this small multigene family, the encoded proteins are structurally similar to the single mammalian insulin peptide (fig. 1). The present comparative analysis of the *dilp* genes across the *Drosophila* phylogeny aimed firstly to determine the role of purifying selection in shaping the *dilp* genes evolution. Secondly, we evaluated the putative role of positive selection acting on *dilp* genes soon after the separation of the Sophophora and *Drosophila* subgenera, given previous observations indicating that adult body size differs between subgenera (Sturtevant 1939; Pitnick et al. 1995; Guirao-Rico and Aguadé 2009) and the possible effect of the insulin-signaling pathway on adult body size. Finally, we evaluated the mode of evolution of the tandem *dilp2* duplication detected in the *D. grimshawi* lineage.

Orthologs of the seven *dilp* genes present in *D. melanogaster* and *D. simulans* were identified in the assembled genomes of the other ten species with the exception of *dilp2* in *D. persimilis*, where it would lie in an unsequenced part of the corresponding scaffold (see supplementary fig. S1, Supplementary Material online). The presence of a single annotated *dilp* copy of each gene was confirmed in all other cases with the exception of *D. grimshawi*, where two tandem copies of *dilp2* were identified (named *dilp2-p* and *dilp2-d*).

The selective pressures acting on *dilp* genes—measured as the nonsynonymous to synonymous divergence ratio ($\omega = dN/dS$)—were analyzed by maximum likelihood (ML) using the curated multiple nucleotide alignments (either for each gene or concatenated for all genes [see supplementary material, Supplementary Material online]) and the best-supported phylogenetic tree topology for all species studied (i.e., considering *D. erecta* and *D. yakuba* as sister species; Pollard et al. 2006; Clark et al. 2007). The comparison of nested evolutionary models $M_{C0}$ and $M_{C3}$ using the concatenated sequences yielded highly significant results (table 1) both for the 12 *Drosophila* species and for a subset including the five more closely related Sophophora species (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, and *D. erecta*). The estimated $\omega$ values were less than 1 for all *dilp* genes (table 1), indicating that they have all been under purifying selection. Moreover, the detected heterogeneity indicates that the strength of selection differed among genes (table 1), with gene *dilp7* the most constrained. When each *dilp* gene was considered separately (table 2), the comparison of the nested branch models M0 and FR yielded in all cases a nonsignificant result, which would support a single $\omega$ value for each gene and, therefore, that the strength of purifying selection acting on each gene had not varied across lineages. The comparison of the nested site models M1 and M2 provided no evidence for positive selection acting at particular codons of any of the seven *dilp* genes.

Given previous observations indicating that adult body size differs between species of the Sophophora and *Drosophila* subgenera (Sturtevant 1939; Pitnick et al. 2009; Pitnick et al. 1995; Guirao-Rico and Aguadé 2009), unlike mammals, has seven insulin-like peptides (DILPS). In *Drosophila*, all seven genes (*dilp*1–7) are single copy in the 12 species studied, except for *D. grimshawi* with two tandem copies of *dilp2*. Our comparative analysis revealed that genes *dilp1–dilp7* exhibit differential functional constraint, which is indicative of some functional divergence. Species of the subgenera Sophophora and *Drosophila* differ in some traits likely affected by the insulin-signaling pathway, such as adult body size. It is in the branch connecting the two subgenera that we found the footprint left by positive selection driving nonsynonymous changes at some *dilp1* codons to fixation. Finally, the similar rate at which the two *dilp2* copies of *D. grimshawi* have evolved since their duplication and the presence of a putative regulatory region highly conserved between the two paralogs would suggest that both copies were preserved either because of subfunctionalization or dose dependency rather than by the neofunctionalization of one of the two copies.

Key words: natural selection, gene duplication, insulin, *Drosophila*.
Synonymous sequence divergence between the two copies of *dilp2* detected in *D. grimshawi* allowed estimating the age of the duplication event at ~10.5 Ma (95% confidence interval, 0.6–22.9; see supplementary material, Supplementary Material online). Divergence between the two copies was lower at nonsynonymous (*K_a* = 0.068) than at synonymous sites (*K_s* = 0.206), an indication of the differential strength of purifying selection on the two site classes. Moreover, their ratio (*ω* = 0.329) was higher than between-orthologs estimates (table 1), suggesting that the level of functional constraint had changed after the duplication. In order to ascertain whether constraint had changed similarly in both copies, an ML analysis using branch models was performed considering the two *dilp2* copies of *D. grimshawi* and the single *dilp2* gene present in the other species studied. Because no heterogeneity among lineages was detected at the *dilp2* gene (table 2), two possibilities were considered: 1) a single *ω* value for the two *D. grimshawi* copies and another for their orthologs (M-2ratio) and 2) a different *ω* value for each of the two *D. grimshawi* copies and for their orthologs (M-3ratio). The ML analysis favored the M-2ratio model (*P* value = 0.0004; supplementary table S1, Supplementary Material online), with an estimated *ω* = 0.374 value for the two *dilp2* genes of *D. grimshawi* and a much lower estimate for their orthologs (*ω* = 0.066). Accordingly, the two *dilp2* copies of *D. grimshawi* would have been subjected to a similar level of functional constraint since duplication, a level that was however lower than for the single *dilp2* gene present in the others species. Comparison of the two DILP2 proteins revealed 13 amino acid differences (eight located in the C peptide and five in the A chain). The nonsynonymous substitutions underlying 8 of the 13 amino acid replacements could be polarized after reconstruction of the ancestral coding region of the two *dilp2* paralogs (Yang 1997). The number of nonsynonymous substitutions did not differ significantly between the *dilp2_p* and the *dilp2_d* branches: seven and five, respectively.

The extent of the duplicated region was established through dot-plot analysis (fig. 2; see supplementary material, Supplementary Material online). This region (~1,200 bp long) includes the entire *dilp2* gene and part of its 5′ and 3′ flanking regions. Comparison of the 5′ flanking region (~500 bp) between copies revealed that divergence at this region (*K = 0.008*) was significantly lower than at synonymous sites of the coding region (*K_s = 0.206*). In *D. melanogaster*, the 5′ flanking region encompasses a fragment that controls the expression of this gene in seven me-

---

**Table 1.** Estimates of the *ω* Value from the Combined ML Analysis of the Seven *dilp* Genes in Drosophila.

<table>
<thead>
<tr>
<th>Alignment</th>
<th><em>dilp1</em></th>
<th><em>dilp2</em></th>
<th><em>dilp3</em></th>
<th><em>dilp4</em></th>
<th><em>dilp5</em></th>
<th><em>dilp6</em></th>
<th><em>dilp7</em></th>
<th><em>P</em> Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete set of species*</td>
<td>0.09</td>
<td>0.08</td>
<td>0.14</td>
<td>0.11</td>
<td>0.24</td>
<td>0.18</td>
<td>0.03</td>
<td>2.86 × 10^{-21}*</td>
</tr>
<tr>
<td>Five species</td>
<td>0.42</td>
<td>0.11</td>
<td>0.08</td>
<td>0.22</td>
<td>0.34</td>
<td>0.28</td>
<td>0.06</td>
<td>4.68 × 10^{-9}</td>
</tr>
</tbody>
</table>

*Note.*—See supplementary material, Supplementary Material online for a description of the models and test performed. *P* values of twice the difference of the log likelihood between the nested models M0 and M3. Statistical significance is indicated in bold.

*It includes those Drosophila species in which all seven *dilp* genes were detected.
dian neurosecretory cells (m-NSCs) in the brain (Ikeya et al. 2002). Location of this fragment in *D. melanogaster* (between sites C0540 and C0146) is partly coincidental with that of the 5’ highly conserved region between the two *D. grimshawi* copies, which raises the possibility that the duplicated upstream region encompasses a regulatory element present in both *D. grimshawi* paralogs.

The similar number of nonsynonymous substitutions in each copy and the differential but similar level of functional constraint detected at both copies would suggest that the duplicates were preserved either because of subfunctionalization or of dose dependency rather than by the neofunctionalization of one of the two copies. If the highly conserved upstream region reflected the presence of a common regulatory element, this regulatory strategy would be compatible with both the subfunctionalization and the dose-dependent hypotheses.

Adult body size is several times larger in *D. grimshawi* and in other Hawaiian drosophilids than in *D. melanogaster* and also larger than in the virilis and repleta species groups (Pitnick et al. 1995). It is worth noting that among *dilp* genes, *dilp2* is the most highly expressed as well as the most potent growth stimulator in *D. melanogaster*. Moreover, its overexpression during development results in larger flies of this species (39% increase in body weight; Brogiolo et al. 2001). It is tempting to speculate that if its duplication in the *D. grimshawi* lineage implied an increase in the *dilp2* gene dose, the duplication might have contributed to the larger body size of this and other Hawaiian species relative to the *D. virilis* and *D. mojavensis* lineages. This scenario needs, however, to be further evaluated through functional analyses.

**Supplementary Material**

Supplementary material, table S1, and figures S1–S3 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

We thank J. Rozas and A. Sánchez-Gracia for comments on the manuscript. We also thank Serveis Científico- Tècnics, Universitat de Barcelona, for automated sequencing facilities. S.G.-R. was supported by a predoctoral fellowship from Ministerio de Educación y Ciencia (Spain). This work was supported by grants BU2004-02253 and BU2007-63229 from Comisión Interdepartamental de Ciencia y Tecnología, Spain, and grants 2005SGR-00166 and 2009SGR-1287 from Comissió Interdepartamental de Recerca i Innovació Tecnològica, Generalitat de Catalunya, Spain, to M.A.

**References**


### Table 2. Analysis of the Individual *dilp* Genes across the Drosophila Phylogeny.

<table>
<thead>
<tr>
<th>Comparison</th>
<th><em>dilp1</em></th>
<th><em>dilp2</em></th>
<th><em>dilp3</em></th>
<th><em>dilp4</em></th>
<th><em>dilp5</em></th>
<th><em>dilp6</em></th>
<th><em>dilp7</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.33)</td>
<td>(0.82)</td>
<td>(0.79)</td>
<td>(0.78)</td>
<td>(0.53)</td>
<td>(0.34)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>M1 versus M2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(0.55)</td>
<td>(1)</td>
<td>(1)</td>
</tr>
<tr>
<td>M1 versus MA</td>
<td>11.60</td>
<td>8.14</td>
<td>0.08</td>
<td>1.59</td>
<td>0</td>
<td>1.24</td>
<td>5.54</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.02)</td>
<td>(0.96)</td>
<td>(0.45)</td>
<td>(1)</td>
<td>(0.54)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>MAfix versus MA</td>
<td>4.19</td>
<td>0.8</td>
<td>0</td>
<td>0.10</td>
<td>0</td>
<td>0.48</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.37)</td>
<td>(1)</td>
<td>(0.75)</td>
<td>(1)</td>
<td>(0.49)</td>
<td>(1)</td>
</tr>
</tbody>
</table>

**Note.**—For each comparison, values in the first and second rows correspond to twice the difference of the log likelihood between two nested models, respectively, and their *P* values (in parentheses). Statistical significance is indicated in bold. See supplementary material, Supplementary Material online for a description of the models and different tests performed.

**Fig. 2.** Dot-plot analysis of the *dilp2* region of *Drosophila grimshawi*. The *D. grimshawi* sequence between genes *dilp1* and *dilp3* is represented on the x (left to right) and y (bottom up) axes, with genes *dilp2*-p and *dilp2*-d indicated in black boxes.


