

LETTER TO EDITOR

**OPEN ACCESS**

Full open access to this and thousands of other papers at <http://www.la-press.com>.

## Comment on “The Molecular Evolutionary Patterns of the Insulin/FOXO Signaling Pathway”

David Alvarez-Ponce<sup>1</sup>, Montserrat Aguadé<sup>2</sup> and Julio Rozas<sup>2</sup>

<sup>1</sup>Laboratorio de Biología Integrativa de Sistemas, Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia, Valencia, Spain. <sup>2</sup>Departament de Genètica, and Institut de Recerca de la Biodiversitat, Universitat de Barcelona, Barcelona, Spain.

Corresponding author email: [jrozas@ub.edu](mailto:jrozas@ub.edu)

---

**Abstract:** Letter to the Editor on Wang M, Wang Q, Wang Z, Zhang X, Pan Y. The molecular evolutionary patterns of the insulin/FOXO signaling pathway. *Evol Bioinform*. 2013;9:1–16. doi: 10.4137/EBO.S10539.

**Keywords:** network evolution, insulin pathway, selective constraint,  $d_N/d_S$

---

*Evolutionary Bioinformatics* 2013;9 229–234

doi: [10.4137/EBO.S11915](https://doi.org/10.4137/EBO.S11915)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article published under the Creative Commons CC-BY-NC 3.0 license.

Wang et al have recently analyzed the pattern of selective pressures acting on genes along the insulin/FOXO pathway of vertebrates.<sup>1</sup> In this study, which is similar to those that we and other authors performed in the Insulin/TOR (IT) signal transduction pathway,<sup>2–5</sup> the authors questioned some of our previous findings. Here we revise previous analyses of the pattern of selective constraints along the IT and other signal transduction pathways and discuss a number of issues that might account for the apparently different results obtained for the IT pathway by the different groups.

Alvarez-Ponce et al<sup>2</sup> analyzed the selective pressures acting on genes involved in the *Drosophila* IT pathway and found that levels of selective constraint exhibit a gradient distribution along the upstream/downstream pathway axis. In particular, levels of selective constraint (measured from the nonsynonymous to synonymous divergence ratio,  $\omega = d_N/d_S$ ) correlate with pathway position (computed as the number of steps required to transduce the signal from the insulin receptor to the rest of pathway components), being the downstream genes the most selectively constrained. We also determined that this trend is not the result of a higher incidence of positive selection in the upstream part of the pathway<sup>2,4</sup> and that the correlation remains significant after factoring out the effect of gene expression level, codon bias, protein length, and number of protein-protein interactions. The analysis was based on 27 *D. melanogaster* genes (representative of 21 gene families), all of which act downstream of the insulin receptor (ie, neither insulin-like peptides, IGFs, nor the insulin receptor were included in this analysis). We identified the orthologs of each gene in another 11 *Drosophila* species and, when necessary, we corrected manually the gene annotations. Remarkably, the distribution of selective constraint levels observed along the IT pathway sharply contrasts with that generally observed in biosynthetic pathways, in which upstream genes tend to be the most selectively constrained.<sup>6–11</sup>

In order to evaluate whether the correlation observed in *Drosophila* is general or, on the contrary, specific to this genus, we performed a similar analysis using six vertebrate genomes.<sup>3</sup> In this case, we analyzed a total of 72 genes (representative of 23 gene families) (Table 1). Our network-level analysis revealed a tendency consistent with that observed in *Drosophila*: levels of

selective constraint exhibited a polarity along the pathway, with downstream genes being the most constrained. This trend, however, was less pronounced than that observed in *Drosophila*, which might be explained, at least partially, by the lower effective population sizes that vertebrates commonly have.<sup>12,13</sup> In fact, overall  $\omega$  values (estimated using the M0 model implemented in the PAML software,<sup>14</sup> which assumes a unique  $\omega$  value for all codons and branches in the phylogeny) do not significantly correlate with pathway position; however, when the analysis is conducted separately for each branch of the phylogeny of the studied vertebrates (using the  $\omega$  values estimated under the free-ratio model, which assumes a separate  $\omega$  ratio for each branch), the sign of the correlation was negative ( $\rho < 0$ ) for most of the branches. This result indicates that, also in vertebrates, levels of selective constraint follow a polarity along the pathway, thus implying that the trend does not represent a feature specific to *Drosophila*. Jovelín and Phillips also studied the levels and patterns of selective constraint along the IT pathway, in this case in *Caenorhabditis*.<sup>5</sup> In their similar network-level analysis based on 13 genes, they also found that  $\omega$  values significantly correlated with pathway position, with downstream genes being the most selectively constrained. Nevertheless, they observed that the trend becomes nonsignificant after discounting the effect of gene expression levels.

Noticeably, a similar distribution of selective constraint levels has been observed in the yeast HOG signal transduction pathway, whose downstream genes are the most selectively constrained.<sup>15</sup> Indeed, this appears to be the most common distribution of levels of selective constraint across signal transduction cascades: analysis of the global human signal transduction network has shown that, on average, genes acting at the downstream part of the network tend to be more selectively constrained than those occupying more highly hierarchical positions.<sup>16</sup>

In their recent study in vertebrates, Wang et al<sup>1</sup> used a gene set that only partially overlaps with that used in our 2011 analysis. Their gene set encompasses a total of 27 genes (representative of 9 gene families) acting at the upper part of the IT pathway (the insulin/FOXO pathway), of which only 15 (representative of 5 gene families) were also used in our study (Table 1).<sup>3</sup> Consistent with previous observations,<sup>2–5</sup> they also detected a negative significant

**Table 1.** Genes analyzed by Alvarez-Ponce et al,<sup>2,3</sup> Jovelín and Phillips,<sup>5</sup> and Wang et al.<sup>1</sup>

Gene family	Pathway position <sup>a</sup>	Alvarez-Ponce et al. 2009 ( <i>Drosophila</i> ) <sup>2</sup>	Alvarez-Ponce et al. 2011 (vertebrates) <sup>3</sup>	Jovelín and Phillips 2011 ( <i>Caenorhabditis</i> ) <sup>5</sup>	Wang et al. 2013 (vertebrates) <sup>1</sup>
IGFBP	−1	—	—	—	6
INS/IGF1	−1	—	—	—	2
INSR	0	—	3	1	1
IRS	1	1	11	1	3
P85	2	1	3	1	3
P110	3	1	4	1	2
MELT	4	1	1	—	—
PDK1	4	1	2	1	— <sup>c</sup>
AKT	5	1	3	1	3
SGK	5	—	—	1	—
PKC	5	—	9	—	—
TSC1	5	1	1	—	—
FOXO	6	1	4	1	3
GSK3	6	1	2	—	—
TSC2	6	1	1	—	—
EIF2B $\epsilon$	7	1	1	—	—
GYS	7	1	2	—	—
MYC	7	1	4	—	—
RHEB	7	1	2	—	—
TOR	8	1	1	1	—
RPTOR	8	—	—	1	—
4EBP	9	1	3	1	—
S6K	9	1	2	—	—
EIF4E	10	7	5	—	—
RPS6	10	1	2	—	—
FOXO1	10	—	—	1	—
HIF	10	—	—	1	—
CYH	— <sup>b</sup>	1	4	—	—
PTEN	— <sup>b</sup>	1	2	—	—
Total		27	72	13	23

**Notes:** Number of genes per family analyzed. <sup>a</sup>Number of steps required to transduce the signal from the insulin or IGF1 receptors to each protein in the pathway. The same numbering as in ref. 2 is used. In the other works, a consistent numbering is used; <sup>b</sup>gene families not included in network-level analyses, as they do not directly participate in the main signaling cascade; <sup>c</sup>Wang et al<sup>1</sup> included in their analyses genes encoding pyruvate dehydrogenase kinases 1–4 and not the 3-phosphoinositide dependent protein kinase-1.

correlation between  $\omega$  estimates and pathway position, even when potentially confounding factors were controlled for. However, they noticed that statistical significance vanished when genes encoding insulin, IGF1, and IGFBP1–6 (genes not included in any of the previous analyses<sup>2–5</sup>) were removed from their analysis. From this observation, the authors conclude that the distribution of selective constraints that they observe across the pathway “is different from [our] previous similar reports.” They claim that their observation that the detected trend vanishes after removing the above-mentioned genes “demonstrates that the relaxed selective constraint at position 1 explains the purifying selection polarity along the IIF pathway,” adding that “the presented correlation [between pathway position and  $\omega$ ] comes from the relaxed selective

constraint operated on genes of position 1, which give us an erroneous perception of truth.”<sup>1</sup> Wang et al’s conclusions, however, are not supported by their observations, as none of the previous analyses included genes encoding insulin, IGF1, and IGFBP1–6 (genes at position 1 in their study) (Table 1).<sup>2–5</sup> Therefore, the polarity that we had detected cannot be caused by the presence of these genes.

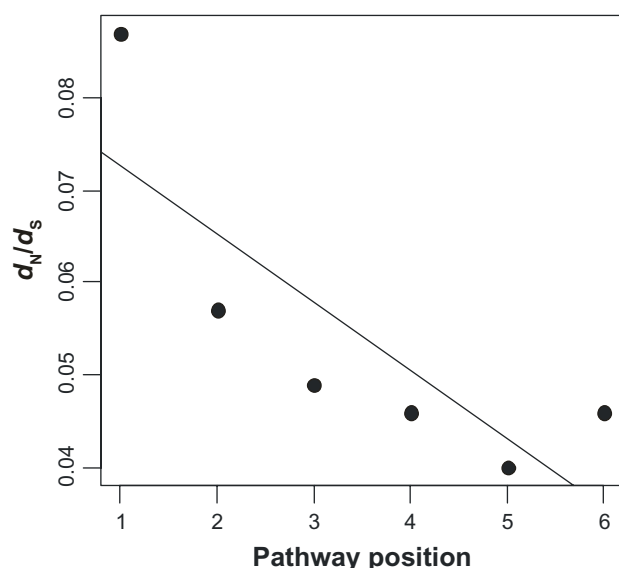
Furthermore, we have reanalyzed our published data sets to evaluate whether the correlation between  $\omega$  and the IT pathway position previously detected vanishes when analyses are restricted to the gene families also analyzed by Wang et al (Table 1). Remarkably, when we reanalyzed our *Drosophila* dataset, we found a significant correlation that is even stronger than that detected in our previous analysis

of the entire pathway<sup>2</sup> (Spearman rank correlation coefficient,  $\rho = -0.898$ ,  $P = 0.015$ ; Kendall rank correlation coefficient,  $\tau = -0.828$ ,  $P = 0.023$ ) (Fig. 1). The reanalysis of our vertebrate dataset shows that the correlation between  $\omega$  and pathway position is negative (negative Spearman and Kendall rank correlation coefficients) in 8 out of the 9 branches of the phylogeny (Table 2). This number of branches with a negative correlation coefficient is higher than the 50% expected under a random distribution of levels of selective constraint along the pathway (binomial test,  $P = 0.039$ ). Therefore, current results in *Drosophila* and vertebrates are consistent with our previously published results and indicate that our previous findings were not the result of using a particular set of genes, since the trend can also be found in the specific part of the pathway studied by Wang et al.<sup>1</sup>

Here we highlight the most important differences between the studies conducted by the different groups, which might account for the observed dissimilarities: quality of gene annotations, genes analyzed, and genes not involved in the main insulin signaling cascade.

## Quality of Gene Annotations

In our previous analyses, we visually inspected all multiple sequence alignments to identify putative



**Figure 1.** Correlation between  $\omega$  estimates and pathway position in the *Drosophila* insulin/TOR pathway. The analysis is restricted to the gene families used both by Alvarez-Ponce et al.<sup>2</sup> and by Wang et al.<sup>1</sup> The  $\omega$  estimates were obtained from Alvarez-Ponce et al.<sup>2</sup> Positions 1–6 correspond, respectively, to genes *chico*, *Pi3K21B*, *Pi3K92E*, *Pk61C*, *Akt1* and *foxo*.

**Table 2.** Correlation between  $\omega$  estimates and pathway position in the vertebrate insulin/TOR pathway.

Phylogenetic branch	$\rho$
a ( <i>Homo sapiens</i> )	0.267
b ( <i>Mus musculus</i> )	-0.272
c ( <i>Bos taurus</i> )	-0.195
d ( <i>Monodelphis domestica</i> )	-0.208
e ( <i>Ornithorhynchus anatinus</i> )	-0.011
f ( <i>Gallus gallus</i> )	-0.362
g (internal branch)	-0.336
h (internal branch)	-0.348
i (internal branch)	-0.497

**Notes:** Analysis restricted to the gene families used both by Alvarez-Ponce et al.<sup>2</sup> and by Wang et al.<sup>1</sup> Estimates of  $\omega$  were obtained from Alvarez-Ponce et al (Dataset 1).<sup>3</sup> Branches are named as in Alvarez-Ponce et al.<sup>3</sup>

gene annotation errors (eg, unannotated exons or exons erroneously annotated in some species). Such annotation errors were, wherever possible, manually fixed using the information of the original genome sequences. After this manual improvement of gene annotations, alignment regions that remained problematic were removed from the analysis. Jovelín and Phillips<sup>5</sup> also manually annotated exon/intron structures.<sup>5</sup> Although Wang et al manually removed “unreliably aligned regions,”<sup>1</sup> they do not mention having corrected annotation errors through manual annotation. Because genome annotations of non-model organisms are mostly the output of automatic computational pipelines, they generally contain a high proportion of errors (See, for example, Tu et al<sup>17</sup> and Collins et al.).<sup>18</sup> Manual correction of such errors is very important, since  $\omega$  estimates are highly sensitive to errors in the gene structure annotation (See, for example, Schneider et al.).<sup>19</sup>

## Genes Analyzed

Wang et al included only 27 genes in their analysis (representative of 9 gene families), all of which act at the upper part of the vertebrate IT pathway, whereas in our vertebrate analysis we used a total of 72 genes (belonging to 23 families), with only 15 (5 families) being shared between both datasets (Table 1).<sup>1,3</sup> Their dataset includes genes encoding insulin, insulin-like growth factor 1 (IGF1), and IGF-binding proteins 1–6 (IGFBP1–6), which were not included in our previous analyses. Conversely, their dataset does not contain genes encoding proteins IGF1R, DOK,

FRS, p110 $\delta$ , PKC, Melted, GSK3, TSC1/2, Rheb, mTOR, S6K, RpS6, 4E-BP, eIF4E, Myc, eIF2B- $\epsilon$ , Glycogen Synthase, PTEN, and Cyh, which were included in our previous analysis (Table 1). The fact that they used a reduced dataset may have decreased the statistical power of their correlation analyses, particularly when genes encoding insulin, IGF1 and IGFBP1–6 (ie, 30% of their dataset) were removed from their analyses. Notice that trimming a small dataset often involves a dramatic reduction in the power of statistical tests, which can result in certain trends becoming statistically nonsignificant, despite their relevance.

### Genes not Involved in the main Insulin Signaling Cascade

According to their report, Wang et al<sup>1</sup> included in their analyses 4 genes encoding 3-phosphoinositide dependent protein kinases (“position 5” in their Fig. 1, downstream of the PI3K complex and upstream of AKT1-3 proteins). However, the accession numbers listed in their additional files 1 and 2 correspond to genes encoding pyruvate dehydrogenase kinases 1–4. Although the activity of the latter enzymes is regulated by insulin, they do not act at “position 5” (in their Fig. 1); indeed, they are not directly involved in the main insulin signal transduction cascade. While insulin activates the 3-phosphoinositide dependent protein kinase-1, it inhibits pyruvate dehydrogenase kinases 1–4.<sup>20–23</sup> The fact that the same acronym (PDK) is often used for both enzymes (even though the gene symbols are *PDPK1* for the 3-phosphoinositide dependent protein kinase-1 and *PDK1–4* for pyruvate dehydrogenase kinases 1–4) likely led Wang et al to confound both gene families.

### Conclusions

In conclusion, the issues raised above might explain the observation by Wang et al<sup>1</sup> that the negative correlation between pathway position and  $\omega$  vanishes after removing some genes from the analysis. These concerns, indeed, may have added some noise to their dataset, reducing the statistical power of their analyses. Therefore, there is no support for claiming that levels of selective constraint do not exhibit a polarity along the upstream/downstream axis of the insulin/TOR pathway.

### Author Contributions

Conceived and designed the experiments: DA-P, MA, JR. Analyzed the data: DA-P. Wrote the first draft of the manuscript: DA-P. Contributed to the writing of the manuscript: DA-P, MA, JR. Agree with manuscript results and conclusions: DA-P, MA, JR. Jointly developed the structure and arguments for the paper: DA-P, MA, JR. Made critical revisions and approved final version: DA-P, MA, JR. All authors reviewed and approved of the final manuscript.

### Funding

Work supported by grants from the Ministerio de Educación y Ciencia of Spain (BFU2010-15484 to JR and BFU2012-35168 to MA) and from the Comissió Interdepartamental de Recerca i Innovació Tecnològica of Spain (2009SGR-1287). JR was partially supported by ICREA Academia (Generalitat de Catalunya), and DA-P by a Juan de la Cierva post-doctoral fellowship from the Spanish Ministerio de Economía y Competitividad (JCI-2011-11089).

### Competing Interests

Author(s) disclose no potential conflicts of interest.

### Disclosures and Ethics

As a requirement of publication the authors have provided signed confirmation of their compliance with ethical and legal obligations including but not limited to compliance with ICMJE authorship and competing interests guidelines, that the article is neither under consideration for publication nor published elsewhere, of their compliance with legal and ethical guidelines concerning human and animal research participants (if applicable), and that permission has been obtained for reproduction of any copyrighted material. This article was subject to blind, independent, expert peer review. The reviewers reported no competing interests.

### References

1. Wang M, Wang Q, Wang Z, Zhang X, Pan Y. The molecular evolutionary patterns of the insulin/FOXO signaling pathway. *Evol Bioinform Online*. 2013;9:1–16.
2. Alvarez-Ponce D, Aguadé M, Rozas J. Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 *Drosophila* genomes. *Genome Res*. 2009;19(2):234–42.
3. Alvarez-Ponce D, Aguadé M, Rozas J. Comparative genomics of the vertebrate insulin/TOR signal transduction pathway: a network-level analysis of selective pressures. *Genome Biol Evol*. 2011;3:87–101.





4. Alvarez-Ponce D, Guirao-Rico S, Orengo DJ, Segarra C, Rozas J, Aguadé M. Molecular population genetics of the insulin/TOR signal transduction pathway: a network-level analysis in *Drosophila melanogaster*. *Mol Biol Evol*. 2012;29(1):123–32.
5. Jovelín R, Phillips PC. Expression level drives the pattern of selective constraints along the insulin/Tor signal transduction pathway in *Caenorhabditis*. *Genome Biol Evol*. 2011;3:715–22.
6. Rausher MD, Miller RE, Tiffin P. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol*. 1999;16(2):266–74.
7. Ramsay H, Rieseberg LH, Ritland K. The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. *Mol Biol Evol*. 2009;26(5):1045–53.
8. Sharkey TD, Yeh S, Wiberley AE, Falbel TG, Gong D, Fernandez DE. Evolution of the isoprene biosynthetic pathway in kudzu. *Plant Physiol*. 2005;137(2):700–12.
9. Livingstone K, Anderson S. Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. *J Hered*. 2009;100(6):754–61.
10. Yu HS, Shen YH, Yuan GX, et al. Evidence of selection at melanin synthesis pathway loci during silkworm domestication. *Mol Biol Evol*. 2011;28(6):1785–99.
11. Ma X, Wang Z, Zhang X. Evolution of dopamine-related systems: biosynthesis, degradation and receptors. *J Mol Evol*. 2010;71(5–6):374–84.
12. Lynch M. *The Origins of Genome Architecture*. Sunderland, MA: Sinauer Associates; 2007.
13. Ohta T. Slightly deleterious mutant substitutions in evolution. *Nature*. 1973;246(5428):96–8.
14. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24(8):1586–91.
15. Wu X, Chi X, Wang P, Zheng D, Ding R, Li Y. The evolutionary rate variation among genes of HOG-signaling pathway in yeast genomes. *Biol Direct*. 2010;5:46.
16. Alvarez-Ponce D. The relationship between the hierarchical position of proteins in the human signal transduction network and their rate of evolution. *BMC Evol Biol*. 2012;12:192.
17. Tu Q, Cameron RA, Worley KC, Gibbs RA, Davidson EH. Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Res*. 2012;22(10):2079–87.
18. Collins JE, White S, Searle SM, Stemple DL. Incorporating RNA-seq data into the zebrafish Ensembl genebuild. *Genome Res*. 2012;22(10):2067–78.
19. Schneider A, Souvorov A, Sabath N, Landan G, Gonnet GH, Graur D. Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biol Evol*. 2009;1:114–8.
20. Majer M, Popov KM, Harris RA, Bogardus C, Prochazka M. Insulin down-regulates pyruvate dehydrogenase kinase (PDK) mRNA: potential mechanism contributing to increased lipid oxidation in insulin-resistant subjects. *Mol Genet Metab*. 1998;65(2):181–6.
21. Huang SM, Hancock MK, Pitman JL, Orth AP, Gekakis N. Negative regulators of insulin signaling revealed in a genome-wide functional screen. *PLoS One*. 2009;4(9):e6871.
22. Giannakou ME, Partridge L. Role of insulin-like signalling in *Drosophila* lifespan. *Trends Biochem Sci*. 2007;32(4):180–8.
23. Oldham S, Hafen E. Insulin/IGF and target of rapamycin signaling: a TOR de force in growth control. *Trends Cell Biol*. 2003;13(2):79–85.