Comparative Analysis of Node Degree on Gene Evolution in the Insulin Signaling Pathway

Alyssa Koehler, Annie Backlund, Abigail Myers, Chinmay P. Rele, Laura K. Reed

BACKGROUND

- A biological pathway is a network of nodes (genes) interacting with each other and other regulating molecules which determines gene expression and therefore overall gene function of the cell
- Node degree refers to the number of connections a gene has to other genes.
- The insulin signaling (INS) pathway is critical for regulating blood glucose levels, cell proliferation, cell survival, and development in *Drosophila* and humans.
- According to Alvarez-Ponce *et al.* 2009, genes located at the top of the INS pathway undergo less evolutionary constraint than genes at the bottom.
- Recent studies (Alverez-Ponce et al. 2017) show that genes with high node degree are under more evolutionary constraint.
- We are conducting a small-scale analysis on the ratio of nonsynonymous to synonymous mutations (dN/dS) of three genes with varying node degree as a network instead of a linear pathway.
- These genes are *Dsor1*, *raptor*, and *GlyS*, which have node degrees of 77, 12, and 3, respectively.
- **Hypothesis: there will be a negative correlation between node degree and gene evolution (shown by dN/dS) in the INS pathway.**
- These genes were annotated and reconciled by students in the Genomics Education Partnership.
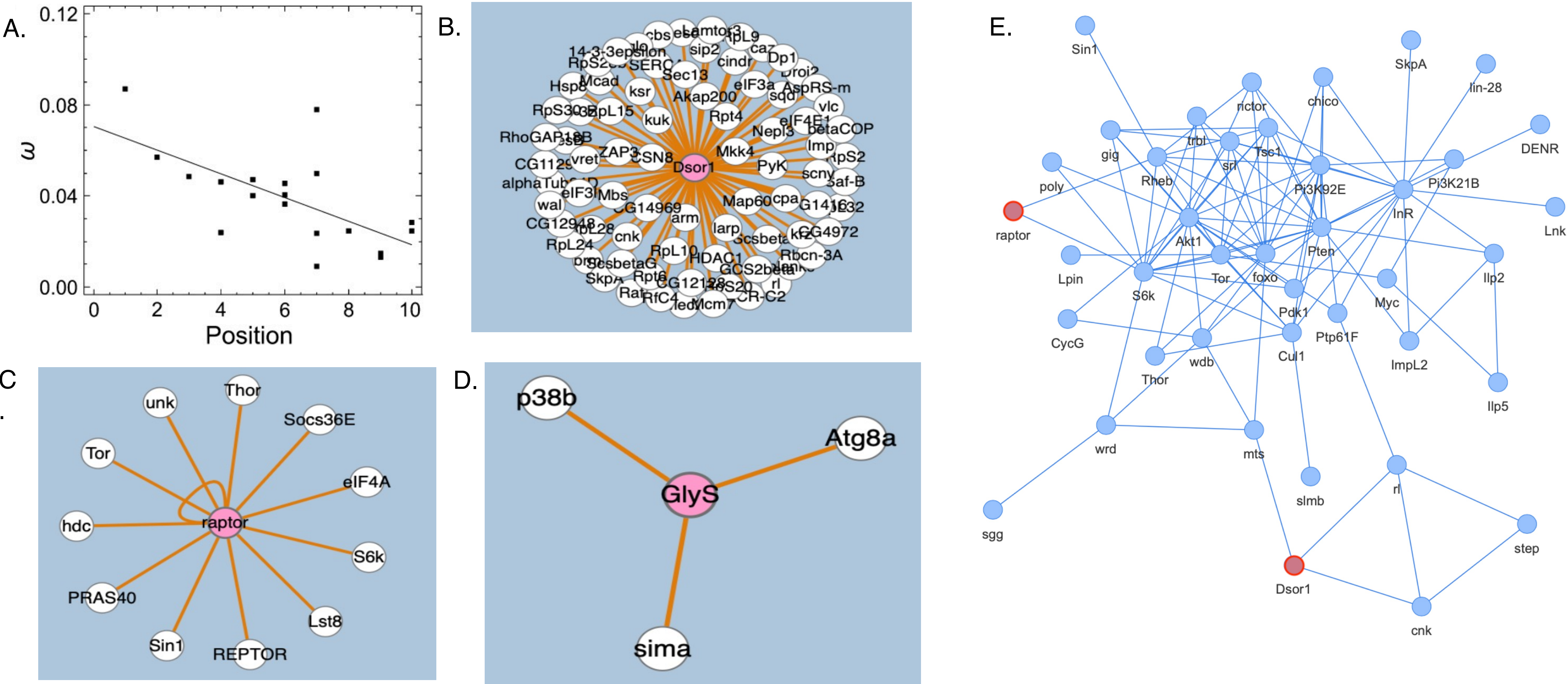
Figure 1

Figure 1 A. The y-axis is dN/dS (ω) = ratio of nonsynonymous to synonymous substitutions and the x-axis is position in the pathway. Each point is a gene in the insulin signaling pathway (Alvarez-Ponce *et al.* 2009). **B.** Overall node degree of *Dsor1* (esyN network diagram, FlyBase). **C.** Overall node degree of *raptor* (esyN network diagram, FlyBase). **D.** Overall node degree of *GlyS* (esyN network diagram, FlyBase). **E.** Connectivity of *Dsor1* and *raptor* (shown in red) in the Insulin Signaling Pathway (Rele C.P. 2022, GitHub).

METHODS

- *Dsor1, GlyS,* and *raptor* gene models were hand-annotated by two or more students for several *Drosophila* species by students participating in the GEP.
- These student models were then reconciled into a final gene model.
- A *Clustal Omega* multiple sequence alignment was performed with sequence data from each target species and D. melanogaster in the UGene app.
- PAML's *codeml* was used to calculate dN/dS in a sliding window and for the entire gene.
  - The sliding window is the dN/dS of every 10 amino acids calculated then graphed according to amino acid position.
  - The dN/dS of the entire gene examines the complete amino acid sequence.
- dN/dS was plotted to identify regions experiencing evolutionary constraint.

DISCUSSION

- *raptor* has the most regions with a high dN/dS, suggesting that it is under the least evolutionary constraint.
- *GlyS* only has one region with a slightly high dN/dS and *Dsor1* only has two regions with a slightly high dN/dS, suggesting that they are under more evolutionary constraint.
- These results do not align with our hypothesis as there is **no correlation between node degree and gene evolution between these genes.** This may be due to only using three genes in our analysis.
- Our analysis does show the specific regions of each gene that are undergoing rapid evolution (shown by high dN/dS) and how they overlap with functional domains of the protein.
- High dN/dS in functional protein domains may alter protein function by altering the amino acid sequence.
- *raptor* has many regions undergoing rapid evolution. Peaks in the dN/dS can be seen in the N-terminal CASPase-like domain and within the WD40 repeats.
- *Dsor1* shows one small dN/dS peak within the protein kinase domain and another one outside the protein domain.
- *GlyS* shows one small dN/dS peak within its glycogen synthase domain.
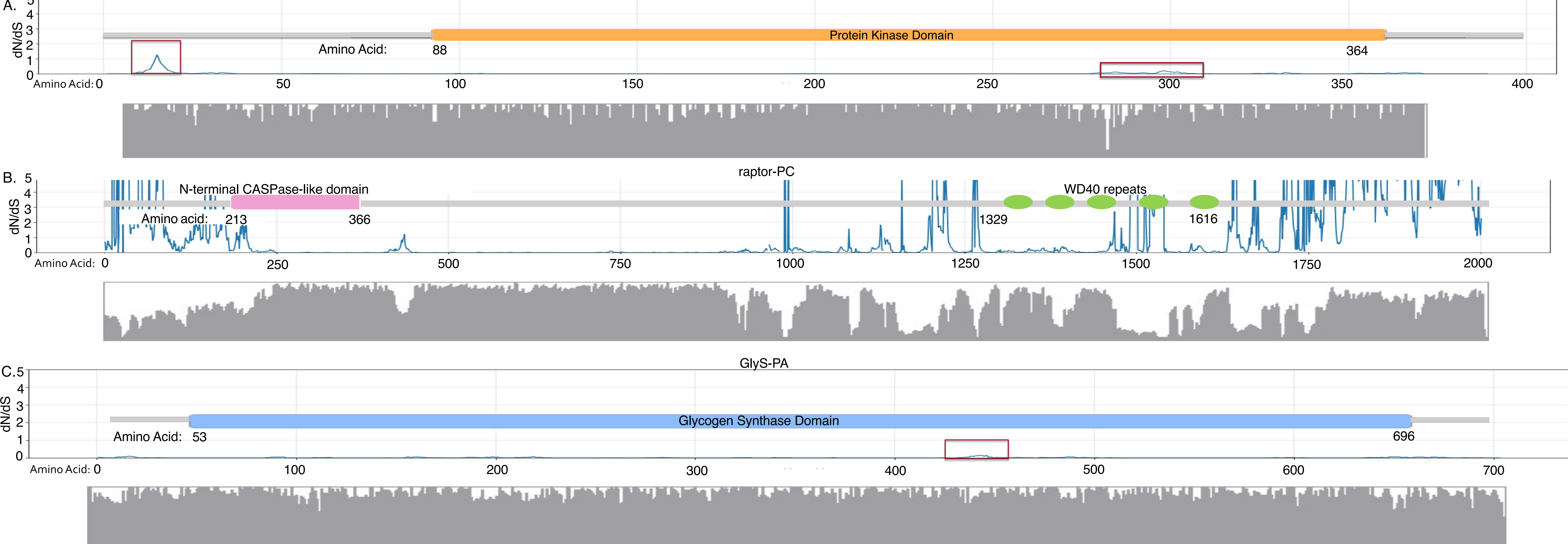
RESULTS

Figure 2

Figure 2: A. Dsor1-PA multiple sequence alignment and functional protein domain. **B.** raptor-PC multiple sequence alignment and functional protein domain. **C.** GlyS-PA multiple sequence alignment and functional protein domain. The top plot for each gene shows the dN/dS on the y-axis and the amino acid position on the x-axis. Peaks in these plots show high rates of evolution via a high ratio of nonsynonymous to synonymous mutations in that region. Each dN/dS plot has the protein domain of the corresponding gene overlapping it (SMART protein domains, FlyBase). These protein domain diagrams align with the amino acid sequence in the dN/dS plot. The bottom plot for each gene shows the amino acid conservation across *Drosophila* species. Gray peaks correspond to amino acid similarity and white valleys correspond to amino acid difference. The deeper the valley the more difference in amino acid sequence in that region across species.

FUTURE DIRECTIONS

64 genes of differing node degree in the insulin signaling pathway are being annotated across 27 *Drosophila* species. These models will have dN/dS is compared between them, showing more information on the impact of node degree on gene evolution.

REFERENCES

- Alvarez-Ponce D., Aguadé M., Rozas J. Network-level molecular evolutionary analyses of the insulin/TOR signal transduction pathway across 12 Drosophila genomes. Genome Res. 2009;19:234–242. doi: 10.1101/gr.084038.108.
- Alvarez-Ponce et al. (2017) Position Matters: Network Centrality Considerably Impacts Rates of Protein Evolution in the Human Protein–Protein Interaction Network, Genome Biology and Evolution, 9;(6):1742–1756
- Rele, C.P., Sandlin, K. M., Leung, W., Reed, L. K. Manual Annotation of Genes within Drosophila Species: the Genomics Education Partnership protocol. bioRxiv 2020.12.10.420521; doi: https://doi.org/10.1101/2020.12.10.420521
- Rele, C. P. omega_analysis. (2021), GitHub Repository, https://github.com/drosophomics/omega_analysis
- Rele, C. P. (2022) Insulin Pathway Evolution. GitHub https://github.com/drosophomics/ins_pathway_evolution
- Hara, K., Maruki, Y., Long, X., Yoshino, K.-I., Oshiro, N., Hidayat, S., Tokunaga, C., Avruch, J. and Yonezawa, K., 2002. Raptor, a Binding Partner of Target of Rapamycin (TOR), Mediates TOR Action. Cell 110, 177–189.
- Takahara, T. and Maeda, T., 2013. Evolutionarily conserved regulation of TOR signaling. The Journal of Biochemistry 154, 1-10. PMID: 23698095