Expression Level Drives the Pattern of Selective Constraints along the Insulin/Tor Signal Transduction Pathway in Caenorhabditis

Richard Jovelin*,[1,2], and Patrick C. Phillips[1]

[1]Department of Biology, Center for Ecology and Evolutionary Biology, University of Oregon

[2]Department of Ecology and Evolutionary Biology, University of Toronto, Ontario, Canada

*Corresponding author: E-mail: richard.jovelin@utoronto.ca.

Accepted: 4 July 2011

Abstract

Genes do not act in isolation but perform their biological functions within genetic pathways that are connected in larger networks. Investigation of nucleotide variation within genetic pathways and networks has shown that topology can affect the rate of protein evolution; however, it remains unclear whether a same pattern of nucleotide variation is expected within functionally similar networks and whether it may be due to similar or different biological mechanisms. We address these questions by investigating nucleotide variation in the context of the structure of the insulin/Tor-signaling pathway in Caenorhabditis, which is well characterized and is functionally conserved across phylogeny. In Drosophila and vertebrates, the rate of protein evolution is negatively correlated with the position of a gene within the insulin/Tor pathway. Similarly, we find that in Caenorhabditis, the rate of amino acid replacement is lower for downstream genes. However, in Caenorhabditis, the rate of synonymous substitution is also strongly affected by the position of a gene in the pathway, and we show that the distribution of selective pressure along the pathway is driven by differential expression level. A full understanding of the effect of pathway structure on selective constraints is therefore likely to require inclusion of specific biological function into more general network models.

Key words: network, aging, molecular evolution, gene expression, selection.

Introduction

Models of evolutionary change, particularly at the molecular level, tend to focus on the effects of mutation, natural selection, and genetic drift operating on genes one at a time. But molecular evolution is actually generated by the manner in which fitness differences at the level of the whole organism are mapped to and from the DNA sequence level via development and physiology. These mapping functions are usually represented in the form of genetic networks. A fundamental question in the field is therefore whether patterns of molecular evolution are best understood by taking the global network context of the genes of interest into account or whether taking a locus-by-locus approach is sufficient. A systematic way of examining the distribution of selective pressure in genetic networks is to investigate the relationship between network structure and nucleotide variation of individual genes interacting within those networks. Large-scale studies in major cellular networks have shown that network topology does indeed affect the rate of protein evolution. Specifically, central and highly connected proteins in the metabolic and protein–protein interaction networks tend to evolve more slowly than proteins at network peripheries (Fraser et al. 2002; Hahn and Kern 2005; Vitkup et al. 2006; Lu et al. 2007). However, central transcription factors tend to evolve faster in the yeast gene regulatory network, despite a similar network topology, suggesting that the distribution of selective pressure within networks depends more specifically on the function of the network under study (Jovelin and Phillips 2009). Thus, it is important to compare the pattern of nucleotide variation in multiple networks in order to ascertain whether or not it is possible to generate general rules for molecular evolution within genetic networks.

Investigation of nucleotide variation in the context of linear metabolic pathways has been particularly informative for addressing this question because the patterns can be interpreted within the context of population genetic models of metabolic flux (Kacser and Burns 1973). Theory predicts that upstream enzymes in linear pathways will evolve greater control over metabolic fluxes and will 1) preferentially fix beneficial mutations during adaptive walks and 2) be under stronger purifying selection as optimal control is reached in the population (Wright and Rausher 2010). Analysis of nucleotide divergence in the anthocyanin pathway in plants supports the prediction that downstream enzymes experience relaxed selection and evolve faster than upstream enzymes (Rausher et al. 1999, 2008; Lu and Rausher 2003). Similar effects of pathway position on nucleotide variation have been shown for the melanin synthesis pathway in silkworms (Yu et al. 2011) as well as the carotenoid (Livingstone and Anderson 2009) and terpenoid (Ramsay et al. 2009) biosynthetic pathways in plants.

It remains unclear, however, whether a similar pattern is to be expected for different types of linear pathways. In signal transduction pathways, for instance, one might expect evolutionary changes to be concentrated in the upstream receptor because receptors interact with the "external" environment, whereas downstream elements might be expected to be under stronger purifying selection because they are located within a more stable cytoplasmic milieu. For example, analysis of the human-signaling network shows that purifying selection increases as a function of cellular localization, from the extracellular space to the nucleus (Cui et al. 2009). However, detailed analyses of various signal transduction pathways in several organisms have revealed heterogeneity in the relationship between the nucleotide rate variation and the position of a gene within the pathway. All potential outcomes appear possible, as these studies have found no relationship to pathway structure (Jovelin et al. 2009), a trend similar to that observed in metabolic pathways in which the most upstream genes evolve more slowly (Riley et al. 2003), and an inverse polarity in which the most downstream genes tend to be more conserved (Alvarez-Ponce et al. 2009, 2010; Wu et al. 2010).

The insulin-signaling (IS) pathway is particularly well suited for addressing this set of questions. It is well characterized and largely functionally conserved across a broad phylogenetic swath including yeast, flies, nematodes, rodents, and humans (Garofalo 2002; Barbieri et al. 2003; Broughton and Partridge 2009). Mutations affecting IS in these organisms have similar phenotypic effects on lifespan, oxidative-stress resistance, lipid storage, and metabolism. The pathway is most famously characterized in *Caenorhabditis elegans*, as mutations in many pathway components can lead to a many-fold increase in lifespan (Kenyon 2005). Here (fig. 1*A*), upon activation of the AAP-1 adaptor subunit and the phosphotidylinositol 3-OH kinase subunit AGE-1 by the insulin/IGF-1 receptor DAF-2, AGE-1 converts PIP_2 into PIP_3 which then recruits the serine/threonine kinases PDK-1, AKT-1, AKT-2, and SGK-1 to the cell membrane. The kinases phosphorylate and control the localization of the transcription factor FOXO/DAF-16 and prevent its entry in the nucleus thereby down-regulating transcription of stress–response genes (fig. 1*A*). Mutations in these factors prevent this phosphorylation and therefore generate constitutive stimulation of the stress–response pathway, which subsequently increases individual lifespan. The IS pathway also interacts with other conserved pathways including the Tor, Ras, JNK, and TGF-β pathways (Shmookler Reis et al. 2009). Investigations of the impact of pathway structure on nucleotide divergence in the insulin/Tor-signaling pathway in Drosophila and vertebrates consistently reveal that they show the reverse-polarity situation, with downstream proteins tending to evolve more slowly than upstream proteins (Alvarez-Ponce et al. 2009, 2010). Is the same pattern of nucleotide variation within functionally conserved pathways expected among distantly related organisms, and if so, will these similar trends be the result of the same or different underlying biological mechanisms?

Here, we examine these questions by investigating nucleotide divergence in Caenorhabditis in the context of the structure of the insulin/Tor-signaling pathway. We find the global pattern of nucleotide substitution to be similar to that seen in Drosophila and vertebrates but that differences in the rate of evolution at synonymous sites suggest that the underlying causes of this pattern are likely to be different. Understanding the relationship between network structure and rates of molecular evolution is therefore likely to depend on a more detailed understanding of the genotype–phenotype map than is revealed by the topology of the genetic network per se.

Materials and Methods

The Insulin/Tor Pathway in *C. elegans*

The pattern of interactions among the *C. elegans* insulin/Tor-signaling pathway genes was obtained from several recent review and research articles (fig. 1*A*) (Jia et al. 2004; Jensen et al. 2006; Mukhopadhyay et al. 2006; Antebi 2007; Braeckman and Vanfleteren 2007; Shaw et al. 2007; Broughton and Partridge 2009; Chen et al. 2009; Shmookler Reis et al. 2009). Within this larger network, we examined nucleotide divergence focusing on proteins that mediate signal transduction in response to stimuli detected by the insulin receptor DAF-2 (fig. 1*B*) in order to investigate how the structure of the insulin/Tor pathway affects protein evolution in nematodes, flies, and vertebrates (Alvarez-Ponce et al. 2009, 2010). We assigned the first position in the pathway to DAF-2 and we counted the number of steps from DAF-2 to determine each protein's position within the pathway (table 1).

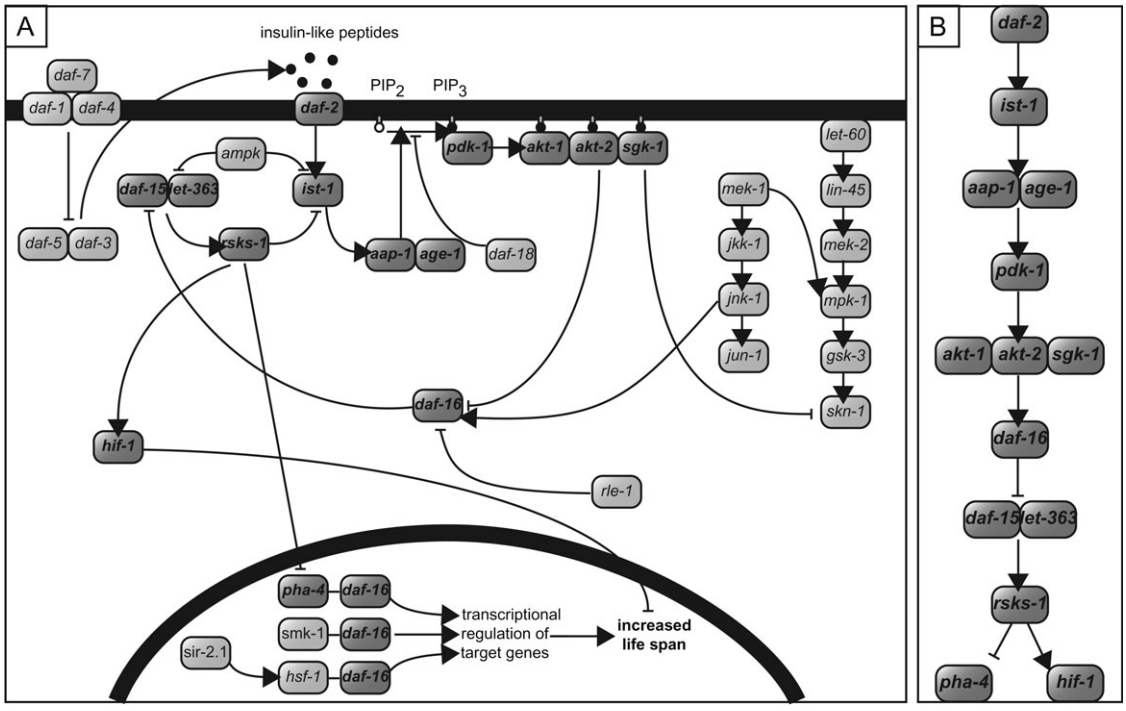

Fig. 1.—(*A*) Activation of the IS pathway negatively regulates the transcription factor DAF-16 by preventing its entry in the nucleus. The IS pathway interacts with multiple signal transduction pathways. Redrawn following (Shmookler Reis et al. 2009), with modifications to include additional interactions specified in Jia et al. (2004) and Chen et al. (2009). (*B*) Linear graph of the insulin/Tor pathway used to investigate the impact of the pathway structure on nucleotide divergence. Here, we focus on proteins that mediate signal transduction in response to stimuli detected by the insulin receptor DAF-2.

Identification of Orthologs and Phylogenetic Analyses

C. elegans orthologs of the insulin/Tor-signaling genes were identified in the *Caenorhabditis briggsae* (Stein et al. 2003), *Caenorhabditis japonica*, *Caenorhabditis brenneri*, *Caenorhabditis remanei*, *Caenorhabditis* sp. 7, *Caenorhabditis* sp. 9, and *Caenorhabditis* sp. 11 (Genome Sequencing Center, Washington University, St. Louis, unpublished data) genome assemblies using the TBlastN program (Altschul et al. 1990). For each gene, we identified only one similar sequence with two exceptions: no clear ortholog of *akt-2* could be identified in any of the genome assemblies and two highly similar sequences of *let-363* were found in

Table 1
Summary of Nucleotide Divergence and Gene Variables Used in the Multivariate Analysis

	Caenorhabditis briggsae–Caenorhabditis sp. 9 Comparisons				*C. briggsae–Caenorhabditis remanei* Comparisons							
Gene	***N***	**d*N***	**d*S***	**ω**	***N***	**d*N***	**d*S***	**ω**	**ENC**	**L**	**Expression Level**	**Position**
daf-2	3909	0.0348	0.3717	0.0936	4656	0.3093	3.4522	0.0896	50.9835	1552	1.99375	1
ist-1	2997	0.0093	0.3073	0.0301	2901	0.1688	2.9045	0.0581	52.5675	967	2.2875	2
aap-1	1566	0.0052	0.1218	0.0428	1539	0.1896	3.2266	0.0588	50.6140	513	NA	3
age-1	3546	0.0030	0.1780	0.0170	3513	0.1867	2.0769	0.0899	50.6090	1171	2.825	3
pdk-1	1863	0.0218	0.1746	0.1248	1854	0.2544	4.3497	0.0585	51.6825	618	4	4
akt-1	1596	0.0099	0.1442	0.0684	1596	0.0856	1.6772	0.0511	50.1770	532	11.25	5
sgk-1	1362	0.0126	0.1216	0.1039	1359	0.0456	1.8861	0.0242	50.7925	453	7.1375	5
daf-16	1584	0.0019	0.0817	0.0226	1557	0.0570	0.7504	0.0759	49.0960	519	12.425	6
daf-15	2148	0.0156	0.3051	0.0510	5325	0.1033	1.9296	0.0535	49.9820	1775	2.725	7
let-363	7617	0.0032	0.2283	0.0142	7734	0.0847	1.5645	0.0541	48.9230	2578	7.475	7
rsks-1	1332	0.0051	0.0982	0.0517	1617	0.0416	1.1936	0.0348	48.6305	539	NA	8
pha-4	1197	0.0183	0.1427	0.1279	1191	0.0956	1.1403	0.0838	48.8010	397	9.5875	9
hif-1	2415	0.0058	0.1508	0.0387	2154	0.0997	0.8055	0.1238	53.9200	718	20.8875	9

Note.—*N* : number of sites analyzed after gaps were removed, ENC: effective number of codons, L: protein length.

C. brenneri, consistent with ~30% of the *C. brenneri* genes having two alleles in the genome assembly (Barrière et al. 2009; Jovelin 2009). We combined exons from the two *Cbn-let-363* alleles because of incomplete sequence for one allele and sequencing errors in the second allele. Intron/exon boundaries were predicted relative the *C. elegans* protein sequence. Some sequences are incomplete due to the fragmented and preliminary nature of the genome assemblies.

Protein sequences were aligned by eye using BioEdit (Hall 1999) and subsequently used to generate codon-based DNA sequence alignments. Sequence alignments are available upon request. Phylogenetic reconstruction for each orthologous group (single-gene analysis) was obtained with MrBayes (Ronquist and Huelsenbeck 2003) applying a JTT model of protein sequence evolution (Jones et al. 1992). Parameter space was searched with four independent runs of 1,000,000 generations each and with four chains. Trees were samples every 100 generations and posterior probabilities were determined after discarding the first 2,500 trees ("burn-in"). We also sought to infer the phylogenetic relationships among the eight Caenorhabditis species using a similar analysis based on concatenated sequences of the insulin/Tor genes. In this analysis, parameter space was searched with four independent runs of 10,000 generations with the same tree sampling frequency and also applying a 25% burn-in. Half of the single-gene analyses (6/13) gave a well supported topology, identical to the one retrieved with the concatenated sequences (fig. 2), and corresponding to the known relationship among the named species (Kiontke et al. 2004; Kiontke and Fitch 2005). Topologies obtained during the other single-gene analyses (7/13) either have poorly supported nodes or conflict with the accepted species tree (Kiontke et al. 2004; Kiontke and Fitch 2005), and only two were identical. In subsequent codon-based tests of selection, we used the topology obtained with the concatenated sequences (fig. 2) and when the single-gene topology was different, we used the topology that best fits the data according to model M0 (see below). Only single-gene topologies for *daf-15*, *pdk-1*, *pha-4*, and *sgk-1* were used after comparing alternative topologies with the likelihood ratio test (LRT).

Codon-Based Sequence Analyses

Maximum likelihood estimates of the rates of nonsynonymous (d*N*) and synonymous (d*S*) substitutions along with the corresponding ratio (ω) were computed between *C. briggsae* and *C. remanei* and between *C. briggsae* and *Caenorhabditis* sp. 9 with the CODEML program in PAML 3.14 (Yang 1997) with model M0, which provides a single estimate of across all sites and lineages. To test for positive selection acting on the insulin/Tor pathway genes, we used an LRT between models M7 and M8 (Yang et al. 2000) and obtained significance of the likelihood ratio statistic $2\Delta l$ by comparison to the χ^2 distribution with 2 degrees of freedom. Models M7 and M8 allow ω to vary among sites according to a beta distribution, estimated with 10 categories, and model M8 has an additional parameter $\omega > 1$. To determine if ω is significantly greater than 1, we compared model M8a in which $\omega = 1$ with model M8 (Swanson et al. 2003) using an LRT with 1 degree of freedom. We also performed an LRT with 1 degree of freedom between models M1a and M2a (Wong et al. 2004). Model M1a allows 2 classes of sites with $\omega < 1$ and $\omega = 1$, respectively, and model M2a has a third class of sites with $\omega > 1$. For all codon-based analyses, ambiguous sites were removed, no molecular clock was assumed, and codon frequencies were estimated from the observed nucleotide frequencies at each codon position.

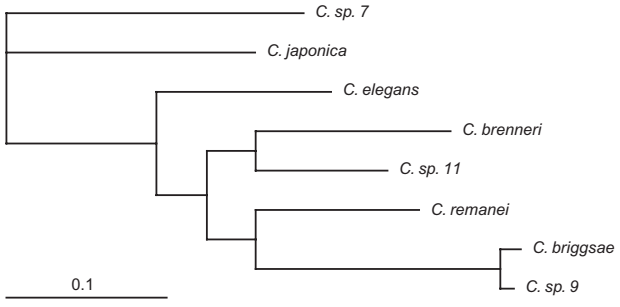

Fig. 2.—Phylogenetic relationships among *Caenorhabditis* species obtained using Bayesian inference with concatenated protein sequences from 13 insulin/Tor-signaling genes. The posterior probability at each node is 1.

Gene-Level Variables

In our investigation of the effect of pathway structure on nucleotide divergence, we also examined the possible confounding effects of several other variables including protein length, codon bias, and expression level. We determined the correlation among variables using Spearman's rank correlation. Protein length is the length of the protein in pairwise alignments between *C. remanei* and *C. briggsae* after gaps were removed. Codon bias is the average effective number of codons (ENC) (Wright 1990) between *C. remanei* and *C. briggsae* orthologs. ENC values were obtained for each orthologous gene with DnaSp 5.10 (Librado and Rozas 2009). Expression level is the average expression measured in *C. elegans* with microarrays at eight time points spanning embryonic development and adulthood (Hill et al. 2000). We also tested the effect of expression level on nucleotide divergence at each separate time point and obtained qualitatively similar results (not shown). Expression level is missing for *aap-1* and *rsks-1*.

Results

The Rate of Nucleotide Substitution Decreases along the Insulin/Tor Pathway

The rate of nonsynonymous changes between *C. briggsae* and *C. remanei* orthologs varies 7.5-fold among the IS genes

(table 1). We tested if the position of a protein within the insulin/Tor pathway could impact the rate of protein sequence evolution and explain the observed variability among IS genes. Variability in nonsynonymous changes is strongly negatively correlated with the position of a protein in the pathway (Spearman's $\rho = -0.638$, $P = 0.018$). The more downstream proteins tend to evolve more slowly (fig. 3), similar to the pattern observed along the insulin/Tor pathway in Drosophila (Alvarez-Ponce et al. 2009) and in vertebrates (Alvarez-Ponce et al. 2010).

This pattern of natural selection could result either from a tendency of the upstream genes to be under positive selection or from increasing purifying selection operating along the pathway. To test between these alternatives, we performed codon-based tests of selection using orthologs from eight Caenorhabditis species. A first LRT favored model M8 (Yang et al. 2000), a model allowing a proportion of sites to evolve under positive selection, for 7 of the 13 genes. Nevertheless, a second LRT showed that ω for this class of sites is significantly greater than 1 for only 1 gene, *pha-4*, located downstream in the pathway. Moreover, comparisons of nearly neutral and positive selection models M1a and M2a (Wong et al. 2004) failed to detect any instance of positive selection (table 2). These results suggest that the observed pattern of variation along the insulin pathway is unlikely to result predominantly from positive selection acting on the upstream genes but may instead be accounted for by increasing levels of purifying selection.

If mutations in the IS genes have different pleiotropic effects, causing selective constraints to be distributed along the insulin/Tor pathway, then we would expect this effect to manifest itself on d*N* and ω but not on d*S*. However, d*S* is not randomly distributed along the insulin pathway (fig. 3) but instead strongly correlates negatively with the position of a protein in the pathway (Spearman's $\rho = -0.815$, $P < 0.001$). In contrast, ω is not correlated (Spearman's $\rho = -0.088$, $P = 0.774$). The strong polarity of d*S* along the pathway and the lack of correlation for ω further suggest that positive selection is unlikely to be a major determinant in the distribution of nucleotide rate divergence and point to increasing levels of purifying selection affecting both d*N* and d*S*. Moreover, pleiotropic constraints due to the cumulative effect of mutations in the insulin/Tor and/or interacting

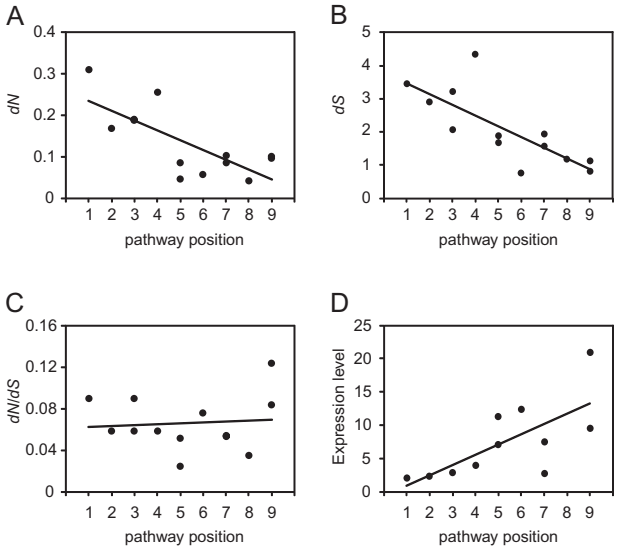

Fig. 3.—Nucleotide substitution is strongly affected by the position of a gene in the insulin/Tor pathway because of increasing purifying selection associated to expression level differences. The rates of nonsynonymous (*A*) and synonymous (*B*) changes are negatively correlated with the position of a gene in the insulin/TOR pathway. However, there is no correlation between and pathway position (*C*). Expression level is strongly correlated with pathway structure and downstream genes tend to be expressed at higher levels (*D*).

Table 2
Results of codon-based tests of selection

	lnL (M7)	lnL (M8)	2Δ*l* (M7–M8)	lnL (M8a)	2Δ*l* (M8a–M8)	lnL (M1a)	lnL (M2a)	2Δ*l* (M1a–M2a)
daf-2	−18114.3	−18109.4	9.8**	−18109.2	−0.4	−18270.7	−18270.7	0
ist-1	−9983.3	−9983.3	0	−9982.6	−1.4	−10102.3	−10102.3	0
aap-1	−8665	−8663.8	2.4	−8663.1	−1.4	−8755.8	−8755.8	0
age-1	−16719.2	−16716.3	5.8	−16716.3	0	−16877	−16877	0
pdk-1	−9340.4	−9340.4	0	−9340.3	−0.2	−9484.1	−9484.1	0
akt-1	−7645.6	−7641.4	8.4*	−7641.4	0	−7720.9	−7720.9	0
sgk-1	−5581.9	−5581.1	1.6	−5581.1	0	−5634.1	−5634.1	0
daf-16	−3889.9	−3888.7	2.4	−3888.7	0	−3916.17	−3916.17	0
daf-15	−9030.7	−9025.9	9.6*	−9025.9	0	−9114	−9114	0
let-363	−26832.9	−26824.8	16.2***	−26824.8	0	−27076.8	−27076.8	0
rsks-1	−5170	−5170	0	−5170	0	−5194.4	−5194.4	0
pha-4	−5293.9	−5289.2	9.4**	−5293.1	7.8*	−5356.7	−5356.7	0
hif-1	−8984.8	−8980.9	7.8*	−8980.9	0	−9020.1	−9020.1	0

LRTs were performed between models M7 and M8, models M8 and M8a and between models M1a and M2a. The LRTs do not show evidence of rampant positive selection acting on the IS genes. With the exception of *pha-4*, ω is not significantly greater than 1 for genes with the best fitting model allowing a proportion of site to evolve under position selection. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

pathways cannot be the sole cause underlying this pattern of selection because such selective constraints would be expected to affect protein sequence evolution only.

In Drosophila and vertebrates, the structure of the insulin/Tor pathway affects the rate of protein sequence evolution but does not seem to have much impact on d*S* (Alvarez-Ponce et al. 2009, 2010). *C. briggsae* and *C. remanei* have diverged long enough such that saturation at synonymous sites could complicate inferences based on d*S* and ω. Therefore, we also estimated nucleotide divergence from another species pair, *C. briggsae* and *Caenorhabditis* sp. 9. These two species split in the recent past, as reflected by their ability to produce fertile hybrid progeny (Woodruff et al. 2010), and do not show saturation at synonymous sites (table 1). The architecture of the insulin pathway has the same effects using these estimates of nucleotide divergence. Both d*N* (Spearman's $\rho = -0.149$, $P = 0.627$) and d*S* (Spearman's $\rho = -0.423$, $P = 0.150$) are negatively correlated with pathway position whereas is not correlated (Spearman's $\rho = 0.066$, $P = 0.830$). Although the correlations are not significant, they are qualitatively similar to those obtained using *C. briggsae* and *C. remanei*. The lack of significance is presumably due to the short divergence time between *C. briggsae* and *Caenorhabditis* sp. 9 and subsequently the lower variance in nucleotide substitution rate among IS genes in this species pair. For instance, the standard deviations for d*N* and d*S* are, respectively, 8 and 12 times lower among *C. briggsae* and *Caenorhabditis* sp. 9 orthlogs than among *C. briggsae* and *C. remanei* orthologs. Overall, then, the effect of pathway position on sequence evolution is not dependent upon the choice of species used to compute nucleotide rate divergence, and nucleotide variation is distributed differently along the insulin pathway in nematodes, flies, and vertebrates.

The Pattern of Selective Constraints along the Insulin/Tor Pathway Is Driven by Differential Expression Level

Because pleiotropic constraint cannot account for the observed pattern of selection along the insulin pathway, we tested if this pattern could be due to the distribution of one or several variables affecting purifying selection (Rocha 2006). Specifically, we tested if the position in the pathway is correlated with expression level, codon bias, and protein length. Protein length (Spearman's $\rho = -0.188$, $P = 0.539$) and codon bias (Spearman's $\rho = -0.505$, $P = 0.078$) do not correlate significantly with pathway position (although the latter is nonetheless fairly negative). However, the level of gene expression in *C. elegans* is strongly correlated with the position that a gene occupies in the pathway (fig. 3*D*; Spearman's $\rho = 0.714$, $P = 0.0136$). Rates of synonymous and nonsynonymous changes are reduced in highly expressed genes, respectively, because of selection for translational accuracy and selection for translational robustness (Drummond et al. 2005). Using *C. elegans* expression as a proxy for other species, rates of nucleotide divergence are highly dependent upon expression level for the IS genes, whether they are measured using *C. briggsae* and *C. remanei* (d*N*: Spearman's $\rho = -0.664$, $P = 0.026$; d*S*: Spearman's $\rho = -0.864$, $P < 0.001$) or *C. briggsae* and *Caenorhabditis* sp. 9 (d*N*: Spearman's $\rho = -0.454$, $P = 0.160$; d*S*: Spearman's $\rho = -0.791$, $P = 0.004$).

Although variation in expression level provides a strong predictor of evolutionary rate in this system, each of the functional variables are correlated with one another, making it impossible to completely isolate their effects. We did not find a significant residual correlation between pathway position and nucleotide divergence after removing the effect of expression level (d*N*: Spearman's $\rho = -0.467$, $P = 0.148$; d*S*: Spearman's $\rho = -0.471$, $P = 0.143$), although the pattern is still negative, and the limited number of pathway components seriously limits the power of a multivariate analysis. Nevertheless, the high dependence of d*S* on expression level and pathway position, which in particular is not expected to be a direct effect of pathway structure, and the absence of correlation between ω and pathway position both strongly argue that expression level has a large influence on evolutionary rates and that the contribution of pathway structure on protein evolution, if any, is likely to be small relative to the effect of expression level.

Altogether, these results suggest that the pattern of nucleotide variation along the insulin pathway is driven by gene expression level rather than by pleiotropic constraints or increased purifying selection in the downstream elements in relation to their function in mediating the cellular response. Moreover, these results point to the need to consider possible confounding variables when analyzing the relationship between pathway structure and evolutionary rates. Importantly, different factors explain the polarity of purifying selection along the well-conserved insulin/Tor pathway in nematodes, flies, and vertebrates (Alvarez-Ponce et al. 2009, 2010), suggesting that the effect of the position of gene within a pathway on protein sequence evolution is largely specific to the broader functional context of the pathway under study.

Discussion

We found that the rate of amino acid replacements correlates negatively with the position of a protein in the insulin/Tor-signaling pathway in Caenorhabditis (fig. 3), similar to the pattern observed in several other signal transduction pathways including the insulin/Tor pathway in Drosophila and vertebrates (Alvarez-Ponce et al. 2009, 2010; Cui et al. 2009; Wu et al. 2010). One possible explanation is that evolutionary changes may preferentially localize to the receptor in signal transduction pathways because such

changes have the potential to affect the entire system and/or that more downstream components may be under stronger purifying selection because they are required to transduce the signal in the cell. Interestingly, the insulin-like receptor DAF-2 is the most divergent protein in the pathway (table 1 and fig. 3). It is conceivable that modifications at the ligand–receptor interaction may reflect adaptation to changing conditions and may play a role in the distribution of selective constraints in signal transduction pathways. Consistent with this interpretation, recent analyses found evidence of adaptive evolution in the insulin receptor and some of its ligands in Drosophila (Guirao-Rico and Aguade 2009, 2011) and a global analysis of human-signaling pathways shows that purifying selection increases from the extracellular space to the nucleus (Cui et al. 2009).

However, a major difference among nematodes, flies, and vertebrates is that selection on synonymous changes is also distributed along the insulin/Tor pathway in worms, while ω does not vary with pathway position (fig. 3). Moreover, in Caenorhabditis, the distribution of purifying selection along the pathway seems to be predominantly the result of the tendency of downstream genes to be expressed at higher levels. The high dependence of both d*N* and d*S* on expression level is consistent with the translational robustness and accuracy hypotheses (Drummond et al. 2005) and is not easily reconciled by the hypothesis of selection pressure acting to maintain the function of downstream signal transduction elements or with pleiotropic constraints in relation to the interaction with other pathways. In Drosophila and vertebrates, the pattern of selection on protein evolution remains after correcting for the effect of gene expression and codon usage (Alvarez-Ponce et al. 2009, 2010). Thus, our results clearly show that the pattern of nucleotide variation can differ among functionally conserved pathways and that the underlying biological causes can also be different.

Previously, we found that the relationship between nucleotide variation and network topology can also be different in large cellular networks sharing similar topological properties (Jovelin and Phillips 2009). These results and those presented here strongly suggest that any rules that govern the evolution of interacting proteins, if they exist, are unlikely to be functions of network structure per se. It is therefore somewhat remarkable that investigations of nucleotide variation in linear metabolic pathways in various organisms do in fact tend to find the same pattern of polarity of selective pressure (Rausher et al. 1999, 2008; Lu and Rausher 2003; Livingstone and Anderson 2009; Ramsay et al. 2009; Yu et al. 2011). In this case, an explicit population genetic model grounded in metabolic flux control theory is available to help predict the effect of the position of a gene on protein evolution (Wright and Rausher 2010). Similarly, predictions for the pattern of selection acting on branch point enzymes in metabolic networks based on metabolic flux control (Eanes 2011) have thus far been supported by the available data (Whitt et al. 2002; Flowers et al. 2007; Greenberg et al. 2008; Yang et al. 2009). The challenge ahead is to build explicit functional models for different types of pathways (including signal transduction pathways), similar to those developed for metabolic pathways (Wright and Rausher 2010), in order to understand the conditions under which natural selection may operate within pathways and networks. This is the framework required to move from anecdotal reports of nucleotide variation among interacting genes to a more predictive network-centered view of natural selection (Wilkins 2007).

Acknowledgments

We thank the Genome Institute at Washington University School of Medicine in St. Louis for releasing unpublished genome sequences and providing the research community with this tremendous resource. This work was supported by grants from the National Science Foundation (DEB-0710378) and the Ontario Ministry of Research and Innovation to R.J. and by grants from the National Institutes of Health (R03-AG029377 and P01-AG022500 to P.C.P.).

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Alvarez-Ponce D, Aguade M, Rozas J. 2009. Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 Drosophila genomes. Genome Res. 19:234–242.

Alvarez-Ponce D, Aguade M, Rozas J. 2010. Comparative genomics of the vertebrate insulin/TOR signal transduction pathway: a network-level analysis of selective pressures. Genome Biol Evol. 3:87–101.

Antebi A. 2007. Genetics of aging in *Caenorhabditis elegans*. PLoS Genet. 3:1565–1571.

Barbieri M, Bonafe M, Franceschi C, Paolisso G. 2003. Insulin/IGF-I-signaling pathway: an evolutionarily conserved mechanism of longevity from yeast to humans. Am J Physiol Endocrinol Metab. 285:E1064–E1071.

Barrière A, et al. 2009. Detecting heterozygosity in shotgun genome assemblies: lessons from obligately outcrossing nematodes. Genome Res. 19:470–480.

Braeckman BP, Vanfleteren JR. 2007. Genetic control of longevity in *C. elegans*. Exp Gerontol. 42:90–98.

Broughton S, Partridge L. 2009. Insulin/IGF-like signalling, the central nervous system and aging. Biochem J. 418:1–12.

Chen D, Thomas EL, Kapahi P. 2009. HIF-1 modulates dietary restriction-mediated lifespan extension via IRE-1 in Caenorhabditis elegans. PLoS Genet. 5:e1000486.

Cui Q, Purisima EO, Wang E. 2009. Protein evolution on a human signaling network. BMC Syst Biol. 3:21.

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A. 102:14338–14343.

Eanes WF. 2011. Molecular population genetics and selection in the glycolytic pathway. J Exp Biol. 214:165–171.

Flowers JM, et al. 2007. Adaptive evolution of metabolic pathways in Drosophila. Mol Biol Evol. 24:1347–1354.

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the protein interaction network. Science 296:750–752.
Garofalo RS. 2002. Genetic analysis of insulin signaling in Drosophila. Trends Endocrinol Metab. 13:156–162.
Greenberg AJ, Stockwell SR, Clark AG. 2008. Evolutionary constraint and adaptation in the metabolic network of Drosophila. Mol Biol Evol. 25:2537–2546.
Guirao-Rico S, Aguade M. 2009. Positive selection has driven the evolution of the Drosophila insulin-like receptor (InR) at different timescales. Mol Biol Evol. 26:1723–1732.
Guirao-Rico S, Aguade M. 2011. Molecular evolution of the ligands of the insulin-signaling pathway: dilp genes in the genus Drosophila. Mol Biol Evol. 28:1557–1560.
Hahn MW, Kern AD. 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol. 22:803–806.
Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 41:95–98.
Hill AA, Hunter CP, Tsung BT, Tucker-Kellogg G, Brown EL. 2000. Genomic analysis of gene expression in *C. elegans*. Science 290:809–812.
Jensen VL, Gallo M, Riddle DL. 2006. Targets of DAF-16 involved in *Caenorhabditis elegans* adult longevity and dauer formation. Exp Gerontol. 41:922–927.
Jia K, Chen D, Riddle DL. 2004. The TOR pathway interacts with the insulin signaling pathway to regulate *C. elegans* larval development, metabolism and life span. Development 131:3897–3906.
Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8:275–282.
Jovelin R. 2009. Rapid sequence evolution of transcription factors controlling neuron differentiation in Caenorhabditis. Mol Biol Evol. 26:2373–2386.
Jovelin R, Dunham JP, Sung FS, Phillips PC. 2009. High nucleotide divergence in developmental regulatory genes contrasts with the structural elements of olfactory pathways in Caenorhabditis. Genetics 181:1387–1397.
Jovelin R, Phillips PC. 2009. Evolutionary rates and centrality in the yeast gene regulatory network. Genome Biol. 10:R35.
Kacser H, Burns JA. 1973. The control of flux. Symp Soc Exp Biol. 27:65–104.
Kenyon C. 2005. The plasticity of aging: insights from long-lived mutants Cell. 120:449–460.
Kiontke K, Fitch DH. 2005. The phylogenetic relationships of Caenorhabditis and other rhabditids. In Wormbook, editor. The C. elegans Research Community. WormBook, Available from: http://www.wormbook.org.
Kiontke K, et al. 2004. Caenorhabditis phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc Natl Acad Sci U S A. 101:9003–9008.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.
Livingstone K, Anderson S. 2009. Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. J Hered. 100:754–761.
Lu C, Zhang Z, Leach L, Kearsey MJ, Luo ZW. 2007. Impacts of yeast metabolic network structure on enzyme evolution. Genome Biol. 8:407.
Lu Y, Rausher MD. 2003. Evolutionary rate variation in anthocyanin pathway genes. Mol Biol Evol. 20:1844–1853.
Mukhopadhyay A, Oh SW, Tissenbaum HA. 2006. Worming pathways to and from DAF-16/FOXO. Exp Gerontol. 41:928–934.
Ramsay H, Rieseberg LH, Ritland K. 2009. The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. Mol Biol Evol. 26:1045–1053.
Rausher MD, Lu Y, Meyer K. 2008. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. J Mol Evol. 67:137–144.
Rausher MD, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol Biol Evol. 16:266–274.
Riley RM, Jin W, Gibson G. 2003. Contrasting selection pressures on components of the Ras-mediated signal transduction pathway in Drosophila. Mol Ecol. 12:1315–1323.
Rocha EP. 2006. The quest for the universals of protein evolution. Trends Genet. 22:412–416.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.
Shaw WM, Luo S, Landis J, Ashraf J, Murphy CT. 2007. The *C. elegans* TGF-beta Dauer pathway regulates longevity via insulin signaling. Curr Biol. 17:1635–1645.
Shmookler Reis RJ, Bharill P, Tazearslan C, Ayyadevara S. 2009. Extreme-longevity mutations orchestrate silencing of multiple signaling pathways. Biochim Biophys Acta. 1790:1075–1083.
Stein L, et al. 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. 1:166–192.
Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol. 20:18–20.
Vitkup D, Kharchenko P, Wagner A. 2006. Influence of metabolic network structure and function on enzyme evolution. Genome Biol. 7:R39.
Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ESt. 2002. Genetic diversity and selection in the maize starch pathway. Proc Natl Acad Sci U S A. 99:12959–12962.
Wilkins AS. 2007. Between "design" and "bricolage": genetic networks, levels of selection, and adaptive evolution. Proc Natl Acad Sci U S A. 104(Suppl 1):8590–8596.
Wong WS, Yang Z, Goldman N, Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168:1041–1051.
Woodruff GC, Eke O, Baird SE, Felix MA, Haag ES. 2010. Insights into species divergence and the evolution of hermaphroditism from fertile interspecies hybrids of Caenorhabditis nematodes. Genetics 186:997–1012.
Wright F. 1990. The 'effective number of codons' used in a gene. Gene 87:23–29.
Wright KM, Rausher MD. 2010. The evolution of control and distribution of adaptive mutations in a metabolic pathway. Genetics 184:483–502.
Wu X, et al. 2010. The evolutionary rate variation among genes of HOG-signaling pathway in yeast genomes. Biol Direct. 5:46.
Yang YH, Zhang FM, Ge S. 2009. Evolutionary rate patterns of the Gibberellin pathway genes. BMC Evol Biol. 9:206.
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556.
Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449.
Yu HS, et al. 2011. Evidence of selection at melanin synthesis pathway loci during silkworm domestication. Mol Biol Evol. 28:1785–1799.

Associate editor: Michael Purugganan